

OECD QSAR Toolbox v.3.4

Step-by-step example of how to build and evaluate a category based on mechanism of action with protein and DNA binding

Outlook

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

Background

- This is a step-by-step presentation designed to take you through the workflow of the Toolbox in building a category and then performing a preliminary evaluation of the category.
- By now you have experience in using the Toolbox so there will be multiple key strokes between screen shots.

Outlook

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

Objectives

- **This presentation demonstrates:**
- Identifying chemicals which could be grouped into a category.
- Conducting a preliminary evaluation of the category.

Outlook

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

Specific Aims

- To examine the workflow of building a category.
- To introduce the user to new functionalities within selected modules.
- To explain the rationale behind each step of the exercise.
- To demonstrate with a practical example how to use the Toolbox to build a category according to the OECD Guidance on Grouping of Chemicals.

Outlook

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

Exercise

- In this exercise we will build a category around the target chemical 3-mercapto-propionic butyl ester.
- We will performed this by first categorizing using Protein binding by OASIS v1.4 and subsequently subcategorize using EcoSAR classification.
- We will perform a preliminary evaluation of the final category for Ames mutagenicity and skin sensitization.
- The predictions will be made by “read-across” analysis.

Exercise

Side-Bar: Developing a Category Based on Mechanism of Action

- First identify the mechanism and mode of action of a representative member of the category, by profiling the chemical.
- If a specific mechanism or mode is identified, then it is recommended to base the category definition on this mechanism or mode.
- Other members of the category can be found by searching for chemicals which have the same mechanism or mode of action.
- The search results can then be refined by eliminating chemicals which are structurally dissimilar.

Exercise

Side-Bar: Developing a Category Based on Mechanism of Action

- If no specific mechanism or mode of action is identified for a representative member of the category, then it is recommended to base the category definition on close structural similarity.
- In this case members of the category can be found by searching for chemicals which are structurally similar to the target chemical.
- The search results can then be refined by eliminating those chemicals which have specific mechanisms or modes of action.

Outlook

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

Workflow of the exercise

- **The Toolbox has six modules which are used in a sequential workflow:**
 - Chemical Input
 - Profiling
 - **Endpoint**
 - Category Definition
 - **Data Gap Filling**
 - Report

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - **Chemical Input**

Chemical Input Overview

- This module provides the user with several means of entering the chemical of interest or the target chemical.
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Chemical Input

Ways of entering chemicals

- **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 **CAS #**
- **Enter 16215-21-7.**
- **Click Search.** (see next screen shot).

Chemical Input

Input target chemical by CAS#

1. Click on CAS#; 2. Enter 16215-21-7; 3. Click Search.

| Selected | CAS | Depiction | Names | CAS/Name | 2D/Name | CAS/2D |
|-----------|------------|----------------------------|-------|----------------------------------------------|----------------------------------------------|--------|
| 1. Yes | 16215-21-7 | <chem>CCCCOC(=O)CCS</chem> | | 1:: High 1:: C 2:: N 3:: T 4:: U | 1:: High 1:: T 2:: N 3:: U 4:: C | |

Chemical Input

Target chemical identity

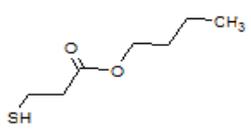
The Toolbox now searches the Toolbox databases and inventories for the presence of the chemical with structure related to the current CAS #. It is displayed as a 2D image.

Search by CAS #

16215-21-7 Tautomeric sets

1

Select All Clear All Invert Selection Selected 1 of 1

| Selected | CAS | Smiles | Depiction | Names | CAS/Name | 2D/Name | CAS/2D |
|-----------|------------|-------------|-------------------------------------------------------------------------------------|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| 1. Yes | 16215-21-7 | CCCCOC(=O)S |  | 1: pro 2: but 3: but 4: pro | 1:: High Qua 1:: Canac 2:: NICNA 3:: TSCA 4:: US HF 2:: High Qua 1:: ECHA 2:: REACI 3:: High Qua 1:: EINEC 4:: Low Qual 1:: US HF | 1:: High Qua 1:: TSCA 2:: NICNA 3:: US HF 4:: Canac 2:: High Qua 1:: ECHA 2:: REACI 3:: High Qua 1:: EINEC 4:: Low Qual 1:: US HF | : High Quality 1:: Canada 2:: ECHA PR 3:: EINECS 4:: NICNAS 5:: REACH E 6:: TSCA 7:: US HPV |

1. Click OK



In case a structure has several CAS numbers or a structure could be related to more than one substance (e.g. in the case of compounds), more than one chemical identity could be retrieved. In this case the user can decide which substance is to be retained for the subsequent workflow.

Chemical Input

Target chemical identity

- You have now selected your target chemical and have its structure.
- Remember from here on the workflow will be based on the structure coded in SMILES.
- **Click** on the box next to “Substance Identity”; this displays the chemical identification information. (see next screen shot).

Chemical Input

Target chemical identity

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Input' module is active, showing a document list on the left with 'CAS: 16215-21-7'. The main workspace displays the chemical structure of the target and a list of substance identity information. The 'CAS Number' field is highlighted with a red circle, containing the value '16215-21-7'. Other fields include 'EINECS:2403435', 'propanoic acid, 3-...', 'butyl 3-mercaptopr...', 'butyl 3-sulfanylpro...', 'propionic acid, 3-m...', 'C7H14O2S', and 'CCCCOC(=O)CCS'. A blue text box at the bottom of the screenshot contains the instruction: 'The workflow on the first module is now complete; click "Profiling" to move to the next module.'

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - **Profiling**

Profiling

Overview

- As you remember “Profiling” refers to the process of retrieving information on the target compound, other than fate and toxicity data.
- Key available information includes likely mechanism(s) of action.
- Background information on a profiler can be viewed by **highlighting** a profiler and **clicking** on “View”.

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 mechanistic and endpoint specific profiling methods should be selected :
 - **Protein binding by OASIS v1.4 – mechanistic grouping**
 - **Protein binding by OECD – mechanistic grouping**
 - **Protein binding alerts for Chromosomal aberration by OASIS v.1.2 – endpoint specific**
 - **Protein binding alerts for skin sensitization by OASIS v1.4 – endpoint specific etc.**

Profiling

Profiling the target chemical (continued)

- **US-EPA New Chemical Categories – predefined**
- **DNA binding by OASIS v.1.4 – mechanistic grouping**
- **DNA binding by OECD - mechanistic grouping**
- **Superfragments – mechanistic grouping**
- **DNA alerts for AMES by OASIS v.1.4 - endpoint specific**
- **DNA alerts for CA and MNT by OASIS v.1.1 - endpoint specific**
- **Aquatic toxicity classification by ECOSAR - endpoint specific**
- **Superfragments – mechanistic grouping**
- **Organic functional groups - empiric**
- **Organic functional groups(nested) - empiric**

Profiling

Profiling the target chemical

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing 'Apply', 'New', and 'Delete' options. The 'Apply' button is circled in red. A red circle highlights the 'General Mechanistic' category in the 'Profiling methods' list, with a '1' callout. Another red circle highlights the 'Apply' button with a '2' callout. The right panel shows a tree view of endpoints and a list of alerts for the target chemical.

1. Check the profilers related to the target endpoint;
2. Click Apply.

Profiling

Profiling the target chemical

The screenshot shows the QSAR Toolbox software interface. The top toolbar includes buttons for 'Profiling' (circled in red with a callout '2'), 'Input', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling methods' panel on the left shows a list of endpoints, with 'Endpoint Specific' (circled in red with a callout '1') selected. The main window displays a tree view of endpoints and a list of results for the selected target chemical.

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 appeared as a dropdown box under the target chemical. (see next screen shot).
- Very specific results are found with the ECOSAR Classification and the Protein Binding profilers.
- These results will be used later in the exercise to build the category.

