A KNOWLEDGE BASED APPROACH OF TOXICITY PREDICTION FOR DRUG FORMULATION

MODELLING DRUG VEHICLE RELATIONSHIPS USING SOFT COMPUTING TECHNIQUES

Pritesh MISTRY

Submitted for the Degree of

Doctor of Philosophy

Faculty of Engineering and Informatics

University of Bradford

2015

Abstract

Pritesh Mistry

A Knowledge Based Approach of Toxicity Prediction for Drug Formulation

Modelling Drug Vehicle Relationships Using Soft Computing Techniques

Keywords: Predictive toxicology, Drug formulation, Random forest, Decision trees, Big data, Data mining, Area-under-curve (AUC)

This multidisciplinary thesis is concerned with the prediction of drug formulations for the reduction of drug toxicity. Both scientific and computational approaches are utilised to make original contributions to the field of predictive toxicology.

The first part of this thesis provides a detailed scientific discussion on all aspects of drug formulation and toxicity. Discussions are focused around the principal mechanisms of drug toxicity and how drug toxicity is studied and reported in the literature. Furthermore, a review of the current technologies available for formulating drugs for toxicity reduction is provided. Examples of studies reported in the literature that have used these technologies to reduce drug toxicity are also reported. The thesis also provides an overview of the computational approaches currently employed in the field of *in silico* predictive toxicology. This overview focuses on the machine learning approaches used to build predictive QSAR classification models, with examples discovered from the literature provided.

Two methodologies have been developed as part of the main work of this thesis. The first is focused on use of directed bipartite graphs and Venn diagrams for the visualisation and extraction of drug-vehicle relationships from large un-curated datasets which show changes in the patterns of toxicity. These relationships can be rapidly extracted and visualised using the methodology proposed in chapter 4.

The second methodology proposed, involves mining large datasets for the extraction of drug-vehicle toxicity data. The methodology uses an area-under-the-curve principle to make pairwise comparisons of vehicles which are classified according to the toxicity protection they offer, from which predictive classification models based on random forests and decisions trees are built. The results of this methodology are reported in chapter 6.

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Pritesh Mistry

Acknowledgements

First and foremost I would like begin by thanking my supervisors, Prof. Daniel Neagu, Dr. Paul Trundle and Dr. Jonathan Vessey. Without their guidance, wisdom, support and humour I find it difficult to see how any of this would have been possible.

A special thank you goes to Lhasa Limited, Leeds, UK, for sponsoring my PhD studies. I have thoroughly enjoyed spending many months interacting and learning from the scientific research staff at Lhasa. Their knowledge in this field of study has been invaluable to the completion of this thesis.

Publications and Presentations

Papers in Peer-Reviewed Journals and Conference Proceedings

Mistry P, Neagu D, Trundle PR and Vessey JD (2015) Using random forest and decision tree models for a new vehicle prediction approach in computational toxicology. *Journal of Soft Computing.* Springer. Published online before print. (doi:10.1007/s00500-015-1925-9)

Mistry P, Palczewska A, Neagu D, Trundle P. (2014) Using computational methods for the prediction of drug vehicles. *14th UK Workshop on Computational Intelligence (UKCI)*. 306-312. IEEE Xplore http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6930194 (doi: 10.1109/UKCI.2014.6930194)

Presentations

Presentation at the 2014 14th UK Workshop on Computational Intelligence (UKCI 2014). *Using Computational Methods for the Prediction of Drug Vehicles*. 8 – 10th September 2014, University of Bradford, UK.

Virtual Presentation at the International Virtual Conference in Toxicology. *Using Computational Methods for the Prediction of Drug Vehicles*. 13 – 17 October 2014, University of Medicine and Pharmacy, Bucharest, Romania.

Table of Contents

List of Figures

List of Tables

Glossary

ALT	Alanine transaminase
ANN	Artificial neural network
AST	Aspartate transaminase
ATP	Adenosine triphosphate
AUC	Area under the plasma concentration time curve
BUN	Blood-urea-nitrogen
C_{max}	Peak plasma concentration of a drug reached over a specified time
CMC	Carboxymethycellulose
Decilitre (dL)	One tenth of a litre
Drug formulation	Technologies that solubilise, encapsulate or adsorb drugs for ease of administration and/or toxicity reduction
Drug vehicle	Agent, often a solvent, oil or polymer used to facilitate drug administration
DT	Decision tree
HPC	Hydroxypropylcellulose
GSH	Glutathione
IC_{50}	Dose of drug required to reduce a process by half
i.m.	Intramuscular
InChI	IUPAC International Chemical Identifier
i.p.	Intraperitoneal
i.v.	Intravenous
kNN	k-Nearest neighbour

LD_{50}	Median lethal dose
LOO	Leave one out
MPT	Mitochondrial permeability transition
MRT	Mean residence time
MTD	Maximum tolerated dose
NCE	New chemical entity
NOAEL	No-observable-adverse-effect-level
QSAR	Quantitative structure activity relationship
QT interval	The time of a heart's electrical cycle between the Q wave and T wave
RBCs	Red blood cells
RES	Reticuloendothelial System or Mononuclear phagocyte system (MPS)
RF	Random forest
s.c.	Subcutaneous
SAR	Structure activity relationships
SMILES	Simplified Molecular Input Line Entry Specifications
SVM	Support Vector Machine
Therapeutic Index	The ratio of the dose required to produce a toxic effect and the dose required for a therapeutic effect
$t_{1/2}$	Elimination half-life
T_{max}	Time taken to reach C_{max}
V_d	Volume of distribution

Vehicle	An inert substance with which a drug is mixed to facilitate administration
V_{ss}	Volume of distribution at steady state
WBCs	White blood cells
Xenobiotic	A chemical substance not normally considered to be endogenous within an organism

"...the awakened sages call a person wise when all his

undertakings are free from anxiety about results..."

Krishna – The Bhagavad Gita

1

Introduction

This thesis is concerned with the application of pharmaceutical drug formulations and their influence on drug toxicity. Drug formulation information is gathered and processed to understand how formulations reduce drug toxicity and whether this knowledge can be used to formulate for the general case.

Original contributions are made in the field of data visualisation, data extraction methods and data classification methods of toxicity data which shows a difference in toxicity with a change in formulation. The effectiveness of these methods is demonstrated using real scientific data.

1.1 Background and Motivation

Toxicity is a major concern implicated in the attrition rates of drug candidates consisting of new chemical entities (NCE). As a result of continued efforts to improve the efficacy and bioavailability of NCEs a shift in attrition rates has been observed from poor efficacy and bioavailability to other stages of the developmental pipeline [1, 2], particularly the toxicity concerns that are apparent today [1]. A clinical phase study of the NCEs for the period of 2011 to 2012 highlighted that failures due to toxicity concerns were higher in phase III and beyond compared with phase II [3]. It was suggested that toxicity as the cause may only become apparent in larger patient trials. The problem is further exacerbated by our need for therapies against more aggressive diseases such as cancer, which carries a developmental failure rate of ~95 % [1]. The financial cost of attrition is large; it was estimated that the cost of bringing a new drug to market in 2000 was $802 million (USD)

with an annual increase of 7.4 % suggested [4], that cost today equates to over $2 billion (USD) spread over 15 years [4-7]. Failure further down the developmental stages are inevitably more costly but we can learn from the continuous improvements made on the bioavailability and efficacy fronts to lead the way for improvements on toxicity and safety issues.

Analysis of the attrition rates per therapeutic area uncovers distinctive trends for which attrition rates are more problematic. Oncology carries the highest failure rates for toxicity reasons [3], largely accounted for by the need for novel mechanisms of action, where proof of concept in the clinic coupled with the cytotoxic nature of the drug candidates is difficult to contain within disease areas. If acceptable efficacy and proof of concept for new oncology drugs can be achieved then the aim is to dissociate this from any toxicity they may inflict.

One approach for lowering toxicity is for medicinal chemists to reformulate the chemical structure during the hit to lead optimisation stage of early drug discovery. Whilst this can be successful it is not without compromise and can affect the efficacy, solubility, permeability or any other desirable property the compound was originally designed for.

An alternative approach is through the use of drug formulations; that is to use innovative pharmaceutical formulations that allow the safer use of drugs without altering their chemical structure. Whilst formulations that improve the solubility, permeability and efficacy of a drug have been brought to market [8, 9], their use for toxicity reduction is not as well established. Demonstrating the functionality of a drug formulation for toxicity reduction can be advantageous during development and when formulating the final product. For cytotoxic drugs which fall under the oncology therapeutic area, it can push the toxicity threshold, expanding the therapeutic index range over which a drug can be safely administered. The advantages translated to the patient are immense, allowing more effective treatment with potentially increased efficacy.

Although formulating for cytotoxic drugs seems to be the most immediate necessity, in principle formulating for other therapeutic areas where toxicity is a concern is also possible. Successfully lowering toxicity of otherwise efficacious drugs would result in huge financial savings enabling drugs to be brought to market sooner.

Formulation efforts that overcome non-toxicological issues such as improving the solubility or permeability of a drug compound are relatively well understood with respect to their

mechanisms of action [10-14]. For toxicity however, the mechanisms of action are not as well defined and as a result, formulating to reduce toxicity can be a hit or miss process. An understanding of the underlying mechanistic principles of how formulations mitigate drug toxicity may allow more informed decisions to be made about formulating for new compounds. This knowledge may also allow rules to be formed for the general case.

Producing such rules would involve obtaining or gathering data from which knowledge can be extracted and modelled for decision making (Figure 1.1). A collection of examples for which formulations have been clinically shown to reduce toxicity for a range of different compounds may allow a database or dataset to be compiled which can be mined and modelled to produce generalised rules. Achieving this would allow NCEs and currently available drugs to be formulated to reduce their toxicity burden on the host.

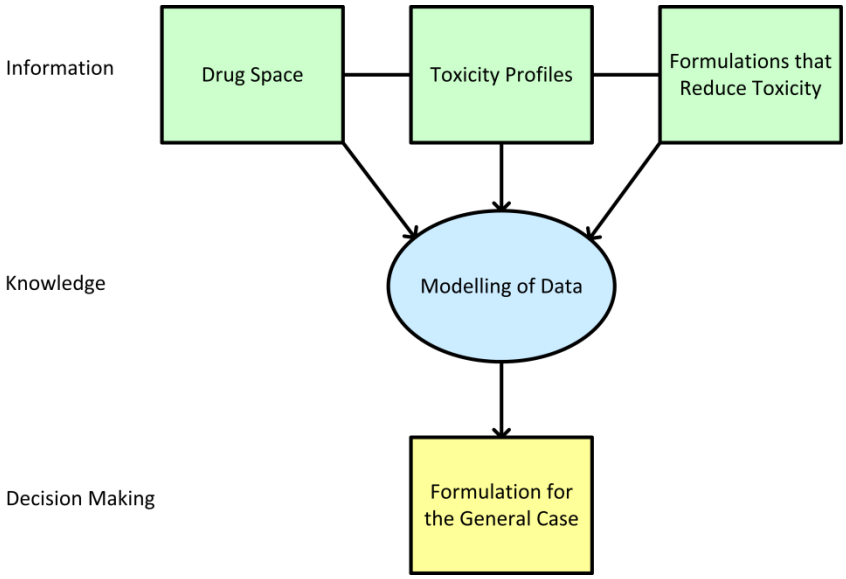


Figure 1.1. Modelling drug formulation data for the general case. Collected data which includes drug information, toxicity profiles and formulations that reduce toxicity can be modelled and decisions made for the general case.

1.2 Problem Description

Drug toxicity is a complex process that involves multiple pathways which manifest into an observable toxicity. Toxicity observed during the developmental process of NCEs becomes costly and time consuming in bringing the right compound to market with an acceptable safety profile. Compounds which display organ, mechanism-based or off-target toxicity are just some of the safety concerns that result in drug development failures. Drug formulations can help reduce some of these unwanted toxicities, but deciding which formulation will be successful is not a straightforward process. The use of computational

intelligence may help with this problem but it is recognised that this is a relatively new area of research.

The current work on using computational intelligence for predictive toxicology purposes focuses on the prediction of specific toxicities [15-19]. This field of research draws enormous amounts of interest from not only the pharmaceutical industry but many other disciplines, owing largely to the potential of reducing or eliminating biological testing and reducing costs substantially. A range of *in silico* systems for predicting a variety of toxicities, such as, genotoxicity, mutagenicity or hepatotoxicity exist today. Such systems are built on computational human reasoning (expert systems), modelling of biochemical profiles (molecular modelling) or the use of experimentally produced data, providing a library of knowledge (data driven). Although this work can provide a foundation to build upon it does not provide a complete framework for which formulation predictions for toxicity reduction can be made.

To do this we first need to understand the physiological and cellular mechanisms that contribute to the manifestation of a toxicity outcome. Generally it is accepted that a toxicity outcome is caused by a cascade of events. These events begin with an initial exposure followed by a multitude of interactions between the invading toxicant and organism which culminates into the observable or measurable toxic effect.

For a formulation to reduce or prevent toxicity it must therefore interfere with some aspect of this toxicity pathway. This could be at the initial exposure event or further down the cascade. Although different chemical compounds will instigate different toxicity pathways there may be some common ground between two or more drug compounds. These drugs could be related by some measure of similarity or completely unrelated. If we can uncover these commonalities then a formulation that reduces toxicity for one compound by interfering with a common pathway should in principle work for the other compound too.

This thesis therefore aims to address the following question:

Can we predict formulations to reduce toxicity?

Discovering common toxicity pathways for different compounds is one approach that will be explored, although the limitations of this method for many drug compounds and in particular NCE's is that the major toxicity pathways are not always fully known. For this

reason alternative approaches are essential to fully explore the possibilities of this problem. The use of QSAR models to relate drug structure to possible toxicity outcomes is the other major area of study this thesis focuses on. If the activity, which in this case is the toxicity, of a compound can be related to its chemical structure then we can use this information to link other chemical structures with similar structural properties. Such models are data dependent and so the right information needs to be present for such models to be built.

1.3 Aims

This thesis aims to address the issues discussed above; more specifically it aims to introduce early workings of an area of research that is relatively unexplored; to build models which are able to predict formulations that would allow for toxicity reduction for the general case. A summary of these aims is given below:

- Explore the underlying principles/pathways of drug toxicity,
- What are the outcomes/endpoints of drug toxicity,
- Discover drug formulations that reduce toxicity,
- Define the current *in silico* computational approaches used in the general field of drug toxicity,
- Obtain and collate scientific data that can be mined, visualised and extracted for interesting data relationships,
- Use machine learning approaches to build classification models that are able to predict formulations for toxicity reduction for the general case,
- Provide scientific interpretations where possible of any models built.

Although this research will focus on the novel approaches in formulation prediction to reduce toxicity, the methodological aspects of this work can in theory be implemented for other areas of research where data driven modelling would be appropriate.

1.4 Methodology and Data

This thesis is concerned with discovering, visualising and building models for the prediction of drug formulations to reduce toxicity. To achieve these aims the following existing methodologies were used:

- Graph theory and Venn diagram representation for visual representation and data extraction.
- The principles of the area under the curve (AUC), to discriminate between the differences in toxicity protection afforded by different drug vehicles.
- Decision tree and Random forest ensemble methods for building classification models.

The data used for the work in this thesis came from several sources which were compiled into a format suitable to aid with answering the main research questions of this thesis. Two main data sources were used for the work in this thesis, they are:

- Data extracted from primary literature sources (journal papers) and compiled into a useable form
- Oncology data obtained from the National Institutes of Health's (NIH) Developmental Therapeutics Program (DTP) [20]

1.5 Thesis Structure

This thesis is arranged into seven chapters and one appendix:

- Chapter 2 – presents a detailed literature review of the scientific aspects of drug toxicity and drug formulation.

This chapter discusses the different types of drug toxicity frequently studied and reported in the literature. The principle mechanisms of drug toxicity are also discussed with regards to their pharmacokinetic and pharmacodynamic actions. A review and discussion of the current technologies available for drug formulations are presented in this chapter, with examples provided of their practical application in reducing drug toxicity. These technologies largely impact the pharmacokinetic or pharmacodynamic actions of a drug to achieve a reduction in toxicity.

- Chapter 3 – presents an overview of the current computational machine learning approaches available for *in silico* classification modelling.

This chapter introduces the various computational machine learning algorithms that have been used in predictive toxicology studies with examples from the literature provided. An overview how predictive models are built is also provided and covers topics such as, data collection, data curation, data representation, chemical data representation, model validation and performance measures.

- Chapter 4 – presents the methodology developed for the visualisation and representation of drug toxicity data using graph theory and Venn diagrams for the extraction of interesting drug-vehicle relationships.

The methodology proposed in this chapter is demonstrated in practice through the use of the NIH dataset, from which interesting drug-vehicle relationships are readily visualised and extracted from a largely un-curated dataset. Mining the graphs thus allows interesting drug-vehicle relationships to be extracted and plotted using Venn diagrams to identify crossover sections that highlight multiple points of interest.

- Chapter 5 – presents a methodology developed for the extraction and classification of toxicity data using the proposed area under the curve with interpolation and extrapolation approach.

In this chapter, aspects of big data analysis and the challenges associated with such work are discussed. A methodology developed, based on the area under the curve principle with interpolation and extrapolation, is applied to the NIH dataset, for the extraction and classification of vehicles to build models with. Twelve pairwise vehicle comparisons are made from which several different binary classification models are built using random forest and decision tress. The results of these models are reported in chapter 6.

- Chapter 6 – presents the results of the 12 binary classification models built using the area under the curve with interpolation and extrapolation approaches.

