

## OECD (Q)SAR Toolbox v.4.4.1

Predicting acute aquatic toxicity to fish of  
*Dodecanenitrile* (CAS 2437-25-4) by taking  
into account tautomerism

# Outlook

- **Background**
- Keywords
- Objectives
- The exercise
- Workflow
- Save prediction

# Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow for filling a data gap for acute aquatic toxicity to fish taking into account tautomerism of the target chemical.

# Outlook

- Background
- **Keywords**
- Objectives
- The exercise
- Workflow
- Save prediction

# Keywords

**TARGET CHEMICAL** - chemical of interest

**MODULE** – a Toolbox module is a section dedicated to specific actions and options

**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)

**TAUTOMER** – Tautomers are structural isomers which are in dynamic equilibrium due to the migration of a proton, accompanied by a switch of a single bond and adjacent double bond.

**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

# Outlook

- Background
- Keywords
- **Objectives**
- The exercise
- Workflow
- Save prediction

# Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
  - Providing tautomeric set for the target chemical
  - Identify analogues for the active tautomeric form
  - Retrieve experimental results available for those analogues
  - Perform trend analysis for the active tautomeric form
  - Assigning of the prediction for the active tautomer to the target chemical
  - Saving the prediction result

# Outlook

- Background
- Keywords
- Objectives
- **The exercise**
- Workflow
- Save prediction



## The Exercise

- In this exercise we will predict *LC50* for fish: *P.promelas* for target chemical *Dodecanenitrile* (CAS 2437-25-4)
- Set of simulated tautomers for the target chemical will be provided
- Analyze the profilers of the tautomeric forms within tautomeric set
- Filling data gaps for active tautomer by trend analysis
- Assign prediction for the tautomeric forms to the target chemical

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
- Save prediction

# Workflow

- **As you know the Toolbox has 6 modules which are typically used in sequence:**
  - Chemical Input
  - Profiling
  - Data
  - Category Definition
  - Data Gap Filling
  - Report

# Chemical Input

## Ways of Entering a Chemicals

### **User Alternatives for input of Chemical:**

#### **A.** Single target chemical

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases

#### **B.** Group of chemicals

- User List/Inventory
- Specialized Databases

# Chemical Input

## Input Screen

- Open the Toolbox.
- The six modules in the workflow are seen listed next to “(Q)SAR TOOLBOX” title.
- **Click** on “Input” (see next screen shot)

# Chemical Input

## Input target chemical by CAS#

1. Click **CAS#**;

2. Enter **2437-25-4**;

3. One structure is found in the Toolbox databases, which is checked by default;

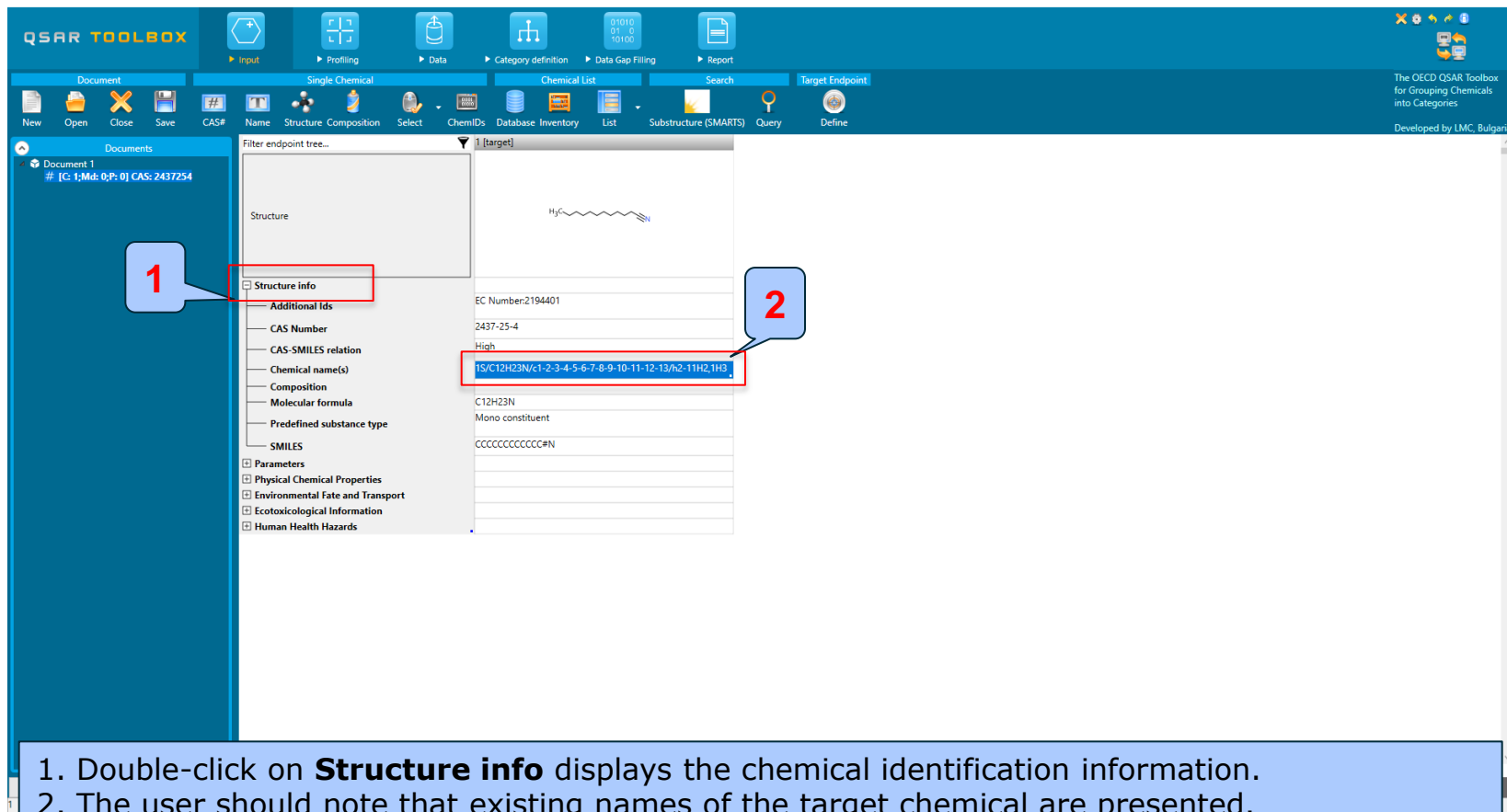
4. Click **OK**.

1	CAS	2437-25-4
	SMILES	CCCCCCCCCCCC#N
	CS Relation	High
	Substance	Mono constituent
	Composition	
	Name	1S/C12H23N/c1-2-3-4-5-6-7-8-9-10-
	Sources	NICNAS Canada DSI

Chemical structure: CCCCCCCCCCCC#N

# Chemical Input

## Target chemical identity



The screenshot shows the QSAR Toolbox interface with the 'Chemical Input' module selected. The 'Structure info' section is expanded, and the 'Additional Ids' list is highlighted. The list contains the following information:

- EC Number: 2194401
- CAS Number: 2437-25-4
- CAS-SMILES relation: High
- Chemical name(s): 1S/C12H23N/c1-2-3-4-5-6-7-8-9-10-11-12-13/a2-11H2,1H3
- Composition: C12H23N
- Molecular formula: Mono constituent
- Predefined substance type: CCCCCCCCCC#N
- SMILES: CCCCCCCCCC#N

1. Double-click on **Structure info** displays the chemical identification information.

2. The user should note that existing names of the target chemical are presented.

The workflow on the first module is now complete, and the user can proceed to the next module.

# Outlook

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



# Define target endpoint Overview

1. Click on **Define Target Endpoint**

2. Select Endpoint "Aquatic Toxicity" and

3. Click **Next**.

# Define target endpoint

## Overview

Select endpoint

Ecotoxicological Information  
Aquatic Toxicity

Effect: Mortality

Duration: 96 h

Test organisms (species): Pimephales promelas

Endpoint: LC50

Selection of additional metadata fields:

Up Down Clear Move

Back Finish

1: Endpoint (LC50)

2: Test organisms (species) (Pimephales promelas)

3: Effect (Mortality)

4: Duration (96 h)

5: Finish button

1. First select "LC 50" from drop- down menu of **Endpoint** field, then consecutively select the following metadata:
2. Field **Test organism (species)** – Pimephales promelas;
3. Field **Duration** – 96 h.
4. Field **Effect** – Mortality
5. Click **Finish**.

# Define target endpoint Overview

The screenshot displays the QSAR Toolbox interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. The 'Define' tab is active. On the left, a 'Documents' panel shows 'Document 1' with CAS# 2437254. The main area features a 'Filter endpoint tree...' window. This window shows a tree structure of endpoints. The 'LC50' endpoint under 'Pimephales promelas' is highlighted in yellow. A callout bubble with the number '1' points to this highlighted endpoint. The main window also displays the chemical structure of H<sub>3</sub>C(CH<sub>2</sub>)<sub>10</sub>CN and a table with one row for the target endpoint.

1. The defined target endpoint is highlighted in yellow

# Outlook

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

## Profiling Overview

- “Profiling” 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 information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

## Profiling

### Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, US-EPA New Chemical categories and clicking on “View” (see next screen shot).

# Profiling

## Side-Bar to Profiling

The screenshot shows the QSAR TOOLBOX Profiling interface. On the left, the 'Profiling methods' list is visible. The 'US-EPA New Chemical Categories' method is selected and highlighted with a red box, with a callout '1' pointing to it. A context menu is open over this selection, with 'View scheme' highlighted by a red box and a callout '2'. The 'Filter endpoint tree...' on the right shows a tree structure. The 'Pimephales promelas' node is highlighted in yellow, with a callout '3' pointing to it. The 'LC50' node is also highlighted in yellow.

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)  
 Benzotriazole-hindered phenols  
 Benzotriazoles (Acute toxicity)  
 Benzotriazoles (Chronic toxicity)  
 beta-Naphthylamines, Sulfonated  
 Boron Compounds  
 Cationic (quaternary ammonium) surfactants  
 Cobalt  
 Dianilines  
 Diazoniums (Acute toxicity)  
 Diazoniums (Chronic toxicity)  
 Dichlorobenzidine-based Pigments  
 Diisocyanates  
 Dithiocarbamates (Acute toxicity)  
 Dithiocarbamates (Chronic toxicity)  
 Epoxides  
Esters (Acute toxicity)  
 Esters (Chronic toxicity)  
 Ethylene Glycol Ethers  
 Hindered Amines  
 Hydrazines and Related Compounds  
 Imides (Acute toxicity)

1. Select US-EPA New Chemical categories
2. **Click** View
3. **Select** "Esters(Acute toxicity)"

# Profiling

## Side-Bar to Profiling

1

Definition Properties Training Set Literature MetaInfo Table Custom Captions Scheme

Category tree

[16] Esters (Acute toxicity)

Structural boundary

2

1

AND

3

ADD  
DEL  
AND  
OR  
NOT

[0] Structure Query Metabolism

Contents

Queries

Add Query Remove

Add Mask

Complex search options

☒ Exact connectivity

☐ Ignore stereo information

☐ Exact match

Queries execution mode All

Mapping

☒ Unique mappings

Max maps 1000

SMARTS

[#6]([#6])(=[#8])[#8R0][#6] Edit

View mode: Facade Navigation mode: Cascade

3

Structural fragment

Left click on any marked atom to explore

1. Select **Definition** tab;
2. The first boundary is a **structural boundary** containing a **structural fragment** (3).



