

OECD (Q)SAR Toolbox v.4.4.1

Manipulation of data matrix and manual transferring of data to the target outside of the data gap filling module

Outlook

- **Background**
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example

Background

- This is a step-by-step presentation designed to introduce the user to the newly created functionalities for manipulation of the data matrix.
- A simple example illustrating how to manually transfer a read-across prediction from source (analogue) to target chemical.

Outlook

- Background
- **Keywords**
- Objectives
- Specific aim
- Manipulation of data matrix
- Example

Keywords

TARGET CHEMICAL - chemical of interest

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**
- Specific aim
- Manipulation of data matrix
- Example

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Rearrangement of columns of the data matrix;
- Filtering of parameters/experimental data or profiling results of the analogues within the category;
- Hide/show gap-filling chart while in Data gap filling module;
- Transferring of data to the target chemical outside Data gap filling module.

Outlook

- Background
- Keywords
- Objectives
- **Specific aim**
- Manipulation of data matrix
- Example

Aim

- To introduce and make the user familiar with:
 - Manipulation of the data matrix;
 - Filtering the data matrix with respect to parameters, experimental data that appear for the selected analogues;
 - Hide/Show the gap filling chart while the user is in the Gap filling module for better screening and analysis of data on the data matrix;
- Transfer of data to target chemical outside Data gap filling.

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- **Manipulation of data matrix**
- Example

Data matrix Overview

- The data matrix window has three main parts:
 - Area with the Endpoint tree (1)
 - Area with the selected chemicals (2) and
 - Area with data (experimental, predicted) (3)

The screenshot shows the QSAR Toolbox Data Matrix window. The interface is divided into three main sections:

- Area with the Endpoint tree (1):** Located on the left, it shows a hierarchical tree of endpoints. The 'Aquatic Toxicity' section is expanded, showing endpoints like Growth, Growth Inhibition, Immobilisation, Intoxication, Mortality, and LC50. The 'Ecotoxicological Information' section is also expanded, showing endpoints like LOEC and NOEC.
- Area with the selected chemicals (2):** Located at the top, it shows a list of selected chemicals. The chemicals are represented by their chemical structures and names.
- Area with data (experimental, predicted) (3):** Located below the selected chemicals, it shows a table of data. The table has columns for each chemical and rows for each endpoint. The data is presented in a grid format, with experimental values in blue and predicted values in white.

Manipulation of data matrix

Implementation in Toolbox

The functionality allows the user manually to manipulate the matrix via:













- 1) Filtering the parameters and/or experimental data and/or profiling results which appear on the data matrix for the selected analogues;
- 2) Reordering the columns with analogues in the category in order to more effectively analyze the data between the target and analogues;
- 3) Transferring of data (experimental/predicted) from analogues to the target chemical outside the gap filling module;
- 4) Hide/Show the data matrix once the user is in the stage of Data gap filling module;

Illustration of the functionalities are shown on the next few slides.

Manipulation of data matrix

Implementation in Toolbox

1) Filtering the data matrix

Filter endpoint tree...	1 [target]	2	3	4	5	6	7	8	9	10	11	12
Structure												
Structure info												
Parameters												
2D												
(Q) Acidic pKa (Chemaxon)	15.6	No value	No value	16.9	15.6	16.6	17.7	15.6				
(Q) Basic pKa (Chemaxon)	-6.94	-4.11	-7.1	-7.02	-6.94	-4.09	-4.2	-6.94				
Acidic pKa (OASIS Consensus)	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	7.67
Acidic pKa (OASIS Electric)	10.3	7.28	9.57	11.2	10.2	10.5	10.3	10.2	7.33	8.49	10.9	5.63
Acidic pKa (OASIS Regression)	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	8.16
Amino acids pKa (OASIS Regression)	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value
BAF	0.81 log(L/kg)	0.05 log(L/kg)	2.09 log(L/kg)	0.76 log(L/kg)	2.6 log(L/kg)	0.1 log(L/kg)	1.52 log(L/kg)	2.36 log(L/kg)	0.03 log(L/kg)	0.91 log(L/kg)	2.93 log(L/kg)	0.16 log(L/kg)
BAF (lower trophic)	0.637 log(L/kg)	0.034 log(L/kg)	1.91 log(L/kg)	0.593 log(L/kg)	2.67 log(L/kg)	0.066 log(L/kg)	1.33 log(L/kg)	2.34 log(L/kg)	0.022 log(L/kg)	0.717 log(L/kg)	2.95 log(L/kg)	0.2 log(L/kg)
BAF (mid trophic)	0.677 log(L/kg)	0.037 log(L/kg)	1.95 log(L/kg)	0.632 log(L/kg)	2.65 log(L/kg)	0.073 log(L/kg)	1.38 log(L/kg)	2.35 log(L/kg)	0.024 log(L/kg)	0.761 log(L/kg)	2.94 log(L/kg)	0.195 log(L/kg)
BAF (upper trophic)	0.807 log(L/kg)	0.047 log(L/kg)	2.09 log(L/kg)	0.757 log(L/kg)	2.6 log(L/kg)	0.099 log(L/kg)	1.52 log(L/kg)	2.36 log(L/kg)	0.026 log(L/kg)	0.905 log(L/kg)	2.93 log(L/kg)	0.161 log(L/kg)
BAF (upper trophic, biotransformation rate is zero)	0.873 log(L/kg)	0.069 log(L/kg)	2.27 log(L/kg)	0.825 log(L/kg)	3.77 log(L/kg)	0.122 log(L/kg)	1.63 log(L/kg)	3.03 log(L/kg)	0.05 log(L/kg)	0.956 log(L/kg)	3.95 log(L/kg)	0.612 log(L/kg)
Basic pKa (OASIS Regression)	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value	No value
Ecotoxicological Information												
Aquatic Toxicity												
Growth	105/123	M: 152 mg/L	M: 145 mg/L	M: 31.7 mg/L	M: 112 mg/L	M: 3.9 mg/L	M: 1.5 mg/L	M: 7.96 mg/L	M: 8.2 mg/L	M: 937 mg/L	M: 191 mg/L	M: 77.2 mg/L
Growth Inhibition	12/24											
Immobilisation	15/15											
Intoxication	1/1											
Mortality												
EC50	5/13											
LC50												
12 h	1/1											
24 h	1/1											
48 h	3/5											
96 h												
Animalia (animals)												
Chordata (chordates)												
Actinopterygii (ray-finned fishes, spiny r...												
Leuciscus idus	5/5											
Oryzias latipes	9/9											
Pimephales promelas	68/68	M: 17.8 mg/L	M: 20.5 mg/L	M: 6.62 mg/L							M: 20 mg/L	M: 5.01 mg/L
Poecilia reticulata	19/19	M: 9.79 mg/L	M: 10.5 mg/L		M: 7.77 mg/L				M: 3.19 mg/L			
LOEC	4/5											
NOEC	4/5											
Physiology	21/28				M: 1.66 mg/L	M: 14.6 mg/L	M: 5.64 mg/L					
Reproduction	10/24					</						

