

OECD (Q)SAR Toolbox v.4.4.1

Example for predicting Skin Sensitization of
a mixture with known components

Outlook

- **Background**
- Keywords
- Objectives
- The exercise
- Workflow
- Save the prediction

Background

This is a step-by-step presentation designed to take the user of the Toolbox through the workflow for prediction of skin sensitization of a mixture.

Outlook

- Background
- **Keywords**
- Objectives
- The exercise
- Workflow
- Save the prediction

Keywords

TARGET CHEMICAL - chemical of interest, in this case it is a mixture with defined components

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

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

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

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

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

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

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

Outlook

- Background
- Keywords
- **Objectives**
- The exercise
- Workflow
- Save the prediction

Objectives

- **This presentation reviews a number of functionalities of the Toolbox:**
 - 2D editor for defining Mixture components
 - Filling data gaps by Independent mode approach

Outlook

- Background
- Keywords
- Objectives
- **The exercise**
- Workflow
- Save the prediction

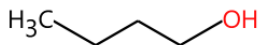
Exercise

➤ In this exercise we will:

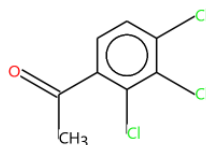
- predict skin sensitization of target substance, which represent a mixture with defined constituents
- Investigate the mode of action for each component of the mixture,
- Gather available experimental data for target chemical,
- Investigate skin sensitization of non-tested component,
- Apply read across for non-tested component, and
- Predict skin sensitization potential of mixture based on experimental data of tested compounds and predicted data of non-tested one.

➤ The target substance will consists of three constituents:

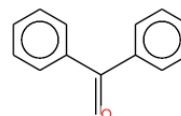
Constituent 1



Constituent 2



Constituent 3



Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
- Save the prediction

Workflow

- **The Toolbox has six modules which are used in a sequential workflow:**
 - Input
 - Profiling
 - Data
 - Category Definition
 - Data Gap Filling
 - Report

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - **Input**

Chemical Input

Overview

- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input

Ways of Entering a mixture

User alternatives for defining mixtures with known compositions:

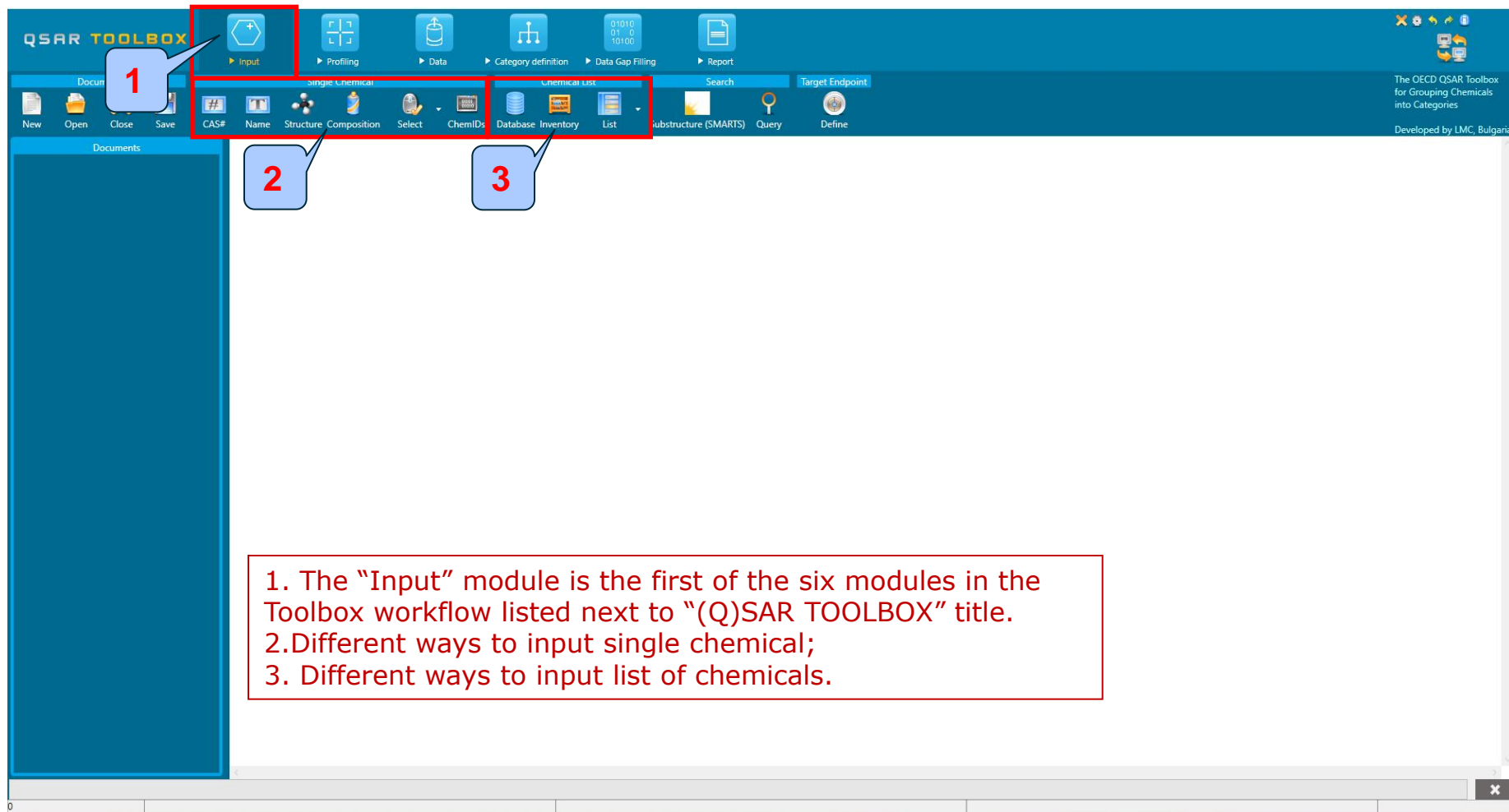
A. Single target substance

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing mixture constituents and defining their quantities
- Select from User List/Inventory/Databases

B. Group of chemicals

- User List/Inventory
- Specialized Databases

Chemical Input Input Screen



1. The "Input" module is the first of the six modules in the Toolbox workflow listed next to "(Q)SAR TOOLBOX" title.
2. Different ways to input single chemical;
3. Different ways to input list of chemicals.

