Computational approaches to support chemical hazard identification: application of the QSAR Toolbox software to case studies

Chiara Laura BATTISTELLI

Environment and Health Department-

Istituto Superiore di Sanità

Content

1. Introduction to in silico methods, focusing on QSAR approaches (such as models, category/grouping, read across, TTC)
2. Introduction to the QSAR Toolbox software:
   - ✓ general introduction (main functionalities)
   - ✓ step by step, with practical use of the tool
3. Practical exercise on case studies with QSAR Toolbox software (in working groups)

Tuesday 17.10.2023 – Marotta Room

9:00 – 10:00	Q&A Case studies of the previous day (Trainers: Dr. Emanuela Testai, Dr. Maura Manganelli, Dr. Simona Scardala Scardala)
10.00 – 11.30	Introduction to *in silico* methods: QSAR, read across, TTC, and the related tools (QSAR Toolbox) characterization (Trainers: Dr. Chiara Battistelli, Dr. Cecilia Bossa, Dr. Olga Tcheremenskaia)
11.30 – 13.00	Presentation of tool functioning (Trainers: Dr. Chiara Battistelli, Dr. Cecilia Bossa, Dr. Olga Tcheremenskaia)
13.00 – 14:30	Lunch
14.30 – 16:30	Practical exercise on case studies, in working groups with QSAR Toolbox software (Trainers: Dr. Chiara Battistelli, Dr. Cecilia Bossa, Dr. Olga Tcheremenskaia)
16.30 – 17:00	Discussion of exercises

Toxicologial studies
(for hazard identification)

Animal testing
in vivo assays

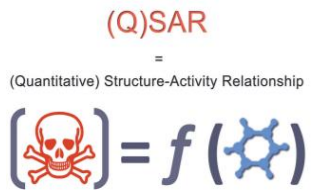
Alternative methods*
(Non animal models)

Experimental models
(*in vitro*, *ex vivo*):
cells, organs, tissues or enzymes

Non testing models
(computational or in silico)
- Speed-up number of evaluated chemicals
- Save of money and time
- Animal welfare

* According the 3Rs principles: *Refinement, Reduction, Replacement* (*Russel and Burch*, 1960)

(Q)SAR approaches



- Methodologies based on the concept that a property (such as toxicity) is a function of chemical structure
- Used to predict (physical chemical or (eco)toxicological) properties, on the basis of chemical structure (Known)
  - **QSAR models:** quantitative structure–activity relationship models. Mathematical equation linking the biological activity to chemical structure, identified by **molecular descriptors** (i.e. physical chemical or other molecular properties)
  - **SAR models**: qualitative structure–activity relationship models (i.e. Structural Alerts)
  - **Chemical grouping and categories**: approach considering more than one chemical at the same time, sharing similar characteristics and physical chemical or (eco)toxicological properties or following a trend, as results of structural similarity **.
  - **Read-Across**: technique used for filling data gaps by predicting endpoint information for one (or more) chemical(s) (*target chemical*), using data for the same endpoint from one or more similar substances (*source chemicals*).

Threshold of Toxicological Concern (TTC): applied to derive a threshold for exposure, below which a toxic effect on human health by the compound is not expected; based on databases on general toxicity

Good-quality data

(*) **Structural alerts (SAs):** functional groups or structural features associated with a potential reactivity for a defined endpoint

(**) **Chemical similarity** is not limited to structural similarity but should consider factors that drive a given toxicity and how these can be linked back to chemical properties or features, e.g. reactivity."

(Q)SAR models

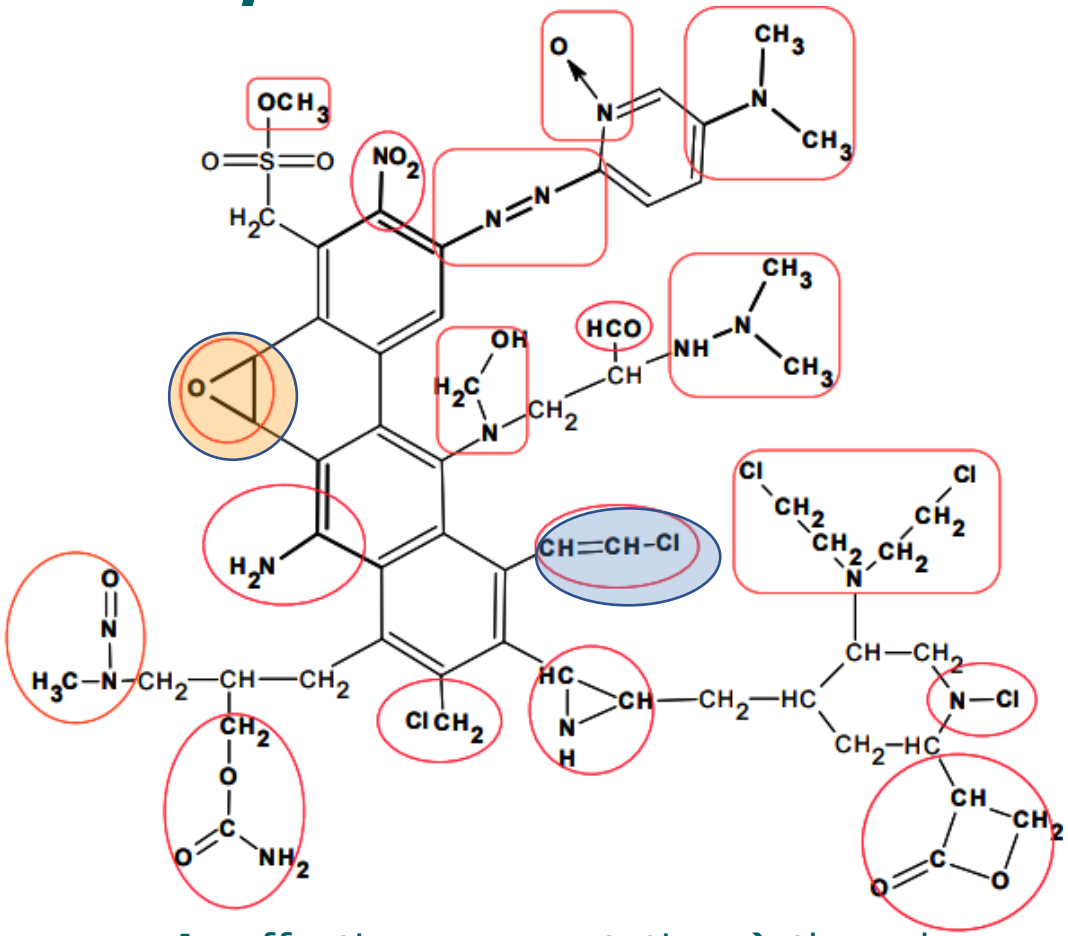
- **Local models**: built on congeneric set of chemicals, i.e., chemicals with similar structure with the same mechanism of action; generally more effective, but with a narrower domain of applicability (strictly related to the set of chemicals used for the model definition)
- **Global models**: often implemented in software tools, ensure a broader applicability, to more than one chemical class.
  - ✓ **Statistical-based**: use machine learning techniques to associate structural features and chemicals activity. These models are data-driven, without expert supervision.
    - ✓ Negative predictions are more accurate.
  - ✓ **Rule-based (expert system):** recognition and codification of functional groups or structural features associated with a potential reactivity for a defined endpoint (**Structural alerts,** SA). SA are rarely defined by an applicability domain (absence of SAs, is not absence of toxicity, but a lack of knowledge)
  - ✓ **Hybrid**: integrate both expert knowledge and statistically derived rules, trying to overcome disadvantages of both approaches.

Structural alerts (SAs): functional groups or structural features associated with a potential reactivity for a defined endpoint

SAR models: the case of SAs for genotoxicity

1. Genotoxic carcinogens: DNA reactivity, direct or after metabolic activation
2. The basis of the alerts: Miller[1] teory, electrophilic species are able to react with nucleophilic sites in DNA
3. Ashby[2] compiled a list of SA (based on experimental data), which if present in a molecule, give it potential reactivity:
   - Direct acyls (acyl halides, ß-lactones..)
   - Direct alkylants (epoxides and aziridines, lactones, nitrites, α-ß unsaturated carbonyls, simple aldehydes, quinones..);
   - Indirect alkylating agents (mono-halogen alkene, hydrazines...).
   - Intercalants and forming adducts (PAHs, aromatic hydrocarbons...)
   - Amino aryls that form adducts (aromatic amines, amides, ..)

1. Miller and Miller, 1981; 2. Ashby et al, 1988



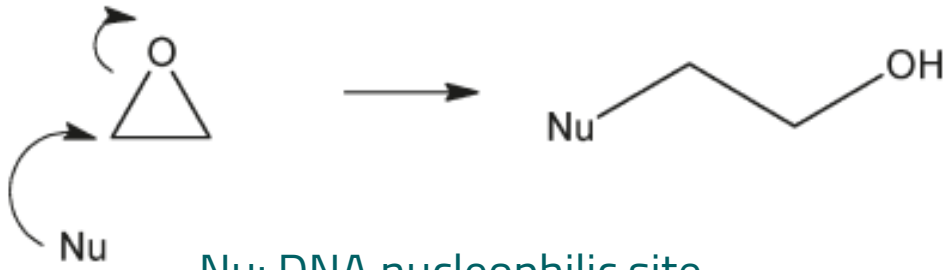
An effective representation → the poly-carcinogen (imaginary molecule)

Each SA codes for a chemical class → specific mechanism of action

ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

Mechanism of direct and indirect alkylants

Epoxide: Direct alkylating agent



Nu: DNA nucleophilic site

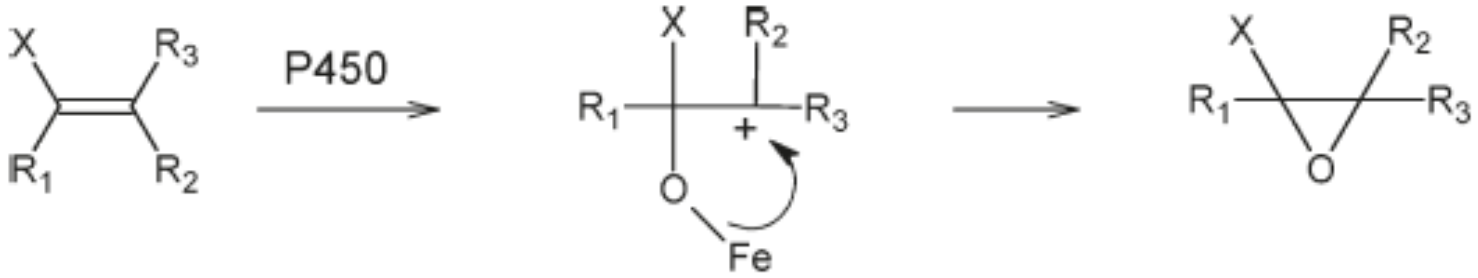
CHEMICAL REVIEWS
dx.doi.org/10.1021/cr100222q | *Chem. Rev.* 2011, 111, 2507–2536
REVIEW
pubs.acs.org/CR

Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology

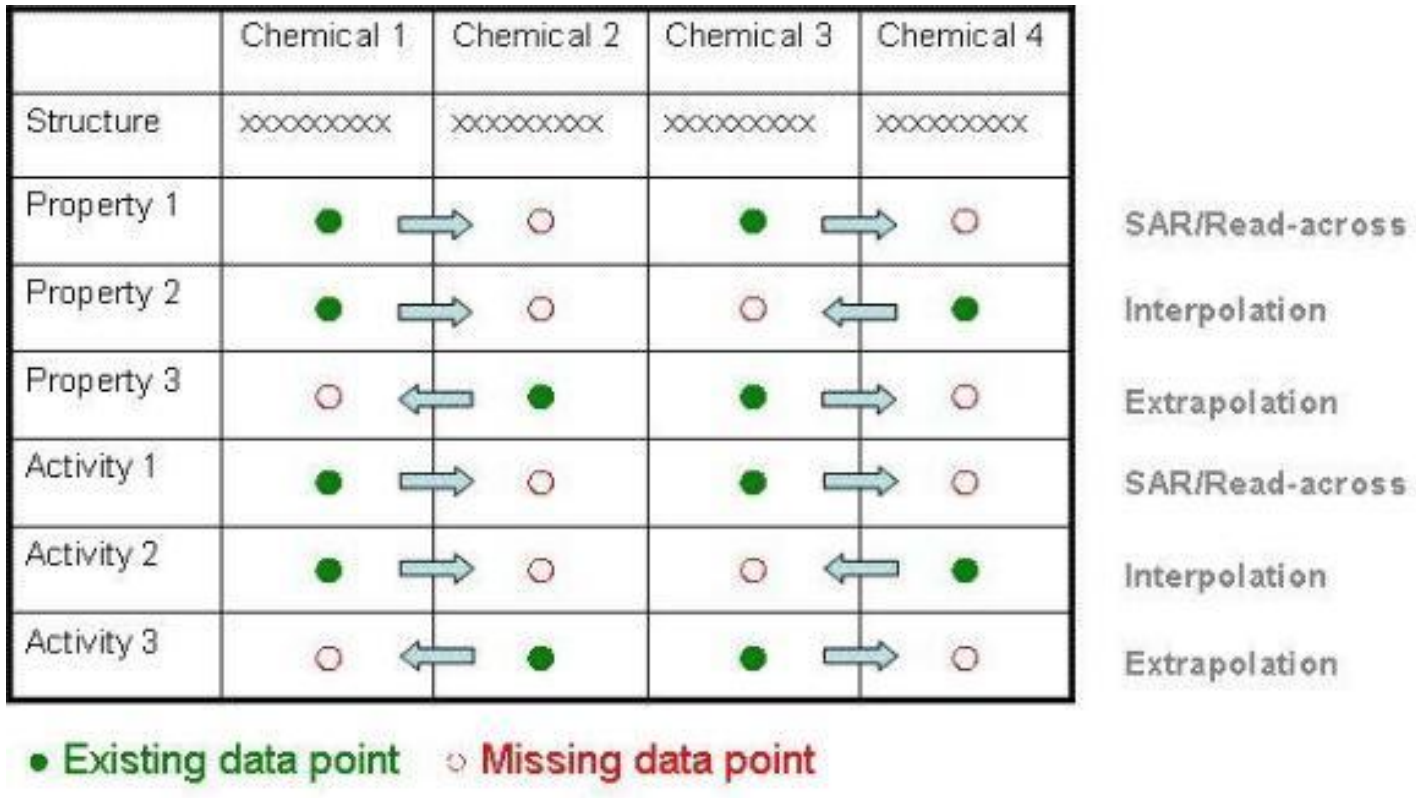
Romualdo Benigni* and Cecilia Bossa

Istituto Superiore di Sanita', Environment and Health Department, Viale Regina Elena, 299 00161 Rome, Italy

Mono halogen alkene: indirect alkylant

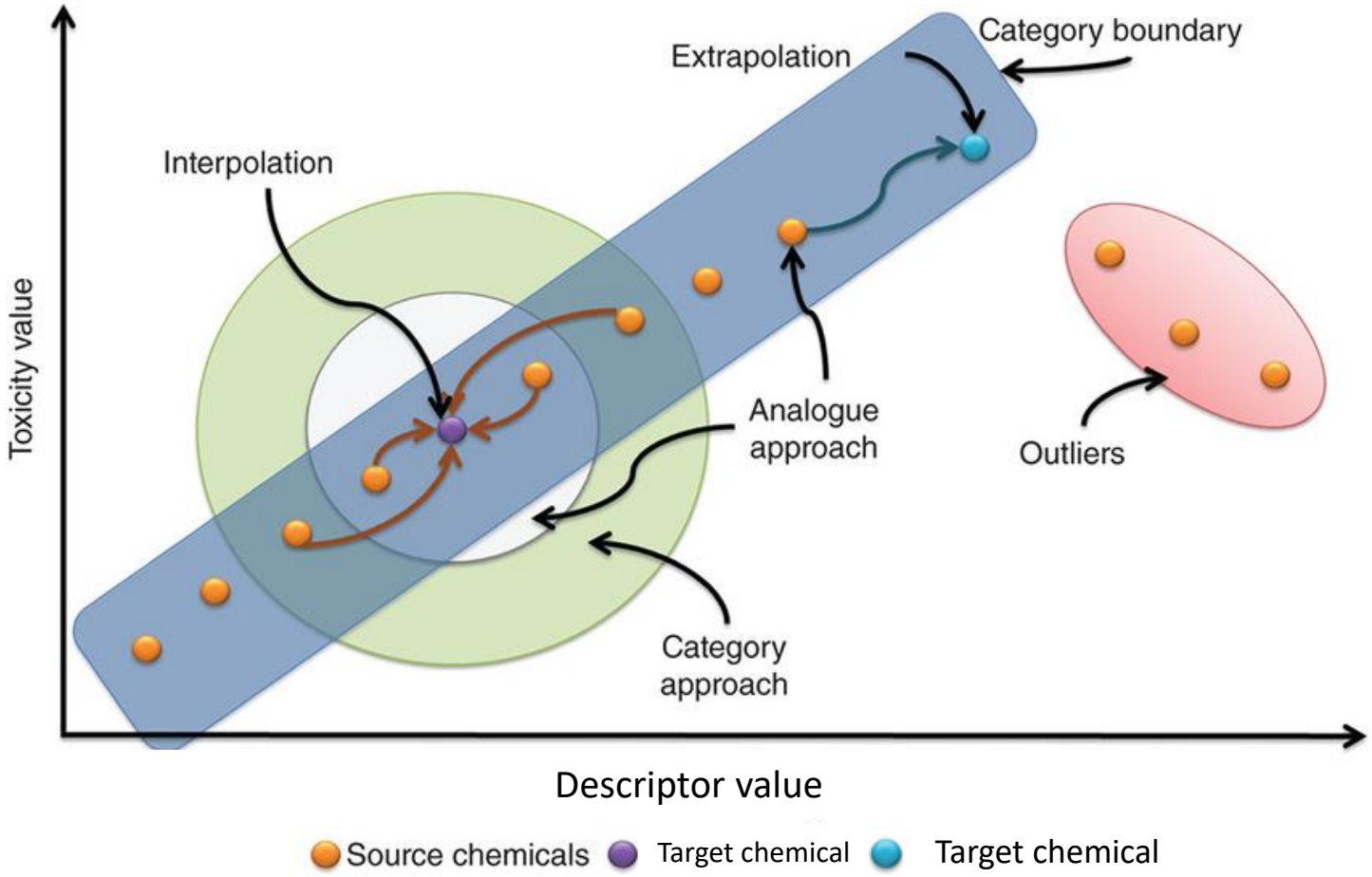


Grouping of substances and Read-across



- Substances that have similar properties may be considered as a **group**, or **category**
- If we have data for one, or more chemicals (**source chemicals**), they can be read-across to fill the data gaps for substances with no data (**target substance**)
- Structural similarity is a pre-requisite (for any grouping and read-across approach).
- Similarities may be due to different factors such as:
  - common functional group;
  - common precursors and/or common breakdown products
  - constant pattern in the changing of the potency of the properties across the group

Grouping /category and Read-across



Category approach: chemicals sharing similar characteristics and (eco)toxicological properties are grouped together. The assessment concerns the category as a whole because data gaps may exist for different category members and different endpoints.

Analogue approach: assessment of one specific chemical, using for the prediction experimental data from others (one or more) similar substances

Trend analysis is a method of predicting toxicity of a chemical by analyzing toxicity trends (increase, decrease, or constant) of tested chemicals.

