

OECD QSAR Toolbox v.4.1

Step-by-step example for predicting Ames mutagenicity by making use of read-across

Outlook

- **Background**
- Objectives
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox in a data-gap filling exercise using read-across based on molecular similarity with data pruning.
- If you are a novice user of the Toolbox you may wish to review the “Getting Started” document available at [www.oecd.org/env/existingchemicals/qsar] as well as go through tutorials 1 and 3.

Outlook

- Background
- **Objectives**
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Objectives

- **This presentation demonstrates a number of functionalities of the Toolbox:**
 - Entering a target chemical by SMILES notation and Profiling
 - Identifying analogues for a target chemical by molecular similarity
 - Retrieving experimental results available for those analogues, and for multiple endpoints
 - Filling data gaps by read-across

Outlook

- Background
- Objectives
- **Specific Aims**
- Read-across
- The exercise
- Workflow of the exercise
- Save the prediction

Specific Aims

- To review the workflow of the Toolbox.
- To reacquaint the user with the six modules of the Toolbox.
- To reacquaint the user with the basic functionalities within each module.
- To introduce the user to new functionalities of selected modules.
- To explain to the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Read-across**
- The exercise
- Workflow of the exercise
- Save the prediction

Read-across & the Analogue Approach

- Remember, read-across is a method that can be used to estimate missing data from a single or limited number of chemicals using the analogue approach.
- In the analogue approach, experimental 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” (i.e., within the same category).

Analogous Chemicals

- Previously you learned that analogous sets of chemicals are often selected based on the hypothesis that the toxicological effects of each member of the set will show a common behaviour.
- For this reason mechanistic profilers and grouping methods have been shown to be of great value in using the Toolbox.
- However, there are cases where the mechanistic profilers and grouping methods are inadequate and one is forced to rely on molecular similarity to form a category.
- The Toolbox allows one to develop a category by using either a mechanistic category like DNA binding or structural similarity.
- Since there is no preferred way of identifying structural similarity, the user is guided to use DNA binding as a first option.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- **The exercise**
- Workflow of the exercise
- Save the prediction

Exercise

- In this exercise we will predict the Ames mutagenicity potential for an untested compound, (n-hexanal) [SMILES CCCCCC=O)], which is 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 empirical similarity, with respect to “Organic functional groups” profiler.
- The prediction itself will be made by “read-across” analysis.

Side-Bar On Mutagenesis

- Mutagens do not create mutations.
- Mutagens create DNA damage.
- Mutations are changes in nucleotide sequence.
- Mutagenesis is a cellular process requiring enzymes and/or DNA replication, thus cells create mutations.

Side-Bar On Mutagenesis

- Mutations within a gene are generally base-substitutions or small deletions/insertions (i.e., frame shifts).
- Such alteration are generally called point mutations.
- The Ames scheme based on strains of *Salmonella* provide the corresponding experimental data.

Side-Bar On Mutagenesis

- The Ames mutagenicity assay (see OECD guideline 471) is designed to assess the ability of a chemical to cause point mutations in the DNA of the bacterium *Salmonella typhimurium*.
- The Ames test includes a number of strains (TA1537, TA1535, TA100, TA98 and TA97) that have been engineered to detect differing classes of mutagenic chemicals.
- The basic test only detects direct acting mutagens (i.e., those chemicals able to interact with DNA without the need for metabolic activation).

Side-Bar on Metabolic Activation

- The inclusion of an S9 mix of rodent liver enzymes is designed to assess those chemicals requiring metabolic activation in order to be mutagenic.
- Typically, chemicals are assayed both without S9 and with S9 with results being reported in a binary fashion
- A positive result in any of the bacterial strains with or without S9 confirms mutagenic potential.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
- Save the prediction

Workflow

- **The 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
- The exercise
- **Workflow of the exercise**
 - **Chemical input**

Chemical Input Overview

- As you leader in the previous tutorials, 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.

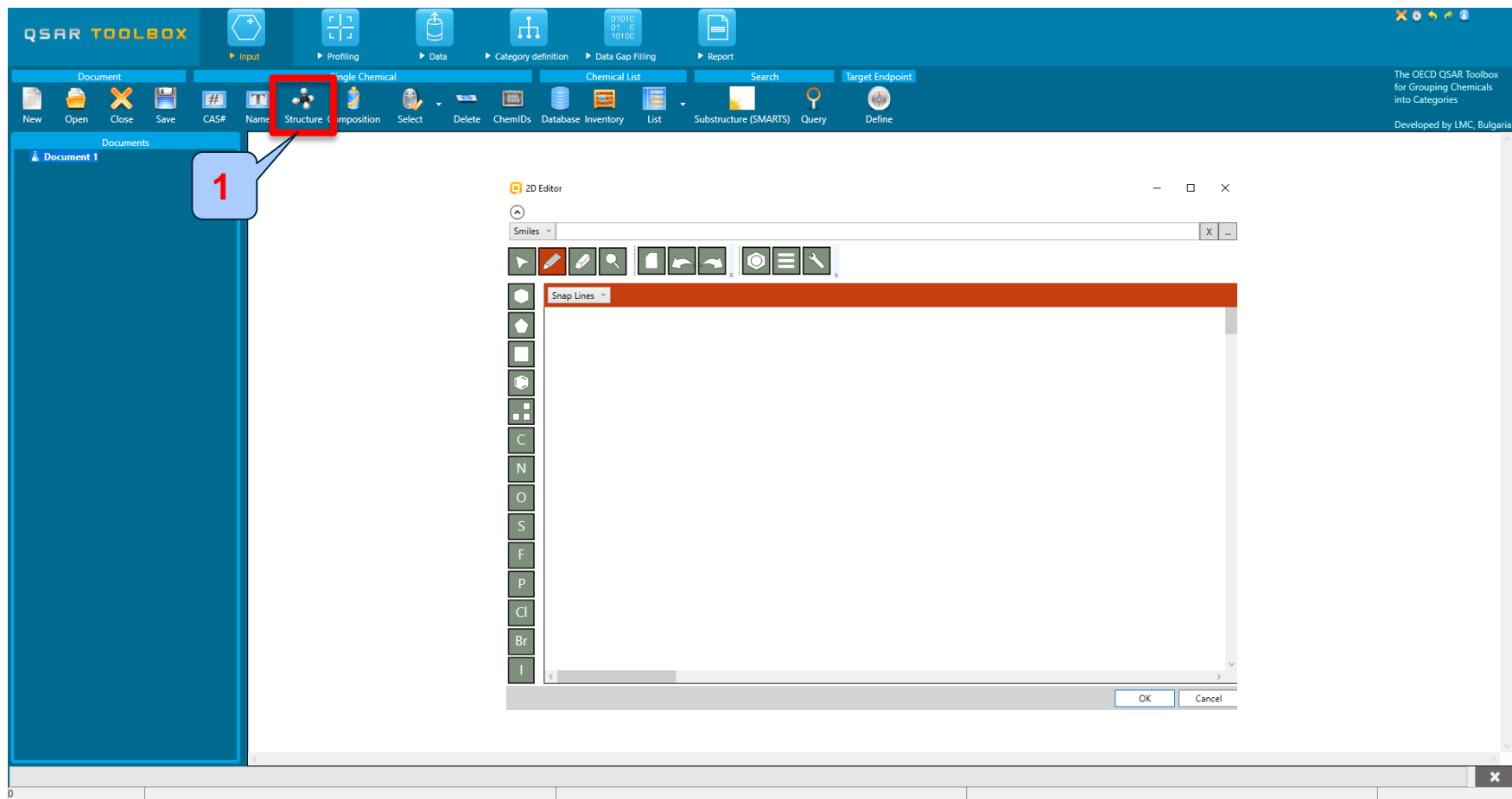
Chemical Input

Ways of Entering a Chemical

- **Remember there are several ways to enter a target chemical and the most often used are:**
 - CAS# ,
 - SMILES (simplified molecular information line entry system) notation, and
 - Drawing the structure
- **Click** on **Structure**.
- This inserts the window entitled "2D editor" (see next screen shot).

Chemical Input

Input target chemical by drawing



1. Click on **Structure**

Chemical Input

Input target chemical by SMILES

- In the area next to "Smiles" **type** CCCCC=O.
- Note as you type the SMILES code the structure is being drawn in the centre of the structure field (see next screen shot).
- **Click** "OK" to accept the target chemical.

Chemical Input

Input target chemical by SMILES

1. **Type** CCCCCCC=O in Smiles window; 2. 2D structure; 3. **Click** OK;

Chemical Input

Input target chemical by SMILES

The Toolbox now searches the Toolbox databases and inventories for the presence of a chemical with structure related to the current SMILES notation. It is depicted as a 2D image.

Two chemicals are found. By default they are unselected. Select the chemical you want by ticking.

Select chemicals

Select All Unselect All Invert Selection Selected 1 of 2

| | | | | |
|-------------------------------------|---|-------------|-----------------------------------------|--|
| <input checked="" type="checkbox"/> | 1 | CAS | 66-25-1 | |
| | | SMILES | CCCCCC=O | |
| | | CS Relation | High | |
| | | Substance | Mono constituent | |
| | | Composition | | |
| | | Name | hexaldehyde Hexanal Hexylaldehyde | |
| <input type="checkbox"/> | 2 | CAS | 110-62-3 | |
| | | SMILES | CCCCCC=O | |
| | | CS Relation | Low | |
| | | Substance | Mono constituent | |
| | | Composition | | |
| | | Name | valeraldehyde | |

OK Cancel

1. **Select** the first chemical by clicking on the tick; 2. **Click** OK.

Chemical Input

Target chemical identity

- You have now selected your target chemical.
- **Expand** "Structure info" field to display chemical identification information (see next screen shot).
- It is important to remember that the workflow is based on the structure coded in SMILES.

Chemical Input

Target chemical identity

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes buttons for 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Profiling' button is highlighted with a blue callout box containing the number '1'. Below the navigation bar, the 'Documents' panel shows 'Document 1' with a 'Search chemical' button. The 'Filter endpoint tree...' panel is open, showing a tree view with 'Structure info' selected and highlighted by a red box. The 'Structure info' section includes fields for 'CAS Number' (66-25-1), 'CAS Smiles relation' (High), 'Chemical name(s)' (hexaldehyde), 'Composition' (C6H12O), 'Molecular Formula' (Mono constituent), and 'Structural Formula' (CCCCC=O). A red circle highlights the 'CAS Number' field. The 'Profiling methods' and 'Metabolism/Transformations' panels are also visible, showing various simulation options.