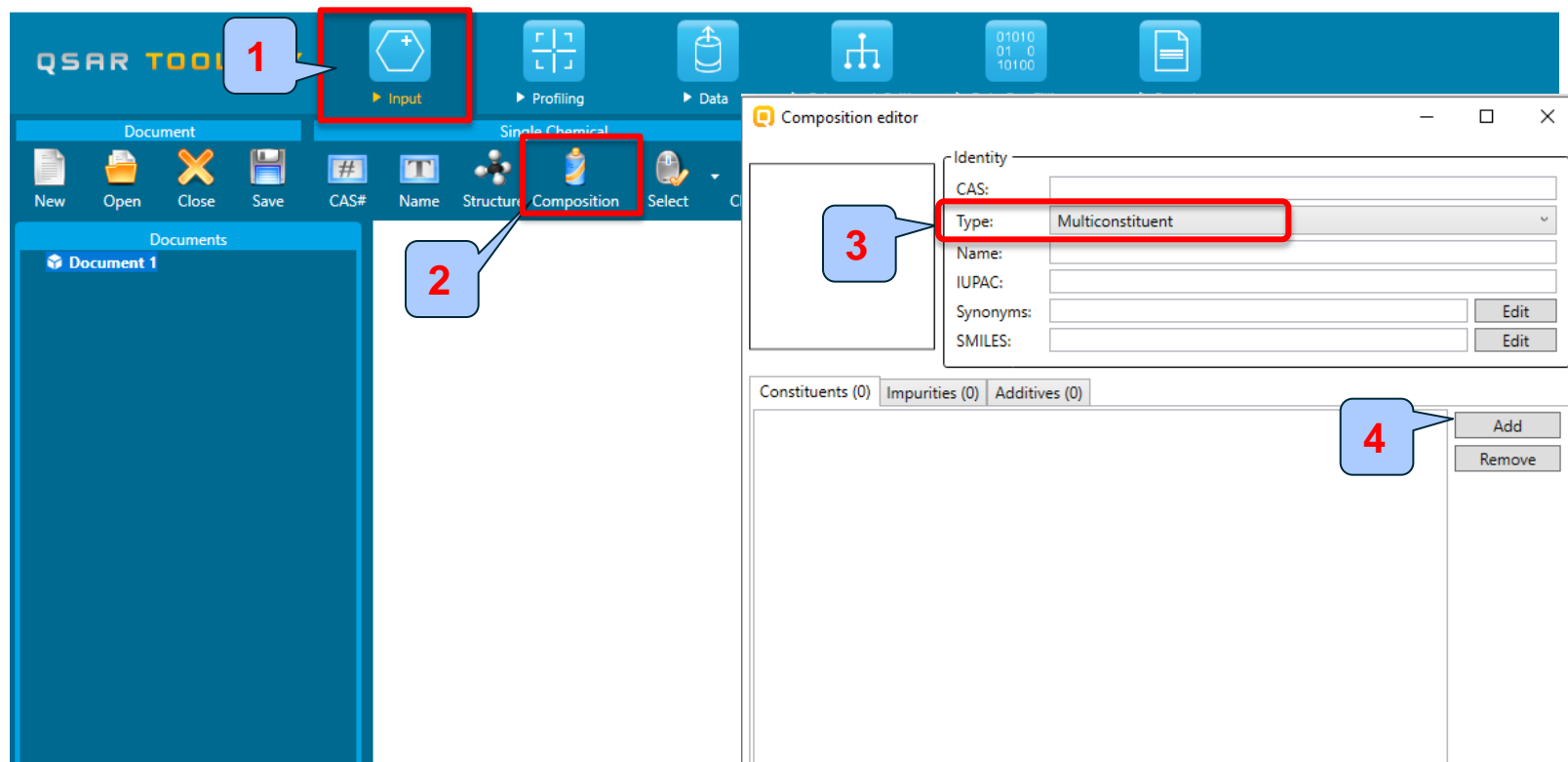
Input

Input a mixture

- In the current example our target substance will be a mixture.
- We will draw its components within the “Composition” tool.

Chemical input

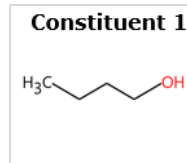
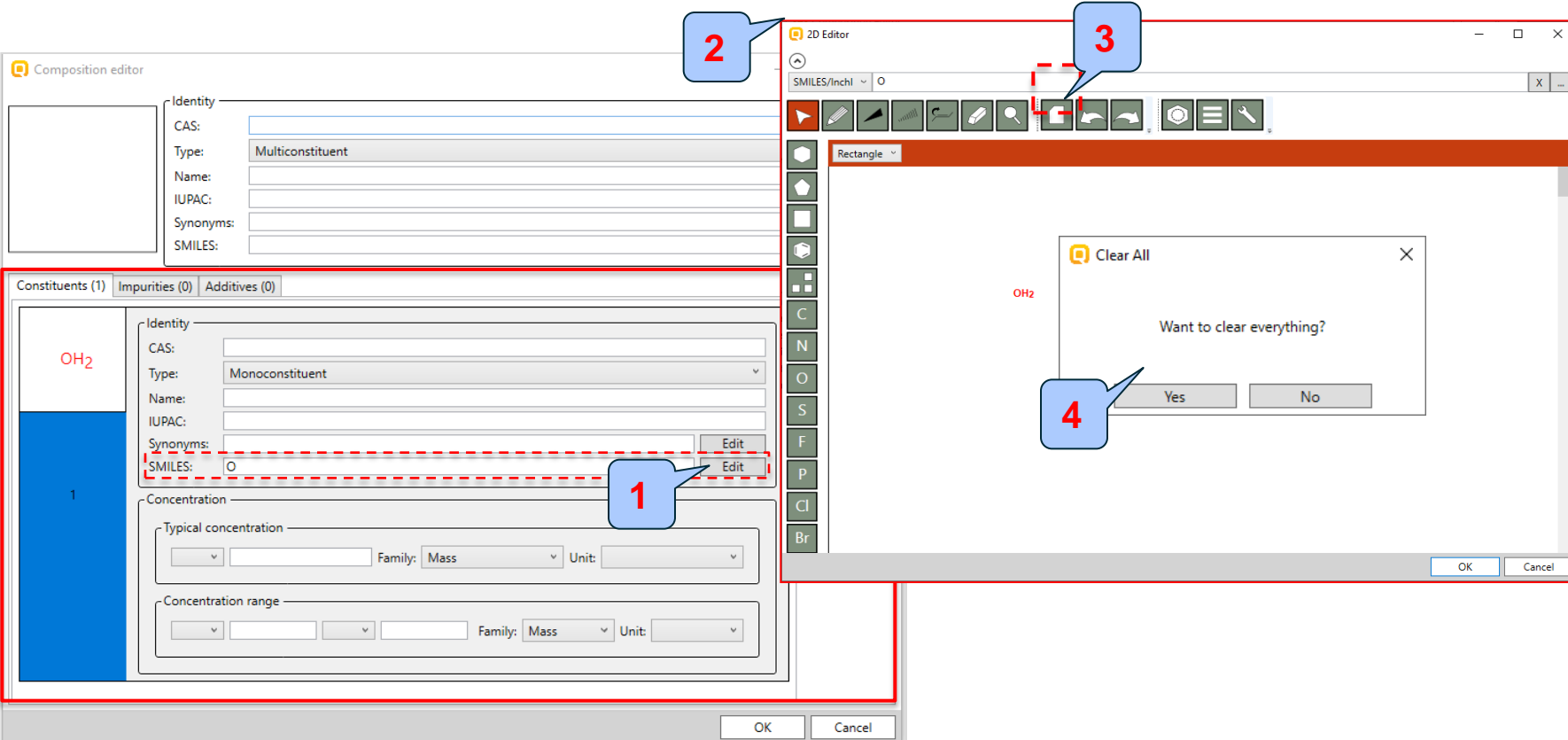
Input target substance by drawing



1. Click on **Input** module;
2. Click on **Composition**;
3. From composition editor select type: **Multiconstituent**;
4. Click **Add** in order to add constituent. Our target substance consists of three constituents, so click three times on the Add button.

