Volume

2

IDEACONSULT

Toxtree User Manual

IDEACONSULT

Toxtree User Manual

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

Version of 02 April 2012

Table of Contents

Introduction 1
Background 3
Main features at a glance 8
Development tools 9
Launching Toxtree 10
Main screen layout 11
Opening/saving a file 13
Classifying/reading the result 14
Decision tree visualisation 15
Decision tree options visualisation 16
Decision tree selection 18
Decision tree editing 22
Splitting the file into groups 25
Structure diagram editor 27
Batch processing 28
A typical usage scenario 29
Command line options 30
Cramer rules specifics 31
Verhaar scheme specifics 34
Modified Verhaar scheme specifics 36
Skin irritation/corrosion rules specifics 37
Eye irritation/corrosion rules specifics 39
Benigni / Bossa rulebase (for mutagenicity and carcinogenicity) specifics 45
SMARTCyp – Cytochrome P450 –mediated metabolism 50
Decision tree editing - typical tasks 51

Creating a simple decision tree 51
Using a decision tree 62

Introduction

Toxtree is a full-featured and flexible user-friendly open source application, which is able to estimate toxic hazard by applying a decision tree approach. Currently it includes the following plug-ins:

- Cramer rules (Cramer G. M., R. A. Ford, R. L. Hall, Estimation of Toxic Hazard - A Decision Tree Approach, J. Cosmet. Toxicol., Vol.16, pp. 255-276, Pergamon Press, 1978);

- Verhaar scheme for predicting toxicity mode of actions (Verhaar HJM, van Leeuwen CJ and Hermens JLM (1992) Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. Chemosphere 25, 471-491);

- A decision tree for estimating skin irritation and corrosion potential, based on rules published in "The Skin Irritation Corrosion Rules Estimation Tool (SICRET), John D. Walker, Ingrid Gerner, Etje Hulzebos, Kerstin Schlegel, QSAR Comb. Sci. 2005, 24, pp. 378-384";

- A decision tree for estimating eye irritation and corrosion potential, based on rules published in "Assessment of the eye irritating properties of chemicals by applying alternatives to the Draize rabbit eye test: the use of QSARs and in vitro tests for the classification of eye irritation, Ingrid Gerner, Manfred Liebsch & Horst Spielmann, Alternatives to Laboratory Animals, 2005, 33, pp. 215-237";

- A decision tree for estimating carcinogenicity and mutagenicity, based on the rules published in the document: "The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree", by R. Benigni, C. Bossa, N. Jeliazkova, T. Netzeva, and A. Worth. European Commission report EUR 23241 EN[1];

- START (Structural Alerts for Reactivity in Toxtree) biodegradation and persistence plug-in is based on 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

[1] http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_23241_EN.pdf

recognized, the system flags the potential persistence or biodegradability of the chemical. Installation and user manual is available online[2];

- Structure Alerts for the in vivo micronucleus assay in rodents, based on the rules, published in the document "Development of structural alerts for the in vivo micronucleus assay in rodents", by Romualdo Benigni, Cecilia Bossa, Olga Tcheremenskaia and Andrew Worth[3], European Commission report EUR 23844 EN;

- Cramer rules with extensions: This plug-in is a copy of the original plug-in, plus minor extensions. Like the Cramer plug-in, this plug-in works by assigning compounds to Class I, II, or III, according to the rules from Cramer, and some extra ones. Several compounds were classified by Munro in 1996[4] as Class I or Class II compounds according to the Cramer rules, even though Munro reported low NOEL values upon oral administration (indicating relatively high toxicity). To overcome such misclassifications, five rules have been introduced to capture the possible toxicity of these compounds;

- Structure Alerts for identification of Michael Acceptors: This plug-in contains structural alerts, able to identify Michael Acceptors, as defined in T. Wayne Schultz, Jason W. Yarbrough, Robert S. Hunter, Aynur O. Aptula (2007) Verification of the Structural Alerts for Michael Acceptors. Chem. Res. Toxicol. 20, 1359–1363;

- Skin sensitization alerts, as per Enoch SJ, Madden JC, Cronin MT, Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach, SAR QSAR Environ Res. 2008; 19(5-6):555-78;

- SMARTCyp - Cytochrome P450 - Mediated Metabolism, implementation of Patrik Rydberg, David E. Gloriam, Jed Zaretzki, Curt Breneman, Lars Olsen, SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism, ACS Med. Chem. Lett., 2010, 1 (3), pp 96–100;

- Kroes TTC decision tree - Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G. (2004). Structure based thresholds of toxicological

---

[2] http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/Toxtree_start_manual.pdf

[3] http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/doc/EUR_23844_EN.pdf

[4] I.C. Munro, R.A. Ford, E. Kennepohl, and J.G. Sprenger, Correlation of structural class with No-Observed-Effect Levels: A proposal for establishing a threshold of concern, Food Chem. Toxicol.34 (1996), pp. 829–867.

concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65–83

- Modified Verhaar scheme for predicting toxicity mode of actions - rules reordered, according to S.J. Enoch, M. Hewitt, M.T.D. Cronin, S. Azam, J.C. Madden, Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree, Chemosphere 73 (2008) 243-248;

- Structural Alerts for Functional Group Identification ISSFUNC (DOI: 10.2788/33281, Catalogue Number: LB-NA-24871-EN-N), Benigni R., O. Tcheremenskaia, and A. Worth, Computational Characterisation of Chemicals and Datasets in Terms of Organic Functional Groups - a New Toxtree Rulebase;

- Protein binding - S. J. Enoch, C. M. Ellison, T. W. Schultz & M. T. D. Cronin, A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity, Critical Reviews in Toxicology, 2011, 1-20.

Toxtree could be applied to datasets from various compatible file types. User-defined molecular structures are also supported - they could be entered by SMILES, InChI, name, CAS or EINECS or by using the built-in 2D structure diagram editor.

The Toxtree application is suitable for a standalone PC. It has been designed with flexible capabilities for future extensions in mind (e.g. other classification schemes that could be developed at a future date). New decision trees with arbitrary rules can be built with the help of graphical user interface or by developing new plug-ins.

Background

Cramer rules

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

The Cramer classification scheme was the first decision tree scheme, implemented in Toxtree. If it is not possible to reduce the rule to a specific substructure search, an extendable list of compounds is used as the input, for making a decision.

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

Cramer rules with extensions

This plug-in is a copy of the original plug-in, plus minor extensions. Like the Cramer plug-in, this plug-in works by assigning compounds to Class I, II, or III, according to the rules from Cramer, and some extra ones. Several compounds were classified by Munro in 1996 as Class I or Class II compounds according to the Cramer rules, even though Munro reported low NOEL values upon oral administration (indicating relatively high toxicity). To overcome such misclassifications, five rules have been introduced to capture the possible toxicity of these compounds.

This plugin was developed by Curious-IT, The Netherlands, on behalf of JRC.

Verhaar scheme

This plugin implements the Verhaar scheme for predicting toxicity mode of action, according to:

Verhaar H.J.M., Van Leeuven C., Hermens J.L.M., Classifying Environmental Pollutants. 1: Structure-Activity Relationships for Prediction of Aquatic Toxicity, Chemosphere, Vol.25, No.4, pp.471-491, 1992.

This plug-in was developed by Ideaconsult Ltd (Sofia, Bulgaria) within the AMBIT project, funded by CEFIC LRI.

Skin irritation and corrosion prediction

Estimates skin irritation and corrosion potential by physicochemical property limits and structural rules, according to:

1. Ingrid Gerner, Kerstin Schlegel, John D. Walker, and Etje Hulzebosc, Use of Physicochemical Property Limits to Develop Rules for Identifying Chemical Substances with no Skin Irritation or Corrosion Potential, QSAR Comb. Sci. 2004, 23, pp.726-733

2. John D. Walker, Ingrid Gerner, Etje Hulzebos, Kerstin Schlegel, The Skin Irritation Corrosion Rules Estimation Tool (SICRET), QSAR Comb. Sci. 2005, 24, pp.378-384

3. Etje Hulzebos, John D. Walker, Ingrid Gerner, and Kerstin Schlegel, Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential, QSAR Comb. Sci. 2005, 24, pp.332-342

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

Eye irritation and corrosion prediction

Estimates eye irritation and corrosion potential by physicochemical property limits and structural rules, according to:

Ingrid Gerner, Manfred Liebsch & Horst Spielmann, Assessment of the eye irritating properties of chemicals by applying alternatives to the Draize rabbit eye test: the use of QSARs and in vitro tests for the classification of eye irritation, Alternatives to Laboratory Animals, 2005, 33, pp. 215-237.

This plug-in was developed by Ivanka Tsakovska and Nina Jeliazkova on behalf of JRC.

Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)

Estimates potential carcinogenicity and mutagenicity, by using: 1) a series of Structural Alerts (SA); and 2) a number of Quantitative Structure-Activity Relationship (QSAR) models. Details on the alerts and QSARs are provided in the document: "The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree", by R. Benigni, C. Bossa, N. Jeliazkova, T. Netzeva, and A. Worth. European Commission report EUR 23241 EN.

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).

Structure Alerts for the in vivo micronucleus assay in rodents

Provides a list of structural alerts for a preliminary screening of potentially in vivo mutagens. These structural alerts are molecular functional groups or substructures that are known to be linked to the positive in vivo micronucleus assay. Details on the alerts are provided in the document "Development of structural alerts for the in vivo micronucleus assay in rodents", by Romualdo Benigni, Cecilia Bossa, Olga Tcheremenskaia and Andrew Worth (European Commission report EUR 23844 EN). Installation and user manual are included in the Toxtree release.

This plug-in was developed by Instituto Superiore di Sanita (Rome, Italy).

Structure Alerts for identification of Michael Acceptors

Identifies Michael Acceptors by structural alerts, defined according to:

T. Wayne Schultz, Jason W. Yarbrough, Robert S. Hunter, Aynur O. Aptula (2007) Verification of the Structural Alerts for Michael Acceptors.Chem. Res. Toxicol. 20, 1359–1363

This plug-in was developed by Ideaconsult Ltd (Sofia, Bulgaria) within the AMBIT XT project, funded by CEFIC LRI.

Skin sensitisation alerts

Identifies skin sensitisation structural alerts, defined according to:

Enoch SJ, Madden JC, Cronin MT, Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach, SAR QSAR Environ Res. 2008; 19(5-6):555-78.

This plug-in was developed by Ideaconsult Ltd (Sofia, Bulgaria) on behalf of Procter & Gamble.

SMARTCyp

SMARTCyp is a method for prediction of sites in a molecule that are labile for metabolism by Cytochromes P450 isoform 3A4. It is also a reactivity model which is applicable to all P450 isoforms. The Toxtree module is a wrapper around SMARTCyp implementation, available at

http://www.farma.ku.dk/index.php/SMARTCyp/7990/0/

The method is published as:

Cytochrome P450 - Mediated Metabolism, implementation of Patrik Rydberg, David E. Gloriam, Jed Zaretzki, Curt Breneman, Lars Olsen, SMARTCyp: A 2D Method for

Prediction of Cytochrome P450-Mediated Drug Metabolism, ACS Med. Chem. Lett., 2010, 1 (3), pp 96–100.

Kroes TTC decision tree

This Toxtree module is an implementation of the decision tree proposed by ILSI Europe to decide whether substances can be assessed by the TTC approach:

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G. (2004). Structure based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65–83.

It requires user input for daily intake and embeds rules from Cramer Rules module and Benigni/Bossa rules for mutagenicity and carcinogenicity for prediction of genotoxic carcinogens.

Modified Verhaar scheme for predicting toxicity mode of actions

In this modified scheme the rules are reordered (starting from Class 4, instead of Class 1), according to:

S.J. Enoch, M. Hewitt, M.T.D. Cronin, S. Azam, J.C. Madden, Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree, Chemosphere 73 (2008) 243-248.

The statistics of applying the Modified Verhaar scheme plugin against the validation dataset, provided in:

Verhaar et al, Chemosphere Volume 40, Issue 8, April 2000, pages 875-883

are summarised in the following table:

Verhaar scheme (modified)	Class 1	Class 2	Class 3	Class 4	Class 5	Accuracy
Expected Class 1	**40**					100%
Expected Class 2	2	**15**		1	4	68%
Expected Class 3	1		**18**	8	11	47%
Expected Class 4			2	**21**	2	84%

The dataset is available at the following URI:

https://toxtree.svn.sourceforge.net/svnroot/toxtree/trunk/toxtree/toxtree-plugins/toxtree-verhaar2/src/test/resources/toxtree/plugins/verhaar2/Verhaar2000.sdf

Main features at a glance

- Toxtree is a standalone software application, implementing the Cramer decision tree, the Verhaar scheme, a decision tree for skin corrosion and irritation prediction, and the Benigni / Bossa rulebase (for mutagenicity and carcinogenicity). Toxtree runs on Microsoft Windows™ operating system as well as on different platforms with Java™ 2 Runtime Environment, Standard Edition 1.4 or newer installed;

- the chemical structures for analysis may be submitted by using an interactive 2D graphical editor, or in a batch mode by using CSV, TXT or SDF file formats;

- the classification results are displayed in a graphical user interface and can be optionally saved as a file (CSV, SDF or TXT) file, together with the classification path explanation (list of applied rules). This provides a transparent audit of the decision path followed;

- users may modify the tree by adding their own structural rules and save a modified tree for future re-use;

- the software provides a flexible and documented (in the doc/src application subfolder) plug-in interface, allowing the integration of different classification schemes developed at a future date;

- the software is packaged in a self-installing file for Windows™, as well as in a ZIP archive;

- the installation procedure is easy and is supplemented with an installation manual (located in the doc application subfolder);

- an user manual is provided in the doc application subfolder;

- documentation of the source code can be found in the doc/src application subfolder;

- the source code of the application is located in the src application subfolder;

- a comprehensive README file is located in the main application folder.

Development tools

The Toxtree application is implemented in Java™. The basic cheminformatics functionality relies on the open source LGPL licensed Java™ library Chemistry Development Kit (CDK). The Integrated Development Environment (IDE) Eclipse, in conjunction with Apache Maven is the main development tool. Some of the Toxtree capabilities are provided through the following open source libraries:

- JChemPaint– a structure diagram editor;
- Ambit[5] – SMARTS and SMIRKS support
- org.xmlcml – CML support;
- gnujaxp – XML support;
- jgrapht – graph algorithms library;
- apache log4j – application logging;
- javax.vecmath – vector and matrix classes;
- OpenBabel[6] - molecule file conversion and pattern matching;
- smi23d[7] - 3D coordinate generation;
- junit – test suite framework;
- L2fprod[8] – GUI components;
- prefuse[9] – decision tree GUI;

[5] http://ambit.sourceforge.net

[6] http://openbabel.sourceforge.net/wiki/Main_Page

[7] http://www.chembiogrid.org/cheminfo/smi23d/

[8] http://www.l2fprod.com/

[9] http://prefuse.org/

Launching Toxtree

In Windows™ platforms, Toxtree can be launched either by using the "Start" menu (Figure 1), 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").

Figure 1: Launching Toxtree on Windows™ platforms

In all platforms (having Java™ 2 Runtime, Standard Edition 1.6 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

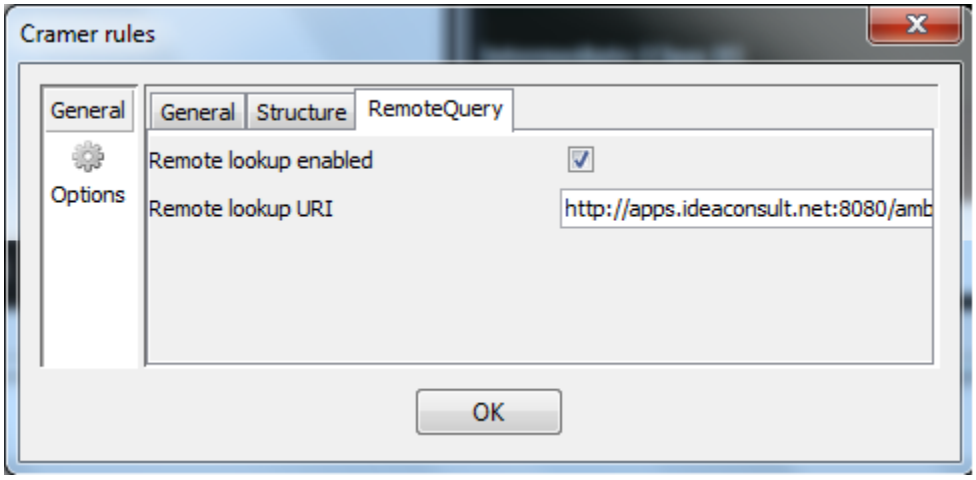
Please, note that in the above mentioned command "java" and "Toxtree-X.Y.Z.jar" should be eventually prefixed with the full path to java and Toxtree on the destination platform.

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 2 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.

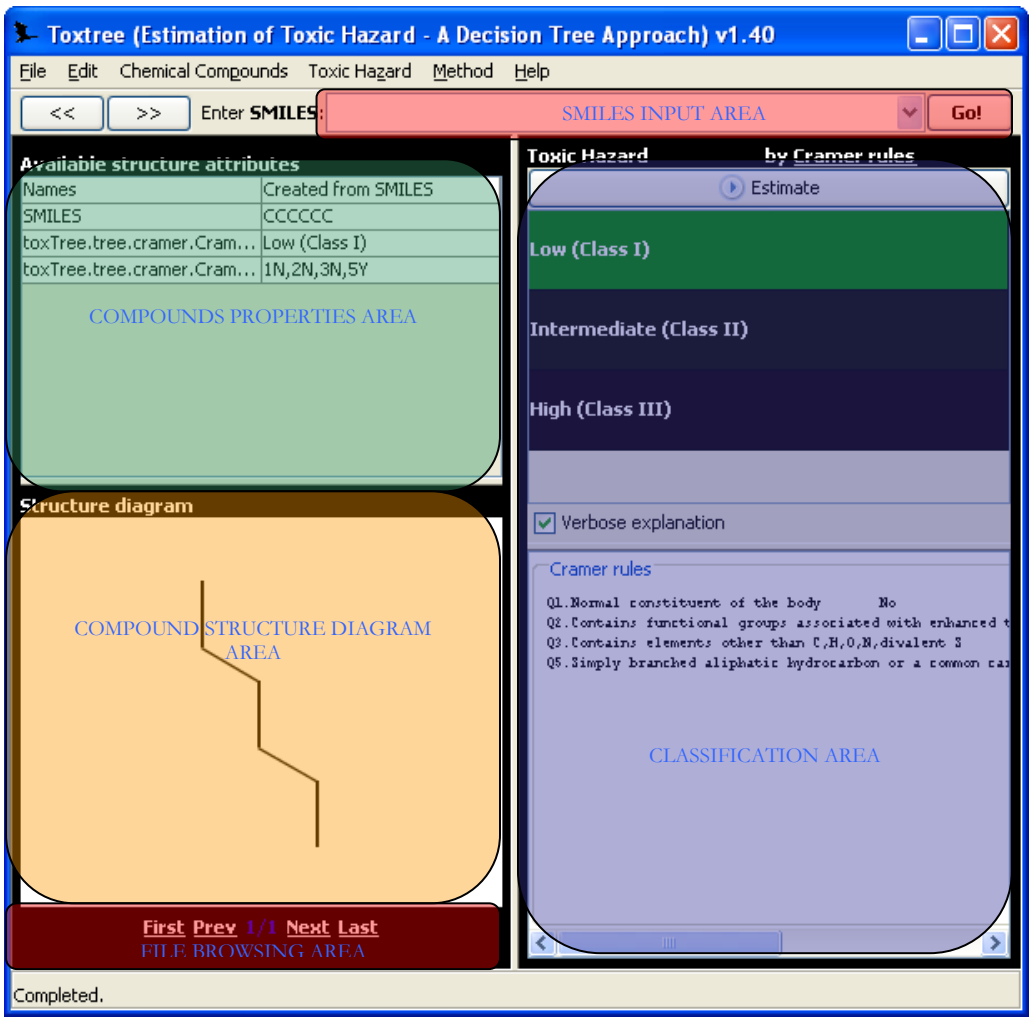


Figure 2: Toxtree main application window

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, as shown on Figure 3. Molecules can be written to SDF, CSV or TXT files, together with their classification data (class & path), by using the "File►Save" menu.

