

OECD QSAR Toolbox v.3.4

Predicting acute aquatic toxicity to fish of
Dodecanenitrile (CAS 2437-25-4) 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 the active tautomeric form
 - Retrieve experimental results available for those analogues
 - Perform trend analysis for the active tautomeric form
 - Assigning of the prediction for the active tautomer to the target chemical
 - Finally saved the 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 *Dodecanenitrile* (CAS 2437-25-4)
- Set of simulated tautomers for the target chemical will be provided
- Analyze the profilers of the tautomeric forms within tautomeric set
- Filling data gaps for active tautomer by trend analysis
- Assign prediction for the tautomeric forms to the target chemical

Outlook

- Background
- 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

Chemical Input

Ways of Entering a Chemicals

User Alternatives for input of Chemical:

A. Single target chemical

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

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input

Input Screen

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

Chemical Input

Input target chemical by CAS#

The screenshot illustrates the steps to input a chemical by CAS number in the QSAR Toolbox. The interface includes a menu bar with 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. A toolbar contains icons for 'CAS#', 'Name', 'Structure', 'Select', 'Delete', 'Query', 'ChemIDs', 'DB', 'Inventory', and 'List'. A 'Documents' panel on the left shows a list of documents. A 'Filter endpoint tree...' window is open, displaying a search for '2437-25-4'. A dialog box titled 'Search by CAS #' is shown, with the input field containing '2437-25-4' and 'OK' and 'Cancel' buttons. A table below the dialog shows the search results, including the CAS number, SMILES string, depiction, and names. A 'Structure' window is also visible, showing the chemical structure of hexane.

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS
1.	2437-25-4	CCCCCCC			1:: High	1:: High	H
					1:: Av	1:: U:	
					2:: D:	2:: Av	
					3:: U:	3:: D:	
					2:: Mode	2:: Mode	
					1: 1:: E:	1: 1:: Pf	
					2: 2:: H:	2: 2:: U:	
					3: 3:: M:	3: 3:: H:	
					4: 4:: Pf	4: 4: M	
					5: 5:: U:	5: 5: E:	
					6: 3:: Low	3: 3: Low	
					1:: Sl	1: 1: Sl	
					4: 4: High	4: 4: High	
					1:: C	1: 1: T	

1. Click on CAS#; **2. Enter** 2437-25-4; **3.** The system identify the structure; **4. OK**

Chemical Input

Target chemical identity

- **Double click** "Substance Identity" displays the chemical identification information.
- The user should note that existing names of the target chemical are presented in different colours. This indicates the reliability of relation CAS-Name for the target chemical(see next screen shots).
- The workflow on the first module is now complete, and the user can proceed to the next module.

Chemical Input

Target chemical identity

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Document', 'Single Chemical', and 'Chemical List'. The 'Input' tab is active, showing a 'Filter endpoint tree...' window with '1 [target]' selected. The main workspace shows the chemical structure of n-undecyl cyanide and its 'Substance Identity' details. A red circle highlights the following information:

- CAS Number: 2437-25-4
- EINECS: 2194401
- Chemical Name: n-undecyl cyanide
- Other names: lauronitrile, dodecanonitrile, dodecanenitrile, undecyl cyanide, c12 nitrile
- Molecular Formula: C12H23N
- Structural Formula: CCCCCCCCCC#N

The interface also shows a 'Documents' panel on the left with a tree view containing 'Document' and 'CAS: 2437-25-4'. The bottom of the window features a 'select filter type ..' dropdown and 'Create' and 'Apply' buttons.

Chemical Input

Target chemical identity

The colour code indicates the reliability of the chemical identifier:

- **Green:** There is a high consistency between the identifier and the structure. This colour is applied if the identifier is the same in several quality assured databases.
- **Yellow:** There is only a moderate reliability between the identifier and the structure. The colour is applied if the identifier is the same in several databases for which the quality assurance could not be established.
- **Red:** There is a poor reliability between the identifier and the structure. The colour is applied if the identifier is allocated to different structures in different databases.

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 main window has a blue header with the 'QSAR TOOLBOX' logo and several icons for different functions: Profiling, Endpoints, Category Definition, Data Gap Filling, and Report. Below the header is a menu bar with 'Profiling' and 'Themes'. 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 'Profiling methods' section, 'US-EPA New Chemical Categories' is selected, with a red circle around it and a callout box containing the number '1'. A red arrow points from the 'View' button to the 'US-EPA New Chemical Categories' entry. The 'Profiling Scheme Browser' window is open, showing a list of chemical categories. 'Esters (Acute toxicity)' is highlighted in blue. A callout box with the number '3' points to this entry. The 'Profile Description' window is also open, showing a detailed textual description for 'Esters Environmental Toxicity'. An arrow points from the text 'Textual description' to the top of this window.

1. **Highlight** the profiler
2. **Click** View
3. **Click** Advance in order to see detailed description of highlighted category (in this case "Esters")

Profiling Side-Bar to Profiling

1. Highlight the profiler

2. Click View

3. Select "Esters(Acute toxicity)"

Profiling

Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information about profilers could be found in “Manual for Getting started” (Chapter 4) published on the OECD website:

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

- Table 4 - 1 in chapter 4 (Manual for getting started) lists a selection of profilers and their relevance for different endpoints of regulatory relevance
- The following profiling schemes are relevant to the **Acute aquatic toxicity**:
 - Aquatic toxicity classification by ECOSAR
 - Acute aquatic toxicity MOA by OASIS
 - Acute aquatic toxicity classification by Verhaar
 - Protein binding alerts by OASIS v1.4
 - Protein binding by OECD
 - Organic function groups – all four profilers are used in the assessment
 - Chemical elements
- More details about profiling schemes used for categorization and collection of analogues is provided in stage “Category formation” on slide 50

Profiling

Profiling the target chemical

- 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 mark disappears) profilers.
- For this example, go through the general and endpoint specific profiling mechanisms and highlight those that apply to acute aquatic toxicity(see next screen shot).

