

Application manual

of OECD QSAR Toolbox v.4

(F1 help)



Table of Contents

QSAR Toolbox End - user Manual	5
Chapter A. Introduction	5
A.1. Objective	6
A.2. How to use this guidance document	7
Chapter B. Welcome	8
B.1. Acknowledgements	8
B.2. What is OECD QSAR Toolbox	9
B.3. Abbreviations	10
Chapter C. User interface	12
C.1. Stages toolbar	12
C.2. Actions toolbar	14
C.3. Stage options panel	16
C.4. Data matrix	22
C.4.1. Endpoint tree area	24
C.4.2. Area with selected chemicals	45
C.4.3. Area with data	47
C.5. Information and active accessories	51
Chapter D. Workflow	53
D.1. Input	54
D.1.1. Single chemical	55
D.1.1.1. Entering a chemical by CAS	55
D.1.1.2. Entering a chemical by chemical name	57
D.1.1.3. Entering a chemical by structure	59
D.1.1.3.1. Add single chemical structure (discrete chemical)	60
D.1.1.3.2. Add single chemical structure (mixture)	63
D.1.1.4. Entering a chemical from the last used chemicals	69
D.1.1.5. Entering a chemical from database	71
D.1.1.6. Entering a chemical from inventory list	75
D.1.1.7. Entering a chemical from file	79
D.1.2. List of chemicals	82
D.1.2.1. Loading a database	82
D.1.2.2. Loading an inventory list	85
D.1.2.3. Loading a custom file	87
D.1.3. Search options	89
D.1.3.1. Substructure (SMARTS)- 2DEditor	90
D.1.3.2. Query tool	95
D.1.4. Define Target Endpoint	116
D.1.4.1. Gene mutation in vitro	124
D.1.4.2. Gene mutation in vivo	127
D.1.4.3. Chromosome aberration in vitro	128
D.1.4.4. Chromosome aberration in vivo	130
D.1.4.5. Other endpoints	132
D.2. Profiling	133
D.2.1. Profiling methods	134
D.2.1.1. Predefined	134
D.2.1.2. General mechanistic	136
D.2.1.3. Endpoint specific	144
D.2.1.4. Empiric	153
D.2.1.5. Toxicological	158
D.2.1.6. Custom	159
D.2.2. Metabolism	160

D.2.2.1. Documented metabolism	160
D.2.2.2. Simulated metabolism	163
D.2.3. Supporting functionalities	168
D.2.4. Explaining profiling methods	172
D.2.5. Explaining profiling results	177
D.2.6. Creating a custom profile	184
D.2.7. Deleting a profile	188
D.3. Data	190
D.3.1. Databases	190
D.3.2. Inventories	216
D.3.3. Collecting data	218
D.3.4. Importing data	221
D.3.4.1. Vertical Import: Database	223
D.3.4.2. Horizontal Import: Database	234
D.3.4.3. Importing Inventory	239
D.3.4.4. Supporting information	242
D.3.5. Exporting data: UCLID 6	244
D.4. Category definition	254
D.4.1. Grouping methods	254
D.4.2. Building category - principles	260
D.4.3. Defining category	261
D.4.3.1. Categorization of single chemical	268
D.4.3.2. Categorization of mixtures	269
D.4.3.3. Categorization using profiling result of hierarchical type profiling scheme	271
D.4.3.4. Categorization with metabolism	276
D.4.3.5. Alert performance	280
D.4.4. Subcategorization	284
D.4.5. Combine categories	288
D.5. Data Gap Filling	290
D.5.1. Data Gap Filling methods	290
D.5.2. Data Gap Filling procedure	292
D.5.3. Data Gap Filling window	295
D.5.4. Common features in three gap filling approaches	298
D.5.4.1. Left-side panel	299
D.5.4.2. Menus of data gap filling	304
D.5.4.2.1. Select /filter data	305
D.5.4.2.1.1. Subcategorize	307
D.5.4.2.1.2. Mark chemical by WS	326
D.5.4.2.1.3. Mark chemicals by descriptor values	329
D.5.4.2.1.4. Filter points by test condition	331
D.5.4.2.1.5. Mark focused chemical	333
D.5.4.2.1.6. Mark focused points	334
D.5.4.2.1.7. Remove marked chemicals/points	337
D.5.4.2.1.8. Clear existing marks	339
D.5.4.2.1.9. Mark outliers	340
D.5.4.2.2. Gap filling approaches	341
D.5.4.2.3. Descriptors/data	342
D.5.4.2.4. Model/(Q)SAR	345
D.5.4.2.5. Calculation options	348
D.5.4.2.5.1. Data usage options	351
D.5.4.2.5.1.1. Prediction examples	357
D.5.4.2.6. Visual options	365
D.5.4.2.7. Information	369

D.5.4.2.8. Miscellaneous	373
D.5.4.3. Right click functionality	375
D.5.5. Specific features of (Q)SARs	376
D.5.6. Automated and Standardized workflows	384
D.5.6.1. Automated workflow	385
D.5.6.2. Standardized workflow	392
D.5.7. Methodologies for estimating toxicity of a set of chemicals	400
D.5.7.1. Example: Quantitative mixtures toxicity prediction	402
D.5.7.1.1. Mixture models	411
D.5.7.1.1.1. Independent MOA	412
D.5.7.1.1.1.1. Illustration	414
D.5.7.1.1.2. Similar MOA	415
D.5.7.1.1.2.1. Illustration	417
D.5.7.1.1.2.2. Example	419
D.6. Report	423
D.6.1. Report window	423
D.6.2. Create a report	424
Chapter E. Options	436
E.1. General	436
E.2. Modules	441
E.3. Profiling	444
E.4. Data Gap Filling	446
E.5. Interface	455
E.6. Data	459

QSAR Toolbox End - user Manual

QSAR Toolbox v.4 QSAR Toolbox v.4

Chapter A. Introduction

[QSAR Toolbox End - user Manual](#) Chapter A. Introduction

The current online help is aimed at providing essential information on using the OECD QSAR Toolbox for Grouping Chemicals into Categories. The main objective of the Toolbox is to allow the user to use (Q)SAR methodologies to group chemicals into categories and to fill data gaps by read-across, trend analysis and (Q)SARs. For in-depth background information on the concept of chemical categories, the user is invited to consult the guidance document for grouping of chemicals published in the Series on Testing and Assessment of the OECD Environment, Health and Safety Publications [OECD (2007); ENV/JM/MONO(2007)28: [http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&ote=env/jm/mono\(2007\)28](http://www.oecd.org/officialdocuments/displaydocument/?doclanguage=en&ote=env/jm/mono(2007)28)].

Additional guidance and training material are available on the dedicated internet site for the QSAR Toolbox [<http://www.qsartoolbox.org>], the internet site for the OECD (Q)SAR Project [<http://www.oecd.org/env/existingchemicals/qsar>] as well as the internet site of the developer of the QSAR Toolbox [<http://toolbox.oasis-lmc.org/>]. The user is invited to regularly consult these internet sites.

The QSAR Toolbox is a project of the Organisation for Economic Co-Operation and Development in collaboration with the European Chemical Agency. It has been developed by the Laboratory of Mathematical Chemistry.

-0-

QSAR Toolbox v.4

A.1. Objective

[QSAR Toolbox End - user Manual](#)

This QSAR Toolbox End-user Manual is designed to give assistance in the use of the QSAR Toolbox software. It is intended to enable any user to learn how to use QSAR Toolbox without having to refer to help desk or training. It also serves as complete reference manual for all QSAR Toolbox features, functions, options and possible paths through the application.

-0-

A.2. How to use this guidance document

Chapter B. Welcome

QSAR Toolbox v.4

B.1. Acknowledgements

[Welcome](#) / B.1. Acknowledgements

B.1. Acknowledgements

The development of the OECD QSAR Toolbox is a large collaborative effort and many scientific teams and stakeholders are donating their skills and tools to be integrated into the Toolbox [see <http://www.oecd.org/env/chemicalsafetyandbiosafety/assessmentofchemicals/donorstotheqsartoolbox.htm>]

-0-

B.2. What is OECD QSAR Toolbox

[Welcome](#) / B.2. What is OECD QSAR Toolbox

B.2. What is OECD QSAR Toolbox

The OECD QSAR Toolbox is a software designed to reduce the use of animals in laboratory tests, reduce the cost for testing and increase the number of chemicals which are assessed for their effects upon human health and the environment. The OECD QSAR Toolbox provides scientific computational methods and information technologies for application of the category approach for filling gap in experimental data that are necessary for hazard and risk assessment. By making use of the system, hazard and risk assessors are able to:

- Use predefined categories, or to refine existing or build new categories.
- Identify analogous chemicals (or category) based on user selected characteristics. Categorize chemicals accounting for their metabolism: rate of disappearance, formation of stable metabolites, formation of high reactive intermediates, deactivation pathways, etc.
- Extract all available experimental or pre-calculated data from local and remote (web) based databases accompanied with information about their reliability: experimental error, analytical or computational method used, replicates, etc.
- Fill the gaps of missing information within the category by making use of chemometrics approaches such as read across, trend analysis, and (Q)SAR models.

QSAR predictions are accompanied with information concerning their mechanistic background, training chemicals, statistics, applicability domain and validity.

The OECD QSAR Toolbox is an expandable application that navigates the information flows between all of the installed components (modules): computational tools, database managers, (Q)SAR libraries, categorization models, etc.

QSAR Toolbox v.4

B.3. Abbreviations

[Welcome](#) / [B.3. Abbreviations](#)

B.3. Abbreviations

AP	Alert performance
AW	Automated workflow
BCF	Bioconcentration factor
EC50	Effective concentration for 50% of the organisms tested
IGC50	Statistically or graphically estimated concentration of test material, under specified concentrations, is expected to cause a 50% inhibition of growth
InChi	IUPAC International Chemical Identifier
Kow	Octanol-water partition coefficient
LC50	Statistically estimated concentration that is expected to be lethal to 50% of a group of organisms tested. Death may be defined by the mortality, intoxication and population effect groups
LOEC	Lowest concentration (LOEC) or level (LOEL) that has a statistically significant adverse effect on the tested organisms
MATC	Maximum Acceptable Toxicant Concentration; a hypothetical threshold concentration that is the geometric mean between the NOEC and LOEC concentration
NOEC	No Observed Effect Concentration/Level; Concentration or dose producing effects not significantly different from responses of controls according to author's reported statistical test
QSAR	Quantitative Structure-Activity Relationship
SAR	Structure-Activity Relationship
SMILES	Simplified Molecular Input Line Entry System
Sw	Water solubility
SW	Standardized workflow
UVBCs	Unknown or Variable Composition, complex reaction products, or

Biological materials

-0-

Chapter C. User interface

QSAR Toolbox v.4

C.1. Stages toolbar

[User interface](#) / C.1. Stages toolbar

C.1. Stages toolbar

The stages toolbar is a steady part of the Toolbox interface. It allows easy navigation between main stages of the program's workflow. Each stage is represented by a toolbar button, which invokes the interface related to the current stage. Some examples are provided below (Figure 1):

The toolbar button (1) and the interface (2) for the stage *"Input"*. (Figure 1)

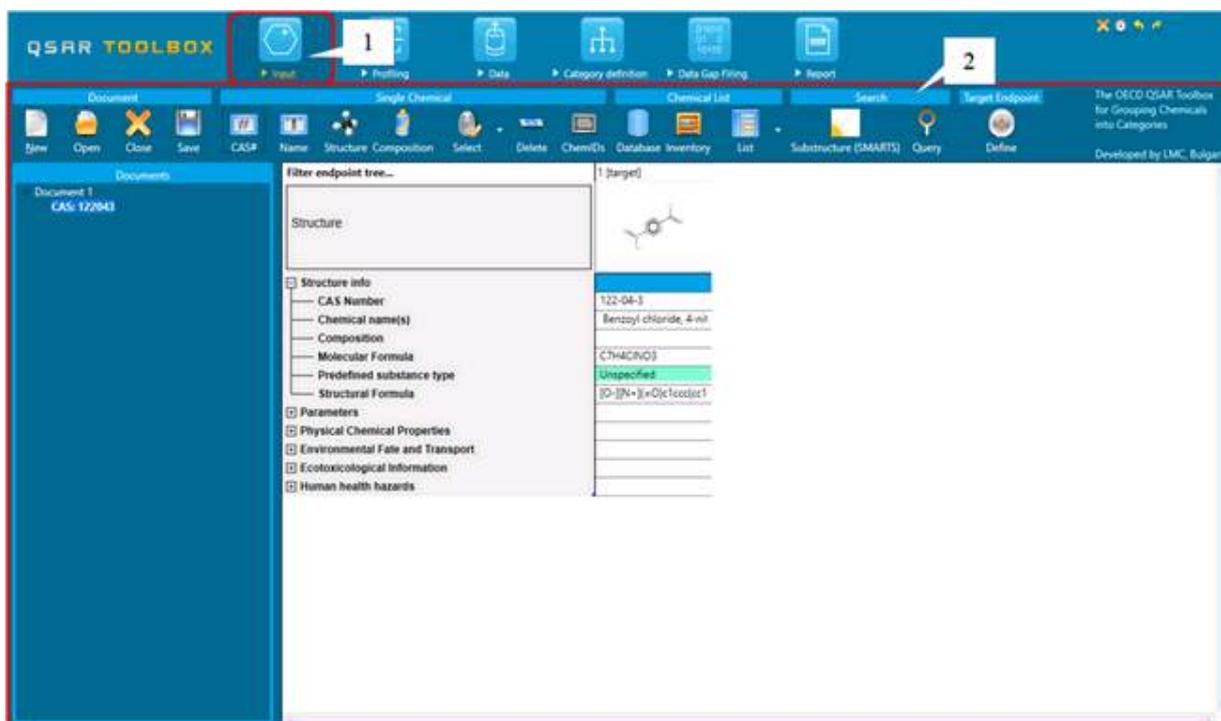


Figure 1

The toolbar button (1) and the interface (2) for the stage *"Data gap filling"*. (Figure 2)

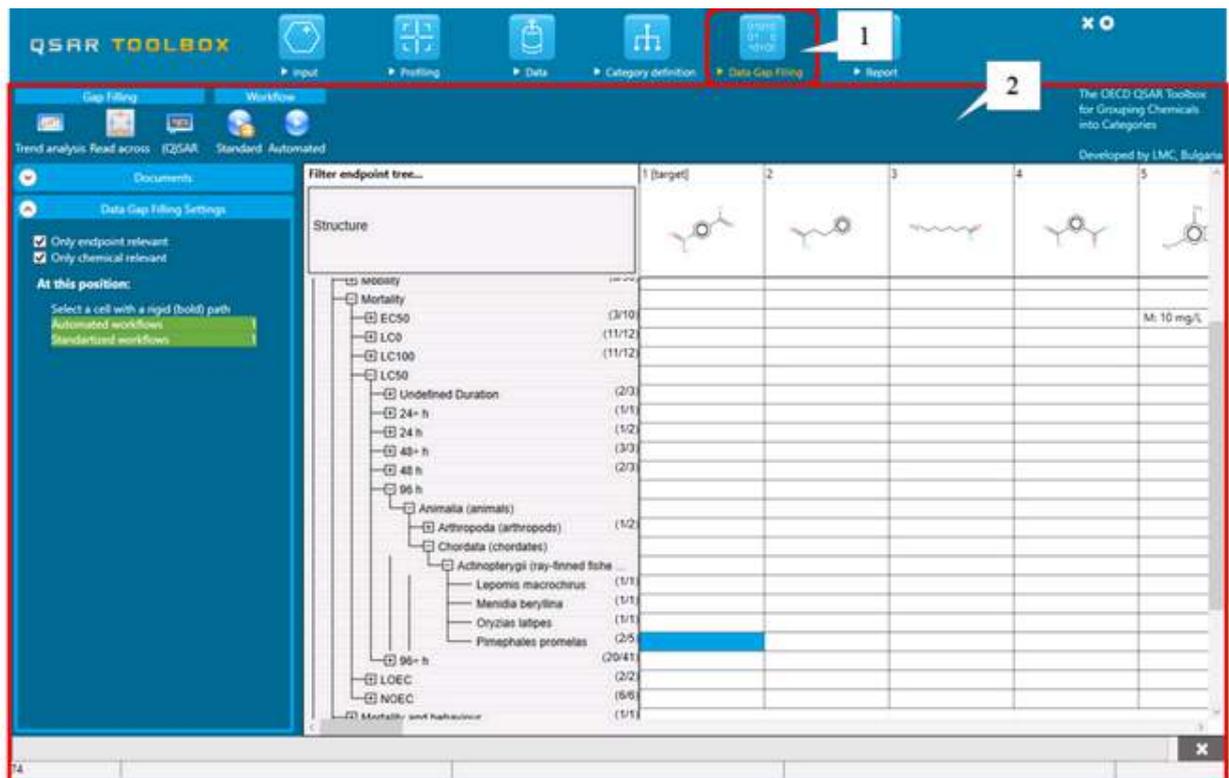


Figure 2

-0-

QSAR Toolbox v.4

C.2. Actions toolbar

[User interface](#) / C.2. Actions toolbar

C.2. Actions toolbar

The actions toolbar provides the basic actions for the stages of the program's workflow. Each stage has its specific actions, this is why the content of the toolbar varies between stages. For the users convenience the actions may be divided into groups. Some examples are provided below (Figure 1-2):

The basic actions of the stage "Input". (Figure 1)

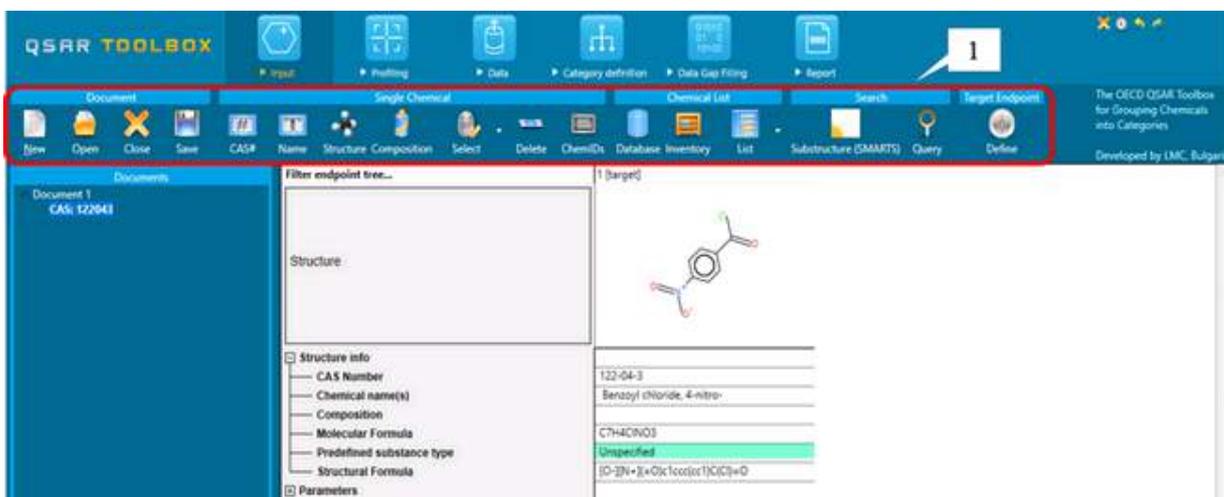


Figure 1

The basic actions of the stage "Data Gap Filling" . (Figure 2)

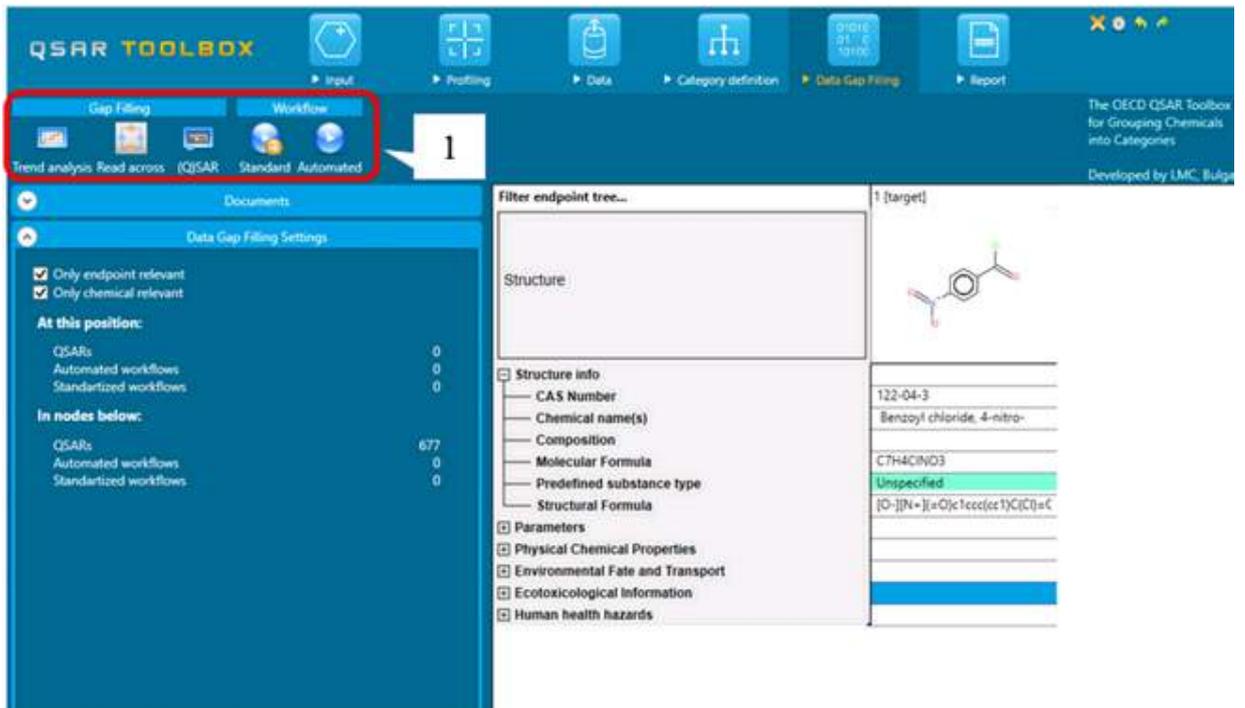


Figure 2

-O-

QSAR Toolbox v.4

C.3. Stage options panel

[User interface](#) / C.3. Stage options panel

C.3. Stage options panel

The stage option panel provides specific content for the current stage and actions related to this content. Each stage has its specific functions and that is why the stage option panel has different content. Some examples are provided below:

1. Input: The stage option panel in the Input stage gives the list with work documents, content of the documents. It also provides three approaches for multiplication of the target structures – multiplication by *tautomerism*, multiplication by *metabolism* and *target multiplication* in case the target is a substance with two or more constituents. Metabolism can be applied via 5 observed and 10 simulated metabolism simulators:

- Observed:
 - Observed Mammalian metabolism
 - Observed Microbial metabolism
 - Observed Rat In vivo metabolism
 - Observed rat liver metabolism with quantitative data
 - Observed Rat Liver S9 metabolism

- Simulated:
 - Autoxidation simulator
 - Autoxidation simulator (alkaline medium)
 - Dissociation simulator
 - Hydrolysis (acidic)
 - Hydrolysis (basic)
 - Hydrolysis (neutral)
 - In vivo Rat metabolism simulator
 - Microbial metabolism simulator
 - Rat liver S9 metabolism simulator
 - Skin metabolism simulator

1. Multiplication by Metabolism – select Skin metabolism simulator

In order to accomplish multiplication of the loaded target structure the user should apply a right click on the target in the stage option panel (1), select Multiplication from the pop-up menu (2), then press Metabolism/ Transformations (3) and finally select a simulator, for example Skin metabolism simulator (4). (Figure 1)

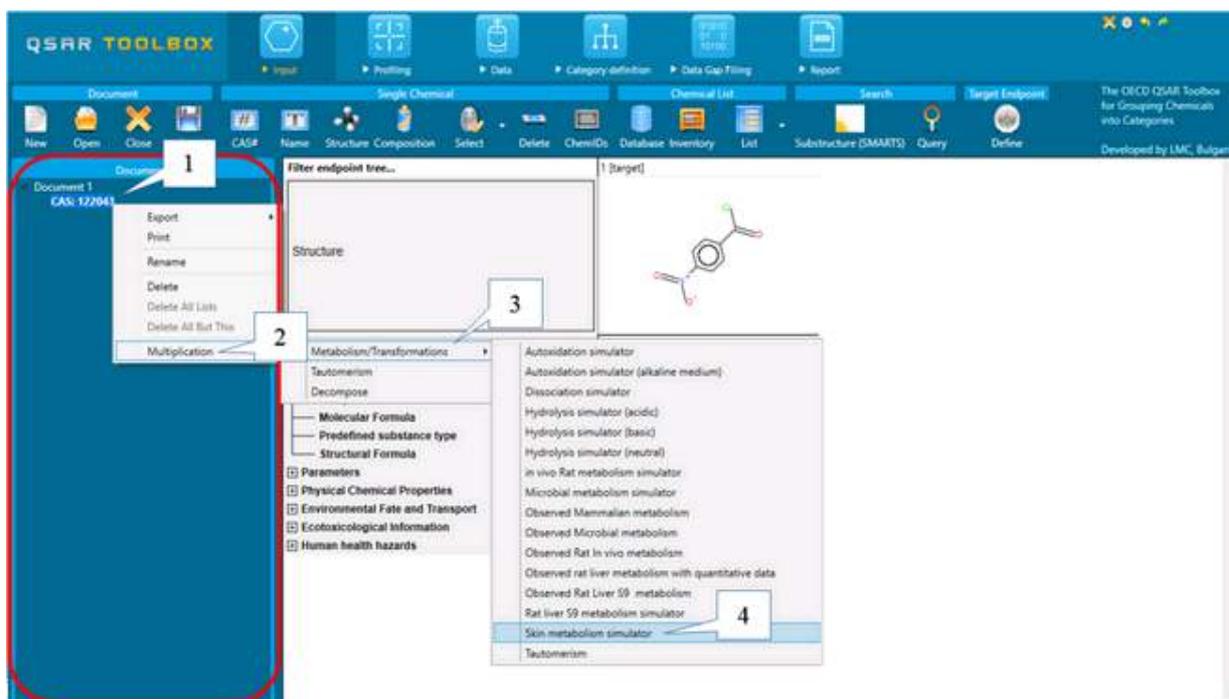


Figure 1

The same procedure could be applied in order to multiply the structure by one of the three observed metabolic simulators.

2. Multiplication by Tautomerism

Right click on the structure (1), select *Multiplication* from the pop-up menu (2) and then press *Tautomerism* (3). (Figure 2)

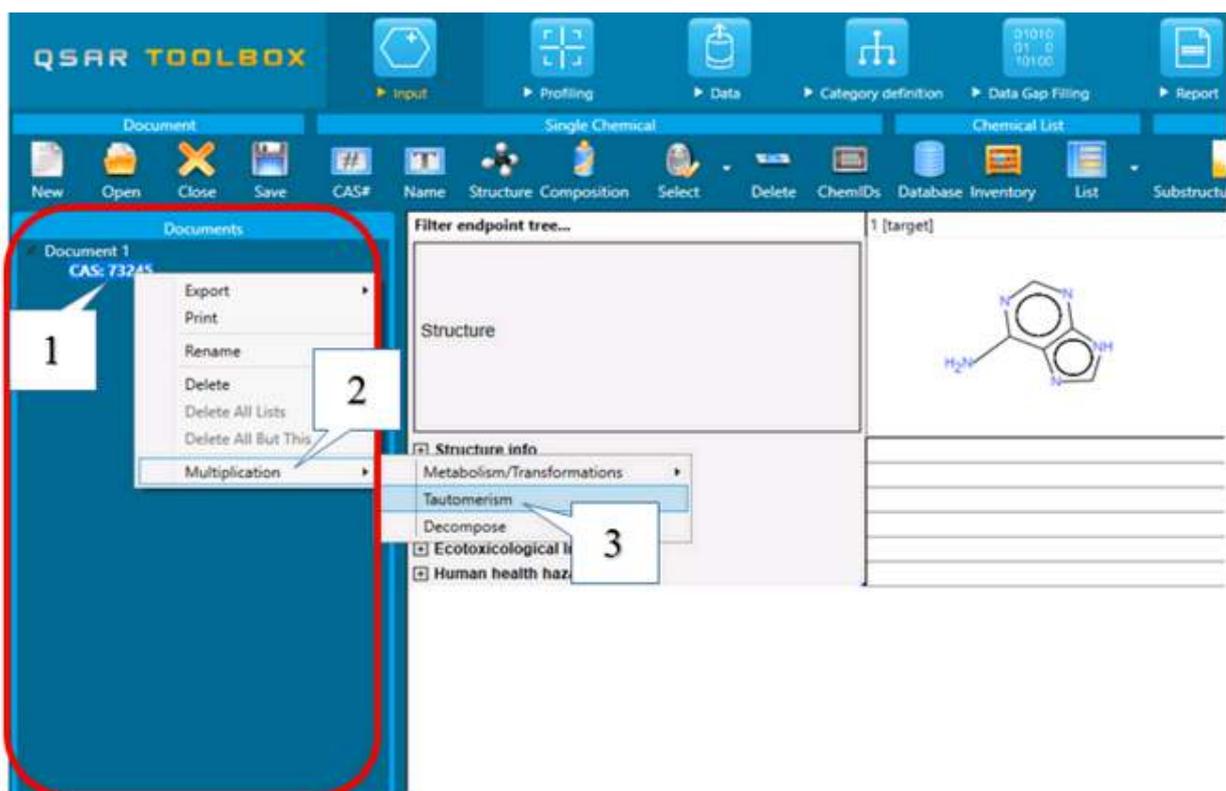


Figure 2

A parent list (1) consists of the whole tautomeric set and some child lists (2) for each of the set`s constituents are created. (Figure 3)

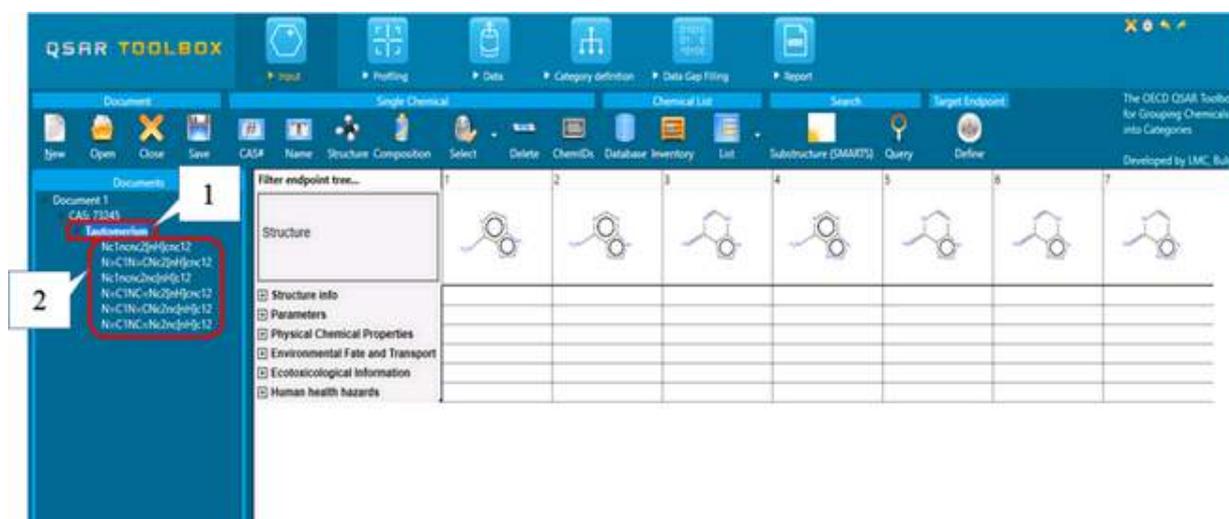


Figure 3

3. Multiplication by Target multiplication

Some of the substances consist of more than one constituents (including

mixtures, presence of additives, impurities, etc.). In order to see all individual components of such kind of substances, the user should apply a right click on the target in the stage option panel (1), select *Multiplication* from the pop-up menu (2) and then select *Target multiplication* (3). (Figure 4) A child-list for each of the available constituents (4) in the substance composition will be created.

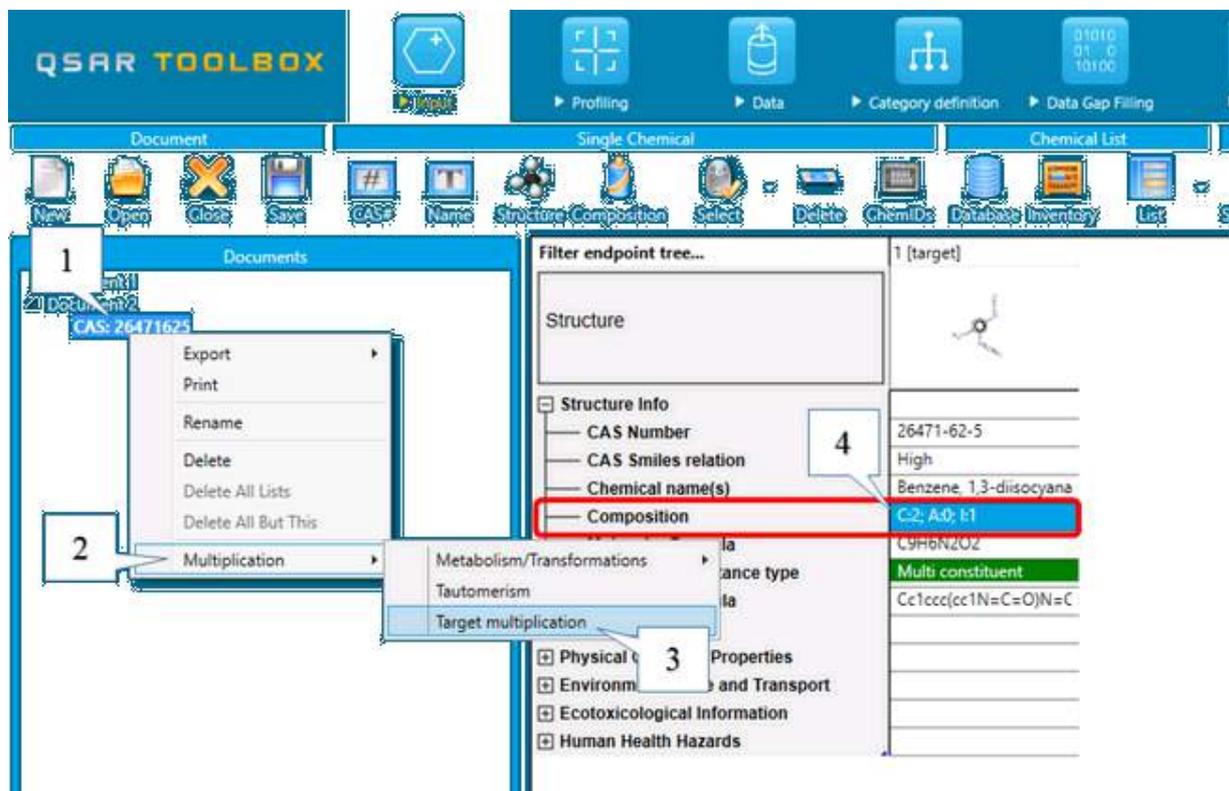


Figure 4

II. Profiling

New functionalities determine possibility to provide profilers relevant to the endpoint (1). The user is able to Color (2) and Group (3) the profilers in the Profiling module by *Relevancy* according to the endpoint of interest. Meaning of the colors can be seen by click on the Legend (5) button. The same functionalities can be used in data gap filling during the subcategorization process. (Figure 5)

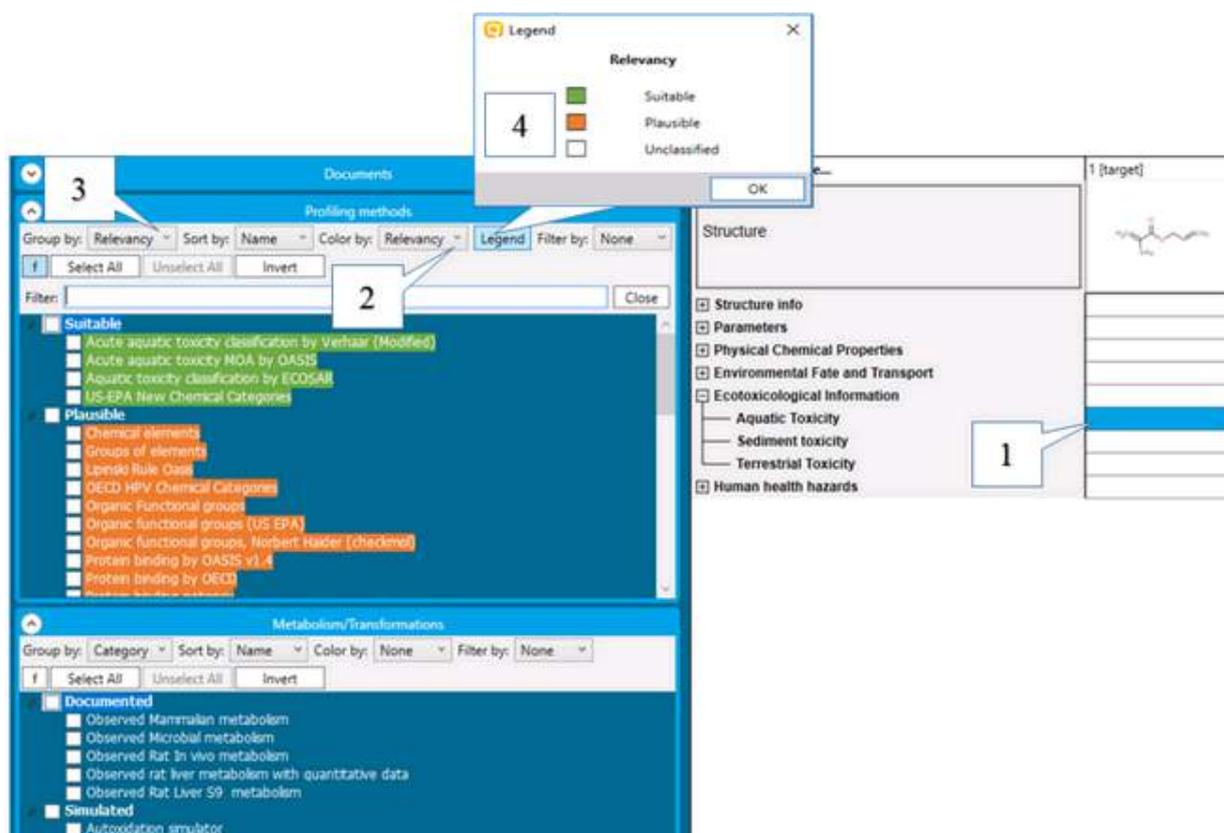


Figure 5

III. Data

Toolbox v.4.0 gives the users the possibility to see which of the databases contain data related to their endpoint of interest (1) by coloring of the databases according to the *Endpoint selected in the data matrix* (2) (or *Target endpoint* if the endpoint is preliminary defined). These of them, which meet the requirement are highlighted in green (3). (Figure 6)

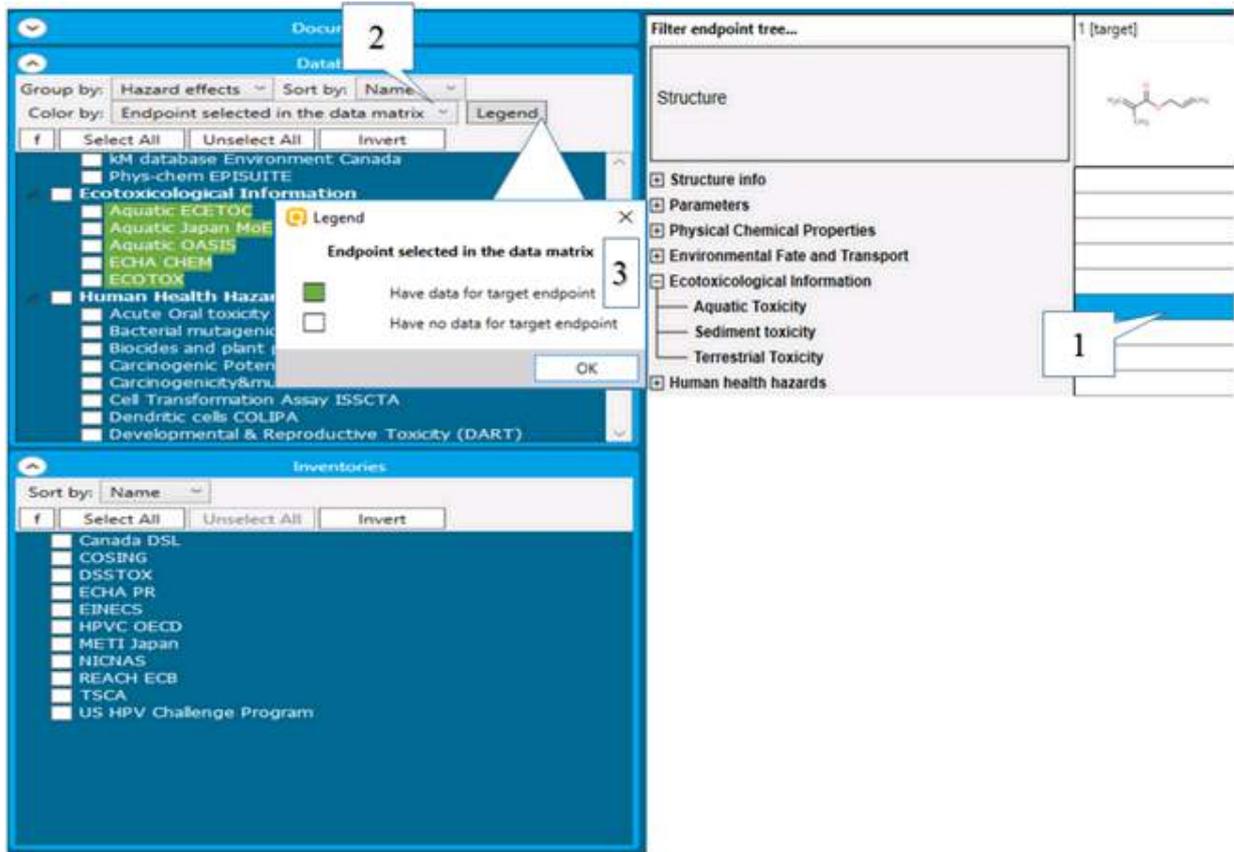


Figure 6

-0-

QSAR Toolbox v.4

C.4. Data matrix

[User interface](#) / C.4. Data matrixC. 4. Data matrix

Below is a snapshot displaying the data matrix window (1). (Figure 1)

The screenshot displays the QSAR Toolbox Data Matrix window. The interface is divided into several sections:

- Filter endpoint tree... (1):** A tree view on the left showing various endpoints such as Structure, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, Aquatic Toxicity (Growth, Growth inhibition, Immobilisation, Mortality, Physiology, Reproduction), Sediment toxicity, Terrestrial Toxicity, and Human health hazards.
- Structure (2):** A column showing chemical structures for seven different compounds.
- Data Matrix (3):** A table with columns for chemical structures and rows for various endpoints. The table contains numerical data for endpoints like AN SV, Growth, Growth inhibition, Immobilisation, Mortality, Physiology, Reproduction, Sediment toxicity, Terrestrial Toxicity, and Human health hazards.

Figure 1

The data matrix window has three main parts (Figure 2):

- Area with the [Endpoint tree](#) (1)
- [Area with the selected chemicals](#) (2) and
- [Area with data](#) (experimental, predicted) (3)

The screenshot displays the OECD QSAR Toolbox application. The interface includes a top navigation bar with icons for 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. Below this, there are sections for 'Documents', 'Databases', and 'Inventories'. The main workspace is divided into several panels:

- Filter endpoint tree:** A tree view on the left showing various toxicity endpoints such as 'Growth', 'Growth inhibition', 'Immobilization', 'Mortality', 'Physiology', 'Reproduction', 'Sediment toxicity', and 'Human health hazards'. A red box labeled '1' highlights this tree.
- Structure:** A panel showing chemical structures. A red box labeled '2' highlights a row of seven chemical structures.
- Data Matrix:** A table displaying numerical data for selected endpoints. A red box labeled '3' highlights a specific row of data.

AW SW	M: 152 mg/L	M: 47.6 mg/L	M: 57.3 mg/L	M: 88.8 mg/L	M: 216 mg/L
(79/89)					
(8/14)					
(12/12)					
(1/1)					
(66/105)	M: 9.79 mg/L		M: 2.18 mg/L	M: 5.01 mg/L	M: 2.71 mg/L
(13/22)			M: 9.95 mg/L		
(5/12)					

Figure 2

-O-

QSAR Toolbox v.4

C.4.1. Endpoint tree area

[User interface](#) / [Data matrix](#) / C.4.1. Endpoint tree area

C.4.1. Endpoint tree area

Content of the Endpoint tree area:

- [Nodes of the Endpoint tree](#)
- [Construction](#)
- [Set tree hierarchy](#)
- [Filtering nodes of the Endpoint tree](#)
- [Tips](#)

4.1.1. Nodes of the Endpoint tree:

The Endpoint tree has five general nodes:

- Substance info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazard

The *Structure info* includes subnodes displaying the substance information of the selected chemical(s) like CAS number; CAS Smiles relation; Chemical Name(s), Composition, Molecular formula, Predefined substance type and Structural formula (Figure 1):

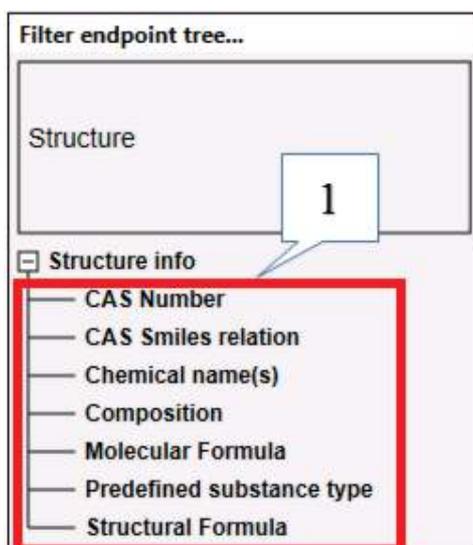


Figure 1

The *Parameters* are separated into two subnodes - 2D and 3D (1). (Figure 2)

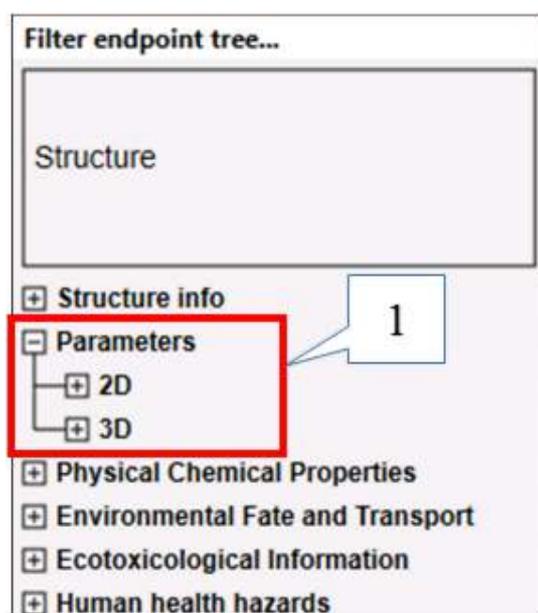


Figure 2

Calculating the desired parameter is possible when the user clicks the right mouse button over the desired parameter (1) and selects one of the available options: to calculate all parameters, to calculate only 2D (3D) parameters or to calculate only the current parameter (2). These options are used for the calculation of the parameter(s) for all chemicals loaded in the data matrix. (Figure 3)

The screenshot displays the 'Filter endpoint tree...' on the left and a table of parameters on the right. The table has one column for the parameter name and one for its status. The 'log Kow' row is highlighted in blue, and a context menu is open over it. The context menu options are:

- Export Data matrix
- Expand branch
- Collapse branch
- Expand All
- Collapse All
- Target endpoint
- Copy Path
- Calculate/extract all parameters for all chemicals
- Calculate/extract all 2D parameters for all chemicals
- Calculate/extract "log Kow" for all chemicals
- Activate AOP

The table of parameters is as follows:

Parameter	Status
Kp (Mackay method)	Not calculated
Kp (Octanol/air (Koa) model)	Not calculated
log BCF max	Not calculated
Log Koa (Air-water partition coefficient model)	Not calculated
Log Koa (Henry's law constant model)	Not calculated
log Kow	Not calculated
logP Multicase	Not calculated
Mean Melting Point	Not calculated
Melting Point (Adapted Job	Not calculated
Melting point (Gold and Og	Not calculated
Molar refraction I	Not calculated
Molar refraction II	Not calculated
Molecular Weight	Not calculated
Number of aromatic bonds	Not calculated
Number of cyclic bonds	Not calculated
Number of double bonds	Not calculated
Number of heavy atoms	Not calculated
Number of nitro groups	Not calculated
Number of rings	Not calculated
Number of single bonds	Not calculated
Number of single bonds (with H)	Not calculated
OVERALL OH Half-life	Not calculated
OVERALL OH rate constant	Not calculated
OVERALL OZONE Half-life	Not calculated
OVERALL OZONE rate constant	Not calculated
phi (Junge-Pankow model)	Not calculated
phi (Mackay model)	Not calculated

Figure 3

Calculation of a parameter for one specific chemical or for all chemicals is possible when user position over the cell of the chemical corresponding to the desired parameter (1) and selects one of the options from the pop-up menu (right mouse click) (2): Calculate/extract or Calculate/Extract all 2D parameters, etc. (Figure 4)

The screenshot displays the 'Filter endpoint tree...' window in the OECD QSAR Toolbox. On the left, a tree view lists various parameters. The 'log Kow' parameter is highlighted in blue. To the right, a table shows the calculation status for each parameter, with 'log Kow' also highlighted in blue. A context menu is open over the 'log Kow' row, listing several actions. Callout box '1' points to the 'log Kow' row in the table, and callout box '2' points to the 'Calculate/extract "log Kow"' option in the context menu.

Parameter	Status
Koc (Log Kow)	Not calculated
Koc (MCI)	Not calculated
Kp (Mackay method)	Not calculated
Kp (Octanol/air (Koa) model)	Not calculated
log BCF max	Not calculated
Log Koa (Air-water partition coefficient model)	Not calculated
Log Koa (Henry's law constant model)	Not calculated
log Kow	Not calculated
logP Multicase	Not calculated
Mean Melting Point	Not calculated
Melting Point (Adapted Joback Method)	Not calculated
Melting point (Gold and Ogle method)	Not calculated
Molar refraction I	Not calculated
Molar refraction II	Not calculated
Molecular Weight	Not calculated
Number of aromatic bonds	Not calculated
Number of cyclic bonds	Not calculated
Number of double bonds	Not calculated
Number of heavy atoms	Not calculated
Number of nitro groups	Not calculated
Number of rings	Not calculated
Number of single bonds	Not calculated
Number of single bonds (with H)	Not calculated
OVERALL OH Half-life	Not calculated

Figure 4

Experimental data available in Toolbox databases is assigned to the other four general nodes and their subnodes. These four nodes are separated in four basic sections depending on the type of the assigned experimental data.

For example results for melting point or partition coefficients (Figure 5) are assigned to the nodes Melting/freezing point or Partition Coefficient, which are sub-nodes of the node Physical Chemical Properties (1),

Filter endpoint tree... 1 [target]

Structure

NC(=O)c1ccccc1

Structure info

Parameters

Physical Chemical Properties

- Autoflammability / Self-ignition temperature
- Boiling point (1/1) M: 290 °C
- Density
- Dissociation Constant (pKa) (1/1) M: 1.82
- Explosive properties
- Flammability
- Flash point
- In chemico (1/1) M: Not reactive at saturation
- Lipid Solubility
- Melting / freezing point (1/1) M: 129 °C**
- Other property
- Oxidation reduction potential
- Oxidising properties
- Particle size
- Partition Coefficient: (1/1) M: 0.64**
- Solubility in organic solvents / fat solubility
- Stability in organic solvents and identity of relev ...
- Surface tension
- Vapour pressure
- Viscosity
- Water solubility (1/1) M: 1.35E+04 mg/L

1

Figure 5

or data associated with the Ames test or Chromosomal aberration are assigned to the nodes Bacterial reverse mutation assay (e.g. Ames Test) and In vitro mammalian chromosome aberration test, which are subnodes of the node Human Health Hazards (2) (Figure 6)

The screenshot shows the 'Filter endpoint tree...' window. The left pane displays a tree structure with the following categories and sub-items:

- Structure
- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human health hazards** (highlighted with a red box)
 - Acute Toxicity
 - Bioaccumulation
 - Carcinogenicity
 - Developmental toxicity / teratogenicity
 - Genetic Toxicity
 - In Vitro
 - Bacterial reverse mutation assay (e.g. Ames (1/19)) M: Negative
 - In vitro mammalian chromosome aberration test (1/3) M: Positive** (highlighted with a red box)
 - In Vivo (1/3) M: Positive
 - Immunotoxicity

The right pane shows a table with one row: '1 [target]' and a chemical structure icon.

Figure 6

4.1.2. Construction of Endpoint tree

The Endpoint tree is constructed in two parts: a predefined and a dynamic part. The predefined part is rigid and cannot be reordered while the dynamic part is flexible and can be reordered. This functionality is implemented due to the diversity of experimental data available from different databases. The predefined part is bolded by default (Figure 7) and the dynamic is not (Figure 8)

- Predefined part (Figure 7)

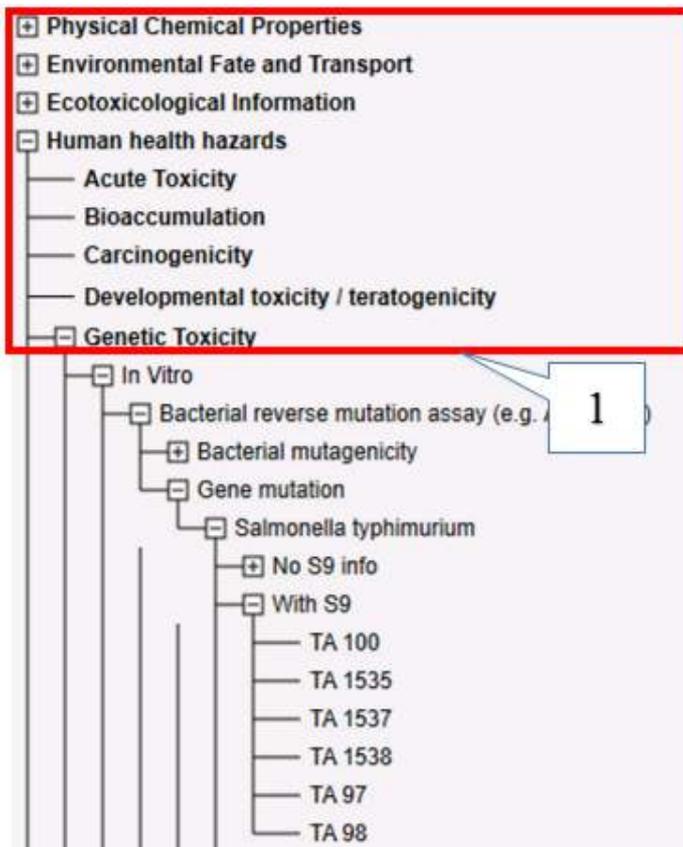


Figure 7

- o Dynamic part (Figure 8)

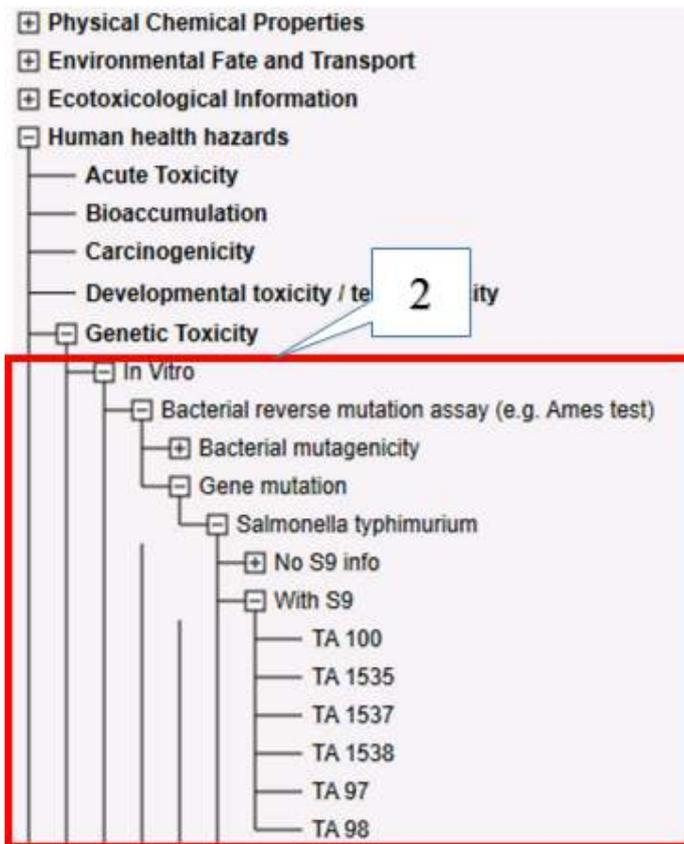


Figure 8

The metadata fields associated with the experimental data is used to build the dynamic part of the endpoint tree. So in this case *in vitro* and *in vivo* are elements of the metadata field called Type of method (1) (Figure 9)

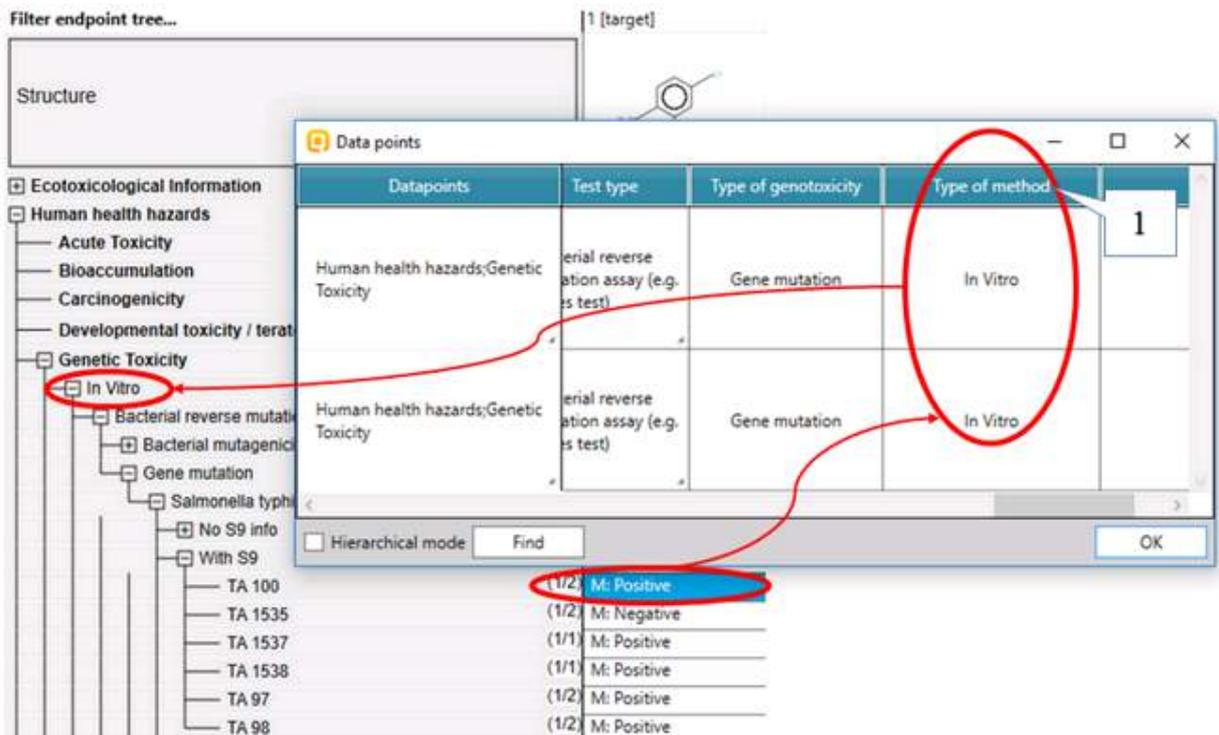


Figure 9

The next node Bacterial reverse mutation assay (e.g Ames test) is the Test type (2) (Figure 10)

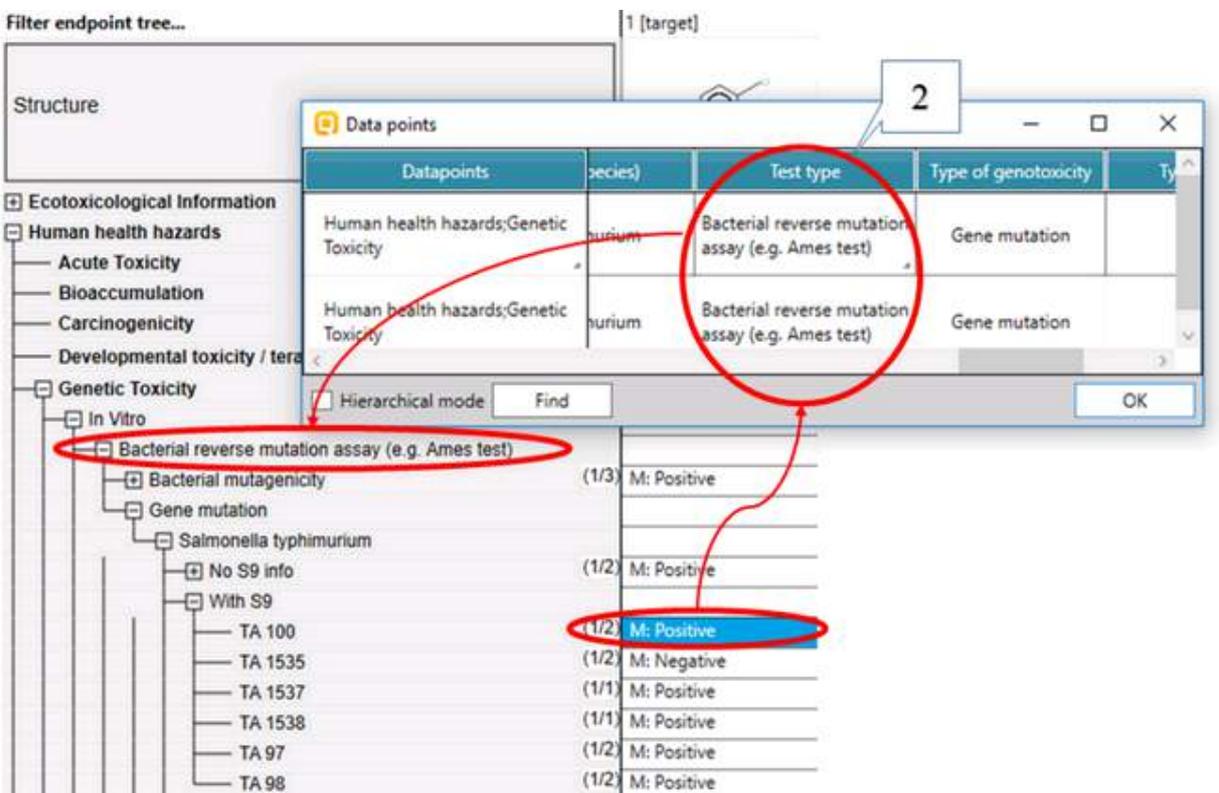


Figure 10

The subsequent two nodes Gene mutation and Salmonella typhimurium are associated with the field Type of genotoxicity and the field Test organism (species) (3). (Figure 11)

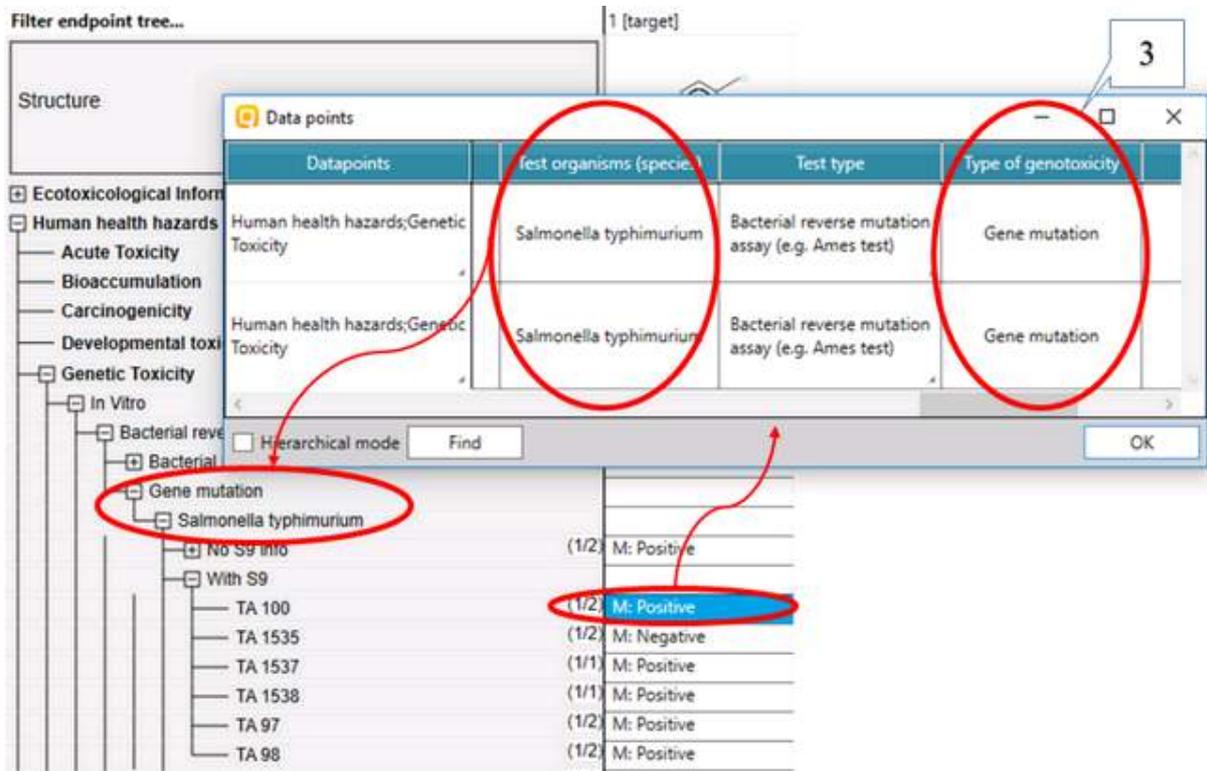


Figure 11

The last two nodes are associated with the following two fields: Metabolic activation and Strain (4). (Figure 12)

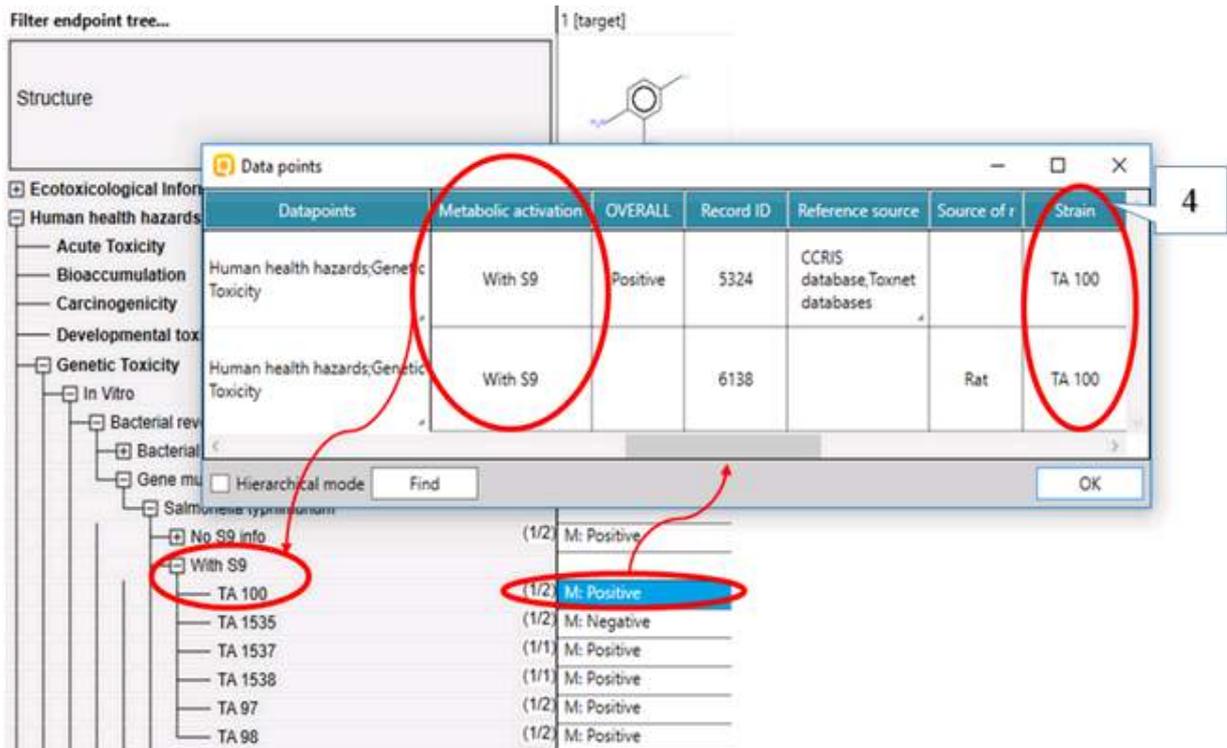


Figure 12

Each of these fields can be reordered using the Set tree hierarchy functionality. This option is available by applying a right mouse click over the node where the corresponding hierarchy should be reordered (1) (or this is the rigid part) and then clicking on Set tree hierarchy (2) from the context menu. The little blue triangles indicate the levels of the nodes where the hierarchy could be set (3). (Figure 13)

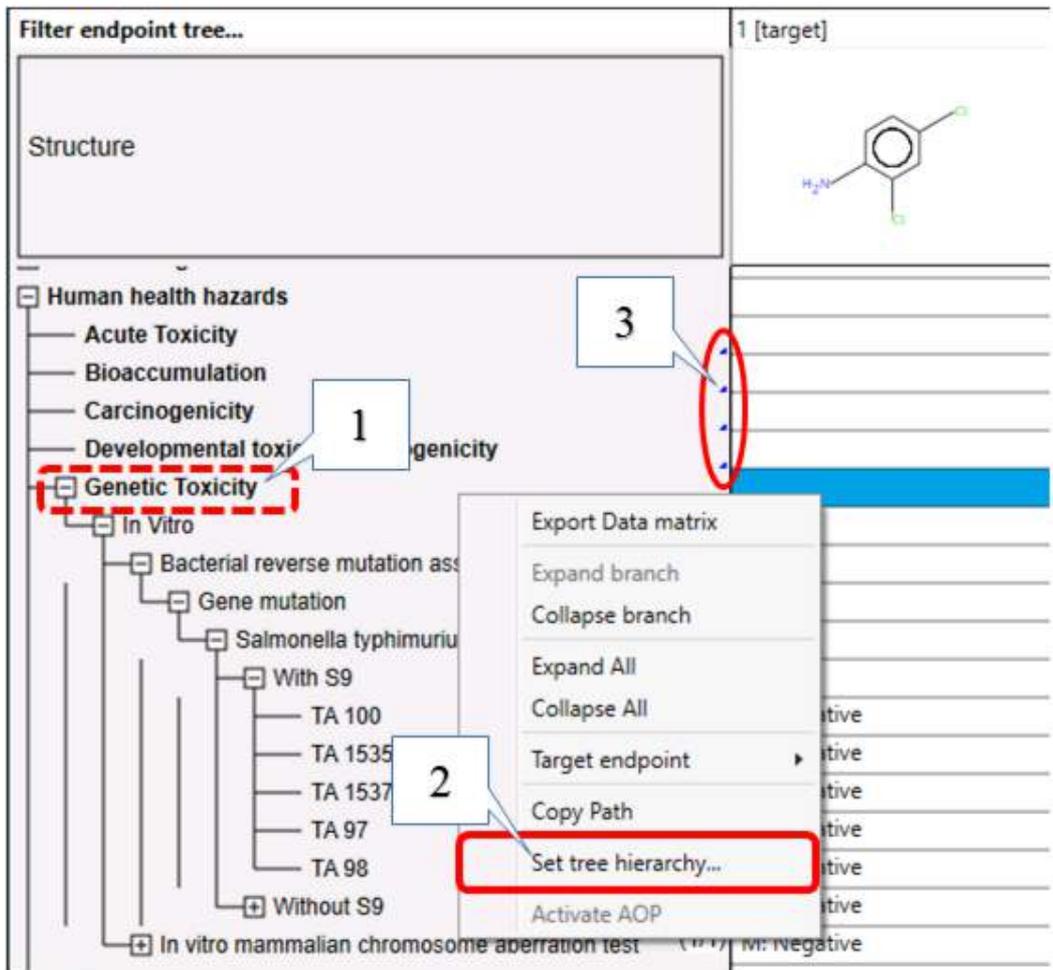


Figure 13

The Set tree hierarchy window appears.

4.1.3. Set tree hierarchy functionality

The Set tree hierarchy window contains two panels: Metadata labels (1) and Sub-nodes (2) (Figure 14)

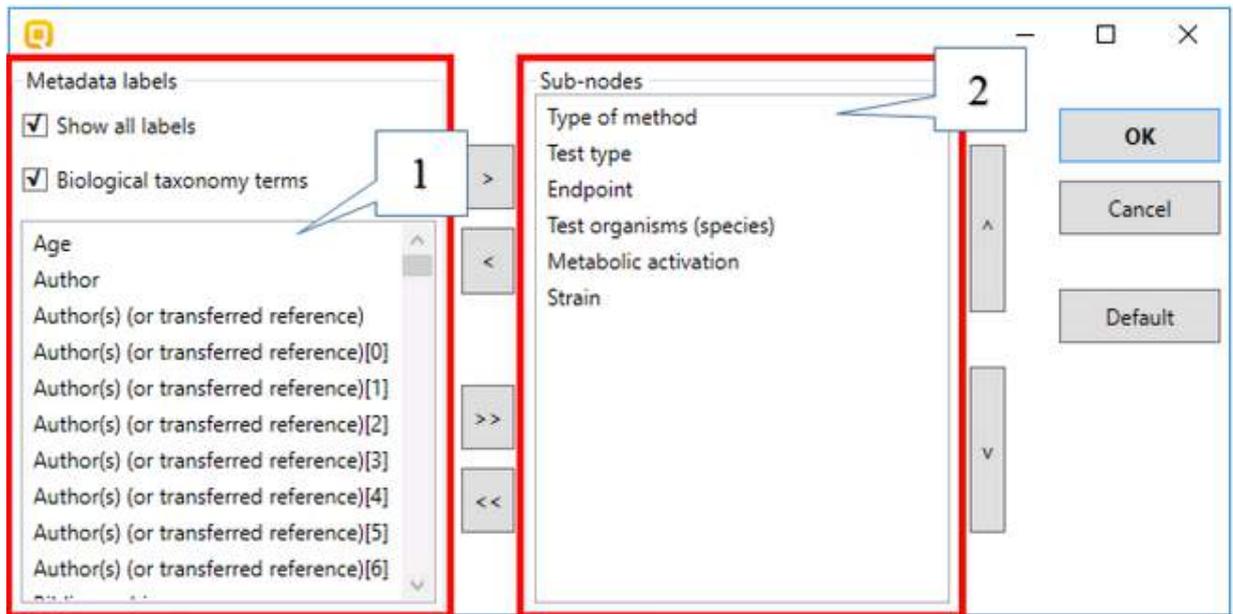


Figure 14

The Toolbox comes with default hierarchy. The panel with Metadata labels contains a list with most usable fields. If the user wants to set another field as a sub-node he/she should check the Show all labels box (1), then the list with all labels available in different databases appears (2). (Figure 15)

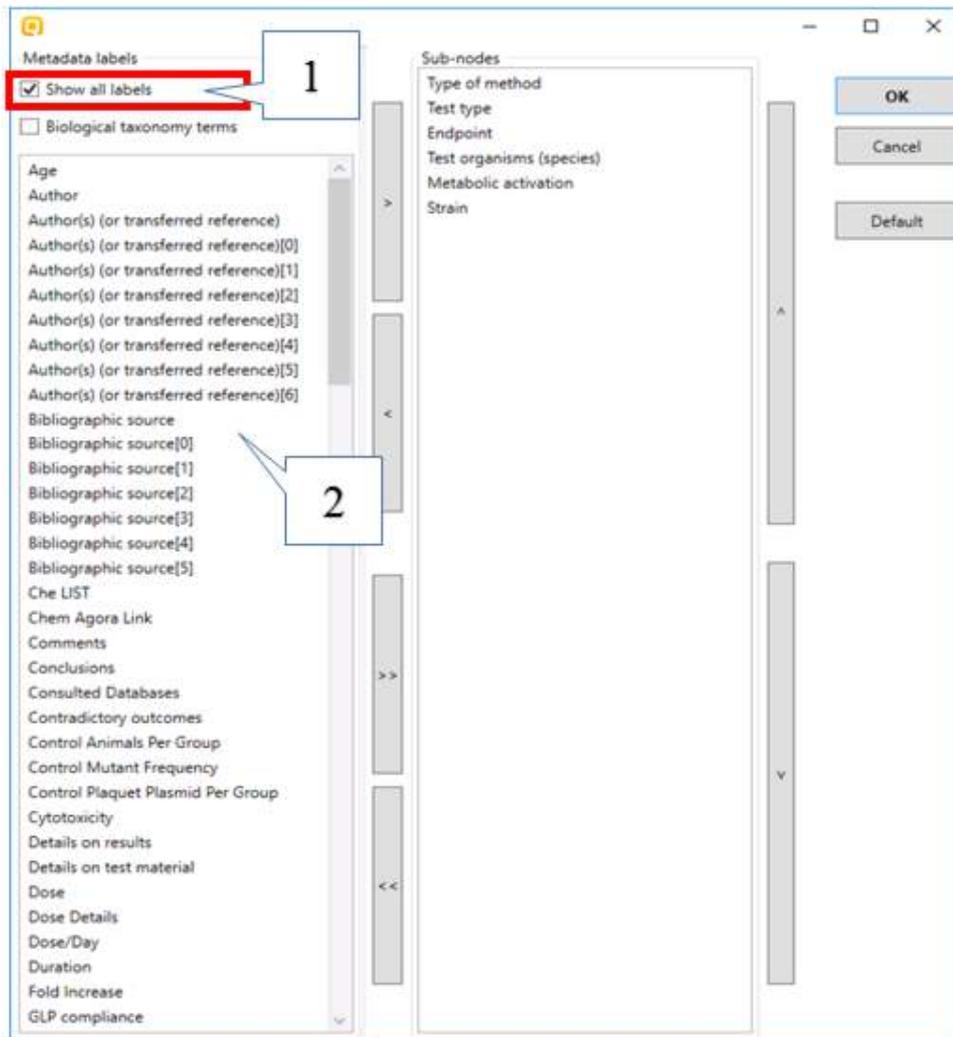


Figure 15

All terms related to the biological taxonomy of the test organisms can be listed in addition by checking the Biological taxonomy terms box (1). Then the terms appears after the available metadata labels (2). (Figure 16)

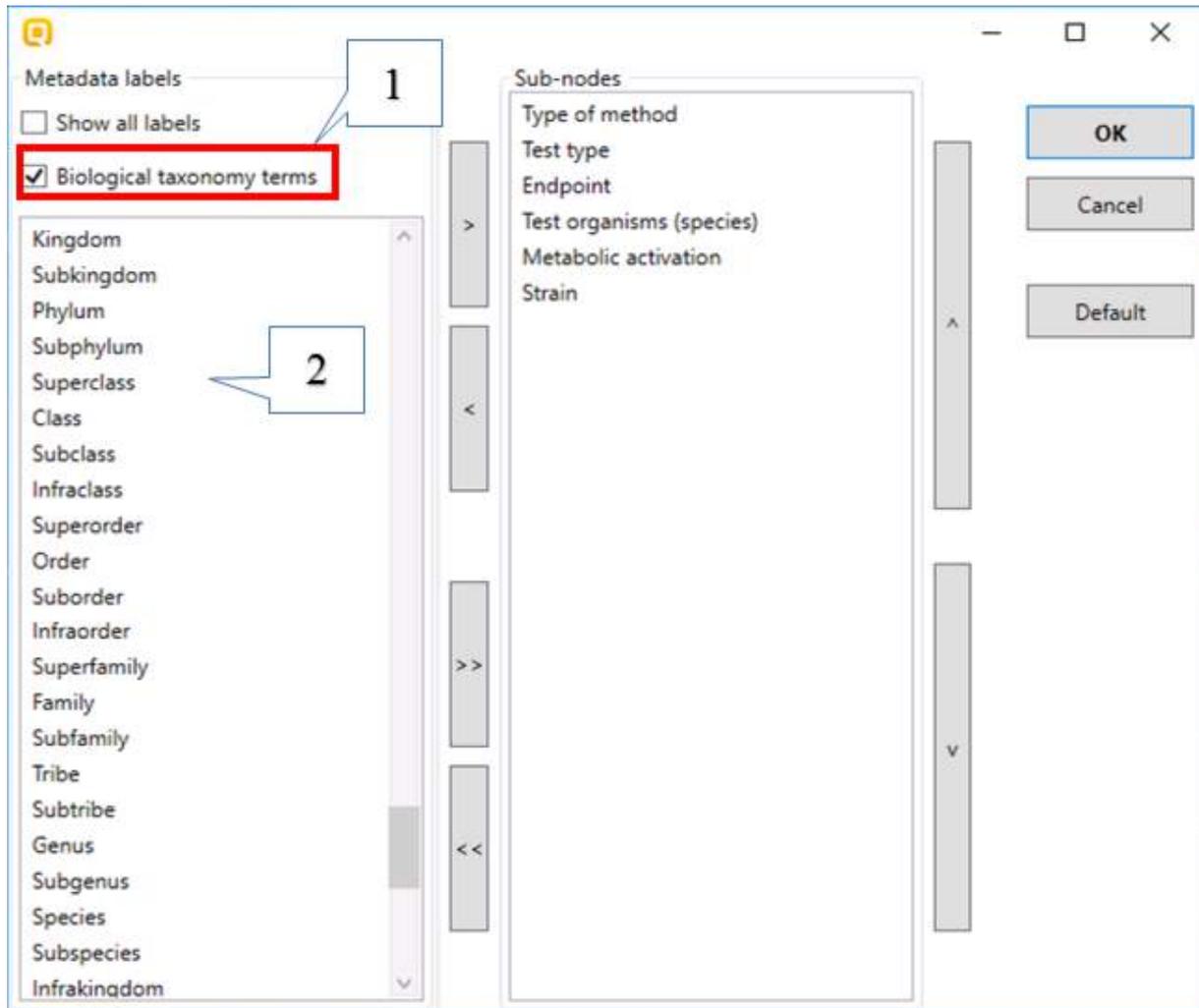


Figure 16

The sequence of fields (1) displayed in the Sub-nodes panel specifies the organization of the nodes of the endpoint tree (2). (Figure 17)

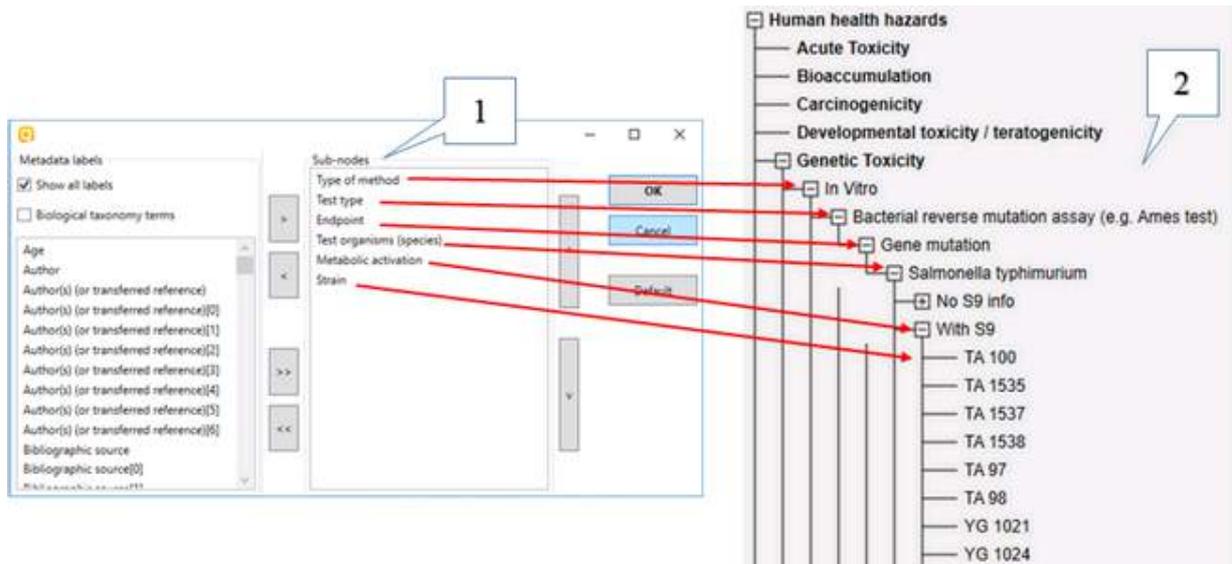


Figure 17

The user can add or remove fields already specified as sub-nodes using the auxiliary buttons (1). The sub-nodes can be reorder using the Up and Down buttons (2). If the user wants to reset the default setting of the endpoint tree then he/she can click the Default button (3). (Figure 18)

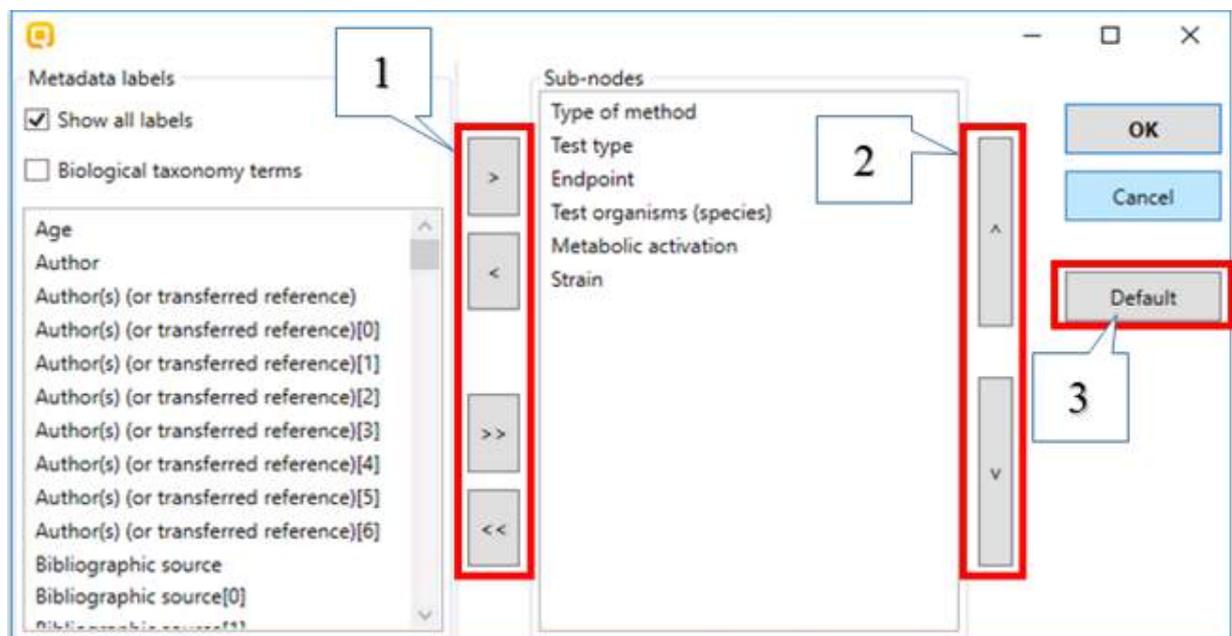


Figure 18

The changes in the endpoint hierarchy are confirmed by pressing the OK button.

4.1.4. Filtering nodes of the Endpoint tree

The nodes of the endpoint tree can be filtered using the Filter endpoint tree... functionality. In order to filter the endpoint tree, the user should write the desired query in the blank field named Filter endpoint tree... (1), where the white field Filter endpoint tree... becomes green colored (2) and then have to press the *Enter* button of the keyboard (3) (for instance write "skin" and press enter, then the endpoint tree is filtered and only nodes related to skin are visible) (Figure 18 -19)

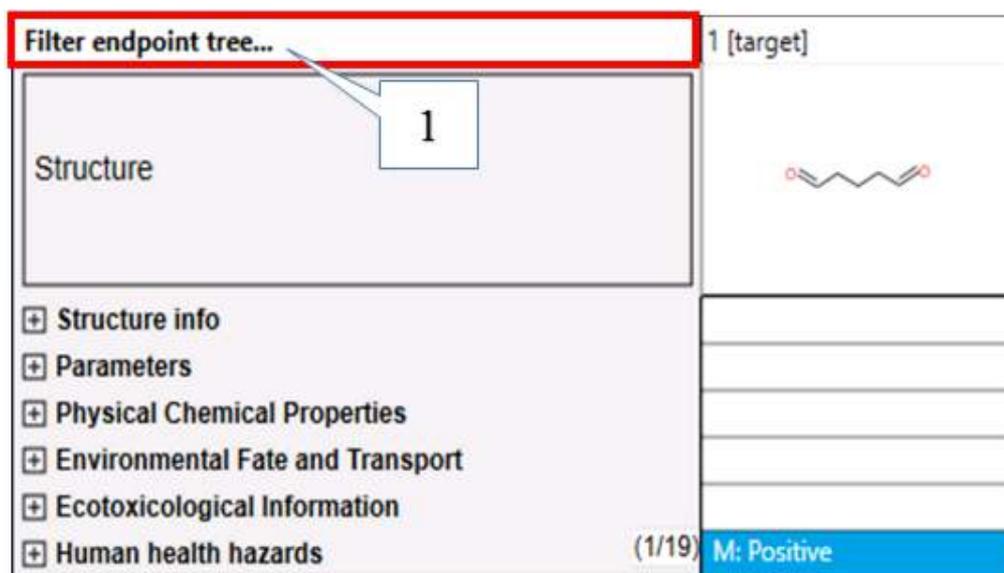


Figure 18

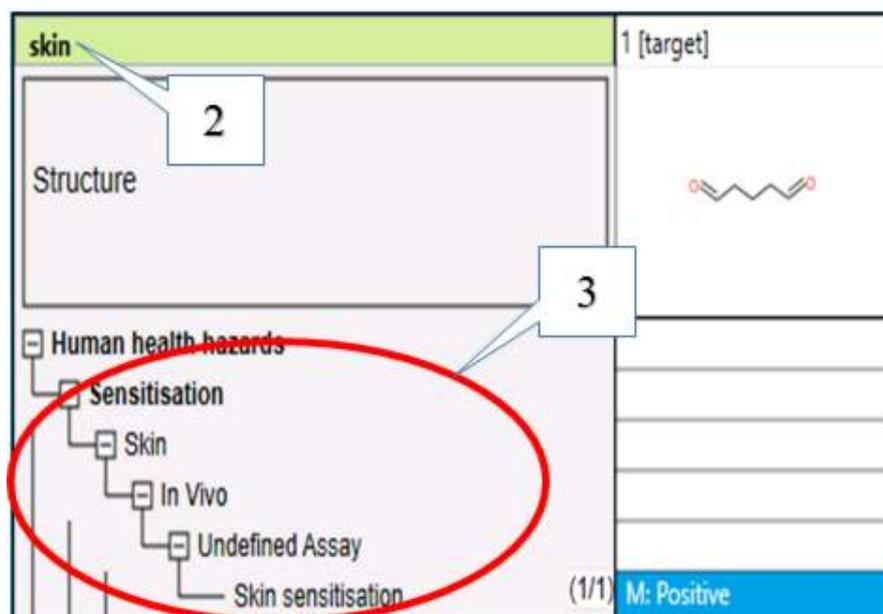


Figure 19

When the user deletes the defined query and press *Enter* button of the keyboard, then the system restore the default settings of the endpoint tree.

4.1.5. Tips

Some additional features are available by applying a right mouse click over the area of the endpoint tree.

- Expand – Expand branch (1) and Expand all (2) options allow to expand the current node or all collapsed nodes on the endpoint tree. The Expand branch option is active only for branches, which could be expanded (3) (Figure 20)

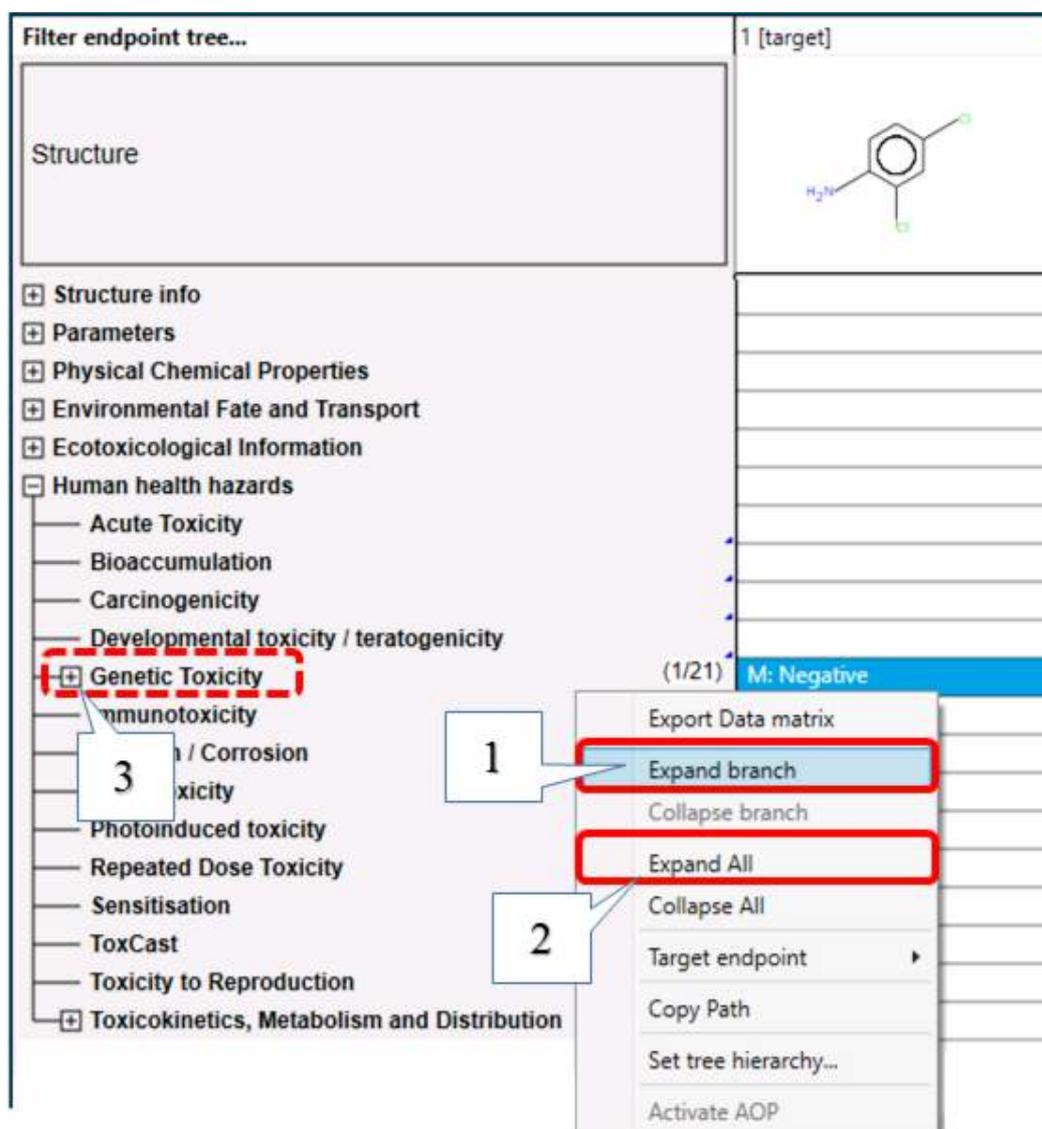


Figure 20

- Collapse – Collapse branch (1) and Collapse all (2) options allow to

collapse the current node or all expanded nodes on the endpoint tree. The Collapse branch option is active only for branches, which are already expanded (3). (Figure 21)

The screenshot displays the 'Filter endpoint tree...' window. On the left, a tree structure is shown under 'Human health hazards'. The 'Genetic Toxicity' node is highlighted with a red dashed box and labeled '3'. A context menu is open over the 'Gene mutatic' node, with 'Collapse branch' and 'Collapse All' options highlighted in red and labeled '1' and '2' respectively. The 'Structure' panel shows a chemical structure of 3,4-dichloroaniline. The '1 [target]' panel shows a table with one row containing '1 [target]'.

Figure 21

- Export Data matrix – this option allows to export the available data from the data matrix (1). (Figure 22)

The screenshot displays the 'Filter endpoint tree...' interface. On the left, a tree structure shows various endpoints under categories like 'Aquatic Toxicity', 'Sediment toxicity', and 'Human health hazards'. The 'Aquatic Toxicity' branch is expanded to show 'Growth', which includes 'EC50', 'IGC50', and '48 h'. '48 h' is further expanded to show 'Protozoa', 'Ciliophora', and 'Ciliatea', with 'Tetrahymena pyriformis' selected. The right pane shows a table with columns for 'AW SW', 'M', and 'M:'. The row for 'Tetrahymena pyriformis' is highlighted in blue, showing 'M: 44.6 mg/L'. A context menu is open over this row, with 'Export Data matrix' highlighted. A callout box labeled '1' points to this option.

AW SW	M	M:
(1/2)		M: 61.6 mg/L
(1/1)		M: 44.6 mg/L

Figure 22

- Copy path – this option allows to copy the endpoint path (1) to the current position (2) (Figure 23)

Filter endpoint tree... 1 [target]

Structure

Nc1ccc(Cl)cc1Cl

- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
 - Aquatic Toxicity
 - Growth
 - EC50 (1/2) M: 61.6 mg/L
 - IGC50
 - 48 h
 - Protozoa
 - Ciliophora
 - Ciliata
 - Tetrahymena pyriformis** (1/1) M: 44.6 mg/L
 - NOEC
 - Growth inhibition
 - Immobilisation
 - Intoxication
 - Mortality
 - Physiology
 - Reproduction
 - Sediment toxicity
 - Terrestrial Toxicity
 - Human health hazards
 - Acute Toxicity
 - Bioaccumulation
 - Carcinogenicity
 - Developmental toxiciv / teratogenicity

AW SW

Export Data matrix

Expand branch

Collapse branch

Expand All

Collapse All

Target endpoint

Copy Path

Activate AOP

1

2

Figure 23

-0-

QSAR Toolbox v.4

C.4.2. Area with selected chemicals

[User interface](#) / [Data matrix](#) / C.4.2. Area with selected chemicals

C.4.2. Area with selected chemicals (2) (Figure 1)

The screenshot displays the QSAR Toolbox interface. On the left, there are panels for 'Documents' and 'Inventories'. The 'Documents' panel shows a tree view of databases including 'Hydrolysis rate constant (QSAR)', 'IR database (Environment Canada)', 'Phys-chem (EPA/SITE)', 'Ecotoxicological Information', and 'Human Health Hazards'. The 'Inventories' panel shows a tree view of various regulatory programs like 'Canada DSL', 'COSLEG', 'DSSTOX', 'ECHA PR', 'EINECS', 'HPV (OECD)', 'METI Japan', 'NORMAS', 'REACH ECH', 'TSCA', and 'US HPV Challenge Program'. The main area is divided into three sections: a 'Filter endpoint tree...' panel on the left, a 'Structure' panel in the middle, and a 'Data Matrix' table on the right. The 'Structure' panel shows a tree view of chemical information categories such as 'Structure info', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Aquatic Toxicity', 'Sediment toxicity', 'Terrestrial Toxicity', and 'Human health hazards'. The 'Data Matrix' table has columns for chemical structures and rows for various endpoints. The table is as follows:

AW SV	1	2	3	4	5	6	7
(7889)	M: 152 mg/L	M: 47.6 mg/L	M: 57.3 mg/L		M: 68.8 mg/L		M: 216 mg/L
(814)							
(1212)							
(1)							
(89-105)	M: 9.79 mg/L		M: 2.16 mg/L	M: 1.01 mg/L		M: 2.71 mg/L	M: 130 mg/L
(1320)			M: 9.95 mg/L				
(512)							

Figure 1

4.2.1. Chemical information

This functionally provides information for the available constituents, additives and impurities (1). The user can see them by right click on the target structure and then select Chemical information (2). New pop-up window appears. It consists of two parts: main (3) and detailed for all components (4). The constituents, additives and impurities of the substance, if any, are located in separated tabs (5). (Figure 2)

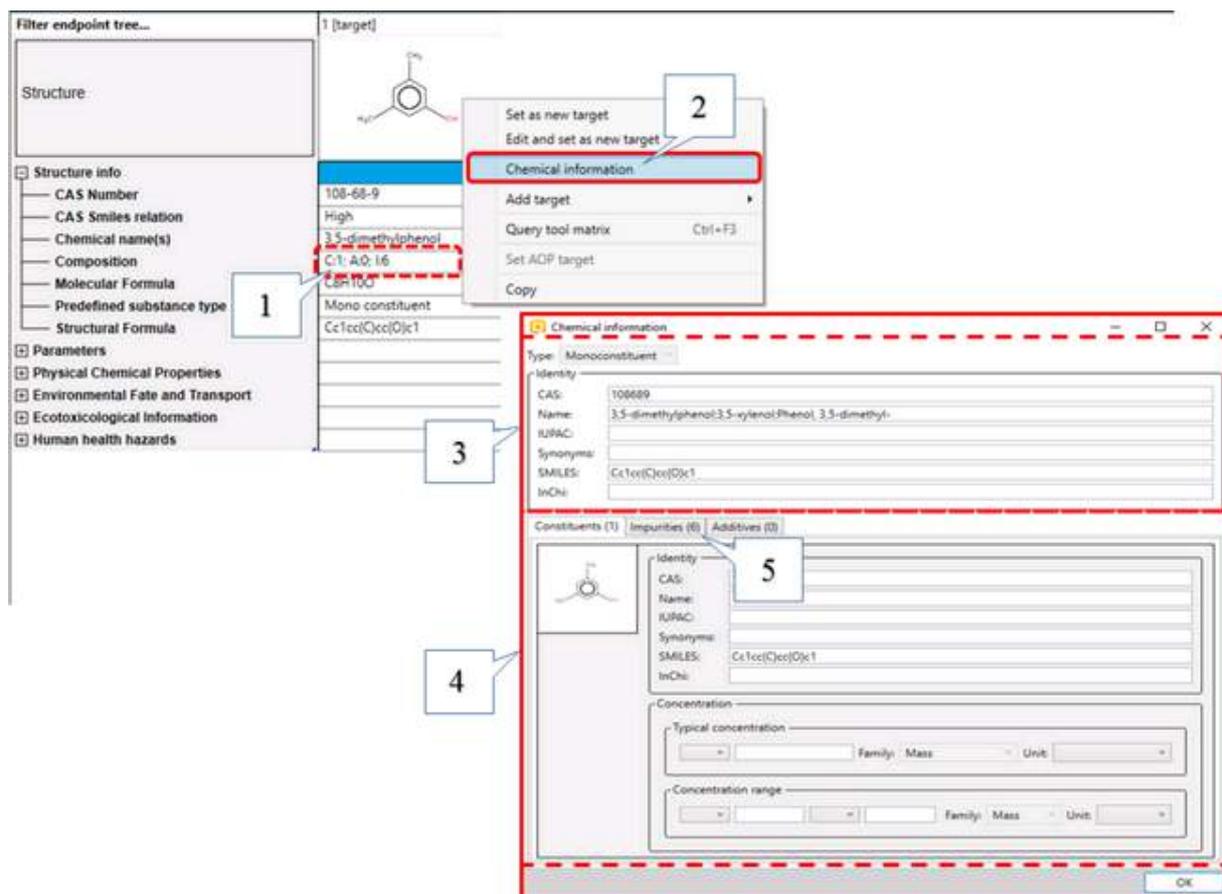


Figure 2

4.2.2. Query tool matrix

The user can search chemicals among these, which are on the data matrix. This could be done by selection of Query tool matrix (1) within the right click options. (Figure 3)

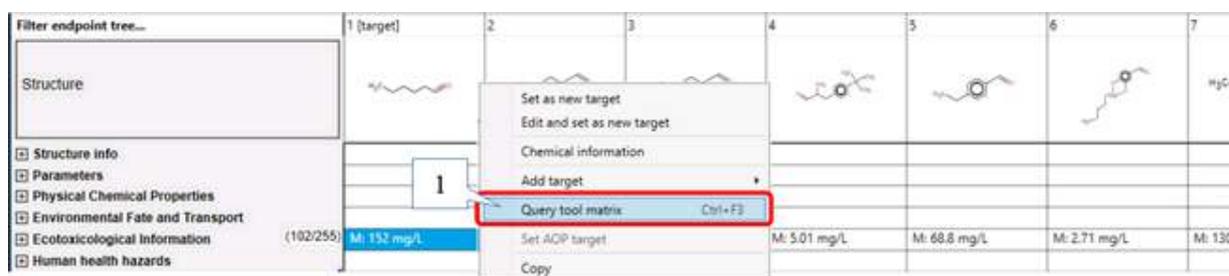


Figure 3

QSAR Toolbox v.4

C.4.3. Area with data

[User interface / Data matrix / C.4.3. Area with data](#)C.4.3. Area with data (3) (Figure 1)

Figure 1

4.3.1. Search

The Search option allows to search the CAS Registry Number or specific Name of a chemical in external sources.

- Search of CAS Number – Right click on the cell with the *CAS Number* (1) automatically invokes a pop-up menu allows the user to search the current CAS within *eChemPortal* or *Google* websites (2). (Figure 2)

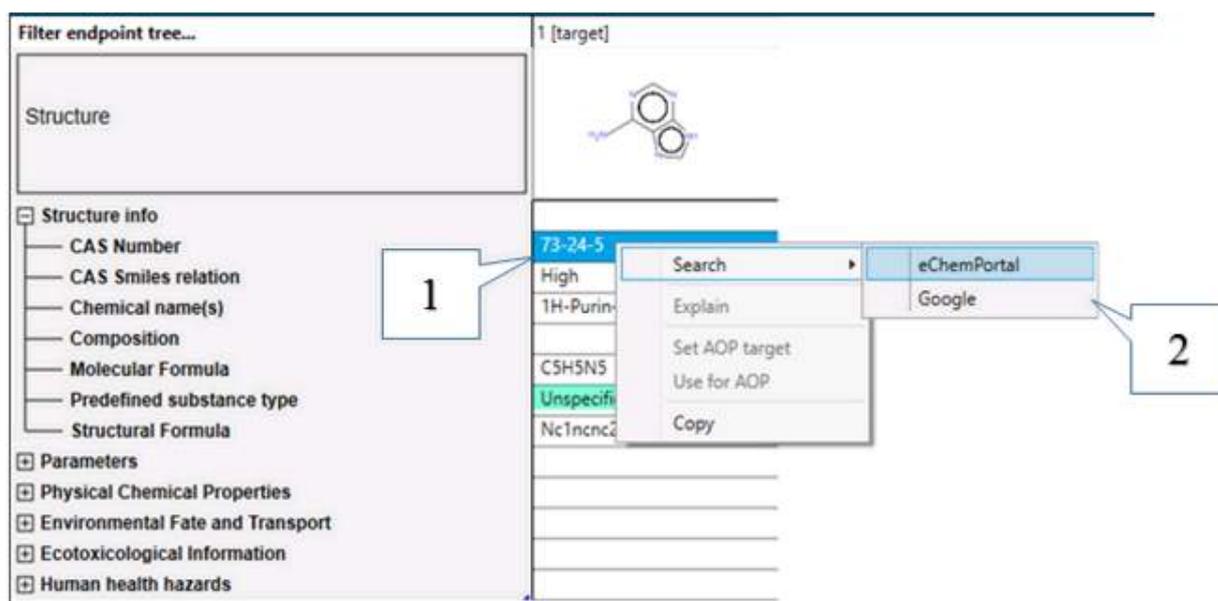


Figure 2

- Search of Chemical name(s) - Right click on the cell with *Chemical name(s)* (1) automatically invokes a pop-up menu allows the user to search one of the available names (2) within *eChemPortal*, *PubChem*, *Wikipedia* or *Google* websites (3). (Figure 3)

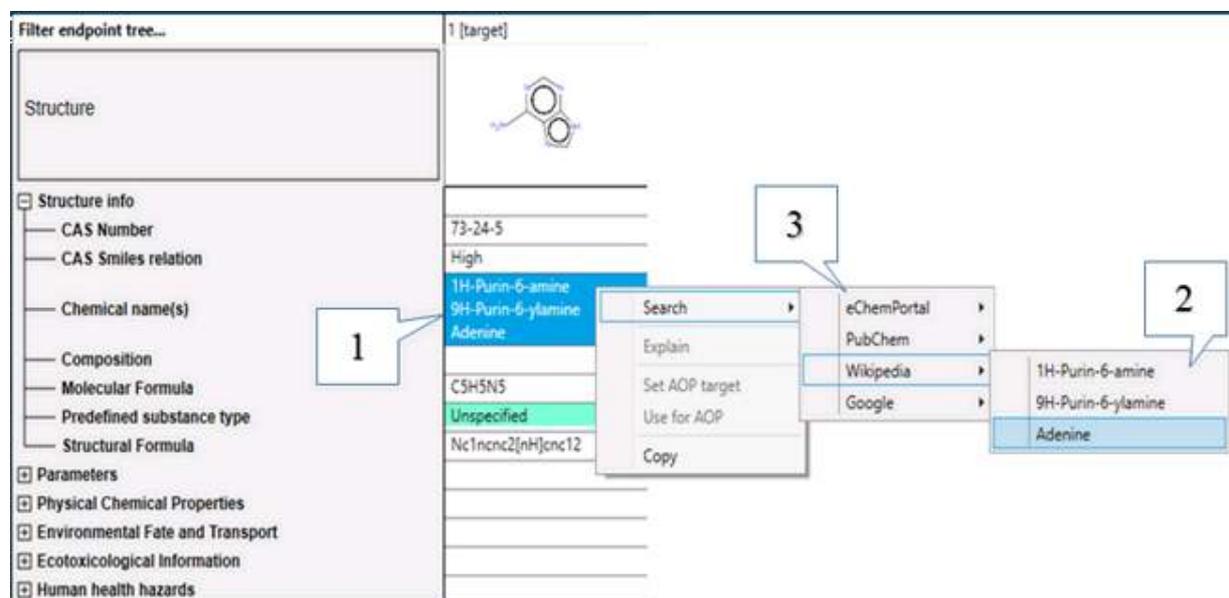


Figure 3

4.3.2. Quality assessment (QA)

The reliability of the CAS/SMILES relation is indicated for all chemicals, which

are available in Toolbox. Double click on the QA or right click and selection of Explain (1) invokes the CAS-SMILES relation window. It contains the following information: the name and type of all sources, in which the chemical present (2); quality of the source (3); information whether the SMILES in the source is assigned from other databases or not (4). (Figure 4)

The screenshot shows the 'CAS-SMILES relation' window with the following table:

Exist in data source	Data source type	Data source quality	Assigned SMILES in data source
Acute Oral toxicity	Database	Distribute to QA	no
Aquatic ECETOX	Database	Distribute to QA	yes
Aquatic OASIS	Database	Distribute to QA	no
Canada DSL	Inventory	Distribute to QA	no
DSSTOX	Inventory	High quality source	no
ECHA CHEM	Database	Distribute to QA	no
ECHA PR	Inventory	Distribute to QA	yes
ECOTOX	Database	Distribute to QA	yes
EINECS	Inventory	High quality source	no
Experimental pKa	Database	Distribute to QA	no
Genotoxicity OASIS	Database	Distribute to QA	no
HPVC OECD	Inventory	Distribute to QA	no
METI Japan	Inventory	Distribute to QA	no
NICNAS	Inventory	Distribute to QA	no
Phys-chem EPISUITE	Database	Distribute to QA	no
REACH ECB	Inventory	High quality source	no
Rep Dose Tox Fraunhofer ITEM	Database	Distribute to QA	no
ToxCastDB	Database	Distribute to QA	no
TSCA	Inventory	Distribute to QA	yes
US HPV Challenge Program	Inventory	Distribute to QA	no

Figure 4

4.3.3. Labels with data

There are some labels for data visualized on data matrix:

- 1) "M" – means measured data, extracted from databases
- 2) "R" – means prediction result obtained as a result of read-across analysis
- 3) "T" – means prediction result obtained as a result of trend analysis
- 4) "Q" – means prediction result obtained as a result of QSAR prediction
- 5) "IMOA" – means prediction result obtained from component based Independent MOA

- 6) "ICOA" – means prediction result obtained from component based Similar MOA

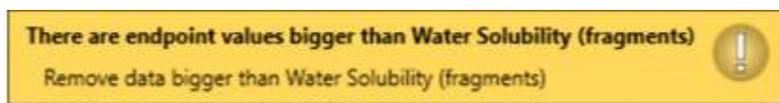
-0-

C.5. Information and active accessories

1. Helpers

New functionality intended to help the user with specific information associated with analogues used in gap filling approach have been implemented. The helpers provide different type of information. Some examples are represented below:

- Providing some specific information for the analogues (e.g. LC50 values of some analogues are bigger than WS of the chemicals). The system compares data values, which are in the volume concentration unit family, against their calculated maximum water solubility in order to detect artificial data points which could be removed.



- warning message, alerting the user:
 - there are analogue with data including qualifiers



- there are analogue with composition (e.g. mixtures, UVCBs, etc.)



- notification message, e.g.:
 - the trend analysis prediction obtained by the current state is acceptable according to the statistics



- the read-across prediction is obtained by more analogues than the default 5. This could be due to chemicals with equal descriptor values (logKow values)



Once the Helpers appear in gap filling stage, click once over the helper to see the message. Click on the helper and slide it to remove it.

2. Display active and background working actions status

Functionality used to follow the progress of the active action or to cancel an active action when an [Automated workflow](#) is run. (Figure 1)

- To see a progress of the active action click on the clock (1)
- To cancel an active action click on Stop button (2)

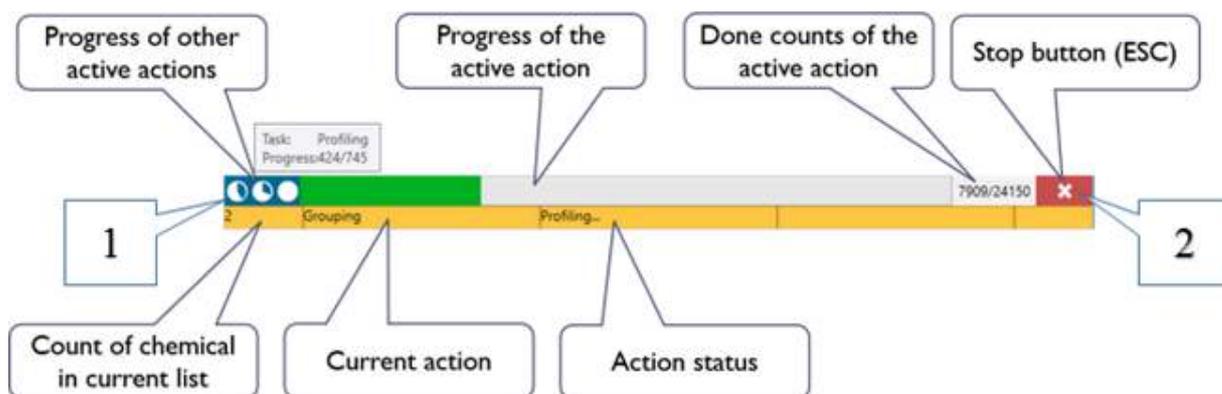


Figure 1

-0-

Chapter D. Workflow

D.1. Input

D.1.1. Single chemical

QSAR Toolbox v.4

D.1.1.1. Entering a chemical by CAS

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.1. Entering a chemical by CAS

D.1.1.1. Entering a chemical by CAS registry number

To enter a chemical by its CAS (Chemical Abstract Service) number simply press the button *CAS #* (1), enter the CAS number of the chemical without hyphens (2), press the button *Search* (3) and once the correct structure appears press button *OK* (4) (Figure 1).

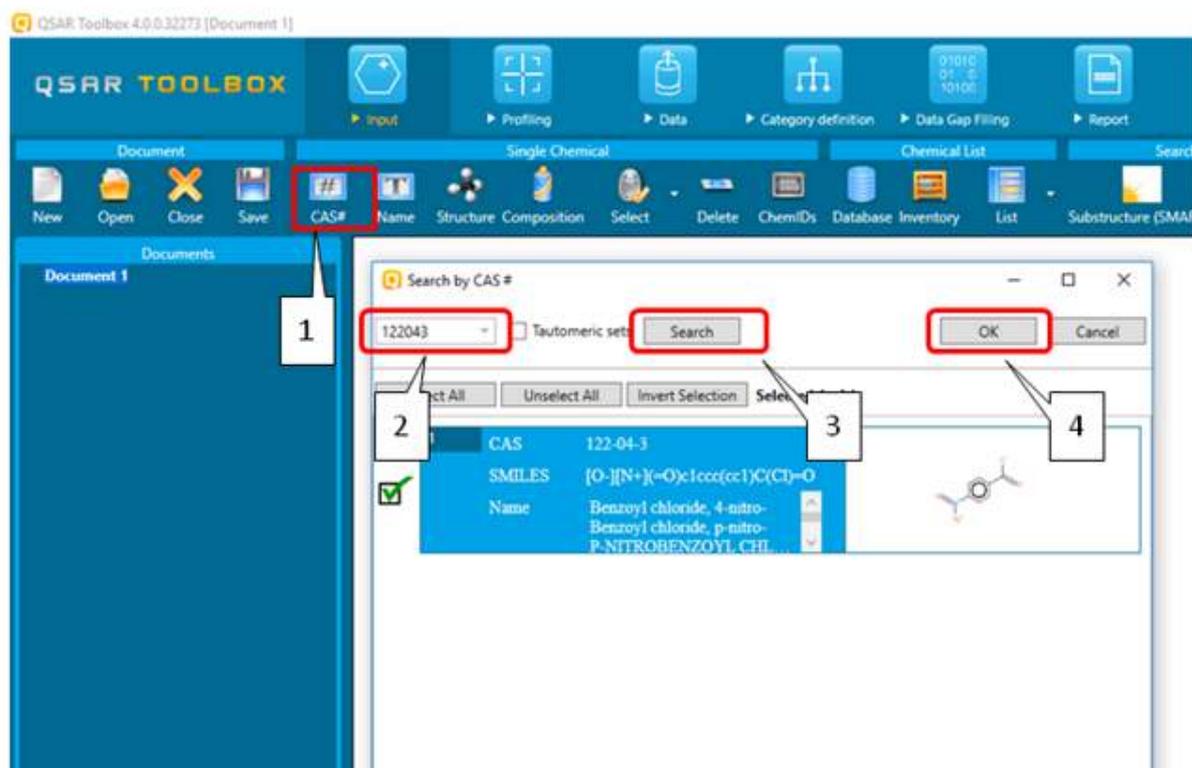


Figure 1

In cases when the CAS # could be related to more than one substance, more than one chemical identity could be retrieved. In these cases the user has to decide which substance to be retained for the subsequent workflow.

Additionally if the CAS # is present in two databases, one of which is ECHA Chem database, two chemicals will appear in the data matrix. Substance type is also taken in account, i.e. if the chemical has different substance type affiliation (applicable to ECHA Chem database only), then each substance type will be considered as an individual chemical and all will be displayed .

-0-

QSAR Toolbox v.4

D.1.1.2. Entering a chemical by chemical name

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.2. Entering a chemical by chemical name

D.1.1.2. Entering a chemical by chemical name

To enter a chemical via its name the user has to select the button *Chemical Name* (1), write the name of the target chemical (or part of it) in the field *Search for name* (2), and select one of the search options: *Exact match*; *Starting with*; *Containing* (3) and press the button *Search* (4) (Figure 1).

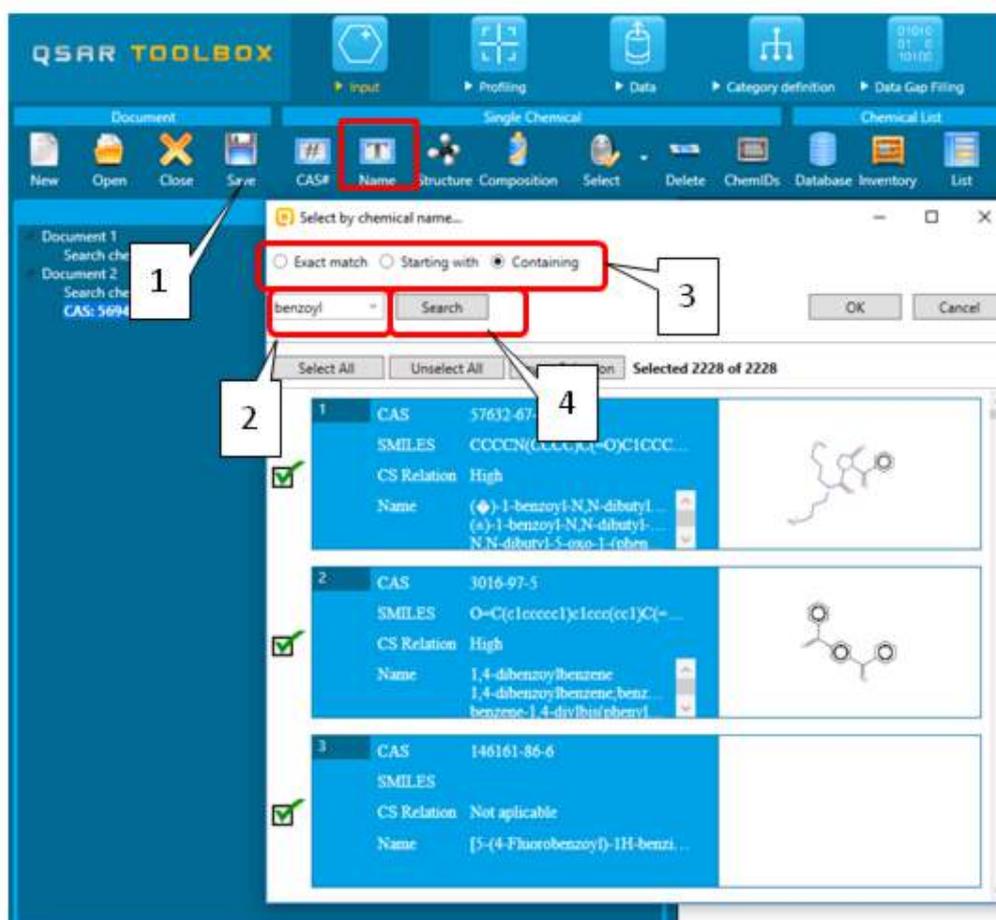


Figure 1

A list of chemical names ordered alphabetically will appear; each name is accompanied by its 2D structure. Select the appropriate chemical from the list of names by clicking over the row with the desired chemical (5) or using the

buttons: Select all; Unselect all; Invert selection (6) and then and then press the button OK (7) (Figure 2).

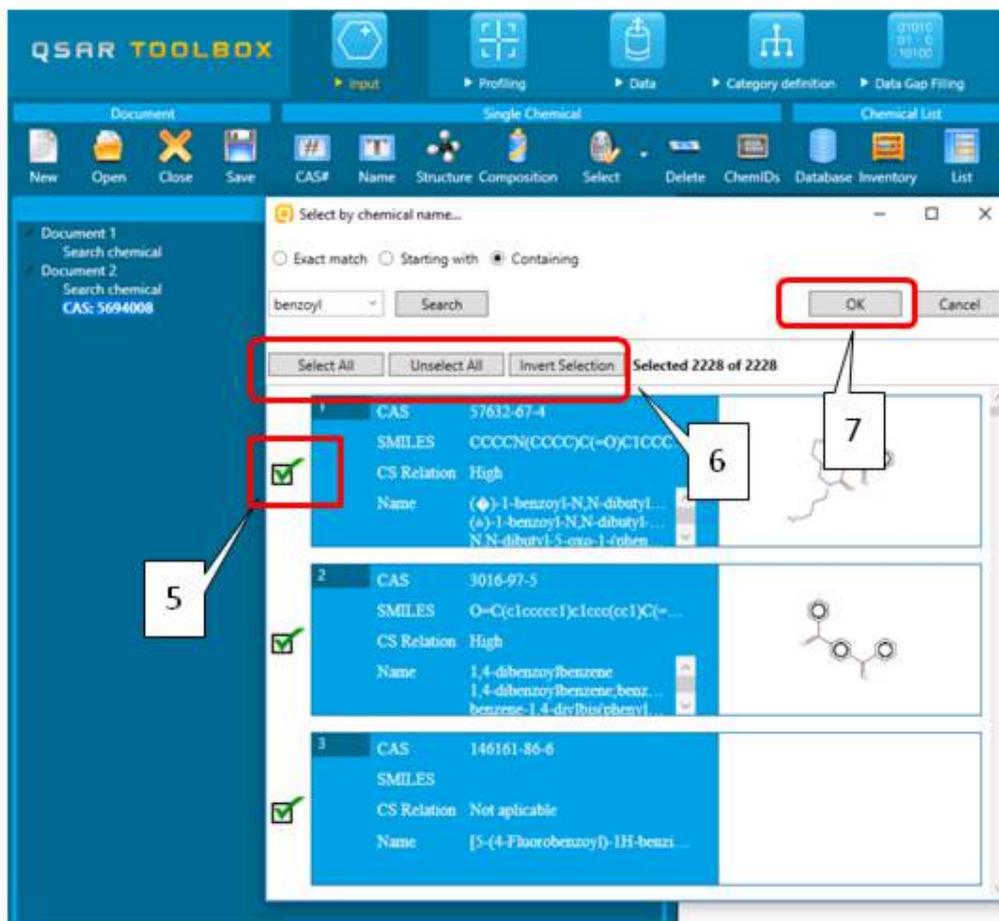


Figure 2

This mode allows entering of chemical identity using "wild cards" search. A maximum of 192 characters is allowed to be used in the field Search for Name (3) (Figure 1).

QSAR Toolbox v.4

D.1.1.3. Entering a chemical by structure

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.3. Entering a chemical by structure

D.1.1.3. Entering a chemical by structure

Three types of chemical structures are possible to be entered by drawing:

- [Single structure \(discrete chemical\)](#)
- [Single structure \(mixture\)](#)

-0-

QSAR Toolbox v.4

D.1.1.3.1. Add single chemical structure (discrete chemical)

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / [Entering a chemical by structure](#)/ D.1.1.3.1. Add single chemical structure (discrete chemical)

D.1.1.3.1. Add single chemical structure (discrete chemical)

Atom connectivity could be defined by drawing of chemical 2D structure. Select the button Structure (1) and then to draw by hand the structure of the target chemical in the Structure Drawing window (3). The corresponding SMILES will be automatically generated with the progress of the drawing. A complete drawing must be confirmed via the button OK (4) (Figure 1).

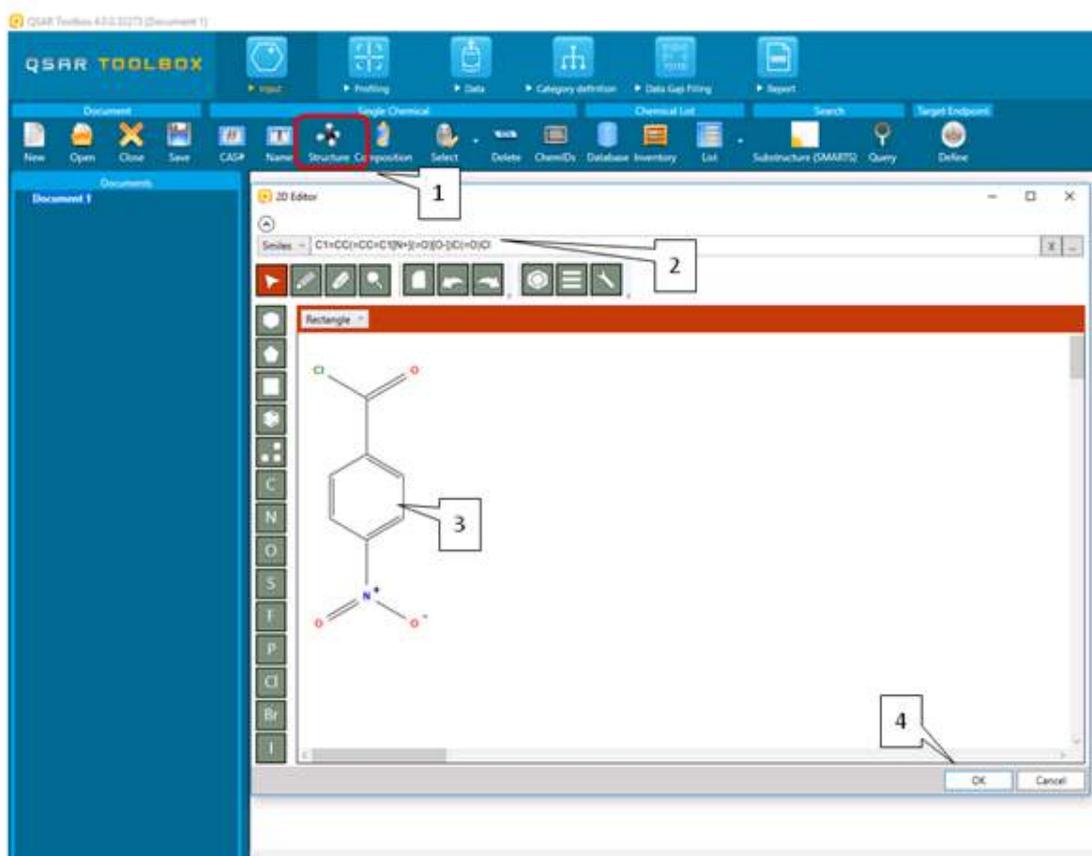


Figure 1

Another way of entering the connectivity is by chemical's SMILES (Simplified Molecular Input Line Entry System). Enter (write or paste) the corresponding code in the field SMILES (2) (Figure 6). If the atom connectivity is coded correctly, the corresponding 2D structure, CAS registry number and chemical name will appear.

In case of incorrect entry code, the incorrect entry will be colored in red in field SMILES(2) will be colored in red and the structure will not be displayed. Short explanation text appears under the SMILES field (3) (Figure 2).

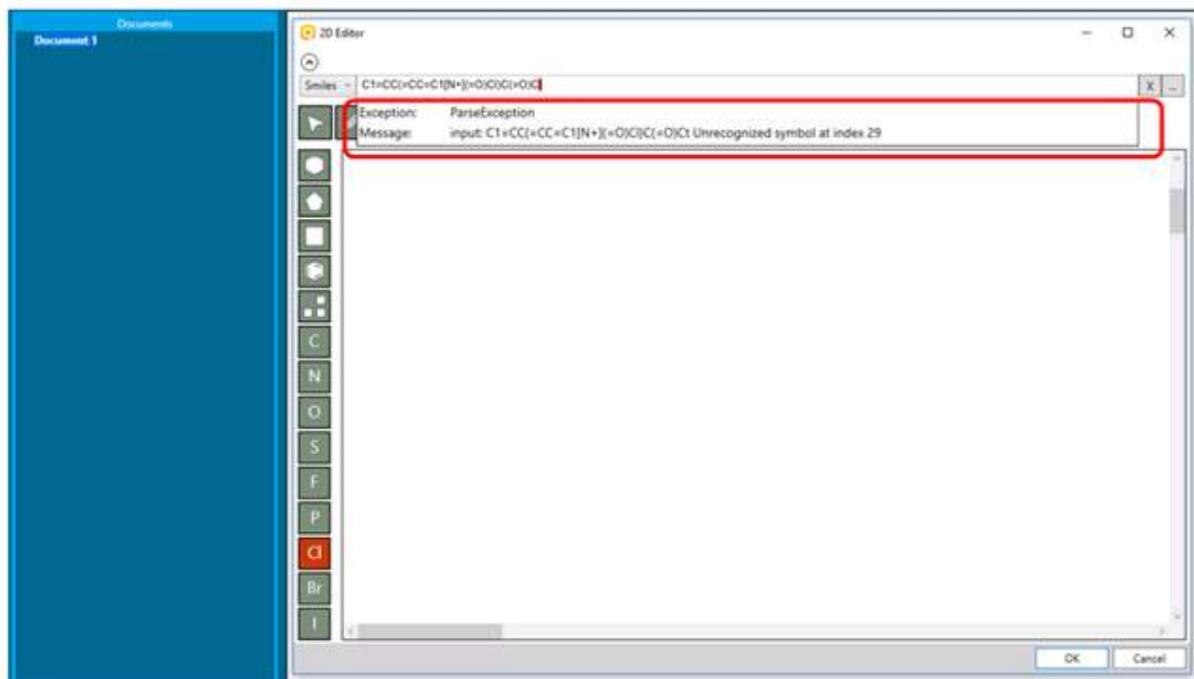


Figure 2

After the drawing the system will search the databases and inventories for the entered SMILES. The identified SMILES code(s) appears, then the user has to select the appropriate chemical in case of more than one SMILES is identified (as in the case of benzene, Figure 8). Finally click OK (Figure 3).

Figure 3

-0-

QSAR Toolbox v.4

D.1.1.3.2. Add single chemical structure (mixture)

Workflow / [Chemicals input](#) / [Single chemical](#) / [Entering a chemical by structure](#) / D.1.1.3.2. Add single chemical structure (mixture)

D.2.1.3.2. Add single chemical structure (mixture)

Composition

An option is provided to include the composition information of the chemical. To do that first select Composition (1) and then select the Type of chemical from the drop-down menu (2). (Figure 1)

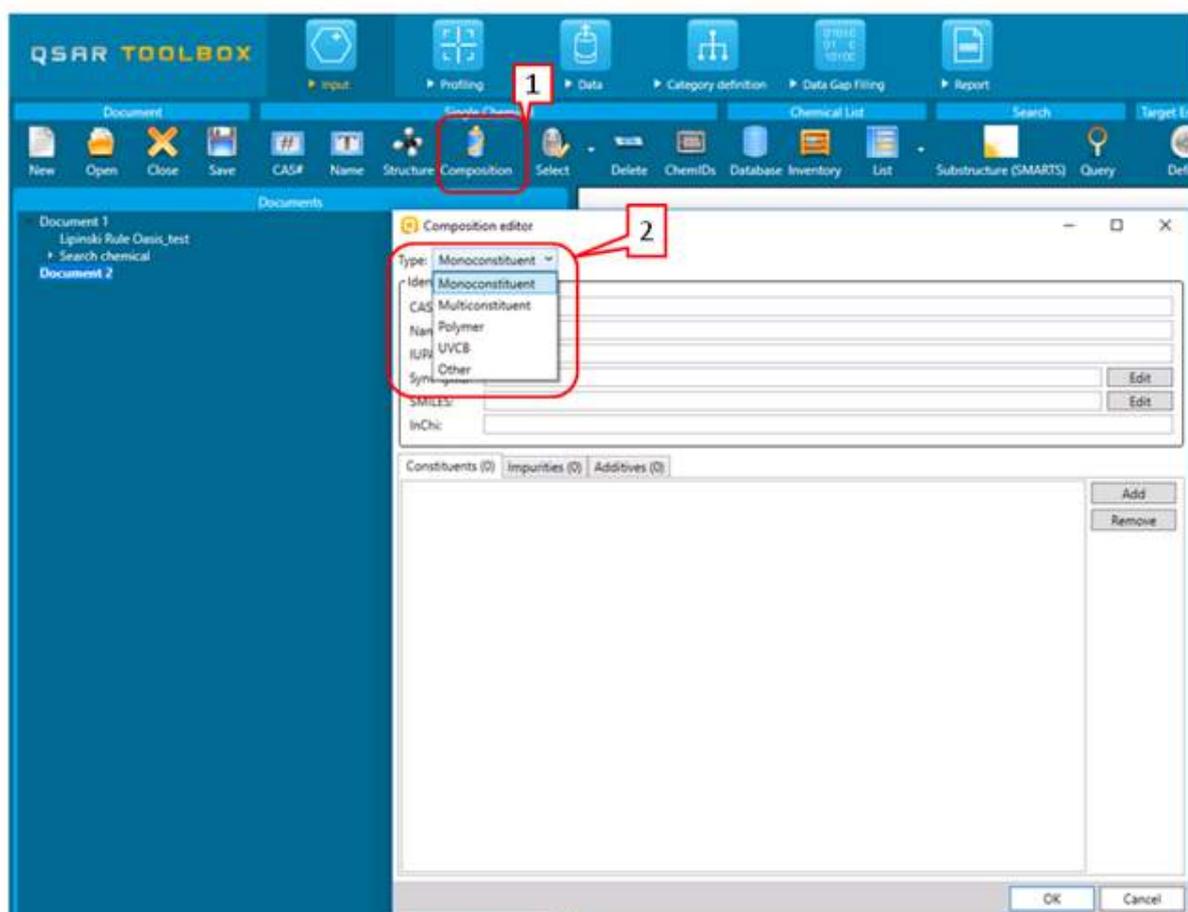


Figure 1

For example, for mixtures, select Multiconstituent, Constituent tab (2) and then Add (3) (Figure 2). Pressing Edit (4) opens 2D Editor where the user can draw the structure or paste its SMILES (5). For more information about 2D Editor please see Section [D 1.3.1 Substructure \(SMARTS\)-2D editor](#). For each component a new dialogue has to be open by pressing Add.

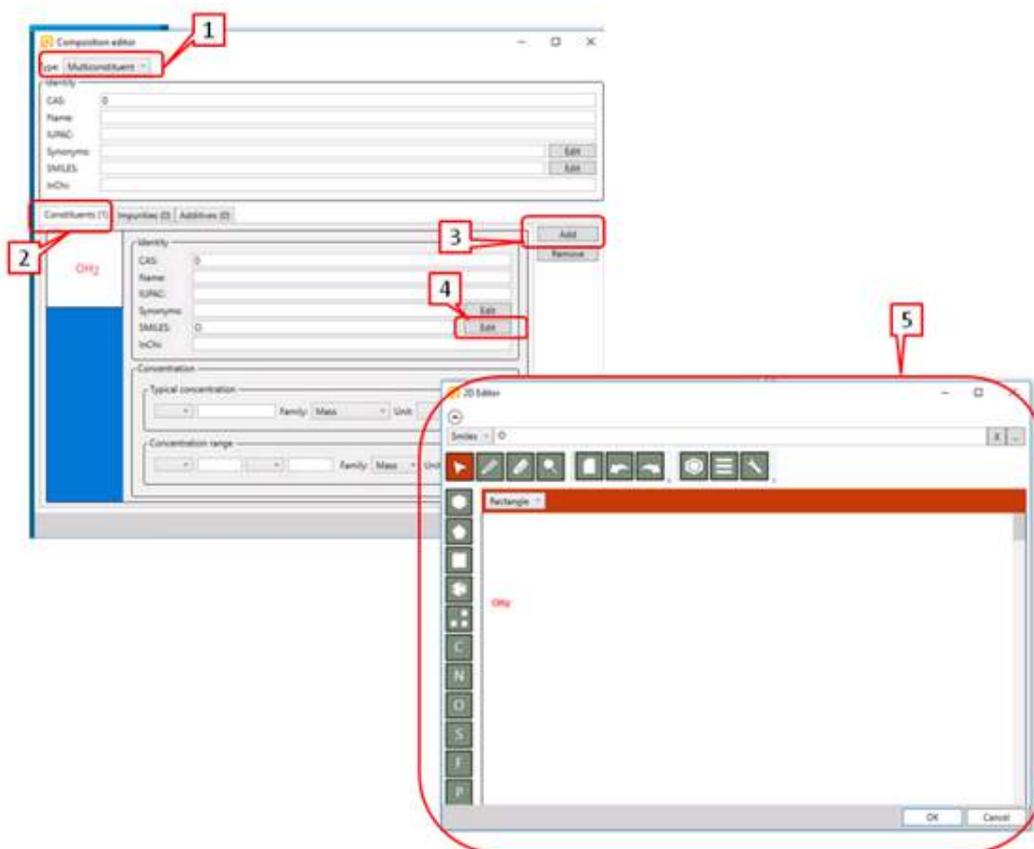


Figure 2

The number of constituents is shown in (1) (Figure 3). Quantitative and qualitative information about additives and impurities can be added (2). The entries can be deleted by pressing Remove. In the identity section (4) there are fields such as CAS RN, Name, which can be filled in by the user.

