

OECD QSAR Toolbox v.3.1

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

Outlook

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

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

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

Outlook

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

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**

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

Implementation of Modeling modes

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

QSAR Toolbox 3.1.0.21 [Document_1]

Document | Single Chemical | Chemical List | **Component Mode**

New Open Close Save CAS# Name Structure Select Delete Query ChemIDs DB Inventory List

Filter endpoint tree... 1 [target] [5] [1]

Structure

Substance Identity
Physical Chemical Properties

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

Document | Single Chemical | Chemical List | **Component Mode**

New Open Close Save CAS# Name Structure Select Delete Query ChemIDs DB Inventory List

Filter endpoint tree... 1 [target] 2 [target,tautomer] 3 [target,tautomer] 4 [target,tautomer] 5 [target,tautomer] 6 [target,tautomer]

Structure

Substance Identity
Physical Chemical Properties
Environmental Estimation and Transport

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 3.1.0.21 interface. The top menu bar includes 'Profiling', 'Input', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing options like 'Apply', 'New', 'View', and 'Delete'. A red circle highlights the 'View' option, with a callout box containing the number '2'. In the left sidebar, under 'Profiling methods', the 'US-EPA New Chemical Categories' is selected, with a callout box containing the number '1'. The main window shows the 'US-EPA New Chemical Categories (Predefined) - Profiling Scheme Browser'. It is divided into two panes: 'List with categories' and 'Textual description'. The 'List with categories' pane shows a scrollable list of chemical classes, with 'Acrylamides' selected. The 'Textual description' pane provides detailed information for the selected category, including its definition, hazard concerns, boundaries, and general testing strategy. Arrows point from the 'View' button to the 'Textual description' pane and from the 'US-EPA New Chemical Categories' in the sidebar to the 'List with categories' pane.

1. Highlight the profiler
2. Click View

Profiling Side-Bar to Profiling

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
 - US-EPA New chemical category
 - Aquatic toxicity classification by ECOSAR
 - Acute aquatic toxicity MOA by OASIS
 - Acute aquatic toxicity classification by Verhaar
 - 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 Profiling methods: Organic functional groups (checked)

2 Profile menu: Apply

3 Right-click context menu: Component Profile Statistic

4 Component Profile Statistic window: Acute aquatic toxicity MOA by OASIS categories of tautomers in the set

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

5 All chemicals: 2D representations of the structures in each category

Number of tautomers in a category bin

- 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; Aquatic US-EPA 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
 - ✓ Aquatic US-EPA ECOTOX

Endpoint Gather data

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Endpoint' menu is open, showing options like 'Gather', 'Import', 'Export', 'Delete', and 'Tautomerize'. The 'Gather' button is circled in red and labeled with a '2'. On the left, the 'Databases' list is visible, with several entries checked with green checkmarks, labeled with a '1'. The main panel displays chemical information for a target endpoint, including a structural formula and a list of chemical names.

- 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 3.1.0.21 interface. The 'Endpoint' workflow is active. In the 'Databases' panel, 'Ecotoxicological Information' is selected, with sub-options like 'Aquatic ECETOC' and 'Aquatic Japan MoE' checked. The 'Filter endpoint tree...' panel shows a tree structure with 'Mortality' expanded to 'LC50' and '96 h'. The 'Datamatrix' table on the right shows a row for 'Pimephales promelas' with a value of '(1/2) M: 24.7 mg/L, 24.8(20.6;29.9) mg/L'. A red circle highlights this row, and a callout bubble with the number '1' points to it. Another red circle highlights the row below it, which contains '(1/1) M: 67.2(45.1;113) mg/L'. The chemical structure of the target is shown as a benzene ring with an amino group (-NH₂) and a nitro group (-NO₂) at the para position, and a methyl group (-CH₃) at the ortho position.

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 shows the QSAR Toolbox interface. In the main window, the 'Filter endpoint tree' on the left has 'US-EPA New Chemical Categories' selected, indicated by a blue callout box with the number 1. The 'Define' button in the top toolbar is circled in red, with a blue callout box containing the number 2. A dialog box titled 'US-EPA New Chemical Categories' is open on the right. It shows a list of 'Target(s) profiles' with 'Anilines (Acute toxicity)' selected and 'Not categorized' highlighted in blue, with a blue callout box containing the number 3. Below the list, a blue arrow button is highlighted with a blue callout box containing the number 4. The dialog also shows options for 'Combine profiles logically' (AND/OR) and 'Invert result' (checkbox).

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 interface. The main window is titled 'QSAR Toolbox 3.1.0.21 [Document_1]'. The 'Category Definition' tab is active. The 'Filter endpoint tree...' shows a chemical structure of Aniline. A callout box labeled '1' points to the 'US-EPA New Chemical Categories' dialog box. This dialog box shows 'Anilines (Acute toxicity)' as the 'Target(s) profiles'. Below, a list of 'All profiles' includes various chemical classes. A callout box labeled '2' points to the 'OK' button in the dialog box.

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 3.1.0.21 interface. The 'Category Definition' step is active. The 'Filter endpoint tree...' window is open, showing a list of endpoints. The 'Ecotoxicological Information' endpoint is selected, and a chemical structure is shown. A warning dialog box is overlaid on the screen, asking 'Do you want to continue?' with 'Yes', 'No', and 'Cancel' buttons. A red speech bubble with the number '1' points to the 'Yes' button.