Manipulation of data matrix Implementation in Toolbox

1) Filtering the data matrix

The screenshot displays the QSAR Toolbox interface. On the left is the 'Filter endpoint tree...' panel with a tree structure of endpoints. In the center is a data matrix with columns labeled [target], 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. On the right is a 'Select' dialog box with a tree of categories and checkboxes. A red arrow points from the 'Select' button in the dialog to the 'Filter endpoint tree...' panel. A red box highlights the 'OK' button in the dialog. A red text box on the right says 'Filter data matrix by select/unselect the corresponding checkboxes'. Another red text box at the bottom right says 'Click OK button to confirm the selection'.

Button for filtering the endpoint tree

Filter data matrix by select/unselect the corresponding checkboxes

Click OK button to confirm the selection

Manipulation of data matrix Implementation in Toolbox

1) Filtering the data matrix

Data matrix after filtering includes the selected items
(parameters/data/profilers) only

The screenshot displays the QSAR Toolbox interface with a data matrix and a 'Select' dialog box. The data matrix has 9 columns, with the first column labeled '1 [target]'. The 'Parameters' section in the left sidebar is highlighted with a red box. The 'Acute aquatic toxicity classification by...' row in the data matrix is highlighted with a red box. The 'Acute aquatic toxicity classification by Verhaar (Modified)' row in the 'Select' dialog is highlighted with a red box. Red arrows point from these highlights to the corresponding rows in the data matrix.

1 [target]	2	3	4	5	6	7	8	9
Structure	<chem>CCCCCCCC</chem>	<chem>O=Cc1ccccc1</chem>	<chem>CC(=O)O</chem>	<chem>CC(=O)O</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
Structure info								
Parameters								
2D								
BCF	0.84 log(L/kg)	0.5 log(L/kg)	1.76 log(L/kg)	0.81 log(L/kg)	1.1 log(L/kg)	0.5 log(L/kg)	1.36 log(L/kg)	2.15 log(L/kg)
Ecotoxicological Information								
Aquatic Toxicity								
Growth								
IGC50								
48 h								
Protozoa								
Ciliophora								
Ciliata								
Tetrahymena pyriformis	97/97							
Profile								
Endpoint Specific								
Acute aquatic toxicity classification by...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...	Class 3 (unspecif...
Acute aquatic toxicity MOA by OASIS	Aldehydes	Aldehydes	Aldehydes	Aldehydes	Aldehydes	Reactive unspecif...	Aldehydes	Aldehydes
Aquatic toxicity classification by ECOSAR	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)	Aldehydes (Mono)	Vinyl/Allyl Aldehydes	Vinyl/Allyl Aldehydes	Aldehydes (Mono)

The 'Select' dialog box shows the following options:

