JRC Scientific and Technical Reports

Computational Characterisation of Chemicals and Datasets in Terms of Organic Functional Groups - a New Toxtree Rulebase

Romualdo Benigni[a], Olga Tcheremenskaia[a] and Andrew Worth[b]

[a] Istituto Superiore di Sanita', Environment and Health Department, Rome, Italy

[b] European Commission - Joint Research Centre, Institute for Health & Consumer Protection, Systems Toxicology Unit, Ispra, Italy

The mission of the JRC-IHCP is to protect the interests and health of the consumer in the framework of EU legislation on chemicals, food, and consumer products by providing scientific and technical support including risk-benefit assessment and analysis of traceability.

Europe Direct is a service to help you find answers
to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/

ABSTRACT

Toxtree is a freely available, user-friendly and extensible software application that is designed to make structure-based predictions for a number of toxicological endpoints and mechanisms of chemical action. The platform has been developed by the Joint Research Centre in collaboration with Ideaconsult Ltd (Sofia, Bulgaria) with a range of modules developed by various contributors. One of the modules developed as an extension to Toxtree is aimed at the identification of organic functional groups in query chemicals. The rulebase consists of 204 organic functional groups recognised by the "Checkmol" program, which was developed by Dr Norbert Haider, University of Vienna. A new Functional Group Profiler, has been coded as a Toxtree module by the Istituto Superiore di Sanita' (Rome, Italy). The Toxtree profiler, called ISSFUNC, can be used to screen and characterise chemicals as a basis for read-across, category formation and (Q)SAR analysis. It can also be used for the global comparison of datasets, such as model training and test sets and chemical inventories.

LIST OF ABBREVIATIONS

ISS	Istituto Superiore di Sanità
ISSCAN	Istituto Superiore di Sanità database on chemical carcinogens
EPA	Environmental Protection Agency (USA)
EU	European Union
JRC	Joint Research Centre
QSAR	Quantitative Structure-Activity Relationships
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals

CONTENTS

1. Introduction

In the context of the recent developments in chemicals regulations and regulatory needs worldwide, the progress in chemoinformatics technology is particularly timely in providing an essential tool to support the chemical assessment process. Until now, the assessment of chemical risks in the European Union (EU) has been largely based on traditional toxicology. However legislative, societal and practical realities (too many chemicals, too few resources) have created new inducements and opportunities to encourage the use and acceptance of "alternative" approaches, which can reduce substantially the need for experimental toxicological testing.

In 2003, the European Commission (EC) adopted a legislative proposal for a new chemical assessment and management system called REACH (Registration, Evaluation and Authorisation of CHemicals). Article 13(1) of the REACH regulation (EC, 2006) states that:

> "Information on intrinsic properties of substances may be generated by means other than tests, provided that the conditions set out in Annex XI are met. In particular for human toxicity, information shall be generated whenever possible by means other than vertebrate animal tests, through the use of alternative methods, for example, in vitro methods or qualitative or quantitative structure-activity relationship models or from information from structurally related substances (grouping or read-across)."

REACH has introduced a dramatic change in the EU regulatory framework - it explicitly provides the basis for the use of structure-activity relationship models, together with other "non-testing" approaches, for predicting the environmental and toxicological properties of chemicals, in the interests of time-effectiveness, cost-effectiveness and animal welfare. This change is increasingly being reflected in other pieces of EU legislation (Worth, 2010).

The science that aims to understand the relationships between chemical structure and the biological activity of molecules is evolving to support three distinct activities: category formation, read-across, and (Quantitative) Structure-Activity Relationship ([Q]SAR) analysis. A chemical category is a group of chemicals whose physicochemical and human health and/or environmental toxicological properties are likely to be similar or follow a regular pattern as a result of structural similarity. If this similarity is recognised with sufficient evidence, all the chemicals in the category can be assessed (and regulated) in the same way. Another approach to fill data gaps is read-across. In the read-across approach, endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered to be "similar" in some way (usually on the basis of structural similarity, but increasingly also on the basis of mechanistic or biological similarity).

Regarding the third approach, the scientific foundation of (Q)SAR models lies in physical organic chemistry, where chemical behaviour and activity are estimated solely from a knowledge of chemical structure. (Q)SAR modeling has been widely used in pharmacology, toxicology and physical chemistry, and its capabilities and limitations are relatively well understood (Hansch & Leo, 1995; Franke & Gruska, 2003; Worth et al., 2007).

The extensive use of estimation techniques such as (Q)SARs, read-across and grouping of chemicals, where appropriate and in a suitably constrained context, has the potential to effect reductions in the use of animals for toxicity assessment. At the same time, all these approaches need to be supported by adequate technological tools. Fortunately, the recent years have witnessed a dramatic progress in the field of manipulation of the chemical structure with computers, ranging from chemical relational databases, to calculation of chemical descriptors and derivation of qualitative and quantitative structure-activity relationships (Chen, 2006; Muchmore et al., 2010).

Among the software tools specifically aimed at supporting the (Q)SAR analyses in the regulatory assessment of chemicals is the expert system Toxtree. Toxtree (http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE) is an open-source, freely available software application that places chemicals into categories and predicts various kinds of toxic effects by applying various decision tree approaches. All estimation methods are structurally-based. The Toxtree platform was developed by Ideaconsult Ltd. (Sofia, Bulgaria) under the terms of a JRC contract (Worth et al., 2007). A range of modules (plug-ins) have also been developed by various contributors. In its present version (2.1.0; June 2011), it contains modules for estimating: a) oral toxicity (Cramer scheme); b) aquatic modes of action (Verhaar scheme); c) skin and eye irritation and corrosion; d) carcinogenicity and *in vitro* and *in vivo* mutagenicity; e) ability to act as Michael acceptors; f) persistence / biodegradation potential (START rulebase); and sites of cytochrome P450-mediated oxidation

All the above rulebases can be used independently to estimate endpoints or properties. Another use is to consider the various rulebases in combination for a wider profiling of the chemicals. In this report, we describe the implementation of a new rulebase for Toxtree, aimed at characterising chemicals in terms of organic functional groups. The rulebase consists of 204 organic functional groups recognized by the "Checkmol" program (http://merian.pch.univie.ac.at/~nhaider/cheminf/cmmm.html) which was developed by Dr Norbert Haider (University of Vienna, Austria).

2. The Functional Groups

The new rulebase identifies the classical organic functional groups (such as carbonyl, nitro or many others) present in the molecules, thus providing a tool to categorise and characterise the "chemical entities" under study (Feldman et al., 2005). The structural features are listed in Appendix 1.

In order to allow the users to discriminate between chemicals with higher and lower structure similarity, the functional groups have been divided into structural features with high specificity (HS) or low specificity (LS), nested hierarchically. For example, the low specificity feature FG75_LS (sulfonic acid derivatives) includes a broad range of compounds and has been further divided into a number of high specificity features: FG75_1_HS (sulfonic acid), FG75_2_HS (sulfonic acid ester), FG75_3_HS (sulfonamide). The high specificity features collect smaller sets of more closely related chemicals. A chemical can have simultaneously LS and HS features. By considering this information, the user can identify sub-categories of more mutually similar compounds and use these (qualitative) results for further analysis.