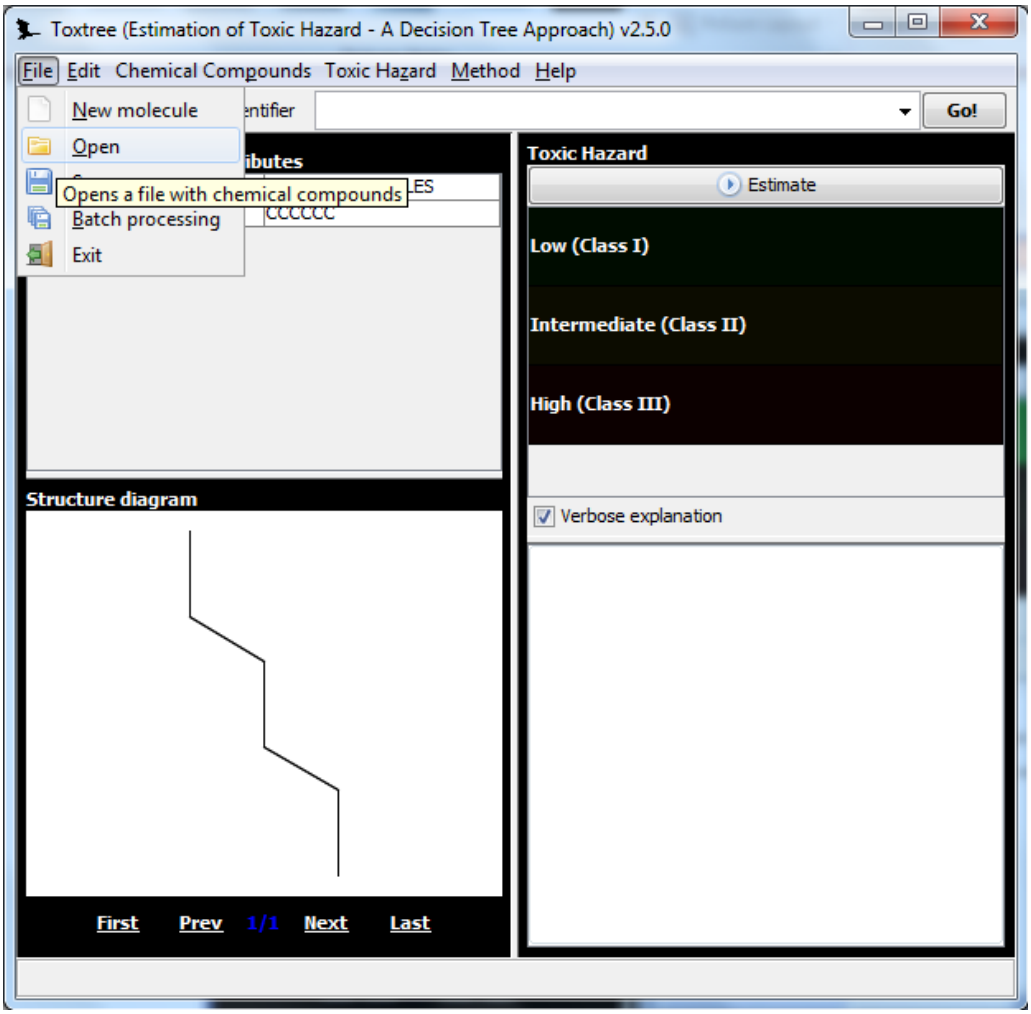


Figure 3: Opening a file with Toxtree

Classifying/reading the result

In order to apply the active decision tree on the current compound (displayed on the left hand side of the main application window), you should press the **Estimate** button, located in the upper part of the **CLASSIFICATION AREA**.

An example of classification result visualisation is show on Figure 4. Hexane is assigned Class I (green highlight), and a verbose text explanation is printed, after pressing the **Estimate** button.

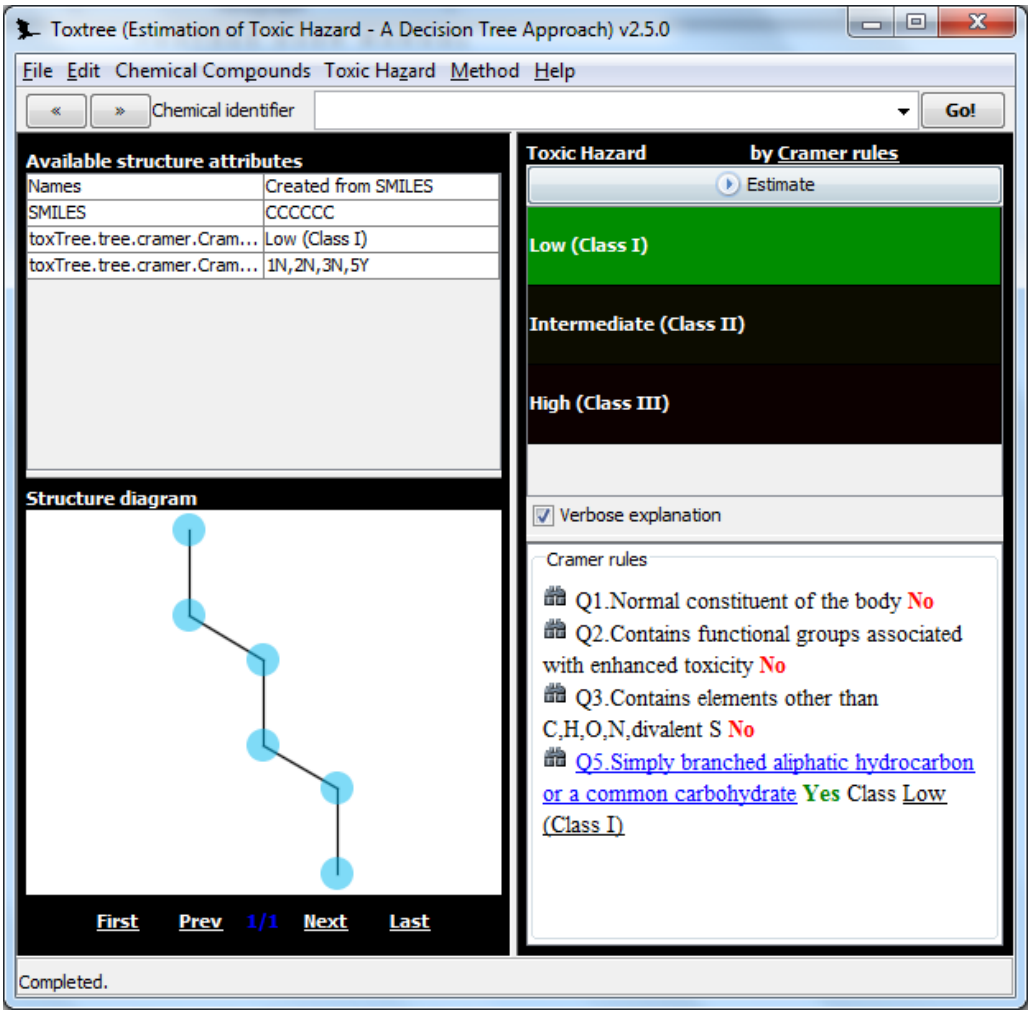


Figure 4: Classification result visualisation

Decision tree visualisation

The current decision tree is accessible through the "Method►View decision tree" menu. It is shown in a child window, as illustrated on Figure 5. 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.

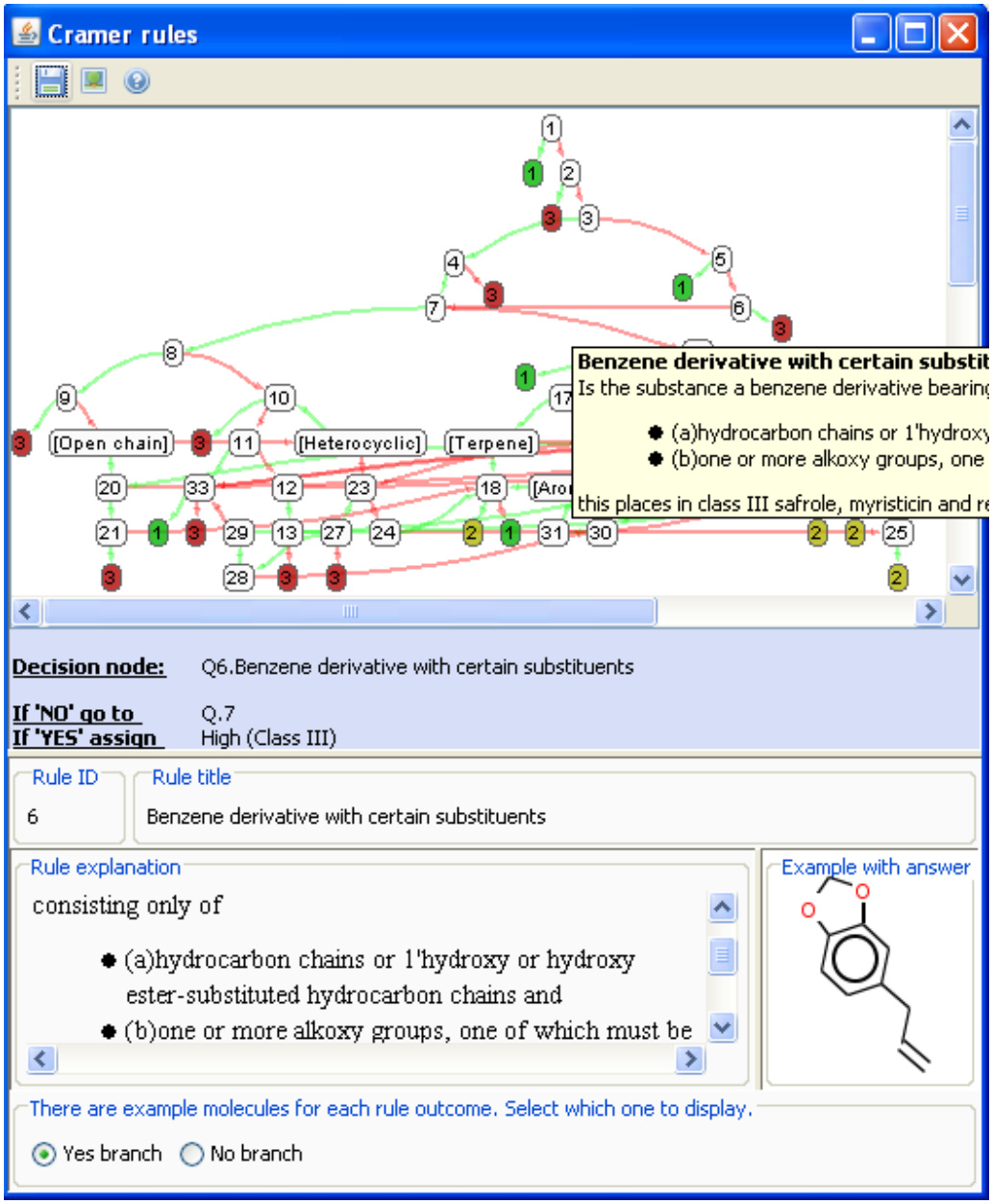


Figure 5: Decision tree visualisation

Decision tree options visualisation

A new menu item, "Method►Decision Tree Options" has been introduced since Toxtree v1.40. The options dialog contains several sections, of which "General" is common for all decision trees, and "Rules" is specific to the currently loaded decision tree, and may be missing, if no rule specific options are available (as for Cramer rules).

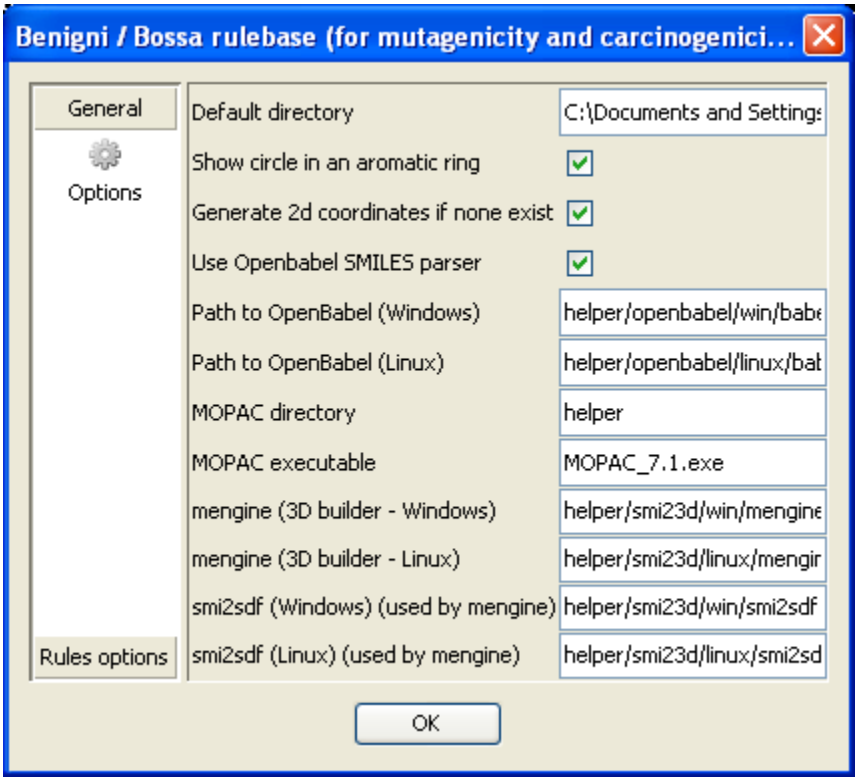


Figure 6: Decision tree options

- **Default directory**: Remembers the directory of the last opened/saved file
  - Default: *Empty*
- **Show circle in an aromatic ring**: Toggles displaying aromatic rings
  - Default: *Checked*
- **Generate 2d coordinates if none exist**: Generate 2D coordinates of the structures, entered as SMILES
  - Default: *Checked*

- **Use Openbabel SMILES parser**: Toggles usage of Openbabel[10] vs. CDK SMILES parser.
    - Default: *Checked*
- **Path to OpenBabel (Windows)**: Specifies the path to Openbabel on Windows platform.
    - Default**:** *helper/openbabel/win/babel.exe*
- **Path to OpenBabel (Linux)**: Specifies the path to OpenBabel on Linux platform.
    - Default**:** *helper/openbabel/linux/babel*
- **MOPAC directory:** Directory where MOPAC resides
    - Default: *helper*
- **MOPAC executable:** Name of the MOPAC executable. Used to calculate electronic descriptors as eHOMO/eLUMO, required by some plug-ins.
    - Default: *MOPAC_7.1.exe*[11]
- **mengine (3D builder - Windows):** MMFF94 force field by mengine[12]. Structures without 3D coordinates are submitted to mengine before running MOPAC. Specifies the path to mengine on Windows platform.
    - Default**:** *helper/smi23d/win/mengine.exe*
- **mengine (3D builder - Linux):** MMFF94 force field by mengine. Structures without 3D coordinates are submitted to mengine before running MOPAC. Specifies the path to mengine on Linux platform.
    - Default**:** *helper/smi23d/linux/mengine*

[10] http://openbabel.org/

[11] OpenMopac, http://openmopac.net/

[12] mengine, http://www.chembiogrid.org/cheminfo/smi23d/

- **smi2sdf (Windows) (used by mengine)**: Generates rough 3D structure[13]. Preparatory step before running mengine. Specifies the path to smi2sdf on Windows platform.
    - **Default:** *helper/smi23d/win/smi2sdf.exe*
- **smi2sdf (Linux) (used by mengine)**: Generates rough 3D structure. Preparatory step before running mengine. Specifies the path to smi2sdf on Windows platform.
    - **Default:** *helper/linux/win/smi2sdf.exe*

These general options are automatically saved in the toxtree.pref file, located in the same directory as *Toxtree-X.YZ.jar*.

Decision tree selection

The "Method►Select decision tree" menu can be used in order to change the active decision tree, as shown on Figure 7. The **Load from file** button enables users to select a different decision tree, which was written from scratch or by editing an existing decision tree.

[13] smi2sdf, http://www.chembiogrid.org/cheminfo/smi23d/

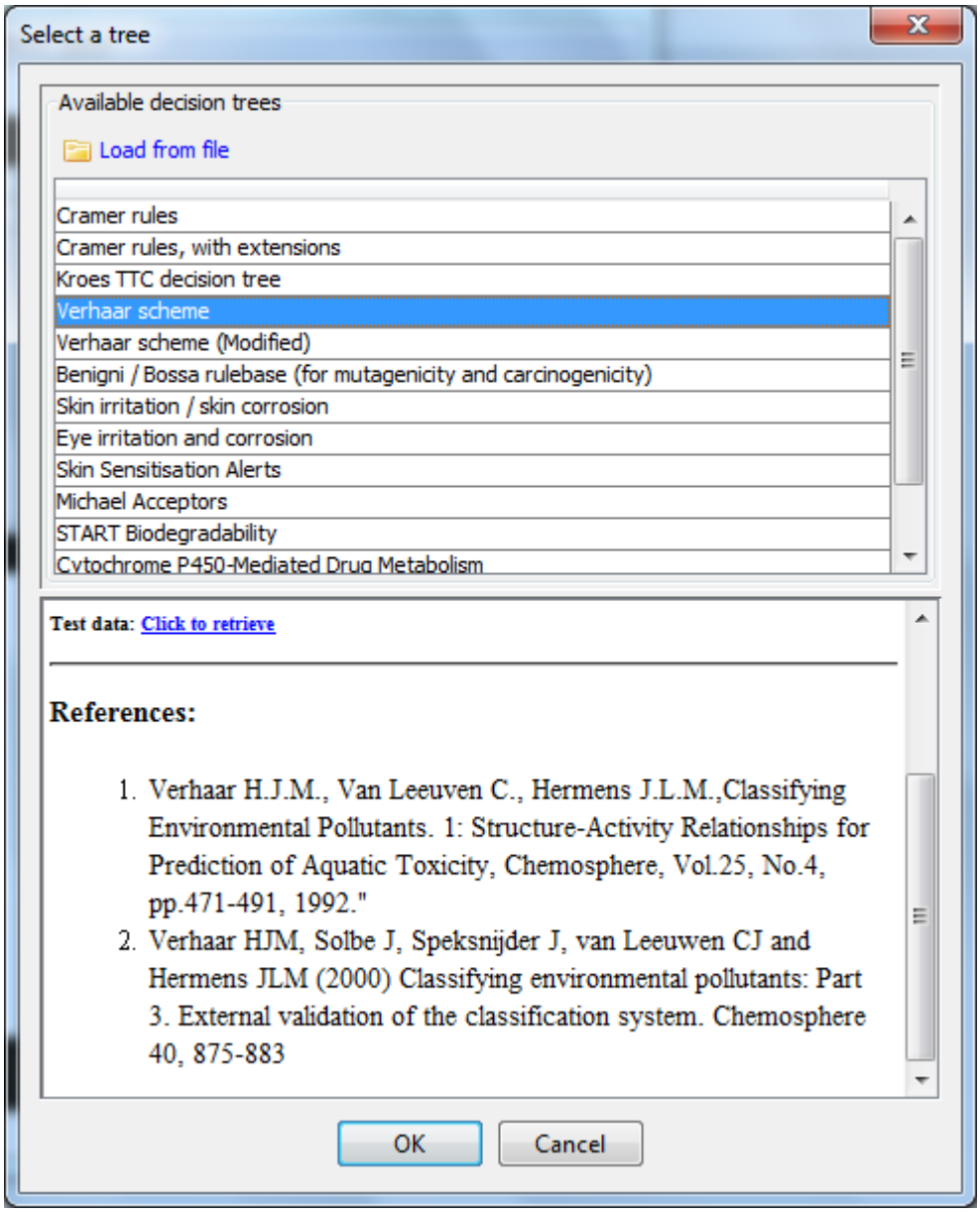


Figure 7: Decision tree selection

When the "Verhaar scheme" option is selected, the Verhaar scheme for predicting toxicity mode of actions is loaded. The tree consists of 5 classes, as shown on Figure 8.

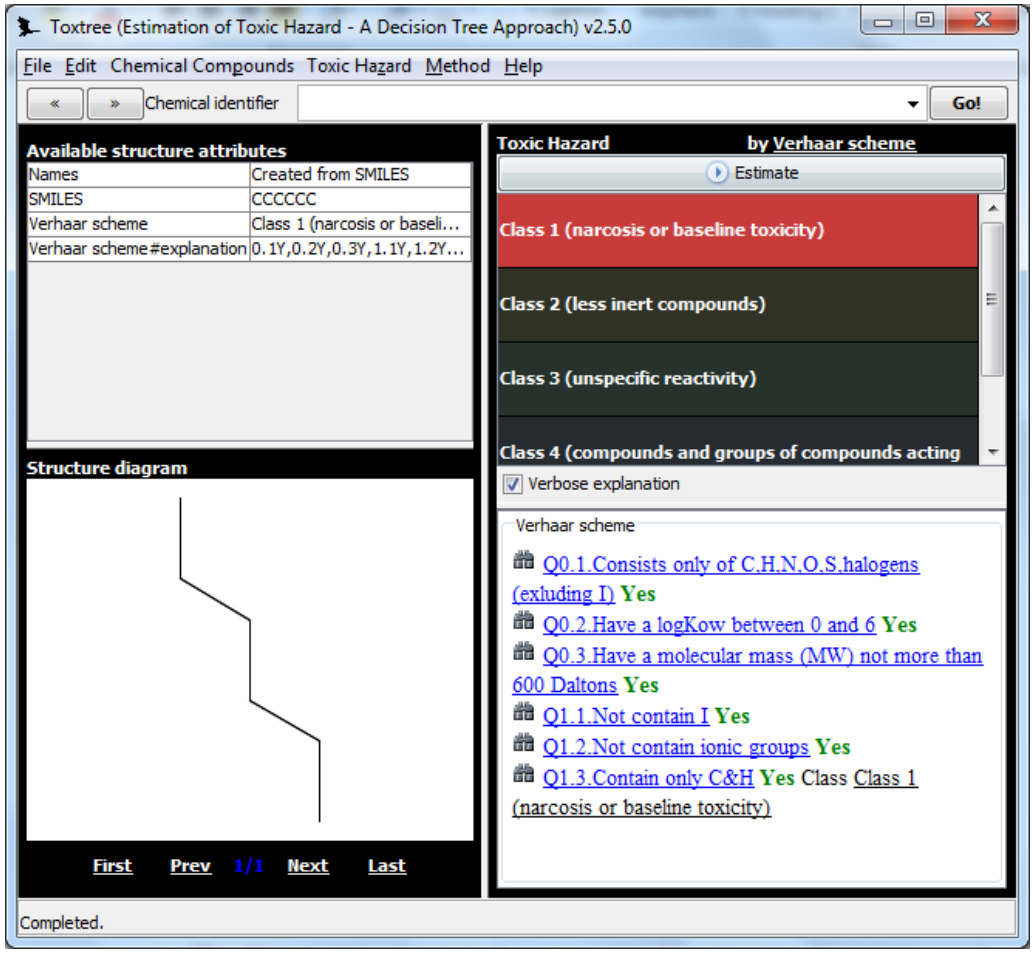


Figure 8: Verhaar scheme

All of the functionality explained for the "Cramer rules" decision tree is valid for any other tree selected.

