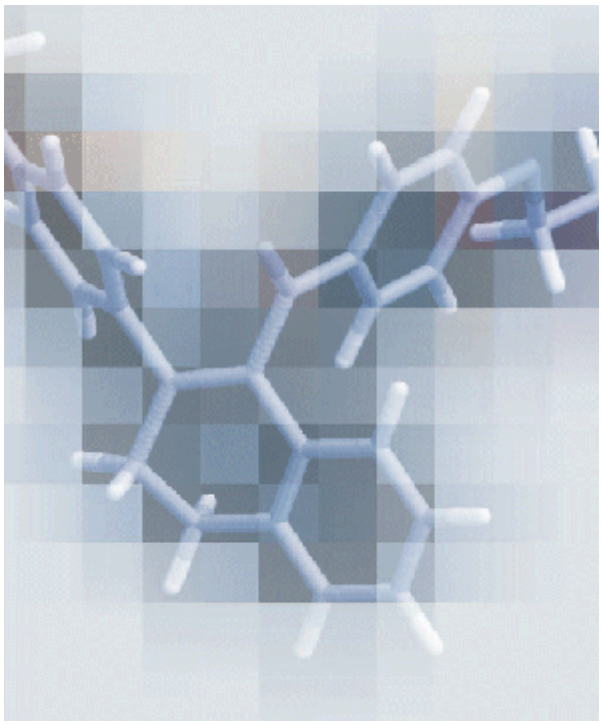
START

START (Structural Alerts for Reactivity in Toxtree)
biodegradation & persistence decision tree

Version 1.0

User Manual



Molecular Networks GmbH Computerchemie
October 2008

http://www.molecular-networks.com

Molecular Networks
Inspiring Chemical Discovery

Henkestr. 91
D-91052 Erlangen
Germany

Phone: +49-9131-815668
Fax: +49-9131-815669

Email: info@molecular-networks.com
WWW: www.molecular-networks.com

(Document version: 1.0-2008-12-16)

Contents

Introduction

The environmental persistence of a chemical is of vital importance in establishing its ecologic effects and environmental fate. A compound that is not persistent is generally considered to provide less risk than a persistent chemical with a similar toxic profile. However, persistence in the environment is an extremely complex phenomenon to model because it depends on the chemical structure, the environmental conditions and, for biodegradation, the ability of available microorganisms to degrade a chemical.

The aim of the START project was the development of a rulebase for the Toxtree software [1] which provides a compilation of structural alerts (SAs, see section “Structural Alerts”) for environmental persistence and biodegradability (see section “Biodegradability and Persistence”). These SAs are molecular functional groups or substructures that are known to be linked to the environmental persistence or biodegradability of chemicals. The rulebase utilizes the SAs in logical decision trees. If one or more SAs embedded in the molecular structure of the chemical are recognized, the system flags the potential persistence or biodegradability of the chemical.

Within the Toxtree software, the rulebase of the START project is encoded as a software program serving as a plug-in to Toxtree. This ensures that users being familiar with the Toxtree software can easily apply the features of the START plug-in to their problems.

Background

Biodegradability and Persistence

Biodegradation is an important process that can result in the loss or transformation of a chemical substance in the environment. Biodegradation of organic chemicals in the environment influences exposure and, hence, it is a key parameter for estimating the risk of long-term adverse effects on biota. Information on the biodegradability of chemicals is an essential requirement for hazard assessment (e.g. for classification and labeling), environmental risk assessment (for chemical safety assessment) and persistency assessments.

Moreover, biodegradability is not only a property or characteristic of a substance, but is also a system's concept, i.e. a system with its conditions determines whether a substance within it is biodegraded. When material is released into the environment, its fate depends upon a whole range of physiochemical processes and its interaction with living organisms. The most stable compound of carbon is carbon dioxide. All more reduced organic compounds are thermodynamically unstable in an oxygen environment and will be randomly attacked by microbial enzymes, provided that they have some structural similarity to naturally occurring substrates. The California Advertising Statute, amended on April 30, 1991, states that a manufacturer cannot claim that a product is biodegradable unless it meets the following definition [2]:

'Biodegradable means that a material has the proven capability to decompose in the most common environment where the material is disposed of within 3 years through natural biological processes into nontoxic carbonaceous soil, water, carbon dioxide or methane.'

