

# A new module for predicting mutagens / carcinogens implemented in Toxtree v. 1.4

C. Bossa<sup>1</sup>, A. Worth<sup>2</sup>, T. Netzeva<sup>3</sup>, N. Jeliaskova-Nikolova<sup>4</sup> and R. Benigni<sup>1</sup>

<sup>1</sup> Istituto Superiore di Sanità, Environment and Health Department, Rome; <sup>2</sup> European Commission, Institute for Health and Consumer Protection, Consumer Product Safety and Quality, Joint Research Centre, Ispra ; <sup>3</sup> European Chemicals Agency ECHA, Helsinki; <sup>4</sup> Ideaconsult Ltd., Sofia.

rbenigni@iss.it

## ABSTRACT

The European Chemicals Bureau (ECB) has developed (through IdeaConsult Ltd.) a software tool called Toxtree that is able to estimate different types of toxic hazards by applying structural rules. This is a freely available application from the ECB website (<http://ecb.jrc.it/QSAR>). Recently, a new module with rules for predicting the carcinogenicity and mutagenicity of chemicals has been implemented. The core of the module is a list of Structural Alerts (SA) for carcinogenicity. The SAs for carcinogenicity are molecular functional groups or substructures known to be linked to the carcinogenic activity of chemicals. As one or more SAs embedded in a molecular structure are recognized, the system flags the potential carcinogenicity of the chemical. The SAs in Toxtree refer mainly to the knowledge on the action mechanisms of genotoxic carcinogenicity (thus they apply also to the mutagenic activity in bacteria), but include also a number of SAs flagging potential nongenotoxic carcinogens. The list of SAs derives from the critical evaluation of different existing proposed sets of SAs. Because of their nature, the SAs have the role of pointing to chemicals potentially toxic, whereas no conclusions or indications about nontoxic chemicals are possible (except by exclusion). Thus the SAs are not a discriminant model on the same ground of the Quantitative Structure-Activity Relationships (QSAR) models, that produce estimates for both positive and negative chemicals. In addition to the SAs, this software includes QSAR models for: 1) the mutagenic activity of aromatic amines in the *Salmonella typhimurium* TA100 strain (Ames test); 2) the carcinogenic activity of the aromatic amines in rodents (summary activity from rats and mice); 3) the mutagenic activity of  $\alpha,\beta$ -unsaturated aldehydes in the *Salmonella typhimurium* TA100 strain (Ames test).

### Carcinogens: mechanisms of action

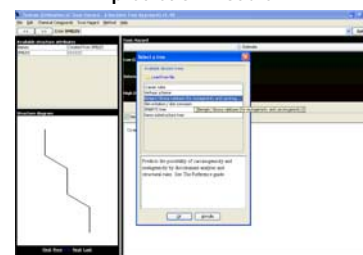
### Toxtree: mutagens/carcinogens prediction module

#### genotoxic carcinogens

- electrophiles or activated to electrophilic reactive intermediates
- damage to DNA
- putatively mutagens

#### epigenetic carcinogens

- no direct damage to DNA
- usually negative in mutagenicity assay



**Structural Alerts (SAs) for genotoxic carcinogens (n=30)**  
chemical functional groups and substructures that have been shown to be associated with DNA reactivity and genotoxic carcinogenicity effects

**Structural Alerts for non-genotoxic carcinogens (n=5)**  
diverse mechanisms with no apparent unifying concept

The mutagens/carcinogens prediction module is based on a series of SAs (implemented as SMARTS) and a number of QSARs. The SAs are accompanied by modulating (detoxifying) factors.

### Toxtree, SAs and QSARs: Possible outcomes

- No presence of SAs for carcinogenicity; Toxtree flags:

No Alerts for cancerogenic activity

- One or more SAs are recognised; Toxtree flags:

Structural Alert for genotoxic carcinogenicity

and/or

Structural Alert for nongenotoxic carcinogenicity

- SAs relative to aromatic amines or  $\alpha,\beta$ -unsaturated aldehydes are recognised, and the chemical goes through QSAR analysis, which may result in a negative or positive outcome; Toxtree flags:

Potential *S. typhimurium* TA100 mutagen based on QSAR

or

Unlikely to be a *S. typhimurium* TA100 mutagen based on QSAR

and/or

Potential carcinogen based on QSAR

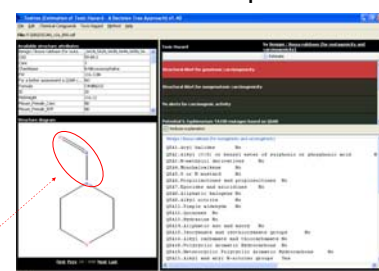
or

Unlikely to be a carcinogen based on QSAR

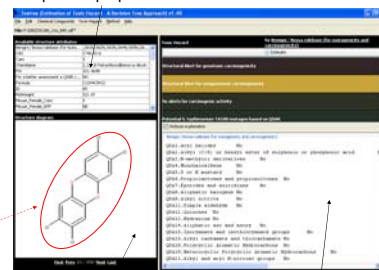
### Toxtree: Structural Alert list

SA_1: Acyl halides	SA_13: Hydrazine	SA_25: aromatic nitroso group
SA_2: alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid	SA_14: Aliphatic azo and azoxy	SA_26: aromatic ring N-oxide
SA_3: N-methylol derivatives	SA_15: isocyanate and isothiocyanate groups	SA_27: Nitro-aromatic
SA_4: Monohaloalkene	SA_16: alkyl carbamate and thiocarbamate	SA_28: primary aromatic amine, hydroxyl amine and its derived esters
SA_5: S or N mustard	SA_17: Thiocarbonyl (nongenotoxic)	SA_28bis: Aromatic mono- and dialkylamine
SA_6 Propiolactones or propiolactones	SA_18: Polycyclic Aromatic Hydrocarbons	SA_28ter: aromatic N-acyl amine
SA_7: Epoxides and aziridines	SA_19: Heterocyclic Polycyclic Aromatic Hydrocarbons	SA_29: Aromatic diazo
SA_8: Aliphatic halogens	SA_20: (Poly) Halogenated Cycloalkanes (nongenotoxic)	SA_30: Coumarins and Furocoumarins
SA_9: Alkyl nitrite	SA_21: alkyl and aryl N-nitroso groups	SA_31a: Halogenated benzene (nongenotoxic)
SA_10: a, b unsaturated carbonyls	SA_22: azide and triazine groups	SA_31b: Halogenated PAH (nongenotoxic)
SA_11: Simple aldehyde	SA_23: aliphatic N-nitro group	SA_31c: Halogenated dibenzodioxins (nongenotoxic)
SA_12: Quinones	SA_24: $\alpha,\beta$ unsaturated aliphatic alkoxy group	

### Toxtree: Examples



Compounds properties area



Compounds structure area

Classification area

### Toxtree: QSAR models

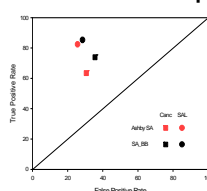
Finely-tuned estimations for chemicals belonging to aromatic amines and  $\alpha,\beta$ -unsaturated aliphatic aldehydes chemical classes

QSAR 6: Mutagenic activity of aromatic amines in *Salmonella typhimurium* TA100 (with S9 metabolic activation)  
 $w = -3.14 \text{ HOMO} + 1.76 \text{ LUMO} + 0.62 \text{ MR2} + 0.75 \text{ MR3} + 1.88 \text{ MR6} + 3.75 \text{ Idist}$   
 $w(\text{mean,Class1}) = 28.42; N1 = 47$  (non-mutagens)  
 $w(\text{mean,Class2}) = 26.44; N2 = 64$  (mutagens)

QSAR 8: Carcinogenicity of aromatic amines in rodents (mice, rats)  
 $w = -3.79 \text{ L(R)} + 3.52 \text{ B5(R)} - 4.12 \text{ HOMO} + 4.41 \text{ LUMO} + 3.09 \text{ MR3} + 2.60 \text{ MR5} + 4.63 \text{ MR6} - 3.49 \text{ I(An)} + 1.80 \text{ I(NO2)} - 1.78 \text{ I(BiBr)}$   
 $w(\text{mean,Class1}) = 27.82; N1 = 12$  (non-carcinogens)  
 $w(\text{mean,Class2}) = 30.34; N2 = 52$  (carcinogens)

QSAR 13: Mutagenic activity of  $\alpha,\beta$ -unsaturated aliphatic aldehydes in *Salmonella typhimurium* TA100 (without S9 metabolic activation)  
 $w = 0.387 \text{ MR} - 3.12 \text{ logP} + 3.23 \text{ LUMO}$   
 $w(\text{mean,Class1}) = 9.69; N1 = 3$  (non-mutagens)  
 $w(\text{mean,Class2}) = 6.37; N2 = 17$  (mutagens)

### Toxtree: The performance of SAs



Agreement of the list of SAs (SA\_BB, black symbols) with the carcinogenicity (Canc) and mutagenicity (SAL) of chemicals in the ISSCAN database<sup>5</sup>. For a comparison the performance of the Ashby SAs (red symbols) in the same database are reported.

Overall Accuracy: SA\_BB 70% 78%  
Ashby SA 65% 78%

## REFERENCES

R.Benigni, C.Bossa, N.G.Jeliaskova, T.I.Netzeva, A.P.Worth. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. EUR 23241 EN. 2008. Luxembourg, Office for the Official Publications of the European Communities. EUR - Scientific and Technical Report Series. download: [http://ecb.jrc.it/documents/QSAR/EUR\\_23241\\_EN.pdf](http://ecb.jrc.it/documents/QSAR/EUR_23241_EN.pdf)

\*Istituto Superiore di Sanità. ISSCAN database - Chemical Carcinogens: Structures and Experimental Data. download: <http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7>  
<http://www.epa.gov/NCCT/dsstoX/OtherPublicDatabases.html>