

OECD QSAR Toolbox v.3.4

Step-by-step example of how to predict Ames mutagenicity for a chemical by a qualitative read-across approach

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 2.

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

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes options like Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. Below the menu bar, there are tabs for Document, Single Chemical, and Chemical List. The 'Structure' button is highlighted with a red box and a callout '1'. The '2D Editor' window is open, showing a grid of chemical fragments for selection. The fragments include a benzene ring, a cyclohexane ring, a cyclopentane ring, a cyclobutane ring, a cyclopropane ring, and a cyclohexane ring with a double bond. The '2D Editor' window also has a toolbar with various drawing tools and a 'Draw' button. The '2D Editor' window title is '2D Editor' and it contains fields for SMILES/InChI and a 'Draw' button. The '2D Editor' window also has a 'Mixture' button and an 'Edit names' button. The '2D Editor' window also has a 'drag the mouse with left button pressed to create bond' instruction and 'OK' and 'Cancel' buttons.

1. Click on **Structure**

Chemical Input

Input target chemical by SMILES

- In the **Aqua-coloured** area next to "SMILES/InChi" **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

The screenshot shows the '2D Editor' window. At the top, the 'SMILES/InChi' field contains the text 'CCCCCC=O', which is highlighted in cyan and labeled with a red '1'. Below this, a 'Templates' panel on the left shows various chemical structures. The main workspace on the right displays the 2D skeletal structure of hexanal, which is circled in red and labeled with a red '2'. At the bottom of the workspace, the 'OK' button is highlighted with a green checkmark and labeled with a red '3'. The 'Cancel' button is also visible. The text 'drag the mouse with left button pressed to create bond' is visible at the bottom left of the workspace. The URL 'oasis-lmc.org' is visible at the bottom right of the workspace.

1. **Type** CCCCCCC=O in SMILES/InChi 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 the chemical with structure related to the current SMILES notation. It is depicted as a 2D image.

Two chemicals are found. All found chemicals are selected by default.

The screenshot shows a 'Select chemicals' dialog box with the following data:

Selected	CAS	Smiles	Depiction	Names	CAS/Name	2D/Name	CAS/2D
1. Yes	66-25-1	CCCCCC=O		1: hexa 2: hexa 3: hexyl	1:: High Qualit 1:: Aquatic 2:: Biodegr 3:: Canada 4:: DSSTO 5:: ECHA P 6:: ECOTO 7:: FINECS	1:: High Qualit 1:: USER D 2:: USER D 3:: TSCA 4:: ECHA P 5:: Aquatic 6:: REACH 7:: NTCNAS	1:: High Qualit 1:: Ac 2:: Ac 3:: Bi 4:: Ca 5:: D 6:: EC 7:: F
2. Yes	110-62-3	CCCCCC=O		1: valer	1:: High Qualit 1:: Acute C 2:: Aquatic 3:: ECHA C 4:: ECHA P 5:: GSH Ex 6:: Genotox 7:: HPVC O	1:: Low Quality 1:: GSH Ex	1:: Low Quality 1:: G

Callout 1 points to the 'Yes' checkbox in the 'Selected' column for the second row. Callout 2 points to the 'OK' button at the bottom right of the dialog box.

1. **Unselect** the second chemical by clicking on the "Yes"; 2. **Click** OK.

Chemical Input

Target chemical identity

- You have now selected your target chemical.
- **Click** on the box next to "Substance Identity"; this displays the 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 toolbar contains several modules: 'Input', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Input' module is currently active, and a red box with the number '1' highlights the 'Input' button. The main window shows a 'Documents' panel on the left with a document named 'Document_1' containing the SMILES string 'CCCCCC=O'. The central panel displays the 'Substance Identity' section, which is expanded to show various identifiers. A red circle highlights the 'EINECS:2006245' entry, which is associated with the chemical name 'hexanal' and the SMILES string 'CCCCCC=O'. The bottom status bar shows '1 Document_1' and '1/0/0'.

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 show 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.
- For this example, the following general mechanistic profiling methods are relevant to genetic toxicity:
 - DNA binding by OASIS v1.4– mechanistic grouping
 - DNA binding by OECD – mechanistic grouping
 - Protein binding by OASIS v1.4 – mechanistic grouping
 - Protein binding by OECD – mechanistic grouping
 - Carcinogenicity (genotox and nongenotox) alerts by ISS - endpoint specific
 - DNA alerts for AMES by OASIS v.1.4 - 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 software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing options like 'Apply', 'New', 'View', and 'Delete'. The 'Apply' button is circled in red. Below the menu, the 'Profiling methods' list is visible, with the 'Endpoint Specific' section highlighted by a red bracket and a callout box labeled '1'. The 'Endpoint Specific' section includes various endpoints such as '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 (BioWIN MITI)', 'Carcinogenicity (genotox and nongenotox) alerts by ISS', 'DART scheme v. 1.0', 'DNA alerts for AMES by OASIS v. 1.4', and 'DNA alerts for CA and MNT by OASIS v. 1.1'. The 'Apply' button is also circled in red, with a callout box labeled '2'. The main window shows a 'Structure' tab with a chemical structure and a list of endpoints to be profiled, including 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', 'Ecotoxicological Information', and 'Human Health Hazards'.

1. Check the profilers related to the target endpoint (see slide 30); **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 box 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 shows the QSAR Toolbox interface with the 'Profiling' menu open. The 'Profile' node is selected in the tree view, and its contents are displayed on the right. A callout box with the number '1' points to the '+' icon next to the 'Profile' node in the tree.

1. Double click on the box  to open the nodes of the tree.

Tree View:

- Substance Profile
 - General Mechanistic
 - DNA binding by OASIS v.1.4
 - DNA binding by OECD
 - Protein binding by OASIS v1.4
 - Protein binding by OECD
 - Endpoint Specific
 - Carcinogenicity (genotox and nongenotox) alerts b...
 - DNA alerts for AMES by OASIS v.1.4
 - in vitro mutagenicity (Ames test) alerts by ISS
 - in vivo mutagenicity (Micronucleus) alerts by ISS
 - Empiric
 - Organic Functional groups

Profile Details:

- No alert found
- Schiff base formers
- Schiff base formers >> Direct Acting Schiff Base Formers
- Schiff base formers >> Direct Acting Schiff Base Formers >> Mono aldehydes
- Schiff base formation
- Schiff base formation >> Schiff base formation with carbonyl compounds
- Schiff base formation >> Schiff base formation with carbonyl compounds >> Aldehydes
- Schiff Base Formers
- Schiff Base Formers >> Direct Acting Schiff Base Formers
- Schiff Base Formers >> Direct Acting Schiff Base Formers >> Mono-carbonyls
- Simple aldehyde (Genotox)
- Structural alert for genotoxic carcinogenicity
- No alert found
- Simple aldehyde
- Simple aldehyde
- Aldehyde

Profiling

Profiles of n-hexanal

The screenshot displays the QSAR Toolbox interface. On the left, the 'Profiling methods' list includes 'DNA binding by OECD' which is checked. The central 'Filter endpoint tree...' shows a tree structure where 'DNA binding by OECD' is expanded. A right-click context menu is open over this node, with the 'Explain' option highlighted. A chemical structure of n-hexanal is shown in the top right of the main window.