Source chemicals: substances with experimental data, considered similar to the target
Target substance: substance to predict (with no experimental data)

Raies and Bajic, 2016, doi: 10.1002/wcms.1240

Threshold of Toxicological Concern (TTC)

- Pragmatic approach used in risk assessment, in the absence of chemical-specific toxicity data
- Used to derive a threshold for exposure, below which a toxic effect on human health is not expected
- Based on a general toxicity database (for oral exposure)*
- Used in some regulatory framework (such as EMA, FDA, EFSA, ECHA, SCCS, …) if exposure is low
- TTC is not strictly a QSAR approach, but a non testing methods which uses structure-activity relationships

Domain of applicability:

- Applicable to organic compounds with known chemical structure
- Not applicable (or recommended) to high potency carcinogens and highly bioaccumulating substances, inorganics, metals, metal containing compounds, polymers, proteins, nanomaterials…

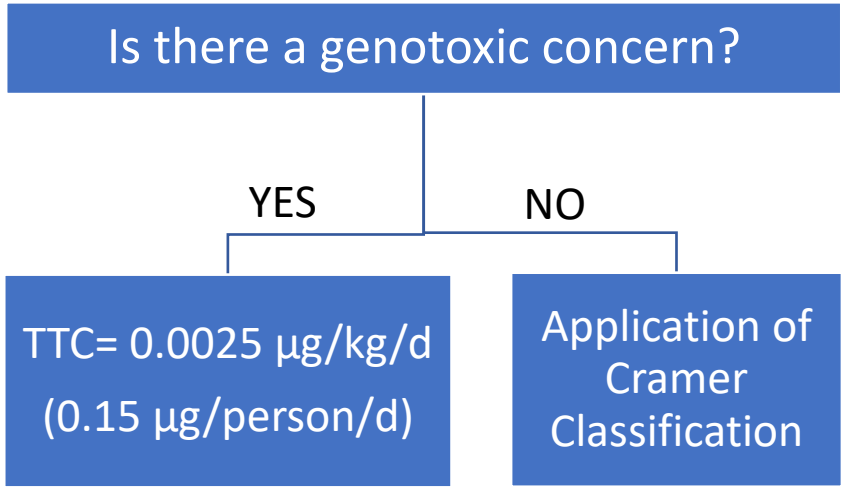
*Munro et al 2008; Kroes et al 2004; Batke et al 2021

TTC: procedure

1) The first step is to assess the genotoxicity by evaluating all available data, and/or applying (Q)SAR approaches in a WoE (Weight of Evidence Approach: integration of different source of data, experimental and estimated data, with an expert supervision)

2) For the potential DNA-reactive mutagens the TTC value of 0.0025 µg/kg/d (0.15 µg/person/day) is considered

3) If No genotoxic concern is assigned, to the substance, the Cramer classification scheme is applied



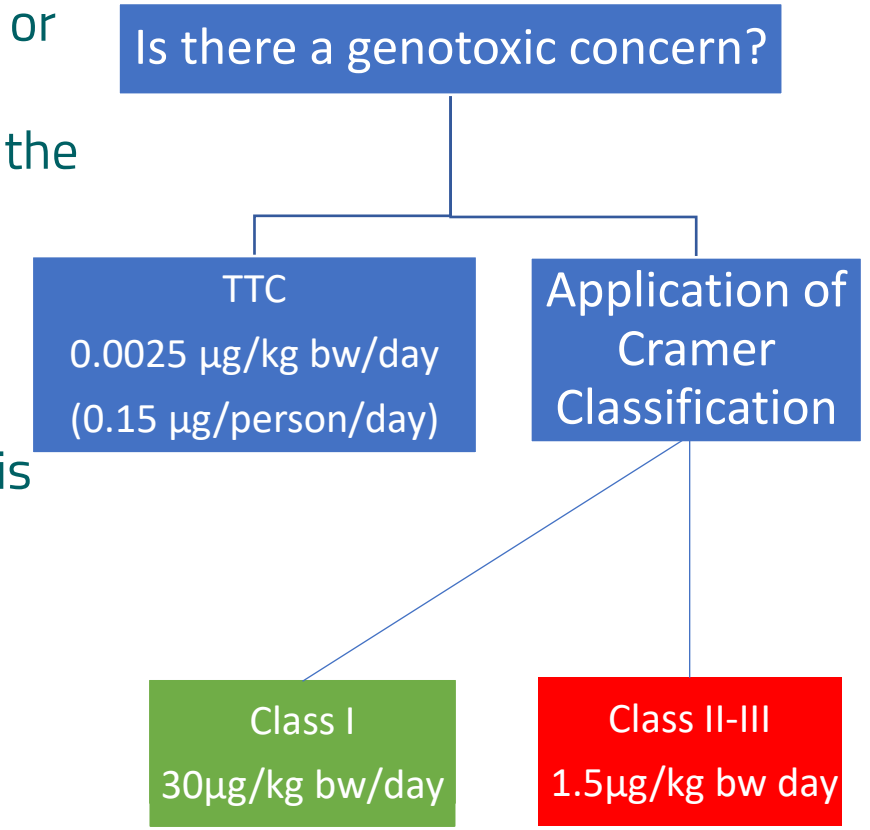
Munro et al 1996; Kroes et al 2004; Batke et al 2021

Cramer classification scheme

- Decision tree is used to categorize non genotoxic chemicals
- The original Cramer decision tree consists of "YES" or "NO" questions or rules (including extension and modification respect the original scheme)
- The answer to each question leads to a final Cramer classification for the chemical in one of three classes:
  - ✓ Class I – low toxicity
  - ✓ Class II – intermediate toxicity
  - ✓ Class III – high toxicity

} together

- Once the Cramer class is determined, a corresponding TTC threshold is chosen: if the chemical is below the TTC threshold, the toxic effect on human health is not expected



- **QSAR Toolbox** can be used to assign the Cramer class (original and extended version)
- It is a crucial step because the interpretation of each rule may vary
  - ✓ expert judgment is often needed

*Cramer et al., 1978; Munro et al 1996

In which regulatory framework (Q)SAR can be used?

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Pharmaceuticals:
Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals

efsa
European Food Safety Authority

Pesticides: Panel on Plant Protection Products and Their Residues

APPROVED: 12 March 2019
doi:10.2903/sp.efsa.2019.EN-1598

Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across

Romualdo Benigni*, Chiara Laura Battistelli**, Cecilia Bossa**, Alessandro Giuliani**, Elena Fioravanzo***, Arianna Bassan***, Mojca Fuart Gatnik***, James Rathman****, Chihae Yang****and Olga Tcheremenskaia**

Cosmetics:
Regulation (EC) No 1223/2009 SCCS safety evaluation of cosmetic ingredients (SCCS1647/22)

European Commission

Scientific Committee on Consumer Safety

SCCS

THE SCCS NOTES OF GUIDANCE FOR THE TESTING OF COSMETIC INGREDIENTS AND THEIR SAFETY EVALUATION

12TH REVISION

Scientific Committees on Consumer Safety

The SCCS adopted this guidance document by written procedure on 15 May 2023

REACH (EU Regulation on the **R**egistration, **E**valuation and **A**uthorisation (restriction) of **C**hemicals). In silico models can be used for:

- Risk assessment
- Classification
- Prioritization

ECHA
EUROPEAN CHEMICALS AGENCY

REACH: EU Regulation on the **R**egistration, **E**valuation and **A**uthorisation (restriction) of **C**hemicals

- Industries are responsible for the **safe use of the chemicals** they produce or import
- **Hazard identification** of substances is the starting point: all the information are needed to be reported in registration dossiers (IUCLID format).
- The higher the tonnage, the more information is required.

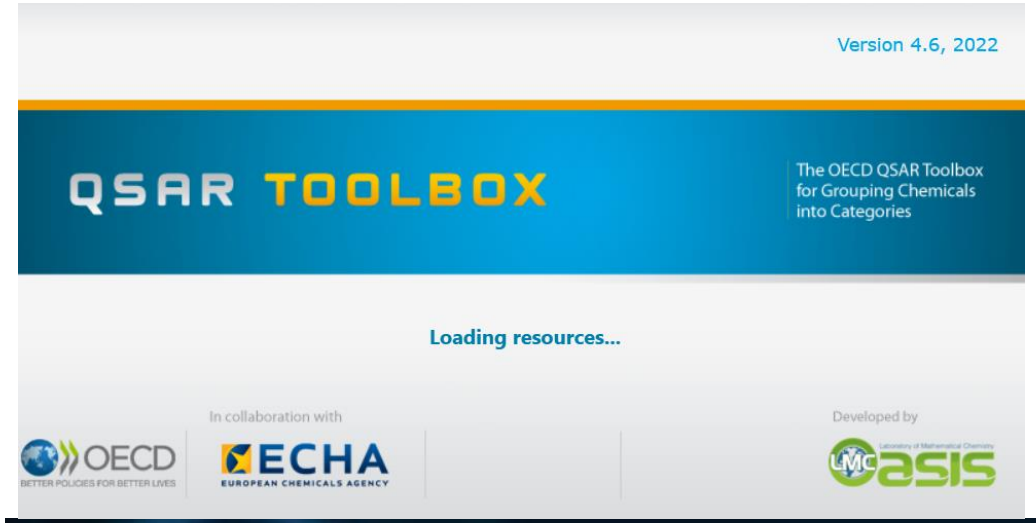
- **Reach** requires **vertebrate** testing as a **last resort** and provides and promotes:
  - ✓ **Data sharing**
  - ✓ **Use of alternative methods, such as (Q)SAR, Read-Across e chemical categories**

  Annex XI reports criteria for (Q)SAR regulatory acceptance (reliable results, fit for purpose, adeguate justified)

To increase the use of in silico methods, improving their regulatory acceptance, ECHA and OECD have funded the QSAR Toolbox software

QSAR Toolbox

- The QSAR Toolbox is a free software application that supports (eco)toxicologists in performing **reproducible** and **transparent** chemical hazard assessment using non-animal methods.
- ECHA e OECD are co-owners (Ver 4.6), developed (IT) and managed by «Laboratory of Mathematical Chemistry» (LMC)
- Website: www.qsartoolbox.org
- Resources: tutorials, manuals, webinar, helpdesk (on line support), forum (public discussion), ontologies https://qsartoolbox.org/support/



QSAR Toolbox: management and contributors

- ✓ ECHA/OECD: co-owner and co-developer
- ✓ LMC: developer, under the OECD and ECHA umbrella
- ✓ LMC/ECHA/OECD: "**QSAR Toolbox Coordination Group"**
- ✓ "**QSAR Toolbox Management Group**": coordination group + experts from industry, authorities , NGOs
  - ✓ Discuss, approve and test software developments
- ✓ Third parties contribute with data, profiler and experience
- ✓ Excellent collaboration between such different entities, such as academic, regulatory and industrial

Supporters or donors (data, profilers and experience)

- OECD
- ECHA
- LMC
- EC
- EURL ECVAM
- US EPA
- Environment Canada
- Health Canada
- NITE Japan
- NIES Japan
- Danish EPA
- UBA Germany
- NICNAS Australia
- DEWNA Australia
- ISS Italy
- Fraunhofer Germany
- BfR Germany
- Ministry of the Envir Japan
- MHLW Japan
- INERIS
- EFSA
- Univ of California
- Cefic
- OASIS
- L'Oreal
- DuPont
- Givaudan
- Dow chemicals
- BASF
- ExxonMobil
- 3M
- Firmenich SV
- SRC, Syracuse
- Unilever
- Multicase
- ChemAxon
- ECETOC
- Mario Negri (Milano)
- ....

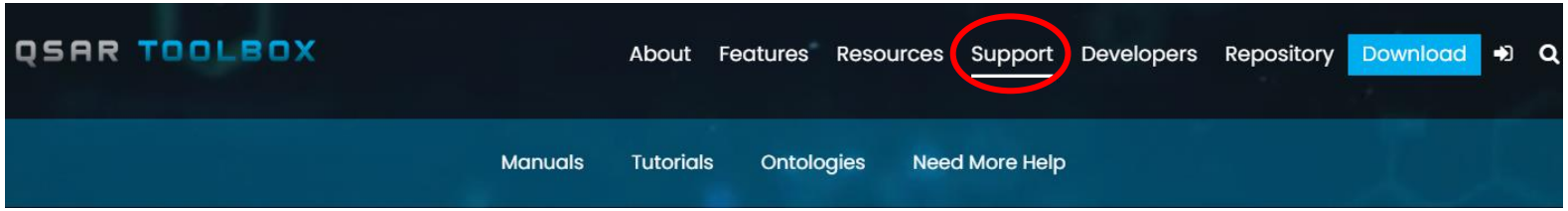
QSAR Toolbox: helps reduce animal testing

- **Prevent duplication of animal tests:** when high quality data are found, there is no need to duplicate the test.
- **Intelligent testing strategies**: by forming categories and identifying data gaps, informed testing strategies can be designed to optimize costs and number of animals required.
- **Predict toxicity using a category approach**: the Toolbox results can be used for data-gap filling and as supporting evidence for read-across cases.
- **Sustainable development and green chemistry:** the toxicity of substances can be predicted even before they are produced, facilitating sustainable product development and green chemistry

QSAR Toolbox: what is the users expertise?

- **IT Tecnhical Skills:** decreased in recent versions, because of the simplified User Interface o Automatic workflow
- **Expertise in Toxicology**: in risk assessment procedure and in the endpoints to predict (e.g. data quality assessment)
- **Expertise in organic and computational chemistry**: to justify and interprete similarity between analogues, to evaluate the QSAR prediction,…

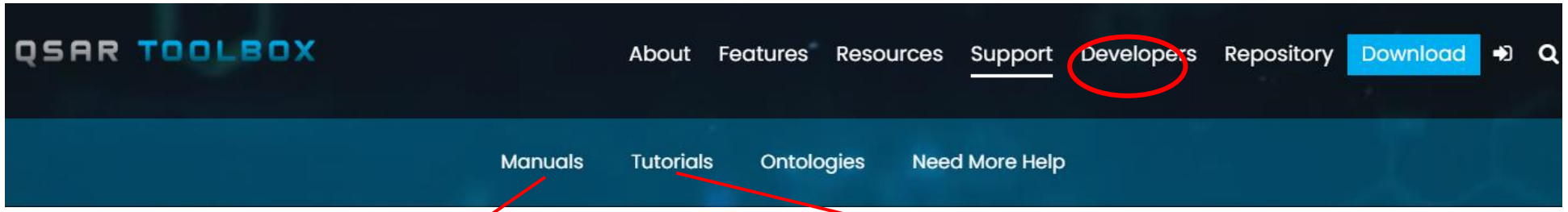
Support section

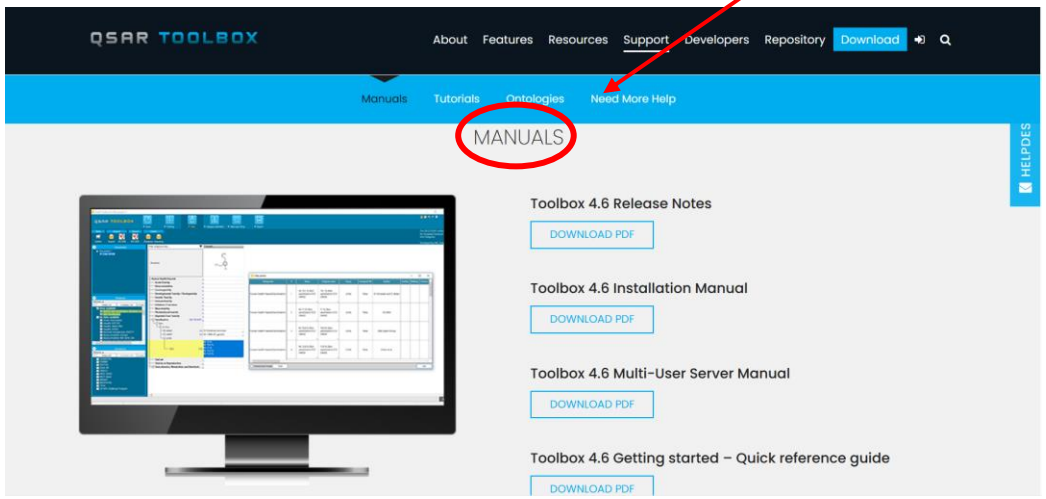


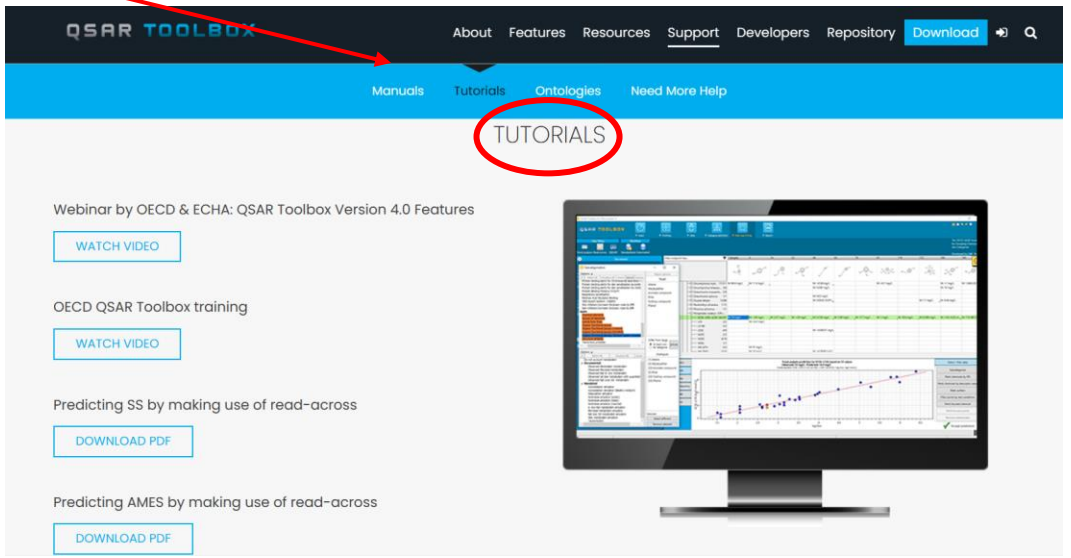
https://qsartoolbox.org/support/

Manuals (installation and user manuals), tutorials (training), ontologies (Controlled toxicological vocabularies and interrelations), more help (helpdesk)…

QSAR Toolbox: support

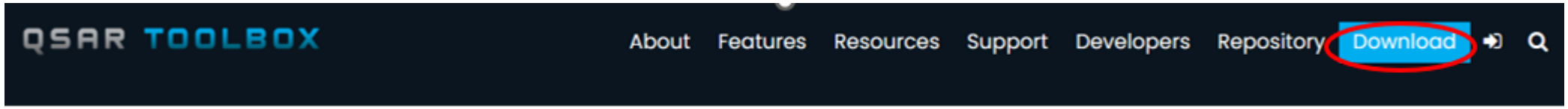


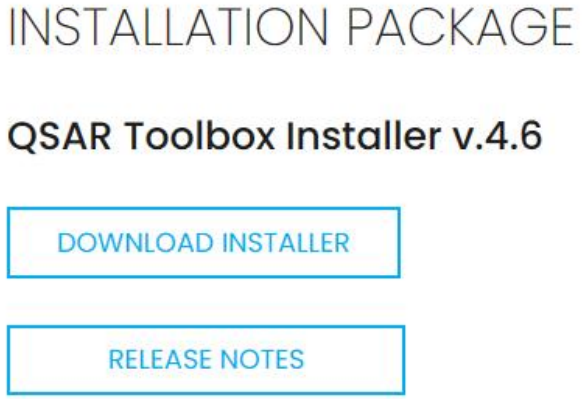




https://qsartoolbox.org/

QSAR Toolbox: download





https://qsartoolbox.org/

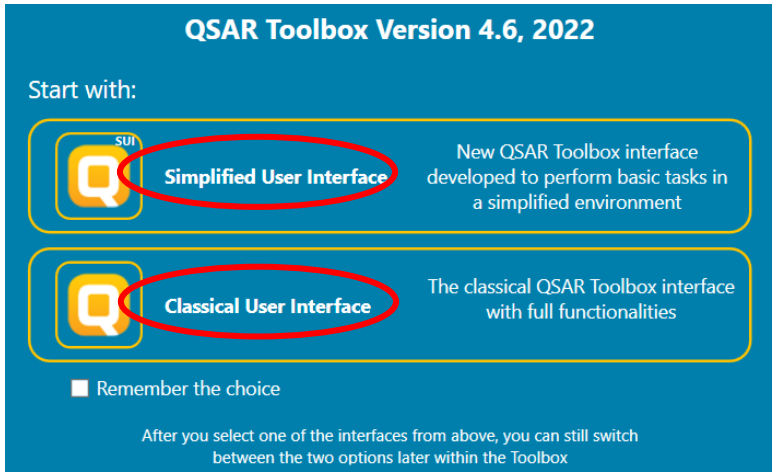
Registration to the web account is required

QSAR Toolbox: interfaces

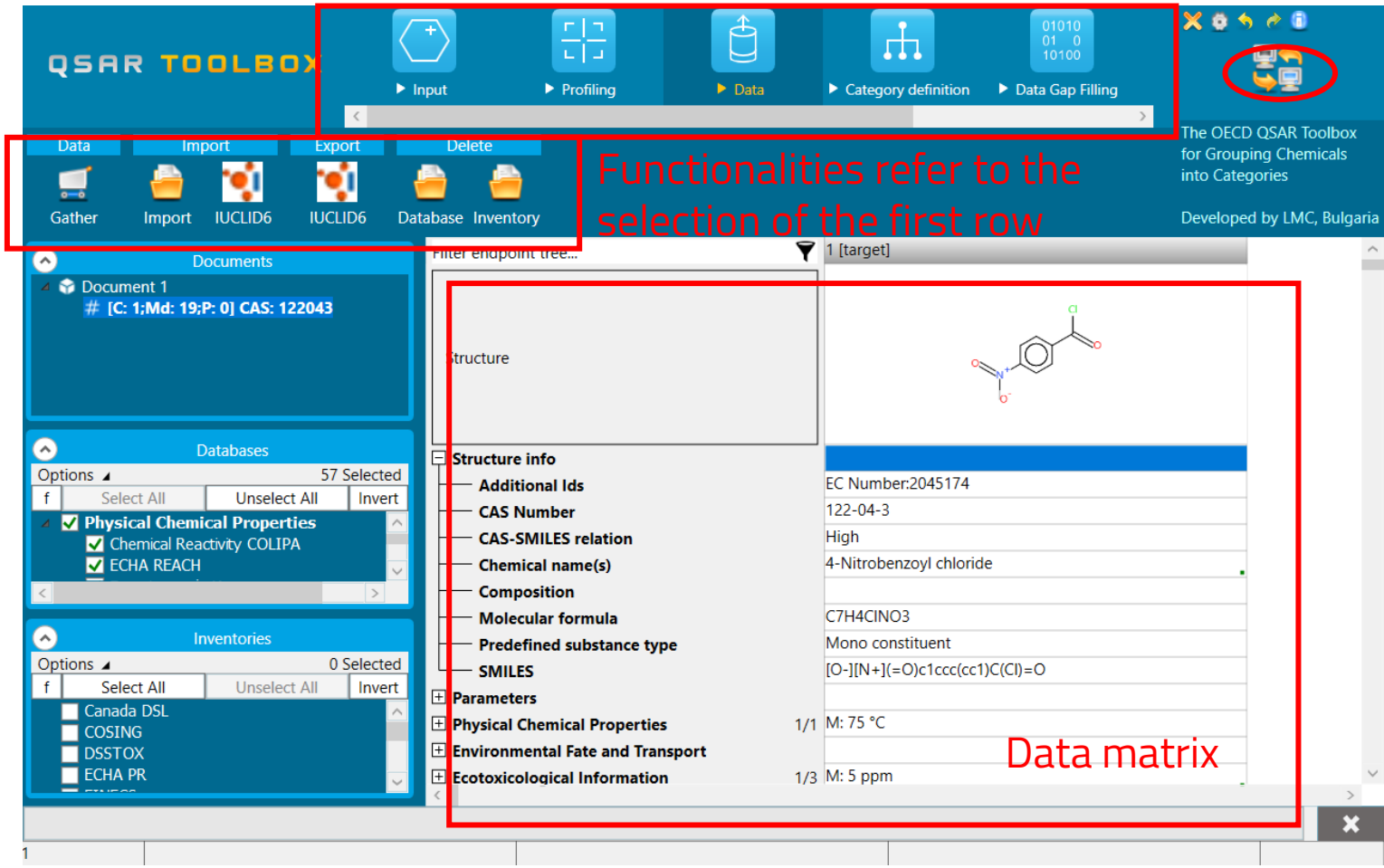
- ✓ **Simplified user interface***: Easy to use but includes only simple functionalities*
- ✓ **Classical user interface***: Main interface. Includes all functionalities but needs training.

