QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.4.1

Step-by-step example for building QSAR model

Outlook

- **Background**
- Keywords
- 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.

Note: Please note that building of custom items (such as profilers, (Q)SAR models as well as importing of custom databases) is only enabled in single user mode.
So, if your Toolbox is installed in multiuser mode, you will be not able to follow this tutorial.

Outlook

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

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. Organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – "group" of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row

(Q)SAR - (Q)SAR models can be used to fill a data gap if no adequate analogues are found for a target chemical

Outlook

- Background
- Keywords
- **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
- Keywords
- 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 target endpoint: 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
- Keywords
- Objectives
- The exercise
- **Workflow of the exercise**

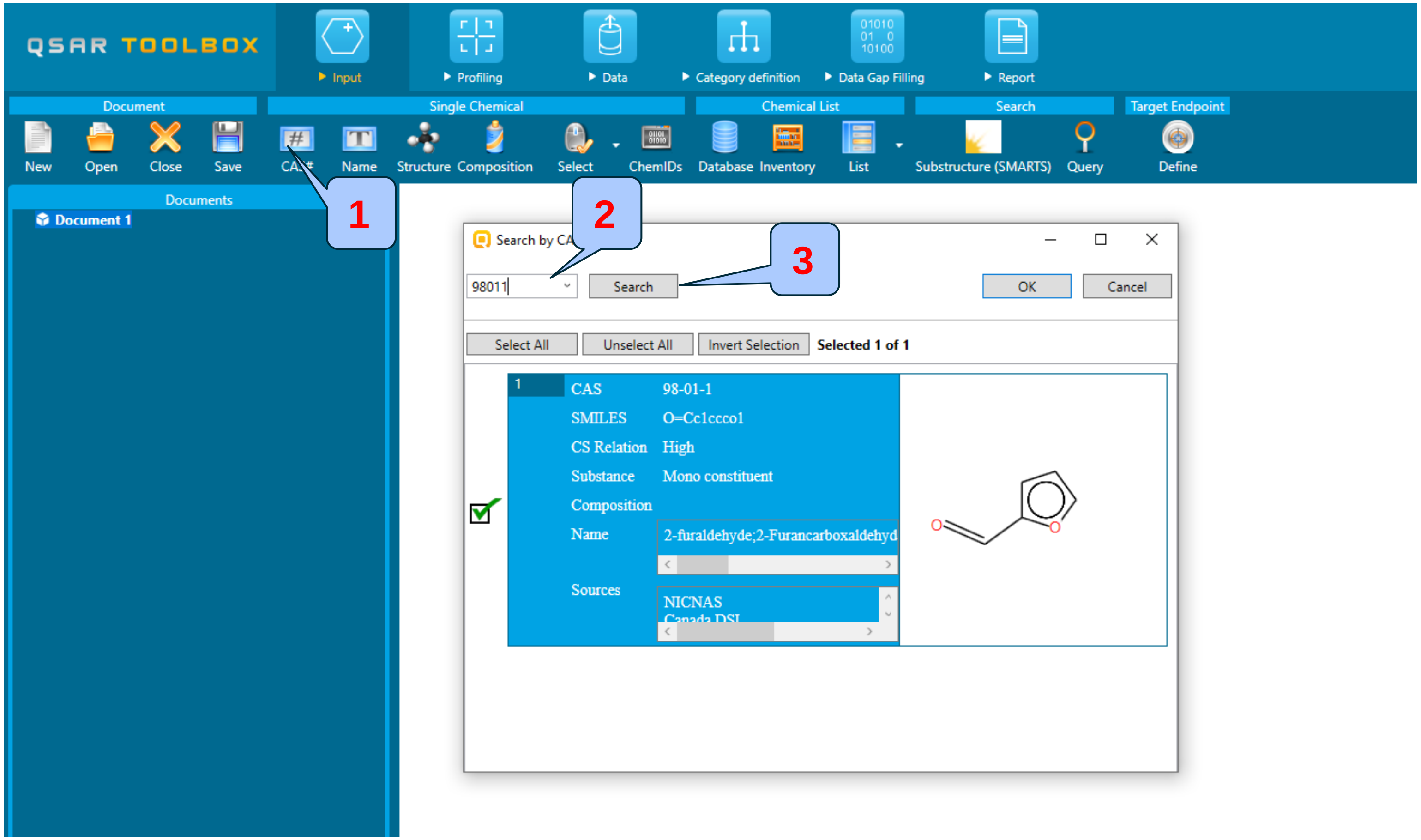
Workflow of the exercise

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

Outlook

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

Input



98011

Selected 1 of 1

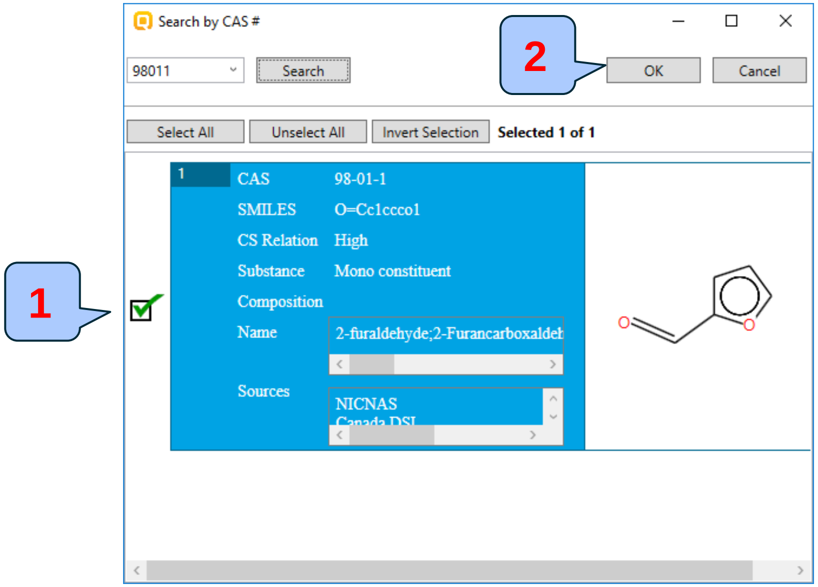
1

CAS	98-01-1
SMILES	O=Cc1ccco1
CS Relation	High
Substance	Mono constituent
Composition	
Name	2-furaldehyde;2-Furancarboxaldehyd
Sources	NICNAS Canada DSL

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.



CAS	98-01-1
SMILES	O=Cc1ccco1
CS Relation	High
Substance	Mono constituent
Composition	
Name	2-furaldehyde;2-Furancarboxaldeh
Sources	NICNAS Canada DSL

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;

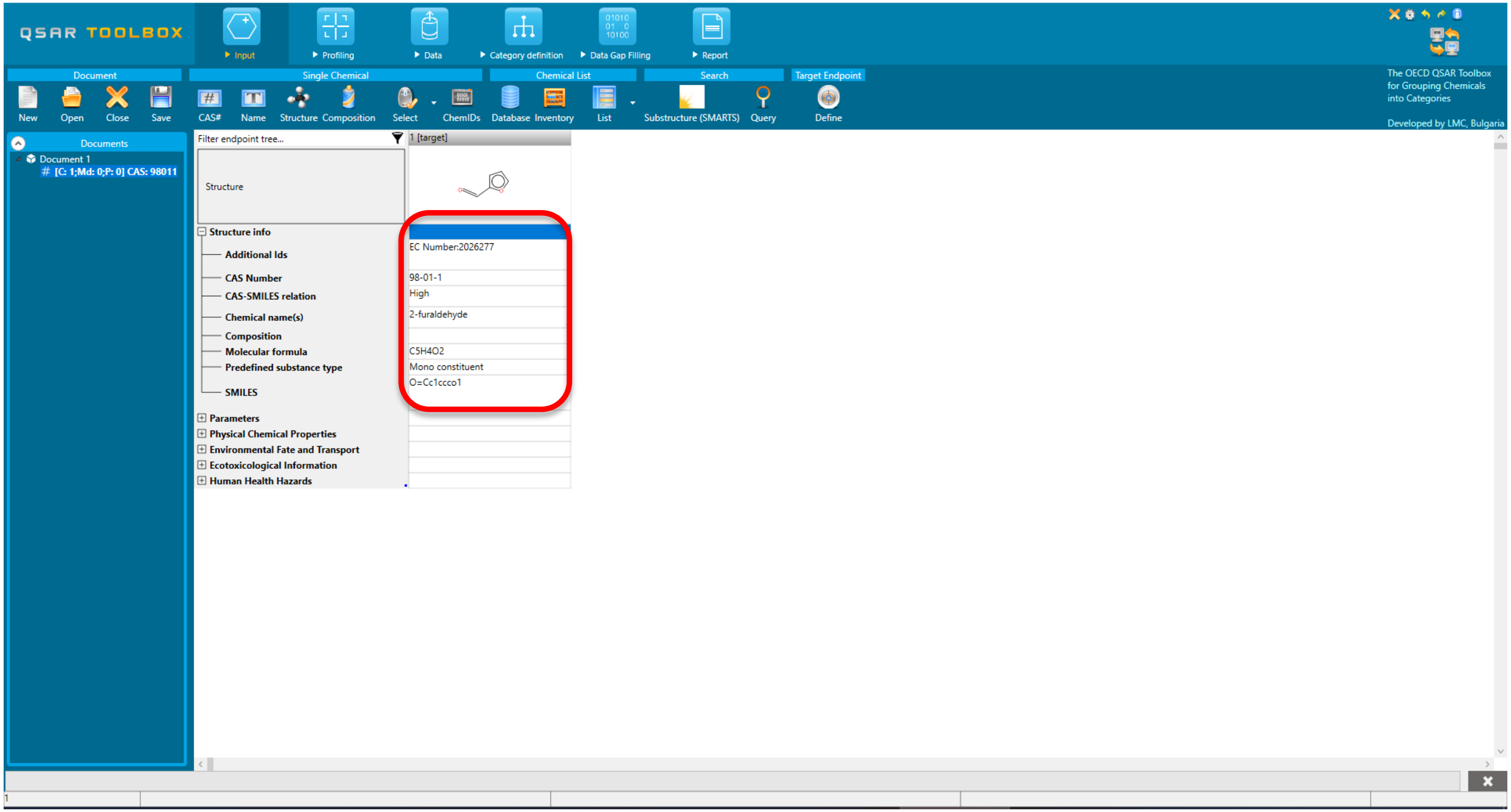
Input
Target chemical identity

- 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



Filter endpoint tree...	1 [target]
Structure	
Structure info	
Additional Ids	EC Number:2026277
CAS Number	98-01-1
CAS-SMILES relation	High
Chemical name(s)	2-furaldehyde
Composition	
Molecular formula	C5H4O2
Predefined substance type	Mono constituent
SMILES	O=Cc1ccco1
Parameters	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow of the exercise**
  - **Input**
    - **Define Target Endpoint**

Input
Define Target Endpoint

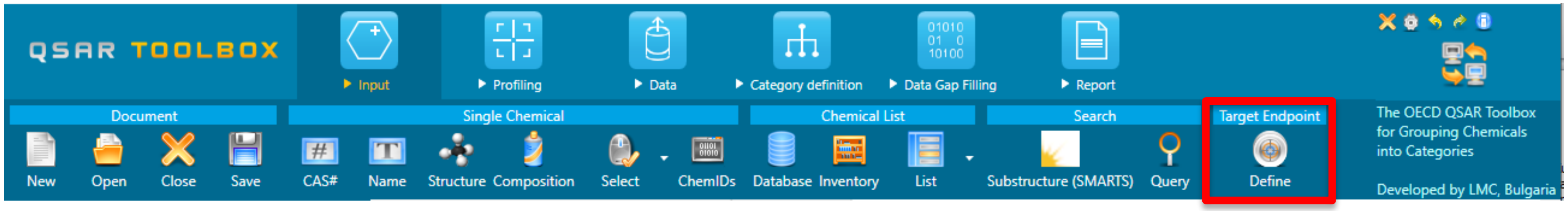
- In this exercise we will build a QSAR model to estimate the following endpoint:

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

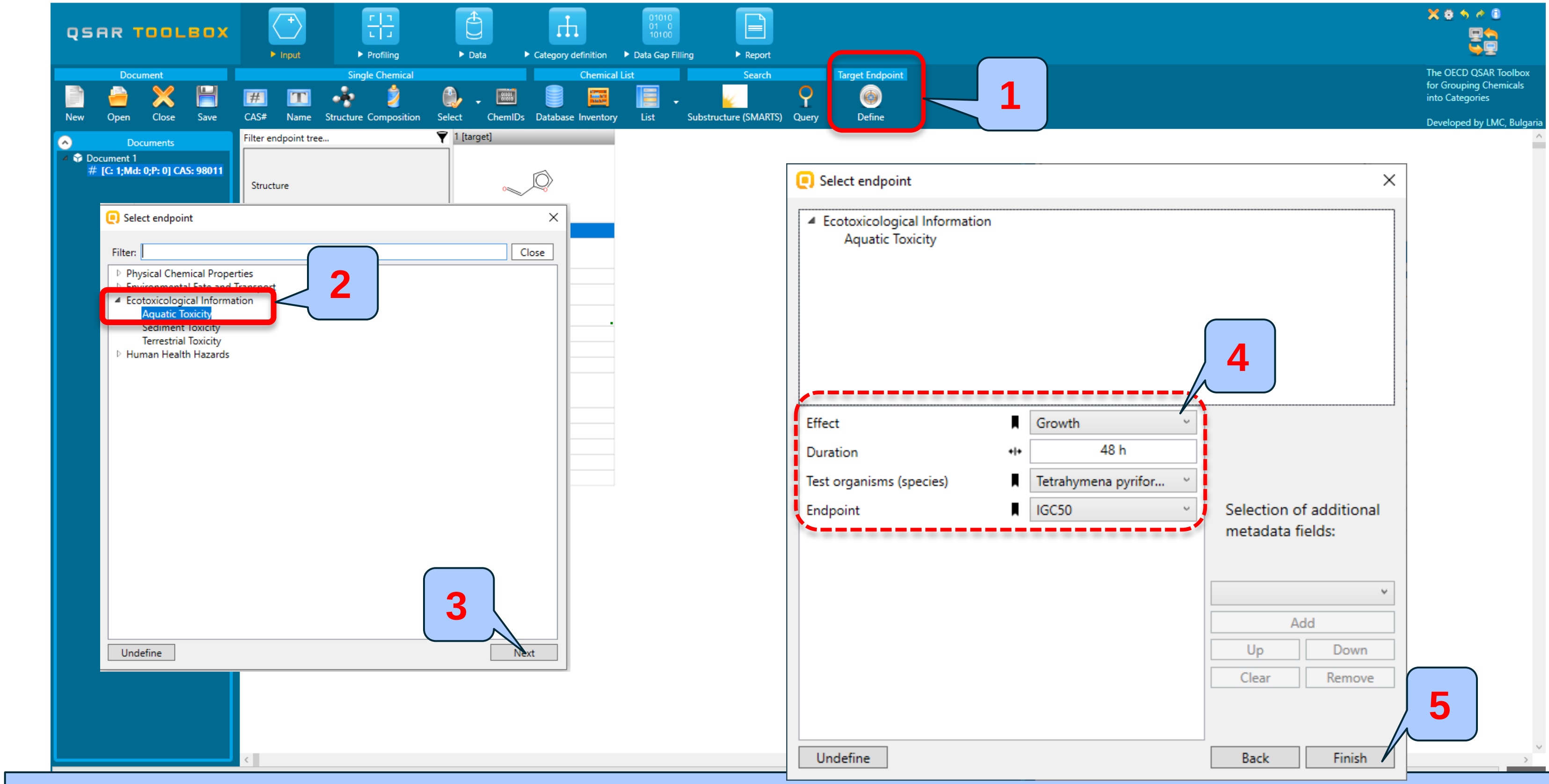
- For defining the target endpoint the "Define target endpoint" functionality is used (see next few slides)

Input
Define target endpoint

- Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, relevancy of the profiles and databases is provided expressed in different highlighting:
  - In green are highlighted the most suitable profilers related to the endpoint and databases including data for the defined target endpoint, while
  - in the orange are colored profilers which are plausible with respect to the defined target endpoint.



Input
Define target endpoint



Select endpoint

Filter: Close

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
  - Aquatic Toxicity
  - Sediment Toxicity
  - Terrestrial Toxicity
- Human Health Hazards

Undefine Next

Select endpoint

- Ecotoxicological Information
  - Aquatic Toxicity

Effect	Growth
Duration	48 h
Test organisms (species)	Tetrahymena pyrifor...
Endpoint	IGC50

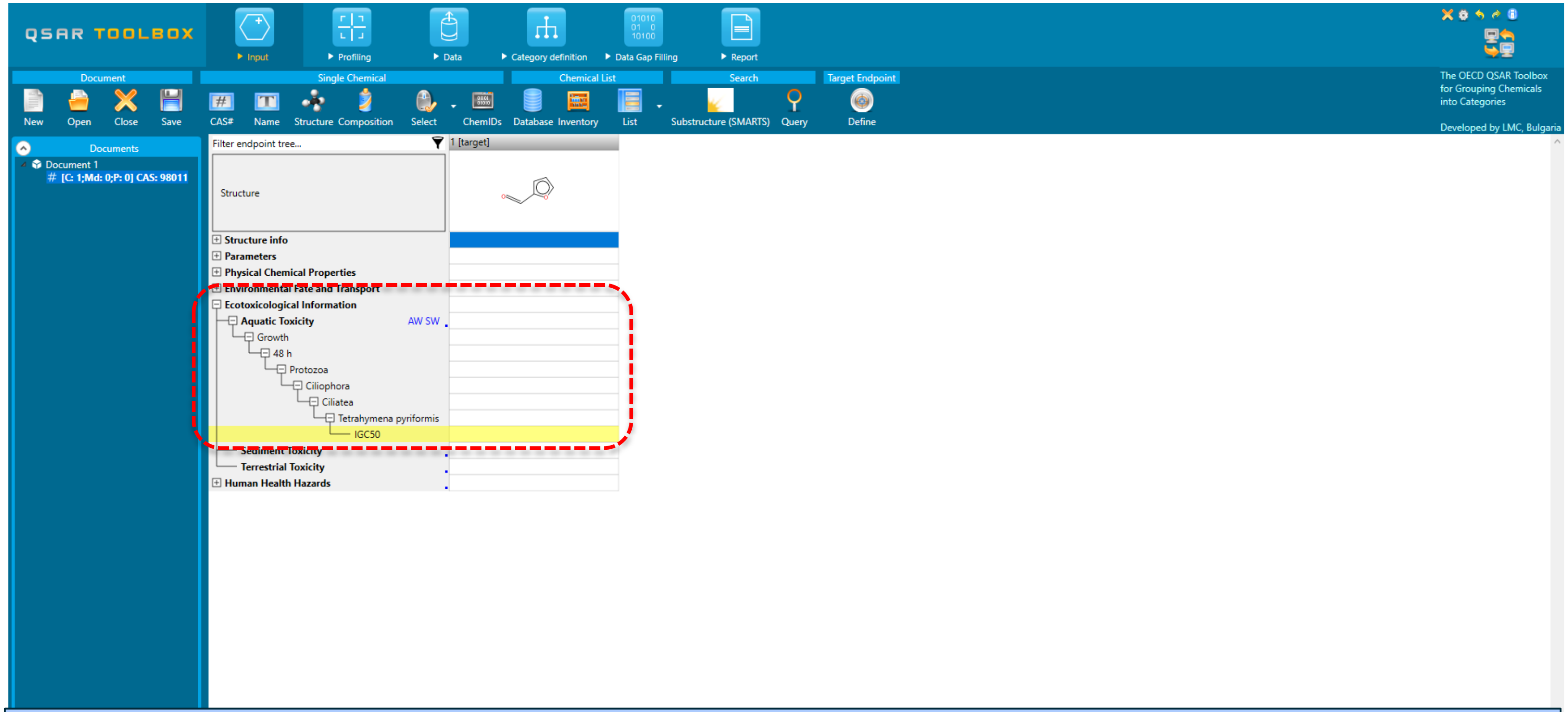
Selection of additional metadata fields:

Add Up Down Clear Remove

Undefine Back Finish

1. Click **Define**; 2. Select **Aquatic Toxicity**; 3.Click **Next** and consecutively add the following endpoint and metadata (4): **Endpoint** – IGC50; Effect – **Growth**; Duration – **48h**; Test organism (species): ***Tetrahymena pyriformis***; 5. Click **Finish**

Input
Define target endpoint



The endpoint tree is automatically expanded to the level of the defined endpoint and the row is highlighted in yellow

⚠ As mentioned above (slide 19) defining the target endpoint lead to highlighting of relevant profilers and databases (see next slides)

Outlook

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

Profiling
Overview

- *"Profiling"* module refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available "profilers" includes likely mechanism(s) of action, wich could be useful in forming categories that include the target chemical.
- "Profilers" are a collection of empirical and mechanism knowledge which could be used to analyse the structural properties of chemicals.
- The "profilers" identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts).
- The "Profiling" module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profilers
- The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances.
- The "profilers" are not (Q)SARs, i.e. they are not prediction models themselves;
- Based on the "profilers' relevancy" (determined by the defined target endpoint), the most suitable and plausible once are getting colour highlighted.

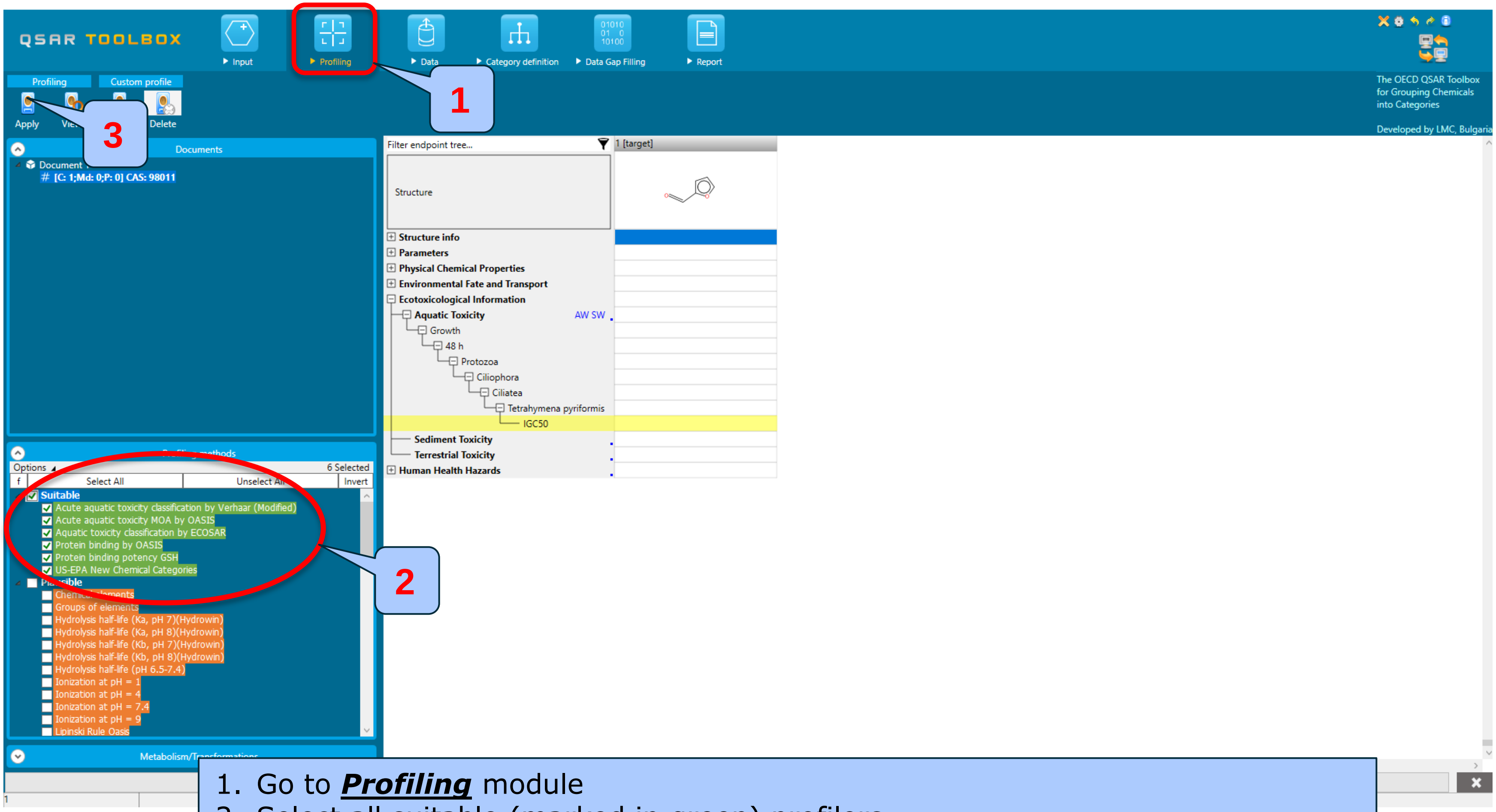
Profiling

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.
- In this case select all green (the most suitable to the target endpoint) profilers – see next slide

Profiling

Profiling the target chemical



Document
[C: 1;Md: 0;P: 0] CAS: 98011

Options — 6 Selected

f | Select All | Unselect All | Invert

- ☑ Suitable
  - ☑ Acute aquatic toxicity classification by Verhaar (Modified)
  - ☑ Acute aquatic toxicity MOA by OASIS
  - ☑ Aquatic toxicity classification by ECOSAR
  - ☑ Protein binding by OASIS
  - ☑ Protein binding potency GSH
  - ☑ US-EPA New Chemical Categories
- □ Plausible
  - □ Chemical elements
  - □ Groups of elements
  - □ Hydrolysis half-life (Ka, pH 7)(Hydrowin)
  - □ Hydrolysis half-life (Ka, pH 8)(Hydrowin)
  - □ Hydrolysis half-life (Kb, pH 7)(Hydrowin)
  - □ Hydrolysis half-life (Kb, pH 8)(Hydrowin)
  - □ Hydrolysis half-life (pH 6.5-7.4)
  - □ Ionization at pH = 1
  - □ Ionization at pH = 4
  - □ Ionization at pH = 7.4
  - □ Ionization at pH = 9
  - □ Lipinski Rule Oasis