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 displays the QSAR Toolbox 3.1.0.21 interface. The main window is titled 'QSAR Toolbox 3.1.0.21 [Document]'. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' menu is expanded, showing options: 'Define', 'Subcategorize', 'Combine', 'Clustering', 'Delete', and 'Delete All'. The 'Define' button is highlighted with a blue callout box containing the number '1'. A 'Define category name' dialog box is open, showing the text 'icity) (US-EPA New Chemical Categories)' in the input field. The background shows the 'Filter endpoint tree...' window with a chemical structure and a list of endpoints including 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'.

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”.
- 314 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 appear in the “Defined Categories” window, along with the number of substances belonging to the category.

[314] Anilines (Acute toxicity) (US-EPA New Chemical Categories)

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: 748 data-points, 255 groups, 115 chemicals

Data points...

	Endpoint	CAS	Structure	Value	Age
<input checked="" type="checkbox"/>	NOEC	95-51-2	<chem>Nc1ccccc1</chem>	3.2 mg/L	
<input checked="" type="checkbox"/>	NOEC	95-51-2	<chem>Nc1ccccc1</chem>	3.2 mg/L	
<input checked="" type="checkbox"/>	NOEC	95-80-7	<chem>Nc1ccc(O)c(O)c1</chem>	1 mg/L	
<input checked="" type="checkbox"/>	NOEC	95-80-7	<chem>Nc1ccc(O)c(O)c1</chem>	1 mg/L	
<input checked="" type="checkbox"/>	NOEC	106-49-0	<chem>Nc1ccc(O)c(O)c1</chem>	3.1 mg/L	
<input checked="" type="checkbox"/>	NOEC	106-49-0	<chem>Nc1ccc(O)c(O)c1</chem>	3.1 mg/L	
<input checked="" type="checkbox"/>	NOEC	95-64-7	<chem>Nc1ccc(O)c(O)c1</chem>	2.9 mg/L	
<input checked="" type="checkbox"/>	NOEC	95-64-7	<chem>Nc1ccc(O)c(O)c1</chem>	2.9 mg/L	

1

2

Select one

Invert

Check All

Uncheck All

OK

Cancel

1. Click Select one and then
2. Click OK

Category Definition

Summary information for Analogues

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. The main window shows a table of chemical analogues with their structures and associated data. The 'Ecotoxicological Inform.' row is highlighted with a red box, indicating available aquatic experimental data for the analogues.

Filter endpoint tree...	1 [target]	2	3	4	5	6	7	8	9
Structure									
Ecotoxicological Inform.	(310/4784) M: 24.7 mg/L, 24.8(...	M: 0.0063 mg/L, 0.0...	M: 0.032 mg/L, 0.0...	M: 0.005 mg/L, 0.0...	M: 0.52 mg/L, 0.81 ...	M: 0.011 mg/L, 0.0...	M: 0.014 mg/L, 0.0...	M: 0.2 mg/L, 0.62 ...	M: 0.0...

Available aquatic experimental data for the analogues represented as tautomeric sets appears on datamatrix.

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 314 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)

Recap

- You have now retrieved the available experimental data on aquatic toxicity for 314 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 displays the QSAR Toolbox interface with the following elements:

- 3**: Callout pointing to the **Apply** button in the top left.
- 2**: Callout pointing to the **Trend analysis** option in the 'Data Gap Filling Method' dropdown.
- 1**: Callout pointing to the experimental data row for *Pimephales promelas* at 96h, where the LC50 value of 24.8 mg/L is highlighted.

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

Target	1 [target]	2	3	4	5	6	7	8
Structure	<chem>Cc1ccc(N)cc1[N+](=O)[O-]</chem>	<chem>Nc1ccccc1</chem>	<chem>Nc1ccc(Cl)cc1</chem>	<chem>Nc1ccc(Cl)cc1</chem>	<chem>Nc1ccc(N)cc1</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>Nc1ccc(N)cc1</chem>	<chem>Nc1ccc(N)cc1</chem>
1 h	(1/1)							
3 h	(1/1)							
6 h	(1/1)							
12 h	(1/1)							
24 h	(12/18)	M: 100,180 mg/L, >...			M: 9.03(8.55,9.53) ...			
48 h	(11/15)	M: 65 mg/L, >...						
72 h	(3/3)							
96 h	(73/194)	M: 24.7 mg/L, 24.8(24.7,24.9) ...	M: 107 mg/L, 75.5(75.5,107) ...	M: 5.7 mg/L, 5.81(5.7,5.9) ...	M: 7.58 mg/L, 6.99(6.99,7.58) ...	M: 1.44E3 mg/L, 1.44E3(1.44E3,1.44E3) ...	M: 158 mg/L, 171(158,171) ...	
7 Days	(1/1)		M: 60.2(53.4;67.9) ...					
21 Days	(1/1)				M: 0.41(0.33;0.7) m...			
LC84	(1/4)				M: 0.62 mg/L, 9 mg...			
LOEC	(2/4)		M: 23.2 mg/L, 2.36 ...		M: 0.034 mg/L, 0.0...			
LOEL	(1/2)				M: 0.0016 mg/L, 0...			
LT50	(2/14)							
MATC	(2/5)		M: 19 mg/L, <2.36 ...		M: 0.0014 mg/L, 0...			
NOEC	(2/4)		M: 15.7 mg/L, 2.1 ...		M: 0.02 mg/L, 0.02...			
NOEL	(1/3)				M: 0.0011 mg/L, 0...			
NR-LETH	(2/2)							

- 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 software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' tab is active, showing a list of chemical structures and their corresponding LC50 values. Below this, a trend analysis plot is shown, titled 'Trend analysis prediction of LC50, making a linear approximation, based on 72 values from 72 analogue chemicals, Observed target value: 24.7 mg/L, Predicted target value: 12.8 mg/L'. The plot shows a positive correlation between log Kow and LC50 (log1 mol/L). The model equation is $LC50 = +3.47 + 0.316 * \log Kow$. On the right side, the 'Accept prediction' panel is visible, with 'Visual options' selected. A red box highlights the 'Visual options' section, and a red circle highlights the 'Show all members of chemical sets' option. Two callout boxes, labeled '1' and '2', point to the 'Visual options' section and the 'Show all members of chemical sets' option, respectively.