3. Use cases

3.1 Basic applications

The most immediate use of the Functional Group rulebase is the identification of similar chemicals. This forms the basis of the read-across procedure, which looks for a few chemicals with characteristics similar to those of a query chemical, whose missing data are to be extrapolated / interpolated from those of the "similar" chemicals. In a similar way, the category approach looks for chemicals to be grouped and assessed together. Another use of the Functional Group rulebase is to identify sets of congeneric chemicals that can be analysed together with a QSAR approach, such as the Hansch or extra-thermodynamic approach.

Whereas the Functional Group rulebase may provide the basic directions in finding similar chemicals, the simultaneous presence in Toxtree of a number of other estimation methods (e.g. Cramer, Verhaar, etc) permits a more sophisticated approach to such similarity searching for predictive toxicology. After a first categorisation through the Functional Group rulebase, it is possible to further sub-categorise by applying one or more of the other rulebases. This should be performed in a stepwise manner, where the functional groups provide the first clustering, and the hazard-based rulebase(s) refine the chemical

category by subdividing it into smaller clusters of chemicals with both chemical and toxicological similarity.

3.2 Comparing chemical databases

 The task of comparing databases of chemicals is of utmost importance. In particular, it provides a means of putting into context and rationalising the performance of a toxicological assay or QSAR when applied to different sets of chemicals. Without anchoring performance statistics to a definition or description of the tested chemicals, it is not possible to give a rational explanation as to why a certain assay or model appears to have a different predictive performance when applied to different sets of chemicals. The chemical space of a QSAR model training set forms the basis of the model applicability domain (Netzeva et al., 2005).

As an example, we show the use of the Functional Group rulebase to compare two databases: a) the classical database of chemicals tested in the rodent carcinogenicity bioassay; and b) a database studied recently within the US EPA's ToxCast Phase I exercise (Martin et al., 2010; Benigni et al., 2010).

The classical experimental carcinogenicity database is ISSCAN v3a (1141 unique chemicals) (Benigni et al., 2008). ISCCAN is available from the ISS website (http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7). The ToxCast data (309 unique chemicals) are contained in a dataset called: "ToxCast Phase I Data (AC50/LEC), downloadable from the ToxCast website (http://www.epa.gov/ncct/toxcast/data_sets.html).

The two sets of chemicals were combined together into an sdf file (with the exclusion of 46 chemicals common to the two databases), and the Functional Group rulebase was applied. A matrix, where each chemical was defined by the presence of the various functional groups, was obtained. The next step was the calculation of a distance matrix among all chemicals, based on the functional group profile. The Jaccard metric for similarity was used (Jaccard, 1912). The $n \times n$ distance matrix was then reduced to 10 Principal Components (PC) (explained variance: 0.81). The last step was the application of Canonical Discriminant Analysis to the PCs, with the aim to separate Toxcast from the ISSCAN chemicals.

Figure 1 displays the distribution of the ISSCAN and ToxCast chemicals along the direction of maximum dissimilarity between the two databases (Canonical Component 1, Squared Canonical Correlation = 0.21). It appears that the functional group composition of the two databases is largely overlapping, with exceptions at the extremes of the X axis.

ISSCAN *versus* ToxCast: functional groups
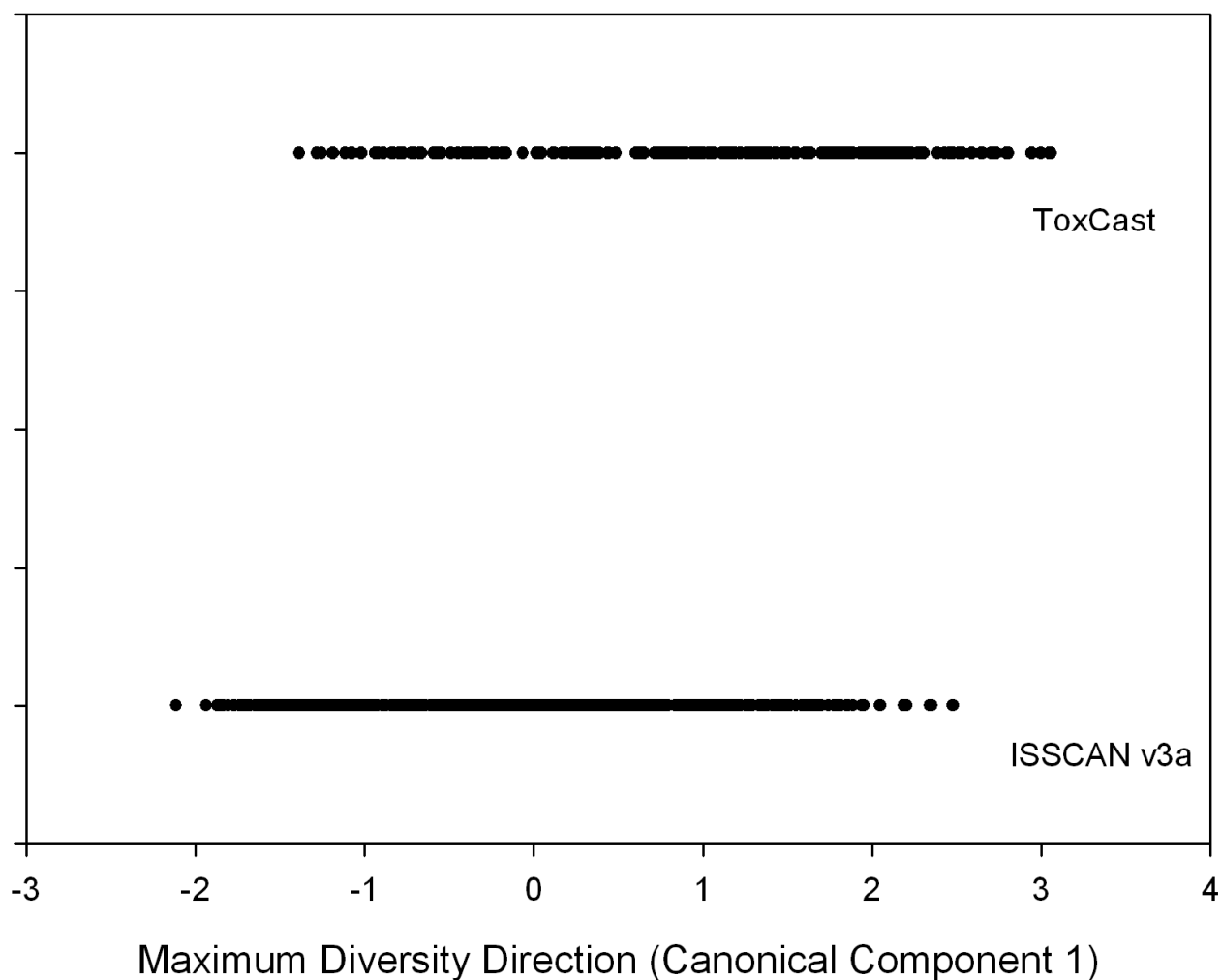
Figure 1 Functional group comparison of the ISSCAN and ToxCast datasets

