

## OECD QSAR Toolbox v.4.1

Step-by-step example on how to predict the skin sensitisation potential approach of a chemical by read-across based on an analogue approach

# Outlook

- **Background**
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

## Background

- This is a step-by-step presentation designed to take the first-time user of the Toolbox through the workflow of a data filling exercise by read-across based on an analogue approach.

# Outlook

- Background
- **Objectives**
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

# Objectives

**This presentation demonstrates a number of functionalities of the Toolbox:**

- Identify analogues for a target chemical.
- Retrieve experimental results available for those analogues.
- Fill data gaps by read-across.

# Outlook

- Background
- Objectives
- **Specific Aims**
- Read across and analogue approach
- The exercise
- Workflow
- Save the prediction result

## Specific Aims

- To introduce to the first-time user the workflow of Toolbox.
- To familiarize the first-time user with the six modules of Toolbox.
- To familiarize the first-time user with the basic functionalities within each module.
- To explain to the first-time user the rationale behind each step of the exercise.

# Outlook

- Background
- Objectives
- Specific Aims
- **Read-across and analogue approach**
- The exercise
- Workflow
- Save the prediction result



# Read-across and Analogue Approach Overview

- A read-across can be used to estimate missing data from a single or limited number of chemicals using an analogue approach. It is especially appropriate for “qualitative” endpoints for which a limited number of results are possible (e.g. positive, negative, equivocal).
- In the analogue approach, endpoint information for a single or small number of tested chemicals is used to predict the same endpoint for an untested chemical that is considered to be “similar”.
- Analogous sets of chemicals are often selected based on the hypothesis that the toxicological effects of each member of the category will show a common behaviour.

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- **The exercise**
- Workflow
- Save the prediction result

## The Exercise

- In this exercise we will predict the skin sensitization potential for an untested compound, (4-nitrobenzoyl chloride) [CAS # 122-04-3], which will be the “target” chemical.
- This prediction will be accomplished by collecting a small set of test data for chemicals considered to be in the same category as the target molecule.
- The category will be defined by the mechanism of protein binding common to all the chemicals in the category.
- The prediction itself will be made by “read-across”.

# The Exercise

## Theoretical considerations on Skin Sensitization

- Allergic contact dermatitis that results from skin sensitization is a significant health concern.
- Skin sensitization is a toxicological endpoint that is complex and conceptually difficult.
- However, there is a growing agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, the mechanisms by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents.

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
- Save the prediction result

# Workflow

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

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - **Chemical 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 Chemical

### User Alternatives for Chemical ID:

#### A. Single target chemical

- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- SMILES (simplified molecular information line entry system) notation/InChi
- Drawing chemical structure
- Select from User List/Inventory/Databases
- Chemical IDs such as EC number, Einecs number
- Substructure search by using SMARTs

#### B. Group of chemicals

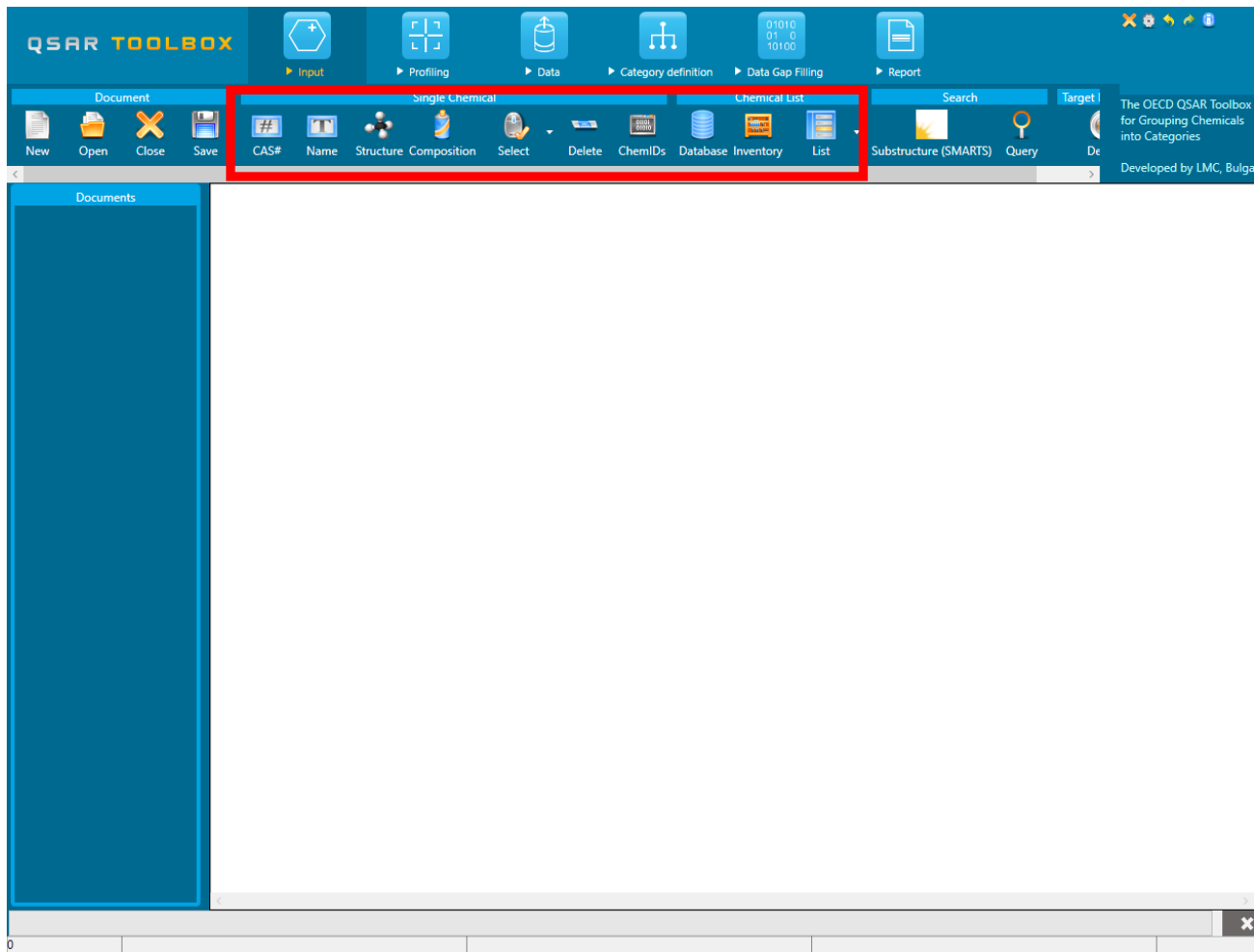
- User List/Inventory
- Specialized Databases