Chemical input

Drawing Constituent 1 of the target mixture

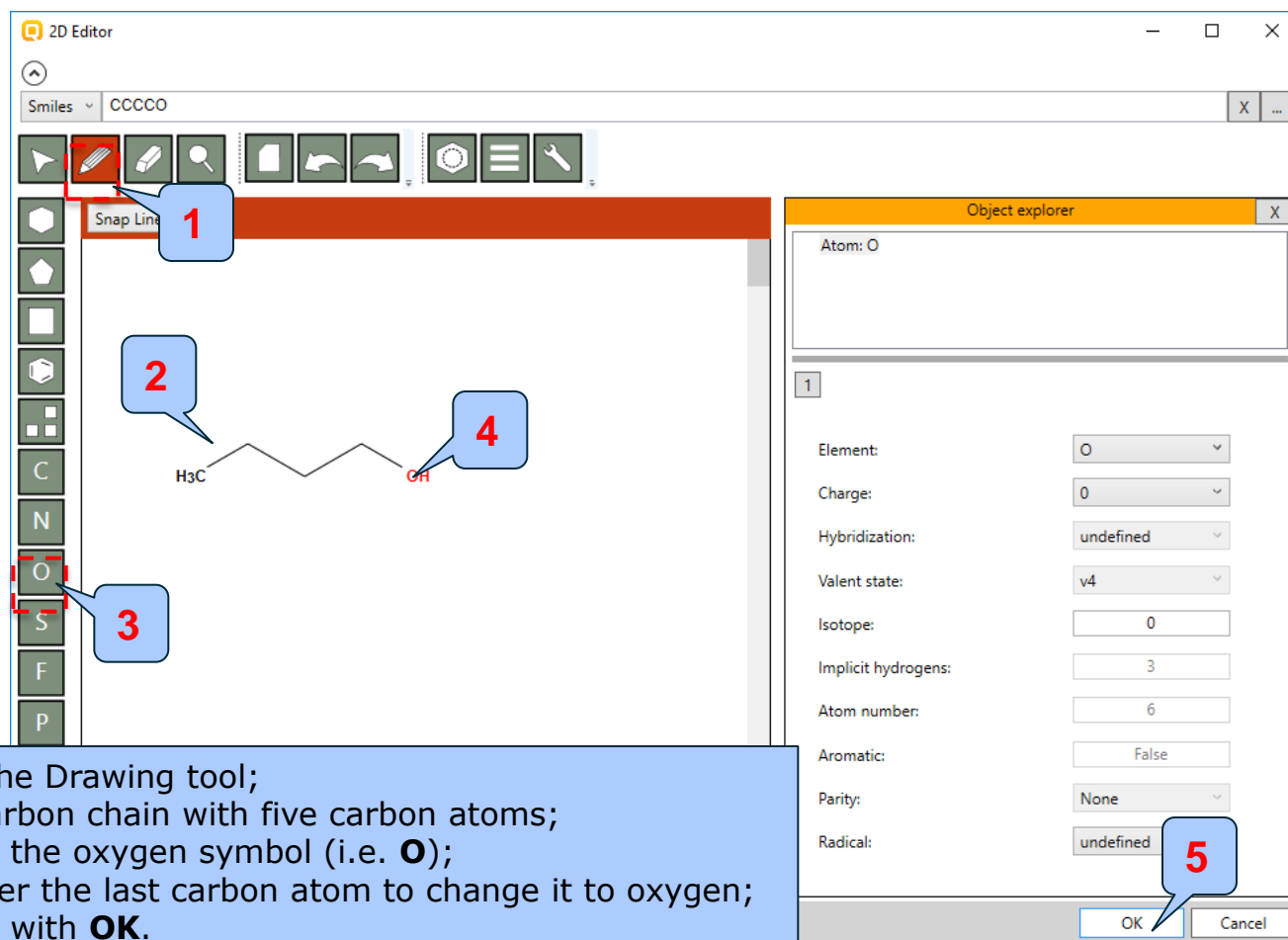
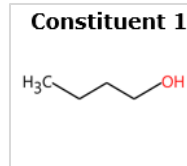



The screenshot shows the 'Composition editor' window with the 'Constituents (1)' tab selected. The first constituent, 'OH2', is highlighted in blue. The 'Identity' section for this constituent shows 'CAS:', 'Type: Monoconstituent', 'Name:', 'IUPAC:', 'Synonyms:', and 'SMILES: O'. The 'Edit' button next to the SMILES field is circled in red and labeled with a blue callout '1'. The '2D Editor' window is open, showing a 'Clean All' dialog box with the text 'Want to clear everything?' and 'Yes' and 'No' buttons. A blue callout '4' points to the 'Yes' button. The 'Clean All' dialog box is also labeled with a blue callout '3' pointing to the 'Clean' button in the toolbar. The '2D Editor' window is labeled with a blue callout '2' pointing to the toolbar. The 'Clean' button in the toolbar is labeled with a blue callout '3'.

1. Click **Edit** on the SMILES row to define the structure of the first constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything.
4. Confirm with **Yes**.

Chemical input

Drawing Constituent 1 of the target mixture



1. Select the Drawing tool;

2. Draw carbon chain with five carbon atoms;

3. Click on the oxygen symbol (i.e. **O**);

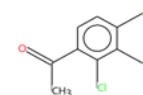
4. Click over the last carbon atom to change it to oxygen;

5. Confirm with **OK**.

Chemical input

Drawing Constituent 2 of the target mixture

Constituent 2

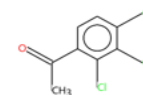


- Now we move down to the second constituent and repeat the same steps:
1. Click **Edit** on the SMILES row to define the structure of the second constituent;
 2. The 2D editor appears;
 3. Click the **Clean button** to clean everything.
 4. Confirm with **Yes**.

Chemical input

Drawing Constituent 2 of the target mixture

Constituent 2



1a Select the *Benzene scaffold* (1a) and paste it into the drawing pane (1b).

2a Select the *Drawing tool* (2a) and draw the connections to the benzene (2b). Second click over a bonds converts it to double bond.

3a Click on the oxygen symbol (i.e. **O**) (3a) and click over the carbon atom connected with double bond (3b).

4a Click on the chlorine symbol (i.e. **Cl**) (4a) and click over the carbon atoms that should be changed (3b).

5 Confirm with **OK** (5).

Smiles: C1=CC(=C(C(=C1C(C)=O)Cl)Cl)Cl

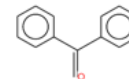
Isotope: 0
 Implicit hydrogens: 3
 Atom number: 6
 Aromatic: False
 Parity: None
 Radical: undefi

OK Cancel

Chemical input

Drawing Constituent 3 of the target mixture

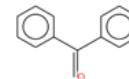
Constituent 3



The screenshot illustrates the steps to add a new constituent to the target mixture. In the 'Composition editor', the 'Constituents (3)' tab is active. The 'Edit' button for the third constituent (labeled '3') is highlighted with a red dashed box and labeled '1'. The '2D Editor' window is open, showing a clean canvas with a 'Rectangle' dropdown and a 'Clean' button highlighted with a red box and labeled '3'. A 'Clear All' dialog box is open, asking 'Want to clear everything?' with 'Yes' and 'No' buttons; the 'Yes' button is labeled '4'.

Now we move down to the last (third) constituent and repeat the same steps:

1. Click **Edit** on the SMILES row to define the structure of the second constituent;
2. The 2D editor appears;
3. Click the **Clean button** to clean everything.
4. Confirm with **Yes**.



Chemical input

Drawing Constituent 3 of the target mixture

2a

1a

3a

3b

2b

1b

4

SMILES/Inchi O=C(C=C1C(C1=CC=CC=C1)=O

Select the Benzene scaffold (1a) and paste it two times in the drawing pane (apply two left clicks) (1b).
Select the Drawing tool (2a) and draw connection between both rings (2b). Double click over a bond converts it to double bond.
Click on the oxygen symbol (i.e. O) (3a) and click over the carbon atom connected with double bond (3b).
Confirm with **OK** (4).