Profiling

Profiles of "butyl 3-sulfanylpropanoate"

The screenshot shows the QSAR Toolbox interface with the 'Profiling' menu selected. The 'Profile' node in the endpoint tree is highlighted with a red circle and a blue callout box containing the number '1'. The right pane displays the profile results for the selected node, including alerts for Radical, SN2, and Esters.

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

In this case there is structural evidence that the target could interact to DNA and proteins and also has esters and thiol alerts in its structure. This step is critical for next grouping of analogues.

Radical
 Radical >> Radical mechanism via ROS formation (indirect)
 Radical >> Radical mechanism via ROS formation (indirect) >> Thiols
 No alert found

SN2
 SN2 >> Interchange reaction with sulphur containing compounds
 SN2 >> Interchange reaction with sulphur containing compounds >> Thiols an...
 SN2
 SN2 >> SN2 reaction at a sulphur atom
 SN2 >> SN2 reaction at a sulphur atom >> Thiols
 No superfragment

Esters
 Thiols and Mercaptans
 No alert found
 No alert found
 No alert found

SN2
 SN2 >> Interchange reaction with sulphur containing compounds
 SN2 >> Interchange reaction with sulphur containing compounds >> Thiols an...

Profiling

Profiles of "butyl 3-sulfanylpropanoate"

The screenshot displays the QSAR Toolbox Profiling interface. The main window shows the 'Profiling results' for 'butyl 3-sulfanylpropanoate'. A red oval highlights the profile 'SN2 >> Interchange reaction with sulphur containing compounds'. A blue callout box with the number '2' points to this profile. Another blue callout box with the number '1' points to the 'Details' button for this profile. The 'Details' panel shows the profile's description: 'SN2 >> Interchange reaction with sulphur containing compounds >> Thiols and disulfide compounds'. The interface also shows a list of 'Profiling methods' on the left and a 'Metabolism/Transformations' section at the bottom.

- 1. Right click** to see why the target is protein binder
- 2. Select "SN2 >> Interchange reaction with sulphur containing compounds >> Thiols and disulfide compounds"**. Then click details(see next screen shot).

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - **Endpoints**

Endpoints

Overview

- Remember, “Endpoints” refer to the electronic process of retrieving the fate and toxicity data that are stored in the Toolbox.
- Data gathering can be executed in a global fashion or on a more narrowly defined basis.
- Since we are forming a category used further to predict mutagenicity and skin sensitisation, we want to query databases containing mutagenic and skin sensitisation measured data in an effort to gather data for the target chemical.
- Please also remember that when querying for members of the category the Toolbox will search for chemicals which are listed in the selected databases.

Endpoints

Expanding the query domain

- When building a category, we are also interested in finding chemicals for which no experimental data are available, but which fit into the category, and thereby could be assessed as part of the category.
- We therefore have to define the relevant inventory in which to search for chemicals that could be grouped with the target compound.
- For example, the user can choose to search for chemicals in a national index like the US-TSCA inventory or EU EINECS or in more restricted inventories like the OECD HPV list.
- Remember that the process of searching in inventories is time consuming.

Endpoints

Case study

- In this example, we conduct an expanded search for chemicals belonging to a category.
- Select the following databases: Bacterial ISSTY; Genotoxicity OASIS; Toxicity Japan; the two Skin sensitization databases
- Among the inventories, select the “OECD HPVC Inventory”.
- **Click** “Gather data”.
- You will find that no experimental results are available for this chemical.

Endpoints Gather data

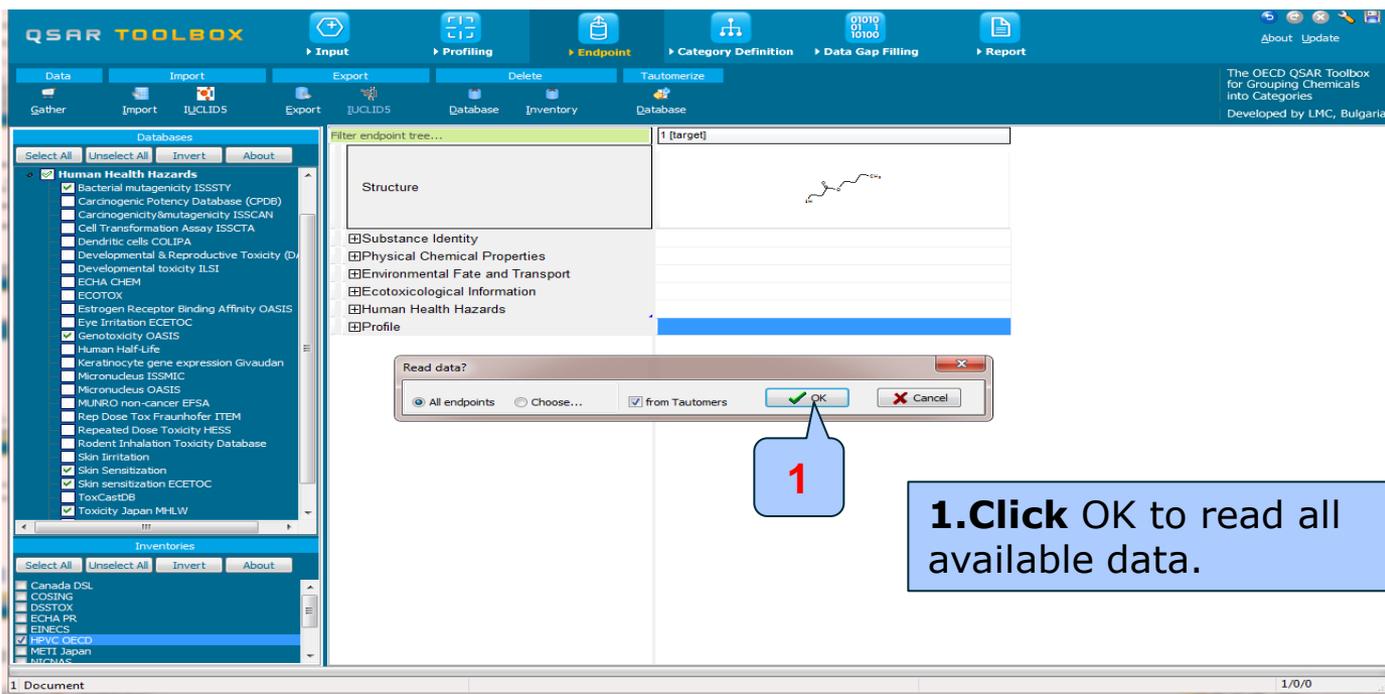
The screenshot displays the QSAR Toolbox interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The toolbar below contains 'Gather', 'Import', 'Export', 'Delete', and 'Tautomerize'. The main window is divided into three panes: 'Databases', 'Filter endpoint tree...', and '1 [target]'. The 'Databases' pane shows a list of databases under 'Human Health Hazards', with 'HPVC OECD' selected. The 'Filter endpoint tree...' pane shows a tree structure with 'Carcinogenicity' selected. The '1 [target]' pane shows a chemical structure. Three red callouts with numbers 1, 2, and 3 point to the 'HPVC OECD' selection, the 'Gather' button, and the 'Gather' button respectively.

1. **Select** databases related to the target endpoint; 2. **Select** HPVC OECD; 3. **Click** Gather.