## Getting Started

- Open the Toolbox.
- The six modules in the workflow are seen listed next to "QSAR TOOLBOX".
- **Click** on "Input" (see next screen shot)

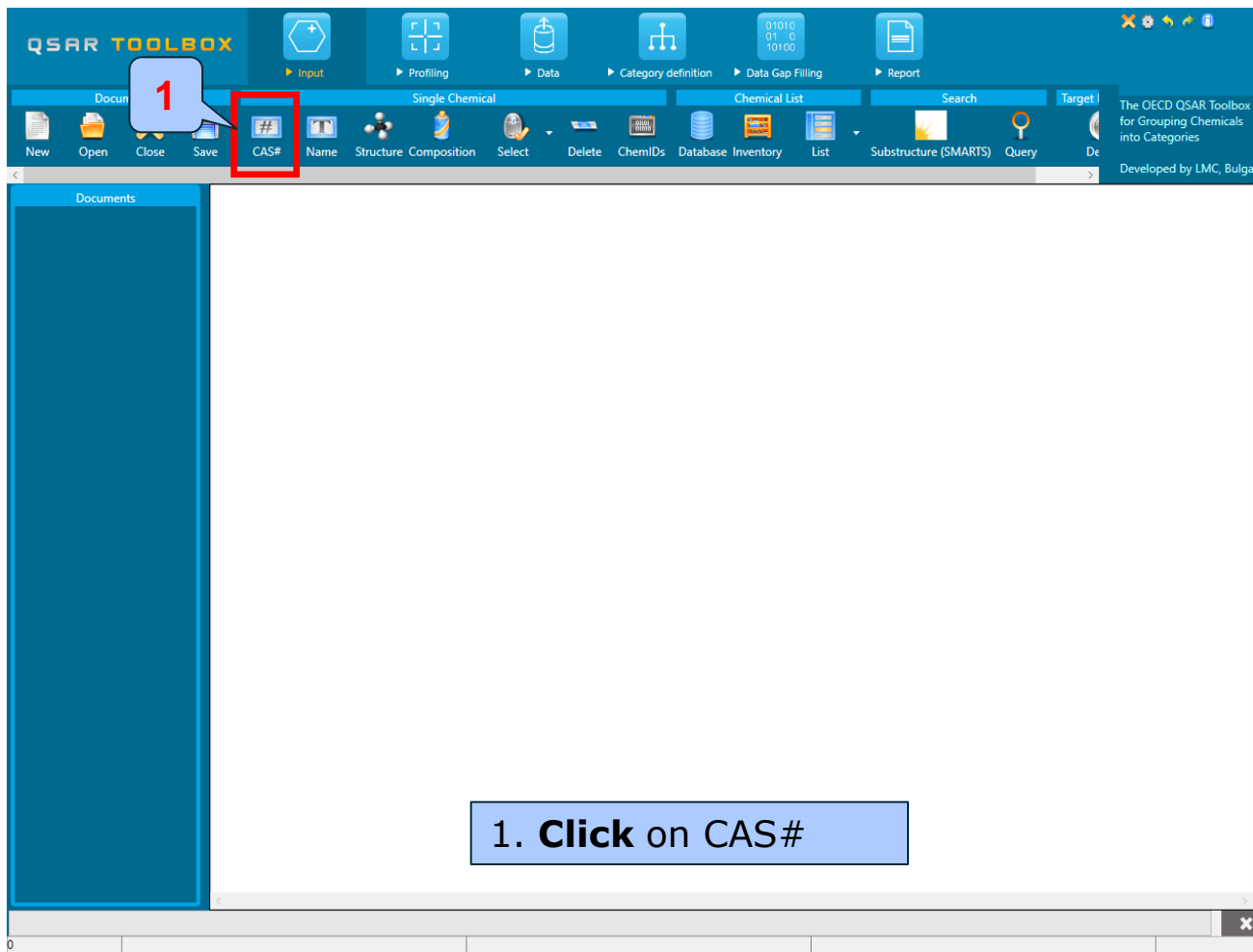
# Chemical Input Screen

## Input screen



# Chemical Input Screen

## Input target chemical by CAS#



# Chemical Input Screen

## Enter CAS# of 4-nitrobenzoyl chloride

The screenshot shows a dialog box titled "Search by CAS #". At the top left, there is a dropdown menu containing the text "122-04-3", which is highlighted with a red rectangular box and a blue callout bubble labeled "1". To the right of the dropdown is a "Search" button, with a blue callout bubble labeled "2" pointing to it. Further to the right are "OK" and "Cancel" buttons, with a blue callout bubble labeled "3" pointing to the "OK" button. Below the search controls, there are buttons for "Select All", "Unselect All", and "Reset Selection", along with the text "Selected 0 of 0". The main area of the dialog box is currently empty.

1. **Enter** the CAS# In the field; 2. **Click** Search button; 3. **Press** OK

# Chemical Input

## Target chemical identity

- Double click “CAS Smiles relation” displays the chemical identification information.
- This indicates the reliability of relation CAS-Name for the target chemical(see next screen shots).
- The workflow on the first module is now complete, and the user can proceed to the next module.

# Chemical Input

## Target chemical identity

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options for Document, Single Chemical, and Chemical List. The main workspace is divided into three sections:

- Documents:** Shows a single document named "Document 1" with the CAS number "# CAS: 122043".
- Filter endpoint tree...:** A tree view on the left lists various chemical properties and endpoints, including Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards.
- 1 [target]:** A table on the right displays the chemical's identity, with a red circle highlighting the highlighted row:
 

122-04-3
High
Benzoyl chloride, 4-nit
C7H4ClNO3
Unspecified
[O-][N+](=O)c1ccc(cc1

# Chemical Input

## Target chemical identity

The code indicates the reliability of the chemical identifier:

- **High:** This reliability corresponds to high reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to at least one high quality data source (database or inventory)
- **Moderate:** This reliability corresponds to moderate reliability of CAS-SMILES relation. The moderate label is assigned if the chemical belongs to three "Distribute to QA" data sources.
- **Low:** This reliability corresponds to poor reliability of CAS-SMILES relation. This label is assigned if the chemical belongs to less than three, but at least one "Distribute to QA" data sources.



# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical 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.

# Profiling Overview

Summary information of the different profilers are provided in the "About"

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Profiling' sub-menu is active, showing 'Apply', 'View', 'New', and 'Delete' options. On the left, a 'Documents' panel shows 'Document 1' with CAS: 122043. Below it, the 'Profiling methods' panel lists various methods, with 'Protein binding alerts for skin sensitization by OASIS' selected. A red callout '1' points to this selection. The 'About' dialog box for this profiler is open, showing its name, short description, and full description. A red callout '2' points to the 'About' button in the dialog's title bar. A red callout '3' points to the 'Close' button in the dialog's title bar. The background shows a chemical structure and a 'Filter endpoint tree'.

