

OECD QSAR Toolbox v.4.1

Step-by-step example of how to evaluate an ad-hoc category of aliphatic amines and to predict an ecotoxicological endpoint

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating of an ad-hoc category.
- You will learn several new functionalities which we be repeated to assure a consistent category is defined.
- It is assumed that you now have some experience in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

- Background
- **Category evaluation - overview**

Category Evaluation

Two phases of evaluation process

Phase I. Evaluation of Category & Data Gap Filling:

- Investigating the structural consistency of an ad hoc category (e.g., a category submitted by an industry consortium to a regulatory assessment program).
- Implementation of an ad hoc category building and data gap filling.

Phase II. Extension of Category & Data Gap Filling:

- Search for other analogues which are consistent with the submitted category.
- Data gap filling using new data matrix.

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - ***Case study***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Case Study

- The submission consists of a category with **19 aliphatic amines**.
- The predicted ecotoxicological endpoint EC50, 48h, *D.magna* of **2-Butanamine(CAS 13952-84-6)** will be reviewed.

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Case study*
 - **Workflow**

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow

The following input workflow is used:

- Input the file for submission as an user list.(In this case input file Aliphatic amines.smi*)
- Evaluate the category applying the following profiling schemes:
 - US-EPA New Chemical Categories
 - Aquatic toxicity classification by ECOSAR
 - Aquatic toxicity MOA of action
 - Organic functional groups (nested)

Aliphatic amines.smi* - file is available with TB installation, located at C:\Program Files (x86)\Common Files\QSAR Toolbox 4.1\Config\Examples

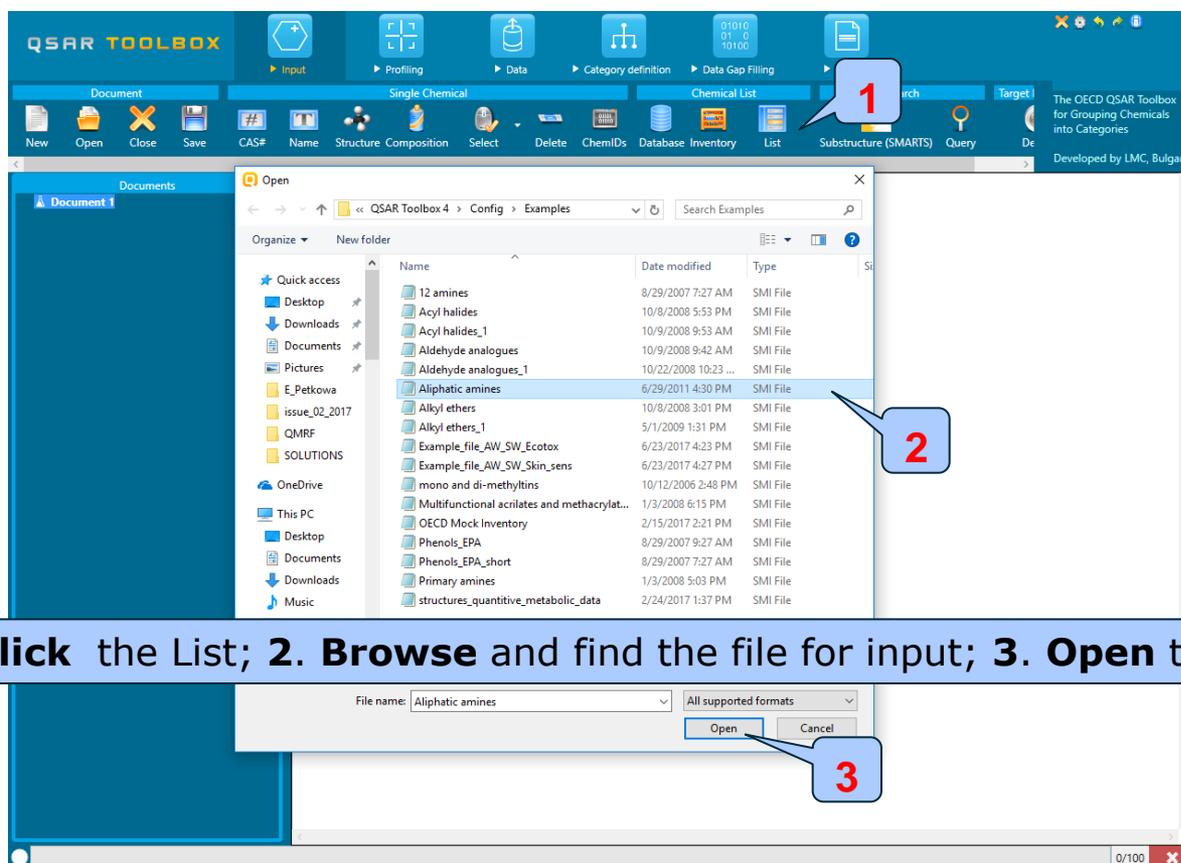
Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Case study*
 - ***Workflow***
 - ***Input***

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Input



1. Click the List; **2. Browse** and find the file for input; **3. Open** the file.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

***Workflow/*Input**

- You have now inserted your chemical list into the system.
- **Click** on the box next to "Structure info"; this displays the chemical identification information (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Input

The screenshot shows the QSAR Toolbox interface. The top menu bar includes Document, Single Chemical, Chemical List, Search, and Target. The main workspace is divided into a left sidebar with a 'Filter endpoint tree...' and a central table of chemical data. A red callout box with the number '1' points to the 'Structure info' section in the left sidebar.

1	2	3	4	5
<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(C)N</chem>
109-73-9	74-89-5	75-04-7	75-31-0	75-64-9
High	High	High	High	High
1-amino-butane	amino-methane	amino-ethane	2-amino-propane	(tert)butylamine
C4H11N	CH5N	C2H7N	C3H9N	C4H11N
Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constitute
CCCCN	CN	CCN	CC(C)N	CC(C)(C)N

1. Click on "Structure info"

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Case study*
 - **Workflow**
 - *Input*
 - **Profiling**

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

- The first step of the category evaluation according to Phase I is the consistency check of the category which starts with evaluation of the **robustness with respect to structural functionalities**. The following schemes could be used for this purpose:
 - US-EPA New Chemical Categories
 - Aquatic toxicity classification by ECOSAR
 - Aquatic toxicity MOA of action
 - Organic functional groups (nested)
- **Select "profiling methods"** by clicking on the boxes before the names of the profilers and **Click "Apply"**. Before selecting the profiling methods unselect all (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

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

- Top Navigation Bar:** Includes icons for Input, Profiling (highlighted with a red circle '1'), Data, Category definition, Data Gap Filling, and Report.
- Left Panel:**
 - Documents:** Shows 'Document 1' containing 'Aliphatic amines'.
 - Profiling methods:** A list of methods with 'Unselect All' highlighted (red circle '2').
 - Metabolism/Transformations:** A list of methods under 'Documented' and 'Simulated' categories.
- Main Panel:**
 - Filter endpoint tree...** A tree view showing 'Structure' and various categories like 'Structure info', 'Parameters', 'Physical Chemical Properties', etc.
 - Table:** A table with 5 columns representing chemical structures and rows for various properties.

