Title: Investigation of the Verhaar scheme for predicting acute aquatic toxicity: improving predictions obtained from Toxtree ver. 2.6

Authors: Claire M. Ellison, Judith C. Madden, Mark. T. D. Cronin and Steven J. Enoch

Corresponding Author: Steven J. Enoch, s.j.enoch@ljmu.ac.uk, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, +44 151 231 2164

Abstract:

Assessment of the potential of compounds to cause harm to the aquatic environment is an integral part of the REACH legislation. To reduce the number of vertebrate and invertebrate animals required for this analysis alternative approaches have been promoted. Category formation and read-across have been applied widely to predict toxicity. A key approach to grouping for environmental toxicity is the Verhaar scheme which uses rules to classify compounds into one of four mechanistic categories. These categories provide a mechanistic basis for grouping and any further predictive modelling. A computational implementation of the Verhaar scheme is available in Toxtree v2.6. The work presented herein demonstrates how modifications to the implementation of Verhaar between version 1.5 and 2.6 of Toxtree have improved performance by reducing the number of incorrectly classified compounds. However, for the datasets used in this analysis, version 2.6 classifies more compounds as outside of the domain of the model. Further amendments to the classification rules have been implemented here using a post-processing filter encoded as a KNIME workflow. This results in fewer compounds being classified as outside of the model domain, further improving the predictivity of the scheme. The utility of the modification described herein is demonstrated through building quality, mechanism-specific Quantitative Structure Activity Relationship (QSAR) models for the compounds within specific mechanistic categories.

Keywords: Verhaar; Toxtree; Aquatic Toxicity; QSAR; Category formation

Highlights:

- The Verhaar scheme as implemented in Toxtree v2.6 has improved performance; results here show 35% fewer compounds misclassified
- The modified Verhaar scheme (Toxtree v2.6) correctly classifies 42% of compounds in test datasets
- A KNIME post-processing filter improves the scheme further resulting in 63% of compounds correctly classified
- QSAR models have been built from compounds in the resultant categories

1. Introduction

Aquatic toxicity studies have traditionally been performed using a variety of vertebrate and invertebrate animals (Walker et al., 1991; Traas and van Leeuwen, 2007). The European REACH legislation (EC, 2006) has required companies to assess fully and report the environmental risks of compounds manufactured or imported in significant quantities (i.e. greater than or equal to one tonne per annum), and hence potentially requiring many tests (Schaafsma et al., 2009). However, alternative approaches have been promoted throughout the implementation of REACH and much research has been published in this area (e.g. (Jaworska et al., 2010; de Haas et al., 2011; Pery et al., 2013; Scholz et al., 2013; Gissi et al., 2014; Patlewicz et al., 2014).

One key aspect of alternative methods is that they should be mechanistically interpretable (McKim et al., 1987). This enables methods to be transparent, credible and supports validation and regulatory acceptance. With regard to applying a mechanistic framework to environmental toxicants, Verhaar and co-workers devised a scheme to assist with the allocation of potential environmental pollutants into mechanisms of action (Verhaar et al., 1992). The scheme utilises 2D chemical structure to classify potential environmental pollutants into one of four categories representing one, or more, mechanisms of action: Class 1 (narcosis or baseline toxicity), Class 2 (less inert compounds), Class 3 (unspecific reactivity) and Class 4 (compounds and groups of compounds acting by a specific mechanism). Grouping potentially allows for predictions of acute toxicity to be made from QSARs (Cronin, 2006), or to establish whether further information may be required for read-across purposes (Koleva et al., 2008). For example, the toxicity of Class 1 and 2 compounds can be predicted using hydrophobicity alone and further testing may not be required. Conversely for classes 3 and 4, the classification scheme is a simple and efficient method to quickly highlight compounds of concern where testing, further research and read-across approaches, possibly within an Integrated Approach to Testing and Assessment (IATA) strategy, may be more appropriate.

Compounds acting as baseline narcotics (Class 1) include saturated aliphatic alcohols and ketones (Ellison et al., 2008). Their mechanism of action has been hypothesised to be related to their ability to accumulate within biological membranes (Roth, 1980). It is possible to predict acute toxic potency values of these compounds using a relevant hydrophobicity (logarithm of the octanol:water partition coefficient (log P)) dependent QSAR model (Könemann, 1981; Veith et al., 1983; Schultz and Tichy, 1993) and there are also indications this relationship may hold for chronic toxicity (Austin and Eadsforth, 2014). This mechanism is termed the ‘baseline’ as all compounds have the potential to act as narcotics, but compounds can show excess toxicity (i.e. a level of toxicity higher than that predicted using hydrophobicity alone) because they contain chemical substructures which facilitate specific mechanism(s).

Compounds acting as polar narcotics (Class 2) exhibit toxicity above the baseline, but can still be modelled using hydrophobicity alone. It has been argued that there is no mechanistic difference between baseline and polar narcotics (Vaes et al., 1998; Escher et al., 2002) but further analysis has subsequently shown that there is a physiological difference between the two mechanisms (Roberts and Costello, 2003) as well as historical evidence from the definition and experimental determination of Fish Acute Toxicity Syndromes (FATS) (McKim et al., 1987). Therefore it is preferable for modelling purposes to treat the two narcosis mechanisms separately (Ellison et al., 2008; Nendza et al., 2014; Su et al., 2014). Roberts and Costello (2003) proposed that the mechanistic difference between the two classes is caused by the hydrophilic, 'polar', part of a compound remaining in the aqueous environment at the outer part of a biological membrane which then limits the compound's position in the membrane. In contrast baseline narcotics pass through fully into the centre of the membrane, where they then accumulate. Thus, compounds in Class 2 should contain a dipole moment significant enough to create distinctive hydrophilic areas within a molecule, which may be brought about by hydrogen bond interactions from an aromatic hydroxyl or amino group.

Compounds acting by reactive (Class 3) mechanisms include those containing specific electrophilic moieties that enable the compound to react with nucleophilic sites on biological macromolecules. These compounds can only be modelled using hydrophobicity alone when there is consistency in the reactivity i.e. a group of compounds with the same reactive functional group but varying chain length; however the addition of an electronic descriptor within specific electrophilic mechanisms can create useful models (Netzeva and Schultz, 2005; Schultz et al., 2007). Also included in this class are molecules which can undergo bioactivation into an electrophilic compound (Hermens, 1990; Lipnick, 1991).

The final set of compounds defined by Verhaar et al, those acting via a specific mechanism (Class 4), is a diverse group which covers all molecules that exhibit toxicity via interactions with certain receptor mediated events. Examples of compounds in this class include organic phosphorus esters which inhibit acetylcholinesterase (Verhaar et al., 1992), and aromatic compounds which can act as weak acid respiratory uncouplers of oxidative phosphorylation (Schultz and Cronin, 1997).

The classes defined by Verhaar et al, therefore, have the potential to group compounds into mechanistically relevant categories to aid modelling, read-across and hence hazard assessment. The Verhaar scheme for classification of environmental pollutants has been coded into a number of pieces of software, with little development or incorporation of new knowledge. In 2008 Enoch and co-workers evaluated the performance of the Verhaar scheme as implemented in the software Toxtree ver. 1.5 (Enoch et al., 2008). A number of misclassifications were noted, and as a result improvements were suggested. It was proposed that these could be achieved by reordering the rules in the system

and implementing additional rules to identify compounds in Classes 3 and 4, as well as refining some of the existing rules.