In particular, ISSCAN contains a large subset of simple aromatic amines that are not present in ToxCast (low values of the Canonical Component, representative examples in Figure 2a), whereas ToxCast contains a subset of complex structures only present in this database (high values of the Canonical Component, representative examples in Figure 2b).
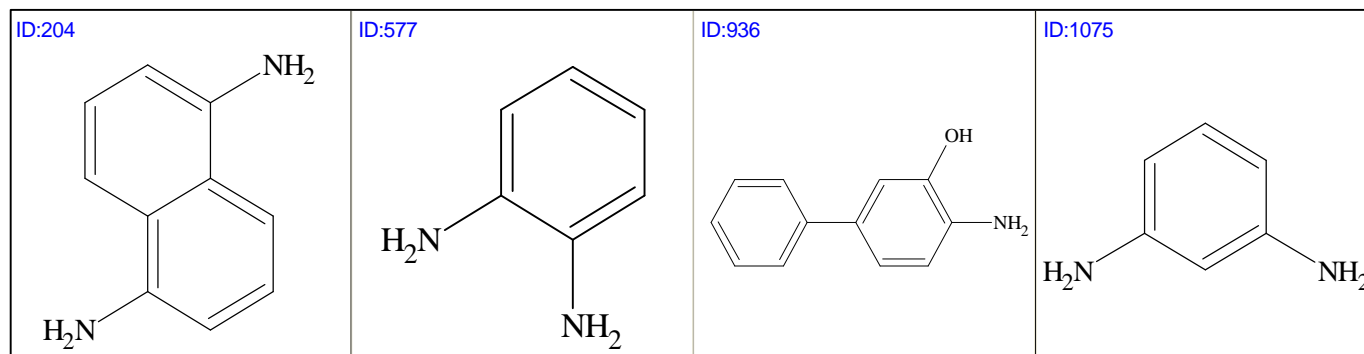
Figure 2a Examples of simple aromatic amines present in ISSCAN but not ToxCast
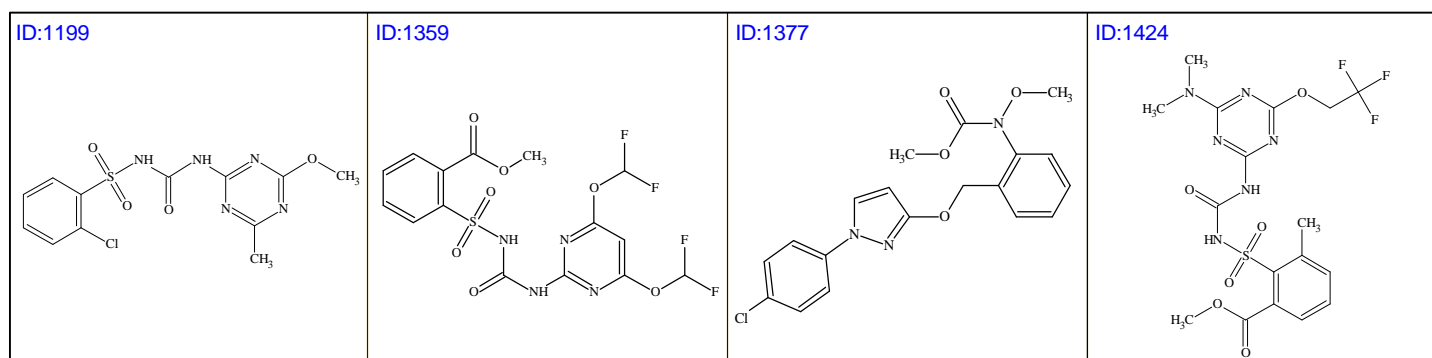
Figure 2b Examples of complex structures present in ToxCast but not ISSCAN

It is interesting to link this observation with observations on the prevalent toxicity mechanisms of the two databases of chemicals. In fact, a recent analysis has estimated that in ISSCAN v3a around 70% of carcinogens have Structural Alerts for genotoxic carcinogenicity, whereas in ToxCast only 35% of carcinogens are thought to act through genotoxic carcinogenicity mechanisms (Benigni et al., 2010). As a matter of fact, the majority of aromatic amines present in ISSCAN, but not in ToxCast (Figure 2a), act by genotoxic mechanisms (Benigni & Bossa, 2011). This shows that the analysis of the functional group distribution in a set of chemicals can provide a powerful means of comparing datasets in terms of their mechanistic toxicology.

4. Summary and Conclusions

In this report, we have presented a new rulebase for Toxtree which provides a means of identifying a set of 204 organic functional groups in query chemicals. The organic functional groups profiler can be used to screen chemical databases and identify "similar" chemicals for the purposes of read-across, category formation and QSAR analysis. It can also be used for the global comparison of chemicals databases / inventories, as illustrated in this study with two chemical databases – ISSCAN and ToxCast. It is anticipated that the organic functional groups profiler will provide a useful means of categorizing chemicals, especially when used in combination with other hazard-based profilers within Toxtree.
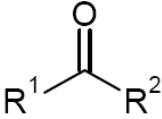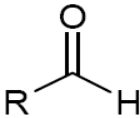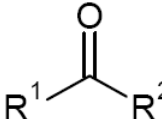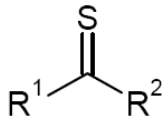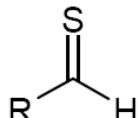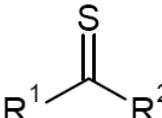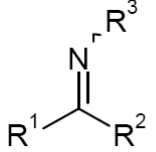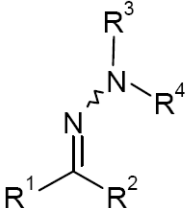
5. Acknowledgements and Disclaimer
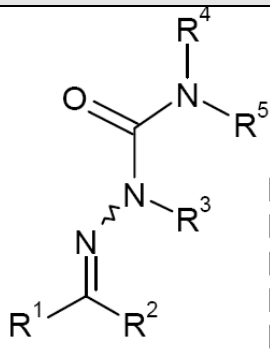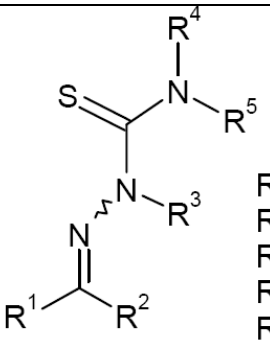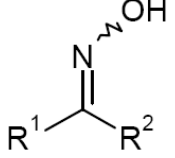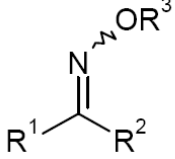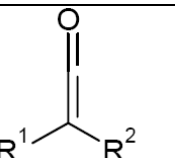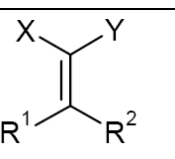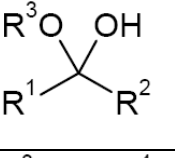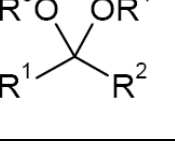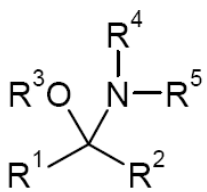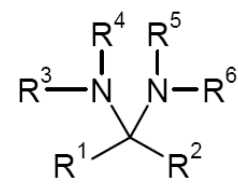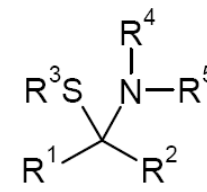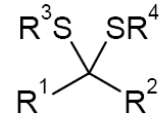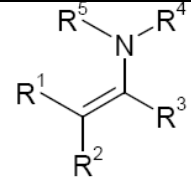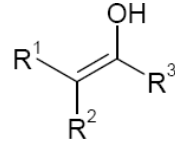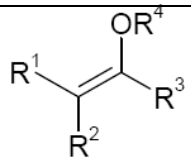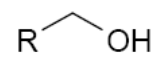
6. References
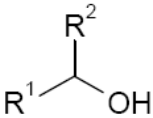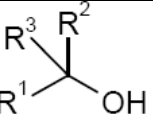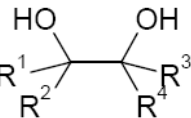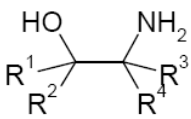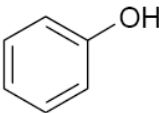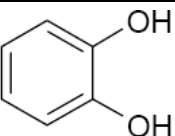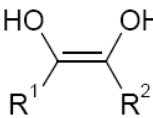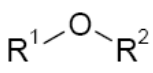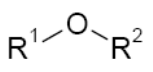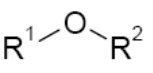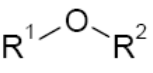
Benigni R & Bossa C (2011). Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chemical Reviews 111, 2507-2536.