Endpoints

Process of collecting data

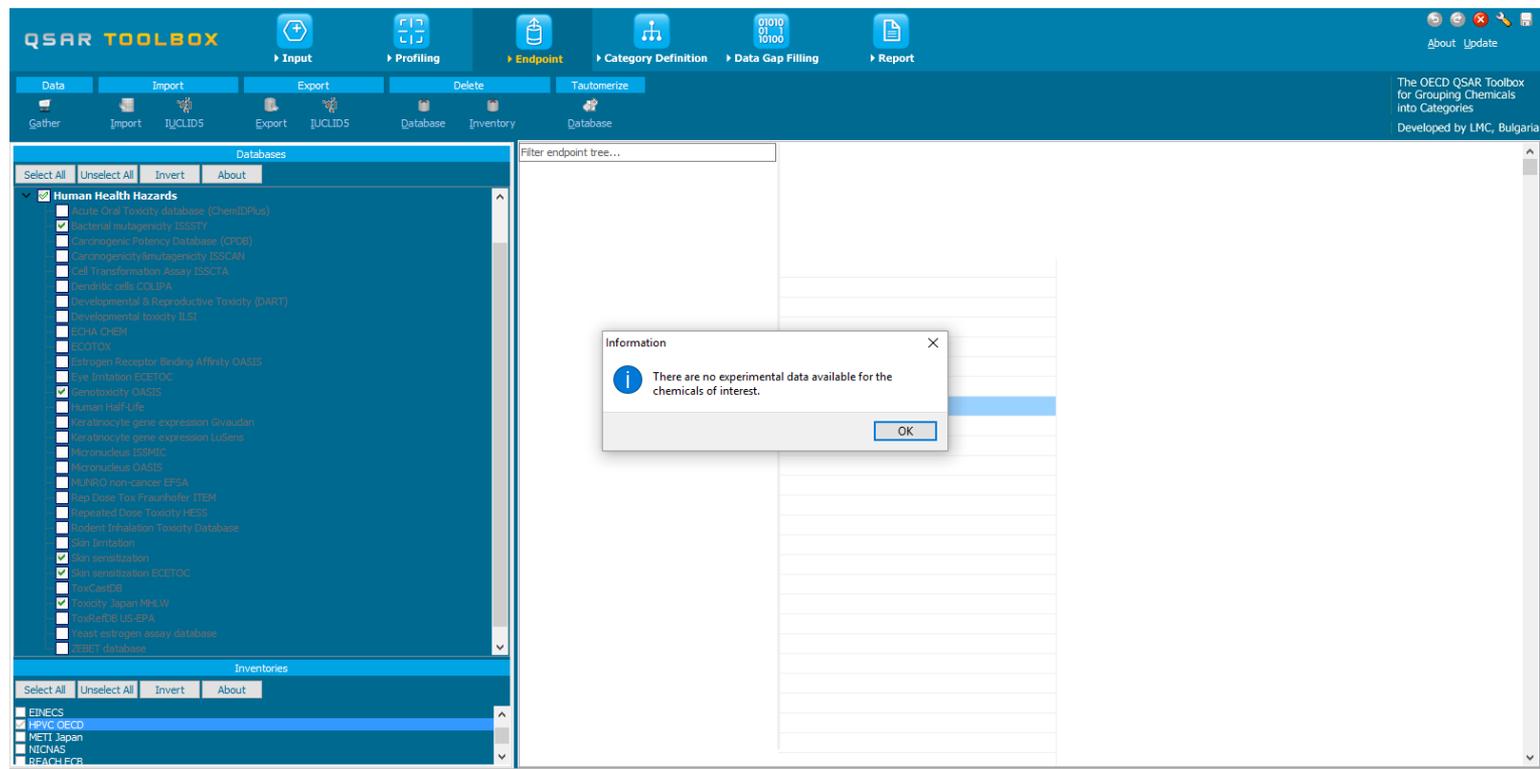
- Toxicity information on the target chemical is electronically collected from the selected datasets.
- A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.



Endpoints

Process of collecting data

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



Endpoints

Recap

- You have entered the target chemical being sure of the correct structure.
- You have profiled the target chemical and found no experimental data is currently available.
- You have checked the databases related to mutagenicity and skin sensitisation experimental data.
- You have defined the inventory in which you want to search for chemicals belonging to the category.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoints
 - **Category definition**

Category definition

Overview

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

Category definition

Grouping methods

- The different grouping methods allow the user to group chemicals into chemical categories according to different measures of “similarity” so that within a category data gaps can be filled by read-across.
- Detailed information about grouping chemical (Chapter 4) could be downloaded from:
<http://www.oecd.org/dataoecd/58/56/46210452.pdf>

Category definition

Case study

- For this example, the user could first select the Protein binding by OASIS v1.4 mechanism of the target chemical and query for all the chemicals with the same mechanism in the selected inventory and databases (see next screen shot).
- The user has first to query according to one profiler and then subcategorise the results step-by-step according to other profilers.

Category definition

Defining Protein binding by OASIS v1.4 category

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Category Definition' menu is open, showing options like 'Define', 'Delete', 'Subcategorize', 'Combine', and 'Clustering'. The 'Define' button is circled in red with a callout '2'. The 'Defined Categories' list on the left shows 'Protein binding by OASIS v1.4' circled in red with a callout '1'. A dialog box in the center asks 'Chemical grouping can be very slow when inventories are selected. Do you want to continue?' with 'OK' circled in blue with a callout '3'. The 'Protein binding by OASIS v1.4' target definition window is open on the right, showing 'Interchange reaction with sulphur containing compounds' circled in red with a callout '4'. The 'All profiles' section lists various chemical reactions.

1. **Highlight** "Protein binding by OASIS v1.4"; 2. **Click** Define, the message that grouping could be slow due to selected inventories appears; 3. **Click** OK; 4. Confirm the category of the target and **click** OK.

Category definition

Defining Protein binding by OASIS v1.4 category

The screenshot displays the QSAR Toolbox software interface. The main window is titled 'Category Definition' and shows a tree view of chemical categories. A dialog box titled 'Define category name' is open, showing the text 'Category name (122 chemicals) Inhibitors (Protein binding by OASIS v1.4)'. A blue callout box with the number '1' points to the 'OK' button. The background shows a tree view of chemical categories and a chemical structure.

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

Category definition

Defining Protein binding by OASIS v1.4 category

- The Toolbox now identifies all chemicals with structural fragment **“Thiols and disulfide compounds”** corresponding to mechanism **“Interchange reaction with sulphur containing compounds”** and domain **“SN2”** by Protein binding by OASIS v1.4 listed in the databases selected under “Endpoints”.
- 122 analogues are identified. Along with the target they form a mechanistic category, used for gap filling.
- The name of the category appears in the “Defined Categories” window, indicating the number of substances belonging to the category.

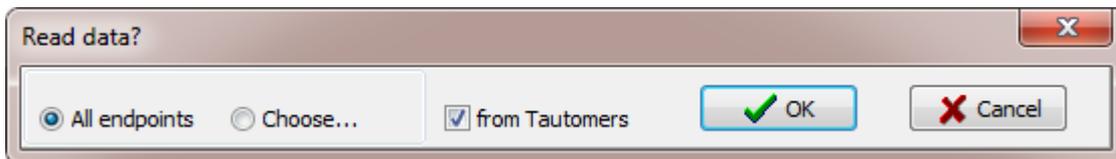
A screenshot of a software interface showing a category definition. The text is displayed in a dark blue box with white text. It shows a dropdown menu with a downward arrow and the word "Document". Below it, the category name is shown: "[122] SN2<AND>SN2 >> Interchange reaction with sulphur containing compounds<AND>SN2 >>".

Document
[122] SN2<AND>SN2 >> Interchange reaction with sulphur containing compounds<AND>SN2 >>

Category definition

Reading data for Analogues

- The Toolbox will now retrieve those chemicals that have the same protein binding mechanism (disulfide formation) as the target compound.
- 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).



