

Toxtree Course QSAR2014

IDEACONSULT

QSAR2014 Toxtree course

© Ideaconsult Ltd.
4 Angel Kanchev St.
1000 Sofia, Bulgaria
Phone +359 886802011 • Email jeliazkova.nina@gmail.com

Version of 15 June 2014

Table of Contents

<i>Download</i>	<i>1</i>
<i>Prerequisites</i>	<i>1</i>
<i>Toxtree distributions</i>	<i>1</i>
<i>Windows™ Toxtree setup</i>	<i>1</i>
<i>Toxtree setup for other OS</i>	<i>2</i>
<i>Launching Toxtree</i>	<i>2</i>
<i>Main screen layout</i>	<i>3</i>
<i>Exercise 1. Cramer rules</i>	<i>5</i>
<i>Exercise 2. Biodegradability</i>	<i>7</i>
<i>Exercise 3. Batch processing</i>	<i>8</i>
<i>Further readings</i>	<i>11</i>
<i>Opening/saving a file</i>	<i>12</i>
<i>Decision tree visualisation</i>	<i>12</i>
<i>Exercise 4. Toxtree online</i>	<i>14</i>

Download

- Open <http://toxtree.sourceforge.net> in a web browser.
- Download Toxtree-2.6.6
http://toxtree.sourceforge.net/download.html#Toxtree_2.6.6

Select “Windows installer” option for Windows , or ZIP archive for Linux or Mac.

Prerequisites

Toxtree requires Java™ 2 Runtime Environment, Standard Edition 1.7 or newer on the target system and it is platform-independent. It runs under any host operating system, which supports Java™ 2 Runtime Environment, Standard Edition.

Toxtree distributions

Toxtree is distributed in two different ways:

- ZIP archive, containing the complete documented Java™ source code, binaries, example lists of compounds, installation manual and user manual;
- Windows™ (NT, 2000, XP or 2003) standalone (offline) automated installer with integrated Java™ 2 Runtime Environment, Standard Edition 1.7 setup;

The ZIP archive is suitable for use on any operating system with Java™ 2 Runtime Environment, Standard Edition 1.7 already installed. It might be useful also for studying the application source code.

The Windows™ installer includes all the contents of the ZIP archive, as well as the ability to check for the presence of Java™ 2 Runtime Environment, Standard Edition 1.7 on the target system. If the installer does not detect Java™ 2 Runtime Environment, Standard Edition 1.7 already installed on the target system, it would attempt to install it, prior to Toxtree setup.

Windows™ Toxtree setup

The Toxtree application is distributed with a fully automated offline installer (Toxtree-vX.Y.Z-setup.exe), compatible with recent versions of the Microsoft Windows (NT, 2000, XP, 2003, 7, 2008) operating system. In order to install Toxtree just run the

installer and follow its instructions. The installer contains all the required packages, including the Java™ 2 Runtime Environment, Standard Edition 1.7 setup.

If the installer does not detect Java™ 2 Runtime Environment, Standard Edition 1.7 or newer on the target system, it will attempt to install it before proceeding with the Toxtree setup. In this case, the user running the installer SHOULD HAVE ADMINISTRATIVE PRIVILEGES on the target system (otherwise the Java™ 2 Runtime Environment, Standard Edition 1.7 setup would be aborted and Toxtree would not be installed).

If the installer detects Java™ 2 Runtime Environment, Standard Edition 1.7 or newer already installed on the target system, it will proceed directly with the Toxtree setup. In this case, ADMINISTRATIVE PRIVILEGES ARE RECOMMENDED BUT NOT NECESSARILY REQUIRED.

When the installer is launched by a user with administrative privileges, it will create Toxtree start menu shortcuts for all the users registered in the target system. Otherwise, start menu shortcuts will be created only for the unprivileged user, who launched the installation.

After a successful installation, Toxtree could be launched from the Start Menu ("Start►All Programs►Ideaconsult►Toxtree-vX.Y.Z►Toxtree-vX.Y.Z").

Toxtree setup for other OS

Toxtree runs under any host operating system, which supports Java™ 2 Runtime Environment, Standard Edition (e.g. Linux, FreeBSD, Solaris, Mac OS, etc...). If you're running such operating system, you should download and use the ZIP archive distribution of Toxtree.

