

OECD QSAR Toolbox v.3.1

Predicting skin sensitization potential of
3,4-dinitrophenol taking into account
tautomerism

Outlook

- **Background**
- Objectives
- The exercise
- Workflow

Background

- This is a step-by-step presentation designed to take the user of the Toolbox through the workflow for filling data gap for skin sensitization taking into account tautomerism of target chemical.

Outlook

- Background
- **Objectives**
- The exercise
- Workflow

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Providing tautomeric set of target chemical
- Identify analogues for a set of tautomers
- Retrieve experimental results available for those analogues
- Fill data gaps by read across

Outlook

- Background
- Objectives
- **The exercise**
- Workflow

The Exercise

- In this exercise we will predict the skin sensitization potential for **(3,4-dinitrophenol) [CAS 577-71-9]**
- Set of simulated tautomers for the target chemical will be provided
- This prediction will be accomplished by collecting a set of similar analogues for set of target and its tautomers
- The initial category will be defined by Protein binding by Oasis v1.1
- Data gap will be filled by read-across

Outlook

- Background
- Objectives
- The exercise
- **Workflow**

Workflow

- **As you know the Toolbox has 6 modules which are typically used in sequence:**
 - Chemical Input
 - Profiling
 - Endpoint
 - Category Definition
 - Data Gap Filling
 - Reporting

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - **Input**

Chemical Input

There are two ways for simulating tautomers of chemicals

- During the process of entering the structure into the system
- Simulating tautomersim of already entered structure

In order to accelerate the workflow all databases are preliminary tautomerized, calculated and profiled. The results are stored in the database.

See next screen shots

Chemical Input

Search by CAS# in tautomerized databases

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The 'CAS#' button in the toolbar is circled in red and labeled with a '1'. A search dialog box is open, showing '577-71-9' in the search field, 'Tautomeric sets' checked, and a 'Search' button labeled with a '4'. The results table shows one entry with CAS# 577-71-9, a chemical structure of 3,4-dinitrophenol, and names '1: 3,4-din' and '2: 3,4-din'. The 'OK' button is labeled with a '5'. A text box at the bottom explains that tautomeric set functionality allows searching for tautomeric forms of target chemical in previously tautomerized databases.

1. **Click** on CAS#

2. **Enter** 577-71-9

3. **Select** Tautomeric sets

4. **Click** Search

5. **OK**

Chemical Input

Target chemical identity

The screenshot shows the QSAR Toolbox interface. In the 'Documents' panel on the left, a chemical structure is selected and circled in red. A red circle with the number '1' points to this structure. A dialog box on the right displays five tautomeric forms of the selected chemical, each with its CAS# (577-71-9) and a chemical structure. A red bracket underlines these five forms, with the text 'Tautomeric forms' below it. The dialog box also includes 'Save to smi', 'Search', and 'OK' buttons.

Target with its tautomeric forms are identified and loaded on a data matrix.
 1. **Double click** over the target structure displays target and its tautomeric forms

Chemical Input

Multiplication a tautomeric set of already defined target

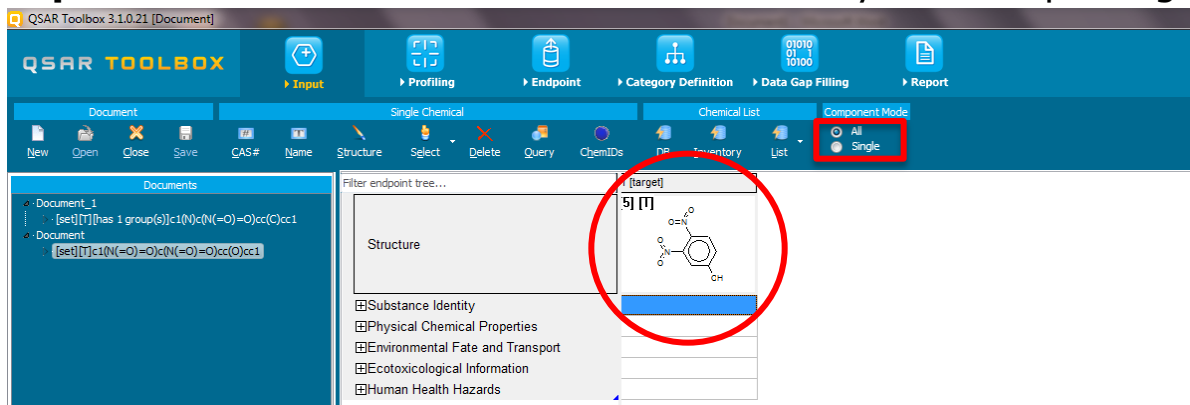
The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', and 'Category'. Below this is a toolbar with icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Select', 'Delete', 'Query', 'ChemIDs', and 'DB'. The main workspace is divided into 'Documents' and 'Single Chemical' panes. In the 'Documents' pane, a tree view shows a document named 'Document_1' containing a SMILES string: c1(N(=O)=O)c(N(=O)=O)cc(O). A right-click context menu is open over this string, with 'Multiplication' selected and 'Tautomerism' highlighted. A chemical structure of a nitrophenol is shown next to the menu. In the 'Single Chemical' pane, a filter endpoint tree shows '1 [target]'. A separate window shows the resulting tautomeric set, starting with '[set][T]c1(N(=O)=O)c(N(=O)=O)cc(O)cc1' followed by several tautomeric SMILES strings.