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

Category definition

Reading 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: 37 data-points, 18 groups, 18 chemicals

Data points...

| | Endpoint | CAS | Structure | Value | Author | |
|-------------------------------------|---------------|----------|-------------------------------------------------------------------------------------|----------|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input checked="" type="checkbox"/> | Gene mutation | 149-30-4 |  | Negative | Kirkland et al. | <input type="button" value="Select one"/> <input type="button" value="Invert"/> <input type="button" value="Check All"/> <input type="button" value="Uncheck All"/> <input checked="" type="button" value="OK"/> <input type="button" value="Cancel"/> |
| <input checked="" type="checkbox"/> | Gene mutation | 149-30-4 | | Negative | Kazius J, McGuire R., Bursi R. | |
| <input checked="" type="checkbox"/> | Gene mutation | 97-77-8 |  | Negative | Kirkland et al. | |
| <input checked="" type="checkbox"/> | Gene mutation | 97-77-8 | | Negative | Kazius J, McGuire R., Bursi R. | |
| <input checked="" type="checkbox"/> | Gene mutation | 137-26-8 |  | Positive | Kirkland et al. | |
| <input checked="" type="checkbox"/> | Gene mutation | 137-26-8 | | Positive | Kazius J, McGuire R., Bursi R. | |
| <input checked="" type="checkbox"/> | Gene mutation | 52-90-4 |  | Negative | National Cancer Institute | |

1

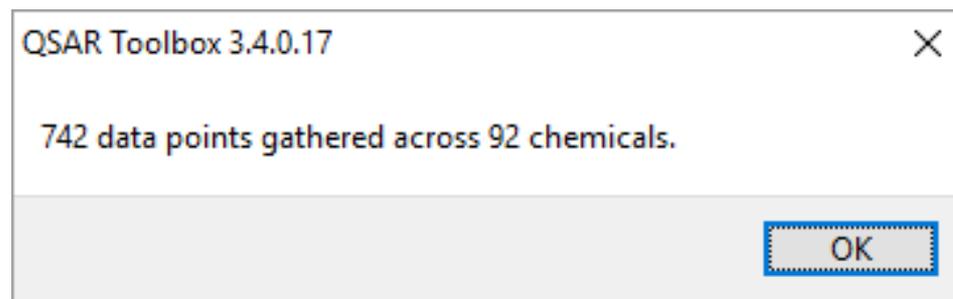
2

1. Click Select one and then 2. Click OK

Category Definition

Read data for Analogues

The system automatically gives indication for the number of gather experimental data points



1. **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 software interface during the 'Category Definition' phase. The main workspace is divided into several panels:

- Grouping methods:** A list of various toxicity endpoints, including 'Acute aquatic toxicity classification by Verhaar (Modified)', 'DNA alerts for Ames by OASIS v.1.4', and 'Skin irritation/corrosion Exclusion rules by BfR'.
- Filter endpoint tree...:** A hierarchical tree of toxicity categories, such as 'Substance Identity', 'Physical Chemical Properties', 'Environmental Fate and Transport', and 'Human Health Hazards'. The 'Human Health Hazards' section is expanded to show 'Acute Toxicity (4/7)', 'Bioaccumulation', 'Carcinogenicity', 'Developmental Toxicity / Teratogen...', and 'Genetic Toxicity'.
- Structure:** A panel showing chemical structures of analogues, labeled '1 [target]', '2', '3', '4', '5', '6', and '7'.
- Data Matrix:** A table with 7 columns representing analogues and rows representing toxicity endpoints. A red box highlights a row in the matrix, indicating experimental results for a specific endpoint across the analogues. The results are summarized as 'M: Positive, Negati...', 'M: Negative, Negat...', 'M: Negative, Negat...', 'M: Negative, Negat...', 'M: Negative, Negative', and 'M: Negative'.

Category definition

Side-bar of experimental data

The screenshot displays the QSAR Toolbox software interface. The main window shows a 'Data points' table with columns for #, Endpoint, Value, Original value, Strain, Test type, Test organisms (species), and Reference source. A 'Data points' dialog box is open, showing a detailed view of the data points. A blue callout box with the number '1' points to a cell in the 'Value' column of the 'Data points' table, which contains the text 'M: Positive, Negati...'. A red circle highlights this cell. The side-bar on the left shows a tree structure of experimental data points, including 'Gene Mutation' and 'Salmonella typhimurium'.

| # | Endpoint | Value | Original value | Strain | Test type | Test organisms (species) | Reference source |
|---|---------------|----------------------------|----------------------------|----------------|---------------------------------------------------|--------------------------|------------------------------|
| 1 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 100 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 2 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 92 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 3 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 98 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 4 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 100 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 5 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 102 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 6 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | TA 92 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | CCRIS database |
| 7 | Gene mutation | Positive (Gene mutation I) | Positive (Gene mutation I) | No strain info | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | Journal of Mutation Research |
| 8 | Gene mutation | Negative (Gene mutation I) | Negative (Gene mutation I) | TA 98 | Bacterial reverse mutation assay (e.g. Ames test) | Salmonella typhimurium | Journal of Mutation Research |

1. Double-click on the cell with measured data to see detailed information for the data points.

Category definition

Side-bar of experimental data

- You have identified a mechanistic category consisting of 122 analogous (**Protein thiol-disulphide interchange**) by Protein binding by OASIS v1.4 classification.
- The available experimental data for these 122 similar chemicals are collected from the previously selected databases under Endpoint section.
- The user can proceed with subcategorisation process.

Category definition

Categorization by ECOSAR

- After the available data has been retrieved, the user can then further subcategorize the results according to “ECOSAR Classification”.
- These steps are summarized in the next screen shot.

Category definition

Subcategorization by ECOSAR by "All categories"

The screenshot displays the QSAR Toolbox software interface. The 'Subcategorization' window is active, showing various grouping methods on the left. The 'ECOSAR' profiler is selected. The 'Adjust options' dialog is open, with 'All categories' selected. The 'Remove' button is highlighted in blue. A 'Define category name' dialog box is open, showing the category name 'Categorized: Aquatic toxicity classification by ECOSAR'. The main window shows a list of chemical structures and their associated hazard categories.

1. **Click** Subcategorize; 2. **Select** ECOSAR profiler;
3. **Select** "All categories"; 4. **Remove**; 5. Confirm the new category by **clicking** OK.

Category definition

Subcategorization by ECOSAR by “All categories”

- **Note** that the target chemical belongs to two ECOSAR classes:
 - Thiols and mercaptans
 - Esters
- The user eliminates all chemicals which do not belong to both these two classes by selecting the “**All categories**” radio-button.
- The result is that only 8 additional category members along with the target are identified (see next screen shot).

Category definition

Subcategorization by ECOSAR by "All categories"

The screenshot displays the QSAR Toolbox interface during the 'Category Definition' process. The left sidebar lists various grouping methods, with 'Endpoint Specific' methods expanded. 'Aquatic toxicity classification by ECOSAR' is highlighted with a red circle. The main workspace shows a table of chemical structures with associated hazard information. A red oval highlights the row for 'Aquatic toxicity classification by ECOSAR', showing '(4/4)' and 'M: Positive' for three different chemical structures.

| Structure | 1 [target] | 2 | 3 | 4 | 5 | 6 | 7 |
|----------------------------------|-------------|---|---|---|-------------|---|-------------|
| | | | | | | | |
| Substance Identity | | | | | | | |
| Physical Chemical Properties | | | | | | | |
| Environmental Fate and Transport | | | | | | | |
| Ecotoxicological Information | (4/4) | | | | | | |
| Human Health Hazards | M: Positive | | | | M: Positive | | M: Positive |
| Profile | | | | | | | |

Category Definition

The result of subcategorization by ECOSAR by
“All categories”

- In this example, the retrieved chemicals have identical mechanistic profiles.
- The number of chemicals retrieved is therefore low.
- One could consider building the category allowing for two subcategories to remain.
- For example, the user could decide to build a category with the same protein binding mechanism but allowing chemicals belonging to either one of the two ECOSAR classes.
- This is done by selecting the radio-button “At least one category” and “pruning” all others (see next screen shot).

Category Definition

The result of subcategorization by ECOSAR by "At least one"

The screenshot displays the QSAR Toolbox software interface. The main window shows a list of grouping methods on the left, with 'Subcategorize' highlighted in red (2). A 'Subcategorization' dialog box is open, showing a list of grouping methods and a radio button for 'At least one category' selected in red (3). A list of categories is shown on the right, with 'Esters Thioles and Mercaptans' highlighted in red (4). A 'Define category name' dialog box is open, showing the category name 'gorized: Aquatic toxicity classification by ECOSAR' entered in the text field (6). A 'Remove' button is highlighted in red (5). A '1' callout points to a category in the main window.