A failure to show biodegradation after prolonged exposure renders the chemical as non-biodegradable or persistent.

Environmental persistence - the ability of certain chemical substances to persist in the environment - is an issue of global concern that requires careful consideration in environmental risk assessment. This is especially true when this ability is coupled with bioaccumulative and toxicological properties. However, assessing environmental persistence of chemical substances is not straightforward. Persistence cannot be directly measured; it can only be inferred from the continued presence of a substance (or of metabolites, when they are persistent) in the environment or the lack of observed biodegradation data in the environment.

Structural Alerts

A Structural Alert (SA) is a molecular functional group or substructure that is known to be linked to a certain chemical property or reactivity. For the START project SAs were collected that are known to be related to the environmental persistence or biodegradability of chemicals. A SA that is related to biodegradability is, e.g., the ester group [3]. Under environmental conditions, esters are rapidly degraded into alcohols and carboxylic acids (see Figure 1)[4].

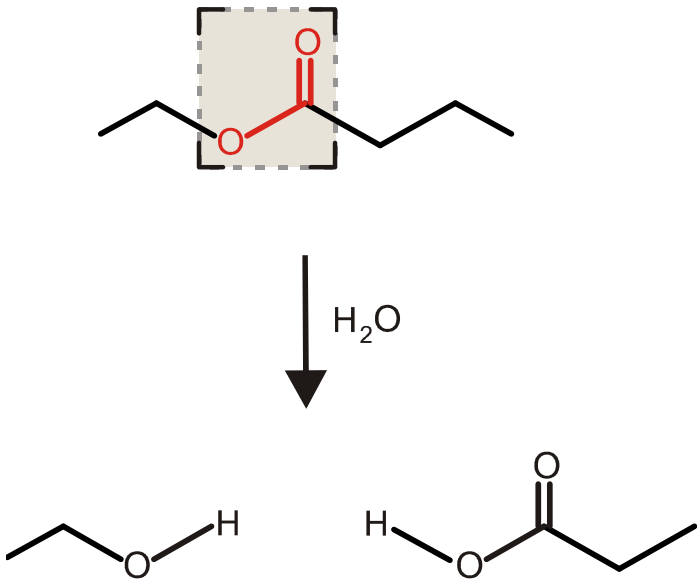


Figure 1 The biodegradation of an ester (SA marked in red) leads to an alcohol and a carboxylic acid.

A set of chemicals that is characterized by the same SA constitutes a family (class) of compounds that share the same mechanism of action. In the example given above, the mechanism of degradation is a simple ester hydrolysis.

The reactivity of a SA can be modulated (increased or reduced) by the remaining part of the molecule in which the SA is embedded. At a coarse-grain level, the modulating effect can be represented by other molecular substructures (e.g., bulky groups ortho to an aromatic amine group) that are known to have an influence on the reactivity of the SA. Usually, the knowledge on the modulating substructures is quite limited for most of the SAs, thus it provides only limited help in deciding if chemicals are biodegradable or persistent.

The START Rulebase

The START rulebase estimates potential biodegradability or environmental persistence, by using a series of Structural Alerts (SAs) in combination with a decision tree. If the substance contains degradable functional groups, it usually can be considered as having the potential to degrade in the compartments to which it partitions. Depending on the weight of evidence that can be gathered on the potential for degradation, a determination can be made as to whether a substance is persistent according to the Persistence and Bioaccumulation Regulations (Government of Canada 2000)[5].

Structural Alters

The SAs included in Toxtree are 32; out of them, 23 SAs refer to environmental persistent chemicals mechanisms of action while 9 SAs refer to easily biodegradable chemicals. Appendix 1 provides the list of the SAs, together with examples of representative chemicals for each of them. The SAs were derived from the *Guidance Manual for the Categorization of Organic and Inorganic Substances on Canada's Domestic Substances List*, which was assembled from the Existing Substances Branch of Environment Canada [6].