1. **Select** the SMILES of the target chemical perform right click on it and then
2. **Select** Multiplication-Tautomerism
3. **Generated** tautomers appear in tree like form

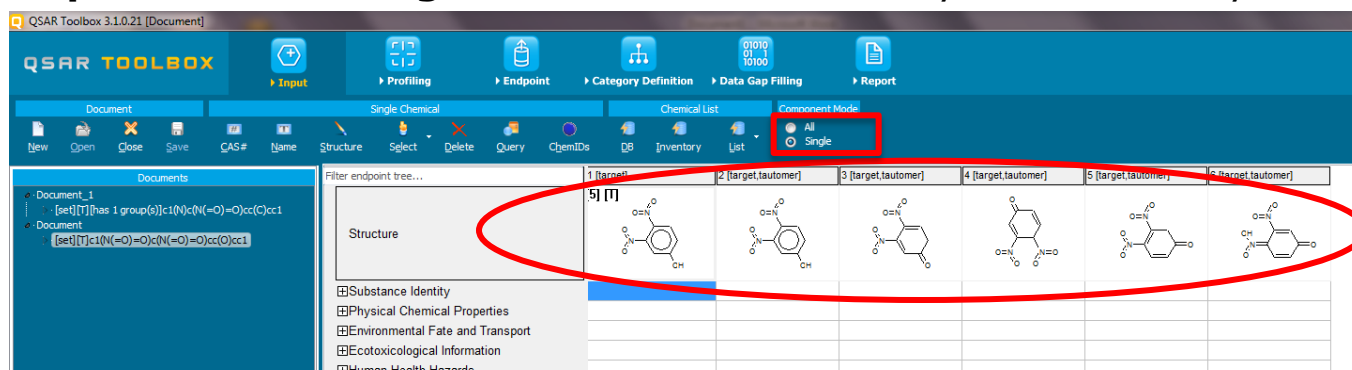
Chemical Input

Implementation of Modeling modes:

- **Component Mode All** – all tautomers are analyzed as a package



- **Component Mode Single** – each tautomer is analyzed individually



Different modes for visualization of tautomeric sets.
A package of target and its tautomeric forms are used in further read across.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - **Profiling**

Profiling Overview

- “Profiling” refers to the electronic process of retrieving relevant information on the target compound, other than environmental fate, ecotoxicity and toxicity data, which are stored in the Toolbox database.
- Available information includes likely mechanism(s) of action, as well as observed or simulated metabolites.

Profiling

Side-Bar to Profiling

- For most of the profilers, background information can be retrieved by highlighting one of the profilers (for example, Protein binding by OASIS v1.1 and clicking on “View” (see next screen shot)).

Profiling Side-Bar to Profiling

1. Highlight the profiler

2. Click View

List with categories

- Protein binding by OASIS v1.1
 - Acylation
 - Ionic
 - Michael addition
 - Nucleophilic addition
 - Radical
 - Free radical formation
 - Organic peroxy compounds
 - Schiff base formation
 - Benzoyl Schiff base formation
 - Benzoylphosphine oxides
 - Nucleophilic cycloaddition
 - Diketones
 - Pyrazolones and pyrazolidones
 - Pyrazolones and pyrazolidinones
 - Schiff base formation with carbonyl compounds
 - Aldehydes
 - alpha-ketoesters
 - SN Vinyl
 - Nucleophilic vinylic substitution on activated halogens
 - Halogenated isothiazolones
 - SN1
 - Carbenium ion formation
 - Azoxy compounds-forming carbenium ion
 - Nucleophilic substitution (SN1) on alkyl (aryl) mercury
 - Mercury compounds
 - SN2
 - Interchange reaction with sulphur containing compounds
 - Thiols and disulfide compounds
 - Nucleophilic substitution at Nitrogen atom
 - N-halogenated diketones or sulfoxides/sulfones
 - N-nitroso compounds
 - N-oxycarbonyl amides
 - Nucleophilic substitution at sp2 Carbon atom