The screenshot displays the 'Identity' and 'Concentration' sections of the OECD QSAR Toolbox application. The 'Identity' section at the top contains fields for CAS (0), Name, IUPAC, Synonyms, SMILES, and InChi, with 'Edit' buttons for Synonyms and SMILES. Below this are buttons for 'Constituents (3)', 'Impurities (0)', and 'Additives (0)'. A red callout '1' points to the 'Constituents (3)' button. A red callout '2' points to the 'Additives (0)' button. A red callout '3' points to the 'Add' and 'Remove' buttons on the right. A red callout '4' points to the 'Identity' section of a smaller, highlighted window below. This smaller window contains fields for CAS, Name, IUPAC, Synonyms, SMILES (with the value CC(=O)c1ccc(C)c(C)c1Cl), and InChi, with 'Edit' buttons for Synonyms and SMILES. Below the smaller window is the 'Concentration' section, which includes 'Typical concentration' and 'Concentration range' sub-sections, each with a dropdown menu, a text input field, a 'Family' dropdown (set to 'Mass'), and a 'Unit' dropdown. At the bottom right are 'OK' and 'Cancel' buttons.

Figure 3

In the concentration section, there are several options to express the quantity of each component (1,2) (Figure 4):

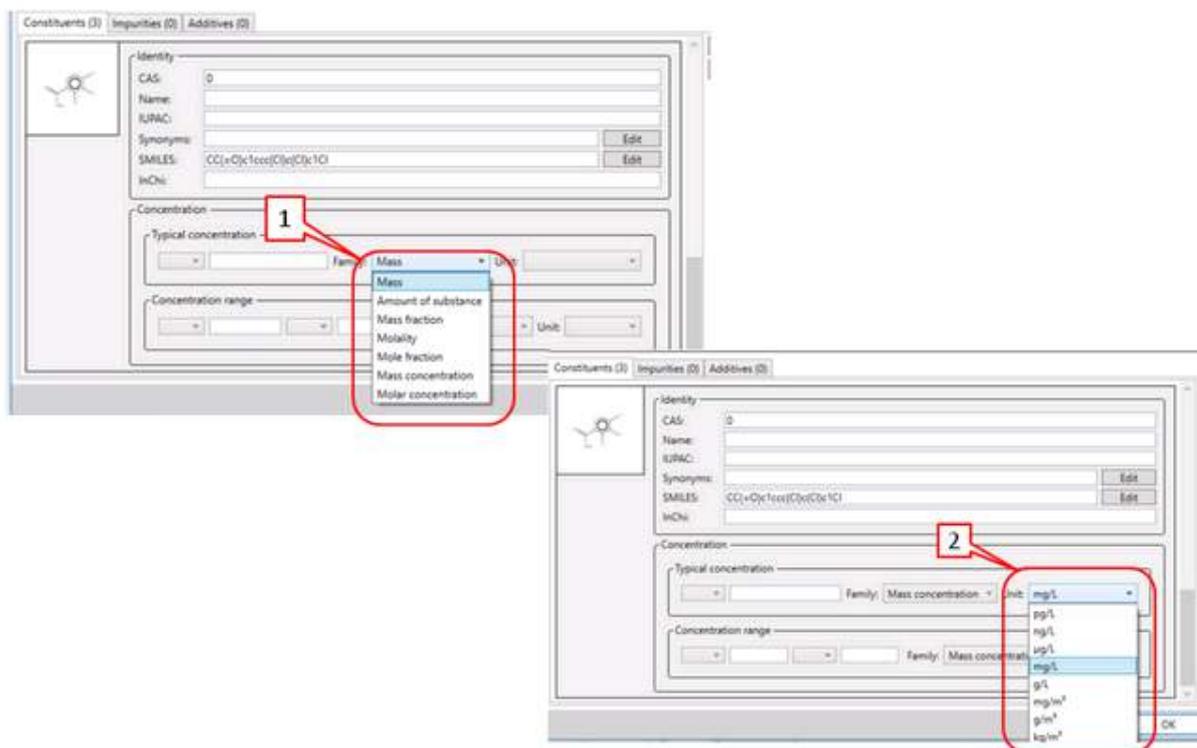


Figure 4

Select from the drop-down menu the qualifier and then the type in the digits (1).

Mass units in both sections (typical concentration and concentration range) has to be the same, even if you do not want to include concentration range (i.e. leave the cells empty) (2) (Figure 5).

Constituents (3) Impurities (0) Additives (0)

Identity

CAS: 0

Name:

IUPAC:

Synonyms: Edit

SMILES: CC(=O)c1ccc(Cl)c(Cl)c1Cl Edit

InChi:

Concentration

Typical concentration

= 100 Family: Mass concentration Unit: mg/L

Concentration range

Family: Mass concentration Unit: mg/L

OK

Figure 5

When all quantities are added press OK. (Figure 6)

Constituents (3) Impurities (0) Additives (0)

Identity

CAS: 0

Name:

IUPAC:

Synonyms: Edit

SMILES: CCCCO Edit

InChi:

Concentration

Typical concentration

= 1 Family: Mass concentration Unit: mg/L

Concentration range

Family: Mass concentration Unit: mg/L

Add

Remove

OK Cancel

Figure 6

The multi-constituent chemical will be displayed in the data matrix (1) (Figure 7).

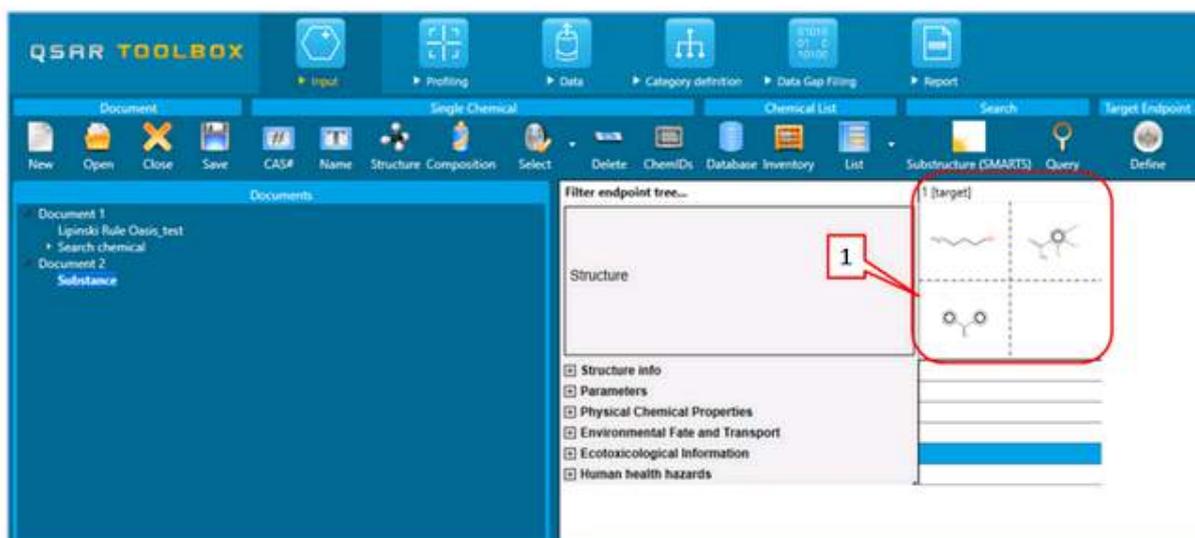


Figure 7

-0-

D.1.1.4. Entering a chemical from the last used chemicals

D.1.1.4. Entering a chemical from the last used chemicals

By pressing the drop-down arrow, you are given an option to select a chemical from (Figure 1) :

- last used chemicals
- database
- inventory
- file



Figure 1

- *Select from last used chemicals*

When you want to select a chemical from *the last used chemicals (F7 shortcut)*, the last 100 chemicals (or less) used will be listed (Figure 2):

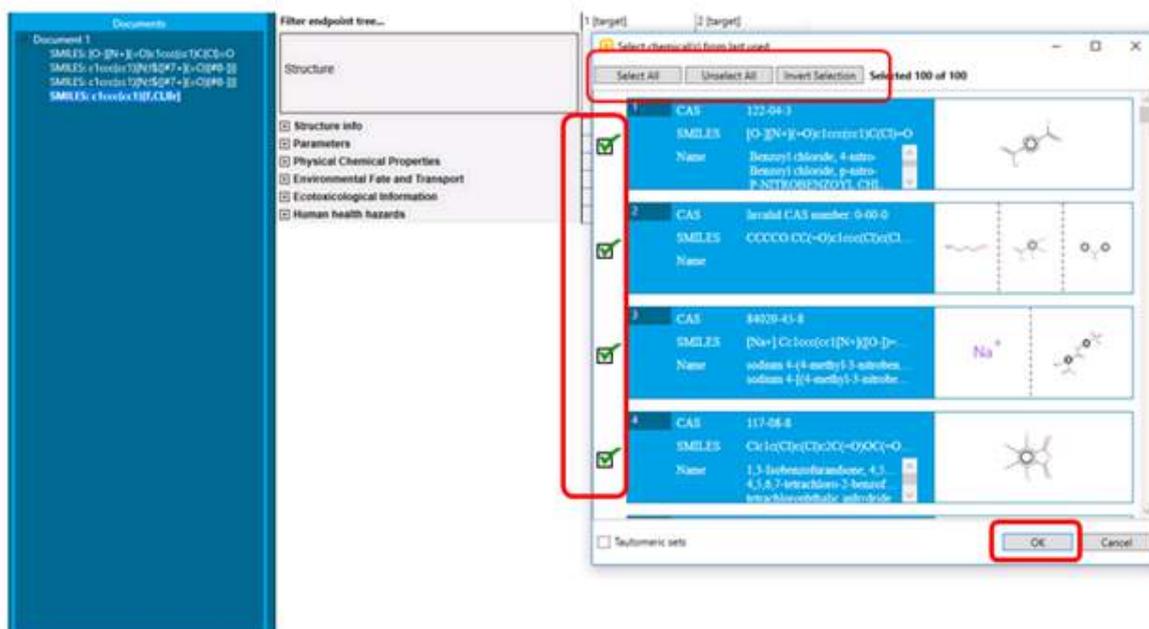


Figure 2

The user can:

- select all chemicals
- unselect all and then select only that (those) chemical(s) that are needed
- invert its selection or
- check/uncheck the chemicals of interest.

Then press OK.

-0-

QSAR Toolbox v.4

D.1.1.5. Entering a chemical from database

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.5. Entering a chemical from database

D.1.1.5. Entering a chemical from database

By pressing the drop-down arrow, you are given an option to select a chemical from (Figure 1) :

- last used chemicals
- database
- inventory
- file



Figure 1

- *Select from database*

Select from database (F8 shortcut) option, allows to select a chemical from all databases included in Toolbox 4.0/ They are grouped based on hazard effects (1). Within the group, they can be sorted by name (1) (Figure 2). When a particular database is selected (e.g. Experimental pKa)(2), an "About" button

is visualized, which when selected gives a short description of the database. Finally, press OK button (4).

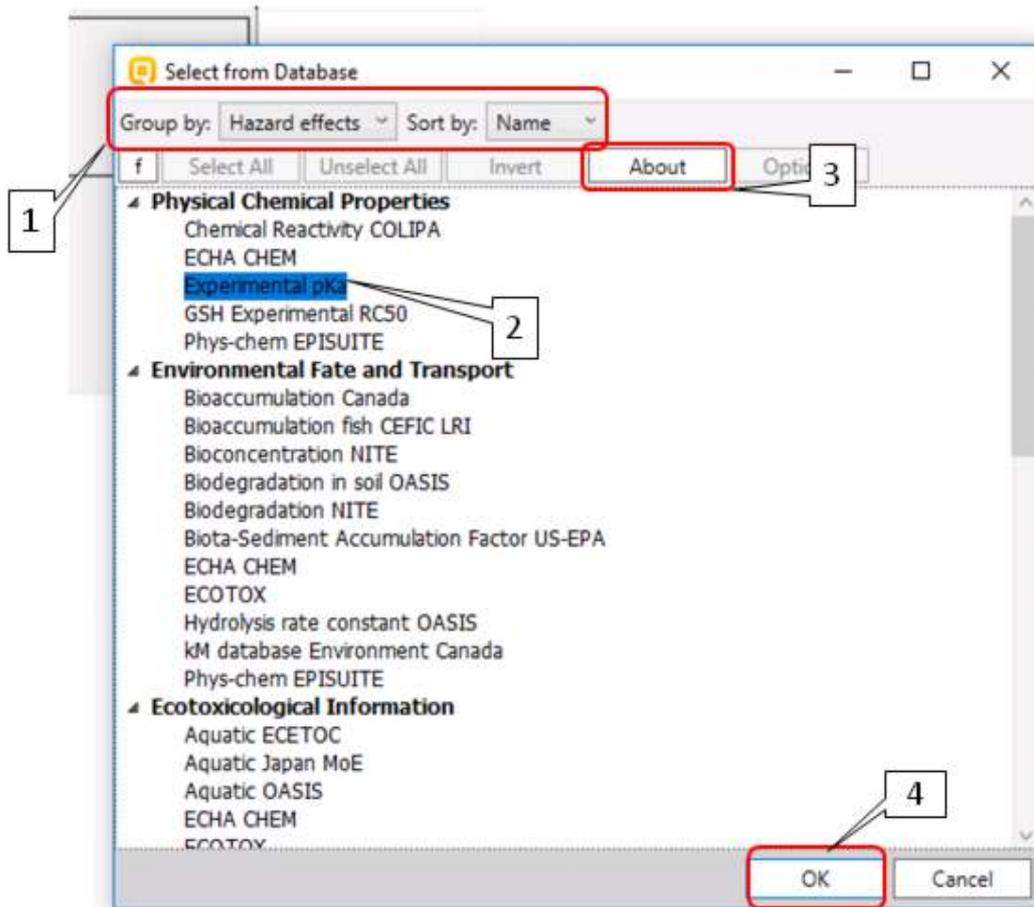


Figure 2

In addition, a search button (Figure 3) (1) is implemented, which allows a quick search to be performed. A field (2) is visualized, where you can write the name or part of the name of a database of interest.

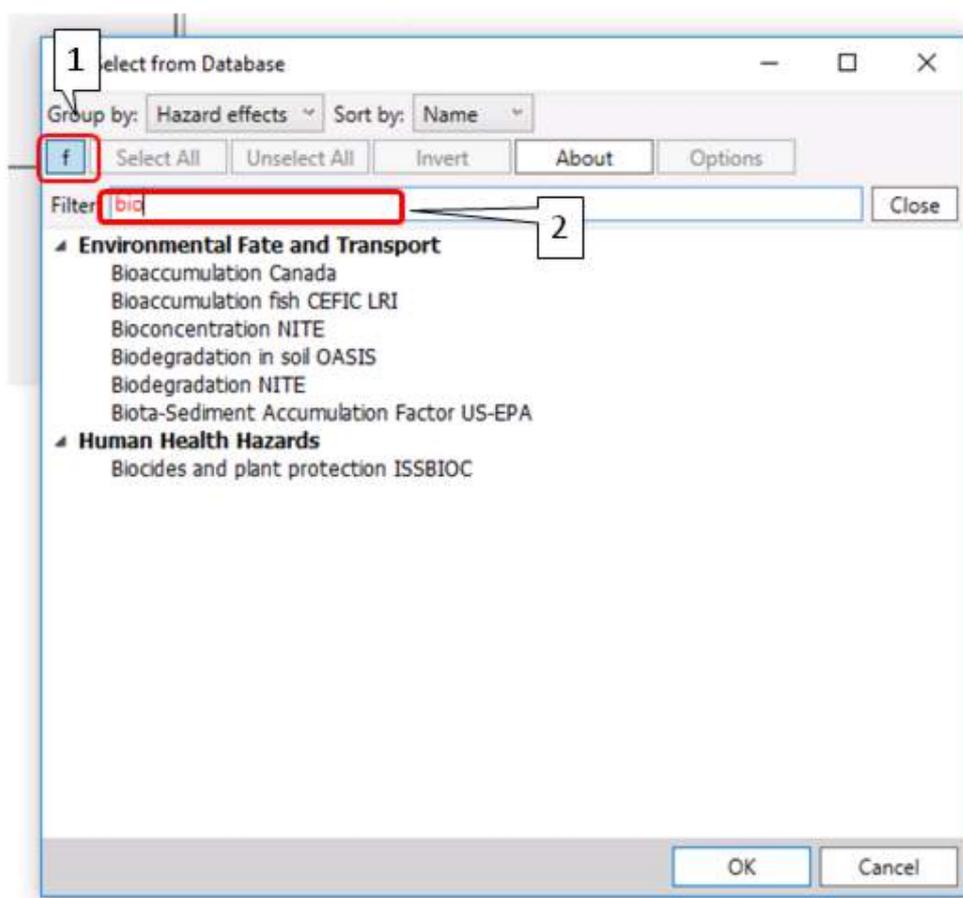


Figure 3

Subsequently a window representing all chemicals from the selected database will appear. The user needs to select the chemical intended to be the target (1) and to confirm the selection via the button OK (2) (Figure 4).

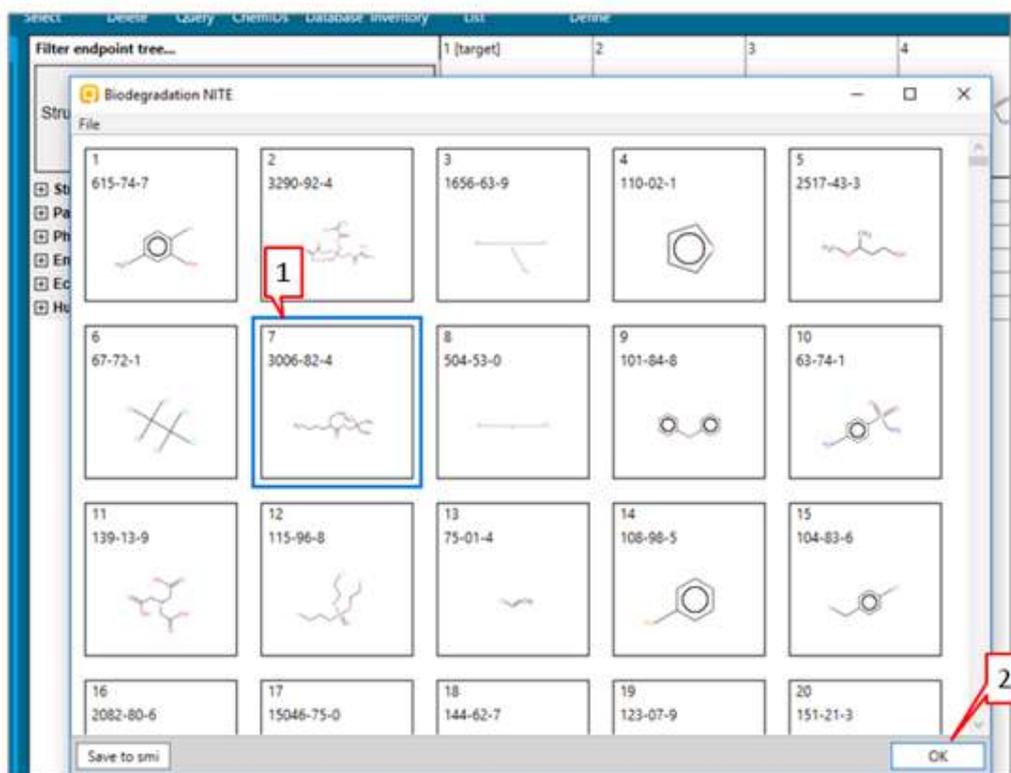


Figure 4

QSAR Toolbox v.4

D.1.1.6. Entering a chemical from inventory list

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.6. Entering a chemical from inventory list

D.1.1.6. Entering a chemical from inventory

By pressing the drop-down arrow, you are given an option to select a chemical from (Figure 1) :

- last used chemicals
- database
- inventory
- file



Figure 1
Select from inventory

Select the target chemical from a list of inventories. A list containing available inventories will appear (Figure 2). The existing inventories can be sorted by name.

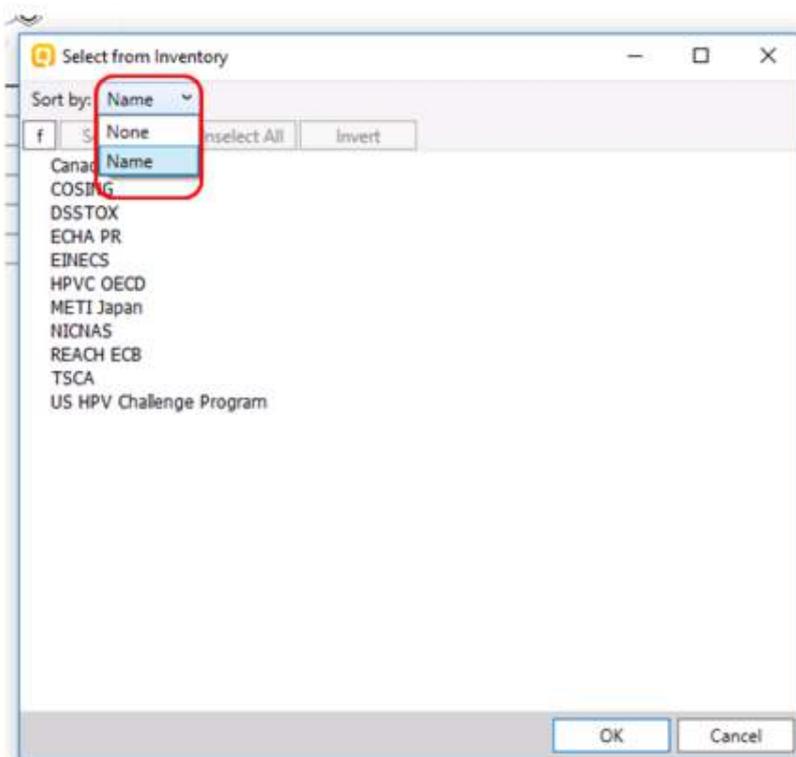


Figure 2

A search button (Figure 3) (1) is implemented, which allows a quick search to be performed. A field (2) is visualized, where you can write the name or part of the name of the inventory of interest. Once the inventory is selected (3), an *About* button is visualized, which gives information about the inventory. Finally press *OK* button to display the inventory.

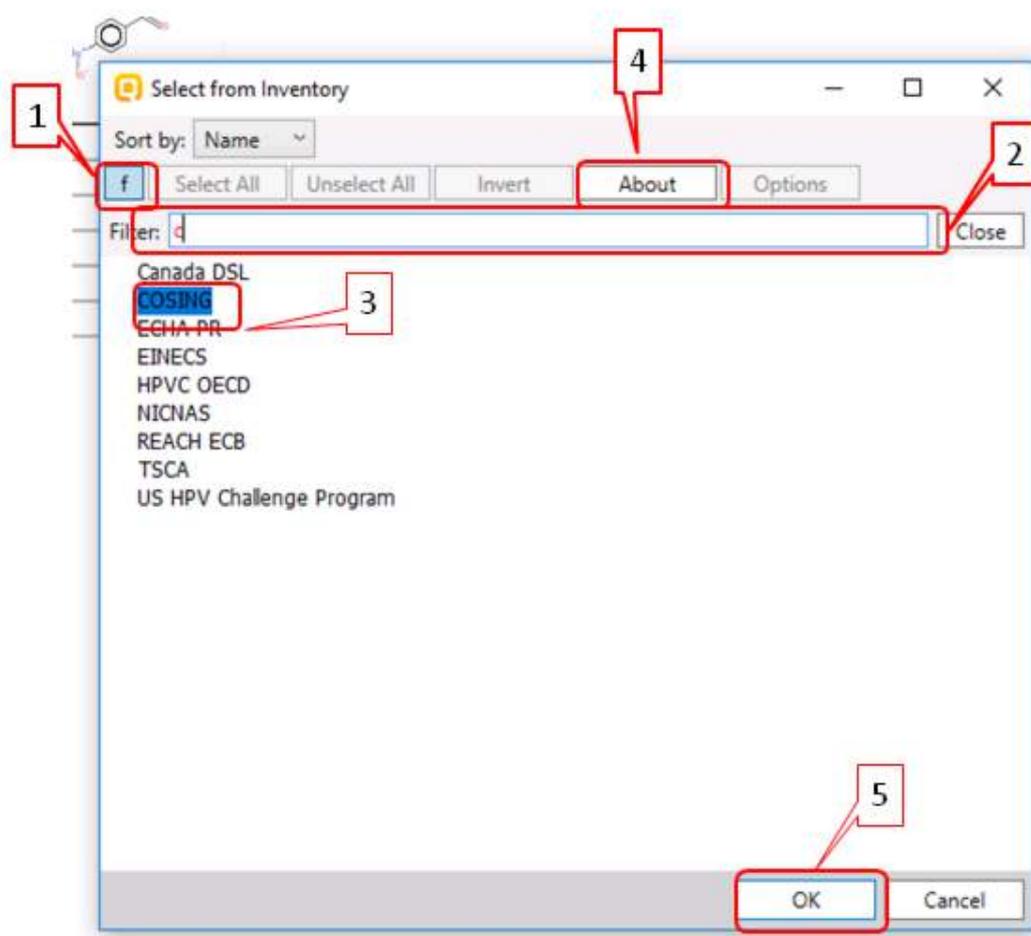


Figure 3

Subsequently a window representing all chemicals from the selected inventory will appear. The user needs to select the chemical intended to be the target (1) and to confirm the selection by pressing button OK (2) (Figure 4).

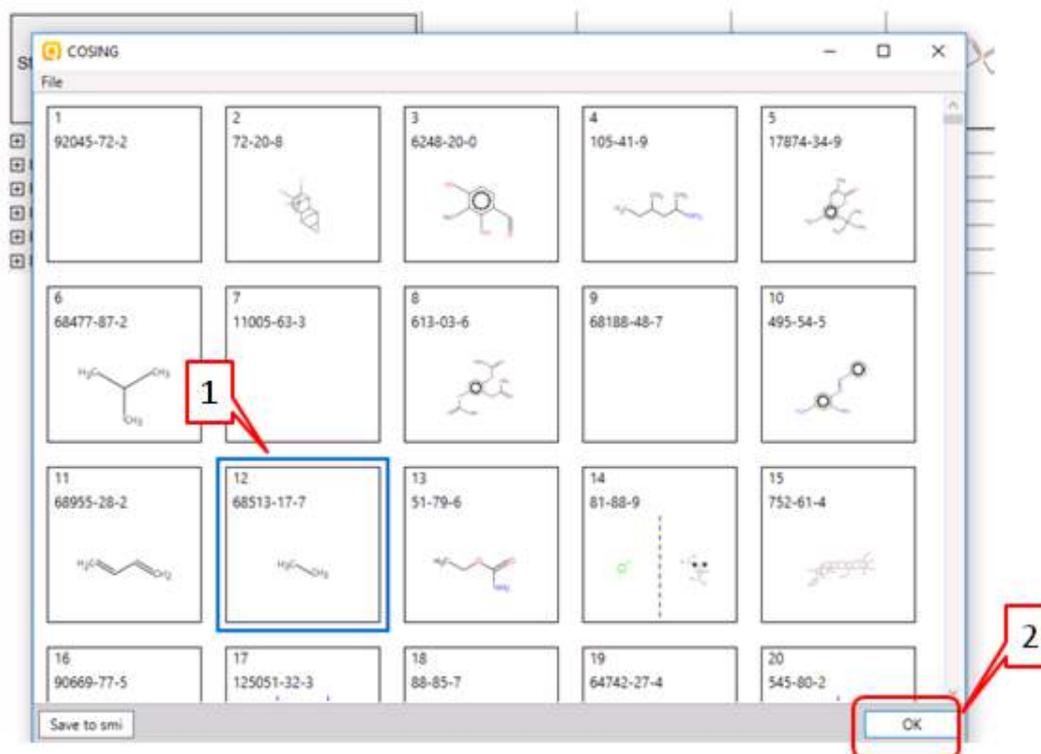


Figure 4

-0-

QSAR Toolbox v.4

D.1.1.7. Entering a chemical from file

[Workflow](#) / [Chemicals input](#) / [Single chemical](#) / D.1.1.7. Entering a chemical from file

D.1.1.7. Entering a chemical from file

By pressing the drop-down arrow, you are given an option to select a chemical from (Figure 1) :

- last used chemicals
- database
- inventory
- file

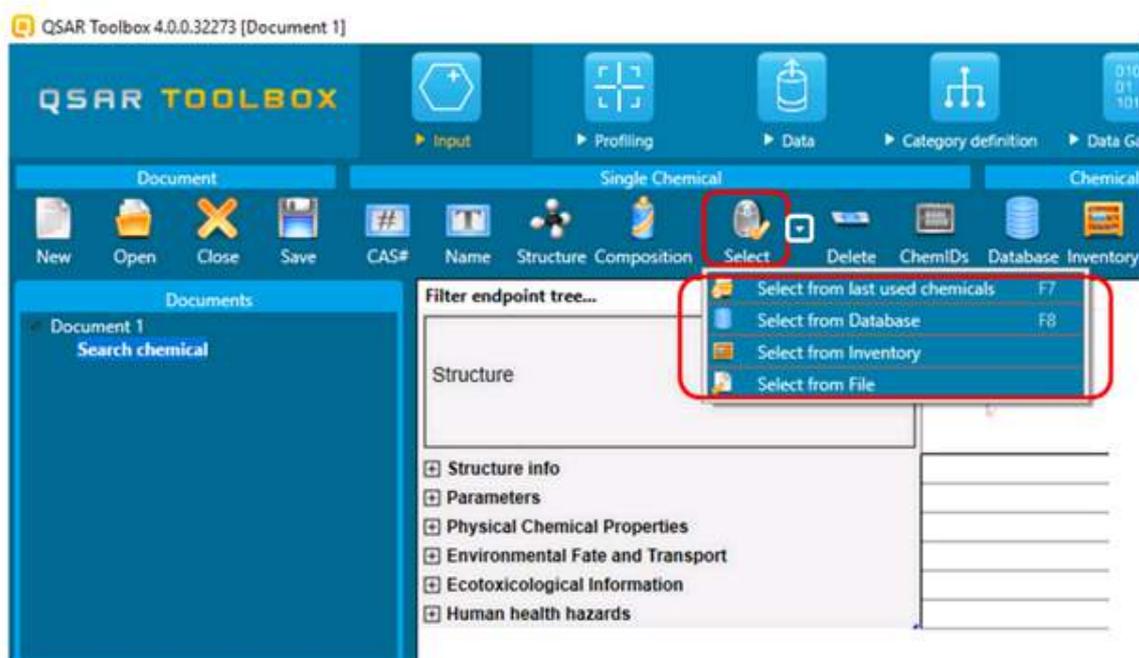


Figure 1

- *Select from file*

The target chemical can be selected from a list of already specified atom connectivity stored in a file. Currently supported by the system atom connectivity file formats are SDF and SMI. Files with SMILES may contain TAB delimited CAS number and chemical name followed by the corresponding SMILES string. The extension of these files is *.smi.

Select the pathway (1), file format (2), and the name of the file (3) (Figure 2).

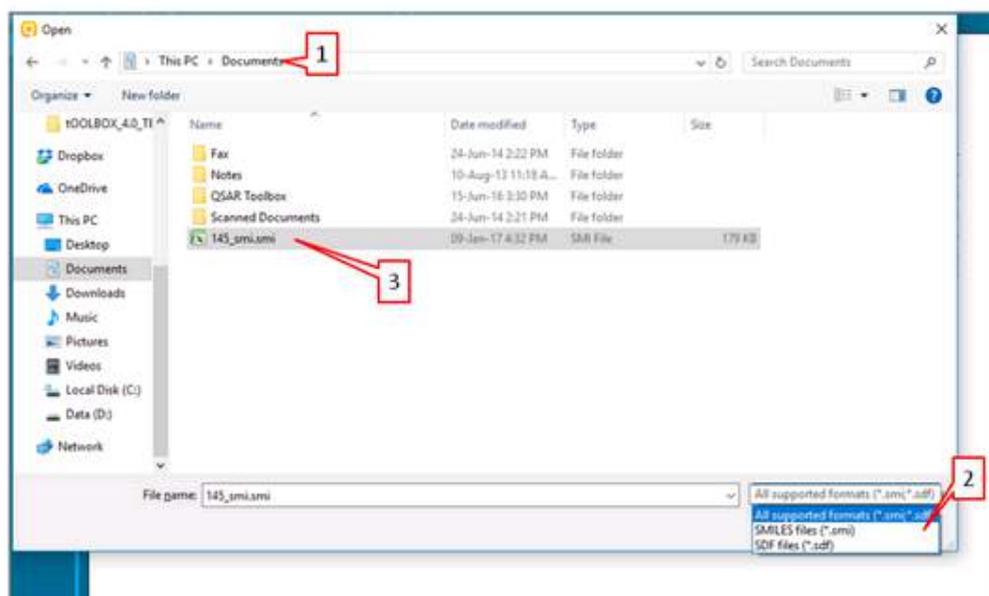


Figure 2

Subsequently a window representing all chemicals from the selected file will appear (Figure 3). Select the chemical intended to be the target (1) and confirm the selection via the OK button (2).

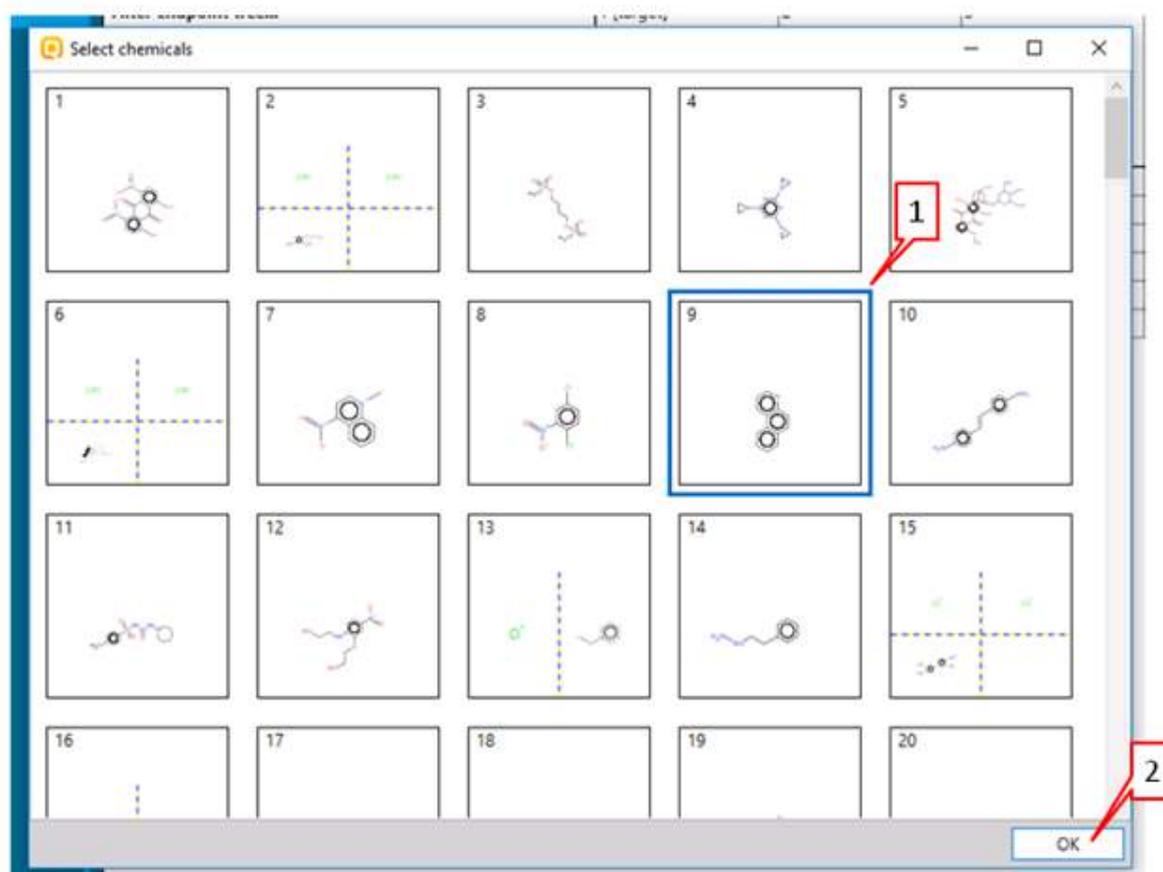


Figure 3

-0-

D.1.2. List of chemicals

QSAR Toolbox v.4

D.1.2.1. Loading a database

[Workflow](#) / [Chemicals input](#) / [List of chemicals](#) / D.1.2.1. Loading a database

D.1.2.1. Loading a database

To load an existing database, the user should press the button Database (1). A list containing the available databases will appear. You will need to select the database (2), click on *About* button (if more info about the database is needed) (3) and then confirm the selection via the button *OK* (4) (Figure 29).

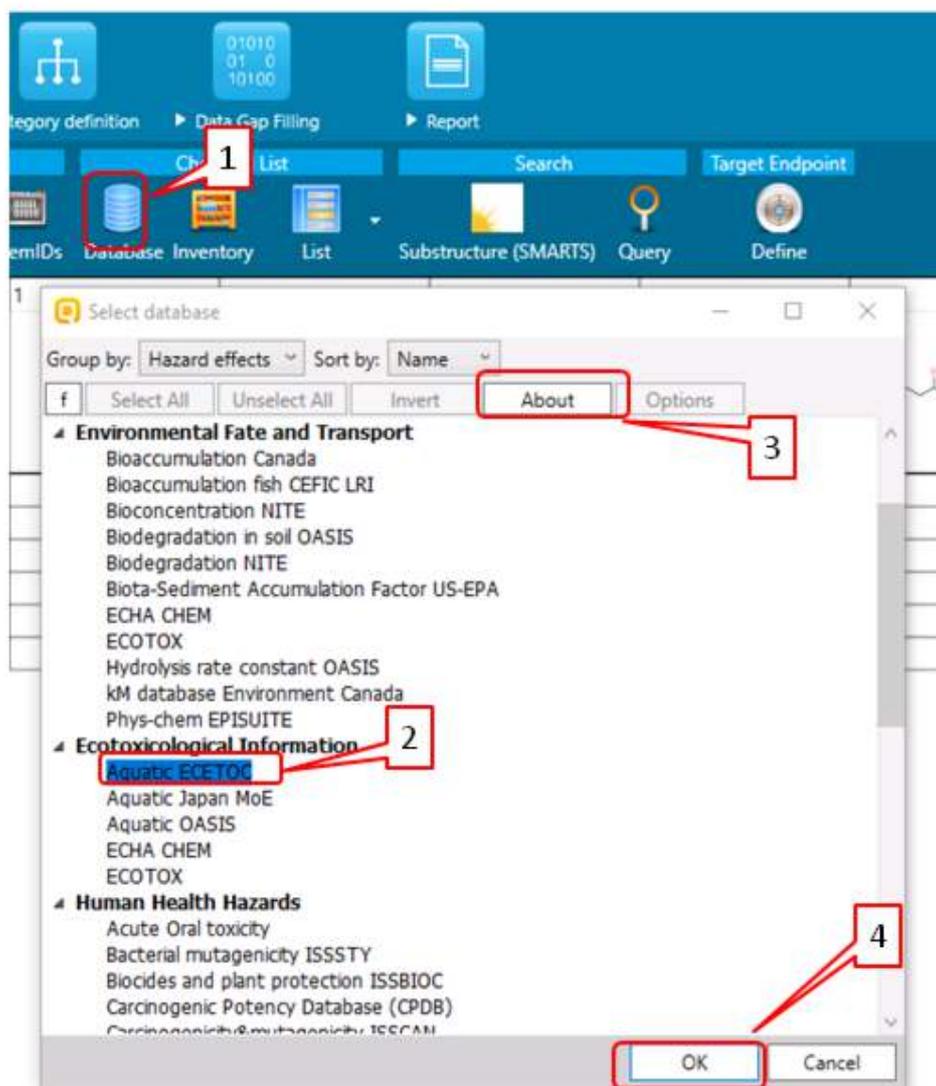


Figure 1

Options intended to help the user when selecting are database are (Figure 2):

- group by hazard effect (1);
- sort be name(2);
- filter option(3), which when selected opens a new field (4) where the user can write part or the whole name of database of interest.

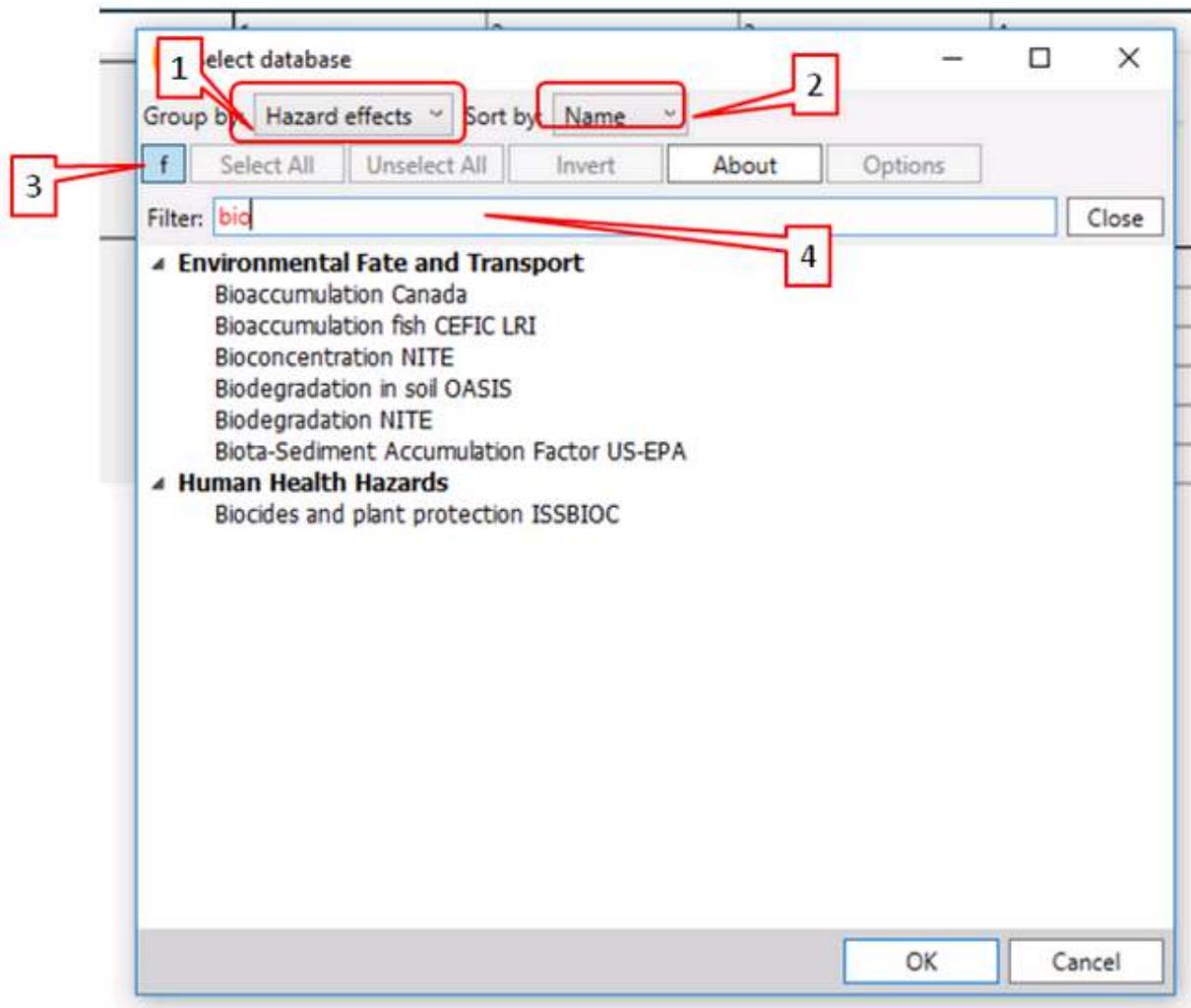


Figure 2

QSAR Toolbox v.4

D.1.2.2. Loading an inventory list

[Workflow](#) / [Chemicals input](#) / [List of chemicals](#) / D.1.2.2. Loading an inventory list

D.1.2.2. Loading an inventory list

To load an existing inventory press the *Inventory* button (Figure 1) (1). A list containing available inventories will appear. You will need to select the inventory (2) from which the target will be chosen and then to confirm the selection via the *OK* button (3).

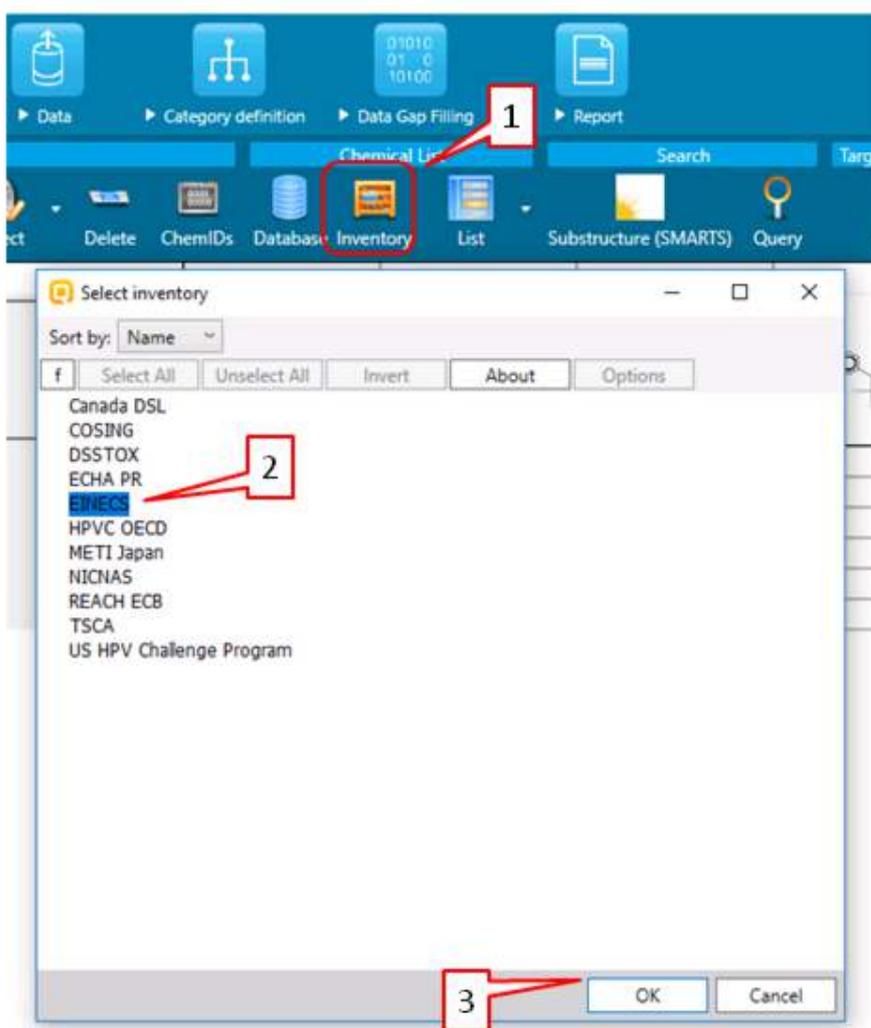


Figure 1

Options intended to help the user when selecting the inventories are (Figure 2):

- sort by name (1);
- filter option (2), which when selected opens a new field (3) where you can write part or the whole name of inventory of interest.

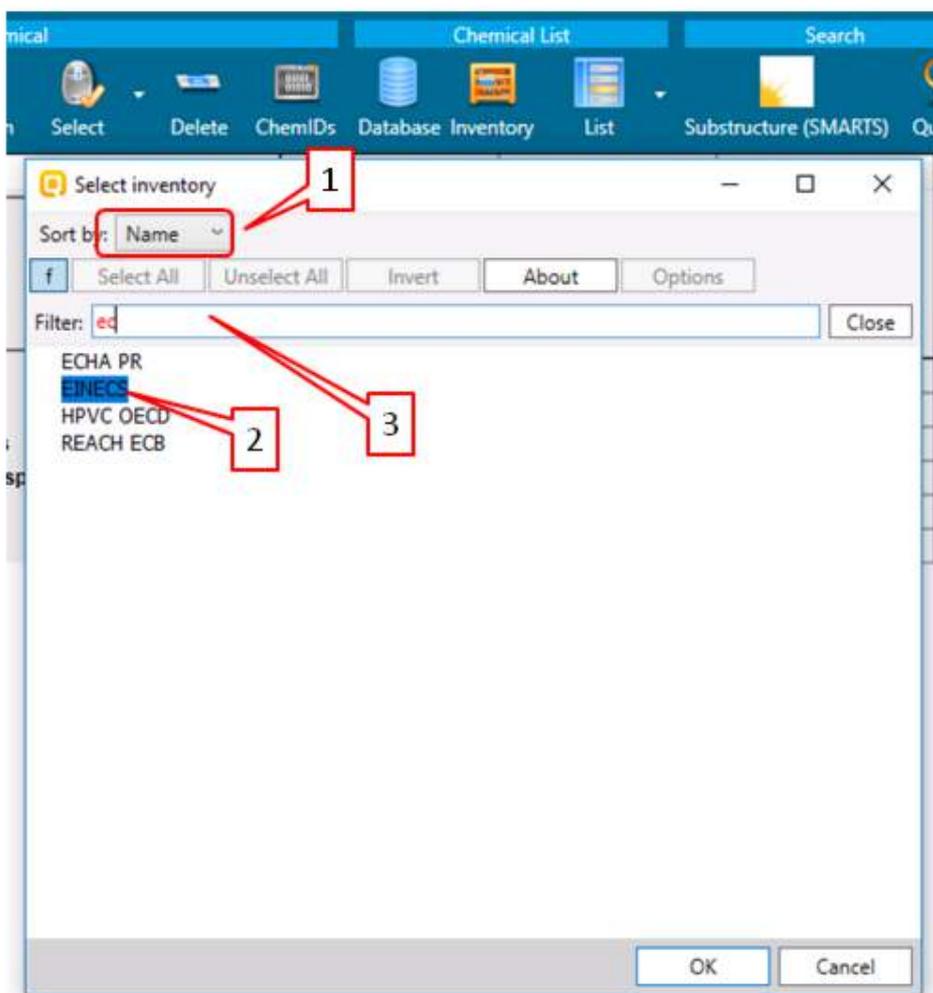


Figure 2

-0-

QSAR Toolbox v.4

D.1.2.3. Loading a custom file

[Workflow](#) / [Chemicals input](#) / [List of chemicals](#) / D.1.2.3. Loading a custom file

D.1.2.3. Loading a custom file

A set of chemicals could be loaded via a list with specified atom connectivity stored in a file. Currently supported by the system atom connectivity file formats are SDF and SMI. Files with SMILES may contain TAB delimited CAS number and chemical name followed by the corresponding SMILES string. Usually the extension of these files is *.smi.

In addition, you have an option to load a file from the last used folder or from Toolbox example files (Figure 1)



Figure 1

Press the button *List* (1), select the pathway (2), name of the file (3) and then press *Open* (4). (Figure 2)

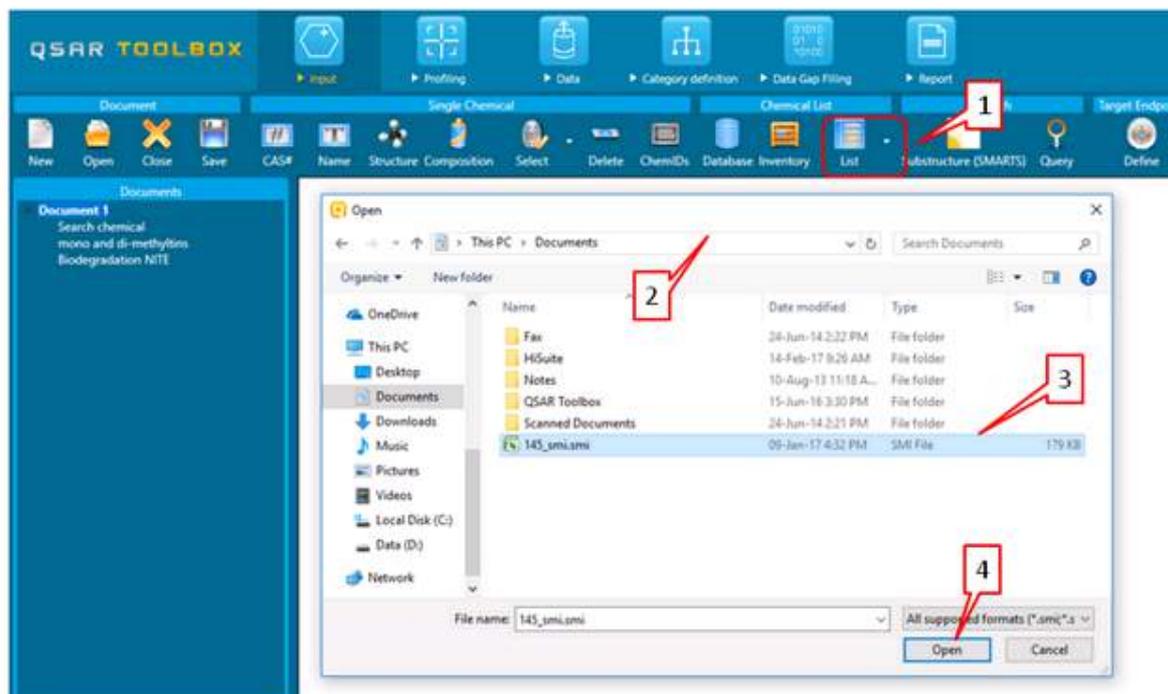


Figure 2

-O-

QSAR Toolbox v.4

D.1.3. Search options

[Workflow](#) / [Chemicals input](#) / D.1.3. Search options

D.1.3. Search options

There are two types searching options:

- [Substructure \(SMARTs\)](#)
- [Query tool](#)

-0-

QSAR Toolbox v.4

D.1.3.1.Substructure (SMARTS)- 2DEditor

[Workflow](#) / [Chemicals input](#) / [Query tool](#)/ D.1.3.1.Substructure (SMARTS)- 2DEditor

D.1.3.1.Substructure (SMARTS)- 2D Editor

The user could select the Substructure (SMARTS) option (Figure 1). This option is useful when the user does not have a specific chemical in mind but rather wants to find a group of chemicals containing the same structural characteristics.



Figure 1

2D Editor incorporating SMART language is visualized in Figure 2. Its main options are:

- Buttons for editing (1): Pencil button (draw), Eraser button (delete), Magnifying glass button (Zoom); Blank page (delete all), Arrows buttons (Undo/Redo)
- SMART type fragments buttons (2) (they all have hints when the user position the cursor on the button):
 - "[" button (enumeration of atoms): allows enumeration of atoms which could be bonded to a given atom in the molecule;
 - "\$" buttons (allows recursive SMART expression): the recursive expression allows to describe the surrounding of an atom. For example it could be used to describe N -atom, which is not part of nitro group;
 - "Rpt" button (repeat fragment): allows to specify how many times an

atom or a fragment could be repeated in the molecule;

- "Rpc" button (replace fragment): gives an option to replace a given atom in the molecule with different chemical elements;

- "Exh" buttons (exhaust fragment): allows only those atoms or fragments listed in the exhaust fragment to be bonded to a given atom.

- Periodic table (3)
- Frequently used atoms (4)
- Selection tool (5) - has to be used every time when the user wants to point out the atom, which characteristics has to be defined.

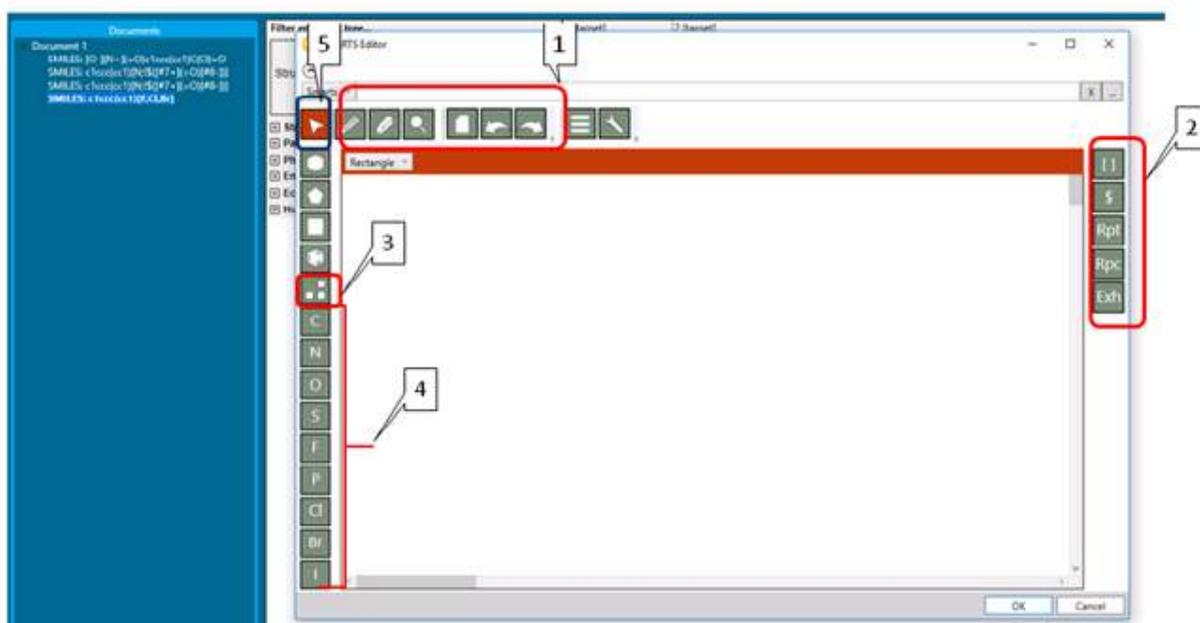


Figure 2

Other options are shown in Figure 3:

- Templates of the four most common rings (1);
- Button (2) opens fragment browser, which contains list of templates of cyclic structures (3) and list of fragments (4). The fragments are the same as the ones on the right in the 2D editor.

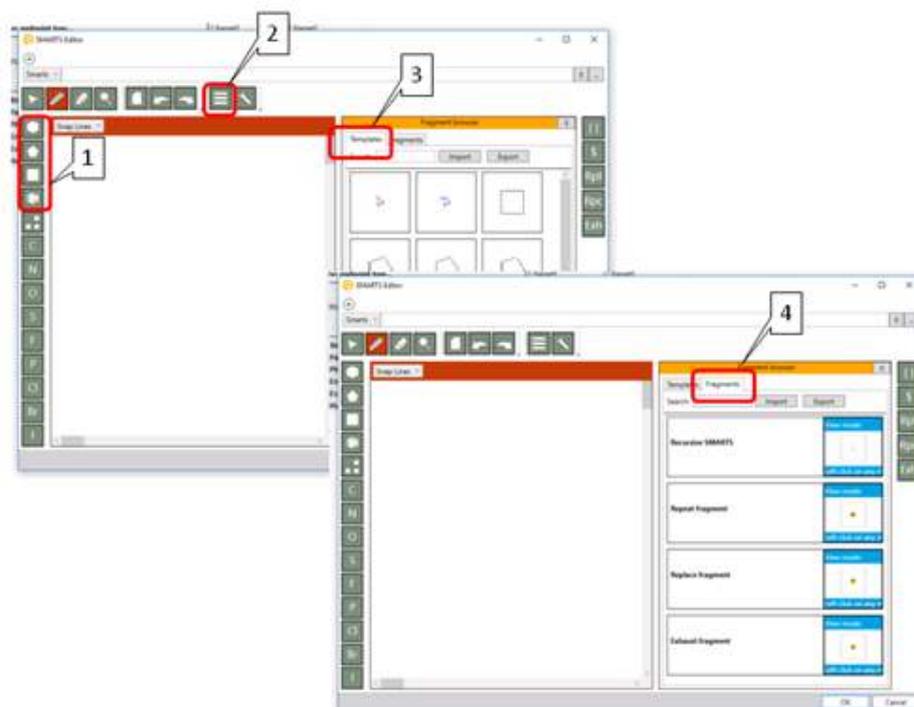


Figure 3

The user can also add their structures to the fragments (Figure 4). After the structure is drawn right click and then select add to library. The library of fragments can also be exported and saved as sfl file and imported later if needed.

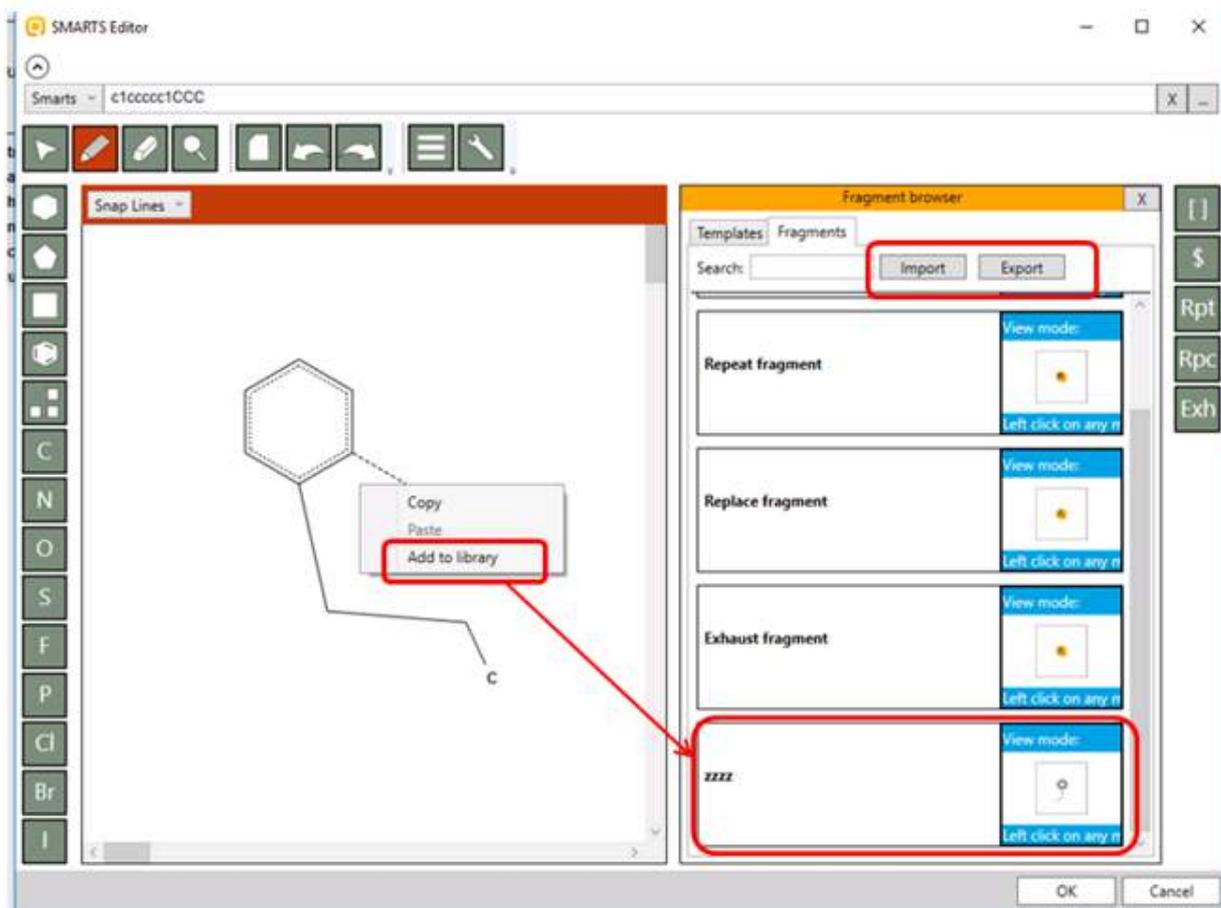


Figure 4

An example of substructure search is shown in Figure 5. Chemicals containing halobenzene moiety are selected from the selected databases and inventories.

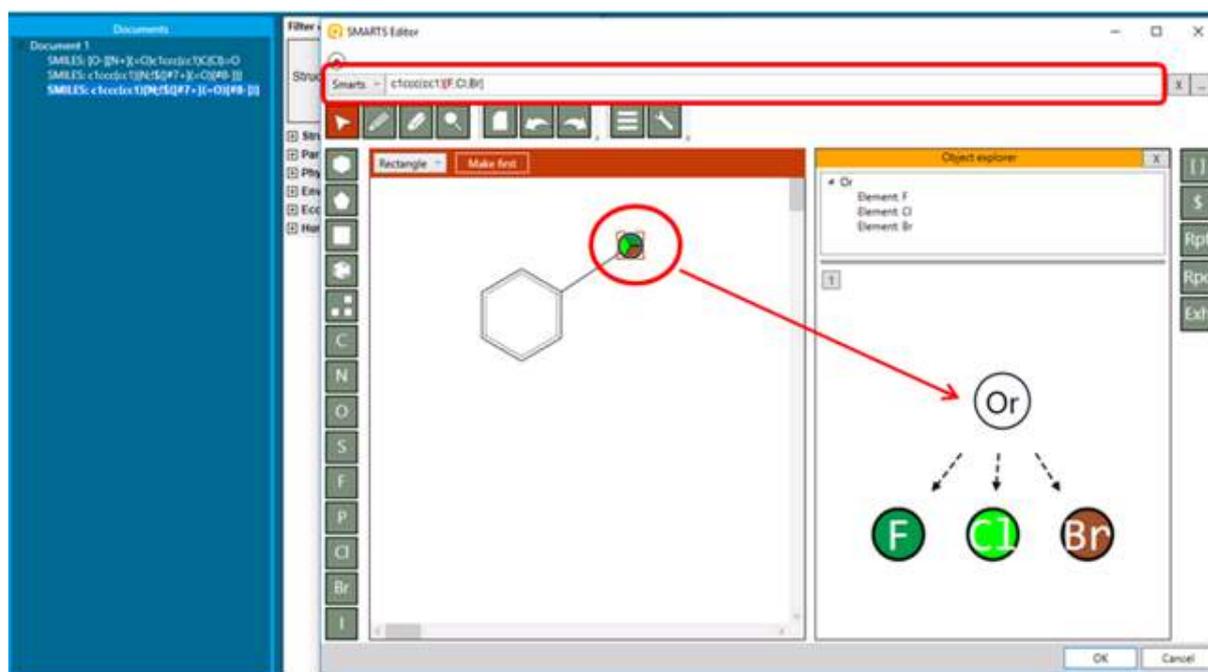


Figure 5

-O-

QSAR Toolbox v.4

D.1.3.2. Query tool

[Workflow](#) / [Chemicals input](#) / [Query tool](#)/ D.1.3.2. Query tool

D.1.3.2. Query tool

Query tool is accessible from Toolbox input panel (1) (Figure 1):

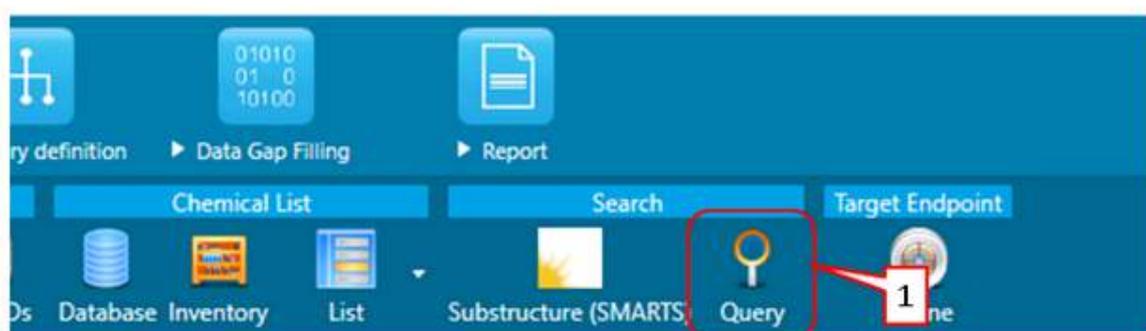


Figure 1

The Query search is performed on selected database and inventories. A message reminds the user to select databases and/or inventories before continuing with their search (Figure 2):

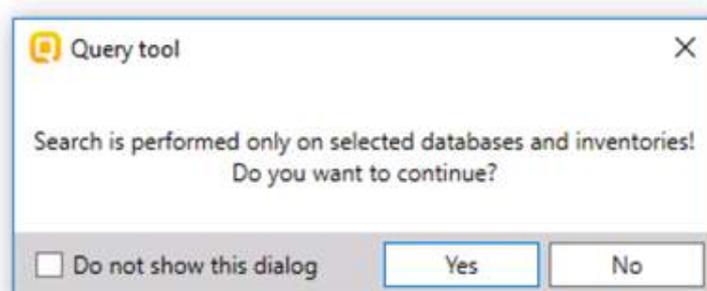


Figure 2

The Query tool main components are (Figure 3):

- o Query edit panel (1)
- o Query tree logic panel (2)

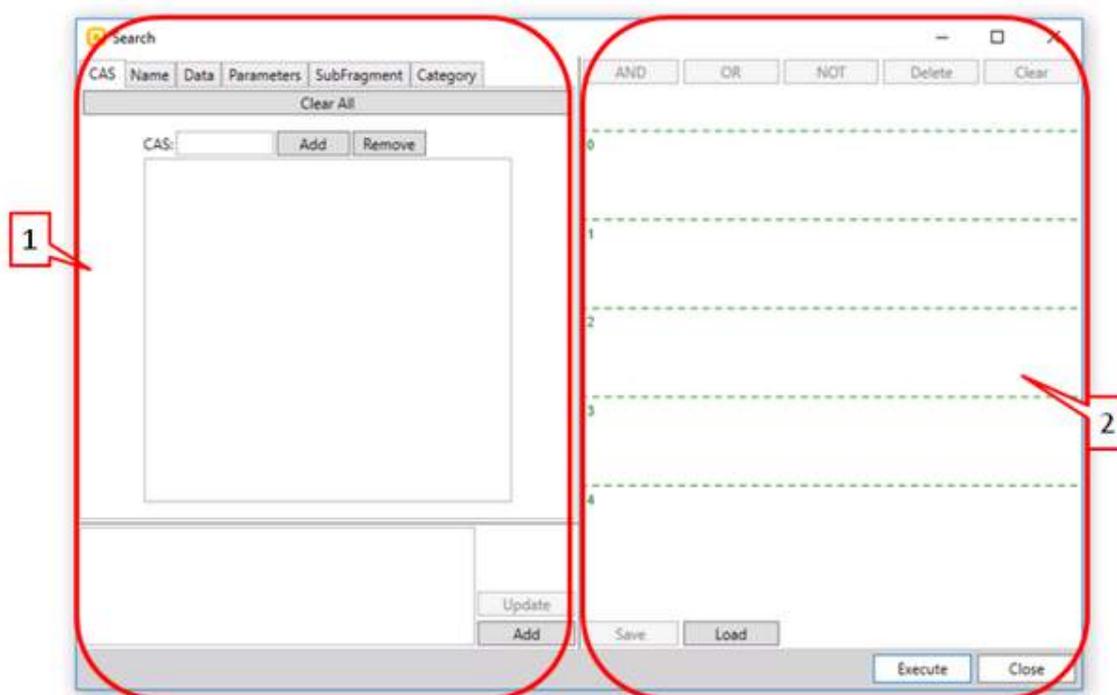


Figure 3

- General features of Query tool edit panel (Figure 4):
 - Main Search tabs (1): CAS, Name, Data, Parameters, SubFragment, Category

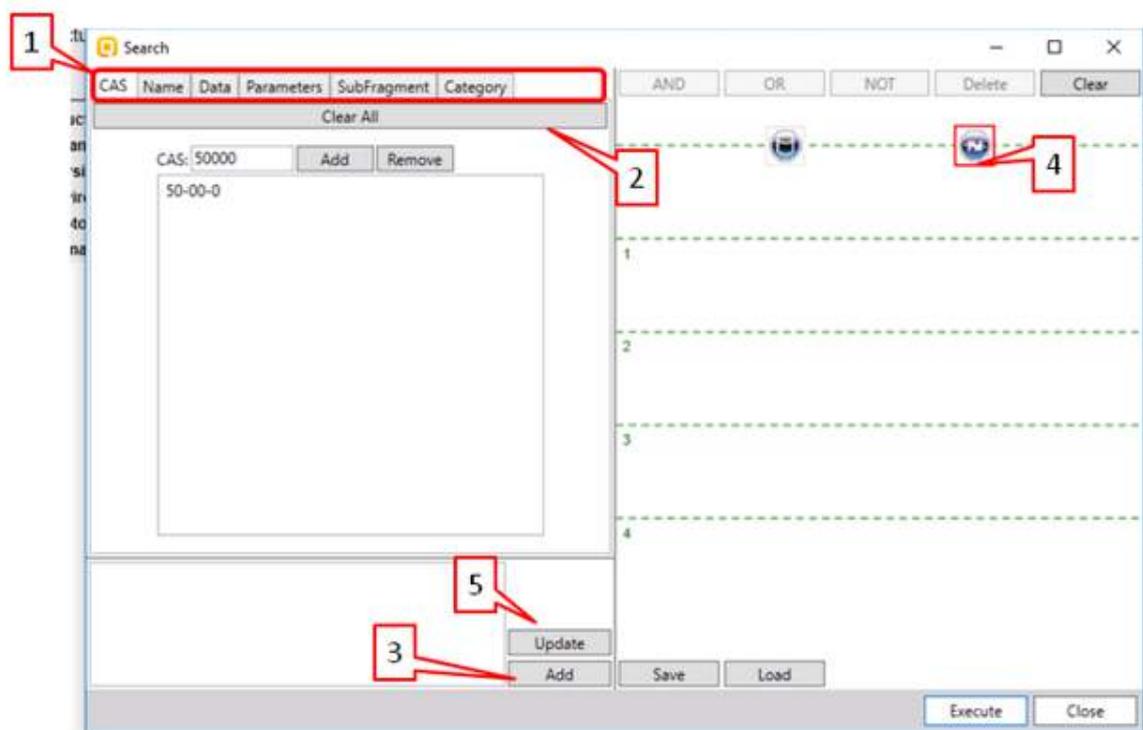


Figure 4

- *Clear All* button (2), which deletes all entries in the search group panel;
- *Add* button, adds all selected search requirements from one tab as a query (4) in the right panel for execution.
- When a query is selected it is circled in red (4). The user has also an option to change the active query by modifying some of the entries and then pressing *Update* (5).
- Right-click on the query will colored it in yellow (1) and the logical operation *Not* (2) can be added (Figure 5). Also the query can be *Deleted* (2). Button *Clear* deletes all queries and logical operation in the panel not only the selected query. All Query and logical operations can be saved by pressing *Save* (3). A saved query can be loaded by pressing *Load* (4).

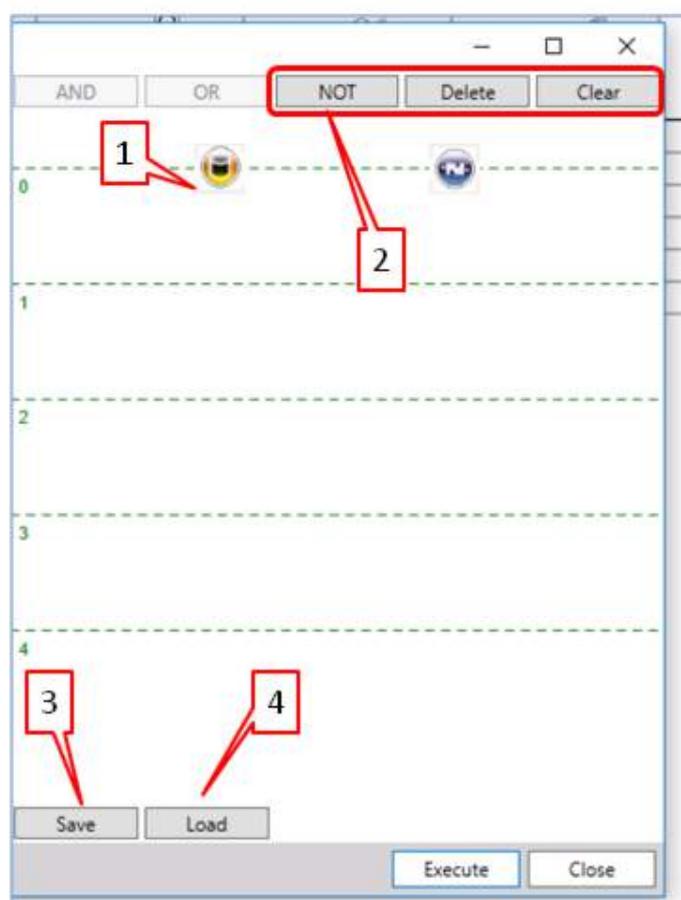


Figure 5

- To insert the logical operation AND or OR, the user has to right-click on the selected queries (colouring them in yellow)(1) and then to select the logical operation (2) (Figure 6). Right-click on the logical operation (colored in red)(4) allows to add NOT or to Deleted (5) it.

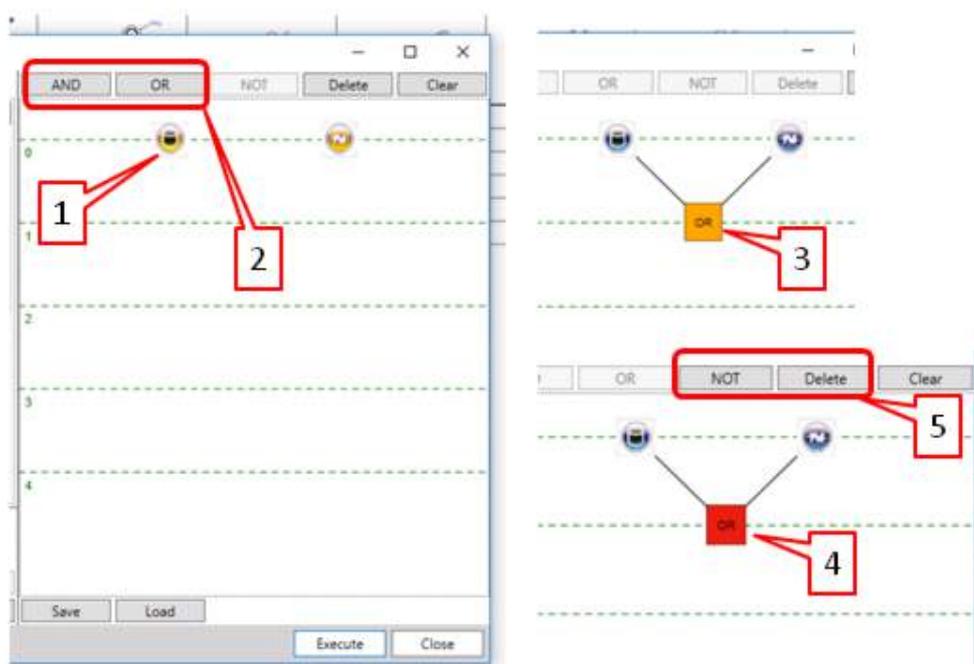


Figure 6

- CAS

In the panel CAS you can do a search in the selected databases or inventories based on CAS RNs. First write the CAS RN in the cell (1 in Figure 7) and press *Add* button. An option to remove CAS numbers is also implemented (2).

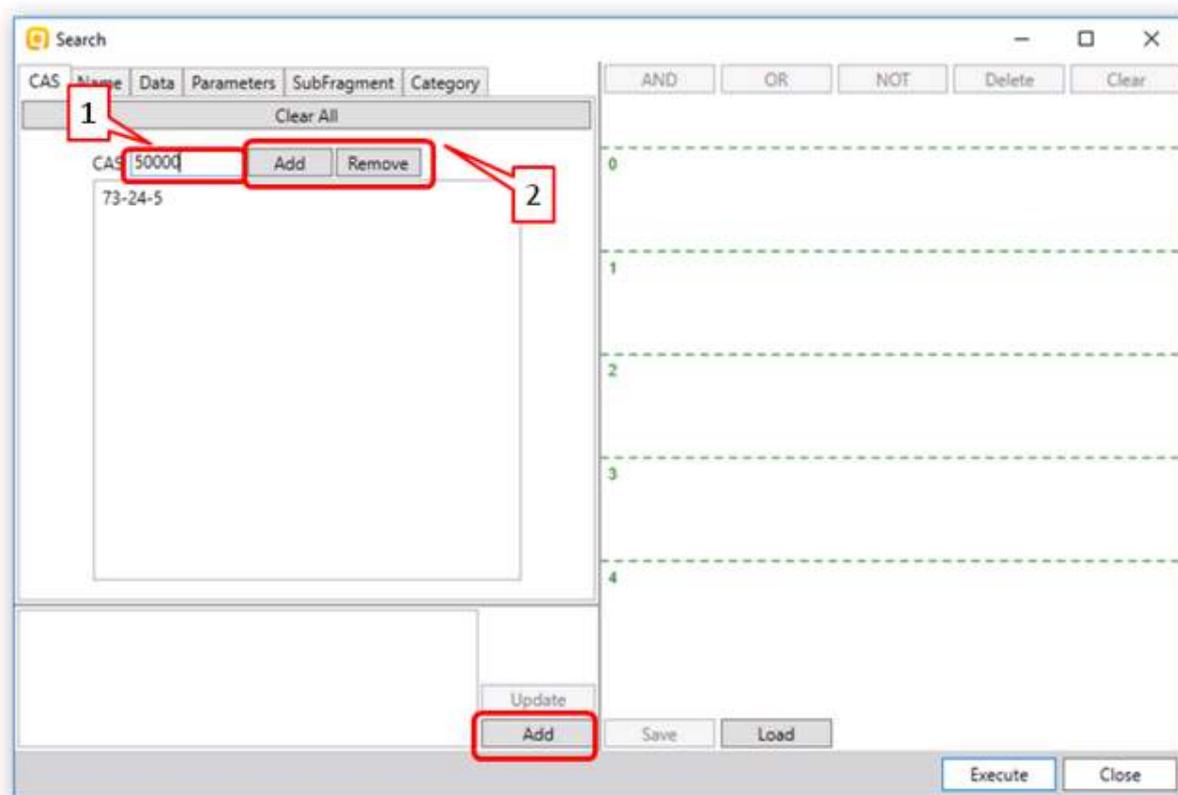


Figure 7

Once the CAS RNs of interest are entered, press add (1), and a new query will appear (2) and then press *Execute* (3) (Figure 8).

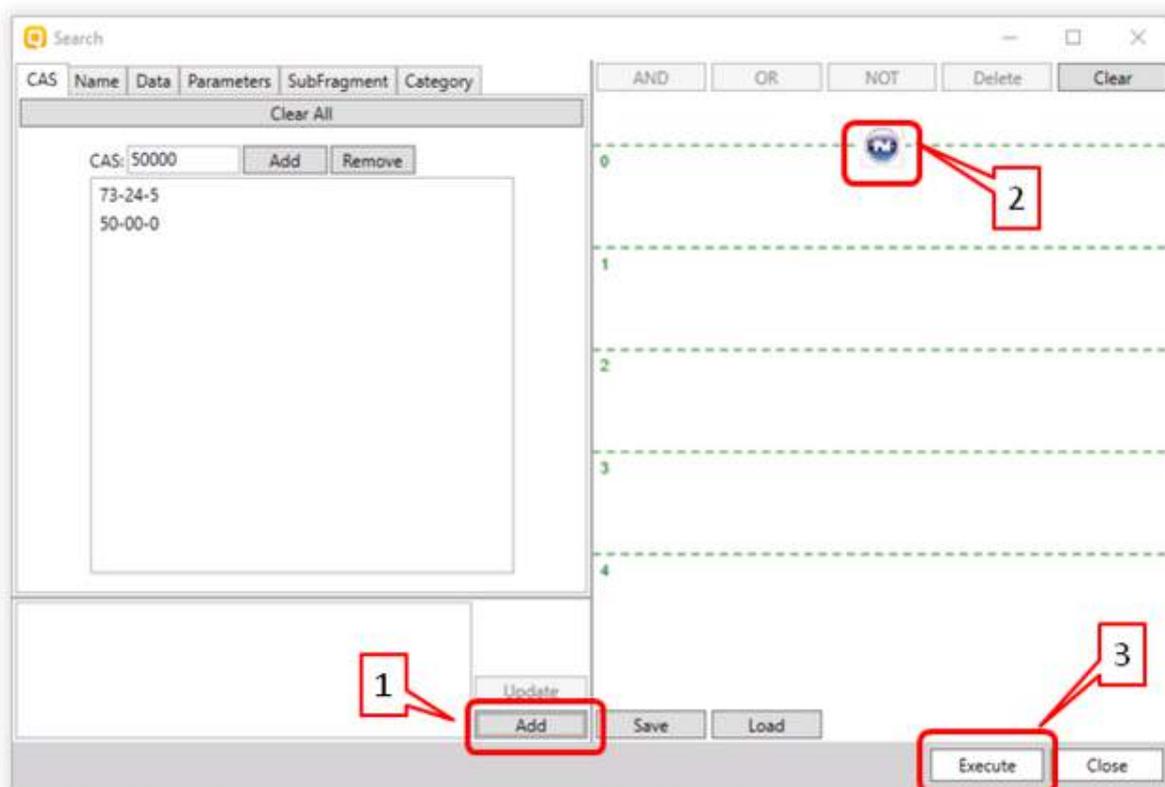


Figure 8

All chemicals corresponding to the CAS RN entered will appear on the matrix (see Figure 9).

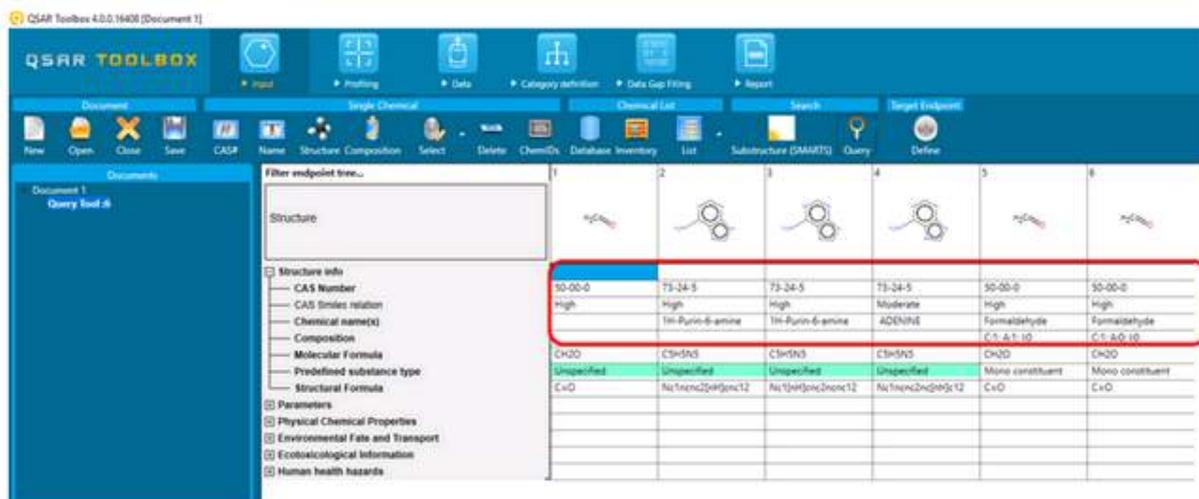


Figure 9

- Name

In the panel Name a search can be performed in the selected databases or inventories (in section Data) based on chemical names. First you need to write the name in the cell (1 in Figure 10), select additional options about the name (2) and then press *Add* button (3). An option to remove the entered names is also implemented (3).

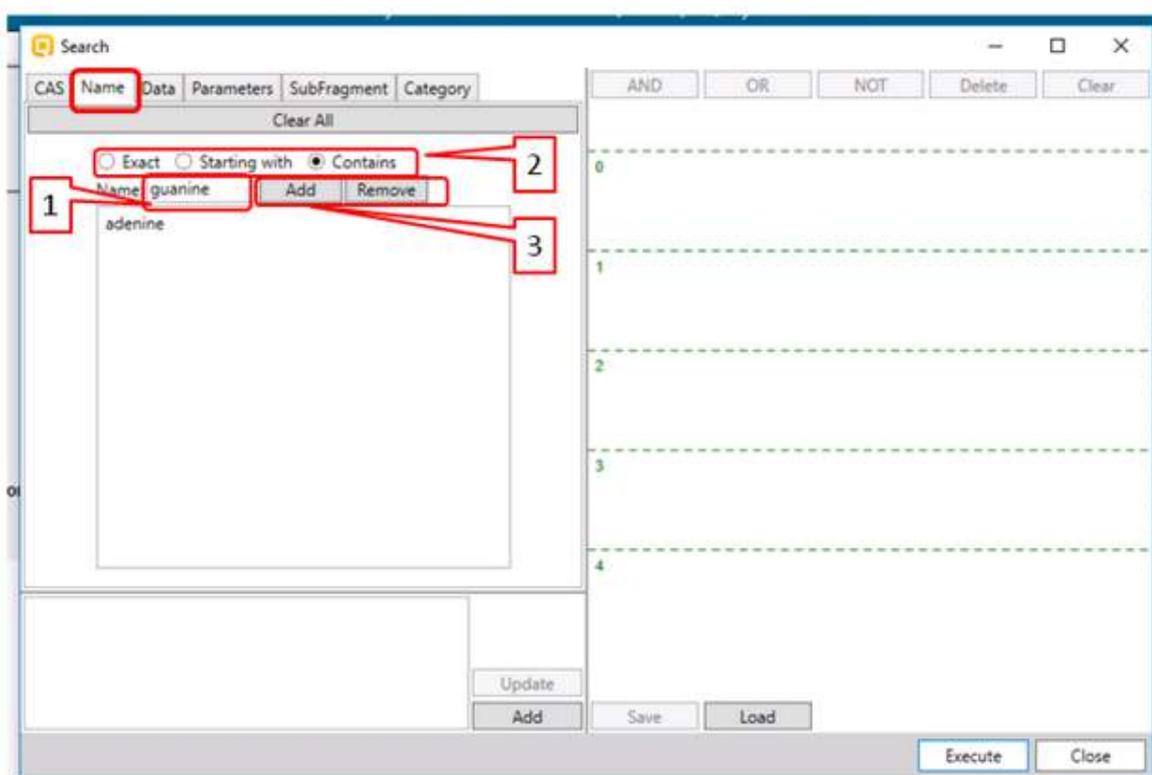


Figure 10

Once the chemical names of interest are entered, you need to press *Add* (1), a new query will appear (2) and then press *Execute* (3). (Figure 11)

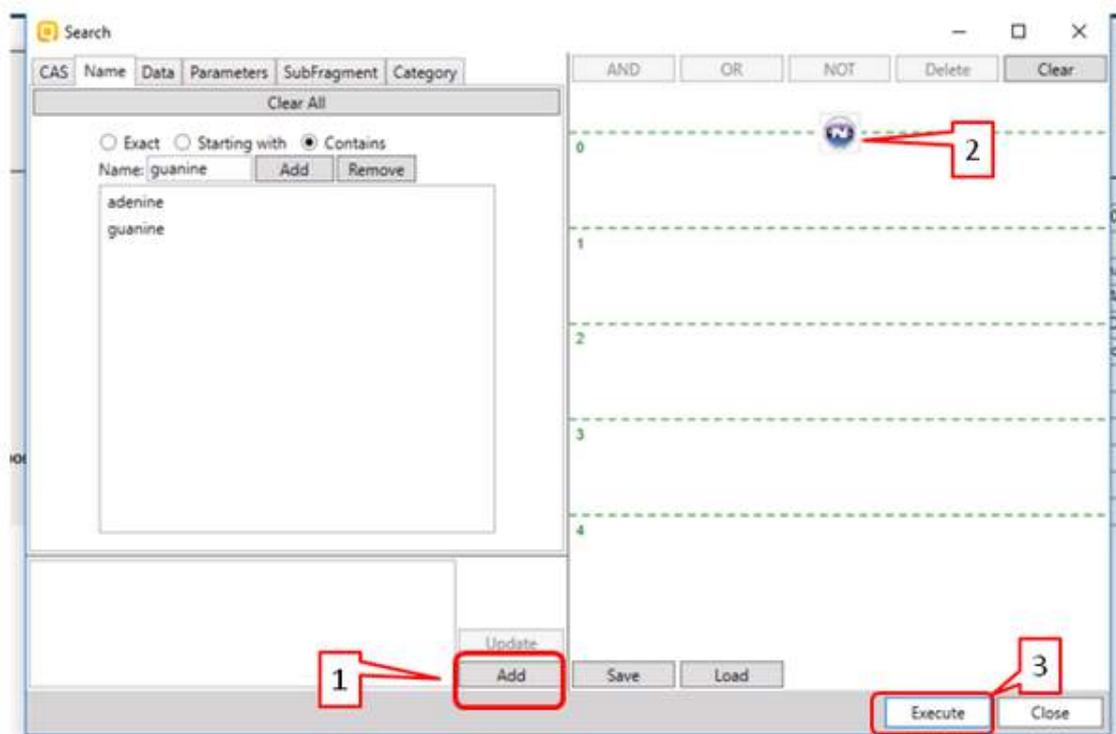


Figure 11

In the example in Figure 12, the program will search for any chemicals containing adenine OR guanine in their names in the selected databases or inventories.

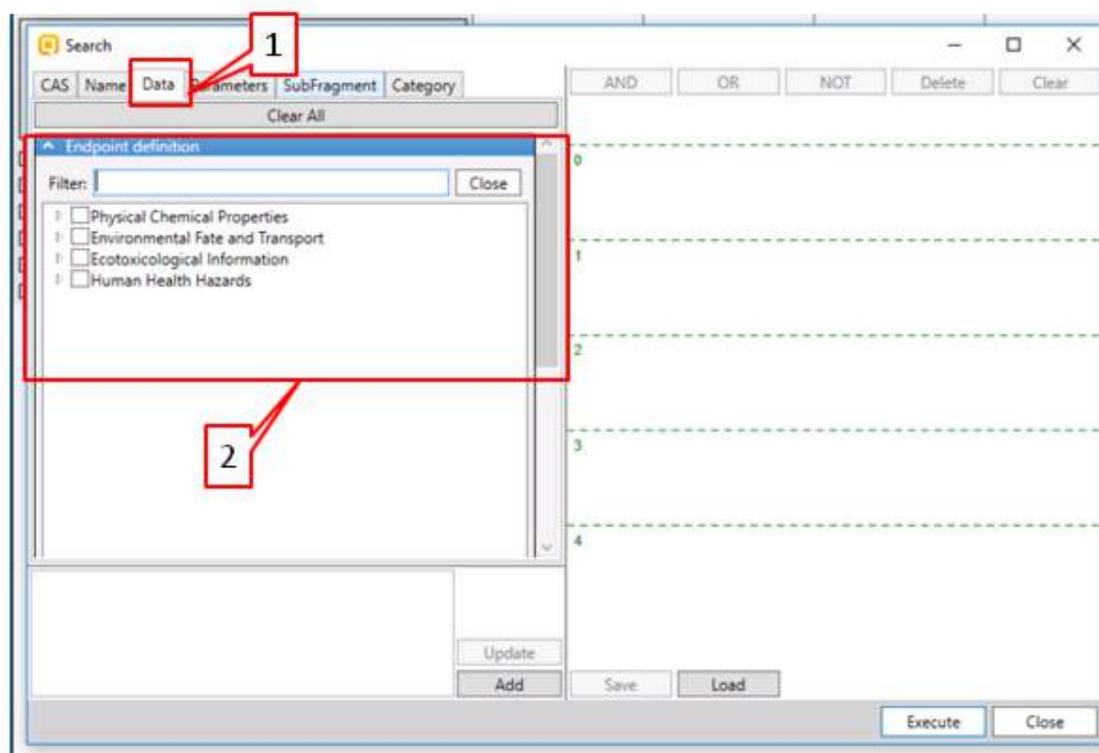


Figure 12

- Data

The endpoints visualized (2) in the panel *Data* (1) are based on the databases that are selected beforehand (Figure 13). If only databases containing Ecotoxicological information are selected for example, then only that node will be displayed.

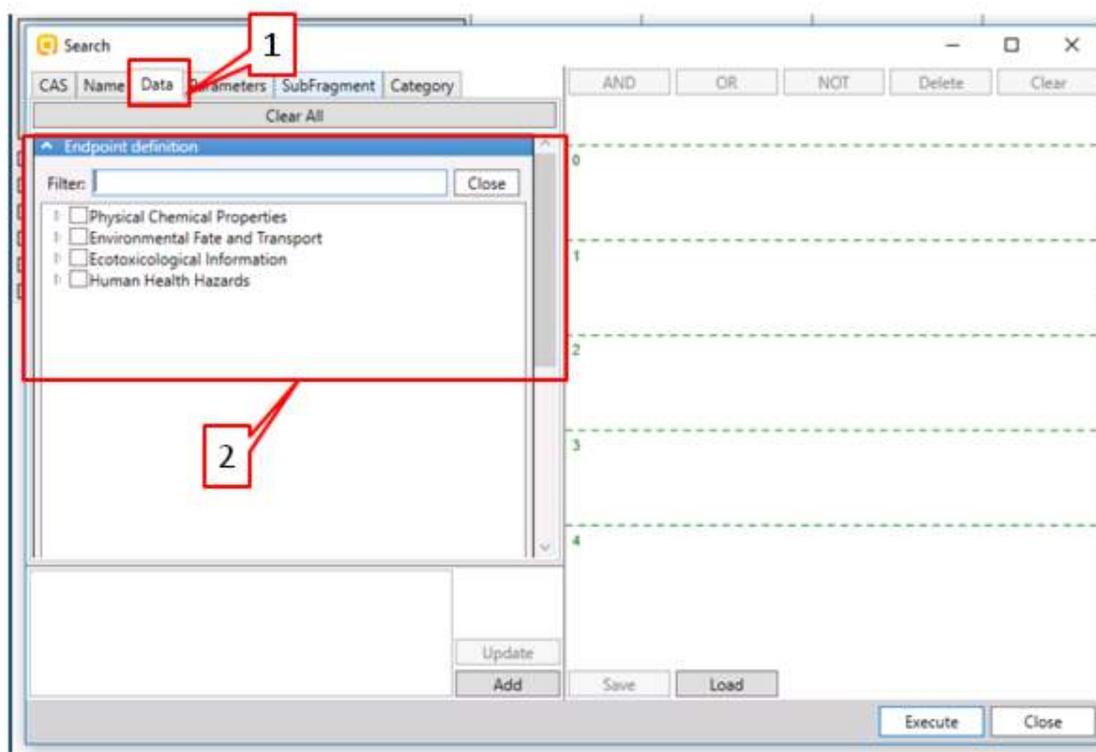


Figure 13

The selection procedure include specification the following steps: the user could specify the:

- endpoint (1)

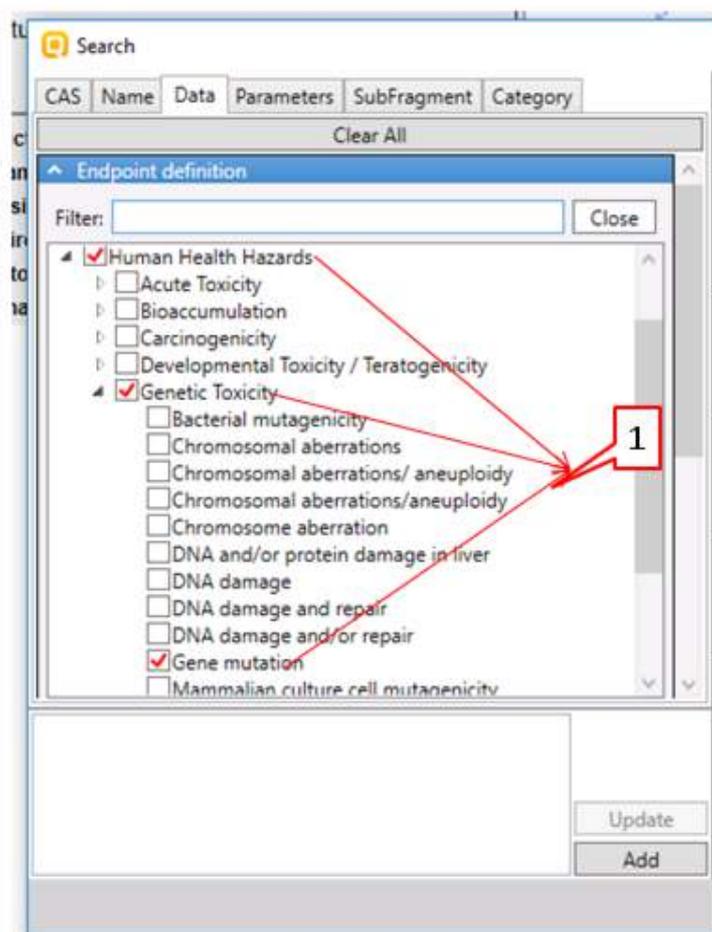


Figure 14

- meta data: from the drop-down arrow (1) select the type of metadata and then press *Add*(2). (Figure 15)

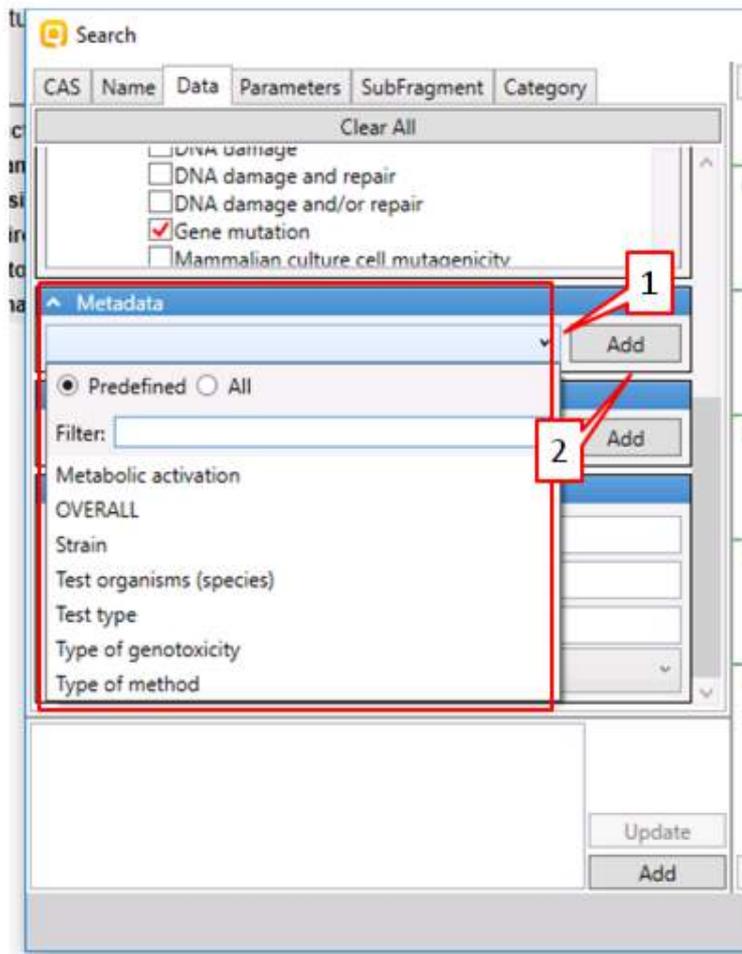


Figure 15

For example, if test organism species is selected (1), then select the type of species (2) and then press Add (3) (Figure 16).

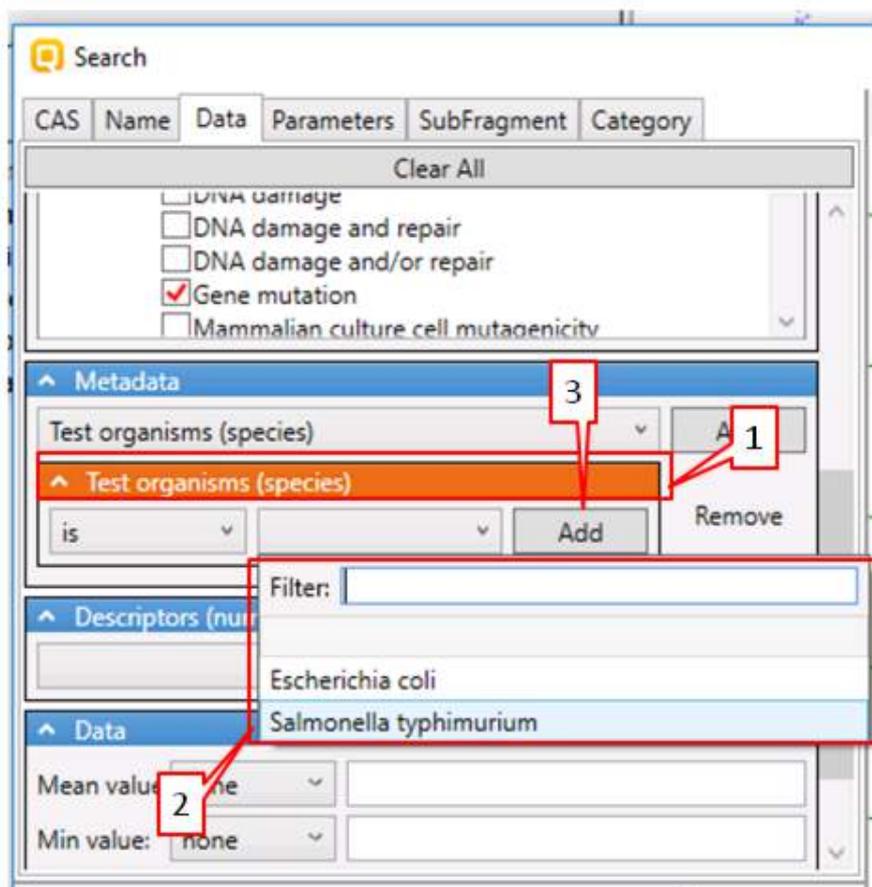


Figure 16

If a descriptor (numerical data), e.g. duration, solubility, etc., is associated with the selected endpoint then it will be visualized in (1) (Figure 17):

The screenshot displays the 'Data' tab of the OECD QSAR Toolbox application. The interface is organized into several sections:

- Metadata:** Contains a 'Test organisms (species)' dropdown menu with an 'Add' button. Below it, a sub-section titled 'Test organisms (species)' is highlighted with a red box. It includes a dropdown menu with 'is', a text input field containing 'Salmonella typ...', and an 'Add' button. A red callout box with the number '1' points to the 'Add' button. Below this, the text 'is 'Salmonella typhimurium'' is displayed with a 'Remove' button.
- Descriptors (numerical metadata):** This section is also highlighted with a red box and contains a dropdown menu and an 'Add' button.
- Data:** This section contains four rows of input fields:
 - Mean value: none (dropdown) and an empty text input field.
 - Min value: none (dropdown) and an empty text input field.
 - Max value: none (dropdown) and an empty text input field.
 - Unit: (dropdown) and an empty text input field.

Figure 17

Selection of an Unit (Scale) (1) and categorical data (2) is also included (Figure 18):

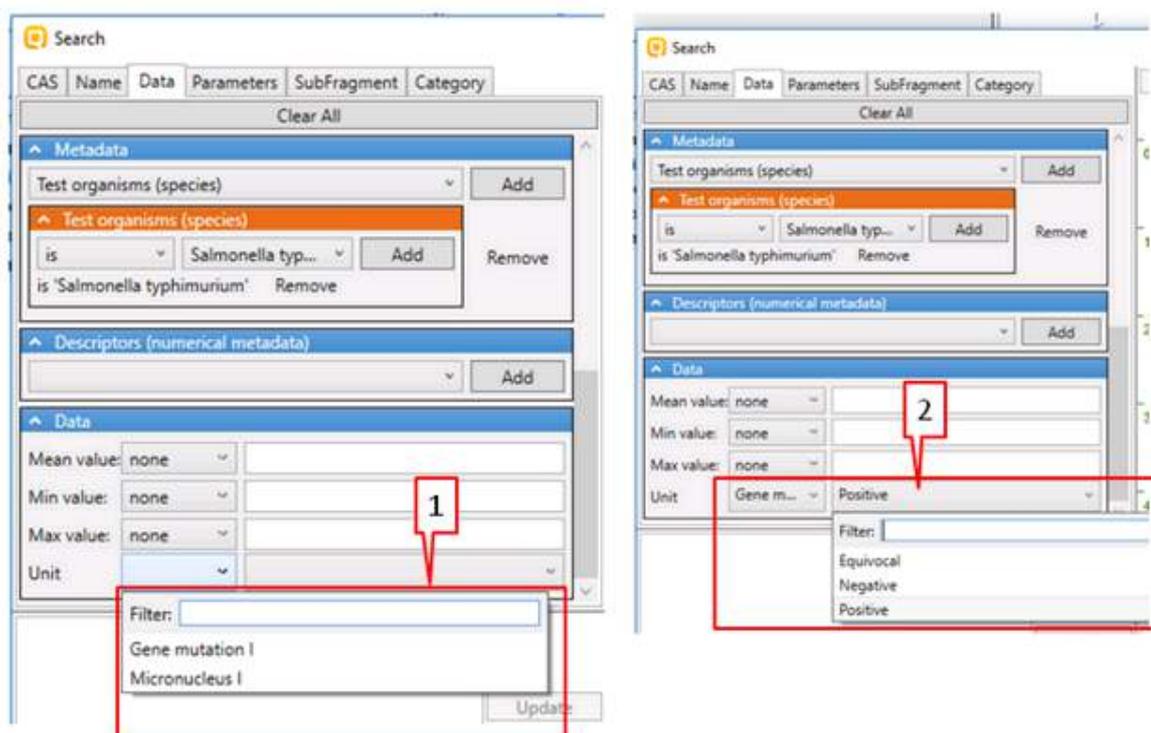


Figure 18

Once all fields are filled, you have to press *Add* (1), a new query will be visualized in the right panel (2) and then press *Execute* (3) or to double click on the query (Figure 19).

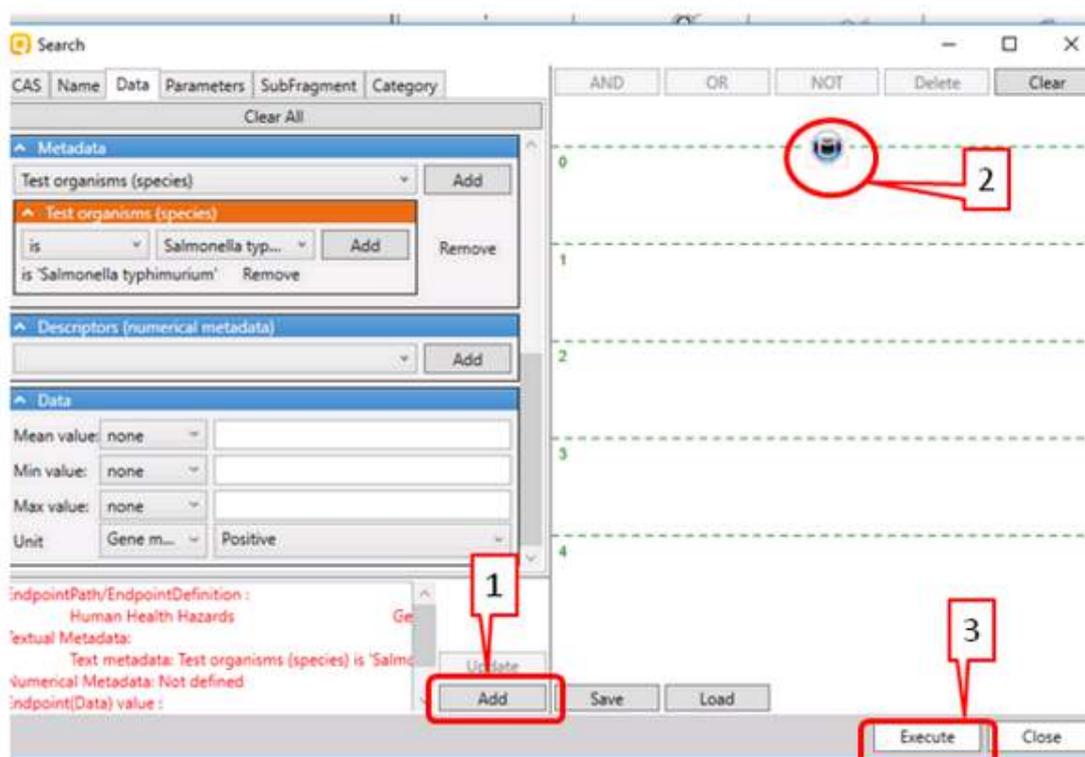


Figure 19

- Parameters panel

The third panel *Parameters* (1) provides the user possibility to search databases by 2D or 3D parameters (2) (Figure 20). The qualifiers are available in order to set different options for searching (3).

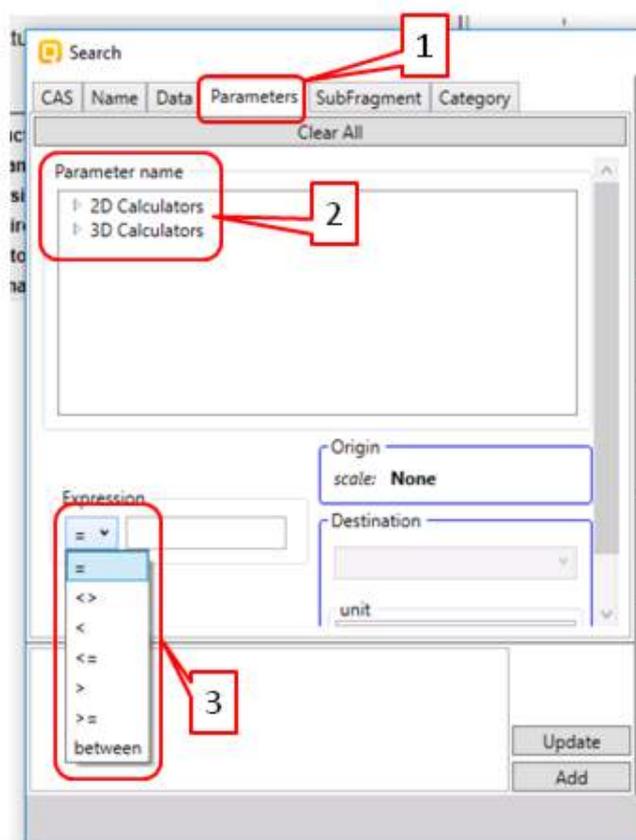


Figure 20

The selection procedure includes the following steps: select a parameter (1), select the qualifier and type the value (2), select the type of expression (3), select the unit (4) (Figure 21).

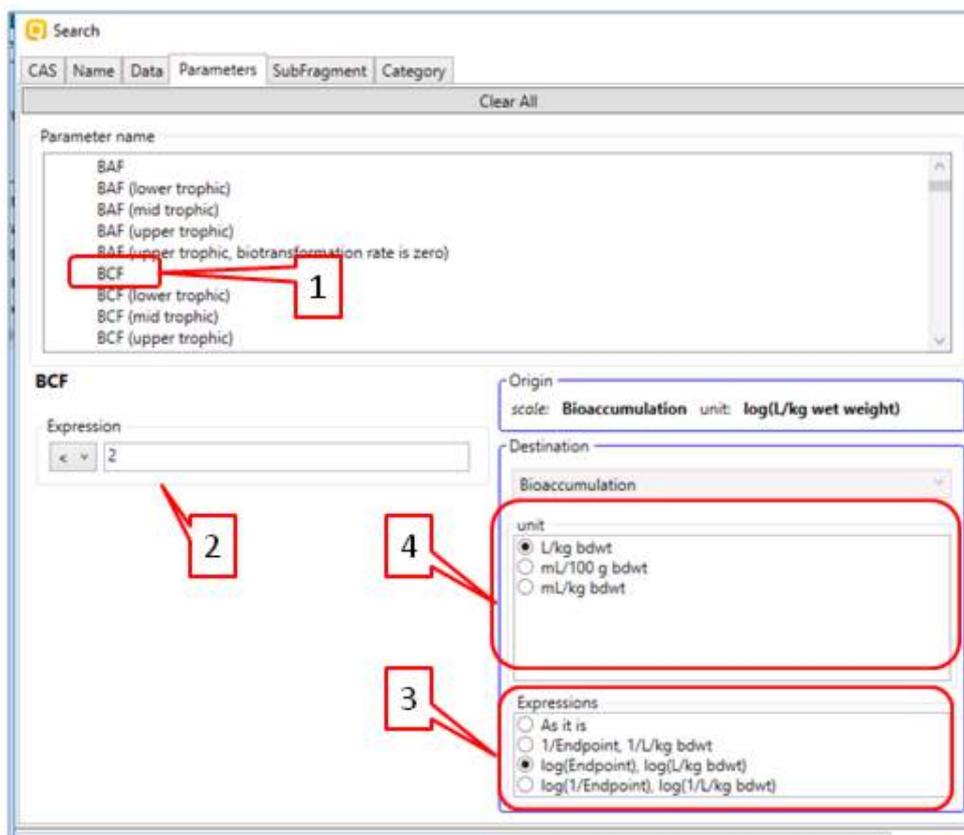


Figure 21

- SubFragment panel

The fourth panel SubFragment (1) allows to search chemical structure, which can be drawn (4) (or paste SMARTS (3)) by using *Add* (2) button (Figure 22). The SMARTS editor is visualized, where the user can paste the subfragment SMARTS (3) or draw it (4). Additional search options are also implemented (5). More information about how to use SMARTS editor could be found in [Substructure \(SMARTS\)- 2D Editor](#).

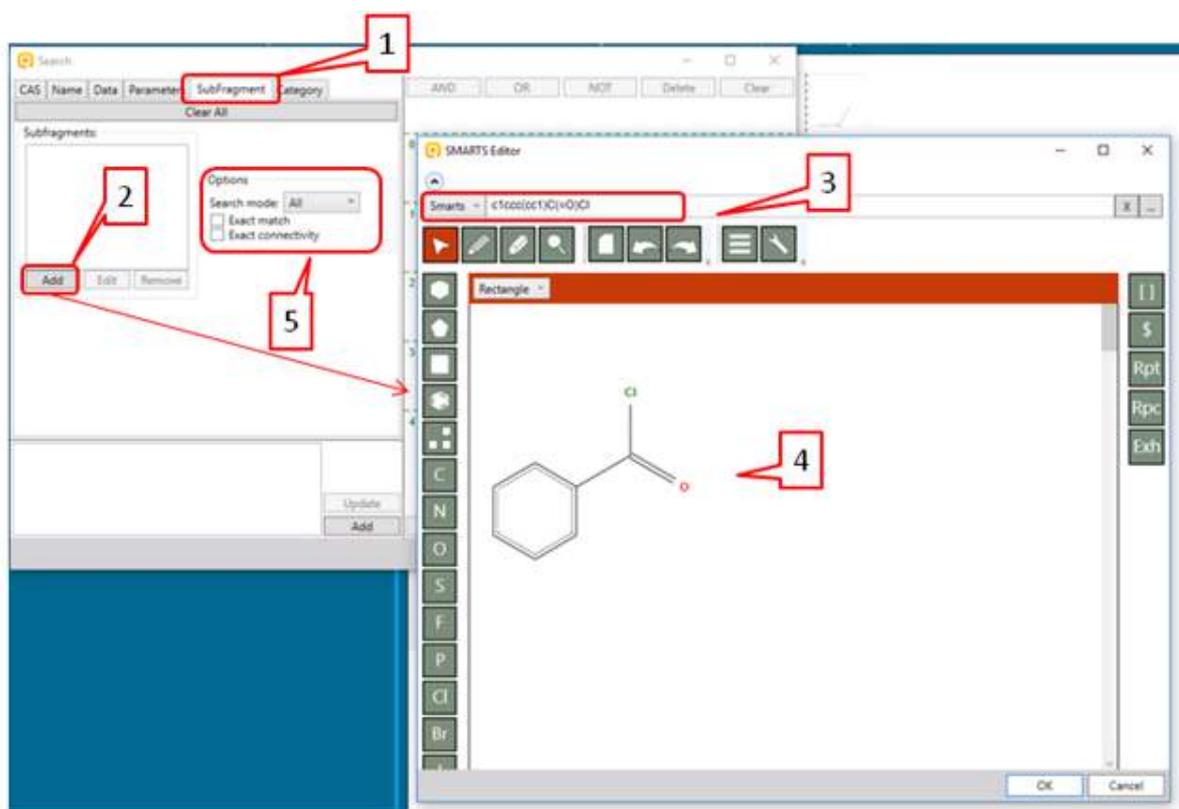


Figure 22

- Category panel

The Category panel allows searching by any of the available profiles (1) (Figure 23). Select the desired category(s) from the panel (2) and move it to the panel (3). Different search options are possible (4).

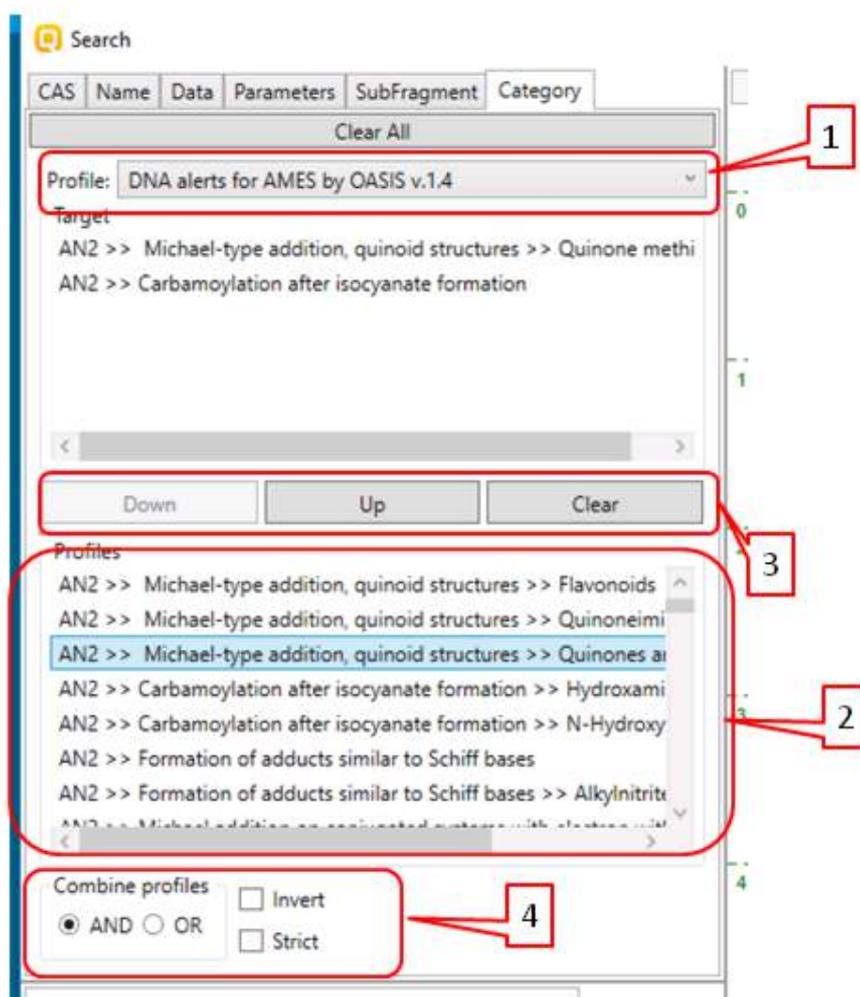


Figure 23

-0-

D.1.4. Define Target Endpoint

D.1.4. Define Target Endpoint

Preliminary defining of target endpoint is possible in Toolbox 4.0. This functionality allows entering the endpoint of interest e.g. EC3, LC50, gene mutation etc. Information for the selected-endpoint relevant profilers and databases could be further included.

Based on the specific metadata information, different relevancy scores could be provided for same endpoint. Calculation of alert performance (AP) is only possible if the target endpoint is preliminary selected.

The main steps of selecting of target endpoint are (Figure 1):

- Press CAS # button (1);
- write the CAS number in (2);
- press Search button(3);
- finally press OK(4).

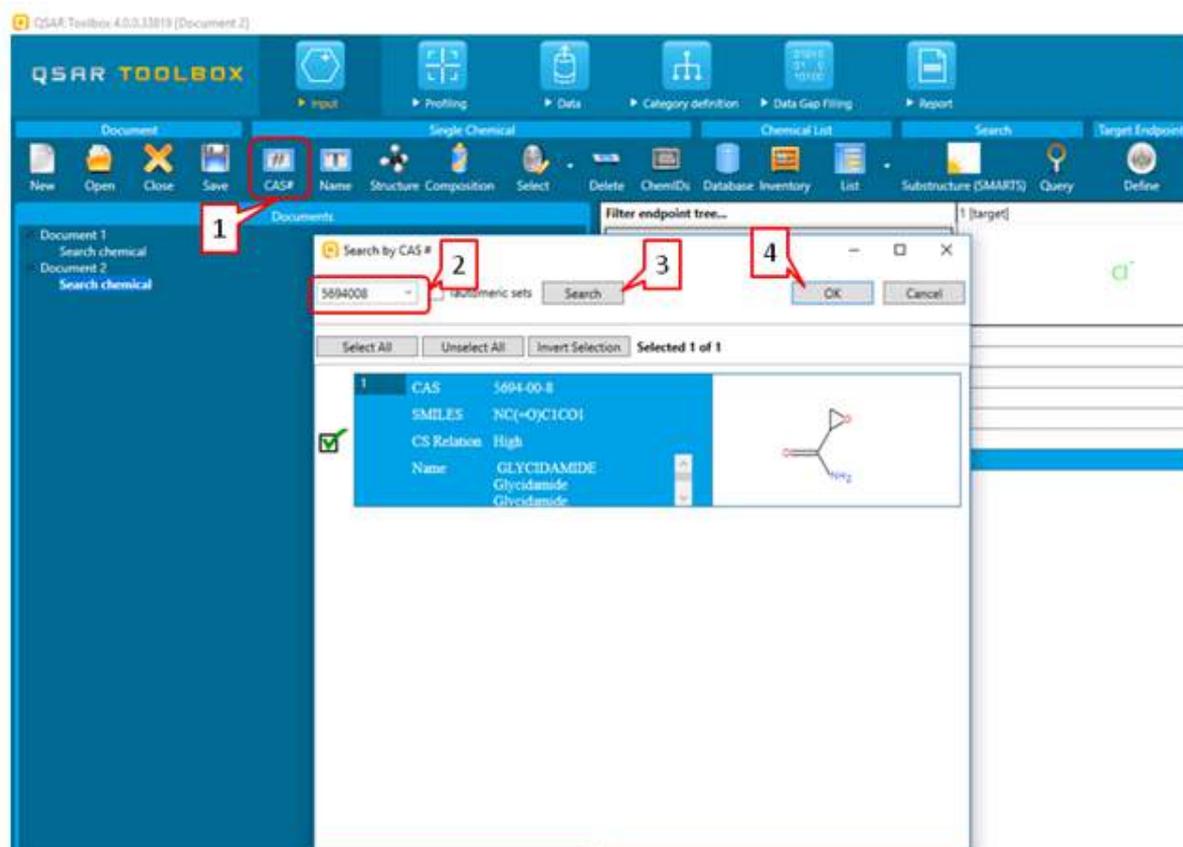


Figure 1

Once the structure is added in the data matrix, select *Define Target Endpoint* (Figure 2)(1). Define endpoint window is visualized (2).

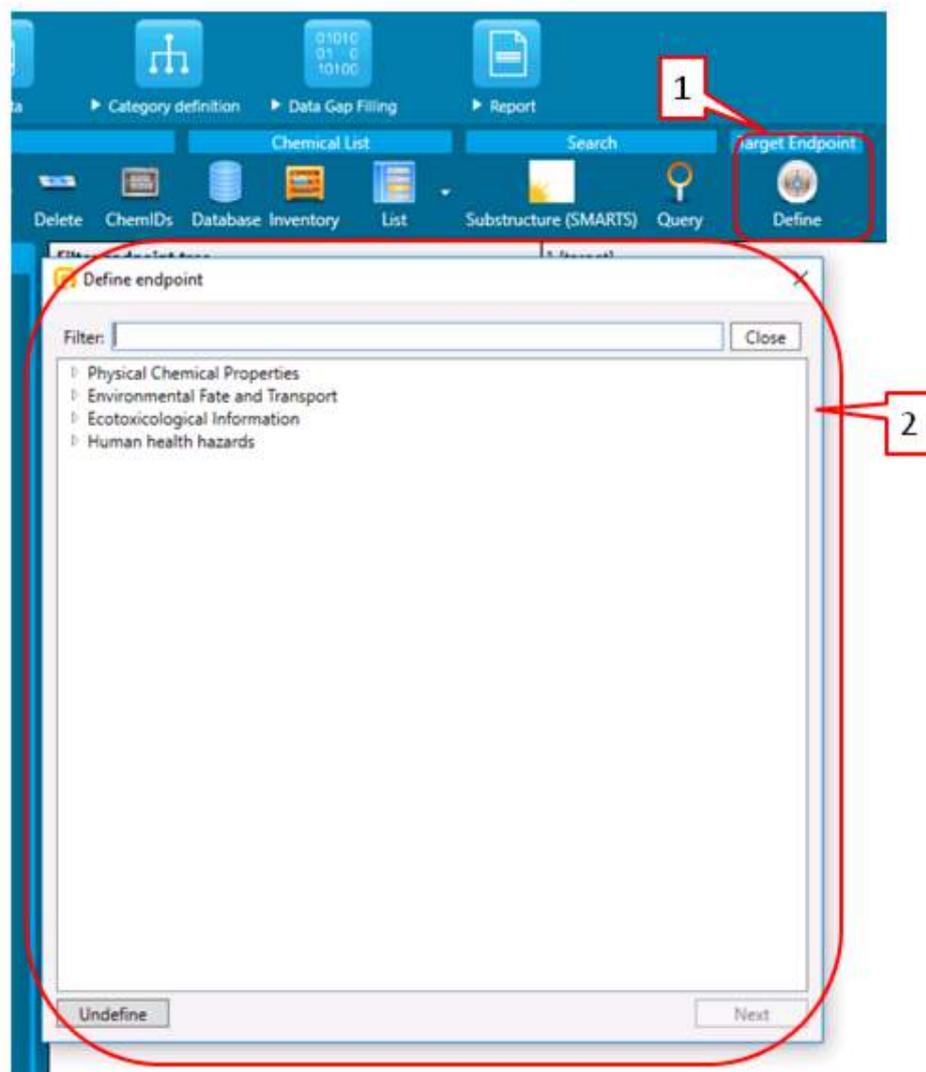


Figure 2

You can select the main level of the endpoint by expanding the tree (1) or type it in the filter field (2) and then press Next (3). (Figure 3)

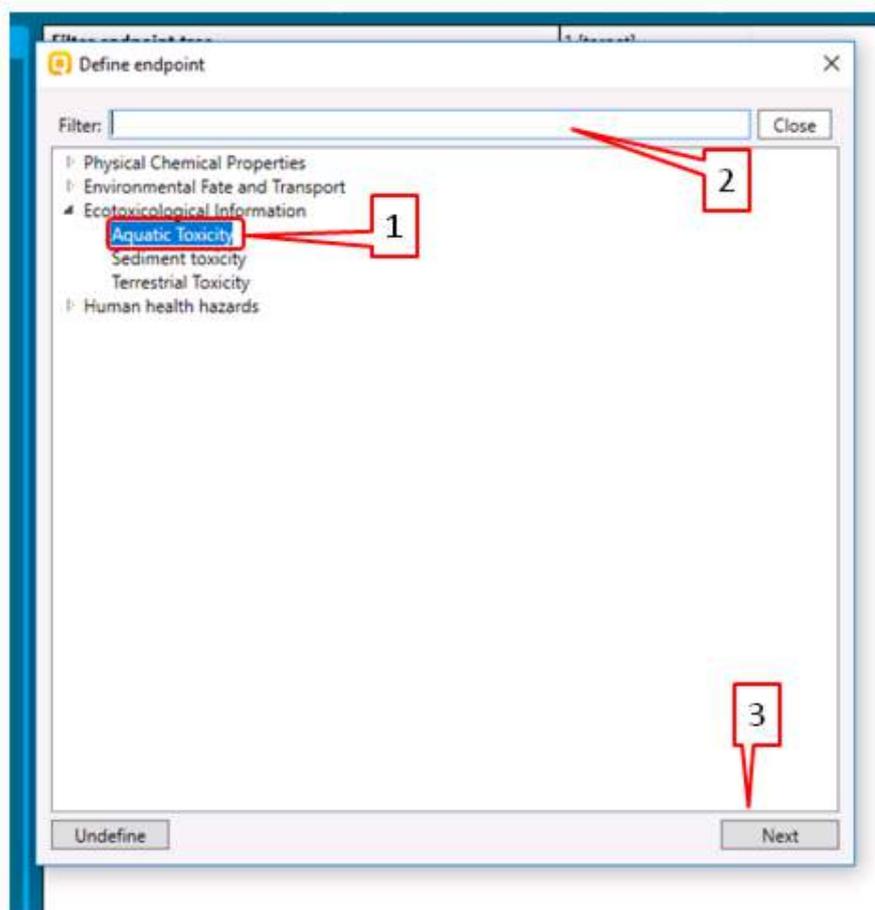


Figure 3

Additional fields for meta data information are applicable, when the user selects LC50 as the endpoint (Figure 4).

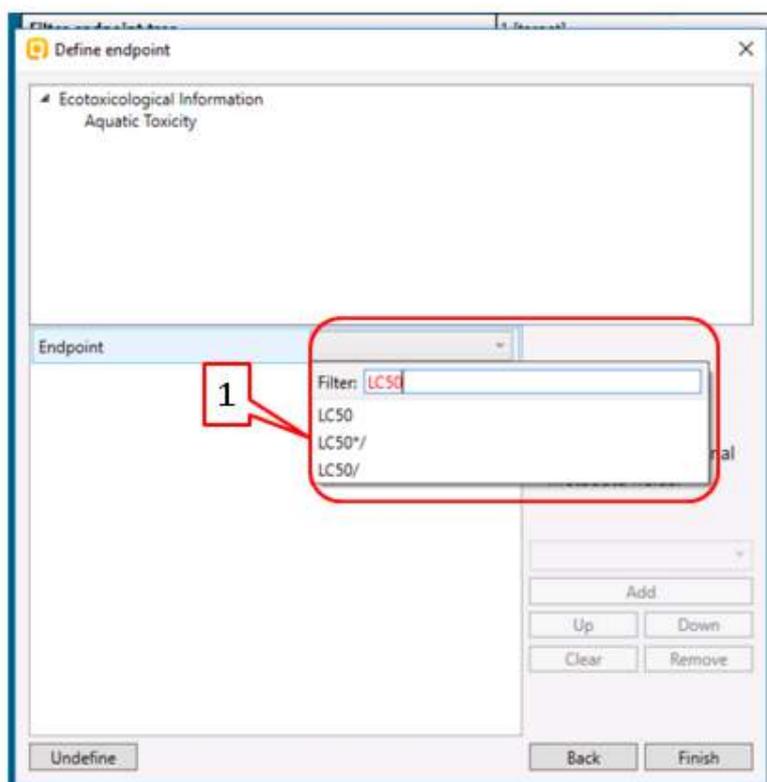


Figure 4

The following steps need to be performed in order for the meta fields to be activated (Figure 5):

- o Metadata fields are selected from the dropdown menu (1)
- o Once the metadata is selected (2) press *Add* (3).
- o From the dropdown menu select the effect, e.g. mortality or use the filter for quick search (4).

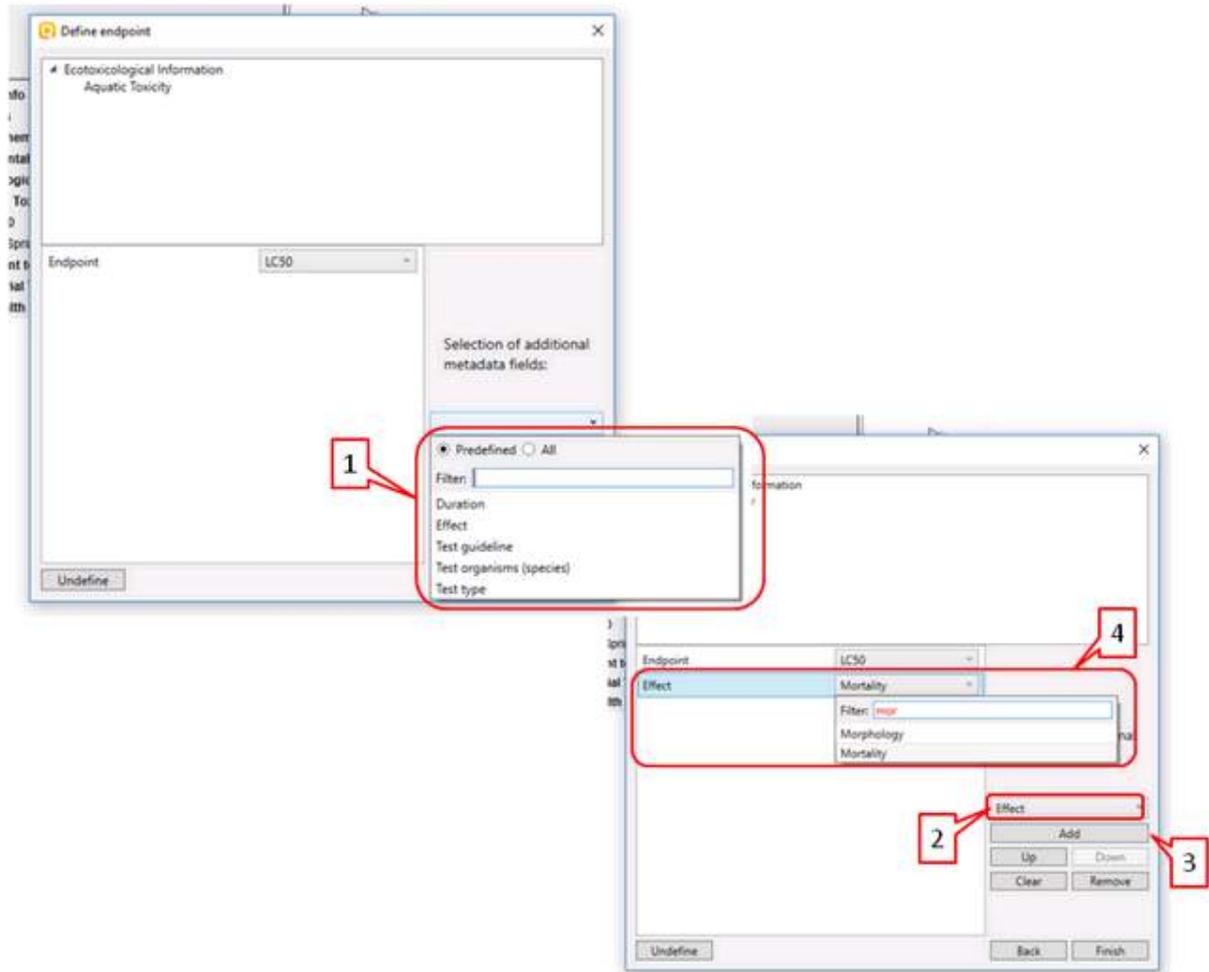


Figure 5

Return back to Metadata field and now select from the drop down list “Test organism (species)” as shown in Figure 6:

- o Select Test organism (species) from the drop down menu (1);
- o Add it as a field – click on *Add* button (2);
- o From the drop-down list with organisms loaded, use the filter and type “pime” and select Pimephales promelas (3);
- o The user can add other metadata fields in the similar way;
- o Once finished with the definition of the fields associated with the selected endpoint press Finish (4).