Chemical input

Target substance identity

The screenshot shows the QSAR Toolbox software interface. The 'Target substance identity' window is open, displaying the following information:

- Structure:** A chemical structure diagram of a multi-constituent mixture.
- Structure info:**
 - Additional Ids: No CAS number, Not applicable
 - CAS Number: Not applicable
 - CAS-SMILES relation: Not applicable
 - Chemical name(s): C3; A0; t0
 - Composition: C25H25Cl3O3
 - Molecular formula: Multi constituent
 - Predefined substance type: CCCCCC
 - SMILES: CCCCC.CC(=O)c1ccccc1c1ccccc1c1ccccc1c1ccccc1c1ccccc1
- Parameters:**
 - Physical Chemical Properties
 - Environmental Fate and Transport
 - Ecotoxicological Information
 - Human Health Hazards

The drawn mixture automatically appears on the data matrix. Note that no CAS number or name is displayed for this chemical. This means the target substance is not listed in the chemical inventories/databases implemented in the Toolbox

Chemical input

Mixture decomposition

- In the current example we will predict the skin sensitization of a mixture based on its constituents.
- A specific option “Decomposition” allows all constituents of a mixture as well as available additives/impurities to be shown in the data matrix.
- Once the constituents are on the data matrix, the user can handle them as individual substances and further, to use them for predicting the whole mixture.

Chemical input

Input mixture

1. Right click over the **Substance**,

2. Select **Multiplication**,

3. Click on **Decomposition**.

Chemical Input

Target chemical identity

- The already drawn target structures automatically appear on the data matrix.
- Note that no CAS number or name is associated with this chemical.
- This means the target chemical is not listed in the chemical inventories/databases available in Toolbox (see next slide).

Chemical Input

Target chemical identity

The screenshot shows the QSAR Toolbox software interface. The 'Target Endpoint' tab is active. In the left sidebar, the 'Documents' tree shows a 'Composition list' with three constituents: [C: 4-Md: 0-P: 0] Constituent #1, [C: 1-Md: 0-P: 0] Constituent #2, and [C: 1-Md: 0-P: 0] Constituent #3. The main area displays a table with chemical structures and their properties. A red box highlights the 'Composition list' and the table data.

| Target chemical (target) | Constituent #1 | Constituent #2 | Constituent #3 |
|---|------------------|------------------|------------------|
| | | | |
| No CAS number | No CAS number | No CAS number | No CAS number |
| Not applicable | Not applicable | Not applicable | Not applicable |
| C3: A0: I0 | C8H5ClO | C4H10O | C13H10O |
| C25H25ClO3 | Mono constituent | Mono constituent | Mono constituent |
| Multi constituent | | | |
| CCCCO.CC(=O)c1ccc(Cl)c(Cl)c1Cl.O=C(c1ccc(Cl)c(Cl)c1)C1=CC=CC=C1 | | | |

The three constituents appear on the data matrix along with the target mixture as well as in a tree-like form.

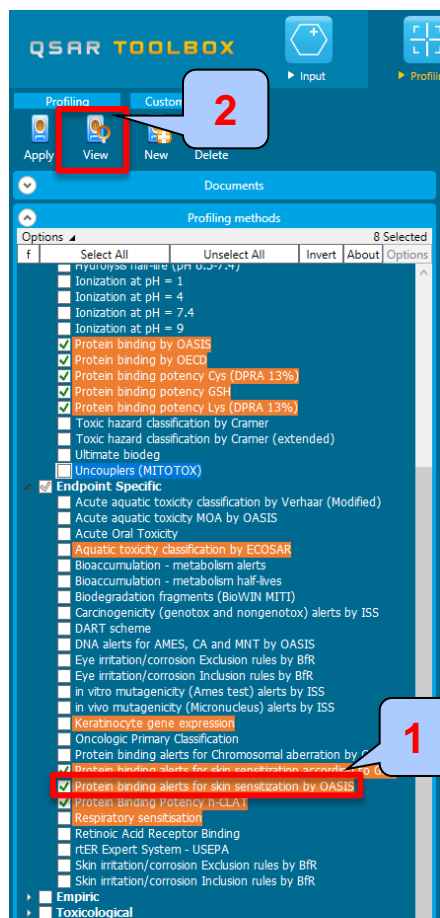
Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - **Profiling**

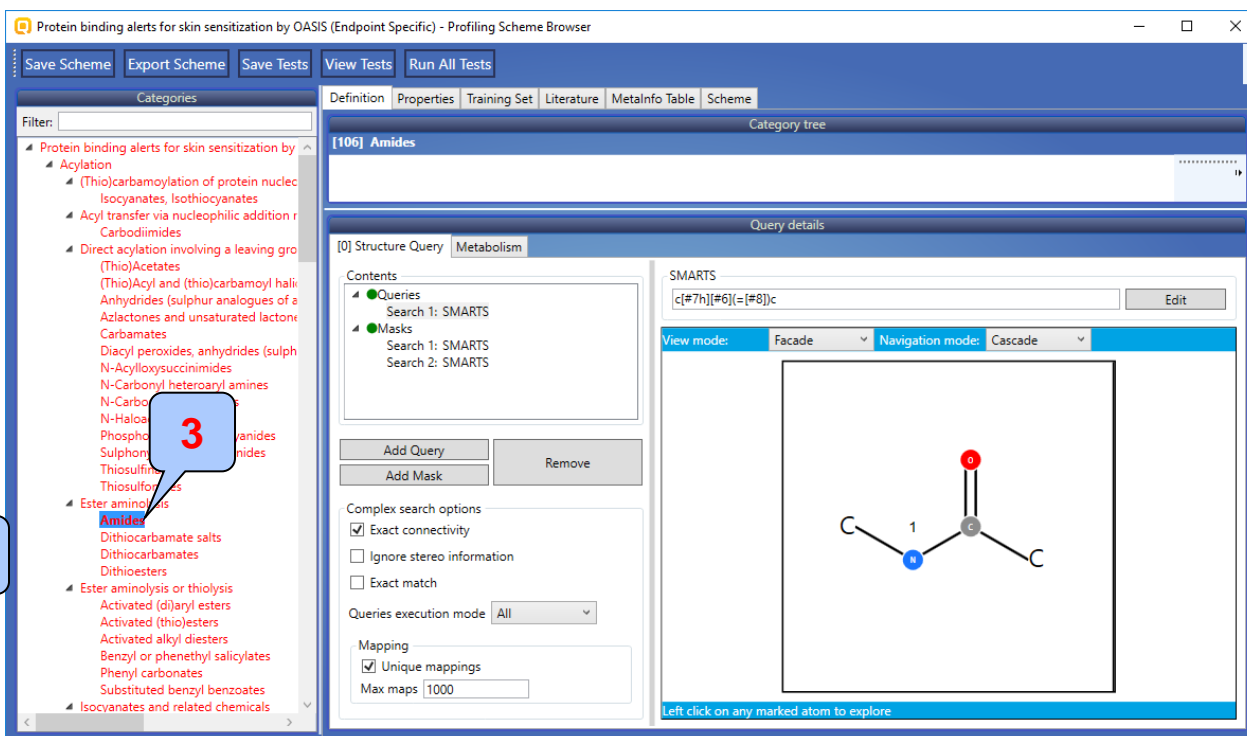
Profiling Overview

- “Profiling” refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.
- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding alerts for SS by OASIS and clicking on “About” or “View” (see next screen shots).

Profiling Side-Bar to Profiling



The **View** button provides more details on the coded knowledge in the profiler.



1. Selected profiler related to the investigated endpoint: **Protein binding alerts for SS by OASIS**;
2. Click on the **"View"** button;
3. Click for example on category **Amides** to see the structural boundaries used to code the knowledge.