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 3.1.0.21 [Document_1]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Filling Apply

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(Vertebrates) Actinopterygii(Fish) Pimephales promelas

Structure	1 [target]	2	3	4	5	6	9	16	17
Pimephal... (73/207)	M: 24.7 mg/L, 24.8(20.6...	M: 107 mg/L, 75.5(...	M: 5.7 mg/L, 5.81(5...	M: 7.58 mg/L, 6.99(...	M: 1.44E3 mg/L, 1...	M: 158 mg/L, 171(1...	M: 1.75 mg/L, 3.24...	M: 31.3 mg/L, 12(7...	M: ...

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 72 values from 72 analogue chemicals, Observed target value: 24.7 mg/L, Predicted target value: 12.8 mg/L,

Model equation: $LC50 = +3.47 + 0.316 * \log Kow$

LC50 (obs., log(1/mmol/L))

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 intercorrelations
- Information
- Miscellaneous

All members of tautomeric sets are displayed on the graph.

Data Gap Filling

Results of Trend analysis

The screenshot displays the QSAR Toolbox interface. At the top, the 'Data Gap Filling' tab is active. Below the navigation bar, a table lists various chemical structures and their corresponding LC50 values. The main window features a scatter plot titled 'Trend analysis prediction of LC50, making a linear approximation, based on 72 values from 72 analogue chemicals, Observed target value: 24.7 mg/L, Predicted target value: 12.8 mg/L, Model equation: $LC50 = +3.47 + 0.316 \cdot \log Kow$ '. The plot shows a positive correlation between log Kow and log(LC50). On the right side, the 'Accept prediction' panel is open, with several settings highlighted by numbered callouts: 1. 'Calculation options' is selected in the left sidebar; 2. 'Data usage' is selected under 'Calculation options'; 3. 'All' is selected under 'Set usage of chemical'; 4. The 'OK' button is highlighted.

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

QSAR Toolbox 3.1.0.21 [Document_1]

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 (Vertebrates) Actinopterygii (Fish) Pimephales promelas

Structure	1 [target]	2	3	4	5	6	9	16	17
Pimephal... (73/207)	M: 24.7 mg/L, 24.8(20.6)	M: 107 mg/L, 75.5	M: 5.7 mg/L, 5.81(5)	M: 7.58 mg/L, 6.99	M: 1.44E3 mg/L, 1	M: 158 mg/L, 171(1)	M: 1.75 mg/L, 3.24	M: 31.3 mg/L, 12(7)	M: 1

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50,
making a linear approximation, based on 205 values from 72 analogue chemicals,
Observed target values: 24.7, 24.8 mg/L, Predicted target value: 12.4 mg/L,
Model equation: $LC50 = +3.42 + 0.348 * \log Kow$

Tautomers within tautomeric sets are displayed

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
- 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

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

Data Gap Filling

Side-Bar of Subcategorisation

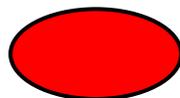
- Remember in the Toolbox, a category retrieved for tautomeric set refers to a group of chemicals with its tautomeric forms which have same profiling result according to one of the profilers listed in the module "Profiling".
- Subcategorisation refers to the process of applying additional profilers to 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.
- Graphical illustration of subcategorization procedure is given on next three slides

Data Gap Filling

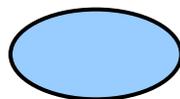
Side-Bar of Subcategorisation

Elimination of single chemical

Features of the target: S_T

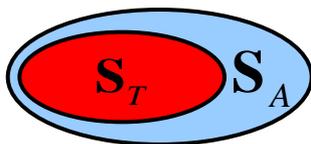


Features of analogue: S_A



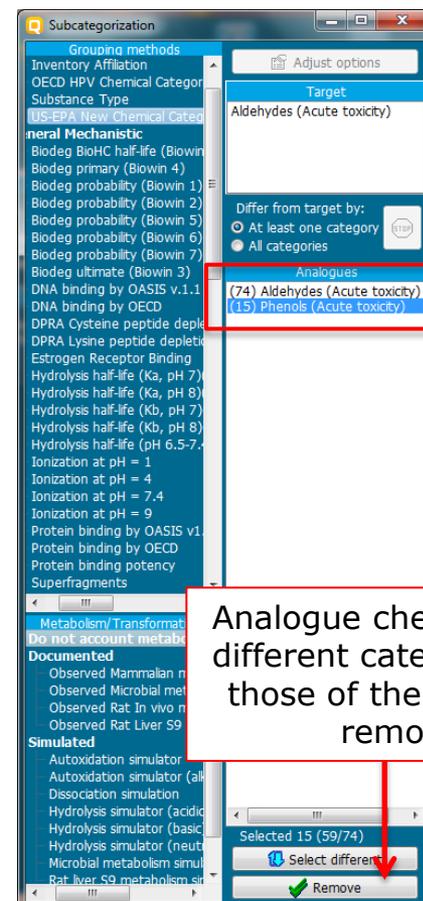
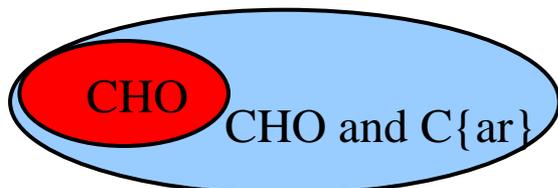
Refinement of category (Sub-categorization)