Textual description

Mechanistic Domain: Schiff base formation
Mechanistic Alert: Schiff base formation with carbonyl compounds
Structural Alert: Aldehydes

This category includes chemicals that can undergo adduct formation with proteins via Schiff base formation with aldehydes. The possible structural alert acting by this mechanism is illustrated below:

$$R-C(=O)-H + Pr-NH_2 \rightleftharpoons R-C(=N-Pr)-H$$

R = any C, H

Aldehydes are highly reactive molecules, many of which are strong sensitizers and their direct conjugation to protein nucleophiles is thought to be responsible. Simple aldehydes react readily with the amino groups of lysine residues on proteins to form imines or Schiff bases. All aliphatic aldehydes can potentially undergo Schiff base formation with a primary amine, which is a reversible reaction (optimal at pH 3-4) and proceeds in two stages via a tetrahedral intermediate.

References:
 Camilla K. Smith, Sharon A.M. Hotchkiss, Allergic Contact Dermatitis: Chemical and Metabolic Mechanisms, 2001, Published by

Profiling Side-Bar to Profiling

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The top menu bar includes 'Input', 'Profiling', 'Endpoint', 'Category Definition', 'Data Gap Filling', and 'Report'. The 'Profiling' menu is open, showing 'Profiling methods', 'New', 'View', and 'Delete'. A red circle highlights the 'View' button, with a red arrow pointing to the 'Profiling methods' sidebar. In the sidebar, 'Protein binding by OASIS v1.1' is highlighted with a red circle and a red arrow pointing to the 'Structural boundaries' window. In this window, the 'Aldehydes' category is highlighted with a red circle. Below the 'Aldehydes' category, a chemical structure of an aldehyde is shown: CC(=O)H. The 'Structural boundaries' window also shows a logical expression: CC(=O)H AND CC(=O)H OR CC(=O)H.

1. Highlight the profiler
2. Click View

Profiling

Profiling the set of target and tautomers

- For this example, the following profilers relevant to **skin sensitization** are used(see next screenshot):
 - Protein binding by OASIS v1.1
 - Protein binding by OECD
 - Protein binding potency
 - Protein binding for skin sensitization by OASIS v1.1
- Select the “Profiling methods” related to the target endpoint by clicking on the box next to the profilers name.
- This selects (a **green** check mark appears) or deselects (**green** check disappears) profilers.

Profiling

Profiling the set of target and tautomers

The screenshot shows the QSAR Toolbox interface. The 'Profiling methods' panel on the left has several checkboxes selected, including 'Protein binding by OASIS v1.1', 'Protein binding by OECD', 'Protein binding potency', and 'Protein binding alerts for skin sensitization by OASIS v1.1'. The main table displays a filter endpoint tree on the left and a grid of results for six target/tautomer sets. A red circle highlights the 'Protein binding by OASIS v1.1' row in the table, where the first column (target) shows 'No alert found' and the other columns show 'Michael addition >>>'. A red arrow points from a text box below to this specific cell.

Filter endpoint tree...	1 [target]	2 [target,tautomer]	3 [target,tautomer]	4 [target,tautomer]	5 [target,tautomer]	6 [target,tautomer]
Structure						
Substance Identity						
Physical						
Human Health Hazards						
Profile						
General Mechanistic						
Protein binding by OASIS v1.1	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found
Protein binding by OECD	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found
Protein binding potency	Extremely reactive (...) Moderately reactive (...) Moderately reactive (...) Not possible to class...	Not possible to cla...	Extremely reactive (...) Extremely reactive (...)	Not possible to cla...	Moderately reactive... Moderately reactive...	Not possible to cla...
Endpoint Specific						
Protein binding alerts for skin sensi...	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found	Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> Michael addition >>> No alert found

The target chemical has no protein binding alert – No Skin Sensitization effect is expected

Profiling

Profiling the set of target and tautomers Profile statistic

Profile statistic

Select profiles: Report Create profiler Stop

Profiles in use: Stack mode Stacked to 100% Group by category

Chemical groups

#	Category	Count	%
1	No alert found	1	20.00
2	Michael addition+Michael addition >> Michael addition on conjugated systems with electron withdrawing group >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-carbonyl compounds with polarized double bonds+Michael addition >> Michael addition on conjugated systems with electron withdrawing group >> Nitroalkenes	3	60.00
3	Michael addition+Michael addition >> Michael addition	1	20.00

Protein binding alerts for skin sensitization by OASIS v1.1

Count

No alert found Michael addition+Mic...

