

OECD QSAR Toolbox v.4.4.1

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

Outlook

- **Background**
- Keywords
- Category evaluation - overview

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox for evaluating an ad-hoc category.
- You will learn several new functionalities which will 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
- **Keywords**
- Category evaluation - overview

Keywords

TARGET CHEMICAL - chemical of interest

MODULE – a Toolbox module is a section dedicated to specific actions and options (e.g. Profiling)

WORKFLOW – the use, in combination, of the different modules (e.g. prediction workflow: from input to report)

PROFILER - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. Organic functional groups), mechanistic (e.g. Protein binding by OECD) and endpoint-specific (e.g. in vitro in vitro mutagenicity (Ames test) alerts by ISS) profilers.

ALERT - the profilers consist of sets of rules or alerts. Each of the rules consists of a set of queries. The queries could be related to the chemical structure, physicochemical properties, experimental data, comparison with the target or list with substances and external queries from other predefined profilers (reference queries).

CATEGORY – “group” of substances sharing same characteristics (e.g. the same functional groups or mode of action). In a typical Toolbox workflow, it consists of the target chemical and its analogues gathered according to the selected profilers

ENDPOINT TREE – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

DATA MATRIX – Table reporting the chemical(s) and data (experimental results, profilers outcomes, predictions). Each chemical is in a different column and each data in a different row

Outlook

- Background
- Keywords
- **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 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
- Keywords
- 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
- Keywords
- 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* from example directory).
- 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 4.4\Config\Examples

Outlook

- Background
- Keywords
- 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. Go to the **Input** module;

2. Click the **List** button and select "**From example directory**";

3. Browse and find the file for input;

4. Click **Open**.

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 displays the QSAR Toolbox interface. The left sidebar shows the 'Structure info' panel expanded, with a red '1' in a blue box highlighting it. The main window shows a table of 13 chemicals with their structures, names, and various identifiers.

1	2	3	4	5	6	7	8	9	10	11	12	13
EC Number:2036-109-73-9	EC Number:2008-74-89-5	EC Number:2008-75-04-7	EC Number:2008-75-31-0	EC Number:2008-75-64-9	EC Number:2011-78-96-6	EC Number:2032-104-75-6	EC Number:2036-108-91-8	EC Number:2039-111-86-4	EC Number:2171-1761-71-3	EC Number:2262-5332-73-0	EC Number:2377-13952-84-6	EC Number:2038-111-68-2
High	High	High	High	High	High	High	High	High	High	High	High	High
1-amino-butane	1S/CH5N/c1-2/h...	amino-ethane	1S/C3H9N/c1-3/h...	(tert)butylamine	1-Amino-2-prop...	1-Hexanamine, 2...	1-Aminocyclohe...	1-aminooctane	1,4-Bis(aminocyc...	1-Propanamine...	(+)-sec-Butylam...	1-aminoheptane
C4H11N	CH5N	C2H7N	C3H9N	C4H11N	C3H9NO	C8H19N	C6H13N	C8H19N	C13H26N2	C4H11NO	C4H11N	C7H17N
Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent
CCCCN	CN	CCN	CC(C)N	CC(C)(C)N	CC(O)CN	CCCC(CO)CN	NC1CCCCC1	CCCCCCCCN	NC1CCC(CC1)CC...	COCCCN	CCC(C)N	CCCCCCC

1. Expand "Structure info"

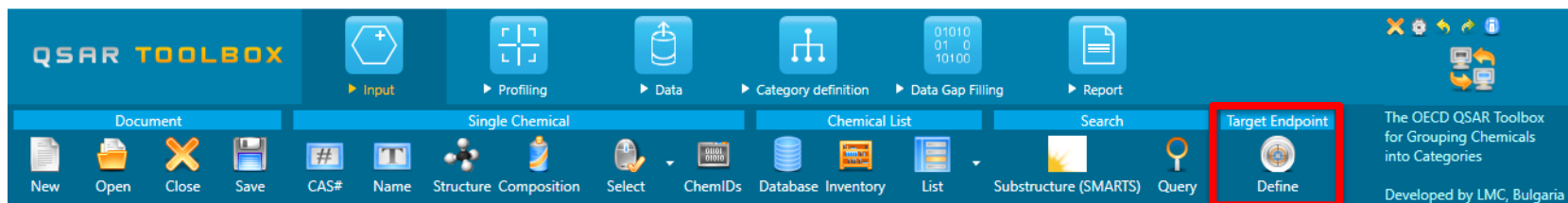
Outlook

- Background
- Keywords
- 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*
 - **Define target endpoint**

Input

Define target endpoint

- Defining of the endpoint allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc., along with specific metadata information. Based on the metadata, relevancy of the profiles and databases is provided expressed in different highlighting:
 - In green are highlighted the most suitable profilers related to the endpoint and databases including data for the defined target endpoint, while
 - In orange are colored profilers which are plausible with respect to the defined target endpoint.



Define target endpoint Overview

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. A 'Select endpoint' dialog box is open, showing a tree view of categories. 'Ecotoxicological Information' is selected, and 'Aquatic Toxicity' is highlighted. A 'Next' button is visible. A second 'Select endpoint' dialog box is open, showing 'Intoxication' as the effect, '48 h' as the duration, and 'Daphnia magna' as the test organism. The 'EC50' endpoint is selected. A 'Finish' button is visible. Numbered callouts 1 through 6 indicate the steps in the workflow.

1. Click on Target Endpoint and define;
2. Select Ecotoxicological Information, Aquatic toxicity;
3. Click **Next**;
4. Select endpoint **EC50** from the drop down menu; and consecutively select the following metadata
5. Test organism(spices) – **Daphnia magna**; Duration – **48 h**; Effect – **Intoxication**;
6. Click **Finish**.

Define target endpoint Overview

The screenshot displays the QSAR TOOLBOX interface with the 'Define' tab selected. The 'Filter endpoint tree...' panel on the left is expanded to show the hierarchy: Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, Aquatic Toxicity, Intoxication, 48 h, Animalia (animals), Arthropoda (arthropods), Branchiopoda (branchiopods), Daphnia magna, EC50. The main table shows 13 chemical structures and their corresponding endpoint data. The row for the 13th chemical is highlighted in yellow.

Structure	1	2	3	4	5	6	7	8	9	10	11	12	13
Structure	<chem>CCCCN</chem>	<chem>CCN</chem>	<chem>CCN</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>	<chem>CC(C)N</chem>
EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...	EC Number:2036...
CAS Number	109-73-9	74-89-5	75-04-7	75-31-0	75-64-9	78-96-6	104-75-6	108-91-8	111-86-4	1761-71-3	5332-73-0	13952-84-6	111-68-2
High	High	High	High	High	High	High	High	High	High	High	High	High	High
Chemical name(s)	1-amino-butane	1,5/CH3N/c1-2/h...	amino-ethane	1,5/C3H9N/c1-3/h...	(tert)butylamine	1-Amino-2-prop...	1-Hexanamine, 2...	1-Aminocyclohe...	1-aminooctane	1,4-Bis(aminocyc...	1-Propanamine,...	(+)-sec-Butylam...	1-aminoheptane
Molecular formula	C4H11N	CH5N	C2H7N	C3H9N	C4H11N	C3H9NO	C8H19N	C6H13N	C8H19N	C13H26N2	C4H11NO	C4H11N	C7H17N
Predefined substance type	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent
SMILES	CCCCN	CN	CCN	CC(C)N	CC(C)N	CC(C)N	CCCC(C)CN	NC1CCCC1	CCCCCCC	NC1CCC(CC1)CC...	CCCCCN	CCC(C)N	CCCCCCN

The endpoint tree is automatically expanded to the level of the defined endpoint and the row is highlighted in yellow

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 (hereafter cited as US-EPA)
 - Aquatic toxicity classification by ECOSAR (i.e. ECOSAR)
 - Aquatic toxicity MOA of action (i.e. MOA)
 - Organic functional groups (nested) (i.e. OFG(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

Profiling

1. Go to **Profiling** module;

2. Select suitable profiler - all green highlighted profilers;

3. Click **Apply** to apply the profilers to the list of chemicals. The results will appear on datamatrix (see next slide)

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Profiling Custom profile

Apply View New Delete

Documents

Document 1
[G: 19;Md: 0;P: 0] Aliphatic amines.smi

Filter endpoint tree...