1. **Select** the name of the profiler, perform **right click** on it and then
2. **Select** About
3. **Close** before proceeding

# Profiling

## Overview

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS) and clicking on “View” (see next screen shot).

# Profiling Overview

1. Highlight the profiler

2. Click View

Protein binding alerts for skin sensitization by OASIS (Endpoint Specific) - Profiling Scheme Browser

Categories

- Protein binding alerts for skin sensitization
  - Acylation
    - (Thio)carbamoylation of protein nu
    - Isocyanates, Isothiocyanates**
    - Acyl transfer via nucleophilic additi
    - Carbodiimides
    - Direct acylation involving a leaving
    - (Thio)Acetates
    - (Thio)Acyl and (thio)carbamoyl
    - Anhydrides (sulphur analogues
    - Azlactones and unsaturated lac
    - Carbamates
    - Diacyl peroxides, anhydrides (su
    - N-Acylloxysuccinimides
    - N-Carbonyl heteroaryl amines
    - N-Carbonylsulfonamides
    - N-Haloacylamides
    - Phosphonyl halides or cyanides
    - Sulphonyl halides or cyanides
    - Thiosulfonates
    - Thiosulfonates
  - Ester aminolysis
    - Amides
    - Dithiocarbamate salts
    - Dithiocarbamates
    - Dithioesters
  - Ester aminolysis or thiolysis
    - Activated (di)aryl esters
    - Activated (thio)esters
    - Activated alkyl diesters
    - Benzyl or phenethyl salicylates
    - Phenyl carbonates
    - Substituted benzyl benzoates
  - Isocyanates and related chemicals
    - Ketenes
    - Ring opening acylation
      - Active cyclic agents
      - beta-Lactams
      - Cyclopropanones

Definition Properties Training Set Literature MetalInfo Table Scheme

[115] Isocyanates, Isothiocyanates

Category tree

Query details

[0] Structure Query Metabolism

Contents

- Queries
- Masks

Search 1: SMARTS

SMARTS

[#6][#7]=[#6]=[#16]

View mode: Facade Navigation mode: Cascade

Left click on any marked atom to explore

# Profiling

## Profiling the target chemical

- The outcome of the profiling determines the most appropriate way to search for analogues (detailed information in Manual for getting started (Chapter 4). <http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- Table 4 - 1 in chapter 4 (Manual for getting started) lists a selection of profilers and their relevance for different endpoints of regulatory relevance.
- For example the following mechanistic and endpoint specific profiling schemes are relevant to the Skin sensitization:
  - Protein binding by OASIS – mechanistic grouping
  - Protein binding alerts for skin sensitization by OASIS – endpoint specific
  - Protein binding alerts for skin sensitization according to GHS – endpoint specific
  - Protein binding by OECD – mechanistic grouping
  - Protein Binding Potency – mechanistic grouping

# Profiling

## Profiling the target chemical

- **Tick the** box of the selected profiling methods related to the target endpoint.
- This selects (a **green** check mark appears) or deselects (**green** check mark disappears) profilers.
- For this example, **tick** all the general mechanistic profilers and **click** on apply (see next screen shot).