Benigni R, Bossa C, Giuliani A & Tcheremenskaia O (2010). Exploring *In Vitro / In Vivo* Correlations: Lessons Learned from Analyzing Phase I Results of U.S. EPA's ToxCast Project. Journal of Environmental Science and Health, Part C: Environmental Carcinogenesis & Ecotoxicology Reviews 28, 272-286.

Benigni R, Bossa C, Richard AM & Yang C (2008). A novel approach: chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity. Annals of the Istituto Superiore di Sanità 44, 48-56. Available at: http://www.iss.it/binary/publ/cont/ANN_08_09%20Benigni.1209032841.pdf

Chen W (2006). Chemoinformatics: past, present, and future. Journal of Chemical Information and Modeling 46, 2230-2255.

European Commission (2006). Regulation (EC) No. 1907/2006 of the European parliament and of the council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 93/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union, vol L136, pp 103-106. Accessible at: http://echa.europa.eu/legislation/reach_legislation_en.asp

Feldman HJ, Dumontier M, Ling S, Haider N & Hogue CWV (2005). CO: a chemical ontology for identification of functional groups and semantic comparison of small molecules. FEBS Letters 579, 4685-4691.

Franke R & Gruska A (2003). General introduction to QSAR. In: Benigni R, editor. Quantitative Structure-Activity Relationhsip (QSAR) models of mutagens and carcinogens. Boca Raton: CRC Press. pp. 1-40.

Hansch C & Leo A (1995). Exploring QSAR. 1. Fundamentals and applications in chemistry and biology. Washington, D.C: American Chemical Society.

Jaccard P (1912). The distribution of the flora in the alpine zone. The New Phytologist XI, 37-50.

Martin MT, Dix DJ, Judson RS, Kavlock RJ, Reif DM, Richard AM, Rotroff DM, Romanov S, Medvedev A, Poltoratskaia N, Gambarian M, Moeser M, Makarov SS & Houck KA (2010). Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chemical Research in Toxicology 23, 578-590.

Muchmore SW, Edmunds JJ, Stewart KD & Hajduk PJ (2010). Cheminformatic tools for medicinal chemists. Journal of Medicinal Chemistry 53, 4830-4941.

Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Khan S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz G, Perkins R, Roberts DW, Schultz TW, Stanton ST, van de Sandt JJM, Tong W, Veith GD & Yang C (2005). Current status of methods for defining the applicability domain of (Quantitative) Structure-Activity Relationships. Alternatives to Laboratory Animals 33,1-19.

Worth AP, Bassan A, de Brujin J, Gallegos Saliner A, Netzeva TI, Pavan M, Patlewicz G, Tsakovska I & Eisenreich S (2007). The Role of the European Chemicals Bureau in Promoting the Regulatory Use of (Q)SAR Methods. SAR and QSAR in Environmental Research 18, 111-125.

Worth AP (2010). The role of QSAR methodology in the regulatory assessment of chemicals, Chapter 13 in Recent advances in QSAR Studies: Methods and Applications. T Puzyn, J Leszczynski & MTD Cronin (Eds). Heidelberg, Germany: Springer. pp. 367-382.

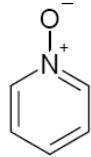# Appendix 1 Structural features incorporated into the Organic Functional Groups Profiler

