OECD QSAR Toolbox v4.0 Simplifying the correct use of non-test methods

Stakeholders' Day IT tool training

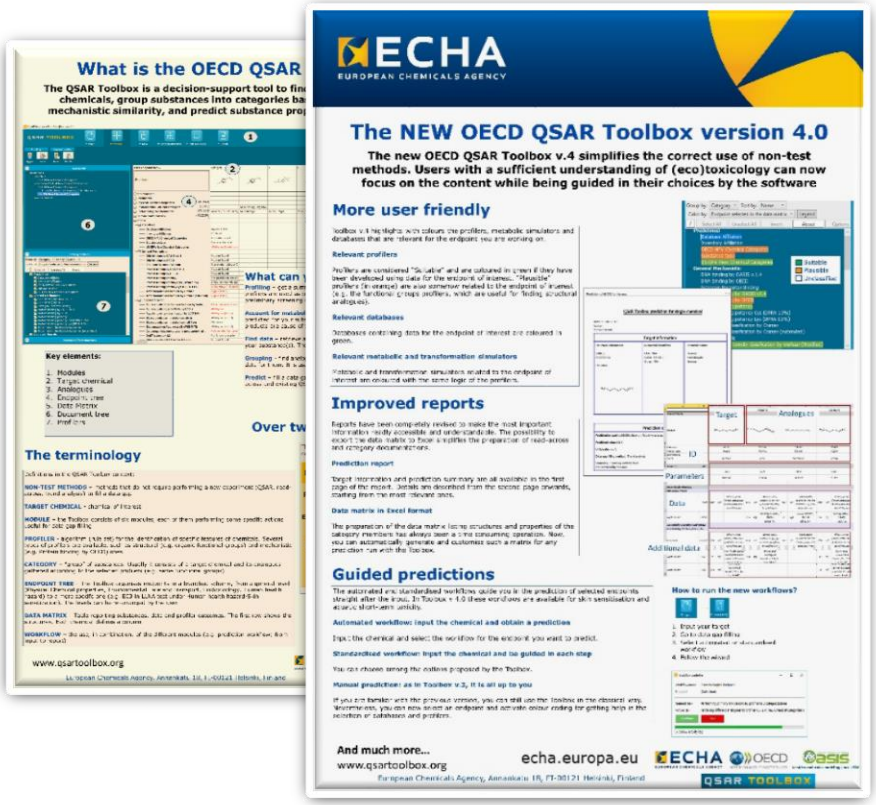
4 April 2017

Tomasz Sobanski
Andrea Gissi
Marta Sannicola

Computational assessment and dissemination
European Chemicals Agency, Helsinki

The new QSAR Toolbox v4.0 is now available!

1. More user friendly
2. Improved reports
3. Guided predictions
4. And much more...



Outline

1. The regulatory context

2. The OECD QSAR Toolbox project

3. Introduction to the tool
    a. General introduction
    b. Toolbox step by step: core functions and new features

4. Practical Applications and Live demo

5. Conclusions

The regulatory context

The regulatory context

- Animal testing as last resort under REACH
- Use of alternatives without compromising the level of protection for human health and environment
- Annex XI general provisions
- Transparency of the methods as a key for regulatory acceptance

Annex XI of the REACH regulation

Art 1.3:

(..)Results of (Q)SARs may be used instead of testing when the following conditions are met:

- results are derived from a (Q)SAR model whose **scientific validity** has been established,
- the substance falls within the applicability domain of the (Q)SAR model,
- results are adequate for the purpose of classification and labelling and/or risk assessment, and
- adequate and reliable documentation of the applied method is provided.

Scientific validity

Five OECD principles for assessing (Q)SAR model's scientific validity:

1. A defined endpoint
2. An unambiguous algorithm
3. A defined domain of applicability
4. Appropriate measures of goodness-of-fit, robustness and predictivity
5. A mechanistic interpretation, if possible

The OECD QSAR Toolbox Project

Facts about QSAR Toolbox

- OECD and ECHA co-own and co-manage the QSAR Toolbox
- International peer review process for developing the system and mechanistic transparency of the results are of key importance for the regulatory acceptance
- The system is freely available and maintained in the public domain by OECD

Many other Toolbox Supporters:

- ✓ OECD
- ✓ European Chemicals Agency
- ✓ US EPA, OPP
- ✓ US EPA, NHEERL
- ✓ Environment Canada
- ✓ Health Canada
- ✓ NITE Japan
- ✓ NIES Japan
- ✓ Danish EPA
- ✓ UBA Germany
- ✓ NICNAS Australia
- ✓ DEWNA Australia
- ✓ ISS Italy
- ✓ Fraunhofer Germany
- ✓ BfR Germany
- ✓ Cefic
- ✓ Oasis
- ✓ L'Oreal
- ✓ DuPont
- ✓ Givaudan
- ✓ Dow chemicals
- ✓ BASF
- ✓ ExxonMobil
- ✓ 3M
- ✓ Firmenich SV
- ✓ SRC, Syracuse
- ✓ Unilever
- ✓ Multicase
- ✓ ChemAxon
- ✓ International QSAR Foundation

QSAR Toolbox: the concept

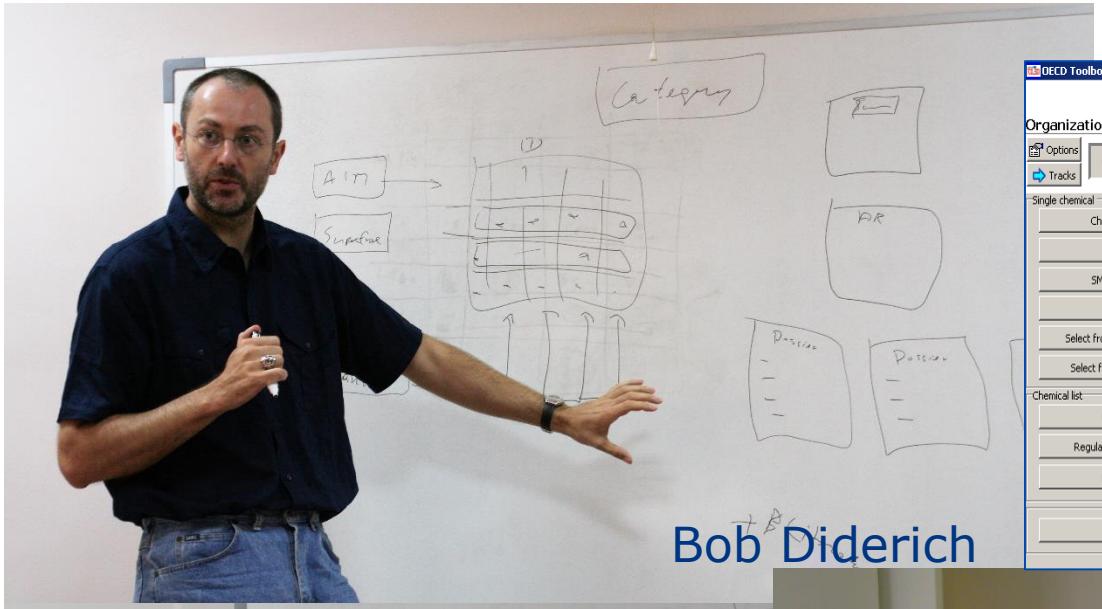
The QSAR Toolbox is a decision supporting tool for hazard assessment - predicting properties is only one of its functionalities.

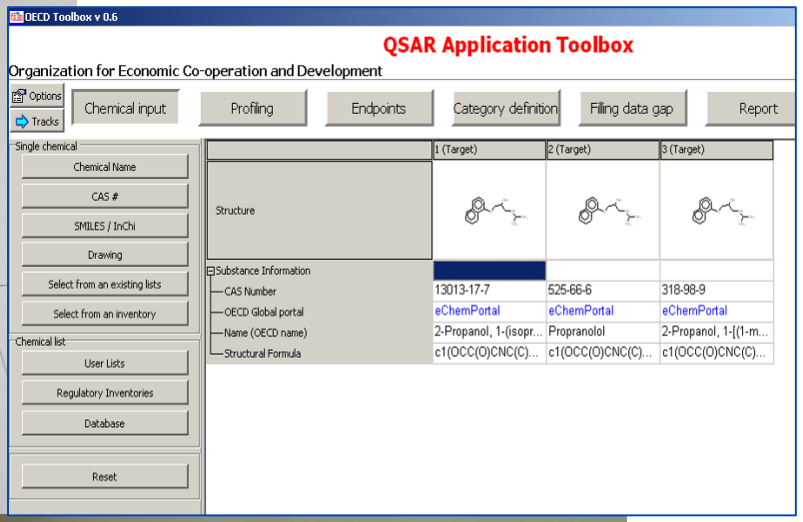
Training sets (categories) for each prediction are defined dynamically, while most of the other (Q)SAR models have static training sets

Each estimated value can be individually justified based on:

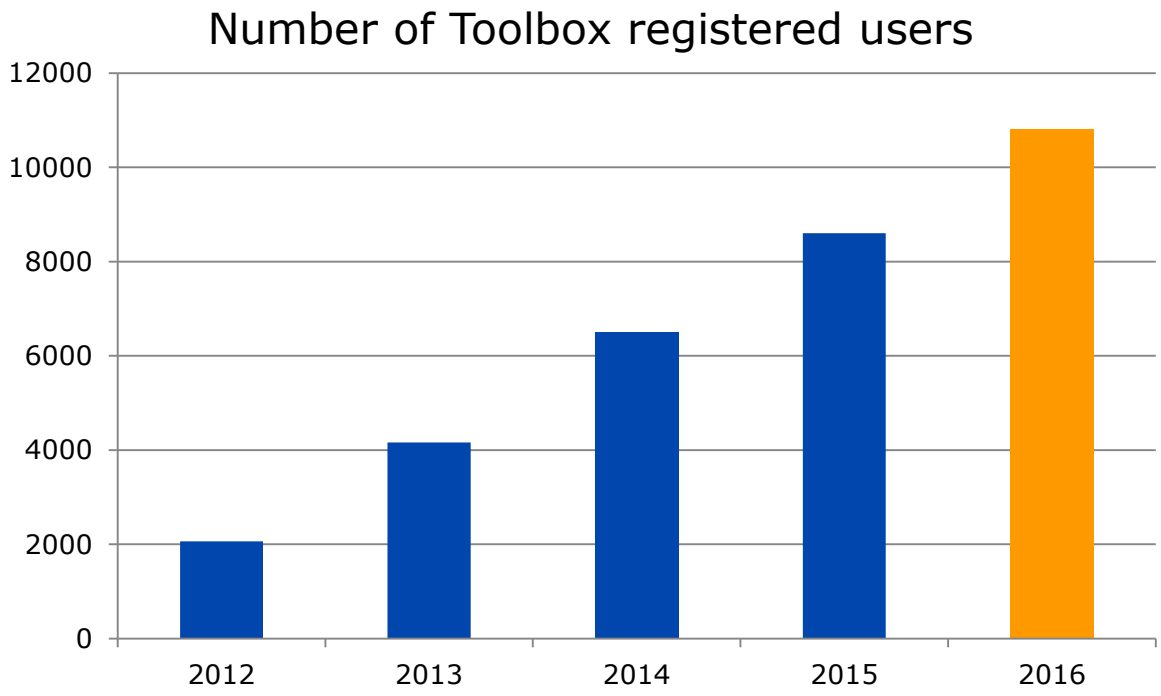
- Category hypothesis (justification)
- Quality and consistency of measured data