- 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

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).
3. The window with chemical profiles appears, click "Details" to see detailed explanation

Profiling

DNA binding by OECD of n-hexanal

1. Structural boundaries of the category; **2.** Definition of the used common fragments; **3.** Mechanistic justification of the category

Outlook

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

Endpoint Overview

- As you should remember, “Endpoints” refer 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).

Endpoint

Case study

- In this example, we limit our data gathering to the common genotoxicity endpoints from databases containing genotoxicity data (Carcinogenicity & Mutagenicity ISSCAN, Genotoxicity OASIS, Micronucleus ISSMIC, Micronucleus OASIS and Toxicity Japan MHLW).

Endpoint Gather data

The screenshot displays the QSAR Toolbox software interface. The top menu bar shows the 'Endpoint' workflow selected. The toolbar below it contains various icons, with the 'Gather' icon circled in red. The main workspace is divided into several panels. On the left, the 'Databases' panel is expanded to show the 'Human Health Hazards' section, which is also circled in red. Within this section, several databases are checked, including 'Cardiogenicity & mutagenicity ISSCA', 'Genotoxicity OASIS', and 'Toxicity Japan MHLW'. A red oval highlights these selected databases. In the center, a chemical structure is displayed, and below it, a list of properties is shown, including 'Substance Identity', 'Physical Chemical Properties', and 'Human Health Hazards'. A red callout box with the number '1' points to the 'Endpoint' menu item. Another red callout box with the number '2' points to the 'Human Health Hazards' section. A third red callout box with the number '3' points to the list of selected databases. A fourth red callout box with the number '4' points to the 'Gather' button in the toolbar. At the bottom of the interface, a blue box contains the following instructions: '1. Click on **Endpoint**; 2. **Expand** the Human Health Hazard section; 3. **Select** databases related to the target endpoint; 4. Click **Gather**'.

Endpoint 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.

The screenshot displays the QSAR Toolbox software interface. The main window is titled '1 [target]' and shows a chemical structure of 1-hexanol. A 'Read data?' dialog box is open, with the 'All endpoints' radio button selected and the 'OK' button highlighted with a red '1' in a blue box. The dialog box also has a 'Choose...' radio button, a 'from Tautomers' checkbox, and 'OK' and 'Cancel' buttons. The main window shows a list of databases on the left and a 'Filter endpoint tree...' panel on the right with a chemical structure of 1-hexanol.

1. Click OK to read all available data

Endpoint

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 displays the QSAR Toolbox 3.4.0.5 interface. The 'Endpoint' menu is active. A dialog box is open in the center, displaying the message: "2 data points gathered across 1 chemicals." with an "OK" button. A blue callout box with the number "1" points to the "OK" button. At the bottom right, a larger blue callout box contains the instruction: "1. Click OK to close the window". The background shows the 'Databases' and 'Inventories' panels on the left and a 'Filter endpoint tree...' search area in the center.

Endpoint 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 displays the QSAR Toolbox interface. On the left, the 'Databases' panel is expanded to show 'Human Health Hazards' and 'Genetic Toxicity'. The 'Filter endpoint tree...' panel in the center shows a tree structure with 'Salmonella typhimurium' selected. The 'Structure' panel on the right shows the chemical structure of Salmonella typhimurium. The 'Data' table on the right shows two rows of data for 'Salmonella typhimurium' with 'M: Negative' results. A callout box with the number '1' points to these two rows. A larger callout box at the bottom right contains the text: '1. There are two negative experimental data for the target chemical'.

Endpoints

Recap

- You have entered the target chemical by SMILES and found it to be n-hexanal with the CAS# [66-25-1].
- You have profiled the target chemical and found 2 experimental data is available for n-hexanal.
- In other words, we will try to reproduce the experimental data by using read-across approach.
- **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
 - Endpoint
 - **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.
- Lets take a moment to review our mechanistic profile of the target chemical (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

Which of the category to be defined?

1. Click on **Category Definition**

In this case n-hexanal has structural evidence that it is has positive DNA binding alert based on general mechanistic DNA profilers. However there is no evidence that the target will elicit positive DNA effect based on endpoint specific DNA profiler. Based on this it is appropriate to identify analogues based on structural similarity with respect to OFG profiler.

Category Definition

Defining Organic functional group

The screenshot illustrates the 'Category Definition' process in the QSAR Toolbox. The main window shows the 'Categorize' tab with the 'Define' button circled in red (callout 2). The 'Grouping methods' list on the left has 'Organic Functional groups' circled in red (callout 1). A 'Filter endpoint tree...' window is open, showing a tree structure where 'Organic Functional groups' is selected (callout 3). A 'Target(s) profiles' dialog box is open, showing 'Aldehyde' selected (callout 3). The 'OK' button in the dialog box is circled in blue (callout 4).

1. **Highlight** "OFG";
2. **Click** Define;
3. The target category is Aldehydes
Confirm the category;
4. **Click** OK

Category Definition

Defining Organic functional group category

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes buttons for Input, Profiling, Endpoint, **Category Definition**, Data Gap Filling, and Report. Below this, a secondary bar shows 'Categorize' and 'Delete' options with sub-buttons like Define, Define with metabolism, Subcategorize, Combine, Clustering, Delete, and Delete All. The main workspace is divided into several panels:

- Left Panel:** 'Grouping methods' and 'Empiric' sections with a list of various chemical and toxicological methods.
- Top Center:** 'Filter endpoint tree...' with a search box containing '1 [target]'.
- Center:** A tree view showing a hierarchy of categories. The 'Aldehyde' category is selected, and a dialog box is open over it.
- Dialog Box:** Titled 'Define category name', it contains the text 'Aldehyde (Organic Functional groups)' in the input field. The 'OK' button is highlighted with a blue callout box containing the number '1'.
- Right Panel:** A list of chemical names, including 'Schiff base formers' and 'Simple aldehyde (Genotox)'. A chemical structure of an aldehyde is visible in the top right corner.