Structure	1	2	3	4	5
Structure	<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(C)N</chem>
CAS Number	109-73-9				
High					
Chemical name(s)	1-amino-butane	amino-methane			
Composition					
Molecular Formula	C4H11N	CH5N			
Predefined substance type		Mono constituent	Mono constituent		
Structural Formula	CCCCN	CN			
Parameters					
Physical Chemical Properties					
Environmental Fate and Transport					
Ecotoxicological Information					
Human Health Hazards					

1. Go to profiling; 2. Select Unselect All.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

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

- Top Bar:** QSAR TOOLBOX logo and navigation icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report.
- Sub-Menu:** Profiling, Custom profile, Apply (circled in red), View, New, Delete.
- Documents Panel:** Profiling methods list with 'Endpoint Specific' checked (marked with a red circle and '1').
- Filter endpoint tree:** Expanded to 'Structure'.
- Structure Info Table:**

Structure	1	2	3	4	5
Structure	<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(C)N</chem>
CAS Number	109-73-9	74-89-5	75-04-7	75-31-0	75-64-9
CAS Smiles relation	High	High	High	High	High
Chemical name(s)	1-amino-butane	amino-methane	amino-ethane	2-amino-propane	(tert)butylamine
Composition					
Molecular Formula	C4H11N	CH5N	C2H7N	C3H9N	C4H11N
Predefined substance type	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent
Structural Formula	CCCCN	CN	CCN	CC(C)N	CC(C)(C)N

1. Check the USEPA, MOA of action, ECOSAR and OFG (nested) profilers; 2. Click Apply.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appeared as a dropdown box under the target chemical.
- The result from profile statistics, can be seen by **right clicking** in the space above the profiler in the endpoint tree and **select Profile statistics** from the dropdown menu (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling

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 Profiling (Apply, View, New, Delete) and Custom profile (New, Delete). The main workspace is divided into several panels:

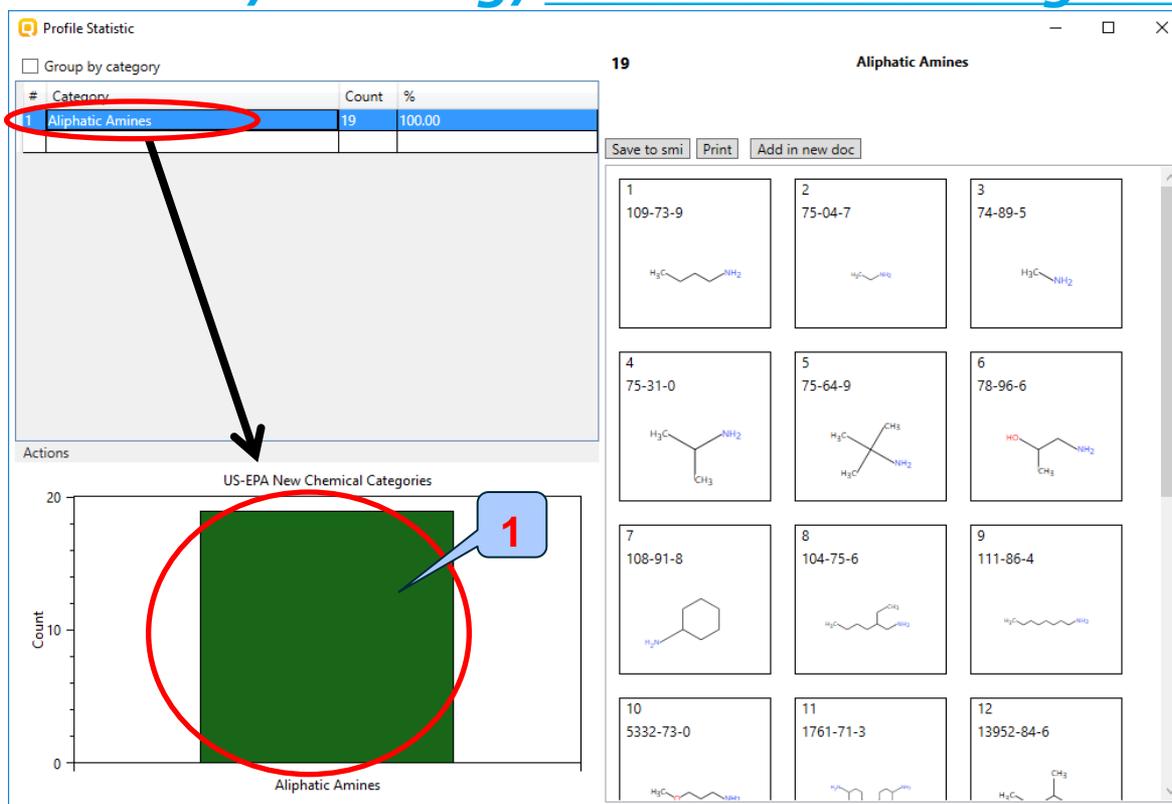
- Documents:** Shows the current document name.
- Profiling methods:** A tree view of various profiling methods. The 'US-EPA New Chemical Categories' is selected under 'General Mechanistic'.
- Filter endpoint tree...:** A tree view showing the selected endpoint, 'Acute aquatic toxicity', under 'US-EPA New Chemical Categories'.
- Structure info table:** A table with 5 columns corresponding to chemical structures. It lists CAS numbers, CAS Smiles relations, chemical names, compositions, molecular formulas, predefined substance types, and structural formulas.
- Structure info table:** A table with 5 columns corresponding to chemical structures. It lists CAS numbers, CAS Smiles relations, chemical names, compositions, molecular formulas, predefined substance types, and structural formulas.
- Profile Statistics table:** A table with 5 columns corresponding to chemical structures. It lists the profile statistics for each structure, such as 'Aliphatic Amines', 'Narcotic Amine', and 'Aliphatic amine, prin'.

A blue callout box with the text "1. Right click above US-EPA profiler; 2. Select Profile Statistics." is positioned over the 'US-EPA New Chemical Categories' in the Filter endpoint tree. A red callout box with the number "2" is positioned over the 'Profile Statistics' option in the context menu.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to US-EPA



1. In this case all 19 chemicals are Aliphatic amines according to US-EPA profiler.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to Aquatic toxicity MOA by OASIS

The screenshot shows the 'Profile Statistic' window for the category 'Narcotic Amine' (19 items). A table lists the category and its count and percentage. Below the table, a bar chart titled 'Acute aquatic toxicity MOA by OASIS' shows a single green bar representing the count of 19. A red circle highlights the bar, and a blue callout box with the number '1' points to it. An arrow points from the 'Narcotic Amine' entry in the table to the bar chart.

#	Category	Count	%
1	Narcotic Amine	19	100.00

Acute aquatic toxicity MOA by OASIS

Count

1

1. In this case all 19 chemicals are Narcotic amines according to MOA by OASIS profiling scheme

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to ECOSAR

Profile Statistic

Group by category

#	Category	Count	%
1	Aliphatic Amines	19	100.00

19 Aliphatic Amines

Save to smi Print Add in new doc

1 109-73-9
CCCCN

2 75-04-7
CCN

3 75-64-9
CC(C)(C)N

4 78-96-6
CC(O)CN

5 74-89-5
CCN

6 75-31-0
CC(C)N

7 104-75-6
CCCCCN

8 111-86-4
CCCCCCN

9 108-91-8
C1CCCCC1N

10 1761-71-3

11 111-68-2

12 5332-73-0

Actions

Aquatic toxicity classification by ECOSAR

1

Count

20

10

0

1. In this case all 19 chemicals are Aliphatic amines according to ECOSAR scheme

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Workflow/Profiling/Statistics according to OFG(nested)

The screenshot displays the 'Profile Statistic' window with a table of categories. A red circle highlights the 'Category' column header. A blue callout box with the number '1' points to the bar chart, which shows a significant peak for the 'Aliphatic amine, primary+Amine, primary+Overlapped groups' category. A blue text box at the bottom states: '1. Aliphatic amine fragment is available.'

#	Category	Count	%
1	Aliphatic amine, primary+Alkane, branched wit	1	5.26
2	Aliphatic amine, primary+Alkane, branched wit	1	5.26
3	Aliphatic amine, primary+Alkane, branched wit	1	5.26
4	Aliphatic amine, primary+Alkane, branched wit	1	5.26
5	Aliphatic amine, primary+Alkane, branched wit	1	5.26
6	Aliphatic amine, primary+Amine, primary+Cycl	2	10.53
7	Aliphatic amine, primary+Amine, primary+Ethe	1	5.26
8	Aliphatic amine, primary+Amine, primary+Ove	9	47.37
9	Aliphatic amine, tertiary+Amine, tertiary+Overl	2	10.53

Organic functional groups (nested)

Aliphatic amine, primary+Amine, primary+Overlapped groups

1. Aliphatic amine fragment is available.

Phase I: Evaluation of Category & Data Gap Filling

Step 1. Investigating the structural consistency of an ad hoc category

Recap

- **Chemicals are defined as:**
 - Aliphatic amines (broader category than primary amines)
 - US-EPA categories, ECOSAR classification
 - The statistics of organic functional groups provide detailed alert description of all 19 structures. However, all 19 chemicals have aliphatic amines fragment.
- **It could be concluded that the category is consistent with respect to structural functionalities (chemicals are empirically similar).**

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Step 2: Investigating the applicability domain of an ad hoc category.*

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

- Boundaries of structural functionalities
 - Aliphatic amines
- Parametric boundaries
 - log Kow (from 0.64 to 7.71)
 - Molecular weight (from 31 to 269 Da)
 - Water solubility (from 0.48×10^{-1} to 1×10^6 mg/l)

You are now ready to extract the 2D and/or 3D parameters (see next screen shot).