From knowledge & experience into the OECD QSAR Toolbox





QSAR Toolbox is getting more and more popular

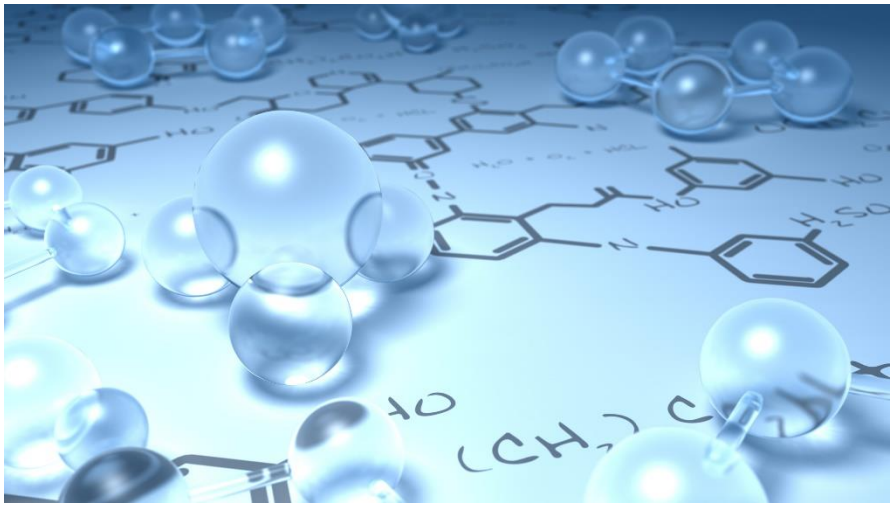


Introduction to the Toolbox

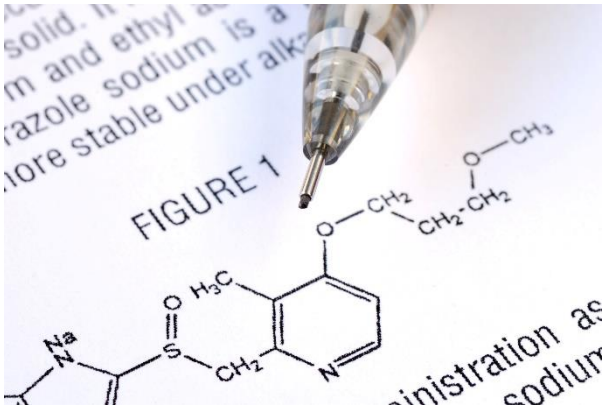
Agenda

General introduction
❖ Predictions and much more
❖ Keywords - Glossary
❖ The interface

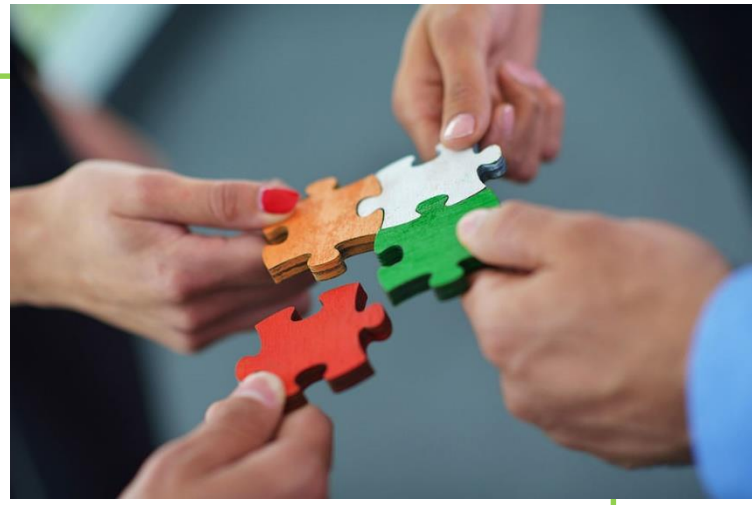
Toolbox step by step: insights
❖ Core functions
❖ New features in version 4.0



Predictions and much more…



QSAR Toolbox: collecting and applying knowledge on chemicals



- **Searching** for available experimental data
- **Profiling** chemicals
- **Grouping** analogues
- **Simulating** metabolites
- **Filling data gaps** with prediction workflows for (eco)toxicological endpoints

Talking about Toolbox…

Keywords:

1. **TARGET CHEMICAL -** chemical of interest

2. **MODULE –** a Toolbox module is a section dedicated to specific actions and options

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

4. **PROFILER** - algorithm (rule set) for the identification of specific features of the chemicals. Several types of profilers are available, such as structural (e.g. organic functional groups) and mechanistic (e.g. Protein binding by OECD) profilers.

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

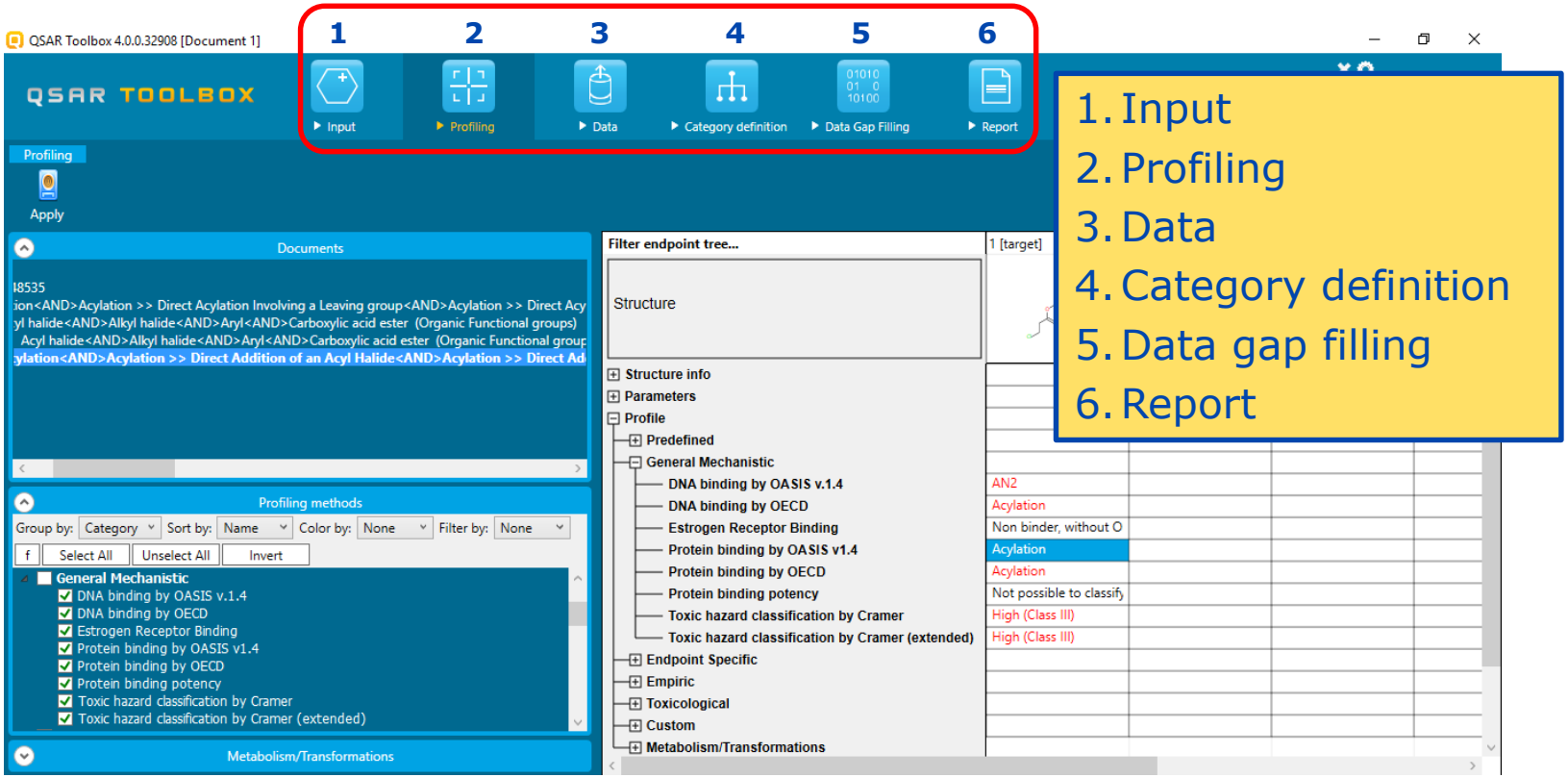
6. **ENDPOINT TREE** – Endpoints are structured in a branched scheme, from a broader level (Phys-Chem properties, Environmental Fate and transport, Ecotoxicology, Human health hazard) to a more detailed one (e.g. EC3 in LLNA test under Human health hazard-Skin sensitization)

