

OECD QSAR Toolbox v.3.4

Predicting acute aquatic toxicity to fish of 4-methyl-2-nitroaniline taking into account tautomerism

Outlook

- **Background**
- 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 data gap for acute aquatic toxicity to fish taking into account tautomerism of target chemical.

Outlook

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

Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
 - Providing tautomeric set of target chemical
 - Identify analogues for a set of tautomers
 - Retrieve experimental results available for those analogues
 - Filling data gap by trend-analysis
 - Save the obtained prediction result

Outlook

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

The Exercise

- In this exercise we will predict *LC50* for fish: *P.promelas* for target chemical
- Set of simulated tautomers for the target chemical will be provided
- This prediction will be accomplished by collecting similar analogues presented with their tautomeric set
- The category will be defined using US-EPA New Chemical Categories
- Data gap will be filled by trend-analysis

Outlook

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

Workflow

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

Outlook

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

Chemical Input

There are two ways for simulating tautomers of chemicals

- During the process of entering the structure into the system
- Simulating tautomersim of already entered structure

In order to accelerate the workflow all databases are preliminary tautomerized, calculated and profiled. The results are stored in the database.

See next screen shots

Chemical Input

Input target chemical by CAS#

1. **Click** on CAS#; 2. **Enter** 89-62-3; 3. **Select** Tautomeric sets; 4. **Click** Search; 5. **OK**

Note: Tautomeric set functionality search tautomeric forms of entered chemical in previously tautomerized databases

Chemical Input

Target chemical identity

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

- Top Menu:** Document, Single Chemical, Chemical List.
- Toolbar:** New, Open, Close, Save, CAS#, Name, Structure, Select, Delete, Query, ChemIDs, DB, Inventory, List.
- Left Panel:** Document tree showing a folder for CAS: 89-62-3.
- Center Panel:** Filter endpoint tree... with a red circle around the target structure and a callout box labeled '1'.
- Right Panel:** Target information including CAS# 89-62-3, EINECS:2019249, and various chemical names like benzenamine, 4-methyl-2-nitroaniline, 4-methyl-2-nitrobenzenamine, 2-nitro-4-methylaniline, and 2-nitro-4-methylphenylamine. The SMILES string Cc1ccc(N)c(N(=O)=O)c1 is also shown.
- Bottom Panel:** A dialog box showing five tautomeric forms of the target chemical, each with its own structure and CAS# 89-62-3. The dialog includes 'Save to smi', 'Search', and 'OK' buttons.

Tautomeric forms

Target with its tautomeric forms are identified and loaded on data matrix.
 1. **Double click** over the target structure displays target and its tautomeric forms

Chemical Input

Implementation of Modeling modes

- Component Mode All** – all tautomeric forms are analyzed in a package

The screenshot shows the QSAR Toolbox interface. The 'Component Mode' dropdown menu is open, and 'All' is selected. The main workspace displays a single chemical structure for the target compound, which is circled in red. The structure is a benzamide derivative with a nitro group and a methyl group.

- Component Mode Single** – each tautomeric form is analyzed individually

The screenshot shows the QSAR Toolbox interface with 'Component Mode Single' selected. The main workspace displays six individual chemical structures, each representing a different tautomeric form of the target compound. These structures are circled in red. The structures are labeled 1 through 6, with 1 being the target and 2-6 being its tautomeric forms.

Different modes for visualization for the set of target and its tautomeric forms is implemented. A package of target and its tautomeric forms are used in further trend analysis.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - **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 displays the QSAR Toolbox software interface. The top navigation bar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The Profiling Methods sidebar on the left contains various predefined and general methods, with 'US-EPA New Chemical Categories' highlighted by a red circle and callout '1'. The 'View' button in the Profiling Methods sidebar is also highlighted by a red circle and callout '2'. The Profiling Scheme Browser window is open, showing a list of categories on the left and a detailed textual description for 'Aldehydes Environmental Toxicity' on the right. The textual description includes information about the functional group, toxicity, and hazard concerns for aldehydes.

1. Highlight the profiler

2. Click View

Profiling Side-Bar to Profiling

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The Profiling side-bar is visible on the left, with the 'View' button highlighted by a red circle and labeled with a '2'. The Profiling Scheme Browser window is open, showing a list of predefined categories. The 'Aldehydes (Chronic toxicity)' category is highlighted with a red circle and labeled with a '1'. The main window shows a structural boundary diagram with various nodes and logical operators (AND, OR, NOT). The 'Boundary Options' section is set to 'Metabolism', and the 'Structural fragment' section shows the chemical structure of an aldehyde: C(H)(=O)C.

1. Highlight the profiler
2. Click View

Profiling

Profiling the set of target and tautomers

- The following primary profilers relevant to the **aquatic toxicity** are used in this example (see next screenshot):
 - OECD HPV Chemical Categories
 - US-EPA New chemical category
 - Aquatic toxicity classification by ECOSAR
 - Acute aquatic toxicity MOA by OASIS
 - Acute aquatic toxicity classification by Verhaar (Modified)
 - Organic function groups – all four profilers are used in the assessment
- Select the “Profiling methods” related to the target endpoint by clicking on the box next to the profilers name.
- This selects (a **green** check mark appears) or deselects (**green** check disappears) profilers.

Profiling

Profiling the set of target and tautomers

1 Acute aquatic toxicity MOA by OASIS categories of tautomers in the set

2 Apply

3 Component Profile Statistic

4 Component Profile Statistic

5 2D representations of the structures in each category

Number of tautomers in a category bin

#	Category	Count	%
1	Basesurface narcotics	1	20.00
2	Phenols and Anilines	1	20.00
3	Reactive unspecified	3	60.00

1. Check the profilers related to acute aquatic toxicity as mention on slide #20

2. Click Apply

3. Perform Right click over the cell with profiling results and

4. Select Component profile statistic

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - **Endpoint**

Endpoint

- “Endpoint” refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- In this example, we limit our data gathering to the common aquatic toxicity endpoints from databases containing aquatic toxicity data - **Aquatic ECETOC; Aquatic OASIS; Aquatic Japan MoE; ECOTOX**
- Data for target chemical and its simulated tautomeric forms is extracted from selected databases if available

Endpoint

- For this example, the following database are relevant to the aquatic toxicity (see next screen shot):
 - ✓ Aquatic ECETOC
 - ✓ Aquatic Japan MoE
 - ✓ Aquatic OASIS
 - ✓ ECOTOX

Endpoint Gather data

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', 'Export', 'Delete', and 'Tautomerize'. The 'Endpoint' workflow is selected. In the 'Databases' panel on the left, several databases under 'Ecotoxicological Information' are checked with green checkmarks. The 'Gather' button is highlighted in the top menu. The main panel shows the chemical structure and identity details for a target compound.

Filter endpoint tree...	1 [target]
Structure	[5] [1]
Structure	
Substance Identity	
— CAS Number	89-62-3
— Chemical IDs	EINECS:2019249
— Chemical Name	benzenamine, 4-m- 4-methyl-2-nitroanil... 2-nitro-p-toluidine 4-methyl-2-nitro-be... p-toluidine, 2-nitro- benzenamine, 4-m- 4-methyl-2-nitrobe... 2-nitro-4-methylanil... C7H8N2O2
— Molecular Formula	C7H8N2O2
— Structural Formula	Cc1ccc(N)c(N(=O))...
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	
Profile	

- 1. Select** databases related to the target endpoint by adding a green check in the box before the database name.
- 2. Click** Gather

Endpoint

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Endpoint' workflow is active, showing a 'Filter endpoint tree...' window. On the left, the 'Databases' panel is visible, with 'Ecotoxicological Information' checked. The tree structure shows 'Actinopterygii (Fish)' selected, with a callout '1' pointing to the data matrix. The data matrix shows experimental data for species like *Lepomis macrochirus* and *Tetrahymena pyriformis*.