# Profiling

## Profiling the target chemical: Example

The screenshot shows the QSAR Toolbox interface. At the top, there are 'Input' and 'Profiling' tabs. Below them are 'Apply', 'View', 'New', and 'Delete' buttons. A callout box with the number '3' points to the 'Apply' button. The 'Documents' panel shows 'Document 1' with CAS: 122043. The 'Profiling methods' panel has an 'Options' menu with 'Select All', 'Unselect All', 'Invert', and 'About'. Under 'Options', there are several categories: 'OECD HPV Chemical Categories', 'Substance type', 'US-EPA New Chemical Categories', and 'General Mechanistic'. The 'General Mechanistic' category is highlighted with a red box and callout box '1'. Below it, a list of methods is shown, with 'Biodegradation primary (Biowin 4)', 'Biodegradation probability (Biowin 1)', 'Biodegradation probability (Biowin 2)', 'Biodegradation probability (Biowin 5)', and 'Biodegradation probability (Biowin 6)' all checked. A red circle highlights this list, with callout box '2' pointing to it. The 'Filter endpoint tree...' panel on the right shows a tree structure with categories like 'Structure info', 'Parameters', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazard'.

1. Check General mechanistic
2. All profilers in this category will be selected
3. Click Apply



# Profiling

## Profiling the target chemical

- The actual profiling will take up to several seconds depending on the number and type of profilers selected.
- The results of profiling automatically appear as a dropdown box under the target chemical (see next screen shot).
- Please note the specific protein-binding profiler – Protein binding by OASIS (see side-bar on sensitisation above).
- This result will be used to search for suitable analogues in the next steps of the exercise.

# Profiling

## Profiling the target chemical

The screenshot displays the QSAR Toolbox interface. On the left, the 'Documents' panel shows 'Document 1' with CAS: 122043. Below it, the 'Profiling methods' and 'Metabolism/Transformations' panels are visible. The central 'Filter endpoint tree...' panel is expanded to 'Protein binding by OASIS', which is highlighted with a red box. A red arrow points from this box to the 'Acylation' result in the '1 [target]' table, which is also circled in red. The table shows various endpoints and their corresponding results for the target chemical.

Endpoint	Result
Biodeg BioHC half-life (Biowin)	No value
Biodegradation primary (Biowin 4)	days - weeks
Biodegradation probability (Biowin 1)	Does NOT Biodegrade f
Biodegradation probability (Biowin 2)	Does NOT Biodegrade f
Biodegradation probability (Biowin 5)	Does NOT Biodegrade f
Biodegradation probability (Biowin 6)	Does NOT Biodegrade f
Biodegradation probability (Biowin 7)	Does NOT Biodegrade f
Biodegradation ultimate (Biowin 3)	weeks - months
DNA binding by OASIS	Radical
DNA binding by OECD	Acylation
Estrogen Receptor Binding	Non binder, without O
Hydrolysis half-life (Ka, pH 7)(Hydrowin)	No value
Hydrolysis half-life (Ka, pH 8)(Hydrowin)	No value
Hydrolysis half-life (Kb, pH 7)(Hydrowin)	No value
Hydrolysis half-life (Kb, pH 8)(Hydrowin)	No value
Hydrolysis half-life (pH 6.5-7.4)	No value
Ionization at pH = 1	No pKa value
Ionization at pH = 4	No pKa value
Ionization at pH = 7.4	No pKa value
Ionization at pH = 9	No pKa value
Protein binding by OASIS	Acylation
Protein binding by OECD	Acylation
Protein binding potency	Not possible to classify
Protein binding potency Cys (DPRA 13%)	Out of mechanistic dom
Protein binding potency Lys (DPRA 13%)	Out of mechanistic dom
Toxic hazard classification by Cramer	High (Class III)
Toxic hazard classification by Cramer (extended)	High (Class III)

1. Go to Protein binding by OASIS to review the profiling results.

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - **Data**

# Data Overview

- “Data” refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion (i.e., collecting all data for all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).

# Data

## Case study

- In this example, we limit our data gathering to a single toxicity endpoint (skin sensitization).
- In this example, we collect data from the databases containing experimental results for Skin sensitisation (Skin sensitisation and Skin sensitisation ECETOC).
- **Click** on “Data” in the Toolbox workflow.
- **Expand the** “Human Health Hazards” section
- **Click** on the box to select that database.
- **Click** on “Gather data” (see next screen shot).