1. Click OK to confirm the name of the category

Category Definition Analogues

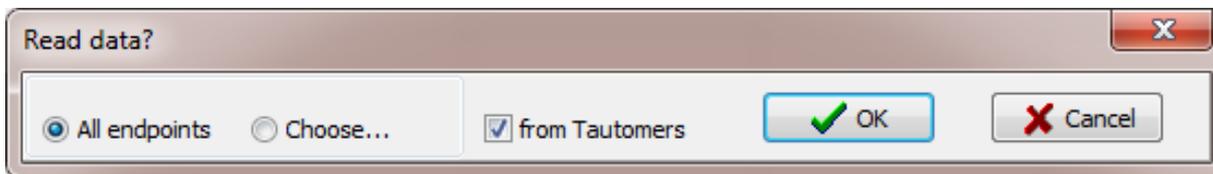
- The Toolbox now identifies all chemicals corresponding to category “Aldehydes” by Organic functional groups listed in the databases selected under “Endpoints”.
- The name of the category appears in the “Defined Categories” window, the number in brackets is the number of substances belonging to the category (107 analogues including the target chemical are identified)



Category Definition

Read data for Analogues

- The Toolbox automatically request 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, as only databases are selected that contain information for genetic toxicity endpoint, so both options give the same results.**

Category Definition

Read data for Analogues

Due to the overlap between the Toolbox databases same data for intersecting chemicals could be found simultaneously in more than one databases. The data redundancy is identified and the user has the opportunity to select either a single data value or all data values.

Repeated values for: 74 data-points, 37 groups, 36 chemicals

Data points...

	Endpoint	CAS	Structure	Value	Author
<input checked="" type="checkbox"/>	Gene mutation	123-11-5		Negative	National Cancer Institute
<input type="checkbox"/>	Gene mutation	123-11-5		Negative	National Cancer Institute
<input checked="" type="checkbox"/>	Gene mutation	104-88-1		Negative	National Cancer Institute
<input type="checkbox"/>	Gene mutation	104-88-1		Negative	National Cancer Institute
<input checked="" type="checkbox"/>	Gene mutation	110-62-3		Negative	National Cancer Institute
<input type="checkbox"/>	Gene mutation	110-62-3		Negative	National Cancer Institute

1. Click Select one and then 2. Click OK

Category Definition

Summary information for Analogues

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

The screenshot displays the QSAR Toolbox 3.4.0.5 interface. The 'Category Definition' tab is active. The left sidebar lists various grouping methods, including 'Empiric' and 'Toxicological'. The central panel shows a tree view of endpoints, with 'In Vitro' and 'Bacterial Reverse Mutation Assay (e.g. Ames Test)' expanded. The right panel is a data matrix with columns labeled 1 through 5. The matrix contains numerical counts and categorical results (e.g., 'M: Negative', '>2E3 mg/kg'). A red box highlights a row in the matrix, indicating experimental results for analogues.

Endpoint	1 [target]	2	3	4	5
Acute Toxicity	(1/1)				M: >2E3 mg/kg
Bacterial Reverse Mutation Assay (e.g. Ames Test)	(2/4)				M: Negative, Negative
Escherichia coli	(81/81)	M: Negative	M: Negative	M: Negative	M: Negative
Salmonella typhimurium	(45/174)	M: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...
With S9	(49/187)	M: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...
Without S9					
In Vitro Mammalian Chromosome Aberrati...	(10/18)				M: Negative, Negative

Category Definition

Side-Bar of experimental data

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' side-bar is active, showing a list of grouping methods on the left and a table of data points in the center. The table has columns for '#', 'Endpoint', 'Value', 'Original value', 'Strain', 'Source of metabolic system', 'Phylum', 'Test organisms (species)', 'Type of method', and 'Type of genotoxicity'. A callout box labeled '1' points to a cell in the 'Value' column containing the text 'M: Negative, Negat...'. Below the table, there are sections for 'Toxicological' and 'Defined Categories'.

#	Endpoint	Value	Original value	Strain	Source of metabolic system	Phylum	Test organisms (species)	Type of method	Type of genotoxicity
1	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 98	rat	Proteobacteria	Salmonella typhimurium	In Vitro	Gene mutation
2	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 100	rat	Proteobacteria	Salmonella	In Vitro	Gene
3	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 1535	rat	Proteobacteria	Salmonella typhimurium	In Vitro	Gene mutation
4	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 1537	rat	Proteobacteria	Salmonella	In Vitro	Gene
5	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 1538	rat	Proteobacteria	Salmonella	In Vitro	Gene
6	Gene mutation	Negative (Gene mutation I)	Negative (Gene mutation I)	TA 97	rat	Proteobacteria	Salmonella	In Vitro	Gene

1. 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 105 analogous (“Aldehydes” by OFG classification) with the target chemical (n-hexanal).
- The available experimental data for these 105 similar chemicals are collected from the previously selected databases under Endpoint 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 double clicking on the nodes of endpoint tree open the tree to the target: **Bacterial reverse mutation (Ames) assay without S9** (*i.e., double click on Human Health Hazards then double 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 software interface. The top navigation bar includes tabs for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The main workspace is divided into several sections:

- Left Sidebar:** 'Grouping methods' section with 'Predefined' and 'General Mechanistic' categories.
- Filter endpoint tree...:** A tree view showing various endpoints. Red callout boxes with numbers 1, 2, 3, and 4 point to specific nodes: 1 points to 'Genetic Toxicity', 2 points to 'In Vitro', 3 points to 'Bacterial Reverse Mutation Assay (e.g. Ames ...)', and 4 points to 'Gene Mutation'.
- Structure:** A section showing chemical structures for the selected endpoints.
- Data Table:** A table with five columns labeled 1 to 5, showing results for various assays. The table includes rows for 'Escherichia coli' and 'Salmonella typhimurium'.

1. Click to Genetic Toxicity after that **2. Click** to In vitro **3. Click** to Bacterial Reverse Mutation Assay (e.g. Ames Test) and finally **4. Click** Gene Mutation

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' menu is active, showing options like 'Define', 'Define with metabolism', 'Subcategorize', 'Combine', 'Clustering', 'Delete', and 'Delete All'. The 'Endpoint tree' is displayed on the left, with a red box and the number '1' highlighting the 'Salmonella typhimurium' node. The right side of the interface shows a table of results for various endpoints, with the 'Salmonella typhimurium' row highlighted in blue.

