

OECD QSAR Toolbox v.4.1

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
 - Data
 - Category Definition
 - Data Gap Filling
 - Report

Chemical Input

Ways of Entering a Chemicals

User Alternatives for input of Chemical:

A. Single target chemical

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

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input

Input Screen

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

Chemical Input

Input target chemical by CAS#

The screenshot shows the QSAR Toolbox software interface. The main window has a menu bar with options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. Below the menu bar is a toolbar with icons for New, Open, Close, Save, and CAS#. A search dialog box is open, showing the input of the CAS# 2437-25-4. The dialog box has a search field, a search button, and OK/Cancel buttons. Below the search field, there are buttons for Select All, Unselect All, and Invert Selection. The search results show a table with the following information:

1	CAS	2437-25-4
	SMILES	CCCCCCCCCCCC#N
	CS Relation	High
	Substance	Mono constituent
	Composition	
	Name	C12 nitrile Dodecanenitrile Dodecanonitrile

The chemical structure of dodecanenitrile is shown as CCCCCCCCCCCC#N. The search results are displayed in a table with a checkmark in the first column. The callouts indicate the following steps: 1. Click on CAS#; 2. Enter 2437-25-4; 3. The system identifies the structure; 4. OK.

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

Chemical Input

Target chemical identity

- Double click “Structure info” displays the chemical identification information.
- The user should note that existing names of the target chemical are presented. 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 options like 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The main window is divided into several panes. On the left, a 'Documents' pane shows 'Document 1' with CAS# 2437254. The central area is titled 'Filter endpoint tree...' and contains a 'Structure' section. Below this, the 'Structure info' section is expanded, showing various fields: 'CAS Number' (2437-25-4), 'CAS Smiles relation' (High), 'Chemical name(s)' (C12 nitrile, Dodecanenitrile), 'Composition' (C12H23N), 'Predefined substance type' (Mono constituent), and 'Structural Formula' (CCCCCCCCCCCC#N). A red circle highlights the 'CAS Number' field. The right side of the interface shows a 'Target Endpoint' section with a '1 [target]' label and a chemical structure diagram.

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 interface. On the left, the Profiling side-bar is visible. Callout 1 highlights the 'US-EPA New Chemical Categories' under the Profiling methods section. Callout 2 points to the 'View Tests' button in the top menu bar. Callout 3 points to 'Esters (Acute toxicity)' in the Categories list. The main workspace shows a 'Parametric boundaries' diagram with three nodes connected to an 'AND' node, and a 'Structural fragment' diagram below it.

1. Highlight the profiler
2. Click View
3. Select "Esters(Acute toxicity)"

Profiling Side-Bar to Profiling

The screenshot displays the QSAR Toolbox interface. On the left, the 'Profiling' sidebar shows a document with CAS: 2437254 and various profiling methods. The 'All Tests' panel on the right is active, with the 'Literature' tab selected. The 'Esters' category is expanded, showing a detailed textual description, hazard concerns, boundaries, and testing strategies. A red circle highlights the 'Literature' tab, and a blue box with the number '1' points to it. An arrow points from the text 'Textual description' to the main text area of the 'Esters' category.

Category: Esters Environmental Toxicology

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

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

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

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

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

General Testing Strategy.

I. Release to Aquatic Ecosystems:

Tier 1. The aquatic base set of environmental toxicity tests will be recommended for aquatic exposures. The acute toxicity tests for fish (40 CFR §797.1400) and daphnids (40 CFR §797.1300) will be done using the flow-through method with measured concentrations; effective concentrations will be based on 100% active ingredients (AI) and mean measured concentrations; the highest treatment concentration on a nominal-basis should not exceed the aqueous solubility limit; and solvent can be used to assist the ester to reach its aqueous solubility limit quicker, but cannot be used to artificially enhance the water solubility of the ester beyond its aqueous solubility limit.

The algal toxicity test (40 CFR §797.1050) should be done with the static method; measured concentrations; effective concentrations based on 100% active ingredients (AI) and mean measured concentrations; statistical analysis of effective concentrations at 24, 48, 72, and 96 hours; test medium with at least 0.300 mg/L EDTA as a final concentration; the highest treatment concentration on a nominal-basis should not exceed the aqueous solubility limit of the ester; and solvent can be used to assist the ester to reach its aqueous solubility limit quicker, but cannot be used to artificially enhance the water solubility of the ester beyond its aqueous solubility limit.

If there is no significant risk from the ester after the results of the environmental base set have been integrated into the risk assessment, then no further testing is recommended. However, if there

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

Profiling

Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues (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 (Modified)
 - Protein binding alerts by OASIS
 - 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

The screenshot shows the QSAR Toolbox Profiling interface. The 'Profiling methods' list on the left includes the following checked items:

- Acute aquatic toxicity MOA by OASIS
- Acute toxicity classification by ECOSA
- Chemical elements

The 'Apply' button in the top left is circled in red. A callout box with the number '2' points to the 'Apply' button, and another callout box with the number '1' points to the checked items in the list.

The right side of the interface shows a 'Filter endpoint tree...' and a 'Structure' panel with a chemical structure and associated data:

Structure	1 [target]
Structure info	
CAS Number	2437-25-4
CAS Smiles relation	High
Chemical name(s)	C12 nitrile
Composition	
Molecular Formula	C12H23N
Predefined substance type	Mono constituent
Structural Formula	CCCCCCCCCCC#N
Parameters	
Substance identity	
Physical Chemical Properties	
Environmental Fate and Transport	
Ecotoxicological Information	
Human Health Hazards	
Profile	

1. **Check** profilers mentioned on #20
2. **Click** Apply

Profiling

Profiling the target chemical

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

(see next screenshot)

Profiling

Profiling the target chemical

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

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

Outlook

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

Data Overview

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

Data

Case study

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

Data Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data' menu is open, showing options like 'Gather', 'Import', 'IUCLID', and 'IUCLID6'. The 'Gather' button is circled in red and labeled with a '4'. In the 'Databases' section, the 'Ecotoxicological Information' folder is expanded, and several sub-items are checked, including 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', 'ECHA CHEM', and 'ECOTOX'. This section is circled in red and labeled with a '2'. The 'Human Health Hazards' folder is also expanded and labeled with a '3'. The 'Filter endpoint tree...' panel shows a single target endpoint with a chemical structure. A blue box at the bottom contains the following instructions:

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

Data

Gather data

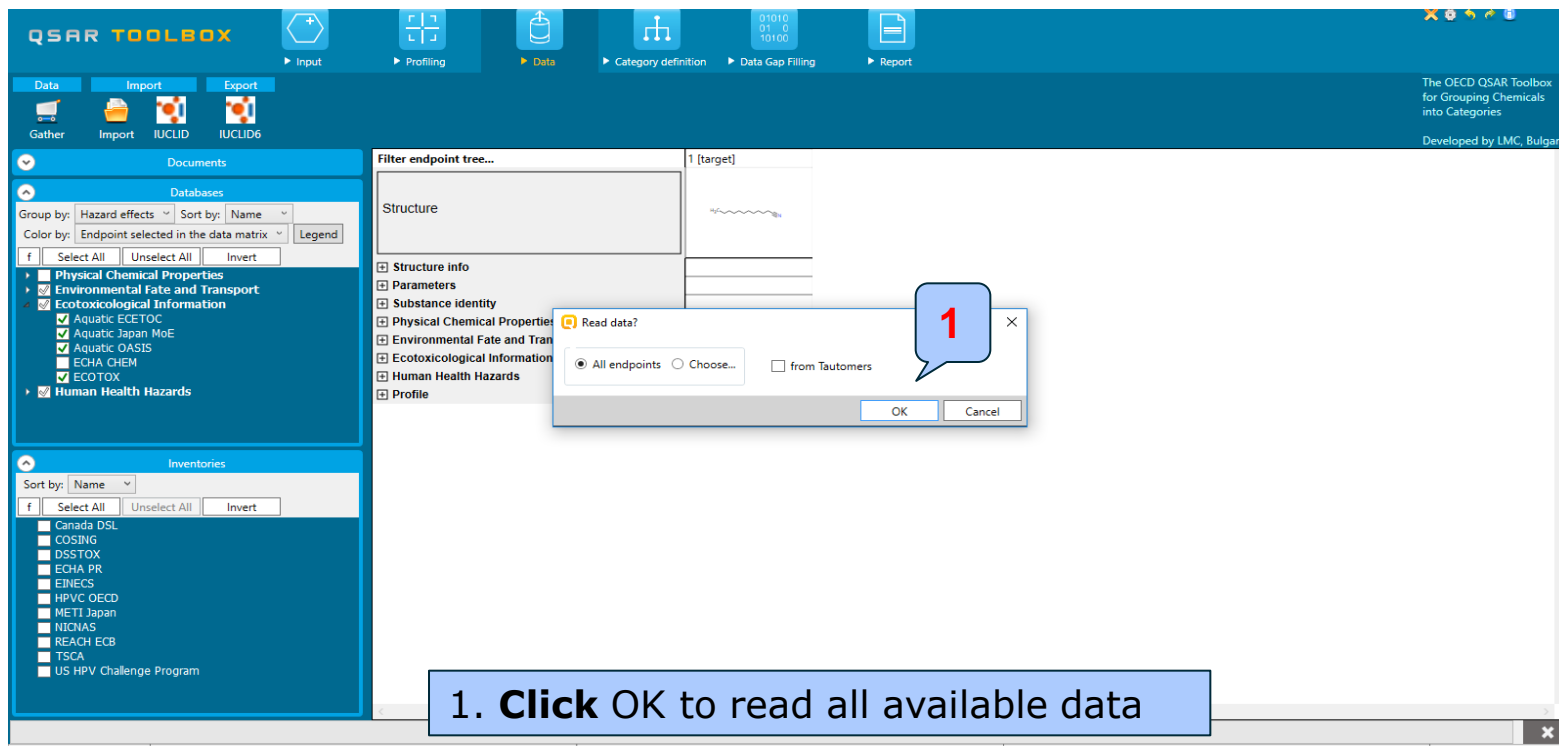
- Toxicity information on the target chemical is electronically collected from the selected dataset(s)
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are Aquatic toxicity OASIS; 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/unece/search?q=revision4>

Data

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.



Data

Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Data', 'Import', and 'Export'. Below this, there are icons for 'Gather', 'Import', 'IUCLID', and 'IUCLID6'. The main workspace is divided into several panels:

- Documents:** Shows a list of documents, with 'pime' selected.
- Databases:** A tree view showing various hazard categories. Under 'Ecotoxicological Information', several sub-categories are checked, including 'Aquatic ECETOC', 'Aquatic Japan MoE', 'Aquatic OASIS', 'ECHA CHEM', and 'ECOTOX'.
- Inventories:** A list of chemical inventories such as 'Canada DSL', 'COSING', 'DSSTOX', etc.
- Structure:** Displays the chemical structure of 'pime' (Pimephales promelas), highlighted with a blue box and the number '1'.
- Filter Tree:** A hierarchical tree view showing the selection of 'pime' and its associated endpoints. The tree is filtered to show data for 'Pimephales promelas'.
- Datamatrix:** A table showing experimental data for the selected endpoint. The table has columns for 'Mortality', 'LC50', and 'M'. The data for 'Pimephales promelas' is highlighted in blue.

Endpoint	Mortality	LC50	M
Behavior	EC50		
	12 h	(1/1)	M: >1.5+2.25 mg/L
	24 h	(1/1)	M: >1.5+2.25 mg/L
	48 h	(1/1)	M: >0.75+1.5 mg/L
	72 h	(1/1)	M: >0+0.75 mg/L
	96 h	(1/1)	M: 0.42 mg/L
Mortality	LC50		
	12 h	(1/1)	M: >1.5+2.25 mg/L
	24 h	(1/1)	M: >1.5+2.25 mg/L
	48 h	(1/1)	M: >0.75+1.5 mg/L
	72 h	(1/1)	M: >0+0.75 mg/L
	96 h	(1/1)	M: >0+0.75 mg/L
Actinopterygii (ray-finned fishes, s...)			
	Pimephales promelas	(1/2)	M: 0.43 (0.4+0.47) mg/L M: 0.425 mg/L

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)

Data Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data' menu is open, showing options like 'Gather', 'Import', 'IUCLID', and 'IUCLID6'. The main workspace displays a hierarchical tree of 'Ecotoxicological Information' with a table of data points. A 'Data points' dialog box is open in the foreground, showing a table with columns: Datapoints, #, Value, Original value, Assigned SMILES, and Source. A blue callout box with the number '1' points to the 'Data' menu item, and another blue callout box with the number '2' points to the 'X' close button on the dialog box.

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

2. Click on the X to close the window

Datapoints	#	Value	Original value	Assigned SMILES	Source
Ecotoxicological Information;Aquatic Toxicity	1	M: 0.43 (0.4+0.47) mg/L (Mass concentration)	0.43 (0.4+0.47) mg/L (Mass concentration)	True	Brooke,L.T., C.E. Northcote
Ecotoxicological Information;Aquatic Toxicity	2	M: 0.425 mg/L (Mass concentration)	2.34E-06 mol/L (Molar concentration)	False	Russom, C., Broderius, J., Hammer

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
 - Data
 - **Handling of tautomerism of target chemical**

Handling of tautomerism of target chemical

The screenshot shows the QSAR Toolbox interface. In the 'Filter endpoint tree...' panel, a red box highlights the 'target' chemical structure. A red arrow points from this box to a table of three tautomeric forms. A red bracket underlines these forms with the text 'Tautomeric forms'. Three numbered callouts (1, 2, 3) indicate the steps: 1. Clicking the 'Input' icon, 2. Right-clicking the target node and selecting 'Multiplication' > 'Tautomerism', and 3. The resulting tautomeric forms.

tautomer #1	tautomer #2	tautomer #3
<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>

1. **Go** to Input
2. **Right click** over the node with SMILES and **select** Multiplication and then Tautomerism
3. Three tautomeric forms are generated for the target chemical

Outlook

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

Handling of tautomerism of target chemical

Profiling set of tautomers

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

Handling of tautomerism of target chemical

Profiling set of tautomers

- The following primary profilers relevant to the **aquatic toxicity** are used in this example(see next screenshot):
 - 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
 - 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

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
 - Data
 - Handling of tautomerism of target chemical
 - Profiling set of tautomers
 - **Focus active tautomer**