# Data Gather data

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Data', 'Import', and 'Export'. The 'Data' menu is open, showing 'Gather', 'Import', 'IUCLID6', and 'IUCLID6'. A red circle highlights the 'Gather' button, with a callout '3'. The 'Documents' panel on the left shows a list of databases. 'Human Health Hazards' is expanded, and 'Skin Sensitization' and 'Skin sensitization ECETOC' are selected, with a red circle and callout '2'. A callout '1' points to the 'Human Health Hazards' section. The 'Filter endpoint tree...' panel on the right shows a tree structure for 'Human Health Hazards' with various sub-endpoints. A chemical structure is shown in the top right of the filter panel.

1. **Expand** the Human Health Hazards section
2. **Select** databases related to the target endpoint
3. **Click** Gather

# Data

## Gather data

- Toxicity information on the target chemical is electronically collected from the selected dataset(s).
- It should be kept in mind that the search for data and analogues is performed only among the chemicals which are listed in the selected databases which in this example are Skin sensitization and Skin sensitization ECETOC .
- In this example, an insert window appears stating there was “no data found” for the target chemical (see next screen shot).

# Data Gather data

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', and 'Export'. The 'Data' tab is active, showing a 'Filter endpoint tree...' dialog box. The dialog box contains a list of endpoints and a table of values. An arrow points to the 'OK' button in the dialog box.

**Filter endpoint tree...**

1 [target]

Structure

Ecotoxicological information

Human Health Hazards

Profile

General Mechanistic

Biodeg BioHC half-life (Biowin)

Biodegradation primary (Biowin 4)

Biodegradation

Biodegradation

Biodegradation

Biodegradation

Biodegradation

DNA binding by OECD

Estrogen Receptor Binding

Hydrolysis half-life (Ka, pH 7)(Hydrowin)

Hydrolysis half-life (Ka, pH 8)(Hydrowin)

Hydrolysis half-life (Kb, pH 7)(Hydrowin)

Hydrolysis half-life (Kb, pH 8)(Hydrowin)

Hydrolysis half-life (pH 6.5-7.4)

Ionization at pH = 1

Ionization at pH = 4

Ionization at pH = 7.4

Ionization at pH = 9

Protein binding by OASIS

Protein binding by OECD

Protein binding potency

Protein binding potency Cys (DPRA 13%)

Protein binding potency Lys (DPRA 13%)

Toxic hazard classification by Cramer

Toxic hazard classification by Cramer (extended)

No value
days - weeks
No value
No value
No value
No value
No value
No value
No value
No value
No value
No value
Basic [0.000 , 10.000]
Basic [0.000 , 10.000]
Basic [0.000 , 10.000]
Basic [0.000 , 10.000]
Basic [0.000 , 10.000]
Acylation
Acylation
Acylation
Not possible to classify
Out of mechanistic dom
Out of mechanistic dom
High (Class III)
High (Class III)

Close the inserted window by **Clicking** on "OK"



# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Data
  - **Category definition**

## Recap

- In module one, you have entered the target chemical CAS RN in order to retrieve the correct structure.
- In the second module, you have profiled the target chemical.
- In the third module, you have found that no experimental data is currently available in the Toolbox for this structure.
- In other words, you have identified a data gap which you would like to fill.
- **Click** on “Category Definition” to move to the next module.

# 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.

# Category Definition

## Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across.
- Detailed information about grouping chemicals could be found at the following link (Chapter 4).  
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- For example, starting from a target chemical for which a specific protein binding mechanism is identified, analogues can be found which can bind to the proteins by the same mechanism and for which experimental results are available.

## Category Definition

### Protein binding by OASIS grouping method

- This is one of the best grouping methods in the Toolbox. It is built on conventional organic chemical reactions and as such is qualitative in character.
- This method is particularly relevant for respiratory and skin sensitization and acute aquatic toxicity, but also for chromosomal aberration and acute inhalation toxicity.

# Category Definition

## Background to Protein binding by OASIS categorization

- This scheme includes 110 categories organized in three level of information:
  - ✓ Level I: Mechanistic Domains
  - ✓ Level II: Mechanistic alerts associated to each mechanistic domain are created on the basis of a common reactive centre being activated by a number of
  - ✓ Level III: A number of structural alerts specifying the substituents to a common reactive centre are made up for each mechanistic alert

# Category Definition

## Background to Protein binding by OASIS categorization

- Each category from level III is presented by defined 2-dimensional structural alerts that is responsible for the eliciting toxic effects, such as skin sensitization which are a result of protein binding.
- The associated chemical reactions are in accordance with existing knowledge on electrophilic interaction mechanisms of various structural functionalities.