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

Extracting 2D and 3D parameters

The screenshot shows the QSAR Toolbox interface. On the left is the 'Filter endpoint tree...' panel with a list of endpoints. A red circle highlights '(Q) Acidic pKa (Chemaxon)'. A context menu is open over this endpoint, with a red circle around the option 'Calculate/extract all 2D parameters for all chemicals'. A callout '1' points to the endpoint in the tree, '2' points to the menu option, and '3' points to the 'Calculate/extract all 2D parameters for all chemicals' option. In the background, a 'Calculate structure' dialog box is open, with a callout '4' pointing to it. A warning message box on the right contains the text: 'If some of the parameters (in case of new chemicals) are not calculated then the message (4) appears. If you want to stop the calculation process, click the **Cancel** button.'

Structure	1	2	3	4
Structure info				
CAS Number				75-31-4
CAS Smiles				High
Chemical name				2-amino-propane
Composition				(tert)butylamini
Metabolism				
Formula	C4H11N	CH5N	C2H7N	C3H9N
Substance type	Mono constituent	Mono constituent	Mono constituent	Mono constituent
Formula	CCCCN	CN	CCN	CC(C)N
2D	Not calculated	Not calculated	Not calculated	Not calculated
(Q) Acidic pKa (Chemaxon)	Not calculated	Not calculated	Not calculated	Not calculated
(Q) Basic pKa (Chemaxon)	Not calculated	Not calculated	Not calculated	Not calculated
BAF	Not calculated	Not calculated	Not calculated	Not calculated
BAF (lower trophic)	Not calculated	Not calculated	Not calculated	Not calculated
BAF (mid trophic)	Not calculated	Not calculated	Not calculated	Not calculated
BAF (upper trophic)	Not calculated	Not calculated	Not calculated	Not calculated
BCF	Not calculated	Not calculated	Not calculated	Not calculated
BCF (lower trophic)	Not calculated	Not calculated	Not calculated	Not calculated
BCF (mid trophic)	Not calculated	Not calculated	Not calculated	Not calculated
BCF (upper trophic)	Not calculated	Not calculated	Not calculated	Not calculated
Bio Half-Life	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation primary (Biowin 4)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation probability (Biowin 1)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation probability (Biowin 2)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation probability (Biowin 5)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation probability (Biowin 6)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation probability (Biowin 7)	Not calculated	Not calculated	Not calculated	Not calculated
Biodegradation ultimate (Biowin 3)	Not calculated	Not calculated	Not calculated	Not calculated
DNA binding by OASIS	Not calculated	Not calculated	Not calculated	Not calculated
DNA binding by OECD	Not calculated	Not calculated	Not calculated	Not calculated
Estrogen Receptor Binding	Not calculated	Not calculated	Not calculated	Not calculated
Hydrolysis half-life (Ka, pH 7)(Hydrowin 1)	Not calculated	Not calculated	Not calculated	Not calculated
Hydrolysis half-life (Ka, pH 8)(Hydrowin 2)	Not calculated	Not calculated	Not calculated	Not calculated
Hydrolysis half-life (Kb, pH 7)(Hydrowin 3)	Not calculated	Not calculated	Not calculated	Not calculated
Hydrolysis half-life (Kb, pH 8)(Hydrowin 4)	Not calculated	Not calculated	Not calculated	Not calculated
Hydrolysis half-life (pH 6.5-7.4)	Not calculated	Not calculated	Not calculated	Not calculated
Ionization at pH = 1	Not calculated	Not calculated	Not calculated	Not calculated
Ionization at pH = 4	Not calculated	Not calculated	Not calculated	Not calculated
Ionization at pH = 7.4	Not calculated	Not calculated	Not calculated	Not calculated
Ionization at pH = 9	Not calculated	Not calculated	Not calculated	Not calculated
Protein binding by OASIS	Not calculated	Not calculated	Not calculated	Not calculated
Protein binding by OECD	Not calculated	Not calculated	Not calculated	Not calculated
Protein binding potency	Not calculated	Not calculated	Not calculated	Not calculated
Protein binding potency Cys (DPRA 13)	Not calculated	Not calculated	Not calculated	Not calculated
Protein binding potency Lys (DPRA 13)	Not calculated	Not calculated	Not calculated	Not calculated
Toxic hazard classification by Cramer	Not calculated	Not calculated	Not calculated	Not calculated
Toxic hazard classification by Cramer (2)	Not calculated	Not calculated	Not calculated	Not calculated
Endpoint Specific				
Acute aquatic toxicity classification by OECD	Not calculated	Not calculated	Not calculated	Not calculated
Acute aquatic toxicity MOA by OASIS	Not calculated	Not calculated	Not calculated	Not calculated
Metabolism/Transformations				
Boiling point	Not calculated	Not calculated	Not calculated	Not calculated

- 1. Right click** above the parameter in the endpoint tree; **2.** The user can calculate all parameters or extract current parameter for all chemicals in the current row. **3.** In this exercise calculate all (2D) parameters.

Phase I: Evaluation of Category & Data Gap Filling

Step 2. Investigating the applicability domain of an ad hoc category

Extracting 2D and 3D parameters - results

QSAR Toolbox 4.0.0.28533 [Document 1]

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

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

Documents

Profiling methods

Options

Select All Unselect All Invert

Predefined

- Database Affiliation
- Inventory Affiliation
- OECD HPV Chemical Categories
- Substance type
- US-EPA New Chemical Categories

General Mechanistic

- Biodeg BioHC half-life (Biowin)
- Biodegradation primary (Biowin 4)
- Biodegradation probability (Biowin 1)
- Biodegradation probability (Biowin 2)
- Biodegradation probability (Biowin 5)
- Biodegradation probability (Biowin 6)
- Biodegradation probability (Biowin 7)
- Biodegradation ultimate (Biowin 3)
- DNA binding by OASIS
- DNA binding by OECD
- Estrogen Receptor Binding
- Hydrolysis half-life (Ka, pH 7)(Hydrowin)
- Hydrolysis half-life (Ka, pH 8)(Hydrowin)
- Hydrolysis half-life (Kb, pH 7)(Hydrowin)
- Hydrolysis half-life (Kb, pH 8)(Hydrowin)
- Hydrolysis half-life (pH 6.5-7.4)
- Ionization at pH = 1
- Ionization at pH = 4
- Ionization at pH = 7.4
- Ionization at pH = 9
- Protein binding by OASIS
- Protein binding by OECD
- Protein binding potency
- Protein binding potency Cys (OPRA 13)
- Protein binding potency Lys (OPRA 13)
- Toxic hazard classification by Cramer
- Toxic hazard classification by Cramer (1)

Endpoint Specific

- Acute aquatic toxicity classification by
- Acute aquatic toxicity MOA by OASIS

Metabolism/Transformations

Filter endpoint tree...

Structure

Structure info

- CAS Number
- CAS Smiles relation
- Chemical name(s)
- Composition
- Molecular Formula
- Predefined substance type
- Structural Formula

Parameters

- 2D
 - (Q) Acidic pKa (Chemaxon)
 - (Q) Basic pKa (Chemaxon)
 - BAF
 - BAF (lower trophic)
 - BAF (mid trophic)
 - BAF (upper trophic)
 - BAF (upper trophic, biotransformation rate is ...)
 - BCF
 - BCF (lower trophic)
 - BCF (mid trophic)
 - BCF (upper trophic)
 - BCF (upper trophic, biotransformation rate is ...)
 - Bio Half-Life
 - Biodeg probability (Biowin 1)
 - Biodeg probability (Biowin 2)
 - Biodeg probability (Biowin 5)
 - Biodeg probability (Biowin 6)
 - Biodeg probability (Biowin 7)
 - BioHC Half-Life
 - Biotransformation Half-Life
 - Bollino point