The workflow on the first module is now complete; **click** on "Profiling" [1] to move to the next module.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - **Profiling**

Profiling

Overview

- As you may remember, “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.
- Available profilers includes likely mechanism(s) of action which have been shown to be useful in forming categories that include the target chemical.

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.
- To help the user to choose suitable profiling methods, a new feature has been developed – see next slide.

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox interface. The top menu bar includes 'Profiling' and 'Custom profile'. The 'Profiling' menu is open, showing options like 'Apply', 'View', 'New', and 'Delete'. A callout box labeled '3' points to the 'Color by' option in the Profiling menu. The 'Filter endpoint tree...' window is open, showing a tree structure of endpoints. A callout box labeled '1' points to the tree, and another callout box labeled '2' points to the 'Human Health Hazards' category. The 'Endpoint Specific' section is expanded, showing various endpoints. A callout box labeled '3' points to the 'Endpoint selected in the data matrix' option in the legend.

1. **Click** on the box to open the nodes of the tree;
2. **Mark the** box in the data matrix related to the target endpoint of chemical;
3. From Profiling / Profiling methods / Options / **Color by** (pop-up menu) select **Endpoint selected in the data matrix**

According to the legend profilers related to the selected endpoint are highlighted in **Green** or **Orange**

Legend

Endpoint selected in the data matrix

- Suitable
- Plausible
- Unclassified

OK

Profiling

Profiling the target chemical

- For this example, the following general mechanistic profiling methods are relevant to genetic toxicity:
 - DNA binding by OASIS – mechanistic grouping
 - DNA binding by OECD – mechanistic grouping
 - Protein binding by OASIS – mechanistic grouping
 - Protein binding by OECD – mechanistic grouping
 - Carcinogenicity (genotox and nongenotox) alerts by ISS - endpoint specific
 - DNA alerts for AMES by OASIS - endpoint specific
 - in vitro mutagenicity (Ames test) alerts by ISS - endpoint specific
 - in vivo mutagenicity (Micronucleus) alerts by ISS - endpoint specific
 - Organic function groups - empiric

Profiling

Profiling the target chemical

- **Select** the “Profiling methods” related to the target endpoint.
- This selects (a **green** check mark appears) or deselects (**green** check disappears) profilers.
- For this example, select the profilers relevant to genetic toxicity (see next screen shot).

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox interface. The top toolbar includes buttons for 'Apply', 'View', 'New', and 'Delete'. A red circle labeled '2' highlights the 'Apply' button. Below the toolbar, the 'Documents' panel shows 'Document 1' with a 'Search chemical' option. The 'Filter endpoint tree...' panel on the right shows a tree structure with 'Structure' selected, displaying chemical information for a target chemical (hexaldehyde). The 'Profiling methods' panel at the bottom shows a list of methods with checkboxes. A red bracket labeled '1' highlights the 'Endpoint Specific' section, which includes the following checked items:

- 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)
- Endpoint Specific**
 - Acute aquatic toxicity classification by Verhaar (Modified)
 - Acute aquatic toxicity MOA by OASIS
 - Aquatic toxicity classification by ECOSAR
 - Bioaccumulation - metabolism alerts
 - Bioaccumulation - metabolism half-lives
 - Biodegradation fragments (BioWIMI MITI)
 - Carcinogenicity (genotox and nongenotox) alerts by ISS
 - DNA T scheme
 - DNA alerts for AMES by OASIS
 - DNA alerts for CA and MNT by OASIS
 - Eye irritation/corrosion Exclusion rules by BFR
 - Eye irritation/corrosion Exclusion rules by OECD

1. Check the profilers related to the target endpoint (see slide 32); **2. Click** Apply

Profiling

Profiling the target chemical

- The actual profiling will take several seconds depending on the number and type of selected profilers.
- The results of profiling automatically appear as a dropdown boxes under the target chemical (see next slide).
- Please note the specific profiling results by DNA, Protein binding, and Organic functional groups.
- These results will be used to search for suitable analogues in the next steps of the exercise.

Profiling

Profiles of n-hexanal

The screenshot displays the QSAR Toolbox interface for profiling n-hexanal. The top navigation bar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this are buttons for Profiling and Custom profile, and a Documents section with a search bar. The main area is divided into a left sidebar with a tree view of endpoints, a central 'Filter endpoint tree...' section, and a right-hand results table. The results table shows various alerts such as 'Schiff base formers', 'Simple aldehyde (Genotox)', and 'Aldehyde'. A red circle highlights a 'Simple aldehyde (Genotox)' entry with a green-white rectangle next to it, and a blue callout box with the number '1' points to it.

Green-white rectangles in the result boxes indicate there is more than one profiling result and the fields need to be expanded (1).

Profiling

Profiles of n-hexanal

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Profiling' section is active, showing a 'Documents' panel on the left and a 'Filter endpoint tree...' panel in the center. The 'Filter endpoint tree...' panel is expanded to show 'DNA binding by OECD'. A red circle highlights the 'Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes' result. A red arrow points from this result to an 'Explain' button. A callout box with the number '1' points to the result, and another callout box with the number '2' points to the 'Explain' button.

- In this case there is structural evidence that the target has positive DNA and Protein binding alert
- This allows to bind covalently to DNA
- This mechanistic information is important for the grouping of analogues.

1. Right click on the box with profiling result by DNA binding by OECD.
2. Left Click on the "Explain" box to see why the target is profiled as "Mono-aldehydes" by DNA binding by OECD (see next slide).

Profiling

Profiles of n-hexanal

The screenshot displays the QSAR Toolbox interface. The top toolbar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The Profiling methods panel on the left shows various endpoints, with 'Empiric' and 'Organic functional groups' selected. The 'Filter endpoint tree...' window shows a tree structure with 'Direct Acting Schiff Base Formers' and 'Mono aldehydes' highlighted in red. The 'Profiling results' window shows a list of results, with 'Direct Acting Schiff Base Formers' and 'Mono aldehydes' highlighted in blue. A 'Details' button is highlighted in red at the bottom of the 'Profiling results' window. A callout box with the number '3' points to the 'Details' button. Another callout box with the number '3' points to the 'Explain' button in the top toolbar.

3. After clicking on the "Explain" a window with chemical profiles appears. Then click "Details" to see detailed explanation

Profiling DNA binding by OECD of n-hexanal

1. Structural boundary of the category; **2.** Definition of the used common fragments; **3.** Mechanistic justification of the category (Literature tab)

1. Structural boundary of the category; **2.** Definition of the used common fragments; **3.** Mechanistic justification of the category (Literature tab)

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - Profiling
 - **Data**

Data Overview

- As you should remember, “Data” refers to the electronic process of retrieving the fate and toxicity data that are stored in the Toolbox database.
- Note, data can be gathered in a global fashion (i.e., collecting all data of all endpoints) or on more narrowly defined settings (e.g., collecting data for a single or limited number of endpoints).

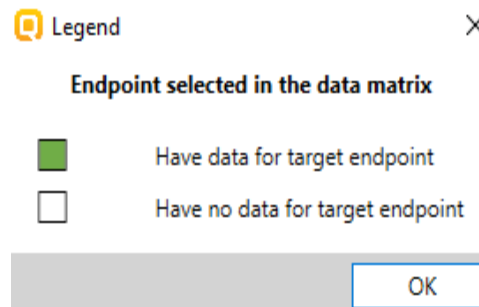
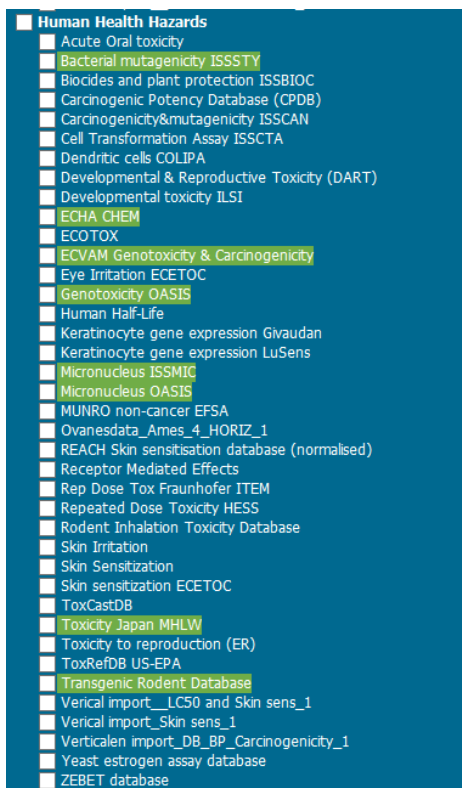
Data Case study

A new functionality for specifying databases containing data with desired endpoint is available

The screenshot displays the QSAR Toolbox software interface. The top navigation bar shows the 'Data' tab selected. The left sidebar contains a 'Databases' menu (callout 4) and an 'Options' menu. The 'Options' menu is open, showing 'Endpoint selected in the data matrix' (callout 2) as the selected option. The main workspace features a 'Filter endpoint tree...' panel on the left and a '1 [target]' panel on the right. The 'Filter endpoint tree...' panel lists various endpoints, with 'Genetic Toxicity' (callout 3) highlighted. The '1 [target]' panel shows the chemical structure of hexanal (CCCCC=O).

1. Click on **Data**;
2. **Click** on the box to open the nodes of the tree;
3. **Mark the** box related to the target endpoint;
4. From Databases / Options / Color by (pop-up menu) select **Endpoint selected in the datamatrix**

Data Case study



- According to the legend in Databases / Options databases containing data related to selected endpoint are highlighted in **Green**

- In our example, we limit our data gathering to the common genotoxicity endpoints from databases containing genotoxicity data (**Bacterial mutagenicity ISSSTY, Genotoxicity OASIS, Micronucleus ISSMIC, Micronucleus OASIS and Toxicity Japan MHLW**) – see next slide.