Analogue has different features of the target:



$$S_A \neq S_T$$

Example



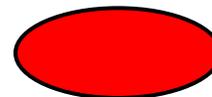
Analogue chemicals with different categories than those of the target are removed

Data Gap Filling

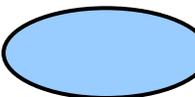
Side-Bar of Subcategorisation

Eliminating set of tautomers

Features of the target tautomeric set: S_T

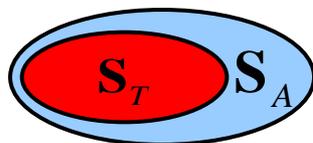


Features of analogues (as tautomeric sets): S_A



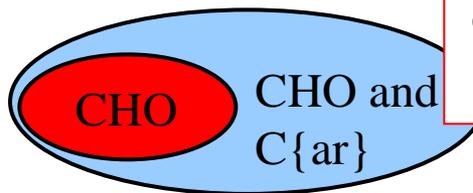
Refinement of category (Sub-categorization)

Analogue as tautomeric set has different features of the target:



$$S_A \neq S_T$$

Example



The screenshot shows the 'Subcategorization' window in QSAR Toolbox. The 'Target' is set to 'Aldehydes (Acute toxicity)'. The 'Differ from target by' options are 'At least one category' and 'All categories'. The 'Analogues' list includes '(39) Aldehydes (Acute toxicity)' and '(14) Phenols (Acute toxicity)'. The 'Trend analysis prediction of LC50' plot shows a scatter of data points with a red regression line. A red box highlights a specific data point on the plot.

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

Data Gap Filling

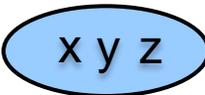
Side-Bar of Subcategorisation

Eliminating tautomers from a tautomeric sets

Features of the target tautomeric set: S_T



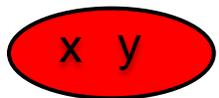
Features of analogues (as tautomeric sets): S_A



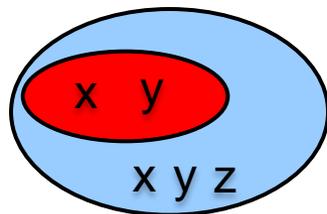
Refinement of category (Sub-categorization)

Tautomers within tautomeric sets have different features of the target:

Tautomeric set of target



Tautomeric set of analogue



$$S_A \neq S_T$$

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

Side-Bar of Subcategorisation

- In this 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

Side-Bar of 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 QSAR Toolbox software interface. On the left, the 'Subcategorization' sidebar is open, showing a list of predefined methods. Callout box 2 points to 'Substance Type' under the 'Predefined' section. The main window shows a data table with columns for target values and chemical structures. Below the table is a scatter plot titled 'Trend analysis prediction of LC50, making a linear approximation, based on 205 values from 72 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 12.4 mg/L, Model equation: $LC50 = +3.42 + 0.348 * \log Kow$ '. Callout box 1 points to the 'Subcategorize' button in the 'Select/filter data' panel on the right. Callout box 3 points to the 'Remove' button at the bottom of the sidebar.

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

The screenshot displays the QSAR Toolbox software interface during a 'Data Gap Filling' process. The main workspace shows a table of chemical structures and their corresponding LC50 values. The table is as follows:

Structure	1 [target]	2	3	4	5	6	9	16	17
<chem>Cc1ccc(cc1)N(=O)=O</chem>	[5] [T]	[1] [T]	[1] [T]	[1] [T]	[9] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]
Pimephal... (73/207)	M. 24.7 mg/L, 24.8 (20.6...	M. 107 mg/L, 75.5(M. 5.7 mg/L, 5.81(5	M. 7.58 mg/L, 6.99(M. 1.44E3 mg/L, 1	M. 158 mg/L, 171(1	M. 1.75 mg/L, 3.24	M. 31.3 mg/L, 12(7	M. 1

Below the table, a trend analysis plot is shown with the following details:

- Trend analysis prediction of LC50, making a linear approximation, based on 193 values from 66 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 13.1 mg/L,**
- Model equation: $LC50 = +3.24 + 0.431 * \log Kow$**

The plot shows a positive linear correlation between log Kow and LC50. The x-axis is labeled 'log Kow' and ranges from -2.00 to 6.00. The y-axis is labeled 'LC50 (mg/L)' and ranges from 2.00 to 6.50. A red regression line is drawn through the data points.

The right-hand panel contains several sections for data management and prediction options:

- 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
 - Remove marked chemicals/points
 - Clear existing marks
- 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**

Data Gap Filling

Subcategorisation by Aquatic toxicity classification by ECOSAR

The screenshot displays the QSAR Toolbox software interface during a 'Data Gap Filling' operation. The main window shows a table of chemical structures and their predicted LC50 values. A trend analysis plot is visible, showing a linear approximation based on 193 values from 66 analogue chemicals. The model equation is $LC50 = +3.24 + 0.431 \cdot \log Kow$. The observed target values are 24.7, 24.8 mg/L, and the predicted target value is 13.1 mg/L. The interface includes a sidebar for 'Subcategorization' with various methods, a 'Trend analysis prediction of LC50' plot, and a 'Accept prediction' panel with a 'Subcategorize' button highlighted by a red circle.