Endpoint	Value
48 h	
96 h	
- Animalia	
- Arthropoda (Invertebrates)	
- Chordata (Vertebrates)	
- Actinopterygii (Fish)	
- <i>Lepomis macrochirus</i>	
- <i>Leuciscus idus</i>	
- <i>Pimephales promelas</i>	(1/2) M: 24.7 mg/L, 24.8
- <i>Poecilia reticulata</i>	
- Undefined Test organisms (species)	
- Mollusca (Invertebrates)	
- Rotifera	
- 14 Days	
- LD50	
- MRCS0	
- Photoinduced Toxicity	
- Physiology	
- Population	
- IC50	
- 60 h	
- Protozoa	
- Ciliophora	
- Ciliata	
- <i>Tetrahymena pyriformis</i>	(1/1) M: 67.2(45.1;113) ...

1. Available experimental data appears on datamatrix.

Recap

- You have entered the target chemical
- You have simulate the tautomeric forms of target chemical
- You have gather data if available and found experimental data for one of tautomeric forms (in our case for entered structure).
- Now you are ready to continue with next step of the workflow “Category definition”.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - **Category definition**

Category Definition Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition.

Category Definition

Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity”.
- Detailed information about grouping chemical (Chapter 4) could be found in document “Manual for Getting started” published on OECD website:

<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers
2. Endpoint specific profilers (for sub-cat)
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements)

Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling
3. More categories of same Phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached

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**

*Another general approach for development of categories for acute aquatic tox endpoints are summarized in document "Strategies for grouping chemicals for data gap filling for acute aquatic toxicity endpoints" posted on OECD Website

Category Definition

Grouping methods

- For this example, specific classifications of target and its tautomers are identified by the following profilers: US-EPA, MOA of action and EcoSAR (phase I)
- For this example analogues identified by US-EPA New chemicals category are used for further data gap filling
- Subsequent search of analogues is applied over the set of tautomers having same categories as those of the target tautomeric set

Category definition is a tool for grouping chemicals, which allows to group chemicals based on different measures of "similarity". For more details see tutorials posted on LMC and OECD website:

<http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm>

<http://superhosting.oasis-lmc.org/products/software/toolbox/toolbox-support.aspx>

Also we strongly recommend training exercises. For more details see:

<http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm>

Category Definition

Side-bar of US-EPA New chemical categories

- US-EPA has been used by the U.S. Environmental Protection Agency to predict the aquatic toxicity of new industrial chemicals in the absence of test data
- US-EPA include classes of chemicals for which sufficient regulatory history has been accumulated
- “Classification by US-EPA” in the Toolbox is used for grouping of chemicals by structural similarity which may have mechanistic meaning. Experience has shown US-EPA to be a robust profiler which makes it a logical choice in an initial profiling scheme.

Category Definition

Defining US-EPA New Chemical category

The screenshot illustrates the steps for defining a new chemical category in the QSAR Toolbox. The 'Define' button in the toolbar is highlighted with a red circle and labeled '2'. In the 'Filter endpoint tree...', the 'US-EPA New Chemical Categories' endpoint is highlighted with a red circle and labeled '1'. A dialog box titled 'US-EPA New Chemical Categories' is open, showing 'Target(s) profiles' with 'Anilines (Acute toxicity)' and 'Not categorized'. The 'Not categorized' option is selected and circled in red, labeled '3'. Below the dialog, a list of 'All profiles' is shown, with an arrow pointing down to the 'Not categorized' option, labeled '4'. The dialog also has 'Combine profiles logically' (AND selected), 'Invert result', and 'Strict' options, along with 'OK' and 'Cancel' buttons.

Chemicals (in this case tautomers) without identified toxic category are classified as "Not categorized" This is the reason to exclude this category from the set of searched analogues.

1. **Highlight** "US-EPA New Chemical Categories "
2. **Click** Define

3. **Select** "Not categorized"
4. **Click** arrow down to remove selected category

Category Definition

Defining US-EPA New Chemical category

The screenshot shows the QSAR Toolbox software interface. The main window displays a hierarchical tree of endpoints under the 'Filter endpoint tree...' tab. A callout box labeled '1' points to the 'Anilines' category in the tree. Another callout box labeled '2' points to the 'OK' button in the 'US-EPA New Chemical Categories' dialog box, which is open over the tree. The dialog box shows a list of profiles, with 'Anilines (Acute toxicity)' selected. The 'OK' button is circled in red.

1. Category **"Anilines"** is used for further categorization
2. **Click OK**

Category Definition

Defining US-EPA New Chemical category

The screenshot shows the QSAR Toolbox software interface. The main window displays a 'Filter endpoint tree...' with a hierarchical structure of biological endpoints. The 'Actinopterygii (Fish)' category is expanded, showing sub-categories like *Lepomis macrochirus*, *Leuciscus idus*, *Pimephales promelas*, and *Poecilia reticulata*. A warning dialog box is displayed, stating: "Warning: You have selected different from target categories! Do you want to continue?". The 'Yes' button in the dialog is circled in red, and a blue callout box with the number '1' points to it.

A notification message informs you that you have selected different categories from those of the target.

1. **Select Yes**

Category Definition

Defining US-EPA New Chemical category

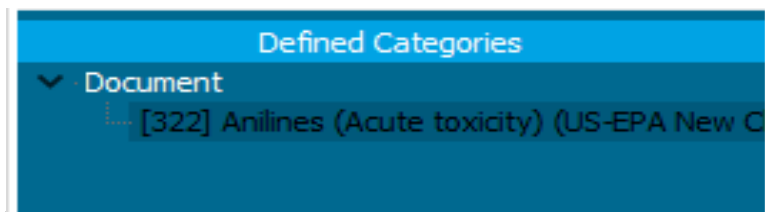
The screenshot shows the QSAR Toolbox software interface. The main window displays a hierarchical tree of endpoints under the heading "Filter endpoint tree...". The tree is organized into several main categories, including "48 h", "96 h", "14 Days", "LD50", "MRC50", "Photoinduced Toxicity", "Physiology", "Population", "IC50", "60 h", and "Tetrahyena pyriformis". A "Define category name" dialog box is open, showing "Category name (322 chemicals)" with the text "acute toxicity (US-EPA New Chemical Categories)" entered. The "OK" button is highlighted with a blue callout box containing the number "1".

1. Click OK

Category Definition

Analogues

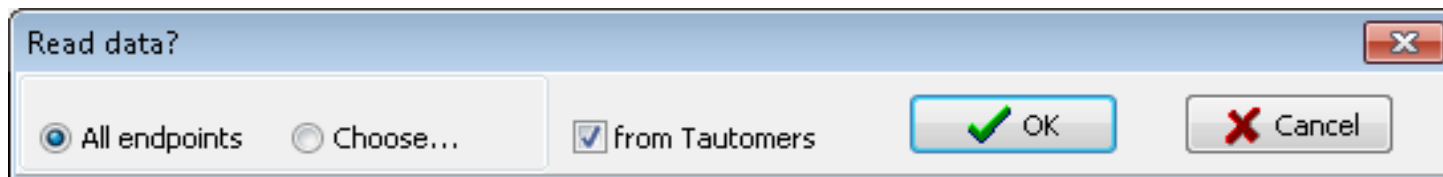
- The Toolbox now identifies all chemicals represented as tautomeric sets corresponding to the US-EPA classification of “Anilines” which are listed in the databases selected under “Endpoint”.
- 322 analogues(tautomeric sets) are identified. Along with the target they form a category (Anilines) which can be used for data gap filling.
- The name of the category appears in the “Defined Categories” window, along with the number of substances belonging to the category.