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

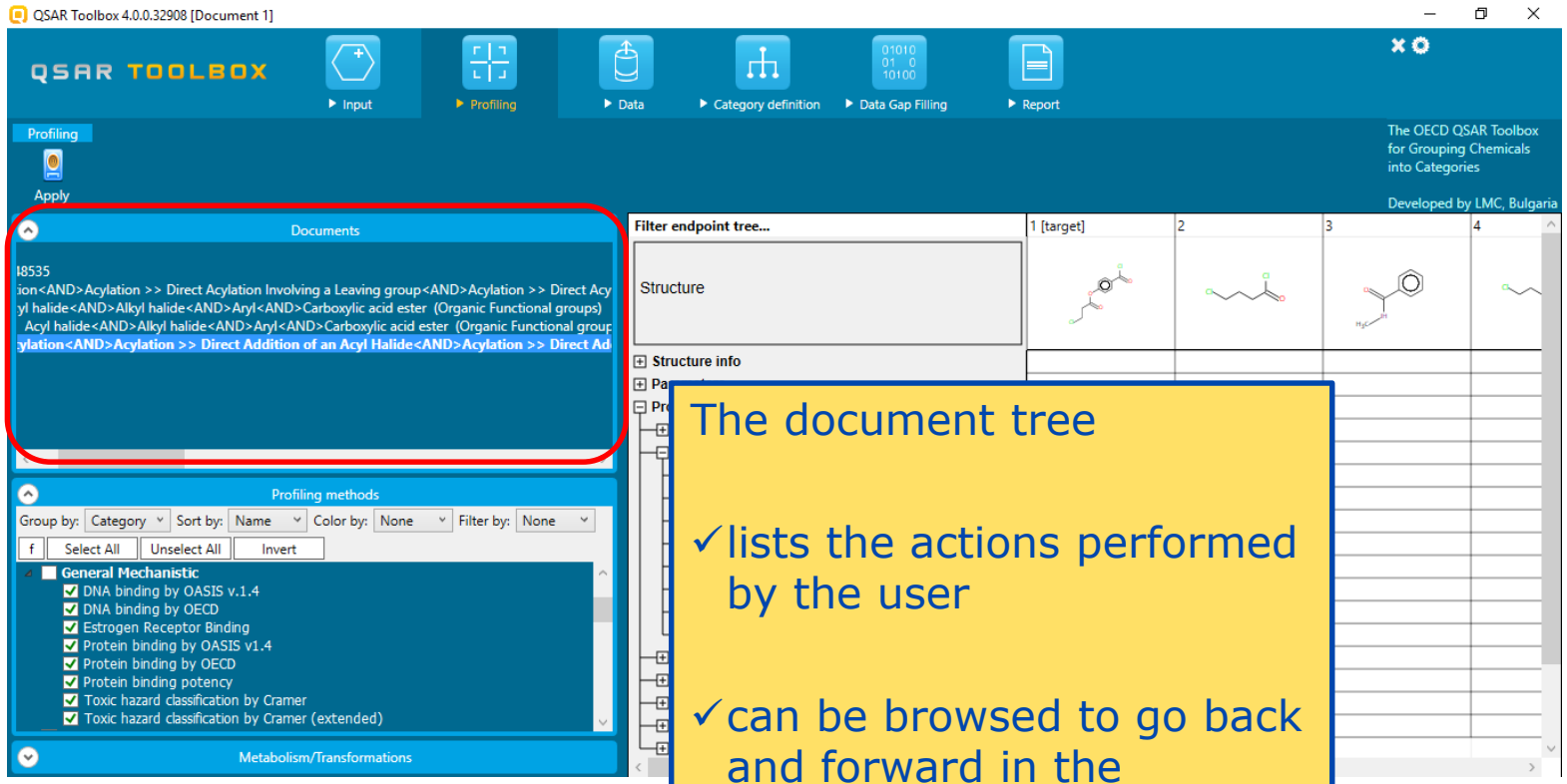
The interface

- ✓ **What it looks like**
- ✓ **Main elements**

✓ Modules


QSAR Toolbox 4.0.0.32908 [Document 1]
1
2
3
4
5
6
QSAR TOOLBOX
Input
Profiling
Data
Category definition
Data Gap Filling
Report
Profiling
Apply
Documents
Profiling methods
Group by: Category
Sort by: Name
Color by: None
Filter by: None
Select All
Unselect All
Invert
General Mechanistic
DNA binding by OASIS v.1.4
DNA binding by OECD
Estrogen Receptor Binding
Protein binding by OASIS v1.4
Protein binding by OECD
Protein binding potency
Toxic hazard classification by Cramer
Toxic hazard classification by Cramer (extended)
Metabolism/Transformations
Filter endpoint tree...
Structure
Structure info
Parameters
Profile
Predefined
General Mechanistic
DNA binding by OASIS v.1.4
DNA binding by OECD
Estrogen Receptor Binding
Protein binding by OASIS v1.4
Protein binding by OECD
Protein binding potency
Toxic hazard classification by Cramer
Toxic hazard classification by Cramer (extended)
Endpoint Specific
Empiric
Toxicological
Custom
Metabolism/Transformations
1 [target]
AN2
Acylation
Non binder, without O
Acylation
Acylation
Not possible to classify
High (Class III)
High (Class III)
1. Input
2. Profiling
3. Data
4. Category definition
5. Data gap filling
6. Report

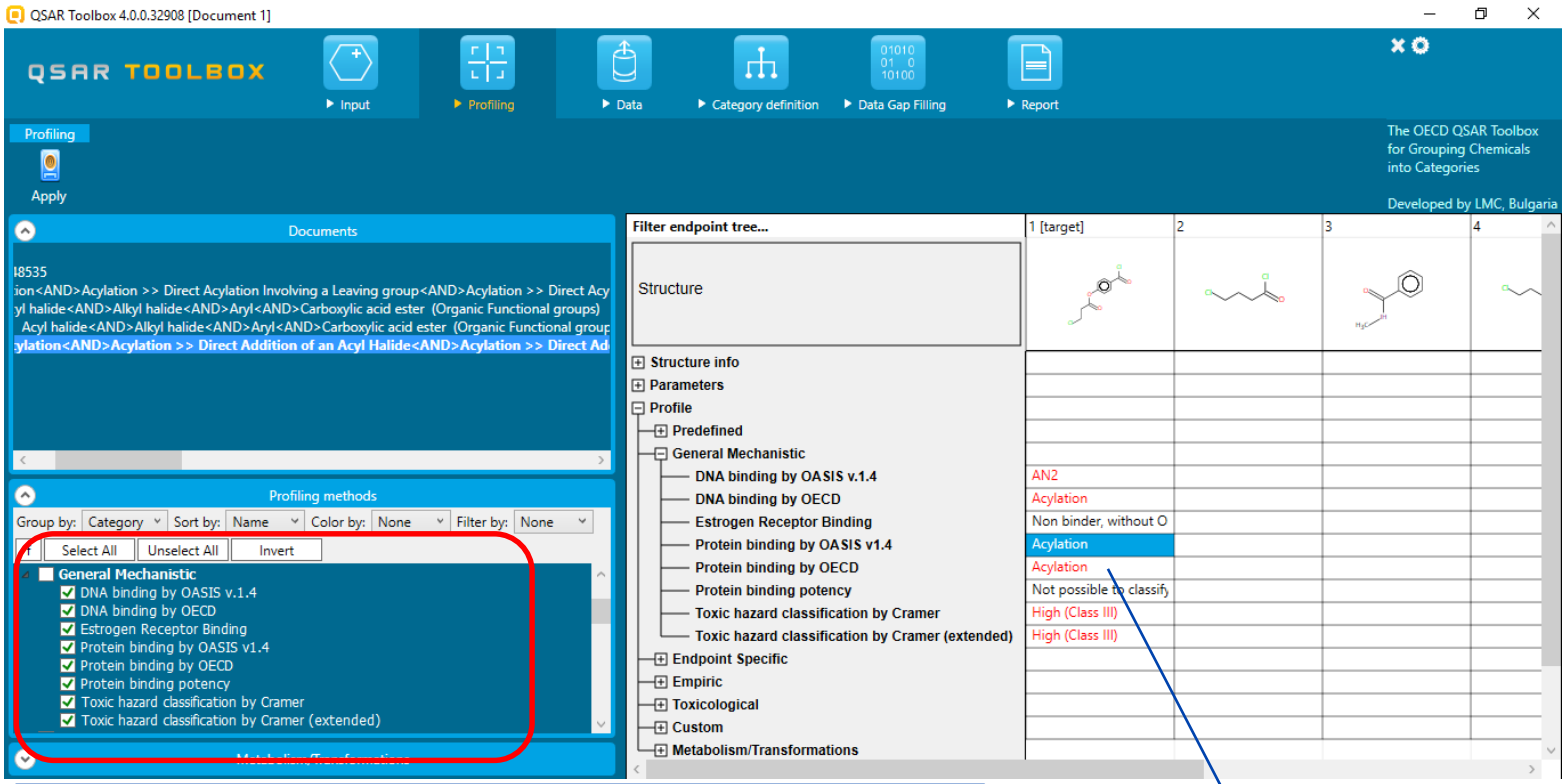
✓ Document tree



The document tree

- ✓lists the actions performed by the user
- ✓can be browsed to go back and forward in the workflow