- ☒ Structure info
- ☒ Parameters
- ☒ 2D
- ☐ 3D
- ☐ Physical Chemical Properties
- ☐ Environmental Fate and Transport
- ☒ Ecotoxicological Information
- ☒ Aquatic Toxicity
- ☒ Growth
- ☐ EC50
- ☒ IGC50
- ☒ 48 h
- ☒ Protozoa
- ☒ Ciliophora
- ☒ Ciliata
- ☒ Tetrahymena pyriformis
- ☐ LOEC
- ☐ NOEC
- ☐ Growth Inhibition
- ☐ Immobilisation
- ☐ Intoxication
- ☐ Mortality
- ☐ Physiology
- ☐ Reproduction
- ☐ Sediment toxicity
- ☐ Terrestrial Toxicity
- ☐ Human Health Hazards
- ☒ Profile
- ☐ Predefined
- ☒ Endpoint Specific
- ☒ Acute aquatic toxicity classification by Verhaar (Modified)
- ☒ Acute aquatic toxicity MOA by OASIS
- ☒ Aquatic toxicity classification by ECOSAR

Manipulation of data matrix Implementation in Toolbox

2) Reordering of the analogues on data matrix

QSAR Toolbox 4.4.1 [Document 1]











The screenshot shows the QSAR Toolbox 4.4.1 interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar contains panels for Documents, Profiling methods, and Metabolism/Transformations. The main area displays a data matrix with columns for different chemical structures and rows for various endpoints. A red dashed box highlights a specific chemical structure in the matrix, and a red arrow points to it, indicating the reordering process.

Left-click on the chemical, hold it and drag it next to the target

Manipulation of data matrix Implementation in Toolbox

3) Transferring of data (exp./predicted) from analogues (metabolites) to the target

Filter endpoint tree...

	Parent chemical...	metabolite #1	metabolite #2	metabolite #3	metabolite #4	metabolite #5	metabolite #6	metabolite #7	metabolite #8	metabolite #9
Structure										
Structure info										
Parameters										
Physical Chemical Properties										
Environmental Fate and Transp										
Ecotoxicological Information										
Human Health Hazards										
Acute Toxicity										
ADME										
Bioaccumulation										
Carcinogenicity										
Developmental Toxicity / Teratogenicity										
Genetic Toxicity										
in Vitro	2/48									
in Vivo										
Micronucleus Assay										
Chromosome aberration	1/5									
Immunotoxicity										
Irritation / Corrosion										
Neurotoxicity										
Photoinduced toxicity										
Repeated Dose Toxicity										
Sensitisation	AW SW AOP									
ToxCast										
Toxicity to Reproduction	AOP									
Toxicokinetics, Metabolism and Distribution										

Methamphetamine (target chemical)

Generated observed *rat in vivo* metabolites of the target chemical

Read-across prediction for the metabolite → **R: Positive**
MS: Positive
MS: Positive
MS: Positive
MS: Positive

Measured data for the metabolite ← MS: Negative
MS: Negative

Manipulation of data matrix Implementation in Toolbox

3) Transferring of data (exp./predicted) from analogues (metabolites) to the target

Methamphetamine (target chemical)

1 Perform right click over the cell with data

2 Select **Transfer to target** from the context menu

3 Select the appropriate scale

4 Confirm by **OK**

5 Select the data to be transferred

6 Click **OK**

7 Selected data is transferred to the target chemical

Possible data inconsistency

Metadata

- Endpoint**
 - ☒ Chromosome aberration (1 chemicals; 5 data)
- Metabolic activation**
 - ☒ No S9 Info (1 chemicals; 2 data)
 - ☒ No S9 Info <OR> Not Applicable (1 chemicals; 1 data)
 - ☒ Undefined Metabolic Activation (1 chemicals; 2 data)
- Native scale/unit**
 - ☒ Chromosome aberration I (Oasis) (1 chemicals; 1 data)
 - ☒ Chromosome aberration V (ECVAM) (1 chemicals; 2 data)
 - ☒ Micronucleus I (1 chemicals; 2 data)
- Strain**
 - ☒ CD-1 (1 chemicals; 1 data)
 - ☒ Undefined Strain (1 chemicals; 4 data)
- Test organisms (species)**
 - ☒ House mouse (1 chemicals; 1 data)
 - ☒ Mammalia (1 chemicals; 1 data)
 - ☒ Undefined Test organisms (species) (1 chemicals; 3 data)
- Test type**
 - Select scale/unit to use
 - ☒ Chromosome aberration I (Oasis) [1 native data and 2 converted]
 - ☐ Chromosome aberration V (ECVAM) [2 native data and 0 converted]
 - ☐ Micronucleus I [2 native data and 0 converted]

Converted data
2 from scale/unit Chromosome aberration V (ECVAM)

OK Cancel

Manipulation of data matrix Implementation in Toolbox

4) Hide/Show the chart while in data gap filling module

Filter endpoint tree... 1 [target] 2 4 11 12 15 16

Structure

2D

- Boiling point
- log Kow
- Molecular Weight
- Vapor Pressure (Antoine method)
- Water Solubility

Physical Chemical Properties

Human Health Hazards

- Carcinogenicity
- Genetic Toxicity
 - in Vitro
 - Bacterial Reverse Mutation Assay (e.g....)
 - Gene mutation
 - Salmonella typhimurium
 - No S9 Info
 - With S9
 - Without S9

363 °C 368 °C 379 °C 352 °C 352 °C 383 °C 387 °C

4.65 5.06 5.55 4.15 4.15 2.69 3.15

223 Da 237 Da 251 Da 209 Da 209 Da 210 Da 220 Da

3.1E-06 mm Hg 2.22E-06 mm Hg 7.94E-07 mm Hg 8.41E-06 mm Hg 8.41E-06 mm Hg 4.89E-07 mm Hg 3.59E-07 mm Hg