Profiling

Profiling the target chemical

1. **Check** profilers mentioned on #21
2. **Click** Apply

Profiling

Profiling the target chemical

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

(see next screenshot)

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is active, showing options like 'Apply', 'New', 'View', and 'Delete'. The 'Profiling methods' panel on the left lists various toxicity and hazard assessment methods, with several checked under 'Empiric'. The 'Filter endpoint tree...' panel shows a hierarchical tree structure. The 'Profile' node is circled in red, and a blue callout bubble with the number '1' points to it. The table on the right displays the results of the profiling, with the 'Profile' row highlighted in blue and a red arrow pointing to it.

Filter endpoint tree...	1 [target]
Structure	
Substance Identity	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	
Profile	
Predefined	
OECD HPV Chemical Categories	Not categorized
US-EPA New Chemical Categories	Not categorized
General Mechanistic	
Protein binding by OASIS v1.4	No alert found
Protein binding by OECD	No alert found
Endpoint Specific	
Acute aquatic toxicity classification by Verhaar (Modified)	Class 5 (Not possible to classify according ...
Acute aquatic toxicity MOA by OASIS	Basesurface narcotics
Aquatic toxicity classification by ECOSAR	Neutral Organics
Empiric	
Chemical elements	Group 14 - Carbon C
Organic Functional groups	Group 15 - Nitrogen N
Organic Functional groups (nested)	Nitrile
Organic functional groups (US EPA)	Nitrile
Organic functional groups, Norbert Haider (checkmol)	Acetylenic Carbon [#C]
	Aliphatic Carbon [CH]
	Aliphatic Carbon [-CH2-]
	Aliphatic Carbon [-CH3]
	Cyano, aliphatic attach [-C#N]
	Nitrile

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

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

Outlook

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

Endpoint Overview

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

Endpoint

Case study

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

Endpoint Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Endpoint' menu is open, showing options: 'Gather', 'Import', 'IUCLID5', 'Export', 'IUCLID5', 'Database', 'Inventory', and 'Database'. The 'Gather' option is circled in red with a callout box containing the number '4'. In the left sidebar, the 'Databases' tree is expanded to 'Ecotoxicological Information', which is also circled in red with a callout box containing the number '2'. Under 'Ecotoxicological Information', several sub-items are checked, including 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', 'ECHA CHEM', and 'ECOTOX'. The 'Human Health Hazards' section is also circled in red with a callout box containing the number '3'. The main window shows a 'Filter endpoint tree...' dialog with a 'Structure' tab and a list of endpoint categories: 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Human Health Hazards', and 'Profile'. The 'Ecotoxicological Information' category is selected. A chemical structure is visible in the 'Structure' tab. A callout box with the number '1' points to the 'Endpoint' menu item.

1. **Click** Endpoint
2. **Expand** the Ecotoxicological Information section
3. **Select** databases related to the target endpoint
4. **Click** Gather

Endpoint Gather data

- Toxicity information on the target chemical is electronically collected from the selected dataset(s)
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are Aquatic toxicity OASIS; Aquatic ECETOC; Aquatic Japan MoE and ECOTOX
- In this example, there is LC50 experimental data for *P. promelas* (96h) for the target chemical (see next screen shots)
- The experimental data for the investigated endpoint falls within the toxic range (less than 1mg/l¹)

¹ **Globally Harmonized System of Classification and Labeling of Chemicals (GHS):**
http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/ST-SG-AC10-30-Rev6e.pdf

Endpoint Gather data

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

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', 'Export', 'Delete', and 'Tautomerize'. The 'Endpoint' menu is open, showing options like 'Gather', 'Import', 'IUCLID5', 'Export', 'IUCLID5', 'Database', 'Inventory', and 'Database'. A 'Read data?' dialog box is overlaid on the interface, with a blue callout bubble containing the number '1' pointing to the 'OK' button. The dialog box has the following options: 'All endpoints' (selected), 'Choose...', and 'from Tautomers' (checked). The 'OK' button is highlighted with a green checkmark, and the 'Cancel' button is highlighted with a red X. The background shows a tree view of chemical properties and a chemical structure.

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

Endpoint Gather data

The screenshot shows the QSAR Toolbox interface with the 'Endpoint' menu selected. The left sidebar shows a filter tree where 'Pime' is selected under 'Ecotoxicological Information'. The main window displays a data matrix for 'Pime' with the following data:

Endpoint	Value
Behavior (1/5)	M: >1.5;>2.25 mg/L, >0;>0.75 mg/L, >1.5;>2.25 ...
Mortality	
LC50	
Undefined Duration	
12 h (1/1)	M: >1.5;>2.25 mg/L
24 h (1/1)	M: >1.5;>2.25 mg/L
48 h (1/1)	M: >0.75;>1.5 mg/L
72 h (1/1)	M: >0;>0.75 mg/L
96 h	
Animalia	
Chordata (Vertebrates)	
Actinopterygii (Fish)	
<i>Pimephales promelas</i> (1/2)	M: 0.425 mg/L, 0.43(0.4,0.47) mg/L
Profile	

1. **Type** "Pime" in the filter tree in order to filter the tree to the investigated endpoint
2. Available experimental data appears on datamatrix (LC50 0.425 mg/l species: *P.promelas*, duration: 96h)

Endpoint Gather data

The screenshot displays the QSAR Toolbox interface during the 'Endpoint' workflow. The 'Databases' panel on the left is expanded to 'Ecotoxicological Information' > 'Aquatic Toxicity' > 'Mortality' > 'LC50'. The 'Structure' window shows the chemical structure of Pimephales promelas. The 'Data points' table is open, showing two rows of data. A callout box with a red '1' points to the cell containing '0.43(0.4,0.47) mg/L' in the 'Value' column. A second callout box with a red '2' points to the 'X' button in the top right corner of the 'Data points' window.