Category Definition

Read 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).



- 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.




Category Definition

Read data for Analogues

Due to overlap between the Toolbox databases for intersecting chemicals the same data may be found simultaneously. Data redundancies are identified and the user has the opportunity to select either a single data value or all data values.

Repeated values for: 842 data-points, 329 groups, 153 chemicals

Data points...

	Endpoint	CAS	Structure	Value	additional_comme
<input checked="" type="checkbox"/>	NOEC	62-53-3		0.004 mg/L	
<input type="checkbox"/>	NOEC	62-53-3		4 micrograms per liter	
<input checked="" type="checkbox"/>	NOEC	95-51-2		0.032 mg/L	
<input type="checkbox"/>	NOEC	95-51-2		0.032 mg/L	
<input type="checkbox"/>	NOEC	95-51-2		32 micrograms per liter	
<input checked="" type="checkbox"/>	NOEC	95-76-1		3.1 micrograms per liter	NR
<input type="checkbox"/>	NOEC	95-76-1		3.1 micrograms per liter	NR
<input type="checkbox"/>	NOEC	95-76-1		3.1 micrograms per liter	NR

Buttons: Select one (1), Invert, Check All, Uncheck All, OK (2), Cancel

1. Click Select one and then
2. Click OK

Category Definition

Summary information for Analogues

The screenshot displays the QSAR Toolbox interface with the 'Category Definition' workflow selected. The left sidebar shows 'Grouping methods' with 'Predefined' and 'General Mechanistic' categories. The main workspace shows a 'Filter endpoint tree...' and a 'Datamatrix' table. The 'Ecotoxicological Information' row is highlighted in red, showing numerical data for various endpoints across 8 columns. A blue callout box at the bottom states: 'Available aquatic experimental data for the analogues represented as tautomeric sets appears on datamatrix.'

Filter endpoint tree...	1 [target]	2	3	4	5	6	7	8	
Structure	[5] [T]	[1] [T]	[1] [T]	[1] [T]	[9] [T]	[1] [T]	[2] [T]	[5] [T]	
Substance Identity									
Physical Chemical Properties									
Environmental Fate and Transport									
Ecotoxicological Information	(308/6261)	M: 24.7 mg/L, 24.8...	M: 0.004 mg/L, 0.0...	M: 0.032 mg/L, 0.1...	M: 0.001 mg/L, 0.0...	M: 845 mg/L, 1.44...	M: 1.29E3 mg/L, 1...	M: 0.014 mg/L, 0.0...	M: 411 r
Human Health Hazards									
Profile									

Recap

- You have identified a category (“anilines”) with the “US-EPA New Chemical Categories” profiler for the target chemical 4-methyl-2-nitroaniline and its tautomeric forms
- The available experimental results for these 322 analogues represented as tautomeric sets have been collected from the selected databases (Aquatic ECETOC, Aquatic Japan MoE, Aquatic USEPA ECOTOX, and Aquatic OASIS).
- But before the user can proceed with the “Filling Data Gap” module, he/she should navigate through the endpoint tree and find the specific gap that will be filled.

Category Definition

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)

Category Definition

Navigation through the endpoint tree

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

- Top Bar:** Navigation tabs for Input, Profiling, Endpoint, Category Definition (active), Data Gap Filling, and Report.
- Filter Box:** Contains the text 'pimephales' with a blue callout '1' pointing to it.
- Endpoint Tree:** A hierarchical tree on the left showing categories like 'Aquatic Toxicity', 'Mortality', and 'Actinopterygii (Fish)'. The entry 'Pimephales promelas' is circled in red, with a blue callout '2' pointing to the '+' icon next to it.
- Main Table:** A table with 8 columns representing different endpoints. The first column is labeled '1 [target]' and contains a chemical structure. Other columns contain numerical data and molecular weights (M).

- 1. Type** "Pimephales promelas" in the filter box, then **press** Enter
- 2. Open** the tree to the target endpoint by **single left** click on the **+** sign

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 322 analogue chemicals of target and its tautomeric forms classified as “anilines” by the “US-EPA New Chemical Categories” profiler.
- You have identified the target endpoint of “96 h LC50 Mortality for *Pimephales promelas*”.
- There is an experimental data for the investigated endpoint, in our exercise we will try to reproduce the experimental data taking into account tautomeric forms of the target
- You are ready to fill in the data gap so click on “Data Gap Filling” (see next screen shot).

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Category definition
 - **Data Gap Filling**

Data Gap Filling

Apply Trend analysis

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

- Top Bar:** Navigation tabs for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Left Sidebar:**
 - 3:** 'Apply' button circled in red.
 - 2:** 'Trend analysis' option selected in the 'Data Gap Filling Methods' section.
 - Target Endpoint:** Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata Actinopterygii Pimephales promelas.
- Main Panel:**
 - Structure:** Chemical structure of Pimephales promelas.
 - Substance Identity:** Pimephales promelas.
 - Ecotoxicological Information:**
 - Aquatic Toxicity:**
 - Behavior (3/14)
 - Development (4/19)
 - Growth (4/57)
 - Mortality (5/7)
 - EC50 (1/1)
 - LC01 (1/4)
 - LC50 (1/2)
 - 1 h (1/2)
 - 3 h (1/2)
 - 6 h (1/1)
 - 12 h (1/1)
 - 24 h (12/19)
 - 48 h (11/15)
 - 72 h (3/3)
 - 96 h (79/211)
 - 1:** M: 24.7 mg/L, 24.8... (highlighted in red)
 - Chordata (Vertebrates)
 - Actinopterygii (Fish)
 - Pimephales promelas (79/211)
 - 5 Days (1/1)
 - 7 Days (2/2)

There is an experimental data for the target, we will try to reproduce it taking into account tautomerism (to check experimental data)

- 1. Highlight** the endpoint box corresponding to *Pimephales promelas*/LC50/96h under the target chemical.
- 2. Select** Trend analysis
- 3. Click** Apply

Data Gap Filling

Results of Trend analysis

The screenshot displays the QSAR Toolbox interface during a trend analysis. The main window shows a scatter plot of LC50 (obs.) in mg/L on the y-axis (ranging from 2.0 to 6.0) against log Kow on the x-axis (ranging from -1.00 to 6.00). A red regression line is fitted to the data points. A specific data point is highlighted in yellow, corresponding to the target chemical, Pimephales promelas, with an observed LC50 of 24.7 mg/L and a predicted LC50 of 11.3 mg/L. The model equation is $LC50 = +3.73 + 0.206 * \log Kow$.

On the right side of the interface, a 'Visual options' menu is open. Callout box 1 points to the 'Visual options' menu item, and callout box 2 points to the 'Show all members of chemical sets' option within that menu.

Visualization of members of chemical sets is possible when **click** on 1. Visual options, then **2. Select** Show all members of chemical sets

Data Gap Filling

Results of Trend analysis

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Filling
Apply

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models

Target Endpoint

Ecotoxicological Information Aquatic Toxicity
Mortality LC50 96 h Animalia Chordata Actinopterygii
Pimephales promelas

Structure

1 [target] 2 3 4 5 6 9 16

5] [1] [1] [1] [1] [9] [1] [1] [1] [1]

Pimephales promelas (79/215) M: 24.7 mg/L, 24.8 M: 75.5(68.4, 83.4) M: 5.68(5.34, 6.04) M: 8.06(7.26, 8.95) M: 1.44E3 mg/L, 1 M: 158 mg/L, 171 M: 1.75 mg/L, 3.24 M: 3

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 78 values from 78 analogue chemicals, Observed target value: 24.7 mg/L, Predicted target value: 11.3 mg/L, Model equation: $LC50 = +3.73 + 0.206 * \log Kow$

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
 - Set units in figure title
 - Set axes ranges
 - Show one point per chemical set
 - Show confidence range
 - Show int
 - Show one point per chemical set
- Inform
- Miscel
 - Hides members of chemical sets (tautomers, mixture components, etc.) and shows the set as a whole.
 - Press F1 for more help.