# Profiling

## Side-Bar to Profiling

The screenshot shows the 'Category tree' on the left with a callout '1' pointing to a puzzle piece icon. The main area is labeled 'Parametric boundary'. Below it, the 'Query details' section shows a 'Parameter Query' for 'Metabolism'. A list of parameter names is shown, with 'log Kow' selected and highlighted with a red box. A callout '2' points to the 'log Kow' entry. The 'Expression' field for 'log Kow' is set to '5'.

The screenshot shows the 'Category tree' on the left with a callout '3' pointing to a puzzle piece icon. The main area is labeled 'Parametric fragment'. Below it, the 'Query details' section shows a 'Parameter Query' for 'Metabolism'. A list of parameter names is shown, with 'Molecular weight' selected and highlighted with a red box. A callout '4' points to the 'Molecular weight' entry. The 'Expression' field for 'Molecular weight' is set to '1E+03'. On the right, there are settings for 'Origin', 'scale: Mass', 'unit: Da', and 'Destination'.

The second boundary (1) is a parameter query containing expression for logKow (2); The third boundary (3) is a parameter query containing expression for Molecular weight (4).

# Profiling

## Side-Bar to Profiling

US-EPA New Chemical Categories (Predefined) - Profiling Scheme Browser

Save Scheme Export Scheme Save Tests View Tests Run All Tests

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)
  - Benzotriazole-hindered phenols
  - Benzotriazoles (Acute toxicity)
  - Benzotriazoles (Chronic toxicity)
  - beta-Naphthylamines, Sulfonated
  - Boron Compounds
  - Cationic (quaternary ammonium) surfactants
  - Cobalt
  - Dianilines
  - Diazoniums (Acute toxicity)
  - Diazoniums (Chronic toxicity)
  - Dichlorobenzidine-based Pigments
  - Diisocyanates
  - Dithiocarbamates (Acute toxicity)
  - Dithiocarbamates (Chronic toxicity)
  - Epoxides
  - Esters (Acute toxicity)**
  - Esters (Chronic toxicity)
  - Ethylene Glycol Ethers
  - Hindered Amines
  - Hydrazines and Related Compounds
  - Imides (Acute toxicity)

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

**Category: Esters**

This category includes all esters, polyesters, vinyl esters, allylic esters, propargylic esters, aliphatic esters, aromatic esters, carboxylic acid esters, and substituted esters. These compounds need to be absorbed to be toxic, therefore, compounds with MWs > 1000 will be excluded from this category. Acute toxicity for esters which are liquids at room temperature is known to be limited by the octanol/water partition coefficient ( $K_{ow}$ ). Above a  $\log K_{ow}$  value of  $\geq 5.0$ , esters show no effects at saturation during 96-h exposures (Veith et al 1984). Esters 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 acute toxicity will be observed at saturation. For solids, the no-effects-at-saturation point has to be determined on a case-by-case basis. The  $K_{ow}$  limit for chronic toxicity is set at a  $\log K_{ow} = 8$  for liquid esters. For solid esters, chronic toxicity testing will determine this  $K_{ow}$  limit.

**Hazard Concerns.** The toxicity for simple esters has been determined through SAR Analysis (Clements 1988). Esters are known to be more toxic than neutral organic chemicals, and this excess toxicity decreases with increasing  $K_{ow}$ . The toxicity for vinyl esters, allylic esters, and propargylic esters is expected to be greater than for simple esters. Again, the additional excess toxicity of these vinyl esters, allylic esters, and propargylic esters is expected to decrease with increasing  $K_{ow}$ .

Members of this category exhibit toxicity ranging from low toxicity (i.e., > 100 mg/L) to high toxicity (i.e., < 1 mg/L) depending on their  $K_{ow}$ , MW, and melting point.

**Boundaries.** There are no known lower boundaries. The upper boundaries will be based on  $K_{ow}$  and MW. Acute toxicity is expected when  $\log K_{ow} < 5.0$ ; no effects at saturation during 96-h exposures when  $\log K_{ow} > 5.0$ . The upper boundary for chronic toxicity is 8.0. MW will be < 1000. The environmental base set of tests will be requested for aquatic releases and the terrestrial base set of tests will be recommended for terrestrial exposures. When the  $\log K_{ow}$  is > 5.0, chronic toxicity testing with fish and daphnids will be recommended.

Fate: Esters are subject to both abiotic and biotic hydrolysis, i.e., ester hydrolysis, and aerobic biodegradation. Aerobic biodegradation is expected to be the dominant route of transformation in the environment.

General Testing Strategy.

I. Release to Aquatic Ecosystems:

1. Click on Literature tab to see mechanistic justification of the category

# Profiling

## Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues
- The following profiling schemes are relevant to the **Acute aquatic toxicity**:
  - Aquatic toxicity classification by ECOSAR
  - Acute aquatic toxicity MOA by OASIS
  - Acute aquatic toxicity classification by Verhaar (Modified)
  - US-EPA New Chemical Categories
- More details about profiling schemes used for categorization and collection of analogues is provided in stage "Category formation" **on slide 51**

# Profiling

## Profiling the target chemical

- Select the “Profiling methods” related to the target endpoint by clicking the box next to the profilers name;
- This selects (a **green** check mark appears) or deselects (**green** check mark disappears) profilers.
- For this example, go through the general and endpoint specific profiling mechanisms and highlight those that apply to acute aquatic toxicity(see next screen shot).

# Profiling

## Profiling the target chemical

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

2

Apply View New Delete

Documents

Profiling methods

Options 4 Selected

Select All Unselect All Invert

1

Suitable

- Acute aquatic toxicity classification by Verhaar (Modified)
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- 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...

1 [target]

Structure

Ecotoxicological Information

Aquatic Toxicity

Mortality

96 h

Animalia (animals)

Chordata (chordates)

Actinopterygii (ray-finned...)

Pimephales promelas

LC50

Sediment Toxicity

Terrestrial Toxicity

Human Health Hazards

1. Check all **green highlighted profilers**;
2. Click **Apply**.

## Profiling

### Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected
- The results of profiling automatically appear as a dropdown box under the target chemical (see next screen shot)
- Please note the endpoint specific profilers and structure based profilers such as US-EPA and ECOSAR
- No structural and endpoint specific alerts have been found for the test compound.

(see next screenshot)

# Profiling

## Profiling the target chemical

The target chemical was not categorized by both OECD and US-EPA profilers. It has no alert found by both protein binding profilers. It is also categorized as "neutral organics and basesurface narcotics" by ECOSAR and MOA of action profilers, which are classes not associated with excess toxicity.

1

Structure

Sediment Toxicity

Terrestrial Toxicity

Human Health Hazards

Profile

Predefined

US-EPA New Chemical Categories

Endpoint Specific

Acute aquatic toxicity classification by...

Acute aquatic toxicity MOA by OASIS

Aquatic toxicity classification by ECOS...

Not categorized

Class 5 (Not possible to classify according to these rules)

Basesurface narcotics

Neutral Organics

1. Double click on **Profile** node to review the profiling results.

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Define Target Endpoint
  - Profiling
  - **Data**



# Data Overview

- “Data” refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).

## Data

### Case study

- In this example, we limit our data gathering to a single toxicity endpoint (acute aquatic toxicity).
- In this example, we collect data from the databases containing experimental results for acute aquatic toxicity (Aquatic toxicity OASIS; ECOTOX and ECHA REACH).
- Click on “Data” in the Toolbox workflow.
- Expand the “Ecotoxicological information” section
- Click on the box to select the relevant databases.
- Click on “Gather data” (see next screen shot).

# Data

## Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar has 'Data' highlighted with a red box and a blue callout '1'. The 'Data' sub-menu is open, showing 'Gather' highlighted with a red box and a blue callout '3'. The 'Databases' list on the left has several databases highlighted in green, with a blue callout '2' pointing to them. The main window shows a chemical structure and a filter endpoint tree.

1. Click **Data**;

2. Select **green highlighted databases** related to the target endpoint;

3. Click **Gather**.

## Data

### Gather data

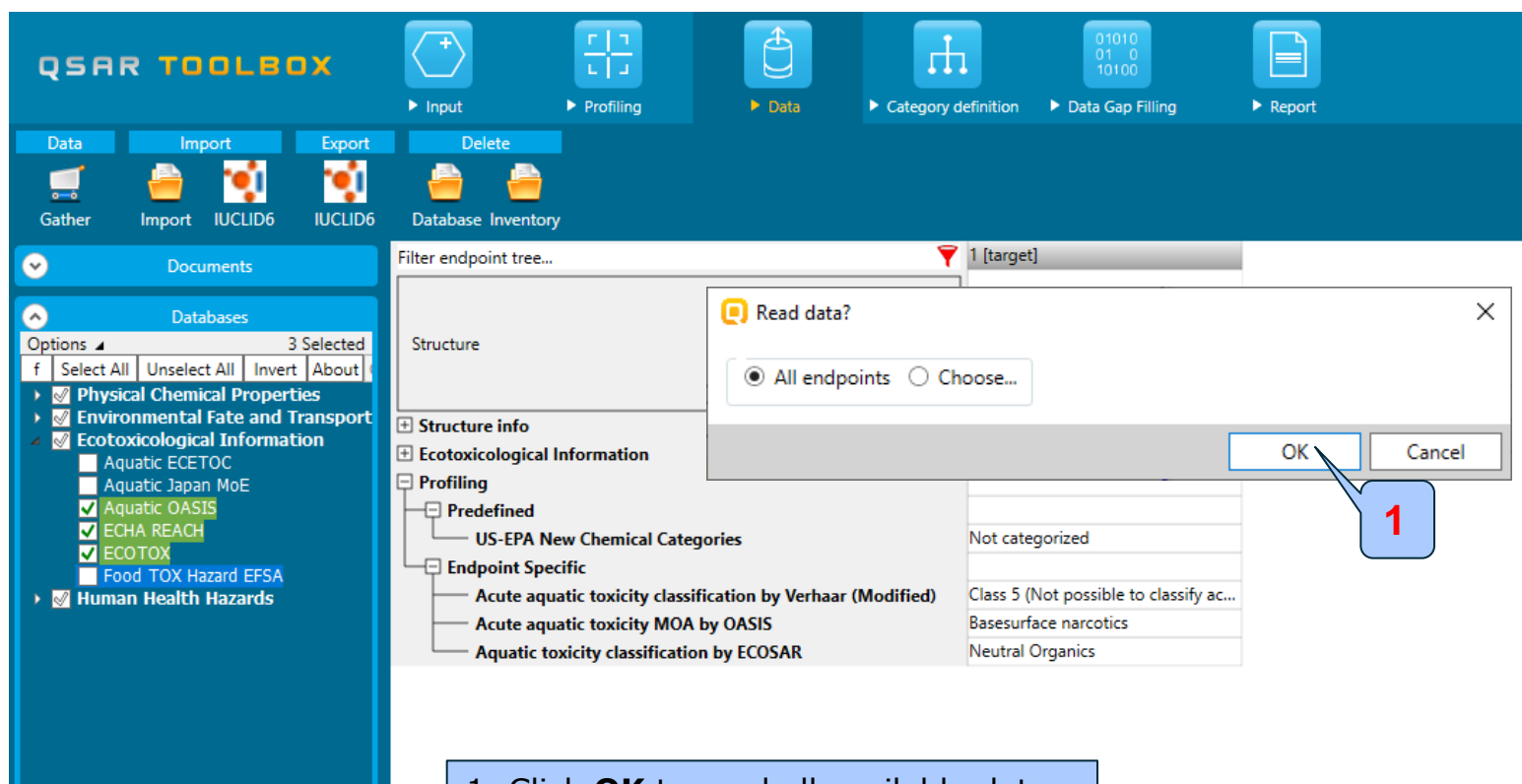
- Toxicity information on the target chemical is electronically collected from the selected dataset(s)
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are Aquatic toxicity OASIS; ECOTOX and ECHA REACH.
- In this example, there is LC50 experimental data for *P. promelas* (96h) for the target chemical (see next screen shots)
- The experimental data for the investigated endpoint falls within the toxic range (less than 1mg/l<sup>1</sup>)