The successful results generated using random forest and decision trees for each of the models are discussed with findings from the decision tree models used to provide scientific interpretation of the results. Furthermore, from the decision tree data it was possible to create partially ordered sets of vehicles for one group of compounds containing aziridine ring structures, from which rules could be formed.

- Chapter 7 – presents the conclusions drawn from the work presented in this thesis and provides suggestions for future work and directions of research.
- Appendix A – includes the NIH dataset used in this thesis, curated drug structures and classification tables for the 12 pairwise comparisons.

2

Literature Review

This chapter provides an informative discussion based on extensive literature reviews covering all aspects of drug toxicity with regards to formulation. It has been structured to provide a thorough scientific understanding of the study of drug toxicity with respect to formulation work, referencing the primary sources of information discovered during literature searches.

The chapter begins with an overview of drug toxicity and the various different types of toxicity that are frequently studied using *in vitro* and *in vivo* scientific experiments. The principal mechanisms of drug toxicity are discussed which show how a drug's toxic profile can be linked to its pharmacokinetic and pharmacodynamics behaviour.

Information about what a drug formulation is and the types of drugs that benefit from drug formulations are discussed. The different technologies currently available for drug formulation which have been demonstrated to reduce toxicity are also discussed and explanations to their mechanism of actions are reported where information was available to do so. References to the literature are also provided where appropriate.

The chapter ends by constructing toxicity protection pathways which help visualise how formulations avert toxicity at various points of an adverse outcome pathway.

2.1 Drug Toxicity

Drug toxicity is an inherent problem associated with pharmaceutical drug therapy. The extent of the toxicity is dependent on many factors making it a multivariate problem. Some of the variables associated with toxicity include species, age, sex, dose, *in vivo* exposure, route of administration, metabolism and excretion. All drugs are toxic to some degree and are considered to exhibit both primary intended effects and secondary unintended effects, known as side effects or adverse effects. Toxicology can be considered as the study of these adverse effects on living organisms at the cellular, biochemical and molecular level. The mechanisms that contribute to a toxicity outcome are related by a series of complex events which begin with an initial exposure (dose) that induces a multitude of physiological and biochemical interactions, culminating into a toxic effect. Understanding the underlying principles of these interactions will allow a drug's toxic profile to be discovered and help ameliorate any adverse effects through formulation efforts.

2.1.1 Toxicity Testing

Experimentally, there are many well designed *in vitro* and *in vivo* toxicology tests that have been developed to study drug toxicity which vary in cost, time and information gain. Toxicity testing comprises of several components which include:

1. A biological system: which can be *in vitro* cell lines or *in vivo* animal models
2. A defined endpoint: which defines what toxicity assessment is to be made (cell viability, hepatoxicity, mutagenicity, ototoxicity etc.)
3. A method of measuring the endpoint: cell survival (haemolysis of RBCs), QT interval (heart), glomerular filtration rate (kidneys)
4. An analysis method to translate the result to a toxicity for example: if glomerular filtration rate is reduced then how much of a reduction is needed for impaired function (i.e. toxicity)
5. A means of translating the *in vitro* and *in vivo* data into relevant human data (allometric scaling)

2.1.1.1 *In Vitro* Toxicity Testing

In vitro toxicity testing typically involves the use of bacterial or mammalian cell cultures to screen for toxicity by assessment of the basal function of cells (i.e. common cellular functions). *In vitro* testing is relatively cheap to perform with a high throughput rate that can produce large number of data quickly. With an increasing emphasis within the scientific community to reduce the use of animals in toxicity testing [21] *in vitro* tests are increasingly used at the initial stages of development. Although *in vitro* tests were originally thought of as "hazard alert" tests, assessing cell viability as an indication of cytotoxicity, they are much more refined and developed today. *In vitro* test assays today are able to quantify metabolic rates, gene expression levels, membrane permeability, mitochondrial function, and even distinguish between the different cell death events of necrosis and apoptosis [22].

In vitro toxicity testing can be performed on many cell types (e.g. RBCs, hepatocytes, kidney epithelial cells and fibroblasts) [23, 24]. Toxicity endpoints may include the inhibition of cell growth, the release of cytosolic enzymes or the consumption of O_2 and ATP. Organ specific toxicities can be assessed in specialised cultured cells pertaining to a particular organ. Membrane integrity or specific cellular metabolism functions can be assessed for toxicity; for instance, the metabolism of glycogen in primary hepatocytes or the beating rate in myocardial cells [25].

Of the formulation work discovered in the literature, it was found that the majority of *in vitro* toxicity testing was largely related to RBC haemolysis measurements [26-28] with much fewer *in vitro* toxicity tests reporting the use of kidney [24, 29], liver [23, 30] or heart cells [31].

2.1.1.2 *In Vivo* Toxicity Testing

In vivo toxicity testing uses laboratory animals to assess haematological and histological parameters [32-34] as a measure of cellular dysfunction which provide indications of organ toxicities. Haematological assessment typically involves the use of biomarkers (short for biological markers) that are objectively measured from a sample of blood (see section 2.3.2). Histological examinations are assessments of organ tissue, which are stained and observed under a microscope to understand the morphology of the toxicity inflicted upon the different cells of a particular organ (see section 2.3.3). *In vivo* testing allows the toxicity to be assessed within a functional biological system, enabling a laboratory to examine different dosing schedules, administration routes and species. *In vivo* testing can provide

real time data over a specified period of time with multipoint toxicity endpoints investigated simultaneously.

The majority of the formulation work found in the literature typically used rats or mice as part of their *in vivo* toxicity testing experiments. Blood cell counts and clinical chemistry parameters were measured to assess various organ toxicity endpoints [33, 35, 36]. Many of the *in vivo* toxicity studies found in the literature additionally assessed the differences in the pharmacokinetics between formulated and non-formulated drugs. The differences observed were then used to provide scientific interpretations for why a formulation was able to reduce the toxicity of a drug [29, 35]. In this thesis, the term 'non-formulated' refers to the alternative administration of the drug used to compare the formulated drug against.

Both *in vitro* and *in vivo* toxicity testing do however come with some disadvantages. For *in vitro* studies the results are difficult to translate into a biological toxicity response. Cells in culture can behave differently to cells *in vivo* and a collective organism toxicity profile cannot be easily constructed from *in vitro* results. Furthermore pharmacokinetic parameters relating to the absorption, distribution, metabolism (biotransformation) and excretion (ADME) which dictate the exposure levels of organs cannot be adequately assessed *in vitro*.

In vivo results are on the whole more informative of how an entire organism behaves when exposed to a toxicant. Biotransformations, membrane permeability, intrinsic intercellular synthetic pathways and inter-organ influences are more likely to be preserved in a living organism than in a cultured cell line [25]. The disadvantage of animal studies however, is that the data do not always correlate with human toxicity data. Different species will sometimes metabolise drugs differently; these differences can result in very different toxicity profiles for the same drug when tested in different species [37, 38]. For this reason it is sometimes difficult to predict the toxic effect in humans from rat or mice data, with a concordance of approximately 43% suggested for the prediction of human toxicity from rodent data [39].

2.2 Types of Toxicity

For NCEs in development there are many toxicity endpoints that need to be satisfied before compounds can progress through the clinical stages prior to market approval. Preclinical toxicological investigations typically include toxicokinetic and non-clinical pharmacokinetic (PK) studies, general toxicity studies, reproduction toxicity studies, genotoxicity studies,

and for drugs that are intended for extended use, an assessment of its carcinogenicity is required [40]. The earliest toxicity considerations for NCEs typically involve investigations of their acute, chronic, sub-acute and sub-chronic toxicities for which there are accepted practises recommended by various regulatory bodies [41-50] that aim to uphold the principles of the 3Rs (Replacement, Reduction and Refinement) [21].

In vitro and *in vivo* studies can then be designed around assessing any specific toxicological concerns such as that of hepatotoxicity, nephrotoxicity and cardiotoxicity amongst others.

Of the various toxicity endpoints that were studied with respect to the formulation work discovered in the literature, a brief overview of these toxicities is provided below. Although alternative interpretations for these toxicities may exist, they are discussed here to reflect how they were collectively described within the literature discovered.

- Acute toxicity – is defined as the adverse effects observed immediately or a short time following (within 24 hours) a single or multiple dose of a substance [51].

- Systemic toxicity – is defined as the toxicity that can be seen in some part of the body or in the behaviour of an animal that is not considered normal behaviour. A reduced body weight or organ weight, apathy, lethargy, anaphylaxis, somnolence and loss of appetite can all be considered as symptoms of systemic toxicity [52].

- Local toxicity – is defined as the toxicity observed at the site of administration as a result of exposure to a toxicant [53, 54].

- Hepatotoxicity – is defined as the clinical and pathological injury inflicted upon the liver which impairs its function when compared to baseline levels. Due to the liver's function as the major site of drug metabolism, and interposed between the gastrointestinal tract and systemic circulation, it is one of the most susceptible organs to drug toxicity [23, 30].

- Nephrotoxicity – is defined as the injury caused by a toxicant on the kidneys resulting in impaired renal function, often measured as a decrease in glomerular filtration rate [24, 29, 33].

- Cardiotoxicity – is defined as impaired or failed heart function as a result of exposure to a toxicant. Cardiotoxicity can result from cardiomyocyte death via apoptosis and

necrosis, or cardiomyocyte dysfunction resulting in arrhythmia, which can be reversible [31, 34].

- Ototoxicity – is the defined as toxicity to the ear and more specifically to the cochlea and auditory nerve. Symptoms of ototoxicity include hearing loss, vertigo and tinnitus [55, 56].

- Pulmonary toxicity – is the defined as the adverse effects resulting in an impairment in lung performance. Pulmonary toxicity can result in breathlessness, fatigue, inflammation and swelling [57, 58].

- Testicular toxicity – is defined as the toxicity which results in a decreased reproductive function. Testicular toxicity may affect spermatogenesis, as well as the hormone levels of testosterone or follicle-stimulating hormone [34, 58, 59].

- Myelotoxicity – is defined as the decrease in the cells responsible for immunity as a result of bone marrow suppression. Since all immune cells originate from haematopoietic cells in the bone marrow then a myelotoxic compound by definition can also be considered immunotoxic. A decrease in the production of leukocytes, erythrocytes, and thrombocytes increases the chances of infection, anaemia and severe bleeding [60, 61].

2.3 Measuring Toxicity

Toxicity can be measured and reported in several ways depending on the type of experiment used to study the toxicity of a drug. *In vitro* and *in vivo* toxicity studies generate different types of toxicity measurements which can be related to the toxicities described above. The toxicity measurements reported in the literature work can be split into three different types: dose measurements, biomarker measurements and histopathological indications. These methods for toxicity can provide distinct information about the nature of a drug's toxicity but may also be interlinked to build an overall toxicity profile of a drug.

2.3.1 Dose Measurements

Dose measurements are used when attempting to quantify how toxic a drug substance is and can be compared by means of dose amount required to cause lethality; for instance if a drug causes lethality in mice at a dose of 10 mg/kg body weight and another causes

lethality at a dose of 0.1 mg/kg body weight, then evidently the latter is much more toxic. Some of the dose measurements that were reported in the literature are listed below:

- MTD – Maximum tolerated dose is the maximum dose administrable that does not result in unacceptable toxicity, organ dysfunction or death [51].

- LD_{50} – The median lethal dose is the dose required to kill half the tested population over a specified period of time [27, 62].

- Cell viability – Cell viability dose measurements are *in vitro* examinations which use RBCs (haemolysis) or primary cell lines that are exposed to a toxicant. The dose at which cell viability is reduced by a certain percentage is reported. The lysis of RBCs is generally considered an indication of acute or systemic toxicity. The viability of primary cell types is an indication of potential organ toxicity *in vivo* [23, 24, 26-31].

- NOAEL – The No-observable-adverse-effect-level (NOAEL) is the dose level at which no adverse effect is observed.

2.3.2 Biomarker Measurements

Biomarkers are endogenous measureable macromolecules that can be measured as an indicator of biological state or integrity. Biomarkers are useful for taking real time measurements over an extended period of time so that long term exposure studies can be investigated. They may be more reproducible as signs of toxicity than other measurements and can be specific not only to an organ but for specific types of injuries. Due to the specialised function of organs such as that of the liver (detoxification of xenobiotics) and kidneys (removal of organic molecules from the blood), they possess specialised cells which express unique enzymes and other biomolecules that allow them to carry out their function. When a drug acts upon an organ to cause toxicity, it frequently results in the necrosis or apoptosis of the organ's cells, the contents of which are released into the bloodstream. These unique biomarkers can then be sampled from the blood and assessed for specific toxicities. Because biomarkers are quantifiable they can be used to assess the severity of the toxicity. The biomarkers that were frequently reported in drug formulation literature are briefly discussed below:

2.3.2.1 Hepatotoxicity

The biomarkers most commonly used to assess drug induced liver injury (DILI) are alanine transaminase (ALT) and aspartate transaminase (AST). Whilst AST is found in other tissue types, ALT is predominantly found in the liver and is therefore a more specific biomarker of liver toxicity. The ratio of AST to ALT is sometimes used to differentiate liver injury from other organ damage. Elevated levels of AST and ALT are clinical indications of liver dysfunction, although ALT levels are sometimes more elevated than AST in certain types of injury, for example in hepatitis [63]. ALT and AST levels do not always correlate well with histopathology data and so alternative biomarkers can be assessed [23, 64, 65].

Glutathione *S*-transferase (GST) is an inducible phase II detoxification enzyme that catalyses the conjugation of glutathione to phase I metabolites. GST is found uniformly distributed around the centrilobular region of the liver whereas the aminotransferases are principally found in the periportal hepatocytes. Furthermore GST shows good correlations between its measured levels and histopathological data [63, 65, 66].

Gamma-glutamyl transpeptidase (GGT) is an enzyme that transfers γ-glutamyl functional groups to amino acids. Although present in the kidneys and pancreas, GGT serum levels can be used as a biomarker for hepatobiliary injury and in particular for cholestasis [67, 68].

Sorbitol dehydrogenase (SDH) catalyses the reaction of converting sorbitol to fructose, it is primarily found in the cytoplasm of hepatocytes and can be considered a specific biomarker hepatocellular injury [69].

2.3.2.2 Nephrotoxicity

Kidney injury results in the reduction of the glomerular filtration rate (GFR), which compromises the kidneys' ability to excrete waste, leading to renal failure [70, 71]. Blood-urea-nitrogen (BUN) and serum creatinine (SCr) are the biomarkers frequently used to assess kidney toxicity. Urea is produced from the digestion of proteins and creatinine from the breakdown of creatine phosphate in muscles. Both urea and creatinine are filtered out by the kidneys for excretion. Elevated levels of BUN or creatinine within the blood are indications of a decrease in glomerular filtration rate which in turn is a sign of renal dysfunction or toxicity [33, 72, 73]. A delayed onset in the reduction of GFR is sometimes an issue when using BUN and SCr as biomarkers, making them non-sensitive and non-specific to injury [63, 74]. They are also influenced by a variety of factors including age,

gender, muscle mass and diet making them variable between individuals and sometimes inconsistent within the same individual [71, 74].

The lysosomal brush border enzyme N-acetyl-β-D-glucosaminidase (NAG) is sometimes preferred and abundantly found in proximal tubular cells [63, 72].

Total protein can also be used as a biomarker of kidney injury as damage to the podocytes present in the Bowman's capsule results in a leakage of plasma proteins. If the reabsorption of these proteins reaches a saturation point due to this injury or if damage to the tubular protein reabsorption complex occurs, proteinuria can result despite regular glomerular filtration rate [63, 75, 76].

2.3.2.3 Myelotoxicity

Biomarkers of bone marrow toxicity can be assessed using cytology and flow cytometry data. Leucocytes, erythrocytes and thrombocytes are cells of the blood that are commonly used as biomarkers for bone marrow toxicity [60]. Since blood cells are produced within the bone marrow a reduction in the cell count from baseline levels would suggest some form of bone marrow suppression or toxicity. Flow cytometry methods are able to classify and differentiate between the major cell lines but unable to evaluate cellular morphology. For this reason flow cytometry is often coupled with cytology studies [35, 59, 60].

2.3.3 Histopathology Indications

Histopathology examinations provide qualitative assessments of the changes in cellular morphology as a result of toxicity. The morphological assessment of the gross and microscopic appearance of organs can provide a broad knowledge of the mechanisms of drug toxicity. Where biomarkers may fail to identify the sites of organ injury, histopathological examinations are more transparent to the location of injury and provide visual data relating to the integrity of the cellular architecture [77]. By examining the morphological changes within the cellular network of organs it is possible to identify the toxicity outcomes those changes relate to [33, 34, 59, 68].

Organ weights and body weight changes are also ways to assess the acute and systemic toxic profile of a toxicant when compared to a control [52]. Although not strictly a histopathological assessment, the weight of organs (tissues) can be reflective of organ toxicities [34, 52, 59]. A reduction in organ weight over a period of time when exposed to a toxicant demonstrates necrosis and loss of cellular matter of that organ. This not only

implies organ susceptibility to a toxicant but can also demonstrate the distribution profile of the toxicant *in vivo*.

2.4 Principal Mechanisms of Drug Toxicity

Before a drug can exert any toxic effect it must reach and come into contact with the site at which it produces toxicity. To do this a drug's physicochemical properties must allow it to be absorbed into the body, avoid rapid elimination and travel to this site. Understanding how a drug travels through a body can provide insights into the toxicity profile of a drug. ADME studies are therefore key to understanding the fate and, more importantly, the toxic manifestations of a drug. Once a drug has reached its intended target it must then undergo a series of molecular interactions which ultimately disrupt the architecture of specialised cells often leading to necrosis or apoptosis of the cell. A drug's toxic profile can therefore be broken down into two principal stages that can be explained using the principles of pharmacokinetics and pharmacodynamics.

