

## OECD QSAR Toolbox v.3.4

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
  - Endpoints
  - 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

#### **B.** Group of chemicals

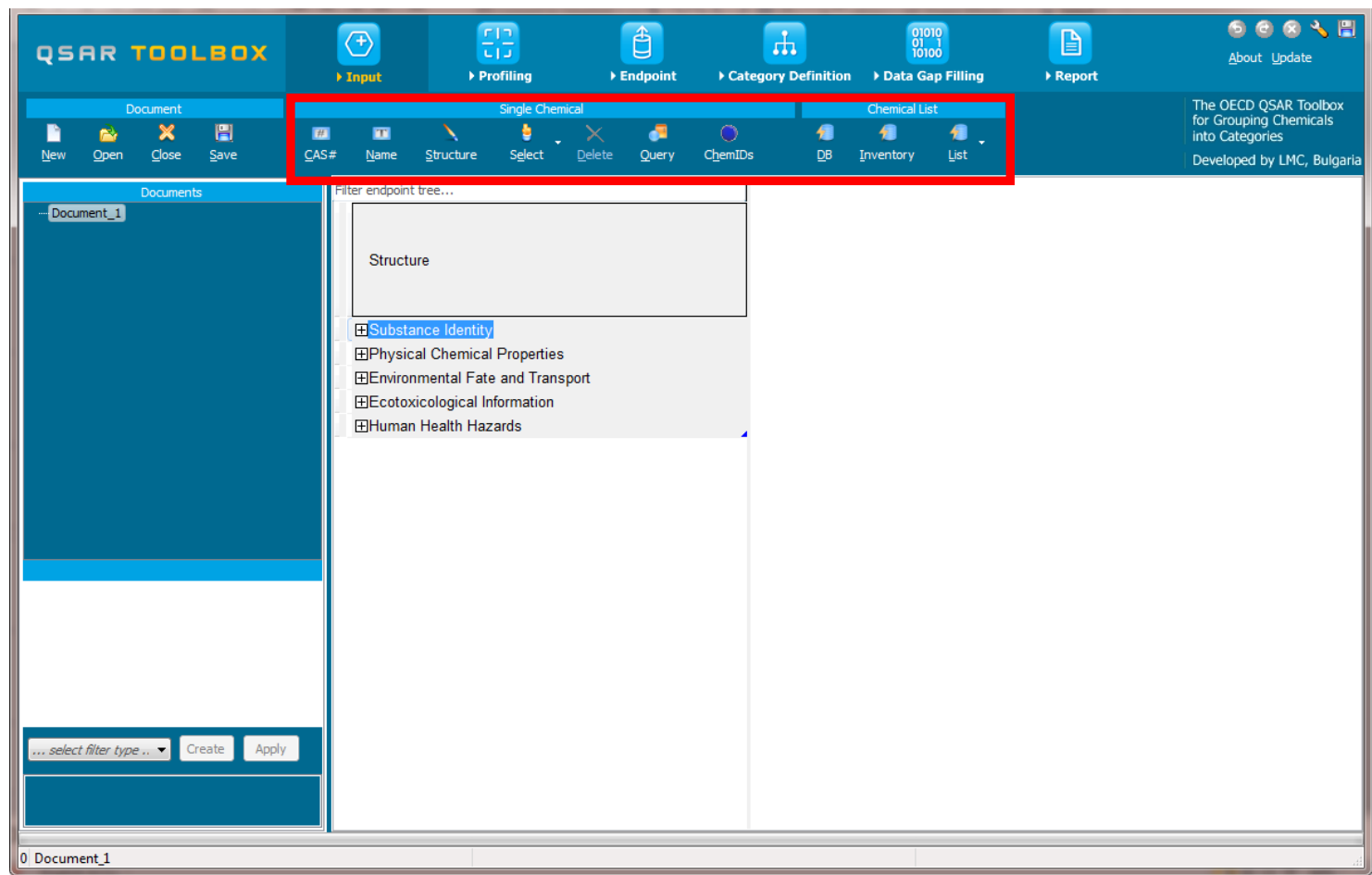
- User List/Inventory
- Specialized Databases