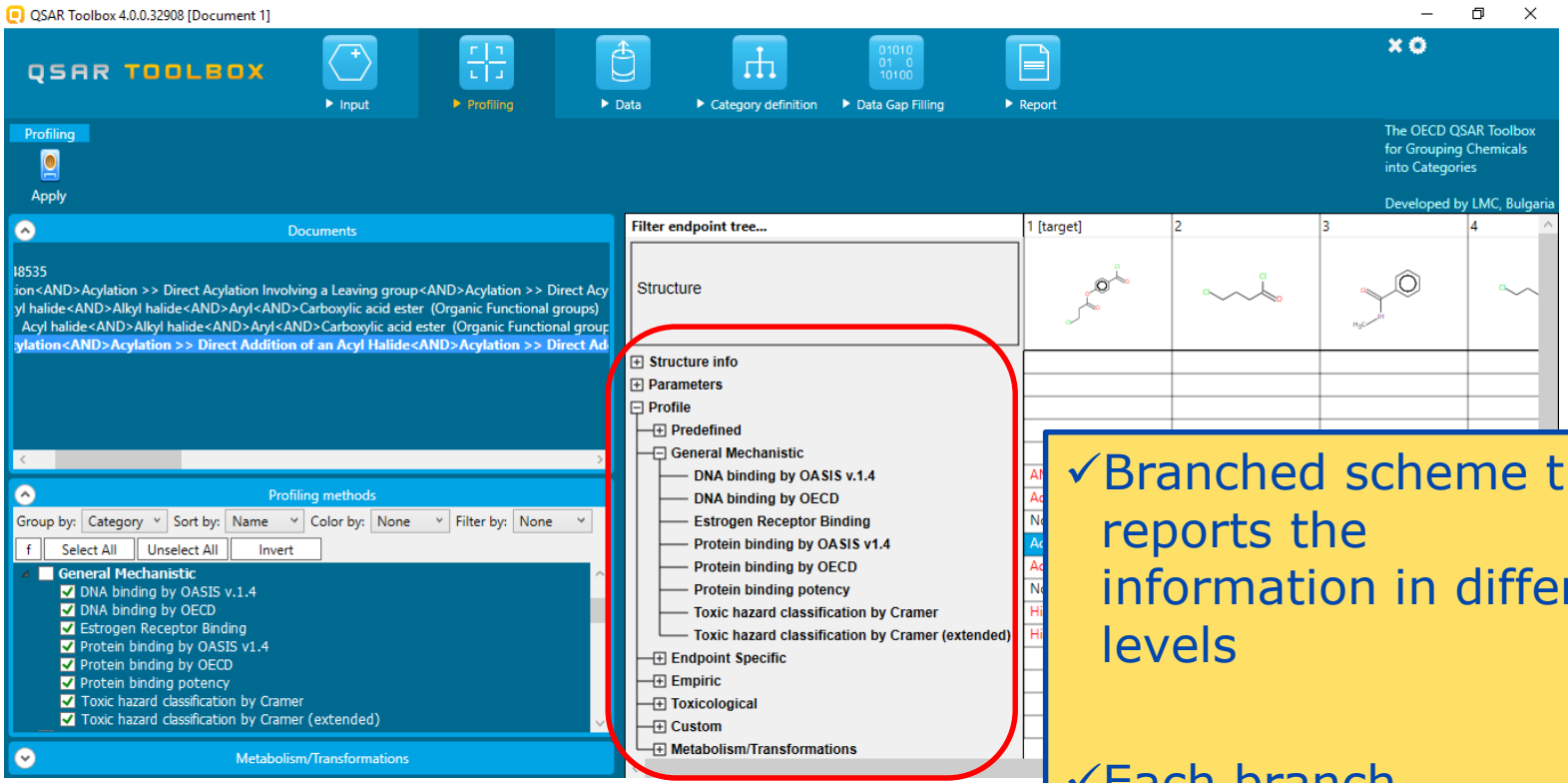
✓ Profilers



The profilers are algorithms for the identification of specific features of the chemicals

Outcome of the profiler

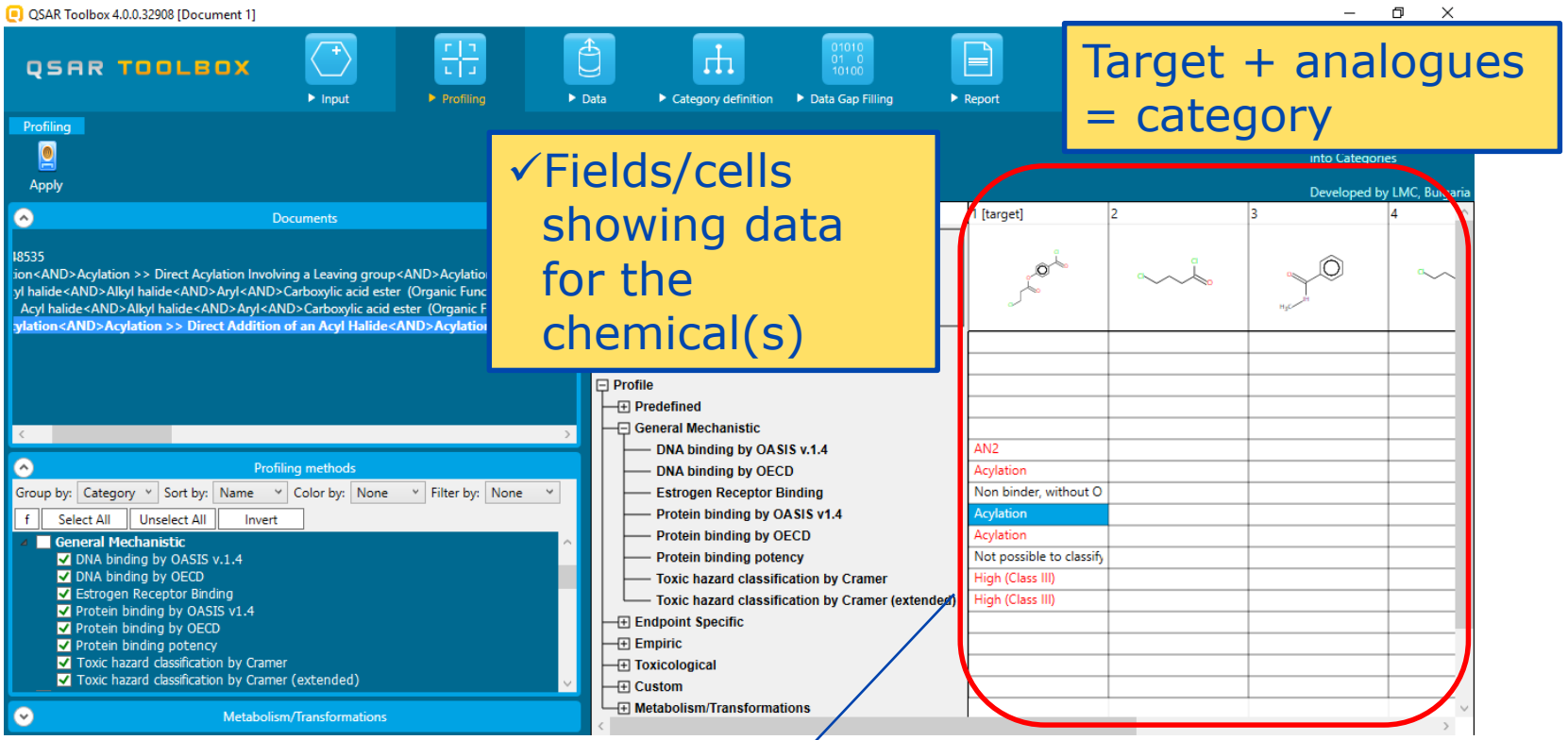
✓ Endpoint tree



✓Branched scheme that reports the information in different levels

✓Each branch corresponds to one cell in the data matrix

✓ Data matrix



NEW

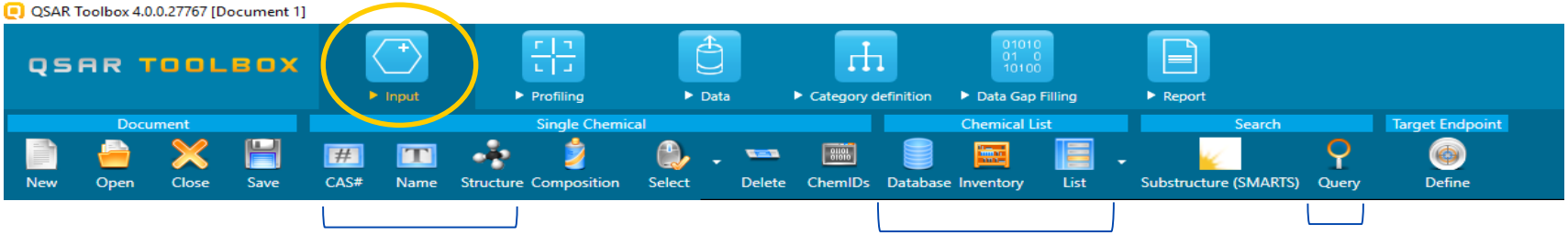
Now exportable in Excel format after making a prediction

Toolbox step by step

Insights into the "modules":

- ✓ **Core functions**
- ✓ **New features in version 4.0**

1. Input



Load the substance(s) into the system

Single chemical:

- Type CAS number
- Write the name
- Draw the structure
- Type SMILES

Chemical list:

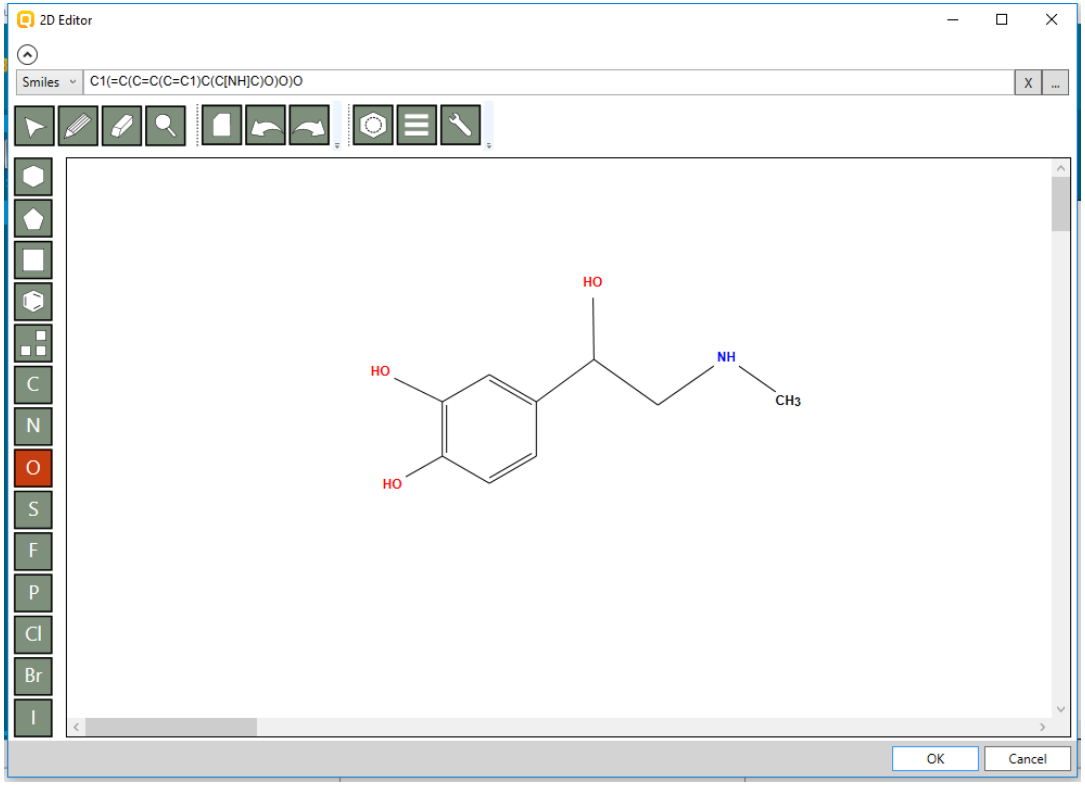
- Select chemicals from a list, database or inventory