# Category Definition

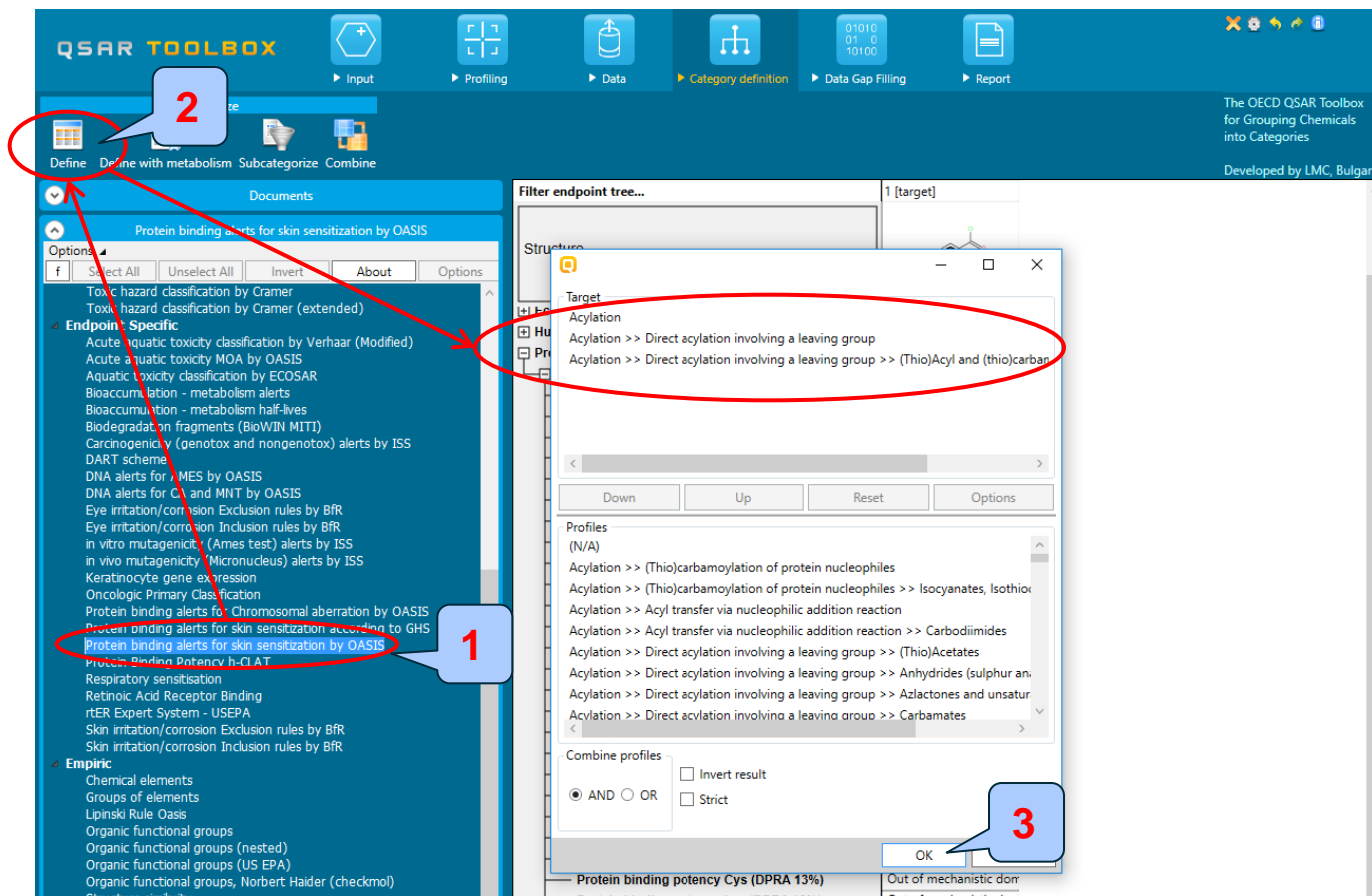
## Background to Protein binding by OASIS categorization

- There is an agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, chemical reactions by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents. So you have mechanistic plausibility for defining your category based on similar protein-binding mechanism.



# Category Definition

## Defining Protein binding alerts for Skin sensitization by OASIS



1. **Highlight** the "Protein binding alerts for skin sensitization by OASIS";
2. **Click** Define;
3. **Click** OK to confirm the defined categories for the target chemical

# Category Definition

## Defining Protein binding by OASIS

The screenshot shows the QSAR Toolbox interface with the 'Category definition' step active. A 'Grouping results' dialog box is open, displaying a list of 9 chemicals found. A blue callout box with the number '1' points to the 'OK' button in the dialog. The background shows a list of endpoints and a table of chemical structures and their properties.

**1. Click OK to confirm the category**

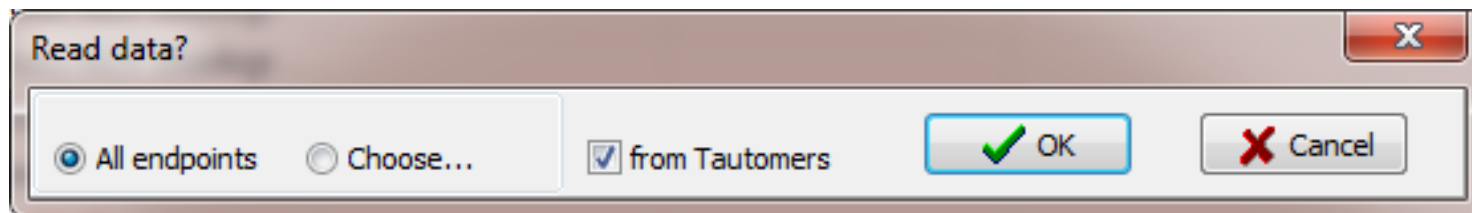
## Category Definition Analogues

- The data is automatically collected.
- Based on the defined category (Acylation<AND>Acylation >> Direct acylation involving a leaving group<AND>Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides) 8 analogues have been identified
- In other words, these 9 compounds along with the target chemical form a category, which can be used for data filling. (see next slide)

## Category Definition

### Read data for Analogues

- The Toolbox automatically requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



# Category Definition

## Summary information for Analogues

- The experimental results for the analogues are inserted into the matrix

The screenshot displays the QSAR Toolbox software interface during the 'Category Definition' process. The main workspace shows a data matrix with columns numbered 1 to 6. Column 1 contains chemical structures. A red box highlights a row in the matrix with the following data: 'AW SW AO (8/8)', 'M: Positive', 'M: Positive', 'M: Positive', 'M: Positive', and 'M: Pt'. A callout box with a question mark icon points to the first cell of this row. A blue text box at the bottom left explains that the callout represents chemical statistics.