Structure	1 [target]	2	3	4	5
Ecotoxicological Information					
Human Health Hazards					
Acute Toxicity (1/1)				M: >2E3 mg/kg	
Bioaccumulation					
Carcinogenicity (10/28)					
Developmental Toxicity / Teratogenicity					
Genetic Toxicity					
In Vitro					
Bacterial Reverse Mutation Assay (e.g. Ames ...)					
Gene Mutation					
Escherichia coli (2/4)				M: Negative, Negative	
Salmonella typhimurium					
No S9 Info (81/81)				M: Negative	M: Negative
With S9 (45/174)	M: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Ne
Without S9 (49/187)	M: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Ne
Undefined Metabolic Activation					
DNA Damage and Repair Assay, Unscheduled ...					
DNA React. (Ashby Fragments)					
In Vitro Mammalian Chromosome Aberr... (10/18)				M: Negative, Negative	
Mammalian Cell Gene Mutation Assay (4/4)					
Sister Chromatid Exchange Assay					
					M: Inconclusive

1. Open the tree to *Salmonella typhimurium*

Category Definition

Navigation through the endpoint tree

The screenshot displays the QSAR Toolbox interface with the 'Category Definition' workflow selected. The 'Endpoint' tree is expanded to show the following path: In Vitro > Bacterial Reverse Mutation Assay (e.g. Ames ...) > Gene Mutation > Escherichia coli > Without S9. The table below shows the results for five target chemicals, with the cell for 'Without S9' in the first target column highlighted in red.

Structure	1 [target]	2	3	4	5
Substance Identity					
Physical Chemical Properties					
Environmental Fate and Transport					
Ecotoxicological Information					
Human Health Hazards					
Acute Toxicity (1/1)					M: >2E3 mg/kg
Bioaccumulation					
Carcinogenicity (10/28)					
Developmental Toxicity / Teratogenicity					
Genetic Toxicity					
In Vitro					
Bacterial Reverse Mutation Assay (e.g. Ames ...)					
Gene Mutation					
Escherichia coli (2/4)					M: Negative, Negative
Salmonella typhimurium					
No S9 Info (81/81)					M: Negative
With S9 (45/174)	M: Negative				M: Negative, Negat...
Without S9 (49/97)	M: Negative	M: Negative, Negat...			M: Negative, Negat...
Undefined Metabolic Activation					
DNA Damage and Repair Assay, Unscheduled ...					
DNA React. (Ashby Fragments)					
In Vitro Mammalian Chromosome Aberr... (10/18)					M: Negative, Negative
Mammalian Cell Gene Mutation Assay (4/4)					
Sister Chromatid Exchange Assay					
In Vivo (8/14)					M: Inconclusiv

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

Category Definition

Recap

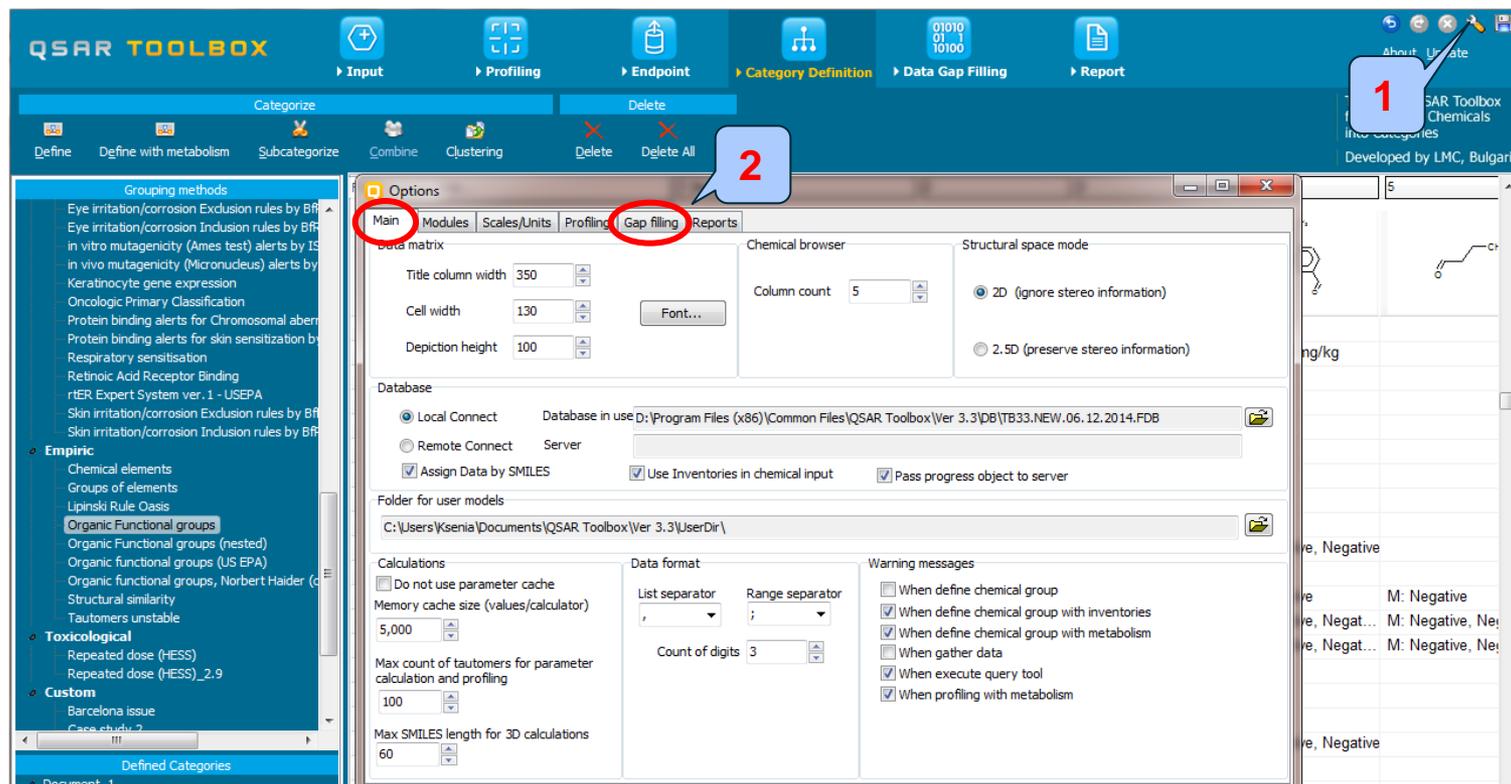
- You have now retrieved the available experimental data on genetic toxicity for 107 chemicals classified as “Aldehydes” by OFG, found in the databases containing mutagenicity data.
- Only 49 out of 107 analogues have experimental mutagenicity data related to the target.
- You are now ready to fill in 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
 - Endpoint
 - Category definition
 - **Data Gap Filling**
 - **Ames without S9**

Data Gap Filling (Ames without S9)

Apply read-across



In order to save model and export data for the analogues in Read-across analysis the user should set the specific options: 1. **Go** to Option; 2. **Open** Gap Filling panel; 3. **Open** Prediction and; 4. **Select** two radio buttons 3 and 5; 5. **Click** OK. (see next two slides)