Classification Scheme

The processing of a query chemical by the START plug-in for Toxtree can result in different outcomes. According to the SAs that are utilized in the START rulebase, chemicals are classified into one of the following three categories:

- Class 1 (easily biodegradable chemical)
- Class 2 (persistent chemical)
- Class 3 (unknown biodegradability)

If one or more SAs for biodegradability is found in the query chemical it is classified as biodegradable (Class 1) and can therefore not be classified as persistent any more. If no SAs for biodegradability are found in the query chemical, the existence of SAs for environmental persistence is evaluated. If one or more SAs for persistence is found in the query chemical it is classified as persistent (Class 2). If neither SAs for biodegradability nor environmental persistence are found in the query chemical, the environmental fate of the chemical cannot be assed with the START rulebase and the chemical is therefore classified as Class 3 (unknown biodegradability).

The Toxtree START plug-in

Toxtree is a Java based decision tree software with a plug-in interface. Therefore, the START plug-in for Toxtree is also Java based and utilizes the Toxtree plug-in interface to make its functionality available to the user. The following two sections describe the installation and the usage of the START plug-in for Toxtree.

Installation

The implementation of the START plug-in follows the standard plug-in mechanism of Toxtree as used in the existing decision tree methods (Verhaar scheme, Cramer rules, etc.). The functionality of the START plug-in is included in the " toxBiodegradation.jar" library file which has to be copied into the plug-in directory of the Toxtree installation as shown in Figure 2.

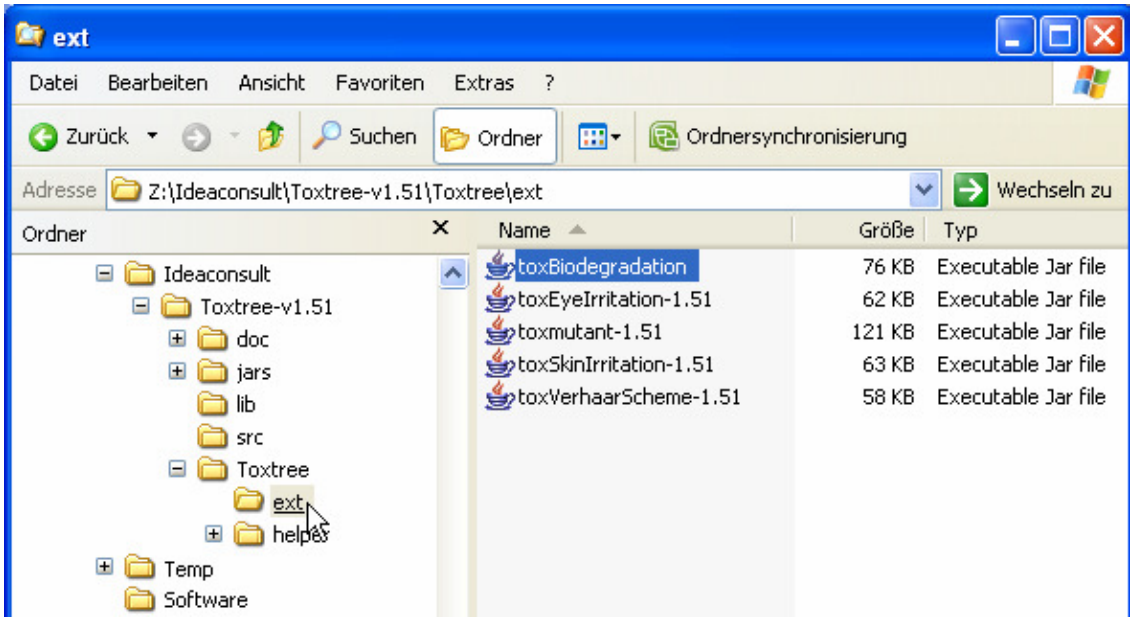


Figure 2 Installation of the START plug-in for Toxtree. The library file "toxBiodegradation.jar" has to be copied into the plug-in directory of the Toxtree installation.

