In Silico Classification of Industrial Chemicals Associated with Acute Aquatic Toxic Action Utilising Molecular Initiating Events

Maria Sapounidou

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

This research programme was carried out in collaboration with the Safety and Environmental Assurance Centre, Unilever Research, Colworth House, Sharnbrook, Bedfordshire, UK.

August 2019

Acknowledgements

First and foremost, I would like to wholeheartedly thank my director of studies Professor Mark T. Cronin for his continuous guidance, and the tremendous amount of time spent to help and provide me constructive feedback from the very beginning. I'm extremely grateful to my supervisor Dr Steve Gutsell for his guidance and valuable insight, and the confidence in me throughout the course of my doctoral studies. Similar thanks must go to my other supervisor Dr Judy C. Madden who has always been willing to help when needed from the very beginning. I would not have been where I am now without their constant support.

I am also very grateful to Dr Geoff Hodges for his constant interest and support for my project. His time, precious feedback and support on expanding and communicating my research throughout my doctoral studies have been vastly appreciated.

I am greatly thankful for the financial support provided to me by Unilever Safety and Environmental Assurance Centre, without which this research would not have been possible.

I would also like to thank all my current and former colleagues in the Cheminformatics group in Liverpool John Moores University, for providing an enjoyable and stimulating working environment. Special thanks to Dr Steven Enoch and Dr Andrew Leach for the interesting discussions, and Dr David Ebbrell, Dr Iva Lukac, and fellow PhD students Julia Pletz and Nicoleta Spinu for being the fantastic colleagues and friends they are.

Thanks must also go to my friends and family, for their support and encouragement. I'm very grateful to my wonderful partner James, who has been my rock from day one. The most special thanks should go to my parents Sofia and Theodosis, χωρίς την ασίγαστη κι άνευ όρων υποστήριξη σας, δεν θα βρισκόμουν εδώ που είμαι. Σας ευχαριστώ και σας αγαπώ πάντα.

Abstract

Chemicals legislation requires the assessment of potential harm to humans and environmental species. There is a move away from testing industrial chemicals in animals to development and use of reliable non-testing methods towards Mode of Action (MOA) determination as an important part of understanding toxicity to aquatic organisms. However, current MOA classification approaches for acute aquatic toxicity endpoints are limited by their chemical and mechanistic domains. This thesis developed an *in silico* classification scheme for aquatic toxicology to enable grouping according to relevant mechanism of action. Over 6,200 publicly available toxicity data for nearly 5,000 chemicals and 10 aquatic species were collated and curated to form the basis of the analysis. In addition, mechanistic information was compiled and organised in three broad domains: narcotic, non-specific reactive and specific mechanisms of action. Where possible, the aquatic toxicology domains were organised around the Molecular Initiating Events (MIEs) of the relevant Adverse Outcome Pathway(s). Utilising the MIE allowed direct linkage between structural chemistry responsible for toxicity and the adverse outcome. Other considerations included information on the MIE target structure, i.e. the interaction, associated chemistry, taxonomic applicability and the sources and type of data (e.g. *in silico*, *in vivo*, *in vitro*) corresponding to MIE/MIE target of interest. Structural alerts were developed for each mechanism and were documented and evaluated using a MIE-centred set of criteria. Approximately 65 MIE and MIE targets relevant to the aquatic flora and fauna were identified. This knowledge extended greatly the classification scheme for non-specific reactive and specific mechanisms of toxicity beyond the currently applied schemes. The chemical structural criteria for class assignment along with transparency in data source and quality were coded in a workflow. This provides the user with an informed MIE prediction that can be used in the application of the AOPs to better understand chemical classification, predict toxicity and support interspecies risk assessment.

Table of Contents

ABSTRACT 2

LIST OF ABBREVIATIONS 5

CHAPTER 1: INTRODUCTION 8

1.1. CHEMICAL SAFETY AND AQUATIC ENVIRONMENT 8
1.2. *IN SILICO* APPLICATIONS FOR RISK ASSESSMENT 11
1.2.1. In silico applications for acute aquatic toxicity prediction 13
1.3. FROM CHEMISTRY-BASED TO MECHANISM-BASED DECISION MAKING 19
1.4. AIM OF THE THESIS 22
1.5. CONCLUSIONS OF CHAPTER 1 23

CHAPTER 2: EVALUATION OF ACUTE TOXICITY DATA FROM THE EXISTING SCIENTIFIC LITERATURE 24

2.1. INTRODUCTION 24
2.2. AIM OF CHAPTER 2 26
2.3. METHODS 27
2.3.1. Data collection and dataset construction 27
2.3.2. Database construction and analysis 31
2.4. RESULTS AND DISCUSSION 31
2.4.1. Calculation of molecular descriptors 36
2.4.2. Interspecies relationships 43
2.5. CONCLUSIONS 49

CHAPTER 3: EVALUATION OF CURRENT *IN SILICO* CLASSIFICATION SCHEMES RELEVANT TO ACUTE AQUATIC TOXIC ACTION 51

3.1. INTRODUCTION 51
3.2. AIM OF CHAPTER 3 53
3.3. METHODS 53
3.3.1. Toxicity data and chemical inventories 53
3.3.2. Classification Schemes 55
3.3.3. Comparison of class assignment among schemes 58
3.3.4. Data analysis 59
3.4. RESULTS AND DISCUSSION 60
3.4.1. Inventories of industrial relevance and mode of action assignment distribution 62
3.4.2. Agreement in mode of action assignment among schemes 66
3.5. CONCLUSIONS 70

CHAPTER 4: COMPILATION OF MOLECULAR INITIATING EVENTS RELATING TO AQUATIC TOXICITY FROM THE LITERATURE 73

4.1. INTRODUCTION 73
4.1.1. Adverse Outcome Pathway 73
4.2. AIM, SPECIFIC OBJECTIVES AND SCOPE OF CHAPTER 4 77
4.3. METHODOLOGY 78
4.3.1. Meek criteria for the evaluation of MIEs 78
4.3.2. Biological information captured from the MIE literature 81
4.3.2.1. Capturing mechanistic information at the molecular initiating event level 81
4.3.2.2. Capturing taxonomic domain of applicability 83
4.3.3. Chemistry of the MIE literature 84
4.3.3.1. MIE derived chemical space 84
4.3.3.2. MIE derived- versus industry relevant- chemical space 85
4.3.4. Clustering of MIEs in mechanistic domains 85

4.4. Results 86
4.4.1. Data Quality 86
4.4.2. Type of MIEs 87
4.4.3. Taxonomic applicability 92
4.4.4. Chemical domain of applicability 93
4.4.5. Mechanistic domains 97
4.5. Discussion 97
4.5.1. Literature search 97
4.5.2. Data quality 98
4.5.2.1. Biological causality 100
4.5.2.2. Taxonomic applicability domain 101
4.5.2.3. Chemical domain of applicability 102
4.5.2 Mechanistic domains 103
4.6. Conclusions 103

CHAPTER 5: DEVELOPMENT OF STRUCTURAL ALERTS FOR MECHANISMS OF AQUATIC TOXIC ACTION BASED ON MECHANISTIC DATA **107**

5.1. Introduction 107
5.2. Aim of Chapter 5 109
5.3. Methodology 110
5.3.1. Data set 110
5.3.2. Development of in silico profilers for mechanistic domain assignment 113
5.3.2.1. Specific Toxicity 114
5.3.2.2. Non-Specific Reactivity 117
5.3.2.3. Narcosis 119
5.3.2.4. Performance of in silico profilers 120
5.4. Results and Discussion 123
5.4.1. Specific Toxicity 123
5.4.2. Non-Specific Reactivity 135
5.4.2.1. Definition of the non-specific reactivity domain 136
5.4.2.2. Links between the non-specific reactivity and specific toxicity domains 142
5.4.3. Narcosis 143
5.4.3.1. Definition of the narcosis domain 144
5.4.3.2. Suggested implementations for narcosis domain 149
5.4.4. Application of in silico profilers for MIE assignment 149
5.5. Conclusions 151

CHAPTER 6: DISCUSSION AND FUTURE WORK 154

6.1. Summary of work 154
6.1.1. Current literature and in silico resources to support hazard identification 155
6.1.2. Using information from Adverse Outcome Pathways to support hazard identification 159
6.2. Future work 162
6.2.1. Extending the chemical coverage of the in silico classification scheme 162
6.2.2. Developing the role of multiple MIEs and biotransformation in the classification scheme 163
6.2.3. Expanding mechanistic knowledge including chronic toxicity 163
6.2.4. Increasing taxonomic applicability 164
6.2.5. Creation of a "smart" computational framework to support safety assessment 165
6.3. Concluding remarks 166

CHAPTER 7: REFERENCES 167

List of Abbreviations

AAT OASIS	Acute Aquatic Toxicity MOA by OASIS
AChR	Acetylcholinesterase receptor
AHAS	Acetohydroxy-acid synthase
AhR	Aryl hydrocarbon receptor
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
AOP-KB	AOP Knowledgebase
AQUIRE	AQUatic toxicity Information Retrieval
ASTER	ASsessment Tools for the Evaluation of Risk
ATP	Adenosine triphosphate
CAS	Chemical Abstract Service
CAS RN	CAS registry number
CDK	Chemistry Development Kit
CNS	Central Nervous System
CSA	Chemical Safety Assessment
CYP	Cytochrome P450 enzymes
Da	Dalton
DB	Database
DHP synthase	Dihydropteroate synthase
DNA	Deoxyribonucleic acid
EC50	Median effective concentration
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
ECOSAR	Ecological Structure Activity Relationships
ECOTOX	Ecotoxicology Database
EPA	Environmental Protection Agency
EPISuite	Estimation Program Interface Suite
EPSP	Enolpyruvylshikimate 3-phosphate
ERA	Environmental Risk Assessment
ERL-D	Environmental Research Laboratory Duluth
EU	European Union
FHM	Fathead minnow

GABA	gamma-Aminobutyric acid
GLP	Good Laboratory Practice
IGC50	50% inhibitory growth concentration
InChI	IUPAC International Chemical Identifier
JRC	Joint Research Centre
KNIME	Konstanz Information Miner
LC50	Median lethal concentration
log K_{ow}	logarithm of the octanol-water partition coefficient
MIE	Molecular Initiating Event
MIE-SA	Molecular Initiating Event - Structural Alert
MOA	Mode of Action
nAChR	nicotinic Acetylcholinesterase Receptor
NADH	Nicotinamide adenine dinucleotide
NADPH	reduced Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Center for Biotechnology Information
NTM	Non-Testing Methods
OASIS	Optimised Approach based on Structural Indices Set
OECD	Organisation for Economic Cooperation and Development
PCB	Polychlorinated biphenyl
pEC50	negative logarithm of the EC50
PPDB	Pesticide Properties DataBase
PPP	Positive Predictability Percentages
PSI/PSII	Photosystem I / Photosystem II
QAAR	Quantitative Activity-Activity Relationships
QSAAR	Quantitative Structure-Activity-Activity Relationships
(Q)SAR / QSAR	(Quantitative) Structure-Activity Relationships
REACH	Registration, Evaluation, Authorisation and restriction of Chemicals
RNA	ribonucleic acid
ROS	Reactive Oxygen Species
SA	Structural Alert
SEAC	Safety and Environmental Assurance Centre
SeqAPASS	Sequence Alignment to Predict Across Species Susceptibility
SMARTS	SMiles ARbitrary Target Specification
SMILES	Simplified Molecular-Input Line-Entry System

TCDD	2,3,7,8-Tetrachlorodibenzodioxin
VLCFA	Very Long Chain Fatty Acid
WEKA	Waikato Environment for Knowledge Analysis
WoE	Weight of Evidence

Chapter 1: Introduction

1.1. Chemical safety and aquatic environment

It is estimated that there are approximately 30 million chemical products on the market globally, comprising at least 100,000 unique chemicals, many of which eventually be released into the environment (Nicolotti *et al*., 2014). The main source of detrimental chemical burden in the environment is of anthropogenic origin followed by inert materials of geophysical nature, fossil fuels and organism by-products (van Leeuwen, 2007a). The principal recipients of this pollutant burden are aquatic ecosystems, home to roughly 300,000 aquatic species (Grosberg *et al*., 2012). The effect of aquatic pollution extends to terrestrial species as those two ecosystems are tightly interlinked through a constant interchange of energy, nutrients and material as well to the air (Soininen *et al*., 2015). However, in a world running on chemicals to ensure wellbeing and quality of life, as well as the ability to feed a growing world population, a full restriction on chemical exposure is an unrealistic goal. Thus, the effect of chemicals on all species must be understood and every effort made to ensure that such exposure is not deleterious.

In order to understand the potential deleterious nature of chemicals, various types of information are required. Traditionally, risk assessment would entail hazard identification and characterisation; assessment of the dose required to cause toxicity based on *in vivo* data; exposure assessment and characterisation of the likely risk to the exposed population; and the final outcome being a prediction of the possible risk to humans or the environment (van Leeuwen, 2007b). Focusing on environmental risk assessment (ERA), the measurement of the potential hazardous effects of chemicals on aquatic species through *in vivo* toxicity tests is the fundamental, as well as the most time and cost effective, component. The focus of this thesis is comprehension of chemical hazard, specifically through the investigation of the acute toxicity of individual chemicals. Whilst this toxicological information is a requirement under many articles of legislation, in reality there is a great paucity of toxicological data for the vast majority of chemicals

released into the environment. For example, in 2002, it was reported that full records of toxicity data for the 2,465 industrial chemicals then registered within European Union were limited to as few as 3%, with an additional 43% of compounds with partial toxicity data (Dearden, 2002).

Aquatic toxicity testing, conducted under Good Laboratory Practice (GLP) protocols has been the predominant source of toxicity data for ERA. A number of factors have to be taken into account in regards to physicochemical properties, dose, species, endpoint and duration (OECD, 2016a, OECD, 2016b, OECD, 2016c), which are summarised as follows:

a) Physicochemical properties for the compound of interest would be either calculated or measured experimentally. For waterborne exposure, water solubility and volatility in solution would be taken into account for the experimental design of further testing (Webb and Morlacci, 2010).
b) Prerequisites for exposure groups are three dose levels (low, medium, high) with additional control groups. Studies are performed either as single dose for short term studies (e.g. acute toxicity) or repeated for long term and continuous exposure (e.g. chronic studies to sublethal concentrations).
c) For aquatic toxicity, the most common test species include fish (e.g. *Oncorhynchus mykiss* (rainbow trout), *Oryzias latipes* (medaka), *Gasterosteus aculeatus* (threespine stickleback)) followed by studies on crustaceans (e.g. *Daphnia magna* (water flea)), algae and cyanobacteria (e.g. *Pseudokirchneriella subcapitata*).
d) The endpoint is defined from the biological responses of the test organism to one or more test concentrations of a chemical over a defined period. Examples of the most common toxicological endpoints are the median lethal concentration (LC_{50}) and the median effective concentration (EC_{50}), for effects such as immobilisation (i.e. for crustaceans) or growth inhibition (i.e. for aquatic plants) effects.

e) Based on their duration, aquatic toxicity studies are divided into those being acute or chronic. Acute toxicity tests span from 1-3 days depending on species, whereas chronic toxicity tests can extend over several weeks or months (e.g. a fish full life-cycle test).

In view of the increasing demands for toxicity assessment, the use of alternative methods to animal testing to be considered for filling the knowledge gaps has been proposed by many organisations including, for instance, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) (ECETOC, 2005). In 2007, the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) legislation entered into force in the European Union and instigated the creation of a new Agency, the European Chemical Agency (ECHA), to administer it. REACH, one of the most complex pieces of legislation in the European Union (EU), aims at transparent and independent decision-making to ensure the protection of human health and environment (European Commission, 2006). The main themes of the chemical safety assessment (CSA) under REACH are preliminary assessment based on available data; hazard identification and assessment; exposure estimation; and risk characterisation for both environment and human health. Bearing in mind the recorded data gaps and the immediate need for toxicity testing to meet these demands (ECHA, 2016), REACH introduced the use of alternative methods to *in vivo* fish testing to assess the ecotoxicity of chemicals (ECHA, 2011b, ECHA, 2014, ECHA, 2017c). While the use of toxicity tests in aquatic invertebrates is implied in cases when no other option is available, the use of *in vitro* and *in silico* (computer-based) techniques to prevent unnecessary testing with fish is encouraged for toxicity assessment.

Chemical safety assessment (CSA) using non-testing methods (NTM) has come to the fore for industrial chemicals with a significant effect on research directed in this area (detailed analysis of the driving forces for use and development of NTM are discussed by Cronin and Yoon (2019)). Guidance documents include information on approved NTMs, such as *in vitro* studies for endpoints of interest (e.g. Table R.7.10 – 7 in Chapter R7c for *in vitro* bioaccumulation studies

(ECHA, 2017a)); and Quantitative Structure-Activity Relationship (QSAR) models (e.g. estimation of median effective concentration for acute aquatic toxicity as found in Chapter R.6 (ECHA, 2008)). There is information available on how to use alternatives to animal testing, including detailed descriptions for methodologies such as the tiered approach (Appendix R.7.12-4 in Chapter 7c (ECHA, 2017a)) on how to integrate *in vivo* and *in vitro* data for risk assessment; read-across (ECHA, 2017b) to derive predictions based on relevant information from analogous substances; and weight of evidence (description in (ECHA, 2011a)) on how to integrate estimated and experimental data.

The predictive strength of any NTM method relies heavily on the reliability of the data on which it is based. Depending on the endpoint of investigation , the type of data can be related to the identity and structure of the chemical, e.g. CAS number, chemical name; data relating to the physico-chemical properties of the chemical, e.g. octanol-water partition coefficient, molecular weight; and data relating to the activity of the chemical in biological systems and/or assays. Criteria to support the systematic reporting of the confidence of ecotoxicology studies have been proposed by Klimisch *et al.* (1997) and Moermond *et al.* (2016). Therefore, a common denominator for the assessment of any *in silico* application is data quality. Chapter 2 in this thesis focuses on retrieving high quality acute aquatic toxicity data relevant to aquatic flora and fauna that are publicly available in literature. Data quality in acute toxicity data is extensively discussed in Chapters 2 and 4.

1.2. *In silico* applications for risk assessment

In silico toxicology is a term used for the application of alternative, non-invasive methods, carried out solely in a computer-based interface to obtain information on the hazard of a substance. Combining the extensive improvements in computational power with a time and cost-effective operation, *in silico* tools form a promising and reliable alternative for hazard assessment. The most prominent *in silico* approaches for the generation of NTM data are (Quantitative) Structure-

Activity Relationships ((Q)SARs), grouping, category formation and read-across methods (ECHA, 2008).

(Q)SARs are theoretical models, designed to predict the physico-chemical properties, human health and/or environmental effects of a substance, based on characteristics of its chemical structure. The most common techniques to develop a (Q)SAR are regression analysis and classification methods (e.g. linear discriminant analysis, classification trees etc.). They have the form of a mathematical model, consisting of three main components: a) the quantitative/qualitative parameter(s) derived from chemical structure, which are linked to b) a quantitative measure of a property or activity (e.g. toxicological endpoint) through c) an algorithm (ECHA, 2008). A crucial component of a reliable (Q)SAR model and its successful application is the accurate definition of its applicability domain; defined as 'the response and chemical structure space in which the model makes predictions with a given reliability' (Netzeva *et al*., 2005).

Applications of category formation and read-across have played a vital role in filling data gaps and decision-making (Enoch, 2010, Spielmann *et al*., 2011). ECHA defined a category as "substances that are structurally similar with physicochemical, toxicological, eco-toxicological and/or environmental fate properties that are likely to be similar or to follow a regular pattern" (ECHA, 2008). Similarly, the Organisation for Economic Co-operation and Development (OECD) defined a category as *"Chemicals whose physical-chemical, toxicological and ecotoxicological properties that are likely to be similar or follow a regular pattern as a result of structural similarity may be considered as a group or category"* (OECD, 2014). Chemical similarity can be defined as chemical shape, the presence of various functional groups or the mechanism through which the chemical initiates an adverse outcome (the molecular initiating event). To facilitate category formation and read-across, OECD launched the OECD QSAR Toolbox (OECD, 2013) and this has been highly successful in promoting this area of science. This software tool contains a number of mechanistic profilers enabling chemical categories to be formed for a range of toxicological endpoints (e.g.

Acute Aquatic Toxicity; DNA binding). Once a category has been developed, predictions can then be made via read-across or trend analysis. Using read-across, data gaps for untested substances can be filled using proxy data from another substance or category of chemicals (ECHA, 2017b). It is important to highlight that an effective read-across application is based on applying selected criteria on high quality data to establish similarity among substances (Cronin and Madden, 2010, Patlewicz *et al*., 2014). This concept provides the main focusses of this thesis where methods to identify similar chemicals are developed and supported by the use of high-quality data (see Chapter 5).

1.2.1. *In silico* applications for acute aquatic toxicity prediction

Common practice in defining a category is the use of structural alerts as a means of determining and demonstrating chemical similarity (e.g. use of a profiler such as that for mitochondrial toxicity (Nelms *et al*., 2015)). A structural alert is defined as a chemical fragment or substructure associated with a particular measured property, effect or biological response (mechanistic or otherwise) (Sushko *et al*., 2012). A collection of structural alerts that define a category or allow for analogue identification can form what is known as an "*in silico* profiler". Such profilers can be used to confidently assign *de novo* compounds to a category or for analogue identification. The OECD QSAR Toolbox contains many such profilers based on effect and mechanistic similarity for a variety of environmental and human health endpoints. The most commonly used and publicly available *in silico* profilers relevant to acute aquatic toxicity are presented below.

Verhaar and co-workers, using literature that captured mechanistic information on narcotic effects (Bradbury *et al*., 1990, Veith and Broderius, 1990, McKim *et al*., 1987), electrophiles (Hermens, 1990), specifically acting compounds (McKim *et al*., 1987), and the hypothesis of a linear correlation between acute toxicity and the logarithm of the octanol-water partition coefficient (log K_{OW}) dating back as far as 1899 (Meyer, 1899, Overton, 1901), developed a set of rules to assign chemicals to mechanistic classes. Each class was associated with a corresponding

QSAR(s) for effect concentration estimation. Thus, the Verhaar scheme classified compounds based on their structure into one of four mechanistic classes:

a) Inert chemicals (Class 1) that demonstrate non-specific baseline acute effects, known as non-polar narcosis. As demonstrated by Könemann (1981) and extended by Verhaar (Verhaar *et al*., 1992), there is a linear correlation between log K_{OW} and fish toxicity for this class.
b) Less inert chemicals (Class 2) that consist of chemicals with higher potency than non-polar narcotics, but still with a non-specific mode of action.
c) Reactive chemicals (Class 3) referring to compounds that act in a non-specific electrophilic manner, targeting biomolecules leading to direct or indirect (bioactivation) toxic effects.
d) Specifically acting chemicals (Class 4) that comprise a diverse chemical space with compounds exhibiting specific interactions with target biomolecules leading to a defined toxic action, such as acetylcholinesterase inhibition and uncoupling of respiratory oxidative phosphorylation.

In an additional class (Class 5) are chemicals that cannot be placed in any of the above four classes. Since its conception, the Verhaar scheme has been challenged, modified and further validated as reported in a number of key studies (Verhaar *et al*., 2000, Enoch *et al*., 2008, Ellison *et al*., 2015). Verhaar *et al.* (2000) externally validated the scheme, primarily using fathead minnow (*Pimephales promelas)* data. The set of rules as described in the original publication along with the suggested modifications following further evaluation in two further studies have been coded into a decision tree in the open source application ToxTree (JRC Computational Toxicology, 2018). The software assigns each chemical, on the basis of its SMILES string, to one of the five classes described above.

In 1991, the United States Environmental Protection Agency (US EPA) published an overview of an in-house expert system, ASTER (ASsessment Tools for the Evaluation of Risk) used for

environmental risk assessment purposes (Russom *et al*., 1991). By design, ASTER combined databases and QSARs to facilitate hazard identification, environmental exposure assessment and ecological risk characterisation. Hazard identification predictions covered both environmental and human health effects, retrieving toxicological information from a range of databases e.g. the Environmental Research Laboratory Duluth (ERL-D) fathead minnow database and AQUatic toxicity Information Retrieval (AQUIRE) database (Nesnow *et al*., 1987). As published in 1997 (Russom *et al*.), ASTER classifies compounds into seven groups relevant to their acute toxic MOA:

a) Narcosis I, comprising aldehydes, anilines (acetophenone and benzamide substructures on aniline ring), phenols (acetophenone and benzamide substructures on phenol ring) and pyridines (with one or more benzene/pyridine rings, or aliphatic carbons connected to the pyridine ring ortho to the nitrogen). Behavioural effects in fish exposed to narcosis I compounds are predominantly depressed locomotory activity with little or no response to outside stimuli (Type I behaviour).
b) Narcosis II, comprising compounds with a benzene ring, anilines (with one para nitro group), phenols (with one amine group or pyridinols) and pyridines. Behavioural effects in fish exposed to narcosis II compounds are hyper- and over- activity to outside stimuli (Type II behaviour).
c) Narcosis III, comprising esters. Behavioural effects in fish exposed to narcosis III esters manifested Type I behaviour.
d) Oxidative phosphorylation uncouplers, comprising anilines (containing more than three halogens or nitro groups), phenols (contain more than three halogens, nitro groups, or azo linkage between phenol and aromatic ring) and pyridines. Behavioural effects in fish when exposed to oxidative phosphorylation uncouplers are Type II behaviour.
e) Reactive electrophiles/proelectrophiles comprising a spectrum of compounds ranging from acetamidophenols, quinolones, epoxides, aziridines, sulphuric, and phosphoric acid esters to diazo compounds, mustards, thiocyanates and diketones. The predominant

acute effect of compounds classified as reactive electrophiles/proelectrophiles is gill irritation along with signs of spontaneous locomotory activity with a high incidence of convulsion, spasms, tetany, scoliosis, lordosis and/or haemorrhaging in the vertebral column (Type III behaviour).

f) Acetylcholinesterase inhibitors, comprising carbamates and organophosphates. Fish exposed to acetylcholinesterase inhibitors exhibit signs of Type III behaviour.

g) Central nervous system (CNS) seizure agents, comprising organochlorines and pyrethroids. Behavioural effects of fish exposed to CNS seizure agents are related to Type III behaviour.

The Russom scheme was part of a US EPA in house expert system. However, an open source application for the prediction of acute aquatic toxicity mode of action (AAT OASIS), heavily based on the Russom scheme, was launched by the Laboratory of Mathematical Chemistry in their OASIS software (oasis-lmc.org) as part of the OECD QSAR Toolbox in April 2010. The OECD QSAR Toolbox is a computational tool for grouping chemicals and reading across missing toxicity values from existing data, with the latest version released in February 2018 (OECD). The Acute Aquatic Toxicity MOA by OASIS (AAT OASIS) profilers assign compounds to different categories according to their acute toxic mode of action (MOA). The seven categories hierarchically ordered MOAs, which are derived from the Russom scheme are:

a) Basesurface narcotics, corresponding to the chemistry and MOA mainly associated with Narcosis I (non-polar narcosis) compounds by the Russom scheme. Chemicals that are not classified as narcotic amines, esters, phenols and anilines, aldehydes, α, β-unsaturated alcohols, and reactive unspecified are included in this category.

b) Narcotic amines, corresponding to the chemistry and MOA mainly associated with Narcosis II (polar narcosis) compounds by the Russom scheme.

c) Esters, corresponding to the chemistry and MOA mainly associated with Narcosis III

compounds by the Russom scheme.

d) Phenols and anilines, a MOA that includes weak and respiratory uncouplers, soft electrophiles and phenols and anilines that elicit Type II behavioural effects.

e) Aldehydes, a chemical class with toxic effects associated with electrophiles that act via covalent bond rearrangement forming Schiff bases with amino groups in biological macromolecules.

f) α, β-Unsaturated alcohols, a chemical class that undergoes metabolic oxidation leading to Michael-type addition reactions and manifests itself as a predominantly reactive mode of toxic action.

g) Reactive unspecified, the most chemically diverse MOA ranging from pyridines with a polar narcosis MOA, electrophiles/protoelectrophiles, halogenated compounds with specific MOA to chemical groups such as amidines, triazines and sulfur-containing compounds with unknown MOA.

Acetylcholinesterase inhibitors and CNS seizure agents are classified under the 'Reactive unspecified' label by AAT, differentiating it both from Verhaar and Russom that classify these compounds in separate classes (i.e. Class 4 in Verhaar and acetylcholinesterase inhibitors, CNS seizure agents in Russom).

The US EPA also provides freely accessible software for the prediction of aquatic toxicity, namely the ECOSAR programme (EPA Ecological Structure Activity Relationships, (US Environmental Protection Agency, 2017)). It is commonly used to predict aquatic (acute and chronic) toxicity to fish, daphnids and green algae. The latest version, v.2.0, was released in October 2017 (https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model). The software assigns compounds based on their chemical class to Structure-Activity Relationships (SARs). The SARs are developed on the principle of the direct correlation between aquatic toxicity and log K_{OW} and how this correlation is reflected for specific chemical

classes and captured for the three taxonomic groups. The basis of QSAR and class assignment is almost exclusively structure-based, as ECOSAR, by design, provides information on the chemical class and proposed aquatic toxicity levels based on linear regression models, rather than mechanistic information (Reuschenbach *et al*., 2008).

ECOSAR, Acute Aquatic Toxicity by OASIS (AAT), Russom and modified Verhaar schemes all base the assignment of chemicals to individual classes on 2D structural information (e.g. SMILES strings). The design of the classes for modified Verhaar, Russom and AAT OASIS is based on a mechanistic framework, harvesting information both from *in silico* and *in vivo* fish studies (key references for Verhaar (Bradbury *et al*., 1990, Lipnick, 1991, Schultz *et al*., 1986, Enoch *et al*., 2008, Ariens, 1986), and Russom/AAT OASIS ((Bradbury *et al*., 1990, Bradbury and Lipnick, 1990, Bradbury and Christensen, 1991, Bradbury *et al*., 1991, McKim *et al*., 1987, Veith *et al*., 1983, Veith and Broderius, 1990)). ECOSAR bases QSAR assignment to predict the acute and chronic toxicity levels of the three distinct taxonomic groups strictly on chemical class assignment with no direct information on mode of action. It is also worth mentioning that despite the strong mechanistic background, chemical species, such as halogenated compounds, expected to be classified as specifically acting compounds, are classified as reactive unspecified/reactive/Class 3. As well as assigning compounds to modes of action, schemes such as Verhaar and Russom are important tools for categorisation and sub-categorisation. Other schemes are also commonly used for sub-categorisation, utilising properties related to bioavailability to support both chemical and mechanistically derived groups. The most well-known is the so-called Lipinski “rule of five” (Lipinski *et al*., 1997) which stated that in the drug discovery setting poor absorption or permeation is more likely when there are more than five hydrogen bond donors, ten hydrogen bond acceptors, the molecular weight is greater than 500 and the calculated log P is greater than 5. The implication is that if a compound has properties that contravene it will be poorly absorbed and therefore have low bioavailability. Rules such as these, and indeed other that are similar, are

fundamentally different to the Verhaar and Russom classification schemes in that they define cut-off criteria based on properties, rather than mechanistic assignment without regard to bioavailability on the basis of structural features.

Chapter 3 focuses on the evaluation of current *in silico* profilers for classification of chemicals according to mechanism of acute aquatic toxic action, chemical coverage relevant to industrial needs and taxonomic applicability.

1.3. From chemistry-based to mechanism-based decision making

In an attempt to integrate mechanistic information for risk assessment and decision-making purposes, Ankley *et al.* (2010) introduced the Adverse Outcome Pathway (AOP) framework. An AOP is defined as a sequence of events from the exposure of an individual to a substance through to an understanding of the adverse effect in individual or population level and is shown schematically in Figure 1.1 (Ankley *et al*., 2010). It is a construct connecting the initiating molecular interaction of a toxicant with a molecular target (molecular initiating event, MIE) with adverse effect(s) at the organism and population level, through single or multiple cellular and organ responses. In practice, an AOP provides a structured approach of plausible connections for risk assessment by organising existing knowledge, engaging both biological and NTM resources.

Although environmental and human health risk assessment share common methodological aspects, their distinction for risk characterisation is important. The novelty of the AOP concept lies in the introduction of a consistent framework for risk assessment for both environmental and human health, empowering risk assessors to integrate and cluster diverse information and identify research needs (Burden et al., 2015). At the same time, many challenges arise, since linearity in an AOP is not always a biological and toxicological reality. It is crucial to define

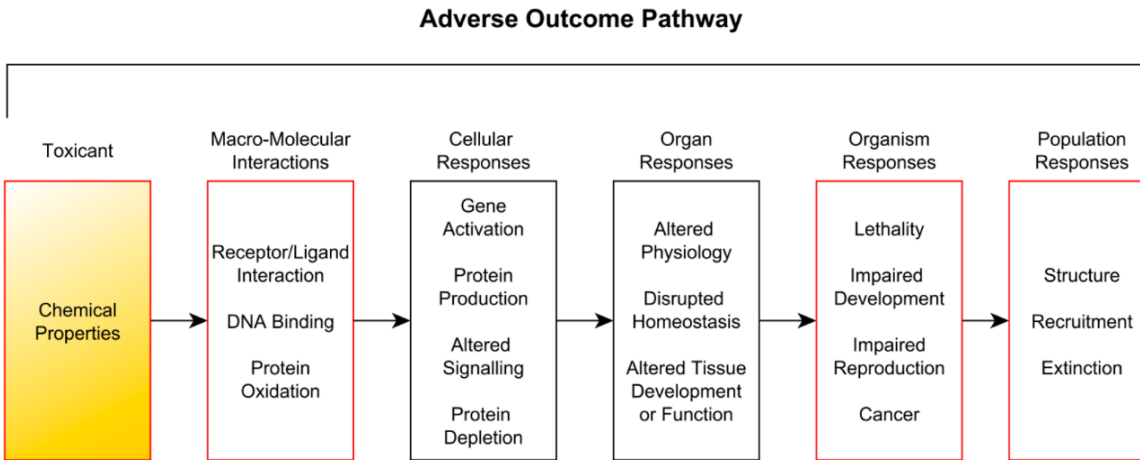


Figure 1.1. Pictorial representation of the adverse outcome pathway framework (modified from Ankley *et al*.(2010))

accurately the MIE linked to the selected adverse outcome inflicted by a substance (Allen *et al*., 2014, Vinken, 2013) so as to derive substantial conclusions. Evidently, a substance might inflict multiple MIEs and adverse effects; therefore, it is not unlikely to construct different AOPs for the same substance, leading to an AOP map (Allen *et al*., 2014). Another advantage of taking the focus away from the AO and to the biochemical interactions at the MIE level is that the additive action of one or multiple substances could potentially be predicted and quantified (Ankley *et al*., 2010, von Stackelberg *et al*., 2015). Accounting for the action of multiple chemicals may require the use of networks to fully explain interaction and allow for true quantification cf. Knapen *et al*. (2018). Whilst networks for AOPs are in their infancy, AOPs are increasingly being seen as a solution to addressing issues such as endocrine disruption (Miller *et al*., 2017)and specific initiating events, such as ecdysone receptor agonism (Song *et al*., 2017). In addition to the effects of organic chemicals to the environment, AOPs have shown use to describe the combination hepatotoxicity effects of three azole fungicides (Heise *et al*., 2018) and for metals (von Stackelberg *et al*., 2015). A further key component of AOPs is the well-defined taxonomic applicability domain, enabling decision-making in population level at the species/taxon level.

There is extensive literature on using weight of evidence (WoE) for developing reliable AOPs. Becker *et al.* (2015) discussed the merits of a tailored weight of evidence approach for data quality within the AOP concept acknowledging in a later publication that, despite caveats for its application within a MOA framework that describes a diverse range of effects, it provides an all-purpose platform to evaluate different aspects of a causal relationship (Becker *et al*., 2017). Bridges *et al.* (2017) highlighted the need for customised WoE criteria, to integrate assay/method-specific considerations depending on the data source. Additionally, Gross *et al.* (2017), using the example of endocrine disruption, discussed the potential of an effect-specific WoE evaluation. Another aspect to be considered for WoE analysis is the life cycle assessment component, highlighted by Gust *et al.* (2016), as there are an emerging number of AOPs for early life effects.

Ågerstrand *et al.* proposed another school of thought for data quality (Moermond *et al*., 2016, Agerstrand and Beronius, 2016) that is based on the Klimisch score system (Klimisch *et al*., 1997). In brief, the Klimisch score assesses the reliability of toxicological studies and is used extensively for regulatory decision making to identify studies performed to GLP conditions (Kase *et al*., 2016, Klimisch *et al*., 1997). Ågerstrand expanded the Klimisch criteria to incorporate data from studies that do not adhere to GLP conditions. Based on this research, there is an online platform (http://scirap.org) that enables the user to evaluate the reliability and relevance of *in vivo* and *in vitro* studies, both for human and environmental health studies (Molander *et al*., 2014). Ågerstrand highlighted the demand for clearer guidance and detailed documentation on how to perform WoE evaluation within a regulatory framework to ensure credibility of the decision-making process (2016).

In Chapter 4, AOP applications, focusing especially on MIEs and data quality for MIEs are extensively discussed.

1.4. Aim of the thesis

The aim of this thesis was to develop a novel classification approach for industrial chemicals to define the domain of mechanisms of acute aquatic toxic action, to be utilised within an Adverse Outcome Pathway (AOP) framework to assess environmental risk. This project was funded by, and performed in collaboration with, Unilever's Safety and Environmental Assurance Centre (SEAC).

The specific objectives of this PhD were:

1. To collect, and evaluate the quality of, acute toxicity data from the existing scientific literature and publicly available resources on industrial organic compounds for relevant aquatic fauna and flora (algae, invertebrate, fish) (Chapter 2).
2. To evaluate the publicly available *in silico* classification schemes for acute aquatic toxic action as proposed by ECHA and OECD for hazard identification (Chapter 3).
3. To investigate the concept of molecular initiating events (MIE) for different mechanisms of aquatic toxic action in publicly available literature by utilising toxicological data from various sources e.g. *in vitro* and high throughput screening data (Chapter 4).
4. To design the chemical space of aquatic toxic action based on mechanistic data derived from MIE knowledge (as discussed in Chapter 5) and develop *in silico* profilers for mechanistic domain and MIE assignment.

1.5. Conclusions of Chapter 1

The increasing number of chemical products on the market has a detrimental environmental impact, stressing the need to further understand their deleterious nature. Traditionally, chemical risk assessment focused on the identification and characterisation of chemical hazard and exposure risk based on *in vivo* toxicity data representative of the aquatic fauna and flora. After the REACH legislation entered into force, the use of non-testing methods was encouraged to support decision-making. Data availability and quality are defining factors for the reliability and predictive strength of non-testing methods, therefore, the Chapters 2 and 4 in this thesis focused on the availability of high-quality data that could support the development of *in silico* applications. Current *in silico* methods to support hazard assessment have been QSAR, grouping, category formation and read-across. Mechanistic approaches assist in hazard identification to fill data gaps. Existing classification methodologies for the assignment of mechanism of action aquatic toxic action (e.g. modified Verhaar, AAT OASIS by QSAR Toolbox) are widely employed. However, they have insufficient coverage by design, since they are based on restrictive chemical domains, and predominantly on fish toxicity studies. Chapter 3 focused on the evaluation of current *in silico* profilers for classification of chemicals using inventories of industrial relevance.

Within the Adverse Outcome Pathway framework, mechanistic information can be utilised to connect the initiating molecular interaction of a toxicant through described events in cellular and organ level with a population response. Chapter 4 explored the availability of high-quality mechanistic data at the MIE level from a wide range of experiments corresponding to various chemistries and taxonomic applicability domains. To ensure data quality, a rigorous data quality process is applied, enforcing MIE-centred criteria. The chemical space derived from the MIE knowledge of Chapter 4 served as a training set to develop an *in silico* profiler for mechanistic domain and MIE assignment with a protocol as described in Chapter 5.

Chapter 2: Evaluation of acute toxicity data from the existing scientific literature

2.1. Introduction

With European Union's implementation of the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation, the demand for acute (and all other relevant) toxicity data amplified. In order to address the needs for Chemical Safety Assessment (CSA) a plethora of online resources with acute toxicity data for multiple endpoints and aquatic species have been made available and many are updated regularly. For instance, governmental agencies around the world maintain databases with toxicity data some of which are summarised in Table 2.1. Recurring problems for these resources are lack of curation and consensus, non-user-friendly interfaces, inconsistencies with structural information and lack of information on the original source and experimental design. Therefore, there is a need and desire to continue to update existing data resources not only in terms of content, but also quality and usability.

Under REACH, the use of alternatives to animal testing methods (e.g. *in vitro*, *in silico*) has been encouraged to facilitate CSA. As thoroughly discussed by Dearden *et al.* (2010), a key prerequisite for good practice in *in silico* modelling is the availability, use and definition of high data quality. To illustrate the problems of, and issues relating to, the assessment of data quality, Przybylak *et al.* (2012) discussed the quality of toxicological data in current resources for hazard assessment. Whilst there are challenges, a series of criteria can be established, for instance it has been suggested a robust assessment of quality should take into consideration, as a minimum study design (e.g. following OECD guidelines), good laboratory practice (GLP) and sufficient data accessibility (e.g. original references) and chemistry (e.g. chemical name, CAS number, purity and source) (Przybylak *et al.*, 2012)).

Table 2.1. Online resources for acute aquatic toxicity data from Governmental Agencies

Organisation	Database	Reference
Environmental Protection Agency, USA (US EPA)	ECOTOX	https://cfpub.epa.gov/ecotox/
	ACToR	https://actor.epa.gov/actor/home.xhtml
Canadian Environmental Agency	CESAR	https://www.canada.ca/en/health-canada/services/chemical-substances.html
European Chemical Agency (ECHA)	ECHA Registered substances	https://echa.europa.eu/information-on-chemicals/registered-substances
	ECHA C&L inventory	https://echa.europa.eu/information-on-chemicals/cl-inventory-database
Organisation for Economic Co-operation and Development (OECD)	eChemPortal	https://www.echemportal.org/
Japanese Authorities	J-CHECK	http://www.safe.nite.go.jp/jcheck/top.action
	JECDB	http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp

Traditionally, *in silico* modelling for acute aquatic toxicity was based on almost exclusively high-quality *in vivo* data from fish studies (exceptions to this included the use of databases for the ciliated protozoan (*Tetrahymena pyriformis)* and bioluminescent bacterium (*Aliivibrio fischeri*). A prime example of such a dataset, that has been used as a 'gold-standard' in QSAR studies, is the Fathead Minnow Acute Toxicity Database (FHMDB) (US Environmental Protection Agency, 2008), comprising 617 acute 96-hour LC_{50} values to *P. promelas* and 225 associated behavioural assessments and 72 joint toxic action experiments -generated by the US EPA Duluth Laboratory in the 1980s. The FHMDB has been used extensively in *in silico* toxicology literature with multiple applications and many QSAR models developed (examples in (Martin *et al*., 2013, Russom *et al*., 1997, Wang *et al*., 2016, Cappelli *et al*., 2015, Veith *et al*., 1983, Verhaar *et al*., 2000)).

With the increasing role of QSAR predictions in decision-making for industrial chemicals, it is crucial to expand the applicability domain of QSARs in both chemistry and at the species level. With the need for increased applicability domains, there is increased concern regarding the quality and reliability of the data. A number of approaches have been proposed to address this concern. Steinmetz *et al.* (2014) suggested a statistical approach to evaluate data quality and the identification of outliers. Using toxicity data derived from the literature for an *in vitro* toxicity assay with *A. fischeri*, Steinmetz *et al.* (2014) introduced a method to assign confidence to 1,944 toxicity values, arguing the importance of repeatability to ensure high quality data. Ruusmann and Maran (2013) proposed a workflow that collected a series of data points from scientific literature and using data for growth inhibition of *T. pyriformis* compiled a database of 2,498 toxicity values for 2,072 compounds. Ruusmann and Maran (2013) discussed in detail different scenarios for data curation on chemical structure (e.g. structure to InChI, CAS number) and toxicity value (e.g. multiple endpoints found in a number of studies). Additionally, there are a number of studies aiming to capture species-specific endpoints (e.g. (Brill *et al.*, 2016, Furuhama *et al.*, 2015a)); and cross-species relationships (e.g. (Wang *et al*., 2016, Basant *et al*., 2016, Furuhama *et al*., 2015b). Within all these resources, there was a need to identify suitable data for this thesis and, where multiple sources of data are available, assess their meaning and the quality and content of the datasets.

2.2. Aim of Chapter 2

The aim of Chapter 2 was to investigate the availability of high quality acute aquatic toxicity data in publicly available resources and the scientific literature for a range of aquatic species relevant to aquatic fauna and flora. The current chapter focused on addressing the following points:

a) Availability of high-quality data for multiple aquatic species;

b) Description of the chemical space captured from available resources;

c) Preliminary analysis of interspecies acute aquatic toxicity relationships.

2.3. Methods

2.3.1. Data collection and dataset construction

The independent data collection process comprised five stages per species or species group:

a) Compilation of references with publicly available acute toxicity data results. Publicly available toxicity data were collected from the following resources US EPA ECOTOX (https://cfpub.epa.gov/ecotox/), eChemPortal (https://www.echemportal.org/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed/), Web of Science (http://apps.webofknowledge.com). For the PubMed and Web of Science databases, toxicological data collection for aquatic fauna and flora was undertaken using the keywords 'qsar', 'dataset' and the respective genera name (i.e. 'fish', 'crustaceans', 'daphnia', 'algae', 'microbiota', 'tetrahymena'). For ECOTOX and eChemPortal, data collection was targeted either per compound of interest or by species, using the advanced search parameters.

b) Evaluation of the references collected and/or tracking original references (i.e. when no information on experimental conditions was provided from the initial reference) for compliance with OECD Test Guidelines was performed. Data retrieval focused on studies from the literature that were performed using species with appropriate OECD Test Guidelines, in GLP conditions, with well described parameters ('Reliable without Restriction' (Klimisch *et al*., 1997)). The protocols that were specifically considered are listed in Table 2.2. Measured effect concentrations are preferred for analysis, due to data availability limitations, both nominal and measured median effect concentrations obtained from the literature were compiled and recorded here in units of mmol/L. Using ACD/Percepta Platform (Development(ACD)/labs, 2017), water solubility values (log S) were calculated, and compared with the available toxicity values; the toxicity values above log S were discarded. For evaluation of toxicity values for surfactant (anionic and cationic) compounds, the main limitation was mixture toxicity due to the range of chain lengths of tested surfactants which

was taken into consideration when surfactant data were further analysed.

Table 2.2. Toxicity assay and endpoint per taxonomic group with associated OECD Test Guideline

Taxonomic group	**Toxicity assay and Endpoint**
Algae	Freshwater Algae and Cyanobacteria, Growth Inhibition Test (OECD, 2016a)
Crustacea	*Daphnia* sp. Acute Immobilisation Test (OECD, 2016b)
Teleostei (fish)	Fish Acute Toxicity Test (OECD, 2016c)

c) Dealing with multiple entries for a single chemical compound per species. When multiple entries were reported for the same chemical compound in a specific species, the original publications were obtained to assess their quality. Adapting a precautionary stance to ensure sensitivity and better representation of the central tendency of the available data, if the difference between entries for the same compound was lower than 10%, the geometrical mean was calculated and used for further analysis, if the difference was greater than 10% the lowest value was selected.

d) Chemical structure identifiers. For all chemicals the chemical name, Simplified Molecular Input Line Entry System (SMILES) string, IUPAC International Chemical Identifier (InChI) and CAS registry number(s) (CAS RN) were retrieved from ChemSpider (http://www.chemspider.com), PubChem (Kim *et al.*, 2016) or generated using Accelrys (2002). For cases, where only one identifier could be found in the initial reference, the original reference was obtained to ensure that the correct substance was documented.

e) Calculation of molecular descriptors. The logarithm of the octanol-water partition coefficient (log K_{OW}) is the main descriptor used in aquatic toxicity prediction as it is a good surrogate and measure of the hydrophobicity of a substance in biological and environmental processes (Madden, 2010). There are different calculation methods for log K_{OW} values with varied results

depending on the compound (e.g. Dearden *et al*., 2007). Calculated log K_{OW} data with experimental validation are generally preferred for analysis, however due to limited availability of experimental log K_{OW} values for the entire inventory, only calculated log K_{OW} values were used for further analyses. Log K_{OW} calculation was conducted using the KOWWIN software within the Estimation Program Interface Suite (EPISuite v.4.10) from US EPA, as it is widely used in a regulatory context and for decision-making. Using the Konstanz Information Miner (KNIME) interface (https://www.knime.com/), the relative molecular mass for each compound was calculated using the open source RDKit Descriptor calculation node (http://www.rdkit.org/C++_Docs/MolDescriptors_8h.html). Another key molecular descriptor that was calculated was the logarithm of the distribution coefficient (log D) i.e. the octanol-water partition coefficient for a compound corrected for the effects of ionisation. Log D was calculated at pH 7.5 and 8.5 using Percepta by ACD/labs (https://www.acdlabs.com/products/percepta/index.php) and there values were verified by ensuring that tested concentrations and experimental EC50 values were below water solubility values.

As the objective of this chapter was to compile high quality acute toxicity datasets for aquatic species, datasets that me*et* all criteria other than OECD guidelines were considered 'Reliable with restrictions' in the context of the Klimisch criteria (Klimisch *et al*., 1997)). Four high quality datasets found in literature were included in the analysis in addition to those found from the literature search described above:

a) Fathead minnow dataset (US Environmental Protection Agency, 2008) has been developed by the by the US EPA Mid-Continental Ecology Division (US Environmental Protection Agency, 2008) and has been the basis for the in-house *in silico* classification scheme developed by Russom *et al.* (1997). All experimental parameters for all compounds are publicly available, along with information relating to organic chemical

class assignments, acute toxicity (96h LC50, mg/L and mmol/L), dose response assessment LC50 Ratio and Excess Toxicity Index), behavioural assessments (Fish Behaviour Test), joint toxicity MOA evaluations of mixtures (MOA Mixture Test), and additional MOA evaluation of fish acute toxicity syndrome (FATS) in rainbow trout. The test parameters are similar to OECD 203 (https://www.oecd-ilibrary.org/environment/test-no-203-fish-acute-toxicity-test_9789264069961-en), however they have been conducted prior to the requirement for GLP certification therefore receive Klimisch score of 2 for reliability (i.e. 'Reliable with restrictions').

b) *Chlorella vulgaris* dataset with results obtained using a Fluorimetric Hydrolase Determination Assay protocol (Netzeva *et al*., 2004). The dataset reported by Netzeva *et al.* (2004) was considered for further analysis as the chemicals tested are highly relevant (e.g. phenols, anilines, benzenes), and testing was performed in the same laboratory on a unicellular algal species, providing all information on the experimental design;

c) *Aliivibrio fischeri* database (as found in supplementary documentation of (Steinmetz *et al*., 2014) performed using the Microtox test, i.e. the inhibition of bioluminescence (ISO, 2016) protocol. The data were considered for further analysis as they were measured using a standardised protocol (i.e. ISO 11348). The high level of standardisation of the method implies little variation may be assumed, in addition Steinmetz *et al.* (2014) have further curated the data to assure high quality; and

d) The *Tetrahymena pyriformis* dataset (as found in supplementary documentation of (Ruusmann and Maran, 2013)) using the *T. pyriformis* population growth impairment assay (Tetratox) (Schultz, 2008). This dataset was considered for further analysis as it contained numerous chemical classes associated with the mechanisms of action under investigation in this thesis e.g. non-polar narcosis, polar narcosis, non-specific reactivity etc., in a unicellular species, where all studies were conducted in the same laboratory and experimental parameters are provided.

For every toxicity value the following information was recorded: chemical name, CAS RN, SMILES string, InChI code, pEC50 (mmol/l), EC50 (mmol/l), EC50(mg/l) (both nominal and measured exposure concentration where available), species, endpoint, exposure time, relative molecular mass, calculated and experimental (when available) log K_{OW} values and the original source reference. The inventory was characterised to identify functional groups using the OECD QSAR Toolbox ver. 4.1 (OECD, 2018a).