Since the publication by Enoch and co-workers in 2008, updates have been made to the Toxtree software and the current version (2.6) is freely available to download (http://Toxtree.sourceforge.net). The aim of the work presented here was to examine the utility of the Verhaar scheme as implemented in Toxtree version 2.6 to assign compounds to the correct mechanism of action as defined by well-established datasets, and to determine whether additional rules would be beneficial to classify compounds correctly. Improvement to the classification performance would aid grouping by the creation of more robust mechanistically interpretable categories thus enabling better and more robust prediction of toxicity.

2. Methods

2.1 Datasets

The data used to assess the performance of the Verhaar scheme as implemented in Toxtree ver. 2.6 were acquired from the supplementary information of Enoch et al (2008). The same data were used so that a direct comparison could be made between Toxtree versions 1.5 (as used by Enoch et al, 2008) and 2.6 (see below). The supplementary information comprised two datasets: a set of 408 compounds tested using *Pimephales promelas* and assigned to mechanisms of action (Russom et al., 1997) and a set of 250 compounds tested using *Tetrahymena pyriformis* and also assigned mechanisms of action (Schultz et al., 1997). The information included: compound names; SMILES strings; toxicity values (LC_{50} and IGC_{50} respectively); assigned mechanism of action (details below); expected Verhaar scheme class (based on assigned mechanism); and Toxtree v1.5 classification. It was assumed that all information provided was correct; no quality analysis was performed on the data and Toxtree v1.5 predictions were not repeated. However it was noted that 23 compounds from the *T. pyriformis* dataset had been recorded against Class 3, whereas their mechanism was actually Class 4. This typographical error was corrected before the data were used.

The *P. promelas* dataset included compounds exhibiting the following mechanisms of toxicity: baseline narcosis (239 chemicals); polar narcosis (36 chemicals); reactive via electrophilic mechanisms (96 chemicals); respiratory uncoupling (12 chemicals); acetylcholinesterase inhibition (16 chemicals); and central nervous system seizure (9 chemicals). A combination of the assessment of whether a chemical exhibited excess toxicity (compared to that which would be predicted if the chemical was a baseline narcotic) together with the presence of structural features known to cause excess toxicity and experimental analysis (behavioural, dose–response and toxicodynamic profiling) was used previously to assign mechanisms of action (Russom et al., 1997).

The *T. pyriformis* dataset included compounds exhibiting the following mechanisms of toxicity: polar narcosis (173 chemicals); reactive via electrophilic mechanisms (27 chemicals); reactive via pro-electrophilic mechanisms (i.e. metabolic activation is required; 27 chemicals); respiratory uncoupling (19 chemicals); and pro-redox cycling (4 chemicals). These mechanisms were previously assigned based on clusters of chemicals identified in a 3D toxic response surface (energy of the Lowest Unoccupied Molecular Orbital (E_{LUMO}), logarithm of the octanol:water partition coefficient (logP) and the inverse logarithm of the 50% Inhibitory Growth Concentration (log IGC_{50}^{-1})). Clusters of chemicals were observed within broad ranges of E_{LUMO} values, where chemicals with lower E_{LUMO} values were classified as potential soft electrophiles, whilst chemicals with higher E_{LUMO} values were classified as polar narcotics. The metabolically converted pro-electrophiles, weak acid respiratory uncouplers and pro-redox cyclers were assigned based on the presence of known structural features and E_{LUMO} values (Schultz et al., 1997). As this dataset did not contain any baseline, non-polar narcotics additional data were included from another publication to ensure all mechanistic categories were represented in both species (Ellison et al., 2008). The data from Ellison et al (2008) comprised the toxicity (log IGC_{50}^{-1}), log P, SMILES and CAS numbers of 64 alcohols and 23 ketones which are accepted to act as baseline narcotics (Könemann, 1981; Veith et al., 1983), and had been tested in the same *T. pyriformis* assay (Schultz, 1997) as the phenols published by Schultz and co-workers (Schultz et al., 1997). The full compilation of all data for both species is available as supplementary information to this paper (which also includes the classifications from Toxtree 1.5, Toxtree 2.6 and the KNIME post-processing filter).

2.2 Software

Toxtree was developed by Ideaconsult Ltd (Sofia, Bulgaria) under the terms of a contract from the European Commission Joint Research Centre (JRC). The software encodes several decision trees and classification schemes useful for analysing the potential toxicity hazards of compounds (Pavan and Worth, 2008). The software is freely available (http://Toxtree.sourceforge.net) and the current version (2.6) includes an updated encoding of the Verhaar scheme under the title "Verhaar scheme (Modified)". All 745 compounds described above were classified using the "Verhaar scheme (Modified)" decision tree through the batch processing functionality of Toxtree v2.6. Additionally the 87 non-polar narcotics taken form Ellison et al (2008) were also processed through the "Verhaar scheme" in Toxtree ver 1.5 to enable to comparison of classifications for these compounds. Structures were entered as SDfiles which were generated from the SMILES strings using MarvinBeans v14 (www.chemaxon.com). The possible outcomes from the scheme have not altered between versions: 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 rules). The first four classes directly relate to the Verhaar classes described above whereas Class 5 can be considered as "out of domain".

KNIME is a freely available analytics platform that allows processes and workflows to be easily encoded (www.knime.org). After the compounds had been processed through Toxtree v2.6 and the data had been analysed (see below) a KNIME workflow was developed to act as a post-processing filter to Toxtree. The aim of the filter was to expand the domain of the Verhaar scheme so that fewer compounds were placed into Class 5.

2.3 Data analysis

The classifications produced by the modified Verhaar scheme as implemented in Toxtree 2.6 were compared with the original expert assigned mechanisms of action for all compounds to assess the performance of the software. If a compound was classified into Class 1, 2, 3 or 4 and this matched the assigned mechanism then this was considered a correct classification, whereas if the class did not match then this was considered an incorrect classification. If a compound was placed into Class 5 then the compound was considered to be outside of the domain of the scheme. The performance of the scheme for each class was assessed by calculating the Positive Predictivity Value (PPV) within each category using the following equation:

$$PPV = Nc / (Nc + Ni)$$

Where Nc is the number of compounds correctly classified and Ni is the number of compounds incorrectly classified.

The performance of the modified Verhaar scheme (as implemented in Toxtree v2.6) was compared to the performance of the Verhaar scheme as implemented in Toxtree v1.5, as reported by Enoch and co-workers (Enoch et al., 2008). Enoch and co-workers did not report the PPV values so these were calculated in the same manner using the data provided in the supplementary information.