1. Double-click on the cell displays metadata information for the observed data

2. Click on the X to close the window

#	endpoint	Value	Original value	Effect	Source	publication_year	published_date	significance_type	significance_code	modified_date	media_type	measurement	other_effect_comments	organism_source	Organism habitat	subhabitat	num_doses
1	LC50	0.425 mg/L	2.34E-6 mol/L	Mortality													
2	LC50	0.43(0.4,0.47) mg/L	0.43(0.4,0.47) milligram per liter	Mortality	Center for Lake Superior Environment	1984	03/12/2014	NA	NA	02/18/2014	fresh water	Mortality	TOXICITY SYMPTOMS	laboratory	Water	Unspecified	6

Recap

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

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - **Handling of tautomerism of target chemical**

Handling of tautomerism of target chemical

Visualization of modeling modes

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

Handling of tautomerism of target chemical

Visualization of modeling modes

The screenshot illustrates the workflow in QSAR Toolbox. Step 1 shows the 'Input' menu being selected. Step 2 shows a right-click context menu over the document tree. Step 3 shows the 'Multiplication' > 'Tautomerism' path being chosen. Step 4 shows a 2D chemical structure being selected from the generated list. A separate window displays three tautomeric forms for the target chemical (CAS# 2437-25-4), labeled 'Tautomeric forms'.

1. **Go** to Input
2. **Right click** over the node with SMILES and **select** Multiplication and then **Tautomerism**
3. **Open** the tree to the end leaf with [set] indication
4. Three tautomeric forms are generated for the target chemical. Double click over the 2D structure with [3][T]

Handling of tautomerism of target chemical

Visualization of modeling modes

- Two component modes are implemented:
 - **Set Mode** - all tautomers are analyzed as a package
 - **Individual Component Mode** - each tautomer is analyzed individually

(see next screen shot)

Handling of tautomerism of target chemical Visualization of modeling modes

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

The screenshot shows the QSAR Toolbox interface with the 'Component Mode' dropdown menu open. The 'All' radio button is selected and circled in red. The main window displays a chemical structure and a list of properties including Substance Identity, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, Human Health Hazards, and Profile. A red circle highlights the chemical structure and the list of properties.

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

The screenshot shows the QSAR Toolbox interface with the 'Component Mode' dropdown menu open. The 'Single' radio button is selected and circled in red. The main window displays four chemical structures, each representing a different tautomeric form of the target chemical. A red oval highlights the four chemical structures.

Different modes for visualization for the set of target and its tautomeric forms is implemented. The latter are visualized when the node with [set] is selected. Single mode is used in further trend analysis.

Outlook

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

Handling of tautomerism of target chemical

Profiling set of tautomers

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

Handling of tautomerism of target chemical

Profiling set of tautomers

- The following primary profilers relevant to the **aquatic toxicity** are used in this example(see next screenshot):
 - 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
 - Protein binding by OASIS v1.4
 - Protein binding by OECD
 - 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.

Handling of tautomerism of target chemical

Profiling set of tautomeric forms

1 Check the profilers related to acute aquatic toxicity as mention on slide #42

2 Click Apply

	1 [target]	2 [target,tautomer]	3 [target,tautomer]	4 [target,tautomer]
Structure	[Chemical Structure]	[Chemical Structure]	[Chemical Structure]	[Chemical Structure]
Substance Identity				
Ecotoxicological Information				
Profile				
Predefined				
OECD HPV Chemical Categories	Not categorized	Not categorized	Not categorized	Not categorized
US-EPA New Chemical Categories	Not categorized	Not categorized	Not categorized	Not categorized
General Mechanistic				
Protein binding by OASIS v1.4	No alert found	No alert found	No alert found	No alert found
Protein binding by OECD	No alert found	No alert found	No alert found	No alert found
Endpoint Specific				
Acute aquatic toxicity classification by Verhaar	Class 5 (Not possible to classif...	Class 5 (Not possi...	Class 5 (Not possi...	Class 5 (Not possi...
Acute aquatic toxicity MOA by OASIS	Basesurface narcotics	Basesurface narco...	Basesurface narco...	Basesurface narco...
Aquatic toxicity classification by ECOSAR	Aliphatic Amines	Not Related to an ...	Neutral Organics	Aliphatic Amines
Empiric				
Chemical elements				
Organic Functional groups				
Organic Functional groups (nested)				
Organic functional groups (US EPA)				
Organic functional groups, Norbert Haider (checkedmol)				
Tautomers unstable				
Toxicological				
Metabolism/Transformations				