Query Tool

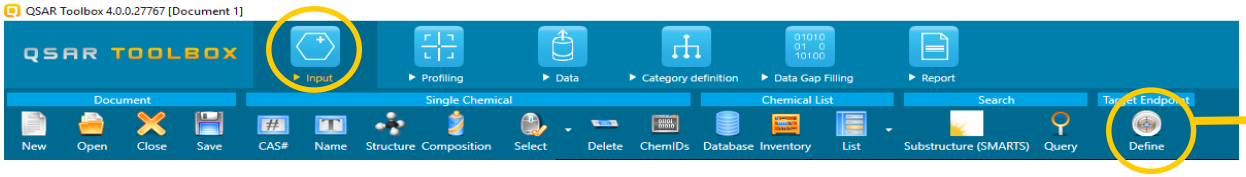
- Search for chemicals by specific criteria such as shared structural fragments

New drawing tool

- ✓ New interface
- ✓ More intuitive
- ✓ Easier to use



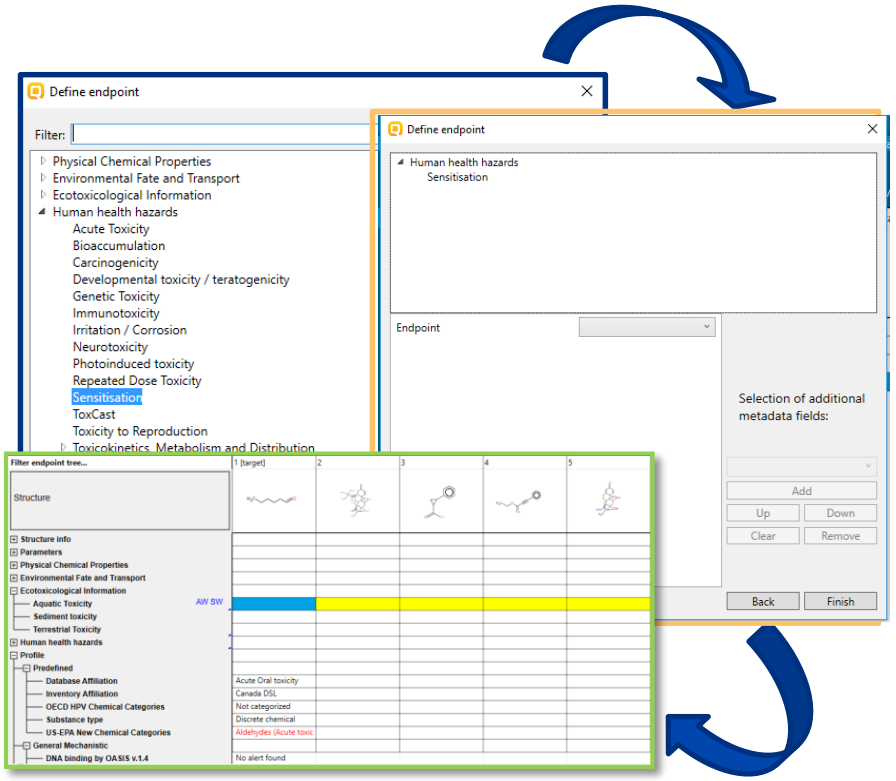
Definition of the target endpoint



In the input module, it is now possible to select the endpoint of interest (e.g. human health hazard, skin sensitization, EC3).

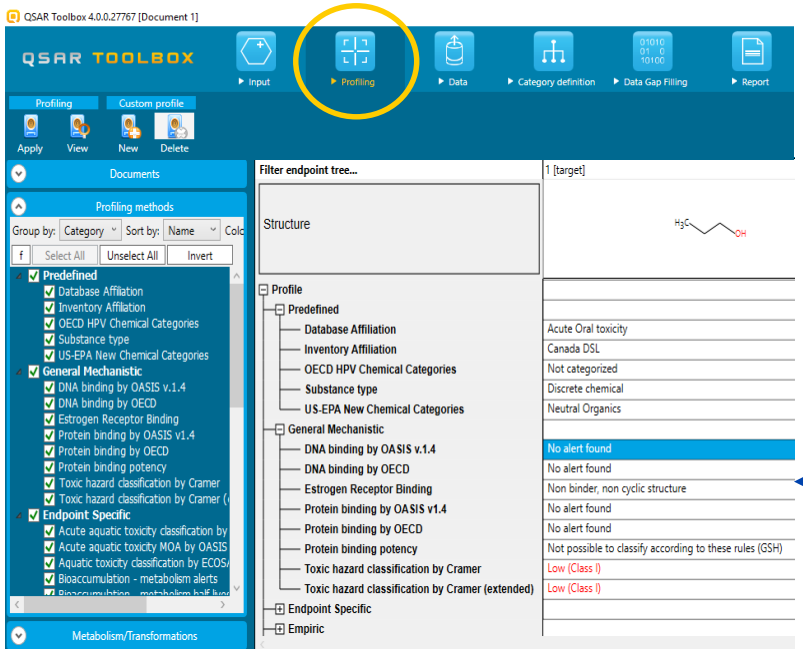
If selected, the QSAR Toolbox can guide the user through the next steps of the workflow:

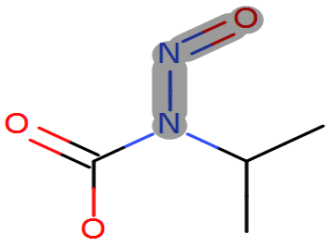
1. In the data matrix, the row corresponding to the selected endpoint will be highlighted
2. In the profiling module, the profilers will be highlighted with different colours depending on their relevance to the endpoint
3. In the data module, the databases containing data for the selected endpoint will be coloured in green



Alternatively, the endpoint can be selected directly from the endpoint tree in the data matrix

2. Profiling





Identify the characteristics of the chemical

- Profile according to different approaches (e.g. **mechanistic** or **structural** points of view)
- **Relevancy**: some profilers can be more relevant to some endpoint (e.g. DNA damage profiler for the endpoint mutagenicity)
- Choose the relevant profilers for your endpoint picking them from the blue list
- The outcome will be shown in the data matrix

Predefined (e.g. OECD HPV Chemical Categories, US-EPA New Chemical Categories)

General Mechanistic (e.g. DNA binding by OECD, Estrogen Receptor binding)

Endpoint specific (e.g. Acute aquatic toxicity classification by Verhaar)

Empiric (e.g. Organic Functional group US EPA)

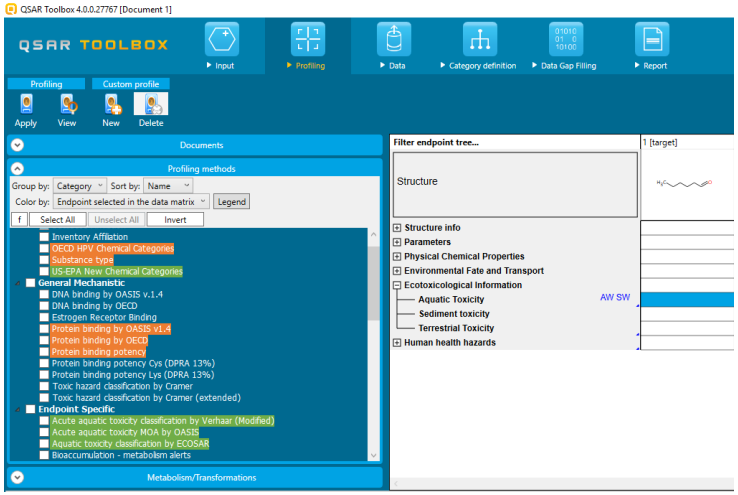
Relevancy of profilers

✓ The Toolbox highlights in different colours the most relevant profilers for the selected endpoint

Suitable – developed using data for the endpoint of interest

Plausible – developed using data somehow related to the endpoint of interest

Unclassified – developed using data not related to the endpoint of interest



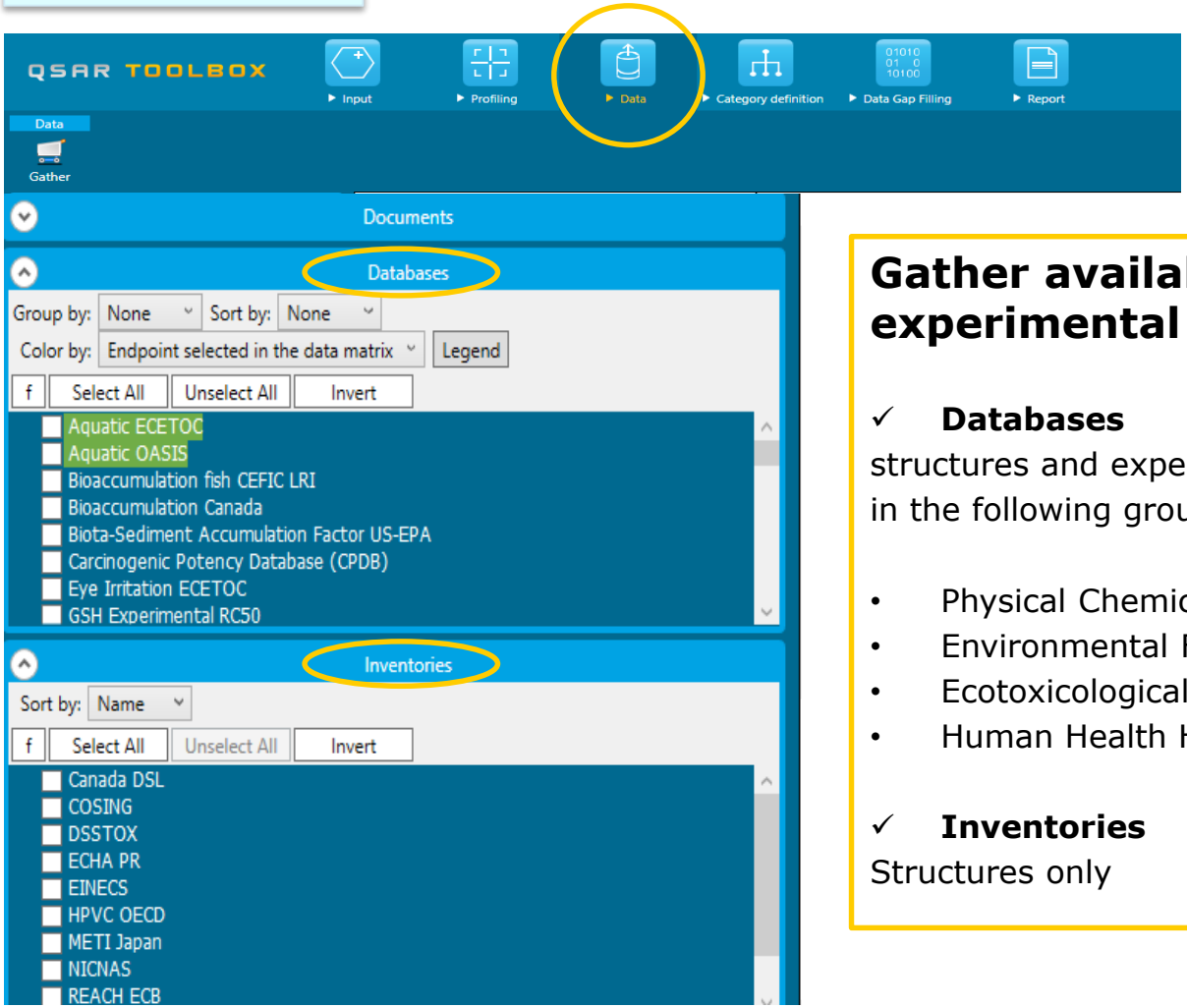
Updated profilers

✓ E.g. DNA binding by OASIS V 1.4; Protein binding by OASIS V 1.4; Acute Aquatic Toxicity MOA BY oasis; Acute aquatic toxicity classification by Verhaar; Organic Functional Groups