Please, note that before attempting to run Toxtree, you should check that Java™ 2 Runtime Environment, Standard Edition 1.7 (or newer) is installed on your system. If Java™ 2 Runtime Environment, Standard Edition is missing (or an older version is present), please obtain (from <http://www.oracle.com/technetwork/java/javase/downloads/jdk7-downloads-1880260.html>) and run the appropriate Java™ 2 Runtime Environment, Standard Edition installer for your system.

Launching Toxtree

In Windows™ platforms, Toxtree can be launched either by using the "Start" menu, or by double clicking on the Toxtree-X.Y.Z.jar file (the full path name is "C:\Ideaconsult\Toxtree-vX.Y.Z\Toxtree\Toxtree-X.Y.Z.jar").

In all platforms (having Java™ 2 Runtime, Standard Edition 1.7 or newer installed), Toxtree can be launched by executing the following command (after decompressing the ZIP archive distribution of Toxtree):

```
java -jar Toxtree-X.Y.Z.jar
```

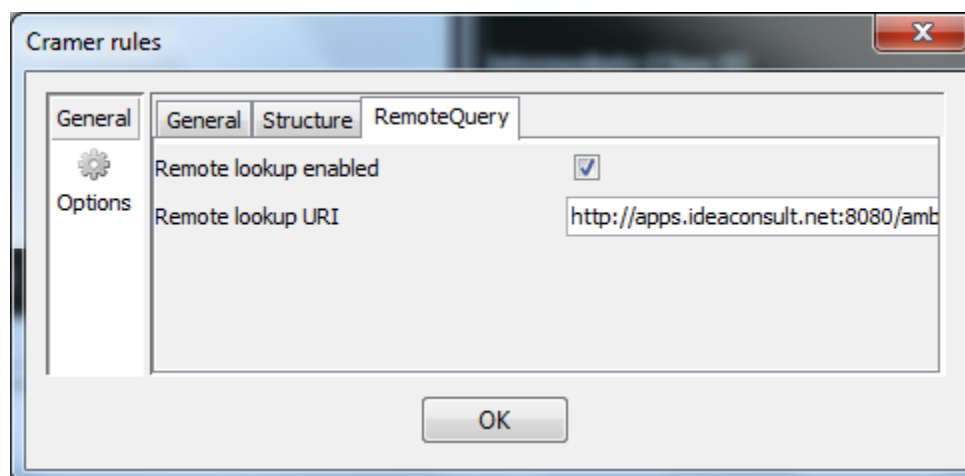
Please, note that in the above mentioned command “java” and “Toxtree-X.Y.Z.jar” should be eventually prefixed with the full path to java and Toxtree on the destination platform.

Main screen layout

The main Toxtree application window comprises a title bar, menu bar, data areas, button bar and status bar. The data areas are highlighted in Figure 1 in several different colours and are labelled in blue.

The CHEMICAL IDENTIFIER INPUT AREA can be used for entering a SMILES string, InChI, IUPAC chemical name, CAS or EINECS.

SMILES, InChI and IUPAC chemical names are parsed by Toxtree and structure is generated and displayed in the compound area. If CAS or EINECS is entered, and the remote queries are enabled, Toxtree will attempt to query a remote OpenTox web service, in order to retrieve the chemical structure. This functionality is disabled by default. It can be enabled by selecting the “Remote lookup enabled” checkbox in the “Method ► Decision tree options ► Options ► Remote Query” menu, as shown in the following screenshot:



Pressing the Go! button draws the structure diagram of the corresponding compound in the COMPOUND STRUCTURE DIAGRAM AREA.

The << and >> buttons on the left hand side of the CHEMICAL IDENTIFIER INPUT AREA are used for navigating the history of entered identifiers. This is also accessible also by means of a drop-down list.

The COMPOUNDS PROPERTIES AREA is used to summarise the available information about the current compound.

The FILE BROWSING AREA provides an easy way to navigate through the list of compounds in the current open file.

The CLASSIFICATION AREA provides access to the Cramer classification results for the current compound.

Pressing the Estimate button starts the classification routines for the current compound (shown both in the COMPOUNDS PROPERTIES AREA and the COMPOUND STRUCTURE DIAGRAM AREA).

The classification result is shown in graphical form (green highlight for class I, yellow highlight for class II and red highlight for class III), as well as in text form.

