

OECD QSAR Toolbox v.3.0

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

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

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

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

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

Exercise

- In this exercise we will predict the Ames mutagenicity potential for an untested compound, (n-hexanal) [SMILES CCCCC=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 structural similarity, in particular "DNA binding by OECD".
- 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., frameshifts).
- 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**

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 3.0.0.840 interface. The 'Structure' button in the 'Single Chemical' section is highlighted with a red box and a callout '1'. The '2D Editor' window is open, showing a grid of chemical structures for selection. A blue callout box at the bottom left contains the text '1. Click on Structure'.

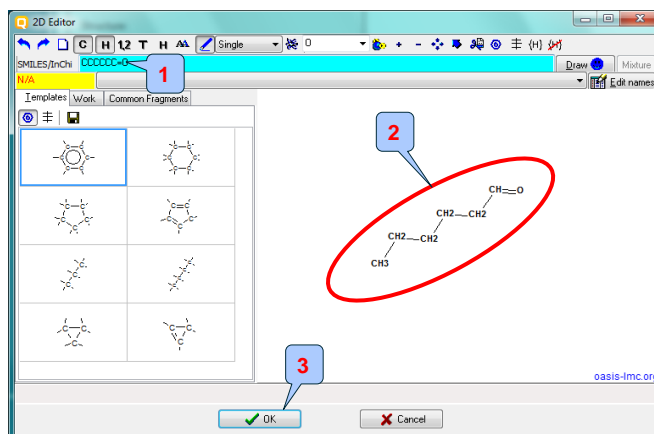
Chemical Input

Input target chemical by SMILES

- In the **Aqua-colored** 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

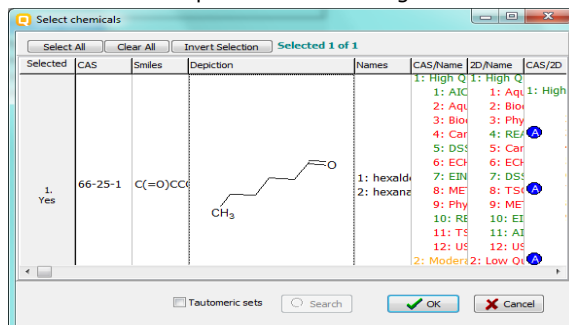
Input target chemical by SMILES



1. **Type** CCCCC=O in SMILES/InChi window; 2. 2D structure; 3. **Click** OK

Chemical Input Target chemical identity

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.



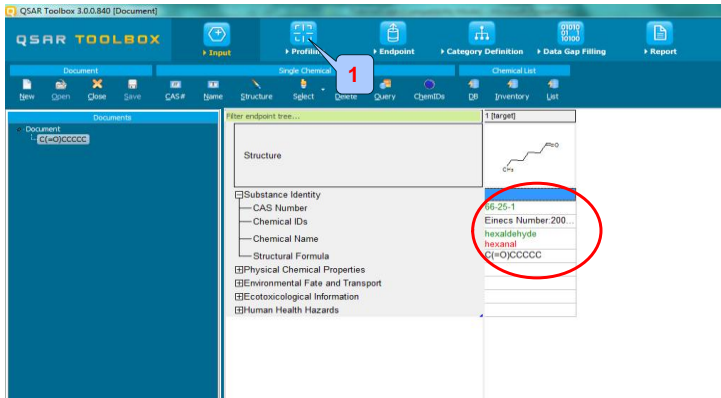
This panel displays QA information for presented chemicals. Click OK to add chemical in data matrix

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.

Q SAR TOOLBOX

Chemical Input Target chemical identity



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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 27

Q SAR TOOLBOX

Outlook

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 28

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:
 - Protein binding by OASIS – mechanistic grouping
 - Protein binding by OECD – mechanistic grouping
 - DNA binding by OASIS v1.1– mechanistic grouping
 - DNA binding by OECD – mechanistic grouping
 - DNA alerts for AMES, MN and CA by OASIS v.1.1
 - Carcinogenicity (genotox and nongenotox) alerts by ISS
 - in vitro mutagenicity (Ames test) alerts by ISS
 - in vivo mutagenicity (Micronucleus) alerts by ISS
 - Organic function groups

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 'Profiling Methods' dialog box is open, displaying a list of profiling methods with checkboxes. A red circle labeled '1' highlights the 'Apply' button at the bottom left. A red circle labeled '2' highlights the 'Apply' button at the top left of the dialog box. The main window shows a chemical structure and various property panels.

