

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 'Search' menu is active, and a search window is open. The search input field contains '2437254'. The search results show a single entry for 'CAS 2437-25-4', which is identified as 'C12 nitrile' (Dodecanenitrile). The chemical structure is displayed as a long chain with a nitrile group. The interface includes a menu bar with options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. A search window is open, showing the input of CAS# 2437254 and the resulting chemical structure of C12 nitrile (Dodecanenitrile). Numbered callouts 1 through 4 guide the user through the steps: 1. Click on CAS#, 2. Enter 2437254, 3. The system identifies the structure, and 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 Document, Single Chemical, Chemical List, Search, and Target Endpoint. The left sidebar shows a list of documents, with 'Document 1' selected, having CAS# 2437254. The main workspace is divided into two panes. The left pane, titled 'Filter endpoint tree...', contains a tree view with categories such as Structure, Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. The right pane, titled '1 [target]', displays the chemical structure and associated data for the selected target. A red circle highlights the 'Structure info' section of the right pane, which contains the following data:

CAS Number	2437-25-4
CAS Smiles relation	High
Chemical name(s)	C12 nitrile Dodecanenitrile
Composition	C12H23N
Molecular Formula	Mono constituent
Predefined substance type	CCCCCCCCCCCC#N
Structural Formula	

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

1. Highlight the profiler

2. Click View

3. Select "Esters(Acute toxicity)"

The screenshot shows the QSAR Toolbox interface. The left sidebar contains the 'Profiling' section with a list of chemical categories. The 'US-EPA New Chemical Categories' is highlighted. The main window displays the 'US-EPA New Chemical Categories (Predefined) - Profiling Scheme Browser'. The 'View Tests' button is circled. The 'Esters (Acute toxicity)' category is selected in the list. The main window also shows a hierarchical tree of categories, including 'Structural boundary' and 'Parametric boundaries', and a 'Query details' section showing a SMARTS query and a chemical structure fragment.

Profiling

Side-Bar to Profiling

The screenshot displays the QSAR Toolbox interface. On the left, the 'Profiling' side-bar is visible, showing a document list with 'Document 1' (CAS: 2437254) and a list of profiling methods including 'OECD HPV Chemical Categories', 'Substance type', 'US-EPA New Chemical Categories', 'General Mechanistic', 'Biodegradation prim', and 'Simulated'. A context menu is open over the 'General Mechanistic' method, showing options: 'Select All', 'Unselect All', 'Invert', and 'View scheme'. The main window shows the 'Literature' tab for the 'Esters' category. A red circle highlights the 'Literature' tab in the top navigation bar, and a blue box with the number '1' points to it. An arrow labeled 'Textual description' points to the text in the 'Literature' tab. The text describes the category of esters, their toxicity, and the need for testing. It also mentions 'Hazard Concerns', 'Boundaries', and 'General Testing Strategy'.

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 $\log K_{ow} > 1000$ 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 $\log K_{ow}$ 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' tab is selected. In the 'Profiling methods' list, 'Acute aquatic toxicity MOA by OASIS' and 'Acute toxicity classification by ECOSAR' are checked. In the 'Metabolism/Transformations' section, 'Chemical elements' is checked. The 'Apply' button is circled in red. A 'Filter endpoint tree...' dialog is open, showing a list of endpoints with '1 [target]' selected. The 'Structure' section shows a chemical structure and various properties like CAS Number (2437-25-4), Molecular Formula (C12H23N), and Structural Formula (CCCCCCCCCCC#N).

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 QSAR Toolbox interface shows the Profiling workflow. The 'Filter endpoint tree...' window is open, displaying a tree structure of endpoints. The 'Profile' node is highlighted, and a red circle and arrow point to it, indicating that the user should double-click on this node to review the profiling results.

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 highlighted with a red circle and a callout '4'. The 'Documents' panel on the left shows the 'Databases' section expanded, with 'Ecotoxicological Information' selected and its sub-items (Aquatic ECETOC, Aquatic Japan MoE, Aquatic OASIS, ECHA CHEM, ECOTOX, and Human Health Hazards) listed. A red circle highlights these sub-items, with callout '2' pointing to the 'Ecotoxicological Information' header and callout '3' pointing to the sub-items. The 'Inventories' section below shows a list of databases including Canada DSL, COSING, DSSTOX, ECHA PR, EINECS, HPVC OECD, METI Japan, NICHAS, REACH ECB, TSCA, and US HPV Challenge Program. The 'Filter endpoint tree...' panel on the right shows a tree structure with 'Structure' selected. The '1 [target]' panel on the right shows a chemical structure.

