Development of *in silico* models to support repeat dose safety assessment of cosmetic ingredients to humans

Mark David Nelms

November 2014

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

Acknowledgements

First and foremost, I would like to thank my director of studies, Dr Steve J. Enoch, for his continuous support, guidance, and feedback throughout the course of my doctoral studies, as well as providing many of the initial ideas for this research. Similar thanks must go to my other supervisors Dr Judy C. Madden and Prof Mark T. Cronin.

I would also like to thank all my colleagues from the QSAR laboratory at Liverpool John Moores University. In particular: Dr Claire Mellor, for sharing her expertise of mitochondria during many useful discussions; and Drs Claire Ellison and Richard Marchese-Robinson, for providing a sounding board for ideas.

Additionally, I greatly appreciate the financial support provided to me by the EU COSMOS project by the European Community's 7th Framework Program and Cosmetics Europe, without which this research would not have been possible.

Special thanks go to Dr Chihae Yang (Altamira, LLC), for providing the standard operating procedure used within Chapter 2, for her valuable discussions, and for kindly allowing me to stay in her home during my time in Columbus, OH; Profs Vera Rogiers and Mathieu Vinken (Vrije Universiteit Brussel), for providing the NOAEL data utilised in Chapter 5; and Prof Terry W. Schultz (University of Tennessee) for providing the *in vitro* and *in chemico* data used in Chapter 6.

Thanks must also go to my friends and family, for their support and encouragement. Final thanks go to Erin Bishop who has provided immense emotional support throughout.

Abstract

Cosmetic products are used daily on a global scale. Therefore, it is necessary to ensure that these products, and their ingredients, do not cause any adverse human health effects under normal usage; to ensure this, risk assessment must be performed. Traditionally, risk assessments are performed *in vivo*, i.e. conducting tests on animals using the chemical(s) of interest. However, over the past decade there has been an increase in research into the use of alternative toxicity testing methods, such as *in vitro*, *in chemico* and *in silico*. Whilst there are a number of alternative techniques that may be employed, no one method can be used in isolation as a full replacement for an *in vivo* test. Therefore, the Adverse Outcome Pathway (AOP) concept is an emerging method by which information provided by the *in vitro*, *in chemico*, and *in silico* approaches can be utilised in an integrated testing strategy. The AOP concept links an upstream molecular initiating event to a downstream adverse outcome, via a number of testable key events.

In silico approaches utilise computers in order to develop predictive models. Within the AOP paradigm *in silico* method work to identify the key features of a chemical (structural alerts) that induce a molecular initiating event (MIE). A collection of structural alerts that induce the same MIE are considered to be an *in silico* profiler. Typically, these *in silico* profilers are supported by associated toxicity, or mechanistic, information pertaining to the ability to induce a specific MIE. The overall aim of the work presented in this thesis was the development of an *in silico* profiler, based upon the hypothesis that the induction of mitochondrial toxicity is a key driver of organ-level toxicity. The research presented herein demonstrates the ability to identify, and develop, two types of structural alert; mechanism- and chemistry-based; that pertain to mitochondrial toxicity. Due to the differences inherent in these two types of alert they should be utilised for different purposes. As such, the main usage of the mechanism-based alerts should be in the formation of chemical categories and subsequent data gap filling via read-across. In comparison, the chemistry-based alerts should be utilised for

the purposes of prioritising chemicals, within an inventory, that should undergo additional testing in *in vitro* and/or *in chemico* assays. It is envisaged that these two types of structural alerts could be used to profile chemical inventories as part of a tiered testing strategy.

Therefore, the future work discussed in detail the need to expand the chemical space covered by the alerts. Additional future work involves utilising experimental information from *in vitro*/*in chemico* assays to verify the mechanism-based alerts and to refine the chemistry-based alerts by discerning mechanistic information associated with them. Furthermore, it is envisaged that these alerts could be incorporated into predictive tools, such as the OECD QSAR Toolbox, to enable their use for screening and prioritisation purposes.

Abbreviations

ADP	Adenosine Diphosphate
AOP	Adverse Outcome Pathway
ATP	Adenosine Triphosphate
BUAV	British Union for the Abolition of Vivisection
CAS	Chemical Abstract Service
COSMOS	Integrated in silico models for the prediction of human repeated dose toxicity of COSMetics to Optimise Safety
DMSO	Dimethyl Sulphoxide
DTND	5,5'-dithio-bis(2-nitrobenzoic acid)
ECEAE	European Coalition to End Animal Experiments
ECHA	European Chemicals Agency
ECVAM	European Centre for the Validation of Alternative Methods
EPA	Environmental Protection Agency
ER	Endoplasmic Reticulum
EU	European Union
FADH	Flavin Adenine Dinucleotide
GMPT	Guinea Pig Maximisation Test
GSH	Glutathione
HNEL	Highest No Effect Level
IATA	Integrated Approach to Testing and Assessment
IGC50	50% inhibitory growth concentration
IMM	Inner Mitochondrial Membrane
IMS	Intermembrane Space

ITS	Integrated Testing Strategy
LEL	Lowest Effect Level
LLNA	Local Lymph Node Assay
LO(A)EL	Lowest Observed (Adverse) Effect Level
MIE	Molecular Initiating Event
MM	Mitochondrial Matrix
MPT	Membrane Permeability Transition
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NADH	Nicotinamide Adenine Dinucleotide
NO(A)EL	No Observed (Adverse) Effect Level
OASIS	Optimised Approach based on Structural Indices Set
OECD	Organisation for Economic Cooperation and Development
PAFA	Priority-based Assessment of Food Additives
PETA	People for the Ethical Treatment of Animals
(Q)SAR	(Quantitative) Structure-Activity Relationships
RC_{50}	50% reactive concentration
REACH	Registration, Evaluation, Authorisation and restriction of Chemicals
ROS	Reactive Oxygen Species
SCCNFP	Scientific Committee on Cosmetic products and Non-Food Products
SCCP	Scientific Committee on Consumer Products
SCCS	Scientific Committee on Consumer Safety
SEURAT	Safety Evaluation Ultimately Replacing Animal Testing
SMARTS	SMiles ARbitrary Target Specification
SMILES	Simplified Molecular-Input Line-Entry System

S_N1	Unimolecular aliphatic Nucleophilic Substitution
S_N2	Bimolecular aliphatic Nucleophilic Substitution
S_NAr	Aromatic Nucleophilic Substitution
SOP	Standard Operating Procedure

Contents

Chapter 1. Introduction ····· 1

1.1. Risk assessment and conventional toxicological testing ····· 1
1.2. European Union regulation ····· 3
1.3. Alternative testing methods ····· 5
1.3.1.Adverse Outcome Pathways and Molecular Initiating Events ····· 6
1.3.2.In silico profilers ····· 7
1.3.3.Category formation and read across ····· 8
1.3.4.Profiling inventories for prioritisation ····· 9
1.3.5.Expert systems ····· 10
1.4. Molecular Initiating Events for repeat dose toxicity ····· 11
1.5. Cellular function and mitochondrial toxicity ····· 12
1.5.1.Cellular function ····· *12*
1.5.2.Cellular toxicity ····· *13*
1.5.2.1. Autophagy ····· *13*
1.5.2.2. Necrosis ····· *13*
1.5.2.3. Apoptosis ····· *14*
1.5.3.*Mitochondria* ····· 14
1.5.3.1 General structure and function ····· 14
1.5.3.2 Oxidative phosphorylation ····· 16
1.6. Mechanisms of mitochondrial dysfunction ····· 18
1.7. SEURAT-1 and the COSMOS project ····· 20
1.8. Research aims of this thesis ····· 22

Chapter 2. Extraction and collation of repeat dose toxicity data from Scientific Committee on Consumer Safety reports into the COSMOS database ····· 24

2.1 Introduction ····· 24
2.2 Method ····· 35
2.3 Application of the COSMOS database using 28- and 90- day repeat dose toxicity data ····· 48
2.3.1 *Introduction to COSMOS database* ····· *49*
2.3.2 Method ····· *50*
2.3.3 Results and discussion of the analysis performed on the dataset ····· *55*
2.4 Conclusions ····· 57

Chapter 3. The development of structural alerts using the ChemoTyper software ·····58

3.1 Introduction ·····58
3.2 Methods ·····61
3.2.1 Data set ·····61
3.2.2 Generation of structural alerts using the ChemoTyper software ·····62
3.3 Results and discussion ·····70
3.3.1 Structural alerts for which an all-encompassing mechanistic hypothesis is possible ·····77
3.3.2 Structural alerts for which an all-encompassing mechanistic hypothesis is not possible ·····78
3.4 Conclusions ·····81

Chapter 4. The use of category formation in the development of an *in silico* profiler for mitochondrial toxicity ·····82

4.1 Introduction ·····82
4.2 Methods ·····86
4.2.1 Data set ·····86
4.2.2 Category formation based upon structural similarity ·····86
4.2.3 Mechanistic hypothesis and the development of alerts ·····87
4.3 Results and discussion ·····88
4.3.1 Profiling and grouping for mitochondrial toxicity as part of the AOP paradigm ·····103
4.4 Conclusions ····· 104

Chapter 5. Development of an *in silico* profiler for categorisation of repeat dose toxicity data of hair dyes ····· 106

5.1 Introduction ····· 106
5.2 Methods ····· 110
5.2.1 Experimental data ·····110
5.2.2 Development of in silico profiler ·····111
5.3 Results and discussion ····· 115
5.3.1 Development of mechanism-based structural alerts for category formation ·····116
5.3.2 Additional chemicals capable of electron cycling ·····126
5.3.3 Mitochondria and repeat dose toxicity ·····131

5.3.4 Adverse Outcome Pathway concept, perspectives and proposed future work····132
5.4 Conclusions ····134

Chapter 6. Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes, and ketones ····· 135

6.1 Introduction ····135
6.2. Method ····145
6.2.1 Data set ····145
6.2.2 In chemico glutathione reactivity ····145
6.2.3 Aquatic toxicity data ····146
6.2.4 In silico predictions ····147
6.2.5 Verification of alerts ····148
6.3 Results and discussion ····149
6.3.1 Relationship between toxicity to Tetrahymena pyriformis and hydrophobicity ····150
6.3.2 Verification of structural alerts ····160
6.4 Conclusions ····161

Chapter 7. Discussion ····· 163

7.1 Progress in developing structural alerts for repeat dose toxicity ····· 163
7.1.1 Summary of work ····163
7.1.2 Conclusions ····167
7.1.3 Proposal of a tiered testing strategy for profiling of chemical inventories ····169
7.2 Prospects for future work ····170
7.2.1 Verification of the developed mechanism-based structural alerts ····171
7.2.2 Use of in vitro/in chemico data to discern mechanistic information for chemistry-based alerts ····172
7.2.3 Development of additional alerts ····172

8. References. ····· 174

9. Appendices. ····· 198

Chapter 1: Introduction

1.1 Risk assessment and conventional toxicological testing

Cosmetic products are used extensively in many communities throughout the world on a daily basis. A cosmetic product is defined by Directive 76/768/EEC as *'any substance or mixture intended to be placed in contact with the various external parts of the human body or with the teeth and the mucous membranes of the oral cavity with a view exclusively or mainly to cleaning them, perfuming them, changing their appearance and/or correcting body odours and/or protecting them or keeping them in good condition'* (EC 2003). Due to their global usage it is essential, both for producers and consumers, that these cosmetic products do not cause any adverse human health effects under normal use conditions. This makes undertaking risk assessment a critical step in product development. Risk assessment is used to assess the potential adverse effects following application or exposure to both human health and the environment of natural and synthetic agents.

There are, traditionally, four components needed in order for a risk assessment to be completed; identification and characterisation of the potential hazard; assessment of the dose required to cause toxicity; assessment of the potential exposure and characterisation of the likely risk to the exposed population; with the final outcome being a prediction of the possible risk to humans or the environment (van Leeuwen 2007). The information required for risk assessment to be undertaken is provided by toxicological studies, typically carried out using animal experiments. The science of toxicology is an inter-disciplinary subject that explores the relationship between chemistry, biology and pharmacology to determine the level of exposure to chemicals that are deemed safe to living organisms. There are three main methods that can be employed to determine the toxicity of a chemical:-

In vivo – tests conducted on humans or animals (animal data are then used to infer toxicity in humans)

In vitro – tests conducted on model system(s) based on either animal or human cells and/or systems

In silico – computer based models that provide predictions of toxicity, using existing *in vivo* or *in vitro* data.

Historically, *in vivo* experimentation is the most frequently employed method in toxicological testing, as data derived from *in vivo* tests are considered the most representative of the endpoint of interest (in this case human toxicity). A toxic endpoint is the effect that a test chemical has on the organ(s) of interest, or on the organism as a whole, during toxicity experimentation. The most relevant endpoints, from a cosmetics perspective, are skin sensitisation, mutagenicity, carcinogenicity, reproductive, and repeat dose toxicity. These endpoints have been tested using a variety of *in vivo* procedures. For example, skin sensitisation testing employs the Local Lymph Node Assay (LLNA) (OECD 2002) and Guinea Pig Maximisation Test (GPMT) (OECD 1992); whilst the rodent carcinogenicity assay is employed to test for carcinogenicity (OECD 2008); and the repeated dose 28- and/or 90-day toxicity studies in rodents test for repeat dose toxicity (OECD 1995, OECD 1998). Chronic toxicity is the observed adverse health effect(s) that occur after the repeated exposure of an organism to a substance on a daily basis for an extended period of time; up to the entire lifespan of the test species. These studies are important as they enable a safe dosage for humans to be discerned through the derivation of a No Observed (Adverse) Effect Level (NO(A)EL); "the highest exposure level at which there are no biologically significant increases in frequency or severity of adverse effects between the exposed population and it appropriate control." (EPA 1995, Lewis *et al* 2002). In order for a NOAEL value to be derived repeat dose toxicity testing must be conducted in, for example, a 28- or 90-day oral, dermal, or inhalation rodent study. This type of study provides the identification of the organ(s) that drive the NO(A)EL value and the derivation of a Lowest Observed (Adverse) Effect Level (LO(A)EL) value; "the lowest exposure level at which there are

biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group." (EPA 1995, Lewis *et al* 2002).

Due to the introduction, and implementation, of two key pieces of European legislation; namely REACH and the 7th amendment to the Cosmetics Directive (discussed below); over the past decade there has been an increased need to replace these *in vivo* repeat dose studies with *in vitro* and/or *in silico* alternatives (EC 2003, EC 2006, EC 2006). Some of this need has arisen from an increase in animal welfare campaigns from organisations such as People for the Ethical Treatment of Animals (PETA) and Lush Cosmetics Company against the use of animal testing. Another factor is the reduction in cost when utilising *in silico* (as compared to *in vivo* or *in vitro*) methods in the primary screening of new products. A single 90-day oral repeat dose rodent study costs between $125,000-175,000 and uses approximately 80 animals (10 animals/sex/group and at least 4 dose groups) per chemical. It is envisaged that alternative techniques will reduce both the monetary and animal costs involved. However, the major contributing factor has been the introduction of legislation coming into force from international bodies, primarily the European Union (EU).

1.2 European Union regulation

In 2007, the EU Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) regulation came into force. The REACH regulation acts to promote Russell and Burch's "3Rs" principle in safety assessment (EC 2006). The 3Rs principles aim to reduce the number of test animals per experiment, to minimise the suffering of the animals by refining the experimental protocol, and to, ultimately, replace animal testing with alternative methods (Russell 1959). Under REACH, any substance produced or imported into the EU at, or above, one tonne per year is required to be registered with the European Chemicals Agency (ECHA). When registering with ECHA companies are obligated to provide details of the substance's properties, along with assessments of the chemicals safety based upon

toxicity testing; with the emphasis on the use of alternative methods and animal experimentation occurring only as a last resort.

Further to the REACH legislation, which encompasses all substances, is the Cosmetic Directive that focuses specifically on cosmetic products and their ingredients (EC 2003). Due to the variety of functions that are performed by cosmetic products their ingredients cover a wide range of chemical space; from low molecular weight chemicals such as cyanoacrylates (used in nail glue) to higher molecular weight chemicals such as sodium laureth sulphate (used in shampoos and toothpastes). Additionally, depending upon their end use, cosmetic ingredients may be inherently reactive towards proteins such as Acid Black 1 (used as a colourant in hair dye products) or unreactive such as dimethicone (used as an emollient in hand/body lotions and hair conditioners). The Cosmetic Directive places the responsibility of cosmetic product safety on the company that is releasing the product onto the market. The aims of the most recent revision to the EU Cosmetic Directive, the 7th amendment introduced in 2004, were two-fold (EC 2003). Firstly, it prohibits the testing of finished cosmetic products on animals after 2004, along with a phasing out of testing of individual cosmetic ingredients as alternative methods became validated by EU regulatory bodies, such as the European Centre for the Validation of Alternative Methods (ECVAM). The deadline for this phasing out passed in 2009 and was imposed whether an alternative method was available or not. The second measure imposes a step-by-step Europe-wide marketing ban if either a finished product or its ingredients have been tested on animals after 2009, in line with the above testing ban for individual ingredients. There were exceptions to this marketing ban (i.e. for repeat-dose, reproductive toxicity, as well as toxicokinetic studies) whereby the final deadline passed in March 2013.

1.3 Alternative testing methods

Since the implementation of both REACH and the 7th amendment of the Cosmetic Directive there has been a marked increase in the effort to find alternative methods for chemical risk assessment. These alternatives have been developed employing *in vitro*, *in chemico* and *in silico* techniques. *In vitro* testing is performed under laboratory conditions on isolated biological organisms, such as cell lines or subcellular components. The Ames test, for example, uses a bacterial strain lacking the ability to produce histidine - grown on a media containing no histidine - to assess the mutagenic potential of compounds (Ames *et al* 1973, OECD 1997). *In chemico* testing is used to assess the ability of a compound to react with, and bind covalently to, important biological macromolecules such as proteins. For example, the glutathione assay is used to assess the ability of a compound to bind to proteins containing a thiol moiety (e.g. cysteine) (Aptula *et al* 2006). *In silico* techniques are computer-based alternatives that enable predictions to be made covering a broad range of endpoints, for example, skin sensitisation and teratogenicity (Enoch *et al* 2008a, Enoch *et al* 2008b, Enoch *et al* 2009, Enoch *et al* 2011a). Aside from reducing the number of animals used within regulatory assessment these testing strategies also have numerous other, obvious, advantages over *in vivo* experimentation. For example, they are less time consuming and more cost-effective. There are a number of *in silico* approaches that can be employed to facilitate predictions regarding a substance's toxicity useful for risk assessment. These *in silico* approaches can be classified as belonging to one of two broad categories; (Quantitative) Structure-Activity Relationships ((Q)SARs) and category formation and read-across. Structure-activity relationships are a method that can be used to relate the structure or (arrangement of) functional groups of a chemical to its biological activity. QSARs are mathematical models that utilise numerical values of physico-chemical properties (also known as descriptors) in order to predict the activity, or potency, of a chemical. In comparison, category formation and read-across is a concept whereby the toxicity of a chemical (with no known toxicological profile) is predicted based upon its similarity to

analogous chemicals, for which toxicological data are available. Whilst there are numerous alternative techniques that can be used to predict various aspects of *in vivo* experimentation, no one method in isolation can replace animal testing. To overcome this problem these alternatives will need to be used in conjunction with one another as part of an integrated testing strategy (ITS) (Hartung *et al* 2013).

1.3.1 Adverse Outcome Pathways and Molecular Initiating Events

A framework is, therefore, required that enables the information provided by these different testing methods to be integrated and organised in a transparent and cohesive manner. To address this, the Adverse Outcome Pathway (AOP) paradigm has been introduced to provide such a framework (Ankley *et al* 2010, OECD 2013, Vinken 2013, Vinken *et al* 2013a, Vinken *et al* 2014). An AOP is a construct that means to establish a mechanistic connection between two anchors; the upstream Molecular Initiating Event (MIE) and the downstream adverse outcome relevant for risk assessment; via a number of key testable events at differing levels of biological organisation (such as the cellular, organ or organism level) (Ankley *et al* 2010, Schultz 2010, OECD 2013, Przybylak and Schultz 2013). The MIE is the primary event that triggers the progression of the AOP towards the adverse event. As such it provides mechanistic information pertaining to the initial interaction between the chemical and the biological system. Upon elucidation of the mechanistic information regarding the MIE it can be associated with structural fragments and, potentially, physico-chemical properties. AOPs are generally represented as a linear construct, as shown in Figure 1.1, with toxicity progressing sequentially from one level to another. However, due to the complex and multi-faceted nature of toxicity, multiple key effects (e.g. at the cellular level) may by induced by one effect upstream (e.g. at the MIE). Therefore, the true pathway may include a number of branching tracks that, ultimately, culminate in the same adverse effect; for example the cholestasis AOP developed by Vinken *et al* (2013b). A number of AOPs have been developed for a wide variety of adverse outcomes including; oestrogen

receptor-mediated reproductive toxicity, cholestasis, weak acid respiratory uncoupling, skin sensitisation and voltage-gated sodium channel-mediated neural toxicity (Ankley *et al* 2010, Schultz 2010, OECD 2011, Landesmann *et al* 2012, OECD 2012). It should be noted that a more comprehensive list of the available AOPs can be found in the AOP wiki via the AOP knowledge base (available from https://aopkb.org/, accessed 17.11.2014). In addition, the OECD has published a guidance document, as a way to standardise this framework, outlining the process by which AOPs should be developed and assessed for their reliability and robustness (OECD 2013).

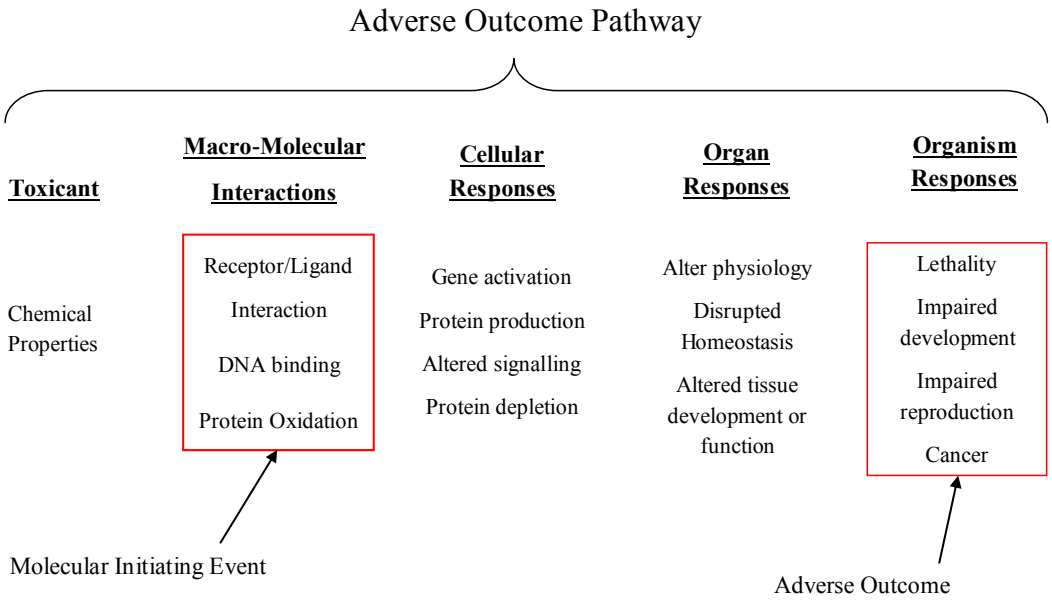


Figure 1.1: Summary of the steps and examples within an adverse outcome pathway (adapted from Ankley *et al* (2010))

1.3.2 In silico profilers

The main contribution that *in silico* techniques provide to the AOP approach is with respect to defining the MIE. This is achieved by identifying chemical structures (or fragments) that are associated with inducing toxicity. These structural fragments form the basis of structural alerts. A collection of structural alerts that culminate in the same outcome, for example DNA or protein binding, are considered to be an *in silico* profiler (Enoch and Cronin 2010, Enoch *et al* 2011b). There are two broad classifications of profilers: mechanistic and non-

mechanistic (described as chemistry-based in this thesis). A mechanistic profiler consists of a (group of) structural alert(s) relating to an MIE that has the potential to induce a specific endpoint. It is essential that each of the structural alerts that comprise a mechanistic profiler have, associated with them, experimental data that provide evidence that demonstrate how the structural alert induces toxicity via the MIE. This experimental data can originate from a variety of sources, such as; (historical) *in vivo*, *in vitro* and/or *in chemico* data present in the available literature; or data generated by in-house *in vitro* and/or *in chemico* experimentation. In comparison, a chemistry-based profiler consists of a (group of) structural alert(s) that have been identified as being associated with inducing a specific endpoint. However, there is no mechanistic information pertaining to how these alerts initiate toxicity. Chemistry-based profilers generally use a chemoinformatics approach, in which the presence of a specific structural fragment is associated with inducing toxicity. Whilst chemistry-based structural alerts have been associated with toxicity, the nature of these alerts can make it difficult to identify the mechanism that initiates the observed toxicity. This is because these alerts may be small fragments found in a variety of chemicals that may induce toxicity via different mechanisms. Due to the inherent differences each type of profiler should be utilised for different purposes; with mechanistic profilers being useful for category formation and subsequent read-across analysis, whilst chemistry-based profilers can be useful for screening large datasets and prioritising those chemicals that should undergo *in vitro* and/or *in chemico* testing first.

1.3.3 Category formation and read across

Chemical category formation is an approach whereby a set of chemicals with similar common properties, or trends in properties, are grouped together into a category (ECHA 2008, OECD 2011). Chemical categories may be developed based upon a similar mechanism of action or, more specifically, the MIE. Therefore, the mechanism-based structural alert(s) can be utilised in the formation of a chemical category, with the same

alert(s) needing to be present in both the target chemical and its analogues. The importance of the category formation approach has led to the development of the Organisation for Economic Co-operation and Development's (OECD) QSAR Toolbox software package (discussed in Chapters 3 and 6). This software tool contains a number of mechanistic profilers enabling chemical categories to be formed for a range of toxicological endpoints (available from www.qsartoolbox.org). Once a chemical category has been developed, predictions can then be made, via read-across, using the premise that similar chemicals should have similar biological and/or chemical activities (Jaworska and Nikolova-Jeliazkova 2007). These read-across predictions involve using existing toxicological data associated with the chemicals in the category in order to predict the activity of the target chemical (for which no toxicological data exist for the endpoint in question). Importantly, depending upon the data that are available for the analogous chemicals, within a specific category, both qualitative and/or quantitative predictions can be made using this approach.

1.3.4 Profiling inventories for prioritisation

A chemical inventory is a library of information, such as chemical name and other identifiers, pertaining to a set of chemicals. In contrast to a database, an inventory does not contain any toxicological data associated with the chemicals. Inventory screening is a process whereby an *in silico* profiler is utilised to quickly identify chemicals, within an inventory, with the potential to induce toxicity, due to the presence of a structural alert. As the chemistry-based structural alerts that compose a chemistry-based profiler do not contain any mechanistic information they are, therefore, less appropriate for use in read-across. This is because the structural fragment that is incorporated in the structural alert may be present in a multitude of different chemicals, each of which may induce toxicity via a different mechanism. However, these profilers can be utilised to screen large data sets, or inventories, for chemicals that contain one, or more, of the structural alerts. Therefore, any chemical that triggers an alert would be hypothesised to have the ability to induce toxicity. As such, the

identification of a chemistry-based alert within a (group of) chemical(s) can be utilised to prioritise these chemicals as requiring to undergo further testing within relevant *in vitro* or *in chemico* assays. It is envisaged that, from this additional testing, mechanistic information pertaining to how the chemical structure initiates the observed toxicity will be elucidated. This information can, thus, be utilised to support the adjustment of the chemical-based alert into a mechanistic alert. Whilst chemistry-based alerts can be used in order to screen an inventory, it should be noted that the use of mechanism-based alerts is preferable; as these would also provide an insight as to the mechanism by which the chemical may induce toxicity.

1.3.5 Expert systems

In addition to the development of novel mechanistic or chemistry-based profilers, there are a number of freely available, and subscription-based, software packages, known as expert systems, such as Toxtree, TIMES-SS, and DEREK Nexus (formerly DEREK for Windows). These expert systems are a repository of expert knowledge containing structural alerts for a variety of endpoints, such as mutagenicity, carcinogenicity, skin sensitisation, and skin irritation. Expert systems use libraries of endpoint specific structural alerts to enable a user to identify chemicals that may have the potential to induce toxicity; based upon the presence of a structural alert in the chemical of interest. Expert systems, such as DEREK Nexus, are utilised by a variety of companies in the pharmaceutical and/or cosmetics industries. These companies use the information held by the expert system to screen data sets, during the early phases of product development, as a method to guide chemical selection for prioritisation for further testing within *in vitro* and/or *in chemico* assays. Additionally, the predictions made by an expert system can be, and are, used in lead optimisation in order to identify those fragments that are required to be removed to reduce toxicity.

1.4 Molecular initiating events for repeat dose toxicity

Repeat dose toxicity testing identifies the adverse effects that are observed after the test organism has been subjected to repeated exposure of a test chemical over a period of time, up to the entire lifespan of the organism. Covalent protein and DNA binding has been shown to be an initiating event in multiple toxicological endpoints such as, skin sensitisation, carcinogenicity, mutagenicity, and hepatotoxicity (de Groot and Noll 1983, Woodward and Timbrell 1984, Aptula and Roberts 2006, Aptula *et al* 2006, Enoch *et al* 2008a, Enoch *et al* 2008b, Enoch and Cronin 2010, Enoch *et al* 2011a, OECD 2012, Hewitt *et al* 2013). As protein and DNA binding have been implicated in initiating numerous toxicological endpoints it is, therefore, plausible that these are also important in inducing repeat dose toxicity. In addition, research has shown that mitochondrial dysfunction is another etiological agent of toxicity, especially of pharmaceutical drugs, in a wide variety of organs, such as the heart, liver, and kidney (Dykens and Will 2007, Dykens and Will 2008, Nadanaciva and Will 2011). This implication of mitochondrial toxicity as a driver in various organ-level toxicities is partly attributable to the withdrawal of a number of pharmaceutical drugs from the market upon observation of mitochondrial dysfunction (Dykens 2008). However, even though there has been an increase in interest of screening chemicals for the ability to induce mitochondrial dysfunction, in comparison to protein/DNA binding, there have been relatively few *in silico* models or structural alerts developed (Zhang *et al* 2009, Naven *et al* 2013, Wallace *et al* 2013). It is for these reasons that the (potential) ability for chemicals to induce mitochondrial toxicity, after repeat exposures, was decided to be the main focus of this thesis.

1.5 Cellular function and mitochondrial toxicity

1.5.1 Cellular function

Every organism is composed of cells. A cell is the simplest collection of matter that can be considered to be alive. Complex, multicellular organisms, such as plants and animals, consist of a number of different organs that perform distinct functions. Organs are composed of specialised (eukaryotic) cells that enable the organ to function correctly. Whilst there are numerous cell types within the human body - such as hepatocytes, myocytes, and nephrocytes – the general structure and composition is well conserved. A typical somatic cell contains many integral internal structures (organelles) that perform specific functions within the cell. There are two categories of organelle: membrane-bound and non-membrane bound. Membrane-bound organelles include, but are not limited to –

- The nucleus, a double-membrane bound organelle, which contains the majority of the cells genetic material, the ability to synthesis and assemble ribosomes, and controls the activity of the other organelles in the cell.
- The endoplasmic reticulum (ER), a single-membrane bound organelle that is continuous with the outer membrane of the nucleus. There are two forms of ER: rough and smooth. Rough ER is associated with ribosomes and is involved in the synthesis and folding of proteins. Smooth ER is not associated with ribosomes and is involved in the synthesis of (phospho)lipids and steroids.
- The Golgi apparatus, a single-membrane bound organelle, which forms vesicles to transport proteins (from the rough ER) throughout the cell or to be secreted into the extracellular space.
- The mitochondria, a double-membrane bound organelle, which is discussed in more detail below.

Non-membrane bound organelles include, but are not limited to: ribosomes, which are involved in protein synthesis; proteasomes, which are used in the degradation of damaged proteins; and the cytoskeleton, which performs a multitude of functions such as maintaining

cell shape, anchoring organelles within the cell and intracellular transport of vesicles and organelles.

1.5.2 Cellular toxicity

Cell death is an essential process within normal human development and homeostasis. However, it can also be initiated by either internal or external xenobiotics or toxins that disrupt normal cell physiology and function. There are three main pathways by which cell death can occur in mammalian cells: autophagy, necrosis, and apoptosis.

1.5.2.1 Autophagy

Autophagy is a highly regulated catabolic mechanism that is activated in response to cellular stressors such as oxidative stress, cellular starvation, irradiation, and/or xenobiotics (Levine *et al* 2008). The morphological characteristics observed during autophagy include the sequestration of cytoplasmic material within double-membraned vesicles (autophagosomes) (Klionsky *et al* 2000). These autophagosomes subsequently fuse with lysosomes degrading the contents of the autophagosome (Fink *et al* 2005, Levine *et al* 2008).

1.5.2.2 Necrosis

Necrosis is considered to be an uncontrolled pathway of cell death that is not energy-dependent (Fink *et al* 2005). Disruption to normal physiological functioning of the cell – such as ATP production, ion transport, and pH balance – may all lead to necrotic cell death. Typical characteristics of necrotic cell death include: swelling and vacuolisation of the cytoplasm, dilation of membrane-bound organelles, and an inflammatory response surrounding the necrotic cell (Fink *et al* 2005).

1.5.2.3 Apoptosis

Apoptosis (programmed cell death) is an essential pathway in maintaining cellular homeostasis within multicellular organisms. The progression of cell death via apoptosis is a well regulated and organised sequence of events that includes fragmentation of DNA, condensation of chromatin, blebbing of the cell membrane and fragmentation into 'apoptic bodies,' and finally engulfment of these 'apoptic bodies' by macrophages or neighbouring cells. There are two main pathways of inducing apoptosis: intrinsic and extrinsic.

The extrinsic pathway is mediated by the binding of death receptor ligands to death receptors present within the plasma membrane, thus, triggering the activation of various downstream caspases (Green *et al* 2004). In contrast, the mitochondria are the main mediators of the intrinsic pathway. The pivotal event in the induction of the intrinsic pathway is the permeabilisation of the mitochondrial outer membrane. There are two mechanisms that can initiate this permeabilisation (Green *et al* 2004). The first mechanism is activated by members of the Bcl-2 family of proteins – which enable the formation of pores in the outer mitochondrial membrane – thereby releasing pro-apoptic molecules (such as cytochrome *c*) that initiate various downstream caspases. The second mechanism of intrinsic pathway initiation involves the induction of mitochondrial permeability transition, which is discussed in more detail in section 1.6.

1.5.3 Mitochondria

1.5.3.1 General structure and function

The mitochondria are organelles present in virtually every cell type of the human body, the exception being mature erythrocytes (Cohen and Gold 2001). The number and shape of mitochondria varies greatly between cell types, with those cells with a higher energy demand such as cardiac and skeletal muscle containing a higher number of mitochondria (Amacher 2005, Pieczenik and Neustadt 2007, Nadanaciva and Will 2011). Whilst the

external morphology of mitochondria can vary between cell types, the constituent parts are conserved (Collins and Bootman 2003, Amacher 2005, Nadanaciva and Will 2011). The basic structure of mitochondria consists of two membranes (the outer and inner membrane) enclosing two components; the intermembrane space and the mitochondrial matrix (Figure 1.2). The outer membrane is relatively smooth and, due to the presence of pores consisting of voltage-dependent anion channels, is permeable to molecules that are less than 5kDa in size (Amacher 2005, Dykens and Will 2008). In contrast, the inner membrane contains multiple invaginations (cristae); is permeable to relatively few molecules including O_2, CO_2, and H_2O; and contains each of the protein complexes that comprise the electron transport chain, ATP synthase (Complex V) and various electron carriers (Dykens and Will 2008, Nadanaciva and Will 2011). Specialised transporting proteins are required in order to enable hydrophilic molecules and inorganic ions to pass across the inner mitochondrial membrane (Amacher 2005, Dykens and Will 2008). Mitochondria are essential for a number of functions vital to a cell's normal functioning and survival. The primary function being the production of approximately 95% of the total adenosine triphosphate (ATP) generated by cells during oxidative phosphorylation; the remainder being made via glycolysis.

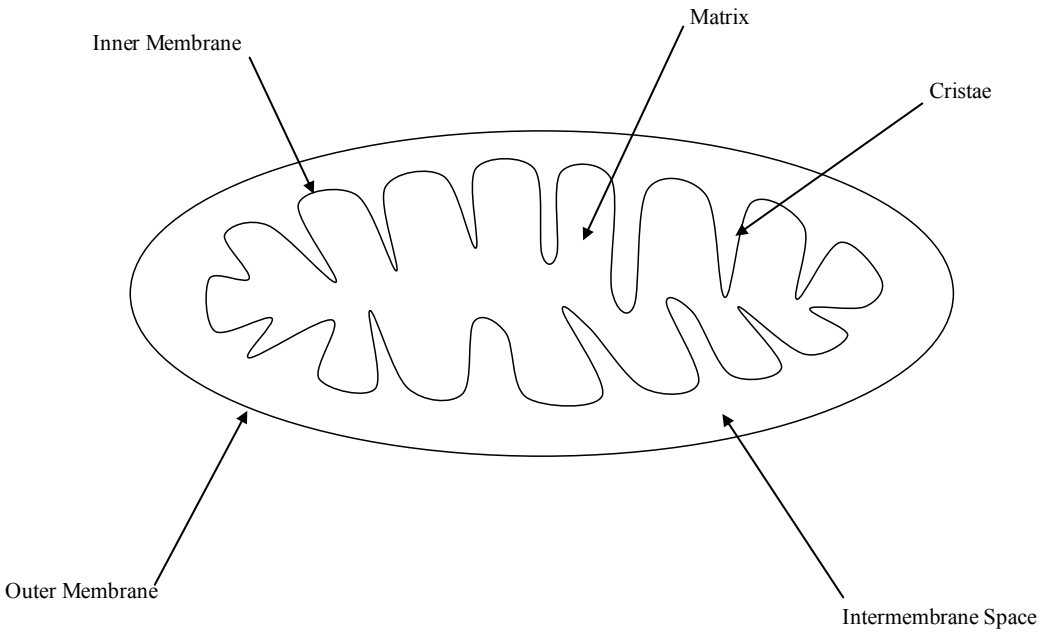


Figure 1.2: Basic structure of mitochondria

1.5.3.2 Oxidative phosphorylation

Oxidative phosphorylation is the process whereby the transfer of electrons, along the electron transport chain, is coupled to ATP synthesis, i.e. the phosphorylation of adenosine diphosphate (ADP) by inorganic phosphate. The electron transport chain is central to the process of oxidative phosphorylation and is comprised of four protein complexes (Complexes I – IV) situated within the inner mitochondrial membrane (Figure 1.3). In addition, Complex V (part of the wider respiratory chain) is the essential protein complex that phosphorylates ADP to produce ATP. Acetyl coenzyme A, generated either by glycolysis in the cytosol or fatty acid β-oxidation in the mitochondria, enters the citric acid cycle (Figure 1.4); whereby, electrons are used to reduce nicotinamide adenine dinucleotide (NAD+) and flavin adenine dinucleotide (FAD) to NADH and $FADH_2$ respectively.

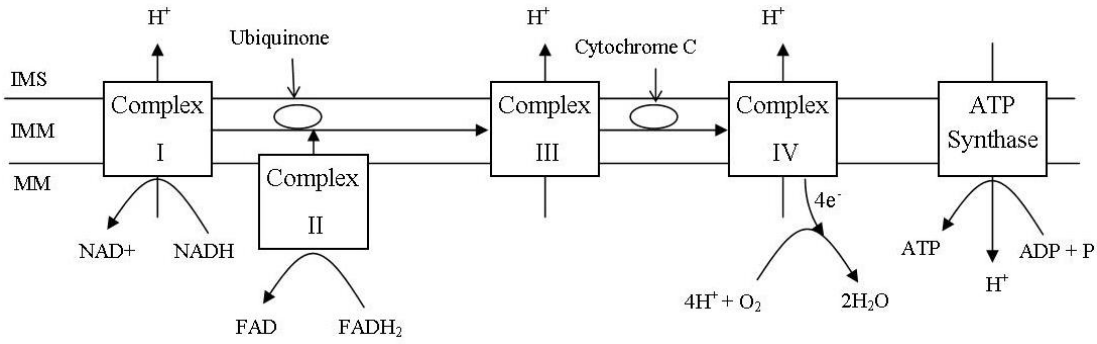


Figure 1.3: An illustration of how the respiratory chain is involved in oxidative phosphorylation

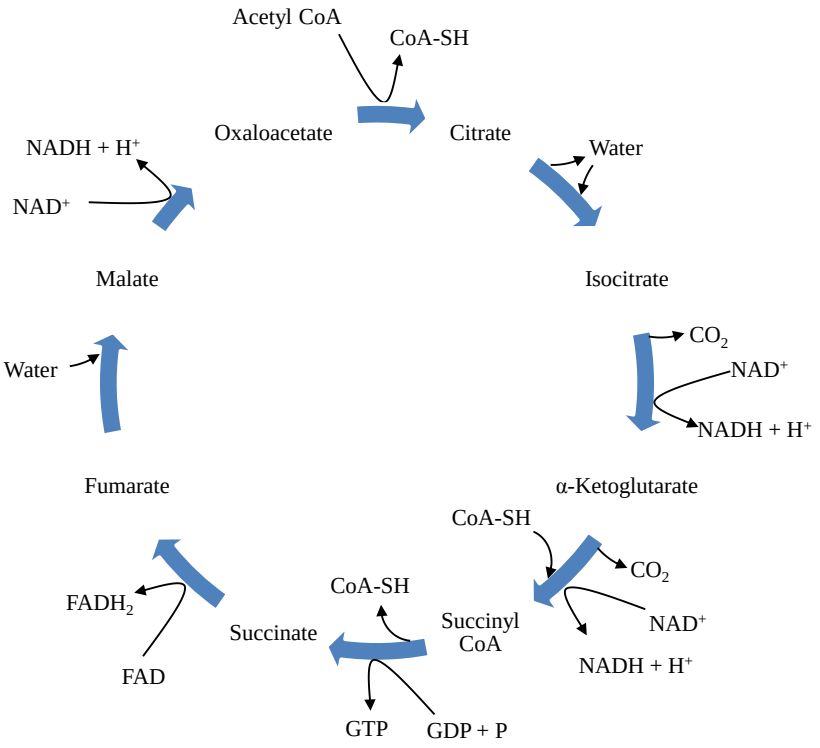


Figure 1.4: The generation of NADH and $FADH_2$ by the citric acid cycle

Complexes I and II are involved in the oxidation of NADH and $FADH_2$ respectively, thereby, providing the input of electrons into the respiratory chain. Both Complex I and II donate electrons to the mobile electron carrier ubiquinone: reducing it to ubiquinol. Ubiquinol, thus, transports the electrons to Complex III, which, in turn, donates the electrons to another electron carrier: cytochrome *c*. Cytochrome *c* subsequently donates the electrons to the final protein complex in the electron transport chain: Complex IV. Finally, Complex IV donates the electrons to molecular oxygen producing water. This process of electron transfer from electron donor to electron acceptor (illustrated in Figure 1.3) can only take place due to the difference in redox potential between the two; with the electron acceptor having a larger redox potential than the electron donor. Complexes I, III and IV use the energy released from the transfer of electrons along the electron transport chain to pump protons out of the mitochondrial matrix into the intermembrane space (Hatefi 1985, Wallace an Starkov 2000, Dykens and Will 2008). Complex V, the terminal complex involved in oxidative

phosphorylation, utilises the electrochemical gradient produced to transfer protons from the intermembrane space back into the mitochondrial matrix. The energy released from this action is used to phosphorylate ADP into ATP (Nadanaciva and Will 2011).

1.6 Mechanisms of mitochondrial dysfunction

There are five main mechanisms that have been associated with the induction of mitochondrial dysfunction: inhibition of the electron transport (respiratory) chain, uncoupling of oxidative phosphorylation, induction of membrane permeability transition, (in)direct inhibition of β-oxidation of mitochondrial fatty acids, and interfering with mitochondrial DNA. Chemicals that inhibit the electron transport chain can do so by either direct binding to the complexes of the electron transport chain or ATP synthase, or by acting as an alternative electron acceptor (Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005). A large number of mitochondrial toxicants have been shown to bind directly to various sites within the electron transport chain complexes, e.g. rotenoids. This binding blocks the ability of the complex to interchange between the oxidised and reduced state, thus blocking electron transfer down the electron transport chain (Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005). Additionally, chemicals that are observed to act as alternative electron acceptors compete with the natural electron carrier(s), ubiquinone and/or cytochrome *c*, in order to liberate electrons from the electron transport chain in order to reduce themselves (Wallace 2003). The inhibition of electron flow along the electron transport chain by both of these mechanisms induces the formation of reactive oxygen species (Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005). The increased production of reactive oxygen species within the mitochondrion results in oxidative stress; ultimately, leading to cell death.

The second mechanism of mitochondrial toxicity is induced by uncoupling of oxidative phosphorylation, due to protons re-entering the mitochondrial matrix through the inner mitochondrial membrane and bypassing ATP synthase (Wallace and Starkov 2000, Chan *et al* 2005, Naven *et al* 2013). During uncoupling of oxidative phosphorylation the flow of

electrons along the electron transport chain and associated translocation of protons into the intermembrane space functions as normal. However, the protons return to the matrix through the inner membrane bypassing ATP synthase, resulting in the loss of ATP production. A number of studies have proposed a general, protonophoric mechanism via which a chemical's uncoupling action can induce mitochondrial toxicity (Wallace and Starkov 2000, Kadenbach 2003, Mehta *et al* 2008, Naven *et al* 2013). Uncouplers act by scavenging a proton from the intermembrane space, migrating across the inner membrane into the matrix due to the membrane potential, where the chemical is deprotonated within the relatively more alkaline mitochondrial matrix. The deprotonated chemical can then return to the intermembrane space continuing the cycle. This assisted transport of protons back into the matrix dissipates the electrochemical potential, resulting in the loss of ATP production and ultimately cell death (Terada 1990, Schonfeld *et al* 1992, Sun and Garlid 1992, Wallace and Starkov 2000, Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005, Spycher *et al* 2008, Cela *et al* 2010). Many chemicals that act to uncouple oxidative phosphorylation are lipophilic weak acids or bases, such as 2,4-dinitrophenol (Wallace and Starkov 2000, Krahenbuhl 2001, Amacher 2005).

The third mechanism of mitochondrial toxicity is the induction of the membrane permeability transition. The membrane permeability transition is an increase in the permeability of the inner mitochondrial membrane to low molecular weight solutes (<1500Da), which leads to a disruption of the electron transport chain, loss of membrane potential, and swelling of both the inner- and outer mitochondrial membranes (Kroemer *et al* 2007, Lemasters *et al* 2009). It is believed the membrane permeability transition is induced by the formation and opening of the membrane permeability transition pore. The membrane permeability transition pore is a mega channel complex consisting of at least three proteins - voltage-dependent anion channel, adenine nucleotide translocator, and cyclophilin D - that spans both of the mitochondrial membranes.

The fourth mechanism is that of (in)direct inhibition of β-oxidation of mitochondrial fatty acids (Krahenbuhl 2001, Amacher 2005). Fatty acid β-oxidation is the multi-step process of breaking down fatty acids into Acetyl Co-A, which can then be used by the citric acid cycle to reduce NAD^+ and FAD into NADH and $FADH_2$. Inhibition of this process reduces the amount of NADH and $FADH_2$ available for oxidative phosphorylation and, therefore, reduces ATP production. In addition, because fatty acids are not broken down by the fatty acid cycle, they accumulate within cells inducing toxicity such as hepatic steatosis (Jaeschke *et al* 2002, Pessayre *et al* 2008).

The final mechanism involves interfering with mitochondrial DNA, for example by inhibition of mitochondrial DNA replication, impairing mitochondrial DNA stability or inhibiting mitochondrial DNA transcription (Amacher 2005, Pessayre *et al* 2008). As mitochondrial DNA encodes for 13 components of the electron transport chain, damage that occurs to mitochondrial DNA can have a wide variety of downstream effects depending upon where it occurs. For example, if the damage leads to ill-formed respiratory chain complexes, electron flow can be inhibited leading to a decrease in the production of ATP and could ultimately lead to cell death (Pessayre *et al* 2008). Given the almost ubiquitous distribution of mitochondria throughout the body, and the central role mitochondria play in regulating normal physiological functioning, it is perhaps not surprising that mitochondrial dysfunction has been suggested as the cause of toxicity within a multitude of organs (Amacher 2005, Dykens and Will 2007, Dykens and Will 2008, Nadanaciva and Will 2011, Pessayre *et al* 2012).

1.7 SEURAT-1 and the COSMOS project

In January 2011, as the final deadline for the Cosmetic Directive's 7th amendment approached, a cluster of European Union projects known as SEURAT-1 (Safety Evaluation Ultimately Replacing Animal Testing) was initiated. SEURAT-1 is a collaboration of six research projects, and one co-ordination project, combining the efforts from over 70

European universities, research institutes and commercial companies from across the EU and US. The five-year goal of SEURAT-1 is the development of solutions for the replacement of current *in vivo* repeat-dose toxicity testing with suitable *in vitro* and *in silico* alternatives. One of the six research projects that form SEURAT-1 is COSMOS. The COSMOS project is jointly funded by both the 7th Framework Programme of the European Commission and Cosmetics Europe, the industry's trade association for cosmetics, toiletries and perfumes. The main objectives of COSMOS are:

- The collation and curation of new sources of toxicological data from regulatory sources and the literature;
- The creation of an inventory of cosmetic ingredients, populated with chemical structures and toxicological data; and
- The development of freely available and accessible computational models to assist with predicting the repeated dose toxicity of cosmetic ingredients towards humans.

Together, the COSMOS project and collaborators within the SEURAT-1 cluster are part of the largest effort to develop alternative tools to help in the safety assessment of chemicals (with a focus on cosmetics related chemicals) within the EU. It is envisaged that these tools could be implemented as a partial replacement of *in vivo* repeat-dose testing. Additionally, the outcomes of each of the six research projects are thought to be able to provide a foundation on which further investigation, and development, surrounding alternative tools and techniques may be made. This PhD project whilst falling under the aims of the COSMOS projects supports the general ethos of the SEURAT-1 cluster by providing frameworks, and tools, for read-across.

1.8 Research aims of this thesis

In keeping with the research aims of the COSMOS project, the overall aim of this thesis was to develop an *in silico* profiler to assist in the safety assessment of repeated dose exposure of cosmetic ingredients to humans. The specific objectives to achieve this aim were to:

i. Collate toxicological data from regulatory submissions for use within the COSMOS database.
   - A standard operating procedure, developed by colleagues from the COSMOS project, is described in detail in Chapter 2. This approach was followed to extract data from EU regulatory submissions and input into the ToxRefDB data entry system, ready for upload into the COSMOS database;

ii. The development of chemistry-based structural alerts that can be used for prioritising chemicals for mitochondrial toxicity.
   - The ChemoTyper software was utilised in Chapter 3 to identify structural fragments, present within a data set from the available literature, associated with mitochondrial toxicity. Subsequently, these structural fragments were used in order to develop chemistry-based alerts;

iii. Utilise qualitative mitochondrial toxicity data in order to identify, and develop, mechanism-based structural alerts for use in category formation.
   - The data set used in Chapter 3 underwent structural similarity, and subsequent mechanistic, analysis in Chapter 4 in order to develop additional mechanism-based structural alerts. This analysis was performed to demonstrate the various approaches that can be used on the same data set to identify structural alerts;

iv. Demonstrate how repeat dose toxicity data, from regulatory submissions, can be used to develop structural alerts capable of category formation.

- The regulatory submission documents contain many valuable toxicological data. Therefore, in Chapter 5, it was demonstrated that data for a variety of chemicals used within hair dye products could be utilised to hypothesise and develop mechanism-based structural alerts;

v. Show how *in vitro* and *in chemico* experimental data can be used in the identification and verification of the chemical domain of structural alerts.

- After the development of structural alerts it is necessary that they be verified. Chapter 6 outlines an approach, using structural alerts related to protein binding and *in vitro* and *in chemico* assay results, which may be used to verify, and further refine, previously developed structural alerts.

Chapter 2: Extraction and collation of repeat dose toxicity data from Scientific Committee on Consumer Safety reports into the COSMOS database

2.1 Introduction

The availability of, ideally, high quality data relating to chemical(s) of interest is a prerequisite for undertaking any *in silico* modelling. *In silico* models can be developed for a multitude of endpoints as long as reliable data are available (Cronin 2010). The type of data that are required is wholly dependent on the endpoint under investigation and the type of model being developed. These data can be split into three broad categories: data relating to the identity and structure of the chemical, e.g. CAS number, chemical name, pictorial representation; data relating to the physico-chemical properties of the chemical, e.g. octanol-water partition coefficient, molecular weight, surface area; and toxicity data relating to the activity of the chemical in biological systems and/or assays. These toxicity data can be further subcategorised as being either quantitative, i.e. identifying the concentration at which an effect is seen, or qualitative, i.e. identifying the presence or absence of an effect. There are a number of ways in which these toxicological data can be utilised in the context of *in silico* approaches, such as the development of a (Q)SAR, as a component of an integrated testing strategy, in the development of a profiler, in category formation and/or read-across. An *in silico* profiler is considered to be a collection of structural alerts that relate to the same outcome. These approaches are discussed in more detail in Chapter 1. The type of data that are required differs depending upon the approach being employed. For example, a 'traditional' (Q)SAR will require a set of continuous quantitative data. Whereas, in the development of a profiler, i.e. a tool that enables the identification of a Molecular Initiating Event associated with a specific endpoint (Chapters 1, 3, 4, and 5), that can be used to screen for a certain endpoint, or organ level toxicity, then a set of qualitative data can be used. In order for these data to be useful for both hazard/risk assessment and modelling purposes

they should, ideally, be of high quality. This is due to the fact that the reliability of a prediction made by an *in silico* model can only be as good as the data used in the model's development, i.e. a high quality prediction cannot be derived from low quality data. A number of schemes have been developed to help assess data quality, with the scheme set out by Klimisch *et al* (1997) becoming the most widely used (Przybylak *et al* 2012). Under the Klimisch scheme the assessment of data quality falls under three headings; adequacy, relevance and reliability; these were defined by Klimisch as (Klimisch *et al* 1997):-

- Adequacy – the definition of the usefulness of data for hazard/risk assessment purposes. When there is more than one set of data for each effect, the greatest weight is attached to the most reliable and relevant
- Relevance – the extent to which data and/or tests are appropriate for particular hazard identification or risk characterisation
- Reliability – evaluation of the inherent quality of a test report or publication relating to preferably standardised methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings

Both the relevance and adequacy of data are context dependent, i.e. relevance identifies if data generated are appropriate for the endpoint of interest (for example the use of a protein binding assay to test for the potential for skin sensitisation), whilst adequacy identifies if data can be used to help inform a risk assessment decision. In contrast, reliability only identifies if the data are plausible in terms of the specific experimental procedure carried out. The final outcome of the Klimisch scheme is the assignment of the data to one of four categories for reliability (Table 2.1).

