

QSAR TOOLBOX

The OECD QSAR Toolbox
for Grouping Chemicals
into Categories

User manual

Getting Started

Getting Started

Document history

| Version | Comment |
|-------------|--|
| Version 1.0 | October 2010: Manual based on QSAR Toolbox version 2.0 |
| Version 2.0 | October 2012: Manual based on QSAR Toolbox version 3.0 |

Issue date: October 2012

Language: English

If you have questions or comments that relate to this document, please send them to ehscont@oecd.org or visit the QSAR Toolbox discussion forum at https://community.oecd.org/community/toolbox_forum

Getting Started

Table of Contents

| | |
|---|-----------|
| Foreword | 4 |
| Introduction | 5 |
| 1 Input | 6 |
| 1.1 Aim of the module "Input" | 6 |
| 1.2 Example | 6 |
| 2 Profiling | 9 |
| 2.1 Aim of the module "Profiling" | 9 |
| 2.2 Example | 12 |
| 3 Endpoints | 14 |
| 3.1 Aim of the module "Endpoints" | 14 |
| 3.2 Example | 20 |
| 4 Category definition | 23 |
| 4.1 Aim of the module "Category definition" | 23 |
| 4.2 Example | 31 |
| 5 Filling data gap | 35 |
| 5.1 Aim of the module "Filling data gap" | 35 |
| 5.2 Example | 38 |
| 6 Report | 46 |
| 6.1 Aim of the module "Report" | 46 |
| 6.2 Example | 46 |

Getting Started

Foreword

The current document is aimed at introducing the user to 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)].

This present "Getting Started" document guides the user step-by-step through the basic functionalities of the Toolbox with the help of an example. Additional guidance and training material will be made 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/chemicalsafety/testingofchemicals/oecdquantitativestructure-activityrelationshipsprojectqsars.htm>] as well as the internet site of the developer of the QSAR Toolbox, the Laboratory of Mathematical Chemistry [<http://toolbox.oasis-lmc.org/>]. The user is invited to regularly consult these internet sites. The (Q)SAR Toolbox has been developed in collaboration with the European Chemical Agency.

Getting Started

Introduction

The Toolbox has multiple functionalities allowing the user to perform a number of operations, for example:

- Identify analogues for a chemical, retrieve experimental results available for those analogues and fill data gaps by read-across or trend analysis.
- Categorize large inventories of chemicals according to mechanisms or modes of action.
- Fill data gaps for any chemical by using the library of QSAR models.
- Evaluate the robustness of a potential analogue for read-across.
- Evaluate the appropriateness of a (Q)SAR model for filling a data gap for a particular target chemical.
- Build QSAR models.

This simple guidance document focuses on the first operation i.e. identify analogues for a chemical, retrieve experimental results available for those analogues and fill data gaps by read-across or trend analysis. The user will be guided step by step through an example to illustrate the functionalities of the Toolbox. The chosen example is filling a data gap for skin sensitization for 4-nitrobenzoyl chloride (CAS No 122-04-3).

The advanced functionalities will be explained in other guidance documents.

Getting Started

1 Input



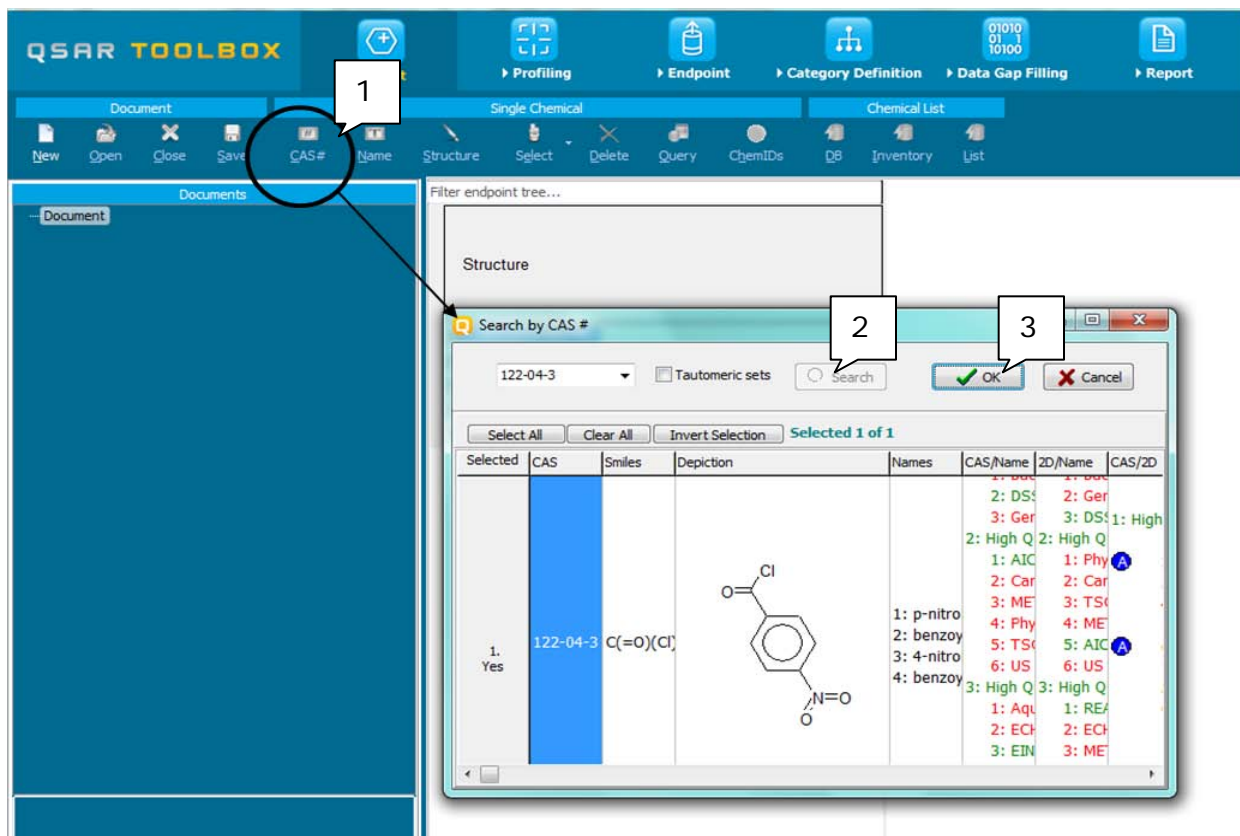
1.1 Aim of the module “Input”

In the “Input” module, the user defines the identity of the target chemical on which to apply the functionalities of the toolbox. The identity of the chemical can be entered by name, CAS number, by drawing the molecule, by SMILES or by selecting it from a list, a database, an inventory or a file.

1.2 Example

In the example selected for this guidance document and as illustrated in the screenshot below, the user can simply press the button **CAS #** (1), enter the CAS number of 4-nitrobenzoyl chloride (CAS No 122-04-3) without hyphens (2), press the button **Search** (3) and once the correct structure appears, press **OK** (4).

Getting Started



In case a structure has several CAS numbers or a structure could be related to more than one substance (e.g. in the case of compounds), more than one chemical identity could be retrieved. In this case the user can decide which substance is to be retained for the subsequent workflow.

The selected substance identity is then added to the data matrix and the user can proceed to the next module (see screenshot below).

Getting Started



The colour code indicates the reliability of the chemical identifier:

Green: There is a high reliability between the identifier and the structure. This colour is attributed if the identifier is the same in several quality assured databases.

Yellow: There is only a moderate reliability between the identifier and the structure. The color is attributed if the identifier is the same in several databases for which the quality assurance could not be established.

Red: There is a poor reliability between the identifier and the structure. The color is attributed if the identifier is allocated to different structures in different databases.

