

ToxAlerts tutorial

Goal:

During the tutorial, we will learn how to screen chemical libraries against toxicological alerts, to identify potentially hazardous compounds and to interpret the results. For accessing this handout and other related materials, please visit <http://docs.eadmet.com/display/EWS2M/>

Motivation:

- Do you have a practically interesting set of molecules? Screen them with a few clicks, analyze the results and publish your screening results in a scientific article.
- Have a set of your own alerts? Share them with the community and get cited.

Datasets:

For speed and convenience, we will not use any external SD-files. We have already uploaded the necessary datasets into OCHEM.

Tutorial 1 – Basic screening for several endpoints

In this tutorial we will learn how to browse alerts, create alert sets, screen chemical libraries against the desired alerts and analyze the results.

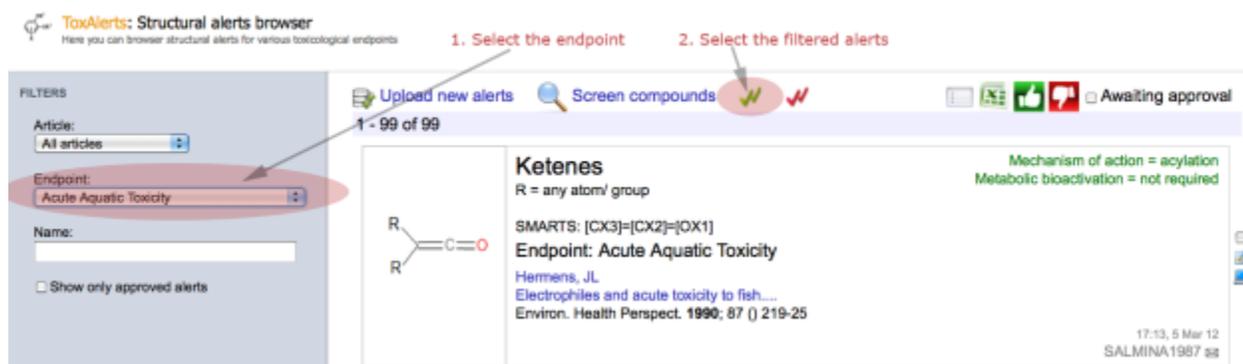
Datasets.

We will use the "High performance volume" dataset. The whole dataset has been split into "HPV low mol. Weight" and "HPV high mol. weight" according to the molecular weight of the structures. Please, use any of the datasets.

1. Choosing the desired alerts

We would like to screen our molecules against five endpoints: *acute aquatic toxicity*, *skin sensitization*, *non-genotoxic carcinogenicity*, *genotoxic carcinogenicity* and *idiosyncratic toxicity*.

- To select the relevant alerts, go to the alerts browser (Menu item: Database ->ToxAlerts -> View available alerts)
- Specify the filter for endpoint "acute aquatic toxicity" (**99 alerts** should be filtered)
- Save these alerts to your set by clicking "select all" icon: 



The screenshot shows the ToxAlerts Structural alerts browser interface. The sidebar on the left has a 'FILTERS' section with 'Endpoint' set to 'Acute Aquatic Toxicity'. The main content area shows a list of alerts, with the first one highlighted: 'Ketenes'. The details for this alert include the chemical structure R-C(=O)-R, the SMILES string [CX3]=[CX2]=[OX1], the endpoint 'Acute Aquatic Toxicity', and the reference 'Hermens, J.L. Electrophiles and acute toxicity to fish... Environ. Health Perspect. 1990; 87 () 219-25'. The top navigation bar has a 'Screen compounds' button with a green checkmark icon next to it, indicating that the alerts have been successfully selected.

- Repeat the same steps the other endpoints: *skin sensitization* (161 alerts), *non-genotoxic carcinogenicity* (5 alerts), *genotoxic carcinogenicity* (117 alerts) and *idiosyncratic toxicity* (35 alerts).
- In the end, you should have selected **417 alerts**.

Now, we are ready to screen molecules against the selected alerts!