3.56 mg/L 1.32 mg/L 0.42 mg/L 11.1 mg/L 11.1 mg/L 195 mg/L 37.7 mg/L

Data matrix

Click the minimize button to hide the gap filling chart (see next slide)

Descriptors

Prediction

Read-across prediction for Gene mutation, based on 5 values
Observed: Positive (x4); Predicted: Positive

Gene mutation

Positive

Equivalocal

Negative

log Kow

Active descriptor X log Kow

Select / filter data

- Subcategorize
- Mark chemicals by WS
- Mark chemicals by descriptor value
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked data
- Clear existing marks

Accept prediction

Manipulation of data matrix Implementation in Toolbox

4) Hide/Show the chart while in data gap filling module

Filter endpoint tree...

Structure

2D

- Boiling point
- log Kow
- Molecular Weight
- Vapor Pressure (Antoine method)
- Water Solubility

Physical Chemical Properties

Human Health Hazards

- Carcinogenicity
- Genetic Toxicity
 - in Vitro
 - Bacterial Reverse Mutation Assay (e.g....)
 - Gene mutation
 - Salmonella typhimurium
 - No S9 Info 3/3
 - With S9 7/29 M: Positive M: Positive M: Positive M: Positive M: Positive M: Negative M: Positive
 - Without S9 7/29 M: Negative M: Negative M: Negative M: Negative M: Negative M: Negative M: Negative
 - Undefined Test organisms... 1/2
 - Mammalian Cell Gene Mutation A... 1/4
 - in Vivo 1/2

- Profile
- Predefined
 - Substance type
- General Mechanistic
- Endpoint Specific
- Empiric
- Metabolism/Transformations
 - Rat liver S9 metabolism simulator
 - Predefined
 - General Mechanistic
 - Endpoint Specific
- DNA alerts for Ames by OASIS

1 [target] 2 4 11 12 15 16

Chemical structures for each column header.

363 °C 368 °C 379 °C 352 °C 352 °C 383 °C 387 °C

4.65 5.06 5.55 4.15 4.15 2.69 3.15

223 Da 237 Da 251 Da 209 Da 209 Da 210 Da 220 Da

3.1E-06 mm Hg 2.22E-06 mm Hg 7.94E-07 mm Hg 8.41E-06 mm Hg 8.41E-06 mm Hg 4.89E-07 mm Hg 3.59E-07 mm Hg

3.56 mg/L 1.32 mg/L 0.42 mg/L 11.1 mg/L 11.1 mg/L 195 mg/L 37.7 mg/L

Data matrix

Discrete chemical Discrete chemical Discrete chemical Discrete chemical Discrete chemical Discrete chemical Discrete chemical

12 metabolite(s) 8 metabolite(s) 16 metabolite(s) 9 metabolite(s) 9 metabolite(s) 5 metabolite(s) 6 metabolite(s)

1 x Radical >> R... 1 x AN2 1 x Radical >> R... 1 x AN2 1 x AN2 1 x AN2 1 x AN2

1 x SN1 >> Nucl... 1 x AN2 >> Carb... 1 x SN1 >> Nucl... 1 x AN2 >> Carb... 1 x AN2 >> Carb... 1 x AN2 >> Carb... 1 x AN2 >> Carb...