Getting Started

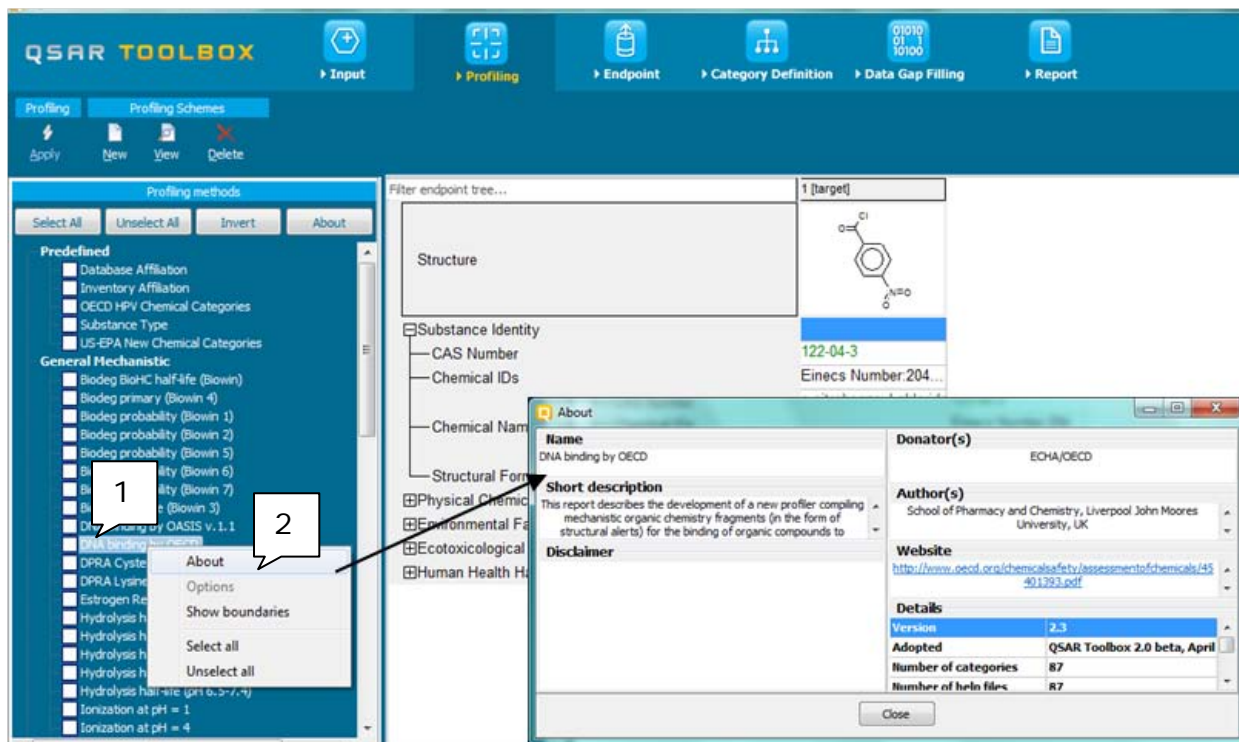
2 Profiling



2.1 Aim of the module “Profiling”

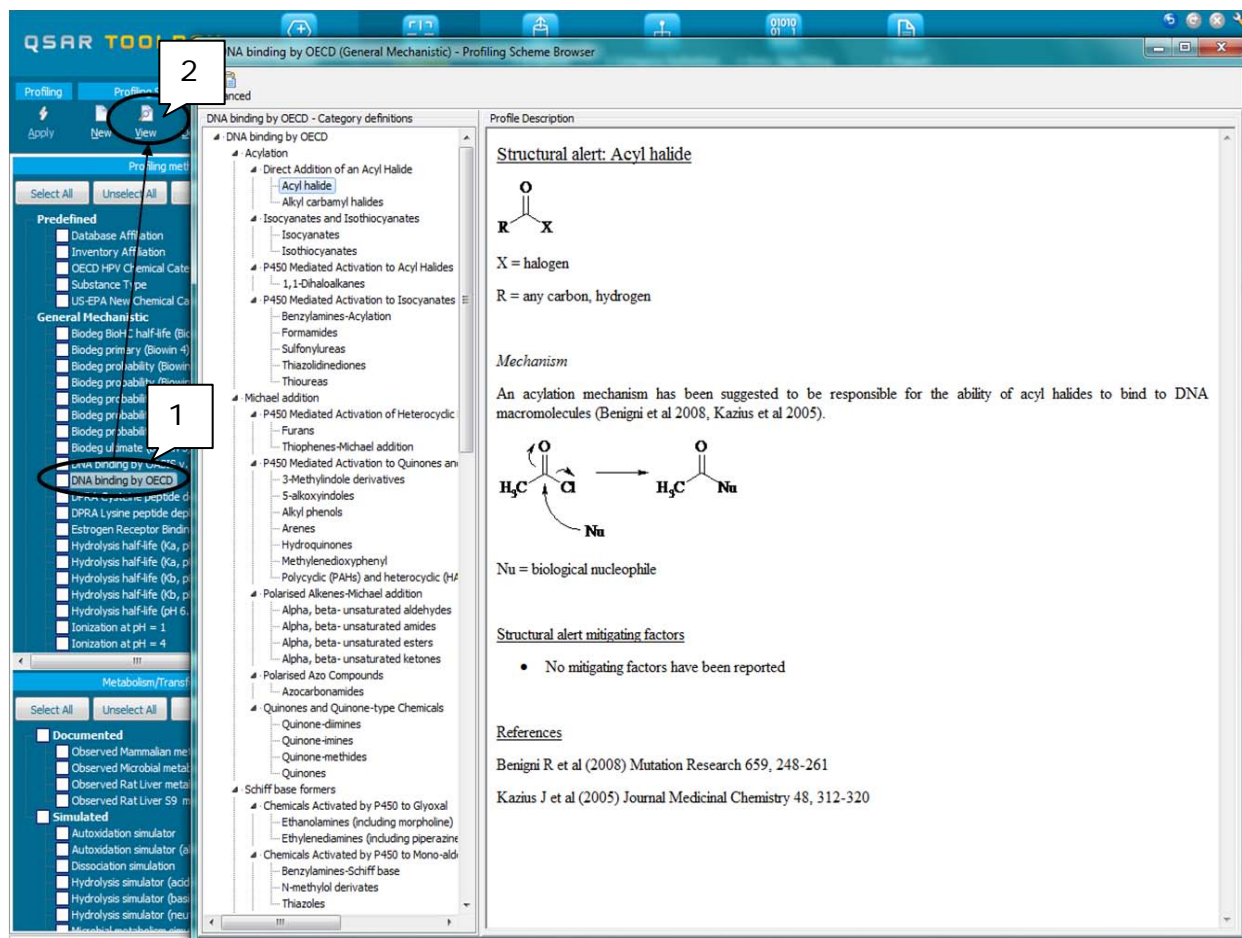
With the module “Profiling”, the user can retrieve information based on the identity of the substance or its structure. The Toolbox currently contains a large list of “profilers” that identify the affiliation to previously defined categories (e.g. the categories defined by the US-EPA for the assessment of new chemicals or the categories assessed within the OECD HPV Chemicals Programme), mechanisms or modes of action as well as observed or simulated metabolites. Summary information can be accessed by selecting the name of the profiler (1) performing a right click on the name of the profiler and then selecting **About** (2).

Getting Started



For most of the profilers, detailed background information can be retrieved by highlighting one of the profilers (1) and pressing the button **View** (2).

Getting Started



The outcome of the profiling also determines the most appropriate way to search for analogues (see Chapter 4). Indeed not all profilers are relevant for all regulatory endpoints. While some general mechanistic profilers such as Protein binding are relevant for many endpoints (skin sensitisation, acute toxicity, aquatic toxicity), some profilers are relevant for building categories for very specific endpoints only. Table 4-1 in chapter 4 lists a selection of profilers and their relevance for different endpoints of regulatory relevance.

Getting Started

2.2 Example

Continuing the example for 4-nitrobenzoyl chloride (CAS No 122-04-3), after checking all the profilers (1) and pressing the button **Apply** (2), the result of the profilers is inserted into the matrix (3), as shown in the following screenshot.

Getting Started

The screenshot displays the QSAR Toolbox software interface. At the top, there are navigation buttons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', and 'Data Gap Filling'. The main window is divided into several panels:

- Profiling Schemes Panel (Left):** Contains a toolbar with 'Apply', 'New', 'View', and 'Delete' buttons. Below it is a list of 'Profiling methods' categorized into 'Predefined' and 'General Mechanistic'. A callout box labeled '1' points to the 'Predefined' section, and another callout box labeled '2' points to the 'Apply' button.
- Filter endpoint tree... Panel (Middle):** Shows a tree view of endpoints. The 'Profile' section is expanded, showing sub-categories like 'Predefined', 'Database Affiliation', 'Inventory Affiliation', 'OECD HPV Chemical Categories', 'Substance Type', and 'US-EPA New Chemical Categories'. The 'General Mechanistic' section is also expanded, listing various biodegradation and DNA binding methods.
- Results Panel (Right):** Displays the results for a specific chemical (1 [target]). At the top, the chemical structure is shown. Below it, a list of endpoints and their corresponding results is provided. For example, under 'US-EPA New Chemical Categories', the result is 'Acid Chlorides'. Other results include 'Not calculated', 'days - weeks', and 'Does NOT Biodegrade Fast'.

The different profilers reveal a number of specific results. The chemical is listed in a number of inventories and fits into specific categories of a number of categorization schemes. For example a specific chemical category according to US-EPA New chemical categories (Acid halides) is recognised for this chemical. By double click on a specific profiling result, the

Getting Started

user can access the reason for identifying these specific mechanisms for the target chemical, as well as the background information regarding the reactions involved.

The screenshot displays the QSAR Toolbox software interface. On the left is a sidebar with 'Profiling' and 'Profiling Schemes' tabs, and a list of 'Profiling methods' including 'Predefined' and 'General Mechanistic' categories. The main window is titled 'Acid Chlorides' and contains several panels: 'Target' (showing a chemical structure), 'Boundaries' (a network diagram), 'Metabonom' (a simulation window), and 'Profile Description' (containing a definition and comments for 'Acid Chlorides'). A callout box labeled '1' points to the 'Acid Chlorides' category in the 'Profile Description' window. Below the 'Profile Description' window, a list of 'Substance Type' and 'US-EPA New Chemical Categories' is visible, with 'Acid Chlorides' highlighted in red.

Getting Started

3 Endpoint



3.1 Aim of the module “Endpoint”

In the module “Endpoints” the user can retrieve experimental results from the resident databases.

Version 3.0 of the Toolbox contains a number of resident databases, as outlined in Table 3-1.

Table 3-1: resident databases and description of their content

| Database | Content : experimental results for |
|--------------------------------------|---|
| Aquatic ECETOC | Aquatic toxicity |
| Aquatic Japan MoE | Aquatic toxicity |
| Aquatic OASIS | Aquatic toxicity |
| Aquatic US-EPA ECOTOX | Aquatic toxicity and bioconcentration |
| Bacterial mutagenicity ISSSTY | Gene mutation, Chromosomal aberration |
| Bioaccumulation CANADA | Bioaccumulation in aquatic organisms |
| Bioaccumulation fish CEFIC LRI | Bioaccumulation in fish |
| Bioconcentration NITE | Bioconcentration factor |
| Biodegradation in soil OASIS | Ready biodegradation in soil |
| Biodegradation NITE | Ready biodegradation |
| Biota-Sediment Accumulation Factor | Biota-sediment accumulation factors (BSAFs) |
| Carcinogenic Potency Database (CPDB) | Long-term animal cancer studies |

Getting Started

| | |
|--|--|
| Carcinogenicity & mutagenicity ISSCAN | Long-term carcinogenicity bioassay on rodents |
| Chemical Reactivity COLIPA | Data for depletion of cysteine and lysine based on peptide reactivity assays |
| Dendritic cells COLIPA | Data associated with two assays: human cell line activation test (h-CLAT) and The myeloid U937 skin sensitisation test (MUSST) |
| ECHA CHEM | The ECHA CHEM database includes information on chemical substances manufactured or imported in Europe. The information originates from the registration dossiers, submitted by companies to the European Chemicals Agency (ECHA) in the framework of the REACH Regulation. The amount of information available for each substance depends on the information requirements specified under REACH - for example, the higher the production volume of the substance, the more information the companies need to provide |
| Estrogen Receptor Binding Affinity OASIS | Estrogen Receptor Binding Affinity, expressed as relative binding affinities in comparison with the estradiol affinity |
| Experimental pKa | pKa values |
| Eye irritation ECETOC | Rabbit eye irritation |
| GSH Experimental RC50 | Abiotic thiol reactivity electrophiles reacting via the Michael addition reaction |
| Developmental toxicity ILSI | dose-response data on treatment-related effects for a range of fetal survival endpoints |
| Repeated Dose Toxicity HESS | Repeated dose toxicity |
| Hydrolysis rate constant OASIS | Experimental data for neutral hydrolysis |
| Cell Transformation Assay ISSCTA | Data for four in vitro cell transformation assays: the SHE (pH 6.7 and pH 7 or higher), the BALB/c 3T3, the C3H/10HT1/2 and the Bhas 42 assays |