2D representations of the structures in each category

Number of tautomers in a category bin

Distribution of target and its tautomers across Protein binding for skin sensitization by OASIS v.1.1

Michael addition+Michael addition >> Michael addition on conjugated systems with electron withdrawing group >> Michael addition on conjugated systems with electron withdrawing group >> alpha,beta-carbonyl compounds with polarized double bonds+Michael addition >> Michael addition on conjugated systems with electron withdrawing group >> Nitroalkenes

3

Save to smi Print... Add in new doc Add as target list

CAS# 577-71-9 CAS# 577-71-9 CAS# 577-71-9

Four tautomeric forms of the target chemical have Protein binding alerts for skin sensitization: "Michael addition/Michael addition on conjugated systems with electron withdrawing group"

Profiling

Profiling the set of target and tautomers

Profile statistic

The screenshot displays the QSAR Toolbox interface. The main window shows a chemical structure of a nitro-substituted phenol derivative. The 'Profile' section is expanded to show 'Protein binding by OECD'. A red circle highlights the 'Michael addition >> Polarisated Alkenes >> Pol' alert, which has a count of 3 and a percentage of 60.00%. Below this, a list of tautomeric forms is shown, with the first three forms circled in red. A red arrow points from this list to a text box that reads: 'Three of the tautomeric forms of the target have positive protein binding alert "Michael addition/Polarised alkenes"'. The 'Profile statistic' window shows a table with the following data:

#	Category	Count	%
1	Michael addition	4	80.00
2	Michael addition >> Polarisated Alkenes	4	80.00
3	Michael addition >> Polarisated Alkenes >> Pol	3	60.00
4	Michael addition >> Polarisated Alkenes >> Pol	3	60.00
5	No alert found	1	20.00

A bar chart titled 'Protein binding by OECD' shows the distribution of counts for these categories: Michael addition (4), Michael addition >> Polarisated Alkenes (4), Michael addition >> Polarisated Alkenes >> Pol (3), and No alert found (1).

At the bottom left, a red box contains the text: 'Distribution of target and its tautomers across Protein binding by OECD'.

Profiling

Profiling the set of target and tautomers

Profile statistic

The screenshot displays the QSAR Toolbox interface. On the left, the 'Documents' panel shows a chemical structure with the SMILES string [set][T]c1(N(=O)=O)c(N(=O)=O)cc(O)cc1. The 'Filter endpoint tree' on the right lists various endpoints, with 'Protein binding potency' selected. The 'Profile statistic' window is open, showing a table of categories and a bar chart. The table lists three categories: 'Not possible to classify according to these rules (G: 3)' with a count of 3 (60.00%), 'Extremely reactive (GSH)+Extremely reactive (GSH- 1)' with a count of 1 (20.00%), and 'Moderately reactive (GSH)+Moderately reactive (G- 1)' with a count of 1 (20.00%). The bar chart shows three bars representing these categories, with the last two bars circled in red. A red box highlights the text 'Two of the five tautomers are "Reactive" by Protein binding potency'. Another red box highlights the text 'Protein binding potency categories of tautomers in the tautomeric set'.

#	Category	Count	%
1	Not possible to classify according to these rules (G: 3)	3	60.00
2	Extremely reactive (GSH)+Extremely reactive (GSH- 1)	1	20.00
3	Moderately reactive (GSH)+Moderately reactive (G- 1)	1	20.00

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - **Endpoint**

Endpoint

- “Endpoint” refer to the electronic process of retrieving the environmental fate, ecotoxicity and toxicity data that are stored in the Toolbox database.
- Data gathering can be executed in a global fashion (i.e., collecting all data of all endpoints) or on a more narrowly defined basis (e.g., collecting data for a single or limited number of endpoints).
- In this example, we limit our data gathering to the common skin sensitization from databases containing skin sensitization data
- Data for target chemical and its tautomeric forms is extracted from selected databases if available

Endpoint

- For this example, the following database are relevant to the skin sensitization(see next screen shot):
 - Skin sensitization
 - Skin sensitization ECETOC

Endpoint Gather data

QSAR Toolbox 3.1.0.21 [Document]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Data Import Export Delete Tautomerize

Gather Import IUCLID5 IUCLID5 Database Inventory Database

Databases

Select All Unselect All Invert About

Human Health Hazards

- Bacterial mutagenicity ISSSTY
- Carcinogenic Potency Database (CPDB)
- Carcinogenicity&mutagenicity ISSCAN
- Cell Transformation Assay ISSCTA
- Dendritic cells COLIPA
- Developmental toxicity ILSI
- ECHA CHEM
- Estrogen Receptor Binding Affinity OASIS
- Eye Irritation ECETOC
- Genotoxicity OASIS
- Keratinocyte gene expression Givaudan
- Micronucleus ISSMIC
- Micronucleus Oasis
- MUNRO non-cancer EFSA
- Rep Dose Tox Fraunhofer ITEM
- Repeated Dose Toxicity HESS
- Rodent Inhalation Toxicity Database
- Skin irritation
- Skin sensitization
- Skin sensitization ECETOC
- Terrestrial US-EPA ECOTOX
- Toxicity Japan MHLW
- ToxRefDB US-EPA
- Yeast estrogen assay database University of Terr

Inventories

Select All Unselect All Invert About

- AICS
- Canada DSL
- COSIVE

Filter endpoint tree...