	1	2	3	4	5
Structure	<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(C)N</chem>
CAS Number	109-73-9	74-89-5	75-04-7	75-31-0	75-64-9
CAS Smiles relation	High	High	High	High	High
Chemical name(s)	1-amino-butane	amino-methane	amino-ethane	2-amino-propane	(tert)butylamin
Composition					
Molecular Formula	C4H11N	CH5N	C2H7N	C3H9N	C4H11N
Predefined substance type	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constitue
Structural Formula	CCCCN	CN	CCN	CC(C)N	CC(C)(C)N
Parameters					
2D					
(Q) Acidic pKa (Chemaxon)	No value	No value	No value	No value	No value
(Q) Basic pKa (Chemaxon)	No value	No value	No value	No value	No value
BAF	0.25 log(L/kg bdwt)	-0.04 log(L/kg bdwt)	-0.02 log(L/kg bdwt)	0.02 log(L/kg bdwt)	0.05 log(L/kg b
BAF (lower trophic)	0.168 log(L/kg bdwt)	-0.02 log(L/kg bdwt)	-0.008 log(L/kg bdwt)	0.018 log(L/kg bdwt)	0.035 log(L/kg l
BAF (mid trophic)	0.186 log(L/kg bdwt)	-0.023 log(L/kg bdwt)	-0.01 log(L/kg bdwt)	0.019 log(L/kg bdwt)	0.038 log(L/kg l
BAF (upper trophic)	0.25 log(L/kg bdwt)	-0.039 log(L/kg bdwt)	-0.02 log(L/kg bdwt)	0.023 log(L/kg bdwt)	0.052 log(L/kg l
BAF (upper trophic, biotransformation rate is ...)	0.28 log(L/kg bdwt)	-0.035 log(L/kg bdwt)	-0.012 log(L/kg bdwt)	0.037 log(L/kg bdwt)	0.066 log(L/kg l
BCF	0.5 log(L/kg bdwt)	0.5 log(L/kg bdwt)	0.5 log(L/kg bdwt)	0.5 log(L/kg bdwt)	0.5 log(L/kg bd
BCF (lower trophic)	0.168 log(L/kg bdwt)	-0.02 log(L/kg bdwt)	-0.008 log(L/kg bdwt)	0.018 log(L/kg bdwt)	0.035 log(L/kg l
BCF (mid trophic)	0.186 log(L/kg bdwt)	-0.023 log(L/kg bdwt)	-0.01 log(L/kg bdwt)	0.019 log(L/kg bdwt)	0.038 log(L/kg l
BCF (upper trophic)	0.25 log(L/kg bdwt)	-0.039 log(L/kg bdwt)	-0.02 log(L/kg bdwt)	0.023 log(L/kg bdwt)	0.052 log(L/kg l
BCF (upper trophic, biotransformation rate is ...)	0.277 log(L/kg bdwt)	-0.035 log(L/kg bdwt)	-0.012 log(L/kg bdwt)	0.036 log(L/kg bdwt)	0.065 log(L/kg l
Bio Half-Life	0.215 d	0.0724 d	0.0963 d	0.125 d	0.175 d
Biodeg probability (Biowin 1)	0.975	0.887	0.88	0.873	0.683
Biodeg probability (Biowin 2)	0.993	0.975	0.97	0.964	0.795
Biodeg probability (Biowin 5)	0.677	0.653	0.661	0.52	0.597
Biodeg probability (Biowin 6)	0.808	0.796	0.8	0.616	0.643
Biodeg probability (Biowin 7)	0.694	0.934	0.96	0.688	0.44
BioHC Half-Life	No value	No value	No value	No value	No value
Biotransformation Half-Life	0.215 d	0.0724 d	0.0962 d	0.125 d	0.175 d
Bollino point	87.7 °C	10.2 °C	36.9 °C	47.3 °C	62.8 °C

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Step 2: Investigating the applicability domain of an ad hoc category.*
 - *Step 3: Reading data for the analogues*

Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

- Next, the Data matrix is constructed by extracting available experimental results for all 19 members of the category.
- Based on preceding category evaluation, no outliers have been identified violating the structural and mechanistic consistency of the category.

Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

The screenshot shows the QSAR Toolbox software interface. The 'Read data?' dialog box is open, with the 'All endpoints' radio button selected. The dialog has 'OK' and 'Cancel' buttons. A blue callout box with the number '3' points to the 'OK' button. In the background, a table displays chemical data for five compounds. A blue callout box with the number '2' points to the 'Gather' button in the top toolbar. Another blue callout box with the number '1' points to the 'Ecotoxicological Information' section in the left sidebar, which is circled in red.

Structure	1	2	3	4	5
Structure	<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(C)N</chem>
CAS Number	109-73-9	74-89-5	75-04-7	75-31-0	75-64-9
CAS Smiles relation	High	High	High	High	High
Chemical name(s)	1-amino-butane	amino-methane	amino-ethane	2-amino-propane	(tert)butylamine
Composition					
Molecular Formula	C4H11N	CH5N	C2H7N	C3H9N	C4H11N
Predefined			Mono constituent	Mono constituent	Mono constituent
Structure			CCN	CC(C)N	CC(C)N
Parameters					
Physical Chem					
Environment					
Ecotoxicology					
Human Health Hazards					
Profile					
Predefined					
US-EPA New Chemical Categories	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amine
Endpoint Specific					
Acute aquatic toxicity MOA by OASIS	Narcotic Amine				
Aquatic toxicity classification by ECOSAR	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amine
Empiric					
Organic functional groups (nested)	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine

1. Select databases related to aquatic toxicity; **2. Click** Gather; **3. Click** OK.

Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data.

Inserting data into the data matrix

QSAR Toolbox 4.0.0.28533 [Document 1]

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

Options

Select All Unselect All Invert

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
 - Aquatic ECETOX
 - Aquatic Japan MoE
 - Aquatic OASIS
 - ECHA CHEM
 - ECOTOX
 - n.f.
- Human Health Hazards

Inventories

Options

Select All Unselect All Invert

- Canada DSL
- COSING
- DSSTOX
- ECHA PR
- EINECS
- HPVC OECD
- Import_Custom Inventory_1
- MIETI Japan
- NICHAS
- REACH ECB
- TSCA
- US HPV Challenge Program

Filter endpoint tree...

Structure

Structure info

Parameters

Physical Chemical Properties

Environmental Fate and Transport (1/10)

Ecotoxicological Information

Aquatic Toxicity AW SW (16/306)

Sediment toxicity

Terrestrial Toxicity (5/11)

Human Health Hazards (1/1)

Profile

Predefined

US-EPA New Chemical Categories

Endpoint Specific

Acute aquatic toxicity MOA by OASIS

Aquatic toxicity classification by ECOSAR

Empiric

Organic functional groups (nested)

	1	2	3	4	5
Structure	<chem>CCCCN</chem>	<chem>CCN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)(Cl)C</chem>
Parameters					
Physical Chemical Properties					
Environmental Fate and Transport (1/10)		M: 3.46 log(L/kg bvd)			
Ecotoxicological Information					
Aquatic Toxicity AW SW (16/306)	M: 53 mg/L	M: 9E-09 ug/cell	M: 226 mg/L	M: 9.9 (7.1+13.6) mg	M: 16 mg/L
Sediment toxicity					
Terrestrial Toxicity (5/11)	M: 578 mg/L	M: 15 mmol			
Human Health Hazards (1/1)					
Profile					
Predefined					
US-EPA New Chemical Categories	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amine
Endpoint Specific					
Acute aquatic toxicity MOA by OASIS	Narcotic Amine				
Aquatic toxicity classification by ECOSAR	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amine
Empiric					
Organic functional groups (nested)	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine, prin	Aliphatic amine

Outlook

- Background
- Category evaluation – overview
- **Phase I. Evaluation of Category & Data Gap Filling**
 - *Step 1: Investigating the structural consistency of an ad hoc category.*
 - *Step 2: Investigating the applicability domain of an ad hoc category.*
 - *Step 3: Reading data for the analogues*
 - *Step 4: Data gap filling for 2-Butanamine*

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

Navigate to the target endpoint: Intoxication, D.magna, EC50, 48h

The screenshot displays the QSAR Toolbox interface. At the top, the 'Data Gap Filling' step is selected in the workflow. The 'Documents' panel on the left shows a tree view where 'Select from: Aliphatic amines' is highlighted with a callout '4'. The main workspace shows a hierarchical tree of 'Ecotoxicological Information' with 'Aquatic Toxicity' expanded to 'Intoxication' -> 'EC50' -> '48 h' -> 'Branchiopoda (branchiopods)' -> 'Daphnia magna'. A callout '2' points to this path. The data table on the right shows chemical structures and numerical values. A red box highlights a cell in the table with a callout '3'. A callout '1' points to the 'daphnia' filter field at the top of the table.