Getting Started

| | |
|---|--|
| Keratinocyte gene expression Givaudan | EC1.5, EC2 and EC3 (the concentrations eliciting a 1.5-, 2- and 3-fold increase in luciferase induction) |
| kM database Environment Canada | In vivo metabolic biotransformation rate constants in aquatic organisms |
| Micronucleus ISSMIC | In vivo micronucleus mutagenicity assay in rodents |
| Micronucleus OASIS | In vivo micronucleus mutagenicity assay in rodents |
| MUNRO non-cancer EFSA | NOEL, LOEL data |
| Genotoxicity OASIS | Experimental results for genotoxicity |
| Phys-chem EPISUITE | Physical-chemical properties |
| Rep Dose Tox Fraunhofer ITEM | Repeated dose toxicity |
| Rodent Inhalation Toxicity Database | Data from rat inhalation studies |
| Skin irritation | Primary Skin Irritation Indices from skin irritation tests |
| Skin sensitization | Skin and respiratory sensitization |
| Skin sensitization ECETOC | Skin sensitization |
| Terrestrial US-EPA ECOTOX | Terrestrial toxicity, mammalian toxicity and bioconcentration |
| Toxicity Japan MHLW | Single dose toxicity test and mutagenicity tests |
| ToxRefDB US-EPA | Chronic toxicity, developmental and reprotoxicity tests |
| Yeast estrogen assay database University of Tennessee-Knoxville (USA) | Relative gene expression of 17 β -estradiol |

Getting Started

Based on the different types of experimental data a new database organization was developed for TB 3.0. The organization of databases follows the organization of predefined nodes of Endpoint tree

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

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

Visualization of distribution of all databases implemented in TB v3.0 is shown on Figure 1.

Getting Started

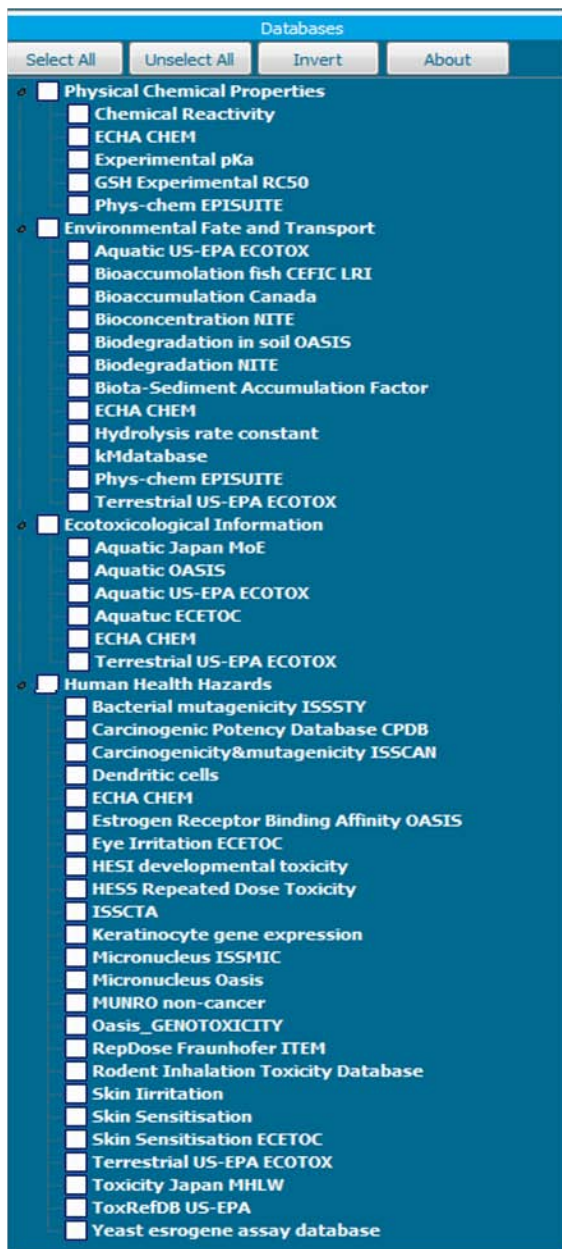
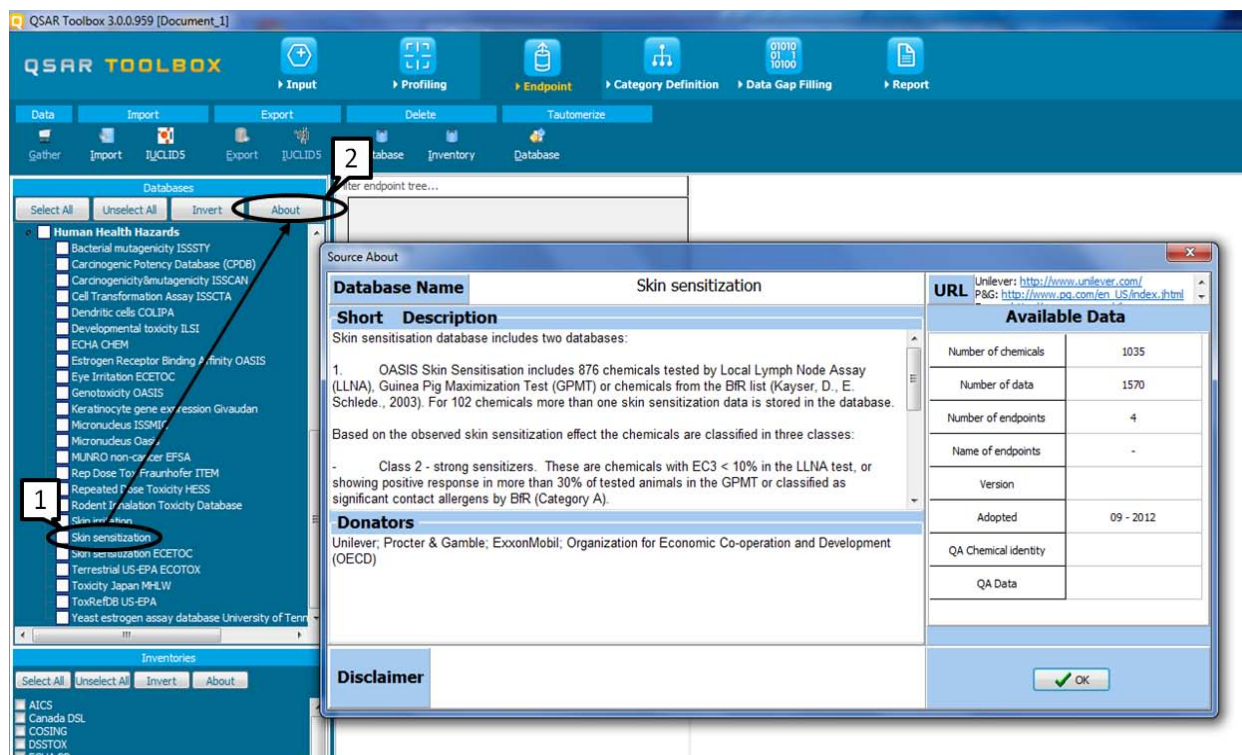


Figure 1

The user can retrieve background information for each database by selecting the name of the database (1) and then clicking the button **About** (2).

Getting Started



The user can choose from which databases data should be retrieved, by checking the box next to the name of the database.

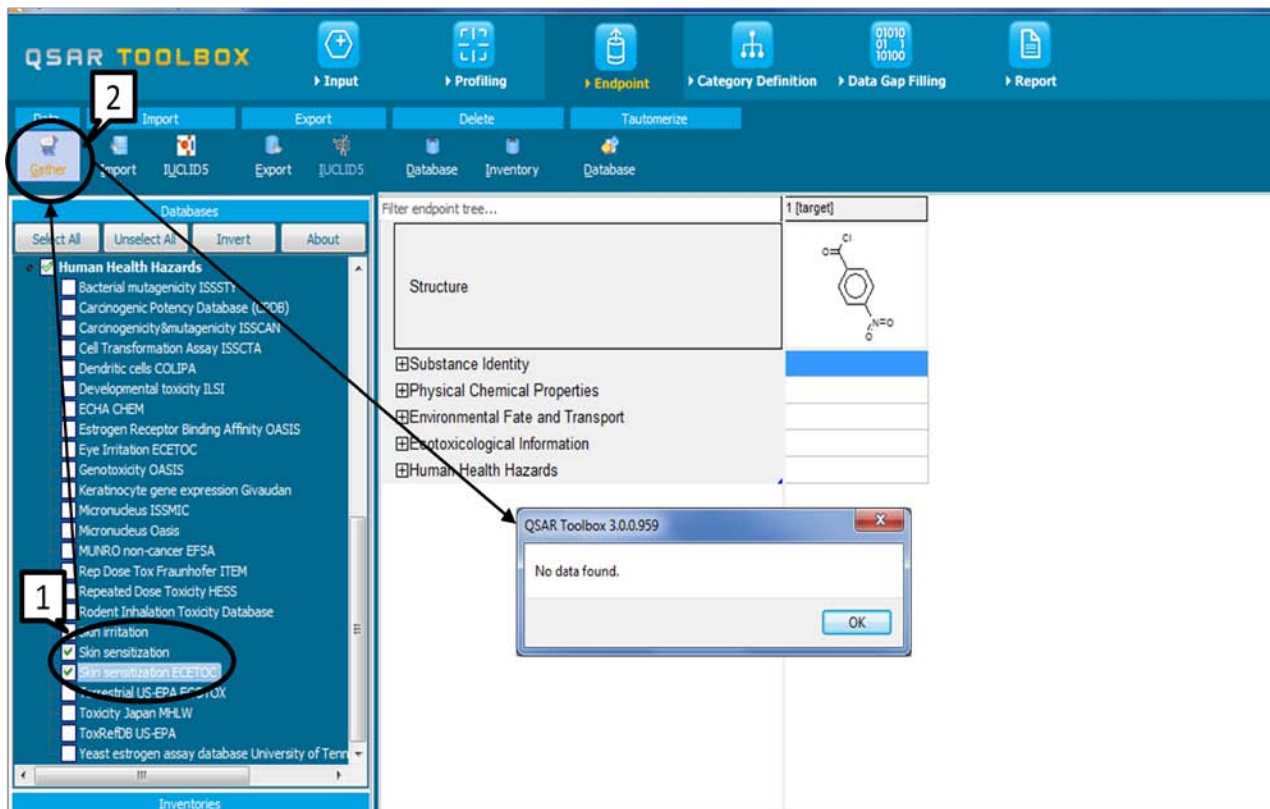
The retrieved data for the target chemical is then inserted into the data matrix for the target chemical (see example below). When the user searches for analogues of the target compound, as outlined in Chapter 4, the results in the selected databases are automatically retrieved and added into the data matrix.