No	Alert ID	Alert Title	Structure
1	FG1	cation	any positive charge
2	FG2	anion	any negative charge
3	FG3_LS	carbonyl compound: aldehyde or ketone	$R^1 = $ H, alkyl, aryl $R^2 = $ H, alkyl, aryl
4	FG3_1_HS	aldehyde	R = H, alkyl, aryl
5	FG3_2_HS	ketone	$R^1 = $ alkyl, aryl $R^2 = $ alkyl, aryl
6	FG4_LS	thiocarbonyl compound: aldehyde or ketone	$R^1 = $ H, alkyl, aryl $R^2 = $ H, alkyl, aryl
7	FG4_1_HS	thioaldehyde	R = H, alkyl, aryl
8	FG4_2_HS	thioketone	$R^1 = $ alkyl, aryl $R^2 = $ alkyl, aryl
9	FG5	imine	$R^1 = $ H, alkyl, aryl $R^2 = $ H, alkyl, aryl $R^3 = $ H, alkyl, aryl
10	FG6	hydrazone	$R^1 = $ H, alkyl, aryl $R^2 = $ H, alkyl, aryl $R^3 = $ H, alkyl, aryl $R^4 = $ H, alkyl, aryl
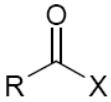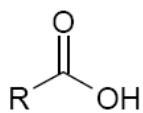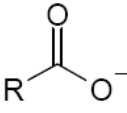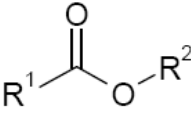
No	Alert ID	Alert Title	Structure
11	FG7	semicarbazone	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl, R^5 = H, alkyl, aryl
12	FG8	thiosemicarbazone	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl, R^5 = H, alkyl, aryl
13	FG9	oxime	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl
14	FG10	oxime ether	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = alkyl, aryl
15	FG11	ketene	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl
16	FG12	ketene acetal derivative	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl X = any hetero atom Y = any hetero atom
17	FG13	carbonyl hydrate	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl
18	FG14	hemiacetal	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = alkyl, aryl
19	FG15	acetal	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = alkyl, aryl R^4 = alkyl, aryl
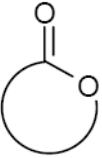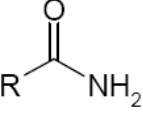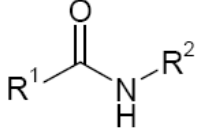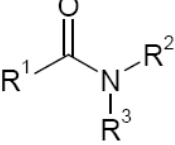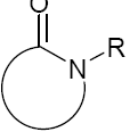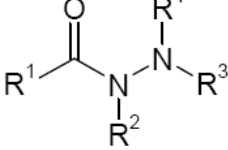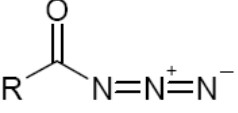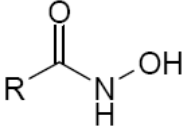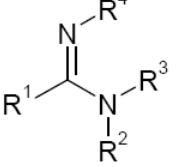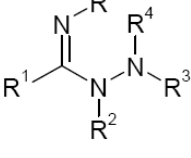
No	Alert ID	Alert Title	Structure
20	FG16	hemiaminal	R^1 = H, alkyl, aryl; R^2 = H, alkyl, aryl; R^3 = H, alkyl, aryl; R^4 = H, alkyl, aryl; R^5 = H, alkyl, aryl
21	FG17	aminal	R^1 = H, alkyl, aryl; R^2 = H, alkyl, aryl; R^3 = H, alkyl, aryl; R^4 = H, alkyl, aryl; R^5 = H, alkyl, aryl; R^6 = H, alkyl, aryl
22	FG18	thiohemiaminal	R^1 = H, alkyl, aryl; R^2 = H, alkyl, aryl; R^3 = H, alkyl, aryl; R^4 = H, alkyl, aryl; R^5 = H, alkyl, aryl
23	FG19	thioacetal	R^1 = H, alkyl, aryl; R^2 = H, alkyl, aryl; R^3 = alkyl, aryl; R^4 = alkyl, aryl
24	FG20	enamine	R^1 = H, acyl, alkyl, aryl; R^2 = H, acyl, alkyl, aryl; R^3 = H, acyl, alkyl, aryl; R^4 = H, acyl, alkyl, aryl; R^5 = H, acyl, alkyl, aryl
25	FG21	enol	R^1 = H, acyl, alkyl, aryl; R^2 = H, acyl, alkyl, aryl; R^3 = H, acyl, alkyl, aryl
26	FG22	enolether	R^1 = H, acyl, alkyl, aryl; R^2 = H, acyl, alkyl, aryl; R^3 = H, acyl, alkyl, aryl; R^4 = alkyl, aryl
27	FG23_LS	alcohol	R—OH R = alkyl, aryl
28	FG23_1_HS	primary alcohol	R = alkyl, aryl
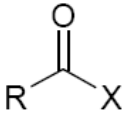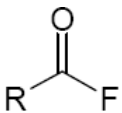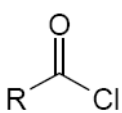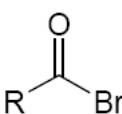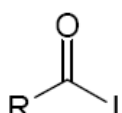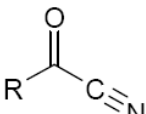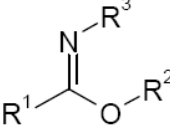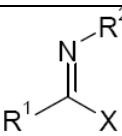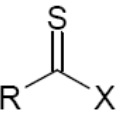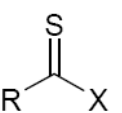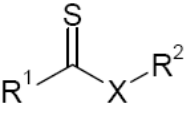
No	Alert ID	Alert Title	Structure
29	FG23_2_HS	secondary alcohol	R^1 = alkyl, aryl R^2 = alkyl, aryl
30	FG23_3_HS	tertiary alcohol	R^1 = alkyl, aryl R^2 = alkyl, aryl R^3 = alkyl, aryl
31	FG23_4_HS	1,2-diol	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl
32	FG23_5_HS	1,2-aminoalcohol	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl
33	FG23_6_HS	phenol	
34	FG23_7_HS	1,2-diphenol	
35	FG23_8_HS	enediol	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl
36	FG24_LS	ether	R^1—O—R^2 R^1 = alkyl, aryl R^2 = alkyl, aryl
37	FG24_1_HS	dialkylether	R^1—O—R^2 R^1 = alkyl R^2 = alkyl
38	FG24_2_HS	alkylarylether	R^1—O—R^2 R^1 = alkyl R^2 = aryl
39	FG24_3_HS	diarylether	R^1—O—R^2 R^1 = aryl R^2 = aryl
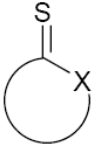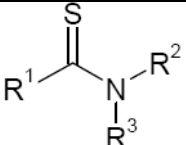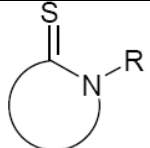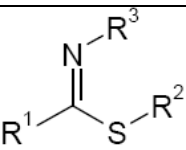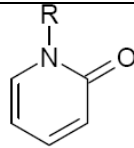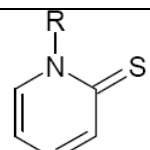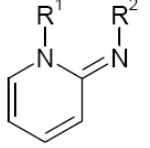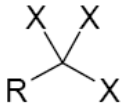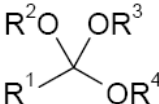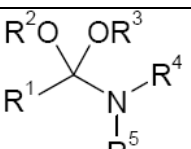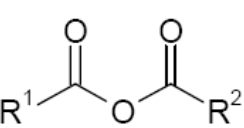
No	Alert ID	Alert Title	Structure
40	FG25	thioether	R^1–S–R^2 R^1 = alkyl, aryl R^2 = alkyl, aryl
41	FG26	disulfide	R^1–S–S–R^2 R^1 = alkyl, aryl R^2 = alkyl, aryl
42	FG27	peroxide	R^1–O–O–R^2 R^1 = alkyl, aryl R^2 = alkyl, aryl
43	FG28	hydroperoxide	R–O–OH R = alkyl, aryl
44	FG29	hydrazine derivative	R^1–N(R^2)–N(R^3)(R^4) R^1 = H, acyl, alkyl, aryl R^2 = H, acyl, alkyl, aryl R^3 = H, acyl, alkyl, aryl R^4 = H, acyl, alkyl, aryl
45	FG30	hydroxylamine	R^1–N(R^2)–O–R^3 R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl
46	FG31_LS	amine	R^1–N(R^3)–R^2 R^1 = alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl
47	FG31_1_HS	primary aliphatic amine	R–NH_2 R = alkyl
48	FG31_2_HS	primary aromatic amine	R–NH_2 R = aryl
49	FG31_3_HS	secondary aliphatic amine	R^1–NH–R^2 R^1 = alkyl R^2 = alkyl
50	FG31_4_HS	secondary mixed amine (aryl alkyl)	R^1–NH–R^2 R^1 = alkyl R^2 = aryl
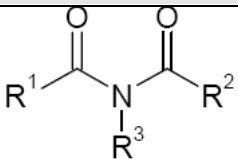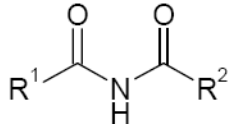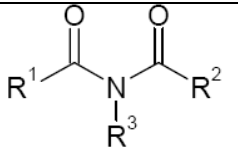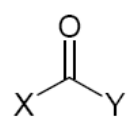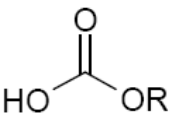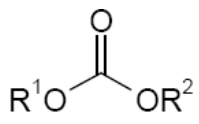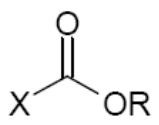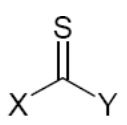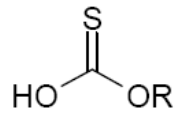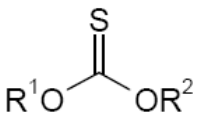
No	Alert ID	Alert Title	Structure
51	FG31_5_HS	secondary aromatic amine	R^1 $\overset{H}{\underset{}{N}}$ R^2 R^1 = aryl R^2 = aryl
52	FG31_6_HS	tertiary aliphatic amine	R^1 $\overset{R^3}{\underset{}{N}}$ R^2 R^1 = alkyl R^2 = alkyl R^3 = alkyl
53	FG31_7_HS	tertiary mixed amine	R^1 $\overset{R^3}{\underset{}{N}}$ R^2 R^1 = alkyl R^2 = aryl R^3 = alkyl, aryl
54	FG31_8_HS	tertiary aromatic amine	R^1 $\overset{R^3}{\underset{}{N}}$ R^2 R^1 = aryl R^2 = aryl R^3 = aryl
55	FG31_9_HS	quaternary ammonium salt	R^4, R^1 $\overset{R^3}{\underset{}{N^+}}$ R^2 R^1 = alkyl, aryl R^2 = alkyl, aryl R^3 = alkyl, aryl R^4 = alkyl, aryl
56	FG32	N-oxide	R^1 $\overset{O^-}{\underset{R^2}{N^+}}$ R^3 R^1 = alkyl, aryl R^2 = alkyl, aryl R^3 = alkyl, aryl
57	FG33_LS	halogen derivative (alkyl or aryl)	R—X X = F, Cl, Br, I R = alkyl, aryl
58	FG33_1_HS	alkyl fluoride	R—F R = alkyl
59	FG33_2_HS	alkyl chloride	R—Cl R = alkyl
60	FG33_3_HS	alkyl bromide	R—Br R = alkyl
61	FG33_4_HS	alkyl iodide	R—I R = alkyl
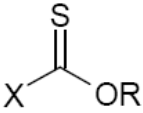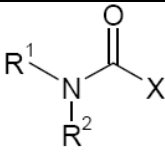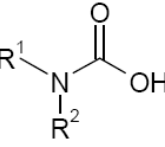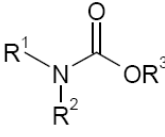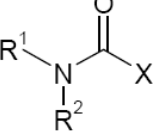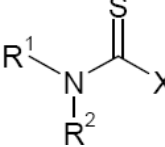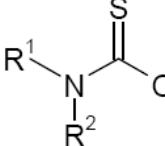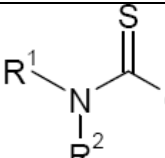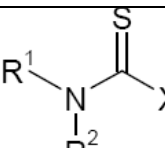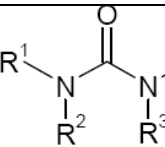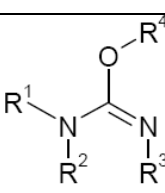
No	Alert ID	Alert Title	Structure
62	FG33_5_HS	aryl fluoride	R—F R = aryl
63	FG33_6_HS	aryl chloride	R—Cl R = aryl
64	FG33_7_HS	aryl bromide	R—Br R = aryl
65	FG33_8_HS	aryl iodide	R—I R = aryl
66	FG34_LS	organometallic compound	R—M M = any metal R = alkyl, aryl
67	FG34_1_HS	organolithium compound	R—Li R = alkyl, aryl
68	FG34_2_HS	organomagnesium compound	R—Mg R = alkyl, aryl
69	FG35_LS	carboxylic acid derivative	R = H, alkyl, aryl X = any hetero atom
70	FG35_1_HS	carboxylic acid	R = H, alkyl, aryl
71	FG35_2_HS	carboxylic acid salt	R = H, alkyl, aryl
72	FG35_3_HS	carboxylic acid ester	R^1 = H, alkyl, aryl R^2 = alkyl, aryl