The results from Toxtree v2.6 were further analysed to identify possible improvements and refinements that could be made to the system. To this end the compounds which were out of the domain of the model (Class 5) were examined to see if they could be made classifiable by the scheme, through refinement of the existing rules. This was performed manually where expert judgement was used for each compound to assess whether it was truly out of the domain of the model, or if it should have been classifiable using an existing rule. The definitions of the rules used in this process were those found in Toxtree software under the menu ‘Method | View decision tree’ along with the more in-depth perspective offered by the original Verhaar publication (Verhaar et al., 1992). Where three or more compounds could be classified correctly by the modification of an existing rule, then this modified rule was implemented in a post-processing filter using the KNIME software. More specifically, structural filters were written using SMARTS patterns in the RD kit ‘Substructure

Structure Filter' node. A minimum of three compounds was utilised to reduce the risk of over-fitting the scheme for the specific compounds present in the datasets used in this analysis.

To assess the utility of the final classifications once all improvements had been implemented, it was investigated whether it would be possible to develop QSARs within a class of compounds all acting via the same mechanism. To this end QSARs were developed for the baseline and polar narcosis as these should be well modelled by log P alone.

3. Results and Discussion

In 2008 Enoch and co-workers assessed the utility of the Verhaar scheme in Toxtree (v1.5) and provided recommendations for improved implementation of the scheme (Enoch et al., 2008). Since then, Toxtree has been updated and version 2.6 has been modified with consideration of the improvements suggested by Enoch and co-workers. This study examined the effect of these modifications using the same datasets as Enoch and co-workers; that is the 408 compounds tested in *Pimephales promelas* (Russom et al., 1997) and 250 phenols tested in *Tetrahymena pyriformis* (Schultz et al., 1997), as well as 87 compounds classified as baseline narcotics and tested in *T. pyriformis* (Ellison et al., 2008). All 745 compounds were assigned a 'true' mechanistic class previously (Russom et al., 1997; Schultz et al., 1997; Ellison et al., 2008) and it was against this classification that the performance of Toxtree 2.6 was assessed. Thus a compound was considered to have received a 'correct' classification when the previously assigned mechanism matched the classification provided by Toxtree.

The comparison of results between the Verhaar scheme as implemented in Toxtree versions 1.5 and 2.6 is available as supplementary information for all data used in this analysis. An initial inspection of these results indicates a marginal improvement; 45% of all compounds were correctly classified in Toxtree v1.5 (note that this figure is an improvement on the figure published by Enoch and co-workers [38%] because of the addition of the 87 baseline narcotics tested in *T. pyriformis*) and in Toxtree v2.6 this figure raises to 49%. However, when considering the number of misclassified compounds, the modified version is significantly outperforming the previous version (196 misclassifications in Toxtree v1.5 compared to 126 misclassifications in Toxtree v2.6; a reduction of 35%). The positive predictivity value of the Verhaar scheme in Toxtree 1.5 was 0.63, whereas the scheme in Toxtree 2.6 has a value of 0.74 thus showing a significant improvement.

The improvement in the classifications provided by Toxtree 2.6 is also apparent when examining the individual groups of compounds, as shown in Table 1. The positive predictivity value (PPV) is greater than 0.7 for three of the four classes, and the PPV value for Class 3 has improved from 0.34 to 0.57. Thus the scheme now performs better over a wider range of mechanisms rather than only performing well when identifying baseline narcotics. The number of compounds correctly classified as Class 1

has increased from 158 to 182. However a greater number of compounds overall now fall into Class 1 and thus the PPV has fallen from 0.95 to 0.83. It is clear that although improvements to the Verhaar classification within Toxtree have occurred between versions 1.5 and 2.6, the performance when using these datasets could be improved further.

Table 1: Number of compounds (in)correctly classified for each class in the Verhaar scheme as implemented in Toxtree versions 1.5 and 2.6, and with additional post-processing filters (Fig. 3) where PPV is the Positive Predictive Value.

		Toxtree 1.5			Toxtree 2.6			Additional Post-processing filter		
		Correct	Incorrect	PPV	Correct	Incorrect	PPV	Correct	Incorrect	PPV
T. pyriformis	Class 1	84	0	1.00	84	25	0.77	87	0	1.00
	Class 2	79	20	0.80	79	19	0.81	152	27	0.85
	Class 3	9	63	0.13	5	17	0.23	32	21	0.60
	Class 4	0	0	N/A	1	1	0.5	9	5	0.64
P. promelas	Class 1	74	8	0.90	98	12	0.89	100	10	0.91
	Class 2	21	23	0.48	19	17	0.53	21	29	0.42
	Class 3	67	82	0.45	56	29	0.66	58	31	0.65
	Class 4	0	0	N/A	20	6	0.77	25	8	0.76
Combined	Class 1	158	8	0.95	182	37	0.83	187	10	0.95
	Class 2	100	43	0.70	98	36	0.73	173	56	0.76
	Class 3	76	145	0.34	61	46	0.57	90	52	0.63
	Class 4	0	0	N/A	21	7	0.75	34	13	0.72

A major problem with the implementation of the Verhhar scheme is the number of compounds which fall into Class 5 (unclassified); overall 34% of compounds are unclassified. Modifications to the

system enabling these compounds to be placed into the correct class would improve the overall performance of the Verhaar scheme. Inspection of the unclassified compounds identified three rules as potential targets for modification:

- Rule 1.5.2 ("Be aliphatic alcohols but not allylic/propargylic alcohols") – a modification would enable correct classification of an additional two compounds from the *P. promelas* dataset and three compounds from the *T. pyriformis* dataset.
- Rule 1.7.1 ("Are halogenated compounds that comply with rule 1.5 ("Contain C, H & O") but not alpha-, beta- halogen substituted compounds") – a modification would enable correct classification of an additional four compounds from the *P. promelas* dataset.
- Rule 2.1 ("Be non- or weakly acidic phenols") – a modification would enable correct classification of an additional 32 compounds from the *T. pyriformis* dataset and two compounds from the *P. promelas* dataset.

The four compounds which are currently unclassified but should be covered by rule 1.5.2 are shown in the Figure 1a. Only four compounds are shown as cyclohexanol has been tested against both *P. promelas* and *T. pyriformis*. All these compounds are aliphatic alcohols and therefore should be covered by rule 1.5.2 but it appears their ring structures are not currently covered by the rule. Therefore a simple change in the implementation of this rule would lead to these compounds being correctly classified.