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

Handling of tautomerism of target chemical

Recap

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

Outlook

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

Handling of tautomerism of target chemical

Focus of active tautomer

This tautomeric form is selected for further trend analysis

1

2

	1 [target]	2 [target,tautomer]	3 [target,tautomer]	4 [target,tautomer]
[3] [7]				
Substance Identity				
Ecotoxicological Information				
Profile				
Predefined				
OECD HPV Chemical Categories	Not categorized	Not categorized	Not categorized	Not categorized
US-EPA New Chemical Categories	Not categorized	Not categorized	Not categorized	Not categorized
General Mechanistic				
Protein binding by OASIS v1.4	No alert found	No alert found	No alert found	No alert found
Protein binding by OECD	No alert found	No alert found	No alert found	No alert found
Endpoint Specific				
Acute aquatic toxicity classification by Ver...	Class 5 (Not possible t...	Class 5 (Not possi...	Class 5 (Not possi...	Class 5 (Not pos
Acute aquatic toxicity MOA by OASIS	Basesurface narcotics	Basesurface narco...	Basesurface narco...	Basesurface nar...
Aquatic toxicity classification by ECOSAR	Aliphatic Amines Neutral Organics Not Related to an Exist...	Not Related to an ...	Neutral Organics	Aliphatic Amines
Empiric				
Organic Functional groups	Aliphatic Amine, primary Alkene Alkyne Allyl Ketenimine Nitrile	Alkene Allyl Ketenimine		
Organic Functional groups (nested)	Aliphatic Amine, primary Alkyne Allyl Ketenimine Nitrile Overlapping groups	Allyl Ketenimine Overlapping groups		

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

1. Right click over the active tautomeric form with positive ECOSAR alert
2. Select Focus from the appeared menu

Handling of tautomerism of target chemical

Focus of active tautomer

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 bar is a toolbar with icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Select', 'Delete', 'Query', 'ChemIDs', 'DB', 'Inventory', and 'List'. The main workspace is divided into several panels:

- Documents:** A tree view showing a document structure. The selected tautomer is circled in red.
- Structure:** A panel showing the chemical structure of the selected tautomer, also circled in red.
- Chemical List:** A table showing the chemical properties of the selected tautomer. The table has the following data:

Property	Value
Substance Identity	
CAS Number	2437-25-4
Chemical IDs	NA
Chemical Name	
Molecular Formula	C12H23N
Structural Formula	CCCCCCCCC#CN
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	
Profile	

The selected tautomer appears in a new data matrix.

Outlook

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

Category Definition

Grouping methods

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

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

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

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

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

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

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

**Broad grouping
Endpoint Non-specific**

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based*

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

**Subcategorization
Endpoint Specific**

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

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

**Subcategorization
Endpoint Specific**

Handling of tautomerism of target chemical

Category definition for active tautomeric form

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

Handling of tautomerism of target chemical

Check databases

The screenshot displays the QSAR Toolbox software interface. At the top, there is a navigation bar with icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below this is a secondary menu with buttons for Data, Import, Export, Delete, and Tautomerize. The main workspace is divided into several panels:

- Databases Panel (Left):** Contains a tree view of database categories. Under 'Ecotoxicological Information', several sub-databases are checked, including 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', 'ECHA CHEM', and 'ECOTOX'.
- Inventories Panel (Bottom Left):** Lists various inventory sources such as 'Canada DSL', 'COSING', 'DSSTOX', 'ECHA PR', 'EINECS', 'HPVC OECD', 'METI Japan', and 'NICNAS'.
- Structure Panel (Top Center):** Shows the chemical structure of the target compound, labeled 'pime'.
- Taxonomic Tree Panel (Center):** A hierarchical tree showing the classification of the chemical. The path 'Actinopterygii (Fish)' > 'Pimephales promelas' is highlighted in blue.
- Properties Panel (Right):** Displays a table with one entry: '1 [target,tautomer]'. Above the table is a small graph showing a sawtooth-like pattern.

Handling of tautomerism of target chemical

Defining ECOSAR category

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

Handling of tautomerism of target chemical

Defining ECOSAR category

1. Highlight "Aquatic toxicity classification by ECOSAR" **2. Click** Define **3. Select** Strict **4. Click** OK to confirm the category **Aliphatic amines** defined by ECOSAR.

Handling of tautomerism of target chemical

Defining ECOSAR category

The screenshot displays the QSAR Toolbox interface during the 'Category Definition' phase. The main window shows a hierarchical tree of classification methods. Under 'Ecotoxicological Information', 'Aquatic Toxicity' is expanded to 'LC50', which is further expanded to '96 h'. A dialog box titled 'Define category name' is open, showing the selected category name 'Strict (Aquatic toxicity classification by ECOSAR)' for 303 chemicals. A blue callout box with the number '1' points to the 'OK' button in the dialog. A banner at the bottom of the screenshot reads '1. Click OK to confirm the name of the category'.

Handling of tautomerism of target chemical

Category analogues

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

Handling of tautomerism of target chemical

Read data for Analogues

- The Toolbox automatically 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.

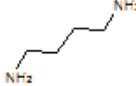
Handling of tautomerism of target chemical

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: 1406 data-points, 422 groups, 168 chemicals

Data points...

	Endpoint	CAS	Structure	Value	additional_comments	application_date	
<input checked="" type="checkbox"/>	NOEC	71-44-3		1 millimolar (millimoles per liter)	SOURCE/COLLECTED FROM VERAVAL COAST, GUJARAT//		<input type="button" value="Select one"/> <input type="button" value="Invert"/> <input type="button" value="Check All"/> <input type="button" value="Uncheck All"/> <input checked="" type="button" value="OK"/> <input type="button" value="Cancel"/>
<input checked="" type="checkbox"/>	NOEC	71-44-3		1 millimolar (millimoles per liter)	SOURCE/COLLECTED FROM VERAVAL COAST, GUJARAT//		
<input checked="" type="checkbox"/>	NOEC	110-60-1		1 millimolar (millimoles per liter)	SOURCE/COLLECTED FROM VERAVAL COAST, GUJARAT//		
<input checked="" type="checkbox"/>	NOEC	110-60-1		1 millimolar (millimoles per liter)	SOURCE/COLLECTED FROM VERAVAL COAST, GUJARAT//		
<input checked="" type="checkbox"/>	NOEC	110-60-1		1 millimolar (millimoles per liter)	SOURCE/COLLECTED FROM VERAVAL COAST, GUJARAT//		