Structure

Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information

Aquatic Toxicity
Intoxication
48 h
Animalia (animals)
Arthropoda (arthropods)
Branchiopoda (branchiopo...
Daphnia magna
EC50

Sediment Toxicity
Terrestrial Toxicity
Human Health Hazards
Profiling

Predefined
US-EPA New Chemical Categories
Endpoint Specific
Acute aquatic toxicity classification by...
Acute aquatic toxicity MOA by OASIS
Aquatic toxicity classification by ECOS...

Profiling methods

Options Select All Unselect All Invert 4 Selected

Suitable
[x] Acute aquatic toxicity classification by Verhaar (Modified)
[x] Acute aquatic toxicity MOA by OASIS
[x] Aquatic toxicity classification by ECOSAR
[x] US-EPA New Chemical Categories

Plausible
[x] Chemical elements
[x] Groups of elements
[x] Hydrolysis half-life (Ka, pH 7)(Hydrowin)
[x] Hydrolysis half-life (Ka, pH 8)(Hydrowin)
[x] Hydrolysis half-life (Kb, pH 7)(Hydrowin)
[x] Hydrolysis half-life (Kb, pH 9)(Hydrowin)
[x] Hydrolysis half-life (pH 6.5-7.4)
[x] Ionization at pH = 1
[x] Ionization at pH = 4
[x] Ionization at pH = 7.4
[x] Ionization at pH = 9
[x] Lipinski Rule Oass
[x] OECD HPV Chemical Categories

Metabolism/Transformations

1 2 3 4 5 6 7 8 9 10 11 12

Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines

Class 2 (less inert... Class 2 (less inert... Class 5 (Not pos... Class 5 (Not pos... Class 2 (less inert... Class 2 (less inert... Class 5 (Not pos... Class 2 (less inert... Class 2 (less inert... Class 2 (less inert... Class 5 (Not pos... Class 2 (less inert...

Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine Narcotic Amine

Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines

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 overall result for the list of chemicals can be seen by **right clicking** in the space near 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 Profiling

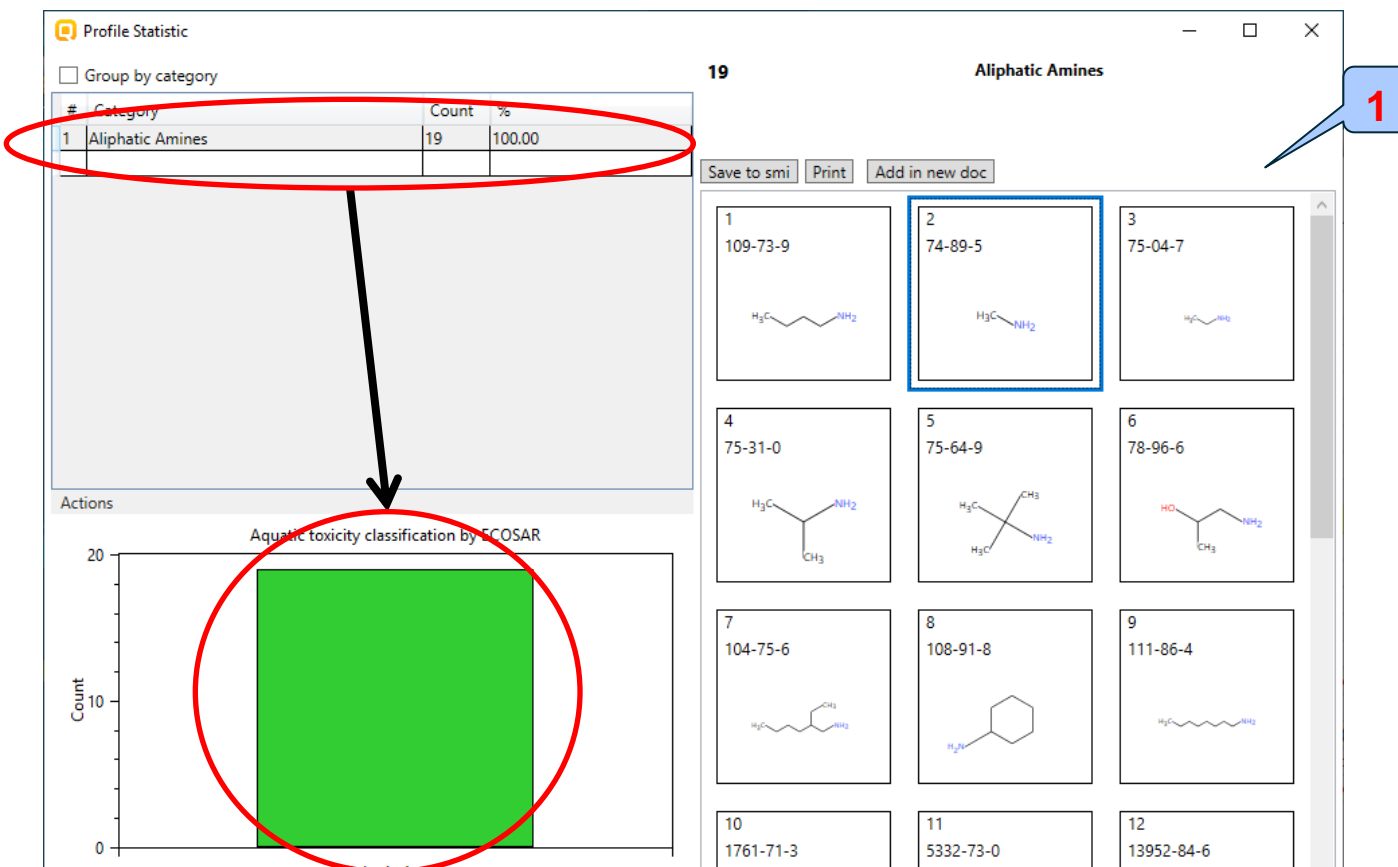
The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows 'Profiling methods' with a list of 5 selected methods: Acute aquatic toxicity classification by Verhaar, Acute aquatic toxicity MOA by OASIS, Aquatic toxicity classification by ECOSAR, US-EPA New Chemical Categories, and Plausible. The 'Filter endpoint tree' on the right shows a tree structure with 'US-EPA New Chemical Categories' highlighted. A context menu is open over this selection, with 'Profile Statistics' chosen. A third window, 'Profile Statistics', is open, showing a table with one category: 'Aliphatic Amines' with a count of 19 and 100.0%.

1. Right click above US-EPA profiler; 2. Select **Profile Statistics**; 3. A window appears with profile statistics for the list of chemicals across "US-EPA" profiler. All the chemicals are "Aliphatic amine"