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

Data

Gather data

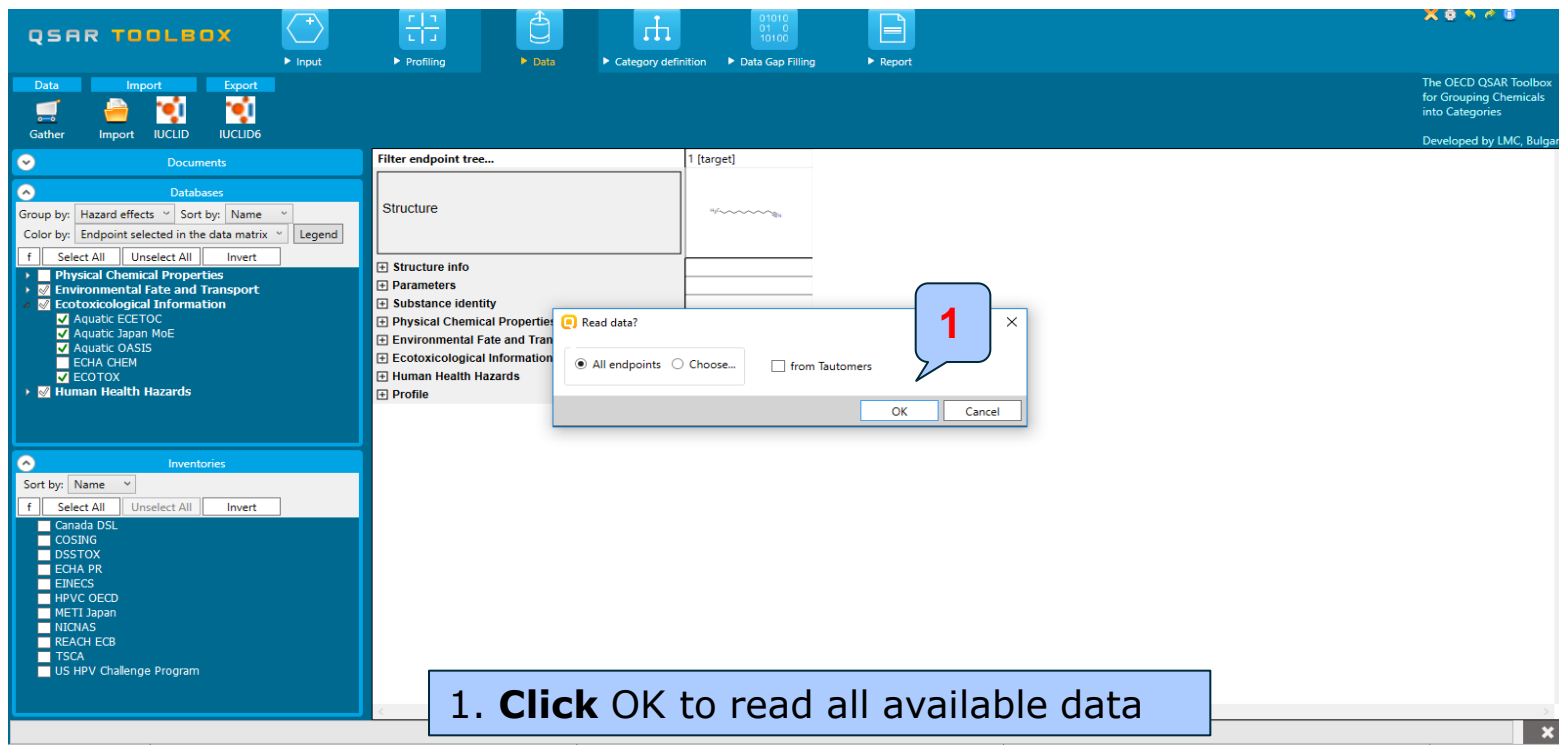
- Toxicity information on the target chemical is electronically collected from the selected dataset(s)
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases, which in this example are Aquatic toxicity OASIS; 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 with the 'Data' tab selected. The left sidebar contains 'Documents' and 'Databases' sections. The 'Databases' section is expanded, showing a list of endpoints under 'Ecotoxicological Information'. The 'pime' endpoint is selected, and a blue callout box with the number '1' points to it in the filter tree. The main panel displays the chemical structure of 'pime' and a table of experimental data for 'Pimephales promelas'.

Endpoint	Value
LC50	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

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Data Import Export

Gather Import IUCLID IUCLID6

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

Documents

Databases

Group by: Hazard effects Sort by: Name

Color by: Endpoint selected in the data matrix Legend

Select All Unselect All Invert

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Inventories

Sort by: Name

Select All Unselect All Invert

Canada DSL

COSING

DSSTOX

ECHA PR

EINECS

HPVC OECD

METI Japan

NICNAS

REACH ECB

TSCA

US HPV Challenge Program

pime

Structure

1

1 [target]

Ecotoxicological Information

Aquatic Toxicity

Behavior

EC50

12 h

24 h

48 h

72 h

96 h

Mortality

LC50

12 h

24 h

48 h

72 h

96 h

Animalia

Chorda

Ac

AW SW

(1/1) M: >1.5+2.25 mg/L

(1/1) M: >1.5+2.25 mg/L

(1/1) M: >0.75+1.5 mg/L

(1/1) M: >0+0.75 mg/L

(1/1) M: 0.42 mg/L

(1/1) M: >1.5+2.25 mg/L

(1/1) M: >1.5+2.25 mg/L

(1/1) M: >0.75+1.5 mg/L

(1/1) M: >0+0.75 mg/L

1

2

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

2. Click on the X to close the window

Data points

Datapoints	#	Value	Original value	Assigned SMILES	
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. North
Ecotoxicological Information/Aquatic Toxicity	2	M: 0.425 mg/L (Mass concentration)	2.34E-06 mol/L (Molar concentration)	False	Russom, C. Broderius, Hammer

Hierarchical mode Find OK

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

1

2

3

Filter endpoint tree...

Structure

Structure info

Parameters

Chemical Properties

and Transport

formation

Information

Profile

tautomer #1

tautomer #2

tautomer #3

Tautomeric forms

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

Profiling set of tautomeric forms

1 Check the profilers related to acute aquatic toxicity as mention on slide #41

2 Click Apply

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

Parent chemical [target]	tautomer #1	tautomer #2	tautomer #3
Structure	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
Structure info			
Parameters			
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information			
Human Health Hazards			
Profile			
Endpoint Specific			
Acute aquatic toxicity classification by Verhaar (Modified)			
Acute aquatic toxicity MOA by OASIS			
Aquatic toxicity classification by ECOSAR			
Empiric			
Organic functional groups			
Organic functional groups (nested)			
Organic functional groups (US EPA)			
Organic functional groups, Norbert Haider (...)			
(2/32) M: >1.5~2.25 mg/L	M: 0.054 mg/L		
Class 5 (Not possible to)	Class 5 (Not possible to)	Class 5 (Not possible to)	Class 5 (Not possible to)
Basesurface narcotics	Basesurface narcotics	Basesurface narcotics	Basesurface narcotics
Neutral Organics	Neutral Organics	Not Related to an Existing	Aliphatic Amines
Nitrile	Nitrile	Allyl	Alkyne

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

1

2

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

1. Right click over the active tautomeric form

2. Focus the chemical

Handling of tautomerism of target chemical

Focus of active tautomer

The screenshot displays the QSAR Toolbox interface. In the 'Documents' panel, 'tautomer #3' is selected. The 'Filter endpoint tree...' panel shows 'Structure' selected. The 'Structure' panel displays the chemical structure of the selected tautomer, which is circled in red. Below the structure, a table shows the following properties:

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

A blue callout box at the bottom 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 menu bar includes options: Data, Import, Export, Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below the menu bar, the 'Databases' panel is active, showing a list of databases with checkboxes for selection. The 'Human Health Hazards' section is expanded, showing 'Acute Oral toxicity' selected. The 'Inventories' section is also expanded, showing 'Canada DSL', 'COSING', 'DSSTOX', 'ECHA PR', 'EINECS', 'HPVC OECD', and 'METI Japan' selected. The 'Filter endpoint tree...' panel on the right shows a tree structure with 'Structure' selected. The '1 [target]' panel on the far right shows a chemical structure. The bottom status bar indicates 'The OECD QSAR Toolbox for Grouping Chemicals into Categories' and 'Developed by LMC, Bulgaria'.

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 interface shows a document tree on the left with 'Document 1' containing 'Tautomerism' and three 'tautomer' entries. The main window displays a list of categories, with 'Aquatic toxicity classification by ECOSAR' highlighted. A 'Filter endpoint tree...' dialog box is open, showing 'Aliphatic Amines' as the target category. The 'Combine profiles' section has 'AND' selected and 'Strict' checked. The 'OK' button is highlighted.