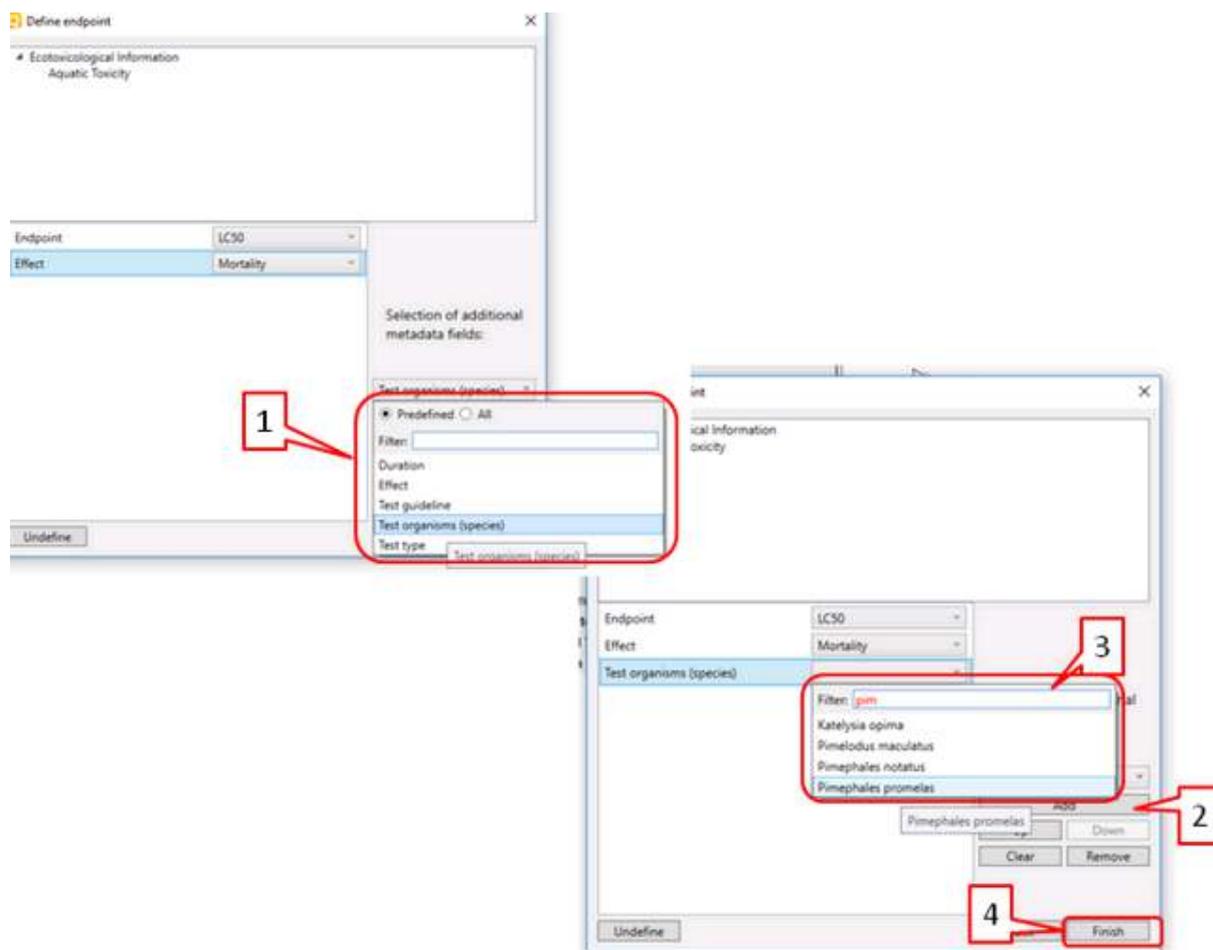


Figure 6

The endpoint tree is expanded and the selected endpoint is highlighted in yellow in the data matrix as illustrated in Figure 7.

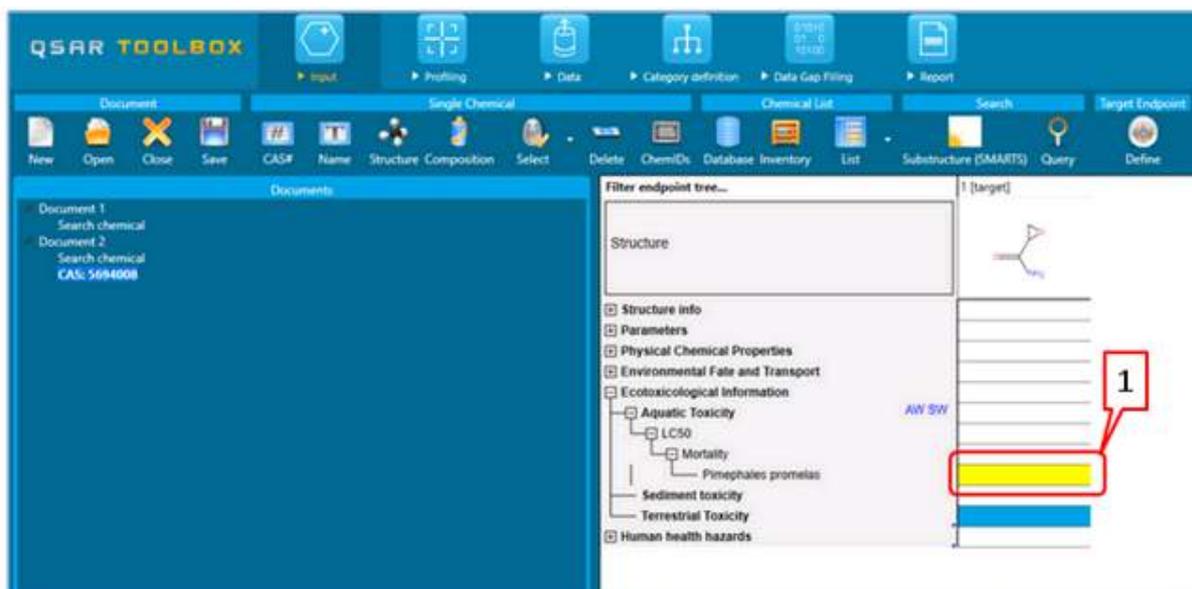


Figure 7

The user can *Undefine* the target endpoint by right-click on the end-point, then select Target endpoint (1), then *Undefine* (2) (Figure 8).

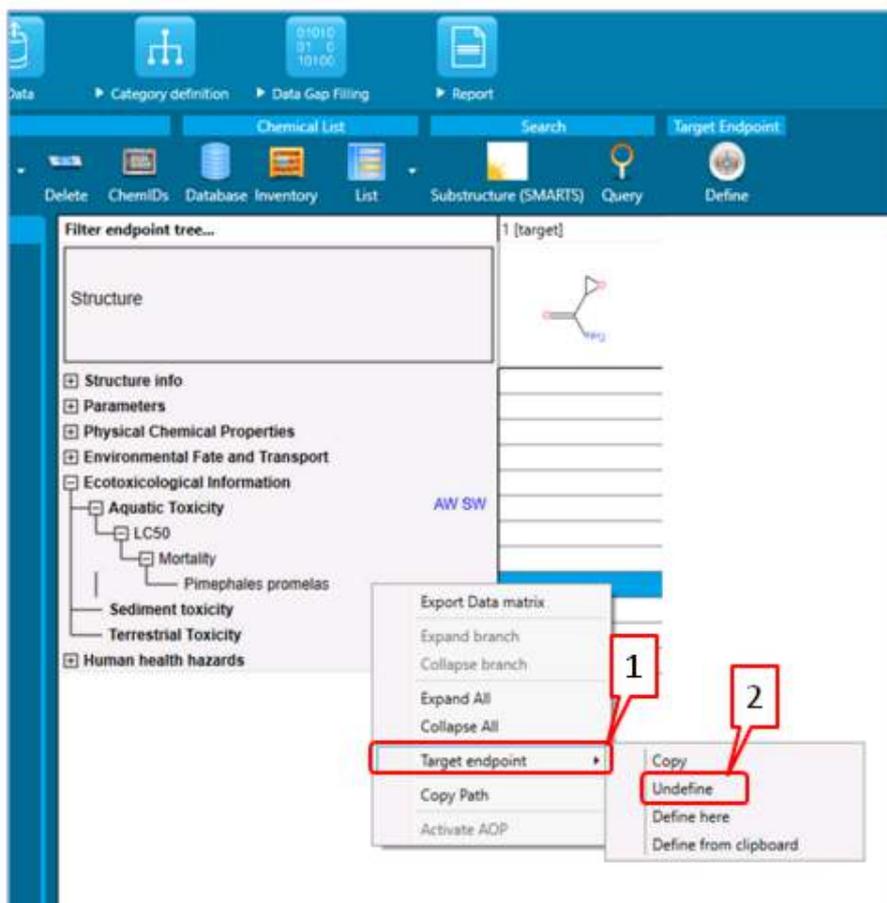


Figure 8

Alternatively, you can go back the Define button (1) the toolbar and then select *Undefine* (2) (Figure 9).

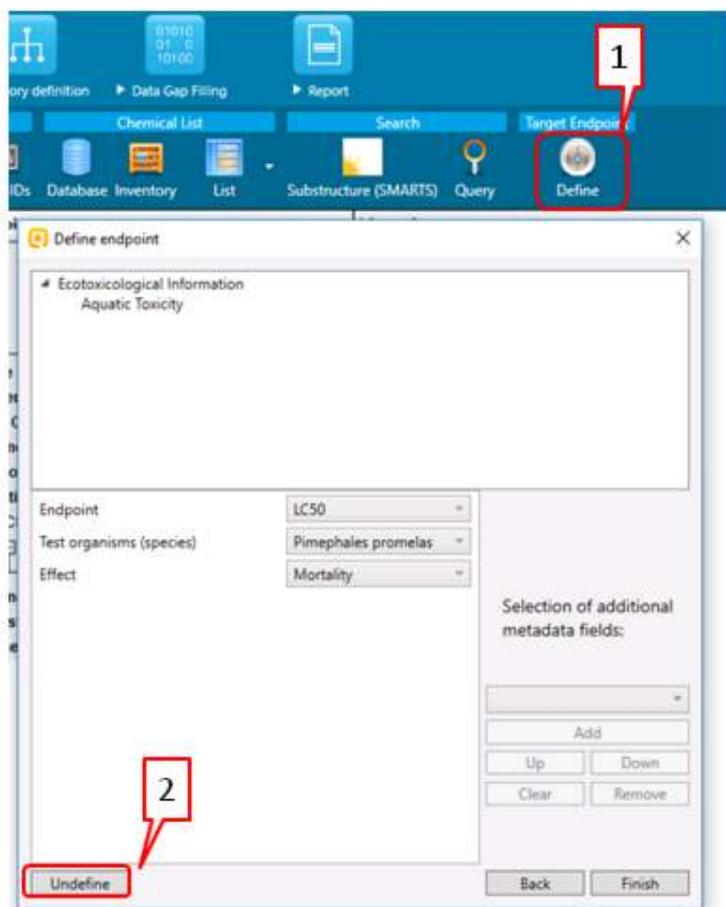


Figure 9

D.1.4.1. Gene mutation in vitro

D.1.4.1. Gene mutation in vitro

Possible combination of *in vitro Gene mutation* pathways are shown below. Those combination will results in highlighting the corresponding databases (See [Databases](#)).

Predefined	Endpoint	Type of method	Test type	Test organisms (species)	Metabolic activation	Strain	Bacterial mutagenicity ISSSTY	Genotoxicity OASIS	ECVAM Genotoxicity and Carcinogenicity	Toxicity Japan MHLW	ECHA CHEM
Genetic Toxicity			Bacterial reverse mutation assay (e.g. Ames test)				relevant DB	relevant DB	relevant DB	relevant DB	relevant DB
Genetic Toxicity	Gene mutation	In Vitro					relevant DB	relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)				relevant DB	relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Salmonella typhimurium			relevant DB	relevant DB		relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Salmonella typhimurium	Without S9		relevant DB	relevant DB		relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Salmonella typhimurium	With S9		relevant DB	relevant DB		relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Salmonella typhimurium	No S9 info			relevant DB			
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g.	Escherichia coli						relevant DB	

			Ames test)								
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)		Without S9		relevant DB	relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)		With S9		relevant DB	relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)		No S9 info			relevant DB	relevant DB		
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Undefined Test organisms (species)					relevant DB		
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)	Salmonella typhimurium		TA 100	relevant DB	relevant DB		relevant DB	
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)			TA 98	relevant DB	relevant DB			
Genetic Toxicity	Gene mutation	In Vitro	Bacterial reverse mutation assay (e.g. Ames test)			TA 98, TA 100		relevant DB			
Genetic Toxicity	Gene mutation	In Vitro	In vitro mammalian cell micronucleus test					relevant DB			
Genetic Toxicity	Gene mutation	In Vitro	Mammalian cell gene mutation assay					relevant DB	relevant DB		

Genetic Toxicity	Gene mutation	In Vitro	Mammalian cell gene mutation assay	With S9			relevant DB	relevant DB		
Genetic Toxicity	Gene mutation	In Vitro	Mammalian cell gene mutation assay		T-lymphoma cell lines		relevant DB			

-0-

D.1.4.2. Gene mutation in vivo

D.1.4.2. Gene mutation in vivo

Possible combination of *in vivo Gene mutation* pathways are shown below. Those combination will results in highlighting the corresponding databases (See [Databases](#)).

Predefined	Endpoint	Type of method	Test type	Test organisms (species)	ECVAM Genotoxicity and Carcinogenicity	Transgenic Rodent Database
Genetic Toxicity	Gene mutation	In Vivo			relevant DB	relevant DB
Genetic Toxicity	Gene mutation	In Vivo	Transgenic Rodent Mutation			relevant DB
Genetic Toxicity	Gene mutation	In Vivo	Transgenic Rodent Mutations		relevant DB	
Genetic Toxicity	Gene mutation	In Vivo		Rat		relevant DB
Genetic Toxicity	Gene mutation	In Vivo		Mouse		relevant DB

-0-

D.1.4.3. Chromosome aberration in vitro

D.1.4.3. Chromosome aberration in vitro

Possible combination of *in vitro Chromosome aberration* pathways are shown below. Those combination will results in highlighting the corresponding databases (See [Databases](#)).

Predefined	Endpoint	Type of method	Test type	Test organisms (species)	Metabolic activation	Strain	Genotoxicity OASIS	ECVAM Genotoxicity and Carcinogenicity	Toxicity Japan MHLW	ECHA CHEM
Genetic Toxicity	Chromosome aberration						relevant DB	relevant DB	relevant DB	relevant DB
Genetic Toxicity	Chromosome aberration	In Vitro					relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Chromosome aberration	In Vitro	In Vitro Mammalian Chromosome Aberration Test				relevant DB	relevant DB	relevant DB	
Genetic Toxicity	Chromosome aberration	In Vitro	In Vitro Mammalian Chromosome Aberration Test	Chinese hamster			relevant DB		relevant DB	
Genetic Toxicity	Chromosome aberration	In Vitro	In Vitro Mammalian Chromosome Aberration Test	Chinese hamster	Without S9		relevant DB			
Genetic Toxicity	Chromosome aberration	In Vitro	In Vitro Mammalian Chromosome Aberration Test	Chinese hamster	With S9		relevant DB			
Genetic Toxicity	Chromosome aberration	In Vitro	In Vitro Mammalian Chromosome Aberration Test	Chinese hamster	No S9 info		relevant DB			
Genetic Toxicity	Chromosome aberration	In Vitro			No S9 info		relevant DB	relevant DB		
Genetic Toxicity	Chromosome aberration	In Vitro	Micronucleus assay					relevant DB		

D.1.4.4. Chromosome aberration in vivo

D.1.4.4. Chromosome aberration in vivo

Possible combination of *in vivo Chromosome aberration* pathways are shown below. Those combination will results in highlighting the corresponding databases (See [Databases](#)).

Predefined	Endpoint	Type of method	Test type	Test organisms (species)	Strain	Genotoxicity OASIS	ECVAM Genotoxicity and Carcinogenicity	Toxicity Japan MHLW	Micronucleus OASIS	Micronucleus ISSMIC
Genetic Toxicity	Chromosome aberration	In Vivo				relevant DB	relevant DB	relevant DB	relevant DB	relevant DB
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay			relevant DB	relevant DB	relevant DB	relevant DB	relevant DB
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Mouse			relevant DB	relevant DB	relevant DB	relevant DB
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Rat			relevant DB		relevant DB	relevant DB
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Rat, mouse			relevant DB			
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Rat, hamster			relevant DB			
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Rabbit					relevant DB	
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Human					relevant DB	
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Mammalia					relevant DB	
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Hamster			relevant DB			

y	n									
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay	Mouse	Bone marrow cells			relevant DB		
Genetic Toxicity	Chromosome aberration	In Vivo	Micronucleus assay		Bone marrow cells			relevant DB		
Genetic Toxicity	Chromosome aberration	In Vivo	In Vivo Mammalian Chromosome Aberration Test					relevant DB		

-0-

D.1.4.5. Other endpoints

D.1.4.5. Other endpoints

Possible combination of some endpoint pathways implemented in Toolbox are shown below. Those combination will results in highlighting of the corresponding databases (See [Databases](#))

Predefined	Endpoint	Type of method	Test type	Test organisms (species)	Genotoxicity OASIS	ECVAM Genotoxicity and Carcinogenicity	Carcinogenic Potency DB (CPDB)	Carcinogenicity and mutagenicity ISSCAN	Biocides and Plant protection ISSBIOC
Genetic Toxicity	DNA and protein damage	In Vivo	Liver micronucleus		relevant DB				
Genetic Toxicity	DNA damage and repair	In Vivo	Single cell gel electrophoresis (comet) assay		relevant DB	relevant DB			
Genetic Toxicity	DNA damage and repair	In Vivo	Unscheduled DNA Synthesis			relevant DB			
Carcinogenicity	Carcinogenicity					relevant DB			
Carcinogenicity	Summary carcinogenicity						relevant DB	relevant DB	relevant DB
Carcinogenicity	Summary carcinogenicity			Mouse			relevant DB	relevant DB	relevant DB
Carcinogenicity	Summary carcinogenicity			Rat			relevant DB	relevant DB	relevant DB
Carcinogenicity	Summary carcinogenicity			Hamster			relevant DB		
Carcinogenicity	TD50						relevant DB	relevant DB	
Carcinogenicity	TD50			Mouse			relevant DB	relevant DB	
Carcinogenicity	TD50			Rat			relevant DB	relevant DB	

D.2. Profiling

D.2.1. Profiling methods

QSAR Toolbox v.4

D.2.1.1. Predefined

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.1. Predefined

D.2.1.1. Predefined

Predefined categories are set of rules for grouping chemicals developed by recognized institutions or organizations and they could be considered as “standard” categorization schemes.

The following predefined categorization schemes are implemented in the Toolbox:

- Database affiliation,
- Inventory affiliation,
- OECD HPV Chemicals categories,
- Substance type,
- US-EPA New Chemicals categories.

Short description for each of the listed profilers is provided below:

Database affiliation

This profiler provides information for database affiliation of chemicals (belonging to database). Belonging to the current database does not mean availability of experimental data. Belonging to database means that the current chemical belong to the corresponding database as a substance identity (availability of SMILES notation)

Inventory affiliation

This profiler provides information for inventory affiliation of chemicals

(belonging to inventory).

OECD HPV Chemicals categories

This profiler is developed based on the OECD list of high production volume chemicals assessed until April 2010. These are chemicals which are produced or imported at levels greater than 1,000 tones per year in at least one member country/region. The profiler helps countries to choose chemicals on which to make a hazard assessment for human health and the environment in the context of the OECD HPV Chemicals Programme.

These hazard assessments, which are a fundamental part of the OECD Existing Chemicals Programme, are based on a relatively restricted set of data elements, the Screening Information Data Set or SIDS. The intention is to screen chemicals for potential hazards, so that resources can be concentrated on undertaking further work on chemicals of concern.

Substance type

The system identifies the following substance types: # - Discrete chemicals # - Dissociating chemicals #- Mixtures# - Polymers # - Chemicals with not defined composition. According to their properties or composition substances are classified into one of these five substance types.

US-EPA New Chemicals categories

The rules coded in the US-EPA New Chemical Program profiler reproduces the original categories cited in the document "TSCA New Chemicals Program (NCP)/ Chemical Categories" - an official document of U.S EPA Office of Pollution Prevention and Toxics. However, not all of the categories have been coded because they are very broad and limits are mostly based on physical considerations and do not include structure based rules. These categories are listed bellow:

- Category: Acid Dyes and Amphoteric Dyes
- Category: Cationic Dyes
- Category: Polyanionic Polymers (& Monomers)
- Category: Polycationic Polymers
- Category: Respirable, Poorly soluble Particulates

-0-

QSAR Toolbox v.4

D.2.1.2. General mechanistic

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.2. General mechanistic

D.2.1.2. General mechanistic

Mechanistic categorization schemes consist of rules for identification of important chemical characteristics based on published knowledge. Practically they rely on certain hypothesis about the studied phenomenon. In this respect, their application is oriented for the phenomena for which they were developed.

The following general mechanistic categorization schemes are implemented in the Toolbox:

- DNA binding by OASIS
- DNA binding by OECD
- Estrogen receptor binding
- Protein binding by OASIS
- Protein binding by OECD
- Protein binding potency
- Protein binding potency Cys (DPRA 13%)
- Protein binding potency Lys (DPRA 13%)
- Biodegradation probability (Biowin 1)
- Biodegradation probability (Biowin 2)
- Biodegradation probability (Biowin 5)
- Biodegradation probability (Biowin 6)
- Biodegradation probability (Biowin 7)
- Biodegradability ultimate (Biowin 3)
- Biodegradability primary (Biowin 4)
- Ultimate Biodeg
- Hydrolysis half-life (Ka, pH7)
- Hydrolysis half-life (Ka, pH8)
- Hydrolysis half-life (Kb, pH7)
- Hydrolysis half-life (Kb, pH8)
- Ionization at pH=1
- Ionization at pH=4
- Ionization at pH=7.4
- Ionization at pH=9
- Toxic hazard classification by Cramer (original)
- Toxic hazard classification by Cramer (with extension)

Short description for each of the listed profilers is provided below:

DNA binding by OASIS v 1.4

The profiler is based on Ames Mutagenicity model part of OASIS TIMES system. The profiler consists of 85 structural alerts responsible for interaction with DNA analyzed in Ames Mutagenicity model. The scope of the profiler is to investigate presence of alerts within target molecules which may interact with DNA. The list of 85 structural alerts has been separated into eight mechanistic domains. Each of the mechanistic domains is separated into mechanistic alerts. 31 of the alerts have been updated. The profiling result assigns a target to the corresponding structural alert, mechanistic alerts and domain.

DNA binding by OECD

This report describes the development of a new profiler compiling mechanistic organic chemistry fragments (in the form of structural alerts) for the binding of organic compounds to DNA. The profiler was created following the mapping of existing structural alerts for mutagenicity and carcinogenicity. The mapping was performed to achieve maximum overlap and usability whilst restricting redundancy in the alerts, and to ensure that the alerts related to the molecular initiating event of covalent DNA binding by OECD. A total of 60 new or re-defined alerts have been created; of these all but two are supported by mechanistic information and meta data. The alerts cross six broad organic chemistry mechanisms and represent the most comprehensive listing of structural alerts for DNA binding by OECD currently available.

Estrogen receptor binding

Estrogen receptor (ER) binding is a molecular initiating event much like protein binding. It is an endpoint where several comprehensive databases exist, which has led to the development of several approaches for using (Q)SARs to predict ER binding and possible subsequent endocrine disruption. The incorporated Toolbox ER binding profiling scheme is based on structural and parametric rules extracted from literature sources and supported by experimental data. The ER-binding profiler classifies chemicals as non binders or binders depending on molecular weight (MW) and structural characteristics of the chemicals:

- Very strong binders: Chemicals with MW between 200 and 500 Da and two rings with a hydroxyl group connected to each of them.
- Strong binders: Chemicals with at least one 5-or 6-members carbon ring with an unhindered hydroxyl or amino group and MW between 200 and 500 Da;
- Moderate binders: Chemicals with at least one 5-or 6-members carbon ring with an unhindered hydroxyl or amino group and MW between 170 and 200 Da;
- Weak binders: Chemicals with at least one 5-or 6-members carbon ring with an unhindered hydroxyl or amino group and MW less than 170 Da;

If the target chemical does not meet some of the structural and parametric requirements listed above it is classified as Non binder:

- Non binder with impaired hydroxyl or amino group;
- Non binder, MW more than 500 Da;
- Non binders without hydroxyl or amino group;
- Non-binder, non-cyclic.

Protein binding by OASIS v 1.4

The protein binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry Bourgas and the partnership of Dr D.Roberts, as a part of the TIMES model to predict skin sensitisation. Under the scope of a research agreement signed in 2007 with Professor Mekenyan (OASIS - LMC), L'Oreal contributed to the assessment and refinement of chemical categories provided in the QSAR Toolbox. The scope of the profiler is to investigate presence of alerts within target molecules responsible for interaction with proteins. The list of 101 structural alerts has been separated into 11 mechanistic domains. Each of the mechanistic domains has been separated into more than 2 mechanistic alerts. The profiling result outcome assigns a target to the corresponding structural alert, mechanistic alerts and domain.

Protein binding by OECD

The Protein binding by OECD profiler was developed by an analysis of direct acting structural alerts based on theoretical organic chemistry (the profiler does not contain metabolically / abiotically activated structural alerts). The alert compilations were analysed in order to place the information contained within the literature into a mechanistic chemistry framework. This mechanistic chemistry can be used as the basis for chemical category formation when utilising the Protein binding by OECD profiler. Within each of the five mechanistic domains related structural alerts have been grouped based on the presence of a common reactivity site into so-called mechanistic alerts. Chemical category formation can be carried out at either the mechanistic alert or structural alert level using this profiler. The protein binding by OECD profiler contains 16 mechanistic alerts covering 52 structural alerts. These data are supported by mechanistic chemistry and references to the scientific literature (the meta data).

Protein binding potency

This profiler is developed on the base of empirical data for thiol reactivity expressed by the in chemico RC50 value. Data are obtained by measuring target chemical covalent binding with the thiol group of glutathione (GSH). The structural alerts for protein binding are extracted from about 400 chemicals comprised within GSH Experimental RC50. All the chemicals have two common electrophilic mechanisms of interaction with GSH – interaction via SN2 and interaction via Michael addition (MA) mechanism. The profiler

contains 49 MA and 46 SN2 categories. The set of 95 structural alerts are separated into five potency categories: Extremely, Highly, Moderately, Slightly reactive and Suspect. Classification of potency categories is as follows: extremely reactive ($RC_{50} < 0.099\text{mmol/L}$); highly reactive ($RC_{50} = 0.100 - 0.990\text{mmol/L}$); moderately reactive ($RC_{50} = 1.000 - 15.000\text{mmol/L}$); slightly reactive ($RC_{50} = 16.000 - 70.000\text{mmol/L}$); suspect ($RC_{50} = 71.000 - 135.000$). The profiling results outcome assigns a target to the corresponding potency category based on matched structural criteria.

Biodegradation probability (Biowin 1)

The linear biodegradation probability of a chemical is calculated by summing the values (fragment coefficients) of each fragment and then adding the summation to a constant coefficient. The probability values are converted into two ranges:

- A Probability Greater Than or Equal to 0.5 indicates --> Biodegrades Fast
- Probability Less Than 0.5 indicates --> Does NOT Biodegrade

Biodegradation probability (Biowin 2)

The biodeg probability (Biowin 2) profiler calculates the Non-Linear Biodegradation Probability of chemicals by summing the values (fragment coefficients) of each fragment and then adding the summation to a constant coefficient. The probability values are converted into two ranges:

- A Probability Greater Than or Equal to 0.5 indicates --> Biodegrades Fast
- Probability Less Than 0.5 indicates --> Does NOT Biodegrade

Biodeg ultimate (Biowin 3)

This model estimates the time required for "complete" ultimate biodegradation. Ultimate biodegradation is the transformation of a parent compound to carbon dioxide and water, mineral oxides of any other elements present in the test compound, and new cell material. Biowin 3 estimates the time required to achieve complete ultimate biodegradation in a typical or "evaluative" aquatic environment.

Biodeg primary (Biowin 4)

This model estimates the time required for primary biodegradation. Primary biodegradation is the transformation of a parent compound to an initial metabolite. Biowin 4 estimates the time required to achieve primary biodegradation in a typical or "evaluative" aquatic environment.

Biodegradation probability (Biowin 5)

Biodeg probability (Biowin 5) profiler is based on Biowin5 model of EPISUIT.

Biowin5 is a predictive model for assessing a compound's biodegradability in the Japanese MITI (Ministry of International Trade and Industry) ready biodegradation test; i.e. OECD 301C. Biodegradability estimates are based upon fragment constants that were developed using multiple linear regression analyses, depending on the model.

- A Probability Greater Than or Equal to 0.5 indicates --> Readily Degradable
- A Probability Less Than 0.5 indicates --> NOT Readily Degradable

Biodegradation probability (Biowin 6)

Biodeg probability (Biowin 6) profiler is based on Biowin6 model of EPISUIT.

Biowin6 is a predictive model for assessing a compound's biodegradability in the Japanese MITI (Ministry of International Trade and Industry) ready biodegradation test; i.e. OECD 301C. Biodegradability estimates are based upon fragment constants that were developed using multiple non-linear regression analyses, depending on the model.

- A Probability Greater Than or Equal to 0.5 indicates --> Readily Degradable
- A Probability Less Than 0.5 indicates --> NOT Readily Degradable

Biodegradation probability (Biowin 7)

Biodeg probability (Biowin 7) profiler is based on Biowin7 model of EPISUIT. Biowin7 estimates the probability of fast biodegradation under methanogenic anaerobic conditions; specifically, under the conditions of the "serum bottle" anaerobic biodegradation screening test. This endpoint is assumed to be predictive of degradation in a typical anaerobic digester.

- A Probability Greater Than or Equal to 0.5 indicates --> Biodegrades Fast
- A Probability Less Than 0.5 indicates --> Does NOT Biodegrade Fast

Ultimate Biodeg

The BOD to half-life profiler classifies chemicals into persistence categories based on experimental BOD data. The profiler converts experimental BOD data (%) into half-lives using first order kinetics.

The following categories have been defined:

- 0-1 day; 1-10 days; 10-100 days; >100 days
- Half-Life = $\ln(2) / (-\ln(1-\text{BOD}/100)) / \text{Day}$

Hydrolysis half-life (Ka, pH7)

The profile Ka Half-Life (pH 7), based on HYDROWIN, estimates hydrolysis half-lives using the total acid-catalyzed hydrolysis rate constants. Half-lives for acid-catalyzed hydrolysis rate constants are calculated at pH 7.

Hydrolysis half-life (Ka, pH8)

The profile Ka Half-Life (pH 8), based on HYDROWIN, estimates hydrolysis half-lives using the total acid-catalyzed hydrolysis rate constants. Half-lives for acid-catalyzed hydrolysis rate constants are calculated at pH 8.

Hydrolysis half-life (Kb, pH7)

The profile Kb Half-Life (pH 7), based on HYDROWIN, estimates hydrolysis half-lives using the total base-catalyzed hydrolysis rate constants. Half-lives for base-catalyzed hydrolysis rate constants are calculated at pH 7.

Hydrolysis half-life (Kb, pH8)

The profile Kb Half-Life (pH 8), based on HYDROWIN, estimates hydrolysis half-lives using the total base-catalyzed hydrolysis rate constants. Half-lives for base-catalyzed hydrolysis rate constants are calculated at pH 8.

Ionization at pH=1

The Ionization at pH=1 profile calculates the fractional concentration of the ionized species of a molecule (Ionization, %) at pH=1. For any given chemical the profile calculates the fractional concentrations of the species resulting from the ionization of the strongest acidic and strongest basic sites only, based on pre-calculated pKa values. The calculated fractional concentrations are used in order to classify chemicals into 10 ranges of variation of Ionization, i.e. [0,10), [10,20) ... [90,100]. The ranges are marked as Acidic or Basic, based on the nature of the ionization site and its pKa value.

Note: The profile is based on a database of calculated pKa values. It could be therefore applied only to chemicals included in the Toolbox database.

Ionization at pH=4

The Ionization at pH=4 profile calculates the fractional concentration of the ionized species of a molecule (Ionization, %) at pH=4. For any given chemical the profile calculates the fractional concentrations of the species resulting from the ionization of the strongest acidic and strongest basic sites only, based on pre-calculated pKa values. The calculated fractional concentrations are used in order to classify chemicals into 10 ranges of variation of Ionization, i.e. [0,10), [10,20) ... [90,100]. The ranges are marked as Acidic or Basic, based on the nature of the ionization site and its pKa value.

Ionization at pH=7.4

The Ionization at pH=7.4 profile calculates the fractional concentration of the ionized species of a molecule (Ionization, %) at pH=7.4. For any given chemical the profile calculates the fractional concentrations of the species resulting from the ionization of the strongest acidic and strongest basic sites

only, based on pre-calculated pKa values. The calculated fractional concentrations are used in order to classify chemicals into 10 ranges of variation of Ionization, i.e. [0,10), [10,20) ... [90,100]. The ranges are marked as Acidic or Basic, based on the nature of the ionization site and its pKa value.

Ionization at pH=9

The Ionization at pH=9 profile calculates the fractional concentration of the ionized species of a molecule (Ionization, %) at pH=9. For any given chemical the profile calculates the fractional concentrations of the species resulting from the ionization of the strongest acidic and strongest basic sites only, based on pre-calculated pKa values. The calculated fractional concentrations are used in order to classify chemicals into 10 ranges of variation of Ionization, i.e. [0,10), [10,20) ... [90,100]. The ranges are marked as Acidic or Basic, based on the nature of the ionization site and its pKa value.

Toxic hazard classification by Cramer (original)

Toxic hazard classification by Cramer profiler is build on the original paper of G.M. Cramer and R.A. Ford Estimation of toxic hazard - a decision tree approach. Food and Cosmetics Toxicology, Volume 16, Issue 6, December 1978, Page 255-276. Categorization rules classifying chemicals into different levels of toxicological concern (when administered orally) are organized in tree-like scheme. Decision tree comprises 33 structural rules and place compounds in one of the three classes:

- Low (Class I)
- Intermediate(Class II)
- High (Class III)

Two of the categories are using external files with 440 compounds in "Common component of food" and 107 compounds in "Normal constituents of body" which are borrowed by ToxTree v.2.1.0

Toxic hazard classification by Cramer (with extension)

Toxic hazard classification by Cramer profiler is build on the original paper of G.M. Cramer and R.A. Ford Estimation of toxic hazard - a decision tree approach. Food and Cosmetics Toxicology, Volume 16, Issue 6, December 1978, Page 255-276. Categorization rules classifying chemicals into different levels of toxicological concern (when administered orally) are organized in tree-like scheme. Decision tree comprises 33 structural rules and place compounds in one of the three classes:

- Low (Class I)
- Intermediate(Class II)
- High (Class III)

Two of the categories are using external files with 440 compounds in "Common component of food" and 107 compounds in "Normal constituents of body" which are borrowed by ToxTree v.2.1.0

Protein binding potency Cys (DPRA 13%)

This profile is built in relation with the implementation of the adverse outcome pathway (AOP) for skin sensitization. It is developed on the base of data derived from Direct Peptide Reactivity Assay (DPRA). The DPRA is a reactivity assay which evaluates the ability of chemicals to react with proteins. As model peptides are used reduced glutathione and two synthetic peptides – lysine and cysteine. The reaction time for both lysine and cysteine is 24 hours. The peptide reactivity is reported as percent peptide depletion. The profile contains 77 structural alerts extracted from about 229 chemicals with experimentally measured cysteine depletion values. The set of 77 structural alerts are separated into three potency categories: DPRA above 21% (DPRA 13%), DPRA less than 9% (DPRA 13%) and Grey zone 9-21% (DPRA 13%). Classification of potency categories is based on analysis published in a collaboration with L`Oreal (Dimitrov et al., 2016).

Protein binding potency Lys (DPRA 13%)

This profile is built in relation with the implementation of the adverse outcome pathway (AOP) for skin sensitization. It is developed on the base of data derived from Direct Peptide Reactivity Assay (DPRA). The DPRA is a reactivity assay which evaluates the ability of chemicals to react with proteins. As model peptides are used reduced glutathione and two synthetic peptides – lysine and cysteine. The reaction time for both lysine and cysteine is 24 hours. The peptide reactivity is reported as percent peptide depletion. The profile contains 73 structural alerts extracted from about 228 chemicals with experimentally measured lysine depletion values. The set of 73 structural alerts are separated into three potency categories: DPRA above 21% (DPRA 13%), DPRA less than 9% (DPRA 13%) and Grey zone 9-21% (DPRA 13%). Classification of potency categories is based on analysis published in a collaboration with L`Oreal (Dimitrov et al., 2016).

-0-

QSAR Toolbox v.4

D.2.1.3. Endpoint specific

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.3. Endpoint specific

D.2.1.3. Endpoint specific

The following endpoint specific categorization schemes are implemented in the Toolbox:

- Acute aquatic toxicity classification by Verhaar
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Bioaccumulation – metabolism alerts
- Bioaccumulation – metabolism half-lives
- Biodegradation fragments (BioWIN MITI)
- Carcinogenicity (genotox and nongenotox) alerts by ISS
- DNA alerts for AMES by OASIS
- DNA alerts for CA and MNT by OASIS
- Eye irritation/corrosion Exclusion rules by BfR
- Eye irritation/corrosion Inclusion rules by BfR
- in vitro mutagenicity (Ames test) alerts by ISS
- in vivo mutagenicity (Micronucleus) alerts by ISS
- Keratinocyte gene expression
- Oncologic Primary Classification
- Protein binding alerts for Chromosomal aberration by OASIS
- Protein binding alerts for skin sensitization by OASIS
- Protein Binding Potency h-CLAT
- Respiratory sensitization
- Retinoic Acid Receptor Binding
- rtER Expert System ver.1 - USEPA
- Skin irritation/corrosion (exclusion rules) by BfR
- Skin irritation/corrosion Inclusion rules by BfR

Short description for each of the listed profilers is provided below:

Acute aquatic toxicity classification by Verhaar (Modified)

The Acute aquatic toxicity classification by Verhaar consists of parametric and structural rules developed by LMC to mimic the Verhaar rules developed by Toxtree software. This classification system is based on modified Verhaar scheme part of Toxtree version 2.5. (Verhaar et al., 1992, Enoch et al., 2008). This system is introduced for chemical categorization purposes or can be used for the prioritization of chemicals for subsequent testing. Moreover, these estimates could be of great value in risk and hazard assessment. It separates a large number of small to intermediate

organic chemicals into four distinct classes that can either be assigned a mode of action, or that can otherwise be assigned quantitative relationships between the structure of the classified chemicals and their acute aquatic toxicity. These four classes are: (1) inert chemicals (baseline toxicity); (2) less inert chemicals, (3) reactive chemicals; and (4) specifically acting chemicals.

Acute aquatic toxicity MOA by OASIS

This profile divides chemicals in different categories according to their acute toxic mode of action (MOA). 2D structural information is used only to identify the MOA of chemicals. Based on theoretical and empiric knowledge the following seven hierarchically ordered MOA are distinguished: Aldehydes; alpha, beta-Unsaturated alcohols; Phenols and Anilines; Esters; Narcotic Amines; Basesurface narcotics.

Aquatic toxicity classification by ECOSAR

The Aquatic Toxicity Classification by ECOSAR profiler consists of molecular definitions developed by LMC and OECD to mimic the structural definitions of chemical classes within the U.S. Environmental Protection Agency's Ecological Structure-Activity Program (ECOSAR™). ECOSAR™ contains a library of class-based SARs for predicting aquatic toxicity, overlaid with an expert decision tree based on expert rules for selecting the appropriate chemical class for evaluation of the compound. ECOSAR™ is currently programmed to identify 118 chemical classes. The profiler is introduced for chemical categorization purposes using the class definitions from ECOSAR™ and not for predicting quantitative toxicity values.

Bioaccumulation – metabolism alerts

Bioaccumulation, as typically evaluated in fish using BCF and BAF measurements and model predictions, is the net result of competing rates of uptake and elimination in an organism. Detailed information on the model for biotransformation in fish included in the BCFBAF model is also available from the help file of the model in EPI Suite version 4.0. The final multiple-linear regression was performed on a matrix containing the number of occurrences of each fragment in each compound plus the logKow and molecular weight of each compound. The solution column of the matrix was the experimental log biotransformation half-life of each compound in days and the Appendix lists the individual fragments. The multiple-linear regression was performed with CoStat™ statistical software. The model estimates screening level whole body primary biotransformation half-lives HL, (day) and biotransformation rate constants kM, (1/day) for discrete organic chemicals in fish. The model contains a large set of unique structural fragments so that it can be broadly applicable to diverse chemical structures; however, these fragments do not reflect the entire domain of possible structural fragments for organic chemicals. A dataset of 632

experimental k_M biotransformation rates in fish (compiled in units of log biotransformation half-lives in days) was divided into a training set of 421 compounds for model derivation and a validation set of 211 compounds for model testing.

Bioaccumulation – metabolism half-lives

Bioaccumulation, as typically evaluated in fish using BCF and BAF measurements and model predictions, is the net result of competing rates of uptake and elimination in an organism. A description of the model for biotransformation in fish included in the BCFBAF model is available in the publication Arnot JA et al., (2009). Detailed information on the model for biotransformation in fish included in the BCFBAF model is also available from the help file of the model in EPI Suite version 4.0.. The model estimates screening level whole body primary biotransformation half-lives, HL, (day) and biotransformation rate constants k_M , (1/day) for discrete organic chemicals in fish. The profiler groups the chemicals into categories of biotransformation rates: very slow, slow, moderate, fast, very fast. This allows the user to identify chemicals that have a similar biotransformation rate. Cut-off values for the 5 categories are as proposed by Arnot JA et al., (2009).

Biodegradation fragments (BioWIN MITI)

BIOWIN estimates the probability of rapid aerobic and anaerobic biodegradation of an organic compound in the presence of mixed populations of environmental microorganisms. Seven separate biodegradation models are included in BIOWIN. Biodegradability estimates are based on fragment coefficients derived from linear or non-linear regression analyses depending on the model. The BIOWIN Biodegradability categorisation scheme is based on the structural fragments used by the MITI Biodegradation Probability Models.

DNA alerts for AMES by OASIS v.1.4

The profiler is based on Ames Mutagenicity model part of OASIS TIMES system. The profiler is based on the 85 structural alerts responsible for interaction of chemicals with DNA extracted from Ames Mutagenicity model. The scope of this profiler is to investigate the presence of alerts within the target molecules responsible for interaction with DNA related to Ames mutagenicity. This profiler accounts for incapability of some chemicals having an alert to interact with DNA due to electronic and steric factors. This is explicitly defined by inhibition masks associated with some alerts. The list of 85 structural alerts has been separated into eight mechanistic domains. Each of the mechanistic domains has been separated into mechanistic alerts. 31 of the alerts have been updated. The profiling result outcome assigns a

target to the corresponding structural alert, mechanistic alerts and domain.

DNA alerts for CA and MNT by OASIS v.1.1

The profiler is based on the 85 structural alerts responsible for interaction of chemicals with DNA extracted from the Chromosomal aberrations model. There is a slight difference between DNA alerts in the in vitro Ames and CA models justified by the different local training set chemicals in both models. The scope of this profiler is to investigate the presence of alerts within the target molecules responsible for interaction with DNA related to Chromosomal aberration and Micronucleus tests. This profiler accounts for incapability of some chemicals having an alert to interact with DNA due to electronic and steric factors. This is explicitly defined by inhibition masks associated with some alerts. The list of 85 structural alerts has been separated into eight mechanistic domains. Each of the mechanistic domains has been separated into mechanistic alerts. The profiling result outcome assigns a target to the corresponding structural alert, mechanistic alerts and domain.

Protein binding alerts for skin sensitization by OASIS v 1.4

The protein binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry Bourgas and the partnership of Dr D.Roberts, as a part of the TIMES model to predict skin sensitisation. The scope of this profiler is to investigate the presence of alerts within the target molecules responsible for interaction with proteins and especially with skin proteins. This profiler accounts for incapability of some chemicals having an alert to interact with skin due to electronic and steric factors. This is explicitly defined by inhibition masks associated with some alerts. The list of 100 structural alerts has been separated into 11 mechanistic domains. Each of the mechanistic domains has been separated into more than 2 mechanistic alerts. The profiling result outcome assigns a target to the corresponding structural alert, mechanistic alerts and domain.

DART scheme v.1.0

This component of The QSAR Toolbox is an adaptation of a framework for identifying chemicals with structural features associated with the potential to act as reproductive or developmental toxicants which is outlined in the following journal article: Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G, Blackburn K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. Chem Res Toxicol. 2013 Dec 16;26(12):1840-61. It was developed on the basis of the combination of known modes of action (MOA) and associated structural features, as well as an empirical association

of structural fragments within molecules of reproductive or developmental toxic (DART) chemicals when MOA information was lacking. The design of this tool is based on a detailed review of 716 chemicals (664 positive, 16 negative, and 36 with insufficient data) that have been evaluated for their DART potential. These chemicals were grouped into 25 different categories, and 129 sub-categories, based on defined receptor binding and chemical properties, and when known, their MOA. After running a chemical through the decision tree, the results will indicate that the chemical of interest is associated with chemical structures known to have DART, or that it is not associated with chemical structures known to have DART, or that it has structural features outside the chemical domain of the DART decision tree. The performance of the decision tree was tested against a group of chemicals not included in the training set, and was able to identify a high percentage of chemicals with known DART effects. The decision tree is not intended to be used as a stand-alone tool, and by design is expected to broadly capture chemicals with features that are similar to chemicals with precedent for reproductive and/ or developmental toxic effects. This tool can be used both as a component of a screening system to identify chemicals of potential concern, and as part of weight of evidence decisions based on structure-activity relationships (SAR), to fill data gaps without generating additional test data. It is expected that the structural groups described in this tool, as well as new chemicals that match these groups, could be used as a starting point for the development of hypotheses for in vitro testing to elucidate MOA and ultimately in the development of refined SAR principles for DART that incorporate MOA (adverse outcome pathways). We anticipate future refinements to the decision tree based on enhanced MOA insight that will facilitate improved discrimination of DART positive versus DART negative chemicals, with the consequent expansion of the chemical domain for which predictions with this tool can be made with confidence. We also anticipate that additional groups will be added to the tree to expand the chemical domain of coverage.

Keratinocyte gene expression

This profile is built in relation with the implementation of the adverse outcome pathway (AOP) for skin sensitization. It is developed on the base of data derived from the KeratinoSens assay, which examined the potential for chemicals to induce the expression of a luciferase reporter gene under control of a single copy of the ARE element of the human AKR1C2 gene stably inserted into immortalized human keratinocytes. Relevance to skin sensitization is inferred from the relationship of Keap1-Nrf2-ARE regulatory pathway and its detection of electrophilic chemicals to sensitization. Based on data derived from the assay three endpoints EC1.5, EC2 and EC3 (the concentrations eliciting a 1.5-, 2- and 3-fold increase in luciferase induction) are reported. No gene induction is observed when EC1.5, EC2, and /or EC3 are >2000. The profile contains 22 structural alerts extracted from about 100 chemicals comprised within EC1.5, EC2 and EC3 data. The set of 22 structural alerts are separated into four categories: very high gene

expression, high gene expression, moderate gene expression and low gene expression. Classification of categories depends on the EC3 values and is as follows: chemicals having very high gene expression ($EC3 < 15 \mu M$); high gene expression ($EC3 = 15 - 50 \mu M$); moderate gene expression ($EC3 = 50 - 100 \mu M$); low gene expression ($EC3 = 100 - 1999 \mu M$).

The profiling results outcome assigns a target to the corresponding potency category based on matched structural criteria.

Eye irritation/corrosion Exclusion rules by BfR

The exclusion rules for eye irritation/corrosion are based on physico-chemical cut-off values to identify chemicals that do not exhibit eye irritation or corrosion potential. The parameters used for defining eye irritation rules are: Lipid Solubility, Octanol Water partition coefficient, Aqueous Solubility, Melting Point and Molecular Weight.

Eye irritation/corrosion Inclusion rules by BfR

This profiler comprised of 17 structural alerts is based on empirically derived structural inclusion rule. It identifies chemicals that show potential for eye irritation and corrosion. The profiler is also based on known mechanisms of action (biochemical reaction within the eye and /or conjunctival tissues). The model is applicable to organic substances with at least 95% purity and contain hetero atoms not other O, N, S, P, Si and halogen.

Carcinogenicity (genotox and nongenotox) alerts by ISS

This profiler is an expanded and updated version of the correspondent module of the software Toxtree. It works as a decision tree for estimating carcinogenicity, based on a list of 55 structural alerts (SAs). Out of them, 35 derive from the Toxtree module and 20 are newly derived. Most of the new SAs are relative to nongenotoxic carcinogenicity, whereas the SAs in the initial list mainly coded genotoxic carcinogenicity. The SAs for carcinogenicity are molecular functional groups or substructures known to be linked to the carcinogenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognised, the system flags the potential carcinogenicity of the chemical.

in vitro mutagenicity (Ames test) alerts by ISS

This profiler is based on the Mutagenicity/Carcinogenicity module of the software Toxtree. It works as a decision tree for estimating in vitro (Ames test) mutagenicity, based on a list of 30 structural alerts (SAs). The SAs for mutagenicity are molecular functional groups or substructures known to be

linked to the mutagenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognised, the system flags the potential mutagenicity of the chemical. The present list of SAs is a subset of the original Toxtree list, obtained by eliminating the SAs for nongenotoxic carcinogenicity.

in vivo mutagenicity (Micronucleus) alerts by ISS

This profiler is based on the ToxMic rulebase of the software Toxtree. This rulebase provides a list of 35 structural alerts (SAs) for a preliminary screening of potentially in vivo mutagens. These SAs are molecular functional groups or substructures that are known to be linked to the induction of effects in the in vivo micronucleus assay. The compilation of SAs for the in vivo micronucleus assay in rodents provided here, is based on both the existing knowledge on the mechanisms of toxic action and on a structural analysis of the chemicals tested in the assay.

Retinoic acid receptor

Retinoic acid receptor (RAR) binding is a molecular initiating event. Retinoids are chemicals that are functionally similar to Vitamin A (retinol), a key endogenous chemical vital for embryonic development and adult homeostasis. In vivo retinol is oxidised to retinoic acid (RA) and is involved in the control of cell proliferation and differentiation. RA modulates gene transcription via interaction with the retinoic acid receptor (RAR) and retinoid X receptor (RXR), each of which is expressed as three isotypes (alpha, beta, and gamma). In the embryo RA regulates germ layer formation, and the organogenesis of the heart, pancreas, digits, ears, eyes and lungs. The influence of retinoids in over activation of the RAR signalling can lead to a range of developmental defects including cardiovascular, CNS, limb, ear, mandible and urogenital defects. The effects on the Wnt appear to be context (including RAR ligand) specific and differ between RAR isoforms [Yasuhara et al., 2010]. The exact mechanisms of teratogenicity are unclear, RAR signalling has a complex role in early embryonic development regulating many gene targets [Kam et al., 2012]. These categories were developed based on an investigation of the published literature to develop an adverse outcome pathway for RAR agonism [under development], and in vitro test data from the cited publications for each category. Active analogues which fall outside the category are assumed to act via a different binding mechanism.

Respiratory sensitization

This profiler is intended to be used for the assessment of respiratory sensitisation potential of low molecular weight chemicals. The profiler has been developed from mechanistic knowledge of the elicitation phase of respiratory sensitisation, thus identifies chemicals able to covalently bind to

proteins in the lung. This mechanistic hypothesis makes the profiler suitable for identifying chemicals capable of inducing respiratory sensitisation via both the skin and lung (as the chemistry (for a given structural alert) must be the same in both the induction and elicitation phases of sensitisation).

Protein binding alerts for Chromosomal aberration v 1.2

The profiler is based on 33 structural alerts accounting for interactions of chemicals with specific proteins, such as topoisomerases, cellular protein adducts, etc. Associated with clear mechanistic justification, these alerts are included as a second reactivity component (complementing DNA reactivity) in the in vitro Chromosomal aberrations OASIS TIMES mutagenicity model. The scope of this profiler is to investigate the ability of target molecules to elicit clastogenicity. Functionalities which bring about steric (or electronic) hindrance in molecules and thus impede interactions with proteins are explicitly defined and associated with some of the alerts as "inhibition" masks.

Oncologic Primary Classification

The OncoLogic Primary Classifier consists of molecular definitions developed by LMC and OECD to mimic the structural criteria of chemical classes of potential carcinogens covered by the U.S. Environmental Protection Agency's OncoLogic™ Cancer Expert System for Predicting the Carcinogenicity Potential. In the QSAR Toolbox, the OncoLogic Primary Classifier is used solely for the purpose of categorization based on the definition of an OncoLogic™ class. The profiler is introduced for categorization purpose and not for predicting carcinogenicity.

rtER Expert System ver.1 - USEPA

The rtER Expert System ver.1 – USEPA profiler consists of molecular definitions mimic the structural criteria of chemical classes potential estrogen receptor-binders covered by US EPA Estrogen Receptor Expert System (ERES) The ERES profiler is an effects-based automated system used to predict estrogen receptor binding affinity. The Estrogen Receptor Expert System (ERES) Profiler is an effects-based automated system used to predict estrogen receptor binding affinity. The ERES was originally developed to address a defined regulatory purpose, specifically for prioritizing chemicals from two specific inventories, food use pesticidal inerts (FI) and antimicrobial pesticides (AM) which do not include any chemicals with steroidal-type chemical structures, and thus not capable of higher affinity ER interactions. This system was built upon a training set of chemicals to cover the defined regulatory inventories using in vitro assays specifically optimized to pick up any indication of binding by testing up to chemical solubility or cytotoxicity within the assays to increase confidence that a chemical predicted negative is unlikely to bind ER. A chemical

class-based approach was designed to allow extrapolating from a limited number of well-characterized TrSet chemicals to a broader inventory of chemicals by employing effects-based chemical category and read-across concepts. The ERES is a logic rule-based decision tree that encodes the experts' mechanistic understanding with respect to both the chemical and biological aspects of the well-defined endpoint, or the ER bioassay domain. The transparency (relationship of predicted chemicals to tested chemicals) and usefulness of the system for the intended purpose (predictions provided for FI and AM chemicals) was emphasized in the approach to develop the ERES. For example, the relationship between relative binding affinity (RBA) and LogKow that was identified for the ERES chemical groups was used within each group to ensure the predicted chemical was bounded by TrSet chemicals. Chemicals falling outside the boundaries of known ability to predict (whether active or inactive) were considered to have "Unknown Binding Potential" (UnkBP). The automated version of the ERES enables users to compare the predicted chemical to TrSet chemicals within each chemical group (i.e., the decision tree node). In the Toolbox, the rtER Expert System ver.1 – USEPA profiler is used for the purpose of categorization based on the structural definitions of the original ERES chemical classes. The rtER Expert System ver.1 – USEPA profiler is introduced for categorization purpose and not for predicting relative binding affinity (RBA).

Skin irritation/corrosion Exclusion rules by BfR

The exclusion rules for skin irritation/corrosion are based on physicochemical cut-off values to identify chemicals that do not exhibit skin irritation or corrosion potential. 1,358 chemicals from the BfR New Chemicals database are used to derive the physicochemical limit rules. These rules are divided into rules applicable to all substances and applicable to specific chemical subclasses (more information is available on the Basic mode of the profiler). The parameters used for defining skin irritation rules are: Lipid Solubility, Surface tension, Octanol Water partition coefficient, Vapour pressure, Aqueous Solubility, Melting Point and Molecular Weight.

Skin irritation/corrosion Inclusion rules by BfR

The profiler contains structural alerts which can be used for positive classification of chemicals causing irritation, corrosion or the combination irritation/corrosion depending on their mechanisms. These 40 organic structural fragments cover the inclusion rules for skin irritation and corrosion as for now.

QSAR Toolbox v.4

D.2.1.4. Empiric

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.4. Empiric

D.2.1.4. Empiric

Empiric categorization schemes include rules used to determine the chemical elements constituting the target chemical, its chemical functionality and structural similarity.

The following empiric categorization schemes are implemented in the Toolbox:

- Chemical elements,
- Groups of elements,
- Lipinski Rule Oasis,
- Organic functional groups,
- Organic functional groups (nested),
- Organic functional groups (US-EPA),
- Organic functional groups, Norbert Haider (checkmol)
- Structure similarity
- Tautomers unstable.

Short description for each of the listed profilers is provided below:

Chemical elements

This profiler contains all chemical elements from Periodic table organized in eighteen groups. Group 1 and 2 contains Alkali and Alkaline earth chemicals. Actinoids and Lanthanoids are presented in two separate Group 3 based on their position in Periodic table. Transition metals stored in Periodic table from 3 to 12 groups are organized in groups in the same way as in the Periodic table group numbers. Metalloids and Metals are presented in Group 13. Carbon along with Sn, Pb and metalloids Si and Ge are located in Group 14. Metals and Metalloids from Group 15 of Periodic table are included in three separate categories. Oxygen, sulfur and Selenium are organized in three separate categories named Group 16. Halogen atoms (F, Cl, Br, I, At) are located in Group 17. Noble gases (He, Ne, Ar, Kr, Xn and Rn) are included in Group 18.

Groups of elements

This profiler contains all chemical elements from Periodic table organized in 9 Categories:

- Alkali and Alkaline Earth chemicals are organized in first two categories.
- Halogen atoms (F, Cl, Br, I, At) are presented in "Halogen" category.
- Metalloids elements Boron (B), Silicon (Si), Germanium (Ge), Arsenic (As), Antimony (Sb), Tellurium (Te), Polonium (Po) are located in category "Metalloids".
- Metals and Non-Metals elements are located in two separate categories.
- Actinoids and Lanthanoids are stored in category "Rare Earth".
- Noble gases (He, Ne, Ar, Kr, Xn and Rn) are included in Category "Noble gases".

All transition metals are organized in one separate category.

Lipinski Rule Oasis

Lipinski's Rule of Five is a rule of thumb to evaluate drug likeness, or determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans. The rule was formulated by Christopher A. Lipinski in 1997, based on the observation that most medication drugs are relatively small and lipophilic molecules. The rule describes molecular properties important for a drug's pharmacokinetics in the human body, including their absorption, distribution, metabolism, and excretion ("ADME"). However, the rule does not predict if a compound is pharmacologically active. The rule is important for drug development where a pharmacologically active lead structure is optimized step-wise for increased activity and selectivity, as well as drug-like properties as described by Lipinski's rule. The modification of the molecular structure often leads to drugs with higher molecular weight, more rings, more rotatable bonds, and a higher lipophilicity.

Organic functional groups

The Organic Functional Groups (OFG) system is designed in order to introduce some classification and systematization of the various functionalities and characteristic structural fragments in organic chemicals from a large database, and identify structurally similar chemicals. Organic functional groups are specific groups of atoms and/or bonds within molecules that determine the characteristic chemical reactions of those molecules. Functional groups are essential part of the characteristic structural fragments of organic molecules. One functional group may change into another one, or a functional group might react with a separate molecule to build up a larger structure. The relative reactivity of a given functional group can be modified by a nearby located other functional groups. Thus functional groups are the principal "reacting units" in organic chemistry, determining the fundamental relationship between structure and

chemical reactivity. The OFG profile is currently subdivided into OFG (general) and OFG (nested), which include the same functional groups. The difference is that OFG (general) displays all functional groups present in the target compounds, while the OFG (nested) do not show the functional groups which are nested in the larger ones (i.e. aldehyde group is nested in carboxylic group and it is not shown with OFG (nested) profiler). Thus any "overlapping" groups in the OFG (nested) system actually form new functional groups.

Organic functional groups (nested)

The Organic Functional Groups (OFG) system is designed in order to introduce some classification and systematization of the various functionalities and characteristic structural fragments in organic chemicals from a large database, and identify structurally similar chemicals. Organic functional groups are specific groups of atoms and/or bonds within molecules that determine the characteristic chemical reactions of those molecules. Functional groups are essential part of the characteristic structural fragments of organic molecules. One functional group may change into another one, or a functional group might react with a separate molecule to build up a larger structure. The relative reactivity of a given functional group can be modified by a nearby located other functional groups. Thus functional groups are the principal "reacting units" in organic chemistry, determining the fundamental relationship between structure and chemical reactivity. The OFG profile is currently subdivided into OFG (general) and OFG (nested), which include the same functional groups. The difference is that OFG (general) displays all functional groups present in the target compounds, while the OFG (nested) do not show the functional groups which are nested in the larger ones (i.e. aldehyde group is nested in carboxylic group and it is not shown with OFG (nested) profiler). Thus any "overlapping" groups in the OFG (nested) system actually form new functional groups.

Organic functional groups (US-EPA)

The 645 structural fragments and correction factors used in the enhanced (Organic Functional Groups (US EPA). Profiler is derived from the U.S. Environmental Protection Agency's KOWWIN fragment library of the EPISuite program. In addition to the base library, additional fragments were added to the atom-fragment set to characterize specific larger ring systems. In general, each non-hydrogen atom (e.g. carbon, nitrogen, oxygen, sulfur, etc.) in a structure is a "core" for a fragment; the exact fragment is determined by what is connected to the atom. Several functional groups are treated as core "atoms"; these include carbonyl (C=O), thiocarbonyl (C=S), nitro (-NO₂), nitrate (ONO₂), cyano (-C/N), and isothiocyanate (-N=C=S). Connections to each core "atom" are either general or specific; specific connections take precedence over general connections. For example, aromatic carbon, aromatic oxygen and aromatic sulfur atoms have nothing but general connections; i.e., the fragment is the same no matter what is connected to the atom. In contrast, there are 5 aromatic nitrogen

fragments: (a) in a five-member ring, (b) in a six-member ring, (c) if the nitrogen is an oxide-type {i.e. pyridine oxide}, (d) if the nitrogen has a fused ring location {i.e. indolizine}, and (e) if the nitrogen has a +5 valence {i.e. N-methyl pyridinium iodide}; since the oxide-type is most specific, it takes precedence over the other four. The aliphatic carbon atom is another example; it does not matter what is connected to -CH₃, -CH₂-, or -CH< , the fragment is the same; however, an aliphatic carbon with no hydrogens has two possible fragments: (a) if there are four single bonds with 3 or more carbon connections and (b) any other not meeting the first criteria. More information on the development of the U.S. Environmental Protection Agency's KOWWIN fragment library along with a list of the fragments, please refer to the EPISuite help menu. The program can be downloaded from the U.S. Environmental Protection Agency's website at: <http://www.epa.gov/oppt/exposure/pubs/episuite.htm>. See the Meylan, W.M. and P.H. Howard. 1995. Atom/fragment contribution method for estimating octanol-water partition coefficients. J. Pharm. Sci. 84: 83-92. Journal article for a more complete description of KOWWIN methodology. The KOWWIN library is also used in the U.S. Environmental Protection Agency's Analog Identification Methodology (AIM) as the basis for the initial atom-fragment structural parsing activity. The Analog Identification Methodology (AIM) is a web-based tool that can be accessed at: <http://aim.epa.gov>.

Organic functional groups, Norbert Haider (checkmol)

The Profile was created on the base of 204 organic functional groups recognized by "Checkmol" program. It analyzes the input molecules for presence of functional group, depending on it the molecules are classified in different categories.

Structure similarity

The molecular similarity is context defined notion, i.e., it depends on the endpoint molecular are compared for. For example, two molecules could be similar with respect to AMES mutagenicity (having same functional groups which could damage DNA) but could be dissimilar with respect to acute aquatic toxicity (not having same functionalities damaging proteins/lipids or same mode of action). Such (dis)similarity is called mechanistic and it is different of the similarity based on abstract molecular features not directly related to functionalities conditioning the interactions with macromolecules (eventually associated with molecular endpoints). The latter is called structural similarity and it has more general character as compared with the mechanistic similarity. The structural similarity is often used in QSAR studies. Also structural similarity could be used as a categorization method used for identifying any structurally similar chemicals based on molecular and atoms characteristics.

Tautomers unstable

The Unstable tautomer profiler is developed on the base of the available data and theoretical calculations for tautomer forms in water and gas

phase. The unstable tautomeric forms are presented as individual categories. The profiler is very useful for Read - Across.

-o-

QSAR Toolbox v.4

D.2.1.5. Toxicological

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.5. Toxicological

D.2.1.5. Toxicological

A repeated dose profiler used to identify the toxicological profiler of chemicals. The profiler contains boundaries based on repeated dose toxicity test data extracted from database of Hazard Evaluation Support System (HESS). the profiler is developed by NITE, METI (Japan) in cooperation with LMC

- Repeated dose (HESS)

Short description for the RDT profiler is provided below:

Repeated dose toxicity (HESS)

The profiler contains category boundaries to be expected to induce similar toxicological effects in repeated dose oral toxicity. These category boundaries were developed based on repeated dose toxicity test data in the database of Hazard Evaluation Support System (HESS). Justification for each category (mechanistic or empirical information) is described.

-0-

QSAR Toolbox v.4

D.2.1.6. Custom

[Workflow](#) / [Profiling](#) / [Profiling methods](#) / D.2.1.6. Custom

D.2.1.6. Custom

Practically, these are the profiling schemes created by the user. In any case, while creating a new profiler the user may choose to place it in any of the above sections.

Example Prioritization Scheme (PBT) is included as a custom profiler. This scheme identifies the persistence, bioaccumulation and toxicity of the target chemical(s) based on their experimental and/or predicted data.

- Example Prioritization Scheme (PBT)

-0-

D.2.2. Metabolism

QSAR Toolbox v.4

D.2.2.1. Documented metabolism

[Workflow](#) / [Profiling](#) / [Metabolism](#) / D.2.2.1. Documented metabolism

D.2.2.1. Documented metabolism

Three databases of collected 400 observed metabolic pathways in microorganisms and mammals are implemented in the Toolbox.

- Observed microbial catabolism

Degradation pathways used by microorganism to obtain carbon and energy from 551 chemicals are stored in a special file format that allows easy computer access to catabolic information. The collection includes the catabolism of C1-compounds, aliphatic hydrocarbons, alicyclic rings, furans, halogenated hydrocarbons, aromatic hydrocarbons and haloaromatics, amines, sulfonates, nitrates, nitro-derivatives, nitriles, and compounds containing more than one functional group. Most of pathways are related to aerobic conditions. Different sources including monographs, scientific articles and public web sites such as the UM-BBD (L.B.M. Ellis, D. Roe, L.P. Wackett. *Nucleic Acids Res.*, 34, D517 (2006); <http://umbbd.msi.umn.edu/>) were used to compile the database.

- Observed mammalian metabolism

Metabolic pathways documented for 100 chemicals with 630 studies in different mammals are stored in a database format that allows easy computer access to metabolic information. The collection includes chemicals with variety of functionalities, aliphatic amines, alkyl and aryl halides, ethers, esters, carbamates, carboxylic acid esters and multifunctional compounds. In vivo and in vitro studies were used to analyze the metabolic fate of chemicals. Metabolic maps with the in vivo studies are predominantly for the collection of studies (347 studies included in 49 maps). Microsomes prevail over the other experimental systems included in the in vitro studies (around 50% of studies). Around 50% of administration routes included the in vivo studies refer to oral route of administration. Different sources including monographs, scientific articles and public web sites were used to compile the database.

- Observed Rat In vivo metabolism

The observed (documented) metabolic pathways for 647 chemicals, extracted from the scientific literature, and associated with the in vivo biotransformations of xenobiotic chemicals in rodents (mostly rats) are stored in a database format that allows easy computer access to the metabolism information. This database includes structurally different chemicals of various functionalities such as aliphatic and aromatic hydrocarbons, halogenated hydrocarbons, alcohols and phenols, carbonyl compounds, carboxylic acids and esters, nitro compounds, amines, organic sulfides, heterocyclic and, mostly, multi-functional chemicals. Fields of applications include industrial chemicals, solvents, monomers, pharmaceuticals, pesticides, some phytochemicals, azo chemicals, etc. Different literature sources, including predominantly scientific publications, monographs, and public websites were used to extract the metabolism information and compile the database.

- Observed Rat Liver S9 metabolism

The documented metabolic pathways for 261 chemicals observed with the use of in vitro experimental systems such as rodent (mostly rat) liver microsomes and S9 fraction are stored in a database format that allows easy computer access to the metabolism information. This database includes structurally different chemicals of various functionalities and fields of application such as aliphatic and aromatic hydrocarbons, halogenated hydrocarbons, carboxylic acids and esters nitro compounds, amines, heterocyclic and multi-functional chemicals, etc. Different literature sources, including mostly scientific publications, monographs, and public web sites were used to extract the metabolism information and compile the database.

- Observed rat liver metabolism with quantitative data

"Implemented quantitative metabolic information and related functionalities" aim to expand and facilitate usage of metabolic information in Toolbox by adding into account quantitative metabolic data. With respect to this task a new database with quantitative metabolic data was created and included in the standard Toolbox installation. Toolbox user interface allows showing details for this kind of information. Also quantitative metabolic data could be used for grouping and filtering of analogues.

After profiling, the user can see the observed metabolic map by the right click options (1) (Figure 1). The label "QTY" (2) indicates that there are some quantitative data for the target/metabolite. Information for the study(ies) extracted from the literature is positioned on the right panel (3). The label, consists of numbers, i.e. "1.14.14.1" in this case, (4) indicate enzymatic information, which could be seen in METAPATH platform.

If the user wants to see the quantities of the parent and/or metabolites,

they have to click on the drop down box over one of the available studies. Automatically the respective metabolite quantities are displayed in tabular manner in the bottom of the window (5) As can be seen quantities of the target with respect to time decreases (e.g. disappearance of the target).

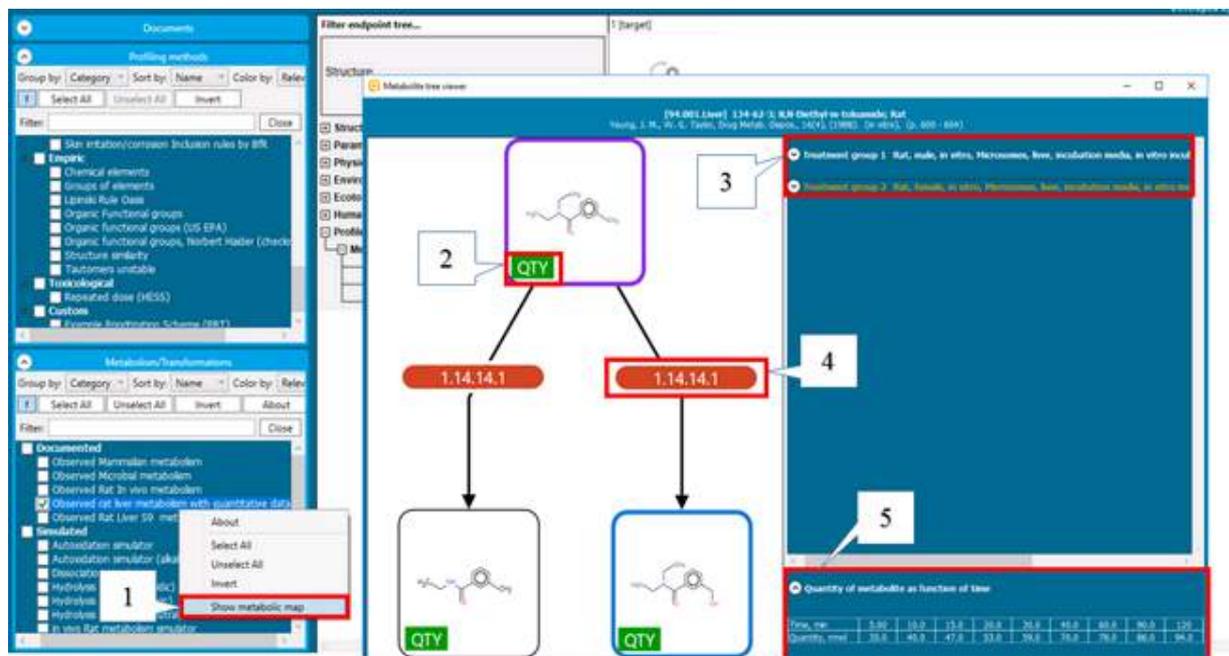


Figure 1

D.2.2.2. Simulated metabolism

[Workflow](#) / [Profiling](#) / [Metabolism](#) / D.2.2.2. Simulated metabolism

D.2.2.2. Simulated metabolism

Simulators of molecular transformations imitating microbial, liver and skin metabolism, as well as abiotic hydroxylation are also implemented in the system. Below a short explanation for each of the simulators is presented:

- Autoxidation simulator

Autoxidation (AU) is a spontaneous, air-induced oxidation of organic molecules. It is a free-radical chain reaction of a chemical with molecular oxygen, resulting in the formation of oxidation products. Among the latter, organic hydroperoxides are regarded as the most important with respect to eliciting adverse effects such as contact allergy. An AU model was therefore developed to simulate the observed AU pathways. To this purpose, a training set of 133 chemicals (terpenes, simple aliphatic and polyethyleneglycol ethers, aldehydes, aminophenols) with published data relative to AU pathways. Data consistency was maintained by collecting data within the following test conditions: air or oxygen exposure, room temperature, atmospheric pressure, AU in bulk or in the presence of different solvents, nearly neutral (pH 7 - 7,5) or slightly alkaline (pH 8 - 9), medium, duration of AU from a few hours to several months.

- Autoxidation simulator (alkaline medium)

Training set of 133 chemicals (terpenes, simple aliphatic and polyethyleneglycol ethers, aldehydes, aminophenols) with published data relative to AU pathways. Data consistency was maintained by collecting data within the following test conditions: air or oxygen exposure, room temperature, atmospheric pressure, AU in bulk or in the presence of different solvents alkaline (pH 10.2-11.5), medium, duration of AU from a few hours to several months. The AU simulator consists of a set of molecular transformations, extracted from the observed AU pathways.

- Dissociation simulator

- Hydrolysis (Acidic) simulator

Hydrolysis (Acidic) simulator predicts hydrolysis products of discrete organic chemicals under the following experimental conditions: acidic pH, ambient or room temperature and atmospheric pressure. The simulator was developed based on data collected from various sources, including articles

and public web sites. The following classes of chemicals are included in the model : epoxides, aziridines, esters, carbamates, halomethanes, selected alkyl halides, anhydrides, dithiocarbamates, isocyanates, isothiocyanates, sulfonyl chloride, lactones, nitriles, amides, N-halamines, carbamates, diketenes, organic peroxides, etc.

- Hydrolysis (Basic) simulator

Hydrolysis (Basic) simulator predicts hydrolysis products of discrete organic chemicals under the following experimental conditions: basic pH, ambient or room temperature and atmospheric pressure. The simulator was developed based on data collected from various sources, including articles and public web sites. The following classes of chemicals are included in the model : sulfonyl halides, organophosphorus compounds, epoxides, aziridines, esters, carbamates, halomethanes, selected alkyl halides, anhydrides, dithiocarbamates, isocyanates, isothiocyanates, sulfonyl chloride, lactones, nitriles, amides, N-halamines, carbamates, diketenes, organic peroxides, etc.

- Hydrolysis (Neutral) simulator

Hydrolysis simulator is an abiotic model and predicts the hydrolysis products of chemicals and their quantities at 28 day under neutral or nearly neutral pH. The model predicts hydrolysis products of discrete organic chemicals under the following experimental conditions: neutral or nearly neutral pH, ambient or room temperature and atmospheric pressure. The simulator was developed based on data collected from various sources, including articles and public web sites. The following classes of chemicals are included in the model : discrete organic chemicals, epoxides, aziridines, esters, carbamates, halomethanes, selected alkyl halides, anhydrides, dithiocarbamates, isocyanates, isothiocyanates, sulfonyl chloride, lactones, nitriles, amides, N-halamines, carbamates, diketenes, organic peroxides, etc.

- In vivo Rat metabolism simulator

The current in vivo rat liver metabolic simulator (transformation table) represents electronically designed set of 609 structurally generalized, hierarchically arranged abiotic and enzymatic transformation reactions, which are characteristic for the metabolism for in vivo experimental systems such as rodent (mostly rat). The principal applicability of this simulator is associated with the reproduction as well as the prediction of the metabolic activation reactions and pathways of xenobiotic chemicals, which may elicit in vitro genotoxicity effects such as bacterial mutagenicity and chromosomal aberrations. Each transformation in simulator consists of source and product structural fragments, and inhibiting "masks". A probability of occurrence is

ascribed to each principal transformation, which determines its hierarchy in the transformation list. A training set of xenobiotic chemicals of a wide structural diversity, with experimentally observed metabolic reactions and pathways has been built, using published data on their liver metabolism. The data on their metabolism are collected mostly from research publications in the field from selected scientific journals, monographs and websites, and are associated with the commonly observed in vivo liver metabolic reactions of chemicals with different structures. The molecular transformations set consists partly of 26 abiotic and 583 enzyme-controlled reactions believed to occur at a very high rate as compared to the duration of the tests, and the highest priority is assigned to these reactions. This subset of reactions includes also transformations of highly-reactive functional groups and intermediates, such as tautomerizations, arene epoxide rearrangements to phenols, etc. On the whole, the simulator contains also 479 enzymatic phase I transformations, such as aliphatic C-oxidation, aromatic C-hydroxylation, oxidative N- and O-dealkylation, epoxidation, ester and amide hydrolysis, carbonyl group reduction, nitro and azo group reduction, N-hydroxylation, oxidative deamination, beta-oxidation, ring cleavage, hydrolytic cleavage, aromatization, decarboxylation, dehalogenation, etc. Additionally, 104 enzymatic phase II transformations, such as glucuronidation, sulfation, glutathione conjugation, N-acetylation, etc. which, unlike the in vitro systems, are believed to occur with high priority in vivo.

- Microbial metabolism simulator

The original CATABOL simulator of microbial metabolism is implemented in the system (J. Jaworska, S. Dimitrov, N. Nikolova, O. Mekenyan. SAR QSAR Environ. Res., 13, 307 (2002); S. Dimitrov, R Breton, D. Mackdonald, J. Walker, O. Mekenyan. SAR QSAR Environ. Res., 13, 445 (2002); S. Dimitrov, V. Kamenska, J.D. Walker, W. Windle, R. Purdy, M. Lewis, O. Mekenyan. SAR QSAR Environ. Res., 15, 69 (2004).). Multiple pathway catabolism is simulated using the abiotic and enzyme-mediated reactions via the hierarchically ordered principal molecular transformations extracted from documented metabolic pathway database. The hierarchy of the transformations is used to control the propagation of the catabolic maps of the chemicals. The simulation starts with the search for match between the parent molecule and the source fragment associated with the transformation having the highest hierarchy. If the match is not found search is performed with the next transformation, etc. When the match is identified, the transformation products are generated. The procedure is repeated for the newly-formed products. Predictability (probability that the metabolite is observed, given that the metabolite is predicted) evaluated on the bases of documented catabolism for 200 chemicals stored in the database of "Observed microbial catabolism" is 83%.

- Rat liver S9 metabolism simulator

The current in vitro rat liver metabolic simulator (transformation table) represents electronically designed set of 509 structurally generalized, hierarchically arranged biotransformation reactions, which are characteristic for the metabolism for in vitro experimental systems such as rodent (mostly rat) liver microsomes and S9 fraction. The principal applicability of this simulator is associated with the reproduction as well as the prediction of the metabolic activation reactions and pathways of xenobiotic chemicals, which may elicit in vitro genotoxicity effects such as bacterial mutagenicity and chromosomal aberrations.

Each transformation in simulator consists of source and product structural fragments, and inhibiting "masks". A probability of occurrence is ascribed to each principal transformation, which determines its hierarchy in the transformation list. A training set of xenobiotic chemicals of a wide structural diversity, with experimentally observed metabolic reactions and pathways has been built, using published data on their metabolism in rodent liver microsomes and S9 fraction. The organic compounds in the training set belong to different classes of industrial chemicals, including single and fused-ring arenes, phenols, haloalkanes and haloarenes, aromatic and aliphatic amines, nitroarenes, alkanes and cycloalkanes, alkenes, ethers, carboxylic acids and their derivatives, halogenated hydrocarbons, alcohols, epoxides, N-nitrosoamines, azo chemicals, etc. The data on their metabolism are collected mostly from research publications in the field from selected scientific journals, monographs and websites, and are associated with the commonly observed in vitro liver metabolic reactions of chemicals with different structures.

The molecular transformations set consists partly of 25 - 30 abiotic and, also, a few enzyme-controlled reactions believed to occur at a very high rate as compared to the duration of the tests, and the highest priority is assigned to these reactions. This subset of reactions includes also transformations of highly-reactive functional groups and intermediates, such as tautomerizations, arene epoxide rearrangements to phenols, etc. On the whole, the simulator contains also 450 – 470 enzymatic phase I transformations, such as aliphatic C-oxidation, aromatic C-hydroxylation, oxidative N- and O-dealkylation, epoxidation, ester and amide hydrolysis, carbonyl group reduction, nitro and azo group reduction, N-hydroxylation, etc. Additionally, 15 – 20 enzymatic phase II transformations, such as glucuronidation, sulfation, glutathione conjugation, N-acetylation, etc. are included with significantly lower priority than phase I ones.

- Skin metabolism simulator

Skin metabolism simulator mimics the metabolism of chemicals in the skin compartment. Given the lack of reported skin metabolism data and the widespread hypotheses is that skin enzymes can metabolize absorbed xenobiotics via reactions analogous to those determined in liver, the simulator was developed as a simplified mammalian liver metabolism simulator. The skin metabolism simulator contains a list of hierarchically

ordered principal transformations, which can be divided into two main types – rate-determining and non-rate-determining. The rate-determining transformations are Phase I and Phase II, such as C-hydroxylation, ester hydrolysis, oxidation, glutathione conjugation, glucuronidation, sulfonation. The non-rate-determining transformations include molecular transformations of highly reactive intermediates. The simulator starts by matching the parent molecule with the reaction fragments associated with the transformation having the highest probability of occurrence. This produces a set of first level metabolites. Each of these derived metabolites is then submitted to the same list of hierarchically ordered transformations, to produce a second level of metabolites. The procedure is repeated until a constraint for metabolism propagation is satisfied (e.g. low probability of obtaining a metabolite or application of Phase II reaction).

The user can apply a selected simulator (1) to a given chemical. The generated metabolite(s) are loaded on data matrix (2). Double clicking over a cell with metabolite(s) evoke a window with displayed metabolite(s) (3). (Figure 1)

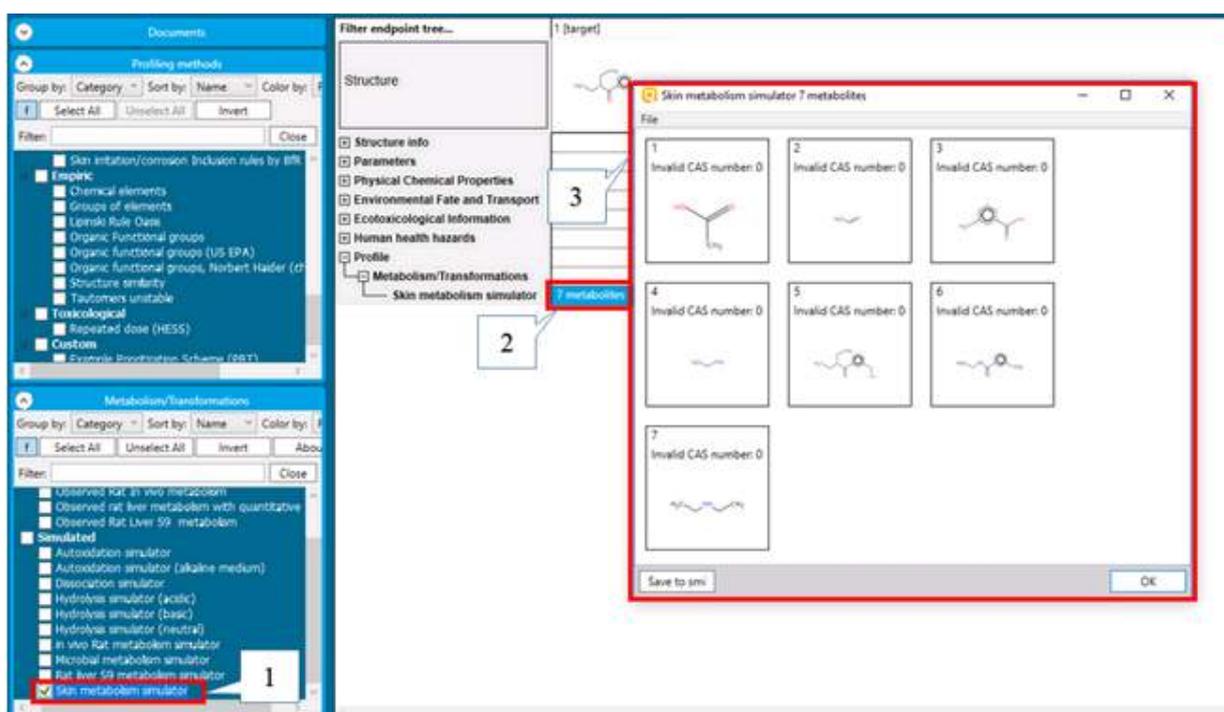


Figure 1

D.2.3. Supporting functionalities

D.2.3. Supporting functionalities

2.3.1. Relevancy of the profiling schemes

Relevancy of a profiler relies on the state how suitable a profiler to build a category is with respect to selected endpoint. Once the endpoint is selected (1) the relevant profilers can be highlighted (2) and grouped (3) by *Endpoint selected in the data matrix*. Meaning of the different colors can be seen by click on the *Legend* button (4). (Figure 1)

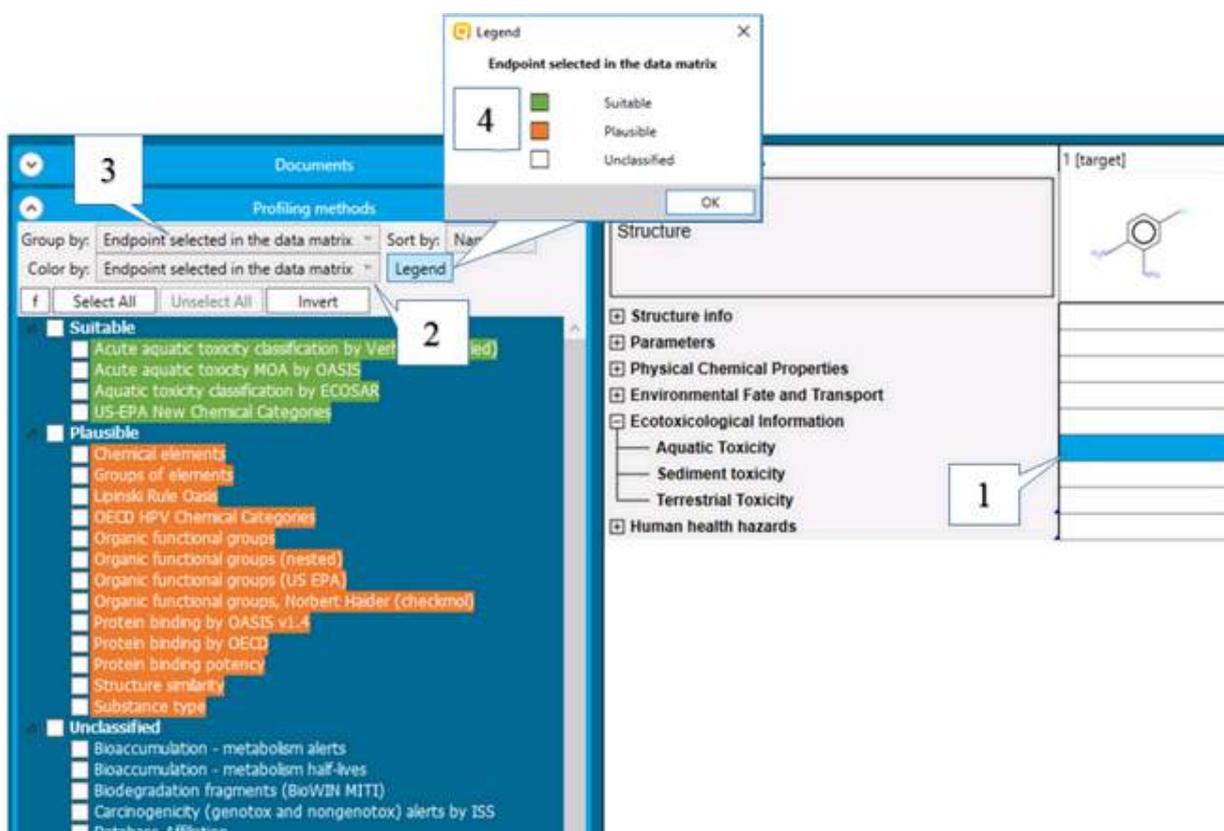


Figure 1

The relation to experimental data and expert knowledge used for the development of the profiler is taken into account in this assessment. Three scores for relevancy are available:

- *Suitable* (highlighted in green) - profiler is developed using data/knowledge for the target endpoint
- *Plausible* (highlighted in orange) - data/knowledge used to build the

profiler is known to be somehow related to the target endpoint

- *Unknown* (not highlighted) - no evidence for relation between used data/knowledge and the target endpoint

If the user move to other position (1) in the endpoint tree, the relevant profilers are changed (2) (Figure 2). The same scores for relevancy can be applied and for simulated and observed metabolic maps (3).

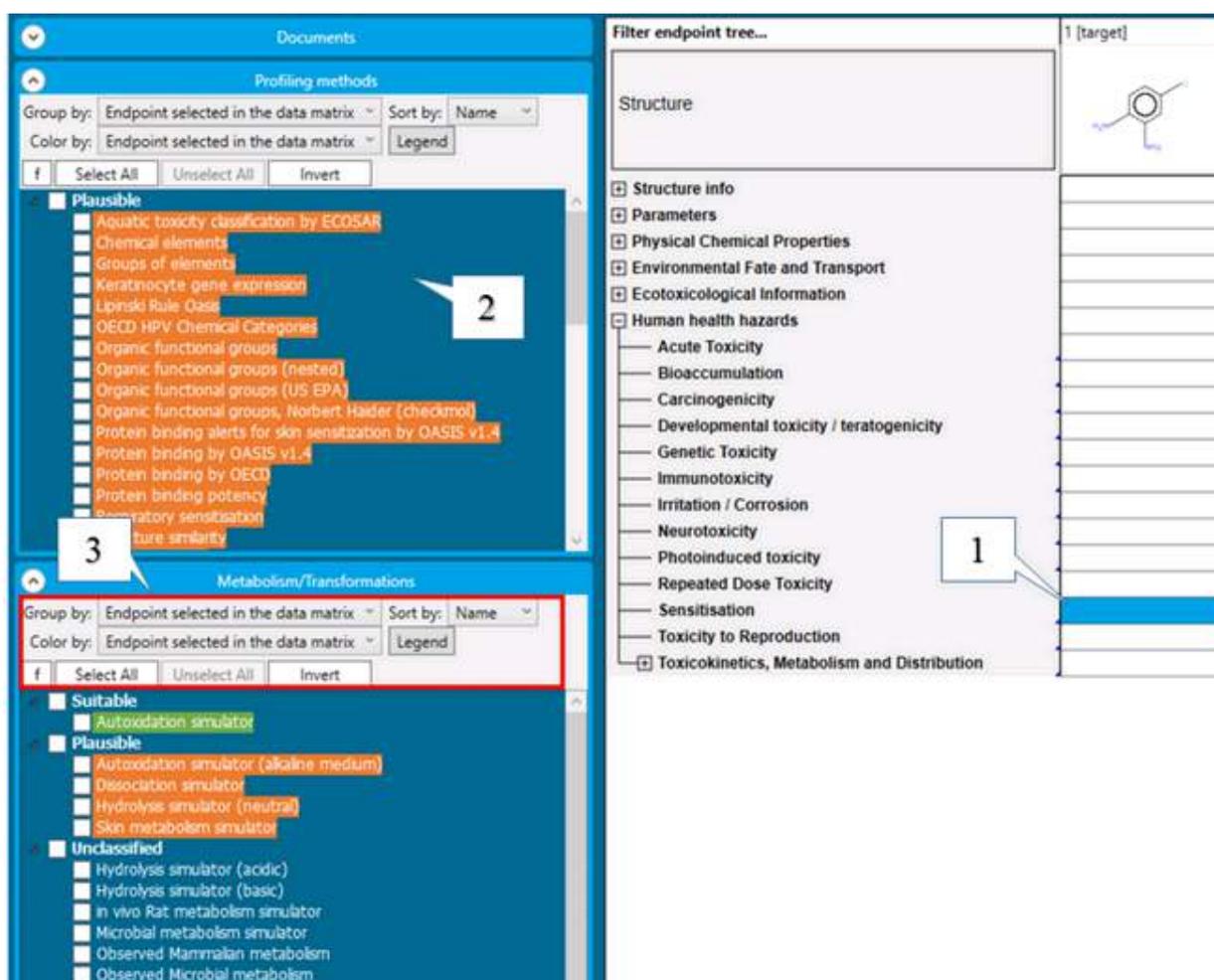


Figure 2

The user can see only the profilers relevant to their previously defined endpoint (1) by coloring and grouping according to the Target Endpoint (2). Selection of this option does not correspond to the position of the endpoint tree (3), but only to the selected endpoint (1). (Figure 3)

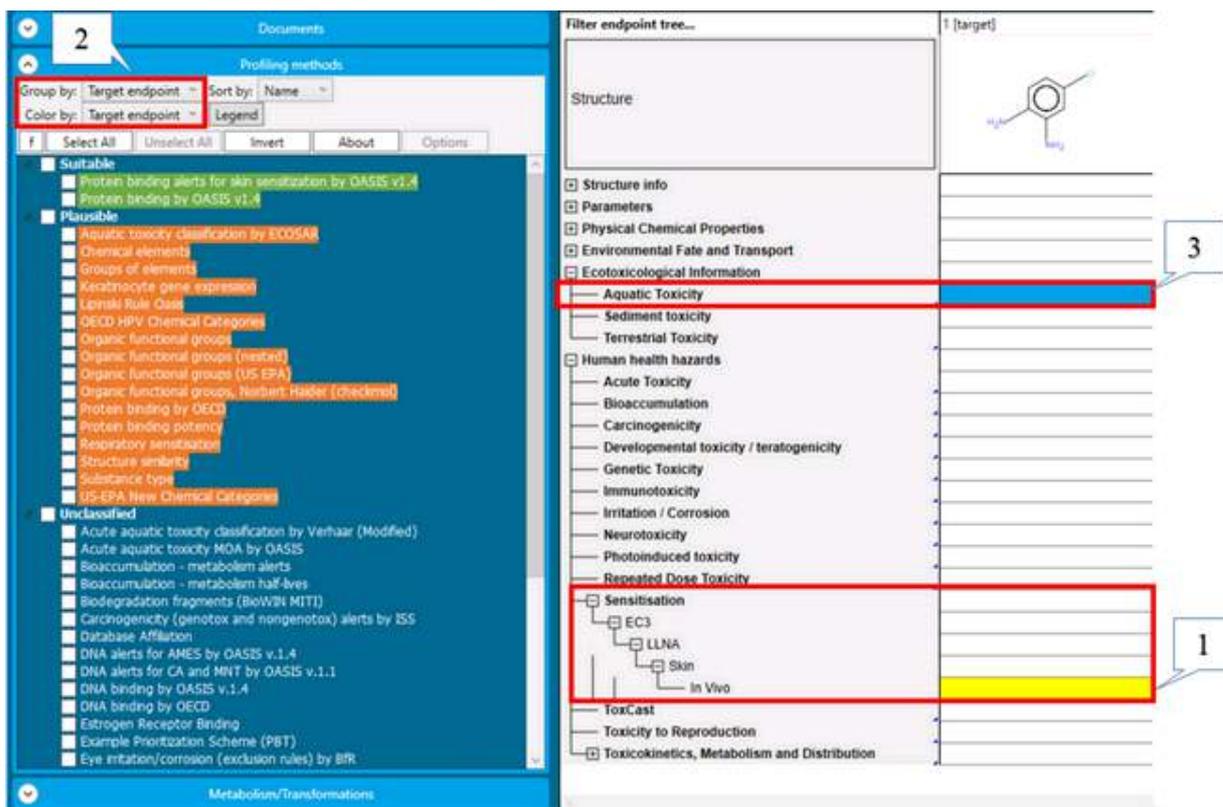


Figure 3

2.3.2. Filtering

Filtering allows the user to find more quickly the profilers and/or simulators of interest. The user have to click on the F button (1) and to write a key word (2). Then the profilers which meet the requirement are filtered (3). (Figure 4)

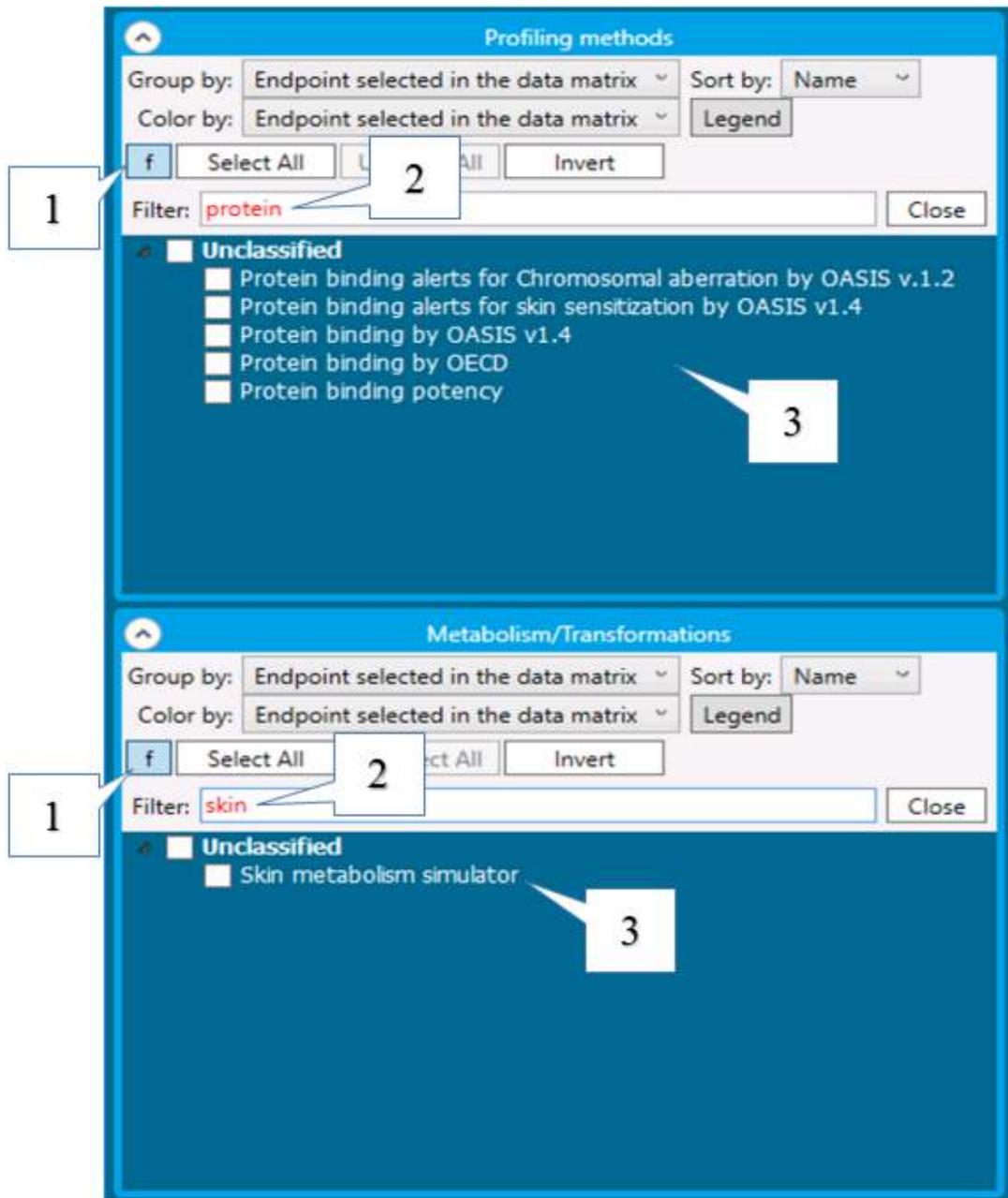


Figure 4

The user can see the query tree and the content of a selected boundary. New categories can be defined, existing ones can be renamed or deleted, rules can be added or deleted, new boundaries can be set. Changes for the profilers that are delivered via the Toolbox installation cannot be saved.

Hierarchical type profiling scheme

A hierarchical organization of categories in profiling scheme (Figure 2) has been developed in Toolbox 3.0. This organization is implemented for the following profiling schemes:

- DNA binding by OASIS
- DNA binding by OECD
- DNA alerts for AMES by OASIS v.1.4
- DNA alerts for CA and MNT by OASIS v.1.1
- Protein binding by OASIS
- Protein binding by OECD
- Protein binding potency
- Protein binding potency Cys (DPRA 13%)
- Protein binding potency Lys (DPRA 13%)
- Keratinocyte gene expression

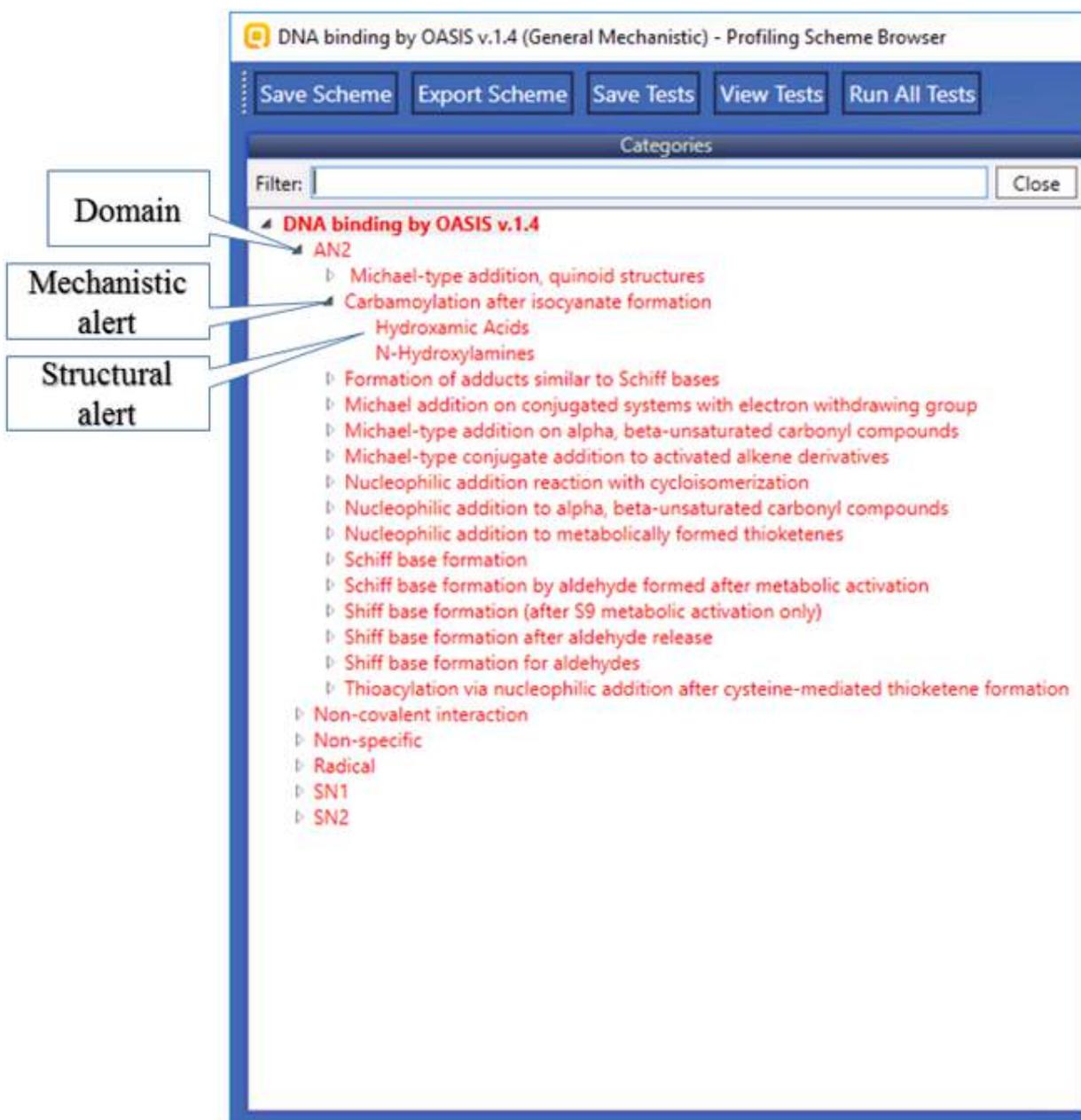


Figure 2

Dendroidal type profiling scheme

The visualization of dendroidal type scheme is illustrated on Figure 3.

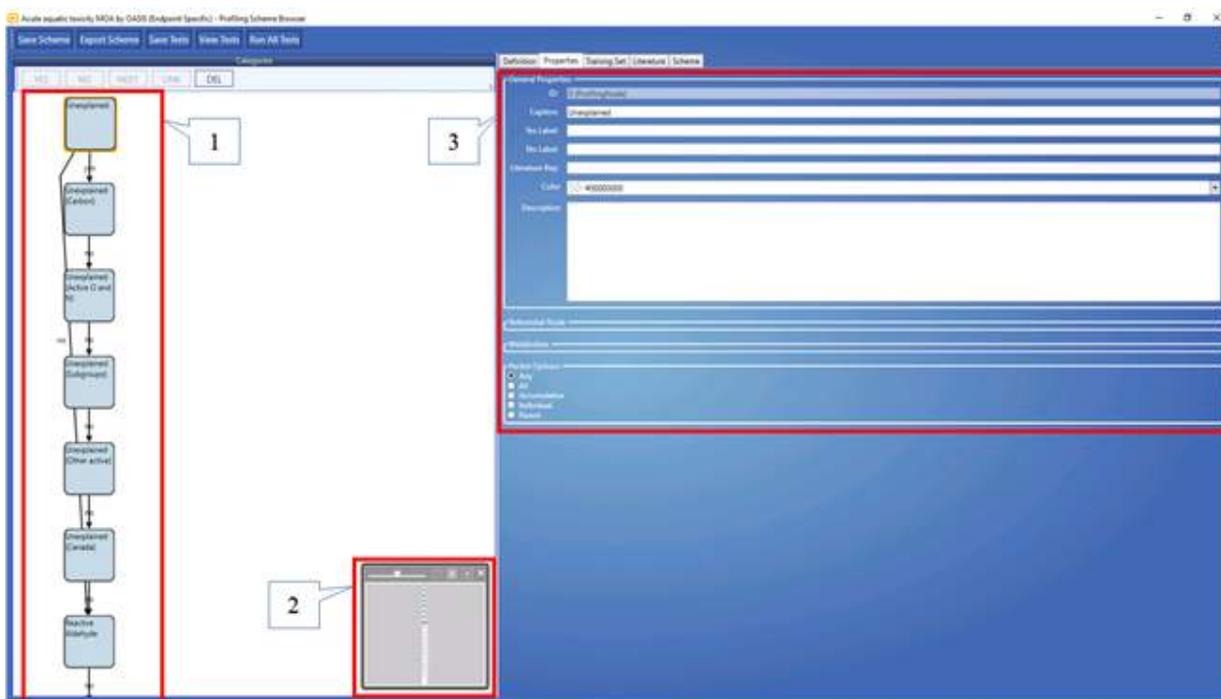


Figure 3

1 – logic panel: here a logic of the profiling scheme organized as a of logical nodes is presented

2 - navigation panel

3 – a panel with properties of nodes

Structure similarity

The Structure similarity profiler is included within the empiric profilers (1) (Figure 4). Structural similarity options in TB 4.0 have been expanded including PubChem substructure features (2). The PubChem generates a binary substructure fingerprint for each chemical structure. A substructure is a fragment of chemical structure. A fingerprint is an ordered list of binary (1/0) bits. Each bit represent a Boolean determination of specific atom or test features used further for similarity neighboring and similarity searching. Seven groups of PubChem features are defined and used:

- Hierarchical element counts;
- Rings;
- Simple atom pairs;
- Simple atom nearest neighbors;

- Detailed atom neighbors;
- Simple SMARTS patterns;
- Complex SMARTS patterns.

Once the PubChem feature has been select percentage of the structural similarity between both example structure has been changed (3). To see similarity details between two structures click on Details (4). A new window appears and the user can see which of the fragments are common (green colored), and which are unique for both chemicals (red colored) (5).

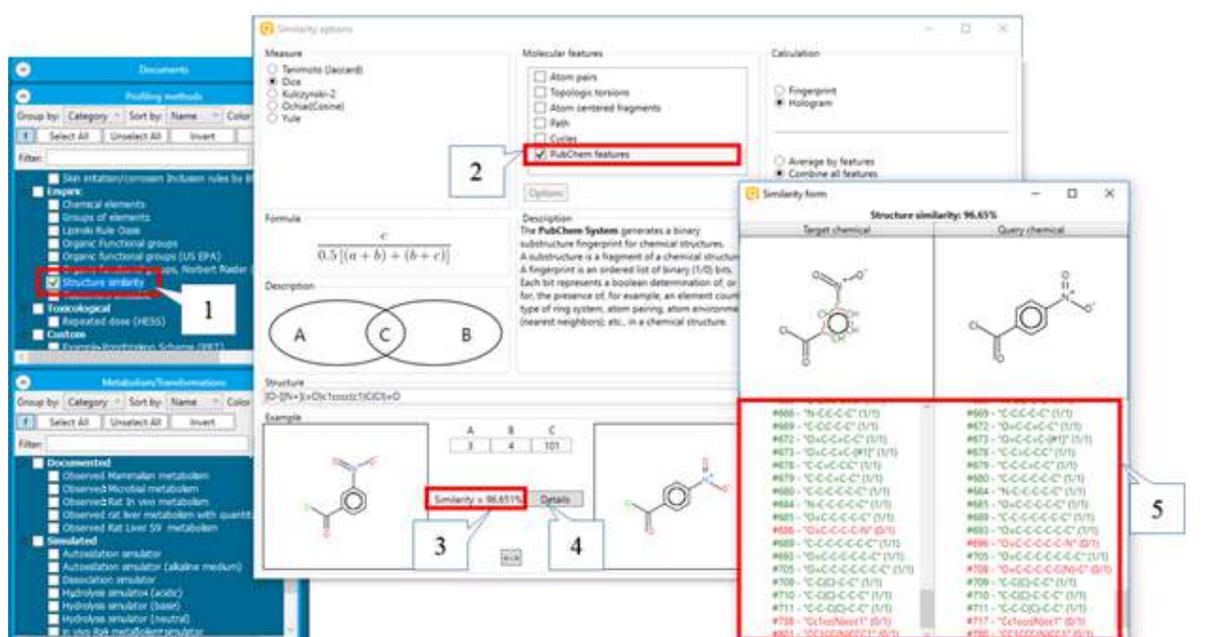


Figure 4

-O-

QSAR Toolbox v.4

D.2.5. Explaining profiling results

[Workflow](#) / [Profiling](#) / D.2.5. Explaining profiling results

D.2.5. Explaining profiling results

Different organization of profiling results displayed on data matrix is available based on the three types profiling methods:

- [Standard explanation](#)
- [Hierarchical explanation](#)

Standard explanation

Profiling results obtained from standard type profiling schemes is shown on Figure 1:

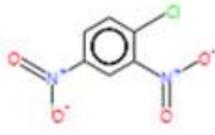
Filter endpoint tree...	1 [target]
Structure	
<input checked="" type="checkbox"/> Structure info	
<input checked="" type="checkbox"/> Parameters	
<input checked="" type="checkbox"/> Physical Chemical Properties	
<input checked="" type="checkbox"/> Environmental Fate and Transport	
<input checked="" type="checkbox"/> Ecotoxicological Information	
<input checked="" type="checkbox"/> Human health hazards	
<input type="checkbox"/> Profile	
<input type="checkbox"/> Predefined	
<input type="checkbox"/> US-EPA New Chemical Categories	Polynitroaromatics (Acute toxicity)
<input type="checkbox"/> General Mechanistic	
<input type="checkbox"/> DNA binding by OASIS v.1.4	Radical Radical >> Radical mechanism via ROS formation (indirect) Radical >> Radical mechanism via ROS formation (indirect) >: SN1 SN1 >> Nucleophilic attack after reduction and nitrenium ion SN1 >> Nucleophilic attack after reduction and nitrenium ion
<input type="checkbox"/> Endpoint Specific	
<input type="checkbox"/> Acute aquatic toxicity MOA by OASIS	Reactive unspecified
<input type="checkbox"/> Aquatic toxicity classification by ECOSAR	Polynitrobenzenes
<input type="checkbox"/> Toxicological	
<input type="checkbox"/> Repeated dose (HESS)	Nitrobenzenes (Hemolytic anemia with methemoglobinemia) Nitrobenzenes (Hepatotoxicity) Rank C

Figure 1

Detailed explanation of the profiling result could be visualized when right click over the cell with profiling result (1) and selects Explain (2). (Figure 2)

1 [target]

Structure

Structure info
 Parameters
 Physical Chemical Properties
 Environmental Fate and Transport
 Ecotoxicological Information
 Human health hazards
 Profile

- Predefined
 - US-EPA New Chemical Categories
- General Mechanistic
 - DNA binding by OASIS v.1.4

- Endpoint Specific
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Toxicological
- Repeated dose (HESS)

Polynitroaromatics (Acute toxicity)

Radical
 Radical >> Radical mechanism via ROS formation (indirect)
 Radical >> Radical mechanism via ROS formation (indirect) >: SN1
 SN1 >> Nucleophilic attack after reduction and nitrenium ion
 SN1 >> Nucleophilic attack after reduction and nitrenium ion

Reactive unspecified

Polynitrobenzenes

Nitrobenzenes (Hemolytic anemia with methemoglobinemia)

Nitrobenzenes (Hepatotoxicity) Rank C

1

2

Explain

Figure 2

As a result of this action a window displaying the boundaries and mechanistic justification of the profiling method appears (1). (Figure 3)

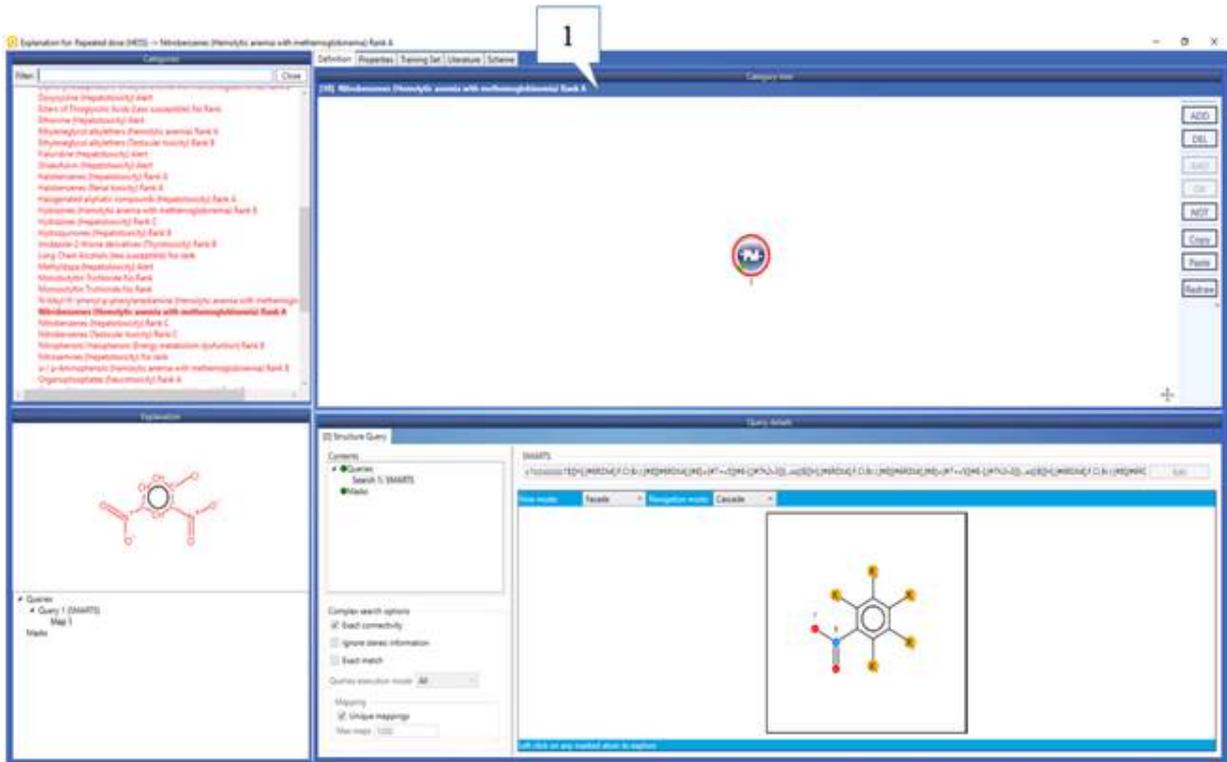


Figure 3

Fields of the profiling results window are explained below (Figure 4):

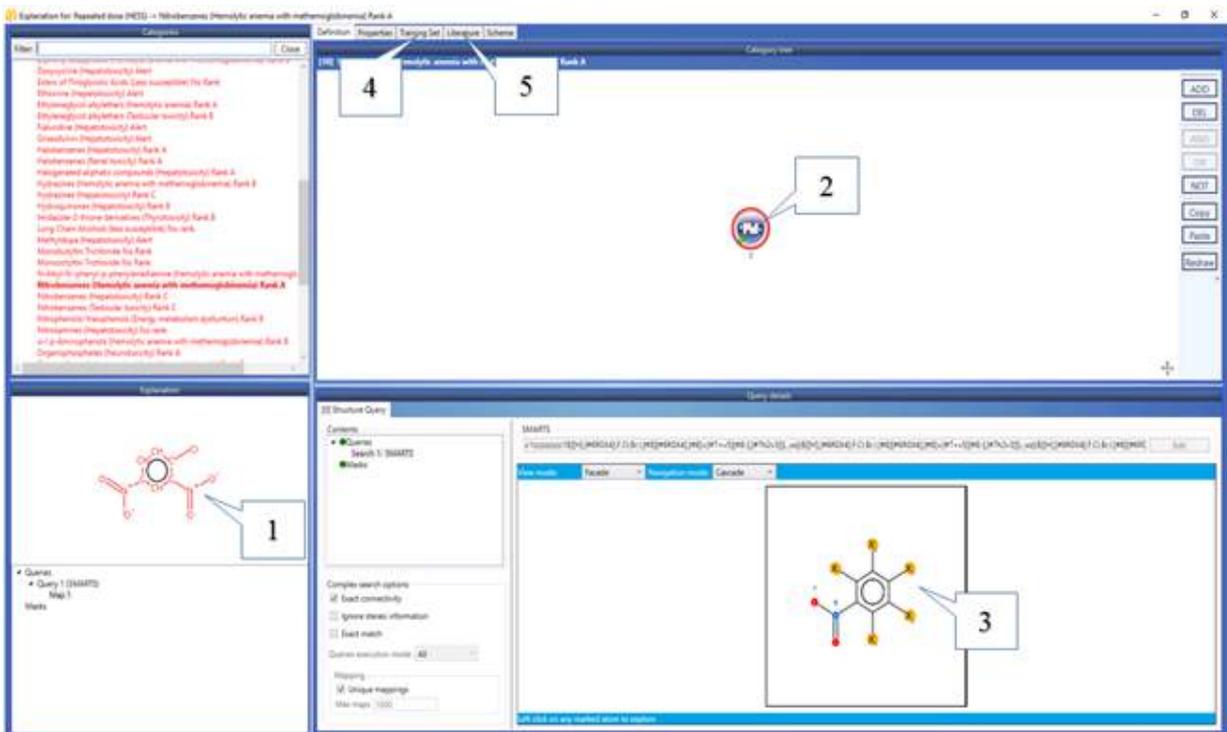


Figure 4

- 1 – Panel with target chemical, if any part of target meets the criteria of the displayed category, then it is highlighted in red.
- 2 – Structural boundary of the displayed category. If the target meets the criteria of the boundaries then it is marked with green tick.
- 3 – Structural fragment used to define the structural boundary of the displayed category
- 4 – Training set chemicals associated to the displayed category
- 5 – Mechanistic justification associated to each of the displayed categories

Training set chemicals are listed in the panel *Training set*. Usually structural boundaries are extracted based on these training set chemicals associated to categories (Figure 5)

The screenshot displays the 'Repeated dose (HESS) (Toxicological) - Profiling Scheme Browser' application. The interface is divided into several panels:

- Categories Panel (Left):** A list of toxicological categories, many of which are highlighted in red. Examples include 'Nitrobenzenes (Hemolytic anemia with methemoglobinemia) Rank A', 'Nitrobenzenes (Hepatotoxicity) Rank C', and 'Nitrobenzenes (Testicular toxicity) Rank C'.
- Training Set Panel (Middle):** A list of chemical identifiers such as '98-95-3 nitrobenzene', '88-72-2 2-nitrotoluene', and '99-08-1 3-nitrotoluene'. A red box labeled '1' points to the 'Training Set' tab in the top navigation bar.
- Chemical Structure Panel (Right):** A chemical structure of nitrobenzene is shown, consisting of a benzene ring attached to a nitro group (NO_2).

Figure 5

The mechanistic justification of the displayed category is located in the panel *Literature*. (Figure 6)

Figure 6

Hierarchical explanation

Based on this hierarchical type organization, the profiling results from obtained from these types scheme is illustrated on Figure 7:

Figure 7

Right click over the Structural alert result (1) and select explain (2) to see details. (Figure 8)

QSAR Toolbox v.4

D.2.6. Creating a custom profile

[Workflow](#) / [Profiling](#) / D.2.6. Creating a custom profile

D.2.6. Creating a custom profile

In addition to the pre-installed categorization schemes, the user can create new profilers, implementing their own logic.

To do this, the user must click on the button New (1). This will bring up a dialog, which requires entering a name for the new categorization scheme (2). (Figure 1)

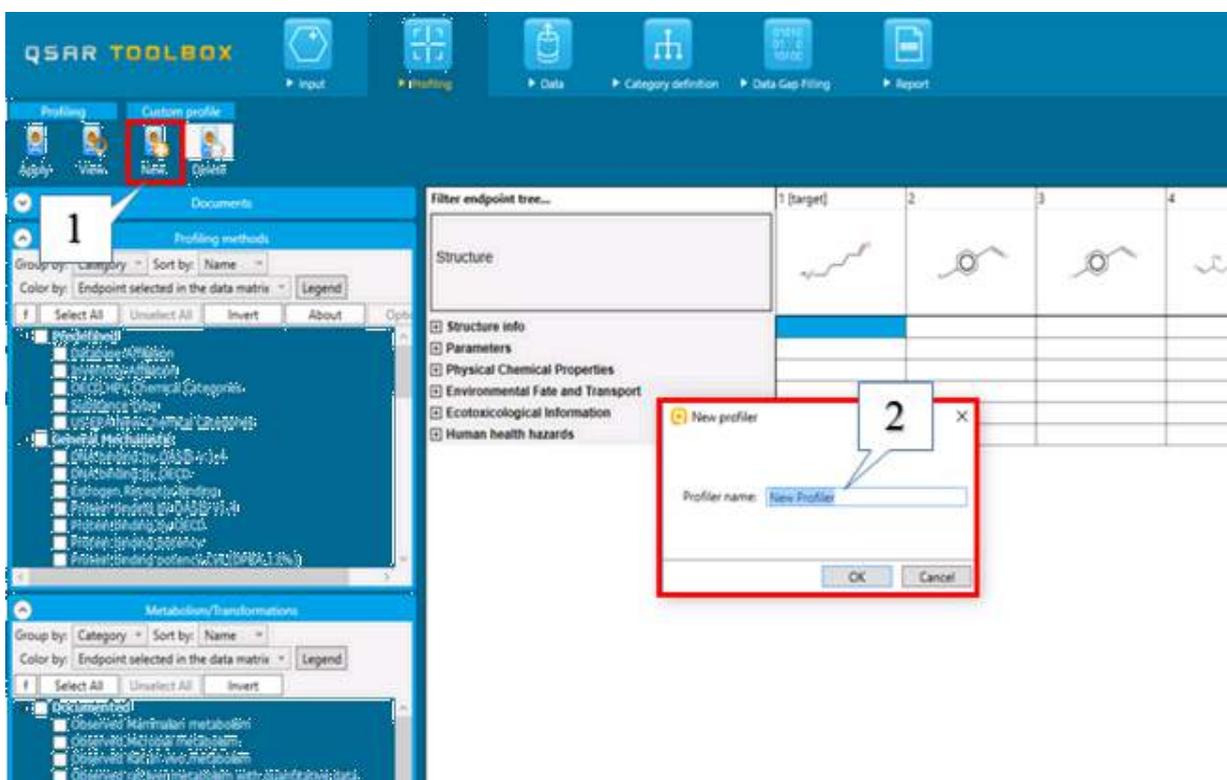


Figure 1

When enter the name, a new window related to the type of scheme appears. The user can select whether its profiling scheme will be linear (or hierarchical) or dendroid (or prioritization scheme) (Figure 2).

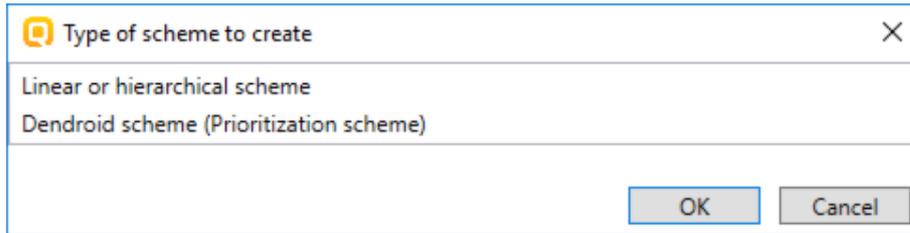


Figure 2

After that the system create a new empty scheme under the Custom section and Profiling Scheme Browser appears.

2.6.2. Adding a new category

To add a new category the user must select Add Category (1) within the right-click options. After entering its name (2) and click on OK, the new category is created. Next, the user has to add its boundaries. (Figure 3)

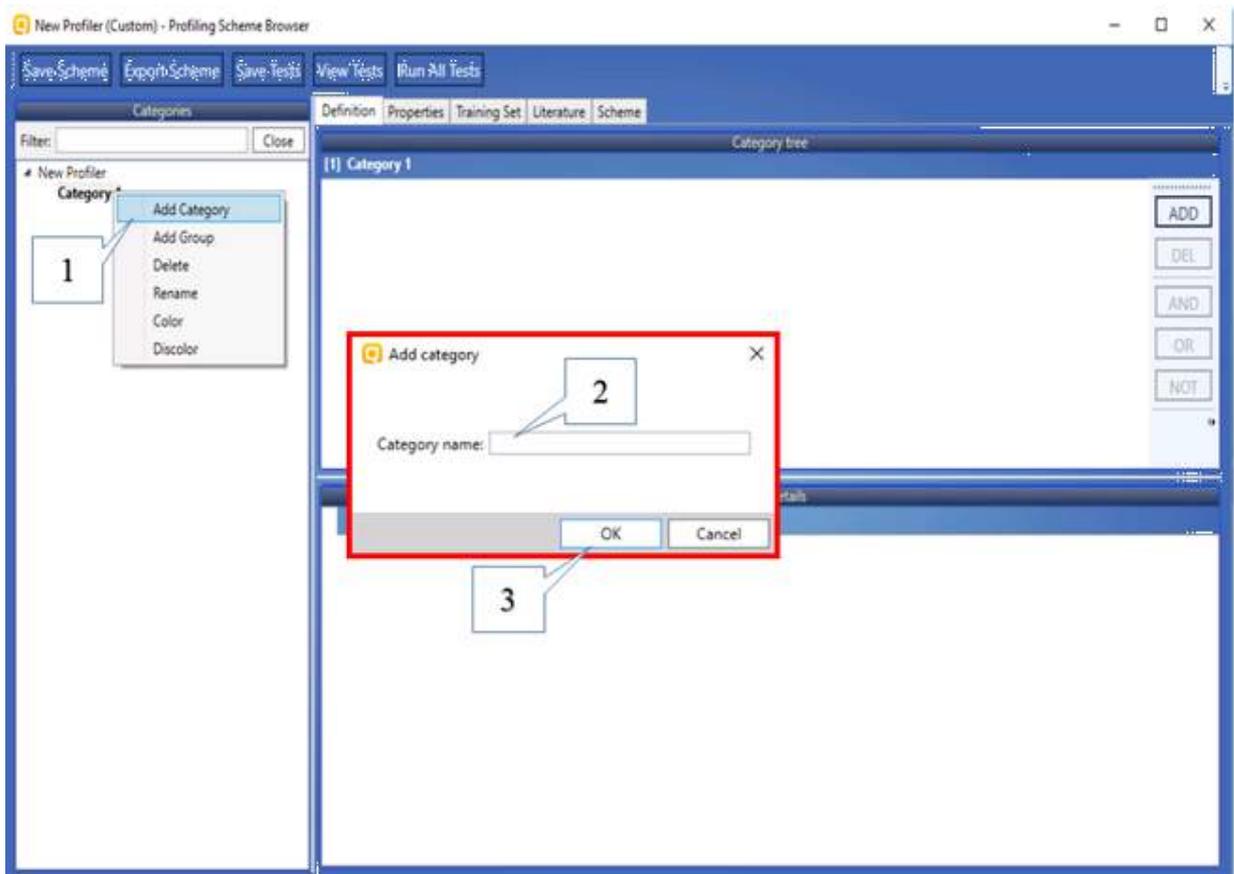


Figure 3

2.6.2. Adding boundaries to a category

To add boundaries, the user must click on the button ADD (1). A list will appear (2) representing the seven types of boundaries that describe a category:

- data,
- label,
- parametric,
- QSAR,
- referential,
- similarity and
- structural.

Depending on the type of category selected, the user will be provided with different options to set the boundaries. For example for parametric boundaries limit values (minimum, maximum or both) are required, for structural boundaries structural fragment(s) are required. (Figure 4)

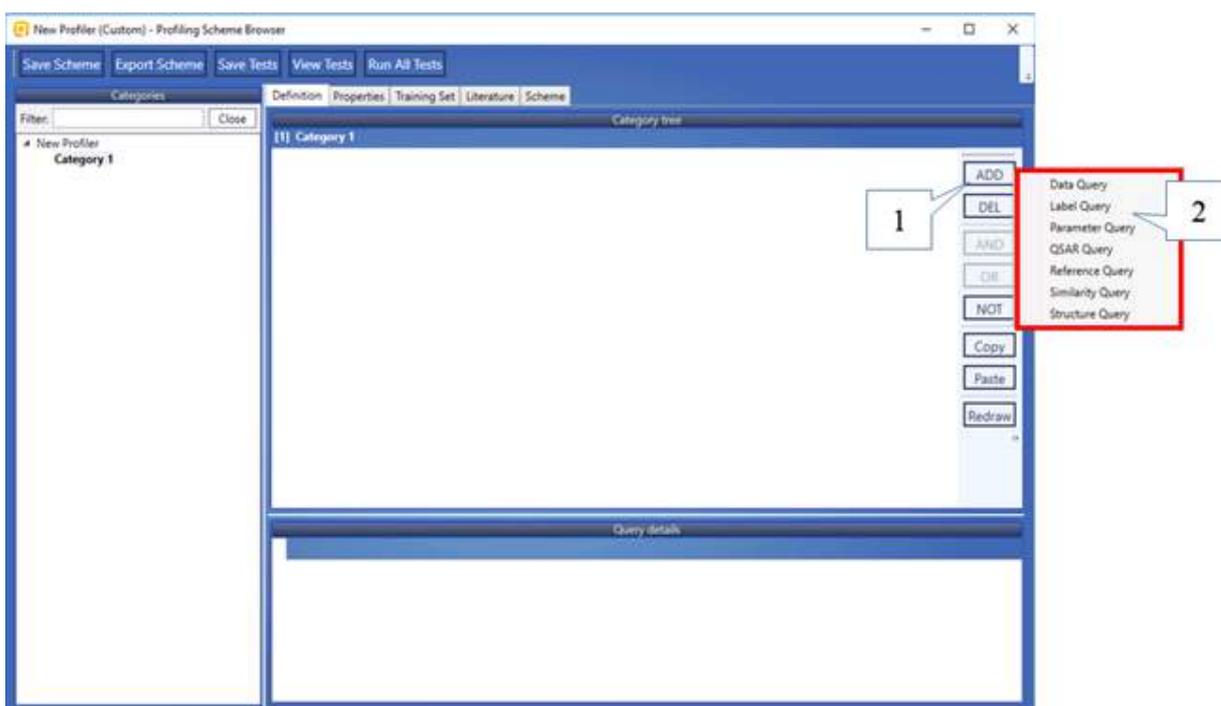


Figure 4

2.6.3. Combining category boundaries

In order to combine logically two or more boundaries, the user must create a query that describes the logic of the category scheme. Each set of boundaries is presented by a query clause (1). Query clauses could be inverted by using the button NOT (2). Additionally, the original clauses or the clauses that are derived from them, could be combined using logical buttons AND (3) or OR

(4).

The picture below presents a category containing two boundary sets - one parametric and one structural. Additionally these boundaries are combined in a way, which defines the following criteria:

- to have MW greater than 100 Da, AND
- NOT to have the fragment C(C)=O in their structure.

After the categories are built, the user have to Save the scheme (5) (Figure 5)

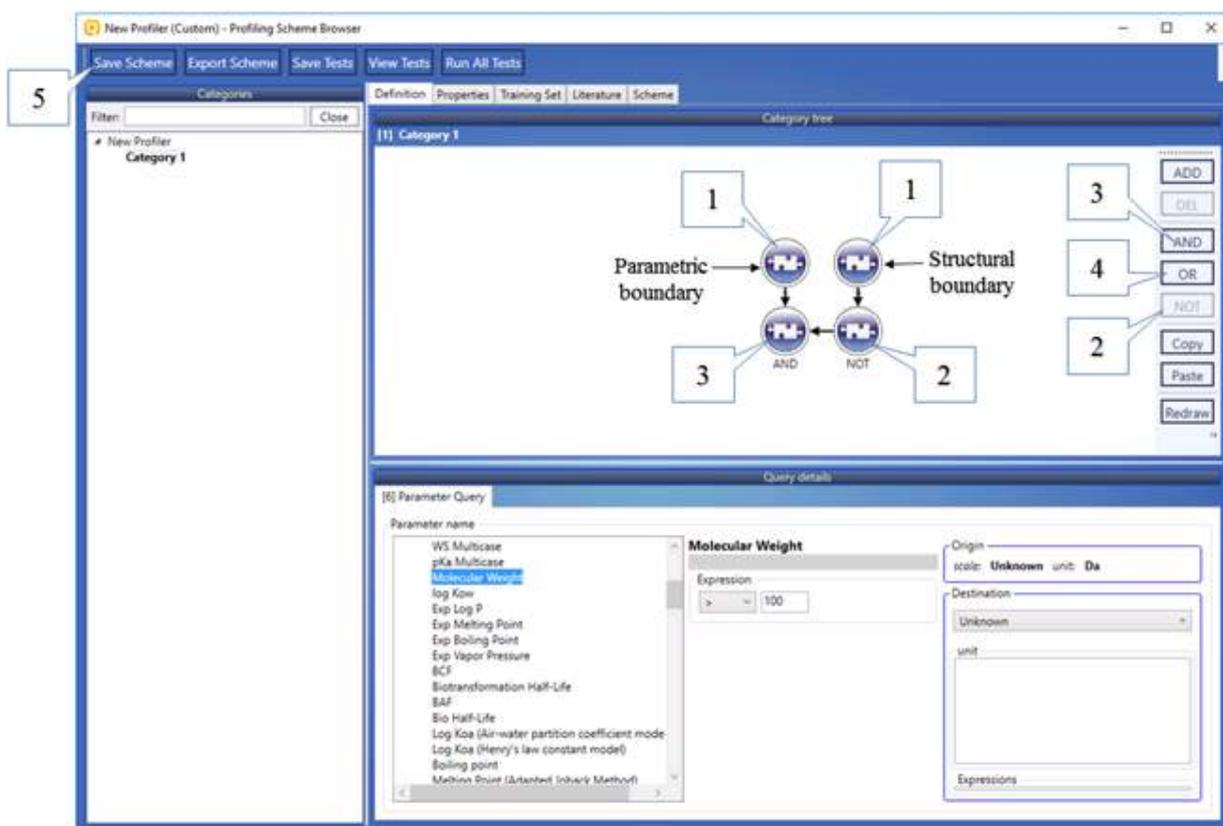


Figure 5

QSAR Toolbox v.4

D.2.7. Deleting a profile

[Workflow](#) / [Profiling](#) / D.2.7. Deleting a profile

D.2.7. Delete a Profiling scheme

The user is allowed to delete a custom profiler ONLY. The available profiling scheme, observed databases and simulators are not available for deletion.

To delete custom profiling scheme, please select the scheme that will be deleted (1), then click Delete scheme button (2). (Figure 1)

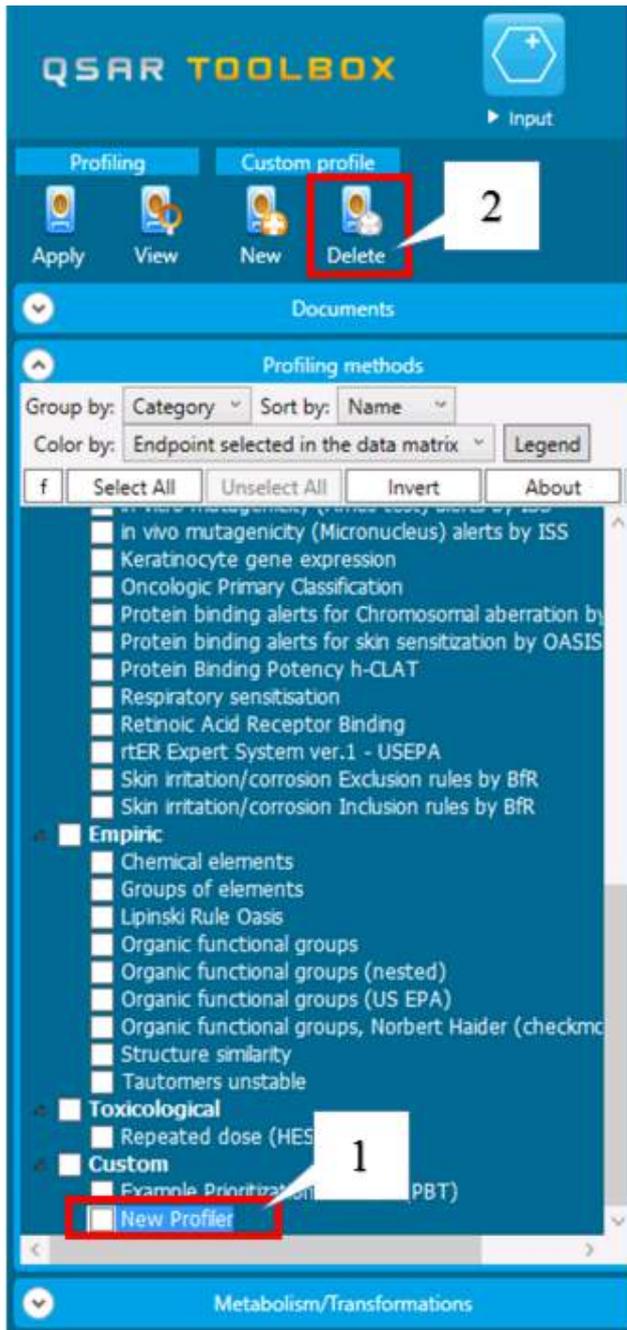


Figure 1

-0-

D.3. Data

QSAR Toolbox v.4

D.3.1. Databases

[Workflow](#) / [Endpoint](#) / D.3.1. Databases

D.3.1. Databases

Each database in Toolbox includes chemicals with measured data. Chemicals are presented with their chemical ID information such as CAS RN, Chemical name and SMILES if available. Measured data (labeled with "M" in the data matrix) could be "quantitative" or "qualitative" type. In this respect some possible conversions of data are developed (scale conversions) in order to save and use all available observed data. For more details about scale and scale conversions, see section [Options/Unit/ Edit Scale](#).