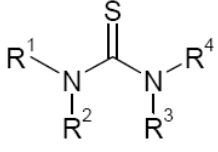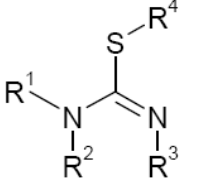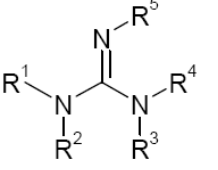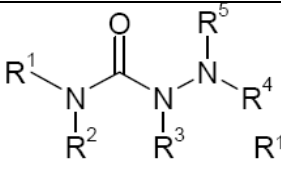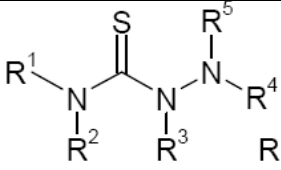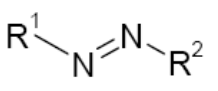| No	Alert ID	Alert Title	Structure
73	FG35_4_HS	lactone	
74	FG35_5_HS	carboxylic acid primary amide	R = H, alkyl, aryl
75	FG35_6_HS	carboxylic acid secondary amide	R^1 = H, alkyl, aryl R^2 = alkyl, aryl
76	FG35_7_HS	carboxylic acid tertiary amide	R^1 = H, alkyl, aryl R^2 = alkyl, aryl R^3 = alkyl, aryl
77	FG35_8_HS	lactam	R = H, alkyl, aryl
78	FG35_9_HS	carboxylic acid hydrazide	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl
79	FG35_10_HS	carboxylic acid azide	R = H, alkyl, aryl
80	FG35_11_HS	hydroxamic acid	R = H, alkyl, aryl
81	FG35_12_HS	carboxylic acid amidine	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl
82	FG35_13_HS	carboxylic acid amidrazone	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl R^4 = H, alkyl, aryl R^5 = H, alkyl, aryl
83	FG36	nitrile	R = H, alkyl, aryl