The "Skin irritation/corrosion" decision tree for estimating skin irritation and corrosion potential by physicochemical property limits and structural rules can be loaded by selecting the corresponding line from the decision tree selection dialog (Figure 7). The result of this operation is displayed in Figure 9.

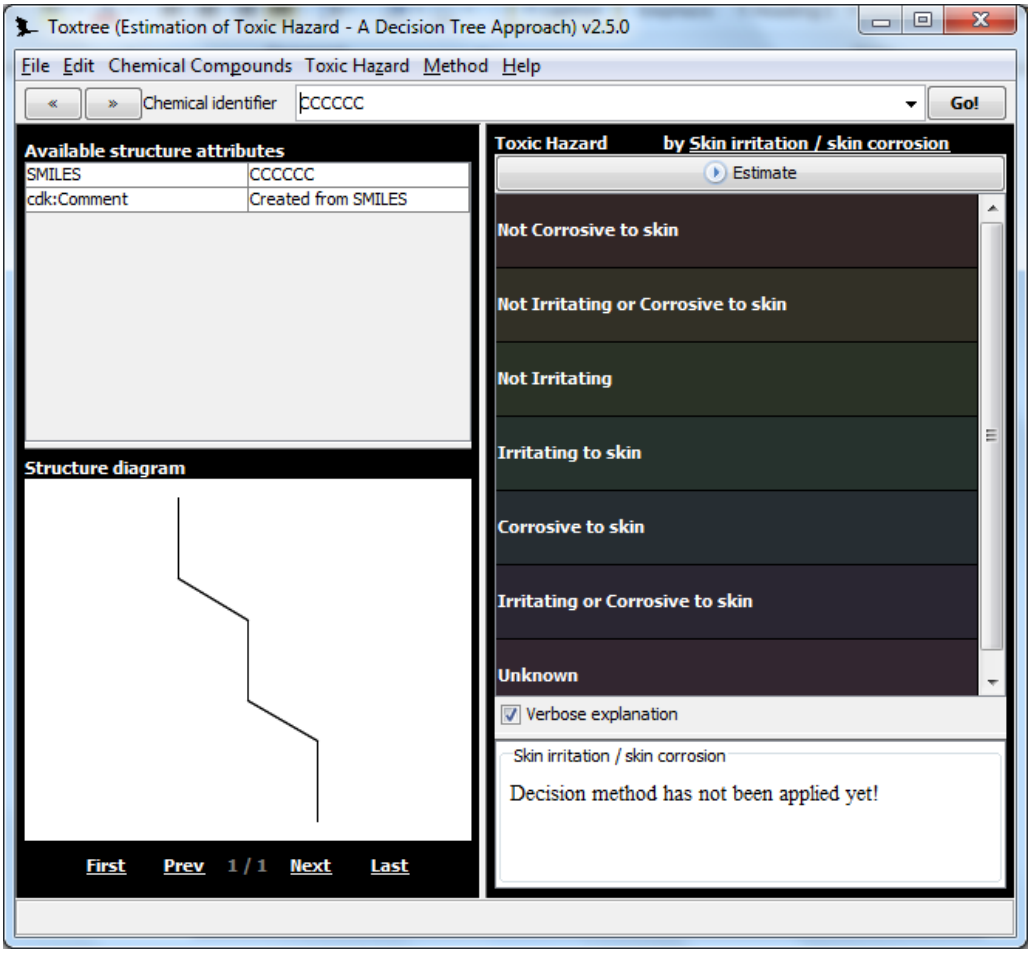


Figure 9: Skin irritation prediction

For more details about skin irritation prediction, please consult the **Skin irritation/corrosion rules specifics** section of the manual.

The "Eye irritation and corrosion" decision tree for estimating eye irritation and corrosion potential by physicochemical property limits and structural rules can be loaded by selecting the corresponding line from the decision tree selection dialog (Figure 7). For more details about eye irritation prediction, please consult the **Skin irritation/corrosion rules specifics** section of the manual.

Selecting the "Benigni / Bossa rulebase (for mutagenicity and carcinogenicity)" option loads the corresponding decision tree as displayed in Figure 10.

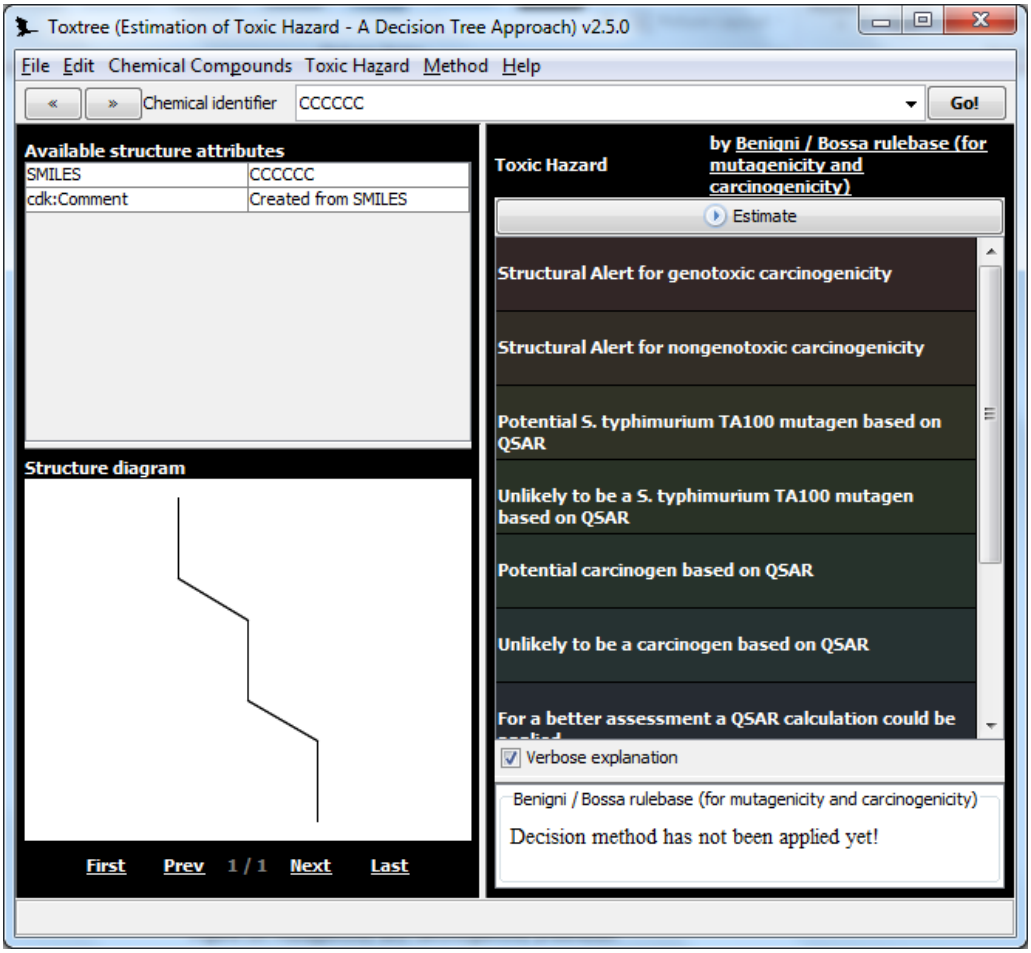


Figure 10: Mutagenicity and carcinogenicity prediction

For more details about mutagenicity and carcinogenicity prediction, please consult the **Benigni / Bossa rulebase (for mutagenicity and carcinogenicity) specifics** section of the manual.

Decision tree editing

The "Method►Edit decision tree" menu can be used in order to edit an existing decision tree or construct a new one from scratch (Figure 11). It provides access to the following 3 submenus:

- the "New decision tree" submenu can be used for constructing a new decision tree from scratch;
- the "Select from list" submenu provides a list of known decision trees, which can be loaded and subsequently edited in the "Decision tree editor";

- the "Load from file" submenu can be used for loading a user-supplied decision tree in the "Decision tree editor".

A copy of the chosen decision tree is loaded in memory for editing and can be subsequently saved for future use.

> *You should always try to give a suitable (self-explaining) name, when saving newly constructed from scratch or modified decision trees.*

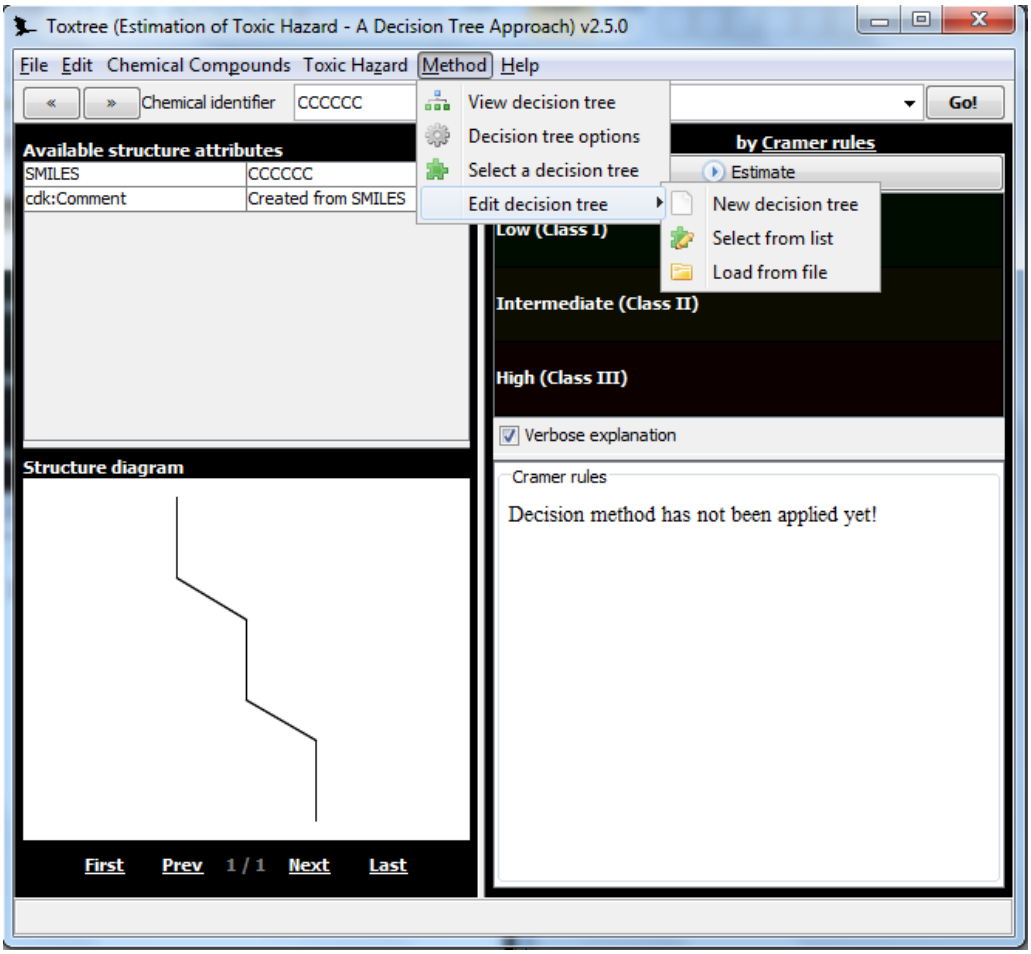


Figure 11: Edit decision tree submenus

The "Decision tree editor" is loaded in a separate (child) window (Figure 12) and provides tools for decision tree editing, as suggested by its title.

The Decision tree editor comprises several areas:

- **Decision tree area** – placing the mouse on a tree node and right-clicking on it provides access to a context menu. This menu can be used for rule editing, as

well as for modifying the "Yes" and/or "No" branches coming out of the current node (highlighted in orange);

- **Rules area** – provides means to add new (supported by Toxtree itself or by an added plug-in) rules, as well as to edit existing rules;
- **Categories area** – add/remove supported categories (classes);
- **Decision node area** – displays details about the currently selected node from the **Decision tree area**. Left-clicking on a node in the **Decision tree area** changes the current node.

Before exiting from the Decision tree editor, users should save the edited decision tree by using the "File▶Save" menu on the upper left corner of the main Decision tree editor window. A reminder is displayed if the user tries to exit the Decision tree editor without having saved his work.

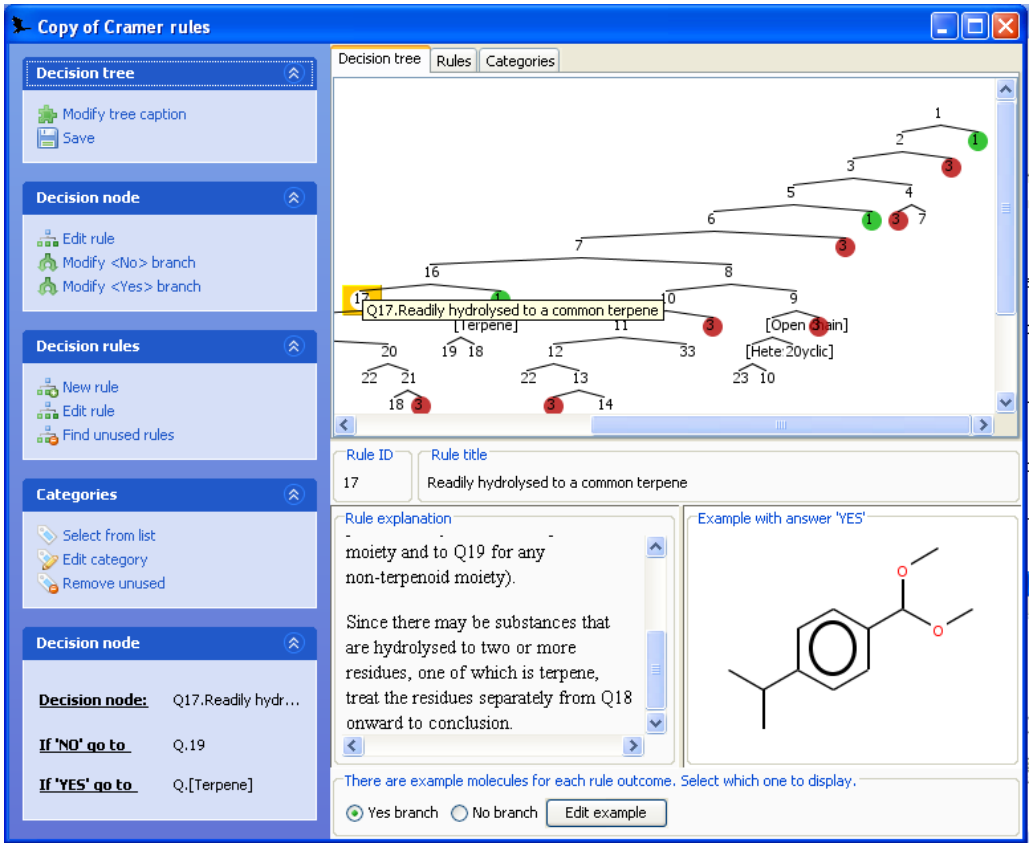


Figure 12: Decision tree editor on a copy of Cramer rules

Detailed instructions how to create and edit a decision tree are provided in section "**Decision tree editing - typical tasks**".

Splitting the file into groups

Toxtree allows to split the file into subsets, defined by decision tree categories. For this purpose:

1. Load a file with chemical compounds (the following examples use the DSSTox EPA Fathead Minnow file EPAFHM_v3b_617_10Apr2006.sdf, downloaded from http://www.epa.gov/ncct/dsstox/sdf_epafhm.html).

2. Run "Toxic Hazard►Estimate All" to apply Cramer rules for all compounds.

3. Click on "Chemical Compounds►Subsets" to display the subsets selection dialog.

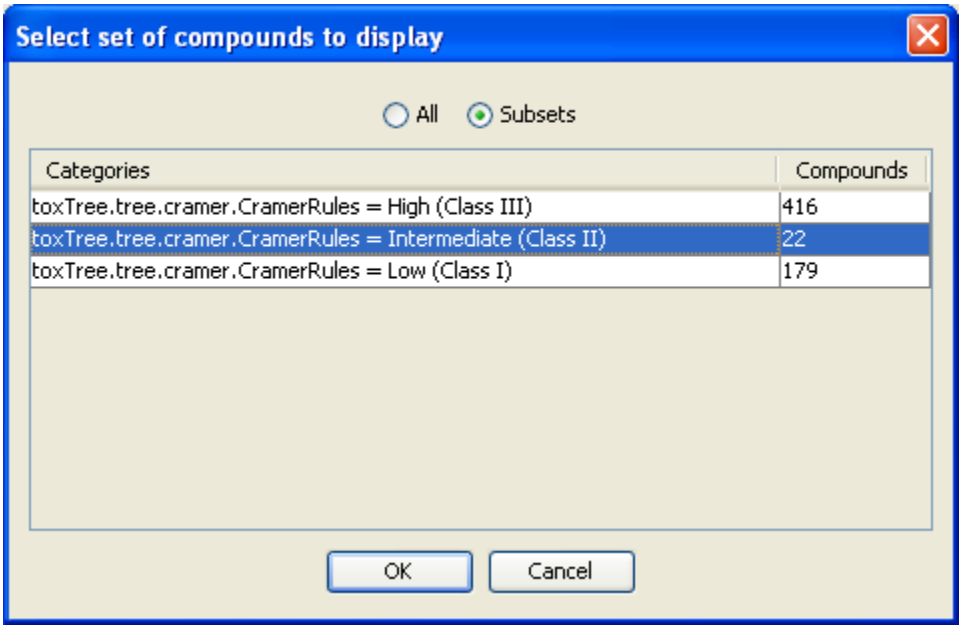


Figure 13: Subsets

The subsets selection dialog (Figure 13) displays three subsets, corresponding to Cramer toxicity classes, and the number of compounds in each subset. Select the second row (Intermediate (Class II)) and click OK. The subset of 22 compounds will be loaded into Toxtree main window (Figure 14).