49 databases , available in TB 4.0, are listed in Table 1.

Table 1. List with databases available in Toolbox 4.0

DATABASES	
Acute Oral toxicity	Human Half-life
Aquatic ECETOC	Hydrolysis rate constant OASIS
Aquatic Japan MoE	Keratinocyte gene expression LuSens
Aquatic OASIS	Keratinocyte gene expression Givaudan
Bacterial mutagenicity ISSSTY	kM database Environment Canada
Bioaccumulation Canada	Micronucleus ISSMIC
Bioaccumulation fish CEFIC LRI	Micronucleus OASIS
Biocides and plant protection ISSBIOC	MUNRO non-cancer EFSA
Bioconcentration NITE	Phis-chemEpisuite
Biodegradation in soil OASIS	Phys-chem EPISUITE
Biodegradation NITE	Receptor mediator effect
Biota-Sediment Accumulation Factor US-EPA	Rep Dose Tox Fraunhofer ITEM

Carcinogenic Potency Database (CPDB)	Repeated Dose Toxicity HESS
Carcinogenicity&Mutagenicity ISSCAN	Rodent Inhalation Toxicity Database
Cell Transformation Assay ISSCTA	Skin irritation
Chemical Reactivity COLIPA	Skin sensitization
Dendritic cells COLIPA	Skin sensitization ECETOC
Developmental toxicity ILSI	Toxcast
ECHA CHEM	Toxicity to reproduction
ECOTOX	ToxRefDB US-EPA
ECVAM genotoxicity and Carcinogenicity	Toxicity Japan MHLW
Experimental pKa	Transgenic rodent database
Eye irritation ECETOC	Yeast estrogen assay database University of Tennessee-Knoxville (USA)
Genotoxicity OASIS	ZEBET database
GSH Experimental RC50	

Based on the different types of experimental data a database organization is developed. The organization of databases follows the organization of predefined nodes of Endpoint tree.

Below are shown the four basic sections according to which the databases are distributed:

- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards

Note: If one database includes endpoints from more than one section, it will be multiplied. Such a database is ECHA CHEM, which is included in all sections.

There are several options implemented, which are intended to help the user

when selecting (a) database(s):

- *Group by option* (1) in Figure 1: allows grouping of databases based on hazard effect or donator

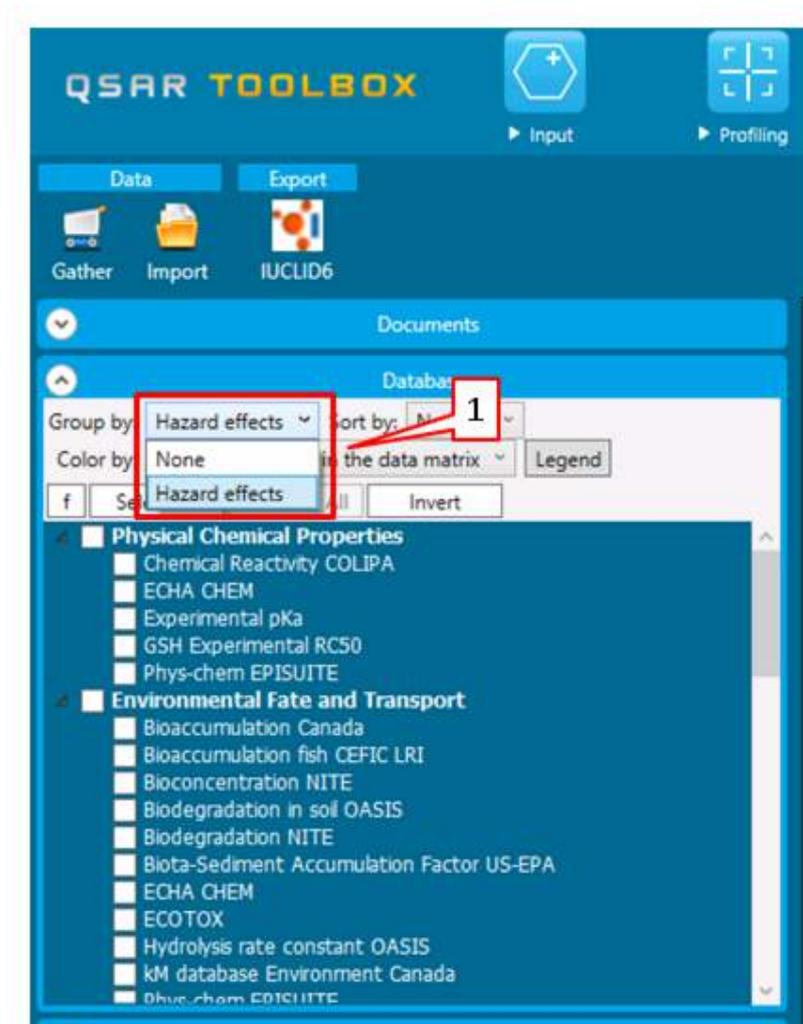


Figure 1

Sort by option (1) in Figure 2: allows sorting by name

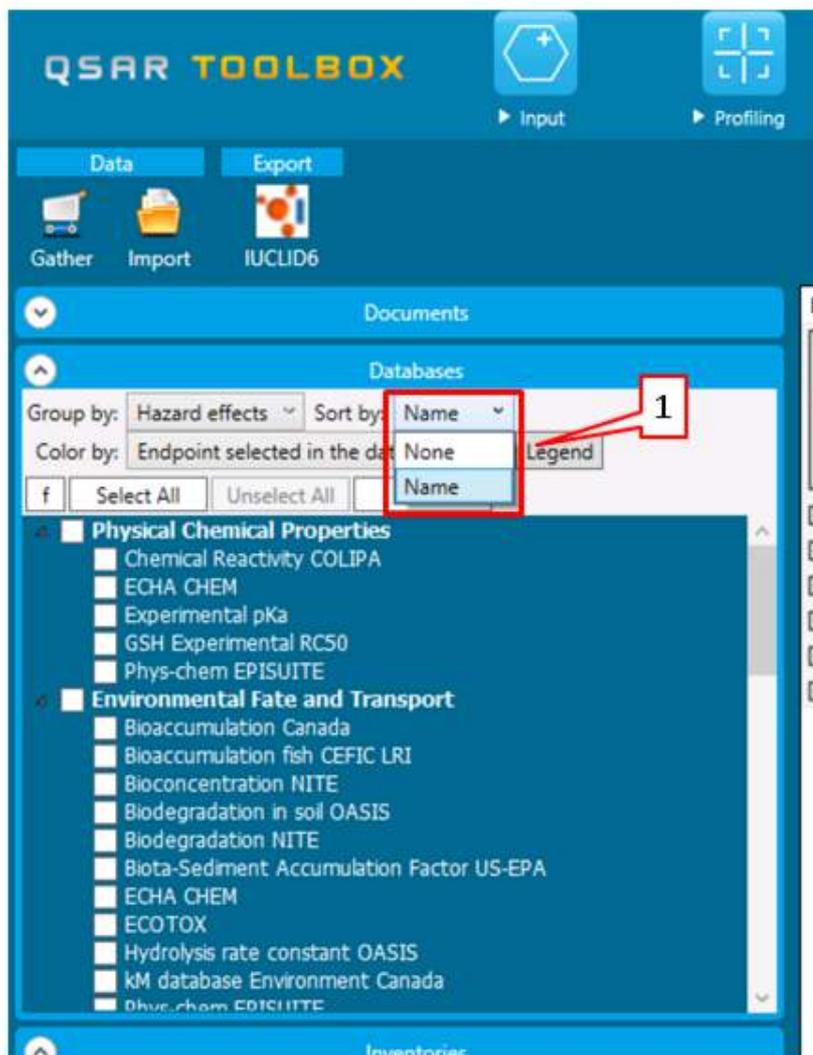


Figure 2

- *Color by option* (1) in Figure 3: allows coloring of the database(s), which have measured data about a selected endpoint. In order for the option to be active, the user has to initially select the endpoint in the data matrix, e.g. Aquatic toxicity (2). Then when selecting Endpoint selected in the data matrix(1), all relevant databases will be colored in green (4). There is also a legend describing the color scheme (3).

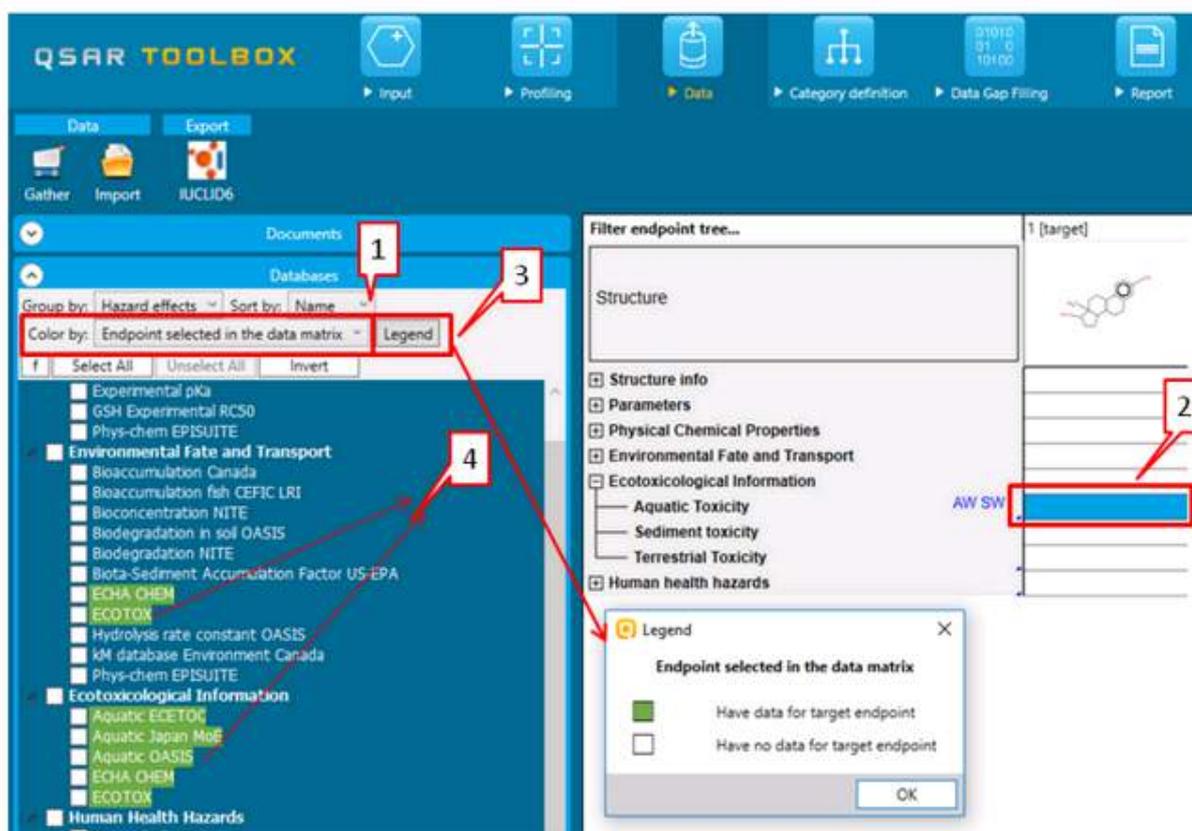


Figure 3

Once the databases are highlighted, the user can select some or all of them (1) and then has to press Gather button (2) (Figure 4). A pop-up window (3) will appear giving the user options to collect data for all or for selected data points. By selecting Choose option (4), the endpoint tree will be visualized and from the drop-down menu arrow the user can select the endpoint of interest, e.g. Aquatic toxicity. Finally, press OK(6).

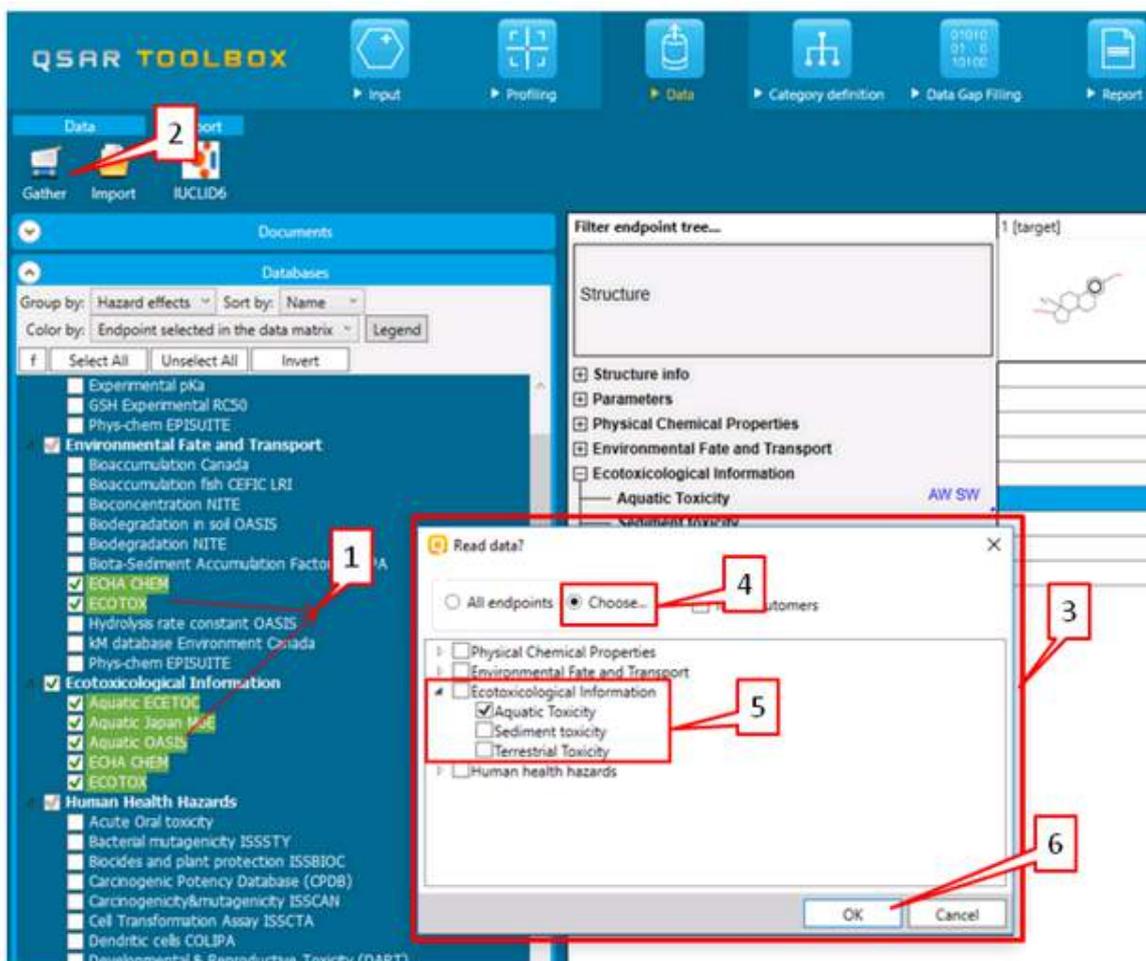


Figure 4

An informative window (1) (Figure 5) will appear giving details about the number of measured data collected for (a) the selected chemical(s). All collected data is shown in the data matrix ((2) in Figure 5) and the user has to open the endpoint tree to visualize the nodes with measured data ((1) in Figure 6).

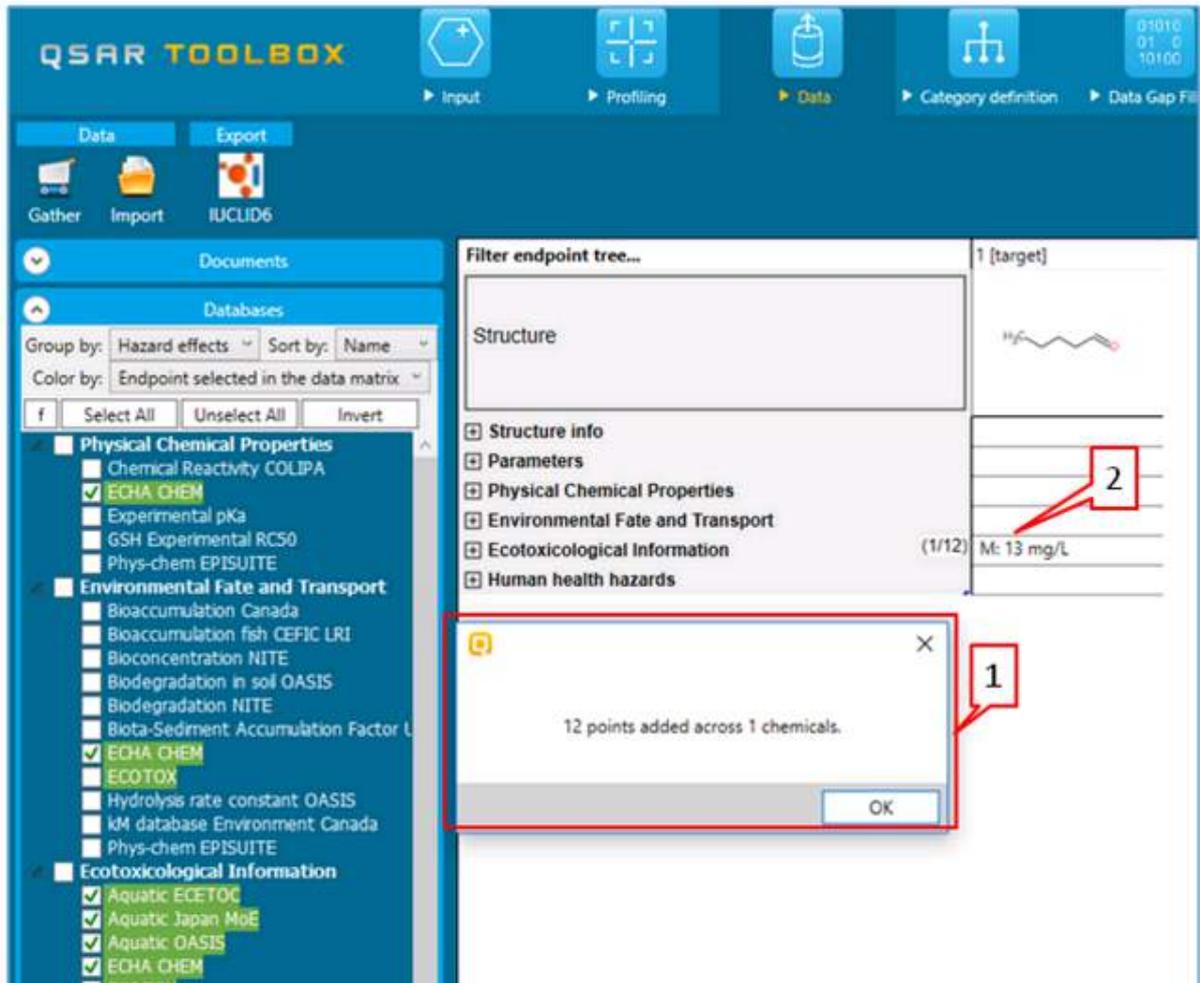


Figure 5

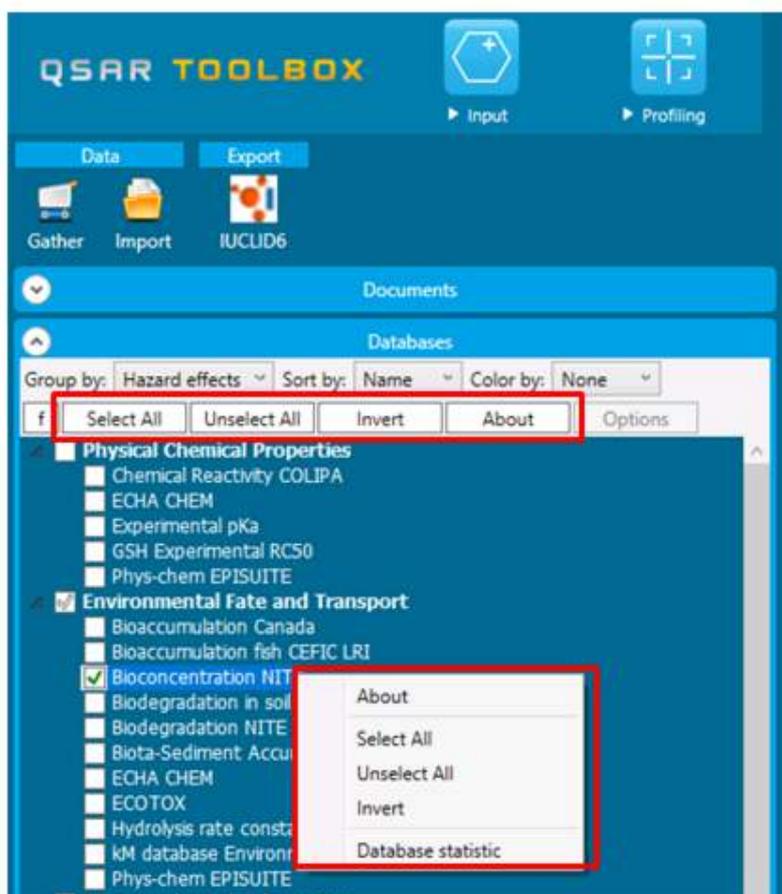


Figure 7

1- Select All- selects all databases. Click Select All button (1), then all databases are selected (2). (Figure 8)

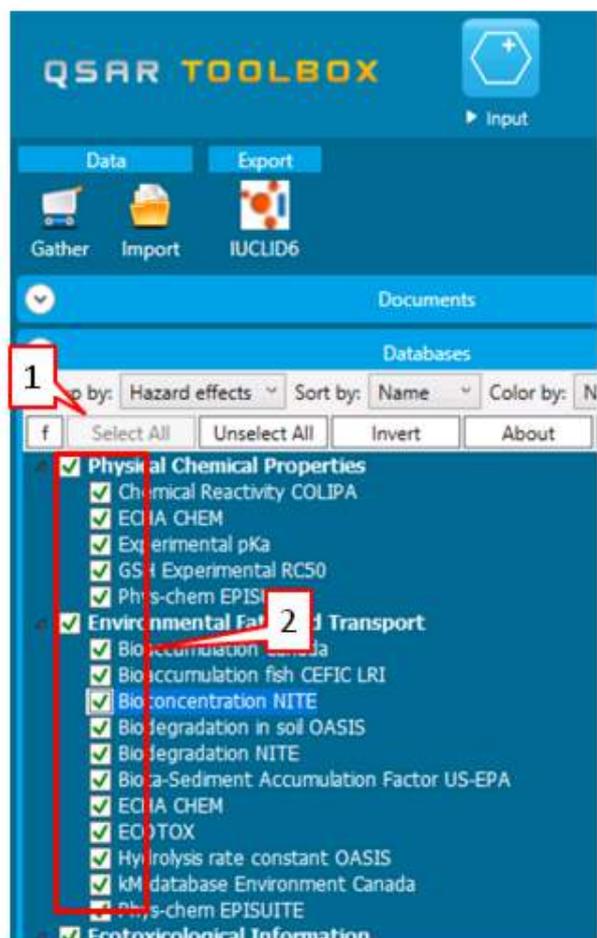


Figure 8

2- Unselect All - unselects all databases. Click *Unselect All* button (1), then all databases are unselected (2). (Figure 9)

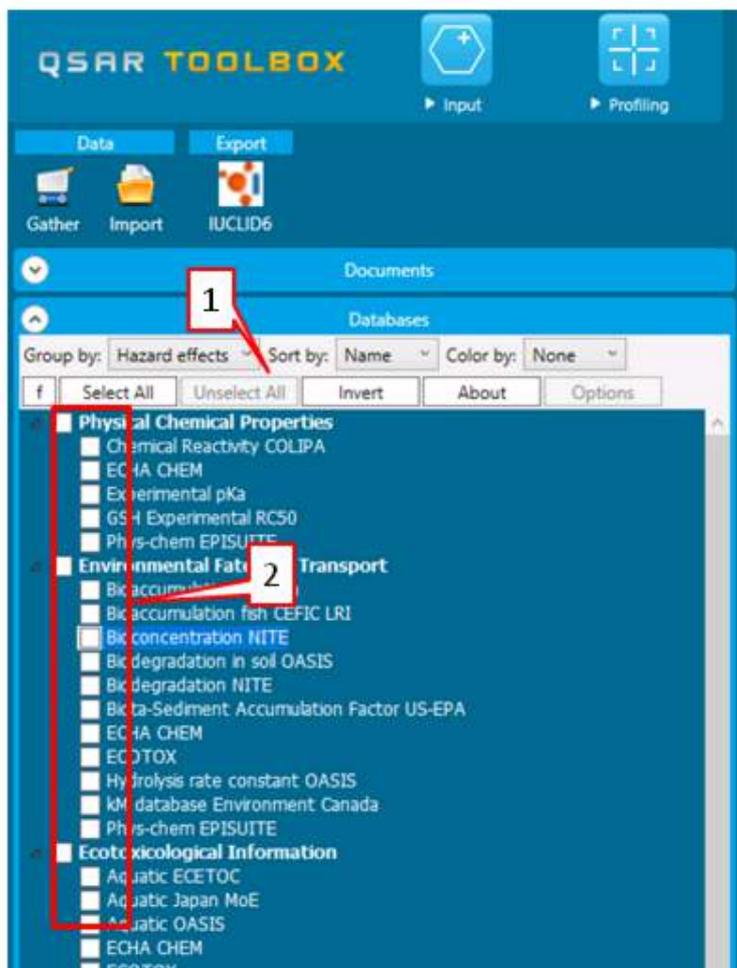


Figure 9

3- Invert - this inverts selection inverts last performed selection of databases. For example if the user selects Bioconcentration NITE (1) and click Invert (2) then the software selects all other databases and unselects Bioconcentration NITE (3). (Figure 10).

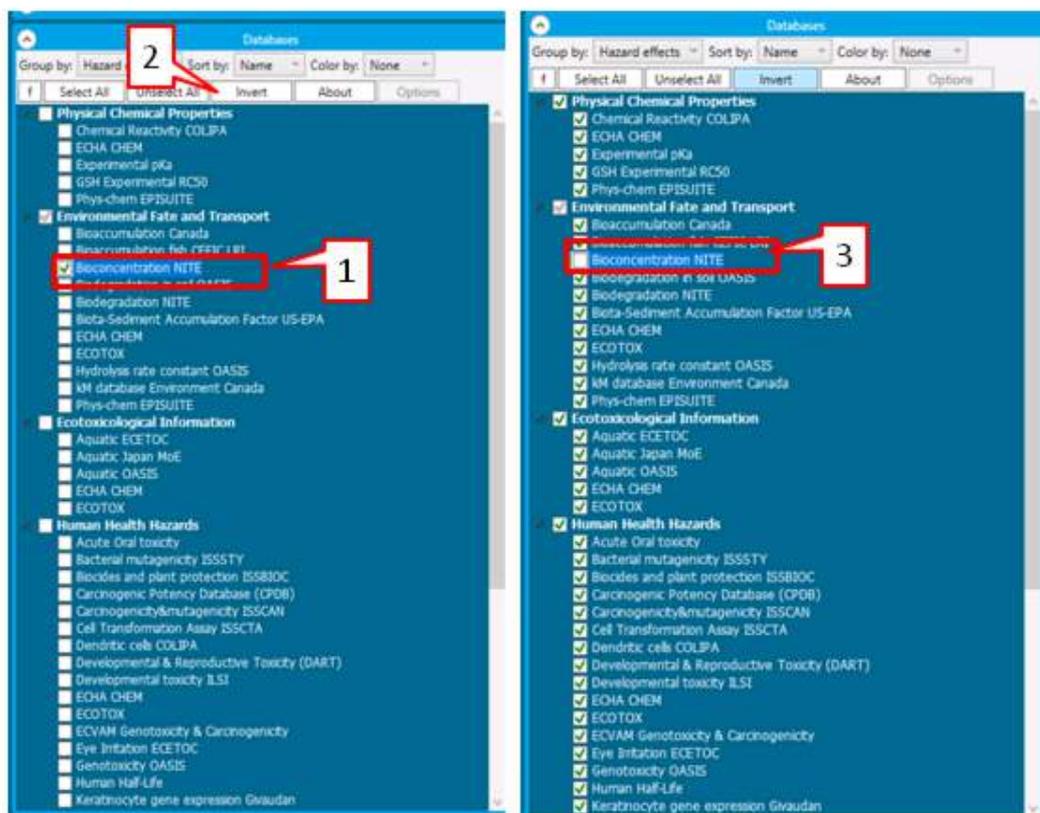


Figure 10

A - About - shows short description of highlighted database (Figure 11). Highlight the database (1); click button About (2), then the short description appears. In the About window (Figure 11), press button Documentation in order to see more details about the database content.

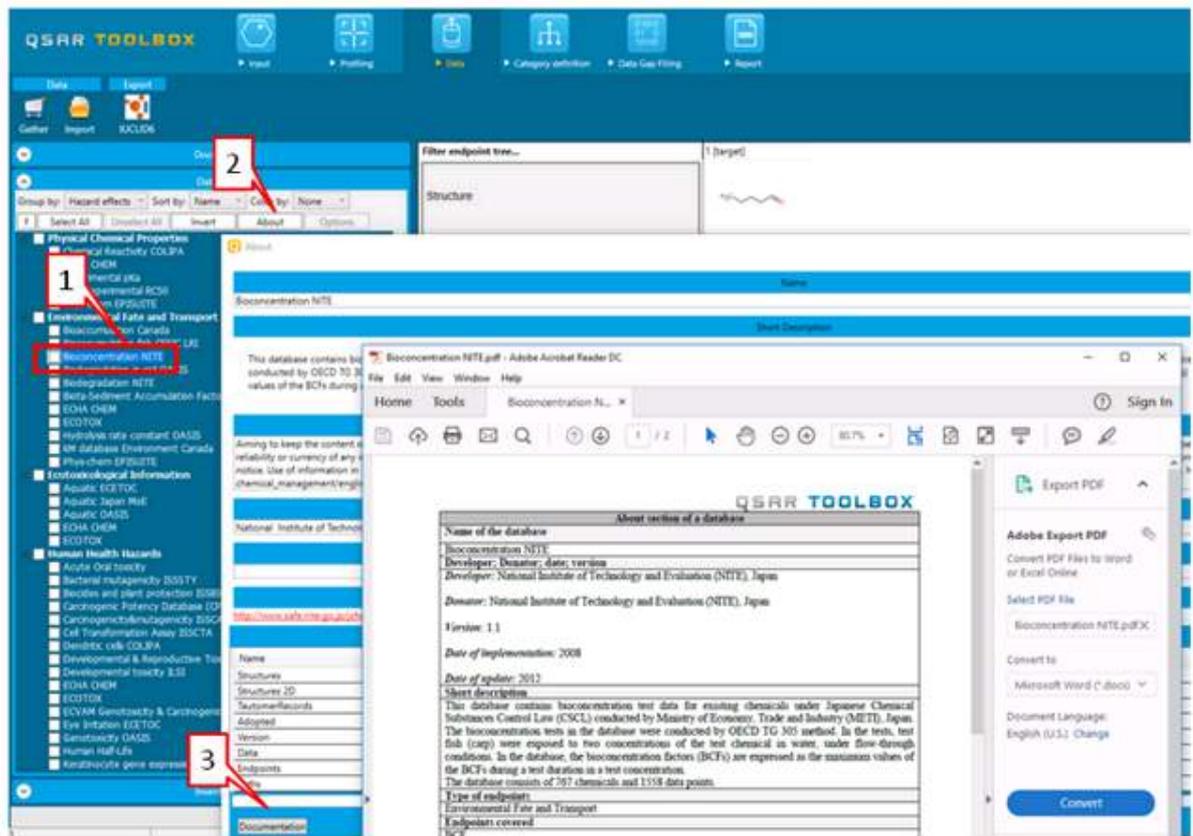


Figure 11

Filter option ((1) in Figure 12) is implemented, for quick search within the databases. When selected a new field for typing is visualized (2), where the user can type in a part of the or the full name of the database of interest.



Figure 12

- *Quality attributes related to databases: Database statistics*

Right click on each database ,e.g. Skin Sensitization (1), opens a window (2), where the user has to select Database statistic(3) (Figure 13)

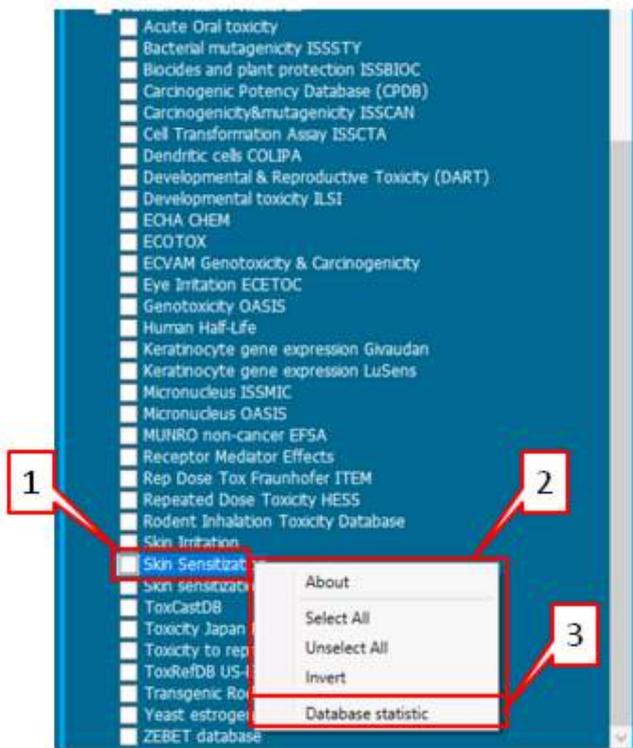


Figure 13

The database statistics of Skin sensitization database is shown in Figure 14.

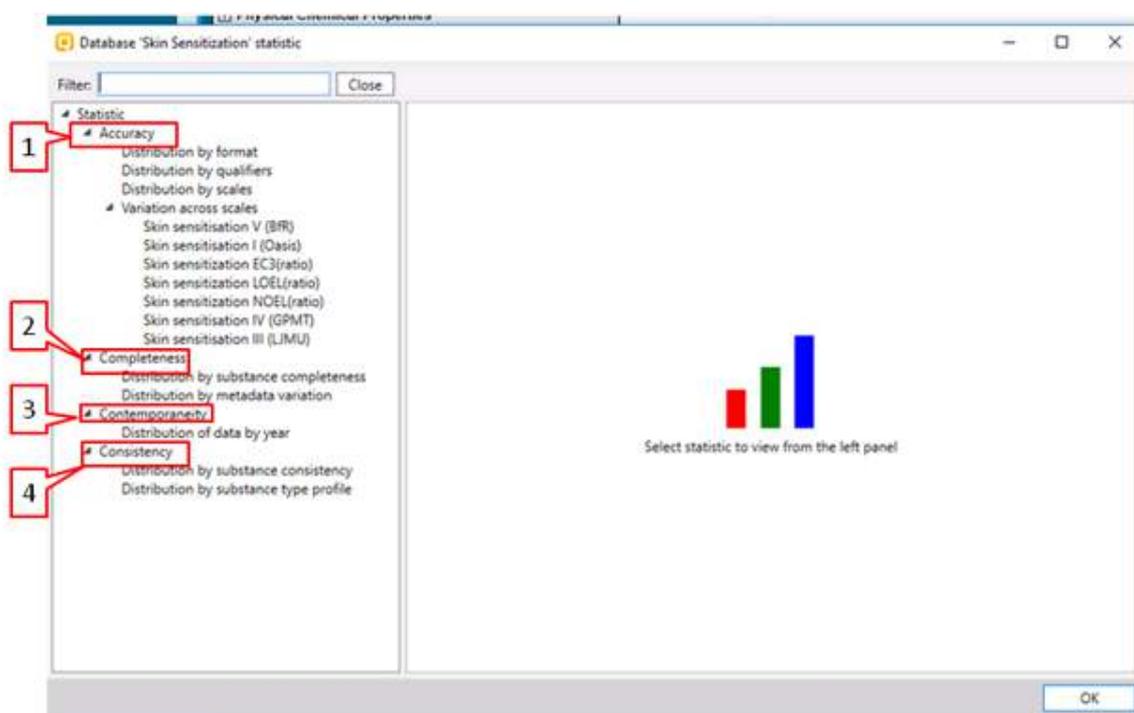


Figure 14

It consists of four attributes: 1) [Accuracy](#) , 2) [Completeness](#) , 3) [Contemporaneity](#) and 4) [Consistency](#).

1) *Accuracy*

Accuracy is the closeness of results of observations to the true values or values accepted as being true. This implies that observations of a phenomenon are usually only considered as estimates of the true value. Because the true value is practically not available, it is very hard to estimate the Accuracy. The following measures of accuracy are implemented:

- Format (categorical or numerical data values (min, max, mean format))
- Qualifiers (<, >, ~)
- Scales (scales, type of scales)
- Variation across the scales (data categories)

Details of Accuracy attribute

- Format - Provide distribution of the data according to their type - numerical or categorical. The numerical data may possess minimal, maximal and mean values. Follow the steps (Figure 15):

- Click on Distribution by format (1)

- Click on the bar and hold to see the number (2) - in this case Skin sensitization database is a compilation of different type of data. It can be seen there are 731 numerical data (left-click on the bar) values and all of them are Mean values. The rest of the data is categorical.

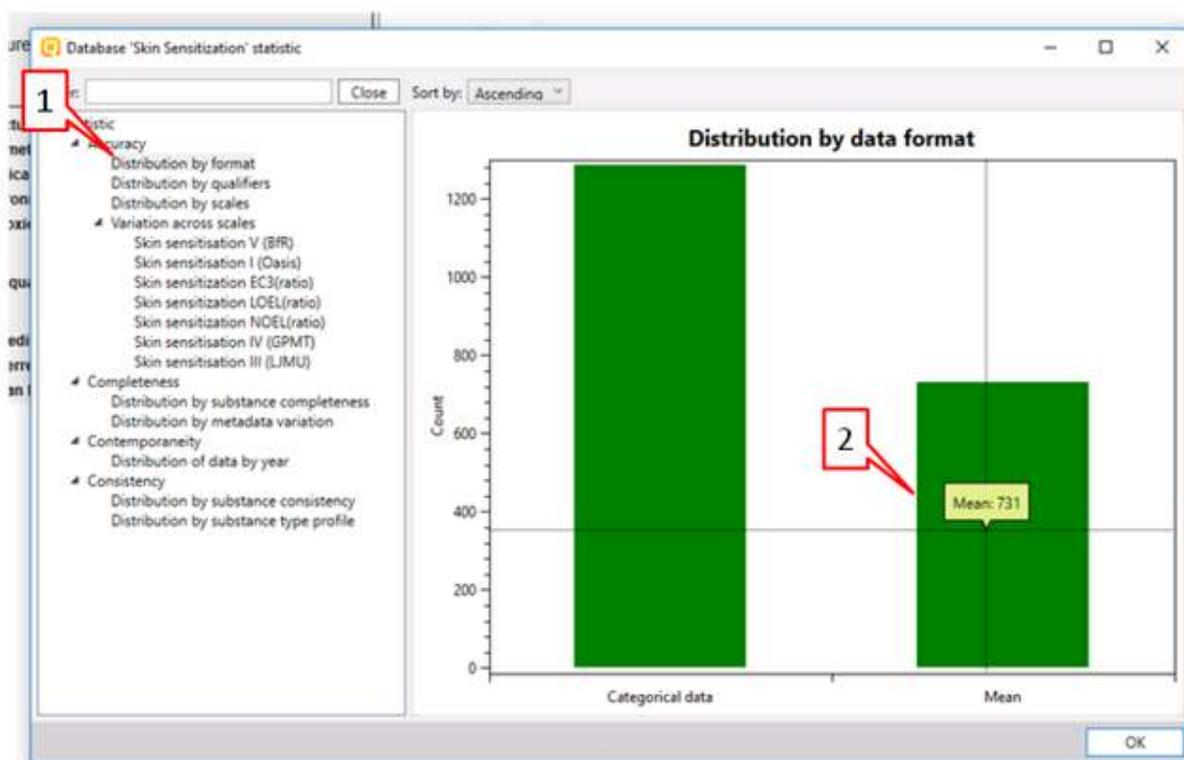


Figure 15

- *Qualifiers* - Useful for numerical data as the categorical data have no qualifiers. The data values possessing qualifiers are treated as absolutely values, currently. Follow the steps (Figure 16):
 - Click on Distribution by qualifiers (2) - the categorical data entries in SS database are 1288 (3).

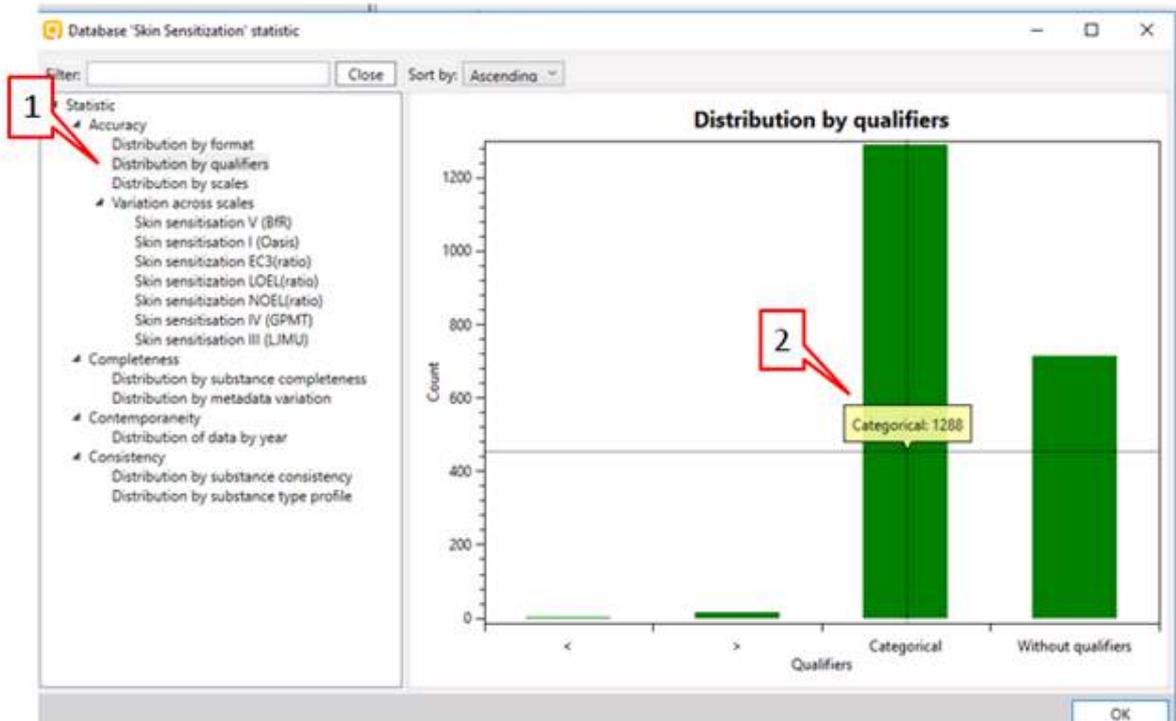


Figure 16

- Scales - Provides data distribution according to their original scales (Figure 17):

Click on *Distribution by scales* (1) - The ratio scale is the most informative scale. Example for a ratio scale is a Skin sensitization IV(GPMT) scale (2)

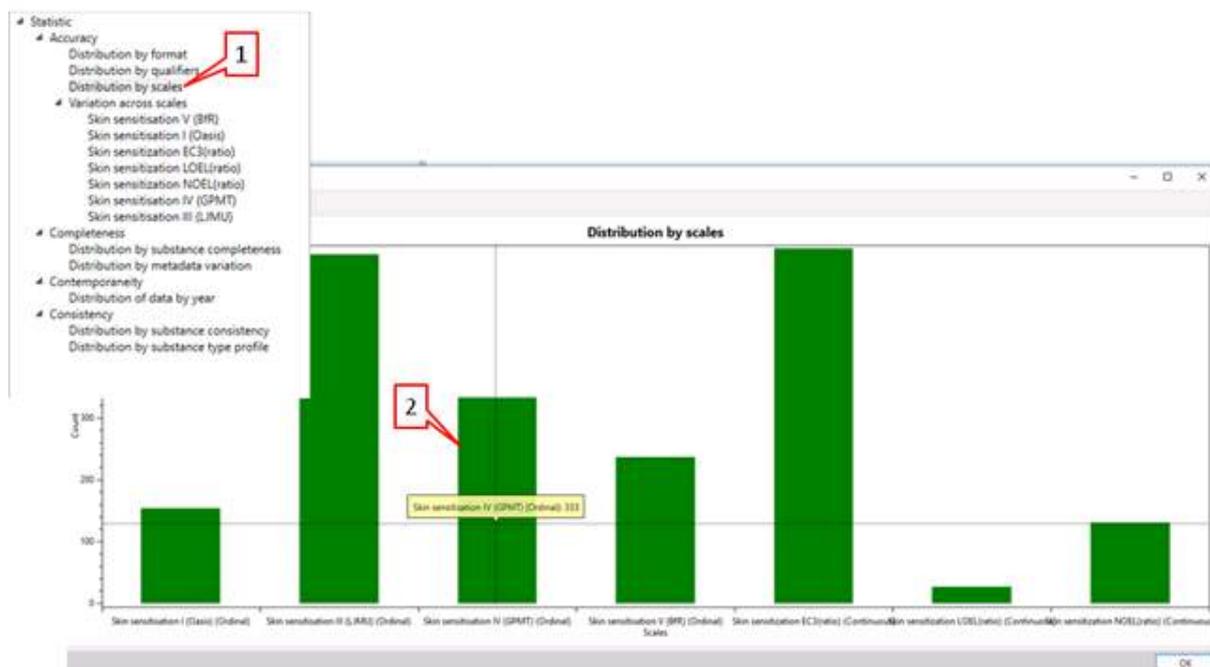


Figure 17

- Variation of data across the scales - This measure of Accuracy is separated to some sub-levels according to the available scales in the original database file. It is currently applicable only to scales with categorical data. Chemicals present in one database could have two or more data related to one scale, e.g. positive, negative. These chemicals are distributed into bars according to their original data. Chemicals with more than one data are divided in One category, Neighboring, Distant and All categories groups.
 - One category - represents the number of chemicals which have several identical data
 - Neighboring categories - represents the number of chemicals, which possess data in close categories (e.g. chemical with weakly positive and strong data);
 - Distant categories - represents the number of chemicals with more than one data where there is a bigger difference in their effects (e.g. one chemical with positive and negative data);
 - All categories - represents the number of chemicals with data in all categories in their original scale.

Follow the steps (Figure 18):

- o Click on variation across scales (1)
- o Click on Skins sensitization III (LJMU) scale (2)
- o As seen most of the data per one chemical are separated into One category (3)

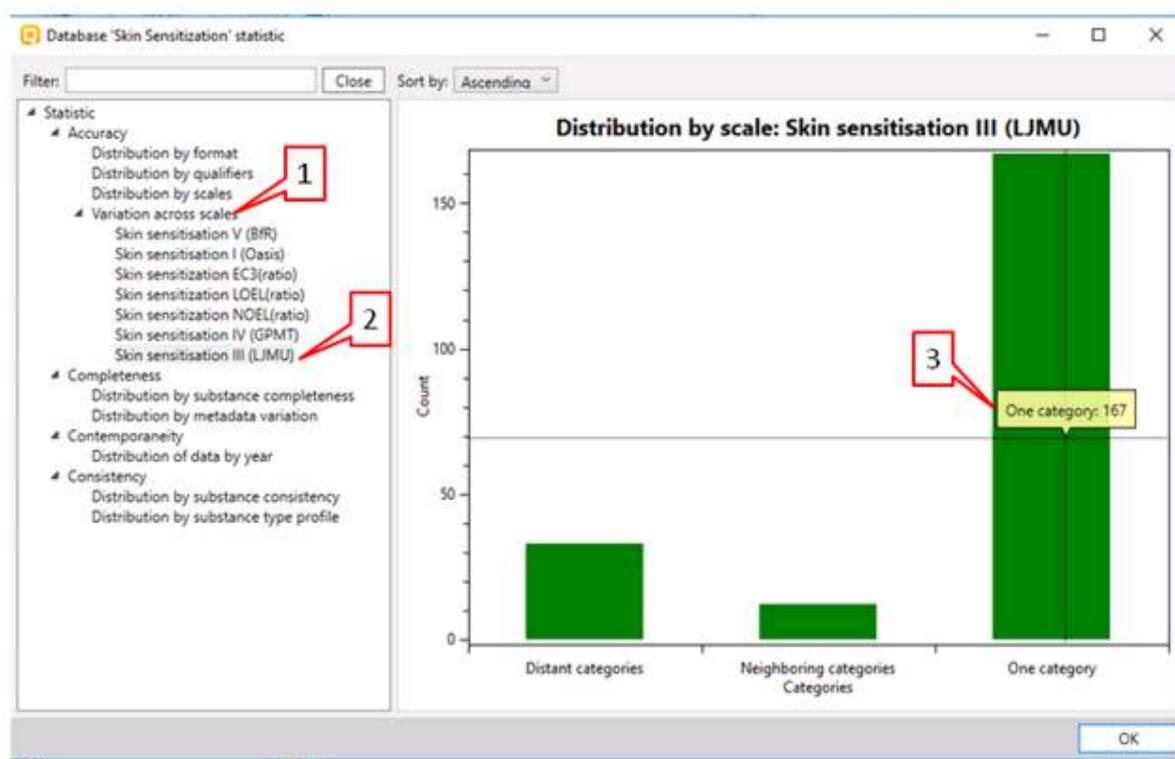


Figure 18

2) Completeness

- *Substance completeness* - Inform the user for the availability and relations of the general substance identifications - CAS RN and SMILES in the selected database. Follow the steps (Figure 19):
 - o Open Completeness level (1)
 - o Click on Distribution by substance completeness (2)
 - o Click on the bar (e.g. Yes: CAS/SMILES) to see the respective number of chemicals having CAS with SMILES (3)



Figure 19

- *Metadata completeness* - Indicate how much of the data possess additional information, e.g. used test organism species. The user could select what type of metadata they want to see after selection of one or more target endpoints. When the needed information is not available in the database there is a column marked as "Undefined". In this case study distribution by assay across SS data will be illustrated. Follow the steps (Figure 20):
 - o Open Completeness level (1)
 - o Click on Distribution by metadata variation (2)
 - o The window "Select endpoint" appears (3)
 - o Open Human health hazard to the level of skin sensitization (4)
 - o Select all available endpoints (5)
 - o Open metadata filed and select "Assay" (6), then click "Add" button (7). Click OK button (8)

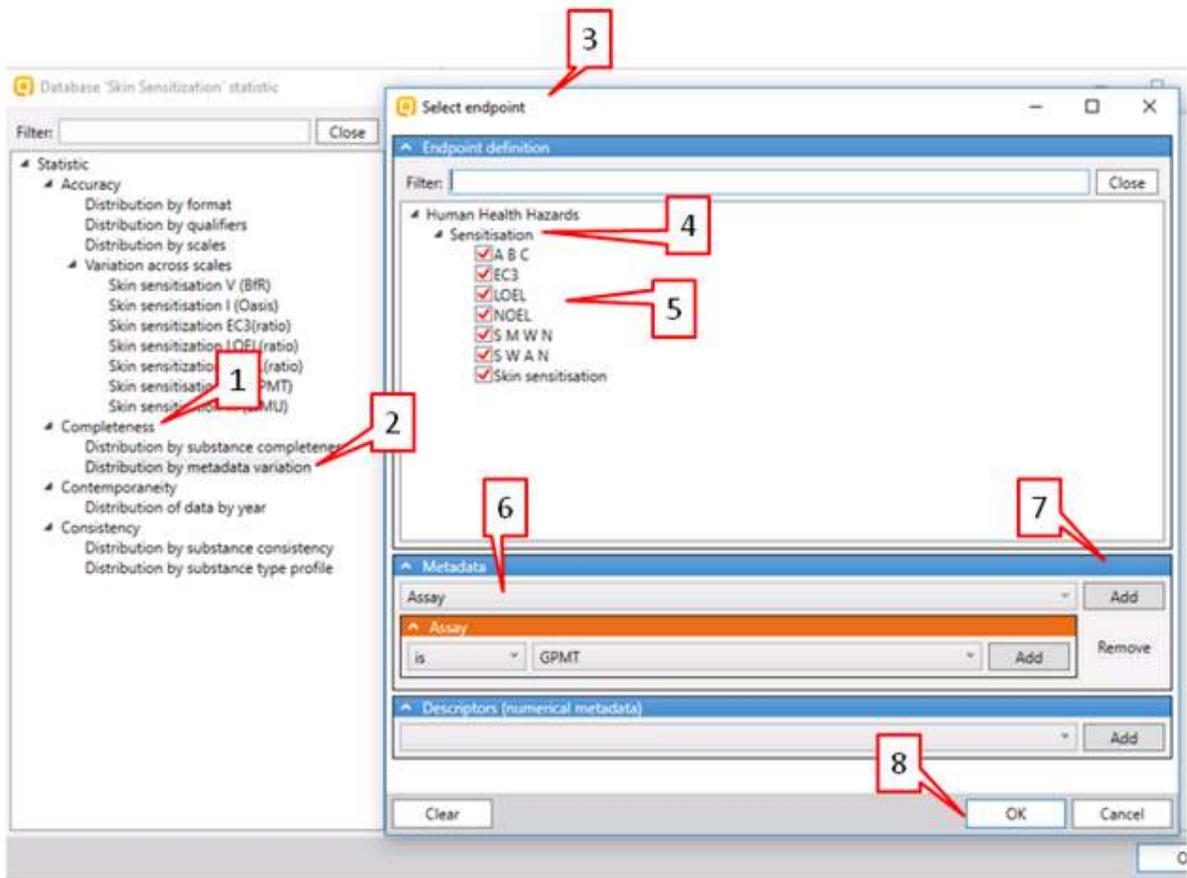


Figure 20

As a result the distribution is as follows, see Figure 21.

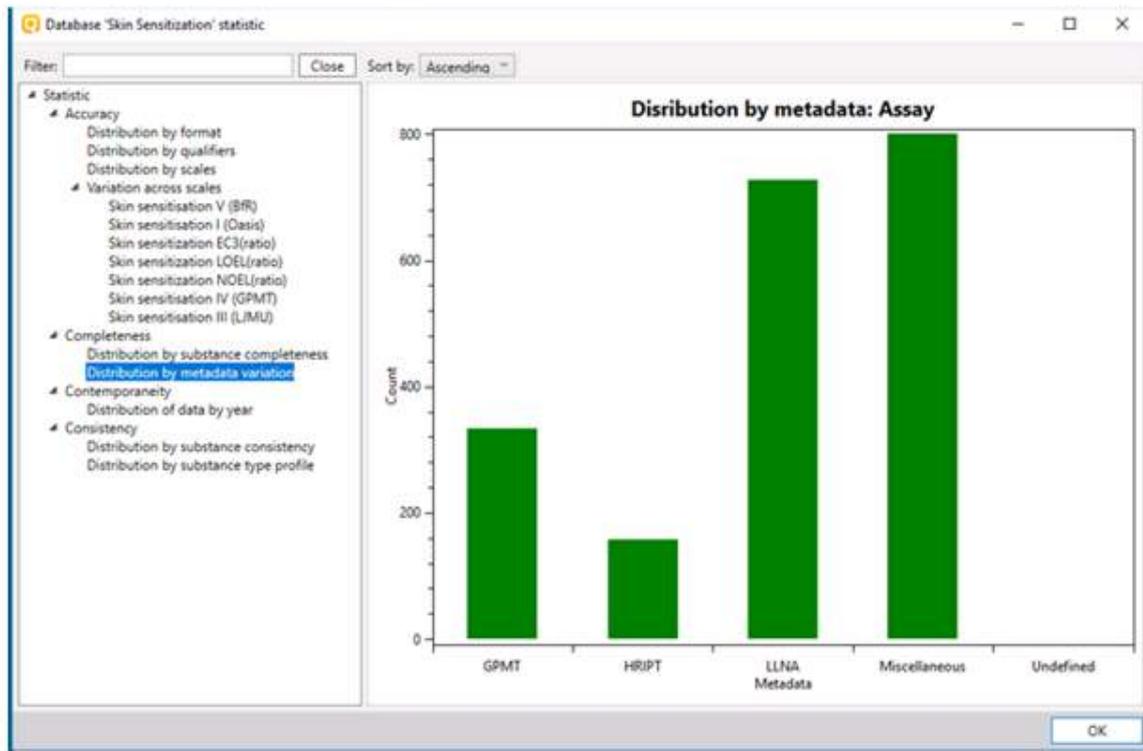


Figure 21

- Contemporaneity

The user could see all of the data in the selected database distributed according to their year of publication. Follow the steps (Figure 22):

- Open Contemporaneity level (1);
- Click on Distribution by year (2);
- Left-click on the bar to see the numbers (3).

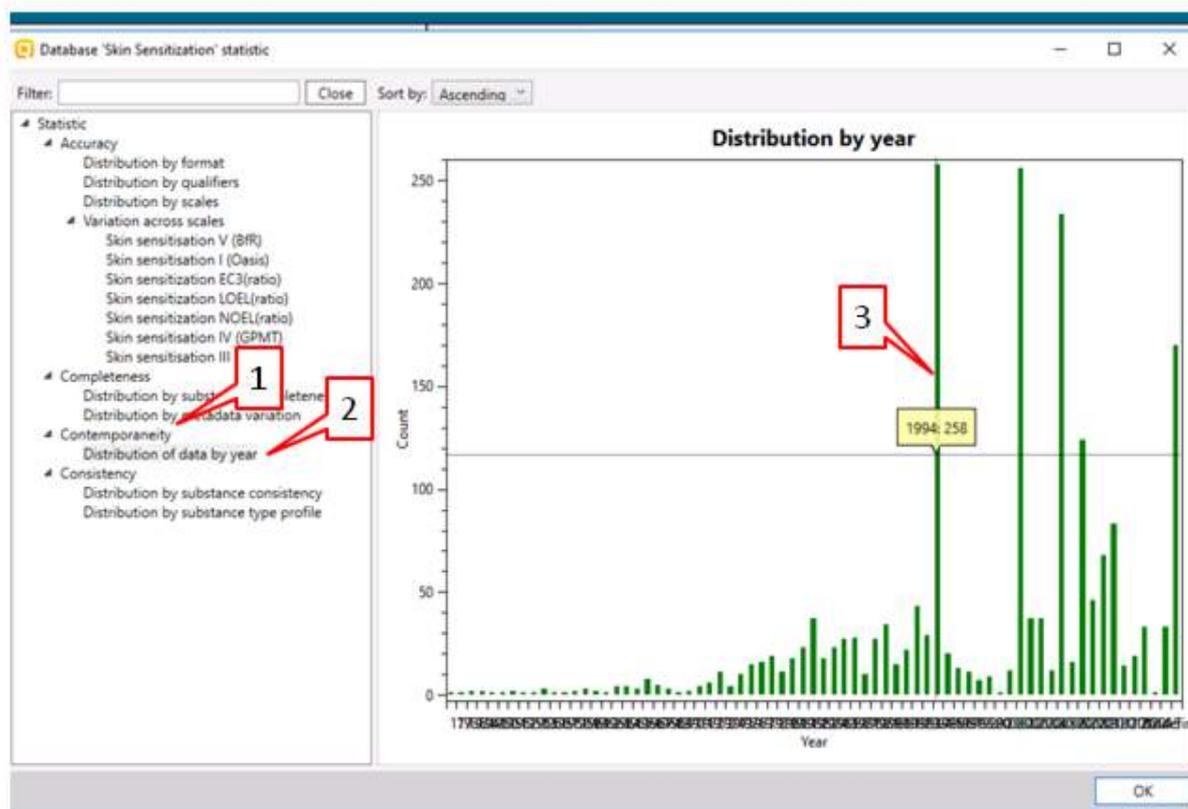


Figure 22

3) Consistency

- *Substance* - Provide information about how many of the SMILES notations have assigned CAS RNs from the current database. It gives indication for a presence of chemicals possessing CAS registry number, which correspond to more than one structure. Follow the steps shown on Figure 23:
 - o Click on Consistency level (1);
 - o Click on Distribution by substance consistency (2);
 - o 1236 CAS RN have associated SMILES (3).

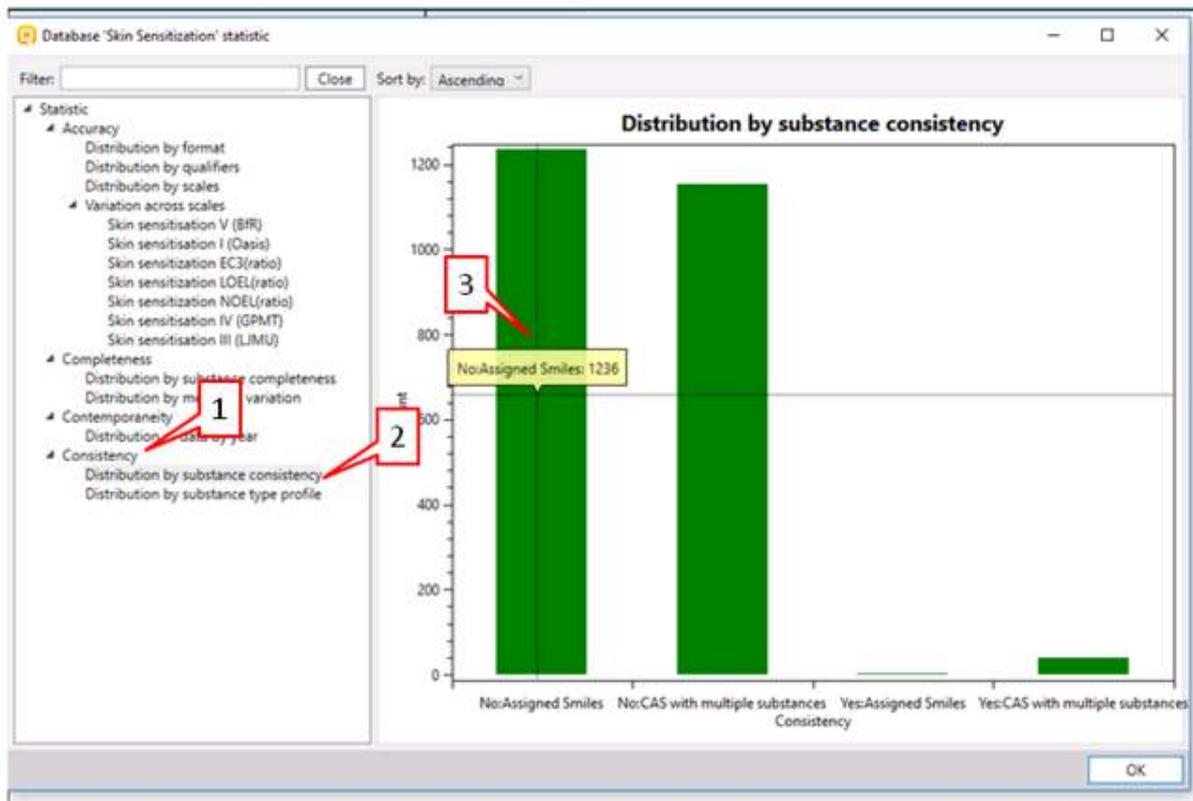


Figure 23

- Substance type profiler - The substance type statistic inform the user whether there are multi-constituent substances or UVCBs across the chemicals in the current database. Follow the steps shown on Figure 24:
 - o Click on Consistency level (1);
 - o Click on Distribution by substance type profiler (2);
 - o 1181 discrete chemicals are identified in Skin sensitization database (3).

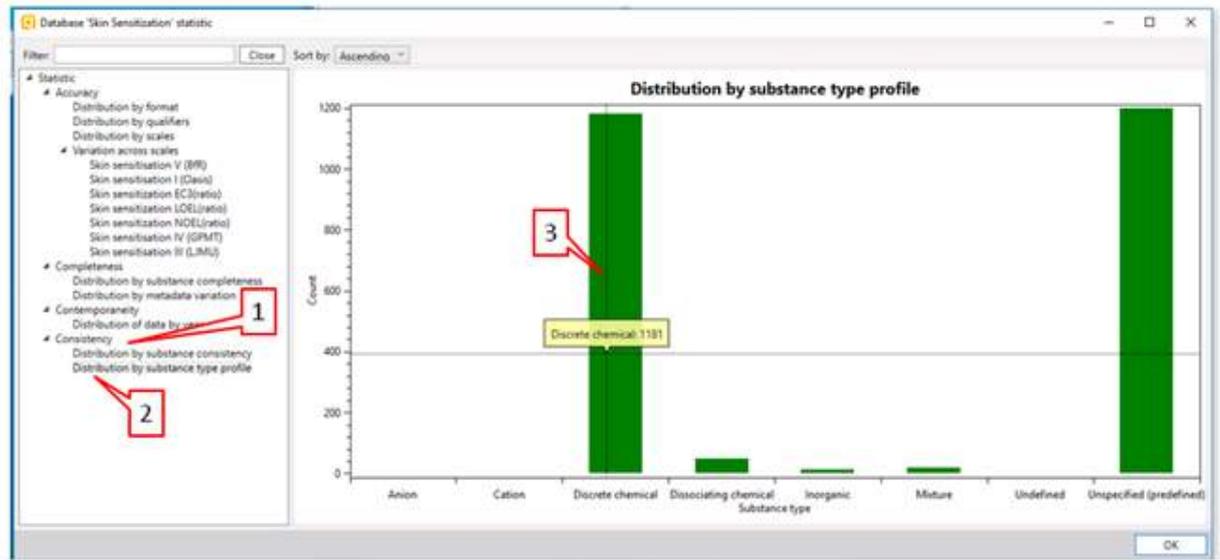


Figure 24

-O-

QSAR Toolbox v.4

D.3.2. Inventories

[Workflow](#) / [Endpoint](#) / D.3.2. Inventories

D.3.2. Inventories

The available inventories in Toolbox 4.0 are listed in [Table 2](#). The databases include chemicals with available experimental data, while the Inventories include chemicals ONLY. Chemicals in the inventories are presented with CAS, Chemical name and SMILES

[Table 2](#). List with databases available in Toolbox 4.0

Inventories	
NICNAS	EINECS
Canada DSL	HPVC OECD
COSING	MITI Japan
DSSTox	REACH_ECB
ECHA PR	TSCA
US HPV Challenge Program	

Same options are available to select, unselect or invert select as in [databases](#) are available. (Figure 1)



Figure 1

- 1- Select All – this selects all inventories
- 2- Unselect All – this unselects all inventories
- 3- Invert – inverts the last performed selection of inventories
- 4- About – displays short description for highlighted inventory. The user has to highlight inventory (1), then click on *About* button (2) and then the Short description window will appear (3) (Figure 2)

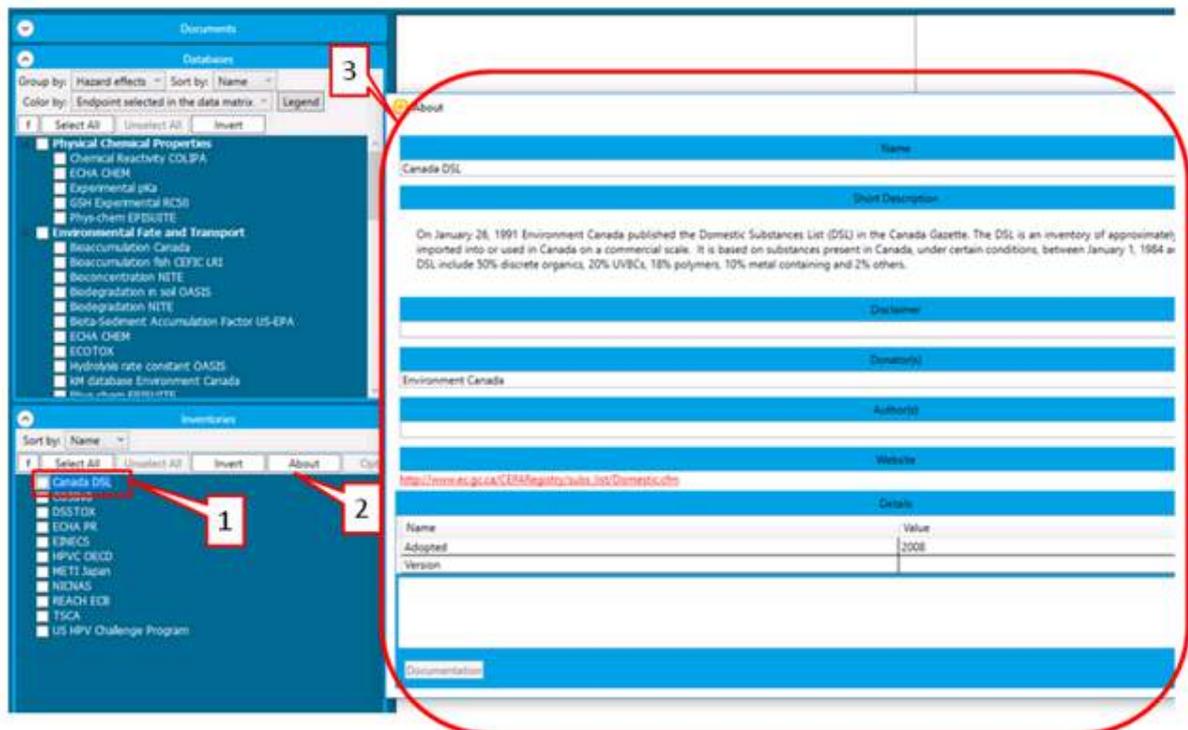


Figure 2

QSAR Toolbox v.4

D.3.3. Collecting data

[Workflow](#) / [Endpoint](#) / D.3.3. Collecting data

D.3.3. Collecting data

Data gathering can be executed in one of two basic ways:

- *Collecting all data for all endpoints.* The user has to click on Select all button (1), then all databases are selected (2), then click on Gather (3) button (Figure 1). All available measured data will be extracted for chemical(s) loaded on data matrix (4). The number of data points for chemicals with experimental data is also shown (5). This process however could be extremely time consuming. A more effective approach will be to narrow down the required databases and endpoints before the execution of the gather data query.

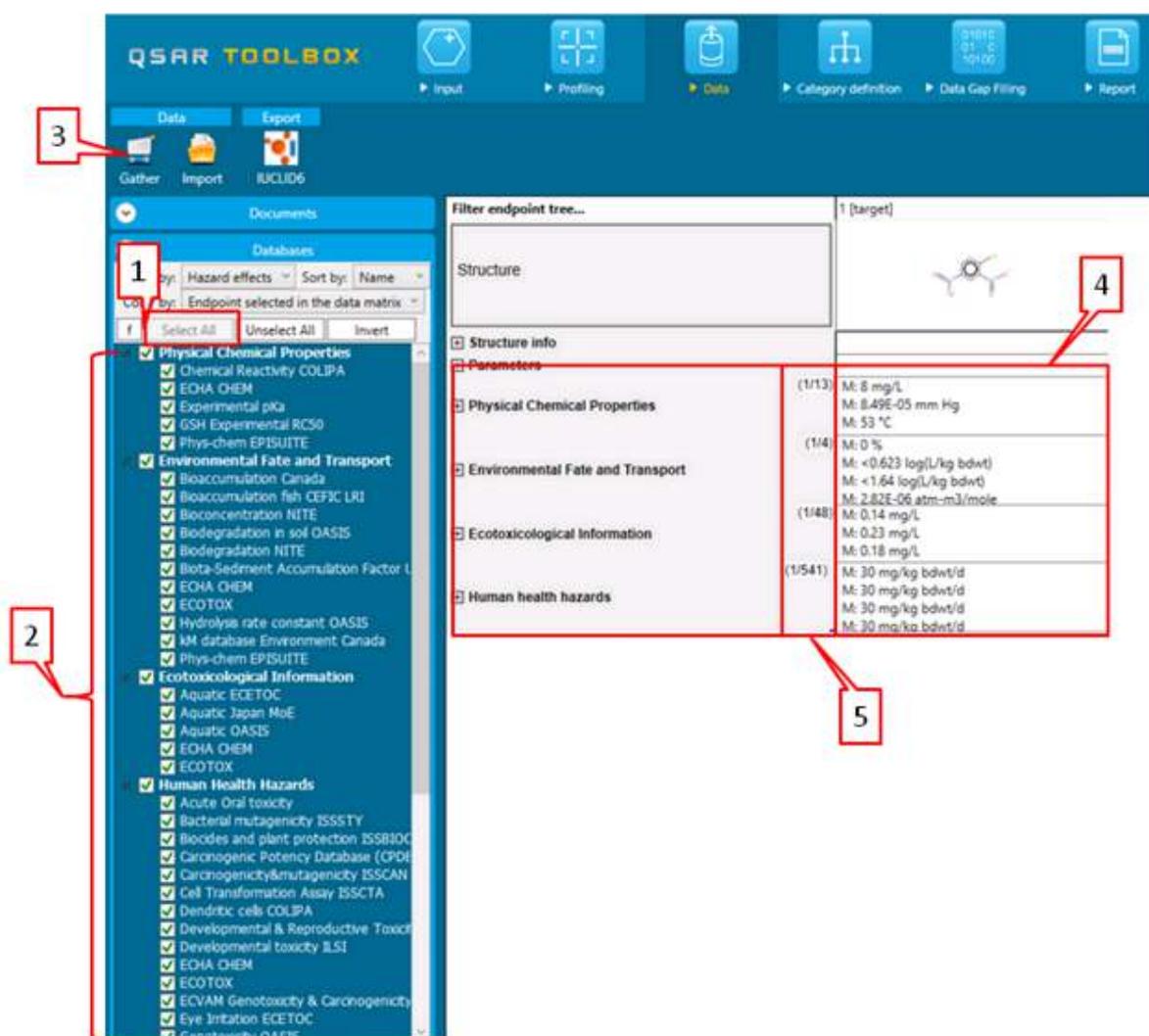


Figure 1

- On a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints): select databases relevant with examined endpoint. The user has to select database(s) related with investigated endpoints (1), secondly he/she has to click on Gather (2) button (Figure 2). Before loading of measured data for chemical(s) on the data matrix, a dialogue window *Read data ?*, prompting the user to select between *all endpoints* (3) or to choose an endpoint (4). Once the relevant endpoints are selected, the user has to click on *OK* (6).

All measured data associated with the selected endpoints for the target chemical will be displayed in the data matrix (7).

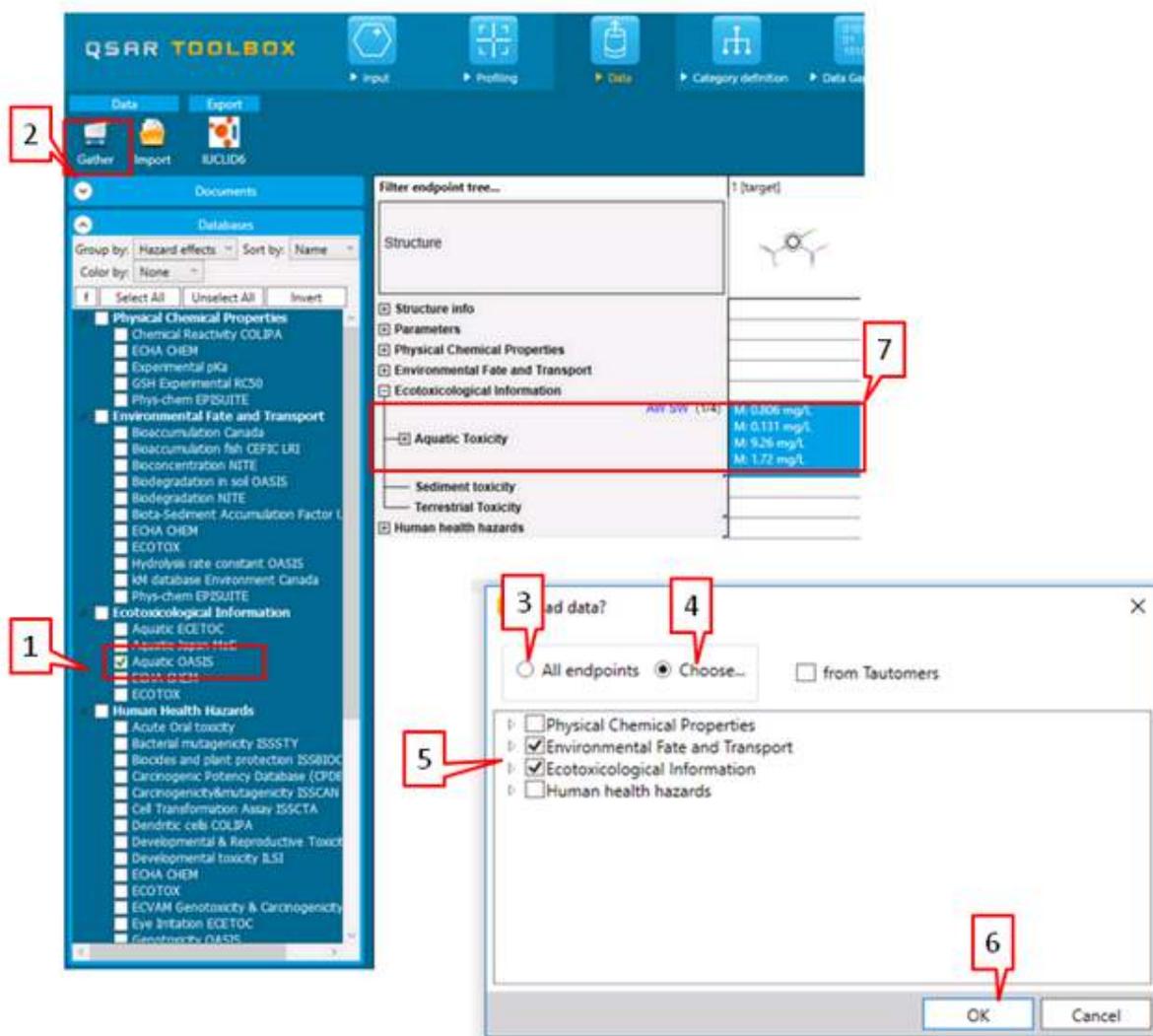


Figure 2

Notes:

1. The Toolbox databases include chemicals with experimental data, while the inventories include only chemicals without experimental data.
2. The databases are profiled and calculated in advance and the results are stored in a cache. This is done in order to accelerate the process of searching analogues with experimental data (defining categories).

-0-

QSAR Toolbox v.4

D.3.4. Importing data

[Workflow](#) / [Endpoint](#) / D.3.4. Importing data

D.3.4. Importing data

The two types of import are available via excel file and tab separated file. The procedures in both cases are the same.

Toolbox 4.0 operates with the following data type (Figure 1):

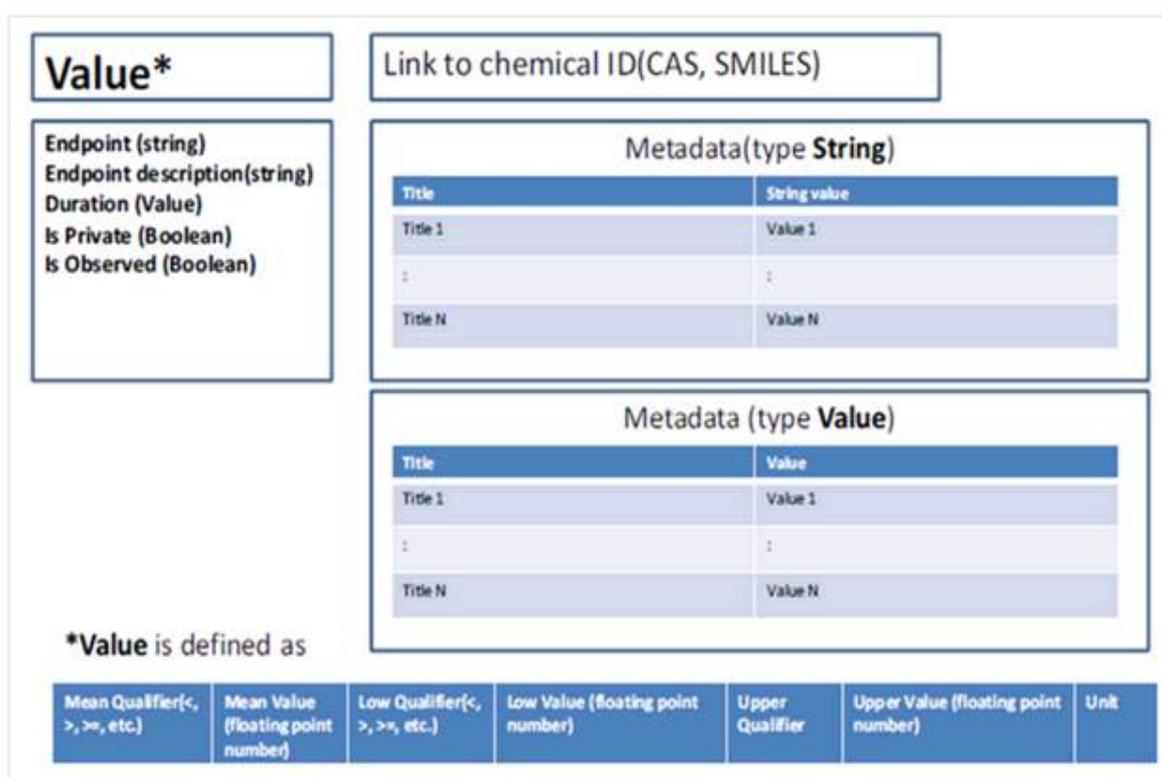


Figure 1

The Import's function is to translate the information in a file (be it XLS or TSF), separate it in different chunks (see the figure above) and write them to the database. The information comprises of connected chemical, numerical and meta-data. In other words the point of the import is to define a list of data points (the number that the user sees in the data-matrix and uses for gap-filling) with its corresponding metadata, namely the additional information on duration, test organisms, endpoint etc.

-0-

D.3.4.1. Vertical Import: Database: Database

D.3.4.1. Vertical Import

Vertical import is appropriate for a set of chemicals with consistent experimental data and the same supporting information (e.g. endpoint, test organism, test condition, author etc.).

The following steps has to be performed in order to import a custom database:

- Click on *Import* button (1) (Figure 1). The import window is displayed (2). Press Open file (3) and select database *Vertical import _BOD and Ames.xlsx* from the example files in Common files/QSAR Toolbox 4/Config/Examples.

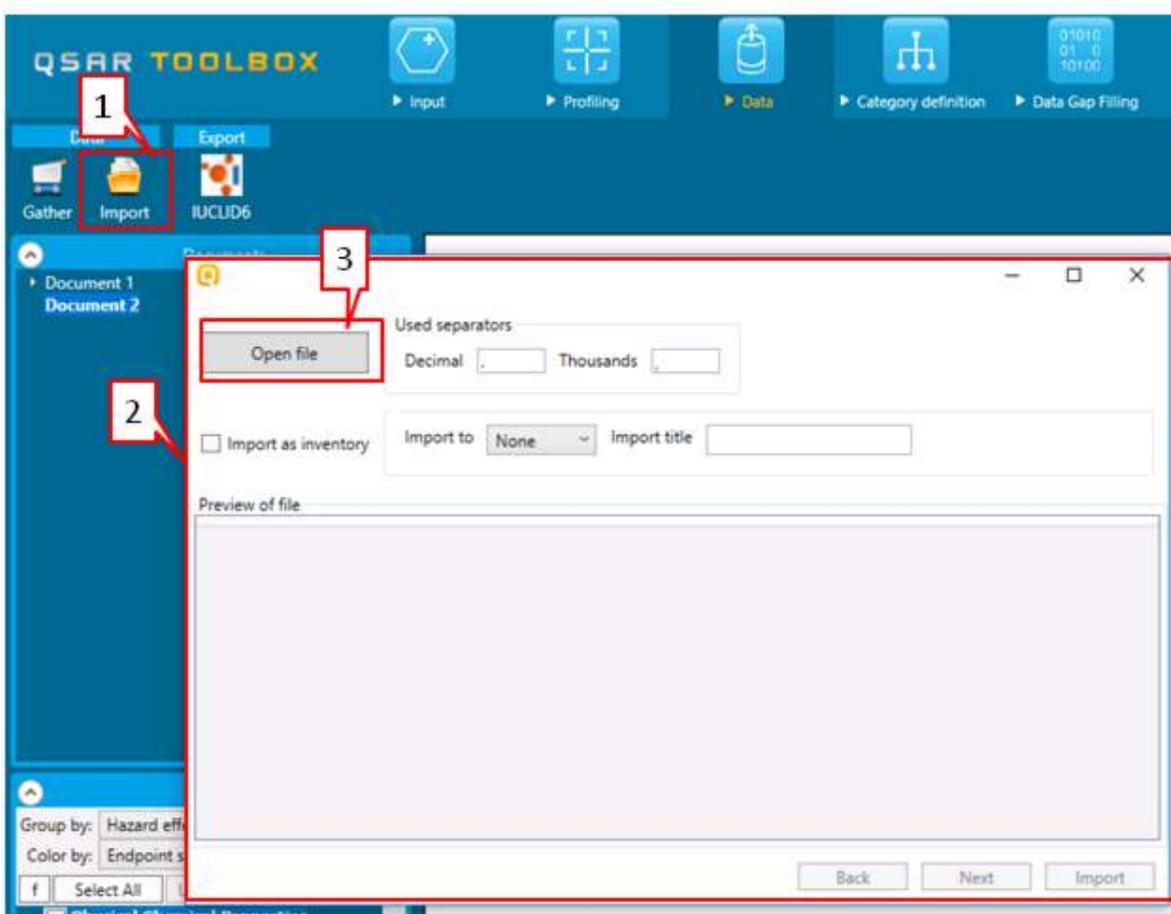


Figure 1

- The name of the database is written in (1) (Figure 2). The information from the file is shown in titled columns (2). Then press *Next* (3).

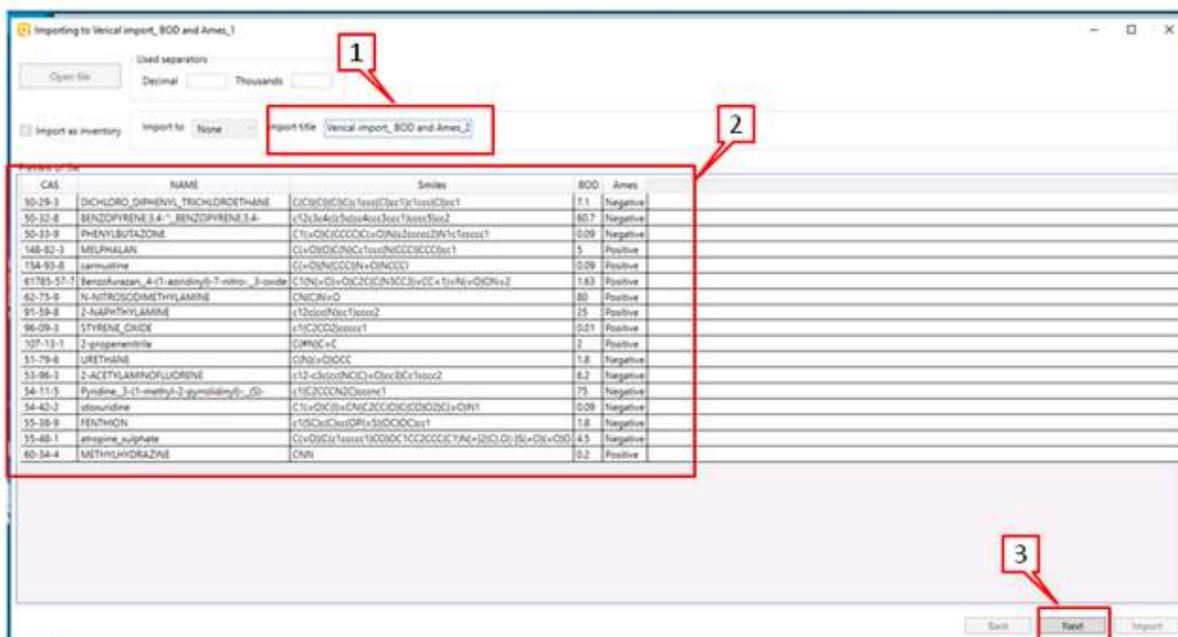


Figure 2

- In Figure 3 select Vertical radio button (1). Define the relevant endpoints associated with BOD (Biological oxygen demand) and Ames (Bacterial reverse mutation assay) (2) by clicking individually on each *No endpoint selected* button under the endpoints names. (3)

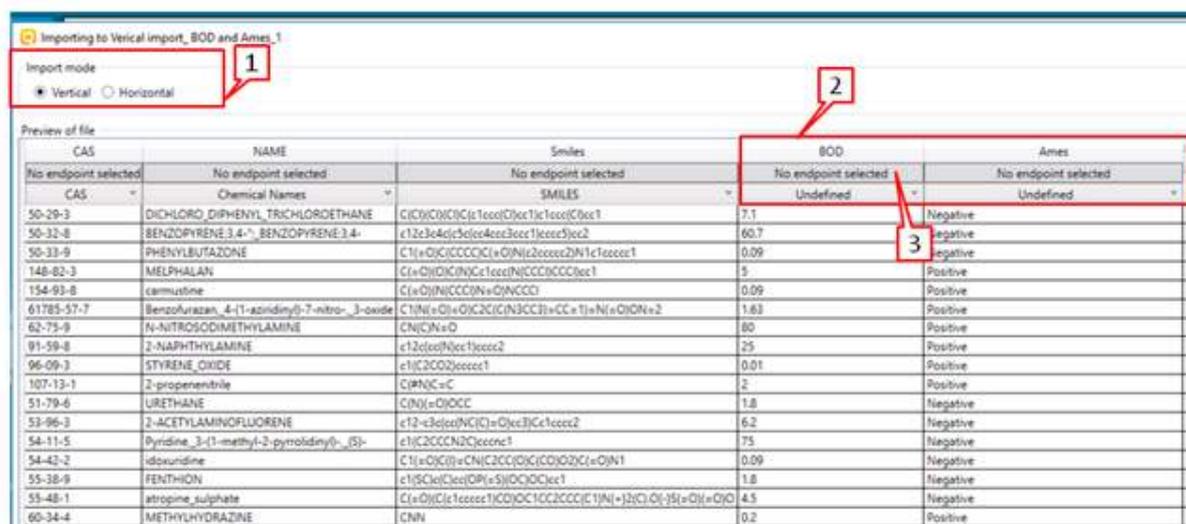


Figure 3

- **BOD**: Define endpoint window is displayed. The user has to select Ready Biodegradability (1) from the endpoint tree, then to select the family

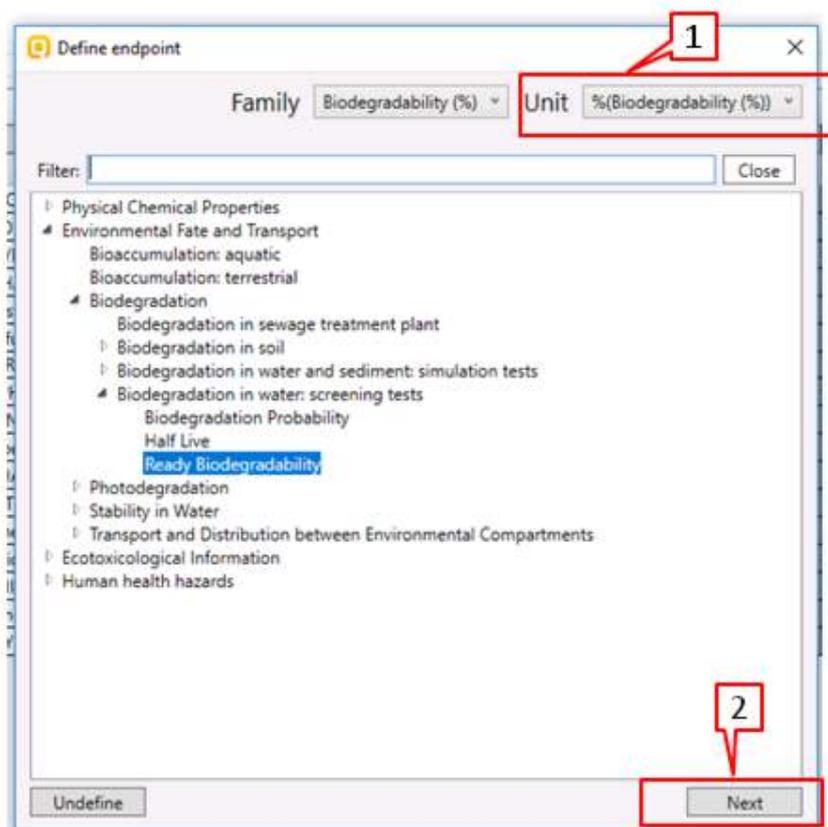


Figure 5

On the next window the user has to select BOD endpoint (1) from the drop-down menu (Figure 6). Then from the drop-down menu of additional metadata (2), they can select, for example, test guideline.

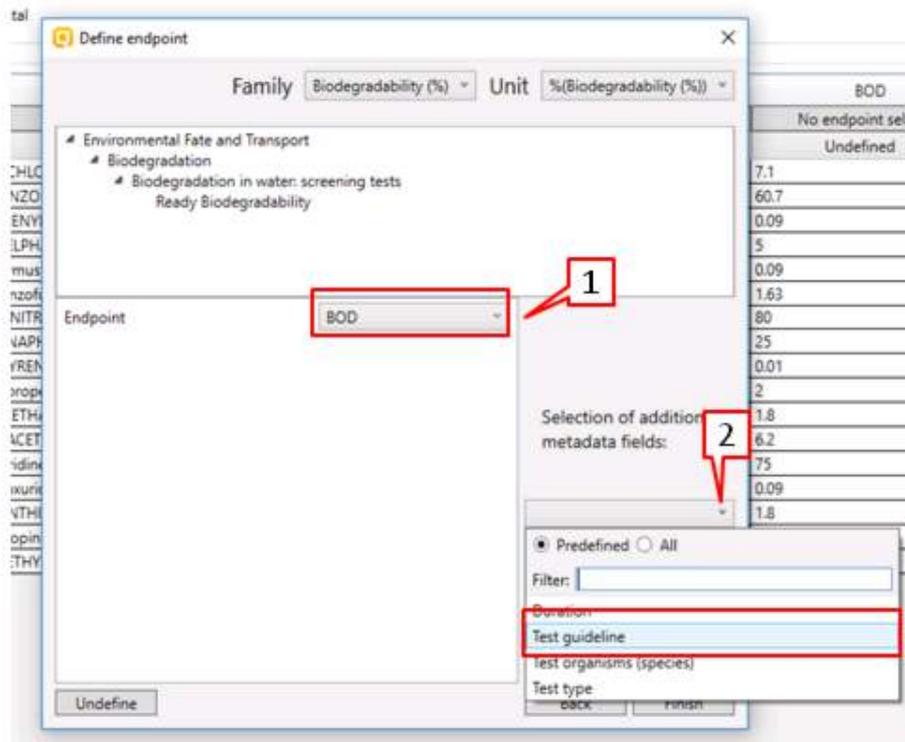


Figure 6

The user has to press *Add* button (1) and then from the drop-down menu to select a test guidance ,e.g. 301 C(2). (Figure 7)

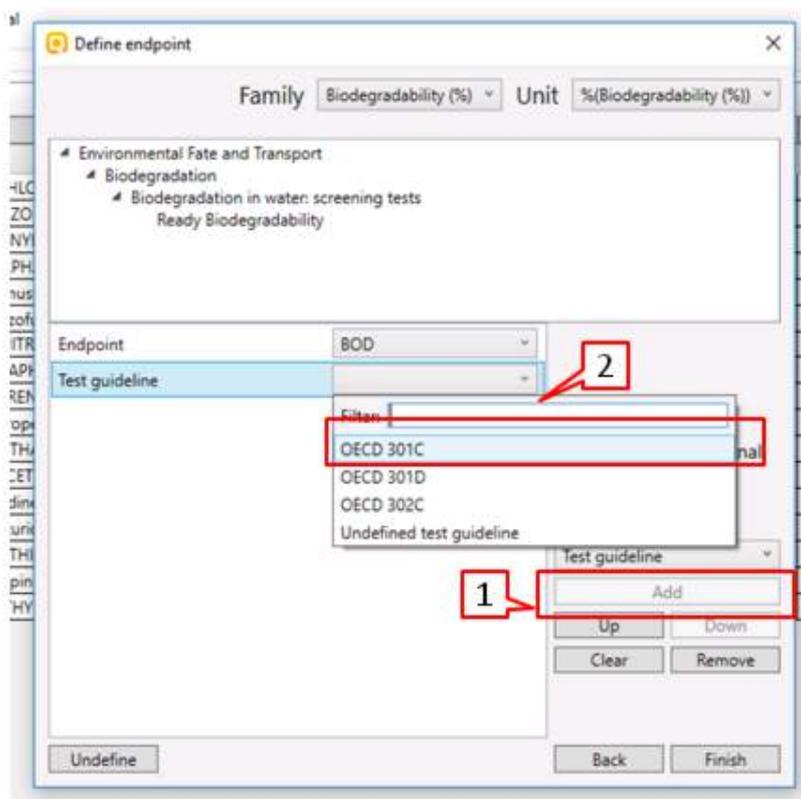


Figure 7

Once all the data fields are filled in the user has to press *Finish* (1) (Figure 8).

Toxicity (1) from the endpoint tree, select the family (Gene mutation I) by using the filter options and finally to press *Next* (3) (Figure 10)

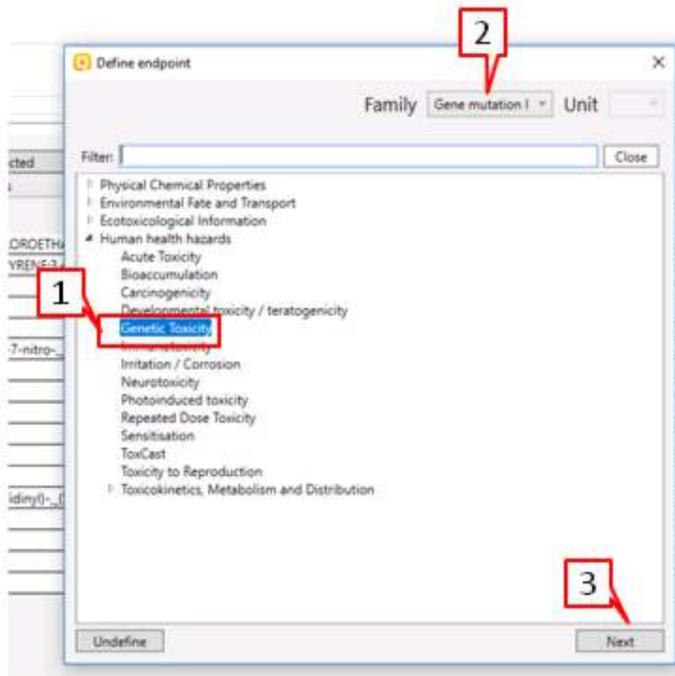


Figure 10

On the next window the user has to select Gene mutation endpoint (1) from the drop-down menu (Figure 11). Then from the drop-down menu of additional metadata (2), they can select other metadata fields and then have to press *Add* (3). Once all the data fields are filled in the user has to press *Finish* (4)

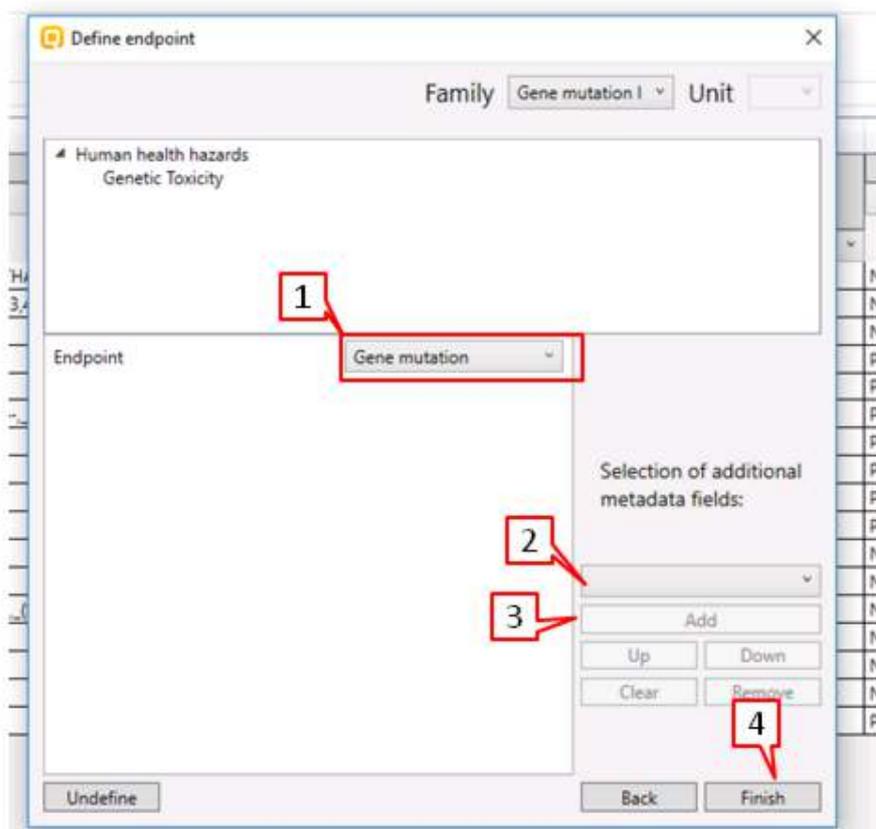


Figure 11

The user edited fields will be visualized in the main table (1) (Figure 12). Finally, the user has to press *Import* (2).

CAS	NAME	Smiles	BOD	Ames
No endpoint selected	No endpoint selected	No endpoint selected	Endpoint=BOD Test guideline=OECD 301C	Endpoint=Gene mutation
CAS	Chemical Names	SMILES	Undefined	Undefined
50-29-3	DICHLORO-DIPHENYL-TRICHLOROETHANE	ClC(Cl)(Cl)C1=CC=CC=C1C1=CC=CC=C1	7.1	Negative
50-32-8	BENZOPYRENE 3,4'-, BENZOPYRENE 3,4-	c12c3c4c5c6c4cc3ccc1ccc2	60.7	Negative
50-33-9	PHENYLBUTAZONE	C1=C(O)C(CCCOC)=O(N)C2=CC=CC=C1C2	0.09	Negative
148-82-3	MELPHALAN	C=C(O)C(N)C=C1C=CC(N)C=CC1	5	Positive
134-93-6	carbamazine	C=C(O)N(CCC)N=C(N)C=O	0.09	Positive
61785-57-7	Benofurazan, 4-(1-aziridinyl)-7-nitro-, 3-oxide	C1N=C(O)C2C(C1N)C(=O)O2C3=CC=C(N)C=C3	1.63	Positive
62-75-9	N-NITROSODIMETHYLAMINE	CN(C)N=O	80	Positive
91-59-8	2-NAPHTHYLAMINE	c12ccc3ccccc3c1	25	Positive
96-09-3	STYRENE OXIDE	c1ccc2occc12	0.01	Positive
107-13-1	2-propenenitrile	C#NCC=C	2	Positive
51-79-6	URETHANE	CN(C)C=O	1.8	Negative
53-96-3	2-ACETYLAMINOFLORENE	c12c3cc(NC(C)=O)ccc3cc1	6.2	Negative
54-11-5	Pyridine, 3-(1-methyl-2-pyrrolidinyl)-, (E)-	c1cc2ccncc2c1	75	Negative
54-42-2	abouridine	C1=C(O)C(C)=CN(C2C(O)C(O)C2)C1=O	0.09	Negative
55-38-9	BENTHION	c1cc2cc(N)cc1	1.8	Negative
55-48-1	atropine sulphate	C=C(O)C1=C(C(=O)OC1C2C(=O)C(C)N)C3=CC=CC=C3	4.5	Negative
60-34-4	METHYLHYDRAZINE	CNN	0.2	Positive

Figure 12

The new database is displayed in the database list (1). Right-click menu (2) is implemented where the user can see the database statistics or deleted it.

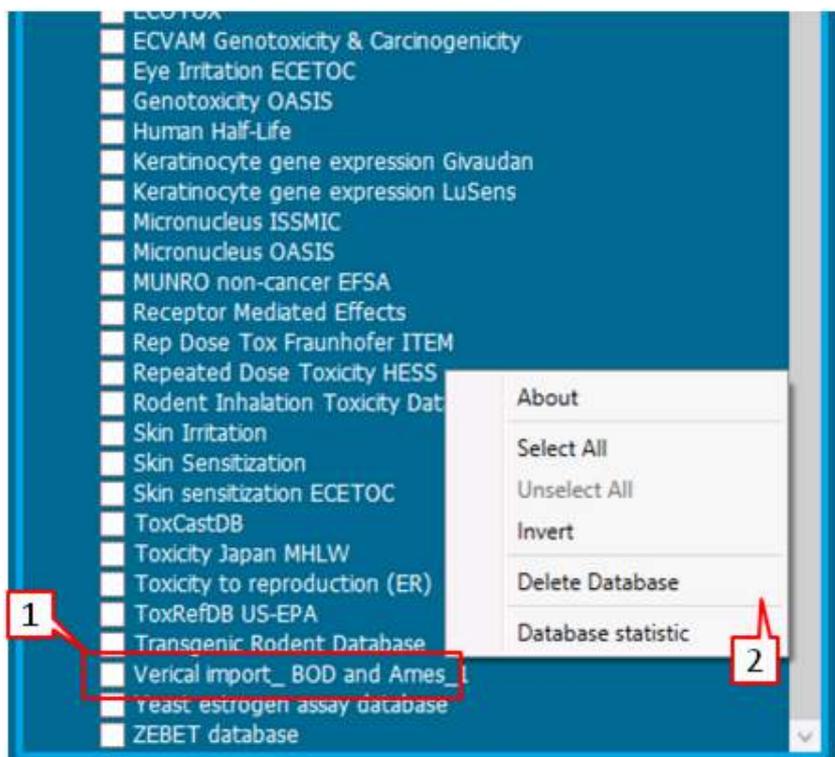


Figure 13

-0-

D.3.4.2. Horizontal Import: Database

D.3.4.2. Horizontal Import

Horizontal import: is appropriate for a set of chemicals with different types of experimental data accompanied with supporting information (endpoints, test condition, test organism, author etc).

Click on Import button (1) (Figure 1). The import window is displayed (2). Press Open file (3) and select database Horizontal import_Genotoxicity.xlsx from the example files in Common files/QSAR Toolbox 4/Config/Examples.

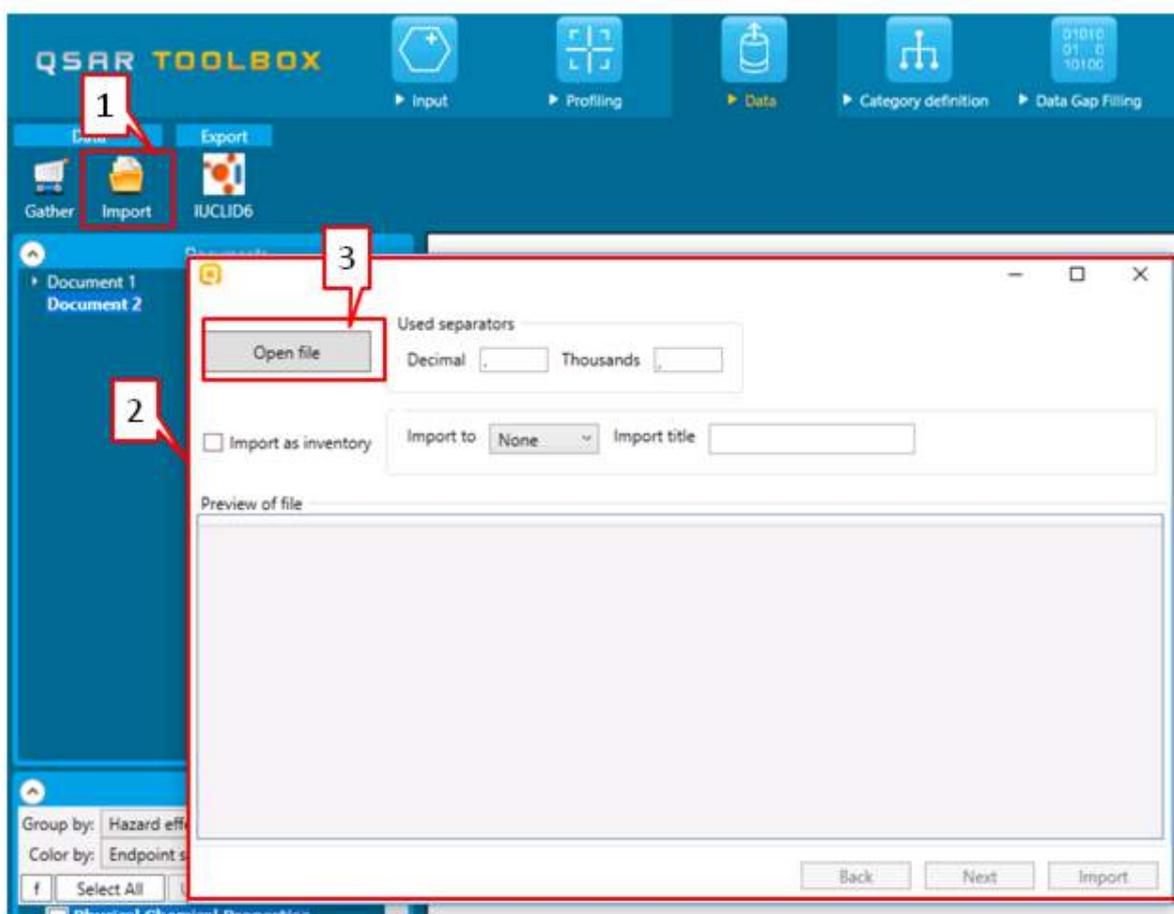


Figure 1

- The name of the database is written in (1). The information from the file is shown in titled columns (2). The user has to press Next (3). (Figure 2)

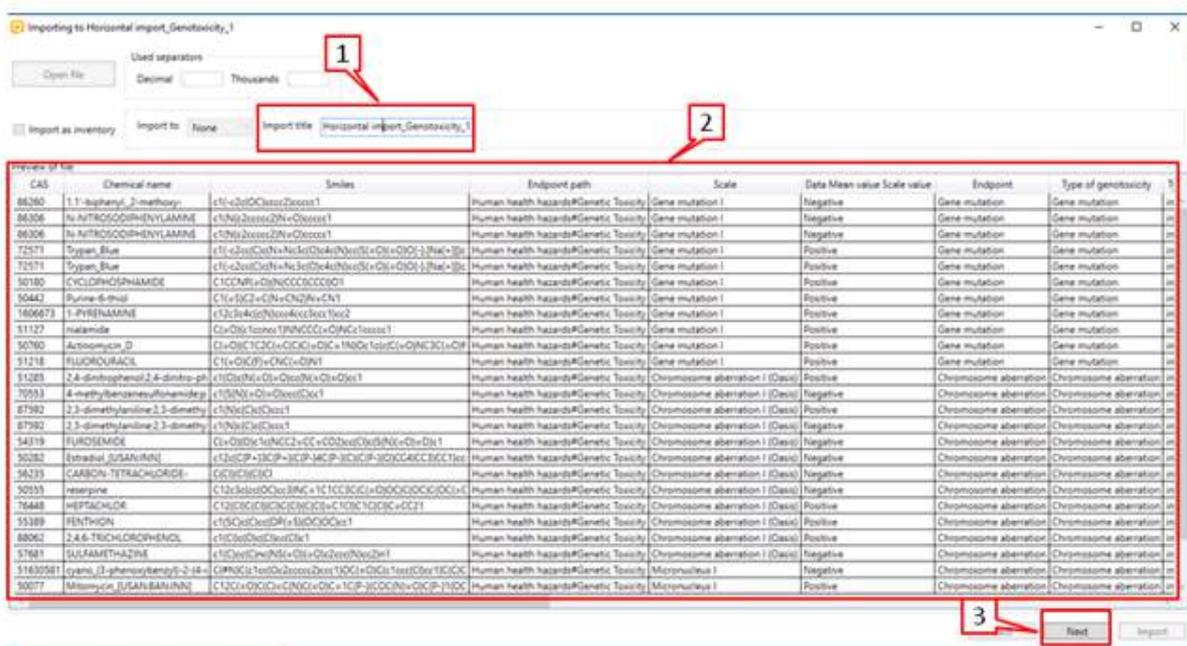


Figure 2

The user has to select Horizontal (1). Red-colored message are displayed above the table, informing the user if some metadata from the imported file is not mapped to a metadata field in Toolbox (2). The user has to manually map the metadata field by the drop-down menu under the metadata name (3). (Figure 3)

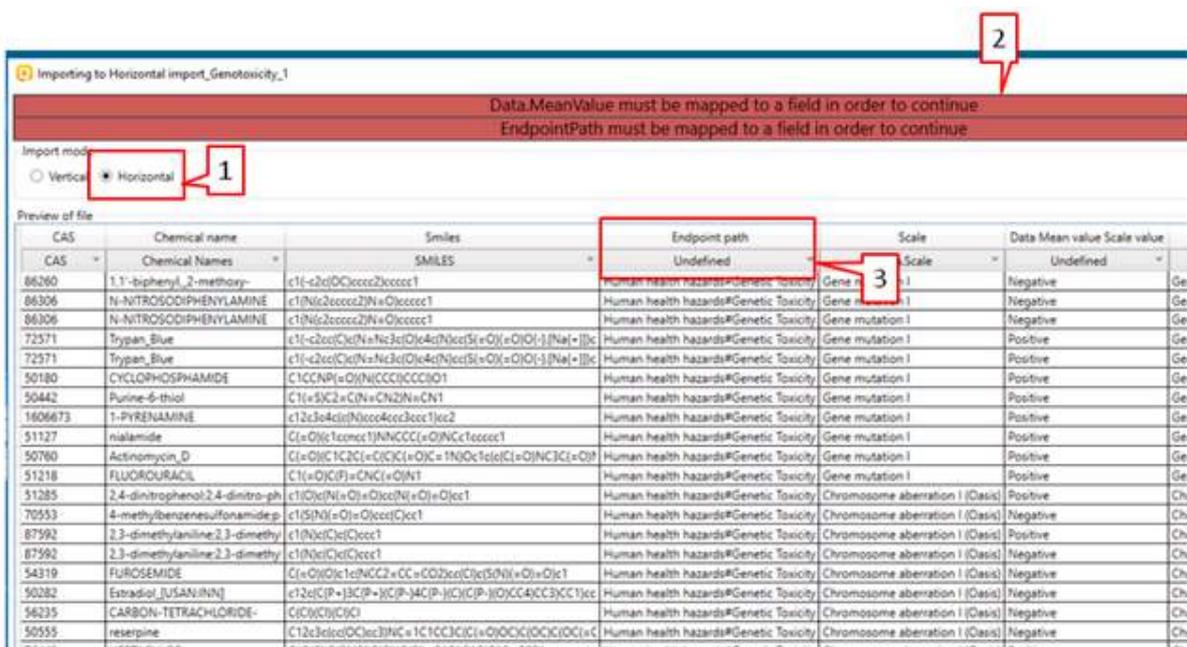


Figure 3

By using the implemented filter (1), find EndpointPath (2) in the drop-down menu and select it (Figure 4).

Data.MeanValue must be mapped to a field in order to continue
EndpointPath must be mapped to a field in order to continue

Initial

Chemical name	Smiles	Endpoint path	Scale	Data Mean value
Chemical Names	SMILES		Data.Scale	Undefine
benyl_2-methoxy-	<chem>c1(-c2c(OC)cccc2)ccccc1</chem>	Filter: end		Negative
OSODIPHENYLAMINE	<chem>c1(N(c2ccccc2)N=O)ccccc1</chem>	Endpoint		Negative
OSODIPHENYLAMINE	<chem>c1(N(c2ccccc2)N=O)ccccc1</chem>	Endpoint 2		Negative
Blue	<chem>c1(-c2cc(C)c(N=Nc3c(O)c4c(N)cc(S(=O)(=O)O(-).[Na(+)]))c</chem>	Endpoint assigned		Positive
Blue	<chem>c1(-c2cc(C)c(N=Nc3c(O)c4c(N)cc(S(=O)(=O)O(-).[Na(+)]))c</chem>	Endpoint comments		Positive
PHOSPHAMIDE	<chem>C1CCNP(=O)(N(CCC)CCC)O1</chem>	Endpoint explanation		Positive
6-thiol	<chem>C1(=S)C2=C(N=CN2)N=CN1</chem>	EndpointPath		Positive
NAMINE	<chem>c12c3c4c(c(N)ccc4ccc3ccc1)cc2</chem>	Gender		Positive
de	<chem>C(=O)(c1ccccc1)NNCCC(=O)NCc1ccccc1</chem>	Organism gender		Positive
tycin_D	<chem>C(=O)(C1C2C(=C(C)C(=O)C=1N)Oe1c(c(C(=O)NC3C(=O)H</chem>	Trend		Positive
JURACIL	<chem>C1(=O)C(F)=CNC(=O)N1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Positive
trophenol;2,4-dinitro-ph	<chem>c1(O)c(N(=O)=O)cc(N(=O)=O)cc1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
ylbenzenesulfonamide;p	<chem>c1(S(N)(=O)=O)ccc(C)cc1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Positive
ethylaniline;2,3-dimethy	<chem>c1(N)c(C)c(C)ccc1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
ethylaniline;2,3-dimethy	<chem>c1(N)c(C)c(C)ccc1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
IMIDE	<chem>C(=O)Oc1c(NCC2=CC=CO2)cc(C)c(S(N)(=O)=O)c1</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
il_[USAN;INN]	<chem>c12c(C(P+)=3C(P+))(C(P-)=4C(P-))(C(P-)=O)CC4)CC3)CC1)cc</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
N-TETRACHLORIDE-	<chem>C(Cl)(Cl)C(Cl)Cl</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative
te	<chem>C12c3c(cc(OC)cc3)NC=1C1CC3C(C(=O)OC)C(OC)C(OC)=C</chem>	Human health hazards#Genetic Toxicity	Chromosome aberration I (Oasis)	Negative

Figure 4

In the similar manner the user has to map Data Mean value (1 in Figure 5). Then they have to press Import (2).

Importing to Horizontal import, Genotoxicity, 1

Import mode
 Vertical Horizontal

Preview of file

CAS	Chemical name	Smiles	Endpoint path	Scale	Data Mean value	Scale value	Endpoint	Type of genotoxicity
CAS	Chemical Names	SMILES	EndpointPath	Data Scale	Data Mean/Value	Endpoint	Type of genotoxicity	
86260	1,1'-biphenyl-2-methoxy-	c1c1c2c3ccccc1	Human health hazards#Genetic Toxicity	Gene mutation	Negative	Gene mutation	Gene mutation	Gene mutation
86306	N-ETHYLOXYDIPHENYLAMINE	c1ccc(ccc2ccccc12)NCC	Human health hazards#Genetic Toxicity	Gene mutation	Negative	Gene mutation	Gene mutation	Gene mutation
86306	N-ETHYLOXYDIPHENYLAMINE	c1ccc(ccc2ccccc12)NCC	Human health hazards#Genetic Toxicity	Gene mutation	Negative	Gene mutation	Gene mutation	Gene mutation
72571	Thyran, Blue	c1c1c2c3ccccc12)NCC	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
72571	Thyran, Blue	c1c1c2c3ccccc12)NCC	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
30180	CYCLOPHOSPHAMIDE	C1CNCP1=ON1CCCCC1	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
30442	Purine-6-thiol	C1=NC2=C(N=CN2)N=C1S	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
1806873	1-PYRENAMINE	c1ccc2c(c1)ccc3ccccc23	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
31127	maleimide	C1=CC=C(C=C1)C(=O)N1C=CC1=O	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
30760	Actinomycin_D	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
31218	FLUOROURACIL	C1=CC=C(C=C1)N=C(N1)C(F)=O	Human health hazards#Genetic Toxicity	Gene mutation	Positive	Gene mutation	Gene mutation	Gene mutation
31285	2,4-dinitrophenol-2,4-dinitro-ph	c1ccc(cc1[N+](=O)[O-])[N+](=O)[O-]	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Positive	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
70553	4-methylbenzamide/Paracetamol	c1ccc(cc1)NC(=O)C	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
87392	2,3-dimethylmaleimide-2,3-dimethyl	c1ccc(cc1)C(=O)N1C=CC1=O	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Positive	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
87392	2,3-dimethylmaleimide-2,3-dimethyl	c1ccc(cc1)C(=O)N1C=CC1=O	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
34319	FUROSENADE	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
30282	Isradipin-3,5-DIETHYL-1,4-DIHYDRO-2,6-DIMETHYL-4-PYRIDINE	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
36235	CARBON- TETRACHLORIDE	ClC(Cl)(Cl)Cl	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
30555	Hexapone	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
76448	HEPTACHLOR	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Positive	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
35389	FENTHION	c1ccc(cc1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Positive	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
38062	2,4,6-TRICHLOROPHENOL	c1ccc(cc1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Positive	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
57861	SULFAMETHAZINE	c1ccc(cc1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Chromosome aberration (Oasis)	Negative	Chromosome aberration (Oasis)	Chromosome aberration	Chromosome aberration
11632581	cyano- (3-phenoxybenzyl)-2-thio-	C#NCC1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Micronucleus	Negative	Chromosome aberration	Chromosome aberration	Chromosome aberration
30077	Misonidazole/SANBANINN	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Micronucleus	Positive	Chromosome aberration	Chromosome aberration	Chromosome aberration
30293	DICHLORO-DIPHENYL-TRICHLORO-	C1=CC=C(C=C1)C(=O)N2C(=O)C1=CC=C(C=C1)N2	Human health hazards#Genetic Toxicity	Micronucleus	Positive	Chromosome aberration	Chromosome aberration	Chromosome aberration
30128	BENZOPYRENE-3,4,9-BENZOPYR-	c1ccc2c(c1)ccc3ccccc23	Human health hazards#Genetic Toxicity	Micronucleus	Positive	Chromosome aberration	Chromosome aberration	Chromosome aberration
66366	1,1'-Azobis(2-methyl-2-propanol)	CN(C)C(=O)N(C)C	Human health hazards#Genetic Toxicity	Gene mutation	Negative	Gene mutation	Gene mutation	Gene mutation

Back Next Import

Figure 5

A message is displayed that the import is successful (Figure 6) . The user has to press OK (1).

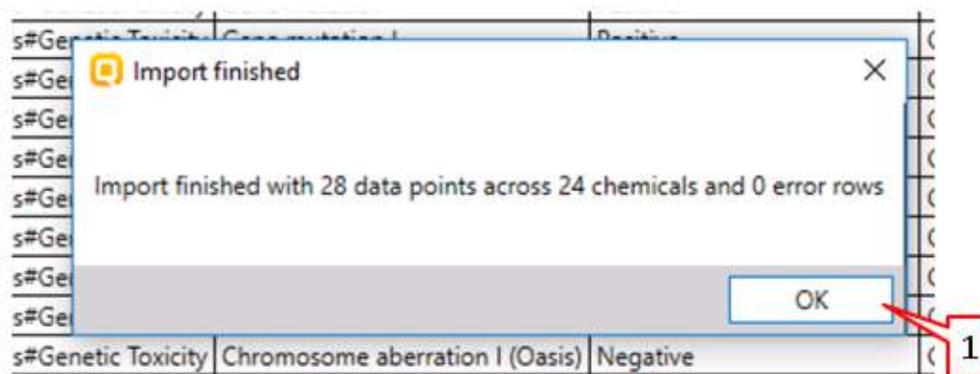


Figure 6

The newly imported database is now included in the list of databases (Figure 7). Right-click menu (2) gives the user several options including database statistics and deletion of the database.

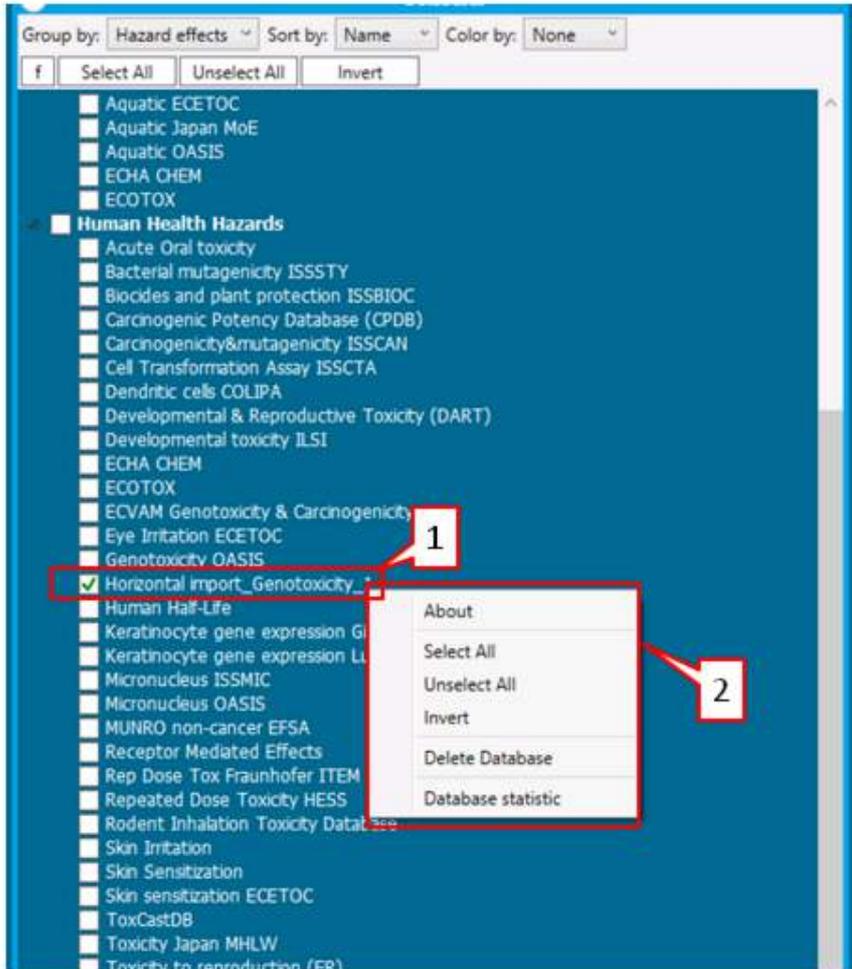


Figure 7

D.3.4.3. Importing Inventory

D.3.4.3. Import of inventory

The following steps need be performed by the user in order to import an inventory:

-- Select *Import* (1), tick *Import as inventory* (2) check box, click on Open file (3), select the inventory (e.g. from Common files/QSAR Toolbox 4/Config/Examples) and then press Open (5) (Figure 1).

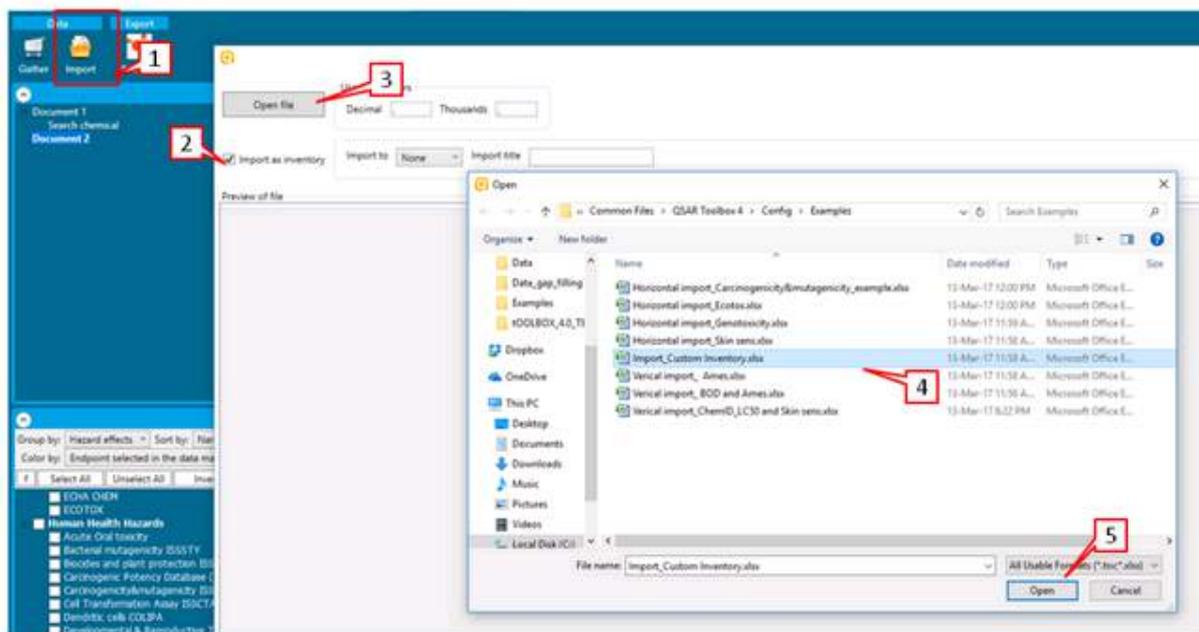


Figure 1

- The import title is written (1) as the content of the imported file is depicted in (2). Then the user has to press *Next* (3). (Figure 2)

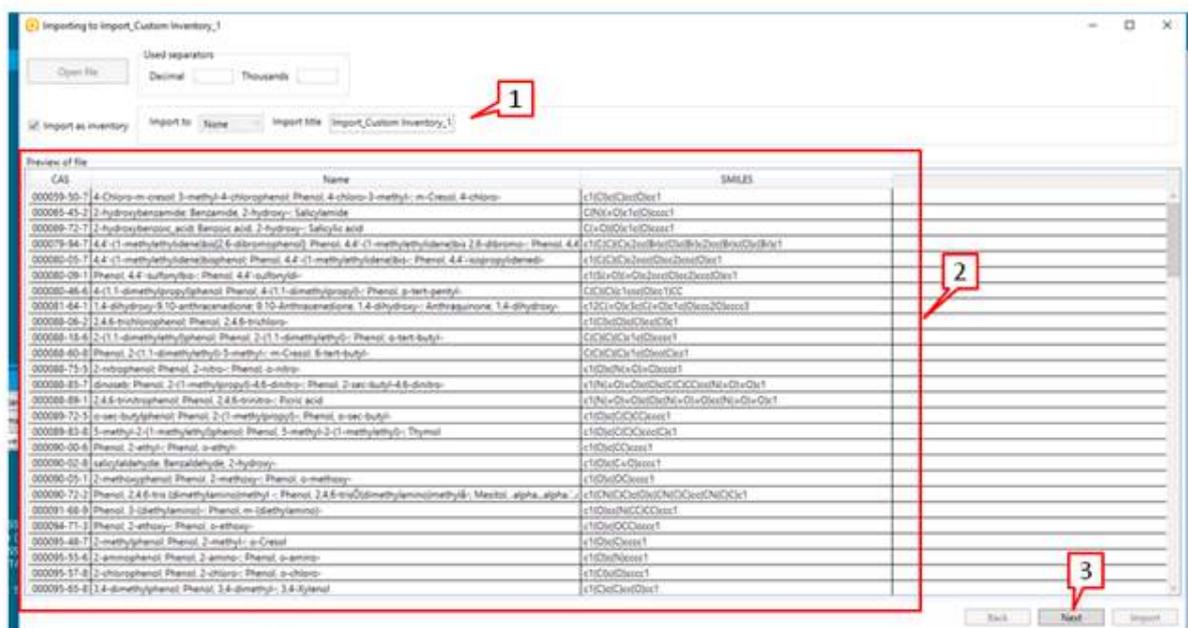


Figure 2

- Click on *Import* (1) and then *OK* (2) on the message confirming that the import is successful. (Figure 3)

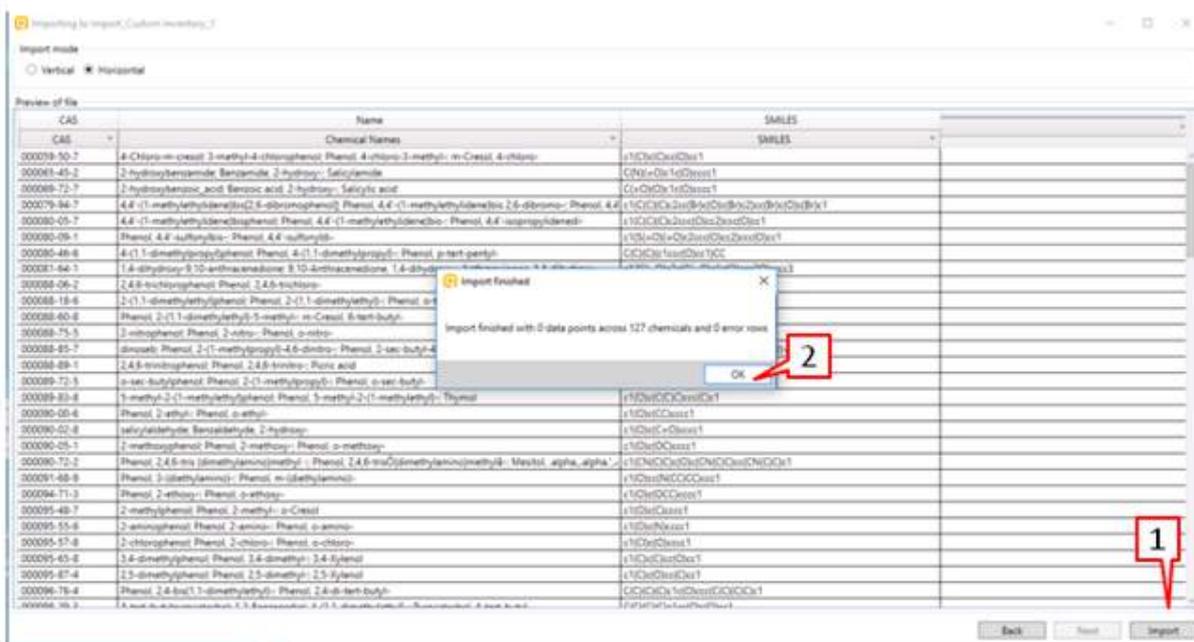


Figure 3

- The newly imported inventory is included inventory list (1) (Figure 4). Right-click menu provide delete options and inventory statistic (2).

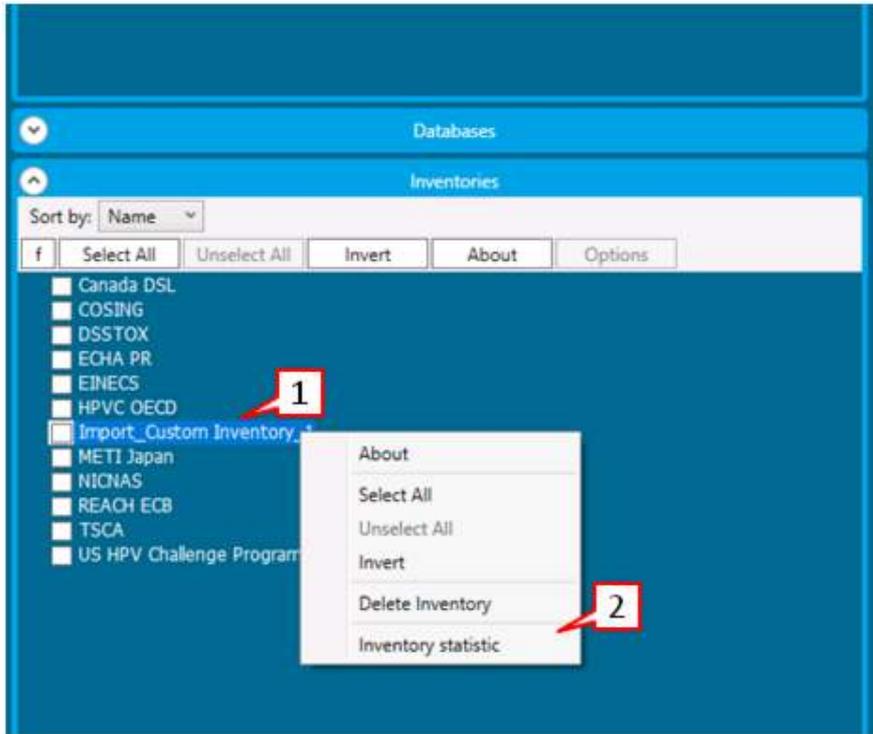


Figure 4

D.3.4.4. Supporting information

D.3.4.4. Supporting information

Some of the most important endpoints are represented in the table below. The information in the table aims to facilitate the users work with importing of new databases.

Hazard effects	Endpoint	Unit family	Unit member (e.g. mg/l)
Physical chemical properties	Boiling point	Temperature	°C;°F; K
	Dissociation constant (pKa)	None	
	Melting/freezing point	Temperature	°C;°F; K
	Partition coefficient: log Kow	None	
	Surface tension	None	mN/m
	Vapour pressure	Pressure	Pa; mmHg; Torr
	Water solubility	Mass concentration	mg/L
Environmental Fate and Transport	Bioaccumulation (e.g. BCF, BAF)	Bioaccumulation	L/kg bdwt
	Biodegradation (BOD)	Biodegradability (%)	%
	Biodegradation (BOD)	Biodegradability (categories)	Ready; Not ready
	Biodegradation (BOD)	Biodegradability (EPI)	Readily Degradable; Not Readily Degradable
Ecotoxicological information	Acute aquatic toxicity (LC50, EC50)	Mass concentration	mg/L; ppm
	Terrestrial Toxicity (EC50, NOEC)	Mass concentration	mg/L; ppm
		Volume concentration	gal/100gal
Human health hazards	Acute toxicity (LD50)	Mass fraction	mg/kg
	Gene mutation	Gene mutation I	positive
			equivocal
			negative
	DNA and protein damage		
	Chromosome aberration	Chromosome aberration I (Oasis)	positive
			negative
	Summary carcinogenicity	Carcinogenicity (ISSCAN)	positive
			equivocal
			negative
	Repeated dose toxicity (LOEL, NOEL)	Administrated dose (mass)	mg/kg bdwt/d
Skin sensitization (all type endpoints)	Skin sensitization II (ECETOC)	positive	
		negative	

	Skin sensitization (EC3)	Skin sensitization EC3(ratio)	%
		Skin sensitisation (Oasis)	Strongly positive
			Weakly positive
			Negative
	Skin sensitization (SMWN)	Skin sensitization (GPMT)	Strong sensitizer
			Moderate sensitizer
			Weak sensitizer
			Non sensitizer
	Skin sensitization (LOEL, NOEL)	Skin sensitization LOEL (ratio)	$\mu\text{g}/\text{cm}^2$
	Skin sensitization (ABC)	Skin sensitization (BfR)	Category A, category B, category C
	Skin sensitization (SWAN)	Skin sensitization (LIMU)	Strongly sensitizing
			Moderately sensitizing
			Ambiguous
			Not sensitizing

-0-

QSAR Toolbox v.4

D.3.5. Exporting data: IUCLID 6

[Workflow](#) / [Endpoint](#) / D.3.5. Exporting data: IUCLID 6

D.3.5. Exporting dataEXPORT of data from the QSAR Toolbox to IUCLID 6

Selected Toolbox predictions can be directly assigned to substances in the IUCLID 6 database. The QSAR Toolbox allows users to export predicted data (by means of the Data Gap Filling tools) to IUCLID 6. This can be done either offline, by creating an *.i6z file which can then be imported into an IUCLID 5 database, or online, by directly connecting to an IUCLID 6 server (via WebServices) and assigning the predicted endpoint data to a selected substance. The second option will be described in here.