## 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

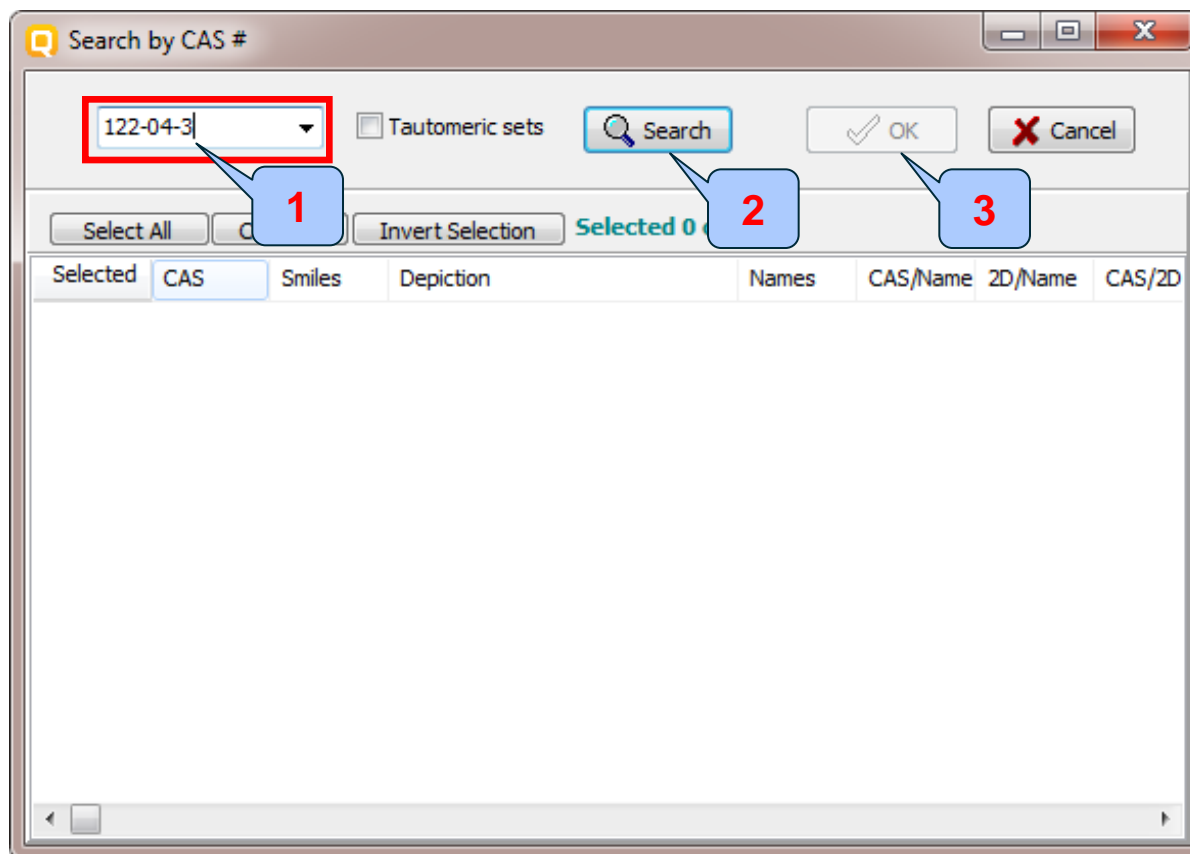
## Input target chemical by CAS#

The screenshot displays the QSAR Toolbox software interface. At the top, the 'QSAR TOOLBOX' logo is visible. The main navigation bar includes tabs for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below this, a sub-menu for 'Document' is open, showing options for 'New', 'Open', 'Close', 'Save', and 'CAS#'. The 'CAS#' option is highlighted with a red box, and a callout bubble with the number '1' points to it. The 'Single Chemical' sub-menu is also visible, showing options for 'Name', 'Structure', 'Select', 'Delete', 'Query', and 'ChemIDs'. The 'Chemical List' sub-menu is also visible, showing options for 'DB', 'Inventory', and 'List'. The main workspace is divided into two panels: 'Documents' on the left and 'Filter endpoint tree...' on the right. The 'Filter endpoint tree...' panel is currently empty.

1. **Click** on CAS#

# Chemical Input Screen

## Enter CAS# of 4-nitrobenzoyl chloride



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

# Chemical Input

## Target chemical identity

- **Double click** "Substance Identity"; this displays the chemical identification information.
- Note that existing in the Toolbox name of target chemical are in different colours (see next screen shot).
- 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 interface with the following components:

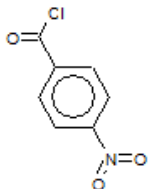
- Top Bar:** QSAR TOOLBOX logo and navigation icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Document Bar:** Document (New, Open, Close, Save), Single Chemical (CAS#, Name, Structure, Select, Delete, Query, ChemIDs), and Chemical List (DB, Inventory, List).
- Documents Panel:** Shows a tree view with 'Document\_1' containing 'CAS: 122-04-3'. The SMILES string O=C(Cl)c1ccc(N(=O)=O)cc1 is visible at the bottom.
- Filter endpoint tree...:** A tree view showing 'Structure' and 'Substance Identity' (expanded). Under 'Substance Identity', the following fields are listed: CAS Number, Chemical IDs, Chemical Name, Molecular Formula, and Structural Formula. Other categories like Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards are collapsed.
- Target Information Panel:** Displays '1 [target]' with a chemical structure of p-nitrobenzoyl chloride. A red circle highlights the following text:
  - 122-04-3
  - EINECS:2045174
  - p-nitrobenzoyl chlo...
  - benzoyl chloride, 4...
  - 4-nitrobenzoyl chlo...
  - benzoyl chloride, p...
  - C7H4ClNO3
  - O=C(Cl)c1ccc(N(=...

# Chemical Input

## Target chemical identity

The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-dimensional depiction.

Explain QA Form

CAS/2D	Names	CAS/Name	2D/Name	CAS/2D	Status
<chem>O=C(Cl)c1ccc[N(=O)=O]cc1</chem> CAS: 122043 	1: p-nitrobenzoyl chlor 2: benzoyl chloride, 4- 3: 4-nitrobenzoyl chlor 4: benzoyl chloride, p-	1:: High Quality 1:: Bacterial mutagenicity 2:: DSSTOX 3:: Genotoxicity OASIS 2:: High Quality 1:: Canada DSL 2:: METI Japan 3:: NICNAS 4:: Phys-chem EPISUITE 5:: TSCA 6:: US HPV Challenge Pro 3:: High Quality 1:: ECHA PR 2:: EINECS 3:: METI Japan 4:: REACH ECB	1:: High Quality 1:: DSSTOX 2:: Genotoxicity OASIS 3:: Bacterial mutagenicity : High Quality 2:: High Quality A 1:: TSCA 2:: NICNAS 3:: Phys-chem EPISUITE A 4:: US HPV Challenge Pro 5:: Canada DSL 6:: METI Japan 3:: High Quality A 1:: ECHA PR 2:: REACH ECB 3:: EINECS 4:: US HPV Challenge Pro	: High Quality 1:: Bacterial mutageni 2:: Canada DSL 3:: DSSTOX 4:: ECHA PR 5:: EINECS 6:: Genotoxicity OASIS 7:: METI Japan 8:: NICNAS 9:: Phys-chem EPISUIT 10:: REACH ECB 11:: TSCA 12:: US HPV Challenge	Base Structure



In case a structure has several CAS numbers or a structure could be related to more than one substance, 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.



# Chemical Input

## Chemical identity

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 applied 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 colour is applied 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 colour is applied if the identifier is allocated to different structures in different databases.

# 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"

**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**

**Mechanistic Domain:** Acylation

**Mechanistic Alert:** Acyl transfer via nucleophilic addition reaction

**Structural Alert:** Carbodiimides

*(The transformation is not confirmed by 3<sup>rd</sup> party Expert vet)*

The chemical is a strong sensitizer as a result of **Protein addition to carbodiimide:**

$$\begin{array}{c} \text{---C---} \\ | \\ \text{---N=C---} \end{array} \xrightarrow{\text{Pr-ZH}} \begin{array}{c} \text{---C---} \\ | \\ \text{---NH---C---} \\ | \\ \text{Z---Pr} \end{array}$$

Z = -S, -NH, -O, -COO

This mechanism could explain the observed skin sensitization potential of dicyclohexylcarbodiimide and diisopropylcarbodiimide assessed by the Mouse Ear Swelling Test (MEST) and the murine Local Lymph Node Assay (LLNA). These chemicals were identified as both irritants and contact sensitizers.

**References:**

... P.C., Griffey S.S., Meade B.J., Contact hypersensitivity to dicyclohexylcarbodiimide and diisopropylcarbodiimide in female B6C3F1 ... Toxicol., 1998, 21(2), 195-206.