- ✓ **Web client**: Latest interface. Runs on all operating systems

*Only runs on Windows



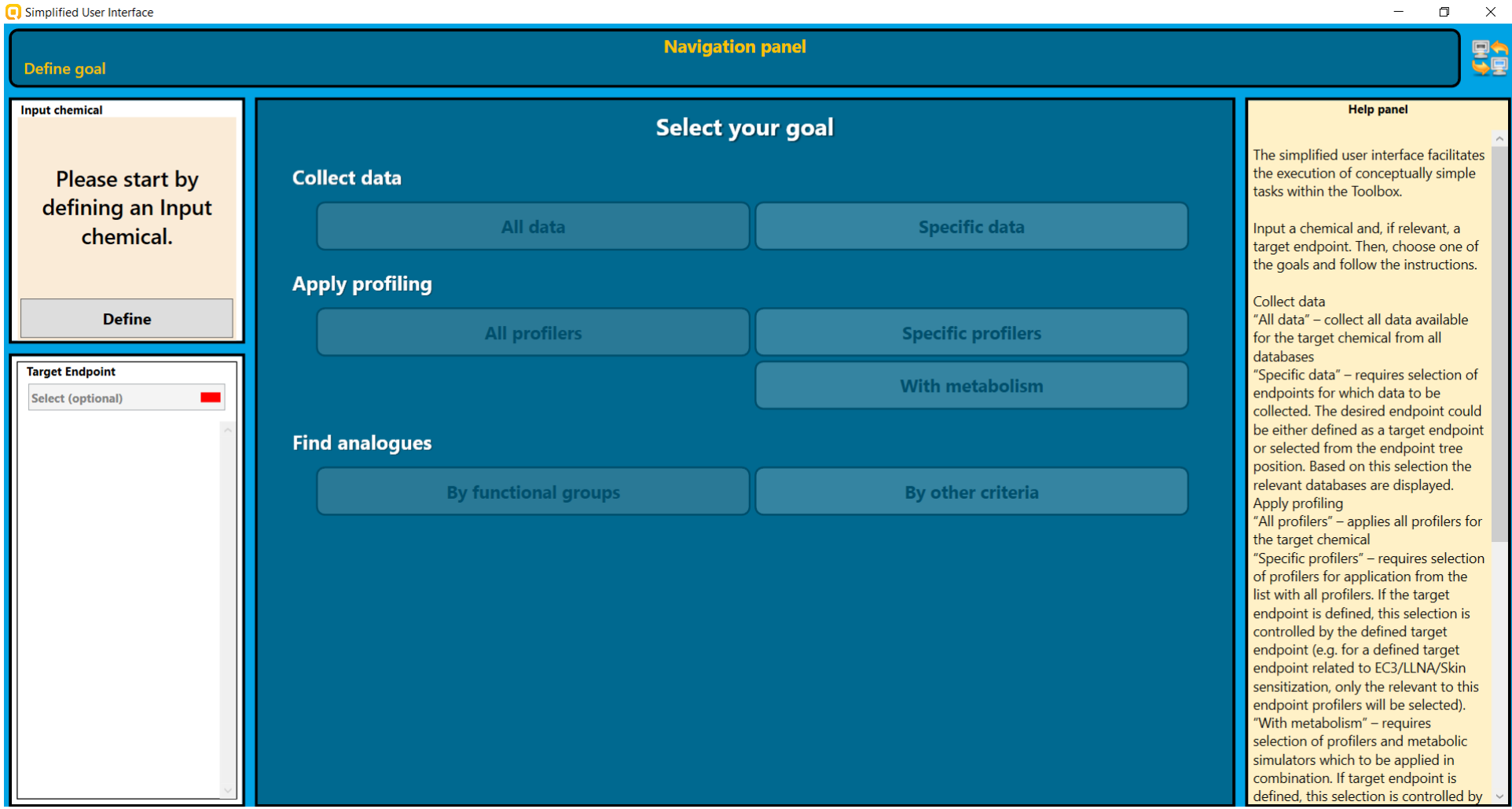
Modules on principal workflow



QSAR Toolbox: Simplified user interface (SUI)

With the simplified user interface the user can:

- Collect data
- Apply profiling
- Find analogues



QSAR Toolbox: key functionalities

The Toolbox consists of a logical and sequential workflow, with the following modules:

1. **Input**: starting point and provides different ways to specify the identity of the target substance and the property under consideration.
2. **Profiling**: contains the knowledge coded in profiling schemes (**profilers**). The profilers identify the affiliation of the target chemical(s) to categories (functional groups/alerts), and include **observed** and **simulated metabolism** and transformation
3. **Data**: include all the data of TB databases. Data can contain chemical information (CAS, name, SMILES), experimental data and supporting information (metadata)
4. **Category definition**: used to group chemicals into a toxicological **category**, according to structural or mechanistic similarity, to be used in read-across or trend analysis
5. **Data Gap Filling**: used to fill a data gap using data from analogues with **trend analysis, read-across or** existing **QSAR models.**
6. **Report**: produce a report for prediction, export the chemicals on the data matrix and related information

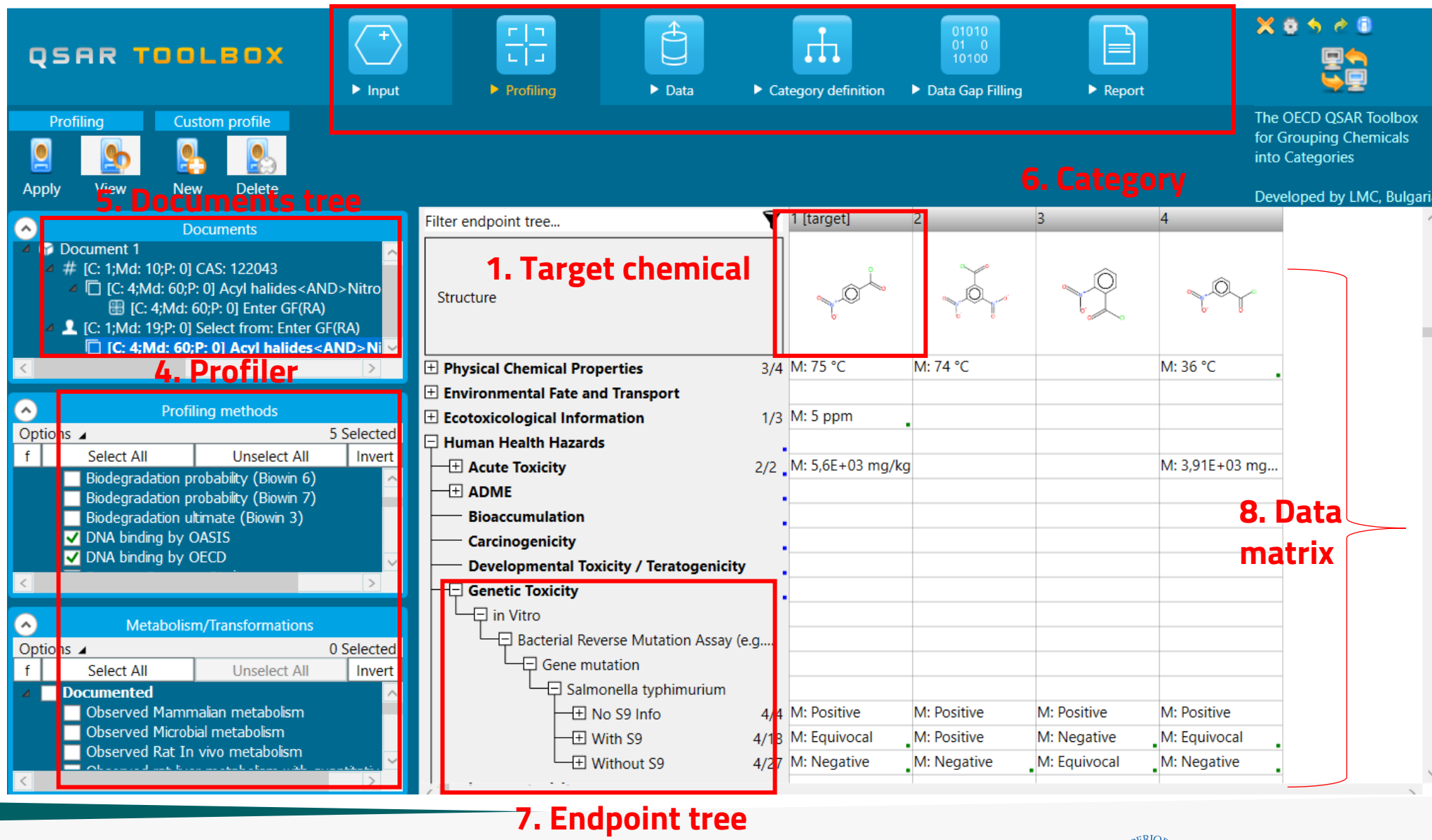
Modules do not necessarily have to be used sequentially

QSAR Toolbox: keywords

- ✓ **Target chemical**: chemical of interest
- ✓ **Module:** section dedicated to a specific action and option (6 modules: input, profiling,…)
- ✓ **Workflow:** the use, in combination, of the different modules (prediction workflow, from input to report)
- ✓ **Profiler**: algorithm (rules set) for the identification of specific features of the chemicals: structural (i.e*. organic functional groups*), general mechanistic (i.e. *protein binding by OECD*), endpoint specific (i.e.*in vitro mutagenicity alerts by ISS*)
- ✓ **Category**: group of substances sharing the same characteristics (e.g. same functional group or mode of action). In a typical Toolbox workflow it consists of the target chemicals and its analogues gathered according to the selected profilers
- ✓ **Endpoint tree**: branched tree scheme, from a broader level (Phys-Chem prop, Environmental fate and transport, Ecotoxicology, Human Health) to a more detailed one (e.g. *in vitro* or *in vivo assay, species and other metadata*)
- ✓ **Data matrix**: table reporting the chemical(s) and data (experimental results, profilers outcomes, preditions). Each chemicals is in a different column, each data in a different row

QSAR Toolbox

1. Target chemical
2. Module
3. Workflow
4. Profiler
5. Document tree
6. Category
7. Endpoint tree
8. Data matrix



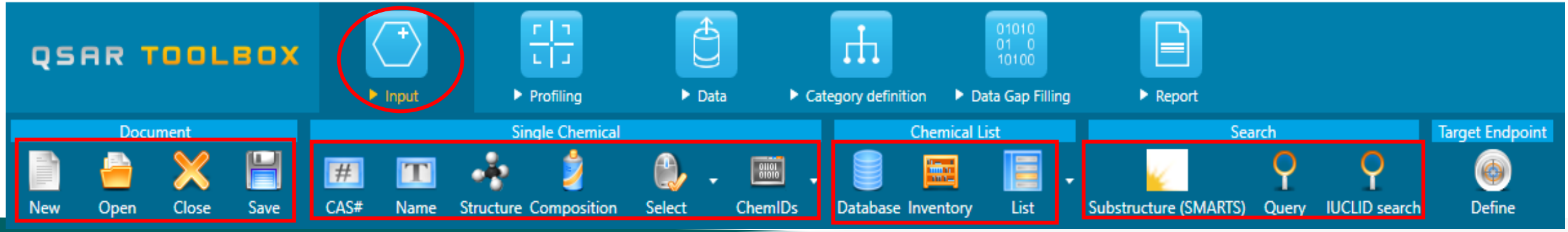
ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

OECD QSAR Toolbox: input

The users are able to:

- ✓ **Open a new or already saved document**
- ✓ **Close or save the current document**
- ✓ **Load a single target chemical** – by CAS, Name, Structure, SMILES* or drawing (including mixtures), select from a file
- ✓ **Load a list of chemicals** – database, inventory, custom file
- ✓ **Customized search** – searching chemicals and/or data within the Toolbox databases. One or more than one criteria combined with logical operators (AND, OR, NOT) could be used.
- ✓ **Search in IUCLID databases** – searching chemicals within the IUCLID databases imported in Toolbox. One or more than one criteria for the composition (e.g. impurity, additive) of the searchable substances could be defined.

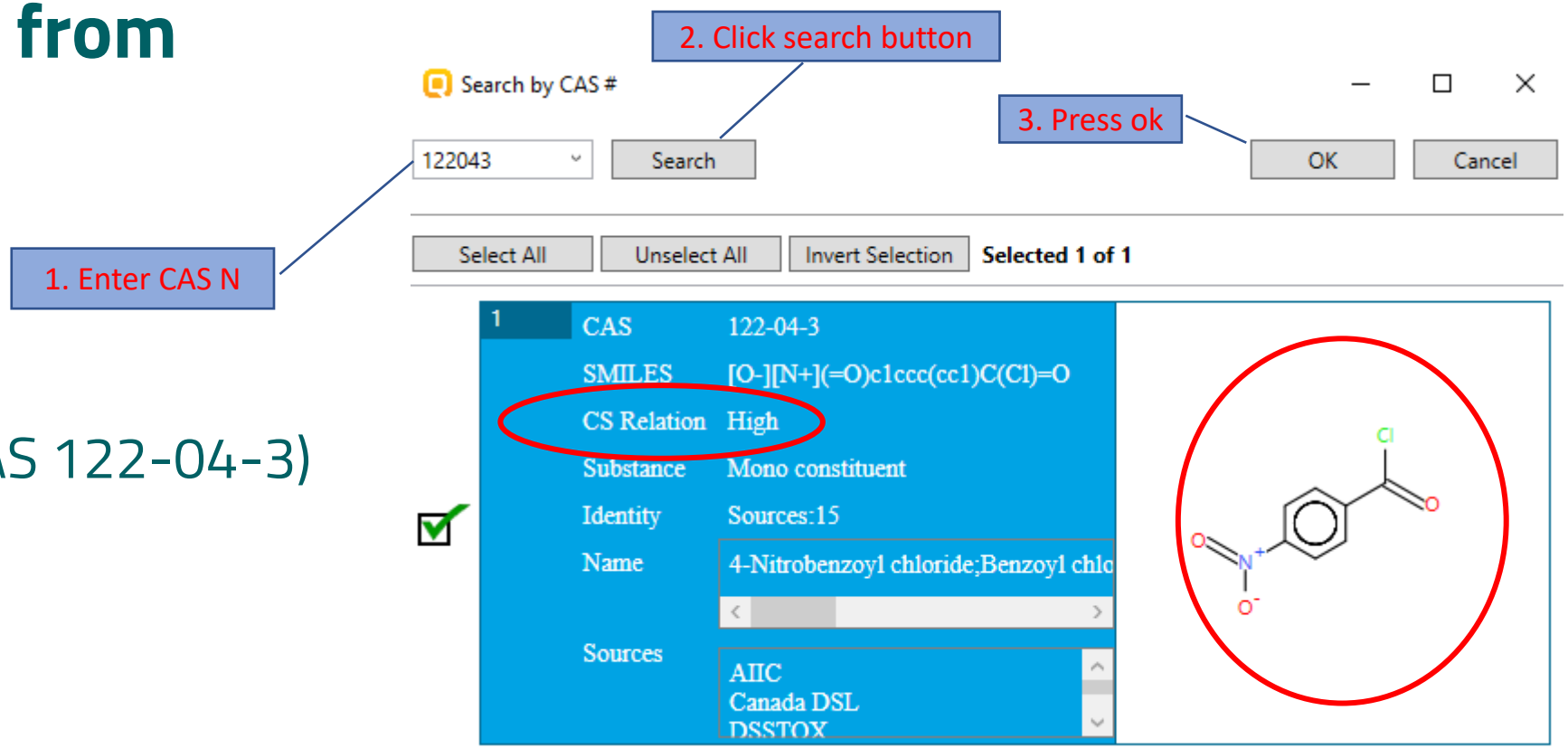
*SMILES notation: string representation of a molecule (Simplified Molecular Input Line Entry Specification)

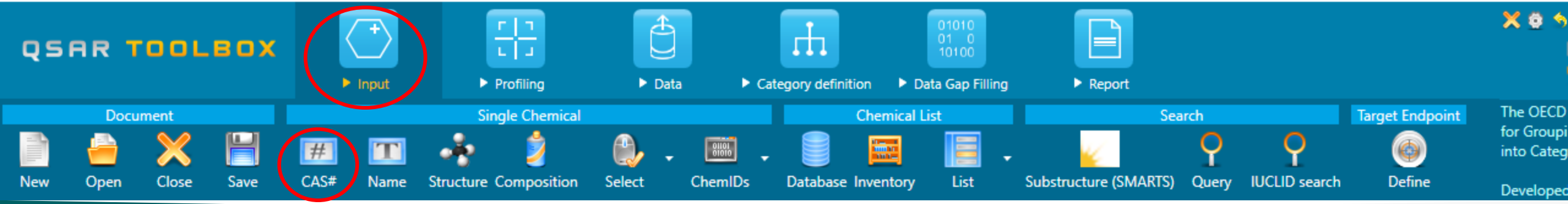


ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

QSAR Toolbox: input from the CAS

Target chemical:
4-nitrobenzoyl chloride (CAS 122-04-3)



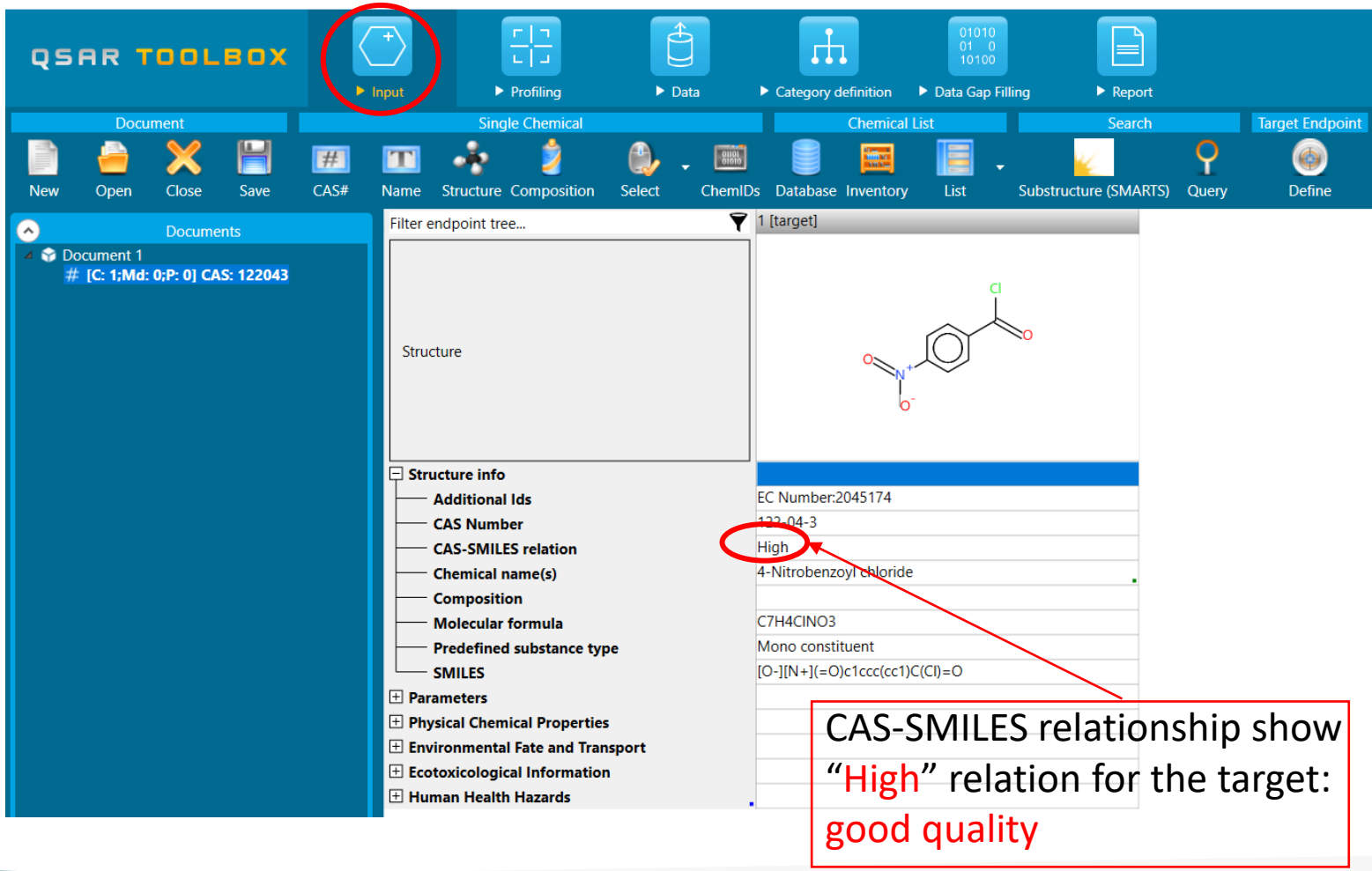


ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT
ENVIRONMENT AND HEALTH

QSAR Toolbox: input (reliability of the target identity)