Data Gap Filling (Ames without S9)

Apply read-across

The screenshot shows the QSAR Toolbox software interface. The main window is titled 'Options' and has several tabs: 'Main', 'Modules', 'Scales/Units', 'Profiling', 'Gap filling', and 'Reports'. The 'Prediction' tab is selected and highlighted with a red circle. A blue callout box with the number '3' points to this tab. The 'Prediction' tab contains the following sections:

- Data consistency criteria:** A list of checkboxes for various criteria, including Age, Assay, Duration, Effect, Endpoint, Metabolic activation, Organ, Route, Route of administration, Strain, Test condition, Test organisms (species), Test type, Type of method, Test guideline, Organ (Tissue), Abnormality, Activity Score, Abnormality, additional_comments, Adduct M+H, Administration period (day), Affected morphology, Affected sex, Agency Responsible Sediment for Data, Analysis direction, Analytical verification of test atmosphere concentrations, Any other information on results incl. tables, and APPLICANT'S SUMMARY AND CONCLUSION (executive summary).
- Default descriptor for X axis:** A dropdown menu set to 'Molecular weight'.
- When starting Gap filling, collect data for the following descriptors:** A list of checkboxes for various descriptors, including (Q) Acidic pKa (Chemaxon), (Q) Basic pKa (Chemaxon), BAF, BAF (lower trophic), BAF (mid trophic), BAF (upper trophic), BAF (upper trophic, biotransformation rate is zero), BCF, BCF (lower trophic), BCF (mid trophic), BCF (upper trophic), BCF (upper trophic, biotransformation rate is zero), Bio Half-Life, Biodeg probability (Biowin 1), Biodeg probability (Biowin 2), Biodeg probability (Biowin 5), Biodeg probability (Biowin 6), Biodeg probability (Biowin 7), BioHC half-life, Biotransformation Half-Life, Boiling Point, Calculated heat of formation, Diameter effective, Diameter maximum, and Diameter minimum.

At the bottom of the 'Options' dialog, there is a button labeled '12 (Q)SAR display/ranking'.

In order to save model and export data for the analogues in Read-across analysis the user should set the specific options: 1. **Go** to Option; 2. **Open** Gap Filling panel; 3. **Open** Prediction and; 4. **Select** two radio buttons 3 and 5; 5. **Click** OK. (see next two slides)

Data Gap Filling (Ames without S9)

Apply read-across

The screenshot shows the 'Options' dialog box in the QSAR Toolbox software. The 'Prediction' tab is selected. The following options are visible:

- Incomplete endpoint data warnings:**
 - using IMO for mixture components
 - using SMO for mixture components
 - using IMO for set of tautomers
 - using SMO for set of tautomers
 - using IMO for set of metabolites/transf. products
 - using SMO for set of metabolites/transf. products
- Accept prediction**
 - When the target chemical is out of the parametric range(s) of analogues:
 - Accept the prediction
 - Do NOT accept the prediction
 - Ask to accept the prediction
 - When making a prediction by external QSAR that has no domain:
 - Accept the prediction
 - Do NOT accept the prediction
 - Ask to accept the prediction
 - When a new regression / categorical model is created, but still not saved:
 - Save the model
 - Do NOT save the model
 - Ask to save the model
- Specify if relevant profiles to the current prediction will be selected (to appear in report):**
 - Select profiles
 - Use default selection
 - Ask to select profiles
- Specify if data for target substance and analogues will be collected from data matrix (to appear in report):**
 - Collect data
 - Do NOT collect data
 - Ask to collect data

At the bottom of the dialog, there are three buttons: 'OK' (with a green checkmark), 'Cancel' (with a red X), and 'Restore default' (with a green refresh icon). A callout box with the number '5' points to the 'OK' button.

In order to save model and export data for the analogues in Read-across analysis the user should set the specific options: 1. **Go** to Option; 2. **Open** Gap Filling panel; 3. **Open** Prediction and; 4. **Select** two radio buttons 3 and 5; 5. **Click** OK

Data Gap Filling (Ames without S9)

Apply read-across

1. Click on Data Gap Filling; **2. Highlight** the data endpoint box corresponding to Ames without S9 under the target chemical (note it is empty); **3. Select** Read across and; **4. Click** Apply; **5.** An insert window alerting you to possible data inconsistencies appears. **Click** OK.

Data Gap Filling (Ames without S9)

Results of Read across

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

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

Data Gap Filling Method

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

Target Endpoint

Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test)
Gene Mutation Salmonella typhimurium Without S9

Structure

1 [target]	2	3	4	5
Without S9 (49/187) M. Negative	M. Negative, Negat.	M. Negative, Negat.	M. Negative, Negat.	M. Negative, Negat.

Descriptors Prediction

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 22 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Negative'

Descriptor X: log Kow

Accept prediction
Return to matrix

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

105 Aldehyde (Organic Functional groups) Create prediction by gap filling 0/1 1/1/0

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 **GREEN** dots (which you will see shortly) represent analogues belonging to different subcategories.

Data Gap Filling (Ames without S9)

Interpretation of the Read across

- Six 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. The interface is divided into several sections:

- Left Sidebar:** A tree view showing various endpoints. Under 'Endpoint Specific', 'DNA alerts for AMES by OASIS v.1.4' is selected and circled in red (callout 3).
- Top Panel:** Navigation buttons for Profiling, Endpoint, Category Definition, Data Gap Filling, and Report. The 'Data Gap Filling' button is highlighted.
- Central Workspace:** Shows a list of chemicals with their predicted values. A scatter plot titled 'Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 22 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Negative'' plots 'log Kow' on the x-axis. Chemical structures are shown above the list.
- Right Sidebar:** A panel with various actions. 'Select/filter data' and 'Subcategorize' are circled in red (callout 1 and 2).
- Bottom Panel:** A list of selected chemicals. The 'Remove' button is circled in red (callout 5).

1. Click Select/filter data; 2. Select Subcategorize;
3. Select DNA alerts for AMES by OASIS v.1.4;
4. Examine dissimilar chemicals and
5. Click Remove to eliminate dissimilar chemical.

Data Gap Filling (Ames without S9)

Subcategorization by OFG (US-EPA)

The screenshot shows the QSAR Toolbox interface for Data Gap Filling. The 'Data Gap Filling Method' is set to 'Read-across'. The target endpoint is 'Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test) Gene Mutation Salmonella typhimurium Without S9'. The main table displays chemical structures and their predicted values for 'Gene mutation' without S9. A scatter plot below the table shows 'Gene mutation (obs.)' vs 'log Kow', with two positive outliers highlighted in red boxes. A callout box with the number '1' points to the 'Subcategorize' option in the 'Accept prediction' panel on the right. The bottom status bar shows '105 Aldehyde (Organic Functional groups)'.