Table 2.1: Reliability categories as defined by Klimisch *et al* (1997)

Code	Category
1	Reliable without restriction "Studies or data from the literature or reports which were carried out or generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline (preferably performed to GLP) or in which all parameters described are closely related/comparable to a guideline method."
2	Reliable with restriction "Studies or data from the literature, reports (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable."
3	Not reliable "Studies or data from the literature/reports in which there are interferences between measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g. unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for an assessment and which is not convincing for an expert judgement."
4	Not assignable "Studies or data from the literature, which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.)"

In order to undertake any of the *in silico* approaches mentioned above, or discussed in Chapter 1, one of the preliminary steps is to identify relevant sources of available data, which fall under two headings; in-house and publically available (Madden 2013). In-house data sources are those held by private companies and, as such, are not readily available to those outside of these companies. Alternatively, there is an abundance of publically available data sources that hold a variety of toxicological information including:

- The scientific literature, whereby journal articles may contain toxicological information for one, or many, chemical(s). These articles may contain data for chemicals carried out at one laboratory as a series of interlinked studies within a specific assay, or they may provide a compilation of data extracted from various articles;

- Internet resources that have collated a number of QSAR and/or toxicity datasets, such as the Cheminformatics and QSAR Society (www.qsar.org), which provides links to an array of available data sets. These resources can be useful for modelling purposes as the data they contain are often times well curated; and
- Regulatory submissions, such as those submitted to the European Chemicals Agency (ECHA) or the Scientific Committee on Consumer Safety (SCCS). Both of which provide information regarding a chemical's toxicity profile in a variety of different assays. The information provided by ECHA is as an online dossier summarising the studies undertaken and their results, whilst the SCCS provide the chemical information as a downloadable PDF.

Whilst these are very useful sources of toxicological data one challenge of these data sources is the perception of a 'lack of data' (Yang *et al* 2006, Richard *et al* 2008). This 'lack of data' covers both the fragmentation of toxicological data across these data sources, and that the data are not usually structured in a manner that enables them to be readily utilised for modelling purposes, i.e. the information is usually contained as free-form text (Cronin 2002, Myshkin *et al* 2012). Free-form text is words and sentences whereby any character can be input by the user. The use of free-form text, within a toxicological database, does not lend itself well to searching and modelling purposes. This is because multiple words could be used to describe the same toxicological effect, e.g. hypersalivation or ptyalism. Therefore, collating toxicological information from these sources can be a time consuming process. To overcome this issue, over the past decade, there has been an increase in the compilation of toxicity data for the development of toxicity databases.

A toxicity database is a large, organised set of data regarding the toxicity profile for certain endpoint(s) of interest that are associated with a chemical structure (Valerio Jr. 2009). These data should be held in an electronic, and searchable, format utilising both a controlled vocabulary and ontology (Yang *et al* 2006). In this instance, i.e. in a toxicity database, a controlled vocabulary is a way of collating multiple terms for the same effect under one

phrase, for example collating fatty liver, hepatic steatosis and non-alcoholic fatty liver under the term liver steatosis. The use of a controlled vocabulary enables one search to be undertaken to retrieve information that would, otherwise, require the use of multiple search terms. This is an important advantage as it only requires the end user to know one search term, rather than attempting to identify all possible terms for a toxicological effect in order not to miss chemicals held in the database. Ontology, meanwhile, is used to identify the hierarchical relationship between the data at various levels within a database (Yang *et al* 2006, Bard 2007, Richard *et al* 2008). This provides a foundation to organise and standardise the chemical and toxicological data and how it relates across various aspects of the database, thereby facilitating easier data retrieval (Hardy *et al* 2012). Another advantage of ontology use is that it enables the unification of data extracted from multiple sources. The classification of the different regions within the liver - i.e. the centrilobular, midzonal, and periportal lobules - is an example of an anatomical ontology. This type of anatomical ontology serves to organise information with regards to the hierarchy of structures within an organ.

A number of freely available, and commercial, toxicity databases and datasets covering a wide variety of toxicological endpoints are highlighted in Table 2.2. However, it should be noted that due to the plethora of databases and datasets available this is not an exhaustive list. As the descriptions in Table 2.2 show, each of the databases and datasets listed holds information pertaining to a variety of toxicities at either the organ or organism level. For example, the TETRATOX dataset (available at http://www.vet.utk.edu/TETRATOX/index.php, accessed 17.11.2014) contains toxicity potency, i.e. 50% inhibition growth concentration, information from a single laboratory, acute toxicity studies performed at the University of Tennessee (Schultz 1997). The advantage of studies performed at the same laboratory is that if the experiments are replicated the results are more likely to be consistent than if they were undertaken at multiple laboratories, i.e. they have a high level of repeatability. A previous study by Hewitt

and colleagues (Hewitt *et al* 2011) demonstrated that the data generated by the TETRATOX assay are of a very high quality, due to the high level of repeatability between replicates, i.e. there is a low level of variability in the data the assay generates. This high level of data quality has led to the TETRATOX dataset, and the information it contains, being utilised in the development of a variety of *in silico* approaches, such as the development of (Q)SARs, and the identification and verification of structural alerts (Schultz *et al* 2007, Ellison *et al* 2008, Nelms *et al* 2013, Richarz *et al* 2013, Rodriguez-Sanchez *et al* 2013). Generally, sources of acute toxicity data are more abundant than those for repeat dose toxicity. The reason for this disparity is due to the extra time, money, and animal usage that are required to perform a repeat dose toxicity study. For example, using information from one laboratory, an acute oral toxicity study, following OECD test guideline 425, uses a maximum of five animals and costs approximately $1700. In comparison, to undertake a 90-day oral toxicity study, according to OECD test guideline 408, uses at least 80 rats (10 rats/sex/group, four groups needed) and costs approximately $164,000. The substantial difference in these three factors makes it difficult to justify performing multiple repeat dose studies. Therefore, outside of the repeat dose databases held by pharmaceutical companies, there are relatively fewer freely, and commercially, available databases that contain repeated dose toxicity testing information. Generally these databases include, but are not limited to, No Observable (Adverse) Effect Level (NO(A)EL) and/or Lowest Observable (Adverse) Effect Level (LO(A)EL) data. Of the databases highlighted in Table 2.2 several are commercially available and contain a number of repeat dose toxicity data, for example Leadscope (www.leadscope.com). There are a number of advantages to using commercial databases, one of the most important, in terms of predictive toxicology, is that of data curation and quality checking, i.e. data contained within a commercial database should be highly curated and have undergone quality assessment. Another advantage is that the database may contain a large number of toxicity data that may not be freely available in the public domain. Additionally, as professional products, the databases themselves are usually well-supported and the data held by the database are updated. However, as these are commercial endeavours,

a disadvantage of these databases is that gaining access to the information held within can, potentially, be quite expensive. In contrast, the main advantage of freely available databases, as their name suggests, is that the information they contain is freely accessible. The main disadvantages of this class of databases are that 1) the data may not have undergone curation or quality checking, and 2) it may be that only information relating to one chemical at a time can be searched, making extracting a dataset of usable data for modelling purposes a time consuming process. It should also be noted that whilst the majority of repeat dose toxicity databases contain NO(A)EL and/or LO(A)EL data, only a limited number also contain organ level effect data associated with the LO(A)EL value, for example Fraunhofer RepDose and Leadscope (Bitsch *et al* 2006). This scarcity of available databases, which are able to link chemical structure to an adverse effect within a specific organ, represents one of the key problems of modelling repeat dose toxicity data.

Without the presence of organ level data, no association can be made between the LO(A)EL and the organ that has been affected. This presents a problem when attempting to identify, and link, structural features of a chemical that are associated with certain adverse effects (e.g. liver fibrosis or kidney necrosis). To overcome this problem, one of the main goals of the COSMOS project (discussed in Chapter 1) is the development of a single, comprehensive, and freely available database that links chemical structure to repeat dose toxicity data, including organ level toxicity data. The inclusion of the organ level data provides modellers with an extremely useful tool that can be used to undertake a variety of *in silico* approaches. In addition, novel repeat dose toxicity data have been extracted from 'Opinions On' reports submitted to the Scientific Committee on Consumer Safety[1] (SCCS), and used to create the oral repeat dose toxicity (oRepeatToxDB) dataset that constitutes part of the COSMOS database. This enables the COSMOS database to be used for a variety of *in silico* approaches, such as; data mining, the development of mechanism based models, grouping chemicals and developing categories for use in read-across and to build profilers. These reports consist of a number of different toxicity studies, including acute, (sub)chronic, and developmental

[1] Formerly the Scientific Committee on Cosmetic Products and Non-food products intended for Consumers (SCCNFP) and the Scientific Committee on Consumer Products (SCCP)

toxicity, that have been undertaken for a specific chemical and submitted to the SCCS in order for a health and safety assessment to be carried out. The SCCS reports are an extremely useful source of repeat dose toxicity data as they provide detailed information on the study design, test substance, species the test was performed on, dosage levels, and route of administration. Possibly the most important information, from a modelling perspective, contained in the reports is the detailed organ level results observed at the different dosage levels. Therefore, as the extraction of this information was carried out by several people from the COSMOS consortium, in order to be consistent it was crucial that a standard operating procedure was followed. In the context of data extraction from the SCCS reports, the standard operating procedure was used to ensure the same level of information was extracted from each report.

The primary aim of this chapter is to describe how the extraction of the data held in the SCCS reports was carried out and entered into the data entry system of the oRepeatTox dataset. In addition, the secondary aim of this chapter is to demonstrate one application of the COSMOS database, i.e. exploiting the 28- and 90-day toxicity data for chemicals in order to ascertain whether data gathered from a 28-day study would negate the need to undergo a 90-day, thereby reducing animal usage.

Table 2.2: A list of freely, and commercially, available databases containing toxicity data (adapted from Madden 2013)

Available database	Description	Reference
Birth Defects Systems Manager (BDSM)	Open-source software to consolidate information regarding developmental toxicity	http://systemsanalysis.louisville.edu/
Carcinogenic Potency Database	Contains standardised information regarding carcinogenic bioassay results for over 6,500 chronic animal cancer tests performed on over 1,500 chemicals held in the available literature	http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CPDB.htm
ChEMBL	Database containing data for over 12 million activities and 1 million assays for over 1.36 million chemicals. An attempt has been made by the developers to standardise the information contained in ChEMBL.	https://www.ebi.ac.uk/chembl/
ChemIDPlus	Web-based database containing information, including physico-chemical and toxicity information on over 400,000 chemicals	http://chem.sis.nlm.nih.gov/chemidplus/
ChemSpider	An amalgamation of data from over 450 data sources and for more than 30 million unique structures, providing physico-chemical information and toxicological data from various species and different routes of administration	http://www.chemspider.com
COSMOSdb	Freely available database containing over 81,000 chemical records and over 44,500 unique chemical structures. Also contains two datasets (US FDA PAFA and oRepeatToxDB) that hold information for 12,538 toxicological studies across 27 endpoints for 1,660 compounds.	http://cosmosdb.cosmostox.eu/
DevTox	Standardises terminology used to describe developmental toxicities and provides a historical control database of developmental toxicity studies	http://www.devtox.org/datintro.htm
Drugs@FDA	US FDA-approved drug products	http://www.fda.gov/Drugs/InformationOnDrugs/ucm135821.htm
DSSTox	US EPA Distributed Structure-Searchable Toxicity database provides a 'public forum for publishing downloadable, structure-searchable, standardized chemical structure files associated with chemical inventories or toxicity data sets of environmental relevance'	http://www.epa.gov/ncct/dsstox/
ECOTOX	US EPA database of single chemical toxicity information for aquatic and terrestrial life	http://cfpub.epa.gov/ecotox/
eTox	**The eTox project is developing a drug safety database from pharmaceutical industry legacy toxicology reports and public toxicology data**	www.etoxproject.eu/

Fraunhofer RepDose	Repeated Dose toxicity is a relational database containing more than 2,200 studies on subacute to chronic toxicity within a variety of routes of administration for about 650 chemicals.	http://www.fraunhofer-repdose.de/
HESS	Hazard Evaluation Support System contains information regarding repeat dose toxicity and toxicity mechanisms. Also supports the evaluation of repeat dose toxicity by utilising category formation	http://www.safe.nite.go.jp/english/kasinn/qsar/hess-e.html
IRIS	Integrated Risk Information System is an aggregation of electronic reports on environmental substances and their potential to cause human health effects	http://cfpub.epa.gov/ncea/iris/index.cfm
ITER	International Toxicity Estimates for Risk Assessment is a database of human health risk values and cancer classifications for over 680 chemicals of environmental concern	http://www.tera.org/iter/
Leadscope	Provides commercial databases containing over 400,000 results covering acute, (sub-) chronic, carcinogenicity, genotoxicity, and reproductive toxicity studies for nearly 180,000 chemicals	http://www.leadscope.com/toxicity_databases/
MDL Toxicity database	A commercially available structure-searchable database containing data from both *in vitro* and *in vivo* studies covering acute, carcinogenicity, mutagenicity and reproductive toxicity studies for over 150,000 chemicals. Also includes information from Registry of Toxic Effects of Chemical Substances (RTECS)	http://www.iop.vast.ac.vn/theor/conferences/smp/1st/kaminuma/ChemDraw/toxicity.html
NTP	National Toxicology Program provides information of agents registered in the US that are of public interest	http://ntp.niehs.nih.gov/
OECD eChemPortal	Access to information on physico-chemical properties, environmental fate and toxicity	webnet3.oecd.org/echemportal/
OECD QSAR Toolbox	The Toolbox is a software application that incorporates data from various data sources for a variety of human health and environmental endpoints. For example it contains repeat dose toxicity information from Fraunhofer RepDose, HESS and ToxRefDB.	http://www.qsartoolbox.org
OSIRIS	The OSIRIS project collated data on aquatic toxicity, carcinogenicity, mutagenicity and repeat dose toxicity	www.osiris.ufz.de
Tox21	US EPA Tox21 is currently screening over 10,000 chemicals at the National Institutes of Health using the ToxCast high throughput screening assays to provide risk assessors with data for use when making decisions about protecting human health and the environment	http://www.epa.gov/ncct/Tox21/
ToxCast	US EPA is using various high throughput screening assays to measure changes in biological activity. Currently ToxCast has evaluated over 2,000 chemicals within over	http://www.epa.gov/ncct/toxcast/

	700 high throughput assay covering roughly 300 signalling pathways	
ToxRefDB	Contains information of over 30 years and $2 billion worth of historical *in vivo* study results including acute, (sub-)chronic, developmental and reproductive endpoints for 474 chemicals ToxRefDB also links with both ACToR and ToxCast databases.	http://www.epa.gov/ncct/toxrefdb/
TETRATOX	A collection of acute aquatic potency data for more than 2,400 industrial organic chemicals	http://www.vet.utk.edu/TETRATOX/index.php
TOXNET	US National Library of Medicine Toxicology Data Network is a group of databases covering chemicals and drug, environmental health, occupational safety, risk assessment and regulations, and toxicology	http://toxnet.nlm.nih.gov/
US EPA ACToR	Aggregation of data from over 1,000 public sources for over 500,000 environmental chemicals. Contains chemical structure, physico-chemical properties, and *in vitro* assay and *in vivo* toxicology data	http://actor.epa.gov/actor/faces/ACToRHome.jsp
US FDA Chemical Estimation Risk Evaluation System (CERES)	A centralised, sustainable data management, and storage, system that will provide support in decision making for both pre- and post-market safety assessment for food ingredients and food contact substances.	http://www.accessdata.fda.gov/FDATrack/track-proj?program=cfsan&id=CFSAN-OFAS-Chemical-Evaluation-and-Risk-Estimation-System
VITIC Nexus	A not for profit database and information management system providing information for a variety of toxicological endpoints including carcinogenicity, mutagenicity, and hepatotoxicity	http://www.lhasalimited.org/products/vitic-nexus.htm

2.2 Method

The harvesting of data from the SCC(NF)P/SCCS reports, recording of data using the ToxRefDB data entry tool, and collation of data into a Microsoft Excel spreadsheet, were performed by the current author and collaborators from the COSMOS project, using the standard operating procedures developed by Chihae Yang (Altamira, LLC) and Vessela Vitcheva (Medical University Sofia, Bulgaria), which are described below.

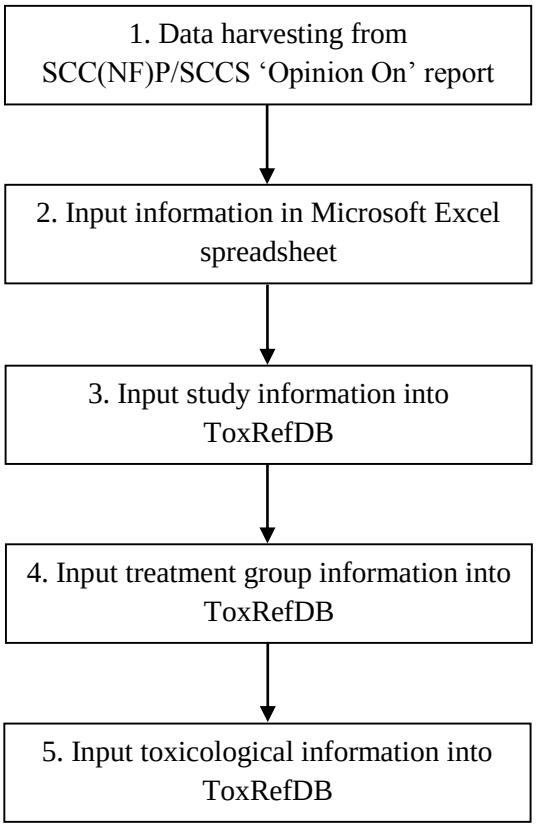


Figure 2.1: Workflow identifying the key steps in the standard operating procedure for repeat dose data collection.

The workflow above (Figure 2.1) outlines the key steps in the standard operating procedure for repeat dose data collection. The first stage is the identification of the study producing the NO(A)EL in the most recent SCC(NF)P/SCCS 'Opinion On' report for the chemical in question. It is important to use the latest report as new studies performed may have a new NO(A)EL value or extra information in them. The second stage is to input the study and

NO(A)EL/LO(A)EL information into the relevant tabs in the Microsoft Excel spreadsheet. Subsequently, the study, treatment group and toxicological effect information were each input separately into the ToxRefDB data entry system. It is important that the data entry system be used to populate the oRepeatTox database as it provides a process by which data harvesters can input the information in a systematic way. In addition, the controlled vocabulary for the input of the toxicological effects ensures consistency when inputting the same effect, which may have multiple terms, under one phrase. This allows for easier searching of the database by the end user as only one search term needs to be known, and utilised, in order to identify all chemicals with data to the specific effect. In total, the SCC(NF)P/SCCS reports for 154 chemicals were used for data mining purposes in order to extract the NO(A)EL, LO(A)EL (if present), and the adverse effects observed at the LO(A)EL. The majority of these chemicals were organic chemicals used in hair dyes, with a smaller proportion being used in sunscreens and as preservatives. These 154 chemicals were divided amongst three researchers, including the current author, each of whom was given approximately 45 reports to harvest. Within this chapter the chemical '3-aminophenol' will be used as an example of how the standard operating procedure was utilised (Figure 2.2) (SCCP 2006).

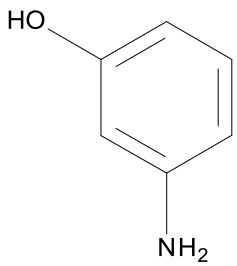


Figure 2.2: Structure of 3-aminophenol

Step 1 Data harvesting from the SCC(NF)P/SCCS 'Opinion On' reports

The following steps were undertaken to extract the NO(A)EL and LO(A)EL values, as well as the adverse effects associated with the LO(A)EL and other dose values.

Step 1.1 Identification of the NO(A)EL and/or LO(A)EL values

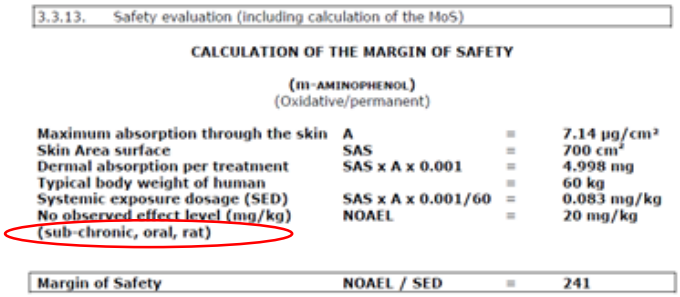
Upon finding the most recent SCC(NF)P/SCCS report for the chemical in question the first requirement was to identify the study that provided the NO(A)EL/LO(A)EL value. This information is usually provided under the heading "Safety evaluation (including calculation of the MoS)" present towards the end of the report (Figure 2.3).

Step 1.2 Identification of the study report used to generate the NO(A)EL/LO(A)EL

The information in parentheses (circled red in Figure 2.3) was used to identify the type of study that generated the NO(A)EL/LO(A)EL value, it was then required that this study be found within the report. It should be noted that only studies undertaken via the oral route of administration were used.

Step 1.3 Extraction of data from the study report

Upon finding the study report that generated the NO(A)EL value, the information contained within it was extracted. The information that was extracted included, but was not limited to: study type, study duration, dosage levels, and effects noted at each dosage level. This information was then entered into both the Excel spreadsheet and the Toxicity Reference DataBase (ToxRefDB) data entry tool, which are described below.



3.3.13. Safety evaluation (including calculation of the MoS)

CALCULATION OF THE MARGIN OF SAFETY

(m-AMINOPHENOL)
(Oxidative/permanent)

Maximum absorption through the skin	A	=	7.14 µg/cm²
Skin Area surface	SAS	=	700 cm²
Dermal absorption per treatment	SAS x A x 0.001	=	4.998 mg
Typical body weight of human		=	60 kg
Systemic exposure dosage (SED)	SAS x A x 0.001/60	=	0.083 mg/kg
No observed effect level (mg/kg) (sub-chronic, oral, rat)	NOAEL	=	20 mg/kg
Margin of Safety	NOAEL / SED	=	241

Figure 2.3: The ‘Safety evaluation’ section for 3-aminophenol extracted from the associated SCCP ‘Opinion On’ report

Step 2 Data recording in Microsoft Excel spreadsheet

The following steps were undertaken to input the information retrieved from the SCC(NF)P/SCCS report into a Microsoft Excel spreadsheet. The spreadsheet contained three datasheets; ‘study info’, ‘noael_loael’, and ‘reference’; where information provided by the report could be input.

Step 2.1 Input of information into ‘study info’ datasheet

The first datasheet was designed to capture information pertaining to the study from which the NO(A)EL/LO(A)EL was derived. The information to be input into this tab included: the study title, species of animal used, duration of study, route of administration, dose range, purity, Klimisch score if present in document, whether study was GLP compliant or not, and if a study guideline was followed.

Step 2.2 Input of information into ‘noael_loael’ datasheet

The second datasheet was designed for the input of data regarding the NO(A)EL and LO(A)EL of the study. The information to be input into this tab included: the source document, the NO(A)EL identified by the SCCS for use within the safety evaluation, the

LO(A)EL of the study (if one could be identified), the target organ, the critical effects observed at the LO(A)EL, and any comments regarding the NO(A)EL/LO(A)EL value.

Step 2.3 Input of information into 'reference' datasheet

The final datasheet was designed for the input of data regarding the reference of the study. The information to be input into this tab included: the type of reference the data were extracted from (e.g. regulatory document), the author of the study, the journal title, the study title, the volume and issue number, the start page, the year of publication, the document source, the report number, the year the study was conducted, the laboratory that performed the study, and if the data were extracted from a book the book title, chapter and publisher were required.

Step 3 Data recording in ToxRefDB data entry tool

The following steps were undertaken to input the information extracted from step one into the ToxRefDB data entry tool. The data entry tool was used so that all the information was input consistently, and so that the controlled vocabulary held within ToxRefDB could be utilised to build a consistent and more easily searchable database. The main page of the data entry tool required information pertaining to (Figure 2.4):

Step 3.1 Study identifiers

- Document source (SCCP)
- Document number (SCCP/**0978**/06)
- The year the study was performed (1996)
- The data usability (OECD Guideline)

Step 3.2 Test material information

- Chemical name (3-aminophenol)
- Purity percentage (100%)
- Lot/batch number (4090202)
- Test material comment (parent compound)

Step 3.3 Study type

- Type of study used (subchronic)
- Study duration (90 days)

Step 3.4 Animal dose information

- Species (rat)
- Strain (Sprague Dawley)
- Route of administration (gavage/intubation)
- Comments, information of dose, vehicle (Doses: 20, 70, 200, 600 mg/kg bw/day. Vehicle: 0.5% methylcellulose solution and 1% d-iso-ascorbic acid)

Step 3.5 Treatment group list

- Information regarding treatment groups and associated effects (described in more detail below)

Step 3.6 Reference

- The reference given in the report for the study in question

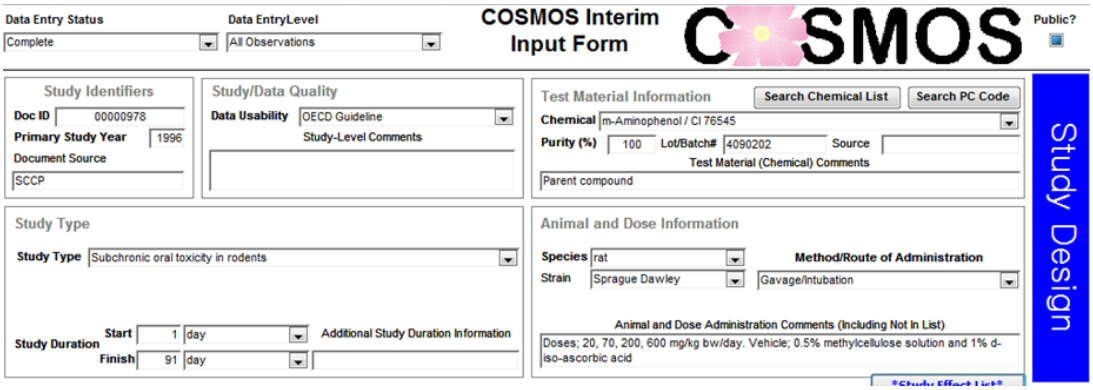


Figure 2.4: Study information required to be input into main page of ToxRefDB data entry system

Step 4 Entering treatment group and toxicological effect level information

These steps were undertaken in order to input the toxicological information at each dosage as described within the study; thereby, associating the toxicological information with the chemical structure

Step 4.1 Populating treatment group form

In order to populate the **Treatment Group list** the **Excel Treatment Group Form** hyperlink was pressed (circled red in Figure 2.5), opening the **Excel Treatment Group form** table. Within this table information related to the gender, dosing period, dose given, dose duration and number of treatment animals were input, if available within the study description (Figure 2.6).

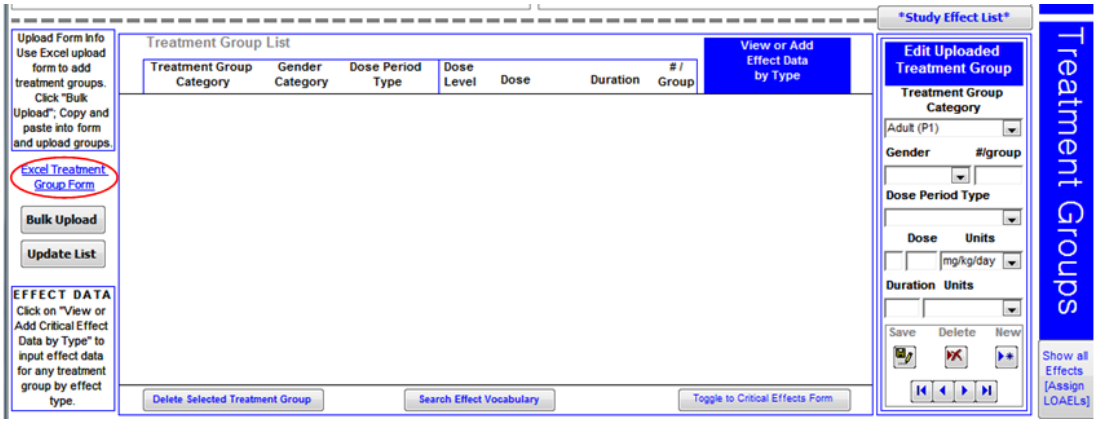
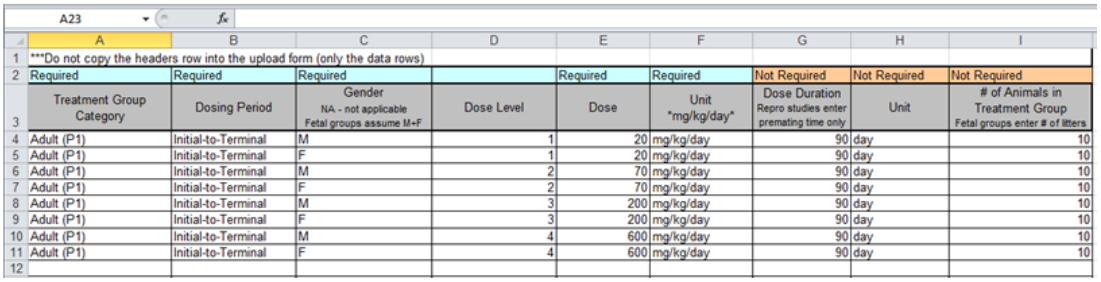


Figure 2.5: Hyperlink used to open the Excel treatment group form table (circled in red).

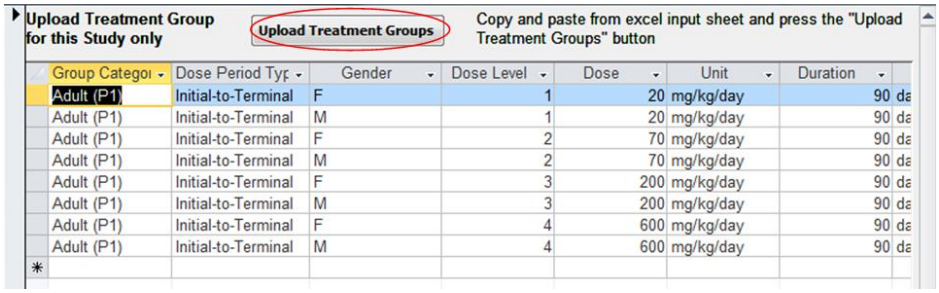


	A	B	C	D	E	F	G	H	I
1	***Do not copy the headers row into the upload form (only the data rows)								
2	Required	Required	Required		Required	Required	Not Required	Not Required	Not Required
3	Treatment Group Category	Dosing Period	Gender NA - not applicable Fetal groups assume M+F	Dose Level	Dose	Unit *mg/kg/day*	Dose Duration Repro studies enter premating time only	Unit	# of Animals in Treatment Group Fetal groups enter # of litters
4	Adult (P1)	Initial-to-Terminal	M	1	20	mg/kg/day	90	day	10
5	Adult (P1)	Initial-to-Terminal	F	1	20	mg/kg/day	90	day	10
6	Adult (P1)	Initial-to-Terminal	M	2	70	mg/kg/day	90	day	10
7	Adult (P1)	Initial-to-Terminal	F	2	70	mg/kg/day	90	day	10
8	Adult (P1)	Initial-to-Terminal	M	3	200	mg/kg/day	90	day	10
9	Adult (P1)	Initial-to-Terminal	F	3	200	mg/kg/day	90	day	10
10	Adult (P1)	Initial-to-Terminal	M	4	600	mg/kg/day	90	day	10
11	Adult (P1)	Initial-to-Terminal	F	4	600	mg/kg/day	90	day	10
12									

Figure 2.6: Excel table completed with treatment group information relating to 3-aminophenol

Step 4.2 Inserting treatment group information into Bulk Upload table

Once all the available treatment information had been entered into the **Excel Treatment Group Form** table it was copied, and pasted, into the **Bulk Upload** table (Figure 2.7). The **Bulk Upload** table was opened by pressing the **Bulk Upload** button present on the main page of the data entry tool (circled red in Figure 2.8).



Group Categor	Dose Period Typ	Gender	Dose Level	Dose	Unit	Duration	
Adult (P1)	Initial-to-Terminal	F	1	20	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	M	1	20	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	F	2	70	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	M	2	70	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	F	3	200	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	M	3	200	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	F	4	600	mg/kg/day	90	da
Adult (P1)	Initial-to-Terminal	M	4	600	mg/kg/day	90	da

Figure 2.7: Bulk upload table where the treatment group information is pasted before being uploaded to the main page of ToxRefDB

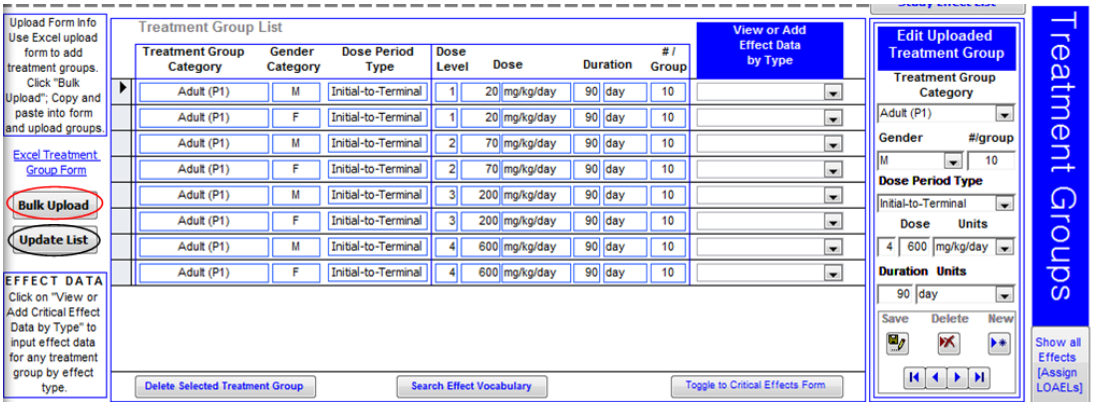


Figure 2.8: The buttons that were pressed in order to open the Bulk Upload table (circled in red) and to update the Treatment Group List (circled in black)

Step 4.3 Updating 'Treatment Group list' on the main page

After the information had been copied into the **Bulk Upload** table the **Upload Treatment Groups** button was pressed (circled red in Figure 2.7). Subsequently, pressing the **Update list** button on the main page uploaded into the **Treatment Group list** information onto the main page (circled black in Figure 2.8). After the **Treatment Group list** table was populated the effects noted within the study could be added.

Step 4.4 Inserting toxicological effects

The **Study Effect list** table was opened, by pressing the **Study Effect List** button on the main page, and the information about each dose group was checked to make sure it was correct before proceeding further.

Step 4.4.1 Entering effect type information

To input the effect level data the first step was to choose an effect from the study report and identify it within the drop down list under **Effect Type Target**, e.g. 'In-Life Observations' (Figure 2.9). The second drop down list underneath Effect Type Target was used to identify the organ affected (for local effects) or to further refine the effect type (for systemic effects), e.g. 'Body Weight'.

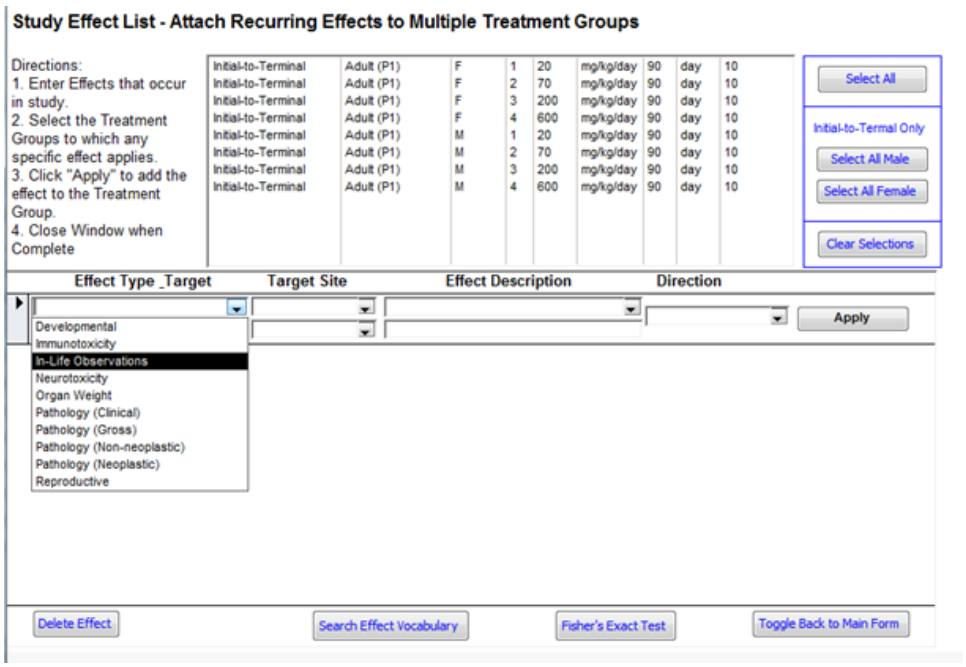


Figure 2.9: The drop down menu showing the categories of effects under the 'Effect Type_Target' heading

Step 4.4.2 Entering target site information

If the effect was observed within a specific region of an organ the **Target Site** drop down menu could be used to further refine where the observation was made within the organ.

Step 4.4.3 Entering effect description and direction of change

Next, in the drop down menu under **Effect Description** the specific effect discussed in the study report was chosen, e.g. 'Body Weight Gain,' the direction of change for the effect was then picked up in the adjacent drop down menu, e.g. 'Decrease' (Figure 2.10).

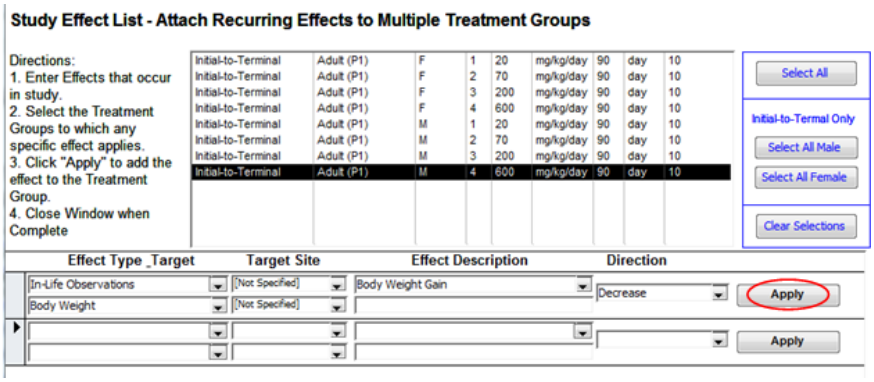


Figure 2.10: Entering effect description and direction of change information

Step 4.4.4 Applying information to treatment groups

Once all of the above information related to the effect was identified, using the drop down menus, the treatment group(s) with which this effect was associated was highlighted and the **Apply** button was pressed (circled red in Figure 2.10).

This procedure was followed for every effect and every direction within the study report. It should be noted that if there was no change for a particular effect for all treatment groups the effect was still included in the **Study Effect list** table. After entering all effects described in the study report the **Treatment Group list** table was closed.

Step 5 Entering the statistical significance and treatment relationship

The steps below were undertaken in order to input the information regarding whether or not an effect, for a particular dose and gender, was deemed to be statistically significant and/or treatment related within the SCC(NF)P/SCCS report.

Step 5.1 Opening the effect input form

Initially, the dose(s) and gender(s) to which a statistically significant and/or treatment related effect was observed were identified in the SCC(NF)P/SCCS document, e.g. only males in the 600 mg/kg bw/day dose showed a decrease in Body Weight Gain. To input this information into the data entry tool the effect type was selected from the drop down menu under the heading **View or Add Effect Data by Type** for the dose and gender in question, e.g. 'In-Life Observations' for the males in the 600 mg/kg bw/day (highlighted in black, Figure 2.11).

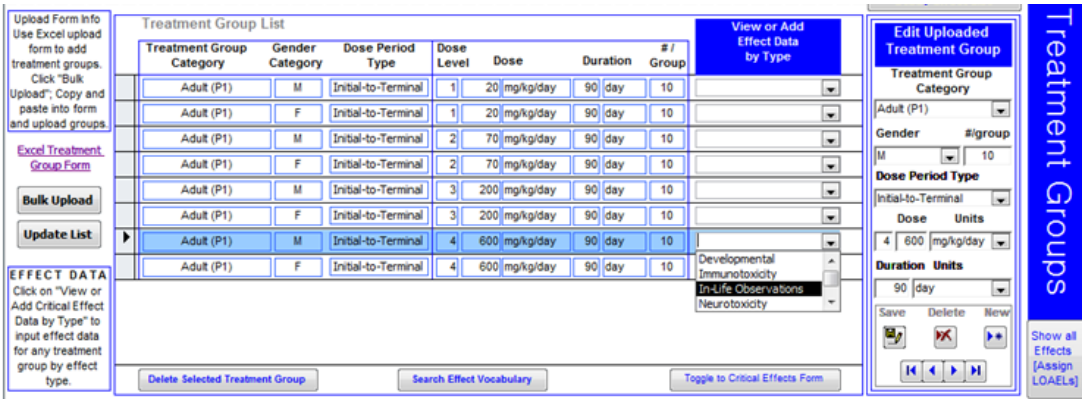


Figure 2.11: Identification of effect that is statistically significant and/or treatment related

Step 5.2 Identifying the endpoint affected

Subsequently, the **Effect Input form** appeared displaying all effects input previously under the **Effect Description** heading, e.g. body weight gain, anaemia, lacrimation, salivation (Figure 2.12). For every effect with an associated change it had to be stated, using the **Study Endpoint** drop down menu, which endpoint was affected, e.g. systemic (circled red in Figure 2.12). For most effects from (sub)chronic studies the endpoint affected was 'systemic', effects on dams from reproductive and developmental were classed as 'maternal' and effects affecting reproduction and development were 'reproductive' and 'developmental' respectively.

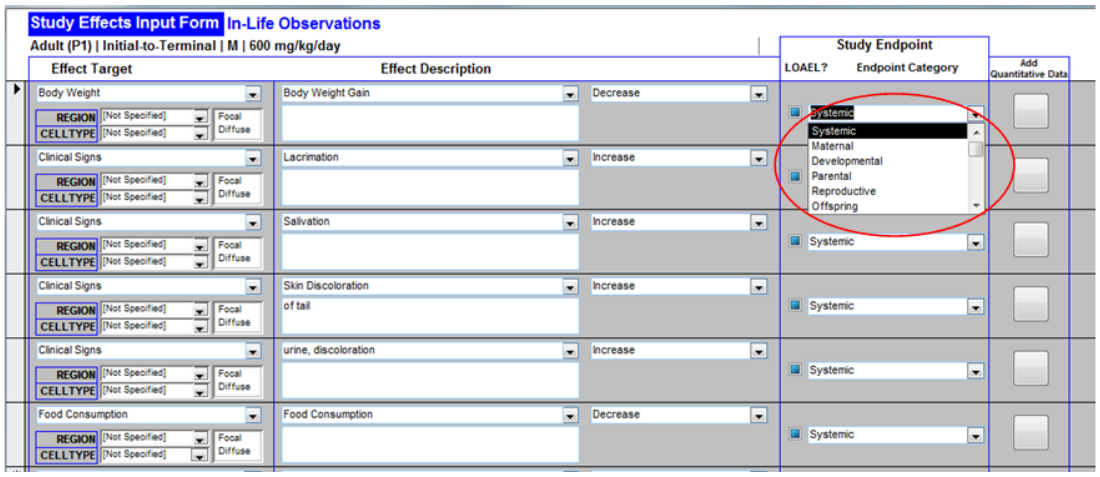


Figure 2.12: The **Effect Input** form with the Study Endpoint drop down menu circled in red

Step 5.3 Opening add quantitative data table and inputting statistical significance and treatment related effect data

Next, the **Add Quantitative Data** table was opened by pressing the corresponding button (circled red in Figure 2.13). It was within this pop-up table where statistical significance and treatment related information were input. If the study report stated the results were statistically significant and/or treatment related the blue box under 'SS' or 'TR' was clicked so that a tick-mark was present (circled red in Figure 2.14). The severity of the observed change was input in the 'severity' box, the 'comments' box was used for any additional information regarding the effect. After all the information was entered for the first effect type the **Add Quantitative Data** form and **Study Effects Input** form were closed by pressing 'Finished.' This procedure was repeated for each dose and gender where a change in the effect was observed.

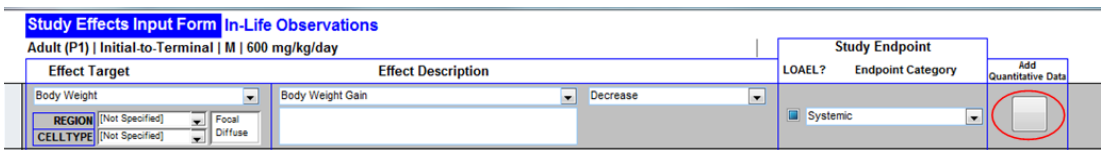


Figure 2.13: The button pressed to open the **Add Quantitative Data** table

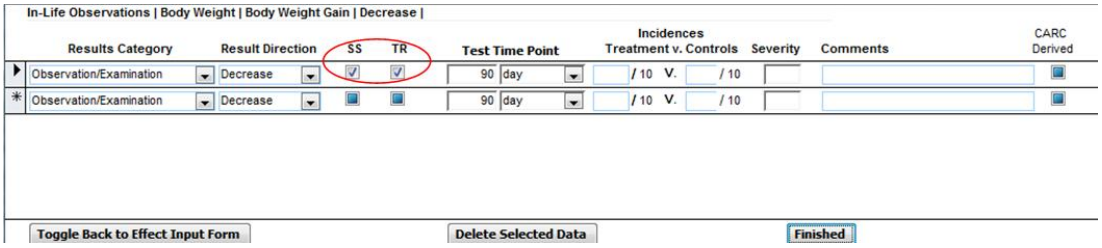


Figure 2.14: **Add Quantitative Data** Table used to identify if an effect was statistically significant and/or treatment related

This standard operating procedure was followed for each of the 154 chemicals that were extracted from the SCC(NF)P/SCCS reports. This procedure has been described in detail so as to illustrate the depth of information that has been captured, and reported, for each individual chemical. Hence, there is a wealth of detailed, novel information within the COSMOS database. The chemical and toxicological data held within the COSMOS database can be exploited for many purposes, one such example is that of ascertaining whether results from 28 day repeat dose studies were protective of results from 90 day repeat dose studies, i.e. if both studies had results at, or above, 1000mg/kg bw/day. This example is discussed in more detail below.

2.3 Application of the COSMOS database using 28- and 90-day repeat dose toxicity data

Whilst the previous section of this chapter discussed how novel toxicological data were extracted from regulatory dossiers and entered into the COSMOS database, the remainder of this chapter will demonstrate one application of how the data held within the COSMOS database can be utilised.

2.3.1 Introduction to COSMOS database

The current release of the COSMOS database provides access to over 12,000 toxicological studies covering 27 different endpoints for more than 1,600 chemicals. These studies are contained within two separate datasets: the US FDA Priority-based Assessment of Food Additives (PAFA) and the oRepeatToxDB. The US FDA PAFA dataset was kindly donated to the COSMOS project by the US FDA Office for Food Safety and contains 12,198 studies. The oRepeatToxDB, meanwhile, contains data for a variety of toxicological effects from 340 *in vivo* repeat dose studies for over 200 chemicals, including those extracted as part of the data harvesting process described above. Due to the quantity of toxicity data held in the COSMOS database and the presence of results for both 28-day and 90-day repeat dose studies therein, researchers at Liverpool John Moores University, including the current author, were approached by researchers from the British Union for the Abolition of Vivisection (BUAV) and the European Coalition to End Animal Experiments (ECEAE). The researchers from BUAV and ECEAE wanted to use this information from repeat dose studies in order to ascertain whether the results from 28-day repeat dose studies were protective of results from 90-day repeat dose studies. Researchers from both the UK Competent Authority for REACH (the Health and Safety Executive) and BUAV have previously undertaken similar studies using data from the Notification of New Substances (NONS) system and eChemPortal (echemportal.org) respectively (HSE 2011, Taylor *et al* 2014). In both instances a No Observed Adverse Effect Level (NOAEL) value of 1000 mg/kg body weight/day (mg/kg bw/day) was used as a profile for low toxicity. In the context of toxicological testing a NOAEL value is considered to be the highest tested dose of a chemical at which no adverse effect is observed. Both studies demonstrated that where a chemical had a profile of low toxicity in a 28-day repeat dose study it was highly likely that the chemical would have a profile of low toxicity in the 90-day repeat dose study, i.e. the results of the 28-day study were protective of the 90-day study. Therefore, the aim of this

section is to demonstrate one of the uses of the data held within the COSMOS database by replicating the studies performed by both the HSE and Taylor *et al* (2014).

2.3.2 Method

In order to maintain consistency with the previous studies conducted the method described by Taylor *et al* (2014) was followed as much as possible; this method is described below.

Step 1.1 Initial selection of the dataset

A property search of the OECD's eChemPortal (www.echemportal.org) was undertaken to identify those chemicals for which there was both 28- and 90-day repeat dose data. The initial search was undertaken using eChemPortal as performing a property search of the ECHA CHEM database was not possible at the time. However, the ECHA CHEM database is the main provider of property data to eChemPortal, therefore, a search of eChemPortal was effectively a search of the ECHA CHEM database. In order to undertake the search of eChemPortal the procedure below was followed:

1) 'Property search' was selected
2) Under the subheading 'Toxicological information' 'repeated dose toxicity:oral' was selected
3) On the Query Block page that opens the search criteria input included:
    a. Under the 'Study result type' dropdown menu only the box marked 'experimental result' was checked
    b. Under the 'Test guideline, Guideline' dropdown menu the following boxes were checked
        i. EPA OPP 82-1 (90-day oral toxicity)
        ii. EPA OPPTS 870.3100 (90-day oral toxicity in rodents)
        iii. EPA OPPTS 870.3150 (90-day oral toxicity in non-rodents)

iv. EPA OTS 798.2650 (90-day oral toxicity in rodents)
v. EU Method B.26 (sub-chronic oral toxicity test: repeated dose 90-day oral toxicity study in rodents)
vi. EU Method B.27 (sub-chronic oral toxicity test: repeated dose 90-day oral toxicity study in non-rodents)
vii. OECD guideline 408 (repeated dose 90-day oral toxicity in rodents)
viii. OECD guideline 409 (repeated dose 90-day oral toxicity in non-rodents)

4) The search criteria were then saved by pressing 'Save' and the query was subsequently executed by pressing 'Execute Query.'

Step 1.2 Inclusion criteria

A manual analysis was then conducted on the initial dataset to identify those chemicals that had oral, rat experimental data for both 28- and 90-day studies. This process involved opening the hyperlink in the 'Results' column for each chemical individually. In every case the results were held in the ECHA CHEM database. Chemicals for which there was no 28- or 90-day experimental study data, i.e. read across was used or no studies were included, were rejected at this stage.

Step 1.3 Exclusion criteria

The dataset was then reviewed to exclude all chemicals that had a 'Key study' with a No Observed Effect Level (NOEL) or NOAEL value of less than 1000mg/kg bw/day. Importantly, if a chemical had a NOAEL of 1000mg/kg bw/day but the NOEL value was lower these chemicals were also excluded. Additionally, where multiple key studies were present a chemical was excluded if any rodent study had a NOAEL below 1000mg/kg bw/day.

Step 1.4 Quality of data

Subsequently, the dataset was reviewed to remove those chemicals with poor quality data for either the 28- or 90-day studies. The criteria as to what constituted poor quality data, in the context of this study, are as follows:

- Any study which did not follow the OECD (or equivalent) guidelines for 28- or 90-day repeat dose testing with regards to duration, species, number of animals or main parameters measure were excluded.
- The study was not conducted up to the limit dose of 1000mg/kg bw/day.
- The study was conducted prior to 1981, as this was when both the OECD test guidelines were introduced.
- A study equivalent to the OECD guidelines but given a Klimisch score of 3 or 4 for any reason.

Step 1.5 Acute toxicity profile

Finally, chemicals were excluded if they did not have toxicity data conducted within acute toxicity, skin or eye irritation, skin sensitisation, or genotoxicity studies.

However, in order to perform the analysis in this chapter a few minor alterations to the method set out by Taylor *et al* (2014) were required. These alterations were required due to the variations in data held within the COSMOS database compared to eChemPortal, these being: the extraction of Highest No Effect Level (HNEL) data, as opposed to NOAEL data; a study completeness/Klimisch score of 'A, B or C' (as defined in the COSMOS database) was accepted in instances where only one HNEL value was present; and no acute toxicity profile was gathered. The alterations were necessary as 1) the current version (v1.0) of the COSMOS database does not contain NOAEL data; 2) if more than one HNEL value above 1000mg/kg bw/day was present the value with the highest Klimisch score was used; and 3)

information pertaining to acute toxic, skin/eye irritation, skin sensitisation and/or genotoxicity was not available for every chemical.

Step 2 Development of the 28- and 90-day repeat dose study dataset from the COSMOS database

The initial search of the COSMOS database was undertaken using the search term 'C'. Within the COSMOS database the search term 'C' identifies all chemicals containing a carbon atom. Additionally, queries were added to identify chemicals that also had a short-term and/or subchronic toxicological studies associated with them; returning a total of 618 chemicals. However, the majority of these chemicals contained only one study and were, therefore, disregarded at this stage. Subsequently, each chemical with more than one study in the '# studies' column was manually analysed in order to identify those chemicals containing toxicity data for both 28- and 90-day repeat dose studies conducted via the oral route in rats; reducing the dataset to 54 chemicals. The data gathered at this stage are available in the Appendix for this chapter (Appendix I). Further analysis was undertaken on the dataset to remove those chemicals that did not have a HNEL value at 1000mg/kg bw/day or greater for both the 28- and 90-day repeat dose study durations; reducing the final dataset to nine chemicals, constituting 15% of the dataset that contained both 28- and 90-day repeat dose study data (Table 2.3). This is in keeping with the previous studies conducted, whereby Taylor *et al* and the HSE identified 10% and 15%, respectively, of the original chemicals were present in the final dataset as having both 28- and 90-day repeat dose study data with a NOAEL value at, or greater than 1000mg/kg bw/day.