Click the minimize button to restore the gap filling chart

Minimize button icon

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- **Example**

Example

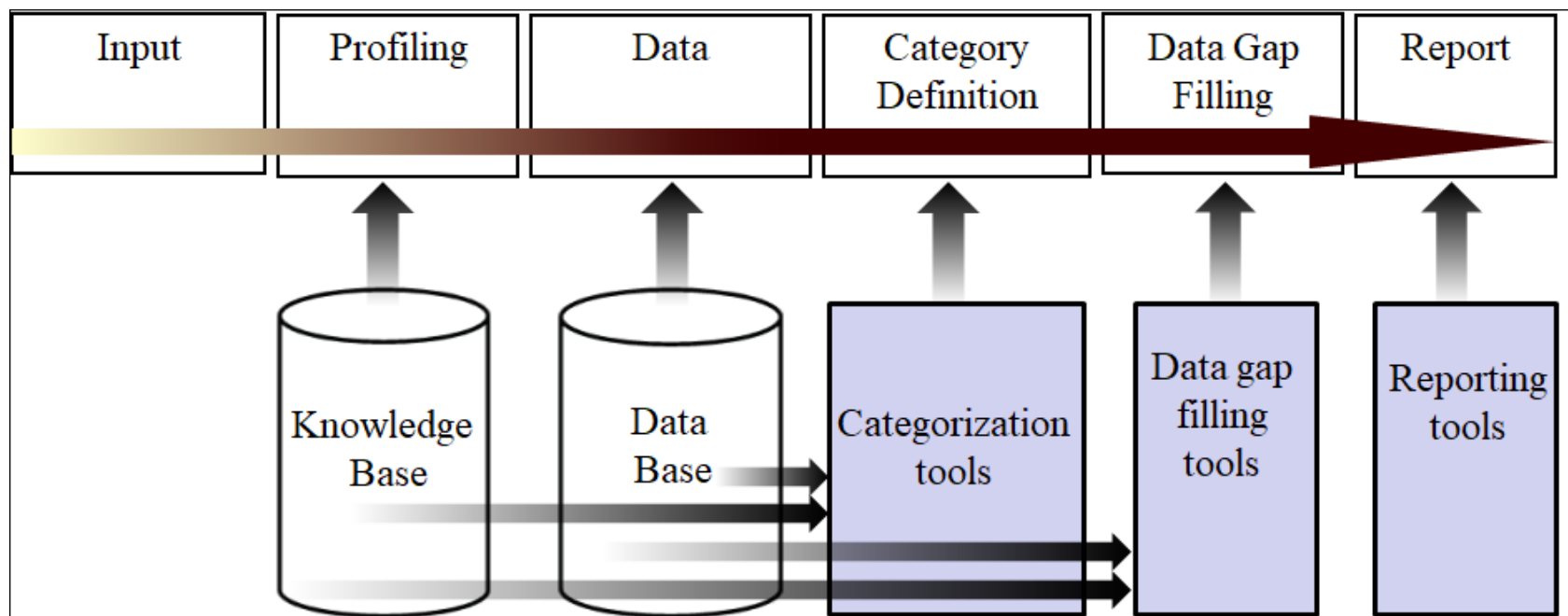
- In this example we will predict *AMES mutagenicity* of Naphthyl N-methylcarbamate [CAS# 27636-33-5], which will be the “target” chemical
- Collect data and profiling results for the target according to the suitable profilers and databases
- Generate *in vitro* rat liver metabolites of the target chemical
- Make read-across prediction for the target by transferring observed data of the preliminary generated metabolites (by *in vitro* rat liver metabolic simulator) to the target outside data gap filling module

Workflow

- **The Toolbox workflow include six modules :**
 - Input
 - Profiling
 - Data
 - Category Definition
 - Data Gap Filling
 - Report
- **In this example we will use only the first three modules in order to fulfil the aims of the example.**

Workflow

Scheme illustrating the Toolbox workflow



Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - **Input**

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.

Input

Entering a target chemical by CAS#

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. The 'Input' icon is highlighted with a red box and labeled '1'. The 'CAS#' button is also highlighted with a red box and labeled '2'. A dialog box titled 'Search by CAS #' is open, showing the input '27636335' and the 'OK' button highlighted with a red box and labeled '3'. The dialog box displays search results for '27636335', including CAS, SMILES, CS Relation, Substance, Composition, Name, and Sources. A chemical structure is shown next to the results.

1. Click **CAS#**

2. Enter **CAS# 27636-33-5**

3. Click **OK**

1	CAS	27636-33-5
	SMILES	<chem>CNC(=O)Oc1cccc2ccccc12</chem>
	CS Relation	Low
	Substance	Mono constituent
	Composition	
	Name	Naphthalenol, Methylcarbamate;N
	Sources	Genotoxicity OASIS ECOTOX

Chemical structure: CNC(=O)Oc1cccc2ccccc12

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - Input
 - **Define target endpoint**

Define target endpoint

Overview

- This functionality allows entering the endpoint of interest e.g., EC3, LC50, gene mutation etc.
- The relevant profiles and databases are getting color highlighted once the targeted endpoint is preliminary defined by this functionality
- There are different ways for defining the target endpoint (via the button from the Input module or by right click from the endpoint tree). For more details press F1 button in order to see the online help.

Define target endpoint

1. Once you are in the **Input** module click **Define** button

2. In the appeared window select **Genetic Toxicity** node part of **Human Health Hazard**

3. Select the specific data from the pop-up menu as shown

4. Click **Finish**

5. The row related to the defined target endpoint is getting yellow highlighted

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- The exercise
 - 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.
- “Profiling” module contains all the knowledge in the system coded in profiling schemes (profilers);
- “Profilers” are a collection of empirical and mechanism knowledge (expertly derived) which could be used to analyse the structural properties of chemicals;
- The “profilers” identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts);
- The “Profiling” module contains also observed and simulated metabolisms/transformations, which could be used in combination with the profiling schemes;
- The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances;
- The “profilers” are not (Q)SARs, i.e. they are not prediction models themselves;
- Based on the defined target endpoint the most relevant profilers are getting colour highlighted: greens are the most suitable, orange are plausible*
- For the purpose of our example only suitable profilers in combination with suitable metabolism simulator are used (see next slide).

**For more details regarding relevancy of profilers see tutorial: [Example for predicting skin sensitization taking into account alert performance](#)*

Profiling

Profiling the target chemical

The screenshot shows the Q SAR Toolbox Profiling interface. The top navigation bar includes buttons for Input, Profiling (highlighted with a red box and callout 1), Data, Category definition, Data Gap Filling, and Report. Below the navigation bar, the left sidebar contains a Documents panel (callout 3) and two panels for Profiling methods and Metabolism/Transformations (both with callout 2). The main area displays a chemical structure (callout 4) and a Filter endpoint tree (callout 5) with a table of results (callout 6).