The Verbose explanation checkbox determines the level of detail of the text-based classification results.

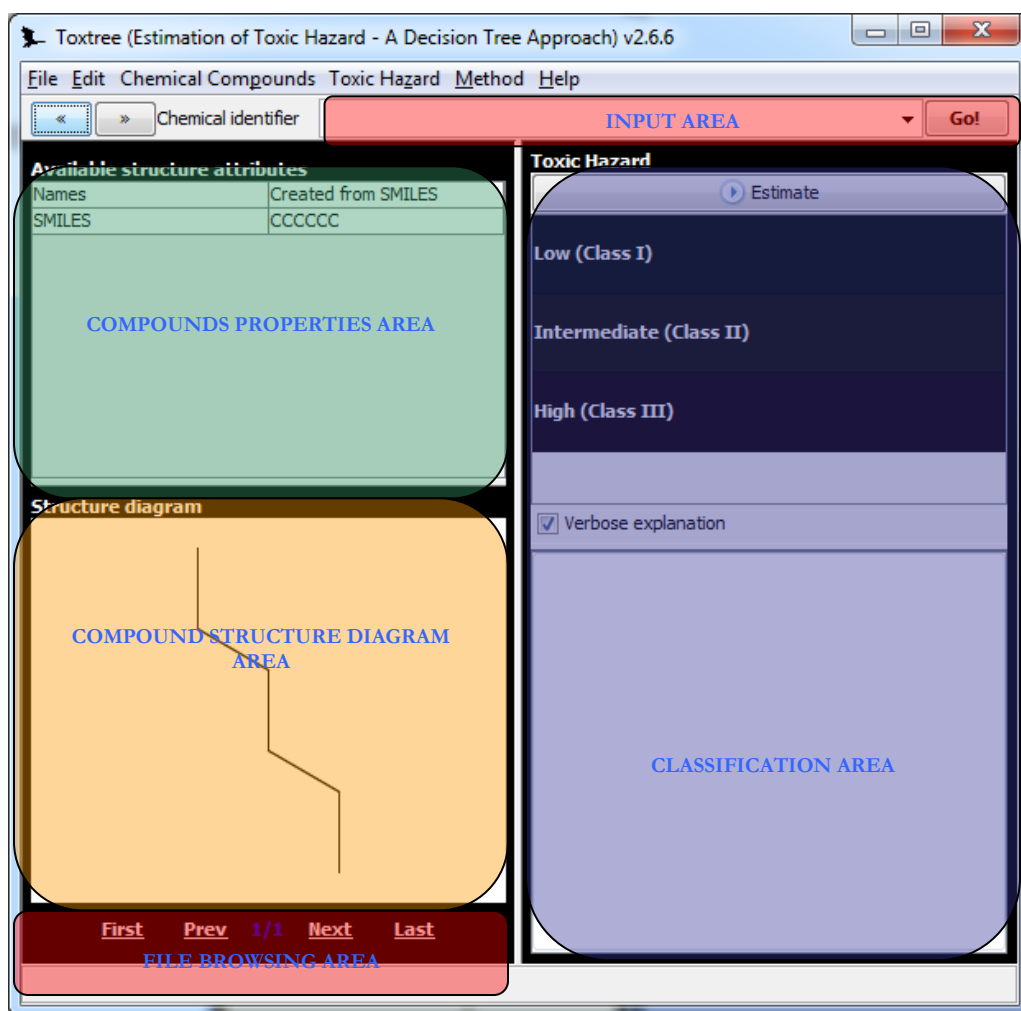


Figure 1: Toxtree main application window

Exercise 1 . Cramer rules

1. launch the application;
2. enter the following SMILES in the INPUT AREA field;
O=C(OCC(OC(=O)c1ccccc1)C)c2ccccc2
3. press "Estimate" (on the right upper part of the application window);
4. read the toxic hazard classification of the structure by application of Cramer rules (low - class I in this case);
5. read the classification explanation (which Cramer rules have been applied);

6. enter the following SMILES in the INPUT AREA field;

CSCCCN=C=S

7. press "Estimate" (on the right upper part of the application window);
8. read the toxic hazard classification of the structure by application of **Cramer rules** (intermediate - class II in this case);
9. read the classification explanation (which Cramer rules have been applied);
10. enter the following SMILES in the INPUT AREA field;

c1ccc2c(c1)cccc2(NC(N)=S)