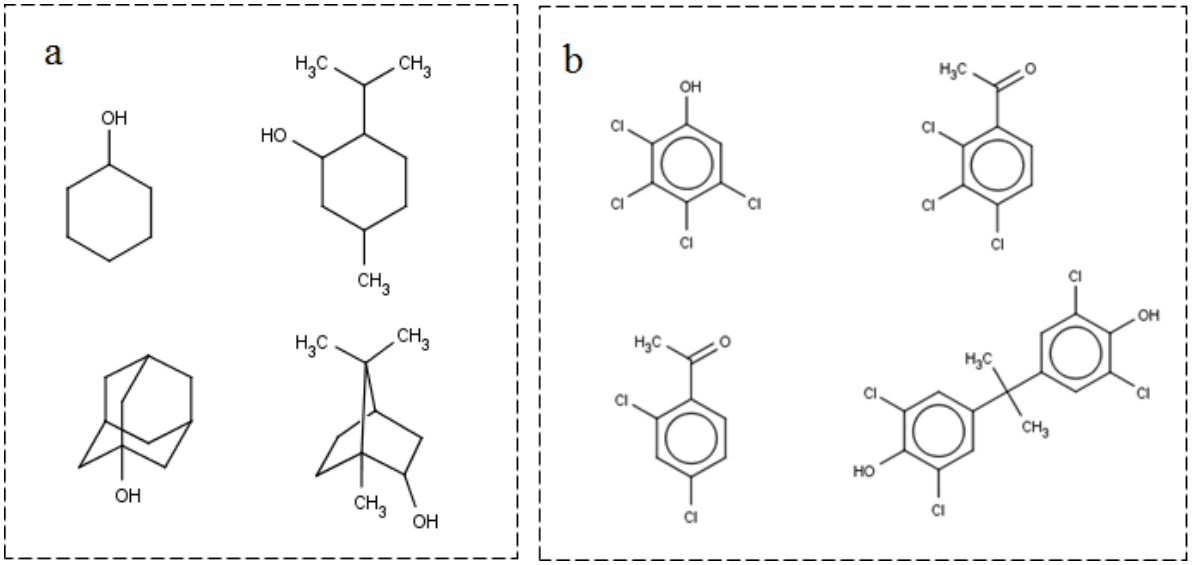

Figure 1. Compounds which (a) do not currently hit rule 1.5.2 but should be covered (cyclohexanol; (1R,2S,5R)-(-)-menthol; 1-adamantanol; and Isoborneol) and (b) do not currently hit rule 1.7.1 but should be covered (2,3,4,5-tetrachlorophenol; 2,3,4-trichloroacetophenone; 2,4-dichloroacetophenone; and 4,4-isopropylidene-bis-2,6-dichlorophenol)

The group of compounds which should be covered by rule 1.7.1 but were assigned to Class 5 are shown in Figure 1b. These compounds are not alpha-, beta- halogen substituted compounds, but

instead contain an aromatic bond where a double bond is expected. Therefore a simple change in the implementation of this rule would lead to these compounds being correctly classified.

Another rule which appears to be not performing as expected is rule 2.1. There are 34 unclassified polar narcotics in total which would be covered by this rule if the strict interpretation of Verhaar's original rule was interpreted with more flexibility. Verhaar's original rule only included phenols with one nitro substituent, and/or one to three chlorine substituents, and/or alkyl substituents. However, all phenols can be contain a dipole and therefore could act as polar narcotics unless they are electrophilic or have the ability to act via a specific mechanism (e.g. respiratory uncoupler of oxidative phosphorylation). In addition, while examining the unclassified compounds it became apparent that there were a large number of reactive phenols which are not currently covered by the scheme. These include many pro-electrophilic phenols which are precursors to quinones (e.g. 2,3-dimethylhydroquinone). Thus implementation of a series of new rules to identify these phenols will lead to an improvement in the ability of the Verhaar scheme to assign this class of chemicals correctly.

The above mentioned additions and alterations to the scheme were implemented through a KNIME workflow employed after processing the compounds through Toxtree v2.6. A schematic representation of the workflow is presented in Figure 2 (the KNIME workflow is available as supplementary information). The updated predictive performance and statistics are shown in Table 1 and Figure 3. There is a marginal increase in the combined predictive performance when using the post-processing filter which arises from a significant improvement in the predictions made for the *T. pyriformis* dataset but this is countered to some extent by the decline in performance in predicting mechanistic assignments for the *P. promelas* dataset.

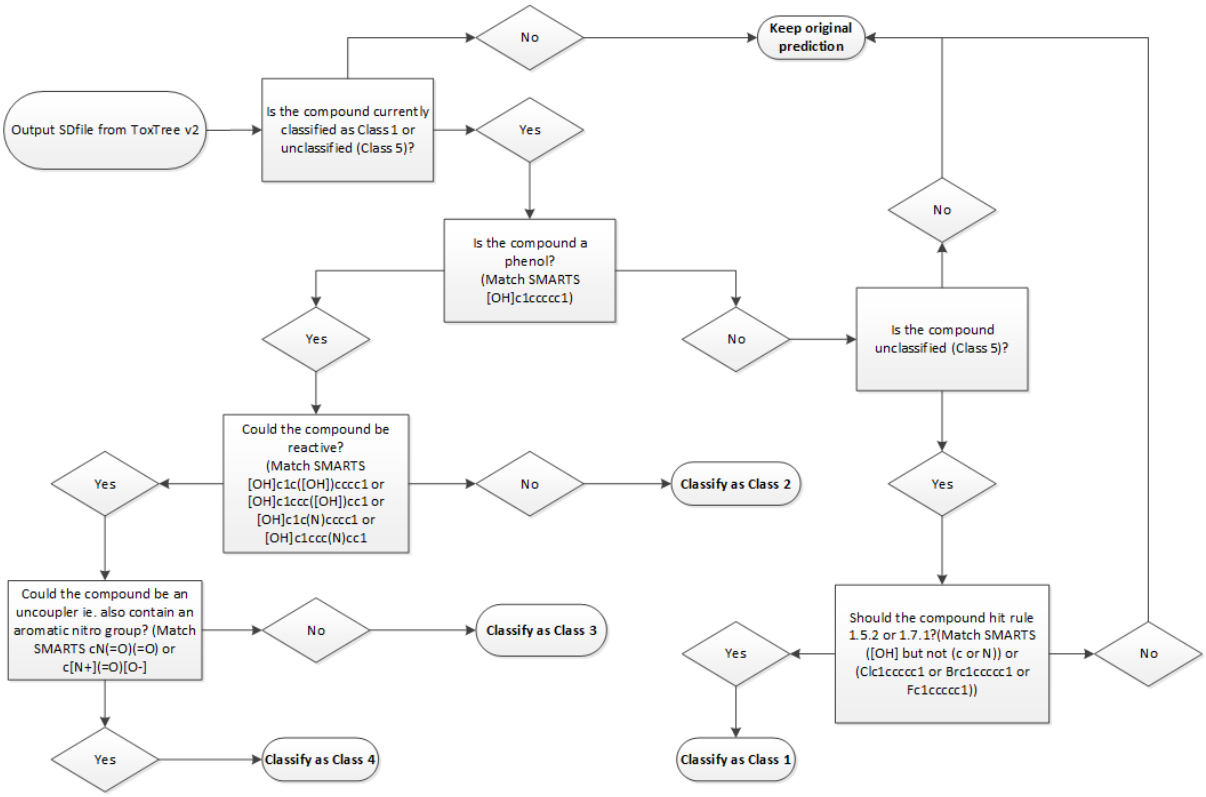

Figure 2. Schematic representation of post-processing filter used to improve classifications