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to ECOSAR



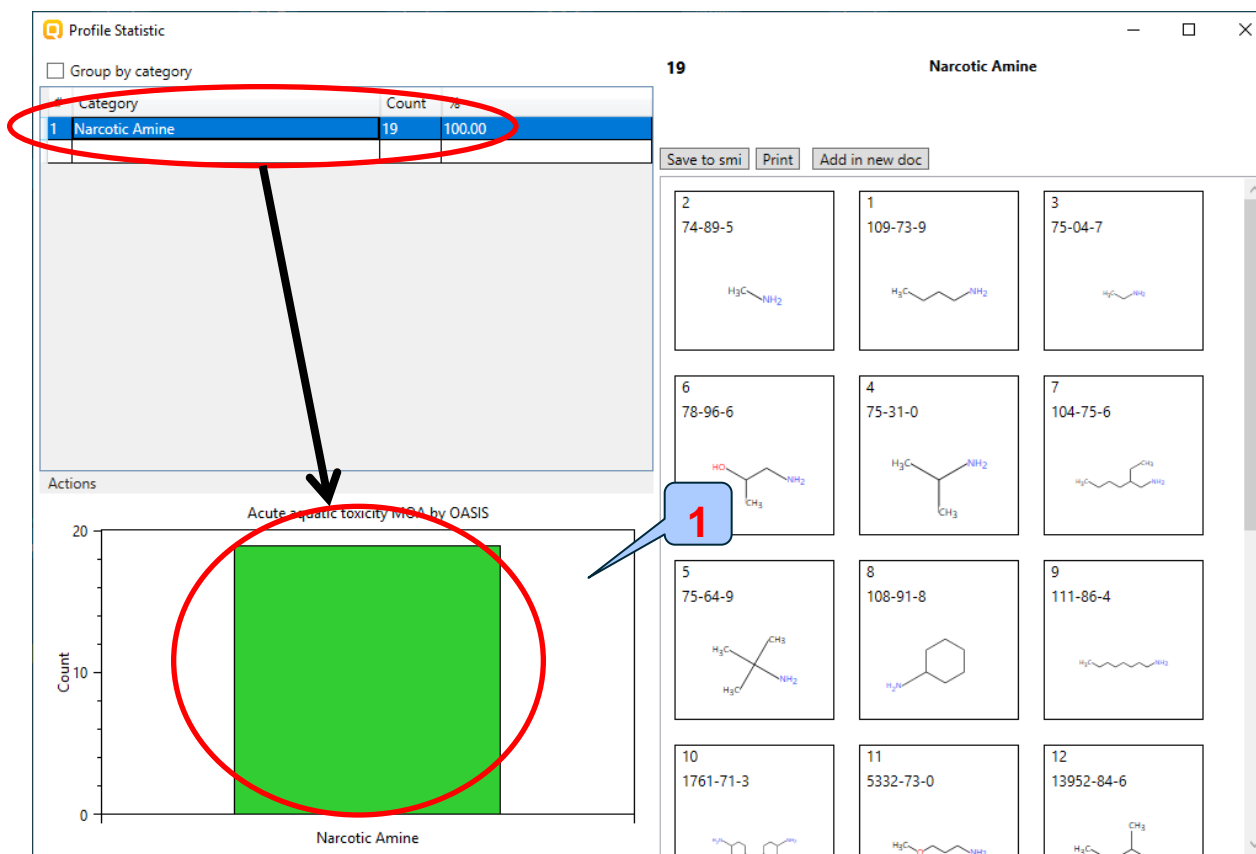
Repeat steps 1 and 2 from slide 24.

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

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to Aquatic toxicity MOA by OASIS

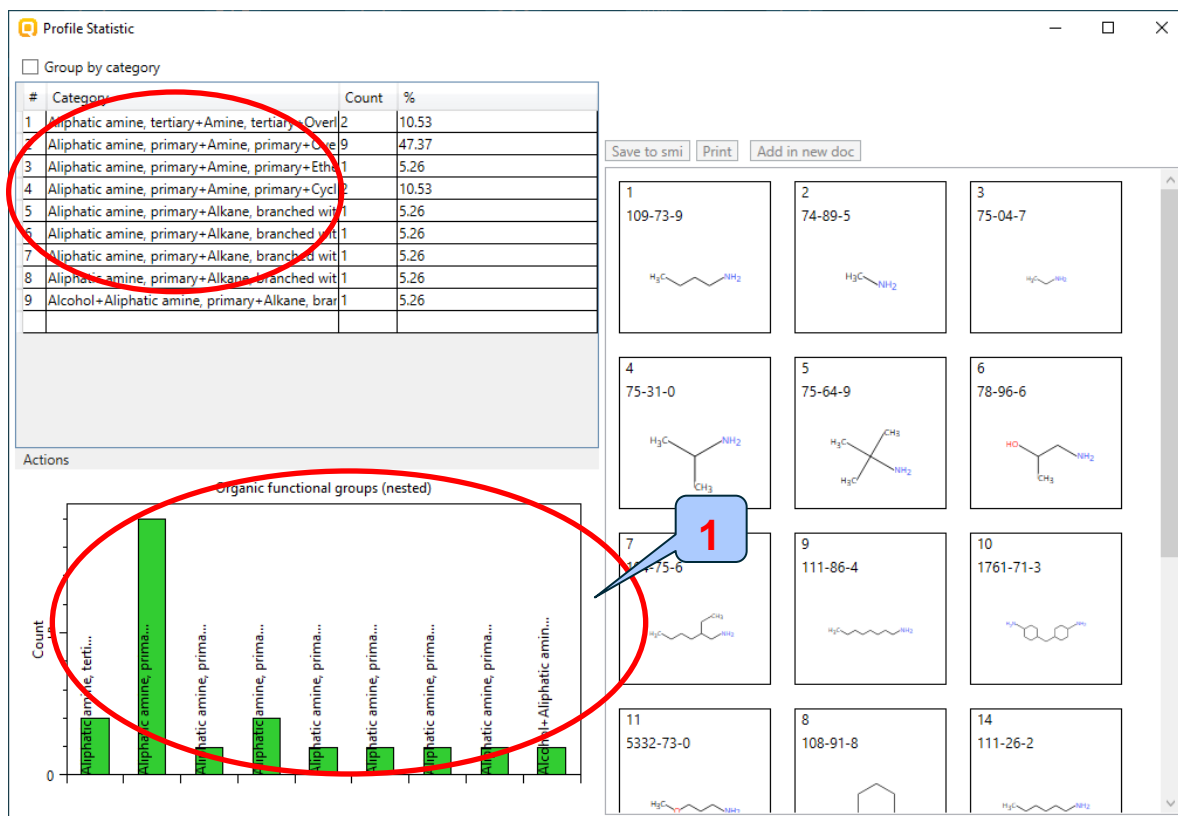


1. In this case all 19 chemicals are Narcotic amines according to MOA by OASIS profiling scheme

Phase I: Evaluation of Category & Data Gap Filling

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

Profiling/Statistics according to OFG (nested)



1. In this case all the chemicals in the category are categorized as "Aliphatic amine," according to OFG (nested) scheme

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- they all are aliphatic amines).**

Outlook

- Background
- Keywords
- 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

In step 2, the applicability domain of the category will be investigated. The domain consist of two layers:

- 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 in order to check the parametric boundaries (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

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Profiling Custom profile

Apply View New Delete

Documents

Document 1

[C: 19; Md: 0; P: 0] Aliphatic amines.smi

Filter endpoint tree...

Structure

Structure info

Parameters

2D

(Q) Acidic pKa (Chemaxon)

(Q) Basic pKa (Chemaxon)

Acidic pKa (OASIS Consensus)

Acidic pKa (OASIS Electric)

Acidic pKa (OASIS Regression)

Amino acids pKa (OASIS Regression)

BAF

BAF (lower trophic)

BAF (mid trophic)

BAF (upper trophic)

BAF (upper trophic, biotransformation...)

Basic pKa (OASIS Regression)

BCF

BCF (lower trophic)

BCF (mid trophic)

BCF (upper trophic)

BCF (upper trophic, biotransformation...)

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

Boiling point

Options

4 Selected

Select All Unselect All Invert

Suitable

Acute aquatic toxicity classification by VEGA

Acute aquatic toxicity MOA by OASIS

Aquatic toxicity classification by ECOSAR

US-EPA New Chemical Categories

Plausible

Chemical elements

Groups of elements

Hydrolysis half-life (K_a, pH 7) (Hydrowin)

Hydrolysis half-life (K_b, pH 5) (Hydrowin)

Hydrolysis half-life (K_b, pH 7) (Hydrowin)

Hydrolysis half-life (K_b, pH 8) (Hydrowin)


Hydrolysis half-life (pH 6.5-7.4)














Ionization at pH = 1

Ionization at pH = 4

Ionization at pH = 7.4

1

1. Expand "Parameters" level then open 2D by click on the box  to.

	1	2	3	4	5	6	7	8	9	10	11	12	13
Structure													
(Q) Acidic pKa (Chemaxon)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
(Q) Basic pKa (Chemaxon)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Acidic pKa (OASIS Consensus)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Acidic pKa (OASIS Electric)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Acidic pKa (OASIS Regression)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Amino acids pKa (OASIS Regression)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BAF	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BAF (lower trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BAF (mid trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BAF (upper trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BAF (upper trophic, biotransformation...)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Basic pKa (OASIS Regression)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BCF	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BCF (lower trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BCF (mid trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BCF (upper trophic)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BCF (upper trophic, biotransformation...)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Bio Half-Life	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biodeg probability (Biowin 1)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biodeg probability (Biowin 2)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biodeg probability (Biowin 5)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biodeg probability (Biowin 6)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biodeg probability (Biowin 7)	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
BioHC Half-Life	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Biotransformation Half-Life	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated
Boiling point	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated	Not calculated

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 displays the QSAR Toolbox interface. On the left, the 'Documents' panel shows 'Document 1' with the file '[C: 19;Md: 0;P: 0] Aliphatic amines.smi'. Below it, the 'Profiling methods' panel is set to 'Options' with '4 Selected' methods, including 'Suitable' and 'Plausible'. The 'Endpoint tree' on the left lists various parameters under '2D' and '3D' categories. A table in the center shows the results for these parameters across five chemical structures. A context menu is open over the table, with the option 'Calculate/extract all parameters for all chemicals' highlighted. Red callouts 1 and 2 indicate the steps: 1. Right click above the parameter in the endpoint tree; 2. The user can calculate all parameters for all chemicals or extract all 2D parameters for all.