# 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 v.1.4 – mechanistic grouping
  - Protein binding alerts for skin sensitization by OASIS v1.4 – 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

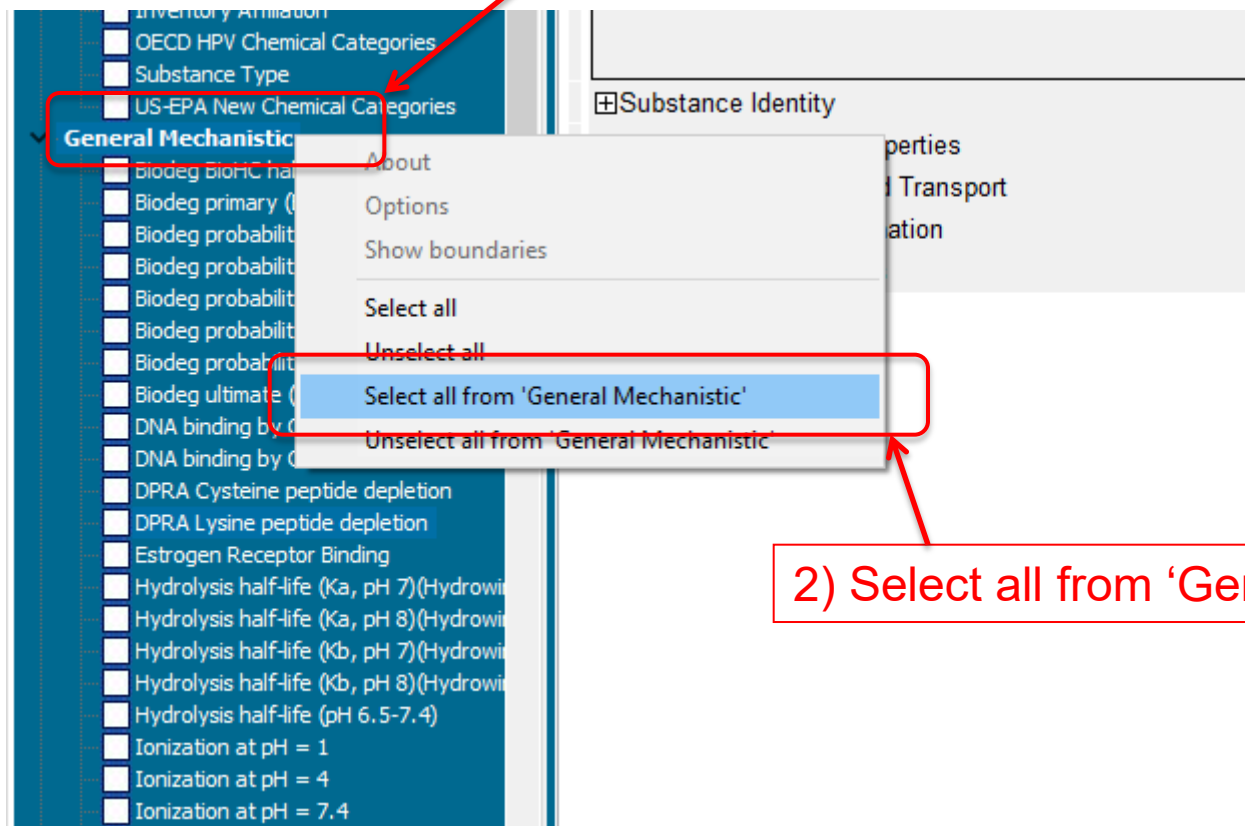
### Profiling the target chemical: Example

- **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

1) Right click on General mechanistic



2) Select all from 'General Mechanistic'

# Profiling

## Profiling the target chemical

The screenshot shows the QSAR Toolbox Profiling interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing 'Apply', 'New', 'View', and 'Delete'. The 'Apply' button is circled in red and labeled with a '1'. Below the menu, the 'Profiling methods' section is visible, with a context menu open over the 'General Mechanistic' section. The context menu options include 'About', 'Options', 'Show boundaries', 'Select all', 'Unselect all', 'Select all from 'General Mechanistic'', and 'Unselect all from 'General Mechanistic''. The 'Select all from 'General Mechanistic'' option is highlighted. The main window displays the 'Structure' of the target chemical, a chemical structure diagram, and a list of properties including '122-04-3', 'EINECS:2045174', 'p-nitrobenzoyl chlor...', 'benzoyl chloride, 4...', '4-nitrobenzoyl chlor...', 'benzoyl chloride, p...', 'C7H4ClNO3', and 'O=C(Cl)c1ccc(N(=...'. A blue callout box at the bottom contains the text '1. 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 Profiling interface. On the left, the 'Profilng methods' list includes various endpoints, with 'Protein binding by OASIS v1.4' selected. The 'Filter endpoint tree...' window shows a list of endpoints, with 'Protein binding by OASIS v1.4' highlighted in a red box. An arrow points from this box to the corresponding result in the table on the right, which is also highlighted in a red oval. The result shows 'Acylation' and 'Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides'.

1. Go to Protein binding by OASIS v1.4 to review the profiling results.

# Outlook

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

# Endpoint Overview

- “Endpoint” 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).