2.4.1 Pharmacokinetics

Pharmacokinetics (PK) is the study of a drug's fate as it travels through an organism. More specifically it is considered as the study of the time course of a drug's absorption, distribution, metabolism and excretion (ADME). Pharmacokinetic studies are important for understanding the body's actions towards a drug so that effective therapy can be applied. The term toxicokinetics is sometimes used when pharmacokinetic principles are applied to study the relationship between the systemic exposure of a drug and its toxicity [50, 78].

ADME studies provide a scientific model based on experimentation and mathematical interpretation to understand how the body responds when challenged with a drug substance. Essentially toxicity can occur at any of the points along the route the drug takes from the point of administration and absorption through to the point of excretion. The site of drug absorption is important since it allows a drug passage into the body. The most important physicochemical properties that affect the rate of absorption are the lipophilicity and solubility of the drug [79]. Absorption generally occurs via three major routes which include the gastrointestinal tract, the lungs and the surface of the skin and can be considered as non-invasive administration routes. Drugs that are difficult to administer non-invasively can be given by alternative routes, including the intravenous (i.v.), intramuscular (i.m.), subcutaneously (s.c.) and intraperitoneal (i.p.) routes. These sites are therefore susceptible to any toxic effects of the drug and dictate how a drug will enter into

the bloodstream. Absorption through the skin takes a drug into the peripheral blood supply whilst absorption through the lungs leads into the pulmonary circulation. For orally administered drugs that enter through the gastrointestinal tract absorption is via the portal vein which leads into the liver. Intravenously administrated drugs have the advantage of entering systemic circulation instantaneously whilst drugs administered intramuscularly, subcutaneously or intraperitoneally require an absorption step before entering systemic circulation. Once within the bloodstream a drug is carried and distributed around the body where it is able to enter specific organs and disrupt cellular functions. Accumulation of sufficient concentrations of drug within organs can lead to specific organ toxicity and dysfunction.

It is important to note that pharmacokinetic measurements are influenced by experimental protocol and considerations need to be taken for several experimental conditions including the dose, route of administration, and species selection. Several key pharmacokinetic parameters are useful when explaining drug toxicity and are discussed briefly below:

Area under the Curve (AUC): The AUC refers to the area under the plasma concentration time curve and is a measure of a drug's concentration from the point of administration to the point where the concentration in the plasma is considered negligible. It can also be used as a measure of drug levels over time in specific tissues. The AUC is useful when comparing different routes of administration since it indicates how quickly a drug gets into and out of systemic circulation and is a good indicator of total drug exposure. A reduction in AUC could therefore be argued as a means to reduce toxicity [80, 81]. For a typical AUC plot following oral administration see Figure 2.1 below, the C_{max} and T_{max} are also established from AUC plots.

C_{max} and T_{max}: The C_{max} is the maximum concentration that is observed over the measured time period and the T_{max} is the time at which this occurred. The C_{max} is one of the most important pharmacokinetic parameters with regards to drug toxicity, with some toxicological concerns showing better correlations with C_{max} levels than with AUC [82]. If the peak exposure dose of a drug exceeds its toxicity threshold then toxicity will occur (Figure 2.1). Maintaining the C_{max} below the drug's toxicity threshold can prevent toxicity [80]. This rationale can be explained by looking at the therapeutic index (TI) of a drug, which is the ratio of the dose required to produce a toxic effect and the dose required for a therapeutic effect. A drug with a high threshold is easier to manage since the margins between toxicity and therapy are wider and so the in vivo drug concentrations do not need

to be closely monitored. Alternatively, for a drug that has a low therapeutic threshold the margins between safety and toxicity are narrow and the C_{max} value of the administered drug must be closely monitored to avoid any toxicity (Figure 2.1).

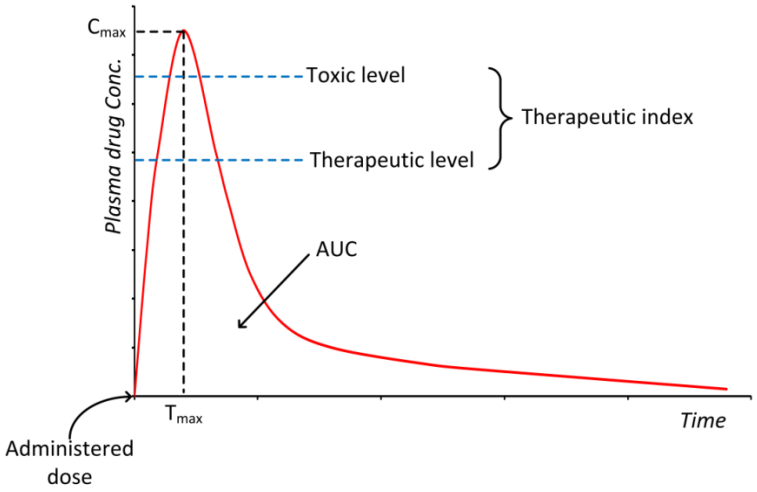


Figure 2.1. Area under the plasma concentration time curve for a single oral dose administered at time zero showing C_{max}, T_{max} and the therapeutic index window. With the C_{max} value exceeding the toxic level it is likely that an adverse outcome will be observed.

Volume of Distribution (V_d): Sometimes referred to as the apparent volume of distribution is the volume of plasma in which the total amount of drug would need to be dissolved to reflect the concentration measured in plasma. It is a reflection of the drug binding in plasma as well as tissues. Since only unbound drug is assumed to pass biological membranes the plasma-protein binding of a drug will affect the volume of distribution. For a drug that has a high volume of distribution, it indicates that a large amount of the drug has reached the intracellular fluid of organs and tissues. This accumulation may leave susceptible organs prone to toxicity issues [83, 84]. Alternatively a drug that results in acute or systemic toxicity and remains largely within the bloodstream, may be safer if it was directed elsewhere [80].

Clearance (Cl): The clearance is a measure of the volume of plasma cleared of drug per unit time. Although the clearance is constant, the amount of drug contained in the clearance volume will change in relation to the plasma drug concentration. It is the sum of all clearance processes such as hepatic metabolism, biliary secretion renal metabolism and renal filtration [85, 86]. Most drugs are eliminated via the liver and kidneys and so hepatic and renal clearances are important in measuring this pharmacokinetic parameter. The clearance rate will have an influence on the mean residence time (MRT) of a drug, which is the average time a drug resides within the body [86]. For a drug that has the potential to

cause toxicity a low clearance rate would augment the situation, allowing the drug to accumulate and exceed its toxicity threshold.

Elimination half-life ($t_{1/2}$): The elimination half-life is defined as the time required for the concentration of a drug to reach half its original value [86]. It is influenced largely by the V_d and the Cl of the drug indicating the half-life can change if these parameters are altered [85]. The half-life is useful to estimate how long therapy may need to be stopped for if toxicity is observed in a patient, assuming the drug follows a linear one-compartmental model. Because the elimination half-life is a quantitative measure of the presence of drug within the body it is commonly used to establish the dosing schedule. The half-life may also help explain the relationship between a drug and tissue accumulation, protein binding and receptor binding, all potential causes of toxicity.

Elimination rate constant (k_e): The elimination rate constant is the amount of change in drug concentration due to elimination over time. It is linked to the V_d given that a large volume would take longer to clear which could translate to toxicity. A zero-order rate constant would suggest that elimination is constant, independent of plasma drug concentration, whereas a first-order rate would indicate a drug's elimination is proportional to its concentration. Although the amount of drug eliminated in a first-order rate changes with plasma drug concentration, the fraction of drug eliminated remains constant. The elimination rate constant therefore represents the fraction of drug eliminated per unit of time. A first-order rate constant would revert back to a zero-order rate when the drug concentration exceeds a limit, for example during an overdose.

Pharmacokinetic parameters describe a dynamic, multi-factorial complex sequence of physiological events that could help explain toxicological outcomes and provide important information which includes:

- The *in vivo* plasma concentration of a drug, which cannot be accurately determined from the dose since not all the dose may be absorbed.
- The concentration of a drug at a particular site which could potentially be a target organ that is susceptible to any toxicity related to that drug.
- The AUC plot of the drug which determines the overall exposure to a drug.
- The maximum *in vivo* concentration reached which can be useful during drug monitoring for adverse effects.

- The elimination kinetics of a drug which determines how long lasting an exposure to a drug may be.

2.4.2 Pharmacodynamics

Pharmacodynamics is the study of the effects a drug has on an organism. To understand toxicity fully we must appreciate both the pharmacokinetic and pharmacodynamic aspects of a drug's toxicity profile. Whilst pharmacokinetic studies are useful for understanding how a drug is absorbed, distributed, metabolised and excreted, they do not fully explain how a drug interacts with biological molecules and membranes to result in a toxic outcome. Pharmacodynamic (or toxicodynamic) studies are useful when explaining the chain of events a drug may instigate leading to some form of toxicity. Upon administration a drug must first reach its site of toxicity before any adverse outcomes can develop. This relationship is explained by the pharmacokinetic parameters described above. Once the drug has reached its intended site a series of biochemical interactions are triggered at the cellular level. These biochemical interactions can be collectively described as an adverse outcome pathway (AOP). For example the conversion of acetaminophen to its toxic metabolite N-acetyl-*p*-benzoquinone imine (NAPQI) by cytochrome P450s leading to the depletion of the endogenous antioxidant glutathione (GSH) allowing NAPQI to interact with biomolecules causing liver toxicity [87-89].

An AOP builds on the more commonly used definition, mode-of-action (MOA), to describe and relate a series of events leading to an adverse outcome. Unlike the MOA, the AOP involves exposure to a chemical and the effects observed on an individual and population level, for example, population decline [90, 91]. Although the literature definition of an AOP is continuously evolving it can be thought of as beginning with an exposure to a chemical, which for *in vivo* studies would be described by the pharmacokinetic parameters of a drug. It ends by describing the impact this exposure will have on a population. The AOP pathway incorporates the toxicity pathway which begins with a molecular initiating event (MIE) and describes the events occurring at a cellular level and the MOA pathway which additionally describes the impact of the toxicity pathway at the organ and individual level [92]. Figure 2.2 conceptualises these pathways showing the overlap of their events and how they can be thought of collectively as an adverse outcome pathway.

Although an AOP spans multiple levels of biological organisation it will initially depend on the pharmacokinetic profile of a drug. For a drug to be cardiotoxic for instance it must first reach the heart before any toxicity pathway can be initiated [93]

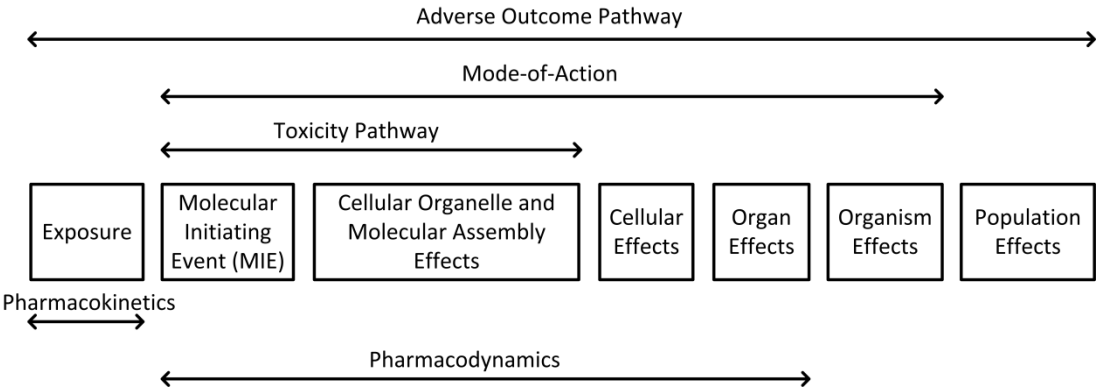


Figure 2.2. Adverse outcome pathway showing the extents of the toxicity pathway and mode-of-action pathway with respect to the toxicity profile of a drug governed by pharmacokinetic and pharmacodynamic principles (Adapted from [90, 91]).

The bioavailability of a drug, which is the amount of a drug reaching systemic circulation and its biodistribution profile, which explains what organs a drug is able to reach, can be explained by the pharmacokinetics of a drug. Once the drug has reached its target organ several key events are described which help sequence the chain of events which subsequently occur. These key events can then be considered as the pharmacodynamics response of a drug. The MIE is the first of these events and is described as the initial interaction between a drug and a biomolecule, such as a ligand-receptor interaction or binding to proteins and nucleic acids [90]. The physicochemical properties of a drug will allow certain MIEs to occur. For this reason the chemical structure of a drug is important and several MIEs are possible, leading to multiple toxicity endpoints. To complicate matters, it is generally accepted that several drugs with vastly different chemical structures may lead to the same toxicity endpoint [92]. Once the MIE has occurred a cascade of downstream events are then activated. Some of these downstream events are described as key events, which relates to a change in biological state that is measurable and leads to the progression of an adverse outcome, for example the depletion of endogenous antioxidants or the activation of inflammatory responses [87, 88, 94, 95]. These key events are therefore quantifiable through some measure and provide a cut off level below which toxicity is not observed [93]. For example in section 2.3.2, biomarkers were discussed as a measure of toxicity. Key events in the progression of liver toxicity can thus be considered as the elevation of the biomarkers ALT and AST. If a significant increase in the concentration of

those biomarkers is detected then sufficient injury to the liver has occurred for there to be a change in the biological state of the liver or its cellular functions. Finally to link the MIE and the key events to an adverse outcome we must define key relationship events. Key relationship events are connecting events that occur throughout an adverse outcome pathway that are not molecular initiating events or key events. They are can be described as the steps that link one key event to another and are identified as one upstream and one downstream event. Through identification of upstream events it may be reasonable to infer or extrapolate the downstream events which may result [90-93].

2.5 Drug Formulation

Active pharmaceutical ingredients (API) are seldom administered alone; rather they are given in combination with one or more non-medicinal agents resulting in a drug formulation. Informed decisions regarding the selection of these agents, sometimes referred to as excipients, can serve specialised functions. Formulations can therefore be considered as carriers that provide some beneficial aid to the drug they carry. This beneficial aid may serve specific functions that improve the therapeutic efficacy of the drug, possibly by improving its solubility, permeability or stability for instance.

Routinely the primary objective of a drug formulation is to ensure the delivery of a drug at a desired location, in the right form, at the correct rate and dose, whilst maintaining the chemical integrity of the compound. Once the pharmacokinetic and pharmacodynamics profiles of a drug are known, formulators can then begin to address innate deficiencies that may hinder the effectiveness of a compound. Formulations therefore profoundly impact the ADME profile of a drug and the associated pharmacodynamic response. Ideally a drug that has been formulated would be dosed orally due to high patient compliance although in practice this is not always possible. Formulation approaches therefore consider administration routes since this can impact the adverse outcome a drug may exhibit.

In the first instance a formulator will make decisions based solely on the molecular properties of a drug (log *P*, pKa, molecular weight) ensuring that the formulation is at least compatible with the drug in question and the excipients used are biocompatible. To add a dimension of complexity to the problem, formulating for toxicity reduction then considers the *in vitro* and *in vivo* toxicity data to ensure the formulation can bring an acceptable element of safety with the clinical use of the drug.

2.6 Drugs with Toxicities Reduced by Formulation

All drugs will present some kind of toxicity provided a sufficient dose is administered [71]. The toxicity of many well studied drugs follows a predictable time course and can be regulated by closely monitoring the dose an individual is exposed to. If an individual shows any signs of toxicity then the dose is reduced or stopped until they sufficiently recover. In principle this practice does work but is tedious in providing effective therapy to a patient. Reducing the administered dose to overcome toxicity may inadvertently reduce the dose to a concentration that is therapeutically ineffective.

Formulating a drug to reduce its toxicity is therefore an interesting challenge for the pharmaceutical community. Formulations that address poor solubility or permeability issues are commonplace within the industry [8, 9] but formulations that avert or reduce toxicity are much less understood and studied.

The decision to formulate for toxicity reduction must arise from the need to allow the safer use of therapeutically effective drugs that are hindered by their toxicity profiles and for which no suitable alternative exists; it is therefore not surprising that cytotoxics and anti-infectives are the major classes of the compounds that have been formulated in this way. Cytotoxics, given their mode of action are naturally some of the most toxic drugs [96, 97] in use today and anti-infectives such as the aminoglycosides are known to be highly neurotoxic and nephrotoxic [98-101]. Other classes of drugs formulated for toxicity reduction include the analgesics [66, 102, 103] and antidiabetics [104].

A list of the more commonly formulated drugs for toxicity reduction that were reported within the literature are listed in Table 2.1. The list provided is by no means exhaustive, rather it is there to provide an overview of the types of drugs that have been formulated for toxicity reduction.

Many of the drugs in Table 2.1 have been formulated using several different approaches, each achieving a reduction in their toxicity when compared to the non-formulated drug.

It is important to state that all the formulation efforts discovered from the literature were either in vitro experiments or in vivo animal experiments. Although in vivo human data for some formulations may exist, they were not easily accessible or widely available.

Table 2.1. List of commonly formulated drugs to ameliorate toxicity and their therapeutic class.