1. Right click above the parameter in the endpoint tree; 2. The user can calculate all parameters for all chemicals or extract all 2D parameters for all.

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

Input Profiling Data Category definition Data Gap Filling Report

Profiling Custom profile

Apply View New Delete

Documents

Document 1

[C: 19; Md: 0; P: 0] Aliphatic amines.smi

Filter endpoint tree...

Structure

Structure Info

Parameters

2D

(Q) Acidic pKa (Chemaxon)

(Q) Basic pKa (Chemaxon)

Acidic pKa (OASIS Consensus)

Acidic pKa (OASIS Electric)

Acidic pKa (OASIS Regression)

Amino acids pKa (OASIS Regression)

BAF

BAF (lower trophic)

BAF (mid trophic)

BAF (upper trophic)

BAF (upper trophic, biotransformation...)

Basic pKa (OASIS Regression)

BCF

BCF (lower trophic)

BCF (mid trophic)

BCF (upper trophic)

BCF (upper trophic, biotransformation...)

Bio Half-Life

Biodegradation probability (Biowin 1)

Biodegradation probability (Biowin 2)

Biodegradation probability (Biowin 5)

Biodegradation probability (Biowin 6)

Biodegradation probability (Biowin 7)

BioHC Half-Life

Biotransformation Half-Life

Boiling point

Exp Boiling Point

Exp Henrys Law Constant

Exp Log P

Exp Melting Point

Exp NO3 rate constant

Options

Select All Unselect All Invert

4 Selected

Suitable

Acute aquatic toxicity classification by Veri

Acute aquatic toxicity MOA by OASIS

Acute toxicity classification by ECOSAR

US-EPA New Chemical Categories

Plausible

Chemical elements

Groups of elements

Hydrolysis half-life (Ka, pH 7)(Hydrowin)

Hydrolysis half-life (Kb, 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

Lipinski Rule Oas

OECD HPV Chemical Categories

Metabolism/Transformations

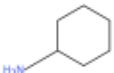



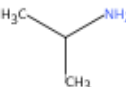
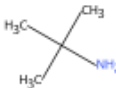
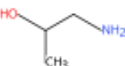

1	2	3	4	5	6	7	8	9	10	11	12	13
No value	No value	No value	No value	No value	No value	15.3	No value	No value	No value	No value	No value	No value
10.5	10.2	10.1	10.2	10.4	10.7	9.6	10.2	10.2	10.8	10.1	10.4	10.2
No value	No value	No value	No value	No value	No value	11.9	No value	No value	No value	No value	No value	No value
8.21	8.16	6.86	8.09	8.13	8.01	7.99	8.23	8.12	8.13	8.05	8.26	8.12
No value	No value	No value	No value	No value	No value	13	No value	No value	No value	No value	No value	No value
No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value
0.59 log(L/kg)	0.25 log(L/kg)	-0.04 log(L/kg)	-0.02 log(L/kg)	0.02 log(L/kg)	0.05 log(L/kg)	-0.05 log(L/kg)	1.75 log(L/kg)	1.82 log(L/kg)	2.24 log(L/kg)	-0.04 log(L/kg)	0.15 log(L/kg)	1.53 log(L/kg)
0.436 log(L/kg)	0.168 log(L/kg)	-0.02 log(L/kg)	-0.008 log(L/kg)	0.018 log(L/kg)	0.035 log(L/kg)	-0.025 log(L/kg)	1.57 log(L/kg)	1.64 log(L/kg)	2.03 log(L/kg)	-0.018 log(L/kg)	0.097 log(L/kg)	1.34 log(L/kg)
0.472 log(L/kg)	0.186 log(L/kg)	-0.023 log(L/kg)	-0.01 log(L/kg)	0.019 log(L/kg)	0.038 log(L/kg)	-0.028 log(L/kg)	1.62 log(L/kg)	1.69 log(L/kg)	2.08 log(L/kg)	-0.021 log(L/kg)	0.108 log(L/kg)	1.39 log(L/kg)
0.594 log(L/kg)	0.25 log(L/kg)	-0.039 log(L/kg)	-0.02 log(L/kg)	0.023 log(L/kg)	0.052 log(L/kg)	-0.046 log(L/kg)	1.75 log(L/kg)	1.82 log(L/kg)	2.24 log(L/kg)	-0.035 log(L/kg)	0.147 log(L/kg)	1.53 log(L/kg)
0.629 log(L/kg)	0.28 log(L/kg)	-0.035 log(L/kg)	-0.012 log(L/kg)	0.037 log(L/kg)	0.066 log(L/kg)	-0.043 log(L/kg)	1.89 log(L/kg)	1.98 log(L/kg)	2.39 log(L/kg)	-0.03 log(L/kg)	0.173 log(L/kg)	1.63 log(L/kg)
10.3	10	9.74	10.1	10.4	10.8	9.04	9.96	10	10.1	8.99	10.4	10
0.65 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	1.53 log(L/kg)	1.58 log(L/kg)	1.82 log(L/kg)	0.5 log(L/kg)	0.5 log(L/kg)	1.36 log(L/kg)
0.436 log(L/kg)	0.168 log(L/kg)	-0.02 log(L/kg)	-0.008 log(L/kg)	0.018 log(L/kg)	0.035 log(L/kg)	-0.025 log(L/kg)	1.57 log(L/kg)	1.64 log(L/kg)	2.03 log(L/kg)	-0.018 log(L/kg)	0.097 log(L/kg)	1.34 log(L/kg)
0.472 log(L/kg)	0.186 log(L/kg)	-0.023 log(L/kg)	-0.01 log(L/kg)	0.019 log(L/kg)	0.038 log(L/kg)	-0.028 log(L/kg)	1.62 log(L/kg)	1.69 log(L/kg)	2.08 log(L/kg)	-0.021 log(L/kg)	0.108 log(L/kg)	1.39 log(L/kg)
0.594 log(L/kg)	0.25 log(L/kg)	-0.039 log(L/kg)	-0.02 log(L/kg)	0.023 log(L/kg)	0.052 log(L/kg)	-0.046 log(L/kg)	1.75 log(L/kg)	1.82 log(L/kg)	2.24 log(L/kg)	-0.035 log(L/kg)	0.147 log(L/kg)	1.53 log(L/kg)
0.623 log(L/kg)	0.277 log(L/kg)	-0.035 log(L/kg)	-0.012 log(L/kg)	0.036 log(L/kg)	0.065 log(L/kg)	-0.043 log(L/kg)	1.85 log(L/kg)	1.93 log(L/kg)	2.29 log(L/kg)	-0.03 log(L/kg)	0.177 log(L/kg)	1.61 log(L/kg)
0.369 d	0.215 d	0.0724 d	0.0963 d	0.125 d	0.175 d	0.025 d	0.723 d	0.755 d	3.85 d	0.064 d	0.171 d	0.614 d
0.854	0.975	0.887	0.88	0.873	0.683	1.02	0.948	0.948	0.955	0.512	0.867	0.955
0.938	0.993	0.975	0.97	0.964	0.795	0.985	0.984	0.984	0.904	0.36	0.956	0.987
0.46	0.539	0.529	0.532	0.46	0.422	0.556	0.477	0.553	0.359	0.503	0.463	0.549
0.528	0.705	0.71	0.708	0.586	0.319	0.687	0.576	0.698	0.169	0.597	0.584	0.7
0.453	0.694	0.934	0.96	0.688	0.44	0.927	0.501	0.798	0.414	0.755	0.714	0.772
No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value
0.369 d	0.215 d	0.0724 d	0.0962 d	0.125 d	0.175 d	0.025 d	0.723 d	0.755 d	3.85 d	0.064 d	0.171 d	0.614 d
145 °C	87.7 °C	10.2 °C	36.9 °C	47.3 °C	62.8 °C	131 °C	166 °C	178 °C	316 °C	113 °C	72.8 °C	157 °C
134 °C	78 °C	-6.3 °C	16.6 °C	31.8 °C	45 °C	160 °C	169 °C	180 °C	320 °C	118 °C	62.5 °C	156 °C
4.16E-06 atm-m...	1.74E-05 atm-m...	1.11E-05 atm-m...	1.23E-05 atm-m...	4.51E-05 atm-m...	3.58E-05 atm-m...	No value	9.52E-05 atm-m...	0.000824 atm-m...	No value	No value	0.000153 atm-m...	No value
1.49	0.97	-0.57	-0.13	0.26	0.4	-0.96	2.82	2.9	No value	No value	0.74	2.57
-17.7 °C	-50 °C	-93.4 °C	-81.2 °C	-95.1 °C	-72.7 °C	25 °C	-76 °C	0 °C	15 °C	No value	No value	-18 °C
No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value