Simulation of metabolites: improved and extended features

✓ Observed rat liver metabolism with quantitative data

3. Data



Gather available structures and experimental results from:

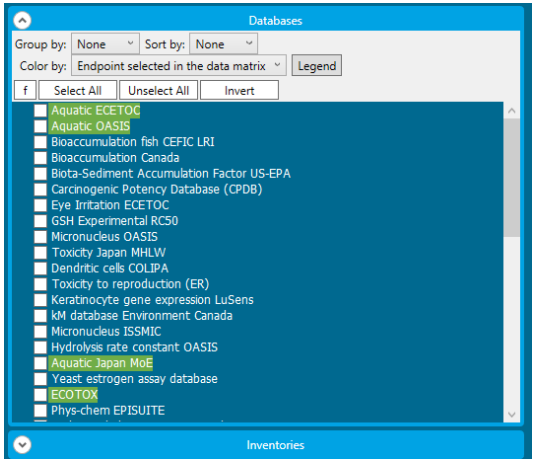
✓ **Databases**
structures and experimental data, organised in the following groups:

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

✓ **Inventories**
Structures only

Relevancy of databases

✓ The Toolbox highlights in different colours the most relevant databases for the endpoint of interests, i.e. the databases that contain experimental data for the that endpoint

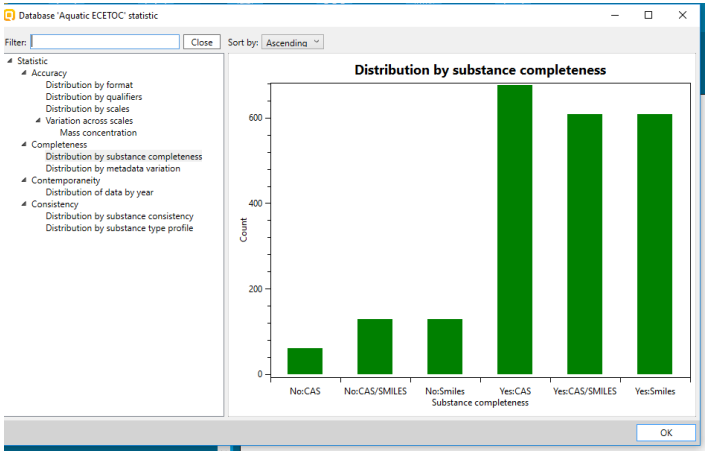


■ Have data for the target endpoint
□ Have no data for the target endpoint

Reliability scores and databases statistics

Quality attributes

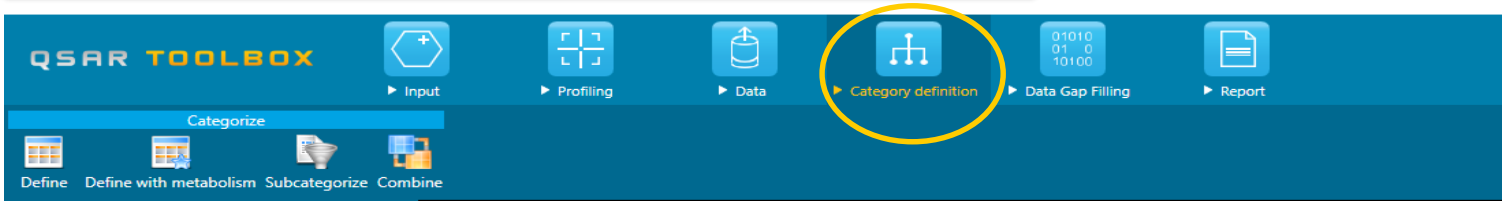
✓ **Accuracy**- distribution of data according to format, qualifiers, scales
✓ **Completeness**-availability of CAS and SMILES, additional info
✓ **Contemporaneity**- year of publication
✓ **Consistency**-how many SMILES are assigned to CAS, multiconstituent or UVCB substances



A huge amount of data in Toolbox 4.0!

Database content	Chemicals	Data points
Physical Chemical Properties	45 238	177 258
Environmental Fate and Transport	9 446	97 469
Ecotoxicological	17 649	856 473
Human Health	30 447	912 687
Total number	**79 204**	**2 043 887**

4. Category definition



Creation of a group of analogues (category) around an input chemical

Usually, grouping according to:

- **Specific mode of action** (e.g. protein binding)
- **Structural similarity** (e.g. functional groups)
- **Predefined categories** (e.g. OECD HPV chemical categories)

The substances are grouped according to the selected profilers' outcome

The substances and data are retrieved from the selected databases ("Data" module)

The profilers have different **relevancy** in relation to the endpoint of interest. Using relevant profilers for the category definitions and subsequent sub-categories leads to the selection of better analogues.

Relevancy of profilers

✓ The relevancy of profilers is also shown in the "Category Definition" module

- ☑ Suitable
- ☑ Plausible
- ☐ Unclassified

Grouping with accounting metabolic transformations - extended function

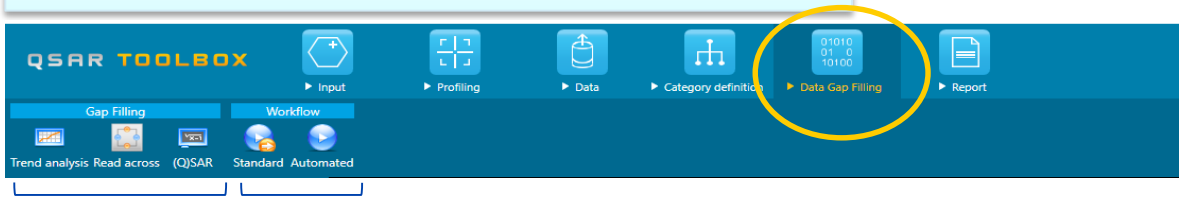
In some cases, the toxicity of a substance could be triggered by its breakdown products or metabolites, even if the parent compound itself might be not toxic.

The Toolbox is able to create categories taking transformations into account.

Analogues could be retrieved according to the following criteria:

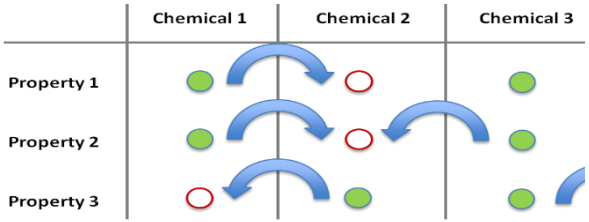
- Parent compound and metabolite with a defined profile
- Metabolite(s) with a defined profile
- Metabolite(s) in common
- Metabolite(s) similar to a defined one

5. Data gap filling

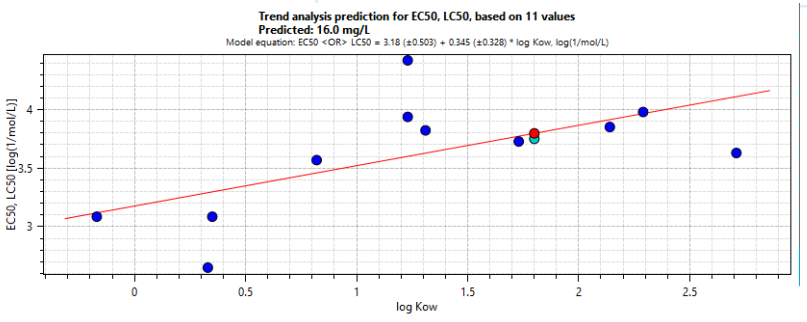


Predict a missing value for the target chemical

Read across	Prediction obtained by using data of the closest analogues in the category. The value can be calculated as average, maximum, minimum, etc.	▪ One or few analogues in the category
Trend analysis	All data form the analogues are used to derive a monovariate regression equation that describes the trend between the endpoint (e.g. LC_{50}) and a selected parameter (by default log Kow). The equation is then applied to fill the data gap for the target chemical.	▪ Many analogues in the category ▪ Not suitable for qualitative endpoints
QSAR models	Use a library of external QSAR models provided and directly accessible from the Toolbox interface (e.g. EPISUITE models)	▪ When adequate analogues for read RA and trend analysis are not found and in WoE considerations



E.g. LC_{50} = f(log Kow)



Prediction workflows in v4.0

How can the Toolbox guide the user to generate a prediction?