Drug	Therapeutic Class	Drug	Therapeutic Class
Acetaminophen	Analgesic	Hamycin	Anti-infective
Albendazole	Anti-infective	Irinotecan	Cytotoxic
Amphotericin B	Anti-infective	Mitomycin C	Cytotoxic
Annamycin	Anti-infective/Cytotoxic	Mitoxantrone	Cytotoxic
Camptothecin	Cytotoxic	Nimesulide	Analgesic
Cantharidin	Cytotoxic	Nimodipine	Anti-ischaemia
Cisplatin	Cytotoxic	Nobiliside A	Cytotoxic
Clofazimine	Anti-infective	Nystatin	Anti-infective
Clotrimazole	Anti-infective	Paclitaxel	Cytotoxic
Cyclosporin	Immunosuppresant	Propofol	Anesthesia
Diclofenac	Analgesic	Repaglinide	Anti-diabetic
Docetaxel	Cytotoxic	Stavudine	Anti-infective
Doxorubicin	Cytotoxic	Topotecan	Cytotoxic
Gentamicin	Anti-infective	N/A	N/A

2.7 Current Methods of Formulation

The formulation approaches that are shown to reduce drug toxicity are discussed in this section. The range of formulation technologies can be loosely grouped into five different types:

1. Encapsulating formulations – E.g. Liposomes, polymeric micelles
2. Micro/nano formulations – E.g. Disks, spheres, suspensions, emulsions, colloids, dispersions
3. Lipid based formulations – E.g. Liposomes, suspensions, emulsions, colloids
4. Complexation formulations – E.g. Cyclodextrins, dendrimers, liposomes, micelles
5. Antidote formulations – E.g. Vitamins, protein extracts, oils, cyclodextrins

Although the formulations can be grouped into the categories above, there exists an element of cross-over between them with the possibility that formulations can be in more than one category. For example, liposomes can be categorised as encapsulating, micro/nano, lipid based and complexation based formulations.

It is possible to group these technologies into four main strategies by which they operate to reduce toxicity. These mechanisms of action are briefly discussed below:

Strategy 1: Active organ protection; provides toxicity relief through the co-administration of other agents which provide a counter measure to the toxic manifestation of a drug. This strategy largely covers the use of antidotal formulations for toxicity reduction.

Strategy 2: Avoiding susceptible targets; which circumvents toxicity by preventing a toxin reaching an organ or other target that is prone to its toxic effects. If a drug can be encased, encapsulated or diverted so that it cannot interact with such an organ then it is reasonable to assume that any adverse effects can be avoided. This strategy uses encapsulating, complexation and lipid based formulations to achieve its mechanism of action.

Strategy 3: Active Targeting; a strategy which actively targets a drug to the site of infection/disease, thereby accumulating in lesser amounts at sites where unwanted toxicities may occur. This strategy uses encapsulating based formulations to target a drug to the site it is required and away from susceptible tissues.

Strategy 4: Availability Enhancement; if we can enhance the availability of a drug (through improved solubility and permeability for example) this dictates that the dose and therefore the exposure of the drug (toxin) can be reduced. Improving solubility of a drug also prevents toxicities relating to aggregation and precipitation in aqueous environments. This strategy makes use of the encapsulating, complexation, lipid based and nano/micro based formulation technologies.

Of the four strategies, strategy 1 would mainly act upon the pharmacodynamic aspects of a drug's toxicity, whereas strategies 2, 3 and 4 would be governed by the pharmacokinetic aspects of a drug's toxicity profile.

An informative discussion about the preparation, practical use and properties of the commonly used technologies for toxicity reduction is given below:

2.7.1 Liposomes

Liposomes are lipid bilayer spherical vesicles first described by Bangham et al. (1964) [105]. They are composed of phospholipids which consist of a glycerol backbone attached to a polar (hydrophilic) phosphate head group and a non-polar (hydrophobic) tail chain. Tail chains vary in length typically from 3 to 24 carbons. Unsaturated tail chains consist of long

fatty acids with at least one carbon-carbon double bond (C=C). Phosphatidylcholines (PC), also known as lecithin are the principal phospholipids used due to their double fatty acid chain which form bilayers in water in preference to micelles [106, 107]. They are derived naturally from soya bean and egg yolk and their relatively low cost makes them an ideal choice for liposome construction. When placed in water the tail chains repel each other and are forced to aggregate forming a lipid bilayer membrane with the tail chains facing inwards and the head groups facing out (Figure 2.3). The tail chains along the perimeter of this membrane remain exposed to the aqueous environment and have an energy associated with them. To minimise this instability the membrane closes forming a spherical vesicle. A further energy penalty is incurred as the bilayer membrane bends, the energy of the system first increases and then decreases as the edges meet and disappear as additional phospholipids are incorporated into the final spherical vesicle which encapsulates an internal aqueous core (Figure 2.3) [108-110]. Phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and phosphatidylserine (PS) make up the other frequently used phospholipids, with sphingolipids, cholesterol and hydrophilic polymer conjugated lipids also contributing to the structure and properties of liposomes [111].

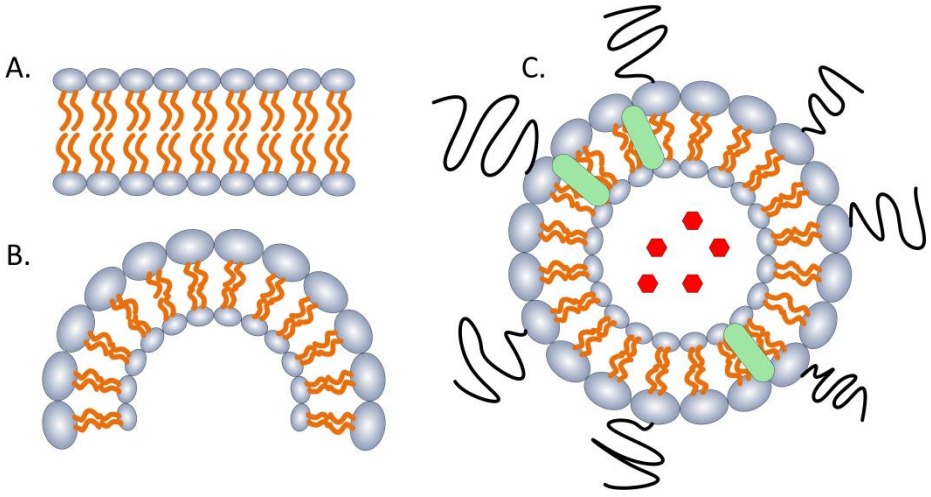


Figure 2.3.Phospholipid assembly in liposome formation. (A) Phospholipid bilayer membrane with outward facing hydrophilic head group and inward facing hydrophobic tail chains. (B) Membrane closing to minimise instability of perimeter tail chains exposed to aqueous environment. (C) Complete assembly of liposome encapsulating hydrophilic drugs (red) within the core and hydrophobic drugs (green) within the lipid bilayer. Attached polymers onto the surface of the liposome allow them to avoid RES uptake and provide targeting abilities.

Using naturally occurring phospholipids in their construction means they are generally considered biologically inert, biodegradable, weakly immunogenic with low intrinsic toxicity [109, 112]. Due to their amphiphilic nature liposomes are able to entrap hydrophilic drugs within their aqueous core whilst hydrophobic drugs are carried within their lipid bilayer

(Figure 2.3). Liposomes may also be used for the effective encapsulation and delivery of drugs intended for topical application, peptides and nucleic acids [113-120].

Liposomes can be formulated to differ in size, charge, composition and lamellarity. Typically they range in size from approximately 20 nm to several micrometres [108, 121]. They can be classed in several ways according to their structure, size, function, composition and preparation method [112]. Uni-lamellar vesicular (ULV) liposomes are constructed of a single membrane bilayer and range from 20 nm to 100 nm in diameter. Multi-lamellar vesicles (MLV) and multi-vesicular (MVL) liposomes are generally larger with a diameter range of up to 1 μm [108, 109, 112]. The choice of liposomal bilayer components influences properties such as the rigidity, fluidity, charge density, permeability, release kinetics and targeting ability with biological components upon administration. For instance unsaturated phospholipids used in the construction of liposomes result in more permeable less stable bilayers, a result of the carbon-carbon double bonds creating a kink in their tail chains which prevents tight packing [108]. Incorporating PS into the bilayer helps to increase the aqueous space of liposomes [122] whilst cholesterol influences the fluidity of the membrane, increasing its rigidity, resulting in a longer retention and slower release rate of encapsulated drugs *in vivo* [123]. For improving the circulation times of liposomes, lipids are conjugated to polymers such as polyethylene glycol (PEG) [51, 109, 124]. Such surface coatings can offer steric hindrance to opsonin adsorption onto bilayers reducing their uptake and removal from circulation by the cells of the reticuloendothelial system (RES) [125, 126]. Ligands can also provide targeting properties to liposomes, allowing more efficient therapy by targeting encapsulated drugs to areas of therapeutic requirement. This also reduces drug wastage by allowing drugs to reach sites not normally accessible by the free drug whilst also avoiding elimination by the RES [35, 127-129].

2.7.1.1 Liposomes in Toxicity Formulations

Liposomes are extremely diverse in structure and capabilities and as a result they have been shown to reduce the toxicity of drugs by the methods described in strategies 2, 3 and 4 of section 2.7. Liposomes can be considered as complexation based, encapsulation based, lipid based and micro/nano based formulations (see section 2.7). Due to the diverse nature of liposome construction many objectives can be fulfilled through liposomal formulation. Liposomal technology used for toxicity reduction has been shown to improve the solubility and permeability of difficult to administer drugs, provide controlled release capabilities,

target drugs to sites of toxicity and also prevent the interaction of a drug with susceptible tissues.

Many classes of drugs have been formulated for toxicity reduction through the use of liposomes. Some of these classes include: antibacterial aminoglycosides, antifungal polyenes and cytotoxic anticancer drugs.

Aminoglycosides are an interesting class of drug that have been successfully formulated for toxicity reduction through liposomal encapsulation. Aminoglycosides such as netilmicin, gentamicin, tobramycin, amikacin, kanamycin, streptomycin, and sisomicin have all been shown to be potent nephrotoxins *in vivo*. They accumulate within the renal cortex, although concentrations are not necessarily correlated to toxicity [130, 131]. Retention of aminoglycosides within the epithelial cell lining of the S1 and S2 segments of the proximal tubules after glomerular filtration results in morphological and functional changes which lead to renal failure and can be detected as a gradual rise in serum creatinine [132, 133].

To mitigate this toxicity several studies have provided evidence for the successful formulation of aminoglycosides with liposomes. It was previously reported that cationic polybasic aminoglycosides have a high affinity for anionic acidic phospholipids such as phosphatidylinositol (PI) found on renal brush border membranes [134]. Mimoso et al. (1997) [62] therefore prepared liposomes to encapsulate netilmicin with a lipid composition of phosphatidylcholine:cholesterol:phosphatidylinositol, at a molar ratio of 5:1:4. When administered intravenously into mice a reported LD_{50} value of >50 mg/kg was reported for the formulated drug when compared to the non-formulated value of 24 ± 4 mg/kg. The authors proposed that a reduction in toxicity was due to the inclusion of phosphatidylinositol into the liposomal membrane which competed for the drug over renal phosphatidylinositol, thereby ensuring the drug avoided interaction with the kidneys. The inclusion of phosphatidylcholine also improved the encapsulation efficiency of the liposome for netilmicin. Pharmacokinetic data reported by the group demonstrated an increase in the half-life ($t_{1/2}$) of the formulated drug and a decrease in the volume of distribution (V_d) when compared to the non-formulated drug. Similar pharmacokinetic profiles were reported for the liposomal encapsulation of gentamicin [135] (Table 2.2).

Table 2.2. Pharmacokinetic data of liposomal encapsulated netilmicin and gentamicin. V_d units – L/kg* or mL (at steady state)†. Data adapted from [62, 135].

Drug/Formulation	Species	Admin route	Dose (mg/kg)	AUC (μg.h/mL)	$t_{1/2}$ (min)	V_d	Reference
Netilmicin (non-formulated)*	Mice	i.v.	4.0	-	12.1	15.7	[62]
Netilmicin-Liposome*	Mice	i.v.	4.0	-	137.6	2.9	
Gentamicin (non-formulated)†	Mice	i.v.	20.0	8.08	61.8	3.68	[135]
Gentamicin-Liposome†	Mice	i.v.	20.0	185.25	228.6	0.59	
Gentamicin (non-formulated)†	Rats	i.v.	20.0	37.51	36.0	0.47	[135]
Gentamicin-Liposome†	Rats	i.v.	20.0	426.55	241.2	0.28	

These data suggest liposomal encapsulation of aminoglycosides markedly alters the pharmacokinetics and reduces toxicity by diverting the drug away from the kidneys, as the reduction in the V_d and increase in AUC compared to free drug indicate. A lower V_d demonstrates the encapsulated drug spends more time in systemic circulation rather than distributed amongst peripheral tissue and organs. This formulation would therefore satisfy the mechanism of action suggested by strategy 2 of section 2.7, whereby toxicity is reduced by avoiding accumulation within susceptible organs (i.e. the kidneys).

Bermudez et al. (1990) [136] reported similar acute toxicity improvements when formulating the aminoglycoside amikacin into liposomes. They proposed that although liposomes are taken up mainly by the liver and spleen, amikacin would be slowly released by phagocytic cells which would result in a high sustained level of drug within the serum; which reflects the pharmacokinetic data reported by Mimoso et al. (1997) [62]. Furthermore biodistribution studies of the aminoglycoside gentamicin encapsulated into liposomes have shown accumulation to be prevalent within the liver but minimal amounts within the kidneys when administered intravenously *in vivo* using animal models [124, 135, 137], reflecting the nephroprotection offered through liposomal encapsulation of aminoglycosides.

Polyene antifungal drugs such as amphotericin B, nystatin and hamycin have also benefited from encapsulation through liposomal formulation. Amphotericin B is currently the only antifungal treatment for systemic mycotic infections in man [138, 139]. However acute toxic adverse effects including, fever, nausea, chills, emesis and hypokalaemia limit the use of the drug. In addition nephrotoxicity, detected by an increase in serum creatinine and urea becomes apparent with continued use resulting in irreversible kidney impairment

[140]. To attenuate this toxicity, several studies have reported promising results using liposomes to ameliorate the toxicity of polyene drugs. For example Szoka et al. (1987) [141] prepared several amphotericin B encapsulating liposomal formulations with different lipid compositions and reported that cholesterol-containing liposomes offered the greatest toxicity protection when injected i.v. into mice, with an LD_{50} of 19 mg/kg compared to 2.3 mg/kg for the non-formulated drug. Interestingly the size of the liposome proved significant, with larger cholesterol-containing liposomes and multilamellar liposomes producing lethality at a lower dose than small unilamellar liposomes.

Acute toxicity studies of liposomal encapsulated amphotericin B that incorporated cholesterol and polyethylene glycol derivatised distearoyl phosphatidylethanolamine (PEG-DSPE) have also been reported with a MTD of the formulated drug to be 13 mg/kg when injected i.v. into mice whereas for the non-formulated drug a MTD of 0.8 mg/kg was recorded [51]. The inclusion of PEG-DSPE was responsible for an increased blood residence time of the liposomes which may have prevented amphotericin B from reaching the kidneys in sufficient enough concentrations to cause toxicity, which suggests that their mechanism of action for toxicity prevention is via strategy 2 discussed in section 2.7. Other studies have supported these findings, providing pharmacokinetic data that shows a reduced accumulation of drug within the kidneys but an increased accumulation within the liver and spleen [140].

From these data it was proposed that liposomal encapsulation reduces the toxicity of amphotericin B by three principal mechanisms: 1) it changes the interaction of amphotericin B with its receptors. It is known that amphotericin B binds preferentially to ergosterol in fungal cells and causes toxicity in animals by binding to cholesterol in mammalian cells [139, 142, 143]. The inclusion of cholesterol into liposomal formulations therefore competes with cholesterol in mammalian cells for amphotericin B, thereby reducing toxicity, 2) it changes the organ distribution of the drug as shown by a reduced accumulation within the kidneys and 3) it alters the rate at which the drug is released from the liposome since cholesterol inclusion provides rigidity and stability, it slows down the rate at which the drug is released from the liposome and becomes free to reach its target receptors.

Other polyenes that have benefited from liposomal encapsulation include hamycin and nystatin. Cholesterol inclusion into liposomal formulations for hamycin were shown *in vitro* to reduce RBC lysis toxicity and *in vivo* improve the MTD of the drug when administered i.v.

to mice [144]. A relatively simplistic liposomal formulation using dimyristoyl phosphatidylcholine (DMPC) and dimyristoyl phosphatidylglycerol (DMPG) to encapsulate nystatin was also shown to reduce the lysis of RBCs *in vitro* when compared to the non-formulated drug. Interestingly the addition of empty liposomes to non-formulated drug significantly reduces the lysis of RBCs *in vitro*. This liposomal formulation of nystatin administered i.v. to mice resulted in a MTD of 16 mg/kg compared to the non-formulated drug which has a MTD of 4 mg/kg [145, 146].

Finally, cytotoxic anticancer drugs are an interesting group of compounds for which liposomal formulation has been shown to improve their therapeutic index and reduce toxicity. Of particular interest are the anthracycline antibiotics that are effective anti-neoplastic agents used in the treatment of adult and paediatric cancers. Anthracyclines such as doxorubicin, epirubicin and annamycin are however limited by toxicities such as haematopoietic suppression, emesis, nausea, alopecia and dose related cardiotoxicity [147]. Although multiple mechanisms are involved in the toxicity pathway of anthracycline induced cardiotoxicity, oxidative stress – resulting in lipid peroxidation and depletion of antioxidants which produces a loss of myofibrils and vacuolisation of myocardial cells – seems to be one of the most common causes of cardiomyopathy and heart failure.