1. Click Select one and then
2. Click OK

Handling of tautomerism of target chemical

Summary information for Analogues

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes tabs for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below the navigation bar is a toolbar with buttons for Define, Define with metabolism, Subcategorize, Combine, Clustering, Delete, and Delete All. The main window is divided into several panes:

- Grouping methods:** A list of methods including Hydrolysis half-life, Ionization at pH, Protein binding by OASIS/OECD, Toxic hazard classification by Cramer, and various aquatic toxicity and carcinogenicity alerts.
- Structure:** A central pane showing the chemical structure of 'pime' and a tree view of categories. The 'Actinopterygii (Fish)' category is highlighted in red, and 'Pimephales promelas (66/133)' is highlighted in blue.
- Datamatrix:** A table showing experimental data for the chemical and its analogues. The highlighted row for 'Pimephales promelas' shows 'M: 21.4 mg/L, 21.4...'.

Available aquatic experimental data for the analogues appears on datamatrix.

Recap

- You have identified a category (“Aliphatic amines”) with the “Acute aquatic toxicity classification by ECOSAR” profiler for the target chemical *Dodecanenitrile* (CAS 2437-25-4)
- The available experimental results for these 303 analogues have been collected from the selected databases (Aquatic ECETOC, Aquatic Japan MoE, 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.

Handling of tautomerism of target chemical

Navigation through the endpoint tree

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

Handling of tautomerism of target chemical

Navigation through the endpoint tree

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

- Top Bar:** Navigation tabs for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Toolbar:** Standard file operations (New, Open, Close, Save) and chemical-specific actions (CAS#, Name, Structure, Delete, Query, ChemIDs, DB, Inventory, List).
- Documents Panel:** Shows the current document structure with chemical formulas like CCCCCCCCCCC#N.
- Search Bar:** Contains the filter text "pime".
- Endpoint Tree:** A hierarchical tree of endpoints. The path is: EC50 (1/1) > LC10 (2/2) > LC50 > Undefined Duration > 1 h (1/1) > 3 h (2/2) > 6 h (1/1) > 12 h (1/1) > 24 h (4/4) > 48 h (4/4) > 72 h (3) > 96 h > Animalia > Chordata (Vertebrates) > Actinopterygii (Fish) > **Pimephales promelas (66/133)**.
- Chemical List:** A table with 6 columns. The first column contains chemical structures, and the second column contains the text "M: 21.4 mg/L, 21.4...".

- 1. Type** "Pimephales promelas" in the filter box or just "Pime", 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 303 analogue chemicals of focused tautomeric form classified as “Aliphatic amines” by the “ECOSAR” profiler.
- You have identified the target endpoint of “96 h LC50 Mortality for *Pimephales promelas*”.
- You are ready to fill in the data gap so click on “Data Gap Filling” (see next screen shots).

Outlook

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

Data Gap Filling

Apply Trend analysis

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The left sidebar has a 'Data Gap Filling' section with three radio buttons: 'Read across', 'Trend analysis', and '(Q)SAR models'. The 'Trend analysis' option is selected. The main workspace displays a hierarchical tree of endpoints for the chemical 'pime'. The endpoint 'Pimephales promelas (66/133)' is highlighted in blue. A red circle highlights the 'Apply' button in the top left, and another red circle highlights the highlighted endpoint box. A red arrow points from the 'Apply' button to the highlighted endpoint box. Three numbered callouts (1, 2, 3) are present: 1 points to the highlighted endpoint box, 2 points to the 'Trend analysis' radio button, and 3 points to the 'Apply' button.

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

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

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

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 pimephales

Structure

1 [target,tautomer] 4 10 13 18 19

Pimephales pimephales (66/133)

M: 21.4 mg/L, 21.4... M: 310 mg/L, 308... M: 266 mg/L, 268... M: 56.9 mg/L, 56.6... M: 22 mg/L, 21.8(1...)

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 66 values from 66 analogue chemicals, Observed target value: N/A, Predicted target value: 3.00 mg/L, Model equation: $LC50 = +2.32 + 0.607 * \log Kow$

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
- Information
- Miscellaneous

303 Aliphatic Amines Strict (Aquatic toxicity classification by ECOSAR) Create prediction by gap filling 0/1 1/10

Data Gap Filling

Side-Bar of Subcategorisation

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

- Chemical elements

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

- Organic functional groups (nested)

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

Subcategorisation steps are demonstrated on the next screen shots.

Data Gap Filling

Subcategorisation by Chemical elements

The screenshot shows the 'Subcategorization' window in the QSAR Toolbox. The interface includes a 'Grouping methods' list on the left, a 'Do not account for' list, and a central 'Trend analysis prediction of LC50' plot. The plot shows a positive correlation between log Kow and LC50 (obs., log1(mol/L)) for 66 values. A red regression line is shown with the equation: $LC50 = +2.32 + 0.607 * \log Kow$. The plot is labeled with a blue callout '1'. The 'Subcategorize' button on the right is also labeled with a blue callout '1'. The 'Chemical elements' option in the 'Grouping methods' list is circled in red and labeled with a blue callout '2'. The 'Remove' button in the 'Do not account for' list is circled in red and labeled with a blue callout '3'.