1 [target]

5] [1]

Structure

O=[N+]([O-])c1ccc(O)c([N+](=O)[O-])c1

- Substance Identity
- Physical Chemical Properties
- Human Health Hazards
- Profile

1. **Expand** the Human Health Hazards section
2. **Select** databases related to the target endpoint by adding a **green** check in the box before the database name.
3. **Click** Gather

Endpoint Gather data

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. The top menu bar includes 'Data', 'Import', 'Export', 'Delete', and 'Tautomerize'. The 'Endpoint' workflow is selected, showing a tree view of experimental data. The 'Human Health Hazards' category is expanded, and 'Skin sensitization' is checked. The 'Inventories' section shows various databases, with 'Skin sensitization' and 'Skin sensitization ECETOC' highlighted. The 'Filter endpoint tree...' window shows a chemical structure and a list of endpoints, with 'GPMT' and 'LLNA' circled in red. The text '(1/1) M: Positive' is visible next to the circled endpoints.

Positive experimental data appears on datamatrix

Recap

- You have entered the target chemical
- You have simulate the tautomeric forms of target chemical
- You have profiled the tautomeric set of the target and identified no protein binding alert for the target. However, four tautomers have positive protein binding alerts
- You have gather data for chemical and its tautomeric forms and found positive experimental data for target.
- It is needed to verify the experimental data by searching for analogs having same functionalities
- Now you are ready to continue with next step of the workflow "Category definition".

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - **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 refining 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”.
- Detailed information about grouping chemical (Chapter 4) could be found in document “Manual for Getting started” published on OECD website:

<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>

Basic guidance for category formation and assessment

Suitable categorization phases:

1. Structure-related profilers
2. Endpoint specific profilers (for sub-cat)
3. Additional structure-related profilers, if needed to eliminate dissimilar chemicals (to increase the consistency of category) (e.g. chemical elements)

Performing categorization:

1. The categorization phases should be applied successively
2. The application order of the phases depend on the specificity of the data gap filling
3. More categories of same Phase could be used in forming categories
4. Some of the phases could be skipped if consistency of category members is reached

Graphical illustration of suitable categorization phases is shown on next slide

Suitable Categorization/Assessment Phases

Phase I. Structure based

- US EPA Categorization
- OECD Categorization
- Organic functional group
- Structural similarity
- ECOSAR

**Broad grouping
Endpoint Non-specific**

Repeating Phase I due to Multifunctionality of chemicals

Phase II. Mechanism based

- DNA binding mechanism
- Protein binding mechanism
- Genotoxicity/carcinogenicity
- Cramer rules
- Verhaar rule
- Skin/eye irritation corrosion rules

**Subcategorization
Endpoint Specific**

Metabolism accounted for

Phase III. Eliminating dissimilar chemicals

**Apply Phase I – for structural dissimilarity
Filter by test conditions – for Biological dissimilarity**

**Subcategorization
Endpoint Specific**

Category Definition

Grouping methods

- For this example, the specific endpoint classification of target and its tautomers is identified by Protein binding profilers. Consistency of the category member is reached and phase I could be skipped (point 4 from performing categorization, slide #33).
- For this example initial group of analogues presented as tautomeric sets is identified by Protein binding by OASIS profiler
- Software search analogues presented as tautomeric sets having same protein binding distribution as those of the target tautomeric set

Category definition is a tool for grouping chemicals. For more details see tutorials posted on LMC and OECD website:

<http://www.oecd.org/env/ehs/risk-assessment/theoecdqsartoolbox.htm>

<http://superhosting.oasis-lmc.org/products/software/toolbox/toolbox-support.aspx>

Category Definition

Defining Protein binding by OASIS v1.1

1. **Highlight** the "Protein binding by OASIS v1.1"

2. **Click** Define

3. **Click** OK to confirm the defined categories for the tautomeric set

Category Definition

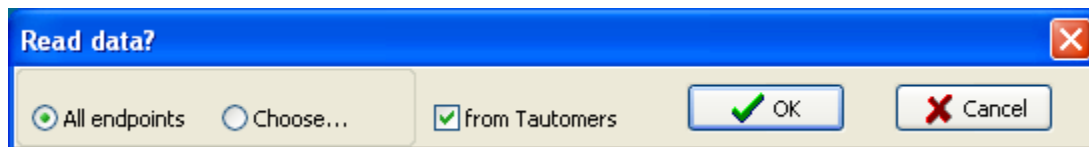
Defining Protein binding by OASIS v1.1

The software identify four chemical (presented as tautomeric set) having same protein binding alerts as the target set
1. Select OK

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, since only databases that contain information for human health hazards endpoints are selected, both options give same results.
- As the Toolbox must search the database, this may take some time.