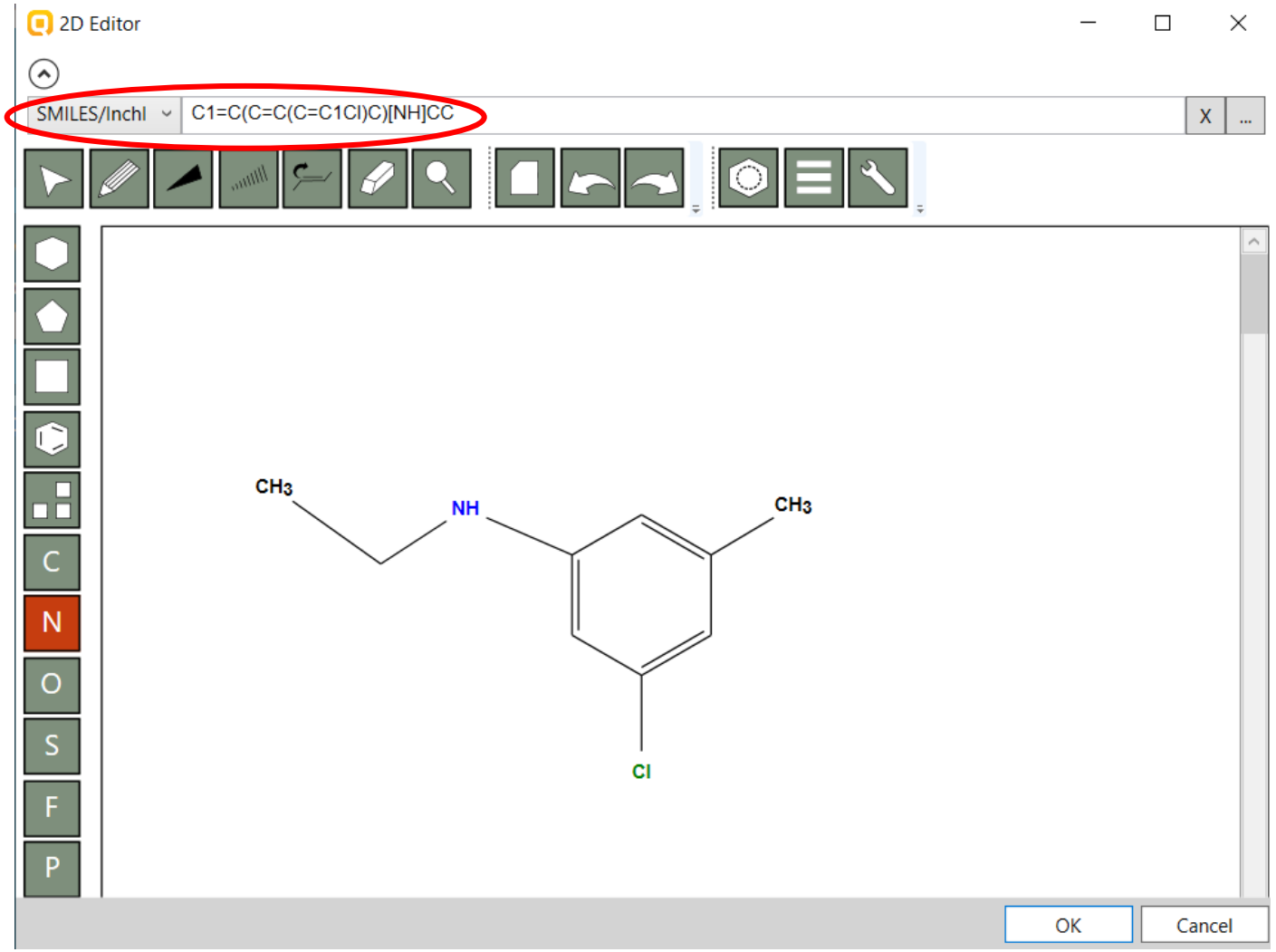
CAS-SMILES relation indicates the reliability of the target identifier

- High: This label is assigned if the chemical belongs to *at least one high quality data source* (database or inventory)
- Moderate: The moderate label is assigned if the chemical belongs to *three or more sources with unknown quality* (marked with "Distribute to QA").
- Low: This label is assigned if the chemical belongs to *less than three, but at least one source with unknown* quality ("Distribute to QA").



QSAR Toolbox: input

- The user can search the chemicals based on:
  - CAS
  - name
  - drawing the 2D Structure (2D editor)
  - pasting or drawing the SMILES or InChI
  - A given property (selecting DB)
  - Subfragment (using SMART)
  - Profilig results

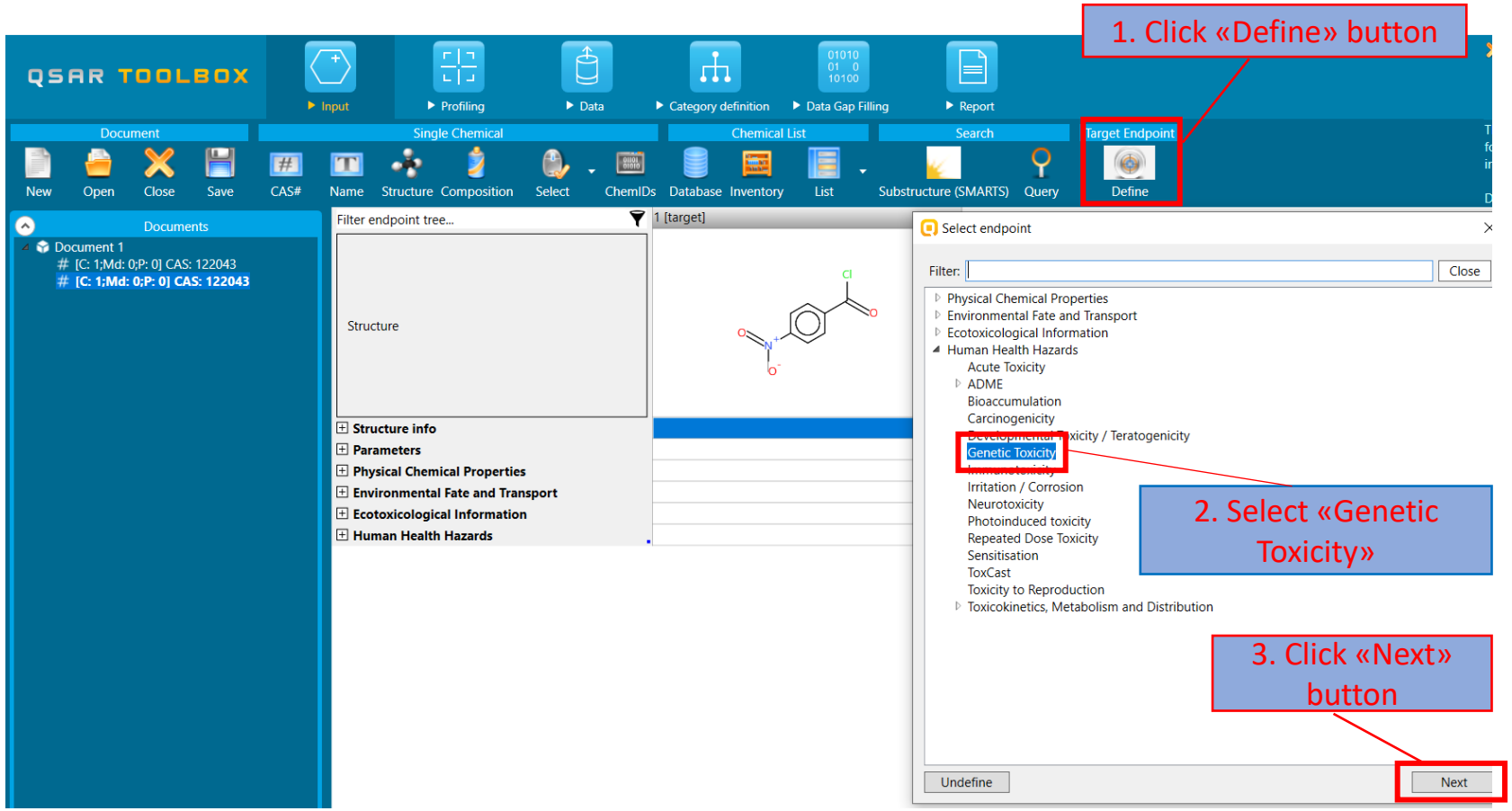


SMILES: Simplified Molecular Input Line Entry System
InChI: International Chemical Identifier
SMART: SMiles ARbitrary Target Specification, extension of SMILES code

QSAR Toolbox: Input (Define target endpoint 1)

Endpoint of interest can be specified, during the input

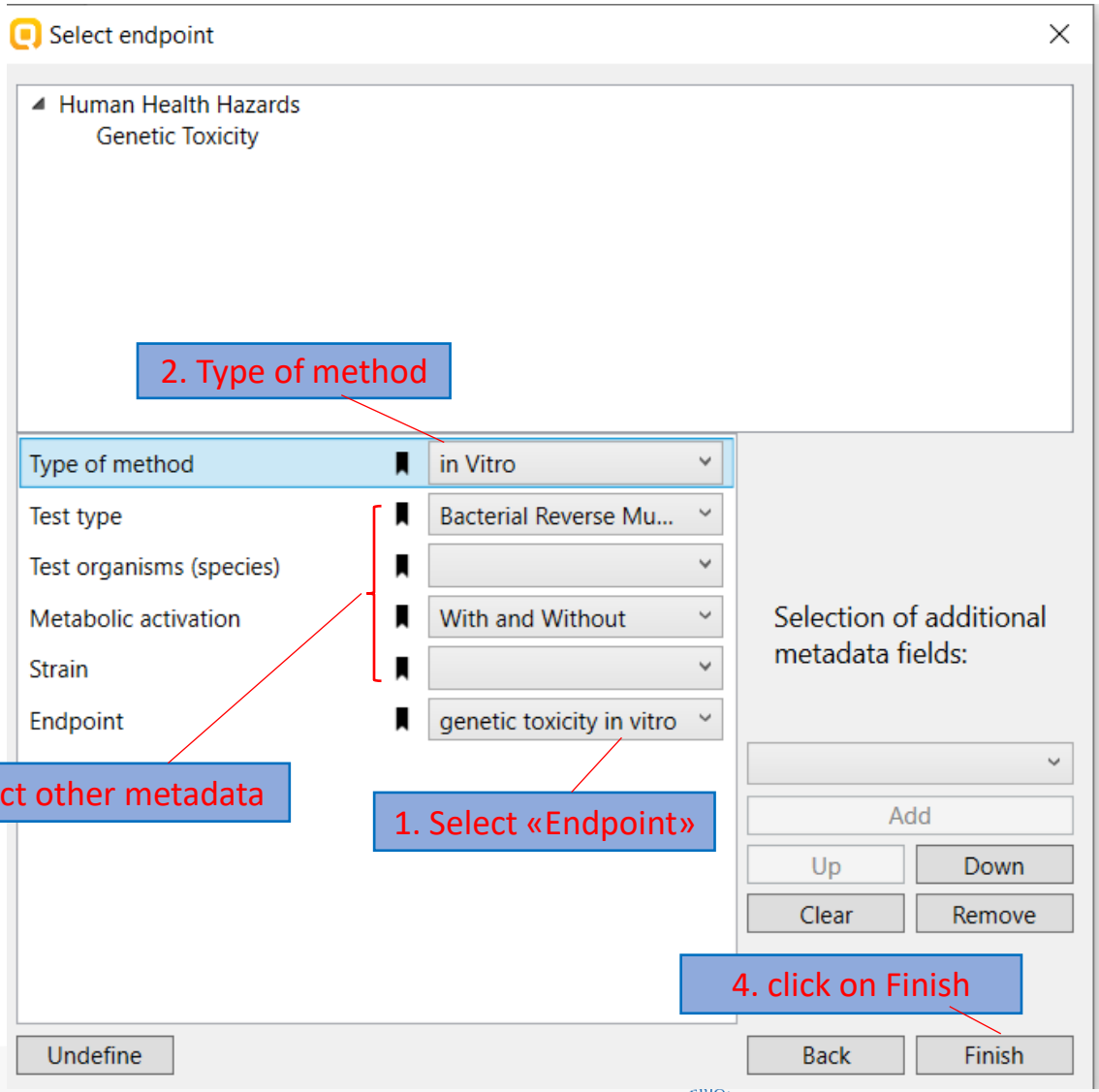
The most relevant profilers and databases will be highlighted with different colors.



OECD QSAR Toolbox: Input (Define target endpoint 2)

1. First click on Endpoint and select the endpoint from the drop-down menu (e.g. "Genetic toxicity *in vitro*")
2. Next select Type of method (e.g. "*In Vitro*")
3. Then consecutively select other metadata: Select "Strain", "Metabolic activation", "Test organ (species)"
4. Finally click on Finish

Decision tree is expanded and the row is yellow



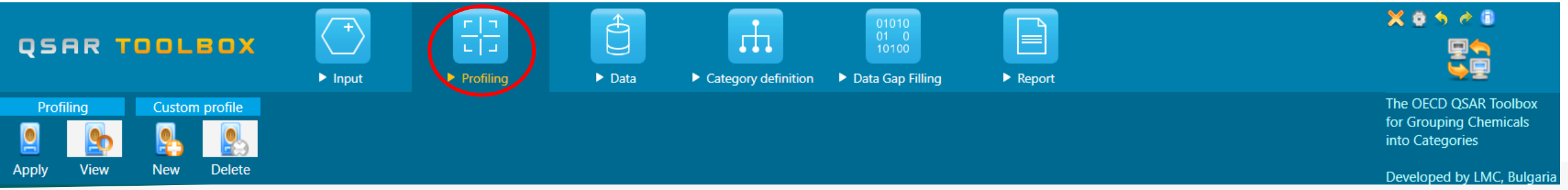
Chiara L Battistelli

ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

QSAR Toolbox: Profiling

- **Profiling**: contains the **structural** or **mechanistic** knowledge of the target, coded in profiling schemes (**profilers**), but not (experimental, (eco)toxicity) data!)
- The **profilers** identify the affiliation of the target chemical(s) to categories (**functional groups/alerts**). Mechanistic justification for the identified alerts is provided.
- The outcome of the profiling determines the most appropriate way to search for analogues, also useful for screening or prioritization of substances
- The "Profiling" module contains also **observed and simulated metabolisms/transformations**, which could be used in combination with the profiling schemes
- The most relevant profilers will be highlighted with different colors, based on the endpoint definition (green or orange)

The profiler results are not a predictions and should not be used as such!



QSAR Toolbox: Defining the endpoint

- **Coloring of the databases**

  Database

  Database contains experimental data for the target (selected) endpoint

- **Coloring of the profiling schemes/metabolic simulators**

  Suitable profiler/metabolic simulator

  Profiler/metabolic simulator is suitable to be used for the target (selected) endpoint

  Plausible profiler/metabolic simulator

  The profiler/metabolic simulator could be used for the target (selected) endpoint

- **Colors in the Data matrix**

  Target endpoint

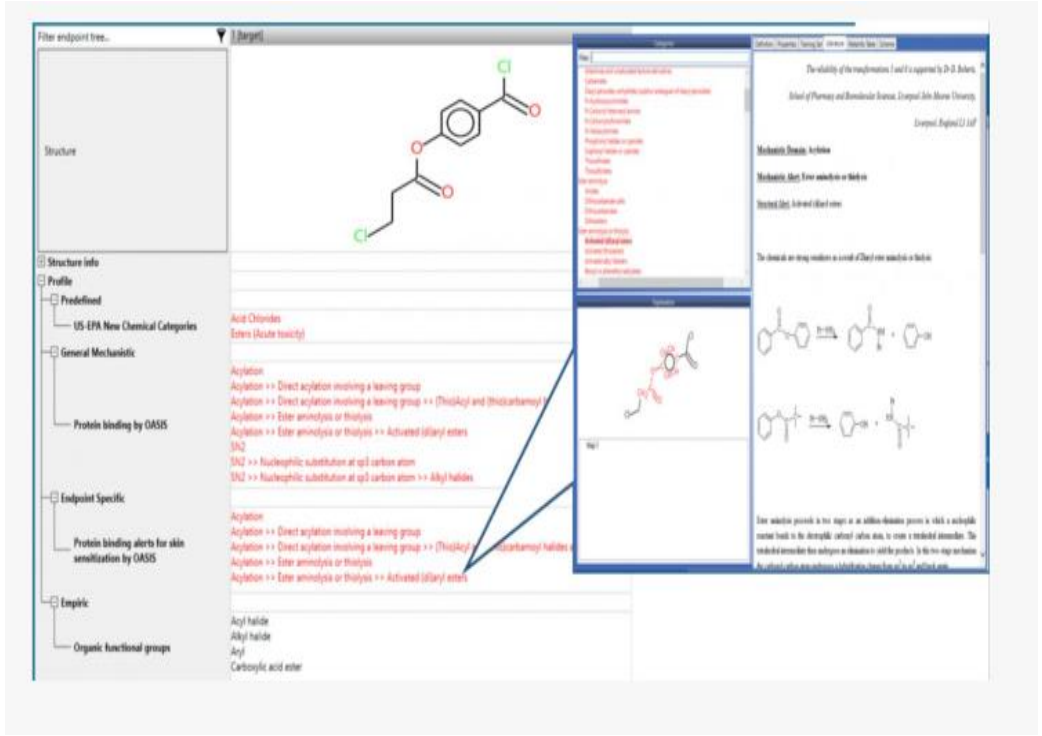
  The row corresponding to the target endpoint is highlighted in yellow on the data matrix

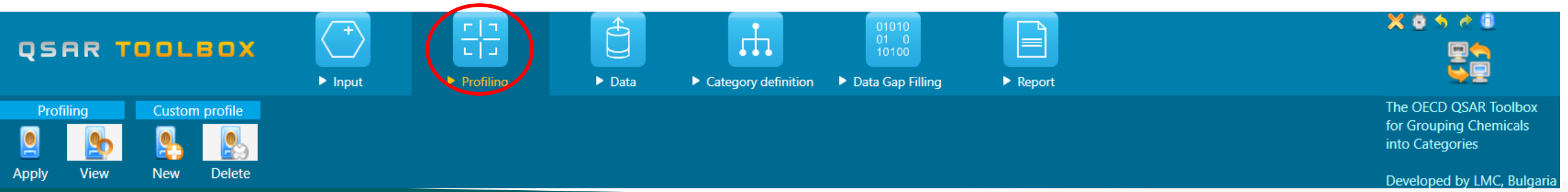
  Supporting endpoint

  The row corresponding to supporting endpoint is highlighted in pale yellow on the data matrix