Chemical statistics presenting the number of chemicals and the available experimental data.

# Category Definition

## Side bar of experimental data

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Category definition' and contains a 'Filter endpoint tree...' section with a 'Structure' dropdown menu. The 'Data points' dropdown menu is open, showing a table of experimental data. The table has the following structure:

Datapoints	#	Value	Original value	Assay
Human Health Hazards;Sensitisation	1	M: Positive (Skin sensitisation II (ECETOC))	Strongly positive (Skin sensitisation I (Oasis))	LLNA

Callout '1' points to the 'Value' column of the first row. Callout '2' points to the 'X' button in the title bar of the 'Data points' dropdown window.

- 1. Double-click** on the cell with measured data to see detailed information;
- 2. Click** on the X to close the dropdown box.

## Category Definition

### Navigation through the endpoint tree

- The user can navigate through the data tree by closing or opening the nodes of the endpoint tree.
- **Click** on the plus sign next to **Human Health Hazards** then **Sensitisation**, followed by **Skin**, **In Vivo** and **LLNA** and finally **EC3**.
- Local lymph node assay is *in vivo* method for assessment of relative skin sensitization potential of chemicals. The potential is expressed as EC3 values.
- In this example, results from skin sensitisation testing for chemicals reacting via nucleophilic substitution of acyl halides are available (see next screen shot).





# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - **Data Gap Filling**

## Recap

- You have identified a mechanistic category (Acylation<AND>Acylation >> Direct acylation involving a leaving group<AND>Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides ) for the target chemical (4-nitrobenzoyl chloride).
- You have now retrieved in the available experimental results on skin sensitisation (EC3) values for eight chemicals with the same mechanism of protein binding as the target compound, which were found in the "Skin Sensitisation" databases.
- The user can now proceed to the next module; click on "Data Gap Filling".

# Data Gap Filling Overview

- “Data Gap Filling” module gives access to three different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
- 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.
  - (Q)SAR models can be used to fill a data gap if no adequate analogues are found for a target chemical.
- In this example, we use read-across.

# Data Gap Filling

## Apply Read across

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling workflow. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The 'Data Gap Filling' button is circled in red and labeled with a '2'. Below the navigation bar, the 'Read across' option is selected. The main workspace is divided into a left sidebar with a 'Filter endpoint tree...' and a central data table. The 'Filter endpoint tree...' shows a hierarchical list of endpoints, with 'EC3' selected. The data table has columns for target chemicals (1-6) and rows for various endpoints. The 'Structure' row shows chemical structures for targets 1-6. The 'EC3' row is circled in red and labeled with a '1', showing '(8/8)' and 'M: Positive' for target 1. A red arrow points from the 'Read across' button to this cell. The bottom of the interface shows 'Data Gap Filling Settings' with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant', and a section for 'At this position:' with a dropdown menu.

**1. Click** on the cell corresponding to “EC3” for the target chemical; **2. Select** Read-across

# Data Gap Filling

## Scale definition

- Skin sensitisation is a “qualitative” endpoint for which the results are presented with categorical data (for example: positive; negative; weak sensitizer; strong sensitizer, etc).
- Skin sensitisation potential of the chemicals came from different authors coded with different names (for example: data from John Moores University of Liverpool are: *Strongly sensitizing, Moderately sensitizing etc.*; data from European centre for Ecotoxicology and Toxicology of chemicals are: *Positive, Negative, and Equivocal*).
- The main purpose of the scales is to unify all data available in the Toolbox databases for a certain endpoint.
- The default scale for Skin Sensitisation is “Skin Sensitisation ECETOC”. It converts all skin data into: Positive and Negative.

# Data Gap Filling Scale definition

**Possible data inconsistency** [Close]

**Native scale/unit**

- Skin sensitisation I (Oasis) (2 data; 2 chemicals)
- Skin sensitization EC3(ratio) (6 data; 6 chemicals)

Gap filling scale/unit

- Skin sensitisation I (Oasis)
- Skin sensitisation II (ECETOC)
- Skin Sensitization (Danish EPA)
- Skin sensitization EC3(ratio)

converted data

2 from scale Skin sensitisation I (Oasis)  
6 from scale Skin sensitization EC3(ratio)

Data 8/8; Chemicals 8/8

OK Cancel

1

**1. Click OK**

## Data Gap Filling Read-across

- The resulting plot is experimental results of all analogues (Y axis) according to a descriptor (X axis) with the default descriptor of log Kow (see next screen shot).
- The **RED** dot represents predicted results for the target chemical.
- The **BROWN** dots represent the experimental results available for the analogues that are used for the read-across.
- The **BLUE** dot represent the experimental results available for the analogues but not used for read-across.