Table 2.3: A comparison of the 28- and 90- day repeat dose studies with Highest No Effect Level data at, or greater than 1000mg/kg bw/day

Chemical name	COSMOS ID	Short-term toxicity study value (mg/kg bw/day)	Length of short-term toxicity study (days)	Sub-chronic toxicity study value (mg/kg bw/day)	Length of sub-chronic toxicity study (days)	Is the short-term study protective of sub-chronic study?
Hydroxypropyl Methylcellulose	CMS-7567	HNEL 10,000	30	HNEL 2339 HNEL 5000 HNEL 6500 HNEL 7700 HNEL 505	90 90 90 90 91	Yes (if using Weight of Evidence) No (if using most protective value)
Ascorbic acid	CMS-108	HNEL 10,000	28	HNEL 2500	90	Yes
Glycyrrhizin, ammoniated	CMS-8524	HNEL 1000 HNEL 1000	30 35	HNEL 500	90	No
Sucrose acetate isobutyrate	CMS-5115	HNEL 2226 HNEL 2592	28 28	HNEL 5300	91	Yes
Maltodextrin	CMS-5576	HNEL 10,000	30	HNEL 3882	90	Yes
Butyl acetate	CMS-1941	HNEL 2000	28	HNEL 600	90	No
Hydroxypropyl cellulose	CMS-10327	HNEL 6000	30	HNEL 5000	90	Yes
Potassium bicarbonate	CMS-1189	HNEL 2132 HNEL 4000	28 28	HNEL 1482 HNEL 2000	91 91	Yes
Polyethylene, oxidised	CMS-34680	HNEL 4650	32	HNEL 5000 HNEL 5000	90 90	Yes

2.3.3 Results and discussion of the analysis performed on the dataset

Overall, the analysis carried out in this chapter identified six (67%), using the most protective HNEL value, or seven (77%), using a Weight of Evidence approach, of the nine chemicals that had a HNEL value at, or greater than, 1000mg/kg bw/day in the 28-day repeat dose study also had a value at, or greater than, 1000mg/kg bw/day in the 90-day repeat dose study, i.e. the 28-day study was protective of the 90-day study. The two chemicals that were not protective within the COSMOS database are butyl acetate and glycyrrhizin (ammoniated) (Table 2.3). The identification of butyl acetate as being non-protective in this chapter correlates to the findings of Taylor *et al*, whereby one of the two non-protective chemical in that study was also an acetate (EC 231-710-0). However, the other chemical (glycyrrhizin) in this chapter does not appear in the ECHA CHEM database and does not correlate with the other non-protective chemical in the Taylor *et al* study, which is a sulfonanilide (EC 649-383-6). These results are marginally below those found previously by the HSE and Taylor *et al* (2014), who identified 100% and 89%, respectively, of the chemicals with 28-day repeat dose day at or greater than 1000mg/kg bw/day were protective of the 90-day repeat dose study respectively, i.e. it also was at or above 1000mg/kg bw/day. This difference in percentage of chemicals that are protective between the previous studies and the analysis presented here could be due to the variances in the total number of chemicals in the final dataset. The analysis conducted in this chapter further supports those results from the HSE (2011) and Taylor *et al* (2014). It can be seen from each of these analyses that only a relatively small proportion of the three databases utilised (10-15%) contain data for both 28- and 90-day repeat dose studies that can be considered to be of low toxicity. Even though this may not be considered to be of great significance, if 15% of the 90-day repeat dose studies required under REACH (described in more detail in Chapter 1), for those chemicals manufactured or imported over 100 tonnes per year, did not have to be performed, due to low toxicity observed in a 28-day study, over €50 million, and approximately 44,000 animals, could be saved (Taylor *et al* 2014). Thus, this would have a major impact not only

on the financial burden of registering chemicals, but additionally on the time and resources required.

In order to identify, and gather data for, those chemicals with both 28- and 90-day toxicity data Taylor *et al* were required to visit different databases, i.e. eChemPortal to search for substances for which there was a 28- and 90-day study and the ECHA CHEM database to retrieve the toxicity data for both study lengths. In comparison, the use of the COSMOS database within this chapter enabled the analysis conducted by Taylor *et al* to be carried out in an easier and less time consuming manner. This was due to the efforts within the COSMOS consortium to compile both chemical and toxicological data from various sources into one database. The presence of both the chemical, and toxicological, data in one easy to search database is one of the key benefits the COSMOS database provides. An additional benefit is the ability to narrow down search results based upon specific endpoints, test systems, routes of exposure or species.

In this instance, it has not been possible to identify chemical classes that are protective in both 28- and 90-day repeat dose studies. This is most likely due to the number of chemicals currently present within the COSMOS database that contain each of the prerequisites described in the method above, these numbers are likely to increase as and when more data become available. In order to identify chemical classes, and therefore develop structural alerts, that are protective further work could include compiling tables of chemicals from the work conducted here, by the HSE, and Taylor *et al* to ascertain if any chemical classes are consistently protective. Alternatively, for those chemicals that are not protective, an investigation could be undertaken to elucidate mechanistic information for both the 28- and 90-day studies in an attempt to understand whether different mechanisms are responsible for initiating the adverse events in the 28-day study compared to the 90-day study. Additionally, an investigation into the mechanistic information pertaining to the adverse effects used to derive the HNEL value may enable a distinction to be identified between those chemicals

that are protective and those that are not protective as part of the same chemical class, thereby potentially facilitating the development of structural alerts for those chemical classes.

2.4 Conclusions

This chapter shows the current need for a single, comprehensive, and freely available database containing chemical structures linked to repeat dose toxicity data. As part of the development of the COSMOS database novel repeat dose data have been extracted from the SCC(NF)P/SCCS reports. These reports have not previously been utilised to extract repeat dose data from in order for it to be collated into a freely available database. In order to maintain consistency, across the data harvesters, whilst extracting the repeat dose data from this source it was essential that both an SOP and a controlled vocabulary were used. The use of these two factors also enables any future data harvesting to proceed with the same consistency, also the controlled vocabulary makes the COSMOS database simpler and easier to search. To maximise the use of the COSMOS database the repeat dose data gathered were not only the NOAEL and LOAEL values themselves, but also the further information concerning the adverse effects observed at the LOAEL. This additional information regarding the adverse effects enables the development of a variety of *in silico* models. One such model is discussed in more detail in Chapter 5. With regards to the investigation undertaken in this chapter having the toxicological data compiled in one easy to use database, as opposed to spread across multiple databases, made this analysis much easier to do. As more toxicological data is input into the COSMOS database it is envisaged that the analyses that can be undertaken will also expand to cover more endpoints.

Chapter 3: The development of structural alerts using the ChemoTyper software

3.1 Introduction

Both Chapters 3 and 4 refer to the use of mitochondrial toxicity data that were extracted from the same journal article, i.e. Zhang *et al* (2009). Whilst the same data are used, different *in silico* approaches have been utilised; in Chapter 3, a software tool (ChemoTyper) comprised of pre-defined structural features has been used. In comparison, Chapter 4 utilises structural similarity, and a subsequent literature search. These methods have been used to demonstrate how using various techniques can enable the identification of different structural alerts from the same data set. When compiled together structural alerts that relate to the same toxicological outcome develop an *in silico* profiler. These profilers fall into two categories: mechanism- and chemistry-based. A mechanism-based profiler is comprised of structural alerts relating to a Molecular Initiating Event (MIE) that are supported by experimental data illustrating how the alert initiates the MIE. In comparison, a chemistry-based profiler consists of structural alerts that have been associated with inducing toxicity; thus, in contrast to mechanism-based profilers, a chemistry-based profiler does not contain mechanistic information relating to how the observed toxicity is initiated. Due to these intrinsic differences, the two types of profiler should be used for different purposes; mechanistic profilers should be used for category formation and read-across, whilst chemistry-based profilers should be used to screen an inventory to prioritise chemicals to undergo (non-animal) testing. A category developed utilising a mechanism-based profiler enables missing toxicological data for a chemical to be filled using available information from analogues within the same category via read-across predictions (Enoch *et al* 2011a).

The general premise of developing a category is based upon chemicals within the category being similar to one another (ECHA 2008, OECD 2011). This similarity can arise from a variety of properties including: structural features, physico-chemical properties or similarity

in the mechanism of action (Chapters 1 and 6). In addition, the subsequent read-across predictions are made based upon the assumption that similar chemicals should have similar chemical and biological activities (ECHA 2008, ECHA 2012). Whilst there are some exceptions, this assumption provides a useful grounding on which to build. Depending upon the data available for the chemicals used in the read-across, i.e. the analogues, the predictions made can either be qualitative or quantitative. A qualitative prediction would enable, for example, a positive/negative or high/low result to be associated with the chemical. In comparison, a quantitative prediction could identify the concentration at which an effect would be expected to occur.

In order for predictions made by read across to be more acceptable, chemical grouping should be based upon a similar mechanism of action, specifically the MIE. The use of this knowledge with regards to the MIE means any prediction is more mechanistically interpretable and, therefore, more acceptable for regulatory purposes. The MIE; as discussed in Chapters 1, 4, and 5 and defined by the AOP paradigm; is the initial interaction between the non-endogenous chemical and the biological system that initiates the perturbation of normal physiological functioning. The chemistry-based structural alerts do not have mechanistic information associated with them. It is, therefore, anticipated that the subsequent *in vitro* and/or *in chemico* testing will enable mechanistic knowledge to be elucidated and, subsequently, associated with the structural alert. Additionally, it should be noted that mechanism-based profilers can also be used for screening and prioritisation purposes. The benefit of this is that the mechanistic information associated with the structural alert can be utilised to guide which *in vitro* and/or *in chemico* test should be performed; based upon the knowledge held by an AOP.

A number of studies have developed profilers, focussing on a variety of organ-level toxicities, and demonstrated their use in chemical category formation and subsequent read-across analysis (discussed in Chapters 4 and 5). The focus of the work undertaken within Chapters 4 and 5 was to develop (mechanism-based) structural alerts that could be collated

together to form a profiler for mitochondrial toxicity. As discussed further in Chapter 6 several profilers have been encoded computationally within the OECD QSAR Toolbox software program. The QSAR Toolbox contains both mechanism- and chemistry-based profilers. The QSAR Toolbox holds a number of mechanism-based profilers covering a variety of endpoints such as skin sensitisation, respiratory sensitisation, acute aquatic toxicity, carcinogenicity, eye irritation and *in vitro* mutagenicity. These mechanistic profilers are used to assign an MIE to a chemical (for example, covalent protein or covalent DNA binding) and, thus, this information can be utilised to develop mechanism-based categories that enable read-across and other structure-activity relationship predictions to be made. In order to justify these predictions the user can view the supporting information that is associated with the structural alert. This supporting information consists of a visual representation of the alert, the rationale as to why this alert was triggered, and the mechanistic information relating the alert to the endpoint in question, for example protein binding. In comparison, the chemistry-based profilers held within the QSAR Toolbox enable chemical categories to be formed based upon the presence of certain structural features, chemical elements or functional groups.

The ChemoTyper software is a freely available chemoinformatics tool that enables a set of chemicals to be searched against a pre-defined set of structural features (denoted as chemotypes in the ChemoTyper software). This application was developed by Molecular Networks GmbH under contract by the US FDA and is freely available from https://chemotyper.org/ (accessed 17.11.2014). The ChemoTyper is a data mining tool that enables a data set to be screened for the presence of certain structural features. The features contained within the ChemoTyper are from the predefined Toxprint library (Yang *et al* 2013) of structural fragments. The resulting outcome of this screening process is a pictorial representation of the 'chemotype' searched for overlaid on top of the whole chemical. Figure 3.1 shows a pictorial representation of the ChemoTyper output; this, coupled with the ability to filter the chemicals based upon structural features, enables a user to readily identify

structural fragments that are associated with toxicity. Therefore, those fragments that are associated with a higher proportion of toxic chemicals can be considered to be associated with (a specific) toxicity, thus, enabling them to be defined as a structural alert. Due to these capabilities the ChemoTyper has, therefore, been proposed to be of use for developing structural alerts.

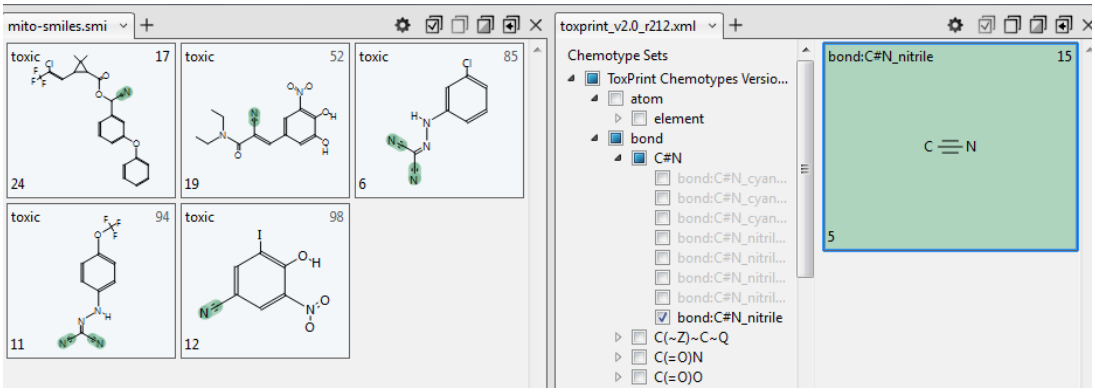


Figure 3.1: Identification of a nitrile chemotype contained within a larger set of chemicals. The nitrile fragment is highlighted green in those chemicals containing the chemotype.

Given the potential that the ChemoTyper has for developing structural alerts, the aim of this chapter was to perform a chemoinformatics analysis of a dataset of chemicals associated with mitochondrial toxicity. This analysis involved identifying structural alerts in the dataset that were associated with mitochondrial toxicity. The work performed in this chapter (along with Chapters 4 and 5) is important as relatively little work has been performed in the area of developing structural alerts for mitochondrial toxicity; especially with regards to AOPs. This issue is discussed further in Chapters 1, 4, and 5.

3.2 Methods

3.2.1 Data set

The 288 chemicals reported by Zhang *et al* (2009) were used as the basis from which to generate the structural alerts associated with mitochondrial toxicity. This article was chosen

for use as it provides one of the largest freely available datasets, for which the chemical list has qualitative mitochondrial toxicity data associated with it. Within this data set 171 chemicals have been reported within the literature as inducing mitochondrial toxicity and were therefore considered to be mitochondrial toxicants. The chemicals with a negative result for mitochondrial toxicity were selected from the FDA-approved drug list, whereby the therapeutic action mechanism, of common and safe oral drugs, was not associated with a mechanism of drug-induced mitochondrial toxicity. Data for the 288 chemicals, including SMILES and toxicity towards mitochondria are available in Appendix II. From this data set the chemical structures were identified in, and the corresponding SMILES strings were extracted from, the Royal Society of Chemistry's ChemSpider website (www.chemspider.com). These chemical structures were subsequently combined with the 'toxic to mitochondria' and 'non-toxic to mitochondria' result from Zhang *et al* (2009) and saved as a .smi file using the chemical visualisation software Marvin View (v6.0.0) developed by Chemaxon (available at www.chemaxon.com/products/marvin/marvinview/, accessed 17.11.2014).

3.2.2 Generation of structural alerts using the ChemoTyper software

The workflow below (Figure 3.2) outlines the steps in the development of the structural alerts within this chapter. This workflow is split into two sections: the first section relates to utilising the ChemoTyper software in order to identify the structural alerts (denoted as chemotypes within the ChemoTyper software) that are associated with mitochondrial toxicants. The second section relates to the generation of SMARTS (SMiles ARbitrary Target Specification; Daylight 2014) patterns using the chemotypes identified in section one. SMARTS is a language utilised when developing structural alerts as it enables the user to identify a specific substructure that may be associated with toxicity.

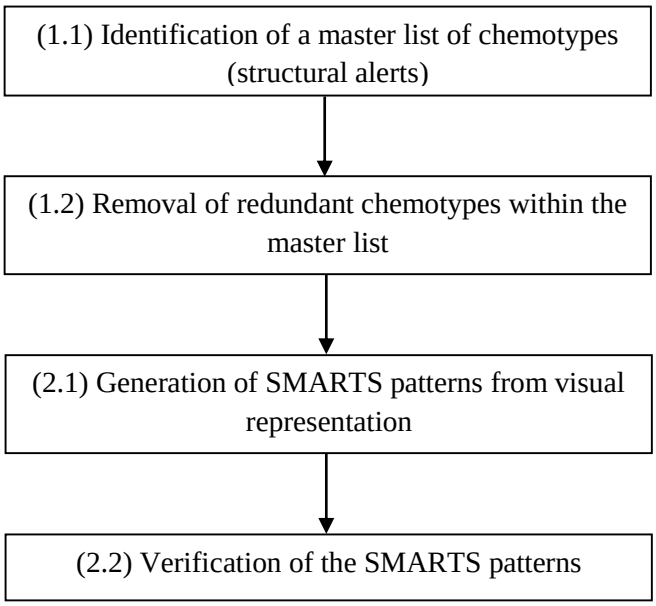


Figure 3.2: Workflow outlining the process undertaken in order to identify and develop the structural alerts

Workflow 1.1. Identification of chemotypes (structural alerts)

The following steps were adhered to in order to develop structural alerts using the freely available ChemoTyper software (v1.0). Additionally, the workflow illustrated in Figure 3.3 shows an outline of how the ChemoTyper was utilised to develop structural alerts.

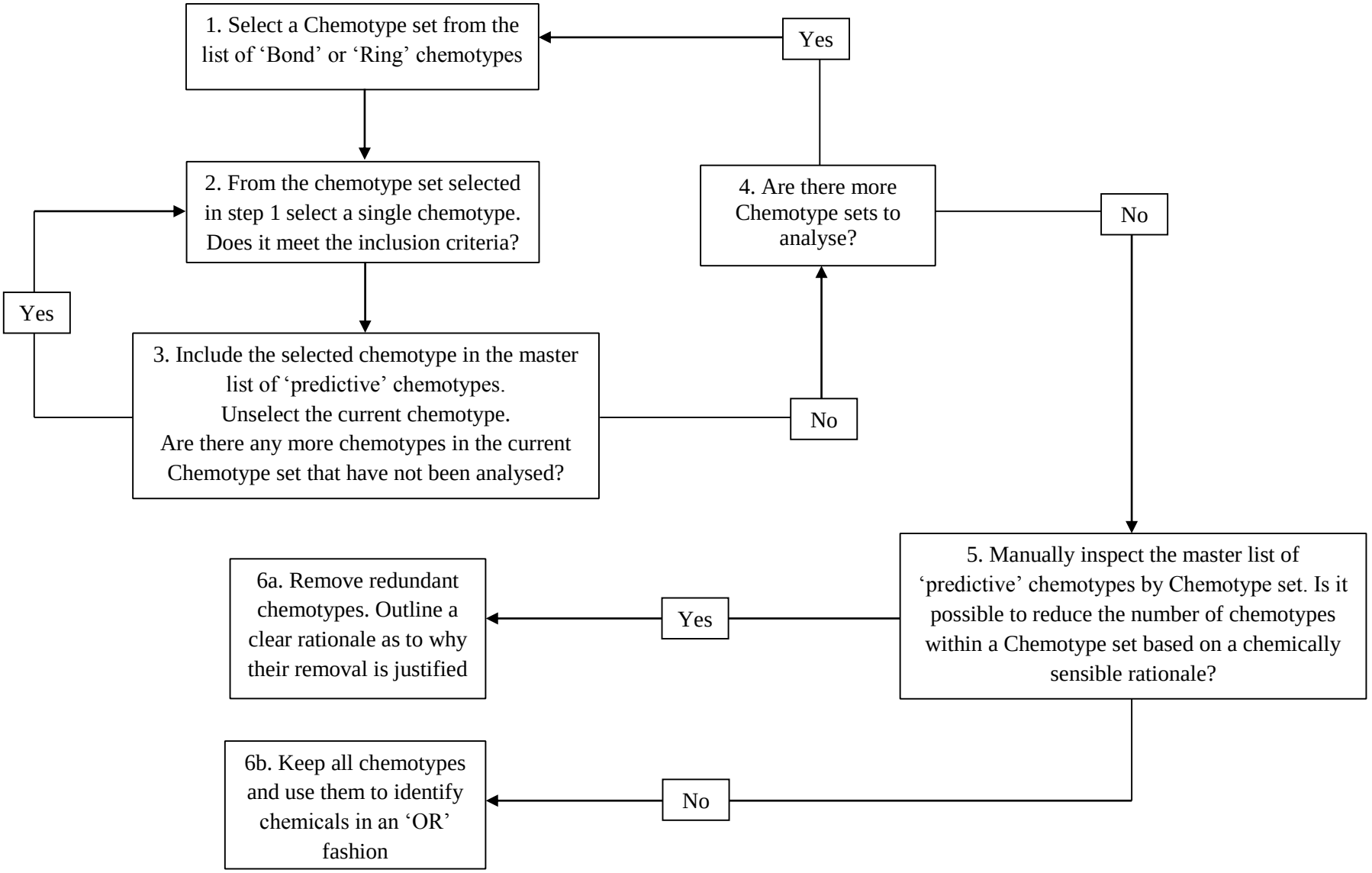


Figure 3.3: Flowchart describing the process of generating the structural alerts using the ChemoTyper software

Step 1 – The Toxprint chemotypes; inherent in the ChemoTyper software; and the data set to be examined; in this instance the data set extracted from Zhang *et al* (2009); were opened within the 'Match' component of the ChemoTyper. The Toxprint chemotypes are the structural fragments against which the data set was analysed.

Step 2 - The drop-down menu adjacent to 'Filter Structures' was set to 'Containing any selected chemotype (OR)' (circled in black, Figure 3.4). This was necessary as it enabled chemicals that contain a chemotype to be rapidly identified, by filtering out those chemicals that did not contain the current chemotype.

Step 3 – A chemotype set was selected from the list present within the ChemoTyper (circled in red, Figure 3.4). From this chemotype set a single chemotype was selected (circled in blue, Figure 3.4).

Step 4 – If the chemotype met the inclusion criteria; i.e. at least three chemicals 'toxic to mitochondria' and no more than a single chemical 'non-toxic to mitochondria' are identified; the chemotype was included in the master list of 'predictive' chemotypes. The use of the inclusion criteria aimed to prevent the development of structural alerts based upon chemotypes that identified a large number of chemicals 'non-toxic to mitochondria', thereby, limiting the number of false positive results predicted by the structural alert.

Step 5 – The current chemotype was unselected and steps 3 and 4 were repeated until all chemotypes had been analysed.

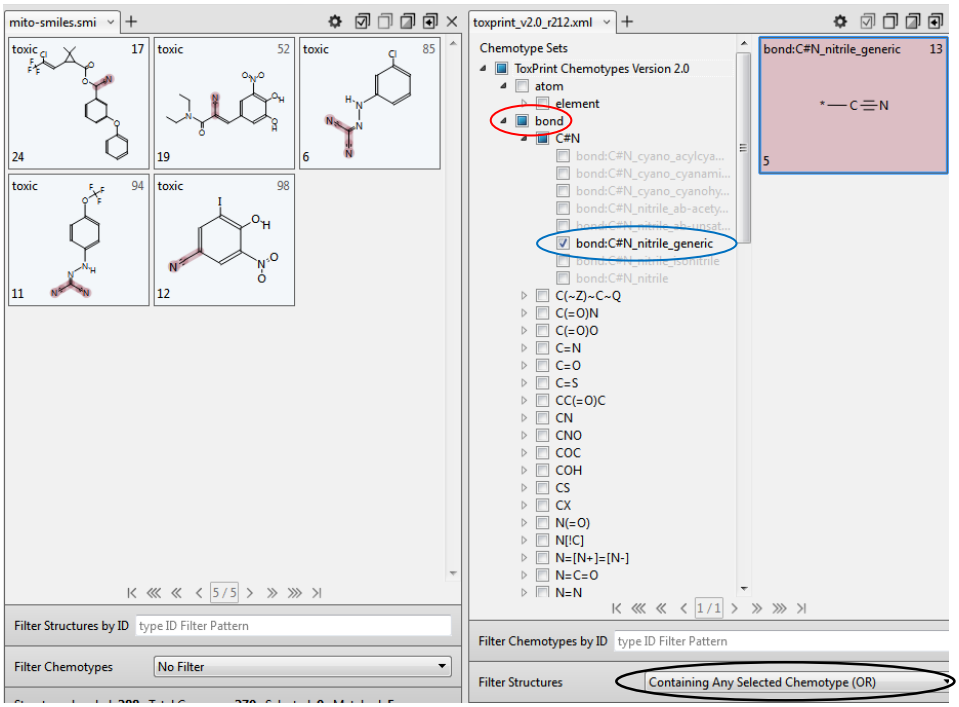


Figure 3.4: A screenshot of the ChemoTyper software. On the right is the chemotype selected. Whilst on the left is a visual representation of each of the chemicals containing the chemotype and whether they were identified by Zhang and colleagues (Zhang *et al* 2009) as being 'toxic to mitochondria' or 'non-toxic to mitochondria.'

Workflow 1.2. Removal of redundant chemotypes within master list

Upon completion of the previous process the master list of 'predictive' chemotypes was inspected so as to remove any redundant chemotypes present (steps 5 and 6 in Figure 3.3). The removal of redundant chemotypes was only undertaken if multiple chemotypes identified the same set of chemicals (or one chemotype identified a sub-set of chemicals identified by a second chemotype). For example, as can be seen in Figure 3.5, there is an overlap between the chemicals identified by the three selected chemotypes (Figure 3.6). Given this overlap, the most specific chemotype ('bond:CX_halide_aromatic-Cl_trihalo_benzene_(1_2_4-)') was selected, with the remaining two being removed from the master list.

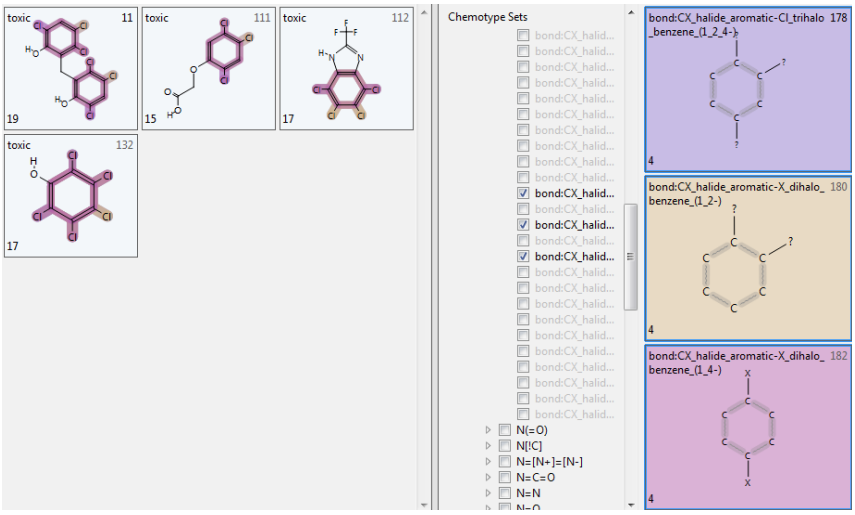


Figure 3.5: Results from the ChemoTyper software, illustrating the redundancy between the three chemotypes

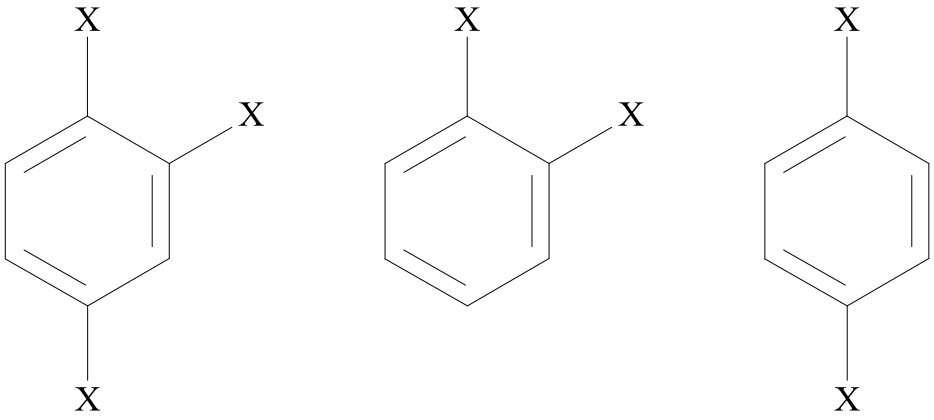


Figure 3.6: A depiction of the 'bond:CX_halide_aromatic-Cl_trihalo_benzene_(1_2_4-),' the 'bond:CX_halide_aromatic-X_dihalo_benzene(1_2-),' and the 'bond:CX_halide_aromatic-X_dihalo_benzene(1_4-)' chemotypes respectively, where X can be any halide atom. The 'bond:CX_halide_aromatic-Cl_trihalo_benzene_(1_2_4-)' chemotype was retained.

Workflow 2.1 Generation of SMARTS patterns from visual representation

The SMARTSeditor software, developed by Universität Hamburg (available from www.zbh.uni-hamburg.de/?id=426, accessed 17.11.2014), was used to generate all SMARTS patterns. SMARTSeditor is a graphic editing tool that enables the user to draw, and edit, a visual representation of a chemical structure, whereby the visual representation is simultaneously converted into a SMARTS pattern. In this instance the chemical phenol will be used as an example of how the SMARTS patterns were generated.

Step 1 – The 'draw' function was used to sketch the structural alert into the SMARTS editor software (circled in red, Figure 3.7).

Step 2 – Any alterations to the sketched structure were made using the 'edit' function. Alterations were required for all structural alerts that contain an aromatic ring system. This is due to aliphatic carbon atoms being inserted as a default by the SMARTSeditor software.

Step 3 – The finalised SMARTS pattern generated was displayed above the visual representation in the SMARTSeditor (circled in blue, Figure 3.7).

Step 4 – The SMARTS pattern generated in Step 3 was, thus, included in the master list of 'predictive' structural alerts.

Step 5 - The visual representation, and associated SMARTS pattern, was deleted from the SMARTSeditor software and Steps 1 to 4 were repeated until SMARTS patterns were generated for all of the structural alerts identified in stage 1.

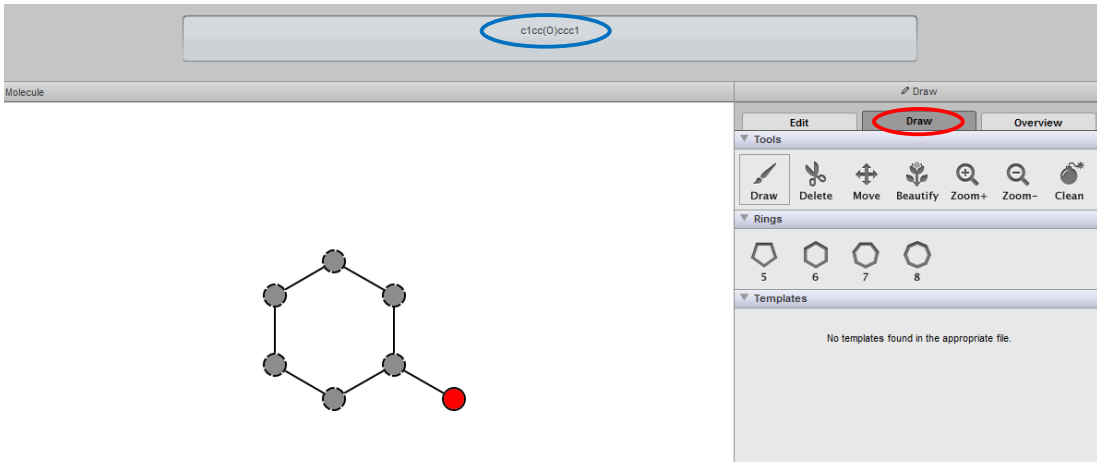


Figure 3.7: The SMARTSeditor program with a visual representation, and SMARTS pattern, for phenol.

Workflow 2.2 Verification of the SMARTS patterns

Each of the SMARTS patterns developed was verified using the 'depictmatch' tool developed by Daylight Chemical Information Systems Inc. (available at http://www.daylight.com/daycgi_tutorials/depictmatch.cgi, accessed 17.11.2014). This was to ensure the SMARTS pattern could correctly identify chemicals containing the structural alerts. The verification process was undertaken as described below.

Step 1 – The SMARTS pattern, generated using the SMARTSeditor, was entered into the depictmatch tool under the heading 'SMARTS' (circled red, Figure 3.8).

Step 2 – A SMILES notation for a chemical containing the structural alert was entered under the heading 'SMILES' (circled blue, Figure 3.8).

Step 3 – The SMARTS pattern was identified as being correct if the depictmatch tool highlighted the structural fragment correlating to the SMARTS pattern in yellow (in the black box Figure 3.8).

Step 4 – If the SMARTS pattern was incorrect, i.e. the visual depiction remained white; the SMARTS pattern was investigated to determine where corrections were required.

Step 5 – Any corrections made were subsequently tested using the depictmatch tool.

Step 6 – Steps 4 and 5 were repeated until the SMARTS pattern correctly identified the structural fragment coded for.

Step 7 –The steps 1 through 6 were repeated for each SMARTS pattern developed in section 2.2.

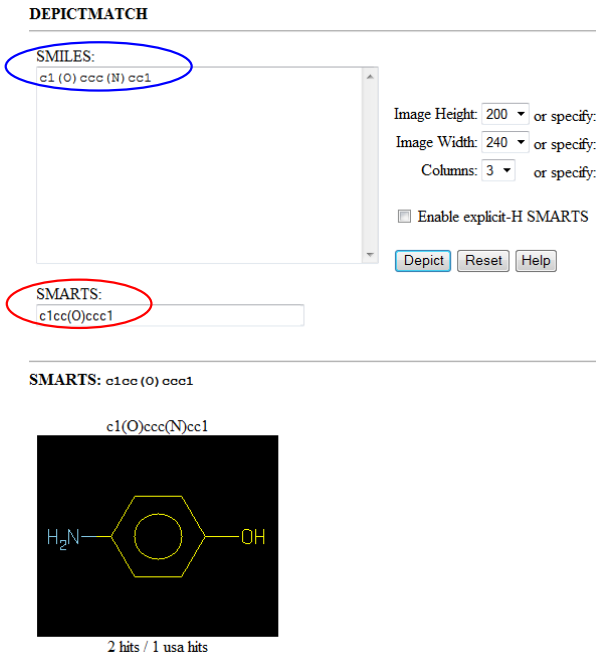


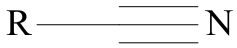
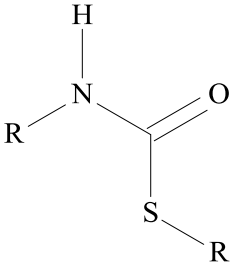
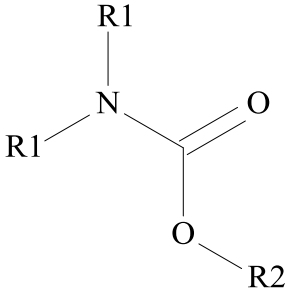
Figure 3.8: Use of the depictmatch tool to identify whether previously created SMARTS patterns correctly identified chemicals contained the structural fragment coded for.

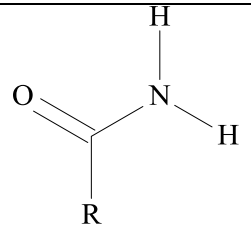
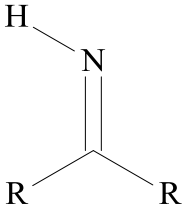
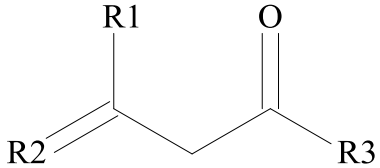
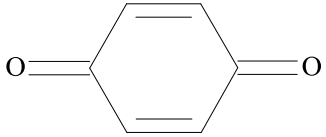
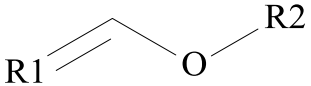
3.3 Results and discussion

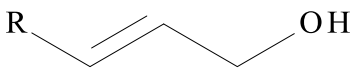
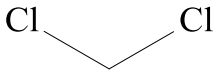
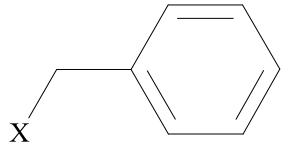
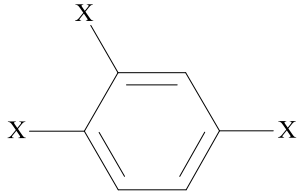
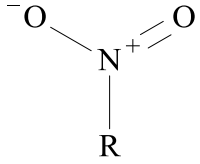
The aim of this chapter was to use the newly developed ChemoTyper software in order to perform chemoinformatics analysis of a dataset of 288 chemicals (Zhang *et al* 2009) and identify structural alerts associated with mitochondrial toxicants. From this

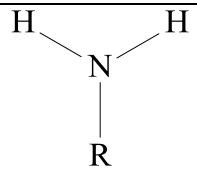
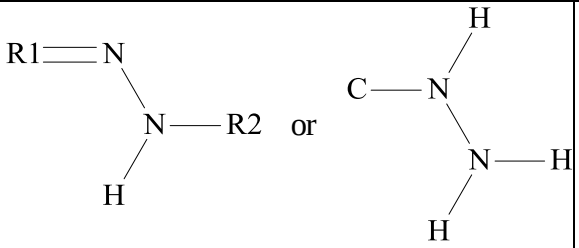
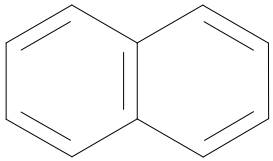
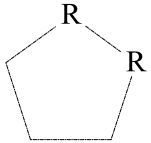
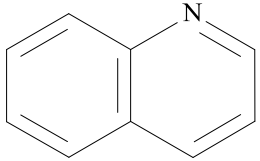
chemoinformatics analysis twenty structural alerts were developed (Table 3.1). In total, these twenty alerts covered 83 chemicals. Of these 83 chemicals, 77 were identified as being mitochondrial toxicants; the remaining six chemicals were identified as not being mitochondrial toxicants. These structural alerts covered twelve of the 61 chemotype sets (under the heading 'bond,' 'chain,' and 'ring') that are present within the ChemoTyper software. A chemotype set consists of a group of chemotypes that contain the same structural fragment. Of these twelve chemotype sets, the 'bond:CX' set contains the most chemotypes that were seen to be associated with mitochondrial toxicants; i.e. four chemotypes (listed in Table 3.1). Structural alert 13 (Table 3.1) identified the highest number of chemicals 'toxic to mitochondria,' i.e. ten chemicals.

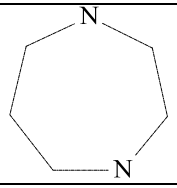
Table 3.1: A list of the chemotypes identified as being associated with mitochondrial toxicants.

Structural alert number	Structural alert	Chemotype name (as denoted within the ChemoTyper software)	Number of chemicals identified*	Is an all-encompassing mechanistic hypothesis possible for the alert?
1	 R = any C atom	bond:C#N_nitrile	5 (5)	No
2	 R = aliphatic C atom	bond:C(=O)N_carbamate_thio_generic	3 (3)	Yes
3	 R1 = aliphatic C or H atom R2 = aliphatic C or N atom	bond:C(=O)N_carbamate	7 (6)	No

4	 R = aliphatic C atom	bond:C(=O)N_carboxamide_(NH2)	7 (6)	No
5	 R = any C atom	bond:C=N_imine_C(connect_H_gt_0)	3 (3)	No
6	 R1 = aliphatic O or H atom or cyclic C atom R2 = sp^2 cyclic C atom R3 = sp^2 cyclic C atom	bond:CC(=O)C_ketone_alkene_cyclic_3-en-1-one	5 (5)	No
7		bond:CC(=O)C_quinone-1_4-benzo	8 (8)	Yes
8	 R1 = sp^2 cyclic C atom R2 = any C atom	bond:COC_ether_alkenyl	4 (4)	No

9	 R = aliphatic C atom	bond:COH_alcohol_allyl	4 (3)	No
10		bond:CX_halide_alkyl-Cl_dichloro_(1_1-)	4 (4)	No
11	 X = any halide atom	bond:CX_halide_alkyl-X_benzyl_alkane	5 (4)	No
12	 X = any halide atom	bond:CX_halide_aromatic-Cl_trihalo_benzene_(1_2_4-)	4 (4)	No
13	 X = any halide atom	bond:CX_halide_generic-X_dihalo_(1_2-)	10 (10)	No
14	 R = aromatic C atom	bond:N(=O)_nitro_aromatic	9 (9)	No

15	 N = aliphatic R = aliphatic S or N atom	bond:N[!C]_amino	4 (3)	No
16	 R1 = aliphatic C atom R2 = aromatic C atom	bond:NN_hydrazine_acyclic_(connect_noZ)	3 (3)	No
17		ring:fused_[6_6]_naphthalene	7 (7)	No
18	 R = aromatic N or O atom	ring:hetero_[5]_Z_1_2-Z	5 (4)	No
19		ring:hetero_[6_6]_N_quinoline	7 (7)	No

20		ring:hetero_[7]_N_diazepine_(1_4-)	3 (3)	No

*The number in column four relates to the total number of chemicals identified by each alert, whilst the number in parenthesis correlates to the number of toxic chemicals identified.

The chemoinformatics analysis showed that two types of structural alert can be identified using the ChemoTyper: 1) well defined alerts for which a mechanistic hypothesis can be defined, 2) more diverse alerts for which a mechanistic hypothesis may not be possible.

3.3.1 Structural alerts for which an all-encompassing mechanistic hypothesis is possible

A mechanistic hypothesis can be attributed to two of the structural alerts identified whilst undertaking this analysis. For example, the group of eight chemicals identified by structural alert 7 (those containing a quinone moiety, Figure 3.9) are likely to induce mitochondrial dysfunction by acting as an alternative electron acceptor, thereby, inhibiting the electron transport chain (Figure 3.10). It has been demonstrated that chemicals containing a quinone moiety can sequester electrons from the electron transport chain by competing with the natural electron carrier: ubiquinone (Wallace and Starkov 2000, Wallace 2003). Upon sequestration of an electron from Complex I the quinone is, itself, reduced to a semi-quinone radical intermediate (Gerwirtz 1999, Wallace 2003). This radical species may transport the electron directly to Complex IV, thus, becoming oxidised back into a quinone. Alternatively, the semi-quinone radical may indirectly induce mitochondrial toxicity by reacting with molecular oxygen, producing reactive oxygen species. Interfering with the electron transport chain in this manner could lead to a multitude of effects such as oxidation of mitochondrial DNA, proteins and/or lipids; and reduction in ATP production (Ohkuma *et al* 2001, Chan *et al* 2005). Additionally, four of the eight chemicals identified by structural alert 7 contain a structure similar to doxorubicin. Chapter 4 discusses in more detail how anthracycline antibiotics act to induce mitochondrial toxicity.

Structural alert 2 identifies a group of three toxic chemicals, each of which contain a thiazolidinedione moiety. This moiety has been identified by Naven *et al* (2013) as inducing mitochondrial toxicity via uncoupling of oxidative phosphorylation due to its ability to act as

a protonophore. Further discussion surrounding the mechanism of toxicity due to the thiazolidinedione moiety can be found in Chapter 4.

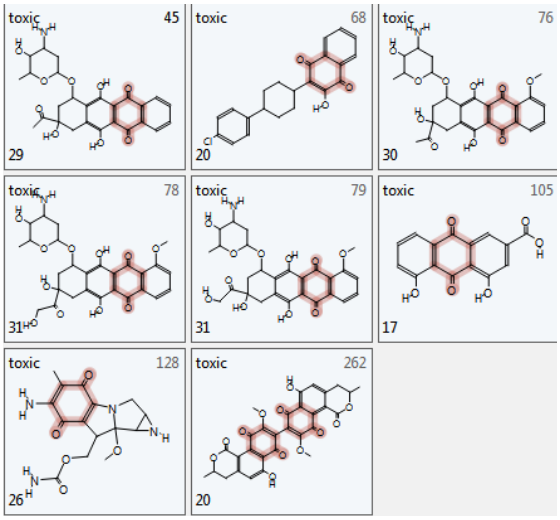


Figure 3.9: A screenshot of the group of chemicals identified by structural alert 7: the quinone moiety is highlighted in red in each of the chemicals.

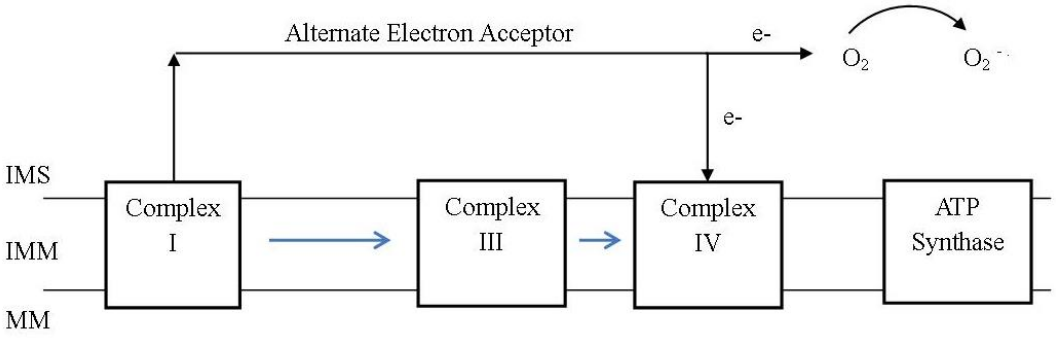


Figure 3.10: An illustration of the mechanism of inhibition of the electron transport chain, via alternative electron cycling, leading mitochondrial toxicity

3.3.2 Structural alerts for which an all-encompassing mechanistic hypothesis is not possible

In contrast, for the remaining alerts, for example the nitro aromatic alert (structural alert 14), it may be more difficult (or not possible) for a mechanistic hypothesis to be identified that encompasses each of the chemicals 'toxic to mitochondria' present in the group. This is due

to the fact that these alerts are broader in scope and, therefore, the chemicals identified are (generally) more diverse (Figure 3.11). This diversity increases the likelihood that the observed toxicity will be induced by a variety of mechanisms. However, within this type of category, where an all-encompassing mechanism may not be readily identifiable, it may still be possible to hypothesise a mechanism for individual chemicals. For example, 2,4-dinitrophenol (circled red, Figure 3.11) is a well-known and well-characterised uncoupler of oxidative phosphorylation (Chan *et al* 2005, Dykens and Will 2008), whilst chloramphenicol (circled blue, Figure 3.11) has been shown to inhibit mitochondrial protein synthesis by binding to the 50S subunit of the ribosome (Kroon and de Vries 1969, Dykens and Will 2008). As there is no single unifying mechanism associated with this class of chemistry-based alerts, these alerts should not be used to develop chemical categories and, subsequently, perform read-across for novel chemicals. In this instance, and throughout this thesis, the use of read-across pertains to hypothesising a mechanism by which (mitochondrial) toxicity is induced, i.e. an MIE, and does not relate to the wider use of read-across in predicting toxicity. However, this type of chemistry-based alert can be used to screen an inventory in order to prioritise the chemicals within it for further (non-animal) testing. This is due to the statistical evidence that chemicals containing one (or more) of these chemistry-based alerts are more likely to be associated with toxicity (Table 3.1). Utilising these types of chemistry-based (rather than mechanism-based) structural alerts, identified in this study, would enable a user to ascertain how many chemicals, within a given inventory, have the potential to induce mitochondrial toxicity. Nonetheless, it is worth noting that chemistry-based structural alerts (such as those identified within this chapter) could be refined to a mechanism-based alert by utilising the results of *in vitro*/*in chemico* testing undertaken to establish a mechanistic hypothesis.

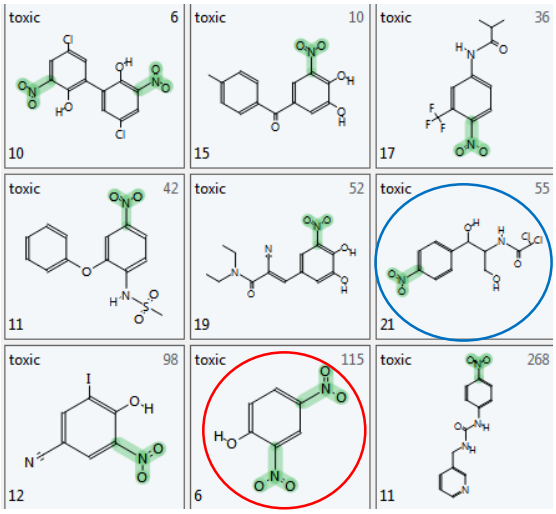


Figure 3.11: A screenshot of the group of chemicals identified by structural alert 14: the nitro aromatic moiety is highlighted in green in each of the chemicals.

The chemoinformatics analysis performed in this chapter shows that the ChemoTyper software has a number of benefits: firstly, it is useful for rapidly identifying structural alerts that are associated with (mitochondrial) toxicity. The ChemoTyper software facilitates this type of analysis due to the visual representation of the structural feature and the chemicals associated with a feature (from a given dataset). In the current chapter the toxicity data associated with the chemicals in the dataset were mitochondrial data; however, other toxicity data could be used. Therefore, this approach could be utilised to develop mechanism- and/or chemistry-based structural alerts for a wide variety of toxicity endpoints where a toxic/non-toxic outcome is available. In addition to allowing for the identification of chemical structures associated with mitochondrial toxicity, the approach laid out within this chapter inherently results in structural alerts with a low false positive prediction rate. This is due to the fact that no more than one chemical, within the category, 'non-toxic to mitochondria' (as identified by Zhang *et al* 2009) could be associated with a structural alert in order for it to be identified as being associated with mitochondrial toxicity. Additionally, the ChemoTyper software enables structural alerts to be defined without the user needing to have any prior mechanistic knowledge of how the alert initiates toxicity. However, this also results in the

main disadvantage of this approach in that a single chemical could, potentially, contain multiple structural alerts making it more difficult to distinguish which of the identified alerts is responsible for inducing toxicity.

3.4 Conclusions

The aim of this chapter was to utilise the ChemoTyper software in order to identify structural alerts associated with mitochondrial toxicity. Overall, twenty structural alerts were developed. Two types of structural alert could be discerned from this analysis: those whereby a mechanistic hypothesis can be defined for all chemicals containing a specific alert; and those where the alert was too broad to be able to hypothesise an all-inclusive mechanism. Whilst this second class of chemistry-based structural alerts should not be used for grouping for read-across purposes (in terms of hypothesising an MIE), they can be of use for screening large data sets of chemicals that will, subsequently, undergo testing within *in vitro*/*in chemico* assays. However, if further testing is undertaken for a subset of chemicals containing a specific chemistry-based alert, it may be possible to associate mechanistic information with a number of the chemistry-based alerts, thereby, enabling them to be rationalised into mechanism-based structural alerts, such as those discussed in subsequent chapters. The development of mechanism based structural alerts and their relative advantages over chemistry-based structural alerts are discussed further in Chapters 4 and 5.

Chapter 4: The use of category formation in the development of an *in silico* profiler for mitochondrial toxicity

4.1 Introduction

Over the past decade a number of changes have occurred in European cosmetic legislation that have led to an increase in efforts to develop alternative methods to traditional animal testing for risk assessment (EC 2003, EC 2006, EC 2006, ECHA 2008, ECHA 2012). These alternatives have been developed employing *in silico*, *in chemico* and *in vitro* methods focussing on replacing or reducing animals used in short- and long-term toxicity tests (Adler *et al* 2011). In order to be relevant, and useful, for regulatory assessment these alternatives should be based upon specific *in vivo* endpoints. Within recent years, interest has grown in developing a greater understanding of toxicity pathways. One such pathway approach is the Adverse Outcome Pathway (AOP) paradigm. An AOP is a framework that means to establish a mechanistic connection between an upstream Molecular Initiating Event (MIE) and a downstream adverse outcome relevant for risk assessment (Ankley *et al* 2010, OECD 2013, Vinken 2013, Vinken *et al* 2013a, Vinken 2014) (Chapter 1). The MIE is the critical event in the progression of an AOP as it provides insight into the initial interaction(s) between the chemical behaviour of the non-endogenous chemical and the biological system that initiates the perturbation of the normal pathway. Elucidation of the mechanistic information relating to specific MIEs enables the identification of structural (and physico-chemical) features of chemicals that are responsible for the interaction with biological macromolecules, thus, facilitating the development of structural alerts.

This process can be labour intensive, another method could be to utilise an automated clustering technique; one such clustering approach is to use the Toxmatch software. Toxmatch is an open source program; developed by Ideaconsult, Sofia; that can be used to group chemicals based upon one of a variety of different similarity indices. The Toxmatch

program contains two molecular graph-based similarity methods: molecular fingerprints and atom environments (Enoch *et al* 2009). Within this chapter, and Chapter 5, the atom environment similarity measure has been utilised. The atom environment approach identifies the "fragments surrounding each atom in a molecule, up to a predefined level" and is calculated according to the explanation in Jaworska and Nikolova-Keliazkova (2007). Subsequently, the program calculates the average Hellinger distance between the atom environments of one chemical and the atom environments of the set. This generates a matrix consisting of similarity scores, between 0 and 1, for each chemical within the data set, with 0 meaning the two chemicals are completely dissimilar and 1 meaning the two chemicals are completely similar. Whilst it is appreciated that more complex clustering approaches may be implemented this is beyond the scope of this thesis.

When combined, multiple structural alerts pertaining to the same MIE form the basis of an *in silico* profiler (Enoch 2010, Enoch and Cronin 2010 Enoch *et al* 2011a, Hewitt *et al* 2013, Przybylak and Schultz 2013) (discussed in more detail in Chapters 1, 3, and 5). The information within an *in silico* profiler can, in turn, be used to develop chemical categories centred on a common MIE (note that multiple MIEs can be initiated by a single chemical). This allows for read-across and data gap filling to be applied. The premise behind these structurally developed categories is that similar chemicals should have similar biological activities and therefore, should have the same MIE. Furthermore, the categories produced using *in silico* profilers can be supported by, and used to prioritise, additional testing using *in vitro* and/or *in chemico* methods, within an integrated testing strategy (ITS) or an integrated approach to testing and assessment (IATA). Such strategies can be used for hazard identification and risk assessment purposes, as well as being incorporated into *in silico* software tools such as the OECD QSAR Toolbox (available at www.qsartoolbox.org, accessed 17.11.2014) (Gutsell and Russell 2013, OECD 2013, Przybylak and Schultz 2013, Vinken 2013).