2.3.2. Database construction and analysis

All toxicity data were organised in tables per species in database format (.accdb), with InChI as the unique identifier per chemical compound. The format of the database enabled easy retrieval of subsets based on criteria of interest (e.g. set of chemicals with available toxicity values in multiple species). Linear regression analysis of subsets of interest and graphic representation of the results were performed using Minitab v.18.1 (2010). Analyses in following chapters utilised datasets from the constructed database of more than 100 compounds per aquatic species.

2.4. Results and Discussion

The main aim of this chapter was to compile high quality robust acute toxicity data for a range of aquatic species. The main inclusion criteria for data curation was compliance with OECD Test Guidelines (see Table 2.2), adequate information on compounds studied (i.e. identifiers), access to original references for information on experimental design, and effective concentration below water solubility to ensure data quality. The four datasets for fathead minnow, *Chlorella vulgaris, Tetrahymena pyriformis* and *Aliivibrio fischeri* used are non-standard tests without OECD Test Guidelines. In respect to the Klimisch criteria, the fathead minnow, *C. vulgaris*, *T. pyriformis*, and *A. fischeri* datasets are data from literature with well described test parameters, well documented and scientifically acceptable for the purposes of QSAR development, falling under the 'Reliable with restrictions' Klimisch category (Klimisch *et al*., 1997). In summary, the database created in this investigation contained data and information for 5,085 unique chemical compounds with

6,243 experimental acute toxicity data across 18 aquatic species – a summary is provided in Table 2.3. A detailed description of the data in Table 2.3 can be found in the Supplementary Information (Appendix 1, .accdb file). The database covers approximately 200 functional groups and is representative of the chemistry of industrial relevance for multiple aquatic species across all prominent levels of the food chain (Table 2.4, Figure 2.1).

Table 2.3. Summary of the acute aquatic toxicity database organised by trophic level or species.

Taxonomic Group	Species	Number of chemicals	EC50 range (min – max, mg/l)	Exposure Time	References
Algae	*Chlorella pyrenoidosa*	11	2.60 – 175	72 h	(Ramos *et al*., 1999)
	Dunaliella tertiolecta	30	0.24 – 289.62	48 h	(Erturk *et al*., 2012)
		30	0.29 – 289.62	72 h	
		30	0.32 – 296.36	96 h	
	Pseudokirchneriella subcapitata	397	0.00005 – 8470	48 h	(Singh *et al*., 2014b, Chen *et al*., 2009, Lee and Chen, 2009)
		201	0.00002 – 36638	72 h	(Aruoja *et al*., 2011, Aruoja *et al*., 2014, Chen *et al*., 2009, Dom *et al*., 2010, Furusjo *et al*., 2006, Furuhama *et al*., 2015c, Lee and Chen, 2009, Pretti *et al*., 2009, Singh *et al*., 2014a, Singh *et al*., 2014b, Zhang *et al*., 2010, Brausch and Rand, 2011)
		44	0.08 – 149.13	96 h	(Aruoja *et al*., 2011, Aruoja *et al*., 2014, Chen *et al*., 2009, Dom *et al*., 2010, Furusjo *et al*., 2006, Furuhama *et al*.,

					2015c, Lee and Chen, 2009, Pretti *et al*., 2009, Singh *et al*., 2014a, Singh *et al*., 2014b, Zhang *et al*., 2010)
	Scenedesmus obliguue	57	0.0000006 – 114.5	48 h	(Singh *et al*., 2014b, Zhang *et al*., 2010)
	Scenedesmus obliquus	38	1.53-489	48 h	(Lu *et al*., 2001)
	Scenedesmus pannonicus	1	-	72 h	(Brausch and Rand, 2011)
	Scenedesmus subcapitatus	1	-	48 h	(Brausch and Rand, 2011)
		1		72 h	
	Scenedesmus vacuolatus	39	0.00007 – 1612.72	24 h	(Singh *et al*., 2014a)
	Selenastrum capricornutum	1	-	96 h	(Brausch and Rand, 2011)
	Skeletonema costatum	1	-	96 h	(Brausch and Rand, 2011)
Crustacean	*Daphnia magna*	844	0.000000002 – 59.87	48 h	(Davies *et al*., 2004, Moosus and Maran, 2011, Roberts *et al*., 2013, Roy and Das, 2013, von der Ohe *et al*., 2005, Zhang *et al*., 2013)
Protozoa	*Tetrahymena pyriformis*	2072	0.06 – 24610.08	40 h	(Ruusmann and Maran, 2013)
Teleostei	*Lepomis macrochirus* (bluegill)	296	0.0005 – 27539.9	96 h	(Barron *et al*., 2015)
	Oncorhynchus mykiss (rainbow trout)	328	0.00006 – 134.15	96 h	(Barron *et al*., 2015)
	Pimephales promelas (fathead minnow)	620	0.0002 – 75200.4	96 h	(Barron *et al*., 2015)

	Poecilia reticulata (guppy)	165	0.0007 – 62602.61	96 h	(Verhaar *et al*., 1992)
Bacteria	*Aliivibrio fischeri*	816	0.001 – 320242.16	5,10,30 min	(Steinmetz *et al*., 2014)

Table 2.4. Chemical space covered by the database, presented are functional groups found in at least 8 compounds within the inventory with additional 151 further functional groups found within 7 or fewer chemicals within the inventory. n = number of compounds with at least one instance of this functional group within its structure

Functional group	n	Functional group	n	Functional group	n
Acetoxy	45	Carboxylic acid	231	No functional group found	68
Acrylate	52	Carboxylic acid ester	435	Organic amide and thioamide	95
Acrylic acids	14	Coumaran	9	Oxazole/Izoxazole	8
Acyl halide	11	Cycloalkane	115	Oxolane	21
Alcohol	395	Cycloalkene	83	Perhalogenated carbon derivatives	15
Aldehyde	140	Cycloketone	71	Phenanthrene	9
Aliphatic amine, primary	102	Diketone	57	Phenol	447
Aliphatic amine, secondary	39	Dinitroaniline	10	Phosphate ester	27
Aliphatic amine, tertiary	64	Disulfide	17	Piperazine	8
Alkane, branched with quaternary carbon	21	Dixydroxyl derivatives	42	Piperidine	10
Alkane, branched with secondary carbon	160	Enol	8	Precursors quinoid compounds	147
Alkane, branched with tertiary carbon	224	Epoxide	23	Pyrazine	11
Alkene	374	Ether	394	Pyridine/Pyridinium ion	213
Alkenyl (hetero)arenes	16	Fluorene	9	Pyrimidine	27
Alkenyl halide	42	Furan	43	Pyrrolidine	14
Alkoxy	178	Fused carbocyclic aromatic	89	Quaternary ammonium salts	67
Alkyl	13	Fused heterocyclic aromatic	70	Quinoid compounds	28
Alkyl (hetero)arenes	410	Fused saturated carbocycles	19	Quinoline/Isoquinoline	29
Alkyl halide	404	Fused unsaturated carbocycles	33	Saturated heterocyclic fragment	166
Alkyl-, alkenyl- and alkynyl (hetero)arenes	431	Fused unsaturated heterocycles	19	Steroids	11
Alkyne	105	Guanidine	10	Sulfate	10
Allyl	247	Haloacetamide	19	Sulfide	57
Amidine	62	Heterocyclic phenols	31	Sulfonamide	15
Amine, primary	347	Hydrazide	11	Sulfone	24
Amine, secondary	74	Hydrazine derivatives	30	Sulfonic acid	13
Amine, tertiary	173	Hydrazone	13	Sulfoxide	11
Ammonium salt	71	Imidazole	62	Surfactants - Anionic	13
Aniline	218	Imide	15	Surfactants - Cationic	11
alpha,beta-Unsaturated aldehyde	27	Imidic acid	60	Terpenes	51
Aromatic amine	60	Isocyanate	13	tert-Butyl	92
Aromatic perhalogencarbons	43	Isopropyl	151	Tetrahydropyran	8

Aryl	2265	Isothiocyanate	45	Thiazole/Isothiazole	12
Aryl halide	693	Ketone	253	(Thio)Phosphonic acid derivatives	12
Benzamide	35	Ketoxime derivatives	11	Thiocarbamate	11
Benzofuran/Dihydrobenzofuran	9	Lactons	17	Thiol	18
Benzothiazole/Benzoisothiazole	9	Methacrylate	23	Thiophene	36
Benzyl	130	Morpholine	16	Thiophosphate	45
Bicycloheptane	15	Naphtalene	89	Thiourea derivatives	33
Biphenyl	49	Naphthoquinone	11	Triazole	14
Bridged-ring carbocycles	31	Nitrile	177	Unsaturated carbocyclic fragment	89
Bridged-ring heterocycles	9	Nitroaliphatic compounds	18	Urea derivatives	13
Carbamate	14	Nitrobenzene	293		

There are a number of caveats relating to the curation of the data presented. The main challenges faced were with regard to the data collection and curation for algal acute toxicity data. At the time of data collection, there were no curated acute toxicity datasets for algal species with more than 50 compounds. From the 74 publications that were available with experimental EC50 values for all aforementioned species, only 20 publications had information available on the exact test species, whether the study design complied with OECD Test Guidelines and the original reference of the study was available and accessible. For only 22% of the dataset, was there information available on whether the EC50 values correspond to measured or nominal values (reported under remarks section per species in Appendix 1). Most commonly, there was no information on the exact test species (e.g (Sanderson and Thomsen, 2009, Escuder-Gilabert *et al*., 2001)); and studies that misreported information from the original studies (e.g. (Das and Roy, 2014)). Katritzky (2001) found a notable example of misreported information in a publication with a *P. reticulata* (guppy) dataset; on checking, the acute toxicity data for the phenolic compounds found in the published dataset come from *P. promelas* (fathead minnow) acute toxicity studies and not guppy. Misreported data were excluded from further analysis. In order to ensure a high level of curation and the level of potential errors, quality control was performed by generating random sets of compounds (n=10% of dataset per species) and checking the reporting of identifiers, calculated molecular descriptors and acute toxicity values.

Datasets with more than 100 compounds were selected for further analysis.

2.4.1. Calculation of molecular descriptors

There is a variety of software for the calculation of molecular descriptors. In an attempt to rely on open source software that may be used for decision-making, all calculations were produced using the US EPA's freely available EPISuite software (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface) and KNIME (https://www.knime.com/). Calculation of descriptors allows for the comparative analysis of the compounds in the database. For all data, calculated log K_{OW} values ranged from -7.53 to 11.47, with a mean of 2.25 and standard deviation 1.61. For individual datasets of toxicity data with more than 100 compounds, frequency distributions of log K_{OW} values for species were created and are presented in Figure 2.1a. In addition, a comparison of the relative frequency distributions is shown in Figure 2.1b. As observed, the log K_{OW} values for all datasets, have comparable frequency distributions and log Kow value ranges.

a)

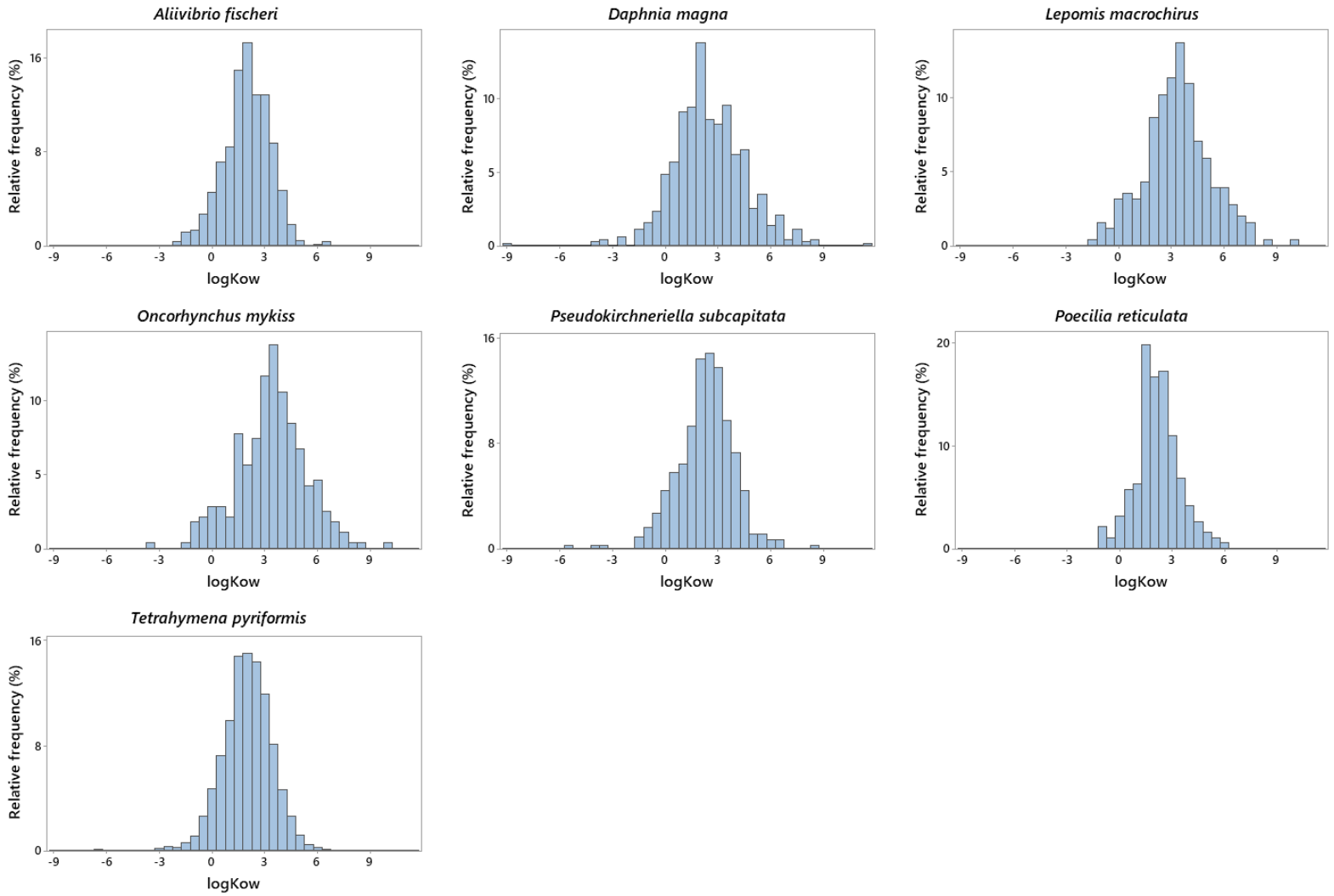


Figure 2.1. Relative frequency (%) of logarithm octanol-water partition coefficient (log K_{ow})values per dataset (when n>100) per species (a), as calculated using EPISuite

b)

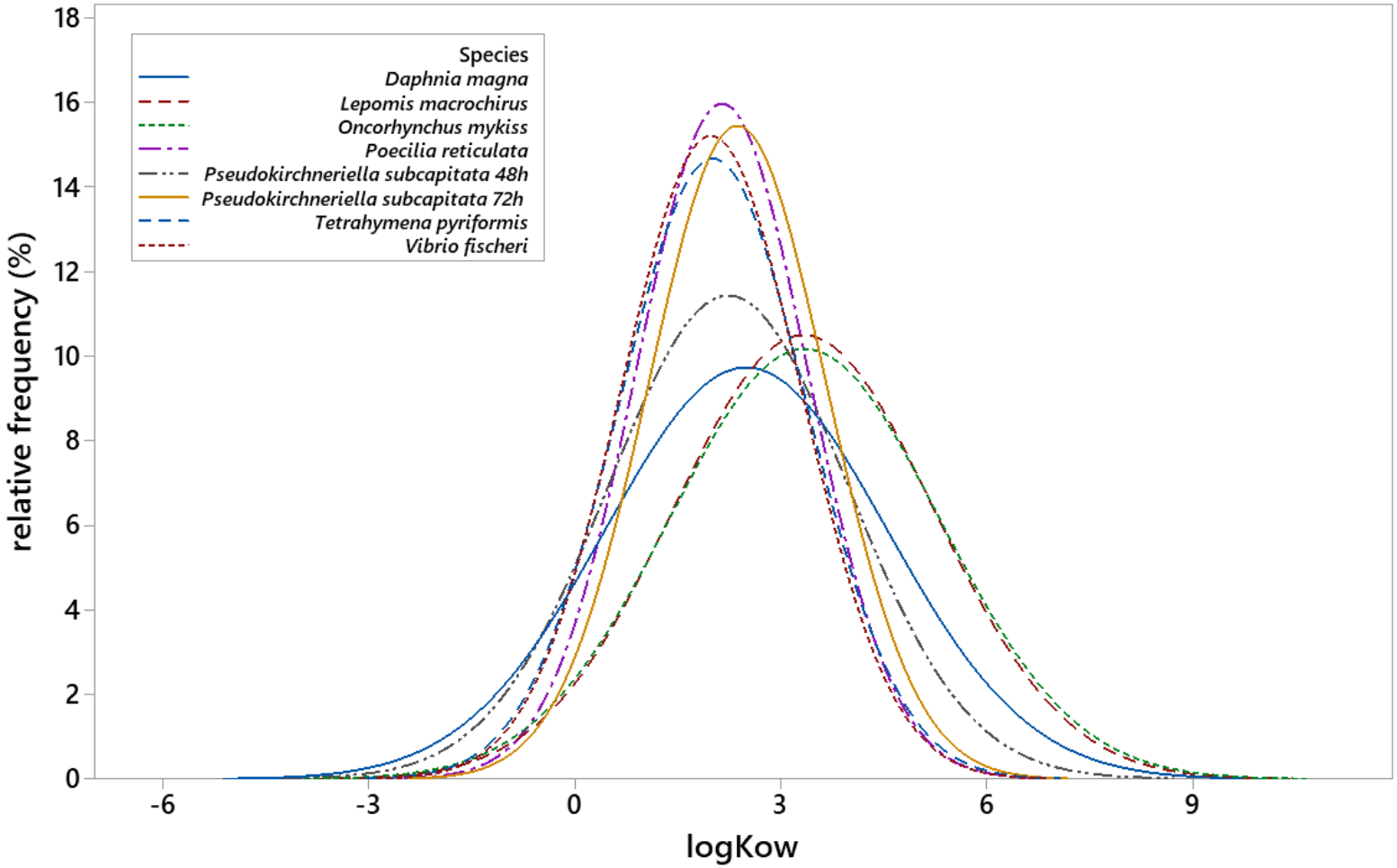


Figure 2.1. (continued) and how relative frequency curves of logarithm octanol-water partition coefficient (log Kow) compare among species (b)

a)

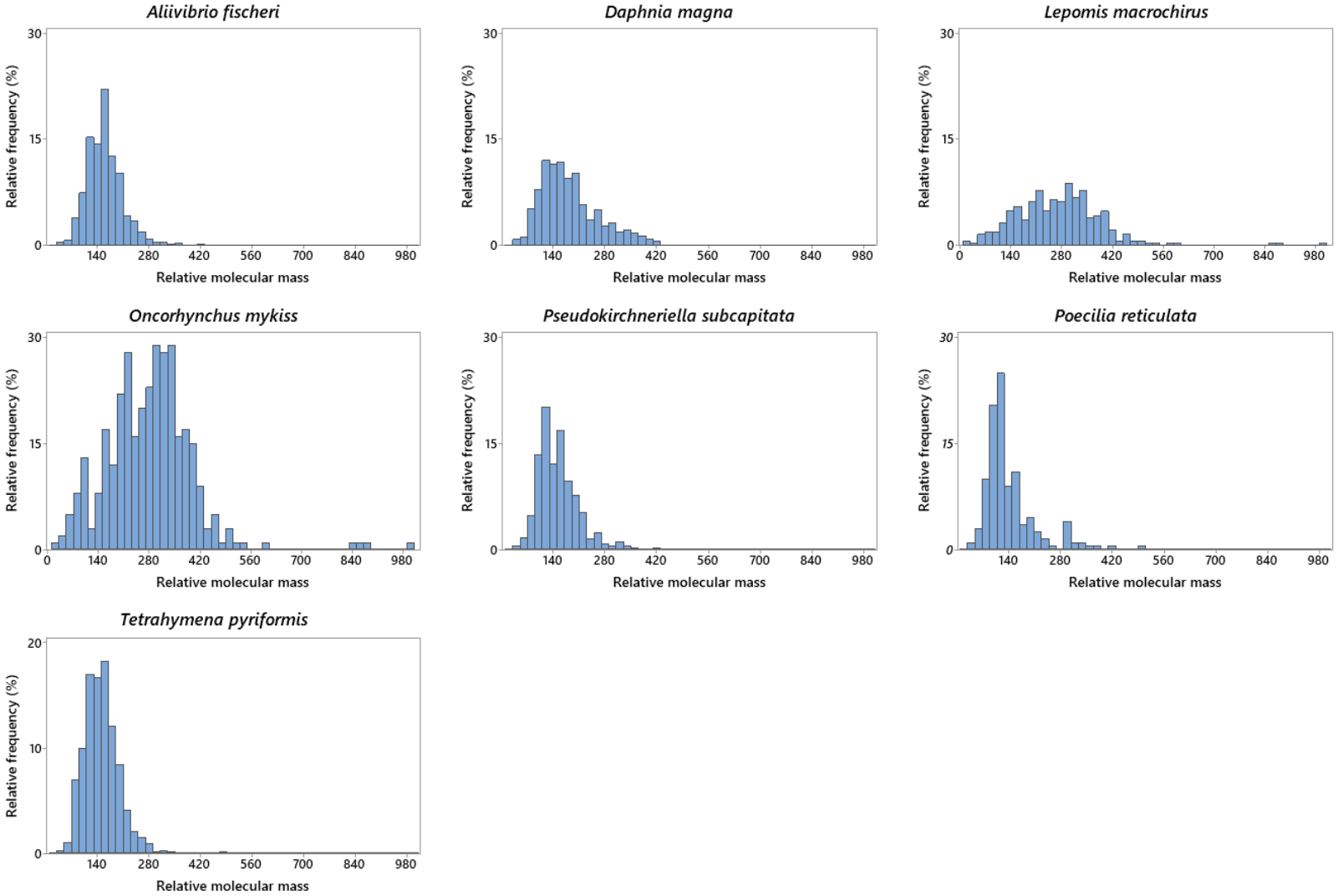


Figure 2.2. Relative frequency (%) of relative molecular mass values per dataset (when n>100) per species (a), as calculated using KNIME

b)

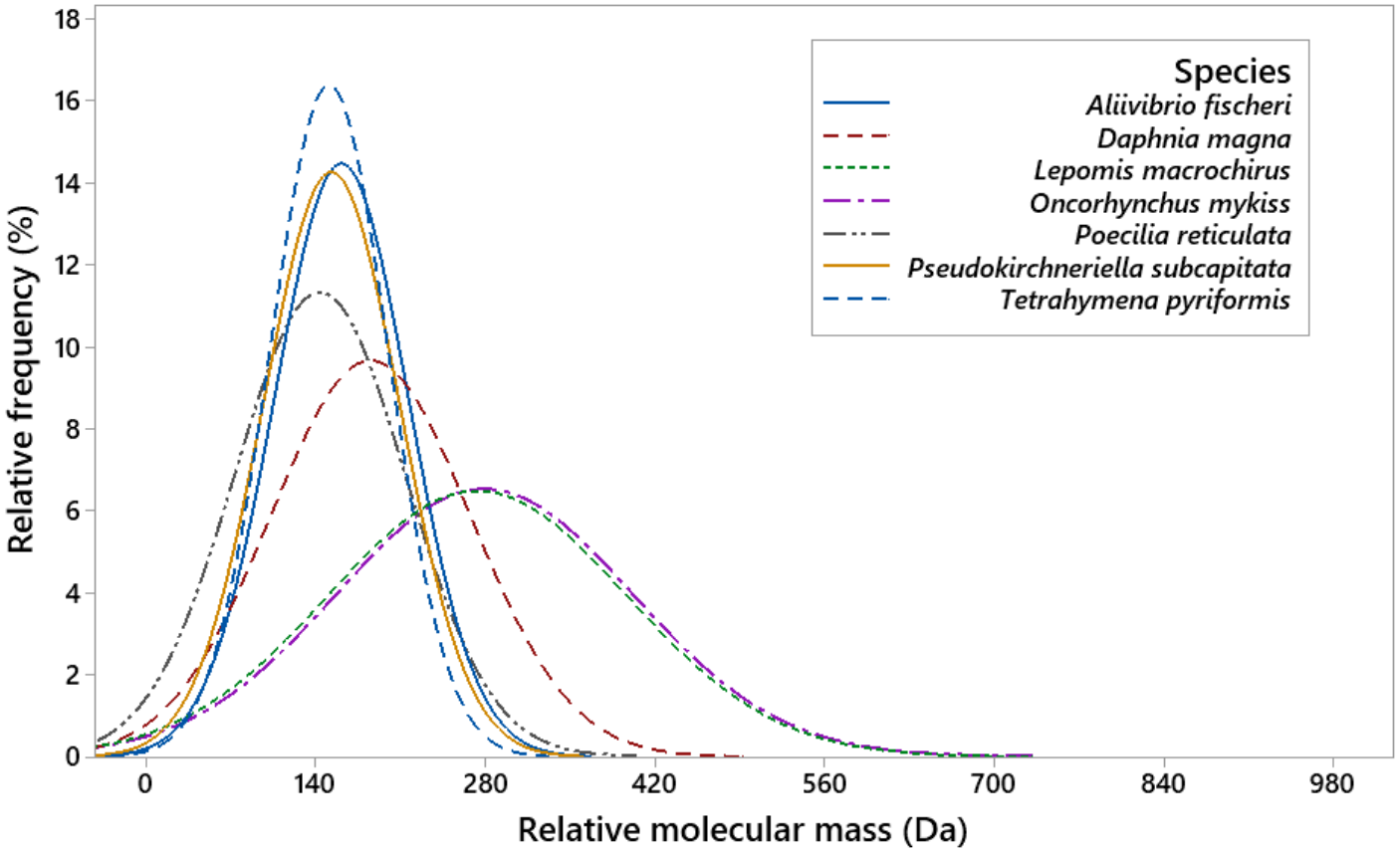


Figure 2.2. (continued) and how relative frequency curves of relative molecular mass compare among species (b)

Similarly, for all data, relative molecular mass values were calculated with values ranging from 17.03 to 1175.16 Da, a mean value of 192.51 and standard deviation of 104.98. For datasets with more than 100 compounds, the frequency distributions of relative molecular mass values per species are presented in Figure 2.2a and a comparison of the relative frequency distributions in Figure 2.2b. For the majority of species, the relative frequency distribution for relative molecular mass is comparable, with the exception of *L. macrochirus* and *O. mykiss*, which have a broader distribution and compounds with a higher relative molecular mass

Utilising these descriptors, log K_{OW} vs pEC50 plots were produced for all compounds with available toxicity data per species and these are presented in Figure 2.3. Traditionally, log K_{OW} vs pEC50 regression relationships are associated with non-specific toxicity (i.e. non-polar and polar narcosis) (Könemann, 1981). Figure 2.3 shows, as anticipated, all species have a trend for increasing toxicity with increasing hydrophobicity as well as a demonstrable baseline.

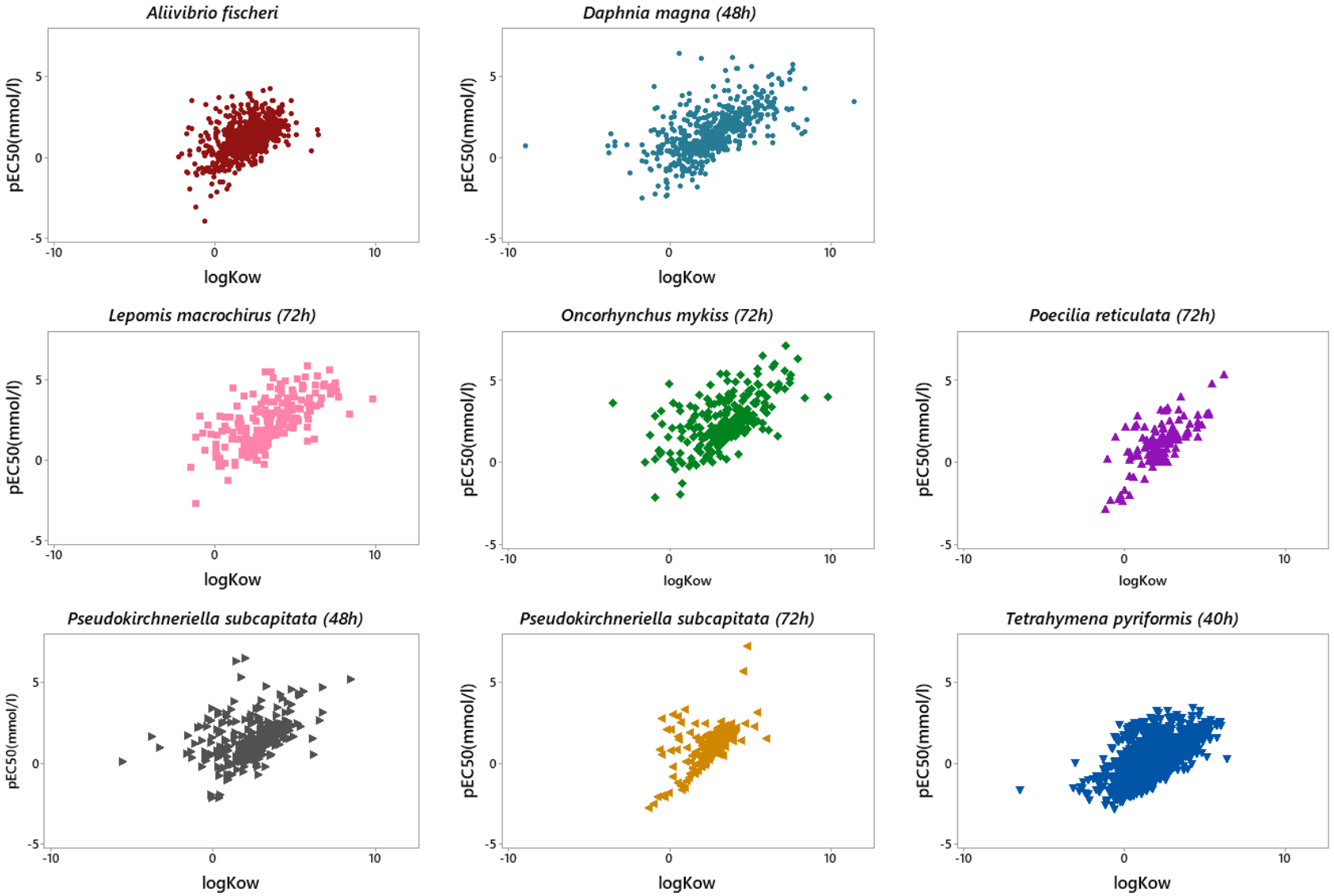


Figure 2.3. log K_{ow} – pEC50 plots for acute aquatic toxicity data per species and duration as collected from literature.

2.4.2. Interspecies relationships

The structure of the database enables easy retrieval of combined acute toxicity data from multiple datasets. Depending on data availability, acute toxicity data from multiple species were retrieved for a set of chemicals and a preliminary interspecies analysis of acute toxicity was performed.

For a set of 215 compounds, acute toxicity data (96h LC50) for two fish species were available. The relationship between the toxicity data for *O. mykiss* and *L. macrochirus* is shown in Figure 2.4. Regression analysis indicated a strong correlation between acute toxicity data for the two species (r^2 = 0.91). The observed R squared value suggests a strong correlation and, along with the small intercept and slope close to one, could suggest that potency among species is similar.

O. mykiss pEC50(mmol/l) = 0.22 + 0.96 * *L. macrohirus* pEC50(mmol/l)

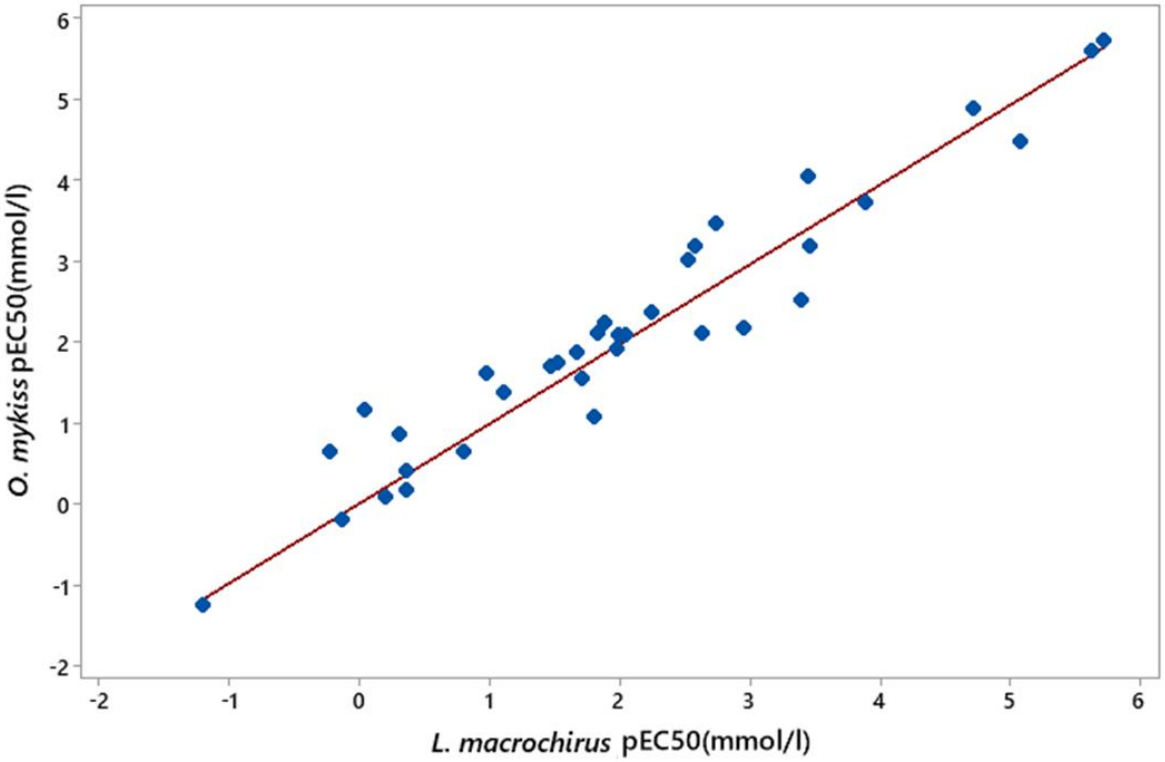


Figure 2.4. Regression analysis of acute toxicity data from two fish species for a set of 215 compounds (r^2 = 90.5 P=0.00)

Similarly, for a set of 48 compounds, acute toxicity data for three fish species (*O. mykiss*, *L. macrochirus, P. promelas*) were retrieved and analysed (Figure 2.5). Results of the regression analyses of the interspecies relationships of toxicity are presented in Table 2.5. All interspecies relationships presented in Table 2.5 are based on datasets covering a range of MOA; with observed low intercepts and slopes approaching unity. This could potentially support interspecies extrapolation of acute toxicity data across these fish species. Both analyses on fish acute toxicity datasets suggest that interspecies extrapolation among the studied species and endpoints can be performed with a high level of confidence. This is partially expected as the fish studies analysed had the same endpoints and were initially found from the same study in literature (i.e. they may have been performed in the same institution by the same operators). The work by Barron *et al*. (2015) aimed to create a gold standard for method development, creating datasets of measured and predicted acute toxicity values for a set of 672 compounds for three fish and four daphnid species.

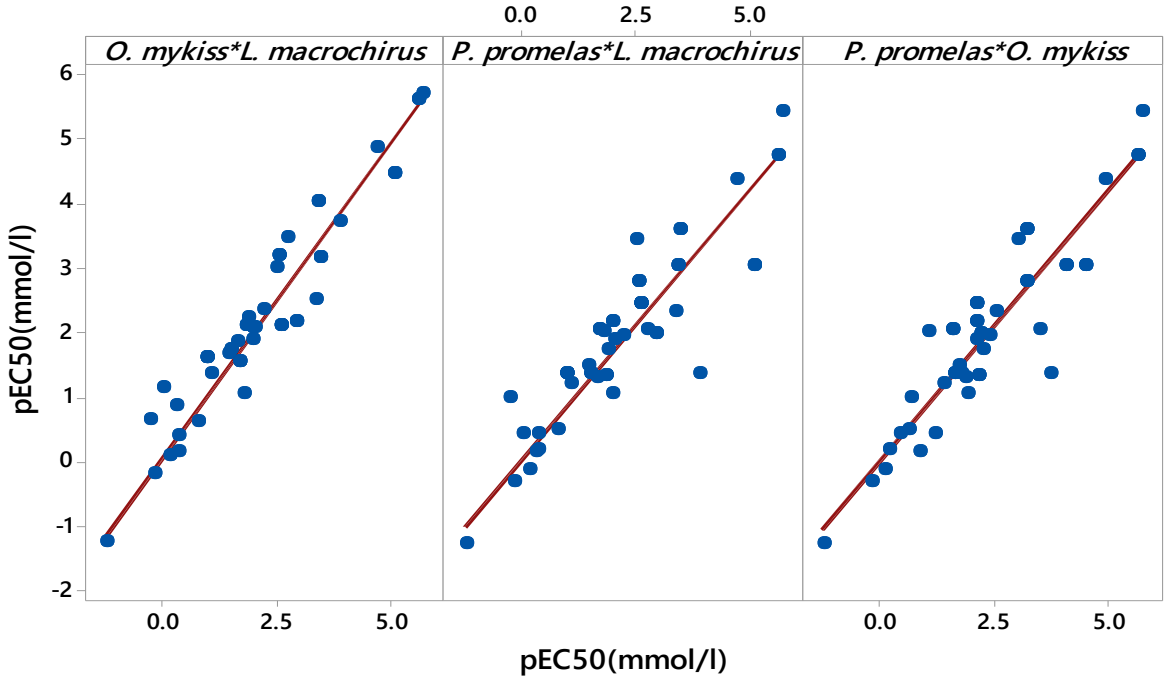


Figure 2.5. Regression analysis of acute toxicity data (96h LC50) from 3 fish species for n=48

Taking a step further, interspecies relationships of acute aquatic toxicity values obtained from different endpoints and genera were investigated. The aim of this analysis was to observe data trends and potential relationships, rather than generate and develop QSARs.

For the acute toxicity data of 91 comparable compounds to *D. magna* (48h EC50) and *P. promelas* (96h LC50) the regression model had $r^2 = 0.74$ and s = 0.70; the plot of the relationship is shown in Figure 2.6. Outliers indicative of *Daphnia* sensitivity (i.e. relatively more toxic to the invertebrate than the fish) were 2,4,6-trichlorophenol, malathion, 4-chloroaniline, aniline, n-methylaniline and 4-ethylaniline. This observation is supported by the literature confirming the hypothesis that aniline-, malathion- and 2,4,6-trichlorophenol-mediated toxicity is more prominent in *Daphnia* than the fathead minnow (anilines (Ramos *et al*., 2002, Cronin *et al*., 2000, Trac *et al*., 2016); malathion (Garza-Leon *et al*., 2017, Toumi *et al*., 2015, Trac *et al*., 2016); 2,4,6-trichlorophenol (Cronin *et al*., 2000, Olaniran and Igbinosa, 2011)). If regression analysis was to be performed without the aforementioned outliers the intercept would be 0.05 and the slope 0.86, giving with r^2=0.88, suggesting a stronger correlation. This highlights that for interspecies extrapolation across species from different taxa, mechanism of action should be considered to inform the prediction (e.g. preliminary evidence indicating that aniline toxicity to daphnids is cholinesterase-mediated (Kienzler *et al*., 2017)).

Table 2.5. Equation and statistics for the interspecies regression analysis on the three fish toxicity datasets (all data in mmol/l)

Species	**Equation and Statistics**
P. promelas	*P. promelas* pEC50 =0.28 + 0.76 * *L. macrochirus* pEC50
L. macrochirus	n = 48; $r^2 = 0.82$; s = 0.57
O. mykiss	*O. mykiss* pEC50 = 0.28 + 0.90 * *L. macrochirus* pEC50
L. macrochirus	n = 48; $r^2 = 0.91$; s = 0.45
O. mykiss	*O. mykiss* pEC50 = 0.26 + 1.03 * *P. promelas* pEC50
P. promelas	n = 48; $r^2 = 0.84$; s = 0.62

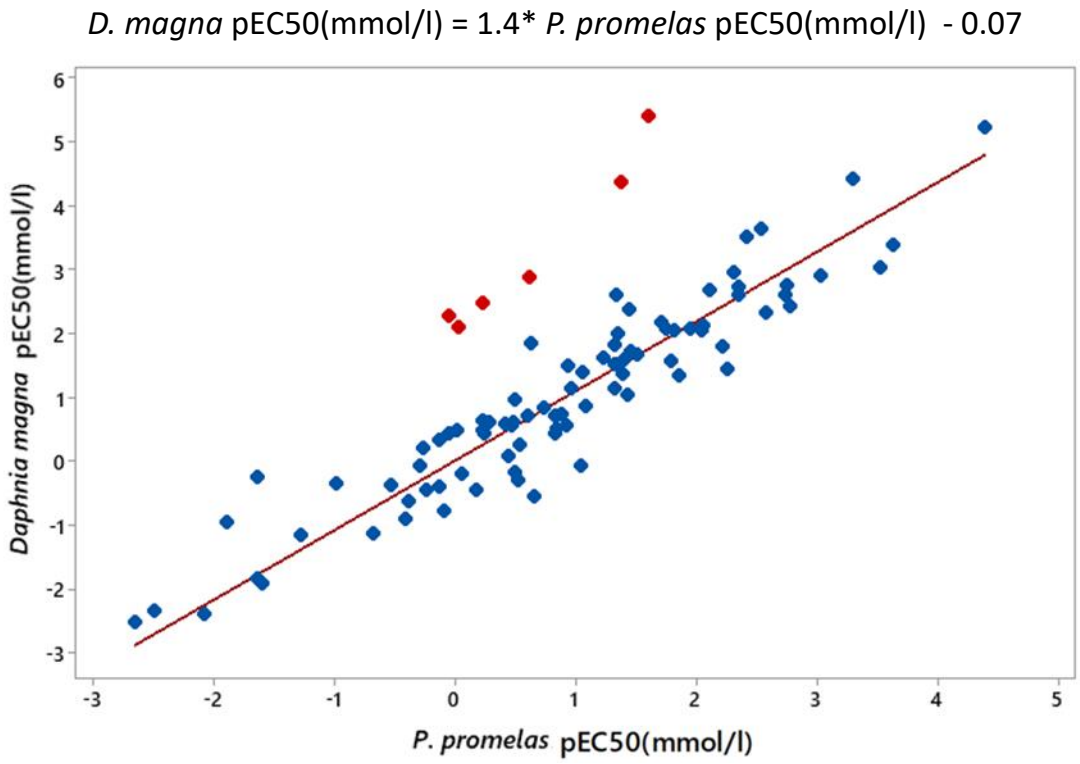


Figure 2.6. Regression analysis of 91 acute toxicity data to *P. promelas* (96h LC50) and *D. magna* (48h EC50). Outlier data points in red.

338 acute toxicity data in common between *P. promelas* (96h LC50) and *T. pyriformis* (40h IGC50) were retrieved and analysed (Figure 2.7). The regression analysis indicated a significant interspecies relationship with r^2 = 0.75 and s = 0.64. The majority of the outliers expressing higher toxicity in *P. promelas* were suspected of acting via an electrophilic mechanism (e.g. n-vinylcarbazole; allyl methacrylate; acrolein; allyl alcohol; 2-butyn-1-ol). The electrophilicity of compounds leading to excess toxicity are often mediated by metabolism, something that could partially explain the observed results if a metabolic pathway is present in fish and absent in *T. pyriformis* (Figure 2.7) (Enoch and Cronin, 2010, Enoch *et al*., 2011). The majority of outliers with higher toxicity to *T. pyriformis* were unspecifically reactive (2,4,6-triiodophenol; 2-propyn-1-ol; 2-nitrophenol; 4-nitrophenol; 2,2'-methylenebis(4-chlorophenol); 3-acetamidophenol; 4-nitroaniline; methyl sulfoxide), suggesting that lack of organism complexity (i.e. defence mechanisms such as glutathione) could be a contributing factor to excess toxicity. All MOA information was retrieved from the FHM database (US Environmental Protection Agency, 2008)). A comprehensive quantitative structure-activity-activity

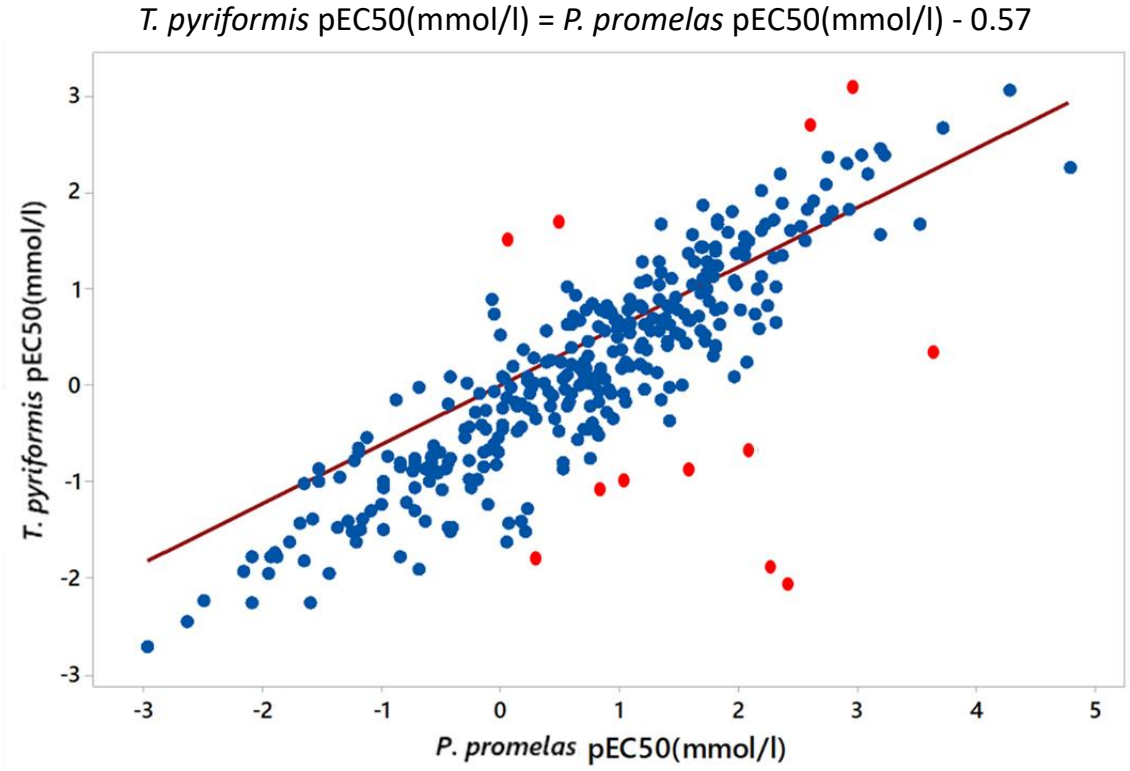


Figure 2.7. Regression analysis of 91 acute toxicity data to *P. promelas* (96h LC50) and *D. magna* (48h EC50). Outlier data points in red.

relationship (QSAAR) study on *P. promelas* (96h LC50) and *T. pyriformis* (40h IGC50) by Kahn *et al.* (2007) is in agreement with the analysis presented here.

Lastly, for 213 chemicals acute toxicity data for *D. magna* (48h EC50) and *P. subcapitata* (72h EC50) were retrieved and analysed (Figure 2.8). The regression was significant with an $r^2 = 0.62$ and s = 0.68. Outliers in Figure 2.8 are the compounds 3,4-dichlorotoluene and 4-(1-methyl-1-phenylethyl)-phenol and phenothiazine. Both *Daphnia* sensitivity to 3,4-dichlorotoluene and 4-(1-methyl-1-phenylethyl)-phenol (i.e. aromatic amines sensitivity) and algal sensitivity to phenothiazine (i.e. photosynthesis inhibition), as identified by the analysis here, is supported by literature demonstrating it is related to specific toxic action (for further details see (Furuhama *et al*., 2015b, Burris and Black, 1983)). A high intercept 0.67 (Figure 2.8) would suggest that, for the compounds from the dataset that were analysed, *D. magna* is more sensitive than *P. subcapitata.*

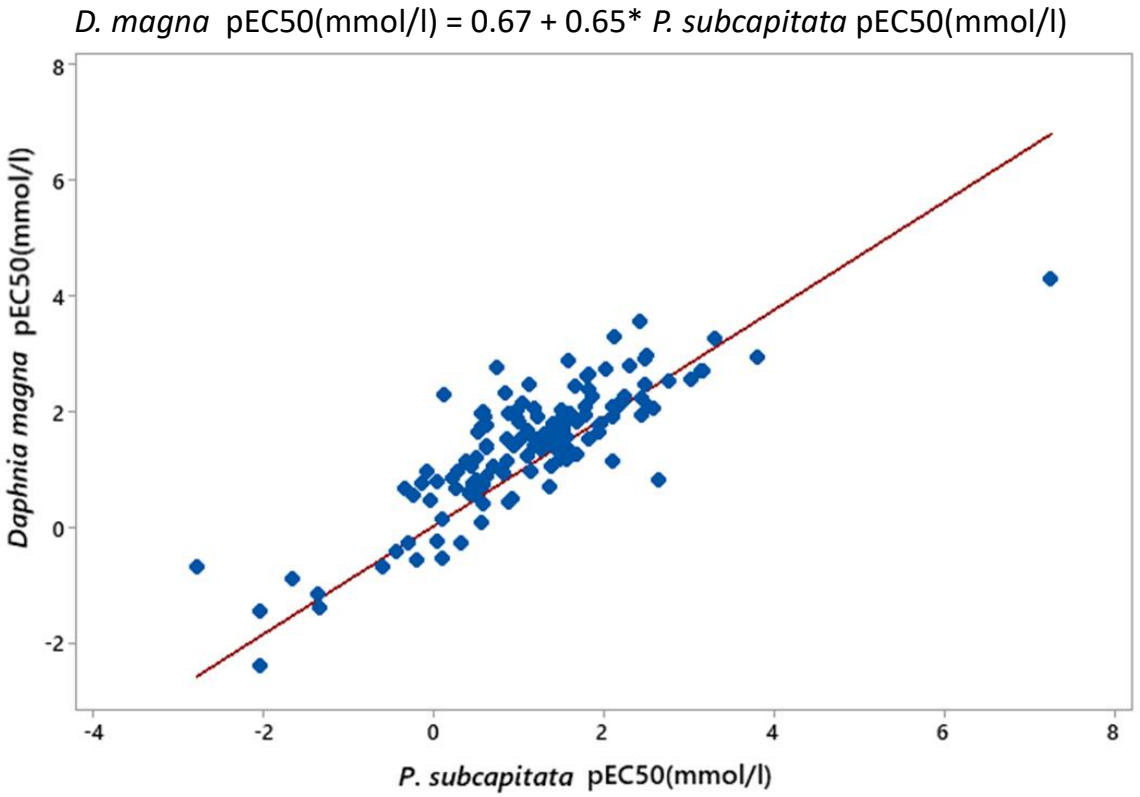


Figure 2.8. Regression analysis of acute toxicity data for *P. subcapitata* (72h EC50) and *D. magna* (48h pEC50) for 213 compounds.

The approach in this chapter was strictly data and chemistry driven, with no mechanistic data (MOA) supporting the analysis. As demonstrated, even though it was feasible to adequately derive reliable Quantitative Activity-Activity Relationships (QAARs), i.e. interspecies relationships, for the prediction of acute toxicity across fish species; this was challenging for cross-genera extrapolation (i.e. *D. magna* – algae). It was also difficult to rationalise the outliers observed without resorting to a mechanistic/MOA-centred explanation. A more mechanistically informed approach in *in silico* modelling and analysis will assist in unlocking the full potential of current knowledge for the aquatic toxicity data, and this is explored in subsequent chapters of this thesis.