Profiling

Side-Bar to Profiling

3 Apply

2

1

Filter endpoint tree...

| Parent chemical [target] | Constituent #1 | Constituent #2 | Constituent #3 |
|---|----------------|----------------|----------------|
| Structure | | | |
| Environmental Fate and Transport | | | |
| Ecotoxicological Information | | | |
| Human Health Hazards | | | |
| Acute Toxicity | | | |
| ADME | | | |
| Bioaccumulation | | | |
| Carcinogenicity | | | |
| Developmental Toxicity / Teratogenicity | | | |
| Genetic Toxicity | | | |
| Immunotoxicity | | | |
| Irritation / Corrosion | | | |
| Neurotoxicity | | | |
| Photoinduced toxicity | | | |
| Repeated Dose Toxicity | | | |
| Sensitisation | | | |
| ToxCast | | | |
| Toxicity to Reproduction | | | |
| Toxicokinetics, Metabolism and Distribution | | | |

Options 22 Selected

Profiling methods

- ☐ Acute aquatic toxicity HPA by OASIS
- ☐ Acute Oral Toxicity
- ☒ Aquatic toxicity classification by ECOSAR
- ☐ Bioaccumulation - metabolism alerts
- ☐ Bioaccumulation - metabolism half-lives
- ☐ Biodegradation fragments (BioWIN MITI)
- ☐ Carcinogenicity (genotox and nongenotox) alerts by DART scheme
- ☐ DNA alerts for AMES, CA and MNT by OASIS
- ☐ Eye irritation/corrosion Exclusion rules by BfR
- ☐ Eye irritation/corrosion Inclusion rules by BfR
- ☐ in vitro mutagenicity (Ames test) alerts by ISS
- ☐ in vitro mutagenicity (Microtubule) alerts by ISS
- ☒ Keratinocyte gene expression
- ☒ Oncogenic Primary Classification
- ☒ Protein binding alerts for chromosomal aberration by
- ☒ Protein binding alerts for skin sensitization according
- ☒ Protein binding alerts for skin sensitization by OASIS
- ☒ Protein Binding Potency h-CLAT
- ☒ Repeated Dose Toxicity
- ☐ Retinoic Acid Receptor Binding
- ☐ rTER Expert System - USEPA
- ☐ Skin irritation/corrosion Exclusion rules by BfR
- ☐ Skin irritation/corrosion Inclusion rules by BfR
- ☒ Empiric
 - ☒ Chemical elements
 - ☒ Groups of elements
 - ☒ Lipinski Rule Oasis
 - ☒ Organic functional groups
 - ☒ Organic functional groups (nested)
 - ☒ Organic functional groups (US EPA)
 - ☒ Organic functional groups, Norbert Haider (checked)
 - ☐ Structural Similarity
 - ☐ Tautomers unstable
- ☐ Toxicological
 - ☐ Repeated dose (HESS)

1

2

3

1. Position the cursor on the level of "Sensitisation";

2. Select the most plausible profilers related to the target endpoint (in our case the orange highlighted);

3. Click **Apply.**

Profiling

Profiling the target substance

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown box under the target substance.
- In this example the target mixture and its constituents are profiled by all profilers defined as plausible (highlighted in orange) for skin sensitization (only endpoint-specific are listed here):
 - Aquatic toxicity classification by ECOSAR;
 - Keratinocyte gene expression;
 - Protein binding alerts for skin sensitization according to GHS
 - Protein binding alerts for skin sensitization by OASIS
 - Protein binding potency h-CLAT
 - Respiratory sensitization

Profiling

Profiling the target substance

Visualizing the profiling results for the target mixture and its individual constituents.

| Filter endpoint tree... | Parent chemical [target] | Constituent #1 | Constituent #2 | Constituent #3 |
|---|---------------------------------|-----------------------------|-----------------------------|-----------------------------|
| Structure | <chem>CCCCCCCC</chem> | <chem>CCCCCCCC</chem> | <chem>CCCCCCCC</chem> | <chem>CCCCCCCC</chem> |
| Environmental Fate and Transport | | | | |
| Predefined | | | | |
| OECD HPV Chemical Categories | Not categorized | Not categorized | Not categorized | Not categorized |
| Substance type | Mixture | Discrete chemical | Discrete chemical | Discrete chemical |
| US-EPA New Chemical Categories | Neutral Organics | Neutral Organics | Neutral Organics | Neutral Organics |
| General Mechanistic | | | | |
| Protein binding by OASIS | Schiff base formation | No alert found | Schiff base formation | No alert found |
| Protein binding by OECD | No alert found | No alert found | No alert found | No alert found |
| Protein binding potency Cys (DPRA 13%) | DPRA less than 9% (DPRA...) | DPRA less than 9% (D... | Out of mechanistic do... | DPRA less than 9% (DP... |
| Protein binding potency GSH | Not possible to classify acc... | Not possible to classify... | Not possible to classify... | Not possible to classify... |
| Protein binding potency Lys (DPRA 13%) | DPRA less than 9% (DPRA... | DPRA less than 9% (D... | DPRA less than 9% (DP... | DPRA less than 9% (DP... |
| Endpoint Specific | | | | |
| Aquatic toxicity classification by ECOS... | Neutral Organics | Neutral Organics | Neutral Organics | Neutral Organics |
| Keratinocyte gene expression | Not possible to classify acc... | Not possible to classif... | Not possible to classify... | Not possible to classify... |
| Protein binding alerts for skin sensitiz... | No alert found | No alert found | No alert found | No alert found |
| Protein binding alerts for skin sensitiz... | Schiff base formation | No alert found | Schiff base formation | No alert found |
| Protein Binding Potency h-CLAT | No alert found | No alert found | No alert found | No alert found |
| Respiratory sensitisation | No alert found | No alert found | No alert found | No alert found |
| Empiric | | | | |
| Chemical elements | Group 14 - Carbon C | Group 14 - Carbon C | Group 14 - Carbon C | Group 14 - Carbon C |
| Groups of elements | Halogens | Non-Metals | Halogens | Non-Metals |
| Lipinski Rule Oasis | Bioavailable | Bioavailable | Bioavailable | Bioavailable |
| Organic functional groups | Alcohol | Alcohol | Aryl | Aryl |
| Organic functional groups (nested) | Alcohol | Alcohol | Aryl | Aryl |

The profiling results for all the constituents are consistent with one exception (Constituent #2). The constituent #2 reacts with proteins via Schiff-base formation according to general and endpoint-specific Protein binding alerts for SS profiler.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - **Data**

Data

- “Data” refers to the electronic process of retrieving the environmental fate, eco-toxicity and toxicity data that are residing in the Toolbox.
- Data gathering can be executed in a global fashion (i.e. collecting all data of all endpoints) or on a more narrowly defined basis (e.g. collecting data for a single or limited number of endpoints).
- In this example, we limit our data gathering to the common Skin endpoints from databases associated with Skin Sensitization endpoint. The relevant databases are become highlighted in green based on the selected target endpoint

Data

The screenshot shows the QSAR Toolbox Data module interface. The top toolbar contains buttons for 'Data' (1), 'Input', 'Data Gap Filling', and 'Report'. The left sidebar shows a 'Documents' panel with a 'Databases' list (4) containing various toxicity databases, some highlighted in green. The main window displays a 'Filter endpoint tree...' (2) with a tree structure where 'Sensitization' is selected. A 'Read data?' dialog box (5) is open, showing a list of endpoints with 'Sensitization' checked. The 'OK' button is visible at the bottom of the dialog.

1. Go to **Data** module;
2. Select node "**Sensitization**" from endpoint tree;
3. Select all green highlighted databases;
4. Click on **Gather**;
5. Select "**Sensitization**" in order to extract only this type of data from selected databases. Click **OK**.

Data

Process of collecting data

The screenshot shows the QSAR Toolbox software interface. The 'Data' tab is selected, and the 'Filter endpoint tree' dialog box is open. The dialog box displays a list of endpoints with a red box highlighting the 'Sensitisation' section. A red arrow points from a text box to the 'Sensitisation' section, indicating that 39 data points are found for two of the three mixture constituents.

39 data points are found for two of the three mixture constituents.