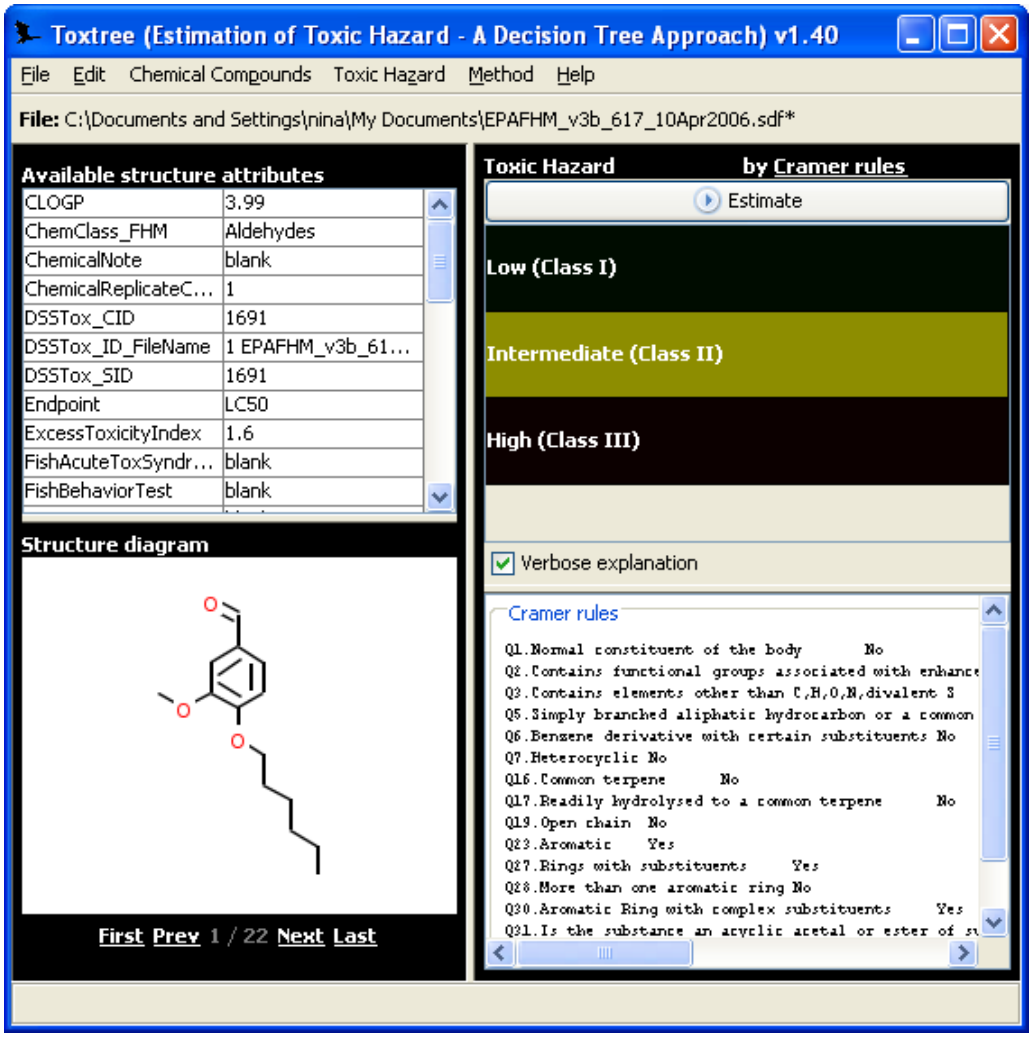


Figure 14: 22 compounds classified as Cramer class II

From this point on, all actions (e.g. File Save, Estimate All) are applied on the subset loaded in the main window. Use the same menu to select another subset or to return to the entire file. For the latest, select "All" options and then click OK (Figure 15).

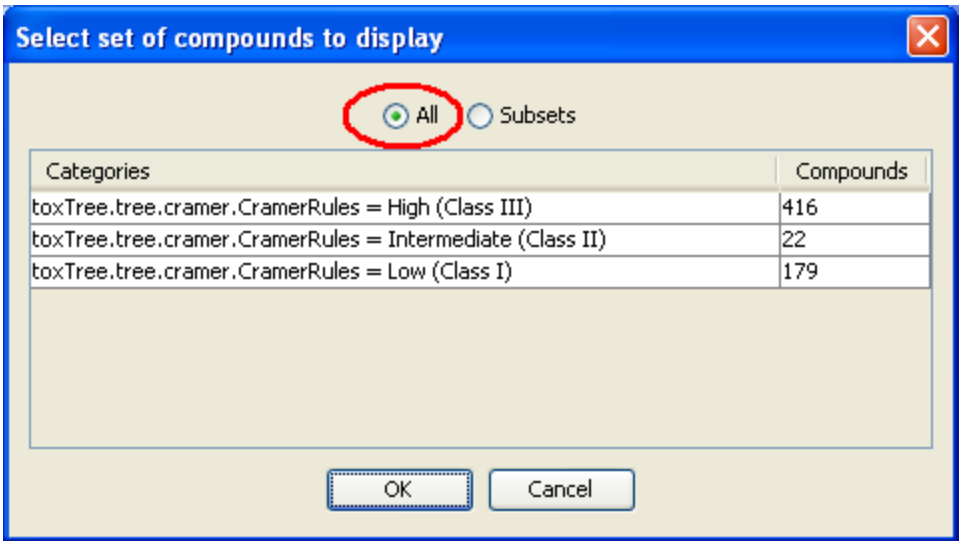


Figure 15: Select "All" option to load the entire file into the main Toxtree window

Structure diagram editor

A 2D structure diagram editor is integrated in Toxtree. It can be accessed through the "Chemical Compounds►Edit compound" menu (Figure 16).

The structure diagram editor provides a convenient way to edit the current compound in Toxtree.

If needed, users could consult the structure diagram editor help and/or tutorial, accessible through the "Help" menu of the editor.

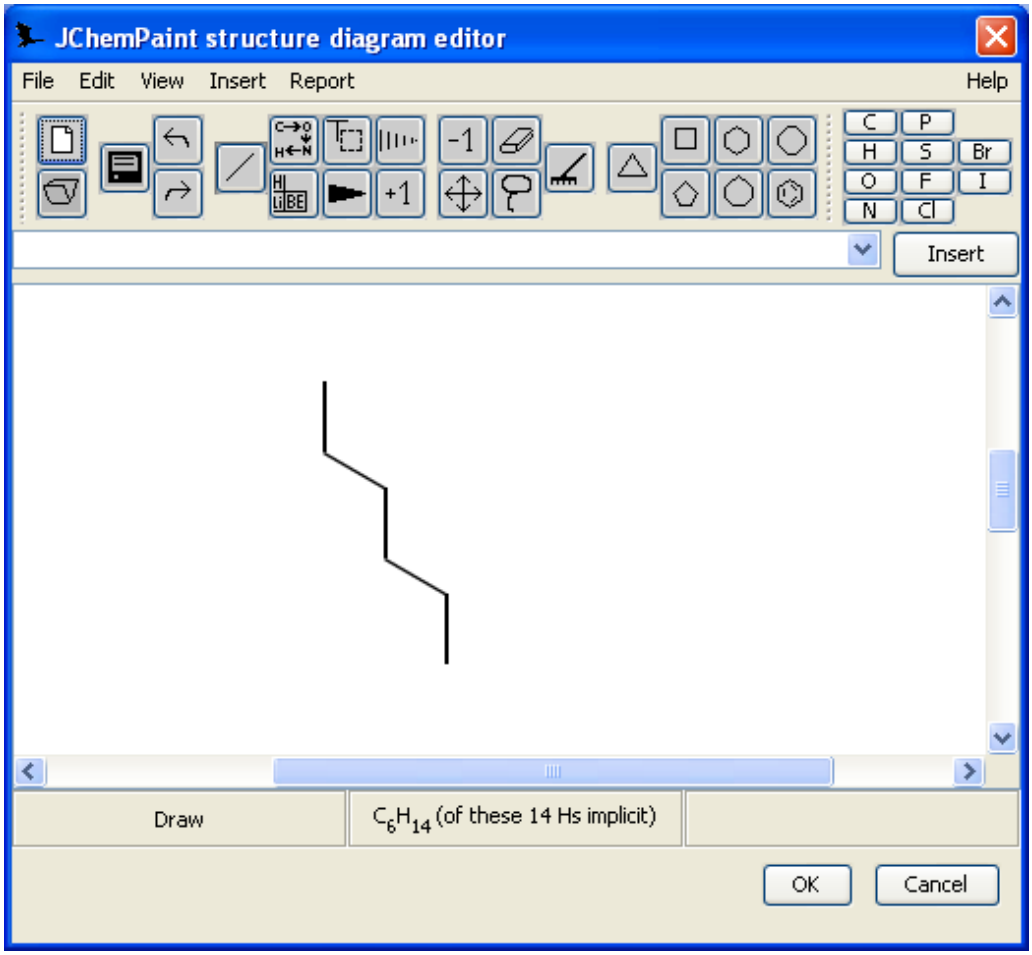


Figure 16: 2D structure diagram editor

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 or aborted upon user request. Batch configuration/state can be saved and loaded at a later time (Figure 17). 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).

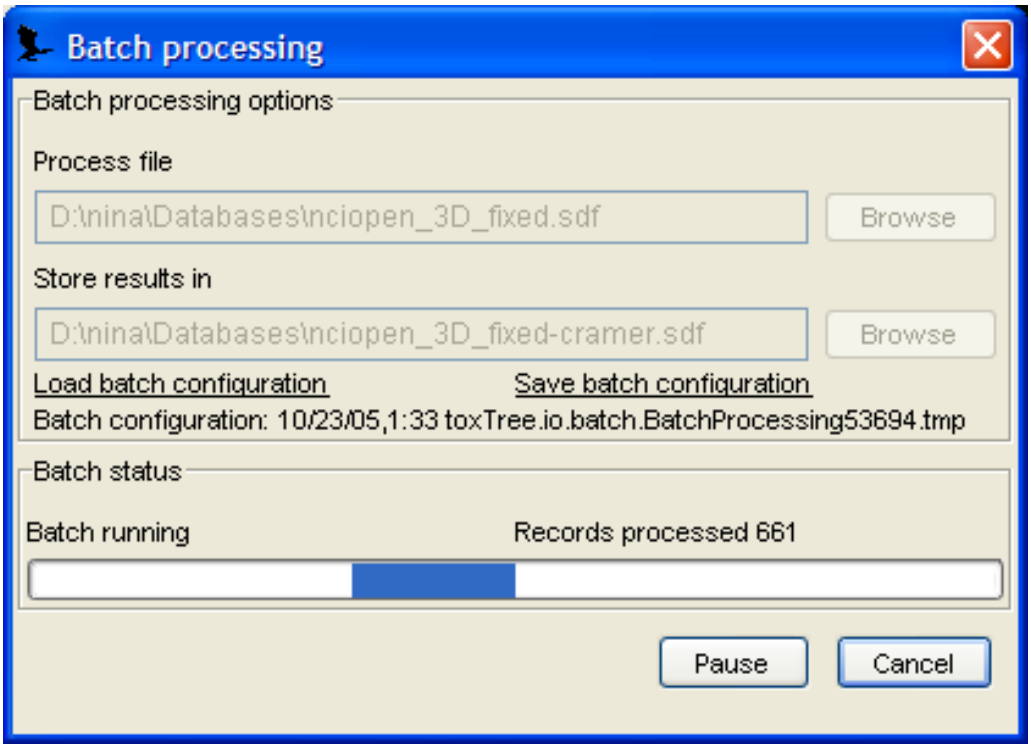


Figure 17: Batch processing

A typical usage scenario

Typical use of Toxtree could follow the following scenario:

1. launch the application;

2. enter a molecular structure in the SMILES field (or open a file, containing molecular structures - compatible file types are: CML, CSV, HIN, ICHI, INCHI, MDL MOL, MDL SDF, MOL2, PDB, SMI, TXT and XYZ). Please, note that CSV files can be read/written by MS Excel™;

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 (either low - class I, intermediate - class II or high - class III);

5. read the classification explanation (which Cramer rules have been applied);

6. go through the list of structures, loaded from a compatible file type (CML, CSV, HIN, ICHI, INCHI, MDL MOL, MDL SDF, MOL2, PDB, SMI, TXT and XYZ) by using the navigation links at the bottom of the left side of the application window (note: CSV files can be read/written by MS Excel™);

7. repeat steps from (3) to (5) for any chosen (in step (6)) molecular structure;

8. consult the decision tree (accessible through the Toxtree "Method►View decision tree" menu);

9. classify all the molecules, loaded from a file, by using the batch processing facility;

10. save the processed molecules, 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™;

11. create (or edit existing) decision trees through the Toxtree "Method►Edit decision tree" menu;

12. change the decision tree used in the estimation through the Toxtree "Method►Select decision tree" menu;

13. modify the current molecule by using the integrated structure diagram editor through the "Chemical Compounds►Edit compound" menu.

Command line options

Toxtree supports the following command line options:

- **-b <file name>** specifies that <file name> should be opened for batch processing;

- **-DtoxTree.debugging=true** turns on verbose console output (useful for debugging purposes, as well as for detailed study of decision tree results), default value - off;

- **-f <file name>** specifies that <file name> should be opened for browsing.

The verbose console output could be redirected to a log file, by using the following command:

java -DtoxTree.debugging=true -jar Toxtree-X.YZ.jar > Toxtree.log

Cramer rules specifics

Lists of compounds

Cramer rules #1 and #22 depend explicitly on user-defined lists of compounds, which are normal constituents of the body or common components of food. We provide example lists of such compounds in the files bodymol.sdf and foodmol.sdf respectively. If the files are removed from the application directory, than a message stating that Q1 and/or Q22 are not implemented is displayed. This is normal behaviour and a reminder, that Q1 and Q22 cannot be implemented by structural rules.

You can check if the files bodymol.sdf and foodmol.sdf are found by Toxtree by using the "Help►Files info" menu (Figure 18).

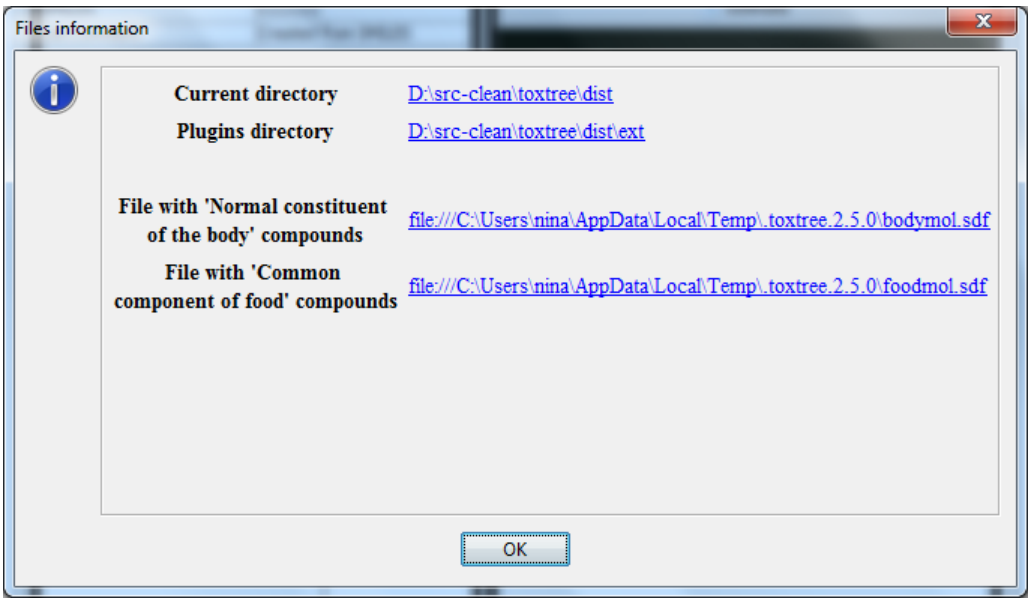


Figure 18: Files information (bodymol.sdf & foodmol.sdf found and used by Toxtree)

The bodymol.sdf and foodmol.sdf files are distributed embedded in Toxtree-{version}.jar, and extracted automatically into a temporary directory, as shown above.

> *Please, note that the bodymol.sdf and foodmol.sdf files are provided mainly as an example. They contain currently only a very limited number of "Normal constituents of the body" and "Common components of food" respectively, following an expert advice. Users should consider expanding these files with appropriate molecules.*

Hydrolysis/metabolic reactions

A limited number of hydrolysis (Cramer rules #15, #17, #29, #30 and #31) and metabolic (Cramer rule #33) reactions are implemented, based on an expert advice. The reactions are stored and read as CML files.

The implemented six hydrolysis reactions are shown on the following figures.

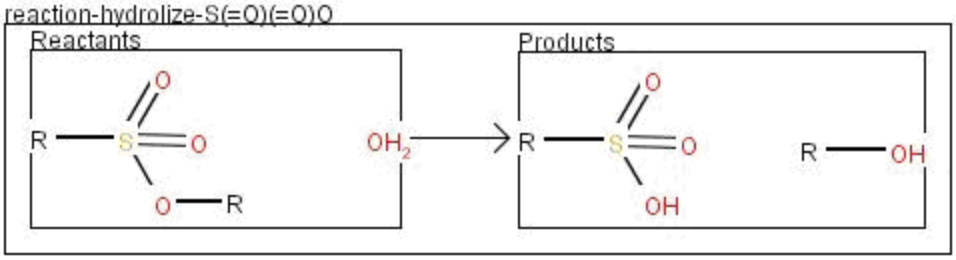


Figure 19: Reaction-hydrolize-S(=O)(=O)O

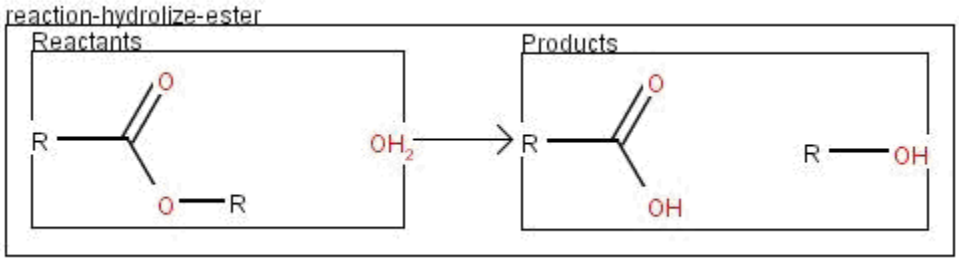


Figure 20: Reaction-hydrolize-ester

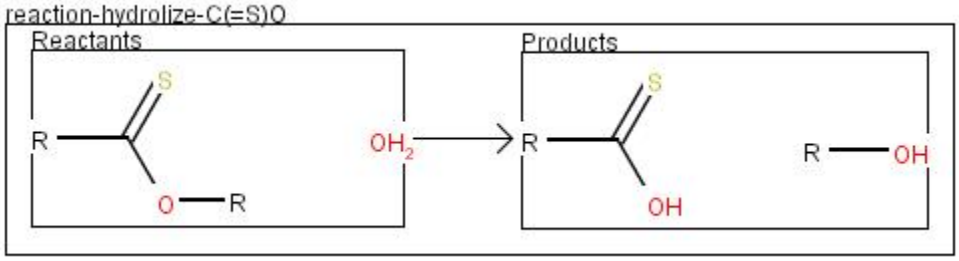


Figure 21: Reaction-hydrolize-C(=S)O

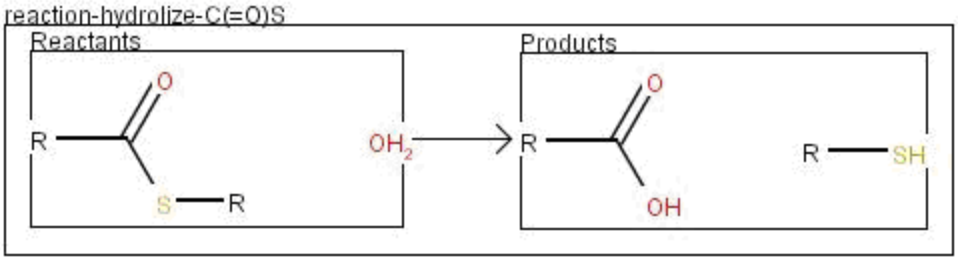


Figure 22: Reaction-hydrolize-C(=O)S

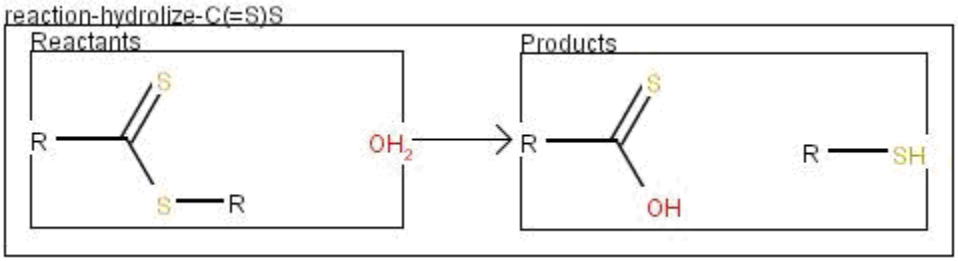


Figure 23: Reaction-hydrolize-C(=S)S

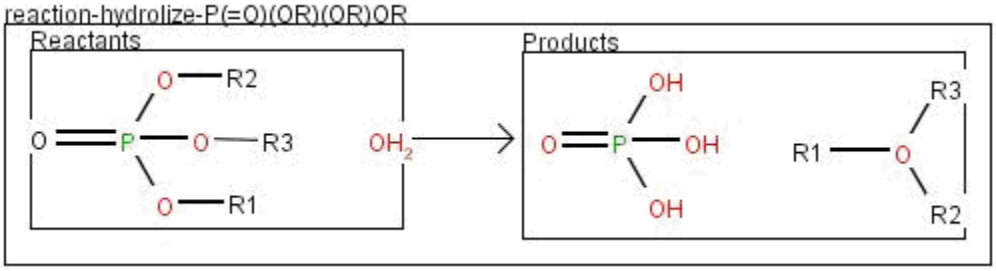


Figure 24: Reaction-hydrolize-P(=O)(OR)(OR)OR

The implemented four metabolic reactions are show on the following figures.