OECD QSAR Toolbox: Profiling

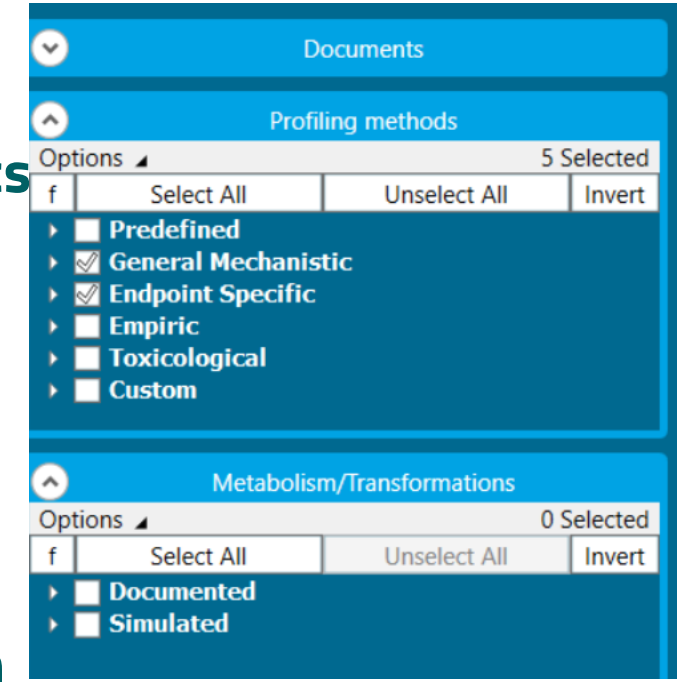
- **Structural profilers** support the identification of structurally similar substances.
- **Mechanistic profilers** provide an understanding of the mode of action, which is key to predict the activity of substances or to form categories based on the mode of action.
- The results of the profilers in the Toolbox, include descriptions and references to scientific papers to explain the outcome

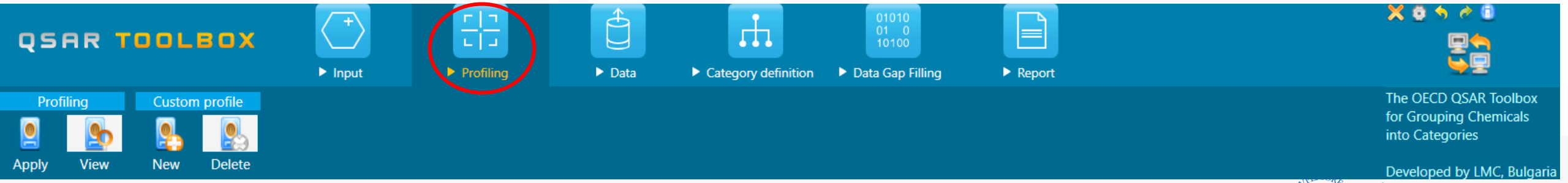




OECD QSAR Toolbox: Profiling

- **Predefined**: include standard categorization schemes e.g. OECD HPV Chemical categories
- **General mechanistic:** consist of rules of **general chemical characteristics** based on published or expert knowledge (e.g. DNA binding by OECD, Estrogen Receptor…)
- **Endpoint specific:** consist of **specific endpoint** based on published or expert knowledge e.g. *in vitro* mutagenicity (Ames test) by ISS
- **Empiric**: e.g. organic functional group US EPA, groups of elements
- **Toxicological**: includes only one profiler, the repeated dose HESS
- **Custom**: schemes based on the user's knowledge. Example Prioritization Scheme (PBT)





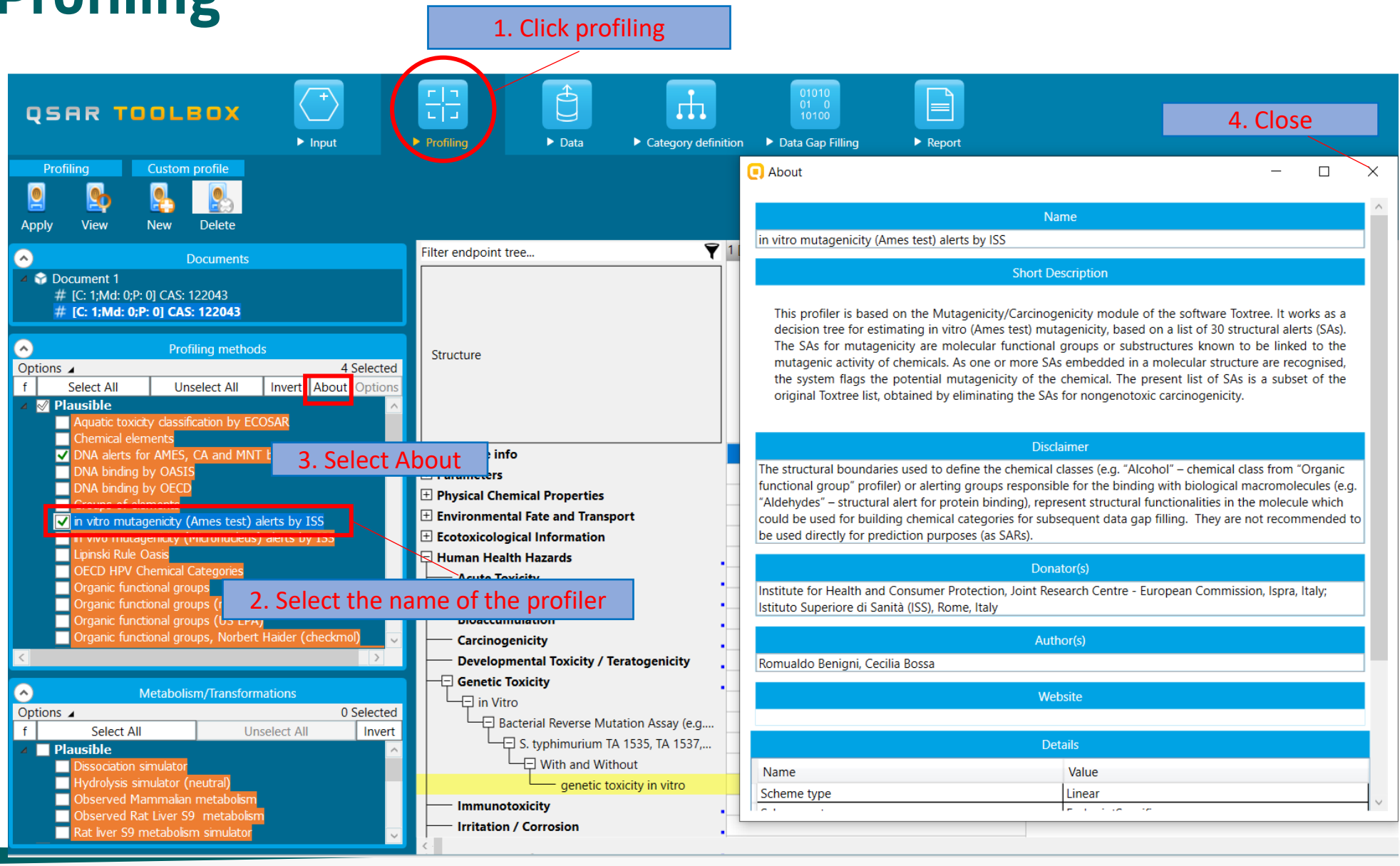
QSAR Toolbox: Profiling

1. Go to Profiling module
2. Select the name of the profiler
3. Select About
4. Close before proceeding

Relevance
Green: more relevant profilers

Orange: plausible profilers, related in some way to the endpoint

Not highlighted: profiler with no relation



QSAR Toolbox: how to select the profiler?

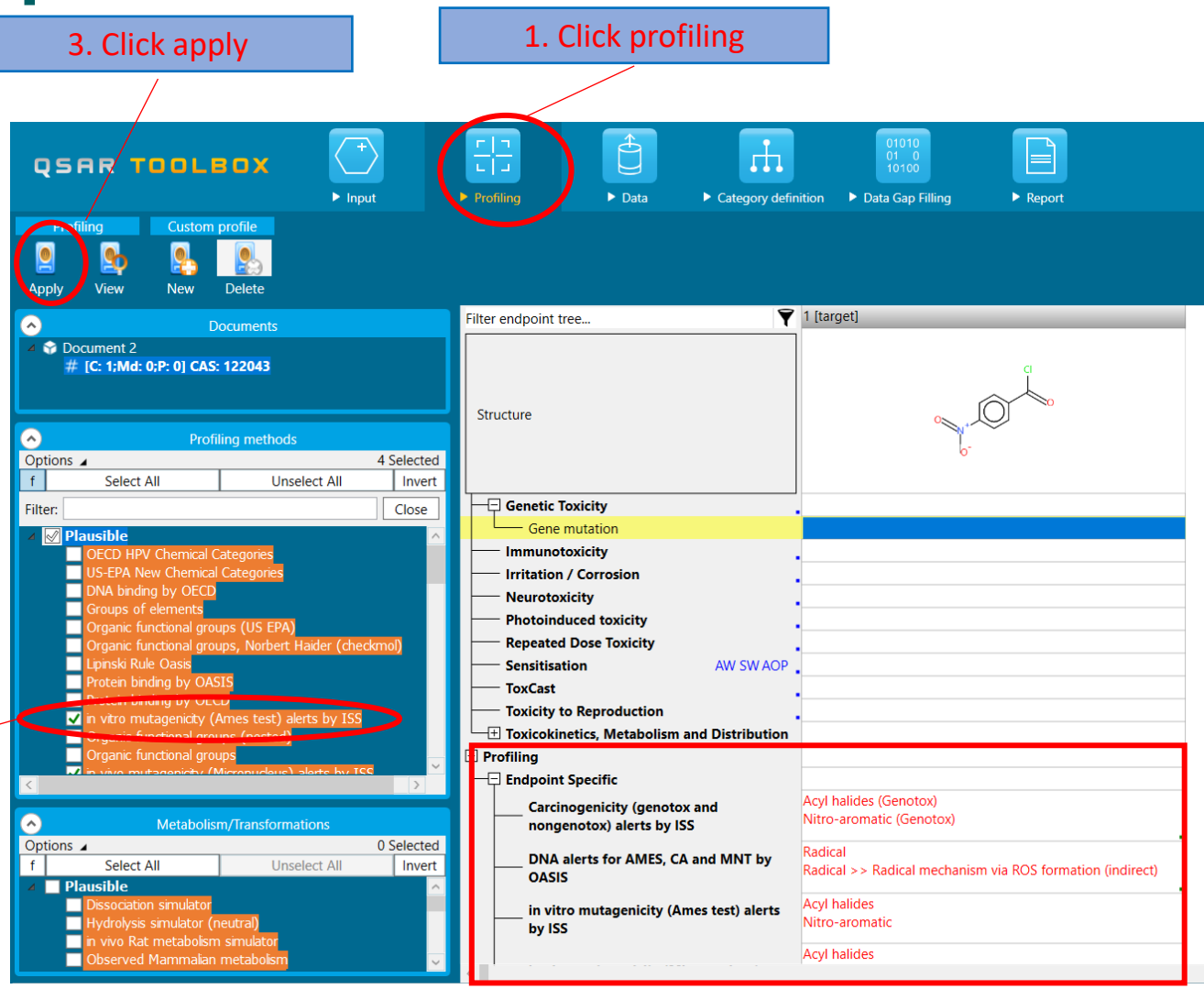
Several approaches can be taken to identify the profilers to be applied:

1) choose the most appropriate ones from mechanistic and/or structural profilers

2) choose the relevance for an endpoint (e.g., in vitro mutagenicity alerts by ISS), also with the help of orange/green colors

3) choose from the list in blue (unguided)

In the matrix the result of applying the profilers is reported



Chiara L Battistelli
ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

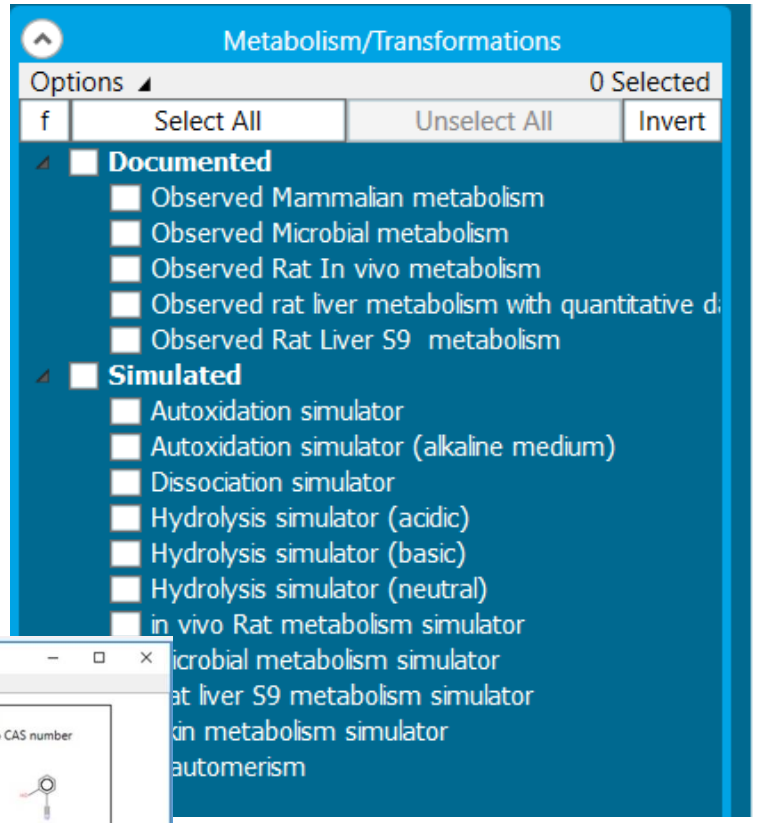
QSAR Toolbox: Profiling, metabolism/transformation

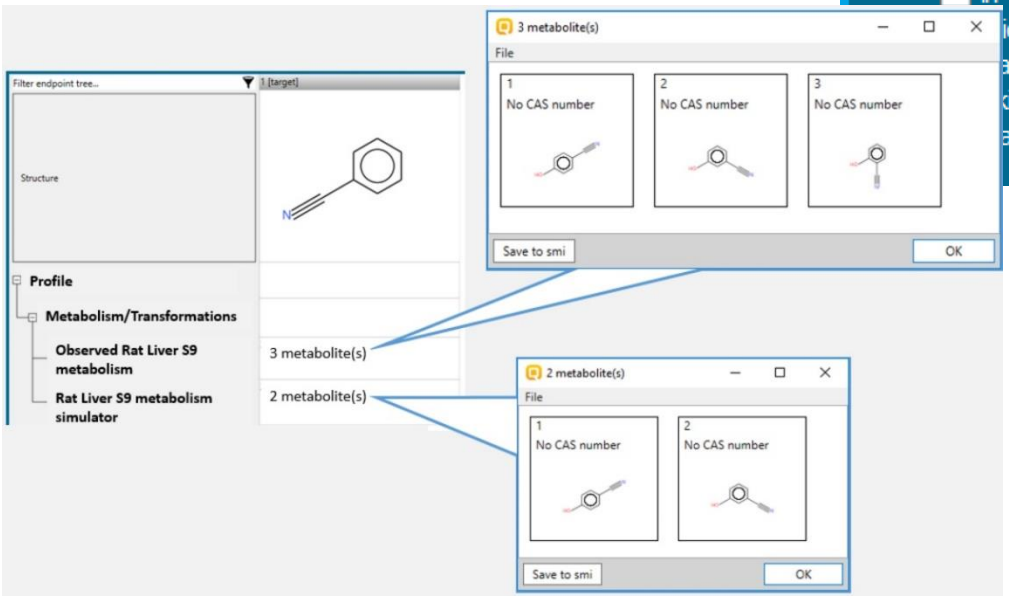
Toolbox can take metabolism into account:

- 5 Databases with observed metabolism (experimental)
- 11 Metabolism simulators (biotic and abiotic)

A screening to all the metabolites for their potential hazard

- Applying the profilers
- Searching through Toolbox databases





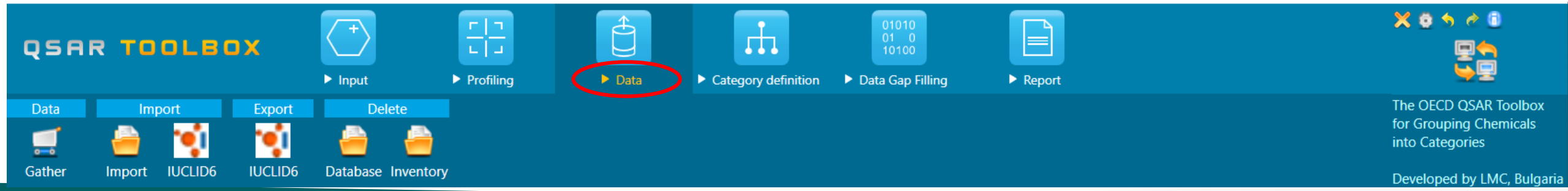
Chiara L Battistelli

ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT
ENVIRONMENT AND HEALTH

QSAR Toolbox: Data

- No need to perform new experimental studies if the information is already (publicly) available
- Toolbox includes experimental data on chemicals from 62 **databases**, including 100,000 chemicals and over 3 million data points
- Experimental data from analogues can also be used for read-across predictions
- **Inventories**: chemicals with ID information, and no experimental data
- Data quality cannot be guaranteed, as data curation is the responsibility of the donor

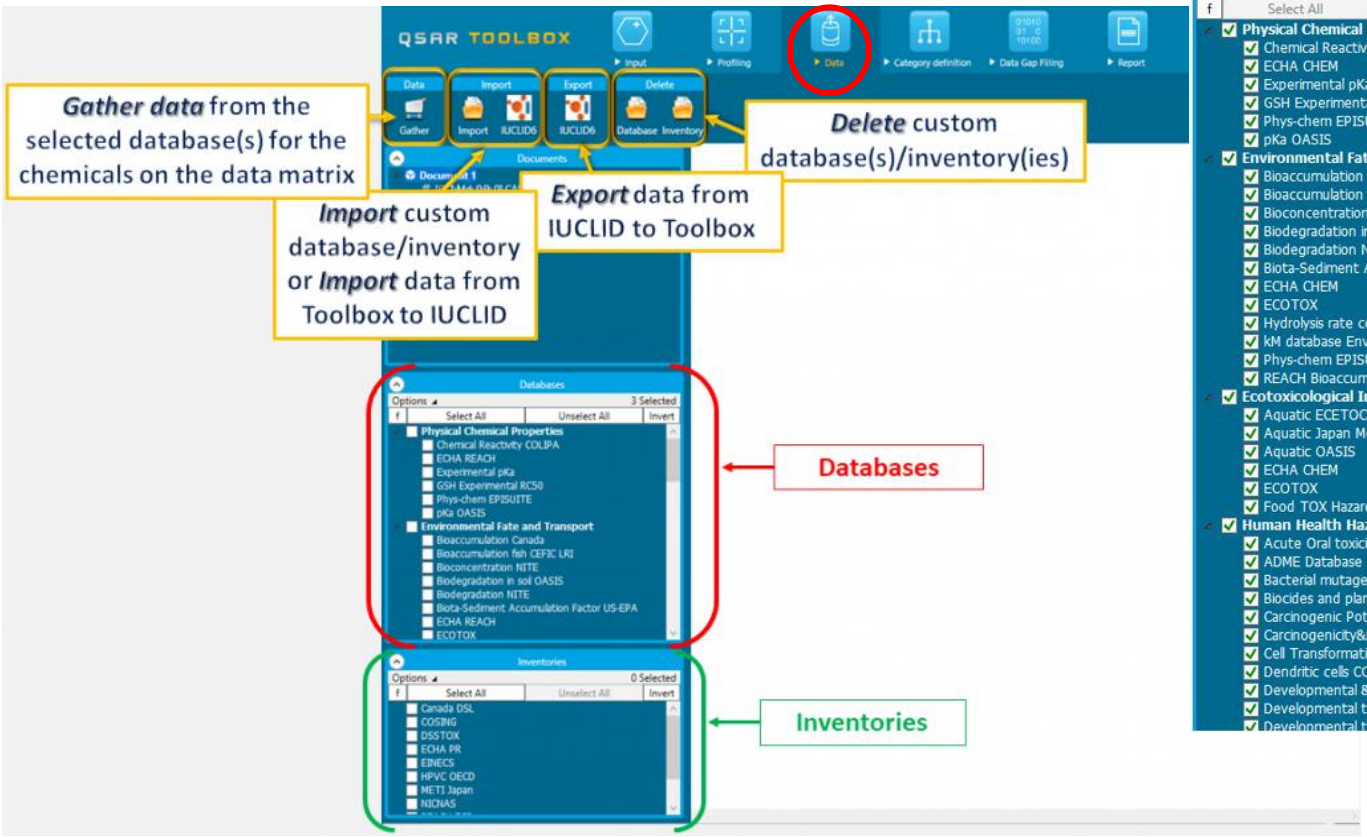
Database content	Chemicals	Data points
Physical chemical properties	50 642	239 949
Environmental fate and transport	15 356	171 861
Ecotoxicological information	23 137	1 349 467
Human health hazards	45 904	1 263 995
Total number	**135 039**	**3 025 272**