All calculated 2D parameters appears on data matrix.

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

Filter endpoint tree...	1	2	3	4	5	6	7	8
Structure								
Parameters								
2D								
log Kow	1.63	0.83	-0.64	-0.15	0.27	0.72	-1.19	2.73
Molecular Weight	99.2 Da	73.1 Da	31.1 Da	45.1 Da	59.1 Da	73.1 Da	75.1 Da	129 Da
Water Solubility	6.4E+04 mg/L	2.03E+05 mg/L	1E+06 mg/L	1E+06 mg/L	8.38E+05 mg/L	6.22E+05 mg/L	1E+06 mg/L	3.68E+03 mg/L

The boundaries of the three phys-chem parameters fall in the following ranges:

1. log Kow (from 0.64 to 7.71)
2. Molecular weight (from 31 to 269 Da)
3. Water solubility (from 0.48×10^{-1} to 1×10^6 mg/l)

Outlook

- Background
- Keywords
- 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 displays the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Input', 'Profiling', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data' module is selected, and the 'Gather' button is highlighted with a red box and a blue callout labeled '3'. The left sidebar shows a tree view of chemical properties, with 'Aquatic Toxicity' and 'EC50' highlighted in green. A blue callout labeled '2' points to the 'Data' section in the sidebar. The main window shows a grid of chemical structures and their corresponding data. A 'Read data?' dialog box is open, with 'All endpoints' selected and 'OK' highlighted with a blue callout labeled '4'.

1. Go to **Data** module; 2. Select the green highlighted databases related to aquatic toxicity EC 50 Daphnia magna; 3. Click **Gather**; 4. Click **OK**.

Phase I: Evaluation of Category & Data Gap Filling

Step 3. Reading data

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

The screenshot displays the QSAR TOOLBOX software interface. On the left, there is a sidebar with 'Documents' and 'Databases' sections. The 'Documents' section shows a file named '[C: 19;Md: 379;P: 0] Aliphatic amines.sm'. The 'Databases' section shows a list of selected databases including 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Aquatic Toxicity', 'Terrestrial Toxicity', and 'Human Health Hazards'. The main window displays a table of experimental data points. A dialog box is overlaid on the table, stating '379 points added across 16 chemicals.' with an 'OK' button. A red circle with the number '1' is placed over the 'OK' button. At the bottom of the screen, a blue box contains the text '1. Click OK'.

379 points added across 16 chemicals.

OK

1

1. Click OK

Outlook

- Background
- Keywords
- 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

- As mentioned before (slide 28) the category is structurally and mechanistically similar
- Hence the data gap could be filled in for analogues from the category. In this case we will fill the gap for 2-Butanamine (this will be our target chemical)
- Before proceeding with filling data gap you should navigate to the column corresponding to the target chemical (see next slide).

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

Navigate to the target chemical

The screenshot displays the QSAR Toolbox software interface. The top toolbar shows the 'Data Gap Filling' module selected. The left sidebar contains a 'Filter endpoint tree...' with various categories like 'Structure', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', and 'Ecotoxicological Information'. The central data table lists chemicals and their corresponding data points. A 'Possible data inconsistency' dialog box is open, showing metadata and scale/unit options. Numbered callouts (1-5) indicate the steps: 1. Click 'Data Gap Filling' in the top toolbar. 2. Find '2-Butanamine' in the data table. 3. Click on the cell for the target endpoint (yellow highlighted row). 4. Click 'Trend analysis' in the left sidebar. 5. Click 'OK' in the 'Possible data inconsistency' dialog box.

1. Go to **Data Gap Filling** module; 2. Found the target chemical (2-Butanamine); 3. Click on the cell corresponding to target endpoint (yellow highlighted row); 4. Click **Trend analysis**; Possible data inconsistency window appears, click **OK** (5)

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine

[illegible]

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 Subcategorization step. The main window shows a list of chemicals and their associated data, including EC50 values and molecular weights. A red box labeled '1' highlights the 'Select / filter data' button in the bottom right panel. A red box labeled '2' highlights the 'Organic functional groups (nested)' profiler in the 'Options' section. A red box labeled '3' highlights the 'Remove selected' button in the 'Simulated' section. The bottom panel shows a scatter plot of EC50 (log10(mg/L)) versus log10(mol/L) with a regression line. Chemical structures are shown for several points on the plot, including 2-butanamine and cyclohexylamine.

There are many profilers to be used for subcategorization. The most suitable profilers (the green ones) does not found any dissimilar chemical. In this respect the structure-based OFG (nested) profiler is used for refining the category: 1. Open **Select/filter data, Subcategorization**; 2. Click on **OFG (nested)** then 3. Click **Remove selected** button

Phase I: Evaluation of Category & Data Gap Filling

Step 4. Data Gap Filling for 2-Butanamine/Predicted result

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Trend analysis Read across (Q)SAR Standardized Automated

Documents

- Document 1
 - [C: 19;Md: 379;P: 0] Aliphatic amines.smi
 - [C: 12;Md: 281;P: 0] Enter GF(TA)
 - [C: 8;Md: 198;P: 0] Subcategorized: Organic functional groups (nested)

Filter endpoint tree...

Structure

- Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchiopods)
 - Daphnia magna
 - EC0 1/2
 - EC100 1/2
 - EC50 7/8
- Mortality 8/68
- Physiology 3/23
- Population 4/28
- Sediment Toxicity 5/12
- Terrestrial Toxicity
- Human Health Hazards
- Profiling
 - Predefined
 - US-EPA New Chemical Categories
 - Endpoint Specific

12 [target]	1	3	9	13	14	16	19
<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>
M: 95 mg/L	M: 260 mg/L	M: 163 (147+180)	M: 1.9 (1.5+2.4)	M: 9.4 (6.9+12)	M: 8.6 (7+11)	M: 56 (27+170)	M: 0.58 (0.39+0.8)
M: 275 (250+301)	M: 24 mg/L	M: 56 mg/L	M: 5.15 mg/L	M: 21.8 (19.7+24)	M: 56.6 (53.8+59)	M: 16.6 mg/L	M: 1.04 mg/L
M: >1E+03 mg/L	M: 0.129 mg/L	M: 0.14 mg/L	M: 0.07 (0+1.28)			M: 1E-09 ug/cell	M: 1.51 mg/L
M: 578 mg/L	M: 15 mmol	M: 2 %				M: 506 mg/L	M: 1.08 mg/cm2
Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines	Aliphatic Amines

Trend analysis prediction for EC50, based on 7 values
Predicted: 79 mg/L
Model equation: $EC50 = 2.34 (\pm 0.337) + 0.822 (\pm 0.154) * \log Kow, \log(1/mol/L)$

EC50 [log(1/mol/L)]

log Kow

Accept prediction

1. Predicted result is 79.0 mg/l

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

trend analysis Read across (Q)SAR Standardized Automated

Documents

- Document 1
 - [C; 19; Md: 378; P: 0] Aliphatic amines.cmi
 - [C; 12; Md: 281; P: 0] Enter GR(TA)
 - [C; 8; Md: 198; P: 0] Subcategorized: Organic functional

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

QSARs	0
Automated workflows	0
Standardized workflows	0

Filter endpoint tree...