Several studies on the liposomal encapsulation of doxorubicin have reported positive findings. Herman et al. (1983) [148] developed positively charged cardiolipin liposomes to encapsulate doxorubicin. Upon administration of a 1.75 mg/kg i.v. dose of the formulated and non-formulated drug given at 3-week intervals over a period of 21 weeks (7 injections) to five beagle dogs they reported a reduction in food consumption and a significant loss in body weight of the non-formulated group. This group also showed signs of toxicity upon administrations, by shaking, staggering and lying down, which was also apparent in the formulated group but to a lesser extent. Alopecia was noted in the non-formulated dogs but was entirely prevented in the liposomal formulation of doxorubicin. This study also reported heart histopathology data and showed lesions consisting of vacuolisation and myofibrillar loss was noted in all 5 dogs of the non-formulated group. In contrast, there were no abnormalities found in the liposomal encapsulated group. No evidence of tissue abnormalities were found in the lungs, liver, kidneys, spleen or small intestine of formulated and non-formulated drug. Bone marrow suppression between the groups was also comparable as a reduction in WBC, RBC, haemoglobin and haematocrit was reported.

Similar data on the cardiotoxicity protection offered though liposomal encapsulation of doxorubicin and annamycin have been reported in other *in vivo* studies using mice and rat models [60, 149, 150]. van Hoesel et al. (1984) [149] additionally reported on the biodistribution of the negatively and positively charged liposomes that were developed for doxorubicin encapsulation. They showed that doxorubicin levels in the heart and kidneys through liposomal formulation were reduced compared to the non-formulated drug which partly explains the cardioprotective nature of doxorubicin liposomal encapsulation (Table 2.3).

From the data in Table 2.3 it becomes apparent that both the negatively and positively charged liposomes were significantly able to reduce concentrations within the heart over a 24hr period when compared to the non-formulated doxorubicin.

Table 2.3. Biodistribution of liposome encapsulated doxorubicin following an i.v. injection of 1 mg/mL into rats. Plasma concentration expressed in µg/L and tissue concentration expressed in mg/kg tissue weight. Data adapted from [149].

Tissue	Doxorubicin (non-formulated)	Doxorubicin in -ve liposomes	Doxorubicin in +ve liposomes
Plasma (4hr)	5.89 ± 0.56	2.04 ± 0.50	35.0 ± 4.06
Heart (4hr)	6.72 ± 0.67	2.53 ± 0.43	0.74 ± 0.26
Kidneys (4hr)	12.97 ± 1.00	4.41 ± 0.67	1.26 ± 0.51
Liver (4hr)	3.76 ± 0.56	18.3 ± 2.5	1.44 ± 0.36
Spleen (4hr)	4.53 ± 0.88	54.6 ± 12.0	2.63 ± 0.93
Plasma (24hr)	2.84 ± 0.57	2.03 ± 0.39	0.44 ± 0.25
Heart (24hr)	3.28 ± 0.86	0.66 ± 0.37	0.58 ± 0.04
Kidneys (24hr)	5.58 ± 1.10	1.76 ± 0.28	0.24 ± 0.05
Liver (24hr)	1.13 ± 0.10	30.2 ± 10.0	1.06 ± 0.42
Spleen (24hr)	4.02 ± 0.57	2.04 ± 0.33	5.06 ± 1.89

Interestingly whilst negatively charged liposomes seemed to accumulate within the liver and spleen, positively charged liposomes remained within the plasma at the 4hr stage but over a 24hr period also accumulated within the liver and spleen.

Liposomal formulation of doxorubicin also provides a good example of targeted delivery for toxicity reduction which corresponds to strategy 3 of the section 2.7. The theory suggests that if a drug can be specifically targeted to a site of disease then it will naturally accumulate to lesser amounts in susceptible tissue thereby avoiding toxicity. This theory was tested by Li et al. (2009) [127], who developed stealth liposomes containing polyethelyene glycol (PEG) and the ligand transferrin. Liposomes containing PEG are able to

avoid uptake by the RES, whilst the transferrin receptor is known to be highly expressed at tumour sites [127]. *In vivo* pharmacokinetic studies of these liposomes in health rats revealed significant changes in the AUC and clearance of the drug from the plasma when a single dose of 5 mg/kg was administered i.v. (Table 2.4).

Table 2.4 shows liposomal encapsulation of doxorubicin with or without conjugated transferrin ligands are able to significantly increase the AUC, MRT and terminal half-life ($t_{1/2}$ (β)) of doxorubicin. The volume of distribution (V_d) and plasma clearance (Cl) was also reduced suggesting doxorubicin in liposomal formulations remained within systemic circulation compared to non-formulated doxorubicin.

Table 2.4. Pharmacokinetics of doxorubicin encapsulated into liposomes containing PEG and transferrin ligands. Rats dosed at 5 mg/kg i.v. Data adapted from [127].

Drug/Formulation	Species	Admin route	AUC (μg/ml) h	V_d (L/kg)	Cl (L/(kg h)	MRT (h)	$t_{1/2}$ (β) (h)
Doxorubicin (non-formulated)	Rats	i.v.	0.538 ± 0.145	4.000 ± 1.000	50.000 ± 1.500	2.431 ± 0.050	5.476 ± 1.690
Doxorubicin-Liposome (containing PEG)	Rats	i.v.	1276.458 ± 195.444	0.150 ± 0.025	0.020 ± 0.005	23.464 ± 1.128	20.584 ± 0.905
Doxorubicin-Liposome (containing PEG & transferrin)	Rats	i.v.	1221.262 ± 80.795	0.205 ± 0.015	0.020 ± 0.000	22.585 ± 2.594	22.238 ± 2.059

To determine whether these changes in pharmacokinetics would also confer better targeting capabilities, the formulations were injected i.v. into tumour bearing mice (HepG2 cells) at a dose of 5 mg/kg. The AUC for various tissues samples was reported over a 96 hour period (Table 2.5).

Table 2.5. Biodistribution study of liposome encapsulated doxorubicin. Tumour bearing mice were administered an i.v dose of 5 mg/kg. AUC tissue units – h.μg/g

Tissue	Doxorubicin (non-formulated)	Doxorubicin-Liposome (containing PEG)	Doxorubicin-Liposome (containing PEG & transferrin)
Heart	114.4	77.6	87.2
Liver	105.8	190.9	252.9
Spleen	180.9	269.3	253.0
Lung	74.5	121.1	110.2
Kidney	120.0	110.4	105.1
Tumour	46.1	166.1	238.7

The biodistribution data in Table 2.5 provides some interesting insights into these liposomal formulations. Both liposomal formulations (with and without transferrin ligands)

showed a significant reduction in the accumulation of drug in the heart compared to non-formulated doxorubicin and would therefore reduce the cardiotoxic burden of the drug. The liver, spleen and lung concentrations were also significantly different for the non-formulated drug compared to the two liposomal formulations. These organs all showed a larger accumulation compared to the non-formulated drug, which possibly suggests a redistribution of drug that would normally accumulate within the heart but is now accumulating within the liver, spleen and lungs. The differences in kidney accumulation were insignificant between the three different administrations. Of particular interest was the accumulation of drug within the inoculated tumour. The study reports that accumulation of drug within the tumour of mice administered liposomes containing both PEG and transferrin, was significantly higher than the non-formulated drug and the PEG only liposomes. These data suggest that transferrin conjugated liposomes encapsulating doxorubicin are able to successfully target tumour sites and avoid accumulation within the heart, thereby reducing the cardiotoxicity of the drug and improving the efficacy [127].

The avoidance of liposomal encapsulated doxorubicin to accumulate within the heart explains the reduction in cardiotoxicity seen with the many liposomal formulations of doxorubicin described above and as such complies with strategies 2 and 3 discussed in section 2.7.

From the liposomal drug formulations that have been shown to reduce drug toxicity it is evident that difference in liposomal charge, size and composition and conjugation are important in achieving toxicity reductions. There are insufficient data to derive a general rule that can be applied to liposomal composition which would imply toxicity protection to a wide variety of drugs, so currently liposomal formulations have to be assessed on a case by case basis.

2.7.2 Polymeric Micelles

Micelles are small spherical structures that spontaneously form when amphiphilic surfactant molecules are placed in water. As with phospholipids aggregating to form liposomes, surfactant molecules in water form aggregates (clustered monomers) with their hydrophilic heads in contact with the water orientating their hydrophobic single tails within the micelle centre forming a core-shell structure [106, 107]. Their ability to self-associate is a result in the decrease in free energy of the system. As the concentration of surfactant molecules within a system increases it approaches a critical micelle concentration (CMC)

[151] where unfavourable energy sequesters the hydrophobic tails away from water and begin to form micelles. Surfactants forming micelles show a strong tendency to do so above the critical micelle temperature, termed the Krafft point (T_k). A single surfactant molecule has limited solubility and below the Krafft point micellization is poor. As the temperature of the system increases however, the solubility increases until the CMC is reached and micelles are formed [152, 153]. Surfactant based micelles however tend to have a relatively high CMC rendering them unstable upon high dilution. For this reason they are sometimes unsuitable for use in biological systems as drug carriers.

Polymeric micelles have therefore emerged as the micelles of choice for use in biological systems. Due to their low CMC values they are less prone to disintegration upon dilution and so make an ideal biomaterial which maintains its integrity in aqueous environments. They are formed from amphiphilic block copolymers that contain a hydrophilic and hydrophobic block which self-assemble to form micelles in aqueous solutions above their CMC [154] (Figure 2.4). As the CMC is reached, the hydrophobic segments of the copolymer associate to minimise contact with water molecules resulting in a hydrophobic core with an outward extended hydrophilic shell known as the corona [151]. Large core-forming segments of block copolymers exhibit strong aggregation tendencies and so high thermodynamic stability. High aggregation of core forming segments may result in high hydrophobic or electrostatic interactions within the core, resulting in lower CMC values making polymeric micelles kinetically stable [151, 155], it has also been suggested that polymeric micelles may not exist in equilibrium with their amphiphilic block copolymer monomers [156] retaining integrity even under high dilution. The hydrophobic cores of polymeric micelles are therefore able to encapsulate hydrophobic drugs with the hydrophilic corona offering the drug protection against an aqueous environment (Figure 2.4). This aids poorly water soluble drugs to overcome bioavailability and solubility issues as well as drug aggregation problems *in vivo* [157]. The predominant shape of micelles is spherical although rods, tubules, needles and hexagonally packed hoops have been reported too [151, 158]. Amphiphilic diblock copolymers are most frequently used to prepare polymeric micelles although triblock and grafted copolymers are other options. The choice of amphiphilic copolymers determines the differences observed in micellar morphologies [159]. Block copolymers are linear polymers of the A-B (diblock) or A-B-A (triblock) arrangement (where A = a hydrophilic block and B = a hydrophobic block) whereas grafted copolymers consist of one copolymer side chain grafted to the main copolymer [160]. Appropriate choice of copolymers determines the physiological and

biological properties of polymeric micelles; this may include lower CMC values which improves stability, increased circulation times, evasion from RES, high drug entrapment and targeted delivery [161-165].

Block copolymers that have been used in drug formulation include Poly(N-vinylpyrrolidone)-*block*-poly(D,L-lactide) as a solubiliser for hydrophobic anticancer drugs [81], poly(ethylene glycol)-*block*-*poly*(ε-caprolactone-co-trimethylenecarbonate) as an encapsulation method for amphotericin B [26] and poly(ethylene glycol)-*block*-poly(D,L-lactic acid) for the encapsulation of multiple poorly soluble drugs including paclitaxel, etoposide and docetaxel [161]. Methods of preparation can vary but for poorly soluble drugs the solid dispersion method, oil-in-water emulsion solvent evaporation method or the dialysis method can be used [165].

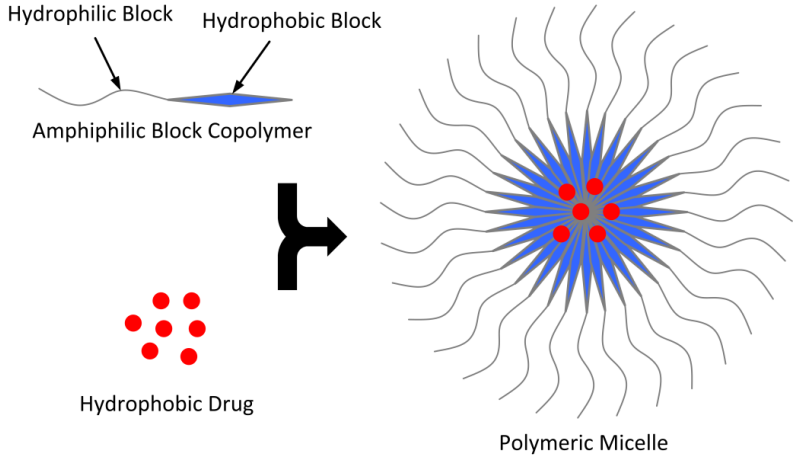


Figure 2.4. Polymeric micelle construction from amphiphilic monomer units. Amphiphilic block copolymer units made of both hydrophilic and hydrophobic blocks self-assemble to form polymeric micelles entrapping hydrophobic drugs within their core, shielded from an aqueous environment by the outer corona shell.

2.7.2.1 Polymeric Micelles in Toxicity Formulations

Polymeric micelles are useful solubilising agents and as such have been used in the formulation of poorly soluble drugs. Furthermore they are now being studied as alternatives to the solubilising agents currently used in practice today, since the solubilising agents themselves can sometimes confer toxicity. Some of the most poorly soluble drugs in use today are the anticancer drugs for which suitable solubilising agents are difficult to find. Traditionally they are formulated using low molecular weight surfactant such as polyethoxylated castor oil (Cremophor EL) or polysorbate 80 (Tween 80). These surfactants however present clinical concerns of their own such as hypersensitivity and neurotoxicity

[166]. Polymeric micelles have therefore emerged as an alternative for the solubilisation of poorly water soluble drugs that have toxicity concerns. Polymeric micelles can achieve high drug loading and protect drugs from premature degradation *in vivo* due to their core shell structure that can shield drugs from metabolism.

Diblock copolymers are particularly well suited for polymeric micelle construction and several studies have demonstrated their capability to solubilise paclitaxel, an anti-cancer drug of the taxene family. Kim et al (2001) developed a polymeric micellar system using the biodegradable amphiphilic diblock copolymer, monomethoxy poly(ethylene glycol)-*block*-poly(D,L-lactide) (mPEG-PDLLA) to solubilise paclitaxel [80]. This polymeric micellar formulation of paclitaxel (PTX-mPEG-PDLLA) was compared to the current clinical formulation of Cremophor EL/ethanol (50:50), known as Taxol, for toxicity, pharmacokinetic and biodistribution differences.

The MTD of PTX-mPEG-PDLLA was assessed in vivo in mice and determined to be 50 mg/kg, in comparison to Taxol which resulted in a MTD of 20 mg/kg. Similarly acute toxicity studies to establish the LD_{50} of the formulations in male rats showed PTX-mPEG-PDLLA to have an LD_{50} of 205.4 mg/kg, compared to 8.3 mg/kg for Taxol.

To understand these toxicity differences the pharmacokinetic behaviour and biodistribution of these formulations was studied *in vivo* in mice bearing murine B16 melanoma cells. Following a single i.v. injection of 50 mg/kg of PTX-mPEG-PDLLA or 20 mg/kg of Taxol the pharmacokinetic profile was fitted by a two-compartment model (Table 2.6).

Table 2.6. Pharmacokinetic data following i.v. administration of PTX-mPEG-PDLLA formulation compared to Taxol in mice. Data adapted from [80].

Drug/Formulation	Admin route	Dose (mg/kg)	$AUC_{0-\infty}$ (µg.h/mL)	C_{max} (µg/mL)	T_{max} (h)	$t_{1/2}(\beta)$ (h)	Cl (L/h/kg)	V_d (L/kg)
Taxol (non-formulated)	i.v.	20.0	94.38	94.08 ± 3.54	0.05	0.34	0.22	0.27
PTX-mPEG-PDLLA	i.v.	50.0	69.75	82.83 ± 4.16	0.05	0.21	0.72	0.88

The pharmacokinetic data in Table 2.6 shows that despite the larger dose of administered PTX-mPEG-PDLLA, the AUC and C_{max} values are in fact lower when compared to Taxol. The authors of the study also suggested that given an equivalent dose of 20 mg/kg, the AUC of PTX-mPEG-PDLLA was shown to be 10 to 20-fold lower than for Taxol, although organ

distributions remained comparable. These data however were not reported in the literature [80].

A similar study solubilising paclitaxel using polymeric micelles made of the diblock copolymer, poly(N-vinylpyrrolidone)-*block*-poly(D,L-lactide) (PTX-PVP-PDLLA) reported similar findings when comparing their micellar formulation to Taxol. The authors determined the MTD of Taxol in mice to be 20 mg/kg whilst their micelle formulation, PTX-PVP-PDLLA, resulted in a MTD of >100 mg/kg. The pharmacokinetic data reported in this study are from equivalent doses of 20 mg/kg, administered i.v. to mice bearing C26 tumours. A non-compartmental open model was fitted (Table 2.7) [81].

Table 2.7. Pharmacokinetic data following i.v. administration of PTX-PVP-PDLLA formulation compared to Taxol in mice. V_d - at steady state, C_0 - extrapolated peak plasma concentration. Data adapted from [81].