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 shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Category definition' step is active. On the left, a 'Documents' panel lists 'Document 1' with CAS: 2437254 and 'Tautomerism' subcategory, including 'tautomer #1 (target)', 'tautomer #2', and 'tautomer #3'. The 'Filter endpoint tree...' panel shows 'Structure' selected. A 'Grouping results' dialog box is open, displaying '371 chemicals found.' and an 'OK' button. A blue callout bubble with the number '1' points to the 'OK' button. The background shows a table with chemical structures and their corresponding ECOSAR categories.

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

Handling of tautomerism of target chemical

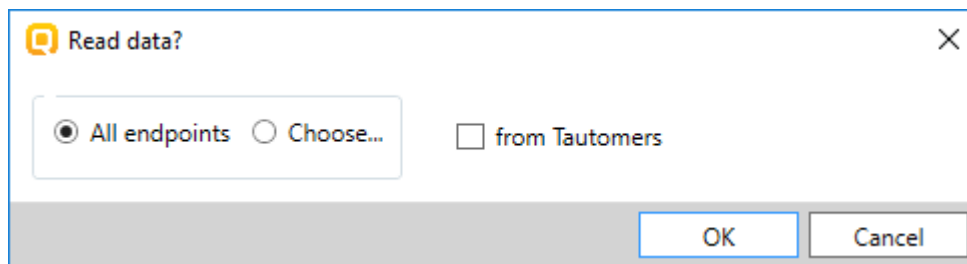
Category analogues

- 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 menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows a 'Documents' panel with a tree view for 'Document 1' (CAS: 2437254) containing 'tautomerism' and 'Aquatic toxicity classification by ECOSAR'. The main workspace shows a 'Filter endpoint tree...' panel on the left and a table of chemical structures on the right. A 'Gather data' dialog box is open in the center, displaying the message: '9414 points added across 369 chemicals.' and an 'OK' button. A red circle with the number '1' is placed over the 'OK' button. The bottom status bar shows '171'.

Handling of tautomerism of target chemical

Summary information for Analogues

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Categorize

Define Define with metabolism Subcategorize Combine

Documents

Document 1

CAS: 2437254

- Tautomerism
 - tautomer #1 (target)
 - tautomer #2
 - tautomer #3
- Aquatic toxicity classification by ECOSAR

Aquatic toxicity classification by ECOSAR

Options

Select All Unselect All Invert About

Toxic hazard classification by Gramer (extended)

Endpoint Specific

- Acute aquatic toxicity classification by Verhaar (Modified)
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Bioaccumulation - metabolism alerts
- Bioaccumulation - metabolism half-lives
- Biodegradation fragments (BioWIN MITI)
- Carcinogenicity (genotox and nongenotox) alerts by ISS
- DART scheme
- DNA alerts for AMES by OASIS
- DNA alerts for CA and MNT by OASIS
- Eye irritation/corrosion Exclusion rules by BFR
- Eye irritation/corrosion Inclusion rules by BFR

Structure

pime

1 [target] 2 3 4 5 6 7

EC50 (6/15)

LC10 (8/19)

LC50

1 h (2/5)

3 h (4/30)

6 h (4/36)

12 h (4/7)

24 h (58/219)

48 h (72/259)

72 h (16/21)

96 h

Actinopterygii (ray-finned fishes, spi ...)

Pimephales promelas (66/145)

26 h (3/9)

LOEC (19/44)

LT50 (3/11)

NOEC (24/67)

NR-LETH (41/90)

NR-ZERO (42/88)

Undefined Endpoint (86/282)

No Effect Coded (39/113)

Reproduction (54/247)

M: 0.0129 (0.0055-0)

M: 0.1 (0.09-0.11) m

M: 0.051 (0.046-0.05)

M: 0.69 (0.61-0.81) mg

M: 25 (22.6-27.6) m

M: 102 (97.9-106) m

M: >10 ppm

M: 0.000287 mg/L

M: 0.074 mg/L

M: 0.096 mg/L

M: 0.0001-0.1 mg/L

M: 10 mg/L

M: >10.1 ppm

M: >10 ppm

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 displays the QSAR TOOLBOX interface. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows the 'Documents' panel with a tree view of the chemical structure and its tautomers. The 'Data Gap Filling Settings' panel is also visible, with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant'. The main panel shows the 'Structure' tab with a tree view of the endpoint tree. The tree is expanded to show the 'Pimephales promelas' endpoint. A red box highlights the 'Pimephales promelas' endpoint in the tree, and a red box highlights the corresponding data row in the table below. The table shows the results of the data gap filling process, including the chemical structure, the endpoint name, and the calculated values for various endpoints.

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

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

QSAR Toolbox 4.0.0.22512 [Document 1]

Menu: Input, Profiling, Data, Category definition, Data Gap Filling, Report

Left sidebar: Documents, Data Gap Filling Settings

Filter endpoint tree...: 1 [target]

Chemicals list:

- Menidia beryllina (2/2)
- Micropterus dolomieu (1/1)
- Micropterus salmoides (1/1)
- Oncorhynchus kisutch (1/1)
- Oncorhynchus mykiss (38/63)
- Oncorhynchus tshawytscha (1/1)
- Oryzias latipes (38/39)
- Pimephales promelas (67/146)
- Poecilia reticulata (28/29)
- Rutilus rutilus (1/1)
- Salmo trutta (1/1)
- Salvelinus namaycush (1/1)
- Sander vitreus (1/1)
- Valamugil engeli (1/1)
- Amphibia (Amphibians) (2/2)
- Mollusca (molluscs, mollusks) (3/4)
- Platyhelminthes (flatworms) (3/4)
- Undefined Kingdom (13/22)
- 100 h (1/1)
- 5 d (4/11)
- 5.83 d (2/2)
- 6 d (1/3)

Possible data inconsistency dialog box:

- Native scale/unit:
 - ☒ mg/L (11 data; 7 chemicals)
 - ☒ mol/L (58 data; 7 chemicals)
 - ☒ ppm (2 data; 1 chemical)
 - ☒ µL/L (8 data; 1 chemical)
 - ☒ µg/L (67 data; 1 chemical)
- Gap filling scale/unit:
 - ☒ log(1/mol/L)
 - ☐ log(1)
 - ☐ mol/L
 - ☐ µL/L
 - ☐ mg/L
 - ☐ ppm
- Data 136/146; Chemicals 65/67
- Buttons: OK, Cancel

- 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 (Q)SAR 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 point

Filter endpoint tree...