There are two 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 to eliminate dissimilar chemical. (see next slides)

Data Gap Filling (Ames without S9)

Subcategorization by OFG (US-EPA)

Subcategorization

Grouping methods

- Carcinogenicity (gen DART scheme v.1.0)
- DNA alerts for AMES
- Eye irritation/corrosi
- Eye irritation/corrosi
- in vitro mutagenicity
- in vivo mutagenicity
- Keratinocyte gene e
- Oncologic Primary Cl
- Protein binding alert
- Protein binding alert
- Respiratory sensitiz
- Retinoic Acid Recept
- rtER Expert System
- Skin irritation/corros
- Skin irritation/corros

Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic Functional g
- Organic Functional g
- Organic functional g
- Organic functional g
- Structural similarity
- Tautomers unstable

Toxicological

- Repeated dose (HE
- Repeated dose (fur

Metabolism/Transformations

Do not account metabol

Documented

- Observed Mammalian me
- Observed Microbial meta
- Observed Rat In vivo me
- Observed Rat Liver S9

Simulated

- Autoxidation simulator
- Autoxidation simulator (
- Dissociation simulation
- Hydrolysis simulator (acid
- Hydrolysis simulator (bas
- Hydrolysis simulator (neu

Adjust options

Target

- Aldehyde, aliphatic attach
- Aliphatic Carbon [CH]
- Aliphatic Carbon [-CH2-]
- Aliphatic Carbon [-CH3]
- Miscellaneous sulfide (=S)

Differ from target by:

- At least one categor
- All categories

Correlation

Analogue

- (3) Alcohol, olefinic attach [-O]
- (23) Aldehyde, aliphatic attach
- (19) Aldehyde, aromatic attach
- (1) Aliphatic Carbon [C]
- (33) Aliphatic Carbon [CH]
- (33) Aliphatic Carbon [-CH2-]
- (3) Aliphatic Carbon [-CH3]
- (2) Aliphatic Carbon [C]
- (2) Aliphatic Carbon [C]
- (1) Aliphatic Carbon [C]
- (1) Aliphatic Carbon [C]
- (5) Chlorine, aromatic attach [-
- (5) Chlorine, olefinic attach [-C
- (2) Ester, aliphatic attach [-C-
- (1) Ester, aromatic attach [-C-
- (1) Hydroxy, aliphatic attach [-
- (3) Hydroxy, aromatic attach [-
- (43) Miscellaneous sulfide (=S)
- (2) Nitro, aromatic attach [-NO
- (43) Olefinic carbon [=CH- or =
- (1) Olefinic carbon [=CH2]
- (1) Ortho-hydroxy to misc. -CO
- (1) Oxycarbonyl compound [CO
- (9) Oxygen, one aromatic atta
- (1) Oxygen, two olefinic attach
- (12) Tertiary Carbon

Selected 34 (9/43)

Select different

Remove

Structure

Without S9 (44/170)

Prediction

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 22 values from 5 neighbour chemicals. Observed target value: 'Negative', Predicted target value: 'Negative'

Accept prediction

Return to matrix

- Select/filter data
- Subcategorize
- Mark chemicals by descriptor value
- Filter points by test conditions
- Mark focused chemical
- Mark focused points
- Remove marked chemicals/points
- Clear existing marks
- Selection navigation
- Gap filling approach
- Descriptors/data
- Model/(Q)SAR
- Calculation options
- Visual options
- Information
- Miscellaneous

105 Aldehyde (Organic Functional groups)

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

- Select** Subcategorize;
- Select** OFG (US-EPA);
- Click** Remove to eliminate dissimilar chemical. (see next slides)

Data Gap Filling (Ames without S9)

Interpretation of the Read across

The screenshot shows the QSAR Toolbox interface during a Data Gap Filling operation. The main window displays a grid of chemical structures (Aldehydes) and their Ames test results. A 'Read across prediction' text box explains the prediction method: 'Read across prediction of Gene mutation, based from the nearest 5 neighbours, based on 21 values from 5 neighbour chemicals, predicted target value: 'Negative', Predicted target value: 'Negative''. A 'Confirm' dialog box is overlaid on the grid, asking if the current model should be saved. A 'Select/filter data' panel is visible on the right. Three numbered callouts (1, 2, 3) highlight key actions: 1. Clicking 'Accept prediction', 2. Clicking 'Yes' in the 'Confirm' dialog, and 3. Clicking 'Edit model' in the 'Edit model' dialog.

Now all analogues are structurally similar (Aldehydes) and negative by the experimental data. The prediction could be accepted by **1. Click** on Accept prediction and If you want to save the model, and use it for further predictions, then **2. Click** Yes and then **3. Edit** the information about the model.

Data Gap Filling (Ames without S9)

Interpretation of the Read across

The screenshot shows the QSAR Toolbox interface during a 'Data Gap Filling' operation. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The main window displays a table of chemical structures and their predicted values for 'Gene mutation'. A 'Confirm' dialog box is overlaid on a scatter plot, asking if additional data should be collected for reporting purposes. A red '1' in a blue callout points to the 'Yes' button in the dialog.

Structure	28	29	32	41
Structure	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>	<chem>CCCCCCCC</chem>
Without S9 (10/33)	M. Negative	M. Negative, Negative	M. Negative, Negat	M. Negative, Negative

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 21 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Negative'

Gene mutation (obs.)

log Kow

Descriptor X: log Kow

Confirm

Do you want to collect additional data for chemicals from data matrix for reporting purposes? (this data will be provided in data matrix tables)

Yes No

1

1. Click OK in order to have in report additional data for the analogues

Data Gap Filling (Ames without S9)

Interpretation of the Read across

The screenshot displays the QSAR Toolbox software interface during a Data Gap Filling operation. The top navigation bar includes buttons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling (active), and Report. The left sidebar shows the 'Data Gap Filling Method' set to 'Read-across' and the 'Target Endpoint' as 'Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g., Ames Test) Gene Mutation Salmonella typhimurium Without S9'. The central workspace shows a chemical structure and a data matrix with columns 28, 29, 32, and 41. A dialog box titled 'Select nodes to be reported in data matrix tables' is open, showing a tree view of hazard categories. Red callouts indicate the following steps: 1. Expanding 'Human Health Hazards (10/93)'; 2. Selecting 'Without S9 (10/33)' under 'Salmonella typhimurium'; 3. Clicking the 'OK' button; 4. Returning to the matrix. The right sidebar contains options for 'Accept prediction', 'Return to matrix', and 'Select/filter data'.