2.5. Conclusions

With the increasing role of computational methods in regulatory decision-making, it is essential to ensure good practice in QSAR development with the starting point being high quality data relevant to the applicability domain of interest (Dearden *et al.*, 2010). This chapter focused on investigating the availability of high-quality data for aquatic species relevant to the aquatic fauna and flora relevant for the chemical space of industrial chemicals.

Assessment of data reliability was a fundamental part of the work presented. Klimisch *et al.* (1997) comprehensively described evaluation criteria for toxicity studies regarding reliability, relevance and adequacy of data. Kase *et al.* (2016) expanded the evaluation criteria presented by Klimisch *et al.* (1997) and proposed a robust set of characteristics to consider regarding reliability, relevance and adequacy of data, while providing additional guidance for data evaluation (Moermond *et al.*, 2016). The most common misreported information on acute toxicity studies as discussed by Przybylak *et al.* (2012) is on experimental parameters, GLP and test chemical. As discussed by Moermond *et al.* (2015), the lipid solubility of the test chemical will reflect on the reliability of the experimental acute toxicity value, therefore, in this analysis for all reported acute toxicity data it was verified that the concentration related to potency fell below water solubility of the compound.

The presented acute toxicity database comprises over 6,200 toxicity data for more than 5,000 compounds (see Appendix 1). It includes data where the experimental parameters are publicly available and identification and characterisation of the test chemical is well documented. The applicability domain of the inventory is quite wide covering a number of chemical groups of industrial interest, e.g. aliphatic compounds, amines, esters, aromatic compounds, a comprehensive list of groups is given in Table 2.4). Only 22% of the acute toxicity data reported in Appendix 1 have publicly available information on whether the potency value is based on nominal or measure concentrations. If it were available, documentation of nominal and measured values as part of the description of the experimental parameters would further inform about the reliability of a study result. The majority of

acute toxicity data included in the database were measured according to GLP ('Reliable without restrictions' by Klimisch criteria (Klimisch *et al*., 1997)), with the exception of four datasets ('Reliable with restrictions' by Klimisch criteria (Klimisch *et al*., 1997)). These four datasets were included because they include chemicals of high ecotoxicological interest (e.g. anilines, phenols, sulfur-containing compounds) and all information on experimental parameters and study design are publicly available.

Datasets of more than 100 compounds were compiled for seven aquatic species from three aquatic genera. These datasets were used to investigate potential interspecies relationships and the potential the data for QSAR development. Preliminary regression analyses on subsets of interest revealed significant interspecies relationships among fish species, fish and *D. magna*, fish and *T. pyriformis*, and *D. magna* and algae (see Fig 2.5-2.8). There has been a number of studies that proposed interspecies QSARs for well-defined chemical applicability domains in functional group level (e.g. (Cassani *et al*., 2013, Furuhama *et al*., 2015b)). Furuhama *et al.* (2015b) proposed simple interspecies quantitative structure-activity-activity relationships (QSAARs) as a potential screening tool for a defined chemical applicability domain. Such approach though is strictly chemistry driven, with no mechanistic data (MOA) supporting information and with limitations on cross-genera extrapolation (i.e. *D. magna* – algae). A more mechanistically informed approach could potentially unlock the full potential of current knowledge chemical knowledge and expand current mechanistic understanding.

Chapter 3: Evaluation of current *in silico* classification schemes relevant to acute aquatic toxic action

3.1. Introduction

Following the implementation of the European REACH legislation in 2006 (European Commission, 2006), the use of alternative methods was encouraged to facilitate risk assessment of chemicals. The guidelines for risk assessment describe in detail a multi-step process that covers human and environmental health, with updated addendums to facilitate risk assessment of new technologies and materials (e.g. the guidance on ecotoxicity of nanomaterials (ECHA, 2017d)). The overall priority of the REACH legislation is a high level of protection for human and environmental health while ensuring that animal testing can be justified only as a last resort. In order to fulfill information requirements along with data waiving and sharing, alternatives (to animal) testing methods such as read-across and category formation, *in vitro* assays, (quantitative) structure-activity relationships ((Q)SARs and weight of evidence are encouraged for every step of the process. Hazard identification and the identification of mode of action (MOA) are among the first steps for environmental risk assessment (ERA), and regulatory bodies (e.g. ECHA) suggest open source *in silico* methods for prediction of acute toxic action are the Verhaar scheme and updates (Verhaar *et al*., 1992, Verhaar *et al*., 2000, Enoch *et al*., 2008) and the Acute Aquatic Toxicity MOA by OASIS (AAT OASIS) scheme that is heavily based on the expert knowledge published by Russom *et al*. (1997) (ECHA, 2008, OECD, 2013). A detailed description of the aforementioned profilers, used for hazard identification, can be found in Chapter 1 (section 1.2.1. *In silico applications for acute aquatic toxicity prediction*).

Both the Verhaar and AAT OASIS *in silico* classification schemes are based on fish acute toxicity data (Verhaar *et al*., 1992, US Environmental Protection Agency, 2008). More recently Barron and co-workers (2015), taking into consideration the limitations of the previous schemes related to mechanistic information for the training set and taxonomic applicability, created a standardised inventory of industrial organic compounds. Their work is published along with an extensive list of

compounds allocated to broad and specific modes of action. To support high confidence in the assignment of mode of action, assessment methods based on chemical structure and literature information included international consensus classifications, QSAR predictions and non-transparent expert judgement. An acute toxicity database is also provided (n=674, measured and predicted) for data to four aquatic species: rainbow trout (*Oncorhynchus mykiss*) (measured n=335, predicted n=329), fathead minnow (*Pimephales promelas*) (measured n=297, predicted n=375), bluegill (*Lepomis macrochirus*) (measured n=304, predicted n=368), and cladocerans (*Daphnia magna, Daphnia pulex, Daphnia*). With this work, Barron and his group aimed to provide a 'gold standard' to support model development for predictive aquatic toxicology. It is worth noting that the use of the terms 'broad mode of action' correspond to what the author, among others (Vonk *et al*., 2009), describes as a functional or anatomical change, at the cellular level, resulting from the exposure of a living organism to a substance; and 'specific mode of action' relates to the term mechanism of action which describes changes at the molecular level.

This chapter focuses on publicly available schemes with a strong mechanistic basis for acute aquatic toxic action relevant to industrial organic chemicals and on understanding the similarities and weaknesses as well as their applicability with a view to further improvement. There are three classification schemes where training sets and the mechanistic basis of suggested domains are publicly available that can be used to generate predictions for compounds of interest: a) the modified Verhaar scheme (Verhaar *et al*., 2000, Verhaar *et al*., 1992), b) the Acute Aquatic Toxicity profile by OASIS in the OECD QSAR Toolbox, and c) the Barron scheme (2015).

3.2. Aim of Chapter 3

The aim of this chapter was to evaluate the scheme proposed by Verhaar, the Acute Aquatic Toxicity profiler by OASIS based on Russom and the Barron scheme for the classification of chemicals according to mechanism of acute aquatic toxic action. This was undertaken to identify the strengths and areas for improvement for these schemes and used a dataset with mechanistic information (i.e. Fathead Minnow Acute Toxicity Database) and inventories containing compounds with no mechanistic information but chemistry relevant to the industrial sector (i.e. the 74,073 substances pre-registered for REACH with ECHA; Acute Toxicity Database (n = 5,085 see Chapter 2). This work aimed to further understanding of the schemes' performance in real-life scenarios. It also evaluated the reliability for model development of the recently published standardised dataset by Barron *et al* (2015).

3.3. Methods

3.3.1. Toxicity data and chemical inventories

Three inventories were used for the analysis in this chapter.

i) The first inventory comprised 462 chemical structures with mechanistic information as extracted from the Fathead Minnow Acute Toxicity Database (referred as the FHM inventory in this chapter) by the US EPA Mid-Continental Ecology Division (US Environmental Protection Agency, 2008). The FHM dataset comprises 618 compounds and includes information relating to organic chemical class assignments, acute toxicity in the fathead minnow (96h LC50, mg/L and mmol/L), dose-response assessments (LC50 Ratio and Excess Toxicity Index), behavioural assessments (Fish Behaviour Test), joint toxicity MOA evaluations of mixtures (MOA Mixture Test), and additional MOA evaluation of fish acute toxicity syndrome (FATS) in rainbow trout. Documentation on the acute toxicity studies populating the FHM dataset includes information on study design, experimental parameters and access to dose response curves. EC50 values are calculated based on measured concentrations and chemistry of interest is identified

clearly with information of the source. As all this information is publicly available, the FHM database is considered as highly reliable. However, since GLP certification it wasn't required at the time of experimentation, it is classified as 'Reliable with restrictions' under Klimisch criteria (Klimisch *et al*., 1997) and CRED scoring on reliability (Moermond *et al*., 2016). The database has been used extensively in *in silico* toxicology as a 'gold standard' for MOA-based, and other types of, QSARs. The FHM dataset was extracted from the US EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database (ftp://ftp.epa.gov/dsstoxftp). There is definitive MOA assignment for only 462 of the 618 compounds in the dataset. Only those structures were used in the analysis. The individual mechanistic classifications are given in Table 3.1. The FHM database is the basis for ASTER, so the MOA class assignment is as suggested by Russom *et al* (1997). For all substances, when not already provided by the source, a set of identifiers (SMILES string, Chemical Name, CAS registry number) was retrieved from Open Babel and Standard InChI Keys from Marvin View 16.10.24 (ChemAxon, 2018).

ii) The second inventory was larger but with no mechanistic information, comprising 5,085 compounds as found in the Acute Toxicity Database. In brief, the Acute Toxicity Database includes information relating to acute toxicity in seven aquatic vertebrate, invertebrate and plant species (96h, 72h and 48h respectively LC50, mg/L and mmol/ L; *A. fischeri* 5-30 min EC50 mmol/L). Data assessment was conducted following reliability criteria as described by Klimisch *et al* (1997) for data 'Reliable without restrictions'. Three datasets that meet all criteria and conducted under ISO protocols rather than GLP (i.e. *A. fischeri, C. vulgaris, T. pyriformis* datasets) were also included in the analysis, as they contained highly relevant information on chemistries of interest and have adequately documented study design, experimental parameters and characterisation of chemistry. The data were collected from the publicly available literature. The data extracted from the Acute Toxicity Database (referred as Acute Toxicity Inventory in this chapter) included chemical structures in SMILES notation, standard InChI keys and CAS registry numbers when available (see details in Appendix 1).

iii) The third inventory contains 74,073 substances as submitted to the European Chemical Agency (ECHA) between 1 June and 1 December 2008 with the intention of pre-registration (referred as the Pre-Registration inventory in this chapter). This list is publicly available to facilitate potential registrants to submit a joint dossier and includes information on chemical structure (SMILES notation) and CAS registry number. The Pre-Registration list was extracted from the OECD QSAR Toolbox ver 4.1 (OECD, 2018) and due to its size saved in Microsoft Access format (.accdb).

Table 3.1. Number of compounds allocated to individual modes of toxic action for the FHM inventory extracted from the Fathead Minnow Acute Toxicity Database from the US EPA (US Environmental Protection Agency, 2008)

Mode of action	**n**
Narcosis – Acrylates	5
Narcosis – Amines	2
Narcosis – Esters	21
Narcosis I	241
Narcosis II	36
Narcosis I & II	13
Reactivity	96
Acetylcholinesterase Inhibition	17
Respiratory blocker/inhibitor	4
Central nervous system seizure/stimulant	15
Uncoupler of oxidative phosphorylation	12

3.3.2. Classification Schemes

For each of the three aforementioned inventories acute toxicity mode of action assignment was generated by applying the following profilers (see detailed description in section 1.2.1.):

a) Modified Verhaar scheme

The substances were assigned to one of four modes of action classes corresponding to non-polar narcosis, polar narcosis, non-specific reactivity and specific toxic action. In the event a compound was not within the applicability domain, it was assigned to Class 5. The Verhaar (modified) decision tree implemented in ToxTree ver 2.6.13 (JRC Computational Toxicology, 2018) was used to determine

Verhaar classifications. Structures were entered in SD format, generated from the SMILES strings using MarvinView 16.10.24 (ChemAxon, 2018). As part of the Verhaar scheme evaluation, post filtering with the KNIME workflow as proposed by Ellison *et al.* (2015) was applied to refine class assignment for phenolic compounds, halogenated compounds and aliphatic alcohols.

b) Acute Aquatic Toxicity Profiler by OASIS (AAT OASIS) based on the Russom scheme

The AAT OASIS profiler was applied and one, or a combination, of the following MOAs was assigned to each compound: Basesurface narcotics, Narcotic Amines, Esters, Phenols and Anilines, Aldehydes, α,β-Unsaturated alcohols, Reactive Unspecified. Compounds were assumed to be Basesurface narcotics if not assigned in any other category. SMILES strings or CAS numbers (when available) were entered in the OECD QSAR Toolbox version 4.1 (August 2017). The compounds were matched with the platform's inventory to ensure the accuracy of the structures.

To input compounds for profiling, the OECD QSAR Toolbox provides the option of entering SMILES strings or CAS registry numbers. The Pre-registration list is within the list of inventories found in the OECD QSAR Toolbox so, as expected, all CAS numbers matched with a single entry, generating one prediction per entry. In the case of the Acute Toxicity Database, when entering files containing SMILES strings of the compounds, 5,085 input entries were matched to over 24,000 compounds. To tackle this, the input file contained CAS registry number when available. Of the 5,085 compounds, no CAS numbers were available for 325 of the compounds. Other significant contributing factors to the inability to classify compounds were the mismatch of the input and output CAS number, where one SMILES string or CAS number corresponds to multiple entries. The QSAR Toolbox provides information on the degree of SMILES-CAS number relation and for this analysis only those rated / annotated as 'High' were included. In cases where SMILES string or CAS number input could not be matched, the entries were excluded from further analysis. There were also entries that were consistently rejected

and discarded by default from the first step, either because they were not within the applicability domain (assigned as (N/A)) or labelled as confidential by the QSAR Toolbox.

c) MOATox classification list as proposed by Barron *et al.* (2015).

Finally, the MOATox classification database is a comprehensive MOA and acute aquatic toxicity dataset. It is not intended as a prediction method, but as a high-quality training set for model development, available as supplementary data in Barron *et al.* (2015) (summarised in Table 3.2). It consists of 1,213 substances assigned to six broad modes of action and 31 specific modes of action. Information on SMILES strings and standard InChI keys was retrieved for all the 1,213 substances listed by Barron *et al.* (2015). As the main focus of this thesis is industrial organic compounds, 30 chemical structures were excluded from the analysis, as they were either mixtures or organometallics. Based on availability, the FHM, Acute Toxicity and Pre-registration inventory were matched with MOATox classification data.

To facilitate the analysis, FHM, Acute Toxicity and Pre-registration inventories were organised in Microsoft Access Format (.accdb). For all compounds per inventory, there was information on a) chemical name, b) SMILES notation, c) standard InChI keys, d) CAS registry when available, suggested classification by e) modified Verhaar, f) AAT OASIS, and when available g) Barron *et al*. (2015).

Table 3.2. Description of mode of action groups in the MOATox classification database (Barron *et al.*, 2015)

Broad Mode of Action	**Specific Mode of Action**	**n**
Narcosis	Non-polar	347
	Polar	61
	Ester	48
	Other	5
Electron transport inhibition	Arsenical respiratory inhibition	22
	Oxidative phosphorylation inhibition	20
	Uncoupling oxidative phosphorylation	54
Reactivity	Acrylate	8

	Alkylation	38
	Carbonyl	13
	Chromate	3
	Cyanate/nitrile	9
	Di/trinitroaromatic	13
	Hydrazine	4
	Other	19
	Phosphide	3
AChE inhibition	Carbamate	74
	Organophosphate	211
Neurotoxicity	Alicyclic GABA antagonism	42
	Diphenyl sodium channel modulation	11
	GABA agonism	16
	nAChR agonism	8
	Other	96
	Pyrazole GABA antagonism	6
	Pyrethroid sodium channel modulation	99
	Sodium channel blocking	3
	Strychnine	4
Iono/Osmoregulatory/Circulatory impairment	Anticoagulation	25
	Metallic iono/osmoregulatory impairment	18
	Methemoglobinemia	6
	Other osmoregulatory	6

3.3.3. Comparison of class assignment among schemes

Each classification scheme assigns compounds in MOA classes that cover comparable general effects (i.e. Narcosis, Reactivity, and Specific Toxicity) but with differences in the terminology and the level of detail per effect. A good example is the Narcosis domain, where Class 1 and 2 by Verhaar corresponds to Narcosis I, Narcosis II, Narcosis III in Russom, Basesurface narcotics, Phenols and anilines, and Esters AAT OASIS. To enable comparison of the assigned MOA between schemes for all compounds, all

suggested mode of action classes suggested by modified Verhaar, Russom, AAT OASIS and MOATox were grouped in three mechanistic domains:

a) Narcosis, non-specific toxicity interactions leading to acute aquatic toxic effects (Schultz *et al*., 1991, Van Wezel and Opperhuizen, 1995, Cronin *et al*., 2000, Roberts and Costello, 2003, Chen *et al*., 2007, Aruoja *et al*., 2011, Aruoja *et al*., 2014)

b) Reactivity, non-specific reactivity interactions targeting biomolecules indiscriminately (e.g. electrophiles/proelectrophiles) (Enoch *et al*., 2011, Enoch and Cronin, 2010, Bradbury *et al*., 1990, Bradbury and Christensen, 1991, Bradbury *et al*., 1991)

c) Specific Toxicity, interactions that involve specific biomolecules and linked with well-defined effects (e.g. carbamates targeting acetylcholinesterase leading to inhibition, anilines acting as uncouplers leading to inhibition of oxidative phosphorylation).

Russom classification is not directly under evaluation in this chapter, as it is an expert system and there is no public access for *de novo* class assignment. An overview of the Russom class assignment is available (Russom *et al*. 1997), and along with previous key publications (listed above) contributed to the grouping in mechanistic domains.

3.3.4. Data analysis

The analysis aimed to evaluate whether (1) the schemes can meet stakeholders' demands using inventories of industrial relevance (i.e. Pre-Registration and Acute Toxicity inventory); (2) performance of MOATox, modified Verhaar and AAT OASIS compared to an inventory with high quality mechanistic information (i.e. FHM inventory), and; (3) the level of concordance in mode of action assignment among schemes using an inventory with mechanistic information (i.e. the FHM inventory).

To address objective (1), descriptive statistical analysis on mode of action assignment was performed for the Acute Toxicity and Pre-Registration inventories. To address objective (2) mode of action

assignment for the compounds found in the FHM inventory were used to investigate the performance of the three schemes. To assess that, the positive predictability percentage (PPP) was calculated per mechanistic group, as follows:

$$PPP = \frac{Nc}{Nc + Ni} * 100$$

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

Lastly, to address objective (3), mechanistic domain assignment of the FHM inventory using Modified Verhaar, Acute Aquatic Toxicity by OASIS profiler and information from MOATox dataset was compared in relation to the mechanistic domain assignment as found in FHM database. To assess that, the level of concordance was calculated per mechanistic group, where Nc is the number of compounds correctly classified by all schemes, and Ni is the number of compounds incorrectly classified.

3.4. Results and Discussion

The aim of this chapter was to assess the performance of widely used classification schemes with a strong mechanistic background, relevant to acute aquatic toxic action. Therefore, the focus fell onto the modified Verhaar scheme (Enoch *et al*., 2008, Verhaar *et al*., 1992, Verhaar *et al*., 2000), AAT OASIS profiler in QSAR Toolbox (v.4.1) and the MOA classifications suggested by Barron *et al.* (2015). These schemes were compared using a dataset with publicly available mechanistic information relevant to aquatic toxicity, the FHM dataset (US Environmental Protection Agency, 2008). To facilitate the analysis and comparison of the results among schemes, the modes of action as assigned by Modified Verhaar, AAT OASIS and Barron, and as found in the FHM dataset, were clustered in three groups as described in Table 3.3.

Table 3.3. Clustering of the individual mode of actions in three mechanistic domains. In the cases where compounds were not within the applicability domain of the scheme, they were flagged as non-classified.

Mechanistic domain	Verhaar	AAT OASIS	Barron	FHM inventory
Narcosis	Class 1	Basesurface narcotics	Narcosis	Narcosis I
	Class 2	Narcotic amine		Narcosis II Narcosis I & II
		Phenols and anilines		Narcosis amines Narcosis – acrylates
		Esters		Narcosis – esters
Reactivity	Class 3	Reactive unspecified	Reactivity	Reactivity
		Aldehydes		
		Alpha, beta-Unsaturated alcohols		
Specific Toxicity	Class 4		Electron transport inhibition	Oxidative phosphorylation uncouplers
			Neurotoxicity	Acetylcholinesterase inhibitors
			Iono/Osmoregulatory/Circulatory impairment	Central Nervous System (CNS) seizure agents
			AChE inhibition	Respiratory blocker/inhibitor
Non-classified	Class 5	(N/A)	-	-
	<blank>	<blank>		

3.4.1. Inventories of industrial relevance and mode of action assignment distribution

As an initial step in the analysis, Figure 3.1 presents the mode of action assignment distribution for Acute Toxicity Database and Pre-Registration inventories when modified Verhaar and AAT OASIS are applied. One factor that should be considered when comparing the performance of modified Verhaar and AAT OASIS is the major design difference between the profilers. Modified Verhaar assigns compounds to Class 5 if they do not fall in any applicability domain of its four MOA classes, whereas the AAT OASIS profiler assigns compounds to Basesurface narcotics if they do not fall in any of the other six classes. As there is no mechanistic information about the analysed datasets, no attempt will be made to assess the validity of the assigned classification at this stage.

The first striking observation from Figure 3.1a and 3.1b is that application of the modified Verhaar scheme produced classifications for 49% or fewer of the compounds. This observation highlights the limitations in chemical coverage relevant to current market. In line with the scientific precautionary principle and based firmly on evidence, rather than absence of it, class assignment by modified Verhaar scheme (Enoch *et al*., 2008, Verhaar *et al*., 1992, Verhaar *et al*., 2000) is confined by the applicability domain of the respective classes. This is a prudent approach and provides a level of confidence to its predictions, within this limited chemical applicability space.

Ellison *et al.* (2015) developed a KNIME implementation after processing through ToxTree v2.6 to refine class assignment for phenolic compounds, halogenated compounds and aliphatic alcohols. However, the class distribution was comparable to that presented in Figure 3.1 and more than half were unclassified (data in Appendix 1). Kienzler *et al.* (2017) also highlighted the high proportion of unclassified compounds by both the Russom and Barron schemes. Similar findings are observed for the Acute Toxicity Database inventory as classified by AAT OASIS (17% non-classified), with the main reasons being SMILES strings to CAS registry number mismatched and applicability domain limitations (e.g. surfactants) of the profiler. On the other hand, all compounds found in REACH Pre-Registration inventory were classified as the inventory was extracted from the QSAR Toolbox and there were no

caveats in SMILES string-CAS registry number match. It is also worth underlining that a very small percentage (n=167) of compounds were assigned as specifically acting toxicants by modified Verhaar, while the AAT OASIS scheme assigned these compounds to the Reactive unspecified mode of action (e.g. methyl carbamate, S-[(4-chlorophenyl)methyl] diethylaminomethanethioate, tris(3,4-dimethylphenyl) phosphate, 2,2,2-trichloro-1,1-bis(4-chlorophenyl)ethanol. The full list is presented in Appendix 1). There is no information on the mechanisms of action for the analysed datasets, therefore, the validity of the class assignments cannot be verified.

For MOATox by Barron *et al.* (2015), the chemical space was solely that of the structures provided. When matching those structures with the Acute Toxicity inventory, there was an overlap of less than 16% (n=727) and with the Pre-Registration inventory less than 2% (n=890). Mechanistic domain distributions as assigned by MOATox for the chemicals included in the Acute Toxicity and Pre-registration inventory subsets are presented in Figure 3.2a and 3.2b respectively.

The MOATox inventory, under the mechanistic domain assignment approach, comprises 39% of compounds in Narcosis domain, 9% in Reactivity domain, and the remaining 52% in Specific Toxicity domain. Taking into account this mechanistic domain distribution, the observed 13-fold increase in assignment to the specific toxicity domain is expected relative to the other schemes. It is worth noting that the MOATox database was intended as a 'gold standard' for model development therefore it was important to cover a wide range of chemistry and adequately cover a range of mechanisms associated with specific toxic action.

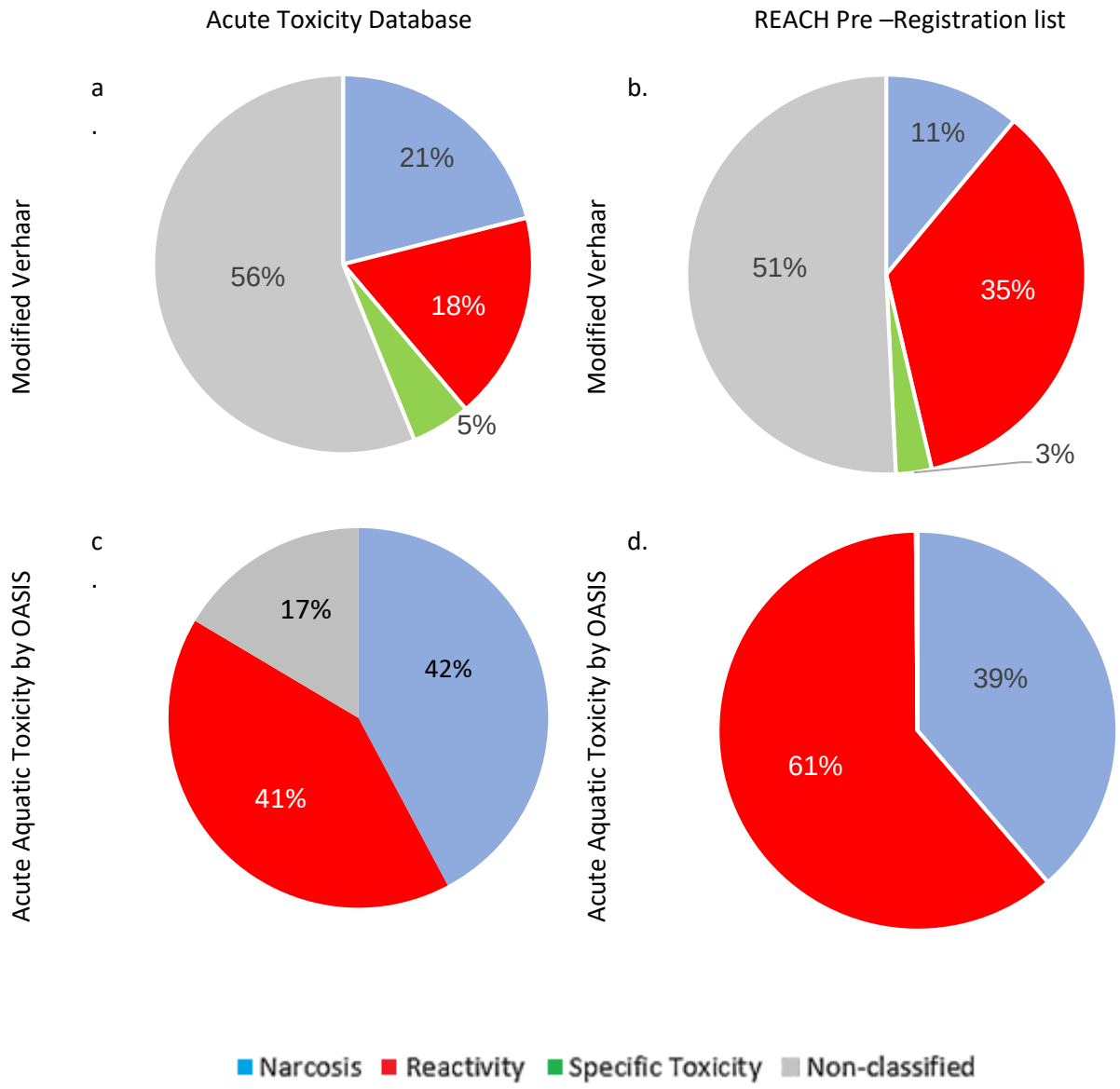


Figure 3.1. Mechanistic domain assignment distribution when modified Verhaar and Acute Aquatic Toxicity by OASIS were applied to the Acute Toxicity inventory (n=4591) (a,c), and REACH Pre-Registration inventory (n=74073) (b,d)

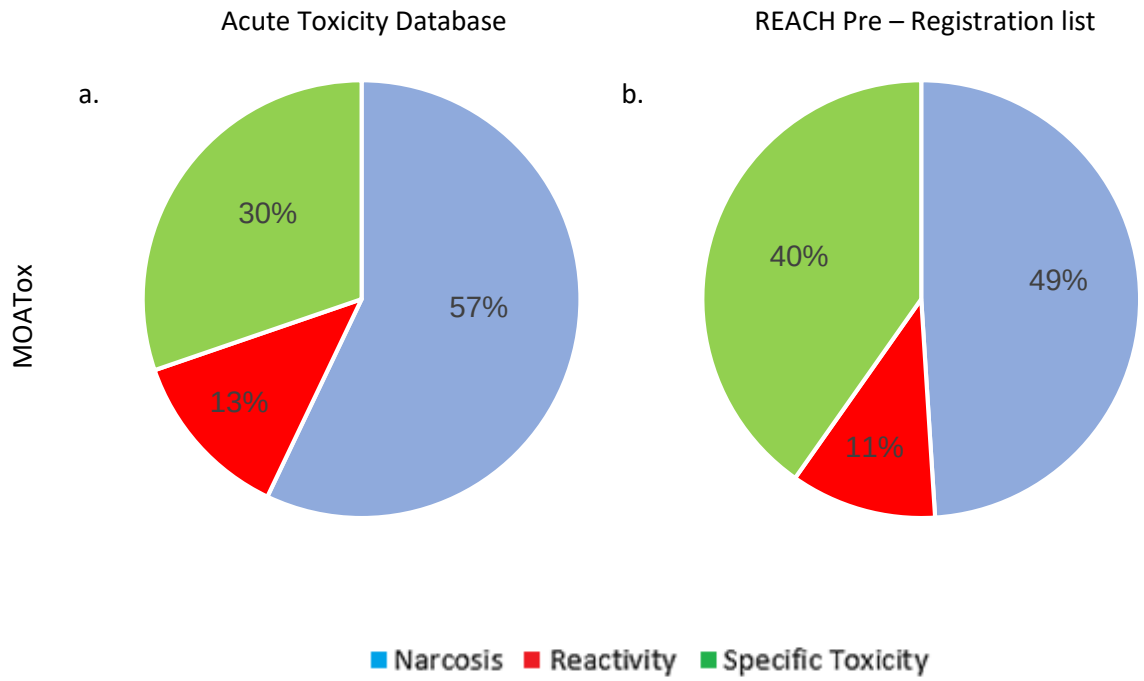


Figure 3.2. Mechanistic domain assignment distribution when Barron scheme is applied to subsets of Acute Toxicity (n=727) and REACH Pre-Registration (n=890) inventory.

MOATox is predominantly populated by specifically acting toxic compounds (56%), followed by compounds falling under the broad mode of action of Narcosis (36%) and reactivity (8%). Considering the imbalance in chemistry distribution amongst the mechanistic domains of the MOATox inventory, the difference in mechanistic domain assignment distribution between Figure 3.1 (Modified Verhaar and AAT OASIS) and Figure 3.2 (MOATox) in assigning compounds in Specific Toxicity domain is quite prominent; 30% for Acute Toxicity Database and 40% for Pre-Registration list are classified as specifically toxic by MOATox compared to Verhaar (3% for Acute Toxicity Database and 5% for Pre-registration list). That was not the case for AAT OASIS, as it does not assign compounds to specific modes of toxic action, but instead includes them under the 'Reactive unspecified' label. Similar observations were made by Kienzler *et al.* (2017) having access to the ASTER and Russom schemes. They highlighted the high proportion of unclassified compounds by the current classification schemes and the limitations in mode of action assignment raised by the Russom and Barron schemes.

The analysis presented highlights the limitations of the design of the applicability domain of the classes, especially those associated with specific toxic action in the case of Modified Verhaar and highlights the importance of further development and redefinition of the specific toxicity applicability domain to meet current industrial demands. Under the precautionary principle, Verhaar *et al.* (1992) have prudently proposed only a handful of chemical classes as definitive specifically acting acute toxicants (i.e. DDT and analogues, (dithio)carbamates, organotin compounds, pyrethroids, organophosphorothionate esters). QSARs and *in silico* methodologies describing specific toxic action in the literature focus predominantly on named chemical classes with well-defined taxonomic applicability (e.g. an interspecies relationship of toxicity for *Daphnia*-fish for benzotriazoles (Cassani *et al*., 2013)). A shift from a chemistry-based, to a target-based, definition of the specific toxicity domain has gained momentum. Examples of such methodologies have been proposed by Hornung et (2014) for low affinity oestrogenic compounds, by Nelms *et al.* (2015) for mitochondrial toxicity and most recently by Boone and di Toro (2019) for a target-focused screening tool for various acute toxicity mechanisms including neurotoxicity. Available mechanistic information could potentially be harvested to serve as a screening tool for acute specific toxic action relevant to molecular targets of interest, following the example of screening tools for reactivity related to MIEs (Enoch *et al*., 2011, Enoch and Cronin, 2010).

3.4.2. Agreement in mode of action assignment among schemes

It is important to have a level of concordance among schemes, as they are intended to facilitate chemical hazard identification as part of risk assessment and potentially contribute in decision making for regulatory bodies. To investigate the consistency of mode of action assignment, the MOATox list was employed as it is the only one with no possibility for *de novo* MOA classification of compounds but comprises a wide variety of chemistry and MOAs. To evaluate the predictability of all three schemes, an inventory with mechanistic information was employed, namely the FHM inventory. The MOATox inventory was profiled using modified Verhaar and AAT OASIS, and using the MOATox

suggested class and mechanistic domain assignment for reference. Positive predictability percentages (PPPs) per domain and scheme are presented in Figure 3.3. The main focus of this analysis was to monitor the consensus in domain assignment. As observed, the highest PPP values are observed for the Reactivity domain assignment between MOATox and modified Verhaar (53.3%) and AAT OASIS (65%), and less when it comes to Narcosis domain assignment (28.4% for modified Verhaar, 41.3% for AAT OASIS). It is worth mentioning at this point, that concordance among all three schemes for narcosis domain assignment is for 63 compounds (PPP=13.7%) and for reactivity domain assignment is for 39 compounds (PPP=36.4%). There is no specific toxicity domain assignment using OASIS. Concordance between Verhaar and MOATox is observed for 50.4% of these compounds. The outcomes presented are in accordance with results presented by Kienzler *et al.* (2017). Only about a third of the compounds assigned as being narcotics or reactives by MOATox, are similarly classified as such by Modified Verhaar and AAT OASIS. Those observations suggest limitations in using multiple schemes to support a weight of evidence approach for class and domain assignment.

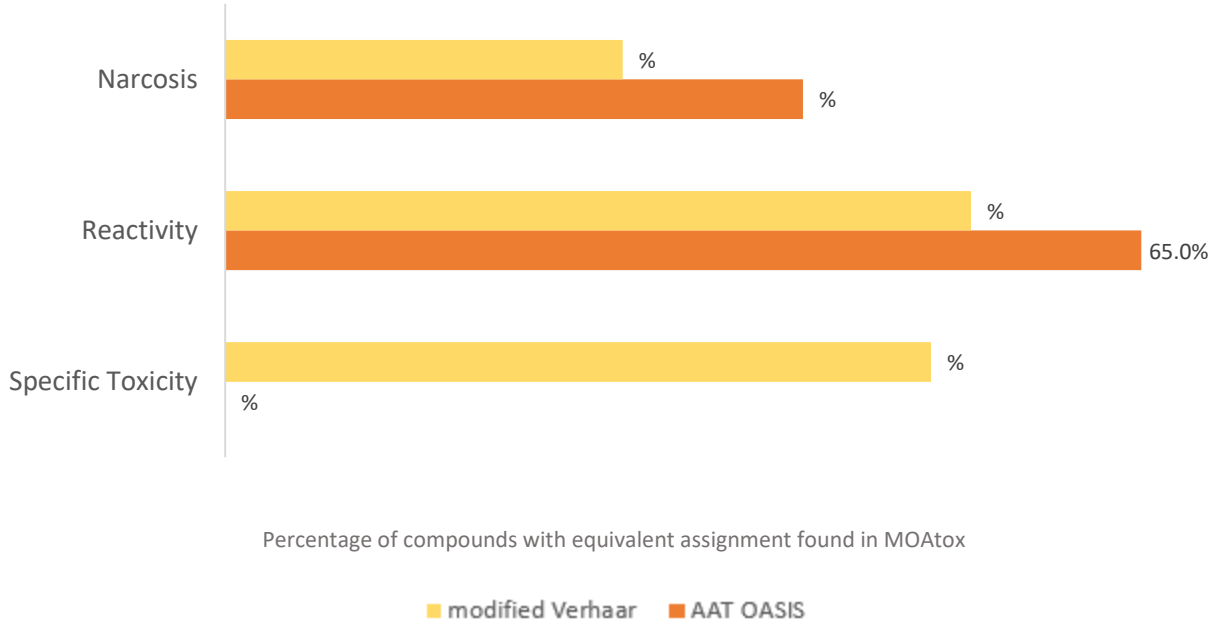


Figure 3.3. Compounds with concordance in mechanistic domain assignment by Modified Verhaar and Acute Aquatic Toxicity by OASIS profiler in relation to class distribution (Positive Predictability Percentages (in colour)) for MOATox dataset by Barron (Narcosis n=461; Reactivity n=110; Specific Toxicity n=612)

Similarly, when employing the FHM inventory, using the true class and mechanistic domain assignment for reference (found in (US Environmental Protection Agency, 2008)), PPPs per scheme and domain were calculated and are presented in Figure 3.4. This analysis aimed to assess the validity of the classifications produced by the AAT OASIS profiler, Modified Verhaar scheme and MOATox. As seen in Figure 3.4, the AAT OASIS profiler outperformed modified Verhaar on correctly assigning compounds in the mechanistic domains of narcosis and reactivity. This observation is partially expected however, as the AAT OASIS profiler is predominantly based on the work by Russom *et al.* (1997) and the ASTER package, both of which used as a training set the FHM database. In the case of specific toxicity though, the limitations of AAT OASIS are emphasised, as there is no applicability domain corresponding to specific toxic action.

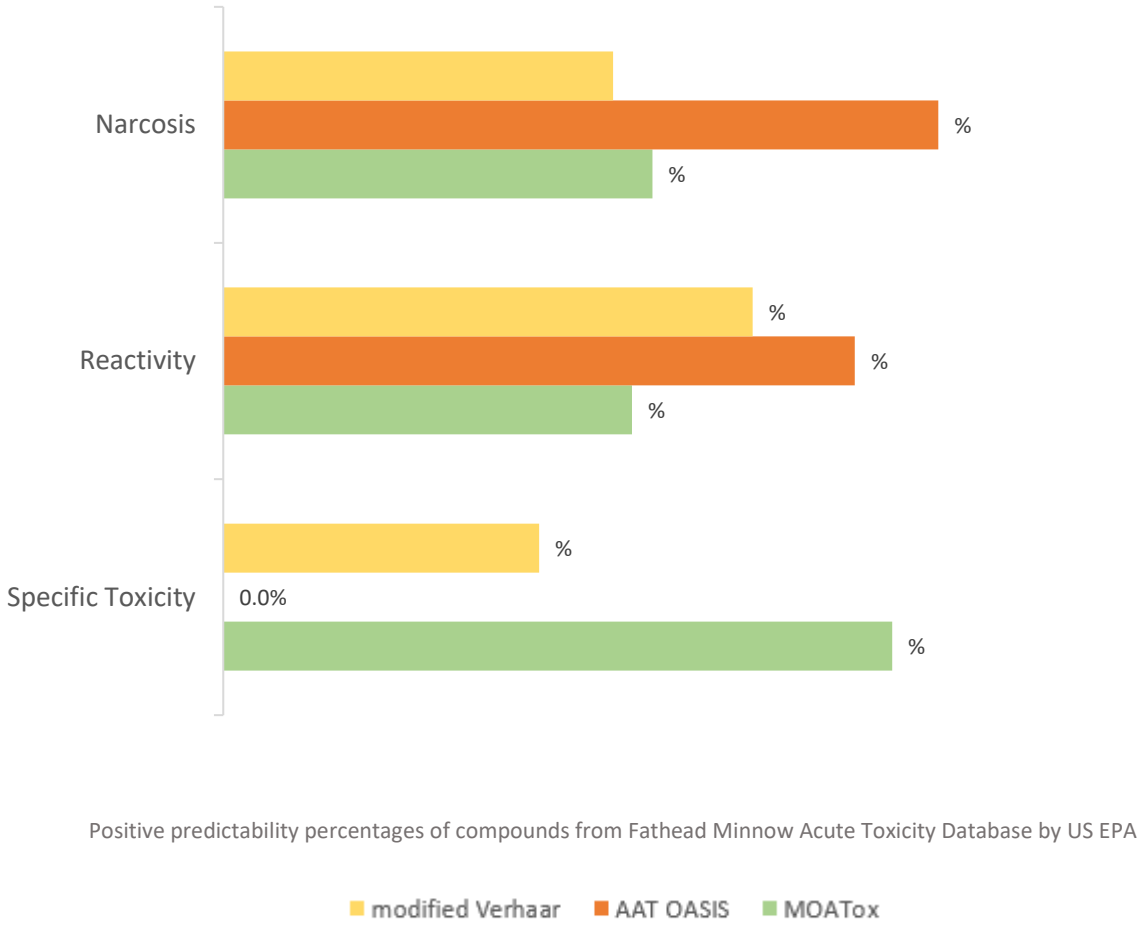


Figure. 3.4 Positive Predictability Percentages for Modified Verhaar (n=462), Acute Aquatic Toxicity by OASIS profiler (n=462) and MOATox dataset (n=241) when applied to the dataset extracted from the Fathead Minnow Acute Toxicity Database by US EPA.

Only 241 of the chemicals in the FHM DB have assignments from MOATox – therefore the assessment is only based on this subset. It is worth noting that of those 241, the expert panel domain assignment found in MOATox publication is in partial agreement with the one found in FHM dataset (PPP values ranging from 45.8 to 75%). For the narcosis and reactivity domain PPP values are 48.1% and 45.8% respectively whereas for the domain of specific toxicity, MOATox assigned correctly 75%. Lastly, the level of concordance among schemes was quantified using the subset of the FHM inventory where class assignments are available for all three schemes (n=241). The percentages of concordance are presented in Figure 3.5 and their comparison with the class distribution of the subset. It is worth noting, that for all mechanistic domains there is an average of 52.7% concordance on correct predictions. This observation should be taken into consideration as it indicates a level of uncertainty. This analysis highlights the importance of combining methods to support *de novo* class assignment of compounds that rely exclusively in one *in silico* classification scheme for acute toxic action; especially for Specific Toxicity MOAs where the level of concordance is less than 50% among the two schemes (i.e. modified Verhaar and MOATox list) that have relevant classes.

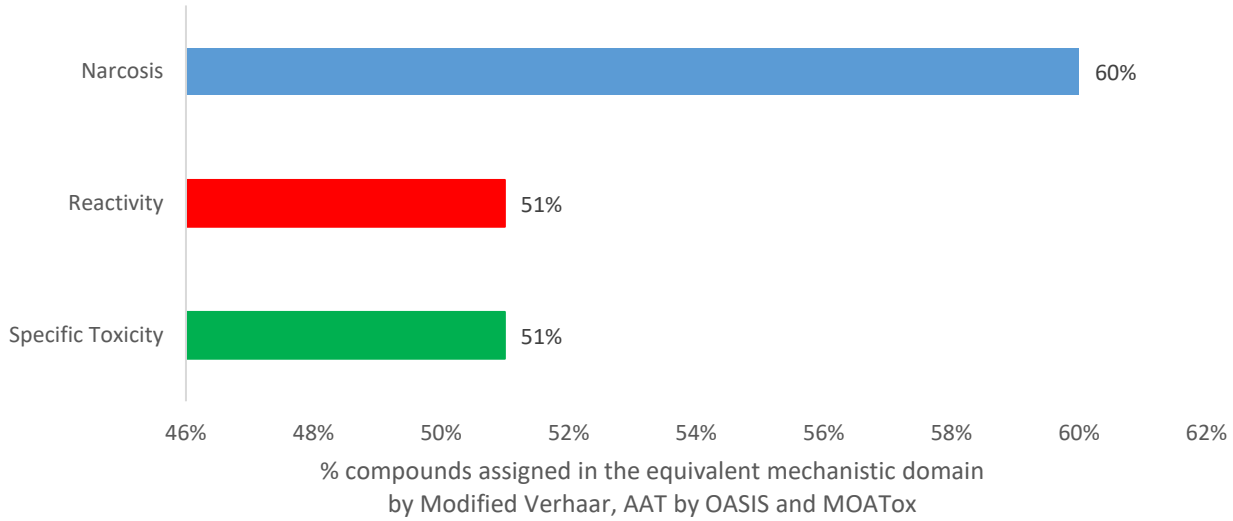


Figure 3.5. Mechanistic domain assignment concordance (expressed as a percentage) amongst Modified Verhaar, Acute Aquatic Toxicity by OASIS profile and MOATox dataset in relation to the subset by the Fathead Minnow Acute Toxicity database (n=241; Narcosis, n=160; Reactivity, n=45; Specific Toxicity, n=36)

3.5. Conclusions

Modified Verhaar (by Verhaar *et al.* (1992, 2000) based on guppy data, externally validated by fathead minnow data), and AAT OASIS (a coded version of Russom *et al.* (1997) based on fathead minnow data) are publicly available *in silico* classification profilers for acute aquatic toxicity relevant to industrial compounds with a strong mechanistic backbone. The use of modified Verhaar and/or AAT OASIS is recommended by regulatory bodies as supporting evidence for decision making. The fathead minnow database (US Environmental Protection Agency, 2008) is populated by high quality (i.e. 'Reliable with restrictions' in Klimisch criteria (Klimisch *et al*., 1997)) and has been used extensively in computational toxicology research as gold standard (e.g. use of fathead minnow acute toxicity data to evaluate the performance of a chemistry-based theoretical model for MOA assignment (Casalegno and Sello, 2013)); used as an external validation set for mechanistic assignment (Verhaar *et al*., 2000)) or for a multitude of QSARs (e.g. using fathead minnow high confidence data, new descriptors are proposed for acute toxicity QSARs (Jia *et al*., 2018)). To support model development Barron *et al.* (2015) proposed a reference dataset, populated by acute toxicity data for multiple fish and cladocera species along with a list of compounds with proposed 'broad' and 'specific' mode of action.

The aim of this chapter was to evaluate these publicly available resources using three different inventories, one supported with high confidence mechanistic information; one with publicly available acute toxicity information for a range of species and no supporting mechanistic information (i.e. the Acute toxicity database; 'Reliable without restrictions' and 'Reliable with restrictions' in terms of Klimisch criteria (Klimisch *et al*., 1997), more information on data collection see section 2.2. and Tables 2.2 and 2.4); and one of high industrial relevance with no mechanistic or otherwise supporting information, (i.e. compounds with the intention of preregistration in ECHA between 1 June and 1 December 2008). The main motivations behind this work was to observe strengths and limitations of current schemes for acute aquatic toxicity and Barron *et al.* (2015), and their level of agreement as they are commonly used to support decision making or data gap filling (e.g. by read-across).

Both schemes and Barron *et al.* (2015) describe general modes of aquatic toxic action in varied levels of detail. Therefore, to facilitate the comparative analysis, all proposed mode of action classes were grouped under these general effects (i.e. Narcosis, Reactivity, Specific Toxic Action) (details found on Table 3.3).

The modified Verhaar scheme has well-defined distinct applicability domains and, under the precautionary principle based on evidence rather than absence of it, assigns compounds only if the chemical of interest falls within the domains of the scheme. This provides confidence to class assignment whilst leading to a substantial percentage of unclassified compounds (Class 5). When modified Verhaar was applied to the acute toxicity database and Pre-registration list, more than 50% of compounds were not classified (see Figure 3.1 for more details). These observations highlight the modified Verhaar scheme's limitations in chemical coverage to meet current industrial needs. In contrast, AAT OASIS does not strictly follow the precautionary principle and, if the chemistry does not fall in any of the other classes, generally assigns compounds to the narcosis domain. Therefore, a narcosis assignment may have lower confidence unless there is other evidence to support it. Lastly, MOATox does not support *de novo* classification, therefore is not possible to assessed predictability. It is worth highlighting there is no transparency on the suggested class assignment, as the information provided was based on expert judgment with no further supporting evidence or explicit justification (Barron *et al*., 2015).

Definition of the specific toxicity domain is varied amongst the different schemes and MOATox. Only five chemical classes are definitively classified as specifically toxic under modified Verhaar (Verhaar *et al.,* 1992). AAT OASIS does not explicitly assign compounds as specifically toxic but rather includes the domain under the 'Reactive unspecified' label (OECD, 2013). MOATox comprises four broad and 18 specific modes of action linked to specific toxicity effects (Barron *et al*., 2015). The findings presented in this Chapter highlighted the limitations in definition of the specific toxicity domain amongst the schemes for *de novo* classification. Current approaches on defining specific toxicity effects focus on

compounds of interest and their respective chemical class (e.g. QSARs for the toxicity of PAHs to cladoceran (Ha *et al*., 2019)) or metabolites (e.g. the QSAR analysis of the toxicity of fluoxetine and its human metabolites to algae (Neuwoehner *et al*., 2009)). An increasing number of studies are employing molecular docking approaches to propose target based methodologies for screening for a target of interest (e.g. structural properties associated with binding to the androgen receptor (Wahl and Smieško, 2018)) and definition of structural features of toxicological relevance (e.g. enantioselectivity of imidazolinone herbicides for acetohydroxy acid synthase (Xie *et al*., 2018)). A more target-based approach would provide a modular and transparent means to populate the specific toxicity domain, and allow more informed decision making on taxonomy (e.g. inhibition of photosynthesis only relevant to photosynthetic organisms) or life stage level (e.g. effects of difenoconazole to the early life stages of zebrafish (Teng *et al*., 2018)).

Weight of evidence approaches under REACH aim to inform decisions and are based on the premise that reliable information from, optimally multiple, independent sources can support a case or decision. Computational methodologies are encouraged to support weight of evidence under REACH (ECHA, 2011). Therefore, it is crucial mode of action assignment by modified Verhaar and AAT OASIS are in agreement to provide confidence to a decision based on mode or mechanistic assessment or grouping. The analysis presented revealed that concordance amongst schemes and MOATox varied between the different domains, with comparable trends in the inventories with known and unknown mechanistic information. Most prominent differences were observed in the assignment of specific toxicity and greater consensus for the narcosis domain. To ensure good practice in using computational predictions for mode of action, considerations on each scheme's limitations and documented uncertainties should be acknowledged and addressed using further, reliable, evidence (Cronin *et al*., 2019).

Chapter 4: Compilation of Molecular Initiating Events Relating to Aquatic Toxicity from the Literature

4.1. Introduction

As described in Chapter 3, with regard to toxicology, the term "mode of action" (MOA) describes a functional or anatomical change, resulting from the exposure of a living organism to an exogenous substance (Rand, 1995). By comparison, a mechanism of toxic action describes changes at the molecular and cellular levels underlying a given mode of action (Escher *et al.*, 2011). Taking as an example the neurotoxic effects of endosulfan, the mode of action is considered to be the reduction of ionotropic GABA receptor chloride channel conductance that leads to a paroxysmal depolarising shift and epileptic seizure, whilst the mechanism of toxic action is binding at the picrotoxin site of the iGABAR chloride channel that leads to a reduction of ionotropic GABA receptor chloride channel conductance (AOP 10, https://aopwiki.org/aops/10). *In silico* classification schemes for acute aquatic toxicity have been designed primarily with a strong basis on chemical structure, often utilising chemical classes for structural alerts for their definition. In Chapter 3, *in silico* classification schemes to assign compounds into groups with common modes action were evaluated. The three schemes investigated were based on chemical structure (Verhaar *et al.*, 1992, Russom *et al.*, 1997, Barron *et al.*, 2015). The evaluation of the three schemes highlighted a number of limitations, notably in chemical coverage and the assignment of specific modes of toxic action.