Structure

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

1 [target] 2 4 10 32 35 36

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

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

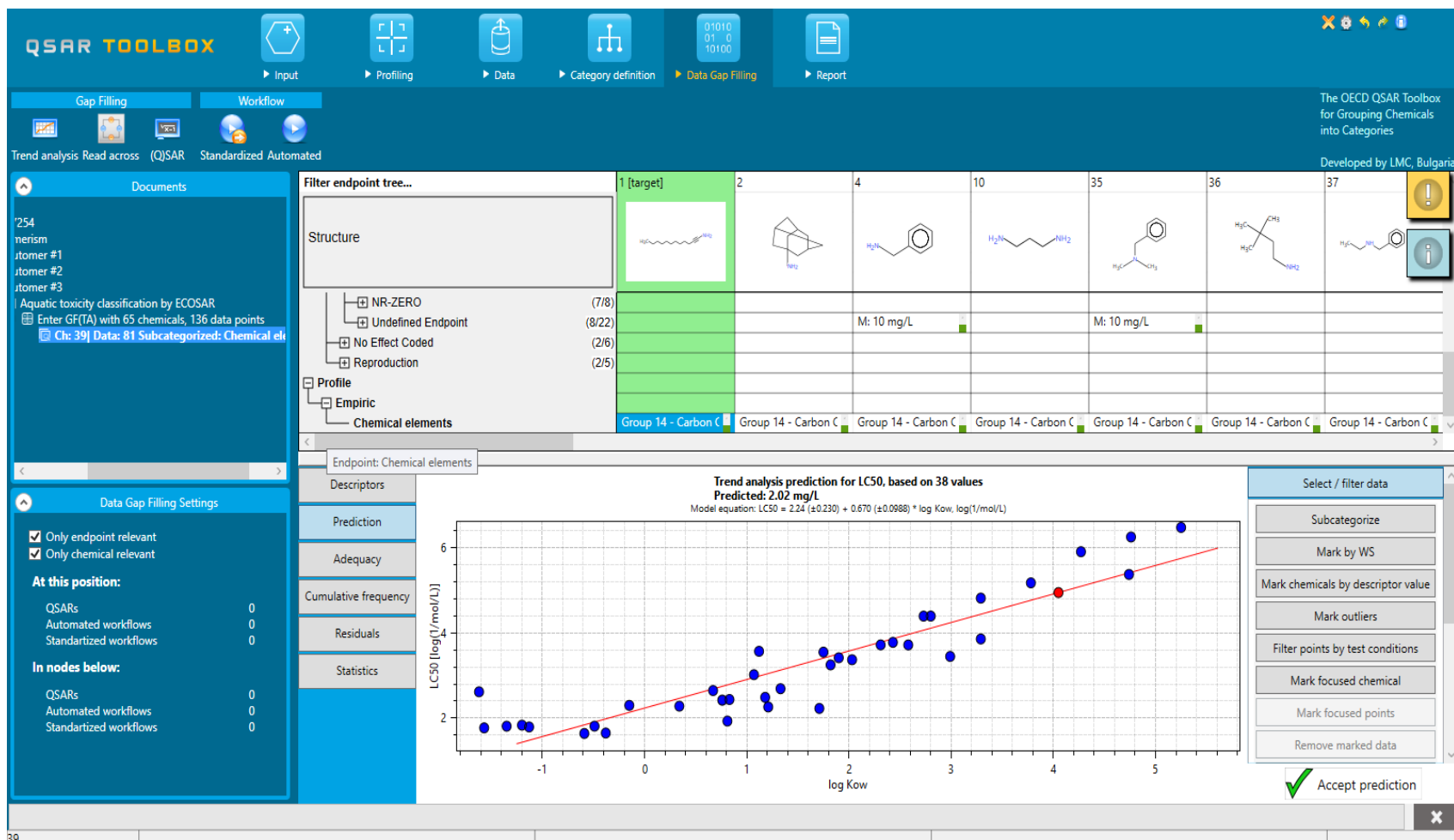
Subcategorisation 1 by Chemical elements

The screenshot displays the QSAR Toolbox software interface during the Subcategorization process. The main window is titled 'Subcategorization' and contains several panels. On the left, there are 'Options' and 'Adjust options' sections. The 'Options' section has a list of endpoints, with 'Chemical elements' highlighted by a red circle and a blue callout labeled '2'. Below this, there are sections for 'Documented' and 'Simulated' endpoints. The 'Adjust options' section on the right shows a list of chemical groups, including 'Group 14 - Carbon C' and 'Group 15 - Nitrogen N'. At the bottom of the 'Subcategorization' window, there are buttons for 'Selected 27 (38/65)', 'Select differ', and 'Remove selected' (highlighted with a red circle and a blue callout labeled '3'). On the far right, a 'Select / filter data' panel is visible, with 'Subcategorize' highlighted by a red circle and a blue callout labeled '1'. The background shows a list of documents on the left and a grid of chemical structures on the right. A log Kow plot is visible at the bottom of the main window.

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

Data Gap Filling

Result of Subcategorisation 1 by Chemical elements



Data Gap Filling

Subcategorisation 2 by OFG (nested)