1 Go to **Profiling**; and unselect all previously checked profiler (click Unselect All);

2 Check the suitable profiles and simulators (the green ones);

3 Click **Apply**;

4 No alert is found in the target structure based on general and endpoint-specific profilers;

5 Seven metabolites are produced for the target after applying the *in vitro* rat liver metabolism simulator;

6 Structural alerts for interaction with DNA are found in the generated metabolites.

Profiling

Explain of profiling results

The screenshot shows the QSAR Toolbox Profiling results window. On the left is a 'Filter endpoint tree...' panel with a search bar containing '1 [target]'. Below it is a 'Structure' panel showing a chemical structure. The main panel displays 'Profiling results' for a target molecule. A right-click context menu is open over the results, with 'Explain' highlighted (annotated with a red box and a blue callout '1'). The results list shows 7 metabolite(s): 5 with no alerts found, 2 x AN2, 2 x Michael-type addition, quinones and trihydroxybenzenes, 2 x Non-covalent interaction, 2 x DNA intercalation, and 2 x Radical. A 'Display chemicals' button is highlighted in the results list (annotated with a red box and a blue callout '2'). A 'File' dialog box is open, showing two metabolites with 'No CAS number' and their chemical structures. The dialog has 'Save to smi' and 'OK' buttons.

1. Right-click over the results and select **Explain** for more details.
2. Right-click over the domain information (e.g. AN2) and select **Display chemicals** to see the metabolites belonging to this domain (in this case this is 'Quinones and Trihydroxybenzenes' alert part of nucleophilic addition type mechanism (AN2)).

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - Input
 - Profiling
 - **Data**

Data Overview

- The “Data” module refers to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox databases.
- 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).
- Once the endpoint is selected, the databases, which contain such type of data are highlighted in green (see next slide).

Data

Collect data

1. Go to **Data** module

2. Select the green highlighted databases only

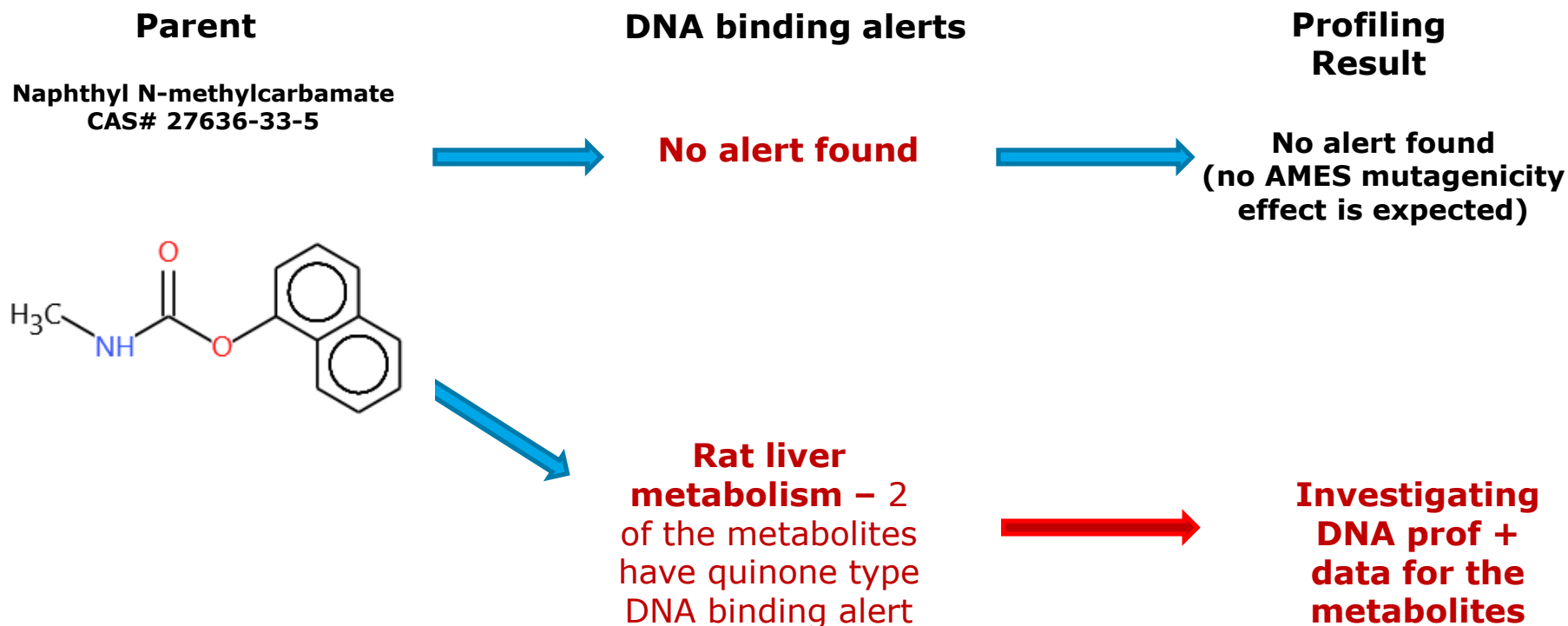
3. Click **Gather**

4. No data has been found for target chemical for the target endpoint

Recap

- The first module (Input), which introduces the target chemical, ensure correctness of the structure.
- The second module (Profiling) shows that there is no DNA binding alert for target chemical itself, but structural alerts responsible for DNA interaction have been found in the generated rat liver metabolites. The latter determines the forthcoming actions of the workflow.
- In the third module (Data), you have found that the target chemical has no data associated with the target endpoint
- Due to the fact that AMES test accounts S9 metabolism the study continues with investigating profiling results and data of generated rat liver metabolites of the target chemical (see next slides).