Different levels of interaction with the Toolbox are available

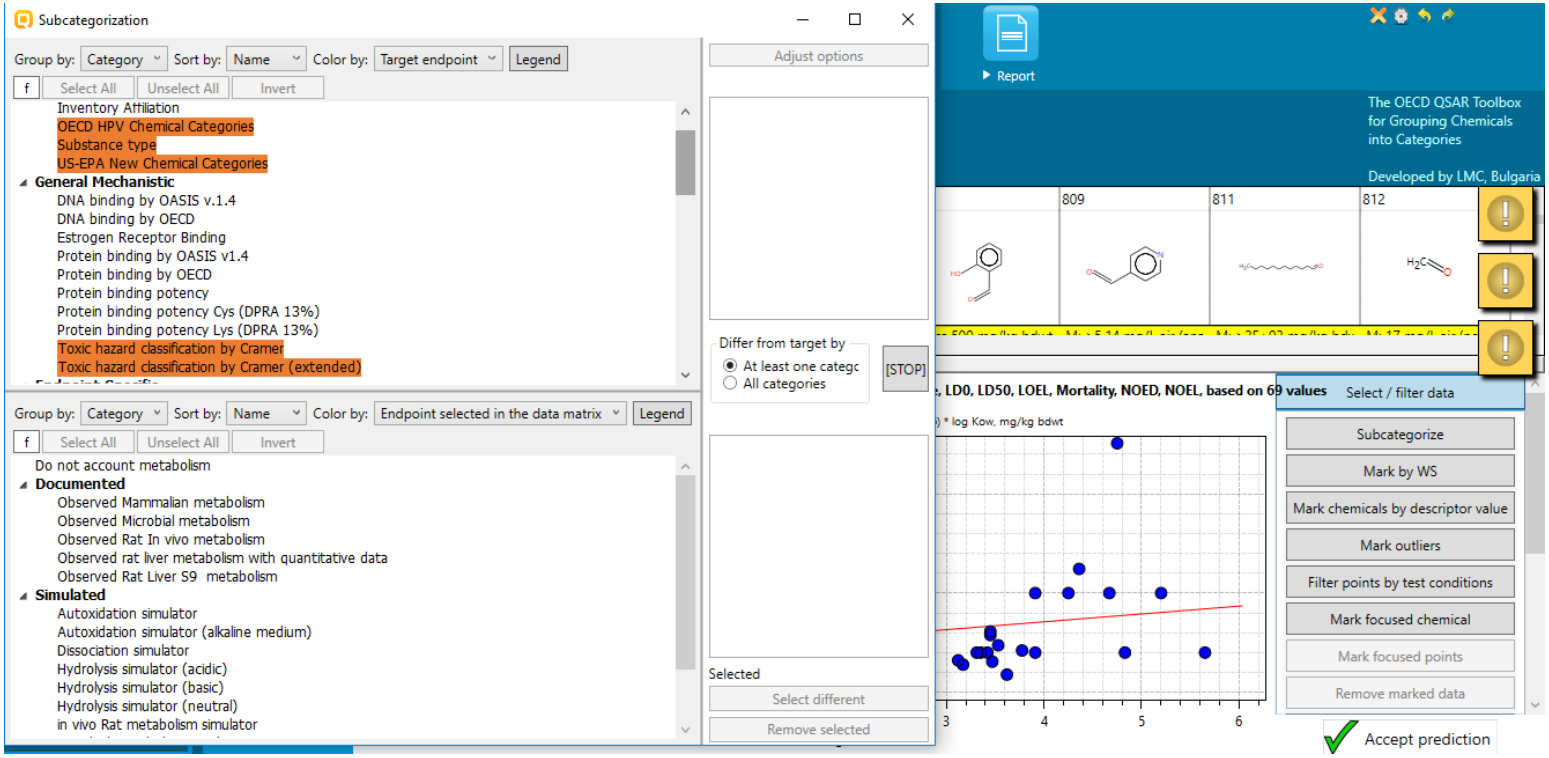
1. **Manual prediction (Classical way)**
2. **Standardized Workflow (SW) – new**
3. **Automated Workflow (AW) - new**

ECHA
EUROPEAN CHEMICALS AGENCY

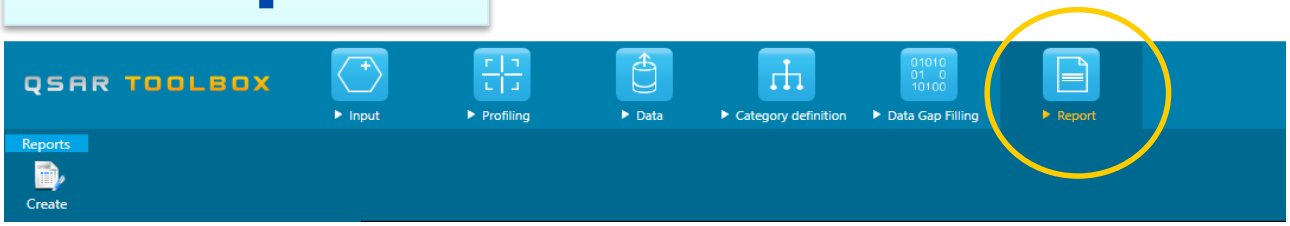
Relevancy of the profiler for subcategorization

✓ When refining a category, the Toolbox highlights the most suitable profilers to choose

- Suitable
- Plausible
- Unclassified



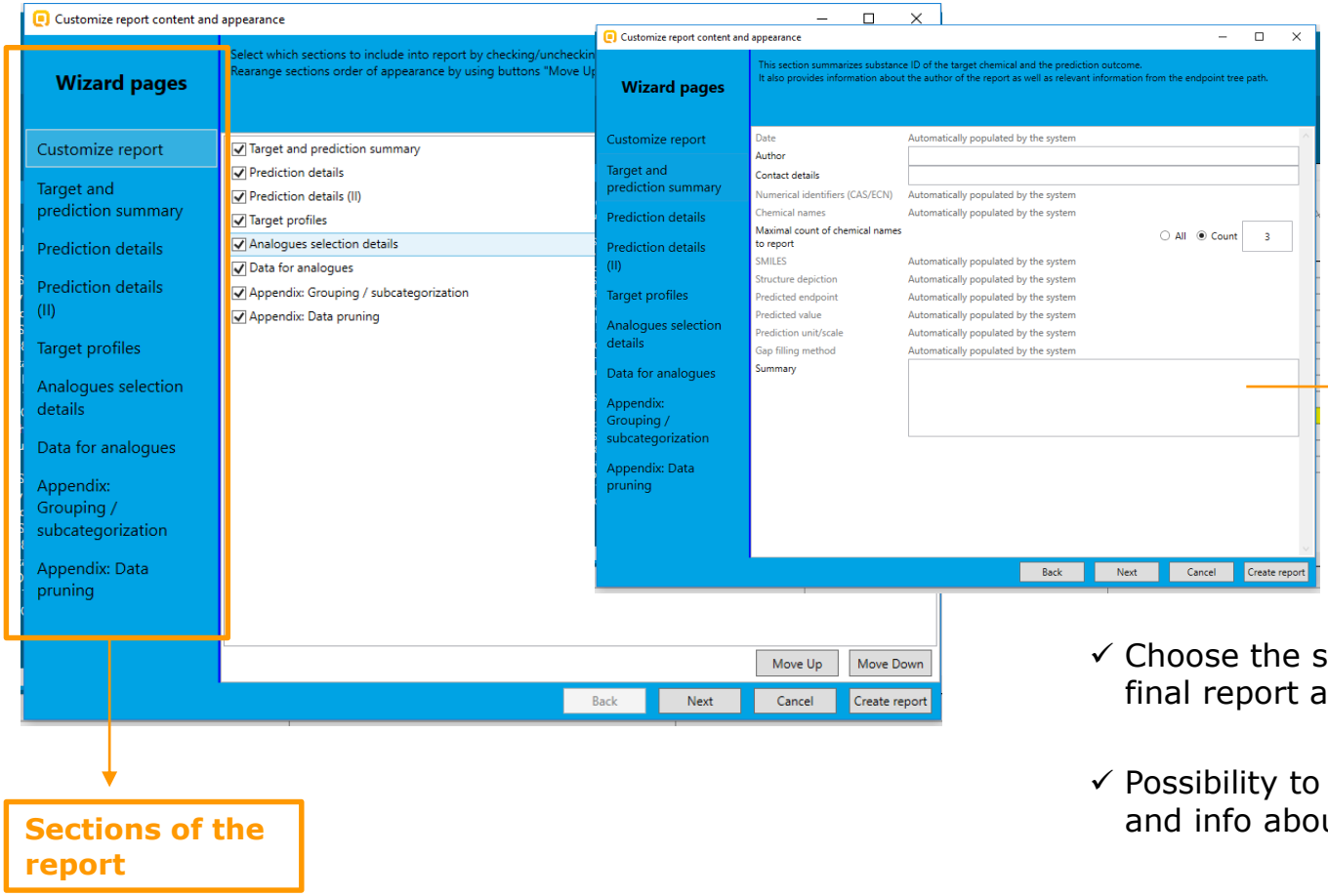
6. Report



Create a report after accepting the prediction

- ✓ The Toolbox can generate reports for the predictions
- ✓ The user needs to critically look at the results and manually complete some of the fields related to comments and explanations

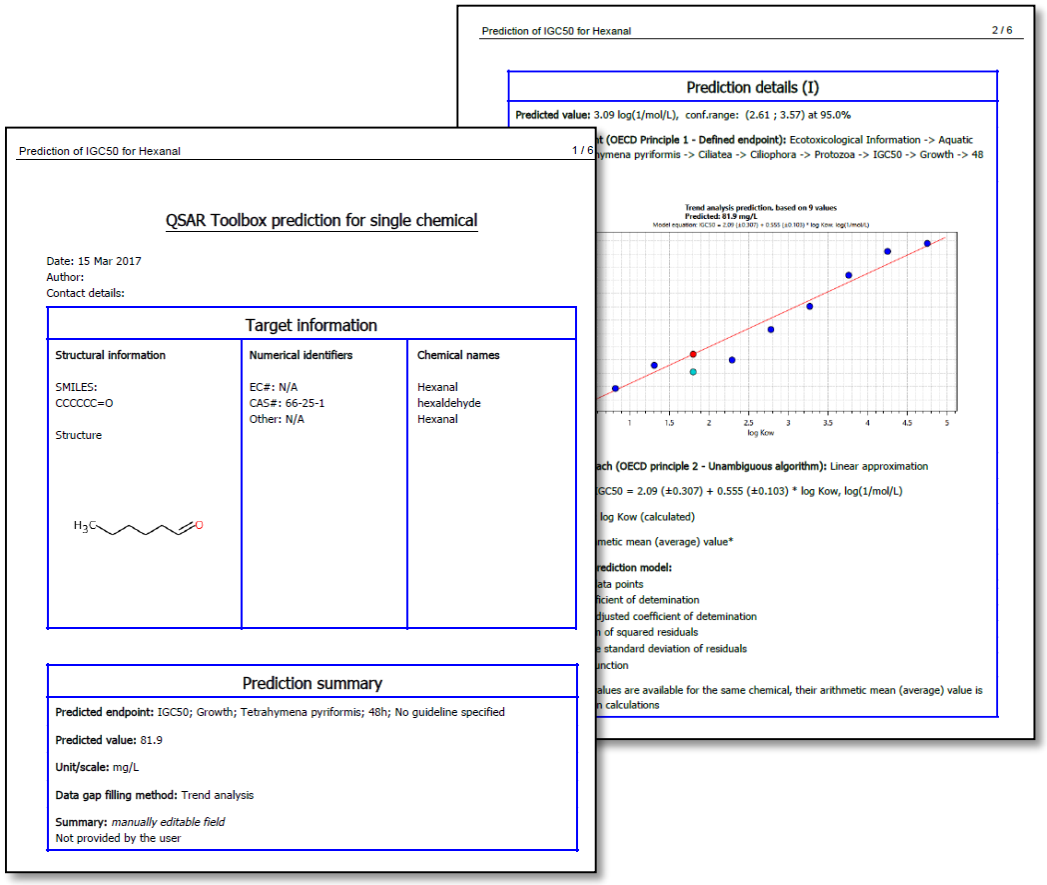
Wizard pages to customize the report



- ✓ Choose the section to include in the final report and their order
- ✓ Possibility to include in the report data and info about the analogues
- ✓ Editable fields for comments and interpretation of the results