Handling of tautomerism of target chemical

Focus of active tautomer

This tautomeric form is selected for further trend analysis

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

1

2

1. Right click over the active tautomeric form

2. Focus the chemical

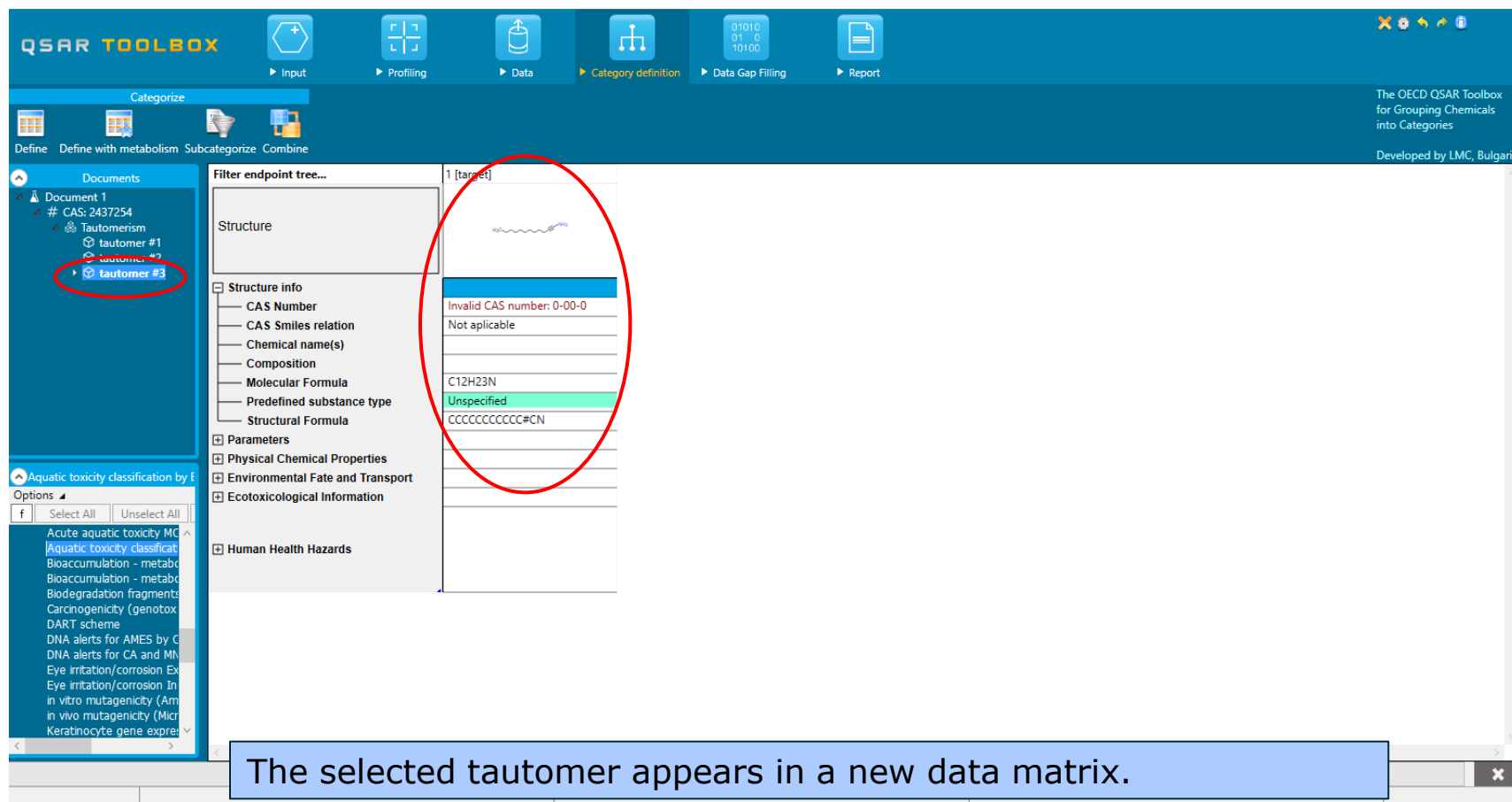
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Documents

- Document 1
 - # CAS: 2437254
 - Tautomerism
 - tautomer #1
 - tautomer #2
 - tautomer #3

Handling of tautomerism of target chemical

Focus of active tautomer



The screenshot displays the QSAR Toolbox interface. On the left, the 'Documents' panel shows a tree view where 'tautomer #3' is selected and circled in red. The 'Filter endpoint tree...' panel in the center shows a chemical structure and a table of properties. The table has the following rows:

Invalid CAS number: 0-00-0
Not applicable
C12H23N
Unspecified
CCCCCCCC#CN

A blue callout box at the bottom of the interface states: "The selected tautomer appears in a new data matrix."

Outlook

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

Category Definition

Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across/trend analysis.
- 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. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, there are buttons for Gather, Import, and IUCLID6. The main workspace is divided into several panels:

- Documents:** Shows a tree view for 'Document 1' containing CAS: 2437254 and three tautomers (tautomer #1, #2, #3).
- Databases:** A list of toxicity databases with checkboxes. Selected databases include:
 - Aquatic ECETOC
 - Aquatic Japan MoE
 - Aquatic OASIS
 - ECHA CHEM
 - ECOTOX
 - Human Health Hazards
 - Acute Oral toxicity
- Inventories:** A list of regulatory inventories with checkboxes. Selected inventories include:
 - Canada DSL
 - COSING
 - DSSTOX
 - ECHA PR
 - EINECS
 - HPVC OECD
 - METI Japan
 - US EPA
- Filter endpoint tree...:** A tree view showing the structure of the filter endpoint tree. The 'Structure' node is expanded, showing a chemical structure. Other nodes include Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information (with sub-nodes for Aquatic, Sediment, and Terrestrial Toxicity), Human Health Hazards, and Profile.

The interface also includes a status bar at the bottom and a window title bar at the top right.

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

The screenshot illustrates the steps to define an ECOSAR category in the QSAR Toolbox. The 'Documents' panel shows a tree view where 'Tautomerism' is expanded. The 'Filter endpoint tree...' dialog box is open, showing 'Aliphatic Amines' as the target. The 'Profiles' list includes 'Acute aquatic toxicity classification by ECOSAR', which is highlighted. The 'Combine profiles' section shows 'Strict' selected. The 'OK' button is circled. Numbered callouts (1-4) indicate the steps: 1. Highlight 'Acute aquatic toxicity classification by ECOSAR', 2. Click 'Define', 3. Select 'Strict', and 4. Click 'OK'.

1. Highlight "Acute 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 software interface during the 'Category definition' step. A 'Grouping results' dialog box is open, showing '371 chemicals found.' and an 'OK' button. A blue callout box with the number '1' points to the 'OK' button. The background shows a 'Filter endpoint tree...' window with a 'Structure' tab selected, and a list of chemical structures in a table. The left sidebar shows a document tree with 'Aquatic toxicity classification by ECOSAR' selected.

1. **Click** OK to confirm the name of the category

Handling of tautomerism of target chemical

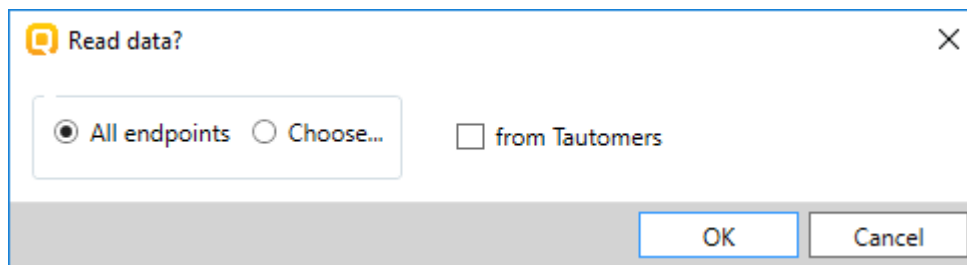
Category analogues

- The Toolbox now identifies all chemicals corresponding to *Aliphatic amines* by ECOSAR listed in the four aquatic databases.
- 371 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.

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, a 'Categorize' section offers options like 'Define', 'Define with metabolism', 'Subcategorize', and 'Combine'. The main workspace is divided into a 'Documents' panel on the left, a 'Filter endpoint tree...' panel, and a grid of chemical structures. A 'Gather data' dialog box is open in the center, displaying the message '9414 points added across 369 chemicals.' and an 'OK' button. A blue callout box with the number '1' points to the dialog box. The bottom-left corner of the interface shows a list of classification options under 'Aquatic toxicity classification by ECOSAR', including 'Acute aquatic toxicity classification by Verhaar' and 'Acute aquatic toxicity MOA by OASIS'.

