

Givaudan[®]

ENGAGING
THE
SENSES



Client Databases – Importation, Issues and Performance
ECHA - OECD [Q]SAR Toolbox Workshop
Geoffrey Hynes – November 2011 - Helsinki

Presentation - Overview

OECD – (Q)SAR Toolbox Brief Overview

OECD (Q)SAR Toolbox Database Importation – Version 1

OECD (Q)SAR Toolbox Database Importation – Version 2

Improving Read-Across Analysis Performance

Further Potential Improvements

Summary

OECD - (Q)SAR Toolbox Brief Overview

OECD (Q)SAR Toolbox:

Released in March 2008 (v.1.0).

Updated October 2010 (v.2.0).

Updated February 2011 (v.2.1).

Toolbox used by: Governments, the chemical industry, other stakeholders

Filling gaps in (eco)toxicity data needed for assessing the hazards of chemicals.

Logical process which allows the grouping chemicals into chemical categories.

Identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical.

Identification of other chemicals that have the same structural characteristics and/or mechanism or mode of action.

Use of existing experimental data to fill the data gap(s).

OECD - (Q)SAR Toolbox Brief Overview

Predominant chemicals within the databases:

Pharmaceutical

Agro chemicals

Published studies

Very limited number of fragrance chemicals with experimental data within the Toolbox to allow read-across or trend analysis.

Many of the read-across and trend analyses results are outside of the domain making them unreliable.

Therefore a propriety database is required.

Database Importation - Fragrance Chemical

The next slide highlights the results from a typical fragrance chemical when run through the OECD Toolbox for read-across analysis.

Database Importation - Fragrance Chemical

Endpoint

Gather data

QSAR Toolbox 2.2.1.1120 [Document_1]

QSAR TOOLBOX

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

Data Import Export Delete

Gather Import Export IUCLIDS Database Inventory

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

1 (Target)

Structure

Filter endpoint tree...

Substance Identity

Physical Chemical Properties

Environmental Fate and Transport

Ecotoxicological Information

Human Health Hazards

Profile

Database Affiliation

Inventory Affiliation

OECD HPV Chemical Categories

Substance Type

US-EPA New Chemical Categories

DNA binding by OASIS

DNA binding by OECD

Estrogen Receptor Binding

Protein binding by OASIS

Protein binding by OECD

Protein Binding Potency

Superfragments

Toxic hazard classification by Cramer (original)

Toxic hazard classification by Cramer (with extension)

Acute aquatic toxicity classification by Verhaar

Acute aquatic toxicity MOA by OASIS

Aquatic toxicity classification by ECOSAR

Bioaccumulation - metabolism alerts

Bioaccumulation - metabolism half-lives

No binding

No Binding

Non binder, without...

No binding

No binding

Not possible to cla...

No superfragment

Low (Class I)

Low (Class I)

Class 1 (narcosis o...

Reactive unspecified

Neutral Organics

Aliphatic alcohol [-...

Methyl [-CH3]

-C=CH [alkenyl hy...

Carbon with 4 singl...

-CH- [linear]

-CH2- [cyclic]

-CH- [cyclic]

Moderate

Aliphatic alcohol f...

QSAR Toolbo... X

No data found.

Choose... from Tautomers

OK Cancel

OK

Do not show this message

OK

1 | Document_1 1/0/0

OECD (Q)SAR Toolbox

Database Importation

Version 1

Database Importation

Simple importation method.

Data were displayed using a predefined endpoints tree.

The data were also required to be associated with a **leaf** node of that endpoints tree.

Fixed system

Data points not specified did not fit neatly in the leaves of the trees.

In these instances the following would be returned in the Toolbox.

Example 1:

Human health hazards#Genetic Toxicity#in vitro#[Undefined Test type](#)#Gene mutation

Database Importation

Example 2:

For the bacterial reverse mutation assay (Ames test) there was no provision for the following *Escherichia coli* strains.

E.coli WP2

E.coli WP2uvrA

E.coli WP2(pKM101)

E.coli WP2uvrA(pKM101)

Potentially limiting for read-across:

In addition, the WP2 series of strains are specifically requested by the Japanese MHW guidelines.