# Endpoint

## 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 “Endpoint” 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).



# Endpoint Gather data

The screenshot displays the QSAR Toolbox interface with the following elements:

- Top Bar:** Navigation icons for Input, Profiling, Endpoint (highlighted), Category Definition, Data Gap Filling, and Report.
- Toolbar:** Action buttons including Gather (circled in red and labeled '3'), Import, Export, Delete, and Tautomerize.
- Left Pane (Data):** A list of endpoints. The 'Human Health Hazards' section is expanded (labeled '1'). Under this section, 'Skin Sensitization' and 'Skin sensitization ECETOC' are checked (labeled '2').
- Right Pane:** Shows a 'Structure' field with a chemical structure of 4-chlorobenzamide. Below it is a tree view of endpoints under 'Human Health Hazards' and 'Profile'. A table on the right lists values for these endpoints, such as 'Not calculated', 'days - weeks', and 'weeks - months'.

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

## Endpoint 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).

# Endpoint Gather data

The screenshot shows the QSAR Toolbox 3.4.0.17 interface. The 'Endpoint' menu is highlighted in the top navigation bar. Below it, the 'Gather' option is selected. An information dialog box is displayed in the center, with the message: "There are no experimental data available for the chemicals of interest." An arrow points to the 'OK' button in the dialog box. The background shows a list of databases and inventories on the left, and a large empty area for the endpoint tree on the right.

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

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - **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 v.1.4 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 v.1.4 categorization

- This scheme includes 146 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 by OASIS v.1.4

The screenshot displays the QSAR Toolbox software interface during the 'Define' step. On the left, a tree view under 'Pre defined' shows 'Protein binding by OASIS v.1.4' highlighted with a red circle and a callout '1'. The top toolbar has the 'Define' button circled in red with a callout '2'. A dialog box titled 'Protein binding by OASIS v.1.4' is open, showing a list of 'Target(s) profiles'. The profile 'Acylation >> Direct acylation involving a leaving group' is selected and circled in red with a callout '3'. The 'OK' button in the dialog is also circled in red with a callout '3'. The main window shows a 'Structure' field with a chemical structure of a target molecule and a list of categories under 'General Mechanistic'.

**1. Highlight** the “Protein binding by OASIS v.1.4”; **2. Click** Define; **3. Click** OK to confirm the defined categories for the target chemical

# Category Definition

## Defining Protein binding by OASIS

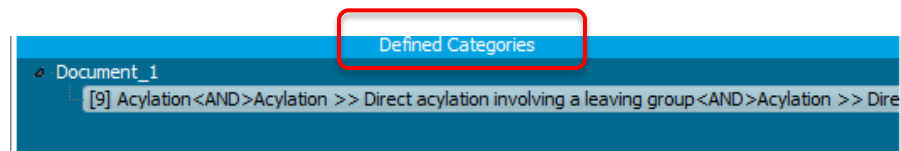
The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', and 'Data Gap Filling'. The 'Category Definition' tab is active, showing a 'Filter endpoint tree...' window with a tree structure. A dialog box titled 'Define category name' is open, with the text 'Es and cyanides (Protein binding by OASIS v1.4)' entered in the 'Category name (9 chemicals)' field. The 'OK' button is highlighted with a red callout box containing the number '1'.

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

# Category Definition

## Analogues

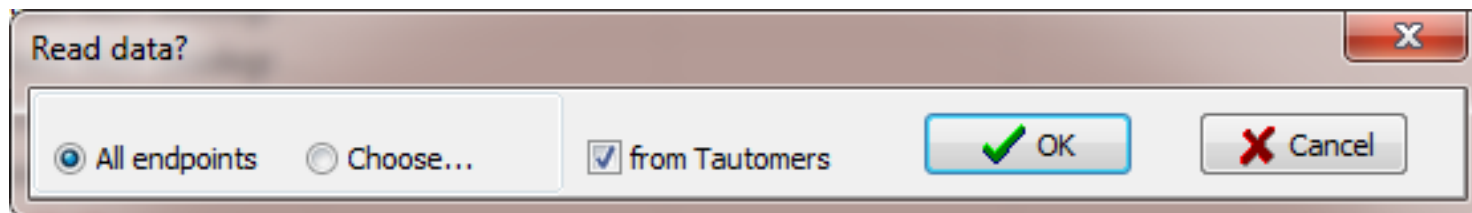
- The data is automatically collected.
- Based on the defined category (Acylation<AND>Acylation >> Direct acylation involving a leaving group<AND>Acylation >> Direct acylation involving a leaving group >> (Thio)Acyl and (thio)carbamoyl halides and cyanides) 8 analogues have been identified
- The name of the category appears in the "Defined Categories" window, indicating the number of substances belonging to the category.
- 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