Handling of tautomerism of target chemical

Summary information for Analogues

The screenshot displays the QSAR Toolbox interface. The main window shows a datamatrix for the chemical 'pime' (CAS: 2437254) and its analogues. The datamatrix columns represent different chemical groups (1-7), and the rows represent various toxicity endpoints. A red box highlights the '96 h' row, which contains data for 'Actinopterygii (ray-finned fishes, spi...)' and 'Pimephales promelas'.

Endpoint	1 [target]	2	3	4	5	6	7
96 h							
Actinopterygii (ray-finned fishes, spi...)							
Pimephales promelas							
26 h							
LOEC							
LT50							
NOEC							
NR-LETH							
NR-ZERO							
Undefined Endpoint							
No Effect Coded							
Reproduction							

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 369 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 icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report.
- Left Panel:** Documents list showing 'Document 1' with CAS: 2437254 and tautomers. Below it, 'Data Gap Filling Settings' are visible.
- Main Area:** A table with columns for chemical structures and various endpoints. A search filter 'pime' is applied. The tree is expanded to 'Pimephales promelas' (6/145), which is highlighted with a red box.
- Annotations:** A blue callout '1' points to the search input area, and another blue callout '2' points to the tree expansion controls.

Endpoint	Count	Value
EC50	(6/15)	
LC10	(8/19)	
LC50		M: 0.0129 (0.0055+0)
1 h	(2/5)	
3 h	(4/30)	
6 h	(4/36)	
12 h	(4/7)	
24 h	(58/219)	M: 0.1 (0.09+0.11) m
48 h	(72/259)	M: 0.051 (0.046+0.057) m
72 h	(16/21)	M: 0.69 (0.61+0.81) mg/L
28 d	(3/6)	
LOEC	(19/44)	M: > 10 ppm
LT50	(3/11)	
NOEC	(24/67)	M: 0.000287 mg/L
NR-LETH	(41/90)	M: 0.074 mg/L
NR-ZERO	(42/88)	M: 0.096 mg/L
Undefined Endpoint	(86/282)	M: 0.0001+0.1 mg/L
No Effect Coded	(39/113)	M: > 10.1 ppm
Reproduction	(54/247)	M: > 10 ppm

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

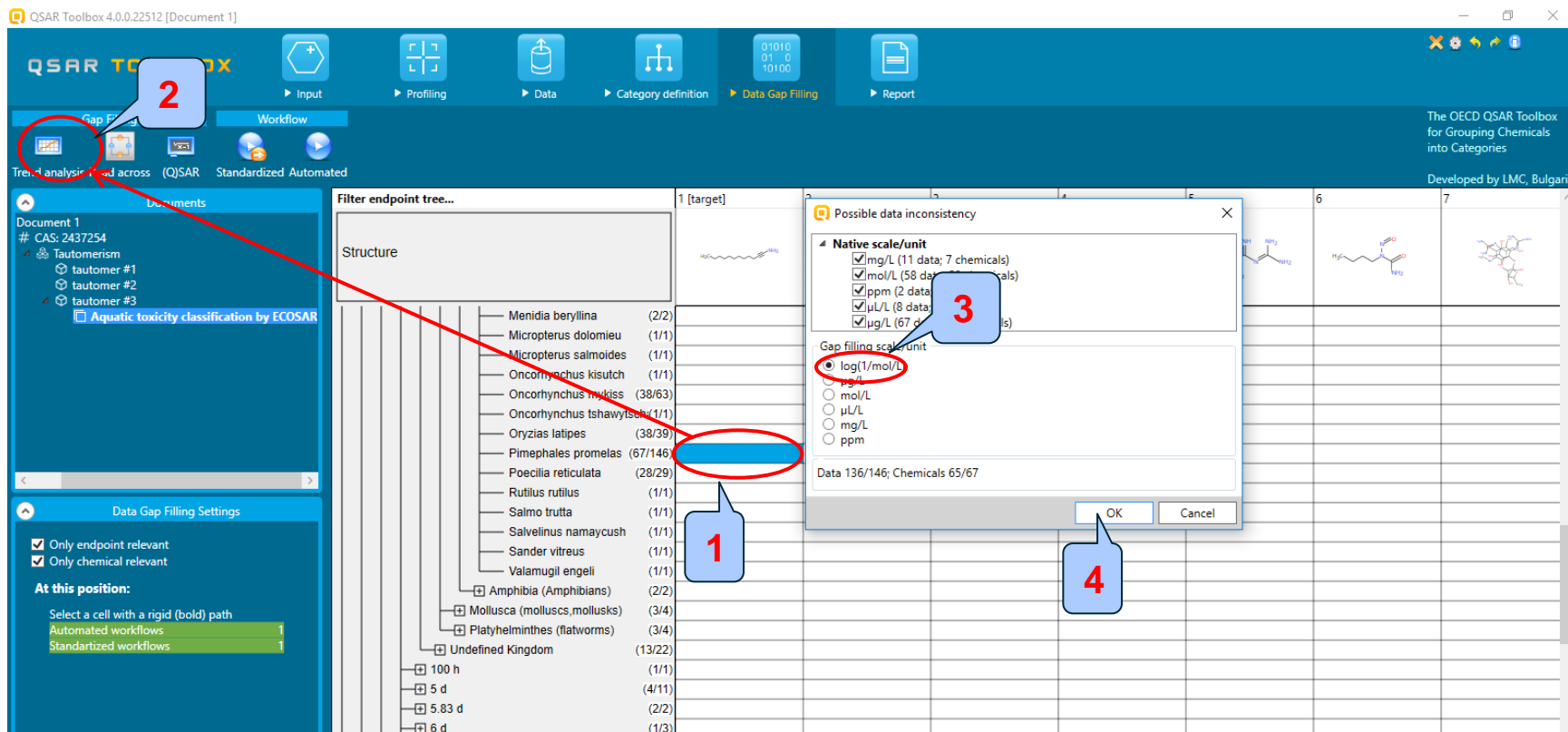
- You have now retrieved the available experimental data on aquatic toxicity for 369 analogue chemicals of focused tautomeric form classified as “Aliphatic amines” by the “ECOSAR” profiler.
- You have identified the target endpoint of “96 h LC50 Mortality for *Pimephales promelas*”.
- You are ready to fill in the data gap so click on “Data Gap Filling” (see next screen shots).

Outlook

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

Data Gap Filling

Apply Trend analysis



1. **Highlight** the endpoint box corresponding to *Pimephales promelas*/LC50/96h under the target chemical.
2. **Select** Trend analysis
3. **Select** scale – log(1 mol/l)
4. **OK**

Data Gap Filling

Results of Trend analysis

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (QSAR) Standardized Automated

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

Documents

54 erism omer #1 omer #2 omer #3 Aquatic toxicity classification by ECOSAR Enter GF(TA) with 66 chemicals, 136 data points

Filter endpoint tree...

Structure

1 [target] 2 4 10 32 35 36

NR-LETH 3.5 h Animalia (animals) Chordata (chordates) Actinopterygii (ray-finned fishes) Pimephales promelas (1/1) NR-ZERO

Descriptors Prediction Adequacy Cumulative frequency Residuals Statistics

Trend analysis prediction for LC50, based on 65 values
Predicted: 2.60 mg/L
Model equation: $LC50 = 2.36 (\pm 0.170) + 0.613 (\pm 0.0754) * \log Kow, \log(1/mol/L)$

LC50 [log(1/mol/L)]

log Kow

Select / filter data
Gap filling approach
Descriptors / data
Model/QSAR
Calculation options
Visual options
Information
Miscellaneous

Accept prediction

66

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 1 by Chemical elements

The screenshot displays the QSAR Toolbox interface during a subcategorization step. The 'Subcategorization' dialog is open, showing the 'Options' pane where 'Chemical elements' is selected under the 'Empirical' category. The 'Adjust options' pane shows a list of chemical groups, with '(2) Group 17 - Halogens Cl' selected. The 'Select / filter data' panel on the right has the 'Subcategorize' button circled in red. The background shows a workflow grid with chemical structures and a log Kow plot at the bottom.