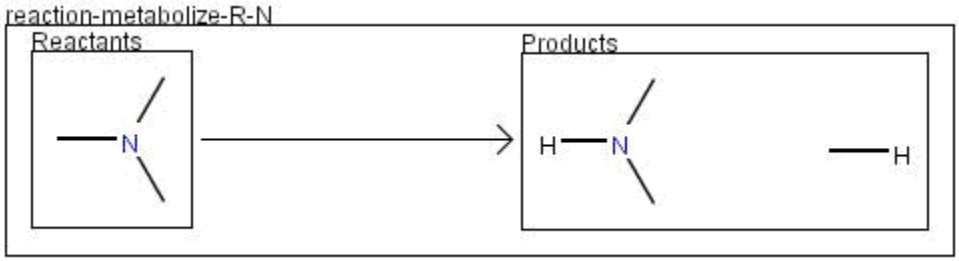


Figure 25: Reaction-metabolize-R-N

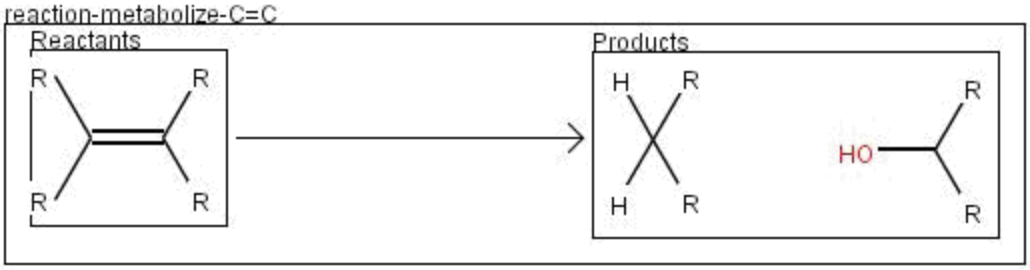


Figure 26: Reaction-metabolize-C=C

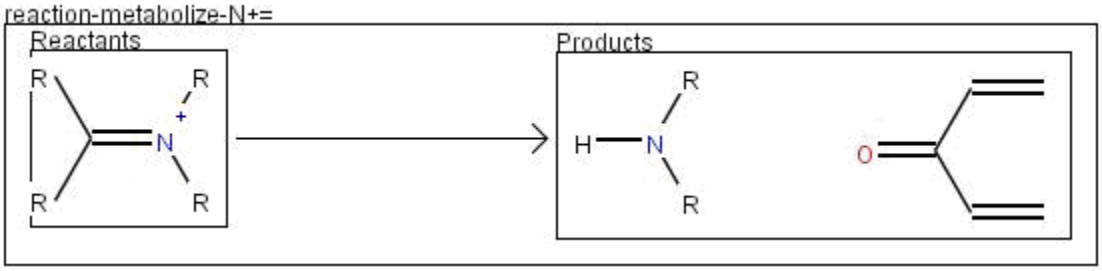


Figure 27: Reaction-metabolize-N+=

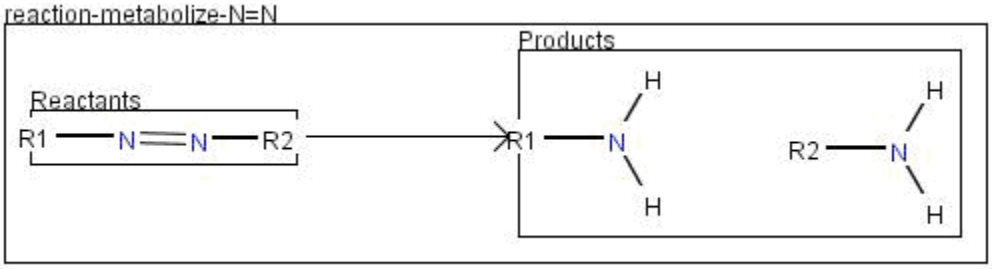


Figure 28: Reaction-metabolize-N=N

Verhaar scheme specifics

The Verhaar scheme classifies compounds in the following categories:

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)

Class 5 (Not possible to classify according to these rules)

The second rule verifies whether the LogP of a compound is within the [0,6] range. LogP is calculated on the fly through the XlogP procedure[14,15], implemented by the CDK library. The implementation was validated by comparison with other LogP implementations[16].

The Verhaar scheme does not define structural rules for Class 4 compounds, but only examples. The implementation was extended to cover as much as possible of the groups of compounds, defined as Class 4.

The implementation of the original Verhaar scheme was considerably improved in Toxtree 2.5.0. The statistics of applying the Verhaar scheme plugin against the validation dataset, provided in:

Verhaar et al, Chemosphere Volume 40, Issue 8, April 2000, pages 875-883

are summarised in the following table:

Verhaar scheme	Class 1	Class 2	Class 3	Class 4	Class 5	Accuracy
Expected Class 1	**40**					100%
Expected Class 2	2	**16**			4	72%
Expected Class 3	6		**16**	5	11	42%
Expected Class 4	3		7	**13**	2	52%

The dataset is available at the following URI:

[14] R. Wang, Y. Fu, and L. Lai. A New Atom-Additive Method for Calculating Partition Coefficients. J. Chem. Inf. Comput. Sci., 37:615–621, 1997.

[15] R. Wang, Y. Gao, and L. Lai. Calculating partition coefficient by atom-additive method. Perspectives in Drug Discovery and Design, 19:47–66, 2000.

[16] Uli Fechner, Kristina Grabowski, QA of the XlogP Descriptor, CDK News - The Newsletter of CDK project, Volume 3/1, March 2004, ISBN 1614-7553, pp. 12-14, http://sourceforge.net/projects/cdk/files/CDK%20News/3_1/cdknews3.1.pdf/download

https://toxtree.svn.sourceforge.net/svnroot/toxtree/trunk/toxtree/toxtree-plugins/toxtree-verhaar2/src/test/resources/toxtree/plugins/verhaar2/Verhaar2000.sdf

Modified Verhaar scheme specifics

This modified scheme differs by reordering the rules, according to:

S.J. Enoch, M. Hewitt, M.T.D. Cronin, S. Azam, J.C. Madden, Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree, Chemosphere 73 (2008) 243-248

The statistics of applying the Modified Verhaar scheme plugin against the validation dataset, provided in:

Verhaar et al, Chemosphere Volume 40, Issue 8, April 2000, pages 875-883

are summarised in the following table:

Verhaar scheme (modified)	Class 1	Class 2	Class 3	Class 4	Class 5	Accuracy
Expected Class 1	**40**					100%
Expected Class 2	2	**15**		1	4	68%
Expected Class 3	1		**18**	8	11	47%
Expected Class 4			2	**21**	2	84%

The dataset is available at the following URI:

https://toxtree.svn.sourceforge.net/svnroot/toxtree/trunk/toxtree/toxtree-plugins/toxtree-verhaar2/src/test/resources/toxtree/plugins/verhaar2/Verhaar2000.sdf

Skin irritation/corrosion rules specifics

The skin irritation/corrosion rules classify compounds into following categories:

- **Not Corrosive**
- **Not Irritating Or Corrosive**
- **Not Irritating**
- **Irritating**
- **Corrosive**
- **Irritating Or Corrosive**
- **Unknown**

The classification is done by physicochemical property limits and structural rules.

Physicochemical properties

The decision rule depends on the following physicochemical properties: molecular weight, LogP, melting point, water solubility, lipid solubility and surface tension.

The Toxtree software calculates on the fly only LogP and Molecular weight, for the rest it expects the values to be read from the file or manually entered by the user.

The file should contain properties with exactly the following names (column names in CSV file of SDF properties in SDF file):

"Vapour Pressure"

"Water Solubility"

"Lipid Solubility"

"Melting Point"

"Surface Tension"

If fields with exactly these names are missing from the file, the software shows a dialog, asking properties to be entered manually for each compound (Figure 29).

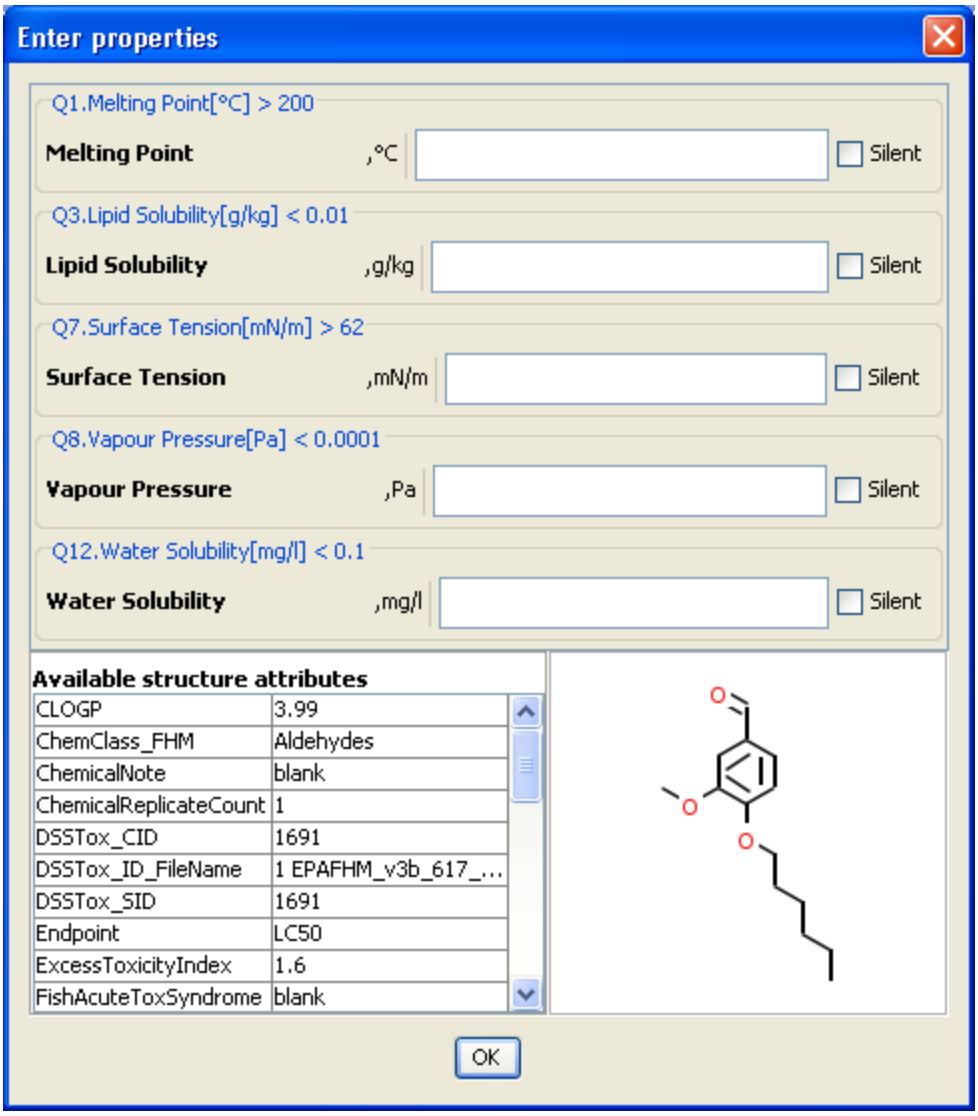


Figure 29: Skin irritation prediction - options

If the values are unknown, you might select the checkbox "Silent". In this case, the result of a silent rule asking for missing properties will always be "No" and therefore the left branch of the tree will be followed. This effectively means that physicochemical rules will be skipped and only structural rules will be applied. Please note that this may result in a low quality prediction.

Eye irritation/corrosion rules specifics

The eye irritation/corrosion caused by a chemical is characterized using the following EU risk phrases:

R36: Irritating to eyes (moderate eye irritation reversible within ≤ 21 days)

R41: Risk of serious damage to eyes (moderate but persistent eye lesions, eye corrosion)

According to international risk assessment guidelines, skin corrosion potential excludes further considerations on a similar hazardous potential to eyes, since the chemical has already proved to have corrosive properties. Thus the following effects induced by local contact to skin are also assumed to be predictive of eye damage:

R34: Causes burns (skin corrosion caused by a 4-hour skin contact)

R35: Causes severe burns (skin corrosion caused by a 3-minute skin contact)

Based on this, the eye irritation/corrosion rules implemented in Toxtree classify compounds into the following categories:

- NOT skin corrosion R34 or R35 — category 1
- NOT lesions R34, R35, R36 or R41 — category 2
- NOT eye irritation R41 — category 3
- NOT eye irritation R36 — category 4
- NOT corrosion R34, R35 or R41 — category 5
- NOT lesions R34, R35 or R36 — category 6
- NOT eye irritation R36 or R41 — category 7
- Serious lesions to the eye R41 — category 8
- Moderate reversible irritation to the eye R36 — category 9
- Skin corrosion R34 or R35 — category 10

- Unknown category 11

Similarly to the skin irritation/corrosion rules, the classification here is done by physicochemical property exclusion rules and structural inclusion rules. Physicochemical exclusion rules are used to identify chemicals with no skin irritation/corrosion or eye irritation/corrosion potential and structural inclusion rules are used to identify chemicals with skin irritation/corrosion or eye irritation/corrosion potential.

Physicochemical exclusion rules

The decision rules depend on the following physicochemical properties: molecular weight, octanol-water partition coefficient LogP, melting point, aqueous solubility, lipid solubility (Table 2). The rules are valid for all groups of chemicals or are specific for the chemicals from a given chemical class (Table 1).

Table 1: Chemical classes for which specific rules are defined

Class (designation)	Empirical Formula
C	CxHyOz
CN	CxHyOzNa
CNHal	CxHyOzNaHalb
CNS	CxHyOzNaSb
CHal	CxHyOzHalb

Table 2: Physicochemical exclusion rules for eye irritation/corrosion as implemented in Toxtree

RuleID	Group	IF parameter	Qualifier	Value	Unit	Category
1	All	m.p. [a]	>	200	°C	1
2	All	logP	>	9		2
3	All	logP	<	-3.1		1
4	All	l.s. [b]	<	0.01	g/kg	1
5	All	a.s. [c]	<	0.000005	g/l	4
6	All	a.s.	<	0.00002	g/l	3
7	All	m.w. [d]	>	650	g/mol	4
8.1	C	m.p.	>	55	°C	1
8.2	C	m.w.	>	380	g/mol	2
8.3	C	a.s.	<	0.0005	g/l	7
8.4	C	a.s.	<	0.0001	g/l	2
9.1	CN	l.s.	<	0.4	g/kg	1
9.2	CN	m.w.	>	290	g/mol	1
9.3	CN	a.s.	<	0.1	g/l	1
9.4	CN	logP	>	4.5		1
10.1	CNHal	logP	>	3.8		5

RuleID	Group	IF parameter	Qualifier	Value	Unit	Category
10.2	CNHal	a.s.	<	0.1	g/l	1
10.3	CNHal	m.w.	>	370	g/mol	1
10.4	CNHal	l.s.	<	400	g/kg	1
10.5	CNHal	a.s.	<	0.004	g/l	3
11.1	CNS	m.w.	>	620	g/mol	6
11.2	CNS	m.p.	>	200	°C	4
11.3	CNS	m.p.	>	50	°C	1
11.4	CNS	log P	<	-2		1
11.5	CNS	log P	>	1.5		4
11.6	CNS	log P	>	3.6		3
11.7	CNS	a.s.	<	0.006	g/l	7
12.1	CHal	m.w.	>	370	g/mol	6
12.2	CHal	m.w.	>	280	g/mol	1
12.3	CHal	m.p.	>	65	°C	1
12.4	CHal	logP	>	4.5		7

[a] m.p. – melting point; [b] l.s. – lipid solubility; [c] a.s. – aqueous solubility; [d] m.w. – molecular weight

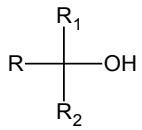
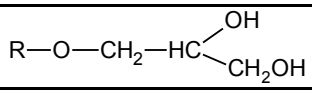
The Toxtree software calculates on the fly only LogP and molecular weight, for the rest it expects to be read from the file or manually entered by the user similarly to skin irritation/corrosion rules.

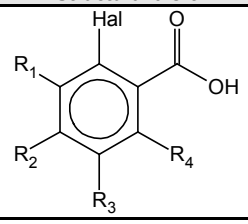
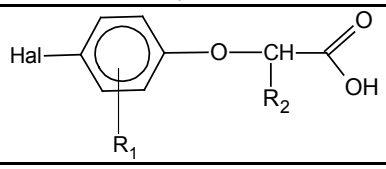
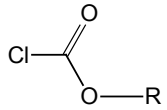
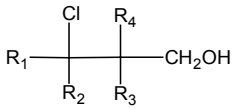
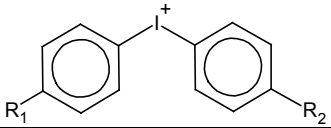
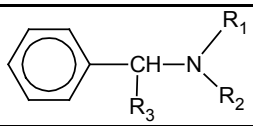
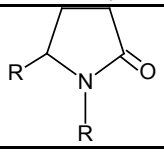
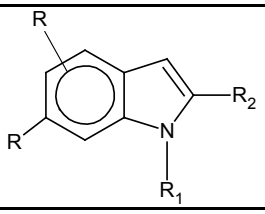
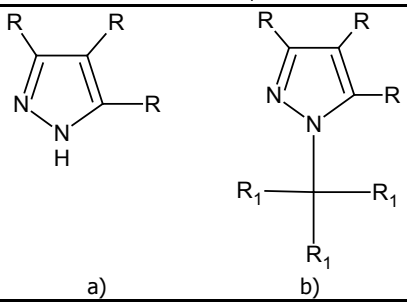
If the values are unknown, you might select the checkbox "Silent". In this case, the physicochemical rules will be skipped and only the structural rules will be applied. Please note that this may result in a low quality prediction.

Structural inclusion rules

The decision rules depend on the structural inclusion rules, given in Table 3, Table 4 and Table 5.

Table 3: Structural inclusion rules for predicting serious local lesions to the eye as implemented in Toxtree (category 8)

RuleID	Structural alert	Chemical class	Limits
13		Aliphatic monoalcohols	R = aliphatic chain R1,2 = H or aliphatic chain C3-C11 - eye damage C12-C14 - eye irritation
14		Aliphatic glycerol monoethers	R = aliphatic chain

RuleID	Structural alert	Chemical class	Limits
15		Derivatives of 2-halogen benzoic acids and corresponding alkali salts	R1-4 = H, aliphatic chain or halogen Hal = F, Cl or Br
16		Halogen benzenes with substituents containing carboxylic acid groups	R1 = H or halogen R2 = aliphatic chain Hal = F, Cl or Br
17		Aliphatic esters of chloro formic acid	R = aliphatic chain
18		Chlorinated aliphatic alcohols	R1 = aliphatic chain R2-4 = H or aliphatic chain
19		Diphenyl iodonium salts	R1 = aliphatic chain R2 = any
20		Derivatives of alpha amino benzene	R1 = H or aliphatic chain R2,3 = any
21		Pyrrolidones	R = H or aliphatic chain
22		Substituted indoles	R=H or OH R1=H or aliphatic ketone R2= any
23	 a) b)	Substituted pyrazoles	R = H, NH2 or aliphatic chain R1 = any

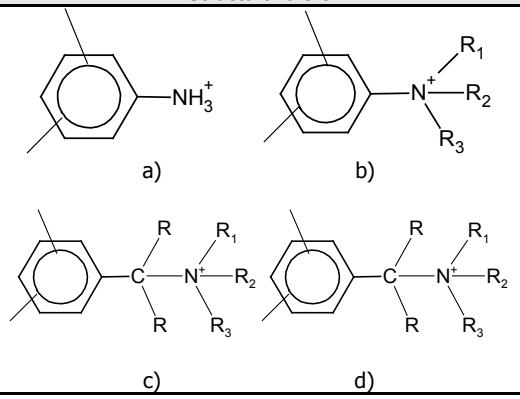
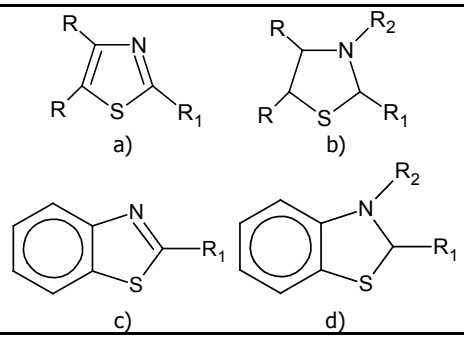
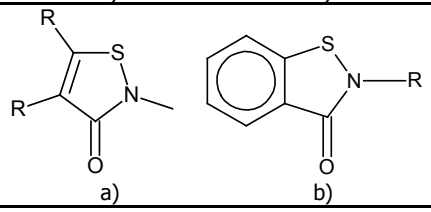
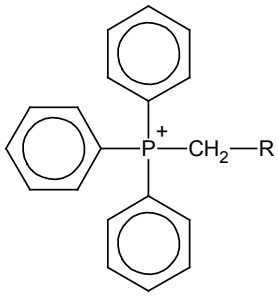
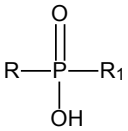
RuleID	Structural alert	Chemical class	Limits
24		Aromatic ammonium salts	R,R3=any; if containing halogen, thio- or sulpho- groups then R41 R1,2 = H or aliphatic chain
25	R—SO_3H	Organic sulphonic salts	R = any (aliphatic or aryl group)
26		Thiazoles and thiazolidines	R = any R1 = H or -C-any R2 = H or aliphatic chain
27		Thiazolones	R = aliphatic chain
28		Triphenylphosphonium salts	R = any
29		Organic phosphinic acids and their derivatives	R = aryl or aliphatic chain R1 = H or -CH2-any