Data Gather data

The screenshot shows the QSAR Toolbox software interface. At the top, there is a navigation bar with icons for 'Input', 'Profiling', 'Data', 'Category definition', and 'Data Gather'. Below this is a toolbar with 'Import' and 'Export' buttons. The 'Gather' button, represented by a folder icon, is circled in red and labeled with a blue callout box containing the number '3'. Below the toolbar is a 'Documents' section and a 'Databases' section. The 'Databases' section has a search field and buttons for 'Select All', 'Unselect All', and 'Invert'. A list of databases is shown, with several checked, including 'Human Health Hazards', 'Bacterial mutagenicity ISSSTY', 'Genotoxicity OASIS', and 'Toxicity Japan MHLW'. A red oval highlights this list, labeled with a blue callout box containing the number '2'. To the right of the database list is a 'Filter endpoint tree...' panel showing a hierarchical tree of endpoints. The 'Human Health Hazards' section is expanded, showing sub-categories like 'Acute Toxicity', 'Bioaccumulation', 'Carcinogenicity', etc. A blue callout box with the number '1' points to the 'Databases' section. At the bottom right, a blue box contains the following instructions:

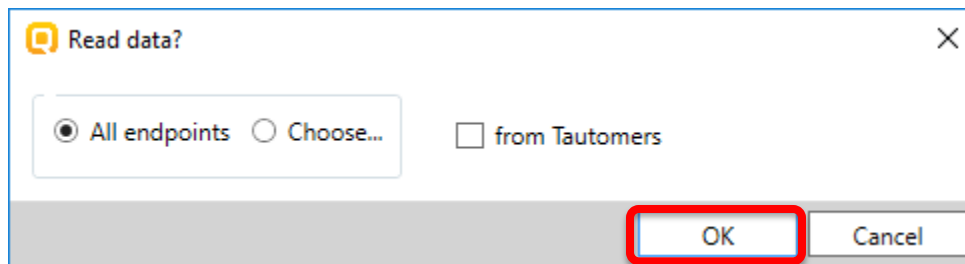
1. **Go to** Databases / Human Health Hazards section;
2. **Select** databases related to the target endpoint (slide 43);
3. Click **Gather**

Data

Process of collecting data

Toxicity information on the target chemical is electronically collected from the selected datasets.

A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.



Click OK to read all available data

Data

Process of collecting data

In this example, an insert window appears stating that there was 2 data points available for the target chemical appears.

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', and 'Data Gap Fill'. The 'Data' step is active. On the left, the 'Databases' panel is open, showing a list of endpoints under 'Human Health Hazards'. The 'Filter endpoint tree...' window is open, displaying a tree view of endpoints. The 'Genetic Toxicity' endpoint is selected, and a sub-window shows 'M: Negative' with '(1/2)' next to it. A dialog box at the bottom center displays the message '2 points added across 1 chemicals.' with an 'OK' button highlighted by a red box.

Data

Process of collecting data

In this example, an insert window appears stating that there was 2 data points available for the target chemical appears. Click OK. Collected data appear on the data matrix.

The screenshot shows the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', and 'Data Gap Fill'. The 'Data' menu is open, showing options like 'Gather', 'Import', 'IUCLID6', and 'IUCLID6'. The 'Documents' and 'Databases' panels are visible on the left. The 'Filter endpoint tree...' panel shows a hierarchical tree of endpoints, with 'Genetic Toxicity' selected. A callout box '1' points to this node. The 'Data matrix' panel on the right shows a table with two rows of data: '(1/1) M: Negative' and '(1/1) M: Negative'. A callout box '2' points to the first row of this data matrix.

1. Click consequently on the boxes to open the nodes of the tree
2. There are two negative experimental data for the target chemical

Data Recap

- You have entered the target chemical by SMILES and found a substance with CAS: 66-25-1 (n-hexanal).
- You have collected two negative experimental data for n-hexanal.
- We will try to reproduce the experimental data by making read-across.
- Click on “Category definition” to move to the next module.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - Profiling
 - Data
 - **Category definition**

Category Definition

Overview

- As stated in the previous tutorials, this module provides the user with several means of grouping chemicals into a toxicologically meaningful category that includes the target molecule.
- Remember, this is the critical step in the workflow of the Toolbox.
- Several options are available in the Toolbox to assist the user in defining the category definition.

Category Definition

Side-Bar on Mutagens

- It is important to remember that mutagens are really cell-damaging agents, which can create a wide array of adverse effects beyond damage to DNA.
- Mechanistic profile of the target chemical is showed in this Section (see next screen shots).

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 chemical (Chapter 4) could be downloaded from:
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>
- For this example, we will start from a broad group based on Organic functional group and after that
- Will refine the category by a specific DNA binding mechanism identified for the target chemical and find analogues which can bind by the same mechanism and for which experimental results are available.

Category Definition

Defining Organic functional group

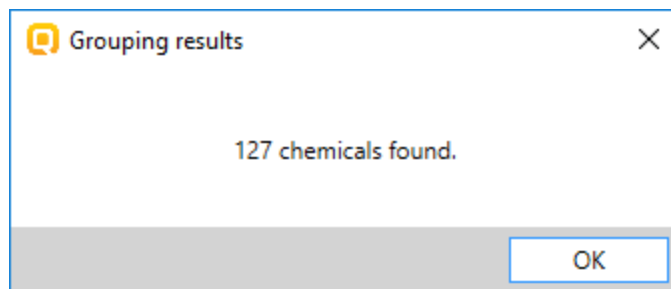
The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition' (highlighted with a red box), and 'Data Gap Fill'. The 'Documents' pane on the left shows a list of categories, with 'Organic functional groups' selected (1). The 'Define' button in the 'Categorize' menu is circled in red (2). The 'Grouping options (Organic functional groups)' dialog box is open, showing a 'Target' field with 'Aldehyde' selected (3). The 'OK' button at the bottom right of the dialog is circled in red (4). The 'Filter endpoint tree...' pane on the right shows a hierarchical tree of endpoints, including 'Structure', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', 'Human Health Hazards', 'Immunotoxicity', 'Irritation / Corrosion', 'Neurotoxicity', 'Photoinduced toxicity', 'Repeated Dose Toxicity', and 'Sensitisation'.

1. **Select** Organic functional groups; 2. **Click** Define; 3. The target category is Aldehydes; 4. **Click** OK

Category Definition

Analogues

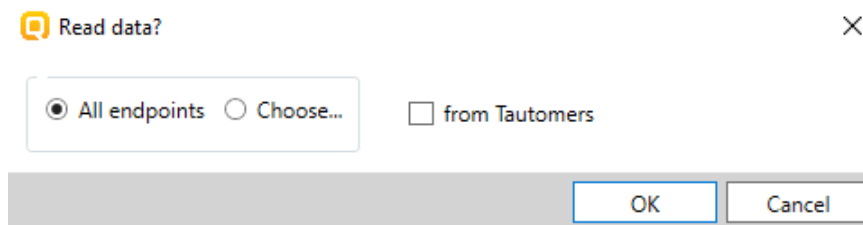
- The Toolbox now identifies all chemicals corresponding to category "Aldehydes" by Organic functional groups listed in the databases selected under "Data".
- A notice with numbers of analogues found (including the target chemical) is appeared.



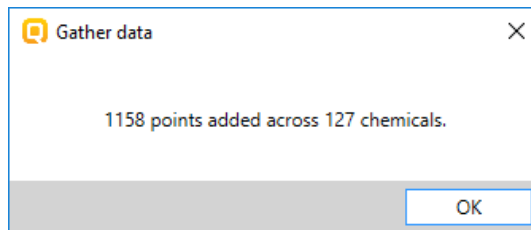
Category Definition

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)



- In this example, because only databases that contain information for genetic toxicity endpoint are selected, both options give the same results.



Category Definition

Summary information for Analogues

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

The screenshot shows the QSAR Toolbox interface with the 'Category definition' workflow selected. The main window displays a matrix of experimental results for various chemical endpoints across seven different chemical structures. A red box highlights a specific row of data for the 'Genetic Toxicity' endpoint.

| Endpoint | 1 [target] | 2 | 3 | 4 | 5 | 6 | 7 |
|--------------------------------------------------|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Structure | | | | | | | |
| Structure info | | | | | | | |
| Parameters | | | | | | | |
| Physical Chemical Properties | | | | | | | |
| Environmental Fate and Transport | | | | | | | |
| Ecotoxicological Information | | | | | | | |
| Human Health Hazards | | | | | | | |
| Acute Toxicity (1/1) | | | | | | | |
| Bioaccumulation | | | | | | | |
| Carcinogenicity | | | | | | | |
| Developmental Toxicity / Teratogenicity | | | | | | | |
| Genetic Toxicity | | | | | | | |
| in Vitro | | | | | | | |
| Bacterial Reverse Mutation Assay (e.g. Ames ...) | | | | | | | |
| Gene mutation | | | | | | | |
| Escherichia coli (2/4) | | | | | | | |
| Salmonella typhimurium | | | | | | | |
| No S9 Info | (87/92) | M: Positive | M: Positive | | M: Positive | | M: Positive |
| With S9 | (86/477) | M: Negative | M: Negative | | M: Negative | M: Negative | M: Negative |
| Without S9 | 111/537 | M: Negative | M: Negative | M: Negative | | M: Negative | M: Negative |
| In Vitro Mammalian Chromosome Aberration (11/24) | | M: Negative | M: Negative | | | M: Negative | |
| Mammalian Cell Gene Mutation Assay (5/5) | | M: Positive | | | | | |
| in Vivo (9/18) | | | | | | | |
| Immunotoxicity | | | | | | | |

Category Definition

Side-Bar of experimental data

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes icons for Input, Profiling, Data, Category definition (highlighted), Data Gap Filling, and Report. Below this, a 'Categorize' sub-menu offers options like 'Define', 'Define with metabolism', 'Subcategorize', and 'Combine'. The left sidebar shows a list of 'Organic functional groups' and 'Options' such as 'Select All', 'Unselect All', and 'Invert'. The main workspace shows a 'Filter endpoint tree...' with a table of endpoints. A 'Data points' dialog box is open, displaying a table of data points for 'Human Health Hazards;Genetic Toxicity'. The table has columns for 'Datapoints', '#', 'Value', 'Original value', and 'Assigned SMILES'. A red box highlights a cell in the 'Assigned SMILES' column containing 'M: Negative'. Below the dialog, a hierarchical tree view shows 'Escherichia coli' and 'Salmonella typhimurium' with associated data counts and values.

| Datapoints | # | Value | Original value | Assigned SMILES |
|---------------------------------------|---|-------------------------------|----------------------------|-----------------|
| Human Health Hazards;Genetic Toxicity | 1 | M: Negative (Gene mutation I) | Negative (Gene mutation I) | False |
| Human Health Hazards;Genetic Toxicity | 2 | M: Negative (Gene mutation I) | Negative (Gene mutation I) | False |

Double-click on the cell with measured data to see detailed information in drop down box.