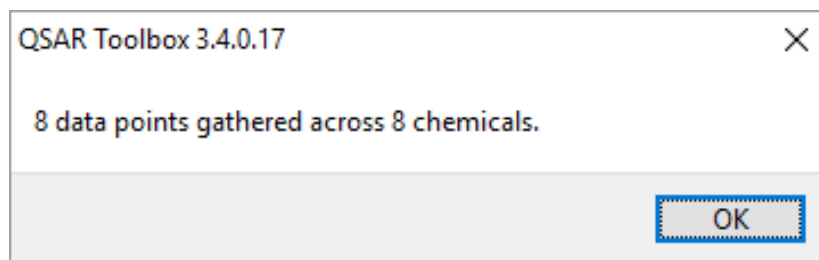
- The Toolbox automatically requests the user to select the endpoint that should be retrieved.
- The user can either select the specific endpoint or by default choose to retrieve data on all endpoints (see below).



# Category Definition

## Read data for Analogues

- Toolbox automatically informs the user for the number of gathered data points across the chemicals in the category




- Click OK to confirm the appeared message



# Category Definition

## Summary information for Analogues

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

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

# Category Definition

## Side bar of experimental data

The screenshot displays the QSAR Toolbox software interface. The main window shows a 'Data points' dropdown menu with a table of experimental data. A callout box labeled '1' points to a cell in the table containing the value 'M: Positive'. Another callout box labeled '2' points to the 'X' button in the top right corner of the dropdown menu.

#	Endpoint	Value	Original value	Organ	Reference source	Phylum (common name)	Phylum	Type of method	Year	Test method / Data	Test organisms (species)	Kingdom	QA (CAS-2D)	Assay	Assigned SMILES	Database name
1	EC3	Positive (Skin sensitisation II (ECETOC))	0.23 % (Skin sensitization EC3 (ratio))	Skin	Unilever	Vertebrates	Chordata	in Vivo	2005	LLNA	mouse	Animalia	High Quality	LLNA	NO	Skin Sensitization

The 'Inventories' sidebar on the left shows a tree structure with 'Skin Sensitization' selected. Below it, a table shows the results of the selection:

(8/8)	M: Positive	M: Positive	M: Positive	M: Positive	M: Positive	M: Positive

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).

# Category Definition

## Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar contains 'Databases' and 'Inventories' sections. The main workspace displays a tree view of endpoints under 'Filter endpoint tree...'. The 'EC3' endpoint is highlighted with a red box and a callout bubble containing the number '1'. The data table below the tree view shows the following information:

Structure	1 [target]	2	3	4	5	6	7
Structure							
Ecotoxicological Information							
Human Health Hazards							
Acute Toxicity							
Bioaccumulation							
Carcinogenicity							
Developmental Toxicity / Teratogenicity							
Genetic Toxicity							
Immunotoxicity							
Irritation / Corrosion							
Neurotoxicity							
Photoinduced Toxicity							
Repeated Dose Toxicity							
Sensitisation							
Skin							
In Chemico							
In Vitro							
In Vivo							
GPMT							
LNA							
EC3	(8/8)	M: Positive	M: Positive	M: Positive	M: Positive	M: Positive	M: Positive
Undefined Assay							
ToxCast							
Toxicity to Reproduction							

1. This is the target endpoint

# Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Chemical Input
  - Profiling
  - Endpoint
  - 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 shows the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar has a 'Filter endpoint tree...' and a 'Target Endpoint' section. The main workspace displays a grid of chemical structures and their corresponding data points. A red arrow points from the 'Read-across' button in the sidebar to a cell in the grid containing '(8/8)'. Callout boxes with numbers 1, 2, and 3 highlight the 'Read-across' button, the 'Apply' button, and the 'Data Gap Filling' menu item, respectively.