1. Check the profilers related to the target endpoint; 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

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories

Profiling Profiles of n-hexanal

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 35

- In this case there is structural evidence that it is a DNA and Protein binding compound.
- This allows to bind covalently to DNA .
- This mechanistic information is important for the grouping of analogues.

- Right click on the box with profiling result by DNA binding by OASIS.
- 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 OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 36

- In this case there is structural evidence that it is a DNA and Protein binding compound.
- This allows to bind covalently to DNA .
- This mechanistic information is important for the grouping of analogues.

- Right click on the box with profiling result by DNA binding by OASIS.
- Left Click on the "Explain" box to see why the target is profiled as "Mono-aldehydes" by DNA binding by OECD (see next slide).
- The window with chemical profiles appears, click "Details" to see detailed explanation

Q SAR TOOLBOX

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 37

Q SAR TOOLBOX

Outlook

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 38

Endpoints 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).

Endpoints Case study

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

Endpoints Gather data

1. Click on **Endpoint**
2. **Expand** the Human Health Hazard section
3. **Select** databases related to the target endpoint
4. Click **Gather**

The OECD QSAR Toolbox for Grouping Chemicals into Categories

Endpoints 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.

1. Click **OK** to read all available data

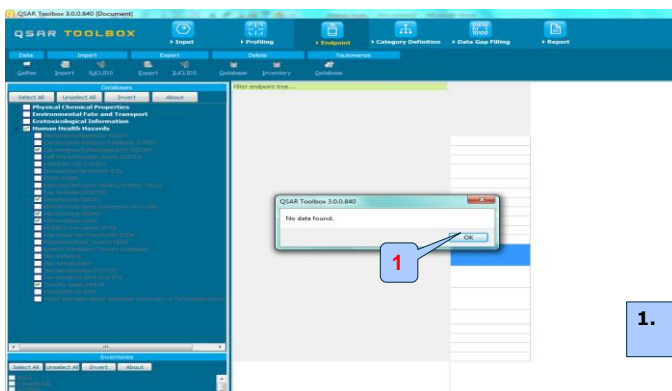
The OECD QSAR Toolbox for Grouping Chemicals into Categories

05.10.2012 42

Endpoints

Process of collecting data

In this example, an insert window appears stating there was "no data found" for the target chemical.



1. **Click** OK to close the window

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 no experimental data is currently available for n-hexanal .
- In other words, you have identified a data gap, which you would like to fill in.
- **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 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 a DNA binding compound. Therefore, grouping by DNA binding by OECD mechanism is possible.

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012

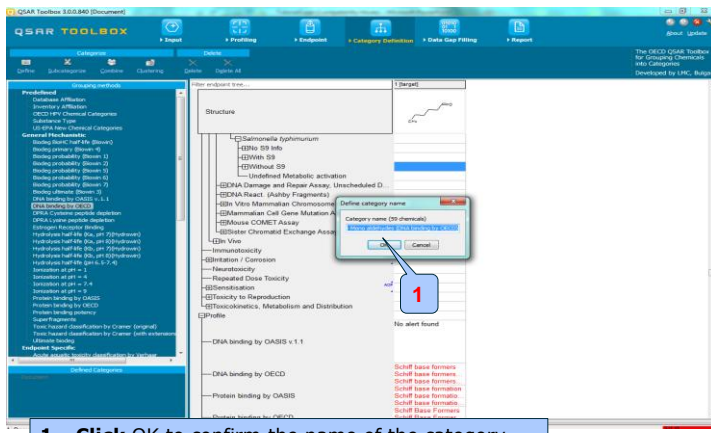
Category Definition

Defining DNA binding by OECD category

1. Highlight "DNA binding by OECD"; 2. Click Define; 3. The target category is Mono-aldehydes belonging to Direct acting Schiff base formers and Schiff base domain, Confirm the category; 4. Click OK