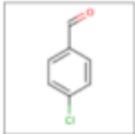
2. Screening molecules against alerts

- Open the screening dialog (Menu item: *Database – ToxAlerts - Screen compounds against alerts*)
- Select the desired set (e.g., molecules by tag, the "HPV low mol. weight" dataset)
- Select the desired alerts (remember, we have selected 417 alerts for five endpoints? Check the "only 417 selected alerts")

Provide the compounds to screen against the structural alerts

Upload compounds from a file
 (SDF/MOL2/SMILES/Excel sheet)

Provide a Name/CAS-RN/SMILES

Draw Molecule
 (click on depiction to the right to draw)
 

Choose a previously prepared set: [...] **1. Select the desired compounds**

Select molecules by a tag: HPV database

Select the structural alerts

Publication **2. Select the desired alerts**

Endpoint

Only approved alerts

Only 417 selected alerts

- Start screening and wait for the calculation task to finish. Can take 1-5 minutes for the set with several thousands molecules.
- You can get back to the task any time later in the browser of pending tasks (Menu item: *Models – View pending tasks*).

3. Analyze the screening results

The screening results dialog shows the potentially hazardous compounds grouped by endpoints, by alerts and by publications. Please, spend some time to click around.

- View the compounds for a particular endpoint (e.g., non-genotoxic carcinogenicity)

ENDPOINTS	
<input type="radio"/> Acute Aquatic Toxicity	1138 compounds
<input type="radio"/> Skin sensitization	880 compounds
<input type="radio"/> Genotoxic carcinogenicity, mutagenicity	631 compounds
<input type="radio"/> Idiosyncratic toxicity (RM formation)	696 compounds
<input checked="" type="radio"/> Non-genotoxic carcinogenicity	82 compounds

- Tag the filtered compound (e.g., "HPV non-genotoxic carcinogens")

View records for the filtered compounds

1 - 15 of 82 items on page 1 of 6 >>

	Halogenated benzene (for Non-genotoxic carcinogenicity in 2008 Benigni) Alkyl halide (for Acute Aquatic Toxicity in 1990 Hermens) Alkyl halides with only C, H and halogen atom (for Acute Aquatic Toxicity in 1990 Hermens) Aryl halide (for Acute Aquatic Toxicity in 1990 Hermens) Halogenated benzylic group (for Acute Aquatic Toxicity in 1990 Hermens)
MoleculeID: M5448	

Tags browser
Tags are convenient way to provide meaning to your data. You can apply multiple tags to articles and properties.

Show tags for **Molecules** Type part of tag name to filter: [search] [create new tag]

Tag editor
Add new tag or edit existing tag

Name: **Type a tag name**
tag name is good and length is 29

Tag type: **Molecules**

This tag is public (accessible by everybody)
 Show this tag in browser of experimental properties

Description:

(min. 50 characters) **Save the tag**

SAVE **CANCEL**

- Access, view or export the tagged structures into an SD-file via the browser of tags (Menu item: *Database – Tags*)

Tutorial 2 – SetCompare using functional groups

In this tutorial, we will use a dataset of ready-biodegradability. We have prepared two sets of molecules containing (a) 717 readily biodegradable compounds and (b) 1221 non-readily biodegradable compounds.

The datasets were taken from the study of Vorberg et. Al "Modeling the biodegradability of chemical compounds using the On-line CHEmical modeling Environment (OCHEM)", *J. Mol. Inf.*, in review.



Question of this tutorial:

What are the **key structural features** distinguishing readily biodegradable compounds from non-readily biodegradable ones?

Follow the steps below:

- Make sure you have the datasets available (Menu item: *Database – Tags*, activate the name filter by typing "ECO")

Tags browser
Tags are convenient way to provide meaning to your data. You can apply multiple tags to articles and properties.

Show tags for **Molecules** Type part of tag name to filter: **ECO** [search] [create new tag]

1 - 2 of 2

	ECO13 - OECD Not readily biodegradable	1221 molecules	OECD Not readily biodegradable ...	svorberg
	ECO13 - OECD Readily biodegradable	717 molecules	ECO13 - OECD Readily biodegrad ...	svorberg

1 - 2 of 2

- Open the SetCompare utility (Menu item: *Models – SetCompare utility*)
- Select the tags with readily- and non-readily biodegradable compounds and click "Next"

The SetCompare utility is experimental. It allows you to compare two sets of molecules based on their structural features. Please, provide the two sets available options below.