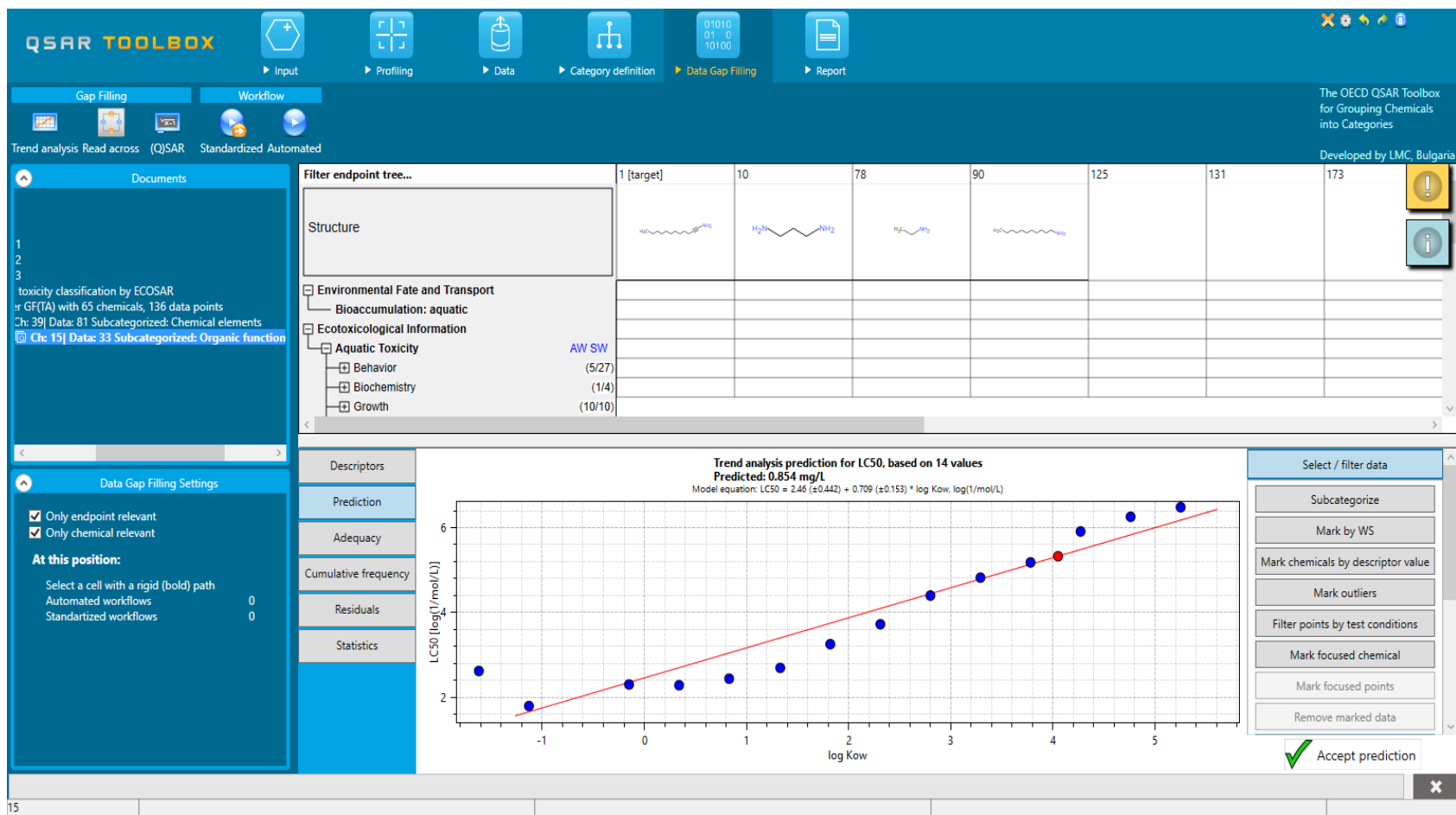
1. Click Subcategorize

2. Select OFG (nested)

3. Click Remove to eliminate dissimilar analogues

Data Gap Filling

Result of Subcategorisation by OFG (nested)



Data Gap Filling

Side-Bar of Subcategorisation

The last subcategorisation procedure 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 light blue; 4. Click Remove to eliminate dissimilar analogues

Subcategorization

Group by: Category Sort by: Name

Color by: Target endpoint Legend

Select All Unselect All Invert

Groups of elements
Lipinski Rule
Organic function
Organic function (nested) (US EPA)
Organic function groups, Norbert Haide
Structure similarity
Tautomers unstable
Toxicological
Repeated dose (HESS)
Custom
Example Prioritization Scheme (PBT)

Adjust options

[90%,100%]

Differ from target by
At least one category
All categories

[STOP]

Groups (nested)

10 78 90 125 131 173 176

Chemical structures: NCCCN, CCN, CCCCN, CCCCCN, CCCCCN, CCCCCN, CCCCCN

M: 10 mg/L

3

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

Chemical structures: NCCCN, CCN, CCCCN

4

Select / filter data

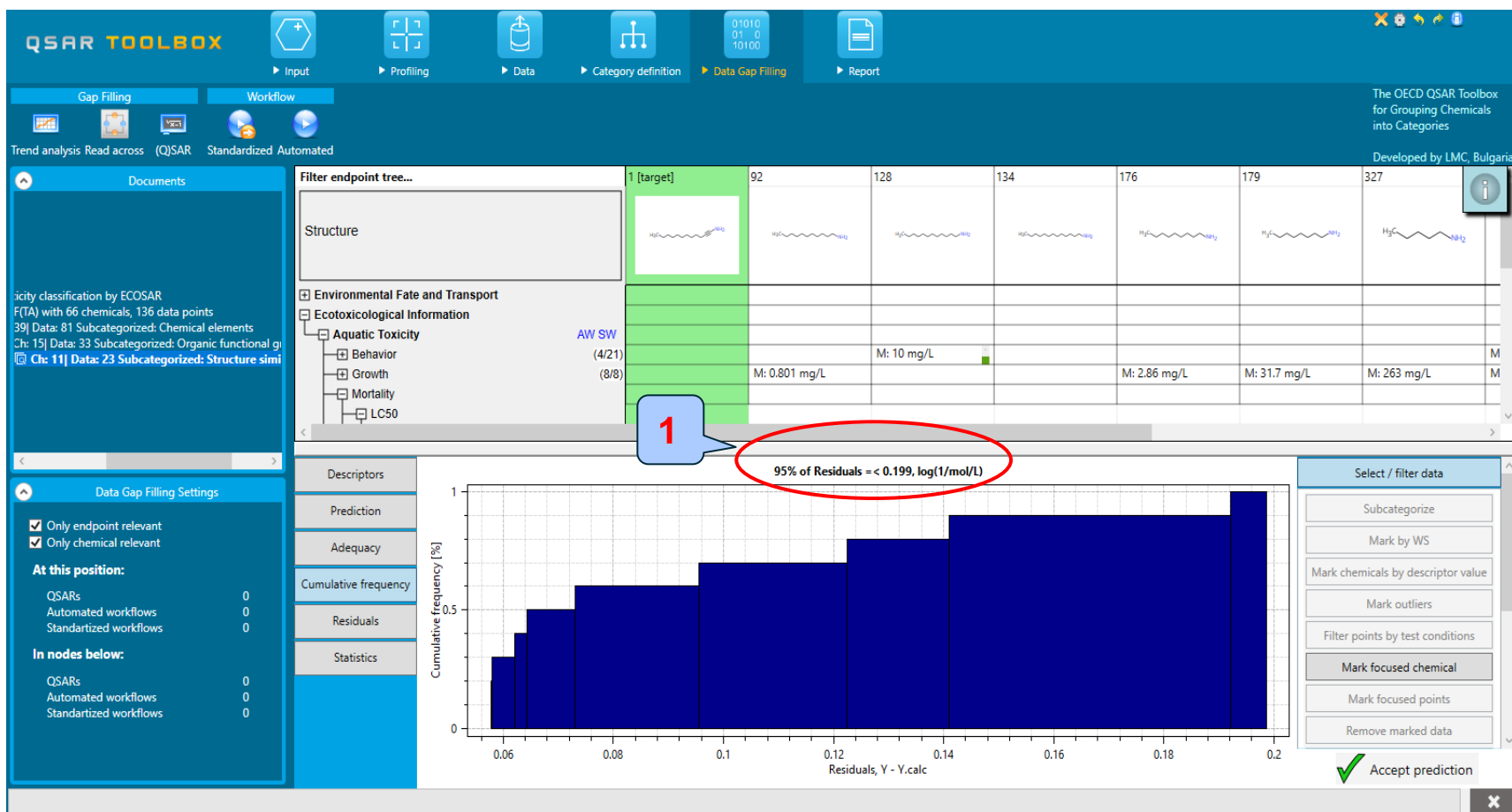
Subcategorize
Mark by WS
Mark chemicals by descriptor value
Mark outliers
Filter points by test conditions
Mark focused chemical
Mark focused points
Remove marked data