3.2 Example

Continuing the example for 4-nitrobenzoyl chloride (CAS No 122-04-3), the user can decide to limit the search to experimental results for sensitization. Currently two databases - Skin sensitisation ECETOC and Skin sensitisation - contain experimental results for sensitisation.

Getting Started

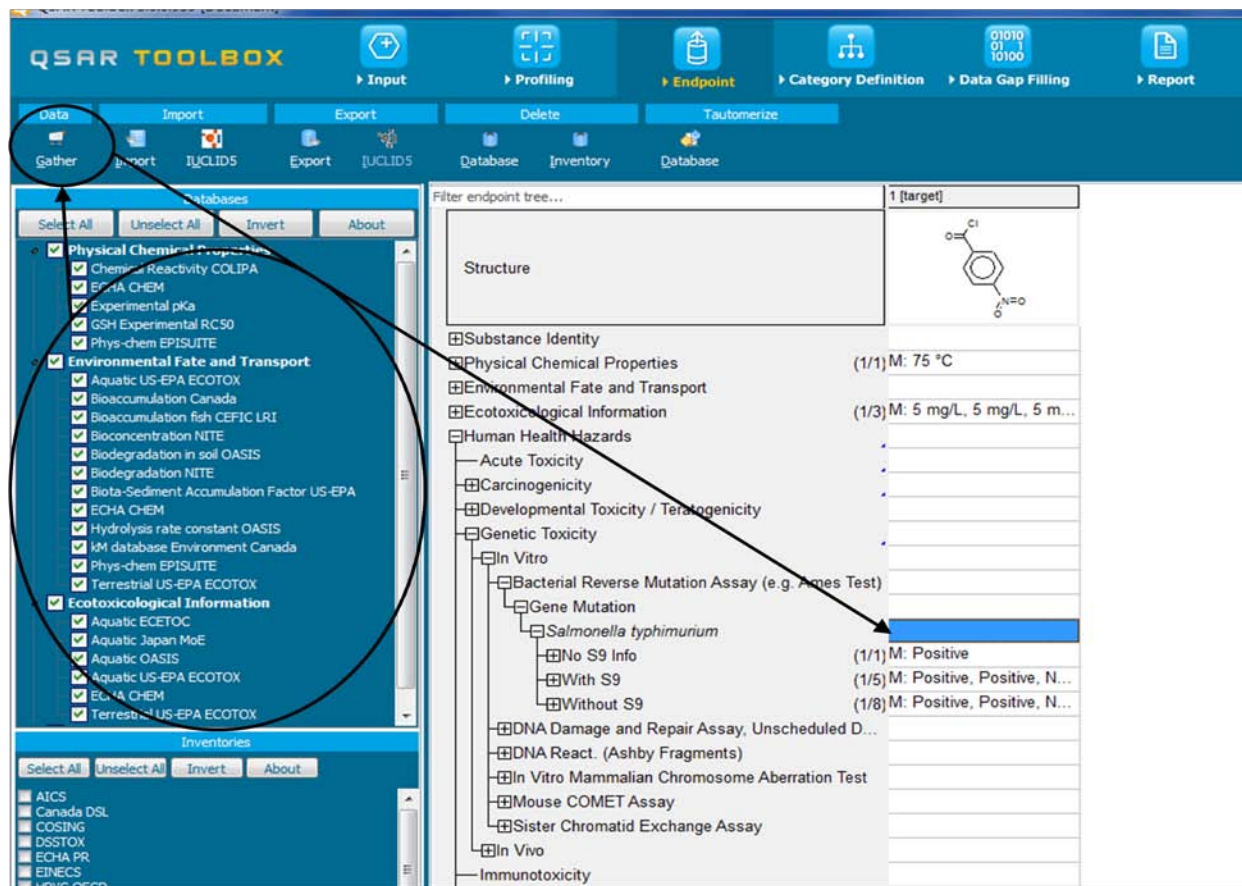
The user may therefore select only these databases (1) and then launch the query by pressing the button **Gather** (2).



No experimental data on sensitisation is available for the target chemical and no data is retrieved.

If the user wants to find out if experimental data is available for other endpoints, s/he may select other databases containing experimental results and re-launch the query.

Getting Started

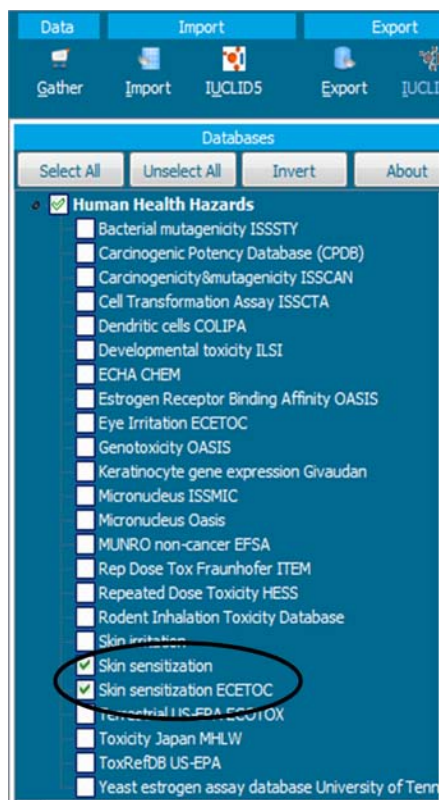
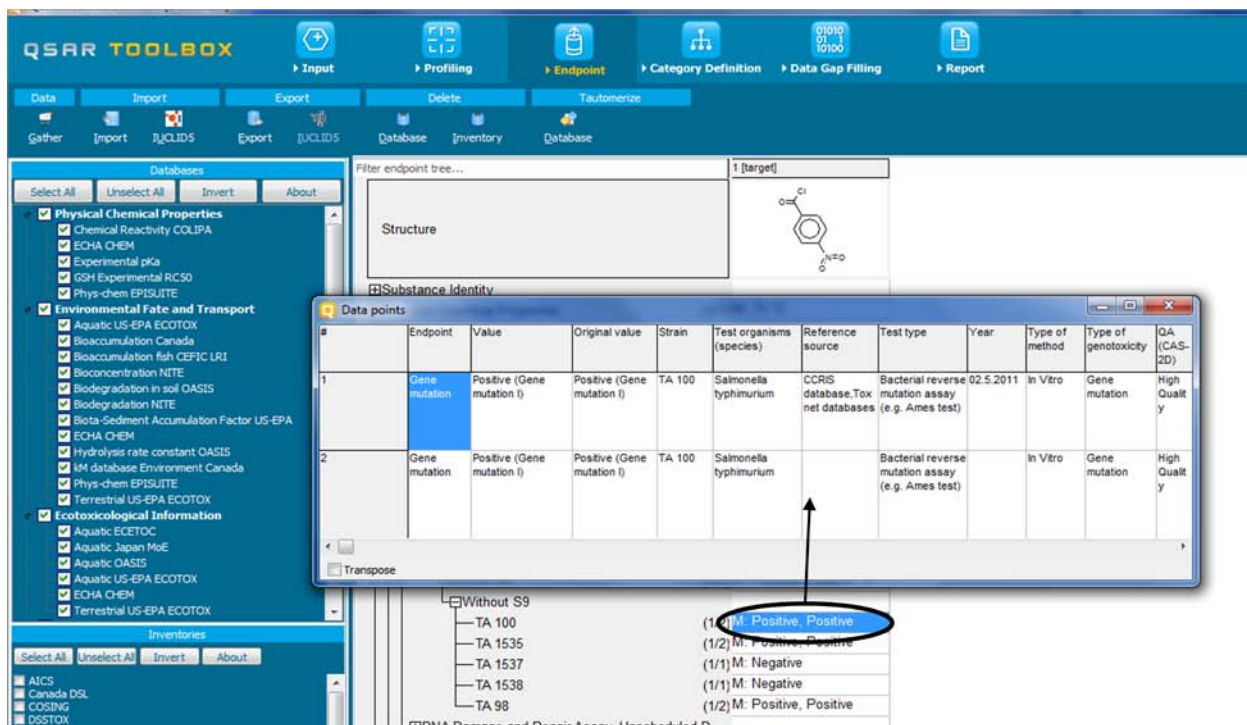


The user can navigate through the data tree by closing or opening the nodes of the tree by double clicking on it. In this example, results from an Ames mutagenicity test are available.

Use the “filter endpoint tree” option in the upper left corner to quickly find the information of interest. For example type the name of the species, duration, endpoint etc to display a selection of data.

By double clicking on a cell in the data matrix, additional information on the test result is made available.

Getting Started



Before continuing to the next module, it is recommended to reset the selection of databases to the databases containing sensitisation results, so as to reproduce the results outlined in the next chapters.

Getting Started

4 Category definition



4.1 The Aim of the module “Category definition”

In the module “Category definition”, the user can group chemicals into chemical categories according to different measures of “similarity”, e.g. specific mechanisms or modes or action or structural similarity, so that within a category data gaps can be filled by read-across or trend analysis. For example within a large inventory, the chemicals can be grouped according to their aquatic toxicity mode of action. Or, starting from a target chemical for which a specific protein binding mechanism is identified, analogues can be found which can bind by the same mechanism and for which experimental results are available. Summary information on some grouping methods is outlined in Table 4-1.