There are two ways to invoke the IUCLID 6 Export Wizard. In either case the user needs to first click on a cell, which contains a prediction.

Then the wizard can be started by clicking on IUCLID Export (1) Wizard in the Endpoint tab (2). (Figure 1)

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes buttons for Input, Profiling, Oxyta (labeled 1), Category definition, Data Gap Filling, and Report. The 'Export' button in the 'Data' section is highlighted with a red box and labeled 2. The main window is divided into several panels:

- Documents:** Shows a document titled 'ent 8' with 'rch Chemical' and 'Aquatic toxicity classification by ECOSAR'. It indicates 'Enter GF(TA) with 29 chemicals, 30 data points' and 'Ch: 15] Data: 15 Subcategorized: Aquatic tox'.
- Databases:** Shows a list of databases with 'Group by: Hazard effects', 'Sort by: Name', and 'Color by: Endpoint selected in the data matrix'. The 'Ecotoxicological Information' section is expanded, showing 'Aquatic ECETOC', 'Aquatic Japan MoE', and 'Aquatic OASIS'.
- Inventories:** Shows a list of inventories with 'Sort by: Name'. The 'Ver_import_Custom Inventory_for tb4' is selected.
- Filter endpoint tree...:** A tree view showing 'Structure' and 'Ecotoxicological Information'. The tree is expanded to show 'Aquatic Toxicity' (AW SW) with sub-endpoints 'Growth (21/22)', 'Mortality (17/19)', and 'LC50'. Under 'LC50', there are '48 h' and '96 h' sub-endpoints. The '96 h' sub-endpoint is further expanded to show 'Animalia (animals)', 'Chordata (chordates)', and 'Actinopterygii (ray-finned) (28/29)'. Under 'Actinopterygii', there are 'Pimephales promelas' and 'Poeilia reticulata (2/2) (5/13)'. The 'Pimephales promelas' sub-endpoint is highlighted in blue.
- Data Table:** A table with three columns labeled '1 [target]', '2', and '3'. The table contains predicted values for various endpoints. The 'Pimephales promelas' sub-endpoint is highlighted in blue, showing a predicted value of 'M: 666 mg/L' and 'T: 386 (37+4.04E+03) mg/L'.

Figure 1

Prepare data

On the first page of the wizard, the user needs to select the prediction(s) that has to be exported (1). This page also allows editing the report information (2). (Figure 2)

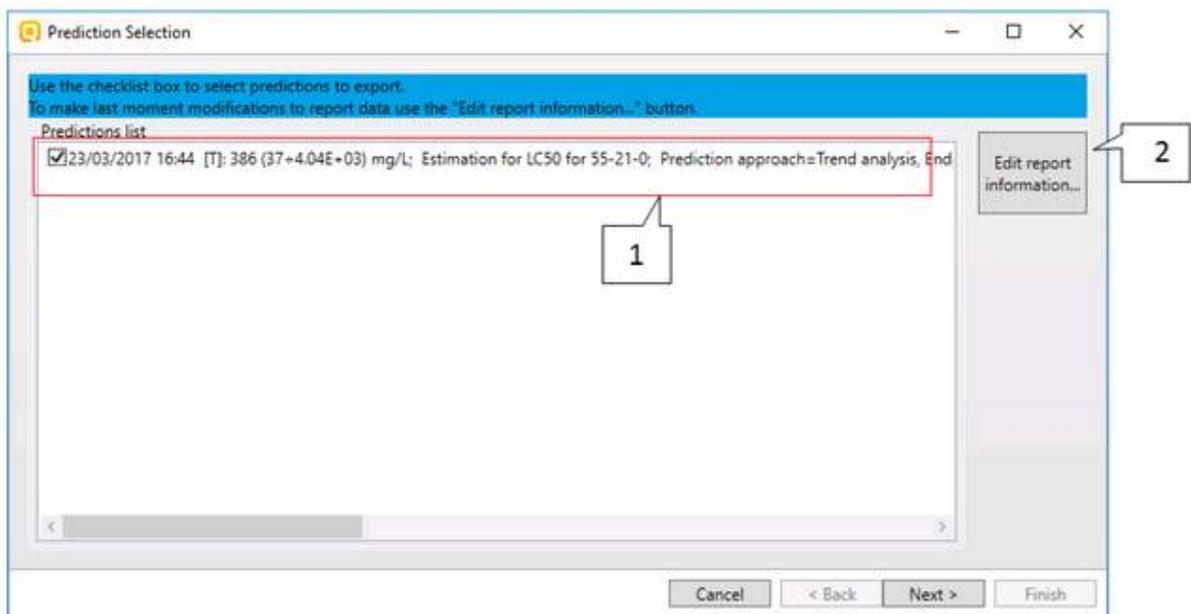


Figure 2

If the user wants to edit or add additional information of fields of the report he/she could specify the information, when click on Edit report information (2). (Figure 2)

Then a new window for editing field of the report appears (1). (Figure 3)

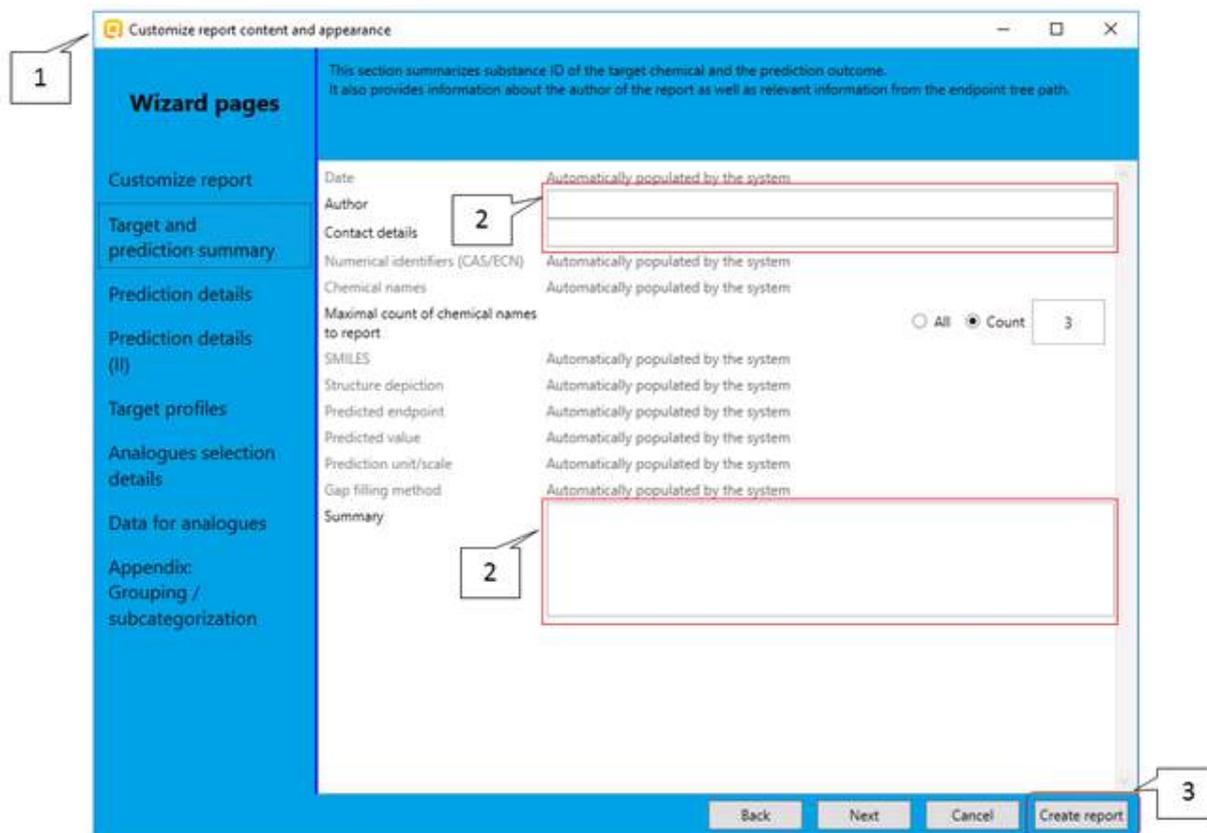


Figure 3

Edit prediction window includes sections such as Substance, General info, Category, Prediction etc., (2) where the user is allowed to edit information for some of the fields (3) or there is an information which is static and could not be changed (4). After adding the information to different fields, then the user has to click OK button (5), and modifications will be saved in the generated report. (Figure 68)

After the user specifies the additions/correction to the report fields, he/she has to click Next (1). (Figure 4)

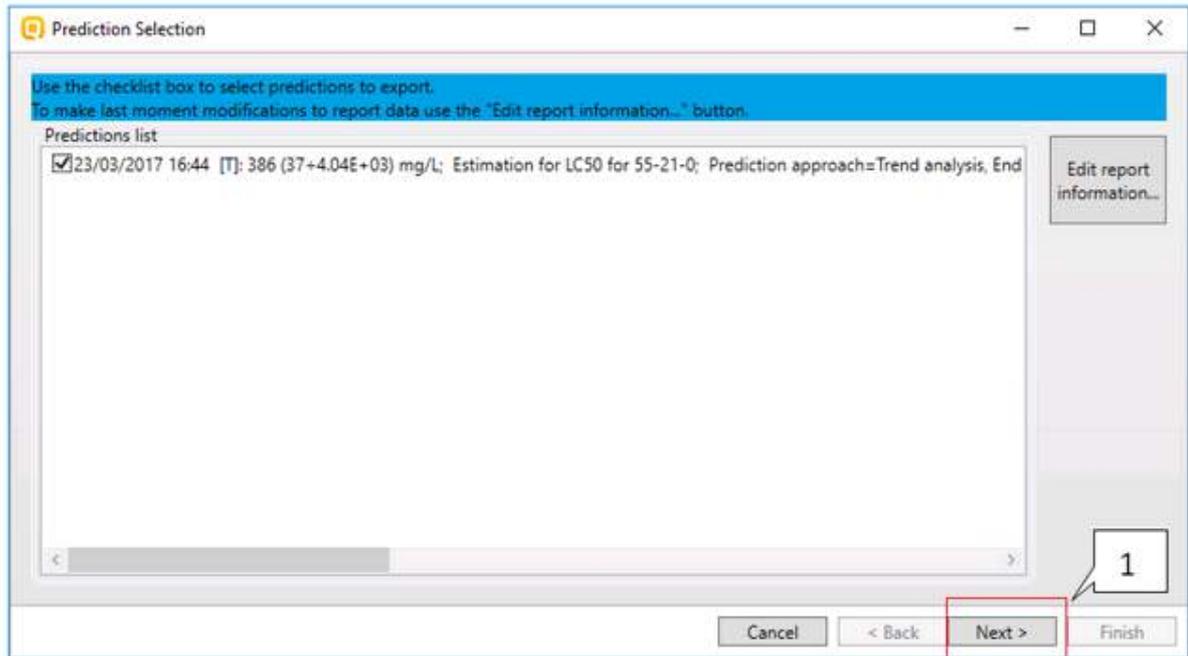


Figure 4

On the next page the user will have to associate each selected prediction (1) with the corresponding Harmonized Template from the list (2). As soon as all predictions have their corresponding template assigned, the Next button will be enabled. (Figure 5)

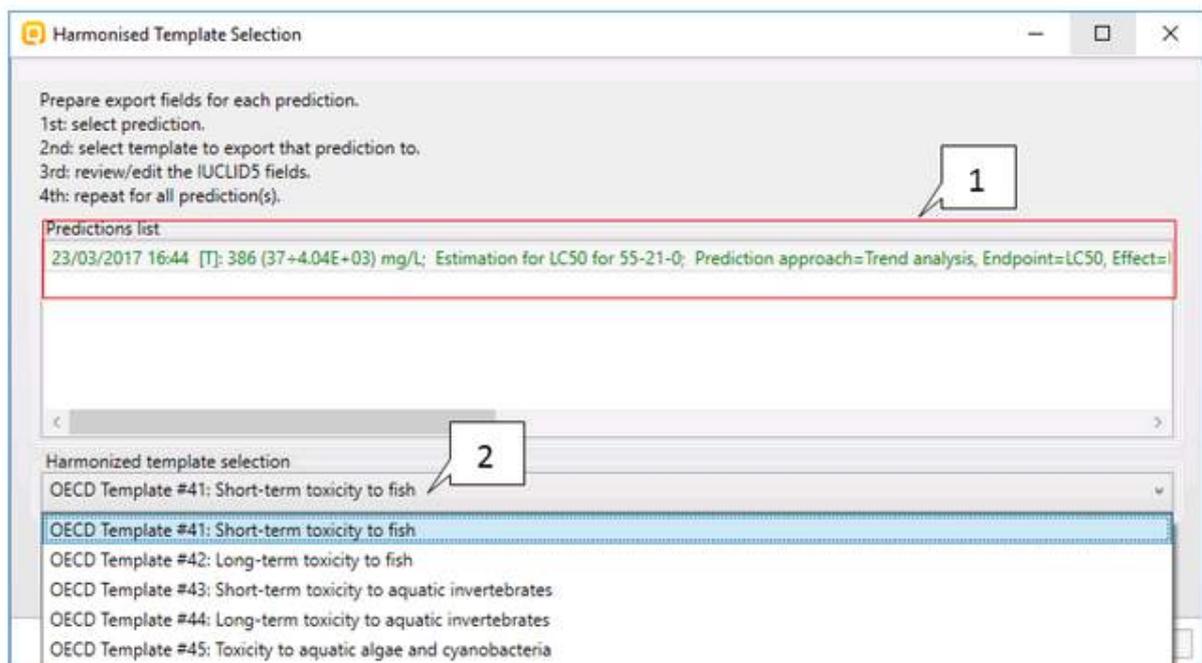


Figure 5

On this stage of the import wizard, the user could review fields for export (1). (Figure 6)

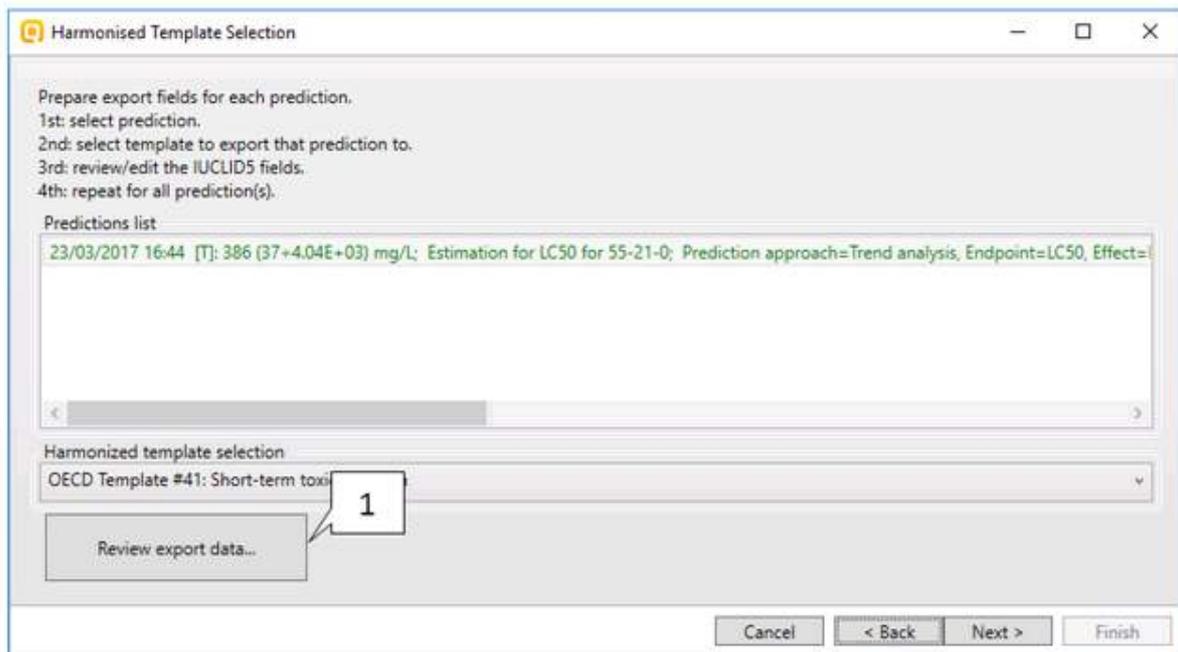


Figure 6

Clicking on Review export data button will evoke an IUCLID6 export (1) window, where the user could review/edit fields marked as mandatory for evaluators (2) and review/edit all other fields (3). Click OK (4) when finishing editing. (Figure 7)

The screenshot shows the 'IUCLID6 Export...' dialog box with the following fields and callouts:

- 1**: Callout pointing to the title bar of the dialog box.
- 2**: Callout pointing to the 'Qualifier of guideline' dropdown menu.
- 3**: Callout pointing to the 'Test guideline' dropdown menu, with the text 'mandatory for evaluators' in red above it.
- 4**: Callout pointing to the 'Water media type' dropdown menu.

Other fields include 'Value from prediction...' for each dropdown, 'Test organisms (species)' with 'Pimephales promelas' selected, and 'Test type'. At the bottom right are 'OK' and 'Cancel' buttons.

Figure 7

Click Next button in order to continue the import wizard (1). (Figure 8)

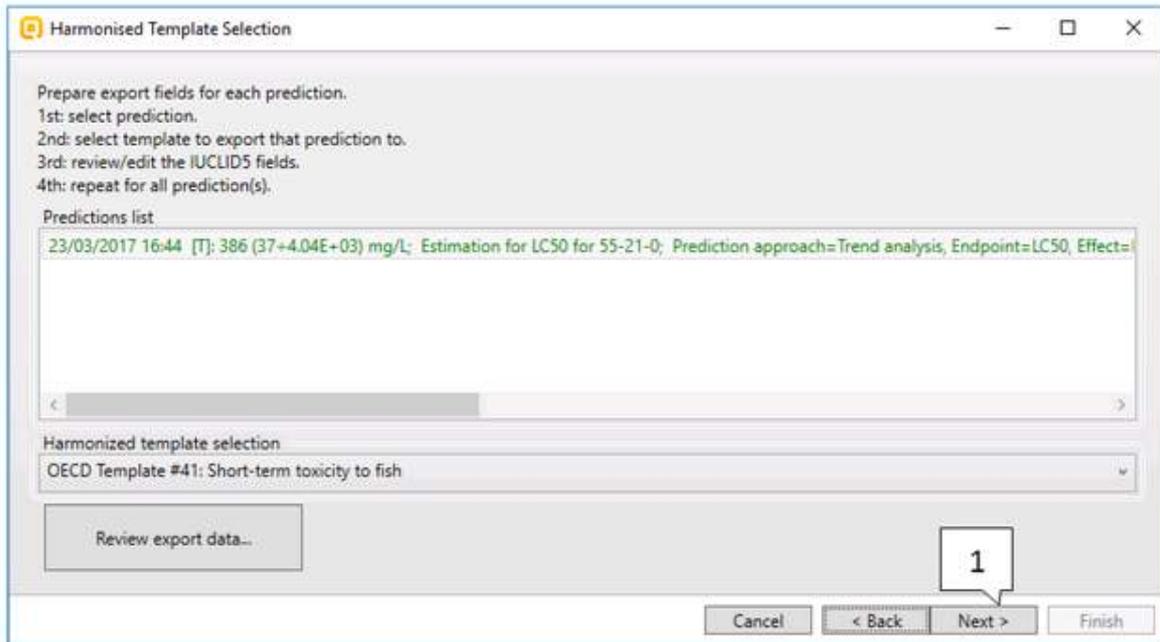


Figure 8

Then the Toolbox connects to the IUCLID 6 server.

On Figure 9 the user needs to provide the connection parameters. These include:

- o WebServices Server: the IP address or DNS name of the running IUCLID 6 server
- o Port: the TCP port that the server is listening on (usually 8080)
- o Username: the IUCLID 5 username valid for the server
- o Password: the corresponding password

After entering all of the above parameters, the user should click *Next* to attempt to connect to the server. Upon successful connection, the server, port and username are saved and can be retrieved later by selecting the server from the list. (Figure 8)

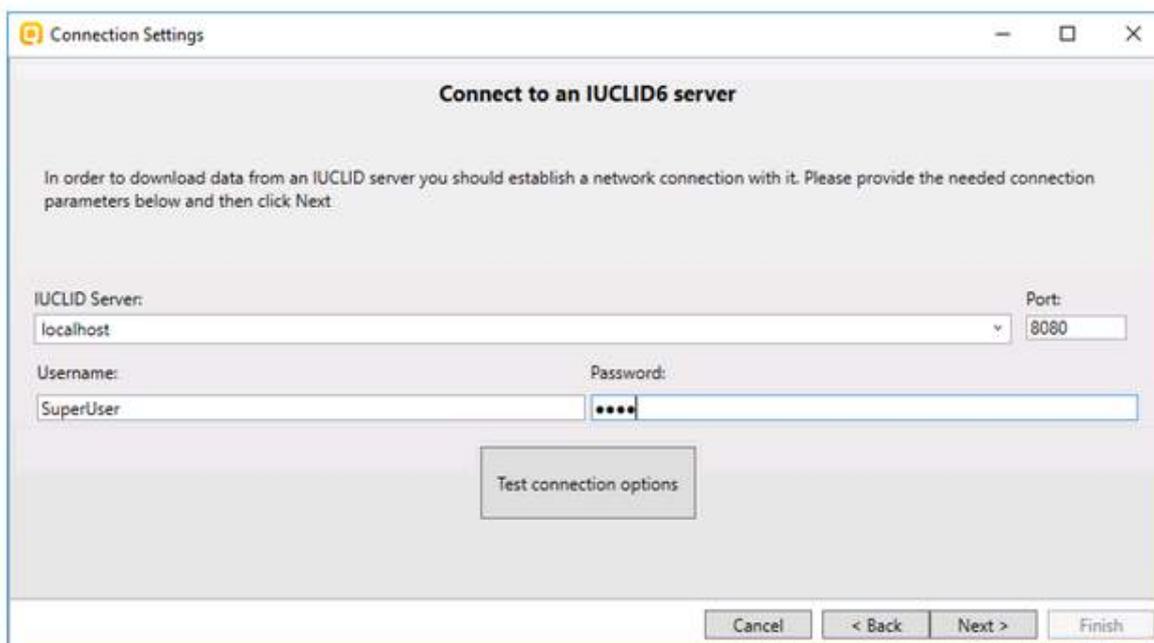


Figure 8

Given that the user has provided the correct parameters and a connection has been established the following screen appears. At this stage the user can click on the Get All Substances button to retrieve a list with all substances found in the corresponding IUCLID 6 database. (Figure 9)

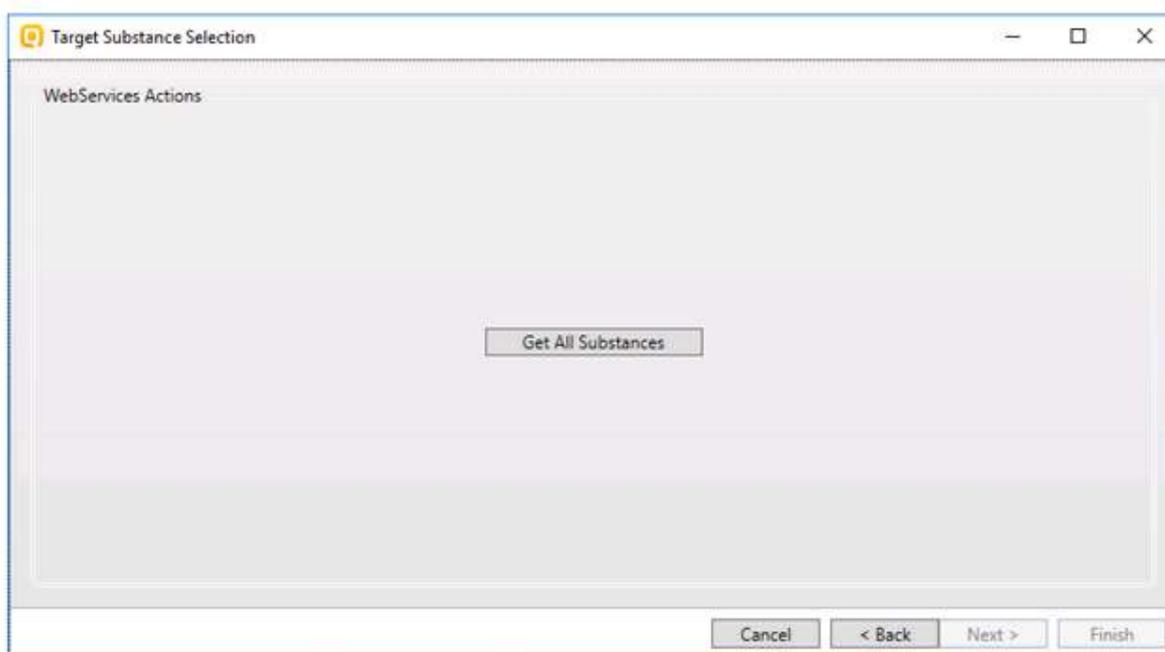


Figure 9

At this stage, the user is presented with the list of all substances found within

the IUCLID 6 database. The endpoint prediction can be assigned to any of these substances. Selecting one of them (1) and clicking Next (2) starts the export process which may take up to a minute. (Figure 10)

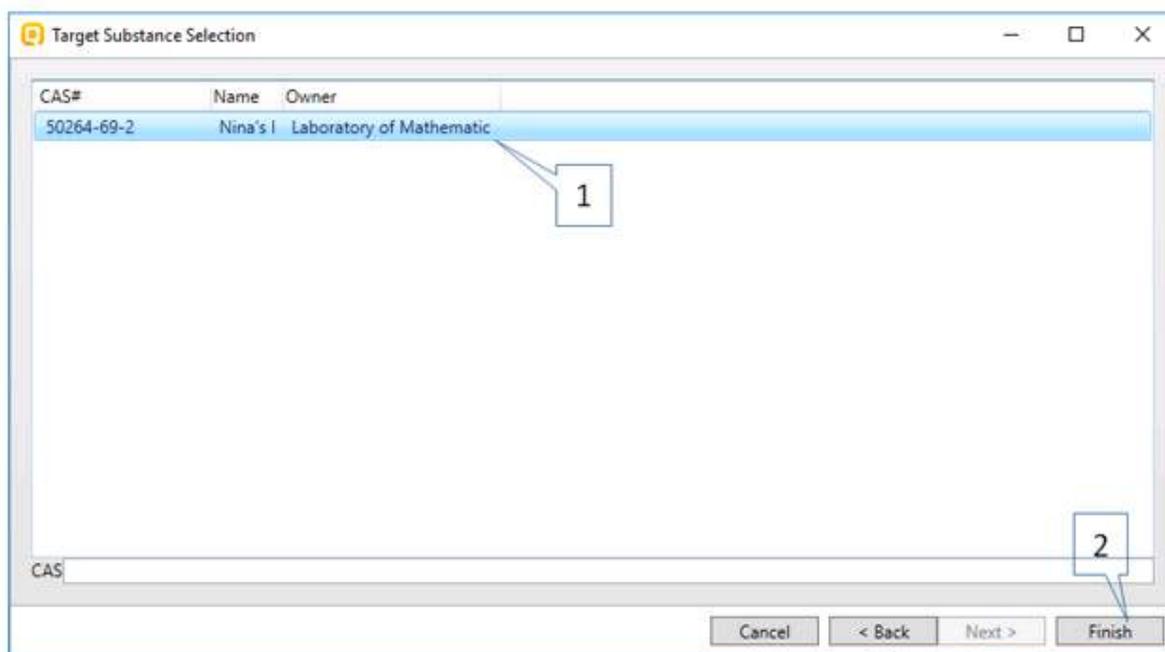


Figure 10

As soon as the export process is finished the user is informed whether data was successfully transferred to IUCLID 6 (Figure 11). In case there is a problem, a detailed log will be displayed which can help administrators/developers identify the issue.

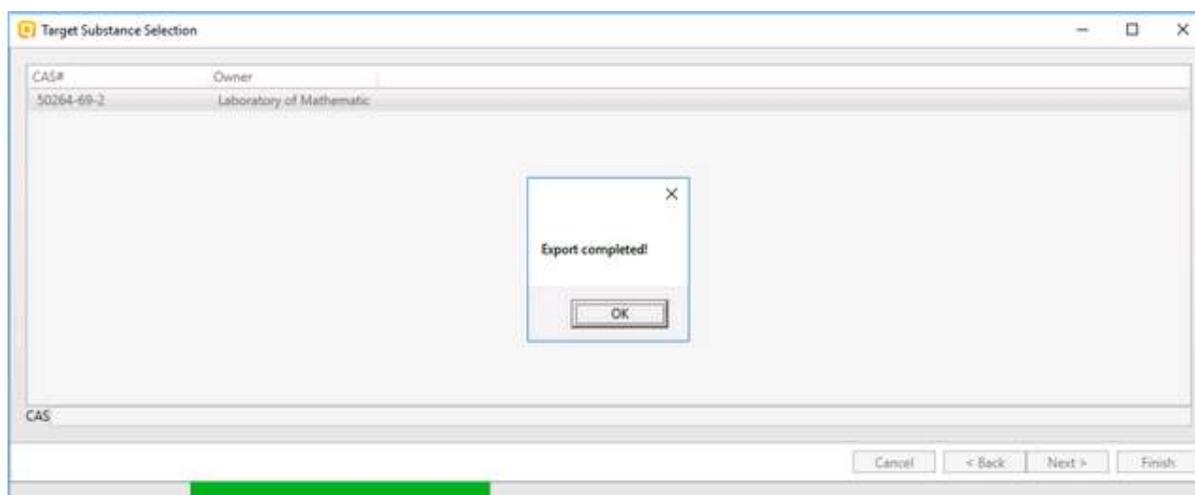


Figure 6

-0-

D.4. Category definition

QSAR Toolbox v.4

D.4.1. Grouping methods

[Workflow](#) / [Category definition](#) / D.4.1. Grouping methods

D.4.1. Grouping methods

List with grouping methods covers the list with profiling methods (Table 1):

Table 1. List with profiling and grouping methods

Profiling method	Grouping method
Predefined	
Database Affiliation	Database Affiliation
Inventory Affiliation	Inventory Affiliation
OECD HPV Chemical Categories	OECD HPV Chemical Categories
Substance type	Substance type
US-EPA New Chemical categories	US-EPA New Chemical categories
General Mechanistic	
Biodeg probability (Biowin1)	Biodeg probability (Biowin1)
Biodeg probability (Biowin2)	Biodeg probability (Biowin2)
Ultimate biode (Biowin3)	Ultimate biode (Biowin3)
Primary biodeg (Biowin4)	Primary biodeg (Biowin4)
Biodeg probability (Biowin5)	Biodeg probability (Biowin5)
Biodeg probability (Biowin6)	Biodeg probability (Biowin6)
Biodeg probability (Biowin7)	Biodeg probability (Biowin7)
BioHC half-life (Biowin)	BioHC half-life (Biowin)
DNA binding by OECD	DNA binding by OECD
DNA binding by OASIS v1.4	DNA binding by OASIS v1.4
Hydrolysis half-life (Ka, pH7)(Hydrowin)	Hydrolysis half-life (Ka, pH7)(Hydrowin)
Hydrolysis half-life (Ka, pH8)(Hydrowin)	Hydrolysis half-life (Ka, pH8)(Hydrowin)
Hydrolysis half-life (Kb, pH7)(Hydrowin)	Hydrolysis half-life (Kb, pH7)(Hydrowin)
Hydrolysis half-life (Kb, pH8)(Hydrowin)	Hydrolysis half-life (Kb, pH8)(Hydrowin)
Hydrolysis half-life (pH 6.5-7.4)	Hydrolysis half-life (pH 6.5-7.4)

Ionization at pH = 1	Ionization at pH = 1
Ionization at pH = 4	Ionization at pH = 4
Ionization at pH = 7.4	Ionization at pH = 7.4
Ionization at pH = 9	Ionization at pH = 9
Protein binding by OASIS	Protein binding by OASIS
Protein binding by OECD	Protein binding by OECD
Protein binding potency	Protein binding potency
Toxic hazard classification by Cramer (original)	Toxic hazard classification by Cramer (original)
Toxic hazard classification by Cramer (with extensions)	Toxic hazard classification by Cramer (with extensions)
Ultimate biodeg	Ultimate biodeg
Endpoint specific	
Acute aquatic toxicity classification by Verhaar	Acute aquatic toxicity classification by Verhaar
Acute aquatic toxicity MOA by OASIS	Acute aquatic toxicity MOA by OASIS
Aquatic toxicity classification by ECOSAR	Aquatic toxicity classification by ECOSAR
Bioaccumulation – metabolism alerts	Bioaccumulation – metabolism alerts
Bioaccumulation – metabolism half-lives	Bioaccumulation – metabolism half-lives
Biodegradation fragments (BioWIN MITI)	Biodegradation fragments (BioWIN MITI)
Carcinogenicity (genotox and nongenotox) alerts by ISS	Carcinogenicity (genotox and nongenotox) alerts by ISS
DNA alerts for Ames by OASIS v.1.4	DNA alerts for Ames by OASIS v.1.4
DNA alerts for CA and MNT by OASIS v.1.1	DNA alerts for CA and MNT by OASIS v.1.1
Eye irritation/corrosion Exclusion rules by BfR	Eye irritation/corrosion Exclusion rules by BfR
Eye irritation/corrosion Inclusion rules by BfR	Eye irritation/corrosion Inclusion rules by BfR
in vitro mutagenicity (Ames test) alerts by ISS	in vitro mutagenicity (Ames test) alerts by ISS
in vivo mutagenicity (Micronucleus) alerts by ISS	in vivo mutagenicity (Micronucleus) alerts by ISS
Keratinocyte gene expression	Keratinocyte gene expression
Oncologic Primary Classification	Oncologic Primary Classification
Protein binding alerts for Chromosomal aberration by OASIS v.1.2	Protein binding alerts for Chromosomal aberration by OASIS v.1.2

Protein binding alerts for skin sensitization by OAS v.1.4	Protein binding alerts for skin sensitization by OAS v.1.4
Respiratory sensitisation	Respiratory sensitisation
Retinoic Acid Receptor Binding	Retinoic Acid Receptor Binding
rER Expert System ver.1 - USEPA	rER Expert System ver.1 - USEPA
Skin irritation/corrosion Exclusion rules by BfR	Skin irritation/corrosion Exclusion rules by BfR
Skin irritation/corrosion Inclusion rules by BfR	Skin irritation/corrosion Inclusion rules by BfR
Empiric	
Chemical elements	Chemical elements
Groups of elements	Groups of elements
Lipinski Rule Oasis	Lipinski Rule Oasis
Organic functional groups	Organic functional groups
Organic functional groups (nested)	Organic functional groups (nested)
Organic functional groups (US EPA)	Organic functional groups (US EPA)
Organic functional groups, Norbert Haider (checkmol)	Organic functional groups, Norbert Haider (checkmol)
Structure similarity	Structure similarity
Tautomers unstable	Tautomers unstable
Toxicological	
Repeated dose (HESS)	Repeated dose (HESS)
Custom	
Example Prioritization Scheme (PBT)	Example Prioritization Scheme (PBT)

Summary background information for some grouping methods is listed in [Table 2](#).

[Table 2](#). Summary information for some grouping methods:

Grouping method	Summary background information	Grouping method	Summary background information
General Mechanistic		Biodegradation Fragments (BIOWIN MITI)	BIOWIN estimates the probability of rapid aerobic and anaerobic biodegradation of an organic compound in the presence of

			mixed populations of environmental microorganisms.
DNA binding by OECD	This grouping method contains categories or chemical methods of DNA binding. This method is particularly relevant for genotoxicity endpoints.	Eye irritation/corrosion exclusion rules by BfR	This grouping method is based on physical-chemical exclusion criteria relevant for eye irritation and corrosion.
DNA binding by OASIS v1.1	This grouping method contains categories or chemical methods of DNA binding. This method is particularly relevant for genotoxicity endpoints.	Eye irritation/corrosion inclusion rules by BfR	This grouping method is based on structural alerts for eye irritation or corrosion.
Estrogen receptor binding	This grouping method contains simple categories for estrogen receptor (ER) binding. This method is relevant for reproductive toxicity endpoints in fish and mammals.	in vivo mutagenicity (Micronucleus) alerts by ISS	This profiler is based on structural alerts known to be linked to the induction of effects in the in vivo micronucleus assay.
Protein binding by OASIS	This grouping method includes categories or chemical mechanisms of protein binding This method is particularly relevant for skin and respiratory sensitization and acute aquatic toxicity, but also for chromosomal aberration and acute inhalation toxicity.	in vitro mutagenicity (Ames test) alerts by ISS	This profiler is based on the Mutagenicity/Carcinogenicity module of the software Toxtree. It works as a decision tree for estimating in vitro (Ames test) mutagenicity, based on a list of 30 structural alerts (SAs).
		Carcinogenicity (genotox and nongenotox) alerts by ISS	This profiler is an expanded and updated version of the correspondent module of the software Toxtree. It works as a decision tree for estimating carcinogenicity, based on a list of 55 structural alerts (SAs). The structural alerts refer mainly to the knowledge on the action mechanisms of genotoxic carcinogenicity (thus they apply also to the mutagenic activity in bacteria), but includes also a number of structural alerts flagging potential non-genotoxic carcinogens
Toxic Hazards Classification by Cramer	This is a historical grouping method where substances are classified into one of three classes for human health assessment.	Oncologic primary classification	The OncoLogic Primary Classifier consists of molecular definitions developed by LMC and OECD to mimic the structural criteria of chemical classes of potential carcinogens covered by the U.S. Environmental Protection Agency's OncoLogic™ Cancer Expert System for Predicting the Carcinogenicity Potential.
Endpoint specific		Skin irritation/corrosion Exclusion rules by BfR	The exclusion rules for skin irritation/corrosion are based on

			physicochemical cut-off values to identify chemicals that do not exhibit skin irritation or corrosion potential.
Acute aquatic toxicity MOA by OASIS	This grouping method classifies chemicals by their acute aquatic toxicity (behavioral) mode of action based on chemicals tested with fathead minnow prior to 1997. This method is specifically relevant for acute aquatic toxicity endpoints. Since this method only delineates modes of action from fathead minnow data, it is best used in combination with other grouping methods linked to acute aquatic toxicity.	Skin irritation/corrosion Inclusion rules by BfR	The profiler contains structural alerts which can be used for positive classification of chemicals causing irritation, corrosion or the combination irritation/corrosion depending on their mechanisms. These 40 organic structural fragments cover the inclusion rules for skin irritation and corrosion as for now.
Aquatic toxicity classification by ECOSAR	ECOSAR is a grouping method for identifying chemical classes. It is especially relevant for aquatic toxicity endpoints. Since this method is the most robust of the mechanistic grouping methods it is often the method of choice.	Empiric	
Acute Aquatic toxicity Classification by Verhaar	This is a historical grouping method according to reactivity based on experience with acute fish toxicity up to 1992. This method is specifically relevant for acute aquatic toxicity endpoints. Since this method only delineates broad classes of reactivity (inert, less inert, reactive, and specifically-acting chemicals), it is best used in combination with other grouping methods linked to acute aquatic toxicity.	Organic functional groups	The Organic Functional Groups (OFG) system is designed in order to introduce classification and systematization of the various functionalities and characteristic structural fragments in organic chemicals from a large database, and identify structurally similar chemicals. Organic functional groups are specific groups of atoms and/or bonds within molecules that determine the characteristic chemical reactions of those molecules.
Bioaccumulation metabolism alerts	These two grouping methods are based on the structural alerts used by the BCFBAF model for bioaccumulation developed for the US EPA. It is particularly relevant for identifying outliers when grouping chemical for read-across for bioaccumulation	Organic functional groups (nested)	The OFG profile is currently subdivided into OFG (general) and OFG (nested), which include the same functional groups. The difference is that OFG (general) displays all functional groups present in the target compounds, while OFG (nested) do not show the functional groups, which are only parts of larger ones. Thus any "overlapping" groups in the OFG (nested) system actually form new functional groups.
Bioaccumulation metabolism Half-lives	–		

When searching for analogues of a target chemical, the outcome of the profiling determines the most appropriate way. The following recommendations can be made:

Recommendations 

If specific mechanisms or modes of action are identified for a target chemical, which are relevant for the investigated endpoint, then it is recommended to search for chemicals which have the same mechanisms or modes of action. The search results can then be refined by eliminating those chemicals which are structurally most dissimilar.

If no specific mechanisms or modes of action are identified for a target chemical, which are relevant for the investigated endpoint, then it is recommended to search for chemicals which are structurally similar to the target chemical. The search results can then be refined by eliminating those chemicals which have specific mechanisms or modes of action.

It should be kept in mind that the search for analogues is performed among the chemicals, which are listed in the selected Databases or inventories listed under Inventories. For example if only the databases "Skin sensitisation ECETOC" and "Skin sensitisation" are selected, the analogue search will only be performed among those chemicals for which experimental data on skin sensitisation are available in those databases. Similarly, the user can decide to expand the search to chemical inventories. For example by selecting the databases "Carcinogenicity & Mutagenicity ISSCAN" and "Genotoxicity OASIS" as well as the inventory "Canada DSL", the Toolbox will query for analogues in those two databases as well as this specific inventory.

Caution 

The inventories contain between 5 000 and 100 000 substances. In order to accelerate the process of identifying the similar analogues only the databases are preliminary profiled and indexed (2D and 3D calculations). In this respect if one is to include one of them in the query this will produce longer calculation times.

D.4.2. Building category - principles

[Workflow](#) / [Category definition](#) / D.4.2. Building category - principles

D.4.2. Building category - principles

When selecting databases and/or inventories to apply category definitions, the following recommendations apply:

- If the aim of the user is to find only analogues for which experimental data are available on specific endpoints, then only those databases that contain results on those endpoints should be selected. No inventory should be selected.
- First step of categorization procedure is to find more structurally similar analogues using non-endpoint specific grouping methods. In this respect structurally based grouping methods which will define more broader group is recommended to be used such as:
 - US EPA Categorization
 - OECD Categorization
 - Organic functional group
 - Structural similarity
 - ECOSAR
- Second step is to refine the broader group using subcategorization procedure. In this step mechanistically based and endpoint specific grouping methods can be used:
 - DNA binding mechanism
 - Protein binding mechanism
 - Genotoxicity/carcinogenicity
 - Cramer rules
 - Verhaar rule
 - Skin/eye irritation corrosion rules
- Final step of subcategorization – apply first step of categorization

QSAR Toolbox v.4

D.4.3. Defining category

[Workflow](#) / [Category definition](#) / D.4.3. Defining category

D.4.3. Defining categories

Procedure for defining categories

The creation of a category is straightforward – the user selects a grouper (1) and presses the Define button (2) (Figure 1)

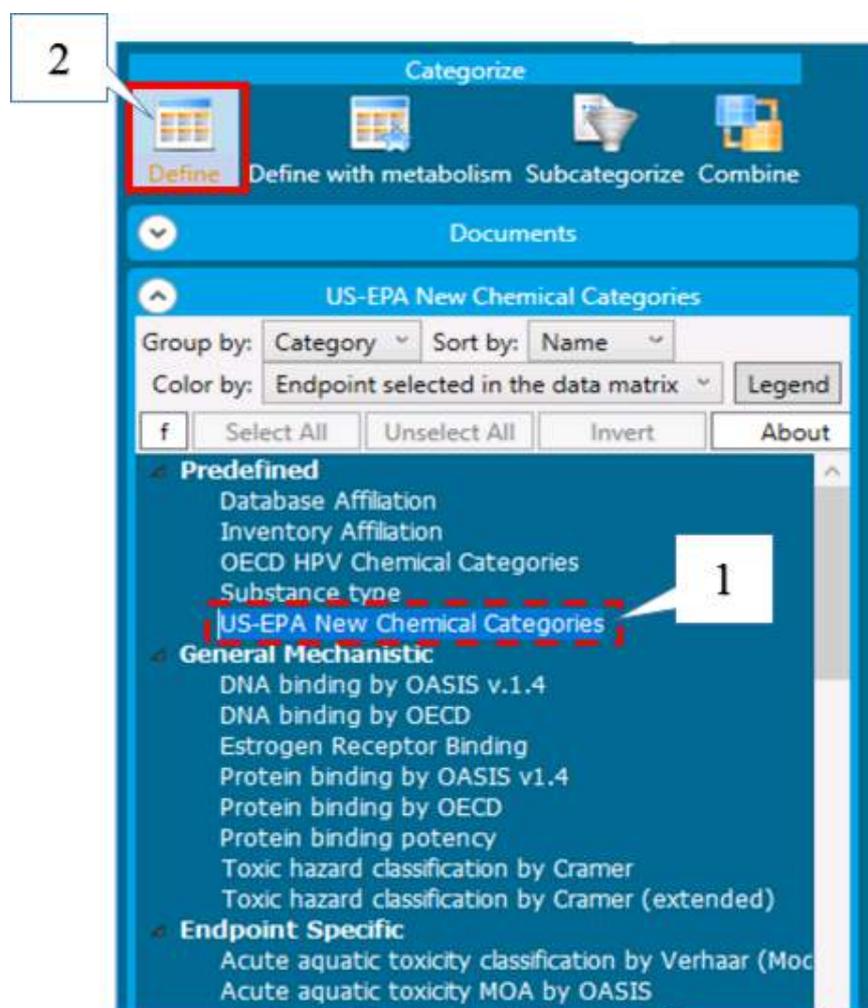


Figure 1

When the grouping method executes it will provide the user with the all the categories in the selected grouping method and the categories of the target chemical (1) (if any) will populate the Target(s) profiles (2) list box. (Figure

2)

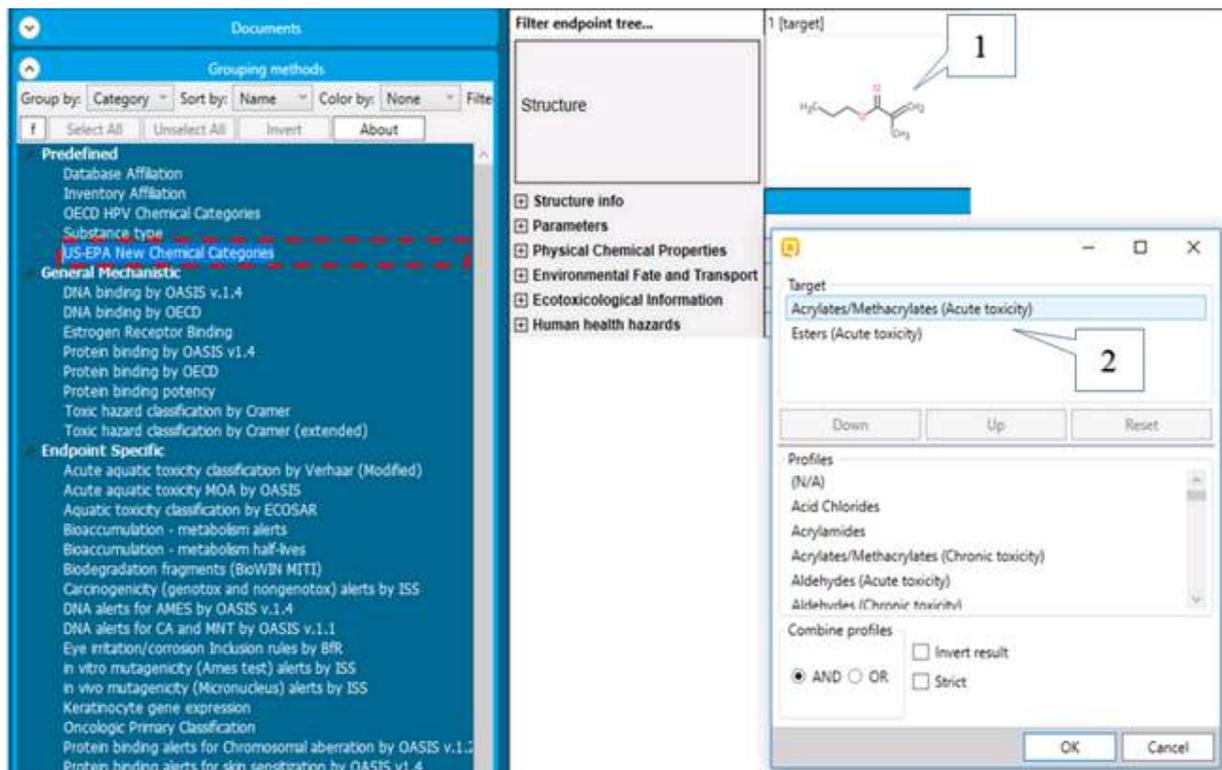


Figure 2

On this stage the user has opportunity to:

- remove targets categories by selecting one of them (1) and moving it in the All profiles (2) panel using the down arrow (3) (Figure 3)

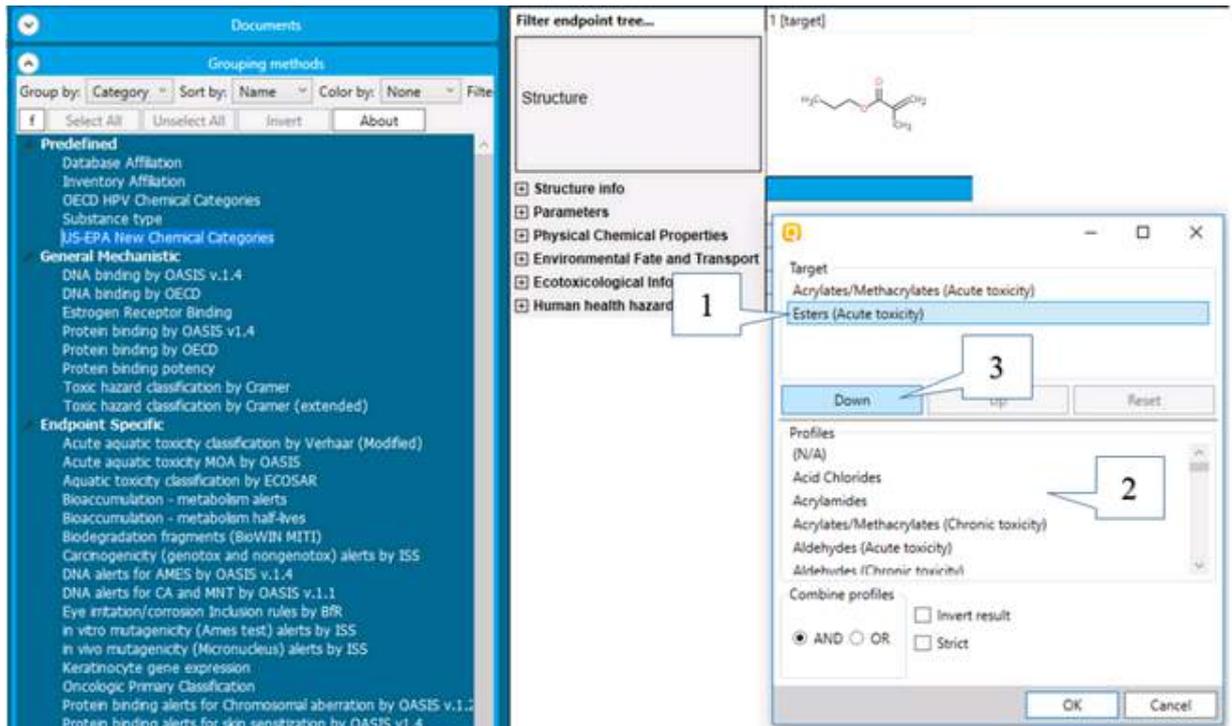


Figure 3

- add categories to the selection of targets profiles by selecting one of them (1) from the All profiles panel (2) and moving it to the Target(s) profiles (4) panel pressing the up arrow (4). (Figure 4)

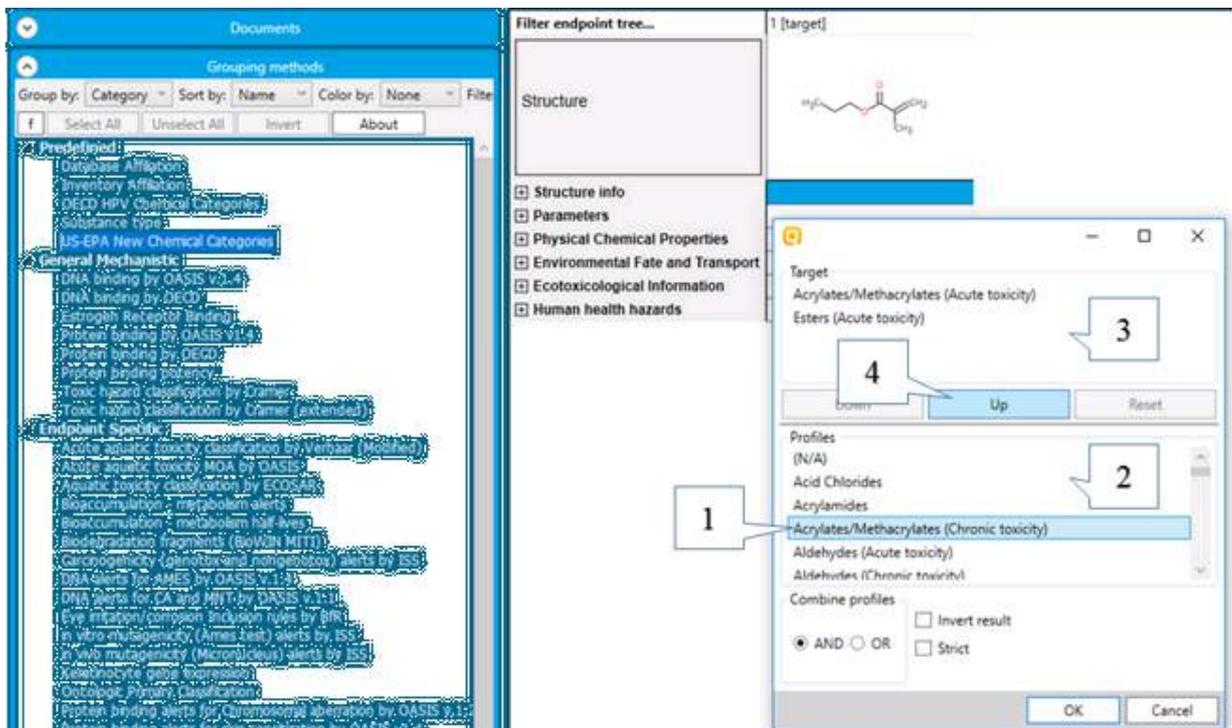


Figure 4

- Combine profiles – this allows the user to choose how to combine targets profiles using logically And/Or operand. If And (1) is chosen then all selected categories have to be presented in the searched chemical(s). If Or (2) is chosen then only one of the selected category is enough. (Figure 5)

By default the AND is selected. (Figure 5)

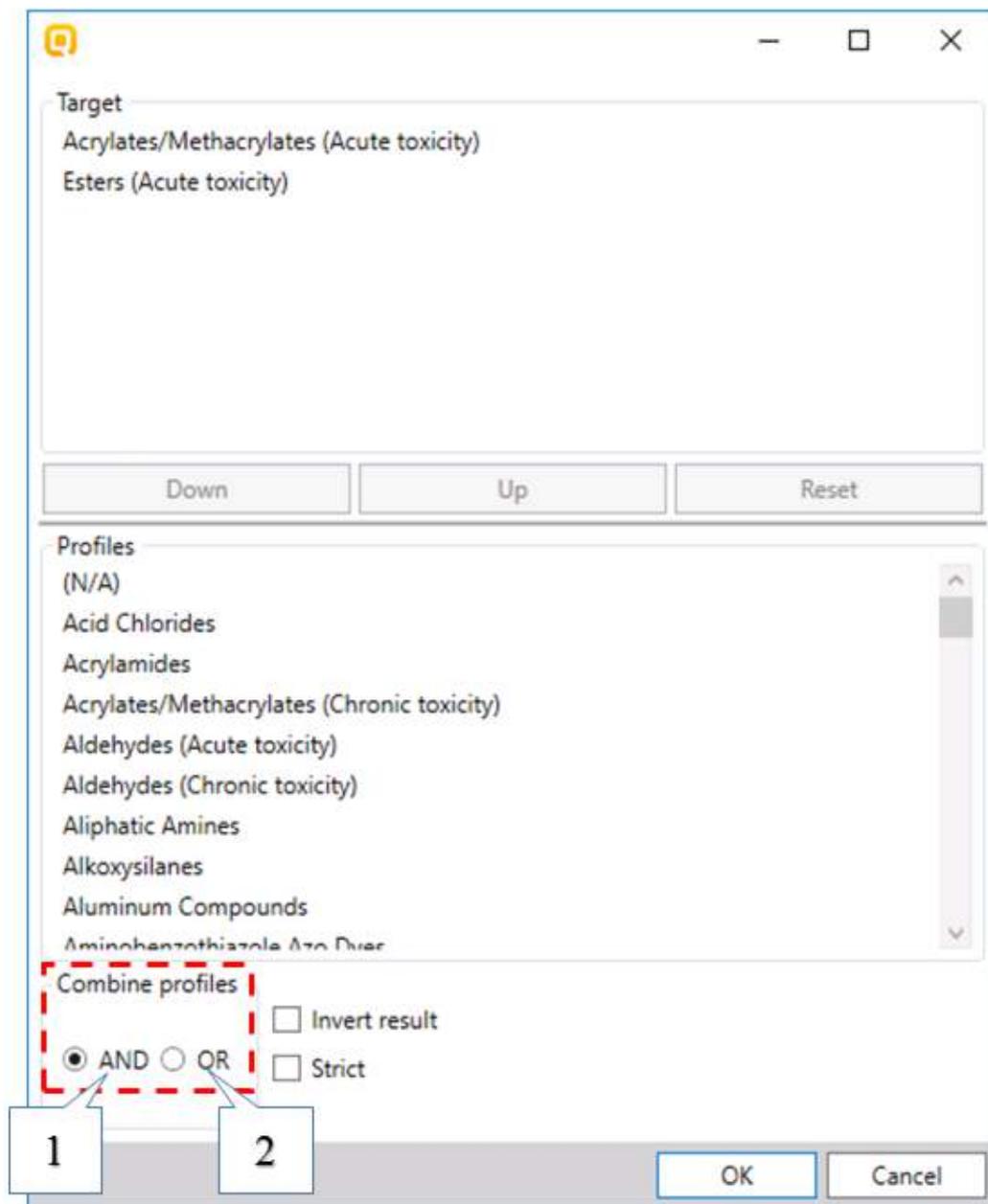


Figure 5

- Invert results (1) – this function searches for chemicals which have profiles different than the targets profiles. On the screenshot below (Figure 6), if the is classified as *Acrylates/Methacrylates (Acute toxicity)* and *Esters (Acute toxicity)* and Inverts results is selected, then the software will search for chemicals having profiles different than the above mentioned.
- Strict option (2) (Figure 6) - If Strict is checked then only the defined categories should be present and not any other.

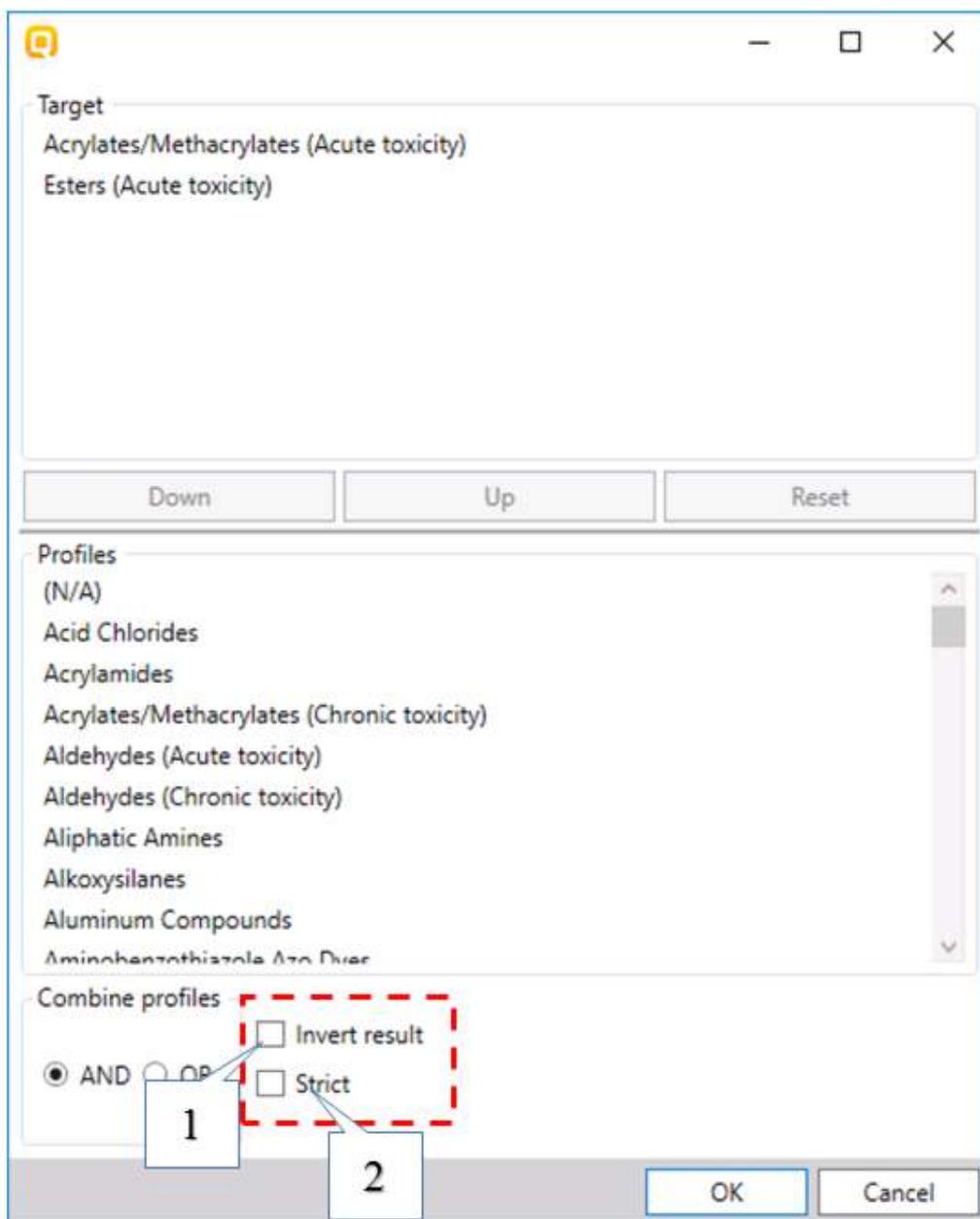


Figure 6

After closing the window for defining the category a category is built and a Grouping results dialog appears (Figure 7).

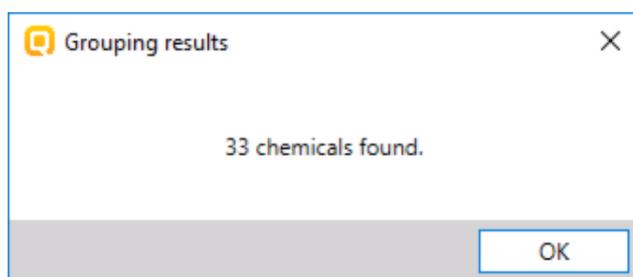


Figure 7

The software identified 33 chemicals from the selected database(s) with same profiles as those of the target chemical.

After closing the Grouping results window, the software automatically commences a gather data action. The user can select the specific endpoint (Choose...) or by default choose to retrieve data for all endpoints (All endpoints) (see below) (Figure 8). If the user has previously selected databases related to the investigated endpoints, then both options will return same results.

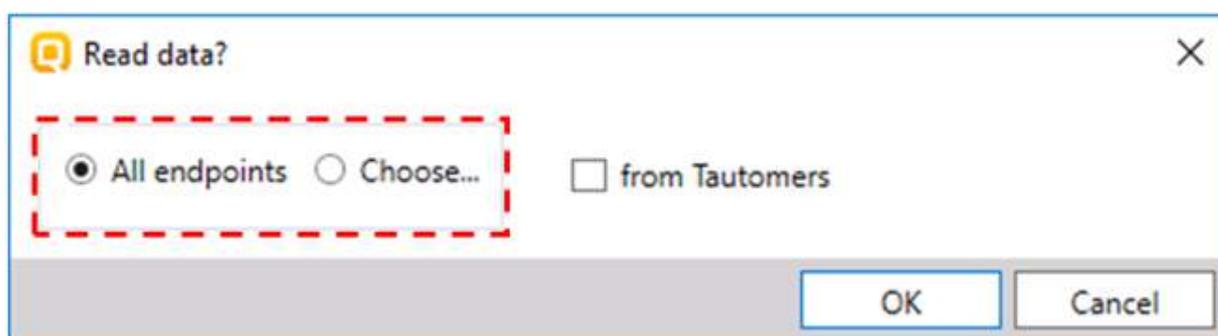


Figure 8

Recommendations

If the user has selected all databases under the Endpoint section, and selects **All endpoint** the gather data operation could be very time consuming due to the diversity of endpoints and size of databases. In this respect the user is recommended to always select only those databases, and endpoint paths, which are related to the investigated endpoints.

Finally after reading data, the defined category appears in the document tree (1). (Figure 9)

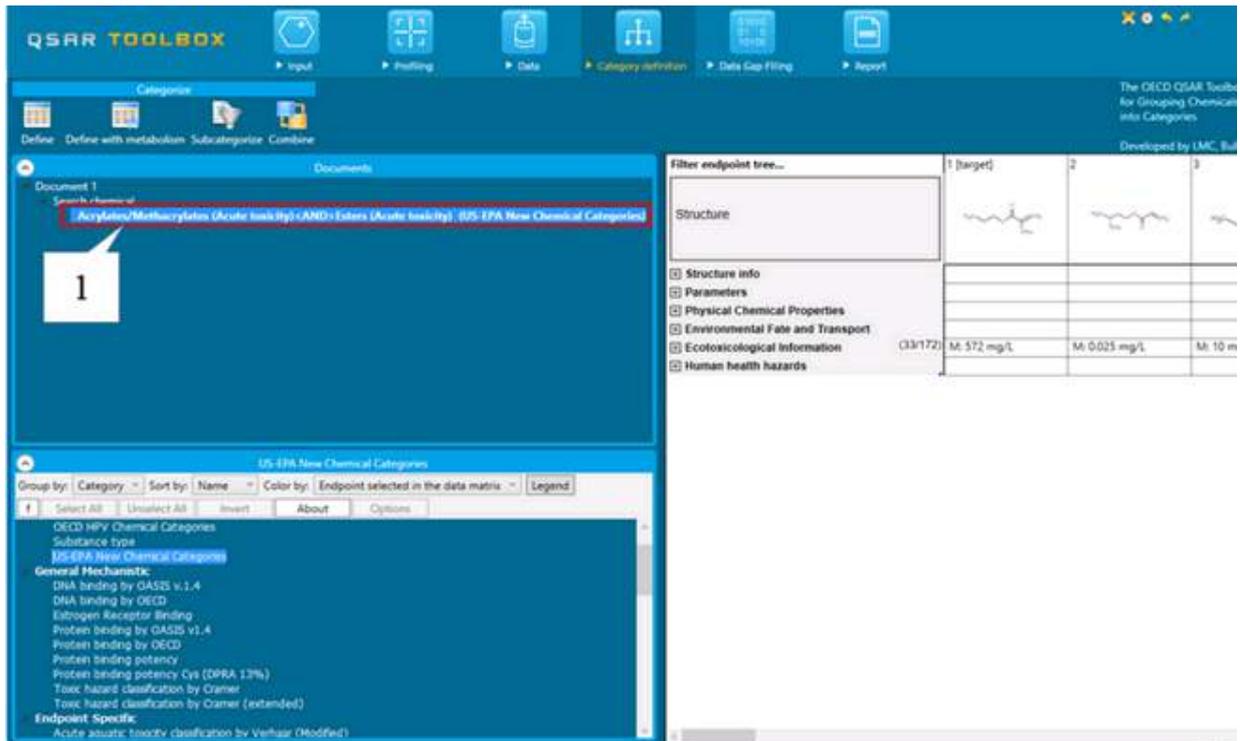


Figure 9

-0-

QSAR Toolbox v.4

D.4.3.1. Categorization of single chemical

[Workflow](#) / [Category definition](#) / [Defining category](#) / D.4.3.1. Categorization of single chemical

D.4.3.1. Defining category of single chemical

Categorizing of single chemical is explained in the section [Procedure for defining categories](#). *(see section Defining category/Procedure for defining categories)*

-0-

QSAR Toolbox v.4

D.4.3.2. Categorization of mixtures

[Workflow](#) / [Category definition](#) / [Defining category](#) / D.4.3.2. Categorization of mixtures

D.4.3.2. Defining category of mixtures

When the user categorizes a set of mixtures all profiles of component of the mixture are taken into account (1) (Figure 1)

The screenshot displays the 'Filter endpoint tree...' panel on the left and the '1 [target]' panel on the right. The 'Filter endpoint tree...' panel includes a 'Structure' section and a tree view with the following categories:

- Structure info
- Parameters
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human health hazards
- Profile
 - Predefined
 - US-EPA New Chemical Categories
 - Endpoint Specific
 - Acute aquatic toxicity MOA by OASIS
 - Aquatic toxicity classification by ECOSAR
 - Empiric
 - Organic Functional groups

The '1 [target]' panel shows chemical structures and a list of categories:

- Neutral Organics
- Basesurface narcotics
- Neutral Organics
- Alcohol
- Aryl
- Aryl halide
- Ketone

A red dashed box highlights the 'Alcohol', 'Aryl', 'Aryl halide', and 'Ketone' categories, with a callout box containing the number '1'.

Figure 1

And in the Categorization panel all profiles of component of the mixture(s) are taken into account for Target(s) profile (1) (Figure 2):

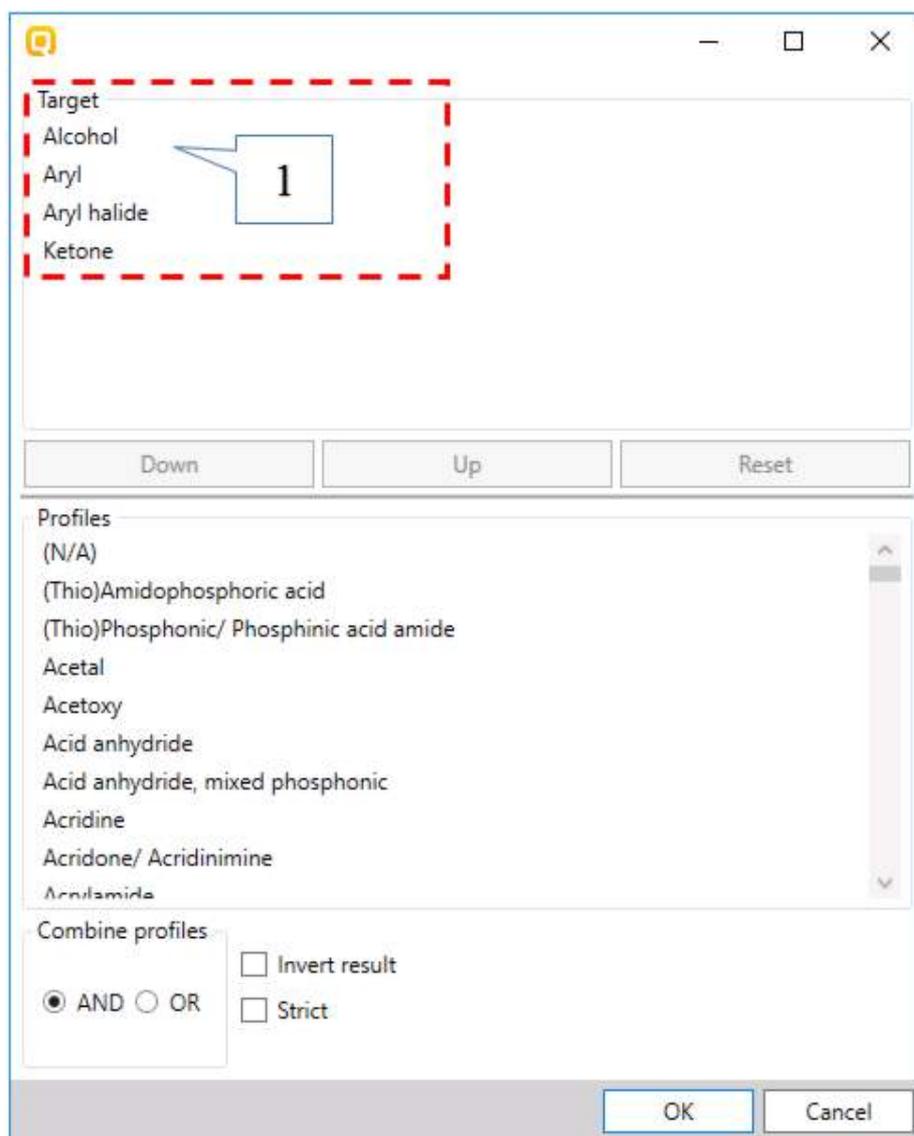


Figure 2

-0-

QSAR Toolbox v.4

D.4.3.3. Categorization using profiling result of hierarchical type profiling scheme

[Workflow](#) / [Category definition](#) / [Defining category](#) / D.4.3.3. Categorization using profiling result of hierarchical type profiling scheme

D.4.3.3. Categorization using profiling result of hierarchical type profiling scheme

Profiling results from hierarchical schemes such as Protein and DNA binding give information for Domain, Mechanistic alert and Structural alerts. Profiling results are visualized hierarchically: (Figure 1)

1 [target]

Structure

Structure info
 Parameters
 Physical Chemical Properties
 Environmental Fate and Transport
 Ecotoxicological Information
 Human health hazards
 Profile

General Mechanistic
 DNA binding by OASIS v.1.4
 Protein binding by OASIS v1.4

SN2
 SN2 >> Alkylation
 SN2 >> Alkylation >> Alkylphosphates, Alkylthiophosphates and Alkylphosphonates
 Acylation
 Acylation >> Ester aminolysis
 Acylation >> Ester aminolysis >> Amides
 SN2
 SN2 >> Nucleophilic substitution at sp³ carbon atom
 SN2 >> Nucleophilic substitution at sp³ carbon atom >> (Thio)Phosphates

Figure 1

- 1) Domain
- 2) Domain >> Mechanistic alert

3) Domain >> Mechanistic alert >> Structural alert

Toolbox gives opportunity to define a category using each of these profiling results.

In case the user applies category "Domain" (Figure 2) for categorization proposes, then the software will search for chemicals which answer the criteria of category "Domain" (e.g. SN2). SN2 includes following categories:

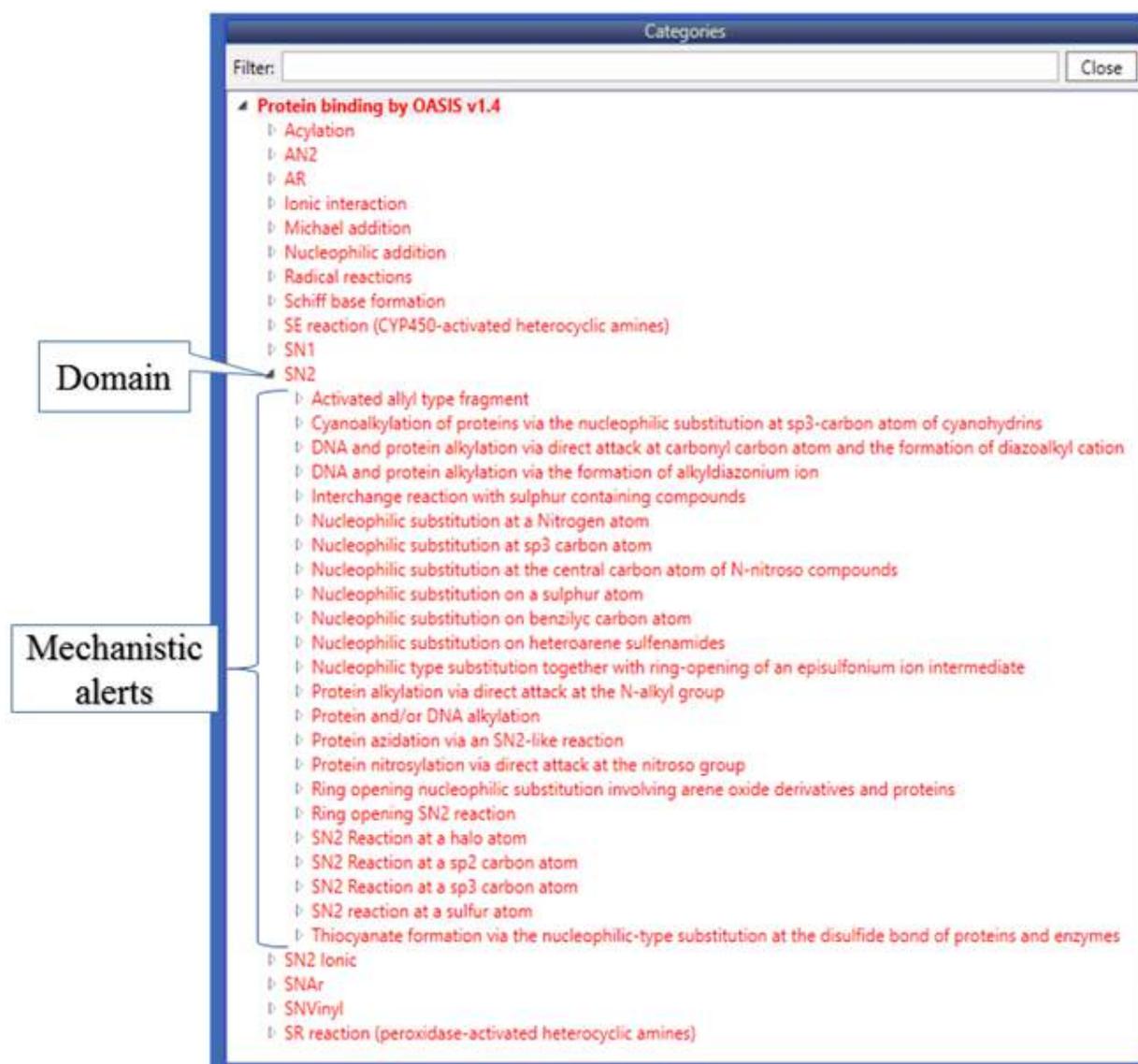


Figure 2

So the defined category will include chemicals classified in one of these (or all together) Mechanistic alerts shown above depending on the logically operant used in defining categories

The defined category “SN2” will be broader (will include more chemicals) than the category “Nucleophilic substitution at sp³ carbon atom” (Mechanistic alert), which includes seven structural alerts. (Figure 3):

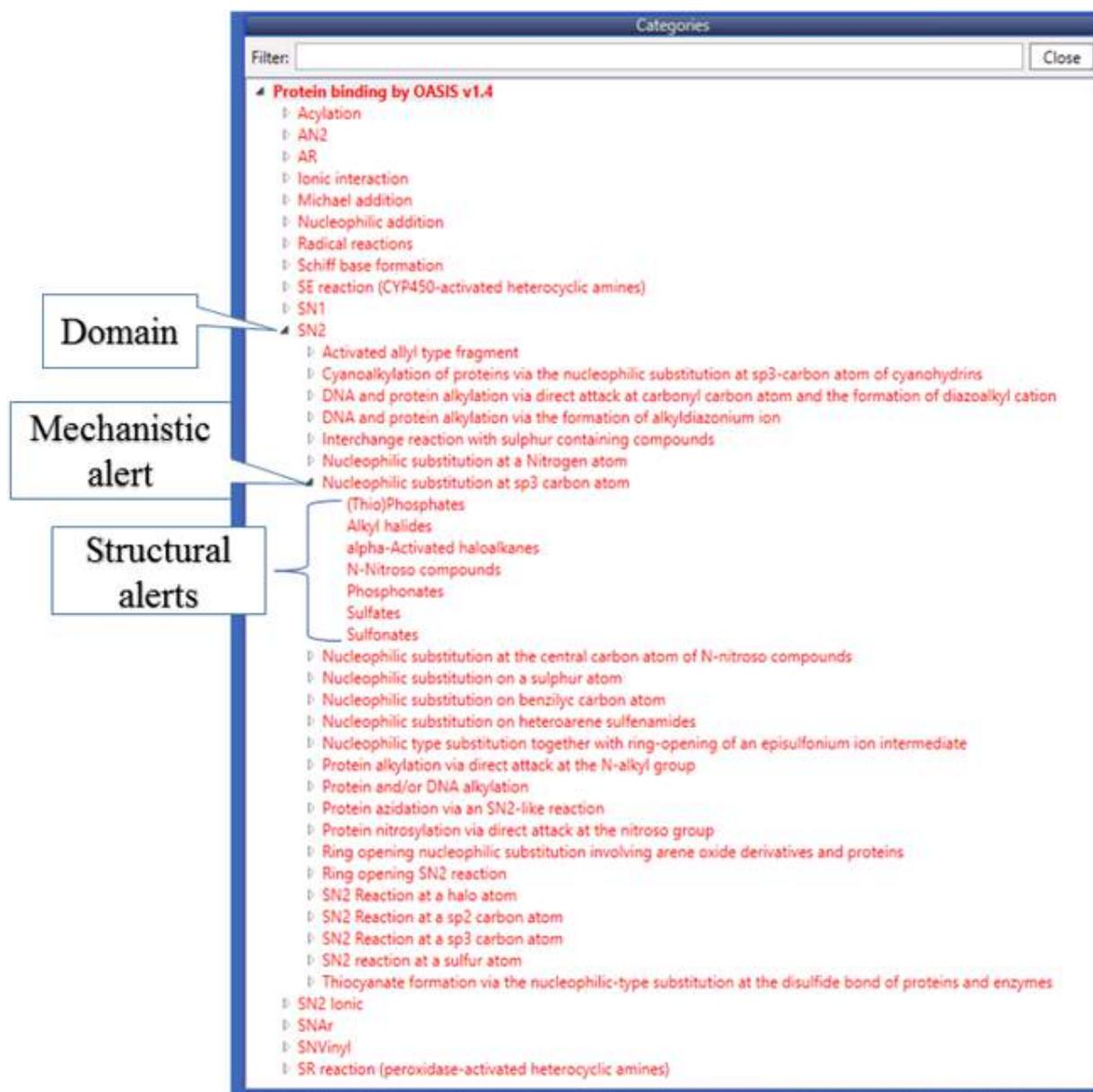


Figure 3

Procedure for categorization

If the user defines category using hierarchical type grouping method, then he/she is allowed to use Domain, Mechanistic alert and Structural alert separately or simultaneously in the categorization procedure. (Figure 4)

The user can remove category related to “Structural alert”, by selecting the

category (1) and moving it down (Figure 4) or he/she can leave it as is (by default):

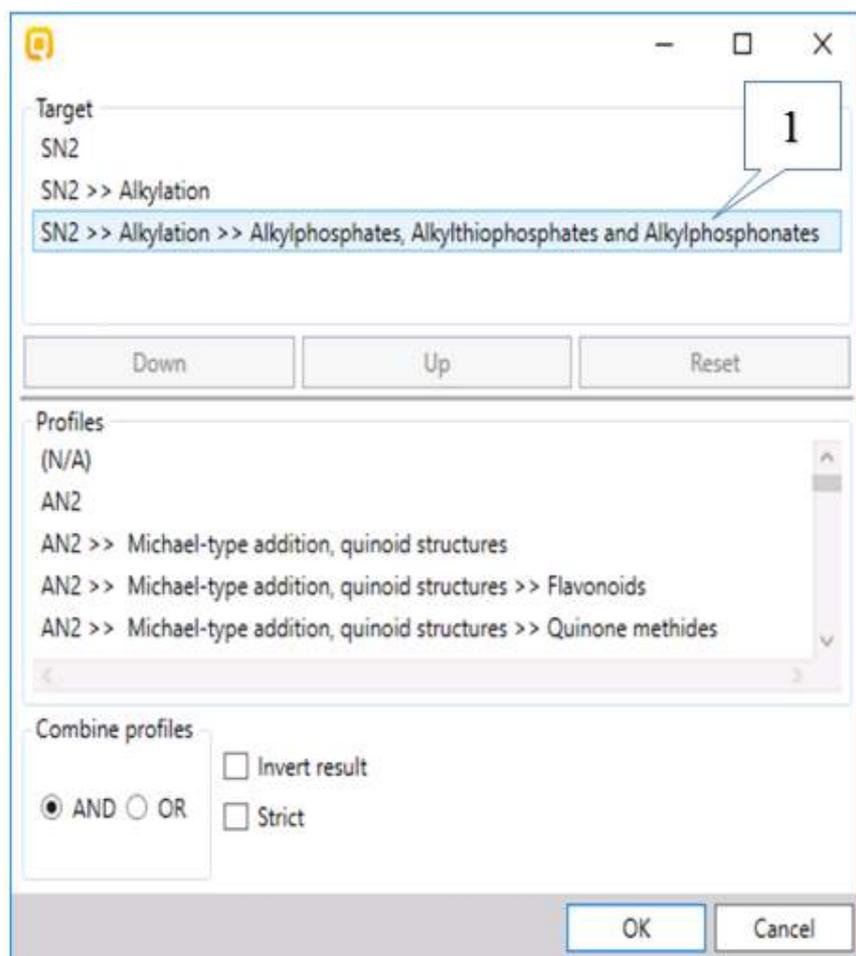


Figure 4

In case the software doesn't identify analogues which answer the criteria of Domain, Mechanistic alert and Structural alert categories combined by AND, then the following message will appear (Figure 5):

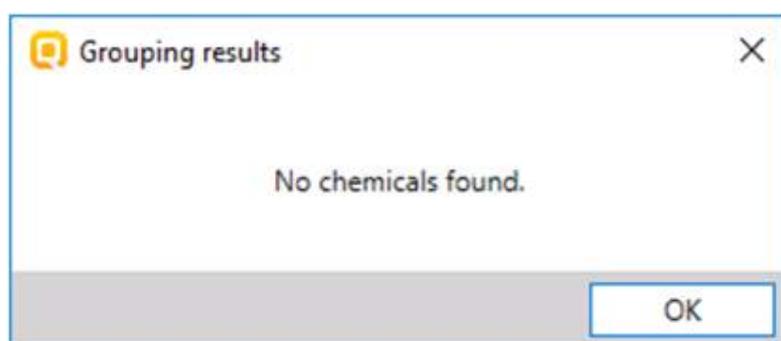


Figure 5

Then the user could expand the category by removing the more specific category (Structural alert) (Figure 6), and use the remaining Domain and Mechanistic alert categories.

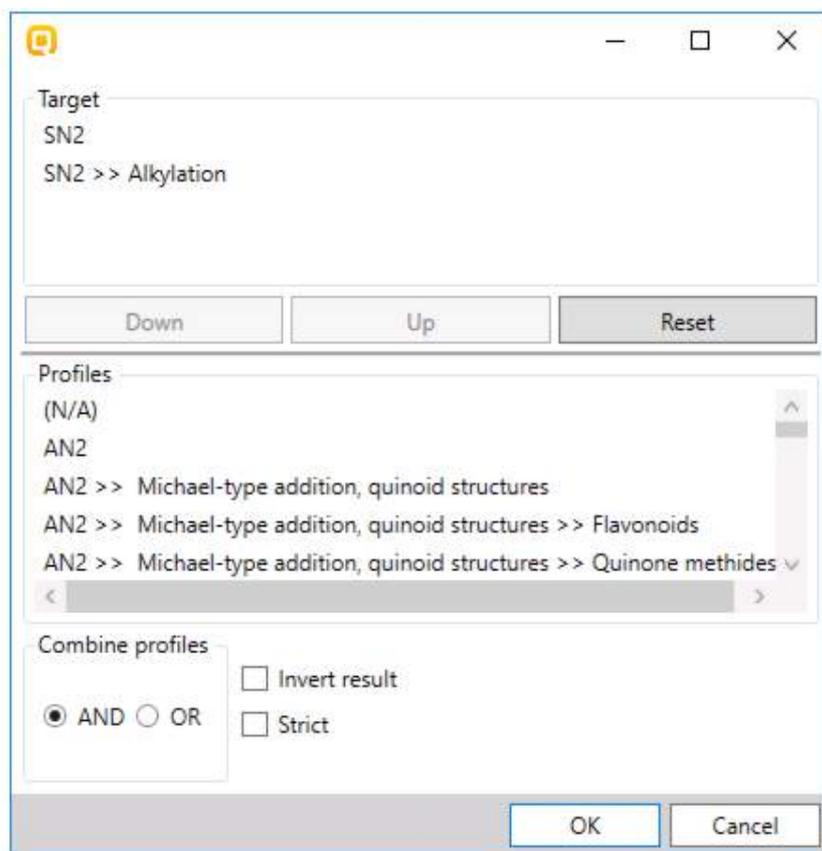


Figure 6

QSAR Toolbox v.4

D.4.3.4. Categorization with metabolism

[Workflow](#) / [Category definition](#) / D.4.3.4. Categorization with metabolism

D.4.3.4. Categorization with metabolism

Grouping with accounting metabolic transformation is a procedure for finding analogues accounting metabolism activation of the chemicals. It has been implemented in Toolbox 3 but improved and extended in Toolbox 4 allowing providing different criteria for separate metabolites.

In order to make a category accounting metabolism, the user have to click on Define with metabolism (1). The “Map similarity options” dialogue appears showing all the generated metabolites of the target chemical by the simulator that was preliminary selected. The dialogue has two subsections which are described below and shown in Figure 1:

- First subsection (2) shows parent and each of the generated metabolites (by the preliminary selected metabolism simulator) is separate rows. This allows defining of different criteria for each of structures for finding analogues.
- Second subsection (3) is working with whole package “parent + metabolites”, i.e. the criteria is provided for the whole package but not for separate metabolite.

A drop down menu (4) is available for each of the structures (in the column “Query”) which allow setting the type of criteria for further looking for analogues.

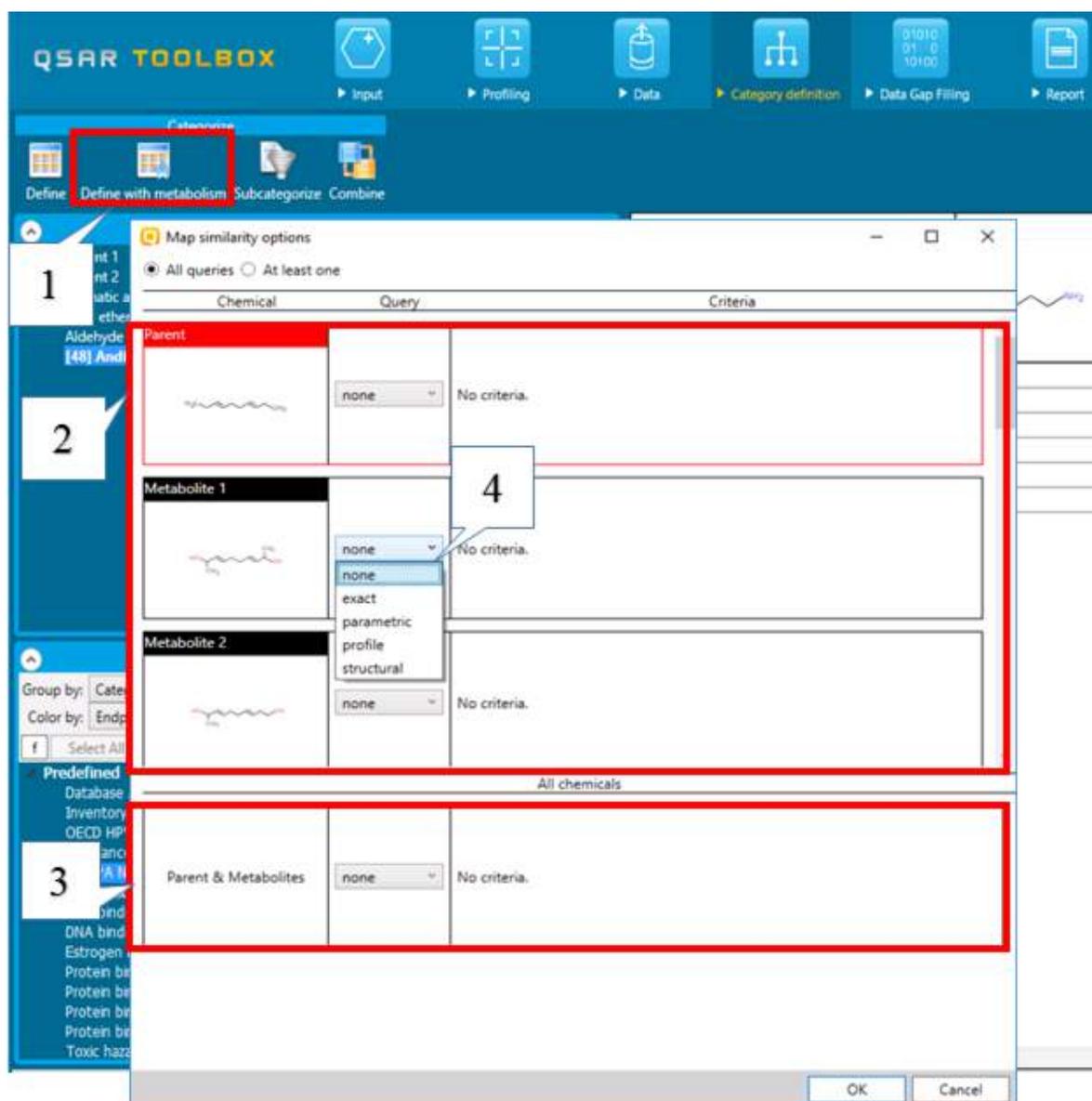


Figure 1

Explanation of different options from the drop down menu

- None – default options; no criteria is set;
- Exact – provides opportunity to search for metabolites in the analogues having exact to the specified metabolite structure; only available for the metabolites and the package “parent + metabolites” but not for the parent chemical;
- Parametric – to have specific value or range of variation of defined parameter (a list with all parameters currently available in the Toolbox is provided);
- Profile – to have specific category by selected profiler (a list with all profilers is provided);
- Structural – to have specific similarity based on the atom centered

fragments.

In short, grouping with metabolism in Toolbox 4.0 allows finding analogues that have:

- parent and its metabolites with defined profile (1)
- metabolite with defined profile (2)
- exact metabolite (3)
- metabolite with defined parameter value (4)
- metabolite similar to defined one (5)
- parent with defined profile, parameter value or structural similar (6)
- combination of above

In addition the user can choose whether all or at least one of the defined criteria to be met (7). (Figure 2)



Figure 2

D.4.3.5. Alert performance

D.4.3.5. Alert performance

Alerts are the main category-building units of many profiling schemes (profilers) and their definition is based on theoretical knowledge and empirical observations. Scores for evaluating performance of alerts have been developed which to inform the user whether the selected alert is suitable for data gap filling. In this respect, calculation of alert performance shows the reliability of a given alert to the selected endpoint.

Alert performance (AP) is used to define how much relevant to a target endpoint an alert is. It reflects the alerts usability for category formation and is applicable for the alerts of the following profilers:

- ✓ *US-EPA New Chemical categories* with respect to *Carcinogenicity endpoint*
- ✓ *Biodeg probability (Biowin 5)* with respect to *Biodegradation endpoints*
- ✓ *Protein binding alerts for skin sensitization by OASIS* with respect to *Skin sensitization endpoint*
- ✓ *DNA alerts for AMES by OASIS* with respect to *AMES Mutagenicity endpoint*

The reliability inform the user whether the selected alert is suitable for category formation. One of the main usages of alerts in Toolbox is to refine preliminary defined chemical category which is associated to specific endpoint. The reliability of alerts is context dependent and are significantly affected by changing the target endpoint, selected databases, selected mode and scale.

In order to use this functionality, the target endpoint needs preliminary to be defined. The databases which will be used for searching of analogues have to be also checked.

When the user select a profiling scheme (1) and click on *Define* (2), a categorization window with the found profiler`s categories in the target structure (3) appears. Then, if the endpoint is preliminary defined (4), an additional section for calculating "*Alert performance*" is designed in this window (5). (Figure 1)

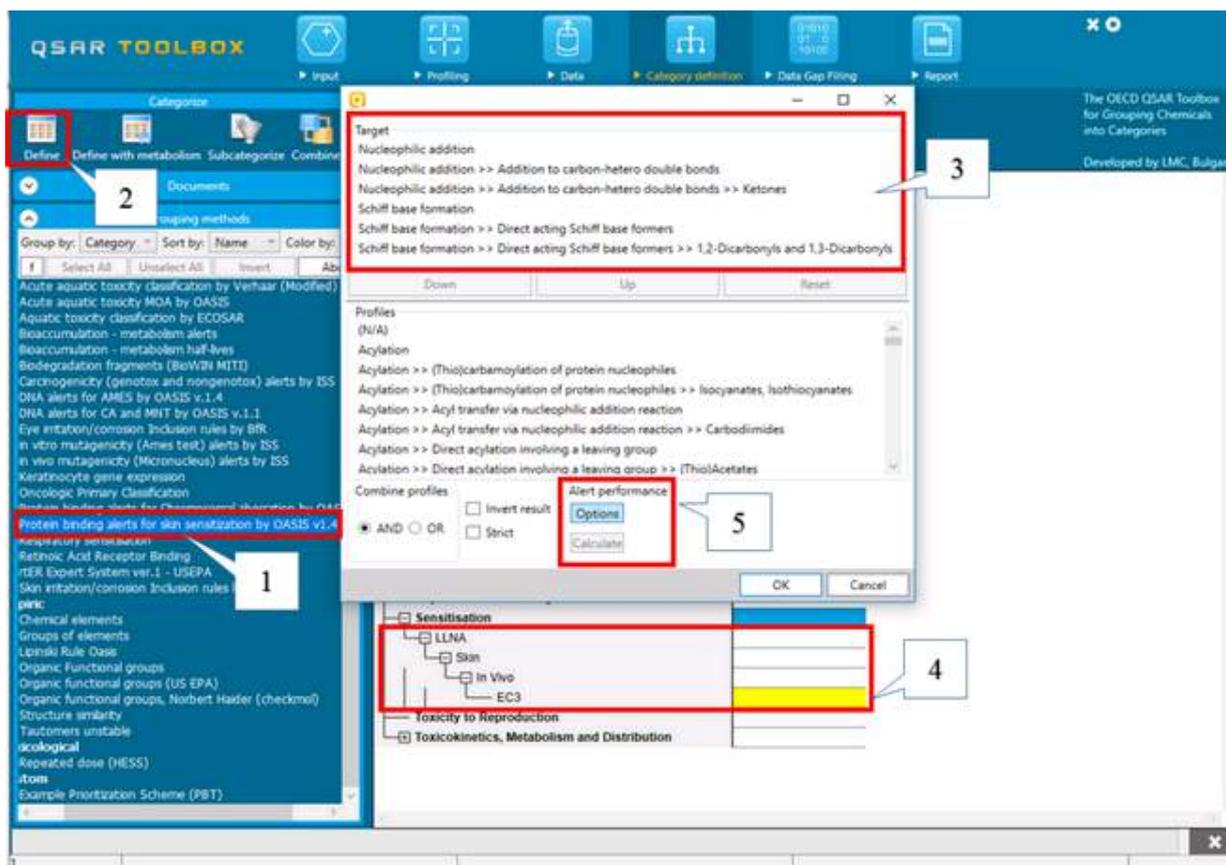


Figure 1

Different options for calculating of *Alert performance* are available within the *Alert performance options* for the selected endpoint. There is a possibility to apply specific scale (1) for the calculation of alert performance (Skin sensitisation II (ECETOC) converts data to Positive and Negative). Additional option for applying different weight of the available data is also provided in a drop-down menu (2). Worst case scenario is set as default (i.e. "Maximal" in this case). (Figure 2)

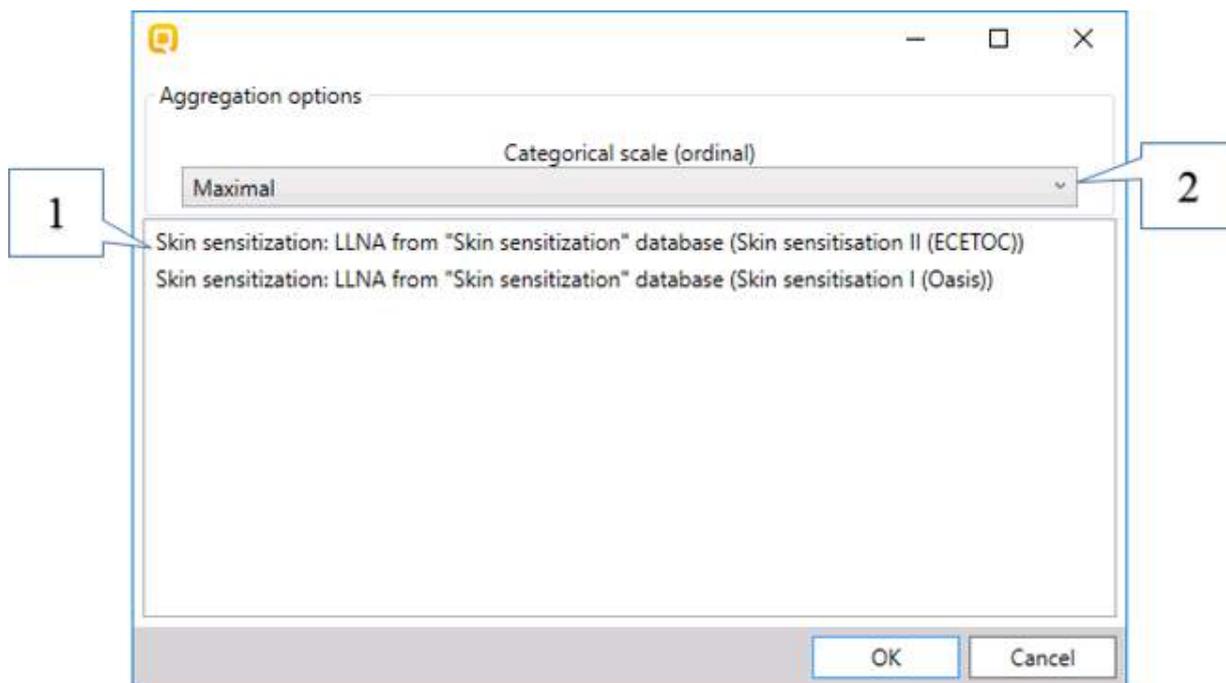


Figure 2

When the options are chosen, performance can be calculated for all alerts found in the target structure or for each of them individually. The performance of an alert represents the number of chemicals (1) with data related to the predefined scale across all chemicals (2) from the selected databases, which have the same alert as the target structure. The distribution of the available data (3) according to a given effect (e.g. positive, negative) in percentages is also provided. (Figure 3)

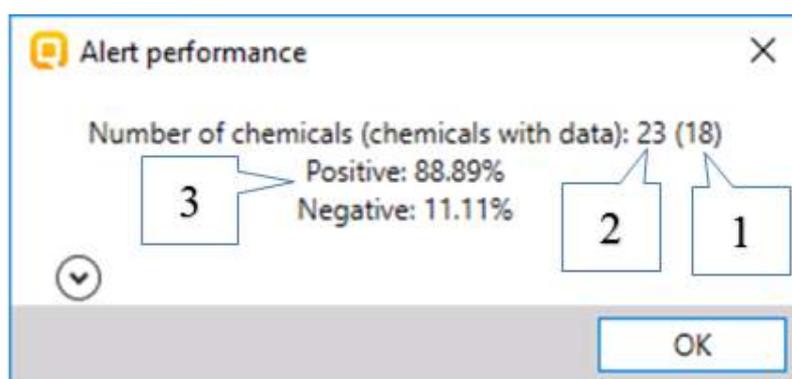


Figure 3

Calculation of alert performance accounting metabolism is also possible.

QSAR Toolbox v.4

D.4.4. Subcategorization

[Workflow](#) / [Category definition](#) / D.4.4. Subcategorization

D.4.4. Subcategorization

The second step of refining the broader category and defining the category of structurally and mechanistically similar analogues is the subcategorization procedure. The user can verify the mechanistic robustness of the analogue approach. If the identification of analogues was performed according to a specific mechanism or mode of action, then the target chemical and the analogues will already have the same relevant mechanisms and modes of action. Nevertheless, the analogues may also have additional mechanisms and modes of action due to additional functional groups in their molecule. In this respect subcategorization procedure is applied to refine the categories (eliminating dissimilar structures).

The broader category can be refined when subcategorization is applied. For example 242 esters are identified by US-EPA category for a target chemical (1) (Figure 23)

When the category is defined the user can check if all these 242 esters have same DNA mechanism of interaction. In this respect the subcategorization procedure is for can be used for checking that all these 216 chemical have same DNA binding mechanism.

Subcategorization panel includes list with same grouping/profiling methods as listed in Profiling section. In the subcategorization procedure metabolism could be accounted for.

After the category is defined the user has to click Subcategorize button (2), and select one of the DNA binding profilers (3) (Figure 1)

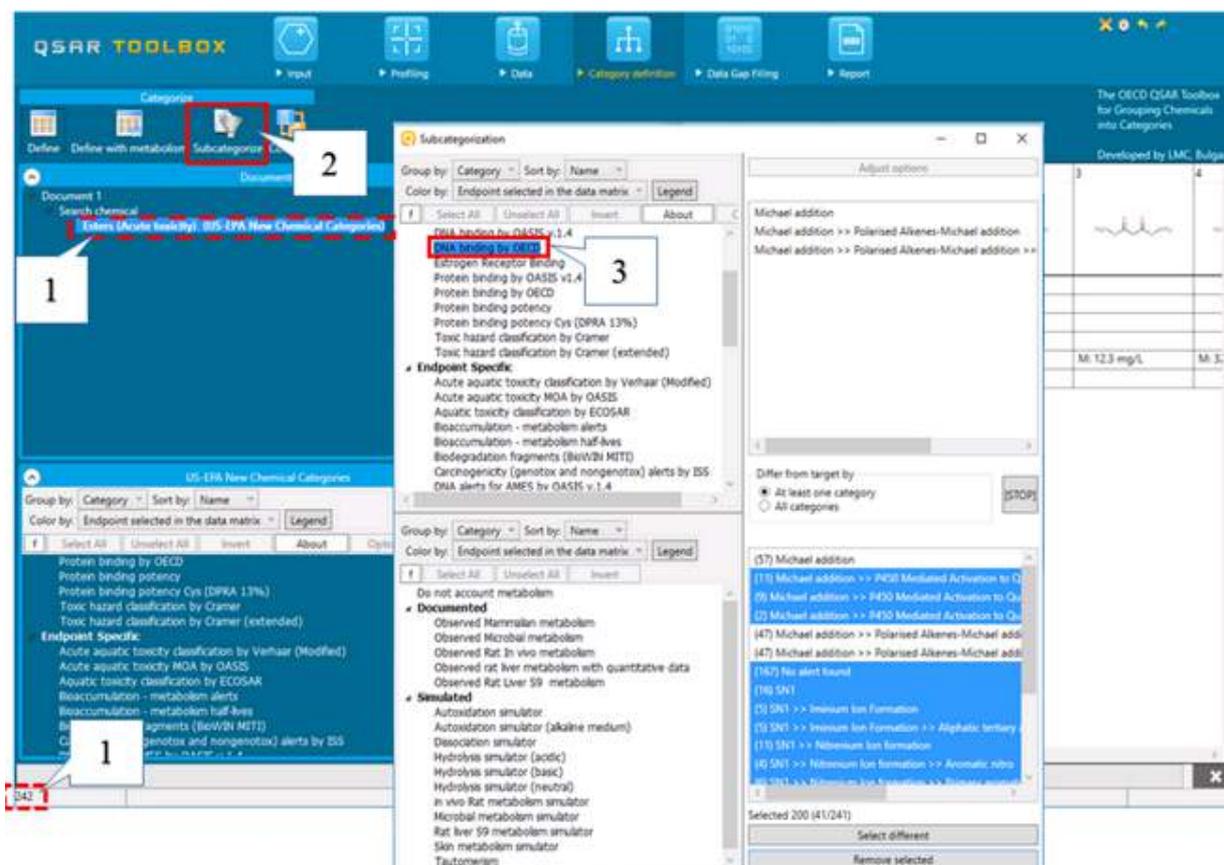


Figure 1

Chemical(s)/analogue(s) which have different mechanism of interaction than of the target chemical are highlighted by a blue background. Others which are not highlighted have at least one category same as those of the target.

If the user selects all categories radio button (1) (Figure 2), then all analogues included in the refined category should have all categories combined by logical conjunction (AND) (2). After removing the dissimilar analogues by clicking on Remove selected button (3), the software defines a new category, which is subcategory of the parent category.

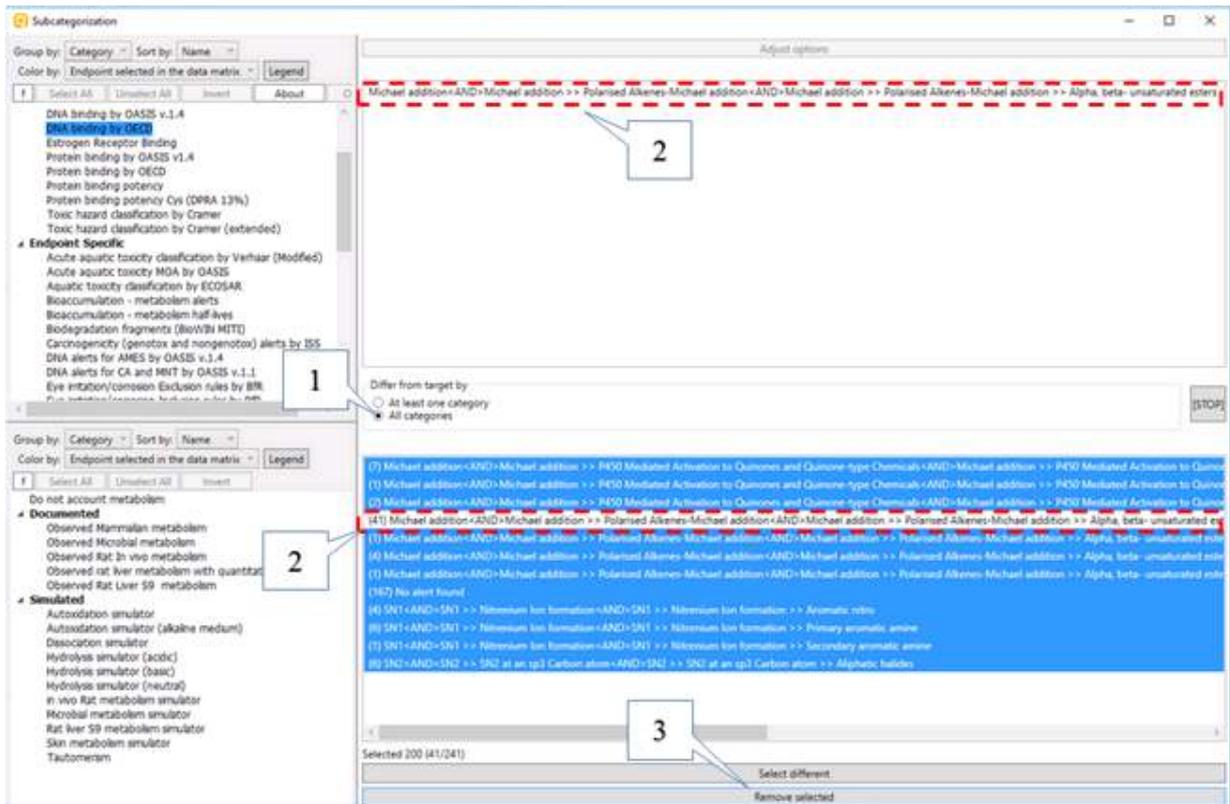


Figure 2

The subcategory appears as a sub-node of the first category (1). New data matrix consists of the chemicals, which remain after the subcategorization is created. (Figure 3)

The screenshot displays the OECD QSAR Toolbox interface. A search result for 'Esters (Acute toxicity) (US-EPA New Chemical Categories)' is shown, with the subcategory 'DNA binding by OECD' highlighted. The 'Filter endpoint tree...' panel is also visible, showing a table of chemical structures and their associated data for four endpoints. The table includes columns for 'Structure', '1 [target]', '2', '3', and '4'. The 'Structure' column shows chemical structures. The '1 [target]' column shows a list of checkboxes for various endpoints: Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information (42/142), and Human health hazards. The '2', '3', and '4' columns show numerical values: M: 572 mg/L, M: 0.36 mg/L, M: 2.3 mg/L, and M: 6.3 mg/L respectively.

Structure	1 [target]	2	3	4
	<input type="checkbox"/> Structure info			
	<input type="checkbox"/> Parameters			
	<input type="checkbox"/> Physical Chemical Properties			
	<input type="checkbox"/> Environmental Fate and Transport			
	<input type="checkbox"/> Ecotoxicological Information (42/142)	M: 572 mg/L	M: 0.36 mg/L	M: 2.3 mg/L
	<input type="checkbox"/> Human health hazards			M: 6.3 mg/L

Figure 3

-0-

QSAR Toolbox v.4

D.4.5. Combine categories

[Workflow](#) / [Category definition](#) / D.4.5. Combine categories

D.4.5. Combine categories

Combining defines a new category by combining logically members from already existing chemical lists. When there are more than one chemical list in the document tree (1), the user can emerge them by clicking on the Combine button (2). Combine categories window with all available lists appears. Some or all lists of chemicals can be selected and then the chemicals could be combined by logically AND/OR (4) operand. Creating of the new category finishes with clicking on OK (5). (Figure 1)

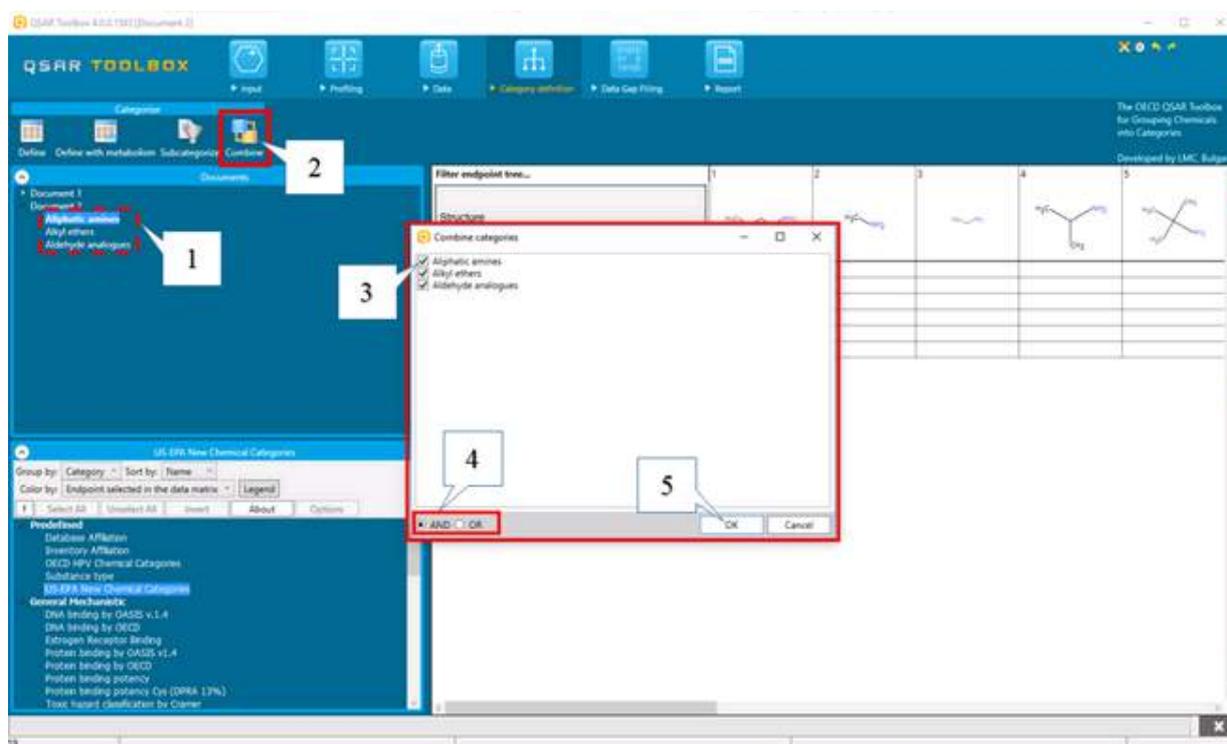


Figure 1

After selection of the combination logic the software defines a new category including chemicals which answer the criteria of the defined category. (Figure 2)

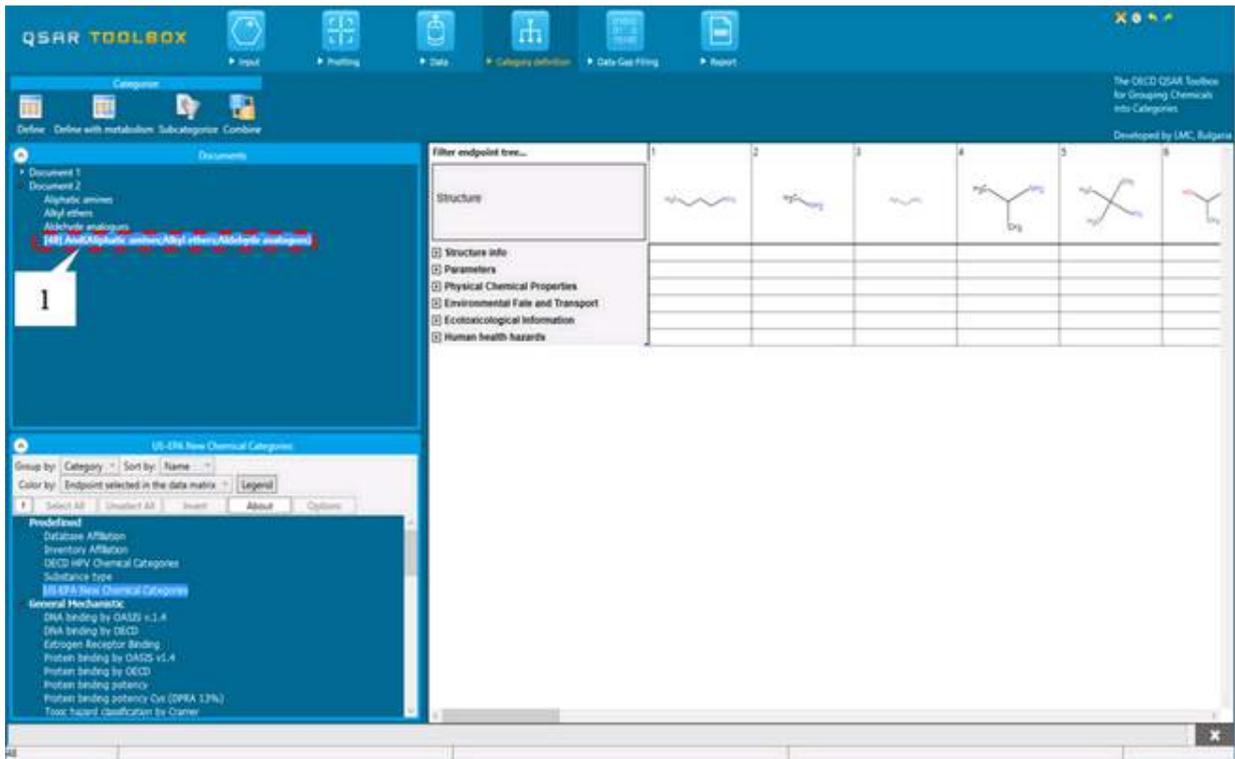


Figure 2

-0-

D.5. Data Gap Filling

QSAR Toolbox v.4

D.5.1. Data Gap Filing methods

[Workflow](#) / [Filling data gaps](#) / D.5.1. Data Gap Filing methods

D.5.1. Data Gap Filling methods

- [Read-across](#)
- [Trend analysis](#)
- [\(Q\)SAR models](#)

Read-across and trend analysis use the available experimental data in the data matrix to fill a data gap. "(Q)SAR models" gives access to a library of external (Q)SAR models which have been integrated into the Toolbox.

The choice of the most relevant data gap mechanism depends on the following considerations:

- Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitization or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
- Trend analysis is the appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
- "(Q)SAR models" can be used to fill a data gap if no adequate analogues are found for a target chemical.

When selecting read-across or trend analysis, the available data in the data matrix is used for filling a data gap. The user can further reduce the data set by using the profilers to eliminate chemicals which have different profiles compared to the target chemical.

It can be distinguished between two scenarios:

- If a specific mechanism or mode of action relevant for the endpoint is identified for the target chemical, then all the analogues considered should have the same mechanism or mode of action.
- If no specific mechanism or mode of action relevant for the endpoint is identified for the target chemical, then none of the structural analogues considered should have specific mechanisms or modes of actions either.

Caution 

Categorical type data such as data for skin sensitization or Ames mutagenicity endpoints are calculated using read across or QSAR methods.

-0-

QSAR Toolbox v.4

D.5.2. Data Gap Filling procedure

[Workflow](#) / [Filling data gaps](#) / D.5.2. Data Gap Filling procedure

D.5.2. Data Gap Filling procedure

As explained in [Category definition](#) section, the Toolbox has identified chemicals which have similar structural functionality as the target chemical and for which experimental results are available. The workflow illustration for the example chemical 4-nitrobenzoyl chloride (CAS No 122-04-3) is presented below. Since the sensitization is a “qualitative” endpoint the data gap can be filled by read-across.

After highlighting the cell (1) in the matrix corresponding to the data gap to be filled, the user has to select the data gap filling methods (2) (Figure 1). Additionally, two workflows options are included: *Standardized* (SW) and *Automated* (AW) (3). They represent algorithms for automated data gap filling. Those algorithms are developed for skin sensitization (LLNA and GPMT data) and acute aquatic toxicity to fish (*Pimephales promelas*, Mortality, LC50, 96 h). Once started, the automated workflows follow the implemented logic and finished with prediction without interaction by the user. The same algorithms have been implemented as standardized execution of data gap filling. The main differences as compared to the AWs are that the domain of application is expanded in the SWs (including other species, durations, etc.) and SWs allow interactions by the user and thus, different selection than those in AW could be done.

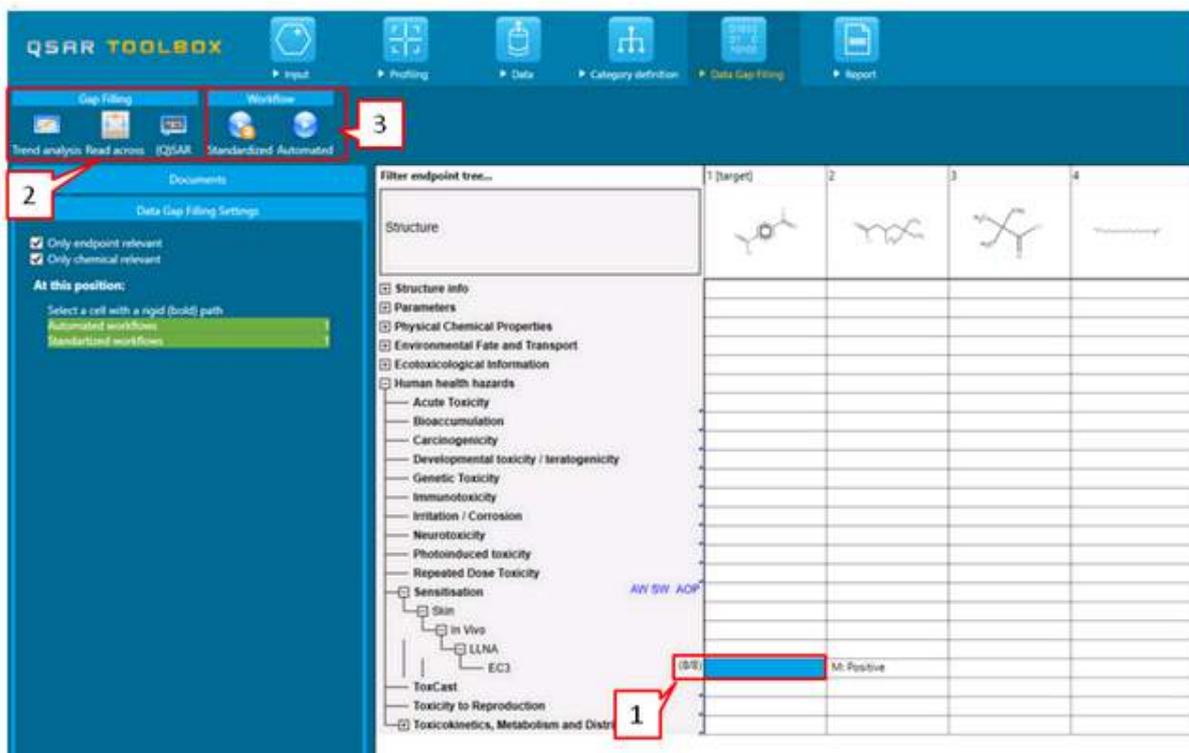


Figure 1

Before entering data gap filling window the possible data inconsistency window appears (1) (Figure 2)

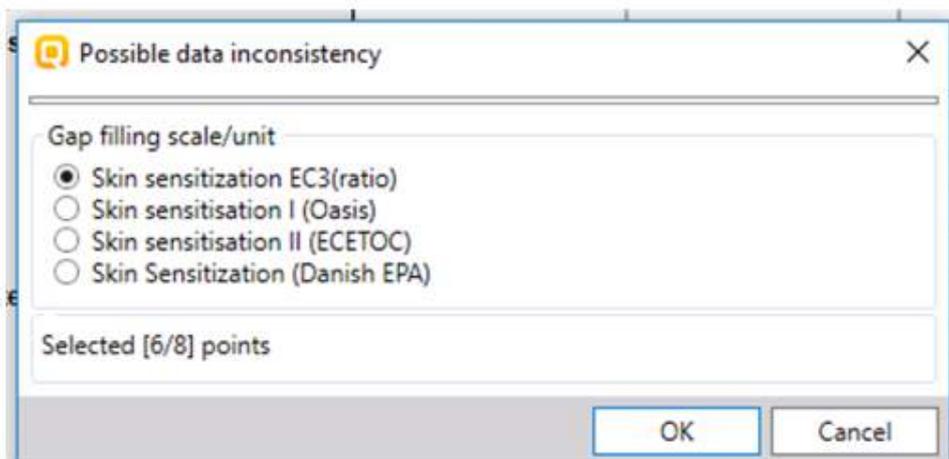


Figure 2

This feature alerts the user for possible data inconsistencies.

In the example illustrated by Figure 3, there are two fields: Metabolic activation (1) and Strain (2). Data included in these two metadata fields are mixed up in data gap filling.

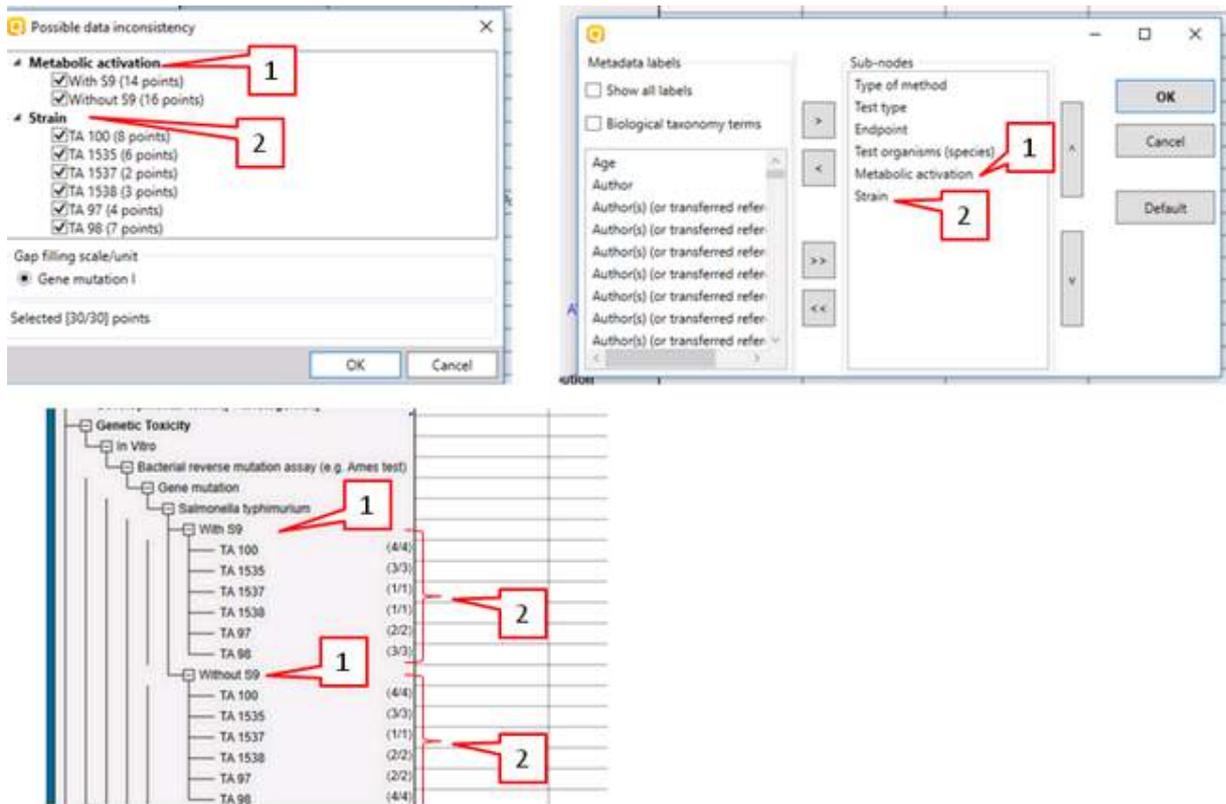


Figure 3

More detailed information about scales is presented in the [About](#) section.

Note  Only one scale/unit is allowed in data gap filling.

The number located on the bottom of the window in Figure 1 (e.g. Selected 8/8) means that 8 data points from a total of 8 data points will enter the data gap filling module.

QSAR Toolbox v.4

D.5.3. Data Gap Filling window

[Workflow](#) / [Filling data gaps](#) / D.5.3. Data Gap Filling window

D.5.3. Data Gap Filling window

The next three snapshots illustrate the different types of gap filling methods:

Read across window:

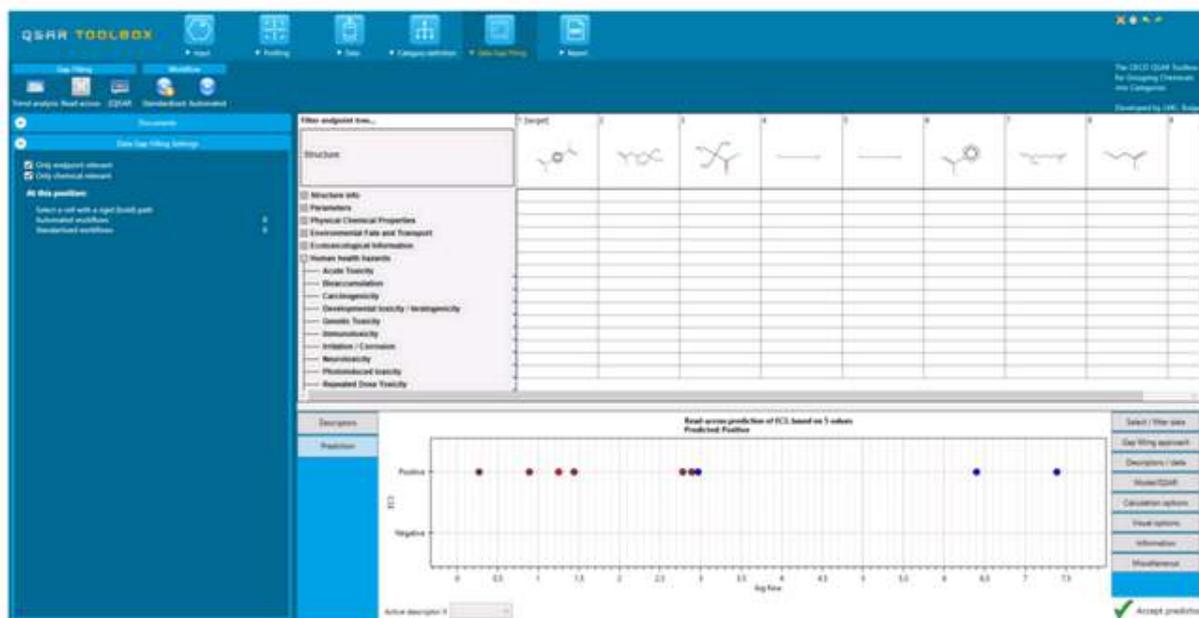


Figure 4

Trend analysis window:

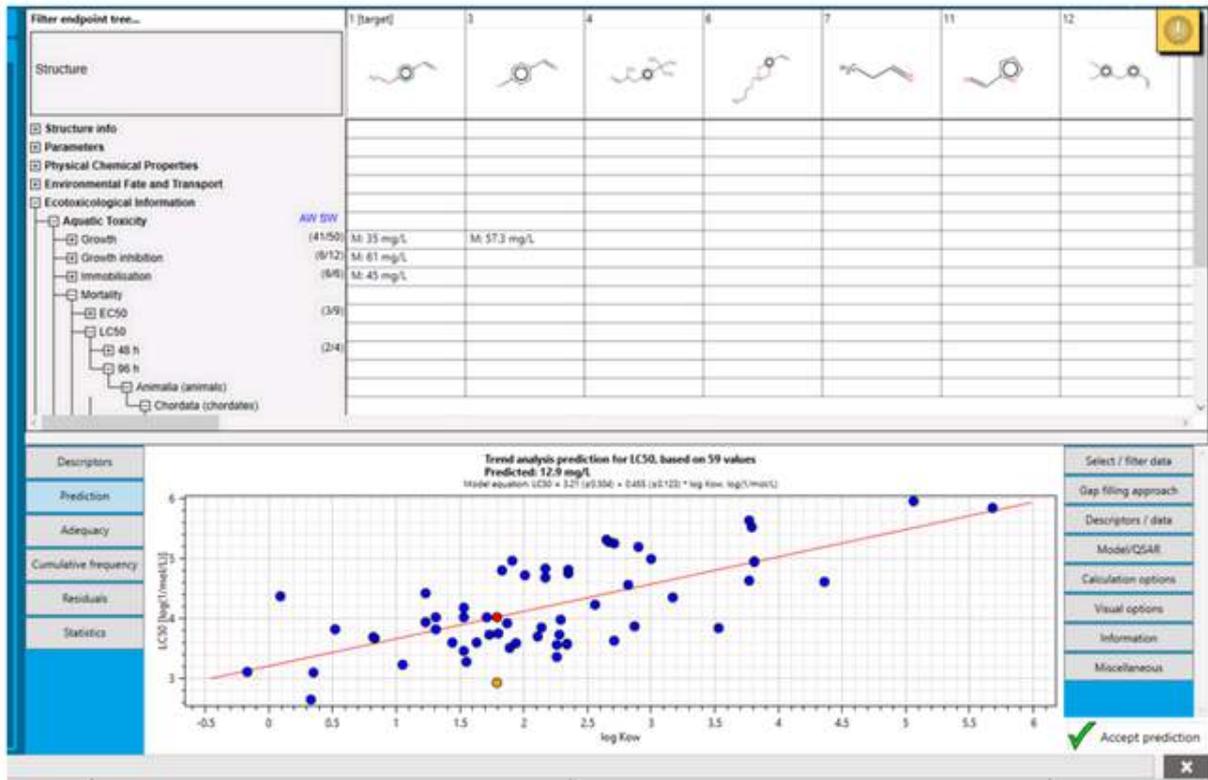


Figure 5

(Q)SAR models window:

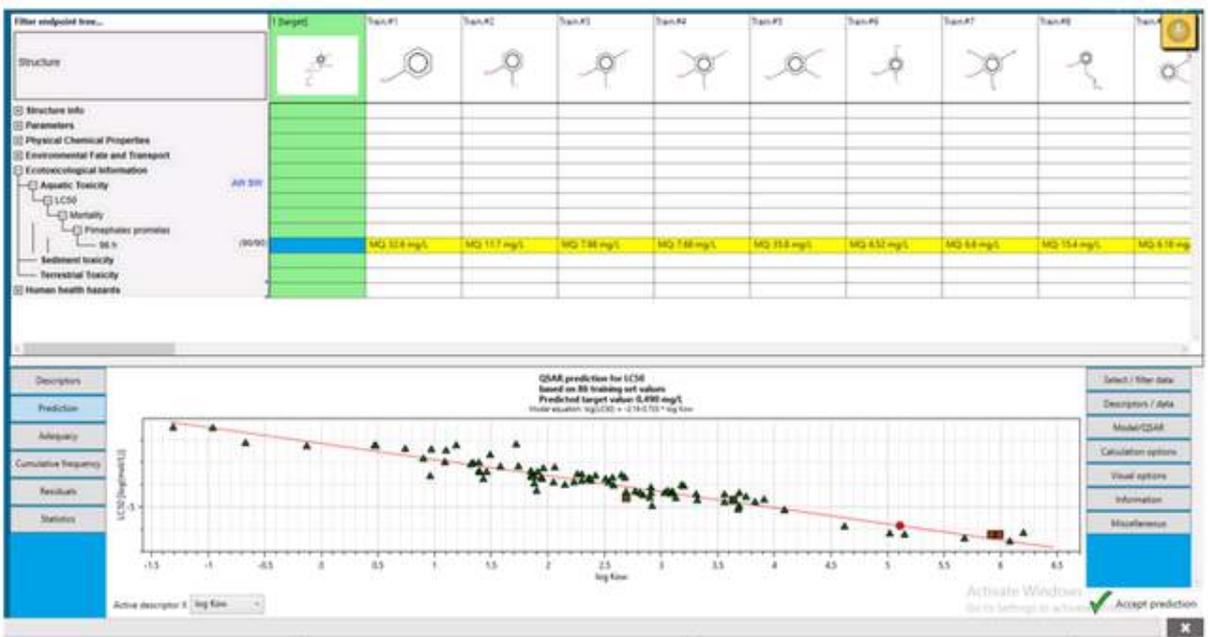


Figure 6

Standardized and automated workflows:

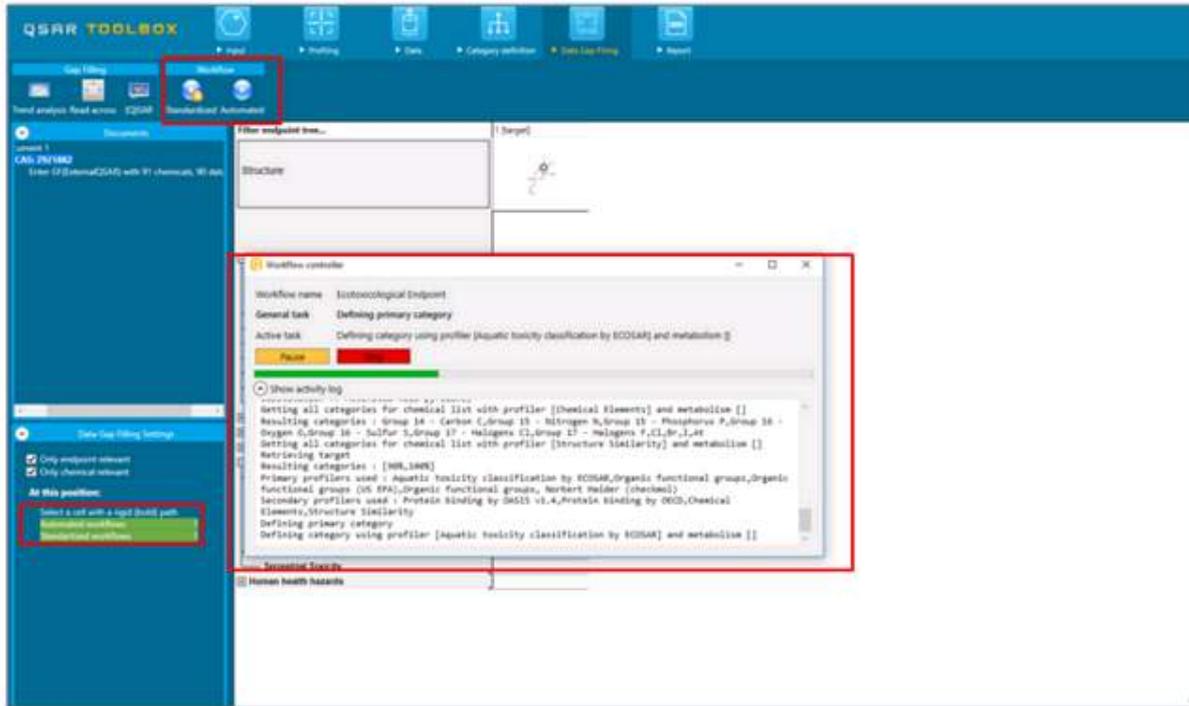


Figure 7

QSAR Toolbox v.4

D.5.4. Common features in three gap filling approaches

[Workflow](#) / [Filling data gaps](#) / D.5.4. Common features in three gap filling approaches

D.5.4. Common features in three gap filling approaches

The three different gap filling approaches, while different, have common features that all share:

- [Left-side panel](#)
- [Menus of data gap filling](#)
- [Right click functionality](#)

-0-

QSAR Toolbox v.4

D.5.4.1. Left-side panel

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#)/ D.5.4.1. Left-side panel

D.5.4.1. Left-side panel

The left-side panel is shown in Figure 1.