Drug/Formulation	Admin route	Dose (mg/kg)	$AUC_{0-\infty}$ (μg.h/g)	C_0 (μg/g)	$t_{1/2}$ (h)	Cl (L/h/kg)	V_d (L/kg)
Taxol (non-formulated)	i.v.	20.0	91.8 ± 10.9	199.2	1.05	0.21	0.22
PTX-PVP-PDLLA	i.v.	20.0	18.6 ± 3.7	45.4	0.56	1.02	0.78

The pharmacokinetic data in Table 2.7 show that at equivalent doses the micellar formulation (PTX-PVP-PDLLA) produces a lower AUC and extrapolated peak plasma concentrations (C_0). Biodistribution data from this study also show that PTX-PVP-PDLLA is able to reduce the concentrations of the drug accumulating within various organs when compared to Taxol (Table 2.8).

Table 2.8. Biodistribution study of PTX-PVP-PDLLA following i.v. administration of 20 mg/kg into mice. $AUC_{0.25-24}$ - h.μg/g. Data adapted from [81].

Tissue	Taxol (non-formulated)	PTX-PVP-PDLLA
Liver	301.9 ± 35.8	260.1 ± 25.0
Spleen	83.8 ± 12.2	46.3 ± 55.4
Kidney	162.3 ± 14.8	61.5 ± 8.6
Lung	99.8 ± 14.7	58.1 ± 5.2
Heart	94.8 ± 9.8	30.3 ± 3.3
Muscle	35.4 ± 3.4	20.7 ± 1.4

At an equivalent dose of 20 mg/kg, the PTX-PVP-PDLLA micellar formulation resulted in less paclitaxel accumulation within all the organs, in comparison to the clinical formulation Taxol.

The results from these studies suggest that a micellar formulation of paclitaxel (PTX-mPEG-PDLLA or PTX-PVP-PDLLA) is able to reduce drug accumulation within organs, although the volume of distribution was reported to increase marginally when compared to Taxol. A reduction in AUC, peak plasma concentration and terminal half-life is observed with these formulations whilst the clearance of the drug is slightly increased. Although the pharmacokinetic and biodistribution data may suggest that the drug is removed quickly from systemic circulation and this naturally confers to an improved toxicity profile, in actual fact the changes in the pharmacokinetic and biodistribution of the drug through micellar formulation allow for a larger dose to be administered without observing any toxicities and would therefore improve the efficacy of the drug. Certainly *in vitro* the micellar formulation was shown to be equipotent against murine C26, EMT-6 and human OVCAR-3 cancer cell lines [81]. The gain in MTD must therefore result in better tolerability of the micellar formulation compared to the Cremophor EL/ethanol formulation.

Other paclitaxel carrying micelles that have been developed include amphiphilically modified chitosan derivatives, consisting of hydrophobic alkyl chains and hydrophilic glycol groups conjugated to a chitosan backbone (PTX-OGC). These micelles were able to achieve a high drug loading capacity and produced no hypersensitivity reactions in guinea pigs when compared to an equivalent dose of Taxol which resulted in anaphylaxis upon administration. The MTD of PTX-OGC micelles in mice was reported to be 100 mg/kg compared to 25 mg/kg for Taxol and the LD_{50} reported as 123.1 mg/kg and 50.7 mg/kg for PTX-OGC and Taxol respectively. Intravenous administration of PTX-OGC at a dose that is half its LD_{50} value resulted in no histopathological changes in the heart, liver, spleen, lung, and kidneys of mice. *In vitro* cytotoxicity studies using human hepatoma HepG2 cells showed micelles without paclitaxel (OGC micelles) to be non-toxic however the Taxol formulation of Cremophor EL/ethanol without paclitaxel resulted in cytotoxicity. The cytotoxicity of drug loaded micelles (PTX-OGC) resulted in an IC_{50} value of 0.28 (± 0.14) µg/mL which was lower than that of Taxol whose IC_{50} was 0.79 (± 0.36) µg/mL. Although Taxol shows strong cytotoxicity it may be partially attributed to the formulation of Cremophor EL/ethanol but in contrast the cytotoxicity of PTX-OGC micelles is probably due to the drug alone [167].

Amphotericin B as discussed in section 2.7.1.1 is a polyene antifungal that presents local and acute toxicity concerns, such as phlebitis, fever, chills and nephrotoxicity. These toxicities are partly related to its affinity for cholesterol in mammalian cell membranes but

also due to its relatively poor aqueous solubility which results in a toxic aggregated form of the drug [168]. To overcome this toxicity due to aggregation, several polymeric micelle formulations have been developed to improve the solubility of the drug. Clinically, amphotericin B is administered in sodium deoxycholate (Fungizone), a bile acid that behaves as a detergent to solubilise the drug, although haemolytic behaviour has been associated with this formulation. Yu et al. (1998) [157, 169] developed polymeric micelles based on poly(ethylene oxide)-block-poly(b-benzyl-L-aspartate) (PEO-PBLA) which solubilise amphotericin B during self-assembly (AmB-PEO-PBLA). *In vitro* haemolysis studies using rat RBCs showed Fungizone to result in 100 % RBC lyses at a drug concentration of 3.0 µg/mL after 30 min of incubation. In contrast, AmB-PEO-PBLA micelles at a concentration of 3.0 µg/mL resulted in no lyses for up to 24 hours, suggesting amphotericin B was gradually released from the micelles [157]. PEO-PBLA micelles without amphotericin B were shown to be non-haemolytic at a concentration of 0.7-1.0 mg/mL [157, 169]. To understand why this toxicity protection is afforded by the micelles, UV/VIS spectrum analysis revealed that AmB-PEO-PBLA micelles produced a spectrum that is characteristic of solubilising monomeric amphotericin B, whereas UV/VIS of Fungizone produced a spectrum that suggested amphotericin B was in an aggregated form [169]. Since the aggregated form of amphotericin B is partly related to its toxicity, the solubilisation of amphotericin B in PEO-PBLA micelles maintains its monomeric form which translates to a reduction in toxicity.

Cisplatin is a platinum based anticancer drug administered intravenously to treat various types of cancer. It consists of a central platinum atom, bonded to two ammonias and two chlorines. Tumouricidal activity results from the displacement of chloride by water during aquation, allowing the drug to interact with DNA and interrupt the mitotic phase of the cell cycle. Due to its low molecular weight, cisplatin is widely distributed into almost all tissue. The clinical use of cisplatin is however hindered by several toxicities associated with it, including ototoxicity, myelotoxicity, and in particular nephrotoxicity [94, 97]. Targeting cisplatin to a tumour site would therefore not only enhance efficacy but also reduce toxicity.

To do this Mizumura et al. (2001) [170] bound cisplatin to the aspartic acid residues of poly(ethylene glycol)-poly(aspartic acid) (PEG-ASP) block copolymer via ligand substitution, which in aqueous media formed polymeric micelles containing cisplatin (PEG-ASP-Cis). They investigated the plasma clearance and biodistribution of PEG-ASP-Cis micelles or non-formulated cisplatin by administering an i.v. dose of 150 µg into mice bearing Lewis lung

carcinoma cells. Whilst the plasma concentration of the non-formulated drug fell to below 5 % of injected dose after 4 hours, the PEG-ASP-Cis injected dose only fell to 42 % after 4 hours. The accumulation of PEG-ASP-Cis within the tumour was also significantly higher than that observed for non-formulated drug, however no difference in the kidney accumulation was observed between PEG-ASP-Cis and non-formulated drug.

Toxicity assessments were made using body weight changes, clinical blood and pathological analysis. Following three 5 mg/kg i.v. injections of either PEG-ASP-Cis or non-formulated drug into mice the body weight changes observed on day 11 were reported. Body weight changes for PEG-ASP-Cis were in fact not significantly different to control animals, with a maximum loss of 3.3 %, whilst the non-formulated drug resulted in a 5 % body weight loss.

Male rats were used for clinical and pathological assessments, following a 10 mg/kg injection of PEG-ASP-Cis or non-formulated drug. Table 2.9 shows the renal function of rats on day 7 after administration.

Table 2.9. Renal function analysis of rats following administration of PEG-ASP-Cis and non-formulated drug after 7 days. Data adapted from [170].

Drug/Formulation	Species	Admin route	Dose (mg/kg)	BUN (mg/mL)	Creatinine (mg/mL)
Cisplatin (non-formulated)	Rats	i.v.	10.0	67.3 ± 20.5	0.82 ± 0.23
PEG-ASP-Cis	Rats	i.v.	10.0	24.0 ± 4.4	0.26 ± 0.09
Control (saline)	Rats	i.v.	-	20.6 ± 2.6	0.24 ± 0.06

The data reported in Table 2.9 suggest that no nephrotoxic effects were apparent following administration of PEG-ASP-Cis, since the renal function parameters are in line with the saline control animals. Non-formulated drug however elevated levels of both BUN and creatinine 7 days after the administration of a 10 mg/kg dose. Renal pathological assessment revealed that rats administered non-formulated drug presented tubular cell degeneration which was not observed in the PEG-ASP-Cis treated rats. The data from this study show how PEG-ASP-Cis is able to accumulate to greater amounts within tumour sites. *In vivo* efficacy experiments reported in the study showed the tumouricidal effects of PEG-ASP-Cis to be as effective as the non-formulated drug but with a reduced nephrotoxic burden [170].

From the data collected on polymeric micelles it is apparent that they achieve their toxicity reduction by strategies 3 and 4 discussed in section 2.7. That is they are able to actively target drugs to the site of disease as seen with cisplatin micelles that are able to

preferentially accumulate within a tumour, or they enhance the availability of the drug as seen with amphotericin B (monomeric vs aggregated form).

2.7.3 Cyclodextrins

Cyclodextrins (CD) are cyclic glucose oligosaccharides, consisting of (α-1,4)-linked α-D-glucopyranose units. Common naturally occurring cyclodextrins consist of 6, 7 or 8 D-glucopyranose units and are referred to as α-, β- and γ-cyclodextrin respectively (Figure 2.5) [171].

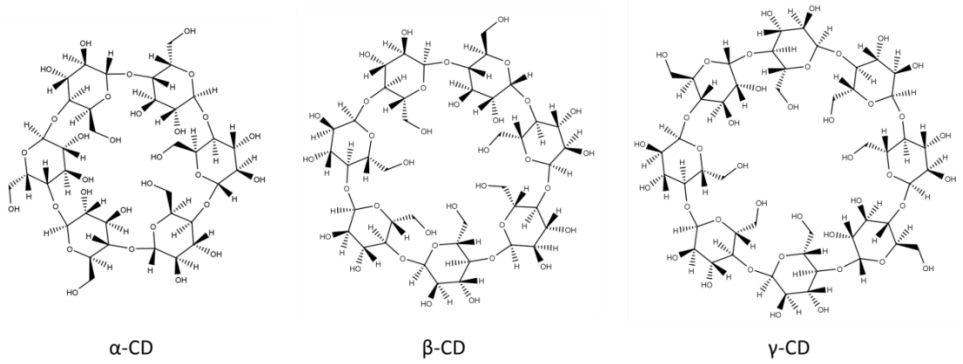


Figure 2.5. α, β and γ-cyclodextrin containing 6, 7 and 8 glucopyranose units respectively.

From the steric arrangement of the glucose units, cyclodextrins form toroidal shaped structures that resemble a truncated cone. The inner side of the torus-like structure is less polar than the outer, resulting in a hydrophilic outer surface and hydrophobic inner core or cavity (Figure 2.6). In aqueous environments the hydrophobic cavity is occupied by water molecules which are energetically unfavourable and are readily substituted with "guest molecules" such as small hydrophobic drug compounds [172]. Ejecting the enthalpy rich water molecules from the cavity decreases the ring strain, producing a more stable energy state [171-173]. This complexation with hydrophobic compounds in aqueous media has been the basis behind the use of cyclodextrins for pharmaceutical applications. Inclusion complexes are formed provided the drug molecules are of the appropriate size and physiochemical properties. Covalent bonds are neither formed nor broken during complexation [174], with complexes readily dissociating in aqueous solution and free drug molecules remaining in equilibrium. Complexes can be formed with the entire drug or a specific portion in a typically 1:1 or 1:2 drug:cyclodextrin ratio [171, 173, 175]. The hydrophilic external hydroxyl groups remain exposed to the aqueous environment and so create a water soluble cyclodextrin-drug complex. Inclusion complexes offer profound effects to the physicochemical properties of the drug they carry. Most commonly they

enhance the solubility, permeability and bioavailability of drug molecules. Because cyclodextrins “lock” a drug within the complex they provide stability from degradation by UV light, heat and oxidation [172].

In crystalline form only the surface molecules of the cyclodextrin can form inclusion complexes. This can be significantly enhanced by heating the cyclodextrin/drug mixture to encourage complexation. For poorly water soluble drugs it is possible to use solvents such as ethanol or diethylether which do not form inclusion complexes themselves, to solubilise drugs and aid complexation.

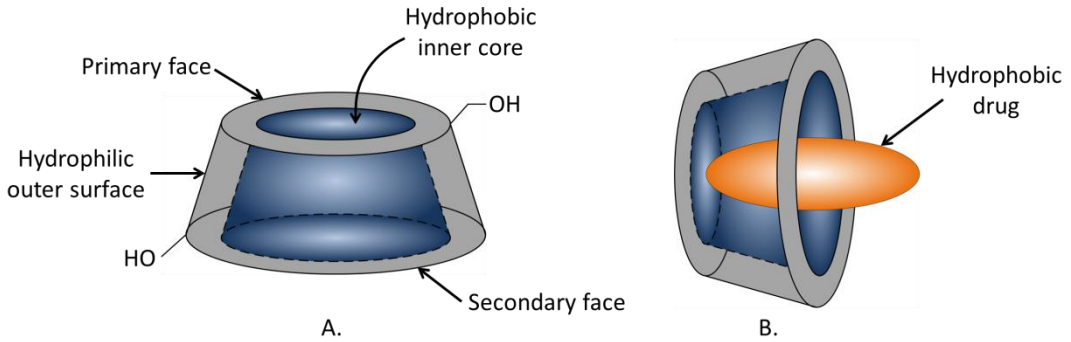


Figure 2.6. Toroidal cyclodextrin shape with a hydrophobic inner core and hydrophilic outer surface (A). 1:1 Cyclodextrin:drug complexation (B).

Since complexation is considered to be in equilibrium with the free drug, it may also be possible to increase the solubility of the drug through ionisation, salt formation and the formation of metal complexes to encourage complexation formation [176]. Several methods describing complexation preparation techniques are described in the literature and include co-precipitation, slurry complexation, and paste complexation [172]. Once a complex is formed and dried it is considered very stable with a long shelf life.

Cyclodextrins are generally considered non-toxic although at sufficiently high concentrations they can cause haemolysis of erythrocytes [173, 177]. When given orally they are poorly absorbed through the gastrointestinal tract and practically resistant to salivary and pancreatic amylases [176, 178]. The chemical structure which contributes to a large number of hydrogen bond donors and acceptors, their molecular weight and very low octanol/water partition coefficient are all characteristics which result in poor membrane permeability [176]. Because of their poor gastrointestinal tract absorption they are considered non-toxic when administered orally [179]. Intra molecular hydrogen bonding between hydroxyl groups contributes to a lower solubility, which can be sufficiently enhanced by the introduction of functional groups at the 2-, 3- and 6-hydroxyl groups of

the glucose residue [173, 176, 179-181]. These changes improve the solubility by breaking the 2-OH-3-OH hydrogen bonds and by preventing crystallisation [173]. With 18 (α-CD), 21 (β-CD) and 24 (γ-CD) substitutable hydroxyl groups there are many possible derivatives that can be made. Hydroxypropyl-β-cyclodextrin (HP-β-CD) and sulfobutyl-β-CD are two such derivatives with satisfactory toxicological profiles that are frequently used [171, 182].

2.7.3.1 Cyclodextrins in Toxicity Formulations

Cyclodextrin complexation reduces toxicity mainly via the mechanisms described by strategies 2 and 4 (see section 2.7). One of the main applications of cyclodextrins in pharmaceutical formulations is to increase the solubility of a drug, although other studies have suggested cyclodextrins may also behave as permeation enhancers [183]. This increase in solubility improves the bioavailability of a drug which can then enter systemic circulation. The rate and extent of the bioavailability can be influenced by adjusting the factors which affect the dissociation equilibrium of the complex. Only the free form of the drug, which in solution will be in equilibrium with the complexed form, is able to penetrate lipophilic membranes to enter systemic circulation (Figure 2.7). Maximum absorption enhancement is therefore most likely to be achieved when the just the right molar concentration of cyclodetrin is used to dissolve the entire drug concentration in solution. An abundance of cyclodextrin would solubilise a drug *in vitro* well, but this abundance *in vivo* would prevent the free form drug dissociating for long enough that it could enter systemic circulation.

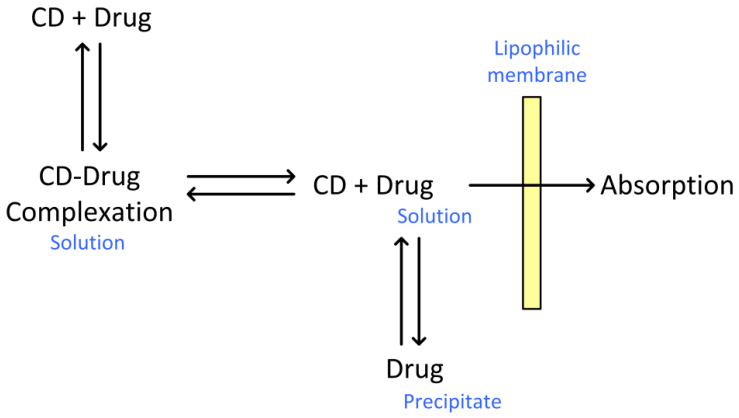


Figure 2.7. Behaviour of cyclodextrin-drug complexation prior to absorption through biological membranes.