## Data Gap Filling

### Interpreting Read-across

- In this example, all results of the analogues are consistent; all the analogues are Skin sensitizers.
- The same positive sensitising potential is therefore predicted for the target chemical.
- Accept the prediction by **clicking** “Accept prediction” (see next screen shot).



## Recap

- The read-across is the appropriate data-gap filling method for “qualitative” endpoints like skin sensitisation. Since all the tested chemicals in the category were positive, it was easy to accept the positive predictions for the target chemical.
- You are now ready to complete the final module and to download the report.
- **Click** on “Report” to proceed to the last module.

# Outlook

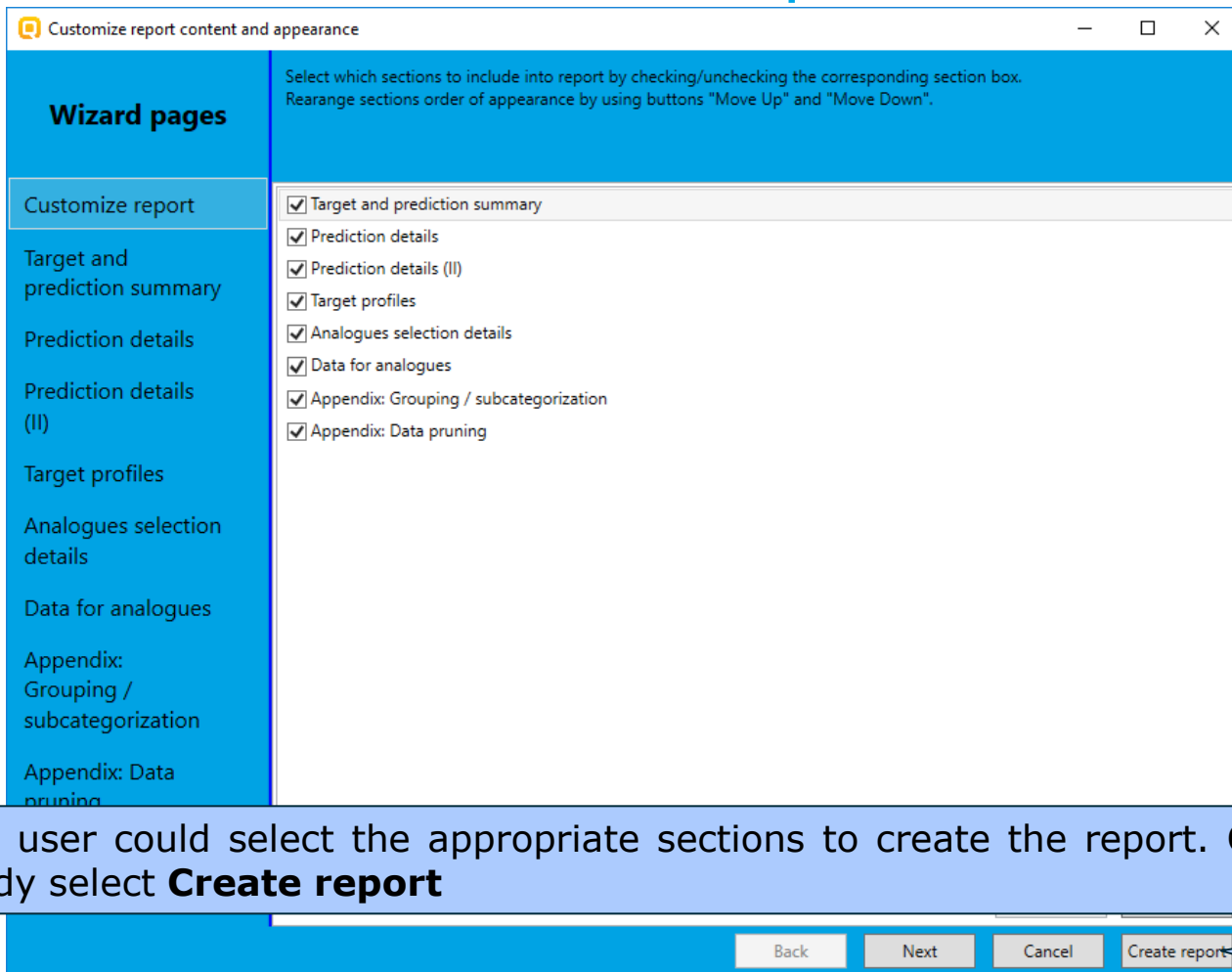
- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Data
  - Category definition
  - Data Gap Filling
  - **Report**

# Report Overview

- The report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report templates as well as a template editor with which users can develop a user defined templates.
- The report can then be printed or saved in different formats.



# Report Generation report



The user could select the appropriate sections to create the report. Once ready select **Create report**

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- Workflow
- **Save prediction**



## Saving the prediction result

- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The 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/Loading the file with TB prediction is shown on next screenshots

# Saving the prediction result

**1. Go to Input section 2. Click on Save button; Define name of the file; 3. Click Save button**

# Open saved file

1

2

3

4

Once the file has been saved **1. Go to Input; 2. Click Open; 3. Find and select file; 4. Click Open**

# Congratulations

- You have now been introduced to the workflow of the Toolbox and completed the tutorial on data gap filling by read-across based on an analogue approach.
- You have been introduced to the six modules of the Toolbox, the basic functionalities within each module and the rationale behind each module.
- Note proficiency comes with practice.