Recap

- We have entered the mixture with defined components.
- The profiling results showed no protein binding alerts for two of the mixture constituents (constituents # 1 and #3). The third constituent (constituent #2) has positive protein binding alerts and could elicit skin sensitization effect.
- Negative experimental data has been found for two of the mixture constituents (constituents # 2 and #3). No experimental data has been found for the third constituent (constituent #1).
- The constituent without experimental data and positive protein binding alert (constituent #1) will be used for further read across analysis. Then, all of the available data – experimental and predicted will be used for skin sensitization prediction of the mixture.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Data
 - **Read across prediction of constituent without data**
 - **Focus constituent without experimental data**

Read across prediction of constituent without data

Focus constituent

The screenshot displays the QSAR Toolbox software interface. On the left, a 'Documents' panel shows a list of databases with '8 Selected' items. The main workspace is divided into several panes. The 'Parent chemical [target]' pane shows a chemical structure. The 'Constituent #1' pane is highlighted with a red dashed box and a blue callout labeled '1'. A right-click context menu is open over 'Constituent #1', with the 'Focus' option selected and highlighted by a red dashed box and a blue callout labeled '2'. The 'Focus' option is underlined in the menu. The 'Constituent #2' and 'Constituent #3' panes show other chemical structures. At the bottom, a table displays data for the selected constituents, including 'MS: >2E+03 µM' and 'MS: GHS criteria not met'.

Selection of constituent # 1 for further read-across analysis

1. Right mouse click over the constituent without experimental data (constituent #1),
2. Select **Focus** from the drop down menu.

Read across prediction of constituent without data

Focus constituent

The screenshot shows the QSAR Toolbox interface. The 'Documents' panel on the left lists 'Document 1' with a tree structure: '[C: 1-Md: 0-P: 0] Substance' -> '[C: 4-Md: 39-P: 0] Composition list' -> '[C: 1-Md: 0-P: 0] Constituent #1' (highlighted with a red circle and the number 1). Below this is the 'Databases' panel with a list of 8 selected databases, including 'Cell Transformation Assay ISSCTA', 'Dendritic cells COLIPA', 'Developmental & Reproductive Toxicity (DART)', 'Developmental toxicity database (CAESAR)', 'Developmental toxicity ILSI', 'ECHA REACH', 'ECOTOX', 'Eye Irritation ECETOC', 'Food TOX Hazard EFSA', 'GARD Skin sensitization', 'Genotoxicity & Carcinogenicity ECVAM', 'Genotoxicity OASIS', 'Genotoxicity pesticides EFSA', 'Human Half-Life', 'Keratocyte gene expression Gvaudan', 'Keratocyte gene expression LuSens', 'Micronucleus ISSMIC', 'Micronucleus OASIS', 'MUNRO non-cancer EFSA', 'REACH Skin sensitization database (normalised)', 'Receptor Mediated Effects', 'Rep Dose Tox Fraunhofer ITEM', 'Repeated Dose Toxicity HESS', 'Rodent Inhalation Toxicity Database', 'Skin Irritation', 'Skin Sensitization', 'Skin sensitization ECETOC', and 'ToxCast'. The main panel shows a 'Filter endpoint tree...' with a chemical structure of a constituent and a list of endpoints including 'Structure info', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Human Health Hazards', 'Acute Toxicity', 'ADME', 'Bioaccumulation', 'Carcinogenicity', 'Developmental Toxicity / Teratogenicity', 'Genetic Toxicity', 'Immunotoxicity', 'Irritation / Corrosion', 'Neurotoxicity', 'Photoinduced toxicity', 'Repeated Dose Toxicity', 'Sensitisation', 'ToxCast', 'Toxicity to Reproduction', and 'Toxicokinetics, Metabolism and Distribution'. A red box highlights the text 'This focused component appeared in separate data matrix'.

1. A documented tree with focused constituent #1 is automatically selected. The workflow could move further with collecting analogues of the focused constituent #1.

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Data
 - **Read across prediction of constituent without data**
 - Focus constituent without experimental data
 - **Define category**

Category Definition Overview

- This module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- This is the critical step in the workflow.
- Several options are available in the Toolbox to assist the user in refining the category definition.
- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity”.

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers.
2. Endpoint specific profilers (for sub-cat).
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements).

Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling
3. More categories of same Phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

**Broad grouping
Endpoint Non-specific**

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

**Subcategorization
Endpoint Specific**

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

**Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity**

**Subcategorization
Endpoint Specific**

Read across prediction of constituent without data

Forming category for studied endpoint

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

Repeating Phase I due to Multifunctionality of chemicals

Broad grouping
Endpoint Non-specific

Phase I categorization in Toolbox

It is not recommended to use "Neutral organic" * as phase I

97 analogues are identified as Aryl halides by OFG

*Neutral organic category include chemicals having different functionalities as alcohols, ketones, ethers etc. In this respect the basic principle that structurally similar chemicals may elicit similar effects would not be preserved, because Neutral organic mixed many different functionalities

Read across prediction of constituent without data

Forming category for studied endpoint

- Based on the above recommendations the OFG is used as initial categorization group
- Refinement of the initial group is based on endpoint-specific protein binding profiler:
 - Protein binding alerts for skin sensitization.

Category definition is a tool for grouping chemicals, which allows to group chemicals based on different measures of “similarity”. For more details see tutorials posted on QSAR Toolbox website:

<https://qsartoolbox.org/support/>

See next slides

Read across prediction of constituent without data

Define category by OFG

Based on above recommendations the OFG is used as an initial group (phase I)

1. Go to **Category definition** module;
2. Select "**Sensitization**" level of endpoint tree;
3. Select **Organic functional groups** (OFG) and click on **Define**;
4. Combination of three organic functional groups has been identified in the target chemical (in our case constituent #2), which will be used for searching analogues, click **OK**;
5. a list of 97 chemicals has been found having all the three categories identified in the target chemical; gather data for the analogues (see next slide)

Read across prediction of constituent without data

Gather data for analogues chemicals

Experimental data for the identified analogues appear on data matrix

228 points added across 64 chemicals.

In the current case we will apply a read-across for *in vivo skin sensitization* without taking into account the specific endpoint. In other words we will use combination of endpoints LLNA and GPMT for read-across and filling data gap (see next slide)

| Test | Points | Chemicals | Read-across | Notes |
|-------------------------------------|--------|-----------|-------------|---|
| in Vivo | 14/21 | 14 | 1 | M: GHS criteria |
| Buehler Test | 14/21 | 14 | 1 | M: GHS criteria |
| Freund's Complete Adjuvant Test | 2/2 | 2 | 0 | M: Negative |
| GPMT | 27/62 | 27 | 6 | M: Negative |
| Intracutaneous Test | 1/1 | 1 | 0 | M: Negative |
| LLNA | 45/126 | 45 | 12 | M: not sensitising, M: GHS criteria n... M: not sensitising |
| Mouse Local Lymph Node Assay (LLNA) | 1/4 | 1 | 4 | M: 20.5 % M: GHS criteria n... M: Negative |

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Data
 - **Read across prediction of constituent without data**
 - **Focus constituent without experimental data**
 - **Define category**
 - **Apply read across**

Read across prediction of constituent without data

Apply read across

The screenshot displays the QSAR Toolbox software interface. The top menu bar has 'Data Gap Filling' highlighted with a red box and a callout '1'. The left sidebar shows 'Read across' circled in red with a callout '3'. The central table shows chemical structures and data points, with the 'In vivo' section circled in red and a callout '2'. The right dialog box, titled 'Possible data inconsistency', shows a list of assays and endpoints. 'Skin sensitization II (ECETOC)' is selected with a callout '4'. The 'OK' button is highlighted with a callout '5'.