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 'Data Gap Filling' module of the QSAR Toolbox. On the left, the 'Subcategorization' panel lists various grouping methods, including 'Aquatic toxicity classification by ECOSAR'. The main workspace shows a table of chemical structures and their predicted values. The table includes columns for chemical structures and their corresponding predicted values (e.g., M: 24.7 mg/L, 24.8(20.6...), M: 107 mg/L, 75.5..., M: 5.7 mg/L, 5.81(5...), M: 7.58 mg/L, 6.99..., M: 158 mg/L, 171(1...), M: 1.75 mg/L, 3.24..., M: 31.3 mg/L, 12(7...), M: 13.4 mg/L, M: 6...).

Below the table, a 'Trend analysis prediction of LC50' plot is shown. The plot displays a scatter of data points with a red regression line. The x-axis is labeled 'log Kow' and ranges from 1.00 to 6.00. The y-axis represents LC50. The plot includes the following text: 'Trend analysis prediction of LC50, making a linear approximation, based on 93 values from 31 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 23.7 mg/L, Model equation: LC50 = +2.31 + 0.778 * log Kow'.

On the right side of the interface, there is a 'Accept prediction' panel with various options for filtering and selecting data, including 'Subcategorize', 'Mark chemicals by WS', 'Mark chemicals by descriptor value', 'Mark outlier points', 'Filter points by test conditions', 'Mark focused chemical', 'Mark focused points', 'Remove marked chemicals/points', 'Clear existing marks', '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', and 'Visual options'.

Data Gap Filling

Subcategorisation by Chemical elements

The screenshot displays the QSAR Toolbox software interface. On the left, the 'Subcategorization' window is open, showing a list of chemical elements and a 'Remove' button. A red callout box with the number '1' points to the 'Chemical elements' section, and another red callout box with the number '2' points to the 'Remove' button. The main window shows a table of chemical structures and their predicted LC50 values. Below the table, a scatter plot titled 'Trend analysis prediction of LC50' shows the relationship between log Kow and LC50. The plot includes a red regression line and a text box stating: 'Trend analysis prediction of LC50, making a linear approximation, based on 93 values from 31 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 23.7 mg/L, Model equation: LC50 = +2.31 + 0.778 * log Kow'. On the right side of the interface, there is a 'Accept prediction' panel with various options for data selection and visualization.

- 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

The screenshot displays the 'Data Gap Filling' module of the QSAR Toolbox. The interface is divided into several sections:

- Left Sidebar:** Contains a list of grouping methods and endpoints. The 'Chemical elements' method is selected, with options for 'At least one chemical element' and 'All categories'.
- Top Panel:** Shows navigation buttons for Profiling, Endpoint, Category Definition, Data Gap Filling (active), and Report.
- Table:** A table listing chemical subcategories with their corresponding chemical structures and predicted values. The table has columns for subcategory ID, name, and predicted values. The first row is highlighted in green.
- Trend Analysis Plot:** A scatter plot showing the relationship between log Kow (x-axis, 1.00 to 6.00) and LC50 (y-axis). A red regression line is shown. The plot title is 'Trend analysis prediction of LC50, making a linear approximation, based on 38 values from 12 analogue chemicals, Observed target values: 24.7, 24.8 mg/L, Predicted target value: 46.9 mg/L, Model equation: LC50 = +2.03 +0.772 * log Kow'.
- Right Panel:** Contains control options for 'Accept prediction', 'Return to matrix', 'Select/filter data', 'Selection navigation', 'Gap filling approach', 'Descriptors/data', 'Model/(Q)SAR', 'Calculation options', and 'Visual options'.

Subcategory	Chemical Structure	Observed Target Values	Predicted Target Value
1 [target]	<chem>Nc1ccc(C)cc1</chem>	24.7 mg/L, 24.8 mg/L	46.9 mg/L
2	<chem>Nc1ccccc1</chem>	107 mg/L, 75.5 mg/L	...
6	<chem>Nc1ccc(C)cc1</chem>	158 mg/L, 171 mg/L	...
44	<chem>Nc1ccc(C)cc1</chem>	0.0614 mg/L, 0 mg/L	...
45	<chem>Nc1ccc(C)cc1</chem>	2.99 mg/L, 2.8 mg/L	...
60	<chem>Nc1ccc(C)cc1</chem>	>20 mg/L	...
67	<chem>Nc1ccc(C)cc1</chem>	36.3 mg/L	...
69	<chem>Nc1ccc(C)cc1</chem>	73 mg/L, 73.65 mg/L	...
74	<chem>Nc1ccc(C)cc1</chem>

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

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. On the left, the 'Subcategorization' window is open, showing a list of target and analogue categories. A red box labeled '1' highlights 'Structure similarity' and 'Tautomers unstable' under the 'Empiric' section. Another red box labeled '2' highlights the 'Apply as filter' checkbox. The main window shows a table of chemical categories with their respective LC50 values. A red box labeled 'Number of tautomers in a category bin' points to the '5' in the first column of the table. Below the table, a trend analysis plot shows the relationship between log Kow and LC50 (obs., log10(mg/L)). The plot includes a red regression line and several data points. The model equation is $LC50 = +2.03 + 0.768 * \log Kow$. The observed target value is 24.7 mg/L, and the predicted target value is 47.5 mg/L.