Category Definition

Recap

- You have identified a category consisting of 127 analogous (“Aldehydes” by OFG classification) with the target chemical (n-hexanal).
- The available experimental data for these 127 similar chemicals are collected from the previously selected databases under Data section.
- The user can proceed with “Filling data gap” module, but before that he/she should navigate through the endpoint tree and find the gap that will be filled in.

Category Definition

Navigation through the endpoint tree

- The user can navigate through the data tree by closing or opening the nodes of the tree.
- In this example, results from genotox testing are available (see next screen shot).
- In this example to see does the target is mutagenic or not, it is recommended to check subsequently the two mutagenic endpoints:
 - Ames without S9
 - Ames with S9
- By clicking on the nodes of endpoint tree open the tree to the target: **Bacterial reverse mutation (Ames) assay without S9** (*i.e.*, *click* on *Human Health Hazards* then *click* on *Genetic Toxicity* followed by *in Vitro* and *Bacterial Reverse Mutation Assay (e.g. Ames Test), Gene Mutation, Salmonella typhimurium, Without S9*) (see next screen shot).

Category Definition

Navigation through the endpoint tree

The screenshot displays the QSAR Toolbox interface with the 'Filter endpoint tree...' window open. The tree is expanded to show the path: Genetic Toxicity > in Vitro > Bacterial Reverse Mutation Assay (e.g. Ames ...) > Gene mutation > Salmonella typhimurium. The right side of the interface shows a table of results for 7 different chemical structures.

| Structure | 1 [target] | 2 | 3 | 4 | 5 | 6 | 7 |
|--------------------------------------------------|---------------------|--------------------------|------------------------|-------------------------|------------------------|---------------------|------------------------|
| Structure | <chem>CCCC=O</chem> | <chem>O=Cc1ccccc1</chem> | <chem>O=C1C=CC1</chem> | <chem>CCCCCCCC=O</chem> | <chem>O=C1C=CC1</chem> | <chem>CCCC=O</chem> | <chem>O=C1C=CC1</chem> |
| Structure info | | | | | | | |
| Parameters | | | | | | | |
| Physical Chemical Properties | | | | | | | |
| Environmental Fate and Transport | | | | | | | |
| Ecotoxicological Information | | | | | | | |
| Human Health Hazards | | | | | | | |
| Acute Toxicity (1/1) | | | | | | | |
| Bioaccumulation | | | | | | | |
| Carcinogenicity | | | | | | | |
| Developmental Toxicity / Teratogenicity | | | | | | | |
| Genetic Toxicity | | | | | | | |
| in Vitro | | | | | | | |
| Bacterial Reverse Mutation Assay (e.g. Ames ...) | | | | | | | |
| Gene mutation | | | | | | | |
| Escherichia coli (2/4) | | | | | | | |
| Salmonella typhimurium | | | | | | | |
| No S9 Info (87/92) | | | | | | | |
| With S9 (86/477) | M: Negative | M: Negative | M: Positive | M: Negative | M: Positive | M: Negative | M: Positive |
| Without S9 (111/537) | M: Negative | M: Negative | | M: Negative | | M: Negative | M: Negative |
| In Vitro Mammalian Chromosome Aberration (11/24) | | M: Negative | | | | M: Negative | |
| Mammalian Cell Gene Mutation Assay (5/5) | | M: Positive | | | | | |
| in Vivo (9/18) | | | | | | | |
| Immunotoxicity | | | | | | | |

1. **Click** to Genetic Toxicity
2. **Click** to *in vitro*
3. **Click** to Bacterial Reverse Mutation Assay (e.g. Ames Test)
4. **Click** Gene Mutation
5. **Open** the tree to *Salmonella typhimurium*

Category Definition

Navigation through the endpoint tree

QSAR TOOLBOX

Input Profiling Data **Category definition** Data Gap Filling Report









Categorize: Define Define with metabolism Subcategorize Combine


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

Documents: Organic functional groups

Options: Select All Unselect All Invert

Filter endpoint tree...

| Structure | 1 [target] | 2 | 3 | 4 | 5 | 6 | 7 |
|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
|  |  |  |  |  |  |  |  |
| Structure info | | | | | | | |
| Parameters | | | | | | | |
| Physical Chemical Properties | | | | | | | |
| Environmental Fate and Transport | | | | | | | |
| Ecotoxicological Information | | | | | | | |
| Human Health Hazards | | | | | | | |
| Acute Toxicity (1/1) | | | | | | | |
| Bioaccumulation | | | | | | | |
| Carcinogenicity | | | | | | | |
| Developmental Toxicity / Teratogenicity | | | | | | | |
| Genetic Toxicity | | | | | | | |
| In Vitro | | | | | | | |
| Bacterial Reverse Mutation Assay (e.g. Ames ...) | | | | | | | |
| Gene mutation | | | | | | | |
| Escherichia coli (2/4) | | | | | | | |
| Salmonella typhimurium | | | | | | | |
| No S9 Info (87/92) | M: Positive | | M: Positive | | M: Positive | | M: Positive |
| With S9 (86/477) | M: Negative | M: Negative | | M: Negative | | M: Negative | M: Negative |
| Without S9 (111/537) | M: Negative | M: Negative | | M: Negative | | M: Negative | M: Negative |
| In Vitro Mammalian Chromosome Aberration(11/24) | M: Negative | | | | | | |
| Mammalian Cell Gene Mutation Assay (5/5) | M: Positive | | | | | | |
| In Vivo (9/18) | | | | | | | |
| Immunotoxicity | | | | | | | |

 In order to examine the target endpoint "Ames without S9", select the cell as shown.

Category Definition

Recap

- You have now retrieved the available experimental data on genetic toxicity for 111 chemicals classified as “Aldehydes” by OFG, found in the databases containing mutagenicity data.
- 537 experimental data for 111 chemicals have been found.
- You are now ready to fill in the data gap and trying to reproduce the experimental data of the target.
- In this example with qualitative mutagenicity data we can only use read-across.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - Profiling
 - Data
 - Category definition
 - **Data Gap Filling**
 - **Ames without S9**

Data Gap Filling

Preparing for Model domain save

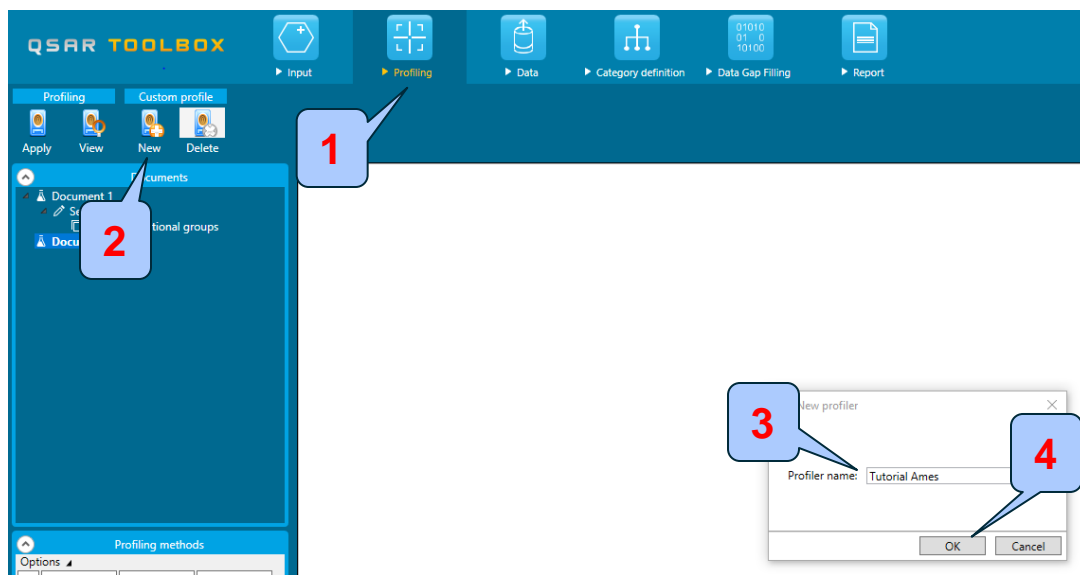
In order to save model and export data for the analogues in Read-across analysis the user should preliminary create a custom profile (see this and the next few files)

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. Below the menu bar, there are tabs for 'Document', 'Single Chemical', 'Chemical List', 'Search', and 'Target Endpoint'. The 'Document' tab is active, showing a 'Documents' panel on the left with 'Document 1' and 'Search chemical' options. The main workspace is divided into a 'Filter endpoint tree...' on the left, a 'Structure' view in the center, and a table of chemical structures on the right. A 'Define document name' dialog box is open in the foreground, with the 'Name' field set to 'Document 2'. The dialog box has 'OK' and 'Cancel' buttons. Four callout boxes with red numbers 1 through 4 indicate the steps: 1. Click 'Input' in the menu bar; 2. Click 'New' in the Document tab; 3. Enter a name in the 'Define document name' dialog box; 4. Click 'OK' in the dialog box.