1. **Type** Daphnia in the filter field; 2. **Navigate** to target endpoint by opening the nodes: EC50, intoxication, 48h
3. **Highlight** the gap under the target chemical which will be filled in; 4. **Select** Trend analysis

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

Navigate to the target endpoint: Intoxication,

The screenshot shows the QSAR Toolbox interface during the 'Data Gap Filling' step. A dialog box titled 'Possible data inconsistency' is displayed, allowing the user to select units for data. The dialog includes the following options:

- Native scale/unit:**
 - mg/L (8 data; 8 chemicals)
 - mol/L (1 data; 1 chemicals)
 - µg/L (2 data; 1 chemicals)
- Gap filling scale/unit:**
 - log(1/mol/L)
 - mol/L
 - mg/L
 - µg/L

The dialog also shows 'Data 11/11; Chemicals 10/10' and 'OK' and 'Cancel' buttons. In the background, a data table is visible with columns for chemical structures and numerical values. A blue highlight is present on a cell containing 'M: 9.4 (6.9+12) mg/L'.

The user will be informed If there is different experimental data

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

Navigate to the target endpoint: Intoxication, D.magna, EC50, 48h

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' step. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' with 'Aliphatic amines' selected and 'Enter GF(IA) with 11 chemicals, 11 data points'. The central 'Filter endpoint tree' shows a tree structure with 'Intoxication' and 'EC50' selected. A table below the tree shows results for 'Daphnia magna' and 'Animalia (animals)'. A red arrow points to a data point in the table. A blue callout box highlights the text 'There is helpful information concerning investigated chemicals'. The bottom section shows a 'Trend analysis prediction for EC50, based on 10 values' with a predicted value of 60.2 mg/L and a model equation: $EC50 = 2.70 (\pm 0.669) + 0.506 (\pm 0.221) * \log Kow, \log(1/mol/L)$. The plot shows a scatter of data points with a red regression line. The bottom right corner has a 'Accept prediction' button with a green checkmark.

There is helpful information concerning investigated chemicals

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/Subcategorize by OFG(nested)

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' step. The main window shows a 'Filter endpoint tree...' dialog with a list of chemical categories and their associated data points. A 'Subcategorization' dialog is open, showing a list of 'Organic functional groups (nested)' and a 'Remove' button. A red box highlights the 'Remove' button, with a '1' next to it. The 'Trend analysis prediction for EC50' graph shows a scatter plot of EC50 (log1/mol/L) versus log Kow (log1/mol/L) with a red regression line. A chemical structure of 2-butanamine is highlighted in the graph. The 'Documents' panel on the left shows a list of documents, including 'Subcategorization'.

1. Remove dissimilar chemicals by clicking on OFG (nested) then **2. Click** Remove button

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/Cumulative frequency

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

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

Filter endpoint tree... 12 [target] 1 2 9 13

Structure

Ecotoxicological Information

Aquatic Toxicity (4/20)

Behavior (4/20)

Intoxication

EC50

24 h

Animalia (animals)

Arthropoda (arthropods)

Branchiopoda (branchiopods)

Daphnia magna (8/9)

48 h

Animalia (animals)

Arthropoda (arthropods)

M: 71 mg/L M: 141 mg/L M: >40+50 mg

M: >180 ppm M: >100 mg/L M: 850 (690+1.05E+ M: >8.3 mg/L M: 20 (8.8+84)

1

95% of Residuals ≤ 0.243, log(1/mol/L)

Cumulative frequency [%]

Select / filter data

Subcategorize

Mark chemicals by WS

Mark chemicals by descriptor value

Mark outliers

Filter points by test conditions

Mark focused chemical

Mark focused points

Remove marked data

1. Cumulative frequency is less than 0.24 log units

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/*Statistics*

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' step. The top navigation bar shows the workflow: Input, Profiling, Data, Category definition, Data Gap Filling (active), and Report. The left sidebar contains 'Documents' and 'Data Gap Filling Settings'. The 'Data Gap Filling Settings' panel has checkboxes for 'Only endpoint relevant' and 'Only chemical relevant', and a section 'At this position:' with options for 'Automated workflows' and 'Standardized workflows'. The main area is divided into a 'Filter endpoint tree...' on the left and a data table on the right. The tree shows 'Ecotoxicological Information' with 'Aquatic Toxicity' expanded to 'Intoxication' and 'EC50'. The data table shows chemical structures and toxicity data for 12 target endpoints. The bottom table displays statistical characteristics and model parameters for the selected endpoint. A red callout bubble with the number '1' points to the 'Statistics' section in the bottom table.

Filter endpoint tree...	12 [target]	1	2	9	13
Structure	<chem>CC(C)CN</chem>	<chem>CCCCN</chem>	<chem>CN</chem>	<chem>CCCCCN</chem>	<chem>CCCCCN</chem>
Ecotoxicological Information					
Aquatic Toxicity (4/20)					
Behavior		M: 71 mg/L	M: 141 mg/L		M: >40+50 mg
Intoxication					
EC50					
24 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna (8/9)	M: >180 ppm	M: >100 mg/L	M: 850 (690+1.05E+)	M: >8.3 mg/L	M: 20 (8.8+84)
48 h					
Animalia (animals)					
Arthropoda (arthropods)					

Descriptors	Statistical characteristics	TA model
Prediction	Number of data points, (N)	7
	Coefficient of determination, (R2)	0.974
	Adjusted coefficient of determination, (R2adj)	0.969
Adequacy	Coefficient of determination - leave one out, (Q2)	N/A
	Sum of squared residuals, (SSR)	0.220
Cumulative frequency	Standard deviation of residuals, (sN)	0.177
	Standard deviation of residuals, (s)	0.210
Residuals	Function, (F)	189
	Fisher threshold for statistical significance, (Fa)	10.6 (95.0%)
Statistics		
	b0	
	- model descriptor	Intercept
	- coeff. value	2.34
	- coeff. range	±0.337
	- significance	No

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/Interpretation of the result

- The structurally similar analogs across category of aliphatic amines is used for data gap filling
- Subcategorization by Organic functional groups(nested) is applied
- The prediction based on the defined category is acceptable.
- The predicted value based on predefined category of aliphatic amines is 79.0 mg/l

Outlook

- Background
- Category evaluation – overview
 - Phase I. Evaluation of Category & Data Gap Filling
 - **Phase II. Extension of the Category& Data gap Filling**

Phase II: Extension Category & Data gap filling

- The extension of the category is performed by using Phase II of the category evaluation process (*Extension of Category & Data Gap Filling*). Other analogues are searched in the Toolbox, which are structurally and mechanistically consistent with the predefined category.
- The structural analogues could be defined using ECOSAR grouping

Phase II: Extension Category & Data gap filling

Step 1: Category definition

The screenshot shows the QSAR Toolbox software interface. A red callout box with the number '1' points to the 'Input' button in the top toolbar. A second red callout box with the number '2' points to the 'Aliphatic amines' entry in the left-hand 'Documents' panel. The main window displays a table with columns for chemical structures and various ecotoxicological parameters.

Structure	10	11	12	13	14
Structure	<chem>CCCCN</chem>	<chem>CCCCCN</chem>	<chem>CCCCCN</chem>	<chem>CCCCCN</chem>	<chem>CCCCCN</chem>
Ecotoxicological Information					
Aquatic Toxicity					
Behavior					
Immobilisation					M: >20+30 mg/L
Intoxication					
EC50					
24 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna	(10/11) mg/L		M: >180 ppm		M: 20 (8.8+84) mg/L
48 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna	(10/11) .5+2.4 mg/L				M: 9.4 (6.9+12) mg/L
EC50/	(1/2)				
IC50	(2/2)				
LC0	(3/3)				
LC100	(3/3)				
LC50	(3/3)				
NOEL	(1/1)		M: 180 ppm		
Mortality	(15/109) mg/L		M: 275 (250+301) m		M: 155 mg/L
Reproduction	(1/2)				

1. Go back to Input section; 2. Select the row with Aliphatic amines

Phase II: Extension Category & Data gap filling

Step 1: Category definition

- Before defining the category, one should specify the endpoint - EC 50 48h *D.magna*

The screenshot displays the QSAR TOOLBOX software interface. The main window shows a tree view of 'Ecotoxicological Information' under 'Aquatic Toxicity'. The tree structure is as follows:

- Ecotoxicological Information
 - Aquatic Toxicity
 - Behavior
 - Immobilisation
 - Intoxication
 - EC50
 - 24 h
 - Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchiopods)
 - Daphnia magna
 - 48 h
 - Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchiopods)
 - Daphnia magna
 - EC5
 - IC50
 - LC0
 - LC100
 - LC50
 - NOEL

A context menu is open over the 'Daphnia magna' node, with the 'Target endpoint' option selected. A blue callout box with the number '2' points to this option. Another blue callout box with the number '1' points to the 'Daphnia magna' node in the tree view.