Category	LC50 (mg/L)
1 [target]	24.7
2	107
6	158
44	0.0614
45	2.99
60	>20
67	36.3

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 displays the 'Subcategorization' window in the QSAR Toolbox. The 'Data Gap Filling' tab is active, showing a table of chemical structures and their associated data. A callout '3' points to the 'Target' dropdown menu, which is set to 'Stable form'. Below the table, a trend analysis plot shows the prediction of LC50 values based on log Kow values. The plot includes a red regression line and several data points. A callout '4' points to the 'Remove' button in the bottom left corner of the plot area. The right-hand panel contains various options for accepting predictions and filtering data.

1 [Target]	2	6	44	45	60	67	69	74
[5] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]	[1] [T]
<chem>Cc1ccc(N)cc1[N+](=O)[O-]</chem>	<chem>Nc1ccccc1</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>CCCCCCCCCCCC</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>Cc1ccc(N)cc1</chem>	<chem>Cc1ccc(N)cc1</chem>
Pimephal... (73/207)	M. 24.7 mg/L, 24.8(20.6...	M. 107 mg/L, 75.5(...	M. 158 mg/L, 171(1...	M. 0.0614 mg/L, 0...	M. 2.99 mg/L, 2.8(2...	M. >20 mg/L	M. 36.3 mg/L	M. 73 mg/L, 73(65...

Trend analysis prediction of LC50,
making a linear approximation, based on 38 values from 12 analogue chemicals,
Observed target values: 24.7, 24.8 mg/L, Predicted target value: 46.9 mg/L,
Model equation: $LC50 = +2.03 + 0.772 * \log Kow$

- 1. 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

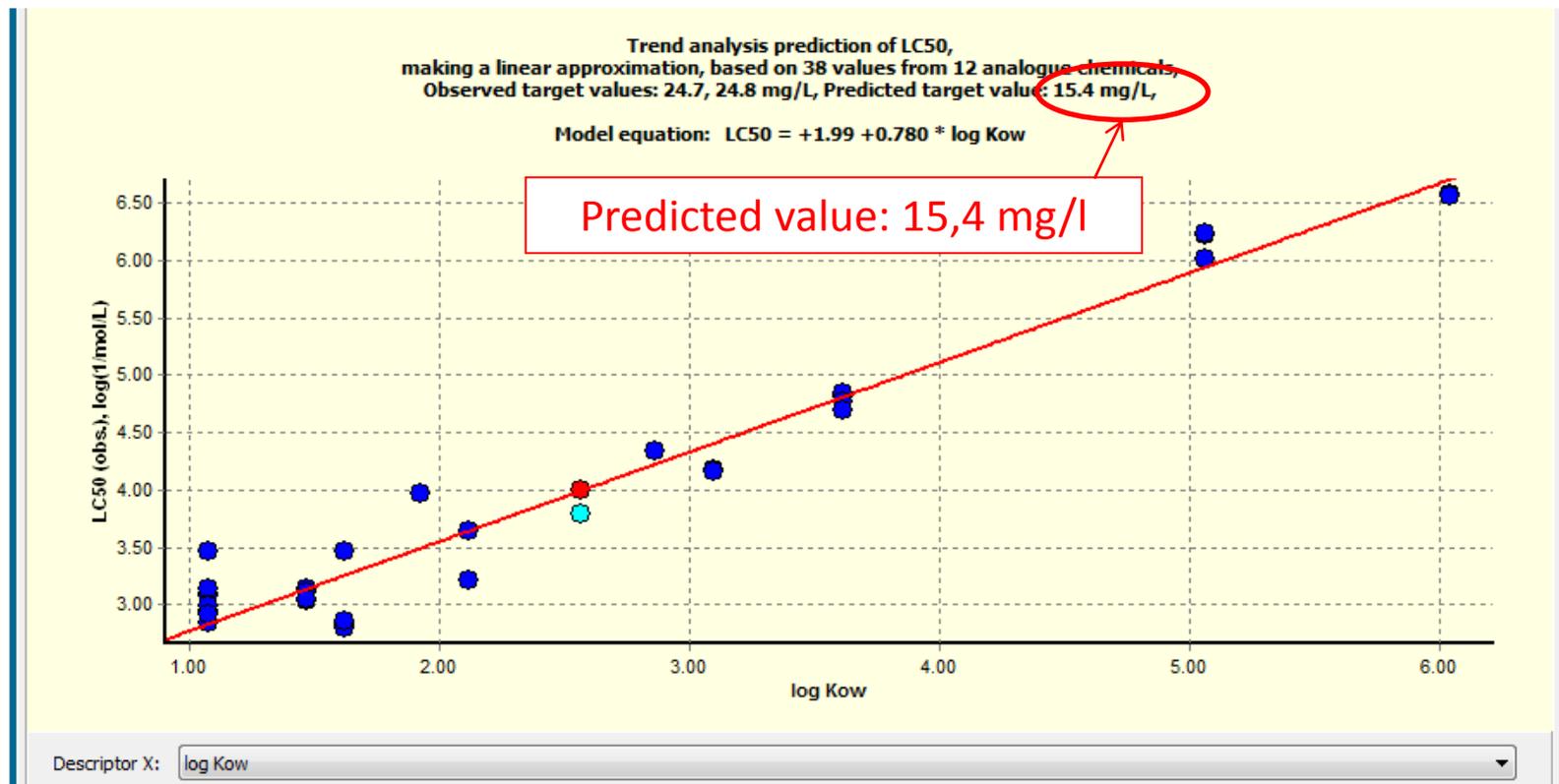
Result of Subcategorisation by Tautomers unstable