11. read the toxic hazard classification of the structure by application of Cramer rules (high - class III in this case);
12. save the processed molecule, together with classification data (class & path) in a file (compatible types are CSV, SDF & TXT).
13. Click on the "Chemical compounds ► Edit compound" menu option. This will launch the JChemPaint¹ structure diagram editor and allows editing the current structure. Edit the structure. Click OK when ready. This will update the structure, displayed on the main screen.
14. Click again on the "Estimate" button. This will apply the decision tree to the new structure.
15. read the toxic hazard classification of the structure by application of **Cramer rules**.
16. read the classification explanation (which structure alerts have been verified); Click on blue links to highlight the structure alerts found.
17. save the processed molecule, together with classification data (class & path) in a file (compatible types are CSV, SDF & TXT). Please, note that CSV files can be read/written by MS Excel™;

Cramer rules background

The threshold of toxicological concern (TTC) is an open research topic with significant practical implications. Two principal approaches exist in the thresholds developed to

¹ More information about JChemPaint can be found at <http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint>

date; a general threshold or a threshold developed in relation to toxicological data or structural information. Thresholds based on structural information have typically been developed by the principles established by Cramer. Chemicals are divided into three structural classes based on a decision tree. This comprises some 33 structural rules and places evaluated compounds into one of three classes:

- Class I substances are simple chemical structures with efficient modes of metabolism suggesting a low order of oral toxicity;
- Class III substances are those that permit no strong initial presumption of safety, or may even suggest significant toxicity or have reactive functional groups;
- and finally, Class II are intermediate.

This plug-in was developed by Ideacon Ltd (Sofia, Bulgaria) on behalf of JRC.

More details at <http://toxtree.sourceforge.net/cramer.html>

Exercise 2 . Biodegradability

Typical use of Toxtree could follow the following scenario:

1. change the decision tree used in the estimation through the Toxtree "Method►Select decision tree" menu; Select **START Biodegradability**
2. enter a the following chemical name in the INPUT AREA field;

2,2'-iminodiethanol

3. press "Estimate" (on the right upper part of the application window);
4. read the toxic hazard classification of the structure by application of **START Biodegradability** (Class1 –easily biodegradable chemical in this case);
5. read the classification explanation (which rules have been applied);
6. enter a the following chemical name in the INPUT AREA field;

(2S,3S,5R)-3-methyl-7-oxo-3-(1H-1,2,3-triazol-1-ylmethyl)-4-thia-1-azabicyclo[3.2.0]heptane-2-carboxylic acid 4,4-dioxide

7. press "Estimate" (on the right upper part of the application window);

8. read the toxic hazard classification of the structure by application of **START Biodegradability** (Class2 – persistent chemical);
9. read the classification explanation (which rules have been applied);
10. enter a the following EINECS number² in the INPUT AREA field;

205-288-3

11. press "Estimate" (on the right upper part of the application window);
12. read the toxic hazard classification of the structure by application of **START Biodegradability** (Class2 – persistent chemical);
13. read the classification explanation (which rules have been applied); The answer Q18.Tertiary amine is positive, click on the blue link “Q18.Tertiary amine” in order to highlight the tertiary amine substructure.
14. consult the decision tree (accessible through the Toxtree "Method►View decision tree" menu);

Background: START (Structural Alerts for Reactivity in Toxtree) biodegradation and persistence plug-in

A compilation of structural alerts for environmental persistence and biodegradability. These structural alerts are molecular functional groups or substructures that are known to be linked to the environmental persistence or biodegradability of chemicals. The rulebase utilizes the structural alerts in logical decision trees. If one or more the structural alerts embedded in the molecular structure of the chemical are recognized, the system flags the potential persistence or biodegradability of the chemical. Installation and user manual is available [online](#).

This plug-in was developed by Molecular Networks GmbH (Erlangen, Germany).