Table 4-1: Summary information for the grouping methods

| Grouping method | Summary background information |
|--------------------------------------|---|
| GENERAL MECHANISTIC GROUPING METHODS | |
| DNA Binding by OECD | This grouping method contains categories or chemical methods of DNA binding. This method is particularly relevant for genotoxicity endpoints. |
| DNA Binding by OASIS V.1.1 | This grouping method contains categories or chemical methods of DNA binding. |

Getting Started

| | |
|---|--|
| <p>DPRA Cysteine peptide depletion</p> | <p>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.</p> |
| <p>DPRA Lysine peptide depletion</p> | <p>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.</p> |
| <p>Estrogen Receptor Binding</p> | <p>This grouping method contains simple categories for estrogen receptor (ER) binding. This method is relevant for reproductive toxicity endpoints in fish and mammals.</p> |
| <p>Protein binding by OASIS v.1.1</p> | <p>This grouping method includes categories or chemical mechanisms of protein binding</p> |
| <p>Superfragment</p> | <p>This grouping method identifies the largest electronically connected atom group. It is relevant for identifying chemicals with differing solvation behavior and can therefore be used to identify outliers for bioavailability.</p> |
| <p>Biodeg probability from 1 to 7(Biowin)</p> | <p>These are models extracted from EPISUIT program related to biodegradability of chemicals</p> |
| <p>Hydrolysis half life (Ka) at pH7 and pH8</p> | <p>These are the profilers based on HYDROWIN, which 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 and pH 8.</p> |
| <p>Hydrolysis half life (Kb) at pH7 and pH8</p> | <p>These are the profilers based on HYDROWIN, which 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 and pH 8.</p> |

Getting Started

| | |
|---|--|
| <p>Ionization at pH 1, 4, 7.4, 9</p> | <p>These profilers calculates the fractional concentration of the ionized species of a molecule (Ionization, %) at different pH=1, 4, 7.4 and 9.</p> |
| <p>ENDPOINT SPECIFIC</p> | |
| <p>Acute Aquatic Toxicity MOA by OASIS</p> | <p>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.</p> |
| <p>Aquatic Toxicity Classification by ECOSAR</p> | <p>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.</p> |
| <p>Acute Aquatic toxicity Classification by Verhaar</p> | <p>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.</p> |
| <p>Bioaccumulation – metabolism alerts</p> | <p>These two grouping methods are based on the structural alerts used by the BCFBAF model for bioaccumulation developed for the US EPA.</p> |
| <p>Bioaccumulation – metabolism half life</p> | |
| <p>Biodegradation Fragments (BIOWIN MITI)</p> | <p>This grouping method is based on the structural alerts for ready biodegradability.</p> |

Getting Started

| | |
|---|--|
| <p>Carcinogenicity (genotox and nongenotox) alerts by ISS</p> | <p>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.</p> |
| <p>DNA alerts for AMES, MN and CA by OASIS v.1.1</p> | <p>This grouping method contains categories or chemical methods of DNA binding. This method is particularly relevant for genotoxicity endpoints.</p> |
| <p>Eye irritation/corrosion exclusion rules by BfR</p> | <p>This grouping method is based on physical-chemical exclusion criteria relevant for eye irritation and corrosion.</p> |
| <p>Eye irritation/corrosion inclusion rules by BfR</p> | <p>This grouping method is based on structural alerts for eye irritation or corrosion.</p> |
| <p>in vitro mutagenicity (Ames test) alerts by ISS</p> | <p>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.</p> |
| <p>in vivo mutagenicity (Micronucleus) alerts by ISS</p> | <p>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.</p> |

Getting Started

| | |
|---|---|
| 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. |
| Protein binding alerts for skin sensitization by OASIS v1.1 | 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 |
| Oncologic Primary Classification | This grouping method is based on structural criteria of chemical classes of potential carcinogens. |
| Skin irritation/corrosion exclusion rules by BfR | This grouping method is based on physical-chemical exclusion criteria relevant for skin irritation and corrosion. |
| Skin irritation/corrosion inclusion rules by BfR | This grouping method is based on structural alerts for skin irritation or corrosion. |
| EMPIRIC | |
| Chemical elements | This profiler contains all chemical elements from Periodic table organized in eighteen groups |
| 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 (nested) | OFG (nested), which include the same functional groups as in OFG. 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. |

Getting Started

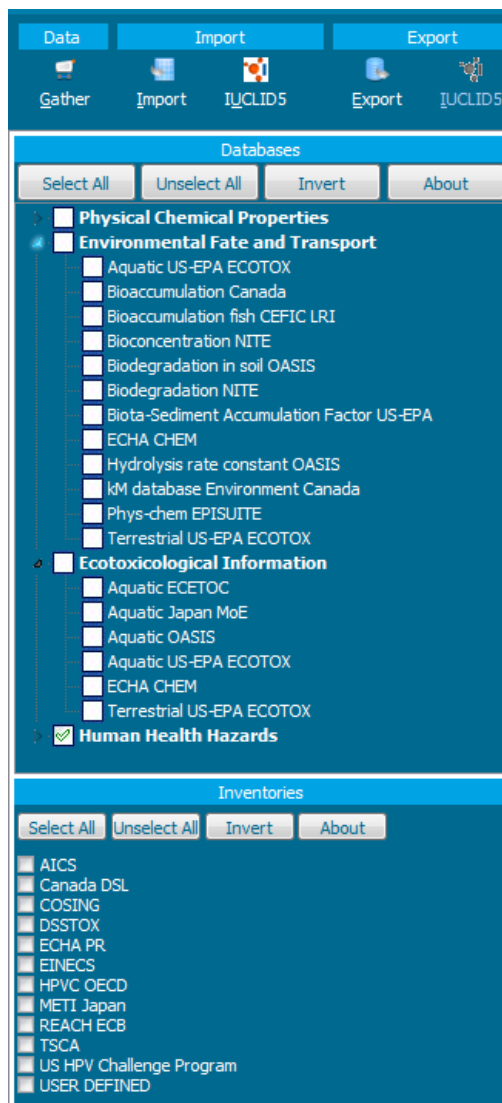
| | |
|--------------------|---|
| Tautomers unstable | The Unstable tautomer profiler is developed on the base of the available data and theoretical calculations for tautomer forms in water and gas phase. The unstable tautomeric forms are presented as individual categories. |
|--------------------|---|



When searching for analogues of a target chemical, the outcome of the profiling (see Chapter 2) determines the most appropriate way. The following recommendations can be made:

- 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.

Getting Started



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.



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

- 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.

Getting Started

- If the aim of the user also is to find analogues without experimental results, but for which the data gaps could be filled in the same way as for the target chemical, then also the appropriate inventory should be selected.
-  The inventories contain between 5 000 and 100 000 substances. Including one of them in the query will induce longer calculation times.

4.2 Example

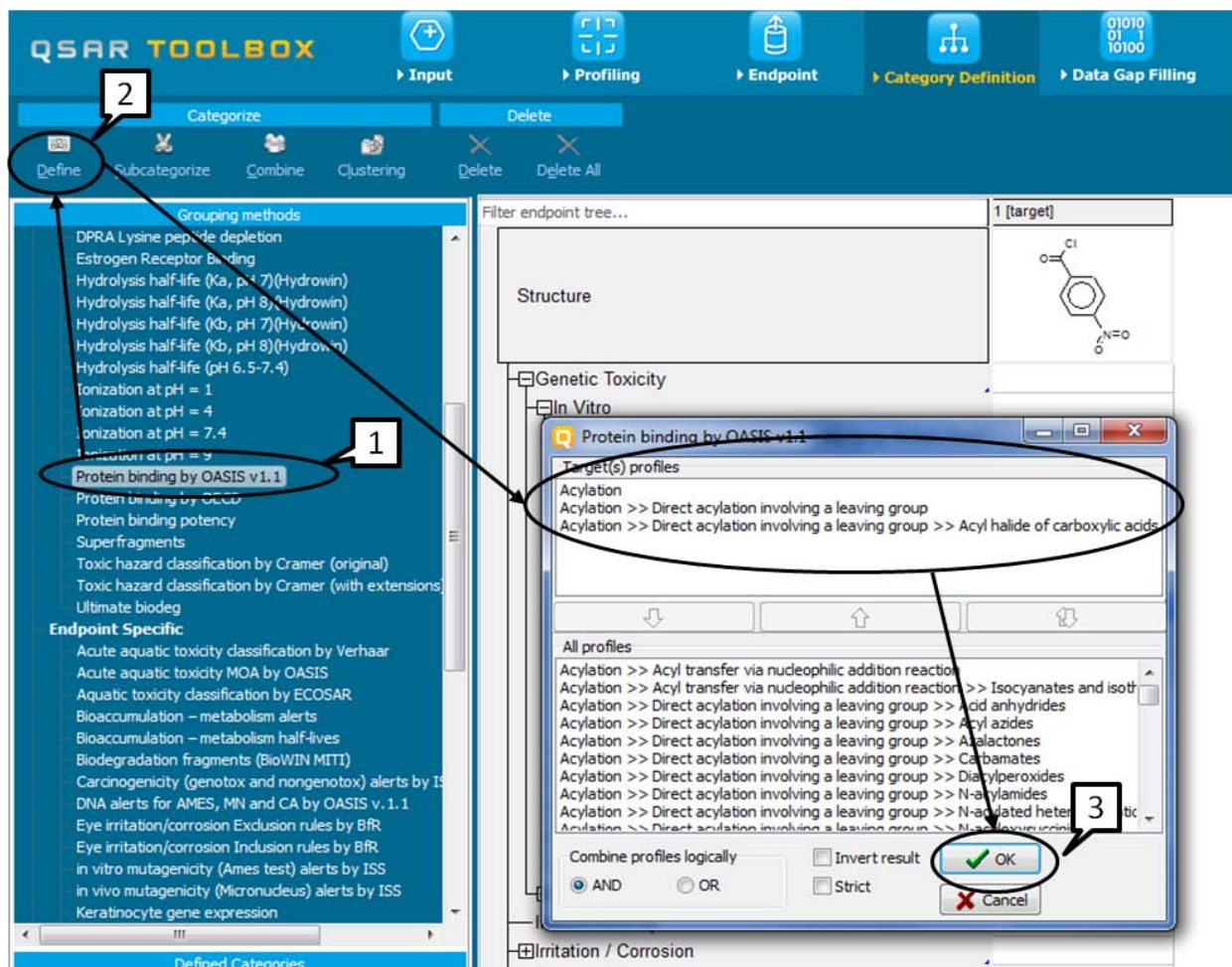
Continuing the example for 4-nitrobenzoyl chloride (CAS No 122-04-3), the aim is to find analogues with available experimental results on sensitisation that can be used for readacross to fill the data gap.