Category Definition

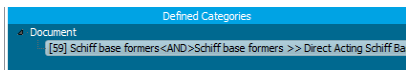
Defining DNA binding by OECD category



Category Definition

Analouges

- The Toolbox now identifies all chemicals corresponding to "Mono-aldehydes" by DNA binding by OECD listed in the databases selected under "Endpoints".
- The name of the category appear in the "Defined Categories" window, the number in brackets is the number of substances belonging to the category (59 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)

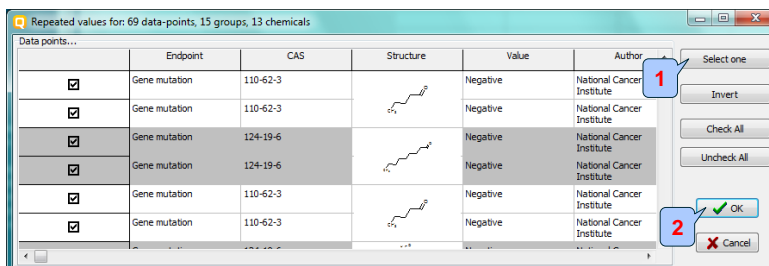


- **In this example, as only databases are selected that contain information for genetic toxicity endpoint, so both options give the same results.**

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.



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 shows the QSAR Toolbox interface with a matrix of experimental results. The matrix has columns for different endpoints and rows for different chemical structures. A red box highlights a cell containing the text: "M. Negative, Negati. M. Negative, Negati. M. Inconclusive, Inc. M. Positive, Positi." This text represents the experimental results for a specific analogue.

Category Definition Side-Bar of experimental data

The screenshot shows the QSAR Toolbox interface with a side-bar of experimental data. A callout box points to a cell in the matrix with the text: "1. Double-click on the cell with measured data to see detailed information in drop down box." This indicates that clicking on the cell will open a detailed view of the experimental data.

Category Definition Recap

- You have identified a mechanistic category consisting of 59 analogous ("Mono-Aldehydes" by DNA binding by OECD classification) with the target chemical (n-hexanal).
- The available experimental data for these 59 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 shows the QSAR Toolbox interface with the endpoint tree expanded. The tree structure is as follows:

- Structure
 - Substance Identity
 - Physical Chemical Properties
 - Environmental Fate and Transport
 - Ecotoxicological Information
 - Human Health Hazards
 - Acute Toxicity
 - Carcinogenicity (3943)
 - Developmental Toxicity / Teratogenicity
 - Genetic Toxicity
 - In Vitro
 - Bacterial Reverse Mutation Assay (e.g. Ames Test)
 - Gene Mutation
 - Salmonella typhimurium (52/166)

1 Target	2	3	4	5	
		M Negative, Negati.	M Negative, Negati.	M Negative, Negati.	M Positive, Positiv

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 3. Click Gene Mutation

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the endpoint tree expanded. The tree structure is as follows:

- Structure
 - Substance Identity
 - Physical Chemical Properties
 - Environmental Fate and Transport
 - Ecotoxicological Information
 - Human Health Hazards
 - Acute Toxicity
 - Carcinogenicity (3943)
 - Developmental Toxicity / Teratogenicity
 - Genetic Toxicity
 - In Vitro
 - Bacterial Reverse Mutation Assay (e.g. Ames Test)
 - Gene Mutation
 - Salmonella typhimurium (49/53)

1 Target	2	3	4	5	
		M Negative	M Negative	M Negative	M Positive, Pos
		M Negative, Negati.	M Negative, Negati.	M Negative, Negati.	M Positive, Pos

1. Open the tree to *Salmonella typhimurium*

Category Definition

Navigation through the endpoint tree

The screenshot shows the QSAR Toolbox interface with the endpoint tree on the left. The 'Genetic Toxicity' category is expanded, and the 'Ames without S9' endpoint is selected and highlighted with a red circle. The table below shows experimental data for 59 chemicals across various endpoints.