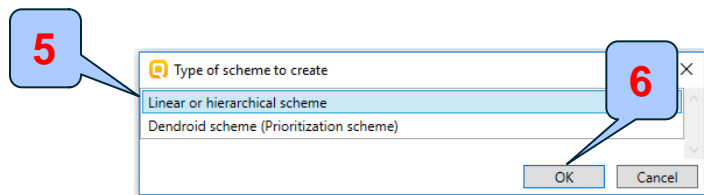
1. Go to Input; 2. Click New; 3. Give a name to the new document; 4. Click OK

Data Gap Filling

Preparing for Model domain save



1. **Go** to Profiling; 2. **Click** New (Custom profile); 3. **Give** a name to the new profile; 4. **Click** OK;



5. For save domain purposes "Linear or hierarchical scheme" should be chosen; 6. **Click** OK;

Data Gap Filling

Preparing for Model domain save

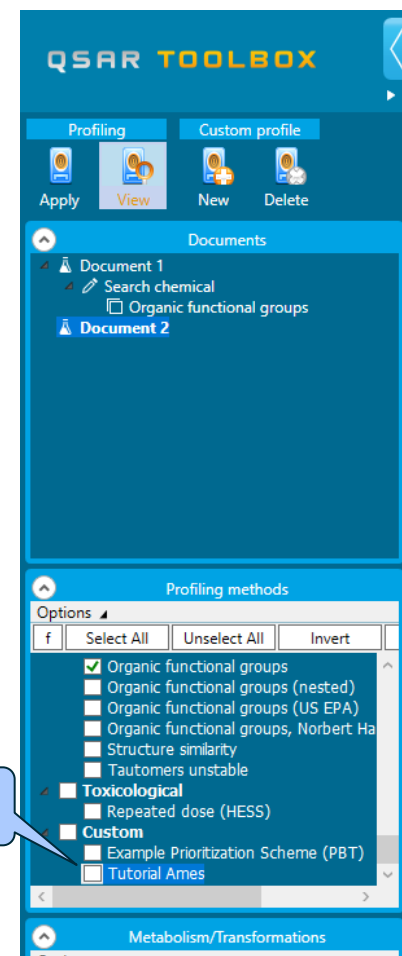
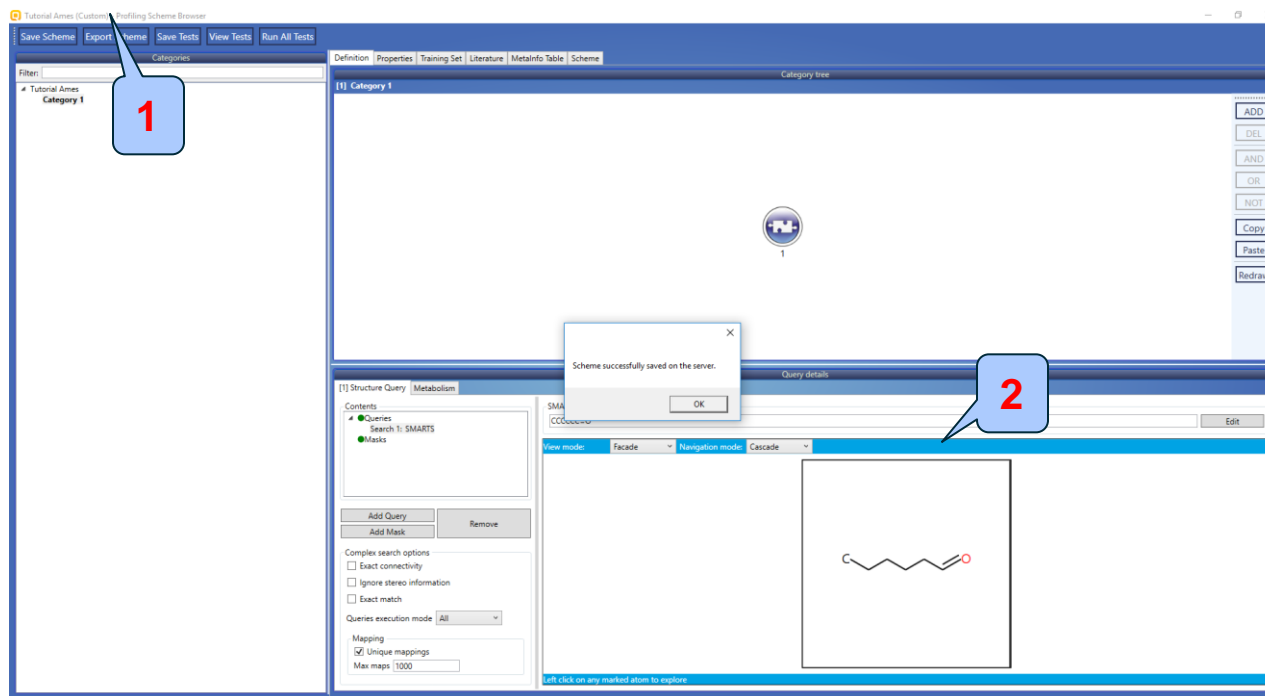
The image shows a sequence of three screenshots from the QSAR Toolbox software, illustrating the steps to add a SMARTS query. The steps are numbered 1 through 7:

- Click Add; 2. Choose Structure query; 3. Click Add query; 4. Radio button should be settled to SMART; 5. Click OK; 6. Mark Search 1: SMART; 7. Click Edit

1. **Click Add**; 2. **Choose Structure query**; 3. **Click Add query**; 4. Radio button should be settled to SMART; 5. **Click OK**; 6. **Mark Search 1: SMART**; 7. **Click Edit**

Data Gap Filling

Preparing for Model domain save



In the SMART Editor **input** the target chemical (see slide 24);

1. **Click** Save scheme;
2. **Click** OK on the message evidencing scheme saving;
3. The newly created profile appears in the Custom section of Profiling methods;

Data Gap Filling (Ames without S9)

Apply read-across

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling operation. The top menu bar shows the 'Data Gap Filling' option selected. The left sidebar contains 'Data Gap Filling Settings' with options for 'Only endpoint relevant' and 'Only chemical relevant'. The central panel shows a 'Filter endpoint tree...' with 'Ames without S9' highlighted. A 'Possible data' dialog box is open, listing various strains with checkboxes for selection. The main data grid shows chemical structures and associated data points.

1. **Click** on Data Gap Filling; 2. The data endpoint box corresponding to Ames without S9 under the target chemical is already **highlighted**; 3. **Click** Read across button; 4. **Choose** desired strains; 5. **Click** OK.

Data Gap Filling (Ames without S9)

Results of Read across

QSAR TOOLBOX

Gap Filling Workflow: Trend analysis, Read across (QSAR), Standardized, Automated

Documents: ent 1, rich chemical, Organic functional groups, Enter GF (RA) with 32 chemicals, 199 data point

Filter endpoint tree...

- GENERIC TOXICITY
 - in Vitro
 - Bacterial Reverse Mutation Assay (e.g. Ames ...
 - Gene mutation
 - Escherichia coli (2/4)
 - Salmonella typhimurium (24)
 - No S9 Info (34/39)
 - With S9 (47/199)
 - Without S9 (52/199)
 - in Vitro Mammalian Chromosome Aberration(9/20)
 - Mammalian Cell Gene Mutation Assay (5/5)
 - in Vivo (6/14)
 - Immunotoxicity
 - Irritation / Corrosion
 - Neurotoxicity
 - Photoinduced toxicity
 - Repeated Dose Toxicity

| Structure | 1 [target] | 2 | 4 | 6 | 8 | 11 | 12 | 13 | 15 | 17 |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | | | | | | | | | |
| | M: Positive | M: Negative | M: Negative | M: Negative | M: Positive | M: Negative | M: Negative | M: Positive | M: Negative | M: Negative |
| | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| | M: Negative | M: Positive | | M: Negative | | | | | | M: Negative |

Data Gap Filling Settings

- Only endpoint relevant
- Only chemical relevant

At this position:

- Select a cell with a rigid (bold) path
- Automated workflows 0
- Standardized workflows 0

Descriptors

Prediction

Read-across prediction for Gene mutation, based on 22 values

Observed: Negative; Predicted: Negative

Active descriptor X: log Kow

- Select / Filter data
- Gap filling approach
- Descriptors / data
- Model/QSAR
- Calculation options
- Visual options
- Information
- Miscellaneous

✔ Accept prediction

52

The OECD QSAR Toolbox for Grouping Chemicals into Categories

July 2017

69

Data Gap Filling (Ames without S9)

Interpreting Read-across

- The resulting plot outlines the experimental Ames results of all analogues (Y axis) according to a descriptor (X axis). Note, Log Kow is on the X-axis; while this descriptor is not significant to Ames data, it is the default descriptor for data gap filing (see next screen shot).
- The **RED** dot represents the predicted value for target chemical (see next screen shot).
- The **PURPLE** dots represent the observed value for the target neighbours (analogues) used for read-across (see next screen shot).
- The **BLUE** dots represent the experimental results available for the analogues but not used for read-across (see next screen shot).
- Please note **LIGHT BLUE** dots (which you will see shortly) represent analogues belonging to different subcategories.

Data Gap Filling (Ames without S9)

Interpretation of the Read across

- Seven of the analogues are mutagenic in the Ames assays without S9, the rest analogues are non-mutagenic
- Non-mutagenic potential (Negative) is, therefore, predicted with confidence for the target chemical.
- However, before data gap filling it is recommended to check the similarity of the analogues used in the prediction (see next screen shot). This is performed in order to assure the category consists of analogues that are both mechanistically and structurally similar.

Data Gap Filling (Ames without S9)

Subcategorization by DNA binding by OASIS (endpoint specific)

The screenshot displays the 'Subcategorization' window in the QSAR Toolbox. On the left, the 'Endpoint Specific' list includes 'DNA alerts for AMES by OASIS', which is highlighted by callout 3. Below this, a list of alerts is shown, with callout 4 pointing to '(1) AN2 >> Schiff base formation by aldehyde formed after metabolic activation'. At the bottom of this list, callout 5 points to the 'Remove selected' button. On the right, a table shows predicted mutation outcomes (M: Negative, M: Positive, M: Inconclusive) for various chemical structures. Callout 1 points to the 'Select / Filter data' menu, and callout 2 points to the 'Subcategorize' option. At the bottom, a scatter plot titled 'Read-across prediction for Gene mutation, based on 22 values' shows 'Gene mutation' on the y-axis (Positive, Equivocal, Negative) and 'log Kow' on the x-axis (ranging from -0.5 to 4.5). Callout 4 also points to a cluster of points in the 'Positive' region of the plot.

1. **Click** Select / filter data; 2. **Select** Subcategorize 3. **Select** DNA alerts for AMES by OASIS (note the same highlighting according profiler relevancy is available) 4. Examine dissimilar chemicals – coloured in light blue on the plot 5. **Click** Remove selected.