OECD (Q)SAR Toolbox

Database Importation

Version 2

Database Importation

Example 3:

New importation method to give more flexibility:

Predefined part (Region)

Dynamic part (Metadata)

However, this meant that for those people already using propriety databases, the following changes needed to be made.

Import Tree Path - Version 1.1

[Toxicoloical Information#Genetic Toxicology \(mutation and chromosome aberrations\)#In Vitro#Ames Mutagenicity#Ames Mutagenicity without S9#rat#AMES Salmonella Typhimurium TA 98](#)

In Vitro – in vitro

Import Tree Path - Version 2.1

[Human health hazards#Genetic Toxicity#in vitro#bacterial reverse mutation assay \(e.g. Ames test\)#Gene mutation#Salmonella typhimurium#without S9#TA 98](#)

Database Importation

Reason for Changes:

Consistency with the OECD Harmonised Templates [www.oecd.org/ehs/templates].

New harmonised templates allow the smooth electronic exchange of test/summary data information between governments and industry and between databases (IUCLID5).

It was therefore logical to implement them in the Toolbox.

Not only the labels were changed. The whole database structure was changed to comply with the OECD Harmonised Templates.

High Level Changes:

“Physical Chemical Properties” unchanged.

“Environmental Fate” to “[Environmental Fate and Transport](#)”.

“Ecotoxicological Information” unchanged.

“Toxicological Information” to “[Human Health Hazards](#)”.

Database Importation

Example 4:

For eye irritation/corrosion it seemed logical to follow the same tree path as set out for Ecotoxicology. However, the following tree path was not recognised:

irritation/corrosion#Eye#Corrosion#LD50#168h

Post discussions on the Toolbox Forum and with LMC gave the correct tree path:

irritation/corrosion#Eye#Corrosion

Database Importation

Example 5: Initial considerations

Initially it seemed relatively easy to import a dataset.

However, this was only for short path descriptions such as:

Environmental Fate and Transport#Biodegradation#Biodegradation in water:
screening tests#% Degradation

It seemed to be difficult to consistently import longer path descriptions such as:

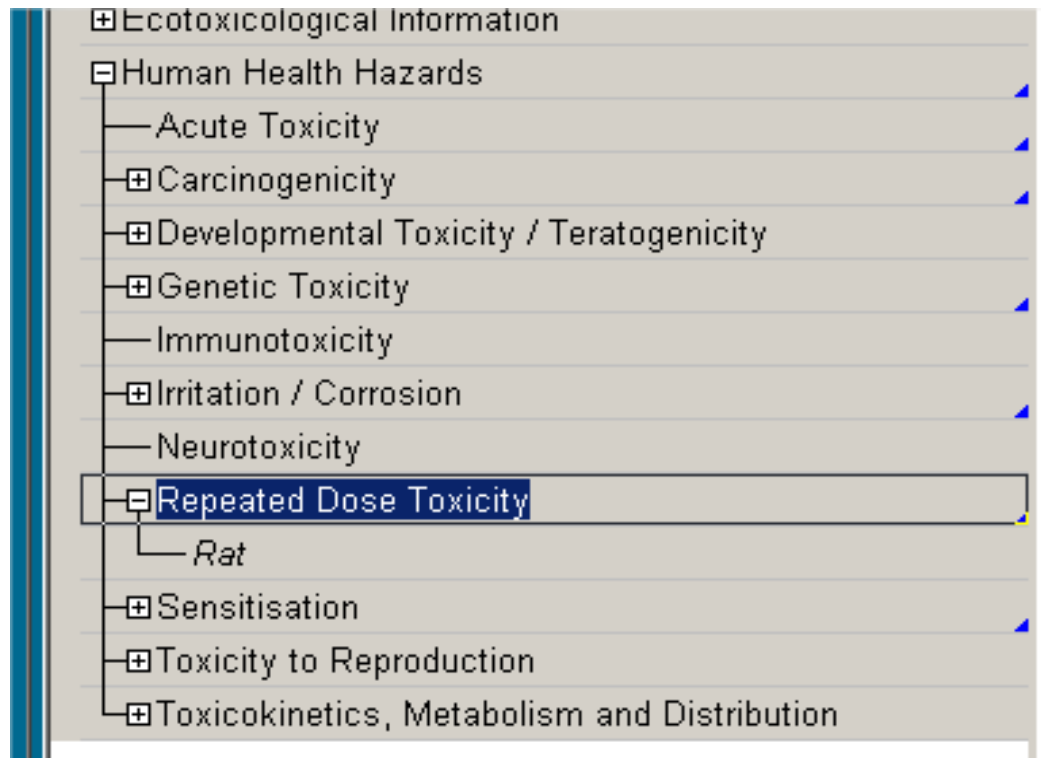
Ecotoxicological#Information#Aquatic Toxicity#Mortality#EC50#48h#Animalia#
Arthropoda(Invertebrates)#Branchiopoda(branchiopods)#Daphnia magna“