All members of tautomeric sets are displayed on the graph.

Data Gap Filling

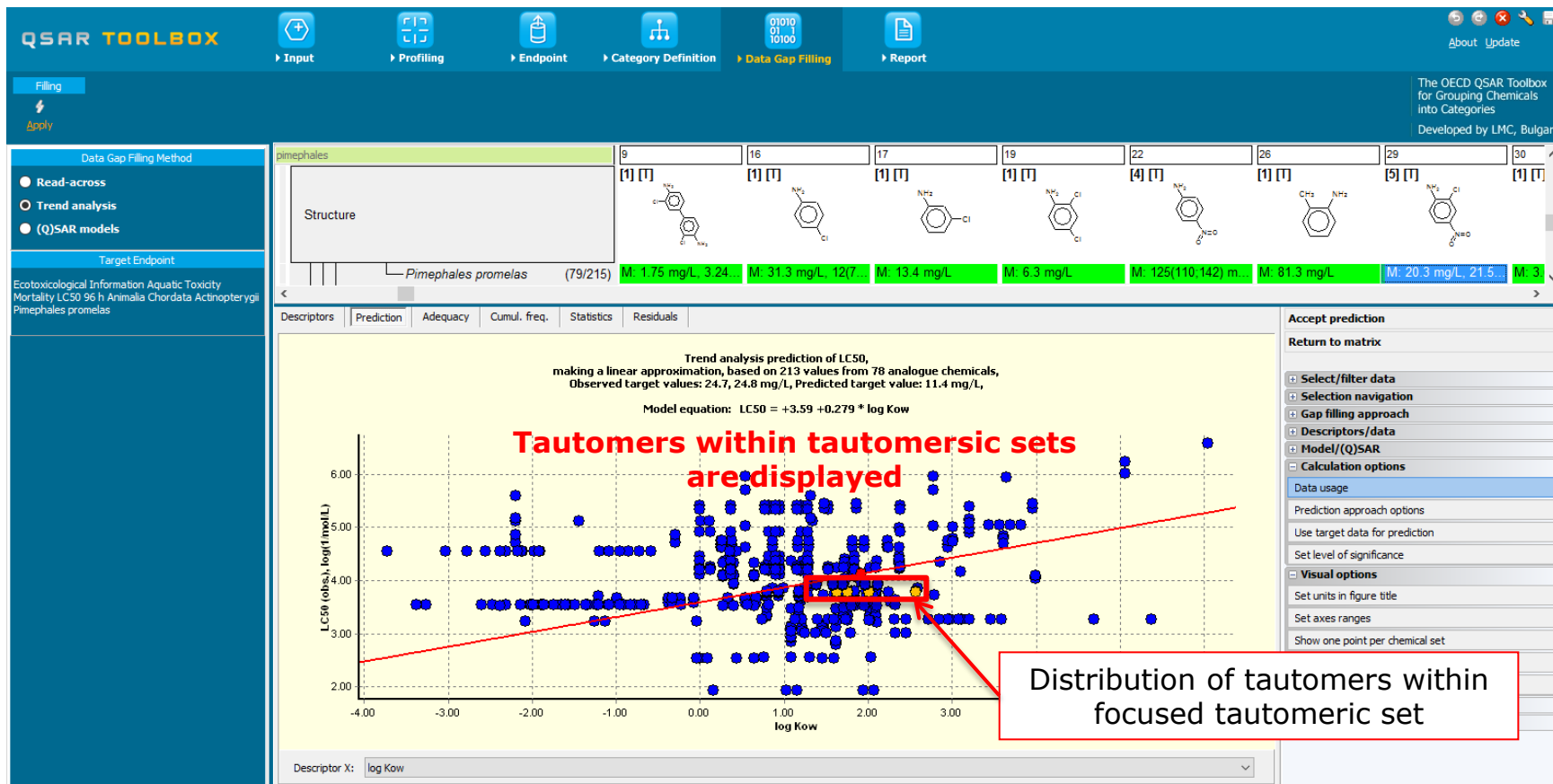
Results of Trend analysis

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' section is active, showing a list of chemicals with their structures and predicted values. Below this is a scatter plot titled 'Trend analysis prediction of LC50, making a linear approximation, based on 213 values from 78 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 11.4 mg/L, Model equation: $LC50 = +3.59 + 0.279 * \log Kow$ '. The plot shows log Kow on the x-axis and LC50 (obs.) log(mg/L) on the y-axis. A red regression line is shown. To the right, the 'Accept prediction' dialog box is open, with 'Calculation options' selected. Four numbered callouts point to specific settings: 1. 'Calculation options' (highlighted with a red box), 2. 'Data usage' (highlighted with a red circle), 3. 'All' (selected radio button), and 4. 'OK' button.

All observed data for chemicals in tautomeric sets could be used in trend analysis when 1. **Open** Calculation options, then 2. **Select** Data usage and 3. **Select** All. Finally 4. **Click** OK

Data Gap Filling

Results of Trend analysis



All members of tautomeric sets are displayed on the graph and all experimental data is taken into account.

Data Gap Filling

Role of Subcategorisation

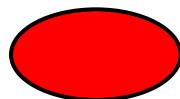
- Remember in the Toolbox, a category retrieved for tautomeric set refers to a group of chemicals with its tautomeric forms which have same profiling result according to one of the profilers listed in the module "Profiling".
- Subcategorisation refers to the process of applying additional profilers to the previously defined category; subcategorization procedure can be applied on:
 - **Single chemical** – eliminate chemicals having different categories than those of the target
 - **Set of tautomers** – eliminate tautomeric sets as a whole having different categories than those of the target tautomeric set
 - **Tautomers within tautomeric set** – eliminate specific tautomers within tautomeric set, which have categories different than those of the target.
- Illustration of subcategorization procedure is given on next three slides. The aim of next three slides is to provide more detailed information on different subcategorization procedures (it has illustrative character only and is not related to our test case)

Data Gap Filling

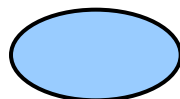
Role of Subcategorisation

Elimination of single chemical

Features of the target: S_T

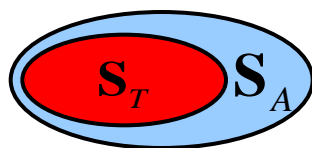


Features of analogue: S_A



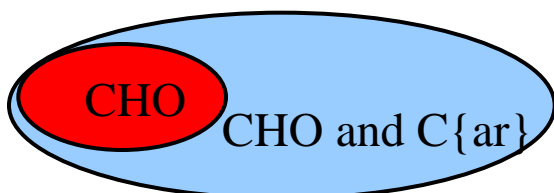
Refinement of category (Sub-categorization)

Analogue has different features of the target:

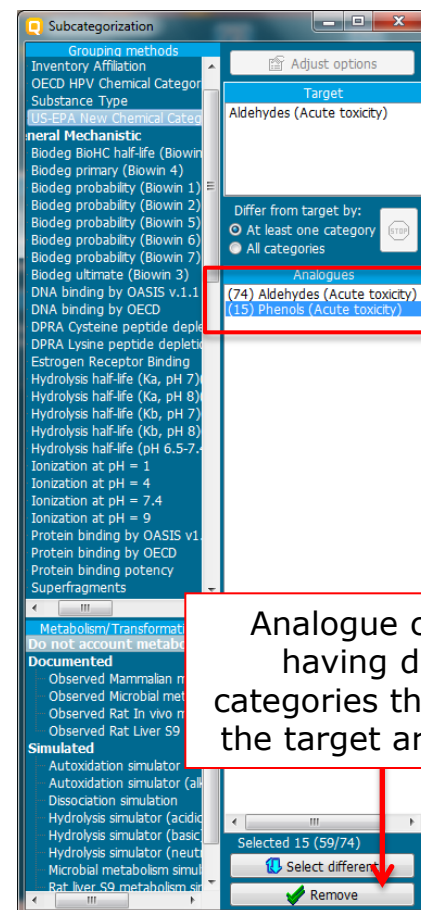


$$S_A \neq S_T$$

Example



Graphical Illustration



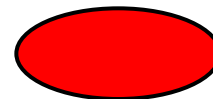
Analogue chemicals having different categories than those of the target are removed

Data Gap Filling

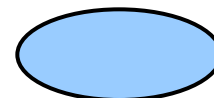
Role of Subcategorisation

Elimination of whole set of tautomers

Features of the target tautomeric set: S_T

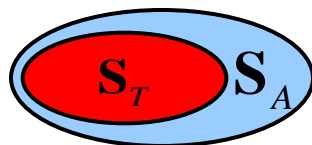


Features of analogues (as tautomeric sets): S_A



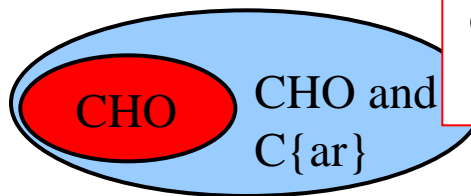
Refinement of category (Sub-categorization)

Analogue represented as tautomeric set has different features of the target:

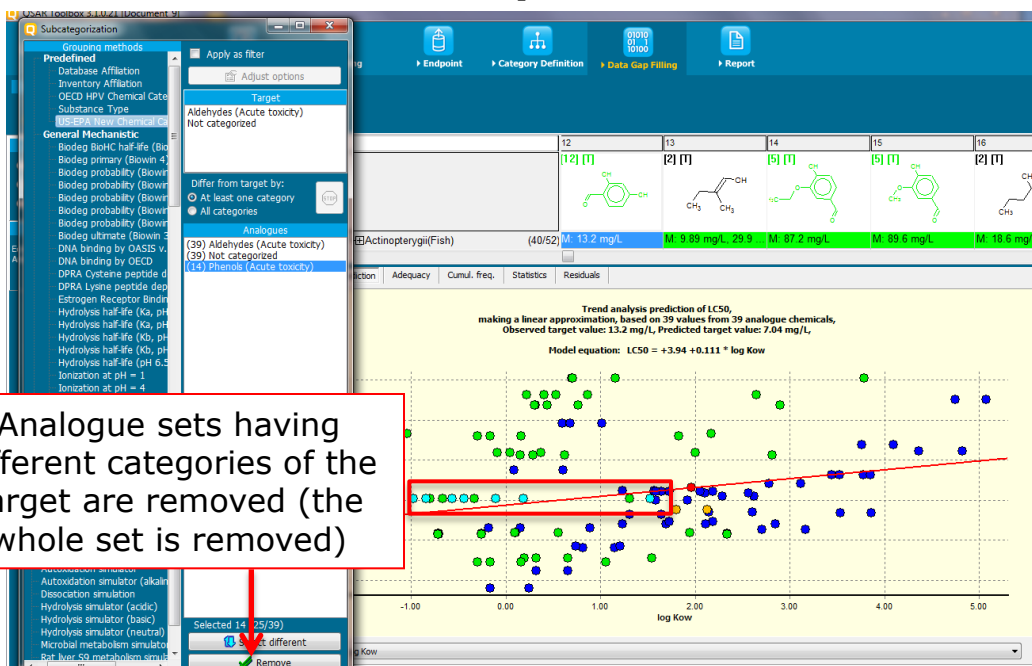


$$S_A \neq S_T$$

Example



Graphical Illustration



Analogue sets having different categories of the target are removed (the whole set is removed)

Data Gap Filling

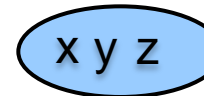
Role of Subcategorisation

Eliminating of single tautomers from from a tautomeric sets
(Apply filter functionality)

Features of the target tautomeric set: S_T



Features of analogues (as tautomeric sets): S_A

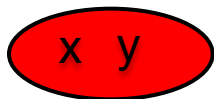


Refinement of category (Sub-categorization)

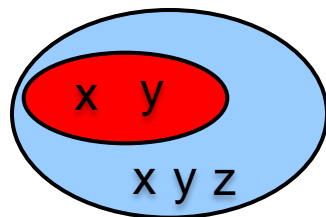
Graphical Illustration

Tautomers within tautomeric sets have different features of the target:

Tautomeric set of target



Tautomeric set of analogue



$$S_A \neq S_T$$

Target: Aldehydes (Acute toxicity), Not categorized

Differ from target by: At least one category, All categories

Analogues: (92) Aldehydes (Acute toxicity), (42) Not categorized, (45) Phenols (Acute toxicity)

Tautomer from tautomeric set having different categories is removed

This tautomeric form is categorized as phenol and is removed from tautomeric set

Data Gap Filling

Subcategorisation

Back to our case study

- In our example, the following subcategorizations are applied in order to eliminate dissimilar tautomeric sets:

- Substance type

The categorisation based on substance type allows keeping among the analogues only those that are of the same chemical type: discrete chemicals, mixtures, polymers, inorganics, organometalics. The current target represented as tautomeric set include discrete chemicals only. Hence the analogues (tautomeric sets) should also be discrete chemicals.

- Aquatic toxicity classification by ECOSAR

The categorization based on mode of action identifies analogues (in this case tautomeric sets) having the same mode of action as the target (i.e phenols and anilines). The analogues (tautomeric sets) having different categories should be eliminated.

Cont'd on next slide

Data Gap Filling

Subcategorisation

- In this example, the following subcategorizations is applied in order to eliminate dissimilar tautomeric sets

- Chemical elements

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

Subcategorisation steps are demonstrated on the next 4 screen shots.

Data Gap Filling Subcategorisation by Substance type

The screenshot displays the OECD QSAR Toolbox interface during a Data Gap Filling session. The central plot shows a trend analysis prediction of LC50 based on 213 values from 78 analogue chemicals. The model equation is $LC50 = +3.59 + 0.279 * \log Kow$. The plot includes a red regression line and data points colored by predicted values. The right sidebar shows the 'Accept prediction' panel with 'Subcategorize' selected. The left sidebar shows the 'Predefined' grouping methods with 'Substance Type' selected. The bottom left shows the 'Do not account metabolism' section with 'Remove' selected.

1. **Click** Subcategorize 2. **Select** Substance type 3. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

Data Gap Filling

Result of Subcategorisation by Substance type

QSAR TOOLBOX

Input Profiling Endpoint Category Definition **Data Gap Filling** Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models**

Target Endpoint

Ecotoxicological Information Aquatic Toxicity
Mortality LC50 96 h Animalia Chordata Actinopterygii
Pimephales promelas

Structure

pimephales

1 [target]	2	3	4	5	6	9	16
[5] [T]	[1] [T]	[1] [T]	[1] [T]	[9] [T]	[1] [T]	[1] [T]	[1]
Pimephales promelas (68/195) M: 24.7 mg/L, 24.8...	M: 75.5(68.4-83.4)	M: 5.68(5.34-6.04)	M: 8.06(7.26-8.95)	M: 1.44E3 mg/L, 1...	M: 158 mg/L, 171(...	M: 1.75 mg/L, 3.24...	M...