- ⊞ Structure info
- ⊞ Parameters
- ⊞ Physical Chemical Properties
- ⊞ Environmental Fate and Transport
- ⊟ Ecotoxicological Information
  - ⊟ Aquatic Toxicity
    - ⊟ Growth
      - ⊟ 48 h
        - ⊟ Protozoa
          - ⊟ Ciliophora
            - ⊟ Ciliatea
              - ⊟ Tetrahymena pyriformis
                - IGC50
  - Sediment Toxicity
  - Terrestrial Toxicity
- ⊞ Human Health Hazards

1. Go to ***Profiling*** module
2. Select all suitable (marked in green) profilers
3. Click **Apply** to apply knowledge of the selected profilers to the target chemical

Profiling

Profiling the target chemical

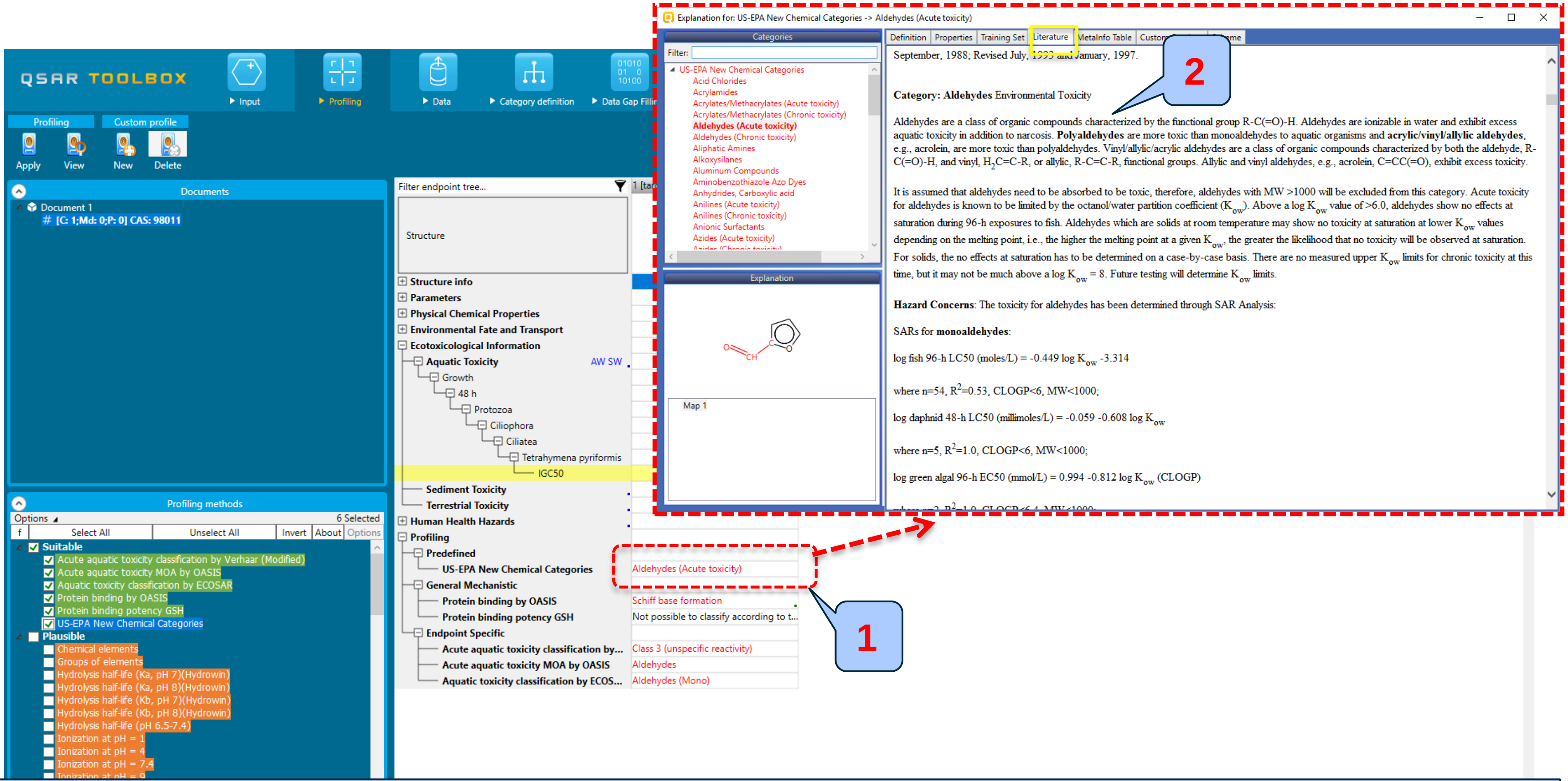
- The actual profiling will take several seconds depending on the number and type of selected profilers.

- The results of profiling automatically appeared as a dropdown box under the target chemical (see next screen shot).

- Green rectangles in some result boxes indicate there is more than one profiling result and the field needs to be expanded.

Profiling

Profiling the target chemical – profiling results


QSAR TOOLBOX
Input
Profiling
Data
Category definition
Data Gap Filling
Profiling
Custom profile
Apply
View
New
Delete
Documents
Document 1
# [C: 1;Md: 0;P: 0] CAS: 98011
Profiling methods
Options
6 Selected
Select All
Unselect All
Invert
About
Options
Suitable
Acute aquatic toxicity classification by Verhaar (Modified)
Acute aquatic toxicity MOA by OASIS
Aquatic toxicity classification by ECOSAR
Protein binding by OASIS
Protein binding potency GSH
US-EPA New Chemical Categories
Plausible
Chemical elements
Groups of elements
Hydrolysis half-life (Ka, pH 7)(Hydrowin)
Hydrolysis half-life (Ka, pH 8)(Hydrowin)
Hydrolysis half-life (Kb, pH 7)(Hydrowin)
Hydrolysis half-life (Kb, pH 8)(Hydrowin)
Hydrolysis half-life (pH 6.5-7.4)
Ionization at pH = 1
Ionization at pH = 4
Ionization at pH = 7,4
Ionization at pH = 9
Filter endpoint tree...
Structure
Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Aquatic Toxicity
AW SW
Growth
48 h
Protozoa
Ciliophora
Ciliatea
Tetrahymena pyriformis
IGC50
Sediment Toxicity
Terrestrial Toxicity
Human Health Hazards
Profiling
Predefined
General Mechanistic
Endpoint Specific
Explanation for: US-EPA New Chemical Categories -> Aldehydes (Acute toxicity)
Categories
Filter:
US-EPA New Chemical Categories
Acid Chlorides
Acrylamides
Acrylates/Methacrylates (Acute toxicity)
Acrylates/Methacrylates (Chronic toxicity)
Aldehydes (Acute toxicity)
Aldehydes (Chronic toxicity)
Aliphatic Amines
Alkoxysilanes
Aluminum Compounds
Aminobenzothiazole Azo Dyes
Anhydrides, Carboxylic acid
Anilines (Acute toxicity)
Anilines (Chronic toxicity)
Anionic Surfactants
Azides (Acute toxicity)
Explanation
O
CH
Map 1
Definition
Properties
Training Set
Literature
MetaInfo Table
2
1

US-EPA New Chemical Categories	Aldehydes (Acute toxicity)
Protein binding by OASIS	Schiff base formation
Protein binding potency GSH	Not possible to classify according to t...
Acute aquatic toxicity classification by...	Class 3 (unspecific reactivity)
Acute aquatic toxicity MOA by OASIS	Aldehydes
Aquatic toxicity classification by ECOS...	Aldehydes (Mono)

September, 1988; Revised July, 1993 and January, 1997.

Category: Aldehydes Environmental Toxicity

Aldehydes are a class of organic compounds characterized by the functional group R-C(=O)-H. Aldehydes are ionizable in water and exhibit excess aquatic toxicity in addition to narcosis. **Polyaldehydes** are more toxic than monoaldehydes to aquatic organisms and **acrylic/vinyl/allylic aldehydes**, e.g., acrolein, are more toxic than polyaldehydes. Vinyl/allylic/acrylic aldehydes are a class of organic compounds characterized by both the aldehyde, R-C(=O)-H, and vinyl, $H_2C=C-R$, or allylic, R-C=C-R, functional groups. Allylic and vinyl aldehydes, e.g., acrolein, C=CC(=O), exhibit excess toxicity.

It is assumed that aldehydes need to be absorbed to be toxic, therefore, aldehydes with MW >1000 will be excluded from this category. Acute toxicity for aldehydes is known to be limited by the octanol/water partition coefficient (K_{ow}). Above a log K_{ow} value of >6.0, aldehydes show no effects at saturation during 96-h exposures to fish. Aldehydes which are solids at room temperature may show no toxicity at saturation at lower K_{ow} values depending on the melting point, i.e., the higher the melting point at a given K_{ow}, the greater the likelihood that no toxicity will be observed at saturation. For solids, the no effects at saturation has to be determined on a case-by-case basis. There are no measured upper K_{ow} limits for chronic toxicity at this time, but it may not be much above a log K_{ow} = 8. Future testing will determine K_{ow} limits.

Hazard Concerns: The toxicity for aldehydes has been determined through SAR Analysis:

SARs for **monoaldehydes**:

log fish 96-h LC50 (moles/L) = -0.449 log K_{ow} -3.314

where n=54, R^2=0.53, CLOGP<6, MW<1000;

log daphnid 48-h LC50 (millimoles/L) = -0.059 -0.608 log K_{ow}

where n=5, R^2=1.0, CLOGP<6, MW<1000;

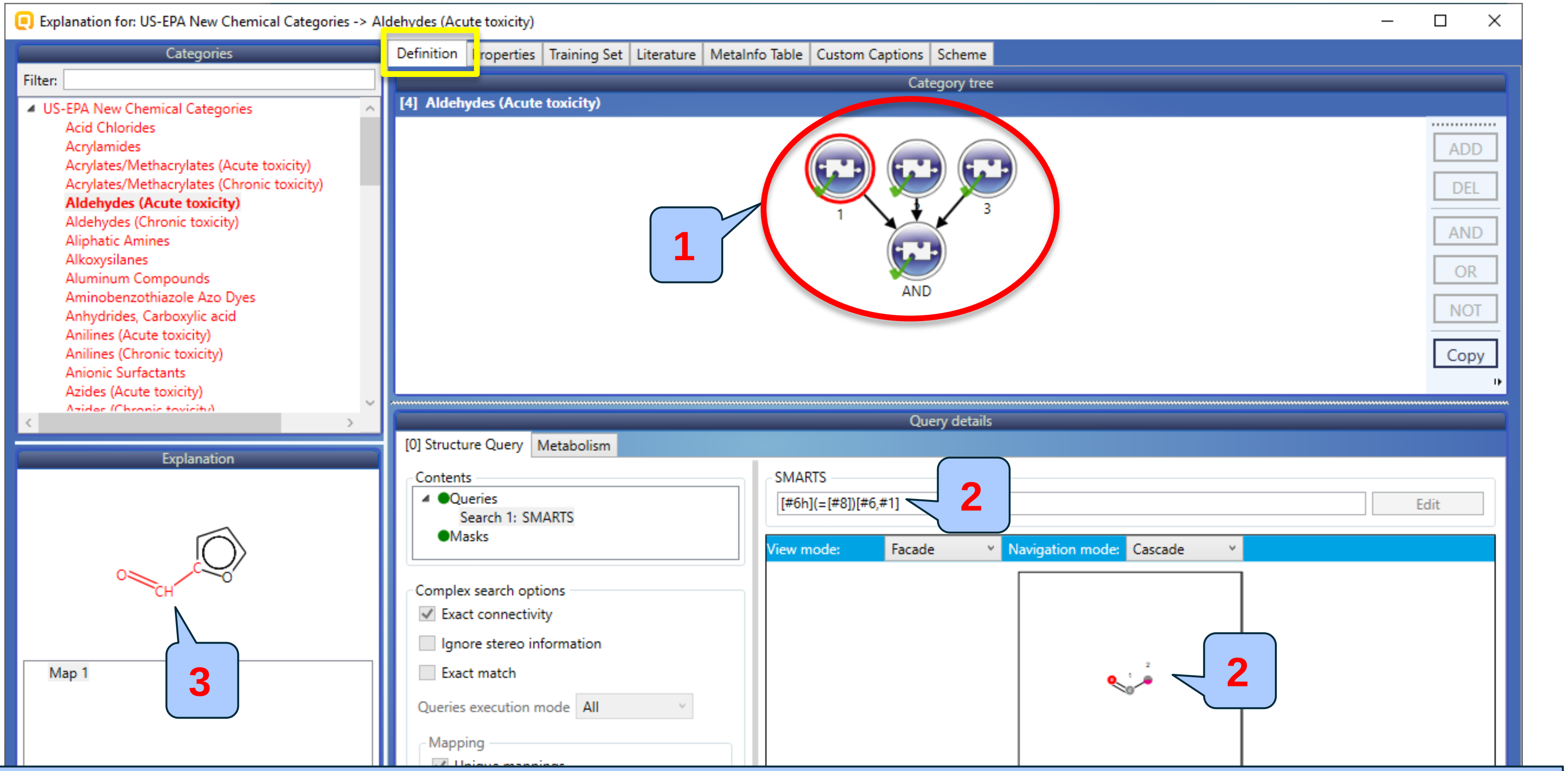
log green algal 96-h EC50 (mmol/L) = 0.994 -0.812 log K_{ow} (CLOGP)

1. Double click on the cell with “Aldehydes (Acute toxicity)” results based on US-EPA Chemical New Chemical Categories to see why the chemical is categorized as aldehyde
2. Literature information is displayed. The knowledge explained here is used for coding the structural boundaries of the category

Continued on next slide

Profiling

Profiling the target chemical – Boundaries of the profilers



Explanation for: US-EPA New Chemical Categories -> Aldehydes (Acute toxicity)

Categories

Filter:

- US-EPA New Chemical Categories
  - Acid Chlorides
  - Acrylamides
  - Acrylates/Methacrylates (Acute toxicity)
  - Acrylates/Methacrylates (Chronic toxicity)
  - **Aldehydes (Acute toxicity)**
  - Aldehydes (Chronic toxicity)
  - Aliphatic Amines
  - Alkoxysilanes
  - Aluminum Compounds
  - Aminobenzothiazole Azo Dyes
  - Anhydrides, Carboxylic acid
  - Anilines (Acute toxicity)
  - Anilines (Chronic toxicity)
  - Anionic Surfactants
  - Azides (Acute toxicity)
  - Azides (Chronic toxicity)

Explanation

Map 1

Definition | Properties | Training Set | Literature | MetaInfo Table | Custom Captions | Scheme

Category tree

[4] Aldehydes (Acute toxicity)

Query details

[0] Structure Query | Metabolism

Contents

- ● Queries
  - Search 1: SMARTS
- ● Masks

Complex search options

☑ Exact connectivity

□ Ignore stereo information

□ Exact match

Queries execution mode: All

Mapping

☑ Unique mappings

SMARTS: [#6h](=[#8])[#6,#1]

View mode: Facade

Navigation mode: Cascade

1. Structural boundaries of the category- Aldehydes (Acute toxicity); The boundaries which are met are ticked with green ✓
2. Definition of the SMARTS used for coding the knowledge; Visualization of the common fragment used for coding the knowledge;
3. The target molecule and highlighted (red) part of the molecule meeting the structure boundary.

Profiling

Profiling results

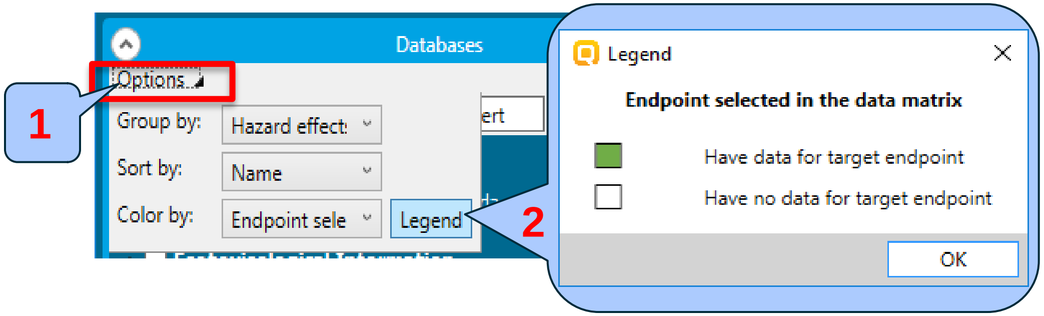
1) In module *Profile*, you have profiled the target chemical according to the suitable profilers (green) related to the target endpoint.
2) The target chemical is categorized as “aldehyde” based on predefined Acute aquatic toxicity US-EPA profiler (hereafter called US-EPA) and the two endpoint-specific profilers (Acute aquatic toxicity classification by ECOSAR (hereafter called ECOSAR) and Acute aquatic toxicity MOA by OASIS (hereafter called MOA))
3) By the endpoint-specific “Acute aquatic toxicity classification by Verhaar” the target is categorized as “Class 3 (unspecific reactivity)”
4) Moreover the target is categorized as “aldehyde” based on Protein binding by OASIS reactiving by Schiff-base formation mechanism
5) In general the target is classified as “aldehyde”
6) All of the above mentioned profilers could be used for categorization purposes (collecting analogues)
7) In this case US-EPA profiler will be used for categorization purpose (primary grouping).

Outlook

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

Data

- *Data* module refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox databases.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (i.e. collecting data for a single or limited number of endpoints).
- Once the endpoint is selected, the relevant databases are highlighted. Meaning of the colors could be seen within the **Options** (1) by click **Legend** (2).



Group by: Hazard effect:
Sort by: Name
Color by: Endpoint sele

Legend

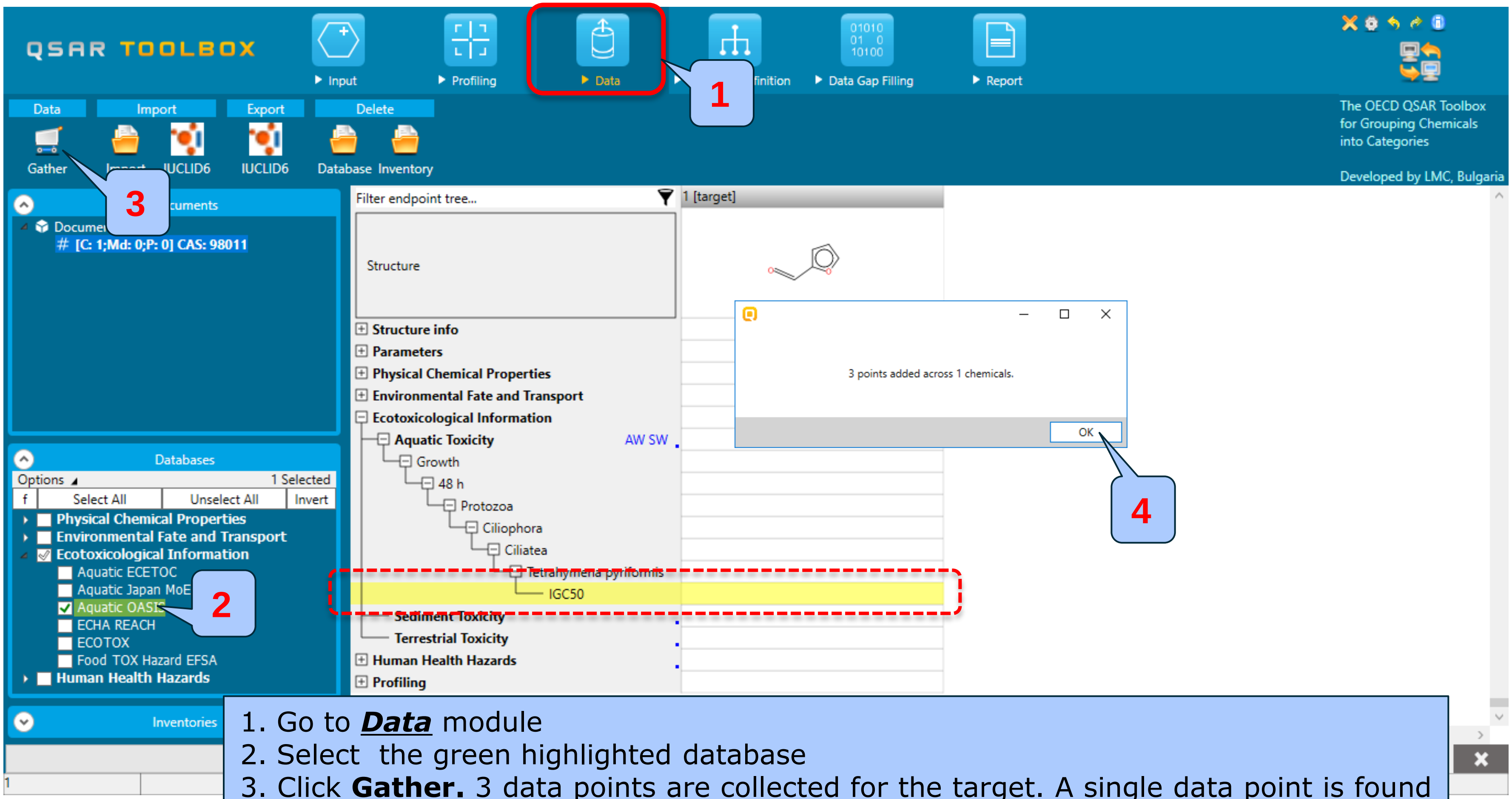
Endpoint selected in the data matrix

🟩 Have data for target endpoint

□ Have no data for target endpoint

- In this example, we limit our data gathering to the databases containing aquatic toxicity data for the defined target endpoint (Aquatic OASIS).

Data
Extracting endpoint values



1. Go to ***Data*** module
2. Select the green highlighted database
3. Click **Gather.** 3 data points are collected for the target. A single data point is found for the target endpoint; We will try to reproduce it.
4. Click **OK**

Outlook

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

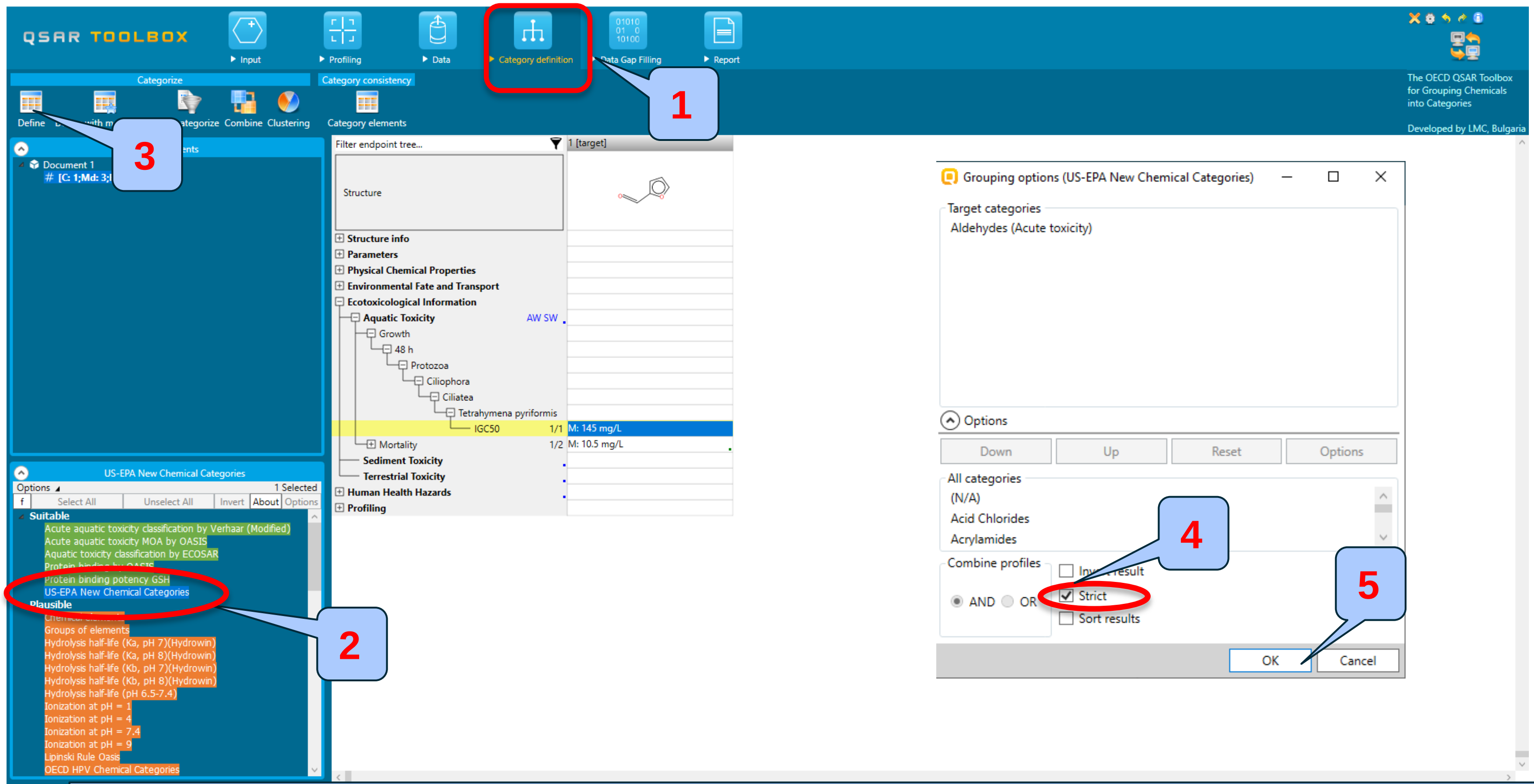
Category definition
Defining US-EPA category

- As mentioned before, 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



Grouping options (US-EPA New Chemical Categories)

Target categories

Aldehydes (Acute toxicity)

Options

Down Up Reset Options

All categories

(N/A)

Acid Chlorides

Acrylamides

Combine profiles

● AND ○ OR

□ Invert result

☑ Strict

□ Sort results

OK Cancel

1. Go to ***Category definition*** module; 2. Highlight **"US-EPA New Chemical Categories"**; 3. Click **Define**; 4. Put a tick in the Strict box (see next screen shot); 5. Click OK to confirm the category **Aldehydes (Acute toxicity);**