The 'Define endpoint' dialog box is open, showing the following settings:

- Effect: Intoxication
- Endpoint: EC50
- Duration: (a,b) 48 h
- Test organisms (species): Daphnia magna

Buttons in the dialog include 'Add', 'Up', 'Down', 'Clear', 'Remove', 'Back', and 'Finish'. The 'Undefined' button is also visible.

1. Right click above "Daphnia magna"; **2. Select** Target endpoint/Define;

Phase II: Extension Category & Data gap filling

Step 1: Category definition

- Before defining the category, one should specify the endpoint - EC 50 48h *D.magna*

The screenshot shows a 'Define endpoint' dialog box with the following configuration:

- Ecotoxicological Information:** Aquatic Toxicity
- Effect:** Intoxication
- Endpoint:** EC50
- Duration:** 48 h
- Test organisms (species):** Daphnia magna
- Additional metadata fields:** (empty list)

Buttons at the bottom include: Undefine, Back, and Finish.

1. Confirm the endpoint by click "Finish"

Phase II: Extension Category & Data gap filling

Step 1: Category definition

- Before defining the category, one should specify the endpoint - EC 50 48h *D.magna*

The screenshot displays the QSAR Toolbox interface during the 'Category definition' phase. The left sidebar shows a document tree with 'Aliphatic amines' selected. The main window is divided into a tree view on the left and a data table on the right. The tree view is expanded to show the endpoint 'EC50' > '48 h' > 'Branchiopoda (branchiopods)' > 'Daphnia magna'. The data table below shows the corresponding data for this endpoint, with the row for 'M: 1.9 (1.5-2.4) mg/L' highlighted in yellow. A text box at the bottom of the screenshot states: 'The row with defined endpoint is highlighted'.

Endpoint	9	10	11	12	13
Behavior (7/30)					M: >20
Immobilisation (1/1)					
Intoxication					
EC50					
24 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna (10/11)					M: >180 ppm
Daphnia magna (10/11)					M: 20
48 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna (10/11)					M: 1.9 (1.5-2.4) mg/L
Daphnia magna (10/11)					M: 9.4
EC50/ (1/2)					
IC50 (2/2) µg/L					
LC0 (3/3)					
LC100 (3/3)					
LC50 (3/3)					
NOEL (1/1)					M: 180 ppm
Mortality (15/109)					M: 278 mg/L
Reproduction (1/2)					M: >20

Phase II: Extension Category & Data gap filling

Step 1: Category definition

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Data Gap Filling' (highlighted with a red '3') and 'Report'. The toolbar contains icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Composition', 'Select', 'Delete', 'ChemIDs', 'Database', 'Inventory', 'List', 'Substructure (SMARTS)', 'Query', and 'Target'. The main workspace is divided into a left sidebar with a tree view of 'Ecotoxicological Information' and a central data table. The tree view shows 'Aquatic Toxicity' expanded to 'Intoxication' and 'EC50', with 'Daphnia magna' selected. The data table has columns for chemical structures (9-13) and rows for various toxicity endpoints. A context menu is open over the '2-Butanamine' entry in the table, with callouts 1, 2, and 3 indicating the steps: 1. Right click, 2. Select 'Set as target', 3. Go to 'Data Gap Filling'.

Structure	9	10	11	12	13
Structure	<chem>CCCCCCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>
Ecotoxicological Information					
Aquatic Toxicity					
Behavior					
Immobilisation					
Intoxication					
EC50					
24 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna					
48 h					
Animalia (animals)					
Arthropoda (arthropods)					
Branchiopoda (branchiopods)					
Daphnia magna					
EC50/					
IC50					
LC0					
LC100					
LC50					
NOEL					
Mortality					
Reproduction					

1. Right click above the target chemical (2-Butanamine); **2. Select Set as target**; **3. Go to Data Gap Filling**

Phase II: Extension Category & Data gap filling

Step 1: Category definition

The screenshot displays the QSAR Toolbox interface during the 'Data Gap Filling' phase. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' workflow is active, showing a 'Filter endpoint tree...' on the left and a '1 [target]' panel on the right. The 'Filter endpoint tree...' is expanded to 'Aquatic Toxicity', which includes endpoints like 'Intoxication', 'EC50', 'Mortality', and 'NOEL'. The '1 [target]' panel shows the chemical structure of 2-Butanamine (CCCC(N)C) and a table of data points for various endpoints. A text box at the bottom of the interface states: "The target chemical 2-Butanamine is loaded in new data matrix".

Phase II: Extension Category & Data gap filling

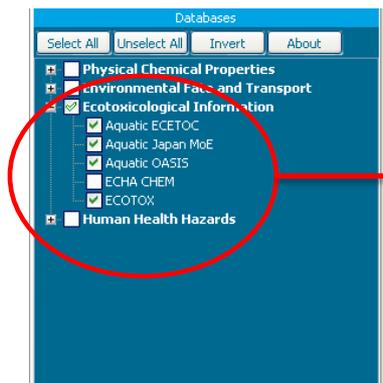
Step 1: Category definition

- The ECOSAR strict category is used to define a broader category used in further analysis.
- The same endpoint: EC 50 48h *D.magna* will be predicted as with the predefined category

Phase II: Extension Category & Data gap filling

Step 1: Category definition

- Define ECOSAR category
- Before defining the category, the following databases related to the predicted endpoint are selected:



Aquatic ECETOC
 Aquatic Japan MoE
 Aquatic OASIS
 ECOTOX

Phase II: Extension Category & Data gap filling

Step 1: Category definition/Defining ECOSAR (strict)

The screenshot displays the QSAR Toolbox software interface during the 'Categorize' phase. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Categorize' sub-menu is open, showing 'Define', 'Define with metabolism', 'Subcategorize', and 'Combine'. A callout '2' points to the 'Define' button. On the left, the 'Documents' pane shows a list of endpoints under 'Aquatic toxicity classification by ECOSAR'. A callout '1' points to 'Aquatic toxicity classification by ECOSAR'. The 'Filter endpoint tree...' dialog is open, showing a tree structure with 'Aquatic Toxicity' selected. A callout '3' points to the 'Strict' radio button under 'Combine profiles'. The 'Target' dialog is also open, showing 'Aliphatic Amines' as the target. A callout '4' points to the 'OK' button.

1. Highlight "Aquatic toxicity classification by ECOSAR"; **2. Click** Define; **3. Select** Strict. **4. Click** OK

Phase II: Extension Category & Data gap filling

Step 1: Category definition/Defining ECOSAR (strict)

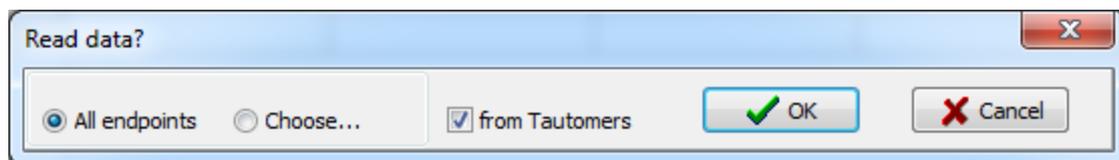
The screenshot shows the QSAR Toolbox interface during the 'Category definition' step. A dialog box titled 'Grouping results' is open, displaying the message '371 chemicals found.' and an 'OK' button. A blue callout box with the number '1' points to the 'OK' button. The background interface includes a 'Filter endpoint tree...' panel on the left, a 'Documents' panel with 'Aquatic toxicity classification by ECOSAR' selected, and a main table with columns for chemical structures and data. The table shows chemical structures in the first column and data like '(1/1) M: 180 ppm' in the second column.

1. Information that 370 analogs are found. Click OK

Phase II: Extension Category & Data gap filling

Step 1: Category definition/Reading data

- The Toolbox will now retrieve those chemicals that have the same ECOSAR functionality as the target compound.
- 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).