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

Data Gap Filling

Result of Subcategorisation 1 by Chemical elements

QSAR TOOLBOX

Input | Profiling | Data | Category definition | Data Gap Filling | Report

Gap Filling | Workflow

Trend analysis Read across (QSAR) Standardized Automated

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

Documents

254
merism
tomer #1
tomer #2
tomer #3
Aquatic toxicity classification by ECOSAR
Enter GF(TA) with 65 chemicals, 136 data points
Ch: 39| Data: 81 Subcategorized: Chemical el

Filter endpoint tree...

Structure

- NR-ZERO (7/8)
- Undefined Endpoint (8/22)
- No Effect Coded (2/6)
- Reproduction (2/5)
- Profile
 - Empiric
 - Chemical elements

1 [target]	2	4	10	35	36	37
		M: 10 mg/L		M: 10 mg/L		
Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C

Endpoint: Chemical elements

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Trend analysis prediction for LC50, based on 38 values
Predicted: 2.02 mg/L
Model equation: $LC50 = 2.24 (\pm 0.230) + 0.670 (\pm 0.0988) * \log Kow, \log(1/\text{mol/L})$

Select / filter data

- Subcategorize
- Mark by WS
- Mark chemicals by descriptor value
- Mark outliers
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked data

Accept prediction

Data Gap Filling

Subcategorisation 2 by OFG (nested)

The screenshot displays the QSAR Toolbox interface during a 'Data Gap Filling' operation. The 'Subcategorization' dialog is active, showing a list of functional groups. 'Organic functional groups (nested)' is highlighted with a red circle and labeled '2'. Below the dialog, a 'Trend analysis prediction for LC50' plot is shown, with a red regression line and the equation: $LC50 = 2.24 (\pm 0.230) + 0.670 (\pm 0.0988) * \log Kow, \log(1/mol/L)$. The predicted LC50 is 2.02 mg/L. A 'Select / filter data' panel on the right has 'Subcategorize' highlighted with a red circle and labeled '1'. A third callout '3' points to the 'Remove selected' button in the dialog.

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 Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Documents

1
2
3

toxicity classification by ECOSAR
or GF(TA) with 65 chemicals, 136 data points
Ch: 39] Data: 81 Subcategorized: Chemical elements
Ch: 15] Data: 33 Subcategorized: Organic function

Filter endpoint tree...

Structure	1 [target]	10	78	90	125	131	173
Environmental Fate and Transport							
Bioaccumulation: aquatic							
Ecotoxicological Information							
Aquatic Toxicity							
Behavior (5/27)							
Biochemistry (1/4)							
Growth (10/10)							

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Trend analysis prediction for LC50, based on 14 values
Predicted: 0.854 mg/L
Model equation: $LC50 = 2.46 (\pm 0.442) + 0.709 (\pm 0.153) * \log Kow, \log(1/\text{mol/L})$

Select / filter data

- Subcategorize
- Mark by WS
- Mark chemicals by descriptor value
- Mark outliers
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked data

Accept prediction

15

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

Subcategorization

Group by: Category Sort by: Name

Color by: Target endpoint Legend

Adjust options

[90%,100%]

Differ from target by

At least one category

All categories

Groups (nested)

Groups (nested)

Do not account metabolism

Documented

- Observed Mammalian metabolism (1) [30%,40%]
- Observed Microbial metabolism (1) [40%,50%]
- Observed Rat In vivo metabolism (2) [50%,60%]
- Observed rat liver metabolism with quant (4) [60%,70%]
- Observed Rat Liver S9 metabolism (2) [70%,80%]

Simulated

- Autoxidation simulator (0/14)
- Autoxidation simulator (alkaline medium)
- Dissociation simulator
- Hydrolysis simulator (acidic)

Remove selected

Trend analysis prediction for LC50, based on 14 values

Predicted: 0.854 mg/L

Model equation: $LC50 = 2.46 (\pm 0.442) + 0.709 (\pm 0.153) * \log Kow, \log(1/\text{mol/L})$

Select / filter data

- Subcategorize
- Mark by WS
- Mark chemicals by descriptor value
- Mark outliers
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked data

Accept prediction

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

Data Gap Filling Result

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (QSAR) Standardized Automated

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

Documents

icity classification by ECOSAR
F(TA) with 65 chemicals, 136 data points
39] Data: 81 Subcategorized: Chemical elements
Ch: 15] Data: 33 Subcategorized: Organic functional group
Ch: 11] Data: 23 Subcategorized: Structure similar

Filter endpoint tree...

Structure

Environmental Fate and Transport

- Bioaccumulation: aquatic
- Ecotoxicological Information
 - Aquatic Toxicity
 - Behavior (4/21)
 - Growth (8/8)
 - Mortality

AW SW

1 [target]	90	125	131	173	176	319
<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>	<chem>CCCCCCCCCCCCCCCCN</chem>
		M: 10 mg/L		M: 2.86 mg/L	M: 31.7 mg/L	M: 263 mg/L

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Data Gap Filling Settings

Only endpoint relevant

Only chemical relevant

At this position:

QSARs: 0

Automated workflows: 0

Standardized workflows: 0

In nodes below:

QSARs: 0

Predicted value: 0.57 mg/l

Trend analysis prediction for LC50, based on 10 values
Predicted: 0.570 mg/L
 Model equation: $LC50 = 1.53 (\pm 0.337) + 0.982 (\pm 0.0686) * \log Kow, \log(1/mol/L)$