Category Definition

Read data for Analogues

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. The main window is titled 'Category Definition' and shows a tree view of endpoints on the left and a data table on the right. The table has columns for 'Structure' and four numbered columns (1-4). A dialog box titled 'Read data?' is open, with 'All endpoints' selected and 'from Tautomers' checked. A blue callout box with the number '1' points to the 'OK' button. A red box at the bottom contains the text 'Read data for tautomeric sets'.

1. Click OK

Read data for tautomeric sets

Category Definition

Summary information for analogues

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. The 'Category Definition' workflow is active, showing a list of databases on the left and a summary table of results on the right. The 'Human Health Hazards' database is selected, and the 'EC3' endpoint is chosen. The summary table shows that all four analogues have positive EC3 data.

Structure	1 [target]	2	3	4
	[5] [7]	[1] [7]	[1] [7]	[1] [7]
	(4/4) M: Positive	M: Positive	M: Positive	M: Positive

All analogues have positive EC3 data

Recap

- You have identified a category of analogues presented as tautomeric sets having same distribution of protein binding alerts as the target tautomeric set
- The available experimental results for these 4 analogues have been collected from the selected databases (Skin sensitization and Skin sensitization ECETOC)
- But before the user can proceed with the “Filling Data Gap” module, he/she should navigate through the endpoint tree and find the specific gap that will be filled

Category Definition

Navigation through the endpoint tree

- The user can navigate through the data tree by closing or opening the nodes of the endpoint tree.
- The data tree is extensive but logically constructed; it can be mastered with a practice.
- In this example, the “EC3” is the target endpoint.
- You can navigate through the endpoint tree: Double-click on the node next to **Human Health Hazards** then effect **Sensitisation**, followed by **Skin**, type of method **In Vivo** and assay **LLNA** and finally **EC3** (see next screen shots)

Category Definition

Navigation through the endpoint tree

QSAR Toolbox 3.1.0.21 [Document]

QSAR TOOLBOX

Input Profiling Endpoint **Category Definition** Data Gap Filling Report

Categorize Delete

Define Subcategorize Combine Clustering Delete Delete All

Grouping methods

EC3

1

Endpoint Specific

- Ultimate biodeg
- Acute aquatic toxicity classification by Verhaar
- Acute aquatic toxicity classification by Verhaar (Modified)
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Bioaccumulation – metabolism alerts
- Bioaccumulation – metabolism half-lives
- Biodegradation fragments (BioWIN MITT)
- Carcinogenicity (genotox and nongenotox) alerts by ISS
- DNA alerts for AMES, MN and CA by OASIS v. 1.1
- Eye irritation/corrosion Exclusion rules by BfR
- Eye irritation/corrosion Inclusion rules by BfR
- in vitro mutagenicity (Ames test) alerts by ISS
- in vivo mutagenicity (Micronucleus) alerts by ISS
- Keratinocyte gene expression
- Oncologic Primary Classification
- Protein binding alerts for skin sensitization by OASIS v.1.1
- rHER Expert System ver. 1 - USEPA
- Skin irritation/corrosion Exclusion rules by BfR
- Skin irritation/corrosion Inclusion rules by BfR

Empiric

- Chemical elements
- Groups of elements
- Lipinski Rule Oasis
- Organic functional groups
- Organic functional groups (nested)
- Organic functional groups (US EPA)
- Organic functional groups, Norbert Haider (checkmol)
- Structure similarity
- Tautomers unstable

Toxicological

Structure

Substance Identity

Human Health Hazards

Sensitisation

Skin

In Vitro

In Vivo

LLNA

EC3

Profile

1 [target]	2	3	4
[5] [0]	[11] [0]	[11] [0]	[11] [0]
(4/4) M: Positive	M: Positive	M: Positive	M: Positive

1. Type "EC3" in the filter box, then press **Enter**
2. **Open** the tree to the target endpoint by single left click on box

Recap

- You have now retrieved the available skin sensitisation data for the four analogues represented by their tautomeric forms.
- You have identified the target endpoint of “Sensitization /Skin/In vivo/LLNA/EC3”.
- You are ready to fill in the data gap, so click on “Data Gap Filling” (see next screen shot).