Endpoint	1	2	3	4	5	6
Structure	<chem>CCCC=O</chem>	<chem>CCCC=O</chem>	<chem>CCCC=O</chem>	<chem>CCCC=O</chem>	<chem>CCCC=O</chem>	<chem>CCCC=O</chem>
Substance Identity						
Physical Chemical Properties						
Environmental Fate and Transport						
Ecotoxicological Information						
Human Health Hazards						
Acute Toxicity						M. Positive, Neg
Developmental Toxicity / Teratogenicity						
Genetic Toxicity						
In Vitro						
Bacterial Reverse Mutation Assay (e.g. Ames Test)						
Gene Mutation						
Salmonella typhimurium						
Ames S9 info	(49/53)		M. Negative	M. Negative	M. Negative	M. Positive, Pos
Ames S9	(17/66)		M. Negative, Negati	M. Negative, Negati	M. Negative, Negati	M. Positive, Pos
Ames without S9	(5/21)		M. Negative, Negati	M. Negative, Negati	M. Negative, Negati	M. Positive, Pos
Undefined Metabolic activation	(5/9)					M. Positive
EDNA Damage and Repair Assay, Unscheduled D						
EDNA React. (Ashby Fragments)						
In Vitro Mammalian Chromosome Aberratio	(5/11)					
Human Vero Cell Gene Mutation Assay	(3/3)					
Mouse Comet Assay						
Sister Chromatid Exchange Assay						
In Vivo	(5/11)					M. Inconclusive, Inc.
Immunotoxicity						
Reproduction / Conception						
Neurotoxicity						
Repeated Dose Toxicity						
Respiratory						
Toxicity to Reproduction						

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 61

Category Definition

Recap

- You have now retrieved the available experimental data on genetic toxicity for 59 chemicals classified as "Monoaldehydes" by DNA binding by OECD, found in the databases containing mutagenicity data.
- Out of 59 only 12 have experimental mutagenicity data related to the target.
- You are now ready to fill in the data gap.
- 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

The screenshot shows the QSAR Toolbox interface. The 'Data Gap Filling' menu item is highlighted with a blue callout '1'. In the left sidebar, 'Read-across' is selected with a blue callout '4'. The main workspace shows a tree of endpoints, with 'Ames without S9' highlighted in blue with a callout '2'. A 'Possible data inconsistency' dialog box is open, showing a list of data points with checkboxes. A callout '3' points to the 'Read-across' option in the dialog. A callout '5' points to the 'OK' button in the dialog.

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.

QSAR TOOLBOX

Data Gap Filling(Ames without S9) Results of Read across

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
65

QSAR TOOLBOX

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 filling (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.

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
66

Q SAR TOOLBOX

Data Gap Filling(Ames without S9) Results of Read across

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
67

Q SAR TOOLBOX

Data Gap Filling(Ames without S9) Interpretation of the Read across

- Two 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.

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
68

QSAR TOOLBOX

Data Gap Filling(Ames without S9) Subcategorization by DNA binding by OASIS

1. Select Select/filter data; **2. Select** Subcategorize; **3. Select** DNA alert for AMES, MN and CA by OASIS; **4. Examine** dissimilar chemicals and **5. Click** Remove to eliminate dissimilar chemical.

QSAR TOOLBOX

Data Gap Filling(Ames without S9) Interpretation of the Read across

Now all analogues are structurally (Aldehydes) similar, then the prediction could be accepted by

- Click** on Accept prediction
If you want to save the model, and use it for further predictions, then
- 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. A dialog box titled 'Read across prediction of Human Health Hazard@Genetic Toxicity' is open. It contains the text: 'Do you want to collect additional data for chemicals from data matrix for reporting purposes? (This data will be provided in data matrix tables)'. There are 'Yes' and 'No' buttons. A blue callout box with the number '4' points to the 'Yes' button. The background shows a graph of Ames test results and chemical structures.

3. Click **Yes** in order to have in report additional data for the analogues

Data Gap Filling(Ames without S9) Interpretation of the Read across

The screenshot shows the QSAR Toolbox interface. A dialog box titled 'Select nodes to be reported in data matrix tables' is open. It has a tree view of nodes on the left and a list of methods on the right. The 'Genetic Toxicity' node is selected. A blue callout box with the number '1' points to the 'Genetic Toxicity' node. Another blue callout box with the number '2' points to the 'Genetic Toxicity' node. A third blue callout box with the number '3' points to the 'OK' button. A fourth blue callout box with the number '4' points to the 'OK' button. The background shows a graph of Ames test results and chemical structures.