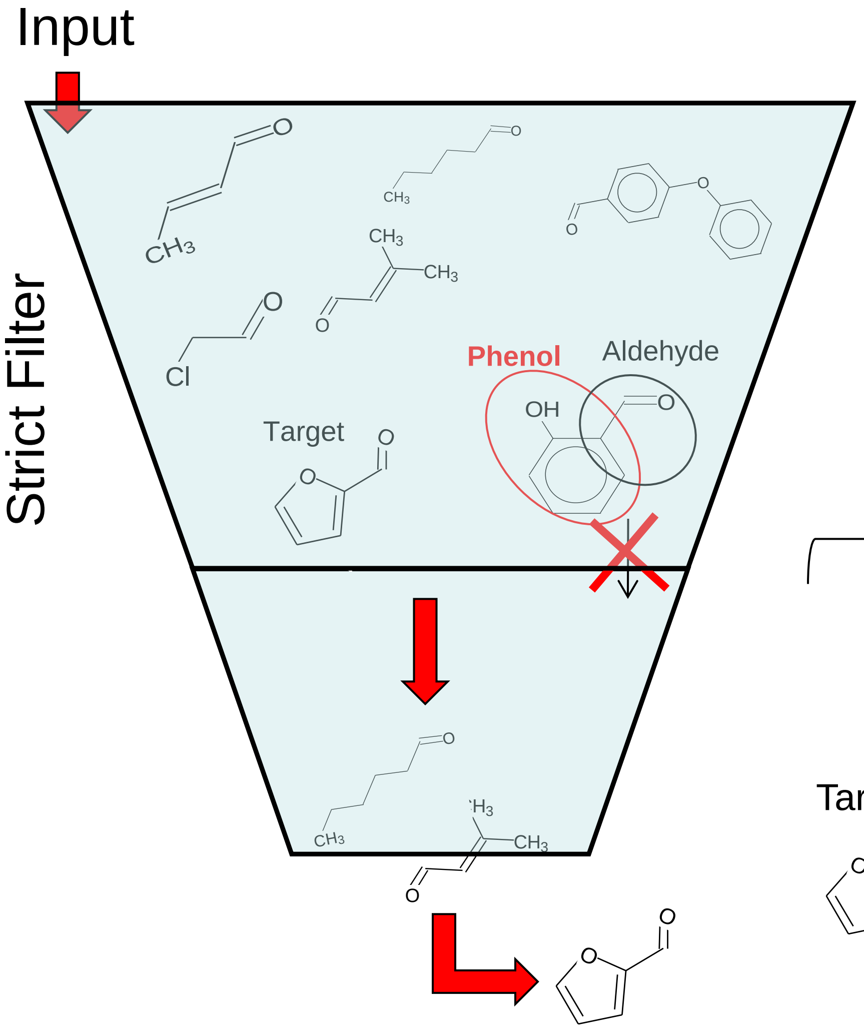
Category definition
Defining US-EPA category strict functionality

- 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

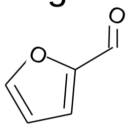
Defining US-EPA category strict functionality



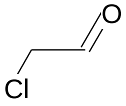
The target and analogues have *Aldehydes* **ONLY** according to US-EPA category

Defined Category

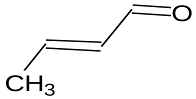
Target



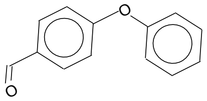
Analogue 1



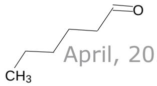
Analogue 2



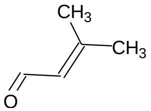
Analogue 3



Analogue 4



Analogue 5



Category definition
Analogues

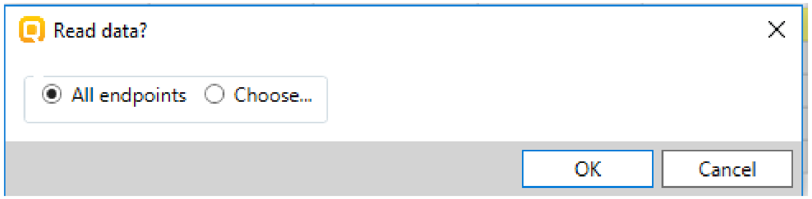
- The Toolbox now identifies all chemicals corresponding to *Aldehydes (Acute toxicity)* by US-EPA listed in the databases selected under “Data”.

- 101 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

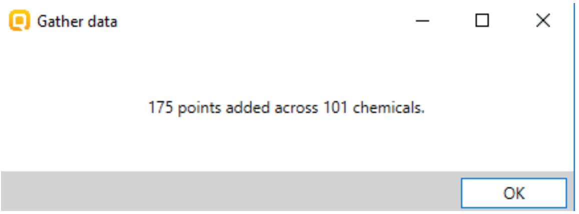
- The Toolbox automatically request the user to select the endpoint that should be retrieved

- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below). Click OK to read all available data. 175 data points are collected for the list of 101 analogues



Read data?

● All endpoints ○ Choose...

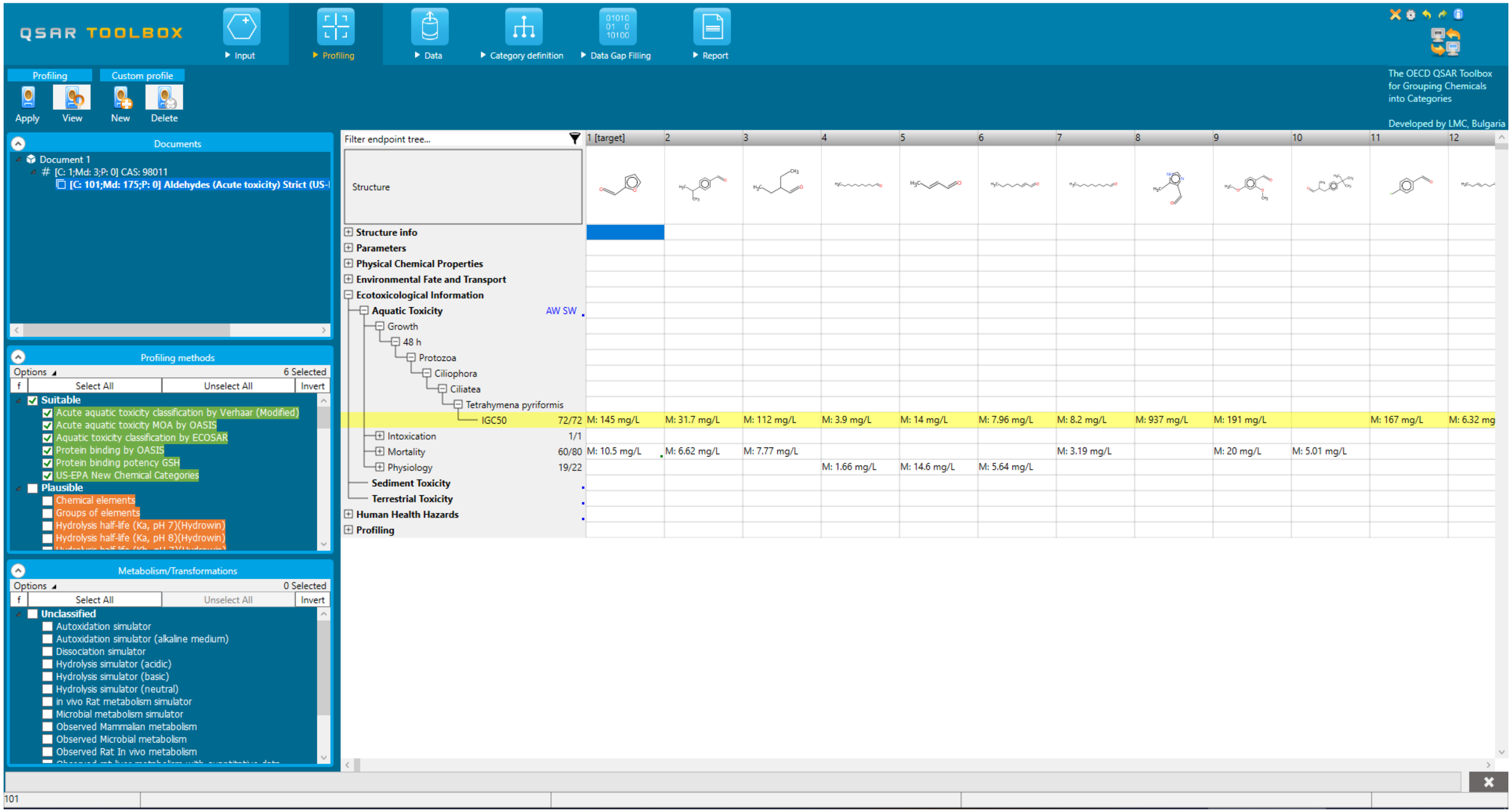


Gather data

175 points added across 101 chemicals.

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.



Documents
- Document 1
  - # [C: 1;Md: 3;P: 0] CAS: 98011
    - [C: 101;Md: 175;P: 0] Aldehydes (Acute toxicity) Strict (US-

Profiling methods

Options 6 Selected

f | Select All | Unselect All | Invert

- ☑ Suitable
  - ☑ Acute aquatic toxicity classification by Verhaar (Modified)
  - ☑ Acute aquatic toxicity MOA by OASIS
  - ☑ Aquatic toxicity classification by ECOSAR
  - ☑ Protein binding by OASIS
  - ☑ Protein binding potency GSH
  - ☑ US-EPA New Chemical Categories
- □ Plausible
  - □ Chemical elements
  - □ Groups of elements
  - □ Hydrolysis half-life (Ka, pH 7)(Hydrowin)
  - □ Hydrolysis half-life (Ka, pH 8)(Hydrowin)

Metabolism/Transformations

Options 0 Selected

f | Select All | Unselect All | Invert

- □ Unclassified
  - □ Autoxidation simulator
  - □ Autoxidation simulator (alkaline medium)
  - □ Dissociation simulator
  - □ Hydrolysis simulator (acidic)
  - □ Hydrolysis simulator (basic)
  - □ Hydrolysis simulator (neutral)
  - □ in vivo Rat metabolism simulator
  - □ Microbial metabolism simulator
  - □ Observed Mammalian metabolism
  - □ Observed Microbial metabolism
  - □ Observed Rat In vivo metabolism

Filter endpoint tree...		1 [target]	2	3	4	5	6	7	8	9	10	11	12
Structure													
⊞ Structure info													
⊞ Parameters													
⊞ Physical Chemical Properties													
⊞ Environmental Fate and Transport													
⊟ Ecotoxicological Information													
⊟ Aquatic Toxicity	AW SW												
⊟ Growth													
⊟ 48 h													
⊟ Protozoa													
⊟ Ciliophora													
⊟ Ciliatea													
⊟ Tetrahymena pyriformis													
IGC50	72/72	M: 145 mg/L	M: 31.7 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 14 mg/L	M: 7.96 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L		M: 167 mg/L	M: 6.32 mg
⊞ Intoxication	1/1												
⊞ Mortality	60/80	M: 10.5 mg/L	M: 6.62 mg/L	M: 7.77 mg/L				M: 3.19 mg/L		M: 20 mg/L	M: 5.01 mg/L		
⊞ Physiology	19/22				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L						
Sediment Toxicity													
Terrestrial Toxicity													
⊞ Human Health Hazards													
⊞ Profiling													

101

Outlook

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

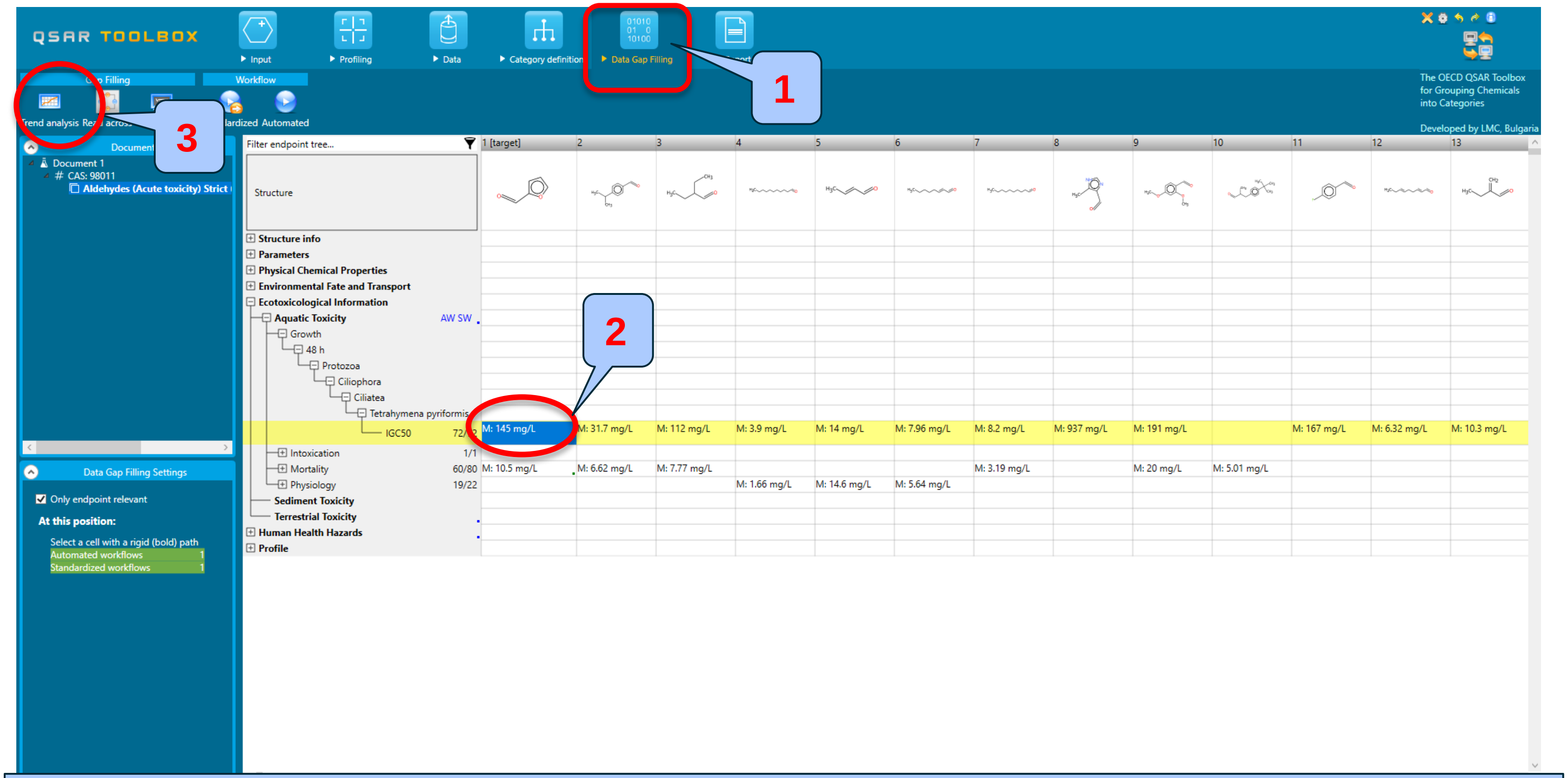
Data Gap Filling

Overview

- *Data Gap Filling* module gives access to five different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
  - Standardized workflow (SW)
  - Automated workflow (AW)
- The most relevant data gap mechanism is used , taking into account the following considerations:
  - *Read-across* is the appropriate data-gap filling method for “qualitative” endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for “quantitative endpoints” (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
  - *Trend analysis* is the appropriate data-gap filling method for “quantitative endpoints” (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
  - *(Q)SAR models* can be used to fill a data gap if no adequate analogues are found for a target chemical.
  - *Automated and standardized workflows* follow preliminary implemented logic. The AW is not affected by the user activities (proceeding or subsequent), while the SW stops at the each step of the workflows allowing the user to make different selection.
- In this example we will use trend analysis.

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

Apply Trend analysis



Document 1
CAS: 98011
Aldehydes (Acute toxicity) Strict

☑ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path

Automated workflows 1

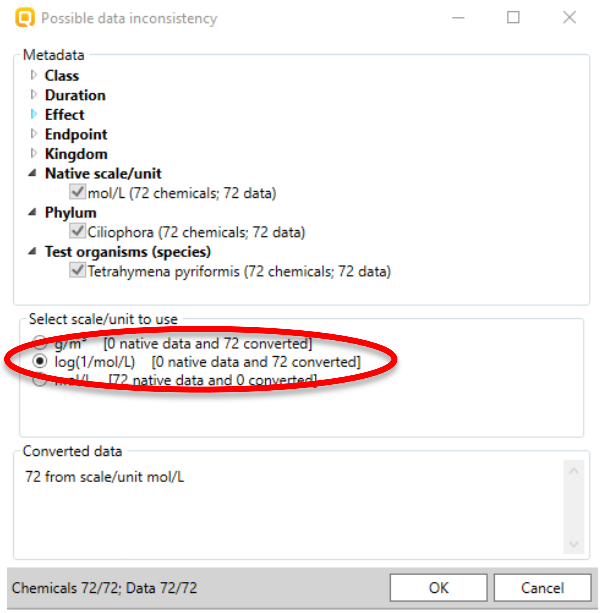
Standardized workflows 1

		1 [target]	2	3	4	5	6	7	8	9	10	11	12	13
Structure														
Structure info														
Parameters														
Physical Chemical Properties														
Environmental Fate and Transport														
Ecotoxicological Information														
Aquatic Toxicity	AW SW													
Growth														
48 h														
Protozoa														
Ciliophora														
Ciliatea														
Tetrahymena pyriformis														
IGC50	72/	M: 145 mg/L	M: 31.7 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 14 mg/L	M: 7.96 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L		M: 167 mg/L	M: 6.32 mg/L	M: 10.3 mg/L
Intoxication	1/1													
Mortality	60/80	M: 10.5 mg/L	M: 6.62 mg/L	M: 7.77 mg/L				M: 3.19 mg/L		M: 20 mg/L	M: 5.01 mg/L			
Physiology	19/22				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L							
Sediment Toxicity														
Terrestrial Toxicity														
Human Health Hazards														
Profile														

1. Go to ***Data Gap Filling;*** 2. Highlight the **data gap** corresponding to target endpoint: 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



Possible data inconsistency

Metadata
- Class
- Duration
- Effect
- Endpoint
- Kingdom
- Native scale/unit
  - ☑ mol/L (72 chemicals; 72 data)
- Phylum
  - ☑ Ciliophora (72 chemicals; 72 data)
- Test organisms (species)
  - ☑ Tetrahymena pyriformis (72 chemicals; 72 data)

Select scale/unit to use

○ g/m^3 [0 native data and 72 converted]
● log(1/mol/L) [0 native data and 72 converted]
○ mol/L [72 native data and 0 converted]

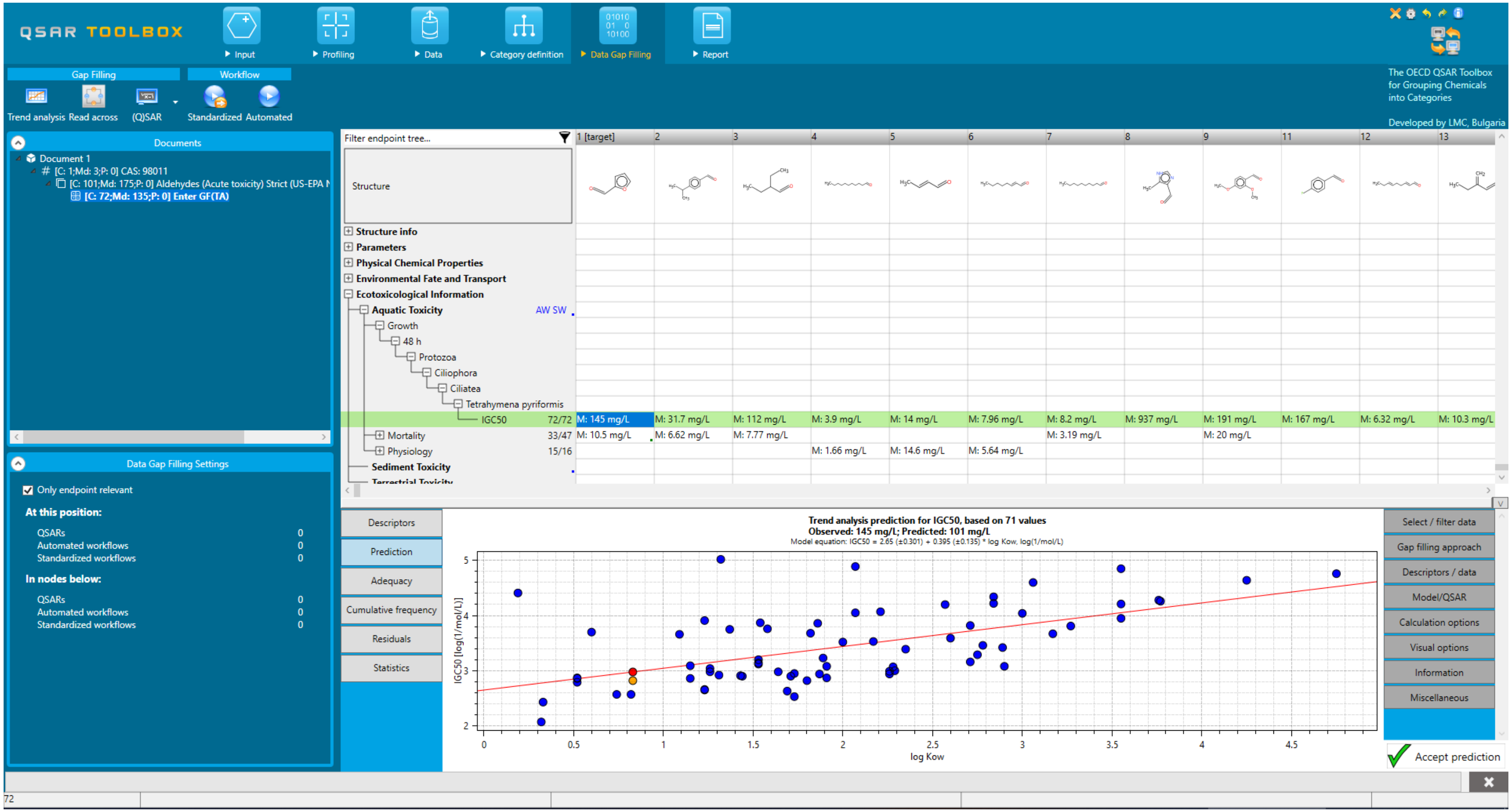
Converted data

72 from scale/unit mol/L

Chemicals 72/72; Data 72/72

- The resulting plot can be seen on next screen shot

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



Document 1
[C: 1;Md: 3;P: 0] CAS: 98011
[C: 101;Md: 175;P: 0] Aldehydes (Acute toxicity) Strict (US-EPA N
[C: 72;Md: 135;P: 0] Enter GF(TA)

☑ Only endpoint relevant

At this position:

QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

QSARs	0
Automated workflows	0
Standardized workflows	0

Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Aquatic Toxicity AW SW
Growth
48 h
Protozoa
Ciliophora
Ciliatea
Tetrahymena pyriformis
Sediment Toxicity
Terrestrial Toxicity

		1 [target]	2	3	4	5	6	7	8	9	11	12	13
IGC50	72/72	M: 145 mg/L	M: 31.7 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 14 mg/L	M: 7.96 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L	M: 167 mg/L	M: 6.32 mg/L	M: 10.3 mg/L
Mortality	33/47	M: 10.5 mg/L	M: 6.62 mg/L	M: 7.77 mg/L				M: 3.19 mg/L		M: 20 mg/L			
Physiology	15/16				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L						

Trend analysis prediction for IGC50, based on 71 values
Observed: 145 mg/L; Predicted: 101 mg/L
Model equation: IGC50 = 2.65 (±0.301) + 0.395 (±0.135) * log Kow, log(1/mol/L)

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 **ORANGE** dot represents the observed data value for the 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*)

Subcategorization

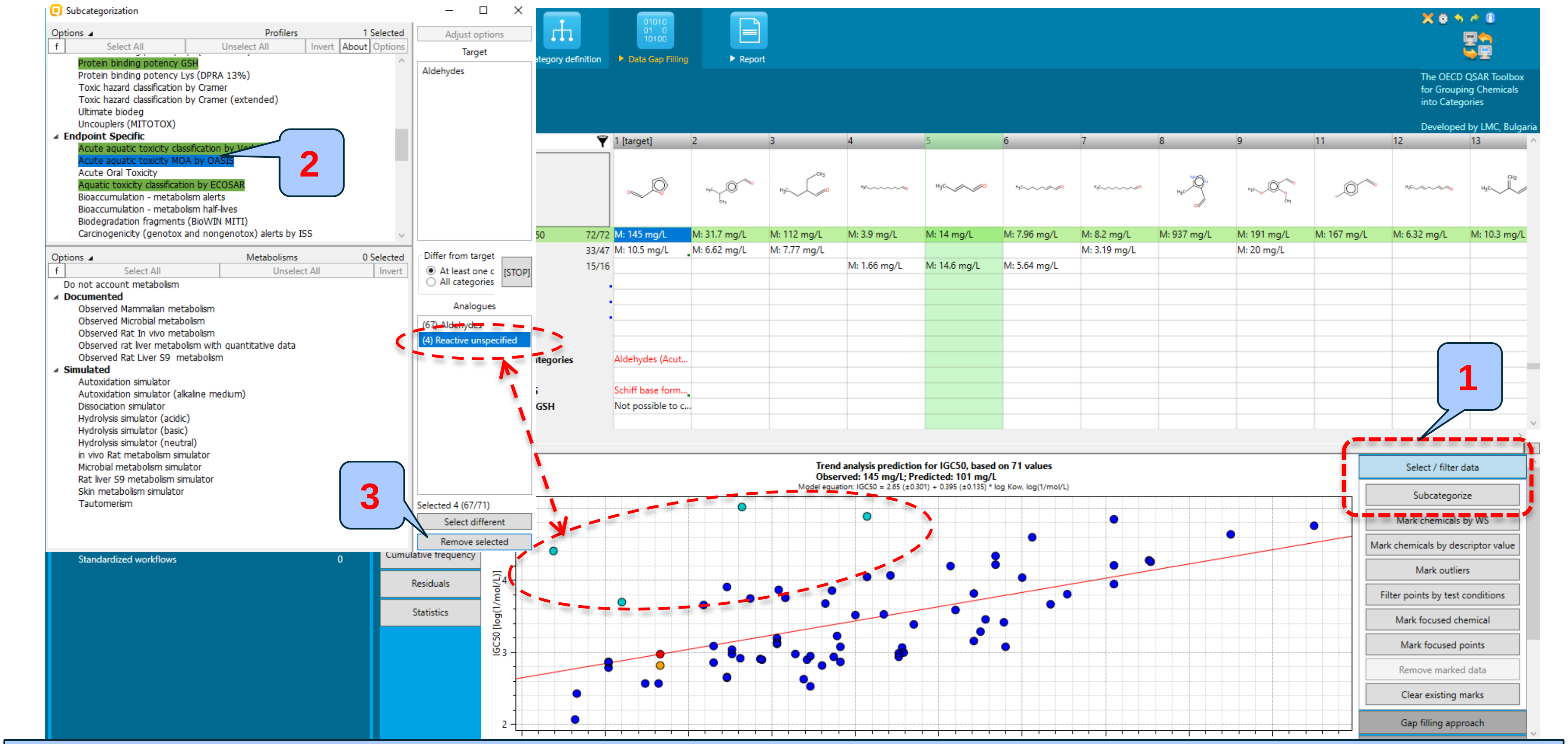
- After the available data has been retrieved, the user can then further subcategorize the results according to the following endpoint-specific subcategorizations:

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

- These steps are summarized in the next screen shots.