Select / filter data

Subcategorize

Mark by WS

Mark chemicals by descriptor value

Mark outliers

Filter points by test conditions

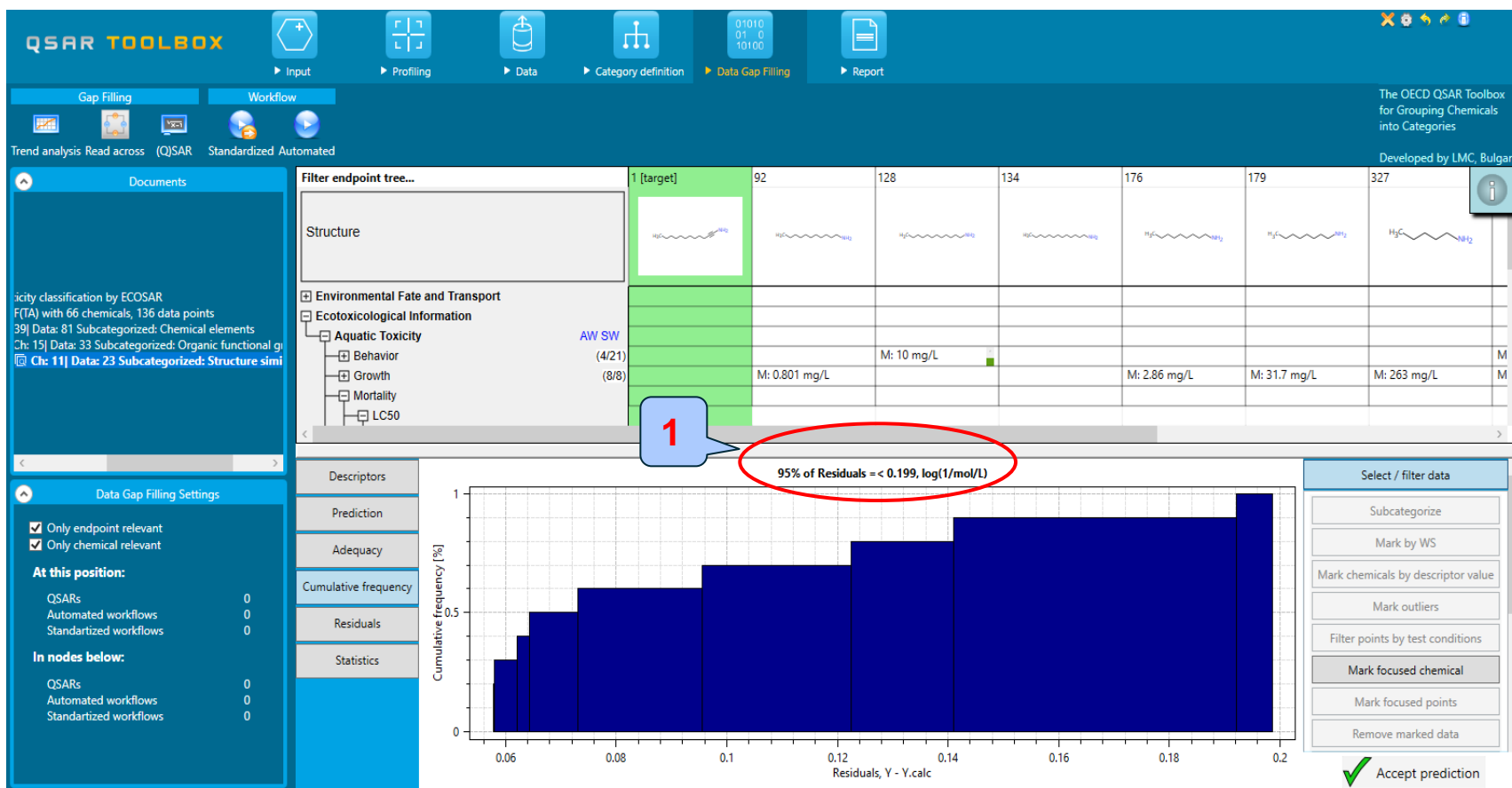
Mark focused chemical

Mark focused points

Remove marked data

Accept prediction

Data Gap Filling Cumulated frequency



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

Data Gap Filling Statistics

The screenshot displays the QSAR Toolbox interface during the Data Gap Filling process. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The main workspace shows a 'Filter endpoint tree...' on the left, a table of data points, and a 'Data Gap Filling Settings' panel. The 'Statistical characteristics' table is as follows:

Statistical characteristics	TA model
Number of data points, (N)	10
Coefficient of determination, (R2)	0.993
Adjusted coefficient of determination, (R2adj)	0.992
Coefficient of determination - leave one out, (Q2)	0.987
Sum of squared residuals, (SSR)	0.140
Standard deviation of residuals, (sN)	0.118
Sample standard deviation of residuals, (s)	0.132
Fisher function, (F)	1.09E3
Fisher threshold for statistical significance, (Fa)	8.09 (95.0%)
b0	
- model descriptor	Intercept
- coeff. value	1.53

The 'Gap filling approach' panel shows the following options:

- Remove marked data
- Clear existing marks
- Gap filling approach: Descriptors / data
- Model/QSAR (circled in red)
- Show domain
- Save model
- Save domain as category
- Calculate Q2 (circled in red)
- Accept prediction (circled in red with a green checkmark)

1. Select Model QSAR
 2. Calculate Q2
 3. The high R2 and Q2 support the reliability of the prediction
 4. Accept prediction

Data Gap Filling

Result of trend analysis

- The analysis of trend analysis shows:
 - The predicted acute aquatic toxicity value is 0.57 mg/l
 - The remaining analogues form robust category of structurally similar analogues (aliphatic amines)
 - The 95% of residuals are in the range of experimental error
 - The high R² and Q² coefficient values support the reliability of the prediction

Data gap filling for focused tautomer

Trend analysis

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. The left sidebar shows a document tree for 'Document 1' with CAS: 2437254, highlighting 'Tautomerism' and its three tautomers. The middle section shows a 'Filter endpoint tree...' with categories like Physical Chemical Properties, Environmental Fate and Transport, and Ecotoxicological Information. The right section is a data matrix table with columns for 'Parent chemical [target]', 'tautomer #1', 'tautomer #2', and 'tautomer #3'. The table contains data for various endpoints, including Aquatic Toxicity (AW SW) and Sediment toxicity. A red circle highlights a predicted value in the 'tautomer #3' column for the 'Actinopterygii (ray-finned fishes.spiny rayed fishes)' endpoint: **T: 0.57 (0.268+1.21) mg**.

The prediction obtained from trend analysis appears on data matrix

Data gap filling for focused tautomer Interpreting Read-across

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

Outlook

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

Handling tautomerism of target chemical

Assigning data to parent chemical

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

- Menu Bar:** Document, Single Chemical, Chemical List, Search, Target Endpoint.
- Toolbar:** New, Open, Close, Save, CAS#, Name, Structure, Composition, Select, Delete, ChemIDs, Database, Inventory, List, Substructure (SMARTS), Query, Define.
- Documents Panel:** Document 1, CAS: 2437254, Tautomerism, tautomer #1, tautomer #2, tautomer #3, Aquatic toxicity classification by ECOSAR, Enter GF(TA) with 66 chemicals, 136 data points, Ch: 39| Data: 81 Subcategorized: Chemical element, Ch: 15| Data: 33 Subcategorized: Organic function, Ch: 11| Data: 23 Subcategorized: Structure.
- Filter endpoint tree:** Structure, Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, Aquatic Toxicity (Behavior, Growth, Growth Inhibition, Immobilisation, Mortality, LC50), Animals (Chordata, Actinopterygii, Oryzias latipes, Pimephales), Sediment toxicity, Terrestrial Toxicity.
- Datamatrix Table:**

Parent chemical [target]	tautomer #1	tautomer #2	tautomer #3
<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
(2/10) M: >1.5+2.25 mg/L	M: >1.5+2.25 mg/L		
(2/2) M: 2.28 mg/L	M: 2.28 mg/L		
(2/4) M: 0.054 mg/L	M: 0.054 mg/L		
(2/2) M: 0.059 mg/L	M: 0.059 mg/L		
(2/2) M: >1.5+2.25 mg/L			
(2/2) M: >1.5+2.25 mg/L			
(2/2) M: >0.75+1.5 mg/L			
(2/2) M: >0+0.75 mg/L			
(2/2) M: 0.84 mg/L	M: 0.84 mg/L		
(2/5) M: 0.43 (0.4+0.47) mg/L	M: 0.425 mg/L		T: 0.57 (0.268+1.21) mg

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

Handling tautomerism of target chemical

Assigning data to parent chemical

- 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 shows the QSAR Toolbox interface with the 'Possible data inconsistency' dialog box open. The dialog has two main sections: 'Native scale/unit' and 'Gap filling scale/unit'. In the 'Native scale/unit' section, three options are checked: 'log(1/mol/L) (1 data; 1 chemicals)', 'mg/L (1 data; 1 chemicals)', and 'mol/L (1 data; 1 chemicals)'. In the 'Gap filling scale/unit' section, the 'log(1/mol/L)' radio button is selected. The dialog also displays 'Data 3/3; Chemicals 2/2' and 'OK' and 'Cancel' buttons. In the background, a data table is visible with the following content:

Chemical	Data	M	T
Oryzias la...	(1/1)	M: 0.84 mg/L	
Pimephal...	(2/3)	M: 0.425 mg/L	T: 0.57 (0.268+)

1. Select parent; 2. **Independent MOA**; 3. Use Scale/unit (log(1/mol/L)); 4. **Click OK.**

Handling tautomersim of target chemical

Assigning data to parent chemical

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (QSAR) Standardized Automated

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

Documents

ent 1
:2437254
Tautomerism
tautomer #1 (target)
tautomer #2
tautomer #3
Aquatic toxicity classification by ECOSAR
Enter GF(TA) with 66 chemicals, 136 data point
Ch: 39| Data: 81 Subcategorized: Chemical
Ch: 15| Data: 33 Subcategorized: Orgar
Ch: 11| Data: 23 Subcategorized: St
Enter GF(IndependentMOA) with 3 chemicals, 3 d

Filter endpoint tree...