**1. Click** on the cell corresponding to "EC3" for the target chemical; **2. Select** Read-across ; **3. Click** Apply

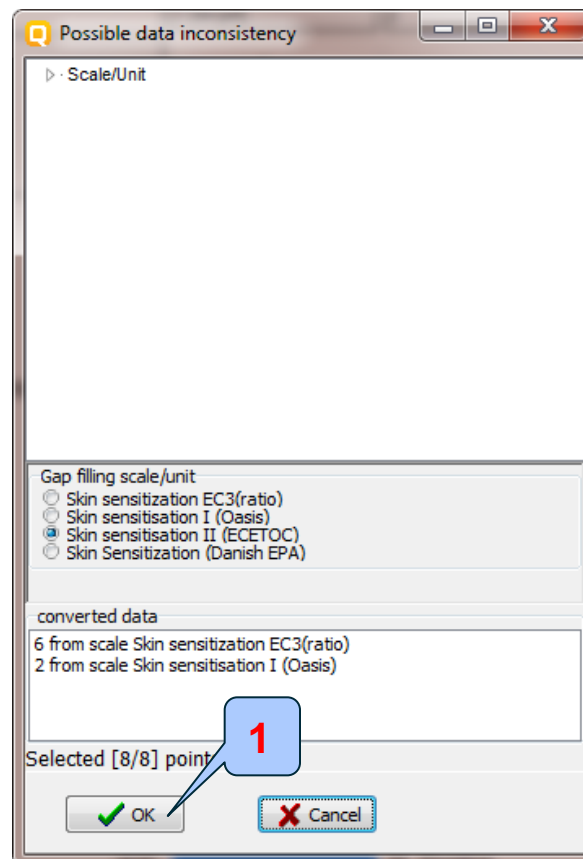


# Data Gap Filling

## 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



**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 Read-across

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

About Update

The OECD QSAR Toolbox for Grouping Chemicals into Categories  
Developed by LMC, Bulgaria

Filling

Apply

Data Gap Filling Method

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

Target Endpoint

Human Health Hazards Sensitisation Skin In Vivo LLNA EC3

Structure

1 [target] 2 3 4

Descriptors Prediction

Read across prediction of EC3, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: N/A, Predicted target value: 'Positive'

EC3 (obs.)

log Kow

Descriptor X: log Kow

Accept prediction

Return to matrix

- Select/filter data
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

9 Acylation<AND>Acylation >> Direct acylation involving a leaving group Create prediction by gap filling 0/1

## 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).

# Data Gap Filling

## Accepting the predicted result

The screenshot shows the QSAR Toolbox interface during a data gap filling operation. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Data Gap Filling Method' with options: Read-across, Trend analysis, and (Q)SAR models. The 'Target Endpoint' is set to 'Human Health Hazards Sensitisation Skin In Vivo LLNA EC3'. The main workspace displays a 'Structure' field with chemical structures, a 'Prediction' tab, and a scatter plot for 'log Kow' vs 'EC3 (obs.)'. An information dialog box is open with the message 'The current prediction was accepted'. A callout box with '1' points to the 'Accept prediction' button, and another callout box with '2' points to the 'OK' button in the dialog. A text box at the bottom left contains the instructions: '1. Click Accept prediction' and '2. Click OK'.

# Data Gap Filling

## Accepting the predicted result

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

The OECD QSAR Toolbox for Grouping Chemicals into Categories  
Developed by LMC, Bulgaria

Structure

1 [target] 2 3 4

Descriptors Prediction

Read across prediction of EC3,  
taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals,  
Observed target value: N/A, Predicted target value: 'Positive'

Positive

EC3 (obs.)

Negative

1.00 2.00 3.00 4.00 5.00 6.00 7.00

log Kow

Accept prediction  
Return to matrix

- Select/filter data
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

9 Acylation<AND>Acylation >> Direct acylation involving a leaving group Create prediction by gap filling 0/100

**1. Click Return to matrix**

## 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
  - Endpoint
  - 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 screenshot displays the QSAR Toolbox software interface. At the top, the 'Report' icon is highlighted with a callout box labeled '1'. Below the main menu, the 'Create' button in the toolbar is circled and labeled with a callout box '3'. The left-hand pane, titled 'Available data to report', shows a tree view where the 'Predictions' folder is expanded, and a specific prediction item is selected, indicated by callout box '2'. The right-hand pane is currently empty, showing the text 'Please select a prediction, (Q...'. The interface also includes a top navigation bar with icons for 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report', and a secondary toolbar with buttons for 'Print', 'Close', 'Save as', 'Register', 'Unregister', 'Update', 'Clone', and 'Design'.