Usage

Once the START plug-in has been installed into Toxtree it has to be activated. This can be done in the menu bar of Toxtree as shown in Figure 3.

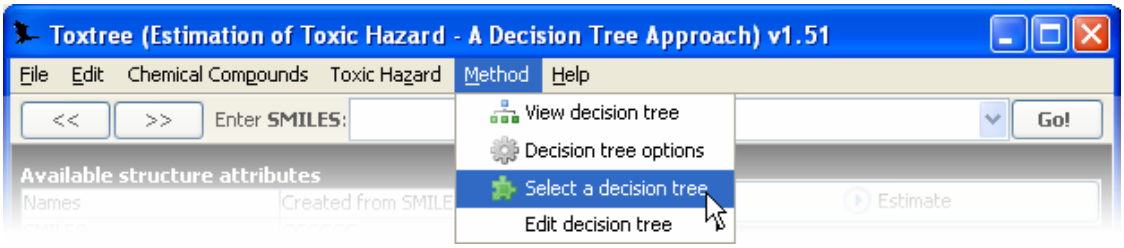


Figure 3 The “Select a decision tree” option in the START menu bar.

Within the “Select a tree” dialog window the START plug-in can be selected from the list of available decision trees (“START Biodegradation”) like any other Toxtree plug-in as illustrated in Figure 4. The START plug-in is now ready to use.

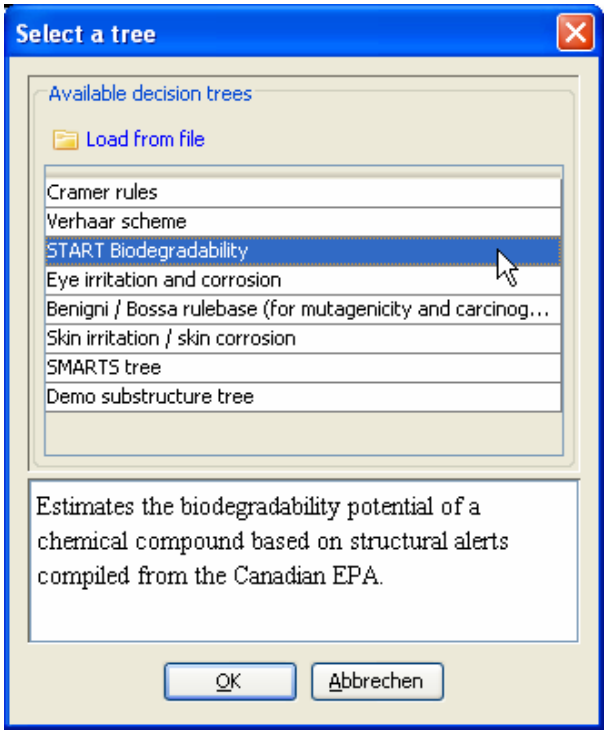


Figure 4 Decision tree selection in the “Select a tree” dialog window of Toxtree.

As shown in Figure 5 the SMILES notation of the chemical under investigation can now be entered into field (A) of the Toxtree main window. In (B) the 2D-depiction of the chemical is shown. The biodegradation potential of the chemical is assessed by pressing the corresponding “Estimate” button (C). The START plug-in classifies the chemical in one of the three categories – “easily biodegradable chemical”, “persistent chemical”, or “unknown biodegradability” (D).