1. **Click** on first category 2. **Click** Subcategorize; 3. **Select** ECOSAR profiler; 4. **Select** "At least one category"; 5. Remove; 6. Confirm name of new category and **click** OK.

Category Definition

The result of subcategorisation by ECOSAR by
“At least one”

- The result of the second subcategorisation is a chemical category with 38 members along with the target (see next screen shot).
- After identifying category members according to a specific mechanism or mode of action it is always necessary to verify whether any of the selected chemicals have additional mechanisms or modes of action, which would make them unsuitable for the category. This can be done by using the “Subcategorisation” procedures.
- For example, there could be chemicals that have specific DNA binding mechanisms, due to additional functional groups in the molecule (this is demonstrated in the next screen shots).

Category Definition

The results of eliminating dissimilar chemicals

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'. The 'Category Definition' button is highlighted with a red circle and a callout box containing the number '1'. Below this, the 'Subcategorization' dialog box is open, showing a list of predefined and general mechanistic categories. The 'DNA binding by OASIS v.1.4' category is highlighted with a red circle and a callout box containing the number '2'. The dialog box also shows options for 'Differ from target by' and 'Analogues'. The main window displays a table of chemical structures and their classifications. The table has columns labeled 2 through 8. The first row shows chemical structures for 'Radical' and 'Thiols'. The second row shows a classification 'M: 1.29E3 mg/kg, ...'. The third row shows 'M: Negative, Negat...' and 'M: Negative, Negat...'. The fourth row shows 'M: Positive' and 'M: Negative, Negat...'. The bottom of the dialog box shows 'Selected 0 (37/37)' and buttons for 'Select different' and 'Remove'.

Category Definition Recap

- In this example, no outliers in terms of mechanism of action are identified and no chemicals will be eliminated from the category.
- The result is a group of chemicals that can bind to protein by the same mechanism (disulfide formation) and that belong to either the ECOSAR class(es) of “Thiols (mercaptans)” or “Esters AND Thiols (mercaptans)”.
- Chemicals with other specific mechanisms or modes of actions have been eliminated so it is expected that the remaining chemicals have similar behaviour for many regulatory endpoints.
- Note that for aquatic toxicity, it is expected that differences in trends could be observed between chemicals belonging to the ECOSAR class(es) of “Thiols (mercaptans)” or “Esters AND Thiols (mercaptans)” and therefore these should be considered as two subcategories.

Category Definition

Preliminary evaluation of the category

- In this particular example, insufficient data is available to fill data gaps and further testing may be necessary.
- Nevertheless, for Ames mutagenicity and sensitisation, the coherence and consistency of the available data can be assessed.
- Regarding point mutation according to the Ames test, the Toolbox has identified 20 chemicals across category consisting of 38 analogues for which results are available (see next screen shot).
- As point mutation is a “qualitative” endpoint, the data gap can be filled by read-across.

Category Definition

Selecting Data Point for AMES mutagenicity

- In this example **navigate** through the endpoint tree by opening the nodes of tree.
- **Highlight** the blank space for “AMES mutagenicity” under the target chemical.
- In this case we mixed all experimental results with different metabolic activation.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category definition
 - **Data gap filling**
 - **Ames**

Data Gap Filling(Ames)

Apply read-across

1. In order to used more qualitative category highlight the category with bigger number of experimental data [38]. **2. Highlight** the cell in data matrix corresponding to mixed Ames under the target chemical. It will be empty. **3. Go** to data gap filling.

Data Gap Filling(Ames)

Apply read-across

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' menu is active. The left sidebar shows 'Data Gap Filling Method' with 'Read-across' selected. The main window displays a tree view of endpoints, with 'Bacterial Reverse Mutation Assay (e.g. Ames ...)' selected. A 'Possible data inconsistency' dialog box is open, showing a list of inconsistencies with checkboxes. The 'OK' button in the dialog is highlighted with a red callout '3'. A red callout '1' points to 'Read-across' in the sidebar, and a red callout '2' points to the 'Apply' button in the top menu bar.

1. Select Read-across; 2. Click Apply; An window alerting you for possible data inconsistencies appears; 3. Click OK.

Data Gap Filling(Ames)

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

Structure

1 [target] 2 5 6 7 8 9 10

Gene Mutation (20/164)

M. Negative, Negat M. Negative, Negat M. Negative, Negat M. Negative, Negat M. Negative, Negative M. Negative, Negat M.

Descriptors Prediction

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 74 values from 5 neighbour chemicals, Observed target value: N/A, 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

38 Subcategorized: Aquatic toxicity classification by ECOSAR Create prediction by gap filling 0/1 1/170

Data Gap Filling(Ames)

Interpretation of Read-across

- The all 20 analogues are non-mutagenic in the Ames assay, except for three chemicals.
- The same non-mutagenic potential (Negative) is therefore, predicted with confidence for the target chemical.
- 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 build a group of mechanistically and structurally similar analogues. Hence, structurally similar analogues interact to DNA at same mechanism.
- Perform subcategorizations by DNA alerts for AMES, MN and CA by OASIS v.1.4

Data Gap Filling(Ames)

Subcategorization by DNA alerts for AMES, MN and CA by OASIS v.1.4

The screenshot displays the QSAR Toolbox interface during a data gap filling process. The 'Subcategorization' window is active, showing a list of grouping methods on the left. A callout box labeled '2' highlights 'DNA alerts for AMES by OASIS v.1.4' in this list. The central panel shows 'No alert found' and 'Analogues' sections. A callout box labeled '3' points to the 'Remove' button at the bottom of the 'Analogues' list. The main matrix view on the right shows chemical structures and their predicted target values. A callout box labeled '1' points to the 'Subcategorize' button in the 'Accept prediction' panel. The main matrix shows a 'prediction of Gene mutation' alert for one of the chemicals.

1. Subcategorization
2. **Select** DNA alerts for AMES, MN and CA by OASIS v.1.4;
3. **Click** Remove to eliminate dissimilar chemical.

Data Gap Filling(Ames)

Results after Subcategorization by DNA alerts for AMES, MN and CA by OASIS v.1.4

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

Structure

1 [target] 2 A 5 6 7 8 9 10

Gene Mutation (18/158)

M Negative, Negat M Negative, Negat M Negative, Negat M Negative, Negat M Negative, Negative M Negative, Negat M

Descriptors Prediction

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 74 values from 5 neighbour chemicals, Observed target value: N/A, Predicted target value: 'Negative'

Descriptor X: log Kow

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

38 Subcategorized: Aquatic toxicity classification by ECOSAR Create prediction by gap filling 17/0

Data Gap Filling(Ames)

Subcategorization by in vitro mutagenicity (Ames test) alerts by ISS

The screenshot displays the QSAR Toolbox interface during a subcategorization task. The main workspace shows a grid of chemical structures and their predicted values for Ames test alerts. A callout box labeled '1' points to the 'Subcategorize' option in the 'Accept prediction' panel. Another callout box labeled '2' points to the 'in vitro mutagenicity (Ames test) alerts by ISS' method in the 'Grouping methods' list. A third callout box labeled '3' points to the 'Remove' button in the 'Analogues' list, which is used to eliminate a chemical that is dissimilar to the target.

1. Subcategorization
2. Select in vitro mutagenicity (Ames test) alerts by ISS;
3. Click Remove to eliminate dissimilar chemical.