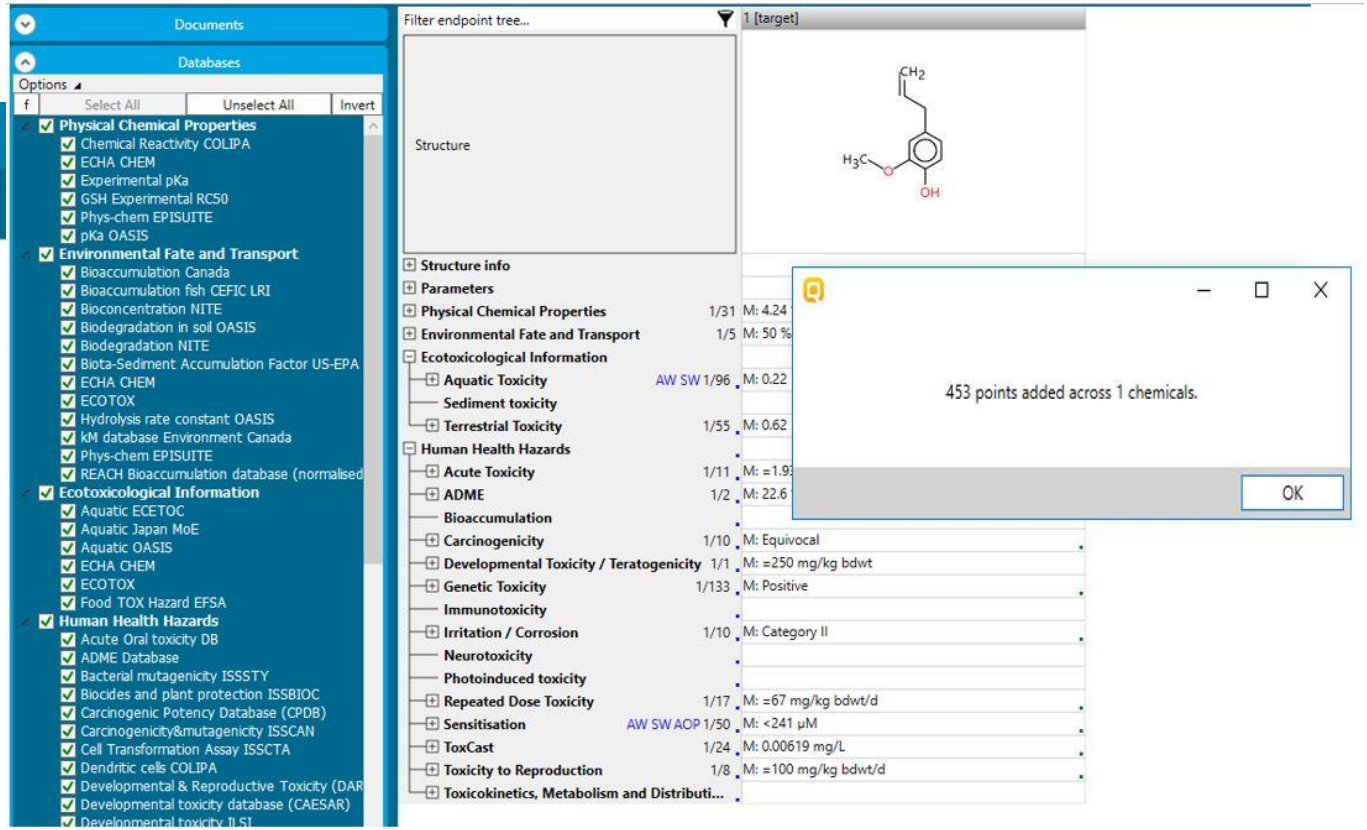


QSAR Toolbox: Data

Database can be selected for:

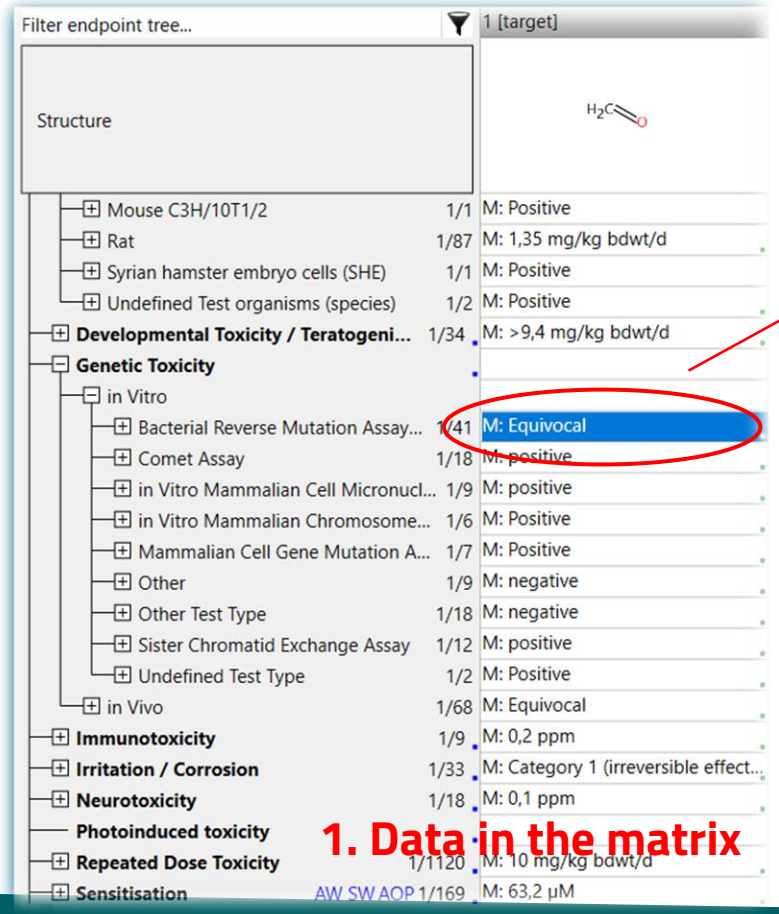
- Search analogs for data gap filling
- Search data for one or more substances

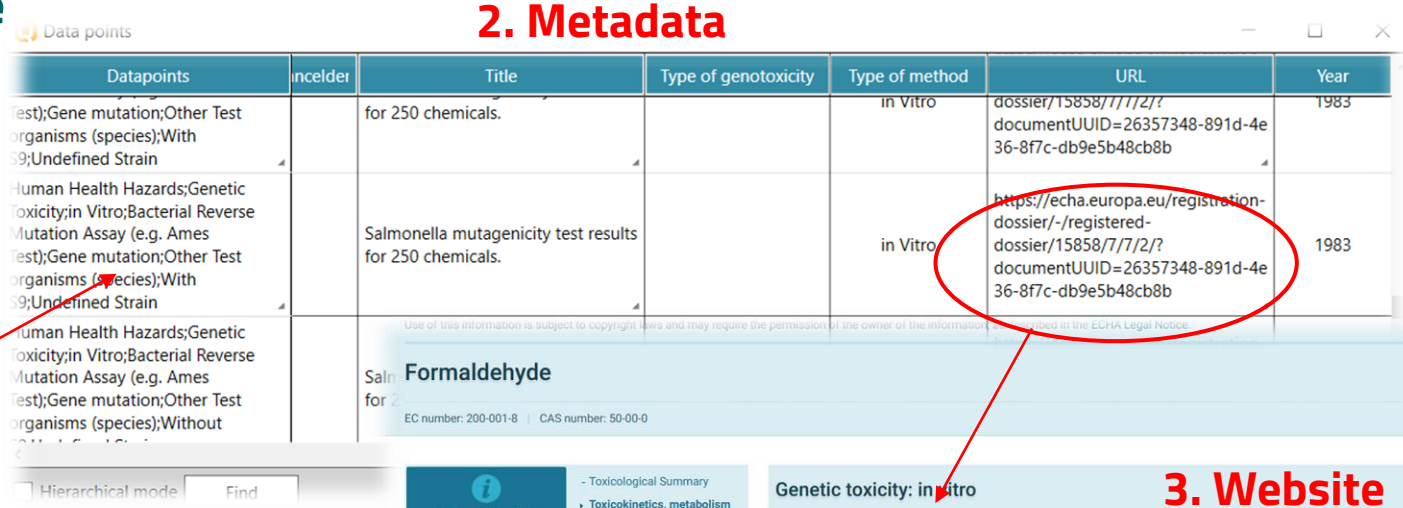


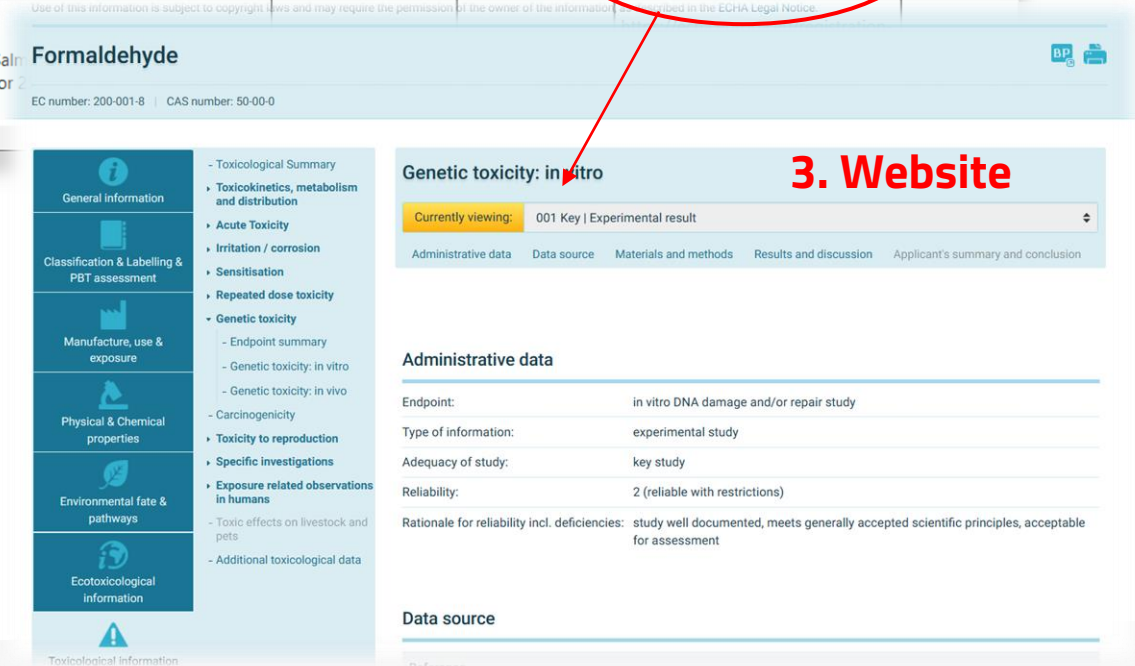


QSAR Toolbox: Data and metadata

1) **Data** are shown in the matrix
2) Click on data to check **metadata** (data that describes data)
3) Click on the link to open the **website**





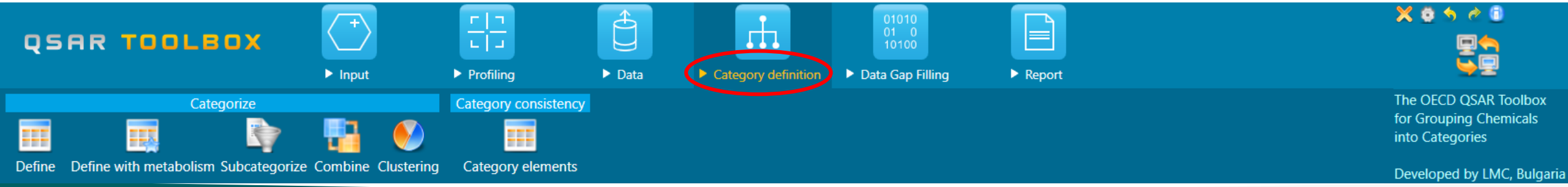


OECD QSAR Toolbox: Category definition

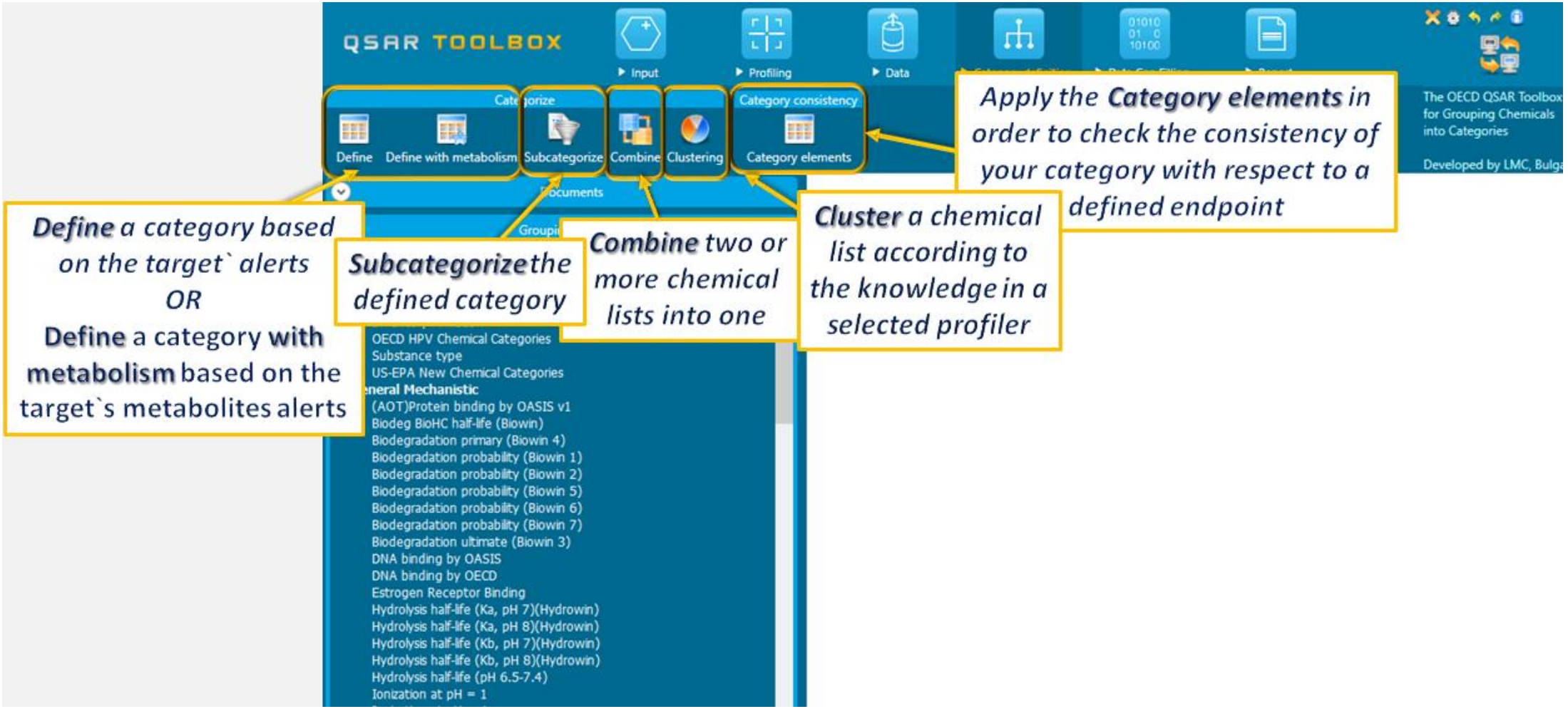
Module for category definition of a target molecule, can be assigned according to:

- structural similiarity (e.g. functional groups)
- mechanistic similarity (e.g. protein binding)
- predefined categories (e.g. OECD HPV Chemical categories)

Substances are grouped through a selected profiler («Profiler»)
Substances and data are searched thorugh the selected («Data»)



QSAR Toolbox: Category definition



QSAR Toolbox: Analogues and categories

Data from analogues can be used to predict the property of a target substance using read-across or trend analysis.

Toolbox can take into account structural, mechanistic and metabolic aspects, to find toxicologically relevant analogues with data on the property of interest.

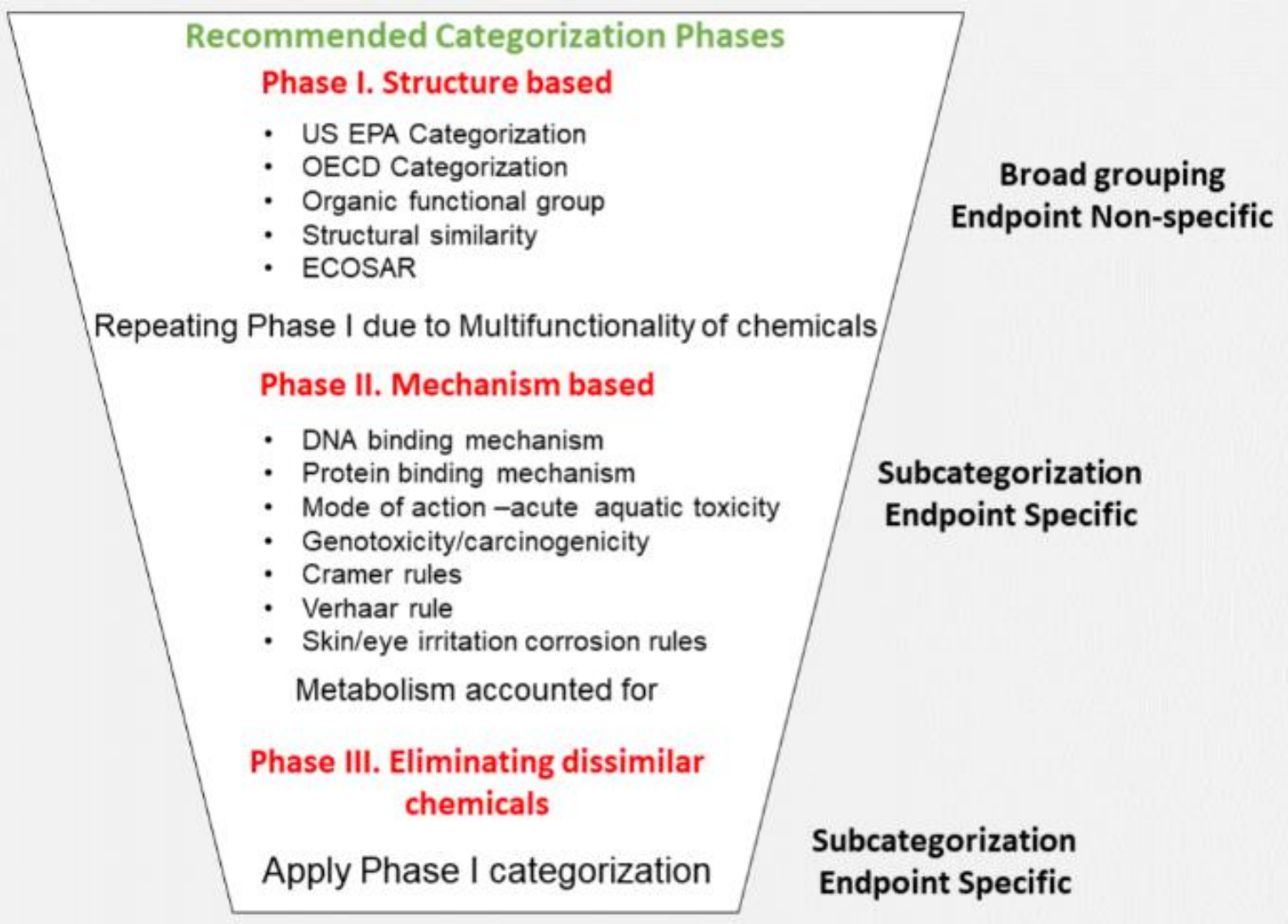
A workflow recomended by ECHA includes different phases: first a categorization structural based, then an endpoint specific, finally a sub categorization with an expert judgment.