Recap



Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - Input
 - Profiling
 - Data
 - **Simulation of rat liver S9 metabolism**

Handling of rat liver S9 metabolism of target chemical

- The next actions in the workflow is metabolizing the target chemical by Rat liver S9 metabolism simulator
- The simulation of the rat liver metabolism of the target chemical is accomplished in section **Input**
- The generated metabolites appear in tree like form (see next slide)

Handling of rat liver S9 metabolism of target chemical

Multiplication of the target chemical

The screenshot displays the QSAR TOOLBOX interface. The top toolbar includes buttons for 'Input', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Input' button is highlighted with a red box and a callout '1'. Below the toolbar, the 'Document' menu is open, showing options like 'New', 'Open', 'Close', 'Export', 'Print', 'Rename', 'Delete', and 'Delete All But This'. The 'Multiplication' option is highlighted with a red box and a callout '2'. The 'Multiplication' submenu is open, showing various simulation options. The 'Metabolism/Transformations' option is highlighted with a red box and a callout '3'. The 'Rat liver S9 metabolism simulator' is selected in the list.

1. Go to **Input** module
2. Click on the level with **CAS #** of the target chemical and right-click on it, then
3. Select **Multiplication / Metabolism/Transformations / Rat liver S9 metabolism simulator**
4. Generated metabolites appear in tree like form and also are aligned next to the target (shown on next slide)

Handling of rat liver S9 metabolism of target chemical

Multiplication of the target chemical

The screenshot displays the QSAR TOOLBOX software interface. The top menu bar includes options like Document, Single Chemical, Chemical List, Search, and Target Endpoint. Below this is a toolbar with icons for New, Open, Close, Save, CAS#, Name, Structure, Composition, Select, ChemIDs, Database, Inventory, List, Substructure (SMARTS), Query, and Define.

On the left, the 'Documents' panel shows a tree structure. A red bracket highlights a list of metabolites generated from a target chemical, labeled '1'. The list includes:

- [C: 1;Md: 1;P: 0] Rat liver S9 metabolism simulator
- [C: 1;Md: 0;P: 0] metabolite #1
- [C: 1;Md: 0;P: 0] metabolite #2
- [C: 1;Md: 0;P: 0] metabolite #3
- [C: 1;Md: 0;P: 0] metabolite #4
- [C: 1;Md: 0;P: 0] metabolite #5
- [C: 1;Md: 0;P: 0] metabolite #6
- [C: 1;Md: 0;P: 0] metabolite #7

In the center, the 'Filter endpoint tree...' panel shows a list of endpoints. A red box highlights the 'Parent chemical...' and the first seven 'metabolite #' columns, labeled '2'. The metabolites are displayed as chemical structures in a grid.

The right side of the interface shows a table with columns for the parent chemical and the seven metabolites. The table is currently empty, with only the chemical structures visible in the first row.

1. Generated metabolites appear in tree like form in the documented tree
2. Also metabolites are aligned next to the target

Next actions are focused on investigating the profilers of the generated metabolites and collecting data for them

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - Input
 - Profiling
 - Data
 - Generation of rat liver S9 metabolism
 - **Profiling and collecting data for metabolites**

Handling of rat liver S9 metabolism of target chemical

Collect data for metabolites

1. Go to **Data** module;

2. The databases related to the defined target endpoint are already selected;

3. Click **Gather**;

4. The data for the parent and metabolites appears in data matrix. Only one of the generated analogues is pure negative chemical according to the collected experimental data