<sup>1</sup> **Globally Harmonized System of Classification and Labeling of Chemicals (GHS):**  
[https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs\\_rev04/English/ST-SG-AC10-30-Rev4e.pdf](https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev04/English/ST-SG-AC10-30-Rev4e.pdf)

# Data

## Gather data

Toxicity information on the target chemical is electronically collected from the selected datasets. A "Read data?" window appears. Now the user could choose to collect "all" or "endpoint specific" data for the target chemical.



1. Click **OK** to read all available data.

# Data Gather data

Filter endpoint tree... 1 [target]

Structure

Animalia (animals)

Chordata (chordates)

Actinopterygii (ray-finned...)

Oryzias latipes 1/1 M: 0.84 mg/L

Pimephales promelas

LC50 1/2 M: 0.425 mg/L

Undefined Kingdom 1/1 M: 0.43 mg/L

Sediment Toxicity

Terrestrial Toxicity

Human Health Hazards

Profile

Predefined

US-EPA New Chemical Categories

Endpoint Specific

Acute aquatic toxicity classification by...

Acute aquatic toxicity MOA by OASIS

Aquatic toxicity classification by ECOS...

Data points

Datapoints	#	Value	Original value	Assigned SMILES
Ecotoxicological Information; Aquatic Toxicity	1	M: 0.425 mg/L (Mass concentration)	2.34E-06 mol/L (Molar concentration)	False
Ecotoxicological Information; Aquatic Toxicity	2	M: 0.43 (0.4-0.47) mg/L (Mass concentration)	0.43 (0.4-0.47) mg/L (Mass concentration)	True

OK

1. Double-click on **the cell** displays metadata information for the observed data (2)
2. Click on the **X** to close the window.

## Recap

- The first module, which introduces the target chemical, ensure correctness of the structure;
- In the second module the target endpoint was defined.
- The third module shows that there is no structural or endpoint specific alerts for target chemical
- In the fourth module, you have found that the target chemical has experimental data for the investigated endpoint
- The study continues with accounting for tautomersim of target chemical trying to explain toxic experimental data of the target chemical (see next slides).

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Define Target Endpoint
  - Profiling
  - Data
  - **Handling of tautomerism of target chemical**



# Handling of tautomerism of target chemical

The screenshot shows the QSAR Toolbox software interface. The 'Input' menu is open, and the 'Multiplication' option is selected. The 'Tautomerism' option is also visible. The 'Information' dialog box is open, showing a parent list with 3 child lists. The table below shows the generated tautomers for the target chemical.

tautomer #1 (target)	tautomer #2	tautomer #3
<chem>CCCCCCCCCCCCCCCC#N</chem>	<chem>CCCCCCCCCCCCCCCC=NC</chem>	<chem>CCCCCCCCCCCCCCCC#N</chem>

1. Go to **Input**
2. Right click over the node with **CAS#** and select **Multiplication** (3) and then **Tautomerism** (4)
5. Three tautomeric forms are generated for the target chemical;
6. Click **OK**.

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - **Profiling set of tautomers**

# Handling of tautomerism of target chemical

## Profiling set of tautomers

- This module identifies profilers of target chemical and its tautomeric forms
- Endpoint specific and structurally based profiles related to acute aquatic toxicity are applied on the set of tautomers
- Profiling results of tautomers are illustrated in Single Component mode
- Click on "Profiling" to go to the required module (see next screen shots)

# Handling of tautomerism of target chemical

## Profiling set of tautomers

- The following primary profilers relevant to the **aquatic toxicity** are used in this example (see next screenshot):
  - US-EPA New chemical category
  - Aquatic toxicity classification by ECOSAR
  - Acute aquatic toxicity MOA by OASIS
  - Acute aquatic toxicity classification by Verhaar
- Select the Profilers highlighted in green which are related to the target endpoint by ticking the boxes next to the profilers name.

# Handling of tautomerism of target chemical

## Profiling set of tautomeric forms

**1** Check the **profilers related to acute aquatic toxicity** (highlighted in green)

**2** Click **Apply**

**Target**

**Tautomeric forms**

**3**

**4**

Parent chemical...	tautomer #2 (target)	tautomer #1	tautomer #3
<chem>CCCCCCCCCCCCCCCC</chem>	<chem>CCCCCCCCCCCCCCCC</chem>	<chem>CCCCCCCCCCCCCCCC</chem>	<chem>CCCCCCCCCCCCCCCC</chem>
Not categorized	Not categorized	Not categorized	Not categorized
Class 5 (Not possible to...)	Class 5 (Not possible to...)	Class 5 (Not possible to...)	Class 5 (Not possible to...)
Basesurface narcotics	Basesurface narcotics	Basesurface narcotics	Basesurface narcotics
Neutral Organics	Neutral Organics	Not Related to an Existing...	Aliphatic Amines

1. Check the **profilers related to acute aquatic toxicity** (highlighted in green)
2. Click **Apply**

The profiling results **(3)** indicates no alerts found for the target chemical. Also classes associated with baseline toxicity (not excess toxicity) have been found for the target. However, there is an endpoint specific alert (Aliphatic amines) **(4)** for one of the simulated tautomeric form. This tautomer has been used in further trend analysis

# Handling of tautomerism of target chemical

## Recap

- The profiling results indicates no endpoint specific or active structural alerts for target chemical
- One of the simulated tautomeric form has positive endpoint specific alert identified by ECOSAR
- The active tautomer is used for further trend analysis
- The next two parts of the exercise will focus the active tautomer and identify the category of similar analogues (see next screenshots).

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - Profiling set of tautomers
    - **Focus active tautomer**

# Handling of tautomerism of target chemical

## Focus of active tautomer

This tautomeric form is selected for further trend analysis

1

2

"Focus" functionality allows the selected tautomer to be used as post target representative of the target chemical

1. Right click over the **active tautomeric form**;  
2. **Focus** the chemical.

Filter endpoint	target chemical [target]	tautomer #1	tautomer #2	tautomer #3 (target) [target]
Structure	<chem>CCCCCCCCCCCCCCCC#N</chem>	<chem>CCCCCCCCCCCCCCCC#N</chem>	<chem>CCCCCCCCCCCCCCCC#N</chem>	<chem>CCCCCCCCCCCCCCCC#N</chem>
72 h	1/1 M: >0÷ <0.75 mg/L			M: >0÷ <0.75 mg/L
96 h				
Animalia (animals)				
Chordata (chordates)				
Actinopterygii (ray-finned...)				
Oryzias latipes	1/1 M: 0.84 mg/L			M: 0.84 mg/L
Pimephales promelas				
Sediment T				M: 0.425 mg/L
Terrestrial				M: 2.34 mg/L
Profiling				
Predefined				
US-EPA New Chemical Categories	Not categorized	Not categorized	Not categorized	Not categorized
Endpoint Specific				
Acute aquatic toxicity classification by...	Class 5 (Not possible to cl...	Class 5 (Not pos...	Class 5 (Not possible to...	Class 5 (Not possible to classify...
Acute aquatic tox				surface narcotics
Aquatic toxicity c				al Organics



# Handling of tautomerism of target chemical

## Focus of active tautomer

The screenshot shows the QSAR TOOLBOX interface. On the left, the 'Documents' panel lists several tautomers, with 'tautomer #3 (target)' selected. The 'Filter endpoint tree...' panel in the center shows a hierarchical tree of endpoints, with 'Physical Chemical Properties' highlighted. The 'Structure' panel on the right displays the chemical structure of the selected tautomer. A red circle highlights the 'Structure' panel and the 'Physical Chemical Properties' endpoint in the filter tree.

The selected tautomer appears in a new data matrix.

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - Profiling set of tautomers
    - Focus active tautomer
    - **Defining category for active tautomer**

# Category Definition

## Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across/trend analysis.

## Basic guidance for category formation and assessment

Usually, a three stages procedure is recommended for building categories for read-across, in Toolbox. The categorization phases could be organized as follows:

- Stage I: Broad and endpoint non-specific primary categorization of chemicals based on their belonging to common chemical classes, predefined categories or being structurally similar
- Stage II: Subcategorization based on mechanisms conditioning the target endpoint thus coming to endpoint specific subset of chemicals reacting by same interaction mechanisms.
- Stage III: Further narrowing down the category based on elimination of chemicals most dissimilar to target one by using additional structure-related profilers

This sequence of stages is not mandatory and depends on the specificity and number of the chemical analogues and target endpoint. Moreover, some of the stages could be skipped if consistency of category members is reached earlier. It is also recommended only primary categorization to be applied in the Category Definition phase of the Toolbox workflow whereas the subcategorization to be applied at Data gap filling phase; thus, one could follow up the effect of subcategorization on the read-across results (having visualization of the endpoint vs. parameter relationship).

The structural similarity is not recommended to be applied as primary categorization. However, often it is needed to be used in the last stage of the subcategorization – for eliminating most dissimilar chemicals. This holds for read-across implementation for any endpoint.