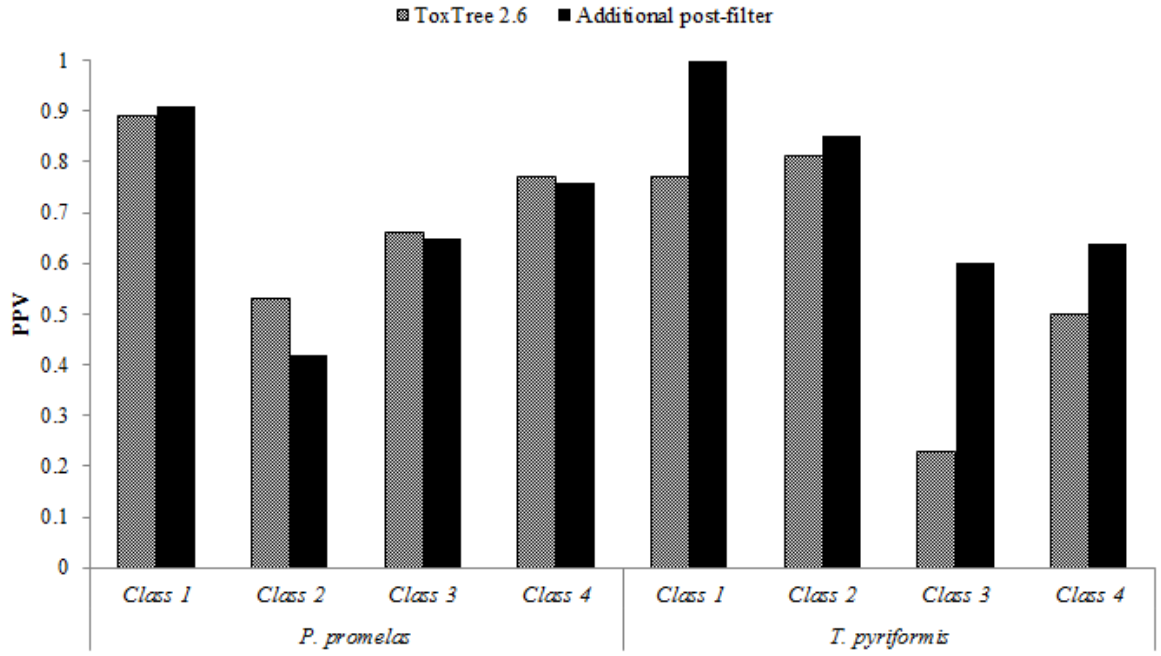

Figure 3. Comparison of positive predictivity values from the Verhaar scheme as implemented in Toxtree v2.6 and those obtained using the additional post-processing filter.

It is important to note that the classification performance of the Verhaar scheme between the datasets cannot be compared directly; the original classifications of the compounds were performed using different methods. The *P. promelas* data were classified using a combination of measured LC_{50} values,

300 animal behaviour during testing and chemical structure (McKim et al., 1987). Conversely, the *T.*
301 *pyriformis* data were classified simply using structure and membership of relevant QSARs. This leads
302 to some discrepancies in the assigned mechanisms between the 69 compounds tested in both species,
303 with only 49 being assigned the same mechanism (Table 2). Thus some of the differences in
304 performance of the Verhaar scheme between the two datasets can be attributed to how the compounds
305 were historically classified into mechanisms.

306 Table 2: Historically assigned mechanisms for compounds tested in both *P. promelas* and *T.*
307 *pyriformis* assays

Compound	Classification	
	P. promelas	*T. pyriformis*
Matching classifications:		
1-butanol	Narcosis I	Non-polar narcosis
1-decanol	Narcosis I	Non-polar narcosis
1-heptanol	Narcosis I	Non-polar narcosis
1-hexanol	Narcosis I	Non-polar narcosis
1-nonanol	Narcosis I	Non-polar narcosis
1-octanol	Narcosis I	Non-polar narcosis
1-pentanol	Narcosis I	Non-polar narcosis
1-propanol	Narcosis I	Non-polar narcosis
2,4-dimethyl-3-pentanol	Narcosis I	Non-polar narcosis
2,4,6-trichlorophenol	Narcosis II	Polar narcosis
2,4-dimethylphenol	Narcosis II	Polar narcosis
2,4-dinitrophenol	Uncoupler	Respiratory uncoupler
2,6-dinitrophenol	Uncoupler	Respiratory uncoupler
2-butanol	Narcosis I	Non-polar narcosis
2-butanone	Narcosis I	Non-polar narcosis
2-chlorophenol	Narcosis II	Polar narcosis
2-decanone	Narcosis I	Non-polar narcosis
2-dodecanone	Narcosis I	Non-polar narcosis
2-ethyl-1-hexanol	Narcosis I	Non-polar narcosis
2-heptanone	Narcosis I	Non-polar narcosis
2-methyl-1-propanol	Narcosis I	Non-polar narcosis

2-methyl-2,4-pentanediol	Narcosis I	Non-polar narcosis
2-methyl-2-propanol	Narcosis I	Non-polar narcosis
2-nonanone	Narcosis I	Non-polar narcosis
2-octanone	Narcosis I	Non-polar narcosis
2-propanol	Narcosis I	Non-polar narcosis
2-tridecanone	Narcosis I	Non-polar narcosis
2-undecanone	Narcosis I	Non-polar narcosis
3,3-dimethyl-2-butanone	Narcosis I	Non-polar narcosis
3-methoxyphenol	Narcosis II	Polar narcosis
3-methyl-2-butanone	Narcosis I	Non-polar narcosis
3-pentanone	Narcosis I	Non-polar narcosis
4,6-dinitro-o-cresol(4,6-dinitro-2-methylphenol)	Uncoupler	Respiratory uncoupler
4-chloro-3-methylphenol	Narcosis II	Polar narcosis
4-chlorocatechol	Reactive	Pro-electrophile
4-chlorophenol	Narcosis II	Polar narcosis
4-methoxyphenol	Narcosis II	Polar narcosis
4-methyl-2-pentanone	Narcosis I	Non-polar narcosis
5-methyl-2-hexanone	Narcosis I	Non-polar narcosis
5-nonanone	Narcosis I	Non-polar narcosis
acetone	Narcosis I	Non-polar narcosis
cyclohexanol	Narcosis I	Non-polar narcosis
cyclohexanone	Narcosis I	Non-polar narcosis
ethanol	Narcosis I	Non-polar narcosis
methanol-rhodamine B	Narcosis I	Non-polar narcosis
o-cresol(2-methylphenol)	Narcosis II	Polar narcosis
pentabromophenol	Uncoupler	Respiratory uncoupler
pentachlorophenol	Uncoupler	Respiratory uncoupler
phenol	Narcosis II	Polar narcosis
Non-matching classifications:		
2,3,4,5-tetrachlorophenol	Narcosis I	Respiratory uncoupler
2,3,6-trimethylphenol	Narcosis I	Polar narcosis
2,4,6-tribromophenol	Narcosis I	Polar narcosis

2,4,6-trimethylphenol	Narcosis I	Polar narcosis
2,5-dinitrophenol	Reactive	Respiratory uncoupler
2,6-di(tert)butyl-4-methylphenol(BTH)	Narcosis I	Polar narcosis
2-hydroxy-4-methoxybenzophenone	Narcosis I	Polar narcosis
salicylamide(2-hydroxybenzamide)	Narcosis I	Polar narcosis
2-nitrophenol	Narcosis II	Soft electrophile
3,5-dibromosalicylaldehyde	Reactive	Polar narcosis
3-ethoxy-4-hydroxybenzaldehyde	Narcosis I	Polar narcosis
4-nitro-3-(trifluoromethyl)-phenol	Narcosis II	Soft electrophile
4-amino-2-nitrophenol	Narcosis II	Soft electrophile
4-nitrophenol	Narcosis II	Soft electrophile
5-bromovanillin	Reactive	Polar narcosis
catechol	Narcosis II	Pro-electrophile
o-vanillin(3-methoxysalicylaldehyde)	Reactive	Polar narcosis
salicylaldehyde(2-hydroxybenzaldehyde)	Reactive	Polar narcosis
tetrachlorocatechol	Uncoupler	Pro-electrophile
vanillin(3-methoxy-4-hydroxybenzaldehyde)	Reactive	Polar narcosis