Structure

1 [target] 2 4

Pimephales promelas (2/3) M: 0.425 mg/L T: 0.57 (0.268+1.21) mg

Sediment toxicity
Terrestrial Toxicity
Human Health Hazards
Profile
Empiric
Chemical elements
Group 14 - Carbon C

Descriptors
Prediction

Data Gap Filling Settings

Only endpoint relevant
 Only chemical relevant

At this position:
Select a cell with a rigid (bold) path
Automated workflows
Standartized workflows

Empirical calculation of LC50, based on 3 values
Predicted: 0.494 mg/L

LC50 [log(1/mg/L)]

log Kow

Active descriptor X log Kow

Select / filter data
Descriptors / data
Calculation options
Visual options
Information
Miscellaneous

1

Accept prediction

1. **Accept prediction**

Handling tautomersim of target chemical

Assigning data to parent chemical

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling

dependent MOA Similar MOA

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

Documents

Document 1
ID: 2437254

Tautomerism

- tautomer #1 (target)
- tautomer #2
- tautomer #3
- Aquatic toxicity classification by ECOSAR
 - Enter GF(TA) with 66 chemicals, 136 data points
 - Ch: 39 | Data: 81 Subcategorized: Chem
 - Ch: 15 | Data: 33 Subcategorized: Or
 - Ch: 11 | Data: 23 Subcategorized: Or
 - Enter GF(IndependentMOA) with 3 chemicals, 3 data points

Data Gap Filling Settings

- Only endpoint relevant
- Only chemical relevant

At this position:

- Select a cell with a rigid (bold) path
- Automated workflows 1
- Standardized workflows 1

Filter endpoint tree...

Structure

Parent chemical [target]	tautomer #1 (target)	tautomer #2	tautomer #3
Aquatic Toxicity			
AW SW			
Behavior (1/5)	M: >0.75+1.5 mg/L		
Growth (1/1)	M: 2.28 mg/L		
Growth Inhibition (1/2)	M: 0.054 mg/L		
Immobilisation (1/1)	M: 0.059 mg/L		
Mortality			
LC50			
12 h (1/1)	M: >1.5+2.25 mg/L		
24 h (1/1)	M: >1.5+2.25 mg/L		
48 h (1/1)	M: >0.75+1.5 mg/L		
72 h (1/1)	M: >0+0.75 mg/L		
96 h			
Animalia (animals)			
Chordata (chordata)			
Actinopterygii			
Oryzias latipes (1/1)	M: 0.84 mg/L		
Pimephales promelas (3/4)	M: 0.425 mg/L		T: 0.57 (0.268-1.21) mg/L
Sediment toxicity			
Terrestrial Toxicity			

1

1. Click on Prediction

Outlook

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

Report

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

Report

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data G, and Report. A callout box with the number '1' points to the 'Report' icon. The 'Documents' sidebar on the left shows a tree view of the project structure, including 'Tautomerism' and 'Aquatic toxicity classification by ECOSAR'. A callout box with the number '3' points to the 'Prediction' icon. The main area is divided into a 'Filter endpoint tree...' on the left and a data table on the right. The table has columns for 'Parent chemical [target]', 'tautomer #1 (target)', 'tautomer #2', and 'tautomer #3'. The 'IMOA: 0.494 mg/L' value in the table is circled in red, with a callout box containing the number '2' pointing to it.

Filter endpoint tree...	Parent chemical [target]	tautomer #1 (target)	tautomer #2	tautomer #3
Structure				
Structure info				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport				
Ecotoxicological Information				
Aquatic Toxicity				
Behavior	(1/5)	M: >0.75+1.5 mg/L		
Growth	(1/1)	M: 2.28 mg/L		
Growth Inhibition	(1/2)	M: 0.054 mg/L		
Immobilisation	(1/1)	M: 0.059 mg/L		
Mortality				
LC50				
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: >0.75+1.5 mg/L		
96 h	(1/1)	M: >0+0.75 mg/L		
Animalia (animals)				
Chordata (chordates)				
Actinopterygii (ray-finned fishes)				
Oryzias latipes	(1/1)	M: 0.84 mg/L		
Pimephales promelas	(3/1)	IMOA: 0.494 mg/L M: 0.425 mg/L M: 0.43 (0.4+0.47) m		T: 0.57 (0.268+1.21) mg/L
Sediment toxicity				

1. Click on section **Report**
2. Select **Prediction**
3. Create prediction report and

Report

1. TB report for multicomponent substance

Prediction report4.pdf x +
file:///D:/TB%204.0/test4_1/Prediction%20report4.pdf

Prediction of LC50 for set of tautomers 1 / 7

QSAR Toolbox prediction for multicomponent substance

Based on observed and predicted data for tautomers

Date: 8 Jul 2017
Author(s):
Contact details:

Target information		
Structural information	Numerical identifiers	Chemical names
SMILES: CCCCCCCCCCC#N	EC#: N/A CAS#: 2437-25-4 Other: N/A	C12 nitrile Dodecanenitrile Dodecanonitrile
Structure		

Do more with Microsoft Edge – the fast, new browser built for Windows 10. Change my default Don't ask again X

Back Next Cancel Create report

Human Health Hazards

Report

The screenshot displays the QSAR Toolbox interface with a prediction report open. The report contains the following information:

- SMILES:** CCCCCCCCCC#N
- Structure:** CCCCCCCCC#N
- EC#:** N/A
- CAS#:** 2437-25-4
- Other:** N/A
- C12 nitrile:** Dodecanenitrile, Dodecanonitrile

The **Prediction summary** section includes:

- Predicted endpoint:** LC50; Mortality; Pimephales promelas; 96h; No guideline specified
- Predicted value:** 0.494 (highlighted with a red circle and callout '1')
- Unit/scale:** mg/L
- Data gap filling method:** Independent mode of action
- Summary:** manually editable field
- Not provided by the user:**