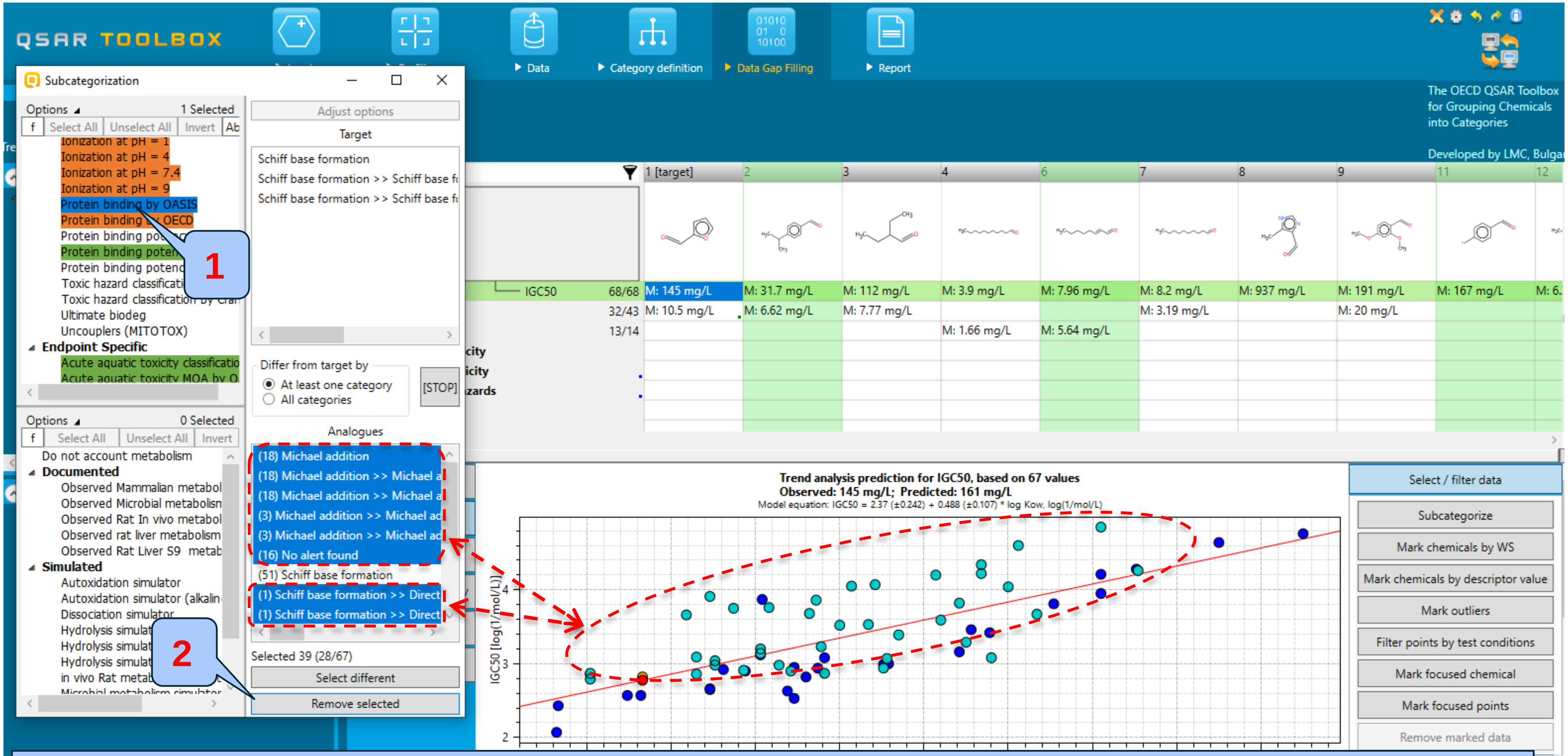
Data Gap Filling (IGC 50 48h of *T. pyriformis*)

Subcategorization 1: Acute aquatic toxicity MOA by OASIS


Subcategorization
Options
Profilers
1 Selected
Select All
Unselect All
Invert
About
Options
Protein binding potency GSH
Protein binding potency Lys (DPRA 13%)
Toxic hazard classification by Cramer
Toxic hazard classification by Cramer (extended)
Ultimate biodeg
Uncouplers (MITOTOX)
Endpoint Specific
Acute aquatic toxicity classification by Verhaar
Acute aquatic toxicity MOA by OASIS
Acute Oral Toxicity
Aquatic toxicity classification by ECOSAR
Bioaccumulation - metabolism alerts
Bioaccumulation - metabolism half-lives
Biodegradation fragments (BioWIN MITI)
Carcinogenicity (genotox and nongenotox) alerts by ISS
Metabolisms
0 Selected
Do not account metabolism
Documented
Observed Mammalian metabolism
Observed Microbial metabolism
Observed Rat In vivo metabolism
Observed rat liver metabolism with quantitative data
Observed Rat Liver S9 metabolism
Simulated
Autoxidation simulator
Autoxidation simulator (alkaline medium)
Dissociation simulator
Hydrolysis simulator (acidic)
Hydrolysis simulator (basic)
Hydrolysis simulator (neutral)
in vivo Rat metabolism simulator
Microbial metabolism simulator
Rat liver S9 metabolism simulator
Skin metabolism simulator
Tautomerism
Adjust options
Target
Aldehydes
Differ from target
At least one c
All categories
[STOP]
Analogues
(67) Aldehydes
(4) Reactive unspecified
Selected 4 (67/71)
Select different
Remove selected
Standardized workflows
0
Cumulative frequency
Residuals
Statistics
Category definition
Data Gap Filling
Report
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria
1 [target]
M: 145 mg/L
M: 10.5 mg/L
M: 31.7 mg/L
M: 6.62 mg/L
M: 112 mg/L
M: 7.77 mg/L
M: 3.9 mg/L
M: 1.66 mg/L
M: 14 mg/L
M: 14.6 mg/L
M: 7.96 mg/L
M: 5.64 mg/L
M: 8.2 mg/L
M: 3.19 mg/L
M: 937 mg/L
M: 191 mg/L
M: 20 mg/L
M: 167 mg/L
M: 6.32 mg/L
M: 10.3 mg/L
72/72
33/47
15/16
Aldehydes (Acut...
Schiff base form...
Not possible to c...
GSH
Trend analysis prediction for IGC50, based on 71 values
Observed: 145 mg/L; Predicted: 101 mg/L
Model equation: IGC50 = 2.65 (±0.301) + 0.395 (±0.135) * log Kow, log(1/mol/L)
IGC50 [log(1/mol/L)]
2
3
4
Select / filter data
Subcategorize
Mark chemicals by WS
Mark chemicals by descriptor value
Mark outliers
Filter points by test conditions
Mark focused chemical
Mark focused points
Remove marked data
Clear existing marks
Gap filling approach
1
2
3

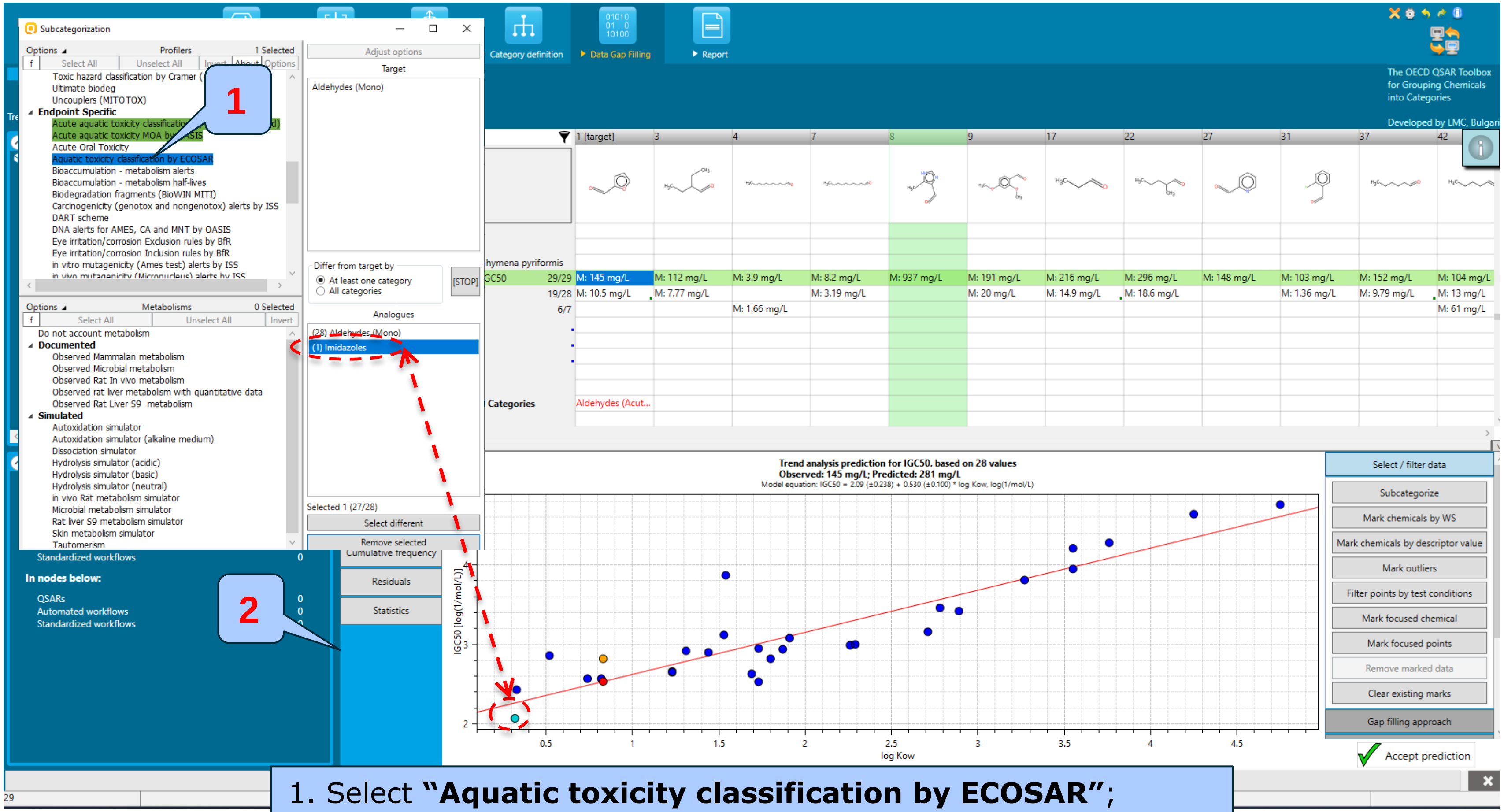
1. Click **Select / filter data**, then **Subcategorize**; 2. Select **"MOA by OASIS"** (Note: the most suitable profilers for subcategorization are again green highlighted); 3. **Click** "Remove selected" to eliminate dissimilar to the target analogues (in this case analogues categorized as "reactive unspecified" based on MOA profiler will be eliminated)

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 analogues.

Data Gap Filling
(IGC 50 48h of *T. pyriformis*)
Subcategorization 3: Aquatic toxicity classification by ECOSAR

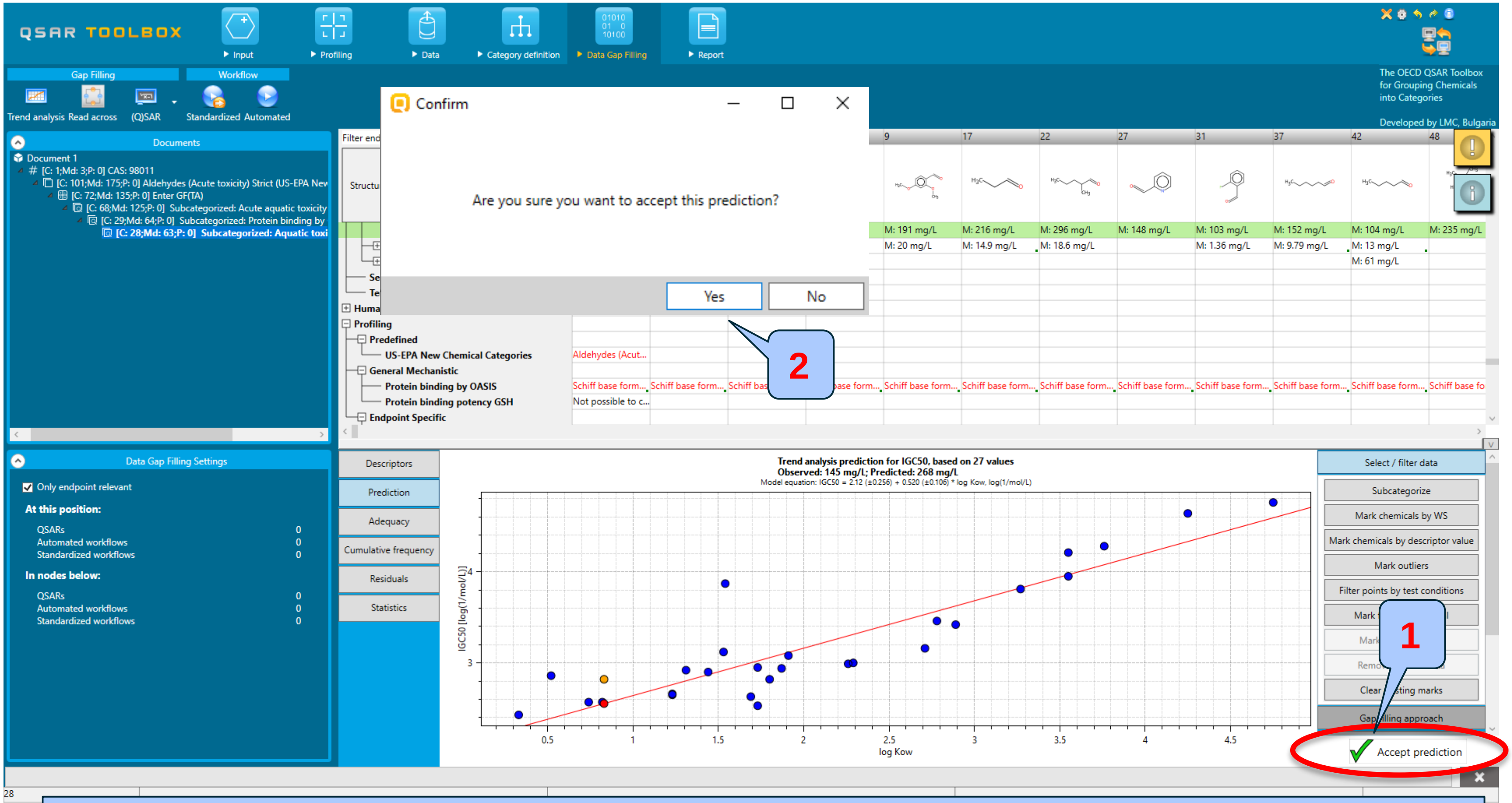

Subcategorization
Options
Profilers
1 Selected
Select All
Unselect All
Invert
About
Options
Toxic hazard classification by Cramer (
Ultimate biodeg
Uncouplers (MITOTOX)
Endpoint Specific
Acute aquatic toxicity classification
Acute aquatic toxicity MOA by OASIS
Acute Oral Toxicity
Aquatic toxicity classification by ECOSAR
Bioaccumulation - metabolism alerts
Bioaccumulation - metabolism half-lives
Biodegradation fragments (BioWIN MITI)
Carcinogenicity (genotox and nongenotox) alerts by ISS
DART scheme
DNA alerts for AMES, CA and MNT by OASIS
Eye irritation/corrosion Exclusion rules by BfR
Eye irritation/corrosion Inclusion rules by BfR
in vitro mutagenicity (Ames test) alerts by ISS
in vivo mutagenicity (Micronucleus) alerts by ISS
Options
Metabolisms
0 Selected
Select All
Unselect All
Invert
Do not account metabolism
Documented
Observed Mammalian metabolism
Observed Microbial metabolism
Observed Rat In vivo metabolism
Observed rat liver metabolism with quantitative data
Observed Rat Liver S9 metabolism
Simulated
Autoxidation simulator
Autoxidation simulator (alkaline medium)
Dissociation simulator
Hydrolysis simulator (acidic)
Hydrolysis simulator (basic)
Hydrolysis simulator (neutral)
in vivo Rat metabolism simulator
Microbial metabolism simulator
Rat liver S9 metabolism simulator
Skin metabolism simulator
Tautomerism
Adjust options
Target
Aldehydes (Mono)
Differ from target by
● At least one category
○ All categories
[STOP]
Analogues
(28) Aldehydes (Mono)
(1) Imidazoles
Selected 1 (27/28)
Select different
Remove selected
1
2
Category definition
Data Gap Filling
Report
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria
Standardized workflows 0
In nodes below:
QSARs 0
Automated workflows 0
Standardized workflows
Cumulative frequency
Residuals
Statistics
1 [target]
3
4
7
8
9
17
22
27
31
37
42
hymena pyriformis
GC50 29/29
19/28
6/7
Categories
Aldehydes (Acut...
Trend analysis prediction for IGC50, based on 28 values
Observed: 145 mg/L; Predicted: 281 mg/L
Model equation: IGC50 = 2.09 (±0.238) + 0.530 (±0.100) * log Kow, log(1/mol/L)
IGC50 [log(1/mol/L)]
log Kow
2
3
4
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Select / filter data
Subcategorize
Mark chemicals by WS
Mark chemicals by descriptor value
Mark outliers
Filter points by test conditions
Mark focused chemical
Mark focused points
Remove marked data
Clear existing marks
Gap filling approach
Accept prediction
29

	1 [target]	3	4	7	8	9	17	22	27	31	37	42
GC50 29/29	M: 145 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L	M: 216 mg/L	M: 296 mg/L	M: 148 mg/L	M: 103 mg/L	M: 152 mg/L	M: 104 mg/L
19/28	M: 10.5 mg/L	M: 7.77 mg/L		M: 3.19 mg/L		M: 20 mg/L	M: 14.9 mg/L	M: 18.6 mg/L		M: 1.36 mg/L	M: 9.79 mg/L	M: 13 mg/L
6/7			M: 1.66 mg/L									M: 61 mg/L

1. Select **"Aquatic toxicity classification by ECOSAR"**;
2. Click **"Remove selected"** to eliminate the single analogue;

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

Results after subcategorisation



Document 1
[C: 1;Md: 3;P: 0] CAS: 98011
[C: 101;Md: 175;P: 0] Aldehydes (Acute toxicity) Strict (US-EPA New
[C: 72;Md: 135;P: 0] Enter GF(TA)
[C: 68;Md: 125;P: 0] Subcategorized: Acute aquatic toxicity
[C: 29;Md: 64;P: 0] Subcategorized: Protein binding by
[C: 28;Md: 63;P: 0] Subcategorized: Aquatic toxi

Confirm

Are you sure you want to accept this prediction?

Profiling
Predefined
US-EPA New Chemical Categories
General Mechanistic
Protein binding by OASIS
Protein binding potency GSH
Endpoint Specific

Data Gap Filling Settings

☑ Only endpoint relevant

At this position:

QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

QSARs	0
Automated workflows	0
Standardized workflows	0

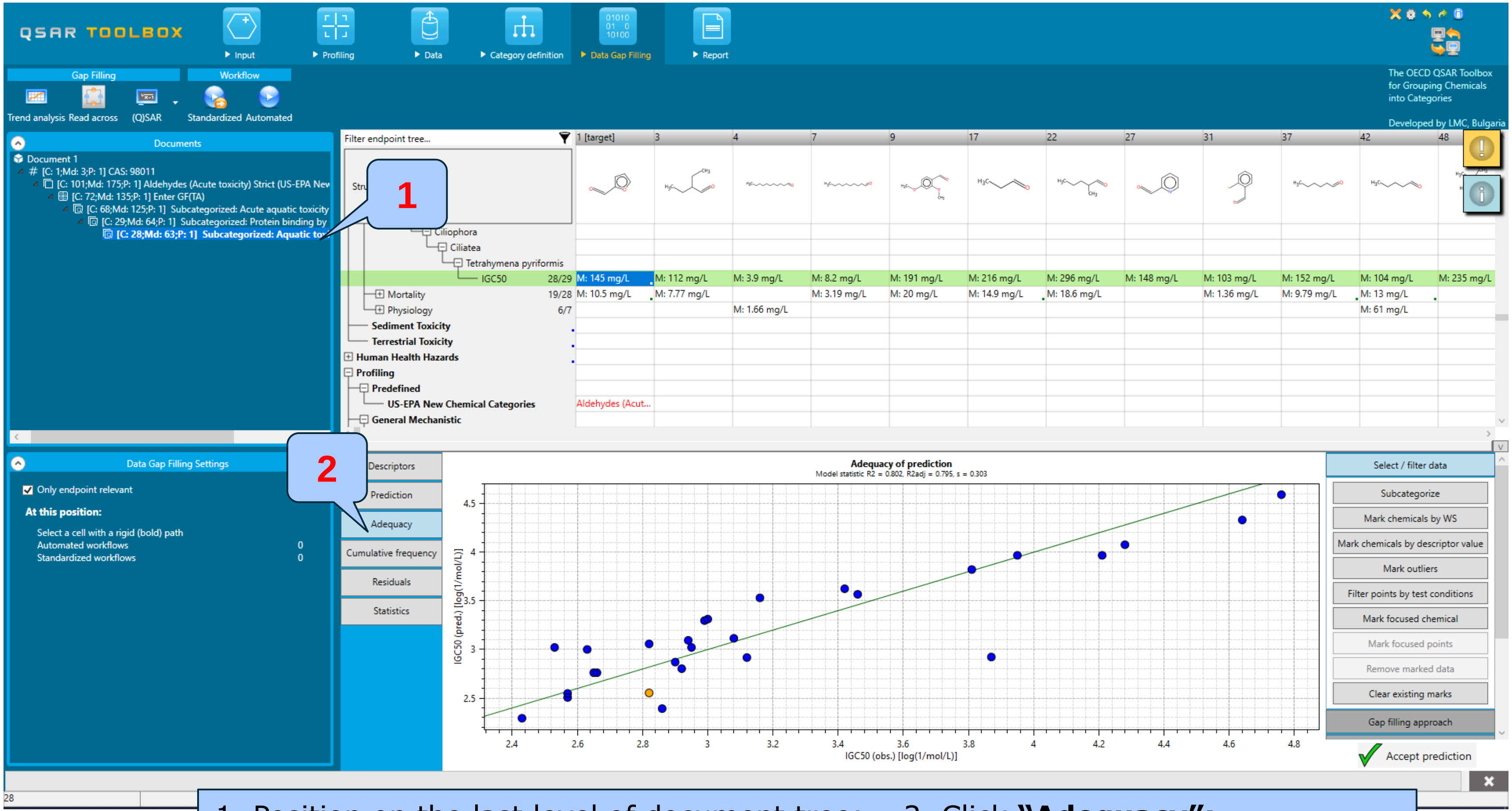
Trend analysis prediction for IGC50, based on 27 values
Observed: 145 mg/L; Predicted: 268 mg/L
Model equation: IGC50 = 2.12 (±0.256) + 0.520 (±0.106) * log Kow, log(1/mol/L)

1. Click **"Accept prediction"**; 2. 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



Document 1
- # [C: 1;Md: 3;P: 1] CAS: 98011
  - [C: 101;Md: 175;P: 1] Aldehydes (Acute toxicity) Strict (US-EPA New
    - [C: 72;Md: 135;P: 1] Enter GF(TA)
      - [C: 68;Md: 125;P: 1] Subcategorized: Acute aquatic toxicity
        - [C: 29;Md: 64;P: 1] Subcategorized: Protein binding by
          - **[C: 28;Md: 63;P: 1] Subcategorized: Aquatic tox**

☑ Only endpoint relevant

At this position:

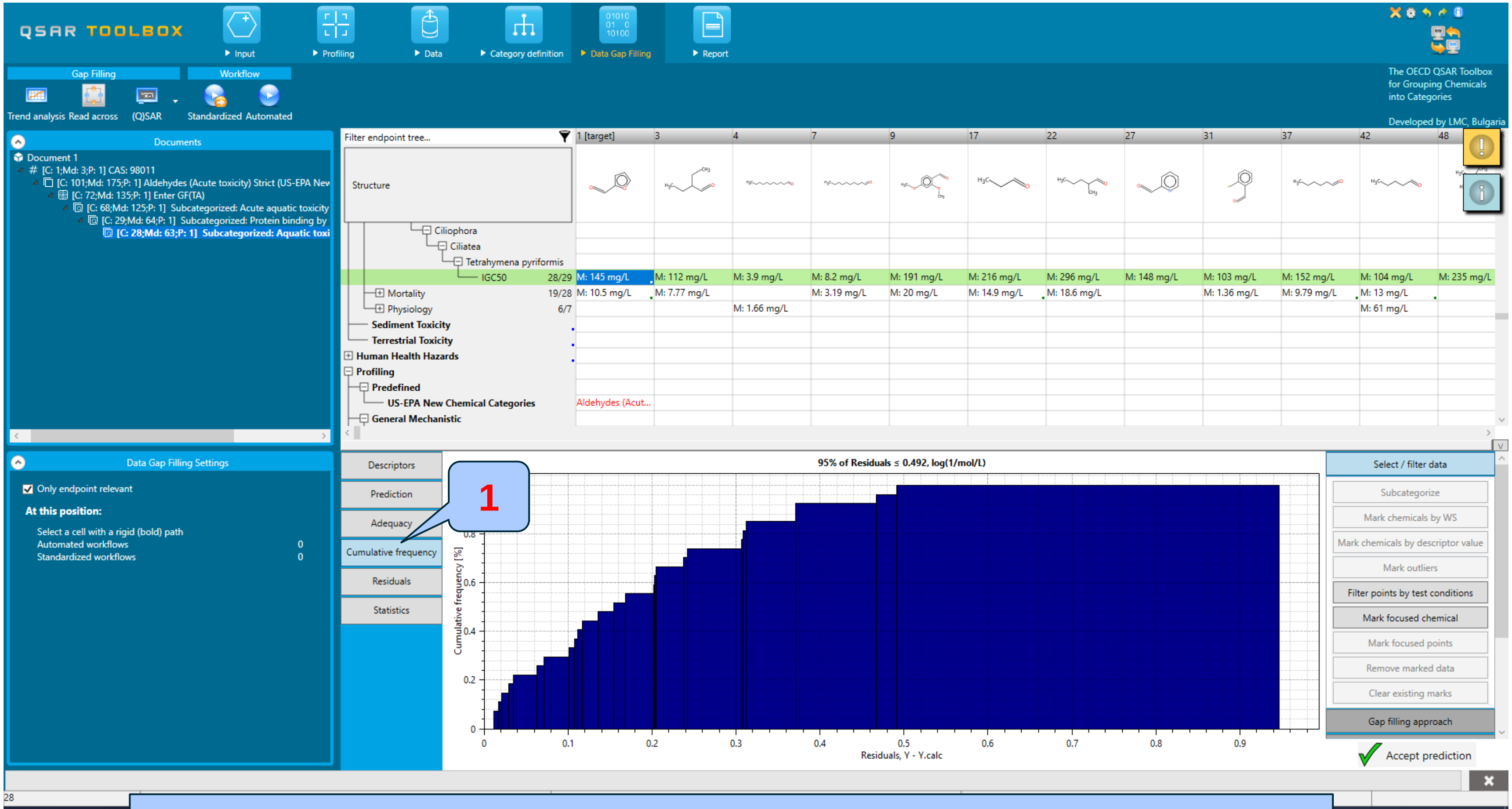
Select a cell with a rigid (bold) path

Automated workflows	0
Standardized workflows	0

		1 [target]	3	4	7	9	17	22	27	31	37	42	48
Ciliophora													
Ciliatea													
Tetrahymena pyriformis													
IGC50	28/29	M: 145 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 8.2 mg/L	M: 191 mg/L	M: 216 mg/L	M: 296 mg/L	M: 148 mg/L	M: 103 mg/L	M: 152 mg/L	M: 104 mg/L	M: 235 mg/L
Mortality	19/28	M: 10.5 mg/L	M: 7.77 mg/L		M: 3.19 mg/L	M: 20 mg/L	M: 14.9 mg/L	M: 18.6 mg/L		M: 1.36 mg/L	M: 9.79 mg/L	M: 13 mg/L	
Physiology	6/7			M: 1.66 mg/L								M: 61 mg/L	
Sediment Toxicity													
Terrestrial Toxicity													
Human Health Hazards													
Profiling													
Predefined													
US-EPA New Chemical Categories		Aldehydes (Acut...											
General Mechanistic													

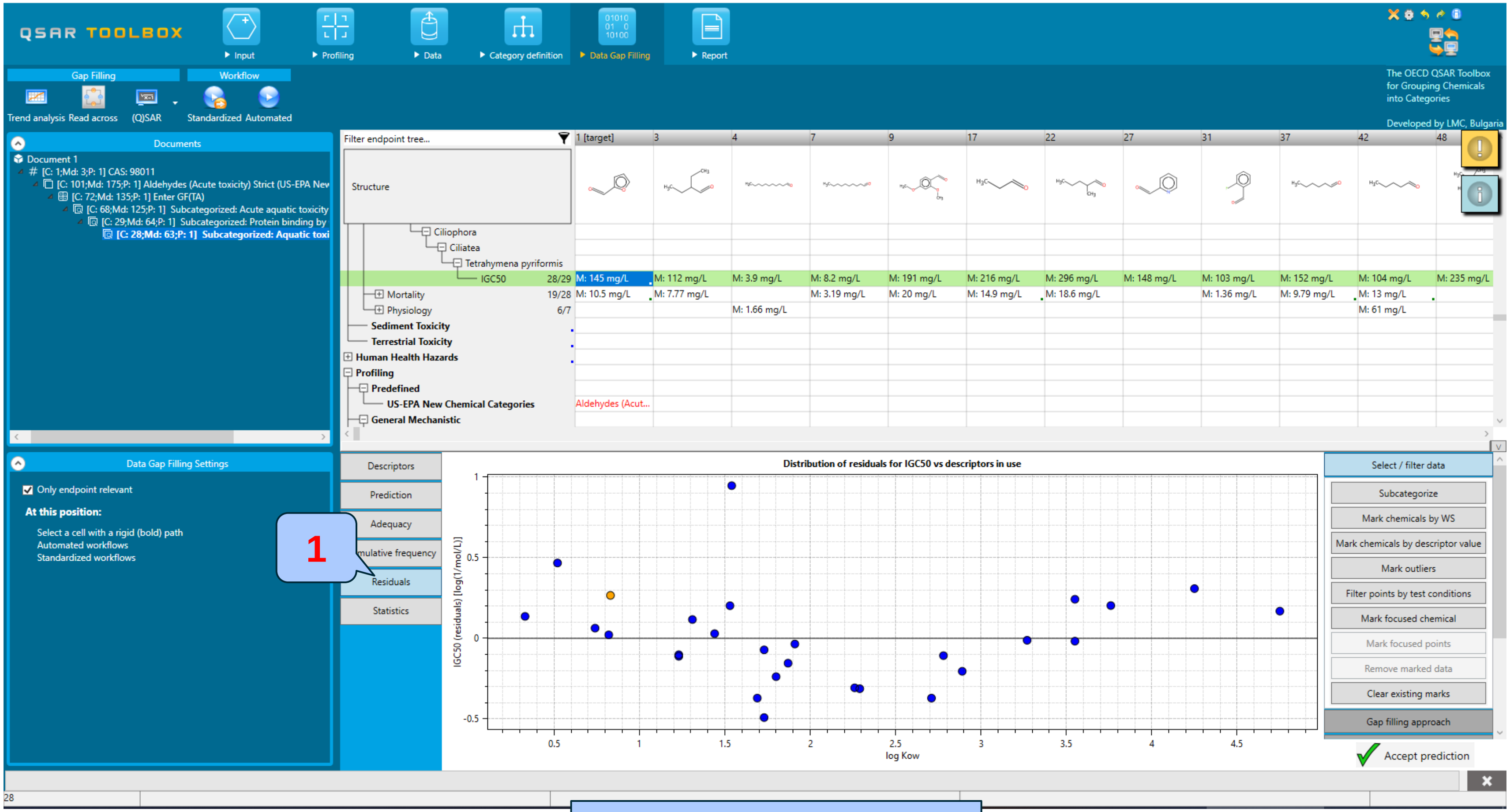
1. Position on the last level of document tree; 2. Click **"Adequacy"**;

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 experimental error.

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



Document 1
- [C: 1;Md: 3;P: 1] CAS: 98011
- [C: 101;Md: 175;P: 1] Aldehydes (Acute toxicity) Strict (US-EPA New
- [C: 72;Md: 135;P: 1] Enter GF(TA)
- [C: 68;Md: 125;P: 1] Subcategorized: Acute aquatic toxicity
- [C: 29;Md: 64;P: 1] Subcategorized: Protein binding by
- **[C: 28;Md: 63;P: 1] Subcategorized: Aquatic toxi**

☑ Only endpoint relevant

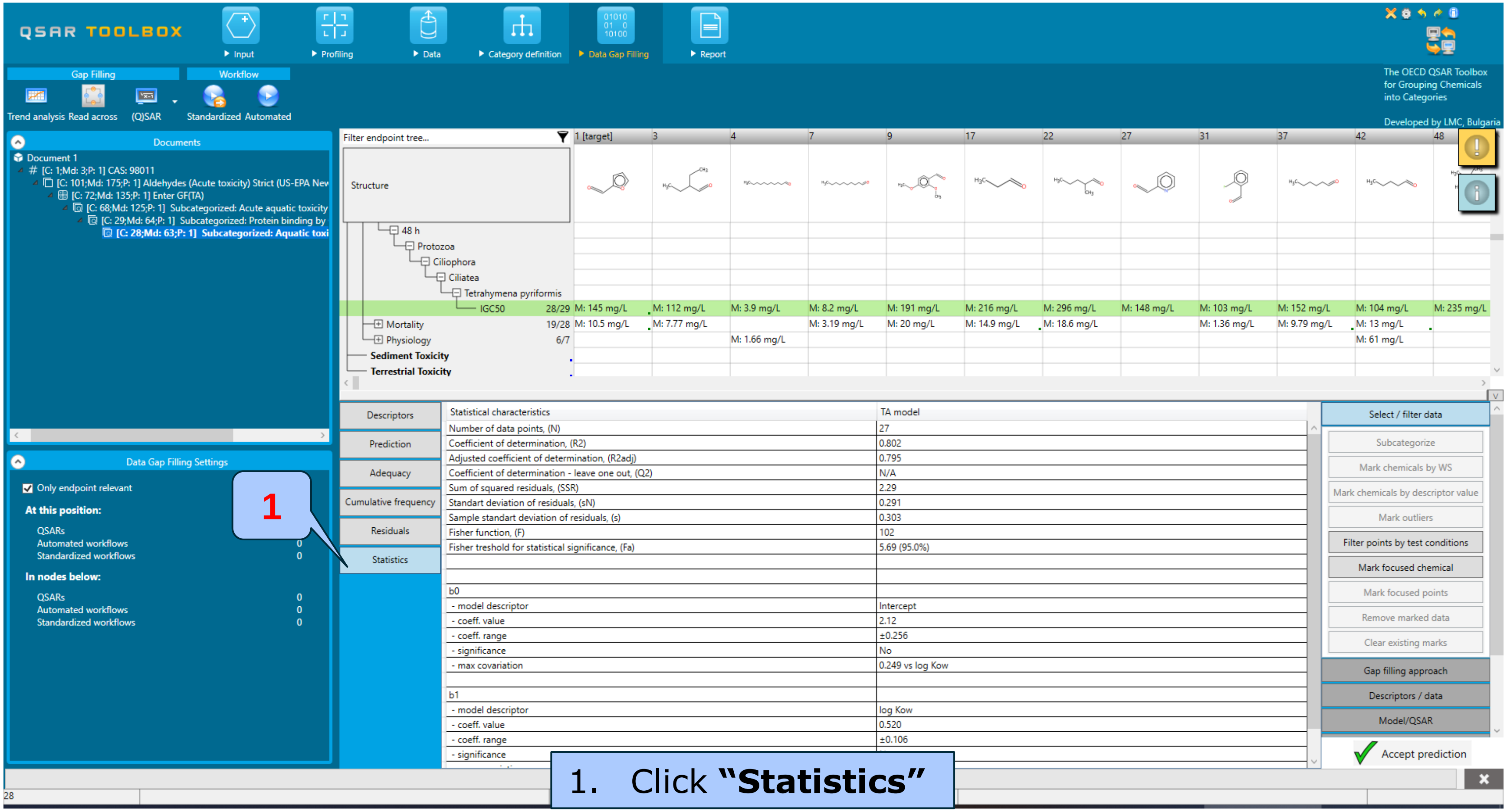
At this position:

Select a cell with a rigid (bold) path
Automated workflows
Standardized workflows

Filter endpoint tree...		1 [target]	3	4	7	9	17	22	27	31	37	42	48
Structure													
Ciliophora													
Ciliatea													
Tetrahymena pyriformis													
IGC50	28/29	M: 145 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 8.2 mg/L	M: 191 mg/L	M: 216 mg/L	M: 296 mg/L	M: 148 mg/L	M: 103 mg/L	M: 152 mg/L	M: 104 mg/L	M: 235 mg/L
Mortality	19/28	M: 10.5 mg/L	M: 7.77 mg/L		M: 3.19 mg/L	M: 20 mg/L	M: 14.9 mg/L	M: 18.6 mg/L		M: 1.36 mg/L	M: 9.79 mg/L	M: 13 mg/L	
Physiology	6/7			M: 1.66 mg/L								M: 61 mg/L	
Sediment Toxicity													
Terrestrial Toxicity													
Human Health Hazards													
Profiling													
Predefined													
US-EPA New Chemical Categories		Aldehydes (Acut...											
General Mechanistic													

1. Click **"Residuals"**

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



Document 1
- # [C: 1;Md: 3;P: 1] CAS: 98011
  - [C: 101;Md: 175;P: 1] Aldehydes (Acute toxicity) Strict (US-EPA New
    - [C: 72;Md: 135;P: 1] Enter GF(TA)
      - [C: 68;Md: 125;P: 1] Subcategorized: Acute aquatic toxicity
        - [C: 29;Md: 64;P: 1] Subcategorized: Protein binding by
          - **[C: 28;Md: 63;P: 1] Subcategorized: Aquatic toxi**

☑ Only endpoint relevant

At this position:

QSARs
Automated workflows 0
Standardized workflows 0

In nodes below:

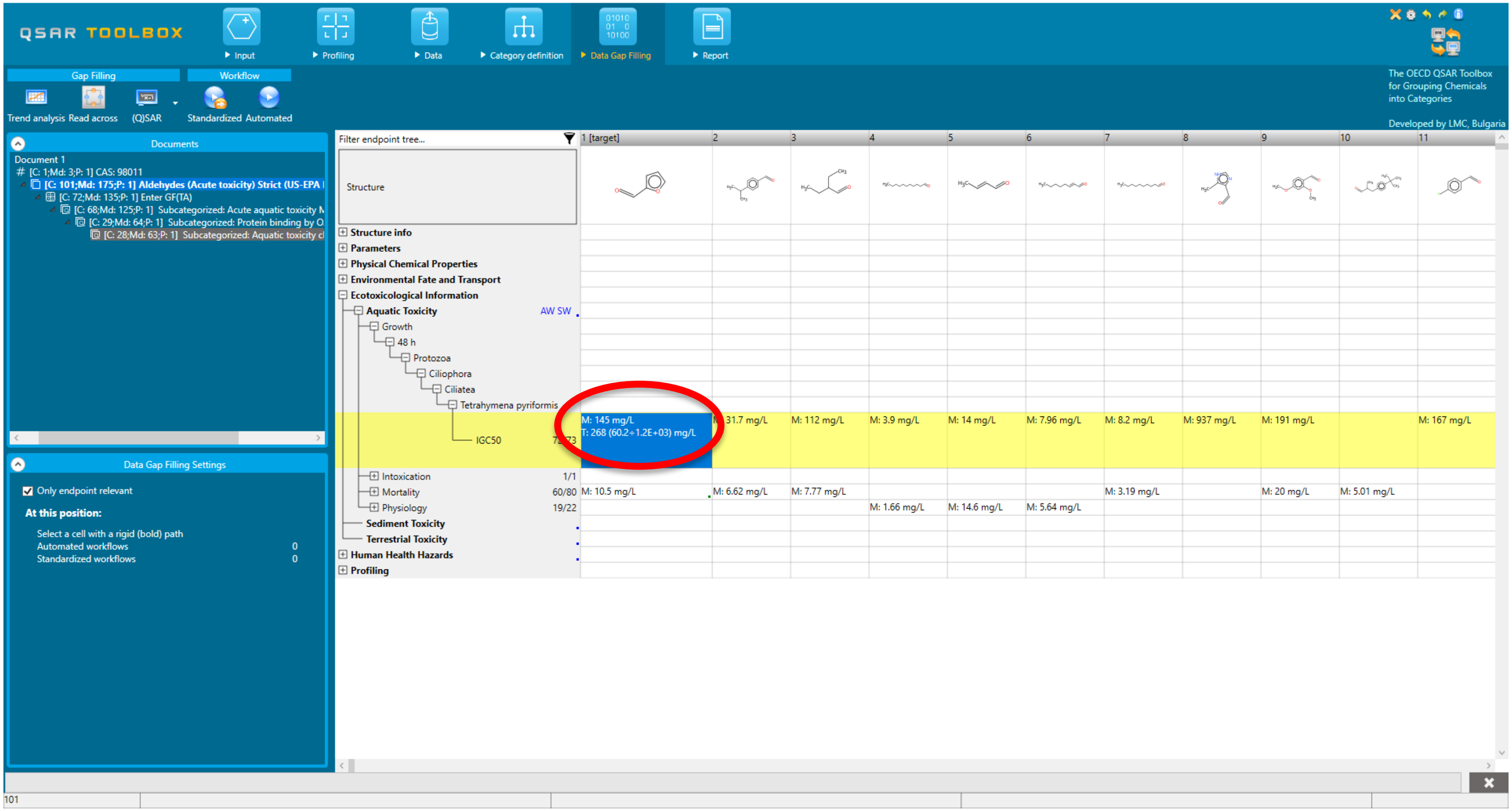
QSARs 0
Automated workflows 0
Standardized workflows 0

Structure		1 [target]	3	4	7	9	17	22	27	31	37	42	48
48 h													
Protozoa													
Ciliophora													
Ciliatea													
Tetrahymena pyriformis													
IGC50	28/29	M: 145 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 8.2 mg/L	M: 191 mg/L	M: 216 mg/L	M: 296 mg/L	M: 148 mg/L	M: 103 mg/L	M: 152 mg/L	M: 104 mg/L	M: 235 mg/L
Mortality	19/28	M: 10.5 mg/L	M: 7.77 mg/L		M: 3.19 mg/L	M: 20 mg/L	M: 14.9 mg/L	M: 18.6 mg/L		M: 1.36 mg/L	M: 9.79 mg/L	M: 13 mg/L	
Physiology	6/7			M: 1.66 mg/L								M: 61 mg/L	
Sediment Toxicity													
Terrestrial Toxicity													

Statistical characteristics	TA model
Number of data points, (N)	27
Coefficient of determination, (R2)	0.802
Adjusted coefficient of determination, (R2adj)	0.795
Coefficient of determination - leave one out, (Q2)	N/A
Sum of squared residuals, (SSR)	2.29
Standart deviation of residuals, (sN)	0.291
Sample standart deviation of residuals, (s)	0.303
Fisher function, (F)	102
Fisher treshold for statistical significance, (Fa)	5.69 (95.0%)
b0	
- model descriptor	Intercept
- coeff. value	2.12
- coeff. range	±0.256
- significance	No
- max covariation	0.249 vs log Kow
b1	
- model descriptor	log Kow
- coeff. value	0.520
- coeff. range	±0.106
- significance	

1. Click **"Statistics"**

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



Document 1
- # [C: 1;Md: 3;P: 1] CAS: 98011
- **[C: 101;Md: 175;P: 1] Aldehydes (Acute toxicity) Strict (US-EPA I**
  - [C: 72;Md: 135;P: 1] Enter GF(TA)
    - [C: 68;Md: 125;P: 1] Subcategorized: Acute aquatic toxicity M
      - [C: 29;Md: 64;P: 1] Subcategorized: Protein binding by O
        - [C: 28;Md: 63;P: 1] Subcategorized: Aquatic toxicity cl

☑ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

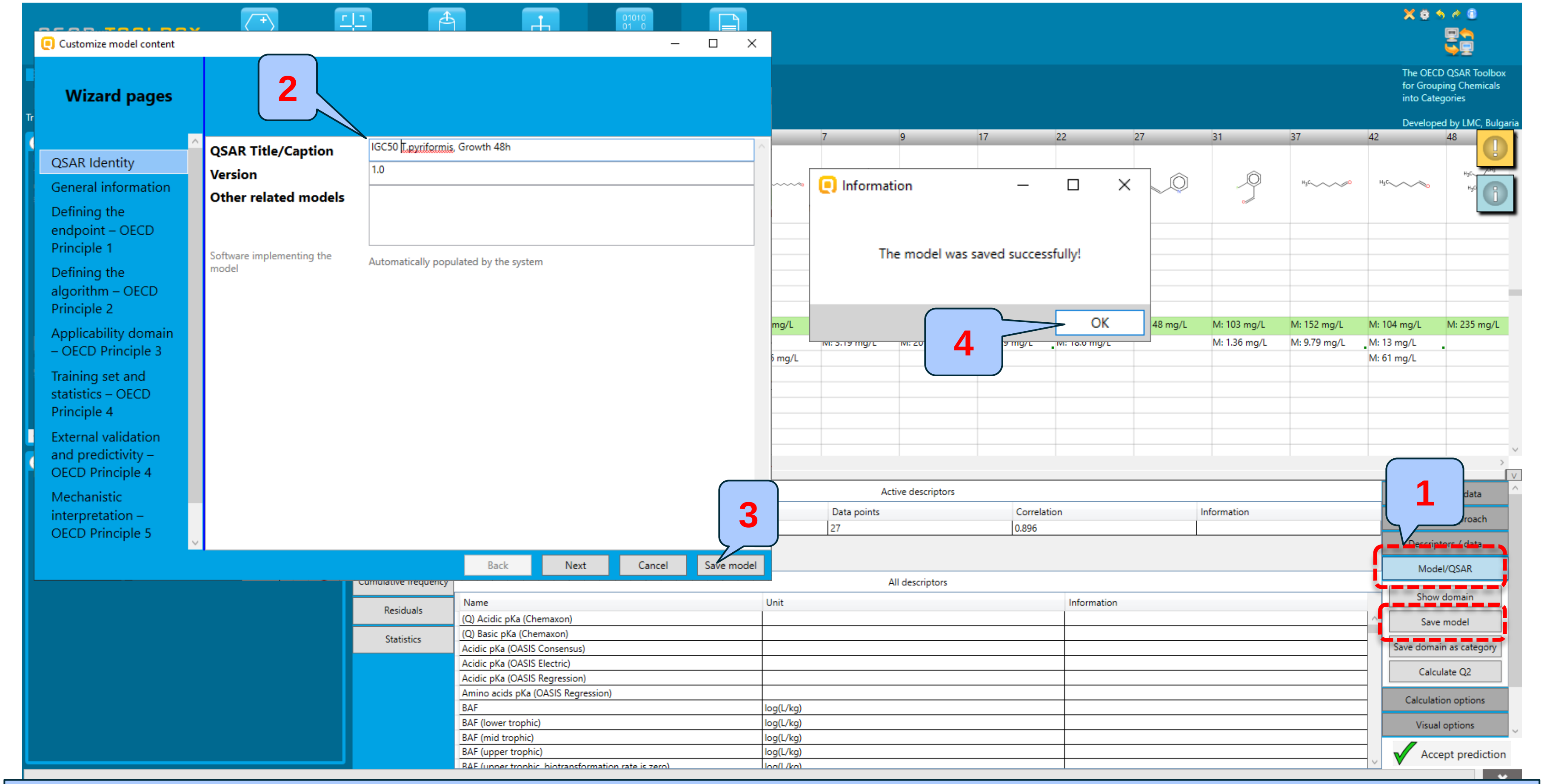
		1 [target]	2	3	4	5	6	7	8	9	10	11
Structure												
⊞ **Structure info**												
⊞ **Parameters**												
⊞ **Physical Chemical Properties**												
⊞ **Environmental Fate and Transport**												
⊟ **Ecotoxicological Information**												
⊟ **Aquatic Toxicity**	AW SW											
⊟ Growth												
⊟ 48 h												
⊟ Protozoa												
⊟ Ciliophora												
⊟ Ciliatea												
⊟ Tetrahymena pyriformis												
IGC50	7[illegible]73	M: 145 mg/L T: 268 (60.2÷1.2E+03) mg/L	M: 31.7 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 14 mg/L	M: 7.96 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L		M: 167 mg/L
⊞ Intoxication	1/1											
⊞ Mortality	60/80	M: 10.5 mg/L	M: 6.62 mg/L	M: 7.77 mg/L				M: 3.19 mg/L		M: 20 mg/L	M: 5.01 mg/L	
⊞ Physiology	19/22				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L					
Sediment Toxicity												
Terrestrial Toxicity												
⊞ **Human Health Hazards**												
⊞ **Profiling**												

101

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



Customize model content

Wizard pages

- QSAR Identity
- General information
- Defining the endpoint – OECD Principle 1
- Defining the algorithm – OECD Principle 2
- Applicability domain – OECD Principle 3
- Training set and statistics – OECD Principle 4
- External validation and predictivity – OECD Principle 4
- Mechanistic interpretation – OECD Principle 5

QSAR Title/Caption: IGC50 T.pyriformis, Growth 48h

Version: 1.0

Other related models

Software implementing the model: Automatically populated by the system

Information

The model was saved successfully!