Irrespective of the differences between datasets, overall the post-processing filters have improved the performance of the Verhaar scheme as implemented in Toxtree v2.6. The post-processing filter has been especially useful in reducing the number of compounds placed into Class 5: reduced from 257 to 130; and thus expanding the applicability domain of the scheme. The slight decrease in the positive predictivity percentages for the *P. promelas* dataset may be offset by the increase in the number of compounds which can now be correctly classified; 204 compounds compared to 193. The 130 compounds which remain out of the domain of the model and thus within Class 5 provide the opportunity to analyse where the Verhaar scheme could be expanded. Indeed Verhaar and co-workers (1992) stated in their original publication that "…this paper is intended as a continuing effort in the development of predictive techniques that can be applied in hazard assessment…" and thus it seems the expansion is well overdue. However, a full study on the expansion of the Verhaar scheme is outside the scope of this paper and therefore these 130 compounds were not analysed further. It is hoped that these compounds will be analysed to elicit the relationships between their structure and mechanism of toxicity, and used in conjunction with other structurally diverse data, in a thorough analysis of where the scheme could be expanded in the future.

The improvement achieved in using the post-processing filters developed here is apparent in the QSAR models built using compounds within these classes. Log P dependent QSAR models for the Class 1 (non-polar narcotics) and Class 2 (polar narcotics) were developed to investigate if the incorrectly classified compounds were outliers. As all the baseline narcotics tested in *T. pyriformis* are classified correctly, the data exactly match the training data used by Ellison and co-workers to produce the following high quality model with no outliers (Ellison et al., 2008):

$$\log IGC_{50}^{-1} = 0.78 \log P - 2.01$$

$$n = 87, r^2 = 0.96$$

Figure 4a demonstrates the relationship between log P and toxicity for those compounds classified as baseline narcotics which have been tested against *P. promelas*. It is apparent that the compounds incorrectly classified are generally outliers compared to the baseline compounds which form the following QSAR:

$$\log LC_{50}^{-1} = 0.89 \log P - 1.87$$

$$n = 100, r^2 = 0.78$$

The two outliers below the line are 1,2-dibromobenzene (log P: 3.77; log LC_{50}^{-1}: -1.13) and amylbenzene (log P: 4.5; log LC_{50}^{-1}: -0.31). These are experimental anomalies which may be attributed to the low water solubility of the compounds. The one significant outlier above the line (log P: 1.18; log LC_{50}^{-1}: 1.73) is 2,3,4-trimethoxyacetophenone, which although correctly classified as a baseline narcotic, shows excess toxicity and may be exhibiting toxicity through another mechanism. The chemical structure reveals that it may be oxidised into the more reactive quinone and could react covalently with proteins via the process presented in Figure 5, or may produce free radicals (Bajot et al., 2011). These outliers have caused the model to be of lower quality to others published in the literature (e.g. Veith et al., 1983; Yuan et al., 2007; Martin et al., 2015) but the model can still be considered useful in demonstrating that the correctly classified compounds are acting via the same, easily modelled, mechanism.

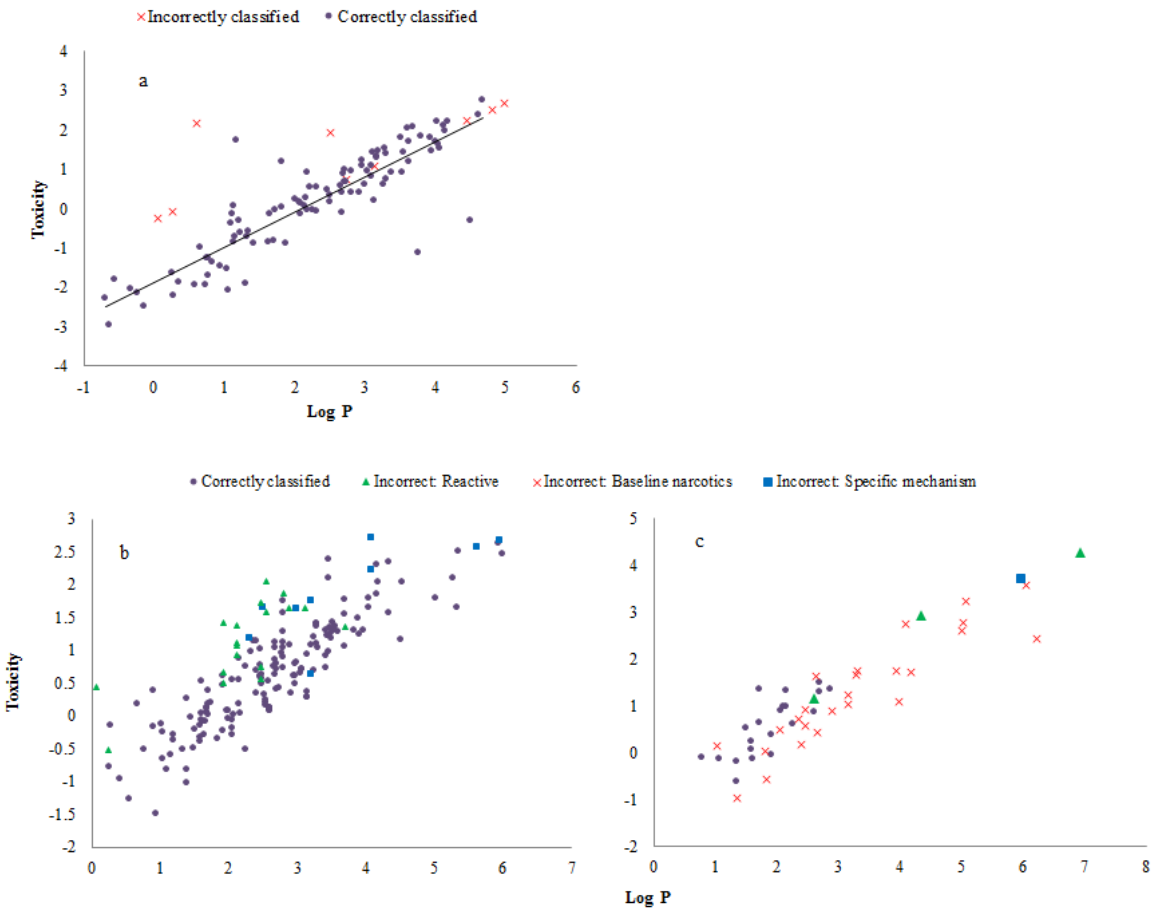

Figure 4. Relationship between (a) toxicity (log LC_{50}^{-1}) and hydrophobicity (log P) for the compounds tested against *P. promelas* and classified into Class 1 (baseline narcotics). A linear relationship for correctly classified compounds is shown; (b) toxicity (log IGC_{50}^{-1}) and hydrophobicity (log P) for the compounds tested against *T. pyriformis* and classified into class 2 (179 compounds) and (c) toxicity (log LC_{50}^{-1}) and hydrophobicity (log P) for the compounds tested against *P. promelas* and classified into class 2 (50 compounds).