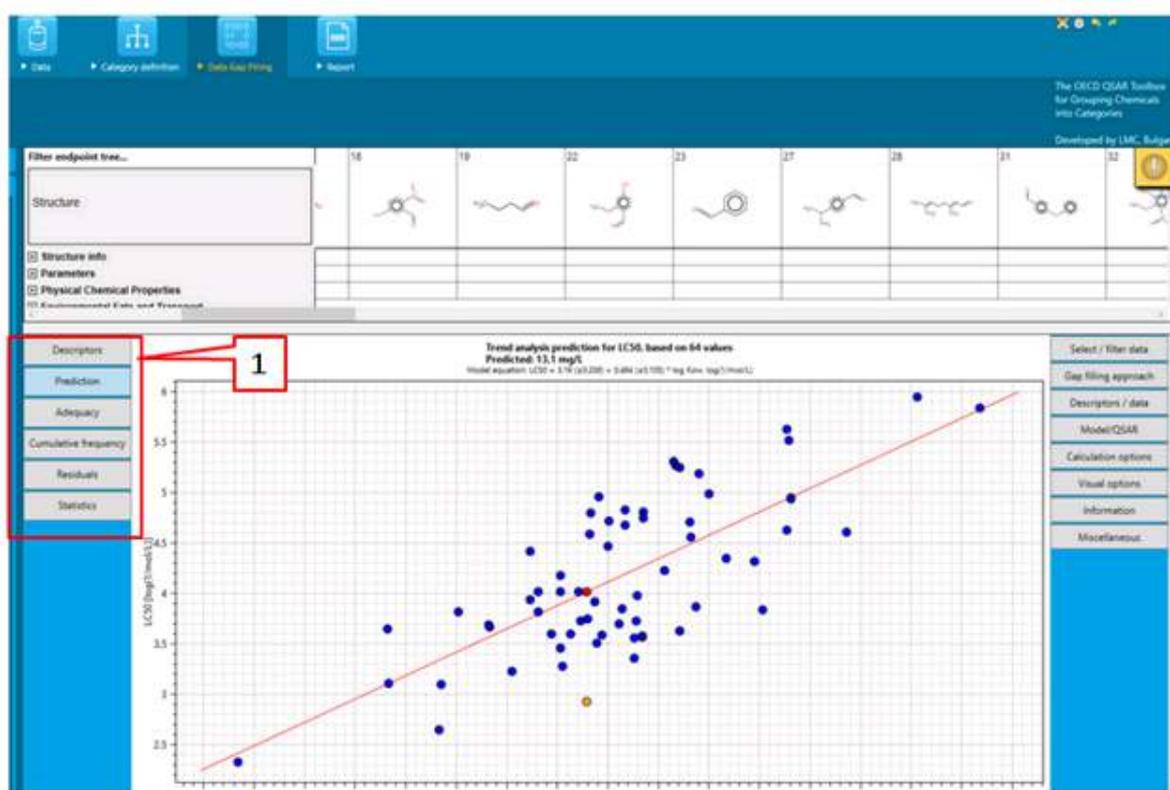


Figure 1

- Descriptors panel

The *Descriptors* panel is shown in Figure 2. Here the user can select the descriptor that he/she likes to see as the X axis of the graph. The Y is the value of the data-points most commonly set in logarithmic scale. The units in Y-axis could be changed in Options, subsection Units. The descriptors panel is available for Read-across, Trend analysis and (Q)SAR models data gap filling approaches. Log Kow is usually the default descriptor. To change a descriptor double click on the active descriptor(1) to remove it and then double-click on the selected descriptor (2) from the bottom list.

Descriptors	Active descriptors			
	Name	Unit	Data points	Correlation
Prediction	log Kow		64	0.746
Adequacy				
Cumulative frequency	All descriptors			
Residuals	Name	Unit	Information	
	Kp (Octanol/air (Koa) model)	m ³ /kg bw/dwt		
Statistics	log BCF max	log(L/kg wet weight)		
	Log Koa (Air-water partition coefficient model)			
	Log Koa (Henry's law constant model)			
	log ₁₀ Multibase			
	LUMO Energy	eV		
	Maximum distance	Å		
	Maximum donor delocalizability	(a.u.) ² /eV		
	Mean Melting Point	°C		
	Melting Point (Adapted Joback Method)	°C		
	Melting point (Gold and Ogle method)	°C		
	Molar refraction I			
	Molar refraction II			
	Molecular Weight	Da		
	Number of aromatic bonds			
Number of cyclic bonds				
Number of double bonds				
Number of heavy atoms				
Number of nitro groups				

Figure 2

- Prediction panel

The *Prediction* panel, shown on Figure 3, is the same for the three data gap filling approaches with only some slight differences - for read-across it highlights the neighbor data-points with respect to descriptor used in Y-axis (by default 5 nearest which can be changed in Prediction approach options), for trend analysis it displays the trend line. For all approaches this panel shows the distribution of data-points, and the predicted value of the target chemical, on a graph.

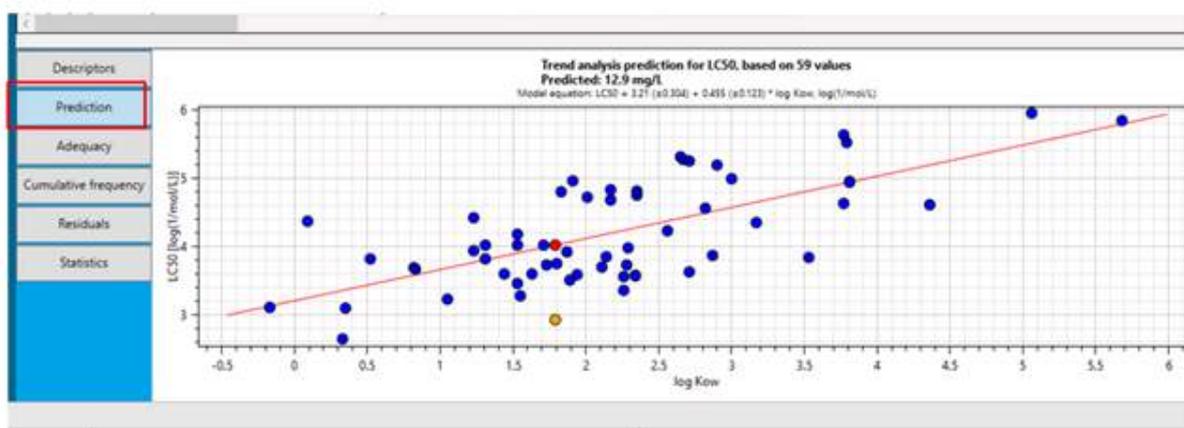


Figure 3

- Adequacy panel

The Adequacy panel, Figure 4, houses the adequacy graph. It is a graph

where on one axis there is the observed value and on the other is the predicted value. Also Coefficient of determination (R^2) and Adjusted Coefficient of determination (R^2_{adj}) (1) are displayed on the top of the graph.

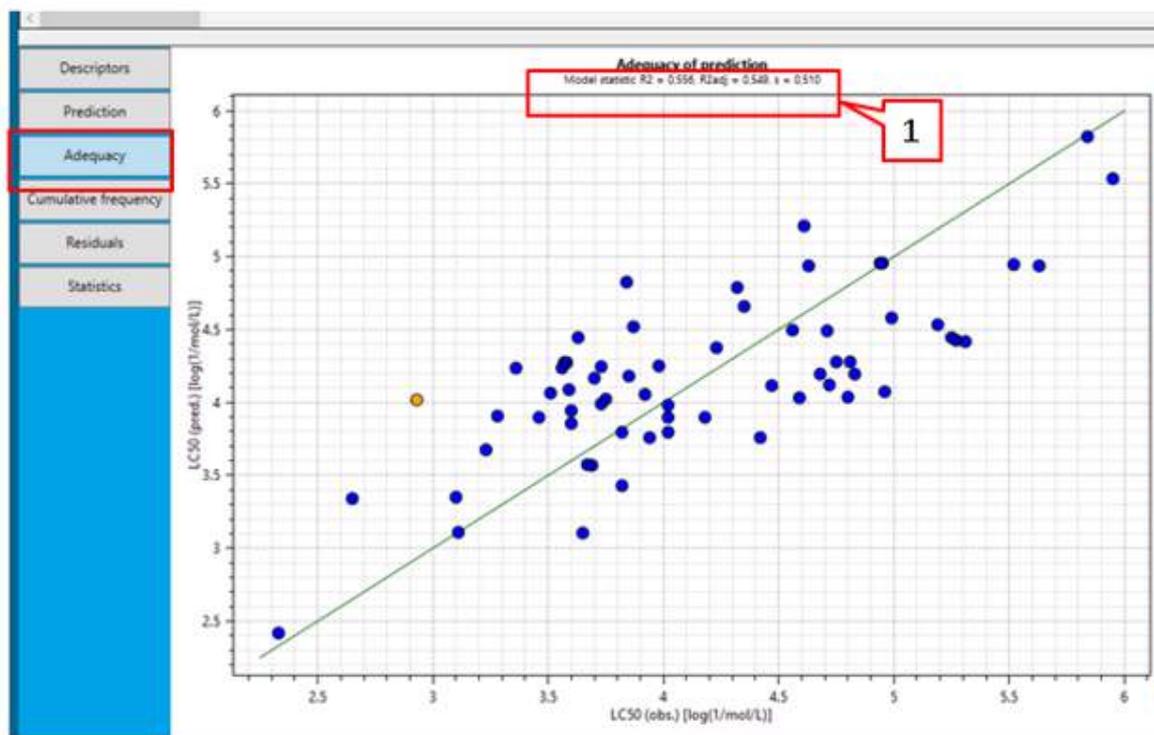


Figure 4

- Cumulative frequency panel

The cumulative frequency is the frequency with which the value of the residual ($EP_{obs} - EP_{calc}$) is less than or equal to a reference residual value.

For example, if cumulative frequency is 70% at residual value of 0.2, then 70% of training set members have residuals less or equal to 0.2. Below is the snapshot with cum. frequency graph. (Figure 5)

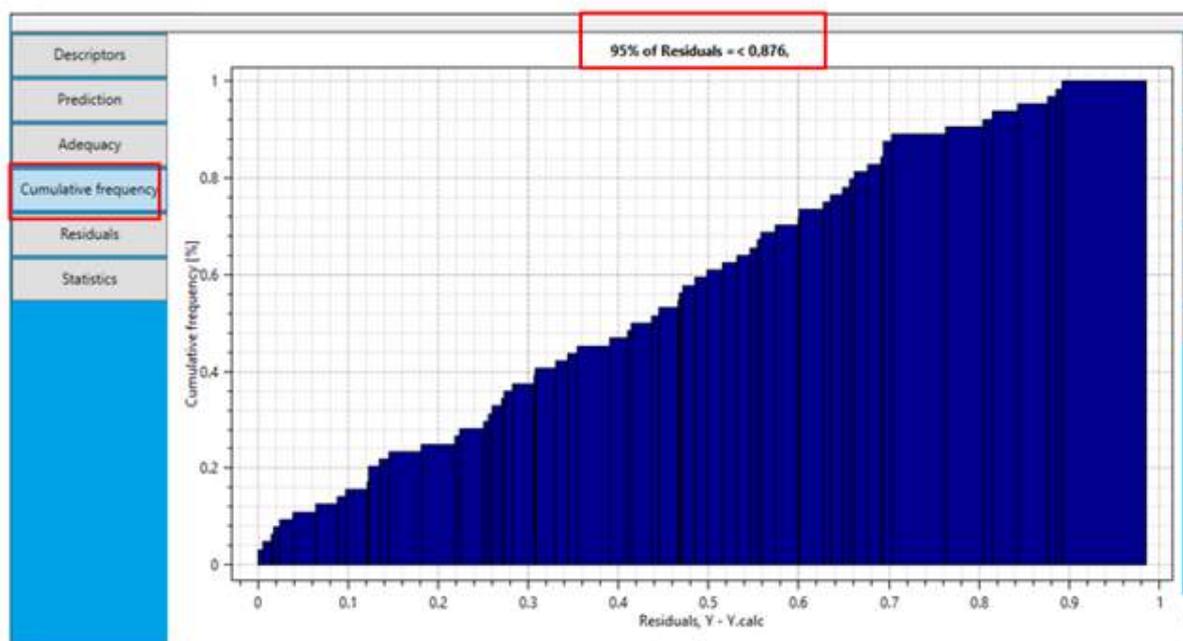


Figure 5

- Residuals panel

This panel includes graph on which the distribution of residuals for a given endpoint versus a specified descriptor is illustrated. (Figure 6)

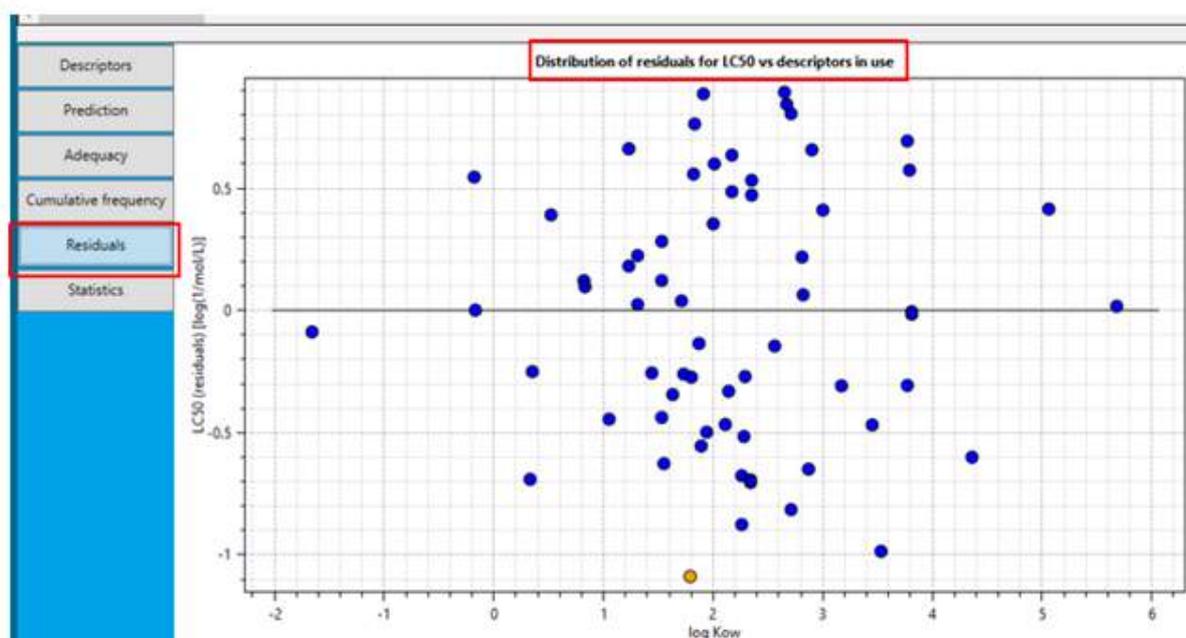


Figure 6

- Statistics panel

This panel shows the statistical characteristics of regression equation (Figure 7). The upper part of the panel includes statistics for regression equation (1), while in the second part coefficients included in the regression equation are shown (2).

Descriptors	Statistical characteristics	TA model
	Number of data points, (N)	64
Prediction	Coefficient of determination, (R2)	0,556
	Adjusted coefficient of determination, (R2adj)	0,549
Adequacy	Coefficient of determination - leave one out, (Q2)	N/A
	Sum of squared residuals, (SSR)	16,2
Cumulative frequency	Standart deviation of residuals, (sN)	0,502
	Sample standart deviation of residuals, (s)	0,510
Residuals	Fisher function, (F)	77,8
Statistics	Fisher treshold for statistical significance, (Fa)	N/A
	b0	
	- model descriptor	Intercept
	- coeff. value	3,19
	- coeff. range	±0,259
	- significance	No
	- max covariation	
	b1	
	- model descriptor	log Kow
	- coeff. value	0,464
	- coeff. range	±0,105
	- significance	No
	- max covariation	

Figure 7

QSAR Toolbox v.4

D.5.4.2. Menus of data gap filling

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / D.5.4.2. Menus of data gap filling

D.5.4.2. Menus of data gap filling

Menus of the data gap filling are shown in Figure 1.

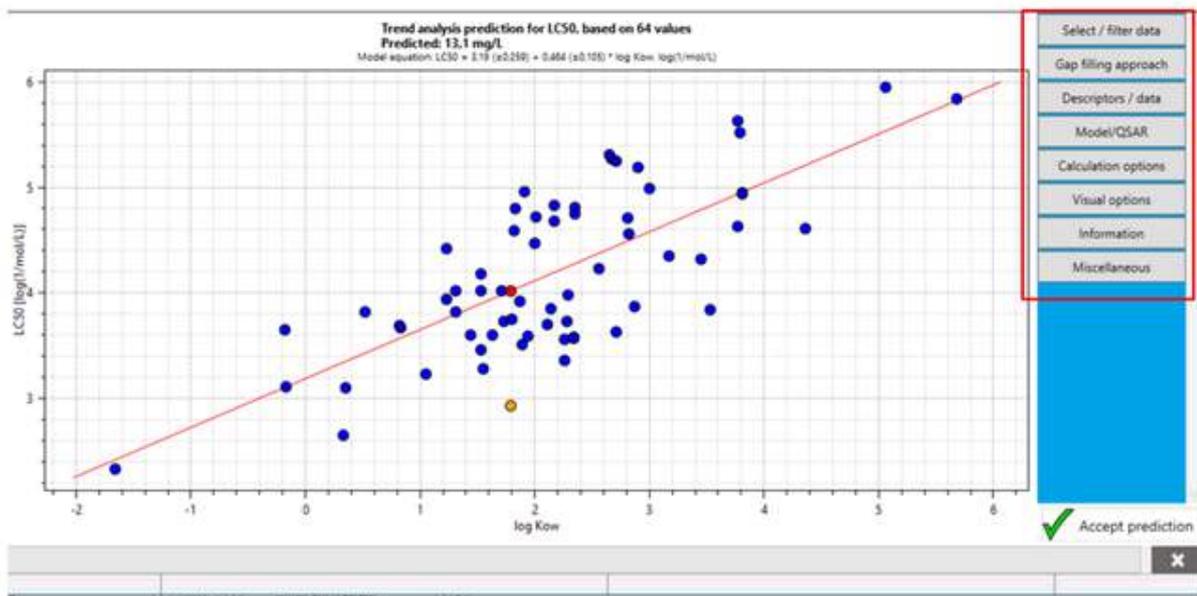


Figure 1 – menus of data gap filling

Clicking on each option opens a menu with more functionalities. They are explained in the sub-sections of D.5.4.2.

-0-

QSAR Toolbox v.4

D.5.4.2.1. Select /filter data

[Workflow / Filling data gaps / Common features in three gap filling approaches/ Menus of data gap filling / D.5.4.2.1. Select /filter data](#)

D.5.4.2.1. Select/filter data

This functionality includes different menus for filtering and selecting data: (Figure 1)

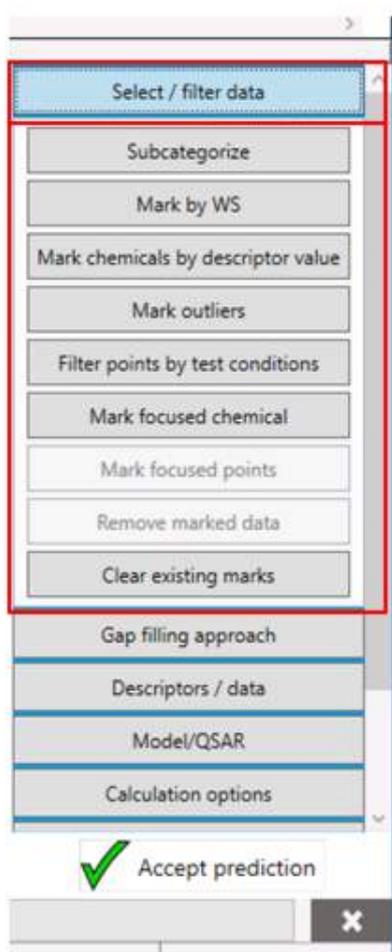


Figure 1

QSAR Toolbox v.4

D.5.4.2.1.1. Subcategorize

[Workflow / Filling data gaps / Common features in three gap filling approaches/ Menus of data gap filling / Select/filter data/ D.5.4.2.1.1. Subcategorize](#)

D.5.4.2.1.1. Subcategorize

Subcategorization is one of the most powerful tools available to the user. It provides the features to refine the broader category into a more consistent group, more pertinent set of chemicals for the user to derive a prediction from using the chemical's properties.

Trend-analysis method for prediction of Acute toxicity (LC50, 96 h, *Pimephalis Promelas*) is used as an example: pressing Subcategorisation (1) opens the Subcategorisation window (2) (Figure 11). Select *Colour by endpoint* selected in the data matrix (3) in order to see which profilers are most suitable for the subcategorisation. A color legend (5) is visualized by pressing *Legend* (4). The user is advised to use those profilers first to narrow down the list of analogues (6).

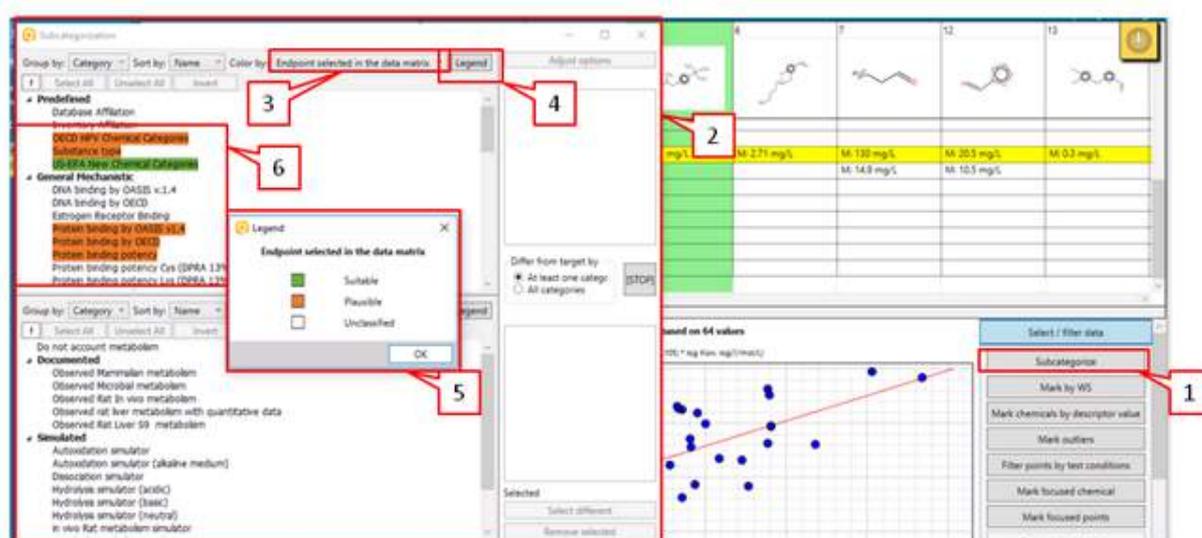


Figure 1

Selecting a profiler US-EPA New chemical category (1) re-profiles the

target and all analogues (Figure 2). The properties of the analogues are then compared with the properties of the target chemical (2). In this particular example, 15 analogues are classified as Phenols and this differentiates them from the target (classified as aldehyde), i.e. they are expected to insert different mechanism of action (the group is coloured in blue) (2). The number of the selected analogues compared to the total number of analogues is shown in 3. They are also coloured on the graph (pale blue dots) and on the data-matrix (green) (4). Pressing *Remove selected* (5) will eliminate those analogues.



Figure 2

The left part of the Subcategorization panel includes all profiling/categorizing methods which can be used for evaluating and refining the category. More detailed information for the endpoint and related subcategorization method is illustrated in Categorization section in the Background information for some grouping methods table. Along with the profiling methods (1) all observed metabolism and simulated transformation tables (2) are included in the Subcategorization panel. (Figure 3)

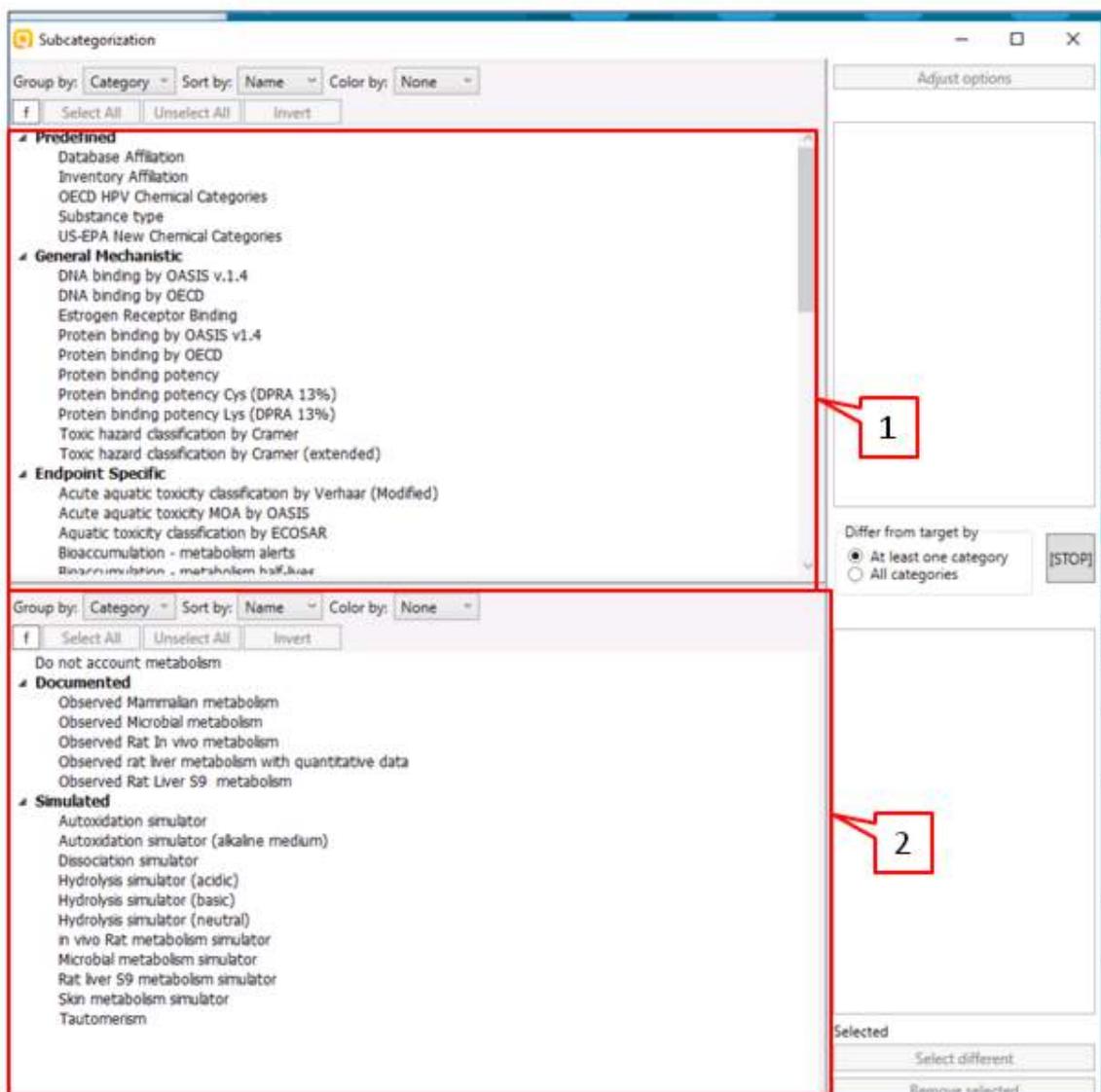


Figure 3

In order to apply some of the available simulators (abiotic/biotic) the user has to select the desired profiling method (1) and click on the related simulator (2) (Figure 4). For the purposes of prediction of skin sensitization the following simulators could be used: Skin sensitization and Autoxidation simulators.

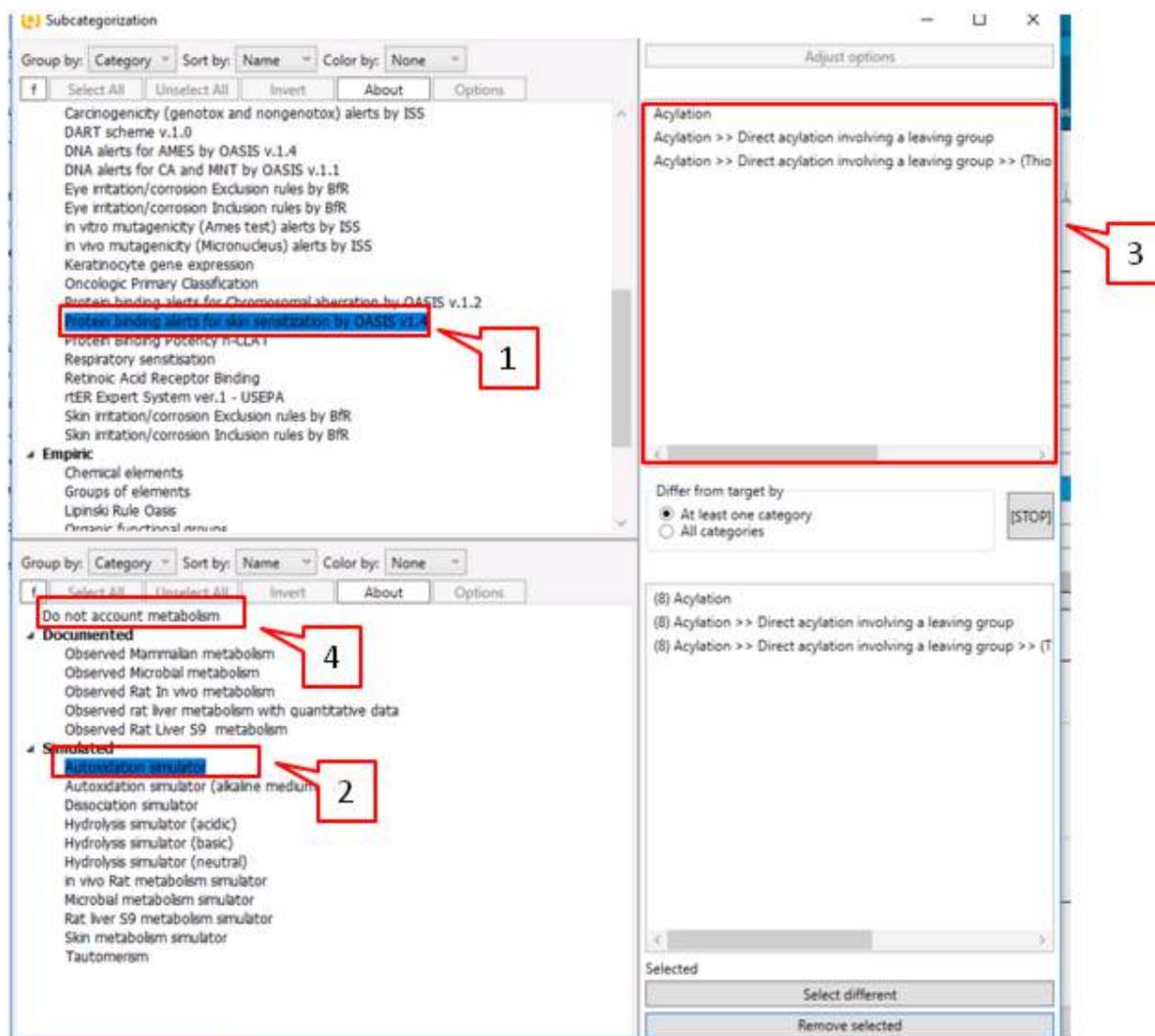


Figure 4

The following simulators are related to prediction of genotoxicity: Rat liver metabolism simulator and Rat liver S9 metabolism simulator.

When a simulator is included in the profiling process the software starts to metabolize the chemicals in the category and applies the selected profile to the package of the target and its metabolites. This process could be time consuming as the Databases are not previously metabolized. The profiling results of a package, target and metabolites, are displayed on the Target panel of Subcategorization window (3) (Figure 4).

The user has to click on *Do not account metabolism* in order to turn off metabolizing and profiling of metabolites (4) (Figure 4).

Note: Keep in mind that all Databases and Inventories are not metabolized.

Selection options are implemented both for the profilers (1) and the simulated/observed metabolism simulators(2) (Figure 5).

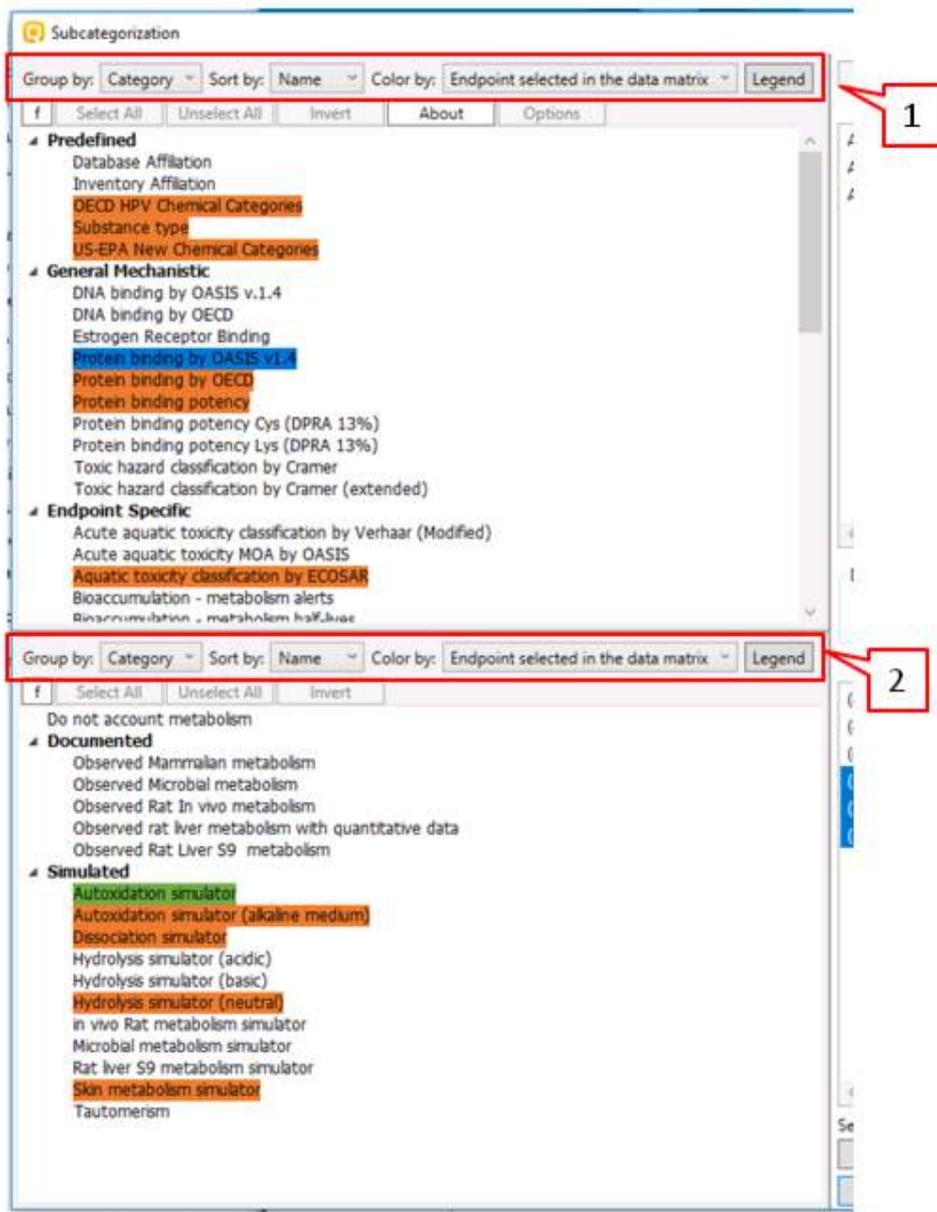


Figure 5

- o *Group by*: allows grouping based on different criteria (see Figure 6)

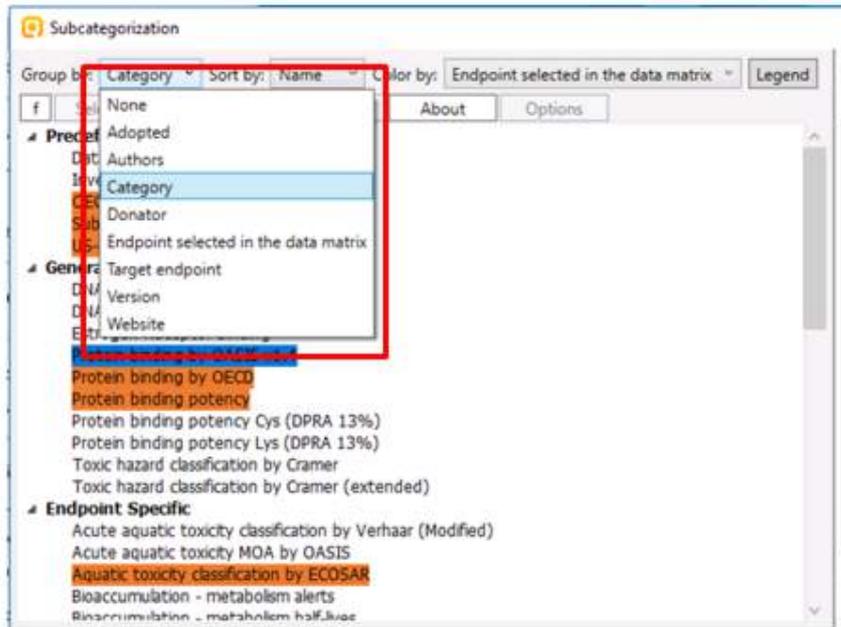


Figure 6

- o *Sort by*: allows sorting based on different criteria (see Figure 7)

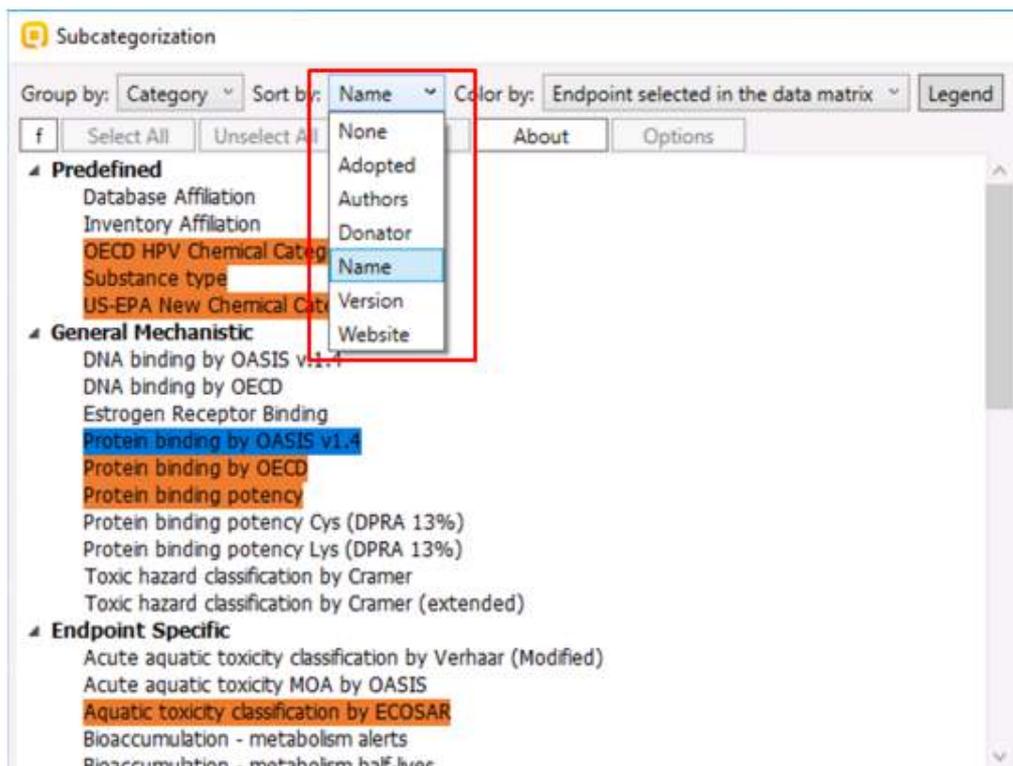


Figure 7

- o *Color by*: colouring of profilers/simulators based on the Endpoint selected in the data matrix(or the Target endpoint)(Figure 8)

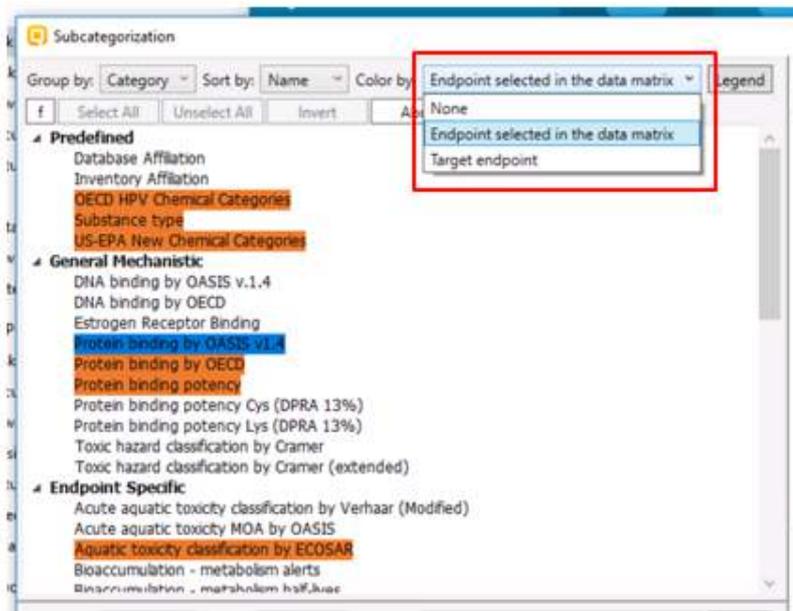


Figure 8

- o *About* - Allows visualization of the information and documentation behind each profiler (Figure 9):
 - Select the profiler (1);
 - Press *About* (2);
 - An *About* window is displayed (3);
 - The full information about the profiler is displayed after selecting Documentation (4).

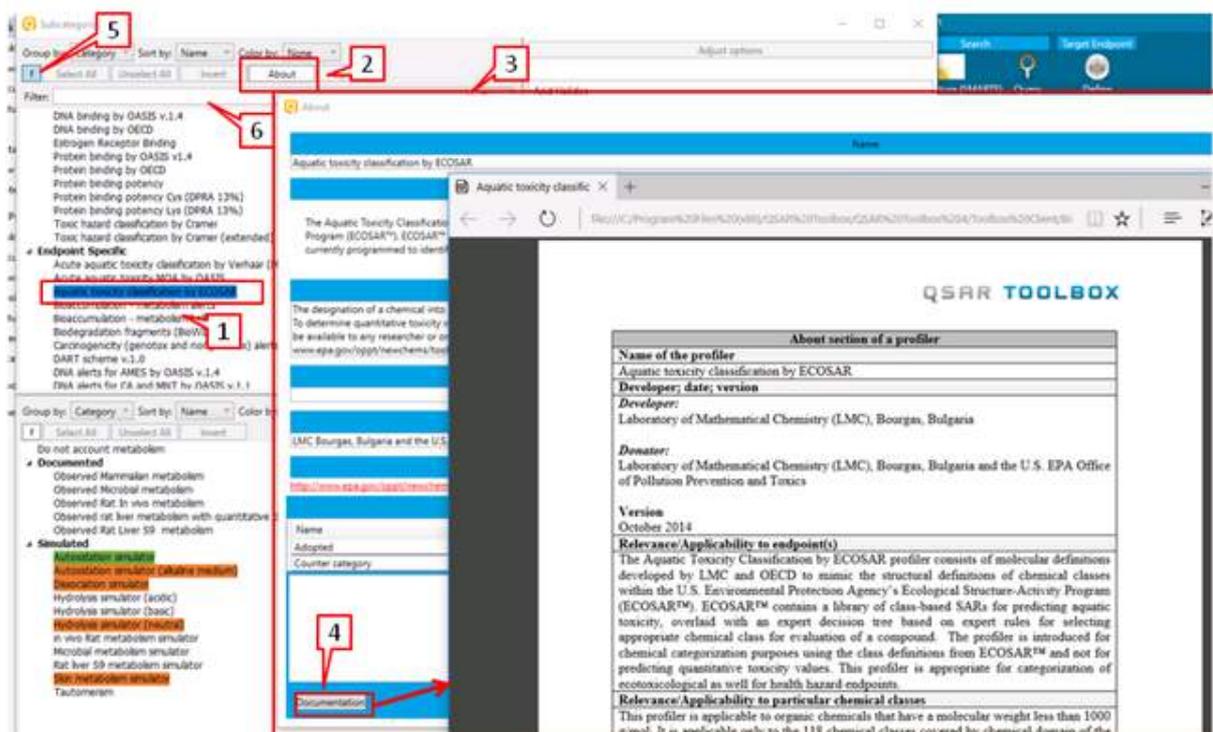


Figure 9

- o *Filter*: allows quick finding of a profiler: Press \mathcal{L} (5) and then type some key words in (6). (Figure 9)

The right panel of Subcategorization window consists of two parts Target and Analogues. Both panels include profiling results of target and analogues. The Target panel includes profiling results (1) across selected profiling method (2) for the target and its metabolites if available (3). (Figure 10)

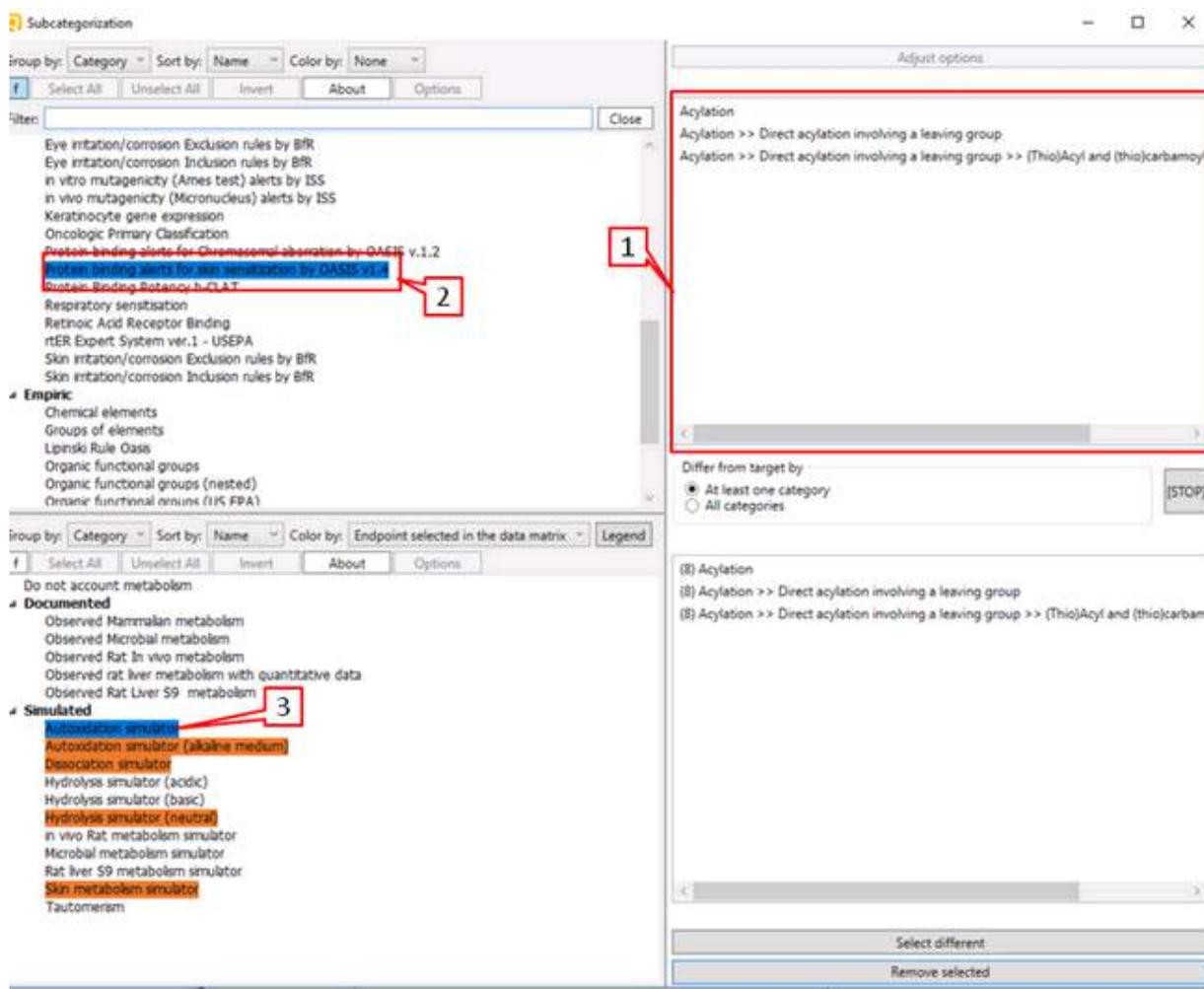


Figure 10

The Analogues panel includes profiling results for analogue chemicals (1), based on the applied profiler (2) and simulator (3). Analogues having different profiling results than those of the target chemical are selected (highlighted by) with a blue background (4). The number in brackets indicates the number of analogues in the respective category (in the example 1 analogue is profiled as alkyl halide). The user can eliminate dissimilar analogues by clicking on *Remove* button (5). (Figure 11)

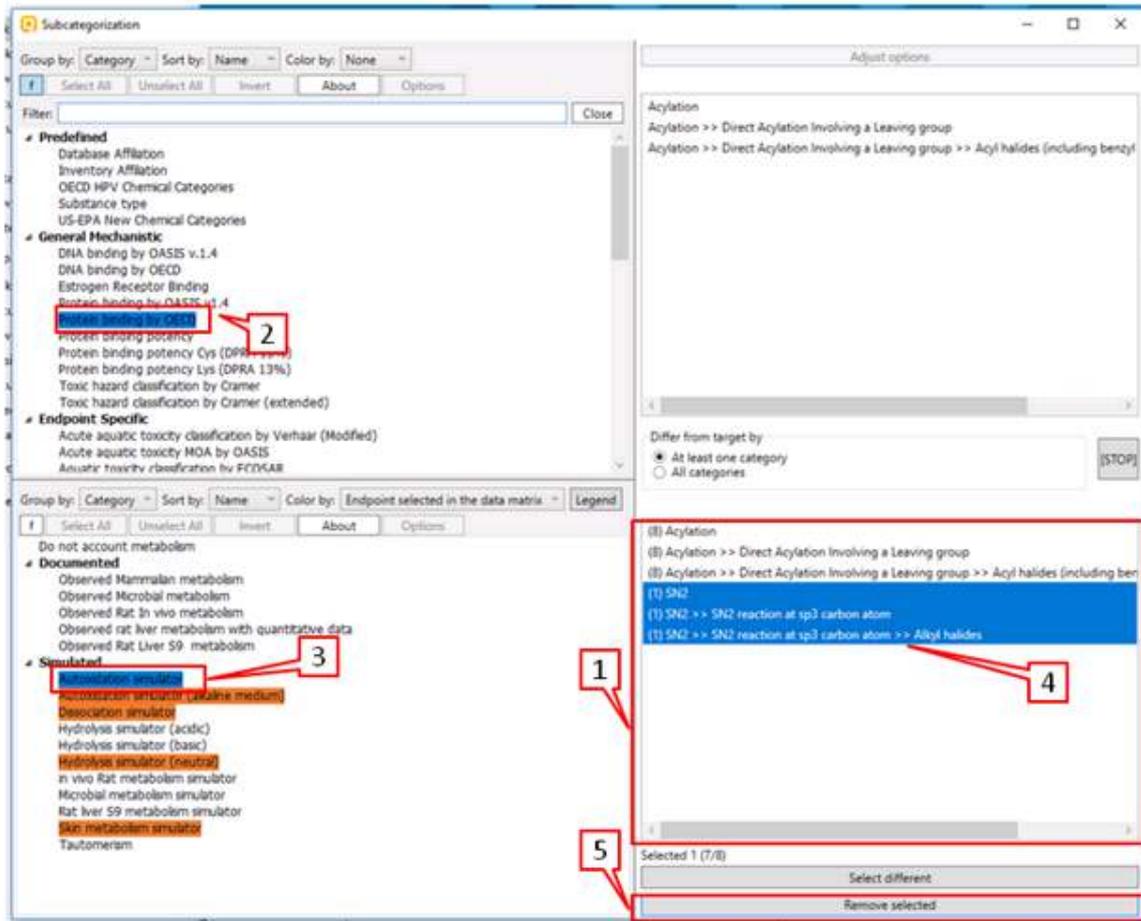


Figure 11

Right click over a category invokes the pop-up menu with some secondary functions (1) (Figure 12)

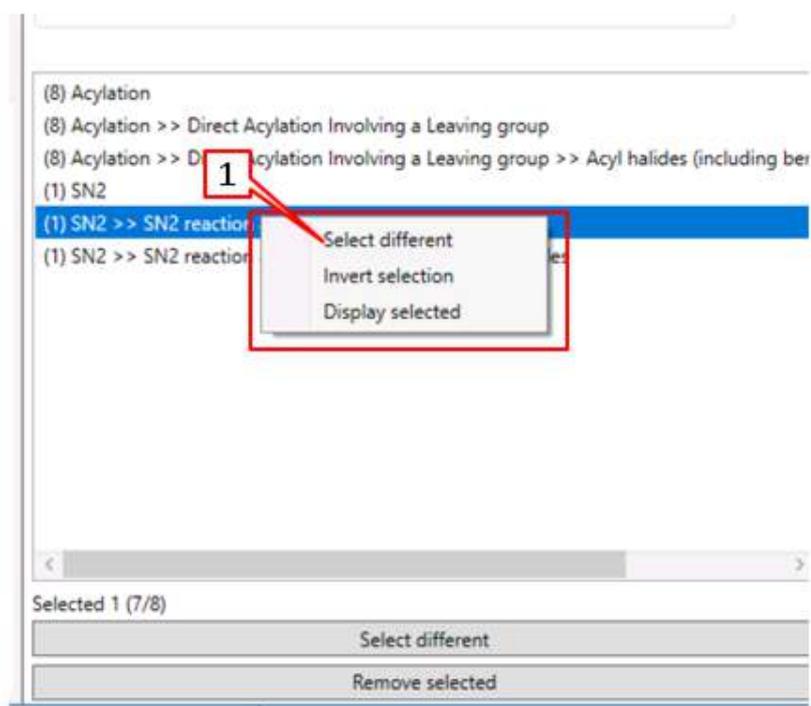


Figure 12

Display selected will show a new window containing all selected analogues (Figure 13):



Figure 13

Select different option. There are two options to differentiate analogues from the target chemical regarding profiling results (Figure 14):

- o Analogues having at least one profiling category different from those of the target (1)
- o Analogues having all categories simultaneously different from those of the target (2)

The first option means that the software will search for analogues having at least one of the (profiling results) categories associated to target chemical.

The second option means that software will search for analogues having all categories (profiling results) simultaneously.

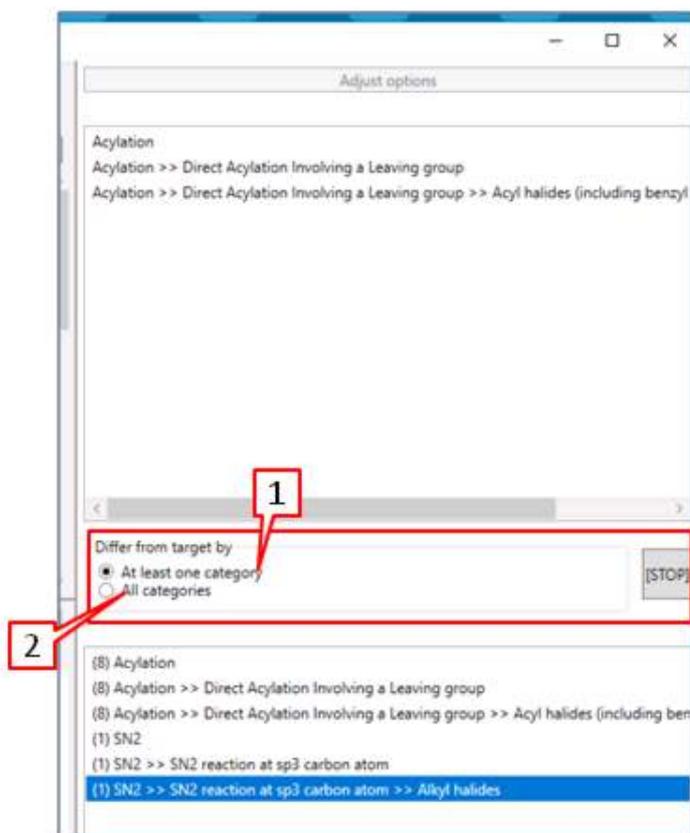


Figure 14

- *Adjust options*

Adjust options button is used for changing settings of the available groupers, for instance the Structure similarity. The structure similarity was developed for identifying chemicals (analogues) based on different levels of structural similarity between the target and analogue chemicals. When the user selects Structure similarity (1) module the software distributes all analogues in bins (2) of similarity (Figure 15).

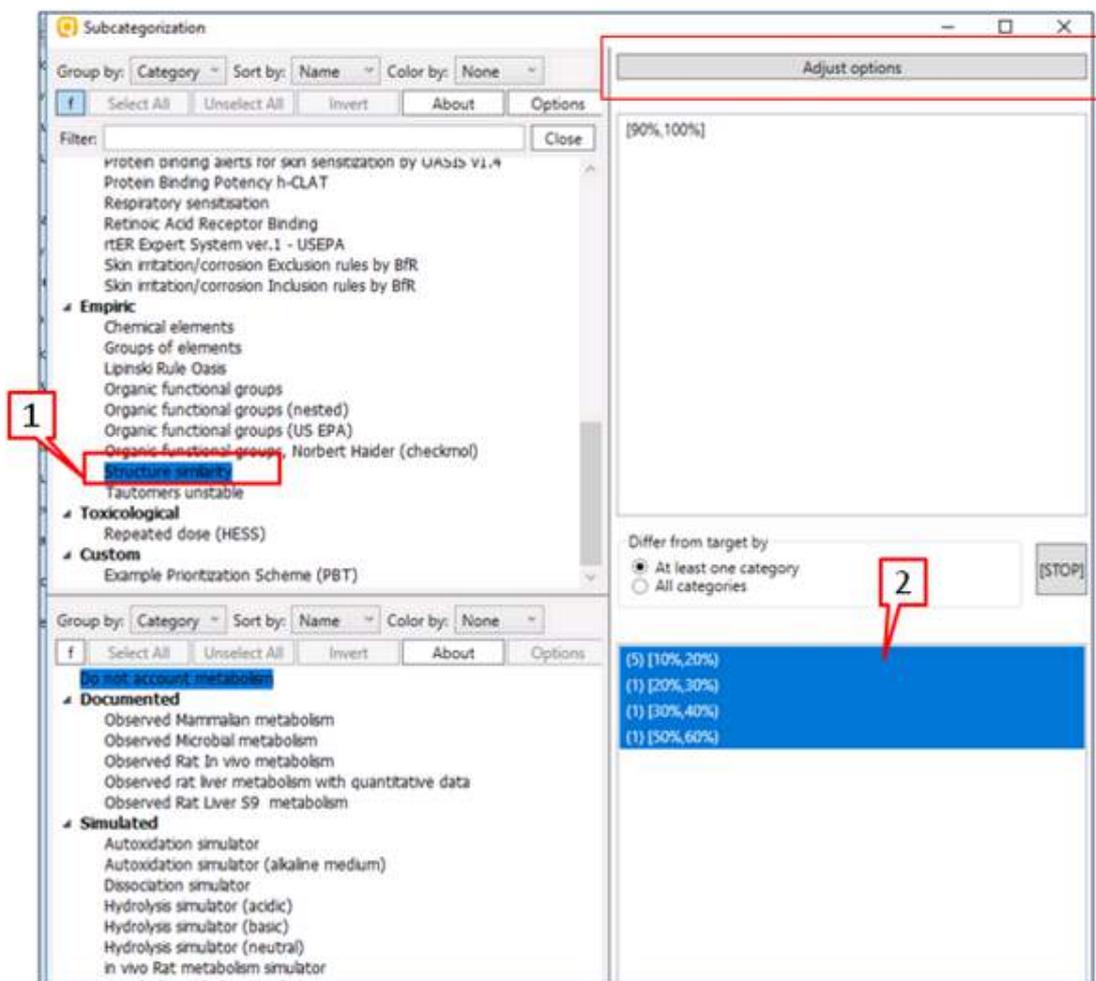


Figure 15

The user could see, which chemicals are displayed in each of the bins by selecting the bin (1), then right-click on it to display the pop-up window(2) (Figure 16). Selection of *Display selected* opens a window containing the structures of the chemicals that fall into that bin. Select the structure (3), then right click on it opens a pop-up window with *Explain* option (4), which when selected opens a new window Profiling results.

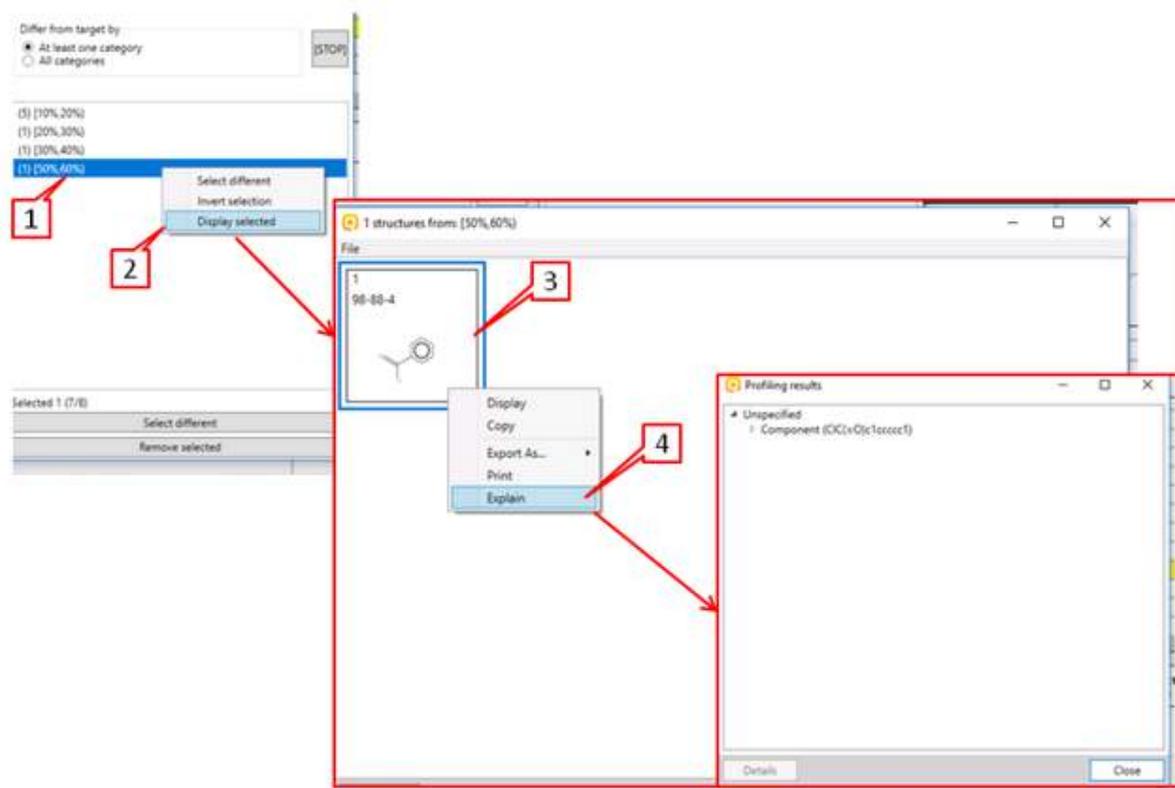


Figure 16

Expand the component level (1), select the similarity range (2) then click on *Details* (3) which opens Similarity form window (4) (Figure 17).

The screenshot displays the 'Profiling results' window with a tree view showing the following structure:

- Unspecified
 - Component (C1C(=O)c1ccccc1)
 - Chemical profile
 - Structure similarity (50%, 60%)

The 'Similarity form' dialog box is titled 'Structure similarity: 57.14%' and contains the following information:

Target chemical	Query chemical
Atom centered fragments (6/12) N(sp3)O[*] (0/1) C(sp3)N(sp3)[*](O)=O (0/1) N(sp3)=O[*] (0/1) CHC(sp3)[*](CH)N(sp3) (0/1) C(sp3)CH[*]CH (2/4) C(sp3)C(sp3)[*](CH)CH (1/1) C(sp3)C(sp3)[*](Cl)=O (1/1) C(sp3)Cl[*] (1/1) C(sp3)=O[*] (1/1)	Atom centered fragments (6/9) C(sp3)Cl[*] (1/1) C(sp3)C(sp3)[*](Cl)=O (1/1) C(sp3)=O[*] (1/1) C(sp3)C(sp3)[*](CH)CH (1/1) C(sp3)CH[*]CH (2/2) CHCH[*]CH (0/3)

Numbered callouts in the image indicate: 1) The similarity percentage (57.14%); 2) The chemical structures of the target and query; 3) The 'Details' tab; 4) The atom-centered fragments lists.

Figure 17

On the top of the window is the percentage of similarity between structures (1). Molecular structure of the target and compared structure is next (2). Features (fragments) (3) used in determining of similarity between two chemical are displayed. Green color (4) is an indication for common feature for the two compared molecules, while the red is an indication for difference (5). (Figure 18)

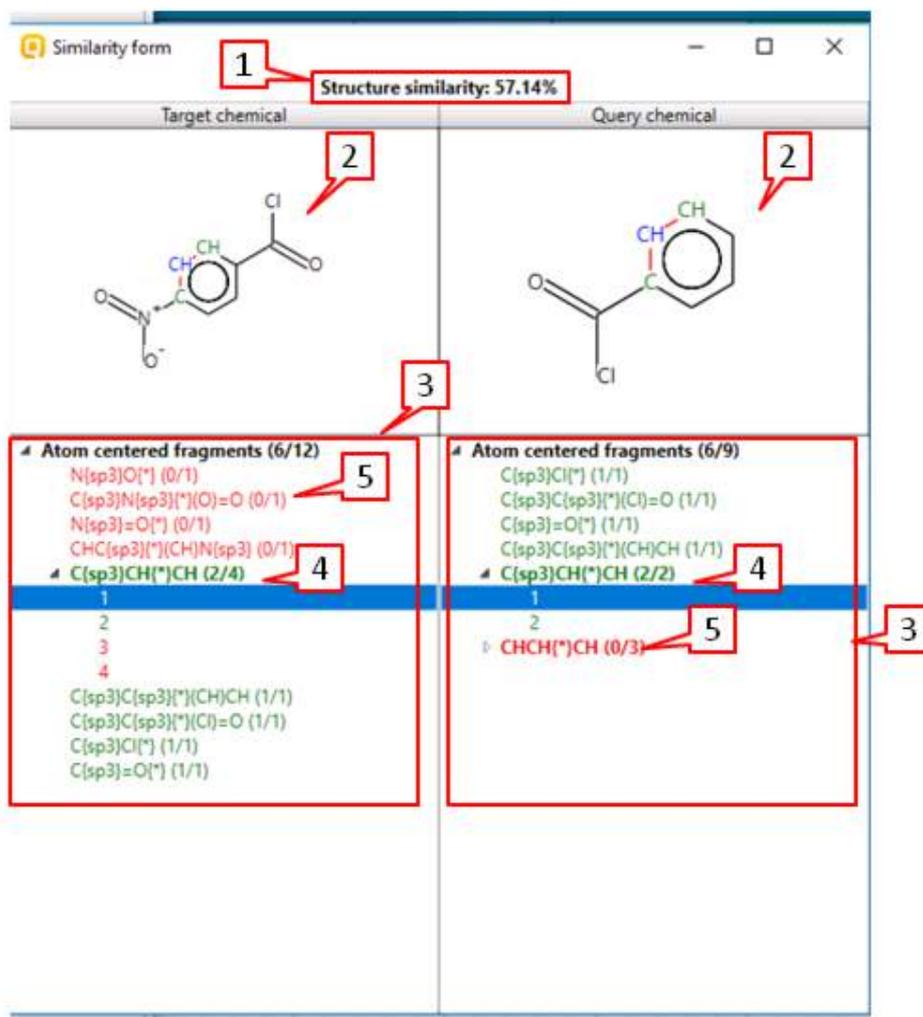


Figure 18

Settings of similarity

A list of settings for similarity calculation is available with a click on *Adjust option* button (1) (Figure 19). The following *Similarity options* are included:

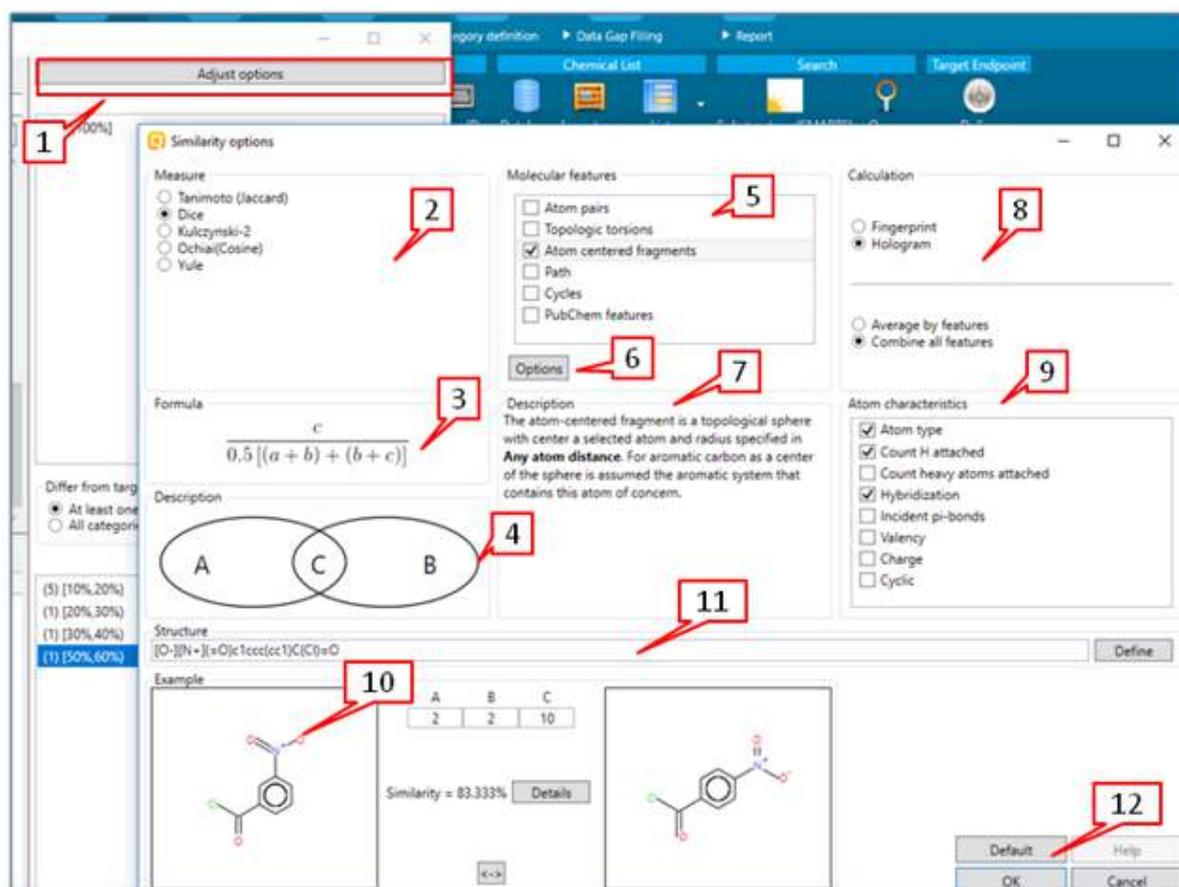


Figure 19

- 2 - Methods for calculation of similarity
- 3 - Panel displaying formula for calculation of similarity
- 4 - Graphical description of overlapping of features of similarity
- 5 - Molecular features used in calculation
- 6 - Options related to corresponding molecular feature
- 7 - Short description of selected molecular feature
- 8 - Calculation options
- 9 - Atom characteristics used in calculation of similarity
- 10 - Example illustrated selected method and features of similarity
- 11 - Field for entering SMILES

- 12 - Set default options

-0-

QSAR Toolbox v.4

D.5.4.2.1.2. Mark chemical by WS

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Select/filter data](#) / D.5.4.2.1.2. Mark chemical by WS

D.5.4.2.1.2. Mark chemical by WS

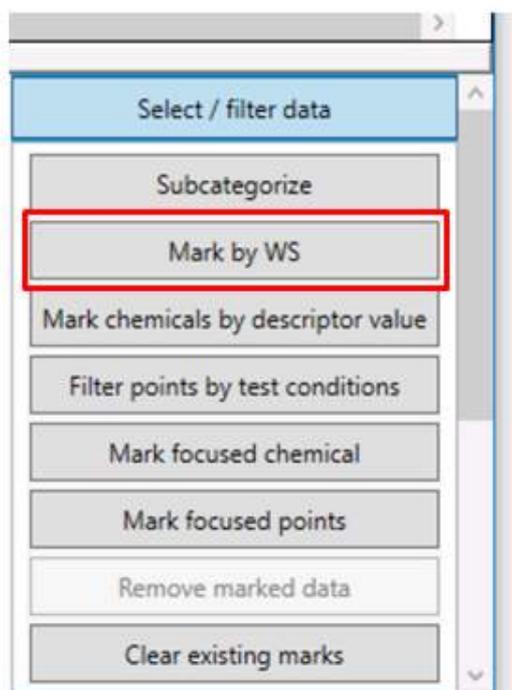


Figure 1

This function compares data values, which are in the volume concentration unit family, against their calculated maximum water solubility in order to detect experimental errors. When this option is selected, a pop-up window is displayed (Figure 2):

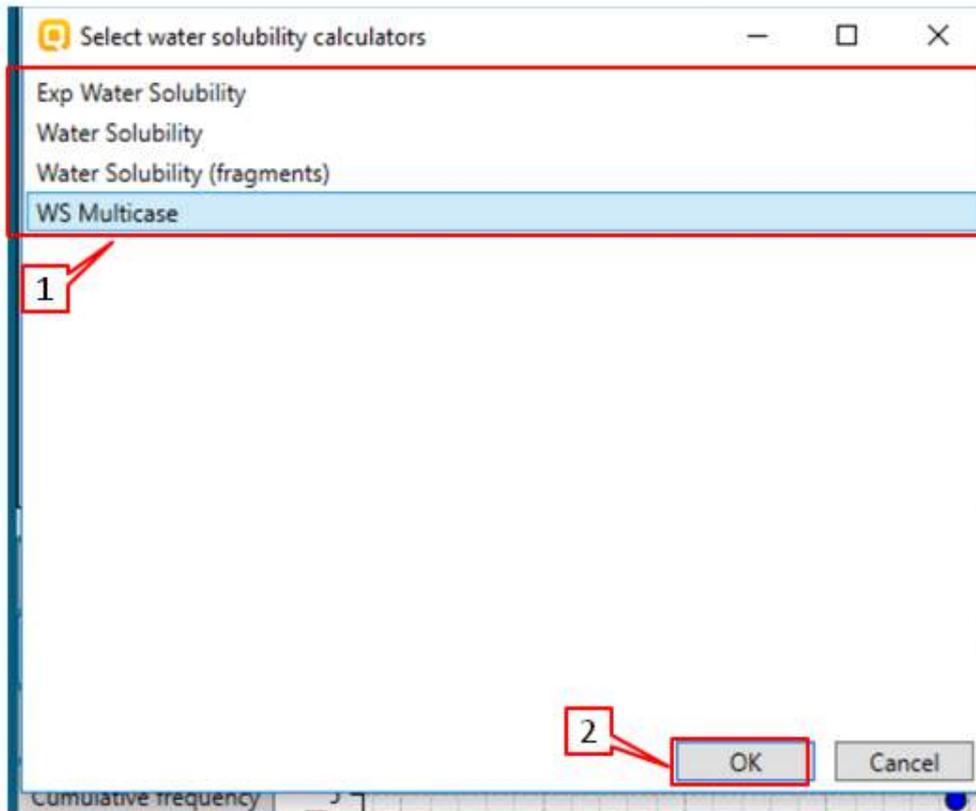


Figure 2

The user has to select one of the available methods (1) and then click OK (2) (Figure 3). The chemicals are then selected (colored in green) (1) (Figure 4).

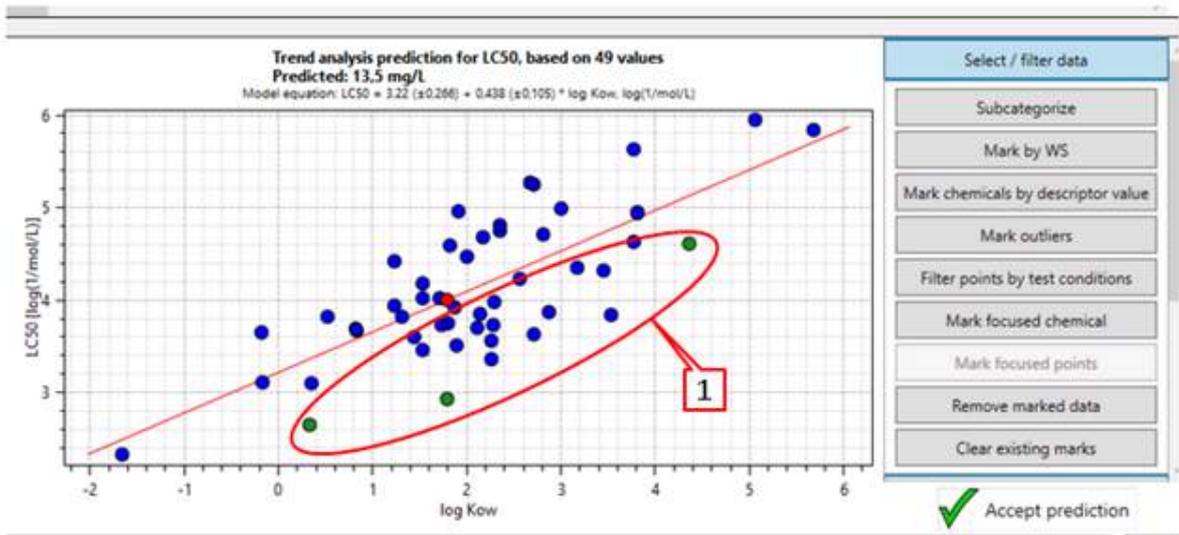


Figure 4

QSAR Toolbox v.4

D.5.4.2.1.3. Mark chemicals by descriptor values

[Workflow / Filling data gaps / Common features in three gap filling approaches/ Menus of data gap filling / Select/filter data/ D.5.4.2.1.3. Mark chemicals by descriptor values](#)

D.5.4.2.1.3. Mark chemical by descriptor values

This function is used to select chemicals that fall within a range of a particular parameter. The user has to click on *Mark chemicals by descriptor value* (1) select one of the *available descriptors* (2), define the desired *range* (3) and then click *OK* (3). (Figure 1)

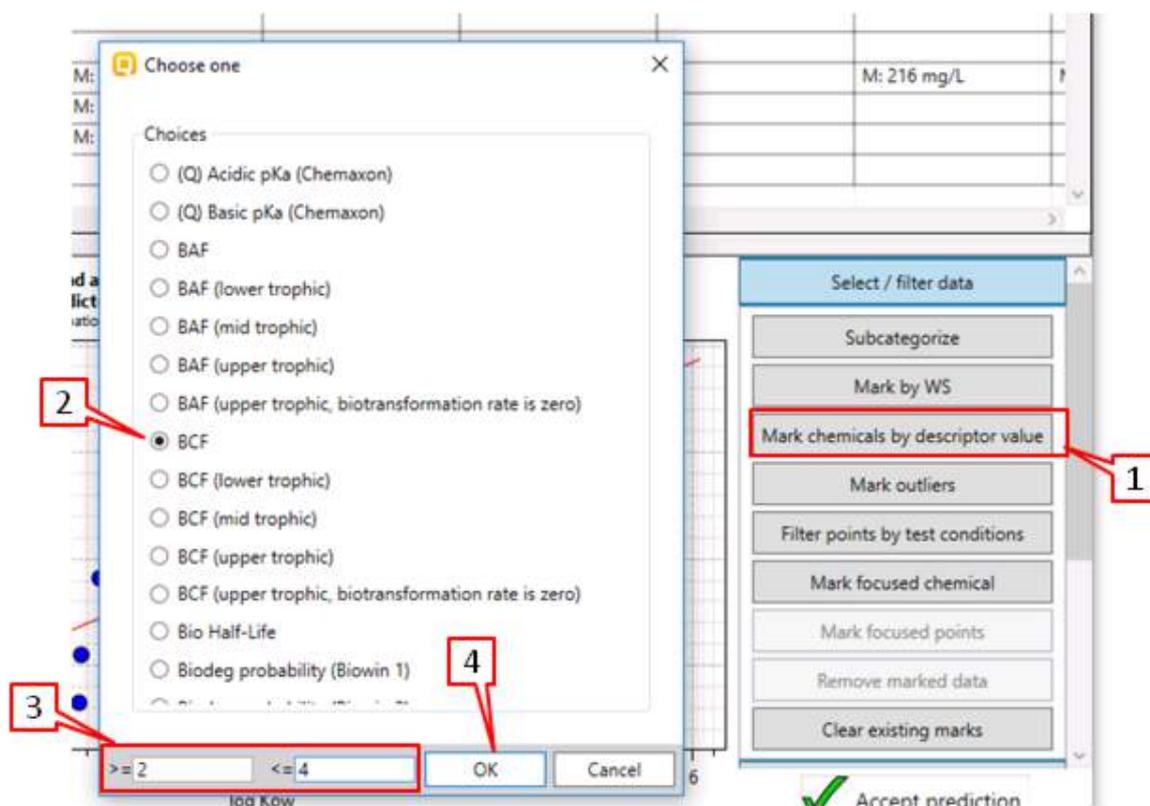


Figure 1

Finally chemical(s) that fall in the defined range are selected (colored in green) (1) (Figure 2).

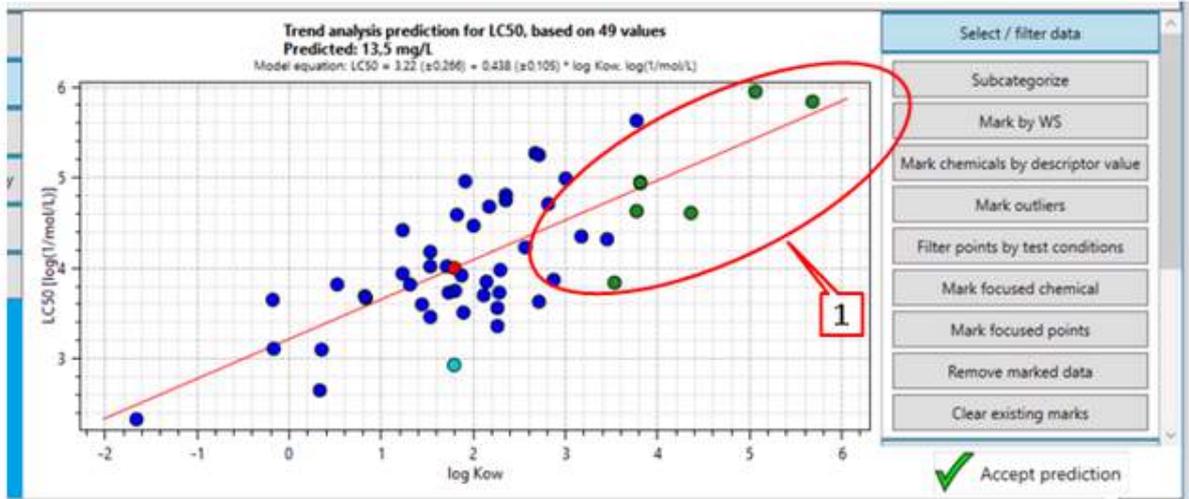


Figure 2

-O-

Note



Filter by test condition functionality removes experimental data for a given chemical.

-0-

QSAR Toolbox v.4

D.5.4.2.1.5. Mark focused chemical

[Workflow / Filling data gaps / Common features in three gap filling approaches/ Menus of data gap filling / Select/filter data/](#) D.5.4.2.1.5. Mark focused chemical

D.5.4.2.1.5. Mark focused chemical

Select a chemical from the graph, colored in pale blue (1), then selecting *Mark focused chemical* (2) marks turns the chemical green (in green) (3) (Figure 1). This is done in cases when the user wants to remove certain data ([see Removed marked chemicals/points](#))

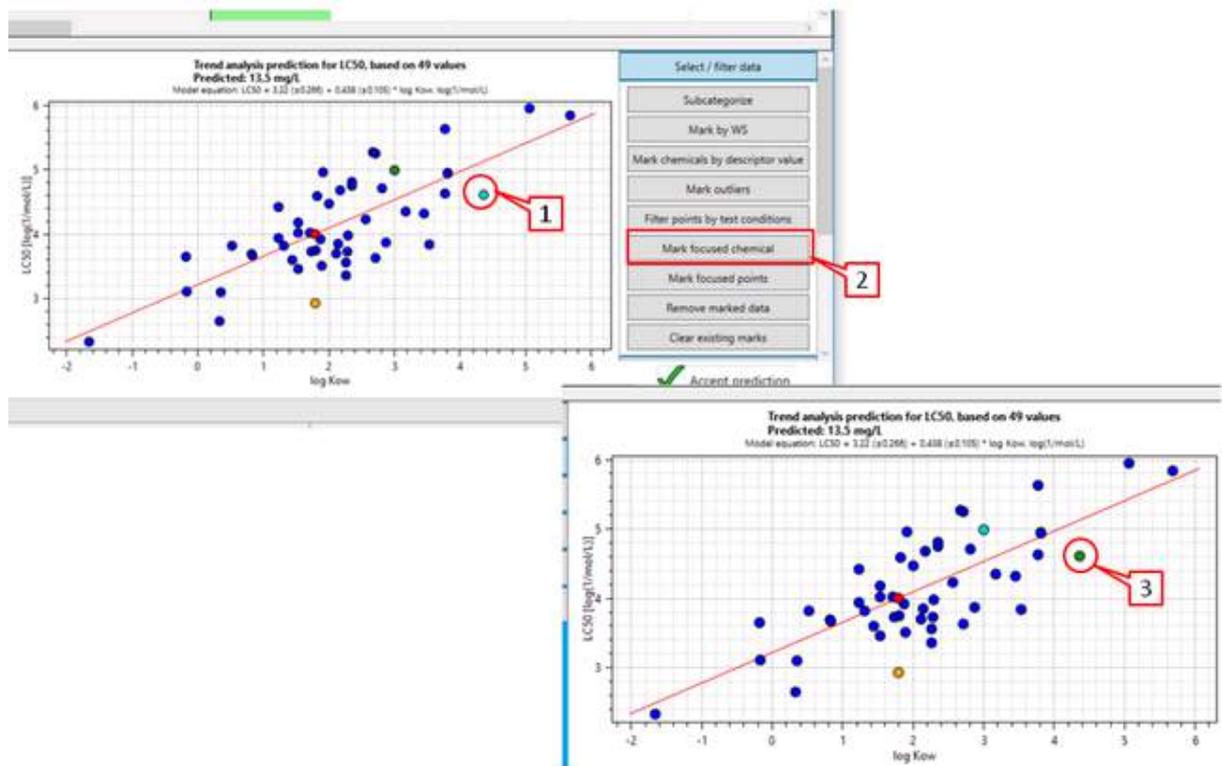


Figure 1

-0-

QSAR Toolbox v.4

D.5.4.2.1.6. Mark focused points

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Select/filter data](#) / D.5.4.2.1.6. Mark focused points

D.5.4.2.1.6. Mark focused points

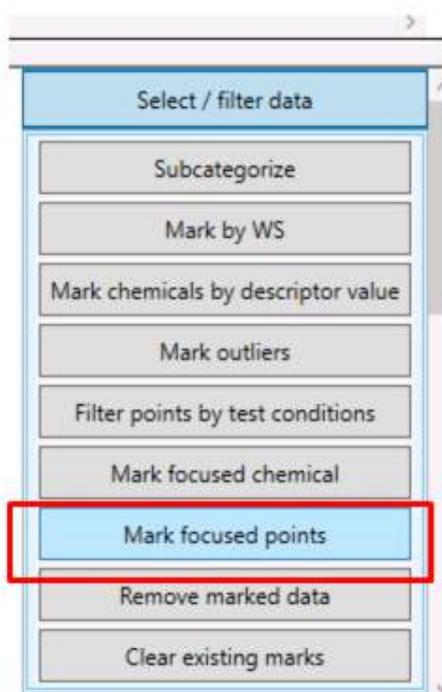


Figure 1

In order to use this functionality the user has to first select a few points in the graph. This is done by using *Region selection tool (2)* (in *Information* sub-level) (1) (Figure 2). This allows selection of a region of points in the graph (3).

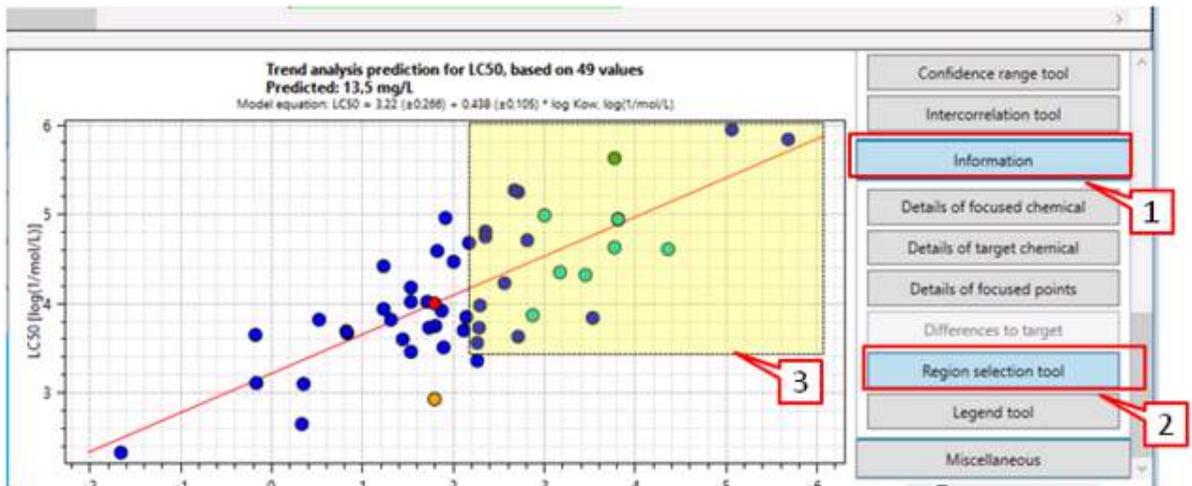


Figure 2

The selected points are colored in pale blue (1). (Figure 3)

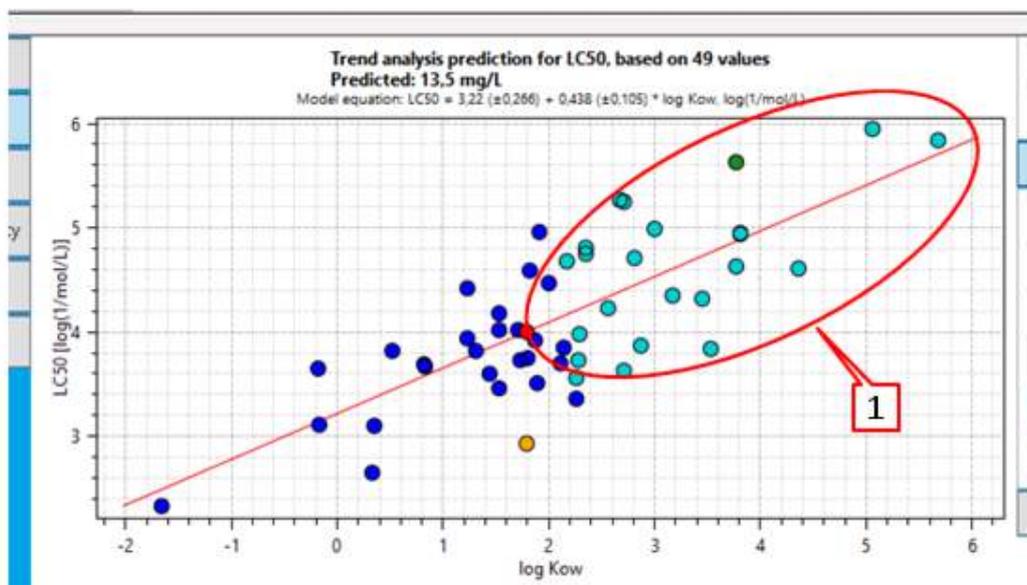


Figure 3

Once the points are focused, press Mark focused points (1). This functionality marks (in green) the currently focused data point(s) (2) (Figure 4).

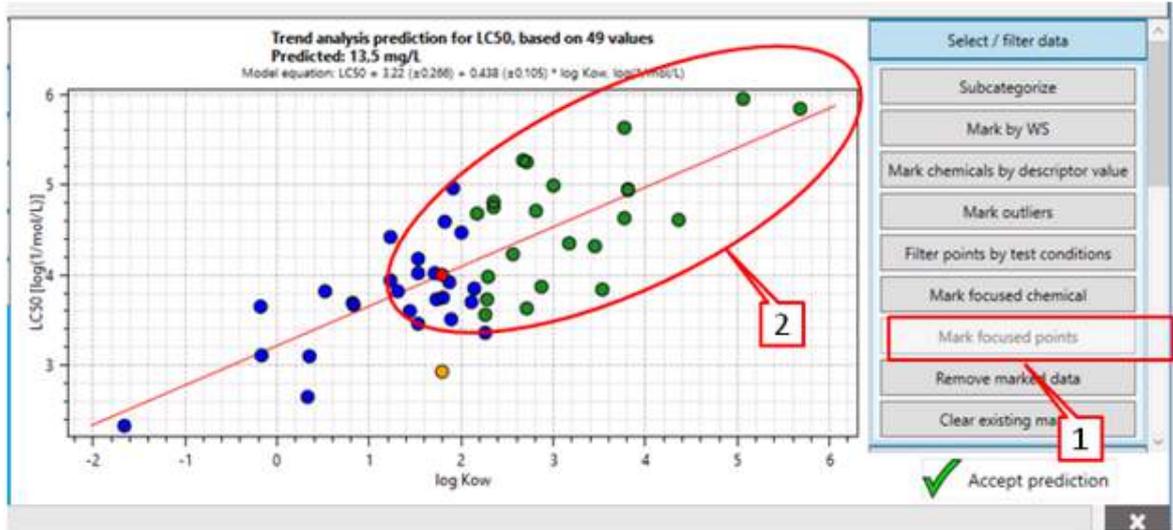


Figure 4

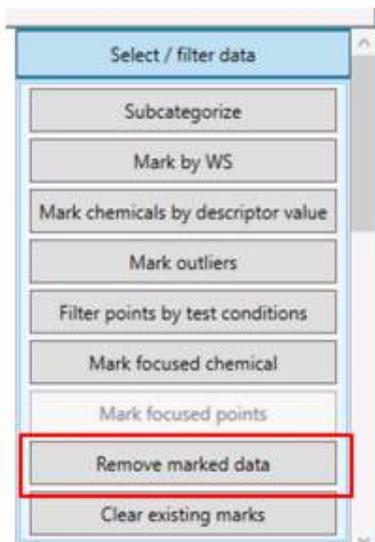
-O-

QSAR Toolbox v.4

D.5.4.2.1.7. Remove marked chemicals/points

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Select/filter data](#) / D.5.4.2.1.7. Remove marked chemicals/points

D.5.4.2.1.7. Remove marked chemical/points



This function (1) removes marked (colored in green.) chemical/data-points (2) (Figure 1)

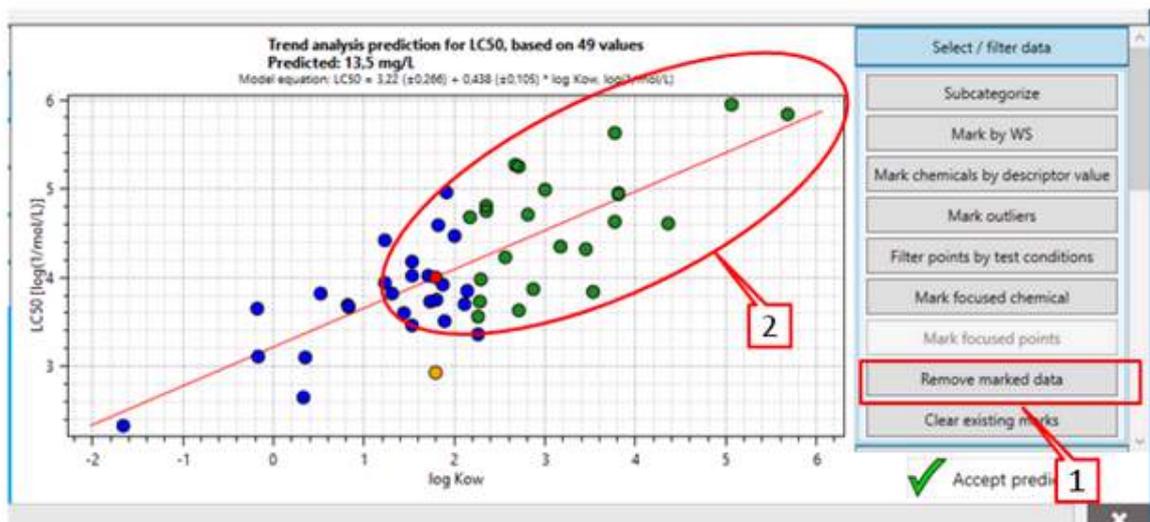


Figure 1

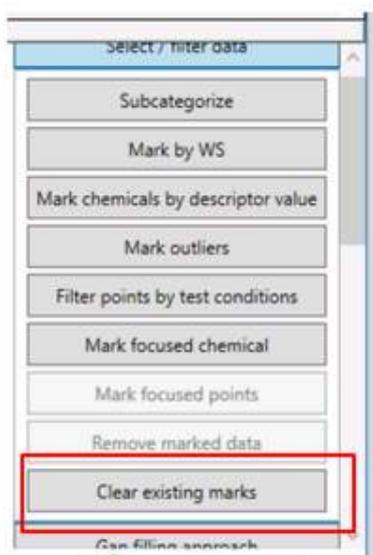
-0-

QSAR Toolbox v.4

D.5.4.2.1.8. Clear existing marks

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Select/filter data](#) / D.5.4.2.1.8. Clear existing marks

D.5.4.2.1.8. Clear existing marks



This function clears the markings of chemicals/data-points.

-0-

D.5.4.2.1.9. Mark outliers

D.5.4.2.1.9. Mark outliers

This function is used to select chemicals that are outliers. The user has to click *Mark outlier* (1) (Figure 1). If there are no outliers, a message will be displayed (2), otherwise the outliers will be coloured in green on the graph.

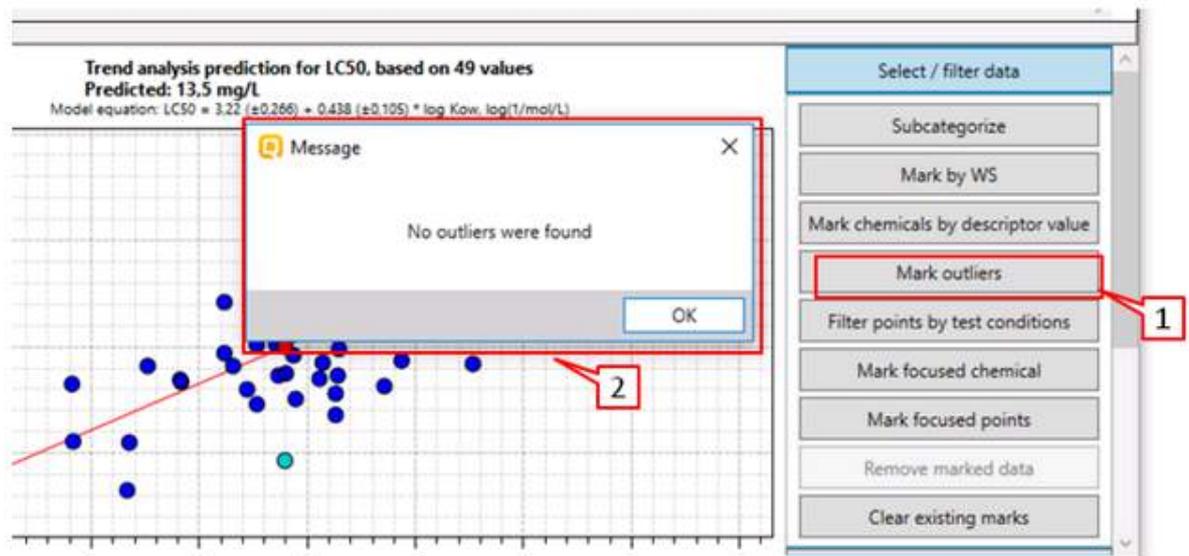


Figure 1

QSAR Toolbox v.4

D.5.4.2.2. Gap filling approaches

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.2. Gap filling approaches

D.5.4.2.2. Data Gap filling approaches

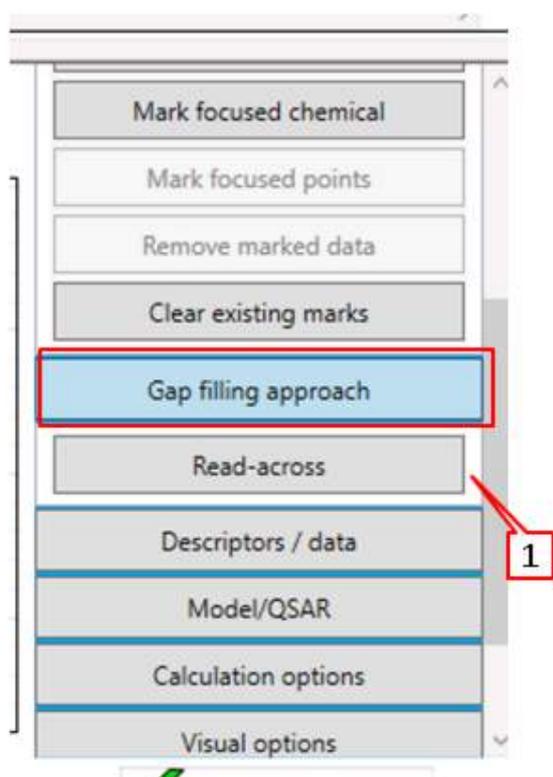


Figure 1

This functionality allows the user to switch between gap filling modules (1). (Figure 1)

- Read-across – switch to read-across.
- Trend analysis – switch to trend analysis.

-0-

QSAR Toolbox v.4

D.5.4.2.3. Descriptors/data

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.3. Descriptors/data

D.5.4.2.3. Descriptors/data

Typical methods for data gap filling explore the correlation/dependency between experimental data for the endpoint of interest and some 2D or 3D parameters. From modelling perspective, the endpoint of interest is the dependent variable while 2D/3D parameters are independent variables. However it is possible experimental data from other endpoints to be used as independent variables in order to investigate and employ correlations between different endpoints. This is done in *Select endpoint tree descriptor* option. (Figure 1)

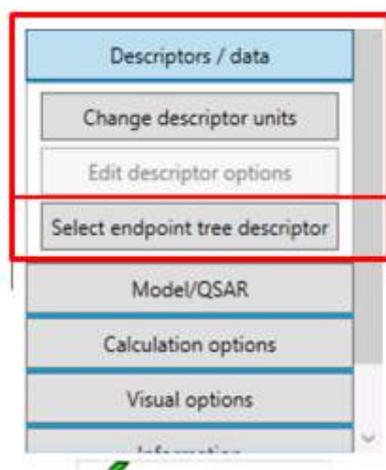


Figure 1

Select endpoint tree descriptor allows comparing of different experimental data for the analogues used in the data gap filling method/ After selecting this feature the system iterates through all data collected on the data matrix and builds an endpoint tree (which is a subset of the original endpoint tree) (1), which contains only the nodes with experimental data (Figure 2). Expand the tree and select the endpoint of interest (1). Then press *OK* (2)

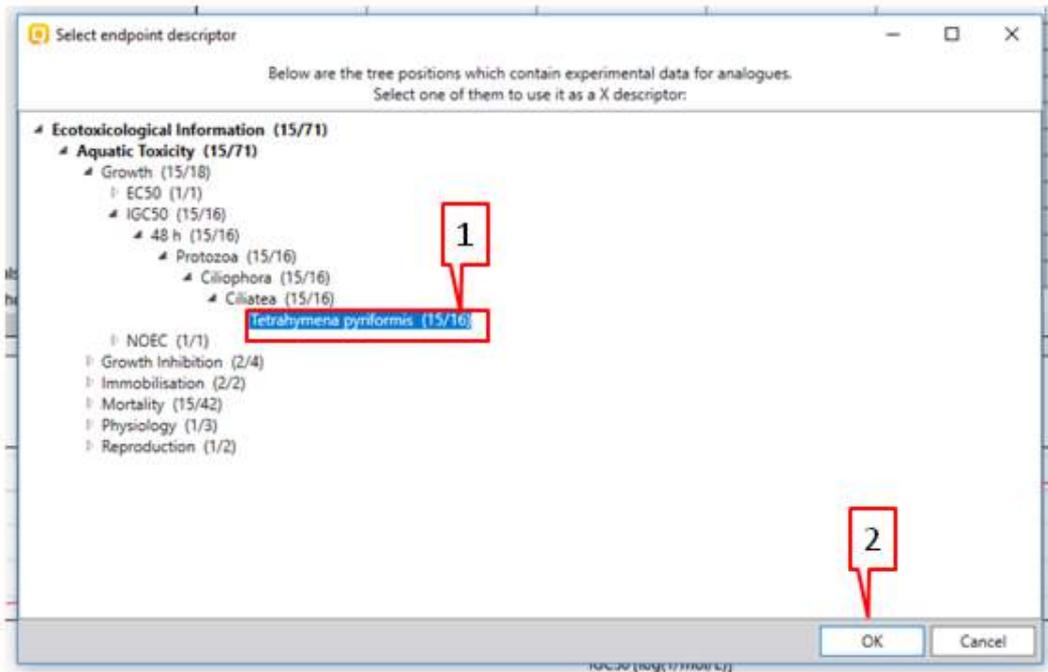


Figure 2

Possible data inconsistency window could be displayed. (Figure 3). Select the scale/ unit and press OK (1).

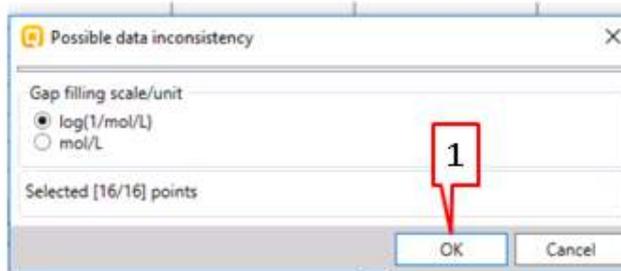


Figure 3

The correlation between the two endpoints is shown on the graph (Figure 4)

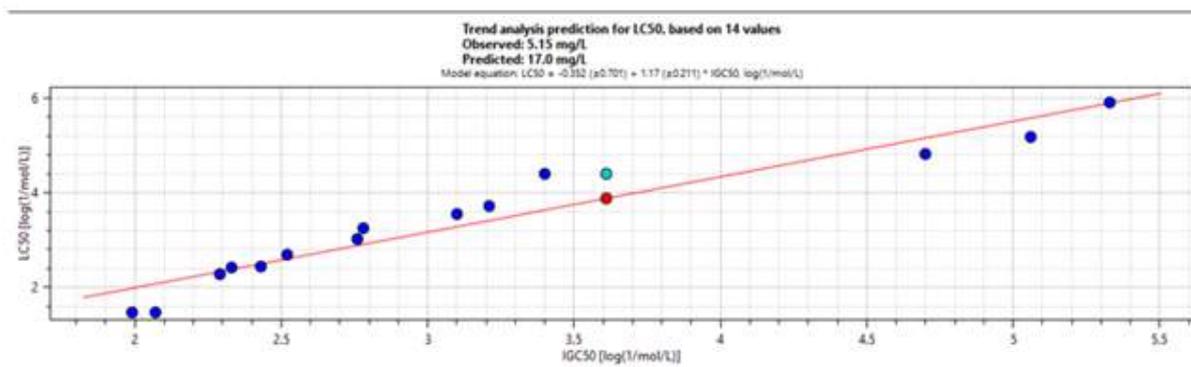


Figure 4

NOTE

The endpoint tree node which is selected in above manner will appear as parameter in the parameters list of the gap filling module and main options form. There is a specific distinction in the way the endpoint tree parameters work compared to the standard (2D/3D) parameters. While the standard parameters return the same value for a given chemical, the endpoint tree parameters return a value based on the experimental data currently available on the data matrix. As at any given moment different databases might be selected, the available experimental data for chemicals might also vary. For example, if the user applies a QSAR based on endpoint tree parameters, the prediction for a given chemical might vary depending on the current selection of databases at the moment of prediction.

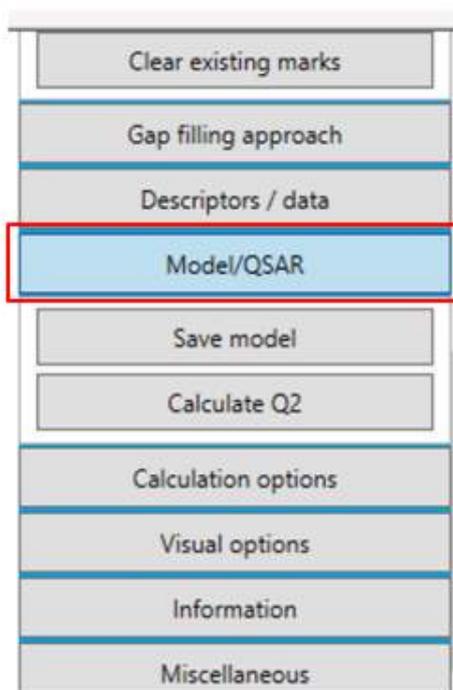
-0-

QSAR Toolbox v.4

D.5.4.2.4. Model/(Q)SAR

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.4. Model/(Q)SAR

D.6.4.3.5. Model/(Q)SAR



This functionality allows to:

- Save model – save the model. A form will be displayed in which the user should fill in all information pertinent to the model (Figure 1).

Customize model content

Wizard pages

- (Q)SAR
- General information
- Endpoint
- Algorithm
- Domain
- Training set
- Test set
- Mechanistic interpretation
- Miscellaneous

Date of QMRF: Automatically populated by the system

QMRF authors and contact details

QMRF update(s)

Model developer(s) and contact details

Date of model development and/or publication

Reference(s) to main scientific papers and/or software

Another QMRF with the same

Back Next Cancel Save model

Figure 1

The user has to select an information section (1) and then to fill in the necessary information (2). Once all the information is filled in, the user has to select Save (3) (Figure 2).

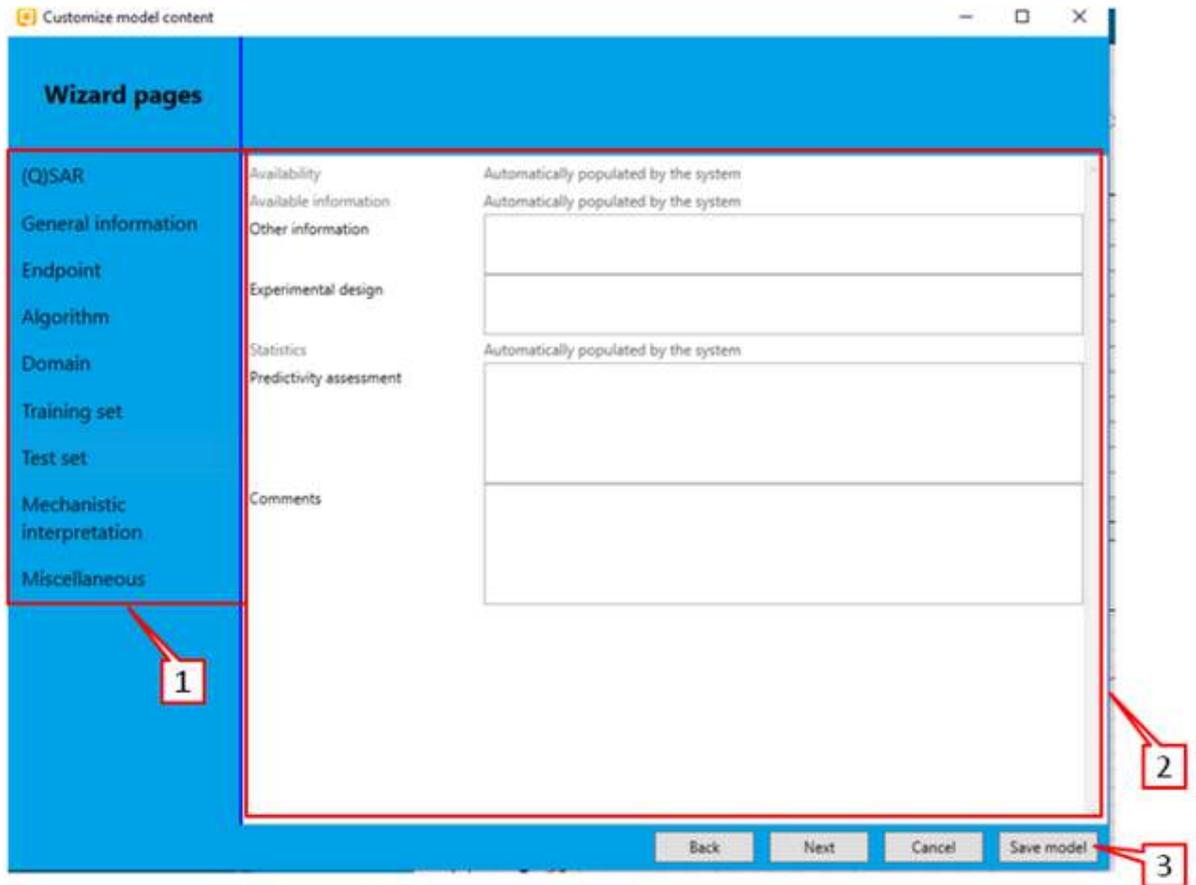


Figure 2

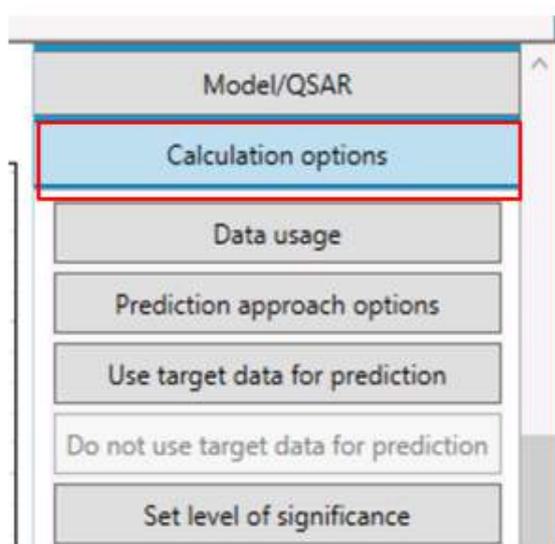
-0-

QSAR Toolbox v.4

D.5.4.2.5. Calculation options

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.5. Calculation options

D.5.4.2.5. Calculation options



- Data usage (1) – this sets the way the Toolbox handles multiple data-points per single chemical (2). (Figure 1)

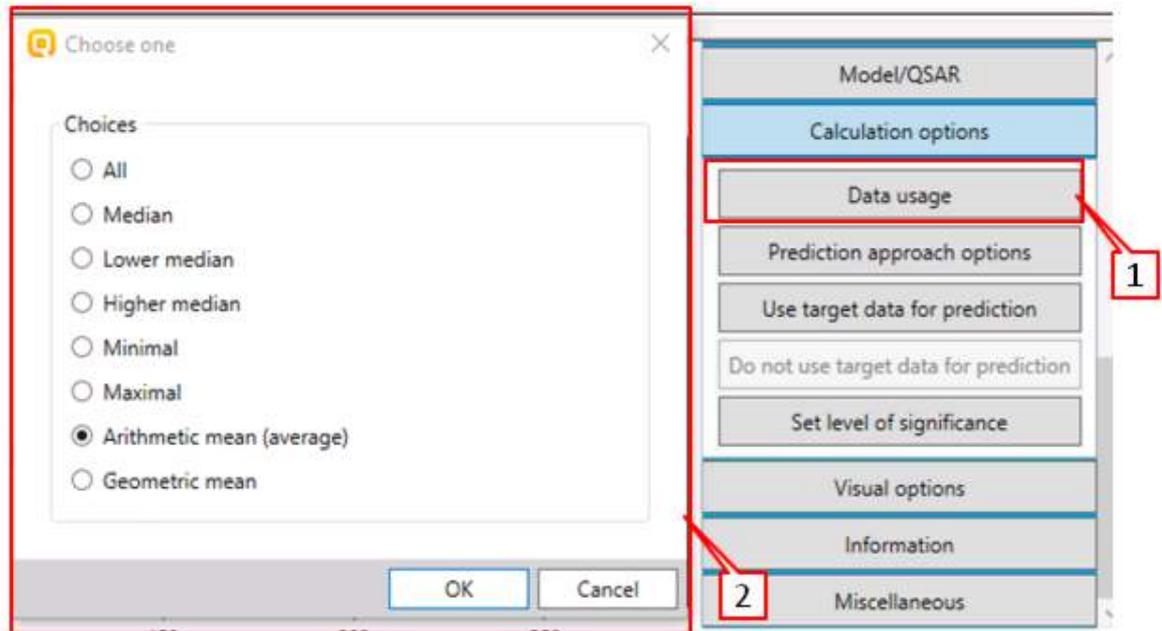


Figure 57

- Prediction approach options (1) – for read-across (2) it sets the way the prediction is approximated – minimal, maximal, average, median, lower median, higher median, mode, lowest mode and highest mode. For trend analysis (3) it sets the approximation type – averaging, linear and quadratic. (Figure 2)

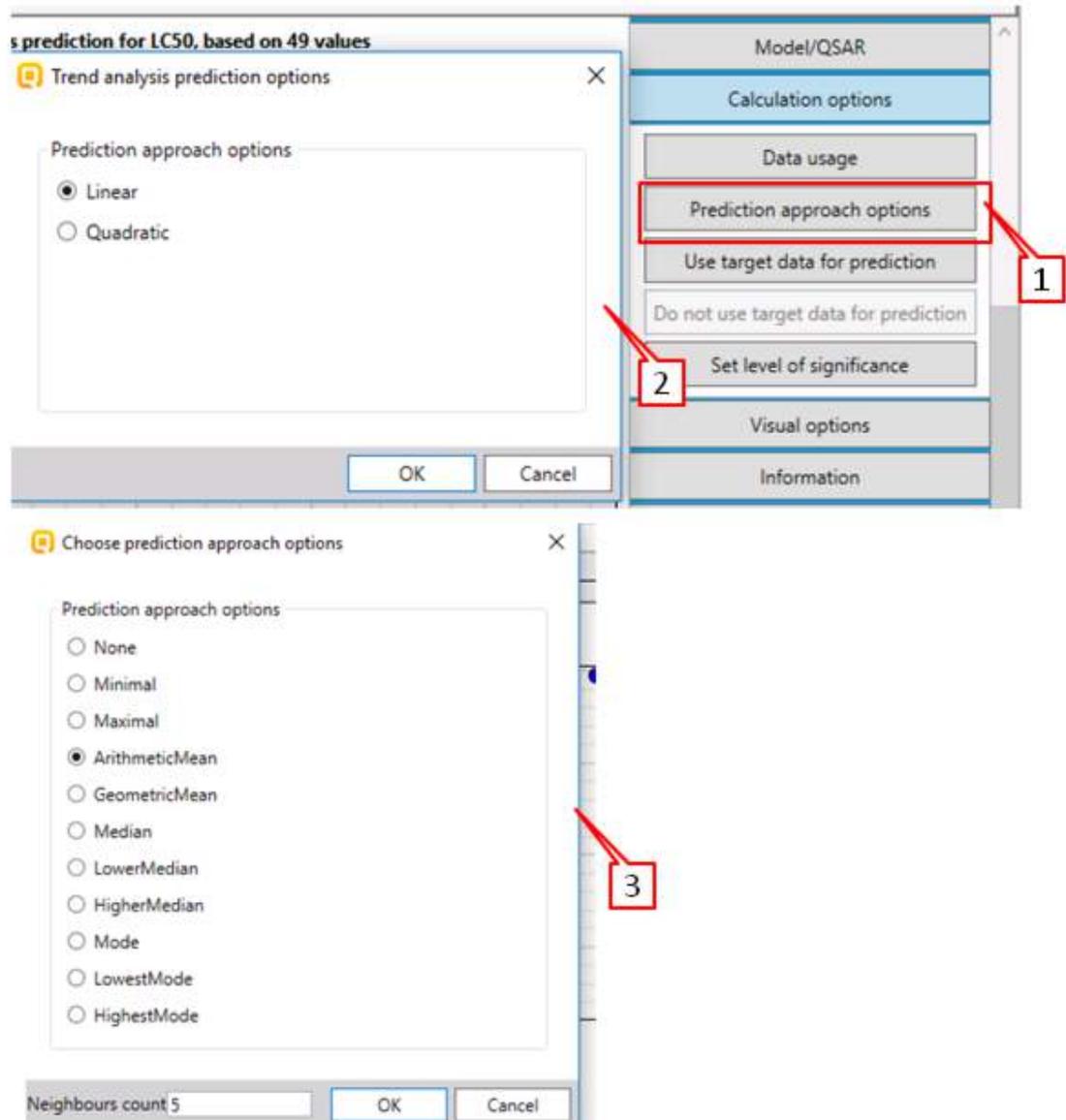


Figure 2

- Use target data for prediction - this functionality allows the user to include observed data of target if available in the gap filling calculation.
- Set level of significance – set the confidence level and standard deviation.

-0-

QSAR Toolbox v.4

D.5.4.2.5.1. Data usage options

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Calculation Options](#) / D.5.4.2.5.1. Data usage options

D.5.4.2.5.1. Data usage options

The following examples describe different approaches for usage of data and making predictions in categorical scales. They are prepared for imaginary scale containing four possible values and three chemicals having more than one observed value for the selected endpoint.

The first picture represents the real observed values for the three chemicals:

1) DataUsage=All, all points will be taken into the calculations.

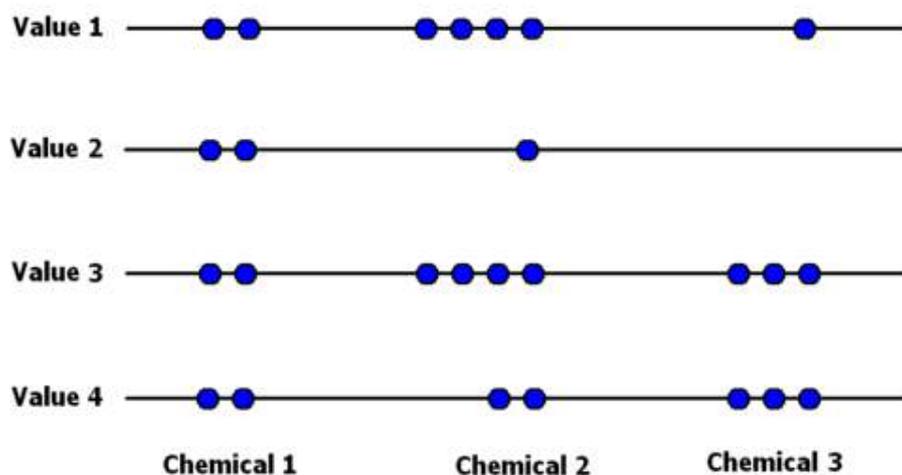


Fig. 1 The real observed values for chemicals (the same picture for using all data points)

The next pictures represent the recalculated values for the three chemicals. The real observed values are blank and the recalculated values are blue.

2) DataUsage=Minimal, one point per chemical is given.

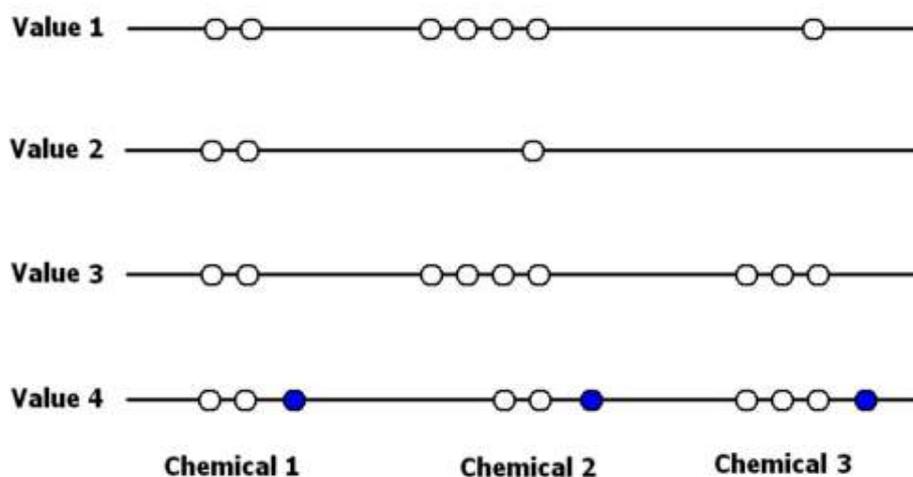


Fig. 2 The recalculated values for chemicals (when using minimal value for each chemical)

3) DataUsage=Maximal, one point per chemical is given.

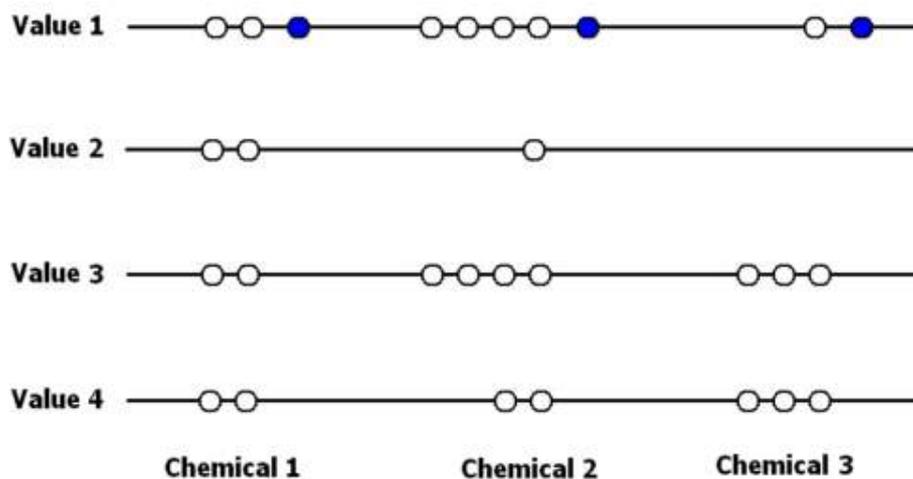


Fig. 3 The recalculated values for chemicals (when using maximal value for each chemical)

4) DataUsage=Median(s), Chemicals 1 have two medians; Chemical 2 and 3 has one median (value 2 is not taken into account here).

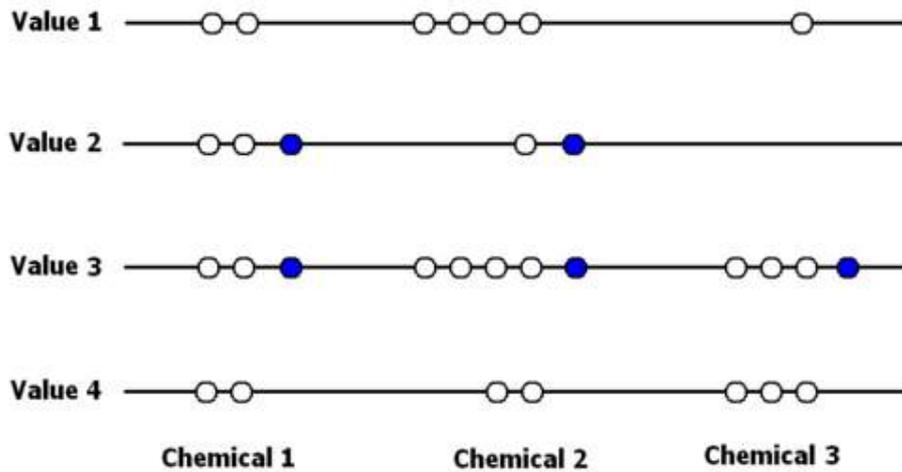


Fig. 4 The recalculated values for chemicals (when using median values for each chemical)

5) DataUsage=Lower median, one point per chemical is given.

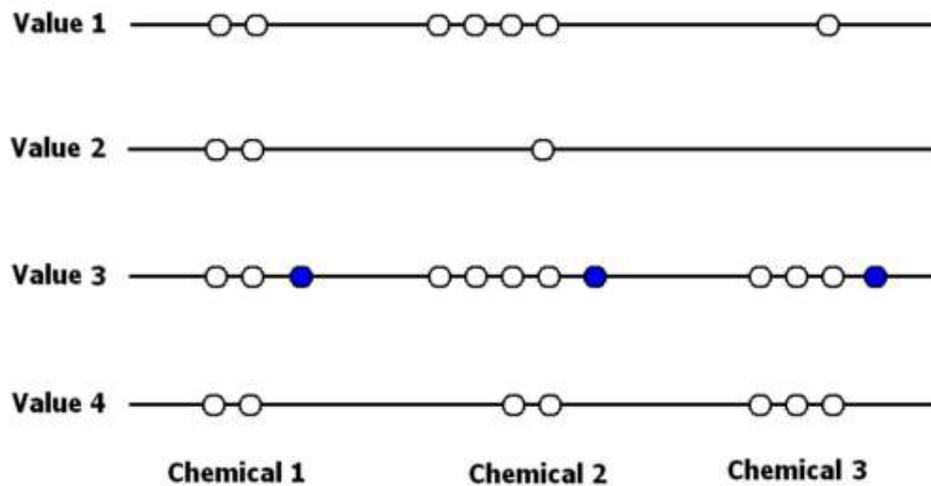


Fig. 5 The recalculated values for chemicals (when using lower median value for each chemical)

6) DataUsage=Higher median, one point per chemical is given.

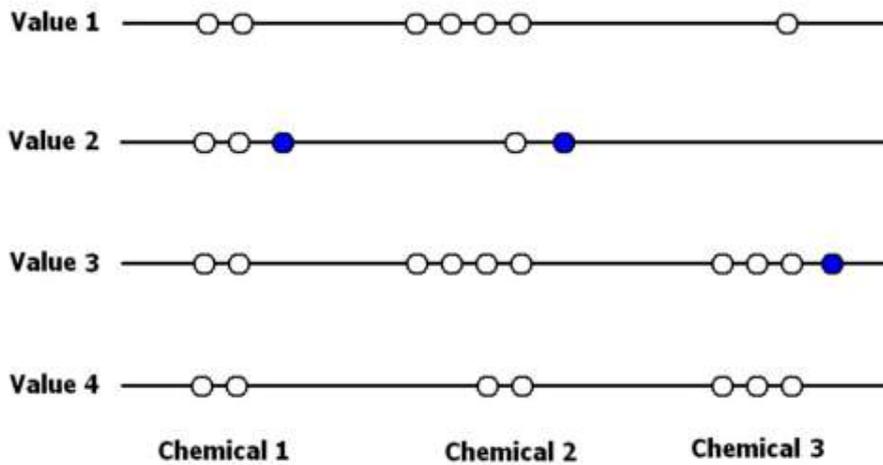


Fig. 6 The recalculated values for chemicals (when using higher median value for each chemical)

7) DataUsage=Mode(s), Chemicals 2 and 3 have two modes; Chemical 1 has four modes.

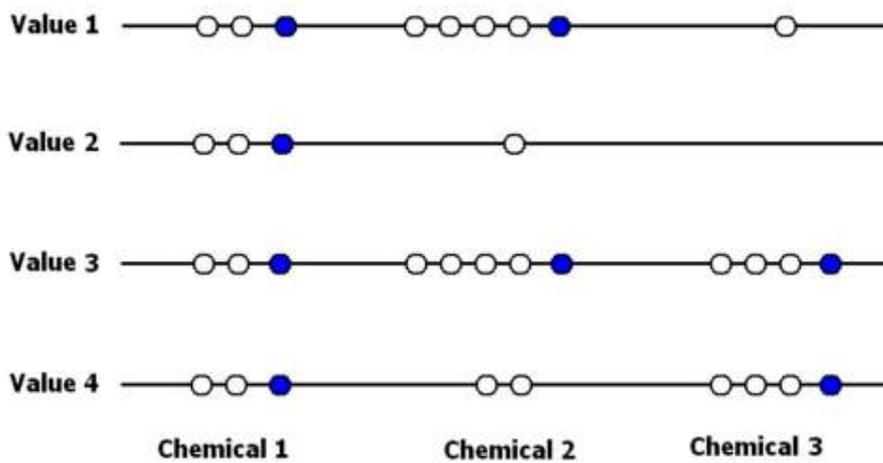


Fig. 7 The recalculated values for chemicals (when using mode values for each chemical)

8) DataUsage=Lowest mode, one point per chemical is given.

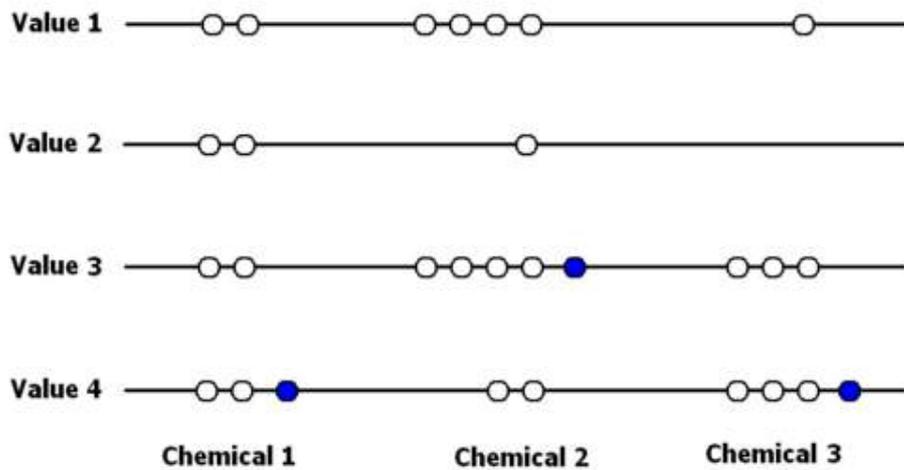


Fig. 8 The recalculated values for chemicals (when using lowest mode value for each chemical)

9) DataUsage=Highest mode, one point per chemical is given.

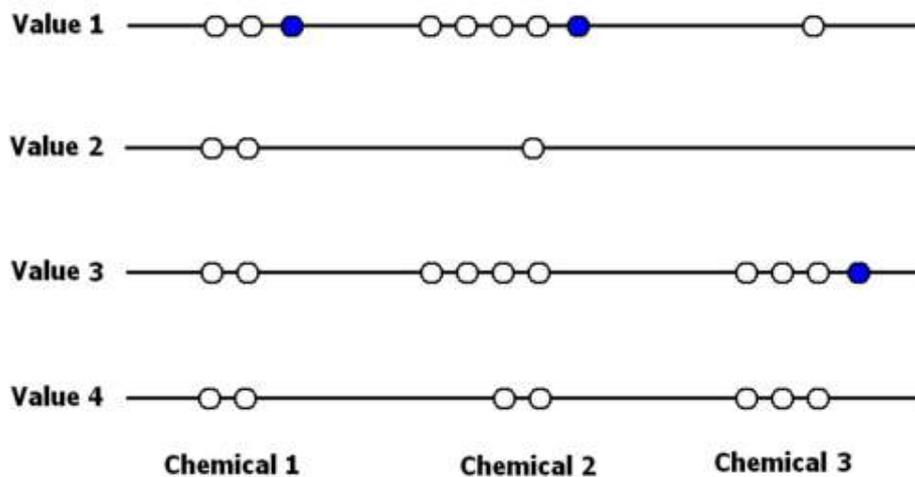


Fig. 9 The recalculated values for chemicals (when using highest mode value for each chemical)

-0-

QSAR Toolbox v.4

D.5.4.2.5.1.1. Prediction examples

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / [Calculation options](#) / [Data usage options](#) / D.5.4.2.5.1.1. Prediction examples

D.5.4.2.5.1.1. Prediction examples

Making predictions

Let's assume that Chemical 2 and 3 are the neighbors that determine the prediction. The various cases shown above will look as follows:

1) DataUsage=All, the prediction value is:

- **Value 4**, when the approximation type is "Minimal"
- **Value 1**, when the approximation type is "Maximal"
- **No value**, when the approximation type is "Median" – Value2 and Value 3 are both medians, so the system cannot make a decision automatically
- **Value 3**, when the approximation type is "Lower median"
- **Value 2**, when the approximation type is "Higher median"
- **Value 3**, when the approximation type is "Mode", "Lowest mode" or "Highest mode" – 7 neighbor points are available for this value; only one mode value is available in this case, so the last three approximation types give the same prediction value.