4.1.1. Adverse Outcome Pathway

In recent years, a novel conceptual framework has been introduced to facilitate hazard characterisation, the adverse outcome pathway (AOP) framework (Ankley *et al.*, 2010). As stated in Chapter 1 (Section 1.3) an AOP is defined as '*a sequence of events from the exposure of an individual to a substance through to an understanding of the adverse effect in individual or population leve*l' and the concept is summarised in Figures 1.1 and 4.1 (Ankley *et al.*, 2010). An AOP is a construct connecting initial interactions at the molecular level, the so-called Molecular Initiating Event (MIE), of a substance with adverse effect(s) at the cellular, tissue, organ or organism and population levels. These linkages

are through single or multiple cellular and organ responses, referred to as Key Events (KEs). A crucial component of an AOP is the quantitative and qualitative Key Event Relationship (KER) between and/or among KEs that leads to the adverse outcome (AO), allowing predictions of toxicity across organisational levels. In practice, an AOP provides a structured approach of plausible connections that may provide information to support hazard identification by organising existing mechanistic knowledge (Ankley *et al.*, 2010). The distinction with mode of action is that AOPs aim to describe quantitatively and qualitatively the relationship between the mechanism of action all through the upper organisational biological levels (Figure 4.1) – as such an AOP is considered to incorporate the (overlapping) information from both the mechanism and mode of action.

The novelty of the AOP concept lies in the introduction of a consistent framework for hazard identification (and potentially hazard and risk assessment) both for environmental and human health effects, enabling risk assessors to integrate and cluster diverse information for decision making, rather than base their decision-making solely on chemistry and population effects (Burden *et al.*, 2015). Whilst there are examples of compounds with a linear relationship between chemistry and organism effect, e.g. cypermethrin that via modulation of sodium channels leads to acute mortality (AOP 96, https://aopwiki.org/aops/ (Davies *et al.*, 2007), this linearity is not realistic of toxicology and it is acknowledged that AOPs exist as interconnected networks, e.g. Villeneuve *et al.* (2014) identified twelve potential AOPs that could converge at a single key event, impaired swim bladder, which leads to reduced young of year survival in fish. From linearity to network, the knowledge from AOPs is very valuable e.g. by accurately defining the molecular event that initiates the adverse outcome(s) of interest (Allen *et al.*, 2014, Vinken, 2013) at the chemical and biological level, it could be possible to derive substantial and usable evidence to support decision making. As MIEs can be part of multiple AOPs, it is possible to construct independent AOPs for the same MIE, leading to an AOP network (Allen *et al.*, 2014). Allen *et al.* (2014) introduced the concept of multiple levels of MIEs, where the chemical-

molecular target is the initial MIE and the subsequent KEs might play the role of the MIE in another AOP branch (an example of this is shown in Figure 4.2).

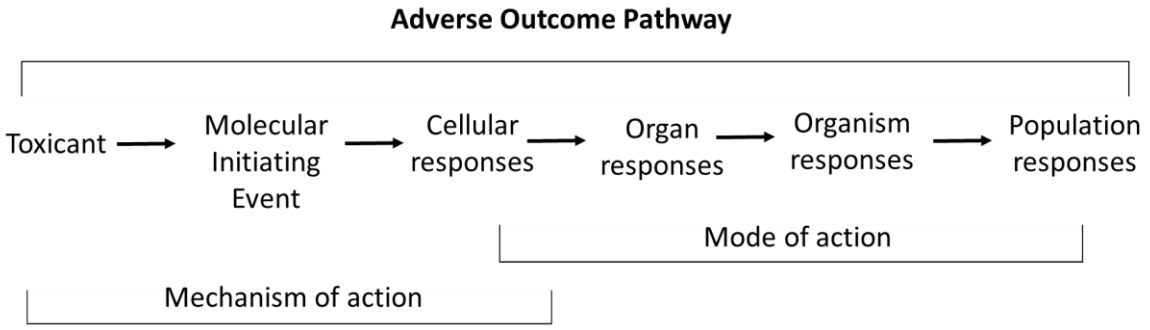


Figure 4.1. Schematic representation of the Adverse Outcome Pathway (AOP) framework in relation to the terms "Mode of action" and "Mechanism of action" (adapted by (Gutsell and Russell, 2013))

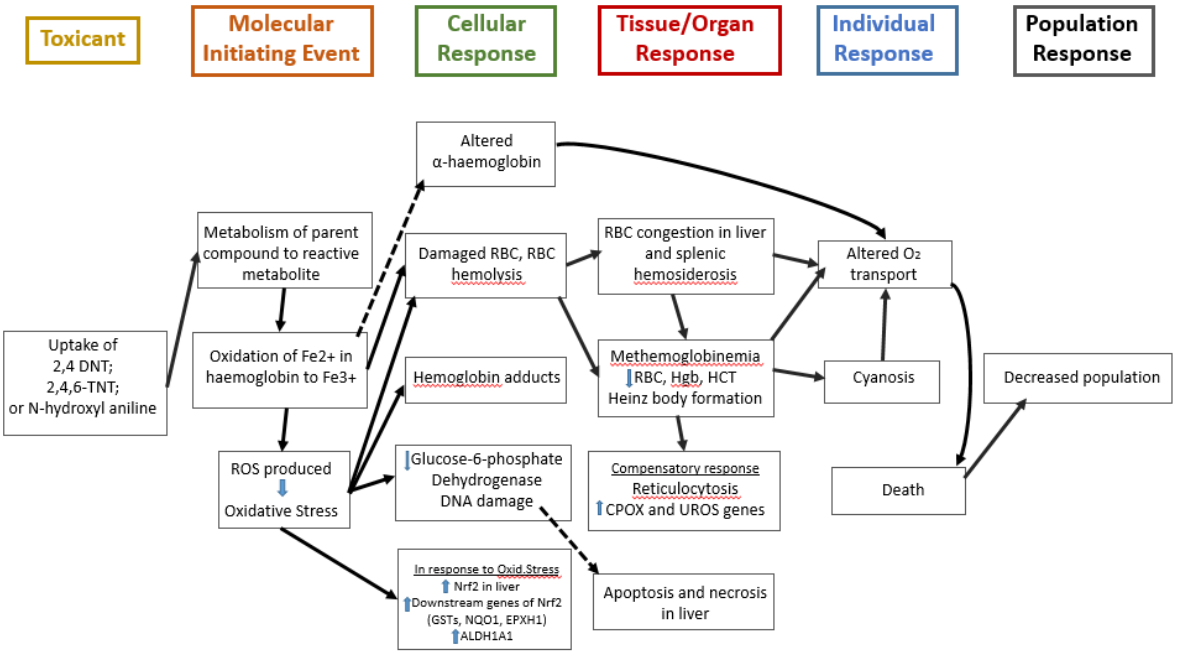


Figure 4.2. Example of an AOP network for decreased population after uptake of 2,4 DNT; 2,4,6-TNT or N-hydroxyll aniline, adopted from in AOP wiki (https://aopwiki.org/system/dragonfly/production/2016/11/29/55bHematotoxicity_due_to_nitroaromatics_and_N-hydroxyl_anilines.jpg), with multiple MIE levels, well described key event relationships leading to a single adverse outcome.

Mechanistic toxicology has been capturing initiating events for toxicity pathways such pyrethroid-mediated inhibition of voltage sodium channels (Davies *et al.*, 2007) and 2,3,7,8-Tetrachlorodibenzodioxin (TCDD)-mediated activation of the AhR pathway (Gooch *et al.*, 1989) for

many years, however, usage of the term “MIE” is much more recent. In order to identify and hence capture MIEs, it is essential to generate new, or identify existing, high quality data that provide the evidence of the ability of a chemical to elicit an interaction as well as the nature of the interaction. To this end, there have been a number of studies dedicated to capturing MIEs and AOP-related knowledge (Nelms *et al.*, 2015, Sanderson *et al.*, 2016, Antczak *et al.*, 2015). The development and assessment of KEs and KERs is described in detail by two guidance documents from the OECD (2013a, 2018). In the context of this thesis, MIE development and assessment can be based on these guidelines by considering an MIE to be a specialised KE. The development of an MIE needs to adhere the same criteria and standards as the KEs and further include a list of stressors and evidence for its biological occurrence. A weight of evidence approach (WoE) is the suggested method to ensure high confidence on data, with the latest guidance (OECD, 2018) implementing new research on weight of evidence (Becker *et al.*, 2015). There is an extensive literature on WoE for AOP practice. Becker *et al.* (2015) discussed the merits of a tailored WoE approach for data quality within the AOP concept. Becker *et al.* (2017) later acknowledged that despite caveats for its application within a MOA framework that describes a diverse range of effects, WoE provides an all-purpose platform to evaluate different aspects of a causal relationship (Becker *et al.*, 2017). Bridges *et al.* (2017) highlighted the need for customised WoE criteria to integrate assay/method-specific considerations depending on the data source. Additionally, Gross *et al.* (2017), using the example of endocrine disruption, discussed the potential of an effect-specific WoE evaluation. Another aspect to be considered for WoE analysis is the life cycle assessment component, highlighted by Gust *et al.* (2016), as there are an emerging number of AOPs for early life effects.

4.2. Aim, specific objectives and scope of Chapter 4

The objective of the thesis was to construct an *in silico* classification scheme to predict acute aquatic toxic action, within the AOP framework. The classification scheme is based around knowledge of MIEs to enable direct linkage to chemical structure. To assist in meeting the overall objective of the thesis, this chapter aimed to capture information on the literature relating to MIE knowledge which incorporated three factors to ensure as broad a range and coverage of MIEs as possible:

a) diverse chemistry,

b) the initiation of a wide range of molecular interactions , and

c) relevance to multiple aquatic vertebrate and invertebrate species.

Knowledge of the MIE is intended to act as the basis for the derivation of the classification scheme – in this Chapter this knowledge resource is referred to as the "training set". The training set of any classification scheme is of great importance, as it defines the quality of the predicted classifications generated. Therefore, the quality of data was ensured at three levels: namely at the biological, chemical and taxonomic applicability levels. MIE-derived knowledge was harvested from publicly available resources and a set of MIE-centred quality criteria were applied, covering well-established mechanistic knowledge and with evidence of biological and chemical causation, to ensure high quality data. The classification scheme was intended to capture mechanisms of toxic action relevant to aquatic fauna and flora, therefore MIEs with defined taxonomic applicability relevant to aquatic invertebrates, aquatic plants and microbiota were within the scope of the literature research.

4.3. Methodology

To build an efficient classification scheme, it is essential that high quality MIE literature is captured for the spectrum of MIEs that covers adequately the chemical space of industrial organic compounds.

4.3.1. Meek criteria for the evaluation of MIEs

The MIE literature review focused on capturing high quality biochemical information for MIEs for adult individuals regardless of sex. Prior to the collation of MIEs a framework was established to evaluate the "quality" of the MIEs in terms of the underlying evidence. To achieve this, the universal mode of action-centred Bradford Hill considerations, developed by Meek *et al.* (2014), were adapted to MIE-centred information and used as selection criteria.

The original Bradford Hill criteria assess nine aspects of an interaction in detail to determine causation to association (Hill, 1965). The criteria provide a means of assessing various aspects of the interaction from specificity of the target, through the sequence of events leading to an adverse outcome, to a quantitative relationship between the event and adverse outcome. A summary of the Bradford Hill criteria is provided in Table 4.1 with more details available in (Hill, 1965). The MIE-centred considerations (Table 4.1) cover seven of the Bradford Hill criteria directly (Table 4.1). The focus of this work was to retrieve well-established (Bradford Hill criterion: *Coherence*) MIEs (Bradford Hill criterion: *Specificity*), with the occurring MIE (Bradford Hill criterion: *Temporality*) having a direct causal link to an adverse outcome (Bradford Hill criterion: *Strength*), with data supported by at least two different sources of lab-, field- or alternative testing experimental evidence (Bradford Hill criterion: *Experiment*). The MIE should be plausible within a defined taxonomic applicability domain

(Bradford Hill criterion: *Plausibility*)), and from well-defined chemistry where analogy could be derived (Bradford Hill criterion: *Analogy*) (Hill, 1965). Repeatability of the causal links was not considered explicitly but rather addressed indirectly by gathering supporting evidence from at least two different sources. The criterion that was not taken into account was biological gradient (Table 4.1 (Hill, 1965)). Supporting evidence were derived, in part, from the fathead minnow acute toxicity database which provided a well-detailed resource of dose response data (US Environmental Protection Agency, 2008). The focus of the work presented was to map plausible biochemical interactions that could be linked to adverse outcomes at multiple taxonomical levels, hence utilising confidently available dose-response information for extrapolation from one species to taxonomical applicability of the MIE as whole wouldn't be appropriate for all MIEs presented.

The adapted Meek considerations were applied with the aim of ensuring that each consideration was met when retrieving MIEs. It is noted that adapted Meek considerations 1, 3, and 4 have multiple equivalent sub criteria, however only one sub criterion per consideration is required for the consideration to be fulfilled.

Table 4.1. Bradford-Hill criteria (Hill, 1965) and adapted Meek considerations for MIEs

Bradford Hill criteria	**Description of the Bradford Hill criteria**	**Adapted Meek considerations**	**Description of the adapted Meek considerations**
Coherence	No conflicting data available on the cause-and-effect interpretation	1. Biological Concordance	1.1 The MIE is well established. 1.2 No conflict with broader biological knowledge.
Specificity	Specificity of the association linking exposure with adverse effect		
Consistency	Repeatability of findings from multiple sources		
Strength	Strength of association by examining underlying mechanisms as found in literature (Fedak *et al.*, 2015)	2. Essentiality of MIE	There are data to support a direct causal link between the MIE and one or multiple AO(s)
Temporality	To establish causal relationship, exposure must precede effect		
Experiment	Availability of experimental data to support causal relationship	3. Concordance of empirical observations	Biological knowledge supports suspected causation based on 3.1 in *vivo* studies 3.2 alternative testing methods (i.e. *in vitro*, *ex vivo*, omics, high-throughput data, *in chemico*)
Plausibility	Evidence on whether the relationship is biologically plausible	4. Consistency among different biological contexts	The taxonomic domain of applicability is defined based on either: 4.1 Lab- and field- based studies 4.2 Alternative testing methods (i.e. *in vitro*, *ex vivo*, omics, high-throughput data, *in chemico*, *in silico*) 4.3 MIE target sequence similarity search
Analogy	*'In some circumstances it would be fair to judge by analogy'*	5. Analogy	The MIE has a defined chemical domain of applicability
Biological gradient	*'if the association is one which can reveal a biological gradient, or dose-response curve, then we should look most carefully for such evidence'*		

4.3.2. Biological information captured from the MIE literature

4.3.2.1. Capturing mechanistic information at the molecular initiating event level

There are a small number of resources which capture AOPs with sufficient detail for the MIE to be retrieved and utilised in this study. Key amongst these are the MIEs (implicitly) approved by the OECD Extended Advisory Group on Molecular Screening and Toxicogenomics (EAGMST) as found in AOP wiki (aopwiki.org). However, as the concept of AOPs is relatively new, there are very few publications or studies designed specifically to capture information on key events and/ or MIEs (e.g. (Feswick *et al.*, 2016)). It is acknowledged that the historical (pre-AOP) literature is much broader. It is also true that whilst the AOP concept was founded in ecotoxicology, the majority of research has been directed towards AOPs for mammalian effects. Thus, the literature review also focused on capturing initiating events relevant for both industrial organic compounds and plant protection products that were not described as MIEs *per se*. The plant protection products were included at this stage as they constitute a group of compounds with specific and very well-defined modes of toxic action well supported, in the main part, by high quality data. Several plant protection products are also used as pharmaceuticals (e.g. antifungals), however, due to data limitations at the time of the data retrieval, pharmaceuticals were excluded from the analysis.

A starting point for the collection of mechanistic information relevant to aquatic toxic action for industrial organic compounds was the book '*The Toxicology of Fishes*' (Di Guilo and Hinton, 2008). In addition, the comprehensive review of mechanisms for organic plant protection products published by Casida (2009) was utilised. More detail on each resource is given below:

- Di Guilo and Hinton (Di Guilo and Hinton, 2008) summarised a very wide range of mechanisms (e.g. receptor-mediated, reactive oxygen species generation, DNA covalent binding) and modes of action (e.g. carcinogenicity, renal toxicity, malformation) relevant to aquatic toxic action targeting, primarily,

fish species. For every mechanism captured from Di Guilo and Hinton (Di Guilo and Hinton, 2008), the source study was checked to verify the documented data and a literature search performed to capture biochemical information on further studies that may be relevant to the MIE.

- Casida (2009) grouped plant protection products (including pesticide and bactericidal compounds) according to target and reaction. The review summarised the proposed grouping of approximately 700 plant protection products based on general modes of action (e.g. non-specific reactivity, specific toxicity) and molecular targets. The plant protection products listed comprised organic compounds, inorganic compounds, metals, mixtures and microbiota. Targets were described at the biomolecular (e.g. nicotinic acetylcholine receptor), cellular, individual (e.g. juvenile growth regulation) or species level. Organic compounds with targets described at the biomolecular level were selected for further analysis.

-AOP wiki (https://aopwiki.org), an online resource, part of the OECD-sponsored AOP Knowledgebase (AOP-KB) effort, which aims to communicate and make publicly available information on AOPs in parallel with encouraging users to author and populate the AOP wiki. Detailed guidelines provide a framework for AOP wiki entries.

A more comprehensive literature search was undertaken using the key phrases 'aquatic toxicology', 'mechanistic toxicology' and 'acute toxicity', further refined for all combinations using the keywords 'fish', 'daphnia', 'algae' from the online resources PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Web of Science (http://webofknowledge.com). More than 8,000 entries were generated, and information was organised as being either associated with known MIEs or being new entries. Only studies which could be described as 'Reliable without restrictions', 'Reliable with restrictions', and 'Relevant without restrictions' according to the Klimisch *et al.* (1997) criteria were considered for data retrieval.

For every MIE entry, the following information was collected:

a) MIE target (biomolecular target)

b) MIE (biochemical interaction)

c) Xenobiotic – chemical name

d) Chemistry identifiers (i.e. Standard InChI Key and SMILES strings for individual compounds; SMARTS strings for functional groups)

e) Reference(s)

4.3.2.2. Capturing taxonomic domain of applicability

As the MIE-based classification scheme will be relevant for aquatic toxic action to all organisms it is crucial to capture molecular events for a range of aquatic species e.g. fish, invertebrates, algae. However, until recently, research in aquatic toxicology focused primarily on *in vivo* fish studies, with limited mechanistic data on other aquatic taxonomic classes (e.g. daphnids, algae). In order to gain an understanding of the diversity of species that could be addressed, the taxonomic applicability domain was captured based either on lab- and/or field- studies found via PubMed ((https://www.ncbi.nlm.nih.gov/pubmed/); Web of Science (http://webofknowledge.com); Pesticide Properties DataBase (PPDB) (University of Hertfordshire, 2018) and ECOTOX (US EPA, 2016).

In addition to the direct observation of effects from tests, evidence has also been gathered relating to similarity of the sequences of the respective MIE targets. For instance, as early as 2013 Lalone *et al.* (2013) proposed species extrapolation for MIEs based on sequence similarity for the biomolecules involved. For MIEs where no lab- and/or field evidence was available, a sequence similarity search was performed for each MIE target. The focus was on capturing the taxonomic order of the a) homology/orthology of the primary amino acid sequence of the MIE target, and b) the similarity of the conserved domains located within the functional unit. An MIE was considered species specific for cases where the homologue/orthologue and conserved domains of the MIE target were present in a

single species, and taxa order(s) specific where homologue/orthologue and conserved domains of the MIE target were present in a one or more taxonomic order(s). The main online tools used to perform this analysis were SeqAPASS (Sequence Alignment to Predict Across Species Susceptibility) (LaLone *et al.*, 2016), OrthoDB (Waterhouse *et al.*, 2013), Homologene (NCBI, 2018b) and Conserved Domains by NCBI (Marchler-Bauer *et al.*, 2015, NCBI, 2018a).

From over 90 MIEs and KEs, only those where taxonomic applicability for aquatic species was captured are presented in this chapter and were selected for further analysis in this and the following chapters.

4.3.3. Chemistry of the MIE literature

4.3.3.1. MIE derived chemical space

MIE-derived chemical space defines the applicability domain of all the chemicals with supporting MIE-related information that were selected for further analysis as described above (i.e. that adhered to the adapted Meek considerations and for which taxonomic applicability was sufficiently captured). Structural identifiers (i.e. SMILES strings, Standard InChI code) were retrieved from ChemSpider (http://www.chemspider.com/) and the logarithm of the octanol-water partition coefficient (log K_{OW}) values were calculated using KOWWIN from EPI Suite software (v 4.10) (US Environmental Protection Agency, 2016). To gain an insight into the chemical class distribution within the chemical space, ECOSAR was applied to the MIE chemical space data set and the assigned chemical classes were recorded (e.g. haloester, acrylamides) (US Environmental Protection Agency, 2017). As ECOSAR can potentially assign multiple classes to a single compound, all classes were considered when the chemical class distribution among MIEs was analysed.

4.3.3.2. MIE derived- versus industry relevant- chemical space

The chemical space covered by the MIEs in the literature was compared to the chemical space of the Acute Toxicity Database (constructed as described in Section 2.3.1) due to its industrial relevance.

The information relevant to the chemical space analysis extracted from the Acute Toxicity Database (referred to as the "Acute Toxicity Inventory" in this chapter) included chemical structures in SMILES notation, standard InChI keys and CAS registry numbers when available.

The Acute Toxicity Inventory was compared in terms of chemical space with that of the compounds for which MIEs from the literature were available using the CheS-Mapper software v2.6.6 (Gütlein *et al.*, 2012, Gütlein *et al.*, 2014). CheS-Mapper maps chemical space and renders all chemical structures entered simultaneously in 3D space based on a set of standard descriptors, facilitating visualisation. The positions do not necessarily reflect feature values; however, the 3D positions aid the user to identify patterns and regularities within the chemical space of interest. The compounds were displayed in relation to 141 features (i.e. log K_{OW}, as calculated by KOWWIN algorithm in EPISuite, molecular weight and CDK Functional Groups (n=139), as calculated by CheS-Mapper software); and embedded into 3D Space using the default Principal Component Analysis 3D Embedder (WEKA) algorithm. No clustering and cluster alignment algorithm methods were applied.

4.3.4. Clustering of MIEs in mechanistic domains

The MIEs are the basis for a novel *in silico* classification scheme for aquatic mechanisms of toxic action. The MIEs retrieved from literature and which adhered to the adapted Meek considerations were clustered into three mechanistic domains based on the type of interaction and MIE target:

a) Narcosis. These mechanisms of action are associated with events that lead to non-specific, non-reactive and reversible toxicity effects and are initiated by unspecific (often poorly defined) interactions with cellular membranes.

b) Non-specific reactivity. These mechanisms are elicited by compounds that are reactive through a variety of potential organic chemistry mechanisms of action based around electrophilic interaction, ROS (reactive oxygen species) generation and, to a lesser extent, nucleophilic interaction.

c) Specific toxicity. A specific mechanism of action was one with a defined target e.g. inhibition of a receptor site in a biomolecule leading to a specific AOP and thus a well-defined adverse outcome; targeting specific enzymes with a well-defined role in a biochemical cascade.

4.4. Results

The literature search for knowledge of MIEs relating to aquatic toxicity had as its starting point the bibliography of fish mechanistic toxicology (Di Guilo and Hinton, 2008). In addition, it considered over 8,000 published studies for identifying known, or potentially new, MIEs. The purpose of this literature review on MIE knowledge was to act as the basis for a novel classification scheme for aquatic toxic action, relevant to industrial organic compounds. Therefore, the focus of this chapter was to create an ''MIE training set'' that would utilise high quality data to encapsulate mechanisms of action relevant to aquatic toxic action for a range of aquatic species, and chemistry of industrial relevance. To this end, the literature review initially identified over 90 initial MIEs with direct biological causality with KEs and/or AOs manifested in multiple aquatic species. This MIE knowledge was supported by data for 814 compounds and covered 31 chemical classes (Appendix 2).

4.4.1. Data Quality

To ensure high quality in data collection, only data from studies was that adhered to GLP standards or provided well-described experimental parameters (‘Reliable without restrictions’ ‘Reliable with restrictions’ and ‘Relevant without restrictions’ by Klimisch *et al.* (1997)) were utilised. To ensure high quality criteria for defining MIEs, MIE-centred adapted Meek considerations were applied. 64 out of 90 MIEs adhered to all considerations. These 64 MIEs are the focus of the analysis in the current and

following chapter. Examples of the application of the adapted Meek considerations for seven MIE entries (six which met the considerations and one which did not are presented in Table 4.2.

4.4.2. Type of MIEs

The 64 MIEs which met the Meek criteria describe a wide variety of molecular interactions. The interactions were causally linked to narcotic effects, ROS generation, P450-mediated toxic action (e.g. AhR mediated toxicity), neurotoxicity, algal photosynthesis inhibition, biosynthesis inhibition (i.e. nucleic acid, fatty acid, protein), mitochondrial toxicity (e.g. respiration inhibition), disruption of cell cycle events (e.g. β-tubulin assembly in mitosis), toxicity effects caused by non-specific DNA and protein covalent binding (e.g. Michael addition, alkylation, SN2 reactions). The 64 MIEs were assigned to one of three mechanistic domains based on the type of MIE target, biological causality, and specificity of the interaction, the domains being: a) narcosis, b) non-specific reactivity, c) specific toxicity. A full listing of the MIEs with details of the compounds bringing about that mechanism, effect, detailed taxonomic applicability etc. is provided in Appendix 2.

Table 4.2. Examples of MIE-centred adapted Meek considerations for two MIEs found from the literature review.

a)

Adapted Meek considerations	**Disruption of membrane integrity**
1. Biological Concordance	Disruption of membranes has been long suggested as the underlining mechanism for narcotic effects (Roberts and Costello, 2003, Veith and Broderius, 1990, Van Wezel and Opperhuizen, 1995) with no conflicting studies suggesting otherwise (**criterion 1.2**).
2. Essentiality of MIE	Behavioural manifestation of narcosis has been described in detail by McKim *et al.* (1987) in rainbow trout exposed to compounds considered baseline toxicants, and how their effects compare to specifically acting and reactive chemistry (**criterion 2**).
3. Concordance of empirical observations	Lipophilicity is, in part, the description of the ability of a compound to dissolve in lipids and fats, which are key components of biological membranes. n-Octanol/water partition coefficient is an expression of lipophilicity and has been used as a key descriptor to efficiently predict median effective acute toxicity in *in silico* studies, within defined chemical domains, for several decades (Könemann, 1981). A number of *in vivo* (Aruoja *et al.*, 2014, Kluver *et al.*, 2016, Li *et al.*, 2015, Worgan *et al.*, 2003) (**criterion 3.1**) and *in silico* studies (Ellison *et al.*, 2008, van Leeuwen *et al.*, 1992, Verhaar *et al.*, 1992, Verhaar *et al.*, 2000) have demonstrated and built on this strong association in multiple biological contexts (**criterion 3.2**).
4. Consistency among different biological contexts	Data supporting a direct link between median effective acute toxicity and lipophilicity within well-defined chemical domain are consistent amongst a range of species (Aruoja *et al.*, 2011, Kluver *et al.*, 2016, Martin *et al.*, 2013, Verhaar *et al.*, 1992, Escher *et al.*, 2002, Nendza *et al.*, 2014, Wang *et al.*, 2016) (**criterion 4.1)**
5. Analogy	Examples 1-octanol, 2-propanol, 2-chloroaniline, 4-nitrotoluene, acetone, aniline, benzene, dichloromethane, diethylene glycol, diethylether, dimethylaminoterephthalate, ethanol (full list can be found in Appendix 2) (Ellison *et al.*, 2008, US Environmental Protection Agency, 2008) (**criterion 5**)
Status	**Included for further analysis as all considerations have been met**

b)

Adapted Meek considerations	**Bipyridylium-mediated redox cycling activity1, 2, 3.1, 3.2, 4.1, 5**
1. Biological Concordance	Paraquat and diquat have been extensively linked with the formation of reactive oxygen species (ROS) leading to redox cycling activity (Tsai, 2013, Di Guilo and Hinton, 2008) (**criterion 1.1**) with no conflicting data found in the literature (**criterion 1.2**).
2. Essentiality of MIE	Paraquat is commonly used in assays as positive control, it has been explicitly demonstrated that exposure to paraquat (Benina *et al.*, 2015) or diquat (Fussell *et al.*, 2011) leads to redox cycling activity (**criterion 2**).
3. Concordance of empirical observations	Both *in vivo* (Banaee *et al.*, 2019, Farmen *et al.*, 2010) (**criterion 3.1**); and *in vitro* studies (Esperanza *et al.*, 2015, Fussell *et al.*, 2011, Ruiz-Leal and George, 2004) (**criterion 3.2**) support the causal link between exposure to paraquat/diquat and redox cycling activity.
4. Consistency among different biological contexts	Toxic effects of paraquat and/or diquat exposure have been demonstrated in multiple species both *in vivo* (e.g. fish (Ayanda *et al.*, 2018, Ma *et al.*, 2018); (aquatic) plants (Benina *et al.*, 2015, Esperanza *et al.*, 2015, Peterson *et al.*, 1997)) (**criterion 4.1)**; and *in vitro* (e.g. (Fussell *et al.*, 2011, Peterson *et al.*, 1997, Sewalk *et al.*, 2000, Afarnegan *et al.*, 2017, Ruiz-Leal and George, 2004, Schaar *et al.*, 2015) (**criterion 4.2**).
5. Analogy	Diquat, Paraquat (Tsai, 2013) (**criterion 5**)
Status	**Included for further analysis as all considerations have been met**

c)

Adapted Meek considerations	AhR activation
1. Biological Concordance	The aryl hydrocarbon receptor (AhR) pathway is a well-established target for TCDD and dioxins, with extensive literature documenting that cytosolic AhR is activated due to covalent binding with TCDD and other dioxin-like compounds (**criterion 1.1**) (Denison and Vella, 1990, Leijs *et al.*, 2018, Lucier *et al.*, 1993, Song and Pollenz, 2002, Suh *et al.*, 2002, Peterson *et al.*, 1993, Rainey *et al.*, 2017), with no conflicting studies suggesting otherwise (**criterion 1.2**).
2. Essentiality of MIE	TCDD toxicity mediated by AhR interaction, amongst others, is linked directly with cardiomyopathy (Mohsenzadeh *et al.*, 2018) and DNA damage leading to apoptosis (Das *et al.*, 2017) (**criterion 2**).
3. Concordance of empirical observations	Data supporting AhR mediated toxicity are based on *in vivo* studies (example of developmental effects in dioxin-exposed zebrafish embryo (Bugel *et al.*, 2013)) (**criterion 3.1**), and alternative testing studies (example of study to investigate apoptotic effect of TCDD based on *in silico* and *in vivo* experiments (Das *et al.*, 2017)) (**criterion 3.2**).
4. Consistency among different biological contexts	AhR mediated toxicity is a conserved mechanism of toxic action among mammals and aquatic species (e.g. an extensive review on cross species differences (Peterson *et al.*, 1993) – **criterion 4.1**; e.g. *in vitro* study on CYP induction due to AhR activation in two fish species (Yuan *et al.*, 2013) – **criterion 4.2**). Several studies have attempted to identify the AhR binding site (e.g. (Pandini *et al.*, 2009)), research can be employed to establish cross-species differences (e.g. (Wang *et al.*, 2013)) (**criterion 4.3**)
5. Analogy	TCDD (Kuiper *et al.*, 2004); polychlorinated biphenyls (Leijs *et al.*, 2018); Nonylphenol (Mortensen and Arukwe, 2007) (**criterion 5**)
Status	**Included for further analysis as all considerations have been met**

c)

Adapted Meek considerations	Catalase binding
1. Biological Concordance	Numerous have suggested the irreversible binding of aminotriazole to catalase mediating toxic action in mammals (e.g. (Darr and Fridovich, 1986, Park et al., 2016) (**criterion 1.1**), with no conflicting data suggesting otherwise (**criterion 1.2**).
2. Essentiality of MIE	Aminotriazole toxicity mediated by catalase is directly linked, among others, with ROS accumulation in mammals (Lee *et al.*, 2018), superoxide accumulation, and subsequent plant death in *Arapidopsis thaliana* and other terrestrial plants (Gechev *et al.*, 2013) (**criterion 2**).
3. Concordance of empirical observations	Mammalian in vivo studies (Lee et al., 2016) (**criterion 3.1**) and in vitro studies (e.g. (Park et al., 2016, Villalobos-García and Hernández-Muñoz, 2017, Douiri et al., 2016) (**criterion 3.2**) supporting aminotriazole/catalase mediated toxicity
4. Consistency among different biological contexts	Only one study suggested aminotriazole mediated oxidative stress due to catalase inhibition in brain of goldfish after intraperitoneal injection (Bagnyukova et al., 2005). Only one in vivo study has suggested the potential of aminotriazole/catalase mediated toxicity in rainbow trout cell line (Dorval and Hontela, 2003). **Limited information on biological occurrence in aquatic species**
5. Analogy	Aminotriazole (Lee et al., 2018)
Status	**Not included in further analysis due to not meeting the criteria for consideration 4**

4.4.3. Taxonomic applicability

The taxonomic applicability of 58 out of 64 MIEs spans more than one taxon (commonly fish and crustacean). The remaining seven MIEs (#32-39, Appendix 2) are algal specific, covering MIEs with causal links to photosynthesis inhibition effects. The collective taxonomic applicability domains span 43 species and eight genera (Table 4.4) (a detailed description of taxonomic applicability per MIE entry can be found in Appendix 2). There were three main types of data or information that either provided sufficient evidence or were combined to investigate taxonomic applicability: a) online resources for toxicity data, b) data from *in vitro* studies, c) *in silico* evidence. The majority of online resources for toxicity data (e.g. Pesticide Properties DataBase (PPDB) (University of Hertfordshire, 2018); ECOTOX (US EPA, 2016)) provide information on the median effective concentration of acute toxicity for a battery of aquatic species (e.g. rainbow trout, daphnids, green algae) and only occasionally information on mode of action, species sensitivity and target site (e.g. alanycarb https://sitem.herts.ac.uk/aeru/ppdb/en/Reports/18.htm). Specifically acting compounds are used extensively in *in vitro* assays as inhibitors of biomolecular targets of interest (e.g. PCBs for inhibition of CYP1A enzymes), so for the corresponding aquatic species cell lines, it was safe to assume that the targets were present in the origin species.

LaLone *et al.* (2013) demonstrated it is possible to compare empirical toxicity data to cross-species predictions based on similarity of key MIEs, using the examples of 17α-ethinyl estradiol on the human oestrogen receptor; permethrin on the mosquito voltage-gated para-like sodium channel; and 17β-trenbolone on the bovine androgen receptor. Separately Lalone *et al.* (2013) demonstrated how *a priori* mechanistic knowledge of an event can facilitate cross-species extrapolation, using the example of the antagonistic effect of spironolactone to androgen receptors as a conserved mechanism of toxic

action in mosquitoes and fish species. Thus, *in silico* data were also employed as supporting evidence to investigate the taxonomic applicability of the MIEs presented (Table 4.3, Appendix 2). A number of online tools (see Methodology) enabled the capture of sequence similarities and orthologous/homologous structures among protein targets for multiple species.

4.4.4. Chemical domain of applicability

The review of literature relevant to MIEs captured a very wide range of chemicals. Figure 4.3 shows the graphical comparison of the chemical space covered by the Acute Toxicity Database inventory with the chemical space described by the MIE literature (i.e. MIE training set); with comparable distribution of compounds in relation to log K_{OW} and molecular weight. From the initial 814 compounds derived from 90 preliminary MIE entries, 709 corresponding to 64 MIEs were included for further analysis. The number of stressors varied per MIE entry, ranging from 1-100 compounds.

To extend the analysis, the chemical space of compounds in the MIE training set was compared with the chemical space of compounds of industrial relevance, namely those from the Acute Toxicity Database inventory (n= 4,660, compiled as described in Chapter 2). The aim was to examine the level of concordance of the two chemical spaces based on features relevant to aquatic toxicity and chemical structure, therefore, 149 features were considered: log K_{OW}, molecular weight and 147 functional groups as suggested by Chemistry Development Kit (CDK). The chemical space is displayed from the principal component scores from the descriptors in Figure 4.3, using CheS-Mapper. The plot of the chemical space shows a clear overlap between the two inventories; with all three mechanistic domains of the MIEs, i.e. narcosis - blue, non-specific reactivity - red, specific toxicity - green, overlapping with the acute toxicity database inventory shown in grey). Due to the difference in size, the inventories of the Acute Toxicity Database (n=5,085) and MIE specific toxicity training set (n=585) are more populated and therefore more visually prominent than the MIE training sets of non-specific reactivity

(n= 29) and narcosis (n=103) in Figure 4.3. Nevertheless, as seen in Figure 4.3, there is a distinct cluster of specifically acting toxicants with low overlap with a cluster of acute toxicity database inventory toxicants that includes pyridines and pyrimidines which are representative of the pesticides and so are not covered in the chemical space of industrial chemicals. This overlap could suggest that the MIE training set captures chemistry representative of compounds of industrial relevance.

Table 4.5. ECOSAR classification for the MIE-derived chemical space and distribution across mechanistic domains.

Chemical Classes as assigned by ECOSAR	Number of compounds (from 709 associated with an MIE)	Narcosis	Non-Specific Reactivity	Specific Toxicity
Acid moiety	34		x	X
Acrylamides	16			X
Aldehydes (Mono)	2			X
Aliphatic Amines	114	x	x	X
Amides	174	x	x	X
Anilines (Hindered)	4			X
Anilines (Unhindered)	32	x	x	X
Benzoylcyclohexanedione	4			X
Benzyl Alcohols	17		x	X
Benzyl Nitriles	33			X
Carbamate Esters	38			X
Carbamate Esters, Phenyl	4			X
Carbonyl Ureas	33		x	X
Epoxides, mono	3		x	X
Esters	258	x	x	X
Esters (phosphate)	52			X
Esters, Dithiophosphates	30			X
Esters, Monothiophosphates	45			X
Halo Acids	4			X
Halo Alcohols	2			X
Halo Ethers	6			X
Halo Ketones (2 free H)	2			X
Haloacetamides	28			X
Halopyrdines	33			X

Hydrazines	17			x	X
Imidazoles	28				x
Imides	16				x
Inorganic Compound	6				x
Ketone alcohols	4				x
Neonicotinoids	6				x
Nereistoxin Analogues	8				x
Neutral Organics	99		x	x	X
Nicotinoids	12				x
Oxetanes	2				x
Peroxy Acids	2				x
Phenol Amines	9		x	x	X
Phenols	23		x	x	X
Polynitroanilines	2				x
Polynitrobenzenes	20				x
Polynitrophenols	2				x
Propargyl Halide	2				x
Pyrazoles/Pyrroles	74				x
Pyrethroids	41				x
Quinones	2			x	
Salt	12				x
Substituted Ureas	34				x
Sulfonyl Ureas	60				x
Thiazolidinones	2				x
Thiocarbamates, Mono	30				x
Thiophenes	14				x
Thioureas	6				x
Triazines, Aromatic	51				x
Triazole Pyrimidine Sulfona	8				x
Triazoles (Non-Fused)	54				x
Vinyl/Allyl Alcohols	20				x
Vinyl/Allyl Esters	18				x
Vinyl/Allyl Ethers	31			x	X
Vinyl/Allyl Halides	48		x		x
Vinyl/Allyl Ketones	30				X
Vinyl/Allyl Nitriles	2				X
Vinyl/Allyl Sulfones	2				X

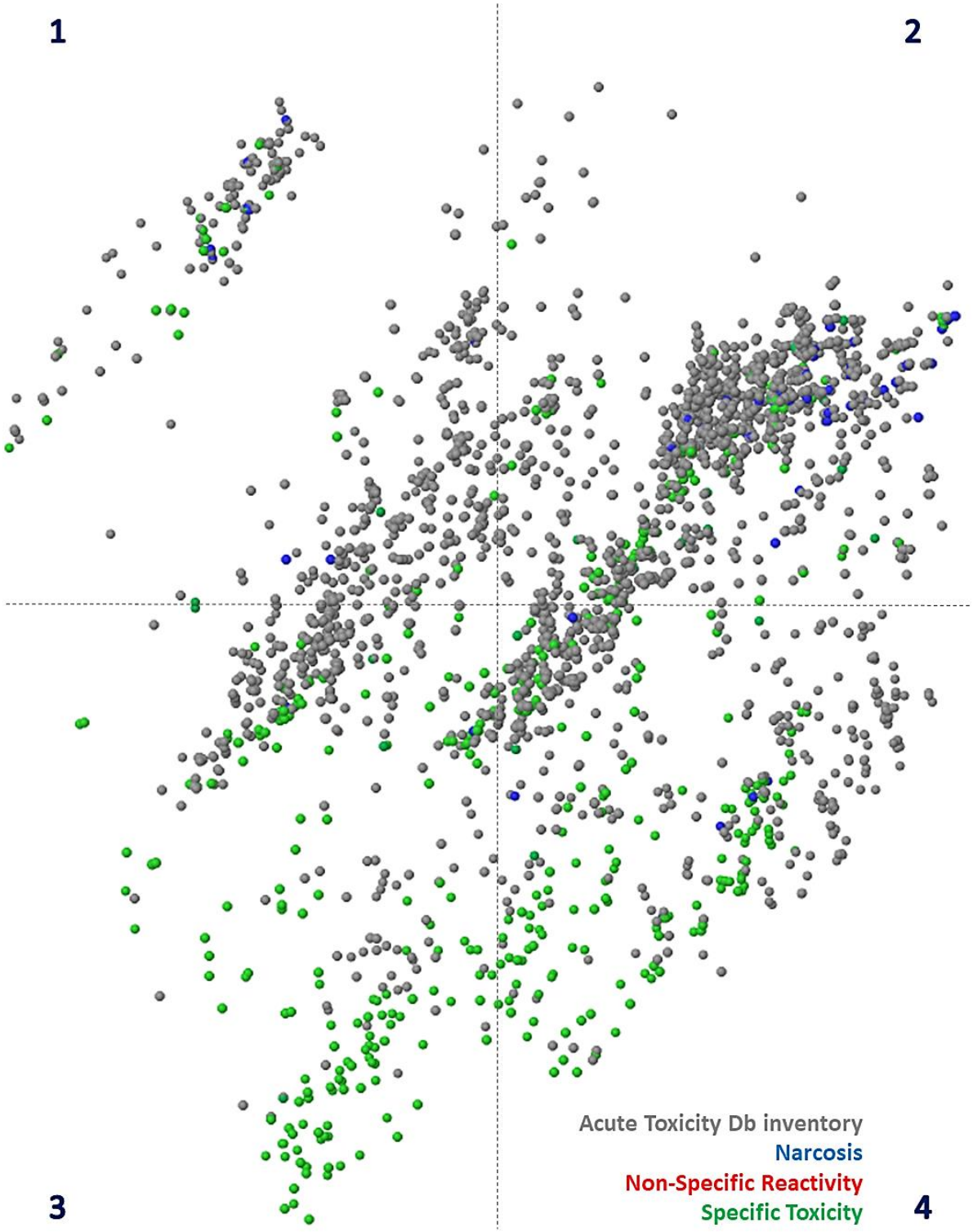


Figure 4.3. Schematic representation of chemical space captured by MIE literature (Narcosis– Blue, n=104; Non-specific reactivity – Red, n=14, restricted visibility on figure due to scale and amount of compounds; Specific toxicity – Green, n=545) versus the chemical space of the inventory of Acute Toxicity Database (Grey, n=5,085) (described in Chapter 2)

4.4.5. Mechanistic domains

The narcosis domain comprised two distinct MIEs. These MIEs were initiated by 103 compounds and 8 chemical classes based on ECOSAR classification (#1,2 in Appendix 2, chemical classes in Table 4.5). The non-specific reactivity domain comprised 11 MIEs initiated by 35 compounds, 14 chemical species (as described by SMARTS strings) and 14 chemical classes based on ECOSAR classifications (#3 – 13 in Table 4.3, chemical classes in Table 4.5). The specific toxicity domain was made up of 51 MIEs initiated by 585 compounds, 5 chemical species and 60 chemical classes based on ECOSAR classification (#15-65 in Table 4.3, chemical classes in Table 4.5). Generally, the MIEs within the narcosis and non-specific reactivity domain are prominent in the majority of aquatic species, whereas the specific toxicity MIEs display species specificity and defined taxonomic applicability.

4.5. Discussion

The work presented in this Chapter focused on compiling mechanistic knowledge from the literature that would be the basis for a novel *in silico* classification scheme for aquatic toxicity. The aim was to include high quality chemically and biologically rich data, creating a training set with a defined applicability domain, associated with a strong mechanistic background and adequate taxonomic coverage, that could be broadly applicable to industrial chemicals.

4.5.1. Literature search

Mechanistic toxicology is a well-established field of research, however, terms such as MIE and KEs have only been used since Ankley's *et al.* (2010) publication on AOPs and even since then rather sparsely. Thus, the literature search entailed primarily traditional mechanistic toxicology research that documented biochemical interactions of xenobiotics with targets at the molecular level. The four main sources of mechanistic data were:

a) Di Guilo and Hinton's (Di Guilo and Hinton, 2008) volume on the mechanistic aquatic toxicology of fish,

b) a comprehensive review on biocide targets (Casida, 2009),

c) online resources (i.e. PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Web of Science (http://webofknowledge.com)), and

d) the AOP wiki (https://aopwiki.org).

In this regard, MIEs were considered to be characterised events that abide by the MIE definition as proposed by Allen *et al.* (Allen *et al.*, 2014): *"A molecular initiating event (MIE) is the initial interaction between a molecule and a biomolecule or biosystem that can be causally linked to an outcome via a pathway."* This enabled a significant number of events to be captured.

4.5.2. Data quality

To ensure high quality information relating to MIEs, data from a variety of experimental methods were considered. OECD guidance on developing adverse outcome events (OECD, 2018) suggests a set of WoE consideration for capturing MIEs and KEs. In brief, the WoE approach is based on the Meek criteria that propose a systematic review of evidence for a causal biological relationship between a cause (e.g. stressor) and observed effect (e.g. MIE) (Hill, 1965). In this chapter, adapted Meek considerations (as described in the Methodology section) were tailored for MIE-relevant data in an all-purpose manner and ensure biological occurrence for defined chemical and taxonomic domains. This work is intended to be the basis for a novel classification approach on aquatic toxicity incorporating MIEs, therefore, the work presented in this Chapter aimed to be inclusive of different data sources (e.g. *in vitro*, omics, *in silico*) to ensure variety in taxonomical and chemical applicability domains; but transparent in documentation of the source to ensure confidence and data quality. The adapted Meek considerations cover biological causality and essentiality for the AO to occur,

concordance of empirical observations and defined taxonomic domain and applicability domain. To ensure inclusivity of empirical observations multiple considerations (see criteria 1,3,4 in Table 4.1) were described with multiple equivalent clauses to encapsulate different experimental methods. A detailed scoring system was considered; however, it would compromise the data inclusivity and heterogeneity, which would not serve the purpose of the scope of this Chapter therefore a simple scoring system based on a pass/fail criterion for each consideration was employed.

Of more than 90 preliminary MIE records based on information for 814 compounds, 64 adhered to the adapted Meek considerations. The main reasons for exclusion were limited evidence on MIE targets (e.g. clofentezine leads to mite growth inhibition with no further information on MIE level) and on taxonomic applicability (e.g. inhibition of UDP-N-acetylglucosamine 1-carboxyvinyltransferase leading to cell wall disruption in bacteria). The exclusion of MIE data did not compromise the purpose of the work presented in this Chapter i.e. to create a *'training set'* for a classification scheme for aquatic toxicity.

All MIEs included for further analysis adhered to the adapted Meek considerations tailored for MIEs, but not all MIEs adhered to all clauses per criterion especially for criteria 3 (concordance of empirical observations) and 4 (consistency among different biological contexts) (see Methodology). MIEs predominantly associated with the AhR pathway and leading to neurotoxic effects were very well described with supporting evidence from *in vivo* and alternative testing methods for taxonomic applicability, and for biological causation (e.g. AhR inhibition, Table 4.2). On the other hand, for highly genus-specific events and non-specific effects there was limited supporting evidence for taxonomic applicability and biological causality respectively. For example, for interference leading to disruption of carotenoid synthesis mediated by phytoene desaturase, taxonomic applicability findings were based solely on an MIE target sequence similarity search. MIE #1 has the largest chemical applicability

domain (n=100), with strong biological causality of chemistry with AO. However, there is little evidence on the MIE target, which is strongly hypothesised to be mediated by disruption of membrane integrity (Ankley *et al.*, 2010, Roberts and Costello, 2003).

4.5.2.1. Biological causality

For all 90 MIEs identified from the review of the four sources, there was evidence of their biological occurrence, either by *in vivo* or alternative testing evidence. Different types of studies were utilised to verify the events, ranging from *in vivo* and physiological responses (e.g. narcosis), *in vitro* and *in chemico* assays (e.g. reactivity), to omics for a battery of gene targets and biochemical interactions (e.g. mitochondrial toxicity).

The MIEs were causally linked with both acute toxic effects (e.g. narcotic effects, irritation, central nervous system seizures, algal growth inhibition) as well as chronic effects (e.g. carcinogenicity, mutagenicity, hepatotoxicity). Traditionally, acute and chronic aquatic toxic effects are studied separately. At the MIE level, biological occurrence of an event is dictated by exposure and toxicokinetics and different effects (either acute or chronic) share the same starting point. As this work is qualitative, it focused on aquatic toxic action and whether there can be an interaction between a stressor and an MIE target. The quantification of AOPs is a longer-term goal for their application, however it is beyond the scope of this work as the current study was not able to draw on information on many quantitative aspects including expression levels of the target, bioavailability and dose of the stressor and provides no information on toxicity levels, to name a few.

Examples where events are initiated but do not manifest AOs are expected and, with implementation of the quantitative component in an AOP network context, may be predictable. Such an example was demonstrated by Li *et al.* (2018) where overexpression of CYP enzymes in *Plutella xylostella* led to resistance in chlorantraniliprole mediated by ryanodine receptor toxicity.

4.5.2.2. Taxonomic applicability domain

Since 2012, research into AOPs has been focused primarily on adverse outcomes relevant to human health. The AOP Wiki is a dynamic platform with an increasing number of users contributing to build its knowledge base. At the time of undertaking the research, 60 MIE entries were found in the AOP Wiki, with only 3 out of 60 MIEs (i.e. alkylation, DNA; increase, ecdysone receptor agonism; AhR activation) being relevant to environmental effects, notably to fish and daphnid species as found under the "taxonomic applicability" section.

Taxonomic applicability domain refers to the species in which the MIE of interest has the potential (i.e. MIE target) to manifest. Adopting again an inclusive approach to define the taxonomic applicability domain for an MIE, it was considered adequate to demonstrate the presence of the MIE target in a species with evidence derived by lab- and field- based studies, alternative testing methods (i.e. *in vitro*, *in chemico*, *in silico*) or homology/orthology of target structures. Combining supporting evidence from a multitude of sources and testing methods was the basis for defining the taxonomic applicability domain of the MIEs leading to the capture of mechanistic information for the 43 individual species and more than 12 genera listed in Table 4.4. The capture of information in this manner enables the design of the applicability domain of a MIE at two levels, taxonomic and chemical. This has the advantage of capturing species-specific effects that current *in silico* applications do not. Current schemes for *de novo* classification (i.e. modified Verhaar; AAT OASIS) have been based on fish studies, limiting the taxonomic applicability of the prediction to fish and by extrapolation to other aquatic species such as invertebrates. This point is particularly crucial to specific and non-specific reactivity toxic action, as these are the type of interactions with the greatest species variability (see taxonomic applicability data among domains in Appendix 2).