- Structure
 - Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchio...
 - Daphnia magna
 - EC0 1/2 M: 95 mg/L
 - EC100 1/2 M: 260 mg/L
 - EC50 7/8 M: >100 mg/L M: 163 (147+180)... M: 1.9 (1.5+2.4)... M: 9.4 (6.9+12)... M: 8.6 (7+11) m... M: 56 (27+170)... M: 0.58 (0.39+0...
 - Mortality 8/68 M: 275 (250+301)... M: 24 mg/L M: 56 mg/L M: 5.15 mg/L M: 21.8 (19.7+24... M: 56.6 (53.8+59... M: 16.6 mg/L M: 1.04 mg/L
 - Physiology 3/23 M: >1E+03 mg/L M: 0.129 mg/L M: 0.07 (0+1.28)...
 - Population 4/28 M: 0.14 mg/L M: 31.1 mg/L
 - Sediment Toxicity
 - Terrestrial Toxicity 5/12 M: 578 mg/L M: 15 mmol M: 2 % M: 506 mg/L M: 1.08 mg/cm2
 - Human Health Hazards
 - Profiling
 - Predefined
 - US-EPA New Chemical Categories
 - Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines Aliphatic Amines

Descriptors Prediction Adequacy Cumulative frequency Residuals Statistics

Cumulative frequency (%)

Residuals, Y - Y_{calc}

95% of Residuals ≤ 0.243, log(1/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

Phase I: Evaluation of Category & Data Gap Filling

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

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis: Read across (QSAR) Standardized Automated

Documents

- Document 1
 - [C: 18;Md: 379;P: 0] Aliphatic amines.smi
 - [C: 12;Md: 281;P: 0] Enter GF(TA)
 - [C: 8;Md: 198;P: 0] Subcategorized: Organic functional

Filter endpoint tree...

Structure

Animalia (animals)

- Arthropoda (arthropods)
- Branchiopoda (branchiopods)
- Daphnia magna
 - EC0
 - EC100
 - EC50
- Mortality
- Physiology
- Population
- Sediment Toxicity
- Terrestrial Toxicity
- Human Health Hazards
- Profiling

	12 [target]	1	3	9	13	14	16	19
Structure	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>	<chem>CCCCN</chem>
EC0	1/2		M: 95 mg/L					
EC100	1/2		M: 260 mg/L					
EC50	7/8	M: >100 mg/L	M: 163 (147+180)	M: 1.9 (1.5+2.4)	M: 9.4 (6.9+12)	M: 8.6 (7+11)	M: 56 (27+170)	M: 0.58 (0.39+0.8)
Mortality	8/68	M: 275 (250+301)	M: 24 mg/L	M: 56 mg/L	M: 5.15 mg/L	M: 21.8 (19.7+24)	M: 56.6 (53.8+59)	M: 16.6 mg/L
Physiology	3/23	M: >1E+03 mg/L	M: 0.129 mg/L				M: 1E-09 ug/cell	M: 1.04 mg/L
Population	4/28	M: 0.14 mg/L	M: 31.1 mg/L	M: 0.07 (0+1.28)				M: 1.51 mg/L
Sediment Toxicity	5/12						M: 506 mg/L	M: 1.08 mg/cm2
Terrestrial Toxicity		M: 578 mg/L	M: 15 mmol	M: 2 %				
Human Health Hazards								
Profiling								

Descriptors

Statistical characteristics

Number of data points, (N)

7

Coefficient of determination, (R2)

0.974

Adjusted coefficient of determination, (R2adj)

0.969

Coefficient of determination - leave one out, (Q2)

N/A

Sum of squared residuals, (SSR)

0.220

Standard deviation of residuals, (sN)

0.177

Sample standard deviation of residuals, (s)

0.210

Fisher function, (F)

189

Fisher threshold for statistical significance, (Fa)

10.0 (95.0%)

b0

- model descriptor

Intercept

- coeff. value

2.34

- coeff. range

±0.337

- significance

No

- max covariation

0.377 vs log Kow

b1

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

Clear existing marks

Gao filling approach

Accept prediction

1. Coefficient of determination is higher - 0.974

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

Phase II: Extending the category & data gap filling

- The extension of the category is performed by using Phase II of the category evaluation process (*Extending the 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: Extending the category & data gap filling

Step 1: Category definition

The screenshot shows the QSAR Toolbox 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, the 'Filter endpoint tree...' panel shows a hierarchical tree of categories. The 'Aliphatic amines' category is selected, highlighted in blue, and labeled with a red '2'. The main table displays chemical structures and their corresponding data points across multiple endpoints. The table has 13 columns, each representing a different chemical structure. The rows represent different endpoints, such as 'Mortality', 'Physiology', 'Population', 'Sediment Toxicity', 'Terrestrial Toxicity', and 'Human Health Hazards'. The data points are numerical values representing the results of the QSAR analysis.

1. Go back to **Input** section; 2. Select the row with **"Aliphatic amines"** from the documented tree

Phase II: Extending the category & data gap filling

Step 1: Category definition

1. Right click above the **target chemical (2-Butanamine)**;

2. Select **Set as new target**; 3. Go to **Data Gap Filling**

Phase II: Extending the category & data gap filling

Step 1: Category definition

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. The main workspace is divided into several panels:

- Documents Panel (Left):** Shows a list of documents. A red dashed box highlights a document named "[C: 1; Md: 8; P: 0] Select from: Aliphatic amines.smi". A blue callout bubble with the number "2" points to this document.
- Data Gap Filling Settings Panel (Bottom Left):** Contains checkboxes for "Only endpoint relevant" and "At this position:". Below these are sections for "In nodes below:" with checkboxes for QSARs, Automated workflows, and Standardized workflows, each with a corresponding "0" value.
- Filter endpoint tree... Panel (Center):** Displays a hierarchical tree of endpoints. The "Aquatic Toxicity" section is expanded, showing sub-endpoints like "Intoxication", "24 h", "48 h", "Animalia (animals)", "Arthropoda (arthropods)", "Branchiopoda (branchiopods)", and "Daphnia magna". The "Daphnia magna" node is highlighted in yellow. A blue callout bubble with the number "1" points to the chemical structure of 2-Butanamine shown above the tree.
- Chemical Structure Panel (Top Right):** Displays the chemical structure of 2-Butanamine, with the formula CCCCN.

1. The target chemical 2-Butanamine is loaded in new data matrix; 2. Moreover a new level appears in the documented tree. The latter allows easy go back and forward to different stages of the workflow.

Phase II: Extending the category & data gap filling

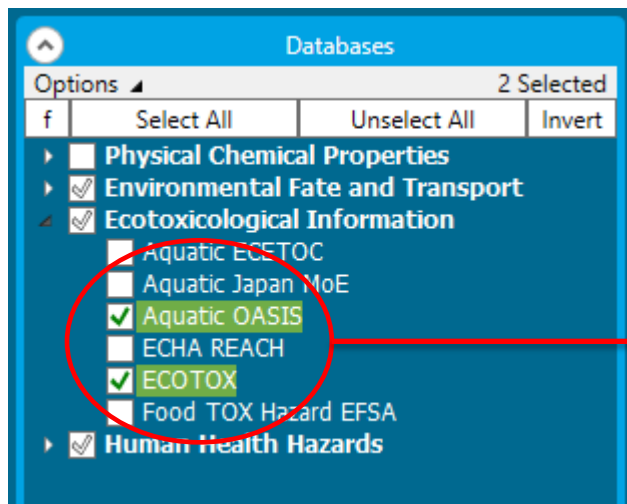
Step 1: Category definition

- The ECOSAR strict category is used to define a broader category used in further analysis. The “Strict” option means that only the defined categories should be present in the found analogues and not any others.
- The same endpoint: EC 50 48h *D.magna* will be predicted as with the predefined category

Phase II: Extending the category & data gap filling

Step 1: Category definition

- Define ECOSAR category
- Before defining the category, the following databases related to the target endpoint are selected (databases highlighted in green are those which have data for the target endpoint):



- Aquatic OASIS
- ECOTOX

Phase II: Extending the category & data gap filling

Step 1: Category definition/Defining ECOSAR (strict)