<http://toxtree.sourceforge.net/start.html>

Exercise 3 . Batch processing

When dealing with large datasets (more than 1000 molecules), Toxtree should be used in batch processing mode. It can be accessed through the “File►Batch processing” menu. Users are invited to **select the input and output files (supported types are CSV, TXT and SDF)**, before starting the batch processing. The batch can be paused

² EINECS lookup requires “Remote lookup enabled”, please consult page 3 on how to enable the remote queries

or aborted upon user request. Batch configuration/state can be saved and loaded at a later time. In case of unexpected interruption (e.g. power failure, hardware failure, operating system failure, etc...), the batch job can be continued from the last previously fully processed record. In order to achieve this, the interrupted batch configuration should be loaded (either from a user-specified file, or from the system TEMP folder, where it is automatically stored if the user has not specified a file).

1. change the decision tree used in the estimation through the Toxtree "Method►Select decision tree" menu; Select **In vitro mutagenicity (Ames test) by ISS**
2. Download the SDF file “Benchmark Data Set for In Silico Prediction of Ames Mutagenicity” from

http://doc.ml.tu-berlin.de/toxbenchmark/tox_benchmark_N6512.sdf

3. consult the decision tree (accessible through the Toxtree "Method►View decision tree" menu);
4. classify all the molecules, loaded from a file, by using the batch processing facility; Select “File►Batch processing”. Select **tox_benchmark_N6512.sdf** as file to be processed and enter a new file name to store the results (e.g. **test.csv**; supported output files are SDF, CSV, TXT). Press “Start” button.

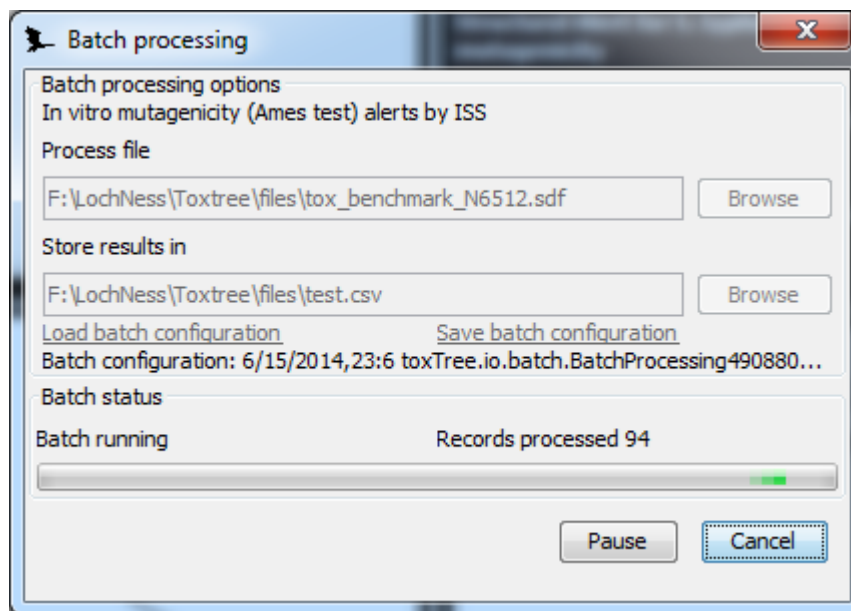


Figure 2: Batch processing

5. Please note the calculation may take a while. The batch process may be cancelled at any time.

The QSAR calculations can be time consuming due to the requirement to calculate electronic descriptors. The software provides options to skip the QSAR calculation, and in this case will assign the category “**For a better assessment a QSAR calculation could be applied**”. Toxtree will show the dialogs at Figure 3 or Figure 4 , if a QSAR is to be calculated.