The limitations of this approach stem from the risk of generalisation of species/genus-specific effects. Such generalisation is, therefore, a factor to be considered when constructing the novel *in silico* classification scheme and the formulation of the generated predictions. An example that highlights this observation is the toxic action of azoxystrobin. Azoxystrobin acts as a specific toxicant by disrupting the mitochondrial respiratory chain by reactive interaction with mitochondrial ubiquinol oxidase at the Qo site in fish and daphnid species (Gao *et al.*, 2014, Lazartigues *et al.*, 2011). It is also known to act in a non-specific reactive manner (i.e. ROS generation) in *Chlorella vulgaris* (Liu *et al.*, 2015), potentially due to lack of homology with the MIE target in the fish/ daphnid species (LaLone *et al.*, 2016). It should be also considered that, with regard to the presence of homologous/orthologous structures and conserved domains, there are differences in the role of biosynthetic cycles among species (e.g. FAD domain superfamily, respiratory chain, regulatory hormones). Taking the example of auxin, it is a plant growth regulator and targeted by auxin mimicking compounds for weed control. Even though algal auxin and its biosynthetic cycle is conserved and could be a potential MIE target, auxin's role in algal physiology is different (Zhang and van Duijn, 2014, Zhang *et al.*, 2016).

The analysis of the MIE literature aimed to capture mechanisms of toxic action for aquatic species. Therefore, MIEs exclusively of human and mammalian health relevance were not documented. It is also important to acknowledge that even though the recorded species and genera are predominantly aquatic species, it does not mean that the MIE described in this work cannot be of mammalian or human health relevance (e.g. MIE targets like voltage-gated channels, acetyl CoA carboxylase).

4.5.2.3. Chemical domain of applicability

The review of literature relevant to MIEs captured a very wide range of chemicals. A graphic representation of the chemical space covered by the Acute Toxicity Database inventory was compared with the chemical space included in the MIE training set in Figure 4.3; indicating that the two

inventories were comparable with regard to structural features, average molecular weight and log K_{ow}. Based on this observation, the MIE training set could be considered a representative sample, and potentially covers the majority of the chemical landscape, of the Acute Toxicity Database. Due to this overlap, it could be argued that the MIEs covered by the MIE training set could suffice as the chemical applicability domain to describe the mechanistic landscape of chemistry with industrial relevance. From the initial 814 compounds derived from 90 preliminary MIE entries, 709 corresponding to 64 MIEs were included for further analysis. The number of stressors varied per MIE entry, ranging from 1-100 compounds.

4.5.2 Mechanistic domains

The MIEs were clustered into three mechanistic domains, a) narcosis, b) non-specific reactivity, and c) specific toxicity. As demonstrated in Chapter 3, a minimalistic approach in mechanistic domain assignment could create a comprehensive mechanistic framework to facilitate the user to undertake further analysis compared to current *in silico* classification schemes (e.g. modified Verhaar). The premise of the MIE domain is defined under a biochemical perspective, where biology and chemistry are equally important for its definition. In the following chapter, each mechanistic domain will be discussed in further detail.

4.6. Conclusions

Traditionally, decision-making for chemical environmental risk assessment using mechanistic information would focus on structural descriptors and physico-chemical properties to derive conclusions on population effects (e.g. Verhaar *et al.* 1992 suggests a mode of action and QSAR for EC50 calculation based on chemical structure). AOPs have provided a framework where the biochemical impact on targets of interest could be accounted (Ankley *et al.*, 2010) and potentially quantified to support decision-making (Conolly *et al.*, 2017). This chapter focused on molecular

initiating events, the first interaction described by AOPs, taking place in molecular level. The work presented focused on retrieving current knowledge on molecular initiating events relevant to aquatic toxicity.

The MIE information has the potential to be the basis for a novel classification scheme, with all information collated from publicly available resources, finding a balance between data inclusion from multiple sources (e.g. *in vivo*, *in vitro*, *in silico* studies) while not compromising data quality and transparency. Available resources from the OECD (https://one.oecd.org/document/ENV/JM/MONO(2016)12/en/pdf) provide guidelines on describing an MIE, which is considered a specialised KE where stressors and MIE target need to be defined. Data assessment as proposed by the OECD focuses on key event relationships (KER) and follows the Bradford Hill principles of causation adapted by Meek *et al.* (2014). The aim of this chapter was to collate high confidence information at the MIE level, with a focus on chemistry and plausibility of biological occurrence. Assessment of the quality of the data for the MIE information was performed following a MIE-centred adaptation of the Meek criteria (Meek *et al.*, 2014) (see Table 4.1). The majority of the Bradford Hill principles (Hill, 1965) are followed (i.e. Coherence, Specificity, Strength, Temporality, Experiment, Plausibility, Analogy) but two (Consistency and Biological gradient) were not considered. For the majority of the MIE information, it is expected that more than two sources should support it. However, due to the nature of agrochemicals, there was limited publicly available information where the chemistry of interest was explicitly used for the study, therefore, it was not included as a cut-off criterion. For similar reasons and due to time limitations, dose-response relationships between the MIE and consequent events were not explored. For MIEs to be considered, they had to abide to all five Meek considerations with supporting studies of high quality (i.e. 'Reliable without restrictions', 'Reliable with restrictions' in Klimisch scoring (1997)). Some of the proposed

criteria had multiple clauses (e.g. Criterion 3, Table 4.1) of equal weight to acknowledge the source of supporting evidence in a transparent fashion.

Reliable documentation on MIE, MIE target, stressors and taxonomic applicability was available for 64 MIEs (detailed list found in Appendix 2). Taxonomic applicability of the majority of MIEs (58 out of 64) span multiple aquatic taxa (see Appendix 2 for full list), except for six MIEs that were taxa specific, targeting the photosynthetic machinery. The chemical applicability domain of the MIE information collated consisted of 709 compounds representing 61 chemical classes (as assigned by ECOSAR).

The current thesis aimed to explore the viability of using MIE information for class assignment and as such, this work deliberately focused on MIEs applicable for both sexes and adult individuals. As proof of concept, it can be potentially expanded to incorporate sex- or life stage-specific events, as there are currently AOPs for early life AO (e.g. early life lethality after AhR activation, https://aopwiki.org/aops/150). A multi-layer approach, incorporating information on taxa, sex and life stage could benefit chemistries that within different biological contexts act by a different mechanism of action. Taking paraquat as an example, it leads to inhibition of Photosystem I (Casida, 2009) and it is also typically used as an inducer of redox cycling activity in *in vitro* assays for non-photosynthetic species (Benina *et al.*, 2015). Another example highlighting the importance of factors to consider other than chemistry and taxonomy, but also life stage, is inhibition of the ecdysone receptor by fenozides (Appendix 2, MIE #64). The ecdysone receptor is a highly specific receptor, with documented orthologues and highly conserved domains across aquatic species (Song *et al.*, 2017). Predominantly found in arthropods and insects, the ecdysone receptor plays a pivotal role in reproduction and moulting (a time sensitive process) in insects (Lenaerts *et al.*, 2019). Nevertheless, stressors initiating inhibition of ecdysone receptor are suggested as inhibitors of an equivalent process in early life *Xenopus* sp. (LaLone *et al.*, 2016). Application of the modified Verhaar scheme classifies fenozides in

Class 5 (Enoch *et al.*, 2008, Verhaar *et al.*, 1992, Verhaar *et al.*, 2000), whereas AAT OASIS classify them as "Reactive unspecified" (OECD, 2013b). Even though the class assignment by AAT OASIS does raise awareness of the potential risks of fenozides, it does not provide a comprehensive picture by taking advantage of all publicly available knowledge. These examples highlight the potential benefits of utilising mechanistic information in a defined taxonomy and life stage to provide confidence and reliability on decision making in different biological contexts.

The MIE information collated in this Chapter covers adequately mechanisms, taxa and chemistry of interest to be further utilised as training set for a novel classification scheme. To do that the MIEs were grouped in three mechanistic domains (i.e. narcosis, non-specific reactivity, specific toxicity) based on the associated downstream effect (e.g. narcosis MIEs associated with baseline toxicity effects) and the type of interaction (e.g. specific toxicity MIEs associated with well-defined targets, leading to specific cascade of events).

Chapter 5: Development of structural alerts for mechanisms of aquatic toxic action based on mechanistic data

5.1. Introduction

Under the European Union's REACH legislation, alternative methods (e.g. *in vitro, in chemico, in silico*) have been promoted to fill data gaps for risk and hazard assessment of compounds with limited *in vivo* data (European Commission, 2006). Of these, there are a number of *in silico* methods to predict toxicity and fate of a chemical (Pavan and Worth, 2008, US Environmental Protection Agency, 2017). A primary concept in *in silico* hazard characterisation is the use of grouping approaches under the premise that similar chemicals tend to have similar chemical and biological activities and that effects may be read-across. As described by ECHA (2008), the formation of a chemical category is based on chemical similarity, with compounds' chemical and biological properties being similar in some respect or following a certain, predictable, pattern. In the same guidance, it is also highlighted that commonality in structural features and consistent trends in their physico-chemical properties could be associated with a common mechanism of action.

In terms of *in silico* techniques for grouping and read-across, a category can be defined by an *in silico* profiler, serving as a tool for class assignment; a prime example of such tool with numerous profilers for a variety of effects is the OECD QSAR Toolbox. The QSAR Toolbox provides a framework for category and group formation using a range of profilers on multiple endpoints (e.g. acute aquatic toxicity, skin sensitisation, bioaccumulation, mutagenicity) and mechanisms of action (e.g. DNA binding, protein binding, retinoic acid receptor binding) that provide mechanistic transparency of the categories formed. The OECD QSAR Toolbox allows for the filling of data gaps by read-across, trend analysis and (Q)SARs.

In silico profilers can generally be either chemical- or mechanistically- based. Structural fragments or structural alerts (SAs) based on 2D chemical sub-structural molecular features have been used widely as a straightforward approach to assign compounds to a particular mode of action (MOA). The creation of a chemically-based profiler relies on data that associate one or more functional group(s) to a specific endpoint. Extensive literature has documented both chemically- and mechanistically-based profilers. For instance, Lepailleur *et al.* (2013) reviewed the evolution of data mining algorithms that aimed to capture chemical fragments from datasets with mechanistic information. They concluded that a set of chemical structural features, rather than a single chemical fragment, is more appropriate to describe chemistry of toxicological relevance (Lepailleur *et al.*, 2013). Nelms *et al.* (2013) and Rodriguez-Sanchez *et al.* (2013) employed a combined approach to verify the toxicological relevance of epoxides, lactones, nitroso, nitros, aldehydes and ketones and cyclic compounds respectively using established profilers from the OECD QSAR Toolbox and *Tetrahymena pyriformis* growth inhibition data.

Development of a mechanistically-based profiler relies on mechanistic data and information from a range of sources (e.g. *in vivo, in vitro, in chemico, in silico*) to define the chemical space of compounds associated with that mechanism of action. Examples of mechanistically-based profilers, to name a few, include a set of SAs developed by Claesson and Minidis (2018) to predict pharmaceutical drugs with reactive metabolites, these SAs were derived from drug structures known to produce reactive metabolites; the *in silico* mitochondrial toxicity profiler developed by Nelms *et al.* (2015) based on drug structures with known mitochondrial toxicity; and a screening workflow for nuclear receptor ligands associated with hepatic steatosis using information from, and knowledge of, the molecular initiating event (Mellor *et al*., 2016).

In order to develop and utilise *in silico* profilers for grouping and read-across, it is prudent to anchor them firmly to the mechanistic basis. The recent advances in AOPs provide an opportunity to do this.

Specifically, as discussed in Chapter 4, the Molecular Initiating Event (MIE) encapsulates the first biochemical interaction at the molecular level initiating the cascades of events leading to an adverse outcome. As such, description of the chemistry associated with the MIE is a good means of defining a grouping and has been the basis of several of the existing *in silico* profilers, as noted above. In Chapter 4, high quality MIE knowledge for acute aquatic toxicity was compiled from the literature and grouped in three mechanistic domains: a) narcosis b) non-specific reactivity, and c) specific toxicity. These three groups harmonise (to some extent at least) the previous approaches of Verhaar, Russom and Barron (see Chapter 3) and provide a basis for grouping compounds.

5.2. Aim of Chapter 5

The aim of this chapter was to annotate the chemistry derived from MIEs for known narcotic, non-specific reactive and specific mechanistic domains such that *in silico* profilers for mechanistic domains could be developed. In order to achieve this, training sets associated with MIEs and mechanistic domains, as extensively presented and discussed in Chapter 4, were utilised to design structural alerts. The structural alerts were organised in an analytics platform (i.e. KNIME) to construct the *in silico* profiler(s).

5.3. Methodology

5.3.1. Data set

MIE knowledge with corresponding chemistry for the range of aquatic species or genera described in detail in Appendix 2 was utilised to perform the analysis within this chapter. All chemical structures were encoded as Simplified Molecular Input Line Entry System (SMILES) strings and all functional groups as Smiles Arbitrary Target Specification (SMARTS) strings (full list in Appendix 2). Information on the MIE, MIE target and MIE domain (in SMILES and/or SMARTS strings) as well as the suggested assignment into the three mechanistic domains was compiled. As discussed in Chapter 4, the MIEs and MIE targets (summarised in Table 5.1) were clustered into three mechanistic domains as follows: narcosis, non-specific reactivity and specific toxicity. The mechanistic information captured covers 24 vertebrate (e.g. fish and amphibians), 15 invertebrate (e.g. crustaceans, protozoa) and 2 algal species (as detailed in Appendix 2).

Table 5.1. Description of the data set of MIE-derived knowledge, clustered in three mechanistic domains: narcosis (#1,2), non-specific reactivity (#3-13), specific toxicity (#15-65).

MIE ID	MIE Target	Molecular Initiating Event (MIE)	MIE domain	
			no. chemicals (in SMILES)	No. chemical groups (in SMARTS)
1	Endoplasmic reticulum calcium ATPase	Inhibition - Calcium mobilisation	5	
2	Cell membrane	Disruption of membrane integrity	100	
Non-Specific Reactivity				
3	DNA	Alkylation	10	5
4	DNA	DNA adduct formation	5	2
5	DNA, protein, lipids (liver)	Electron reduced metabolite mediated ROS formation	7	1

6	Biological nucleophiles	Michael addition		6
7	DNA, protein, lipids	ROS reactivity mediated by episulfonium ion	2	1
8	DNA, protein, lipids	ROS reactivity mediated by nitrenium	5	2
9	DNA, protein, lipids	Quinone-mediated redox cycle activity	2	1
10	DNA, protein, lipids	Procyanin-mediated redox cycle activity	1	
11	DNA, protein, lipids	Fenton-chemistry-like generation of hydroxyl radical	1	
12	DNA, protein, lipids (liver)	ROS reactivity mediated by superoxide anion and peroxide	1	
13	DNA, proteins, lipids (chloroplast; lung)	Bipyridyl-mediated redox cycle activity	2	1
Specific Toxicity				
15	Aryl hydrocarbon receptor (AhR)	Covalent binding leading to activation	8	5
16	Calmodulin receptor	Covalent binding leading to inhibition	1	
17	Androgen Receptor	Binding – Agonistic effect	3	
18	Aromatase (CYP19s)	Interference with electron transfer via the cytochrome P450 haem group of the aromatase enzyme	3	
19	Cyclooxygenase	Covalent interference leading to inhibition	11	
20	CYP enzymes	Binding leading to decreased activity	5	1
21	CYP1A	Binding leading to inhibition	3	
22	Acetylcholinesterase	Reversible binding leading to inhibition	26	3
23	Acetylcholinesterase	Irreversible binding leading to inhibition	65	
24	Nicotinic acetylcholine receptor	Competitive agonist	10	
25	Nicotinic acetylcholine receptor	Allosteric agonist	2	
26	Nicotinic acetylcholine receptor	Antagonist leading to inhibitory neurotoxicity	4	1
27	Voltage gated sodium channel	Modulator leading to deactivation	42	2
28	Voltage gated sodium channel	Binding during the slow inactivated state of the channel leading to deactivation	2	
29	GABA-gated chloride channel	Non-competitive antagonism leading to channel deactivation	4	
30	Glutamate-gated chloride channel	Activation	6	
31	Ryanodine and ryanodine-like receptors	Binding leading to cytosolic Ca2+ transients	3	

32	Photosystem II, triazine site	Binding leading to inhibition	24
33	Photosystem II, nitrile site	Binding leading to inhibition	6
34	Photosystem II, urea site	Binding leading to inhibition	20
35	Photosystem I	Redox interference leading to electron diversion	2
36	Protoporphyrinogen IX oxidase	Inhibiting interference leading to decrease in activity	26
37	Phytoene desaturase	Interference leading to disruption of carotenoid synthesis	7
38	Lycopene cyclase	Interference leading to disruption of carotenoid synthesis	1
39	4-Hydroxyphenyl pyruvate di-oxygenase	Binding leading to disruption of carotenoid synthesis	8
40	EPSP (enolpyruvylshikimate 3-phosphate) synthase	Binding leading to amino acid biosynthesis disruption	2
41	Acetolactate synthase ALS (acetohydroxy-acid synthase AHAS)	Binding leading to amino acid biosynthesis inhibition	50
42	Glutamine synthase	Inhibitory interaction leading to inhibition of amino acid synthesis	2
43	Microtubule	Interaction with multiple sites leading to disruption of microtubule assembly and cell division	14
44	Microtubule	Interaction with multiple sites leading to disruption microtubule assembly and mitosis	3
45	DHP synthase	Inhibition of enzyme and blocking the conversion of 4-aminobenzoic acid to 7,8 - dihydropteroate	1
46	Acetyl-CoA carboxylase	Binding leading to inhibition of fatty acid synthesis	20
47	Fatty acid synthesis	Interaction with components of the cycle (not acetyl-CoA carboxylase) leading to inhibition	19
48	VLCFA (very long chain fatty acid) synthesis	Interaction with components of the process leading inhibition and disruption of cell division	21
49	Cellulose synthase	Interaction leading to cell wall (cellulose) synthesis inhibition	5
50	14a-demethylase (CYP51)	Binding indicative of coordination between the triazole N-4 and the haem Fe3+ leading to inhibition of the 14a-demethylation during ergosterol biosynthesis	36
51	Squalene epoxidase	Hydrogen bonding leading to inhibition of ergosterol and cholesterol biosynthetic pathways (terbinafine site: hydroxyl group of Tyr90)	3
52	RNA polymerase I	Interaction leading to disruption of nucleic acid biosynthesis	7

53	DNA topoisomerase type II	Disruptive interaction leading to nucleic acid biosynthesis inhibition	1
54	Protein biosynthesis	Disruptive interaction leading to inhibition of protein synthesis	4
55	β-Tubulin assembly in mitosis	Disruptive interaction leading to mitosis and cell division inhibition	9
56	Sterol 24-C-methyltransferase	Binding leading to inhibition of phosphatidyl choline synthesis	4
57	NADH oxireductase	Binding to membrane subunit of the enzyme (coupling site I) leading to inhibition of the mitochondrial respiratory chain	3
58	Complex I electron transport	Interference with electron transport chain leading to inhibition of mitochondrial respiratory chain	5
59	Succinic dehydrogenase (mitochondria)	Interaction with the complex II of the enzyme leading to inhibition of mitochondrial respiratory chain	22
60	Ubiquinol oxidase at Qo site	Reactive interaction targeting complex III site of mitochondrial respiratory chain leading to inhibition	20
61	complex IV, proton gradient (mitochondria)	Uncoupling via disruption of proton gradient within complex IV leading to inhibition of the oxidative phosphorylation	7
62	Respiration, ATP synthase	Inhibition	9
63	Juvenile hormone	Mimicking effect leading to increased male neonate production	5
64	Ecdysone receptor	Agonistic effect leading to premature moulting	4
65	Chitin biosynthesis	Disruption of chitin biosynthesis leading to premature moulting	13

5.3.2. Development of *in silico* profilers for mechanistic domain assignment

In silico profilers were developed for each mechanistic domain described in Table 5.1, as presented below. The narcosis profiler was constructed under the premise that if compounds are not within the non-specific reactivity or specific toxicity domains then they are within the narcosis domain. Therefore, the domains of specific toxicity and non-specific reactivity were defined first. The profilers were organised in a KNIME workflow to optimise the MIE assignment and minimise misclassification in the narcosis domain.

5.3.2.1. Specific Toxicity

The Specific Aquatic Toxicity domain (entries #15-65 in Table 5.1) was defined using structural alerts assigned to MIEs and MIE targets utilising the following five step process:

1) Visualisation: The training set was displayed a) in Marvin View (v.17.28.0), that comes as part of a chemical editor by ChemAxon (https://chemaxon.com/products/marvin), and/or b) in relation to a pre-defined set of structural features, as found in, the ChemoTyper (v. 1.1), a freely available chemoinformatics tool (https://chemotyper.org/).
2) Development: Structural fragments within the training set based on chemical similarity (i.e. identified as a set of chemical fingerprints/structural features as compared with the pre-defined set using Chemotyper (v.1.1.) for n>8 and/or empirically for n=<8) were captured. The protocol followed for SA development using the Chemotyper is thoroughly described by Nelms (2014). In summary, a set of chemotypes were identified, and redundant chemotypes (e.g. C–C) were removed. Based on these findings a set of structural alerts (SAs) was encoded in SMARTS strings for each MIE using Marvin Sketch by ChemAxon (https://www.ncbi.nlm.nih.gov/guide/chemicals-bioassays/). For MIEs with chemical groups in their training set (e.g. pyrethroids), corresponding SMARTS were created directly using Marvin Sketch by ChemAxon.
3) Refinement: The SAs were further refined by MIE information from publicly available resources on structural features related to activity (depending on data availability). A targeted literature review using keywords from the MIE of interest was performed in the free search engine PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Web of Science (https://clarivate.com/products/web-of-science/) to capture information on structural features related to activity. Furthermore, the chemical applicability domain was enriched (when data were available) using the 'PubChem BioAssay' resource (https://www.ncbi.nlm.nih.gov/pcassay/).

PubChem Bioassay contains bioactivity screens of chemical substances providing a searchable description of each bioassay. When available, MIEs were matched with bioassays found in PubChem Bioassay and the associated chemistry was added in the MIE domain and compared and contributed in further refinement of developed SAs.

4) Evaluation: A set of SAs corresponding to a single MIE was compared to the set of chemicals derived by all MIEs (n=694) to evaluate coverage and specificity of SAs for the MIE of interest. The main objective of this work was to map the chemical space of the MIE entries with an adequate level of specificity; therefore the evaluation of the alerts was performed as a set for each MIE (# MIE-alerts) rather than individually. The evaluation was performed using an in-house workflow in KNIME, an analytics platform interface (Berthold *et al.*, 2007) to ensure that they identified a minimum of 50% of the MIE domain. The workflow consisted of three parts (Fig. 1):
    a. Input and translators: the input should be a text file (.txt) with information for 694 compounds (i.e. SMILES strings, Chemical Name, MIE) followed by two translator nodes (i.e. Molecule Type Cast and RDKit from Molecule) that bring the entries in a readable format for the next step. For entries that fail the translation step, an alternative SMILES strings format should be provided.
    b. RDKit Substructure filter: Setting as a variable one SA of interest in SMARTS strings per RDKit Substructure filter node, the filter imposes the SA on each entry of the input. For a set of SAs corresponding to a single MIE, the appropriate number of RDKit Substructure Filter nodes were added in the workflow.
    c. Output: A prediction label is added in the entries that match the SA of interest (using the 'Rule Engine' metanode) and displayed using the 'Interactive Table' node, where all input information (i.e. SMILES strings, Chemical Name, MIE) are displayed along with the chemical

structures. Finally, the alerts were given numerical labels with the first number indicating the MIE entry and the second the number of the alert (e.g. 17.1).

For the cases of MIE assignment of compounds out of the intended MIE domain as collated before (work described in Chapter 4), a targeted literature search was conducted to investigate whether the SAs had reliably predicted secondary effects (examples discussed in the Results and Discussion section). In cases of low specificity or underperformance (i.e. less than 50% predictability), steps 2 and 3 were repeated.

5) Inclusion: To increase coverage and enforce the predictability of the *in silico* profiler, two strategies were followed. The first strategy was to combine MIEs with overlapping MIE domains, with *de novo* labelling derived from the number of the original entries (e.g. #15.21). Alternatively, chemical fragments with well-known specific toxic activity relevance, developed using steps 1-4 but failing to display MIE-specificity (e.g. carbamates) were pooled together under the label 'Specific toxicity', sharing a common 'MIE entry' number in their ID.

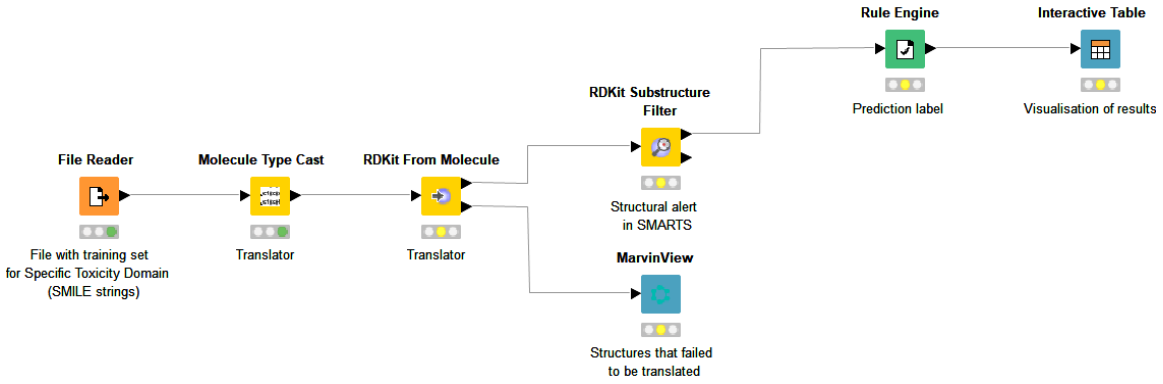


Figure 5.1. In-house workflow to evaluate the performance of structural alerts. It comprises a file reader to input the training set for the specific toxicity domain as SMILES strings (.txt); two translator nodes to bring the input in appropriate format for the RDKit Substrustruce filter to impose the SA of interest for each entry of the input file; and two nodes to prepare (i.e. add a prediction label) and display the entries that have been successfully matched with the SA of interest. For SMILES strings that fail to be translated in RDKit format, a node to display these structures was implemented and the input was amended based on these findings.

5.3.2.2. Non-Specific Reactivity

There are extensive resources, SAs and *in silico* profilers for the mechanistic characterisation for non-specific reactivity in the literature, especially for DNA and protein covalent binding reactions. Instead of developing SAs *de novo*, online resources and tools (e.g. ToxAlerts (https://ochem.eu/alerts/home.do); OECD QSAR Toolbox) were evaluated to retrieve available SAs with a strong mechanistic basis. These are described in more detail below.

ToxAlerts (v.3.0.69.3 accessed August 2018) is an open source structural alert browser, where SAs are organised based on a set of filters (e.g. chelating agents) or endpoints (e.g. acute aquatic toxicity, skin sensitisation, non-genotoxic compounds). For every structural alert entry, there is information on the corresponding functional group, associated endpoint or filter and literature reference (Sushko *et al.*, 2012a, Sushko *et al.*, 2012b).

The OECD QSAR Toolbox (v.4.1) contains four profilers for DNA/protein binding; two donated by the OECD (i.e. DNA binding by OECD, protein binding by OECD), two developed by the Laboratory of Mathematical Chemistry (i.e. DNA binding by OASIS, protein binding by OASIS) (Mekenyan *et al*., 2004, Serafimova *et al.*, 2007).

The profilers with the strongest mechanistic basis were:

a) DNA binding by OECD, developed by Liverpool John Moores University. It is a compilation of 60 (new or re-defined) mechanistic fragments in the form of SAs for DNA binding. They cover six broad organic chemistry mechanisms and denote the most comprehensive listing of structural alerts related to the molecular initiating event of DNA covalent binding (OECD, 2010, Enoch and Cronin, 2010).
b) Protein binding by OECD, developed by Liverpool John Moores University. The profiler consists of 16 mechanisms covering 52 structural alerts, with data supported by mechanistic chemistry and

up-to-date literature that moulded a mechanistic chemistry framework. The alerts were grouped into five mechanistic domains based on the presence of a common reactivity site into so-called mechanistic alerts (Enoch *et al.*, 2011).

Both profilers generate 3-tier predictions corresponding to the type of interaction – mechanism (e.g. acylation); description of the interaction (e.g. ring opening acylation) or chemical groups (e.g. isocyanates and related chemicals); and structural alert (e.g. ketenes).

Non-specific reactivity is the most well defined mechanistic domain with numerous studies providing information on mechanisms such as Michael addition (e.g. (Rodriguez-Sanchez *et al*., 2013, Schultz *et al*., 2005), alkylation (e.g. alkylation, DNA, https://aopwiki.org/events/97; alkylation, protein, https://aopwiki.org/events/244). The MIE literature presented in Chapter 4 provides a chemistry framework for documented structural features with documented non-specific reactivity toxicological relevance for aquatic species. Reviews from Enoch *et al.* (Enoch and Cronin, 2010, Enoch *et al*., 2011) have detailed mechanistic information at the MIE level for a very wide range of non-specific reactivity events. These reviews (Enoch and Cronin, 2010, Enoch *et al.*, 2011) have been the basis for the DNA binding by OECD and protein binding by OECD profilers, developed to facilitate chemical category formation and implemented in the OECE QSAR Toolbox. These profilers were selected to be the basis for defining the MIE non-specific reactivity domain. Alerts were selected from the OECD documents (OECD, 2010, OECD, 2011, Enoch *et al.*, 2010, Enoch *et al.*, 2011) for the profilers on DNA binding by OECD and protein binding by OECD respectively per MIE (entries #3-14 (Table 2)), as follows:

a) Using the OECD QSAR Toolbox 4.2 (OECD, 2018), the DNA binding by OECD and protein binding by OECD profilers were applied to the MIE domains for MIE entries #3-14 (in SMILES strings). The 3-tier predictions generated (e.g. SN2> Episulfonium Ion Formation> Mustards) were compared with the available MIE information. For cases where the suggested 3-tier prediction

matched the MIE of interest, the structural alerts defining the third tier of the prediction (e.g. mustards) were retrieved.

b) For the training sets in SMARTS strings, the chemical groups suggested by the literature (e.g. quinones) were matched with mechanisms of comparable structural feature-activity relevance (e.g. quinones and quinone-type chemicals), and the corresponding alerts were retrieved (e.g. benzoquinones, pyranones, quinone-diimine, quinone-imine, quinone-methides).

For instances of discrepancies of translation of SMARTs from the OECD QSAR Toolbox to RDKit in KNIME, documentation of the equivalent versions of the OECD QSAR Toolbox profilers (DNA binding by OECD, protein binding by OECD) in KNIME (Covalent protein binding alerts, Covalent DNA binding alerts as found in http://knimewebportal.cosmostox.eu/) was consulted. The retrieved SAs were organised per MIE domain and given an Alert ID with first part indicating the MIE they correspond to (i.e. assign) and the second serving as a unique identifier.

5.3.2.3. Narcosis

The last domain aims to capture mechanistic knowledge relating to narcosis and narcotic effects. There are studies hypothesising the exact mechanism of action at the cellular level (e.g. Roberts and Costello (2003)) but limited number of studies capturing MIE related data (i.e. Antzcak *et al.* (2015)). The narcosis domain consists of MIE entries #1 and #2 (Table 2). There was a chemical overlap between MIE domain #1 and MIE domain #2, leading to merging the two entries (narcosis MIE training set of n=103). The narcosis profiler was constructed under the premise that if compounds are captured by the non-specific reactivity or specific toxicity rules then they are within the narcosis domain. A set of excluding criteria was implemented, based on results of preliminary analysis and using Chemotyper (v.1.1.) to identify chemistry related to compounds falsely assigned to the narcosis domain (i.e. false

positives); and incorrectly assigned to other domains ahead of narcosis (i.e. false negatives); as well as the literature on narcosis (Ellison *et al.*, 2008).

Taxonomic applicability of narcosis domain includes fish (i.e. *D. rerio, L. macrochirus, O. mykiss, P. promelas, P. reticulata*) and crustacean species (i.e. *D. magna*). A study by Aruoja *et al.* (2014) demonstrated that the toxicity of non-polar chemicals to an algal was well correlated with hydrophobicity. Li *et al.* (2015) also highlighted the potential of using algae species as a surrogate for fish lethality, namely for baseline and less inert compounds while acknowledging species sensitivities need to be considered for interspecies extrapolation relevant to narcosis. The MIE narcosis training set is rather diverse, and as suggested by Li *et al.* (2015), it is prudent to consider species sensitivities before interspecies extrapolation. Due to time constraints, such analysis was not possible to be conducted; therefore, it was decided not to include algal species in the taxonomic applicability of the narcosis domain on the grounds of the precautionary principle.

5.3.2.4. Performance of in silico profilers

All SAs and *in silico* profilers were organised in a KNIME workflow. Following the template of the in-house workflow to evaluate structural alert performance, the main components of the workflow were an input node (i.e. SMILES strings in .txt format), translator nodes, one Rule engine metanode for MIE assignment and two output nodes (Figure 2a). The Rule engine metanode consisted of three metanodes, one per mechanistic domain (Figure 2b). Each mechanistic domain metanode comprised a Table creator node containing all SAs (in SMARTS strings) of the respective mechanistic domain along with information on the respective SAs (Alert ID, MIE, MIE target) followed by a translator; an RDKit substructure filter node; and a Rule engine metanode adding three prediction labels 'MIE Target', 'MIE', 'Mechanistic Domain' (Figure 2c). The Table Creator node for the narcosis domain contained all SAs for MIEs 3-65 along with specific toxicity specific SAs (#66) and a set of exclusion criteria. The RDKit

substructure filter node had two input ports, one variable port for SAs and one chemical structure port for compounds to be filtered (Figure 2c). The RDKit substructure filter node had two output ports, for specific toxicity and non-specific reactivity the output was the port 'Molecules matching filter', and for narcosis the port 'Non-matching filter'. For the specific toxicity metanode, there were two levels of filtering; one level with the MIE-specific alerts for 48 MIEs, followed by a level of domain-specific alerts for the compounds with no MIE assignment.

Evaluation of SAs per MIE was performed as a group (#MIE SAs) rather than individually, as the objective of this work was to map effectively the chemical space per MIE domain. Performance of the *in silico* profilers and their specificity in MIE assignment was assessed by their application to the list of compounds derived from all MIEs. The equation used to calculate SAs performance was:

$$\%\ Structural\ alerts\ performance\ = \frac{correctly\ assigned\ compounds\ included\ in\ the\ MIE\ domain}{number\ of\ compounds\ in\ the\ MIE\ domain} * 100$$

As mentioned, there were cases of compounds assigned to multiple MIEs within the MIE training set (n=694), capturing potential secondary mechanisms of action. Therefore, evaluation of the SAs was focused on their respective MIE domains rather than the predictability of the SAs within the MIE training set in its entirety. The structure of the scheme deliberately allows compounds to be assigned to multiple MIEs across the specific toxicity and non-specific reactivity domains but not in the narcosis domain i.e. it is not possible for a compound to fall within both the narcosis and specific toxicity domain or non-specific reactivity domain, therefore, the analysis of the results was performed per domain.

Lastly, to evaluate the profiler the number of correctly assigned, false negative and false positive compounds assigned in the Narcosis domain was calculated.

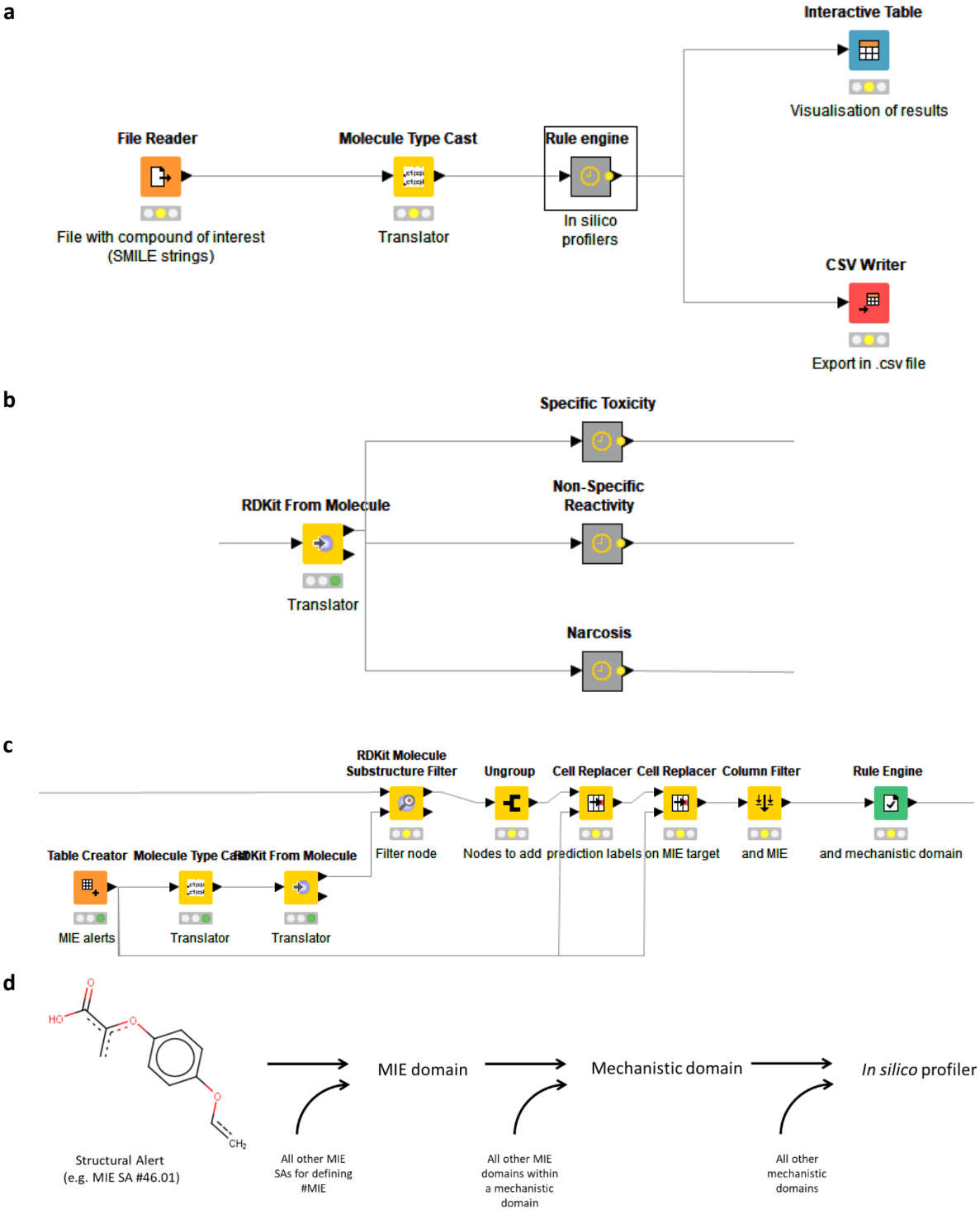


Figure 5.2. Layout of the KNIME workflow (a), consisting an input node (File reader), a translating node (Molecule Type Cast), an *in silico* profiling metanode (Rule engine), and output nodes (Interactive Table; CSV writer). Within the Rule engine metanode (b), there are three *in silico* profilers for MIE assignment for specific toxicity, non-specific reactivity and/or Narcosis domains. The basic structure of each profiler per domain is presented in (c) and a summary on the principle of MIE assignment is described in (d).

5.4. Results and Discussion

The objective of this chapter was to map the chemical space of the MIE literature (as described in Chapter 4) and create an *in silico* profiler to assign chemical compounds to MIEs. This was achieved by writing, or collecting, structural alerts from the literature, to define the chemical space of each MIE. The MIEs were grouped into three mechanistic domains, with the MIE alerts organised into *in silico* profilers that allow for assignment of a chemical into an MIE and mechanistic domain. The alerts and their analysis are summarised below.

5.4.1. Specific Toxicity

For the purposes of this thesis, specific toxicity refers to a mechanism of action with a defined molecular target leading to a specific AOP and thus a well-defined adverse outcome. From the compilation of MIEs as described in Chapter 4, 48 MIEs were found to be associated with specific toxic action. The literature research focused on capturing information on MIEs and the corresponding chemistry, with 548 compounds and 12 functional groups (e.g. carbamates) (referred to as the MIE training set) associated with the 48 MIEs. For each MIE, a set of alerts was developed based on the corresponding chemistry of the MIE of interest (the MIE domain). Following a five step method, as described in the Methodology Section, SAs were developed incorporating information from both the MIE domain and mechanistic literature. Examples of how these steps were applied are presented in Table 5.2.

This led to 127 structural alerts (SAs, full list in SMARTS in Appendix 2). In Appendix 2, an additional 'MIE entry' (MIE #66, specific toxicity) covers SAs for MIE domains and chemical classes well-known for specific toxic action but with misclassification in MIE-assignment.

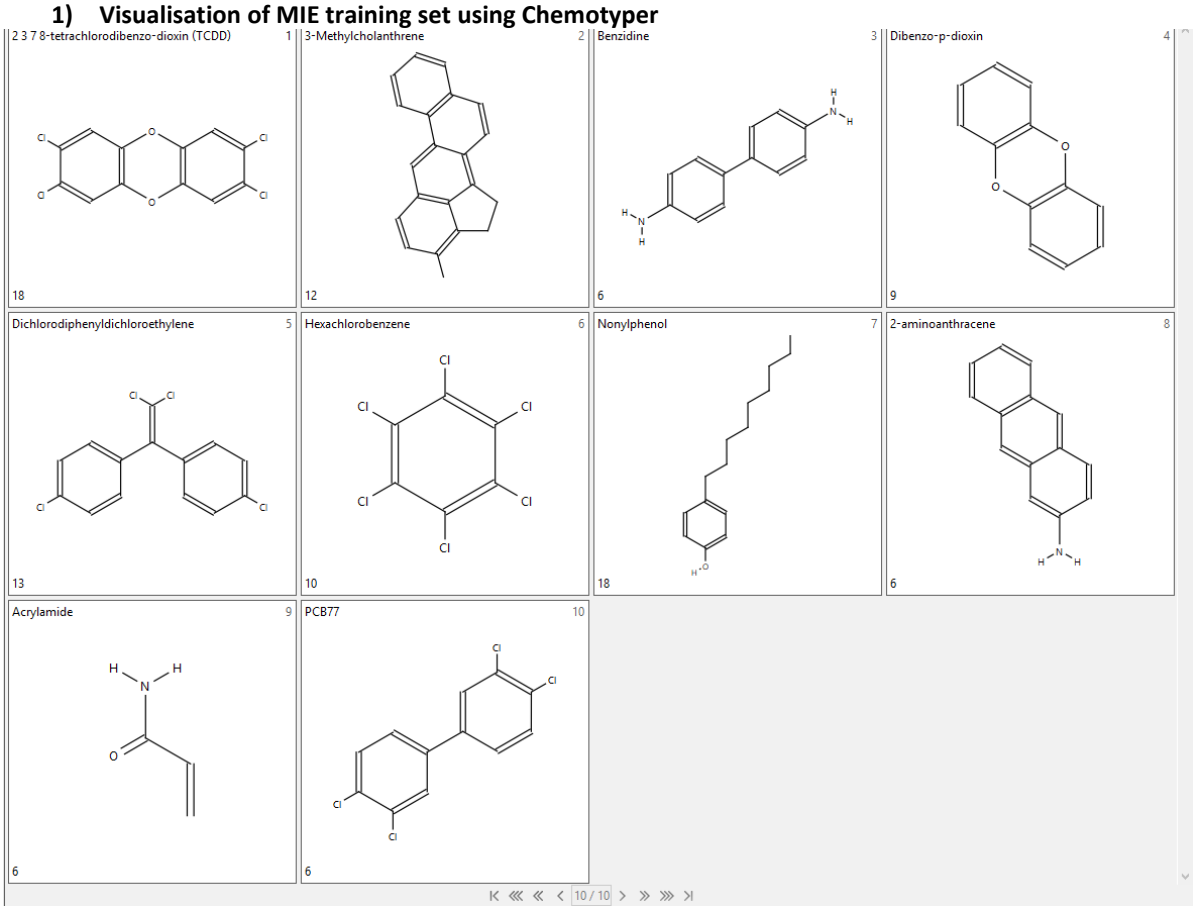
Table 5.2. Examples using the five-step method described in Section 5.3.2.1 for structural alert development incorporating MIE information for 1. AhR activation and 2. microtubule interactions.

Example 1

MIE #15.21 AhR activation, covalent binding leading to induction of CYP1A expression

---

1) Visualisation of MIE training set using Chemotyper



The applicability domain also includes the following chemical species halogenated aromatic hydrocarbon, indoles, polychlorinated biphenyls, polychlorinated dibenzofurans, polycyclic aromatic hydrocarbons (PAHs).

---

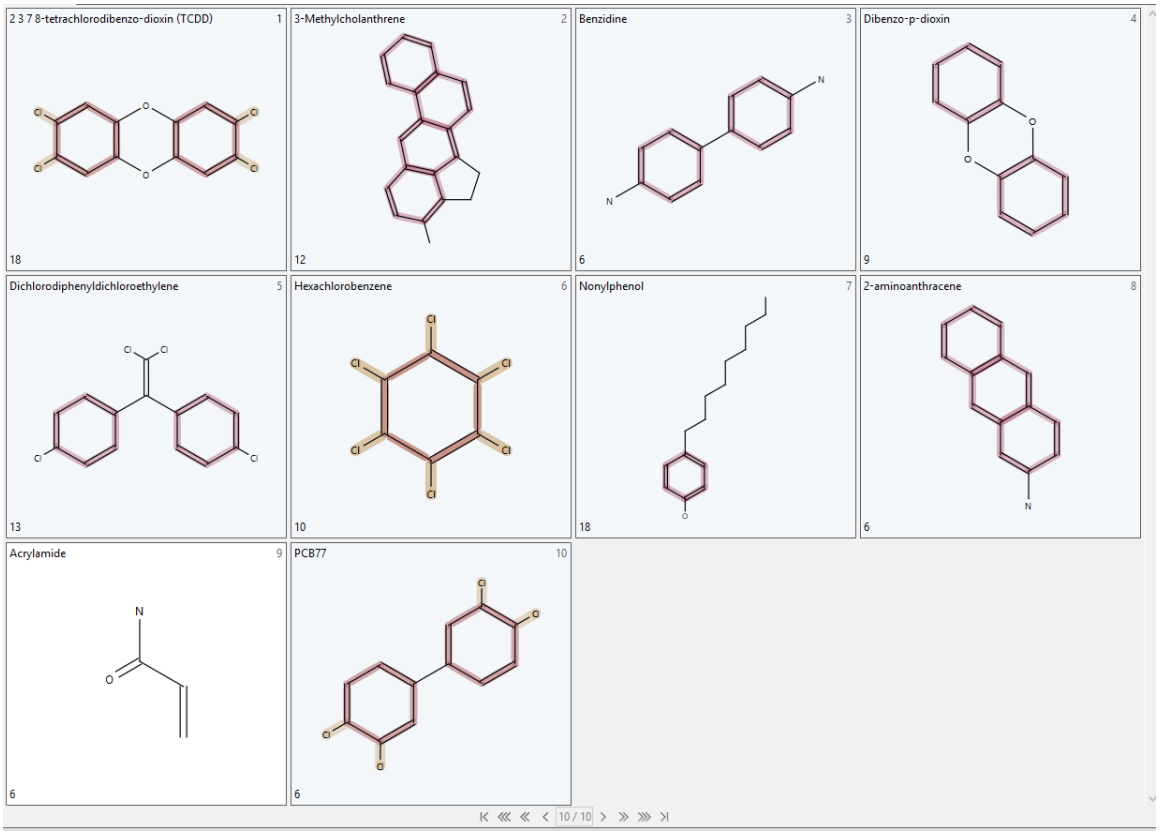
2) Development of structural fragments using findings from the Chemotyper

Chemotyper analysis highlighted five key structures related to AhR activation:

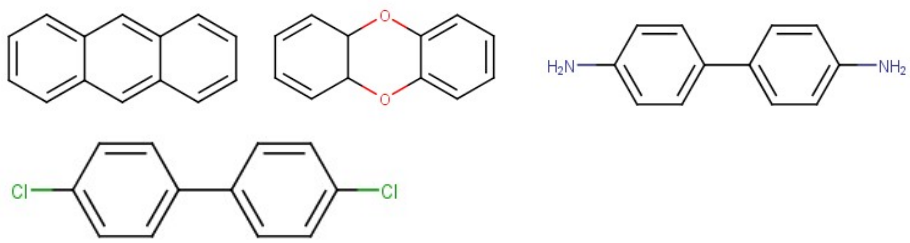
F1	ring:aromatic_benzene	9/10
F2	bond:CX_halide_aromatic-X_generic	4/10

F3	bond:X[any]_halide	4/10
F4	bond:CX_halide_aromatic-X_dihalo_benzene_(1_2-)	3/10
F5	bond:CX_halide_generic-X_dihalo_(1_2-)	3/10

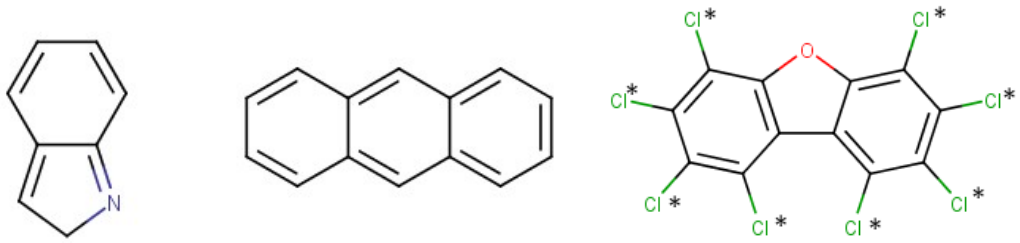
The fragments F2, F3 and F4, F5 describe similar structures therefore the fragments F1, 2 and 4 were selected.



Based on these findings the following alerts were directly designed using Marvin Sketch by ChemAxon.

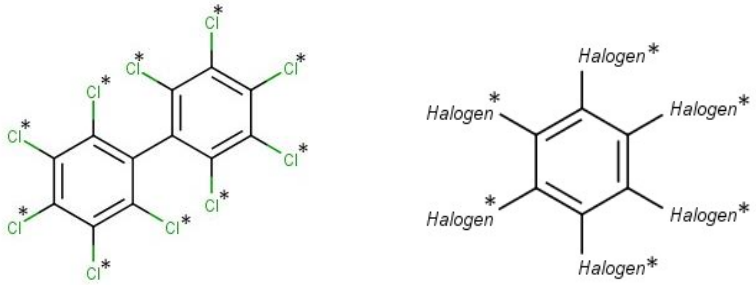


Structural expressions of the chemical species within the training set are presented below (for SMARTS strings , see Appendix 2)



indoles polycyclic aromatic hydrocarbons polychlorinated dibenzofurans

*potential substitution



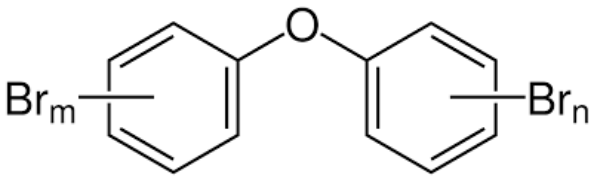
polychlorinated biphenyls

*potential substitution

---

3) Refinement

A targeted literature review revealed polybrominated diphenyl ethers as potential agonists of AhR, with *in vivo* and *in silico* supporting evidence (Jiang *et al.*, 2016, Yang *et al.*, 2017, Zhang *et al.*, 2018b). Due to the limited data availability, it was decided not to include this fragment at this time.