1. Go to **Data Gap filling** module;
2. Click on the cell corresponding to Skin Sensitization in Vivo (i.e. in this case we will combine all the data stored under "In vivo" level);
3. Click on **Read-across** button;
4. Select scale/unite **Skin sensitization II(ECETOC)**;
5. Click **OK** (in this case we mix all endpoints and assays).

Read across prediction of constituent without data

Apply read across

The screenshot displays the QSAR Toolbox software interface. The main window shows a 'Filter endpoint tree' on the left with various toxicity endpoints. A 'Choose one' dialog box is open, allowing selection of a prediction mode. The 'Maximal' option is selected and highlighted with a red dashed box and a blue callout labeled '2'. The 'OK' button is also highlighted with a red dashed box. Below the dialog, a scatter plot titled 'Read-across prediction for EC3, Skin sensitisation, based on 12 values' shows data points for 'Positive' and 'Negative' outcomes. On the right, a 'Calculation options' panel is visible, with 'Data usage' selected and highlighted with a red dashed box and a blue callout labeled '1'. The 'Use target data for prediction' option is also visible.

Apply worst case scenario. In this case we need to take the maximal values for data usage (in other words if there are multiple data points for one chemical, such as negative and positive, then the positive data will be taken for read-across prediction)

1. Open **Calculation options** and click on **Data usage**;
2. Select Maximal and click **OK**

Read across prediction of constituent without data

Subcategorization

- The initial category could be refined by subcategorizing the analogues according to the “Protein binding alerts for skin sensitization by OASIS” and Structural similarity profilers.
- These steps are summarized in the next screen shots.

Read across prediction of constituent without data

Subcategorization by Protein binding alert for SS

The screenshot displays the OECD QSAR Toolbox interface. On the left, the 'Subcategorization' panel shows a list of alerts. Alert 2, 'Protein binding alerts for skin sensitization by OASIS', is highlighted with a red circle and a callout labeled '2'. Below it, the 'Metabolisms' section shows a list of metabolic pathways, with 'Skin metabolism simulator' selected. A callout labeled '3' points to the 'Selected 42 (5/47)' list. On the right, the 'Analogues' panel shows a grid of chemical structures. A red box highlights the text: 'The read-across is based on four analogs with positive and negative SS data, reacting with proteins by the same mechanism as the target chemical (Schiff-base formation)'. Below this, a scatter plot titled 'Read-across prediction for EC3. Skin sensitisation, based on 5 values' shows 'Predicted: Negative'. The plot has 'log Kow' on the x-axis (ranging from -5 to 12) and 'EC3 Skin sensitisation' on the y-axis (Positive and Negative). A red circle highlights the data points. On the far right, a panel labeled '1' shows the 'Select / filter data' options, including 'Subcategorize', 'Mark chemicals by WS', 'Mark chemicals by descriptor value', 'Filter points by test conditions', 'Mark focused chemical', 'Mark focused points', 'Remove marked data', and 'Clear existing marks'. The 'Accept prediction' button is at the bottom right.

1. Select filter data/Subcategorize
2. Select **Protein binding alerts for skin sensitization by OASIS**.
3. **Remove selected** to eliminate dissimilar chemicals, reacting by different protein binding mechanisms;

1. **Select filter data/Subcategorize**
2. Select **Structural similarity**.
3. Select first two bins with most dissimilar analogues ([20%,30%] and [30%,40%])
4. **Remove selected** to eliminate dissimilar chemicals;
5. Click **Accept prediction**

Read across prediction of constituent without data

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. Below this is a 'Workflow' section with 'Gap Filling' and 'Standardized Automated'. The main window shows a 'Documents' panel on the left with a tree view of chemical constituents. The central area displays a 'Filter endpoint tree...' with a list of endpoints and their associated data. A red dashed box highlights the 'GPMT <OR> LLNA' endpoint, which is marked as 'R: Positive'. A red text box at the bottom states: 'The read-across prediction for the constituent #1 generates a new level of the endpoint tree.'

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Read across prediction of constituent without data
 - **Filling data gap for skin sensitization of mixture**

Data Gap Filling Overview

- “Data Gap Filling” module gives access to two different data gap filling tools:
 - **Independent MOA-** all components are with different mode of action
 - **Similar MOA-** all components are with similar mode of action
- More details about different MOA could be found in F1 help
- In this particular case all components of the current mixture are with dissimilar mode of action. In this respect Independent MOA is applied

Data Gap Filling Independent MOA

Assumption – combined effect can be calculated from the effects caused by the individual mixture components by following the statistical concept of independent random events

Mixture response:
$$E(C_{Mix}) = 1 - \prod_{i=1}^N [1 - E(C_i)]$$

$E(C_{Mix})$ - the effect provoked by the total mixture

$E(C_i)$ - the effects that the individual components would cause if applied singly at that concentration at which they are present in the mixture

Problem - dose-response relationships are practically unknown

Data Gap Filling

Case study

- In this particular case components of the current mixture have different modes of action (constituent #1 and #3 have same mode, they do not interact with proteins (see slide 32), however constituent #2 interacts with proteins via Schiff-base mechanism). In this respect Independent MOA is applied;
- Application of Independent MOA for this case study is illustrated on the next slides

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

The screenshot shows the QSAR Toolbox interface. The top toolbar has icons for Profiling, Data, Category definition, Data Gap Filling, and Report. The sidebar on the left shows a document tree with 'Composition list' highlighted. The main workspace displays a table with columns for Parent chemical, Constituent #1, Constituent #2, and Constituent #3. A red dashed box highlights a row in the table, and a red arrow points to it from a text box. Another red arrow points to a text box containing experimental data for skin sensitization.

Here are the experimental data for Skin sensitization for two of the mixture's constituents

Here is the Read across prediction for the constituent without data

1. Once you are in the Data gap filling module, collapse the endpoint tree to the level of "in Vivo" as shown on the snapshot above;
2. Click on the cell corresponding to this mixture;
3. Click on **Composition list** (highlighted) from documented three;
4. The latter action automatically activates two modes of action: **Independent MOA** and **Similar MOA**.

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

1 Click on the cell corresponding to the Skin Sensitization;

2 Select **Independent MOA**;

3 Select **Skin sensitization II (ECETOC)**;

4 Click **OK**.

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Documents

- Document 1
 - Substance
 - Composition list
 - Constituent #1
 - Ketone<AND>Aryl<AND>Aryl halide (Organic)
 - Enter GF(RA) with 30 chemicals, 66 data poi
 - Data usage options are changed to: Ma
 - Ch: 3| Data: 2 Subcategorized: Protei
 - Constituent #2
 - Enter GF(IndependentMOA) with 4 chemicals, 6

Filter endpoint tree...

Structure

Photoinduced toxicity

Repeated Dose Toxicity

Sensitisation

Skin

in Vitro 1/12

in Vivo 3/6

ToxCast

Toxicity to Reproduction

Toxicokinetics, Metabolism and Distributi...

Parent chemical [target]

Constituent #1

Constituent #2

Constituent #3

MS: >2E+03 µM

MS: Negative

R: Positive

MS: Category C

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

QSARs 0

Automated workflows 0

Standardized workflows 0

In nodes below:

QSARs 0

Automated workflows 0

Standardized workflows 0

Descriptors

Prediction

Empirical calculation of A B C, EC3, Skin sensitisation, based on 6 values

Predicted: Negative

A B C EC3, Skin sensitisation

log Kow

Active descriptor X log Kow

Select / filter data

Descriptors / data

Calculation options

Visual options

Information

Miscellaneous

Accept prediction

Read across is applied for the mixture (assuming Independent Mode of Action)

Filling data gap for skin sensitization of mixture

Applying Independent Mode of Action

Documents

- Document 1
 - [C: 1;Md: 0;P: 0] Substance
 - [C: 4;Md: 39;P: 1] Composition list
 - [C: 1;Md: 0;P: 1] Constituent #1
 - [C: 97;Md: 228;P: 1] Ketone<AND>Aryl<
 - [C: 48;Md: 169;P: 1] Enter GF(RA)
 - [C: 48;Md: 169;P: 1] Data usage o
 - [C: 6;Md: 10;P: 1] Subcategory
 - [C: 4;Md: 7;P: 1] Subcategory
 - [C: 1;Md: 0;P: 0] Constituent #2
 - [C: 1;Md: 0;P: 0] Constituent #3
 - [C: 4;Md: 39;P: 1] Enter GF(IndependentMOA)
 - [C: 4;Md: 39;P: 1] Data usage options

Filter endpoint tree...