A number of *in silico* profilers have been developed for a variety of organ-level toxicities, such as skin sensitisation, respiratory sensitisation, genotoxicity, protein binding and hepatotoxicity (Enoch *et al* 2008a, Enoch *et al* 2008b, Enoch and Cronin 2010, Enoch *et al* 2011a, Hewitt *et al* 2013, Sakuratani *et al* 2013a, Sakuratani *et al* 2013b, Vinken *et al* 2013a, Vinken *et al* 2013b). However, very few have dealt with toxicity induced by mitochondrial dysfunction (Zhang *et al* 2009, Naven *et al* 2013, Wallace *et al* 2013). This is, in part, due to the number of mechanisms by which a chemical could induce mitochondrial dysfunction (Nadanaciva and Will 2011). An additional complication is that a single chemical might have the ability to induce more than one of these mechanisms, making it difficult to define a single MIE within the AOP paradigm. Over the past decade, interest in screening chemicals for an ability to induce mitochondrial dysfunction has increased (Dykens and Will 2008, Nadanaciva and Will 2011). This is, in part, due to the withdrawal of a number of pharmaceuticals from the market after observed mitochondrial dysfunctions (Wallace and Starkov 2000, Brunmair *et al* 2004, Rolo *et al* 2004, Chan *et al* 2005, Dykens and Will 2007, Dykens *et al* 2007). Toxicity to mitochondria has led to such withdrawals as these are important organelles present within almost every cell type of the body, the exception being mature erythrocytes (Cohen and Gold 2001, Pieczenik and Neustadt 2007). Previous research has shown that mitochondrial dysfunction may be induced by a range of chemicals and has been linked to a variety of organ toxicities within kidney, liver and nervous tissues (Wallace and Starkov 2000, Brunmair *et al* 2004, Rolo *et al* 2004, Chan *et al* 2005, Dykens and Will 2007, Dykens *et al* 2007). The most susceptible tissues to mitochondrial dysfunction are those containing a higher concentration of mitochondria or those exposed to a higher concentration of chemical: such as the liver, kidneys and heart (Amacher 2005, Dykens and Will 2007, Dykens and Will 2008, Nadanaciva and Will 2011). Five general mechanisms of mitochondrial dysfunction have been identified (Krahenbuhl 2001, Amacher 2005): inhibitors of the electron transport chain and ATP synthase (Complex V), uncouplers of oxidative phosphorylation, opening of the membrane permeability transition pore,

inhibition of fatty acid β-oxidation, and oxidation or inhibition of mitochondrial DNA (discussed in more detail in Chapter 1).

As an example of the importance of mitochondrial toxicity approximately 35%, of more than 500 pharmaceutically relevant chemicals, have been shown to be directly involved in impairing normal mitochondrial functioning by inhibition of the electron transport chain and/or by uncoupling of oxidative phosphorylation (Dykens and Will 2008). Additionally, there are chemicals that can induce mitochondrial toxicity via alternative mechanisms, such as inducing the membrane permeability transition, inhibition of β-oxidation of mitochondrial fatty acids, or interfering with mitochondrial DNA. Briefly stated, chemicals that inhibit the electron transport chain can do so by either direct binding to the complexes of the electron transport chain or ATP synthase or by acting as an alternative electron acceptor (Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005). The inhibition of electron flow along the electron transport chain by both of these mechanisms induces the formation of reactive oxygen species resulting in oxidative stress (Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005). Uncouplers of oxidative phosphorylation induce mitochondrial toxicity by shuttling protons into the mitochondrial matrix, via the inner mitochondrial membrane, bypassing ATP synthase. This assisted transport of protons back into the matrix dissipates the electrochemical potential, resulting in the loss of ATP production and, ultimately, cell death (Terada 1990, Schonfeld *et al* 1992, Sun and Garlid 1992, Wallace and Starkov 2000, Krahenbuhl 2001, Amacher 2005, Chan *et al* 2005, Spycher *et al* 2008, Cela *et al* 2010). Induction of the membrane permeability transition increases the permeability of the inner mitochondrial membrane to low molecular weight solutes (<1500Da), leading to a disruption of the electron transport chain, loss of membrane potential, and swelling of both the inner- and outer mitochondrial membranes (Kroemer *et al* 2007, Lemasters *et al* 2009). Inhibition of β-oxidation of mitochondrial fatty acids reduces the amount of NADH and $FADH_2$ available for oxidative phosphorylation that, in turn, reduces ATP production (Pessayre *et al* 2008). Mitochondrial DNA encodes 13 components of the electron transport chain, damage

that occurs to mitochondrial DNA can have a variety of downstream effects depending upon where it occurs (Amacher 2005, Pessayre *et al* 2008). It should be noted, however, that there is the potential that multiple, competing, mechanisms could initiate mitochondrial toxicity observed for a single (group of) chemical(s), i.e. one chemical may induce several MIEs.

Given the importance of mitochondria within most cell systems, and the wide range of organ-level toxicities that may arise from mitochondrial dysfunction, the aim of this chapter was to utilise structural similarity, and subsequent information in the available literature, to identify structural alerts that could be combined to form an *in silico* profiler. Consequently, this profiler could be incorporated into software tools, to enable large datasets to be screened to identify chemicals with the ability to induce mitochondrial toxicity.

4.2 Methods

4.2.1 Data set

The data set extracted from Zhang *et al* (2009), discussed in Chapter 3, was utilised to perform the clustering analysis within this chapter. Given the lack of supporting mechanistic information to confirm the presence, or absence, of mitochondrial toxicity additional analysis was carried out, as detailed below.

4.2.2 Category formation based upon structural similarity

All chemical structures were encoded into Simplified Molecular Input Line Entry System (SMILES) strings, neutralised and had salts removed. Each of the SMILES strings was extracted from the Royal Society of Chemistry's ChemSpider website (http://www.chemspider.com/). Similarity calculations were implemented within the freely available Toxmatch software (v1.07) using the atom environment nearest neighbour

approach, generating a data matrix with a Tanimoto similarity score for each chemical to all others within the data set. Subsequently, in-house code was implemented within Microsoft Excel that identified analogues with a similarity index of 0.6 or greater; this was used in order to develop categories for the chemicals within the dataset with two, or more, analogues. Further analysis was undertaken upon those categories that met the following criteria: they contained three or more chemicals and at least one mitochondrial toxic chemical.

4.2.3 Mechanistic hypothesis and the development of alerts

Once categories had been developed using structural similarity a detailed search of the available literature was undertaken to elucidate the mechanistic information behind the molecular initiating event, along with other downstream key events, leading to the disruption of the mitochondria. This mechanistic information was subsequently utilised to support the definition of a structural alert suitable for grouping chemicals. These structural alerts were defined by identifying the common fragment present within each of the chemicals found to have positive mitochondrial toxicity according to literature information associated with them. Any additional information regarding the limits of the fragment found during the literature search, such as the requirement for an electron withdrawing group or the type of bond needed (e.g. a tertiary amine), was used to refine the structural alert further. The resulting alerts were subsequently defined as SMARTS patterns (www.daylight.com). The process of how the SMARTS patterns were developed is described in detail in Chapter 3. A structural alert was only developed if information linking category members to mitochondrial toxicity was present within the scientific literature. The benefit of undertaking the analysis for each category is that it enabled the chemical space associated with a known, and tested, mechanism of mitochondrial toxicity to be identified. The development of chemical categories and identification of additional mechanistic information from the literature was crucial in addressing the limitations of the information in the original dataset.

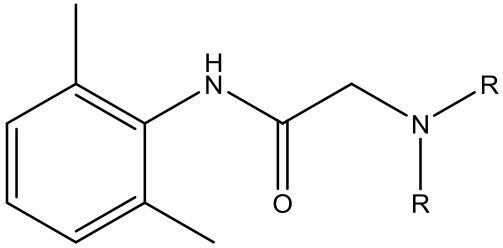
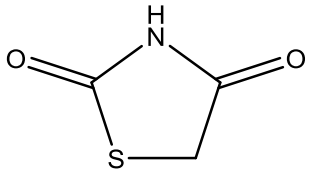
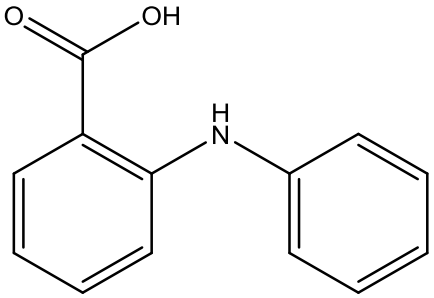
4.3 Results and discussion

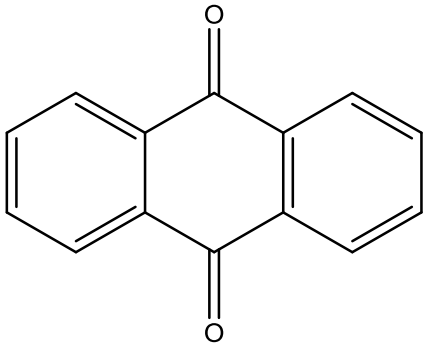
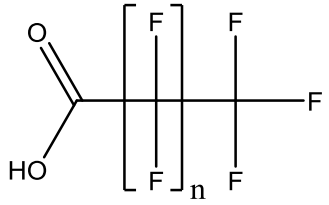
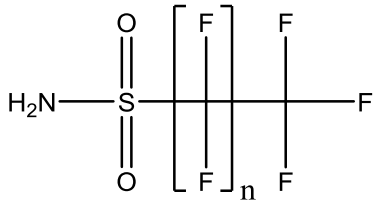
The aim of this chapter was to develop an *in silico* profiler for mitochondrial toxicity based around clearly defined mechanistic information. This was achieved by grouping chemicals based upon their structural similarity, followed by a literature search to elucidate mechanistic information for the chemicals in categories associated with toxicity to mitochondria. Overall, 35 of the 288 chemicals were identified as belonging to categories containing toxic chemicals: local anaesthetics, anti-anginal, and anti-arrhythmic; anti-diabetic drugs; non-steroidal anti-inflammatory drugs; anthracycline antibiotics; hypolipodemic drugs; bile acids; anti-histaminic, anti-psychotic and anti-emetic drugs; and β-blockers. A summary of the categories developed within this chapter is shown in Table 4.1. In total, eight structural alerts were formed: two separate molecular initiating events for the hypolipodemic drugs category were identified, whilst no structural alert for the β-blocker category could be defined. A summary of the associated structural alerts developed within this chapter is shown in Table 4.2. These structural alerts cover five mechanisms of mitochondrial toxicity: inhibition of the electron transport chain, alternative electron acceptance, initiation of the death receptor pathway, uncoupling of oxidative phosphorylation and induction of the membrane permeability transition.

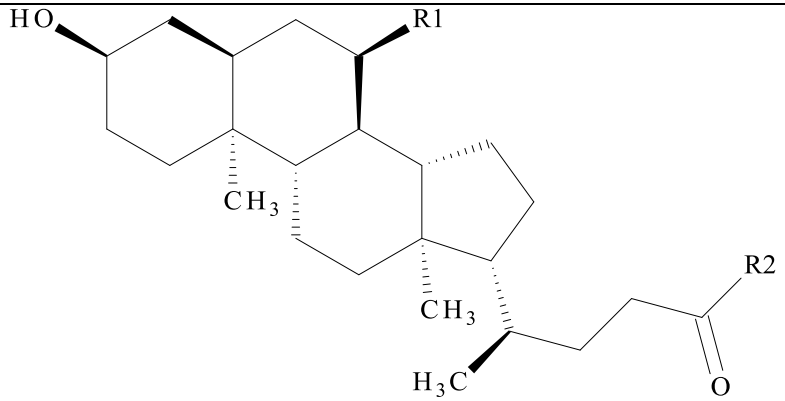
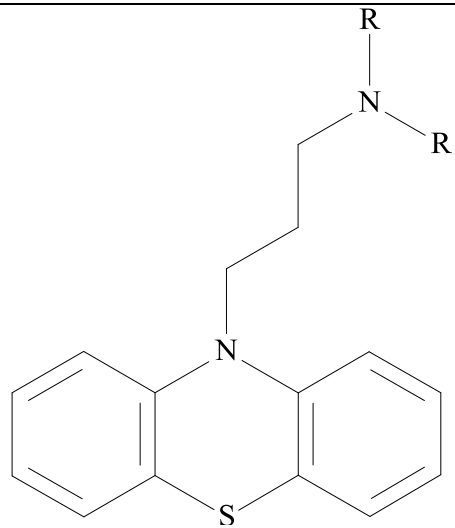
Table 4.1: Chemicals grouped into categories using structural similarity and their associated mitochondrial toxicity

Category	Name	Mitochondrial toxicity
1	Lidocaine	Positive
	Bupivacaine	Positive
	Etidocaine	Positive
	Ropivacaine	Positive
	Ranolazine	Positive
	Tocainide	Negative
2	Rosiglitazone	Positive
	Pioglitazone	Positive
	Troglitazone	Positive
3	Mefenamic acid	Positive
	Flufenamic acid	Positive
	Tolfenamic acid	Positive
4	Daunorubicin	Positive
	Doxorubicin	Positive
	Epirubicin	Positive
	Idarubicin	Positive
5	Perfluorodecanoic acid	Positive
	Perfluorooctanoic acid	Positive
	Perfluorooctane-sulphonamide	Positive
6	Cholic acid	Positive
	Chenodeoxycholic acid	Positive
	Deoxycholic acid	Positive
	Glycocholic acid	Positive
	Lithocholic acid	Positive
	Taurocholic acid	Positive
7	Promethazine	Negative
	Chlorpromazine	Positive
	Fluphenazine	Positive
	Mequitazine	Negative
	Methdilazine	Negative
	Thiethylperazine	Negative
	Trimeprazine	Negative
8	Alprenolol	Negative
	Atenolol	Positive
	Propranolol	Positive

Table 4.2: Structural alerts developed in the current chapter

Category	Structural alert	Number of mitochondrial toxic chemicals in the category (mitotoxic/total)	Mechanism(s) associated with the structural alert
1	 R = CH_2, CH_3	5/6	Uncoupling of oxidative phosphorylation
2		3/3	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation
3		3/3	Uncoupling of oxidative phosphorylation

4		4/4	Inhibition of the electron transport chain
5a	 n = 5-11	2/2	Induction of mitochondrial membrane permeability transition
5b	 n = 5-11	1/1	Uncoupling of oxidative phosphorylation

6	 R1 = OH, H R2 = OH, NHR (R = CH_2)	6/6	Inhibition of the electron transport chain and the extrinsic pathway
7	 R = CH_2, CH_3	2/7	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation

A more detailed discussion surrounding each of the chemical categories and the mechanism by which mitochondrial toxicity is initiated is presented below.

Category 1: Local anaesthetics, anti-anginal and anti-arrhythmic

The local anaesthetics category consisted of six analogues, four of which are anaesthetics, with ranolazine and tocainide being an anti-anginal and anti-arrhythmic respectively. All but one of the chemicals, tocainide, has been shown to exhibit toxicity towards mitochondria enabling a single structural alert to be defined (Table 4.2). The structural alert is supported by a number of studies that have shown that such chemicals affect mitochondrial metabolism by uncoupling oxidative phosphorylation (Dabadie *et al* 1987, Wallace and Starkov 2000, Dippenaar 2007, Mehta *et al* 2008, Cela *et al* 2010). This uncoupling has been suggested to be mediated by both the protonophoric properties and the pKa of these chemicals. As the pKa is relatively similar to the intracellular pH, the level of protonated and deprotonated chemical is roughly at equilibrium. The presence of deprotonated chemical within the intermembrane space means that protons can be scavenged. Subsequently, the protonated chemical can combine with a hydrophobic anion to form a neutral ion-pair complex, which can then migrate across the inner mitochondrial membrane into the matrix, where the complex dissociates and the proton is released. Both the chemical and the hydrophobic anion then return to the intermembrane space, continuing the cycle. This assisted transport of protons back into the matrix dissipates the electrochemical potential, resulting in a loss of ATP production and ultimately cell death (Terada 1990, Schonfeld *et al* 1992, Sun and Garlid 1992, Sztark *et al* 1997, Wallace and Starkov 2000, Mehta *et al* 2008, Cela *et al* 2010). It has been suggested that bupivacaine, and other highly lipophilic anaesthetics, can also act to uncouple oxidative phosphorylation via the mechanism outlined above without the need to complex with a lipophilic anion (Dabadie *et al* 1987). The structural alert defined for this category can be seen in Table 4.2. The presence, and pKa, of the tertiary amine group is thought to be responsible for the ability of these chemicals to scavenge protons within the intermembrane space. Therefore, the lack of a tertiary amine group offers an

explanation as to why no mitochondrial toxicity has been associated with tocainide and also allows for further refinement of the structural alert.

Category 2: Anti-diabetic drugs

This category consists of three thiazolidinediones: pioglitazone, rosiglitazone, troglitazone: each of which were identified as inducing mitochondrial toxicity. Thiazolidinediones are the major orally administered drugs used in the treatment of Type 2 (non-insulin dependent) diabetes. These drugs are used to improve insulin sensitivity and lower blood glucose levels within diabetic patients. Many of the thiazolidinediones have been suspected of initiating hepatotoxicity via mitochondrial dysfunction (Dykens and Will 2007). For example, troglitazone was withdrawn from the world market in 2000 due to hepatotoxicity observed in a number of patients (Chan *et al* 2005, Mehta *et al* 2008).

Research into the thiazolidinediones suggests the chemicals within this category elicit their mitochondrial dysfunction by inhibiting the electron transport chain and uncoupling oxidative phosphorylation (Brunmair *et al* 2004, Dykens and Will 2007, Mehta *et al* 2008, Naven *et al* 2013). These drugs have been shown to inhibit the activity of mitochondrial complexes, the main target being Complex I (Brunmair *et al* 2004, Chan *et al* 2005, Nadanaciva *et al* 2007a, Nadanaciva *et al* 2007b, Mehta *et al* 2008). These chemicals subsequently induce mitochondrial swelling and decrease the membrane potential across the inner mitochondrial membrane, in turn inducing mitochondrial permeability transition (Masubuchi 2006, Nadanaciva *et al* 2007a, Nadanaciva *et al* 2007b). Additionally, thiazolidinediones have been shown to uncouple oxidative phosphorylation in a manner similar to that described above for the chemicals within category one (Brunmair *et al* 2004, Naven *et al* 2013). The structural alert defined for this category is shown in Table 4.2. It is thought that the properties that enable the thiazolidinediones to bind to the nuclear PPAR-gamma receptor confers the ability to bind to Complex I (Brunmair *et al* 2004). Additionally,

the heterocyclic properties of the ring system are thought to enable the thiazolidinedione to cycle between a protonated and deprotonated form conferring the ability to transport protons across the inner mitochondrial membrane thereby uncoupling oxidative phosphorylation (Naven *et al* 2013).

Category 3: Nonsteroidal anti-inflammatory drugs

The third category comprises three chemicals: mefenamic acid, flufenamic acid, and tolfenamic acid: each of which has been identified as being able to induce mitochondrial toxicity. Each of these three chemicals are part of a group of nonsteroidal anti-inflammatory (NSAIDs). NSAIDs are some of the most widely used pharmaceutical drugs on the market that are used for their analgesic, anti-pyretic and anti-inflammatory properties to reduce and relieve symptoms for a variety of conditions. In order for the anti-inflammatory properties associated with NSAIDs to be present a carboxylic acid moiety is needed (Mehta *et al* 2008). The carboxylic acid moiety acts to inhibit cyclooxygenase activity, an enzyme responsible for the production of mediators of the inflammatory response, thereby reducing the level of inflammatory signalling. Previous research substantiates the positive mitochondrial toxicity result for each chemical within this category. A variety of literature sources identify each of these chemicals as having the ability to uncouple oxidative phosphorylation via a similar mechanism as that described above for the lidocaine category (Uyemura *et al* 1997, Masubuchi *et al* 1998, Moreno-Sanchez *et al* 1999, Boelsterli 2002, Siraki *et al* 2005). However, due to their lipophilicity, these chemicals do not necessarily need to be associated with a separate hydrophobic anion in order to translocate into the mitochondrial matrix. A single structural alert could be developed for this category, as shown in Table 4.2. The carboxylic acid moiety, which is required for the anti-inflammatory properties of the NSAIDs, is believed to also be required to induce the uncoupling ability of this group of chemicals (Mehta *et al* 2008).

Category 4: Anthracycline antibiotics

Anthracycline antibiotics are a group of hydroxylated tetracycline quinones with a duanosamine sugar sidechain attached. One of the category members, doxorubicin, is one of the most widely used anti-neoplastic drugs within the U.S. (Wallace 2003). Structural similarity identified three similar chemicals. A number of studies have shown that the anthracycline antibiotics cause mitochondrial dysfunction by acting as alternative electron acceptors interfering with, and inhibiting, the electron transport chain, leading to oxidative stress. This occurs because under normal physiological conditions anthracyclines are usually deprotonated and can permeate across the outer mitochondrial membrane. Once within the intermembrane space these chemicals disrupt the electron transport chain by sequestering an electron from Complex I and are thus reduced to a semiquinone radical intermediate (Gerwirtz 1999, Wallace 2003). These semiquinone radicals subsequently interact with molecular oxygen present within the mitochondria, producing reactive oxygen species (ROS), including hydroxyl and superoxide anion radicals. Downstream these ROS lead to a variety of effects such a mitochondrial permeability transition induction and oxidative damage of DNA, proteins and lipids (Kappus 1986, Ohkuma *et al* 2001, Kim *et al* 2003). Due to the high level of similarity between the chemicals it can be assumed that the mechanism of action is conserved throughout the category. Analysis of the literature enabled a structural alert to be defined based upon the quinone-type moiety (Table 4.2).

Category 5: Hypolipodemic drugs

Perfluorinated chemicals have been widely used in a variety of commercial and pharmaceutical products, such as flame retardants, surfactants and hypolipidemic drugs. These hypolipidemic drugs induce the proliferation of peroxisomes and thus increase β-oxidation of fatty acids. Three perfluorinated chemical analogues; perfluorodecanoic acid, perfluorooctanoic acid, perfluorooctane sulphonamide; were identified as having a high level

of similarity. However, despite the high levels of similarity between the chemicals multiple mechanisms were seen to induce mitochondrial dysfunction. This highlights the need to undertake mechanistic analysis of the categories as structural similarity on its own is not enough. As is shown with this category slight variations in structure have the potential to induce different mechanistic pathways. Accordingly, information in the literature suggests that for this category there are two potential mechanisms by which the perfluorinated chemicals elicit their mitochondrial toxicity; uncoupling of oxidative phosphorylation and induction of the mitochondrial membrane permeability transition pore.

Perfluorooctane sulphonamide has been shown to uncouple oxidative phosphorylation *in vitro* via a protonophoric mechanism, similar to that described above, within various species (Schnellmann 1990, Schnellmann and Manning 1990, Starkov and Wallace 2002, Wallace *et al* 2013). In comparison to *p*-trifluromethoxyphenylhydrazone, one of the most potent uncouplers, perfluorooctane sulphonamide has been known to uncouple oxidative phosphorylation with a potency of a similar magnitude. It has been suggested that the pKa and ionisability of the amino acid moiety, in conjunction with the relatively high lipophilicity of the chemical, enables the shuttling of protons across the inner mitochondrial membrane into the matrix, dissipating the membrane potential (Starkov and Wallace 2002). In addition, perfluorooctane sulphonamide is one of a very limited number of uncoupling chemicals that does not contain a ring structure (Schnellmann and Manning 1990).

In contrast, the perfluoroalkyl acids are believed to induce the mitochondrial membrane permeability transition at lower concentrations, whilst higher concentrations can uncouple oxidative phosphorylation (Langley 1990, Keller *et al* 1992, Starkov and Wallace 2002). It has been observed that perfluorodecanoic acid forms reactive oxygen species (ROS), hydrogen peroxide and peroxynitrite anion (Kleszczynski *et al* 2009). The presence of elevated ROS levels initiates oxidative stress within mitochondria. Oxidative stress has been shown to induce the membrane permeability transition (MPT) (Kowaltowski *et al* 2001, Battaglia 2005). The MPT is an increase in permeability of the inner mitochondrial

membrane to low molecular weight solutes (<1500 Daltons). The subsequent influx of solutes into the matrix instigates swelling of the inner and outer membranes causing disruption of the electron transport chain and a release of apoptotic proteins such as cytochrome *c* (Kleszczynski 2009, Kleszczynski and Skladanowski 2011, Wallace *et al* 2013). The uncoupling action of the two perfluoroalkyl acids is similar to that described for category one. Two alerts were defined due to two distinct MIEs being identified (Table 4.2).

Previous research has shown there to be an increase in toxicity concomitant to an increase in alkyl side chain length up to C_{12}, with the most marked increase in toxicity (a five- to ten-fold increase) occurring between C_6 and C_8 perfluoroalkyl acids and sulphates (Wallace *et al* 2013). An unsubstituted amide fragment has been shown to be required in order for uncoupling by perfluorinated sulphonamides to occur: fully substituted sulphonamides, which lack the protonated amide moiety, were found to lack the ability to uncouple oxidative phosphorylation (Starkov and Wallace 2002). The carboxylic acid moiety of the perfluoroalkyl acids chemicals is thought to be responsible for the uncoupling action of these chemicals at higher concentrations. However, it is unclear which fragments are required in the induction of the MPT.

Category 6: Bile acids

The bile acid category consists of three secondary bile acids: chenodeoxycholic acid, glycocholic acid and taurocholic acid: and two conjugated bile acids: deoxycholic acid and lithocholic acid: all with a high level of similarity to the primary bile acid; cholic acid. Bile acids are one of the main constituents of bile and are synthesised from cholesterol by hepatocytes. They play a vital role in multiple functions within both the liver and intestines, the main function being the sequestration of fats within micelles for excretion. Bile acids have been shown to decrease the membrane potential of mitochondria, alongside a decrease in state 3 respiration and an increase in state 4 respiration. The specific cellular mechanism

of bile acid-induced toxicity has not been elucidated. However, both the intrinsic (mitochondrial) and extrinsic (death receptor) pathways have been implicated in the disruption of normal mitochondrial function.

Intrinsic pathway

The intrinsic apoptotic pathway results in mitochondrial dysfunction due to an increase in intracellular stress. Hydrophobic bile acids have been shown to inhibit the electron transport chain by decreasing the activity of complexes I, III and IV, resulting in a decrease in state 3 respiration and a concomitant generation of ROS (Krahenbuhl *et al* 1994, Winklhofer-Roob *et al* 1996, Sokol *et al* 2001, Yerushalmi *et al* 2001, Palmeira and Rolo 2004, Perez 2009). It has been proposed that the inhibition of complex III leads to a subsequent electron leak through the ubiquinone-complex III site and a concomitant ROS generation (Winklhofer-Roob *et al* 1996, Yerushalmi *et al* 2001). The increased oxidant stress may then cause the induction of the MPT by oxidation of the thiol sites on the membrane permeability transition pore (Sokol *et al* 2001). Induction of the MPT triggers the release of cytochrome *c*, thus, stimulating the translocation of Bax to the mitochondrial membrane, stimulating further release of cytochrome *c* (Spivey *et al* 1993, Rodrigues *et al* 1999, Yin and Ding 2003). Cytochrome *c* is also able to initiate downstream caspase activation events discussed below (Yerushalmi *et al* 2001, Yin and Ding 2003, Palmeira 2004, Taylor *et al* 2008, Perez 2009).

Extrinsic pathway

Mitochondrial dysfunction and apoptosis initiated via the extrinsic pathway results from extracellular signals triggering downstream caspase activity. The oxidative stress generated by bile acids induces an increased presentation of Fas receptor within the plasma membrane, following phosphorylation of the Fas receptor by the epidermal growth factor receptor (Faubion *et al* 1998, Qiao *et al* 2001, Perez 2009). Upon presentation of Fas within the plasma membrane Fas agonists can interact with the receptor, initiating the formation of the death-inducing signalling complex (DISC) and subsequent activation of caspase-8. In turn,

caspase-8 activates caspases-3 and -7 triggering a caspase cascade that culminates in apoptosis (Faubion *et al* 1998, Jaeschke *et al* 2002, Yin and Ding 2003, Taylor *et al* 2008, Perez 2009). Additionally, caspase-8 can initiate the intrinsic pathway via proteolytic cleavage of Bid. Truncated Bid activates Bax and Bak proteins present on the mitochondria via oligomerisation and induction of MPT. The activated Bak and Bax proteins form channels within the mitochondria releasing additional cytochrome c. Cytosolic cytochrome c causes the assembly of apoptotic protease-activating factor-1 (APAF-1) and caspase-9, thus, activating caspase-9. Upon caspase-9 activation a proteolytic caspase cascade is initiated ultimately leading to cell death (Yin and Ding 2003, Taylor *et al* 2008). The generation of ROS, induction of MPT and activation of the caspase cascade seem to be essential steps within both pathways to initiate mitochondrial perturbation and apoptosis. Therefore, it seems likely that both pathways work synergistically to induce mitochondrial dysfunction. Together these chemicals enabled the definition of a single structural alert; based on the steroid structure to be defined (Table 4.2).

Category 7: Anti-histaminic, anti-psychotic and anti-emetic drugs

Phenothiazines are a group of heterocyclic chemicals composed of a nitrogen and a sulphur atom joining two benzene rings. These chemicals are widely used in the treatment of mental disorders, such as schizophrenia, psychosis, and anxiety, as well as conferring anti-histaminic and anti-emetic action. This category comprises seven chemicals: chlorpromazine, fluphenazine, mequitazine, methdilazine, promethazine, thiethylperazine, and trimeprazine: five of which have been identified have been identified as being non-toxic, whilst the remaining two have been shown to be toxic (Table 4.1). It is important to rationalise the mixed toxicity results for the chemicals within this category. A number of studies in the literature report toxicity induced by chlorpromazine and fluphenazine (Saito *et al* 1982, Lucas-Heron *et al* 1994, Balijepalli *et al* 1999, Chan *et al* 2005, Nadanaciva *et al* 2007a,

Mehta *et al* 2008) that corroborate the data in the current chapter (obtained from Zhang *et al* 2009).

Both chlorpromazine and fluphenazine have been observed to inhibit mitochondrial respiration within brain and liver tissues (Guth *et al* 1964). This toxicity was induced by binding to, and inhibiting, Complex I of the electron transport chain (Chan *et al* 2005, Mehta *et al* 2008, Nadanaciva and Will 2011). Further investigation revealed that chlorpromazine is also capable of impairing mitochondrial function by inhibiting Complex IV and acting as an uncoupler of oxidative phosphorylation (Matsubara and Hagihara 1968, Eto *et al* 1985, Mehta *et al* 2008). Eto *et al* (1985) noted the addition of the chlorine atom increases and alters the mechanism by which mitochondrial toxicity occurs, i.e. chlorpromazine acts as an uncoupler of oxidative phosphorylation at low concentrations and an electron transport chain inhibitor at higher concentrations.

Each chemical within this category contains a phenothiazine fragment. This class of drugs were found to cause toxicity towards mitochondria by inhibiting oxidative phosphorylation within liver mitochondria (Gallagher *et al* 1965). Due to this conserved fragment it can be hypothesised that the other chemicals within this category may elicit toxicity via a similar mechanism. Research into promethazine has shown that it can act as an uncoupler of oxidative phosphorylation by impeding both state 3 and state 4 respiration and intramitochondrial potassium ion compartmentalisation at high and low concentrations (Matsubara and Hagihara 1968, Eto *et al* 1985). Further investigation into chlorpromazine reveals that this chemical elicits its electron transport chain inhibitor action by inhibiting Complex V of the electron transport chain. The associated structural alert is as shown in Table 4.2. Based upon information in the literature the tertiary amine moiety is required in order to initiate the uncoupling of oxidative phosphorylation; whilst the phenothiazine fragment with an associated electron-withdrawing group leads to an increase in toxicity (Cela *et al* 2010, Cruz *et al* 2010, Eto *et al* 1985, Matsubara and Hagihara 1968, Terada 1990).

Category 8: β-blockers

Alprenolol, propranolol and atenolol are a group of (non-)selective β-blockers used in the treatment of hypertension. As can be seen in Table 4.1 two chemicals, atenolol and propranolol, were reported within Zhang *et al* as inducing mitochondrial toxicity, whilst the remaining chemical, alprenolol, was reported as being negative for mitochondrial toxicity. Propranolol has been seen to inhibit, via non-competitive binding, Complex V of the respiratory chain (Wei *et al* 1985, Almotrefi and Dzimiri 1992). Chemicals that inhibit Complex V can do so by binding to one of two subunits (F_0/F_1) that comprise the ATP synthase enzyme, thus blocking the passage of protons back into the mitochondrial matrix (Wei *et al* 1985). Together the membrane-bound F_0 and matrix protruding F_1 subunits are responsible for catalysing both the synthesis and hydrolysis of ATP. Wei *et al* have described previously that propranolol binds to the Mg^{2+}-ATPase (F_0 subunit) of Complex V inhibiting state 3 respiration. It has also been seen that the potency of ATPase inhibition induced by propranolol is of the same order of magnitude as its ability to inhibit other membrane-bound enzymes (Almotrefi and Dzimiri 1992). Therefore, this inhibitory effect induced by propranolol is due to its membrane stabilising activity and its ability to bind to the lipophilic F_0 subunit of Complex V. In contrast, atenolol, a relatively more hydrophilic drug, has been shown to act via stimulating Complex V activity. The decrease in lipophilicity and, therefore, a decrease in ability to penetrate and interact with membrane macromolecules is pertinent to the decrease in inhibitory potency of atenolol (Almotrefi and Dzimiri 1992). Additionally, results from Almotrefi and Dzimiri (1992) suggest that atenolol may interact with the more hydrophilic subunit (F_1) of Complex V, resulting in mitochondrial toxicity by stimulating the hydrolysis of ATP to ADP and inorganic phosphate (Almotrefi and Dzimiri 1992). As propranolol and atenolol elicit their mitochondrial toxicity via separate, and contrasting, mechanisms a structural alert could not be defined. In order to overcome this, further testing is required to elicit more information regarding the mitochondrial toxicity of other, structurally similar, β-blockers. This category

further highlights the necessity of undertaking mechanistic analysis of categories formed using structural similarity prior to defining a structural alert.

Categories one through seven have enabled the development of structural alerts for mitochondrial toxicity, as the chemicals in the same category initiate the same toxicity pathway. These structural alerts may be used for read-across purposes within risk assessment. Meanwhile, category eight highlights an area where further investigation, and testing, is needed based around the differing pathways initiated by structurally similar chemicals.

4.3.1 Profiling and grouping for mitochondrial toxicity as part of the AOP paradigm

The ability to predict organ-level toxicity will become increasingly important to the long term goal of replacing animal use in determining a Lowest Observed (Adverse) Effect Level (LO(A)EL). Traditionally, LO(A)ELs are identified after undertaking a 28- or 90-day repeated dose study, with the lowest dose initiating a treatment related adverse effect in an organ(s) producing the LO(A)EL value. However, as no animal testing is permissible for cosmetic ingredients in Europe alternatives are required. As discussed previously this requirement has led to increased interest in the understanding of toxicity pathways and in the development of AOPs. As such the structural alerts that have been developed in this chapter are intended for use in chemical risk assessment within the AOP paradigm. Importantly for the data within the current chapter, it has been reported that toxicity to a number of organs is likely to be driven by toxicity to mitochondria (Amacher 2005, Dykens and Will 2007, Nadanaciva and Will 2011, Vinken *et al* 2013a, Vinken *et al* 2013b). Therefore, in order for a full AOP to be developed, further investigation into the organ(s) affected is required; this was, however, beyond the scope of the current chapter. The main outcome from the current chapter is that the structural alerts defined enable chemicals to be grouped into mechanistically-based categories based around the knowledge of a number of key MIEs for

mitochondrial toxicity. The resulting categories can thus be used for either prioritisation of chemicals for further *in vitro* testing or, where sufficient *in vivo* data exist, for read-across predictions of organ-level toxicity (from, for example, repeat dose toxicity testing). In terms of predicting organ-level toxicity in the future it is likely that additional steps in the AOP will need to be investigated within *in vitro* assays using a range of organ specific cell lines. For example, the use of primary human renal proximal tubule epithelial cells in the MTT assay to investigate nephrotoxicity due to mitochondrial dysfunction. This will enable a mechanistically-based weight of evidence to be constructed based around the AOP. Currently, chemistry-based grouping methods such as those outlined above offer the most immediate solution to risk assessment without using animals.

4.4 Conclusions

The aim of this chapter was to develop an *in silico* profiler for mitochondrial toxicity based around clearly defined mechanistic information utilising structural similarity and chemical category formation. The analysis resulted in the development of eight chemical categories and the definition of eight (mechanism-based) structural alerts. Of the alerts developed within this chapter, seven have not been defined in terms of the mechanism by which they initiate mitochondrial toxicity; whilst, the remaining alert (thiazolidinedione) has been identified previously by Naven *et al* (2013). Importantly, these structural alerts were derived using mechanistic information in the available literature to elucidate knowledge of a number of key Molecular Initiating Events that disrupt the normal functioning of mitochondria. It is envisaged that structural alerts, such as those defined in this chapter, will be combined with other alerts pertaining to mitochondrial toxicity; such as those within Chapters 3 and 5 and in the available literature; to develop a single profiler (discussed in Chapter 7). This profiler could be useful for grouping chemicals into categories, thus, enabling predictions to be made regarding mitochondrial toxicity. Additionally, the work discussed in Chapters 3 and 4

demonstrates how different *in silico* tools can be utilised in the identification of structural alerts.

Chapter 5: Development of an *in silico* profiler for categorisation of repeat dose toxicity data of hair dyes[†]

5.1 Introduction

Each year millions of people worldwide use hair dye products. It is estimated that over one third of women aged 18 or over, and approximately ten percent of men aged 40 or over, in the United States and Europe, use at least one type of hair dye product (Huncharek and Kulpelnick 2005). Hair dyes can be separated into three classes: temporary, semi-permanent, and permanent. Permanent, or oxidative, hair dyes are the most widely used class of hair dyes, accounting for approximately 80% of the hair colouring product market in the US and EU (Corbett *et al*. 1999, Cosmetics Europe 2014). This class of hair dyes is different to the other two classes in respect to their composition: oxidative dyes require a chemical reaction between a primary intermediate and a coupler in order to generate the coloured dye on/in the hair (Nohynek *et al*. 2010). The primary intermediates are normally aryl diamine or aminophenol compounds substituted at either the *ortho*- or *para*- position, such as *p*-aminophenol. In contrast, couplers are normally aryl aminophenol or diphenols substituted at the *meta*- position, such as resorcinol. In the presence of a developer, such as hydrogen peroxide, the primary intermediate is oxidised and reacts with the coupler to produce a coloured aromatic dye (Nohynek *et al*. 2010) (Figure 5.1).

[†]This chapter is based on a published article by Nelms et al (2014)

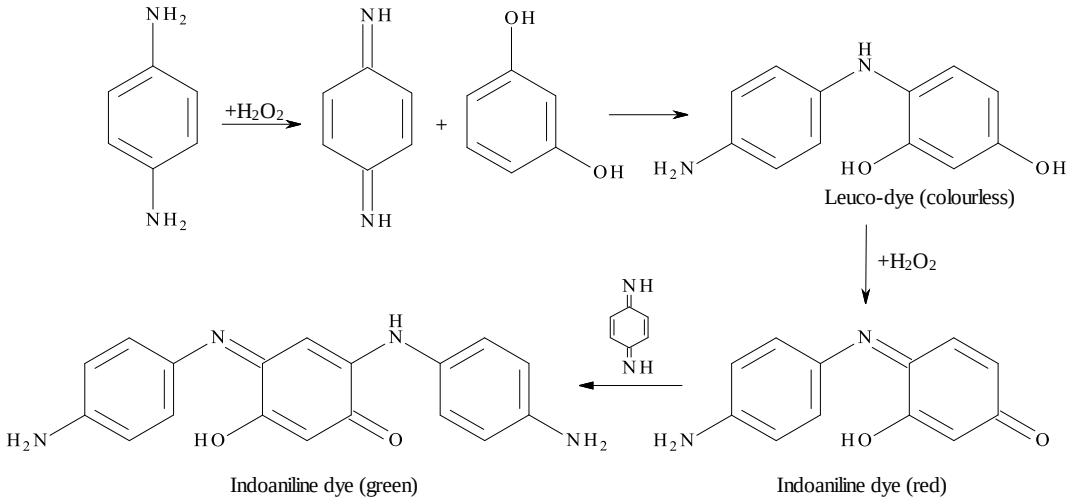


Figure 5.1: A suggested reaction pathway showing how the oxidative hair dye phenylenediamine (primary intermediate) results in a coloured dye in the presence of resorcinol (a coupler) and hydrogen peroxide (adapted from Nohynek *et al* 2010).

Typically, hair dye products contain between 0.05-2% primary intermediate, with the higher the percentage producing a darker shade of dye. In comparison, temporary and semi-permanent hair dyes are typically acidic or basic chemicals that bind to the proteins of hair and do not use developers or couplers. Typical classes of temporary and semi-permanent hair dyes include anthraquinones and nitroaminophenols respectively. The large number of people exposed to, and the reactivity of, hair dye products has led to them becoming some of the most widely studied cosmetic ingredients. A number of studies, both *in vitro* and *in vivo*, have raised concern about the carcinogenic potential of certain members of chemicals used within hair dye products (Baan *et al*. 2008; Freudenthal *et al*. 1999; Gago-Dominguez *et al*. 2001; IARC 2010; Skipper *et al*. 2010).

Previously, safety assessments for cosmetic ingredients, including hair dyes, would have been made, at least in part, using data from *in vivo* experimentation. However, significant changes in the European cosmetic and chemical legislations during the last decade have concentrated efforts in the development of alternative methods for safety testing purposes (EC 2003; EC 2007). The Adverse Outcome Pathway (AOP) paradigm has emerged as a

promising approach in that it enables key events in the pathway that leads to a toxicological outcome to be identified (Ankley *et al*. 2010; Vinken 2013; Vinken *et al*. 2013a). Key amongst these events is the Molecular Initiating Event (MIE), which has been the focus for the development of *in silico* profilers (Przybylak and Schultz 2013). These profilers define the chemical features associated with a given MIE in terms of collections of structural alerts and are intended to be used to categorise chemicals based on a common MIE (Enoch *et al*. 2011a; Enoch *et al*. 2013b; Enoch and Roberts 2013; Przybylak and Schultz 2013; Sakuratani *et al*. 2013a; Sakuratani *et al*. 2013b; Vinken 2013; Vinken *et al*. 2013a) (discussed in more detail in Chapters 1, 3 and 4). The development of mechanism-based *in silico* profilers suitable for category formation is a time-consuming, literature-intensive process. Previous research leading to the establishment of *in silico* profilers for toxicological endpoints such as skin and respiratory sensitisation utilised a mechanistic hypothesis as a starting point for structural alert development (Enoch *et al*. 2008; Enoch *et al*. 2012b). However, for complex endpoints such as organ-specific toxicity for which knowledge relating to possible MIEs is lacking, a chemoinformatics approach, coupled with *a posteri* mechanistic rationalisation, has been shown to be successful (Hewitt *et al*. 2013). Given the complexity of potential mechanisms driving oral repeat dose toxicity, the current chapter employed the latter approach using the protocol described hereafter. The mechanism-based categories of chemicals that result from such AOP-derived profilers are applicable to predict hazard via read-across and, hence, assist in the filling of data gaps. In addition, these groupings also form the basis for the more in-depth analysis that is required for an overall risk assessment. In such a situation, additional testing using *in vitro* and/or *in chemico* methods to assess other key steps in the AOP is required. The ability to group chemicals into mechanism-based categories using *in silico* profilers enables *in vitro* and/or *in chemico* assays to be developed to enable the prioritisation of chemicals (Gutsell and Russell 2013).

In order to generate structural alerts and, thus, mechanism-based chemical categories information pertaining to the endpoint, and chemicals, of interest are required. With respect

to the work undertaken in this chapter, and as part of the wider goal of the COSMOS project (discussed in Chapter 1), information relating to the repeat dose toxicity of cosmetic ingredients is required. One available source of toxicological data associated with cosmetic ingredients are the 'Opinion On' reports published by the Scientific Committee on Consumer Safety (SCCS) and its predecessors, the Scientific Committee on Cosmetic products and Non-Food Products intended for consumers (SCCNFP) and the Scientific Committee on Consumer Products (SCCP). The reports are generated for cosmetic substances for which some concern exists with regards to human health (e.g. colourants, preservatives, UV-filters and hair dyes) and contain data for a variety of toxicological endpoints, such as; skin irritation, acute toxicity, carcinogenicity and (sub-)chronic repeat dose studies. These reports usually contain No Observable Adverse Effect level (NOAEL)-values, and Lowest Observable Adverse Effect Level (LOAEL)-values generated by the repeat dose studies. NOAEL and LOAEL values are determined upon the completion of various repeated dose toxicity studies, such as (sub-) chronic, developmental or reproductive toxicity (discussed in more detail in Chapter 1). These data, ideally the NOAEL, are used by the SCCS, within the 'Opinion On' reports, in order to calculate the margin of safety (Figure 2.3, Chapter 2). Clearly, such data could provide a useful starting point for developing MIEs and identifying the chemistry required for the grouping of chemicals for read-across.

In particular for hair dyes, high quality toxicological data became available as a consequence of the step-wise strategy of the European Commission to regulate all hair dyes listed as substances in cosmetic products. The trigger for this action was the major concern of the scientific community for a putative link between the use of hair dyes and the development of cancer, with a focus on leukaemia and bladder cancer (Gago-Dominguez *et al*. 2001, IARC, Baan 2008, Huncharek 2005, Nohynek 2004, Skipper 2010). As such, industry was required to submit safety dossiers for hair dye components and possible mixtures for evaluation by the Scientific Committee on Consumer Safety and its predecessors. Despite the requirement

to assess the toxicity of hair dyes, few *in silico* models or structural alerts for their toxic effects, or rationale for their grouping, are currently available.

Therefore, the aim of this chapter is to develop an *in silico* profiler from a retrospective analysis of oral repeat dose toxicity data, available for hair dyes, retrieved from the Scientific Committees 'Opinions On' reports published between 2000 and 2013. These data were used to group hair dyes based upon structural similarity, with subsequent mechanistic analysis being undertaken using information from the peer reviewed literature. This mechanistic information, relating these structural alerts to potential MIEs, is important as it provides evidence for the interaction between the chemical and the biological system. The profiler could, thus, be used for a variety of process including screening data sets to identify chemicals of concern or to prioritise those chemicals that should undergo *in chemico*/*in vitro* testing first.

5.2 Methods

5.2.1 Experimental data

NOAEL values from oral 90-day rat studies for 94 hair dyes were extracted from the SCC(NF)P/SCCS 'Opinion On' reports published between 2000 and 2013, and provided to the current author, by Professors Vera Rogiers and Mathieu Vinken from Vrije Universiteit Brussel. Chemical names, CAS numbers and chemical structures were also taken from these reports. These data were used in the chemoinformatics analysis, described in more detail below, leading to the development of mechanism-based structural alerts. All data are available within Appendix III.

5.2.2 Development of in silico profiler

The workflow below (Figure 5.2) outlines the key steps in the development of the *in silico* profiler within this chapter. Step (1) is the formation of the initial chemical categories, based upon the similarity of one chemical to another within the data set. Step (2) is to identify the structural fragment that is conserved within each of the chemicals that populate the chemical category. Subsequently, the conserved fragment is encoded into a SMARTS pattern and used to identify further chemicals from the data set that were missed by the structural similarity analysis (Step (3)). It was also at this stage that LOAEL values, and the associated adverse effects, were extracted from the SCC(NF)P/SCCS reports: this information was utilised when attempting to identify the potential MIE for each category. Step (4) involved using the peer-reviewed literature, in conjunction with the adverse effect information, to identify a potential MIE for each category that had been developed in the previous step. Step (5) is to utilise the mechanistic knowledge of the potential MIE in order to identify additional structural alerts capable of triggering the same MIE. Whilst also collating each of the alerts developed in to one *in silico* profiler.

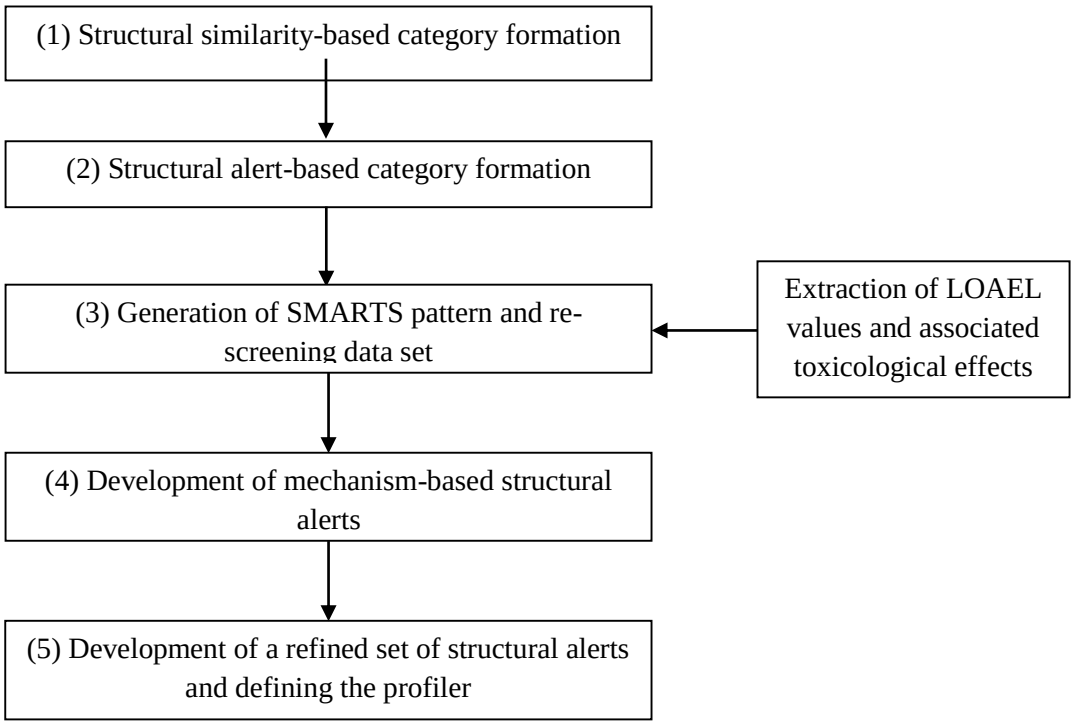


Figure 5.2: Workflow identifying the key steps in the procedure to develop an *in silico* profiler

Step 1 Structural similarity-based category formation

All chemical structures were encoded as SMILES strings, neutralised and salts removed *prior* to chemical similarity analysis. Structural similarity of each chemical to all others in the dataset was calculated using the atom environments/Tanimoto coefficient approach as implemented in the freely available Toxmatch software (V1.07) (available from http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/laboratories-research/predictive_toxicology/qsar_tools/toxmatch, accessed 17.11.2014) (discussed further in Chapter 4). Categories were developed for each chemical in the dataset using an in-house code implemented in Excel software that identified analogues with a similarity index of 0.7 or greater. The cut-off value of 0.7 was adapted from previous research by Enoch *et al* (2009), who identified that a value of 0.6 produced 'meaningful' categories for a diverse set

of chemicals. This adaptation was made in this chapter as the chemicals contained within the dataset were all hair dye chemicals and, therefore, assumed to have a relatively higher level of structural similarity given their use. Categories containing three or more chemicals were selected for further analysis.

Step 2 Structural alert-based category formation

Each similarity-based category containing three of more chemicals was inspected visually in order to identify key structural fragments present in all category members. This structural fragment was then encoded as a SMARTS pattern.

Step 2.1 Generation of SMARTS patterns from visual representation

The procedure, described in more detail in Chapter 3, was followed in order to develop SMARTS patterns from the structural fragments identified above.

Step 3 Use of initial SMARTS patterns to re-profile dataset

Subsequently, an in-house workflow, developed in the software package KNIME (v2.8.2), was utilised to re-profile each chemical in the dataset against these structural alerts (Figure 5.3). The re-profiling was carried out in order to expand the groupings to include chemicals that were not found by the structural similarity analysis. This is an important step in the protocol as pure structural similarity-based categories are frequently unable to detect chemicals containing the key structural fragments. Additionally, LOAEL data, and the associated adverse effects, were extracted from the SCC(NF)P/SCCS reports for each chemical within the categories formed.

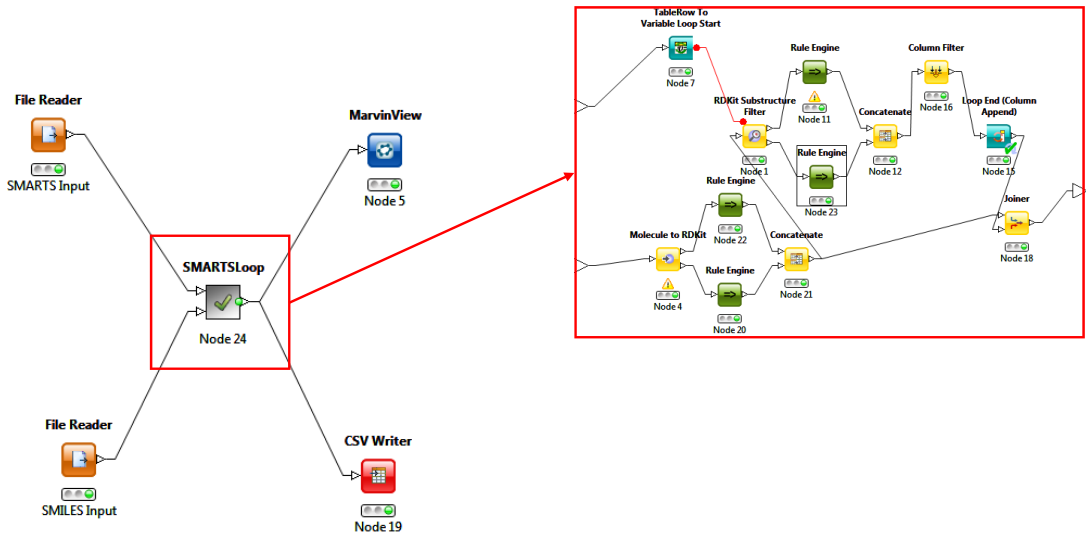


Figure 5.3: An overview of the workflow that screens a data set of chemicals for mitochondrial toxicity developed using the software program KNIME

Step 4. Development of mechanism-based structural alerts

Each of the structural alerts were then subjected to a mechanistic analysis involving detailed literature work in order to outline an MIE for the corresponding category members. This mechanistic analysis involved establishing potential MIEs related to chronic toxicity and linking them to the chemistry of the structural alerts. The literature work entailed performing keyword searches within a variety of scientific journal databases; including ScienceDirect and Web of Science; and using the Google Scholar search engine to find relevant full-text journal articles. The keywords used within these searches included:

- The name of the conserved structural fragment for each category or the common (or IUPAC) name of each of the chemicals within the category, and
- The type of toxicity observed or the organ in which the toxicity was observed

A search including the keyword ‘mitochondria’, in addition to the above keywords, was undertaken separately. This additional keyword was used as it was believed that

mitochondrial toxicity may explain the observation of toxicity in multiple organs by chemicals in the same category. Structural alerts were only considered as robust if a clear correlation between their chemistry and an MIE identified from relevant scientific literature could be established.