(image by wikidata.org)

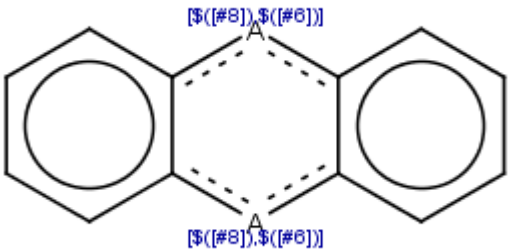
polybrominated diphenyl esters

---

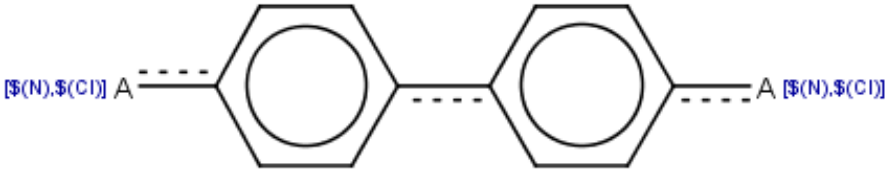
4) Evaluation

Taking into account all the information collected the following SAs were applied to the MIE training set (n=694) using a KNIME workflow as described in Figure 5.1.

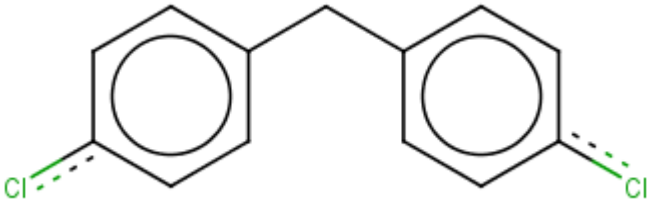
SA1



SA2



SA3



The combination of SA1 and SA2 correctly identified 6 out of 10 of the MIE #15.21 training set. It also identified as potential AhR agonists 2 compounds from the reactivity domain (MIE #4), dibenzo[a,l]pyrene and 3-methylcholanthrene which have been associated with AhR mediated CYP1 induction (Lewis and Lake, 1996, Schreiber and Podack, 2009).

SA3 successfully identified 1 out of 10 of the MIE#15.21 training set. However it lacked specificity as it misclassified multiple compounds from other MIE domains.

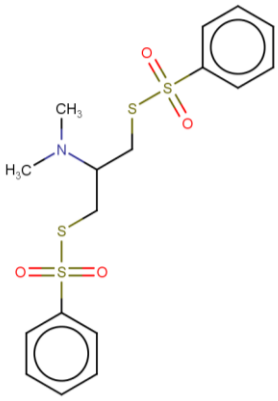
SA1, SA2 and the SAs to describe the chemical species associated with the MIE #15.21, as presented above constitute the #MIE SAs 15.21 (for full list see Appendix 2).

Example 2

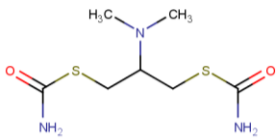
MIE # 43, Interaction with multiple sites of the microtubule leading to assembly disruption and cell division

---

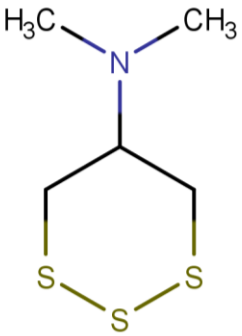
1) Visualisation of MIE training set using Marvin View



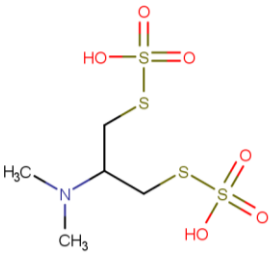
Bensultap



Cartap



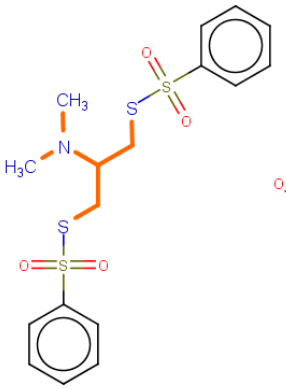
Thiocyclam



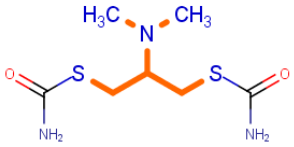
Thiosultrap-sodium

---

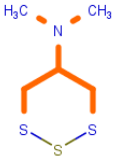
2) Empirical development of structural fragments



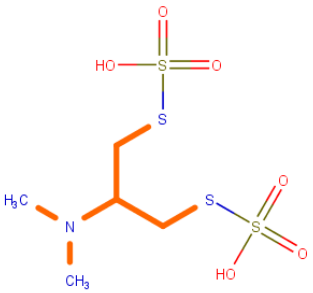
Bensultap



Cartap



Thiocyclam



Thiosultrap-sodium

Identification of a fragment based on chemical similarity as highlighted on structures above.

3) Refinement

No evidence of additional chemistry associated with this MIE.

4) Evaluation

When the fragment above was applied to the MIE training, it assigned only the four compounds to MIE #26.

The 48 MIE domains comprising the Specific toxicity domain ranged in size (i.e. 1-65, average 12 compounds per MIE domain) covering a wide range of chemical classes (more details on chemistry see Chapter 4). MIE-SAs (i.e. groups of SAs corresponding to a single MIE domain) were developed to identify a minimum 50% of chemicals from the training set assigned to that MIE domain. Due to the variable number of chemicals within an MIE domain, there was no restriction on the number of SAs comprising a MIE-SAs.

The development of SAs was based on chemical similarity within the MIE domain training set and mechanistic information. One of the objectives in SA development was MIE-specificity. Abundance or lack of chemical similarity has been the main caveat in SA development; the approaches to tackle limitations in SA development discussed below.

High overlap among MIE domains was observed for the MIEs relating to the Aryl hydrocarbon Receptor (AhR) pathway and CYP1A mediated MIEs, as well as MIEs describing acetylcholinesterase inhibition. The AhR pathway is a very well established mechanism that is predominantly (Larigot *et al.* 2018) but, not exclusively, activated by CYP1A induction (Abron *et al.* 2018, Mortensen and Arukwe 2007). The overlap between these two MIEs is partially expected as CYP1A expression is predominantly controlled by AhR activation after exposure to organic pollutants in multiple aquatic species (e.g. fish (Mortensen and Arukwe, 2007, Yuan *et al.*, 2013); frogs (Zimmermann *et al.*, 2008); invertebrates (Kim *et al.*, 2015)). Based on this knowledge, it may be assumed that they are part of the same AOP, with AhR activation as the MIE and induction of CYP1A expression as a key event. Therefore, they have been merged (MIE #15.21 AhR activation (leading to CYP1A expression), Appendix 2). Another example of MIEs being merged due to high overlap among MIE domains is acetylcholinesterase inhibition (MIEs #22, #23). Acetylcholinesterase inhibition is a well-described interaction (Čolović *et al.*, 2013b), with irreversible (#22, Table 5.1) and reversible (#23, Table 5.1) effects mediated predominantly by organophosphates and pyrethroids respectively. Despite mechanistic evidence on the different MIEs, there is a high degree of overlap between the two MIE domains (#22, #23); therefore, they were merged for the purposes of SA development into one (#22.23, Appendix 2).

Conversely, there was little similarity amongst the chemicals within MIE domains # 17, 19, 56, 57, 58 (Table 5.1), which restricted MIE-specific SA development. In some cases, this resulted in a SA being

associated with a single chemical. To tackle this issue, the compounds from these five MIE domains (i.e. # 17, 19, 56, 57, 58) were grouped under an additional 'MIE entry' under the name 'specific toxicity' (for more details see Appendix 2).

Another limitation in adequately assigning compounds to MIEs strictly based on 2D descriptors was the chemical class of carbamates. Carbamates are a relatively reactive chemical class known to interact reversibly with acetylcholinesterase (Čolović et al., 2013). Carbamates target a range of MIE targets (e.g. acetylcholinesterase, nicotinic acetylcholinesterase, PSII triazine site, DHP synthase, β-tubulin assembly in mitosis, fatty acid synthesis (Casida, 2009)) with high species-specific selectivity. Taking the example of thiocarbamates, they are associated with cholinesterase inhibition (Čolović et al., 2013), however, selected thiocarbamates (e.g. dimepiperate) are also associated with inhibition of fatty acid synthesis in algae (Casida, 2009). Ideally, this high selectivity (in MIE target and taxonomic applicability) would be reflected within the MIE-SAs for compounds (i.e. carbamates, thiocarbamates, dithiocarbamates) to be correctly assigned to MIEs, as it could support decision-making on joint toxicity or mixture assessment. However, this cannot be the case using the current approach to SA development for the SAs identifying carbamates (i.e. carbamates, thiocarbamates, dithiocarbamates) in MIE specific fashion. They were, however, included in the 'specific toxicity' MIE (#66, Appendix 2), to ensure inclusivity and correct mechanistic domain assignment.

To evaluate the performance of the MIE assignment, the MIE training set for the specific toxicity domain (n=568) was used as a test set, due to its high specificity of the targets and taxonomic applicability. Even though there are datasets with known mechanistic information in literature, they are taxa specific (e.g. the Fathead Minnow Acute Toxicity database) and/or do not capture information at MIE level for targets relevant to aquatic vertebrates (e.g. fish) and invertebrates (e.g. *Daphnia*, algae). The assessment of the predictability of MIE alerts was performed driven by correct MIE-

assignment (e.g. set of MIE SAs) and not by assessing performance of individual alerts per MIE of interest. Taking the example of MIE #15 (nAChR agonism (neonicotinoids)), the MIE domain comprises 10 compounds and there are two MIE-SAs with a predictability of 70% (i.e. 7 out of 10 compounds correctly assigned). This percentage reflects the predictability of correct MIE assignment by either of the two MIE #15 SAs (MIE-SA), over the individual SAs which are 30% and 40% respectively. Predictability per MIE-SA for correct MIE and mechanistic domain assignment is presented in Table 5.3.

On average, each MIE-SAs consists of three SAs with 79.1% correct class assignment of the respective MIE training set. The average ratio of individual SAs to compounds within the MIE specific toxicity domain is 1:4. For MIE-assignment only (#15.21-65, 116 alerts in Appendix 2), 78.8% of compounds associated with specific toxicity were correctly assigned. Expanding the analysis to MIE- and mechanistic domain assignment (i.e. including MIE entry #66) an additional 10.4% of the training set was correctly assigned (89.2% of the training set in total). After development of all alerts (n=127), only 6% of the MIE specific toxicity training set was misclassified to the incorrect MIE, whilst still classified in the specific toxicity domain, with the remaining 4.8% not be assigned to any MIE.

The training set for the specific toxicity domain consisted of varied chemistry with well-defined targets. It is expected these compounds may have more than one MIE target, however as part of the SA development, only the primary target was taken into account. During SA development and analysis, the profiler correctly identified secondary MIE effects of compounds within the training set that match evidence in the literature. Examples of compounds with secondary MIE targets identified through SA development were diclofop-methyl, metosulam, asulam and endosulfan. Diclofop-methyl is associated with the inhibition of acetyl-CoA carboxylase (#46, Appendix 2) and there is evidence that it is also associated with endocrine disruption mediated by juvenile hormone mimicking (#63,

Appendix 2) (Meyer *et al.*, 2006). Metosulam leads to inhibition of amino acid biosynthesis via binding to acetohydroxy-acid synthase (AHAS) (#41, Appendix 2) and there is also evidence on depigmentation effects via disruption of the carotenoid synthesis in *Vicia faba* plants (Badr *et al.*, 2013) (#38, Appendix 2). Asulam inhibits DHP synthase, leading to inhibition of the conversion of 4-aminobenzoic acid to 7,8-dihydropteroate, a key reaction part of the reductive Acetyl-CoA pathway. MIE-assignment suggested that asulam has an inhibitory role in mitosis by preventing microtubular assembly (# 44, Appendix 2). Endosulfan acts as an antagonist on the GABA-gated chloride channel (# 29, Appendix 2) and literature provides evidence and supports the identified disruptive effect on respiration and ATP levels (#62, Appendix 2) (Rainey *et al.*, 2017).

The work presented explores the potential of utilising current MIE knowledge to extrapolate and reliably derive conclusions among taxa and chemical domains. The current approach serves as a proof of concept and relies solely on the strength of high quality data and simple methodologies for definition of chemical domains using fragments to explore whether MIE derived SAs could be used as the basis for class assignment.

A rounded definition of specific toxicity MIE training sets could also include Lipinski's rule of five for drug-likeness, where additional rules related to molecular weight, octanol-water partition coefficient and number of atoms with hydrogen bond potential could be implemented (Lipinski *et al.*, 1997). The addition of Lipinski could assist in screening out compounds with low oral bioavailability.

Table 5.3. Performance of the set of SAs associated with a single MIE (MIE SAs) for MIEs within the specific toxicity domain, along with the descriptive information on the training set per MIE and correctly assigned compounds

# MIE ID	Training set		SAs	Compounds	% MIE-SAs
	compounds	Chemical groups	(n)	correctly assigned	performance
15.21	10	8	16	7	70.0
16	1			1	100.0
18	3		1	2	66.6
20	5	1	2	5	100.0
22.23	65	3	5	48	73.8
24	10		2	7	70.0
25	2		1	1	50.0
26	4	1	1	4	100.0
27	42	2	3	39	88.6
28	2		1	1	50.0
29	4		2	4	100.0
30	6		2	3	50.0
31	3		1	2	66.7
32	24		4	18	75.0
33	6		3	4	66.7
34	20		4	19	95.0
35	2		1	1	50.0
36	26		2	15	57.7
37	7		4	6	85.7
38	1		1	1	100.0
39	8		2	4	50.0
40	2		1	2	100.0
41	50		7	46	92.0
42	2		1	2	100.0
43	14		2	11	78.6
44	3		2	2	66.7
45	1		1	1	100.0

46	20	5	19	95.0
47	19	4	18	94.7
48	21	5	15	71.4
49	5	2	4	80.0
50	36	5	27	75.0
51	3	1	2	66.7
52	7	2	7	100.0
53	1	1	1	100.0
54	4	2	3	75.0
55	9	2	6	66.7
59	22	2	14	63.6
60	20	3	15	75.0
61	7	2	5	71.4
62	9	3	6	66.7
63	5	2	5	100.0
64	4	1	4	100.0
65	13	1	10	76.9
66		10	55	85.4

5.4.2. Non-Specific Reactivity

Non-specific reactivity, as discussed previously, entails interactions targeting biological macromolecules in a non-discriminatory fashion, associated with oxidative stress (Lushchak, 2016). Reactivity is observed in fish as gill irritation (Drummond *et al.*, 1986), whereas no information was available on the target sites and phenotypes in aquatic invertebrates. Non-specific reactivity interactions have been widely described at the MIE level (Enoch *et al.*, 2011, Enoch and Cronin, 2010). A number of studies experimentally verifying *in silico* findings e.g. electrophilic and proelectrophilic interactions with the ciliated protozoan *T. pyriformis* (e.g. (Nelms *et al.*, 2013, Richarz *et al.*, 2014, Rodriguez-Sanchez *et al.*, 2013)). As a result, the OECD QSAR Toolbox profilers for covalent binding to

DNA and Protein ((OECD, 2010, OECD, 2011) were developed with the aim of capturing events at the molecular level (i.e. MIE).

As discussed in Chapter 4 (summary in Table 5.1), 12 MIEs associated with non-specific reactivity were retrieved from the aquatic toxicology literature, with information on the type of reaction (e.g. Michael addition), chemistry (i.e. 30 compounds and 17 functional groups, MIE Reactivity training set; details in Appendix) and sufficient evidence of biological occurrence (phenotypic and/or molecular) in multiple aquatic taxa.

5.4.2.1. Definition of the non-specific reactivity domain

The DNA and protein binding profilers by OECD capture, in high detail, the potential of a reactive event to occur between the xenobiotic and a biological (macro)molecule. The DNA and protein binding profilers by OECD were applied to the MIE Reactivity training set (n = 36 compounds) and, based on given mechanistic alert assignment from the profilers, 191 SAs were retrieved from relevant documentation (OECD, 2010, OECD, 2011). A schematic representation of the SA retrieval is presented in Figure 5.3, followed by two worked examples (MIE#4 and #13, Table 5.1) on how SA selection was performed.

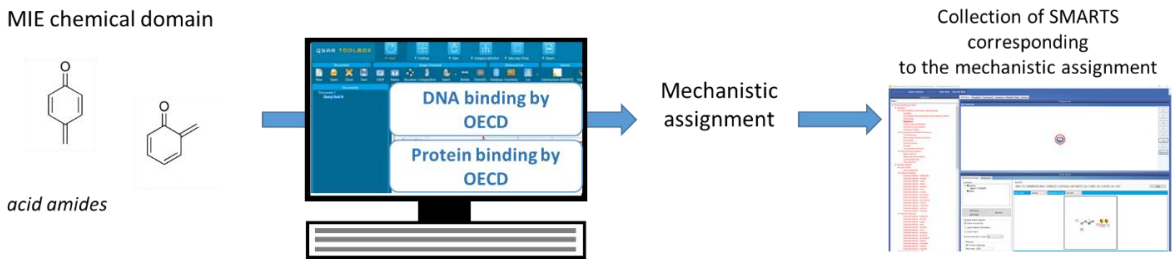
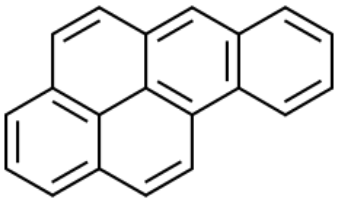


Figure 5.3. Outline of selection process for MIEs of reactivity domain.

Taking the example of oxidated metabolite-mediated DNA adduct formation, targeting DNA (MIE #4, Table 5.1), the MIE training set is presented in Figure 5.4 below.

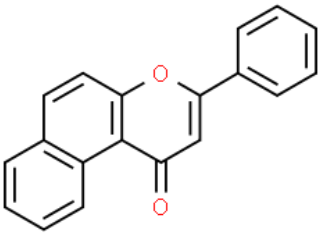
Compounds

a)



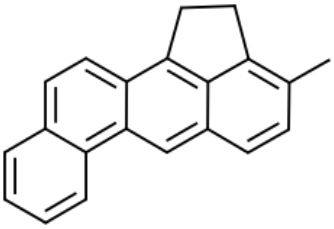
Benzo[a]pyrene

b)



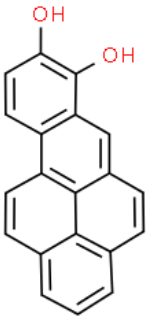
Beta-naphthoflavone

c)



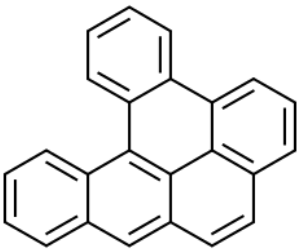
3-methylcholanthrene

d)



Benzo[a]pyrene-7,8-diol

e)



Dibenzo[a,l]pyrene

Chemical species

a) polycyclic aromatic hydrocarbons (PAHs) with so called 'bay region'

b) polycyclic aromatic hydrocarbons (PAHs) with so called 'fjord regions'

Figure 5.4. Training set of MIE #4, Oxidated metabolite-mediated DNA adduct formation

For the compounds within the MIE non-specific reactivity training set, DNA binding by OECD and Protein binding by OECD profilers were applied (example in Figure 5.5).

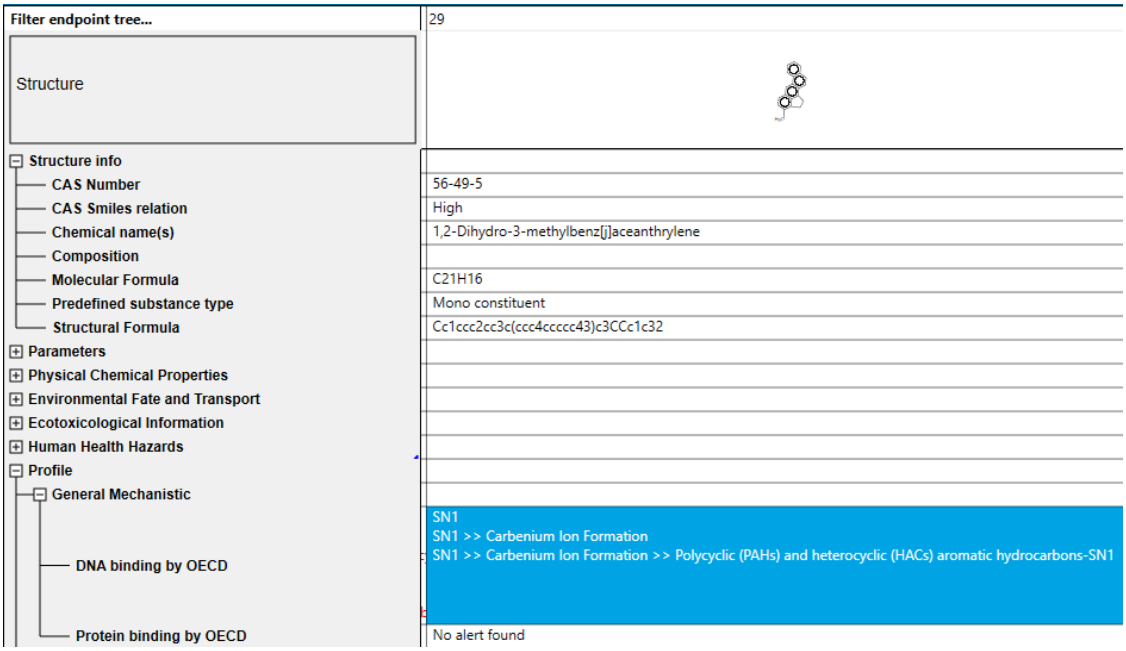


Figure 5.5. Example of mechanistic assignment after application of DNA binding by OECD and Protein binding by OECD on 3-methylcholanthrene

The compounds were assigned by mechanism of action (e.g. SN1), mechanistic domain (e.g. Carbenium Ion Formation) and structural alert (e.g. Polycyclic (PAHs) and heterocyclic (HACs) aromatic hydrocarbons-SN1). Based on the assignment of the mechanistic domain, all the associated SMARTS, as found in the corresponding profiler (e.g. DNA binding by OECD), were selected (Figure 5.7) and checked for compatibility with the RDkit node in KNIME.

The DNA binding by OECD and Protein binding profilers by OECD are also coded as KNIME workflows (http://www.cosmostox.eu/what/knime/) (Enoch and Cronin, 2010, Enoch *et al.*, 2011, OECD, 2008, ECHA, 2008). For SMARTS not compatible with RDkit node in KNIME, such as [42] Polycyclic (PAHs) and heterocyclic (HACs) aromatic hydrocarbons-SN1, [ch]1[ch]c2c([ch][ch]1)[a~1][a~1][a~1]1[a~1]2[a~1][a~1][a~1][a~1]1.$[$[[#7]]..4,[#6]]~1, the quivalent as suggested by the supporting documentation of the KNIME workflows that use the same node (i.e. RDKit) was selected (i.e. [cH]1[cH][cH]c2c(ccc3ccccc23)[cH]1).

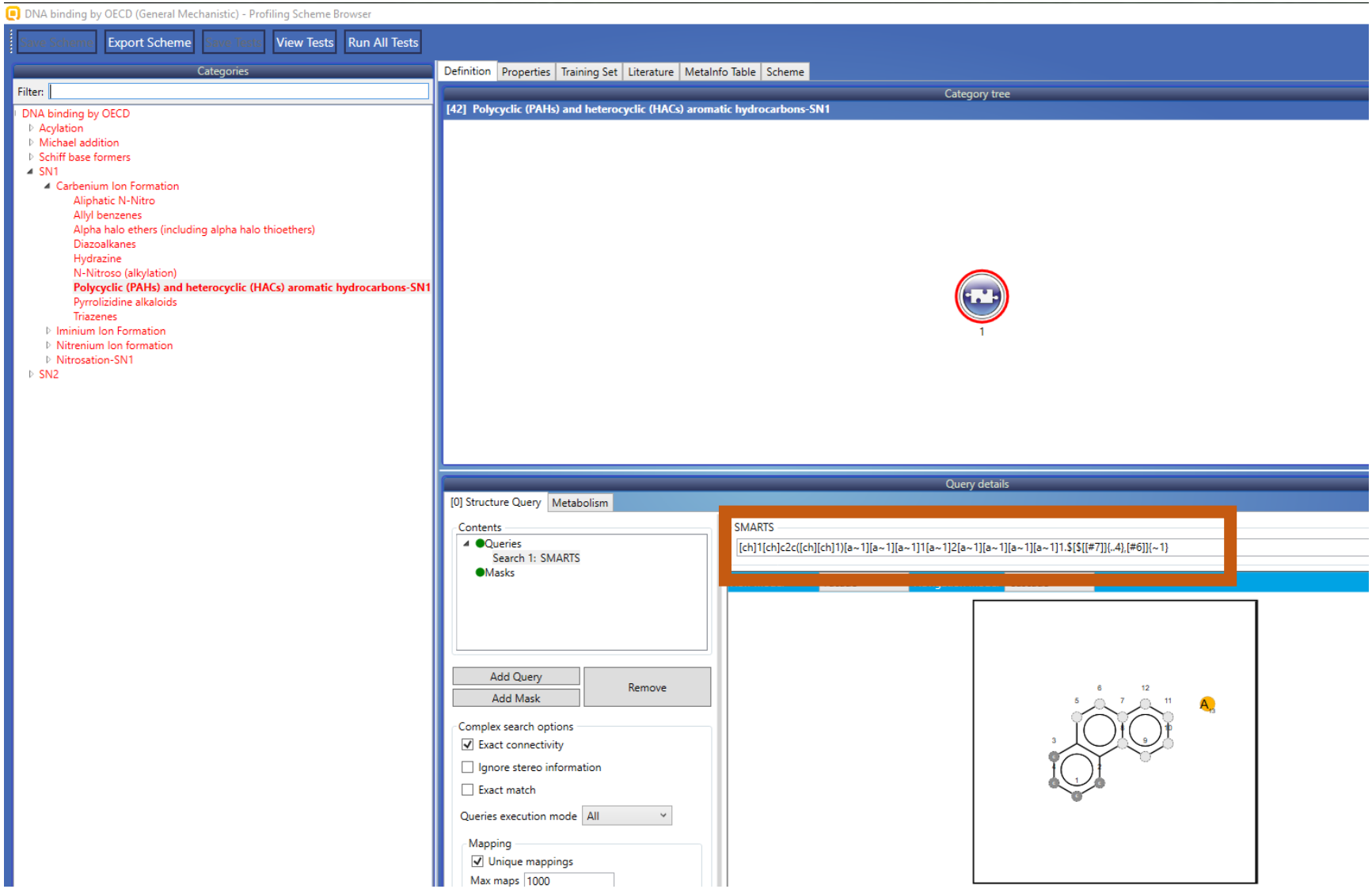


Figure 5.7. Example of SMARTS retrieval based on mechanistic domain and structural alert assignment by DNA profiler

For chemical species within the MIE non-specific reactivity training set (e.g. coumarins), structural alerts fitting the description as found in literature were selected. Using the example of 'polycyclic aromatic hydrocarbons (PAHs), with the so-called 'bay region', the SMARTS string [cH]1[cH][cH]c2c(ccc3ccccc23)[cH]1 was selected:

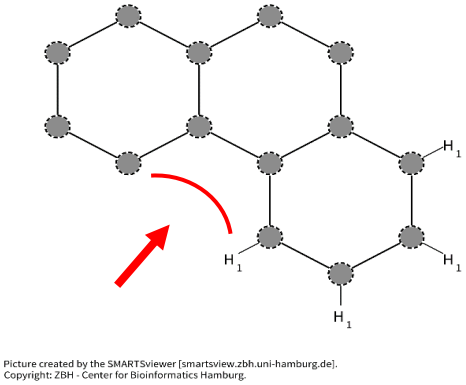


Figure 5.6. Structural representation of bay region in polycyclic aromatic hydrocarbons (PAHs)

The same steps were taken for all compounds and chemical species constituting the MIE reactivity domain, with the corresponding retrieved SAs organised based on MIE. The only exception to this was MIE #13, which describes bipyridyl-mediated redox cycle activity in biomolecules in chloroplast and lung. The training set of MIE #13 consists of paraquat and diquat and when DNA binding by OECD, and Protein binding by OECD was applied no alert was assigned. Therefore, based on the structures of the molecules an alert was developed as shown in Figure 5.8.

Appendix 2 lists the SAs retrieved which correspond to MIEs within the non-specific reactivity domain (MIE #2-14). SA selection is based on high quality chemical mechanistic data (i.e. (Enoch and Cronin, 2010, Enoch *et al.*, 2011, OECD, 2011, OECD, 2010)) with supporting evidence on biochemical manifestation and biological occurrence (i.e. MIE knowledge).

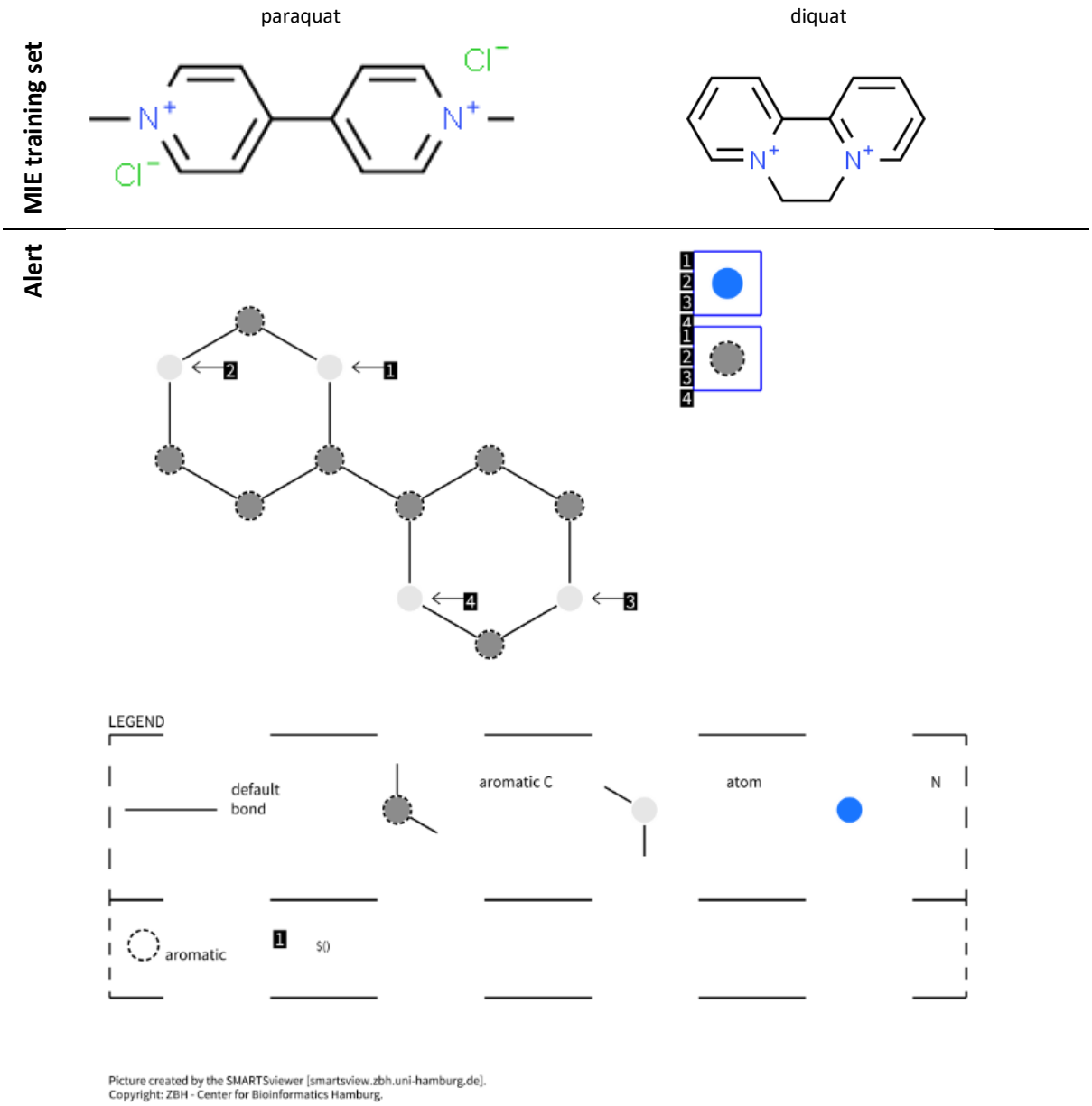


Figure 5.8. Training set for MIE #13, and developed alert to describe the set

On average, every MIE-SAs consists of 17 SAs with 100% correct class assignment of the respective MIE training set. The overall ratio of individual SAs to compounds within the MIE non-specific reactivity domain is 4:1. For MIE domains comprised of SMARTS strings exclusively, they were the basis for SAs development and were included in the corresponding MIE SA (i.e. #6 and #14, Table 5.4). The SAs

retrieved readily define the respective MIE domains (i.e. compounds and functional groups), with 100% in alert performance for all MIE-SAs (Table 5.4). MIE #12 for ROS formation mediated by CYP2E1 was excluded due to limitations in the training set (n=1, ethanol). However, acetaldehyde, the reactive product of CYP2E1-mediated ethanol oxidation, falls within the non-specific reactivity domain with MIEs describing P450-mediated interactions (e.g. MIE #5, Appendix 2).

Table 5.4. Performance of MIE-SAs for MIEs within the non-specific reactivity domain, along with the descriptive information on the training set per MIE and correctly assigned compounds

#	Training set		SAs (n)	Compounds correctly assigned	% SAs performance
	SMILES	SMARTS			
3	10	5	39	10	100
4	4	2	20	4	100
5	5	1	27	5	100
6		1	50	*	100
7	1	1	8	1	100
8	5	2	24	5	100
9	2	1	18	1	100
10	1		1	1	100
11	1		3	1	100
13	2		1	2	100

* n/a for MIEs with training set consisting of exclusively SMARTS

5.4.2.2. Links between the non-specific reactivity and specific toxicity domains

DNA and protein binding are significant events for specific toxicity and non-specific reactivity. Therefore, compounds may well initiate events related to both specific toxicity and non-specific reactivity. 209 compounds from the specific toxicity training set were also assigned in multiple MIEs within the non-specific reactivity domain. The most common reactive MIE associated with specifically acting compounds was ROS formation (liver) (MIE #5, Appendix 2) followed by Redox cycle activity (Fenton-chemistry-like generation of hydroxyl radical (MIE #11, Appendix 2). Notably, compounds assigned to both domains were predominantly compounds with MIEs related to respiratory toxicity (i.e. succinic dehydrogenase (mitochondria), complex IV, proton gradient (mitochondria), complex I electron transport, respiration, ATP synthase); photosynthesis (i.e. Photosystem II- urea site,

Protoporphyrinogen IX oxidase, Phytoene desaturase, ubiquinol oxidase at Qo site, Photosystem I); and fatty acid synthesis (i.e. VLCFA (very long chain fatty acid) synthesis, Acetyl-CoA carboxylase, 14a-demethylase (CYP51), squalene epoxidase (flavoprotein monooxygenase)). The aforementioned MIE targets are traditionally herbicide targets (Casida, 2009), with the corresponding chemistry usually considered to fall within the non-specific reactivity domain, which agrees with the findings presented here. With a number of MIEs within the specific toxicity domain being taxa or group specific (e.g. algal specific MIEs on photosynthesis inhibition), it is crucial to include other toxicity pathways of the same chemistry for other aquatic species.

These observations highlight the benefit of the presented work as it contains information across many species and allows investigation of species-specific as well as more general MIE/AOPs, where it could potentially identify additive effects for a defined taxonomic domain, and quantified within a quantitative AOP framework.

5.4.3. Narcosis

Narcosis refers to a non-specific toxic effect from exposure to a xenobiotic. For humans, narcosis is defined as a general non-specific CNS depression (Rocco *et al.*, 2019) occurring, for example, after ethanol consumption (Spanagel, 2009), inert gas inhalation (e.g. nitrogen narcosis (Rostain and Lavoute, 2016) or exposure to anaesthetics (e.g. propofol). Narcotic effects in humans have been associated with inhibition of neurotransmission, with ethanol associated with disruption of cys-loop ligand-gated ion channels and ionotropic glutamate receptors (Abrahao *et al.*, 2017), nitrogen narcosis been linked with desensitisation of GABA(A) receptor activity (Rostain and Lavoute, 2016), and propofol acting as a positive modulator of the GABA(A) receptor (Chen *et al.*, 2019). For aquatic species, on the other hand, narcosis has a rather broad meaning due to high taxonomic diversity.

Narcosis manifests as a set of well-described physiological responses in fish (Bradbury *et al.*, 1990), effects on swimming behaviour in daphnid species (Bownik, 2017) and it is speculated to disrupt cell function in algae (McCarty *et al.*, 2013). This thesis investigated the potential of using MIE knowledge to facilitate class assignment for multiple aquatic species. Therefore, at this stage, and due to time constraints, it was decided not to investigate and expand the narcosis domain to algal species, while acknowledging it is possible.

5.4.3.1. Definition of the narcosis domain

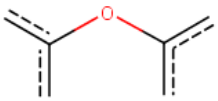
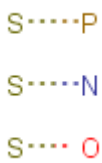
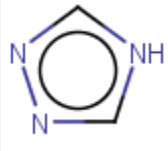
MIE selection: Solely based on available MIE knowledge at the time of MIE data collection, the narcosis domain comprises two MIEs, with the corresponding chemistry covering a wide spectrum of chemical classes (Table 5.1, further details in Appendix 2 and Chapter 4).

Definition of chemical space: The chemical space of narcosis is rather diverse including aliphatic and aromatic hydrocarbons, halogenated hydrocarbons, alcohols, ethers, ketones, weak acids and bases, and aliphatic nitro compounds (Verhaar *et al.*, 1992, Ellison *et al.*, 2008). There was a high structural overlap between the two MIEs of the Narcosis domain. As there was not enough information at the MIE level for distinction between the two, the MIE domains were merged (n=103). Due to the merge of the two MIE entries within the domain, only the mechanistic domain assignment was generated.

Profiler development: The profiler was built under the premise that compounds not acting via a non-specific reactive or specific toxic action are considered within the Narcosis domain. This strategy was applied with the intention to define, as clearly as possible, the applicability domain of narcosis. In preliminary analysis, there were 141 compounds which did not fall into either the specific toxic action or non-specific reactivity domains. Of these, 96 were considered to be narcotic and 45 were not. The most common structures found within the incorrectly classified compounds, with toxicological relevance as supported by literature, were included as exclusion criteria to avoid further

misclassifications. This process resulted in the reduction of the false positive assignment to 15 as shown with supporting literature data in Table 5.5. The exclusion SAs are associated with reactivity and specific toxic action (i.e. not in MIE level) and could potentially be implemented in the specific toxicity and non-specific reactivity domain, providing the MIE, MIE target and taxonomic applicability can be adequately defined. After implementation of the five SMARTS strings presented in Table 5.5, the final profiler was applied to the MIE training set (n=694) and the results are presented in Table 5.6.

Table 5.5. Implemented structures (a-e) for exclusion from narcosis domain

	SMARTS	Chemical structure	Mechanistic background
a.	cc(c)Oc(c)c		9% of the compounds in MIE specific toxicity training set contain this structure at least once. It is associated with herbicidal activity (Kouji *et al.*, 1989); and NADPH-dependent, CYP450-catalysed O-dealkylation, potentially leading to formaldehyde formation (Miyazawa *et al.*, 2001)
b.	FC(F)(F)		14% of the compounds in MIE specific toxicity training set contain this structure at least once. Even though fluoroalkanes are considered stable compounds, unsaturated fluocarbons are considered more unstable, and due to the highly polarised (C-F) bonds, potentially reactive (O'Hagan, 2008, Lemal, 2004)
c.	[$(S~P),$(S~N),$(S~O)]		21% of the compounds in MIE specific toxicity training set contain this structure at least once. Organic thiophosphates, sulfamides and sulfoxides are associated with non-specific reactivity events (Enoch *et al.*, 2011, Enoch and Cronin, 2010c)
d.	c1ncnn1		10% of the compounds in the specific toxicity training set contain this structure at least once. 1,2,4-triazoles are associated with antifungal activity (Pautus *et al.*, 2006, Peyton *et al.*, 2015)

e.	C#C		3% of the compounds in MIE specific toxicity training set contain this structure at least once. Alkyns are associated with reactivity due to a (good) leaving group (Verhaar *et al.*, 1992)

Profiler performance: As shown in Table 5.7, 80.6% of the compounds in the narcosis domain were correctly assigned within that domain, with 13 compounds being misclassified out of the domain (the full list is reported in Table 5.7) and 15 misclassified within the domain (reported in Table 5.8). Assignment in the narcosis domain is exclusive, whereas compounds can be assigned in both the specific toxicity and non-specific reactivity domains if they fall within the respective MIE domains.

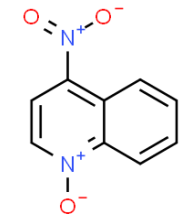
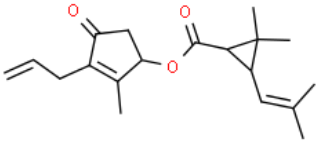
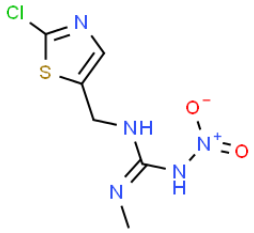
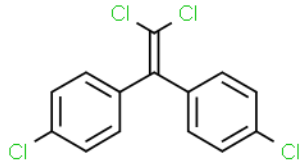
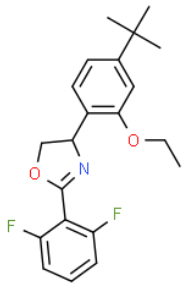
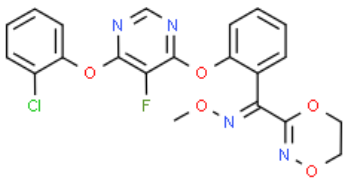
Table 5.6. Performance of MIE-SAs for MIEs within the narcosis domain as applied in MIE training set (n=694), along with the descriptive information on the training set per MIE and correctly assigned compounds

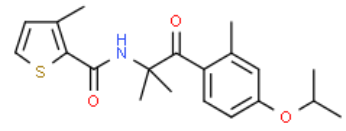
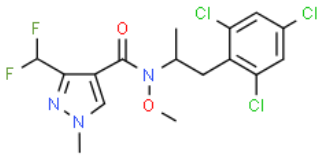
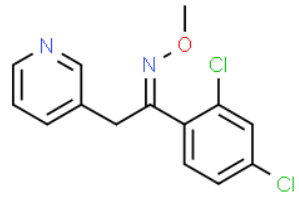
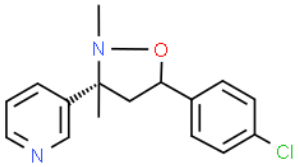
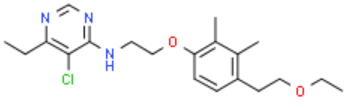
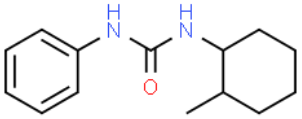
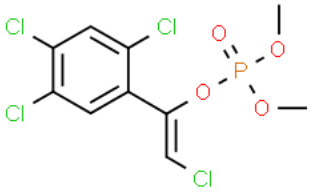
Narcosis training set (SMILES)	Compounds correctly assigned	False negatives	False positive	True negatives
103	83 (80.6%)	13	15	576

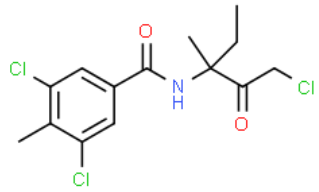
Table 5.7. List of compounds that should have been assigned in the narcosis domain (false negatives).

Compound	Supporting evidence on mechanistic background
1,1,2,2-tetrachloroethane 1,1-dichloroethane 1,2,3-trichloropropane 1,2-dichloroethane 1,2-dichloropropane Benzene m-Xylene p-Xylene Pentachloroethane	Median effective concentration (EC50) for compounds acting via narcosis can be directly calculated by well-established QSARs (Verhaar *et al.*, 1992). Toxicity ratio calculation by Verhaar *et al.* (1992) for fish and conclusions from Ellison *et al.* (2008) for *T. pyriformis* provide sufficient evidence to justify assignment of chloroalkanes in the narcosis domain.
1-octanol Tripropargylamine	In a study presented by McKim et at (1987), fish exposed to 1-octanol and tripropargylamine expressed behavioural responses associated with narcosis. *In silico* evidence support assignment in the narcosis domain for *D. magna* (Zhang *et al.*, 2013)
Dimethylaminoterephthalate (Dimethyl p-phthalate) Phenyl-4-aminosalicylate	Classification of dimethyl p-phthalate in the narcosis domain is supported by *in vivo* data ((US Environmental Protection Agency, 2008) and QSAR analysis using *T. pyriformis* data (Ellison *et al.*, 2008)

Table 5.8. List of compounds that should not have been assigned in the narcosis domain (false positives).

Chemical structure	Compound	Supporting evidence on mechanistic background
	4-nitroquinoline 1-oxide	Associated with DNA damage (https://www.cdc.gov/niosh-rtecs/VC200B20.html) in multiple species via alkylation (Carvalho *et al.*, 2012)
	Allethrin	Pyrethroids are well-established modulators of voltage gated sodium channel leading to deactivation, with available studies in multiple aquatic species (Murayama *et al.*, 1972, Wang *et al.*, 1972, Zhorov and Dong, 2017, Davies *et al.*, 2007)
	Clothianidin	Neonicotinoids are well-established competitive agonists of nicotinic acetylcholine receptor (Casida and Durkin, 2013, Casida, 2018) with supporting evidence on effects on non-target aquatic species (e.g. (Miles *et al.*, 2017))
	Dichloro-diphenyl-dichloroethylene (DDE)	DDE has been associated with AhR activation (Gaspar-Ramirez *et al.*, 2015), interference with DNA methylation patterns (Olsvik and Softeland, 2018) and oxidative phosphorylation (Elmore and La Merrill, 2019).
	Etoxazole	A study by Nauen and Smagghe (2006) has demonstrated that insecticidal mode of action of etoxazole is chitin biosynthesis inhibition, with data supporting *D. magna* sensitivity (Chang *et al.*, 2019).
	Fluoxastrobin	Fluoxastrobin, as other strobilurin fungicides (Zhang *et al.*, 2018a), act via targeting complex III site of mitochondrial respiratory chain leading to inhibition ubiquinol oxidase at Qo site (Casida, 2009), with toxic effects observed in daphnids (Cui *et al.*, 2017) and algal species (Liu *et al.*, 2015).

 	Isofetamid Pydiflumetofen	Isofetamid and pydiflumetofen act by interacting with the complex II of the enzyme leading to inhibition of mitochondrial respiratory chain succinic dehydrogenase (Casida, 2009). Orthologs of succinic dehydrogenase are present in multiple aquatic species (Waterhouse *et al.*, 2013). Isofetamid is a phenyl-oxo-ethyl thiophene amide, a chemical class associated with electrophilic reactivity targeting DNA (Enoch and Cronin, 2010a).
 	Pyrifenox Pyrisoxazole	Pyrifenox acts via binding on 14a-demethylase (CYP51) leading to inhibition of the 14a-demethylation during ergosterol biosynthesis (Casida, 2009). Orthologs of CYP51 are present in multiple aquatic species (Waterhouse *et al.*, 2013, Lepesheva and Waterman, 2007).
	Pyrimidifen	Pyrimidifen acts via interfering with the electron transport chain leading to inhibition of mitochondrial respiratory chain, complex I electron transport (Casida, 2009). It belongs to aminopyrimidines, a chemical class not associated with narcosis (Verhaar *et al.*, 1992, Verhaar *et al.*, 2000).
	Siduron	Siduron acts via inhibition of photosynthesis, targeting the urea site of photosystem (Casida, 2009). Siduron contains a phenyl-urea fragment, which is associated with formation of nitrenium ion leading to DNA adduct formation (Enoch and Cronin, 2010a).
	Tetrachlorvinphos	Tetrachlorvinphos belongs to the organophosphates, a chemical class strongly associated with irreversible binding to acetylcholinesterase leading to inhibition (Čolović *et al.*, 2013a).

	Zoxamide	Zoxamide is applied as a cell division inhibitor, acting via disruption of β-tubulin assembly (Casida, 2009), with supporting evidence from Young and and Slawecki (2001).

5.4.3.2. *Suggested implementations for narcosis domain*

For the purposes of this thesis, the narcosis profiler was primarily based on MIE data as a proof of concept. The domain could be further refined by incorporating additional knowledge that is not strongly linked with MIEs. Examples of such knowledge could be *in silico* studies on chemical classes for defining polar and non-polar narcosis applicability domains (e.g. polar effects of anilines and propargylic alcohols to *P. subcapitata* (Chen *et al.*, 2007, Chen *et al.*, 2012); or the on the narcotic effect of chlorobenzenes to *T. pyriformis* (Zhang *et al.*, 2012)). A more systematic approach in defining the narcosis domain, using *in silico* studies, would further refine the mechanistic domain assignment avoiding misclassification of chloroalkanes and xylene (supporting data by (van Leeuwen *et al.*, 1992, Verhaar *et al.*, 1992)); phenols (supporting data by (Ellison *et al.*, 2015)); octanol and tripropargylamine (supporting data by (McKim *et al.*, 1987)). Such implementations would resolve discrepancies in the assignment of compounds such as those noted in Table 5.7 as being out of the domain. Moreover, refinement of the specific toxicity domain to reduce the number of false positive compounds would further increase the predictive power of both the specific toxicity and narcosis profiler.

5.4.4. Application of *in silico* profilers for MIE assignment

The *in silico* profilers are based on data from 65 MIEs with well-defined chemistry and taxonomic applicability. The MIE profilers capture events relevant for vertebrates, invertebrates and algae (Table 5.1) or are taxa specific, with the generated assignments providing information on the MIE, MIE target

and mechanistic domain label. As mentioned previously, compounds assigned in the non-specific reactivity or specific toxicity domain could not be assigned to the narcosis domain. For MIE domains utilised for this analysis, there is extensive supporting information on taxonomic applicability. At this stage, taxonomic applicability information was not a contributing factor for the development of the profilers; something that has proven problematic for compounds within the narcosis domain and simultaneously in non-specific reactivity or specific toxicity domain. For example, studies on acute toxicity of aniline and chloroaniline have demonstrated their narcotic effect in fish, algae and *V. fischeri* (Aruoja *et al.*, 2011, Aruoja *et al.*, 2014, Bradbury *et al.*, 1993) and the high sensitivity of daphnids (Ramos *et al.*, 2002, Yan *et al.*, 2005). However, they are assigned only in the narcosis domain. This discrepancy could be resolved with the profiler factoring taxonomic applicability for its development, with multi-tiered profiling for the different high levels of organisation. Such an approach would allow for expansion of the taxonomic applicability to algae, while enabling interspecies extrapolation to identify MIE targets (e.g. from mammals to fish on neuronal receptors), which is currently restricted due to the way taxonomic applicability is defined.

Furthermore, there is limited high-quality MIE information on aquatic toxicity in current literature, apart from the collated information presented in Chapter 4. Therefore, external validation of the profilers was not possible. Evaluation of the developed profilers was performed using the MIE training set. Another challenge in externally validating the profilers arises due to their broad taxonomic applicability. As demonstrated in Chapter 3 and discussed in Chapter 1, the use of *in silico* MIE profilers in conjunction with *in silico* classification schemes on acute aquatic toxic action (e.g. Modified Verhaar scheme, Acute Aquatic Toxicity by OASIS) would ensure good practice to build weight of evidence.