Data Gap Filling(Ames)

Results after Subcategorization by in vitro mutagenicity (Ames test) alerts by ISS

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

Structure

1 [target] 2 5 6 7 8 9 10

Gene Mutation (17/147)

M: Negative, Negat M

Descriptors Prediction

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 64 values from 5 neighbour chemicals, Observed target value: N/A, Predicted target value: 'Negative'

Descriptor X: log Kow

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

Data Gap Filling(Ames)

Subcategorization by Organic functional groups (nested)

The screenshot displays the QSAR Toolbox interface during the Data Gap Filling process. The 'Subcategorization' window is active, showing a list of chemical structures and their predicted categories. The 'Grouping methods' list on the left includes 'Organic functional groups (nested)'. The 'Subcategorization' window shows a list of chemical structures with their predicted categories, and a 'Remove' button at the bottom. The main window shows a grid of chemical structures with their predicted categories, and a scatter plot of log Kow values. A text box in the center provides instructions: '1. Subcategorization', '2. Select Organic functional groups (nested);', '3. Click Remove to eliminate dissimilar chemical.'

Data Gap Filling(Ames)

Results after Subcategorization by Organic functional groups (nested)

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

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

Target Endpoint: Human Health Hazards Genetic Toxicity In Vitro Bacterial Reverse Mutation Assay (e.g. Ames Test) Gene Mutation

Structure

1 [target] 6 7 10 21 37

Gene Mutation (5/34) M: Negative, Negat M: Negative, Negat M: Negative, Negat M: Negative, Negative M: Negative

Read across prediction of Gene mutation, taking the highest mode from the nearest 5 neighbours, based on 34 values from 5 neighbour chemicals, Observed target value: N/A, Predicted target value: 'Negative'

Gene mutation (obs)

log Kow

Descriptor X: log Kow

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

38 Subcategorized: Aquatic toxicity classification by ECOSAR Create prediction by gap filling 1/10

Data Gap Filling(Ames)

Interpretation of Read-across

- All results of the category members are consistent. They all are negative in the Ames test. The available results for point mutation therefore appear to confirm the adequacy of the category.
- The same exercise can be performed for skin sensitisation (see next screen shot).

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category definition
 - **Data gap filling**
 - Ames
 - **Skin Sensitization**

Data Gap Filling(Skin sensitization) Apply Read-across

The screenshot shows the QSAR Toolbox interface during a 'Data Gap Filling' operation. The 'Data Gap Filling Method' is set to 'Read-across'. The 'Filter endpoint tree...' on the left shows 'Skin sensitisation II (ECETOC)' selected. The main workspace displays a table of data points for various chemical structures. A dialog box titled 'Possible data inconsistency' is open, showing options for 'Gap filling scale/unit' and 'converted data'. The 'OK' button in the dialog is highlighted with a red circle and callout 4.

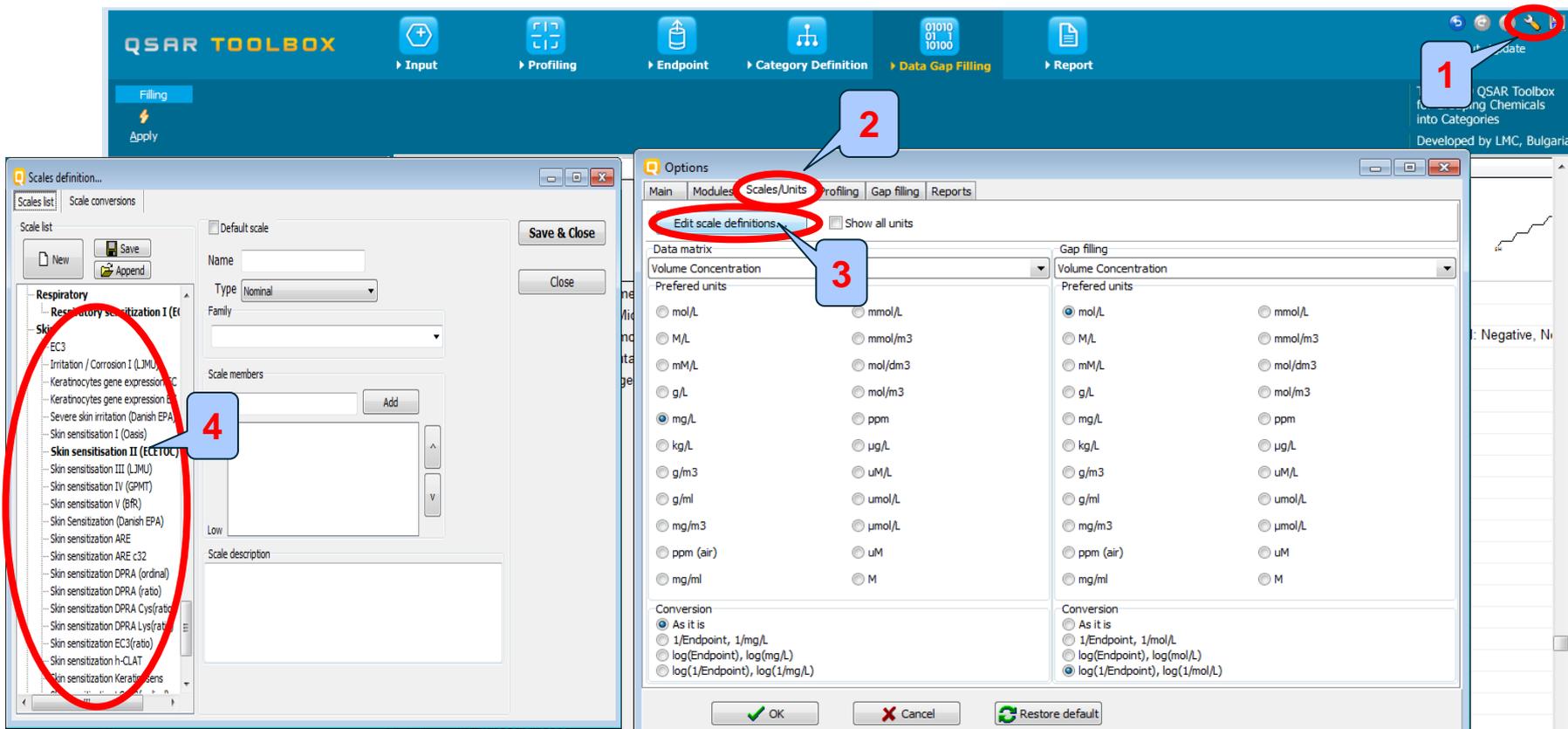
1. Highlight the data endpoint box corresponding to Skin sensitisation under the target chemical, note it will be empty; **2. Select** Read-across; **3. Click** Apply, an insert window alerting you for possible data inconsistencies appears (this issue is related to Scales, see more details on next screen shot); **4. Click** OK.

Data Gap Filling(Skin sensitization)

Apply Read-across Scales

- This window shows all available scales corresponding to skin experimental data.
- The checked scale is the default one. This means that all other are converted into the default one.
- To see scale details go to: **Options**; Click on **Edit scale definitions** button. 
- This conversion is performed in Toolbox in order of standardize skin sensitisation experimental data.

Data Gap Filling(Skin sensitization) Result Read-across



1. Click Options; 2. Select Options and Click Units; 3. Click Edit scale definitions inconsistencies 4. See Skin.

Data Gap Filling (Skin sensitization) Result Read-across

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Profiling Profiling Schemes

Apply New View Delete

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

1 [target] 2 23 30 31 34

Structure

EC3 (5/5) M. Positive M. Positive M. Negative M. Positive M. Positive

Descriptors Prediction

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

EC3 (obs.)

log Kow

Descriptor X: log Kow

Accept prediction
Return to matrix

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

38 Subcategorized Aquatic toxicity classification by ECOSAR Create prediction by gap filling 0/1 17/70

Data Gap Filling (Skin sensitization) Result Read-across

QSAR Toolbox 3.4.0.17 - chemical details for 'SC1=Nc2cccc2O1'

Smiles: SC1=Nc2cccc2O1
 Chem. name(s): 3h-benzoxazole-2-thione; 2(3h)-benzoxazolethione; 2-mercaptobenzoxazole; benzoxazole-2-thiol; 1,3-benzoxazole-2-thione
 CAS No.: 2382-96-9