Step 5 Development of a refined set of structural alerts and in silico profiler

The final stage in the analysis was to use the mechanistic knowledge to extend the applicability domain of the structural alerts enabling an *in silico* profiler to be developed. This analysis involved identifying additional structural alerts capable of triggering the same MIEs based on chemical information. The mechanistic rationale for these additional structural alerts was supported by evidence drawn from the scientific literature. All structural alerts identified in this chapter were then converted into SMARTS patterns and collated into an *in silico* profiler that allowed chemicals capable of causing the same MIE to be assigned to a single category. In keeping with the development of previous *in silico* profilers present in the literature, the structural alerts were described within the resulting *in silico* profiler based on commonality of the underlying chemistry.

5.3 Results and discussion

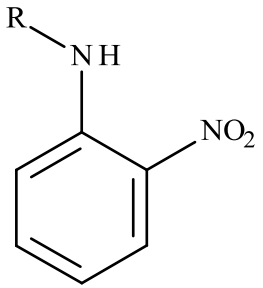
The aim of this chapter was to develop an *in silico* profiler suitable for chemical categorisation of oral repeat dose toxicity data of hair dyes. The analysis involved utilising chemical similarity to identify groups of chemicals from a dataset of 94 hair dyes. The similarity analysis (conducted in step 1) identified four categories of hair dyes containing either a 2-nitroaminobenzene, 4-nitroaminobenzene, aromatic azo or anthraquinone moieties. These key structural fragments were used to develop a mechanistic hypothesis for the MIE for each category. This analysis resulted in the definition of four structural alerts related to

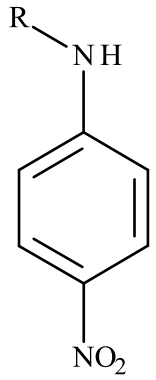
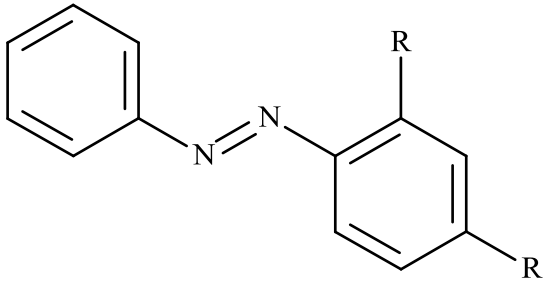
the ability of aromatic chemicals to disrupt mitochondrial function due to their free radical chemistry. This mechanistic chemistry allowed an *in silico* profiler containing a refined set of structural alerts to be defined. The resulting *in silico* profiler assigned 56 of the 94 chemicals in the dataset to a mechanism-based chemical category. However, further experimental analysis is required to identify additional key steps to allow an AOP (or AOPs) to be defined.

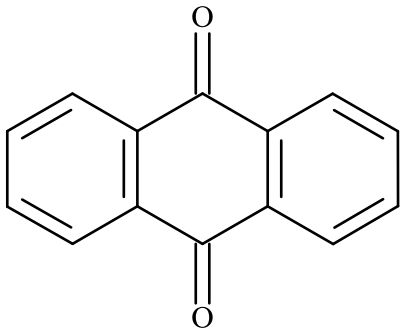
5.3.1 Development of mechanism-based structural alerts for category formation

The chemoinformatics analysis identified four similarity-based categories in the dataset, a category is defined here as a cluster containing three or more analogues. These included 2-nitroaminobenzenes, 4-nitroaminobenzenes, aromatic azos and anthraquinones. In all datasets, a structural alert was defined based on the key fragment in each of the clusters. These structural alerts were used to identify additional related chemicals not identified by the structural similarity analysis. This re-profiling is a crucial step in the development of mechanism-based structural alerts when using chemical similarity to cluster the initial dataset as related chemicals are frequently omitted. The resulting structural alerts and the number of analogues identified using them to re-screen the data are summarised in Table 5.1.

Table 5.1: Structural alerts identified from the similarity analysis carried out on the 93 hair dye chemicals

Structural alert	Key structural fragment	Number of analogues	Mechanism(s) associated with the structural fragment
2-nitroaminobenzenes	 R = hydrogen, carbon	21	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation

4-nitroaminobenzenes	 R = hydrogen, carbon	6	Inhibition of the electron transport chain and uncoupling of oxidative phosphorylation
Aromatic azos	 R = at least one must be NH_2, NH	9	Inhibition of the electron transport chain

Anthraquinones		5	Inhibition of the electron transport chain

Category 1: 2-nitroaminobenzene and 4-nitroaminobenzene and refined pro-quinone structural alerts

A total of 26 chemicals were identified using the 2-nitroaminobenzene and 4-nitroaminobenzene structural alerts, with one chemical (HC Yellow No. 10) triggering both alerts. The use of these chemicals was split between those chemicals that were used in both semi-permanent and permanent hair dye products, as well as those that were solely used in semi-permanent hair dye products. The 'Opinion On' reports show that these chemicals induce a variety of different toxicities within multiple organs, with no one organ predominantly exhibiting toxicity (Appendix III). One example is that of HC Orange No. 3 that has been seen to induce toxicity in the kidney, liver, and spleen, alongside increasing enzyme levels (SCCNFP 2003). The nitro group in these chemicals can be readily reduced to an amino moiety by nitroreductase via a hydroxylamine intermediate in the gut and the liver resulting in the production of 1,2- and/or 1,4-diaminobenzenes (Gorontzy *et al*. 1993; Roldan *et al*. 2008). These chemicals are then prone to oxidation to the corresponding 1,2- and/or 1,4-phenylenediamines (Figure 5.4).

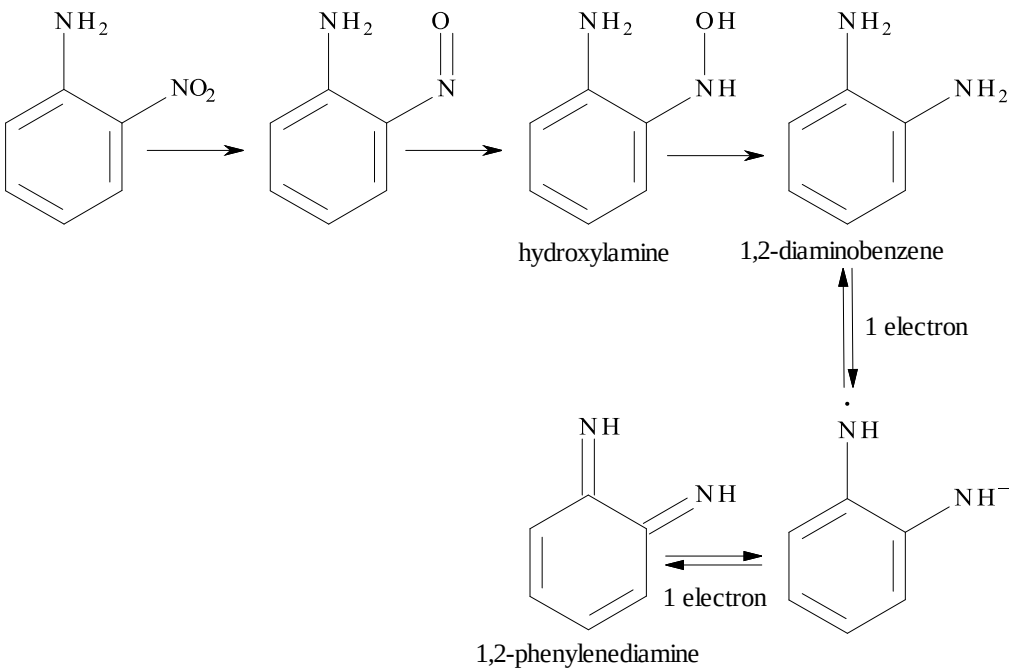


Figure 5.4: Reduction of 2-nitroaminobenzene to the corresponding 1,2-diaminobenzene and then subsequent oxidation to a 1,2-phenylenediamine

Importantly, the conversion of 1,2-diaminobenzenes into 1,2-phenylenediamines is reversible implying that these chemicals are capable of cycling electrons. This also holds true for the corresponding 1,4-diaminobenzenes. It is known that this electron cycling mechanism allows these types of chemicals to interfere with the electron transport chain within the mitochondria (Chapter 1) (Wallace and Starkov 2000). The mechanism leading to disruption could therefore involve the 1,2-diaminobenzene moiety within a chemical accepting an electron from respiratory Complex I. This could reduce the 1,2-diaminobenzene moiety to a 1,2-phenylenediamine which thereafter could transport the electron several steps down the respiratory chain directly into Complex VI. The release of the electron would then oxidise 1,2-phenylenediamine back to a 1,2-diaminobenzene allowing the process to be repeated in a cyclic fashion (Figure 5.5). This disruption ultimately could lead to a reduction in mitochondrial membrane potential and a subsequent reduction in ATP production (Bironaite *et al*. 1991; Chan *et al*. 2005; Munday 1992; Wallace and Starkov 2000).

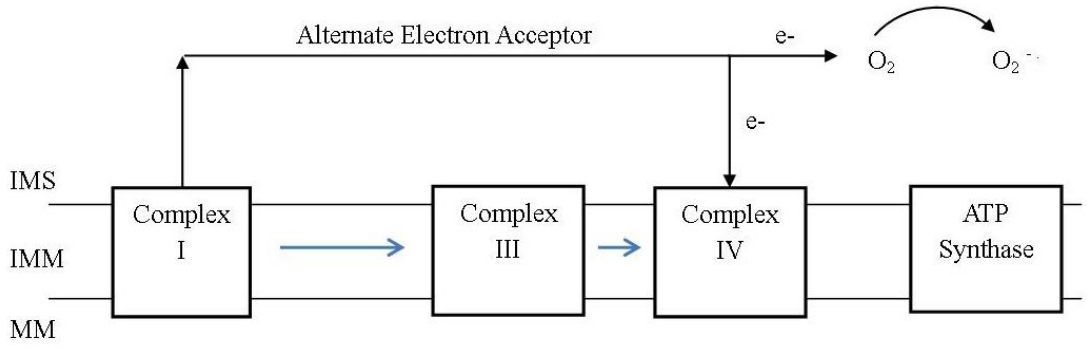


Figure 5.5: Electron cycling process leading to disruption of the respiratory chain in the mitochondria due to the presence of an alternate electron acceptor

The aromatic amine moiety of the reduction products is also known to induce uncoupling of oxidative phosphorylation via a protonophoric mechanism (Chapter 1) (Terada 1990a) (Figure 5.6). The deprotonated form of these compounds scavenges a free proton from the Intermembrane Space (IMS). Upon protonation the compound is able to migrate across the Inner Mitochondrial Membrane (IMM) into the Mitochondrial Matrix (MM). Due to the increased alkaline environment within the matrix the proton dissociates and the deprotonated compound returns to the intermembrane space enabling the cycle to continue. The continuation of this cycle increases oxygen consumption and heat production, alongside a reduction in the electrochemical gradient and ATP production (Chan *et al*. 2005; Pessayre *et al*. 2012; Terada 1990a; Wallace and Starkov 2000). Therefore, it is suggested that both mechanisms might contribute to the observed mitochondrial dysfunction.

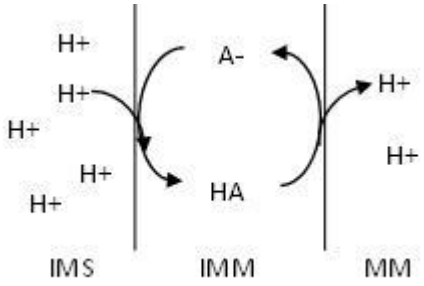


Figure 5.6: Cycling of the compound within the inner mitochondrial membrane (IMM), scavenging hydrogen ions from within the intermembrane space (IMS) and transporting them to the mitochondrial matrix (MM).

Based upon the mechanistic chemistry identified for the initial 2-nitroaminobenzene and 4-nitroaminobenzene structural alerts, discussed above, the alert was refined to cover pro-quinone chemicals substituted with one, or more, hydroxyl, nitro, or primary/secondary amine groups (or a combination thereof) (Table 5.2). This refinement could be made due to the extensive additional mechanistic chemistry knowledge in the wider literature relating to the types of chemicals that are readily converted to the corresponding quinones (Enoch *et al* 2011b, Kalgutkar *et al*. 2005). The refinement of this alert, to cover a wider spectrum of pro-quinones, significantly extended the number of chemicals assigned to the category: identifying twelve additional chemicals. As with the initial chemicals that comprised this category the additional chemicals are mainly used within both semi-permanent and permanent hair dye products. Again these chemicals have been seen to induce adverse effects within a variety of organs; however, the majority of these additional chemicals have an effect within the kidney.

Category 2: Anthraquinone and refined quinone structural alerts

The structural alert based on the anthraquinone moiety identified a total of five chemicals in the dataset. These chemicals are all used within semi-permanent hair dye products, with one chemical (acid blue 62) also being used in temporary hair dyes products. As with the

previous category a range of toxicities was observed, including increase in kidney weight, decreased body weight, increased alanine aminotransferase, increased cholesterol and decrease in motor activity (Appendix III). The majority of these toxicities were observed in the chemical acid blue 62 (SCCP 2005). These chemicals have also been shown to be capable of disrupting the electron transport chain in mitochondria by transporting electrons from respiratory Complex I directly to Complex IV (Henry and Wallace 1995; Kitani *et al*. 1981). This process is similar to that outlined for 1,2- and 1,4-diaminobenzenes in that the anthraquinone moiety accepts an electron from Complex I to become a semi-quinone radical. This radical species could transport an electron directly to Complex IV, being oxidised back to the anthraquinone in the process (Figure 5.7). Again, this reaction is reversible allowing the anthraquinone moiety to cycle electrons repeatedly from respiratory Complex I to Complex IV. In addition to acting as direct electron transport agents, the production of the semi-quinone radical has also been suggested to cause indirect mitochondrial toxicity due to their ability to react with molecular oxygen to produce reactive oxygen species. The chemical species include hydroxyl and superoxide radicals that are capable of evoking widespread damage to mitochondrial DNA, proteins and lipids (Kappus 1986; Ohkuma *et al*. 2001).

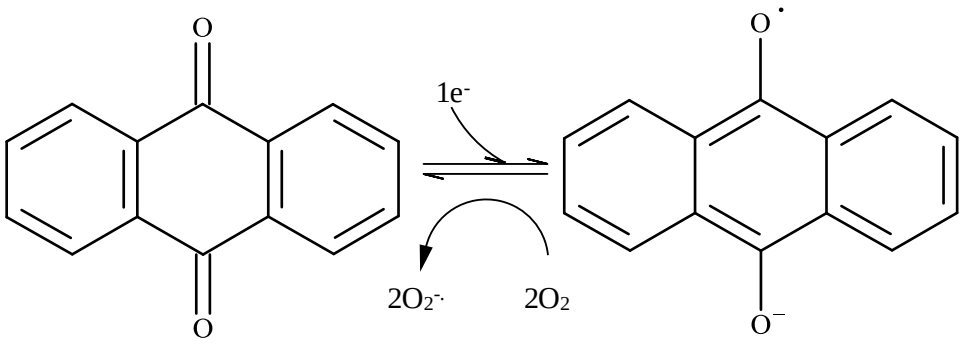


Figure 5.7: Activation of the anthraquinone moiety into a semi-quinone radical

The information pertaining to the mechanistic chemistry of the anthraquinone structural alert enabled the alert to be refined in order to cover chemicals containing a quinone moiety, without the necessity to be bound to two benzene rings (Table 5.2). This refinement was based on the related chemistry quinones exhibit and the proven ability of these chemicals to disrupt the respiratory chain in mitochondria via the same mechanism as described for the anthraquinones (Henry and Wallace 1995; Kitani *et al* 1981; Scatena *et al*. 2007). The refinement of this alert identified two further chemicals within the data set with the potential to induce toxicity towards mitochondria: lawsone and HC Green No. 1. As with the chemicals containing the anthraquinone moiety both lawsone and HC Green No. 1 are used within semi-permanent hair dye products. In addition, these two chemicals were seen to induce multiple adverse effects, such as decreased erythrocyte count, increased hypokalemia and increase triglycerides (Appendix III).

Category 3: Aromatic azo structural alert

The final structural alert identified from the similarity analysis related to chemicals containing an aromatic azo moiety and identified six chemicals from the dataset. Four of these six chemicals are used solely within semi-permanent hair dye products, a further one chemical (Disperse Red 17) is used in both semi-permanent and permanent hair dye products, whilst the sixth (Basic Brown 16) is solely used in permanent hair dye products. In comparison to the previous two categories, the number of adverse effects is reduced. In addition, the adverse effects exhibited by these chemicals are primarily observed in the circulatory system, with effects including, but not limited to decreased haemoglobin, increased blood phosphorus and decreased haematocrit (Appendix III). Chemicals containing an aromatic azo linkage are readily reduced to the free amine by the enzyme azoreductase (Nam and Renganathan 2000). The presence of an additional nitro, amine or hydroxyl group in the 2- or 4-position on at least one of the aromatic rings could result in the

possibility of the production of a 1,2- or 1,4-diaminobenzene moiety (Figure 5.8). This moiety might then act as an electron cycling agent resulting in the disruption of the respiratory chain in the mitochondria, as outlined previously for the 2-nitroaminobenzene and 4-nitroaminobenzene clusters.

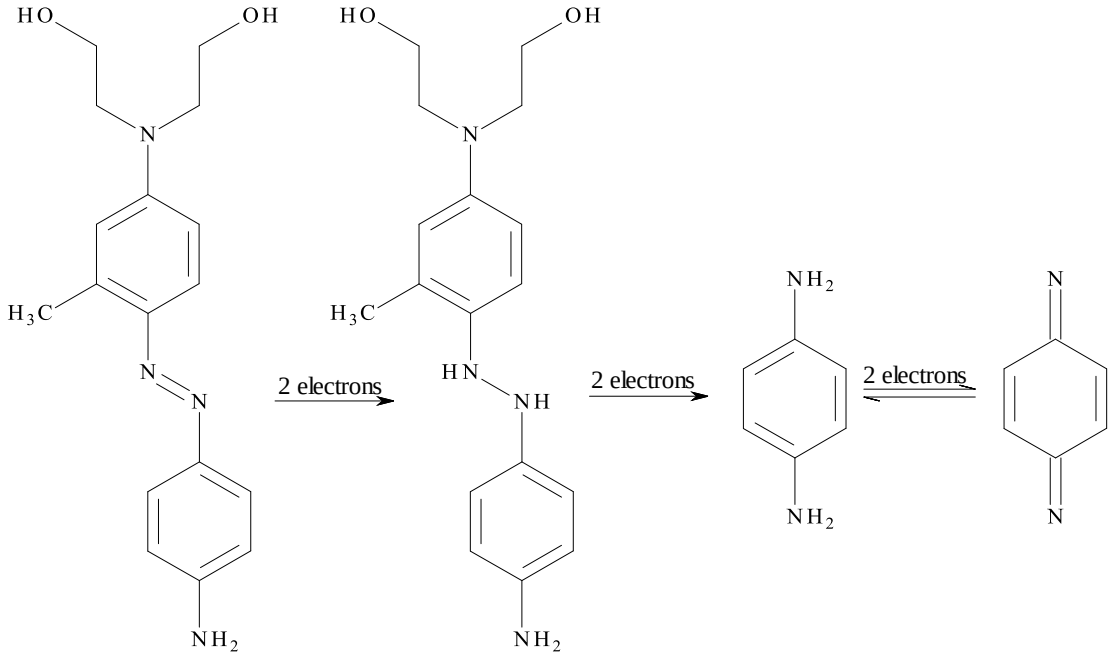


Figure 5.8: Reduction of aromatic azo compounds producing a 1,4-diaminobenzene and then subsequent oxidation to 1,4-phenylenediamine capable of cycling electrons (using the hair dye HC Yellow N$^{o.}$ 7 as an example)

5.3.2 Additional chemicals capable of electron cycling

Category 4: *Meta*-substituted benzene alert

The mechanistic chemistry outlined for the four structural alerts identified from the similarity analysis suggests that the ability to cycle electrons might represent a key MIE for mitochondrial toxicity for aromatic chemicals of this type. The mechanistic analysis further suggests that chemicals capable of forming free radicals could trigger this type of MIE resulting in toxicity. Therefore, it was possible to develop an additional structural alert based

around this mechanistic chemistry to increase the applicable chemical space relating to the MIE with respect to electron cycling. The additional alert in question relates to the *meta*-substituted benzene alert present in Table 5.2. This alert identified four chemicals as being contained in this category (Appendix III). Each of these four chemicals are used in both semi-permanent and permanent hair dye products. As with the previous categories the chemicals in the *meta*-substituted benzene category have been seen to induce a variety of adverse effects, such as an increase in centrilobular hepatotrophy, increase in kidney and liver necrosis, increase in bilirubin, and an increase in kidney and liver degeneration (Appendix III). Due to the 1,3-alignment of the substituents these *meta*-substituted chemicals are not able to form a quinone-type species. Therefore, the mechanistic chemistry is somewhat different to the structural alerts discussed above. However, it has been reported that these chemicals are capable of causing toxicity via a free radical mechanism (Aptula *et al*. 2009) (Figure 5.9). Thus, the inclusion of this additional alert can be justified based on the hypothetical mechanistic rationale that a key MIE for mitochondrial toxicity could be electron cycling due to free radical formation.

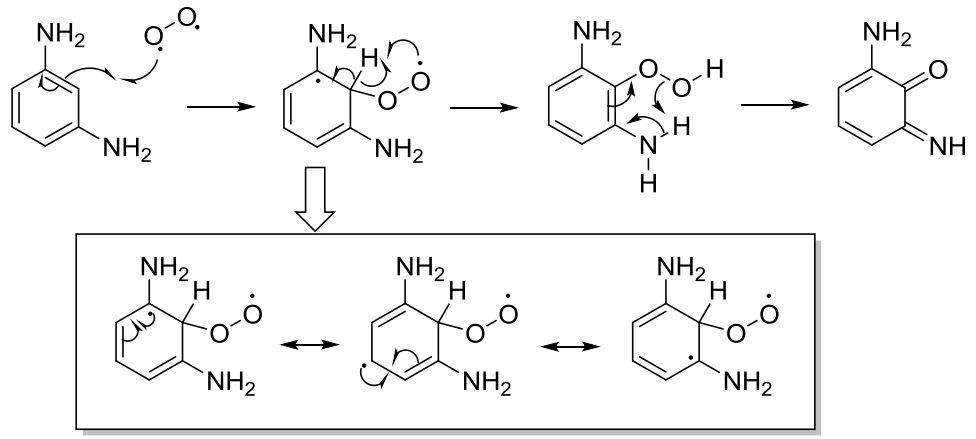
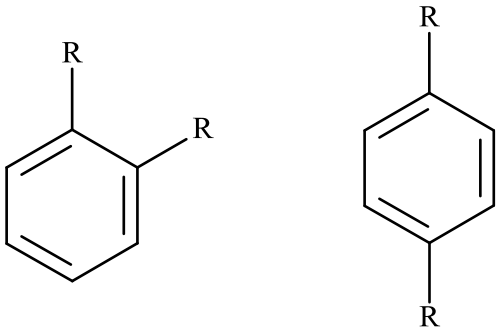
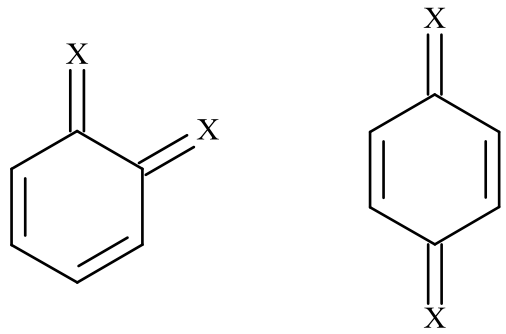
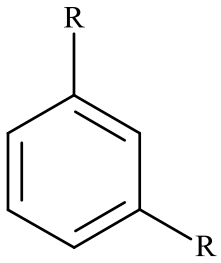
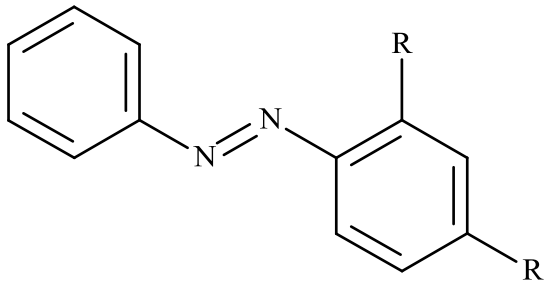


Figure 5.9: Proposed oxidation of 1,3-diaminobenzene resulting in the production of free radical species capable of inducing mitochondrial dysfunction (an analogous mechanism is possible for the 1,3-dihydroxybenzene and 3-hydroxyaminobenzene containing chemicals)

Table 5.2: Refined set of structural alerts capable of free radical cycling chemistry (NOAEL values relate to 90-day oral rat studies)

Name	Key structural features	Number of chemicals	oral NOAEL ranges (mg/kg/day)	oral LOAEL ranges (mg/kg/day)	Figure
Pro-quinones (R = OH, NH_2, NH, NO_2)		37	1.4 – 250.0	4.2 - 800	5.4

Quinones (X = NH,O)		7	2.0 – 200.0	7 - 940	5.7
Meta-substituted benzenes (R = NH_2, OH)		4	50.0 – 100.0	33 - 316	5.9

Aromatic azo (R = NH_2, NH, OH)		8	0.3 – 52.6	12 - 99	5.8

5.3.3 Mitochondria and repeat dose toxicity

The hypothetical mechanistic analysis presented above suggests that chemicals capable of free radical chemistry might disrupt the respiratory chain in the mitochondria leading to chronic toxicity. This is in keeping with previous research into the cardiotoxicity of anthracyclines upon extended low dose exposure (Montaigne *et al*. 2012). This adverse reaction has been shown to be related to mitochondrial dysfunction which results in the activation of a number of protein kinases. The MIE for this toxicity has been suggested to involve the ability of the quinone moiety within these drugs to form a semi-quinone radical and thus cycle electrons (Figure 5.7). In addition, these chemicals have been shown to form a variety of reactive oxygen species - such as hydroxyl and superoxide anion radicals - also capable of disrupting the normal function of mitochondria. These reactive oxygen species can subsequently induce damage to (mitochondrial) DNA, proteins, and lipids and may initiate membrane permeability transduction (Kappus 1986, Ohkuma *et al* 2001, Kim *et al* 2003). It has also been suggested that mitochondrial dysfunction is a key driver in chronic toxicity (Kovacic 2001a; Kovacic 2001b; Porceddu *et al*. 2012; Vinken *et al*. 2013a). A recent study also outlined how for the same chemical the mechanism driving toxicity can change on-going from acute to chronic exposure (Nikam *et al*. 2013). The importance of mitochondrial dysfunction as a driver of chronic toxicity has recently also led to the definition of a number of structural alerts, one of which (2-aminonitrophenol) was included in the current chapter (Naven *et al*. 2013).

Detailed analysis of the repeat dose data highlights that a variety of adverse effects within multiple organs are associated with the LOAEL values for chemicals assigned to each category (available in the Appendix III). This variability in the toxicity profile adds weight to the hypothesis that the observed toxicity might have been initiated by mitochondrial dysfunction. This is due to the fact that mitochondria are present within most organ systems, performing a number of roles vital to normal cellular functioning. There is an extensive body

of literature outlining a range of chemicals that inhibit the mitochondrial physiology resulting in toxicity at the organ level (Dykens 2008b). Typically, the most susceptible organs are those containing a higher concentration of mitochondria, those exposed to a higher concentration of the compound and/or those with a higher aerobic energy demand, such as the liver, kidney and cardiac muscle (Amacher 2005; Dykens 2007a; Dykens 2008b). In addition, it has been recognised by the pharmaceutical industry that mitochondrial dysfunction may be a cause of numerous toxicities within a variety of organs, and has led to the withdrawal of a number of therapeutic drugs (Amacher 2005; Dykens 2007a; Dykens 2008b; Pessayre *et al*. 2012).

5.3.4 Adverse Outcome Pathway concept, perspectives and proposed future work

The analysis presented above outlines how structural alerts related to potential MIEs could be derived. The main focus of this type of analysis is the development of the mechanistic chemistry relating the structural alerts to a possible MIE. This focus is a process that involves an in-depth survey of relevant scientific literature in support of the mechanistic hypothesis made, enabling *in silico* profilers to be developed for a given MIE. The current chapter has resulted in the development of a profiler capable of identifying chemicals that could cycle electrons and, thus, potentially lead to the disruption of the respiratory chain in the mitochondria. An important aspect of the on-going development of *in silico* profilers is the experimental verification of the mechanistic hypothesis, which increases confidence in the prediction of an MIE for an untested chemical. Such analysis has been recently undertaken for the *in silico* profilers relating to covalent protein binding in the OECD QSAR Toolbox (Enoch *et al*. 2012a; Enoch *et al*. 2013a; Nelms *et al*. 2013; Rodriguez-Sanchez *et al*. 2013). In terms of the current chapter, future work would consist of testing of a representative number of hair dyes/chemicals from each of the categories outlined to cause mitochondrial toxicity in an *in vitro* experimental set-up. In the longer term, the applicability

domain of the *in silico* profiler could then be much better defined through the use of directed and intelligent testing of compounds using assays appropriately defined by the key events of the AOP. Work undertaken in Chapter 6 demonstrates how *in vitro* and *in chemico* assay results can be utilised to verify, and refine, structural alerts.

To be able to predict oral repeat dose toxicity reliably, it is necessary, in addition to defining the applicability domain of the *in silico* profiler and by extension the MIE associated with the profiler, to generate extensive knowledge of subsequent key events in the AOP leading to toxicity. This requirement is highlighted by the broad range of oral NOAEL values for the categories derived in the current chapter which vary between one and two orders of magnitude (Table 5.2 for the ranges, Appendix III for each chemical within each category). Importantly, these values show the limitations of the *in silico* profilers ability to predict oral repeat dose toxicity. Assuming no additional information is available, the most realistic prediction for an untested chemical, assigned to one of the categories, would be to state that the oral NOAEL value would be likely to fall within the range of the values for the other category members, i.e. perform read-across. However, even this type of prediction may not be appropriate, in that the new untested chemical could be capable of altering a downstream key event in the AOP in a different manner to the remaining category members. It is also possible that the chemical may have a different toxicokinetic and/or dynamic profile to the other category members. It is therefore essential that the mechanistic information relating to the MIE contained within an *in silico* profiler is complimented with information derived from other existing *in vivo* data, *in vitro*, *in silico* or *in chemico* tests designed to target other key events in the AOP (and relating to toxicokinetics and dynamics). Only when a significant proportion of this information is available will the estimation of values such as NOAELs become possible without using animal models.

5.4 Conclusions

This chapter proposes an *in silico* profiler for chemicals used as hair dyes capable of causing mitochondrial dysfunction. It is based on a retrospective analysis of oral repeat dose toxicity data for 94 hair dye chemicals and is intended for use in grouping and category formation. It is important to note that the proposed profiler does not predict oral repeat dose toxicity; instead it provides arguments for a key molecular initiating event that might be responsible for initiating an adverse outcome pathway leading to chronic toxicity. In order to be more widely applicable for mitochondrial toxicity the structural alerts developed within this chapter will need to be combined with additional alerts; both for electron cycling and for other molecular initiating events leading to mitochondrial dysfunction, for example those discussed in Chapter 4. However, at present this *in silico* profiler can still be useful for identifying novel chemicals, containing the moieties identified herein, with the ability to induce mitochondrial toxicity via inhibition of the electron transport chain due to electron cycling (Chapter 1). Thereby, enabling the user to identify chemicals within a data set that should be prioritised to undergo testing within *in vitro* or *in chemico* assays. This work generally shows that detailed mechanistic analysis is required for the development of *in silico* profilers and explains how such analysis can be used to identify potential molecular initiating events. Clearly future *in vitro* and/or *in chemico* work must be undertaken to outline additional key events in the biological pathway before a relevant and complete adverse outcome pathway could be established (discussed in Chapter 7).

Chapter 6: Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitrosos, nitros, aldehydes, and ketones*

6.1 Introduction

An *in silico* profiler consists of a series of chemical fragments, derived from knowledge of mechanistic organic chemistry, known as structural alerts. These structural alerts can be used to group chemicals into categories based upon the knowledge of a well-defined molecular initiating event (Enoch *et al* 2008b, Enoch *et al* 2011a). The structural alerts developed in the previous chapters (Chapters 3, 4, and 5) focussed on the ability to induce mitochondrial toxicity via different molecular initiating events. This mechanistic approach to grouping allows for interpretable predictions to be made for novel chemicals for which toxicological data are absent. It is therefore important that the structural alerts within profilers are well defined in terms of the chemical space in which they can be applied and can make reliable predictions (also known as the applicability domain). A key advantage of developing structural alerts with a well-defined chemical space is that there is less likelihood that a chemical will be incorrectly assigned to a category. That is to say, the profiler will be less likely to assign non-toxic chemicals as being toxic or assigning chemicals acting via different mechanisms to the same category.

Previous research has shown that the mechanistic domains pertaining to protein binding described below can be encoded computationally into *in silico* profilers (Enoch 2010, Enoch and Cronin 2010, Enoch *et al* 2011a, Enoch *et al* 2011b). One such computational program is the Organisation for Economic Co-operation and Development (OECD) QSAR Toolbox. The OECD QSAR Toolbox contains a number of features including: a metabolic simulator, which enables predictions to be made regarding a chemicals potential metabolites in a variety of test systems; a database containing a vast quantity of acute and repeat dose toxicity data, provided by several commercial and governmental sources such as the

*This chapter is based on an a published article by Nelms *et al* (2013)

European Chemicals Agency, the US Environmental Protection Agency, and Fraunhofer Institute. Additionally, the QSAR Toolbox provides a method to group chemicals together, into categories, based upon structural and/or mechanistic similarities, and enables structure-activity relationships, such as read-across, to be used to fill data gaps for chemical hazard assessment. Furthermore, the QSAR Toolbox contains a number of *in silico* profilers covering a variety of (eco)toxicological endpoints, such as bioaccumulation, carcinogenicity, DNA binding, whilst also including two profilers for protein binding; the Optimised Approach based on Structural Indices Set (OASIS) and the OECD profilers.

Traditionally, a chemical's toxicological profile has been assessed via *in vivo* testing. However, since the introduction of chemical legislation, such as the European Union's REACH (Registration, Authorisation, restriction and Evaluation of Chemicals) (EC 2006a, EC 2006b) and the 7th amendment to the Cosmetic Directive (EC 2003), a number of alternative techniques have been promoted (Schultz 2010, Adler 2011) (discussed in more detail in Chapter 1). It is envisaged that these alternatives will generate information that can be compiled, and used as part of a weight of evidence, in order to aid hazard and risk assessments to be performed for chemicals lacking *in vivo* data.

One of the simplest *in silico* techniques that can be applied to the toxicological assessment of chemicals is the creation of chemical categories (ECHA 2008, OECD 2011). Chemical categories can be developed based upon similarity across a group of chemicals for a variety of properties, such as structural features, physico-chemical properties or mechanism of action. These categories can thus be utilised to make predictions based upon the premise that similar chemicals should have similar chemical and biological activities (ECHA 2008, ECHA 2012). Thus, data for tested chemicals in a particular category can be used to fill data gaps for untested chemicals (as discussed in Chapter 1). One of the better ways to assess chemical similarity is to apply the principles of mechanistic chemistry to group substances by their ability to undergo a common reaction (Schwöbel *et al* 2011). Of particular relevance to category formation are electrophilic reactions, whereby a covalent bond(s) is formed

between an electron-poor non-endogenous chemical (an electrophile) and an electron-rich biological target (a nucleophile) (Schwöbel *et al* 2011). An electrophile is a chemical that, during chemical reactions, is attracted to an electron-rich centre, acting as a Lewis acid in so much as an electron pair is accepted from a nucleophile in order to form a covalent bond. In contrast, a nucleophile is a chemical attracted to an electron-deficient centre that donates an electron pair within a covalent bond, i.e. a Lewis base.

It is well known that there are a range of electrophilic reactions that target a variety of biological nucleophiles, such as proteins, lipids or other related electron-rich molecules (Schwöbel *et al* 2011). It is, therefore, important to consider the selectivity of an electrophile toward a specific nucleophile. This can be explained by the classification of electrophiles and nucleophiles according to their chemical "hardness" and "softness". Briefly stated, hard electrophiles bind preferentially to hard nucleophiles while soft electrophiles bind preferentially to soft nucleophiles. This concept of "like-reacts-with-like", whilst not absolute, reflects the fact that dissimilarity in electrophilic hardness/softness results in a higher potential energy barrier for the electrophilic reaction and subsequently lower chemical reactivity.

Hard electrophiles are molecules with low polarisability having their electron deficiency localised as a positive (or partially positive) electrostatic charge. In contrast, soft electrophiles are molecules with high polarisability having their electron deficiency spread over a larger area of the molecule (resulting in a negative or partially negative charge). In this context polarisability can be defined as the amount by which the electron cloud, of an electrophile, can be deformed in the presence of a nucleophilic anion (or lone pair of electrons), with the electron cloud of soft electrophiles deforming easier, and to a greater extent, than hard electrophiles. Nucleophilic targets in biological molecules typically include electron-rich heteroatoms such as sulphur, nitrogen, and oxygen. Sulphur is at the soft end and oxygen is at the hard end of the nucleophilic range with nitrogen being intermediate. This means that harder electrophiles prefer to react with the in-ring oxygen and nitrogen of

nucleic acids, while softer electrophiles prefer to react with sulphur and nitrogen of amino acids in proteins (Schultz *et al* 2006). Thus, nucleophilic sites related to protein and DNA binding are, in order of increasing hardness: thiol group of cysteine (and glutathione), sulphur atoms of methionine, primary amino groups (e.g. of lysine), secondary amino groups of histidine, primary amino groups of purine bases (e.g. adenine), ring N-atoms of purine and pyrimidine bases, O-atoms of purine and pyrimidine bases, and O-atoms of phosphate (Schwöbel *et al* 2011). Figure 6.1 shows the biological nucleophiles identified above in order of increasing hardness.

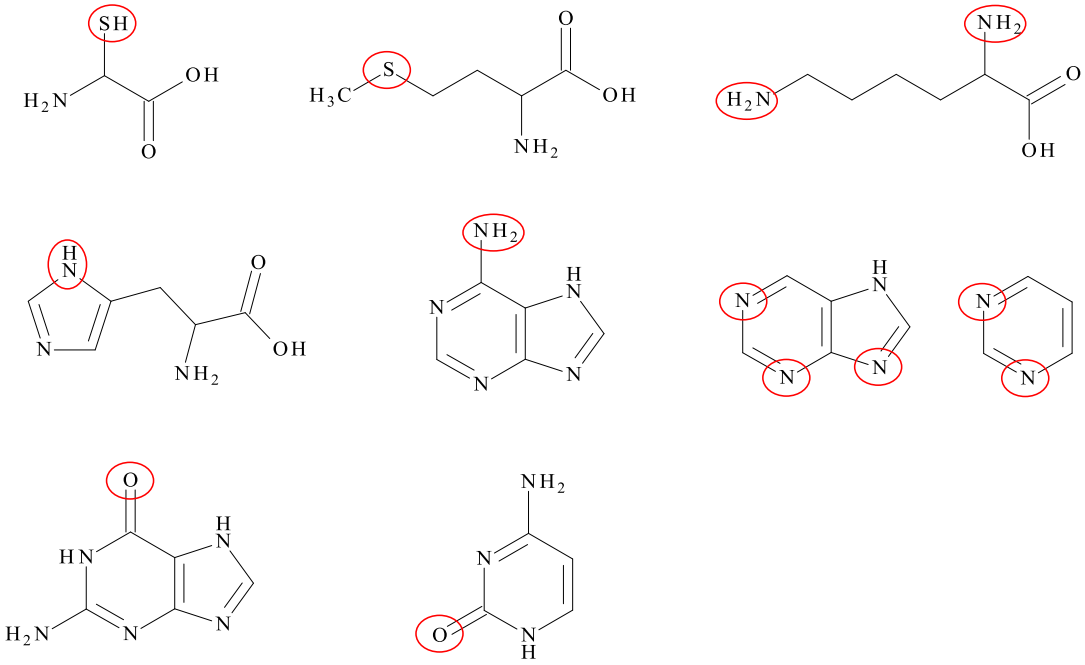


Figure 6.1: Nucleophilic sites related to protein and DNA binding (circled in red), in order of increasing hardness from left to right and top to bottom.

Earlier investigations into the use of mechanistic chemistry to define the reactions that occur when a non-endogenous chemical electrophile covalently binds with a biological macromolecule outlined six mechanistic domains (Aptula and Roberts 2006, Enoch *et al* 2011b). These mechanistic domains are; Michael addition, aromatic nucleophilic

substitution (S_NAr), unimolecular aliphatic nucleophilic substitution (S_N1), bimolecular aliphatic nucleophilic substitution (S_N2), Schiff base formation, and acylation.

Michael addition reactions occur between a biological nucleophile and an alkene or alkyne polarised by an electron withdrawing group (α,β-unsaturated alkenes and alkynes). The presence of the electron withdrawing group is key to the reactivity as it draws electron density away from the β-carbon, therefore, making it more positively charged than it would otherwise be within an unpolarised alkene or alkyne (Figure 6.2). In addition, the electron withdrawing group also stabilises the negative charge that develops on the α-carbon in the intermediate, with more electronegative groups increasing reactivity (Aptula and Roberts 2006). Finally, substituents surrounding the β-carbon reactive centre have been shown to be influential on the rate of reactivity, with chemicals highly substituted being less reactive than unsubstituted chemicals (Koleva *et al* 2008, Schwöbel *et al* 2010a, Schwöbel *et al* 2010b, Enoch and Roberts 2013b). This can be rationalised in terms of the accessibility of the β-carbon, with more highly substituted carbons being less accessible to attack by the nucleophile.

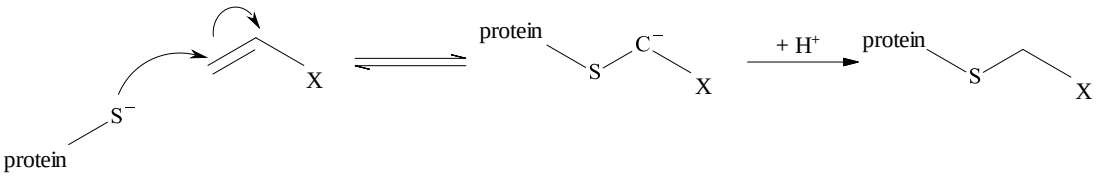


Figure 6.2: Scheme showing the Michael addition reaction (X = polarising group e.g. aldehyde, ketone, ester or amide)

Both unimolecular and bimolecular nucleophilic substitution reactions (S_N1 and S_N2 respectively) involve an aliphatic carbon, nitrogen, sulphur or halogen atom, with an electronegative leaving group attached, being attacked by a biological nucleophile. Unimolecular nucleophilic substitution reactions occur in two steps, the first step involves the formation of a cationic intermediate. The second step involves the cationic intermediate

being attacked by the biological nucleophile (Figure 6.3). In contrast, S_N2 reactions occur in one step via a transition state (Figure 6.4). Typically, these two mechanisms are in competition with one another, with the preferred mechanism for a given chemical being based upon two key factors: the steric hindrance at the reactive centre and the amount of stabilisation provided to the cationic intermediate by the surrounding substituents. For example, tertiary halides react via an S_N1 mechanism as the bulky substituents surrounding the reactive centre act to prevent any back-side attack by the nucleophile due to their steric bulk. In addition, the presence of the alkyl groups helps to stabilise the cationic intermediate due to the inductive effect. In contrast, primary and secondary halides react via an S_N2 mechanism, as the reactive centre is less sterically hindered coupled with decreased stabilisation of the (potential) cationic intermediate due to the lower inductive effect from the substituents.

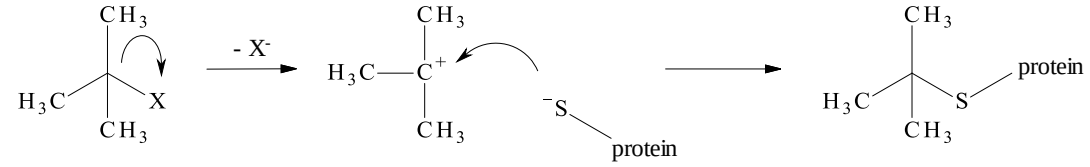


Figure 6.3: Scheme showing the S_N1 reaction (X = leaving group e.g. halogen)

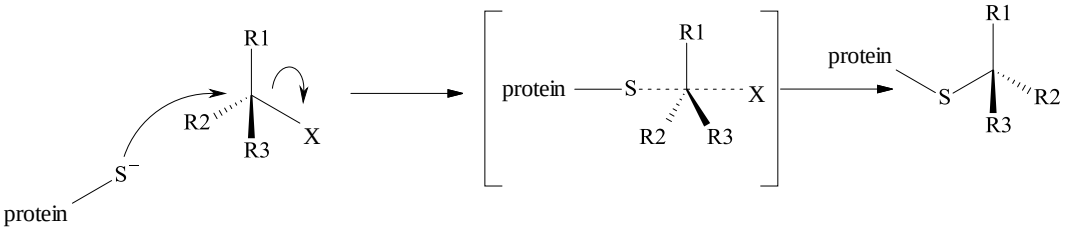


Figure 6.4: Scheme showing the S_N2 reaction (X = leaving group e.g. halogen)

Substituted aromatic chemicals can react with biological nucleophiles via nucleophilic aromatic substitution (S_NAr). As with the previously discussed S_N2 mechanism, the S_NAr reaction involves an initial nucleophilic attack producing a resonance stabilised negatively charged transition state. This is followed by the elimination of the leaving group to produce

the substituted adduct (Figure 6.5). In order for an S_NAr reaction to be viable, at least two electron withdrawing groups are required *ortho* and *para* to the leaving group, whether as substituents or present within the benzene ring, meta-substituted aryl halides have been shown to be non-reactive via S_NAr (Aptula and Roberts 2006, Enoch *et al* 2011b). The rate of reaction for the S_NAr mechanism is dependent upon the number of electron withdrawing groups present, i.e. the more electron withdrawing groups that are substituted on (or within) the benzene ring the more reactive the chemical (Figure 6.6), this is due to the added stability that is conferred onto the carbanion intermediate by the electron withdrawing groups (Enoch *et al* 2012a).

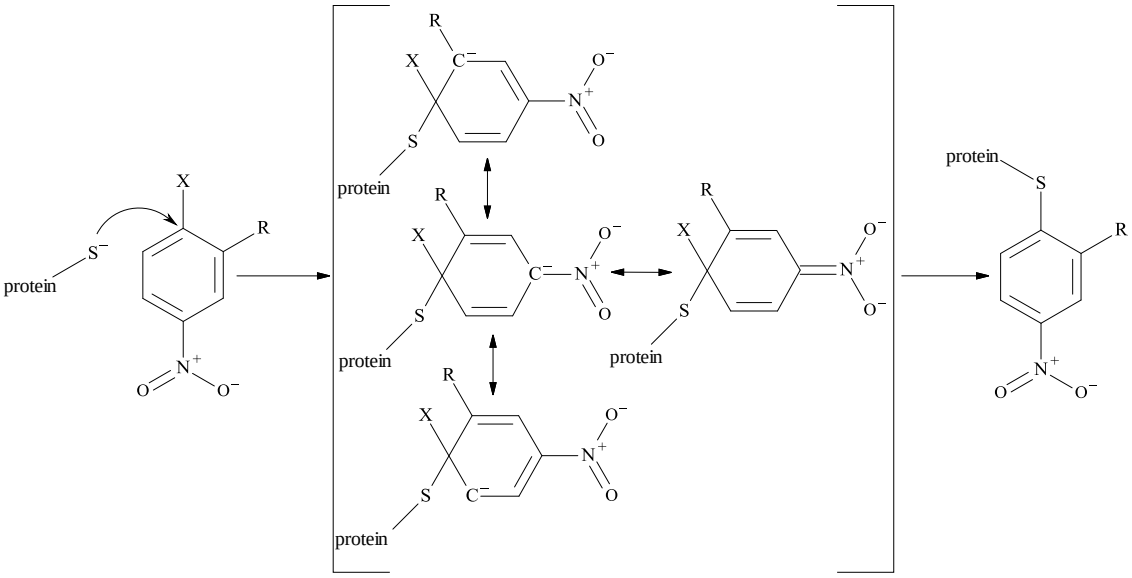


Figure 6.5: Scheme showing the S_NAr reaction, using a nitro group as an electron withdrawing group (X = leaving group e.g. halogen, R = suitable electron withdrawing group e.g. nitro group)

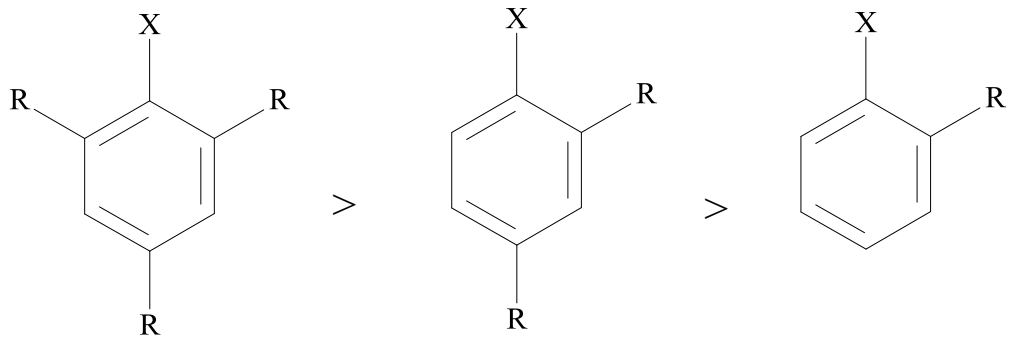


Figure 6.6: The effect multiple electron withdrawing groups can have on protein binding affinity. Additional withdrawing groups increase toxicity due to added stability conferred onto the carbanion intermediates by electron withdrawing groups (X = leaving group e.g. halogen, R = electron withdrawing group e.g. nitro group)

Schiff base formation occurs when a primary amine of a lysine acts as a nucleophile by attacking the electrophilic carbon of the carbonyl moiety in an aldehyde or ketone (Figure 6.7). The first stage of Schiff base formation involves the lone pair of electrons on the amine reacting with the carbonyl carbon to form a tetrahedral hemiaminal intermediate. Upon protonation of the hydroxyl group, the lone pair of electrons on the amine nitrogen atom forms a double bond with the neighbouring carbon, expelling the newly formed water molecule. The final stage is the deprotonation of the iminium ion and production of the final Schiff base product, or imine. The rate of Schiff base formation is understood to increase as the side chain of the aldehyde or ketone becomes more electronegative. In contrast, a decrease in Schiff base formation has been seen for aldehydes directly bound to a benzene ring. It is thought this is due to the additional activation energy required to overcome the resonance stabilisation conferred to the aldehyde moiety by the benzene ring (Patlewicz *et al* 2001). Additionally, increasing the steric bulk of substituents surrounding the carbonyl moiety decreases reactivity by hindering the ability of the nucleophile to attack the carbonyl carbon.

As α,β-unsaturated aldehydes have the potential to react via both Schiff base formation and Michael addition it can be problematic to distinguish which reaction mechanism is more likely to occur. However, the likelihood of whether a Schiff base or Michael addition reaction occurs is dependent upon: the saturation of the aldehyde, the accessibility of the β-carbon, and the nucleophile present. For example, if the aldehyde is fully saturated, or the β-carbon is sterically hindered by bulky substituents groups, and the nucleophile is an amine a Schiff base reaction will be favoured. In contrast, if the aldehyde is unsaturated, with little or no steric bulk around the β-carbon, and the nucleophile is a sulphur group a Michael addition reaction will be favoured.

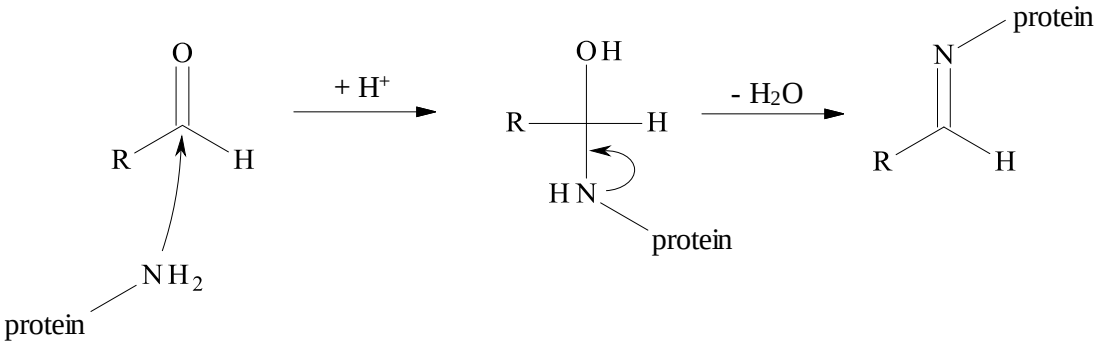


Figure 6.7: Scheme showing the Schiff base reaction

Acylation reactions are similar to Schiff base formation in that the nucleophile attacks the carbonyl carbon moiety. However, in an acylation reaction the carbonyl carbon atom is attached to an electronegative group that, during the course of the reaction, acts as a leaving group. As can be seen in Figure 6.8 the nucleophile attacks the carbonyl group resulting in four coordinate, tetrahedral, transition state bonds. There are three main factors that affect the rate of an acylation reaction. The first is the electronegativity of the leaving group, with more electronegative groups inducing a quicker reaction. This results in both an increase in the partial positive charge present on the carbonyl moiety, as the additional electron withdrawing group further polarises the carbon atom, and the increase in electronegativity of

the leaving group allows it to more readily accept the negative charge as it is expelled. The second factor is the amount of steric bulk surrounding the carbonyl moiety, an increase in the bulk of substituents surrounding the carbonyl moiety decreases the rate of reaction by reducing the ability of the nucleophile to attack the carbonyl carbon. Finally, as the resonance between the carbonyl group and the electronegative leaving group increases the reactivity will decrease, due to the electron density being more greatly dispersed decreasing the δ^+ charge present on the carbonyl carbon.