1. Predicted value

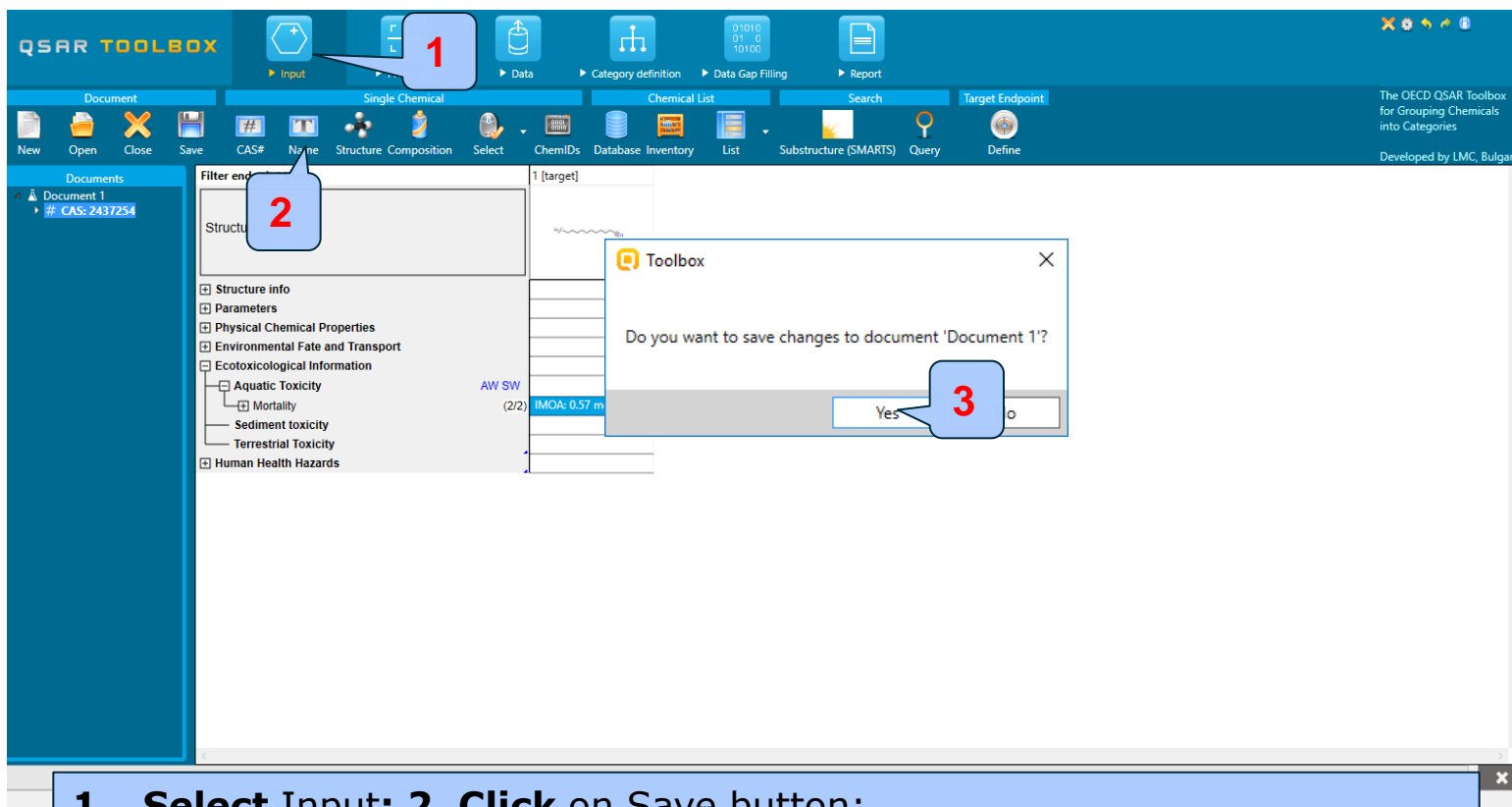
Outlook

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

Saving the prediction result

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

Saving the prediction



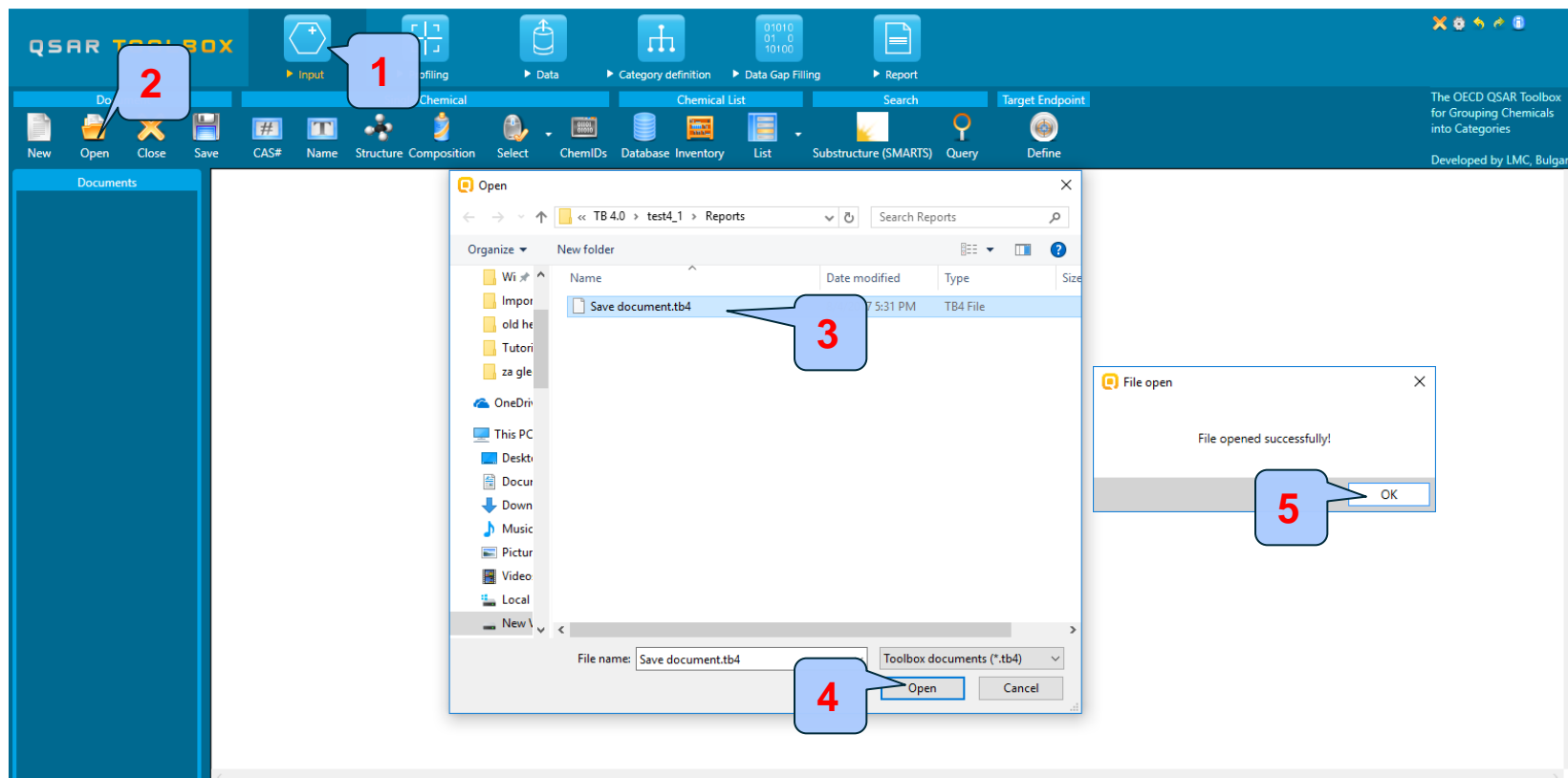
1. Select Input;
2. Click on Save button;
3. Click Yes

Saving the prediction

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. The main toolbar contains icons for New, Open, Close, Save, CAS#, Name, Structure, Composition, Select, ChemIDs, Database Inventory, List, Substructure (SMARTS), Query, and Define. On the left, a 'Filter endpoint tree...' panel shows a tree structure with categories like Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. A 'Save document' dialog box is open in the center, showing the file name 'Save document' and the file type 'QSAR Toolbox documents (*.tb4)'. A 'File save' confirmation message is shown on the right, stating 'File saved successfully!' with an 'OK' button. Three callouts (1, 2, and 3) highlight the 'File name' field, the 'Save' button, and the 'OK' button, respectively.

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

Open saved file



Once the file has been saved **1. Go** to Input; **2. Click** Open; **3. Find** and **select file**; **4. Click** Open **5. OK**