1. **Go** to the Report section; 2. **Expand** Prediction in the "Available data to report" window; 3. **Click** Create

# Report Generation report

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Reports' and 'Repository' tabs. The 'Reports' tab is active, showing a list of actions: Create, Print, Close, Save as, Register, Unregister, Update, Clone, and Design. The main workspace is divided into three panes:

- Left Pane:** 'Available data to report' showing a tree view with 'Predictions' (containing '[1] 17.06.2016 10:28 [R]: Positive; Estimation for EC3 for CAS 122-04-3; Domain: In domain'), '(Q)SARs', and 'Categories'. Below this is 'Available report templates' with 'Standard (predefined)' (containing 'QSAR Toolbox Prediction Report (TPRF v.3.4)') and 'Custom (user defined)' (containing 'Editable copy of QSAR Toolbox Prediction Report (TPRF v.3.4)').
- Middle Pane:** 'Prediction [1]' showing a large empty area.
- Right Pane:** 'Prediction of EC3 for p-nitrobenzoyl chloride' with a page number '1 / 2'. The main content area displays the title 'QSAR Toolbox prediction for single chemical' and a footer note: 'The template of the current report is based on "GUIDANCE DOCUMENT ON THE'.

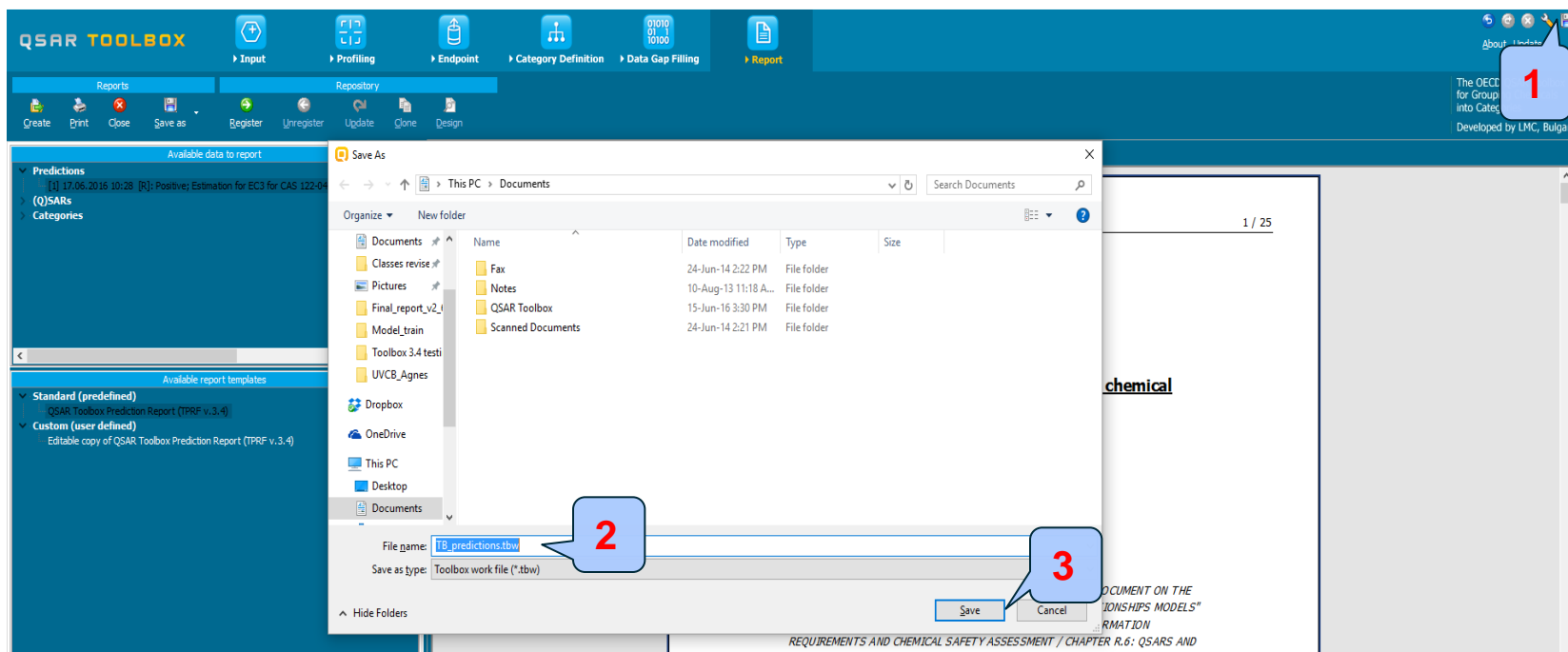
# Outlook

- 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. Click** on Save button;  
**3. Click** Save button

**2. Define** name of the file;

# Open saved file

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below the menu bar, there are tabs for 'Chemical' and 'Chemical List'. A 'Select file' dialog box is open, showing a file list with columns for Name, Date modified, Type, and Size. The file 'TB\_predictions.tbw' is selected. The 'Open' button is highlighted. Numbered callouts (1-4) indicate the steps: 1. Go to Input; 2. Click Open; 3. Find and select file; 4. Click Open.

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.