No	Alert ID	Alert Title	Structure
84	FG37_LS	acyl halide	R = H, alkyl, aryl; X = F, Cl, Br, I
85	FG37_1_HS	acyl fluoride	R = H, alkyl, aryl
86	FG37_2_HS	acyl chloride	R = H, alkyl, aryl
87	FG37_3_HS	acyl bromide	R = H, alkyl, aryl
88	FG37_4_HS	acyl iodide	R = H, alkyl, aryl
89	FG38	acyl cyanide	R = H, alkyl, aryl
90	FG39	imido ester	R^1 = H, alkyl, aryl; R^2 = alkyl, aryl; R^3 = H, alkyl, aryl
91	FG40	imidoyl halide	R^1 = H, alkyl, aryl; R^2 = H, alkyl, aryl; X = F, Cl, Br, I
92	FG41_LS	thiocarboxylic acid derivative	R = H, alkyl, aryl; X = any hetero atom
93	FG41_1_HS	thiocarboxylic acid	R = H, alkyl, aryl; X = OH, SH
94	FG41_2_HS	thiocarboxylic acid ester	R^1 = H, alkyl, aryl; R^2 = alkyl, aryl; X = O, S

No	Alert ID	Alert Title	Structure
95	FG41_3_HS	thiolactone	X = O, S
96	FG41_4_HS	thiocarboxylic acid amide	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl
97	FG41_5_HS	thiolactam	R = H, alkyl, aryl
98	FG42	imidothioester	R^1 = H, alkyl, aryl R^2 = alkyl, aryl R^3 = H, alkyl, aryl
99	FG43	oxohetarene	R = H, alkyl, aryl
100	FG44	thioxohetarene	R = H, alkyl, aryl
101	FG45	iminohetarene	R^1 = H, alkyl, aryl R^2 = H, alkyl, aryl
102	FG46_LS	orthocarboxylic acid derivative	R = H, alkyl, aryl X = OH, alkoxy, aryloxy, (substituted) amino, etc.
103	FG46_1_HS	carboxylic acid orthoester	R^1 = H, alkyl, aryl R^2, R^3, R^4 = alkyl, aryl
104	FG46_2_HS	carboxylic acid amide acetal	R^1 = H, alkyl, aryl R^2, R^3 = alkyl, aryl R^4, R^5 = H, alkyl, aryl
105	FG47	carboxylic acid anhydride	R^1, R^2 = H, alkyl, aryl

No	Alert ID	Alert Title	Structure
106	FG48_LS	carboxylic acid imide	R^1, R^2 = H, alkyl, aryl R^3 = H, alkyl, aryl, ...
107	FG48_1_HS	carboxylic acid unsubstituted imide	R^1, R^2 = H, alkyl, aryl
108	FG48_2_HS	carboxylic acid substituted imide	R^1, R^2 = H, alkyl, aryl R^3 = anything but H
109	FG49	CO_2 derivative (general)	any carbon with 4 valences to hetero atoms
110	FG50_LS	carbonic acid derivative	X, Y = any hetero atom
111	FG50_1_HS	carbonic acid monoester	R = alkyl, aryl
112	FG50_2_HS	carbonic acid diester	R^1, R^2 = alkyl, aryl
113	FG50_3_HS	carbonic acid ester halide	R = alkyl, aryl X = F, Cl, Br, I
114	FG51_LS	thiocarbonic acid derivative	X, Y = any hetero atom
115	FG51_1_HS	thiocarbonic acid monoester	R = alkyl, aryl
116	FG51_2_HS	thiocarbonic acid diester	R^1, R^2 = alkyl, aryl

No	Alert ID	Alert Title	Structure
117	FG51_3_HS	thiocarbonic acid ester halide	(structure) R = alkyl, aryl X = F, Cl, Br, I
118	FG52_LS	carbamic acid derivative	(structure) R^1, R^2 = H, alkyl, aryl X = OH, alkoxy, aryloxy, halogen
119	FG52_1_HS	carbamic acid	(structure) R^1, R^2 = H, alkyl, aryl
120	FG52_2_HS	carbamic acid ester (urethane)	(structure) R^1, R^2 = H, alkyl, aryl R^3 = alkyl, aryl
121	FG52_3_HS	carbamic acid halide	(structure) R^1, R^2 = H, alkyl, aryl X = F, Cl, Br, I
122	FG53_LS	thiocarbamic acid derivative	(structure) R^1, R^2 = H, alkyl, aryl X = OH, alkoxy, aryloxy, halogen
123	FG53_1_HS	thiocarbamic acid	(structure) R^1, R^2 = H, alkyl, aryl
124	FG53_2_HS	thiocarbamic acid ester	(structure) R^1, R^2 = H, alkyl, aryl R^3 = alkyl, aryl
125	FG53_3_HS	thiocarbamic acid halide	(structure) R^1, R^2 = H, alkyl, aryl X = F, Cl, Br, I
126	FG54	urea	(structure) R^1, R^2, R^3, R^4 = H, alkyl, aryl
127	FG55	isourea	(structure) R^1, R^2, R^3, R^4 = H, alkyl, aryl

No	Alert ID	Alert Title	Structure
128	FG56	thiourea	R^1, R^2, R^3, R^4 = H, alkyl, aryl
129	FG57	isothiourea	R^1, R^2, R^3, R^4 = H, alkyl, aryl
130	FG58	guanidine	R^1, R^2, R^3, R^4, R^5 = H, alkyl, aryl
131	FG59	semicarbazide	R^1, R^2, R^3, R^4, R^5 = H, alkyl, aryl
132	FG60	thiosemicarbazide	R^1, R^2, R^3, R^4, R^5 = H, alkyl, aryl
133	FG61	azide	R—N=N$^+$=N$^-$ R = alkyl, aryl
134	FG62	azo compound	R^1, R^2 = alkyl, aryl
135	FG63	diazonium salt	R—N=N$^+$ R = alkyl, aryl
136	FG64	isonitrile	R—N$^+$≡C$^-$ R = alkyl, aryl
137	FG65	cyanate	RO—C≡N R = alkyl, aryl

No	Alert ID	Alert Title	Structure
138	FG66	isocyanate	$R-N=C=O$ R = alkyl, aryl
139	FG67	thiocyanate	$RS-C\equiv N$ R = alkyl, aryl
140	FG68	isothiocyanate	$R-N=C=S$ R = alkyl, aryl
141	FG69	carbodiimide	$R^1-N=C=N-R^2$ R^1, R^2 = H, alkyl, aryl
142	FG70	nitroso compound	$R-N=O$ R = alkyl, aryl
143	FG71	nitro compound	$R-N(=O)=O$ R = alkyl, aryl
144	FG72	nitrite	$RO-N=O$ R = alkyl, aryl
145	FG73	nitrate	$RO-N(=O)=O$ R = alkyl, aryl
146	FG74_LS	sulfuric acid derivative	$X-S(=O)(=O)-Y$ X, Y = any hetero atom
147	FG74_1_HS	sulfuric acid	$HO-S(=O)(=O)-OH$
148	FG74_2_HS	sulfuric acid monoester	$RO-S(=O)(=O)-OH$ R = alkyl, aryl