**Graphical illustration of suitable categorization phases is shown on next slide**

## Suitable Categorization/Assessment Phases

### Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

**Broad grouping  
Endpoint Non-specific**

**Repeating Phase I due to Multifunctionality of chemicals**

### Phase II. Mechanism based\*

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

**Subcategorization  
Endpoint Specific**

**Metabolism accounted for**

### Phase III. Eliminating dissimilar chemicals

**Apply Phase I – for structural dissimilarity  
Filter by test conditions – for Biological dissimilarity**

**Subcategorization  
Endpoint Specific**

## Handling of tautomerism of target chemical

### Category definition for active tautomeric form

- In this exercise, the active tautomer is classified as: Aliphatic amine by ECOSAR category (phase I)
- Searching for similar analogues of the selected active tautomeric form is accomplished using ECOSAR category
- Searching for similar analogues is accomplished using four acute aquatic toxicity databases: Aquatic toxicity OASIS; ECOTOX and ECHA REACH.
- Before defining the category make sure that three aquatic aquatic databases (1) have been selected (see next screenshot)

# Handling of tautomerism of target chemical

## Check databases

The screenshot displays the QSAR Toolbox software interface. The 'Databases' panel on the left shows a list of databases with 'Aquatic toxicity OASIS', 'ECOTOX', and 'ECHA REACH' selected. A red box highlights these three options, and a red circle with the number '1' points to it. The 'Filter endpoint tree...' panel in the center shows a tree structure with 'Aquatic Toxicity' selected. The 'Structure' panel on the right shows a chemical structure. The top toolbar includes buttons for 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The bottom status bar indicates 'The OECD QSAR Toolbox for Grouping Chemicals into Categories' and 'Developed by LMC, Bulgaria'.

1. Check Aquatic toxicity OASIS; ECOTOX and ECHA REACH.

# Handling of tautomerism of target chemical

## Defining ECOSAR category

- The category ECOSAR (strict) is used;
- **Strict** functionality means that the software will identify analogues having ONLY the categories of the target (e.g. aliphatic amines) and will exclude the analogues having any other categories
- Select Aquatic toxicity classification by ECOSAR category
- Click Define (see next screen shots)



# Handling of tautomerism of target chemical

## Defining ECOSAR category

**1** Highlight "Aquatic toxicity classification by ECOSAR"

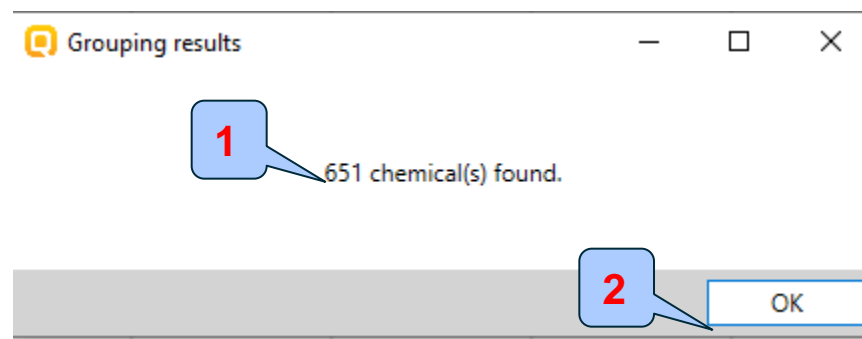
**2** Click **Define**

**3** Select **Strict**

**4** Click **OK** to confirm the category **Aliphatic amines** defined by ECOSAR.

# Handling of tautomerism of target chemical

## Defining ECOSAR category



1. 651 chemicals having Aliphatic amines category are found in the selected databases.
2. Click **OK**

## Handling of tautomerism of target chemical

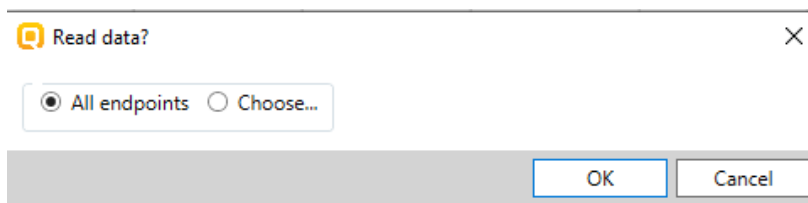
### Category analogues

- The Toolbox now identifies all chemicals corresponding to *Aliphatic amines* by ECOSAR listed in the four aquatic databases.
- 651 analogues including the target chemical are identified; they form a mechanistic category named “**Aliphatic amines**”, which will be used for further data gap filling.
- The experimental data for analogues in the category appears on datamatrix

## Handling of tautomerism of target chemical

### Read data for Analogues

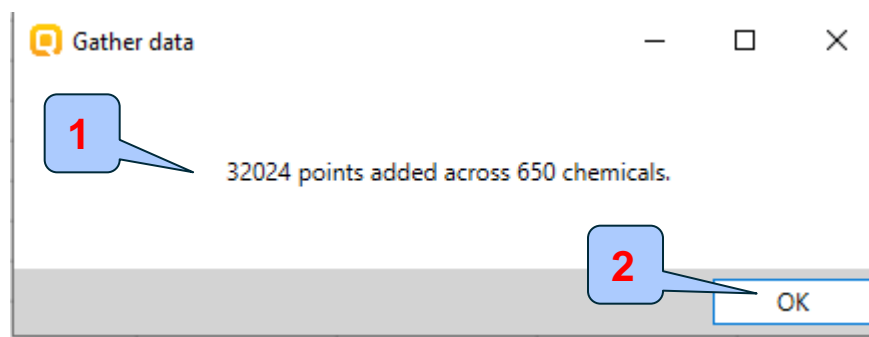
- The Toolbox automatically requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



- In this example, since only databases that contain information for ecotoxicological endpoints are selected, both options give the same results.
- As the Toolbox must search the database, this may take some time.

# Handling of tautomerism of target chemical

## Read data for Analogues



1. 32024 data values across 650 chemicals are found.
2. Click **OK**.

Available aquatic experimental data for the analogues appears in the data matrix.

## Recap

- You have identified a category (“Aliphatic amines”) with the “Acute aquatic toxicity classification by ECOSAR” profiler for the target chemical *Dodecanenitrile* (CAS 2437-25-4)
- The available experimental results for these 650 analogues have been collected from the selected databases (Aquatic ECETOC, ECOTOX and ECHA REACH).
- But before the user can proceed with the “Data Gap Filling” module, he/she should navigate through the endpoint tree and find the specific gap that will be filled.

# Handling of tautomerism of target chemical

## Navigation through the endpoint tree

- The user can navigate through the data tree by opening (or closing) the nodes of the tree.
- The data tree is extensive but logically constructed; it can be mastered with a practice.
- In this example, the “96 h LC50 Mortality for *Pimephales promelas*” is the target endpoint.
- You can navigate through the endpoint tree by typing the species “*Pimephales promelas*” in the “Filter endpoint tree...” box and double click on Aquatic Toxicity, Mortality, LC50, 96 h, Animalia, etc to *Pimephales promelas* - the specific endpoint (see next screenshot)



# Handling of tautomerism of target chemical

## Navigation through the endpoint tree

The screenshot displays the QSAR TOOLBOX interface. The top menu bar includes options like Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows a 'Documents' panel with a tree view of chemical categories, including 'Aliphatic Amines Strict (A)' and 'Tautomerism'. The main workspace is divided into a 'Filter endpoint tree...' panel on the left and a grid of chemical structures on the right. The 'Filter endpoint tree...' panel shows a hierarchical list of endpoints, with 'LC50' highlighted in yellow. The grid on the right shows chemical structures for various endpoints, with the 'LC50' row highlighted in yellow and a red box around it. The 'LC50' row shows the chemical structure of the target chemical, which is a tautomer of the chemical in the 'LC50' row.

Documents

- Document 1
  - # [C: 1;Md: 52;P: 0] CAS: 2437254
    - [C: 4;Md: 52;P: 0] Tautomerism
      - [C: 1;Md: 0;P: 0] tautomer #1
        - [C: 651;Md: 32024;P: 0] Aliphatic Amines Strict (A)
          - [C: 1;Md: 0;P: 0] tautomer #2
            - [C: 1;Md: 52;P: 0] tautomer #3 (target)

Filter endpoint tree...

Structure

- Micropterus dolomieu 1/1
- Micropterus salmoides 1/2
- Oncorhynchus kisutch 1/1
- Oncorhynchus tshawytscha 84/200
- Oncorhynchus tshawytscha 1/1
- Oryzias latipes 23/38
- Perca fluviatilis 1/1
- Pimephales promelas
  - EC50 2/2
  - LC0 1/1
  - LC10 2/2
  - LC100 2/2
  - LC50 83/191
  - LOEC 1/1
  - NOEC 6/6
  - NOEL 1/1
  - NR-ZERO 5/5
  - Undefined End... 2/3
- Pleuronectes platessa 2/2
- Poecilia reticulata 35/43
- Rutilus rutilus 1/1
- Salmo trutta 1/1
- Salvelinus namaycush 1/1
- Sander vitreus 1/2
- Valamugil engeli 1/1
- Tetrapoda 6/19
- Mollusca (molluscs,mollusks) 3/6
- Platyhelminthes (flatworms) 3/4
- Undefined Kingdom 30/70
- > 96 h 1/1
- 0-<200 h 1/1
- 100 h 1/1
- 4.42 d 2/2
- 4.5 d 1/2

Options

Aquatic toxicity classification by ECOSAR

Options

Select All Unselect All Invert

0 Selected

Suitable

- Acute aquatic toxicity classification by Verhaar (Modified)
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- US-EPA New Chemical Categories

Plausible

- Chemical elements
- Groups of elements
- Hydrolysis half-life (K<sub>a</sub>, pH 7)(Hydrowin)
- Hydrolysis half-life (K<sub>a</sub>, pH 8)(Hydrowin)
- Hydrolysis half-life (K<sub>b</sub>, pH 7)(Hydrowin)
- Hydrolysis half-life (K<sub>b</sub>, 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
- Loonski Rule Cases
- OECD HPV Chemical Categories
- Organic functional groups

## Recap

- You have now retrieved the available experimental data on aquatic toxicity for 83 analogue chemicals of focused tautomeric form classified as “Aliphatic amines” by the “ECOSAR” profiler.
- You have identified the target endpoint of “96 h LC50 Mortality for *Pimephales promelas*”.
- You are ready to fill in the data gap so click on “Data Gap Filling” (see next screen shots).