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Category definition
 - **Data Gap Filling**

Data Gap Filling

Apply Read across analysis

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The 'Data Gap Filling' menu is active. In the left sidebar, 'Read-across' is selected under 'Data Gap Filling Method'. The main table displays seven chemical structures in columns 1 through 7. A filter endpoint tree on the left shows 'EC3' selected under 'Sensitisation'. A callout '1' points to the 'EC3' data point in the table, which is circled in red. Callout '2' points to the 'Read-across' method in the sidebar, and callout '3' points to the 'Apply' button in the top left.

Filter endpoint tree...	1 [target]	2	3	4	5	6	7
Structure	<chem>O=[N+]([O-])c1ccc(O)cc1</chem>	<chem>Nc1ccc(O)cc1[N+](=O)[O-]</chem>	<chem>O=C(O)c1ccc(O)cc1[N+](=O)[O-]</chem>	<chem>Oc1ccc(Cl)cc1[N+](=O)[O-]</chem>	<chem>Nc1ccc(O)cc1[N+](=O)[O-]</chem>	<chem>Nc1ccc(O)cc1[N+](=O)[O-]</chem>	<chem>Nc1ccc(O)cc1[N+](=O)[O-]</chem>
Substance Identity							
Physical Chemical Properties							
Environmental Fate and Transport							
Ecotoxicological Information							
Human Health Hazards							
Acute Toxicity							
Carcinogenicity							
Developmental Toxicity / Teratogenicity							
Genetic Toxicity							
Immunotoxicity							
Irritation / Corrosion							
Neurotoxicity							
Repeated Dose Toxicity							
Sensitisation							
Skin							
In Chemico							
In Vitro							
In Vivo							
GPMT							
LLNA							
EC3	(1/1) M: Positive	M: Positive	M: Negative	M: Positive		M: Positive	M: Positive

- 1. Highlight** the data endpoint box corresponding to "EC3" under the target chemical.
- 2. Select** Read-across
- 3. Click** Apply

Data Gap Filling Result of Read-across

The screenshot shows the 'Data Gap Filling' workflow in the OECD QSAR Toolbox. The 'Subcategorization' panel on the left is open, with 'Protein binding by OASIS v1.1' selected. The main window displays a table of chemical structures and their predicted values for 'Protein binding'. A graph below the table shows the distribution of log Kow values for the target set and the predicted set. A yellow callout box highlights the text: 'The tautomeric sets of analogues have same distribution of Protein binding alerts as the target set'. A red callout box '1' points to the 'Select/filter data' option in the 'Return to matrix' panel on the right.

1 [target]	2	3	4
[5] [1]	[1] [1]	[1] [1]	[1] [1]
<chem>O=[N+]([O-])c1ccc(O)cc1</chem>	<chem>O=[N+]([O-])c1ccc(O)cc1</chem>	<chem>Nc1ccc(O)cc1</chem>	<chem>Nc1ccc(O)cc1</chem>
(4/4) M. Positive	M. Positive	M. Positive	M. Positive

Read across prediction of EC3, taking the highest mode from the nearest 5 neighbours, based on 3 values from 3 neighbour chemicals, Observed target value: 'Positive', Predicted target value: 'Positive'

The tautomeric sets of analogues have same distribution of Protein binding alerts as the target set

1. **Select** Select/Filter data and then **Subcategorization**
2. **Select** Protein binding by OASIS v1.1

Data Gap Filling

Interpreting Read-across

- In this example, all results of the analogues are consistent; all the analogues are Skin sensitizers.
- The tautomeric sets of analogues have same distribution of Protein binding alerts as the target set.
- The same positive sensitising potential is therefore predicted for the target chemical.
- Accept the prediction by **clicking** “Accept prediction” (see next screen shot).

Data Gap Filling Result of Read-across

QSAR Toolbox 3.1.0.21 [Document]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Filing Apply

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 Sensitisation Skin In Vivo LLNA EC3

EC3	1 [target]	2	3	4
Structure	<chem>O=[N+]([O-])c1ccc(O)cc1</chem>	<chem>Oc1ccc(N)cc1</chem>	<chem>O=[N+]([O-])c1ccc(N)cc1</chem>	<chem>Oc1ccc(N)cc1</chem>
EC3	(4/4) M: Positive	M: Positive	M: Positive	M: Positive

Descriptors Prediction

Read across prediction of EC3,
taking the highest mode from the nearest 5 neighbours, based on 3 values from 3 neighbour chemicals,
Observed target value: 'Positive', Predicted target value: 'Positive'

Information

The current prediction was accepted

OK

Accept prediction
Return to matrix

Select/filter data
Selection naviga
Gap filling appro
Descriptors/dat
Model/(Q)SAR
Calculation options
Visual options
Information
Miscellaneous

1. Select Accept prediction
2. Click OK and then 3. Click Return to matrix

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Input
 - Profiling
 - Endpoint
 - Category definition
 - Data Gap Filling
 - **Report**

Report

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The top navigation bar contains icons for Input, Profiling, Endpoint, Category Definition, Data Gap Filling (selected), and Report. The left sidebar has a 'Filling' section with an 'Apply' button. The main area is divided into three panes: a 'Data Gap Filling Method' sidebar, a 'Filter endpoint tree...' pane, and a main table of predictions.