1. Expand the Human Health hazard section and **select** without S9; **3. Click** OK; **4. Return** to matrix

Data Gap Filling (Ames without S9) Results

- By accepting the prediction the data gap is filled.
- By clicking on "Return to Matrix", the user can close the read-across for the current endpoint and 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 shows the QSAR Toolbox interface with the 'Data Gap Filling' method selected. The 'Target Endpoint' is 'Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test) Gene Mutation Salmonella typhimurium With S9'. The results table shows the following data:

Endpoint	1 [target]	2	3	4	5
Escherichia coli (2/4)				M: Negative, Negative	
Salmonella typhimurium					
No S9 Info (81/81)		M: Negative	M: Negative	M: Negative	M: Negative
With S9 (45/174)	M: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...
Without S9 (49/188)	M: Negative R: Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...
Undefined Metabolic Activation					
DNA Damage and Repair Assay, Unsch...					
DNA React. (Ashby Fragments)					
In Vitro Mammalian Chromosome...				M: Negative, Negative	
Mammalian Cell Gene Mutation Assay(4/4)					
Sister Chromatid Exchange Assay					
In Vivo (6/12)					M: Inconclusive
Immunotoxicity					
Irritation / Corrosion					
Neurotoxicity					
Photoinduced Toxicity					
Repeated Dose Toxicity					
Sensitisation					
ToxCast					
Toxicity to Reproduction					
Toxicokinetics, Metabolism and Distribution					
Profile					

Callout 1 points to the 'M: Negative' result for the first endpoint (target). Callout 2 points to the 'M: Negative, Negat...' result for the second endpoint, indicating a data gap.

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
 - Endpoint
 - Category definition
 - **Data Gap Filling**
 - Ames without S9
 - **Ames with S9**

Data Gap Filling (Ames with S9)

- We do this the same way as with Ames without S9.
- Make sure **Data Gap Filling** is highlighted.
- Highlight the **data endpoint box**; this time corresponding to **Ames with S9**. Again the box under the structure is empty.
- Select **Read across** and Click **Apply**.
- As before an insert window alerting you to **possible data inconsistencies** appears. Click **OK** (see next screen shot).

Data Gap Filling (Ames with S9) Apply read-across

1 Read-across

2 Apply

3 OK

1. If you have trouble review slide number 68.

105 Aldehyde (Organic Func... 1/1/0

Data Gap Filling (Ames with S9)

Results of Read across

The screenshot displays the QSAR Toolbox interface during the Data Gap Filling process. The top navigation bar includes: **Input**, **Profiling**, **Endpoint**, **Category Definition**, **Data Gap Filling** (active), and **Report**. The left sidebar shows the **Data Gap Filling Method** set to **Read-across** and the **Target Endpoint** as **Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test) Gene Mutation Salmonella typhimurium With S9**.

The main workspace shows a **Structure** field with a chemical structure and a table of 5 neighboring chemicals. The table headers are 1 [target], 2, 3, 4, and 5. The table content shows chemical structures and their predicted target values: M: Negative, M: Negative, Negat..., M: Negative, Negat..., M: Negative, Negat..., and M: Negative, Negat....

Below the table is a scatter plot titled **Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 23 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Negative'**. The y-axis is **Gene mutation (obs.)** with categories Positive, Equivocal, and Negative. The x-axis is **log Kow** with values from 0.00 to 4.00. The plot shows a cluster of points at the bottom (Negative) and a few points at the top (Positive).

The status bar at the bottom indicates: **105 Aldehyde (Organic Functional groups)**, **Create prediction by gap filling**, **0/1**, and **1/1/0**.

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 subcategorisation.
- Subcategorisation 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 subcategorisation 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 'Subcategorization' window. The top navigation bar includes 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Endpoint Specific' methods, with 'DNA alerts for AMES' selected. The central workspace shows chemical structures and a scatter plot of 'Gene mutation (obs)' vs 'log Kow'. The right sidebar has 'Accept prediction' and 'Subcategorize' options. Red callouts 1-5 highlight key steps in the workflow.

1. **Select** Select/Filter data
2. **Click** Subcategorize
3. **Select** DNA alerts for AMES by OASIS v.1.4 (endpoint specific)
4. **Select** Rat liver metabolism simulator.
5. **Click**

Remove

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

- As with Ames without S9, we want to refined the category by subcategorisation with OFG (US-EPA)
- Select **Select/filter data**
- Select **Subcategorize**
- Select **Organic functional groups (US-EPA)**
- Look for dissimilar chemicals
- Click **Remove** to eliminate dissimilar chemical.

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

The screenshot displays the 'Subcategorization' window of the QSAR Toolbox. The interface is divided into several sections:

- Left Sidebar:** Lists various grouping methods under categories like 'Toxicological', 'Experimental', 'Custom', and 'Metabolism/Transformations'. A callout box labeled '1' points to the 'Do not account metabolism' option.
- Target List:** A list of chemical features to be used for subcategorization. A callout box labeled '2' points to 'Organic functional groups'.
- Central Workspace:** Shows chemical structures and their predicted categories. A callout box labeled '3' points to the 'Remove' button at the bottom of the 'Selected' list.
- Right Sidebar:** Contains control options for the prediction process, including 'Accept prediction', 'Return to matrix', and 'Selection navigation'.

The central plot shows a scatter plot of 'Gene mutation (obs.)' versus 'log Kow'. The plot title reads: "Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 20 values from 5 neighbour chemicals, Observed target value: 'Negative', Predicted target value: 'Negative'".

1. Click on Do not account metabolism; 2. Select OFG (US-EPA); 3. Click Remove

Data Gap Filling (Ames with S9)

Result of read-across

Now all 5 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

1. **Click** on Model/(Q)SAR and then;
2. **Click** on Save domain as category
3. Since a custom profiler has previously been defined, **highlight** custom profiler and
4. **Click** OK.

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

The screenshot shows the QSAR Toolbox interface with the 'Data Gap Filling' method selected. The 'Read-across' method is chosen, and the target endpoint is 'Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test) Gene Mutation Salmonella typhimurium With S9'. The main window displays a list of chemicals with their predicted categories. A dialog box is open for defining a category name, and a '3' callout points to the 'Accept prediction' button in the right-hand panel.

1. Type a name for the category in the "Name" box; **2. Click** OK; **3. Click** Accept prediction and Return to matrix.

Data Gap Filling (Ames with S9)

Result of read-across

The screenshot shows the QSAR Toolbox interface during a Data Gap Filling operation. The 'Data Gap Filling' tab is active, showing a list of chemicals with their predicted target values. A 'Confirm' dialog box is open, asking if the user wants to save the model. The plot shows 'Gene mutation (obs.)' on the y-axis and 'log Kow' on the x-axis. The predicted target value is 'Negative', matching the observed target value.

1. Click Accept prediction and then click **Yes** if you want to save the model, otherwise click **NO**; **2. Click** NO; **3. Return** to matrix.