Cyclodextrins therefore serve two purposes when used in formulations for toxicity reduction. 1) They entrap a poorly water soluble drug through complexation, thus

preventing precipitation and aggregation issues and 2) Through complexation they prevent large concentrations of drug interacting with local tissues.

These two properties of cyclodextrins satisfy their mode of action as described by strategies 2 and 4 in section 2.7.

To demonstrate these principles in practice there are several pieces of published work which show how the toxicity of a drug is ameliorated through cyclodextrin complexation. It is interesting to note that the majority of these studies used cyclodextrin formulations for the prevention of toxicities resulting from nonsteroidal anti-inflammatory drugs (NSAID).

An *in vitro* study on the haemolysis of RBCs, demonstrated that complexation of chlorpromazine and flufenamic acid with β-cyclodextrin protected RBCs from lysis, accounting this protection to a reduced drug concentration within the cells [184]. Furthermore, *in vivo*, using rabbits it was shown that chlorpromazine complexed with β-cyclodextrin resulted in a decrease in muscular tissue damage at the site of injection. This reduction in local toxicity was again ascribed to the reduction in the affinity of chlorpromazine for tissue membrane due to its affinity to preferentially form inclusion complexes [185].

The formulation of nonsteroidal anti-inflammatory drugs through cyclodextrin complexation has been studied and reported extensively in the literature. NSAIDs such as indomethacin, naproxen and diclofenac are known to cause adverse reactions in the gastrointestinal tract when administered orally, a condition that is exacerbated if the drug is poorly soluble or poorly permeable. As a consequence of poor absorption, levels of drug within the gut build up to a sufficient enough concentration to result in irritability and toxicity of the gastrointestinal mucosa [186]. Similar local toxicity issues are observed when NSAIDs are administered subcutaneously (s.c.) [187] or intramuscularly (i.m.) [188].

When administered orally, NSAIDs are reported to inhibit prostaglandin production through the inhibition of cyclooxygenases (COX-1 and COX-2). COX-1 is critical for maintaining healthy tissue and COX-2 is involved in the anti-inflammatory pathway of the gastrointestinal mucosa. These NSAIDs have been shown to significantly decrease the mucosal prostaglandin E_2 concentration when administered orally to rats [189]. This deficiency leads to an increased gastric motility provoking severe gastric lesions and ulceration [189, 190]. Cyclodextrin complexation with NSAIDs would therefore seem a logical formulation given that complexation would prevent the NSAIDs from interacting

with the cyclooxygenases and enhance absorption into systemic circulation. Several studies have reported such findings in the literature. Indomethacin, etodolac, naproxen, rofecoxib, flurbiprofen and 4-Biphenylacetic Acid (felbinac) have all been formulated using various cyclodextrin derivatives and administered orally to rats. When compared to the non-formulated drug, an enhancement in solubility, dissolution rate and bioavailability was reported. Gastrointestinal tolerability was also improved with lesions and ulcerations to gastrointestinal mucosa reduced in the formulated drugs [186, 191-195].

From these studies on NSAIDs we can conclude that cyclodextrins aid with the solubility of poorly soluble NSAIDs by reducing aggregation and prevent interaction with biological membranes such as the mucosa membrane thus preventing cyclooxygenases inhibition. To understand pharmacokinetic changes that result from cyclodextrin formulations several studies have reported their findings (Table 2.10). From these data it becomes apparent that regardless of the species used or the route of administration, a cyclodextrin formulation which reduces local toxicity does so because it is able to get the drug into systemic circulation quicker, indicated by the increase in AUC and C_{max} of the cyclodextrin formulated drug compared to the non-formulated drug [186-188].

Table 2.10. Pharmacokinetic changes through cyclodextrin formulation. AUC units - µg.h/mL* or ng.h/mL†. Cmax units - µg/mL* or ng/mL†. Administered dose amounts and AUC time periods have been omitted for simplicity reasons. Data adapted from [186-188].

Drug/Formulation	Species	Admin route	AUC	C_{max}	T_{max} (h)	$t_{1/2}$ (h)	Reference
Ricobendazole (non-formulated)*	Sheep	s.c.	36.7 ± 9.2	1.34 ± 0.31	9.6 ± 2.9	8.5 ± 3.4	[187]
Ricobendazole-HP-β-CD*	Sheep	s.c.	54.5 ± 15.3	2.90 ± 0.77	5.0 ± 0.6	5.5 ± 2.8	
Nimodipine (non-formulated)†	Rabbits	i.m.	31.90 ± 8.78	9.13 ± 1.78	0.78 ± 0.21	-	[188]
Nimodipine-2HP-β-CD†	Rabbits	i.m.	82.55 ± 8.92	23.18 ± 2.94	0.35 ± 0.06	-	
4-Biphenylacetic Acid (non-formulated)*	Rats	Oral	10736	70.2 ± 1.1	-	-	[186]
4-Biphenylacetic Acid-HP-β-CD*	Rats	Oral	14444	108.6 ± 2.5	-	-	

This was attributed to the ability of cyclodextrin to inhibit precipitation post-administration allowing rapid drug absorption when compared to the non-formulated drug. Because of this rapid absorption, local toxicity at the site of administration was reduced. Interestingly the T_{max} and $t_{1/2}$ of these formulations are shorter, which is also a consequence of the rapid absorption in comparison to the non-formulated drug. Because the formulated drug will enter systemic circulation sooner it is therefore eliminated faster when compared to the

non-formulated drug which will enter systemic circulation slower due to precipitation at the site of administration [187, 188]. These pharmacokinetic data reinforce the reasoning behind how cyclodextrin formulations reduce the local toxicity concerns of NSAIDs. Similar pharmacokinetic changes have been reported for other classes of drugs when complexed with cyclodextrins [175].

2.7.4 Dendrimers

Dendrimers are hyperbranched 3D tree-like molecules made up of polymers such as poly(amidoamine) (PAMAM) [196]. They possess a highly ordered structure with a high degree of molecular uniformity, narrow molecular weight distribution and multifunctional terminal surface properties. PAMAM dendrimers are synthesised from an alkyldiamine 'core' such as ethylene diamine (EDA) reacted with methyl acrylate by nucleophilic addition (Michael addition) to form a branched intermediate. Subsequent reaction of this intermediate with EDA would produce a G0 (generation zero) dendrimer with four NH_2 surface groups (Figure 2.8) whilst the use of ethanolamine would produce a G0 dendrimer flanked by four OH surface groups. Higher generation dendrimers (G1, G2, G3 etc.) containing exponentially more branching points are possible with further methyl acrylate reactions terminating with an amine, resulting in a spherical nanomolecule with a high surface polycationic charge. Depending on the polymer used dendritic growth is limited from reaching higher generations due to the branching arms which lead to steric hindrance.

Dendrimers are synthesised or grown in several ways. A core that is linear in structure will give rise to a rod like dendrimer whilst a tetrahedral core produces a more spherical dendrimer. Since dendrimers are made from polymers they are considered biologically safe [197], in particular ammonium core PAMAM generation 3, 4 and 5 dendrimers whose shape and structure is said to be mimicry of globular proteins such as insulin and haemoglobin [198]. Dendrimers additionally have a high degree of solubility and possess a large number of surface groups that can serve specific purposes such as carrying insoluble drugs to avert aggregation related toxicities [199].

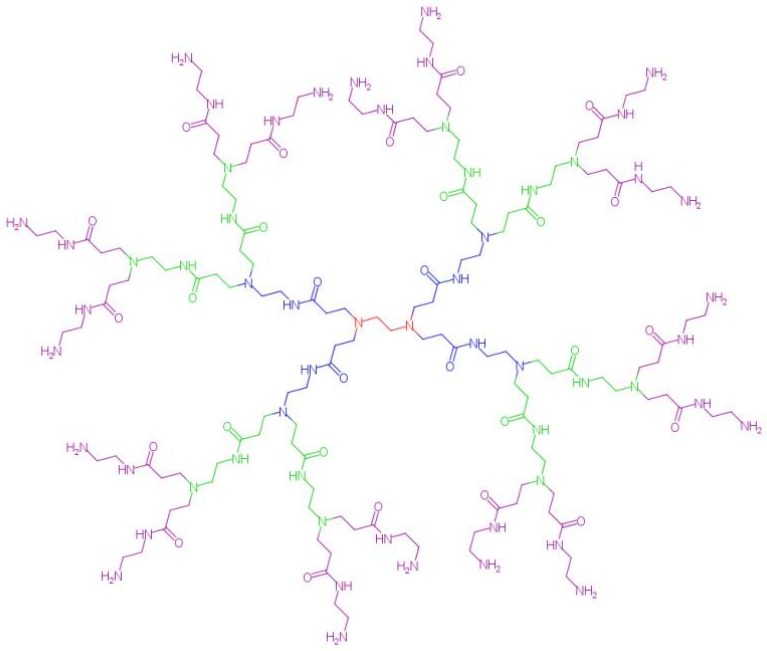


Figure 2.8. PAMAM G2 dendrimer showing ethylene diamine (EDA) core (red) reacted with methyl acrylate to produce G0 (red+blue), G1 (red+blue+green) and G2 dendrimer (red+blue+green+purple).

2.7.4.1 Dendrimers in Toxicity Formulation

The *in vivo* hepatotoxicity of methotrexate and 6-mercaptopurine was assessed by repeated dose toxicity assessment through formulation via melamine based dendrimers. After solubilising these anticancer drugs with dendrimers, a dose of 2 or 3.5 mg/kg of methotrexate or 6-mercaptopurine respectively was administered i.v. to mice for three consecutive days. 48 hours after the last dose the mice were sacrificed and blood samples assessed for hepatotoxic indications. The non-formulated administration of both methotrexate and 6-mercaptopurine elevated the serum ALT levels in mice when compared to a saline control, whilst the dendrimer formulated drugs resulted in lower ALT serum concentrations that were statistically different to the non-formulated groups but not significantly different to the saline controls. Dendrimers alone did not result in any serum ALT elevations either. These results suggest that dendrimers based on melamine are able to prevent the hepatotoxicity of methotrexate and 6-mercaptopurine. Although the exact cause of toxicity reduction is not reported it is thought that the noncovalent interaction of drug with the dendrimer prevents rapid diffusion of the free drug into systemic circulation thereby not exceeding a toxic threshold [200].

The anthracycline doxorubicin is another candidate for toxicity reduction via dendrimer formulation. In a study using fifth generation polypropylene imine (PPI) dendrimers conjugated to folic acid, doxorubicin loaded formulations (PPI-FA-Dox) were assessed for *in vitro* and *in vivo* toxicity [28]. The haemolytic toxicity of the formulation was assessed *in vitro* using human RBCs and compared against non-formulated doxorubicin. At equivalent doses, PPI-FA-Dox resulted in 4.28 % RBC haemolysis whilst the non-formulated drug resulted in 16.37 % RBC haemolysis. Interestingly PPI dendrimer formulations without folic acid conjugates (PPI-Dox) resulted in 18.35 % RBC haemolysis, demonstrating that the folic acid conjugates played an important role in haemolytic protection. It was proposed that the folic acid conjugates shielded doxorubicin from interaction with RBC surfaces which resulted in reduced RBC lysis.

Haematological studies of the PPI-FA-Dox formulation was also analysed *in vivo* using rats. Seven daily i.v. injections were administered of PPI-FA-Dox, non-formulated doxorubicin and PPI-Dox at a dose of 250 mg/mL. After 15 days, blood samples from the animals were collected and analysed for haematological cell counts. The analysis showed PPI-FA-Dox formulations were able to significantly reduce the reduction of RBCs and increase in WBCs that was typical of the non-formulated drug. An increase in monocyte and lymphocyte counts observed with non-formulated drug was also prevented with PPI-FA-Dox. PPI-Dox without folic acid conjugates was unable to prevent these toxicities.

The *in vitro* drug release profiles of the two dendrimer formulations (PPI-FA-Dox and PPI-Dox) were compared using PBS (phosphate buffered saline) at a pH of 7.4 and 6.4. PPI-Dox resulted in the release of 59.46 and 74.36 % drug from the formulation in PBS at pH 7.4 and 6.4 respectively. PPI-FA-Dox produced a slower release profile resulting in the release of 52.36 and 68.24 % drug release at pH 7.4 and 6.4 respectively.

Cell inhibition studies also revealed an interesting mechanism of action for the PPI-FA-Dox formulations. Using MCF-7 human breast cancer cells the IC_{50} value of PPI-FA-Dox was reported as 0.45 µg/mL while PPI-Dox and non-formulated doxorubicin resulted in an IC_{50} value of 0.75 and 0.70 µg/mL respectively. Fluorescence analysis showed a higher uptake of the PPI-FA-Dox formulation into MCF-7 cells which was accounted for due to the folate residues on the surface of PPI-FA-Dox dendrimer formulations.

Together these results suggest that PPI-FA-Dox results in a higher uptake into MCF-7 cells due to the overexpression of folate receptors on these cells, suggesting a targeted

mechanism for the improved IC_{50} efficacy of the drug. The quicker release of drug from both PPI-Dox and PPI-FA-Dox formulations at a pH of 6.4 compared to pH 7.4 may also contribute to the improved efficacy of the drug. Since the microenvironment around tumour sites is generally below pH 7 this may confer improved drug targeting and efficacy *in vivo*.

A dendrimer formulation bound to doxorubicin has been tested *in vivo* and reported by Kaminskas et al. They prepared a polylysine based dendrimer conjugated with PEG and doxorubicin (PolyL-PEG-Dox) [52]. The MTD of the PolyL-PEG-Dox formulation in mice was reported as 10 mg/kg that resulted in no more than 10 % body weight loss. The MTD of the non-formulated doxorubicin was 4.5 mg/kg. The systemic toxicity and cardiotoxicity effects of non-formulated doxorubicin were also reduced by the PolyL-PEG-Dox formulation even when administered at a 2-fold higher dose in mice [52].

Dendrimer formulations are therefore able to reduce the toxicity of a drug via the strategies 2, 3 and 4 discussed in section 2.7.

2.7.5 Micro/Nano Formulations

Micro and nanoformulations as their name implies are formulations in the micro (1-1000 µm) and nano (1-1000 nm) size range. They are produced in much the same way although nano scaled formulations require much more robust methods. By convention, liposomes, cyclodexterins, polymeric micelles and dendrimers can be classed as nanoparticles given that their physical size is often in the nano range. This section however focuses on the micro/nano technologies that are formulated as emulsions, particles, suspensions, colloids and solid dispersions. Of these the micro/nano emulsions and particles are most widely used to formulate drugs for toxicity reduction.

Microemulsions are formed by mixing oil, water and surfactants, forming phases that are in equilibrium with each other [201]. Microemulsions can be water-continuous, oil-continuous or bi-continuous depending on the concentrations and drugs formulated [202]. An oil-in-water (o/w) microemulsion for example is a colloidal dispersion consisting of oil droplets and surfactant dispersed in a continuous water phase. As with micelles (see section 2.7.2), the surfactant's hydrophobic tails associate to reduce the thermodynamic instability when in contact with water. The hydrophilic head groups then protrude and contact the aqueous phase. Oil droplets are carried within the hydrophobic interior of the

micelle and may contain a hydrophobic drug that becomes dispersed within the continuous aqueous phase.

Nanoemulsions by contrast are considered to be in a non-equilibrium state making them thermodynamically unstable systems [202], However the destabilisation of nanoemulsions is slow, typically months [201] and they are resistant to moderate dilution and temperature changes [203], with their small size providing stability and avoiding Ostwald ripening and therefore emulsion degradation [201, 204]. A nanoemulsion can therefore be described as a conventional emulsion that contains particles of a nano size, typically 300 - 500 nm in diameter [201, 203]. Nano-emulsions can be oil-in-water (o/w) or water-in-oil (w/o) dispersions, depending on whether the water is the continuous phase and the oil is the dispersion phase or vice versa [204]. The major distinction between micro- and nanoemulsions is therefore the physical size of the dispersion phase and the thermodynamic stability of the system. Oil-in-water systems are used to formulate poorly water soluble drugs and disperse them into an aqueous media for administration.

Lastly, nanoformulations are frequently produced as nanoparticles, which are described as solid colloidal particles in the nano scale range [203]. They are made from various macromolecules to which drugs are adsorbed, entrapped or covalently bound. Synthetic polymers, such as poly(lactic acid) (PLA), poly(glycolic acid) (PGA) and their copolymer poly(lactide-co-glycolide) (PLGA) [205, 206] or Pluronic F127 [207] as well as natural macromolecules such as chitosan, polysaccharides and gelatin are frequently used to form nanoparticles [203]. PLA and PGA are considered biocompatible, have good stability and can sustain drug release. Nanoparticles have the added advantage that their sheer size also makes them ideal for intravenous delivery of drugs without the risk of forming aggregates causing an embolism or removal via the RES. Nanoparticles can be synthesised to deliver hydrophobic and hydrophilic drugs as well as proteins, vaccines and macromolecules.

2.7.5.1 Micro/Nano Particles in Toxicity Formulation

The platinum based anticancer drug cisplatin has been formulated using biodegradable poly(D,L-Lactic-co-glycolic) acid (PLGA) to form cisplatin bound nanoparticles (Cis-PLGA) (Figure 2.9) that were shown to be nephroprotective to the adverse effects of the drug [208]. This formulation was prepared by a water-oil-water emulsion solvent evaporation method.

The formulation was tested in mice bearing the DHD/K12PROb adenocarcinoma colon cell line. Following a 5 mg/kg i.p. dose once per week for two cycles, of formulated and non-formulated drug administered to mice the study showed Cis-PLGA was able to reduce elevations in BUN that were recorded with the non-formulated drug.