As indicated above, the profiling result determines the most appropriate way to search for analogues. The most meaningful profiler for skin sensitisation is "protein binding". The profiling results (see Chapter 2) have revealed that the target chemical can bind to protein by a specific mechanism (nucleophilic substitution on acyl halide of carboxylic acids). The most appropriate way forward is therefore to search for chemicals that can bind to protein by the same mechanism and for which experimental results on sensitisation are available. As seen in section 3.2, the appropriate databases "Skin sensitisation ECETOC" and "Skin sensitisation" are selected.

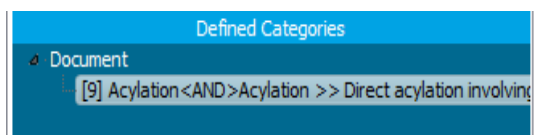
The user can then select the appropriate grouping method "Protein binding by OASIS"

(1) and press the **Define** button (2) as outlined below. The user is asked to confirm the choice of categories to be queried by pressing **OK** (3) in the pop-up window.

Getting Started



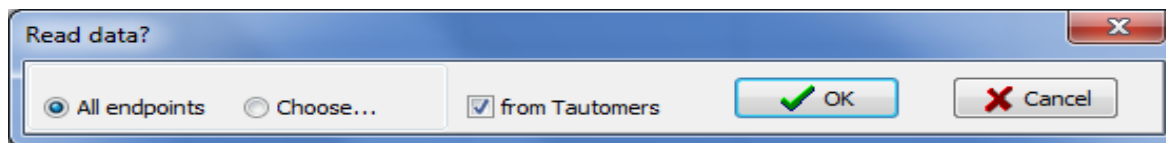
The Toolbox will now retrieve those chemicals that have the same protein binding mechanism than the target compound and for which data is available in the previously selected databases. In this example, 9 structures (including the target chemical) are retrieved.



The name of the category will appear in the “Defined Categories” window, also indicating the number of substances belonging to the category.

Getting Started

The Toolbox will automatically request the user to select the endpoint that should be retrieved. The user can either select the specific endpoint (skin sensitisation in this case) or by default choose to retrieve data on all endpoints (see below).



In this example, as only databases are selected that contain information on only one endpoint, both options give the same results.

The experimental results for the analogues are inserted into the matrix. To receive background information for a result, the user can double-click on the corresponding cell in the matrix.

| # | Endpoint | Value | Original value | Organ | Reference source | Institution and country | Year |
|---|----------|--|--|-------|---|-------------------------|------|
| 1 | EC3 | Positive (Skin sensitisation I (ECETOC)) | Strongly positive (Skin sensitisation I (Oasis)) | Skin | Regulatory Toxicology and Pharmacology, 48: 225-239 | LMC,BUL | 2007 |

Getting Started

The user can then proceed to the module “Filling data gap”.

Getting Started

5 Data Gap Filling



5.1 Aim of the module “Filling data gap”

The aim of the module “Filling Data Gap” is to give access to three different data gap filling tools:

- Read-across
- Trend analysis
- (Q)SAR models

Read-across and trend analysis use the available results 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. Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:

- Read-across is the appropriate data-gap filling method for “qualitative” endpoints like skin sensitisation 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.

Getting Started

- “(Q)SAR models” can be used to fill a data gap if no adequate analogues are found for a target chemical.

5.2 Example

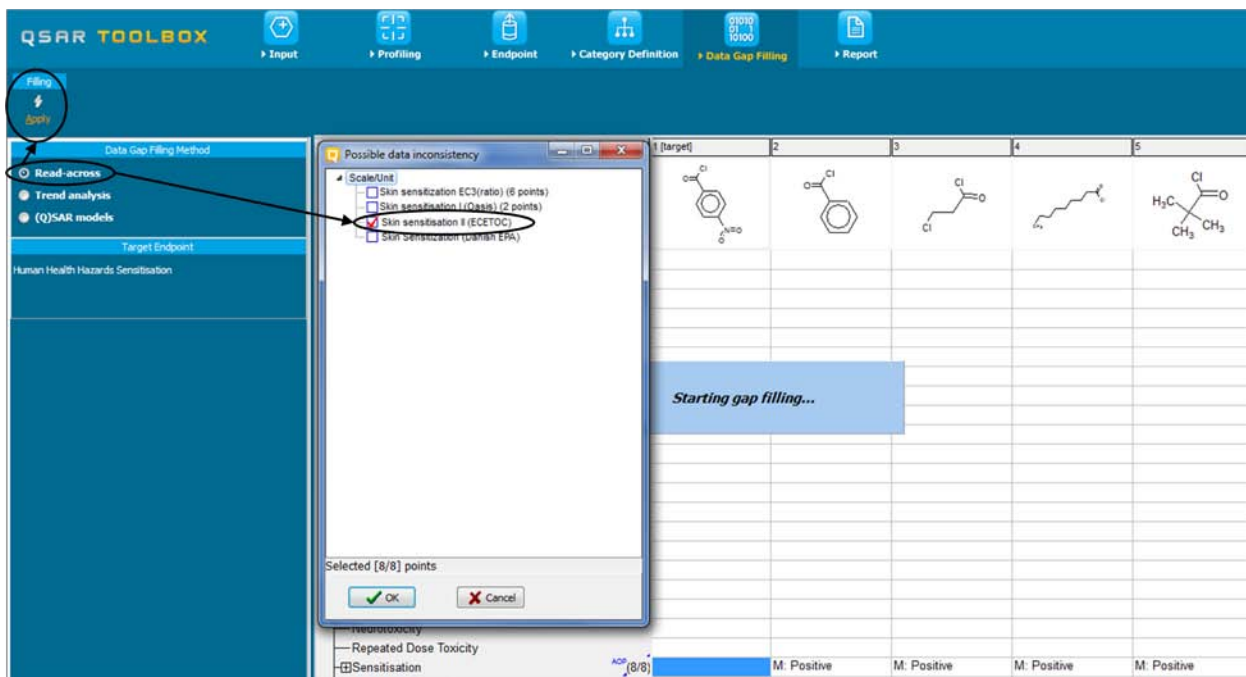
If the identification of analogues was performed according to a specific mechanism or mode of action, the target and source chemicals will already have the same relevant mechanisms and modes of action. The workflow illustration for the example chemical 4- nitrobenzoyl chloride (CAS No 122-04-3) is continued below.

As shown above, the Toolbox has identified seven additional chemicals which have the same protein binding mechanism as the target chemical and for which experimental results are available. Sensitisation is a “qualitative” endpoint, therefore the data gap can be filled by read-across.

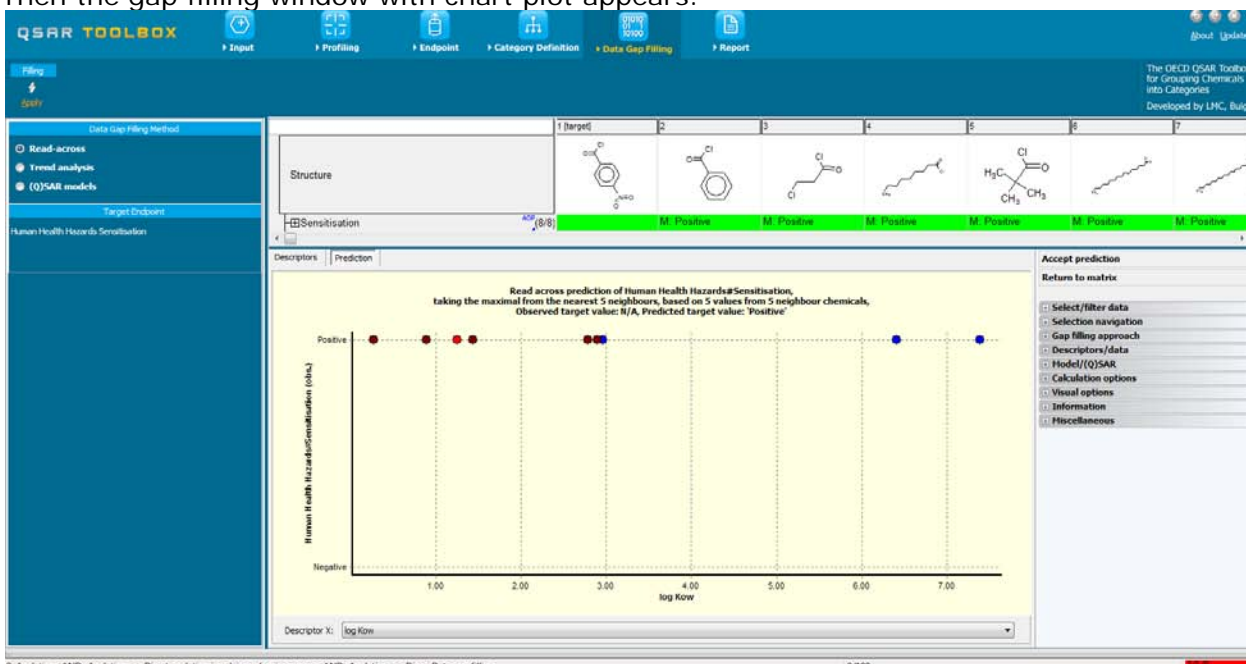
After highlighting the cell in the matrix corresponding to the data gap to be filled, the user can select the data gap filling methods (1) and then click on **Apply** (2).