5.5. Conclusions

This chapter presents a novel *in silico* classification approach utilising MIE knowledge relevant to aquatic toxicity, based on high-quality MIE data collated in Chapter 4. The work presented was intended to serve as a proof of concept for utilising exclusively MIE information to support *in silico* classification. In Chapter 3, the need to further expand *in silico* assignment in specific toxicity classes for aquatic species was highlighted in the context of are a number of studies for the definition of non-specific reactivity events down to MIE level (Ebbrell *et al.*, 2016, Enoch and Cronin, 2010, Enoch *et al.*, 2011). MIEs from the literature were clustered into three mechanistic domains: narcosis, non-specific reactivity and specific toxicity. Limited information is available at the MIE level for narcosis, with only two events available in literature at the time of data retrieval. Therefore, at this stage, the *in silico* profiler was constructed in a similar fashion to Acute Aquatic Toxicity profiler by OASIS (OECD, 2013, Russom *et al.*, 1997), where chemistry that is not associated with non-specific reactivity and specific toxicity MIEs would be assigned to the narcosis domain. Implementation of the narcosis domain with further inclusive and exclusive criteria (as discussed in Section 5.4.3.2) could reduce the probability of false positive classification and provide more confidence in narcosis domain assignment.

Starting with definition of the specific toxicity domain, this was based on 48 MIEs associated with specific toxic action. The training set for specific toxicity MIEs consists of 548 compounds and 12 chemical classes (e.g. polychlorinated biphenyls). The premise for defining the respective MIE domains was structural similarity and the identification of characteristic fragments specific for the interaction (i.e. MIE) of interest. For every MIE, chemical fingerprints were identified using the Chemotyper (https://chemotyper.org/) or empirically depending on the training set size. The MIEs were then evaluated for coverage and specificity and the optimum MIE SAs set selected to describe the MIE domain (see Table 5.2 for worked examples). This process was repeated for all MIEs. Compounds and

MIE sub-domains, where development of MIE-specific SAs was not possible due to lack of chemical similarity or high overlap with multiple MIE domains, were grouped for SA development leading to specific toxicity domain assignment only. In summary, 127 structural alerts and 89.2 % of compounds associated with specific toxicity have been correctly assigned. One of the challenges in the definition of the specific toxicity domain was the lack of chemical similarity within the MIE domain. Taking of MIE #15 (AhR activation) and MIE #21 (CYP1A induction) as examples, their respective domains had a high overlap. After consideration of associated chemistry and studies in multiple species indicating that AhR activation precedes CYP1A induction (e.g fish (Mortensen and Arukwe, 2007, Yuan *et al.*, 2013); frogs (Zimmermann *et al.*, 2008); invertebrates (Kim *et al.*, 2015)), the MIEs were merged and the definition of the chemical space was performed based on merged data (see Table 5.2). The current approach relies solely on the strength of high quality MIE data with implementations utilising other classification approaches, such as Lipinski's rule of five for drug-likeness (Lipinski *et al.*, 1997), further refine domain definition.

Non-specific reactivity is the most well-characterised domain, with available literature capturing mechanistic information at the MIE level (Ebbrell *et al.*, 2016, Enoch and Cronin, 2010, Enoch *et al.*, 201) with detailed supporting documentation for chemical definition of these mechanisms (OECD, 2008, OECD, 2010, OECD, 2011, ECHA, 2008). MIEs and MIE domains with non-specific reactivity provided the biological framework to select SAs from literature (see example in section 5.4.2.2), leading to 191 alerts to readily define the domain. Application of the non-specific reactivity profiler to the MIE training set assigned compounds with predominantly herbicidal activity associated with species specific MIEs in the non-specific reactivity domain. These findings highlight the potential of using MIEs to support class assignment as they can inform decision-making in both mechanistic domain and taxa level.

There is extensive literature on anaesthesia and narcosis effects in human (e.g. studies on ethanol and inert gas exposure (Rostain and Lavoute, 2016, Harris *et al.*, 2008)); however, there is little mechanistic information on manifestation of narcosis in aquatic species. The narcosis domain comprises two MIEs, with very diverse corresponding chemistry (Appendix 2). The narcosis profiler was developed solely based on available MIE knowledge at the time of MIE data collection and under the premise that compounds not acting via a non-specific reactive or specific toxic mechanism of action are assumed to be narcotic. Such an approach led to 80.6% correct assignment of the MIE narcosis domain. Inclusion criteria were developed with a strong chemical basis, as suggested by several *in silico* studies (e.g. (van Leeuwen *et al.*, 1992, Verhaar *et al.*, 1992)), and further refinement of the specific toxicity domain. In comparison to the performance of current approach, the most recent analysis of the modified Verhaar scheme by Ellison *et al.* (2015), modified Verhaar (as found in Toxtree v2.6) was applied to 280 compounds associated with narcosis (182 non-polar narcosis, and 98 polar narcosis), 207 compounds were classified correctly (average 74%, 83% for non-polar narcosis; 73% for polar narcosis). These results highlight that, despite data limitations on mechanistic level, a modular mechanistic approach could be employed to define the narcosis domain and potentially provide evidence for decision-making. Integrating knowledge from chemistry-driven *in silico* and *in vivo* studies would further refine such an approach and provide confidence in data to support hazard identification and risk assessment.

Chapter 5 has developed a novel MIE-based *in silico* classification scheme with all SAs describing the chemical domains are supported by evidence on biological occurrence. The presented work highlights the merits of an MIE-based classification approach, by integrating knowledge from multiple sources while extending and improving current paradigms.

Chapter 6: Discussion and Future Work

This chapter provides a summary and brief discussion of the main conclusions from Chapters 2 to 5. In addition, concepts for the future work required to develop the *in silico* classification scheme further are presented. The description of future work is focused on improvements to the design structural alerts derived from MIE knowledge; profiler design for respective domains; data quality; and how the scheme can be used to facilitate hazard identification for aquatic toxicity.

6.1. Summary of work

Recent European legislation has legally acknowledged and endorsed the principle of 'the polluter pays' with regard to providing the information to ensure safety of environmental exposure, while basing decision making under the scientific precautionary principle for compound registration and ERA to industry. This process has highlighted the lack of information regarding the effects of many chemicals on the environment (details on ERA in Chapter 1). At the same time, the use of *in vivo* animal testing was emphasised as a last resort, with recommendations for data to support hazard identification to be derived from available resources or non-invasive (e.g. *in silico*) methods (Van den Brink et al., 2018). Thus, there has been a shift towards a range of alternative techniques to traditional testing to provide information that will be suitable for regulatory purposes (Kienzler et al., 2017a).

Within the context of providing non-test solutions to hazard identification, this thesis is founded on over four decades of fundamental research to develop chemistry-based knowledge regarding the effects of chemicals in the aquatic environment. The general premise is that the activity or effects of any molecule can be understood with regard to its chemical structure. Once this knowledge has been developed, it can be applied to make assumptions and predictions about other, similar, molecules. The simple concept has magnified from the development of simple QSARs, exemplified by

Könemann's (Konemann, 1981) fundamental model for the toxicity of non-polar narcosis, into a bewildering array of, on the one hand, multivariate techniques, machine learning and artificial intelligence (Welborn et al., 2018) and on the other hand read-across approaches, bringing together various strands of evidence to make a knowledge-based prediction (Ahlers et al., 2019). There has long been a desire to computerise these models, making it easy for the user to enter a chemical structure and apply the knowledge of experts in a transparent manner. The prediction of ecotoxicological effects is separated from toxicological effects in that the most frequently used software (ECOSAR from the US EPA) has been freely available for approximately 25 years (Clements et al., 1995). At the heart of the majority of these modelling approaches (ECOSAR being a notable exception being class-based) is a foundation in mechanistic underpinning. Verhaar *et al* (1992) were amongst the first to provide a scheme that integrated mechanistic knowledge at that time into a chemistry-based decision tree. The Verhaar scheme has become an "industry standard" and served the industrial and regulatory users well for many years. However, it has seldom been updated or reconstituted. In the context of 21st Century Toxicology there are now many data and informatics resources that can be applied to extend the Verhaar scheme. There is considerable interest in revitalising the scheme to assist in ecotoxicological assessment of chemicals, notably through grouping and read-across, but also extending to other species (Ashauer and Jager, 2018, Kienzler et al., 2017a). Thus, this thesis has seized the opportunity to rethink the concept of mechanistic assignment of chemicals and apply concepts around the MIE, supported where possible by data from New Approach Methodology (NAMs), to create a novel *in silico* tool.

6.1.1. Current literature and *in silico* resources to support hazard identification

There is a need to develop adequate high quality data resources to support modelling and the development of the *in silico* classification scheme proposed in this thesis. Whilst separate datasets for aquatic toxicity exist, there is no one single resource that is suitable in terms of its chemical, species

and mechanistic coverage or that has been suitably curated. With a view to resolving this issue, Chapter 2 focused on investigating the availability of high quality acute toxicity data for aquatic species relevant for the chemical space of industrial chemicals from available literature resources. Great emphasis was placed on collating data for species relevant to aquatic fauna and flora as decision-making is expected to be reflective of the aquatic biota rather than a single aquatic taxon. Evaluating and ensuring the quality of data was a vital component of the work presented in Chapter 2. In order to assess data quality, the availability, accessibility and compliance with OECD guidelines of the original studies were evaluated prior to entering into the database. Data curation extended from the assessment of chemical structure and highlighted poor documentation for a multitude of toxicity data with a lack, or misreporting, of information on test species, test conditions and tested chemistry commonplace. Analysis of the availability of toxicity data confirmed relatively few high quality toxicity studies relevant to aquatic invertebrates (e.g. daphnid and algal species) compared to aquatic vertebrates (e.g. fish species), partially due to no regulatory constraints on non-animal testing.

The toxicity data collated were organised in a database format (referred to as the Acute Toxicity Database), comprising over 6,200 toxicity data for seven species amassed for nearly 5,000 compounds. However, whilst comparatively large, the size of the database inventory in this thesis (n=5,085) corresponded to fewer than 10% of the 74,073 substances as submitted to the European Chemical Agency (ECHA) between 1 June and 1 December 2008 with the intention of pre-registration (referred as the Pre-Registration Inventory). This illustrates the well-known issues of data gaps as well as providing an illustration of why alternative methods of data gap filling, not necessarily involving testing, are required. In order to determine the usability of the available data an analysis was performed of the chemical space covered from the inventory of the database revealing wide chemical variability, and comparable relative frequency of average molecular weight, log K_{OW} among studied

species (see Table 2.4 in Chapter 2 and Section 2.2.1.). Moreover, preliminary regression analysis was performed that revealed significant interspecies relationships among fish species (for details see Chapter 2, Figure 2.6 and Table 2.5); between fish and crustacean species (for details see Chapter 2, Figure 2.7); and between algal and crustacean species (for details see Chapter 2, Figure 2.8). As an initial finding, the prevalence of interspecies relationships could potentially provide a means of data gap filling when sufficient information is available.

The requirement for non-invasive methods (e.g. *in silico, in vitro* approaches) for regulatory and product development purposes has resulted in a considerable impact on this research area, especially with regard to the development of new, and evaluation of existing, techniques. One of the key approaches, especially in the context of this this is the use of category formation, by applying *in silico* profilers, and read-across to fill data gaps (Spielmann et al., 2011). *In silico* profilers relevant to acute toxicity and and available in the OECD QSAR Toolbox are a) the modified Verhaar scheme and b) the Acute Aquatic Toxicity (AAT) profiler by OASIS. These profilers classify compounds into mode of action (MOA) classes based on their chemical structure (for details, see Section 1.2.1). Both profilers are based on data and information from acute toxicity fish studies, potentially limiting their taxonomic applicability to fish species alone, with their relevance to classification to other taxa unknown or not considered. Barron *et al* (2015a) attempted to expand the taxonomic applicability domain for *in silico* MOA profiling, through the publication of a high quality training set of information intended to support mechanistic understanding. The training set proposed by Barron *et al* (Barron et al., 2015a) consisted of an extended list of compounds (referred as MOATox) with broad and specific MOAs assigned by an expert panel; and experimental and calculated acute toxicity data for three fish species and daphnid species. The MOA classes utilised in each of these separate schemes capture different levels of detail, from general MOA description for modified Verhaar, moderately specific (i.e. AAT OASIS) to highly specific (i.e. MOATox) (more details on MOA classes per scheme in Section 1.2.1). This inevitably draws

the question of what level of detail is appropriate within a classification scheme as risk assessment moves to a more mechanistically (or AOP) based paradigm. In order to address this overarching question, the objective of Chapter 3 was to evaluate the performance and accuracy of MOA assignment by the schemes when applied to a dataset with known (or putative) MOAs. One of the key findings from this research was the conclusion to rationalise and organise the mechanistic information for aquatic toxicology into three broad mechanistic domains, with a firm basis on the existing schemes. The three broad mechanistic domains were a) narcosis, b) non-specific reactivity and c) specific toxicity. The advantage to this approach is that along with building on the existing framework, it simplifies the overall approach and allows for detailed mechanistic characterisation within each domain i.e. for narcosis: non-polar narcosis, polar narcosis, ester narcosis etc.; for non-specific reactivity: the established reactivity domains; for specific toxicity: other detailed mechanisms of action based around known MIEs which cause elevated toxicity or chronic effects. The scheme proposed in this thesis is also analogous to other contemporary schemes for organising mechanistic information (Kienzler et al., 2017a).

The analysis of existing schemes for MOA assignment revealed gaps in knowledge relating to the assignment of compounds to the specific toxicity domain by the modified Verhaar and AAT by OASIS schemes. This shortcoming is, to some extent at least, to be expected due to the limited applicability domain of the *in silico* profilers for specific toxicity mechanisms. Taking the Verhaar scheme as an example, the Class 4 alerts represent little more than the organophosphates and carbamates. Moreover, the analysis highlighted the limited level of agreement amongst the three schemes (47-60% for the three domains). The limited agreement in the classification schemes is, at first sight, somewhat of a surprise. However, the schemes have been developed with different aims and objectives and hence have different strengths. Verhaar is the most historic scheme, and hence limited in its applicability domain; AAT by OASIS and MOATox have been developed further. In addition to

these factors, limitations in the chemical and taxonomic applicability domains were observed, with many compounds not being classified, especially for compounds (e.g. surfactants) that had been addressed previously. Thus, the lack of domain coverage and inconsistency in the classification motivated further development of the classification scheme in this thesis.

6.1.2. Using information from Adverse Outcome Pathways to support hazard identification

Chapters 2 and 3 demonstrated the limitations of the current schemes to classify MOA for aquatic toxicity in terms of the range of taxa addressed and the chemical and mechanistic coverage. The improvement of mechanistic classification schemes has taken advantage of the shift in focus of toxicology to a greater reliance on mechanism of action supported by AOPs. A fundamental component of the AOP is the Molecular Initiating Event (MIE) – and this knowledge was applied widely in this thesis to inform the definition of chemistry based profilers. The MIE represents the biochemical interaction between a xenobiotic and the first point of contact in a biological environment. The MIE anchors the chemistry of the compounds to the cascade of events leading to an adverse outcome (see details in Sections 1.3. and 4.1.1); hence, the relationship between chemistry and adverse outcome is cemented.

The thesis developed profilers around the MIE, extending the methods and hypotheses of workers such as Allen *et al* (2014, 2015, 2016). In this context, the MIE was assumed to be the biochemical reference point which could be related to the possibility and occurrence of toxicity being manifested at the molecular level. It is recognised that compounds may initiate one or more interactions with biological targets. As such, this thesis focused on describing these chemical-toxicological landscapes as a map of interactions. In Chapter 4, a comprehensive list of MIEs for aquatic toxicology was identified from the literature. Key information was recorded for each MIE including the MIE target, associated chemistry, detail of the potential to initiate the event, as well as taxonomic applicability.

MIEs that were not taxa specific (e.g. excluding mechanisms such as the inhibition of photosynthesis which is relevant for aquatic plants only), were often pertinent to both aquatic vertebrate and invertebrate species (see Appendix 2 for a detailed description). To ensure inclusivity of chemistry and MIEs for further analysis, both i*n vivo* and NTM data resources were utilised to collate information on MIEs for the new classification scheme. Moreover, to ensure high data quality, an MIE-centred set of criteria was applied to evaluate the information on 90 preliminary retrieved MIE entries. In total the literature search revealed a diverse list of 64 MIEs, relevant to 43 aquatic species (20 vertebrates and 23 invertebrates) and over 700 compounds (the MIE training set) (for a detailed description of collated MIE information see Chapter 4 and Appendix 2). The MIEs were clustered based on the type of interaction into the three mechanistic domains: a) narcosis, b) non-specific reactivity, and c) specific toxicity greatly extending the current knowledge.

The work described in Chapter 5 allowed for the construction of *in silico* profilers for the three mechanistic domains. Using MIEs to populate mechanistic domains and considering the associated chemistry and taxonomic applicability as part of the domain allowed for the creation of a strong and

highly specified mechanistic basis for category formation. The new classification scheme will facilitate better predictions not only at the MOA-level but also at the species level. It also enables the capture of species-specific information related to chemical classes, for instance *D. magna* is known to be more sensitive than fish species to anilines (Ramos et al., 2002)). The same type of information can be captured for individual compounds e.g. atrazine is considered to be a narcotic toxicant to aquatic invertebrates (US Environmental Protection Agency, 2008) but will act specifically towards aquatic plants (see Appendix 2, MIE #32)). The *in silico* profiler was initially built on the chemistry associated with the information collated on the MIE (Chapter 4). The definition of the chemical space was performed by creating 2-D structural alerts (SAs), which were dedicated to each MIE domain (#MIE-

SAs). For instance, development of MIE-SAs for the specific toxicity domain followed a five-step approach (as described in Section 5.3.2.1.); whereas to describe the non-specific reactivity domain MIE-SAs were retrieved from relevant literature (Enoch et al., 2011, Enoch and Cronin, 2010). Similarly, to AAT OASIS, compounds that did not fall in the non-specific reactivity or specific toxicity domains were considered to fall in the narcosis domain. The reasons for defining the narcosis domain indirectly were the lack of chemical similarity and structural analogy of the MIE narcosis training set. The very broad nature of the narcosis domain means that it is not possible to sufficiently describe it by 2D SAs. This issue is exaggerated by the lack of information on the MIE(s) that drive narcosis. This latter point i.e. the MIEs associated with narcosis has long caused problems with its definition, whilst it is a fundamental aspect of acute aquatic toxicology little is definitively known about the MIE with concept varying from non-specific membrane perturbation (Van Wezel and Opperhuizen, 1995) to inhibition of calcium channels (Antczak et al., 2015). In order to assist usability and assure transparency, all profilers were coded computationally in a KNIME interface. The collated MIE training set was used to assess performance of the new *in silico* classification scheme, as it is unique in literature for the amount and level of MIE information relevant to aquatic toxicology. The *in silico* profiler developed adequately assigned compounds in the correct MIEs and mechanistic domains, with 84% of compounds correctly assigned to the narcosis domain, 100% of compounds correctly assigned to the non-specific reactivity domain, and 89% of compounds correctly assigned as being specific toxicants. These performance statistics demonstrate the improvement of the classification scheme in the current thesis above the historical schemes (see Chapter 3). The breadth of the coverage of the new scheme both in terms of chemistry but also mechanisms of action exemplifies this.

6.2. Future work

The major outputs from this thesis include a significant aquatic toxicological data resource, an updated and expanded approach to the classification of chemicals for their effect to environmental species and the definition of the underpinning knowledge in the context of AOPs. Whilst there has been significant progress, inevitably this project has created a number of opportunities for further work in this area, which are summarised below.

6.2.1. Extending the chemical coverage of the *in silico* classification scheme

The methodologies to develop the *in silico* profilers in Chapter 5 relied solely on MIE relevant knowledge to create the SAs. It is inevitable that greater knowledge will become available on the three mechanistic domains (narcosis, non-specific reactivity and specific toxicity) including additional mechanistic data.

The narcosis domain is a prime example where further information could be included in the profiler. For instance, sources of this could be the acute toxicity information from the current literature (as provided in Chapter 2) and the *in silico* profilers relevant to acute toxicity (i.e. modified Verhaar; AAT by OASIS; developed as described in Chapter 5). From these it would be possible to identify compounds within the narcosis domain – and even to extend this to consideration of different species (see Section 6.2.4). Creating such a definitive dataset of narcotic compounds would allow for the definition the narcosis chemical domain. This could be assisted by utilising the well-established correlations of toxicity with log K_{OW} to ensure the definition of the domain. Ellison *et al* (Ellison et al., 2008) have provided an example of how this could be attempted previously for the definition of the narcotic domain to *Tetrahymena pyriformis*.

With regard to the reactivity domain, this has been well-defined in previous work by Enoch *et al* (2010) and Enoch and Cronin (2011) and is currently a lower priority. However, a key aspect here will be

continuing to capture the transformation (biotic and abiotic) of compounds into reactive compounds which otherwise may be misclassified.

As mechanistic knowledge is uncovered about specific mechanisms of toxicity, especially relating to chronic toxicity (see Section 6.2.3), this could also be utilised to expand the coverage of the of the specific toxicity domain. There are clear linkages here with the development of AOPs and understanding of MIEs.

6.2.2. Developing the role of multiple MIEs and biotransformation in the classification scheme

There is an appreciation that multiple MIEs could be part of the same AOP (Allen et al., 2014, Enoch and Cronin, 2010, Enoch et al., 2011). Such information was discussed in Chapter 4 where two MIEs within the same AOP are causally linked to an AO. This concept is especially relevant for biotransformation steps with numerous examples in the literature e.g. CYP biotransformation of benzo(a)pyrene leading to ROS generation; bupirimate and its transformation products dimethirimol and ethimirol which are pyrimidine fungicides and have been considered as having two separate MIEs (Padilla-Sanchez *et al*., 2012). As information on multiple, and transformation-led, MIEs becomes available, it could be implemented in the schemes.

6.2.3. Expanding mechanistic knowledge including chronic toxicity

As noted in Section 6.2.1, there is a need to identify MIEs (as presented in Chapter 4, or the literature) linked, or potentially linked, with chronic adverse effects i.e. non-lethal from long-term low dose exposure. This stems from the assumption that MIEs leading to AOs (chronic and acute), in the biochemical sense, can be narrowed down to defined multitude of biochemical interactions. In the next stage, MIEs and associated chemistry could be linked to chronic or acute AOs. That would be particularly challenging for compounds within the Non-Specific Reactivity and Specific Toxicity domains. Raimondo *et al* (2007) and Wang *et al* (2017) have demonstrated the difference in MOAs

between acute and chronic exposure for metals, narcotics, pesticides, and other organic chemicals. Within the analysis MIEs #62-65 are associated with chronic toxicity in the stage of metamorphosis; further analysis on their respective training sets could potentially reveal the differences between acute and chronic toxic action with a strong mechanistic basis, to serve as supporting evidence for acute-to-chronic extrapolation.

6.2.4. Increasing taxonomic applicability

For the purposes of this thesis, taxonomic applicability was not included when defining the MIE domains and designing the profilers. However, taxonomic applicability could be implemented and play a pivotal role in the design of *in silico* profilers. For instance, taking atrazine as an example, it should be allocated to the narcosis domain for aquatic vertebrates (US Environmental Protection Agency, 2008; Barron *et al*., 2015) but specific toxicity (mediates photosynthesis inhibition) for aquatic plants (Majewska et al., 2018). Definition of the taxonomic applicability domain of MIEs would allow for the generation of predictions with a higher level of information and decrease misclassified compounds. Inclusion of taxonomic-, or even genera- / species-, level information will greatly increase the complexity of the classification schemes. Practically, this would translate into multi-tier profiling. The number of tiers would vary based on available information with each tier representing the taxa/species for the individual MIE. In addition, further information such as species sensitivity distribution (SSD) data could be incorporated.

By introducing taxonomy in the predictions generated, a long-term goal is to extend the current *in silico* profiler to mammalian toxicology. Taking examples of MIE targets such as acetyl-CoA carboxylase or RNA polymerase, which have highly conserved domains within species, or mechanistic domains such non-specific reactivity, defining the chemical applicability domain of MIEs in a defined taxonomy, would allow for the expansion of the framework from environmental to human health purposes.

Another important direction would be to identify conserved domains within MIE targets which are linked directly to MIEs (e.g. proteome targets), rather than specific MIE targets (e.g. biomolecules). This would allow for the definition of MIEs at not only the molecular level, but also at the proteome and genome levels. This approach would allow for the incorporation of omics data to define MIEs and act as supporting evidence for taxonomic applicability or identifying potential novel targets.

6.2.5. Creation of a "smart" computational framework to support safety assessment

The ultimate aim of reliable safety assessment without the use of vertebrates will require a number of technologies and approaches to be successful. Computational approaches will be at the heart of this new paradigm in safety assessment. This thesis has provided a "fledgling" *in silico* classification scheme which, while usable, does have acknowledged limitations in domains and mechanistic and species-specific applicability. Thus, the long-term goal is to develop computational tools that will lead the user through a series of steps, or independently, to perform the analysis, from the entry of a chemical structure through to a decision on the safety, or otherwise, of a chemical in a particular scenario. This will require integrated informatics, modelling and data retrieval (including experimental measurements). The mechanistic framework initiated in work such as that described in this thesis is fundamental to the success of this approach and will be embedded through the assessment of hazard in the safety decision.

6.3. Concluding remarks

There is a need to maintain the speed of development of new approaches to predict the toxicity of chemicals whether it be from QSAR or read-across. However, current publicly available resources on which to develop and evaluate *in silico* approaches have limited high quality acute toxicity data for multiple aquatic species and fewer chronic toxicity data. The *in silico* classification schemes for MOA assignment available for grouping that are relevant to aquatic toxicity do not consistently meet current industrial and regulatory demands, mainly due to limitations in chemical and taxonomic applicability. To address these issues, in this thesis, the use of three mechanistic domains was found to cover a significant proportion of mechanistic space. AOPs were found to be an excellent starting point to source MIEs to develop the novel *in silico* profiler. In particular, the MIE within the AOP framework served bridge between the gap between chemistry at the molecular level and the adverse outcome.

Chapter 7: References

ABRAHAO, K. P., SALINAS, A. G. & LOVINGER, D. M. 2017. Alcohol and the brain: Neuronal molecular targets, synapses, and circuits. *Neuron,* 96**,** 1223-1238.

AFARNEGAN, H., SHAHRAKI, A. & SHAHRAKI, J. 2017. The hepatoprotective effects of aquatic extract of Levisticum officinale against paraquat hepatocyte toxicity. *Pakistan Journal of Pharmaceutical Sciences,* 30**,** 2363-2368.

AGERSTRAND, M. & BERONIUS, A. 2016. Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. *Environ International,* 92-93**,** 590-596.

AHLERS, J., NENDZA, M. & SCHWARTZ, D. 2019. Environmental hazard and risk assessment of thiochemicals. Application of integrated testing and intelligent assessment strategies (ITS) to fulfil the REACH requirements for aquatic toxicity. *Chemosphere,* 214**,** 480-490.

ALLEN, T., LIGGI, S., GOODMAN, J., GUTSELL, S. & RUSSELL, P. 2015. Toward an MIE atlas - tools for toxicity prediction. *Abstracts of Papers of the American Chemical Society,* 250.

ALLEN, T. E. H., GOODMAN, J. M., GUTSELL, S. & RUSSELL, P. J. 2014. Defining Molecular Initiating Events in the Adverse Outcome Pathway framework for risk assessment. *Chemical Research in Toxicology,* 27**,** 2100-2112.

ALLEN, T. E. H., GOODMAN, J. M., GUTSELL, S. & RUSSELL, P. J. 2016. A History of the Molecular Initiating Event. *Chemical Research in Toxicology,* 29**,** 2060-2070.

ANKLEY, G., BENNETT, R., ERICKSON, R., HOFF, D., HORNUNG, M., JOHNSON, R., MOUNT, D., NICHOLS, J., RUSSOM, C., SCHMIEDER, P., SERRRANO, J., TIETGE, J. & VILLENEUVE, D. 2010. Adverse Outcome Pathways: A conceptul framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry,* 29**,** 730-741.

ANTCZAK, P., WHITE, T. A., GIRI, A., MICHELANGELI, F., VIANT, M. R., CRONIN, M. T., VULPE, C. & FALCIANI, F. 2015. Systems biology approach reveals a calcium-dependent mechanism for basal toxicity in *Daphnia magna*. *Environmental Science & Technology,* 49**,** 11132-11140.

AOP-WIKI. 2018. *AOP-Wiki* [Online]. Available: https://aopwiki.org [Accessed 1 April 2018].

ARIENS, E. J. 1986. Receptors: a tool in drug development. *Innovative Approaches in Drug Research***,** 9-22.

ARUOJA, V., MOOSUS, M., KAHRU, A., SIHTMAE, M. & MARAN, U. 2014. Measurement of baseline toxicity and QSAR analysis of 50 non-polar and 58 polar narcotic chemicals for the alga *Pseudokirchneriella subcapitata*. *Chemosphere,* 96**,** 23-32.

ARUOJA, V., SIHTMAE, M., DUBOURGUIER, H. C. & KAHRU, A. 2011. Toxicity of 58 substituted anilines and phenols to algae *Pseudokirchneriella subcapitata* and bacteria *Vibrio fischeri*: Comparison with published data and QSARs. *Chemosphere,* 84**,** 1310-1320.

ASHAUER, R. & JAGER, T. 2018. Physiological modes of action across species and toxicants: the key to predictive ecotoxicology. *Environmental Science-Processes & Impacts,* 20**,** 48-57.

AYANDA, I. O., OLASEHINDE, G. I. & AJAYI, A. A. 2018. Toxicity of sublethal concentrations of glyphosate and paraquat herbicide in the african catfish (*Clarias gariepinus*). *International Journal of Agriculture and Biology,* 20**,** 1359-1364.

BADR, A., ZAKI, H., GERMOUSH, M. O., TAWFEEK, A. Q. & EL-TAYEB, M. A. 2013. Cytophysiological impacts of Metosulam herbicide on *Vicia faba* plants. *Acta Physiologiae Plantarum,* 35**,** 1933-1941.

BAGNYUKOVA, T. V., VASYLKIV, O. Y., STOREY, K. B. & LUSHCHAK, V. I. 2005. Catalase inhibition by amino triazole induces oxidative stress in goldfish brain. *Brain Research,* 1052**,** 180-186.

BANAEE, M., TAHERY, S., HAGHI, B. N., SHAHAFVE, S. & VAZIRIYAN, M. 2019. Blood biochemical changes in common carp (*Cyprinus carpio*) upon co-exposure to titanium dioxide nanoparticles and paraquat. *Iranian Journal of Fisheries Sciences,* 18**,** 242-255.

BARRON, M. G., LILAVOIS, C. R. & MARTIN, T. M. 2015. MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. *Aquatic Toxicology,* 161**,** 102-107.

BASANT, N., GUPTA, S. & SINGH, K. P. 2016. Modeling the toxicity of chemical pesticides in multiple test species using local and global QSTR approaches. *Toxicology Research,* 5**,** 340-353.

BECKER, R. A., ANKLEY, G. T., EDWARDS, S. W., KENNEDY, S. W., LINKOV, I., MEEK, B., SACHANA, M., SEGNER, H., VAN DER BURG, B., VILLENEUVE, D. L., WATANABE, H. & BARTON-MACLAREN, T. S. 2015. Increasing scientific confidence in Adverse Outcome Pathways: Application of tailored Bradford-Hill Considerations for evaluating weight of evidence. *Regulatory Toxicology and Pharmacology,* 72**,** 514-537.

BECKER, R. A., DELLARCO, V., SEED, J., KRONENBERG, J. M., MEEK, B., FOREMAN, J., PALERMO, C., KIRMAN, C., LINKOV, I., SCHOENY, R., DOURSON, M., POTTENGER, L. H. & MANIBUSAN, M. K. 2017. Quantitative weight of evidence to assess confidence in potential modes of action. *Regulatory Toxicology and Pharmacology,* 86**,** 205-220.

BENINA, M., RIBEIRO, D. M., GECHEV, T. S., MUELLER-ROEBER, B. & SCHIPPERS, J. H. M. 2015. A cell type-specific view on the translation of mRNAs from ROS-responsive genes upon paraquat treatment of *Arabidopsis thaliana* leaves. *Plant Cell and Environment,* 38**,** 349-363.

BERTHOLD, M. R., CEBRON, N., DILL, F., GABRIEL, T. R., KOTTER, T., MEINL, T., OHL, P., SIEB, C., THIEL, K. & WISWEDEL, B. 2007. KNIME: The Konstanz Information Miner. *Studies in Classification, Data Analysis, and Knowledge Organization (GdKL, 2007).* Springer.

BOWNIK, A. 2017. *Daphnia* swimming behaviour as a biomarker in toxicity assessment: A review. *Science of the Total Environment,* 601**,** 194-205.

BRADBURY, S. P., CARLSON, R. W., NIEMI, G. J. & HENRY, T. R. 1991. Use of respiratory cardiovascular responses of rainbow trout (*Oncorhynchus Mykiss*) in identifying acute toxicity syndromes in fish. 4.Central Nervous System seizure agents. *Environmental Toxicology and Chemistry,* 10**,** 115-131.

BRADBURY, S. P. & CHRISTENSEN, G. M. 1991. Inhibition of alcohol-dehydrogenase activity by acetylenic and allylic alcohols - concordance with *in vivo* electrophile reactivity in fish. *Environmental Toxicology and Chemistry,* 10**,** 1155-1160.

BRADBURY, S. P., DADY, J. M., FITZSIMMONS, P. N., VOIT, M. M., HAMMERMEISTER, D. E. & ERICKSON, R. J. 1993. Toxicokinetics and metabolism of aniline and 4-chloroaniline in medaka (*Oryzias latipes*). *Toxicology and Applied Pharmacology,* 118**,** 205-214.

BRADBURY, S. P., HENRY, T. R. & CARLSON, R. W. 1990. Fish Acute Toxicity Syndromes in the development of mechanism-specific QSARs. In: W. Karcher, J. Devillers (Eds.), Practical Applications of Quantitative Structure-activity Relationships (QSAR) in Environmental Chemistry and Toxicology, Kluwer Academic Publishers, Dordrecht, The Netherlands (1990), pp. 295-315

BRADBURY, S. P. & LIPNICK, R. L. 1990. Structural properties for determining mechanisms of toxic action - Introduction. *Environmental Health Perspectives,* 87**,** 181-182.

BRAUSCH, J. M. & RAND, G. M. 2011. A review of personal care products in the aquatic environment: Environmental concentrations and toxicity. *Chemosphere,* 82**,** 1518-1532.

BRIDGES, J., SAUER, U. G., BUESEN, R., DEFERME, L., TOLLEFSEN, K. E., TRALAU, T., VAN RAVENZWAAY, B., POOLE, A. & PEMBERTON, M. 2017. Framework for the quantitative weight-of-evidence analysis of 'omics data for regulatory purposes. *Regulatory Toxicology and Pharmacology,* 91**,** S46-S60.

BRILL, J. L., BELANGER, S. E., CHANEY, J. G., DYER, S. D., RAIMONDO, S., BARRON, M. G. & PITTINGER, C. A. 2016. Development of algal interspecies correlation estimation models for chemical hazard assessment. *Environmental Toxicology and Chemistry,* 35**,** 2368-2378.

BUGEL, S. M., WHITE, L. A. & COOPER, K. R. 2013. Inhibition of vitellogenin gene induction by 2,3,7,8-tetrachlorodibenzo-p-dioxin is mediated by aryl hydrocarbon receptor 2 (AHR2) in zebrafish (*Danio rerio*). *Aquatic Toxicology,* 126**,** 1-8.

BURDEN, N., SEWELL F., ANDERSEN, M. E., BOOBIS, A., CHIPMAN, J. K., CRONIN, M. T. D., HUTCHINSON, T. H., KIMBER, I., & WHELAN, M. 2015. Adverse Outcome Pathways can drive non-animal approaches for safety assessment. *Journal of Applied Toxicology*, 35, 971-975.

BURRIS, J. E. & BLACK, C. C. 1983. Inhibition of coral and algal photosynthesis by Ca-antagonist phenothiazine drugs. *Plant Physiology,* 71**,** 712-715.

CAPPELLI, C. I., CASSANO, A., GOLBAMAKI, A., MOGGIO, Y., LOMBARDO, A., COLAFRANCESCHI, M. & BENFENATI, E. 2015. Assessment of *in silico* models for acute aquatic toxicity towards fish under REACH regulation. *SAR and QSAR in Environmental Research,* 26**,** 977-999.

CARVALHO, J. G., NOGUTI, J., DA SILVA, V. H. P., DEDIVITIS, R. A., FRANCO, M. & RIBEIRO, D. A. 2012. Alkylation-induced genotoxicity as a predictor of DNA repair deficiency following experimental oral carcinogenesis. *Journal of Molecular Histology,* 43**,** 145-150.

CASALEGNO, M. & SELLO, G. 2013. Determination of toxicant mode of action by augmented top priority fragment class. *Journal of Chemical Information and Modeling,* 53**,** 1113-1126.

CASIDA, J. E. 2009. Pest toxicology: The primary mechanisms of pesticide action. *Chemical Research in Toxicology,* 22**,** 609-619.

CASIDA, J. E. 2018. Neonicotinoids and other insect nicotinic receptor competitive modulators: progress and prospects. *Annual Review of Entomology,* 63, 125-144.

CASIDA, J. E. & DURKIN, K. A. 2013. Neuroactive insecticides: Targets, selectivity, resistance, and secondary effects. *Annual Review of Entomology*, 58, 99-117.

CASSANI, S., KOVARICH, S., PAPA, E., ROY, P. P., VAN DER WAL, L. & GRAMATICA, P. 2013. *Daphnia* and fish toxicity of (benzo)triazoles: Validated QSAR models, and interspecies quantitative activity-activity modelling. *Journal of Hazardous Materials,* 258**,** 50-60.

CHANG, W. X., NIE, J. Y., YAN, Z., WANG, Y. J. & FAROOQ, S. 2019. Systemic stereoselectivity study of etoxazole: Stereoselective bioactivity, acute toxicity, and environmental behavior in fruits and soils. *Journal of Agricultural and Food Chemistry,* 67**,** 6708-6715.

CHEMAXON 2018. MarvinView (version 16.10.24) , chemical structure viewer module by ChemAxon.

CHEN, C. Y., KO, C. W. & LEE, P. I. 2007. Toxicity of substituted anilines to *Pseudokirchneriella subcapitata* and quantitative structure-activity relationship analysis for polar narcotics. *Environmental Toxicology and Chemistry,* 26**,** 1158-1164.

CHEN, C. Y., KUO, K. L. & FAN, J. W. 2012. Toxicity of propargylic alcohols on green alga-*Pseudokirchneriella subcapitata*. *Journal of Environmental Monitoring,* 14**,** 181-186.

CHEN, C. Y., WANG, Y. J. & YANG, C. F. 2009. Estimating low-toxic-effect concentrations in closed-system algal toxicity tests. *Ecotoxicology and Environmental Safety,* 72**,** 1514-1522.

CHEN, Z. W., BRACAMONTES, J. R., BUDELIER, M. M., GERMANN, A. L., SHIN, D. J., KATHIRESAN, K., QIAN, M. X., MANION, B., CHENG, W. W. L., REICHERT, D. E., AKK, G., COVEY, D. F. & EVERS, A. S. 2019. Multiple functional neurosteroid binding sites on GABA(A) receptors. *PLOS Biology,* 17**,** 1-27.

CLAESSON, A. & MINIDIS, A. 2018. Systematic approach to organizing structural alerts for reactive metabolite formation from potential drugs. *Chemical Research in Toxicology,* 31**,** 389-411.

CLEMENTS, R. G., NABHOLZ, J. V., ZEEMAN, M. G. & AUER, C. M. 1995. The application of structure-activity relationships (SARs) in the aquatic toxicity evaluation of discrete organic chemicals. *SAR and QSAR in Environmental Research,* 3**,** 203-15.

CONOLLY, R. B., ANKLEY, G. T., CHENG, W. Y., MAYO, M. L., MILLER, D. H., PERKINS, E. J., VILLENEUVE, D. L. & WATANABE, K. H. 2017. Quantitative Adverse Outcome Pathways and their application to predictive Toxicology. *Environmental Science & Technology,* 51**,** 4661-4672.

CRONIN, M. & YOON, M. 2019. Chapter 5.3 - Computational Methods to Predict Toxicity. *In:* BALLS, M., COMBES, R. & WORTH, A. (eds.) *The History of Alternative Test Methods in Toxicology.* Elsevier. 287-300.

CRONIN, M. T. D., MADDEN, J. C., YANG, C. & WORTH, A. P. 2019. Unlocking the potential of *in silico* chemical safety assessment – A report on a cross-sector symposium on current opportunities and future challenges. *Computational Toxicology,* 10**,** 38-43.

CRONIN, M. T. D., ZHAO, Y. H. & YU, R. L. 2000. pH-dependence and QSAR analysis of the toxicity of phenols and anilines to *Daphnia magna*. *Environmental Toxicology,* 15**,** 140-148.

CUI, F., CHAI, T. T., LIU, X. X. & WANG, C. J. 2017. Toxicity of three strobilurins (kresoxim-methyl, pyraclostrobin, and trifloxystrobin) on *Daphnia magna*. *Environmental Toxicology and Chemistry,* 36**,** 182-189.

DARR, D. & FRIDOVICH, I. 1986. Irreversible inactivation of catalase by 3-amino-1,2,4-triazole. *Biochemical Pharmacology,* 35**,** 3642-3642.

DAS, D. N., PANDA, P. K., SINHA, N., MUKHOPADHYAY, S., NAIK, P. P. & BHUTIA, S. K. 2017. DNA damage by 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced p53-mediated apoptosis through activation of cytochrome P450/aryl hydrocarbon receptor. *Environmental Toxicology and Pharmacology,* 55**,** 175-185.

DAS, R. N. & ROY, K. 2014. Predictive modeling studies for the ecotoxicity of ionic liquids towards the green algae *Scenedesmus vacuolatus*. *Chemosphere,* 104**,** 170-176.

DASSAULT SYSTEMES BIOVIA, Accelrys Draw, v.4.1; Accelrys, San Diego: Dassault Systemes, 201

DAVIES, J., WARD, R. S., HODGES, G. & ROBERTS, D. W. 2004. Quantitative Structure-Activity Relationship modeling of acute toxicity of quaternary alkylammonium sulfobetaines to *Daphnia magna*. *Environmental Toxicology and Chemistry,* 23**,** 2111-2115.

DAVIES, T. G., FIELD, L. M., USHERWOOD, P. N. & WILLIAMSON, M. S. 2007. DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life,* 59**,** 151-62.

DEARDEN, J., CRONIN, M. & LAPPIN, D. 2007. A comparison of commercially available software for the prediction of pKa. *Journal of Pharmacy and Pharmacology*. 59, 16.

DEARDEN, J. C. 2002. Prediction of environmental toxicity and fate using quantitative structure-activity relationships (QSARs). *Journal of the Brazilian Chemical Society,* 13**,** 754-762.

DEARDEN, J. C., CRONIN, M. T. D. & KAISER, K. L. E. 2010. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research,* 20**,** 241-266.

DENISON, M. S. & VELLA, L. M. 1990. The hepatic Ah receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin: species differences in subunit dissociation. *Arch Biochem Biophys,* 277**,** 382-388.

DEVELOPMENT(ACD)/LABS, A. C. 2017. ACD/Percepta Platforms v.2017.1.

DI GUILO, R. T. & HINTON, D. E. 2008. *The Toxicology of Fishes*, CRC Press. Boca Raton, FL.

DIAS, E., COSTA, F. G. E., MORAIS, S. & PEREIRA, M. D. 2015. A review on the assessment of the potential adverse health impacts of carbamate pesticides. *Topics in Public Health***,** 197-212.

DIMITROV, S. & MEKENYAN, O. 2010. An Introduction to Read-Across for the Prediction of the Effects of Chemicals. *In*: CRONIN, M. T. D. & MADDEN, J. (eds.) *In Silico Toxicology, Principles and Applications.* The Royal Society of Chemistry. 372-384.

DOM, N., KNAPEN, D., BENOOT, D., NOBELS, I. & BLUST, R. 2010. Aquatic multi-species acute toxicity of (chlorinated) anilines: Experimental versus predicted data. *Chemosphere,* 81**,** 177-186.

DORVAL, J. & HONTELA, A. 2003. Role of glutathione redox cycle and catalase in defense against oxidative stress induced by endosulfan in adrenocortical cells of rainbow trout (*Oncorhynchus mykiss*). *Toxicol Appl Pharmacol,* 192**,** 191-200.

DOUIRI, S., BAHDOUDI, S., HAMDI, Y., CUBÌ, R., BASILLE, M., FOURNIER, A., VAUDRY, H., TONON, M. C., AMRI, M., VAUDRY, D. & MASMOUDI-KOUKI, O. 2016. Involvement of endogenous antioxidant systems in the protective activity of pituitary adenylate cyclase-activating polypeptide against hydrogen peroxide-induced oxidative damages in cultured rat astrocytes. *Journal of Neurochemistry,* 137**,** 913-930.

DRUMMOND, R. A., RUSSOM, C. L., GEIGER, D. L. & DEFOE, D. L. 1986. Behavioral and morphological changes in fathead minnow (*Pimephales promelas*) as diagnostic endpoints for screening chemicals according to mode of action. *In:* POSTON, T. M. & PURDY, R. (eds.) *Aquatic Toxicology and Environmental Fate, Ninth Volume.* Philadelphia, PA. pp. 415-435.

EBBRELL, D. J., MADDEN, J. C., CRONIN, M. T., SCHULTZ, T. W. & ENOCH, S. J. 2016. Development of a fragment-based *in silico* profiler for michael addition thiol reactivity. *Chemical Research in Toxicology,* 29**,** 1073-1081.

ECETOC. 2005. Alternative Testing Approaches in Environmental Safety Assessment, Technical report No. 97 (http://www.ecetoc.org/wp-content/uploads/2014/08/ECETOC-TR-097.pdf). Brussels.

ECHA. 2008. Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals. Available from: http://guidance.echa.europa.eu/docs/guidance_document/. Helsinki.

ECHA. 2011a. Guidance on information requirements and chemical safety assessment Chapter R.4: Evaluation of available information, ECHA-2011-G-13-EN (https://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf/d6395ad2-1596-4708-ba86-0136686d205e) Helsinki.

ECHA. 2011b. The Use of Alternatives to Testing on Animals for the REACH Regulation (https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2011_en.pdf/9b0f7e93-4d61-401d-ba2c-80b3b9faaf66). Helsinki.

ECHA. 2014. The Use of Alternatives to Testing on Animals for the REACH Regulation, Second report under Article 117(3) of the REACH Regulation (https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2014_en.pdf/587d000c-688e-4cdd-9f59-f7d7aacc677b) Helsinki.

ECHA. 2016. Evaluation under REACH Progress Report 2015, ECHA-15-R-20-EN. Helsinki.

ECHA. 2017a. Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7c: Endpoint specific guidance, Version 3.0 (https://echa.europa.eu/documents/10162/13632/information_requirements_r7c_en.pdf/e2e23a98-adb2-4573-b450-cc0dfa7988e5). Helsinki.

ECHA. 2017b. Read-Across Assessment Framework (RAAF) (https://echa.europa.eu/documents/10162/13628/raaf_en.pdf). Helsinki.

ECHA. 2017c. The use of alternatives to testing on animals for the REACH regulation, Third report, ECHA-17-R-02-EN, (https://echa.europa.eu/documents/10162/13639/alternatives_test_animals_2017_en.pdf). Helsinki.

ECHA. 2017d. Guidance on information requirements and chemical safety assessment, Appendix R7-1 for nanomaterials applicable to Chapter R7b - Endpoint specific guidance.

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.

ELLISON, C. M., MADDEN, J. C., CRONIN, M. T. D. & ENOCH, S. J. 2015. Investigation of the Verhaar scheme for predicting acute aquatic toxicity: Improving predictions obtained from Toxtree ver. 2.6. *Chemosphere,* 139**,** 146-154.

ELMORE, S. E. & LA MERRILL, M. A. 2019. Oxidative Phosphorylation Impairment by DDT and DDE. *Frontiers in Endocrinology,* 10**,** 8.

ENOCH, S. J. 2010. Chemical Category Formation and Read-Across for the Prediction of Toxicity. In: Puzyn T., Leszczynski J., Cronin M. (eds) Recent Advances in QSAR Studies. Challenges and Advances in Computational Chemistry and Physics, vol 8. Springer, Dordrecht. pp. 209-219.

ENOCH, S. J. & CRONIN, M. T. D. 2010. A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Critical Reviews in Toxicology,* 40**,** 728-748.

ENOCH, S. J., ELLISON, C. M., SCHULTZ, T. W. & CRONIN, M. T. D. 2011. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology,* 41**,** 783-802.

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.

ERTURK, M. D., SACAN, M. T., NOVIC, M. & MINOVSKI, N. 2012. Quantitative structure-activity relationships (QSARs) using the novel marine algal toxicity data of phenols. *Journal of Molecular Graphics & Modelling,* 38**,** 90-100.

ESCHER, B. I., ASHAUER, R., DYER, S., HERMENS, J. L., LEE, J. H., LESLIE, H. A., MAYER, P., MEADOR, J. P. & WARNE, M. S. 2011. Crucial role of mechanisms and modes of toxic action for understanding tissue residue toxicity and internal effect concentrations of organic chemicals. *Integrated Environmental Assessment and Management,* 7**,** 28-49.

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.

ESCUDER-GILABERT, L., MARTIN-BIOSCA, Y., SAGRADO, S., VILLANUEVA-CAMANAS, R. M. & MEDINA-HERNANDEZ, M. J. 2001. Biopartitioning micellar chromatography to predict ecotoxicity. *Analytica Chimica Acta,* 448**,** 173-185.

ESPERANZA, M., CID, A., HERRERO, C. & RIOBOO, C. 2015. Acute effects of a prooxidant herbicide on the microalga *Chlamydomonas reinhardtii*: Screening cytotoxicity and genotoxicity endpoints. *Aquatic Toxicology,* 165**,** 210-221.

EUROPEAN COMMISSION 2006. Regulation (EC) No 1907/2006 of the European parliament and of the council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *In:* COMMISSION, E. (ed.) *No 1907/2006.*

FARMEN, E., OLSVIK, P. A., BERNTSSEN, M. H. G., HYLLAND, K. & TOLLEFSEN, K. E. 2010. Oxidative stress responses in rainbow trout (*Oncorhynchus mykiss*) hepatocytes exposed to pro-oxidants and a complex environmental sample. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology,* 151**,** 431-438.

FEDAK, K. M., BERNAL, A., CAPSHAW, Z. A. & GROSS, S. 2015. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerging Themes in Epidemiology,* 12**,** 14.

FESWICK, A., LOUGHERY, J. R., ISAACS, M. A., MUNKITTRICK, K. R. & MARTYNIUK, C. J. 2016. Molecular initiating events of the intersex phenotype: Low-dose exposure to 17alpha-ethinylestradiol rapidly regulates molecular networks associated with gonad differentiation in the adult fathead minnow testis. *Aquatic Toxicology,* 181**,** 46-56.

FURUHAMA, A., HASUNUMA, K. & AOKI, Y. 2015a. Interspecies quantitative structure-activity relationships (QSARs) for eco-toxicity screening of chemicals: the role of physicochemical properties. *SAR and QSAR in Environmental Research,* 26**,** 809-830.

FURUHAMA, A., HASUNUMA, K. & AOKI, Y. 2015b. Interspecies quantitative structure-activity-activity relationships (QSAARs) for prediction of acute aquatic toxicity of aromatic amines and phenols. *SAR and QSAR in Environmental Research,* 26**,** 301-323.

FURUSJO, E., SVENSON, A., RAHMBERG, M. & ANDERSSON, M. 2006. The importance of outlier detection and training set selection for reliable environmental QSAR predictions. *Chemosphere,* 63**,** 99-108.