It has been commented in the forum that this appeared to be down to the metadata tags as it was not obviously clear which tags referred to which points on the endpoint path.

Database Importation - Metadata

Predefined – Metadata

No template, how do you find the correct path for example “Repeated Dose Toxicity” when there is no Metadata present?



Database Importation - Metadata

Search by using a known chemical such as “sucrose” or “benzene” to obtain the correct Metadata and there are 139 and 22649 respective chemical returns.

Human Health Hazards	
Acute Toxicity	
Carcinogenicity	(1/1) M: Negative
Developmental Toxicity / Teratogenicity	
Genetic Toxicity	(1/14) M: Negative, Positiv...
Immunotoxicity	
Irritation / Corrosion	
Neurotoxicity	
Repeated Dose Toxicity	
<i>Mus musculus</i>	
Undefined Route	
Undefined Organ	
Growth	
NOEL	(1/3) M: 1E4 mg/kg/d, 1...
Mortality	(1/3) M: 1E4 mg/kg/d, 1...
NOEL	
Reproduction	(1/2) M: 1E4 mg/kg/d, 1...
NOEL	
Rat	
Sensitisation	

However, for “Repeated Dose Toxicity, there are still “Undefined Route” and “Undefined Organ” information within the OECD Toolbox. [Future Read-Across Issues?](#)

Database Importation - Metadata

In addition:

Not all species/strains are consistent for the Repeated Dose Toxicity endpoint.

You cannot efficiently read-across for say all **rat** or all **mouse** data as they are on the same level.

Would this be a useful change to this endpoint?

Common Name Tag

Strain Tag

Mouse	
Drinking Water	(2/15)
Feed	(38/394)
Gavage	(15/231)
Inhalation	(10/167)
Mus musculus	(57/771)
Mustela sp.	
Undefined Route	(1/4)
Myotis lucifugus	
Undefined Route	(1/1)
Onychomys leucogaster	
Undefined Route	
Undefined Organ	(1/3)
Onyctolagus cuniculus	
Undefined Route	(1/3)
Ovis aries	
Undefined Route	(3/7)
Peromyscus maniculatus	(1/1)
Rat	(128/18251)
Rattus norvegicus	
Undefined Route	(59/1847)
Rattus rattus	(3/13)
Reithrodontomys megalotis	(1/1)
Sus scrofa	(4/58)
Sus sp.	(1/2)
Tadarida brasiliensis	(1/2)
Thomomys talpoides	(1/8)
Vulpes vulpes	(1/3)
Sensitisation	(137/176)

Database Importation - Excel Database

Excel database size

For version 1 of the OECD Toolbox, the excel column width for all the “Exported” data was stored in approximately 52 columns (A – BA).

However, in version 2 of the OECD Toolbox, it is not uncommon to have data stretching over 162 columns (A – FG).

Therefore, when you create your own proprietary database, it seems prudent to limit the amount of data you enter as it can become extremely time consuming.

Currently the Givaudan Central Database contains 35 columns worth of relevant data which can be interrogated during the read-across process.