A “Possible data inconsistency” window appears indicating all available scales. Detailed information for each of scale is available in **Options**, subsection **Units**. All possible conversions of scales are performed. Being less informative scale, **Skin sensitisation ECETOC** is selected as default scale. The user is advised to perform read-across using **Skin sensitisation ECETOC** scale. Confirm the scale **Skin sensitisation ECETOC** by clicking **OK** button (3).

Getting Started



Then the gap filling window with chart plot appears:



Getting Started



Select the cell referring to the endpoint of interest and your target substance in the data matrix before applying one of the data gap filling techniques.

The resulting graph plots the existing experimental results of all analogues (Y axis) according to a descriptor (X axis). The default descriptor is log Kow. The dark red dots on the graph represent the experimental results available for the analogues and which are used for the read-across. The blue dots on the graph represent the experimental results available for the analogues and which are not used for the read-across, because they are furthest away from the target chemical, based on the X-axis descriptor. The red dot represents the estimated result for the target chemical based on the read-across from the analogues.

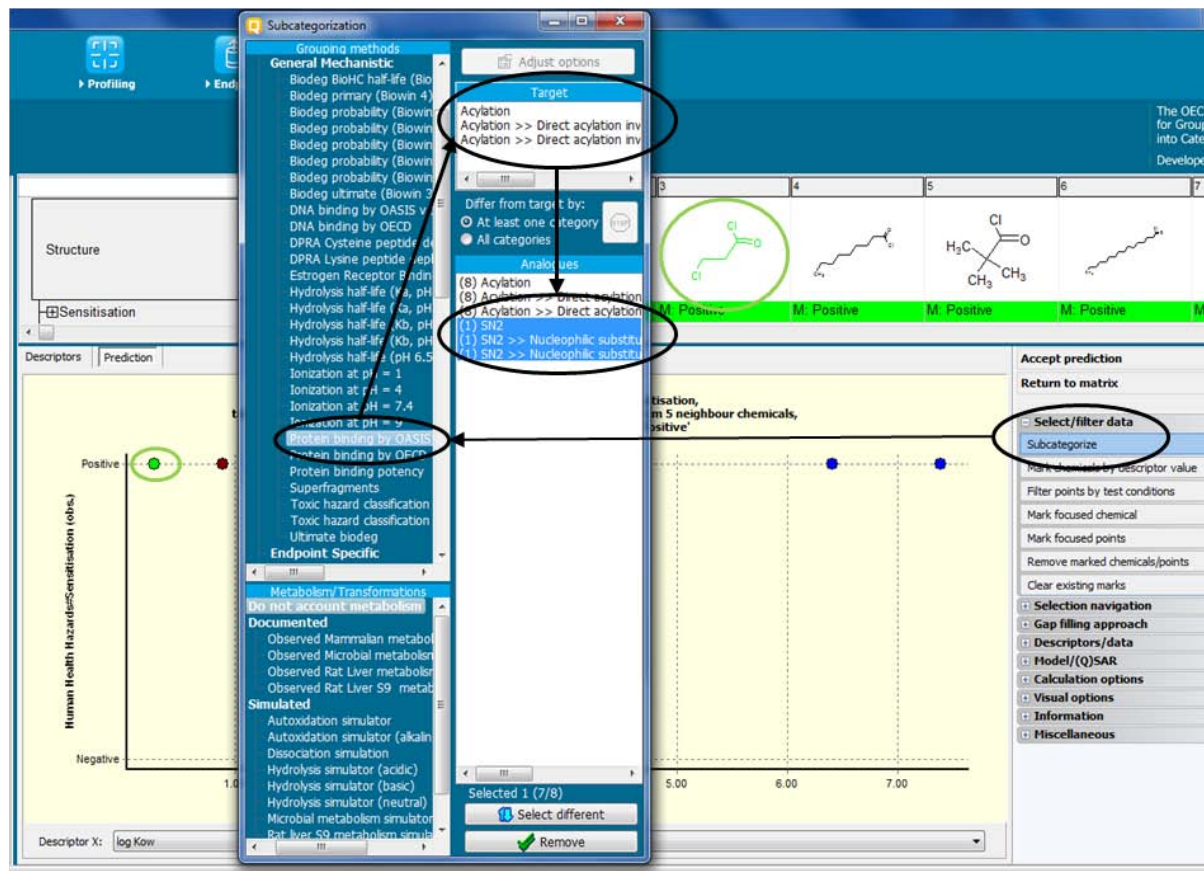
In this particular example, all results of the analogues are positive. The same sensitizing potential is therefore also predicted for the target chemical. By default, the Toolbox averages the result of the 5 "nearest" analogues (as defined by the X-axis descriptor) to estimate the result for the target chemical.

The user can then verify the mechanistic robustness of the analogue approach. As indicated above, if the identification of analogues was performed according to a specific mechanism or mode of action, 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.

This can be verified by opening **Select / Filter Data** menu and re-profiling the list of identified analogues by clicking on **Subcategorize** (1) and choosing **Protein Binding by OASIS v.1.1** (2). The properties of the analogues are then compared with the properties of the target chemical (3). As seen from the picture below, there is one analogue with different

Getting Started

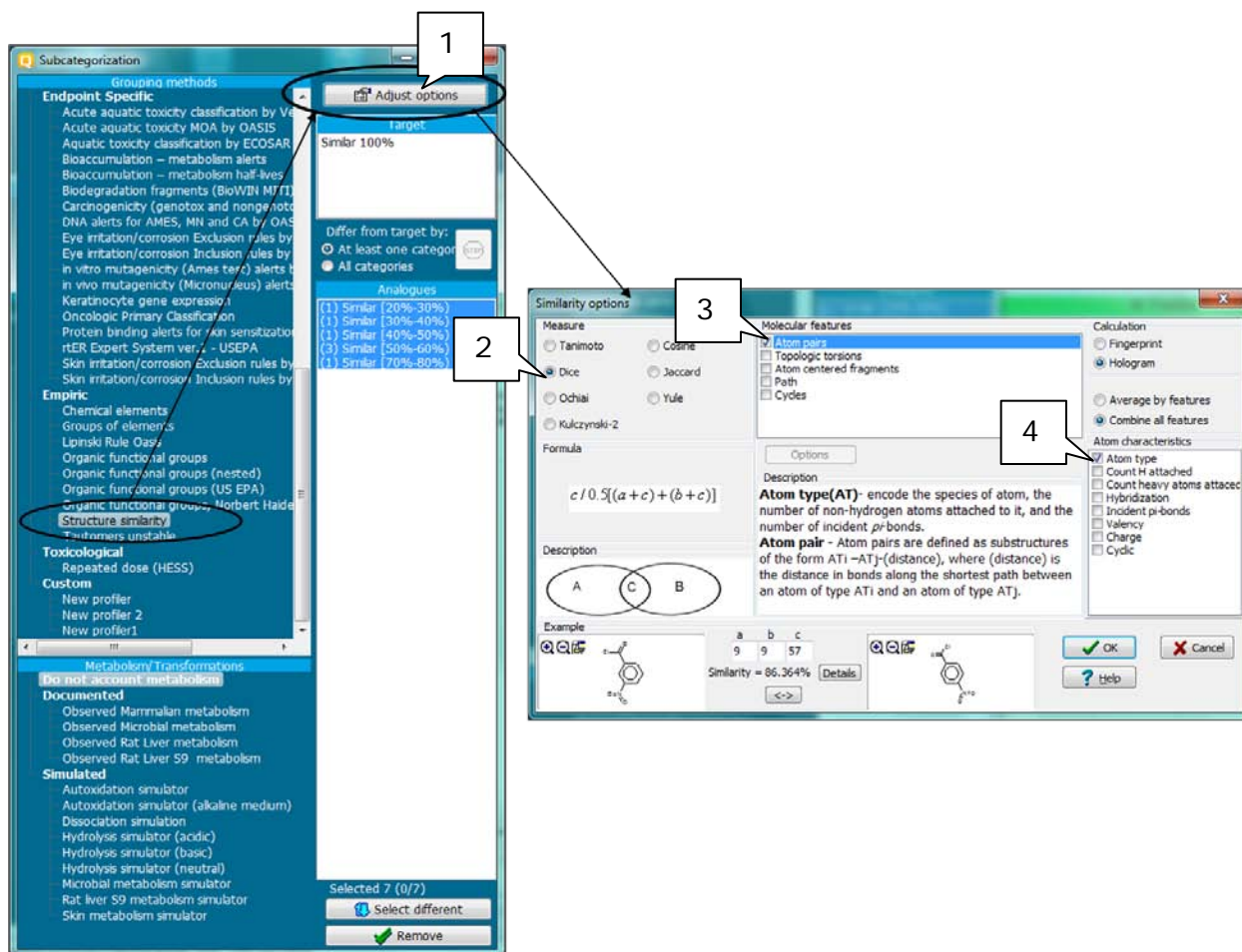
than the target protein binding mechanism (4). The user could eliminate dissimilar chemical by click **Remove** button (5).