FUSSELL, K. C., UDASIN, R. G., GRAY, J. P., MISHIN, V., SMITH, P. J. S., HECK, D. E. & LASKIN, J. D. 2011. Redox cycling and increased oxygen utilization contribute to diquat-induced oxidative stress and cytotoxicity in Chinese hamster ovary cells overexpressing NADPH-cytochrome P450 reductase. *Free Radical Biology and Medicine,* 50**,** 874-882.

GAO, A. H., FU, Y. Y., ZHANG, K. Z., ZHANG, M., JIANG, H. W., FAN, L. X., NAN, F. J., YUAN, C. G., LI, J., ZHOU, Y. B. & LI, J. Y. 2014. Azoxystrobin, a mitochondrial complex III Qo site inhibitor, exerts beneficial metabolic effects *in vivo* and *in vitro*. *Biochimica et Biophysica Acta,* 1840**,** 2212-21.

GARZA-LEON, C. V., ARZATE-CARDENAS, M. A. & RICO-MARTINEZ, R. 2017. Toxicity evaluation of cypermethrin, glyphosate, and malathion, on two indigenous zooplanktonic species. *Environmental Science and Pollution Research,* 24**,** 18123-18134.

GASPAR-RAMIREZ, O., PEREZ-VAZQUEZ, F. J., SALGADO-BUSTAMANTE, M., GONZALEZ-AMARO, R., HERNANDEZ-CASTRO, B. & PEREZ-MALDONADO, I. N. 2015. DDE and PCB 153 independently induce aryl hydrocarbon receptor (AhR) expression in peripheral blood mononuclear cells. *Journal of Immunotoxicology,* 12**,** 266-272.

GECHEV, T., MEHTEROV, N., DENEV, I. & HILLE, J. 2013. A simple and powerful approach for isolation of *Arabidopsis* mutants with increased tolerance to H2O2-induced cell death. *Methods in Enzymology,* 527**,** 203-20.

GOOCH, J. W., ELSKUS, A. A., KLOEPPER-SAMS, P. J., HAHN, M. E. & STEGEMAN, J. J. 1989. Effects of ortho- and non-ortho-substituted polychlorinated biphenyl congeners on the hepatic monooxygenase system in scup (*Stenotomus chrysops*). *Toxicology and Applied Pharmacology,* 98**,** 422-433.

GROSBERG, R. K., VERMEIJ, G. J. & WAINWRIGHT, P. C. 2012. Biodiversity in water and on land. *Current Biology,* 22**,** R900-R903.

GROSS, M., GREEN, R. M., WELTJE, L. & WHEELER, J. R. 2017. Weight of evidence approaches for the identification of endocrine disrupting properties of chemicals: Review and recommendations for EU regulatory application. *Regulatory Toxicology and Pharmacology,* 91**,** 20-28.

GUST, K. A., COLLIER, Z. A., MAYO, M. L., STANLEY, J. K., GONG, P. & CHAPPELL, M. A. 2016. Limitations of toxicity characterization in life cycle assessment: Can adverse outcome pathways provide a new foundation? *Integrated Environmental Assessment and Management,* 12**,** 580-590.

GUTSELL, S. & RUSSELL, P. 2013. The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment. *Toxicology Research,* 2**,** 299-307.

GÜTLEIN, M., KARWATH, A. & KRAMER, S. 2012. CheS-Mapper - Chemical Space Mapping and Visualization in 3D. *Journal of Cheminformatics,* 4**,** 7.

GÜTLEIN, M., KARWATH, A. & KRAMER, S. 2014. CheS-Mapper 2.0 for visual validation of (Q)SAR models. *Journal of Cheminformatics,* 6**,** 41.

HA, H., PARK, K., KANG, G. & LEE, S. 2019. QSAR study using acute toxicity of *Daphnia magna* and *Hyalella azteca* through exposure to polycyclic aromatic hydrocarbons (PAHs). *Ecotoxicology,* 28**,** 333-342.

HARRIS, R. A., TRUDELL, J. R. & MIHIC, S. J. 2008. Ethanol's molecular targets. *Science Signaling,* 1**,** 7.

HEISE, T., SCHMIDT, F., KNEBEL, C., RIEKE, S., HAIDER, W., GEBUREK, I., NIEMANN, L. & MARX-STOELTING, P. 2018. Hepatotoxic combination effects of three azole fungicides in a broad dose range. *Archives of Toxicology,* 92**,** 859-872.

HERMENS, J. L. 1990. Electrophiles and acute toxicity to fish. *Environmental Health Perspectives,* 87**,** 219-225.

HILL, A. B. 1965. The Environment and Disease: Association or Causation?. *Proceedings of the Royal Society of Medicine,* 58**,** 295-300.

BIOVIA. 2019. Minitab 19. Available from (https://www.minitab.com/en-us/).

ISO 2016. ISO 11348-3:2007 - Water quality -- Determination of the inhibitory effect of water samples on the light emission of *Vibrio fischeri* (Luminescent bacteria test) -- Part 3: Method using freeze-dried bacteria.

JIA, Q., ZHAO, Y., YAN, F. & WANG, Q. 2018. QSAR model for predicting the toxicity of organic compounds to fathead minnow. *Environmental Science and Pollution Research,* 25**,** 35420-35428.

JIANG, L., LIU, G. J. & LI, Q. 2016. Binding mechanism analysis between AhR protein and typical polybrominated diphenyl ether via molecule docking. *Proceedings of the 2016 International Conference on Artificial Intelligence and Engineering Applications,* 63**,** 464-467.

JRC COMPUTATIONAL TOXICOLOGY 2018. Toxtree (https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxtree, accessed in 2018).

KAHN, I., MARAN, U., BENFENATI, E., NETZEVA, T. I., SCHULTZ, T. W. & CRONINS, M. T. D. 2007. Comparative quantitative structure-activity-activity relationships for toxicity to *Tetrahymena pyriformis* and *Pimephales promelas*. *ATLA-Alternatives to Laboratory Animals,* 35**,** 15-24.

KASE, R., KORKARIC, M., WERNER, I. & AGERSTRAND, M. 2016. Criteria for Reporting and Evaluating ecotoxicity Data (CRED): comparison and perception of the Klimisch and CRED methods for evaluating reliability and relevance of ecotoxicity studies. *Environmental Sciences Europe,* 28**,** 7.

KATRITZKY, A. R., TATHAM, D. B. & MARAN, U. 2001. Theoretical descriptors for the correlation of aquatic toxicity of environmental pollutants by quantitative structure-toxicity relationships. *Journal of Chemical Information and Computer Sciences,* 41**,** 1162-1176.

KIENZLER, A., BARRON, M. G., BELANGER, S. E., BEASLEY, A. & EMBRY, M. R. 2017a. Mode of Action (MOA) assignment classifications for ecotoxicology: An evaluation of approaches. *Environmental Science and Technology* 51, 10203-10211.

KIENZLER, A., HALDER, M. & WORTH, A. 2017b. Waiving chronic fish tests: possible use of acute-to-chronic relationships and interspecies correlations. *Toxicological & Environmental Chemistry,* 99**,** 1129-1151.

KIM, B. M., RHEE, J. S., HWANG, U. K., SEO, J. S., SHIN, K. H. & LEE, J. S. 2015. Dose- and time-dependent expression of aryl hydrocarbon receptor (AhR) and aryl hydrocarbon receptor nuclear translocator (ARNT) in PCB-, B[a]P-, and TBT-exposed intertidal copepod *Tigriopus japonicus*. *Chemosphere,* 120**,** 398-406.

KIM, S., THIESSEN, P. A., BOLTON, E. E., CHEN, J., FU, G., GINDULYTE, A., HAN, L., HE, J., HE, S., SHOEMAKER, B. A., WANG, J., YU, B., ZHANG, J. & BRYANT, S. H. 2016. PubChem substance and compound databases. *Nucleic Acids Research,* 44**,** D1202-D1213.

KLIMISCH, H. J., ANDREAE, M. & TILLMANN, U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology,* 25**,** 1-5.

KLUVER, N., VOGS, C., ALTENBURGER, R., ESCHER, B. I. & SCHOLZ, S. 2016. Development of a general baseline toxicity QSAR model for the fish embryo acute toxicity test. *Chemosphere,* 164**,** 164-173.

KNAPEN, D., ANGRISH, M. M., FORTIN, M. C., KATSIADAKI, I., LEONARD, M., MARGIOTTA-CASALUCI, L., MUNN, S., O'BRIEN, J. M., POLLESCH, N., SMITH, L. C., ZHANG, X. & VILLENEUVE, D. L. 2018. Adverse Outcome Pathway networks I: Development and applications. *Environmental Toxicology and Chemistry,* 37**,** 1723-1733.

KÖNEMANN, H. 1981. Quantitative Structure-Activity-Relationships in fish toxicity studies, 1. Relationship for 50 industrial pollutants. *Toxicology,* 19**,** 209-221.

KOUJI, H., MASUDA, T. & MATSUNAKA, S. 1989. Action mechanism of diphenyl ether herbicides – stimulation of 5-aminolevulinic acid-synthesizing system activities. *Pesticide Biochemistry and Physiology,* 33**,** 230-238.

KUIPER, R. V., BERGMAN, A., VOS, J. G. & VAN DEN BERG, M. 2004. Some polybrominated diphenyl ether (PBDE) flame retardants with wide environmental distribution inhibit TCDD-induced EROD activity in primary cultured carp (*Cyprinus carpio*) hepatocytes. *Aquatic Toxicology,* 68**,** 129-39.

LALONE, C. A., VILLENEUVE, D. L., BURGOON, L. D., RUSSOM, C. L., HELGEN, H. W., BERNINGER, J. P., TIETGE, J. E., SEVERSON, M. N., CAVALLIN, J. E. & ANKLEY, G. T. 2013. Molecular target sequence similarity as a basis for species extrapolation to assess the ecological risk of chemicals with known modes of action. *Aquatic Toxicology,* 144-145**,** 141-54.

LALONE, C. A., VILLENEUVE, D. L., CAVALLIN, J. E., KAHL, M. D., DURHAN, E. J., MAKYNEN, E. A., JENSEN, K. M., STEVENS, K. E., SEVERSON, M. N., BLANKSMA, C. A., FLYNN, K. M., HARTIG, P. C., WOODARD, J. S., BERNINGER, J. P., NORBERG-KING, T. J., JOHNSON, R. D. & ANKLEY, G. T. 2013. Cross-species sensitivity to a novel androgen receptor agonist of potential environmental concern, spironolactone. *Environmental Toxicology and Chemistry,* 32**,** 2528-2541.

LALONE, C. A., VILLENEUVE, D. L., LYONS, D., HELGEN, H. W., ROBINSON, S. L., SWINTEK, J. A., SAARI, T. W. & ANKLEY, G. T. 2016. Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A web-based tool for addressing the challenges of cross-species extrapolation of chemical toxicity. *Toxicological Sciences,* 153**,** 228-245.

LAZARTIGUES, A., FRATTA, C., BAUDOT, R., WIEST, L., FEIDT, C., THOMAS, M. & CREN-OLIVE, C. 2011. Multiresidue method for the determination of 13 pesticides in three environmental matrices: water, sediments and fish muscle. *Talanta,* 85**,** 1500-1507.

LEE, J. N., DUTTA, R. K., MAHARJAN, Y., LIU, Z. Q., LIM, J. Y., KIM, S. J., CHO, D. H., SO, H. S., CHOE, S. K. & PARK, R. 2018. Catalase inhibition induces pexophagy through ROS accumulation. *Biochemical and Biophysical Research Communications,* 501**,** 696-702.

LEE, J. N., KIM, S. G., LIM, J. Y., DUTTA, R. K., KIM, S. J., CHOE, S. K., SO, H. S. & PARK, R. 2016. 3-Aminotriazole protects from CoCl2-induced ototoxicity by inhibiting the generation of reactive oxygen species and proinflammatory cytokines in mice. *Archives of Toxicology,* 90**,** 781-91.

LEE, P. Y. & CHEN, C. Y. 2009. Toxicity and Quantitative Structure-Activity Relationships of benzoic acids to *Pseudokirchneriella subcapitata*. *Journal of Hazardous Materials,* 165**,** 156-161.

LEIJS, M. M., ESSER, A., AMANN, P. M., SCHETTGEN, T., HEISE, R., FIETKAU, K., GUBE, M., MERK, H. F., KRAUS, T. & BARON, J. M. 2018. Expression of CYP1A1, CYP1B1 and IL-1 beta in PBMCs and skin samples of PCB exposed individuals. *Science of the Total Environment,* 642**,** 1429-1438.

LEMAL, D. M. 2004. Perspective on fluorocarbon chemistry. *Journal of Organic Chemistry,* 69**,** 1-11.
LEPAILLEUR, A., POEZEVARA, G. & BUREAU, R. 2013. Automated detection of structural alerts (chemical fragments) in (eco)toxicology. *Computational and Structural Biotechnology Journal,* 5**,** e201302013.
LEPESHEVA, G. I. & WATERMAN, M. R. 2007. Sterol 14α-demethylase Cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochimica et Biophysica Acta,* 1770**,** 467-77.
LEWIS, D. F. V. & LAKE, B. G. 1996. Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: Rationalization of CYP1A substrate specificity in terms of active site amino acid residues. *Xenobiotica,* 26**,** 723-753.
LI, J. J., TAI, H. W., YU, Y., WEN, Y., WANG, X. H. & ZHAO, Y. H. 2015. Comparison of toxicity of class-based organic chemicals to algae and fish based on discrimination of excess toxicity from baseline level. *Environmental Toxicology and Pharmacology,* 40**,** 292-299.
LI, X., LI, R., ZHU, B., GAO, X. & LIANG, P. 2018. Overexpression of cytochrome P450 CYP6BG1 may contribute to chlorantraniliprole resistance in *Plutella xylostella* (L.). *Pest Management Science,* 74**,** 1386-1393.
LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* 23**,** 3-25.
LIPNICK, R. L. 1991. Outliers - their origin and use in the classification of molecular mechanisms of toxicity. *Science of the Total Environment,* 109**,** 131-153.
LIU, L., ZHU, B. & WANG, G. X. 2015. Azoxystrobin-induced excessive reactive oxygen species (ROS) production and inhibition of photosynthesis in the unicellular green algae *Chlorella vulgaris*. *Environmental Science and Pollution Research,* 22**,** 7766-7775.
LU, G. H., YUAN, X. & ZHAO, Y. H. 2001. QSAR study on the toxicity of substituted benzenes to the algae (*Scenedesmus obliquus*). *Chemosphere,* 44**,** 437-440.
LUCIER, G. W., PORTIER, C. J. & GALLO, M. A. 1993. Receptor mechanisms and dose-response models for the effects of dioxins. *Environmental Health Perspectives,* 101**,** 36-44.
LUSHCHAK, V. I. 2016. Contaminant-induced oxidative stress in fish: a mechanistic approach. *Fish Physiology and Biochemistry,* 42**,** 711-747.
MA, J. G., LI, Y. Y., LI, W. G. & LI, X. Y. 2018. Hepatotoxicity of paraquat on common carp (*Cyprinus carpio*). *Science of the Total Environment,* 616**,** 889-898.
MADDEN, J. 2010. Introduction to QSAR and other *in silico* methods to predict toxicology. *In:* CRONIN, M. & MADDEN, J. (eds.) *In Silico Toxicology: Principles and Applications.* The Royal Society of Chemistry. pp. 11-30.
MAJEWSKA, M., HARSHKOVA, D., GUSCIORA, M. & AKSMANN, A. 2018. Phytotoxic activity of diclofenac: Evaluation using a model green alga *Chlamydomonas reinhardtii* with atrazine as a reference substance. *Chemosphere,* 209**,** 989-997.
MARCHLER-BAUER, A., DERBYSHIRE, M. K., GONZALES, N. R., LU, S., CHITSAZ, F., GEER, L. Y., GEER, R. C., HE, J., GWADZ, M., HURWITZ, D. I., LANCZYCKI, C. J., LU, F., MARCHLER, G. H., SONG, J. S., THANKI, N., WANG, Z., YAMASHITA, R. A., ZHANG, D., ZHENG, C. & BRYANT, S. H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research,* 43**,** D222-D226.
MARTIN, T. M., GRULKE, C. M., YOUNG, D. M., RUSSOM, C. L., WANG, N. N. Y., JACKSON, C. R. & BARRON, M. G. 2013. Prediction of aquatic toxicity mode of action using linear discriminant and random forest models. *Journal of Chemical Information and Modeling,* 53**,** 2229-2239.
MCCARTY, L. S., ARNOT, J. A. & MACKAY, D. 2013. Evaluation of critical body residue data for acute narcosis in aquatic organisms. *Environmental Toxicology and Chemistry,* 32**,** 2301-2314.
MCKIM, J. M., BRADBURY, S. P. & NIEMI, G. J. 1987. Fish Acute Toxicity Syndromes and their use in the QSAR approach for hazard assessment. *Environmental Health Perspectives,* 71**,** 171-186.

MEEK, M. E., PALERMO, C. M., BACHMAN, A. N., NORTH, C. M. & JEFFREY LEWIS, R. 2014. Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. *Journal of Applied Toxicology,* 34**,** 595-606.

MEKENYAN, O., DIMITROV, S., SERAFIMOVA, R., THOMPSON, E., KOTOV, S., DIMITROVA, N. & WALKER, J. D. 2004. Identification of the structural requirements for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals I: TA100 model. *Chemical Research in Toxicology,* 17**,** 753-766.

MELLOR, C. L., STEINMETZ, F. P. & CRONIN, M. T. D. 2016. Using Molecular Initiating Events to develop a structural alert based screening workflow for nuclear receptor ligands associated with hepatic steatosis. *Chemical Research in Toxicology,* 29**,** 203-212.

MEYER, H. H. 1899. Welche eigenschaft der anasthetica bedingt inre Narkotische wirkung? *Archiv für experimentelle Pathologie und Pharmakologie,* 42**,** 109-118.

MEYER, K. J., REIF, J. S., VEERAMACHANENI, D. N. R., LUBEN, T. J., MOSLEY, B. S. & NUCKOLS, J. R. 2006. Agricultural pesticide use and hypospadias in eastern Arkansas. *Environmental Health Perspectives,* 114**,** 1589-1595.

MILES, J. C., HUA, J., SEPULVEDA, M. S., KRUPKE, C. H. & HOVERMAN, J. T. 2017. Effects of clothianidin on aquatic communities: Evaluating the impacts of lethal and sublethal exposure to neonicotinoids. *PLOS One,* 13**,** e0194634.

MILLER, M. M., MCMULLEN, P. D., ANDERSEN, M. E. & CLEWELL, R. A. 2017. Multiple receptors shape the estrogen response pathway and are critical considerations for the future of *in vitro*-based risk assessment efforts. *Critical Reviews in Toxicology,* 47**,** 564-580.

MIYAZAWA, M., SHINDO, M. & SHIMADA, T. 2001. Oxidation of 1,8-cineole, the monoterpene cyclic ether originated from *Eucalyptus polybractea*, by Cytochrome P450 3A enzymes in rat and human liver microsomes. *Drug Metabolism and Disposition,* 29**,** 200-205.

MOERMOND, C. T., KASE, R., KORKARIC, M. & AGERSTRAND, M. 2016. CRED: Criteria for reporting and evaluating ecotoxicity data. *Environmental Toxicology and Chemistry,* 35**,** 1297-309.

MOHSENZADEH, M. S., ZANJANI, B. R. & KARIMI, G. 2018. Mechanisms of 2,3,7,8-tetrachlorodibenzo-p-dioxin- induced cardiovascular toxicity: An overview. *Chemico-Biological Interactions,* 282**,** 1-6.

MOLANDER, L., ÅGERSTRAND, M., BERONIUS, A., HANBERG, A. & RUDÉN, C. 2014. Science in Risk Assessment and Policy (SciRAP): an online resource for evaluating and reporting *in vivo* (eco)toxicity studies. *Human and Ecological Risk Assessment: An International Journal,* 21, 753-762.

MOOSUS, M. & MARAN, U. 2011. Quantitative structure-activity relationship analysis of acute toxicity of diverse chemicals to *Daphnia magna* with whole molecule descriptors. *SAR and QSAR in Environmental Research,* 22**,** 757-774.

MORTENSEN, A. S. & ARUKWE, A. 2007. Interactions between estrogen- and Ah-Receptor signalling pathways in primary culture of salmon hepatocytes exposed to nonylphenol and 3,3',4,4'-tetrachlorobiphenyl (Congener 77). *Computational Hepatology.* 6, 2.

MURAYAMA, K., ABBOTT, N. J., NARAHASH.T & SHAPIRO, B. I. 1972. Effects of allethrin and condylactis toxin on kinetics of sodium conductance of crayfish axon membranes. *General Pharmacology,* 3**,** 391-400.

NAUEN, R. & SMAGGHE, G. 2006. Rapid report - Mode of action of etoxazole. *Pest Management Science,* 62**,** 379-382.

NCBI, N. C. F. B. I. 2018a. *Conserved Domains Database (CDD) and Resources* [Online]. Available: https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml [Accessed April 2018].

NCBI, N. C. F. B. I. 2018b. Home - HomoloGene (https://www.ncbi.nlm.nih.gov/homologene/).

NELMS, M. D. 2014. Development of *in silico* models to support repeat dose safety assessment of cosmetic ingredients to humans. Doctoral Thesis, Liverpool John Moores University.
NELMS, M. D., CRONIN, M. T. D., SCHULTZ, T. W. & ENOCH, S. J. 2013. Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones. *SAR and QSAR in Environmental Research,* 24**,** 695-709.
NELMS, M. D., MELLOR, C. L., CRONIN, M. T., MADDEN, J. C. & ENOCH, S. J. 2015. Development of an *in silico* profiler for mitochondrial toxicity. *Chemical Research in Toxicology,* 28**,** 1891-1902.
NENDZA, M., MULLER, M. & WENZEL, A. 2014. Discriminating toxicant classes by mode of action: 4. Baseline and excess toxicity. *SAR and QSAR in Environmental Research,* 25**,** 393-405.
NESNOW, S., ARGUS, M., BERGMAN, H., CHU, K., FRITH, C., HELMES, T., MCGAUGHY, R., RAY, V., SLAGA, T. J., TENNANT, R. & E., W. 1987. Chemical carcinogens: a review and analysis of the literature of selected chemicals and the establishment of the Gene-Tox Carcinogen Data Base. *Mutation Research,* 185**,** 1-195.
NETZEVA, T. I., DEARDEN, J. C., EDWARDS, R., WORGAN, A. D. P. & CRONIN, M. T. D. 2004. QSAR analysis of the toxicity of aromatic compounds to *Chlorella vulgaris* in a novel short-term assay. *Journal of Chemical Information and Computer Sciences,* 44**,** 258-265.
NETZEVA, T. I., WORTH, A. P., ALDENBERG, T., BENIGNI, R., CRONIN, M. T. D., GRAMATICA, P., JAWORSKA, J. S., KAHN, S., KLOPMAN, G., MARCHANT, C. A., MYATT, G., NIKOLOVA-JELIAZKOVA, N., PATLEWICZ, G. Y., PERKINS, R., ROBERTS, D. W., SCHULTZ, T. W., STANTON, D. T., VAN DE SANDT, J. J. M., TONG, W. D., VEITH, G. & YANG, C. H. 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM Workshop 52. *ATLA-Alternatives to Laboratory Animals,* 33**,** 155-173.
NEUWOEHNER, J., FENNER, K. & ESCHER, B. I. 2009. Physiological modes of action of fluoxetine and its human metabolites in algae. *Environmental Science & Technology,* 43**,** 6830-6837.
NICOLOTTI, O., BENFENATI, E., CAROTTI, A., GADALETA, D., GISSI, A., MANGIATORDI, G. F. & NOVELLINO, E. 2014. REACH and *in silico* methods: an attractive opportunity for medicinal chemists. *Drug Discovery Today,* 19**,** 1757-1768.
O'HAGAN, D. 2008. Understanding organofluorine chemistry. An introduction to the C-F bond. *Chemical Society Reviews,* 37**,** 308-319.
OECD. 2008. Guidance on Grouping of Chemicals. Available from http://www.oecd.org/LongAbstract/0,3425,en_2649_34379_39404999_119829_1_1_37407,00.html).
OECD. 2010. Report of the expert consultation on scientific and regulatory evaluation of organic chemistry mechanism-based structural alerts for the identification of DNA binding chemicals (http://www.oecd.org/ dataoecd /10/59/45401393.pdf) Paris.
OECD. 2011. Report of the expert consultation on scientific and regulatory evaluation of organic chemistry mechanism-based structural alerts for the identification of protein binding chemicals (http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2011)9&doclanguage=en). Paris: OECD.
OECD. 2013a. Guidance document on developing and assessing adverse outcome pathways. Available from: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282013%296&doclanguage=en .
OECD. 2013b. QSAR Toolbox, User Manual, Strategies for grouping chemicals to fill data gaps to assess acute aquatic toxicity endpoints. Available from: https://www.oecd.org/chemicalsafety/Acute%20aquatic%20toxicity.pdf

OECD. 2014. Guidance on grouping of chemicals (http://www.oecd.org/chemicalsafety/risk-assessment/chapter3dataevaluation.htm). 2nd edition.
OECD. 2016a. OECD Guidelines for the Testing of Chemicals, Section 2, Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Available from: https://www.oecd-ilibrary.org/environment/test-no-201-alga-growth-inhibition-test_9789264069923-en
OECD. 2016b. OECD Guidelines for the Testing of Chemicals, Section 2, Test No. 202: Daphnia sp. Acute Immobilisation Test. Available from: https://www.oecd-ilibrary.org/environment/test-no-202-daphnia-sp-acute-immobilisation-test_9789264069947-en
OECD. 2016c. OECD Guidelines for the Testing of Chemicals, Section 2, Test No. 203: Fish, Acute Toxicity Test. Available from: https://www.oecd-ilibrary.org/environment/test-no-203-fish-acute-toxicity-test_9789264069961-en
OECD. 2018a. QSAR Toolbox. 4.2. Available from: https://qsartoolbox.org/
OECD, 2018b. User's handbook supplement to the guidance document for developing and assessing AOPs. Available from: https://doi.org/10.1787/2415170X
OLANIRAN, A. O. & IGBINOSA, E. O. 2011. Chlorophenols and other related derivatives of environmental concern: Properties, distribution and microbial degradation processes. *Chemosphere,* 83**,** 1297-1306.
OLSVIK, P. A. & SOFTELAND, L. 2018. Metabolic effects of p,p-DDE on Atlantic salmon hepatocytes. *Journal of Applied Toxicology,* 38**,** 489-503.
OVERTON, C. E. 1901. Studien über die Narkose zugleich ein Beitrag zur allgemeinen Pharmakologie. *Gustav Fischer, Jena*.
PANDINI, A., SOSHILOV, A. A., SONG, Y., ZHAO, J., BONATI, L. & DENISON, M. S. 2009. Detection of the TCDD binding-fingerprint within the Ah receptor ligand binding domain by structurally driven mutagenesis and functional analysis. *Biochemistry,* 48**,** 5972-83.
PARK, Y. S., UDDIN, M. J., PIAO, L., HWANG, I., LEE, J. H. & HA, H. 2016. Novel role of endogenous catalase in macrophage polarization in adipose tissue. *Mediators of Inflammation,* 2016**,** 8675905.
PATLEWICZ, G., BALL, N., BECKER, R. A., BOOTH, E. D., CRONIN, M. T. D., KROESE, D., STEUP, D., VAN RAVENZWAAY, B. & HARTUNG, T. 2014. Food for thought ... Read-Across approaches - misconceptions, promises and challenges ahead. *Altex-Alternatives to Animal Experimentation,* 31**,** 387-396.
PAUTUS, S., YEE, S. W., JAYNE, M., COOGAN, M. P. & SIMONS, C. 2006. Synthesis and CYP26A1 inhibitory activity of 1- benzofuran-2-yl-(4-alkyl/aryl-phenyl)-methyl -1H-triazoles. *Bioorganic & Medicinal Chemistry,* 14**,** 3643-3653.
PAVAN, M. & WORTH, A. P. 2008. Publicly-accessible QSAR software tools developed by the Joint Research Centre. *SAR and QSAR in Environmental Research,* 19**,** 785-799.
PETERSON, H. G., BOUTIN, C., FREEMARK, K. E. & MARTIN, P. A. 1997. Toxicity of hexazinone and diquat to green algae, diatoms, cyanobacteria and duckweed. *Aquatic Toxicology,* 39**,** 111-134.
PETERSON, R. E., THEOBALD, H. M. & KIMMEL, G. L. 1993. Developmental and reproductive toxicity of dioxins and related compounds: cross-species comparisons. *Critical Reviews in Toxicology,* 23**,** 283-335.
PEYTON, L., GALLAGHER, S. & HASHEMZADEH, M. 2015. Triazole antifungals: a review. *Drugs of Today* 51, 705-712.
PRETTI, C., CHIAPPE, C., BALDETTI, I., BRUNINI, S., MONNI, G. & INTORRE, L. 2009. Acute toxicity of ionic liquids for three freshwater organisms: *Pseudokirchneriella subcapitata*, *Daphnia magna* and *Danio rerio*. *Ecotoxicology and Environmental Safety,* 72**,** 1170-1176.

PRZYBYLAK, K. R., MADDEN, J. C., CRONIN, M. T. D. & HEWITT, M. 2012. Assessing toxicological data quality: basic principles, existing schemes and current limitations. *SAR and QSAR in Environmental Research,* 23**,** 435-459.

RAIMONDO, S., MONTAGUE, B. J. & BARRON, M. G. 2007. Determinants of variability in acute to chronic toxicity ratios for aquatic invertebrates and fish. *Environmental Toxicology and Chemistry,* 26**,** 2019-2023.

RAINEY, N. E., SARIC, A., LEBERRE, A., DEWAILLY, E., SLOMIANNY, C., VIAL, G., ZELIGER, H. I. & PETIT, P. X. 2017. Synergistic cellular effects including mitochondrial destabilization, autophagy and apoptosis following low-level exposure to a mixture of lipophilic persistent organic pollutants. *Scientific Reports,* 7.

RAMOS, E. U., VAAL, M. A. & HERMENS, J. L. M. 2002. Interspecies sensitivity in the aquatic toxicity of aromatic amines. *Environmental Toxicology and Pharmacology,* 11**,** 149-158.

RAMOS, E. U., VAES, W. H. J., MAYER, P. & HERMENS, J. L. M. 1999. Algal growth inhibition of *Chlorella pyrenoidosa* by polar narcotic pollutants: toxic cell concentrations and QSAR modeling. *Aquatic Toxicology,* 46**,** 1-10.

RAND, G. M. 1995. *Fundamentals of Aquatic Toxicology: Effects, Environmental Fate and Risk Assessment* (2nd Ed.), CRC Press. Boca Raton, FL. p. 1148.

REUSCHENBACH, P., SILVANI, M., DAMMANN, M., WARNECKE, D. & KNACKER, T. 2008. ECOSAR model performance with a large test set of industrial chemicals. *Chemosphere,* 71**,** 1986-1995.

RICHARZ, A. N., SCHULTZ, T. W., CRONIN, M. T. D. & ENOCH, S. J. 2014. Experimental verification of structural alerts for the protein binding of sulfur-containing compounds. *SAR and QSAR in Environmental Research,* 25**,** 325-341.

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

ROBERTS, D. W., ROBERTS, J. F., HODGES, G., GUTSELL, S., WARD, R. S. & LLEWELLYN, C. 2013. Aquatic toxicity of cationic surfactants to *Daphnia magna*. *SAR and QSAR in Environmental Research,* 24**,** 683-693.

ROCCO, M., PELAIA, P., DI BENEDETTO, P., CONTE, G., MAGGI, L., FIORELLI, S., MERCIERI, M., BALESTRA, C., DE BLASI, R. A., SCARATOZZO, D., GALA, F., SCREPANTI, L., DIVING, A., NICOLINI, S., MESA, S. & INVESTIGATORS, R. P. 2019. Inert gas narcosis in scuba diving, different gases different reactions. *European Journal of Applied Physiology,* 119**,** 247-255.

RODRIGUEZ-SANCHEZ, N., SCHULTZ, T. W., CRONIN, M. T. D. & ENOCH, S. J. 2013. Experimental verification of structural alerts for the protein binding of cyclic compounds acting as Michael acceptors. *SAR and QSAR in Environmental Research,* 24**,** 963-977.

ROSTAIN, J. C. & LAVOUTE, C. 2016. Metabolically inert gas narcosis in the central nervous system. *Comprehensive Physiology,* 6**,** 1579-1590.

ROY, K. & DAS, R. N. 2013. QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards *Daphnia magna*. *Journal of Hazardous Materials,* 254**,** 166-178.

RUIZ-LEAL, M. & GEORGE, S. 2004. An *in vitro* procedure for evaluation of early stage oxidative stress in an established fish cell line applied to investigation of PHAH and pesticide toxicity. *Marine Environmental Research,* 58**,** 631-635.

RUSSOM, C. L., ANDERSON, E. B., GREENWOOD, B. E. & PILLI, A. 1991. ASTER - in integration of the aquire database and the QSAR system for use in ecological risk assessments. *Science of the Total Environment,* 109**,** 667-670.

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.

RUUSMANN, V. & MARAN, U. 2013. From data point timelines to a well curated data set, data mining of experimental data and chemical structure data from scientific articles, problems and possible solutions. *Journal of Computer-Aided Molecular Design,* 27**,** 583-603.

SANDERSON, H. & THOMSEN, M. 2009. Comparative analysis of pharmaceuticals versus industrial chemicals acute aquatic toxicity classification according to the United Nations classification system for chemicals. Assessment of the (Q)SAR predictability of pharmaceuticals acute aquatic toxicity and their predominant acute toxic mode-of-action. *Toxicology Letters,* 187**,** 84-93.

SANDERSON, P. N., SIMPSON, W., CUBBERLEY, R., ALEKSIC, M., GUTSELL, S. & RUSSELL, P. J. 2016. Mechanistic understanding of molecular initiating events (MIEs) using NMR spectroscopy. *Toxicology Research,* 5**,** 34-44.

SCHAAR, C. E., DUES, D. J., SPIELBAUER, K. K., MACHIELA, E., COOPER, J. F., SENCHUK, M., HEKIMI, S. & VAN RAAMSDONK, J. M. 2015. Mitochondrial and cytoplasmic ROS have opposing effects on lifespan. *PLOS Genetics,* 11**,** 24.

SCHREIBER, T. H. & PODACK, E. R. 2009. A critical analysis of the tumour immunosurveillance controversy for 3-MCA-induced sarcomas. *British Journal of Cancer,* 101**,** 381-386.

SCHULTZ, T. W. 2008. TETRATOX: *Tetrahymena Pyriformis* population growth impairment endpoint surrogate for fish lethality. *Toxicology Methods,* 7, 289-309.

SCHULTZ, T. W., HOLCOMBE, G. W. & PHIPPS, G. L. 1986. Relationships of quantitative structure-activity to comparative toxicity of selected phenols in the *Pimephales Promelas* and *Tetrahymena Pyriformis* test systems. *Ecotoxicology and Environmental Safety,* 12**,** 146-153.

SCHULTZ, T. W., LIN, D. T. & MICHELLE ARNOLD, L. 1991. QSARs for monosubstituted anilines eliciting the polar narcosis mechanism of action. *Science of The Total Environment,* 109-110**,** 569-580.

SCHULTZ, T. W., YARBROUGH, J. W. & JOHNSON, E. L. 2005. Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *SAR and QSAR in Environmental Research,* 16**,** 313-322.

SERAFIMOVA, R., TODOROV, M., PAVLOV, T., KOTOV, S., JACOB, E., APTULA, A. & MEKENYAN, O. 2007. Identification of the structural requirements for mutagencitiy, by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model. *Chemical Research in Toxicology,* 20**,** 662-676.

SEWALK, C. J., BREWER, G. L. & HOFFMAN, D. J. 2000. Effects of diquat, an aquatic herbicide, on the development of mallard embryos. *Journal of Toxicology and Environmental Health-Part A-Current Issues,* 62**,** 33-45.

SINGH, K. P., GUPTA, S. & BASANT, N. 2014a. Predicting toxicities of ionic liquids in multiple test species - an aid in designing green chemicals. *RSC Advances,* 4**,** 64443-64456.

SINGH, K. P., GUPTA, S., KUMAR, A. & MOHAN, D. 2014b. Multispecies QSAR modeling for predicting the aquatic toxicity of diverse organic chemicalsi for regulatory toxicology. *Chemical Research in Toxicology,* 27**,** 741-753.

SOININEN, J., BARTELS, P., HEINO, J., LUOTO, M. & HILLEBRAND, H. 2015. Toward more integrated ecosystem research in aquatic and terrestrial environments. *BioScience,* 65**,** 174-182.

SONG, Y., VILLENEUVE, D. L., TOYOTA, K., IGUCHI, T. & TOLLEFSEN, K. E. 2017. Ecdysone receptor agonism leading to lethal molting disruption in arthropods: review and Adverse Outcome Pathway development. *Environmental Science & Technology,* 51**,** 4142-4157.

SONG, Z. & POLLENZ, R. S. 2002. Ligand-dependent and independent modulation of aryl hydrocarbon receptor localization, degradation, and gene regulation. *Molecular Pharmacology,* 62**,** 806-816.
SPANAGEL, R. 2009. Alcoholism: a systems approach from molecular physiology to addictive behavior. *Physiological Reviews,* 89**,** 649-705.
SPIELMANN, H., SAUER, U. G. & MEKENYAN, O. 2011. A critical evaluation of the 2011 ECHA reports on compliance with the REACH and CLP regulations and on the use of alternatives to testing on animals for compliance with the REACH regulation. *ATLA-Alternatives to Laboratory Animals,* 39**,** 481-493.
STEINMETZ, F. P., ENOCH, S. J., MADDEN, J. C., NELMS, M. D., RODRIGUEZ-SANCHEZ, N., ROWE, P. H., WEN, Y. & CRONIN, M. T. D. 2014. Methods for assigning confidence to toxicity data with multiple values identifying experimental outliers. *Science of the Total Environment,* 482**,** 358-365.
SUH, J., JEON, Y. J., KIM, H. M., KANG, J. S., KAMINSKI, N. E. & YANG, K. H. 2002. Aryl hydrocarbon receptor-dependent inhibition of AP-1 activity by 2,3,7,8-tetrachlorodibenzo-p-dioxin in activated B cells. *Toxicology and Applied Pharmacology,* 181**,** 116-123.
SUSHKO, I., SALMINA, E., POTEMKIN, V. & TETKO, I. V. 2012a. ToxAlerts: Online database of toxicological structure alerts. *Abstracts of Papers of the American Chemical Society,* 244.
SUSHKO, I., SALMINA, E., POTEMKIN, V. A., PODA, G. & TETKO, I. V. 2012b. ToxAlerts: A web server of structural alerts for toxfic chemicals and compounds with potential adverse reactions. *Journal of Chemical Information and Modeling,* 52**,** 2310-2316.
TENG, M., ZHU, W., WANG, D., QI, S., WANG, Y., YAN, J., DONG, K., ZHENG, M. & WANG, C. 2018. Metabolomics and transcriptomics reveal the toxicity of difenoconazole to the early life stages of zebrafish (*Danio rerio*). *Aquatic Toxicology,* 194**,** 112-120.
TOUMI, H., BOUMAIZA, M., MILLET, M., RADETSKI, C. M., CAMARA, B. I., FELTEN, V. & FERARD, J. F. 2015. Investigation of differences in sensitivity between 3 strains of *Daphnia magna* (crustacean Cladocera) exposed to malathion (organophosphorous pesticide). *Journal of Environmental Science and Health Part B-Pesticides Food Contaminants and Agricultural Wastes,* 50**,** 34-44.
TRAC, L. N., ANDERSEN, O. & PALMQVIST, A. 2016. Deciphering mechanisms of malathion toxicity under pulse exposure of the freshwater cladoceran *Daphnia magna*. *Environmental Toxicology and Chemistry,* 35**,** 394-404.
TSAI, W.-T. 2013. A review on environmental exposure and health risks of herbicide paraquat. *Toxicological and Environmental Chemistry,* 95**,** 197-206.
UNIVERSITY OF HERTFORDSHIRE 2018. PPDB: Pesticide Properties DataBase. Available from: https://sitem.herts.ac.uk/aeru/ppdb/
US ENVIRONMENTAL PROTECTION AGENCY. 2008. EPA Fathead Minnow Acute Toxicity. Distributed Structure-Searchable Toxicity (DSSTox) Database.
US ENVIRONMENTAL PROTECTION AGENCY. 2016. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.
US ENVIRONMENTAL PROTECTION AGENCY. 2017. Ecological Structure Activity Relationships, ECOSAR (version 2) https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model.
US ENVIRONMENTAL PROTECTION AGENCY. 2016. *ECOTOX.* Available from: https://cfpub.epa.gov/ecotox/ [Accessed April 2016].
VAN DEN BRINK, P. J., BOXALL, A. B. A., MALTBY, L., BROOKS, B. W., RUDD, M. A., BACKHAUS, T., SPURGEON, D., VEROUGSTRAETE, V., AJAO, C., ANKLEY, G. T., APITZ, S. E., ARNOLD, K.,

BRODIN, T., CANEDO-ARGUELLES, M., CHAPMAN, J., CORRALES, J., COUTELLEC, M.-A., FERNANDES, T. F., FICK, J., FORD, A. T., GIMENEZ PAPIOL, G., GROH, K. J., HUTCHINSON, T. H., KRUGER, H., KUKKONEN, J. V. K., LOUTSETI, S., MARSHALL, S., MUIR, D., ORTIZ-SANTALIESTRA, M. E., PAUL, K. B., RICO, A., RODEA-PALOMARES, I., ROEMBKE, J., RYDBERG, T., SEGNER, H., SMIT, M., VAN GESTEL, C. A. M., VIGHI, M., WERNER, I., ZIMMER, E. I. & VAN WENSEM, J. 2018. Toward sustainable environmental quality: Priority research questions for Europe. *Environmental Toxicology and Chemistry,* 37**,** 2281-2295.

VAN DER POEL, P., BROOKE, D.N., VAN LEEUWEN, C.J. 2007a. Emissions of chemicals to the environment. *In:* VAN LEEUWEN, C. & VERMEIRE, T. (eds.) *Risk Assessment of Chemicals: An Introduction.* Springer. pp. 37-71.

VERMEIRE, T.G., BAARS, A.J., BESSEMS, J.G.M., BLAAUBOER, B.J., SLOB, W., MULLER, J.J.A. 2007b. Toxicity testing for human health risk assessment. *In:* VAN LEEUWEN, C. & VERMEIRE, T. (eds.) *Risk Assessment of Chemicals, An Introduction.* Springer. pp. 227-280

VAN LEEUWEN, C. J., VAN DERZANDT, P. T. J., ALDENBERG, T., VERHAAR, H. J. M. & HERMENS, J. L. M. 1992. Application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment. 1.Narcotic industrial pollutants. *Environmental Toxicology and Chemistry,* 11**,** 267-282.

VAN WEZEL, A. P. & OPPERHUIZEN, A. 1995. Narcosis due to environmental pollutants in aquatic organisms - residue-based toxicity, mechanisms, and membrane burdens. *Critical Reviews in Toxicology,* 25**,** 255-279.

VEITH, G. D. & BRODERIUS, S. J. 1990. Rules for distinguishing toxicants that cause Type-I and Type-II narcosis syndromes. *Environmental Health Perspectives,* 87**,** 207-211.

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., SOLBE, J., SPEKSNIJDER, J., VAN LEEUWEN, C. J. & HERMENS, J. L. M. 2000. Classifying environmental pollutants: Part 3. External validation of the classification system. *Chemosphere,* 40**,** 875-883.

VERHAAR, H. J. M., VAN LEEUWEN, C. J. & HERMENS, J. L. M. 1992. Classifying environmental pollutants. 1. Structure-Aactivity-relationships for prediction of aquatic toxicity. *Chemosphere,* 25**,** 471-491.

VILLALOBOS-GARCÍA, D. & HERNÁNDEZ-MUÑOZ, R. 2017. Catalase increases ethanol oxidation through the purine catabolism in rat liver. *Biochemical Pharmacology,* 137**,** 107-112.

VILLENEUVE, D., VOLZ, D. C., EMBRY, M. R., ANKLEY, G. T., BELANGER, S. E., LÉONARD, M., SCHIRMER, K., TANGUAY, R., TRUONG, L. & WEHMAS, L. 2014. Investigating alternatives to the fish early-life stage test: a strategy for discovering and annotating adverse outcome pathways for early fish development. *Environmental Toxicology and Chemistry,* 33**,** 158-69.

VINKEN, M. 2013. The adverse outcome pathway concept: a pragmatic tool in toxicology. *Toxicology,* 312**,** 158-165.

VON DER OHE, P. C., KUHNE, R., EBERT, R. U., ALTENBURGER, R., LIESS, M. & SCHUURMANN, G. 2005. Structural alerts - A new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay. *Chemical Research in Toxicology,* 18**,** 536-555.

VON STACKELBERG, K., GUZY, E., CHU, T. & HENN, B. C. 2015. Exposure to mixtures of metals and neurodevelopmental outcomes: a multidisciplinary review using an adverse outcome pathway framework. *Risk Analysis,* 35**,** 971-1016.

VONK, J. A., BENIGNI, R., HEWITT, M., NENDZA, M., SEGNER, H., VAN DE MEENT, D. & CRONIN, M. T. D. 2009. The use of mechanisms and modes of toxic action in integrated testing strategies:

the report and recommendations of a workshop held as part of the European Union OSIRIS integrated project. *ATLA-Alternatives to Laboratory Animals,* 5**,** 557-571.
WAHL, J. & SMIEŠKO, M. 2018. Endocrine disruption at the androgen receptor: employing molecular dynamics and docking for improved virtual screening and toxicity prediction. *International Journal of Molecular Sciences,* 19, 1784.
WANG, C. M., SCUKA, M. & NARAHASHI, T. 1972. Mechanism of negative temperature coefficient of nerve blocking action of allethrin. *Journal of Pharmacology and Experimental Therapeutics,* 182**,** 442-453.
WANG, X. H., FAN, L. Y., WANG, S., WANG, Y., YAN, L. C., ZHENG, S. S., MARTYNIUK, C. J. & ZHAO, Y. H. 2017. Relationship between acute and chronic toxicity for prevalent organic pollutants in *Vibrio fischeri* based upon chemical mode of action. *Journal of Hazardous Materials,* 338**,** 458-465.
WANG, X. H., YU, Y., HUANG, T., QIN, W. C., SU, L. M. & ZHAO, Y. H. 2016. Comparison of toxicities to *Vibrio fischeri* and fish based on discrimination of excess toxicity from baseline level. *PLOS One,* 11**,** 17.
WANG, Y., WANG, Q., WU, B., LI, Y. & LU, G. 2013. Correlation between TCDD acute toxicity and aryl hydrocarbon receptor structure for different mammals. *Ecotoxicology and Environmental Safety,* 89**,** 84-88.
WATERHOUSE, R. M., TEGENFELDT, F., LI, J., ZDOBNOV, E. M. & KRIVENTSEVA, E. V. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research,* 41**,** D358–D365.
WEBB, T. & MORLACCI, L. 2010. Calculation of physico-chemical and environmental fate properties. *In:* CRONIN, M. & MADDEN, J. (eds.) *In Silico Toxicology: Principles and Applications.* The Royal Society of Chemistry. 118-147.
WELBORN, M., CHENG, L. & MILLER, T. F. 2018. Transferability in machine learning for electronic structure via the molecular orbital basis. *Journal of Chemical Theory and Computation,* 14**,** 4772-4779.
WORGAN, A. P., DEARDEN, J. C., EDWARDS, R. , NETZEVA, T. I. & CRONIN, M. T. D., 2003. Evaluation of a novel short-term algal toxicity assay by the development of QSARs and inter-species relationships for narcotic chemicals. *QSAR & Combinatorial Science,* 22**,** 204-209.
XIE, J., ZHAO, L., LIU, K., GUO, F., GAO, L. & LIU, W. 2018. Activity, toxicity, molecular docking, and environmental effects of three imidazolinone herbicides enantiomers. *Science of the Total Environment,* 622-623**,** 594-602.
YAN, X. F., XIAO, H. M., GONG, X. D. & JU, X. H. 2005. Quantitative structure-activity relationships of nitroaromatics toxicity to the algae (*Scenedesmus obliguus*). *Chemosphere,* 59**,** 467-471.
YANG, J., ZHAO, H. & CHAN, K. M. 2017. Toxic effects of polybrominated diphenyl ethers (BDE 47 and 99) and localization of BDE-99-induced cyp1a mRNA in zebrafish larvae. *Toxicology Reports,* 4**,** 614-624.
YOUNG, D. H. & SLAWECKI, R. A. 2001. Mode of action of zoxamide (RH-7281), a new oomycete fungicide. *Pesticide Biochemistry and Physiology,* 69**,** 100-111.
YUAN, L., LV, B., ZHA, J., WANG, Z., WANG, W., LI, W. & ZHU, L. 2013. New cytochrome P450 1B1, 1C1, 2Aa, 2Y3, and 2K genes from Chinese rare minnow (*Gobiocypris rarus*): Molecular characterization, basal expression and response of rare minnow CYP1s and CYP2s mRNA exposed to the AHR agonist benzo[a]pyrene. *Chemosphere,* 93**,** 209-216.
ZHANG, C., ZHOU, T. T., WANG, J., ZHANG, S., ZHU, L. S., DU, Z. K. & WANG, J. H. 2018a. Acute and chronic toxic effects of fluoxastrobin on zebrafish (*Danio rerio*). *Science of the Total Environment,* 610**,** 769-775.

ZHANG, L., JIN, Y. R., HAN, Z. H., LIU, H. L., SHI, L. H., HUA, X. X., DOERING, J. A., TANG, S., GIESY, J. P. & YU, H. X. 2018b. Integrated *in silico* and *in vivo* approaches to investigate effects of BDE-99 mediated by the nuclear receptors on developing zebrafish. *Environmental Toxicology and Chemistry,* 37**,** 780-787.

ZHANG, S., DE BOER, A. H. & VAN DUIJN, B. 2016. Auxin effects on ion transport in *Chara corallina*. *J Plant Physiology,* 193**,** 37-44.

ZHANG, S. & VAN DUIJN, B. 2014. Cellular auxin transport in algae. *Plants (Basel),* 3**,** 58-69.

ZHANG, T., LI, X., MIN, X. M., FANG, T. T., ZHANG, Z. J., YANG, L. & LIU, P. 2012. Acute toxicity of chlorobenzenes in *Tetrahymena*: Estimated by microcalorimetry and mechanism. *Environmental Toxicology and Pharmacology,* 33**,** 377-385.

ZHANG, X., QIN, W., HE, J., WEN, Y., SU, L., SHENG, L. & ZHAO, Y. 2013. Discrimination of excess toxicity from narcotic effect: Comparison of toxicity of class-based organic chemicals to *Daphnia magna* and *Tetrahymena pyriformis*. *Chemosphere,* 93**,** 397-407.

ZHANG, X. J., QIN, H. W., SU, L. M., QIN, W. C., ZOU, M. Y., SHENG, L. X., ZHAO, Y. H. & ABRAHAM, M. H. 2010. Interspecies correlations of toxicity to eight aquatic organisms: Theoretical considerations. *Science of the Total Environment,* 408**,** 4549-4555.

ZHOROV, B. S. & DONG, K. 2017. Elucidation of pyrethroid and DDT receptor sites in the voltage-gated sodium channel. *Neurotoxicology,* 60**,** 171-177.

ZIMMERMANN, A. L., KING, E. A., DENGLER, E., SCOGIN, S. R. & POWELL, W. H. 2008. An aryl hydrocarbon receptor repressor from *Xenopus laevis*: Function, expression, and role in dioxin responsiveness during frog development. *Toxicological Sciences,* 104**,** 124-134.

ČOLOVIĆ, M. B., KRSTIĆ, D. Z., LAZAREVIĆ-PAŠTI, T. D., BONDŽIĆ, A. M. & VASIĆ, V. M. 2013. Acetylcholinesterase inhibitors: Pharmacology and toxicology. *Current Neuropharmacology.* 11, 315–335.