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - Profiling set of tautomers
    - Focus active tautomer
    - Defining category for active tautomer
  - **Trend analysis of the focused tautomer**

# Data Gap Filling

## Apply Trend analysis

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Gap Fill', 'Workflow', 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar contains 'Documents' and 'Data Gap Filling Settings'. The main area displays a 'Filter endpoint tree...' with a list of chemicals and their endpoints. A 'Possible data inconsistency' dialog box is open on the right, showing metadata and scale/unit options. Numbered callouts 1 through 4 highlight specific steps in the workflow.

**1.** Highlight the endpoint box corresponding to *Pimephales promelas*/LC50/96h under the target chemical.

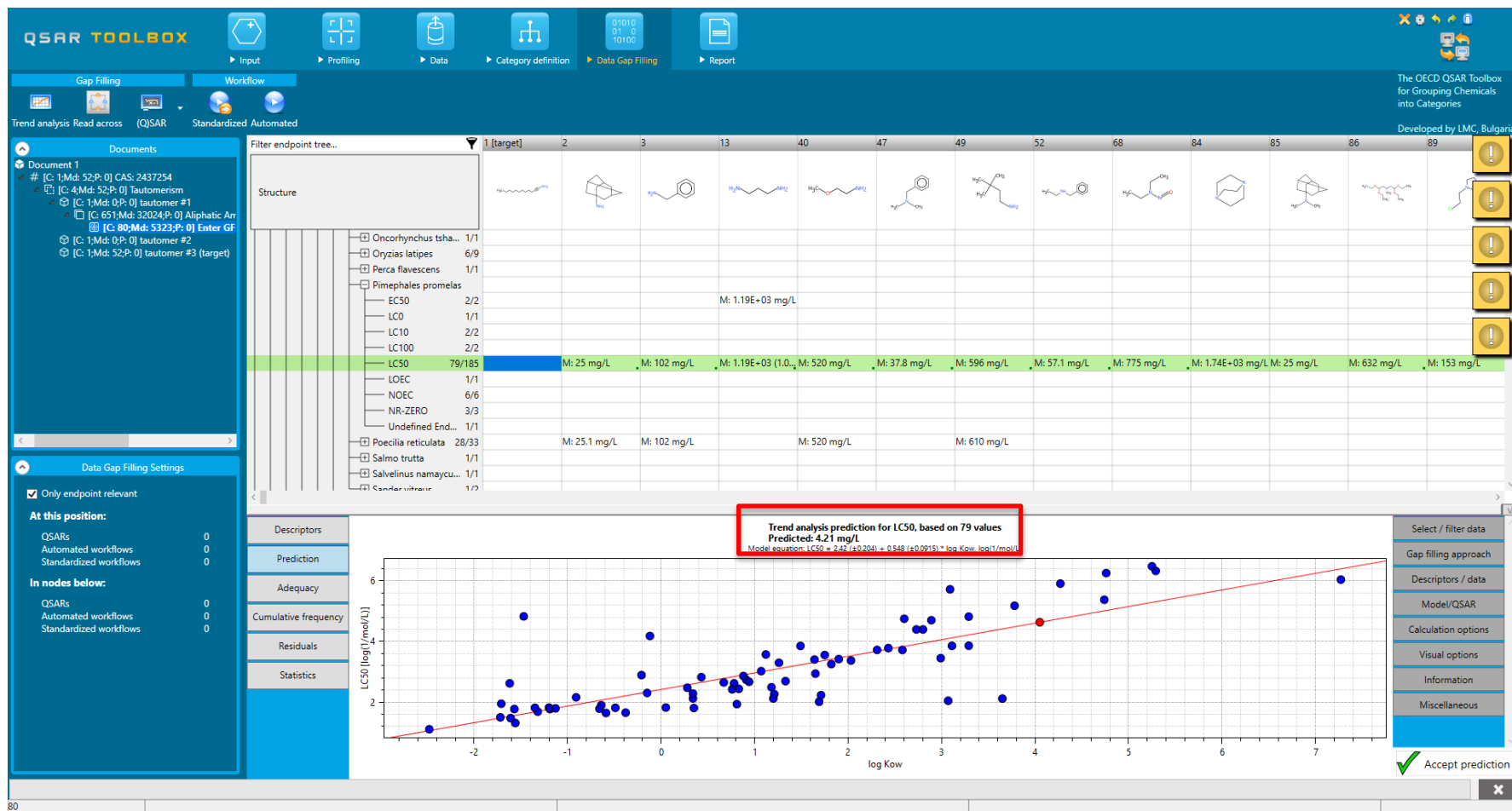
**2.** Select **Trend analysis**

**3.** Select **scale** – log(1 mol/l)

**4.** Click **OK**.

# Data Gap Filling

## Results of Trend analysis



# Data Gap Filling

## Side-Bar of Subcategorisation

- In this example, the following subcategorizations are applied in order to eliminate dissimilar analogues (phase II):

- Chemical elements

The categorisation based on Chemical elements allows keeping among the analogues only those that have the same chemical elements as the target chemical (target tautomeric form).

- Organic functional groups (nested)

Subcategorization by OFG (nested) eliminates dissimilar analogues with respect to structural functionalities. This subcategorization will eliminate structurally dissimilar analogues such as aromatic amines.

Subcategorisation steps are demonstrated on the next slides.

# Data Gap Filling

## Subcategorisation 1 by Chemical elements

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. The interface is divided into several panels:

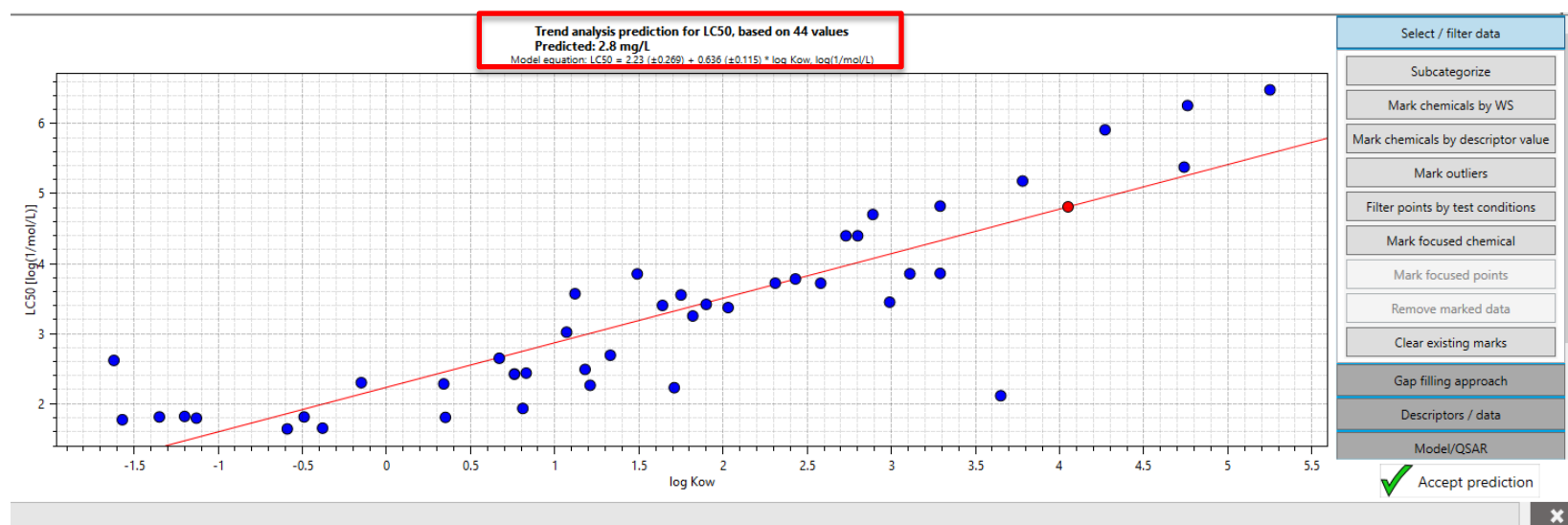
- Options Panel (Left):** Contains a tree view of categorization options. Under the 'Empiric' section, 'Chemical elements' is highlighted with a red box and a callout '2'. Other options include 'Groups of elements', 'Lipinski Rule Oasis', 'Organic functional groups', 'Organic functional groups (nested)', 'Organic functional groups (US EPA)', 'Organic functional groups, Norbert Haider (checkmol)', 'Structure similarity', and 'Tautomers unstable'. The 'Toxicological' section includes 'Repeated dose (HESS)', and the 'Custom' section includes 'Example Prioritization Scheme (PBT)'.
- Adjust options Panel (Top Center):** Shows 'Target' categories: 'Group 14 - Carbon C' and 'Group 15 - Nitrogen N'. Below, it asks to 'Differ from target by' with options 'At least one category' (selected) and 'All categories'. A '[STOP]' button is present.
- Analogues Panel (Bottom Center):** Lists 79 analogues. The first few are '(79) Group 14 - Carbon C', '(79) Group 15 - Nitrogen N', and '(33) Group 16 - Oxygen O'. A blue box highlights the first three items, with a callout '3' pointing to the list.
- Scatter Plot (Center):** A plot of 'log(1/mol/L)' vs. an unlabeled x-axis. Data points are colored blue, red, and green. A red regression line is shown. A callout '1' points to the plot area.
- Right Panel:** A 'Select / filter data' panel with buttons: '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', and 'Gap filling approach'. A green checkmark and 'Accept prediction' button are at the bottom.

Below the interface, a blue box contains the following steps:

1. Click Subcategorize
2. Select **Chemical elements**
3. Click **Remove** to eliminate dissimilar analogues

# Data Gap Filling

## Result of Subcategorisation 1 by Chemical elements





# Data Gap Filling

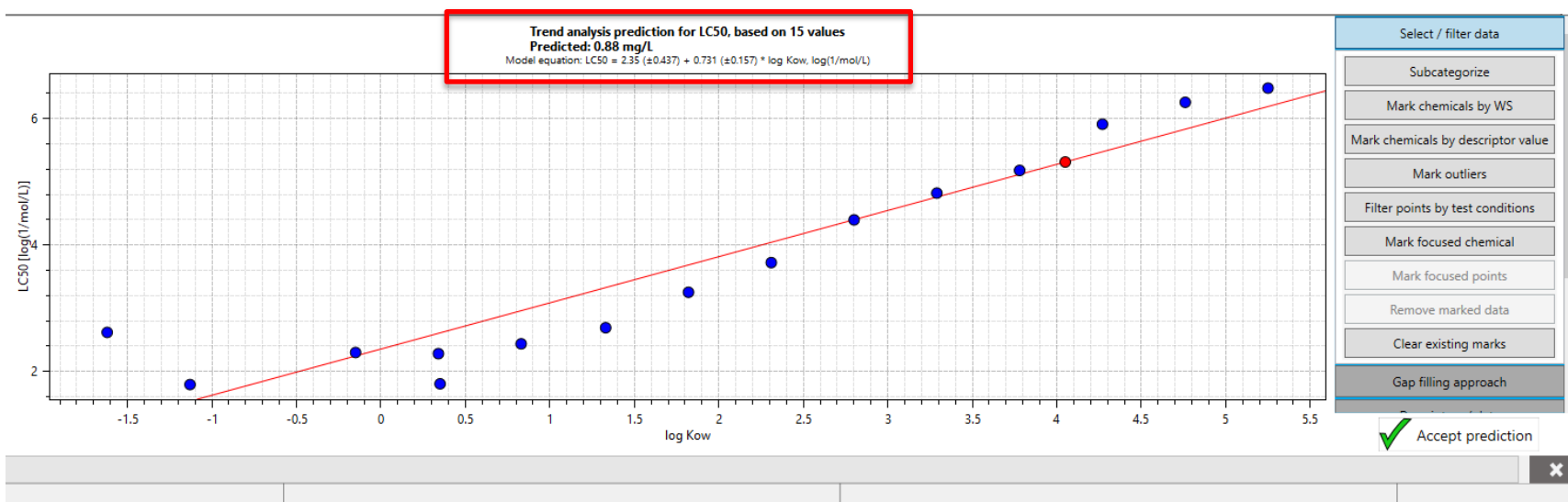
## Subcategorisation 2 by OFG (nested)