New Report format

- ✓ Completely renewed
- ✓ Clear, simple and straightforward



ECHA
EUROPEAN CHEMICALS AGENCY

Exportable data matrix in Excel format

Target Chemical

Analogues

Parameters

Endpoint data

Additional endpoints data

	A	B	Target chemical			Analogue #1			Analogue #2			Analogue #3		
1														
2	**Substance identity**													
3	Structure													
4	CAS number		66-25-1			123-38-6			124-19-6			123-72-8		
5	Chemical name		Hexanal			PROPANAL			NONANAL			CCCC=O		
6	Other identifier		CCCCCC=O			CCC=O			CCCCCCCCC=O			CCCC=O		
9	Parameters	unit												
10	Boiling point	°C	132.2			59.78			196.48			84.82		
11	Molecular Weight	Da	100.15528			58.07764			142.23292			72.10352		
12														
13	**Measured and predicted data**													
14	***Data used for prediction***													
	ent	endpoint	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, typ of method, assay, strai test guideline, year*
	oxicity	IGC50				216	mg/L	Toxicology methods, 7, 289-309. Schultz, T.W.	22	mg/L	Toxicology methods, 7, 289-309. Schultz, T.W.	194	mg/L	Toxicology methods, 7 289-309. Schultz, T.W.
16														
17	***Ecotoxicological Information/Aquatic Toxicity; LC50; Mortality; Pimephales promelas; 96 h***													
	t	endpoint	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, type of method, assay, strain, test guideline, year*	value	unit	*reference, species, duration, test type, typ of method, assay, strai test guideline, year*
19	Aquatic Toxicity	LC50	22	mg/L	Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI;332 p	130	mg/L	Environmental Toxicology and Chemistry, Vol. 16, No. 5, pp. 948–967				14.7	mg/L	Environmental Toxicology and Chemistry, Vol. 16, No 5, pp. 948–967
20	Aquatic Toxicity	LC50	17.8	mg/L	Environmental Toxicology and Chemistry, Vol. 16, No. 5, pp. 948–967							16	mg/L	Center for Lake Superior Environmental Studies University of Wisconsin-Superior, Superior

Practical applications

Installation

- The Toolbox requires the installation of a PostgreSQL database, the QSAR Toolbox application itself and Metapath

- Read the Installation manual before starting!

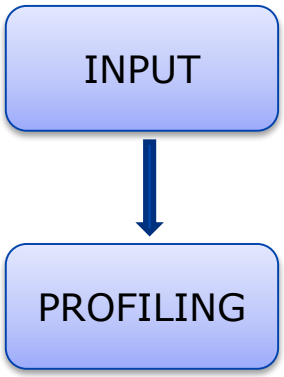
What can you do with the Toolbox?

- The live demo will demonstrate how modules can be used in different combinations (workflows) for various purposes:

1. Profiling
2. Metabolism
3. Data
4. Category
5. Prediction

Profiling

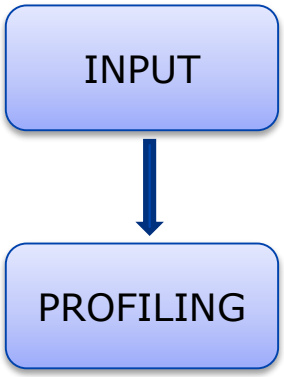
- The "Profiling" workflow gives you a summary of the relevant properties of your substance(s)
- The profilers are used as criteria to find analogues, but they are also useful for preliminary screening or prioritisation of substances



1. Input one or more substances of interest
2. Go to the Profiling module
3. Select all profilers and execute them
4. Read the outcome (colour code helps, red for alerts)
5. Export the table (useful for batch calculations)

Metabolism

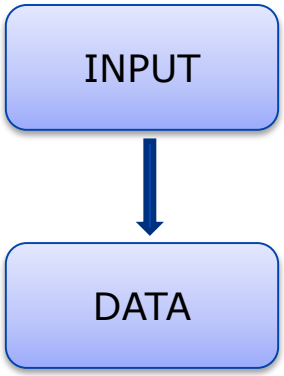
- The "Metabolism" workflow gives you the metabolic and abiotic products observed or predicted for your substance
- Sometimes, metabolites and transformation products cause toxicity



1. Input one or more substances of interest
2. Go to the Profiling module
3. Select all the choices available in the Metabolism windows and execute
4. Read the outcome (the number of metabolites is indicated in the cells, double click on results to see the structures)
5. Run workflows for the metabolites and/or export the structures (if needed)

Data

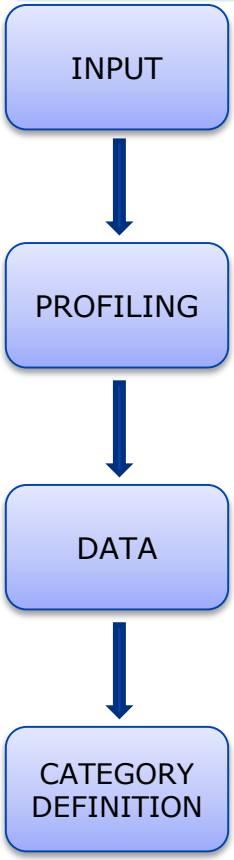
- The "Data" workflow gives you all experimental information available in the Toolbox for your substance(s)
- The Toolbox also includes details and references of the study



1. Input one or more substances of interest
2. Go to the Data module
3. Select all databases and execute
4. Read the outcome
5. Double click on the data to read more experimental details about that specific result
6. Export the table (if needed)

Category

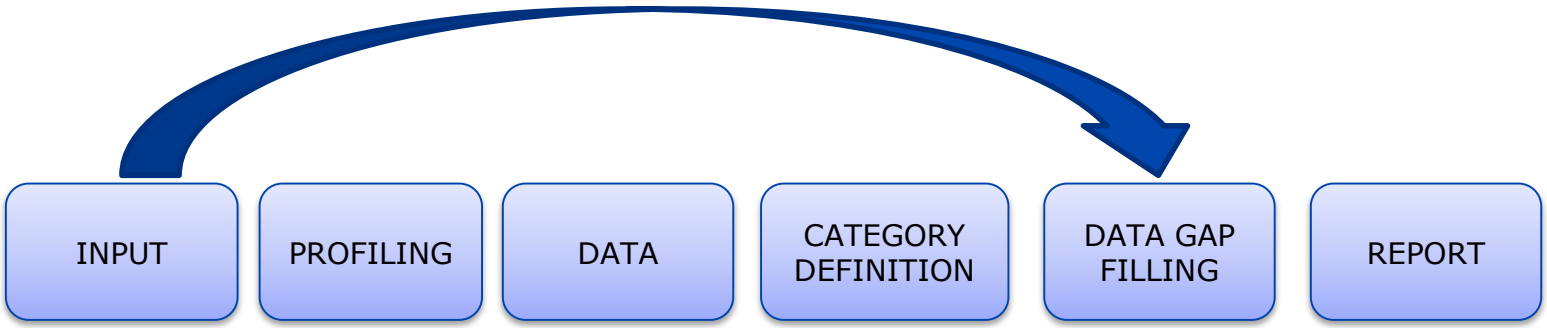
- The "Category" workflow helps you in finding analogues for your substance and the available experimental data for them
- It is useful for identifying analogues and data gaps



1. Input the substance of interest
2. Go to the Profiling module, select all profilers and execute
3. Go to the Data module, select all databases and execute
4. Go to the Category definition module, choose one profiler of interest and click "Define". The Toolbox will find analogues that share the same profiler outcome
5. Repeat the previous point if you want to sub-categorise
6. Read the outcome (and export the table, if needed)

Prediction

- The Prediction workflow uses all the modules in the Toolbox to fill data gaps with trend-analysis or read-across
- The new automatic workflow directly connects Input and Data gap filling



How to run the new workflows?

1. Input your target
2. Go to data gap filling
3. Select automated or standardised workflow
4. Follow the wizard

The different workflows

Automatic workflow: input the chemical and obtain a prediction

- Input the chemical and select the workflow for the endpoint you want to predict

Standardised workflow: input the chemical and be guided in each step

- You can choose among the options proposed by the Toolbox

Manual prediction: as in Toolbox v.3, all up to you

- If you were familiar with the previous version, you can still use the Toolbox in the same way. Nevertheless, you can now select an endpoint and activate colour coding for getting help in the databases and profilers

Useful resources

- ECHA's website with practical examples: https://echa.europa.eu/it/support/oecd-qsar-toolbox
- QSAR Toolbox Website: https://www.qsartoolbox.org/

Upcoming trainings on the Toolbox:

- Webinars (one hosted by Chemical Watch planned for the end of April, more from ECHA and OECD will come)
- OECD QSAR Toolbox Training (by LMC, commercial, 12-16 June 2017) http://www.reachmonitor.com/index.php?lang=2&aptd=4&id=55

Live demo of Toolbox 4.0

Conclusions

- The Toolbox is not just another QSAR predictor, it is complete decision supporting system for hazard assessment, therefore the user needs sufficient understanding of (eco)toxicology.

- The Toolbox combines experimental data with knowledge based profilers and transformation simulators allowing adjustment of the prediction to particular needs.

- The Toolbox is transparent, therefore predictions can be easily verified.

Acknowledgments

Special thanks to the Development team from *Laboratory of Mathematical Chemistry led by Prof. Mekenyan for the determination and hard work dedicated to have Toolbox v4.0 released*

ECHA
EUROPEAN CHEMICALS AGENCY

LIVERPOOL JOHN MOORES UNIVERSITY

Thank you!

tomasz.sobanski@echa.Europa.eu
andrea.gissi@echa.europa.eu
marta.sannicola@echa.europa.eu

Subscribe to our news at
echa.europa.eu/subscribe

Follow us on Twitter
@EU_ECHA

Follow us on Facebook
Facebook.com/EUECHA