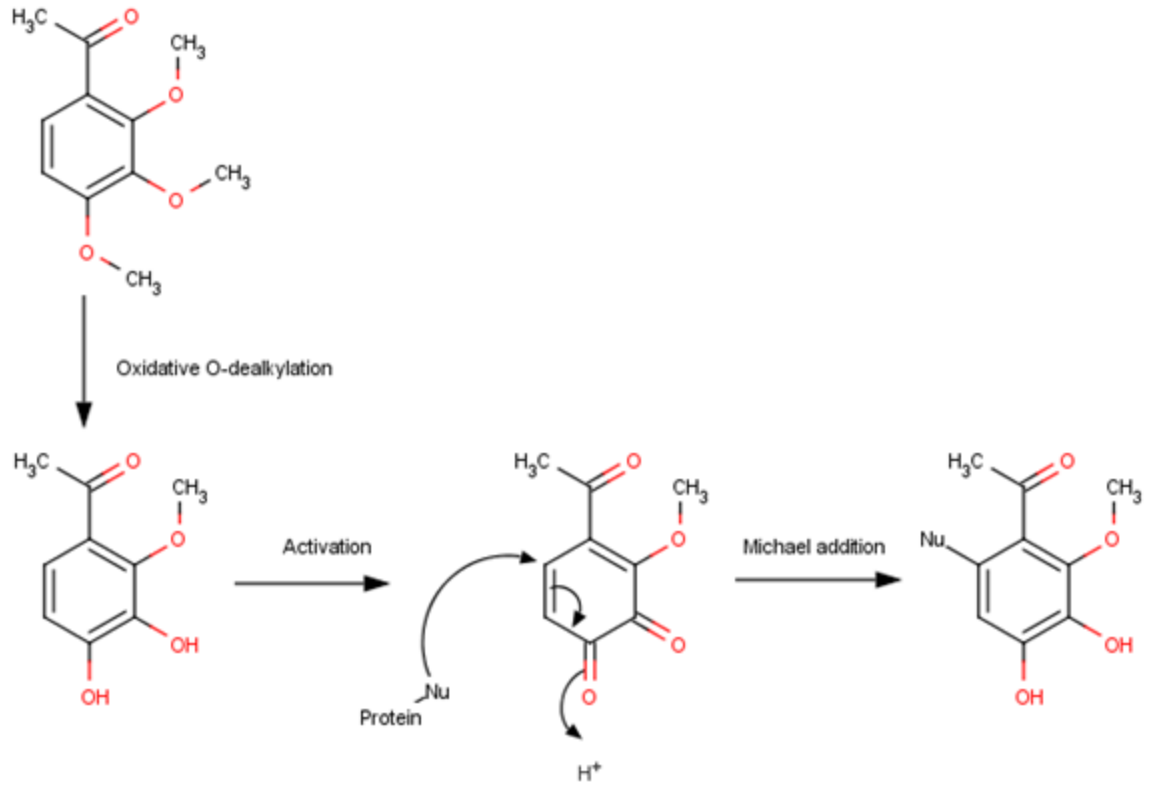

Figure 5. Proposed mechanistic rationale for 2,3,4-trimethoxyacetophenone exhibiting excess toxicity

The compounds tested in *T. pyriformis* and classified as polar narcotics demonstrate a similar pattern with the compounds acting via reactive, or specifically assigned mechanisms of action, being outliers to the general trend (Figure 4b). The same is not true of the compounds tested using *P. promelas* and classified into class 2 (Figure 4c). However, the "true" mechanisms of the misclassified compounds are different in this instance. Unlike the *T. pyriformis* example, the majority of compounds misclassified into class 2 are baseline narcotics. Their toxicity shows a clear trend with hydrophobicity and, as expected, the compounds lie below the correctly classified polar narcotics.

4. Conclusion

The Verhaar scheme is a useful method for assigning compounds into broad categories to assist with hazard identification. The implementation of the scheme in Toxtree means that it can be easily accessed and used by a wide range of scientists in regulatory agencies, industry and academia. This paper demonstrates that the updated implementation of the scheme in Toxtree v2.6 offers increased performance compared to previous versions. However, this research has shown changes to three of the rules in Toxtree v2.6 enabled additional improvements in the scheme to be achieved. The suggested rule improvements will enable scientists to assign compounds to mechanism-based categories suitable for hazard identification with a greater degree of confidence when using the Verhaar scheme.

Acknowledgements

This study is funded by the AlterREACH project (Norwegian Research Council project no. 196318).

References

Regulation (EC) No 1907/2006 of the European Parliament and 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 793/93 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. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF (accessed 25.11.2014).

Austin, T.J., Eadsforth, C.V., 2014. Development of a chronic fish toxicity model for predicting sub-lethal NOEC values for non-polar narcotics. SAR and QSAR in Environmental Research 25, 147-160.

Bajot, F., Cronin, M.T.D., Roberts, D.W., Schultz, T.W., 2011. Reactivity and aquatic toxicity of aromatic compounds transformable to quinone-type Michael acceptors. SAR and QSAR in Environmental Research 22, 51-65.

Cronin, M.T.D., 2006. The role of hydrophobicity in toxicity prediction. Current Computer-Aided Drug Design 2, 405-413.

de Haas, E.M., Eikelboom, T., Bouwman, T., 2011. Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR (TM). SAR and QSAR in Environmental Research 22, 545-559.

Ellison, C.M., Cronin, M.T.D., Madden, J.C., Schultz, T.W., 2008. Definition of the structural domain of the baseline non-polar narcosis model for *Tetrahymena pyriformis*. SAR and QSAR in Environmental Research 19, 751-783.

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

Escher, B.I., Eggen, R.I.L., Schreiber, U., Schreiber, Z., Vye, E., Wisner, B., Schwarzenbach, R.P., 2002. Baseline toxicity (narcosis) of organic chemicals determined by in vitro membrane potential measurements in energy-transducing membranes. Environmental Science & Technology 36, 1971-1979.

405 Gissi, A., Gadaleta, D., Floris, M., Olla, S., Carotti, A., Novellino, E., Benfenati, E., Nicolotti, O.,
406 2014. An Alternative QSAR-Based Approach for Predicting the Bioconcentration Factor for
407 Regulatory Purposes. Altex-Alternatives to Animal Experimentation 31, 23-36.

408 Hermens, J.L.M., 1990. Electrophiles and acute toxicity to fish. Environmental Health Perspectives 87,
409 219-225.

410 Jaworska, J., Gabbert, S., Aldenberg, T., 2010. Towards optimization of chemical testing under
411 REACH: A Bayesian network approach to Integrated Testing Strategies. Regulatory Toxicology and
412 Pharmacology 57, 157-167.