The screenshot displays the QSAR Toolbox software interface. On the left, the 'Options' sidebar is open, showing 'Empiric' and 'Toxicological' categories. Under 'Empiric', 'Organic functional groups (nested)' is selected and highlighted with a red box and callout 2. Below this, the 'Metabolisms' section is visible. In the center, the 'Adjust options' panel shows 'Target' and 'Analogues' lists. The 'Analogues' list contains several chemical categories, with 'Aliphatic amine, primary' and 'Aliphatic amine, secondary' highlighted. A red box and callout 3 point to the 'Remove selected' button at the bottom of the 'Analogues' list. On the right, a table displays chemical data with columns for chemical structures and numerical values. Below the table, a scatter plot shows 'log Kow' on the x-axis and 'LC50 [log1/mg/L]' on the y-axis, with a red regression line. A red box and callout 1 point to the 'Subcategorize' button in the right sidebar. At the bottom left, a blue box contains the following instructions:

1. Click **Subcategorize**;
2. Select **OFG (nested)**;
3. Click **Remove** to eliminate dissimilar analogues.

# Data Gap Filling

## Result of Subcategorisation by OFG (nested)



# Data Gap Filling

## Side-Bar of Subcategorisation

The last subcategorisation procedure aims to check and eliminate structurally dissimilar chemicals based on structural similarity

- Structural similarity

The options of structural similarity used in the last subcategorization step are as follows: Dice, Atom centred fragments(ACF), atom features: Atom type; Count H attached; Hybridizations;

Analogues with similarity less than 30 % have been eliminated.

See next two slides.

# Data Gap Filling

## Subcategorisation by Structural similarity

Most dissimilar analogues (0-30 %) are highlighted in green in the data matrix. Most of them are dialiphatic amines and short chain aliphatic amines

1. Select **Structure similarity**;

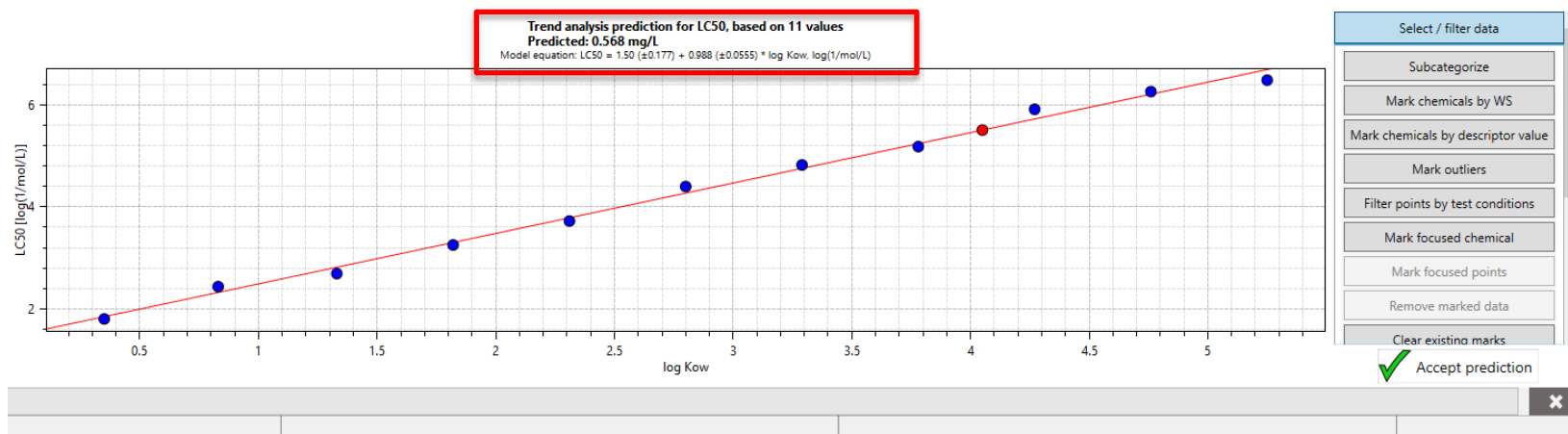
2. Manually select categories between **0 and 30%** (hold Ctrl button and select categories);

3. Dissimilar analogues are highlighted in light blue;

4. Click **Remove** to eliminate dissimilar analogues

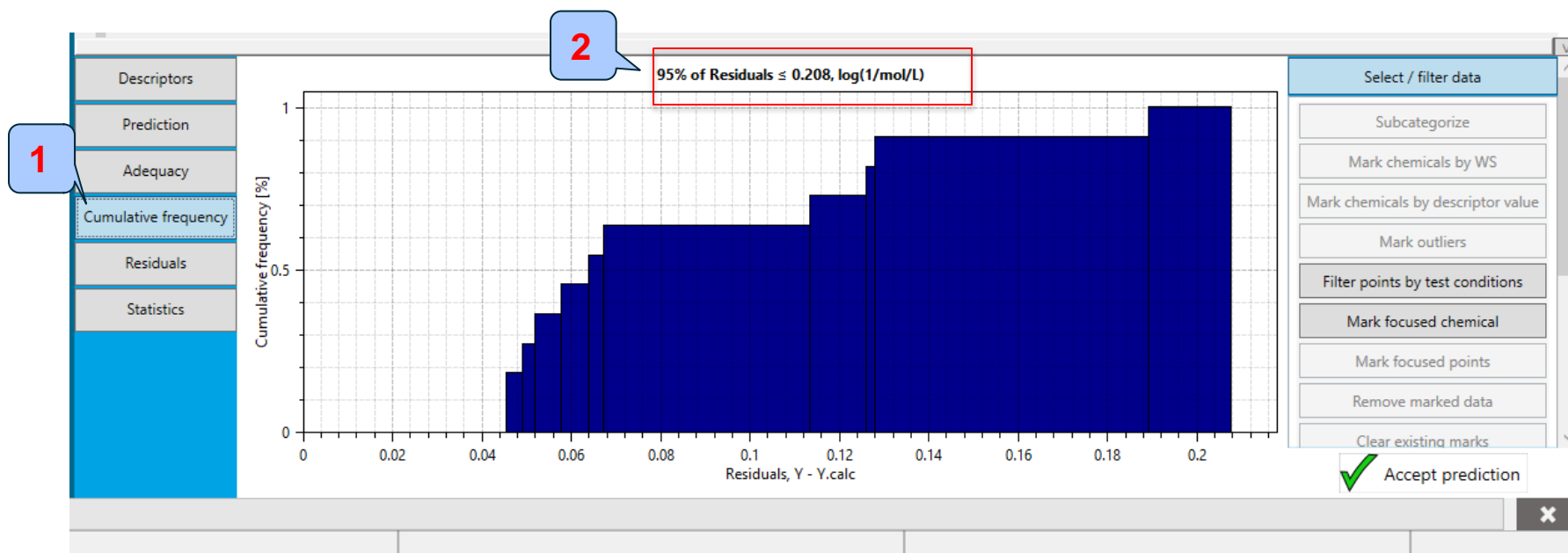
# Data Gap Filling Result

Predicted value:  
0.57 mg/l



# Data Gap Filling

## Cumulated frequency



1. Click **Cumulative frequency**;
2. 95% of residuals are in the range of experimental error

# Data Gap Filling Statistics

Descriptors	Statistical characteristics	TA model	
	Number of data points, (N)	11	1
Prediction	Coefficient of determination, (R <sup>2</sup> )	0.994	3
	Adjusted coefficient of determination, (R <sup>2</sup> <sub>adj</sub> )	0.994	
Adequacy	Coefficient of determination - leave one out, (Q <sup>2</sup> )	0.991	
	Sum of squared residuals, (SSR)	0.143	
Cumulative frequency	Standart deviation of residuals, (sN)	0.114	
	Sample standart deviation of residuals, (s)	0.126	2
Residuals	Fisher function, (F)	1.62E3	
	Fisher treshold for statistical significance, (Fa)	7.21 (95.0%)	
Statistics			
	b0		
	- model descriptor	Intercept	
	- coeff. value	1.50	
	- coeff. range	±0.177	
	- significance	No	
	- max covariation	0.325 vs log Kow	4

Descriptors / data  
 Model/QSAR  
 Show domain  
 Save model  
 Save domain as category  
 Calculate Q<sup>2</sup>  
 Calculation options  
 Visual options  
 Information  
 Miscellaneous  
☒ Accept prediction

1. Select **Model (Q)SAR**;
2. Calculate **Q<sup>2</sup>**;
3. The high **R<sup>2</sup>** and **Q<sup>2</sup>** support the reliability of the prediction;
4. **Accept prediction.**

# Data Gap Filling

## Result of trend analysis

- The analysis of trend analysis shows:
  - The predicted acute aquatic toxicity value is 0.57 mg/l;
  - The remaining analogues form robust category of structurally similar analogues (aliphatic amines);
  - The 95% of residuals are in the range of experimental error;
  - The high R<sup>2</sup> and Q<sup>2</sup> coefficient values support the reliability of the prediction;



# Data gap filling for focused tautomer Trend analysis

The screenshot displays the QSAR Toolbox software interface during a Data Gap Filling workflow. The top menu bar includes options for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The main workspace shows a data matrix with chemical structures in the first column and numerical data in subsequent columns. A red circle highlights a specific data point in the matrix, and a yellow highlight is visible on the row containing this point. The left sidebar shows the 'Documents' panel with a tree view of the project structure, including 'Document 1' and its sub-items. The 'Data Gap Filling Settings' panel is also visible, showing options for 'Only endpoint relevant' and 'At this position:'.