Outlook

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

Report Overview

- Report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report templates as well as a template editor with which users can define their own user defined templates.
- The report can then be printed or saved in different formats. (see next screen shot).

Report Generate Report

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Report' menu is highlighted with a callout box labeled '1'. Below the menu bar, there are two panes: 'Available data to report' and 'Available report templates'. In the 'Available data to report' pane, a prediction is selected and circled in red, with a callout box labeled '2'. The 'Available report templates' pane shows 'Standard (predefined)' and 'Custom (user defined)' options. The 'Create' button in the 'Reports' toolbar is circled in red, with a callout box labeled '3'. The main window displays a report titled 'Prediction of Gene mutation for hexanal' with a page number '1 / 27'. The report content includes the title 'QSAR Toolbox prediction for single chemical' and a paragraph of text: 'The template of the current report is based on "GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS MODELS" published by OECD (September, 2007) and "GUIDANCE ON INFORMATION REQUIREMENTS AND CHEMICAL SAFETY ASSESSMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS" published by ECHA (May, 2008). The report provides information about the target substance, chemical characteristics used for the grouping, the resulting boundaries of the group of chemicals (applicability'.

1. **Select** "Report"; 2. **Select** the current prediction from "Available data to report" window, and then 3. **Click** Create .

Outlook

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

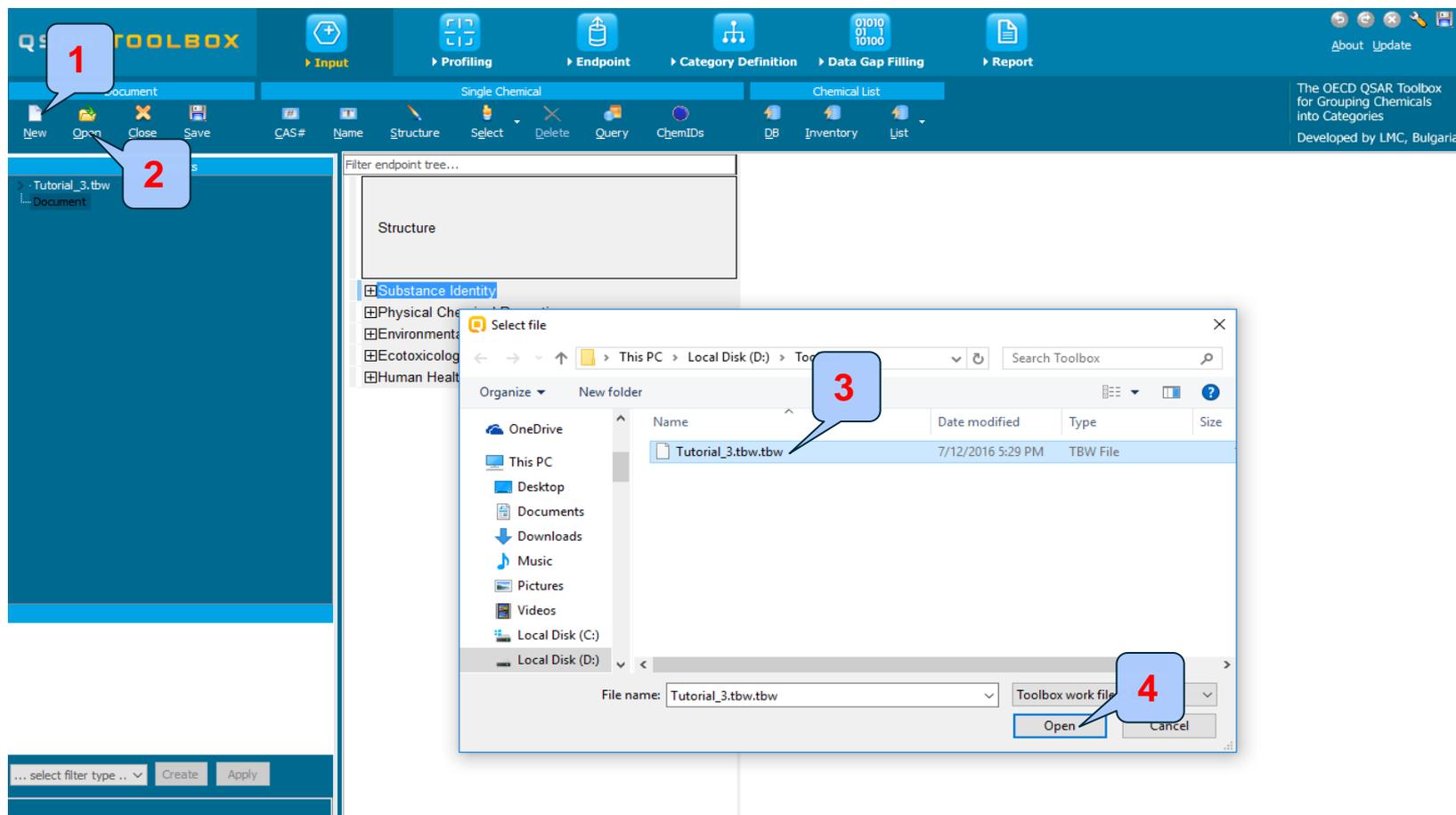
Saving the prediction result

- 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

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

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 software interface. The top menu bar includes options like 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. Below this is a toolbar with icons for 'New', 'Open', 'Close', 'Save', and various chemical manipulation tools. The main workspace is divided into several panes: a 'Documents' list on the left, a 'Filter endpoint tree...' pane, a 'Structure' pane, and a 'Chemical List' table. An 'Information' dialog box is open in the center, displaying a success message and details about interface changes. A blue callout bubble with the number '1' points to the 'OK' button in the dialog box.

Documents

- RA without metabolism
 - CAS: 97-53-0
 - Document
- Tutorial_3_Ames mutagenicity.tbw
 - SMILES: CCCCC=O

Chemical List

Filter endpoint tree...	1 [target]	2	3	4	5	6
Structure						
Substance Identity						
Physical Chemical Properties						
Environmental Fate and Transport						
Ecotoxicological Information						
Human Health Hazards (105/511)	M: Negative, Negative R: Negative, Negative	M: Negative, Negat...	M: Negative, Negat...	M: Negative, Negat...	M: Inconclusive, N...	M: Negative

Information

The file was executed successfully

The following interface changes were made during execution:

- Database "Bacterial mutagenicity ISSSTY" in branch "Human Health Hazards" was selected/unselected

OK

1. The file is opened successfully

1. 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 function in the Toolbox, especially using different profilers in subcategorizing the category of the target chemical.
- Remember proficiency in using the Toolbox will only come with practice.