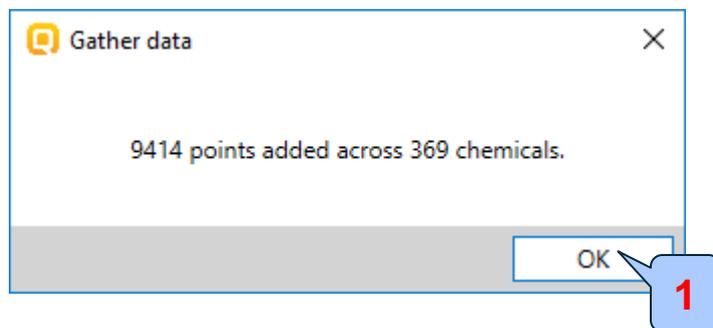


- In this example, as only databases are selected that contain information for aquatic toxicity endpoint, both options give the same results.

Phase II: Extension Category & Data gap filling

Step 1: Category definition/Reading data

The system automatically gives indication for the number of gather experimental data points



1. **Click OK**

Outlook

- Background
- Category evaluation – overview
 - Phase I. Evaluation of Category & Data Gap Filling
 - **Phase II. Extension of the Category& Data gap Filling**
 - Step 1: Category definition
 - Step 2: Navigate to the target endpoint – this is already done, so it is skipped

Outlook

- Background
- Category evaluation – overview
 - Phase I. Evaluation of Category & Data Gap Filling
 - **Phase II. Extension of the Category& Data gap Filling**
 - Step 1: Category definition
 - Step 2: Navigate to the target endpoint
 - Step 3: Data Gap Filling

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' phase. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows 'Documents' and 'Data Gap Filling Settings'. The main area is divided into a 'Filter endpoint tree...' on the left and a data table on the right. A red circle with the number '1' highlights the 'Trend analysis' button in the top-left corner of the main area.

Filter endpoint tree...

- Structure
- Ecotoxicological Information
 - Aquatic Toxicity **AW SW**
 - Accumulation (11/23)
 - Avoidance (4/4)
 - Behavior (50/353)
 - M: 10 mg/L
 - Biochemistry (27/280)
 - M: 10 mg/L
 - Cell(s) (15/35)
 - M: 500+8E+03 mg/L
 - Development (30/118)
 - Ecosystem Process (1/4)
 - M: 1+102 mg/L
 - Enzyme(s) (16/155)
 - Feeding Behavior (13/38)
 - M: 435 mg/L
 - Genetics (22/360)
 - Growth (92/321)
 - M: 1 mg/L
 - Growth Inhibition (33/66)
 - Histology (19/78)
 - Hormone(s) (9/93)
 - Immobilisation (36/37)
 - Immunological (3/3)
 - Injury (9/53)
 - Intoxication
 - EC50
 - 24 h (40/46)
 - M: 96 (89+10)
 - 48 h
 - Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchiopods)
 - Ceriodaphnia dubia (4/6)
 - Daphnia magna (59/68)
 - M: 25.2 (19.6+31.7) ppm
 - Daphnia pulex (1/2)
 - Insecta (insects) (1/1)
 - Malacostraca (1/5)
 - Maxillopoda (1/1)
 - M: 1.2 (1+1.4)
 - Mollusca (molluscs,mollusks) (2/5)

Data Gap Filling Settings

- Only endpoint relevant
- Only chemical relevant

At this position:

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

1. Select Trend analysis

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling

Possible data inconsistency

Native scale/unit

- mg/L (35 data; 35 chemicals)
- mol/L (4 data; 4 chemicals)
- ppb (5 data; 5 chemicals)
- ppm (21 data; 14 chemicals)
- μM (1 data; 1 chemicals)
- μg/L (2 data; 1 chemicals)

Gap filling scale/unit

- log(1/mol/L)
- mg/L
- mol/L
- ppb
- ppm
- μM
- μg/L

Data 42/68; Chemicals 41/59

OK Cancel

The user will be informed If there is different experimental data. Click OK

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling

The screenshot displays the QSAR Toolbox interface during the Data Gap Filling step. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The workflow menu shows 'Gap Filling' and 'Workflow' options. The main workspace is divided into several panels:

- Filter endpoint tree...:** A tree view showing various endpoints such as IC50, LC0, LC100, LC50, LOEC, NOEC, NOEL, Morphology, Mortality, and Reproduction.
- Data Table:** A table with columns for 'Structure', 'M', and 'M:'. A red arrow points from a data point in the table to the trend analysis plot.
- Trend analysis prediction for EC50, based on 41 values:** A scatter plot showing the relationship between log Kow and EC50. The predicted value is 34.8 mg/L. The model equation is $EC50 = 2.93 (\pm 0.397) + 0.518 (\pm 0.102) * \log Kow, \log(1/mol/L)$.
- Warning Boxes:** Several yellow boxes provide information about endpoint values and gap filling states. For example, 'There are 4 endpoint values for 4 chemicals bigger than WS calculated by "Water Solubility (fragments)"' and 'The current gap filling state contains data with qualifiers'.
- Callout Box:** A blue box at the bottom of the screenshot states: 'There is helpful information concerning investigated chemicals'.
- Accept Prediction:** A green checkmark and the text 'Accept prediction' are visible in the bottom right corner.

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Subcategorize by Water solubility

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' phase. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' and 'Data Gap Filling Settings'. The central area features a 'Filter endpoint tree' and a 'Structure' view. The main data table displays chemical structures and their properties, with a scatter plot at the bottom showing 'log Kow' vs. 'log Kow'. A dialog box 'Select water solubility calculators' is open, showing 'Water Solubility (fragments)' selected. Numbered callouts (1-4) highlight key actions: 1. 'Mark chemicals by WS' button in the right sidebar; 2. Selection of 'Water Solubility (fragments)' in the dialog; 3. 'OK' button in the dialog; 4. Notifying helper messages at the top right of the data table.

1. Click Mark chemical by WS;
2. Select Water solubility (fragments);
3. Click OK.
4. The same steps will be reproduced if you click on the link available in the notifying helper

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Subcategorize by Water solubility

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' phase. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' with a list of chemical groups and 'Data Gap Filling Settings' with options for endpoint and chemical relevance. The central area features a 'Filter endpoint tree...' panel on the left and a main data table on the right. The table lists chemical structures and their corresponding endpoint values. An 'Information' dialog box is open, displaying the message 'Marked 4 data point(s) for 4 chemical(s)'. Below the table, a 'Trend analysis prediction for EC50' graph is shown, with a red circle highlighting four data points. The 'Subcategorize' panel on the right contains buttons for marking chemicals by water solubility (WS), descriptor value, outliers, and focused chemicals, as well as options to remove marked data or clear existing marks. Callouts 1, 2, and 3 indicate the steps: 1. Marking chemicals in green in the table; 2. Clicking 'OK' in the 'Information' dialog; 3. Clicking 'Remove marked data' in the 'Subcategorize' panel.

1. Four chemicals are marked in green; 2. Click OK; 3. Remove marked chemicals

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Subcategorize by OFG(nested)

The screenshot displays the QSAR Toolbox software interface during the subcategorization phase. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows the 'Subcategorization' dialog box with various options categorized into Empiric, Toxicological, and Custom. A callout box labeled '2' points to the 'Organic functional groups (nested)' option. The central area shows a table of chemical structures and their corresponding data points. A callout box labeled '1' points to the 'Subcategorize' button in the right sidebar. The bottom plot shows a trend analysis prediction for EC50 based on 37 values, with a predicted value of 38.5 mg/L. A callout box labeled '3' points to the 'Remove selected' button in the bottom right corner of the dialog box.

1. Click Subcategorize; 2. Select Organic functional groups(nested); 3. Remove dissimilar chemicals

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Subcategorize by Lipinski rules

The screenshot displays the QSAR Toolbox software interface during the subcategorization phase. The main window shows a table of chemical data with columns for various properties. A 'Subcategorization' dialog box is open on the left, with 'Lipinski Rule Based' selected under the 'Empirical' category (1). Below this, a list of 'Not bioavailable' chemicals is shown (2). A smaller window displays four chemical structures (3). At the bottom, a scatter plot shows EC50 (log(L/m)) values, with a red circle highlighting outliers (4).