413 Koleva, Y.K., Madden, J.C., Cronin, M.T.D., 2008. Formation of categories from structure-activity
414 relationships to allow read-across for risk assessment: Toxicity of α,β-unsaturated carbonyl
415 compounds. Chemical Research in Toxicology 21, 2300-2312.

416 Könemann, H., 1981. Quantitative structure-activity-relationships in fish toxicity studies. 1.
417 Relationship for 50 industrial pollutants. Toxicology 19, 209-221.

418 Lipnick, R.L., 1991. Outliers - Their origin and use in the classification of molecular mechanisms of
419 toxicity. Science of the Total Environment 109, 131-153.

420 Martin, T.M., Young, D.M., Lilavois, C.R., Barron, M.G., 2015. Comparison of global and mode of
421 action-based models for aquatic toxicity. SAR and QSAR in Environmental Research 26, 245-262.

422 McKim, J.M., Bradbury, S.P., Niemi, G.J., 1987. Fish actute toxicity syndromes and their use in the
423 QSAR approach to hazard assessment. Environmental Health Perspectives 71, 171-186.

424 Nendza, M., Muller, M., Wenzel, A., 2014. Discriminating toxicant classes by mode of action: 4.
425 Baseline and excess toxicity. SAR and QSAR in Environmental Research 25, 393-405.

426 Netzeva, T.I., Schultz, T.W., 2005. QSARs for the aquatic toxicity of aromatic aldehydes from
427 *Tetrahymena* data. Chemosphere 61, 1632-1643.

428 Patlewicz, G., Ball, N., Becker, R.A., Booth, E.D., Cronin, M.T.D., Kroese, D., Steup, D., van
429 Ravenzwaay, B., Hartung, T., 2014. Read-across approaches - misconceptions, promises and
430 challenges ahead. Altex 31, 387-396.

431 Pavan, M., Worth, A.P., 2008. Publicly-accessible QSAR software tools developed by the Joint
432 Research Centre. Sar and Qsar in Environmental Research 19, 785-799.

Pery, A.R.R., Schuurmann, G., Ciffroy, P., Faust, M., Backhaus, T., Aicher, L., Mombelli, E., Tebby, C., Cronin, M.T.D., Tissot, S., Andres, S., Brignon, J.M., Frewer, L., Georgiou, S., Mattas, K., Vergnaud, J.C., Peijnenburg, W., Capri, E., Marchis, A., Wilks, M.F., 2013. Perspectives for integrating human and environmental risk assessment and synergies with socio-economic analysis. Science of the Total Environment 456, 307-316.

Roberts, D.W., Costello, J.F., 2003. Mechanisms of action for general and polar narcosis: A difference in dimension. QSAR and Combinatorial Science 22, 226-233.

Roth, S.H., 1980. Membrane and cellular actions of anesthetic agents. Federation Proceedings 39, 1595-1599.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environmental Toxicology and Chemistry 16, 948-967.

Schaafsma, G., Kroese, E.D., Tielemans, E.L.J.P., van de Sandt, J.J.M., van Leeuwen, C.J., 2009. REACH, non-testing approaches and the urgent need for a change in mind set. Regulatory Toxicology and Pharmacology 53, 70-80.

Scholz, S., Sela, E., Blaha, L., Braunbeck, T., Galay-Burgos, M., Garcia-Franco, M., Guinea, J., Kluver, N., Schirmer, K., Tanneberger, K., Tobor-Kaplon, M., Witters, H., Belanger, S., Benfenati, E., Creton, S., Cronin, M.T.D., Eggen, R.I.L., Embry, M., Ekman, D., Gourmelon, A., Halder, M., Hardy, B., Hartung, T., Hubesch, B., Jungmann, D., Lampi, M.A., Lee, L., Leonard, M., Kuster, E., Lillicrap, A., Luckenbach, T., Murk, A.J., Navas, J.M., Peijnenburg, W., Repetto, G., Salinas, E., Schuurmann, G., Spielmann, H., Tollefsen, K.E., Walter-Rohde, S., Whale, G., Wheeler, J.R., Winter, M.J., 2013. A European perspective on alternatives to animal testing for environmental hazard identification and risk assessment. Regulatory Toxicology and Pharmacology 67, 506-530.

Schultz, T.W., 1997. Tetratox: *Tetrahymena pyriformis* population growth impairment endpoint - A surrogate for fish lethality. Toxicology Methods 7, 289-309.

Schultz, T.W., Cronin, M.T.D., 1997. Quantitative structure-activity relationships for weak acid respiratory uncouplers to *Vibrio fisheri*. Environmental Toxicology and Chemistry 16, 357-360.

Schultz, T.W., Ralston, K.E., Roberts, D.W., Veith, G., Aptula, A.O., 2007. Structure-activity relationships for abiotic thiol reactivity and aquatic toxicity of halo-substituted carbonyl compounds. SAR and QSAR in Environmental Research 18, 21-29.

Schultz, T.W., Sinks, G.D., Cronin, M.T.D., 1997. Identification of mechanisms of toxic action of phenols to Tetrahymena pyriformis from molecular descriptors. in: Chen, F., Schuurmann, G. (Eds.). Quantitative Structure-Activity Relationships in Environmental Sciences - Vii. Setac Press, Pensacola, pp. 329-342.

Schultz, T.W., Tichy, M., 1993. Structure-toxicity relationships for unsaturated alcohols to *Tetrahymena-pyriformis* - C5 and C6 analogs and primary propargylic alcohols. Bulletin of Environmental Contamination and Toxicology 51, 681-688.

Su, L.M., Liu, X., Wang, Y., Li, J.J., Wang, X.H., Sheng, L.X., Zhao, Y.H., 2014. The discrimination of excess toxicity from baseline effect: Effect of bioconcentration. Science of the Total Environment 484, 137-145.

Traas, T.P., van Leeuwen, C.J., 2007. Ecotoxicological effects. in: van Leeuwen, C.J., Vermeire, T.G. (Eds.). Risk Assessment of Chemicals: An introduction. Springer, Dordrecht, The Netherlands, pp. 281-356.

Vaes, W.H.J., Ramos, E.U., Verhaar, H.J.M., Hermens, J.L.M., 1998. Acute toxicity of non-polar versus polar narcosis: Is there a difference? Environmental Toxicology and Chemistry 17, 1380-1384.

Veith, G.D., Call, D.J., Brooke, L.T., 1983. Structure toxicity relationships for the fathead minnow, *Pimephales promelas*: Narcotic industrial-chemicals. Canadian Journal of Fisheries and Aquatic Sciences 40, 743-748.

Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental-pollutants. 1. Structure-activity-relationships for prediction of aquatic toxicity. Chemosphere 25, 471-491.

Walker, C.H., Greig-Smith, P.W., Crossland, N.O., Brown, R., 1991. Ecotoxicology. in: Balls, M., Bridges, J., Southee, J. (Eds.). Animals and Alternatives in Toxicology; Present status and future prospects. Macmillan Academic and Professional Ltd, Hampshire.

Yuan, H., Wang, Y. Cheng, Y., 2007. Local and global quantitative structure-activity relationship modeling and prediction for baseline toxicity. Journal of Chemical Information and Modeling 47, 159-169.