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50,
making a linear approximation, based on 193 values from 67 analogue chemicals,
Observed target values: 24.7, 24.8 mg/L, Predicted target value: 13.5 mg/L,
Model equation: $LC50 = +3.20 + 0.445 * \log Kow$

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
 - Subcategorize
 - Mark chemicals by WS
 - Mark chemicals by descriptor value
 - Mark outlier points
 - Filter points by test conditions
 - Mark focused chemical
 - Mark focused points
- Selection navigation
 - Gap filling approach
 - Descriptors/data
 - Model/(Q)SAR
 - Calculation options
 - Data usage
 - Prediction approach options
 - Use target data for prediction
 - Set level of significance
- Visual options
- Information
- Miscellaneous

Data Gap Filling

Subcategorisation by Aquatic toxicity classification by ECOSAR

The screenshot displays the QSAR Toolbox software interface during a Data Gap Filling process. The main window shows a 'Target' selection area with 'Anilines (Unhindered)' chosen. Below this, a 'Trend analysis prediction of LC50' plot is visible, showing a scatter plot of log Kow values versus LC50 values with a red regression line. The plot includes the equation: $LC50 = +3.20 + 0.445 * \log Kow$. The plot also shows 'Observed target values: 24.7, 24.8 mg/L' and 'Predicted target value: 13.5 mg/L'. The right sidebar contains a 'Select/filter data' menu with 'Subcategorize' highlighted. The left sidebar shows a list of 'Grouping methods' with 'Aquatic toxicity classification' selected. A 'Selected' list at the bottom shows 35 items, with a 'Remove' button highlighted.

1. **Click** Subcategorize
2. **Select** Aquatic toxicity classification by ECOSAR
3. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

Data Gap Filling

Result of Subcategorisation by Aquatic toxicity classification by ECOSAR

The screenshot displays the QSAR Toolbox software interface during a Data Gap Filling operation. The main window is titled "Data Gap Filling" and shows a scatter plot of log Kow (x-axis, 1.50 to 6.00) versus LC50 (y-axis, 1.50 to 6.00). The plot contains numerous blue data points and a red linear regression line. The regression equation is displayed as $LC50 = +2.29 + 0.784 * \log Kow$. Above the plot, a table lists 32 chemical analogues, including "Anilines (Unhindered)", with their respective log Kow and LC50 values. The interface also features a sidebar on the left with various grouping methods and a right-hand panel with options for "Accept prediction", "Return to matrix", and "Select/filter data".

Data Gap Filling

Subcategorisation by Chemical elements

The screenshot displays the QSAR Toolbox interface during the 'Data Gap Filling' process. On the left, the 'Grouping methods' sidebar is visible, with a callout '1' pointing to the 'Chemical elements' section. The main workspace shows the 'Target' selection panel with 'Group 14 - Carbon C', 'Group 15 - Nitrogen N', and 'Group 16 - Oxygen O' selected. Below this, a table lists chemical groups with their respective counts and molecular weights. A scatter plot titled 'Trend analysis prediction of LC50' shows the relationship between log Kow and LC50, with a red regression line and the equation $LC50 = +2.29 + 0.784 * \log Kow$. A callout '2' points to the 'Remove' button at the bottom of the interface.

1. **Select** Chemical elements
2. **Click** Remove to eliminate dissimilar tautomeric sets (the whole set is removed)

Data Gap Filling

Result of Subcategorisation by Chemical elements

Grouping methods

- Superfragments
- Toxic hazard classification
- Toxic hazard classification
- Ultimate biodeg

Endpoint Specific

- Acute aquatic toxicity clas
- Acute aquatic toxicity MO
- Aquatic toxicity classificati
- Bioaccumulation - metabo
- Bioaccumulation - metabo
- Biodegradation fragments
- Carcinogenicity (genotox.
- DART scheme v.1.0
- DNA alerts for AMES by O
- DNA alerts for CA and MN
- Eye irritation/corrosion Ex
- Eye irritation/corrosion Ind
- in vitro mutagenicity (Ames)
- in vivo mutagenicity (Micro
- Keratinocyte gene expres
- Oncologic Primary Classific
- Protein binding alerts for
- Protein binding alerts for
- Respiratory sensitisation
- Retinoic Acid Receptor Bir
- rtER Expert System ver.1
- Skin irritation/corrosion Ex
- Skin irritation/corrosion In

Empiric

- Chemical elements

Apply as filter

Adjust options

Target

- Group 14 - Carbon C
- Group 15 - Nitrogen N
- Group 16 - Oxygen O

Differ from target by:

At least one category

All categories

Correlation

Analogues

- (13) Group 14 - Carbon C
- (13) Group 15 - Nitrogen N
- (4) Group 16 - Oxygen O

Selected 0 (13/13)

Select different

Remove

Endpoint Category Definition Data Gap Filling Report

About Update

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

1 [target]	2	6	22	35	36	69
[5] [1]	[1] [1]	[1] [1]	[4] [1]	[1] [1]	[1] [1]	[1] [1]

Chemical structures: Nc1ccc(O)cc1, Nc1ccccc1, Nc1ccc(O)cc1, Nc1ccc(O)cc1, CCCCCCCCc1ccc(O)cc1, CCCCc1ccc(O)cc1, Nc1ccc(O)cc1

Chemicals: *haloes promelas* (14/43) | M: 24.7 mg/L, 24.8... | M: 75.5(68.4, 83.4) ... | M: 158 mg/L, 171(... | M: 125(110, 142) m ... | M: 0.0614 mg/L, 0 ... | M: 2.99 mg/L, 2.8(... | M: >20 mg/L

quacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 41 values from 13 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 46.3 mg/L, Model equation: $LC50 = +2.04 + 0.771 * \log Kow$

Accept prediction

Return to matrix

Select/filter data

- Subcategorize
- Mark chemicals by WS
- Mark chemicals by descriptor value
- Mark outlier points
- Filter points by test conditions
- Mark focused chemical
- Mark focused points

Selection navigation

- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options

Data usage

- Prediction approach options
- Use target data for prediction
- Set level of significance

Visual options

Information

Miscellaneous

Data Gap Filling

Side-Bar of Subcategorisation

The last subcategorisation procedure eliminates unstable tautomeric forms from given tautomeric sets. This elimination is possible with respect to

- Tautomers unstable profiler

The categorisation based on Tautomers unstable allows keeping among the set of analogues only those thautomeric forms that are stable. For toxic effects conditioned by less specific interactions (such as mortality, growth inhibition, immobilization, etc.) the stable tautomeric forms appear to be the dominant toxicants. The tautomeric sets of target chemical and analogues include stable and unstable tautomeric forms. Based on the above recommendation the set of analogues and the target should contain only stable tautomeric forms. In this respect filtering the tautomeric sets should be applied. ("Apply filter option" should be selected)

See next two slides

Data Gap Filling

Subcategorization by Tautomers unstable

1 Select Tautomers unstable profiler

2 Check Apply as filter

Number of tautomers in a category bin

Category	LC50 (mg/L)
(1) [target]	4.7 mg/L, 24.8
(2) []	M: 75.5(68.4;83.4)
(6) []	M: 158 mg/L, 171(
(22) []	M: 125(110;142) m
(35) []	M: 0.0614 mg/L, 0
(36) []	M: 2.99 mg/L, 2.8(
(59) []	M: >20 mg/L

Trend analysis prediction of LC50, making a linear approximation, based on 41 values from 13 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 46.3 mg/L, Model equation: $LC50 = +2.04 + 0.771 * \log Kow$

1. Select Tautomers unstable profiler

In this case both tautomeric sets of target and analogues have same unstable tautomeric forms. The user should manually select unstable tautomeric forms in order to remove them, because from the system's point of view all labels are equal and the system cannot prefer the label "stable" to other (unstable ones). (see next slide)

2. Check Apply as filter

Data Gap Filling

Subcategorization by Tautomers unstable

The screenshot illustrates the 'Data Gap Filling' process in the QSAR Toolbox. On the left, the 'Grouping methods' sidebar is visible, with 'Tautomers unstable' selected under 'Empiric'. The main interface shows the 'Data Gap Filling' step, where a target chemical (Cyclohexadinenines form - 1,5-H shift) is compared against its analogues. A trend analysis plot shows the prediction of LC50 based on log Kow values, with a model equation of $LC50 = +2.04 + 0.771 * \log Kow$. The plot includes observed target values (24.7, 24.8 mg/L) and a predicted target value (46.3 mg/L). A 'Selected 3 (10/13)' list is shown at the bottom left, and a 'Remove' button is highlighted with a blue callout '4'. A blue callout '3' points to the 'Apply as filter' checkbox in the 'Target' section.