Accept prediction

71

Data Gap Filling

Cumulated frequency



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

Data Gap Filling Statistics

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

The OECD QSAR Toolbox for Grouping Chemicals into Categories

Developed by LMC, Bulgaria

Documents

Filter endpoint tree...

Structure

Environmental Fate and Transport

Ecotoxicological Information

Aquatic Toxicity

Behavior (4/21)

Growth (8/8)

Mortality

LC50

1 [target]

92

128

134

176

179

327

Statistical characteristics

Number of data points, (N)

Coefficient of determination, (R2)

Adjusted coefficient of determination, (R2adj)

Coefficient of determination - leave one out, (Q2)

Sum of squared residuals, (SSR)

Standard deviation of residuals, (sN)

Sample standard deviation of residuals, (s)

Fisher function, (F)

Fisher threshold for statistical significance, (Fa)

b0

- model descriptor

- coeff. value

TA model

10

0.993

0.992

0.987

0.140

0.118

0.132

1.09E3

8.09 (95.0%)

Intercept

1.53

Remove marked data

Clear existing marks

Gap filling approach

Descriptors / data

Model/QSAR

Show domain

Save model

Save domain as category

Calculate Q2

Accept prediction

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

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Document Single Chemical Chemical List Search Target Endpoint

New Open Close Save CAS# Name Structure Composition Select Delete ChemIDs Database Inventory List Substructure (SMARTS) Query Define

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

Document 1
CAS: 2437254
Tautomerism
tautomer #1
tautomer #2
tautomer #3
Aquatic toxicity classification by ECOSA
Enter GF(TA) with 66 chemicals, 136
Ch: 39| Data: 81 Subcategorized:
Ch: 15| Data: 33 Subcategorized:
Ch: 11| Data: 23 Subcategorized:

Filter endpoint tree...

Structure

Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information

Aquatic Toxicity

Behavior
Growth
Growth Inhibition
Immobilisation
Mortality
LC50
12 h
24 h
48 h
72 h
96 h
Animalia (animals)
Chordata (chordates)
Actinopterygii (ray-finned fishes, spiny rayed fishes)
Oryzias latipes
Pimephales promelas
Sediment toxicity

Parent chemical [target]

tautomer #1

tautomer #2

tautomer #3

AW SW

(2/10) M: >1.5+2.25 mg/L M: >0.75+1.5 mg/L

(2/2) M: 2.28 mg/L M: 2.28 mg/L

(2/4) M: 0.15 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 M: >1.5+2.25 mg/L

(2/2) M: >1.5+2.25 mg/L M: >1.5+2.25 mg/L

(2/2) M: >0.75+1.5 mg/L M: >0.75+1.5 mg/L

(2/2) M: >0+0.75 mg/L M: >0+0.75 mg/L

(2/2) M: 0.84 mg/L M: 0.84 mg/L

M: 0.425 mg/L M: 0.425 mg/L

M: 0.43 (0.4+0.47) mg/L M: 0.43 (0.4+0.47) mg/L

(3/5)

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

Document: 2437254
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
12 h
24 h
48 h
72 h
96 h
Animalia (animals)
Chordata (chordates)
Actinopterygii (ray-finned fishes)
Oryzias latipes
Pimephales
Sediment toxicity
Terrestrial Toxicity

Parent chemical [target]	tautomer #1	tautomer #2	tautomer #3
Structure			
Structure info			
Parameters			
Physical Chemical Properties			
Environmental Fate and Transport			
Ecotoxicological Information			
Aquatic Toxicity			
Behavior			
Growth			
Growth Inhibition			
Immobilisation			
Mortality			
LC50			
12 h			
24 h			
48 h			
72 h			
96 h			
Animalia (animals)			
Chordata (chordates)			
Actinopterygii (ray-finned fishes)			
Oryzias latipes			
Pimephales			
Sediment toxicity			
Terrestrial Toxicity			

TA prediction coincide with measured data

2. The prediction of the tautomeric form is assigned to the last SMILES within the set;

1. The trend analysis prediction appears on datamatrix;

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

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** Independent MOA

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. The top toolbar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows a document tree with 'Tautomers' highlighted. The central 'Filter endpoint tree...' pane shows a hierarchical list of endpoints. A 'Possible data inconsistency' dialog box is open, showing 'Native scale/unit' and 'Gap filling scale/unit' options. A table at the bottom displays data for 'Oryzias latipes' and 'Pimephales promelas'. Numbered callouts (1-4) indicate the steps: 1. Select parent; 2. Independent MOA; 3. Use Scale/unit (log(1/mol/L)); 4. Click OK.

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

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

Pimephales promelas (2/3)

Sediment toxicity
Terrestrial Toxicity
Human Health Hazards
Profile
Empiric
Chemical elements

1 [target] 2 4

M: 0.425 mg/L T: 0.57 (0.268+1.21) mg

Group 14 - Carbon C

Descriptors

Prediction

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

1. Accept prediction

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

1

Accept prediction

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 interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data G, and Report. The Report icon is highlighted with a red circle and a blue callout box containing the number 1. The left sidebar shows a tree view of documents, with 'Prediction' highlighted by a red circle and a blue callout box containing the number 3. The main window displays a table of results for various endpoints, including Aquatic toxicity classification by ECOSAR. A red circle highlights the 'IMOA: 0.494 mg/L' value in the table, with a blue callout box containing the number 2.