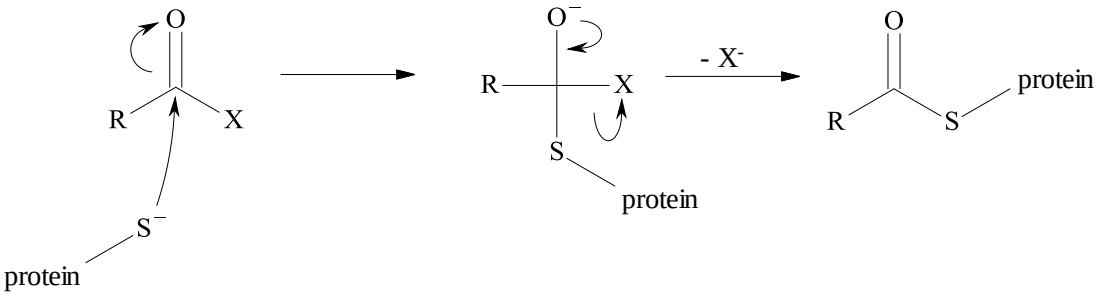


Figure 6.8: Scheme showing the acylation reaction (X = leaving group e.g. halogen)

Whilst these principles are important for all structural alerts (i.e. not just the protein binding alerts described above), the aim of this chapter was to highlight the importance of using experimental data obtained from *in chemico* and *in vitro* assays to verify the structural alerts within the OASIS and OECD *in silico* profilers for protein binding in the OECD QSAR Toolbox. Additionally, the data obtained from these assays were utilised to identify chemical space where new structural alerts were needed, or existing alerts needed to be refined. The chemical space of the structural alerts for seven chemical classes (epoxides, lactones, nitrosos, nitros, aldehydes, ketones and ring-strained hydrocarbons) was investigated using data from an *in chemico* glutathione (GSH) reactivity assay and an *in vitro* growth inhibition assay.

6.2 Method

6.2.1 Data Set

Thirty three chemicals containing an aldehyde, epoxide, ketone, lactone, nitroso, nitros or a strained hydrocarbon ring moiety were evaluated within two separate assays: an *in chemico* assay used to measure the reactivity of a chemical towards the thiol group present within glutathione, and an *in vitro* assay that measures the concentration of a chemical required to inhibit the growth of *Tetrahymena pyriformis* by 50%. The *in vitro* assay can be used to identify chemicals that react with proteins via each of the six mechanisms described above; this is due to the presence of multiple protein types within *Tetrahymena pyriformis*. In contrast, the *in chemico* assay, as it only contains one (tri)peptide (glutathione), can identify chemicals that act via all except the Schiff base reaction mechanism; this is due to the absence of a lysine moiety within glutathione that is required for Schiff base formation to occur. All tested chemicals were purchased from commercial sources (SigmaAldrich.com or Alfa.com) in the highest purity available (95% minimum) and were not further purified prior to testing. Both the *in chemico* and *in vitro* assays were performed by colleagues from the University of Tennessee: the data generated were obtained by following the protocols detailed below.

6.2.2 In chemico glutathione reactivity

Reactivity with the thiol group of glutathione (GSH) was measured in a simple and rapid spectrophotometric-based assay with the free thiol quantified by its reaction with 5,5'-dithio-bis(2-nitrobenzoic acid) (DTNB) with the absorption of the product measured at 412 nm (Schultz 2005, Enoch 2012a). Briefly, experiments were performed by a) freshly preparing GSH at 1.375 mM by dissolving 0.042 g of reduced GSH into 100 ml of phosphate buffer at pH 7.4; b) freshly preparing stock solutions of each tested chemical by dissolving the test chemical in dimethyl sulphoxide (DMSO) to which phosphate buffer was subsequently

added; and c) combining the correct amounts of GSH solution, stock solution, and buffer to bring the final concentration of thiol to 0.1375 mM, in a manner so the concentration of DMSO in the final solution was always <10%.

Following range-finding experiments, subsequent experiments were performed with concentrations adjusted to 90, 80, 60, 40, 20 and 10% of the stock solution. Associated with each assay was a control containing GSH and a blank without GSH. The RC_{50} values (the concentration giving 50% reaction in a fixed time of 2h) were determined from nominal chemical concentrations (dependent variable) and absorbance normalised to the control (independent variable) using Probit Analysis of the Statistical Analysis System (SAS) software (SAS Institute, Cary, NC).

Chemicals with a RC_{50} value of greater than 135 mM were considered non-reactive. Additional reactivity testing was performed on selected epoxides that were not reactive in the more polar environment, as described above. In these cases, the reactivity assessment was performed in a medium of 50% methanol and 50% buffer with all other parameters being the same as described above. This modification increased the solubility of test substances without altering inherent reactivity. The comparability between reactivity measurements made in aqueous solution and methanol has been demonstrated previously (Enoch *et al* 2012a).

6.2.3 Aquatic toxicity data

The protocol described by Schultz (Schultz 1997) was utilised by colleagues at the University of Tennessee to obtain the 50% inhibitory growth concentration (IGC_{50}) of the common ciliate *Tetrahymena pyriformis* for chemicals shown in Table 6.1. This assay was conducted over a 40 h period with the population density of *Tetrahymena pyriformis* cells being spectrophotometrically quantified at 540 nm as the test endpoint. Two controls, the first control containing no test material and *Tetrahymena pyriformis* and the second

containing neither test material nor *Tetrahymena pyriformis*, were used in order to indicate the suitability of the medium and also to help interpret the results produced under test conditions. Each chemical was tested in triplicate; these triplicates consisted of a minimum of five different concentrations of the test chemical. Following the 40h incubation and the spectrophotometric quantification of the population density of *Tetrahymena pyriformis* for each test condition, the IGC_{50} value was calculated (in millimolar units) by absorbance normalised to controls (independent variable) and the nominal concentration of the toxicant (dependent variable) using the Probit Analysis in SAS (Statistical Analysis Software) software (SAS Institute, Cary, NC).

6.2.4 In silico predictions

The chemical structure of each chemical was checked against various identifiers, i.e. chemical name, CAS and/or other identifiers, before the chemicals were profiled using the OASIS and OECD protein binding profilers individually. The chemicals were profiled using the OASIS and OECD protein binding profilers in version 3.0 of the OECD QSAR Toolbox (available from www.qsartoolbox.org). If a structural alert was triggered for a chemical in one, or both, of the profilers then a chemical was deemed to have the ability to covalently bind to proteins. For chemicals where an alert for protein binding was identified, the electrophilic mechanism was recorded. As neither of the two assays described above consider metabolism only chemicals with the potential to directly act as electrophiles were investigated, meaning that structural alerts relating to metabolism were not investigated.

6.2.5 Verification of alerts

The *in chemico* and *in vitro* assays were subsequently used to verify the *in silico* predictions. A structural alert in a profiler was deemed to be correct if it was in agreement with one, or both, of the assay results, i.e. reactive to GSH and/or demonstrating toxicity to *Tetrahymena* that was significantly different from baseline. Those chemicals that exhibited toxicity significantly different from baseline demonstrate that factors other than hydrophobicity are driving toxicity, i.e. an electrophilic reaction. The calculated narcosis baseline was developed by Ellison *et al* (2008) from a linear regression analysis of toxicity data, within the *Tetrahymena pyriformis* assay, for 87 saturated alcohols and ketones contained within a larger dataset of 517 chemicals covering a variety of chemical classes, including, but not limited to, amides, esters, haloalkanes and sulphides. The results of this linear regression analysis were used to develop Equation (1). In this chapter a chemical was judged to exhibit toxicity significantly different from baseline, within the scatterplot of log $1/IGC_{50}$ data against log P (Figure 6.9) produced using Minitab 16.2.2, if there was a ≥1 log unit deviation from the calculated baseline toxicity model developed by Ellison *et al* (2008). Within this scatterplot log P was calculated using the KOWWIN (v1.68) application available in the EPISuite software (freely available at www.epa.gov)

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

$$n = 87,\ r^2 = 0.96,\ s = 0.20,\ F = 2131$$

Where, *n* is the number of observations; r^2 is the square of the correlation coefficient adjusted for degrees of freedom; *s* is the SE on the estimate; *F* is Fishers statistic.

Additional analysis was undertaken to discern the reasoning behind any discrepancy, i.e. comparing the results from the *in silico* profilers and the *in chemico* and *in vitro* assays to understand how and why variances occurred.

6.3 Results and discussion

The aim of this chapter was to highlight the importance of using experimental data to verify the structural alerts within the OASIS and OECD *in silico* profilers for protein binding in the OECD QSAR Toolbox. In addition, the experimental data were used to define new, or refine existing, structural alerts contained within the profilers. Seven classes of chemicals with various functional groups were studied: epoxides, lactones, nitrosos, nitros, aldehydes, ketones and ring-strained hydrocarbons. Despite the fact that the two profilers were developed separately there is still a high degree of overlap between them due to the underlying data from which they were developed (Patlewicz *et al* 2007, Enoch *et al* 2011b). The reactivity information was supplemented with that from the presence or absence of excess toxicity (in this instance an indication of cellular protein binding) in the *Tetrahymena pyriformis* growth inhibition assay. The *in silico*, *in chemico* and *in vitro* data are summarised in Table 6.1. These data showed, as is consistent with organic reaction chemistry, chemicals within the epoxide, nitroso, nitros, lactone and aldehyde and di-ketone chemical classes to be capable of a covalent reaction with proteins. No reactivity was observed in either *in vitro* or *in chemico* assay for the mono-ketone and ring-strained hydrocarbon classes; this is in agreement with the absence of a structural alert within both of the *in silico* profilers for these chemical classes.

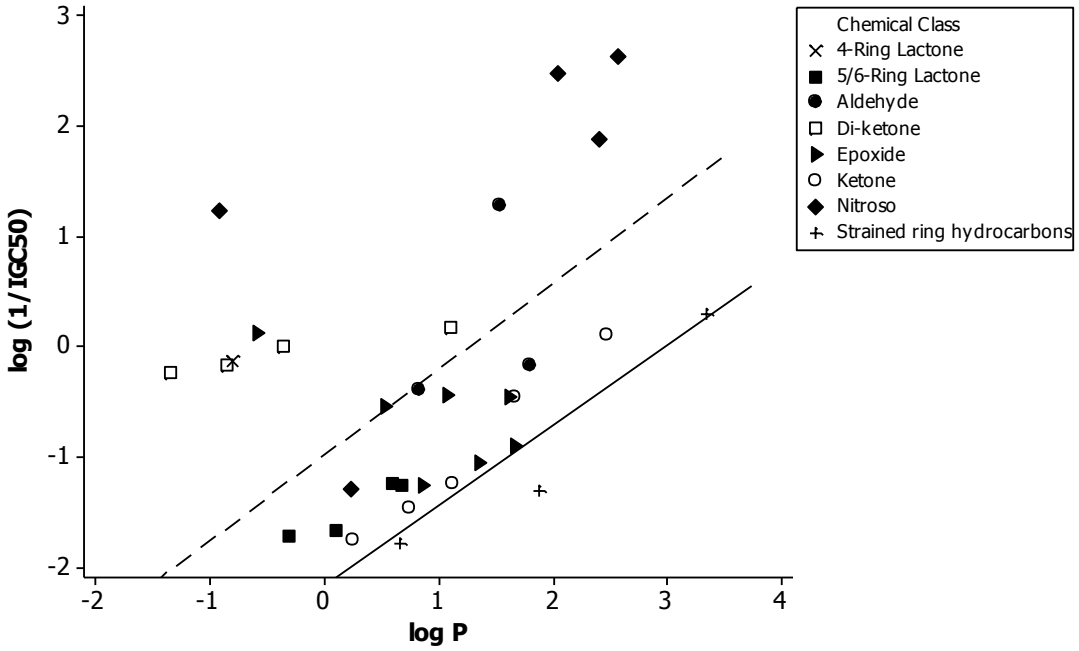


Figure 6.9: Plot of log ($1/IGC_{50}$) to *T. pyriformis* against log P for each chemical category. The solid line represents baseline toxicity as defined by Ellison *et al* (2008), whilst the dashed line represents baseline toxicity plus one log unit (i.e. excess toxicity if above the dashed line) as defined by Nendza *et al* (2007).

6.3.1 Relationship between toxicity to Tetrahymena pyriformis and hydrophobicity

The relationship of acute (cyto)toxicity (IGC_{50}) with hydrophobicity (log P) shows that as hydrophobicity increases there is an associated increase in (cyto)toxicity (Figure 6.9). The plot shows that a significant number of chemicals have excess toxicity. A verdict of excess toxicity was assigned if a chemical's toxicity was at least one log unit above the calculated baseline (Nendza *et al* 2007, Ellison *et al* 2008). Excess toxicity is an indicator, although not absolute proof, of reactivity. This is due to there being multiple mechanisms by which excess toxicity could be initiated, these being: weak acid respiratory uncoupling, precursor to soft electrophiles, precursor to redox cyclers, and soft electrophiles (Schultz *et al* 1996). Importantly, there are a number of reasons why a chemical may be intrinsically reactive yet

not exhibit excess toxicity in the *Tetrahymena pyriformis* assay. The main reason why an intrinsically reactive chemical may not exhibit excess toxicity in the *in vitro* assay relates to a chemical's hydrophobicity. Firstly, the hydrophobicity of the chemical may be such that the chemical is not soluble in the test system, i.e. the chemical is too hydrophobic to be soluble in the aquatic environment of the *Tetrahymena pyriformis* assay. Alternatively, a chemical may be reactive in the *in vitro* assay but still appear on the narcosis baseline due to the hydrophobicity of the chemical, i.e. the chemical's log ($1/IGC_{50}$) and lop P values may coincidentally intersect on the narcosis baseline. In such a case inspection of a series of related chemicals often reveals their toxicity to be dependent on reactivity rather than hydrophobicity. Therefore, this information is useful and should be considered to be part of a weight of evidence to understand the reactivity of a chemical.

Table 6.1: Summary of experimental and *in silico* data (S_N2 = bimolecular nucleophilic substitution, AC = acylation, SB = Schiff base formation, Non = no alert, NucA = nucleophilic addition NR = not-reactive, Non-cov = non covalent interaction, NSDB = not significantly different from baseline, XS = excess toxicity)

ID	Class	Chemical Name	LogP	*In silico* profiler		Log $1/RC_{50}$	*In chemico*	Log $1/IGC_{50}$	Toxicity to *T. pyriformis*
				OASIS	OECD				
1	Epoxides	1,2-Epoxybutane	0.86	S_N2	S_N2	-1.73	Reactive	-1.25	NSDB
2		1,2-Epoxypentane	1.35	S_N2	S_N2	-1.68	Reactive	-1.05	NSDB
3		1,3-Butadiene diepoxide	-0.60	S_N2	S_N2	-1.06	Reactive	0.13	XS
4		Cyclohexene oxide	1.66	S_N2	S_N2	-1.48	Reactive	-0.90	NSDB
5		Glycidyl isopropyl ether	0.53	S_N2	S_N2	-1.74	Reactive	-0.54	NSDB
6		Glycidyl n-butyl ether	1.08	S_N2	S_N2	-1.52	Reactive	-0.43	NSDB
7		Glycidyl phenyl ether	1.61	S_N2	S_N2	-1.39	Reactive	-0.46	NSDB
8	Lactones	β –Propiolactone	-0.80	AC	AC	-0.20	Reactive	-0.13	XS
9		γ-Butyrolactone	-0.30	Non	AC	NR	NR	-1.72	NSDB
10		γ-Caprolactone	0.60	Non	AC	NR	NR	-1.24	NSDB
11		γ-Valerolactone	0.11	Non	AC	NR	NR	-1.67	NSDB
12		δ-Valerolactone	0.19	AC	AC	NR	NR	NR	NSDB
13		ε-Caprolactone	0.68	AC	AC	NR	NR	-1.26	NSDB
14	Nitrosos	1-Nitrosopyrrolidine	0.23	Non	Non	NR	NR	-1.28	NSDB
15		2-Nitrosotoluene	2.41	NucA	Non	1.55	Reactive	1.88	XS
16		2-Nitroso-1-naphthol	2.56	NucA	Non	0.77	Reactive	2.62	XS
17		N,N-dimethyl-4-nitrosoaniline	2.04	NucA	Non	1.20	Reactive	2.48	XS
18	Nitros	4-Nitropyridine N-oxide	-0.90	Non	Non	-0.40	Reactive	1.23	XS
19	Aldehydes	Butyraldehyde	0.82	SB	SB	Non-cov	NR	-0.38	XS
20		Hexylaldehyde	1.80	SB	SB	Non-cov	NR	-0.17	XS
21		Phenyl acetylaldehyde	1.54	SB	SB	Non-cov	NR	1.29	XS

22	Ketones	1-Phenyl-1,2-propanedione	1.11	SB	Non	Non-cov	NR	0.18	XS
23		2,3-Butanedione	-1.30	SB	SB	Non-cov	NR	-0.23	XS
24		2,3-Pentanedione	-0.90	SB	SB	Non-cov	NR	-0.16	XS
25		3,4-Hexanedione	-0.40	SB	SB	Non-cov	NR	0.00	XS
26		2-Butanone	0.26	Non	Non	NR	NR	-1.75	NSDB
27		3-Pentanone	0.75	Non	Non	NR	NR	-1.46	NSDB
28		Acetophenone	1.67	Non	Non	NR	NR	-0.46	NSDB
29		Cyclohexanone	1.13	Non	Non	NR	NR	-1.23	NSDB
30		Cyclopropyl phenyl ketone	2.47	Non	Non	NR	NR	0.11	NSDB
31	Strained ring hydrocarbons	Cyclopropane methanol	0.66	Non	Non	NR	NR	-1.78	NSDB
32		Cyclopropyl benzene	3.34	Non	Non	NR	NR	0.29	NSDB
33		Dicyclopropyl benzene	1.87	Non	Non	NR	NR	-1.31	NSDB

Epoxides

The seven epoxide-containing chemicals were profiled by the OECD and OASIS profilers as being reactive towards proteins via an S_N2 mechanism (chemicals 1-7 in Table 6.1, mechanism as shown in Figure 6.10). Inspection of the experimental data for these seven chemicals showed them to all be reactive towards GSH, confirming the *in silico* predictions. In contrast, only 1,3-butadiene di-epoxide and glycidyl isopropyl ether showed excess toxicity when tested in the *Tetrahymena pyriformis* assay. The experimental toxicity values for the remaining epoxides are not significantly different from those predicted from equation (1). Inspection of the glutathione reactivity data for these chemicals indicates that the rate of covalent bond formation stays relatively consistent across the group even as log P increases (Table 6.1). Thus, when investigated together, the toxicity for this group of chemicals is predicted by their ability to covalently protein bind rather than the hydrophobicity of the chemical. Although the epoxides fall close to the narcosis baseline toxicity is not driven by hydrophobicity but rather by covalent bond formation, which is reasonably consistent irrespective of log P. This finding is in keeping with several previous structural alert studies that did not consistently identify aliphatic epoxides as causing excess toxicity (von der Ohe *et al* 2005, Blaschke *et al* 2010).

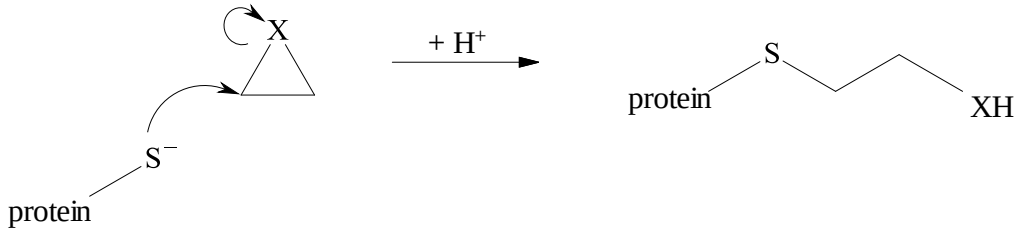


Figure 6.10: Scheme for ring opening S_N2 reaction for epoxides (X = NH, O, S)

Lactones

Lactone ring systems contain an electrophilic carbonyl centre that can undergo an acylation ring opening reaction. The level of ring strain governs the reactivity of the lactone ring, with more highly strained systems being more reactive. This is due to the release of ring strain contributing significantly to the ease of ring opening (Figure 6.11). The OECD profiler identified a potential acylation mechanism in all six chemicals, whilst the OASIS profiler identified only three chemicals capable of the same mechanism (chemicals 8-13 in Table 6.1). In contrast, the GSH reactivity data showed only the 4-membered lactone ring to be reactive (chemical 8 in Table 6.1). In addition, this chemical is the only one in this class to show excess toxicity in the *T. pyriformis* assay. The results show that under biologically relevant conditions 5- and 6-membered lactone rings are not reactive. This is because 5- and 6-membered lactones do not benefit from the release of ring strain upon ring opening; therefore, the activation energy required for acylation is greater than for 4-membered lactones (Hemminki 1981). Previous research has also shown that 4-membered ring lactones can undergo either acylation or S_N2 mechanism depending on whether the nucleophile is hard or soft (Figure 6.11) (Uittenbogaard *et al* 2011). This has clear implications for the results of the *in silico* profiling, suggesting that both mechanisms need to be included for this class of chemicals. In addition, the *in silico* predictions indicating 5- and 6-membered lactone ring systems are reactive are incorrect. This additional mechanistic information needs including in the *in silico* profilers.

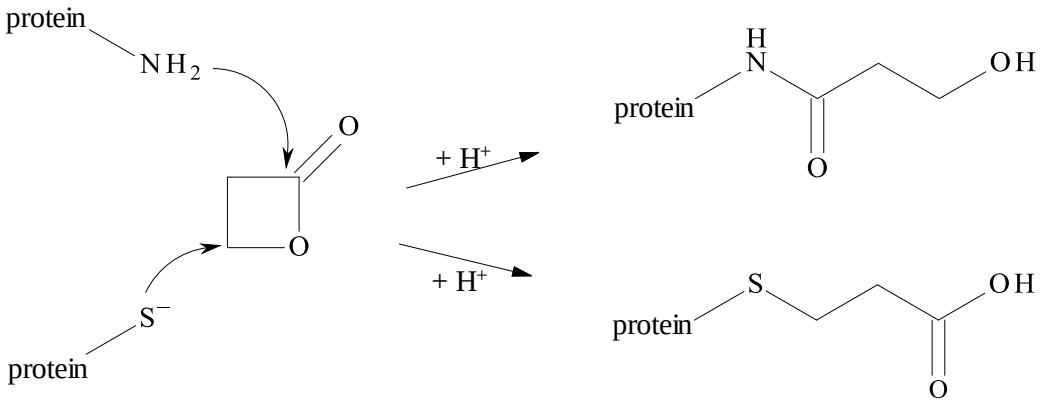


Figure 6.11: Scheme showing the potential ring opening acylation reactions for chemicals containing a lactone ring system (β-propiolactone is shown)

Nitrosos / nitros

Profiling of the nitroso chemicals in this class resulted in the OASIS profiler identifying the three of them as being potentially reactive towards proteins via nucleophilic addition (Figure 6.12). In contrast, the OECD profiler did not identify any of the chemicals as being potentially reactive. Inspection of the experimental data showed all of the nitroso chemicals to be reactive towards GSH and to show excess toxicity towards *Tetrahymena pyriformis*, except 1-nitrosopyrrolidine (chemicals 14-17 in Table 6.1). This chemical is not reactive due to it containing an N-nitroso rather than a C-nitroso moiety. The N-nitroso moiety in this chemical cannot undergo the direct nucleophilic addition reaction shown in Figure 6.12. The fact that the ring system of this chemical is aliphatic rather than aromatic plays a role in its lack of reactivity. An aliphatic ring system cannot stabilise the intermediate resulting in a significant increase in the activation energy. These prevent the reaction from occurring in the *in chemico* and *in vitro* test systems, as well as under biologically relevant conditions. These results show that the OASIS profiler correctly profiles chemicals of this type, whilst the chemistry within the OECD profiler needs expanding. The final chemical in this class, 4-niropyridine-N-oxide, contains an aromatic nitro and is profiled as being non-reactive by both the OASIS and OECD and profilers (chemical 18 in Table 6.1). In contrast, the

experimental data show this chemical to be reactive and exhibit excess toxicity. There are two potential reasons for this observed reactivity: the first is that an adduct may be produced between the 4-nitropyridine N-oxide and the sulphide group of GSH. The second potential reason is that, as 4-nitropyridine N-oxide is an oxidising agent, GSH depletion may occur due to oxidation of the thiol group to form a disulphide bond. In order to elucidate the mechanism further experimental work is required with additional chemicals containing a pyridine N-oxide moiety.

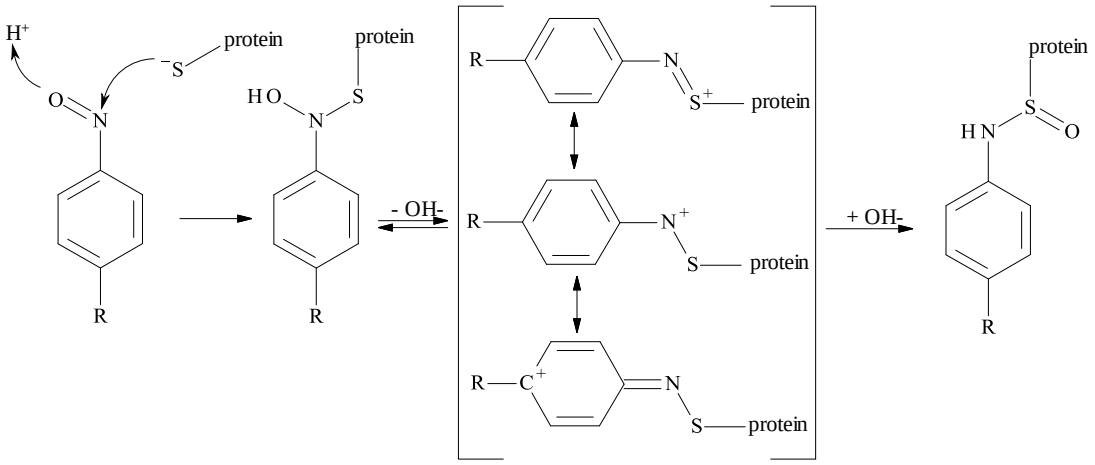


Figure 6.12: Scheme showing the nucleophilic addition reaction (showing key resonance forms) for aromatic nitroso chemicals with nucleophilic moiety of glutathione (R = N=O, N, O or CH_3)

Aldehydes and ketones

The dataset contained 12 chemicals with either an aldehyde or ketone functional group (chemicals 19-30 in Table 6.1). Profiling these chemicals with the OASIS profiler resulted in the aldehydes and di-ketones being identified as being potentially capable of forming a Schiff base (chemicals 19-25, mechanism shown in Figure 6.13). The OECD profiler identified all but one of this sub-set of chemicals as also being Schiff base formers, the exception being 1-phenyl-1,2-propanedione (chemical 22). The reactivity data for these

chemicals showed none of them were capable of reacting with GSH. In contrast, phenyl acetaldehyde and the di- ketones exhibited excess toxicity in the *T. pyriformis* assay. This discrepancy can be understood by the requirement for a free lysine side chain in order for a Schiff-base reaction to occur. Given that GSH does not contain a suitable lysine unit chemicals acting via Schiff base formation do not show reactivity in the GSH based assay. In addition, only the most reactive aldehyde showed excess toxicity, with the toxicity of the remaining two aldehydes being not significantly different from baseline. Importantly, the Schiff base mechanism of toxicity for simple aldehydes and di-ketones highlights the need for the development of *in chemico* methods capable of assessing nitrogen-based chemical reactivity. Such methods would not be limited by the effect of narcosis-driven toxicity for chemicals with either low reactivity or a high hydrophobicity (or a combination of the two). Additionally, the discrepancy with this class of chemicals highlights the need to use multiple experimental assays in order to define the chemical space of *in silico* profilers.

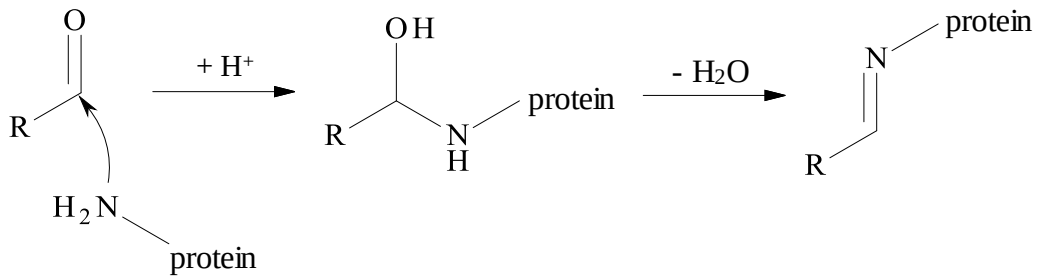


Figure 6.13: Scheme showing reaction between an aldehyde and an amine functional group leading to the formation of a Schiff base. An analogous reaction can occur for chemicals containing a 1,2-di-ketone moiety (R = alkyl or carbonyl group).

The final set of chemicals in this class were simple mono-ketones, these were profiled as being non-reactive by both the OASIS and OECD profilers (chemicals 26-30 in Table 6.1). The *in silico* profiling results are in keeping with the experimental data that showed these

chemicals to be non-reactive to GSH and not to exhibit excess toxicity. These results are as expected due to the decreased reactivity towards Schiff base formation that mono-ketones display compared to aldehydes (Roberts *et al* 2006). This is understood in terms of the relative reactivity of the aldehyde carbonyl group compared to the carbonyl group in a ketone. The decreased reactivity observed for the mono-ketones, in comparison to aldehydes, can be explained by two factors: firstly, the additional alkyl group present at the carbonyl carbon adds steric bulk to the reactive site; secondly, the additional alkyl group also acts to donate electrons onto the carbonyl carbon via a plus inductive effect. Both of these factors contribute to the decrease in reactivity observed within mono-ketones by making the reactive site less accessible and reducing the partial positive charge present on the carbonyl carbon respectively. This can be seen clearly for the 1,2-di-ketones each of which show significant excess toxicity, which as discussed is an indication of covalent reactivity.

Strained Ring Hydrocarbons

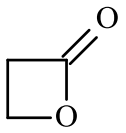
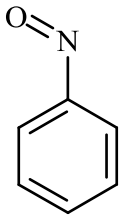
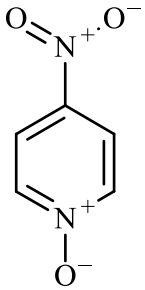
The final group of chemicals investigated were three hydrocarbon chemicals containing a cyclopropyl moiety. The three chemicals were profiled as being non-reactive by both the OASIS and OECD profilers as a consequence of not triggering a structural alert. The *in silico* predictions were confirmed by the experimental data that showed no reactivity towards GSH and no excess toxicity in the *T. pyriformis* assay (chemicals 31-33 in Table 6.1). This can be rationalised in terms of the absence of an electrophilic moiety in each of these chemicals. Without an electrophilic moiety these chemicals lack the ability to form a covalent bond with the proteins present within the test systems, hence the observed results.

6.3.2 Verification of structural alerts

The experimental data outlined can be used to verify the structural alerts in each of the profilers investigated. Additionally, the data can be used to identify new, and refine existing, structural alerts present within each of the profilers (Table 6.2). A well-defined chemical space is important as it gives the user greater confidence in the predictive capabilities of the profiler, so long as the query chemical falls within the chemical space used in the development of the profiler. The results show that the chemical space for the structural alerts relating to epoxides and aldehydes are well defined in both profilers. In contrast, the alerts related to lactones (lactones structural alert in Table 6.2) need refining. This refinement is two-fold, firstly the profilers need to identify the 4-membered lactones as having the ability to act via either acylation or S_N2 depending upon the nucleophile. The second refinement relates to the removal of the structural alerts for 5- and 6-membered lactones from both profilers. This is to prevent the profilers incorrectly identifying non-reactive 5- and 6-membered lactones as being reactive. In addition, there is the potential for both profilers to add a new structural alert to cover 4-nitropyridine N-oxide as neither identified this chemical as being capable of depleting glutathione. However, further experimental work must be conducted before a mechanistic rationale can be associated with this alert. Finally in contrast to the OASIS profiler, the OECD profiler fails to identify nitroso containing chemicals and 1-phenyl-1,2-propanedione as being reactive. Thus, the OECD profiler needs additional structural alerts, relating to aromatic nitroso containing chemicals, that covers the nucleophilic addition reaction chemistry exhibited by these types of chemicals (nitrosos structural alert within Table 6.2). The data also show the importance of using multiple profilers in category formation as, with the exception of 4-nitropyridine-N-oxide, all of the potentially reactive chemicals were identified by at least one of the profilers. The use of multiple profilers, to investigate the same endpoint, offers two key benefits over using a single profiler in isolation. The first is a wider coverage of the chemical space, meaning there will be a greater likelihood that chemicals, within a given dataset, will fall under the

boundary of at least one of the profilers. The second benefit is an increase in the confidence of the profiling results when each of the *in silico* profilers agree.

Table 6.2: Summary of the chemical space for the structural alerts investigated

Structural alert	Profiler correct		Additional alerts required
	OASIS	OECD	
Epoxides	Yes	Yes	None required
Lactones	No	No	Yes (refinement): 4-ring lactones only 
Nitrosos	Yes	No	Yes (OECD profiler only): aromatic nitroso 
Nitros	No	No	Yes 4-Nitropyridine-N-oxide 
Aldehydes	Yes	Yes	None required
Ketones	Yes	Yes	None required

6.4 Conclusions

The aim of this chapter was to illustrate the importance of using *in vitro* and *in chemico* assay data to verify structural alerts: in this instance the alerts held within the OASIS and OECD protein binding profilers in the OECD QSAR Toolbox. However, it is envisaged that a similar process, using relevant *in vitro*/*in chemico* assays, could be performed in order to

verify structural alerts pertaining to other initiating events, for example the alerts identified in Chapters 3, 4, and 5 (discussed in Chapter 7). An additional aim was to utilise the experimental data to identify where new alerts were required, or to suggest refinements to existing alerts, in order to extend the chemical space of the profilers. The structural alerts investigated related to acylation, Schiff base formation and the S_N2 mechanisms for covalent bond formation. The results showed that the structural alerts for epoxides, 4-membered ring lactones, nitroso, aldehydes, and di-ketones were correctly profiled by at least one of the *in silico* profilers. In addition, all of the chemicals that the experimental data showed to be non-reactive were correctly profiled by both profilers; seen by the absence of an alert for the non-reactive chemicals. However, the experimental data also highlighted the need for modification to existing structural alerts or the addition of new structural alerts to one, or both, of the profilers. The results outline the importance of using experimental data to define the chemical space of the structural alerts associated with *in silico* profilers. Furthermore, iterative refinement of profilers based upon experimental data as they become available leads to an increase in confidence, which is necessary for any regulatory use. Finally, the utility of using multiple profilers in category formation has been highlighted, i.e. the concomitant use of multiple profilers increases the confidence of profiling results and widens the chemical space that is covered.

Chapter 7: Discussion

The first section of the final chapter of this thesis will summarise, and discuss, the main conclusions of the research presented within Chapters 2 to 6. A full discussion of the results can be found within each of the respective chapters. The second section will focus on the future work that is required in order for progress to be made with regards to developing structural alerts for repeat dose toxicity; more specifically toxicity initiated via mitochondrial dysfunction; and how this work could be developed into a practical tool for end-users.

7.1 Progress in developing structural alerts for repeat dose toxicity

7.1.1 Summary of work

From the outset, the work presented in this thesis has been focussed on the development of an *in silico* profiler that can be utilised to assist in the safety assessment of chemicals upon repeated exposure. This need has arisen due to the introduction of EU legislation, such as REACH and the 7th amendment to the Cosmetics Directive. As part of this legislation, more traditional *in vivo* toxicity testing cannot be used as part of the safety assessment for cosmetic products or their ingredients. Therefore, alternative techniques are required to ensure the continued safety of these products for consumers. Therefore, Chapter 1 introduced the broad area of *in silico* toxicology, with a focus on Adverse Outcome Pathways and category formation, and the impact the EU regulation has had on driving research in the area of *in silico* toxicology since its inception and implementation over the past decade.

The availability, and accessibility, of relevant toxicological data, including observed adverse physiological effects, is a necessity prior to the development of any *in silico* models or profilers. There is a multitude of commercially, and freely, available databases that hold a wide variety of toxicological data. However, there is a need for a single, comprehensive, and freely available database containing repeat dose toxicity data associated with chemical structures. Chapter 2 described how novel repeat dose toxicity data were extracted from EU regulatory (SCC(NF)P/SCCS) reports and, subsequently, input into the ToxRefDB data entry tool in order to be uploaded (by other partners in COSMOS) into the COSMOS database. These data were harvested following a standard operating procedure, developed by colleagues on the COSMOS project. This SOP provided a consistent method by which each of the data harvesters were to extract, and input, the data. Information extracted from the reports included the NOAEL and/or LOAEL values and the histopathological findings observed whilst undertaking the experiment. The investigation performed at the end of Chapter 2 examined whether the results from 28 day repeat dose toxicity studies are protective of results from 90 day repeat dose studies held in the COSMOS database. The outcome of this investigation identified that for six of the nine chemicals (66%) the 28 day study was protective of the 90 study, i.e. if the toxicity value for the 28 day study was over 1000 mg/kg bw/day the toxicity value for the 90 day study was also over 1000 mg/kg bw/day. The percentage of those chemicals that were protective within this investigation was marginally under those found previously by the HSE and Taylor et al (2014). This may, however, be explainable by the variances in the total number of chemicals in the final datasets between the investigation performed in Chapter 2 and those carried out by the HSE and Taylor et al (2014). Therefore, the results from the analysis performed in this chapter support the findings by the HSE and Taylor et al (2014). This investigation, in conjunction with the previous work set out by the HSE and Taylor et al (2014), could have a major impact both financially, and in terms of animal usage, with regards to those chemicals manufactured (or imported) above 100 tonnes per year under REACH. In addition, this investigation demonstrated that having toxicological values, and histopathological findings,

compiled within the same database enabled this type of analysis to be performed more easily than if multiple databases were used.

In Chapter 3 a chemoinformatics analysis was performed on a set of chemicals, from the scientific literature, with associated qualitative data pertaining to mitochondrial toxicity. This analysis was undertaken using the freely available data mining software, ChemoTyper, which contains the ToxPrint library of predefined structural fragments. The analysis performed in this chapter found two types of structural alerts could be identified utilising this software; 1) well-defined alerts that could be associated with a mechanistic hypothesis, 2) more diverse alerts for which it was not possible to hypothesise a mechanism for the entire category. Overall, a total of twenty alerts were developed. Of these alerts, it was possible to hypothesise a mechanism encompassing all chemicals identified by the ChemoTyper for two alerts. For the remaining alerts it was not possible to hypothesise a mechanism that encompassed all the chemicals identified as 'toxic to mitochondria' within the group. In addition, this chapter also outlined the use of these different types of structural alerts; with mechanism-based structural alerts being intended for category formation and read-across, whilst chemistry-based alerts could be used for the purposes of screening and prioritisation of an inventory. The inherent differences in the two types of alert make them suitable for different purposes. Mechanistic information associated with the mechanism-based alerts provides additional support to both the development of chemical categories and subsequent read-across predictions for toxicity. In comparison, chemistry-based alerts, whilst lacking mechanistic information, are associated with toxicity, therefore, they can be used to identify chemicals, within an inventory, for which further *in vitro*/*in chemico* testing may be appropriate.

The work in Chapter 4 focussed on utilising structural similarity and category formation to re-analyse the data set from Chapter 3. Overall, 35 chemicals in the data set were identified as belonging to categories containing mitochondrial toxicants: local anaesthetics, anti-anginal, and anti-arrhythmic (6 chemicals); anti-diabetic drugs (3 chemicals); non-steroidal

anti-inflammatory drugs (3 chemicals); anthracycline antibiotics (4 chemicals); hypolipodemic drugs (3 chemicals); bile acids (6 chemicals); anti-histaminic, anti-psychotic and anti-emetic drugs (7 chemicals); and β-blockers (3 chemicals). A total of eight mechanism-based alerts were developed covering five initiating events; inhibition of the electron transport chain, alternative electron acceptance, initiation of the death receptor pathway, uncoupling of oxidative phosphorylation and induction of the membrane permeability transition. Additionally, the work carried out in Chapters 3 and 4 demonstrated that no one approach can be utilised to identify all possible structural alerts. Therefore, it is envisaged that these techniques will be used in combination to cover as large a chemical space as possible.

The toxicological information provided by the regulatory dossiers for 94 hair dye chemicals, published by the SCC(NF)P/SCCS, was utilised in Chapter 5 to develop mechanism-based structural alerts. These dossiers are a valuable, yet currently under-used, source of toxicological data. The analysis performed in Chapter 5 expanded on the work undertaken in Chapter 4 by identifying additional mechanism-based alerts associated with mitochondrial dysfunction. A total of four mechanism-based alerts were identified covering pro-quinones (37 chemicals), quinones (7 chemicals), meta-substituted benzenes (4 chemicals), and aromatic azo compounds (8 chemicals). Each of these alerts is associated with inducing mitochondrial toxicity via a single Molecular Initiating Event (MIE): inhibition of the electron transport chain. The alerts identified in Chapter 5 broadened the chemical space regarding those chemicals that have the potential to induce mitochondrial toxicity via inhibition of the electron transport chain. These alerts can be utilised to either screen an inventory for prioritisation or to develop chemical categories, from which read-across predictions could be made regarding to a chemical's ability to initiate mitochondrial toxicity. In order to expand the use of this profiler for mitochondrial toxicity additional alerts are required for (alternative) MIEs to account for the toxic potential of the remaining chemicals. The work undertaken in Chapters 4 and 5 demonstrate the vital importance of the available

literature in providing the mechanistic information necessary for developing mechanism-based structural alerts. Chapter 5 reiterates the use of mechanistic structural alerts for the purposes of grouping and profiling. Furthermore, the work carried out within this chapter highlights the usability and usefulness of the information contained within regulatory reports, such as those published by the SCC(NF)P/SCCS.

Upon development of structural alerts it is necessary to substantiate that each of the alerts correctly identifies chemicals with the potential to instigate an MIE. Other alternative techniques, such as *in vitro* and/or *in chemico* assays, can be used to verify the alerts developed are correct. However, it should be noted that whilst these alerts may be correct other factors, such as the internal concentration or metabolism, may mean the MIE is not induced *in vivo*. Finally, Chapter 6 illustrates the importance of using data generated by *in vitro* and *in chemico* assays to verify, and refine, structural alerts. These alternative techniques are important as they provide experimentally derived mechanistic information. In turn, this information can be utilised to verify the correct mechanism is associated with an alert, whilst also providing support for possible refinements. In Chapter 6 the structural alerts investigated relate to the protein binding alerts in the OECD QSAR Toolbox. However, a similar process could be undertaken to verify other structural alerts, such as those pertaining to mitochondrial toxicity identified in Chapters 3, 4, and 5.

7.1.2 Conclusions

Mitochondrial toxicity has been implicated as one of the key drivers of various organ-level toxicities (Dykens and Will 2008). The work presented within this thesis has identified a variety of structural alerts that have the ability to induce mitochondrial toxicity. This was achieved by utilising various *in silico* approaches in order to analyse two data sets; the first containing qualitative mitochondrial data, and the second containing quantitative toxicity data from regulatory submissions. Overall, a total of 31 structural alerts were developed. Of

these 31 alerts, twelve were mechanism-based alerts, whilst the remaining 19 were chemistry-based alerts.

Since the implementation of both REACH and the 7th amendment to the Cosmetics Directive there has been an increased interest in the use of alternative techniques to provide information for safety assessments. Whilst numerous alternative techniques are available no one method can replace animal testing outright. Therefore, the AOP paradigm has been suggested as a framework by which data from *in vitro*, *in chemico*, and *in silico* methods can be integrated in order to better organise the toxicological information provided by these techniques. An AOP is a scheme that links an MIE, via testable events at different levels of biological organisation, to an adverse effect relevant for risk assessment. The MIE provides mechanistic information regarding the initial interaction between the chemical and the biological system. *In silico* methods contribute to the AOP paradigm by identifying the chemical structures that are associated with inducing an MIE. As discussed in previous chapters (Chapters 1, 3, 4 and 5) the twelve mechanism-based alerts are structural fragments that have information, from experimental studies, pertaining to the mechanism by which toxicity is initiated associated with them. When developing mechanism-based alerts it is essential that an in-depth analysis of the relevant scientific literature, and/or information generated from *in vitro* and/or *in chemic*o assays are utilised. This is due to the fact that these experimental data provide critical evidence demonstrating how the structural alert(s) induce the observed toxicity via the MIE. These mechanism-based alerts can, subsequently, be aggregated together to construct a mechanism-based *in silico* profiler. The main use of a mechanism-based profiler should be to develop chemical categories centred on the ability to induce the same MIE. Thus, these chemical categories can be used for read-across purposes in order to fill data gaps for a novel chemical utilising the toxicological data present for the analogous chemicals within the category. The use of the mechanistic information associated with the MIE enables any read-across prediction made to be more mechanistically interpretable and, therefore, more robust.

In comparison, the 19 chemistry-based alerts have been identified as being associated with inducing mitochondrial toxicity. However, no accompanying mechanistic information is (currently) available that relates the alert with how mitochondrial toxicity is initiated. Given the lack of mechanistic information, chemistry-based alerts should not be used in the development of chemical categories or subsequent read-across analysis, with respect to hypothesising the molecular initiating event of (mitochondrial) toxicity. This is due to the fact that whilst the chemistry-based alerts have been associated with toxicity, it may not be possible to identify a (single) mechanism by which all chemicals containing the alert initiate toxicity, as discussed in Chapter 3. However, chemistry-based alerts can be useful for screening, and prioritising, chemicals held in an inventory that should undergo further testing within *in vitro*/*in chemico* assays. This additional testing could, potentially, enable both the mechanistic information relating to a structural alert to be elucidated; thereby, converting it into a mechanism-based alert; and/or enable the alert to be refined. Additionally, it should be noted that mechanism-based alerts can also be used in order to screen an inventory. The results of this screening would have the added benefit of informing a user as to the type of *in vitro*/*in chemico* tests that should be performed first.

7.1.3 Proposal of a tiered testing strategy for profiling of chemical inventories

It is envisaged that the mechanism-based alerts (developed in Chapters 4 and 5) will be used in conjunction with the chemistry-based alerts (developed in Chapter 3) as part of a two tiered testing strategy for profiling chemical inventories. The first tier would comprise profiling the chemicals against the mechanism-based alerts. Whilst these alerts are likely to trigger only a small proportion of the chemicals in a data set they provide a high level of confidence as to the potential of these chemicals to be mitochondrial toxicants. This confidence is derived from the mechanistic information, which is associated with the chemical structure, being available in the literature and being used in the development of the

mechanism-based structural alert. Thus, this information can be used to support the identification of potential mitochondrial toxicants. Additionally, if a chemical triggered a mechanism-based alert the chemical would be of a high priority to test, with the mechanistic information directing the *in vitro*/*in chemico* assay(s) that should be undertaken The second tier of this profiling strategy would be the use of the chemistry-based structural alerts. In comparison to the mechanism-based alerts above, the chemistry-based alerts are likely to trigger a significantly larger proportion of the chemicals in the inventory, thereby, increasing the chemical space covered; this is due to the more generic nature of the chemistry-based alerts. Whilst a larger number of chemicals are identified there is less confidence surrounding the identification of potential mitochondrial toxicants. This is due to the fact that these chemistry-based alerts do not have any mechanistic information associated with their ability to induce toxicity. If a chemical triggered a chemistry-based alert the chemical would be of concern. However, due to the absence of inherent mechanistic information and the lower confidence in the assignment of a 'toxic' prediction a larger quantity of *in vitro*/*in chemico* testing, and/or supporting information from the available literature, would be needed to ascertain the toxic potential of the chemical.

7.2 Prospects for future work

The work carried out in this thesis has enabled the development of a number of mechanism- and chemistry-based structural alerts relating to mitochondrial toxicity. However, further work that is needed centres on the verification, and refinement, of the alerts presented in the previous chapters, alongside developing additional alerts using other data sets, such as ToxCast. This further work is described in detail below.

7.2.1 Verification of the developed mechanism-based structural alerts

Data present in the available scientific literature has enabled twelve mechanism-based structural alerts to be developed within this thesis (Chapters 4 and 5). The work outlined in Chapter 6 demonstrated the importance of utilising data generated from *in vitro*/*in chemico* assay in order to verify structural alerts, in this instance for protein binding. Therefore, it is envisaged that future work could involve the use of *in vitro* and/or *in chemico* assays relevant for mitochondrial toxicity to verify the mechanism-based alerts developed in previous chapters. This would involve taking a representative sample of chemicals, containing one of the alerts developed, and utilising the hypothesised mechanism for each alert to guide the mitochondrial assay(s) that the chemicals should be tested in. For example, the representative chemical(s) for the local anaesthetics category should be tested within assays that identify uncouplers of oxidative phosphorylation, such as the use of oxygen-sensitive fluorescent probes (Hynes et al 2008). Briefly stated, the level of fluorescence emitted by the oxygen-sensitive dye changes with the presence or absence of, either intracellular or extracellular, molecular oxygen, i.e. a decreased presence of molecular oxygen (due to an increase in electron transport chain activity) increases the level of fluorescence, thereby, identifying the mechanism as uncoupling of oxidative phosphorylation (Hynes et al 2008). In comparison, the representative chemical(s) for the anti-diabetic drugs (thiazolidinedione) category should be tested within assays that identify inhibitors of the electron, such immunocapture-based assays (Nadanaciva 2008). These immunocapture-based assays use 96-well plates coated with monoclonal antibodies against Complexes I, II, IV and V to measure the activity of each of these complexes after exposure to the test chemical to help identify which complex is inhibited (Nadanaciva 2008). Subsequently, the results from these assays would be used in order to verify whether the hypothesised mechanism associated with the structural alert is correct. Additionally, this information could be utilised to further refine the developed structural alert.

7.2.2 Use of in vitro/in chemico data to discern mechanistic information for chemistry-based alerts

The work performed in Chapter 3 has enabled the development of 19 chemistry-based alerts that do not have a mechanistic hypothesis associated with them. This is due to the chemistry-based alerts being, relatively, more broad, thereby, identifying more diverse chemicals. Thus, this diversity increases the likelihood that a variety of mechanisms will initiate the observed toxicity. In addition to verifying mechanism-based alerts, the data generated by *in vitro* and/or *in chemico* assays could be utilised to discern mechanistic information associated with a chemical structure. This could, in turn, permit the chemistry-based structural alerts to be refined to mechanism-based alerts. Thus, enabling chemicals that contain the same chemistry-based alert but that initiate different MIEs to be discerned. The use of such assays would allow the refinement of the chemistry-based alert into a mechanism-based alert by associating a mechanistic hypothesis with the alert.

7.2.3 Development of additional alerts

Whilst the work carried out in this thesis developed a number of mechanism- and chemistry-based structural alerts it is acknowledged that the chemical space covered by these alerts is relatively small. Therefore, future work could entail the use of other data sets containing mitochondrial toxicity data, for example ToxCast, in order to develop additional mechanism- and chemistry-based alerts. It is envisaged that processes similar to those described in Chapter 3, 4, and 5 could be used in order to develop these alerts. Subsequently, any alerts that are identified using these data sets should undergo the verification, and refinement, testing within the *in vitro* and/or *in chemico* assays described above. Thus, it is expected that these additional alerts would expand the chemical domain of the *in silico* profiler developed within this thesis.

It is envisaged that the alerts developed in this thesis (together with any alerts derived in the future) will be incorporated into predictive tools or software; such as workflows in KNIME or as mechanistic alerts within the OECD QSAR Toolbox. These tools could, subsequently, be utilised to screen chemical inventories in order to help identify chemicals with the potential to induce mitochondrial toxicity. Furthermore, the use of these alerts, in conjunction with other alternative techniques and information available in the literature, could be utilised in the development, and verification, of Adverse Outcome Pathways that extend up to physiological effects at the organism level.