The 'Filter endpoint tree...' pane shows a tree view of endpoints. Under 'Human Health Hazards', 'Sensitisation', and 'Skin', 'LLNA' is selected. A callout '1' points to this selection.

The main table has columns for 'Structure', '1 [target]', '2', '3', and '4'. The 'Structure' column shows chemical structures. The '1 [target]' column contains a prediction with a structure and the text '(4/5) R: Positive'. A context menu is open over this prediction, with a callout '2' pointing to the 'Report' option.

Structure	1 [target]	2	3	4
<chem>O=C1C=CC(=C1)C(=O)N</chem>	[11] [1]	[11] [1]	[11] [1]	[11] [1]
<chem>Nc1ccc(O)cc1</chem>				
<chem>Nc1ccc(N)cc1</chem>				
<chem>Nc1ccc(O)cc1</chem>				
<chem>Nc1ccc(O)cc1</chem>				
			M: Positive	M: Positive

- 1. Select prediction
- 2. Right Click and **Select** Report

Report

The screenshot shows the QSAR Toolbox 3.1.0.21 interface. The main window displays a report titled "QSAR Toolbox prediction based on read-across" for the prediction of EC3 for 3,4-dinitrophenol. A callout box with the number "1" points to the "Summary" section of the report.

Summary

Toxicity of the target chemical (Positive) is predicted from category members using read-across based on 3 values (Positive x3) from 3 nearest neighbours compared by prediction descriptors. Category members are set of tautomers and are selected based on the profile of the target chemical and/or its tautomers. Only chemicals having experimental data are listed in the category.

The target set of tautomers FALLS within applicability domain of the prediction (see Section 4.3 for details).

The descriptor values for the target chemical and the category members in case they are set of tautomers, set of metabolites or mixtures are calculated using the following rule(s):

1. "log Kow" - taking the weighted average value

The endpoint data is selected from the following database(s):

1. Skin sensitization
2. Skin sensitization ECETOC


Below is a summary table for endpoint & descriptor values for the target set of tautomers and the category members.
Experimental values from data matrix are presented in bold font. Recalculated endpoint values (if required by selected data usage option in Gap Filling) are presented in italic font. Recalculated endpoint values based on experimental data only are presented in bold and italic font.

	<i>Endpoint(s)</i>	<i>Descriptor(s)</i>
	Sensitisation	log Kow
	Skin sensitisation II (ECETOC)	-
<i>Target chemical & its</i>	<i>Positive</i>	<i>0.144</i>

1. Summary information for the prediction of tautomeric set

Report

The screenshot displays the QSAR Toolbox 3.1.0.21 interface. The main window shows a prediction report for 3,4-dinitrophenol. The report content is as follows:

1 Prediction of EC3 for 3,4-dinitrophenol 8 / 50
i. Predicted value (model result):
 Positive
j. Predicted value (comments): *manually editable field*
 Not provided by the user
4.3. Applicability domain (OECD Principle 3): **2**
 The target set of tautomers FALLS within applicability domain
 (see Section 3.1.b for detailed description of the domain)

4.4. Uncertainty of the prediction (OECD Principle 4): *manually editable field*
 Not provided by the user

At the bottom of the screenshot, two blue callout boxes are present:

- 1. Predicted value**
- 2. Applicability domain**

Report

The screenshot shows the QSAR Toolbox software interface. The main window displays a report titled "QSAR Toolbox prediction based on read-across" for "Prediction of EC3 for 3,4-dinitrophenol". The report includes an "APPENDIX 7 - Chemical components" section. A red oval highlights the text "Tautomer No.1 of target chemical and its tautomers:", and a green box highlights the text "used in prediction". Below this, a list of four items is shown: 1. CAS number: 577-71-9; 2. Other regulatory numbers: Not reported; 3. Chemical name(s): Not available; 4. Structural formula: (with a partial chemical structure shown). The left sidebar shows available data to report (Predictions, QSARs, Categories) and available report templates (Standard and Custom).

Additional Appendix 7 list tautomers of target and analogue chemicals used in read-across
 Also an information about which tautomer is used in the RA prediction is provided.