This was decided by reviewing the following default metadata available in the QSAR Toolbox – User Manual – Database Import Wizard
Version 1 released April 2011
Table 1

Database Importation - Metadata Fields

Default label of metadata field	Examples of metadata field
Endpoint	LC 50, EC10, EC 50, LOEL, NOEL, Skin sensitisation, Carcinogenicity, Ames, Chromosomal aberration, Estrogen receptor binding....
Duration	years, months, days, hours, minutes, seconds...
Test organisms (species)	Daphnia magna, Lepomis symmetricus, Oncorhynchus mykiss, Poecilia reticulata, Tetrahymena pyriformis....
Effect	Immobilization, Mortality, Reproduction....
Effect type	Maternal toxicity, Developmental toxicity, Fetotoxicity, Embryotoxicity
Metabolic activation	with S9, without S9, no S9 info, with and without
Sexual maturation (offspring)	Male, Female, Male/Female...
Strain	TA 98, TA 100, TA 104, New Zealand White, Swiss, Fischer 344/DuCrj
Test type	bacterial reverse mutation assay (e.g. Ames test), in vitro mammalian cell micronucleus test, bacterial gene mutation assay, acute, subacute, chronic...., developmental, static, semi-static, flow-through
Type of genotoxicity	Gene mutation, Chromosomal aberration, DNA damage and/or repair, genome mutation
Type of method	in vivo, in vitro, other
Organ	Lung, Liver
Route	oral, inhalation, dermal, implantation, intramuscular, intraperitoneal

OECD (Q)SAR Toolbox

Improving Performance

Improving Performance - Example 1

Issues ???

Fragrance substances are defined as discrete chemicals and are not defined within the OECD or US-EPA categories

Category	Value
OECD HPV Chemical Categories	(N/A)
Substance Type	Discrete chemical
US-EPA New Chemical Categories	(N/A)
Estrogen Receptor Binding	Non binder, without...
Protein binding by OASIS	No binding
Protein binding by OECD	No binding

Improving Performance - Example 1

Example: Skin Sensitization

The screenshot shows the QSAR Toolbox 2.2.1.1120 interface. A dialog box titled "Define category name" is open, displaying "Category name (46654 chemicals)" and "Discrete chemical (Substance Type)". The background shows a tree view of chemical categories, with "Skin Sensitisation" selected under "Human Health Hazards".

Issues ???

1) 46654 substances 15 mins!

2) Need to refine search

3) Change database search

Improving Performance - Example 1

Issues ???

- 1) 46654 substances 15 mins!
- 2) Need to refine search
- 3) Change database search

Use only relevant databases under the Endpoint Tab to improve profiling speed

- 4) 1222 substances 3 mins

Improving Performance - Example 2

The screenshot shows the QSAR Toolbox 2.2.1.1120 interface. The 'Inventories' panel is open, displaying a list of inventory categories with checkboxes. All checkboxes are unchecked. The categories listed are: AICS, Canada DSL, COSING, DSSTox, ECHA PR, EINECS, HPVC OECD, MITI Japan, REACH_ECB, TSCA, and US HPV Challenge Program. An orange arrow points from the 'Inventories' panel to the text annotation on the right.

Inventories:
Unchecked

Inventories
The inventories contain only CAS, Name and SMILES and will slow the toolbox down dramatically .

OECD (Q)SAR Toolbox

Interfacing

Version 2

Interfacing - Version 2

Interface with other software:

Due to the consistency with the OECD Harmonised Templates the following additional interactions are/will be available

Current Interactions:

The Toolbox will allow the import/export with IUCLID. Although IUCLID 5.3 is not yet supported but will be added in version 3.0 of the Toolbox.

Reports generated comply with Agency formats

- QSAR Model Report Format (QMRF)
- QSAR Prediction Report Format (QPRF)

Future Interactions:

Data exchange with eChemPortal.

OECD (Q)SAR Toolbox

Further Potential Improvements

Further Potential Improvements

- Automation : Run series of structures
: Based on either CAS, SMILES, name
- Ability to save predictions : New database
: Automatically saved
- Databases : Centralised company databases

OECD (Q)SAR Toolbox

Summary

Summary

Version 2 has further improved the usability of the OECD [Q]SAR Toolbox.

For certain industries, the use of a propriety database is essential to be within the read-across domain.

The importation of propriety databases is relatively easy, however the metadata tags for the endpoint tree path can be difficult to set up initially.

There are still some issues with undefined routes and organs within the databases.

Selection of the correct databases can greatly increase the speed of analysis when completing read-across.

The potential improvements and additional interactions are exciting for version 3 of the OECD [Q]SAR Toolbox.

Presentation - Overview

Thank you for listening.

Any Questions?

Givaudan[®]

ENGAGING
THE
SENSES