3. **Hold** Ctrl button and select unstable forms from the target tautomeric set, then the system automatically will select unstable tautomeric forms from analogues sets
4. **Click** Remove to eliminate dissimilar tautomers from tautomeric sets

Data Gap Filling

Result of Subcategorisation by Tautomers unstable

The screenshot shows the QSAR Toolbox interface during a Data Gap Filling operation. The 'Target' is 'Stable form'. The 'Differ from target by' section is set to 'All categories'. The 'Analogues' section shows 13 'Stable form' analogues. A trend analysis plot is displayed with the equation $LC50 = +1.97 + 0.785 * \log Kow$. A red circle highlights the 'Accept prediction' button in the right-hand panel.

1. Accept prediction

Data Gap Filling Result

The screenshot displays the 'Data Gap Filling' module of the QSAR Toolbox. The main window shows a trend analysis prediction for LC50 based on 13 analogues. The predicted target value is 15.6 mg/L, which is circled in red. A red box highlights the text 'Predicted value: 15,6 mg/l'.

Trend analysis prediction of LC50,
 making a linear approximation, based on 41 values from 13 analogue chemicals,
 Observed target values: 24.7, 24.8 mg/L, Predicted target value: 15.6 mg/L,
 Model equation: $LC50 = +1.97 + 0.785 * \log Kow$

The graph shows a positive linear correlation between log Kow and LC50. The x-axis is labeled 'log Kow' and ranges from 1.50 to 6.00. The y-axis represents LC50 values. A red regression line is shown with several data points (blue dots) and one outlier (red dot).

Table of Analogues:

Chemical Name	LC50 (mg/L)
Chlores promelas	24.7
...	...
...	24.8
...	158
...	125
...	0.0614
...	2.99
...	>20

Right Panel (Accept prediction):

- Return to matrix
- Select/filter data
 - Subcategorize
 - Mark chemicals by WS
 - Mark chemicals by descriptor value
 - Mark outlier points
 - Filter points by test conditions
 - Mark focused chemical
 - Mark focused points
- Selection navigation
 - Gap filling approach
 - Descriptors/data
 - Model/(Q)SAR
- Calculation options
 - Data usage
 - Prediction approach options
 - Use target data for prediction
 - Set level of significance
- Visual options
- Information
- Miscellaneous

Data Gap Filling

Cumulated frequency

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models**

Target Endpoint

Ecotoxicological Information Aquatic Toxicity
Mortality LC50 96 h Animalia Chordata Actinopterygii
Pimephales promelas

Structure

1 [target] 2 6 22 35 36 59 69

(1/5) (1) (1) (1) (1/4) (1) (1) (1) (1)

Pimephales promelas (14/43) M: 24.7 mg/L, 24.8... M: 75.5(68.4, 83.4) M: 158 mg/L, 171... M: 125(110, 142) m... M: 0.0614 mg/L, 0... M: 2.99 mg/L, 2.8... M: >20 mg/L

Descriptors Prediction Adequacy Cumul. fr Residuals

1

95% of Residuals $\leq 0.446, \log(1/\text{mol/L})$

Accept prediction

Return to matrix

- Select/filter data
 - Subcategorize
 - Mark chemicals by WS
 - Mark chemicals by descriptor value
 - Mark outlier points
 - Filter points by test conditions
 - Mark focused chemical
 - Mark focused points
- Selection navigation
 - Gap filling approach
 - Descriptors/data
 - Model/(Q)SAR
 - Calculation options
 - Data usage
 - Prediction approach options
 - Use target data for prediction
 - Set level of significance
- Visual options
- Information
- Miscellaneous

1. 95% of residuals are in the range of experimental error

Data Gap Filling Statistics

The screenshot shows the QSAR Toolbox interface for Data Gap Filling. The main window displays a list of chemicals, including 'Pimephales promelas', with their structures and predicted values. A callout box labeled '1' points to the 'Statistics' tab, which displays a table of statistical characteristics for the TA model. The 'Coefficient of determination, (R2)' is highlighted with a red circle and a value of 0.956.

Statistical characteristics	TA model
Number of data points, (N)	41
Coefficient of determination, (R2)	0.956
Adjusted coefficient of determination, (R2adj)	0.933
Coefficient of determination - leave one out, (Q2)	0.951
Coefficient of correlation for external set, (r2)	-
Sum of squared residuals, (SSR)	2.51
Standard deviation of residuals, (sN)	-
Sample standard deviation of residuals, (s)	0.253
Fisher function, (F)	842
Fisher threshold for statistical significance, (Fa)	5.75
b0	
- model descriptor	Intercept
- coeff. value	1.97
- coeff. range	± 0.15
- significance	Yes
- max. covariation	0.166 (vs b1)
b1	
- model descriptor	log Kow
- coeff. value	0.785
- coeff. range	± 0.053
- significance	Yes
- max. covariation	0.166 (vs b0)

1. Coefficient of determination is high

Data Gap Filling

Summary on implementation of tautomers in trend analysis

- For toxic effects conditioned by less specific interactions (such as mortality, growth inhibition, immobilization, etc.) the stable tautomeric forms appear to be the dominant toxicants.
- **Recommendation:** to use the most stable tautomers for representation of the chemicals

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Category definition
 - Data Gap Filling
- **Report**

Report

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Filling Apply

Data Gap Filling Method

- Read-across
- Trend analysis
- (Q)SAR models

Target Endpoint

Ecotoxicological Information Aquatic Toxicity
Mortality LC50 96 h Animalia Chordata Actinopterygii
Pimephales promelas

Structure	1 [target]	2	3	4	5	6	7	8
	[5] [1]	[1] [1]	[1] [1]	[1] [1]	[9] [1]	[1] [1]	[2] [1]	[5]
3 h	(1/2)							
6 h	(1/1)							
12 h	(1/1)							
24 h	(12/19)	M: 100;180 mg/L, ...			M: 9(7.5;11) mg/L...			
48 h	(11/15)	M: 65 mg/L, >135;...			M: 7(3.6;8.8) mg/L...			
72 h	(3/3)	M: ~135 mg/L						
96 h								
Animalia								
Chordata (Vertebrates)								
Actinopterygii (Fish)								
Pimephales promelas	(79/2) (6)	M: 24.7 mg/L, 24.8... T: 15.6(4.9;49.7) m...	M: 75.5(68.4;83.4) ...	M: 5.68(5.34;6.04) ...	M: 8.06(7.26;8.95) ...	M: 1.44E3 mg/L, 1...	M: 158 mg/L, 171(...	
5 Days	(1/1)				M: 5.91 mg/L			
7 Days	(2/2)				M: 0.35 mg/L			
21 Days	(1/1)				M: 0.41(0.33;0.7) ...			
LC84	(1/4)				M: 0.62 mg/L, 17 ...			
LOEC	(3/7)				M: 0.034 mg/L, 0.0...			
LOEL	(1/2)				M: 0.0016 mg/L, 0...			
LT50	(2/14)							
MATC	(4/9)				M: 0.0014 mg/L, 0...			
NOEC	(4/10)				M: 0.02 mg/L, 0.02...			
NOEL	(1/3)				M: 0.0011 mg/L, 0...			
NR-LETH	(5/6)				M: 0.157 mg/L			
NR-ZERO	(3/3)							
Population	(1/4)				M: 0.026 mg/L, 0.0...			