Data Gap Filling (Ames without S9) Subcategorization by OFG (US-EPA)

The screenshot displays the QSAR Toolbox interface for subcategorization. The 'Subcategorization' window is active, showing the 'Adjust options' tab where 'Aliphatic Carbon [-CHO]' is selected. The 'Options' tab shows 'Organic functional groups (US-EPA)' selected. The 'Read-across prediction' plot shows 'Gene mutation' vs 'log Kow' with three positive outliers. The 'Select / filter data' panel is visible on the right.

There are three positive outliers which are quite dissimilar by the target and could be eliminated by OFG (US-EPA)

1. **Select** Subcategorize; 2. **Select** OFG (US-EPA); 3. **Click** Remove selected

Data Gap Filling (Ames without S9)

Interpretation of the Read across

The screenshot shows the QSAR Toolbox interface during a 'Data Gap Filling' workflow. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Documents' and 'Data Gap Filling Settings' with checkboxes for 'Only endpoint relevant' and 'Only chemical relevant'. The main workspace is divided into a 'Filter endpoint tree...' on the left, a 'Read-across prediction for Gene mutation' plot in the center, and a 'Select / filter data' panel on the right. The plot shows a scatter of points with 'log Kow' on the x-axis and 'Gene mutation' on the y-axis. A red callout box with the number '1' points to the 'Accept prediction' button in the bottom right corner of the plot area.

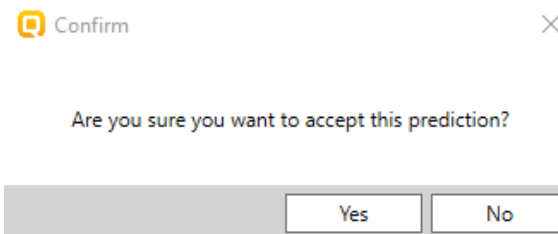
Now all analogues are structurally similar (Aldehydes) and negative by the experimental data. The prediction could be accepted by:

1. **Click** on "Accept prediction";

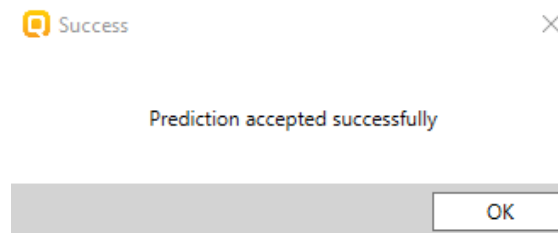
Data Gap Filling (Ames without S9)

Interpretation of the Read across

- A message appears:



- Clicking on No allows to continue with the subcategorization
- Click Yes to accept prediction



Data Gap Filling (Ames without S9) Results

- By accepting the prediction the data gap is filled.
- The read-across for the current endpoint is finished.
- The screen is returned back to the Matrix
- The user can proceed with the workflow for the second endpoint, which in this case will be “Ames with S9” (see next screen shot).

Data Gap Filling (Ames without S9) Results

The screenshot displays the QSAR Toolbox interface for Data Gap Filling. The main window shows a grid of results for 10 different endpoints. The left sidebar contains a 'Filter endpoint tree...' and 'Data Gap Filling Settings'. The settings are configured to show only endpoint and chemical relevant data. The main table shows results for various endpoints, including Acute Toxicity, Bioaccumulation, Carcinogenicity, and Genetic Toxicity. Two callout boxes highlight specific results: '1' points to a prediction of 'M: Negative' for the 'in Vitro Mammalian Chromosome Aberration' endpoint, and '2' points to a data gap for the 'Bacterial Reverse Mutation Assay (e.g. Ames ...)' endpoint.

1. This is the prediction for the first endpoint.
2. This is the data gap for the second endpoint.

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - Profiling
 - Data
 - Category definition
 - **Data Gap Filling**
 - Ames without S9
 - **Ames with S9**

Data Gap Filling (Ames with S9)

- Do this the same way as with Ames without S9.
- Make sure **Data Gap Filling** is highlighted.
- Highlight the **data endpoint box**; Now it should correspond to **Ames with S9**.
- Select **Read across**.
- Choose desired strains.

Data Gap Filling (Ames with S9)

Results of Read across

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

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

Documents

Document 1
Search chemical
Organic functional groups
Enter GF(RA) with 52 chemicals, 199 data point
Ch: 47|Data: 182 Subcategorized: DNA aler
Ch: 10|Data: 33 Subcategorized: Organ
Enter GF(RA) with 48 chemicals, 200 data p

Filter endpoint tree...

Structure

| 1 [target] | 2 | 4 | 6 | 8 | 11 | 12 | 13 | 15 | 17 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | | | | | | | | | |
| M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| | M: Negative | | | | | | | | |
| | M: Positive | | | | | | | | |

With S9 (48/200)
Without S9 (47/185)
In Vitro Mammalian Chromosome Aberration(9/20)
Mammalian Cell Gene Mutation Assay (4/4)
in Vivo (6/14)
Immunotoxicity
Irritation / Corrosion
Neurotoxicity
Photoinduced toxicity
Repeated Dose Toxicity
Sensitisation
ToxCast
Toxicity to Reproduction
Toxicokinetics, Metabolism and Distribution
Profile
Predefined

AW SW AOP

Data Gap Filling Settings

Only endpoint relevant
 Only chemical relevant

At this position:
Select a cell with a rigid (bold) path
Automated workflows 0
Standardized workflows 0

Descriptors
Prediction

Read-across prediction for Gene mutation, based on 26 values
Observed: Negative; Predicted: Negative

Active descriptor X log Kow

Select / filter data
Gap filling approach
Descriptors / data
Model/QSAR
Calculation options
Visual options
Information
Miscellaneous

Accept prediction

Data Gap Filling (Ames with S9)

Results of Read across

- As with Ames without S9, before accepting the estimated result for the target chemical, by read-across the user should refined the category by subcategorization.
- Subcategorization refers to the process of applying additional profilers to the previously defined category, identifying chemicals which have differing profiling results and eventually eliminating these chemicals from the category.
- In this example, we are going to use several different profilers to repeatedly subcategorise the data set.

Data Gap Filling (Ames with S9) Side Bar of Subcategorization

The analogues which are dissimilar to the target chemical with respect to:

- **DNA binding alerts (endpoint specific) taking into account liver metabolism** – The categorization based on this profiler identifies analogues having same DNA binding alerts as the target after metabolic activation
- **Organic functional groups (US-EPA)** – The categorization based on this profiler identifies analogues having the same organic functional groups

can be removed from the initial list of analogues previously defined by OFG.

Data Gap Filling (Ames with S9)

Subcategorization by DNA alerts taking into account liver metabolism

- As with Ames without S9, we want to refined the category by subcategorization with DNA binding by OASIS, taking into account liver metabolism
- Select **Select/filter data**
- Select **Subcategorize**
- Select **DNA binding alert**
- Select **Rat Liver S9 metabolism simulator**
- Look for dissimilar chemicals
- Click **Remove** to eliminate dissimilar chemical.

Data Gap Filling (Ames with S9)

Subcategorization by DNA binding alerts taking into account Rat liver metabolism

The screenshot displays the QSAR Toolbox interface during a subcategorization process. The 'Subcategorization' window is open, showing a list of alert categories. Callout 3 points to 'DNA alerts for AMES by OASIS'. The 'Adjust options' window is also open, showing a list of alerts. Callout 5 points to the 'Remove selected' button. The 'Read-across prediction' plot shows 'Gene mutation' on the y-axis and 'log Kow' on the x-axis. Callout 2 points to the 'Subcategorize' button in the 'Select / filter data' panel. Callout 1 points to the 'Subcategorize' button in the 'Select / filter data' panel. Callout 4 points to the 'Observed Rat Liver S9 metabolism simulator' in the bottom window.

1. **Select** Select/Filter data 2. **Click** Subcategorize 3. **Select** DNA alerts for AMES by OASIS 4. **Select** Rat liver metabolism simulator from the bottom window 5. **Click** Remove selected

Data Gap Filling (Ames with S9) Subcategorization by OFG (US-EPA)

- As with Ames without S9, we want to refined the category by subcategorization with OFG (US-EPA)
- Select **Select/filter data**
- Select **Subcategorize**
- Select **Organic functional groups (US-EPA)**
- Unselect **Do not account metabolism**
- Look for dissimilar chemicals
- Click **Remove** to eliminate dissimilar chemical.

Data Gap Filling (Ames with S9) Subcategorization by OFG (US-EPA)

The screenshot shows the 'Subcategorization' window with three numbered callouts:

- 1**: Points to the 'Organic functional groups (US EPA)' option in the 'Empiric' section of the 'Options' pane.
- 2**: Points to the 'Do not account metabolism' option in the 'Simulated' section of the 'Options' pane.
- 3**: Points to the 'Remove selected' button in the 'Adjust options' dialog box.

The 'Adjust options' dialog shows a list of categories with 35 items selected (9/44). The 'Differ from target by' options are 'At least one category' and 'All categories'. The 'Read-across prediction for Gene mutation, based on 26 values' plot shows 'Observed: Negative; Predicted: Negative'.