1. **Expand** the Human Health hazard section
2. **Select** Genetic toxicity
3. **Click** OK
4. 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**
 - 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 it 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. If you have trouble review slide number 64.

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
77

Data Gap Filling (Ames with S9)

Results of Read across

The OECD QSAR Toolbox for Grouping Chemicals into Categories
05.10.2012
78

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:

- **Organic functional groups** – The categorization based on this profiler identifies analogues having the same organic functional groups.
- **Structural similarity** – The categorization based on this identifies the most structurally similar chemicals (In this case refine analogues below 50%).

can be removed from the initial list of analogues previously defined by DNA binding by OECD.

Data Gap Filling(Ames with S9) Subcategorization by Organic functional groups

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

Data Gap Filling(Ames with S9) Subcategorization by Organic functional groups

1. If you have trouble review slide number 68.

Data Gap Filling(Ames with S9) Subcategorization by structural similarity

- While it is the method of last resort Toolbox provides the user with the option for subcategorizing by structural similarity.
- This is done by using the “Structural similarity” profiler and then setting the percent similarity desired (see next screen shot).

Data Gap Filling(Ames with S9) Subcategorization by structural similarity

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

- Adjust options dialog:** Shows 'Similar 100%' and 'All categories' selected. A red circle '2' highlights the 'Adjust options' button.
- Similarity options dialog:** Shows 'Structural similarity' selected under 'Measure'. Under 'Options', 'Any atoms distance' is selected. A red circle '1' highlights the 'Similarity' dropdown menu.
- Main window:** Displays a list of chemical structures and their predicted Ames S9 results. The 'Ames S9' column shows '12/55'.

1. Select Structural similarity;
2. Adjust the options of similarity as shown in the next slide

QSAR TOOLBOX

Data Gap Filling(Ames with S9) Subcategorization by structural similarity

Similarity options

Similarity options

Measure

Tanimoto Cosine

Dice Jaccard

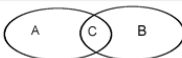
Ochiai Yule

Kulczynski-2


Formula

$$c / 0.5[(a+c) + (b+c)]$$

Description



Example



a	b	c
2	2	10

Similarity = 83.333% [Details](#)

Molecular features

Atom pairs

Topologic torsions

Atom centered fragments

Path

Cycles

Options

Description

The atom-centered fragment is a topological sphere with center a selected atom and radius specified in **Any atom distance**. For aromatic carbon as a center of the sphere is assumed the aromatic system that contains this atom of concern.

Calculation

Fingerprint

Hologram

Average by features

Combine all features

Atom characteristics

Atom type

Count H attached

Count heavy atoms attached

Hybridization

Incident pi-bonds

Valency

Charge

Cyclic

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 85

QSAR TOOLBOX

Data Gap Filling(Ames with S9) Subcategorization by structural similarity

QSAR Toolbox 3.0.0.940 [Document]

Subcategorization

Target

Similar 100%

Differ from Target by

At least one cat. All categories

Similarity groups

(1) Similar (90%-100%)

(2) Similar (50%-60%)

(3) Similar (30%-40%)

(4) Similar (10%-20%)

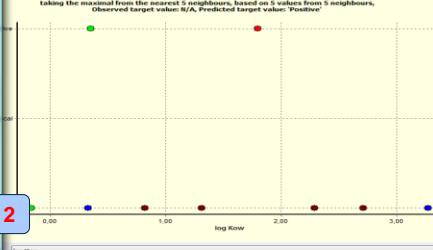
(5) Similar (0%-10%)

Accept prediction

Select different Remove

Read across prediction of Human Health Hazard@Genetic Toxicity.

Taking the maximal from the nearest 5 neighbours, based on 5 values from 5 neighbours, Observed target value: 'N/A', Predicted target value: 'Positive'



1. Select the analogues with less than 50% similarity by highlighting the lower % groups;

2. Click Remove to eliminate dissimilar chemicals (i.e., the ones marked in green).

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 86

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

Now all 5 analogues are structurally similar, then the prediction could be accepted or saved as a category(domain) in the profiler 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

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 88

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

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

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 91

QSAR TOOLBOX

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.

The OECD QSAR Toolbox for Grouping Chemicals into Categories 05.10.2012 92