Table 4: Structural inclusion rules for predicting moderate reversible irritation to the eye as implemented in Toxtree (category 9)

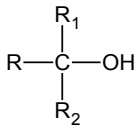
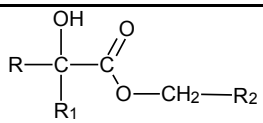
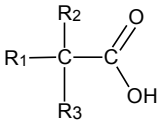
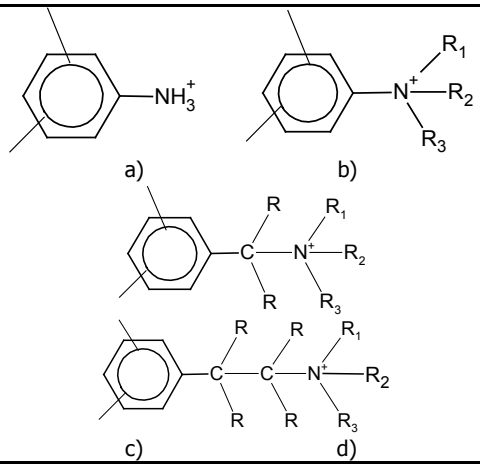
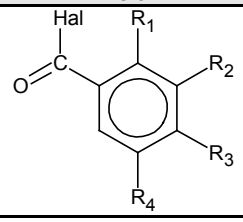
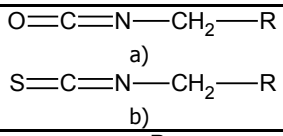
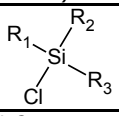
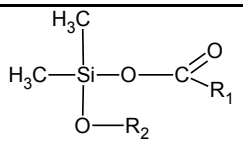
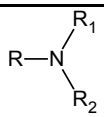
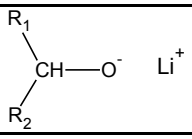
RuleID	Alert	Chemical class	Limits
30		Aliphatic monoalcohols	R=aliphatic chain $R_{1,2}$=H or aliphatic chain C_{12}-C_{14} - eye irritation > C_{15} - only slight eye irritation
31		Aliphatic alpha hydroxyesters	R,R_2 = aliphatic chain; R_1 = H or aliphatic chain
32		Aliphatic carboxylic acids	R_{1-3}=aliphatic chains with or without further oxygen containing substituents; C_6-C_{14} - strong eye irritation > C_{20} - only slight eye irritation
33		Aromatic ammonium salts	R,R_3=any; if containing halogen, thio- or sulpho groups then R41 $R_{1,2}$=H or aliphatic chain

Table 5: Structural inclusion rules for predicting skin corrosion as implemented in Toxtree (category 10)

RuleID	Alert	Chemical class	Limits
34		Substituted benzoic acid halides	Hal = Cl or F R_{1-4} = any
35		Aliphatic iso(thio)cyanates	R=aliphatic chain
36		Chlorosilanes	R_{1-3} = any (e.g., further halogen)
37		Mixed oxy-carboxysilanes	$R_{1,2}$ = any
38		Aliphatic amines	R = aliphatic chain which may contain ether functions $R_{1,2}$ = H or aliphatic chain
39		Alkali salts of aliphatic alcohols	R_1 = H or aliphatic chain R_2 = aliphatic chain

Benigni / Bossa rulebase (for mutagenicity and carcinogenicity) specifics

The processing of a query chemical can give rise to a limited number of different outcomes, namely: a) no presence of SAs for carcinogenicity; b) one or more SAs are recognized; c) SAs relative to aromatic amines or αβ-unsaturated aldehydes are recognized, and the chemical goes through QSAR analysis, which may result in a negative or positive outcome. The system flags either outcome through one, or a combination of several labels, as follows:

- **No alerts for carcinogenic activity** – no SAs have been recognized by the system;
- **Structural Alert for genotoxic carcinogenicity** – the system recognizes the presence of one or more SAs, and specifies a genotoxic mechanism;
- **Structural Alert for nongenotoxic carcinogenicity** – the system recognizes the presence of one or more SAs, and specifies a nongenotoxic mechanism;
- **Potential *S. typhimurium* TA100 mutagen based on QSAR** – assigned according to the output of QSAR6 or QSAR13;
- **Unlikely to be a *S. typhimurium* TA100 mutagen based on QSAR** – assigned according to the output of QSAR6 or QSAR13;
- **Potential carcinogen based on QSAR** – assigned according to the output of QSAR8 (aromatic amines);
- **Unlikely to be a carcinogen based on QSAR** – assigned according to the output of QSAR8 (aromatic amines);
- **For a better assessment a QSAR calculation could be applied** – assigned when one of QSAR6, QSAR8 or QSAR13 is applicable, but the user chooses not to apply a QSAR.

If the query chemical belongs to the classes of aromatic amines or αβ-unsaturated aldehydes, the appropriate QSAR is applied. A QSAR provides a more refined assessment than SAs, and should be given higher importance in a weight-of-evidence scheme. Thus, a QSAR analysis might point to an estimated lack of toxic effects, in spite of the presence of SAs.

All molecular descriptors for the QSARs are calculated on the fly by the software. LogP is calculated through the XlogP procedure[17,18], implemented by the CDK library. The implementation has been validated by comparison with other LogP implementations[19]. EHOMO and ELUMO are calculated by launching OpenMopac

---

[17] R. Wang, Y. Fu, and L. Lai. A New Atom-Additive Method for Calculating Partition Coefficients. J. Chem. Inf. Comput. Sci., 37:615–621, 1997.

[18] R. Wang, Y. Gao, and L. Lai. Calculating partition coefficient by atom-additive method. Perspectives in Drug Discovery and Design, 19:47–66, 2000.

[19] Uli Fechner, Kristina Grabowski, QA of the XlogP Descriptor, CDK News - The Newsletter of CDK project, Volume 3/1, March 2004, ISBN 1614-7553, pp. 12-14, http://sourceforge.net/projects/cdk/files/CDK%20News/3_1/cdknews3.1.pdf/download

7.1[20]. An additional validation step was performed by comparing the results with the original values presented by the authors of the QSARs.

In order to reproduce the values of the descriptors present in the original papers, the following correction factors are applied to the calculated values (the corrected values are used for predictions):

QSAR13

$MR_{QSAR13} = 0.8718 * MR – 2.3452$

$LogP_{QSAR13} = 0.99738 * LogP – 0.10589$

$ELUMO_{QSAR13} = 1.07907 * ELUMO – 0.01463$

QSAR6

$EHOMO_{QSAR6} = 1.03383 * EHOMO + 0.30348$

$ELUMO_{QSAR6} = 0.98963 * ELUMO – 0.04037$

QSAR8

$EHOMO_{QSAR8} = 0.88239 * EHOMO – 1.0381$

$ELUMO_{QSAR8} = 0.96239 * ELUMO – 0.01521$

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 30 or Figure 31, if a QSAR is to be calculated.

[20] Available at http://openmopac.net/Downloads/MOPAC_7.1executable.zip

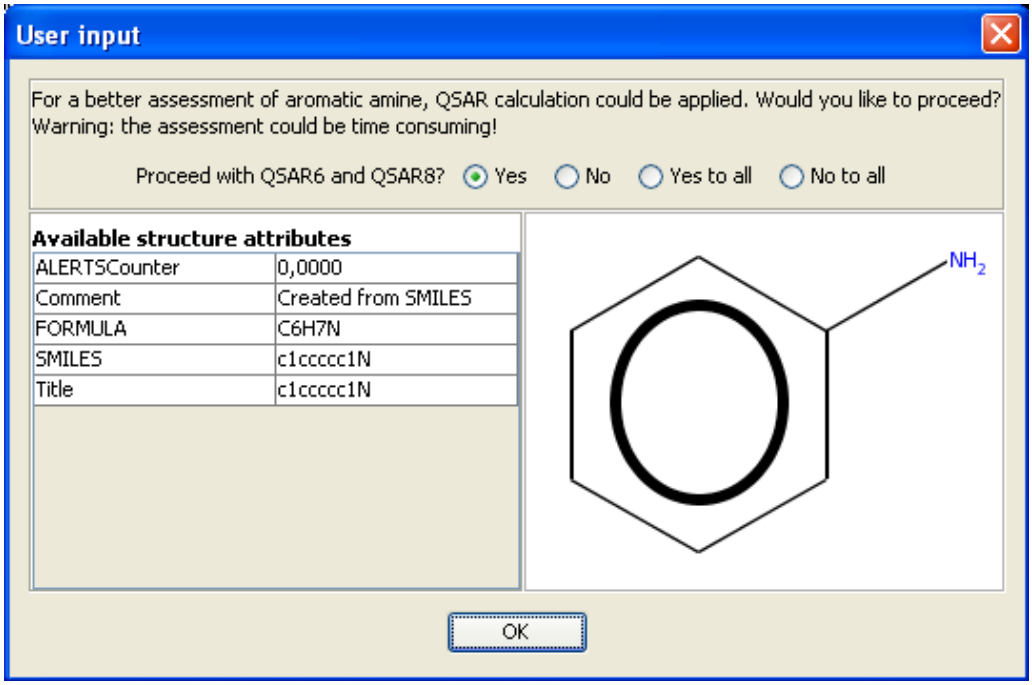


Figure 30: Options for QSAR6 and QSAR8 calculations

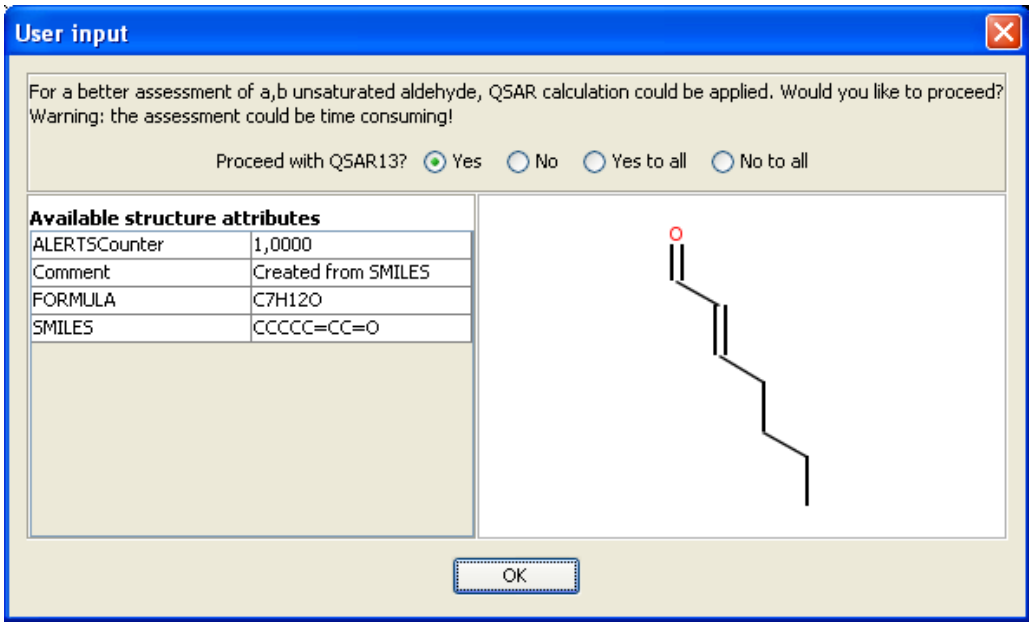


Figure 31: Options for QSAR13 calculations

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.

The rule options can also be accessed through the Rules section of the "Method▶Decision Tree Options" menu (Figure 32).

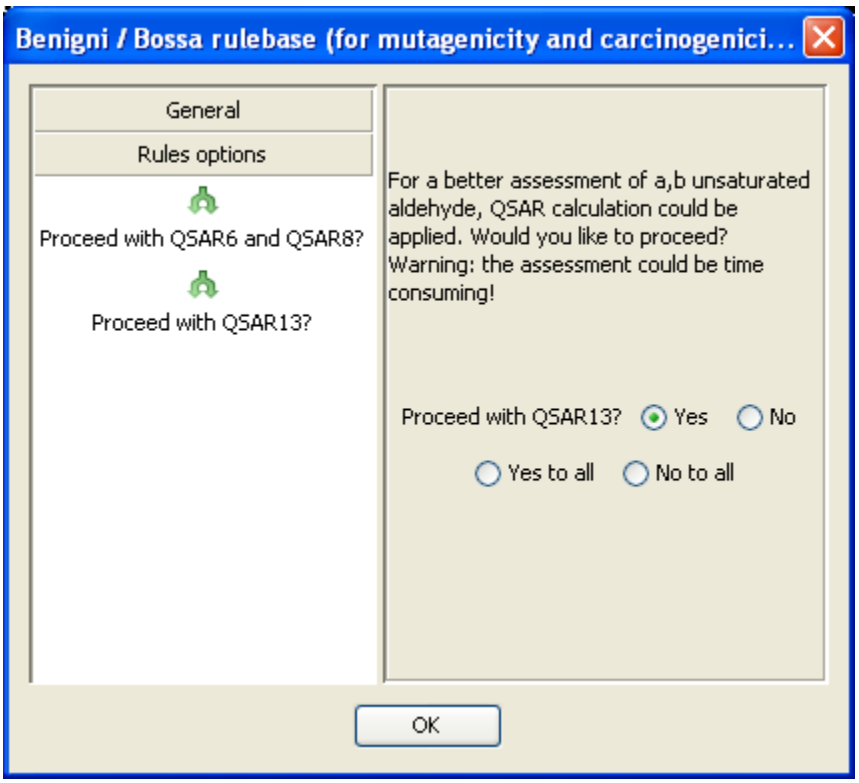


Figure 32: Decision tree specific options

SMARTCyp – Cytochrome P450 – mediated metabolism

Prediction of sites of metabolism is performed by SMARTCyp, as published in:

Patrik Rydberg, David E. Gloriam, Jed Zaretzki, Curt Breneman, Lars Olsen, SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism, ACS Med. Chem. Lett., 2010,1 (3), pp 96 100

Patrik Rydberg, David Gloriam and Lars Olsen, The SMARTCyp cytochrome P450 metabolism prediction server, Bioinformatics, 2010, 26, 2988-2989

Toxtree 2.5.0 adds metabolite prediction, based on sites of metabolism, predicted by SMARTCyp and a set of relevant reactions, defined as SMIRKS. Clicking on the link *"Q1. SMARTCyp primary sites of metabolism"* highlights the site(s) of metabolism and shows an additional window, with the generated metabolite(s). The "Copy molecule" button transfers the metabolite into the main screen, where it can be further processed as usual.

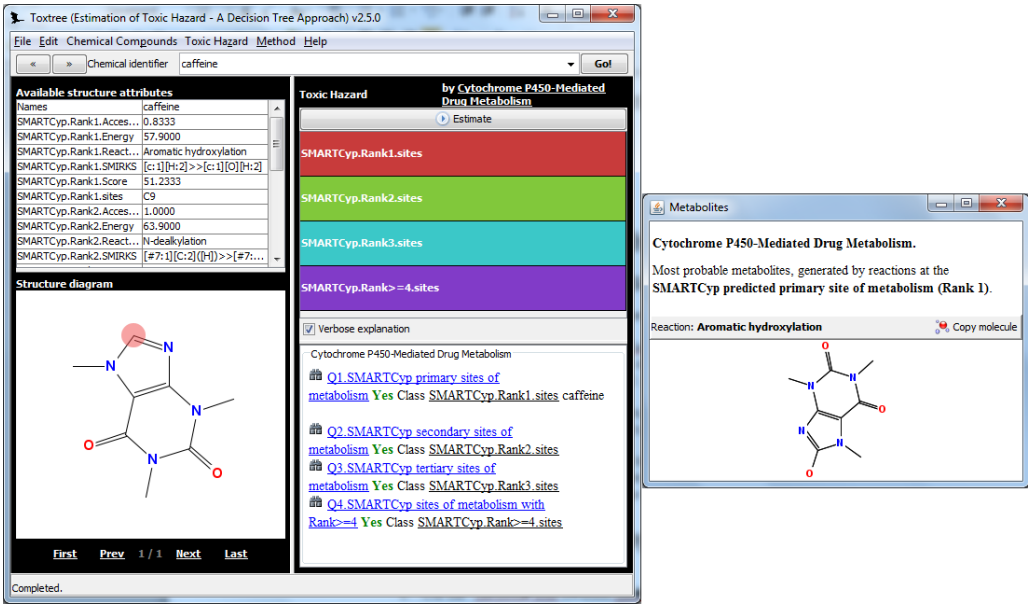


Figure 33: SMARTCyp plugin

Decision tree editing - typical tasks

Creating a simple decision tree

1. Use the "Method►Edit Decision tree►New Decision Tree" menu option to create a new (empty) decision tree. This will launch the decision tree editor, as shown on Figure 34.

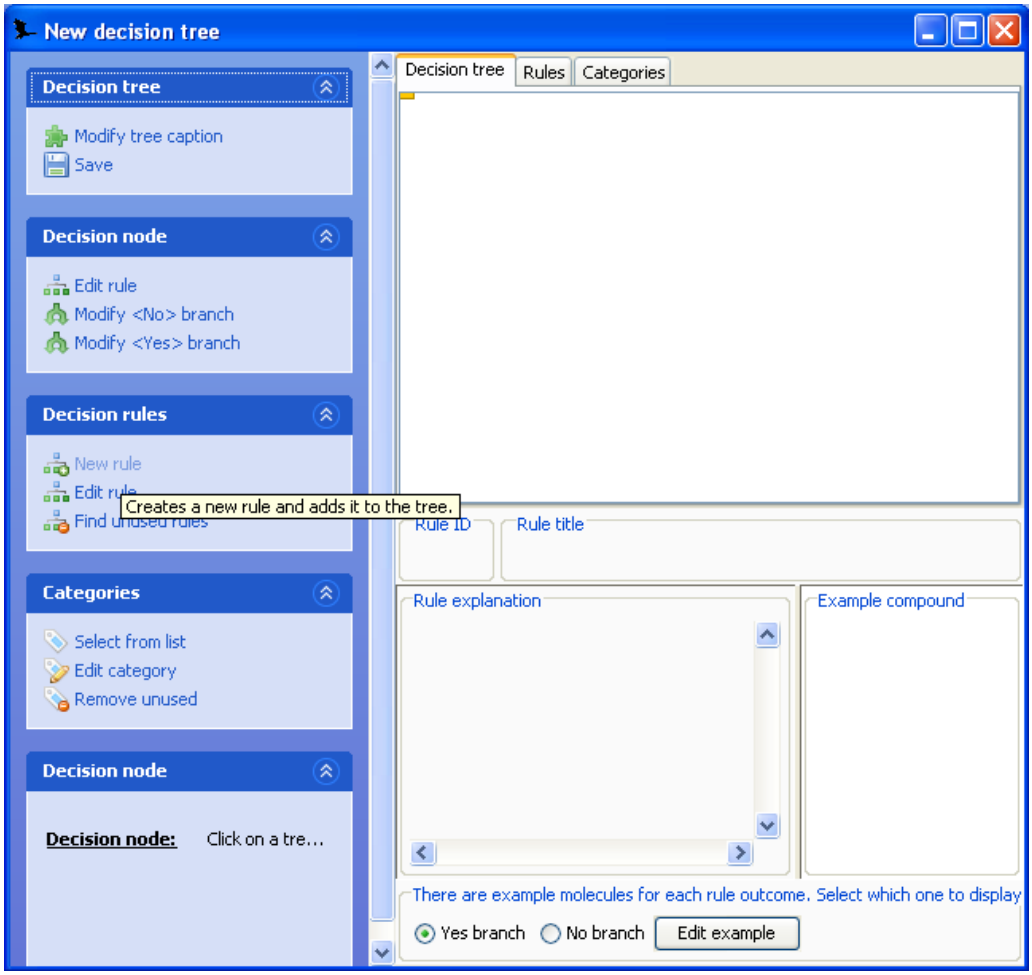


Figure 34: Decision tree editor on empty decision tree

2. Click the "Decision rules►New Rule" menu option to create a new rule. This will launch the decision rule wizard, as shown on Figure 35.