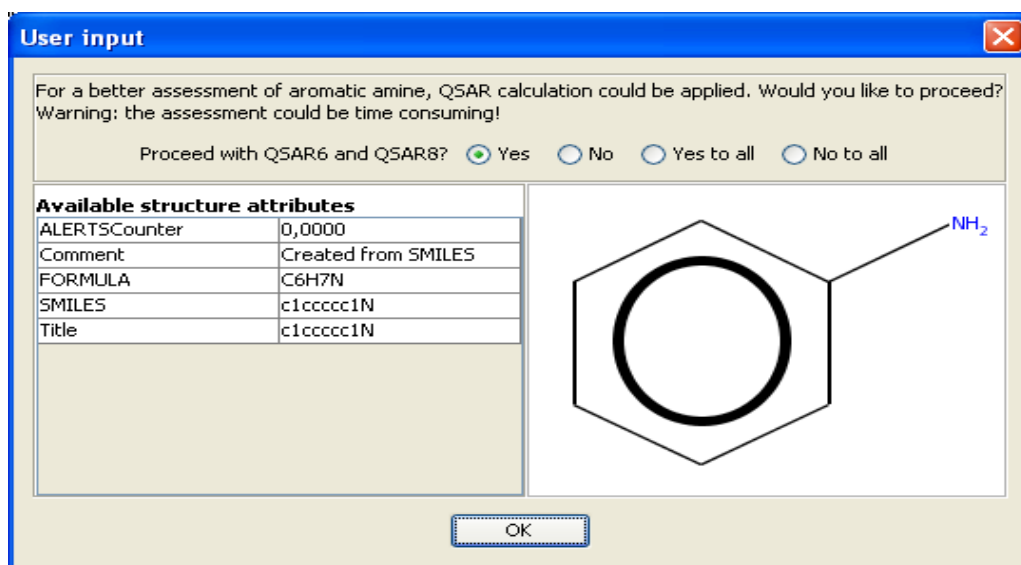


Figure 3: Options for QSAR6 and QSAR8 calculations

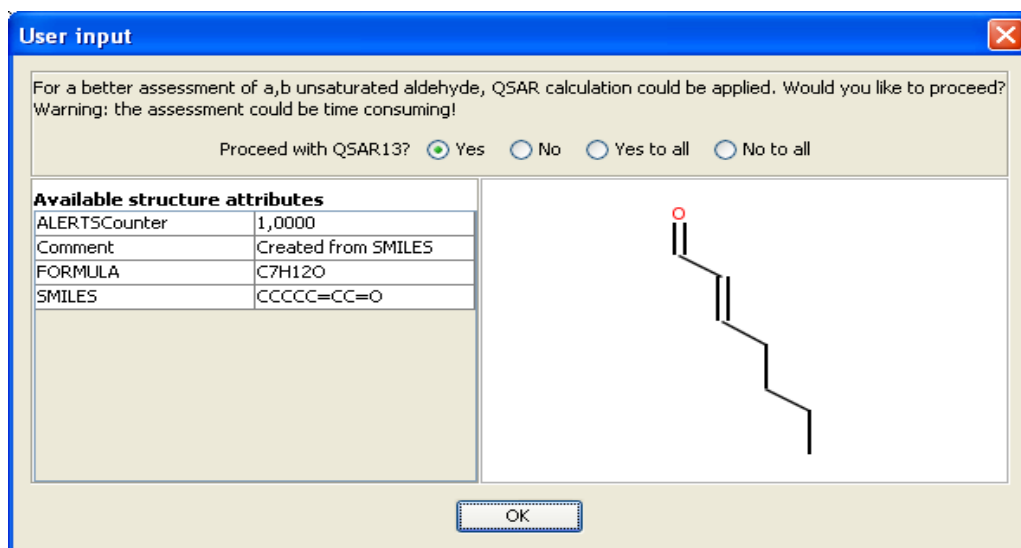


Figure 4: Options for QSAR13 calculations

The available options have the following semantics:

- **Yes:** The QSAR will be calculated for the current chemical, and the question dialog will appear for each subsequent chemical
 - **No:** The QSAR will not be calculated for the current chemical, and the question dialog will appear for each subsequent chemical
 - **Yes to all:** The QSAR will be calculated for the current chemical and for all subsequently processed chemicals, for which it is applicable. The question dialog will not appear anymore, unless the option is changed via the “Method ► Decision Tree Options” menu.
 - **No to all:** The QSAR will be not calculated for the current chemical and for all subsequently processed chemicals, for which it is applicable. The question dialog will not appear anymore, unless the option is changed via the “Method ► Decision Tree Options” menu.
6. Once the batch is completed, the results are stored in “test.csv” (comma delimited file), as shown on Figure 5 .

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1	SMILES	Source	Activity	WDI_Name	REFERENCE_MC_Exam_MC_Pred	DEREK_A	DEREK_P	Molecular Weight	Structural Alerts	Potential	Unlikely to be a test	Error when tested	SAS0	SAS1	SAS2	SAS3	SAS4	SAS5	SAS6	SAS7	SAS8	SAS9	SAS10
2	O=C1C=CC(=O)C=C1	CCRS	0	MILAMIN	0	0	0	108.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
3	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
4	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
5	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
6	CC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	106.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
7	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
8	CC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	106.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
9	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
10	NC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	107.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
11	CC1=CC=CC=C1	CCRS	1	MILAMIN	0	0	0	106.094	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO

Figure 5: Sample CSV file with prediction results.