**Documents**

- Document 1
  - # [C: 1;Md: 52;P: 0] CAS: 2437254
    - [C: 4;Md: 52;P: 1] Tautomerism
      - [C: 1;Md: 0;P: 1] tautomer #1
        - [C: 651;Md: 32024;P: 1] Aliphatic
          - [C: 80;Md: 5323;P: 1] Enter GF(T)
            - [C: 45;Md: 2526;P: 1] Subcat
              - [C: 16;Md: 1209;P: 1] Sut
                - [C: 12;Md: 716;P: 1] S

**Data Gap Filling Settings**

☒ Only endpoint relevant

**At this position:**

Select a cell with a rigid (bold) path

Automated workflows 0

Standardized workflows 0

**Filter endpoint tree...**

Structure

- Leuciscus idus 82/261
- Menidia beryllina 2/5
- Micropterus dolom... 1/1
- Micropterus salmoi... 1/2
- Oncorhynchus kis... 1/1
- Oncorhynchus... 84/200
- Oncorhynchus tsha... 1/1
- Oryzias latipes 23/38
- Perca flavescens 1/1
- Pimephales promelas
  - EC50 2/2
  - LC0 1/1
  - LC10 2/2
  - LC100 2/2
  - LC50 84/92
  - LOEC 1/1
  - NOEC 6/6
  - NOEL 1/1
  - NR-ZERO 5/5
  - Undefined End... 2/3
- Pleuronectes plates... 2/2
- Poecilia reticulata 35/43
- Rutilus rutilus 1/1
- Salmo trutta 1/1
- Salvelinus namaycu... 1/1
- Sander vitreus 1/2
- Valamugil engeli 1/1
- Tetrapoda 6/19
- Mollusca (molluscs,mollusks) 3/6
- Platyhelminthes (flatworms) 3/4
- Undefined Kingdom 30/70
- >96 h 1/1
- 0+ <200 h 1/1
- 100 h 1/1

**1 [target]**

2

3

4

5

6

7

8

9

10

11

12

13

M: >21.5+ <46.4...

M: 1E+04 mg/L

M: 0.568 (0.281+1.15) mg/L

M: 25 (22.6+27.6) M: 102 (97.9+10...

M:

M:

M: 25.1 mg/L M: 102 mg/L

The prediction obtained from trend analysis appears on data matrix.

## Data gap filling for focused tautomer

### Interpreting Read-across

- In this example, all analogues are aliphatic amines
- All analogues exhibit toxic effect to fish (*P.promelas*)
- The same toxic effect is therefore predicted for the target (i.e. focused tautomer).
- The prediction of tautomer is further transferred to the parent chemical using “**Transfer to target**” (see next screen shots)

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - Profiling set of tautomers
    - Focus active tautomer
    - Defining category for active tautomer
    - Trend analysis of the focused tautomer
  - **Assigning prediction of tautomer to parent**

# Handling tautomerism of target chemical

## Assigning data to parent chemical

The screenshot shows the QSAR Toolbox software interface during the 'Data Gap Filling' process. The 'Documents' panel on the left lists the current project and its sub-projects. The 'Filter endpoint tree' on the left shows the selected endpoints for the analysis. The central data matrix table displays the results of the analysis, including the parent chemical and its tautomers. A red box highlights a specific data point in the table, and a blue callout points to it. A red box also highlights a specific data point in the table, and a blue callout points to it.

**Documents:**

- current 1
- [C: 1Md: 52P: 0] CAS: 2437254
- [C: 4Md: 52P: 1] Tautomerism
  - [C: 1Md: 0P: 1] tautomer #1
    - [C: 651Md: 32024P: 1] Aliphatic Amines Strict (Aqua)
    - [C: 80Md: 5323P: 1] Enter GF(TA)
    - [C: 45Md: 2526P: 1] Subcategorized: Chemical
    - [C: 16Md: 1209P: 1] Subcategorized: Orga
    - [C: 17Md: 716P: 1] Subcategorized: St
  - [C: 1Md: 0P: 0] t
  - [C: 1Md: 52P: 0]

**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

**Filter endpoint tree...**

- Structure
- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
  - Aquatic Toxicity
    - Behavior
    - Growth
    - Growth Inhibition
    - Growth Rate
    - Mobility
    - Mortality
      - 12 h
      - 24 h
      - 48 h
      - 72 h
      - 96 h
  - Animalia (animals)
    - Chordata (chordates)
      - Actinopterygii (ray-finned...
      - Oryzias latipes
      - Pimephales promelas
    - LC50
  - Undefined Kingdom
- Sediment Toxicity
- Terrestrial Toxicity
- Human Health Hazards
- Profiling
  - Predefined
    - US-EPA New Chemical Categories
  - Endpoint Specific
    - Acute aquatic toxicity classification by...
    - Acute aquatic toxicity MOA by OASIS

**Parent chemical [target] tautomer #1 tautomer #2 tautomer #3 (target) [target]**

Parent chemical [target]	tautomer #1	tautomer #2	tautomer #3 (target) [target]
Structure	Structure	Structure	Structure
1/15 M: non flammable			M: non flammable
1/1 M: 100 %			M: 100 %
1/5 M: >0+ <0.75 mg/L			M: >0+ <0.75 mg/L
1/1 M: 2.28 mg/L			M: 2.28 mg/L
1/1 M: 0.0125 mg/L			M: 0.0125 mg/L
1/2 M: 0.054 mg/L			M: 0.054 mg/L
1/1 M: 0.059 mg/L			M: 0.059 mg/L
1/1 M: >1.5+ <2.25 mg/L			M: >1.5+ <2.25 mg/L
1/1 M: >1.5+ <2.25 mg/L			M: >1.5+ <2.25 mg/L
1/1 M: >0.75+ <1.5 mg/L			M: >0.75+ <1.5 mg/L
1/1 M: >0+ <0.75 mg/L			M: >0+ <0.75 mg/L
1/1 M: 0.84 mg/L			M: 0.84 mg/L
2/1 M: 0.425 mg/L			M: 0.425 mg/L
2/1 M: 0.43 (0.4+0.47) mg/L			M: 0.43 (0.4+0.47) mg/L
1/1 M: 2.34 mg/L			M: 2.34 mg/L
1/18 M: 50 mg/kg bdwt/d			M: 50 mg/kg bdwt/d
Not categorized			Not categorized
Class 5 (Not possible to classif...			Class 5 (Not possible to classify accord...
Basesurface narcotics			Basesurface narcotics

**TA prediction coincide with measured data**

**1**

**2**

1. The trend analysis prediction appears in data matrix;
2. The prediction of the tautomeric form is assigned to the last SMILES within the set;

# Handling tautomerism of target chemical

## Assigning data to parent chemical

**1** Go to **Input**

**2** Select the cell of the **tautomer**, right-click and click **"Transfer to target"**(3)

**3**

The screenshot shows the QSAR Toolbox interface with the following components:

- Top Bar:** QSAR TOOLBOX logo and navigation icons for Input, Profiling, Data, Category definition, and Data Gap Filling (highlighted with callout 1).
- Left Sidebar:**
  - Documents:** A list of chemical documents, including 'current 1' and 'Tautomerism'.
  - Data Gap Filling Settings:** A panel with checkboxes for 'Only endpoint relevant' and 'At this position'.
- Central Table:** A table with columns for 'Parent chemical [target]', 'tautomer #1', 'tautomer #2', and 'tautomer #3 (target) [target]'. It displays chemical structures and various endpoints like 'M: non flammable', 'M: 100 %', and 'M: 0.425 mg/L'.
- Filter endpoint tree:** A tree view on the left of the table showing various endpoints like 'Structure', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Aquatic Toxicity', 'Sediment Toxicity', 'Terrestrial Toxicity', 'Human Health Hazards', 'Profiling', and 'Endpoint Specific'.
- Context Menu:** A right-click menu is shown over a cell in the 'tautomer #3' column, with the 'Transfer to target' option highlighted (callout 3). Other options include 'Explain', 'Delete prediction', 'Explain prediction', 'Set AOP target', 'Use for AOP', and 'Copy'.

# Handling tautomersim of target chemical

## Assigning data to parent chemical

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling

Independent MOA Similar MOA

Documents

current 1  
[C: 1Md: 52P: 1] CAS: 2437254  
[C: 4Md: 52P: 2] Tautomerism  
[C: 1Md: 0P: 1] tautomer #1  
[C: 651Md: 32024P: 1] Aliphatic Amines Strict (Aqua...  
[C: 80Md: 5323P: 1] Enter GF(TA)  
[C: 45Md: 2526P: 1] Subcategorized: Chemical  
[C: 16Md: 1209P: 1] Subcategorized: Orga  
[C: 12Md: 716P: 1] Subcategorized: St  
[C: 1Md: 0P: 0] tautomer #2  
[C: 1Md: 52P: 1] tautomer #3 (target)

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

Select a cell with a rigid (bold) path

Automated workflows 0

Standardized workflows 0

Filter endpoint tree...

Structure

Structure info  
Parameters  
Physical Chemical Properties  
Environmental Fate and Transport  
Ecotoxicological Information

Aquatic Toxicity

Behavior  
Growth  
Growth Inhibition  
Growth Rate  
Mobility  
Mortality  
12 h  
24 h  
48 h  
72 h  
96 h

Animalia (animals)  
Chordata (chordates)  
Actinopterygii (ray-finned...  
Oryzias latipes  
Pimephales promelas

LC50

Undefined Kingdom

Sediment Toxicity  
Terrestrial Toxicity  
Human Health Hazards  
Profiling

Predefined  
US-EPA New Chemical Categories  
Endpoint Specific  
Acute aquatic toxicity classification by...  
Acute aquatic toxicity MOA by OASIS

Parent chemical [target] tautomer #1 tautomer #2 tautomer #3 (target) [target]

1/15 M: non flammable  
1/1 M: 100 %  
1/5 M: >0+ <0.75 mg/L  
1/1 M: 2.28 mg/L  
1/1 M: 0.0125 mg  
1/2 M: 0.054 mg/l  
1/1 M: 0.059 mg/l  
1/1 M: >1.5+ <2.2  
1/1 M: >1.5+ <2.2  
1/1 M: >0.75+ <1  
1/1 M: >0+ <0.75 mg/L  
M: 0.84 mg/L  
M: 0.425 mg/L  
M: 0.43 (0.4+0.47) mg/L  
M: 2.34 mg/L  
M: 50 mg/kg bdwt/d  
Not categorized  
Class 5 (Not possible to classif...  
Basesurface narcotics

M: >0+ <0.75 mg/L  
M: non flammable  
M: 100 %  
M: >0+ <0.75 mg/L  
M: 2.25 mg/L  
M: <2.25 mg/L  
M: <1.5 mg/L  
M: >0+ <0.75 mg/L  
M: 0.84 mg/L  
M: 0.425 mg/L  
M: 0.43 (0.4+0.47) mg/L  
M: 2.34 mg/L  
M: 50 mg/kg bdwt/d  
Not categorized  
Class 5 (Not possible to classif...  
Basesurface

Success

A prediction has been created successfully.