 - Structure
 - Developmental Toxicity
 - Genetic Toxicity
 - Immunotoxicity
 - Irritation / Corrosion
 - Neurotoxicity
 - Photoinduced toxicity
 - Repeated Dose Toxicity
 - Sensitisation
 - Skin
 - in Chemico 1/8 MS: GHS criteria
 - in Vitro 1/17 MS: >2E+03 µM
 - in Vivo 3/14 R: Positive MS: GHS criteria
 - Undefined Type of Method 1/1
 - ToxCast
 - Toxicity to Reproduction
 - Toxicokinetics, Metabolism and Distribution

Choose one

Choices

 - ☐ Mode
 - ☐ Lowest mode
 - ☐ Highest mode
 - ☐ Median
 - ☐ Lower media
 - ☐ Higher media
 - ☐ Minimal
 - ☒ Maximal
 - ☐ All

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

| Method | Count |
|------------------------|-------|
| QSARs | 0 |
| Automated workflows | 0 |
| Standardized workflows | 0 |

In nodes below:

| Method | Count |
|------------------------|-------|
| QSARs | 0 |
| Automated workflows | 0 |
| Standardized workflows | 0 |

Descriptors

| Descriptor | Prediction |
|------------------------------|------------|
| A B C EC3 Skin sensitisation | Positive |

Plot

A B C EC3 Skin sensitisation

log Kow

Empirical calculation of A B C, EC3, Skin sensitisation, based on 3 values
Predicted: Positive

Consecutive steps:

 1. Calculation options;
 2. Data usage;
 3. Maximal data and click OK;
 4. Accept prediction.

Read across prediction for the mixture (labeled as IMO) based on predicted (marked as R) and experimental data of the mixture constituents (marked with MS) appears on data matrix

Outlook

- Background
- Keywords
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Read across prediction of constituent without data
 - Filling data gap for skin sensitization of mixture
 - **Generating report for mixture**

Report

- Remember the report module allows the user to generate a report on the predictions performed with the Toolbox.
- The report can be printed or saved in different formats.
- Generating the report is shown on next screenshots.

Report

The screenshot illustrates the steps to generate a report in the QSAR Toolbox. The main window shows the 'Report' section in the top menu (1). The 'Prediction' button in the left sidebar is highlighted (3). The 'Generated report files' dialog box (6) lists the files generated for the mixture and individual constituent. The 'Customize report content and appearance' wizard (4) is open, showing the 'Prediction' section. A warning dialog (5) is displayed, indicating that some chemicals or data are not available for prediction. The 'Create report' button is highlighted (4).

Summary report for the whole mixture

Individual report for the constituent for which a read-across was performed

Generated report files

The following files were generated:

- Prediction for mixture
 - Prediction report
 - Category report
 - Data matrix
- Individual prediction #1
 - Prediction report
 - Category report

PDF file containing the prediction report

Open Save as

Customize report content and appearance

Wizard pages

- Prediction for mixture
 - Customization
 - Customize report
 - Prediction
 - Target and prediction summary
 - Prediction details (I)
 - Target profiles
 - Mixture components used for prediction
 - Category
 - Category definition and members
 - Consistency check
 - Options
 - Data matrix
 - Options

Warning

Highlighted is an individual report (report for a mixture, or report for one of its components). Expand the tree on the left and customize the appearance of the selected individual report.

Due to restrictions some chemicals or data are not available for prediction or in supplementary files

OK

This is an individual report for prediction for mixture

To customize the report appearance navigate through report section using [Back] and [Next] buttons

Back Next Cancel Create report

1. Go to **Report** section;
2. Click on the cell corresponding to IMOA prediction;
3. Click **Prediction** button. A wizard appears where the user could customize the sections;
4. Click **Create report**.
5. Click **OK** on the appeared message; Two reports are generated: one summary report for the mixture and one for the read-across prediction for the constituent #1.
6. Select **prediction of mixture** and click **Open** button.

Report

Prediction of A B C, EC3, Skin sensitisation for mixture

1 / 7

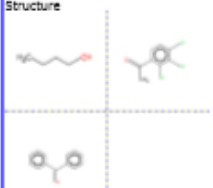
QSAR Toolbox prediction for multicomponent substance

Based on observed and predicted data for mixture components

Date: 13 Apr 2020

Author(s):

Contact details:

| Target information | | |
|---|--|----------------|
| Structural information | Numerical identifiers | Chemical names |
| <p>SMILES:</p> <chem>CCCCO.CC(=O)c1ccc(Cl)cc1</chem> <chem>c1ccc(Cl)cc1.O=C(c1ccc(Cl)cc1)c1ccccc1</chem> | <p>CAS#: No CAS number</p> <p>Other: N/A</p> | |
| <p>Structure</p>  | | |

| Prediction summary | | |
|---|-----------------|--|
| <p>Predicted endpoint: A B C, EC3, Skin sensitisation; No effect specified; No species specified; No duration specified; No guideline specified</p> <p>Predicted values: Positive</p> <p>Unit/scale: Skin Sensitization (Danish EPA)</p> <p>Data gap filling method: Independent mode of action</p> <p>Summary: manually editable field</p> <p>Not provided by the user</p> | | |
| | Quantity, mol % | In Vivo, Skin Sensitization (Danish EPA) |
| Target mixture | - | |

Prediction of A B C, EC3, Skin sensitisation for mixture

2 / 7

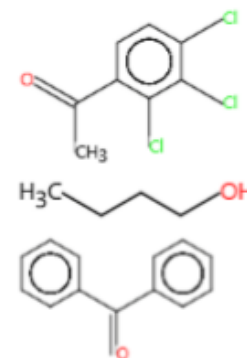
| | | |
|----------------------|------|------------|
| Mixture component #1 | 33.3 | Positive |
| Mixture component #2 | 33.3 | Positive |
| Mixture component #3 | 33.3 | Category C |

NOTE: Predicted values are written in *italic*

Mixture component #1
CC(=O)c1ccc(Cl)cc1

Mixture component #2
CCCCO

Mixture component #3
O=C(c1ccccc1)c1ccccc1



Information of the mixture's constituents

Prediction summary

Outlook

- Background
- Keywords
- Objectives
- The exercise
- Workflow
- **Save the prediction result**

Saving the prediction result

- Saving functionality allows storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions, etc.
- This functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving the file with TB prediction is illustrated on next screenshot.

Saving the prediction result

The screenshot shows the QSAR Toolbox software interface. The 'Input' module is selected in the top toolbar (callout 1). The 'Save' button in the top toolbar is circled (callout 2). A 'Save document' dialog box is open, showing the 'Documents' folder (callout 3). The 'File name' field is set to 'Tutorial 10' (callout 4). The 'Save' button in the dialog is highlighted (callout 5). The 'Open' button in the top toolbar is also visible (callout 6).

1. Go to **Input** module;

2. Click **Save** button;

3. Browse to the folder on your PC;

4. Give **name of the file**;

5. Click **Save** button. The file is saved and could be opened later by using **Open** button (6).