1

2

Report

1. Click on the cell with prediction
2. Perform **Right click** and **Select Report**

Report

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below this is a toolbar with actions such as Create, Print, Close, Save as, Register, Unregister, Update, Clone, and Design. On the left, there are panels for 'Available data to report' (Predictions, QSARs, Categories) and 'Available report templates' (Standard predefined, Custom user defined). The main window shows a report titled 'Prediction of LC50 for benzenamine, 4-methyl-2-nitro-' with page number '3 / 47'. The report content includes a section header 'QSAR Toolbox prediction based on trend analysis', a sub-header 'Prediction of LC50 for benzenamine, 4-methyl-2-nitro-', and a section titled 'Summary' which is circled in red. A blue callout box with the number '1' points to this 'Summary' section.

1

Prediction of LC50 for benzenamine, 4-methyl-2-nitro- 3 / 47

QSAR Toolbox prediction based on trend analysis

Prediction of LC50 for benzenamine, 4-methyl-2-nitro-

Summary

Toxicity of the target chemical (15.6 mg/L) is predicted from category members using trend analysis based on 41 values within the range 0.0614 - 171 mg/L from 13 category members. Category members are set of tautomers and are selected based on the profile of the target chemical and/or its tautomers. Only chemicals having experimental data are listed in the category.

The target set of tautomers FALLS within applicability domain of the prediction (see Section 4.3 for details).

The descriptor values for the target chemical and the category members in case they are set of tautomers, set of metabolites or mixtures are calculated using the following rule(s):

1. "log Kow" - taking the weighted average value

The data used for calculating the current prediction is taken from 41 experimental values selected from the following database(s):

1. Aquatic OASIS
2. ECOTOX

Below is a summary table for endpoint & descriptor values for the target set of tautomers and the first 10

1. Summary information for tautomer prediction

Report

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Reports Repository

Create Print Close Save as Register Unregister Update Clone Design

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Available data to report

- Predictions
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
 - QSAR Model Reporting Format (QMRF v. 3.4)
 - QSAR Toolbox Category Report (CCRF v. 3.3)
 - QSAR Toolbox Prediction Report (TPRF v. 3.3)
- Custom (user defined)
 - Editable copy of QSAR Model Reporting Form
 - Editable copy of QSAR Toolbox Category Report
 - Editable copy of QSAR Toolbox Prediction Report

Prediction [1]

d. Model version:
Not applicable

e. Reference to QMRF:
Not applicable

f. Input for prediction (target chemical):
SMILES

g. Descriptor and endpoint values for target chemical (if applicable):

Descriptor(s)	
log Kow	2.57
Endpoint (dep. variable)	
Ecotoxicological Information;	24.7 mg/L
Aquatic Toxicity	24.8 mg/L

h. Additional chemicals eliminations (not determined by domain):
Not available

i. Additional data eliminations (not determined by domain):
Not available

j. Predicted value (model result):
15.6 mg/L

k. Predicted value (comments): *manually editable field*
Not provided by the user

4.3. Applicability domain (OECD Principle 3):
The target set of tautomers FALLS within applicability domain (see Section 3.1.b for detailed description of the domain)

1

j. Predicted value (model result):
15.6 mg/L

1. Predicted value

Report

Prediction of LC50 for benzenamine, 1-methyl-2-nitro 9 / 47

1

Tautomeric filter applied to one of the target tautomers

4.4. Uncertainty of the prediction (OECD Principle 4): *manually editable field*

Adequacy of prediction

Adequacy of prediction
Model statistic: R2 = 0.956, R2adj = 0.955, s = 0.233

1. Applicability domain

Report

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Reports Repository

Create Print Close Save as Register Unregister Update Clone Design

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Available data to report

- Predictions
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
 - QSAR Model Reporting Format (QMRF v.3.4)
 - QSAR Toolbox Category Report (CCRF v.3.0)
 - QSAR Toolbox Prediction Report (TPRF v.3.0)
- Custom (user defined)
 - Editable copy of QSAR Model Reporting Form
 - Editable copy of QSAR Toolbox Category Report
 - Editable copy of QSAR Toolbox Prediction Report

Prediction [1]

QSAR Toolbox prediction based on trend analysis

Prediction of LC50 for benzenamine, 4-methyl-2-nitro-

APPENDIX 7 - Chemical components

Tautomer No.1 of target chemical:

1

used in prediction

- CAS number:**
89-62-3
- Other regulatory numbers:**
Not reported
- Chemical name(s):**
Not available
- Structural formula:**

CH₃

1. Additional information indicates which tautomer is used in trend analysis

Outlook

- Background
- Objectives
- The exercise
- Workflow
- **Save the 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 result

The screenshot shows the QSAR Toolbox software interface. A 'Save As' dialog box is open, allowing the user to save a prediction result. The dialog shows the file name 'Totorial 14.tbw' and the save type 'Toolbox work file (*.tbw)'. The background report template includes sections for 'trend analysis', 'methyl-2-nitro-', 'ponents', and 'used in prediction'. A chemical structure with a CH₃ group is visible at the bottom.

1. **Click** on Save button; 2. **Define** name of the file; 3. **Click** Save button

Open saved file

The screenshot displays the QSAR Toolbox interface. A 'Select file' dialog box is open, showing a file list with 'Tutorial 14.tbw' selected. The dialog box has a 'File name' field containing 'Tutorial 14.tbw' and an 'Open' button. The background shows a chemical structure and a table of results.

Chemical Structure	Category	Value
<chem>Cc1ccc(N)cc1</chem>	LC84	(1/4)
	LOEC	(3/7)
	LOEL	(1/2)
	LT50	(2/14)
	MATC	(4/9)
	NOEC	(4/10)
	NOEL	(1/3)
	NR-LETH	(5/6)
	NR-ZERO	(3/3)
	Population	(1/4)

Once the file has been saved **1. Go to Input; 2. Click Open; 3. Find and select file; 4. Click Open**

Open saved file

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below the menu, there are tabs for 'Document', 'Single Chemical', and 'Chemical List'. The main workspace is divided into several sections: a left sidebar for 'Documents', a central area for 'Filter endpoint tree...', and a right area for a 'Chemical List' table. A 'Read from DB...' button is visible in the center. At the bottom, a status bar shows 'Loading file "Totorial 14.tbw" ...' with a red callout bubble containing the number '1'.

During loading a file and reproducing steps of the prediction an indication message appears (1)

Open saved file

The screenshot displays the QSAR Toolbox software interface. The main window shows a table with 6 columns representing different chemical entities. The first column is labeled '1 [target]' and contains a chemical structure of 2-nitroaniline. The other columns contain various chemical structures and their corresponding physical and chemical properties.

1 [target]	2	3	4	5	6
[5] [1]	[1] [1]	[1] [1]	[1] [1]	[9] [1]	[1] [1]
M: 24.7 mg/L, 24.8... T: 15.6(4.9;49.7) m...	M: 0.004 mg/L, 0.0...	M: 0.032 mg/L, 0.1...	M: 0.001 mg/L, 0.0...	M: 845 mg/L, 1.44...	M: 1.29

An 'Information' dialog box is overlaid on the table, displaying the message: 'The file was executed successfully'. A red circle with the number '1' points to the 'OK' button in the dialog box.

When the process finish a message appears. Click OK to close it (1)