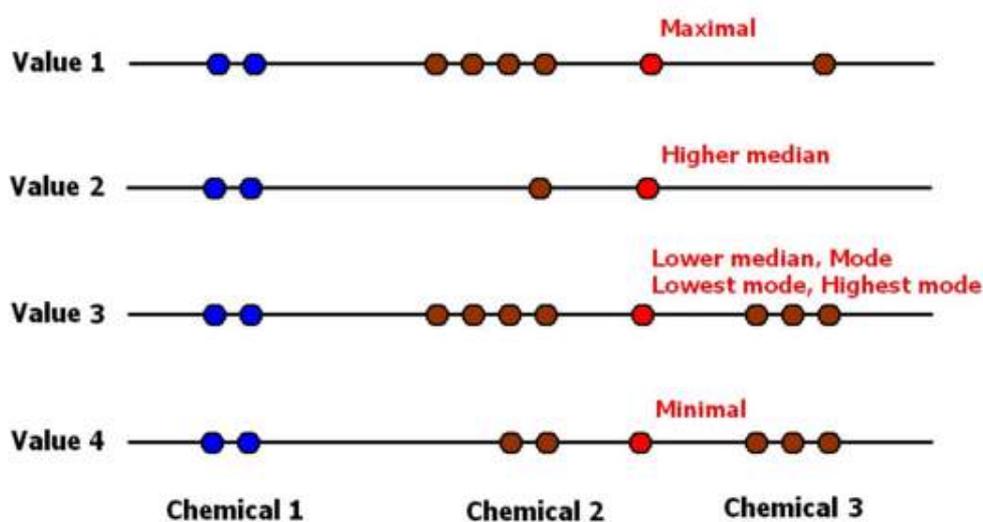


Fig. 1 The prediction values when using all data points

2) DataUsage=Minimal, the prediction value is:

- Value 4, for all approximation types.

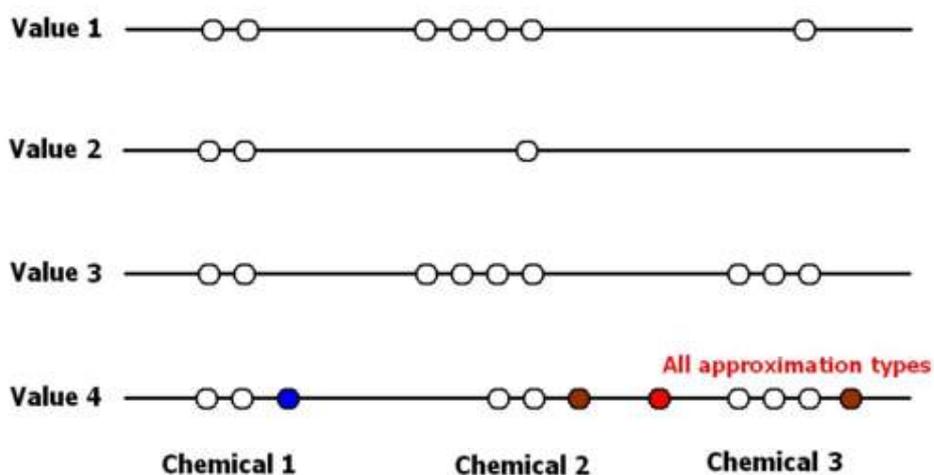


Fig. 2 The prediction value when using minimal value for each chemical

3) DataUsage=Maximal, the prediction value is:

- Value 1, for all approximation types.

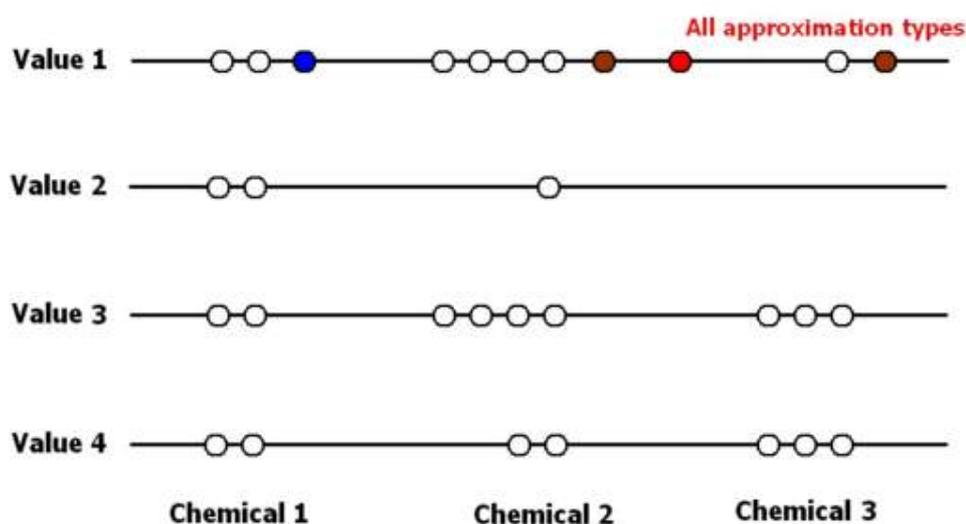


Fig. 3 The prediction value when using maximal value for each chemical

4) DataUsage=Median(s), the prediction value is:

- **Value 3**, when the approximation type is "Minimal"
- **Value 2**, when the approximation type is "Maximal"
- **No value**, when the approximation type is "Median" – Value2 and Value 3 are both medians, so the system cannot make a decision automatically
- **Value 3**, when the approximation type is "Lower median"
- **Value 2**, when the approximation type is "Higher median"
- **Value 3**, when the approximation type is "Mode", "Lowest mode" or "Highest mode" – 2 neighbor points are available for this value; only one mode value is available in this case, so the last three approximation types give the same prediction value

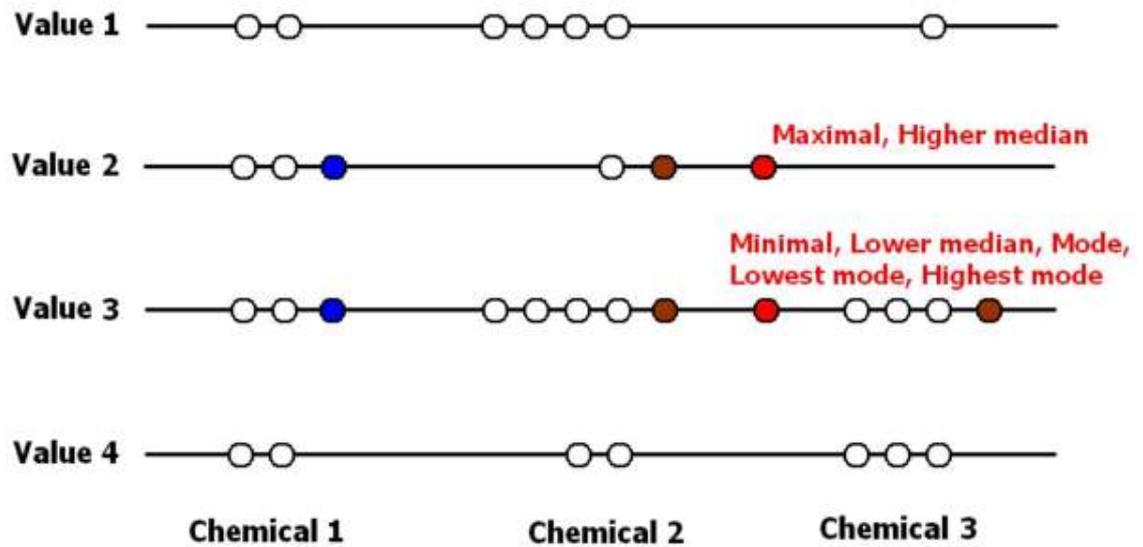


Fig. 4 The prediction values when using median values for each chemical

5) DataUsage=Lower median, the prediction value is:

Value 3, for all approximation types.

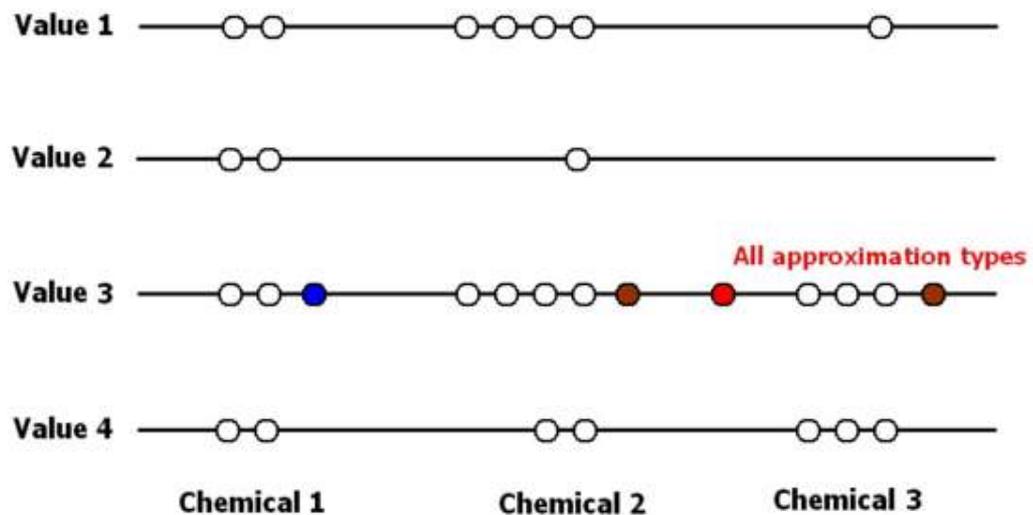


Fig. 5 The prediction value when using lower median value for each chemical

6) DataUsage=Higher median, the prediction value is:

- Value 3, when the approximation type is "Minimal"
- Value 2, when the approximation type is "Maximal"
- No value, when the approximation type is "Median" – Value2 and Value 3 are both medians, so the system cannot make a decision automatically
- Value 3, when the approximation type is "Lower median"
- Value 2, when the approximation type is "Higher median"
- No value, when the approximation type is "Mode" – Value2 and Value 3 are both modes, so the system cannot make a decision automatically
- Value 3, when the approximation type is "Lowest mode"
- Value 2, when the approximation type is "Highest mode"

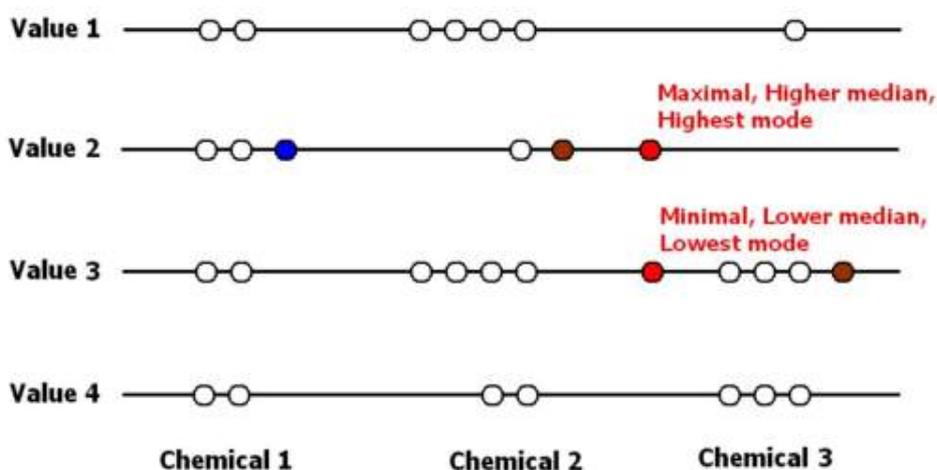


Fig. 6 The prediction value when using higher median value for each chemical

7) DataUsage=Mode(s), the prediction value is:

- Value 4, when the approximation type is "Minimal"
- Value 1, when the approximation type is "Maximal"

- **Value 3**, when the approximation type is “Median”, “Lower median” and “Higher median” – only one median value is available in this case, so these three approximation types give the same prediction value

- **Value 3**, when the approximation type is “Mode”, “Lowest mode” or “Highest mode” – 2 neighbor points are available for this value; only one mode value is available in this case, so the last three approximation types give the same prediction value.

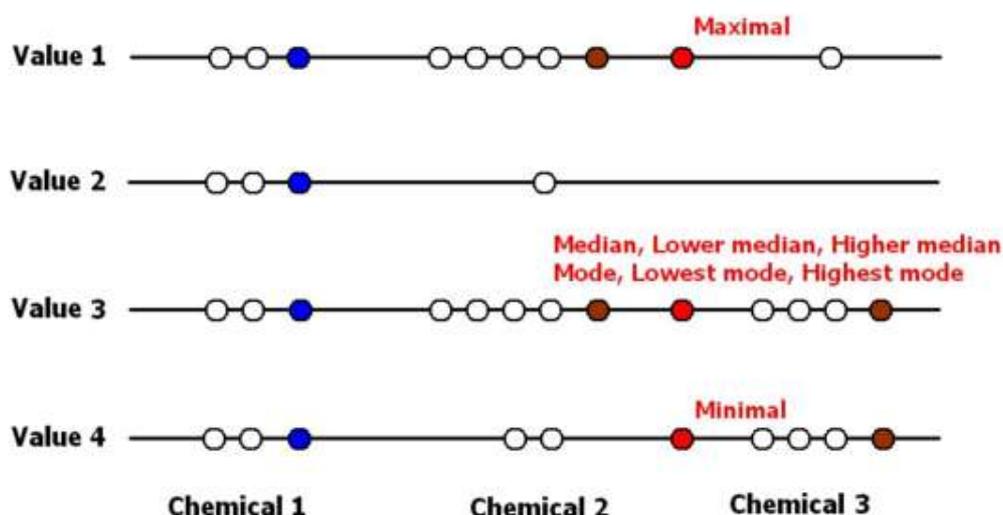


Fig. 7 The prediction values when using mode values for each chemical

8) DataUsage=Lowest mode, the prediction value is:

- **Value 4**, when the approximation type is “Minimal”
- **Value 3**, when the approximation type is “Maximal”
- **No value**, when the approximation type is “Median” – Value 3 and Value 4 are both medians, so the system cannot make a decision automatically
- **Value 4**, when the approximation type is “Lower median”
- **Value 3**, when the approximation type is “Higher median”
- **No value**, when the approximation type is “Mode” – Value 3 and Value 4 are both modes, so the system cannot make a decision automatically

- Value 4, when the approximation type is "Lowest mode"
- Value 3, when the approximation type is "Highest mode".

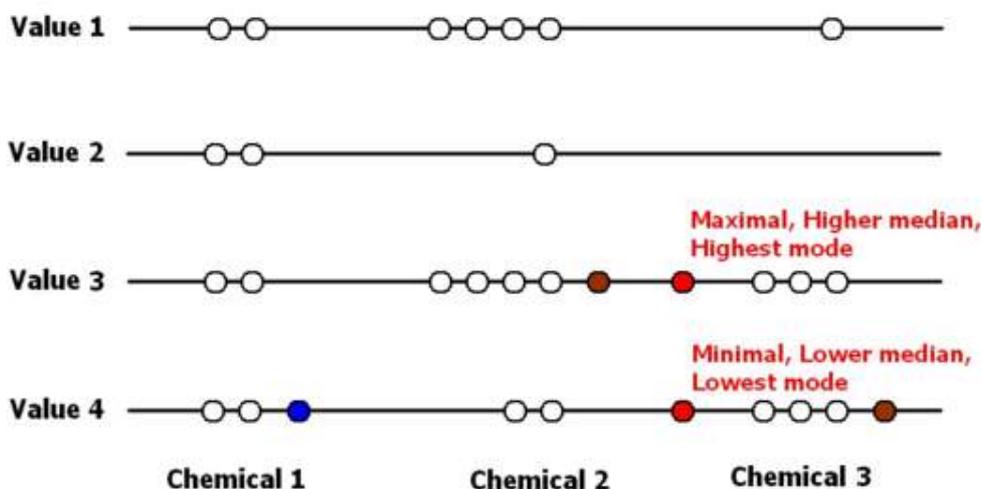


Fig. 8 The prediction value when using lowest mode value for each chemical

- 9) DataUsage=Highest mode, the prediction value is:
- Value 3, when the approximation type is "Minimal"
 - Value 1, when the approximation type is "Maximal"
 - No value, when the approximation type is "Median" – Value 1 and Value 3 are both medians (Value 2 is not taken into account here), so the system cannot make a decision automatically
 - Value 3, when the approximation type is "Lower median"
 - Value 1, when the approximation type is "Higher median"
 - No value, when the approximation type is "Mode" – Value 1 and Value 3 are both modes, so the system cannot make a decision automatically
 - Value 3, when the approximation type is "Lowest mode"
 - Value 1, when the approximation type is "Highest mode".

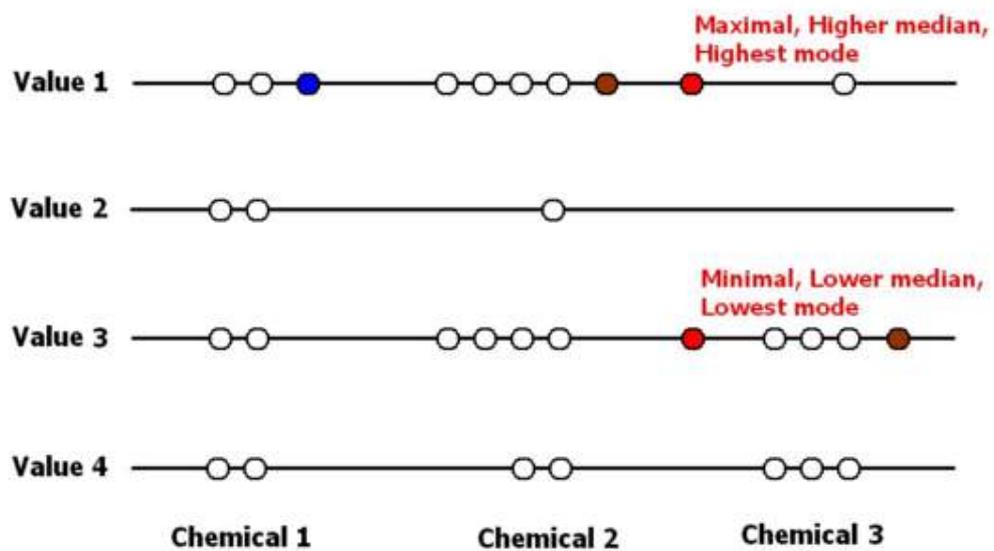


Fig. 9 The prediction value when using highest mode value for each chemical

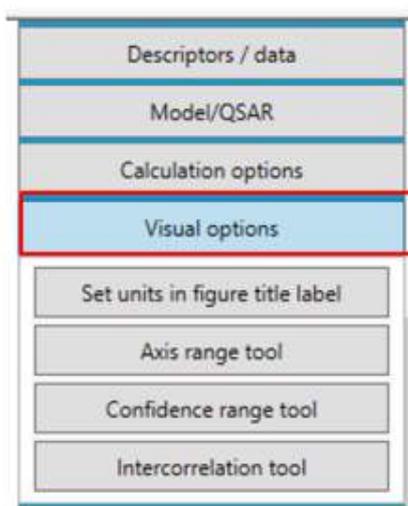
-0-

QSAR Toolbox v.4

D.5.4.2.6. Visual options

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.6. Visual options

D.5.4.2.6. Visual options



- Set units in figure title – set the visualization options. (Figure 1)

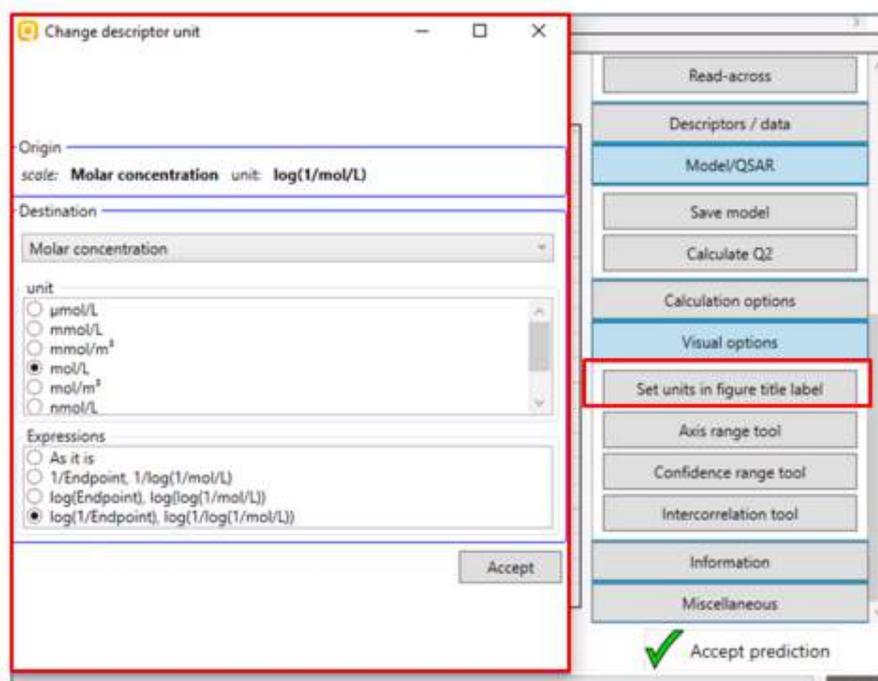


Figure 1

- Set axes ranges – allows manual setting of X and Y axis ranges. (Figure 2)

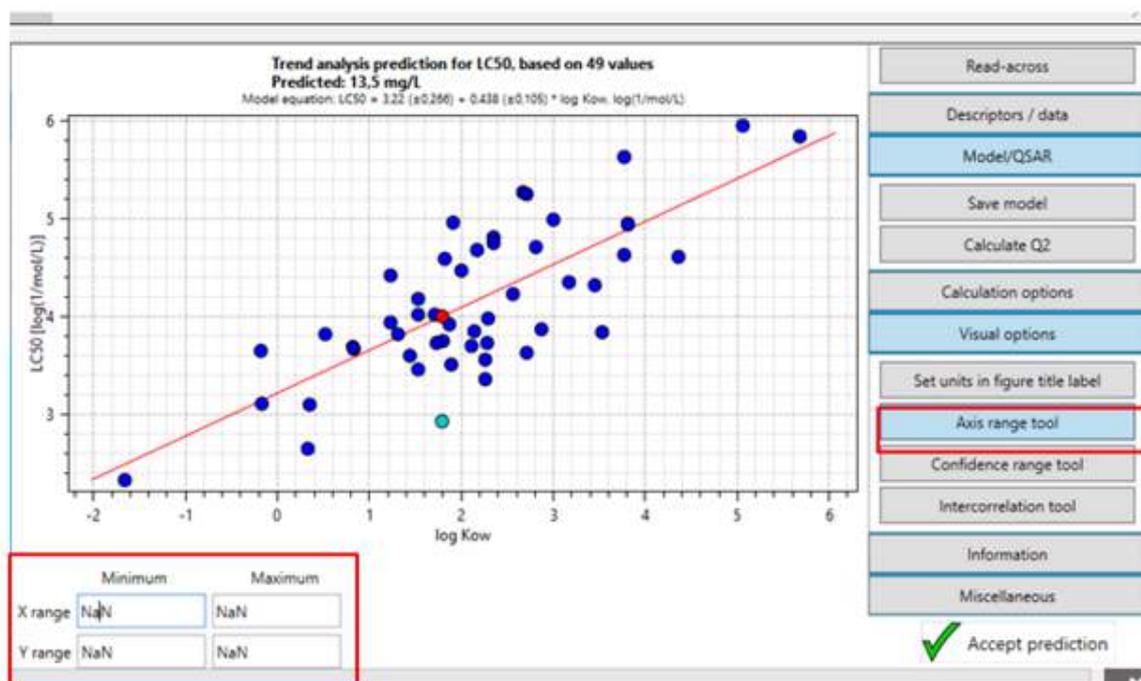


Figure 2

- Confidence range tool – show/hide confidence range in the prediction panel (1). The inside range shows confidence range of regression equation (2), while the outside range shows confidence range of individual prediction(3) (Figure 3)

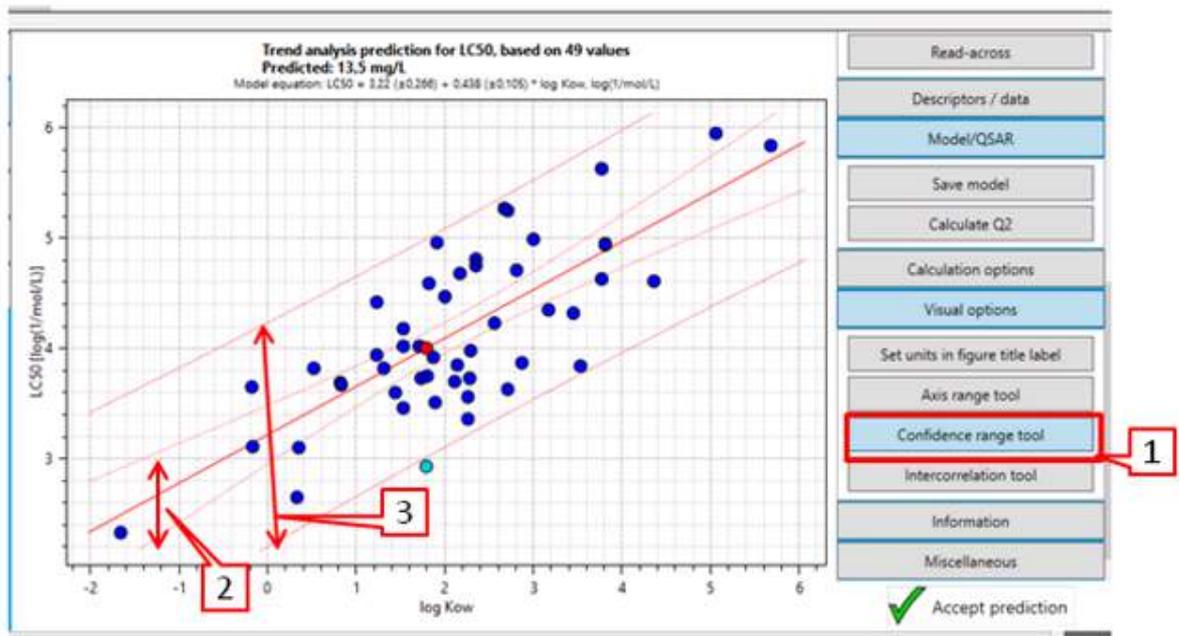


Figure 3

- Intercorrelation tool (1)–shows the inter-correlations panel. When this button is clicked, the points on the graph disappear, and the user is given a choice to select two descriptors (2), which correlation he/she would want to estimate. Once the descriptors are selected the user has to click on Intercorrelation (3). The correlation is shown in the graph (4). (Figure 4)

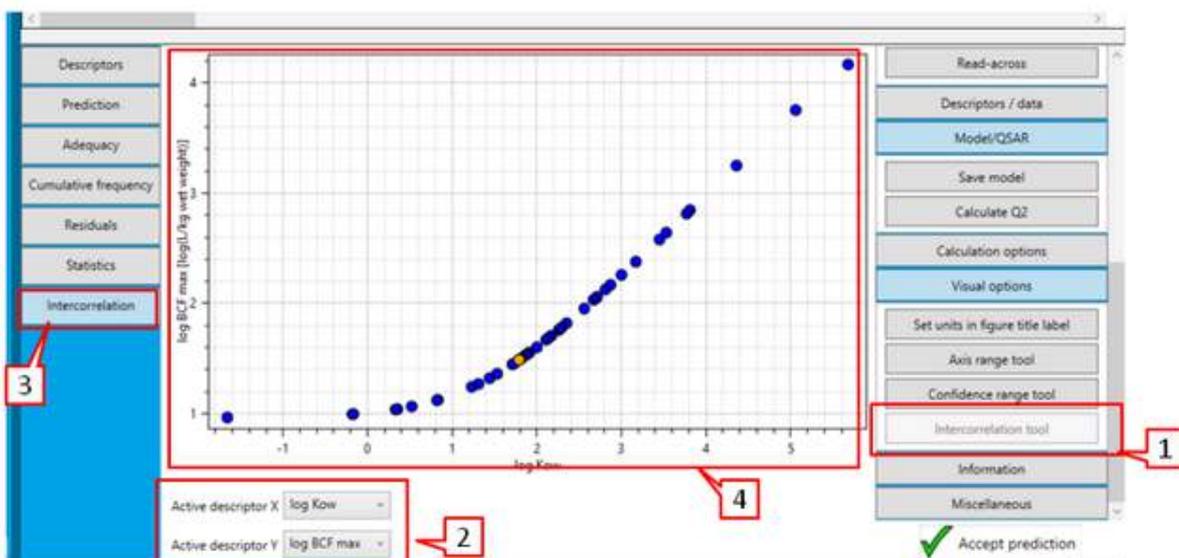


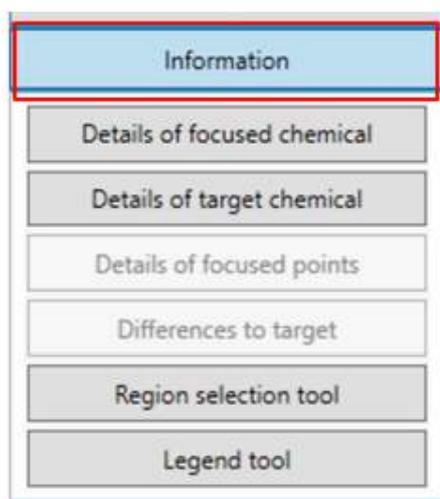
Figure 4

QSAR Toolbox v.4

D.5.4.2.7. Information

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / [Menus of data gap filling](#) / D.5.4.2.7. Information

D.5.4.2.7. Information



- Details of focused chemical (1) – show additional details about the selected data-point's chemical (2). This window (3) includes Chem ID information , structure of chemical, panel with calculated descriptors , panel with metadata. (Figure 1)

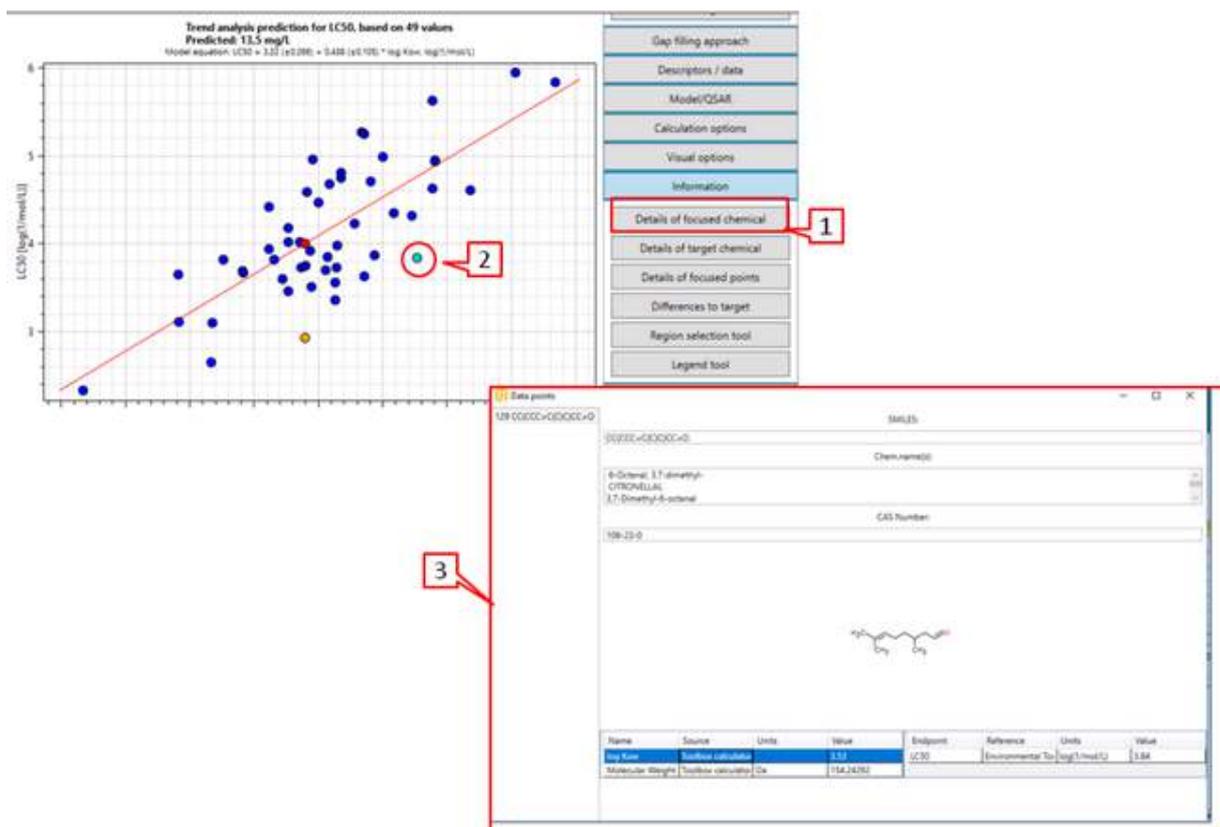


Figure 1

- Details of target chemical – show additional information about the target chemical.
- Differences to target (1)– this shows the differences between the selected data-point (2) and target chemical with respect to all available Profiling methods. The profiling method(s) for which there are some differences are colored in bordeaux (3) when *Color by different* (4) is selected. (Figure 2)

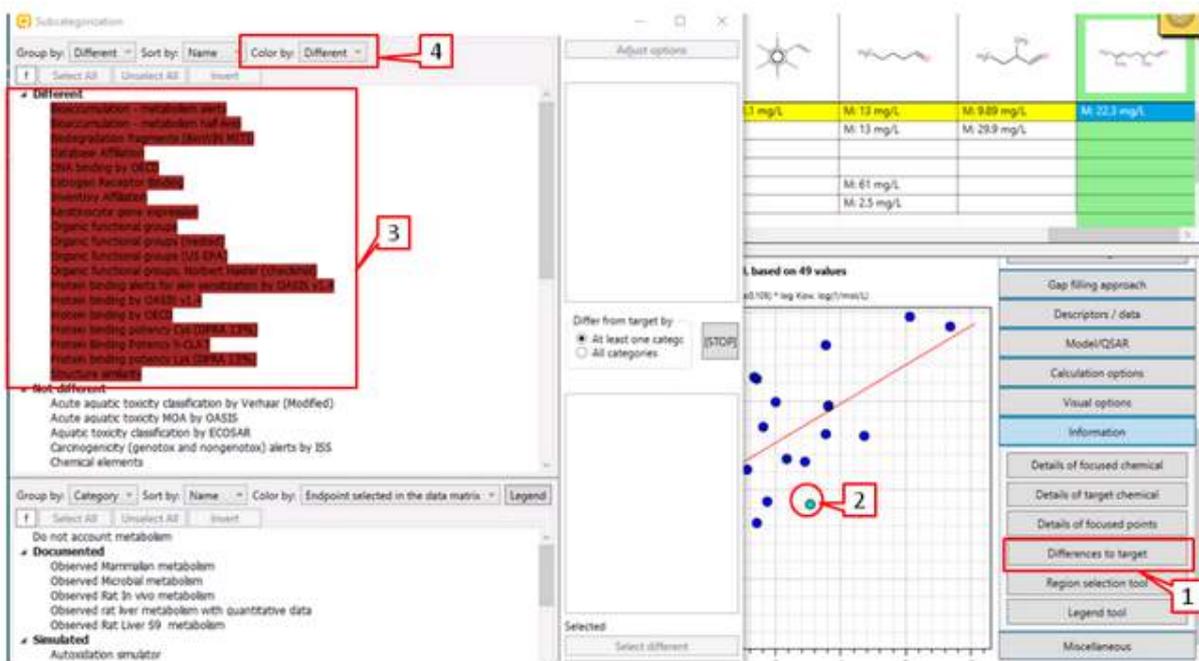


Figure 2

- Region Selection Tool – sometimes there are chemicals with same logKow values in a category of two or more chemicals, presented as a dots, to be one behind the other. This function allows the user to see all the chemicals (dots) within one region. The user has to click on *Region Selection tool* (1), then to drag the mouse (left mouse button) in order to specify the rectangular region (2) (Figure 3). Then use *right-click options* for each focused point (see [Right-click options](#)).

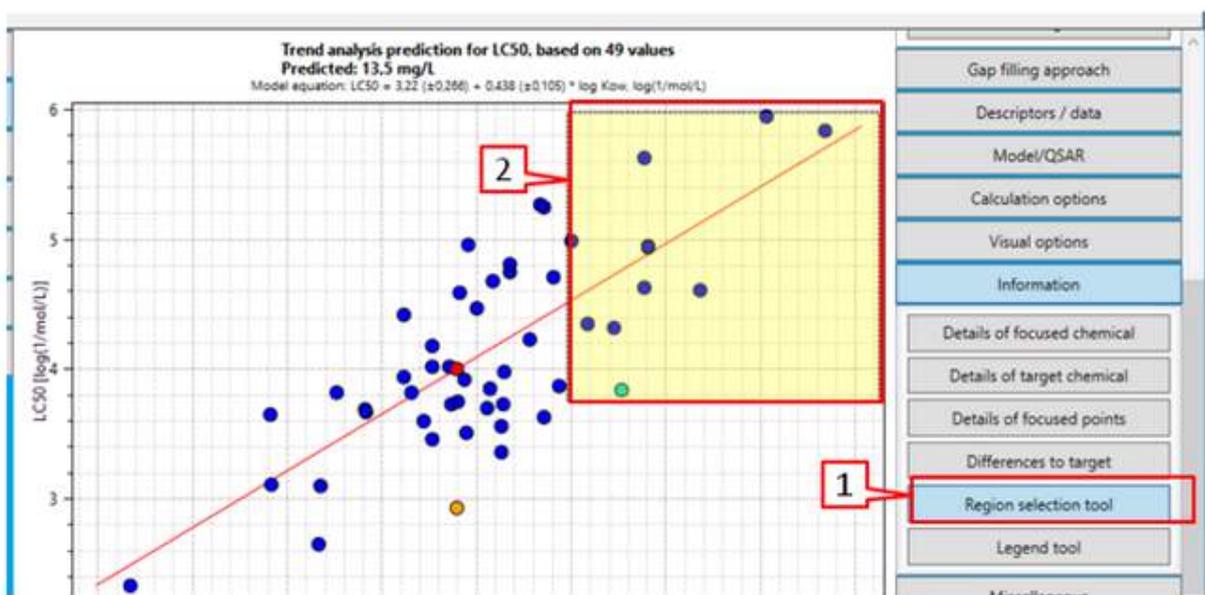


Figure 3

- Legend tool (1) – show/hide the chart's legend (2). (Figure 4)

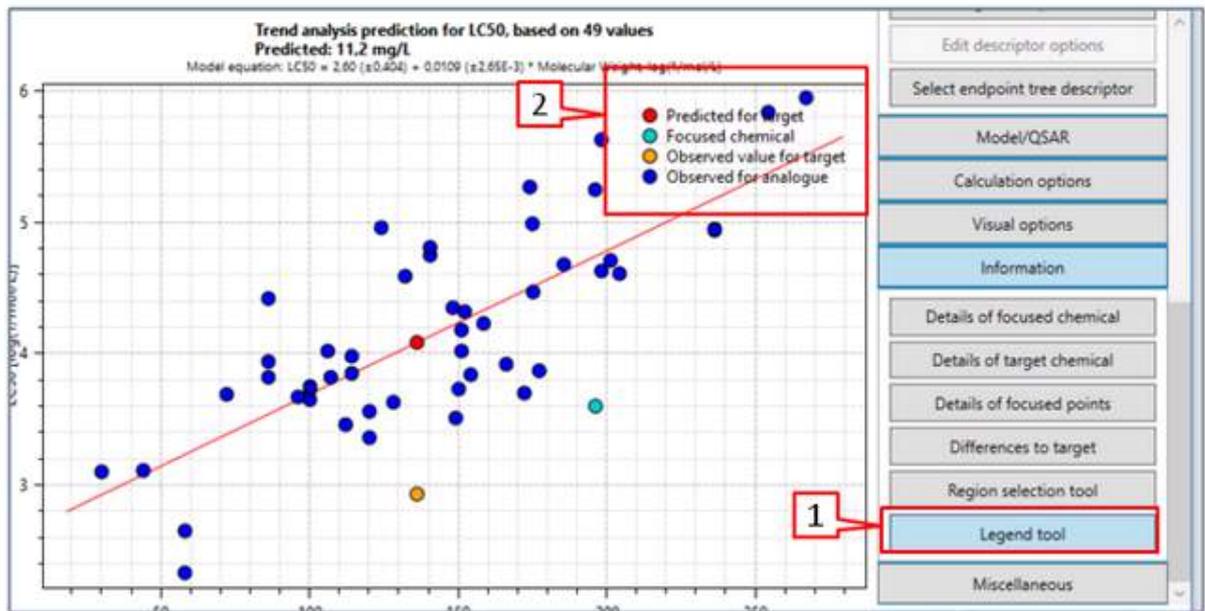


Figure 4

QSAR Toolbox v.4

D.5.4.2.8. Miscellaneous

[Workflow / Filling data gaps / Common features in three gap filling approaches/ Menus of data gap filling / D.5.4.2.8. Miscellaneous](#)

D.5.4.2.8. Miscellaneous

Miscellaneous (1) (Figure 1): *Save picture* (2) option is implemented. The user has to select the file name (3) and then to Save (4) it.

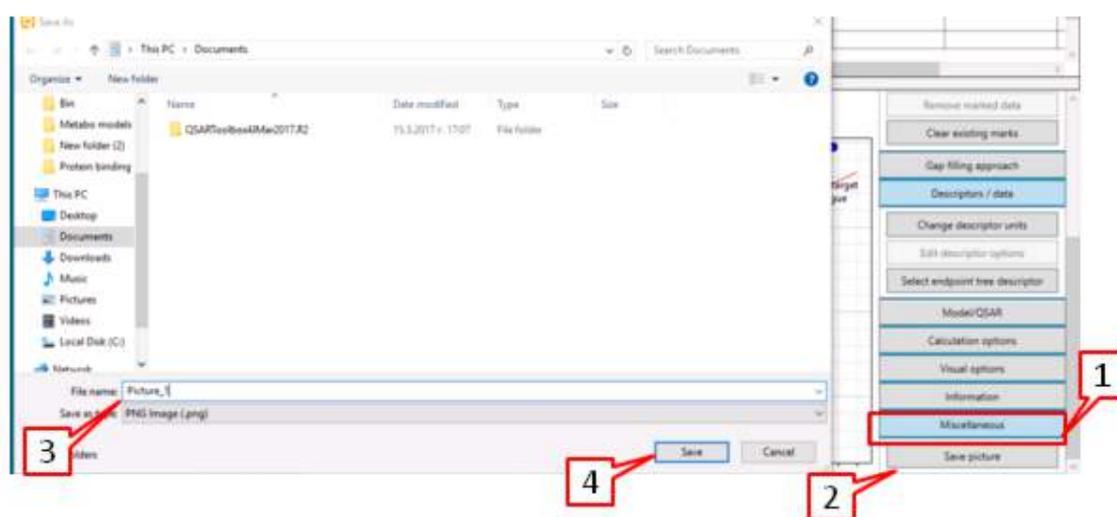


Figure 1

QSAR Toolbox v.4

D.5.4.3. Right click functionality

[Workflow](#) / [Filling data gaps](#) / [Common features in three gap filling approaches](#) / D.5.4.3. Right click functionality

D.5.4.3. Right click functionality

Another common feature for the three gap filling approaches is right click menus (1) in Figure 1. Right click on a selected point gives some of the options implement in the [Menus of data gap filling](#).



Figure 1

-0-

QSAR Toolbox v.4

D.5.5. Specific features of (Q)SARs

[Workflow](#) / [Filling data gaps](#) / D.5.5. Specific features of (Q)SARs

D.5.5. Specific features of (Q)SARs

742 QSAR are implemented in Toolbox 4.0. Clicking on the main toxicological levels (1) in the endpoint tree displayed the number of QSARs relevant to the nodes, which belong to that level (Figure 1):

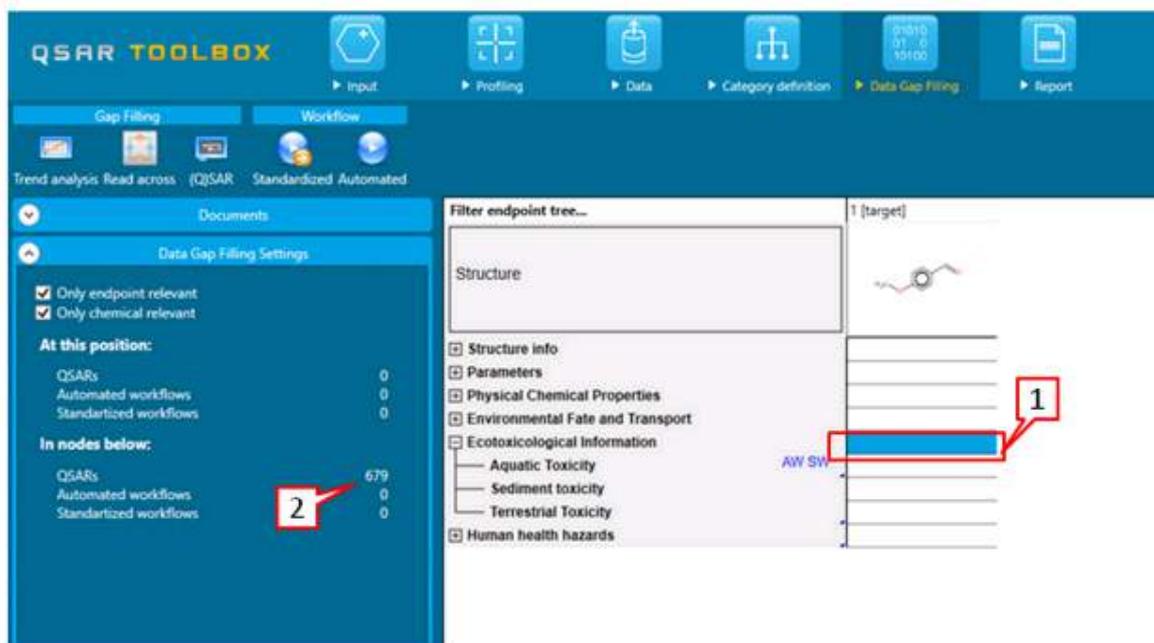


Figure 1

When selecting a lower node, e.g. Aquatic Toxicity (1), information about the number of QSARs at the position (2) is displayed on the left panel under At this position (Figure 2).

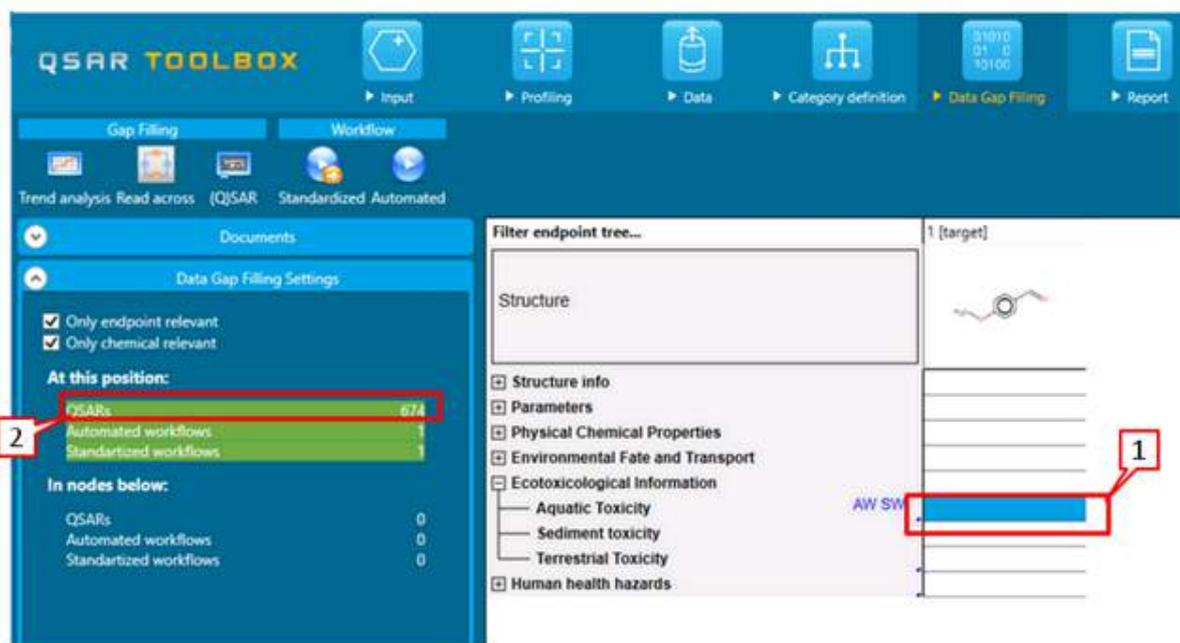


Figure 2

If the user unchecks only endpoint relevant (1) then the number (2) will represent all available QSARs in the software regardless of the endpoint.



Figure 3

The following steps has to be executed in order to run a QSAR (Figure 4):

- select the relevant endpoint (1);
- check Only chemical relevant (2). This will reduce the number of applied QSARs (5) based on the functionalities of the target chemical. If the user uncheck this option then all QSARs relevant for the endpoint will be applied regardless of the functionalities present in the target chemical.
- Press (Q)SAR (3);
- A pop-up window will appear (4), with list of all QSAR(s) that have been executed and their predictions(6)

Figure 4 shows the QSAR Toolbox interface with the following steps highlighted:

- 1: Selecting the relevant endpoint (Aquatic Toxicity).
- 2: Checking 'Only chemical relevant' in the Data Gap Filling Settings.
- 3: Pressing the QSAR button.
- 4: A pop-up window showing the list of QSAR models and their predictions.
- 5: The Predicted column in the QSAR models table.
- 6: The Class column in the QSAR models table.

QSAR name	Predicted	Class	Duration
M4 - LC50 - Pimephales promelas (fathead minnow)	47.8 mg/L		96 h
M3 - LC50 - Pimephales promelas (fathead minnow)	182 mg/L		96 h
M2 - LC50 - Pimephales promelas (fathead minnow)	87.9 mg/L		96 h
M1 - LC50 - Pimephales promelas (fathead minnow)	31.0 mg/L		96 h
ECOSAR: FISH 96 h LC50 Mortality Aldehydes (Monol)	10.0 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	96 h
ECOSAR: DAPHND 48 h LC50 Mortality Aldehydes (Monol)	18.0 mg/L	Branchiopoda (branchiopods)	48 h
ECOSAR: GREEN ALGAE 96 h EC50 Aldehydes (Monol)	26.1 mg/L		96 h
ECOSAR: FISH 96 h LC50 Mortality Aldehydes (Monol)	1.85 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	96 h
ECOSAR: DAPHND 96 h LC50 Mortality Aldehydes (Monol)	2.52 mg/L	Branchiopoda (branchiopods)	96 h
ECOSAR: GREEN ALGAE 96 h EC50 Aldehydes (Monol)	8.02 mg/L		96 h
ECOSAR: Fish (SW) 96 h LC50 Mortality Aldehydes (Monol)	14.2 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	96 h
ECOSAR: Fish (SW) 96 h LC50 Mortality Aldehydes (Monol)	1.66 mg/L	Actinopterygii (ray-finned fishes,spiny rayed fishes)	96 h
Photoinduced toxicity of PAHs	Not Phototoxic		

Figure 4

Right click on the name of a (Q)SAR (1), displays a pop-up menu (2) where the user can select to display the training set (3) or the test set. (Figure 5)

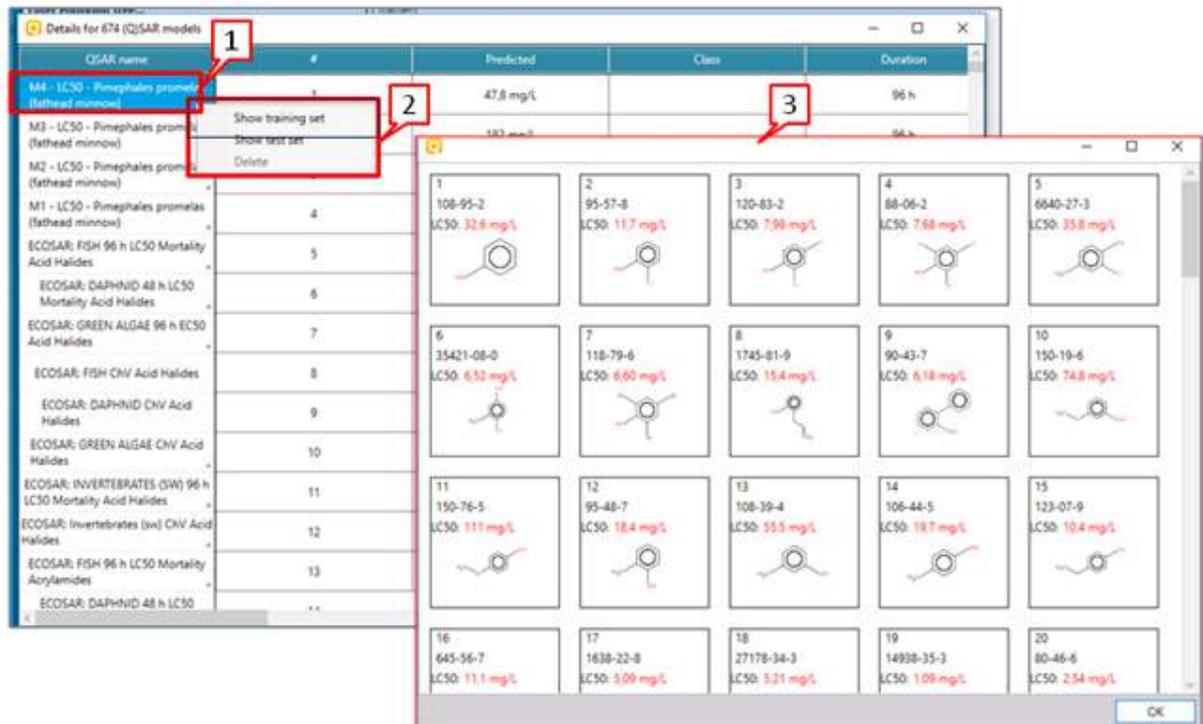


Figure 5

If the user wants to accept a given prediction (1) a pop-up window with three options appears (2) (Figure 6)

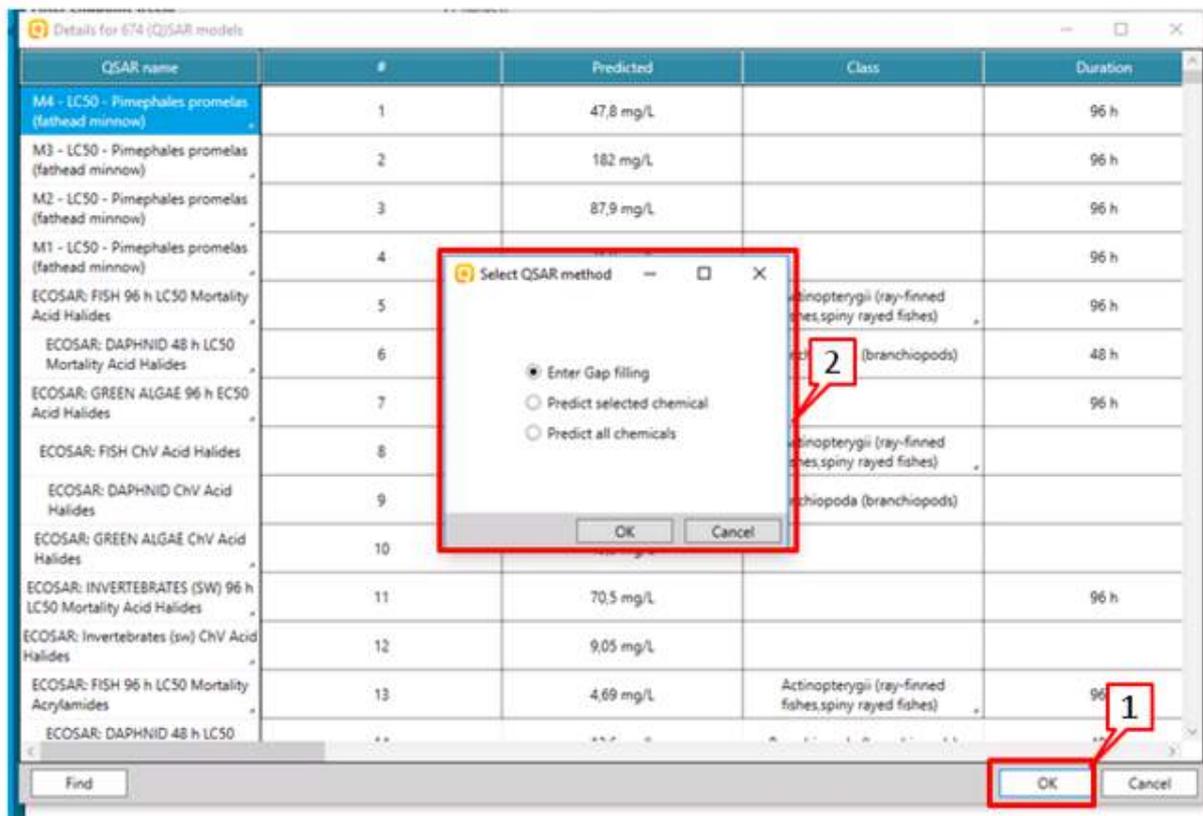


Figure (6)

- Enter Gap filling: shows the training set chemicals (1) of the QSAR, the prediction for the target chemical (2) and the regression graph (3) (Figure 7). Press Legend (4) to see the legend of the points colouring . Once the prediction is accepted (5) it will be shown in the data matrix.

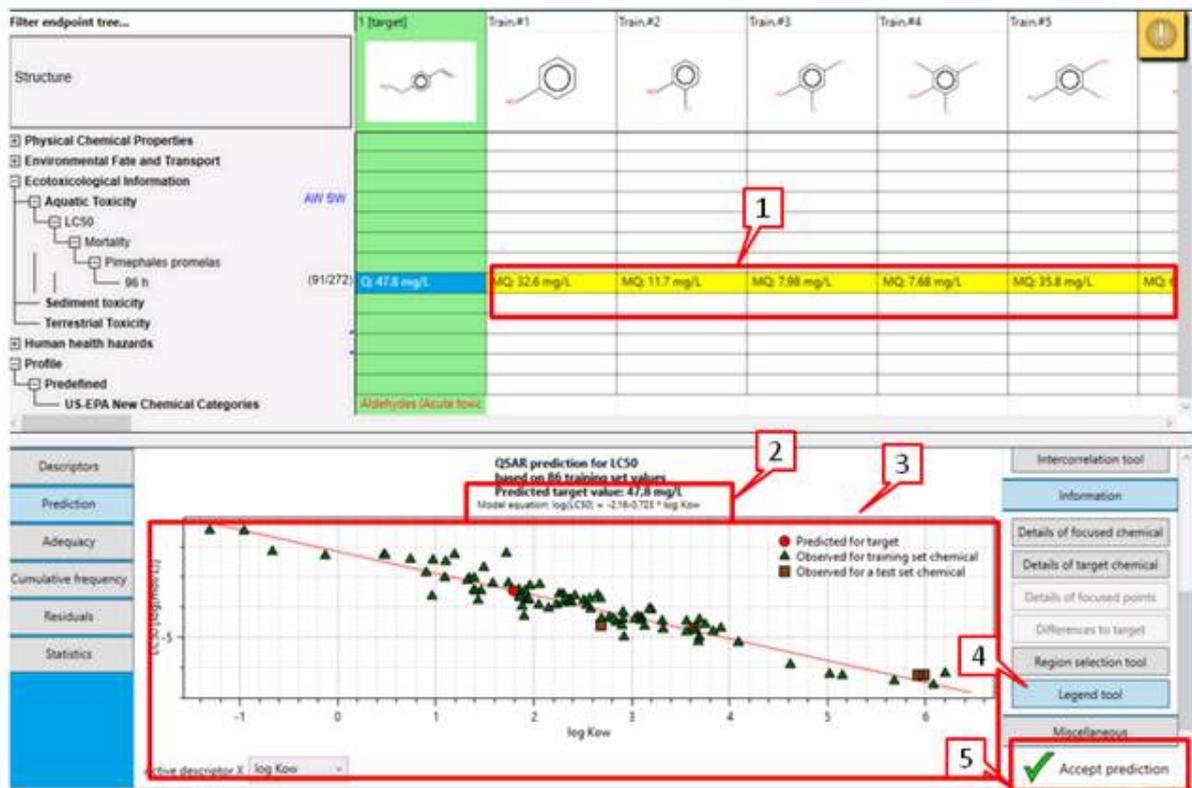


Figure 7

- Predict Selected chemical: Select the (Q)SAR (1), then *Predict selected chemical* (2) and finally the prediction is displayed on the data matrix (3).

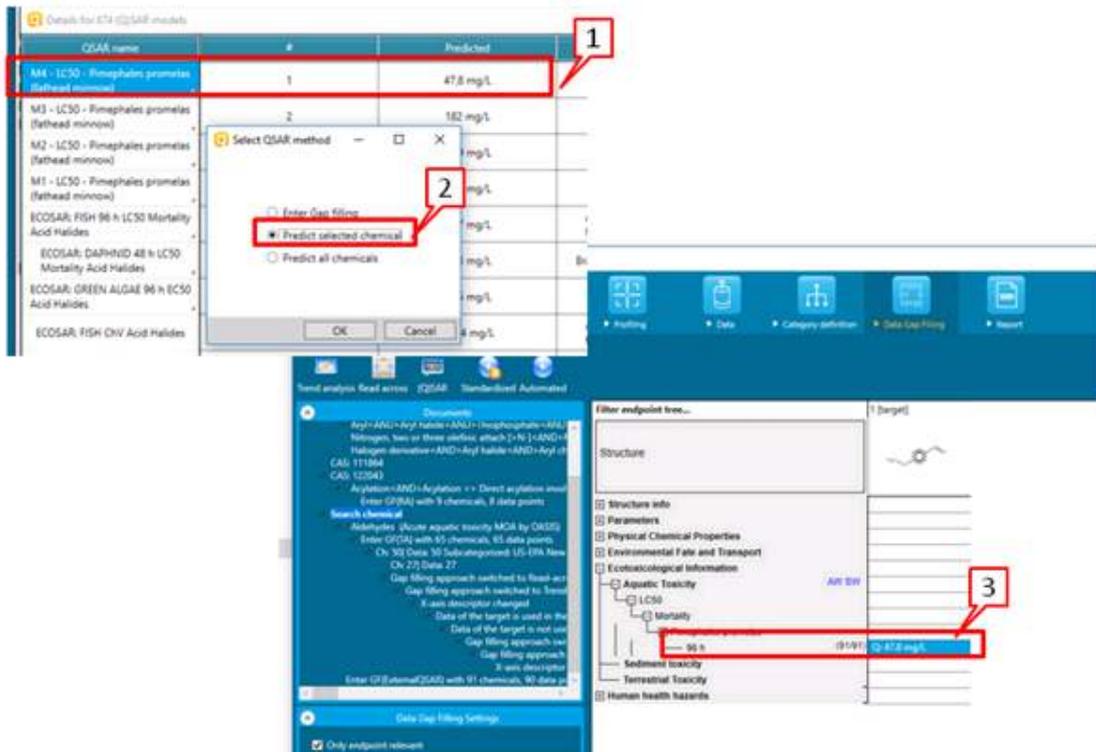


Figure 8

- Predict all chemicals (Figure 9): Select the (Q)SAR (1), then Predict all chemical (2) and finally the prediction is displayed on the data matrix (3).

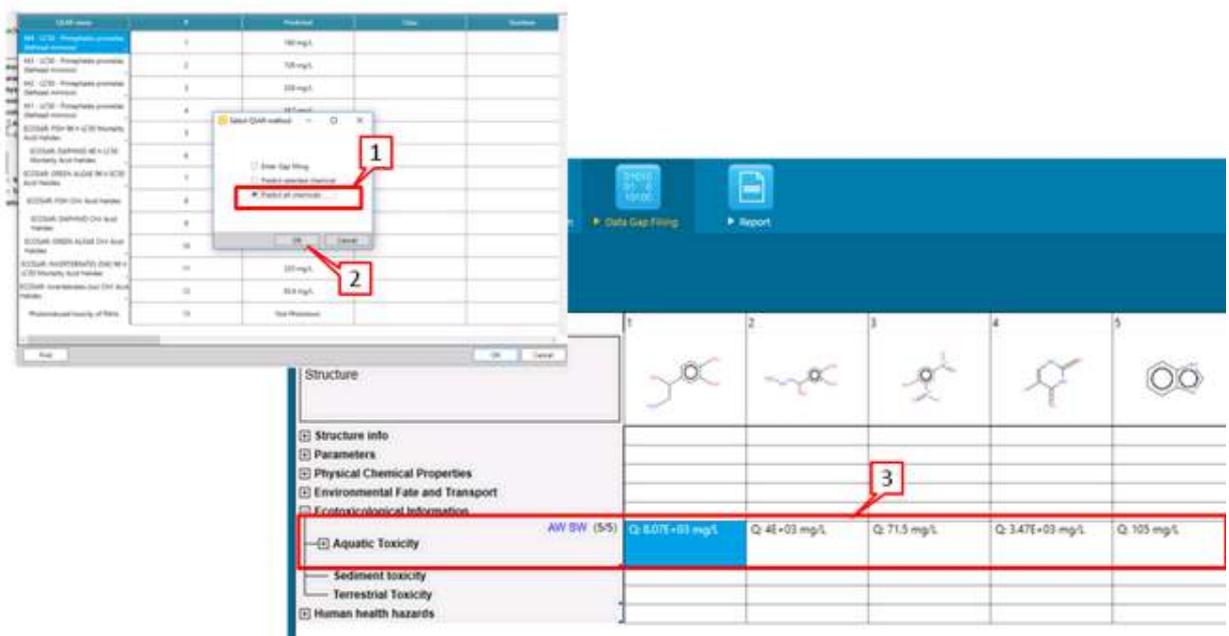


Figure 9

-0-

D.5.6. Automated and Standardized workflows

D.5.6. Automated and Standardized workflows

Algorithms for automated data gap filling have been developed for skin sensitization (LLNA and GPMT data) and acute aquatic toxicity to fish (*Pimephales promelas*, Mortality, LC50, 96 h). Once started, the automated workflows (AWs) follow the implemented logic and finished with prediction without interaction by the user. The same algorithms have been implemented as standardized execution of data gap filling (SW). The main differences as compared to the AWs are that the domain of application is expanded in the SWs (including other species, durations, etc.) and SWs allow interactions by the user and thus, different selection than those in AW could be done.

-0-

D.5.6.1. Automated workflow

D.5.6.1. Automated workflow

I. Automated workflow for Ecotox endpoint for a single chemical

The automated workflow (AW) has been developed for a one endpoint only:

- Endpoint: concentration of a toxic substance lethal to half of the test animals – LC50
- Effect – Mortality
- Species – fish (*P. promelas*)
- Duration – 96 h

For execution of the automated workflow, please follow the steps described below (Figure 1):

- Go to Data Gap filling (1);
- Select *Automated* (2);
- Click on the Ecotoxicological Endpoint (3) radio-button in the pop-up window (4);
- Finally press *OK* (5).

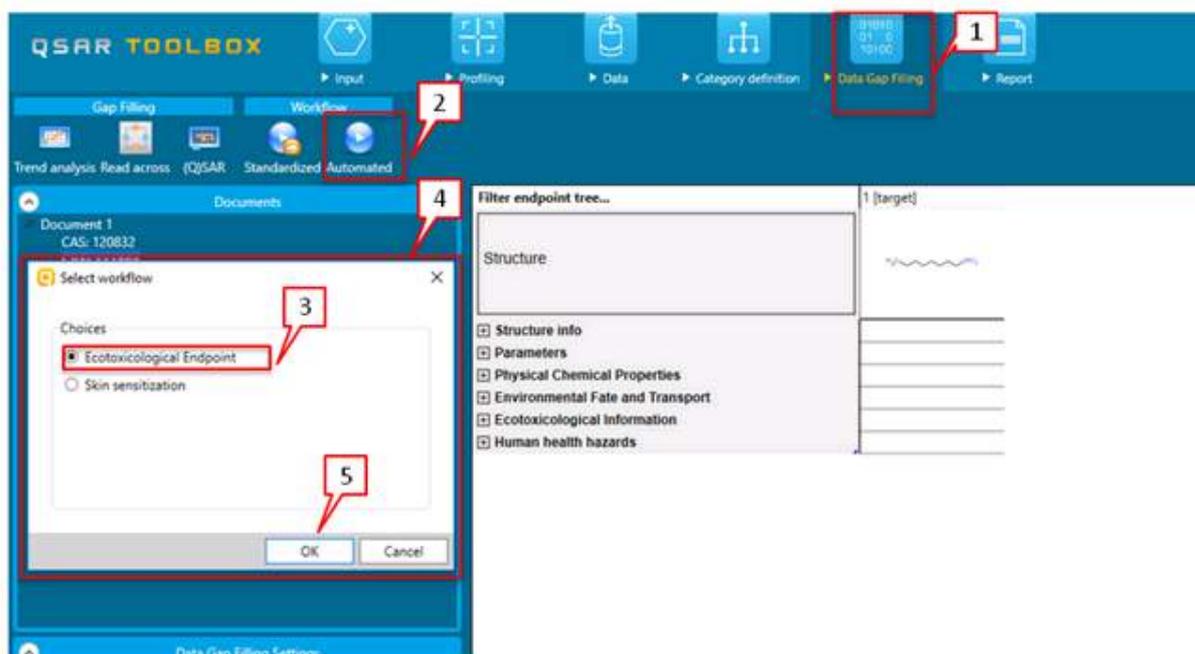


Figure 1 Activation of AW for Ecotoxicological endpoint

A dialogue window gives the user a choice to select the endpoint path (1, 2)

(Figure 2)

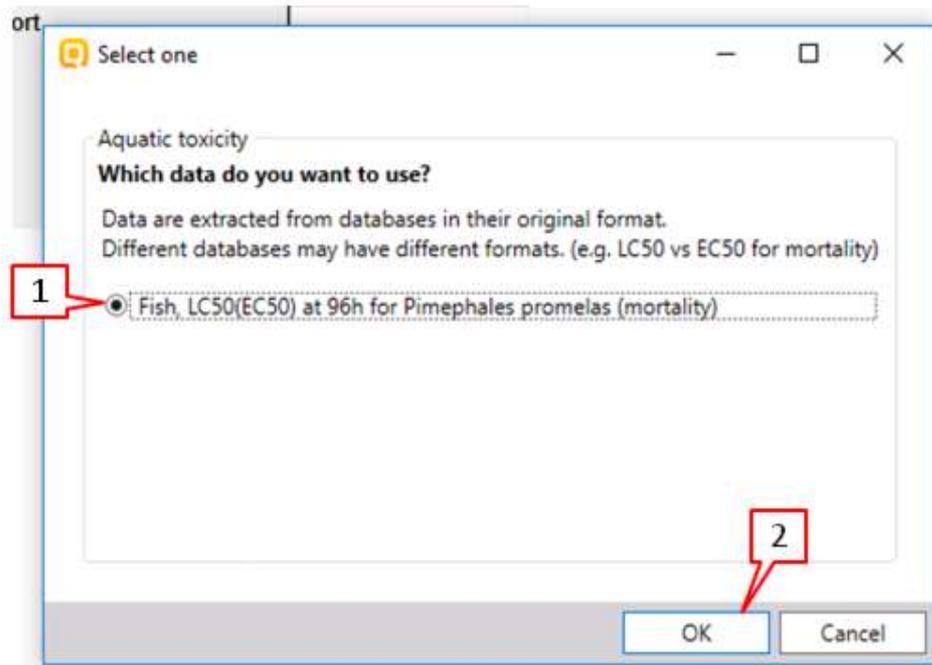


Figure 2

The next window that appears is the so-called Workflow controller (Figure 3). It has two main navigation buttons Pause and Stop (1). Below the caption of the workflow name (2) is included the general task (3) indicating about the currently performed action and the active task (5) providing information for the performed subtask of the general task. The sequence of steps applied during workflow execution is summarized in the activity log. Click on *Show activity log* (5) to see the sequence of steps.

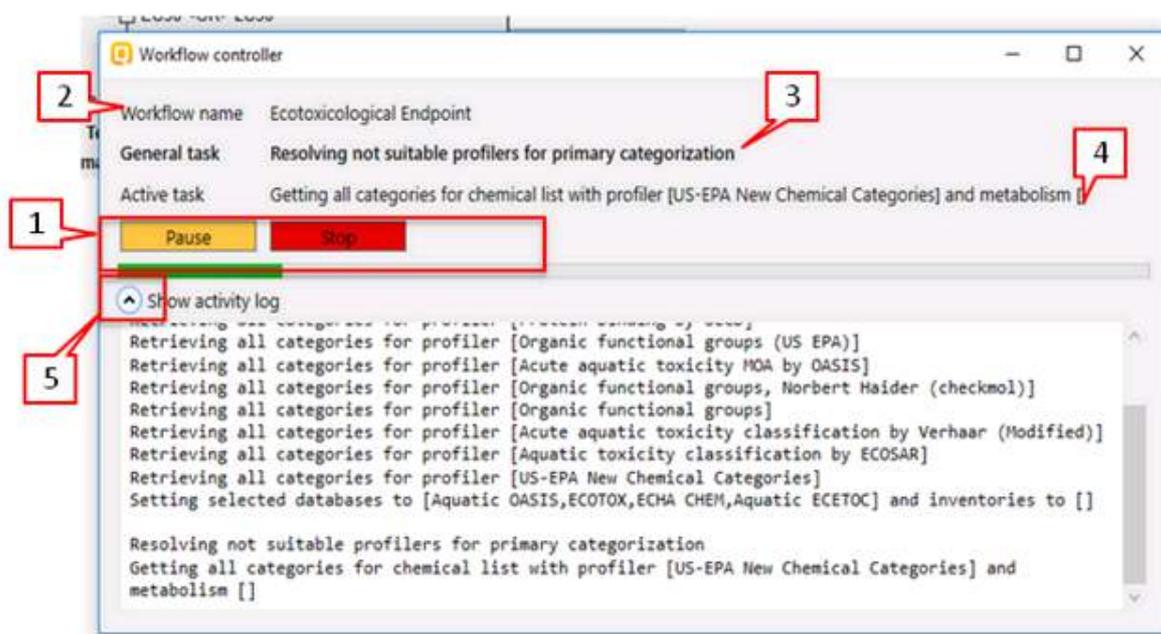


Figure 3

Sequences of steps are applied consequentially automatically during execution of AW; no user activity needed here.

Once the AW finishes, a message *"Prediction accepted successfully"* appears (1) and the progress bar is completely filled (2) (Figure 4). Click OK (3) and then close the workflow controller window by X button (4).

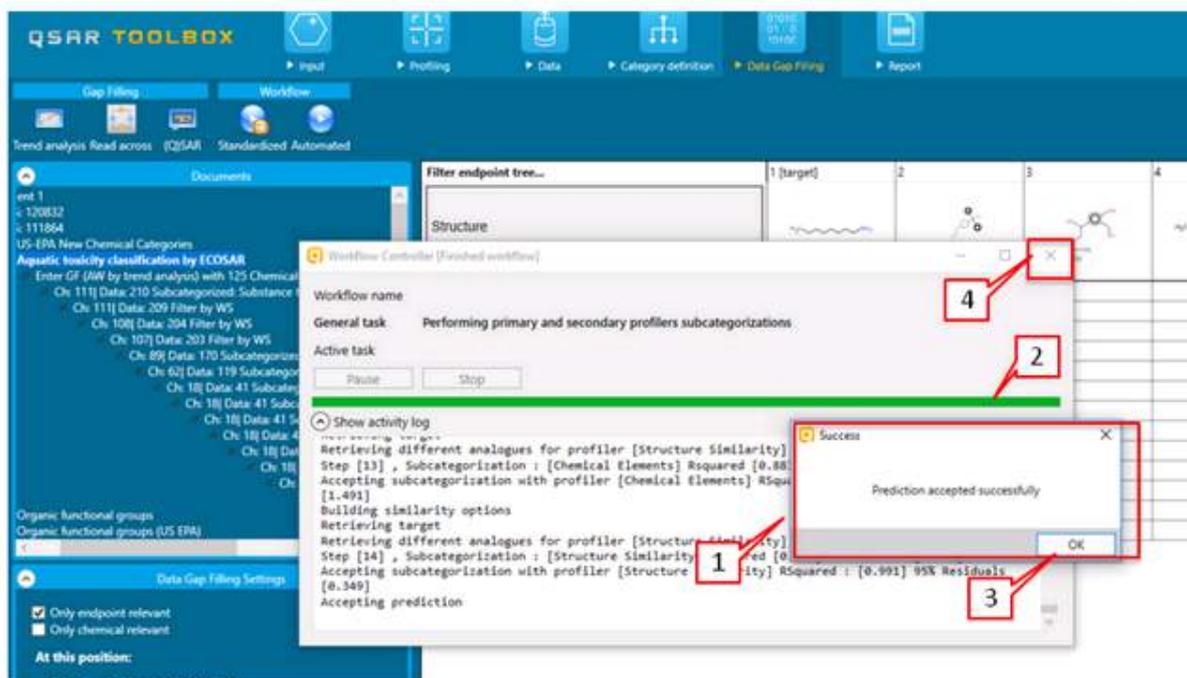


Figure 4

In case the prediction does not answer the criteria for acceptance of the prediction or not enough data is collected for primary grouping then the following messages appear: "No enough data to build primary group" or "We couldn't find a valid answer "(1) (Figure 5): click OK (1) and close the workflow controller (2).

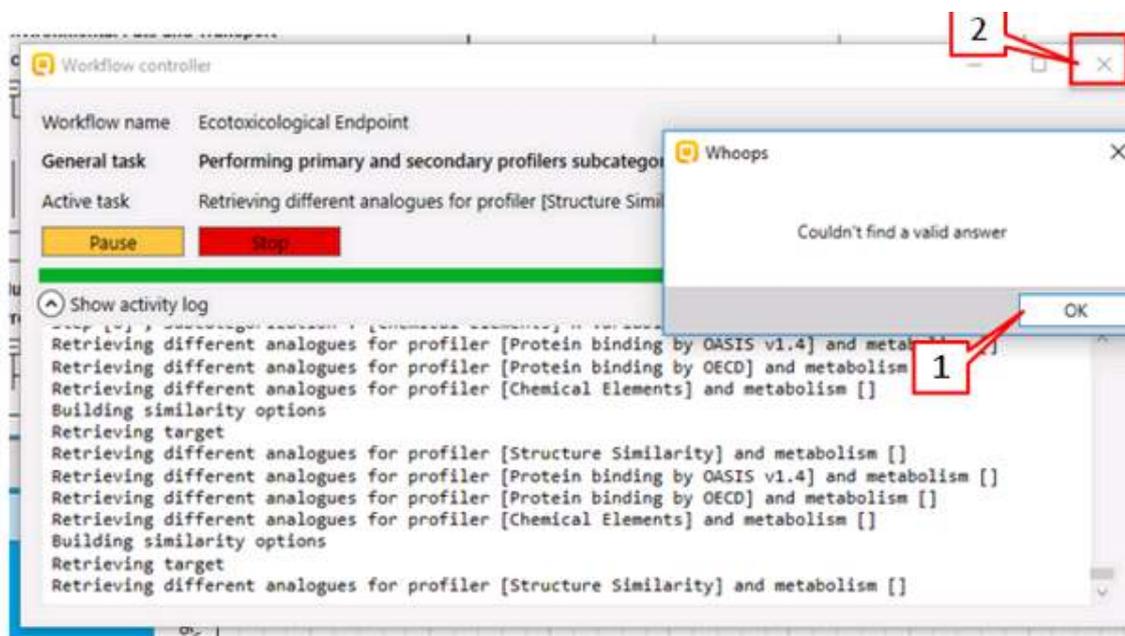


Figure 5

II. Automated workflow for Ecotoxicological endpoints for a batch of chemicals

See [Input options](#) on how to input a set of chemicals.

Then select Data Gap filling (1), press *Automated* (2), select *Ecotoxicological Endpoint* (3) from the pop-up window (4) and then press OK (5). (Figure 6)

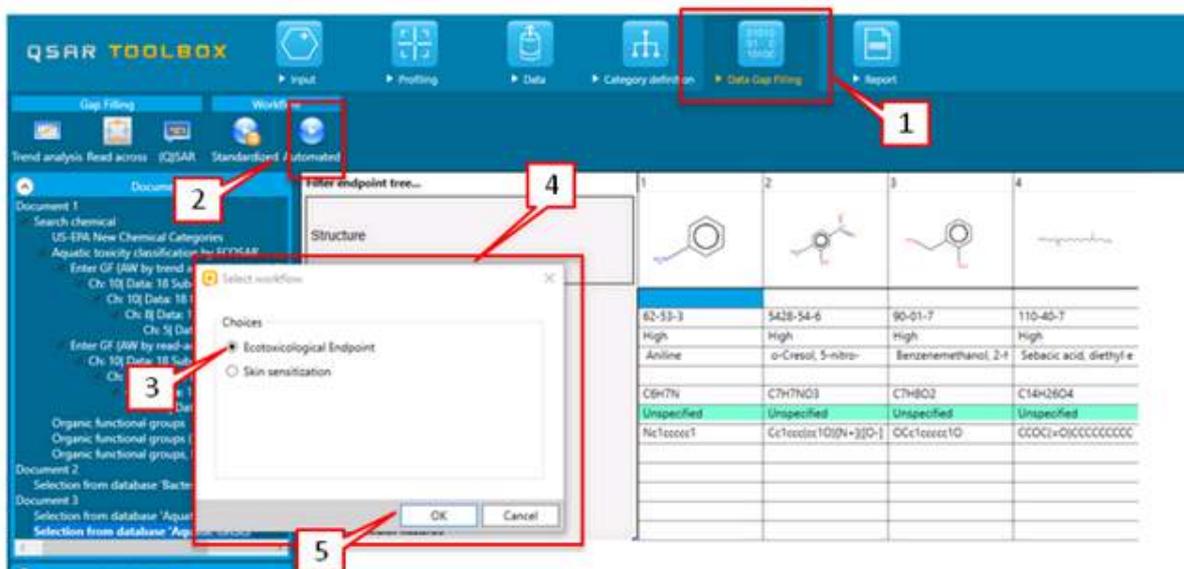


Figure 6

A pop-up window (1) prompts the user to select the endpoint path (2):

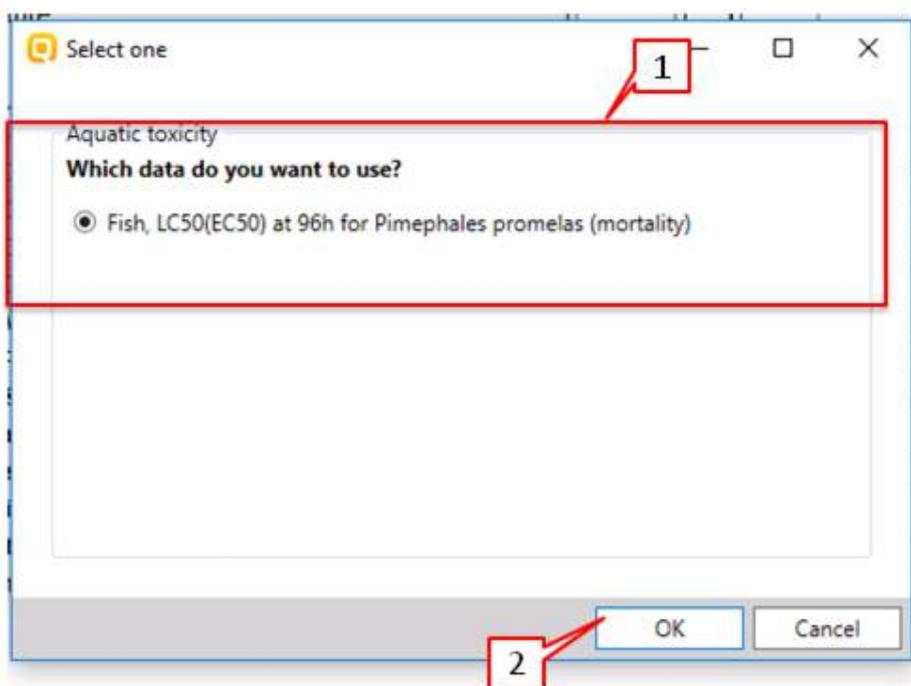


Figure 7

A Workflow controller window appears, which is not active (1) (Figure 8). The pop-window *Select range* (2) is displayed, where the user has to select the range of chemicals from the set, which has to be predicted. Finally the user has to press *OK* (3).

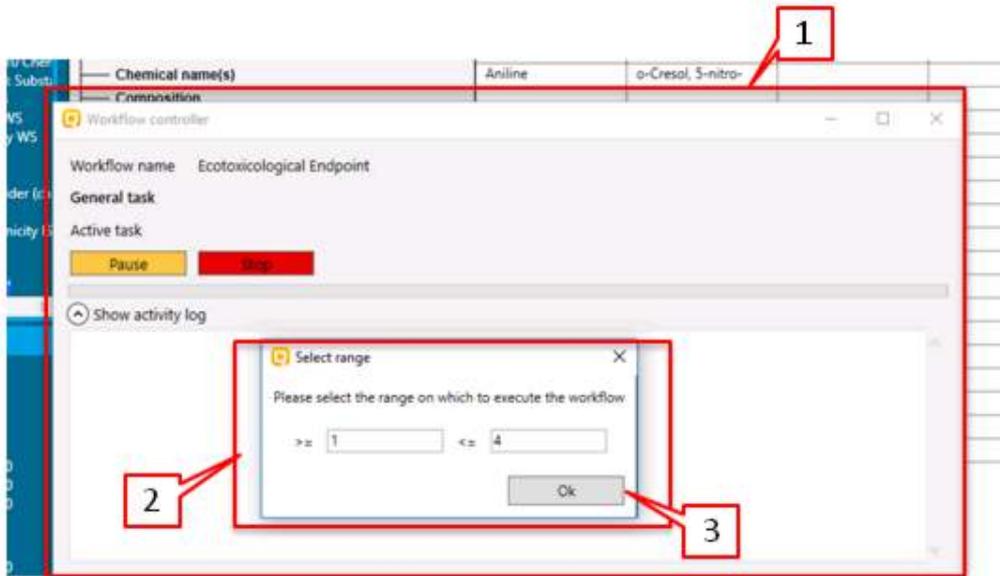


Figure 8

When the workflow finishes there is an indication in the workflow controller(1) (Figure 9). Also the progress bar is completely filled (2). The predictions are displayed on the matrix (3). There is also an indication that 4 out of 4 chemicals are predicted.

Finally, the user has to close the workflow window by pressing X button (4).

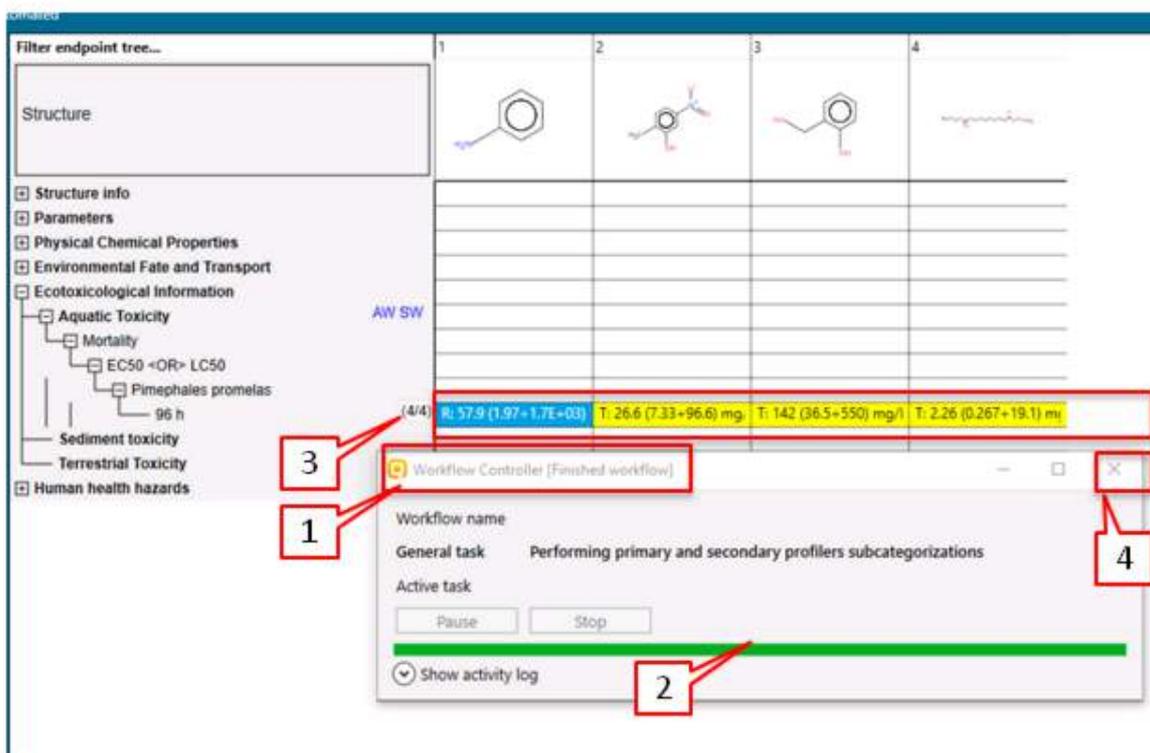


Figure 9

-0-

D.5.6.2. Standardized workflow (SW)

D.5.6.2. Standardized workflow (SW)

Example for application of SW when trend analysis is applied :

Target endpoint: LC50, 96h all fish (CAS# 111-86-4)

The following steps have to be applied: go to Data Gap filling module (1), click on Standardized button (2), select Ecotoxicological endpoint (3) from the pop-up window(4) and finally press OK (5). (Figure 1)

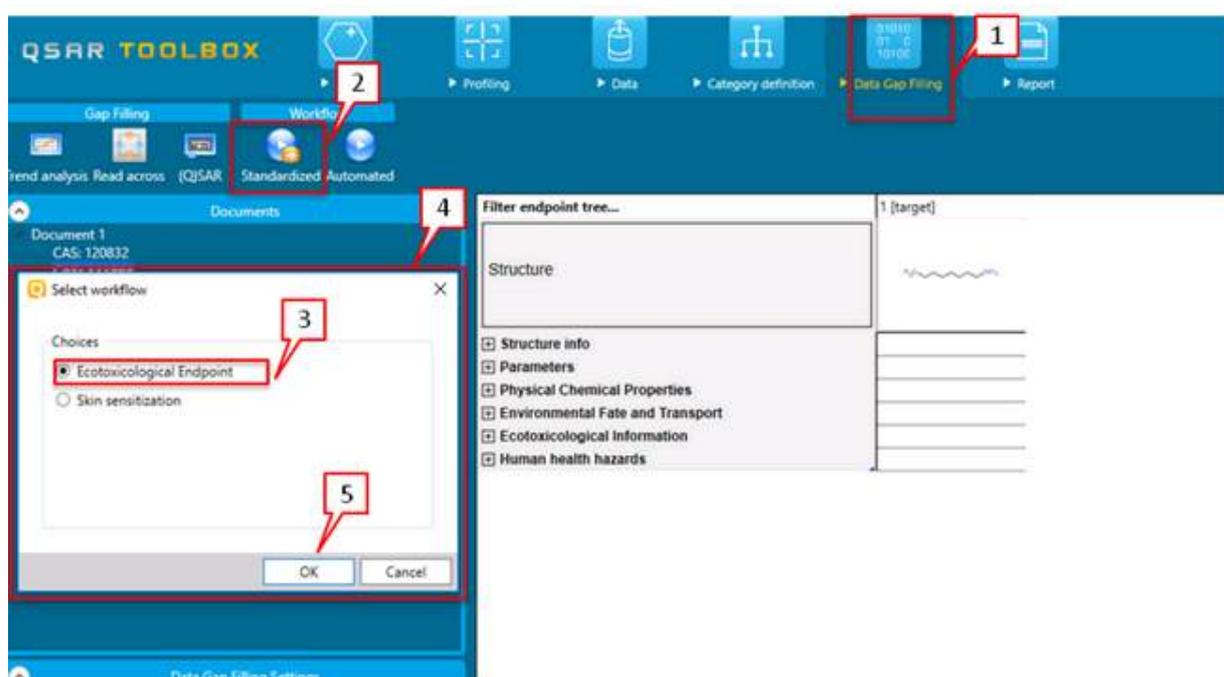


Figure 1

A dialogue window appears (Figure 2) prompting the user to select the précised endpoint. The user has to select the endpoint (1) and then to press OK (2).

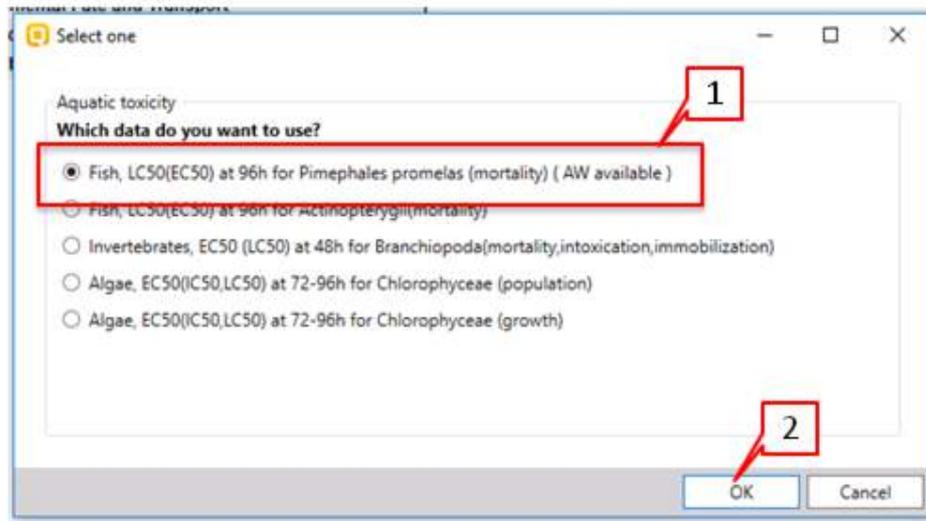


Figure 2

Once the target endpoint is selected the Workflow controller window appears (Figure 3).



Figure 3

In a few sections after Workflow controller appears an additional window with list of databases for selection (Figure 4). Databases used in AW are indicated. The user has to select the databases (2) and then to press OK(3).

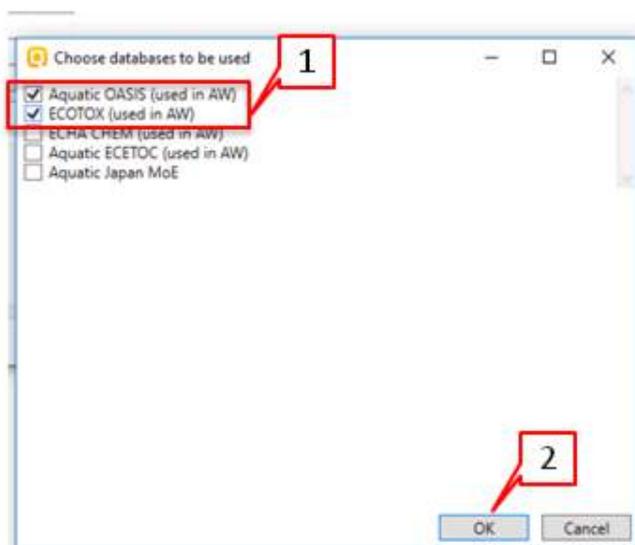


Figure 4

After the selection of databases, the workflow applies each of the primary grouping profilers to the target chemical. As a result a list with all applied profilers is provided with the corresponding chemicals/data statistics (Figure 5). The user has to select one profiler (1) and then to press OK (2).

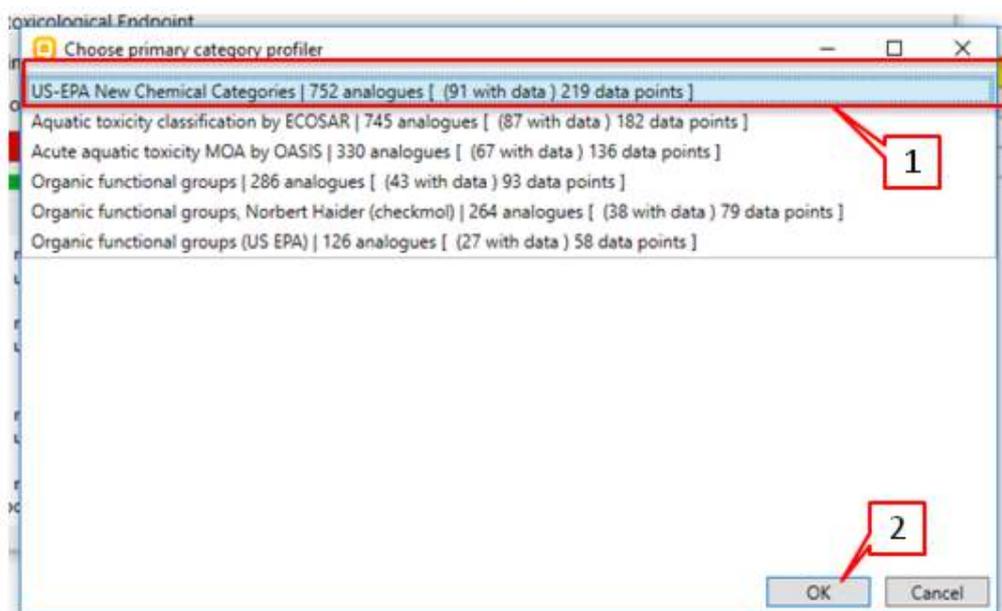


Figure 5

When the subcategorisation starts, a Subcategorisation window (1) appears, which is separated into two sections: primary grouping and secondary grouping (2) (Figure 6). When the Color by (3) is set to Outcome, the profilers

are coloured . The meaning of the coloring is explained by a color chart (4) visualized when clicking on the Legend button (5).

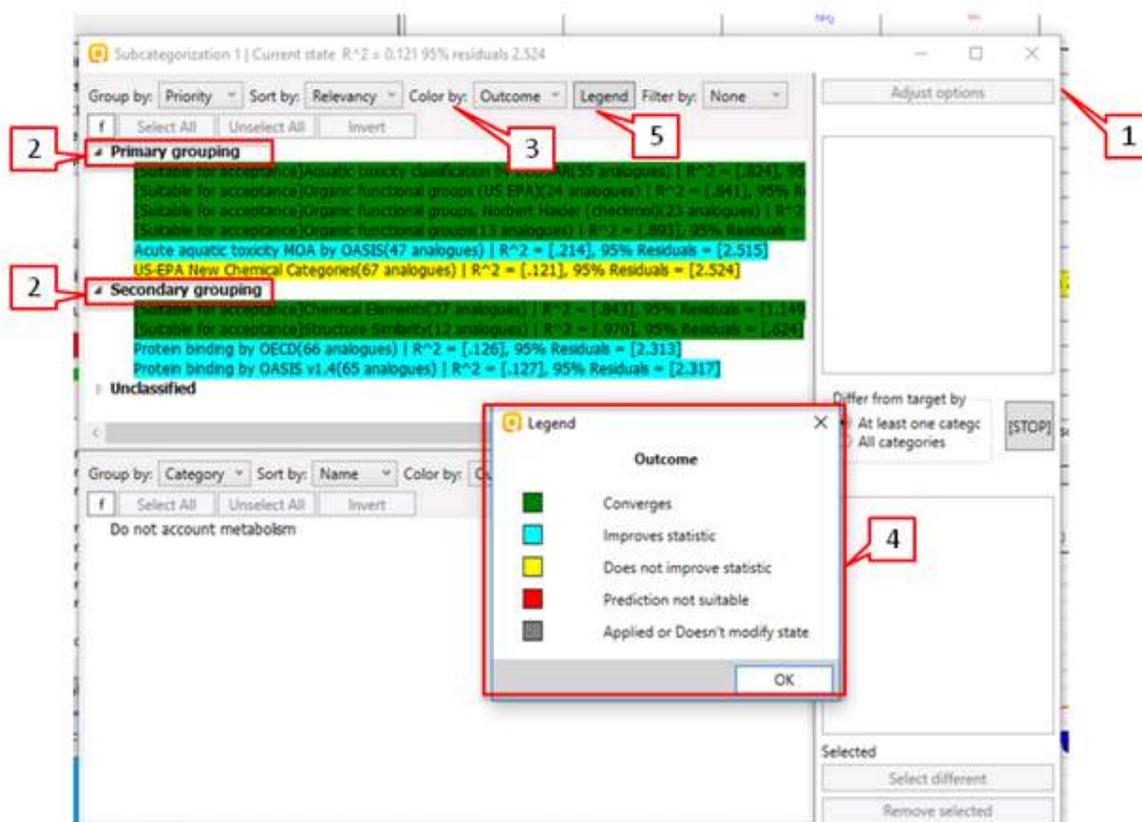


Figure 6

Color legend:

Green – application of the profiler converges of the criteria for acceptance the prediction

Blue – application of the profiler improve statistic only (e.g. increase Correlation coefficient R2 or decrease 95% of residuals)

Yellow – application of the profiler does not change the statistics

Red – criteria for acceptance the subcategorization not reached

Grey – already applied profiler

Other features of the Subcategorisation window are (Figure 7):

- Statistical criteria of the current state (1)
- Criteria for acceptance of prediction (2)
- Indication for profiler suitability (3)
- Navigation bar allowing different ordering of the profilers based on selected criteria (4)

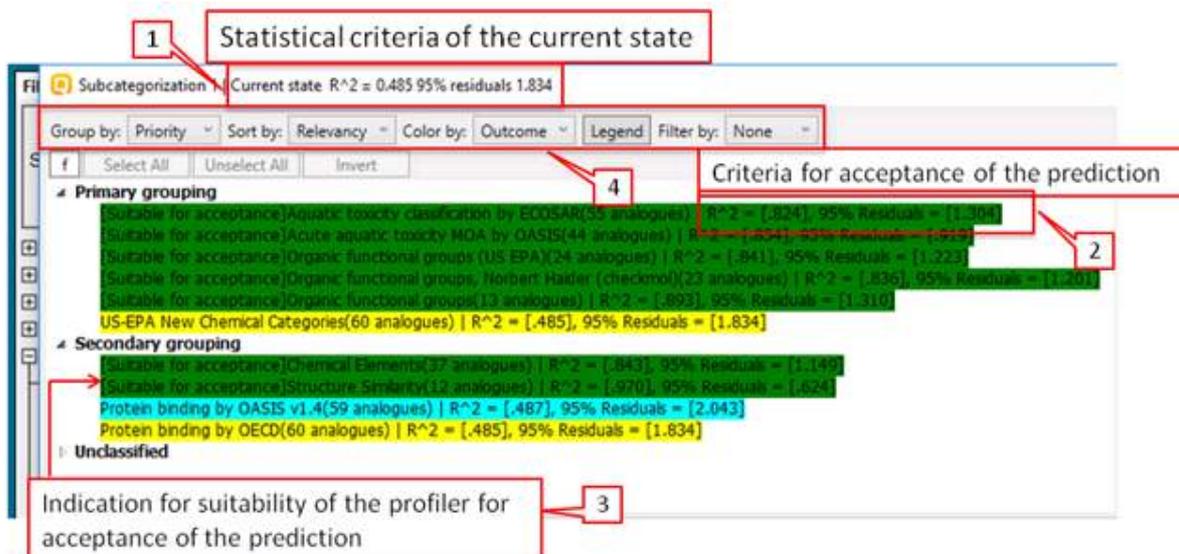


Figure 7

Click on the first profiler in the primary indicated as suitable in order to see, which different functionalities present in the analogues compared to the target chemical. (Figure 8). They are highlighted in blue. Press Remove selected (3).

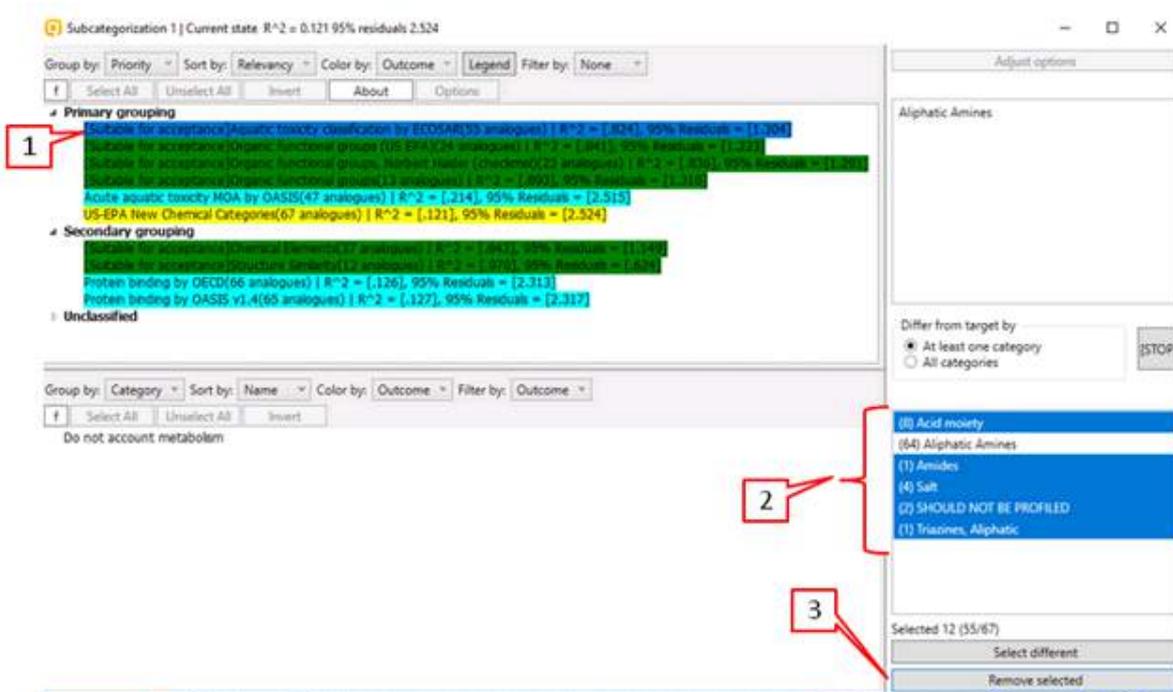


Figure 8

The systems remove different analogues. If the results satisfies the criteria for acceptance of the prediction the following message will be displayed (1) (Figure 9):

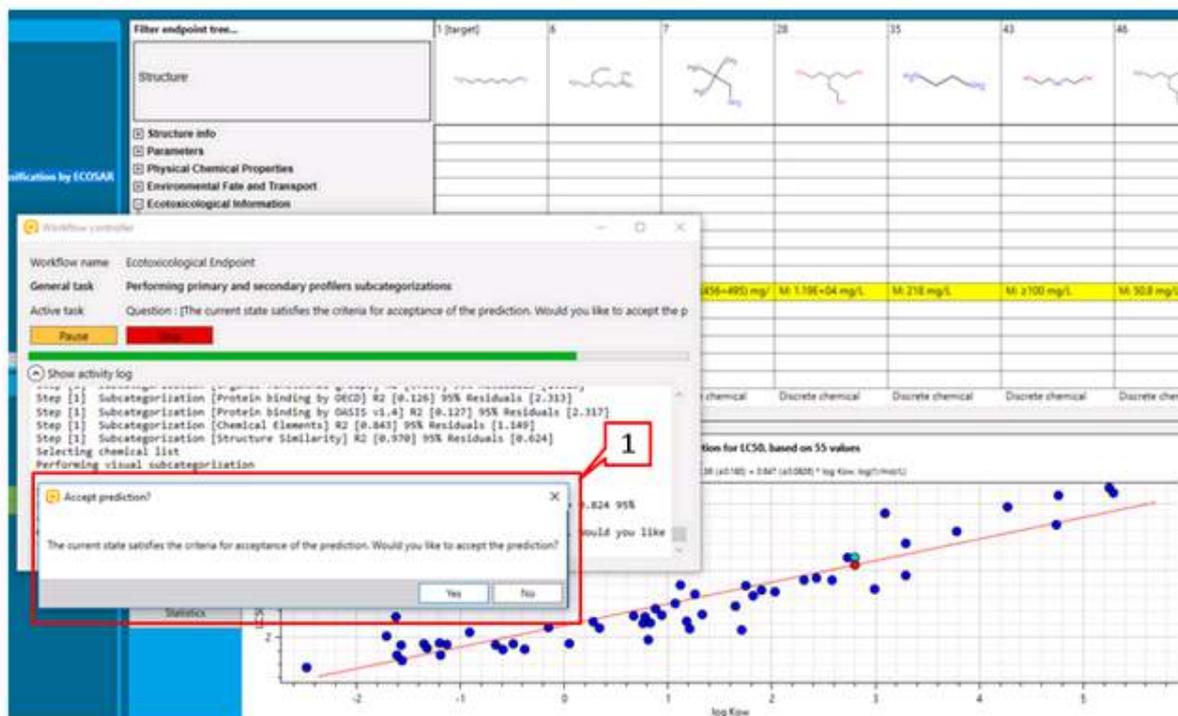


Figure 9

The user can press Yes and accept the prediction or Press No and continue with the other profilers in the subcategorisation until he/she is satisfied with the criteria of the prediction.

After pressing Yes, the following message is displayed:

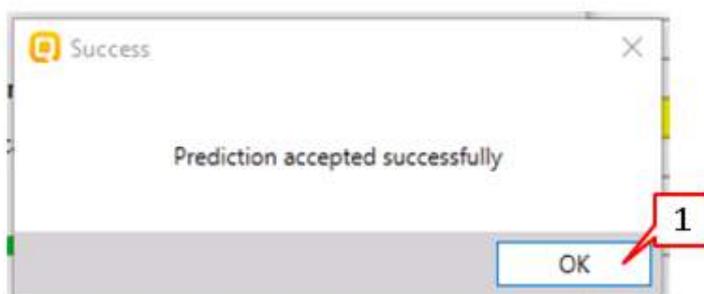


Figure 10

The user has to click on Stop button (1) in the Workflow controller and to confirm that she/he wants to exit the workflow (2) (Figure 11).

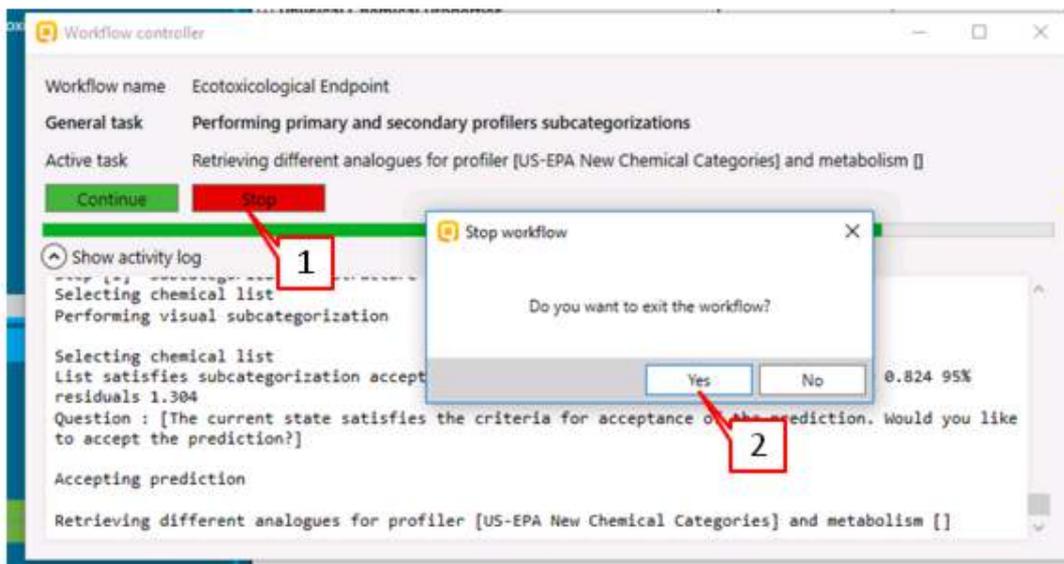


Figure 11

A *Finished workflow* (1) message is written in the workflow controller. Finally press X button (2) to close the controller (Figure 12).

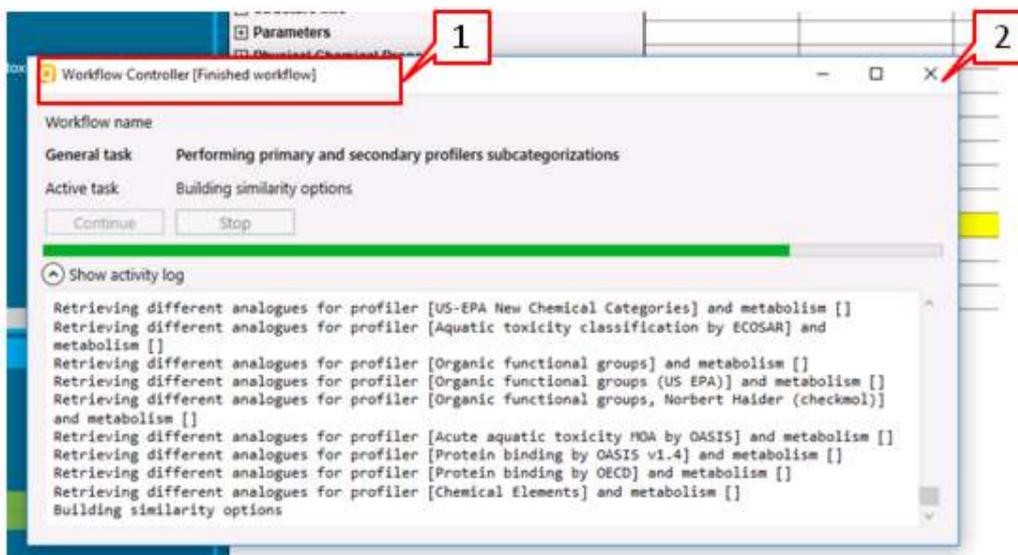


Figure 12

The result is displayed on the data matrix (1) (Figure 13). The workflow finishes on the document level of the primary subcategorisation (2).

The screenshot displays the OECD QSAR Toolbox interface. On the left, a 'Documents' pane shows a tree structure with 'Document 2' selected. A red box labeled '2' highlights the 'US-EPA New Chemical Categories' folder. The main area is divided into a 'Filter endpoint tree...' on the left and a data table on the right. The filter tree shows 'Aquatic Toxicity' expanded to 'Mortality' and 'EC50 <OR> LC50'. A red box labeled '1' highlights the 'Rimephales promelas' entry. The data table below shows a value of 'T: 8.64 (0.544 - 137) mg/L' in a yellow-highlighted cell.

Filter endpoint tree...	1 [target]	2
Structure	<chem>CCCCCCCC</chem>	<chem>OC(=O)C</chem>
Structure info		
Parameters		
Physical Chemical Properties		
Environmental Fate and Transport		
Ecotoxicological Information		
Aquatic Toxicity		
Mortality		
EC50 <OR> LC50		
96 h		
Rimephales promelas		
(91/220)	T: 8.64 (0.544 - 137) mg/L	
Sediment toxicity		
Terrestrial Toxicity		
Human health hazards		

Figure 13

-O-

QSAR Toolbox v.4

D.5.7. Methodologies for estimating toxicity of a set of chemicals

[Workflow / Filling data gaps/ D.5.7. Methodologies for estimating toxicity of a set of chemicals](#)

D.5.7. Methodologies for estimating toxicity of a set of chemicals

Three methodologies for estimating toxicity of set of chemicals are developed:

- [Independent mode](#) (Dissimilar action)
- [Similar mode](#) (Dose concentration)
- Specific models

Both concepts (independent action and dose/concentration addition) are based on the assumption that chemicals in a mixture do not influence each other's toxicity, i.e. they do not interact with each other at the biological target site. Such chemicals can either elicit similar responses by a common or similar mode of action, or they act independently and may have different endpoints and/or different target organs.

Both concepts have been suggested as default approaches in regulatory risk assessment of chemical mixtures.

Independent action (response addition, effects addition) occurs if chemicals act independently from each other, usually through different modes of action that do not influence each other.

Dose/concentration addition (similar action, similar joint action) occurs if chemicals in a mixture act by the same mechanism/mode of action, and differ only in their potencies. In principle, doses or concentrations of the single components are added after being multiplied by a scaling factor that accounts for differences in the potency of the individual substances.

The effect of a mixture of similarly acting compounds is equivalent to the effects of the sum of the potency-corrected (adjusted) doses/concentrations of each compound.

Specific models – This methodology has the aim to use QSAR models developed on a basis of set of chemicals (mixtures) for purposes of mixture toxicity prediction. This section is under development.

Based these methodologies for handling set of chemicals, different ways for handling of tautomers, mixtures and metabolites are developed.

Caution 

In case the gap filling is entered with set of chemicals with undefined quantities of the components equimolar quantities for all components are assumed for the gap filling calculations.

-0-

QSAR Toolbox v.4

D.5.7.1. Example: Quantitative mixtures toxicity prediction

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures](#)/D.5.7.1. Example: Quantitative mixtures toxicity prediction

D.5.7.1. Example: Quantitative mixture toxicity prediction

Defined mixtures are handled as part of the structure multiplication of parent chemical.

Three new options for prediction of mixtures are available based on the mode of action of the constituents:

- *Acting Independently* (with different mode of action)
- *Acting Similarly* (with same mode of action)

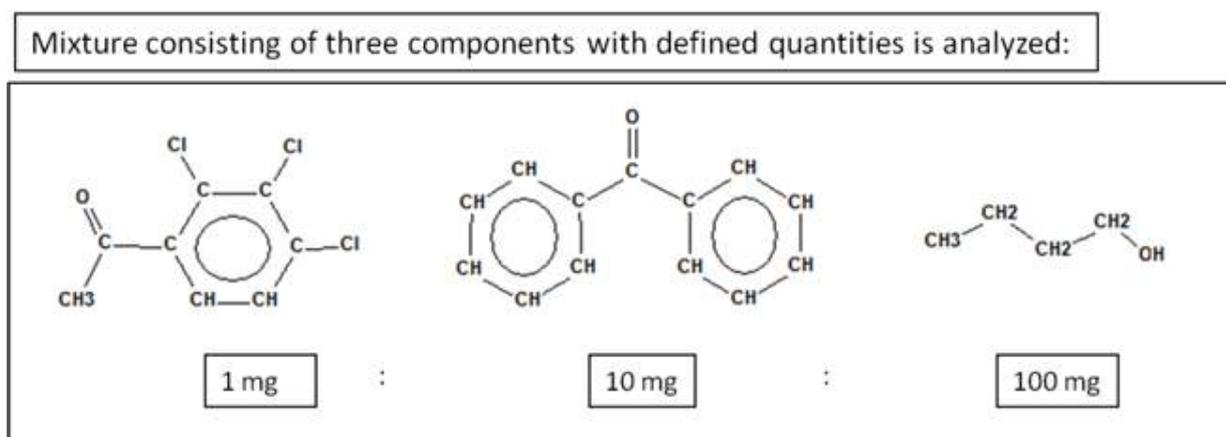


Figure 1

Predicting aquatic toxicity of mixtures:

The procedure is described by below:

- *Input of a mixture.* See [Input section](#) about instructions how to draw a mixture and input the quantity of its constituents in Toolbox 4.0. Once the mixture is drawn in it is displayed on the data matrix (1) (Figure 2). In order to predict the mixture, first it has to be decompose: right-click on the substance in the document tree(2), then select

Multiplication (3), then Decompose (4).

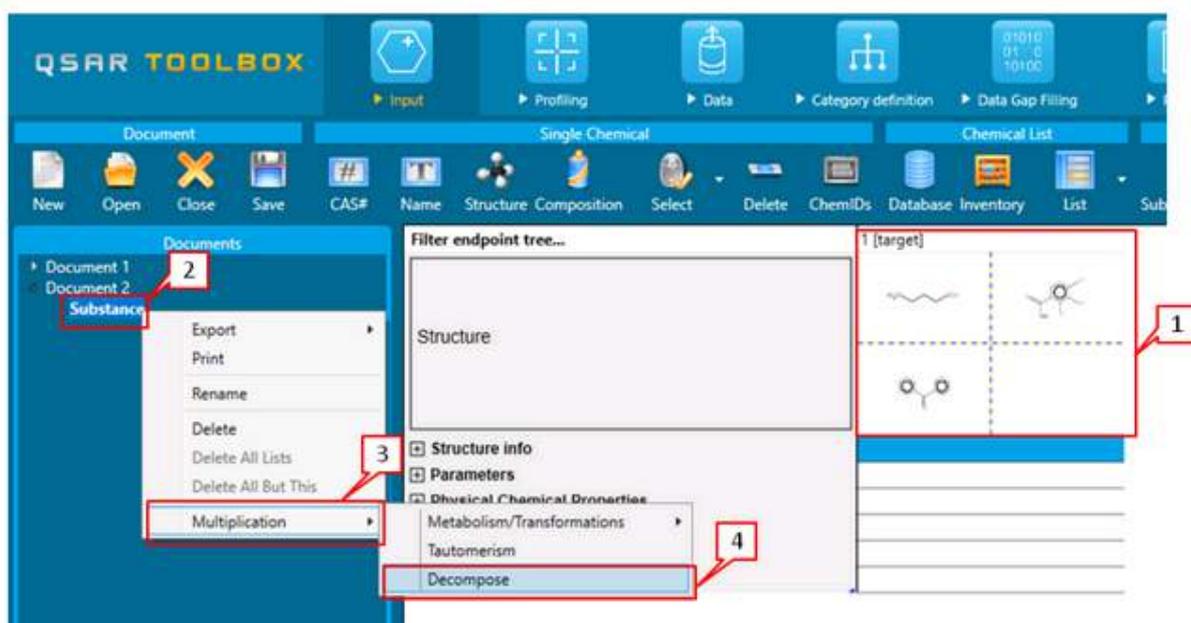


Figure 2

A message is displayed, informing the user about the number of components that are being generated. The user has to press OK. (Figure 3)

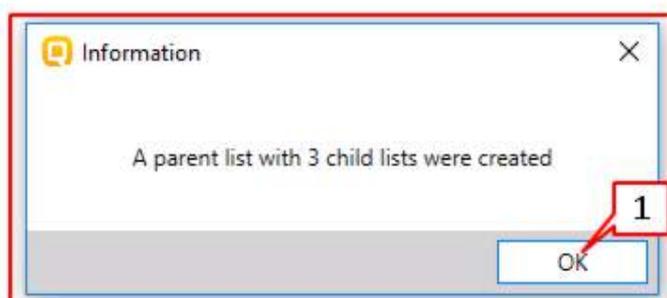


Figure 3

The mixture and its three components are shown on the data matrix (Figure 4):

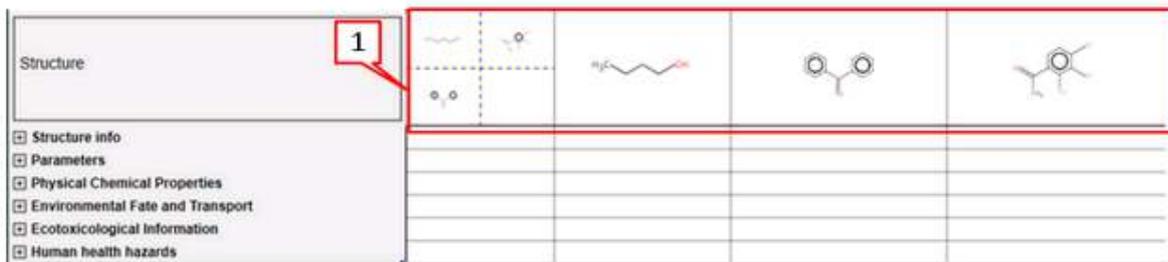


Figure 4

- *Profiling the components of the mixture*: go to Profiling panel (1), select the following profilers: US-EPA New chemical category, Acute aquatic toxicity MOA by OASIS, Acute toxicity classification by ECOSAR (2), then press Apply (3). As it can be seen, all components have the same mode (3). (Figure 5)

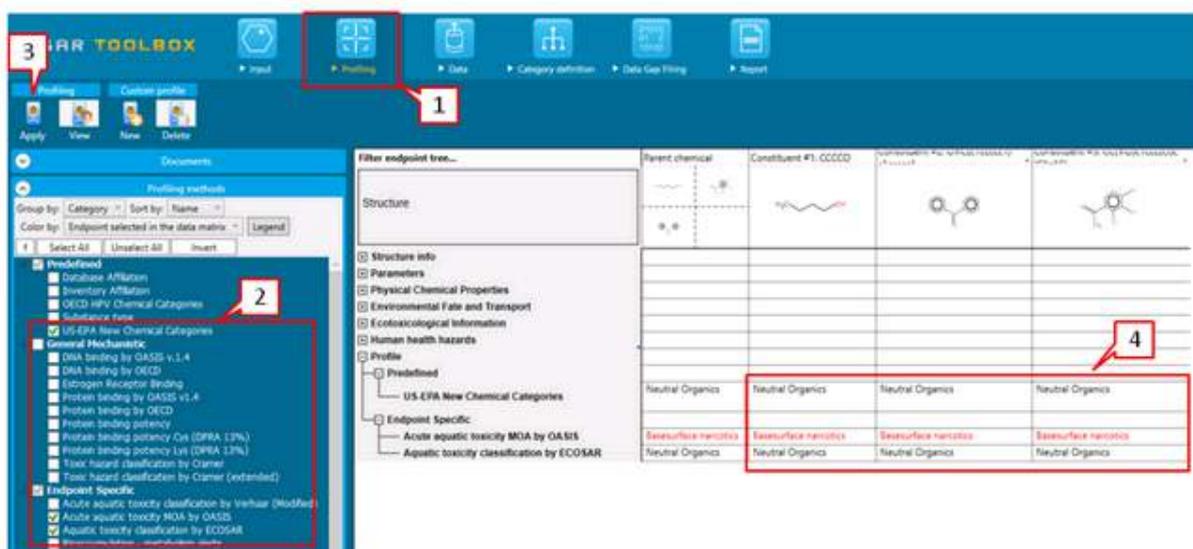


Figure 5

- *Gathering experimental data* : Go to Data (1), select the ecotoxicological databases (2), press Gather (3). A pop-up window appears where the user has to select the ecotoxicological data only and then to click OK. (Figure 6)

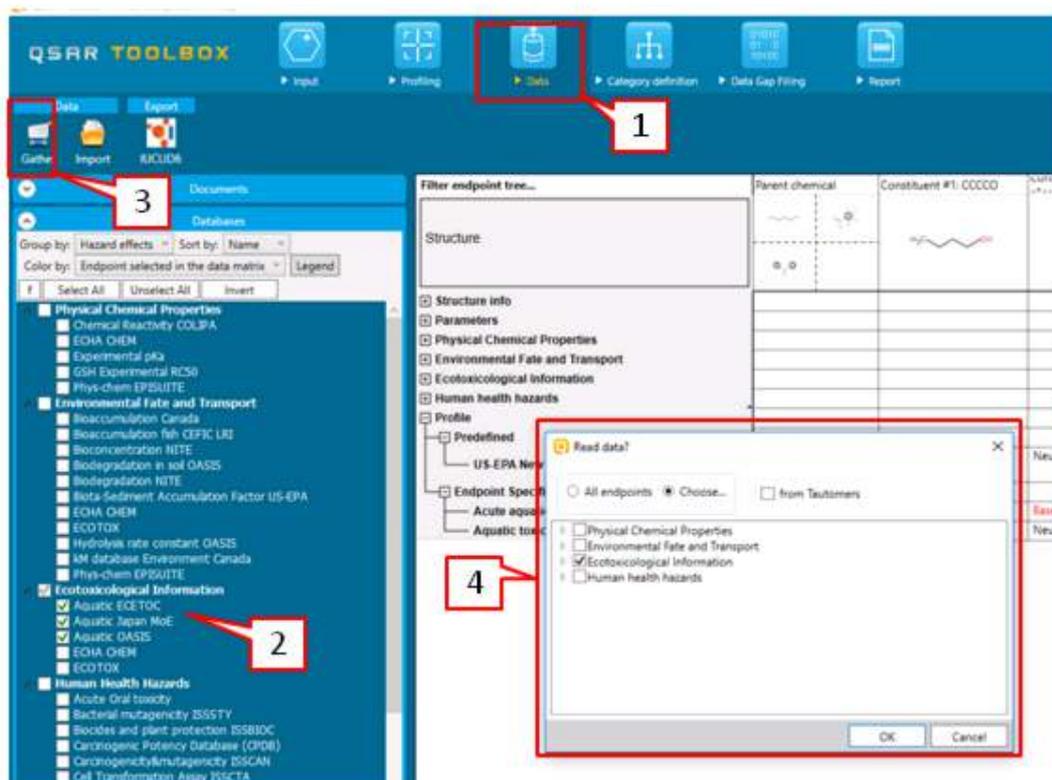


Figure 6

- *Data gap filling approach using Similar mode of action:* In this particular case study Similar mode of action is applied for calculation purposes based on the fact that the investigated mixture has defined quantities and all component have same mode of action. The user has to go to Data gap Filling module (1), click on the cell related to LC 50, 96h, Pimephales promelas (2) for the mixture and then select Similar MOA (3) (Figure 7).

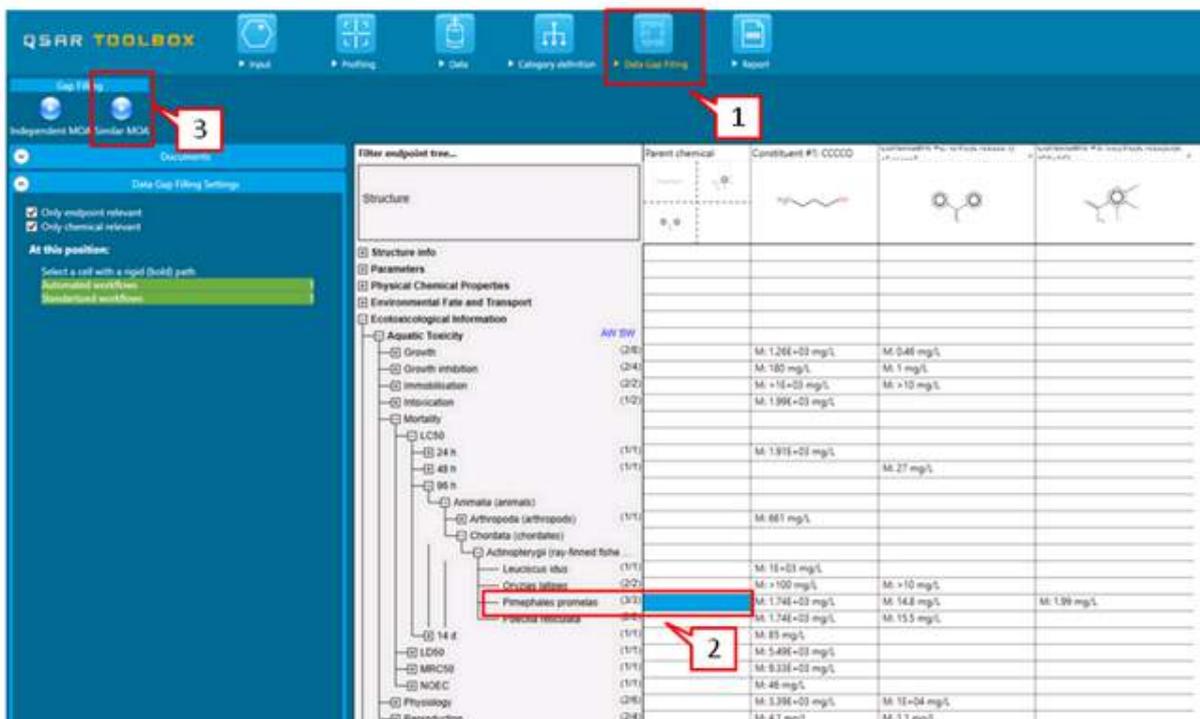


Figure 7

The prediction result accounts for quantities of each component and uses dose concentration calculation for prediction of LC 50(1) (Figure 8). Then the user can accept the prediction (2).

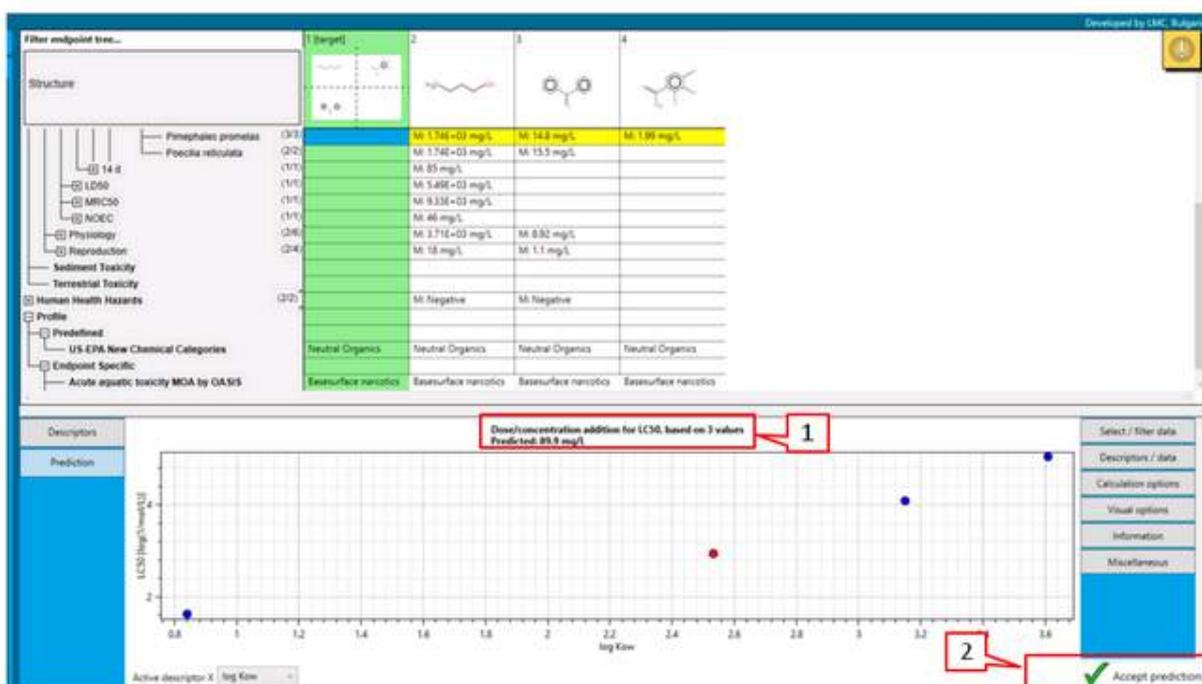


Figure 8

The prediction is displayed on the data matrix (Figure 9)

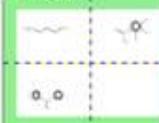
Filter endpoint tree...		1 [target]	2	3	4
Structure					
48 h	(1/1)			M: 27 mg/L	
96 h					
Animalia (animals)					
Arthropoda (arthropods)	(1/1)		M: 661 mg/L		
Chordata (chordates)					
Actinopterygii (ray-finned fishes)					
Leuciscus idus	(1/1)		M: 1E+03 mg/L		
Oryzias latipes	(2/2)		M: >100 mg/L	M: >10 mg/L	
Pimephales promelas	(4/4)	SMOA: 89.9 mg/L	M: 1.74E+03 mg/L	M: 14.8 mg/L	M: 1.99 mg/L
Poecilia reticulata	(2/2)		M: 1.74E+03 mg/L	M: 15.5 mg/L	
14 d					
LD50	(1/1)		M: 85 mg/L		
MRC50	(1/1)		M: 5.49E+03 mg/L		
NOEC	(1/1)		M: 9.33E+03 mg/L		
Physiology	(2/6)		M: 46 mg/L		
Reproduction	(2/4)		M: 3.71E+03 mg/L	M: 8.92 mg/L	

Figure 9

- If the user prefers to use *Independent mode of action* the following steps have to be performed: go to Data gap Filling module (1), click on the cell related to LC 50, 96h, Pimephales promelas (2) for the mixture and then select Independent MOA (3) (Figure 10):

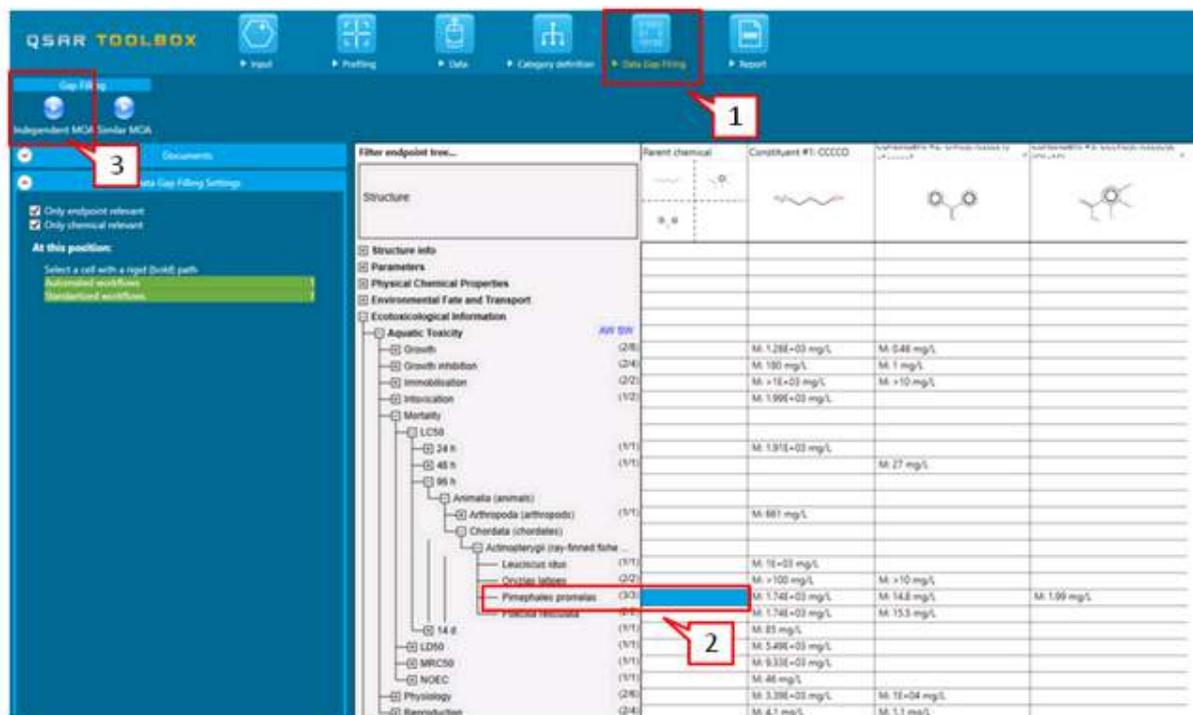


Figure 10

A *Possible data inconsistency* window is displayed where, the user needs to select the unit, that wants to be used in the data gap filling: select the unit (1) than press *OK* (2) (Figure 11).

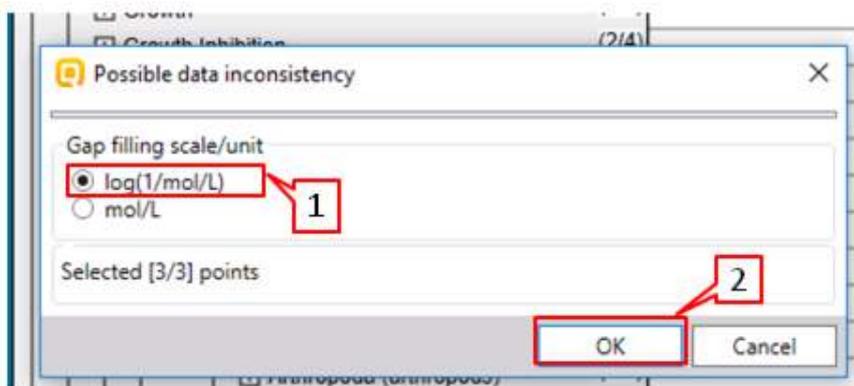


Figure 11

The prediction result (1) is shown above the graph (Figure 12). Then the user can accept the prediction (2).