No	Alert ID	Alert Title	Structure
149	FG74_3_HS	sulfuric acid diester	$R^1O-S-OR^2$ (with two $=O$) R^1, R^2 = alkyl, aryl
150	FG74_4_HS	sulfuric acid amide ester	R^1O-S-N with R^3, R^2 (two $=O$) R^1 = alkyl, aryl; R^2, R^3 = H, alkyl, aryl
151	FG74_5_HS	sulfuric acid amide	$HO-S-N$ with R^2, R^1 (two $=O$) R^1, R^2 = H, alkyl, aryl
152	FG74_6_HS	sulfuric acid diamide	$R^1, R^2-N-S-N-R^4, R^3$ (two $=O$) R^1, R^2, R^3, R^4 = H, alkyl, aryl
153	FG74_7_HS	sulfuryl halide	$X-S-Y$ (two $=O$) X = F, Cl, Br, I; Y = any hetero atom
154	FG75_LS	sulfonic acid derivative	$R-S-X$ (two $=O$) R = alkyl, aryl; X = any hetero atom
155	FG75_1_HS	sulfonic acid	$R-S-OH$ (two $=O$) R = alkyl, aryl
156	FG75_2_HS	sulfonic acid ester	R^1-S-OR^2 (two $=O$) R^1, R^2 = alkyl, aryl
157	FG75_3_HS	sulfonamide	R^1-S-N with R^3, R^2 (two $=O$) R^1 = alkyl, aryl; R^2, R^3 = H, alkyl, aryl
158	FG75_4_HS	sulfonyl halide	$R-S-X$ (two $=O$) R = alkyl, aryl; X = F, Cl, Br, I
159	FG76	sulfone	R^1-S-R^2 (two $=O$) R^1, R^2 = alkyl, aryl

No	Alert ID	Alert Title	Structure
160	FG77	sulfoxide	R^1—S(=O)—R^2 R^1, R^2 = alkyl, aryl
161	FG78_LS	sulfinic acid derivative	R—S(=O)—X R = alkyl, aryl X = any hetero atom
162	FG78_1_HS	sulfinic acid	R—S(=O)—OH R = alkyl, aryl
163	FG78_2_HS	sulfinic acid ester	R^1—S(=O)—OR^2 R^1, R^2 = alkyl, aryl
164	FG78_3_HS	sulfinic acid halide	R—S(=O)—X R = alkyl, aryl X = F, Cl, Br, I
165	FG78_4_HS	sulfinic acid amide	R^1—S(=O)—N(R^2)—R^3 R^1 = alkyl, aryl R^2, R^3 = H, alkyl, aryl
166	FG79_LS	sulfenic acid derivative	R—S—X R = alkyl, aryl X = any hetero atom
167	FG79_1_HS	sulfenic acid	R—S—OH R = alkyl, aryl
168	FG79_2_HS	sulfenic acid ester	R^1—S—OR^2 R^1, R^2 = alkyl, aryl
169	FG79_3_HS	sulfenic acid halide	R—S—X R = alkyl, aryl X = F, Cl, Br, I
170	FG79_4_HS	sulfenic acid amide	R^1—S—N(R^2)—R^3 R^1 = alkyl, aryl R^2, R^3 = H, alkyl, aryl

No	Alert ID	Alert Title	Structure
171	FG80_LS	thiol	R—SH R = alkyl, aryl
172	FG80_1_HS	alkylthiol	R—SH R = alkyl
173	FG80_2_HS	arylthiol	R—SH R = aryl
174	FG81_LS	phosphoric acid derivative	X, Y, Z = O, N, Hal residue
175	FG81_1_HS	phosphoric acid	
176	FG81_2_HS	phosphoric acid ester	R = alkyl, aryl; X, Y = any O, N, Hal residue
177	FG81_3_HS	phosphoric acid halide	X = F, Cl, Br, I; Y, Z = any O, N, Hal residue
178	FG81_4_HS	phosphoric acid amide	R^1, R^2 = H, alkyl, aryl; X, Y = any O, N, Hal residue
179	FG82_LS	thiophosphoric acid derivative	X, Y, Z = any O, N, Hal residue
180	FG82_1_HS	thiophosphoric acid	
181	FG82_2_HS	thiophosphoric acid ester	R = alkyl, aryl; X, Y = any O, N, Hal residue

No	Alert ID	Alert Title	Structure
182	FG82_3_HS	thiophosphoric acid halide	$X = F, Cl, Br, I$ Y, Z = any O, N, Hal residue
183	FG82_4_HS	thiophosphoric acid amide	R^1, R^2 = H, alkyl, aryl X, Y = any O, N, Hal residue
184	FG83_LS	phosphonic acid derivative	R = alkyl, aryl X, Y = any O, N, Hal residue
185	FG83_1_HS	phosphonic acid	R = alkyl, aryl
186	FG83_2_HS	phosphonic acid ester	R^1, R^2 = alkyl, aryl X = any O, N, Hal residue
187	FG83_3_HS	phosphine	R^1, R^2, R^3 = alkyl, aryl
188	FG83_4_HS	phosphinoxide	R^1, R^2, R^3 = alkyl, aryl
189	FG84_LS	boronic acid derivative	R = alkyl, aryl X, Y = any O, N, Hal residue
190	FG84_1_HS	boronic acid	R = alkyl, aryl
191	FG84_2_HS	boronic acid ester	R^1, R^2 = alkyl, aryl X = any O, N, Hal residue
192	FG85	alkene	R^1, R^2, R^3, R^4 = H, alkyl, aryl

No	Alert ID	Alert Title	Structure
193	FG86	alkyne	R^1————R^2 R^1, R^2 = H, alkyl, aryl
194	FG87	aromatic compound	any aromatic carbocyclic or heterocyclic structure, including cyclopentadienyl anion, tropylium cation, fulvene, tropone, pyridone-type lactams, etc.
195	FG88	heterocyclic compound	any cyclic structure with at least one non-carbon atom incorporated
196	FG89	alpha-aminoacid	R^1, R^2 = H, alkyl, aryl
197	FG90	alpha-hydroxyacid	R = H, alkyl, aryl

Abstract

Toxtree is a freely available, user-friendly and extensible software application that is designed to make structure-based predictions for a number of toxicological endpoints and mechanisms of chemical action. The platform has been developed by the Joint Research Centre in collaboration with Ideaconsult Ltd (Sofia, Bulgaria) with a range of modules developed by various contributors. One of the modules developed as an extension to Toxtree is aimed at the identification of organic functional groups in query chemicals. The rulebase consists of 204 organic functional groups recognised by the "Checkmol" program, which was developed by Dr Norbert Haider, University of Vienna. A new Functional Group Profiler has been coded as a Toxtree module by the Istituto Superiore di Sanita' (Rome, Italy). The Toxtree profiler, called ISSFUNC, can be used to screen and characterise chemicals as a basis for read-across, category formation and (Q)SAR analysis. It can also be used for the global comparison of datasets, such as model training and test sets and chemical inventories.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-24871-EN-N

JRC
EUROPEAN COMMISSION

Publications Office