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+0.75 mg/L		
96 h				
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

Document 1
CAS: 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: Chemical
Ch: 15| Data: 33 Subcategorized: Organism
Ch: 11| Data: 23 Subcategorized: Structure
Enter GF(IndependentMOA) with 3 chemicals, 136 data points
Select from: Tautomerism
Aquatic toxicity classification by ECOSAR
Enter GF(TA) with 66 chemicals, 136 data points
Ch: 39| Data: 81 Subcategorized: Chemical
Ch: 15| Data: 33 Subcategorized: Organism
Ch: 11| Data: 23 Subcategorized: Structure

Prediction report4.pdf

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

1

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: CCCCCCCCCCCC#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

Back Next Cancel Create report

Human Health Hazards

1. TB report for multicomponent substance

Report

Document 1
CAS: 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: Chemical
Ch: 15| Data: 33 Subcategorized: Organ
Ch: 11| Data: 23 Subcategorized: Str
Enter GF(IndependentMOA) with 3 chemicals, 136 data points
Select from: Tautomerism
Aquatic toxicity classification by ECOSAR
Enter GF(TA) with 66 chemicals, 136 data points
Ch: 39| Data: 81 Subcategorized: Chemical
Ch: 15| Data: 33 Subcategorized: Organ
Ch: 11| Data: 23 Subcategorized: Str

SMILES:	EC#:	C12 nitrile
CCCCCCCCCCCC#N	N/A	Dodecanenitrile
Structure	CAS#: 2437-25-4	Dodecanonitrile
<chem>CCCCCCCCCCCC#N</chem>	Other: N/A	

Prediction summary

Predicted endpoint: LC50; Mortality; Pimephales promelas; 96h; No guideline specified