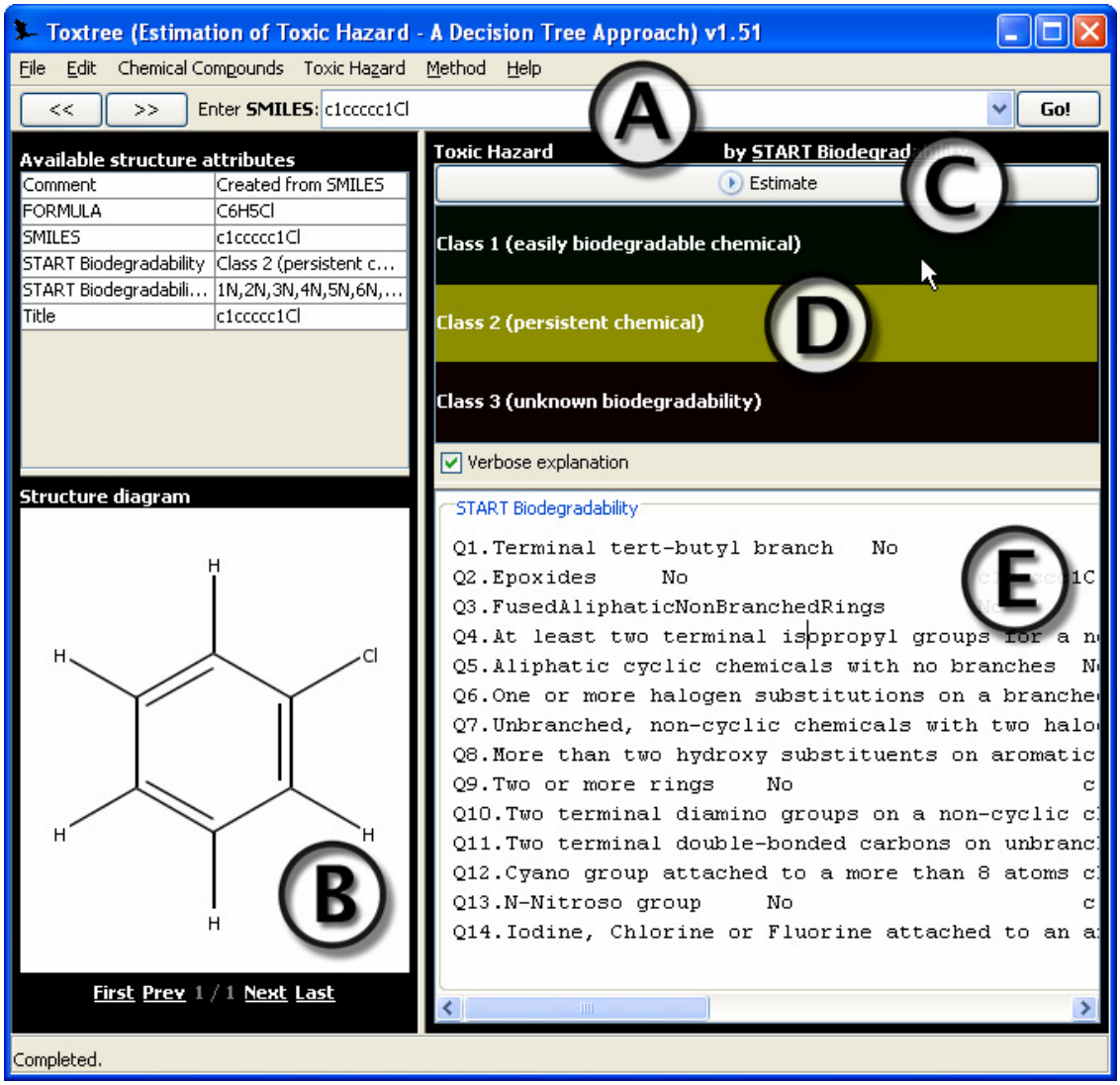


Figure 5 Components of the Toxtree main window: (A) field for the input of a chemical structure in the SMILES notation, (B) chemical structure preview, (C) “Estimate” button to start the evaluation of the chemical, (D) result of the evaluation, (E) detailed description of the evaluation process.

The procedure for assigning any of these labels is outlined below:

1. The compound under investigation is searched for the presence of all structural alerts notifying a persistent chemical and the number of found alerts is kept.
2. If at least one alert notifying an easily degradable chemical is found, the compound under investigation is declared as “easily biodegradable chemical”.
3. If no alerts for notifying an easily degradable chemical are found and there is at least one alert notifying a persistent chemical the compound under investigation is declared as “persistent chemical”.
4. If no alerts were found at all, the compound under investigation is declared as “unknown biodegradability”.

Finally, the user is presented with a detailed summary of all alerts found in the compound under investigation (E).