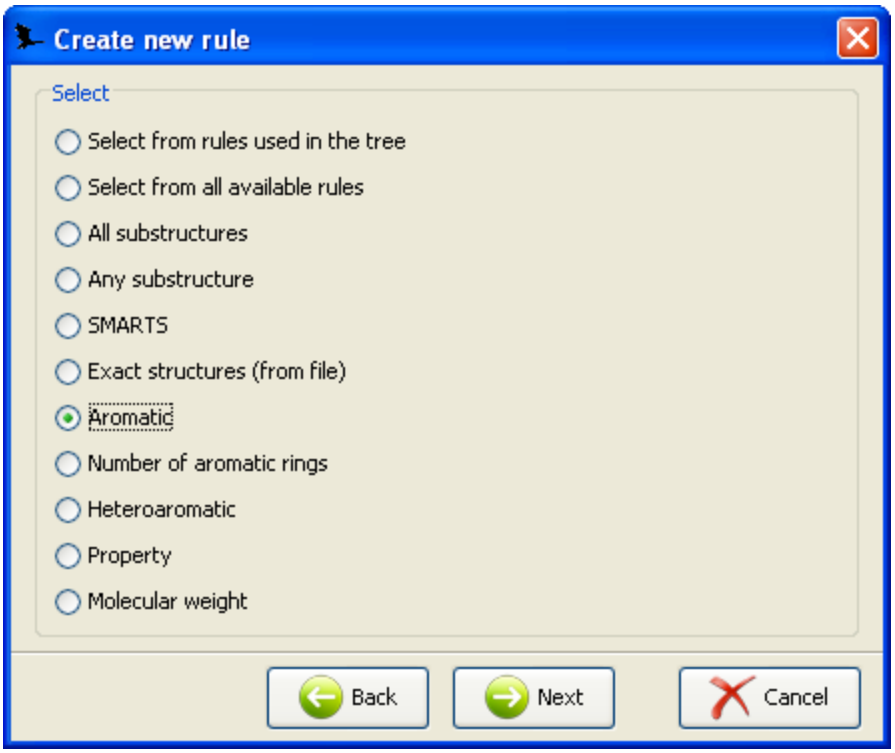


Figure 35: Decision tree editor – new rule wizard

3. Click on the "Aromatic" option to create a rule that will verify if the substance is aromatic. Then click OK. The next page of the wizard will appear, as shown on Figure 36.

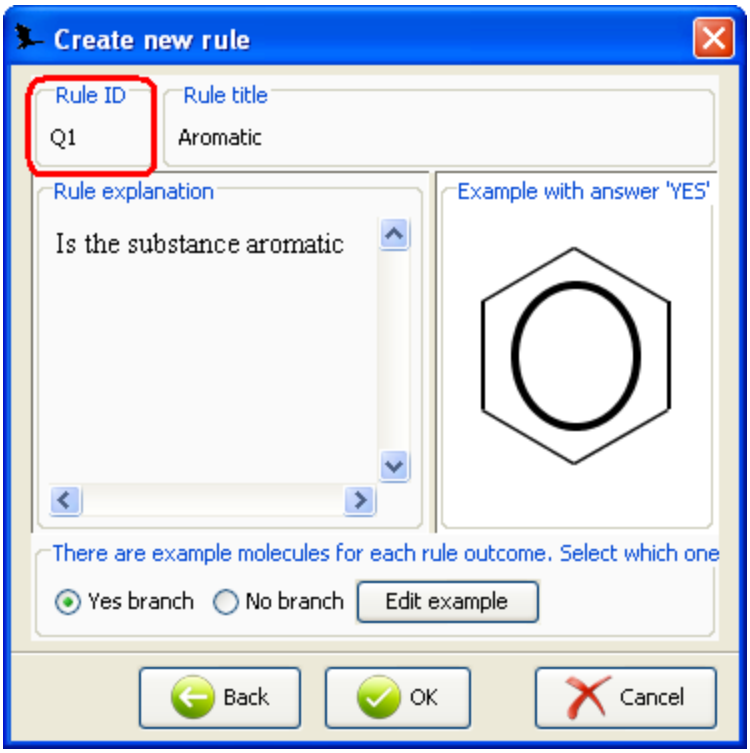


Figure 36: Decision tree editor – aromatic rule editor

4. Click on the "Rule ID" field, and type in "Q1". This will be used as a rule identifier. You could also modify the rule's title, explanation and examples. When finished, click OK. The rule will be added as a top node of the tree as shown on Figure 37. Clicking on the rule will highlight it with orange colour and will display decision node details on the left and decision rule details at the bottom.

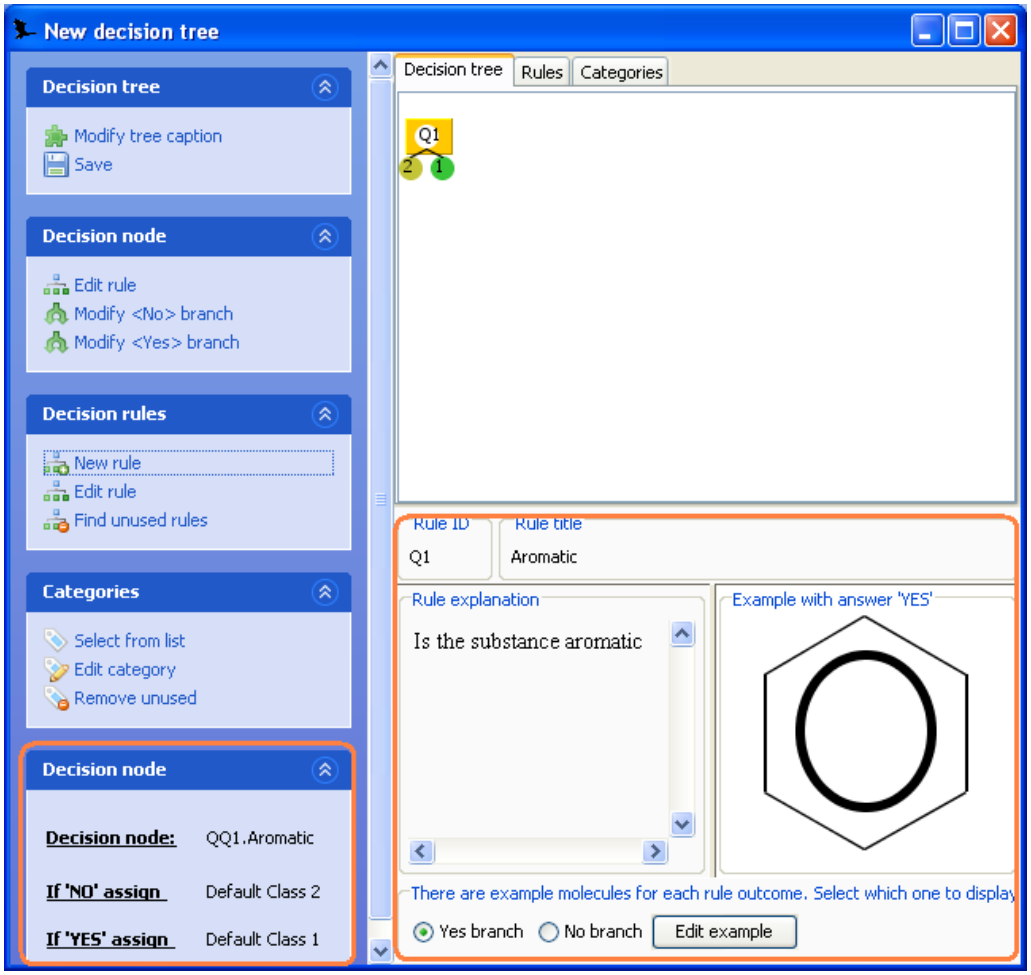


Figure 37: Decision tree editor – new rule added as a top node

5. Click on the "Categories" tab. This will show a list of categories used so far in the tree. Click on "Default class 2" (second row), as shown on Figure 38.

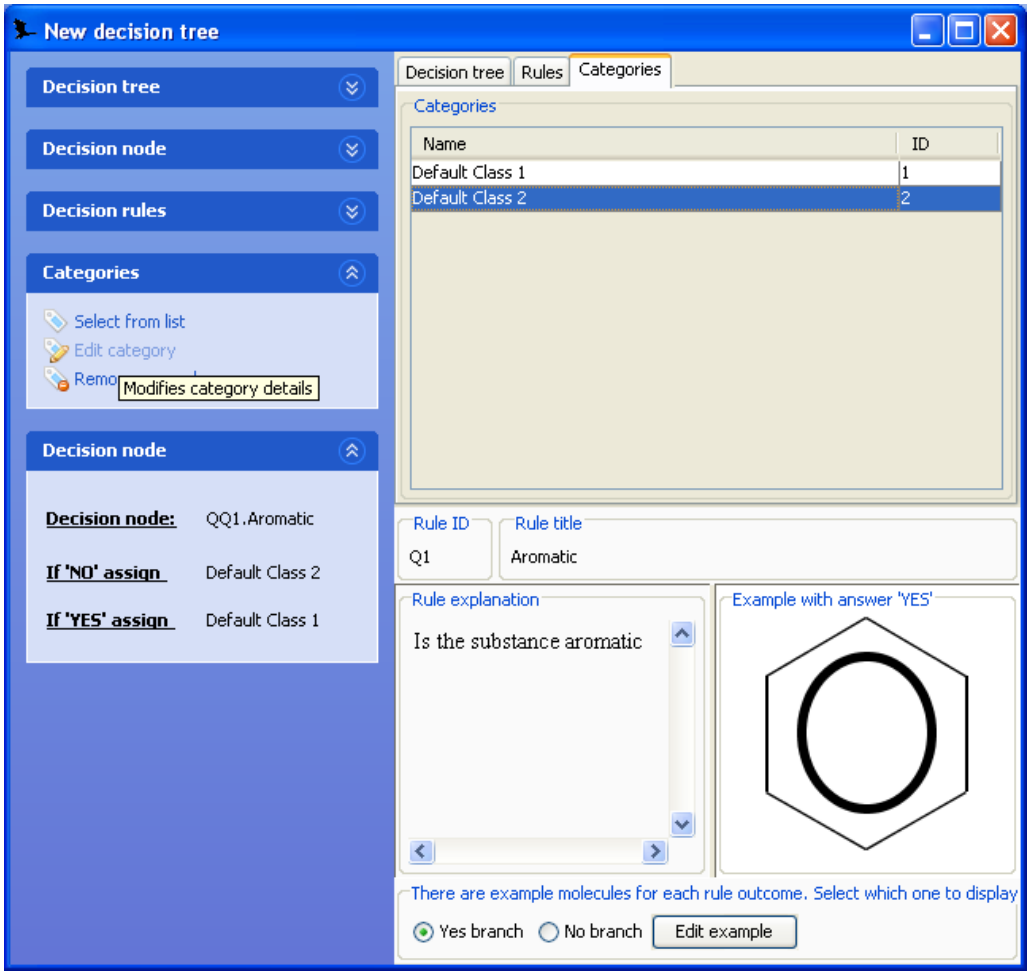


Figure 38: Decision tree editor – modifying a category

6. Then click on "Categories►Edit category" menu option on the left. This will launch the category editor, as shown on Figure 39. Type in "Not aromatic" as category title and similar text in the explanation field. Click OK when ready.

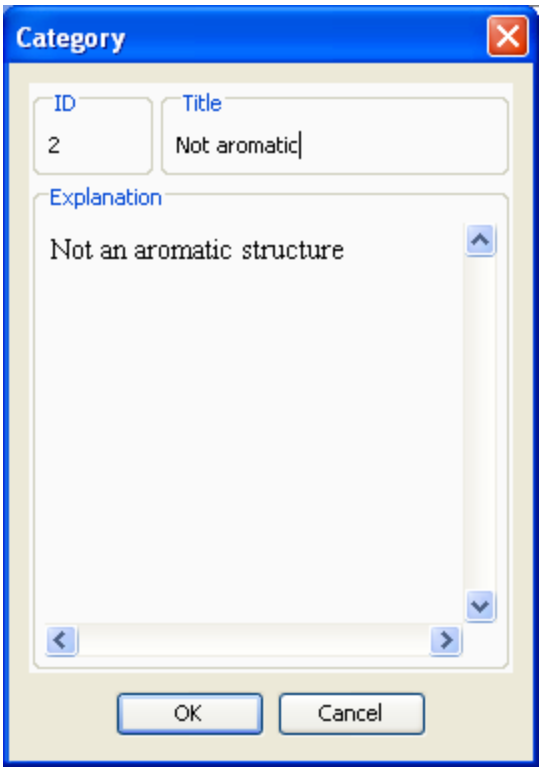


Figure 39: Decision tree editor – category editor

7. Right click on the decision node in order to invoke a popup menu, as shown on Figure 40. These menu options allow editing the rule and modifying the left and right branches. The left branch is followed when the answer of the rule is "No" and the right branch is followed if the answer is "Yes".

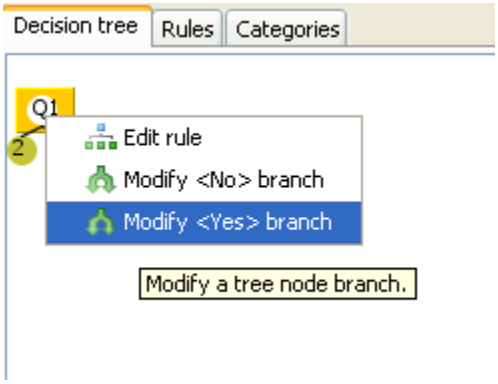


Figure 40: Decision tree editor – popup menu

8. Proceed with modifying the right branch, which in the current tree will be followed when the structure is aromatic. Click on the "Modify <YES>

branch" menu option. This will launch a wizard, as shown on Figure 41. Select the "Rule" option and click "Next".

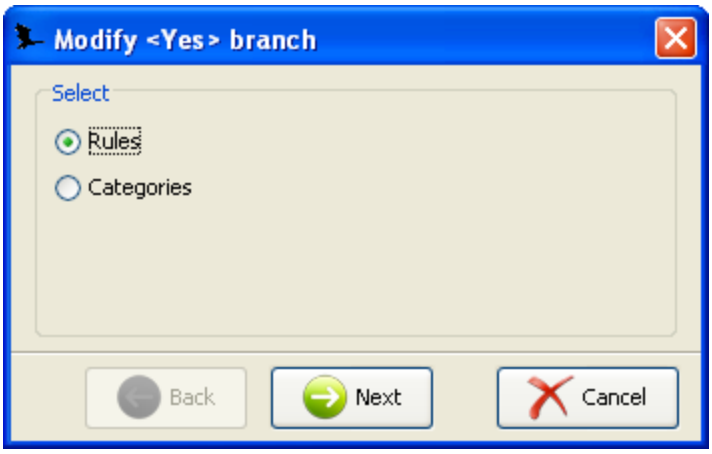


Figure 41: Decision tree editor – branch wizard

9. The next wizard page shows the same options as in Figure 35. Click on the "Heteroaromatic" option and follow the wizard's instructions. At the end a new node will be added to the tree, as shown on Figure 42.

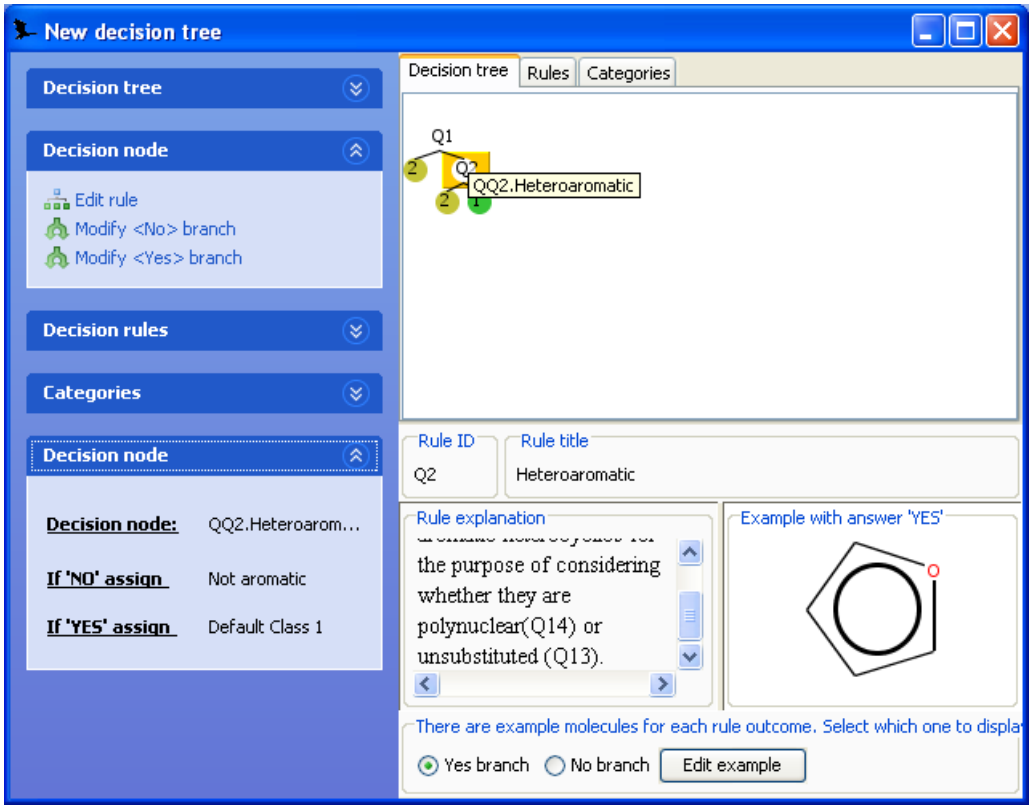


Figure 42: Decision tree editor – decision tree with two nodes

10. Proceed with modifying the left branch, which in the current tree will be followed when the structure is aromatic but not heteroaromatic. Right click on the new node and invoke the "Modify <NO> branch" menu option. The same wizard as in Figure 41 will appear. Select the "Category" option and click "Next".

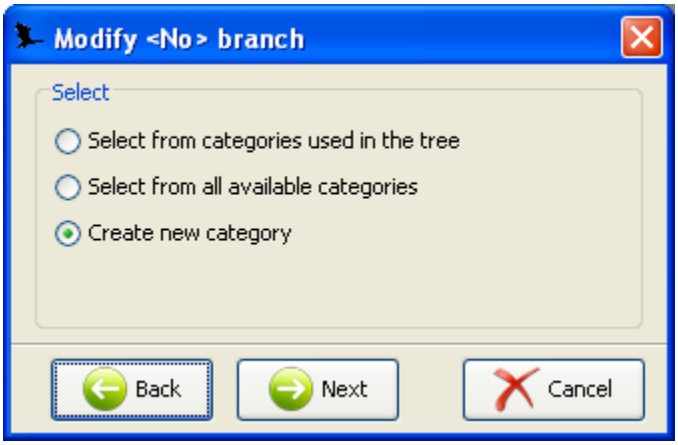


Figure 43: Decision tree editor – category options

11. Select the "Create new category" from the three options, shown on Figure 43, and click "Next". A category editor will appear, as in Figure 39. Type in "Aromatic, but not heteroaromatic" as a title and "The structure is not heteroaromatic" in the explanation field. Then click OK. This will change the left branch of the second node.

12. Finally, we will modify the result that will be assigned if the structure is heteroaromatic. Right click on the second (Heteroaromatic) node and invoke the "Modify <YES> branch" menu option. The same wizard as in Figure 41 will appear. Select the "Category" option and click "Next".

13. From the category wizard (Figure 43) select the first option – "Select from categories used in this tree" and click "Next". The next page displays the categories used in the tree. Select "Default class 1", as shown on Figure 44, and follow the wizard's instructions.

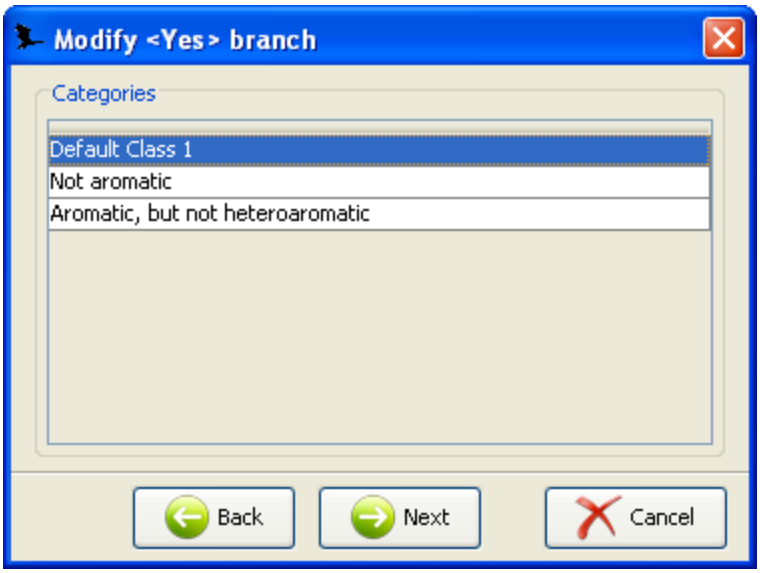


Figure 44: Decision tree editor – list of categories used in the tree

The next page will be a category editor as in Figure 39. Type in "Heteroaromatic" in the title field. Click OK when ready. The resulting tree should look like as shown on Figure 45 (2 nodes, 3 categories).