Current subcategorization: Database Affiliation

Chemical Reactivity COLIPA
 ECOTOX
 Experimental pKa
 GSH Experimental RC50
 Keratinocyte gene expression Givaudan
 Phys-chem EPISUITE
 Skin sensitization

| Descriptor | Units | Value | Endpoint reference | Units |
|------------------|-------|-------|----------------------|-------|
| log Kow | | 2.41 | Dermatitis, 21(1); 8 | |
| Molecular weight | Da | 151 | | |

Descriptor X: log Kow

In this case there is one chemical with negative experimental data,
1. Double click on a dot to see detailed information
2. Click difference to target

Data Gap Filling (Skin sensitization) Result Read-across

The screenshot displays the QSAR Toolbox interface for skin sensitization data gap filling. The 'Subcategorization' sidebar on the left lists various grouping methods, with 'Similar [0%-10%]' highlighted in red and labeled '1'. Below it, 'Selected 2 (3/5)' is shown, with a 'Remove' button labeled '2'. The main workspace shows a list of chemicals with their predicted categories: (5/5) M: Positive, M: Positive, M: Negative, M: Positive, and M: Positive. The 'Read across prediction' plot shows a horizontal line at log Kow = 0.00, with several data points plotted. The text above the plot reads: 'Read across prediction of EC3, taking the highest mode from the nearest 5 neighbours, based on 5 values from 5 neighbour chemicals, Observed target value: N/A, Predicted target value: Positive'.

The red colour profilers are indication for analogues having categories different to the target.

1. In this case, the subcategorization by **Structural similarity (default options)** is applied
2. Analogues with similarity less than 20% are removed.

Data Gap Filling (Skin sensitization)

Accept the prediction

The screenshot displays the QSAR Toolbox software interface. The top navigation bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The left sidebar shows 'Data Gap Filling Method' with options: Read-across (selected), Trend analysis, and (Q)SAR models. The 'Target Endpoint' is 'Human Health Hazards Sensitisation Skin In Vivo LLNA EC3'. The central workspace shows a 'Structure' field with four chemical structures and their corresponding predictions: (4/4) R. Positive, M. Positive, M. Positive, and M. Positive. A plot titled 'Read across prediction of EC3' shows 'EC3 (obs.)' on the y-axis (Positive/Negative) and 'Descriptor X: log Kov' on the x-axis. A dialog box is open with the message 'The current prediction was accepted' and an 'OK' button. The right sidebar has 'Accept prediction' and 'Return to matrix' options. A blue callout box at the bottom contains the instructions: '1. Click Accept the prediction; 2. Click OK; 3. Click Return to the matrix.'

Data Gap Filling Recap

- Based on the profiling results of a target chemical, you have built a category with two subcategories.
- You have gathered available experimental results for the members of the category.
- You have performed a preliminary evaluation of the category based on the available experimental data.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category definition
 - **Data gap filling**
 - Ames
 - Skin Sensitization
 - **Evaluation by QSAR model**

Data Gap Filling

Preliminary evaluation using a QSAR model

- The robustness of the category could be further evaluated with the help of external QSARs from the Toolbox library of QSARs.
- To access the available models for a given endpoint, **highlight** a cell in the matrix for a given endpoint (e.g. Sensitisation>>skin) and **click** on “(Q)SAR models”
- The list of available QSAR models related to the given endpoint appear in the box “QSAR models” (see next screen shot).
- In this example, only one model “DB Danish EPA Skin sensitisation” is available.

Data Gap Filling

Preliminary evaluation using a QSAR model

The screenshot shows the QSAR Toolbox interface during a Data Gap Filling operation. The 'Data Gap Filling Method' sidebar on the left has three options: 'Read-across', 'Trend analysis', and '(Q)SAR models', with the latter being selected and circled in red. The 'Target Endpoint' sidebar shows 'Human Health Hazards Sensitisation Skin In Vivo Undefined Assay Skin Sensitisation'. The main area displays a 'Structure' tree on the left and a data matrix on the right. The matrix has columns for chemical structures and rows for various endpoints. A cell in the 'Skin Sensitisation' row is highlighted with a red circle and labeled '1'. The '(Q)SAR models' option is labeled '2', and the 'Skin sensitisation (Danish EPA DB)' model is labeled '3'.

1. Highlight a cell in data matrix associated with Skin Sensitization endpoint; **2. Select** (Q)SAR models; **3. Only** Skin sensitisation (Danish EPA) is available (by default).

Data Gap Filling

Background information on the external QSAR model

- Before applying a QSAR model it is recommended to consult its documentation.
- Perform **right click** over the name of the model and select "Model about".
- A window with summary information on the available models for that endpoint will appear (see next screen shot).

Data Gap Filling

Background information on the external QSAR model

The screenshot shows the QSAR Toolbox interface with the 'Data Gap Filling' tab active. On the left, under 'Relevant (Q)SAR models', the model 'Skin sensitisation (Danish EPA DB)' is highlighted. A right-click context menu is open over this model, with 'Model About' selected. An 'About' dialog box is open, displaying the following information:

- Name:** Estimated DB (Q)SAR
- Short description:** The results from this model are based on the Danish (Q)SAR Database. To support the regulatory assessment of chemicals, the Danish Environmental Protection Agency (EPA) constructed a (Q)SAR database comprising predictions made by some 70 models for about 166,000 organic chemicals for a wide range of different endpoints. In 2004, a
- Disclaimer:**
- Donator(s):** Danish EPA
- Author(s):** Danish EPA
- Website:** <http://qsar.food.dtu.dk>
- Details:**
 - Adopted:** Toolbox 2.0. October 2010

Three numbered callouts indicate the steps: 1. Right-click on the model; 2. Select 'Model About' from the context menu; 3. Click the 'Close' button on the 'About' dialog.

1. Perform right click over the model; 2. Select Model About; 3. Click Close.

Data Gap Filling

Applying of the model to the members of the category

- The model can be used to evaluate the category by applying it to all the chemicals in the category and analysing the results.
- To apply the model simultaneously to all the chemicals:
 - in the category, **select** the model, **right-click** upon it and select "Predict Endpoint" and "All chemicals".
 - in the domain of the model, **select** the model, **right-click** upon it and select "Predict Endpoint" and "All chemicals in domain" (see next screen shot).

Data Gap Filling

Applying of the model to the members of the category

The screenshot illustrates the 'Data Gap Filling' process in the QSAR Toolbox. The interface is divided into several sections:

- Top Bar:** Navigation tabs for Input, Profiling, Endpoint, Category Definition, Data Gap Filling, and Report.
- Left Panel:** 'Data Gap Filling Method' section with radio buttons for 'Read-across', 'Trend analysis', and '(Q)SAR models'. Below it, 'Target endpoint' is set to 'Human Health Hazards Sensitisation Skin In Vivo Undefined Assay Skin Sensitisation'. A 'Relevant (Q)SAR models' list is shown with a red callout '1' pointing to a model.
- Center Panel:** 'Filter endpoint tree...' showing a hierarchical list of endpoints. A red callout '2' points to the 'Predict' option in the context menu.
- Right Panel:** A grid showing chemical structures and their predicted results. A red callout '3' points to the 'Predict Chemicals in Domain' option in the context menu.

1. Select the model and right click; 2. Select Predict; 3. Select Predict chemicals in Domain.

Data Gap Filling

Results from the model

- For all chemicals in the category which are also in the applicability domain of the model, estimations are generated and inserted into the data matrix, preceded by the letter “Q”.
- Estimations are generated for 20 chemicals out of 36 in the category (see next screen shot).