1. Go to **Category definition** module; 2. Highlight **"Aquatic toxicity classification by ECOSAR"**; 3. Click **Define**; 4. Select **Strict**. 5. Click **OK**

Phase II: Extending the category & data gap filling

Step 1: Category definition/Defining ECOSAR (strict)

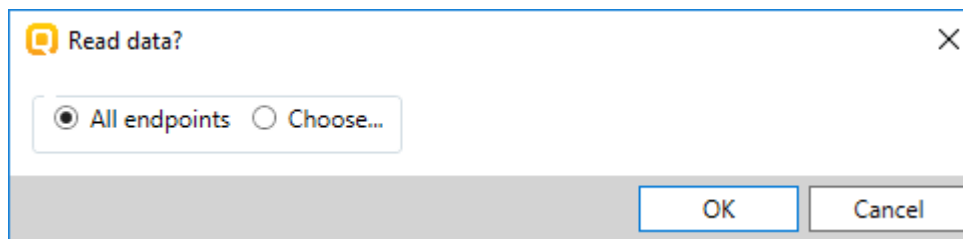
The screenshot shows the QSAR Toolbox software interface. The 'Category definition' tab is active, displaying a 'Filter endpoint tree' on the left and a list of chemical structures on the right. A 'Grouping results' dialog box is open in the center, showing '370 chemical(s) found.' and an 'OK' button. A red callout bubble with the number '1' points to the 'OK' button. The background shows the 'Category definition' tab with a 'Filter endpoint tree' and a list of chemical structures.

1. A message appears informing that 370 analogs are found. Click **OK**

Phase II: Extending the 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 requests 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).

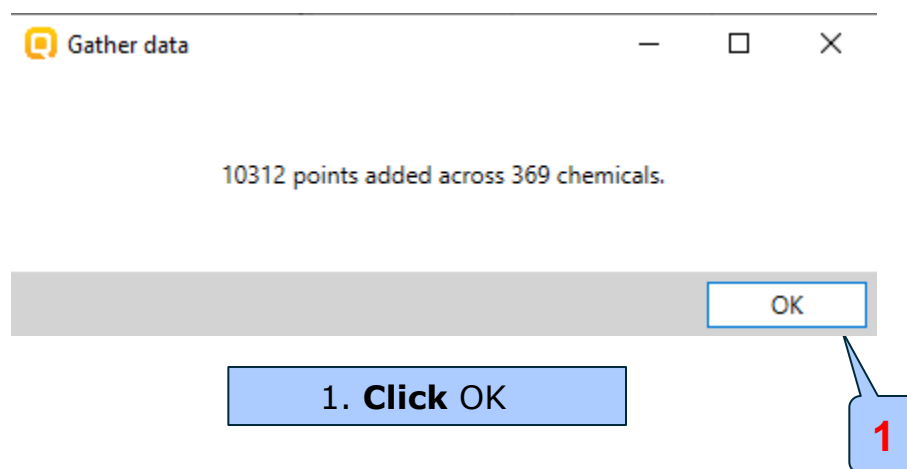


- In this example, as only databases are selected that contain information for aquatic toxicity endpoint, both options give the same results.

Phase II: Extending the category & data gap filling

Step 1: Category definition/Reading data

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



Outlook

- Background
- Keywords
- Category evaluation – overview
 - Phase I. Evaluation of Category & Data Gap Filling
 - **Phase II. Extending the Category& Data gap Filling**
 - Step 1: Category definition
 - Step 2: Navigate to the target endpoint – this is already done, so it is skipped
 - **Step 3: Data Gap Filling**

Phase II: Extending the category & data gap filling

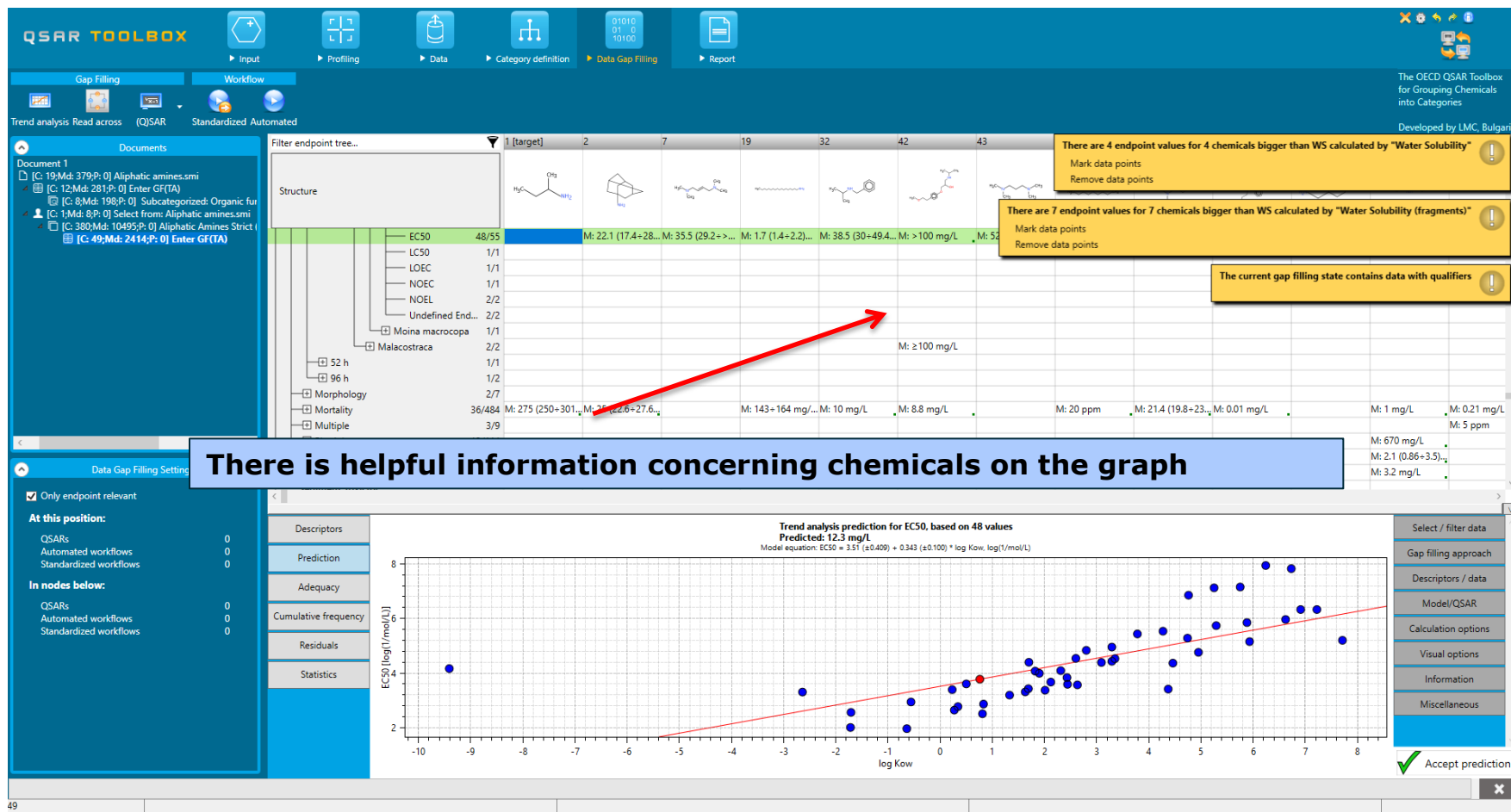
Step 3: Data Gap Filling

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' menu is highlighted with a red box and a callout '1'. Below it, the 'Trend analysis' option is also highlighted with a red box and a callout '1'. The main workspace shows a 'Filter endpoint tree...' on the left, a central table of chemical data, and a 'Data Gap Filling Settings' panel on the bottom left. A 'Possible data inconsistency' dialog box is open on the right, showing metadata for 'Class', 'Duration', 'Effect', 'Endpoint', and 'Kingdom'. It lists 'Native scale/unit' options (mg/L, mol/L, ppb, ppm, µg/L) and 'Converted data' options (g/m³, log(1/mol/L), mg/L, mol/L, mol/m³, ppb). The dialog box has 'OK' and 'Cancel' buttons, with a callout '2' pointing to the 'OK' button. The background table shows chemical structures and their corresponding data points, with some cells highlighted in yellow.