	Data points	Correlation	Information
	27	0.896	

All descriptors

Name	Unit	Information
(Q) Acidic pKa (Chemaxon)		
(Q) Basic pKa (Chemaxon)		
Acidic pKa (OASIS Consensus)		
Acidic pKa (OASIS Electric)		
Acidic pKa (OASIS Regression)		
Amino acids pKa (OASIS Regression)		
BAF	log(L/kg)	
BAF (lower trophic)	log(L/kg)	
BAF (mid trophic)	log(L/kg)	
BAF (upper trophic)	log(L/kg)	
BAF (upper trophic, biotransformation rate is zero)	log(L/kg)	

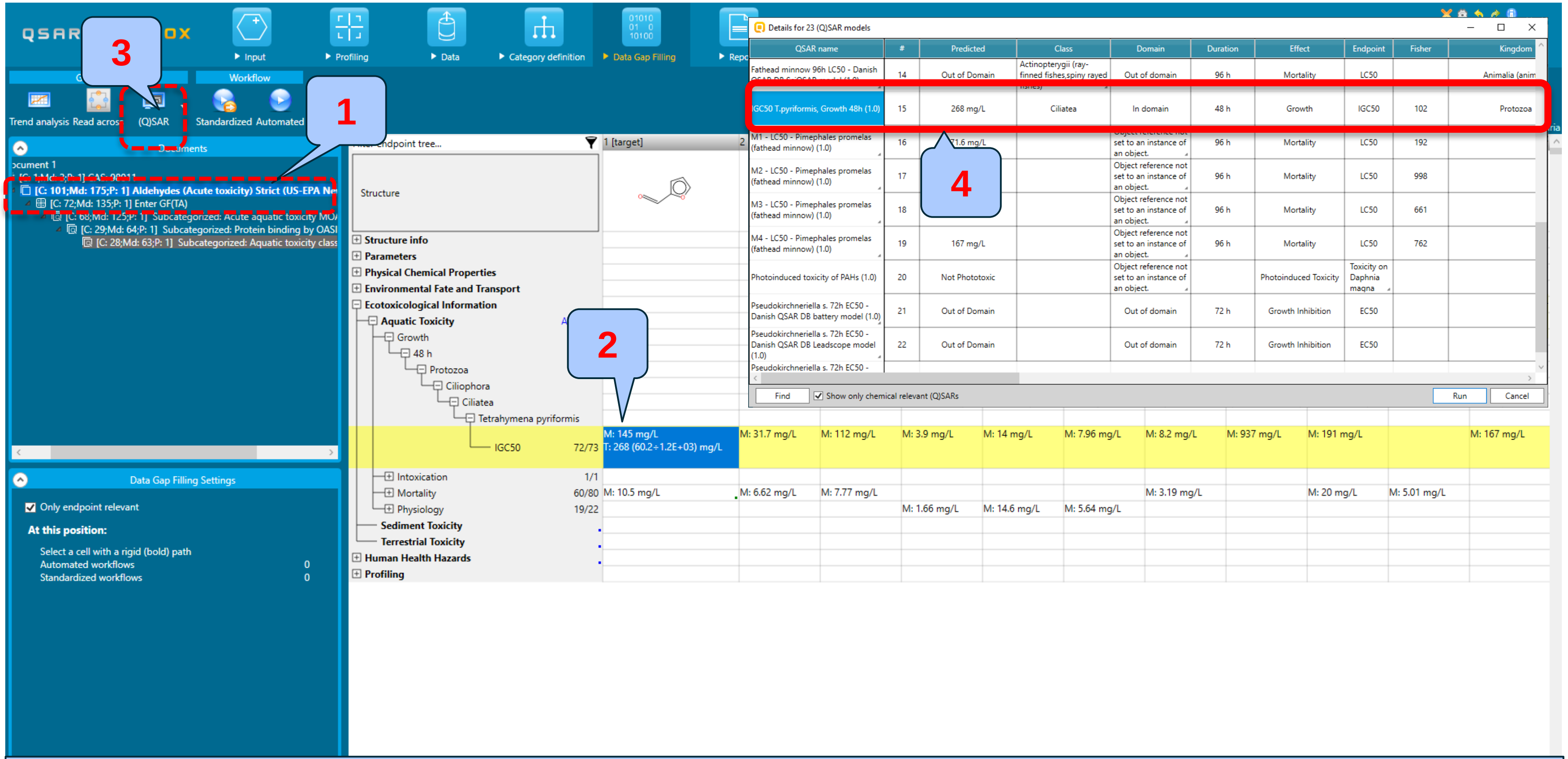
1. Click **"Model/QSAR"**, then **"Save model"; **2. Type name of the model and fill the fields in the wizard if necessary (Use Next/Back buttons to navigate within it); 3. Click **"Save model"**; 4. Click **OK.**

Outlook

- Background
- Keywords
- 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?



Details for 23 (Q)SAR models

QSAR name	#	Predicted	Class	Domain	Duration	Effect	Endpoint	Fisher	Kingdom
Fathead minnow 96h LC50 - Danish QSAR DB SciQSAR model (1.0)	14	Out of Domain	Actinopterygii (ray-finned fishes,spiny rayed fishes)	Out of domain	96 h	Mortality	LC50		Animalia (anim
IGC50 T.pyriformis, Growth 48h (1.0)	15	268 mg/L	Ciliatea	In domain	48 h	Growth	IGC50	102	Protozoa
M1 - LC50 - Pimephales promelas (fathead minnow) (1.0)	16	71.6 mg/L		Object reference not set to an instance of an object.	96 h	Mortality	LC50	192	
M2 - LC50 - Pimephales promelas (fathead minnow) (1.0)	17			Object reference not set to an instance of an object.	96 h	Mortality	LC50	998	
M3 - LC50 - Pimephales promelas (fathead minnow) (1.0)	18			Object reference not set to an instance of an object.	96 h	Mortality	LC50	661	
M4 - LC50 - Pimephales promelas (fathead minnow) (1.0)	19	167 mg/L		Object reference not set to an instance of an object.	96 h	Mortality	LC50	762	
Photoinduced toxicity of PAHs (1.0)	20	Not Phototoxic		Object reference not set to an instance of an object.		Photoinduced Toxicity	Toxicity on Daphnia magna		
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB battery model (1.0)	21	Out of Domain		Out of domain	72 h	Growth Inhibition	EC50		
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB Leadscope model (1.0)	22	Out of Domain		Out of domain	72 h	Growth Inhibition	EC50		
Pseudokirchneriella s. 72h EC50 -									

Find ☑ Show only chemical relevant (Q)SARs Run Cancel

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

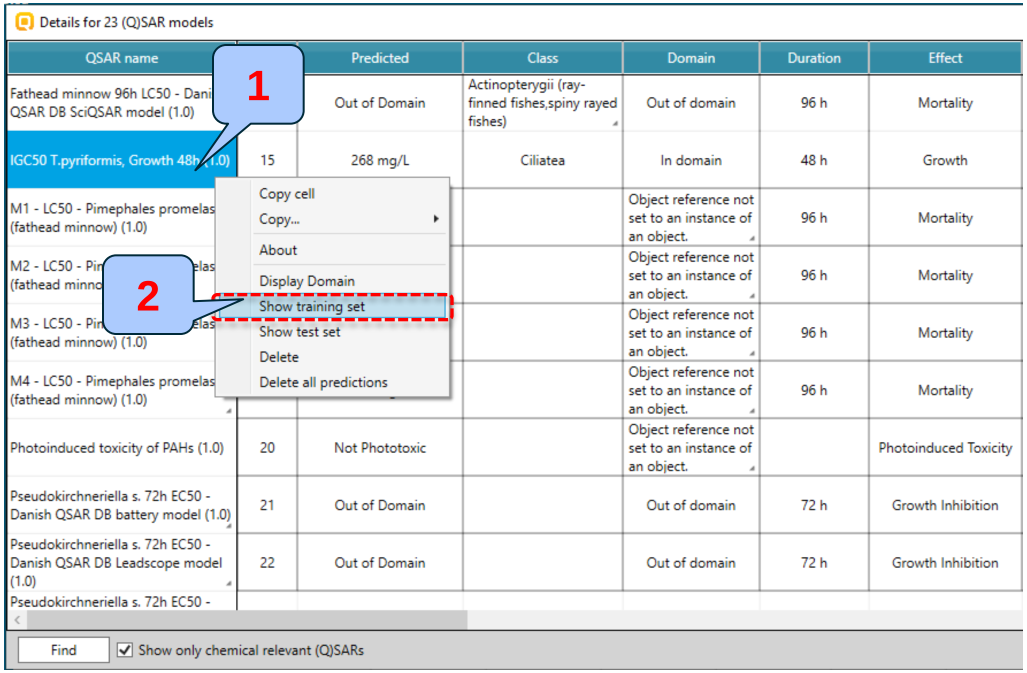
Data Gap Filling
How to see the derived QSAR?

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 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

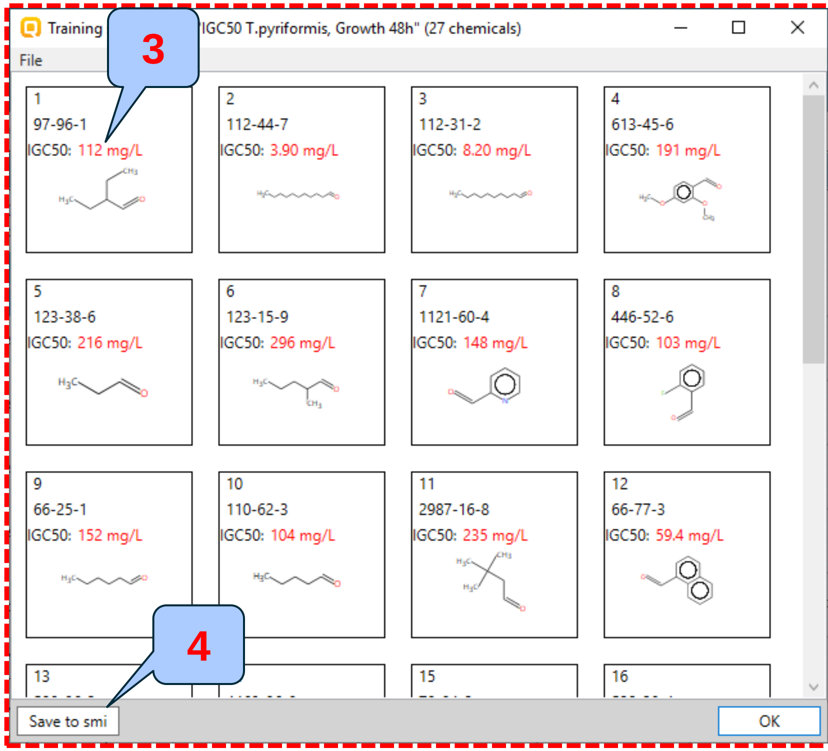
Visualisation of the training set



Details for 23 (Q)SAR models

QSAR name		Predicted	Class	Domain	Duration	Effect
Fathead minnow 96h LC50 - Dani QSAR DB SciQSAR model (1.0)		Out of Domain	Actinopterygii (ray-finned fishes,spiny rayed fishes)	Out of domain	96 h	Mortality
IGC50 T.pyriformis, Growth 48h (1.0)	15	268 mg/L	Ciliatea	In domain	48 h	Growth
M1 - LC50 - Pimephales promelas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality
M2 - LC50 - Pi elas (fathead minno				Object reference not set to an instance of an object.	96 h	Mortality
M3 - LC50 - Pi elas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality
M4 - LC50 - Pimephales promelas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality
Photoinduced toxicity of PAHs (1.0)	20	Not Phototoxic		Object reference not set to an instance of an object.		Photoinduced Toxicity
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB battery model (1.0)	21	Out of Domain		Out of domain	72 h	Growth Inhibition
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB Leadscope model (1.0)	22	Out of Domain		Out of domain	72 h	Growth Inhibition
Pseudokirchneriella s. 72h EC50 -						

☑ Show only chemical relevant (Q)SARs



Training "IGC50 T.pyriformis, Growth 48h" (27 chemicals)

File

1 97-96-1 IGC50: 112 mg/L

2 112-44-7 IGC50: 3.90 mg/L

3 112-31-2 IGC50: 8.20 mg/L

4 613-45-6 IGC50: 191 mg/L

5 123-38-6 IGC50: 216 mg/L

6 123-15-9 IGC50: 296 mg/L

7 1121-60-4 IGC50: 148 mg/L

8 446-52-6 IGC50: 103 mg/L

9 66-25-1 IGC50: 152 mg/L

10 110-62-3 IGC50: 104 mg/L

11 2987-16-8 IGC50: 235 mg/L

12 66-77-3 IGC50: 59.4 mg/L

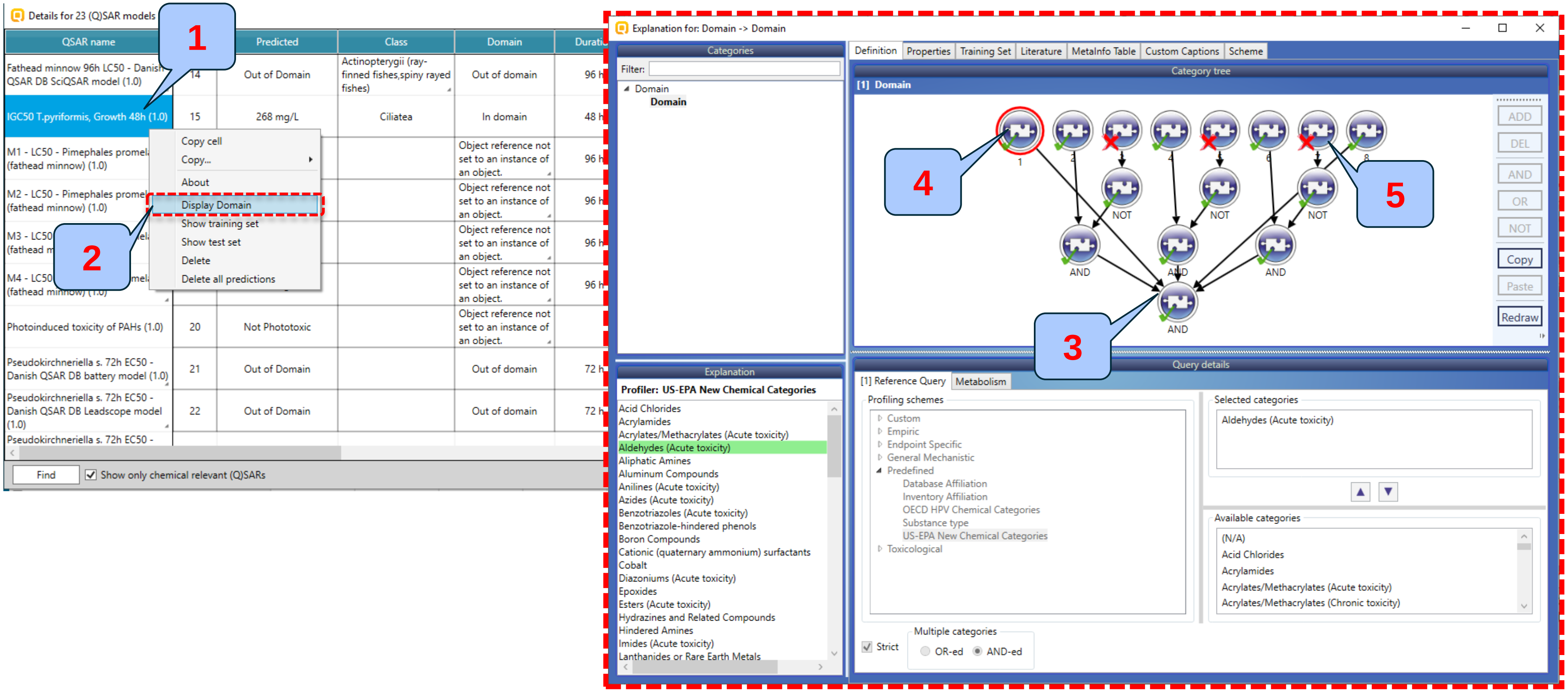
13

15

16

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



Details for 23 (Q)SAR models

QSAR name		Predicted	Class	Domain	Durati
Fathead minnow 96h LC50 - Danish QSAR DB SciQSAR model (1.0)	14	Out of Domain	Actinopterygii (ray-finned fishes,spiny rayed fishes)	Out of domain	96 h
IGC50 T.pyriformis, Growth 48h (1.0)	15	268 mg/L	Ciliatea	In domain	48 h
M1 - LC50 - Pimephales promel (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h
M2 - LC50 - Pimephales promel (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h
M3 - LC50 (fathead m				Object reference not set to an instance of an object.	96 h
M4 - LC50 (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h
Photoinduced toxicity of PAHs (1.0)	20	Not Phototoxic		Object reference not set to an instance of an object.	
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB battery model (1.0)	21	Out of Domain		Out of domain	72 h
Pseudokirchneriella s. 72h EC50 - Danish QSAR DB Leadscope model (1.0)	22	Out of Domain		Out of domain	72 h
Pseudokirchneriella s. 72h EC50 -					

☑ Show only chemical relevant (Q)SARs

Profiler: US-EPA New Chemical Categories

Acid Chlorides
Acrylamides
Acrylates/Methacrylates (Acute toxicity)
Aldehydes (Acute toxicity)
Aliphatic Amines
Aluminum Compounds
Anilines (Acute toxicity)
Azides (Acute toxicity)
Benzotriazoles (Acute toxicity)
Benzotriazole-hindered phenols
Boron Compounds
Cationic (quaternary ammonium) surfactants
Cobalt
Diazoniums (Acute toxicity)
Epoxides
Esters (Acute toxicity)
Hydrazines and Related Compounds
Hindered Amines
Imides (Acute toxicity)
Lanthanides or Rare Earth Metals

Profiling schemes
- Custom
- Empiric
- Endpoint Specific
- General Mechanistic
- Predefined
  - Database Affiliation
  - Inventory Affiliation
  - OECD HPV Chemical Categories
  - Substance type
  - US-EPA New Chemical Categories
- Toxicological

Selected categories
Aldehydes (Acute toxicity)

Available categories
(N/A)
Acid Chlorides
Acrylamides
Acrylates/Methacrylates (Acute toxicity)
Acrylates/Methacrylates (Chronic toxicity)

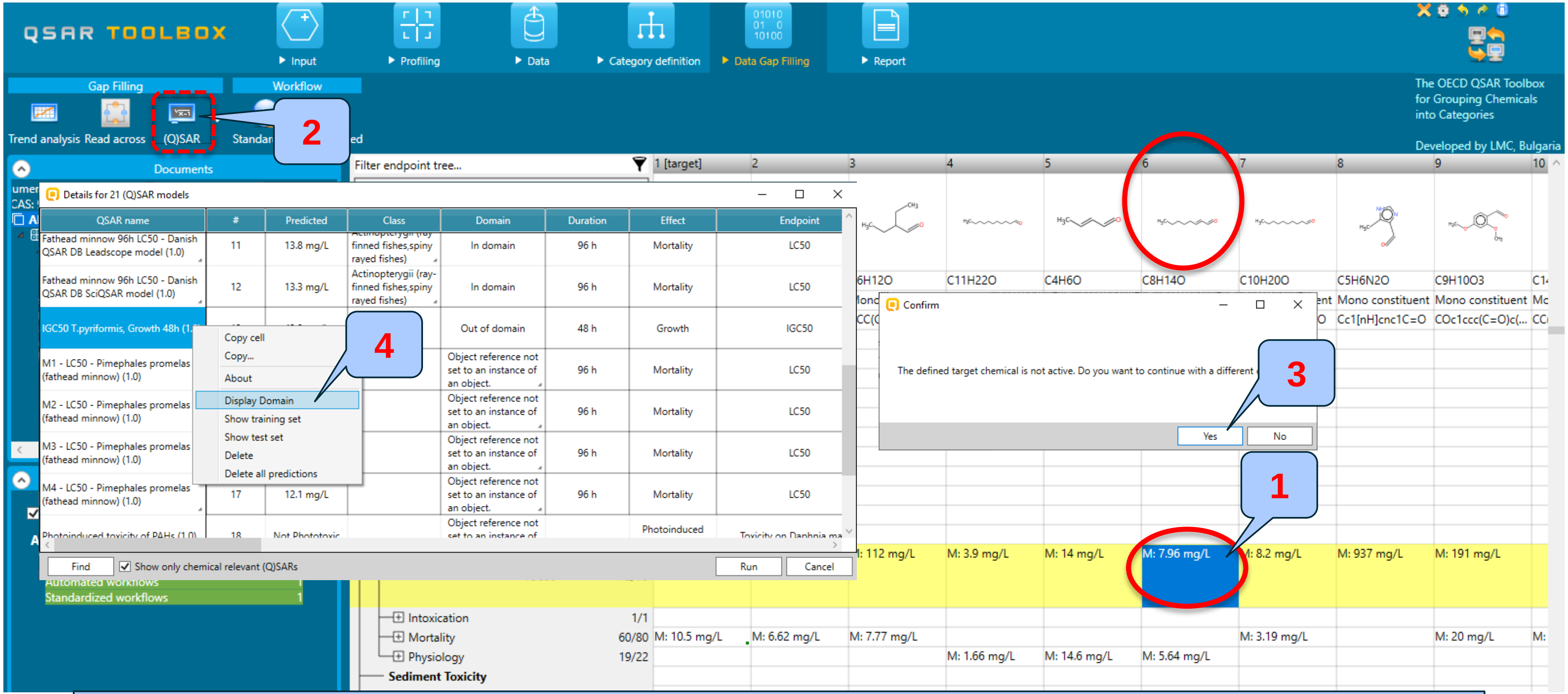
☑ Strict

Multiple categories
○ OR-ed ● AND-ed

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 of whether a chemical is in the domain of the model



Details for 21 (Q)SAR models

QSAR name	#	Predicted	Class	Domain	Duration	Effect	Endpoint
Fathead minnow 96h LC50 - Danish QSAR DB Leadscope model (1.0)	11	13.8 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	In domain	96 h	Mortality	LC50
Fathead minnow 96h LC50 - Danish QSAR DB SciQSAR model (1.0)	12	13.3 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	In domain	96 h	Mortality	LC50
IGC50 T.pyriformis, Growth 48h (1.				Out of domain	48 h	Growth	IGC50
M1 - LC50 - Pimephales promelas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality	LC50
M2 - LC50 - Pimephales promelas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality	LC50
M3 - LC50 - Pimephales promelas (fathead minnow) (1.0)				Object reference not set to an instance of an object.	96 h	Mortality	LC50
M4 - LC50 - Pimephales promelas (fathead minnow) (1.0)	17	12.1 mg/L		Object reference not set to an instance of an object.	96 h	Mortality	LC50
Photoinduced toxicity of PAHs (1.0)	18	Not Phototoxic		Object reference not set to an instance of		Photoinduced	Toxicity on Daphnia ma

Confirm

The defined target chemical is not active. Do you want to continue with a different

1. Highlight the cell of one of the analogues (e.g., chemical # 6 in the data matrix; 2. Click on **"(Q)SAR"**; 3. A message informs you that the QSAR is applied not on the target chemical. Click **Yes**; 4. **Right click** above the model and **Left click** on Display domain (see next screen shot).

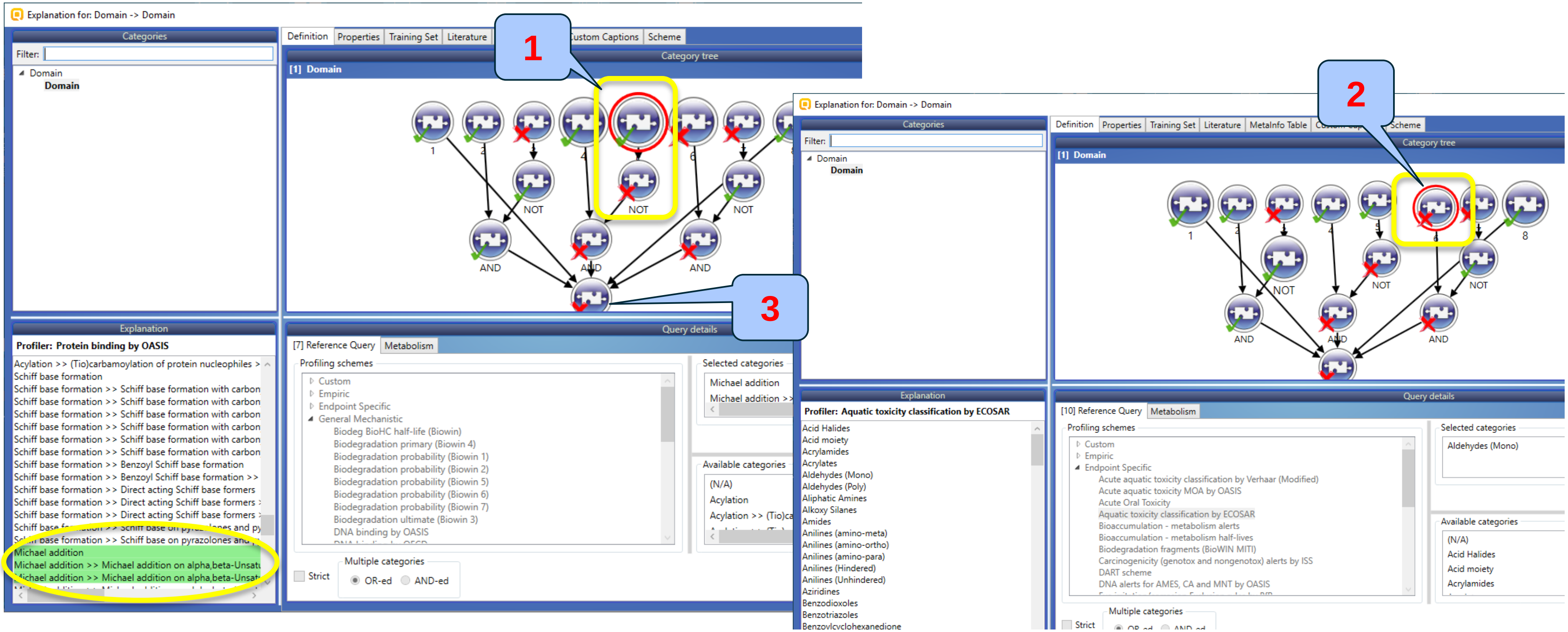
Data Gap Filling

Visualisation of 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) and to be not “reactive unspecified” (boundary 3)
- It can react with protein by Schiff-base formation (boundary 4) and should not belong to any of the eliminated mechanistic domains according to Protein binding by OASIS (boundary 5):
  - Michael addition (α,β-Aldehydes, Conjugated systems with electron withdrawing groups) (boundary 5)
  - SNAr (Activated aryl and heteroaryl compounds) (boundary 5)
  - Schiff base formation (Bis aldehydes, Di-substituted α,β-unsaturated aldehydes and Aromatic carbonyl compounds) (boundary 5)
- The chemical should be an “aldehyde” as required by Aquatic toxicity classification by ECOSAR (boundary 6) and not to be “imidazoles” (boundary 7).
- Another requirement is Log Kow to be >=0.308 and <= 4.77 (boundary 8):