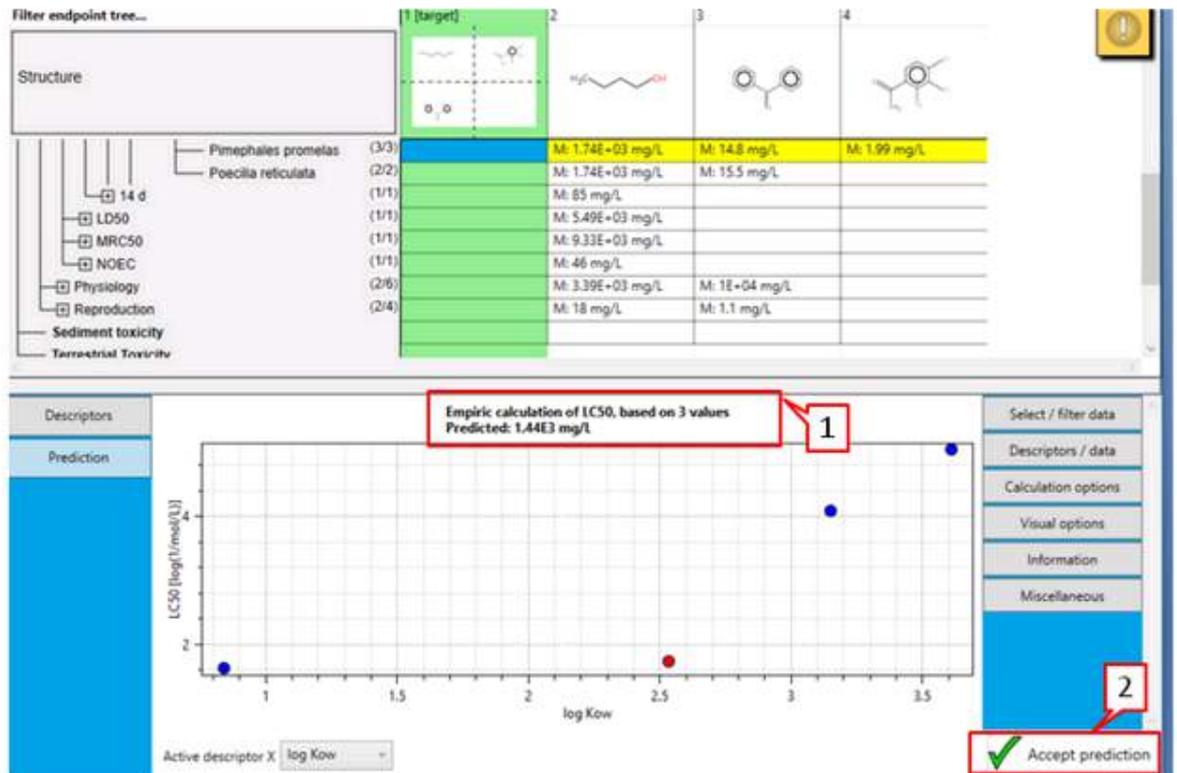


Figure 12

The prediction is displayed (1) on the data matrix (Figure 13):

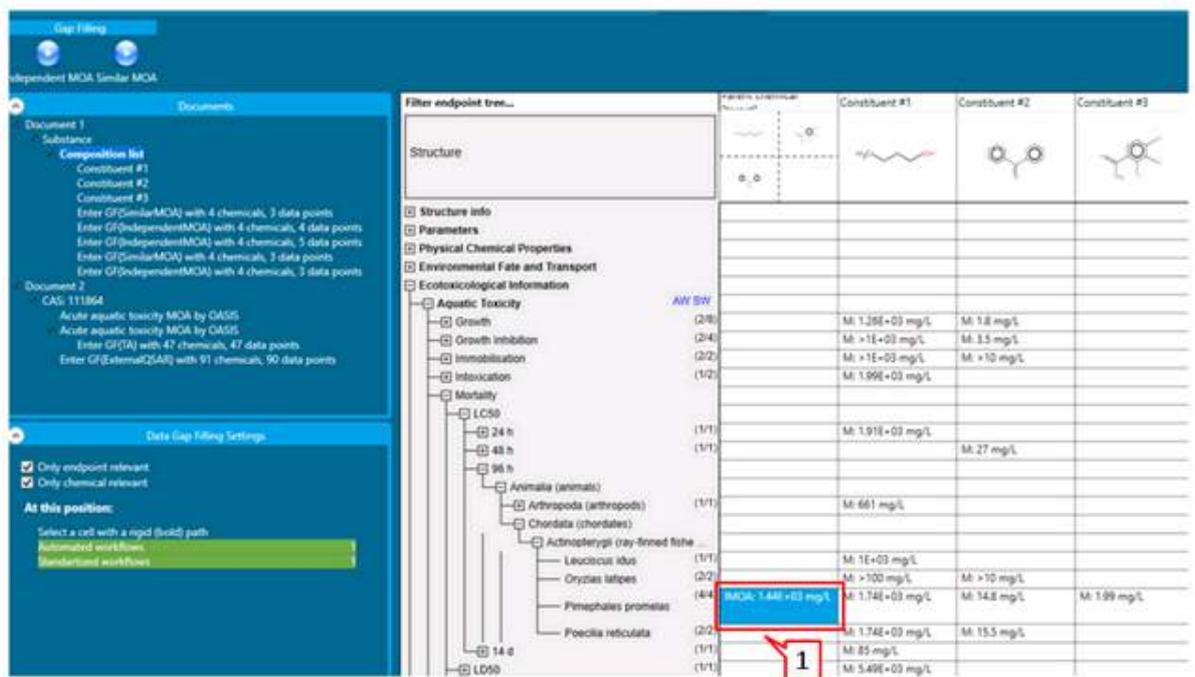


Figure 13

-0-

QSAR Toolbox v.4

D.5.7.1.1. Mixture models

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures/ Quantitative mixtures toxicity prediction/D.5.7.1.1. Mixture models](#)

Mixture models: we are calculating mixtures using two modes of actions:

[Independent MOA](#)

[Similar MOA](#)

-0-

QSAR Toolbox v.4

D.5.7.1.1.1. Independent MOA

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures/ Quantitative mixtures toxicity prediction/ Mixture models /](#)
D.5.7.1.1.1. Independent MOA

Mixture models for components acting by independent MOA: (response addition, effects addition) occurs if chemicals act independently from each other, usually through different modes of action that do not influence each other.

Assumption – combined effect can be calculated from the effects caused by the individual mixture components by following the statistical concept of independent random events

Mixture response:

$$E(C_{Mtx}) = 1 - \prod_{i=1}^N [1 - E(C_i)]$$

$E(C_{Mtx})$ - the effect provoked by the total mixture

$E(C_i)$ - the effects that the individual components would cause if applied singly at that concentration at which they are present in the mixture

Problem - dose-response relationships are practically unknown

Approximation type - based on the undefined dose-response relationships we are using the following approximating types (fig. 1):

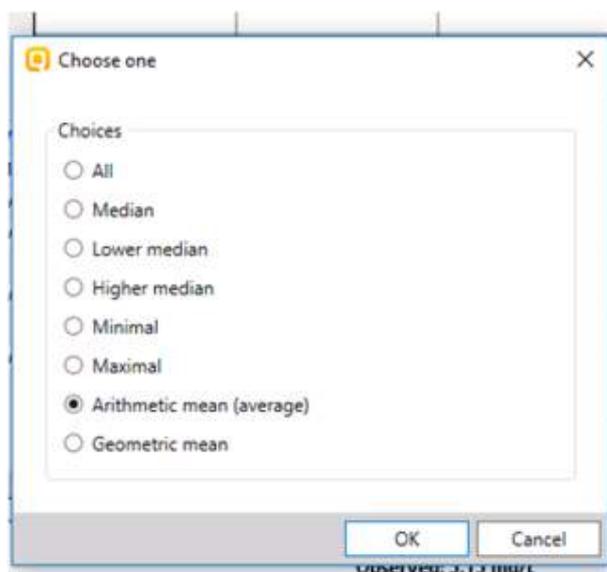


Figure 1

-0-

QSAR Toolbox v.4

D.5.7.1.1.1.1. Illustration

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures/ Quantitative mixtures toxicity prediction/ Mixture models / Independent MOA/D.5.7.1.1.1.1. Illustration](#)

Mixture models for components acting by independent MOA:

Illustration:

$$y_{Mixture} = f(y_j, j = 1, 2, 3, \dots)$$

f- Mean, Mode, Median, Max, Min, etc.

Arithmetic

$$y_{Mixture} = \sum_{n=1}^N x_n y_n$$

Geometric

$$y_{Mixture} = \prod_{n=1}^N y_n^{x_n}$$

Harmonic

$$y_{Mixture} = \frac{\sum_{n=1}^N x_n}{\sum_{n=1}^N \frac{x_n}{y_n}}$$

-0-

QSAR Toolbox v.4

D.5.7.1.1.2. Similar MOA

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures/ Quantitative mixtures toxicity prediction/ Mixture models / D.5.7.1.1.2. Similar MOA](#)

Mixture models for components acting by similar MOA

Assumption – components in a mixture contribute to the joint effect, in proportion to their prevalence and individual potency

- Components act at the same target site
- Components act by the same mechanism
- Components have similar effect (rather than mechanism)

$$RPF_j^{(i)} = \frac{ED_{resp}^{(i)}}{ED_{resp}^{(j)}}$$

Relative potency factor -

i – index (reference) chemical

ED_{resp} – dose (concentration) of a chemical that cause a specified response (fraction of animals that respond, fractional change in a measured physiological value, etc.)

Chemical Equivalent Dose (Concentration) -

$$CED_j^{(i)} = RPF_j^{(i)} d_j$$

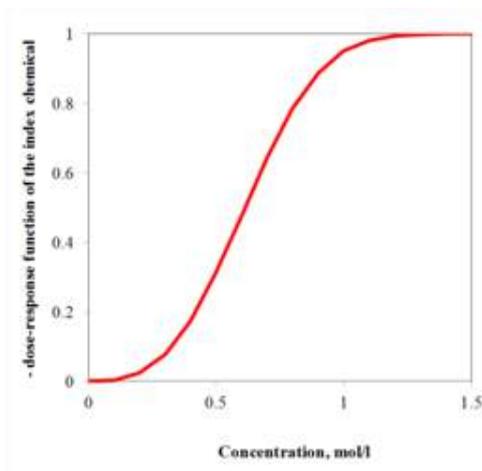
Dose (concentration) of the reference chemical i that will cause the same effect as chemical j at dose (concentration) d_j

Index Chemical Equivalent Dose (Concentration) -

$$ICED = \sum_{j=1}^J CED_j^{(i)} = \sum_{j=1}^J RPF_j^{(i)} d_j$$

Equivalent dose (concentration) of the reference chemical / that will cause the same effect as the mixture

Toxic effect of mixture - response (fraction of animals that respond, fractional change in a measured physiological value, etc.) as a result of exposure to mixture



$$Effect^{Mixture} = f_i(ICED)$$

f_i - dose-response function of the index chemical

-0-

QSAR Toolbox v.4

D.5.7.1.1.2.1. Illustration

[Workflow / Filling data gaps/ Data Gap Filling approaches using different modes of handling chemical structures/ Quantitative mixtures toxicity prediction/ Mixture models / Similar MOA /D.5.7.1.1.2.1. Illustration](#)

Mixture models for components acting by similar MOA

Illustration:

Reference chemical: Component 1 ($i = 1$)

Fig.1.

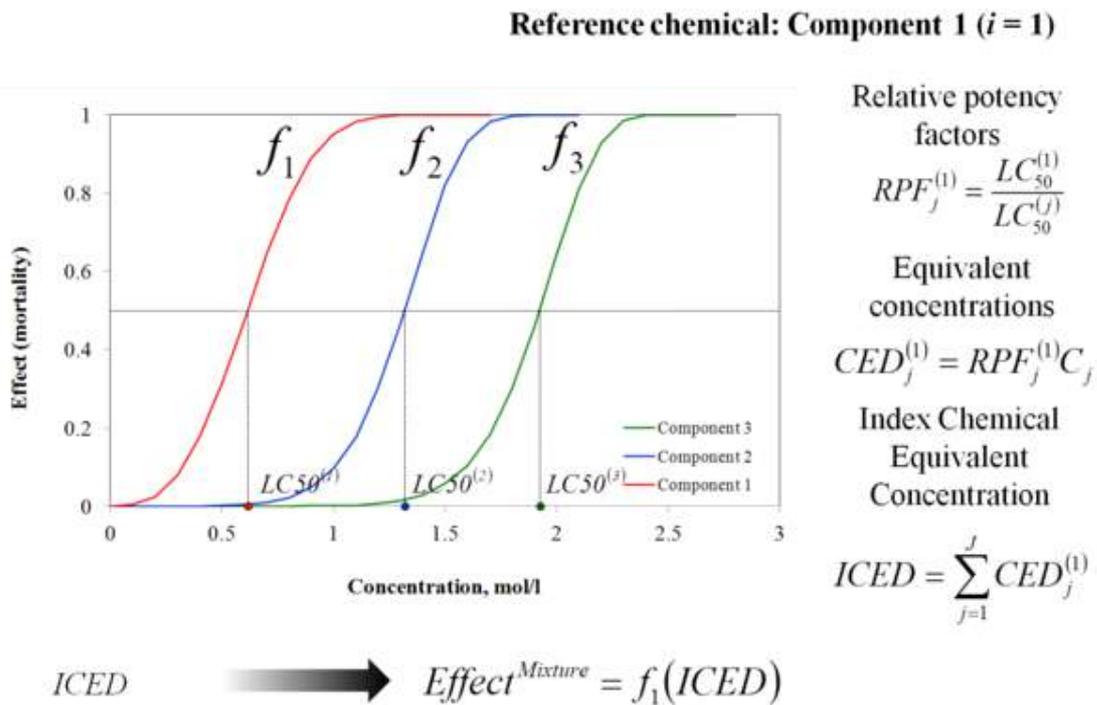
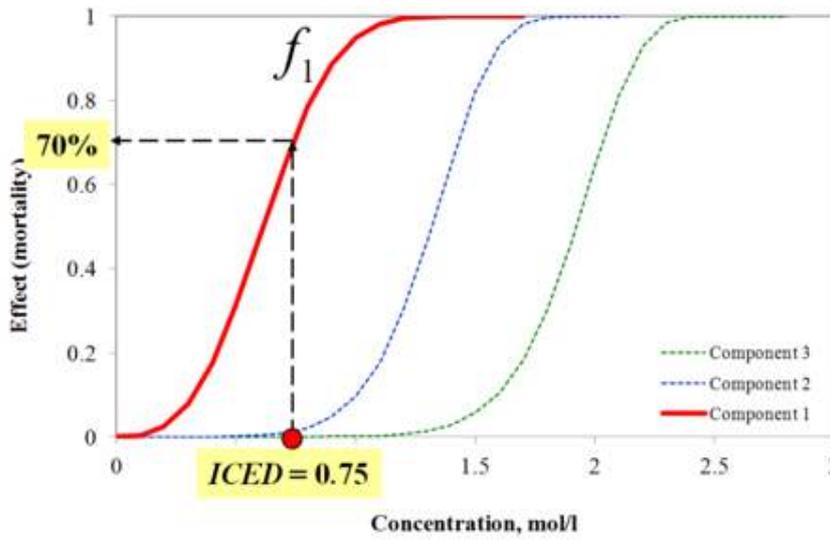


Fig.2

Reference chemical: Component 1 ($i = 1$)



Relative potency factors

$$RPF_j^{(1)} = \frac{LC_{50}^{(1)}}{LC_{50}^{(j)}}$$

Equivalent concentrations

$$CED_j^{(1)} = RPF_j^{(1)} C_j$$

Index Chemical Equivalent Concentration

$$ICED = \sum_{j=1}^J CED_j^{(1)}$$

$$ICED = 0.75 \longrightarrow Effect^{Mixture} = f_i(ICED) \approx 70\% \text{ mortality}$$

-o-

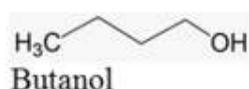
QSAR Toolbox v.4

D.5.7.1.1.2.2. Example

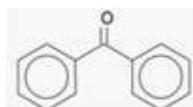
[Workflow](#) / [Filling data gaps](#) / [Data Gap Filling approaches using different modes of handling chemical structures](#) / [Quantitative mixtures toxicity prediction](#) / [Mixture models](#) / [Similar MOA](#) / D.5.7.1.1.2.2. Example

Mixture models for components acting by similar MOA

Example

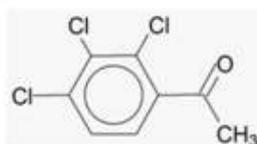
Three component mixture**Weight composition: $w_1 : w_2 : w_3 = 100 : 10 : 1$, kg**

$LC_{50}^{(1)} = 1740 \text{ mg/l} = 0.023 \text{ mol/l}$; $MW = 74 \text{ Da}$
90 w% ~ 95.8 mol%



Benzophenone

$LC_{50}^{(2)} = 15 \text{ mg/l} = 8.32 \times 10^{-5} \text{ mol/l}$; $MW = 182 \text{ Da}$
9 w% ~ 3.9 mol%

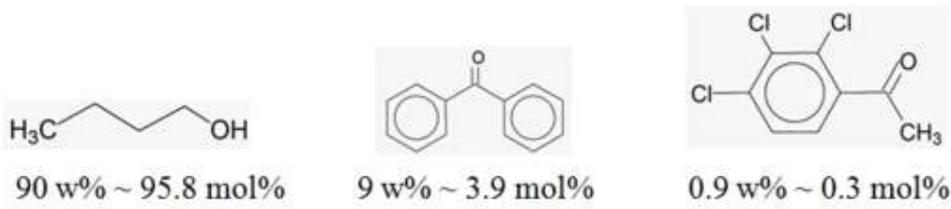


2',3',4'-Trichloroacetophenone

$LC_{50}^{(3)} = 1.99 \text{ mg/l} = 8.91 \times 10^{-6} \text{ mol/l}$; $MW = 223 \text{ Da}$
0.9 w% ~ 0.3 mol%

Figure 1. Compositions of mixture

Acute toxicity of 150 mg/l substance solution?



$$C_1 = 150 \cdot 0.90 = 135.1 \dots \text{ mg/l} = 0.00182 \text{ mol/l}$$

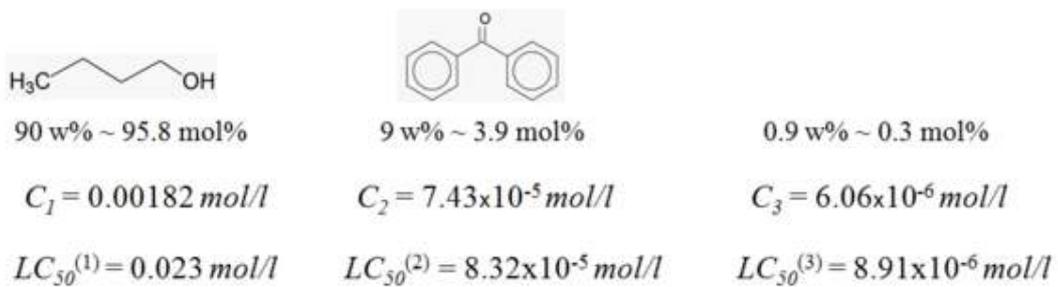
$$C_2 = 150 \cdot 0.09 = 13.51 \dots \text{ mg/l} = 7.43 \times 10^{-5} \text{ mol/l}$$

$$C_3 = 150 \cdot 0.009 = 1.351 \dots \text{ mg/l} = 6.06 \times 10^{-6} \text{ mol/l}$$

Mixture concentration: $C_M = 150 \text{ mg/l} = 0.0019 \text{ mol/l}$

Figure 2. Determine concentrations of components

Acute toxicity of 150 mg/l substance solution?



$$RPF_1^{(1)} = 1$$

$$RPF_2^{(1)} = 282$$

$$RPF_3^{(1)} = 2630$$

Index Chemical Equivalent Concentration $ICED = \sum_{j=1}^3 RPF_j^{(1)} C_j = 0.039 \text{ mol/l}$

$ICED = 0.039$ \longrightarrow $Effect^{Mixture} = f_1(ICED)$ **Unfortunately f_1 is unknown**

$$ICED = 0.039 \text{ mol/l} > LC_{50}^{(1)} = 0.023 \text{ mol/l}$$

the mortality caused by mixture will be > 50%

Figure 3. Calculating RPF (relative potency factors) for each components

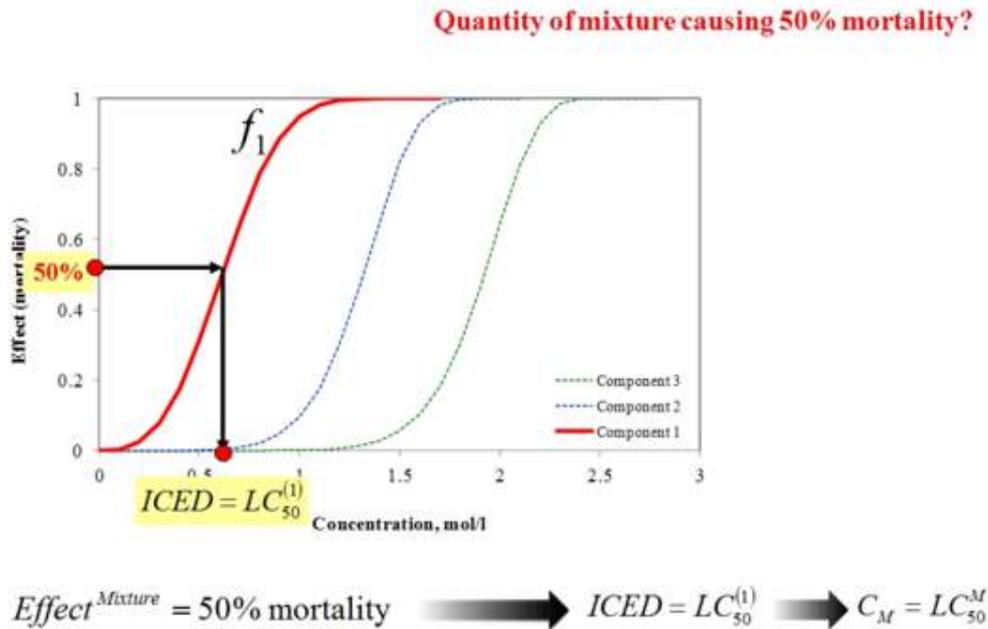


Figure 4. Illustrating formula for calculate quantity of mixture causing 50% mortality

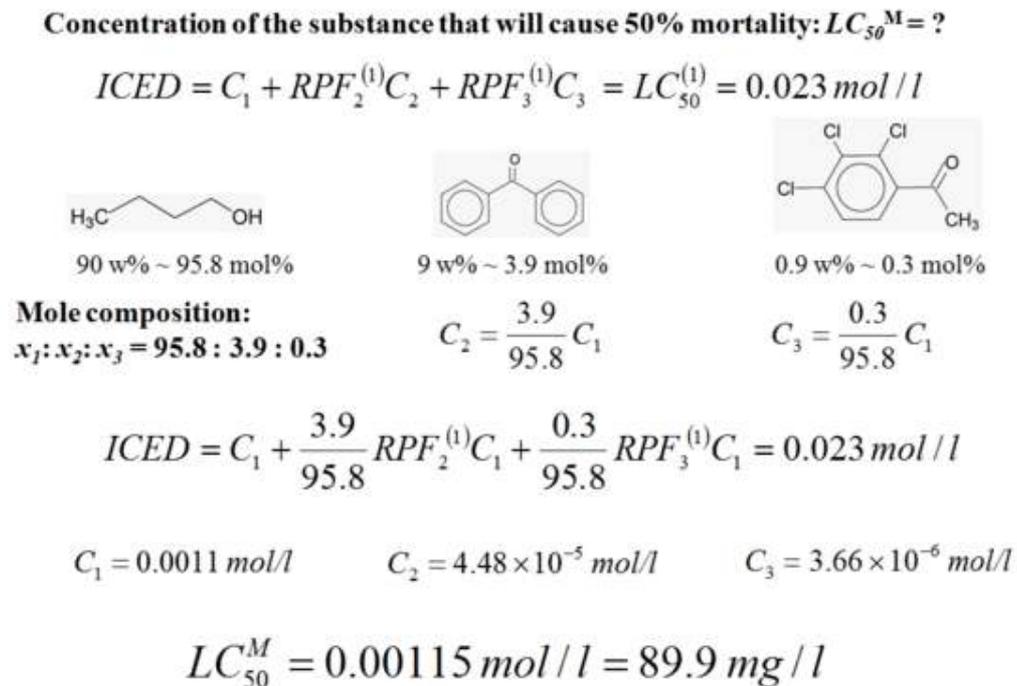


Figure 5. Illustration for calculation of concentration of the substance that will cause 50% mortality

-0-

D.6. Report

QSAR Toolbox v.4

D.6.1. Report window

[Workflow](#) / [Reporting](#) / D.6.1. Report window

D.6.1. Report window

The report section of the toolbar is simplified in QSAR Toolbox v.4. It consists of only one button for generation of a prediction report (1). (Figure 1)



Figure 1

-0-

QSAR Toolbox v.4

D.6.2. Create a report

[Workflow](#) / [Reporting](#) / D.6.2. Create a report

D.6.2. Create a report

The Create button generates reports for the performed prediction for the target chemical and used analogues.

In order to create a report the user needs to select a cell with prediction (1) and click on the Create (2) button. If more than one predicted result is available in the selected cell, a pop-up window appears. The user can choose a prediction (3) and then click on OK button (4). (Figure 1)

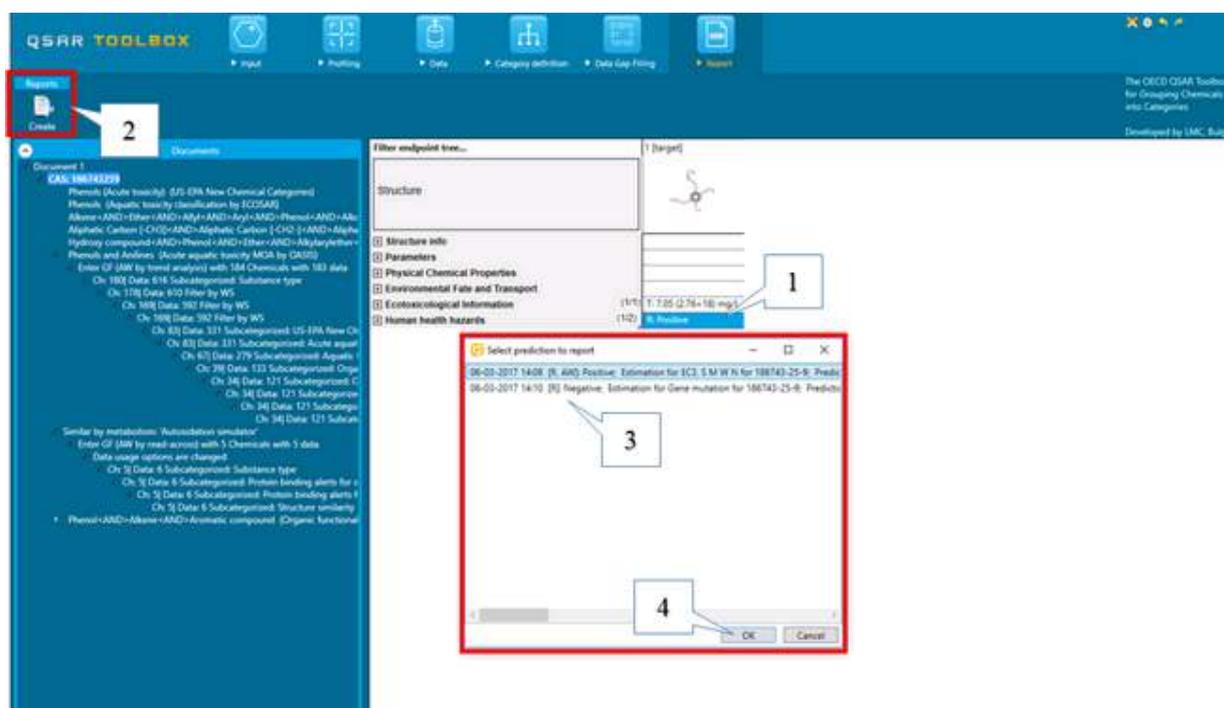


Figure 1

After the prediction is selected the Report wizard appears. The user can customize the sections, which will be included into the report (1) and go through the sections using the *Back* and *Next* buttons (2). (Figure 2)

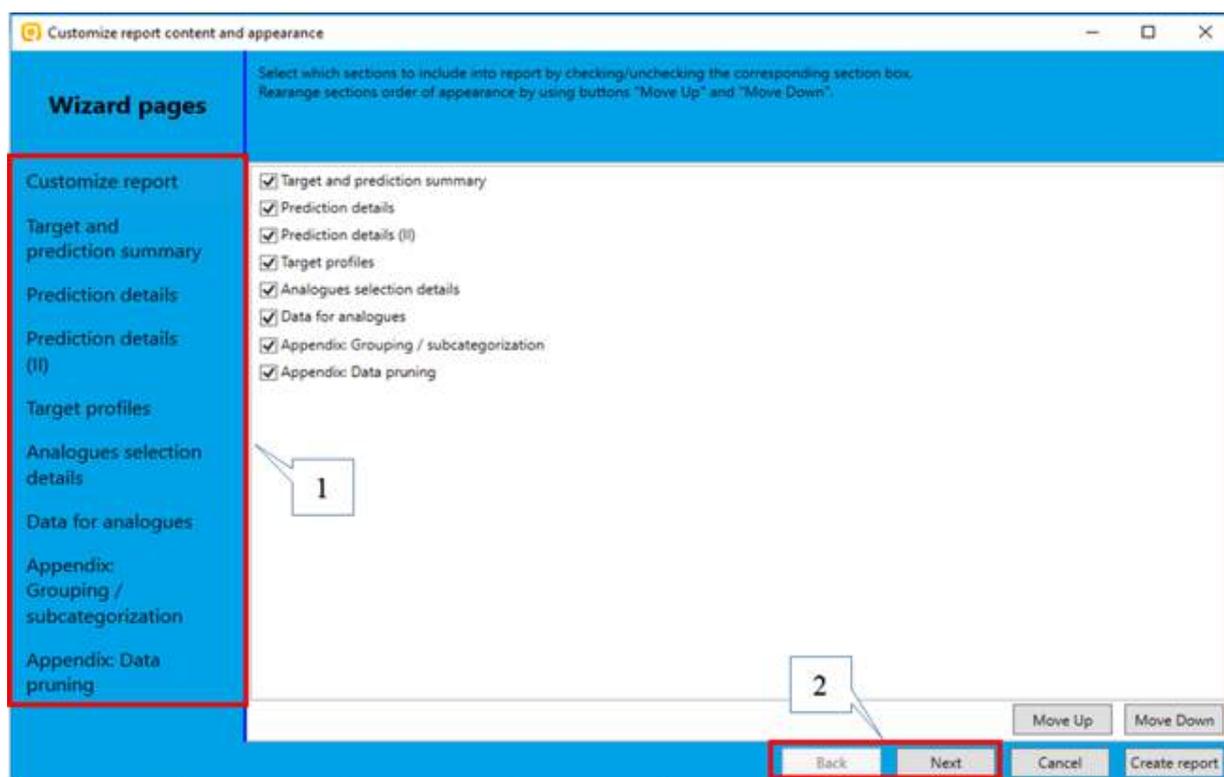


Figure 2

- Section Customized report - the user is able to include or exclude the sections in the report by check/ uncheck (1) the corresponding box. Also order of appearance of sections could be changed by Move Up/Down buttons (2). (Figure 3)

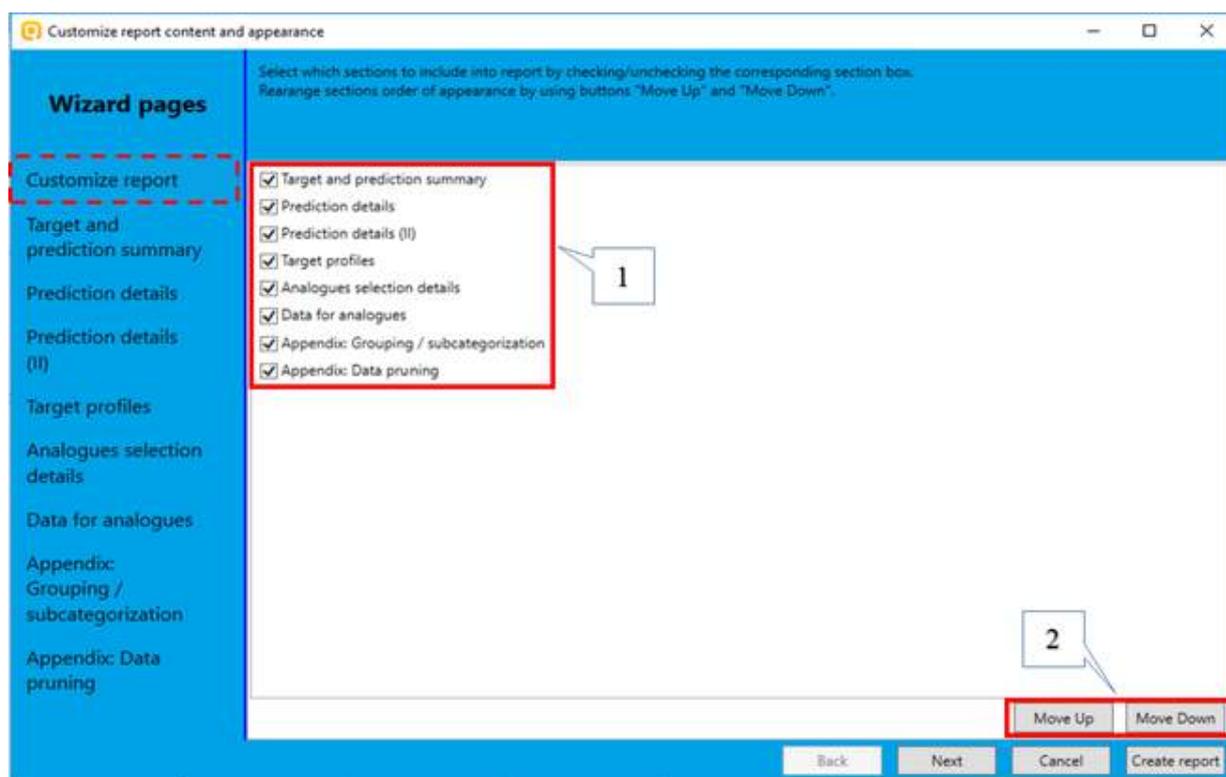


Figure 3

- Section Target and prediction summary – This section includes substance ID of the target chemical and the prediction outcome. Fields which are automatically populated by the system are indicated. Here the user could add information for the author, contact details and summary information (1). Because the target could have a lot of synonyms, additional option related to the maximal count of chemicals names to report is also created. The user can choose to see all available chemical names of the target in the report by selecting of All radio button (2) or to point out the exact number of the names by selecting of Count radio button (3). (Figure 4)

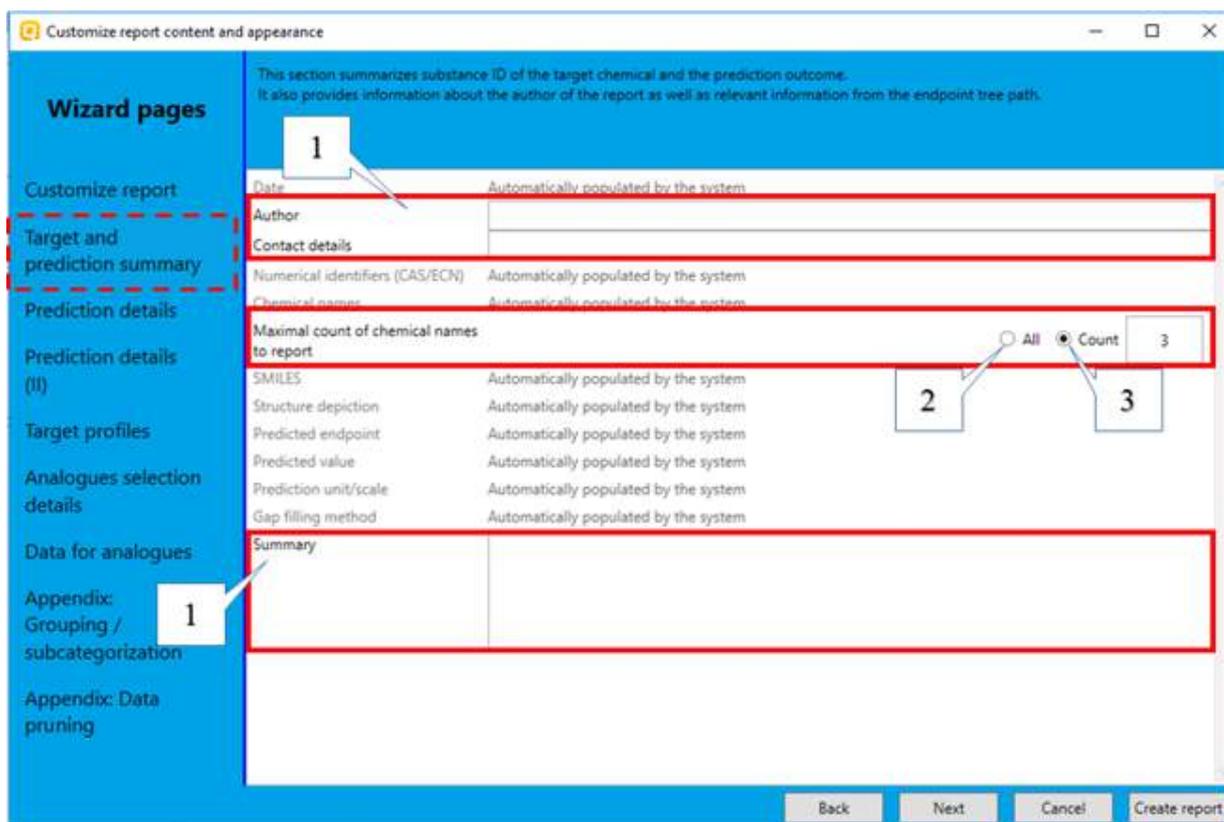


Figure 4

- Sections Prediction details and Prediction details (II) – section *Prediction details* provides details about the prediction and its reliability. *Prediction details (II)* is optional and it provides specific information about the prediction depending on the gap filling approach. The Prediction details information is automatically populated by the system (1), while the Prediction details (II) section consists of fields for manually filling (2). (Figure 5)

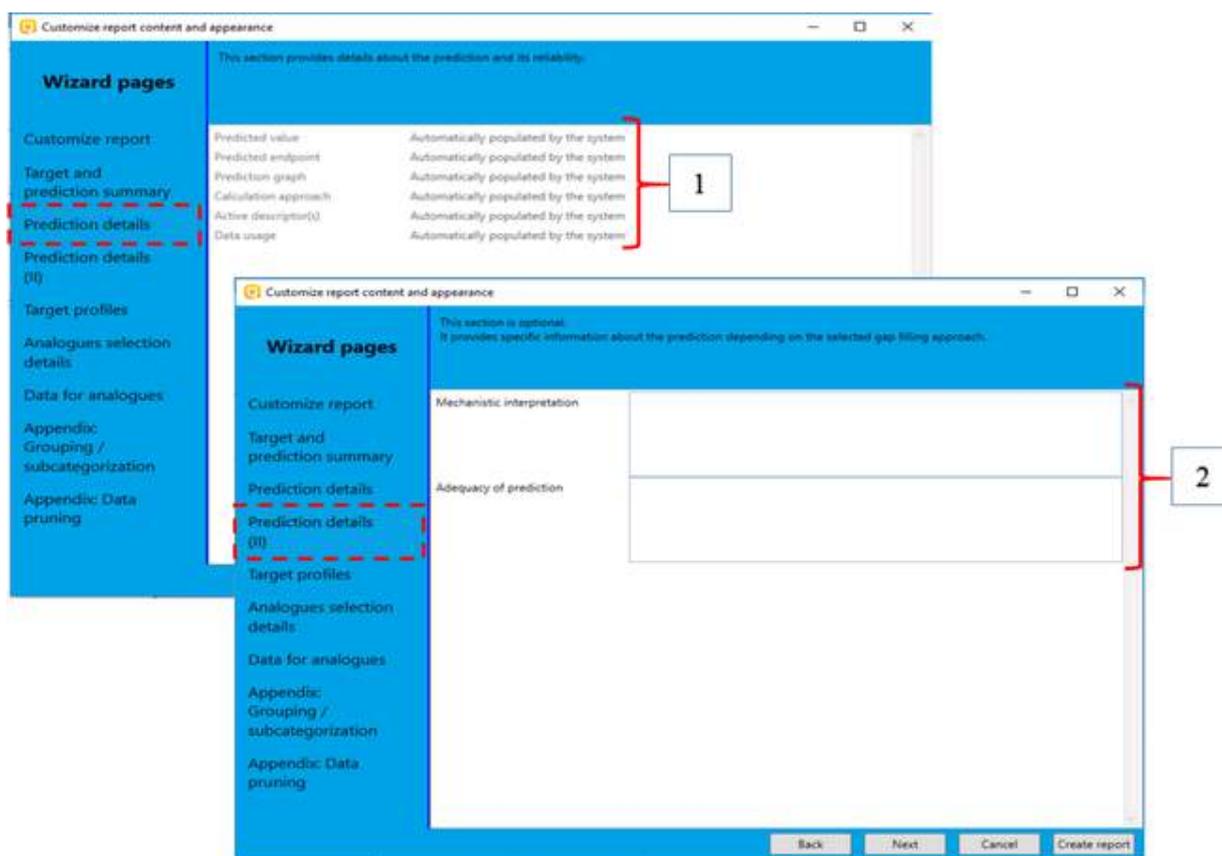


Figure 5

- Section Target profiles – this section summarize profiles used for the prediction. In the first subsection “Profiles” system provides profilers used for primary categorization and subcategorizations in a way they are applied (1). List with profiles could be expanded by Adding new or some could be deleted etc. (2). By default option *Show in report* is checked (3). This is optional setting. Also list with profiles could be *exported in an Excel format* (4). (Figure 6)

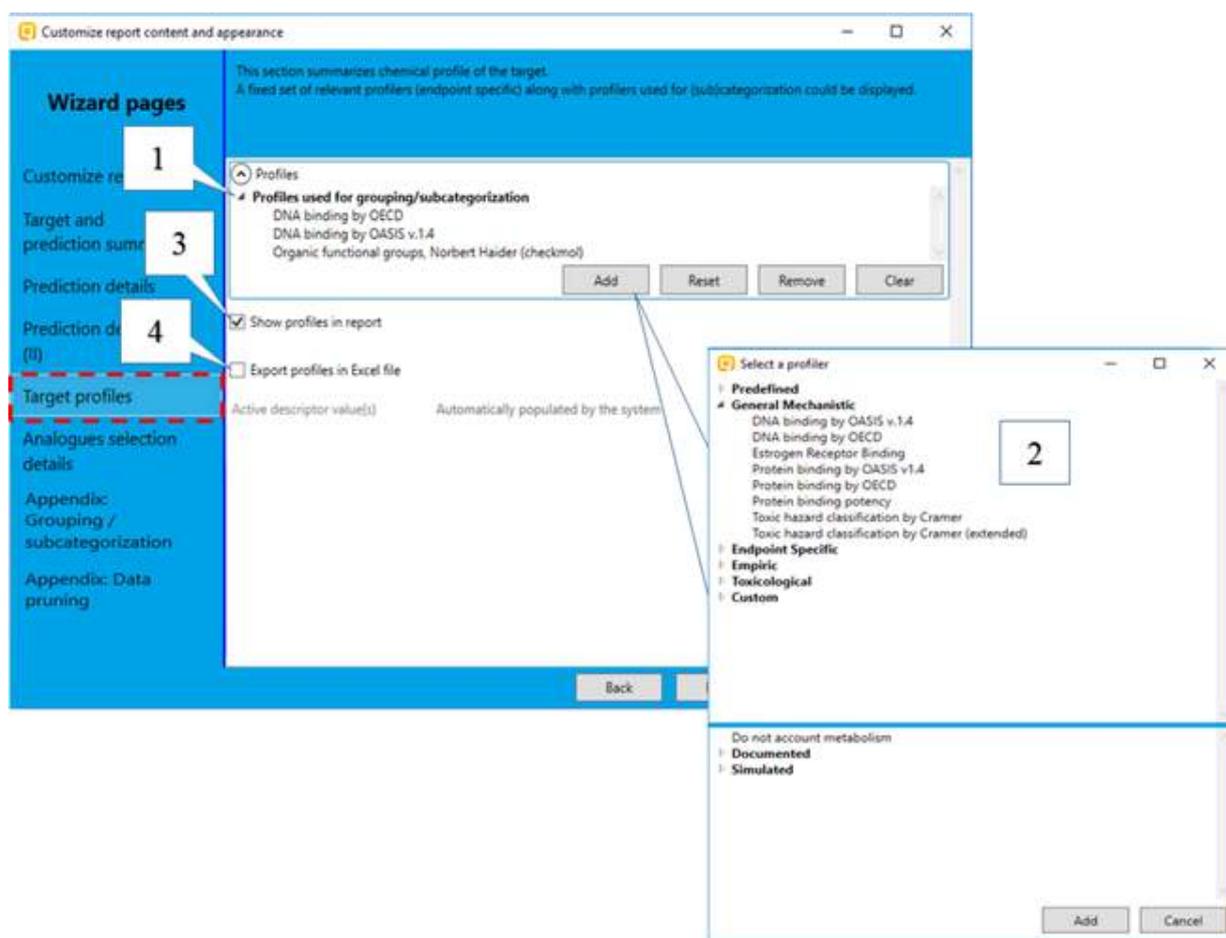


Figure 6

- Section Analogues selection details – This section illustrates how analogues were selected. It displays selected databases, category boundaries and applicability domain. It is given automatically. (Figure 7)

Note:

The whole domain information is under development currently.

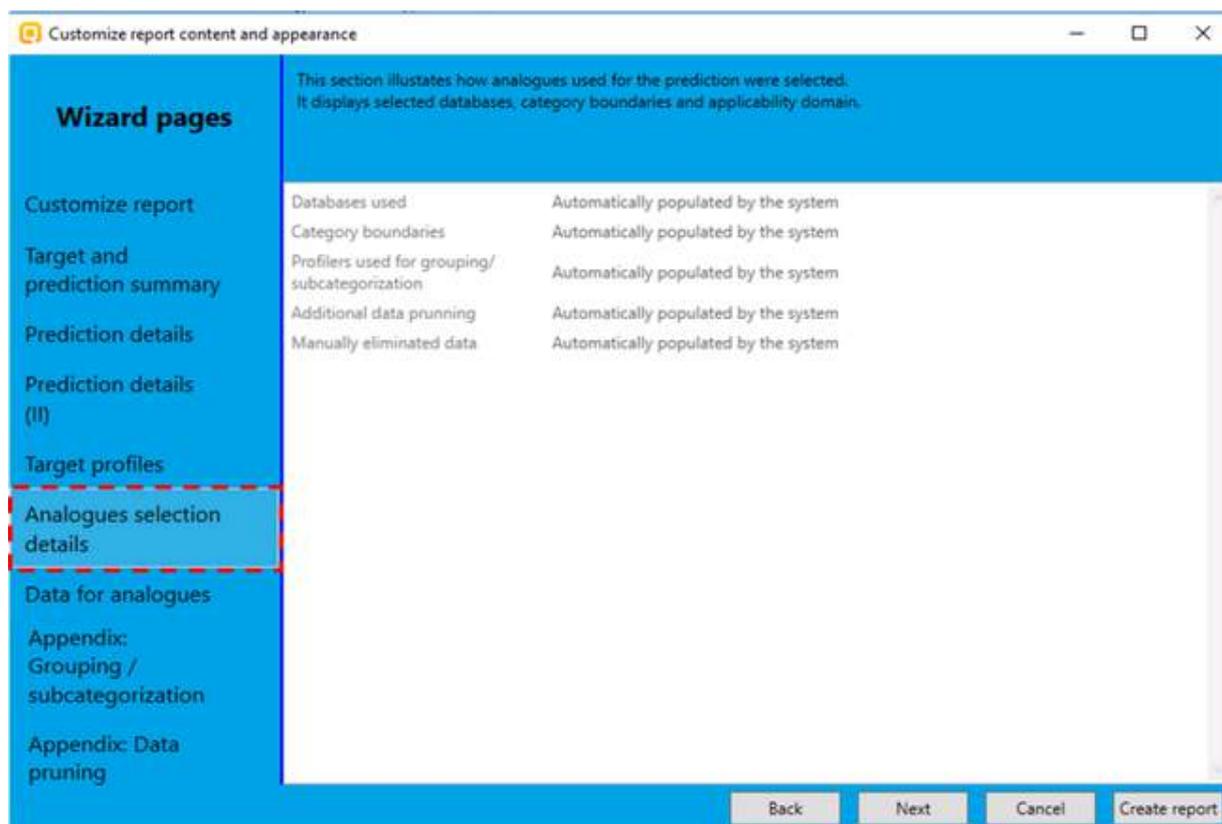


Figure 7

- Section Data for analogues – this section provides details information about the analogues used for obtaining the prediction including parameters, profilers and experimental data. Once opening the respective menu a list with available parameters appears (1). List with profilers used for deriving the prediction are presented in subsection Profiles (2), the list could be added, removed etc. (3). Experimental data for the analogues could be added in the subsection “Include experimental data” (4), by clicking on Add (5) button. The user can select endpoint (6) (e.g. EC3, using filter menu or expand the tree and find the desired endpoint) and they are also able to add metadata fields from the *Metadata field* section (7). Selection of data finishes with clicking on the *OK* button (8). (Figure 8)

Note:

Keep in mind that experimental data for the analogues used for obtaining the prediction are provided automatically by the system.

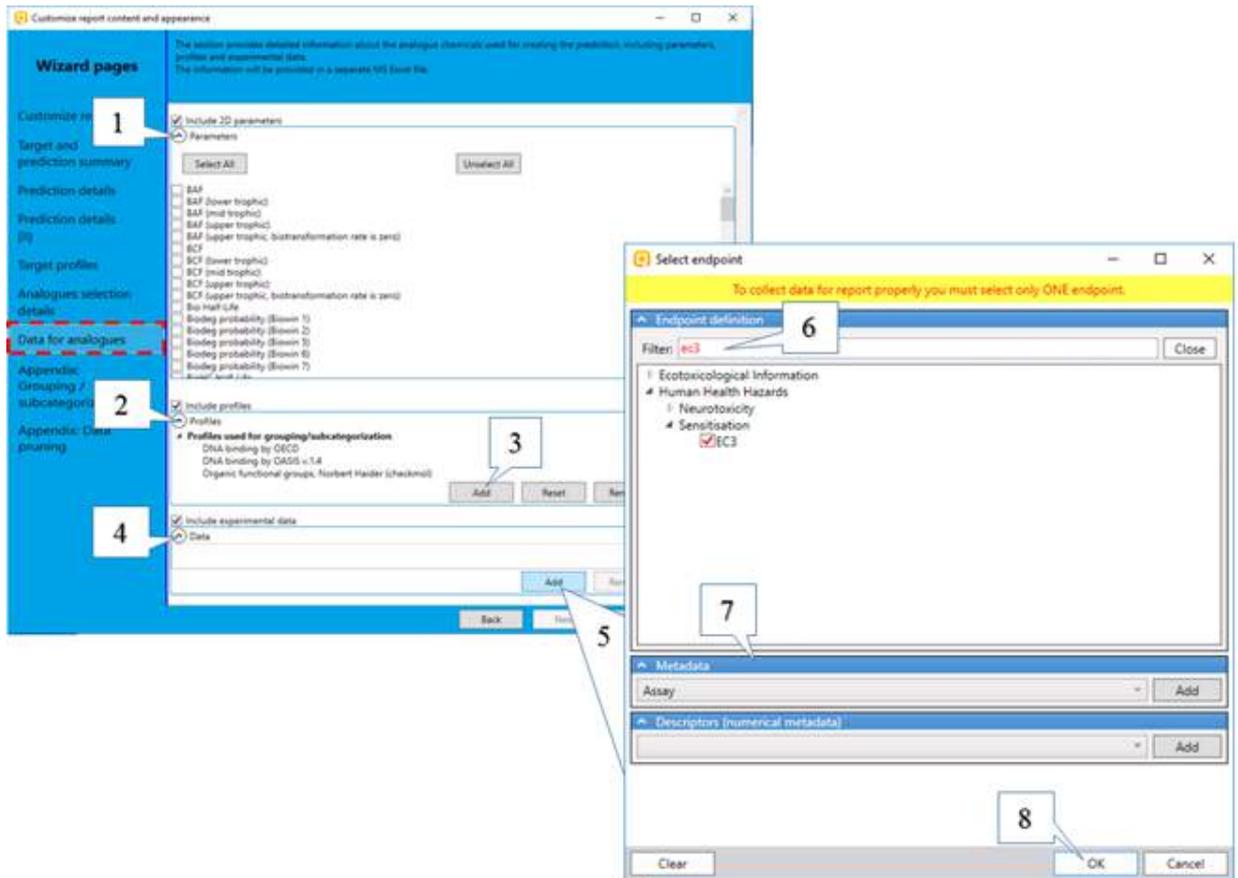


Figure 8

- Sections Appendix: Grouping / subcategorization and Appendix: Data pruning - These sections include the categories used for primary categorization as well as the categories and metadata removed during the subcategorization automatically populated by the system (1-2). The user can provide justification for the manually eliminated data (3). (Figure 9) The Appendix: Data pruning section appears only when manual workflow is executed.

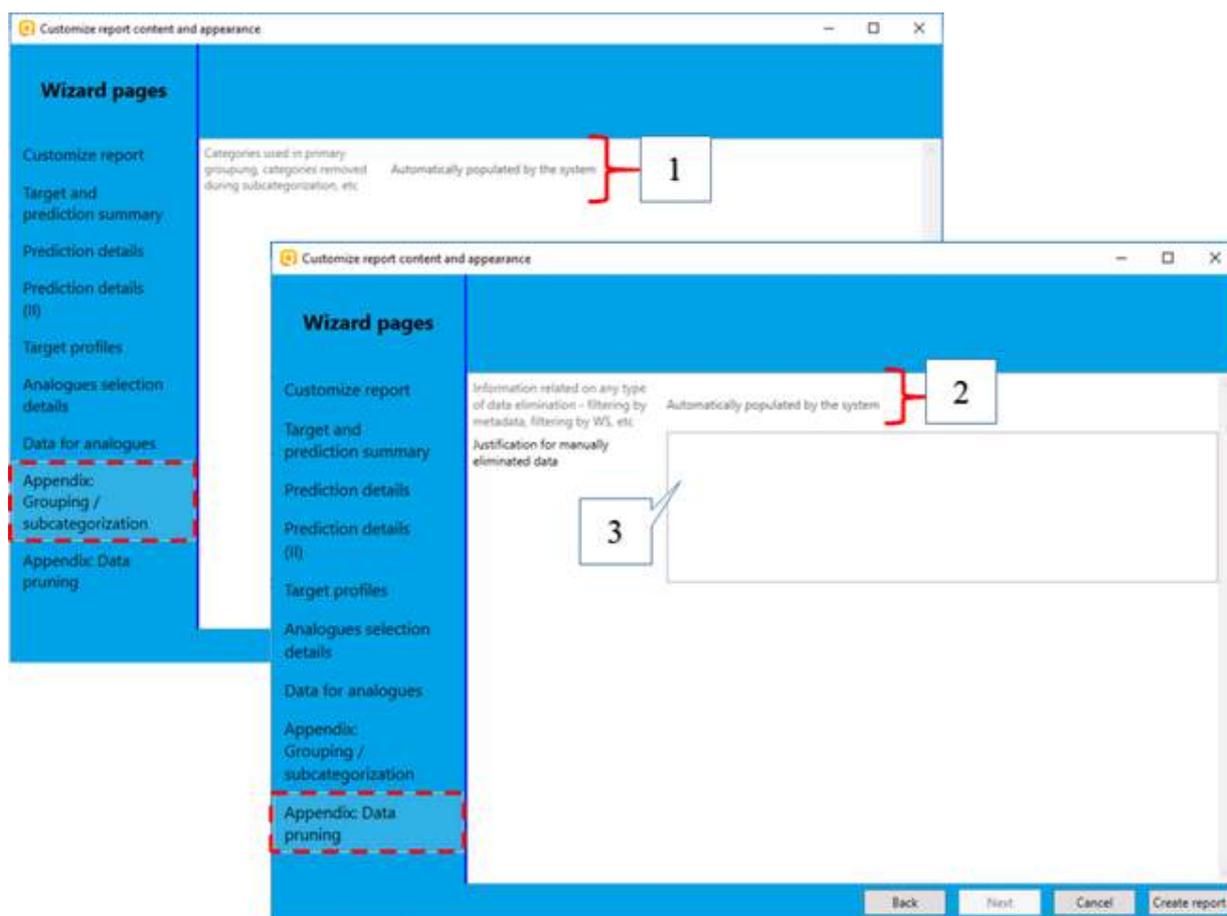


Figure 9

To generate the report click on Create report button (1). Additional window appears with two options: Prediction report or Data matrix report (2). The user can open and/or save the reports (3). (Figure 10)

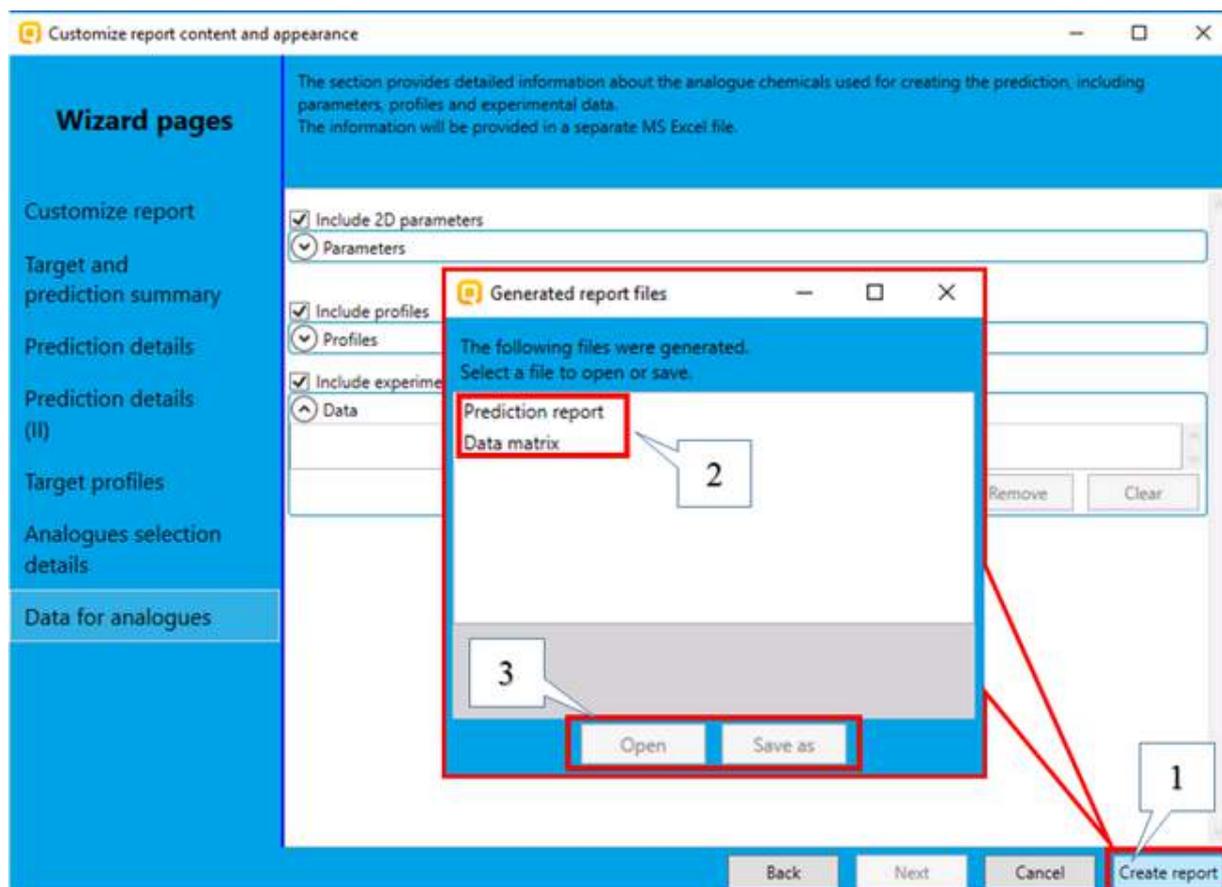


Figure 10

The *Prediction report* (Figure 10a) is a PDF file containing the prediction information related to the target. The *Data matrix* report (Figure 10b) is a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.

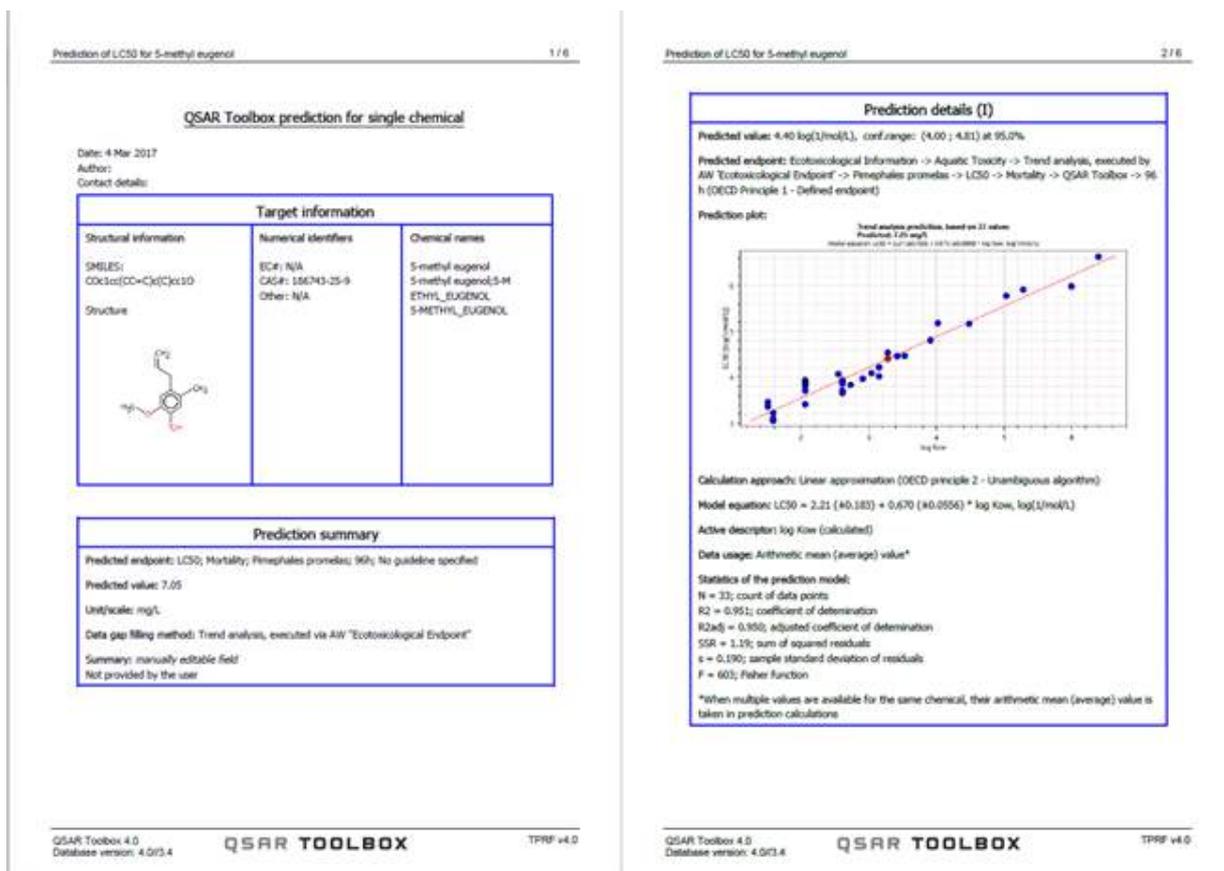


Figure 10a

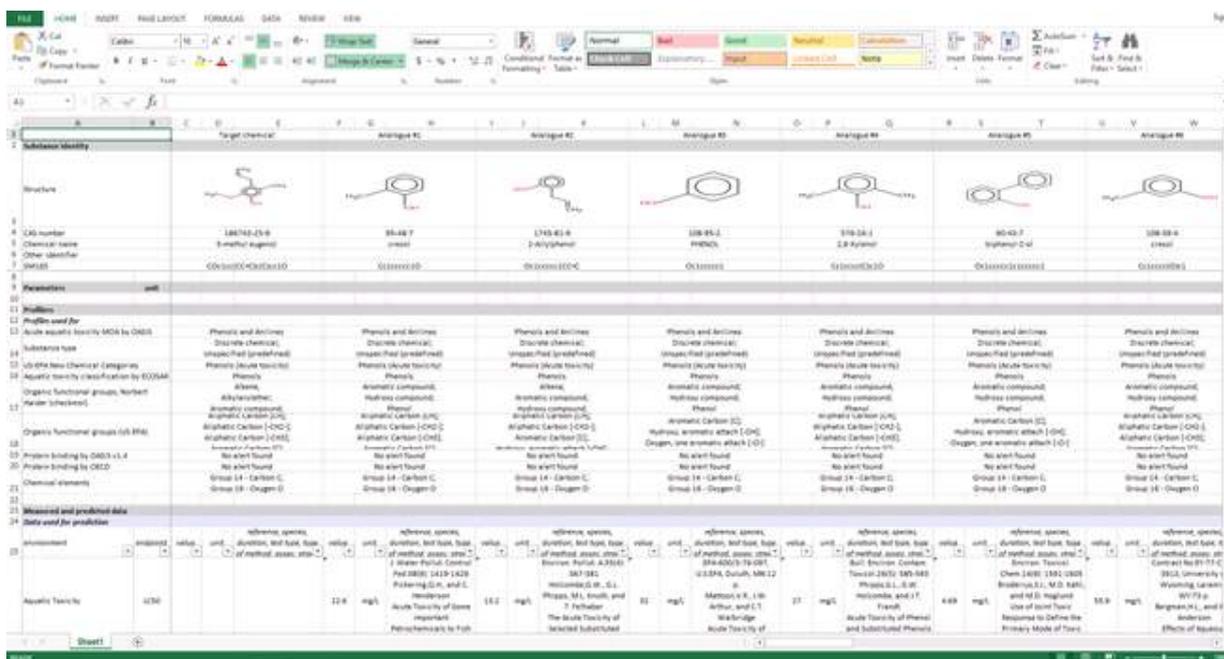


Figure 10b

-0-

Chapter E. Options

QSAR Toolbox v.4

E.1. General

[Options](#) / E.1. General

E.1. General

General settings allows user to modify server settings (local or remote)(1) and calculation settings (2) (Figure 4)

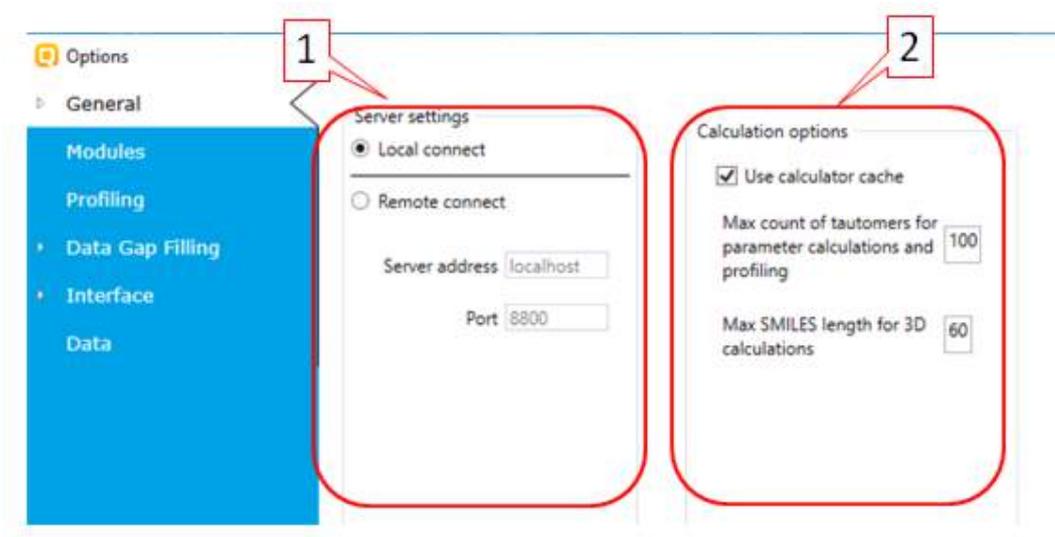


Figure 1

(1) Server settings:

1. If the Local connect option is checked Toolbox with the database (FDB file) stored in the same PC, where the Toolbox is installed. The installation of the database is accomplished during the process of Toolbox installation. Here the user is provided the possibility to change the working FDB file.

2. If the Remote connect option is selected the Toolbox connects to a server to use its database.

3. The user has to type in server address and Port.

(2) Calculation options:

1. The option Use calculator cache provides possibility to perform calculation of 2D and 3D parameters with/without using values calculated in advance and stored in the Toolbox cache. The user should be careful when this option is checked, because the process of calculations could be extremely slow in some cases.

2. Max count of tautomers option is responsible for the number of tautomers generated for the specific structure which are subject to profiling and 2D and 3D calculation. The user could change the threshold for the number of tautomeric forms which will be subject of calculations and profiling.

3. Max SMILES length for 3D calculations gives the user opportunity to modify that threshold.

- Scale editor

Scale editor (2) is visualized by expanding the General (1) options (Figure 2). In the Scale editor panel (3), all scales implements in the Toolbox 4.0 are visualized.

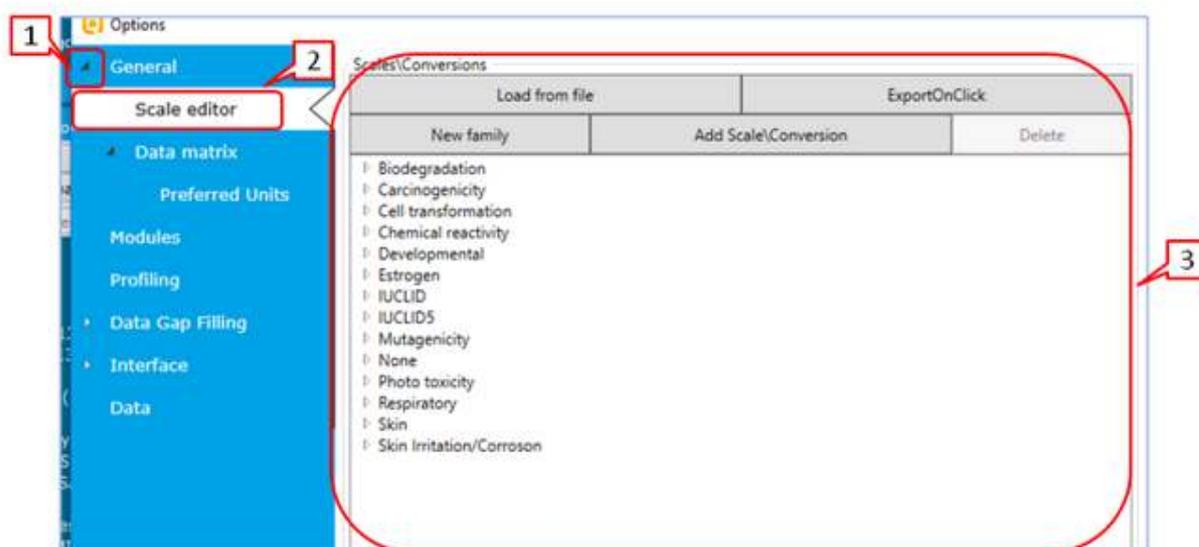


Figure 2

The user is given an option to Add new family scale (1,2) in Figure 3:

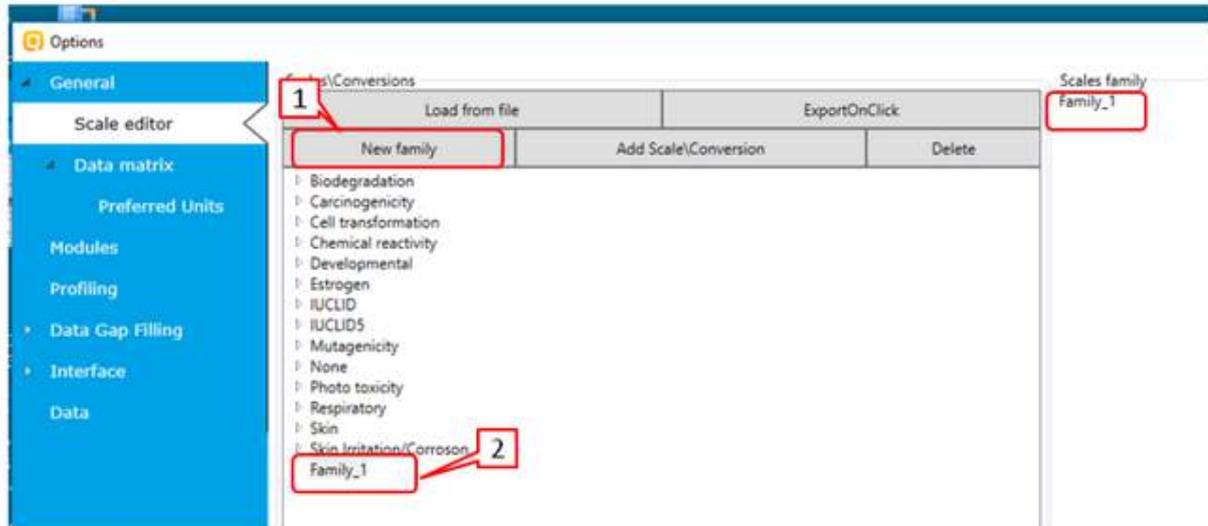


Figure 3

Then a new scale can be added to the family (1,2) in Figure 4:

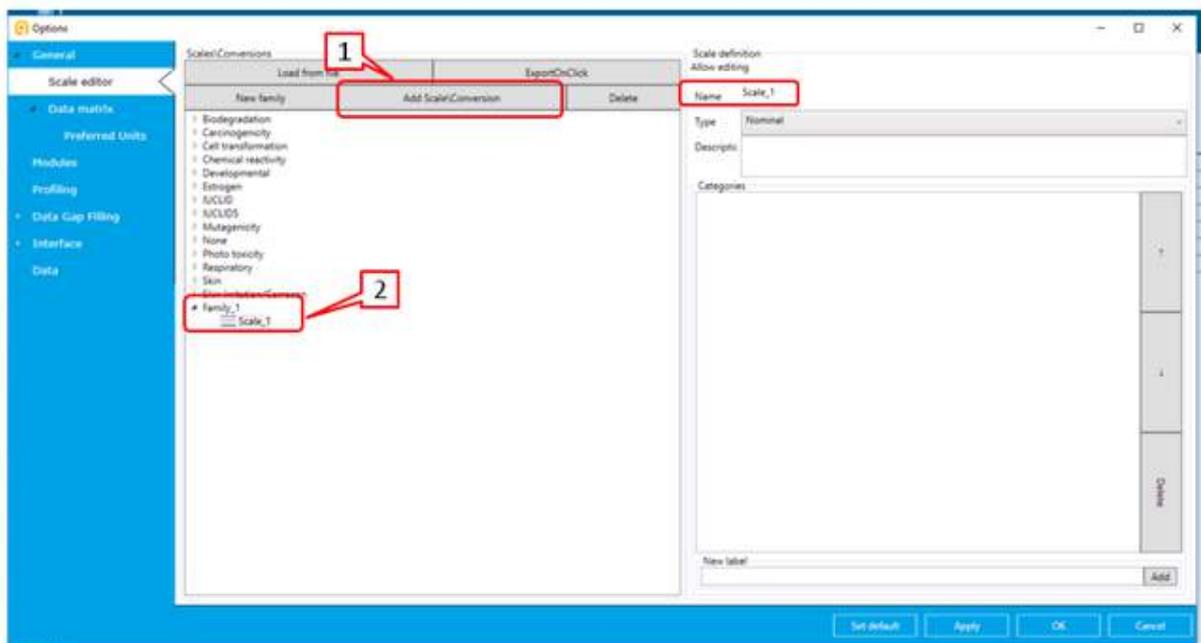


Figure 4

Expansion of the existing families of scales shows all the scale belonging to this family and their conversion as well (if any)(1) (Figure 5). Also the user can Load (2) a scale or Export a scale (3).

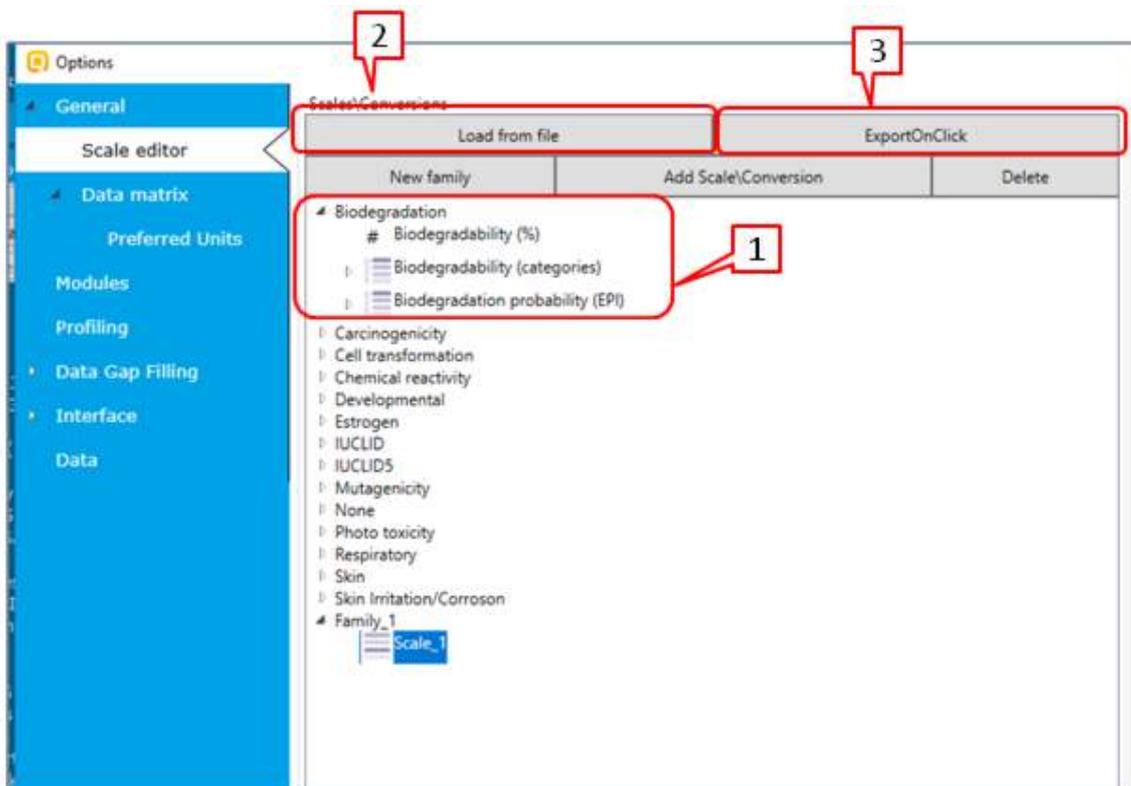


Figure 5

- Data matrix

Data matrix options (1) is sublevel of general options. The main right panel gives the dimension parameters of the matrix (2) (Figure 6)

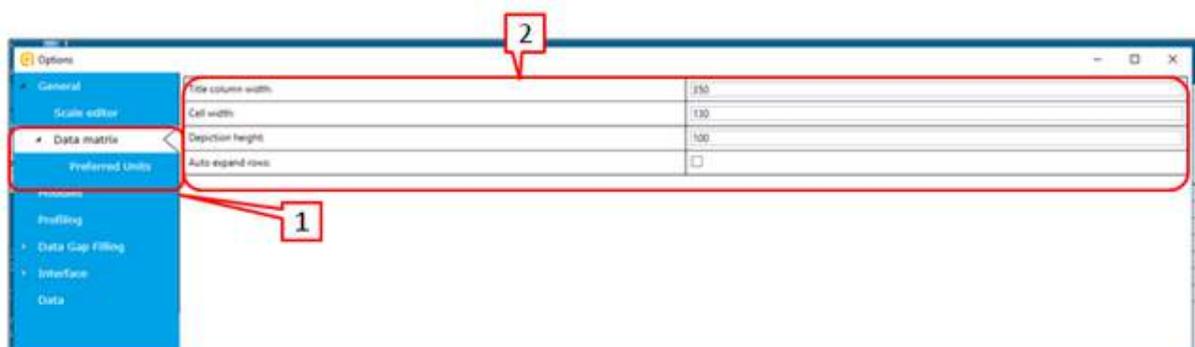


Figure 6

The sublevel of data matrix options are the Preferred units (1)(Figure 7):

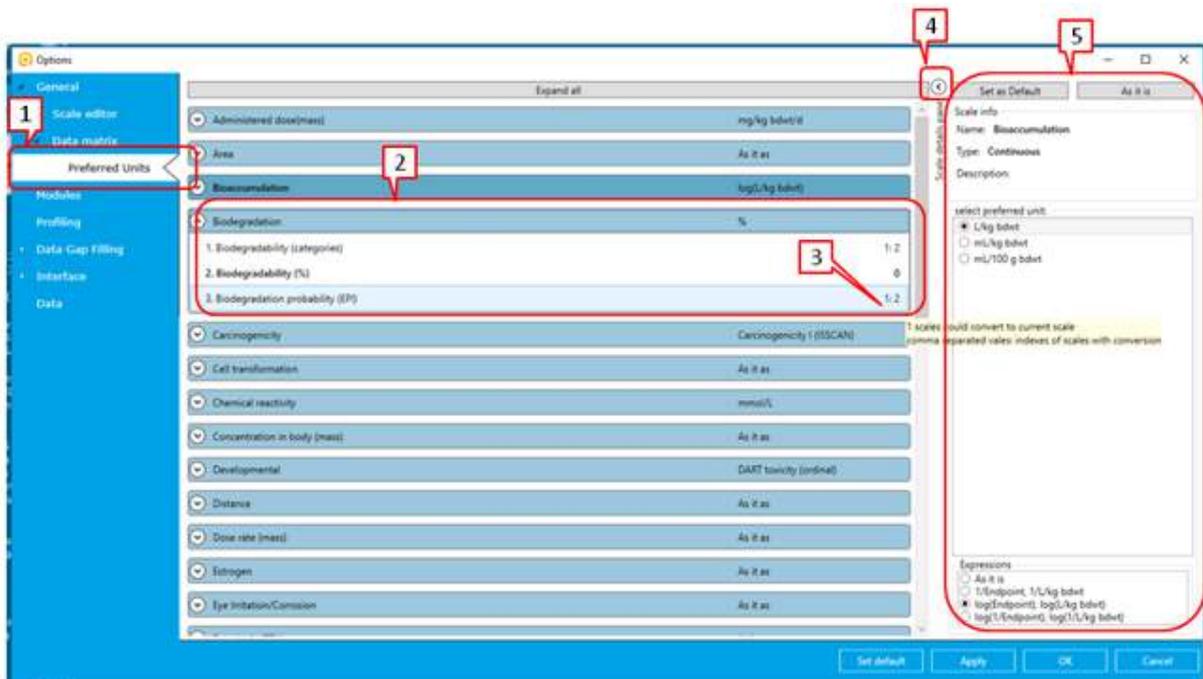


Figure 7

Here, the biodegradation units are given as an example:

- Expand the triangle next to the name in order to see all scales which can be used (2);
- The scale, which is bolded , Biodegradability (%), is the default scale;
- The number at the end (3) shows the scale conversion. For example Biodegradation probability scale is converted to scale number 2, which is biodegradability %.
- Scale details panel (5) can be open by clicking on (4). The panel allows selecting of different units and expressions.

QSAR Toolbox v.4

E.2. Modules

[Options](#) / E.2. Modules

E.2. Modules

The Modules panel contains list with all modules available in Toolbox – Calculators, Databases, Inventories, Profilers, (Q)SAR models, Metabolism etc. (Figure 1).

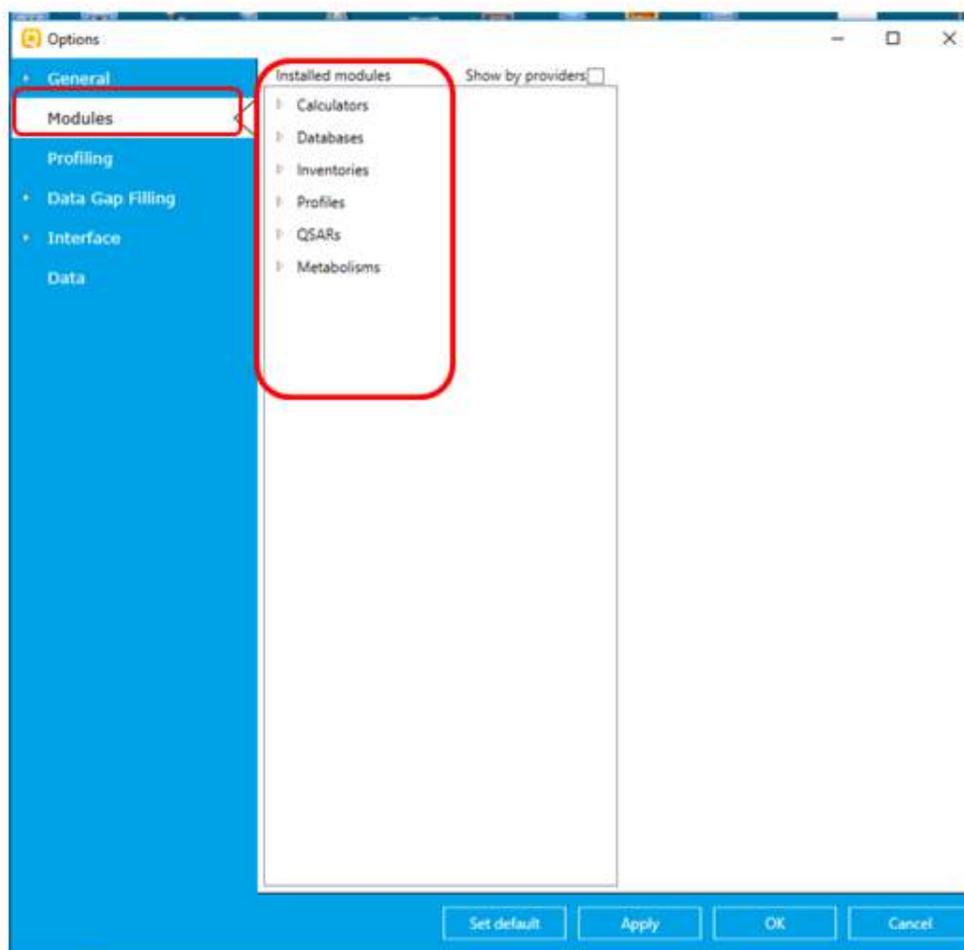


Figure 1

Double click on the expandable triangle to visualize the modules content . For example, double click on the calculators provides two list with all 2D and 3D calculators available in Toolbox (Figure 2).

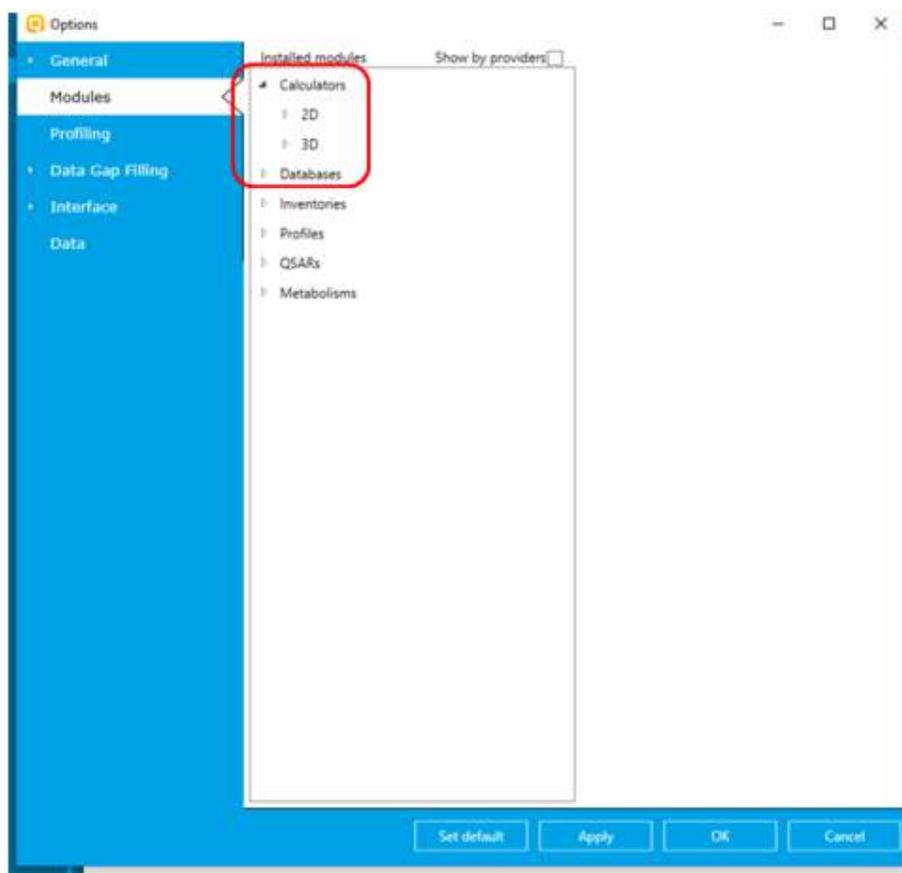


Figure 2

The next figure (Figure 3) represents the node Calculators in more detail:

QSAR Toolbox v.4

E.3. Profiling

[Options](#) / E.3. Profiling

E.3. Profiling

The general options associated with the *Profiling module* are shown in Figure 1:

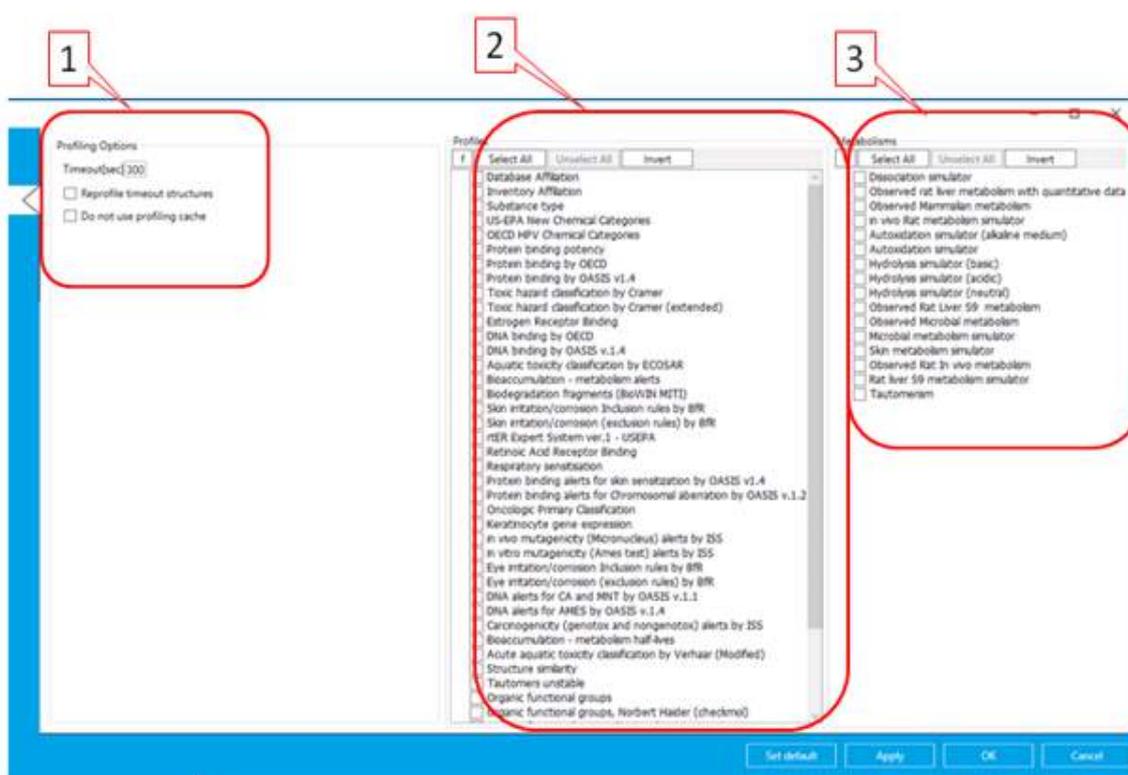


Figure 1

(1) - *Timeout* – here the user could increase or decrease the default maximum time for profiling of a structure.

Note: The profiling of some big and complex structure could take a lot of time which is the reason the timeout option was implemented. In case the time for profiling of that type of structures exceeds the set time the profiling result for these structures is !Timeout.

Reprofile timeout structures – this option is related to the previous. It allows the profiling the Time out structures if the Timeout threshold had been increased.

Profiling option related to the profiling calculation. Here the user can select not to use the profiling cache. The user should be careful when selecting not to use the cache as the process of profiling of large sets of chemicals is quite slow. This also affects the categorization process as it relies on the stored profiling results cache.

Note: Selecting of this option is recommended in cases when check the results of newly created profiling scheme or check modifications of already existing scheme

(2) - Allows the user to choose profilers which will be selected every time , when the software starts.

(3) - Allows the user to choose metabolism simulators, which will be selected every time, when the software starts.

-0-

QSAR Toolbox v.4

E.4 Data Gap Filling

[Options](#) / E.4 Data Gap Filling

E.4. Data gap filling

Data gap filling options contains four main sub-levels (Figure 1) (1): Preferred Units, Calculation, Endpoint tree, Descriptors

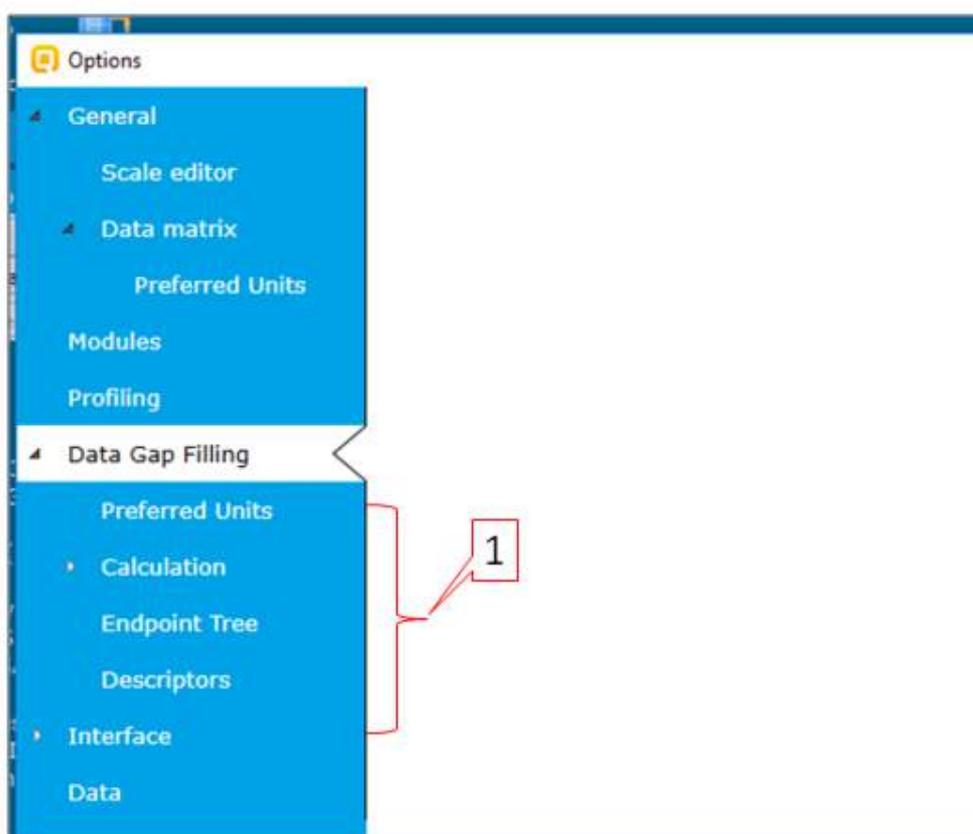


Figure 1

- Preferred Units (1) in Figure 2

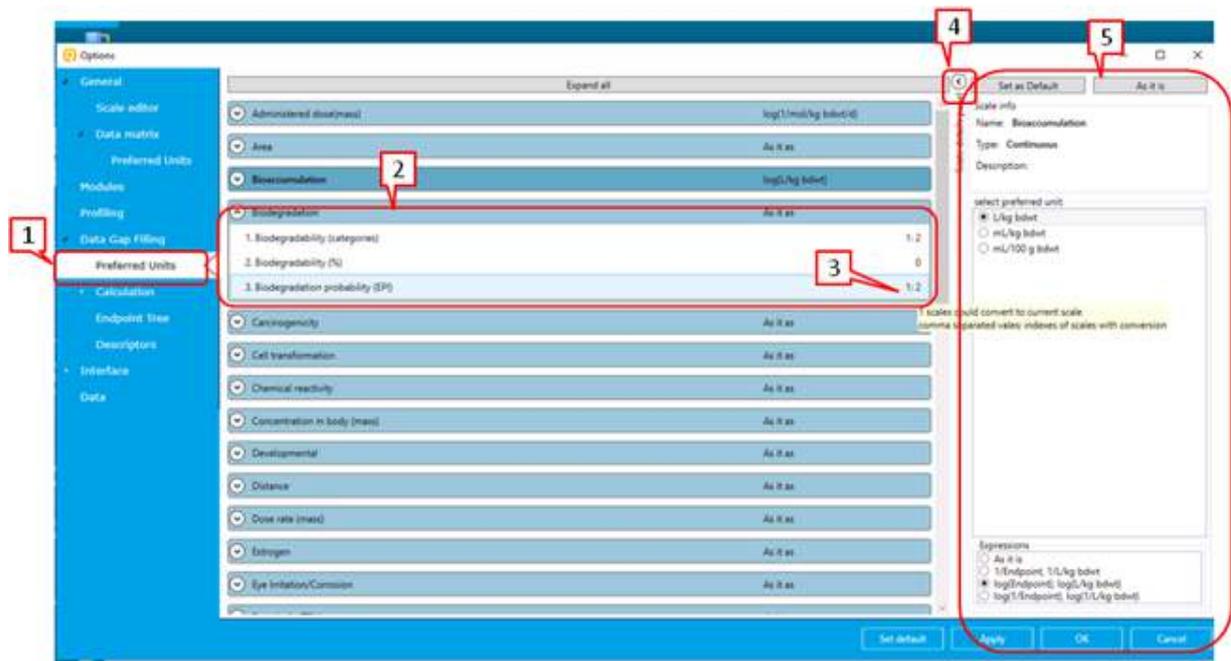


Figure 2

Biodegradation units are given as an example (Figure 2):

- Expand the triangle next to the name in order to see all scales which can be used (2);
- The bolded scale, Biodegradability (%), is the default scale;
- The number at the end shows the scale conversion. For example Biodegradation probability scale is converted to scale number 2, which is Biodegradability %.
- Scale details panel (5) can be open by clicking on (4). The panel allows the user to select different units and expressions.

- Calculation

In Calculation section (1), the user can select the different calculation options for each data-gap filling method (2) (Figure 3).

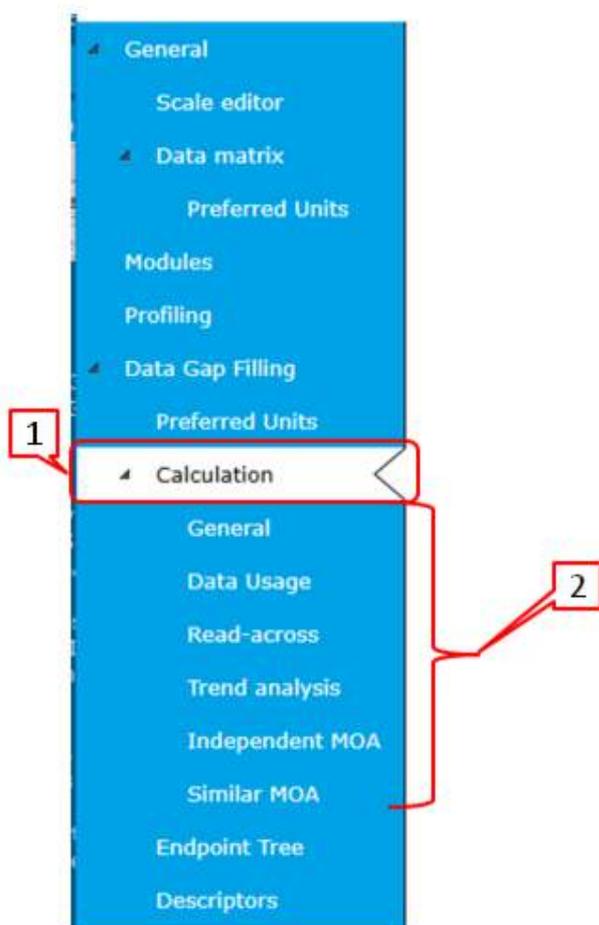


Figure 3

- *General (1)*: contains the general options (2) associated with all methods (Figure 4)

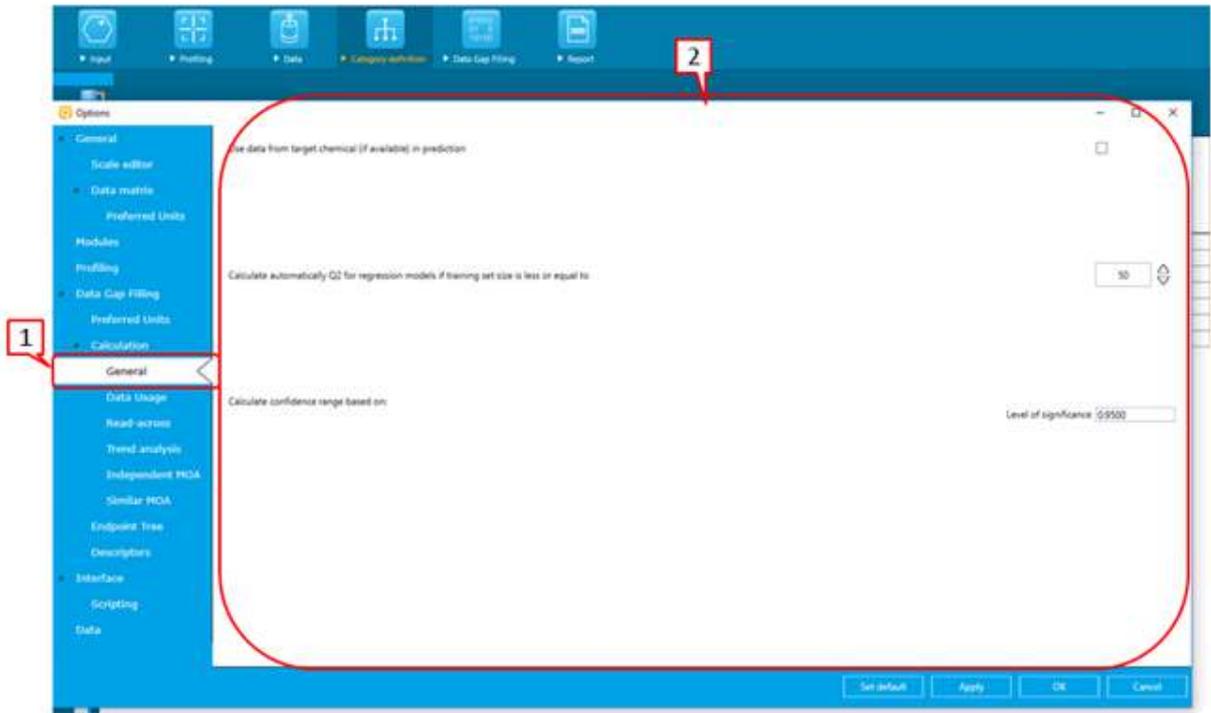


Figure 4

- *Data usage (1)*: shows the type of scales used in all data gap filling methods and their default settings (2) (Figure 5).

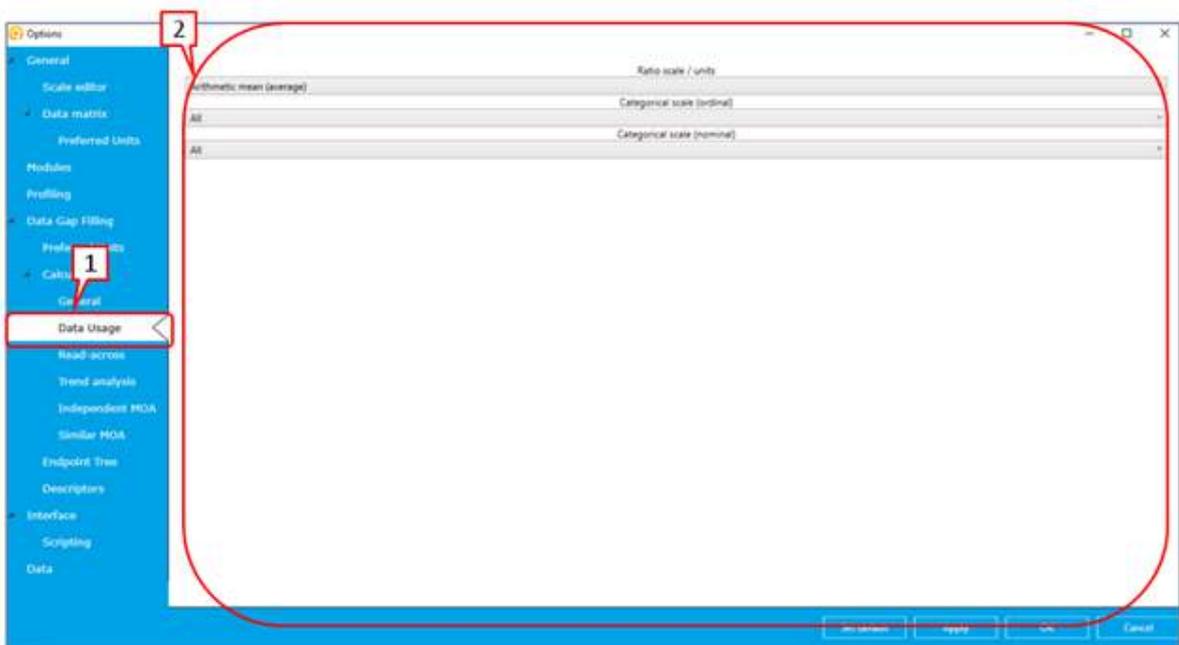


Figure 5

The user can modify them by selecting the drop-down menu of each scale (1 in Figure 6).

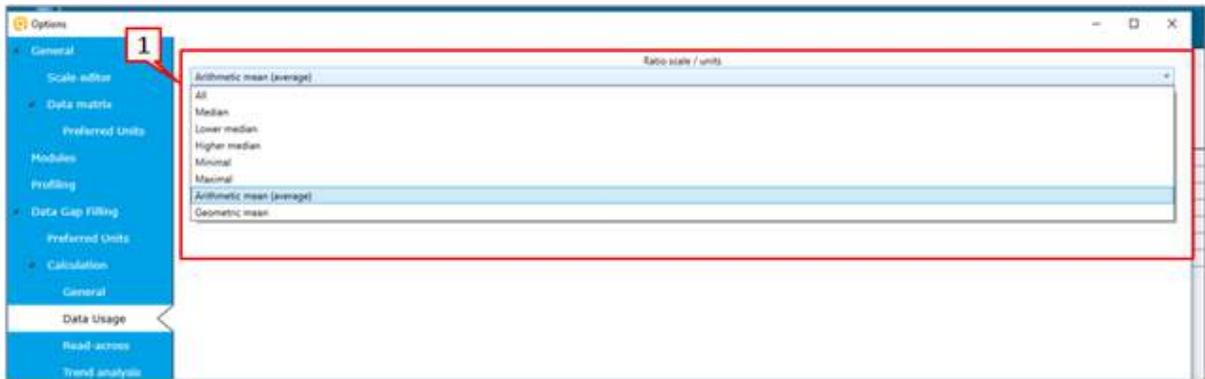


Figure 6

- *Read-across (1)*: shows the default settings for data usage of the nearest analogues used in the read-across prediction (2). The settings and the number of the nearest analogues used in the read-across can be modified by the user (Figure 7).

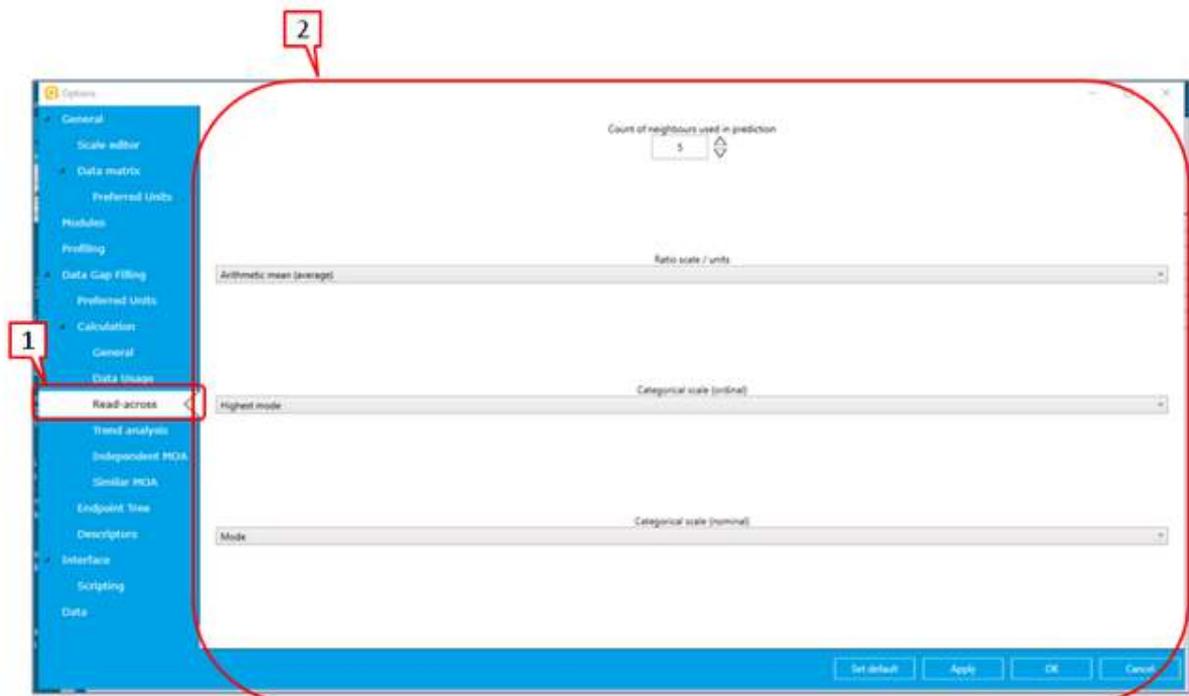


Figure 7

- *Trend-analysis (1)*: gives the type of approximation used in the trend-analysis method: linear or quadratic (2) (Figure 8)

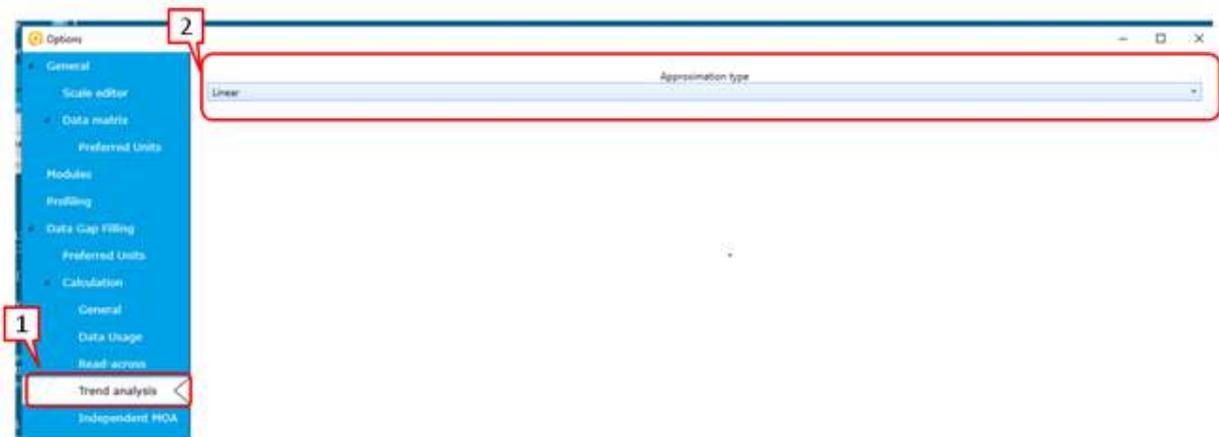


Figure 8

- Independent MOA (Mode of action) (1): This data-gap filling method is used for mixtures. Its options are set in order to represent the worst-case scenario (2) (Figure 9).

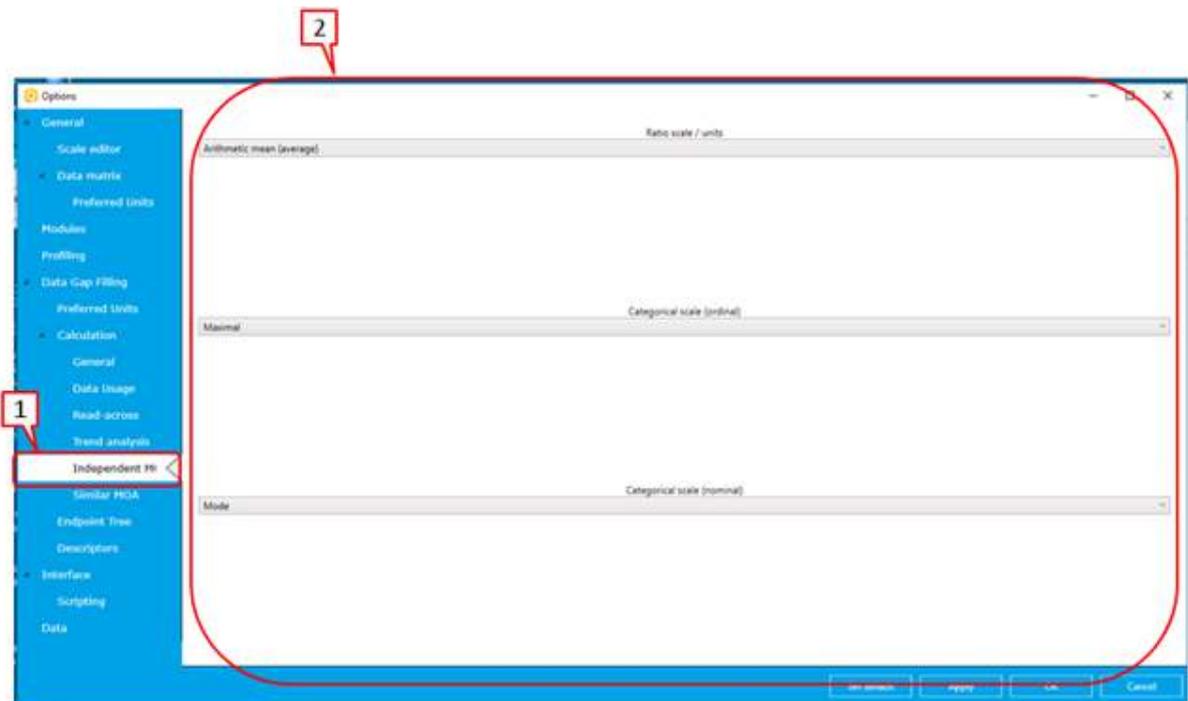


Figure 9

- Similar MOA (Mode of action)(1) (Figure 10): This data-gap filling method is used for mixtures. Only one setting is allowed (2), which is selected in order to take into account the quantities of the individual components.

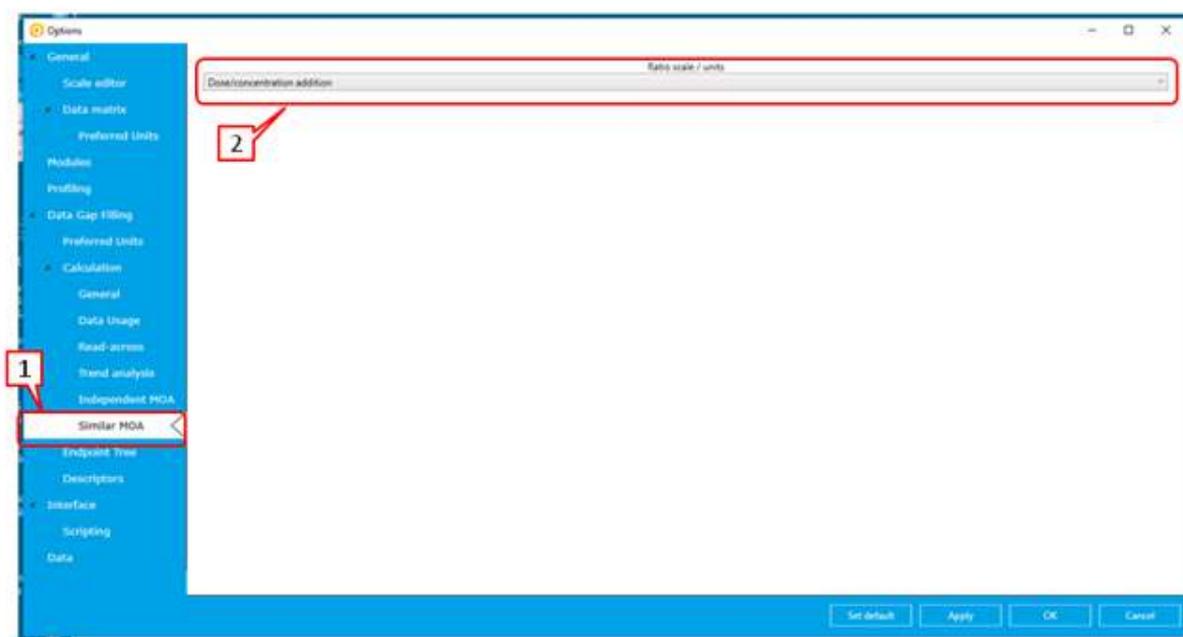


Figure 10

- Endpoint tree (1): In this section the user can select a preferred descriptor on the x axis of the prediction for each endpoint. (Figure 11):
 - Expand Default (2);
 - Select an endpoint (3);
 - Set the descriptor (4). By default the descriptor on x axis is set to log Kow.

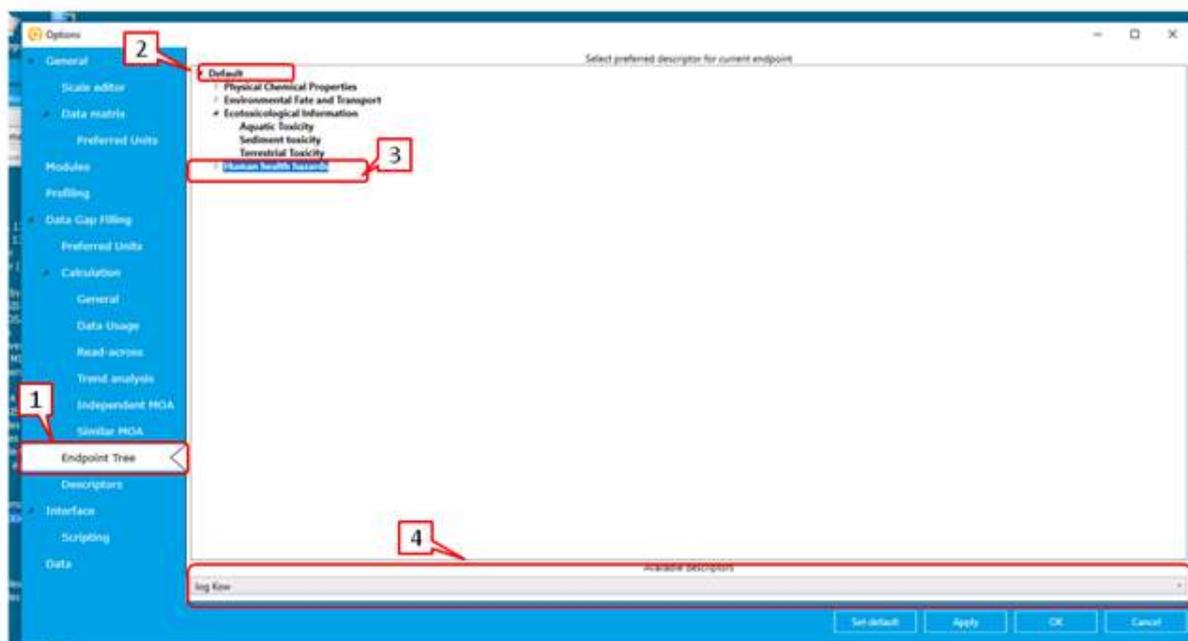


Figure 11

- Descriptors (1) (Figure 12): In this section the user can select the preferred units and expressions of the descriptors used on the x axis of the prediction:
 - Select the descriptor (2);
 - Open Scale details panel (3);
 - Select the preferred Unit and expressions (4).

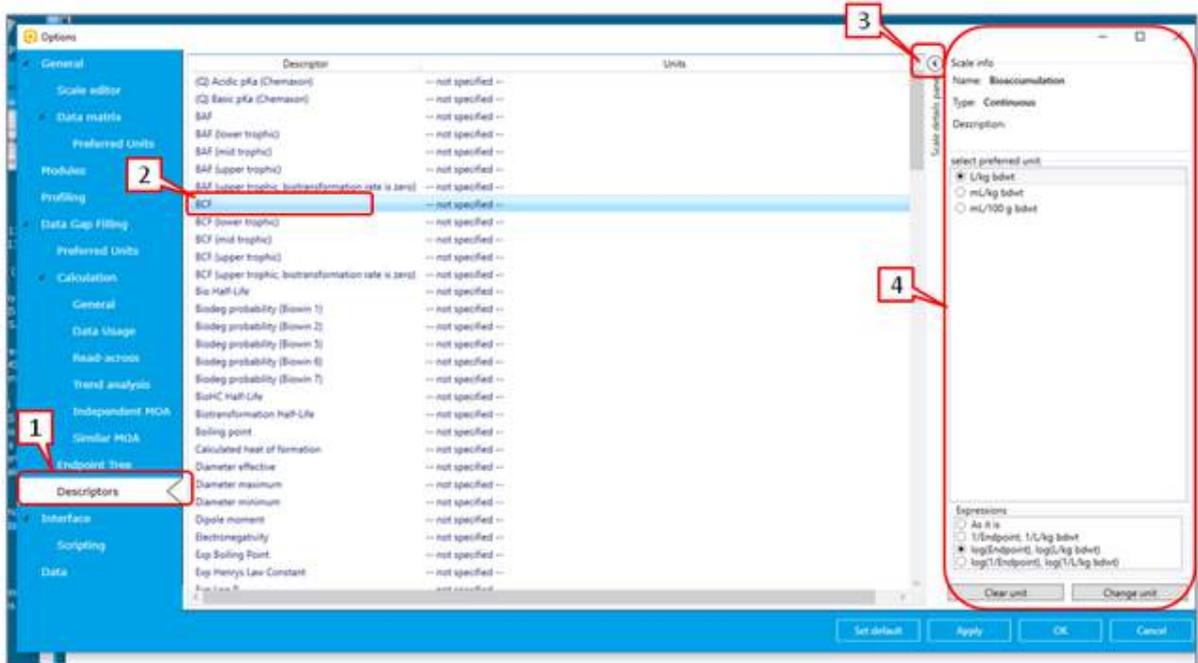


Figure 12

QSAR Toolbox v.4

E.5. Interface

[Options](#) / E.5. Interface

E.5. Interface

This options (1) allows changing some feature of the interface (2) (Figure 1):

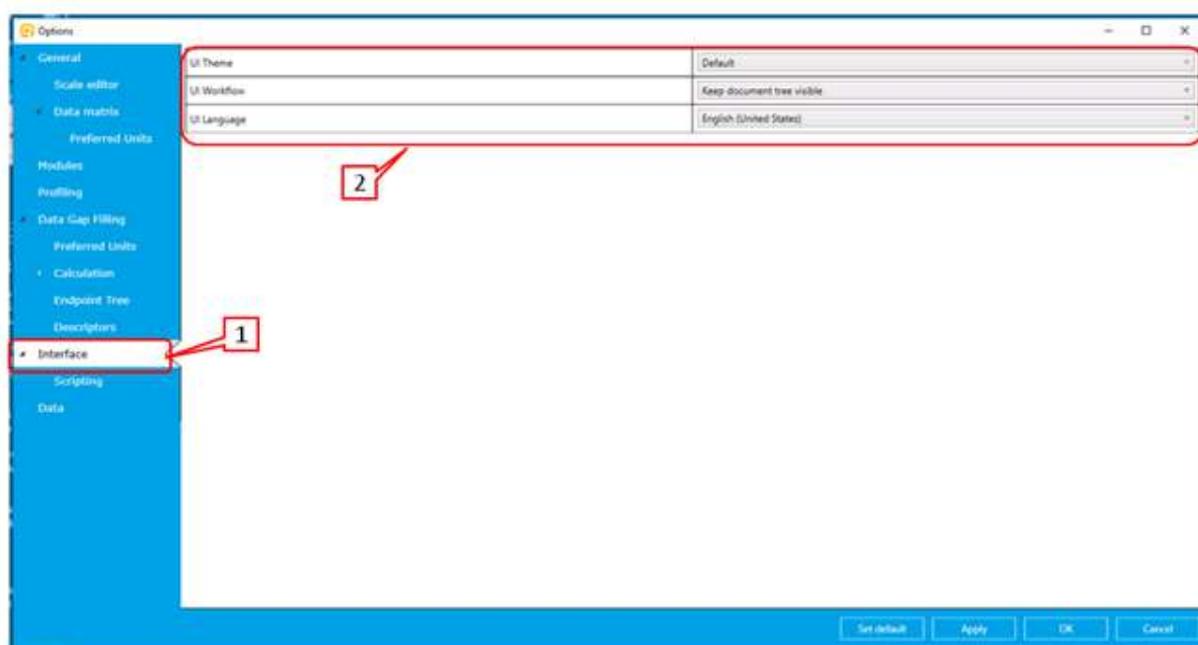


Figure 1

The following settings are available:

- UI (User Interface) Theme: several themes are available;
- UI (User Interface) Workflow: controls the right-hand panel during execution of workflows. It gives options to view/hide the document tree(1) and to view/hide the data-gap filling settings panel (2) (Figure 2)

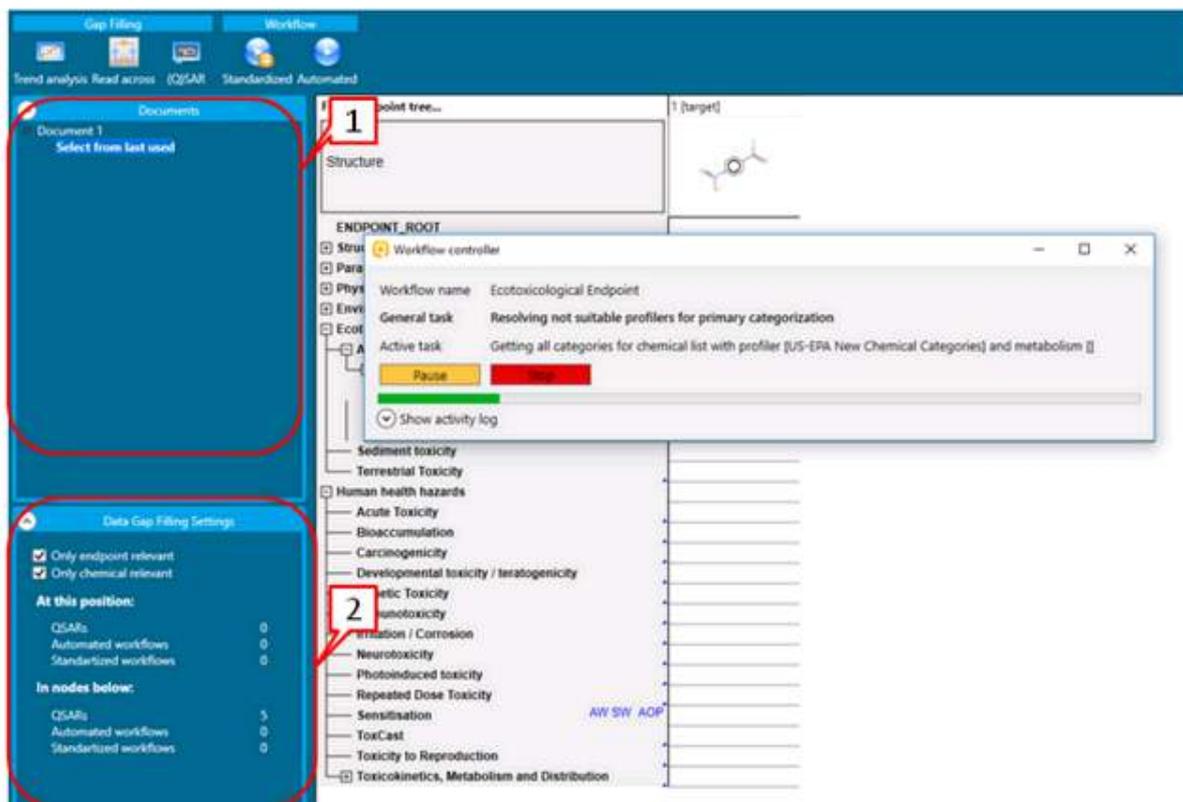


Figure 2

- Scripting

Section *Scripting* (1) (Figure 3) contains settings (2) related to the interface of the automated and standard workflows available in Data gap filling. *Enable logging* option is checked by default as it allows user to check the progress of the workflow during its execution (1 in Figure 4) .

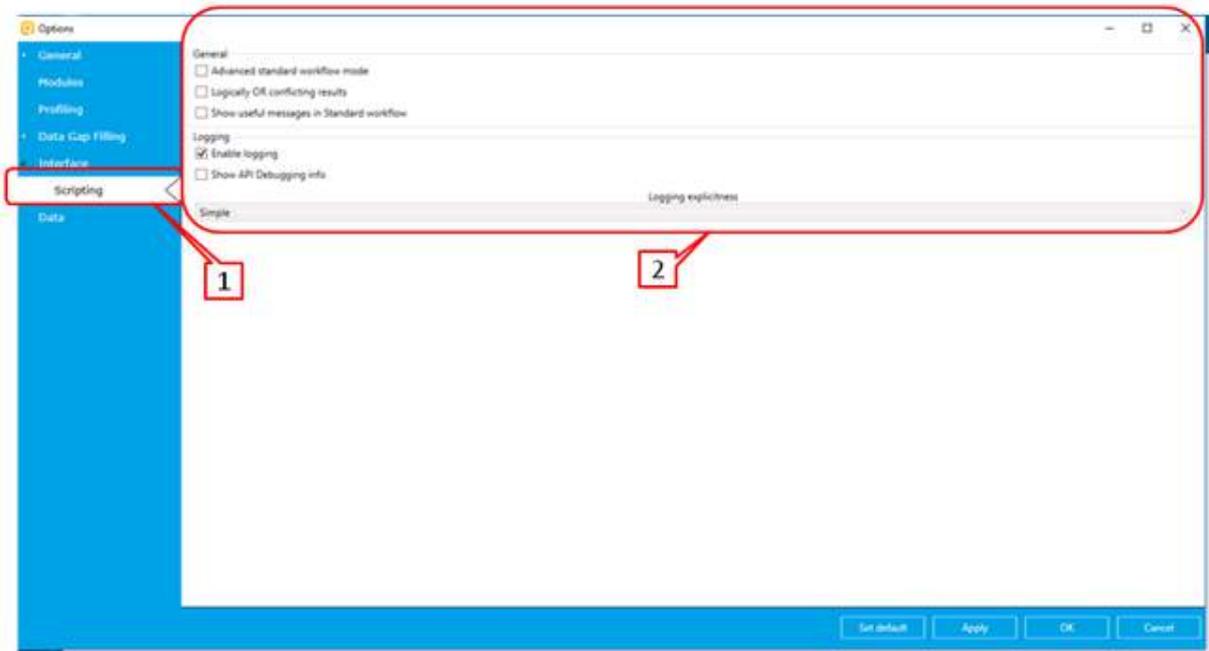


Figure 3

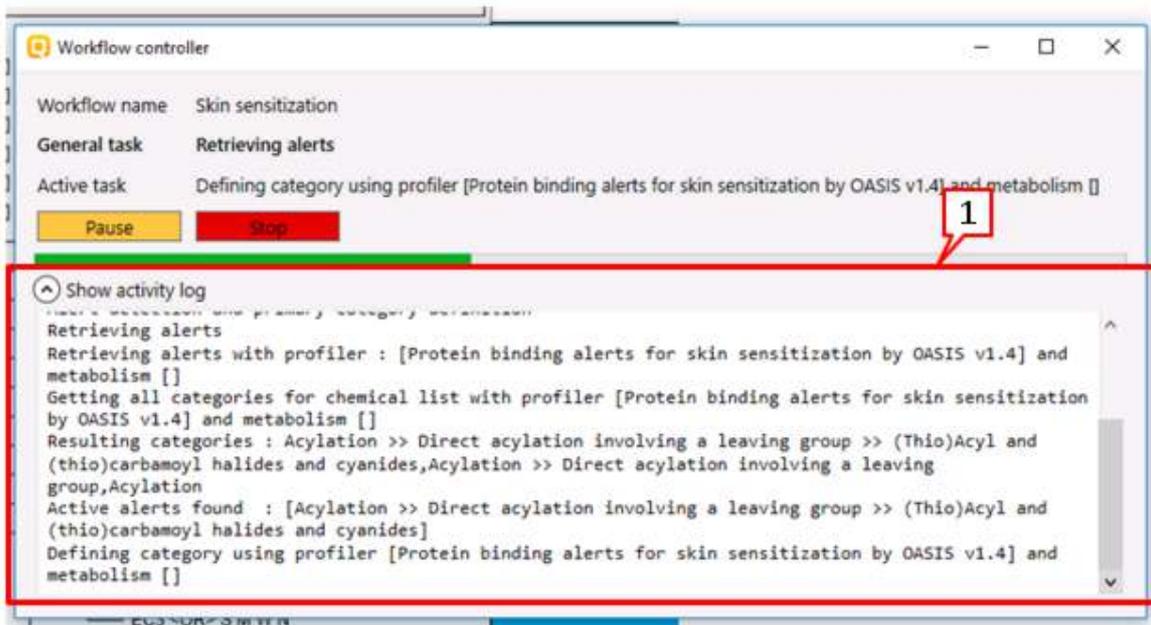


Figure 4

Index

- A -

A.1. Objective 6

- B -

B.1. Acknowledgements 8
B.2. What is OECD QSAR Toolbox
B.3. Abbreviations 10

- C -

C.1. Stages toolbar 12
C.2. Actions toolbar 14
C.3. Stage options panel 16
C.4. Data matrix 22
C.4.1. Endpoint tree area 24
C.4.2. Area with selected chemicals 45
C.4.3. Area with data 47
C.5. Information and active accessories 51
Chapter A. Introduction 5
Chapter B. Welcome 8

- D -

D.1. Documents
D.1.1.3.1. Add single chemical structure (discrete chemical) 60
D.1.1.3.2. Add single chemical structure (mixture) 63
D.1.1.4. Entering a chemical from the last used chemicals 69
D.1.2.1. Loading a database 82
D.1.2.2. Loading an inventory list 85
D.1.2.3. Loading a custom file 87
D.1.3.1. Substructure (SMARTS)- 2DEditor 90
D.1.3.2. Query tool 95
D.1.3. Search options 89
D.1.4. Define Target Endpoint 116
D.1.4.1. Gene mutation in vitro 124
D.1.4.2. Gene mutation in vivo 127
D.1.4.3. Chromosome aberration in vitro 128
D.1.4.4. Chromosome aberration in vivo 130
D.1.4.5. Other endpoints 132
D.2.1.1. Entering a chemical by CAS ___
D.2.1.1. Predefined 134
D.2.1.2. Entering a chemical by chemical name ___
D.2.1.2. General mechanistic 136
D.2.1.3. Endpoint specific 144
D.2.1.3. Entering a chemical by structure ___
D.2.1.4. Empiric 153
D.2.1.5. Toxicological 158

- D.2.1.6. Custom 159
 - D.2.2. Metabolism 160
 - D.2.2.1. Documented metabolism 160
 - D.2.2.2. Simulated metabolism 163
 - D.2.3. Supporting functionalities 168
 - D.2.4. Explaining profiling methods 172
 - D.2.5. Explaining profiling results 177
 - D.2.6. Creating a custom profile 184
 - D.2.7. Deleting a profile 188
 - D.3.1. Databases 190
 - D.3.2. Inventories 216
 - D.3.4.1. Vertical Import: Database 223
 - D.3.4.2. Horizontal Import: Database 234
 - D.3.4.3. Importing Inventory 239
 - D.3.4.4. Supporting information 242
 - D.4.3. Defining category 261
 - D.4.3.1. Categorization of single chemical 268
 - D.4.3.2. Categorization of mixtures 269
 - D.4.3.3. Categorization using profiling result of hierarchical type profiling scheme 271
 - D.4.3.4. Categorization with metabolism 276
 - D.4.3.5. Alert performance 280
 - D.4.4. Subcategorization 284
 - D.4.5. Combine categories 288
 - D.5.1. Data Gap Filing methods 290
 - D.5.2. Data Gap Filing procedure 292
 - D.5.3. Data Gap Filing window 295
 - D.5.4. Common features in three gap filling approaches 298
 - D.5.4.1. Left-side panel 299
 - D.5.4.2. Menus of data gap filling 304
 - D.5.4.2.1. Select /filter data 305
 - D.5.4.2.1.1. Subcategorize 307
 - D.5.4.2.1.2. Mark chemical by WS
 - D.5.4.2.1.3. Mark chemicals by descriptor values 329
 - D.5.4.2.1.4. Filter points by test condition 331
 - D.5.4.2.1.5. Mark focused chemical 333
 - D.5.4.2.1.6. Mark focused points 334
 - D.5.4.2.1.7. Remove marked chemicals/points 337
 - D.5.4.2.1.8. Clear existing marks 339
 - D.5.4.2.1.9. Mark outliers 340
 - D.5.4.2.2. Gap filling approaches 341
 - D.5.4.2.3. Descriptors/data 342
 - D.5.4.2.4. Model/(Q)SAR 345
 - D.5.4.2.5. Calculation options 348
 - D.5.4.2.5.1. Data usage options 351
 - D.5.4.2.5.1.1. Prediction examples 357
 - D.5.4.2.6. Visual options 365
 - D.5.4.2.7. Information 369
 - D.5.4.2.8. Miscellaneous 373
 - D.5.4.3. Right click functionality 375
 - D.5.5. Specific features of (Q)SARs 376
 - D.5.6. Automated and Standardized workflows 384
 - D.5.6.1. Automated workflow 385
 - D.5.6.2. Standardized workflow 392
 - D.5.7. Methodologies for estimating toxicity of a set of chemicals
 - D.5.7.1. Example: Quantitative mixtures toxicity prediction
 - D.5.7.1.1. Mixture models 411
 - D.5.7.1.1.1. Independent MOA 412
 - D.5.7.1.1.1.1. Illustration 414
 - D.5.7.1.1.2. Similar MOA 415
 - D.5.7.1.1.2.1. Illustration 417
 - D.5.7.1.1.2.2. Example 419
- D.6.1. Report window 423
- D.6.2. Create a report 424

- E -

- E.1. General 436
- E.2. Modules 441
- E.3. Profiling
- E.4 Data Gap Filling 446
- E.5. Interface 455
- E.6. Data 459