The screenshot shows the QSAR Toolbox interface during a 'Data Gap Filling' operation. On the left, the 'Subcategorization' panel is open, showing a list of grouping methods. Under the 'Toxicological' section, 'Tautomers unstable' is selected. The main window displays a table of chemical structures and their predicted LC50 values. A red circle highlights the 'Accept prediction' button in the right-hand panel. Below the table is a scatter plot titled 'Trend analysis prediction of LC50' showing a positive linear correlation between log Kow and LC50. The plot includes a red regression line and data points for various chemicals.

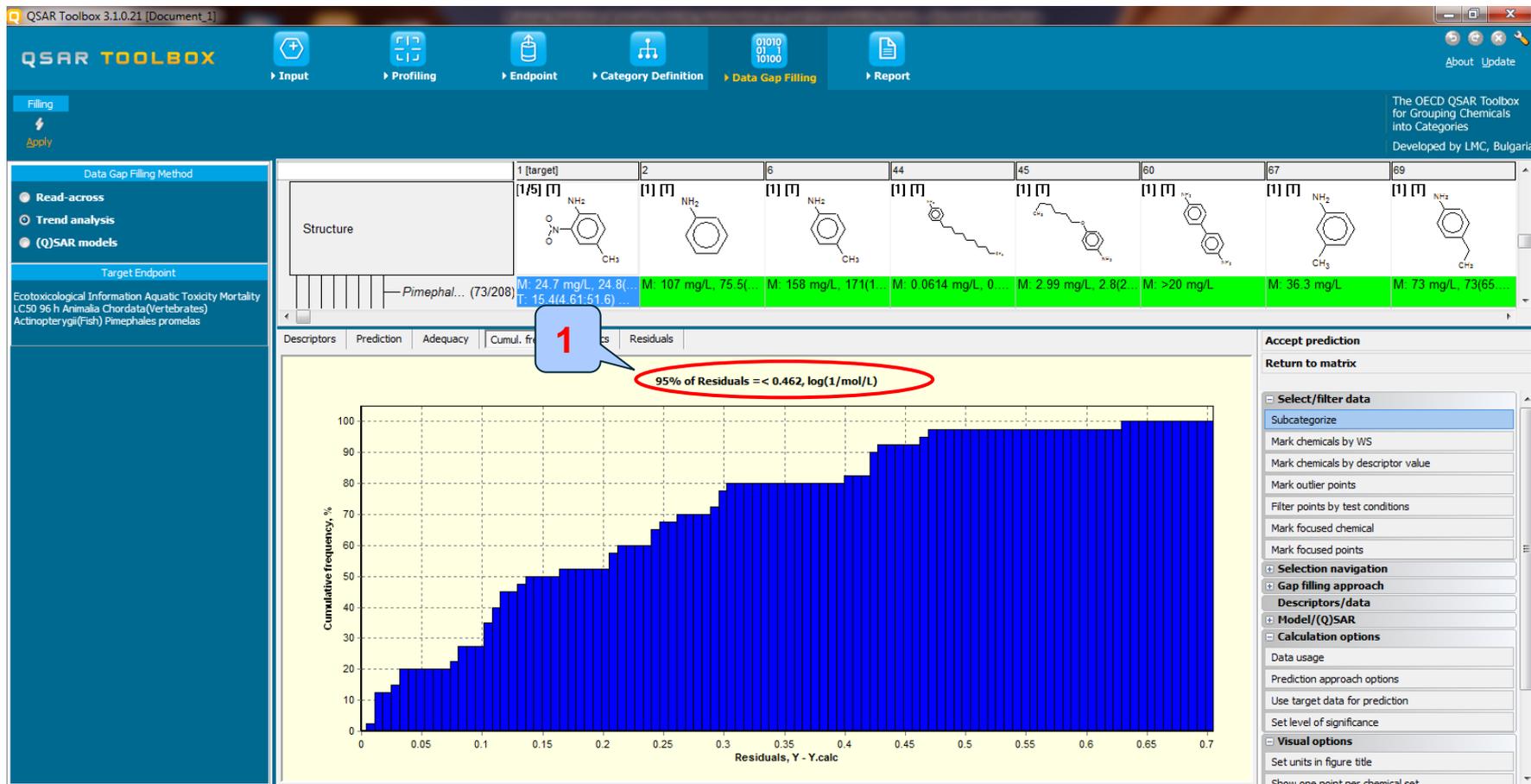
1. Accept prediction

1. Accept prediction

Data Gap Filling Result



Data Gap Filling Cumulated frequency



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

Data Gap Filling Statistics

QSAR Toolbox 3.1.0.21 [Document_1]

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(Vertebrates) Actinopterygii(Fish) Pimephales promelas

Structure	1 [target]	2	6	44	45	60	67	69
	M: 24.7 mg/L, 24.8(1) T: 15.4(4 61,51 6)	M: 107 mg/L, 75.5(1)	M: 158 mg/L, 171(1)	M: 0.0614 mg/L, 0...	M: 2.99 mg/L, 2.8(2)	M: >20 mg/L	M: 36.3 mg/L	M: 73 mg/L, 73(65)