Getting Started

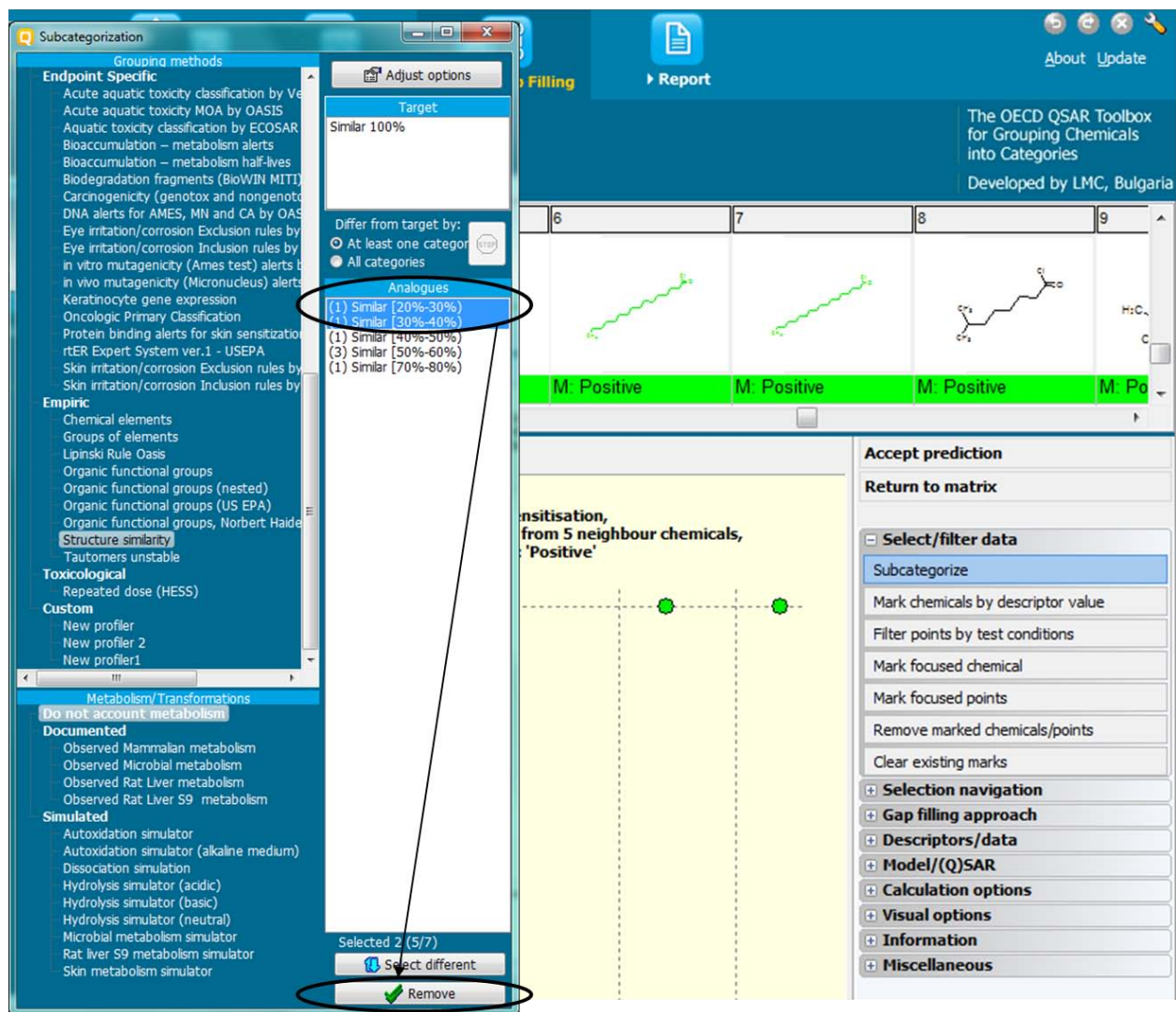
Furthermore, if relevant, the user can then identify those chemicals within the list of analogues that have the lowest structural similarity to the target chemical, so that they can be removed from the list of analogues if relevant.

The user can **Adjust** (1) the similarity options to **Dice** method (2), select **Atom pairs** (3) from the specific molecular features and specify **Atom type** (4) from the atom characteristics.



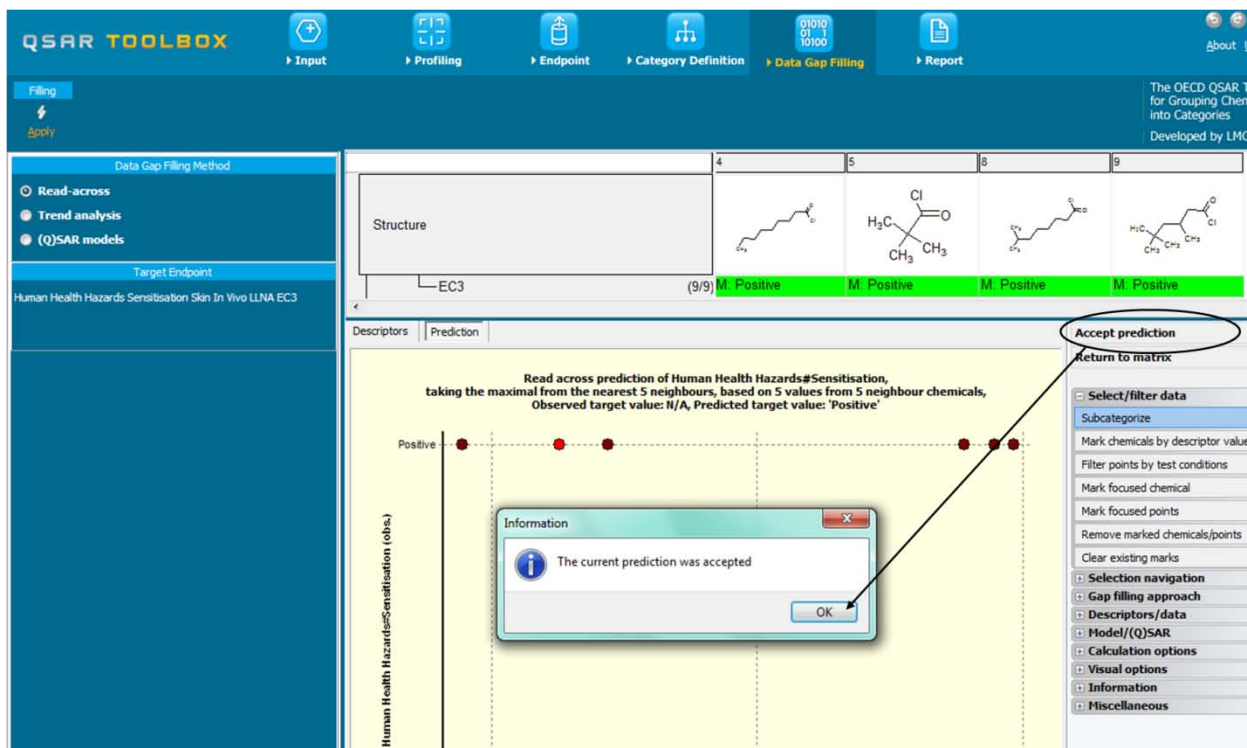
In this specific example, there is only one chemical that has a very high structural similarity with the target compound. The user can decide whether to keep only this analogue for the read-across or whether to keep also the other analogues.

Getting Started



Once the user is satisfied that the analogue approach is sufficiently robust, both from a point of view of mechanism or mode of action and structural similarity, the prediction can be accepted by clicking on **Accept prediction** and clicking **OK**, whereby the result is inserted into the data matrix.

Getting Started



! The origin of the result value presented in the matrix is indicated upfront with a capital letter: “M” stands for “measured”, “R” for “read-across”, “T” for “trend analysis” and “Q” for QSAR model.

By clicking on **Return to Matrix**, the user can close the read-across and proceed with the workflow.

! The user can export predictions to IUCLID 5 by moving the mouse in the column of the target substance and clicking the right mouse button. A wizard will guide the user through the different steps. These steps include the selection of the predicted value to be exported, editing report information such as information on the substance, category, prediction and adequacy, selecting an OECD harmonized template, completion of endpoint study record information such as reliability, interpretation of

Getting Started

interpretation of results, etc., and finally choosing the export format which can be via xml (*.i5z file) or directly via Webservice.

Getting Started

6 Report

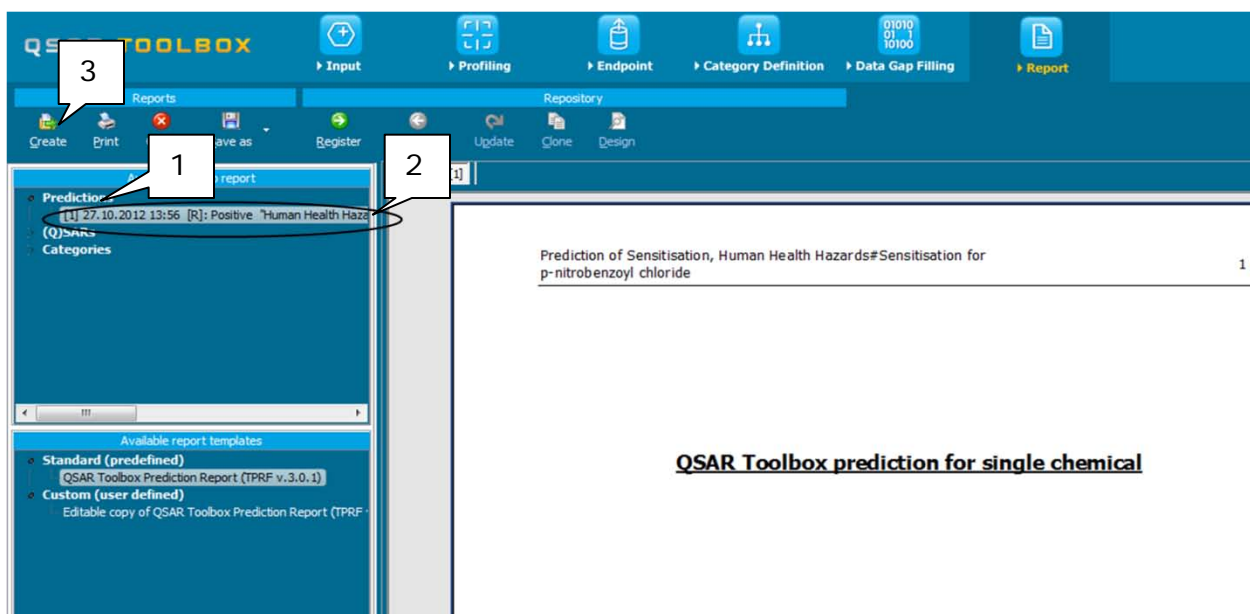


6.1 Aim of the module "Report"

With the module "Report", the user can generate reports on any of the predictions performed with the Toolbox. The Toolbox contains predefined report templates as well as a template editor with which users can define their own user defined templates.

6.2 Example

To generate a report for the prediction for skin sensitisation for 4-nitrobenzoyl chloride, the user has open the **Predictions** in the window **Available data to report** by double click (1) select the specific prediction (2), and then click the button **Create** (3).



Getting Started**OECD**

2, rue André Pascal
75775 Paris Cedex 16
France

Tel.: +33 1 45 24 82 00

Fax: +33 1 45 24 85 00

ehscont@oecd.org

<http://echa.europa.eu>