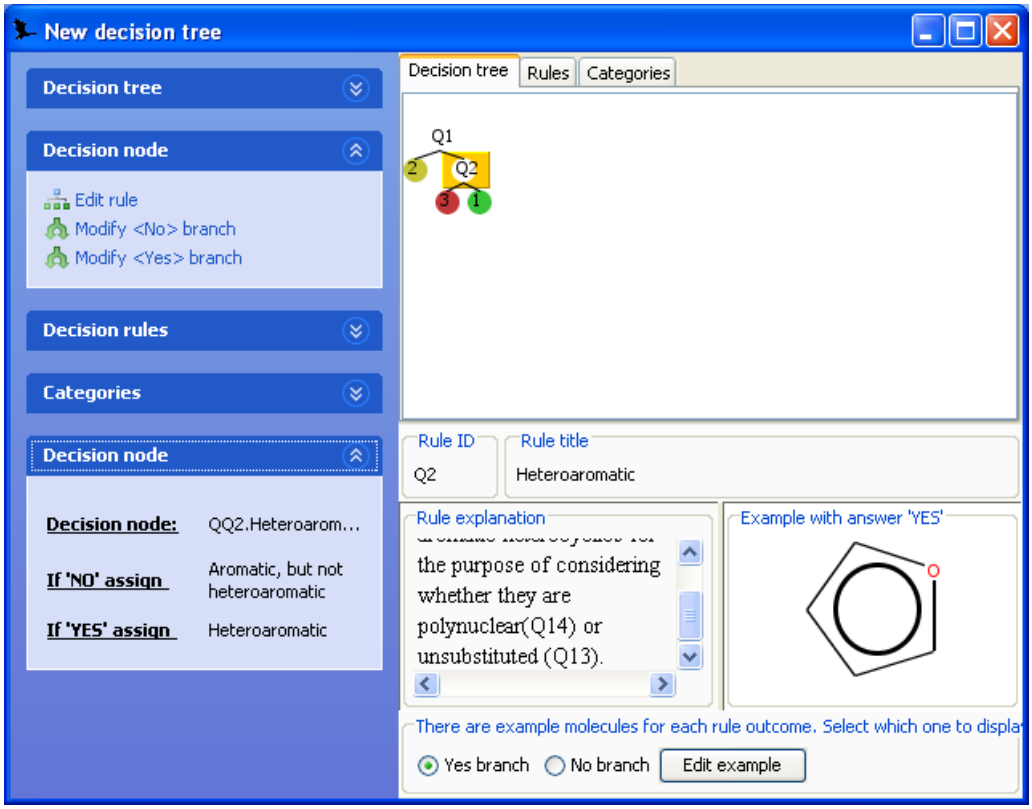


Figure 45: Decision tree editor – the resulting tree (2 nodes, 3 categories)

14. Use the "Decision Tree►Modify tree caption" menu option on the left. This will ask you to enter the tree's title. Type in "Aromatic substances" and click OK.

15. Save the file as *.tree or *.tml file and close the decision editor. You might then load the tree by "Method►Select Decision Tree" menu option Figure 46.

    *The ***.tree** file is a binary file and normally .tree files are incompatible between releases. A new XML based format ***.tml** was introduced, in order to avoid incompatibility of .tree files. The ***.tml** file is the preferred format, while ***.tree** format is likely to be abandoned in subsequent releases.*

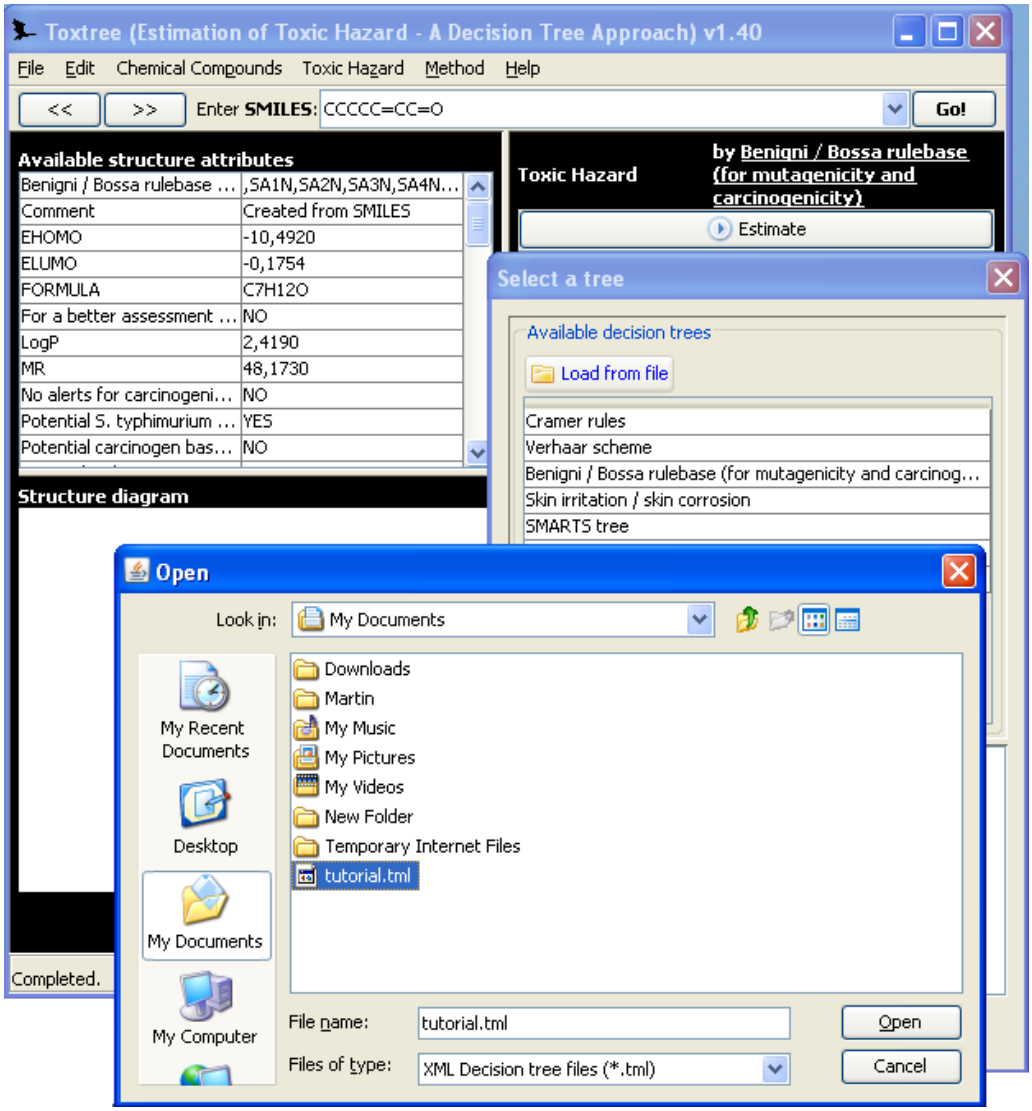


Figure 46: Loading the new decision tree from a file

16. The decision tree will be loaded as a default decision method in Toxtree. The classification area will reflect the categories, defined in the tree (Figure 47).

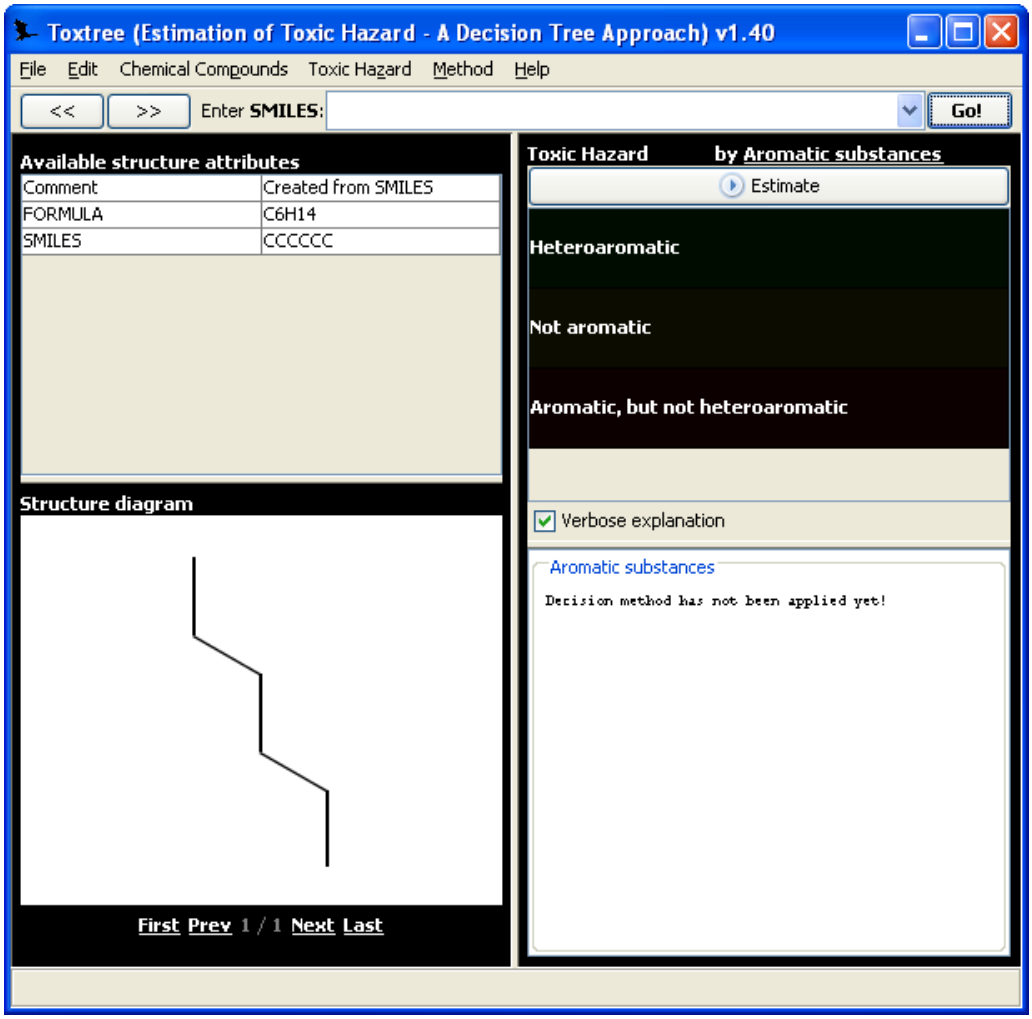


Figure 47: The new "Aromatic substances" decision tree displayed on the main Toxtree screen

Using a decision tree

1. Apply the decision tree, created in the previous section, to the default structure (hexane) by clicking on the "Estimate" button. This applies the decision tree to the current structure. The result says that the compound is not aromatic, as shown on Figure 48. The result is also assigned as a property of the compound and will be saved along with other properties when using the "File►Save" menu option.

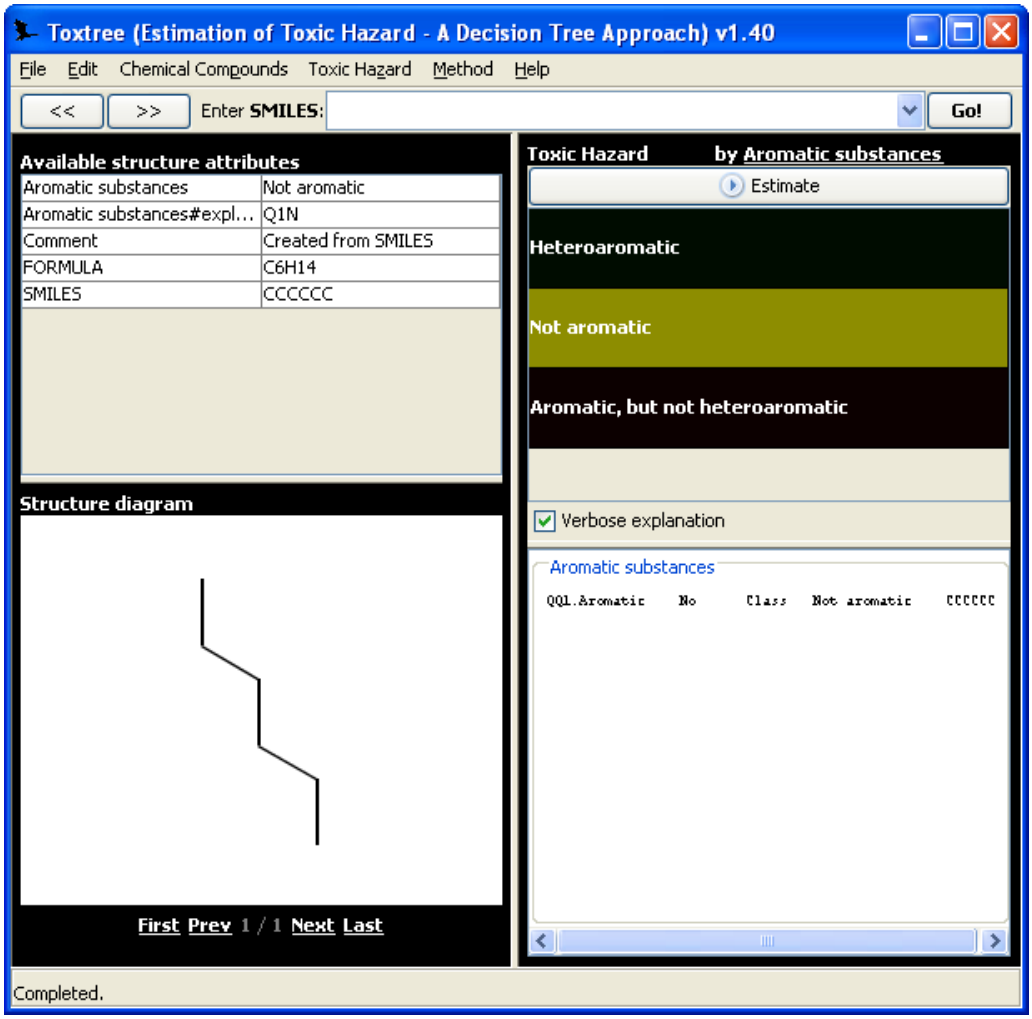


Figure 48: The result of applying the new tree to Hexane

2. Click on the "Chemical compounds►Edit compound" menu option. This will launch the JChemPaint[21] structure diagram editor and allows editing the current structure. Add a benzene ring to the chain, as shown on Figure 49. Click OK when ready. This will update the structure, displayed on the main screen.

[21] More information about JChemPaint can be found at http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint

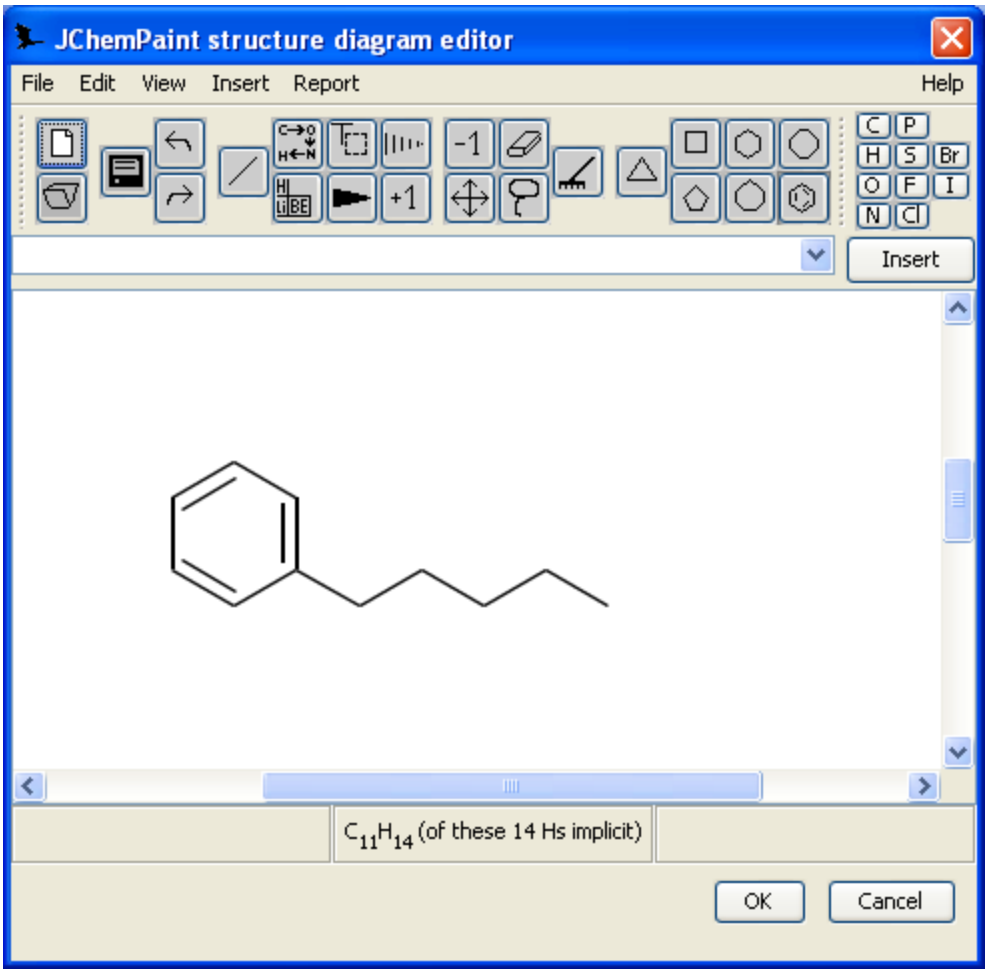


Figure 49: JChempaint structure diagram editor

3. Click again on the "Estimate" button. This will apply the decision tree to the new structure. The result is shown on Figure 50.

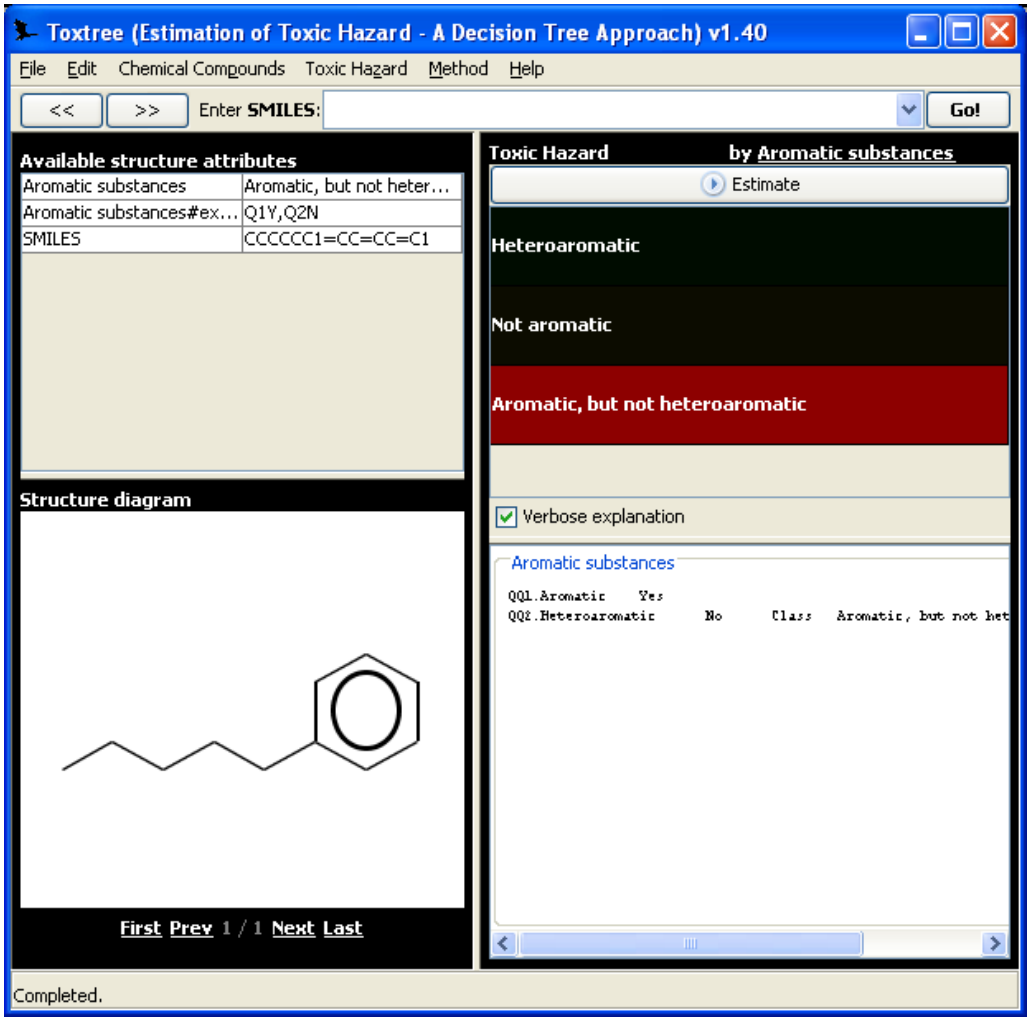


Figure 50: The result of applying the new tree to the new structure