8: References

Adler S, Basketter D, Creton S, et al (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects—2010. *Arch Toxicol* 85(5):367-485

Almotrefi A, and Dzimiri N (1992) Effect of beta-adrenoceptor blockers on mitochondrial ATPase activity in guinea pig heart preparations. *Eur J Pharmacol* 215:231-236

Amacher DE (2005) Drug-associated mitochondrial toxicity and its detection. *Curr Med Chem* 12:1829-1839

Ames BN, Lee FD, and Durston WE (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. *PNAS USA* 70(3): 782 - 786

Ankley GT, Bennett RS, Erickson RJ, et al. (2010) Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol and Chem* 29(3):730-741

Aptula AO, and Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. *Chem Res Toxicol* 19(8): 1097-1105

Aptula AO, Enoch SJ, and Roberts DW (2009) Chemical mechanisms for skin sensitisation by aromatic compounds with hydroxy and amino groups. *Chem Res Toxicol* 22(9):1541-1547

Aptula AO, Patlewicz G, Roberts DW, and Schultz TW (2006) Non-enzymatic glutathione reactivity and *in vitro* toxicity: A non-animal approach to skin sensitization. *Toxicol in Vitro* 20: 239-247

Baan R, Straif K, Grosse Y, et al. (2008) Special Report: Policy. Carcinogenicity of some aromatic amines, organic dyes, and related exposures. *The Lancet Oncology*. 9(4):322-323

Balijepalli S, Boyd MR, and Ravindranath V (1999) Inhibition of mitochondrial complex I by haloperidol: the role of thiol oxidation. *Neuropharmacology* 38:567-577

Bard J (2007) Systems developmental biology: the use of ontologies in annotating models and in identifying gene function within and across species. *Mamm Genome* 6-7: 402-411

Battaglia V (2005) Oxidative Stress Is Responsible for Mitochondrial Permeability Transition Induction by Salicylate in Liver Mitochondria. *J Biol Chem* 280:33864-33872

Bironaite DA, Cenas NK, and Kulys J (1991) The rotenone-insensitive reduction of quinones and nitrocompounds by mitochondrial NADH: ubiquinone reductase. *Biochim Biophys ACTA* 1060:203-209

Bitsch A, Jacobi S, Melber C, et al (2006) REPDOSE: A database on repeated dose toxicity studies of commerical chemicals - A multifunctional tool. *Regul Toxicol Pharm* 46: 202-210

Blaschke U, Paschke A, Rensch I, and Schuurmann G (2010) Acute and Chronic Toxicity toward the Bacteria Vibrio fischeri of Organic Narcotics and Epoxides: Structural Alerts for Epoxide Excess Toxicity. *Chem Res Toxicol* 23: 1936-1946

Boelsterli U (2002) Mechanisms of NSAID-induced hepatotoxicity: Focus on Nimesulide. *Drug Safety* 25:633-648

Brunmair B, Staniek K, Gras F, Scharf N, et al (2004) Thiazolidinediones, like metformin, inhibit respiratory complex I. A common mechanism contributing to their antibiabetic actions? *Diabetes* 53:1052-1059

Cela O, Piccoli C, Scrima R, Quarato G, et al (2010) Bupivacaine uncouples the mitochondrial oxidative phosphorylation, inhibits respiratory chain complexes I and III and enhances ROS production: results of a study on cell cultures. *Mitochondrion* 10:487-496

Chan K, Truong D, Shangari N, and O'Brien PJ (2005) Drug-induced mitochondrial toxicity. *Expert Opin Drug Met* 1:655-669

Cohen BH, and Gold DR (2001) Mitochondrial cytopathy in adults: What we know so far. *Clev Clin J Med* 68:625-642

Collins TJ, and Bootman MD (2003) Mitochondria are morphologically heterogeneous within cells. *J Exp Biol* 206: 1993 - 2000

Corbett JE, Sharma RK, and Dressler WE (1999) Cosmetic toxicology. In: Marquardt H, Schafer SG, McClellan RO, Welsch F (eds.), *Toxicology*. Academic Press, San Diego, CA, USA. P 899-918

Cosmetics Europe https://www.cosmeticseurope.eu/safety-and-science-cosmetics-europe/products-and-ingredients/hair-dyes.html (accessed 28.09.2014)

Cronin MTD (2002) The current status and future applicability of quantitative structure-activity relationships (QSARs) in predicting toxicity. *ATLA* 30(Suppliment 2): 81-84

Cronin MTD (2010). Finding the data to develop and evaluate (Q)SARs and populate categories for toxicity prediction. In: Cronin MTD, and Madden JC (ed) *In Silico Toxicology: Principles and Applications*. Cambridge, UK, Royal Society of Chemistry

Cruz TS, Faria PA, Santana DP, Ferreira JC, et al (2010) On the mechanisms of phenothiazine-induced mitochondrial permeability transition: Thiol oxidation, strict Ca^{2+} dependence, and cyt *c* release. *Biochem Pharmacol* 80:1284-1295

Dabadie P, Bendriss P, Erny P, and Mazat JP (1987) Uncoupling effects of local anesthetics on rat liver mitochondria. *FEBS Letters* 226:77-82

de Groot H, and Noll T (1983) Halothane hepatotoxicity: Relation between metabolic activation, hypoxia, covalent binding, lipid peroxidation and liver cell damage. *Hepatology* 3(4): 601 - 606

Dippenaar JM (2007) Local anaesthetic toxicity. *SAJAA* 13(3):23-28

Dykens JA (2008) Preface. In: Dykens JA, and Will, Y. (ed) *Drug-Induced Mitochondrial Dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

Dykens JA and Will Y (2007) The signifcance of mitochondrial toxicity testing in drug development. *Drug Discov Today* 12:777-785

Dykens JA and Will Y (2008) *Drug-induced mitochondrial dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

Dykens JA, Marroquin LD, and Will Y (2007) Strategies to reduce late-stage drug attrition due to mitochondrial toxicity. *Expert Rev Mol Diagn* 7:161-175

EC (2003) Directive 2003/15/EC of the European Parliament and of the Council (Cosmetics Directive 7th Amendment) *Official Journal of the European Union*

EC (2006a) Corrigenda of Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) *Official Journal of the European Union* EC 1907/2006

EC (2006b) Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and

Restriction of Chemicals (REACH) *Official Journal of the European Union* EC 1907/2006

EC (2007) REACH technical guidance. Available at http://ecb.jrc.it/reach/rip/ (accessed

ECHA (2008) Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. Available at: http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

ECHA (2012) Practical guide 6: How to report read-across and categories. 1-30. Available at: http://echa.europa.eu/documents/10162/13655/pg_report_readacross_en.pdf

Ellison CM, Cronin MTD, Madden JC, and Schultz TW (2008) Definition of the structural domain of the baseline non-polar narcosis model for Tetrahymena pyriformis. *SAR QSAR Environ Res* 19: 751-783

Enoch SJ (2010) Chemical category formation and read-across for the prediction of toxicity. In: Puzyn T, Leszczynski J, and Cronin MTD (ed) *Recent Advances in QSAR studies: Methods and applications*. Springer, Dordrecht, Heidelberg, London, New York. 209 - 219

Enoch SJ, and Cronin MTD (2010) A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Crit Rev Toxicol* 40:728-748

Enoch SJ, and Roberts DW (2013a) Approaches for grouping chemicals into categories. In: Cronin MTD, Madden JC, Enoch SJ, Roberts DW (eds) *Chemical Toxicity Prediction: Category Formation and Read-Across.* The Royal Society of Chemistry, Cambridge, UK, p 30-43

Enoch SJ, and Roberts DW (2013b) Predicting skin sensitisation potency for Michael acceptors in the LLNA using quantum mechanics calculations. *Chem Res Toxicol* 26: 767 - 774

Enoch SJ, Cronin MTD, and Ellison CM (2011a) The use of a chemistry based profiler for covalent DNA binding in the development of chemical categories for read-across for genotoxicity. *ATLA* 39:131-145

Enoch SJ, Cronin MTD, and Schultz TW (2012a) The definition of the applicability domain relevant to skin sensitisation for the aromatic nucleophilic substitution electrophilic mechanism. *SAR QSAR Environ Res* 23:649-663

Enoch SJ, Cronin MTD, and Schultz TW (2013a) The definition of the toxicologically relevant applicability domain for the S_NAr reaction for substituted pyridines and pyrimidines. *SAR QSAR Environ Res* 24:385-392

Enoch SJ, Cronin MTD, Madden JC, and Hewitt M (2009) Formation of Structural Categories to Allow for Read-Across for Teratogenicity. *QSAR Comb Sci* 28(6-7): 696-708

Enoch SJ, Cronin MTD, Schultz TW, and Madden JC (2008a) Quantitative and mechanistic read across for predicting the skin sensitisation potential of alkenes acting via Michael addition. *Chem Res Toxicol* 21:513-520

Enoch SJ, Ellison CM, Schultz TW, and Cronin MTD (2011b) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit Rev Toxicol* 41:783-802

Enoch SJ, Madden JC, and Cronin MTD (2008b) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ Res* 19:555-578

Enoch SJ, Przybylak KP, and Cronin MTD (2013b) Category Formation Case Studies. In: Cronin MTD, Madden JC, Enoch SJ, Roberts DW (eds) *Chemical Toxicity Prediction: Category Formation and Read-Across*. vol 17. RSC, Cambridge, UK

Enoch SJ, Seed MJ, Roberts DW, et al (2012b) Development of mechanism-based structural alerts for respiratory sensitisation hazard identification. *Chem Res Toxicol* 25:2490-2498

EPA (1995) The use of the benchmark dose approach in health risk assessment. EPA/630/R-94/007

Eto K, Fukuda T, Araki Y, Inoue B, and Ogata M (1985) Effect of tricyclic drugs on mitochondrial membrane. *Acta Med Okayama* 39:289-295

Faubion WA, Guicciardi ME, Miyoshi H, et al (1998) Toxic bile salts induce rodent hepatocyte apoptosis via direct activation of Fas. *J Clin Invest* 103:137-145

Freudenthal RI, Stephens E, and Anderson DP (1999) Determining the potential of aromatic amines to induce cancer of the urinary bladder. *Int J Toxicol* 18:353-359

Fink SL, and Cookson BT (2005) Apoptosis, pyroptosis, and necrosis: Mechanistic description of dead and dying eukaryotic cells. *Infect Immun* 73(4):1907-1916

Gago-Dominguez M, Castelao JE, Yuan J, et al (2001) Use of permanent hair dyes and bladder cancer risk. *Int J Cancer* 91:575-579

Gallagher CH, Koch JH, and Mann DM (1965) The effect of phenothiazine on the metabolism of liver mitochondria. *Biochem Pharmacol* 14:789-797

Gerwirtz DA (1999) A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. *Biochem Pharmacol* 57:727-741

Gorontzy T, Kuver J, and Blotevogel K (1993) Microbial transformation of nitroaromatic compounds under anaerobic conditions. *J Gen Microbiol* 139:1331-1336

Green DR, and Kroemer G (2004) The pathophysiology of mitochondrial cell death. *Science* 305:626-629

Guth PS, and Sprites MA (1964) The phenothiazine tranquilizers: biochemical and biophysical actions. *Int Rev Neurobiol* 7:231-278

Gutsell S, and Russell P (2013) The role of chemistry in developing understanding of adverse outcome pathways and their application in risk assessment. *Toxicol Res* 2(5):299-307

Hardy B, Apic G, Carthew P, et al. (2012) Toxicology ontology perspectives. *ATLA* 29: 139-156

Hartung T, Luechtefeld T, Maertens A and Kleensang,A (2013) Intergrated testing strategies for safety assessments. *ATLA* 30(1): 3 - 18

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem* 54:1015-1069

Hemminki K (1981). Reactions of beta-propiolactone, beta-butyrolactone and gamma-butyrolactone with nucleic acids. *Chem-Biol Interact* 34: 323 - 331

Henry TR, and Wallace KB (1995) Differential mechanisms of induction of the mitochondrial permeability transition by quinones of varying chemical reactivities. *Toxicol App Pharmacol* 134:195-203

Hewitt M, Cronin MTD, Rowe PH, and Schultz TW (2011) Repeatability analysis of the Tetrahymena pyriformis population growth impairment assay. *SAR QSAR Environ Res* 5-6: 621-637

Hewitt M, Enoch SJ, Madden JC, Przybylak KP, and Cronin MTD (2013) Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechansism(s) of action. *Crit Rev Toxicol* 43(7):537-558

HSE (2011) A Weight-of-Evidence approach for justifying the waiving the 90-day repeat dose toxicity test under REACH, ECHA/MSC-16/2011/002. Helsinki (19 January 2011)

Huncharek M, and Kupelnick B (2005) Personal use of hair dyes and the risk of bladder cancer: Results of a meta-analysis. *Public Health Reports* 120(1):31-38

Hynes J, O'riordan TC, and Papkovsky DB (2008) Use of oxygen-sensitive fluorescent probes for the assessemtn of mitochondrial function. In: Dykens JA, and Will, Y (ed) *Drug-Induced Mitochondrial Dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

IARC (2010) IARC monographs on the evaluation of carcinogenic risks to humans. Some aromatic amines, organic dyes, and related exposures. *Volume 99*. International Agency for Research on Cancer, Lyon, France

Jaeschke H, Gores GJ, Cederbaum AI, et al (2002) Mechanisms of Hepatotoxicity. *Toxicol Sci* 65:166-176

Jaworska J, and Nikolova-Jeliazkova N (2007) How can structural similarity analysis help in category formation? *SAR QSAR Environ Res* 18(3-4): 195-207

Kadenbach B (2003) Intrinsic and extrinsic uncoupling of oxidative phosphorylation. *Biochim Biophys ACTA - Bioenergetics* 1604(2): 77-94

Kalgutkar AS, Gardner I, Obach RS, et al. (2005) A comprehensive listing of bioactivation pathways of organic functional groups. *Curr Drug Met* 6:161-225

Kappus H (1986) Overview of enzyme systems invovled in bio-reduction of drugs and in redox cycling. *Biochem Pharm* 335:1-6

Keller BJ, Yamanaka H, and Thurman RG (1992) Inhibition of mitochondrial respiration and oxygen-dependent hepatotoxicity by six structurally dissimilar peroxisomal proliferating agents. *Toxicology* 71:49-61

Kim J-S, He L, and Lemasters JJ (2003) Mitochondrial permeability transition: a common pathway to necrosis and apoptosis. *Biochem Bioph Res Co* 304(3):463-470

Kitani A, So Y, and Miller LL (1981) An electrochemical study of the kinetics of NADH being oxidised by diimines derived from diaminobenzenes and diaminopyrimidines. *J Am Chem Soc* 103:7636-7641

Kleszczynski K, and Skladanowski AC (2011) Mechanism of cytotoxic action of perfluorinated acids. III. Disturbance in Ca(2)+ homeostasis. *Toxicol Appl Pharmacol* 251:163-168

Kleszczynski K, Stepnowski P, and Skladanowski AC (2009) Mechanism of cytotoxic action of perfluorinated acids II. Disruption of mitochondrial bioenergetics. *Toxicol Appl Pharmacol* 235:182-190

Klimisch H-J, Andreae M, and Tillmann U (1997) A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Reg Toxicol Pharmacol* 25: 1-5

Klionsky DJ, and Emr SD (2000) Autophagy as a regulated pathway of cellular degradation. *Science* 290:1717-1721

Koleva YK, Madden JC, and Cronin MTD (2008) Formation of categories from structure-activity relationships to allow read-across for risk assessment: Toxicity of alpha,beta-unsaturated carbonyl compounds. *Chem Res Toxicol* 21: 2300-2312

Kovacic P, Jacintho JD (2001a) Mechanisms of carcinogenesis: Focus on oxidative stress and electron transfer. *Curr Med Chem* 8:773-796

Kovacic P, Jacintho JD (2001b) Reproductive toxins: Pervasive theme of oxidative stress and electron transfer. *Curr Med Chem* 8:863-892

Kowaltowski AJ, Castilho RF, Vercesi AE (2001) Mitochondrial permeability transition and oxidative stress. *FEBS Letters* 495:12-15

Krahenbuhl S (2001) Mitochondria: important target for drug toxicity? *J Hepatol* 34:334-336

Krahenbuhl S, Talos C, Fischer S, and Reichen J (1994) Toxicity of bile acids on the electron transport chain of isolated rat liver mitochondria. *Hepatology* 19:471-479

Kroemer G, Galluzzi L, and Brenner C (2007) Mitochondrial membrane permeabilization in cell death. *Physiol Rev* 87:99-163

Kroon AM and de Vries H (1969) The effects of chloramphenicol on the biogenesis of mitochondria of rat liver *in vivo*. *FEBS Letters* 3: 208 - 210

Landesmann B, Goumenou M, Munn S, and Whelan M (2012) Description of prototype modes-of-action related to repeated dose toxicity. *JRC Scientific and Policy Report 75689*

Langley AE (1990) Effects of perfluoro-n-decanoic acid on the respiratory activity of isolated rat liver mitochondria. *J Toxicol Environ Health* 29:329-336

Lemasters JJ, Theruvath TP, Zhong Z, and Nieminen AL (2009) Mitochondrial calcium and the permeability transition in cell death. *Biochim Biophys ACTA* 1787:1395-1401

Levine B and Kroemer G (2008) Autophagy in the pathogenesis of disease. *Cell* 132:27-42

Lewis RW, Billington R, Debryune E, et al (2002) Recognition of adverse and nonadverse effects in toxicity studies. *Toxicol Pathol* 30: 66-74

Lucas-Heron B, Schmitt N, and Ollivier B (1994) Mdx mouse skeletal muscle: Could a mitochondrial factor be responsible for the absence of progressive necrosis? *Neuroscience* 129:97-100

Madden JC (2013). Sources of chemical information, toxicity data and assessment of their quality. In: Cronin MTD, Madden JC, Enoch SJ, and Roberts DW (ed) *Chemical Toxicity Prediction Category Formation and Read-Across*. Cambridge, UK, Royal Society of Chemistry

Masubuchi Y (2006) Metabolic and non-metabolic factors determining troglitazone hepatotoxicity: A review. *Drug Metab Pharmacok* 21:347-356

Masubuchi Y, Saito H, and Horie T (1998) Structural requirements for the hepatotoxicity of nonsteroidal anti-inflammatory drugs in isolated rat hepatocytes. *J Pharmacol Exp Ther* 287:208-213

Matsubara T, and Hagihara B (1968) Action mechanism of phenothiazine derivatives on mitochondrial respiration. *J Biochem* 63:156-164

Mehta R, Chan K, Lee O, Tafazoli S, and O'Brien PJ (2008) Drug-associated mitochondrial toxicity. In: Dykens JA, and Will, Y (ed) *Drug-Induced Mitochondrial Dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

Montaigne D, Hurt C, and Neviere R (2012) Mitochondria death/survival signalling pathways in cardiotoxicity induced by anthracyclines and anticancer-targeted therapies. *Biochem Res Int* 2012:ID 951539

Moreno-Sanchez R, Bravo C, Vasquez C, et al (1999) Inhibition and uncoupling of oxidative phosphorylation by Nonsteroidal Anti-Inflammatory Drugs. *Biochem Pharmacol* 57:743-752

Munday R (1992) Mitochondrial oxidation of p-phenylenediamine derivatives *in vitro*: Structure-activity relationships and correlation with mytoxic activity *in vivo*. *Chem-Biol Interact* 82:165-179

Myshkin E, Brennan R, Khasanova T, et al (2012) Prediction of organ toxicity endpoints by QSAR modelling based on precise chemical-histopathology annotations. *Chem Biol Drug Des* 80: 406-416

Nadanaciva S (2008) OXPHOS Complex Activity Assays and Dipstick Immunoassays for Assessment of OXPHOS Protein Levels. In: Dykens JA, and Will, Y (ed) *Drug-Induced Mitochondrial Dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

Nadanaciva S, and Will Y (2011) New insights in drug-induced mitochondrial toxicity. *Curr Pharm Design* 17:2100-2112

Nadanaciva S, Bernal A, Aggeler R, Capaldi R, and Will Y (2007a) Target identification of drug induced mitochondrial toxicity using immunocapture based OXPHOS activity assays. *Toxicol in Vitro* 21:902-911

Nadanaciva S, Dykens JA, Bernal A, Capaldi RA, and Will Y (2007b) Mitochondrial impairment by PPAR agonists and statins identified via immunocaptured OXPHOS complex activities and respiration. *Toxicol Appl Pharm* 223:277-287

Nam S, and Renganathan V (2000) Non-enzymatic reduction of azo dyes by NADH. *Chemosphere* 40:351-357

Naven RT, Swiss R, Klug-Mcleod J, Will Y, and Greene N (2013) The development of structure-activity relationships for mitochondrial dysfunction: Uncoupling of oxidative phosphorylation. *Toxicol Sci* 131(1):271-278

Nelms MD, Cronin MTD, Schultz TW, and Enoch SJ (2013) Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones. *SAR QSAR Environ Res* 24:695-709

Nelms MD, Ates G, Madden JC, Vinken M, Cronin MTD, Rogiers V, Enoch SJ (2014) Proposal of an *in silico* profiler for categorisation of repeat dose toxicity data of hair dyes. *Arch Toxicol*. DOI: 10.1007/s00204-014-1277-8. Available at: http://link.springer.com/article/10.1007%2Fs00204-014-1277-8 (accessed 17.11.2014)

Nendza M, and Muller M (2007) Discriminating toxicant classes by mode of action: 3. Substructure indicators. *SAR QSAR Environ Res* 18: 155-168

Nikam A, Patankar JV, Lackner C, et al. (2013) Tansition between acute and chronic hepatotoxicity in mice is associated with impaired energy metabolism and induction of mitochondrial heam oxygenase-1. *PLoS ONE* 8:e66094

Nohynek GJ, Antignac E, Re T, and Toutain H (2010) Safety assessment of personal care products/cosmetics and their ingredients. *Toxicol Appl Pharm* 234:239-259

OECD (1992) OECD Guidelines for Skin Sensitisation: Guinea Pig Maximisation Test OECD 406. Available at: http://www.oecd-ilibrary.org/environment/test-no-406-skin-sensitisation_9789264070660-en (accessed 16.11.2014)

OECD (1995) OECD Guidelines for Repeated Dose 28-day Oral Toxicity Study in Rodents OECD 407. Available at: http://www.oecd-ilibrary.org/environment/test-no-407-repeated-dose-28-day-oral-toxicity-study-in-rodents_9789264070684-en (accessed 16.11.2014)

OECD (1997) OECD Guideline for testing of chemicals: Bacterial Reverse Mutation Test. OECD 471. Available at: http://www.oecd-ilibrary.org/environment/test-no-471-bacterial-reverse-mutation-test_9789264071247-en (accessed 17.11.2014)

OECD (1998) OECD Guidelines for Repeated Dose 90-day Oral Toxicity Study in Rodents. OECD 408. Available at: http://www.oecd-ilibrary.org/environment/test-no-408-repeated-dose-90-day-oral-toxicity-study-in-rodents_9789264070707-en (accessed 17.11.2014)

OECD (2002). OECD Guidelines for Skin Sensitisation: Local Lymph Node Assay. OECD 429. Available at: http://www.oecd-ilibrary.org/environment/test-no-429-skin-sensitisation_9789264071100-en (accessed 17.11.2014)

OECD (2008). Draft guideline for rodent carcinogenicity tests. OECD 451. Available at: *www.oecd.org/chemicalsafety/testing/41753121.pdf* (accessed 17.11.2014)

OECD (2011). Report of the workshop on using mechanistic information on forming chemical categories. Series on testing and assessment No. 138. ENV/JM/MONO(2011)8. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2011)8&doclanguage=en (accessed 17.11.2014)

OECD (2012). The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins part I: Scientific evidence. Series on testing and assessment No. 168. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en (accessed 17.11.2014)

OECD (2013). Guidance on developing and assessing Adverse Outcome Pathways. Series on testing and assessment ENV/JM/MONO(2013)6. Available at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2013)6&doclanguage=en (accessed 17.11.2014)

Ohkuma Y, Hiraku Y, and Kawanishi S (2001) Sequence-specific DNA damage induced by carcinogenic danthron and anthraquinone in the presence of Cu(II), cytochrome P450 reductase and NADPH. *Free Rad Res* 34:595-604

Palmeira CM, and Rolo AP (2004) Mitochondrially-mediated toxicity of bile acids. *Toxicology* 203:1-15

Patlewicz G, Basketter DA, Smith CK, et al (2001) Skin-sensitisation structure-activity relationships for aldehydes. *Contact Dermatitis* 44: 331 - 336

Patlewicz G, Dimitrov SD, Low LK, et al. (2007) TIMES-SS - A promising tool for the assessment of skin sensitisation hazard. A characterisation with respect to the OECD validation principles for (Q)SARs and an external evaluation for predicitivity. *Reg Toxicol Pharm* 48(2): 225 - 239

Perez MJ (2009) Bile-acid-induced cell injury and protection. *World J Gastroentero* 15:1677

Pessayre D, Berson A, and Fromenty B (2008) Features and mechanisms of drug-induced liver injury. In: Dykens JA, and Will, Y (ed) *Drug-Induced Mitochondrial Dysfunction*. John Wiley and Sons, Inc., Hoboken, New Jersey, USA

Pessayre D, Fromenty B, Berson A, et al. (2012) Central role of mitochondria in drug-induced liver injury. *Drug Metab Rev* 44:34-87

Pieczenik SR, and Neustadt J (2007) Mitochondrial dysfunction and molecular pathways of disease. *Exp Mol Pathol* 83:84-92

Porceddu M, Buron N, Roussel C, et al (2012) Prediction of liver injury induced by chemicals in human with a multiparametric assay on isolated mouse liver mitochondria. *Toxicol Sci* 129:332-345

Przybylak KR, Madden JC, Cronin MTD, and Hewitt M (2012) Assessing toxicological data quality: basic principles, existing schemes and current limitations. *SAR QSAR Environ Res* 23: 435-459

Przybylak KR, Schultz TW (2013) Informing Chemical Categories through the Development of Adverse Outcome Pathways. In: Cronin MTD, Madden JC, Enoch SJ, Roberts DW (eds) *Chemical Toxicity Prediction: Category Formation and Read-Across.* vol 17. RSC, Cambridge, UK, p 44-71

Qiao L, Studer E, Leach K, McKinstry R, et al (2001) Deoxycholic Acid (DCA) causes ligand-independent activation of Epidermal Growth Factor Receptor (EGFR) and Fas receptor in primary hepatocytes: Inhibition of EGFR/Mitogen-activated protein kinase-signalling module enhances DCA-induced apoptosis. *Mol Biol Cell* 12:2609-2645

Richard AM, Yang C, and Judson RS (2008) Toxicity data informatics: Supporting a new paradigm for toxicity prediction. *Toxicol Mech Meth* 18: 103-118

Richarz AN, Schultz TW, Cronin MTD, and Enoch SJ (2013) Experimental verification of structural alerts for the protein binding of sulfur-containing compounds. *SAR QSAR Environ Res* 25(4): 325-341

Roberts DW, Aptula AO, and Patlewicz G (2006) Mechanistic applicability domains for non-animal based predictions of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitisation. *Chem Res Toxicol* 19: 1228 - 1233

Rodrigues CMP, Ma X, Linehan-Stieers C, et al (1999) Ursodeoxycholic acid prevents cytochrome c release in apoptosis by inhibiting mitochondrial membrane depolarisation and channel formation. *Cell Death Differ* 6:842-854

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 QSAR Environ Res* 24(11):963-977

Roldan MD, Perez-Reinado E, Castillo F, and Moreno-Vivian C (2008) Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *Microbiol Rev* 32:474-500

Rolo AP, Palmeira CM, Holy JM, and Wallace KB (2004) Role of mitochondrial dysfunction in combined bile acid-induced cytotoxicity: The switch between apoptosis and necrosis. *Toxicol Sci* 79:196-204

Russell, W. M. S. (1959). *The Principles of Humane Experimental Technique*. London, Methuen and Co.

Saito K, Kakei M, Uchimura S, Kashima T, and Tanaka H (1982) Toxic effects of chlorpromazine on red and white muscles in rats: an ultrastructural study. *Toxicol Appl Pharm* 65:347-353

Sakuratani Y, Zhang HQ, Nishikawa S, et al. (2013a) Hazard Evaluation Support System (HESS) for predicting repeated dose toxicity using toxicological categories. *SAR QSAR Environ Res* 24:351-363

Sakuratani Y, Zhang HQ, Nishikawa S, et al. (2013b) Categorization of nitrobenzenes for repeated dose toxicity based on adverse outcome pathways. *SAR QSAR Environ Res* 24:35-46

Scatena R, Bottoni P, Botta G, Martorana GE, Giardina B (2007) The role of mitochondria in pharmacotoxicology: A re-evaluation of an old, newly emerging topic. *Am J Physiol Cell Ph* 293:C12-C21

SCCNFP (2003) Opinion of the Scientific Committee on Cosmetic products and Non-Food Products intended for consumers concerning HC Orange No. 3, *SCCNFP/0676/03*

SCCP (2005) Scientific Committee on Consumer Products Opinion On Acid Blue No.62. *SCCP/0878/05*

SCCP (2006) Scientific Committee on Consumer Products Opinion On m-aminophenol. *SCCP/0978/06*

Schnellmann RG (1990) The cellular effects of a unique pesticide sulfluramid (N-ethylperfluorooctane sulphonamide) on rabbit renal proximal tubules. *Toxicol in Vitro* 4:71-74

Schnellmann RG, and Manning RO (1990) Perfluorooctane sulfonamide: a structurally novel uncoupler of oxidative phosphorylation. *Biochem Biophys ACTA* 1016:344-348

Schonfeld P, Sztark F, Slimani M, Dabadie P, and Mazat JP (1992) Is bupivacaine a decoupler, a protonophore or a proton-leak-inducer. *FEBS Letters* 304:273-276

Schultz TW (1997). TERATOX: Tetrahymena Pyriformis Population Growth Impairment Endpoint - A Surrogate for Fish Lethality. *Toxicol Mech Meth* 7: 289-309

Schultz TW (2010). Adverse Outcome Pathways: A way of linking chemical structure to *in vivo* toxicological hazards. In: Cronin MTD, and Madden JC. *In silico Toxicology: Principles and Applications.* Cambridge, UK, Royal Society of Chemistry

Schultz TW, Carlson RE, Cronin MTD, et al (2006) A conceptual framework for predicting toxicity of reactive chemicals: Models for soft electrophilicity. *SAR QSAR Environ Res* 17(4): 413-428

Schultz TW, Hewitt M, Netzeva TI, and Cronin MTD (2007) Assessing applicability domains of toxicological QSARs: Definition, confidence in predicted values, and the role of mechanisms of action. *QSAR Comb Sci* 26(2): 238-254

Schultz TW, Sinks GD, and Cronin MTD (1996) Chapter 23: Identification of mechanisms of toxic action of phenols to Tetrahymena pyriformis from molecular descriptors. In: Chen F, and Schuurmann G. (eds) *Quantitative Structure-Activity Relationships in environmental sciences - VII*. Pensacola, Florida, U.S.A, Society of Environmental Toxicology and Chemisty

Schultz TW, Yarbrough JW, and Johnson EL (2005) Structure-activity relationships for te reactivity of carbonyl-containing with glutathione. *SAR QSAR Environ Res* 16: 313-322

Schwöbel J, Madden JC, and Cronin MTD (2010a) Examination of Michael addition reactivity towards glutathione by transition state calculations. *SAR QSAR Environ Res* 21: 693 - 710

Schwöbel J, Wondrousch D, Koleva YK, et al (2010b) Prediction of Michael type acceptor reactivity toward glutathione. *Chem Res Toxicol* 23: 1576 - 1585

Schwöbel JAH, Koleva YK, Bajot F, et al. (2011) Measurement and estimation of electrophilic reactivity for predictive toxicology. *Chem Rev* 111: 2562-2596

Siraki AG, Chevaldina T, and O'Brien PJ (2005) Application of quantitative structure-toxicity relationships for acute NSAID cytotoxicity in rat hepatocytes. *Chem-Biol Interact* 151:177-191

Skipper PL, Kim MY, Patty Sun H.-L, Wogan GN, Tannenbaum SR (2010) Monocyclic aromatic amines as potential human carcinogens: old is new again. *Carcinogenesis*. 31(1):50-58

Sokol RJ, Straka MS, Dahl R, Devereaux MW, et al (2001) Role of oxidant stress in the permeability transition induced in rat hepatic mitochondria by hydrophobic bile acids. *Pediatr Res* 49:519-531

Spivey JR, Bronk SF, and Gores GJ (1993) Glycochenodeoxycholate-induced lethal hepatocellular injury in rat hepatocytes. Role of ATP depletion and cytosolic free calcium. *J Clinical Invest* 92:17-24

Spycher S, Smejtek P, Netzeva TI, and Escher BI (2008) Towards a class independent Quantitative Structure-Activity Relationship model for uncouplers of oxidative phosphorylation. *Chem Res Toxicol* 21:911-927

Starkov AA, and Wallace KB (2002) Structural determinants of fluorochemical-induced mitochondrial dysfunction. *Toxicol Sci* 66:244-252

Sun X, and Garlid KD (1992) On the mechanism by which bupivacaine conducts protons across the membranes of mitochondria and liposomes. *J Biol Chem* 267:19147-19154

Sztark F, Ouhabi R, Dabadie P, and Mazat JP (1997) Effects of the local aesthetic bupivacaine on mitochondria energy metabolism, change from uncoupling to decoupling depending on the respiration state. *Biochem Mol Biol Int* 43:997-1003

Taylor K, Andrew DJ, Rego L (2014) The added value of the 90-day repeated dose oral toxicity test for industrial chemicals with a low (sub)acute toxicity profile in a high quality dataset. *Reg Toxicol Pharm* 69: 320-332

Taylor RC, Cullen SP, and Martin SJ (2008) Apoptosis: controlled demolition at the cellular level. *Nat Rev Mol Cell Biol* 9:231-241

Terada H (1990) Uncouplers of oxidative phosphorylation. *Environ Health Persp* 87:213-218

Terada H, Shima O, Yoshida K, and Shinohara Y (1990) Effects of the local anaesthetic bupivacaine on oxidative phosphorylation in mitochondria. Change from decoupling to uncoupling by formation of a leakage type ion pathway specific from H+ in cooperation with hydrophobic anions. *J Biol Chem* 265:7837-7842

Uittenbogaard JP, Zomer B, Hoogerhout P, and Metz B (2011) Reactions of beta-propiolactone with nucleobase analogues, nucleosides, and peptides. *J Biol Chem* 286: 36198 - 36214

US EPA http://www.epa.gov/risk_assessment/glossary.htm#n

Uyemura SA, Santos AC, Mingatto FE, Jordani MC, and Curti C (1997) Diclofenac sodium and mefenamic acid, potent inducers of the membrane permeability transition in renal cortex mitochondria. *Arch Biochem Biophys* 342:231-235

Valerio Jr. L G (2009) *In silico* toxicology for the pharmaceutical sciences. *Toxicol Appl Pharm* 241: 356-370

van Leeuwen CJ (2007) Toxicity testing for human health risk assessment. In: van Leeuwan CJ, and Vermeire TG (eds) *Risk assessment of chemicals, An Introduction.* Springer: 227-274

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

Vinken M, Landesmann B, Goumenou M, Vinken S, et al (2013a) Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. *Toxicol Sci* 136:97-106

Vinken M, Maes M, Vanhaecke T, and Rogiers V (2013b) Drug-induced liver injury: mechanisms, types and biomarkers. *Curr Med Chem* 20:3011-3021

Vinken M, Whelan M, and Rogiers V (2014) Adverse outcome pathways: hype or hope? *Arch Toxicol* 88:1-2

von der Ohe PC, Kuhne R, Ebert RU, et al (2005) Structural alerts: a classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay. *Chem Res Toxicol* 18: 536-555

Wallace KB (2003) Doxorubicin-induced cardiac mitochondrionopathy. *Pharmacol Toxicol* 93:105-115

Wallace KB, and Starkov AA (2000) Mitochondrial targets of drug toxicity. A*nn Rev Pharm Toxicol* 40:353-388

Wallace KB, Kissling GE, Melnick RL, Blystone CR (2013) Structure–activity relationships for perfluoroalkane-induced *in vitro* interference with rat liver mitochondrial respiration. *Toxicol Lett* 222:257-264

Wei Y-H, Lin TN, Hong CY, and Chiang B (1985) Inhibition of the mitochondrial Mg2+-ATPase by propranolol. *Biochem Pharmacol* 34:911-917

Winklhofer-Roob BM, McKim JM, Jr, Devereaux MW, and Sokol RJ (1996) Characterization of site of reactive oxygen species (ROS) generation in mitochondria exposed to glycochenodeoxycholic acid [Abstract]. *Hepatology* 24(part2):338A

Woodward KN, and Timbrell JA (1984). Acetylhydrazine hepatotoxicity: The role of covalent binding. *Toxicology* 30: 65 - 74

Yang C, Arvidson K, Detroyer A, et al (2013). Chemotypes: A new structure representation standard for incorporating atom/bond properties into structural alerts for toxicity effects and mechanisms. *The Toxicologist, Supplement to Toxicological Sciences* 132(1): 183

Yang C, Benz RD, and Cheeseman MA (2006). Landscape of current toxicity databases and database standards. *Curr Opin Drug Di De* 9(1): 124-133

Yerushalmi B, Dahl R, Devereaux MW, Gumpricht E, and Sokol RJ (2001) Bile acid-induced rat hepatocyte apoptosis is inhibited by antioxidants and blockers of the mitochondrial permeability transition. *Hepatology* 33:616-626

Yin X-M, and Ding WX (2003) Death receptor activation-induced hepatocyte apoptosis and liver injury. *Curr Mol Med* 3:491-508

Zhang H, Chen QY, Xiang ML, Ma CY, et al (2009) *In silico* prediction of mitochondrial toxicity by using GA-CG-SVM approach. *Toxicol in Vitro* 23:134-140

9: Appendices

Appendix I. List of the 54 chemicals in the COSMOS DB with short-term and sub-chronic toxicity values (Chapter 2).

Appendix II. Data set extracted from Zhang et al (2009) used to develop the chemical categories and subsequent mechanism-based structural alerts (Chapters 3 and 4)

Appendix III. 90-day repeat dose toxicity data associated to each compound within the categories developed (Chapter 5)

Appendix IV. Copies of published work

Appendix I. List of the 54 chemicals in the COSMOS DB with short-term and sub-chronic toxicity values (Chapter 2). (Abbreviations: HNEL – Highest No Effect Level, LEL – Lowest Effect Level.)

Chemical name	COSMOS ID	Short-term toxicity study value (mg/kg bw/day)	Length of short-term toxicity study (days)	Sub-chronic toxicity study value (mg/kg bw/day)	Length of sub-chronic toxicity study (days)
FD&C Yellow Nº. 5	CMS-1476	LEL 15	30	HNEL 2500 LEL 5000	91 91
Vitamin A Acetate	CMS-1255	HNEL 34 HNEL 40 LEL 51 LEL 80 LEL 20 LEL 250 HNEL 16	28 28 28 28 30 30 35	HNEL 44 LEL 88	90 90
Ethylene Glycol Monobutyl Ether	CMS-3639	HNEL 500 LEL 1000	35 (MOUSE) 35 (MOUSE)	LEL 225 LOAEL 615 HNEL 553 LEL 676	90 (MOUSE) 90 (MOUSE) 91 (MOUSE) 91 (MOUSE)
DL-Phenylalanine	CMS-3235	HNEL 8000 LEL 5000	28 28	LEL 5000	84
Hydroxypropyl Methylcellulose	CMS-7567	LEL 3846 HNEL 10,000 LEL 25,000	30 (DOG) 30 30	HNEL 1250 HNEL 2004 HNEL 2257 LEL 8364 HNEL 2339 HNEL 5000 HNEL 6500 HNEL 7700 LEL 8209 HNEL 505 LEL 1020	90 (DOG) 90 (DOG) 84 84 90 90 90 90 90 91 91

Glycyrrhizin, ammoniated	CMS-8524	HNEL 1000 HNEL 1000 LEL 2000	30 35 35	HNEL 500	90
Resorcinol	CMS-1253	HNEL 260.7	28	LOAEL (♂) 32 LOAEL (♀) 65 HNEL 260 LEL 520	90 90 91 91
o-Cresol	CMS-1797	HNEL 280 LEL 763 HNEL 266 LEL 861	28 (MOUSE) 28 (MOUSE) 28 28	HNEL 794 (MOUSE) LEL 1500 (MOUSE) HNEL 247 LEL 510	90 90 90 90
Benzoic acid	CMS-147	LEL 1100 LEL 3000	35 35	HNEL 4000	90
Monosodium glutamate	CMS-927	LEL 16700	35	LEL 6000	90
Allyl isothiocyanate	CMS-49	HNEL 25 LEL 25 LEL 75 HNEL 300 HNEL 0.5	28 28 28 30 30	HNEL 25	91
Xylitol	CMS-9788	LEL 12100	28	LEL 5000	91
Ethylene oxide/propylene oxide copolymer	CMS-13925	HNEL 300 LEL 0.5 LEL 480	30 30 30	HNEL 40 HNEL 500 HNEL 1000 HNEL 1000 HNEL 1000 LEL 200 LEL 5000	90 90 90 90 90 90 90
Ferrous lactate	CMS-13965	LEL 625 LEL 5000	28 28	LEL 313	90
L-Phenylalanine	CMS-8300	HNEL 5000 LEL 7000	28 28	LEL 5000	84

Furfural	CMS-663	HNEL 96 LEL 120 HNEL 2.5 LEL 25	28 28 35 35	LOAEL (♂) 11 HNEL 90 LEL 180	90 91 91
Butylated hydroxyanisole	CMS-223	HNEL 50 LEL 150 LEL 2000 LEL 2000 LEL 2000 LEL 1000 LEL 1000	28 28 28 28 28 32 35	HNEL 125 HNEL 2000 LEL 125 LEL 1000 LEL 2000 LEL 2000 HNEL 500 LEL 2000	90 90 90 90 90 90 91 91
L-Aspartic acid	CMS-2493	HNEL 50 HNEL 1000 HNEL 1000	28 28 28	HNEL 715 LEL 1417	90 90
Polymaleic acid	CMS-14369	HNEL 100	28	HNEL 400	90
Annatto extract (Bixa Orellana)	CMS-3887	LEL 2000	28	HNEL 1000 HNEL 69 LEL 204	90 91 91
Sucrose acetate isobutyrate	CMS-5115	HNEL 2226 HNEL 2592	28 28	HNEL 5300	91
Propylene glycol	CMS-1220	LEL 1600 LEL 1600 LEL 1600	35 (CAT) 35 (CAT) 35 (CAT)	LEL 2750	91 (CAT)
Choline chloride	CMS-329	LEL 1000	25	HNEL 500 LEL 1350	90 90
Maltodextrin	CMS-5576	HNEL 10,000	30	HNEL 3882	90
p-Cresol	CMS-1848	LEL 2000	28	HNEL 50 LEL 175 LOAEL 175	91 91 91
Butyl acetate	CMS-1941	HNEL 2000	28	HNEL 600	90

		LEL 800	30	LEL 2000	90
Sucralose	CMS-7967	LEL 4000 HNEL 1000 HNEL 2000 LEL 2500 LEL 5000	26 28 28 28 28	LEL 2000	89
Hydroxypropyl cellulose	CMS-10327	HNEL 6000	30	HNEL 5000	90
Agar (Gelidium spp.)	CMS-13069	LEL 25,000	28	HNEL 2500	90
Vitamin A	CMS-3312	LEL 30	28	HNEL 0.24 LEL 0.48	90 90
Vitamin A Palmitate	CMS-1256	LEL 60	26	LEL 33	91
Potassium bicarbonate	CMS-1189	HNEL 2132 HNEL 4000 LEL 4385	28 28 28	HNEL 1482 HNEL 2000 LEL 2500 LEL 3133 LEL 4300	91 91 91 91 91
Canthaxanthin	CMS-2585	HNEL 313	30	HNEL 50	90
Niacin	CMS-952	LEL 1000	30	HNEL 19 LEL 49	90 90
Cellulose, methyl	CMS-7542	HNEL 50 HNEL 5000	28 30	HNEL 5000 LEL 25,000 LEL 5000	90 90 95
Polyethylene, oxidised	CMS-34680	HNEL 4650	32	HNEL 5000 HNEL 5000	90 90
Tributyl acetylcitrate	CMS-4878	LEL 1780	28	HNEL 100	90
Ascorbic acid	CMS-108	HNEL 625 HNEL 10,000 LEL 27300	28 (MOUSE) 28 28	HNEL 3750 LEL 7500 HNEL 2500 LEL 2500 LEL 5000	91 (MOUSE) 91 (MOUSE) 90 90 90
Allyl Heptanoate	CMS-10786	LOAEL 100	28	LOAEL 49.6	90

4-pyrimidinol, 2,5,6-triamino-, monosulfate	CMS-46638	LOAEL 1000	28	LOAEL 1000	90
Formaldehyde	CMS-653	HNEL 25 LEL 20 LEL 125	28 28 28	HNEL 50 HNEL 100 LEL 100 LEL 150	90 90 90 90
β-Carotene	CMS-261	LEL 1000 HNEL 5 HNEL 5 HNEL 80	28 30 30 35	HNEL 3127 LEL 1000	90 91
Adipic acid	CMS-1627	HNEL 440 HNEL 2400 HNEL 3400	28 28 28	HNEL 1700 LEL 8000	90 90
3-ethyl-2-hydorxy-2-cyclopenten-1-one	CMS-10298	HNEL 800	30	LEL 100	90
Quinine hydrochloride	CMS-10282	HNEL 250 LEL 25	28 28	HNEL 10 HNEL 40 HNEL 60 LEL 40 LEL 85 LEL 85 LEL 100	90 90 90 90 90 90 90
Ethylenediamine	CMS-1857	HNEL 250	28	HNEL 250 LEL 1000 HNEL 100 LEL 200	90 90 91 91
D-Limonene	CMS-797	LEL 75 HNEL 100 LEL 25 LEL 200 LEL 400 LEL 277	27 28 28 28 28 30	LEL 150 LOAEL (♂) 30 LOAEL (♂) 150 LOAEL (♀) 2400 HNEL 30 LEL 75	89 90 90 90 91 91

				LEL 150	91
Phenol	CMS-1136	LEL 1.8 HNEL 86.2	28 (MOUSE) 28	HNEL 750 LEL 2500 HNEL 300 LEL 1000	91 (MOUSE) 91 (MOUSE) 91 91
Caffeine	CMS-240	LEL 40 HNEL 100 LEL 50 LEL 500	28 (MOUSE) 28 28 28	HNEL 80 LEL 122 LOAEL 87 HNEL 163 LEL 279 LOAEL 280	90 (MOUSE) 90 (MOUSE) 90 (MOUSE) 90 90 90
p-Ethylphenol	CMS-1937	HNEL 100 LEL 300	28 28	HNEL 342	90
Ethylene Glycol Monoethyl Ether	CMS-3637	LEL 150	28	HNEL 740 LEL 1890 HNEL 92.9 LEL 122 LEL 185.8	90 90 91 91 91
Butylated hydroxytoluene	CMS-224	LEL 300 HNEL 25 HNEL 250 HNEL 1000 LEL 250 LEL 500 LEL 700 LEL 1500 LE 100	28 (DOG) 28 28 28 28 28 28 28 30	HNEL 2000 HNEL 200 LEL 25 LEL 200 LEL 500	90 (DOG) 90 90 90 90
Isoeugenyl Methyl Ether	CMS-15322	HNEL 91 LEL 264	31 31	HNEL 6	91
Sodium benzoate	CMS-144	LEL 2000	28	HNEL 8000 LOAEL 6290 LEL 1250	90 90 91

Appendix II. Data set extracted from Zhang et al (2009) used to develop the chemical categories and subsequent mechanism-based structural alerts (Chapters 3 and 4)

(+ve - mitochondrial toxicant, -ve - non-mitochondrial toxicant)

Chemical ID	Chemical Name	Toxicity	SMILES
1	Menadione	+ve	O=C\2c1c(cccc1)C(=O)/C(=C/2)C
2	Pentamidine	+ve	[H]/N=C(\N)/c1ccc(cc1)OCCCCCOc2ccc(cc2)/C(=N/[H])/N
3	Nalidixic acid	+ve	CCn1cc(c(=O)c2c1nc(cc2)C)C(=O)O
4	Diazepam	+ve	CN1c2ccc(cc2C(=NCC1=O)c3ccccc3)Cl
5	Clofibric acid	+ve	Clc1ccc(OC(C(=O)O)(C)C)cc1
6	Niclofolan	+ve	Clc2cc(c(O)c(c1c(O)c([N+]([O-])=O)cc(Cl)c1)c2)[N+]([O-])=O
7	Lonidamine	+ve	Clc1ccc(c(Cl)c1)Cn3nc(c2ccccc23)C(=O)O
8	Sulofenur	+ve	Clc1ccc(cc1)NC(=O)NS(=O)(=O)c2ccc3c(c2)CCC3
9	Ciprofloxacin	+ve	c1c2c(cc(c1F)N3CCNCC3)n(cc(c2=O)C(=O)O)C4CC4
10	Tolcapone	+ve	[O-][N+](=O)c2cc(C(=O)c1ccc(cc1)C)cc(O)c2O
11	Hexachlorophene	+ve	Clc1c(c(O)c(Cl)cc1Cl)Cc2c(O)c(Cl)cc(Cl)c2Cl
12	Buquinolate	+ve	CCOC(=O)c1cnc2cc(c(cc2c1O)OCC(C)C)OCC(C)C
13	Menoctone	+ve	O=C1c3ccccc3C(\O)=C(/C1=O)CCCCCCCCCC2CCCCC2
14	Amquinate	+ve	O=C\2c1c(cc(c(c1)CCC)N(CC)CC)N/C=C/2C(=O)OC
15	Decoquinate	+ve	O=C\2c1c(cc(OCC)c(OCCCCCCCCCC)c1)N/C=C/2C(=O)OCC
16	Tioxaprofen	+ve	Clc3ccc(c2oc(SC(C(=O)O)C)nc2c1ccc(Cl)cc1)cc3

17	Cyhalothrin	+ve	CC1([C@@H]([C@H]1C(=O)OC(C#N)c2cccc(c2)Oc3ccccc3)/C=C(\C(F)(F)F)/Cl)C
18	Myxothiazol	+ve	O=C(N)\C=C(\OC)[C@H](C)[C@@H](OC)/C=C/c1nc(sc1)c2nc(sc2)[C@H](\C=C\C=C\C(C)C)C
19	Haloperidol	+ve	c1cc(ccc1C(=O)CCCN2CCC(CC2)(c3ccc(cc3)Cl)O)F
20	Indomethacin	+ve	Cc1c(c2cc(ccc2n1C(=O)c3ccc(cc3)Cl)OC)CC(=O)O
21	Phenytoin	+ve	c1ccc(cc1)C2(C(=O)NC(=O)N2)c3ccccc3
22	Lindane	+ve	Cl[C@H]1[C@H](Cl)[C@@H](Cl)[C@@H](Cl)[C@H](Cl)[C@H]1Cl
23	Procaine	+ve	O=C(OCCN(CC)CC)c1ccc(N)cc1
24	Methionine	+ve	CSCCC(C(=O)O)N
25	Fluphenazine	+ve	c1ccc2c(c1)N(c3cc(ccc3S2)C(F)(F)F)CCCN4CCN(CC4)CCO
26	Salicylic acid	+ve	c1ccc(c(c1)C(=O)O)O
27	Valproic acid	+ve	O=C(O)C(CCC)CCC
28	Phenformin	+ve	N(=C(/N=C(\N)N)N)\CCc1ccccc1
29	Lidocaine	+ve	CCN(CC)CC(=O)Nc1c(cccc1C)C
30	Butacaine	+ve	O=C(OCCCN(CCCC)CCCC)c1ccc(N)cc1
31	Metformin	+ve	[H]/N=C(/N=C(N)N)\N(C)C
32	Buformin	+ve	N(=C(/N)N)\C(=N\CCCC)N
33	Porfiromycin	+ve	O=C\1C/2=C(\C(=O)/C(N)=C/1C)[C@H]([C@]3(OC)N\2C[C@@H]4N([C@H]34)C)COC(=O)N
34	Anthralin	+ve	O=C2c1c(O)cccc1Cc3c2c(O)ccc3
35	Propofol	+ve	Oc1c(cccc1C(C)C)C(C)C
36	Flutamide	+ve	O=C(Nc1cc(c(cc1)[N+]([O-])=O)C(F)(F)F)C(C)C

37	Ibuprofen	+ve	CC(C)Cc1ccc(cc1)C(C)C(=O)O
38	Zidovudine	+ve	Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)N=[N+]=[N-]
39	Ropivacaine	+ve	O=C(Nc1c(cccc1C)C)[C@H]2N(CCC)CCCC2
40	Etidocaine	+ve	O=C(Nc1c(cccc1C)C)C(N(CC)CCC)CC
41	Nabumetone	+ve	O=C(C)CCc1ccc2c(c1)ccc(OC)c2
42	Nimesulide	+ve	[O-][N+](=O)c2cc(Oc1ccccc1)c(cc2)NS(=O)(=O)C
43	Ciprofibrate	+ve	CC(C)(C(=O)O)Oc1ccc(cc1)C2CC2(Cl)Cl
44	Fluoxetine	+ve	CNCCC(c1ccccc1)Oc2ccc(cc2)C(F)(F)F
45	Idarubicin	+ve	O=C2c1c(O)c5c(c(O)c1C(=O)c3ccccc23)C[C@@](O)(C(=O)C)C[C@@H]5O[C@@H]4O[C@H]([C@@H](O)[C@@H](N)C4)C
46	Nicorandil	+ve	c1cc(cnc1)C(=O)NCCO[N+](=O)[O-]
47	Carvedilol	+ve	COc1ccccc1OCCNCC(COc2cccc3c2c4ccccc4[nH]3)O
48	Nefazodone	+ve	Clc4cccc(N3CCN(CCCN1/N=C(\N(C1=O)CCOc2ccccc2)CC)CC3)c4
49	Didanosine	+ve	c1nc2c(c(n1)O)ncn2[C@H]3CC[C@H](O3)CO
50	Stavudine	+ve	O=C1/C(=C\N(C(=O)N1)[C@@H]/2O[C@@H](\C=C\2)CO)C
51	Nevirapine	+ve	Cc1ccnc2c1NC(=O)c3cccnc3N2C4CC4
52	Entacapone	+ve	[O-][N+](=O)c1cc(\C=C(/C#N)C(=O)N(CC)CC)cc(O)c1O
53	Lamivudine	+ve	c1cn(c(=O)nc1N)[C@@H]2CS[C@@H](O2)CO
54	Diclofenac	+ve	c1ccc(c(c1)CC(=O)O)Nc2c(cccc2Cl)Cl

55	Chloramphenicol	+ve	c1cc(ccc1[C@H]([C@@H](CO)NC(=O)C(Cl)Cl)O)[N+](=O)[O-]
56	Minocycline	+ve	CN(C)c1ccc(c2c1C[C@H]3C[C@H]4[C@@H](C(=C(C(=O)[C@]4(C(=C3C2=O)O)O)C(=O)N)O)N(C)C)O
57	Naproxen	+ve	C[C@@H](c1ccc2cc(ccc2c1)OC)C(=O)O
58	Butylated Hydroxy	+ve	
59	Linezolid	+ve	O=C1O[C@@H](CNC(=O)C)CN1c3cc(F)c(N2CCOCC2)cc3
60	Celecoxib	+ve	O=S(=O)(c3ccc(n1nc(cc1c2ccc(cc2)C)C(F)(F)F)cc3)N
61	Efavirenz	+ve	FC(F)(F)[C@@]3(C#CC1CC1)OC(=O)Nc2ccc(Cl)cc23
62	Rosiglitazone	+ve	O=C1NC(=O)SC1Cc3ccc(OCCN(c2ncccc2)C)cc3
63	Chlorpromazine	+ve	CN(C)CCCN1c2ccccc2Sc3c1cc(cc3)Cl
64	Mefenamic acid	+ve	O=C(O)c2c(Nc1cccc(c1C)C)cccc2
65	Flufenamic acid	+ve	FC(F)(F)c1cc(ccc1)Nc2ccccc2C(=O)O
66	Tolfenamic acid	+ve	Clc2cccc(Nc1ccccc1C(=O)O)c2C
67	Fenofibrate	+ve	O=C(c1ccc(Cl)cc1)c2ccc(OC(C(=O)OC(C)C)(C)C)cc2
68	Atovaquone	+ve	c1cc2c(cc1)C(=O)C(=C(C2=O)[C@H]3CC[C@@H](CC3)c4ccc(cc4)Cl)O
69	Troglitazone	+ve	O=C1NC(=O)SC1Cc4ccc(OCC3(Oc2c(c(c(O)c(c2CC3)C)C)C)C)cc4
70	Tamoxifen	+ve	O(c1ccc(cc1)/C(c2ccccc2)=C(\c3ccccc3)CC)CCN(C)C
71	Tetracycline	+ve	C[C@]1(c2cccc(c2C(=O)C3=C([C@]4([C@@H](C[C@@H]31)[C@@H](C(=C(C4=O)C(=O)N)O)N(C)C)O)O)O)O
72	Dieldrin	+ve	ClC5(Cl)[C@]3(Cl)C(\Cl)=C(\Cl)[C@@]5(Cl)[C@H]4[C@H]1C[C@H]([C@@H]2O[C@H]12)[C@@H]34