Categorization structural based (non specific endpoint)

Sub-categorization endpoint specific

Sub-categorization
Expert judgment

QSAR Toolbox: ECHA recommended categorization



Expert judgment

QSAR Toolbox: Category definition

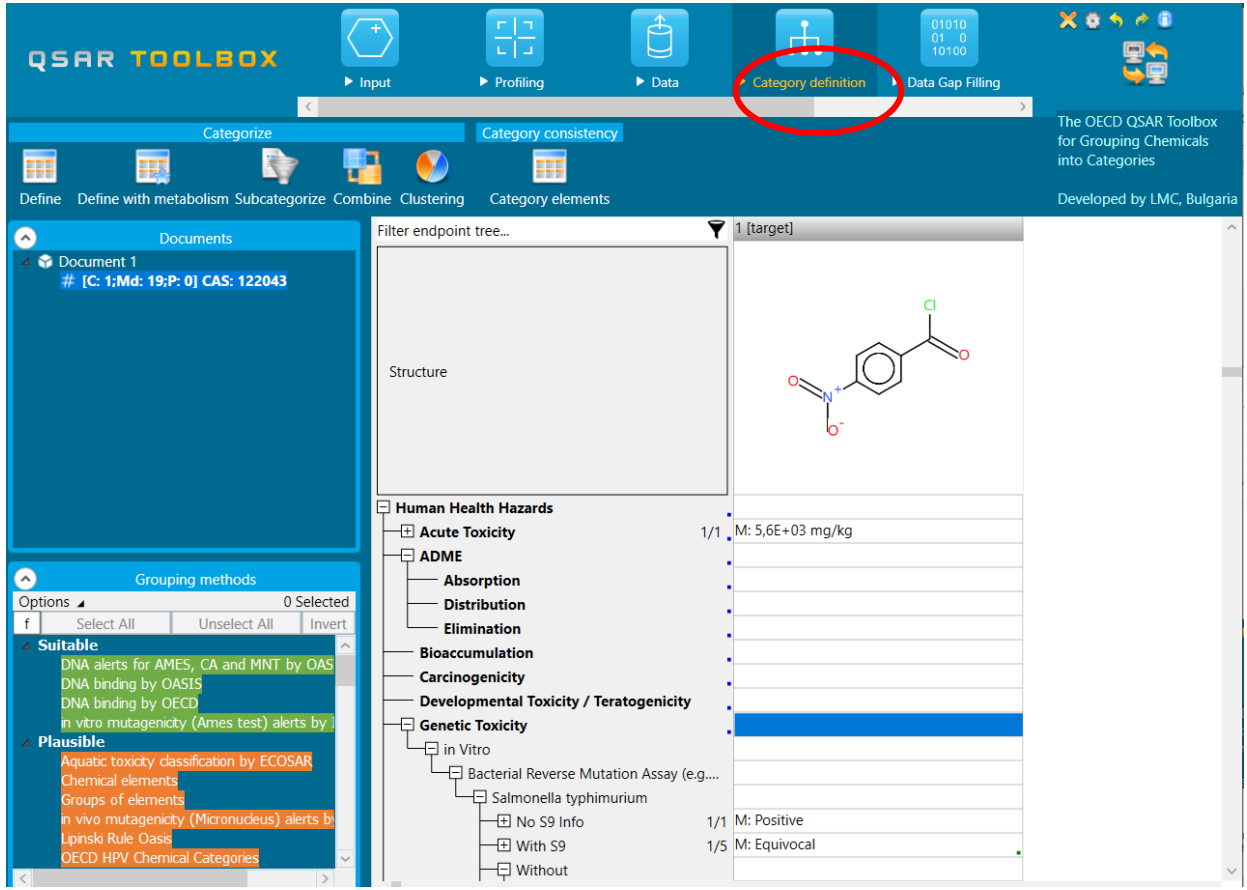
Relevance of the profiler in the «Category definition» can be:

- Green, for relevant profiler
- Orange, for plausible profilers
- With no colour, for profilers with no connection

Metabolism profiler can be applied to take into account in the category definition

Analogues can be searched according these criteria:

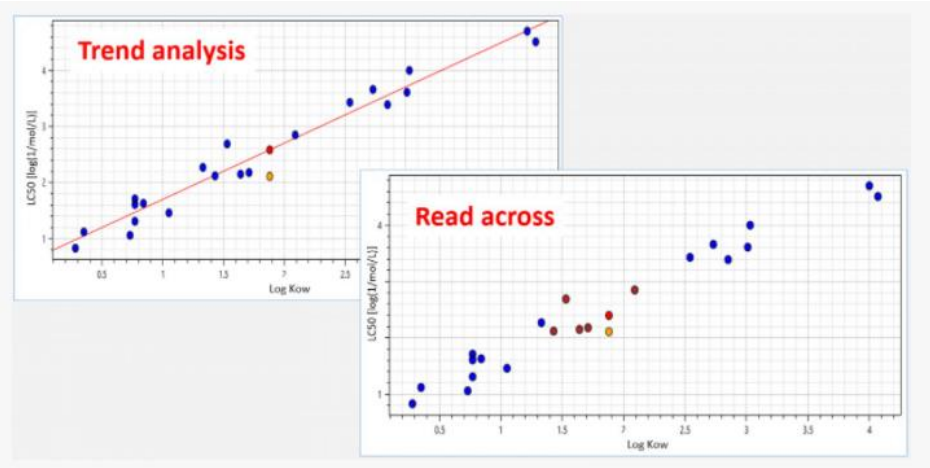
- Parents and metabolites with the same profiler
- Parents and metabolites with a defined profiler
- Common metabolites
- Metabolites similar to a specific compound



QSAR Toolbox: Data Gap Filling

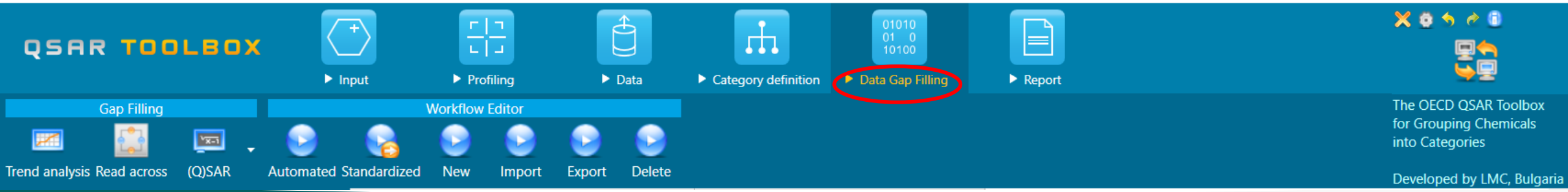
Prediction of the data for the target chemical:

- **Read-across**: using some of the experimental data for some of the closest analogs, identified with the Toolbox
- **Trend analysis:** using all the data from the analogs to derive the regression equation, used to derive the target data
- **QSAR models**: EPIsuite, ECOSAR, DK QSAR database



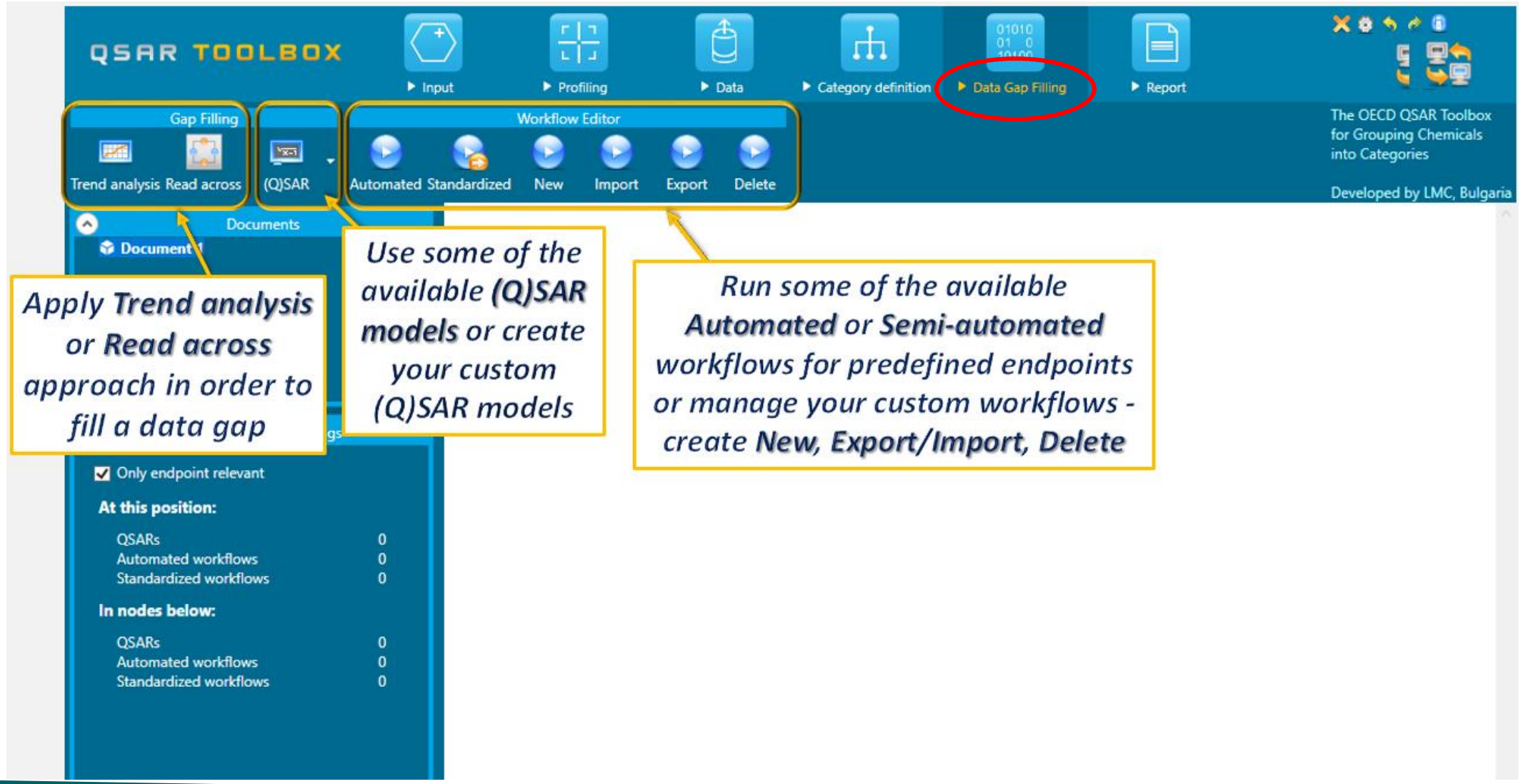
Trend analysis: method of predicting toxicity of a chemical by analyzing toxicity trends of tested chemicals

Read-across: technique of predicting toxicity of the target chemical using data from source chemicals



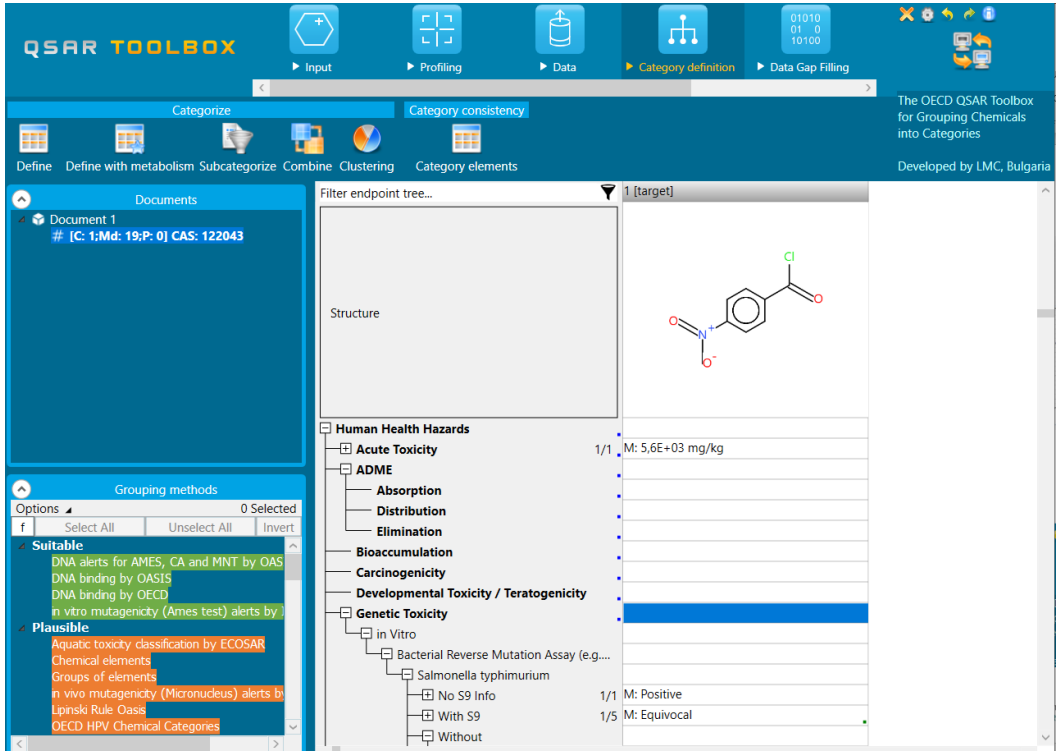
DEPARTMENT ENVIRONMENT AND HEALTH

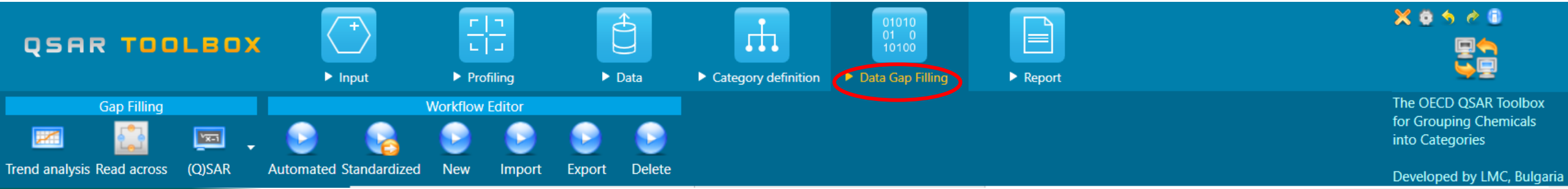
QSAR Toolbox: Data Gap Filling



QSAR Toolbox: Data Gap Filling

- **Relevance** of the profiler used in the sub-categorization, to remove analogues from the prediction is highlighted by colors
- **Workflow**
  - ✓ Automated (fully decided by the TB)
  - ✓ Standardized (partial user selection)
  - ✓ Manual (fully decided by the user)





OECD QSAR Toolbox: Report

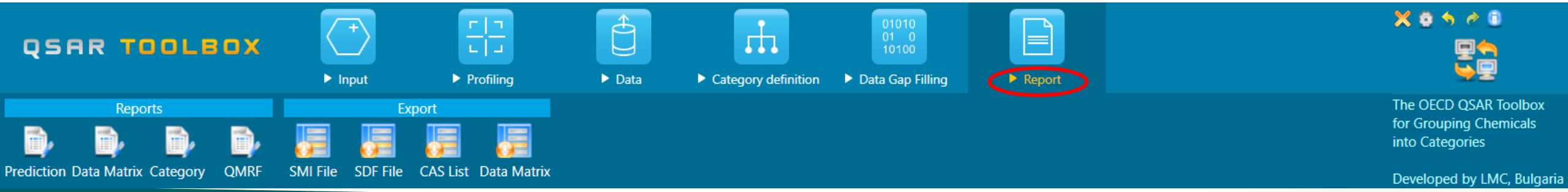
The report module creates a document after accepting the prediction in the previous module. Three different reports can be created:

1. Prediction report
2. Category report
3. Data matrix

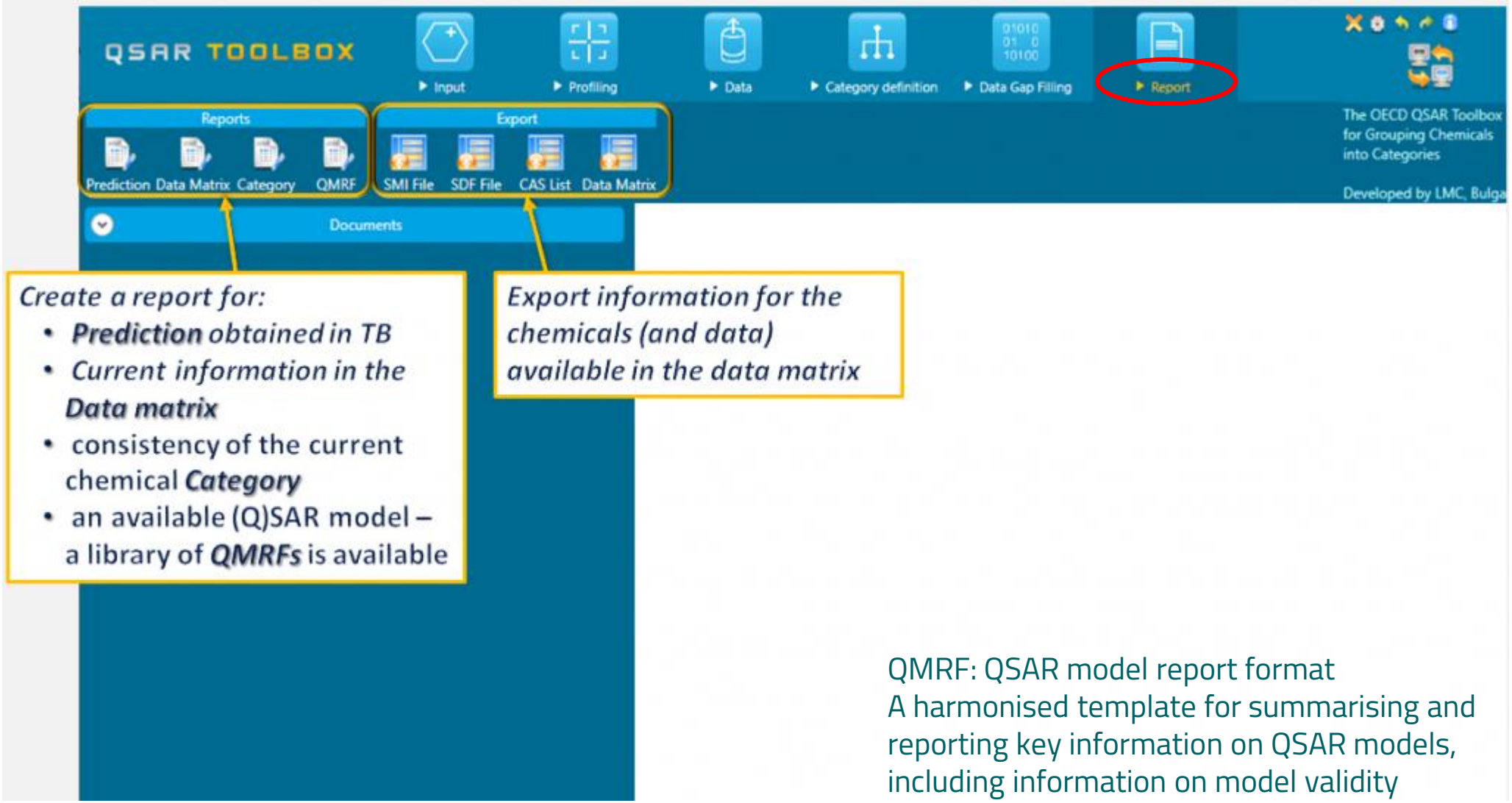
All these reports provide the basis for justifying the reliability of the prediction that needs to be critically reviewed by the users and can be manually completed for comments and explanations

In this module the user can:

- ✓ Choose which section to include in the final report, and in what order
- ✓ Include information and data from analogs in the report
- ✓ Enter comments and interpretation of results in the editable fields



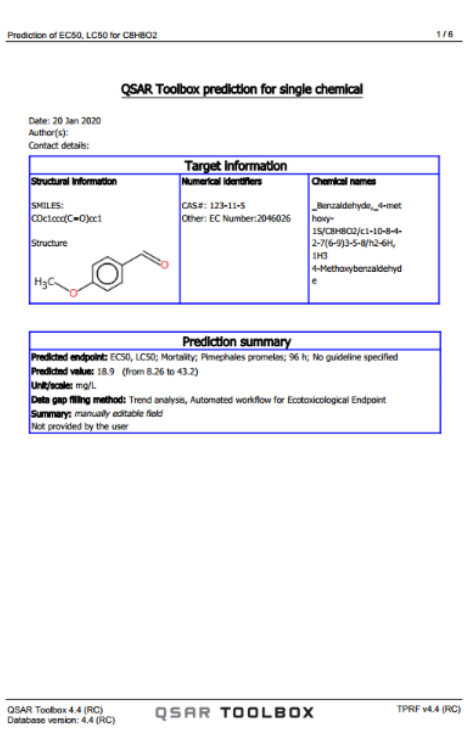
OECD QSAR Toolbox: Report



QMRF: QSAR model report format
A harmonised template for summarising and reporting key information on QSAR models, including information on model validity

QSAR Toolbox: report

1. Prediction report



Prediction of EC50, LC50 for C8H8O2 1 / 6

QSAR Toolbox prediction for single chemical

Date: 20 Jan 2020
Author(s):
Contact details:

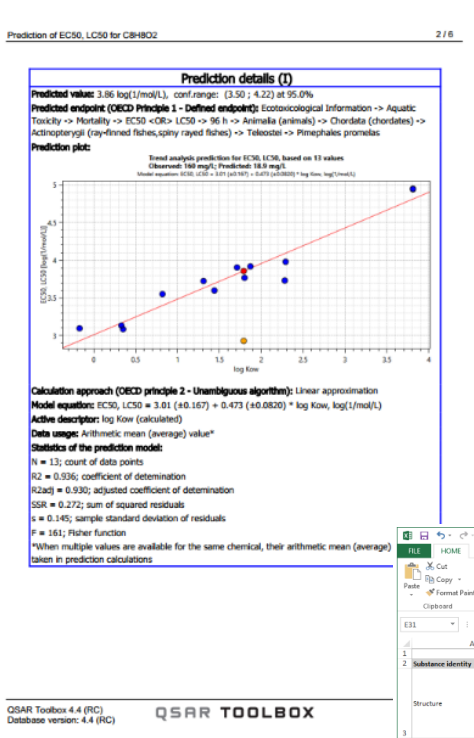
Target information

Structural information	Numerical identifiers	Chemical names
SMILES: COc1ccc(C=O)cc1 Structure	CAS#: 123-11-5 Other: EC Number:2046026	_Benzaldehyde,_4-methoxy- 1S/C8H8O2/c1-10-8-4-2-7(6-9)3-5-8/h2-6H,1H3 4-Methoxybenzaldehyde

Prediction summary

Predicted endpoint: EC50, LC50; Mortality; Pimephales promelas; 96 h; No guideline specified
Predicted value: 18.9 (from 8.26 to 43.2)
Unit/scale: mg/L
Data gap filling method: Trend analysis, Automated workflow for Ecotoxicological Endpoint
Summary: *manually editable field*
Not provided by the user