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

Data Gap Filling

Result of Subcategorisation by Chemical elements

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,tautomer] 4 10 13 18 19 20

Pimephales promelas (40/81)

M. 21.4 mg/L, 21.4 M. 310 mg/L, 308 M. 266 mg/L, 268 M. 56.9 mg/L, 56.6 M. 22 mg/L, 21.8(1. M. 5.15 mg/L

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 40 values from 40 analogue chemicals, Observed target value: N/A, Predicted target value: 3.10 mg/L, Model equation: $LC50 = +2.21 + 0.632 * \log Kow$

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
- Visual options
- Information
- Miscellaneous

Data Gap Filling Subcategorisation by OFG (nested)

The screenshot displays the QSAR Toolbox software interface during a Data Gap Filling session. The main window shows a list of chemical structures with their corresponding predicted LC50 values. A trend analysis plot is visible, showing the relationship between log Kow and LC50 (obs., log10(mol/L)). The plot includes a red regression line and data points. The interface is annotated with three callout boxes:

- 1:** Points to the 'Subcategorize' button in the right sidebar under the 'Accept prediction' section.
- 2:** Points to the 'Organic Functional groups (nested)' option in the left sidebar under the 'Empiric' section.
- 3:** Points to the 'Remove' button at the bottom of the left sidebar.

The trend analysis plot shows the following equation: $LC50 = +2.21 + 0.632 * \log Kow$. The observed target value is N/A, and the predicted target value is 3.10 mg/L. The plot displays data points for various chemical structures, with a red regression line indicating the trend.

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

Data Gap Filling

Result of Subcategorisation by OFG (nested)

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

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

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 pimephales

Structure

1 [target,tautomer] 10 13 18 19 20 22

Pimephales pimephales (18/40)

M. 310 mg/L, 300(M. 266 mg/L, 260	M. 56.9 mg/L, 56.6	M. 22 mg/L, 21.8(1	M. 5.15 mg/L, 5.19	M. 2.17 mg/L
-------------------	------------------	--------------------	--------------------	--------------------	--------------

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 18 values from 18 analogue chemicals, Observed target value: N/A, Predicted target value: 0.929 mg/L, Model equation: $LC50 = +2.39 + 0.715 * \log Kow$

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
 - Go back
 - Go forward
 - Go to first
 - Go to last
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

303. Aliphatic Amines Strict (Aquatic toxicity classification by ECOSAR) Create prediction by gap filling 1/10

Data Gap Filling

Side-Bar of Subcategorisation

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

- Structural similarity

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

Analogues with similarity less than 30 % have been eliminated

See next two slide

Data Gap Filling

Subcategorisation by Structural similarity

Most dissimilar analogues are highlighted in green. Most of them are dialiphatic amines and short chain aliphatic amines

Developed by LMC, Bulgaria

1. Select Structure similarity; 2. Manually select categories between 0 and 30% (hold Ctrl button and select categories); 3. Dissimilar analogues are highlighted in green; 4. Click Remove to eliminate dissimilar analogues

1. **Select** Structure similarity; 2. Manually select categories between 0 and 30% (hold Ctrl button and select categories); 3. Dissimilar analogues are highlighted in green; 4. **Click** Remove to eliminate dissimilar analogues

Data Gap Filling Result

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 Animals Chordata Actinopterygii Pimephales promelas

Structure

pime 25 33 36 40 65 100

Structure

Pimephales pro... (11/25) M: 0.102 mg/L, 0.1 M: 1.04 mg/L, 1.04 M: 0.066 mg/L, 0.0 M: 0.211 mg/L, 0.2 M: 178 mg/L, 177 M: 5.15 mg/L, 5.28

Descriptors Prediction Adequacy Cumul. freq. Statistics Residuals

Trend analysis prediction of LC50, making a linear approximation, based on 11 values from 11 analog chemicals, Observed target value: N/A, Predicted target value: 0.551 mg/L.

Model equation: $LC50 = +1.55 + 0.979 * \log Kow$

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

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
 - Go back
 - Go forward
 - Go to first
 - Go to last
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

Predicted value: 0.55 mg/l

303 Aliphatic Amines Strict (Aquatic toxicity classification by ECOSAR) Create prediction by gap filling 1/1/0

Data Gap Filling

Cumulated frequency

1. 95% of Residuals = $\leq 0.210, \log(1/\text{mol/L})$

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

Data Gap Filling Statistics

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

Pimephales pro... (11/25) M: 0.102 mg/L, 0.1 M: 1.04 mg/L, 1.04 M: 0.066 mg/L, 0.0 M: 0.211 mg/L, 0.2 M: 178 mg/L, 177 M: 5.15 mg/L, 5.28

Descriptors Prediction Adequacy Cumul. freq. **Statistics** Residuals

Statistical characteristics		TA model
Number of data points, (N)		11
Coefficient of determination, (R2)		0.991
Adjusted coefficient of determination, (R2adj)		0.990
Coefficient of determination - leave one out, (Q2)		0.985
Coefficient of correlation for external set, (r2)		-
Sum of squared residuals, (SSR)		0.173
Standard deviation of residuals, (sN)		-
Sample standard deviation of residuals, (s)		0.139
Fisher function, (F)		992
Fisher threshold for statistical significance, (Fa)		7.71

b0

- model descriptor	Intercept
- coeff. value	1.55
- coeff. range	± 0.23
- significance	Yes
- max. covariation	0.389 (vs b1)

b1

- model descriptor	log Kow
- coeff. value	0.979
- coeff. range	± 0.070
- significance	Yes
- max. covariation	0.389 (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
 - Visual options
 - Information
 - Miscellaneous