Background : A decision tree for estimating in vitro mutagenicity (Ames test). Benigni, R., Bossa C., Tcheremenskaia O. (2013) In vitro cell transformation assays for an integrated, alternative assessment of carcinogenicity: a data-based analysis. *Mutagenesis* 2013;28(1):107-16.

This plug-in was developed by Istituto Superiore di Sanita (Rome, Italy) and Ideaconult Ltd (Sofia, Bulgaria) on behalf of JRC.

<http://toxtree.sourceforge.net/ames.html>

Further readings

User manuals in the **doc** folder

On Windows : C:\Ideaconult\Toxtree-v2.6.6\doc

Opening/saving a file

Toxtree can open CML, CSV, HIN, ICHI, INCHI, MDL MOL, MDL SDF, MOL2, PDB, SMI, TXT and XYZ file types.

Please, note that CSV files can be read/written by MS Excel™. Attention should be paid to cells type (should be 'text', otherwise MS Excel™ could interpret/show incorrectly their values). Also, note that input of CSV and TXT requires a column with "SMILES" heading in order the structure to be read. All other fields are optional, will be read as molecule properties and displayed as such.

You can open a supported type of file by using the “File►Open” menu. Molecules can be written to SDF, CSV or TXT files, together with their classification data (class & path), by using the “File►Save” menu.

Decision tree visualisation

The current decision tree is accessible through the “Method►View decision tree” menu. It is shown in a child window. Rule details are printed after clicking on the respective tree nodes. Leaves are coloured according to classes (I - green, II - yellow, III - red). There are example molecules for each rule outcome (“Yes” or “No”), selectable by radio buttons.