73	Sulindac	+ve	O=S(c1ccc(cc1)\C=C3/c2ccc(F)cc2\C(=C3C)CC(=O)O)C
74	Amiodarone	+ve	Ic1cc(cc(I)c1OCCN(CC)CC)C(=O)c2c3ccccc3oc2CCCC
75	Saquinavir	+ve	O=C(N)C[C@H](NC(=O)c1nc2c(cc1)cccc2)C(=O)N[C@@H](Cc3ccccc3)[C@H](O)CN5[C@H](C(=O)NC(C)(C)C)C[C@@H]4CCCC[C@@H]4C5
76	Daunorubicin	+ve	C[C@H]1[C@H]([C@H](C[C@@H](O1)O[C@H]2C[C@@](Cc3c2c(c4c(c3O)C(=O)c5cccc(c5C4=O)OC)O)(C(=O)CO)O)N)O
77	Nelfinavir	+ve	O=C(c1cccc(O)c1C)N[C@@H](CSc2ccccc2)[C@H](O)CN4[C@H](C(=O)NC(C)(C)C)C[C@@H]3CCCC[C@@H]3C4
78	Doxorubicin	+ve	C[C@H]1[C@H]([C@H](C[C@@H](O1)O[C@H]2C[C@@](Cc3c2c(c4c(c3O)C(=O)c5cccc(c5C4=O)OC)O)(C(=O)CO)O)N)O
79	Epirubicin	+ve	O=C2c1c(O)c5c(c(O)c1C(=O)c3cccc(OC)c23)C[C@@](O)(C(=O)CO)C[C@@H]5O[C@@H]4O[C@H]([C@H](O)[C@@H](N)C4)C
80	Ritonavir	+ve	CC(C)c1nc(cs1)CN(C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc2ccccc2)C[C@@H]([C@H](Cc3ccccc3)NC(=O)OCc4cncs4)O
81	2,4-Dichlorophenoxylacetic acid	+ve	Clc1cc(Cl)ccc1OCC(=O)O
82	2-Methylhamine	+ve	O(C=2/C=C\C1=C3/C=C\N(\C(=C3\N=C1C=2)C)C)C
83	Aldicarb	+ve	O=C(O\N=C\C(SC)(C)C)NC
84	Betulinic acid	+ve	O=C(O)[C@@]54[C@@H]([C@@H]3[C@@]([C@]1([C@@H]([C@]2(C)[C@@H](CC1)C(C)(C)[C@@H](O)CC2)CC3)C)(C)CC4)[C@H](C(=C)\C)CC5
85	CCCP	+ve	Clc1cc(N\N=C(/C#N)C#N)ccc1

86	DAPI	+ve	[H]/N=C(/c1ccc(cc1)c2cc3ccc(cc3[nH]2)/C(=N/[H])/N)\N
87	DASPEI	+ve	c2(\C=C\c1[n](cccc1)CC)ccc(N(C)C)cc2
88	DDD	+ve	Oc4ccc3cc(SSc2ccc1c(ccc(O)c1)c2)ccc3c4
89	DDT	+ve	Clc1ccc(cc1)C(c2ccc(Cl)cc2)C(Cl)(Cl)Cl
90	Dequalinium	+ve	c12ccccc1c(cc([n+]2CCCCCCCCCC[n+]4c3ccccc3c(N)cc4C)C)N
91	Diethylstilbestrol	+ve	CC/C(=C(/CC)\c1ccc(cc1)O)/c2ccc(cc2)O
92	Ellipticine	+ve	Cc1c2ccncc2c(c3c1[nH]c4c3cccc4)C
93	Ethidium bromide	+ve	CC[n]1c2cc(ccc2c3ccc(cc3c1c4ccccc4)N)N
94	FCCP	+ve	FC(F)(F)Oc1ccc(cc1)N/N=C(\C#N)C#N
95	JC 1	+ve	Clc1cc2N(/C(N(c2cc1Cl)CC)=C\C=C\c4[n](c3cc(Cl)c(Cl)cc3n4CC)CC)CC
96	Methyltriphenyl-phosphonium	+ve	c1(ccccc1)[P+](c2ccccc2)(c3ccccc3)C
97	MPCU	+ve	Clc2ccc(NC(=O)NS(=O)(=O)c1ccc(cc1)C)cc2
98	Nitroxynil	+ve	Ic1cc(C#N)cc([N+]([O-])=O)c1O
99	Nonylacridine Orange	+ve	c3(ccc2cc1ccc(N(C)C)cc1[n+](c2c3)CCCCCCCCC)N(C)C
100	o-Phenanthroline	+ve	c1cc2ccc3cccnc3c2nc1
101	Perfluorodecanoic acid	+ve	FC(F)(C(F)(F)C(=O)O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
102	Perfluorooctane-sulfonamide	+ve	FC(F)(C(F)(F)S(=O)(=O)N)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
103	Perfluorooctanoic acid	+ve	FC(F)(C(F)(F)C(=O)O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F
104	PK 11195	+ve	Clc3ccccc3c2nc(cc1ccccc12)C(=O)N(CC)CC)C
105	Rhein	+ve	O=C2c1cccc(O)c1C(=O)c3c2cc(cc3O)C(

			=O)O
106	Rhodamine 6G	+ve	CCOC(=O)c4ccccc4C=1c3cc(C)c(cc3O/C/2=C/C(=N/CC)C(/C)=C\C=1\2)NCC
107	Rhodamine 123	+ve	O=C(OC)c4ccccc4C=1c3c(OC=2C=1\C=C/C(=[NH2])/C=2)cc(cc3)N
108	Rhodamine B	+ve	CCN(CC)c1ccc2c(c1)oc-3cc(=[N](CC)CC)ccc3c2c4ccccc4C(=O)O
109	TBTP	+ve	O=P(SCCCC)(SCCCC)SCCCC
110	Tebufenpyrad	+ve	Clc2c(nn(c2C(=O)NCc1ccc(cc1)C(C)(C)C)C)CC
111	Trichlorophen-Oxyacetic acid	+ve	Clc1cc(OCC(=O)O)c(Cl)cc1Cl
112	TTFB	+ve	FC(F)(F)c2nc1c(Cl)c(Cl)c(Cl)c(Cl)c1n2
113	Victoria blue B	+ve	C/[N](C)=C1\C=C/C(C=C1)=C(\c4ccc(Nc2ccccc2)c3ccccc34)c5ccc(cc5)N(C)C
114	Zalcitabine	+ve	O=C1/N=C(/N)\C=C/N1[C@@H]2O[C@@H](CC2)CO
115	2,4-Dinitrophenol	+ve	c1cc(c(cc1[N+](=O)[O-])[N+](=O)[O-])O
116	Amytal	+ve	O=C1NC(=O)NC(=O)C1(CCC(C)C)CC
117	Antimycin A	+ve	O=CNc1cccc(c1O)C(=O)N[C@@H]2C(=O)O[C@H]([C@H](OC(=O)CC(C)C)[C@H](C(=O)O[C@@H]2C)CCCCCC)C
118	Azidothymidine	+ve	Cc1cn(c(=O)[nH]c1=O)[C@H]2C[C@@H]([C@H](O2)CO)N=[N+]=[N-]
119	Carbaryl	+ve	O=C(Oc2cccc1ccccc12)NC
120	Chenodeoxycholic acid	+ve	C[C@H](CCC(=O)O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)C
121	Cholic acid	+ve	C[C@H](CCC(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C

122	Deguelin	+ve	O=C2c5c(O[C@@H]3COc1cc(OC)c(OC)cc1[C@H]23)c4\C=C/C(Oc4cc5)(C)C
123	Deoxycholic acid	+ve	C[C@H](CCC(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2CC[C@H]4[C@@]3(CC[C@H](C4)O)C)O)C
124	Diazinon	+ve	S=P(OCC)(OCC)Oc1nc(nc(c1)C)C(C)C
125	Diphenylamine	+ve	c1ccc(cc1)Nc2ccccc2
126	Glycocholic	+ve	C[C@H](CCC(=O)NCC(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C
127	Lithocholic	+ve	O=C(O)CC[C@H]([C@H]1CC[C@@H]2[C@]1(C)CC[C@H]4[C@H]2CC[C@@H]3C[C@H](O)CC[C@@]34C)C
128	Mitomycin C	+ve	CC1=C(C(=O)C2=C(C1=O)N3C[C@H]4[C@@H]([C@@]3([C@@H]2COC(=O)N)OC)N4)N
129	MPTP	+ve	c2c(/C1=C/CN(C)CC1)cccc2
130	Naproxen	+ve	C[C@@H](c1ccc2cc(ccc2c1)OC)C(=O)O
131	Paraquat	+ve	C[n+]1ccc(cc1)c2cc[n+](cc2)C
132	Pentachlorophenol	+ve	Clc1c(O)c(Cl)c(Cl)c(Cl)c1Cl
133	Perhexiline	+ve	N3C(CC(C1CCCCC1)C2CCCCC2)CCCC3
134	Rotenone	+ve	CC(=C)[C@H]1Cc2c(ccc3c2O[C@@H]4COc5cc(c(cc5[C@@H]4C3=O)OC)OC)O1
135	Ranolazine	+ve	O=C(Nc1c(cccc1C)C)CN3CCN(CC(O)COc2ccccc2OC)CC3
136	Chloroquine	+ve	Clc1cc2nccc(c2cc1)NC(C)CCCN(CC)CC
137	Atenolol	+ve	CC(C)NCC(COc1ccc(cc1)CC(=O)N)O
138	Amineptine	+ve	O=C(O)CCCCCCNC3c1ccccc1CCc2c3cccc2

139	Tianeptine	+ve	Clc1cc2c(cc1)C(c3c(N(C)S2(=O)=O)ccc c3)NCCCCCCC(=O)O
140	Quinidine	+ve	O(c4cc1c(nccc1[C@H](O)[C@@H]2N3 CC[C@@H](C2)[C@@H](/C=C)C3)cc4)C
141	Abacavir	+ve	n3c1c(ncn1[C@H]2/C=C\[C@@H](CO) C2)c(nc3N)NC4CC4
142	Cerivastatin	+ve	O=C(O)C[C@H](O)C[C@H](O)/C=C/c 1c(nc(c(c1c2ccc(F)cc2)COC)C(C)C)C(C)C
143	Cinnarizine	+ve	c1c(cccc1)C(c2ccccc2)N3CCN(CC3)C/C =C\c4ccccc4
144	Flunarizine	+ve	c1ccc(cc1)/C=C/CN2CCN(CC2)C(c3ccc (cc3)F)c4ccc(cc4)F
145	Ketoconazole	+ve	O=C(N5CCN(c4ccc(OC[C@@H]1O[C @](OC1)(c2ccc(Cl)cc2Cl)Cn3ccnc3)cc4) CC5)C
146	Bupivacaine	+ve	CCCCN1CCCCC1C(=O)Nc2c(cccc2C)C
147	Clofibrate	+ve	Clc1ccc(OC(C(=O)OCC)(C)C)cc1
148	Prilocaine	+ve	O=C(Nc1ccccc1C)C(NCCC)C
149	Gemfibrozil	+ve	Cc1ccc(c(c1)OCCCC(C)(C)C(=O)O)C
150	Fenoprofen	+ve	O=C(O)C(c2cc(Oc1ccccc1)ccc2)C
151	Risperidone	+ve	Cc1c(c(=O)n2c(n1)CCCC2)CCN3CCC(CC3)c4c5ccc(cc5on4)F
152	Amphetamine	+ve	CC(Cc1ccccc1)N
153	Deoxycholic acid	+ve	C[C@H](CCC(=O)O)[C@H]1CC[C@@ H]2[C@@]1([C@H](C[C@H]3[C@H]2 CC[C@H]4[C@@]3(CC[C@H](C4)O)C)O)C
154	Taurocholic	+ve	C[C@H](CCC(=O)NCCS(=O)(=O)O)[C @H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C @@]3(CC[C@H](C4)O)C)O)O)C
155	Phentolamine	-ve	Oc3cc(N(c1ccc(cc1)C)CC/2=N/CCN\2)c cc3

156	Dopamine	-ve	c1cc(c(cc1CCN)O)O
157	Guanethidine	-ve	N(=C(\N)N)\CCN1CCCCCCC1
158	Diphenhydramine	-ve	O(CCN(C)C)C(c1ccccc1)c2ccccc2
159	Promethazine	-ve	S2c1ccccc1N(c3c2cccc3)CC(N(C)C)C
160	Oxymetazoline	-ve	Oc1c(c(c(cc1C(C)(C)C)C)CC/2=N/CCN\2)C
161	Tropicamide	-ve	CCN(Cc1ccncc1)C(=O)C(CO)c2ccccc2
162	Warfarin	-ve	CC(=O)CC(c1ccccc1)c2c(c3ccccc3oc2=O)O
163	Phenindione	-ve	O=C2c1ccccc1C(=O)C2c3ccccc3
164	Trimeprazine	-ve	S2c1ccccc1N(c3c2cccc3)CC(C)CN(C)C
165	Tripelennamine	-ve	n1ccccc1N(CCN(C)C)Cc2ccccc2
166	Bromodiphen-Hydramine	-ve	Brc1ccc(cc1)C(OCCN(C)C)c2ccccc2
167	Trimethobenzamide	-ve	O=C(c1cc(OC)c(OC)c(OC)c1)NCc2ccc(OCCN(C)C)cc2
168	Diphenylpyraline	-ve	O(C(c1ccccc1)c2ccccc2)C3CCN(C)CC3
169	Benzphetamine	-ve	N(C)(Cc1ccccc1)[C@@H](C)Cc2ccccc2
170	Metoclopramide	-ve	Clc1cc(c(OC)cc1N)C(=O)NCCN(CC)CC
171	Methoxamine	-ve	O(c1ccc(OC)cc1C(O)C(N)C)C
172	Phenprocoumon	-ve	OC=1c3ccccc3OC(=O)C=1C(CC)c2ccccc2
173	Crotamiton	-ve	O=C(N(c1ccccc1C)CC)/C=C/C
174	Alprostadil	-ve	O=C1C[C@@H](O)[C@H](/C=C/[C@@H](O)CCCCC)[C@H]1CCCCCCC(=O)O
175	Diphenidol	-ve	OC(c1ccccc1)(c2ccccc2)CCCN3CCCCC3
176	Methdilazine	-ve	S2c1ccccc1N(c3c2cccc3)CC4CCN(C)C4
177	Disopyramide	-ve	O=C(N)C(c1ncccc1)(c2ccccc2)CCN(C(C)C)C(C)C

178	Azatadine	-ve	n4c3\C(=C1/CCN(C)CC1)c2ccccc2CCc3ccc4
179	Pentoxifylline	-ve	CC(=O)CCCCn1c(=O)c2c(ncn2C)n(c1=O)C
180	Dobutamine	-ve	Oc1ccc(cc1O)CCNC(C)CCc2ccc(O)cc2
181	Alprenolol	-ve	O(c1ccccc1C\C=C)CC(O)CNC(C)C
182	Naltrexone	-ve	O=C4[C@@H]5Oc1c2c(ccc1O)C[C@H]3N(CC[C@]25[C@@]3(O)CC4)CC6CC6
183	Prazosin	-ve	O=C(N3CCN(c2nc1cc(OC)c(OC)cc1c(n2)N)CC3)c4occc4
184	Minaprine	-ve	n2nc(NCCN1CCOCC1)c(cc2c3ccccc3)C
185	Guanfacine	-ve	Clc1cccc(Cl)c1CC(=O)\\N=C(/N)N
186	Tocainide	-ve	O=C(Nc1c(cccc1C)C)C(N)C
187	Nadolol	-ve	OC(CNC(C)(C)C)COc1cccc2c1C[C@H](O)[C@H](O)C2
188	Midodrine	-ve	O=C(NCC(O)c1cc(OC)ccc1OC)CN
189	Flecainide	-ve	FC(F)(F)COc2cc(C(=O)NCC1NCCCC1)c(OCC(F)(F)F)cc2
190	Fenoldopam	-ve	Clc1c3c(cc(O)c1O)C(c2ccc(O)cc2)CNCC3
191	Dapiprazole	-ve	n1nc(n2c1CCCC2)CCN4CCN(c3ccccc3C)CC4
192	Enalapril	-ve	O=C(O)[C@H]2N(C(=O)[C@@H](N[C@H](C(=O)OCC)CCc1ccccc1)C)CCC2
193	Lisinopril	-ve	c1ccc(cc1)CC[C@@H](C(=O)O)N[C@@H](CCCCN)C(=O)N2CCC[C@H]2C(=O)O
194	Quinapril	-ve	O=C(OCC)[C@@H](N[C@H](C(=O)N2[C@H](C(=O)O)Cc1c(cccc1)C2)C)CCc3ccccc3
195	Ondansetron	-ve	O=C3c2c1ccccc1n(c2CCC3Cn4ccnc4C)C

196	Ridogrel	-ve	FC(F)(F)c2cccc(C(=N\OCCCCC(=O)O)/c1cccnc1)c2
197	Benazepril	-ve	O=C(OCC)[C@@H](N[C@@H]2C(=O)N(c1ccccc1CC2)CC(=O)O)CCc3ccccc3
198	Ramipril	-ve	CCOC(=O)[C@H](CCc1ccccc1)N[C@@H](C)C(=O)N2[C@H]3CCC[C@H]3C[C@H]2C(=O)O
199	Trandolapril	-ve	O=C(OCC)[C@@H](N[C@H](C(=O)N1[C@H](C(=O)O)C[C@H]2CCCC[C@H]12)C)CCc3ccccc3
200	Granisetron	-ve	CN4[C@@H]1CCC[C@H]4C[C@H](C1)NC(=O)c3nn(C)c2ccccc23
201	Acamprosate	-ve	O=S(=O)(O)CCCNC(=O)C
202	Moexipril	-ve	O=C(OCC)[C@@H](N[C@H](C(=O)N2[C@H](C(=O)O)Cc1c(cc(OC)c(OC)c1)C2)C)CCc3ccccc3
203	Dofetilide	-ve	O=S(=O)(Nc1ccc(cc1)CCN(CCOc2ccc(cc2)NS(=O)(=O)C)C)C
204	Brimonidine	-ve	Brc2c1nccnc1ccc2N/C3=N/CCN3
205	Losartan	-ve	Clc1nc(n(c1CO)Cc4ccc(c2ccccc2c3nnnn3)cc4)CCCC
206	Arbutamine	-ve	Oc1ccc(cc1O)[C@@H](O)CNCCCCc2ccc(O)cc2
207	Valsartan	-ve	CCCCC(=O)N(Cc1ccc(cc1)c2ccccc2c3[nH]nnn3)[C@@H](C(C)C)C(=O)O
208	Atropine	-ve	CN3[C@H]1CC[C@@H]3C[C@@H](C1)OC(=O)C(CO)c2ccccc2
209	Levorphanol	-ve	Oc3ccc4C[C@H]1N(C)CC[C@@]2(CCCC[C@@H]12)c4c3
210	Pilocarpine	-ve	O=C2OC[C@H](Cc1n(cnc1)C)[C@@H]2CC
211	Triamcinolone	-ve	C[C@]12C[C@@H]([C@]3([C@H]([C@@H]1C[C@H]([C@@]2(C(=O)CO)O)O)CCC4=CC(=O)C=C[C@@]43C)F)O
212	Spironolactone	-ve	O=C5O[C@@]4([C@@]3([C@H]([C@@H]2[C@H](SC(=O)C)C/C1=C/C(=O)

			CC[C@]1(C)[C@H]2CC3)CC4)C)CC5
213	Fluocinonide	-ve	O=C(OCC(=O)[C@]45OC(O[C@@H]5C[C@@H]2[C@@]4(C[C@H](O)[C@]3(F)[C@]/1(/C=C\C(=O)\C=C\1[C@@H](F)C[C@@H]23)C)C)(C)C)C
214	Betamethasone	-ve	O=C(CO)[C@]3(O)[C@]2(C[C@H](O)[C@]4(F)[C@@]/1(\C(=C/C(=O)\C=C\1)CC[C@H]4[C@@H]2C[C@@H]3C)C)C
215	Fluorometholone	-ve	O=C(C)[C@]3(O)[C@]2(C[C@H](O)[C@]4(F)[C@@]/1(\C(=C/C(=O)\C=C\1)[C@@H](C)C[C@H]4[C@@H]2CC3)C)C
216	Methyldopa	-ve	O=C(O)[C@@](N)(Cc1cc(O)c(O)cc1)C
217	Nicergoline	-ve	Cn1cc2c3c1cccc3[C@]4(C[C@H](CN([C@@H]4C2)C)COC(=O)c5cc(cnc5)Br)OC
218	Captopril	-ve	C[C@H](CS)C(=O)N1CCC[C@H]1C(=O)O
219	Olopatadine	-ve	O=C(O)Cc2ccc1OCc3c(C(\c1c2)=C\CCN(C)C)cccc3
220	Tirofiban	-ve	O=S(=O)(N[C@H](C(=O)O)Cc2ccc(OCCCCC1CCNCC1)cc2)CCCC
221	Palonosetron	-ve	O=C5N([C@H]2C1CCN(CC1)C2)C[C@@H]4c3c5cccc3CCC4
222	Scopolamine	-ve	OC[C@H](c1ccccc1)C(=O)O[C@@H]2C[C@H]3N(C)[C@@H](C2)[C@@H]4O[C@H]34
223	Desloratadine	-ve	c1cc2c(nc1)C(=C3CCNCC3)c4ccc(cc4CC2)Cl
224	Betanidine	-ve	N(=C(/NCc1ccccc1)NC)\C
225	Sibutramine	-ve	Clc1ccc(cc1)C2(C(N(C)C)CC(C)C)CCC2
226	Thiethylperazine	-ve	S(c2cc1N(c3c(Sc1cc2)cccc3)CCCN4CCN(C)CC4)CC
227	Mequitazine	-ve	c1ccc2c(c1)N(c3ccccc3S2)CC4CN5CCC

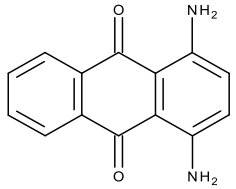
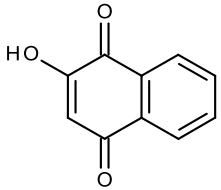
			4CC5
228	Terfenadine	-ve	OC(c1ccccc1)(c2ccccc2)C4CCN(CCCC(O)c3ccc(cc3)C(C)(C)C)CC4
229	Nabilone	-ve	O=C3CC[C@@H]1[C@H](c2c(OC1(C)C)cc(cc2O)C(C)(C)CCCCCC)C3
230	Acitretin	-ve	Cc1cc(c(c(c1/C=C/C(=C/C=C/C(=C/C(=O)O)/C)/C)C)C)OC
231	Azelastine	-ve	Clc1ccc(cc1)CC\3=N\N(C(=O)c2ccccc2/3)C4CCCN(C)CC4
232	Famotidine	-ve	c1c(nc(s1)N=C(N)N)CSCC/C(=N/S(=O)(=O)N)/N
233	Loratadine	-ve	CCOC(=O)N1CCC(=C2c3ccc(cc3CCc4c2nccc4)Cl)CC1
234	Sibutramine	-ve	Clc1ccc(cc1)C2(C(N(C)C)CC(C)C)CCC2
235	Remikiren	-ve	O=S(=O)(C(C)(C)C)C[C@H](C(=O)N[C@H](C(=O)N[C@@H](CC1CCCCC1)[C@@H](O)[C@@H](O)C2CC2)Cc3cncn3)Cc4ccccc4
236	Deserpidine	-ve	O=C(OC)[C@H]6[C@H]4C[C@@H]3c2nc1ccccc1c2CCN3C[C@H]4C[C@@H](OC(=O)c5cc(OC)c(OC)c(OC)c5)[C@@H]6OC
237	Irbesartan	-ve	CCCCC1=NC2(CCCC2)C(=O)N1Cc3ccc(cc3)c4ccccc4c5[nH]nnn5
238	Fexofenadine	-ve	O=C(O)C(c1ccc(cc1)C(O)CCCN2CCC(CC2)C(O)(c3ccccc3)c4ccccc4)(C)C
239	Eprosartan	-ve	CCCCc1ncc(n1Cc2ccc(cc2)C(=O)O)/C=C(\Cc3cccs3)/C(=O)O
240	Dipyridamole	-ve	C1CCN(CC1)c2c3c(c(nc(n3)N(CCO)CCO)N4CCCCC4)nc(n2)N(CCO)CCO
241	Fosinopril	-ve	O=C(CP(=O)(CCCCc1ccccc1)OC(OC(=O)CC)C(C)C)N2C[C@@H](C[C@H]2C(O)=O)C3CCCCC3
242	Candoxatril	-ve	O=C(N[C@H]1CC[C@H](CC1)C(O)=O)C4(C[C@@H](COCCOC)C(=O)Oc2cc

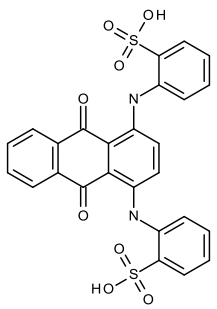
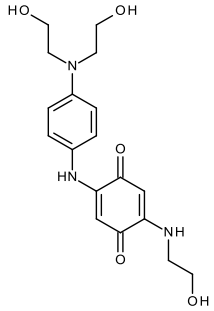
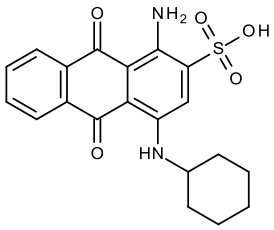
			3CCCc3cc2)CCCC4
243	Lactulose	-ve	O[C@H]2[C@H](O[C@@H]1O[C@H](CO)[C@H](O)[C@H](O)[C@H]1O)[C@H](O[C@]2(O)CO)CO
244	Rescinnamine	-ve	O=C(OC)[C@H]6[C@H]4C[C@@H]3c2nc1cc(OC)ccc1c2CCN3C[C@H]4C[C@@H](OC(=O)\C=C\c5cc(OC)c(OC)c(OC)c5)[C@@H]6OC
245	Tobramycin	-ve	C1[C@@H]([C@H]([C@@H]([C@H]([C@@H]1N)O[C@@H]2[C@@H]([C@H]([C@@H]([C@H](O2)CO)O)N)O)O)O[C@@H]3[C@@H](C[C@@H]([C@H](O3)CN)O)N)N
246	Argatroban	-ve	O=C(O)[C@@H]3N(C(=O)[C@@H](NS(=O)(=O)c1cccc2c1NCC(C2)C)CCC/N=C(\N)N)CC[C@@H](C)C3
247	Trimethaphan	-ve	O=C2N(C4C[S+]1CCCC1C4N2Cc3ccccc3)Cc5ccccc5
248	Bretylium	-ve	Brc1ccccc1C[N+](CC)(C)C
249	Clidinium	-ve	O=C(OC2C1CC[N+](CC1)(C)C2)C(O)(c3ccccc3)c4ccccc4
250	Marinol	-ve	CCCCCc1cc(c2c(c1)OC([C@H]3[C@H]2C=C(CC3)C)(C)C)O
251	Methantheline	-ve	O=C(OCC[N+](CC)(CC)C)C2c3c(Oc1c2cccc1)cccc3
252	Oxyphenonium	-ve	O=C(OCC[N+](CC)(CC)C)C(O)(c1ccccc1)C2CCCCC2
253	Propantheline	-ve	O=C(OCC[N+](C(C)C)(C(C)C)C)C2c3c(Oc1c2cccc1)cccc3
254	Meperidine	+ve	O=C(OCC)C2(c1ccccc1)CCN(C)CC2
255	Isoniazid	+ve	O=C(NN)c1ccncc1
256	Acetaminophen	+ve	CC(=O)Nc1ccc(cc1)O
257	Propranolol	+ve	CC(C)NCC(COc1cccc2c1cccc2)O
258	Clozapine	+ve	CN1CCN(CC1)C2=Nc3cc(ccc3Nc4c2cccc4)Cl

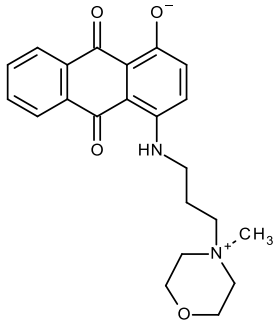
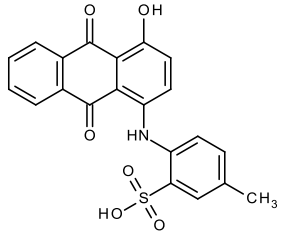
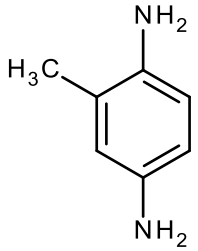
259	Piroxicam	+ve	CN1C(=C(c2ccccc2S1(=O)=O)O)C(=O)Nc3ccccn3
260	Ribavirin	+ve	c1nc(nn1[C@H]2[C@@H]([C@@H]([C@H](O2)CO)O)O)C(=O)N
261	Pioglitazone	+ve	O=C1NC(=O)SC1Cc3ccc(OCCc2ncc(cc2)CC)cc3
262	Xanthomegnin	+ve	COC=2C(=O)c1c6C(=O)OC(C)Cc6cc(O)c1C(=O)C=2\C5=C(/OC)C(=O)c4c(c(O)cc3CC(C)OC(=O)c34)C5=O
263	Malachite green	+ve	OC(c1ccc(N(C)C)cc1)(c2ccccc2)c3ccc(N(C)C)cc3
264	Methyl benzoquate	+ve	O=C\3c2c(cc(OCc1ccccc1)c(c2)CCCC)N/C=C/3C(=O)OC
265	MKT-077	+ve	O=C2C(\S/C(=C\c1cccc[n+]1CC)N2CC)=C3/Sc4ccccc4N3C
266	MTT	+ve	Cc1c(sc(n1)[n+]2nc(nn2c3ccccc3)c4ccccc4)C
267	Safranine O	+ve	n1c4c([n+](c2c1cc(c(N)c2)C)c3ccccc3)cc(c(c4)C)N
268	Vacor	+ve	[O-][N+](=O)c1ccc(cc1)NC(=O)NCc2cccnc2
269	Aspirin	+ve	CC(=O)Oc1ccccc1C(=O)O
270	Meloxicam	+ve	Cc1cnc(s1)NC(=O)C2=C(c3ccccc3S(=O)(=O)N2C)O
271	Cysteamine	-ve	SCCN
272	Phentermine	-ve	NC(Cc1ccccc1)(C)C
273	Triprolidine	-ve	n3c(\C(=C\CN1CCCC1)c2ccc(cc2)C)cccc3
274	Flucytosine	-ve	FC=1\C=N/C(=O)NC=1N
275	Tolmetin	-ve	O=C(c1ccc(n1C)CC(=O)O)c2ccc(cc2)C
276	Timolol	-ve	O[C@H](COc1nsnc1N2CCOCC2)CNC(C)(C)C

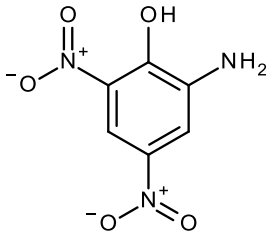
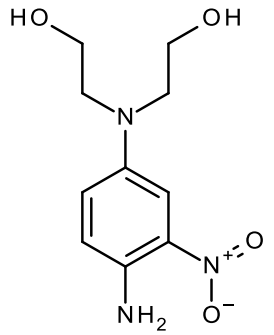
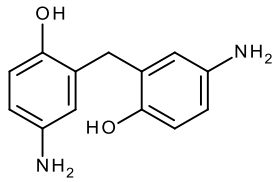
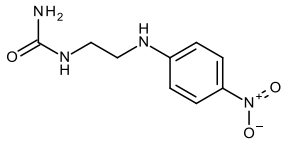
277	Cisapride	-ve	Clc1cc(c(OC)cc1N)C(=O)NC3CCN(CCCOc2ccc(F)cc2)CC3OC
278	Dexmedetomidine	-ve	n1cc(nc1)C(c2c(c(ccc2)C)C)C
279	Mannitol	-ve	C([C@H]([C@H]([C@@H]([C@@H](CO)O)O)O)O)O
280	Clotrimazole	-ve	c1ccc(cc1)C(c2ccccc2)(c3ccccc3Cl)n4ccnc4
281	Levothyroxine	-ve	c1c(cc(c(c1I)Oc2cc(c(c(c2)I)O)I)I)C[C@@H](C(=O)O)N
282	Telmisartan	-ve	CCCc1nc2c(cc(cc2n1Cc3ccc(cc3)c4ccccc4C(=O)O)c5nc6ccccc6n5C)C
283	Natamycin	-ve	OC(=O)[C@@H]3[C@@H](O)C[C@@]2(O)C[C@@H](O)C[C@H]4O[C@@H]4/C=C/C(=O)O[C@H](C)C\C=C\C=C\C=C\C=C\[C@H](OC1O[C@H](C)[C@@H](O)[C@H](N)[C@@H]1O)C[C@@H]3O2
284	Dolasetron	-ve	O=C5CN4[C@@H]1C[C@H]5C[C@H]4C[C@H](C1)OC(=O)c3cnc2ccccc23
285	Ramelteon	-ve	O=C(NCC[C@H]3c2c(ccc1OCCc12)CC3)CC
286	Tolazoline	-ve	N\1=C(\NCC/1)Cc2ccccc2
287	Levocarnitine	-ve	[O-]C(=O)C[C@@H](O)C[N+](C)(C)C
288	Apraclonidine	-ve	Clc1c(c(Cl)cc(N)c1)N/C2=N/CCN2

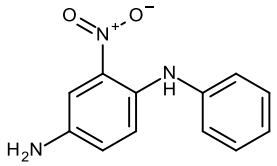
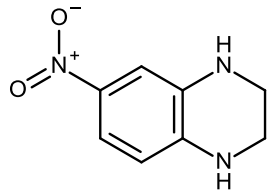
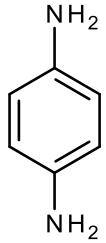
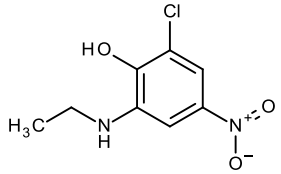
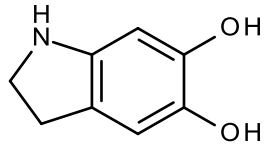
Appendix III. 90-day repeat dose toxicity data associated to each compound within the categories developed (Chapter 5). (Abbreviations: AAT – Alanine aminotransferase, APTT – Activated Partial Thromboplastin Time, AST – Aspartate aminotransferase, BWG – Body Weight Gain, GI – GastroIntestinal, MCH – Mean Corpuscular/cell Haemoglobin, MCV – Mean Corpuscular/cell Volume, PT – Prothrombin Time, RBC – Red Blood Cell.)

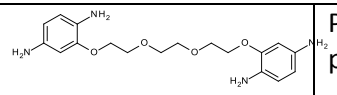
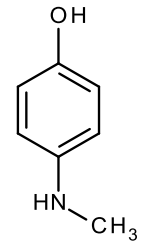
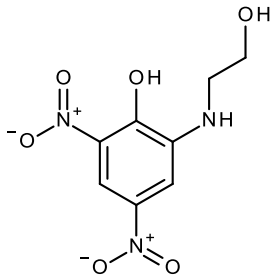
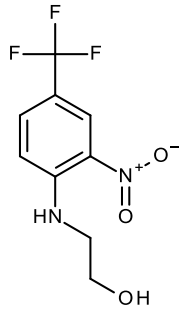
Compound Number	Category Name	Structure	Compound Name	NO(A)EL (mg/kg bw/day)	LO(A)EL (mg/kg bw/day)	Adverse effects used to derive LO(A)EL within SCC(NF)P and SCCS opinions
1	Quinone		Disperse Violet 1	2	20	↑ Centrilobular/Midzonal hepatocyte hypertrophy ↑ Triglycerides (♀) ↑ Cholesterol ↓ Motor activity
2	Quinone		Lawsone	2	7	↓ Erythrocyte count (♀) ↓ Blood urea (♀) ↓ Albumin:Globulin ratio (♀) ↑ Bilirubin (♀) ↑ Kidney weight (♀) ↓ Blood glucose (♂) ↑ Triglycerides (♂) ↑ Haematopoiesis, spleen (♂) ↑ (Multi)focal ulceration of mucosa, forestomach ↑ Interstitial oedema, forestomach

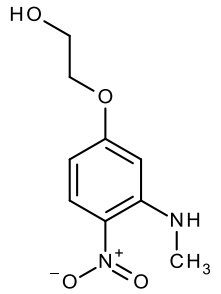
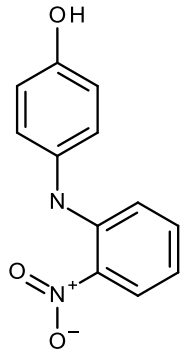
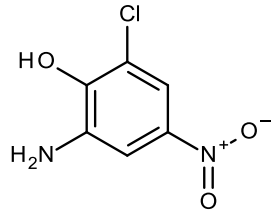
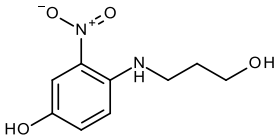
3	Quinone		Acid Green 25	100 (95 due to active ingredient)	300 (285 due to active ingredient)	↑ Kidney weight
4	Quinone		HC Green No. 1	100	300	↓ Food consumption (♀) ↓ Body weight (♀) ↑ Hypokalemia ↑ Oliguria (♂)
5	Quinone		Acid Blue 62	300 (160 due to active ingredient)	1000 (534 due to active ingredient)	↑ Kidney weight ↑ Liver weight ↑ Ptyalism ↑ Tubular nephrosis ↑ Centrilobular hepatocyte hypertrophy ↑ Blood Urea ↑ Albumin ↑ Cholesterol ↑ AAT ↓ Body weight ↓ Glucose

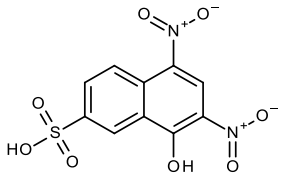
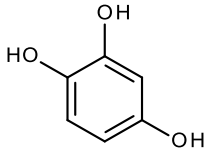
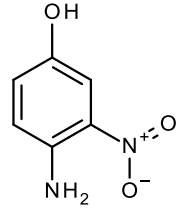
6	Quinone		Hydroxyanthraquinone Aminopropyl Methyl Morpholinium Methosulfate	200	800	↓ Absolute thymus weight (♀) ↓ Body weight (♂) ↓ Relative thymus weight
7	Quinone		Acid Violet 43	300 (282 due to active ingredient)	1000 (940 due to active ingredient)	↑ PT ↑ APTT
8	Pro-quinone		Toluene-2,5-diamine	10	20	↑ AST ↑ Mononuclear cell infiltrates, diaphragm ↑ Mononuclear cell infiltrates, eye ↑ Mononuclear cell infiltrates, thigh ↑ Mononuclear cell infiltrates, tongue ↑ Muscular degeneration, diaphragm ↑ Muscular degeneration, thigh ↑ Muscular degeneration, tongue ↑ Muscular regeneration, diaphragm

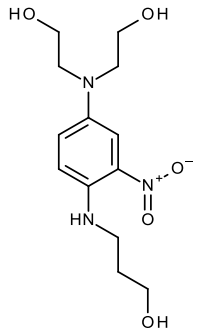
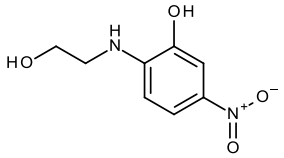
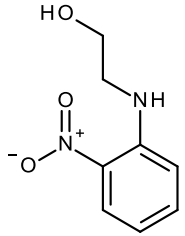
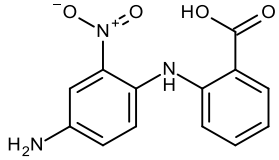
9	Pro-quinone		Picramic Acid	5 (3.2 due to active ingredient)	15 (9.4 due to active ingredient)	↑ Ulceration of GI tract ↑ Inflammation of GI tract ↑ Fibrosis of GI tract ↑ Tubular cell swelling ↑ MCV ↑ MCH ↑ Reticulocyte count
10	Pro-quinone		HC Red No. 13	No NO(A)EL	5 (4.2 due to active ingredient)	↑ Creatinine (♀) ↑ Kidney weight ↑ PT (♂) ↓ Albumin:Globulin ratio (♀) ↓ Glucose (♀) ↓ MCH (♂) ↓ MCV
11	Pro-quinone		2,2'-Methylenebis-4-aminophenol	5	15	↑ Cast formation, kidney ↑ Thickened basement membrane, kidney ↑ Tubular basophilia, kidney ↑ Tubular degeneration, kidney
12	Pro-quinone		4-Nitrophenyl aminoethylurea	5	25	↓ RBC count ↓ Haemoglobin concentration ↑ MCV ↑ Reticulocyte count ↑ Extramedullary haematopoiesis, spleen ↑ Haemosiderosis (♀) ↓ Packed cell volume (♂)

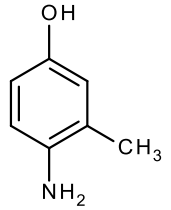
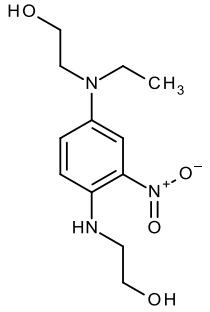
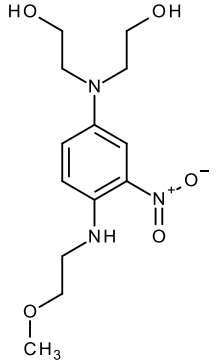
13	Pro-quinone		HC Red No. 1	5	20	↓ Erythrocytes (♀) ↑ Leukocytes (♀) ↑ Lymphocytes (♀) ↓ Thymus weight (♂) ↑ MCH (♂)
14	Pro-quinone		Tetrahydro-6-nitroquinoxaline	5	25	↑ Ptyalism ↑ Liver weight ↑ Spleen weight
15	Pro-quinone		*p*-Phenylenediamine	8	16	↑ Myodegeneration, skeletal muscle
16	Pro-quinone		2-Chloro-6-ethylamino-4-nitrophenol	10	30	↑ Liver weight
17	Pro-quinone		Dihydroxyindoline	10	20	↑ Pigmentation, kidney

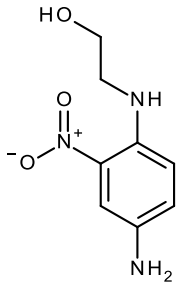
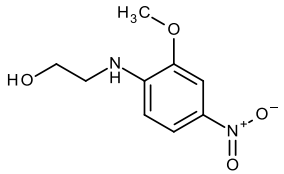
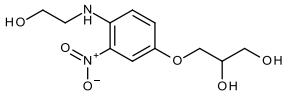
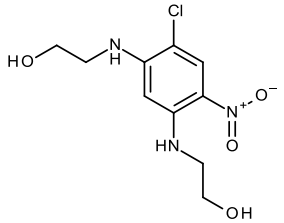
18	Pro-quinone		PEG-3-2',2'-di-p-phenylenediamine	10	25	↑ Intracellular pigmentation, kidney tubules ↑ Pigmentation, thyroid epithelium ↑ Pigmentation, duodenum
19	Pro-quinone		*p*-Methylaminophenol sulphate	10	30	↑ Tubular epithelial degeneration, kidney ↑ Single cell necrosis, kidney ↓ Specific gravity (♂) ↑ Urinary volume (♂)
20	Pro-quinone		2-Hydroxyethyl picramic Acid	15	60	↑ Protein cylinders, kidneys ↑ Activation of thyroid epithelial cells
21	Pro-quinone		HC Yellow No. 13	21	90	↑ Degeneration, Islet cells ↑ Inflammation, endocrine pancreas ↑ Fibrosis, endocrine pancreas ↑ Serum cholesterol (♂)

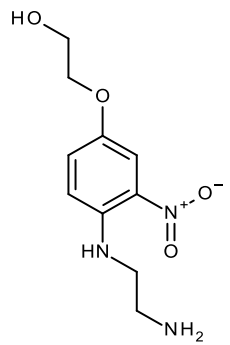
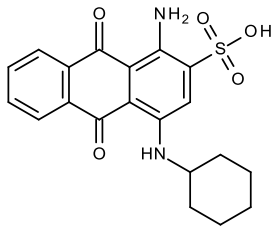
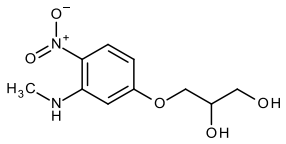
22	Pro-quinone		3-Methylamino-4-nitrophenoxyethanol	25	100	↑ Ptyalism ↓ Lymphoid in thymus
23	Pro-quinone		HC Orange No.1	25	No LO(A)EL	
24	Pro-quinone		2-Amino-6-chloro-4-nitrophenol	30	90	↓ Body weight ↑ Kidney weight
25	Pro-quinone		4-Hydroxypropylamino-3-nitrophenol	30	90	↑ Thyroid weight ↓ AST

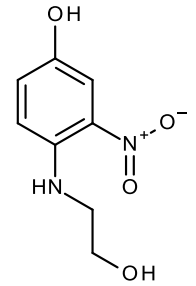
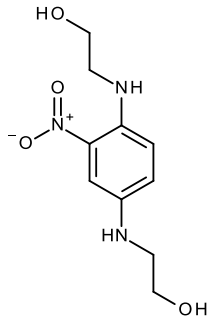
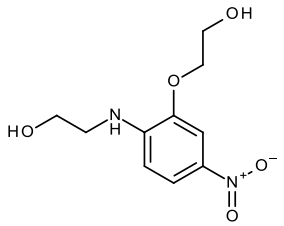
26	Pro-quinone		Acid Yellow 1	30	100	↑ Mean absolute reticulocyte ↑ Haematopoiesis ↑ Lesions, caecum ↑ Lesions, intestine ↑ Lesions, liver ↑ Lesions, spleen ↑ Haemosiderosis (♀) ↑ MCV (♀) ↑ Spleen weight (♂)
27	Pro-quinone		1,2,4-Trihydroxybenzene	50	100	↑ Piloerection ↑ Ptyalism ↑ Mean RBC volume ↑ MCH ↑ Platelets ↓ Haematocrit ↓ RBC count ↓ Haemoglobin ↑ Kidney weight ↑ Liver weight ↑ Spleen weight ↓ Body weight (♂)
28	Pro-quinone		4-Amino-3-nitrophenol	50	250	↑ Liver weight (♂)

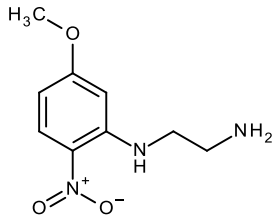
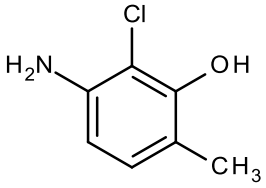
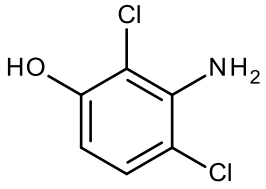
29	Pro-quinone		HC Violet No. 2	50	200	↑ Liver weight ↓ RBC ↓ PT
30	Pro-quinone		HC Yellow No. 11	50	200	↑ Acidophilic globules in cortical tubular epithelium ↑ Liver weight (♀) ↑ Kidney weight ↓ Thymus weight ↓ Creatinine
31	Pro-quinone		HC Yellow No. 2	50	No LO(A)EL	
32	Pro-quinone		2-Nitro-4-amino-diphenylamine-2'-carboxylic acid	60	180	↑ Thrombocytes ↑ Water consumption (♀)

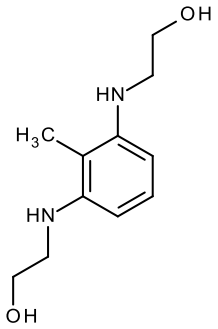
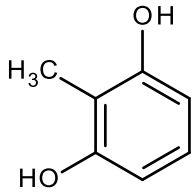
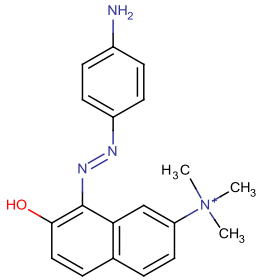
33	Pro-quinone		4-Amino-*m*-cresol	60	120	↑ Spleen weight
34	Pro-quinone		HC Blue No. 12	60	No LO(A)EL	
35	Pro-quinone		HC Blue No. 11	80	160	↑ Kidney weight ↑ Vacuolated tubular cell

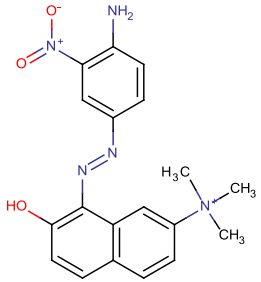
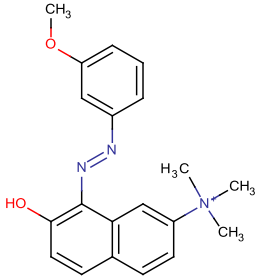
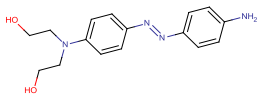
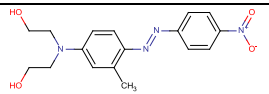
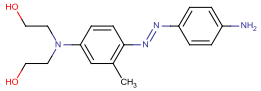
36	Pro-quinone		HC Red No. 3	90	250	↓ Body weight
37	Pro-quinone		2-Hydroxyethylamino-5-nitroanisole	100	500	↑ Liver weight ↑ Spleen weight ↑ PT ↑ Fibrinogen level ↑ Blood urea nitrogen ↑ AAT (♂) ↑ Urinary volume
38	Pro-quinone		HC Orange No. 3	100	300	↑ Kidney weight ↑ Liver weight ↑ Spleen weight ↑ AAT ↑ AST
39	Pro-quinone		HC Yellow No. 10	100	500	↑ Staining, body ↑ Staining, fur ↑ Body weight ↑ Ptyalism ↑ Food consumption ↑ Liver weight ↑ Spleen weight (♂)

40	Pro-quinone		HC Orange No. 2	150	500	↑ Ptyalism ↓ BWG ↓ Food consumption ↓ Blood glucose
41	Pro-quinone		Acid Blue 62	300 (160 due to active ingredient)	1000 (534 due to active ingredient)	↑ Kidney weight ↑ Liver weight ↑ Ptyalism ↑ Tubular nephrosis ↑ Centrilobular hepatocyte hypertrophy ↑ Blood urea ↑ Albumin ↑ Cholesterol ↑ AAT ↓ BWG ↓ Glucose
42	Pro-quinone		2-Nitro-5-glyceryl methylaniline	200	800	↑ Ptyalism ↑ Vacuolated pancreatic cells ↑ Vacuolated renal tubular cells ↑ Tubular nephrosis ↑ Piloerection ↑ Hunched back ↑ Hypokinesia ↑ Bilateral opacity ↑ Adrenal weight

						↑ Kidney weight ↑ Liver weight ↓ BWG
43	Pro-quinone		3-Nitro-*p*-hydroxyethylaminophenol	200	No LO(A)EL	
44	Pro-quinone		N,N'-bis(hydroxyethyl)-2-nitro-*p*-phenylenediamine	240	720	↑ Kidney weight ↑ Liver weight ↓ Activity (♀) ↓ Ataxia (♀) ↑ Ptyalism (♀) ↑ Ocular discharge (♀) ↑ Lethargy (♀) ↑ Hunched posture (♀) ↑ Triglycerides (♂) ↑ Urea (♂) ↑ Urinary specific gravity
45	Pro-quinone		HC Yellow No. 4	250	500	↓ Body weight ↑ Thyroid lesions ↑ Uterine lesions (♀) ↑ Kidney lesions (♂) 1 Mortality

46	Pro-quinone		HC Yellow No. 9	250	No LO(A)EL	
47	Meta-hydroquinone		5-amino-6-chloro-*o*-cresol	No NO(A)EL	100 (33 after adjustment factor of 3)	↑ Centrilobular hepatotrophy, liver ↑ MCV ↑ Mean corpuscular Hb (♀) ↑ MCH concentration (♀)
48	Meta-hydroquinone		3-Amino-2,4-dichlorophenol	80	160	↑ Liver degeneration ↑ Liver necrosis ↑ Foci mononuclear cell infiltration ↑ Kidney degeneration ↑ Kidney necrosis ↑ Tubular epithelial cell hypertrophy ↑ Phosphorus (♂) ↑ Sodium (♂) ↑ Chloride (♂)

49	Meta-hydroquinone		2,6-Dihydroxyethylaminotoluene	100	316	↑ Bilirubin ↑ Urobilinogen ↓ Serum creatinine (♀)
50	Meta-hydroquinone		2-Methylresorcinol	100	200	↑ Clonic spasms ↑ Ptyalism ↑ Scratching movements ↑ Body weight (♂) ↑ Liver weight (♂) ↑ AST (♂) ↑ AAT (♂)
51	Aromatic azo		Basic Brown 16	50 (32 due to dye content)	150 (99 due to dye content)	↓ BWG (♂)

52	Aromatic azo		Basic Brown 17	60 (46 due to dye content)	120 (93 due to dye content)	↑ Extramedullary haemopoiesis
53	Aromatic azo		Basic Red 76	20	60	↓ RBC (♂) ↓ Haemoglobin ↓ Haematocrit (♂) ↓ MCH concentration (♀)
54	Aromatic azo		Disperse Black 9	100 (52.6 due to dye content)	No LO(A)EL	
55	Aromatic azo		Disperse Red 17	10 (4 due to dye content)	30 (12 due to dye content)	↑ Spleen weight
56	Aromatic azo		HC Yellow N°. 7	10	40	↑ Kidney weight ↑ Bilateral discolouration of fundus ↑ Pytalism ↑ Tubular basophilia ↑ Blood phosphorous (♀)

						↓ Blood glucose (♀) ↑ Blood sodium (♂)

Appendix IV. Copies of published articles

One. Nelms MD, Cronin MTD, Enoch SJ, Schultz TW (2013) Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones. SAR QSAR Enviro Res 24(9): 695-709

Two. Nelms MD, Ates G, Madden JC, Vinken M, Cronin MTD, Rogiers V, Enoch SJ (2014) Proposal of an *in silico* profiler for categorisation of repeat dose toxicity data of hair dyes. *Arch Toxicol*. Available at: http://link.springer.com/article/10.1007%2Fs00204-014-1277-8 (accessed 17.11.2014)