1. The high R2 and Q2 support the reliability of the prediction

Data Gap Filling

Result of trend analysis

- The analysis of trend analysis shows:
 - The predicted acute aquatic toxicity value is 0.55 mg/l
 - The remaining analogues form robust category of structurally similar analogues (aliphatic amines)
 - The 95% of residuals are in the range of experimental error
 - The high R² and Q² coefficient values support the reliability of the prediction

Data Gap Filling

Accept the prediction

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling session. The main window shows a scatter plot of observed LC50 values (log(1/mol/L)) versus log Kow. A red regression line is fitted to the data points. An information dialog box is open, stating "The current prediction was accepted" with an "OK" button. A sidebar on the right, titled "Accept prediction", contains a "Return to matrix" section with a "Select/filter data" button. Three callout boxes with numbers 1, 2, and 3 point to the "Select/filter data" button, the "OK" button, and the "Return to matrix" section, respectively.

1. Accept prediction

2. Click OK

3. Return to matrix

1. Accept prediction 2. Click OK 3. Return to matrix

Data gap filling for focused tautomer

Trend analysis

The screenshot displays the QSAR Toolbox interface during the 'Data Gap Filling' step. The 'Data Gap Filling Method' is 'Trend analysis'. The 'Target Endpoint' is 'Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata Actinopterygii Pimephales promelas'. The 'Structure' column shows the chemical structure of pime. The 'Data Matrix' table shows a prediction for the 'Pimephales p.' row, with a value of 0.551 (0.254, 1.1) circled in red. A text box at the bottom states: 'The prediction obtained from trend analysis appears on data matrix'.

Structure	1 [target,tautomer]	2	3	4	5	6	7
EC50 (1/1)							
LC10 (2/2)							
LC50							
Undefined Duration							
1 h (1/1)							
3 h (2/2)							
6 h (1/1)							
12 h (1/1)							
24 h (4/4)							
48 h (4/4)							
72 h (3/3)							
96 h							
Animalia							
Chordata (Vertebrates)							
Actinopterygii (Fish)							
Pimephales p. (67/134)	T: 0.551(0.254, 1.1)						
28 days post-hatch (2/2)							
LOEC (3/3)							
LT50 (1/3)							
NOEC (7/12)							
INR-LETH (1/1)							

M: 21.4 mg/L, 21.4...

303 Aliphatic Amines Strict (Aquatic toxicity classification by ECOSAR)

1/0/0

Data gap filling for focused tautomer Interpreting Read-across

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

Outlook

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

Handling tautomerism of target chemical

Assigning data to parent chemical

1: T. 0.551(0.254, 1.19) mg/L

2: [c1]([CCCCCCCCC=C-N])

3: [c1]([CCCCCCCCC=C-N])



4: [c1]([CCCCCCCCC=C-N])

5: T. 0.425 mg/L 0.4

TA prediction coincide with measured data

1. The trend analysis prediction appears on datamatrix; 2. The prediction of the tautomeric form is assigned to the last SMILES within the set; 3. **Click** on the first SMILES in order to go back to the set; 4. All tautomeric forms within the set are visualized on data matrix The TA prediction coincide with experimental data.; 5. **Click** on the cell related to the parent chemical

Handling tautomerism of target chemical

Assigning data to parent chemical

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling operation. The top navigation bar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling (highlighted with callout 1), and Report. The left sidebar shows the 'Data Gap Filling Method' set to 'Independent MOA' (callout 3) and the 'Target Endpoint' as 'Mortality LC50 96 h Animalia Chordata Actinopterygii Pimephales promelas'. The main workspace shows a hierarchical tree of endpoints for the chemical 'pimep'. A table below the tree displays data points for various endpoints. The cell containing the parent chemical's prediction, 'M: 0.425 mg/L, 0.43(0.4,0.4)', is circled in red and labeled with callout 2. The 'Apply' button in the top left is circled in red and labeled with callout 4.

Structure	1 [target]	2 [target,tautomer]	3 [target,tautomer]	4 [target,tautomer]
[3] [T]				
Morphology				
Mortality				
EC50				
LC10				
LC50				
1 h				
3 h				
6 h				
12 h	(1/1) M: >1.5;>2.25 mg/L			
24 h	(1/1) M: >1.5;>2.25 mg/L			
48 h	(1/1) M: >0.75;>1.5 mg/L			
72 h	(1/1) M: >0;>0.75 mg/L			
96 h				
Animalia				
Chordata (Vertebrates)				
Actinopterygii (Fish)				
Pimephales pro.	(2/3) M: 0.425 mg/L, 0.43(0.4,0.4)			T: 0.551(0.254;1.1...
28 days post-hatch				
LOEC				
LT50				
NOEC				
NR-LETH				
NR-ZERO				
No Effect Coded				

1. Go to Data Gap filling 2. **Select** the cell of the parent; The independent MOA is used to transfer the prediction to the parent chemical 3. **Select** Independent mode; 4. **Click** Apply

Handling tautomerism of target chemical

Assigning data to parent chemical

- The following actions (steps) are used for assigning data to parent chemical:
 - Accept prediction
 - Return to matrix
- Independent mode of action is formally used for transferring the value from metabolite to the target chemical.
 - Independent MOA- all components are with different mode of action
 - Similar MOA- all components are with similar mode of action. The quantities of the components are taken into account*
- Final prediction for the parent compound labeled as CI (Component based Independent mode) (see next screen shot)

*Additional information for both MOA could be found in "Tutorial 2 Prediction of Acute fish for mixtures" posted on OECD and LMC website: http://www.oecd.org/chemicalsafety/risk-assessment/Tutorial_12_TB%203.2.pdf