Statistics

Statistical characteristics	TA model
Number of data points, (N)	38
Coefficient of determination, (R2)	0.955
Adjusted coefficient of determination, (R2adj)	0.953
Coefficient of determination - leave one out, (Q2)	0.950
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)	0.264
Fisher function, (F)	757
Fisher threshold for statistical significance, (Fa)	5.79

b0

- model descriptor	Intercept
- coeff. value	1.99
- coeff. range	± 0.16
- significance	Yes
- max. covariation	0.169 (vs b1)

b1

- model descriptor	log Kow
- coeff. value	0.780
- coeff. range	± 0.056
- significance	Yes
- max. covariation	0.169 (vs b0)

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
 - Set units in figure title
 - Show one point per chemical set

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

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The top menu bar includes Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The left sidebar shows the 'Data Gap Filling Method' (Read-across, Trend analysis, QSAR models) and 'Target Endpoint' (Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata(Vertebrates) Actinopterygii(Fish) Pimephales promelas). The main area displays a table with columns for chemical structures and various endpoints. A context menu is open over a cell containing a prediction for Pimephales promelas, with the 'Report' option highlighted. Red circles and blue callouts labeled '1' and '2' indicate the steps: clicking the cell and then selecting 'Report'.

Structure	1 [target]	2	3	4	5	6	7
<chem>Cc1ccc(N)cc1[N+](=O)[O-]</chem>	[5] [T]	[1] [T]	[1] [T]	[1] [T]	[9] [T]	[1] [T]	[2] [T]
Cyprinella lutrensis	(1/1)						
Cyprinodon variegatus	(1/1)						
Cyprinus carpio	(2/2)						
Danio rerio	(9/20)	M: 32;33 mg/L, 57...	M: 5.23 mg/L	M: 8.59 mg/L, 8.5(7...			
Esox lucius	(2/2)						
Esox masquinongy	(1/1)						
Gambusia affinis	(1/1)						
Gasterosteus aculeatus	(1/1)						
Ictalurus punctatus	(9/22)						
Jordanella floridae	(2/2)						
Lates calcarifer	(1/3)						
Lepomis cyanellus	(1/1)						
Lepomis macrochirus	(15/35)						
Leuciscus idus	(4/5)						
Morone saxatilis	(4/8)						
Oncorhynchus kisutch	(1/3)						
Oncorhynchus mykiss	(18/43)						
Oryzias latipes	(27/31)						
Perca flavescens	(2/4)						
Perca fluviatilis	(1/1)						
Pimephales promelas	(73/109)	M: 24.7 mg/L, 24.8(20.6;24.8) mg/L T: 15.5(4.48;53.7) mg/L					
Pleuronectes platessa	(1/1)						
Poecilia reticulata	(36/44)	M: 115 mg/L, 115(1...)	M: 6.25 mg/L	M: 6.6 mg/L, 9 mg/...			M: 20.4 mg/L
Pomatoschistus microps	(1/1)						
Salmo salar	(1/1)						

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

Report

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The main window displays a report titled "QSAR Toolbox prediction based on trend analysis" for the "Prediction of LC50 for benzenamine, 4-methyl-2-nitro". A callout box with the number "1" points to the "Summary" section of the report.

Summary

Toxicity of the target chemical (15.4 mg/L) is predicted from category members using trend analysis based on 38 values within the range 0.0614 - 171 mg/L from 12 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 endpoint data is selected from the following database(s):

1. Aquatic ECETOC
2. Aquatic Japan MoE
3. Aquatic OASIS
4. Aquatic US-EPA ECOTOX

Below is a summary table for endpoint & descriptor values for the target set of tautomers and the first 10 category members. Experimental values from data matrix are presented in bold font. Recalculated endpoint values (if required by selected data usage option in Gap Filling) are presented in italic font. Recalculated endpoint values based on experimental data only are presented in bold and italic font.

	<i>Endpoint(s)</i>	<i>Descriptor(s)</i>
	Aquatic Toxicity	log Kow

1. Summary information for tautomer prediction

Report

Available data to report

- Predictions
 - [1] 14.02.2013 15:27 [T]: 15.4(4.61;51.6) m
- (Q)SARs
- Categories

Available report templates

- Standard (predefined)
 - QSAR Toolbox Prediction Report (TPRF v.3.1)
- Custom (user defined)
 - Editable copy of QSAR Toolbox Prediction Rep

Prediction [1]

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	24.7 mg/L
Information# Aquatic Toxicity	24.8 mg/L

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

i. Predicted value (model result):
15.4 mg/L

j. Predicted value (comments):
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)

manually editable field

QSAR Toolbox 3.1.0.21 Database version: 3.4.4/3.1.2 QSAR TOOLBOX TPRF v.3.1.1.31004

1. Predicted value

2. Applicability domain

Report

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The main window displays a prediction report titled "QSAR Toolbox prediction based on trend analysis" for "benzenamine, 4-methyl-2-nitro-". The report includes a section for "APPENDIX 7 - Chemical components" with a list of information:

- Tautomer No.1 of target chemical and its tautomers:** (circled in red)
- 1. CAS number:** 89-62-3
- 2. Other regulatory numbers:** Not reported
- 3. Chemical name(s):** Not available
- 4. Structural formula:**

A green box labeled "used in prediction" (with a callout bubble containing the number 1) is positioned next to the tautomer information. The left sidebar shows available report templates, including "Standard (predefined)" and "Custom (user defined)".

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