Data Gap Filling

Visualisation of whether a chemical is in the domain of the model



Profiler: Protein binding by OASIS

Acylation >> (Tio)carbamoylation of protein nucleophiles >
Schiff base formation
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Schiff base formation with carbon
Schiff base formation >> Benzoyl Schiff base formation
Schiff base formation >> Benzoyl Schiff base formation >>
Schiff base formation >> Direct acting Schiff base formers
Schiff base formation >> Direct acting Schiff base formers >
Schiff base formation >> Direct acting Schiff base formers >
Schiff base formation >> Schiff base on pyrazolones and py
Schiff base formation >> Schiff base on pyrazolones and p
Michael addition
Michael addition >> Michael addition on alpha,beta-Unsatu
Michael addition >> Michael addition on alpha,beta-Unsat

Profiling schemes
- Custom
- Empiric
- Endpoint Specific
- General Mechanistic
  - Biodeg BioHC half-life (Biowin)
  - Biodegradation primary (Biowin 4)
  - Biodegradation probability (Biowin 1)
  - Biodegradation probability (Biowin 2)
  - Biodegradation probability (Biowin 5)
  - Biodegradation probability (Biowin 6)
  - Biodegradation probability (Biowin 7)
  - Biodegradation ultimate (Biowin 3)
  - DNA binding by OASIS

Selected categories
Michael addition
Michael addition >>

Available categories
(N/A)
Acylation
Acylation >> (Tio)ca

□ Strict
Multiple categories
● OR-ed ○ AND-ed

Profiler: 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

Profiling schemes
- Custom
- Empiric
- Endpoint Specific
  - Acute aquatic toxicity classification by Verhaar (Modified)
  - Acute aquatic toxicity MOA by OASIS
  - Acute Oral Toxicity
  - Aquatic toxicity classification by ECOSAR
  - Bioaccumulation - metabolism alerts
  - Bioaccumulation - metabolism half-lives
  - Biodegradation fragments (BioWIN MITI)
  - Carcinogenicity (genotox and nongenotox) alerts by ISS
  - DART scheme
  - DNA alerts for AMES, CA and MNT by OASIS

Selected categories
Aldehydes (Mono)

Available categories
(N/A)
Acid Halides
Acid moiety
Acrylamides

□ Strict
Multiple categories
● OR-ed ○ AND-ed

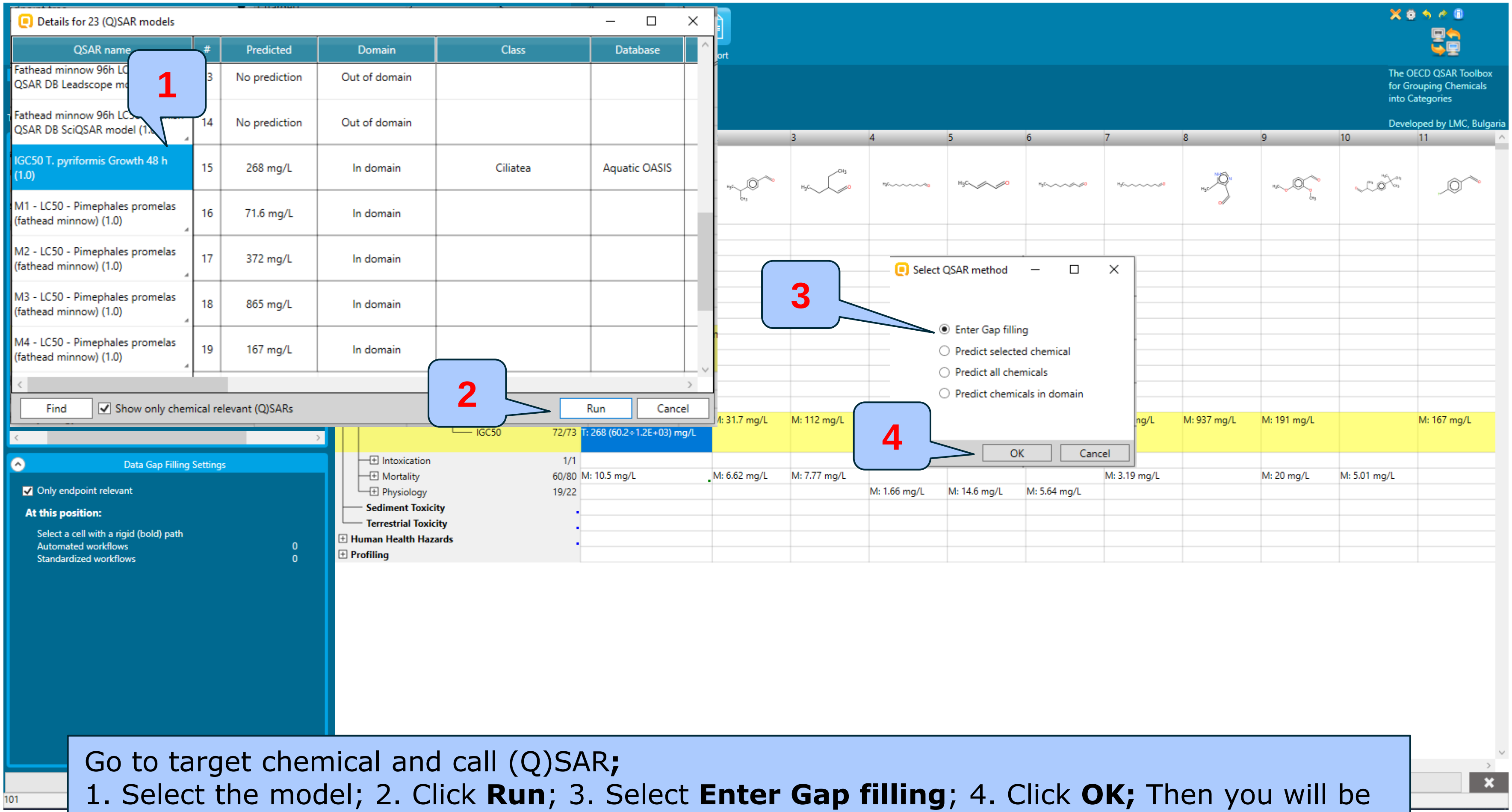
The target chemical is out of the model domain due to:

1) Belonging to "Michael addition" mechanism by "Protein binding by OASIS" profiler, which have been eliminated from the domain (negated by logical "NOT") (boundary 5)
2) The chemical is not an "aldehyde" as requested by ECOSAR profiler (boundary 6).

⚠ 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



Details for 23 (Q)SAR models

QSAR name	#	Predicted	Domain	Class	Database
Fathead minnow 96h LC QSAR DB Leadscope m	3	No prediction	Out of domain		
Fathead minnow 96h LC QSAR DB SciQSAR model (1.	14	No prediction	Out of domain		
IGC50 T. pyriformis Growth 48 h (1.0)	15	268 mg/L	In domain	Ciliatea	Aquatic OASIS
M1 - LC50 - Pimephales promelas (fathead minnow) (1.0)	16	71.6 mg/L	In domain		
M2 - LC50 - Pimephales promelas (fathead minnow) (1.0)	17	372 mg/L	In domain		
M3 - LC50 - Pimephales promelas (fathead minnow) (1.0)	18	865 mg/L	In domain		
M4 - LC50 - Pimephales promelas (fathead minnow) (1.0)	19	167 mg/L	In domain		

Find ☑ Show only chemical relevant (Q)SARs Run Cancel

Select QSAR method

● Enter Gap filling
○ Predict selected chemical
○ Predict all chemicals
○ Predict chemicals in domain

OK Cancel

Data Gap Filling Settings

☑ Only endpoint relevant

At this position:

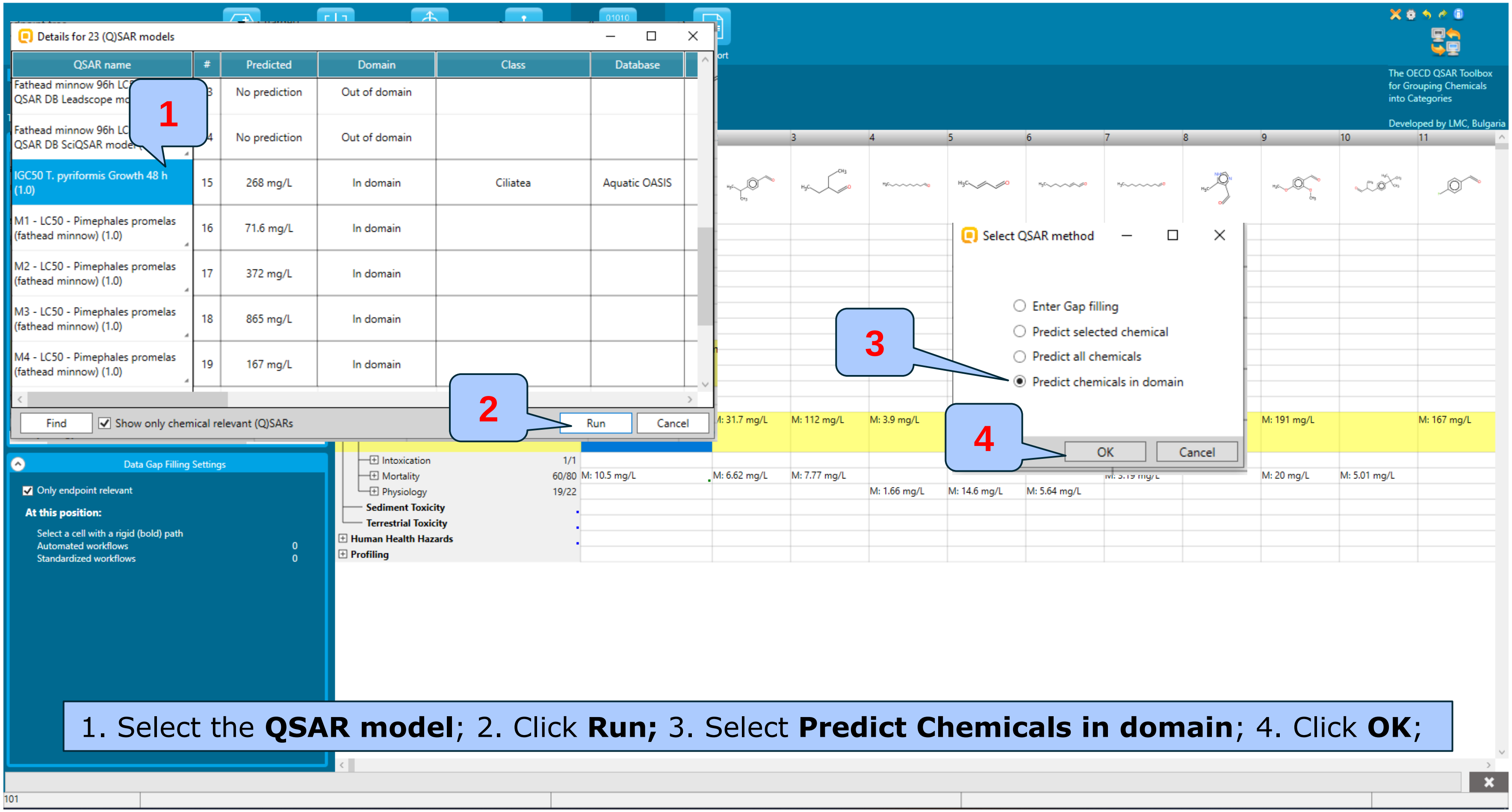
Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

IGC50 72/73 T: 268 (60.2÷1.2E+03) mg/L
Intoxication 1/1
Mortality 60/80 M: 10.5 mg/L
Physiology 19/22
Sediment Toxicity
Terrestrial Toxicity
Human Health Hazards
Profiling

Go to target chemical and call (Q)SAR**;**
1. Select the model; 2. Click **Run**; 3. Select **Enter Gap filling**; 4. Click **OK;** Then you will be transferred automatically to **Gap filling** and can operate (not shown);

Data Gap Filling

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



Details for 23 (Q)SAR models

QSAR name	#	Predicted	Domain	Class	Database
Fathead minnow 96h LC QSAR DB Leadscope mo	3	No prediction	Out of domain		
Fathead minnow 96h LC QSAR DB SciQSAR mode	4	No prediction	Out of domain		
IGC50 T. pyriformis Growth 48 h (1.0)	15	268 mg/L	In domain	Ciliatea	Aquatic OASIS
M1 - LC50 - Pimephales promelas (fathead minnow) (1.0)	16	71.6 mg/L	In domain		
M2 - LC50 - Pimephales promelas (fathead minnow) (1.0)	17	372 mg/L	In domain		
M3 - LC50 - Pimephales promelas (fathead minnow) (1.0)	18	865 mg/L	In domain		
M4 - LC50 - Pimephales promelas (fathead minnow) (1.0)	19	167 mg/L	In domain		

Find ☑ Show only chemical relevant (Q)SARs Run Cancel

Select QSAR method

○ Enter Gap filling
○ Predict selected chemical
○ Predict all chemicals
● Predict chemicals in domain

OK Cancel

Data Gap Filling Settings

☑ Only endpoint relevant

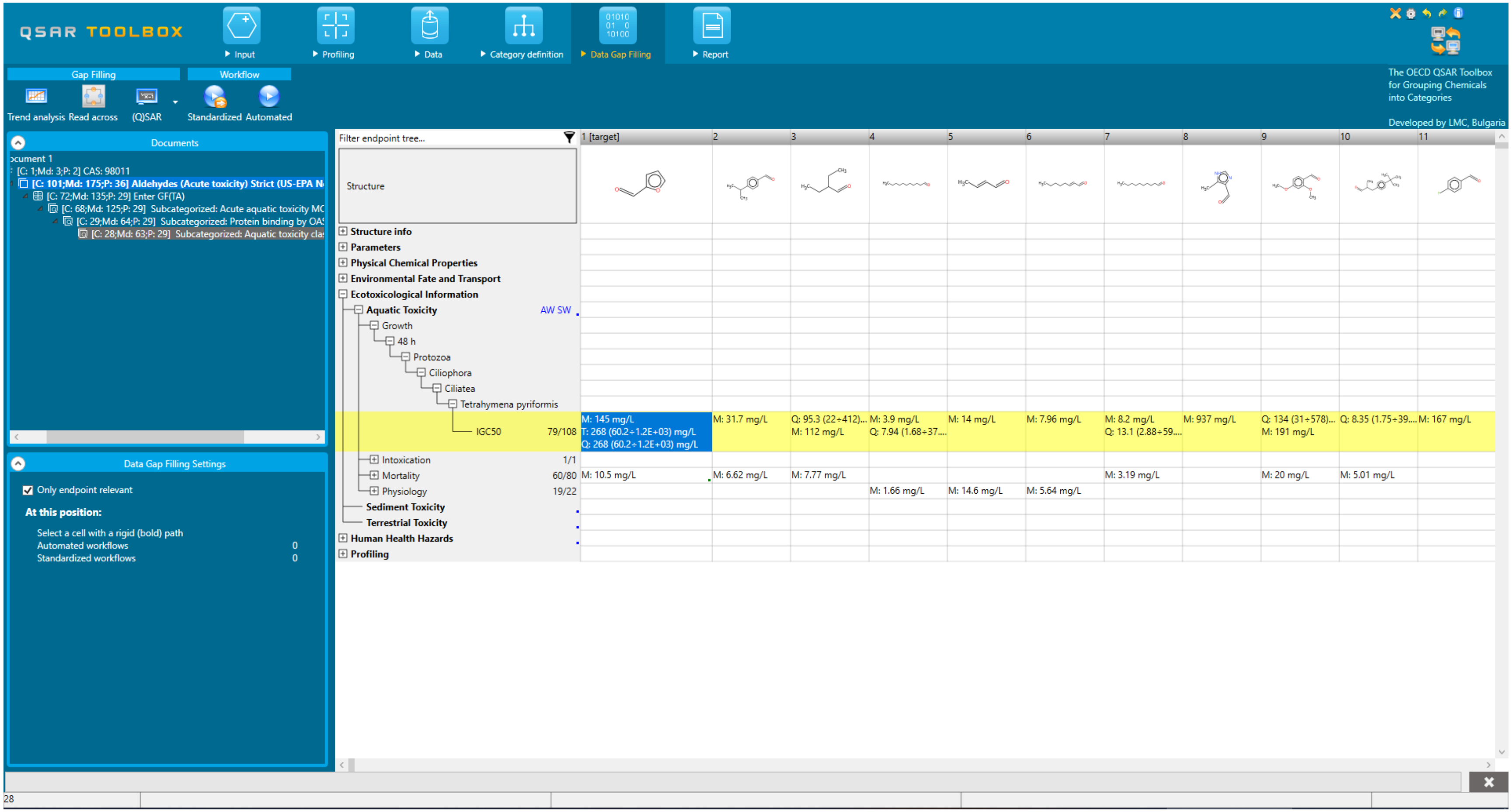
At this position:

Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

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

Data Gap Filling

Perform prediction for chemicals in domain



Documents

ocument 1
- [C: 1;Md: 3;P: 2] CAS: 98011
- **[C: 101;Md: 175;P: 36] Aldehydes (Acute toxicity) Strict (US-EPA N**
  - [C: 72;Md: 135;P: 29] Enter GF(TA)
    - [C: 68;Md: 125;P: 29] Subcategorized: Acute aquatic toxicity MC
      - [C: 29;Md: 64;P: 29] Subcategorized: Protein binding by OAS
        - [C: 28;Md: 63;P: 29] Subcategorized: Aquatic toxicity clas

Data Gap Filling Settings

☑ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

- **Structure info**
- **Parameters**
- **Physical Chemical Properties**
- **Environmental Fate and Transport**
- **Ecotoxicological Information**
  - **Aquatic Toxicity**
    - Growth
      - 48 h
        - Protozoa
          - Ciliophora
            - Ciliatea
              - Tetrahymena pyriformis
    - Intoxication
    - Mortality
    - Physiology
  - **Sediment Toxicity**
  - **Terrestrial Toxicity**
- **Human Health Hazards**
- **Profiling**

		1 [target]	2	3	4	5	6	7	8	9	10	11
IGC50	79/108	M: 145 mg/L T: 268 (60.2÷1.2E+03) mg/L Q: 268 (60.2÷1.2E+03) mg/L	M: 31.7 mg/L	Q: 95.3 (22÷412)... M: 112 mg/L	M: 3.9 mg/L Q: 7.94 (1.68÷37....	M: 14 mg/L	M: 7.96 mg/L	M: 8.2 mg/L Q: 13.1 (2.88÷59....	M: 937 mg/L	Q: 134 (31÷578)... M: 191 mg/L	Q: 8.35 (1.75÷39....	M: 167 mg/L
Intoxication	1/1											
Mortality	60/80	M: 10.5 mg/L	M: 6.62 mg/L	M: 7.77 mg/L				M: 3.19 mg/L		M: 20 mg/L	M: 5.01 mg/L	
Physiology	19/22				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L					

Outlook

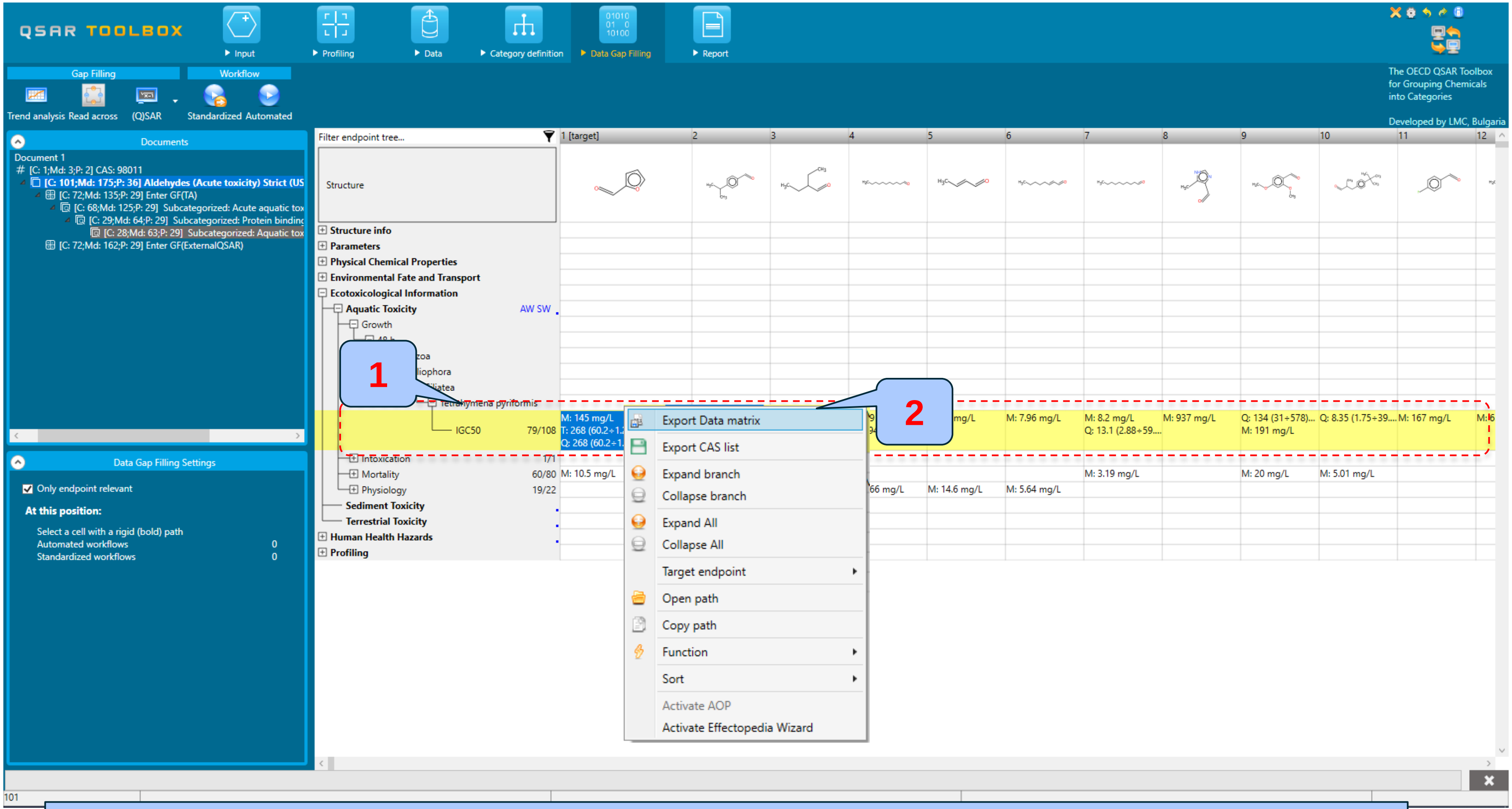
- Background
- Keywords
- Objectives
- The exercise
- **Workflow of the exercise**
  - Input
  - Profiling
  - Data
  - Category definition
  - Data gap filling
    - QSAR model
    - **Export QSAR prediction**

Export QSAR results

- The QSAR 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



Document 1
[C: 1;Md: 3;P: 2] CAS: 98011
[C: 101;Md: 175;P: 36] Aldehydes (Acute toxicity) Strict (US
[C: 72;Md: 135;P: 29] Enter GF(TA)
[C: 68;Md: 125;P: 29] Subcategorized: Acute aquatic tox
[C: 29;Md: 64;P: 29] Subcategorized: Protein binding
[C: 28;Md: 63;P: 29] Subcategorized: Aquatic tox
[C: 72;Md: 162;P: 29] Enter GF(ExternalQSAR)

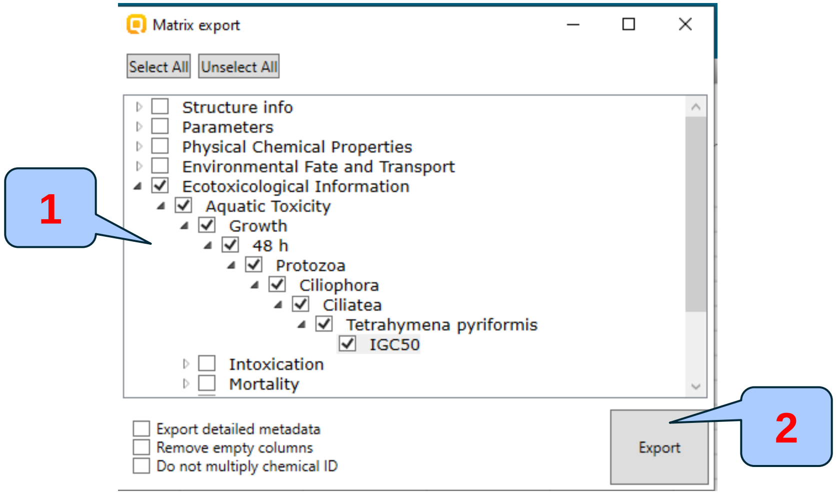
☑ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

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).

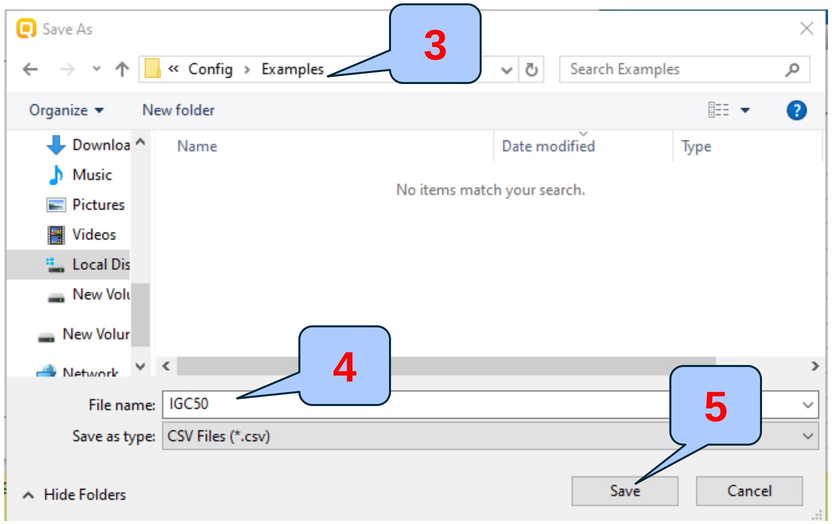
Export QSAR results



Matrix export

- □ Structure info
- □ Parameters
- □ Physical Chemical Properties
- □ Environmental Fate and Transport
- ☑ Ecotoxicological Information
  - ☑ Aquatic Toxicity
    - ☑ Growth
      - ☑ 48 h
        - ☑ Protozoa
          - ☑ Ciliophora
            - ☑ Ciliatea
              - ☑ Tetrahymena pyriformis
                - ☑ IGC50
    - □ Intoxication
    - □ Mortality

□ Export detailed metadata
□ Remove empty columns
□ Do not multiply chemical ID



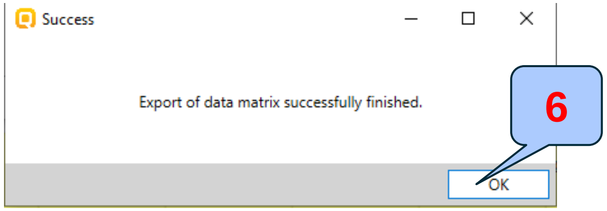
Save As

« Config › Examples

No items match your search.