References

[1] The Toxtree software, ECB, 2008, Website (http://ecb.jrc.it/qsar/qsar-tools/).

[2] S. Harold, Biodegradability: Review of the Current Situation, 1993, Lubrizol Corporation.

[3] Reaction Mechanisms in Environmental Organic Chemistry, R.A. Larson; E.J. Weber, Lewis Publishers, 1994, pp 124-160.

[4] Environmental Organic Chemistry, R.P. Schwarzenbach; P.M. Gschwend; D.M. Imboden, Wiley-Interscience Publishers, 1993, pp 372-387.

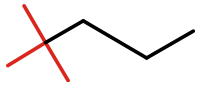
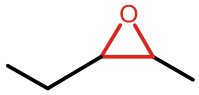
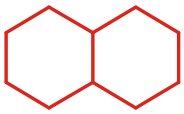
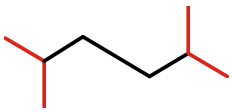
[5] Guidance Manual for the Categorization of Organic and Inorganic Substances on Canada's Domestic Substances List, Existing Substances Branch,Environment Canada, 2003, pp 10-11.

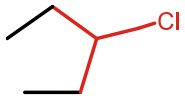
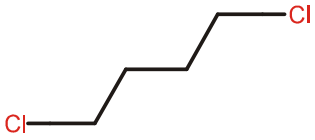
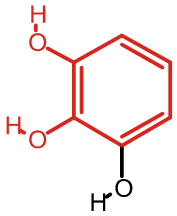
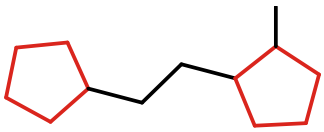
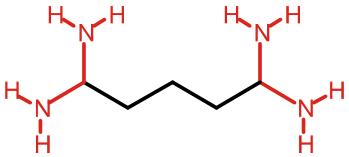
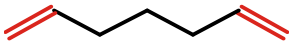
[6] Guidance Manual for the Categorization of Organic and Inorganic Substances on Canada's Domestic Substances List, Existing Substances Branch,Environment Canada, 2003, pp 89-90.

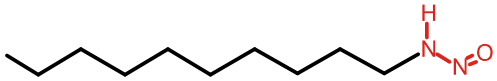
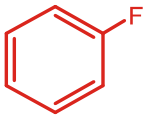
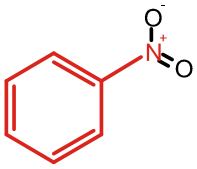
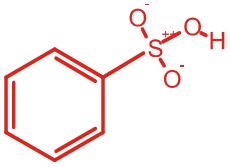
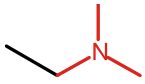
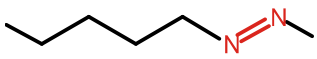
Appendix

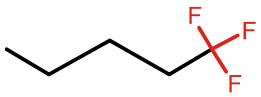
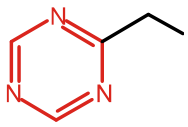
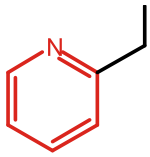
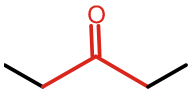
Structural Alerts included in the TOXTREE START plug-in

Structural Alerts for Environmental Persistence

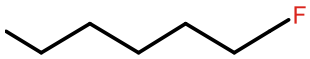
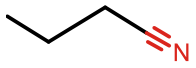
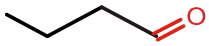
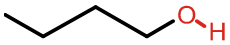
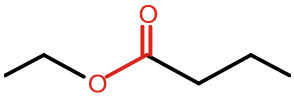
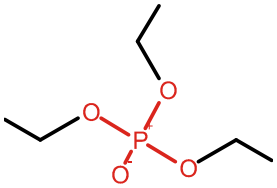
RuleID	Structural Alert (SA)	Example (SA highlighted)
1	terminal t-butly group	
2	*epoxy group*	
3	*aliphatic fused ring, non-branched*	
4	*two or more isopropyl groups, non-cyclic*	
5	*aliphatic, cyclic, non-branched*	