Chemicals with very long chain could be removed from the category due to their non-bioavailability. **1. Select** Lipinski rules; **2. Double click** to see "Not- bioavailable" chemicals; **3. Close** the appeared window; **4. Remove** dissimilar chemicals

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Prediction result

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

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

Trend analysis Read across (Q)SAR Standardized Automated

Documents

aliphatic amines
toxicity classification by ECOSAR
F(TA) with 42 chemicals, 42 data points
38] Data: 38 Filter by WS - Water Solubility (fragments
Ch: 16] Data: 16 Subcategorized: Organic functional g
Ch: 12] Data: 12 Subcategorized: Lipinski Rule

Filter endpoint tree...

Structure

Daphnia magna (11/12)

| | | | | |
|------------------------|------------------------|------------------------|------------------------|------------------------|
| 1 [target] | 92 | 128 | 177 | 180 |
| <chem>CC(C)N</chem> | <chem>CCCCCCCCN</chem> | <chem>CCCCCCCCN</chem> | <chem>CCCCCCCCN</chem> | <chem>CCCCCCCCN</chem> |
| M: 0.5 (0.4+0.61) mg/L | M: 0.026 (0.018+0.036) | M: 1.6 (1.2+2) mg/L | M: 1.9 (1.5+2.4) | |
| M: 180 ppm | M: 0.211 mg/L | M: 10 mg/L | M: 2.17 mg/L | M: 31 mg/L |
| M: 278 mg/L | M: 5 ppm | M: 5 ppm | | |
| | | M: 1.28 mg/L | | M: 0.937 mg/L |
| | | | M: 1 mL | M: 2 % |

LC0 (1/1)
LC100 (1/1)
LC50 (1/1)
NOEL (1/1)
Mortality (12/89)
No Effect Coded (2/6)
Physiology (4/28)
Population (5/29)
Sediment toxicity
Terrestrial Toxicity (7/28)
Human Health Hazards
Profile

Descriptors
Prediction
Adequacy
Cumulative frequency
Residuals
Statistics

Trend analysis prediction for EC50 values
Predicted: 77.8 mg/L
Model equation: $EC50 = 2.34 (\pm 0.324) + 0.838 (\pm 0.012) \cdot \log(1/\text{mol/L})$

Select / filter data

Subcategorize
Mark chemicals 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

Prediction result is 77.8 mg/l.

Phase II: Extension Category & Data gap filling

Step 3: Data Gap Filling/Accept the Prediction result

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' phase. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' and 'Data Gap Filling Settings'. The 'Data Gap Filling Settings' panel is active, showing options for 'Only endpoint relevant' and 'Only chemical relevant'. The 'Filter endpoint tree...' panel shows a tree structure with 'Daphnia magna' selected. The data table below shows results for 'Daphnia magna' with columns for 'M' and 'M' values. A 'Confirm' dialog box is open, asking 'Are you sure you want to accept this prediction?'. The scatter plot shows 'EC50 [log(1/mg/L)]' vs 'log Kow, log(1/mo/L)' with a regression line. The 'Select / filter data' panel is on the right. A blue banner at the bottom contains instructions: '1. Click on Accept prediction. 2. Click Yes'.

1. Click on Accept prediction. 2. Click Yes

Outlook

- Background
- Category evaluation – overview
 - Phase I. Evaluation of Category & Data Gap Filling
 - Phase II. Extension of the Category& Data gap Filling
- Save the prediction result

Saving the prediction result

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

Saving the prediction result

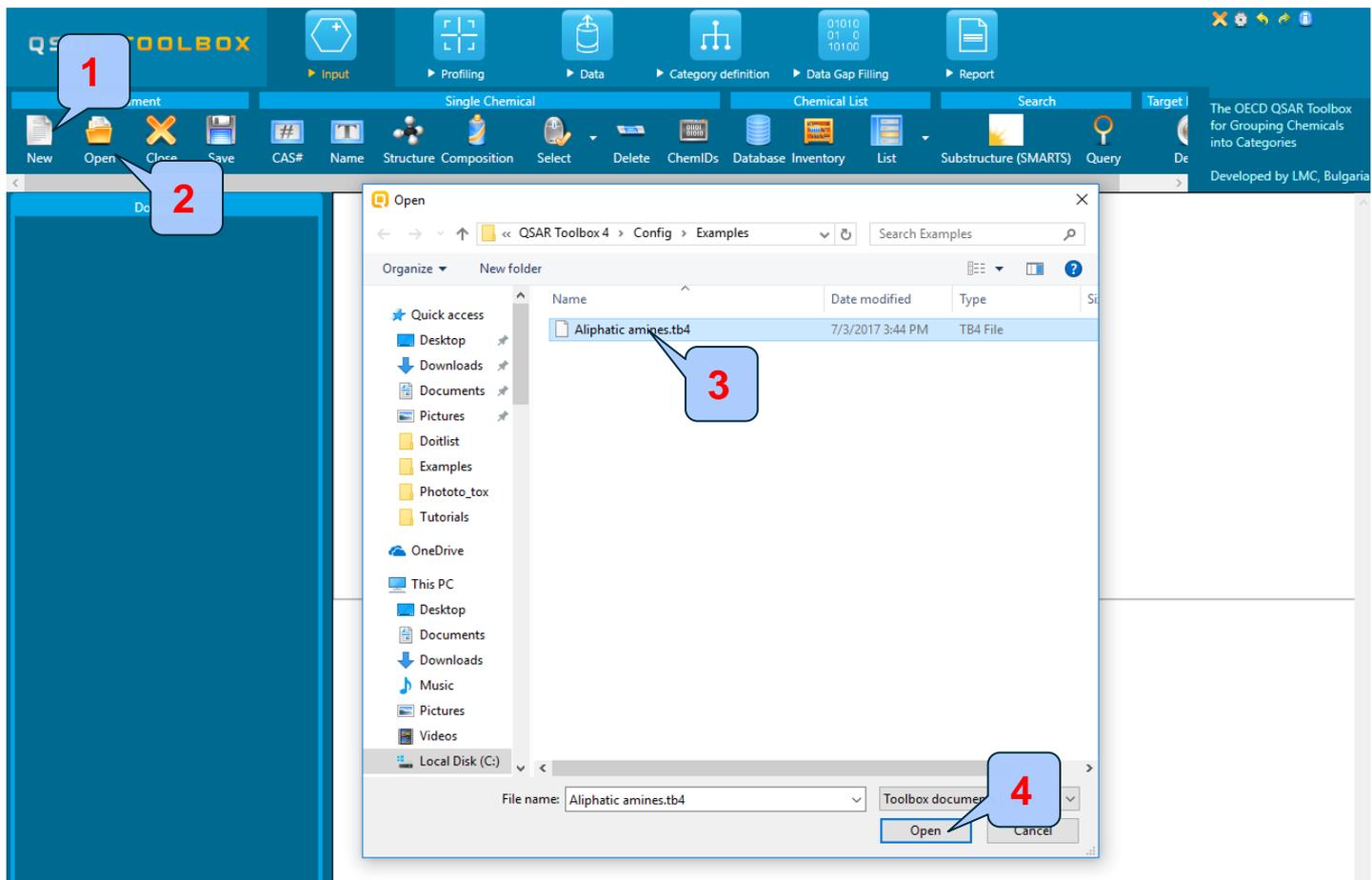
1. Go to Input section

2. Click on Save button

3. Define name of the file;

4. Click Save button

Open saved file



1. **Create** new document
2. **Click** Open;
3. **Find** and **select** file;
4. **Click** Open

Open saved file

The screenshot shows the QSAR Toolbox interface. At the top, there is a menu bar with options like 'Document', 'Single Chemical', 'Chemical List', 'Search', and 'Target'. Below the menu bar is a toolbar with icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Composition', 'Select', 'Delete', 'ChemIDs', 'Database', 'Inventory', 'List', 'Substructure (SMARTS)', and 'Query'. The main workspace is divided into several panels. On the left, there is a 'Documents' panel showing a list of files. In the center, there is a 'Filter endpoint tree...' panel with a tree view of endpoints. To the right of the tree is a data table with columns for 'Structure', 'Daphnia magna', and various endpoints. A 'File open' dialog box is open in the center, displaying the message 'File opened successfully!' and an 'OK' button. A red callout box with the number '1' points to the 'OK' button. At the bottom, there is a 'Trend analysis prediction' plot showing a scatter plot of EC50 (log(1/mo/L)) versus log Kow, with a red regression line. The plot is titled 'Trend analysis prediction based on 41 values' and shows a predicted value of 34.8 mg/L. A green checkmark and the text 'Accept prediction' are visible at the bottom right of the plot area.

1. The file is opened successfully 1. **Click OK**