OK

1. Click OK.

# Handling tautomersim of target chemical

## Assigning data to parent chemical

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling

Independent MOA Similar MOA

Documents

current 1  
[C: 1Md: 52P: 1] CAS: 2437254  
[C: 4Md: 52P: 2] Tautomerism  
[C: 1Md: 0P: 1] tautomer #1  
[C: 651Md: 32024P: 1] Aliphatic Amines Strict (Aqua)  
[C: 80Md: 3323P: 1] Enter GF(TA)  
[C: 45Md: 2526P: 1] Subcategorized: Chemical  
[C: 16Md: 1209P: 1] Subcategorized: Orga  
[C: 12Md: 716P: 1] Subcategorized: St  
[C: 1Md: 0P: 0] tautomer #2  
[C: 1Md: 52P: 1] tautomer #3 (target)

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

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

Filter endpoint tree...

Parent chemical [target] tautomer #1 tautomer #2 tautomer #3 (target) [target]

Structure

Structure info  
Parameters  
Physical Chemical Properties  
Environmental Fate and Transport  
Ecotoxicological Information

Aquatic Toxicity

Behavior  
Growth  
Growth Inhibition  
Growth Rate  
Mobility  
Mortality

12 h  
24 h  
48 h  
72 h  
96 h

Animalia (animals)  
Chordata (chordates)  
Actinopterygii (ray-finned...  
Oryzias latipes  
Pimephales promelas

LC50

Undefined Kingdom

Sediment Toxicity  
Terrestrial Toxicity  
Human Health Hazards  
Profiling  
Predefined  
US-EPA New Chemical Categories

1/15 M: non flammable  
1/1 M: 100 %  
1/5 M: >0+ <0.75 mg/L  
1/1 M: 2.28 mg/L  
1/1 M: 0.0125 mg/L  
1/2 M: 0.054 mg/L  
1/1 M: 0.059 mg/L  
1/1 M: >1.5+ <2.25 mg/L  
1/1 M: >1.5+ <2.25 mg/L  
1/1 M: >0.75+ <1.5 mg/L  
1/1 M: >0+ <0.75 mg/L  
1/1 M: 0.84 mg/L  
1/1 M: 2.34 mg/L  
1/18 M: 50 mg/kg bdwt/d  
Not categorized

M: 0.425 mg/L  
M: 0.43 (0.4+0.47) mg/L  
R: 0.568 (0.281+1.15) mg/L

1

M: 0.568 (0.281+1.15) mg/L

M: 0.425 mg/L  
M: 0.43 (0.4+0.47) mg/L  
R: 0.568 (0.281+1.15) mg/L

M: 2.34 mg/L  
M: 50 mg/kg bdwt/d  
Not categorized

1. The Prediction is transferred to the target.

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
    - Profiling set of tautomers
    - Focus active tautomer
    - Defining category for active tautomer
    - Trend analysis of the focused tautomer
  - Assigning prediction of tautomer to parent
  - **Report**



## Report

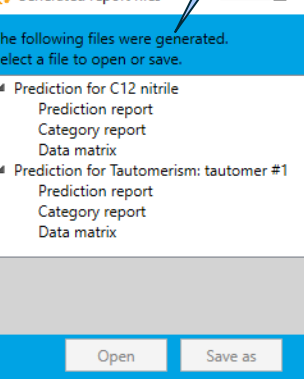
- Remember the report module allows you to generate a report on the predictions performed with the Toolbox. This module contains predefined report templates as well as a template editor with which users can define their own user defined templates. The report can then be printed or saved in different formats.
- The report consist of two sections:
  - Summary report for the whole tautomeric set
  - Report for the individual prediction obtained for the active tautomeric form
- Generating the report is shown on next screenshots

# Report

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report (marked with a red '1'). Below the menu bar, the left sidebar shows 'Documents' and 'Exports' sections. The 'Exports' section includes 'Prediction' (marked with a red '3'), Data Matrix, Category, QMRF, SMI File, SDF File, CAS List, and Data Matrix. The main window shows a 'Filter endpoint tree...' dialog with a tree structure. The tree includes endpoints like Growth Rate, Mobility, Mortality, and LC50. The 'LC50' endpoint is highlighted in yellow (marked with a red '2'). The right side of the dialog shows the results for the selected endpoint, including a chemical structure and a table of values.

Endpoint	Value
Growth Rate	1/2 M: 0.054 mg/L
Mobility	1/1 M: 0.059 mg/L
Mortality	1/1 M: >1.5+ <2.25 mg/L
12 h	1/1 M: >1.5+ <2.25 mg/L
24 h	1/1 M: >1.5+ <2.25 mg/L
48 h	1/1 M: >0.75+ <1.5 mg/L
72 h	1/1 M: >0+ <0.75 mg/L
96 h	1/1 M: 0.84 mg/L
Animalia (animals)	
Chordata (chordates)	
Actinopterygii (ray-finned fishes, spiny r...	
Oryzias latipes	1/1 M: 0.425 mg/L
Pimephales promelas	1/3 M: 0.43 (0.4+0.47) mg/L R: 0.568 (0.281+1.15) mg/L
LC50	1/1 M: 2.34 mg/L
Undefined Kingdom	
Sediment Toxicity	
Terrestrial Toxicity	

1. Click on section **Report**;
2. Select the **Prediction** (marked with "R");
3. Click **Prediction**.



Generated report files

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

- Prediction for C12 nitrile
  - Prediction report
  - Category report
  - Data matrix
- Prediction for Tautomerism: tautomer #1
  - Prediction report
  - Category report
  - Data matrix

Open Save as

- 91

# Report

Prediction of LC50 for C12 nitrile


1 / 6

## QSAR Toolbox prediction for single chemical

Date: 14 Apr 2020

Author(s):

Contact details:

Target information		
Structural information	Numerical identifiers	Chemical names
SMILES: <chem>CCCCCCCCCCCC#N</chem>	CAS#: 2437-25-4 Other: EC Number:2194401	15/C12H23N/c1-2-3-4- 5-6-7-8-9-10-11-12-1 3/h2-11H2,1H3 C(CCCCCC)CCCCC#N C12 nitrile
Structure 		

1

Prediction summary
Predicted endpoint: LC50; Mortality; Pimephales promelas; 96 h; No guideline specified Predicted value: 0.568 (from 0.281 to 1.15) Unit/scale: mg/L Data gap filling method: Read-across analysis Summary: manually editable field Not provided by the user

1. Predicted value

# Outlook

- Background
- Keywords
- Objectives
- The exercise
- Workflow
  - Input
  - Profiling
  - Data
  - Handling of tautomerism of target chemical
  - Assigning prediction of tautomer to parent
  - Report
- **Save prediction**

## Saving the prediction result

- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving/Loading the file with TB prediction is shown on next screenshots

# Saving the prediction

The screenshot shows the QSAR Toolbox interface with a 'Toolbox' dialog box open. The dialog box contains the text: 'Do you want to save changes to document 'Document 1'?'. Below the text are two buttons: 'Yes' and 'No'. A blue callout box with the number '3' points to the 'Yes' button. Another blue callout box with the number '2' points to the 'Save' button in the top toolbar. A third blue callout box with the number '1' points to the 'Input' button in the top toolbar. The background shows a chemical structure and a table of predicted values for various endpoints.

Endpoint	Parent chemical [target]	tautomer #1	tautomer #2	tautomer #3 (target) [target]
Structure	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
Structure info				
Parameters				
Physical Chemical Properties	1/15 M: non flammable			
Environmental Fate and Transport	1/1 M: 100 %			
Ecotoxicological Information				
Aquatic Toxicity				
Behavior	1/5 M: >0+ <0.75 mg/L			
Growth	1/1 M: 2.28 mg/L			
Growth Inhibition	1/1 M: 0.0125 mg/L			
Growth Rate	1/2 M: 0.054 mg/L			
Mobility	1/1 M: 0.059 mg/L			
Mortality				
12 h	1/1 M: >1.5+ <2.25 mg/L			
24 h	1/1 M: >1.5+ <2.25 mg/L			
48 h	1/1 M: >0.75+ <1.5 mg/L			
72 h	1/1 M: >0+ <0.75 mg/L			
96 h				
Animalia (animals)				
Chordata (chordates)				
Actinopterygii (ray-finned...)	1/1 M: 0.84 mg/L			
Oryzias latipes				M: 0.84 mg/L
Pimephales promelas				
LC50	2/4 M: 0.425 mg/L M: 0.43 (0.4+0.47) mg/L R: 0.568 (0.281+1.15) mg/L	T: 0.568 (0.281+1.15) mg/L		M: 0.425 mg/L M: 0.43 (0.4+0.47) mg/L R: 0.568 (0.281+1.15) mg/L
Undefined Kingdom	1/1 M: 2.34 mg/L			M: 2.34 mg/L
Sediment Toxicity				
Terrestrial Toxicity				
Human Health Hazards	1/18 M: 50 mg/kg bdwt/d			M: 50 mg/kg bdwt/d
Profiling				
Predefined				
US-EPA New Chemical Categories	Not categorized	Not categorized		Not categorized

1. Select **Input**;
2. Click on **Save** button;
3. Click **Yes**.

# Saving the prediction

The screenshot shows the QSAR Toolbox interface with a 'Save document' dialog box open. The dialog box is titled 'Save document 'Document 1'' and shows the file name 'Tutorial 15' and the save type 'Toolbox documents (\*.tb4)'. A red callout bubble with the number '1' points to the 'Save' button. To the right, a 'File save' confirmation message box says 'File saved successfully!' with an 'OK' button. A red callout bubble with the number '2' points to the 'OK' button. The background shows the main interface with a tree view on the left and a table of results on the right.

Chemical	Endpoint	Value	Unit	MS	TS	TS (0.281+1.15)
Oryzias latipes	2/2	M: 0.84 mg/L		MS: 0.84 mg/L		
Pimephales promelas		M: 0.425 mg/L		MS: 0.425 mg/L		
LC50	3/6	M: 0.43 (0.4+0.47) mg/L		MS: 0.43 (0.4+0.47) mg/L		
Undefined Kingdom	2/2	M: 2.34 mg/L		MS: 2.34 mg/L		
Sediment Toxicity						
Terrestrial Toxicity						
Human Health Hazards	2/36	M: 50 mg/kg bdwt/d		MS: 50 mg/kg b...		
Profile						
Predefined						
US-EPA New Chemical Categories		Not categorized				
Endpoint Specific						
Acute aquatic toxicity classification by...		Class 5 (Not possible to classi...				
Acute aquatic toxicity MOA by OASIS		Basesurface narcotics				
Aquatic toxicity classification by ECOS...		Neutral Organics				

**1. Save the file;**  
**2. Click OK.**



# Open saved file