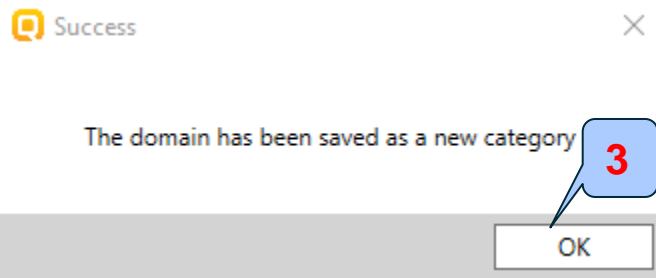
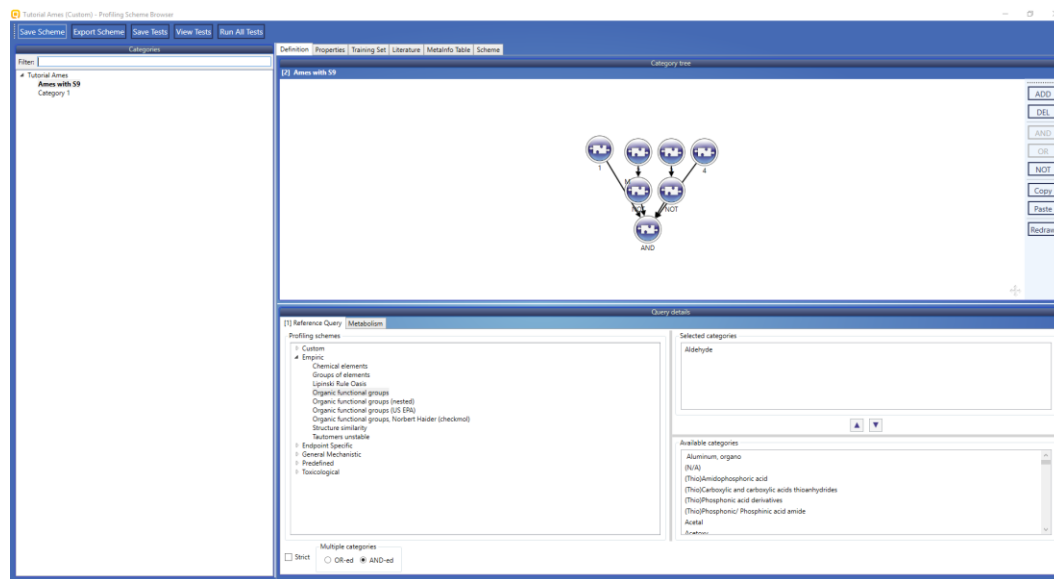
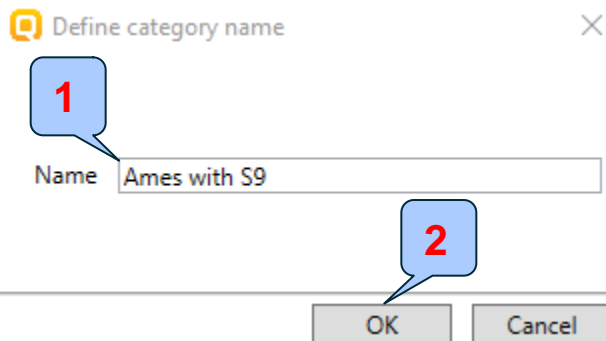
| log Kow | Gene mutation |
|---------|---------------|
| -0.2 | Negative |
| 0.4 | Positive |
| 0.8 | Negative |
| 1.0 | Negative |
| 1.2 | Negative |
| 1.3 | Negative |
| 1.4 | Negative |
| 1.5 | Negative |
| 1.6 | Negative |
| 1.7 | Negative |
| 1.8 | Negative |
| 1.9 | Negative |
| 2.0 | Negative |
| 2.1 | Negative |
| 2.2 | Negative |
| 2.3 | Negative |
| 2.4 | Negative |
| 2.5 | Negative |
| 2.6 | Negative |
| 2.7 | Negative |
| 2.8 | Negative |
| 2.9 | Negative |
| 3.0 | Positive |
| 3.1 | Negative |
| 3.2 | Negative |
| 3.3 | Negative |
| 3.4 | Negative |
| 3.5 | Negative |
| 3.6 | Negative |
| 3.7 | Negative |
| 3.8 | Negative |
| 3.9 | Negative |
| 4.0 | Negative |
| 4.1 | Negative |
| 4.2 | Negative |
| 4.3 | Negative |
| 4.4 | Negative |
| 4.5 | Negative |

1. **Select** OFG (US-EPA); 2. **Click** on Do not account metabolism; 3. **Click** Remove selected; Finally close Subcategorization window

Data Gap Filling (Ames with S9) Result of read-across

1. A blue helper appears in top right of the screen. By clicking on it the message can be read. Now all 6 analogues are structurally and mechanistically similar, then the prediction could be accepted or saved as a category (domain) in the custom profiler, which could be used further for screening purposes. This could be done by
2. **Deselect** Select / filter data 3. **Click** on Model/(Q)SAR; 4. **Click** on Save domain as category 5. Since a custom profiler has previously been defined, **highlight** custom profiler and 6. **Click** OK.

Data Gap Filling (Ames with S9) Result of read-across



1. **Type** a name of the category in the "Name" box; 2. **Click** OK; 3. **Click** OK on the message evidencing domain saving; 4. **The result** can be seen through Profiling / Profiling methods / Right click on the Custom profile / View scheme. The individual steps are presented in different boundaries.

Data Gap Filling (Ames with S9)




Result of read-across

The screenshot displays the QSAR Toolbox interface during a Data Gap Filling operation. The central table shows a grid of chemical structures and their predicted outcomes for various endpoints. A 'Confirm' dialog box is overlaid on the table, asking 'Are you sure you want to accept this prediction?'. A 'Success' message box is also present, indicating 'Prediction accepted successfully'. The bottom right corner features a 'Accept prediction' button, and the bottom center has a plot showing 'Gene mutation' vs 'log Kow'.

1. **Click** Accept prediction
2. Click **Yes** if you want to Accept prediction
3. Click **No** if you decided to continue Subcategorization.
4. **Click** OK on the message

Data Gap Filling (Ames with S9)

Result of read-across

| Filter endpoint tree... | 1 [target] | 4 | 6 |
|------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Structure |  |  |  |
| Gene mutation | | | |
| Salmonella typhimurium | | | |
| No S9 Info (6/7) | | | |
| With S9 (10/38) | M: Negative R: Negative | M: Negative | M: Negative M: Negative |
| Without S9 (10/34) | M: Negative R: Negative | M: Negative | M: Negative M: Negative |
| in Vitro Mammalian Chromosome Aberration (2/5) | | | M: Negative |
| Mammalian Cell Gene Mutation Assay (1/1) | | | |
| in Vivo (2/7) | | | |
| Immunotoxicity | | | |
| Irritation / Corrosion | | | |

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- **Workflow of the exercise**
 - Chemical input
 - Profiling
 - Data
 - Category definition
 - Data Gap Filling
 - **Report**

Report Overview

- Report module could generate *Prediction report* and *Data matrix report* for each of predictions performed with the Toolbox.
- It is necessary just the cell with the desired prediction to be marked and then to click on the respective button.
- Report module contains predefined reports. The user can choose how to customize them.
- The *Prediction* and *Data matrix report* can be saved respectively in PDF / MS Excel formats. (see next screen shot).

Report Generate Report

The screenshot shows the QSAR Toolbox interface with the 'Report' menu highlighted. A callout '1' points to this menu. Below, a prediction table is visible with a callout '2' pointing to a cell containing 'M: Negative' and 'R: Negative'. The table also shows chemical structures and target information.

| Structure | 1 [target] | 4 | 6 | 12 | 35 |
|------------------------------------------------|----------------------------|-------------|-------------|-------------|-------------|
| Salmonella typhimurium | | | | M: Negative | M: Negative |
| No S9 info (6/7) | | | | M: Negative | M: Negative |
| With S9 (10/38) | M: Negative R: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| Without S9 (10/34) | M: Negative R: Negative | M: Negative | M: Negative | M: Negative | M: Negative |
| In Vitro Mammalian Chromosome Aberration (2/5) | | | | M: Negative | M: Negative |
| Mammalian Cell Gene Mutation Assay (1/1) | | | | M: Negative | M: Negative |
| In Vivo (2/7) | | | | M: Negative | M: Negative |
| Immunotoxicity | | | | | |
| Irritation / Corrosion | | | | | |

1. **Select** Report;
2. **Mark** the desired prediction cell;
3. **Click** Prediction;
4. **Customize** the report;
5. **Choose** type of the report;

The screenshot shows the 'Customize report content and appearance' dialog box. A callout '4' points to the 'Wizard pages' section on the left. The main area shows a list of sections to be included in the report, with checkboxes for each.

Wizard pages

- Customize report
- Target and prediction summary
- Prediction details
- Prediction details (II)
- Target profiles
- Analogues selection details
- Data for analogues
- Appendix: Grouping / subcategorization
- Appendix: Data pruning

Selected sections (checked):

- Target and prediction summary
- Prediction details
- Prediction details (II)
- Target profiles
- Analogues selection details
- Data for analogues
- Appendix: Grouping / subcategorization
- Appendix: Data pruning

Buttons: Move Up, Move Down, Back, Next, Cancel, Create report

The screenshot shows the 'Generated report files' dialog box. A callout '6' points to the 'Prediction report' file in the list. The dialog also shows a 'Data matrix' file and a 'PDF file containing the prediction report'.

The following files were generated.
Select a file to open or save.

- Prediction report
- Data matrix
- PDF file containing the prediction report

Buttons: Open, Save as

Report

Generated report files


Prediction of Gene mutation for Hexanal

1 / 10

QSAR Toolbox prediction for single chemical

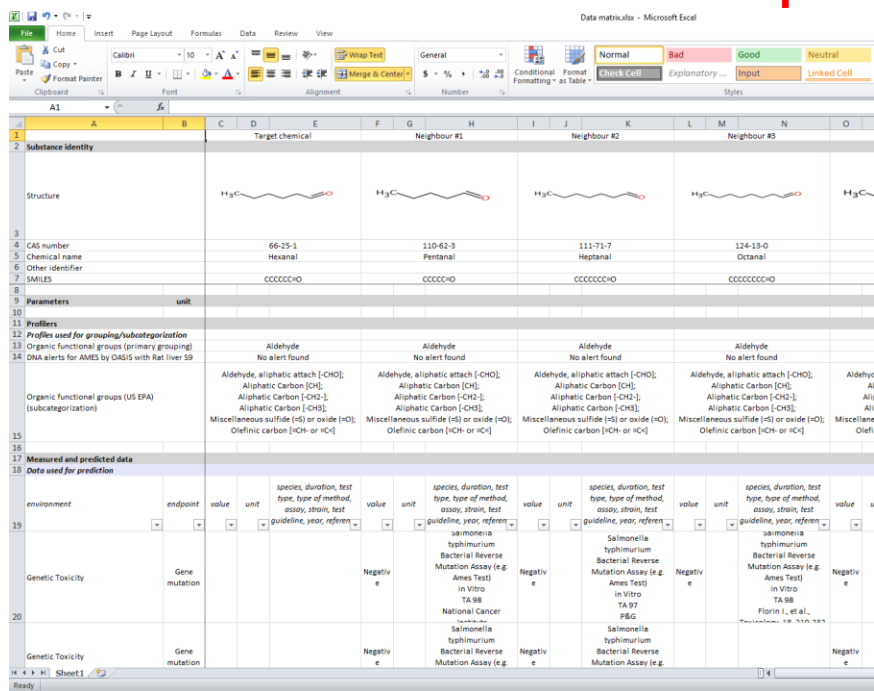
Date: 19 Jul 2017
 Author(s):
 Contact details:





Prediction report

| Target information | | |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|-----------------------------------------|
| Structural Information | Numerical Identifiers | Chemical names |
| SMILES: <chem>CCCCCC=O</chem> Structure  | EC#: N/A CAS#: 66-25-1 Other: N/A | hexaldehyde Hexanal Hexylaldehyde |

| Prediction summary |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Predicted endpoint: Gene mutation; No effect specified; Salmonella typhimurium; No duration specified; No guideline specified Predicted value: Negative Unit/scale: Gene mutation I Data gap filling method: Read-across analysis Summary: manually editable field Not provided by the user |

Data matrix report



| | Target chemical | Neighbour #1 | Neighbour #2 | Neighbour #3 | | | | | | | | | | | |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|------|---------------------------------------------------------------------------------------------|----------|------|-----------------------------------------------------------------------------------------------------------|----------|------|-----------------------------------------------------------------------------------------------------------|----------|------|
| Substance Identity | | | | | | | | | | | | | | | |
| Structure |  |  |  |  | | | | | | | | | | | |
| CAS number | 66-25-1 | 110-62-3 | 111-71-7 | 124-13-0 | | | | | | | | | | | |
| Chemical name | Hexanal | Pentanal | Heptanal | Octanal | | | | | | | | | | | |
| Other identifier | | | | | | | | | | | | | | | |
| SMILES | <chem>CCCCCC=O</chem> | <chem>CCCCC=O</chem> | <chem>CCCCCCC=O</chem> | <chem>CCCCCCCC=O</chem> | | | | | | | | | | | |
| Parameters | unit | | | | | | | | | | | | | | |
| Profiles | | | | | | | | | | | | | | | |
| Organic functional groups (US EPA) | Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [-CH-]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (S) or oxide (O); Olefinic carbon [-CH= or =C-] | Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [-CH-]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (S) or oxide (O); Olefinic carbon [-CH= or =C-] | Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [-CH-]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (S) or oxide (O); Olefinic carbon [-CH= or =C-] | Aldehyde, aliphatic attach [-CHO]; Aliphatic Carbon [-CH-]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Miscellaneous sulfide (S) or oxide (O); Olefinic carbon [-CH= or =C-] | | | | | | | | | | | |
| Measured and predicted data | | | | | | | | | | | | | | | |
| environment | endpoint | value | unit | species, duration, test type, type of method, assay, strain, test guideline, year, referen | value | unit | species, duration, test type, type of method, assay, strain, test guideline, year, referen | value | unit | species, duration, test type, type of method, assay, strain, test guideline, year, referen | value | unit | species, duration, test type, type of method, assay, strain, test guideline, year, referen | value | unit |
| Genetic Toxicity | Gene mutation | | | Salmonella typhimurium Bacterial Reverse Mutation Assay (e.g. Ames Test) In Vitro TA 98 National Cancer | Negative | | Salmonella typhimurium Bacterial Reverse Mutation Assay (e.g. Ames Test) In Vitro TA 98 P&G | Negative | | Salmonella typhimurium Bacterial Reverse Mutation Assay (e.g. Ames Test) In Vitro TA 98 Florin I, et al., | Negative | | Salmonella typhimurium Bacterial Reverse Mutation Assay (e.g. Ames Test) In Vitro TA 98 Florin I, et al., | Negative | |

Outlook

- Background
- Objectives
- Specific Aims
- Read-across
- The exercise
- Workflow of the exercise
- **Save the prediction**

Saving the prediction result

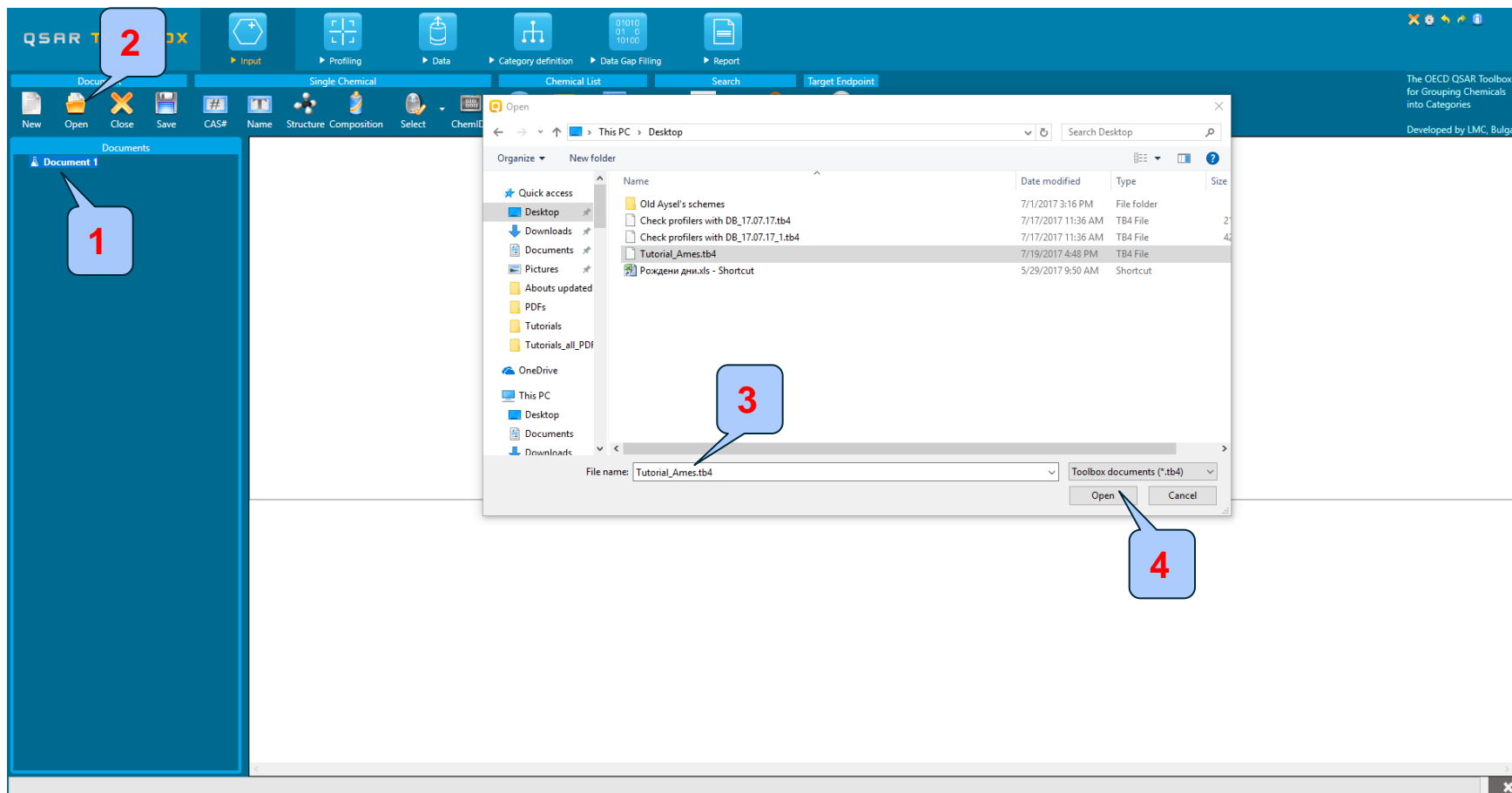
- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document 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

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Document', 'Single Chemical', 'Chemical List', 'Search', and 'Target Endpoint'. The toolbar contains icons for 'New', 'Open', 'Close', 'Save', 'CASF', 'Name', 'Structure', 'Composition', 'Select', 'ChemIDs', 'Database', 'Inventory', 'List', 'Substructure (SMARTS)', 'Query', and 'Define'. The left sidebar shows a document list with a 'Save' button highlighted by callout 2. The main workspace shows a chemical structure of Salmonella typhimurium and a prediction plot for 'Gene mutation' vs 'log Kow'. A 'Save document 'Document 1'' dialog box is open, showing the file name field and 'Save' button highlighted by callout 5. A 'Do you want to save changes to document 'Document 1'?' dialog box is also visible, with 'Yes' highlighted by callout 3. A file explorer window is open, showing the file name field highlighted by callout 4. The bottom right corner shows a 'Calculation options' panel with 'Accept prediction' checked.

1. **Go** to Input section;
2. **Click** on Save button;
3. **Click** OK;
4. **Define** name of the file;
5. **Click** Save button (Message "File saved successfully")

Open saved file



1. **Create** new document; 2. **Click** Open; 3. **Find** and **select** file; 4. **Click** Open;

Open saved file

The screenshot displays the QSAR Toolbox interface. The top menu bar includes Document, Single Chemical, Chemical List, Search, and Target Endpoint. The left sidebar shows a file tree with documents like 'Ames.tb4' and 'Enter GF(RA) with 52 chemicals, 199 data points'. The main workspace contains a 'Filter endpoint tree...' on the left and a table of chemical structures on the right. Two dialog boxes are overlaid: 'File open' (labeled '1') and 'Consistency check Info: OK' (labeled '2'). Below the table is a scatter plot titled 'Read-across prediction for Gene mutation, based on 22 values' with 'Observed: Negative; Predicted: Negative'. The plot shows data points for 'Gene mutation' (Positive, Equivocal, Negative) against 'log Kow' (ranging from -0.5 to 4.5). A legend on the right side of the plot includes 'Select / filter data', 'Gap filling approach', 'Descriptors / data', 'Model/QSAR', 'Calculation options', 'Visual options', 'Information', and 'Miscellaneous'. A green checkmark and the text 'Accept prediction' are visible at the bottom right of the plot area.

The file is opened successfully 1. **Click** OK; 2. **Click** OK;

Congratulation

- By now you should feel comfortable with the six basic modules of the Toolbox and how they form the work flow of the Toolbox.
- In this tutorial you have now been introduced to several additional functions in the Toolbox, especially using different profilers in subcategorizing the category of the target chemical.
- Remember proficiency only comes with practice.