Handling tautomersim of target chemical

Assigning data to parent chemical

The screenshot displays the QSAR Toolbox interface. At the top, there is a navigation bar with icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below this, a sidebar on the left contains options for Data Gap Filling Method (Independent MOA, Similar MOA, Specific models) and Target Endpoint (Ecotoxicological Information Aquatic Toxicity, Mortality LC50 96 h Animalia Chordata Actinopterygii, Pimephales promelas). The main workspace shows a prediction plot for 'log Kow' (Descriptor X) versus 'LC50 (obs.), log(1 mol/L)'. An 'Information' dialog box is open, displaying the message 'The current prediction was accepted' and an 'OK' button. Three red callouts are present: '1' points to the 'Accept prediction' button in the right-hand panel; '2' points to the 'OK' button in the dialog box; and '3' points to the 'Return to matrix' button in the right-hand panel.

1. Accept prediction 2. Click OK 3. Return to matrix

Outlook

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

Report

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

Report

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes tabs for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows the 'Data Gap Filling Method' (Independent MOA, Similar MOA, Specific models) and 'Target Endpoint' (Ecotoxicological Information, Aquatic Toxicity, etc.). The main workspace shows a hierarchical tree of endpoints under 'Mortality' and 'Reproduction'. A prediction cell is highlighted with a red circle and a blue callout '1'. A right-click context menu is open over this cell, with the 'Report' option highlighted by another red circle and a blue callout '2'.

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

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

Prediction [2]

Prediction of LC50 for n-undecyl cyanide 1 / 44

1

QSAR Toolbox prediction for multicomponent substance

(uses single component mode for handling of target tautomers)

The template of the current report is based on "GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS MODELS" published by OECD (September, 2007) and "GUIDANCE ON INFORMATION REQUIREMENTS AND CHEMICAL SAFETY ASSESSMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS" published by ECHA (May, 2008).
The report provides information about the target substance, chemical characteristics used for the grouping, the resulting boundaries of the group of chemicals (applicability)

4 Document_2 3/0/0

1. TB report for multicomponent substance

Report

The screenshot shows the QSAR Toolbox software interface. The main window displays a report titled "QSAR Toolbox prediction for set of tautomers based on independent mode of action for tautomers". The report content includes:

Prediction of LC50 for n-undecyl cyanide

Summary

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

The data used for calculating the current prediction is taken from 1 Toolbox prediction.

Below is a summary table for endpoint & descriptor values for the target set of tautomers and the 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>
	Ecotoxicological Information, Aquatic Toxicity
	mg/L
<i>Target chemical & its tautomers</i>	-
<i>Tautomer No. 1</i>	0.551

The interface also shows a sidebar with "Available data to report" and "Available report templates". A callout box with the number "1" points to the "Summary" section of the report.

1. Summary information for prediction

Report

The screenshot displays the QSAR Toolbox Report interface. The top navigation bar includes options like Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The left sidebar shows available data to report and report templates, categorized into Standard (predefined) and Custom (user defined). The main content area shows a table for endpoints and detailed prediction results for Prediction [2].

Endpoint (dep. variable)	
Ecotoxicological Information; Aquatic Toxicity	-

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):
0.551 mg/L

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

4.3. Applicability domain (OECD Principle 3):
The target substance of current prediction is IN DOMAIN, because the target substances in individual predictions are in domain
Below is the list of domain classification for the individual predictions (for details see the related prediction reports)

Individual component prediction No.1:
Target substance is IN DOMAIN

4.4. Uncertainty of the prediction (OECD Principle 4): *manually editable field*
Not provided by the user

4.5. Chemical and biological mechanisms (OECD Principle 5):
Profiling results for the target substance:

Callout 1 points to the predicted value (0.551 mg/L). Callout 2 points to the applicability domain section (4.3).

1. Predicted value
2. Applicability domain

The target chemical is "In domain", because the prediction of the single tautomer is "In domain".

Report

The screenshot shows the QSAR Toolbox software interface. The main window displays a report titled "Prediction [2]". The report content is as follows:

QSAR Toolbox prediction based on trend analysis

Prediction of LC50 for CCCCCCCCCC#CN
(individual component prediction #1)

Summary

Toxicity of the target chemical (0.551 mg/L) is predicted from category members using trend analysis based on 11 values within the range 0.0657 - 267 mg/L from 11 category members. Category members are single chemicals or mixtures and are selected based on the profile of the target chemical. Only chemicals having experimental data are listed in the category.

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

The data used for calculating the current prediction is taken from 25 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 chemical 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>
	Ecotoxicological Information; Aquatic Toxicity	log Kow

At the bottom of the interface, a blue box contains the text: "1. Report for individual component".

Outlook

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

Saving the prediction result

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

Saving the prediction

The screenshot shows the QSAR Toolbox software interface. The main window displays a report titled 'Prediction on trend analysis' for the chemical 'CCCCCCC#CN'. The report includes a prediction number and a list of category members. A 'Save As' dialog box is overlaid on the report, showing the file name 'Tutorial 20' and the save type 'Toolbox work file (*.tbw)'. Three callout boxes with numbers 1, 2, and 3 point to the 'Save' button, the file name field, and the 'Save' button respectively.

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

Open saved file

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Filtering', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Input' menu is open, and the 'Open' option is highlighted. A 'Select file' dialog box is open, showing a list of files in the 'TB documents' folder. The file 'Tutorial 20.tbw' is selected. The 'Open' button in the dialog box is highlighted. A callout box at the bottom of the screenshot provides the following instructions:

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