1. Go to **Data Gap filling** and select **Trend analysis**; 2. The user will be informed If there is different experimental data. Click **OK**.

Phase II: Extending the category & data gap filling

Step 3: Data Gap Filling



Phase II: Extending the category & data gap filling

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

The screenshot displays the OECD QSAR Toolbox software interface during the Subcategorization process. The interface is divided into several sections:

- Left Sidebar:** Contains 'Options' and 'Metabolisms' tabs. Under 'Options', the 'Empiric' section is expanded, showing a list of categories. 'Organic functional groups (nested)' is highlighted with a red box and a blue callout '2'. Under 'Metabolisms', the 'Simulated' section is expanded, showing a list of simulation types. 'Alkyl-, alkenyl- and alkylnyl (hetero)an' is highlighted with a red box and a blue callout '3'.
- Subcategorization Window:** A central window titled 'Subcategorization' with 'Options' and 'Adjust options' tabs. It shows a 'Target' list and a 'Differ from target by' section. A 'STOP' button is visible.
- Main Data Table:** A table with columns for chemical structures and predicted values. The table is divided into two main sections: 'Target' and 'Analogues'. The 'Target' section shows chemical structures and their predicted values. The 'Analogues' section shows chemical structures and their predicted values. A 'Trend analysis prediction for EC50, based on 48 values' is displayed below the table, with a predicted value of 12.3 mg/L. A red line indicates the trend.
- Right Panel:** A panel titled 'Select / filter data' with buttons for 'Subcategorize', 'Mark chemicals by descriptor value', 'Mark outliers', 'Filter points by test conditions', 'Mark focused chemical', 'Mark focused points', and 'Remove marked data'. A green checkmark and 'Accept prediction' button are at the bottom.

In order to refine the obtained category with analogues few subcategorizations are applied: 1. Open **Select/Filter data**, then click **Subcategorize**; 2. Select **Organic functional groups(nested)**; 3. Chemicals which are structurally dissimilar (highlighted in light blue on the chart) are removed.

Phase II: Extending the category & data gap filling

Step 3: Data Gap Filling/Subcategorize by Lipinski rules

The screenshot displays the OECD QSAR Toolbox interface. On the left, the 'Subcategorization' panel is open, showing the 'Empiric' section with 'Lipinski Rule Based' selected. A red callout '1' points to this selection. Below, the 'Custom' section has 'Metabolisms' selected, with a red callout '2' pointing to the 'Do not account metabolism' option. At the bottom of the 'Subcategorization' panel, a red callout '4' points to the 'Remove selected' button. In the center, a 'Trend analysis prediction for EC50' plot is shown, with a red circle highlighting a cluster of points at high log Kow values. A red arrow points from this cluster to a window titled '6 structures from: Less bioavailable', which is highlighted by a red callout '3'. The plot shows a positive correlation between log Kow and EC50, with a model equation: $EC_{50} = 2.78 (\pm 0.867) + 0.640 (\pm 0.193) * \log Kow, \log(1/mol/L)$. The predicted EC50 is 39.4 mg/L.

Chemicals with very long chain could be removed from the category due to their less-bioavailability. In this respect: 1. Select **Lipinski rules**; 2. Double click to see "Less-bioavailable" chemicals; 3. Close the appeared window; 4. **Remove selected** 6 chemicals

Phase II: Extending the category & data gap filling

Step 3: Data Gap Filling/Prediction result

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

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

Document 1

- [C: 19;Md: 379;P: 0] Aliphatic amines.smi
- [C: 12;Md: 281;P: 0] Enter GF(TA)
- [C: 8;Md: 198;P: 0] Subcategorized: Organic fu...
- [C: 1;Md: 8;P: 0] Select from: Aliphatic amines.smi
- [C: 380;Md: 10495;P: 0] Aliphatic Amines Strict
- [C: 49;Md: 2414;P: 0] Enter GF(TA)
- [C: 18;Md: 282;P: 0] Subcategorized: Or...
- [C: 12;Md: 262;P: 0] Subcategoriz...

Filter endpoint tree...

Structure

- Animalia (animals)
 - Arthropoda (arthropods)
 - Branchiopoda (branchiopo...
 - Daphnia magna
 - EC0
 - 1/2
 - 1/2
 - EC50
 - 11/12
- Mortality
 - 2/6
 - 4/28
 - Physiology
 - Population
 - 5/29
- Sediment Toxicity
 - Terrestrial Toxicity
 - Human Health Hazards
 - 7/29

Descriptors

Prediction

Adequacy

Cumulative frequency

Residuals

Statistics

Data Gap Filling Settings

Only endpoint relevant

At this position:

QSARs 0

Automated workflows 0

Standardized workflows 0

In nodes below:

QSARs 0

Automated workflows 0

Standardized workflows 0

1 [target] 84 127 175 177 328 329 337 364 365 369 373

M: 0.5 (0.4+0.61)... M: 0.026 (0.018+... M: 1.6 (1.2+2) m... M: 1.9 (1.5+2.4)... M: 56 (27+170)... M: 0.58 (0.39+0... M: 94 (6.9+12)... M: >100 mg/L M: 8.6 (7+11) m... M: >100 mg/L M: 163 (147+180...

M: 275 (250+301... M: 0.21 mg/L M: 0.0995 mg/L M: 2.16 (2.02+2... M: 5.15 mg/L M: 16.6 mg/L M: 1.04 mg/L M: 21.8 (19.7+24... M: 24 mg/L M: 56.6 mg/L M: 0.121 mg/L M: 56 mg/L

M: 5 ppm M: 5 ppm

M: 1.28 mg/L M: 0.07 (0+1.28)... M: 1.51 mg/L M: >1E+03 mg/L M: 0.14 mg/L M: 4E-09 ug/cell M: 0.129 mg/L M: 31.1 mg/L

M: 0.1 mL M: 2 % M: 506 mg/L M: 1.08 mg/cm2 M: 578 mg/L M: 8.7 mmol M: 15 mmol

Trend analysis prediction for EC50, based on 11 values
Predicted: 77.8 mg/L
Model equation: $EC50 = 2.34 (\pm 0.324) + 0.838 (\pm 0.116) * \log Kow, \log(1/mol/L)$

1

2

Accept prediction

1. Prediction result is 77.8 mg/L

2. Click on **Accept prediction**.

Outlook

- Background
- Keywords
- 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

The screenshot illustrates the steps for saving a prediction result in the QSAR Toolbox. The interface shows the 'Input' section selected in the top menu. The 'Documents' panel on the left lists several documents, with 'Document 1' selected. The 'tree...' panel in the center shows a hierarchical structure of chemical categories. The 'target' input field on the right is set to 'M: 0.00'. The 'Save document' dialog box is open, showing the file name 'Tutorial 4' and the save type 'Toolbox documents (*.tb4)'. The 'Save' button is highlighted. A 'Toolbox' dialog box is also open, asking 'Do you want to save changes to document 1?' with 'Yes' and 'No' buttons.

1. Go to **Input** section 2. Click on **Save** button 3. Click **Yes**; 4. Define name of the file and **Save** the workflow.

Open saved file

The screenshot shows the QSAR Toolbox software interface. The 'Open' dialog box is open, displaying a file explorer view. The file 'Tutorial 4.tb4' is selected in the 'Toolbox documents (*.tb4)' folder. The 'Open' button is highlighted. The background interface shows the 'Documents' pane on the left, the 'Structure' pane in the center, and a 'Table' pane on the right. The 'Documents' pane lists various documents, including 'ment 1', '19-Md: 379-P: 1', 'Aliphatic an', 'C: 12-Md: 281-P: 1', 'Enter GF', 'Md: 198-P: 1', 'Subc', '8-P: 1', 'Select from:', '80-Md: 10495-P: 1', '49-Md: 2414-P: 1', 'C: 18-Md: 282-P: 1', 'C: 12-Md: 26'. The 'Structure' pane shows a chemical structure of a molecule. The 'Table' pane shows a list of chemical properties and their values.

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

All the steps done during the workflow are getting executed consequentially. Once ready a message appears that the file is opened successfully, otherwise a warning message will inform you for the mistakes.