1 Select the compounds in the first set :	2 Select the compounds in the second set :
<p>Upload compounds from a file (SDF/MOL2/SMILES/Excel sheet) <input type="text"/> <input type="button" value="Browse..."/></p>	<p>Upload compounds from a file (SDF/MOL2/SMILES/Excel sheet) <input type="text"/> <input type="button" value="Browse..."/></p>
<p>Provide a Name/CAS-RN/SMILES <input type="text"/></p>	<p>Provide a Name/CAS-RN/SMILES <input type="text"/></p>
<p>Draw Molecule (click on depiction to the right to draw) </p>	<p>Draw Molecule (click on depiction to the right to draw) </p>
<p>Choose a previously prepared set: <input type="button" value="[-]"/></p>	<p>Choose a previously prepared set: <input type="button" value="[-]"/></p>
<p><input checked="" type="radio"/> Select molecules by a tag: ECO13 - OECD Not readily biodegradable</p>	<p><input checked="" type="radio"/> Select molecules by a tag: ECO13 - OECD Readily biodegradable</p>
<input type="button" value="Next >>"/>	

Select the sets you would like to compare

- Now we have to choose the descriptors used to compare our sets. We will use ToxAlerts' functional groups. Make sure you **uncheck** all descriptor types except of "Structural alerts". Select the "functional groups" in the endpoint filter.

Select the molecular descriptors:

Suggested descriptors:

- E-state [W](#)
- ALogPS (2) [W](#)
- GSFragment (1138) [W](#)
- Dragon v. 6.0 (4885/3D) [W](#)
- ISIDA fragments [W](#)
- ADRIANA.Code (211/3D) [W](#)
- CDK descriptors (246/3D) [W](#)
- 'Inductive' descriptors (54/3D) [W](#)
- MERA descriptors (529/3D) [W](#)
- MERSY descriptors (42/3D) [W](#)
- Chemaxon descriptors (499/3D) [W](#)
- QNPR [W](#)
- Spectrophores (144/3D) [W](#)

Additional or obsolete descriptors:

- OEState [W](#)
 - MolPrint [W](#)
 - Dragon v. 5.4 (1630/3D) [W](#)
 - Dragon v. 5.5 (3190/3D) [W](#)
 - Structural alerts (ToxAlerts) [W](#)
- Uncheck everything except ToxAlerts
- Select "functional groups"

Aromatize structures: Chemaxon General

Select alerts:

Publication: All articles

Endpoint: Functional groups

Only approved alerts

- MOPAC descriptors (21/3D) [W](#)
- ShapeSignatures (3D) [W](#)

Experimental descriptors (use only if you know how to use them):

- AMBIT Descriptors [W](#)
- ISIDA fragments (2011) [W](#)
- Chiral Descriptors (/3D) [W](#)
- Scaffold Hunter Descriptors [W](#)
- Functional Groups [W](#)
- ETM descriptors [W](#)
- DockingDescriptors (pre-pre-alfa) [W](#) *Not supported by your installation*
- Experimental values of other properties [W](#) *Not supported by your installa*

Outputs of other models [W](#)

[Add a model]

Next >>

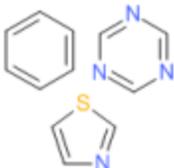
Start the calculations

- Again, you can access your calculation task any time later from the browser of pending tasks (Menu item: *Models – View pending tasks*)

Finally, when your calculation is completed, review and analyze the results!

The screen with results will show you the structural features that are statistically over-represented in the one or the other set.

This exemplary study shows you, for example, that Halogens are over-represented in non-readily biodegradable compounds.

Descriptor	In set 1 (1221 molecules)	In set 2 (717 molecules)	p-Value
<p>Halogens</p> <p>F Cl Br</p> <p>I At</p>	384	49	4.08E-41
<p>Halogens are specific for non-readily biodegradable compounds</p> <p>Click to see the "exceptions"</p>			
R—X	355	40	1.03E-40
	773	236	8.22E-39
	229	331	-4.11E-37
<p>Carboxylic acid derivatives are specific for readily biodegradable compounds</p> <p>The structural feature depiction. Click to see details.</p>			
R—X	220	11	2.96E-34



How to interpret these results? This is up to you as a researcher!

Conclusion

The tutorials demonstrate a fundamental difference of ToxAlerts (and the alerts-based analysis) from traditional QSAR: **interpretability** of the former approach.

While QSAR approaches are, in general, more powerful, universal and accurate, the alerts-based approach allows to obtain more interpretable results.

Thus, alerts-based analysis is a useful utility in the toolbox of a chemoinformatician that can complement the conventional QSAR techniques.