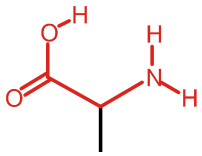
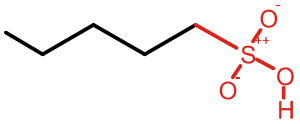
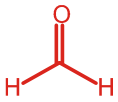
RuleID	Structural Alert (SA)	Example (SA highlighted)
6	*one or more halogens, branched*	
7	*two halogens, non-cyclic, non-branched*	
8	*two or more hydroxyl groups attached to an aromatic ring*	
9	*two or more cycles*	
10	*two terminal diamino groups, non-cyclic*	
11	*two terminal double bonded carbon atoms*	
12	*cyano group attached to a long chain (more than 8 atoms)*	

RuleID	Structural Alert (SA)	Example (SA highlighted)
13	*N-nitroso group*	
14	*fluorine, chlorine, or iodine attached to an aromatic ring*	
15	*nitrogen attached to an aromatic ring*	
16	*aromatic sulphonic acid*	
17	*ether group, aliphatic*	
18	*tertiary amino group*	
19	*azo group*	

RuleID	Structural Alert (SA)	Example (SA highlighted)
20	*trifluromethyl group*	
21	*triazine ring*	
22	*pyridine ring*	
23	*keto group*	

Structural Alerts for Biodegradability

RuleID	Structural Alert (SA)	Example (SA highlighted)
24	*one halogen, non-branched*	
25	*nitrile group*	
26	*aldehyde group*	
27	*hydroxyl group*	
28	*ester group*	
29	*phosphoric ester group*	

RuleID	Structural Alert (SA)	Example (SA highlighted)
30	*amino acid*	
31	*aliphatic sulphonic acid group*	
32	*formaldehyde*	

The XML based Toxtree START Plugin

The XML based Toxtree START Plugin is an extension of the Toxtree START plug-in which facilitates the extension of the Toxtree rulebase without coding new rules in JAVA classes. Therefore, the XML based Toxtree START Plugin uses a rulebase that is encoded in an XML file. This XML files provides all information that is needed to encode the rulebase, e.g., SMARTS pattern to define SAs as illustrated in Figure 6.

New rules can easily be added by extending the XML file that encodes the rules. The XML file is shipped as part of the XML based Toxtree START plug-in and as such is available at any time. When the XML based Toxtree START plug-in is activated in Toxtree, the XML rulebase is loaded and can then be applied by the user.

```
<!--
        Document : alerts.xml
        Author : Molecular Networks
        Description: Set of structural alerts for the START plugin
-->
<alerts xsi:schemaLocation="http://www.molecular-networks.com/schema/start.xsd start.xsd">
        <degradable>
                <alert id="1" editable="no" containsAll="yes">
                        <smarts id="1" negate="yes" count="1" countType="atleast">
                                [$([C][C;!R]([C])[C])]
                        </smarts>
                        <smarts id="2" negate="yes" count="1" countType="atleast">[R]</smarts>
                        <smarts id="3" negate="no" count="1" countType="exact">[#6][F,Cl,Br,I]</smarts>
                        <title>
                                Unbranched chemicals with one halogen substitutions
                        </title>
                        <explanation>
                                Unbranched chemicals with one halogen substitutions are
                                associated with high biodegradability
                        </explanation>
                        <negativeExample>CC(CC)CCCCCl</negativeExample>
                        <positiveExample>CCCCCCCF</positiveExample>
                </alert>
        </degradable>
        <persistent>
                <alert id="2" editable="no" containsAll="yes">
                        <smarts id="1" negate="no" count="1" countType="atleast">
                                [$([C]([CH3])([CH3])([CH3])([*]))]
                        </smarts>
                        <title>Terminal tert-butyl branch</title>
                        <explanation>
                                Compounds, which contain a terminal tert-butyl branch are persistent
                        </explanation>
                        <negativeExample>CCCCC</negativeExample>
                        <positiveExample>C(C)(C)(C)CCC</positiveExample>
                </alert>
        </persistent>
</alerts>
```

Figure 6 A simplified example of a knowledge base stored in an XML format.