Data Gap Filling Results from the model

The screenshot displays the QSAR Toolbox software interface during the 'Data Gap Filling' process. The main window is divided into several sections:

- Navigation Bar:** Includes icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling (active), and Report.
- Left Sidebar:** Contains 'Data Gap Filling Method' (Read-across, Trend analysis, QSAR models), 'Target Endpoint' (Human Health Hazards, Sensitisation, Skin, In Vivo, Undefined Assay, Skin Sensitisation), 'Run ECOSAR' button, 'Relevant (Q)SAR models' (CREATE A NEW QSAR, Skin sensitisation (Danish EPA DB)), and 'Rank Models' button.
- Filter endpoint tree...:** A hierarchical tree view showing endpoints such as In Vivo, Immunotoxicity, Irritation / Corrosion, Neurotoxicity, Photoinduced Toxicity, Repeated Dose Toxicity, Sensitisation, Respiratory Tract, Skin, In Chemico, In Vitro, In Vivo, GPMT, LLNA, EC3, Miscellaneous, Undefined Assay, Skin Sensitisation, Undefined Endpoint, ToxCast, and Toxicity to Reproduction.
- Main Results Table:** Displays chemical structures in columns 1-6. Row 1 shows a chemical structure with a 'Q: Positive' prediction for 'Skin Sensitisation'. Other rows show 'M: Positive' and 'Q: Positive' predictions for various endpoints.
- Information Dialog:** A pop-up window stating 'Predicted 20 out of 36 chemicals' with an 'OK' button.
- Status Bar:** Shows '36 Subcategorized: Aquatic toxicity classification by ECOSAR 1', 'Create predictions by QSAR', '0/1', and '1/1/0'.

Data Gap Filling

Interpretation of the Results

- Those chemicals which are in its applicability domain, are predicted to be skin sensitisers, with the exception of cyclohexanethiol (Equivocal).
- Overall the model tends to confirm the evaluation of the category based on the available experimental data, namely that the chemicals in this category are probably skin sensitisers, except for some chemicals which are not bioavailable.
- Based on this evaluation, it could be envisioned to limit the testing plan for this endpoint to identify those chemicals which are not skin sensitisers due to low bioavailability (e.g. $\text{Log } K_{ow} > 6$).

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- **Workflow of the exercise**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category definition
 - Data gap filling
 - **Report**
 - Save the prediction result

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.

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 open, showing options like 'Create', 'Print', 'Close', 'Save as', 'Register', 'Unregister', 'Update', 'Clone', and 'Design'. A red circle highlights the 'Create' button, with a callout '2'. Below the menu, there are two main panels. The left panel, titled 'Available data to report', contains a list of predictions. A red oval highlights this list, with a callout '1' pointing to a specific prediction entry. The right panel, titled 'Prediction [4]', shows a preview of a report page. The report title is 'QSAR Toolbox prediction for single chemical'. The report content includes a reference to OECD guidance documents and a description of the report's purpose.

1. Select the prediction in the window Available data to report, then **2. Click** Create.

Outlook

- Background
- Objectives
- Specific Aims
- The exercise
- Workflow of the exercise
- **Saving the prediction result**

Saving the prediction result

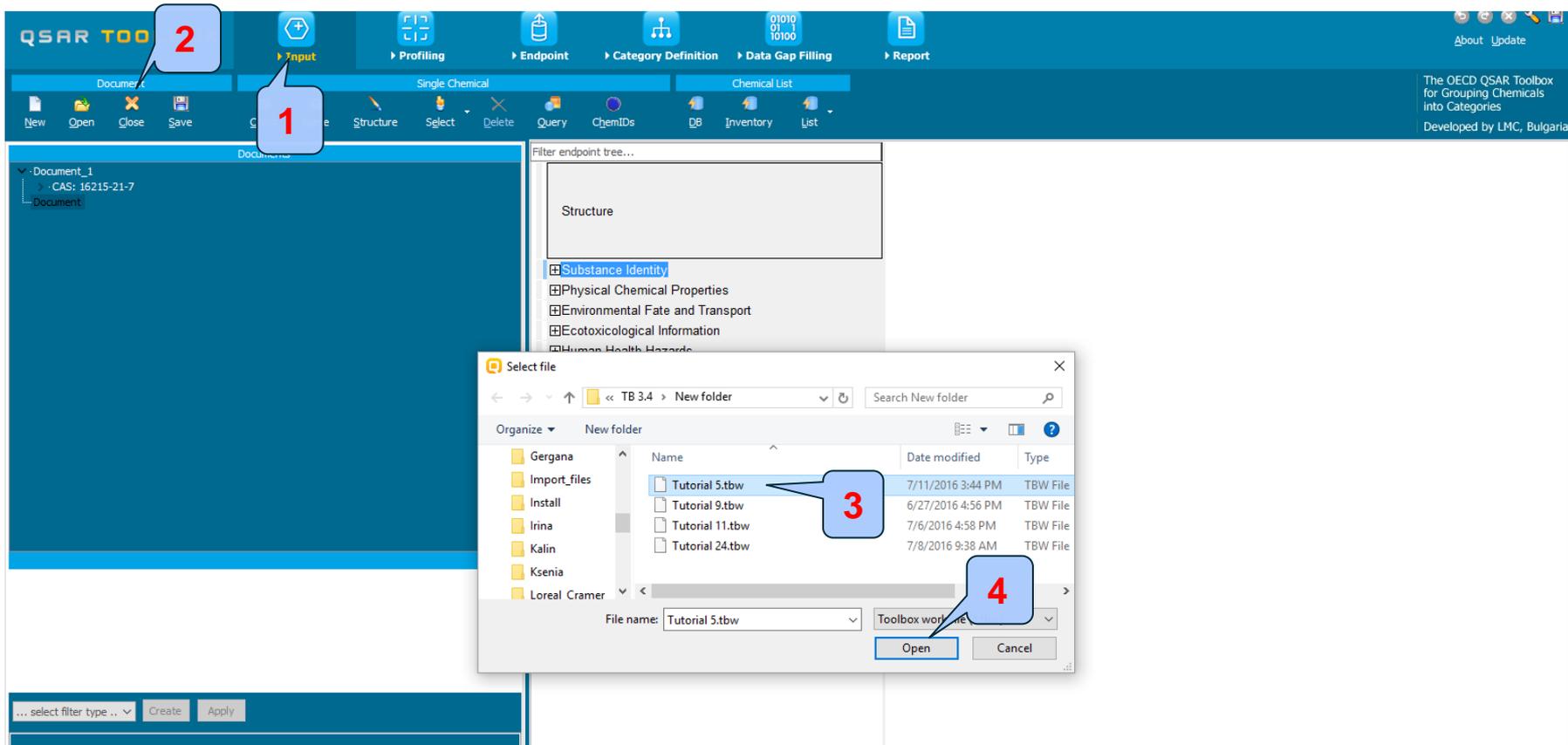
- This functionality allow storing/restoring the current state of Toolbox documents including loaded chemicals, experimental data, profiles, predictions etc, on the same computer. The functionality is implemented based on saving the sequence of actions that led to the current state of the Toolbox document and later executing these actions in the same sequence in order to get the same result(s).
- Saving/Loading the file with TB prediction is shown on next screenshots

Saving the prediction result

The screenshot shows the QSAR Toolbox interface with a report titled "Prediction of Skin Sensitisation for cyclohexanethiol" open. A "Save As" dialog box is overlaid on the report, showing a file explorer view of a "New folder" in the "TB 3.4" directory. The dialog contains a list of files: Tutorial 9.tbw, Tutorial 11.tbw, and Tutorial 24.tbw. The "File name" field is set to "Tutorial 5.tbw" and the "Save as type" is "Toolbox work file (*.tbw)". Three callouts are present: Callout 1 points to the "Save" button at the bottom right of the dialog; Callout 2 points to the "File name" text box; Callout 3 points to the "Save" button at the bottom of the dialog. The background report shows a table of predictions and a section titled "GUIDANCE DOCUMENT ON THE ACTIVITY RELATIONSHIPS MODELS" with a sub-section "GUIDANCE ON INFORMATION MANAGEMENT / CHAPTER R.6: QSARS AND GROUPING OF CHEMICALS" published by ECHA in May 2008.

1. Click on Save button; 2. Define name of the file; 3. Click Save button

Open saved file



Once the file has been saved **1. Go to Input; 2. Click Open; 3. Find and select file; 4. Click Open**