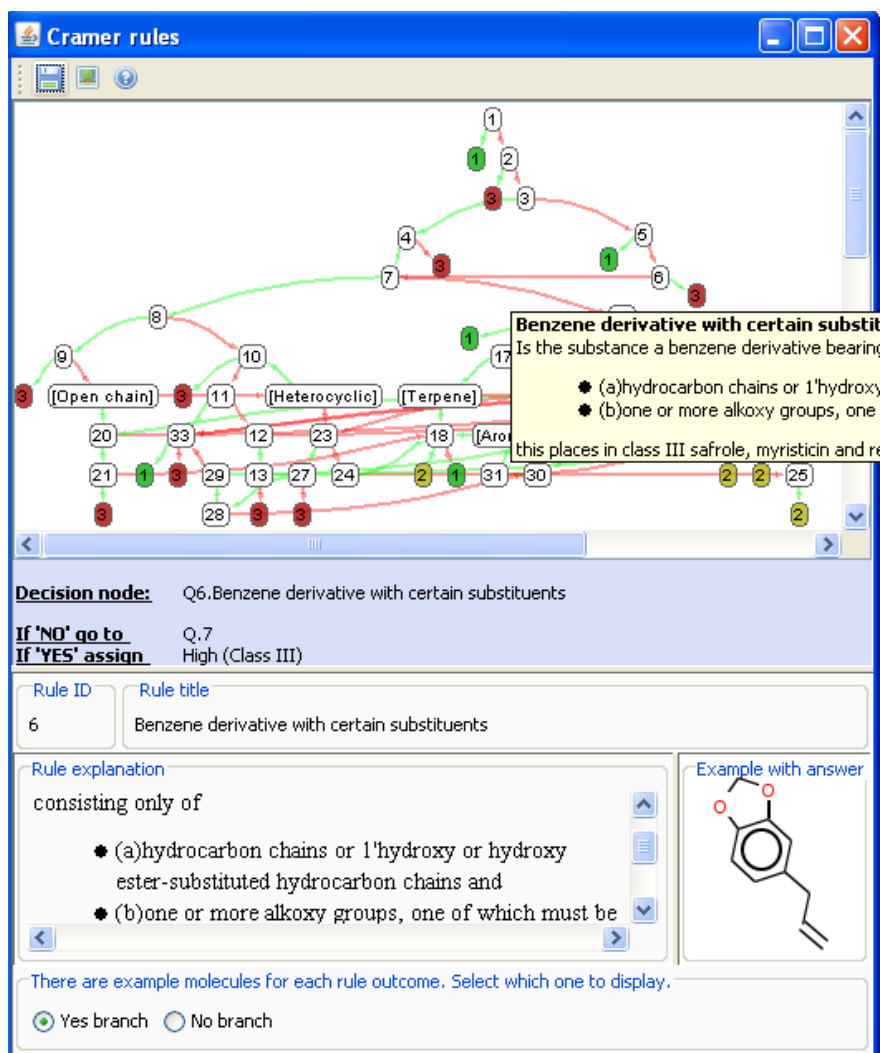


Figure 6: Decision tree visualisation


Exercise 4 . Toxtree online

Open the following link in a web browser

<http://toxtree.sourceforge.net/predict/>

Enter chemical identifier in the text box (e.g. 50-00-0) and press “Query”.

Web edition [Tweet](#) [in](#) [Share](#) [toxtree.sf.net](#) [g+](#) [g](#) [1 circle](#)

 **Toxtree - Toxic Hazard Estimation by decision tree approach**

50-00-0 QUERY

Available structure attributes

CasRN	50-00-0
EC number	200-001-8
IUPAC name	formaldehy
Names	formaldehy
SMILES	C=O[H]C[
Std. InChI	InChI=1S/C
Std. InChI key	WSFSSNU
REACH regis...	30.11.2010

Cramer rules
Cramer rules | Low (Class I)

Extended Cramer rules
Cramer rules... | Low (Class I)

Verhaar scheme for predicting toxic
Verhaar sche... | Class 3 (ur)

Verhaar scheme (modified) for pred
Verhaar sche... | Class 3 (ur)

Protein binding
Alert for SNA... | NO
Alert for Schif... | YES
Alert for Mich... | NO
Alert for Acyl... | NO
Alert for SN2... | NO
No protein bi... | NO

Toxicity prediction modules

Cramer rules Auto [?](#)

Intermediate (Class II)
Low (Class I) ✓
High (Class III)

1. Normal constituent of the body Yes Class Low (Class I)

Extended Cramer rules Auto [?](#)

Intermediate (Class II)
Low (Class I) ✓
High (Class III)

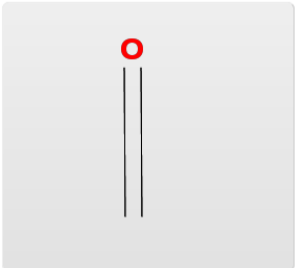
1. Normal constituent of the body Yes Class Low (Class I)

Verhaar scheme for predicting toxicity mode of action Auto [?](#)

Class 5 (Not possible to classify according to these rules)
Class 1 (narcosis or baseline toxicity)
Class 2 (less inert compounds)
Class 3 (unspecific reactivity) ✓
Class 4 (compounds and groups of compounds acting by a specific mechanism)

0.1. Consists only of C,H,N,O,S halogens (excluding I)[S, C, N, O, X] Yes
0.2. Have a logKow between 0 and 6 Yes
0.3. Have a molecular mass (MW) not more than 600 Daltons Yes
1.1. Not contain [I] Yes
1.2. Not contain ionic groups Yes
Q1.3. Contain only C&H No
Q1.4. Contain only C,H and halogen No
1.5. Contain C,H & O[C, O] Yes
Q1.5.1. Be linear ethers or monocyclic mono ethers, but not epoxides or peroxides No
Q1.5.2. Be aliphatic alcohols but not allylic/propargylic alcohols No
Q1.5.3. Be alcohols with aromatic moieties, but NOT phenols or benzylic alcohols No
Q1.5.4. Be ketones, but not alpha-, beta unsaturated ketones (e.g. 1-butenone or acetophenone) No
Q2.1. Be non- or weakly acidic phenols No
Q2.2. Be anilines with one nitro substituent and/or one to three chlorine substituents, and/or alkyl substituents No
Q2.3. Be mononitroaromatics with one or two chlorine substituents and/or alkyl substituents No
Q2.4. Be primary alkyl amines (containing only C,H,N) No

Structure diagram



Click on the “?” button next to Toxicity prediction modules, in order to launch the toxicity estimation.