File name: IGC50

Save as type: CSV Files (*.csv)



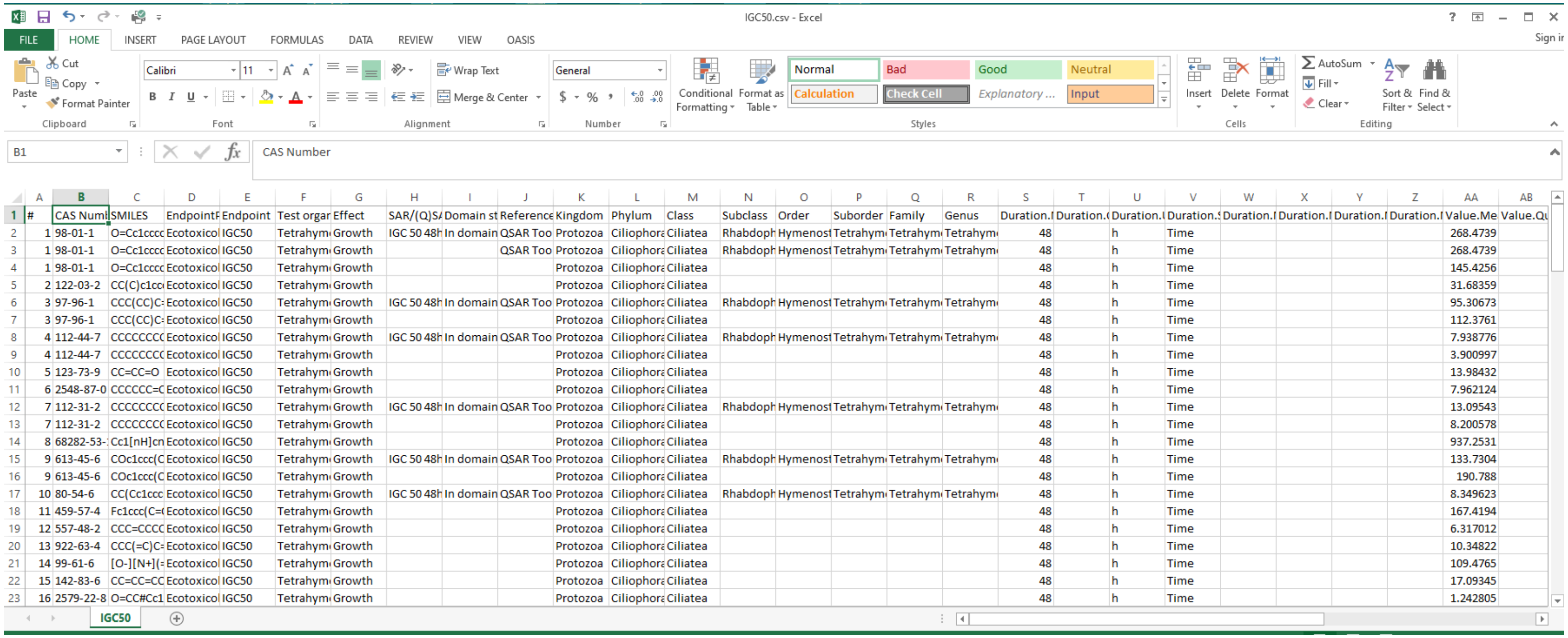
Success

Export of data matrix successfully finished.

1. The nodes from the tree associated with QSAR predictions which will be exported are selected with check marks; 2. Click **Export**; 3. Browse to save the file 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.

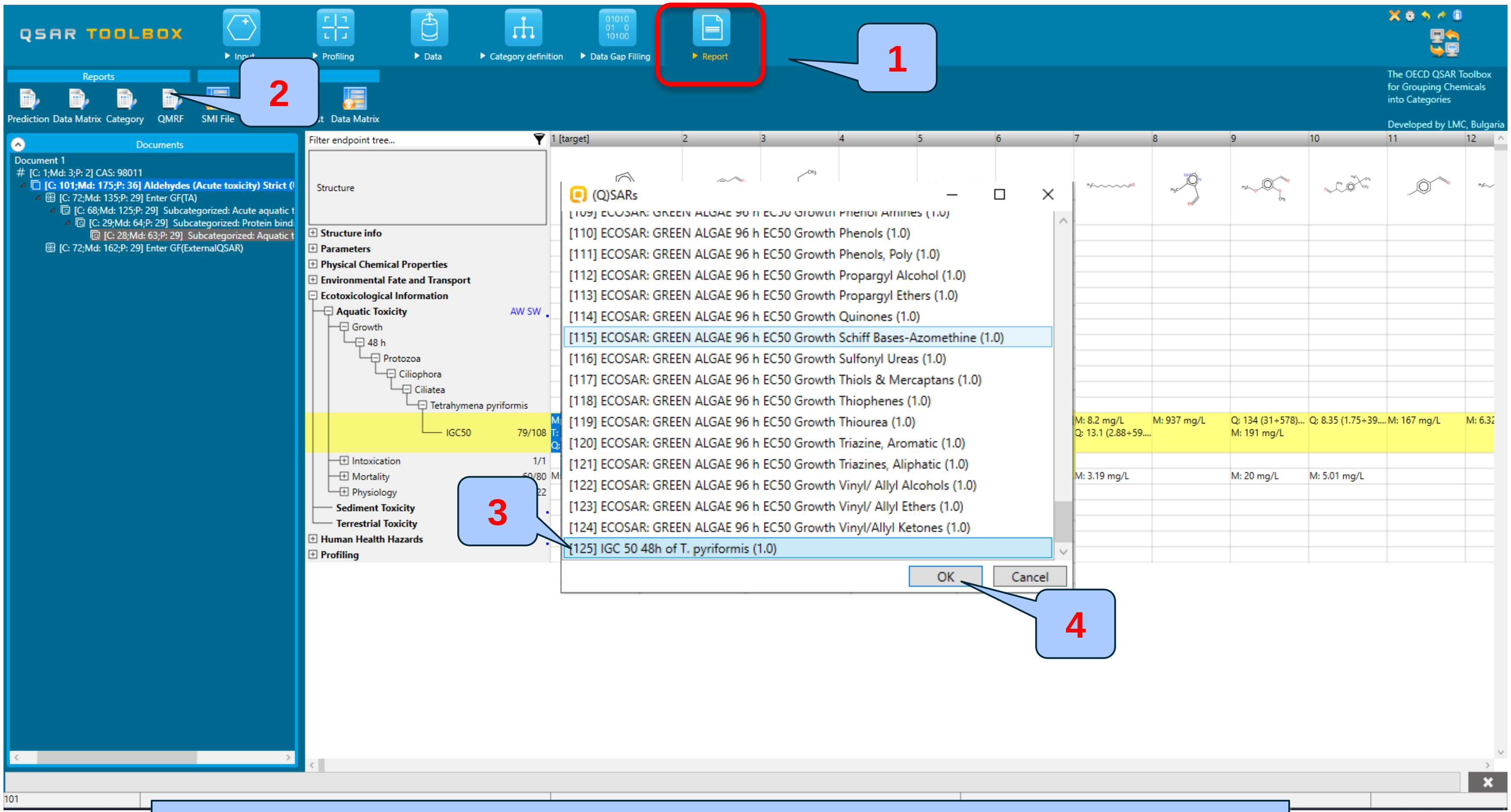


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	#	CAS Numl	SMILES	EndpointF	Endpoint	Test orgar	Effect	SAR/(Q)SA	Domain st	Reference	Kingdom	Phylum	Class	Subclass	Order	Suborder	Family	Genus	Duration.I	Duration.(	Duration.U	Duration.S	Duration.I	Duration.I	Duration.I	Duration.I	Value.Me	Value.Qu
2	1	98-01-1	O=Cc1cccc	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					268.4739	
3	1	98-01-1	O=Cc1cccc	Ecotoxicol	IGC50	Tetrahym	Growth			QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					268.4739	
4	1	98-01-1	O=Cc1cccc	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					145.4256	
5	2	122-03-2	CC(C)c1cc	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					31.68359	
6	3	97-96-1	CCC(CC)C=	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					95.30673	
7	3	97-96-1	CCC(CC)C=	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					112.3761	
8	4	112-44-7	CCCCCCC(	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					7.938776	
9	4	112-44-7	CCCCCCC(	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					3.900997	
10	5	123-73-9	CC=CC=O	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					13.98432	
11	6	2548-87-0	CCCCCC=(	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					7.962124	
12	7	112-31-2	CCCCCCC(	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					13.09543	
13	7	112-31-2	CCCCCCC(	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					8.200578	
14	8	68282-53-:	Cc1[nH]cn	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					937.2531	
15	9	613-45-6	COc1ccc(C	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					133.7304	
16	9	613-45-6	COc1ccc(C	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					190.788	
17	10	80-54-6	CC(Cc1ccc	Ecotoxicol	IGC50	Tetrahym	Growth	IGC 50 48h	In domain	QSAR Too	Protozoa	Ciliophora	Ciliatea	Rhabdoph	Hymenost	Tetrahym	Tetrahym	Tetrahym	48		h	Time					8.349623	
18	11	459-57-4	Fc1ccc(C=(	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					167.4194	
19	12	557-48-2	CCC=CCCC	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					6.317012	
20	13	922-63-4	CCC(=C)C=	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					10.34822	
21	14	99-61-6	[O-][N+](=	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					109.4765	
22	15	142-83-6	CC=CC=CC	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					17.09345	
23	16	2579-22-8	O=CC#Cc1	Ecotoxicol	IGC50	Tetrahym	Growth				Protozoa	Ciliophora	Ciliatea						48		h	Time					1.242805	

Outlook

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

Report



Document 1
[C: 1;Md: 3;P: 2] CAS: 98011
[C: 101;Md: 175;P: 36] Aldehydes (Acute toxicity) Strict (
[C: 72;Md: 135;P: 29] Enter GF(TA)
[C: 68;Md: 125;P: 29] Subcategorized: Acute aquatic t
[C: 29;Md: 64;P: 29] Subcategorized: Protein bindi
[C: 28;Md: 63;P: 29] Subcategorized: Aquatic t
[C: 72;Md: 162;P: 29] Enter GF(ExternalQSAR)

Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Aquatic Toxicity AW SW
Growth
48 h
Protozoa
Ciliophora
Ciliatea
Tetrahymena pyriformis
IGC50 79/108
Intoxication 1/1
Mortality 60/80
Physiology 22
Sediment Toxicity
Terrestrial Toxicity
Human Health Hazards
Profiling

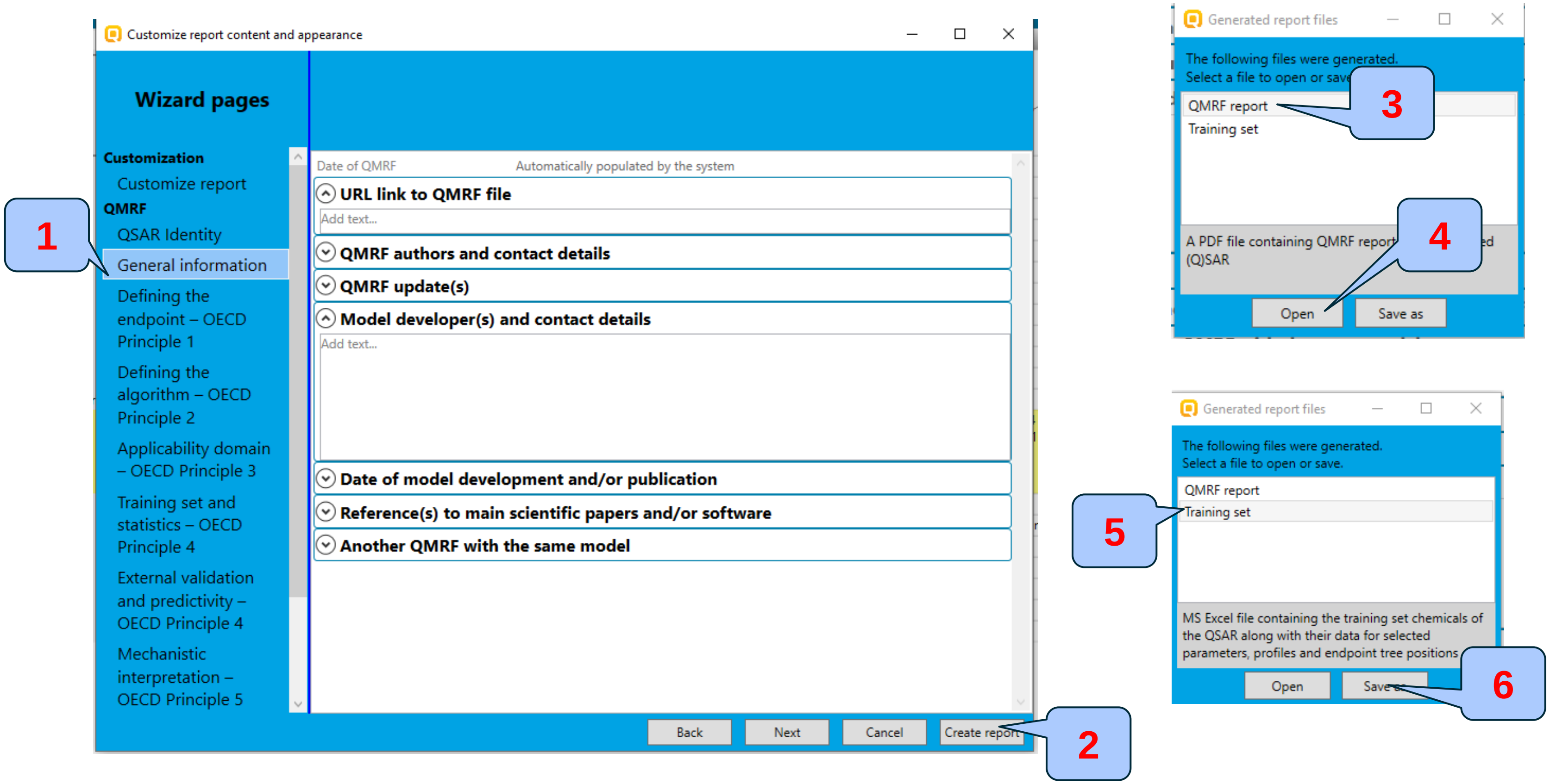
M: 8.2 mg/L
Q: 13.1 (2.88÷59....
M: 937 mg/L
Q: 134 (31÷578)...
M: 191 mg/L
Q: 8.35 (1.75÷39....
M: 167 mg/L
M: 6.32
M: 3.19 mg/L
M: 20 mg/L
M: 5.01 mg/L

(Q)SARs

[109] ECOSAR: GREEN ALGAE 96 h EC50 Growth Phenol Amines (1.0)
[110] ECOSAR: GREEN ALGAE 96 h EC50 Growth Phenols (1.0)
[111] ECOSAR: GREEN ALGAE 96 h EC50 Growth Phenols, Poly (1.0)
[112] ECOSAR: GREEN ALGAE 96 h EC50 Growth Propargyl Alcohol (1.0)
[113] ECOSAR: GREEN ALGAE 96 h EC50 Growth Propargyl Ethers (1.0)
[114] ECOSAR: GREEN ALGAE 96 h EC50 Growth Quinones (1.0)
[115] ECOSAR: GREEN ALGAE 96 h EC50 Growth Schiff Bases-Azomethine (1.0)
[116] ECOSAR: GREEN ALGAE 96 h EC50 Growth Sulfonyl Ureas (1.0)
[117] ECOSAR: GREEN ALGAE 96 h EC50 Growth Thiols & Mercaptans (1.0)
[118] ECOSAR: GREEN ALGAE 96 h EC50 Growth Thiophenes (1.0)
[119] ECOSAR: GREEN ALGAE 96 h EC50 Growth Thiourea (1.0)
[120] ECOSAR: GREEN ALGAE 96 h EC50 Growth Triazine, Aromatic (1.0)
[121] ECOSAR: GREEN ALGAE 96 h EC50 Growth Triazines, Aliphatic (1.0)
[122] ECOSAR: GREEN ALGAE 96 h EC50 Growth Vinyl/ Allyl Alcohols (1.0)
[123] ECOSAR: GREEN ALGAE 96 h EC50 Growth Vinyl/ Allyl Ethers (1.0)
[124] ECOSAR: GREEN ALGAE 96 h EC50 Growth Vinyl/Allyl Ketones (1.0)
[125] IGC 50 48h of T. pyriformis (1.0)

1. Go to ***Report*** module; 2. Click **QMRF**;
3. Select the name of the user-defined QSAR model; 4. Click **OK**;

Report



Customize report content and appearance

Wizard pages

Customization

Customize report

QMRF

QSAR Identity

General information

Defining the endpoint – OECD Principle 1

Defining the algorithm – OECD Principle 2

Applicability domain – OECD Principle 3

Training set and statistics – OECD Principle 4

External validation and predictivity – OECD Principle 4

Mechanistic interpretation – OECD Principle 5

Date of QMRF — Automatically populated by the system

URL link to QMRF file

Add text...

QMRF authors and contact details

QMRF update(s)

Model developer(s) and contact details

Add text...

Date of model development and/or publication

Reference(s) to main scientific papers and/or software

Another QMRF with the same model

Generated report files

The following files were generated.
Select a file to open or save.

QMRF report
Training set

A PDF file containing QMRF report ... (Q)SAR

Generated report files

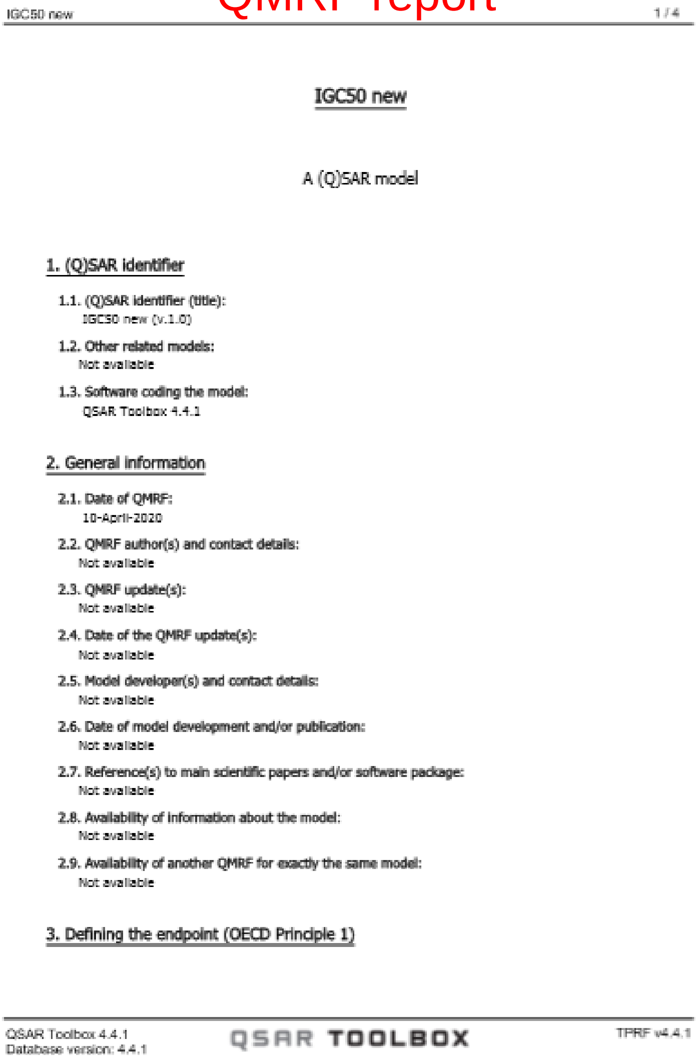
The following files were generated.
Select a file to open or save.

QMRF report
Training set

MS Excel file containing the training set chemicals of the QSAR along with their data for selected parameters, profiles and endpoint tree positions

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

Report



QMRF report

IGC50 new 1 / 4

IGC50 new

A (Q)SAR model

1. (Q)SAR identifier

1.1. (Q)SAR identifier (title):
IGC50 new (v.1.0)

1.2. Other related models:
Not available

1.3. Software coding the model:
QSAR Toolbox 4.4.1

2. General information

2.1. Date of QMRF:
10-April-2020

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

2.3. QMRF update(s):
Not available

2.4. Date of the QMRF update(s):
Not available

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

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

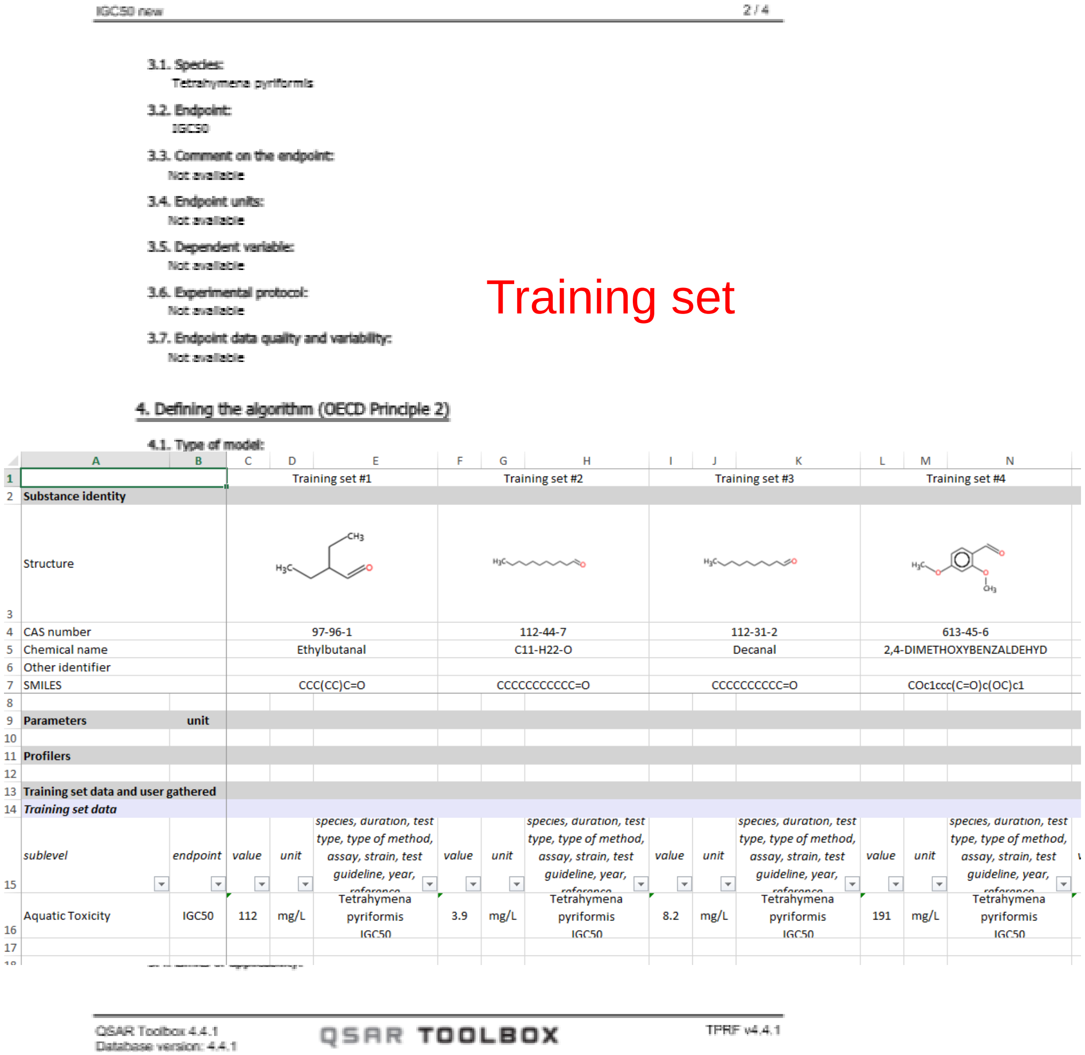
2.7. Reference(s) to main scientific papers and/or software package:
Not available

2.8. Availability of information about the model:
Not available

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

3. Defining the endpoint (OECD Principle 1)

QSAR Toolbox 4.4.1
Database version: 4.4.1
QSAR TOOLBOX
TPRF v4.4.1



IGC50 new 2 / 4

3.1. Species:
Tetrahymena pyriformis

3.2. Endpoint:
IGC50

3.3. Comment on the endpoint:
Not available

3.4. Endpoint units:
Not available

3.5. Dependent variable:
Not available

3.6. Experimental protocol:
Not available

3.7. Endpoint data quality and variability:
Not available

Training set

4. Defining the algorithm (OECD Principle 2)

4.1. Type of model:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			Training set #1			Training set #2			Training set #3			Training set #4		
2	**Substance identity**													
3	Structure													
4	CAS number		97-96-1			112-44-7			112-31-2			613-45-6		
5	Chemical name		Ethylbutanal			C11-H22-O			Decanal			2,4-DIMETHOXYBENZALDEHYD		
6	Other identifier													
7	SMILES		CCC(CC)C=O			CCCCCCCCCCC=O			CCCCCCCCCC=O			COc1ccc(C=O)c(OC)c1		
8														
9	**Parameters**	**unit**												
10														
11	**Profilers**													
12														
13	**Training set data and user gathered**													
14	***Training set data***													
15	*sublevel*	*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*
16	Aquatic Toxicity	IGC50	112	mg/L	Tetrahymena pyriformis IGC50	3.9	mg/L	Tetrahymena pyriformis IGC50	8.2	mg/L	Tetrahymena pyriformis IGC50	191	mg/L	Tetrahymena pyriformis IGC50
17														

QSAR Toolbox 4.4.1
Database version: 4.4.1
QSAR TOOLBOX
TPRF v4.4.1

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.