Parent chemical...	metabolite #1	metabolite #2	metabolite #3	metabolite #4	metabolite #5	metabolite #6	metabolite #7
<chem>N#Cc1ccccc1</chem>	<chem>CCN</chem>	<chem>CC(O)N</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>
Structure	Structure	Structure	Structure	Structure	Structure	Structure	Structure
Structure info	Structure info	Structure info	Structure info	Structure info	Structure info	Structure info	Structure info
Parameters	Parameters	Parameters	Parameters	Parameters	Parameters	Parameters	Parameters
Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties	Physical Chemical Properties
Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport	Environmental Fate and Transport
Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information	Ecotoxicological Information
Human Health Hazards	Human Health Hazards	Human Health Hazards	Human Health Hazards	Human Health Hazards	Human Health Hazards	Human Health Hazards	Human Health Hazards
Acute Toxicity	Acute Toxicity	Acute Toxicity	Acute Toxicity	Acute Toxicity	Acute Toxicity	Acute Toxicity	Acute Toxicity
ADME	ADME	ADME	ADME	ADME	ADME	ADME	ADME
Bioaccumulation	Bioaccumulation	Bioaccumulation	Bioaccumulation	Bioaccumulation	Bioaccumulation	Bioaccumulation	Bioaccumulation
Carcinogenicity	Carcinogenicity	Carcinogenicity	Carcinogenicity	Carcinogenicity	Carcinogenicity	Carcinogenicity	Carcinogenicity
Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity	Developmental Toxicity / Teratogenicity
Genetic Toxicity	Genetic Toxicity	Genetic Toxicity	Genetic Toxicity	Genetic Toxicity	Genetic Toxicity	Genetic Toxicity	Genetic Toxicity
in Vitro	in Vitro	in Vitro	in Vitro	in Vitro	in Vitro	in Vitro	in Vitro
Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)	Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)
Gene mutation	Gene mutation	Gene mutation	Gene mutation	Gene mutation	Gene mutation	Gene mutation	Gene mutation
5/78	MS: Negative MS: Negative MS: Negative	MS: Equivocal MS: Negative MS: Negative	MS: Equivocal MS: Negative MS: Negative	MS: Equivocal MS: Negative MS: Negative	MS: Positive MS: Positive	MS: Negative MS: Negative MS: Negative	MS: Negative MS: Negative MS: Negative
20 Negative data (-/+S9)	3 Positive (-/+S9); 13 Neg; 1 Equivocal	7 Positive (-/+S9); 10 Neg; 1 Equivocal	2 Positive(+S9); 19 Negative	2 Positive data (-S9)			

1. Go to **Data** module;
2. The databases related to the defined target endpoint are already selected;
3. Click **Gather**;
4. The data for the parent and metabolites appears in data matrix. Only one of the generated analogues is pure negative chemical according to the collected experimental data

Handling of rat liver S9 metabolism of target chemical

Profiling the package of metabolites

1. Go to **Profiling**;

2. Check suitable profilers (green) related to the target endpoint;

3. Unselect Rat liver S9 metabolism simulator

4. Click **Apply**;

5. The profiling results appears in the data matrix.

Profiling results related to the most suitable profilers of package parent and metabolites

Parent chemical...	metabolite #1	metabolite #2	metabolite #3	metabolite #4	metabolite #5	metabolite #6	metabolite #7
Structure	<chem>CC(N)=O</chem>	<chem>CC(=O)O</chem>	<chem>CC1=CC=CC=C1</chem>	<chem>CC1=CC=CC=C1</chem>	<chem>CC1=CC=CC=C1</chem>	<chem>CC1=CC=CC=C1</chem>	<chem>CC1=CC=CC=C1</chem>
Environmental Fate and Transport							
Ecotoxicological Information							
Human Health Hazards							
Acute Toxicity							
ADME							
Bioaccumulation							
Carcinogenicity							
Developmental Toxicity / Teratogenicity							
Genetic Toxicity							
in Vitro							
Bacterial Reverse Mutation Assay (e.g., Salmonella typhimurium)							
Gene mutation	5/78	MS: Negative MS: Negative MS: Negative	MS: Equivocal MS: Negative MS: Negative	MS: Equivocal MS: Negative MS: Negative	MS: Positive MS: Positive		MS: Negative MS: Negative MS: Negative
in Vitro Mammalian Chromosome...	1/1	M: Positive					
Immunotoxicity							
Irritation / Corrosion							
Neurotoxicity							
Photoinduced toxicity							
Repeated Dose Toxicity							
Sensitisation							
ToxCast							
Toxicity to Reproduction							
Toxicokinetics, Metabolism and Distribution							
Profiling							
General Mechanistic							
DNA binding by OASIS	No alert found	No alert found	No alert found	AN2	AN2	No alert found	No alert found
DNA binding by OECD	No alert found	No alert found	No alert found	Michael addition	Michael addition	Michael addition	Michael addition
				AN2	AN2	No alert found	No alert found
				Quinones	Quinones	No alert found	No alert found

Outlook

- Background
- Keywords
- Objectives
- Specific aim
- Manipulation of data matrix
- Example
 - Input
 - Profiling
 - Data
 - Generation of rat liver S9 metabolism
 - Profiling and collecting data for metabolites
 - **Transferring data to the target outside data gap filling module**

Hold (1) and **drag** the chemical next to the target.
The profiling results (2) and the data (3) are available for the parent and metabolites (see next slide).

The structural alert for interaction with DNA for metabolite #4 coincide with the identified positive experimental data of the metabolite as parent (3)

Structural alerts for interaction with DNA identified in the generated metabolites

Transferring experimental data of metabolite to the target outside data gap filling

1 Right-click the cell with observed data of the metabolite #4;

2 Select **Transfer to target**;

3 Select the data point to be transferred to the parent;

4 Click **OK** button.

The positive data points of the metabolite #4 could be transferred to the target chemical. A single data point could be transferred at a time

Read-across prediction based on the positive observed data of the metabolite #4 appeared for the target chemical

Congratulations!

- You have now been introduced to the Data matrix manipulation options;
- You have now been introduced to the transfer of a read-across prediction to the target chemical outside of the gap filling module.
- Note, proficiency comes with practice!