QSAR Toolbox 4.4 (RC)
Database version: 4.4 (RC)
QSAR TOOLBOX
TPRF v4.4 (RC)



Prediction of EC50, LC50 for C8H8O2 2 / 6

Prediction details (I)

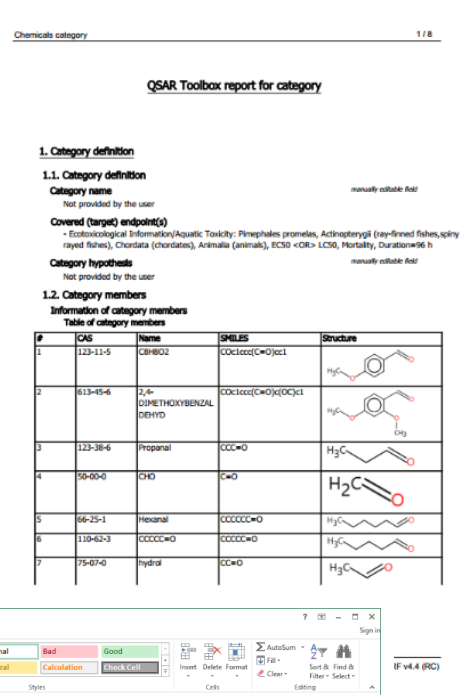
Predicted value: 3.86 log(1/mol/L), conf.range: (3.50 ; 4.22) at 95.0%
Predicted endpoint (OECD Principle 1 - Defined endpoint): Ecotoxicological Information -> Aquatic Toxicity -> Mortality -> EC50 <OR> LC50 -> 96 h -> Animalia (animals) -> Chordata (chordates) -> Actinopterygii (ray-finned fishes,spiny rayed fishes) -> Teleostei -> Pimephales promelas
Prediction plot:

Trend analysis prediction for EC50, LC50, based on 13 values
Observed: 160 mg/l; Predicted: 18.9 mg/l

Calculation approach (OECD principle 2 - Unambiguous algorithm): Linear approximation
Model equation: EC50, LC50 = 3.01 (±0.167) + 0.473 (±0.0820) * log Kow, log(1/mol/L)
Active descriptor: log Kow (calculated)
Data usage: Arithmetic mean (average) value*
Statistics of the prediction model:
N = 13; count of data points
R2 = 0.936; coefficient of detemination
R2adj = 0.930; adjusted coefficient of determination
SSR = 0.272; sum of squared residuals
s = 0.145; sample standard deviation of residuals
F = 161; Fisher function
*When multiple values are available for the same chemical, their arithmetic mean (average) taken in prediction calculations

QSAR Toolbox 4.4 (RC)
Database version: 4.4 (RC)
QSAR TOOLBOX

2. Category report



Chemicals category 1 / 8

QSAR Toolbox report for category

1. Category definition

1.1. Category definition

Category name *manually editable field*
Not provided by the user

Covered (target) endpoint(s)
- Ecotoxicological Information/Aquatic Toxicity: Pimephales promelas, Actinopterygii (ray-finned fishes,spiny rayed fishes), Chordata (chordates), Animalia (animals), EC50 <OR> LC50, Mortality, Duration=96 h

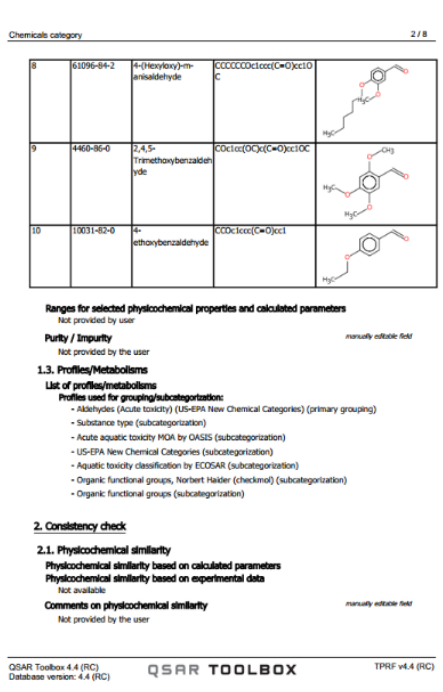
Category hypothesis *manually editable field*
Not provided by the user

1.2. Category members

Information of category members
Table of category members

#	CAS	Name	SMILES	Structure
1	123-11-5	C8H8O2	COc1ccc(C=O)cc1	
2	613-45-6	2,4-DIMETHOXYBENZALDEHYD	COc1ccc(C=O)c(OC)c1	
3	123-38-6	Propanal	CCC=O	
4	50-00-0	CHO	C=O	
5	66-25-1	Hexanal	CCCCCC=O	
6	110-62-3	CCCCC=O	CCCCC=O	
7	75-07-0	hydrol	CC=O	

TPRF v4.4 (RC)



Chemicals category 2 / 8

8	61096-84-2	4-(Hexyloxy)-m-anisaldehyde	CCCCCCOc1ccc(C=O)cc1OC	
9	4460-86-0	2,4,5-Trimethoxybenzaldehyde	COc1cc(OC)c(C=O)cc1OC	
10	10031-82-0	4-ethoxybenzaldehyde	CCOc1ccc(C=O)cc1	

Ranges for selected physicochemical properties and calculated parameters
Not provided by user

Purity / Impurity *manually editable field*
Not provided by the user

1.3. Profiles/Metabolisms

List of profiles/metabolisms
Profiles used for grouping/subcategorization:
- Aldehydes (Acute toxicity) (US-EPA New Chemical Categories) (primary grouping)
- Substance type (subcategorization)
- Acute aquatic toxicity MOA by OASIS (subcategorization)
- US-EPA New Chemical Categories (subcategorization)
- Aquatic toxicity classification by ECOSAR (subcategorization)
- Organic functional groups, Norbert Haider (checkmol) (subcategorization)
- Organic functional groups (subcategorization)

2. Consistency check

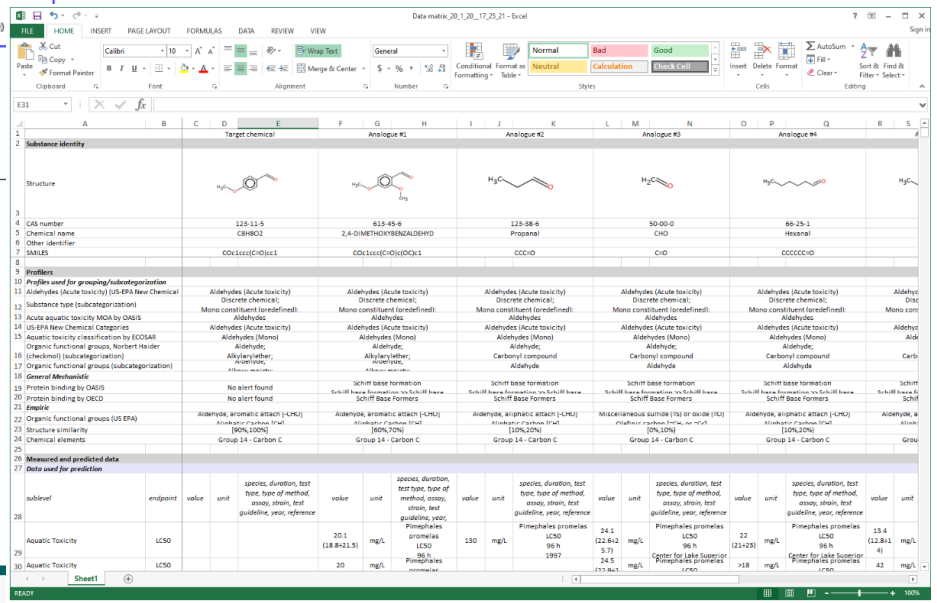
2.1. Physicochemical similarity

Physicochemical similarity based on calculated parameters
Physicochemical similarity based on experimental data
Not available

Comments on physicochemical similarity *manually editable field*
Not provided by the user

QSAR Toolbox 4.4 (RC)
Database version: 4.4 (RC)
QSAR TOOLBOX
TPRF v4.4 (RC)

3. Data matrix



Data matrix_20_1_20__17_25_21 - Excel

	Target chemical	Analogue #1	Analogue #2	Analogue #3	Analogue #4
Substance identity					
Structure					
CAS number	123-11-5	613-45-6	123-38-6	50-00-0	66-25-1
Chemical name	C8H8O2	2,4-DIMETHOXYBENZALDEHYD	Propanal	CHO	Hexanal
Other identifier					
SMILES	COc1ccc(C=O)cc1	COc1ccc(C=O)c(OC)c1	CCC=O	C=O	CCCCCC=O
Profilers					
Profiles used for grouping/subcategorization					
Aldehydes (Acute toxicity) (US-EPA New Chemical	Aldehydes (Acute toxicity) Discrete chemical; Mono constituent (predefined):	Aldehydes (Acute toxicity) Discrete chemical; Mono constituent (predefined):	Aldehydes (Acute toxicity) Discrete chemical; Mono constituent (predefined):	Aldehydes (Acute toxicity) Discrete chemical; Mono constituent (predefined):	Aldehydes (Acute toxicity) Discrete chemical; Mono constituent (predefined):
Substance type (subcategorization)	Aldehydes	Aldehydes	Aldehydes	Aldehydes	Aldehydes
Acute aquatic toxicity MOA by OASIS	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)
US-EPA New Chemical Categories	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)
Aquatic toxicity classification by ECOSAR	Aldehydes;	Aldehydes;	Aldehydes;	Aldehydes;	Aldehydes;
Organic functional groups, Norbert Haider (checkmol) (subcategorization)	Alkylaryl ether;	Alkylaryl ether;	Carbonyl compound	Carbonyl compound	Carbonyl compound
Organic functional groups (subcategorization)	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde
General Mechanistic					
Protein binding by OASIS	No alert found	Schiff base formation	Schiff base formation	Schiff base formation	Schiff base formation
Protein binding by OECD	No alert found	Schiff Base Formers	Schiff Base Formers	Schiff Base Formers	Schiff Base Formers
Empiric					
Organic functional groups (US EPA)	Aldehyde, aromatic attach (-CHO)	Aldehyde, aromatic attach (-CHO)	Aldehyde, aliphatic attach (-CHO)	Miscellaneous sulfide (TS) or oxide (TO)	Aldehyde, aliphatic attach (-CHO)
Structure similarity	[90%,100%]	[60%,70%]	[10%,20%]	[0%,10%]	[10%,20%]
Chemical elements	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C	Group 14 - Carbon C
Measured and predicted data					
Data used for prediction					
sublevel / endpoint / value / unit / species, duration, test type, type of method, assay, strain, test guideline, year, reference					
Aquatic Toxicity LC50		20.1 (18.8+21.5) mg/L Pimephales promelas LC50 96 h 1997	130 mg/L Pimephales promelas LC50 96 h 1997	24.1 (22.6+25.7) mg/L Pimephales promelas LC50 96 h Center for Lake Superior	22 (21+23) mg/L Pimephales promelas LC50 96 h Center for Lake Superior
Aquatic Toxicity LC50		20 mg/L Pimephales promelas		24.5 mg/L Pimephales promelas	>18 mg/L Pimephales promelas

	Pimephales promelas LC50 96 h Center for Lake Superior	Pimephales promelas LC50 96 h
Aquatic Toxicity LC50	13.4 (12.8+14.4) mg/L	42 mg/L

Sheet1
READY

User can customize the report, inserting comments and interpreting the data in the editable fields

QSAR Toolbox: Take home messages

- **Free software application**: ECHA, OECD and LMC, contribution from donors of data, profilers, models
- Supports (eco)toxicologists in performing **reproducible** and **transparent** chemical hazard assessment using non-animal methods
- Source of **QSAR models, existing data, metabolic and mechanistic information** can be used to fill data gaps
- The **profiler results** are **not predictions,** but they can contribute to find suitable analogues with data to build read-across, or they can be used in a WoE approach
- Simplified user interface and automatic worlkflow can help user, together with support section (manuals, forum, help-desk) in the website www.qsartoolbox.org
- TB should not be considered a Black Box, in which a user uncritically enter an input and get an output, without expert supervision
- Specific expertise is required e.g., in Toxicology/computational and organic chemistry/IT skills: expert judgement is required in critical step such as for the identification of the analogues, the relevance and reliability of the data

→Practical exercises with QSAR Toolbox

Application of TTC approach for a 'data poor' chemical

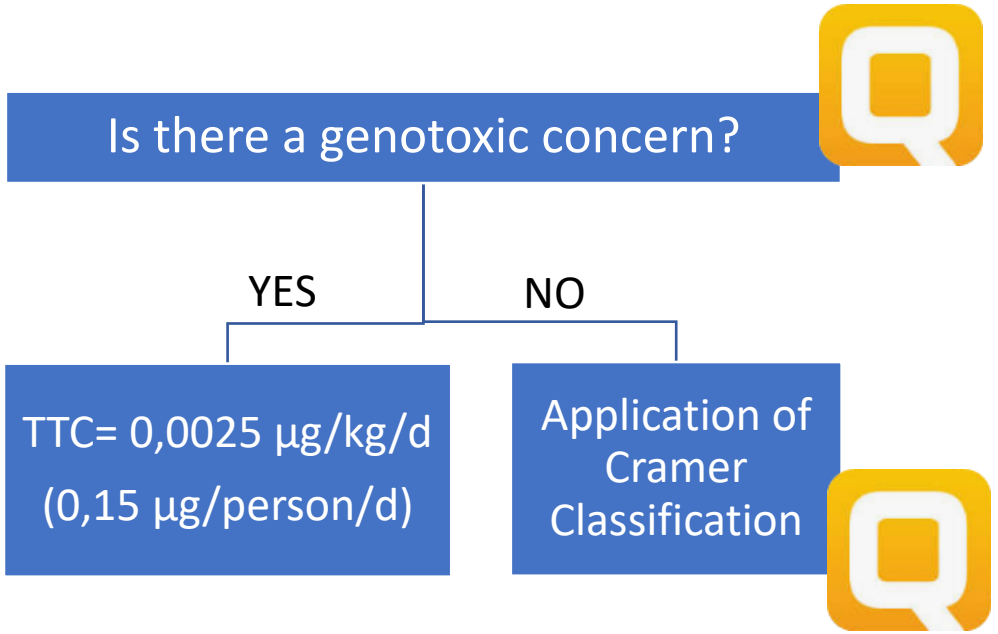
Hands-on exercise with the QSAR Toolbox

Cecilia BOSSA
Environment and Health Department-Istituto Superiore di Sanità

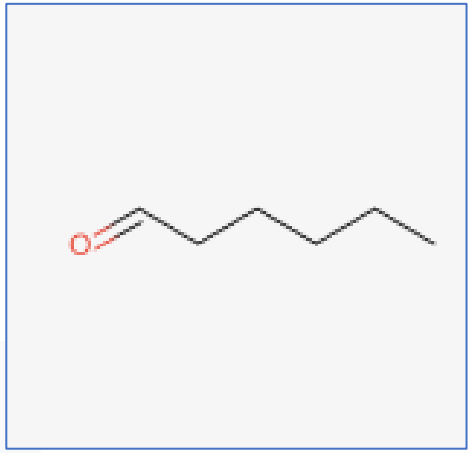
Practical exercises with QSAR Toolbox

Application of TTC approach for a 'data poor' chemical

1) The first step is the **genotoxicity** assessment
   - For the potential DNA-reactive mutagens the TTC value of 0,0025 µg/kg/d (0.15 µg/person/day) is considered
   - If No genotoxic concern is assigned, to the substance, the Cramer classification scheme is applied



n-hexanal



Leibniz Institute for Food Systems Biology at the Technical University of Munich

WaterTOP COST Action (CA18225)

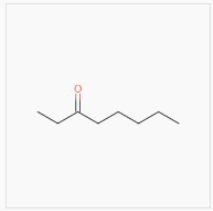
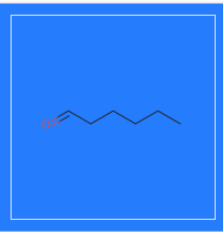
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

WaterTOP COST Action (CA18225)
for Water Taste and Odor Problems

About Publication Flavor Wheel

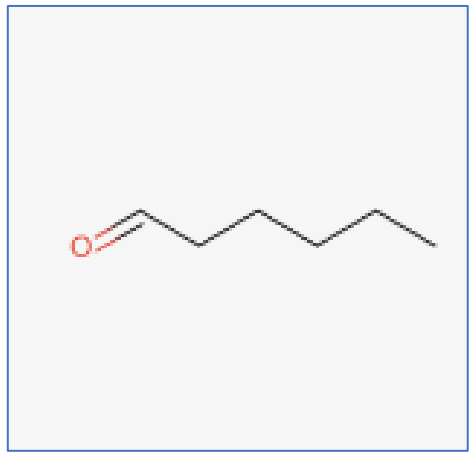
CSV PDF Print Column visibility ▾ Show 10 entries Search: hexanal

Compound	Structure	MW	Quality	OT	Links	Source
Search Compound	Structure Search Tool	Min Max	Search Quality	Min Max	Search Links	Search Source
3-Octanone		128.21	mouldy (details)	39 µg/L(details)	PubChem:246728 FSBI-DB	
Hexanal		100.16	grassy, fatty, lettuce heart (details)	4.5 µg/L(details)	PubChem:6184 FSBI-DB	algae (details)

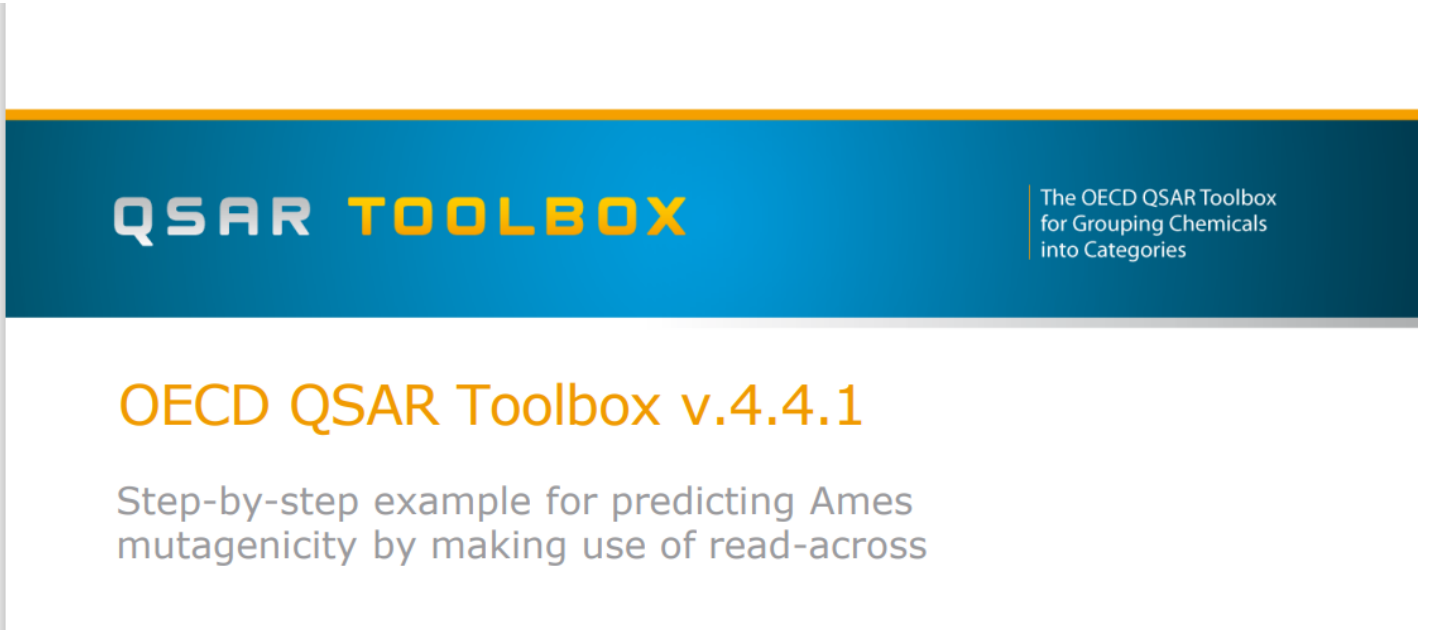
OT=4,5 µg/L

ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT ENVIRONMENT AND HEALTH

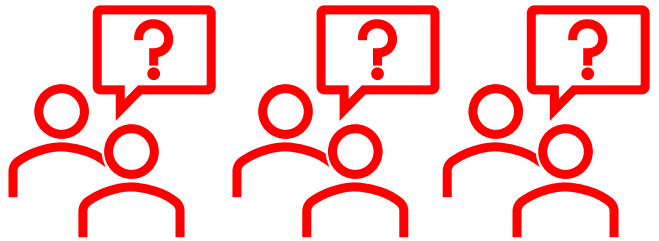
QSAR Toolbox: mutagenicity prediction by Read Across approach



n-hexanal



https://qsartoolbox.org/wp-content/uploads/2020/04/Tutorial_2_Predicting-AMES-by-making-use-of-read-across.pdf



Chiara L Battistelli
and Cecilia Bossa

ISTITUTO SUPERIORE DI SANITÀ
DEPARTMENT
ENVIRONMENT AND HEALTH