Predicted value: 0.494

Unit/scale: mg/L

Data gap filling method: Independent mode of action

Summary: manually editable field
Not provided by the user

1. Predicted value

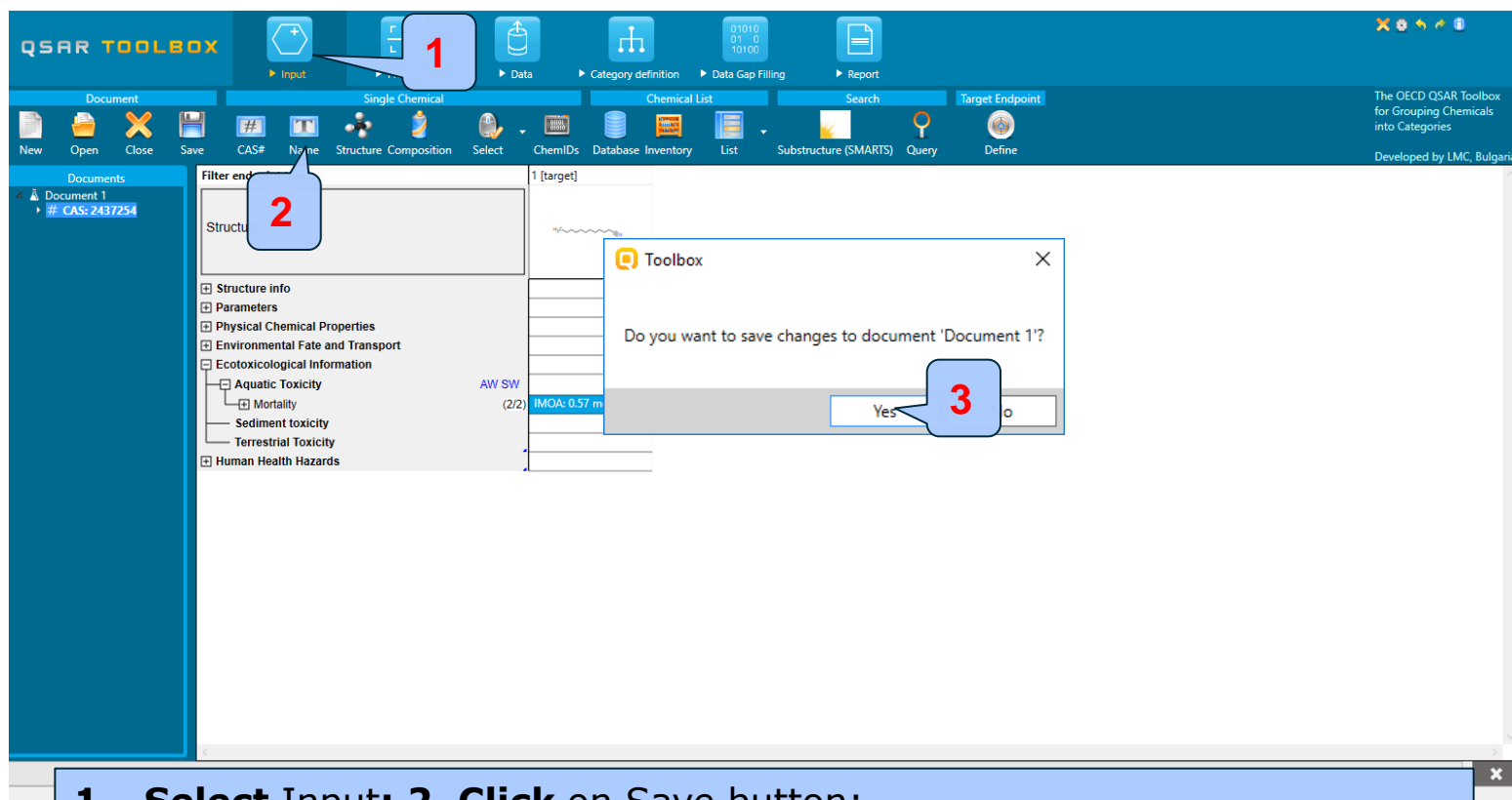
Outlook

- 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

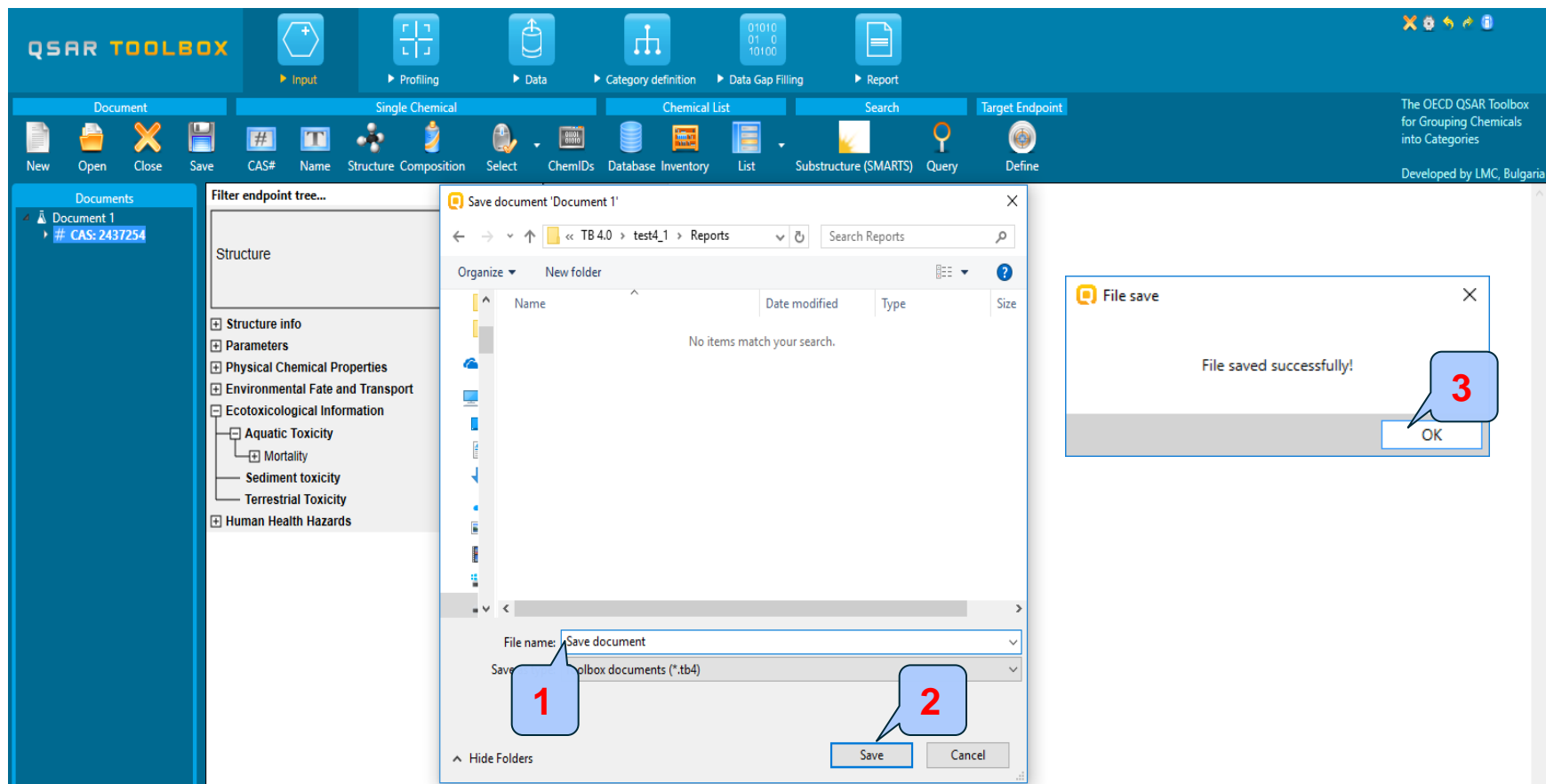
- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document 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



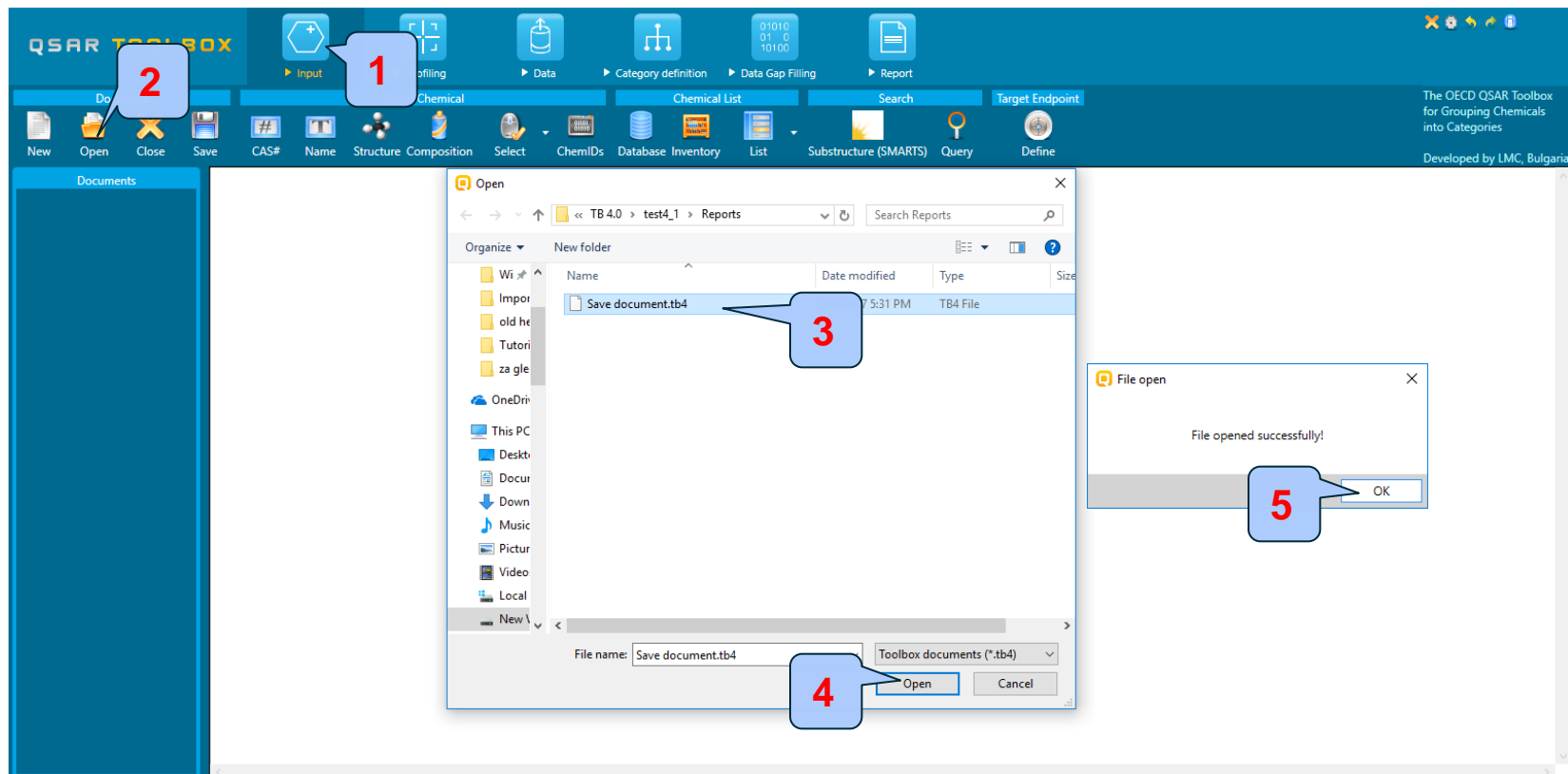
1. Select Input; 2. Click on Save button;
3. Click Yes

Saving the prediction



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**