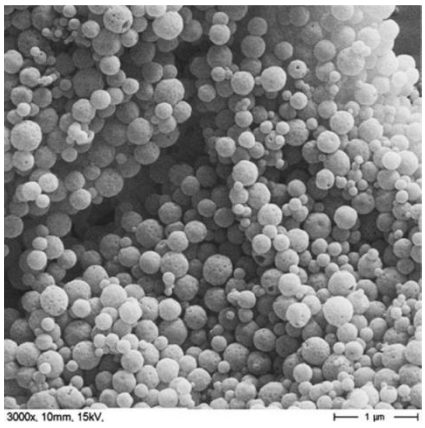


Figure 2.9. Scanning electron micrograph of cisplatin-PLGA nanoparticles prepared by a water-oil-water emulsion solvent evaporation method. Image reproduced from [208].

Table 2.11 shows that on the 8th day the BUN levels of the non-formulated drug reached 42.1 ± 3.2 mg/dL, compared to Cis-PLGA levels of 19.9 ± 1.9 mg/dL which were comparable to control animals. The changes in body weight were also favourable for the Cis-PLGA group with no body weight loss measured, whereas for the non-formulated drug result in approximately 10 % body weight loss when compared to the control group.

Table 2.11. BUN plasma levels (mg/dL) following 5 mg/kg i.v. administration of Cis-PLGA and non-formulated cisplatin. Data represents the mean ± SD of 6 animals. Data adapted from [208].

Time (Days)	Control (saline)	Empty PLGA nanoparticles	Cisplatin (non-formulated	Cis-PLGA
Basal	20.1 ± 1.6	19.8 ± 3.3	21.1 ± 1.2	20.0 ± 0.6
7	20.3 ± 1.5	21.7 ± 1.9	18.5 ± 2.6	19.6 ± 1.7
7.25	21.3 ± 2.9	22.2 ± 0.7	27.5 ± 1.2	20.0 ± 0.6
8	20.1 ± 2.5	21.8 ± 2.1	42.1 ± 3.2	19.9 ± 1.9
9	21.5 ± 2.4	21.4 ± 2.3	35.7 ± 3.9	20.3 ± 2.1
11	21.3 ± 1.3	21.5 ± 0.8	20.9 ± 1.5	19.6 ± 0.4

Both the non-formulated drug and Cis-PLGA reduced the growth of the developing tumour with no significant difference between them, suggesting the efficacy of the Cis-PLGA formulation was maintained yet nephrotoxicity was reduced. No statistical difference on the tumour growth was observed between the control and empty nanoparticle groups either, suggesting PLGA nanoparticles alone have no tumouricidal properties. A statistical

difference was found between the two cisplatin groups and the two non-cisplatin groups in terms of tumour growth.

The biodistribution and pharmacokinetics of the Cis-PLGA formulation were also reported and highlighted some interesting changes that result from nanoparticle formulation. Table 2.12 shows the change in the pharmacokinetics that becomes apparent with a Cis-PLGA formulation when compared to the non-formulated drug.

Table 2.12. Pharmacokinetics of cisplatin formulated in PLGA nanoparticles and non-formulated free drug. Data adapted from [208]

Drug/Formulation	Species	Admin route	$AUC_{0\text{-}21}$ (mg.day/L)	C_{max} (mg/L)	T_{max} (min)
Cisplatin (non-formulated)	Mice	i.p.	3.77 ± 0.33	10.83 ± 0.95	5
Cis-PLGA	Mice	i.p.	2.80 ± 0.02	0.33 ± 0.05	60

The changes in pharmacokinetics that result from a PLGA nanoparticle formulation are striking, a statistical difference was found in the AUC, C_{max} and T_{max} values when compared to the non-formulated drug. Of particular interest is the C_{max} value of the Cis-PLGA formulation that was only 3 % of the C_{max} of the non-formulated drug. The T_{max} was also prolonged to 60 min from 5 min for the Cis-PLGA formulation and non-formulated drug respectively.

Biodistribution profiles of Cis-PLGA and non-formulated drug showed the main organ for uptake of Cis-PLGA was the spleen, whilst a reduced uptake was observed in the lungs and liver when compared to non-formulated drug. Drug uptake into the kidneys was similar for both Cis-PLGA and non-formulated drug but the reduced nephrotoxicity of Cis-PLGA may be reflected by the slow release of the drug from PLGA nanoparticles that accumulate within the spleen. The significantly lower C_{max} and prolongation of T_{max} demonstrate that a sustained release of the drug from the Cis-PLGA formulation from within the spleen is the most likely reason for the nephroprotective properties of the nanoparticle formulation.

Microemulsions are an alternative to the nano-formulations that have been shown to be successful at ameliorating toxicity. For example Brime at al. [209] has reported on lecithin-based oil-in-water microemulsions of amphotericin B that demonstrated a reduced toxicity profile of the drug. A lecithin based microemulsion that incorporated amphotericin B was administered to mice at a dose of 1 mg/kg i.v. for 5 consecutive days and compared to the clinical sodium deoxycholate formulation of amphotericin B (Fungizone). Histopathological examinations of mice administered Fungizone showed tubular lesions of the kidneys,

including focal necrosis of the epithelial cells that occurred in the Henle loop and distal convoluted tubules. No significant histopathological changes were observed in animals treated with the microemulsion formulation. Furthermore, no significant changes were recorded at a dose of 3 mg/kg for the microemulsion formulation either, demonstrating a higher tolerability of this formulation. A 3 mg/kg dose of Fungizone was not tested. At the tested dose of 1 mg/kg no significant changes in the blood creatinine levels between the microemulsion and Fungizone were reported, although histopathological reports confirmed nephrotoxicity, creatinine may not be a sensitive enough biomarker for the detection of renal injury at low doses. Pharmacokinetic parameters assessed in rabbits demonstrated the differences between the formulations, with the microemulsion formulation producing a lower AUC, C_{max}, MRT and V_{ss} than that of Fungizone. The Cl was reported to be slightly higher for the microemulsion than for Fungizone (Table 2.13). This may demonstrate a faster elimination of the microemulsion incorporated amphotericin B from the plasma via RES uptake, followed by redistribution as seen with other lipid-based amphotericin B formulations [140, 210].

Table 2.13. Pharmacokinetic profiles of a microemulsion amphotericin B formulation and Fungizone administered at 1 mg/kg i.v. into rabbits. All values are mean ± SD. *Non-compartmental analysis. † Two-compartmental analysis. Data adapted from [209].

Drug/Formulation	Species	Admin route	$AUC_{0-\infty}$ (mg.h/L)*	C_{max} (mg/L)*	MRT (h)*	$t_{1/2}$ (h)*	Cl (L/h)†	V_{ss} (L)†
Amphotericn B (Fungizone)	Mice	i.v.	32.280 ± 7.310	3.890 ± 0.480	27.100 ± 4.800	19.840 ± 3.140	0.082 ± 0.010	2.035 ± 0.170
Amphotericn B Microemulsion	Mice	i.v.	21.890 ± 5.170	2.920 ± 0.540	16.230 ± 0.200	12.320 ± 0.480	0.113 ± 0.006	1.890 ± 0.160

This lipid based microemulsion formulation of amphotericin B may offer toxicity protection in a similar manner to other lipid based formulation of amphotericin B, principally by reducing the exposure of the drug to mammalian membranes as discussed in section 2.7.1.1.

Mitoxantrone is an antineoplastic agent that results in local systemic toxicity issues at therapeutic doses with intraperitoneal injections reported to cause abdominal pain and bowel obstruction. The effects reported on the abdominal cavity are thought to develop as a result of high local drug concentrations at the site of administration. Luftensteiner et al. therefore developed albumin loaded mitoxantrone microspheres (Mxn-Alb) with high stability, biocompatibility and controlled release properties. Rats administered 30 mg/m^2 surface area i.p. at 3-weekly intervals with Mxn-Alb were shown to survive significantly longer than rats administered the non-formulated drug. Mxn-Alb administered animals

were also shown to gain weight over a period of 21 days whilst animals administered non-formulated drug slightly lost weight [211].

The pharmacokinetic profiles of the Mxn-Alb formulation were compared to the non-formulated drug following a single i.p. dose of 30 mg/m2 surface area (Table 2.14). The data show that following i.p. administration of Mxn-Alb the plasma AUC and C_{max} are reduced compared to the non-formulated drug. The T_{max} is also prolonged to 8 hours for the Mxn-Alb formulation, compared to 1 hour for the non-formulated drug [212]. These data show that albumin microspheres achieve a reduction in toxicity by controlling the release of the drug from the formulation.

Table 2.14. Pharmacokinetic profile of mitoxantrone in albumin microspheres and non-formulated drug at an i.p. dose of 30 mg/m2 surface area. Data adapted from [212].

Drug/Formulation	Species	Admin route	Dose (mg/m^2 surface area)	AUC_{0-72} (ng.h/mL)	C_{max} (µg/mL)	T_{max} (h)
Mitoxantrone (non-formulated)	Rats	i.p.	30	463.5	70.2	1
Mitoxantrone albunim microspheres	Rats	i.p.	30	442.8	20.3	8

Lastly, lipid based nano-particles are an alternative approach to nano-based formulations, that have been widely reported in the literature. Lipid nanoparticles form a hydrophobic core made from a monolayer of phospholipids and can entrap poorly soluble drugs, modulating their release behaviour *in vivo* and improving their therapeutic index [29]. Jung et al. formulated several lipid based nanoparticles using a spontaneous emulsification, solvent evaporation method to entrap amphotericin B (LNP-AmpB); plain, anionic and PEG based lipid nano-particles were developed and compared to the clinical formulation of amphotericin B in sodium deoxycholate (Fungizone).

In vitro haemolysis studies performed on rat erythrocytes showed that all the LNP-AmpB formulations showed very little haemolysis up to a concentration of 50 µg/mL amphotericin B, whereas Fungizone resulted in 100 % haemolysis at an equivalent concentration. These results suggest that the lipid matrices of LNP-AmpB formulations reduce the direct interaction of the drug with erythrocytes resulting in less haemolysis, a similar finding is reported with liposomal based amphotericin B formulations (see section 2.7.1.1). Furthermore due to the reported nephrotoxic implications of amphotericin B the study also tested the *in vitro* cytotoxicity of these LNP-AmpB formulations against human kidney cells (293 cells). At a concentration of 100 µg/mL Fungizone lowered cell viability to

approximately 20 %, whereas the anionic and PEG based LNP-AmpB formulations resulted in no effect on cell viability.

The *in vivo* pharmacokinetic profiles of these formulations were compared against Fungizone using rats. LNP-AmpB formulations and Fugizone were administered i.v. at a dose of 3 mg/kg body weight. A one-compartment model was fitted and showed that LNP-AmpB formulations increased the AUC and Cmax of amphotericin B when compared to Fungizone (Table 2.15). The PEG LNP-AmpB formulations however significantly increased the AUC and Cmax of amphotericin B, suggesting that the PEG stabilised amphotericin B in systemic circulation and reduced opsonisation or RES uptake due to low plasma protein binding. The extended half-life of PEG LNP-AmpB formulations *in vivo* also suggests that amphotericin B is slowly released from the lipid nanoparticles.

Table 2.15. Pharmacokinetic properties of lipid based nanoparticle formulations entrapping amphotericin B, administered at a dose of 3 mg/kg i.v. into rats. Data adapted from [29].

Drug/Formulation	Species	Admin route	$AUC_{0\text{-}24h}$ (µg.h/mL)	C_{max} (ng/mL)	$t_{1/2\ (0\text{-}24h)}$ (h)
Amphotericn B (Fungizone)	Rat	i.v.	1.084	0.224	3.825
Anionic LNP-AmpB	Rat	i.v.	2.598	0.907	1.526
PEG LNP-AmpB	Rat	i.v.	19.448	4.021	5.591

It is worth noting that these pharmacokinetic results are in contrast to the lecithin based microemulsion formulation [209] discussed earlier in this section. Although the pharmacokinetics was assessed in rabbits the microemulsion formulation reduced the plasma AUC and C_{max} values. This may be accounted for by the different animal models used or the differences in micro and nano formulation properties. Both studies however reported a toxicity reduction that was accounted for by the reduced interaction of amphotericin B with its site of toxicity.

Nano/micro formulations are therefore able to reduce the toxicity of a drug through strategy 2 as discussed in section 2.7. By controlling the release of a drug the formulation is effectively preventing a toxic dose from interacting with its site of toxicity.

2.7.6 Antidotes

Antidote formulations do not generally involve drug encapsulation, solubilisation, polymer attachment or targeting techniques for toxicity reduction. Rather they are co-administrative agents that aid in ameliorating a drug's toxicity. They achieve this by

interfering with or interrupting adverse outcome pathways, out competing ligands for active sites, thereby preventing molecular initiating events (MIE), promoting endogenous cellular defences, up/down regulating biochemical pathways or acting as scavengers of reactive species.

Antidote formulations are interesting since they impact the pharmacodynamic toxicity aspects of a drug. Although pharmacokinetic modulation through antidote formulation may be possible [213], the majority of the antidotes modulate some aspect of a drug's adverse outcome pathway (AOP). These antidotes can therefore be integrated into the AOP profile of a drug providing an understanding of the mechanism for their reduction in toxicity. Understanding the toxicity pathway of a drug may additionally help translate antidotes that are effective with one drug to another provided they share some commonality in their toxicity pathway.

Antidotes cover a vast range of compounds that cannot be described collectively. They comprise vitamins, plant extracts, seed oils and other drugs. Since the pharmacokinetic profile of a drug is largely unaffected by antidote formulations, these formulations most probably reduce the toxicity manifestation of a drug through strategy 1 discussed in section 2.7.

2.7.6.1 Antidotes in Toxicity Formulation

To understand how antidotes achieve their toxicity reductions it is important to understand in some detail the toxic pathway of a specific drug. For this reason this section will focus on the antidote formulations that have successfully reduced the hepatotoxicity of acetaminophen.

Acetaminophen (paracetamol) has a well-documented metabolism pathway, with new additions continuously discovered and reported in the literature. The major parts of acetaminophen's toxic pathway are shown in Figure 2.10. Briefly, at therapeutic doses acetaminophen undergoes extensive hepatic metabolism with about 85 % of a therapeutic dose undergoing phase II metabolism and safely excreted via glucuronidation and sulphation. The remaining drug undergoes phase I metabolism via a direct two electron oxidation step and is converted into the toxic metabolite N-acetyl-*p*-benzoquinone imine (NAPQI) by cytochrome P450 enzymes [214, 215]. The principal enzyme in this metabolism is CYP2E1 although others such as CYP1A2, CYP3A4 and CYP2D6 have been reported in the literature [88, 214, 216]. At therapeutic doses, NAPQI is detoxified by reduced glutathione

(GSH), an endogenous antioxidant and excreted via the kidneys. At toxic doses however or at times of impaired liver function, GSH becomes rapidly depleted leaving NAPQI to bind with cellular proteins resulting in the formation of reactive oxygen species (ROS) and reactive nitrogen species (RNS). NAPQI is also implicated with mitochondrial protein binding which depletes mitochondrial GSH resulting in further ROS production. As intracellular ROS increases it activates ASK-1 (Apoptosis signal-regulating kinase 1), a mitogen-activated protein kinase kinase kinase (MAP3K), which in turn activates JNK (c-Jun N-terminal kinases). The translocation of JNK to mitochondria induces mitochondrial permeability transition (MPT) pore opening which diminishes mitochondrial integrity resulting in cell death [216-218].

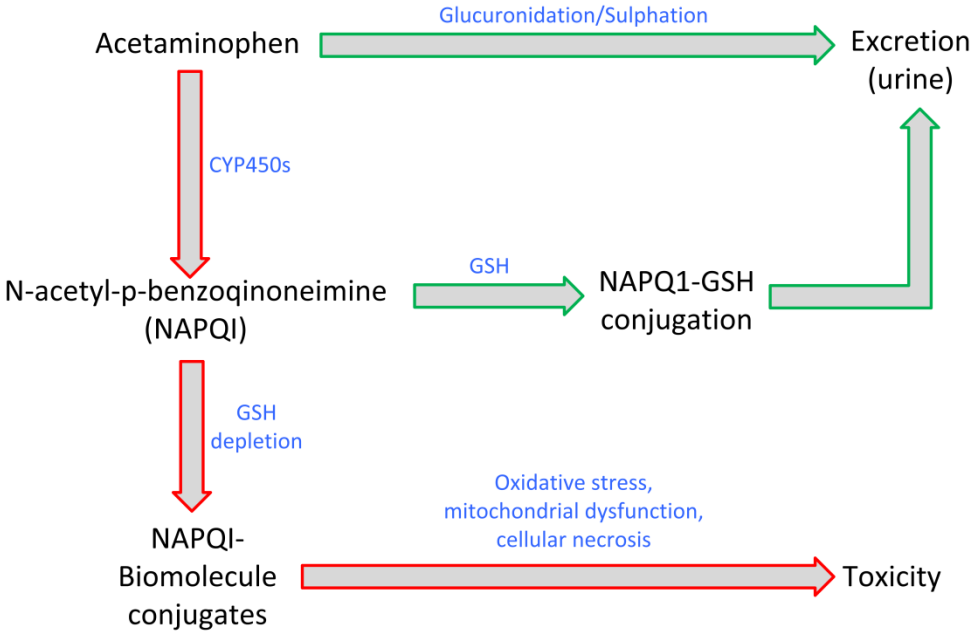


Figure 2.10. Metabolism of acetaminophen. At toxic doses, NAPQI is not sufficiently detoxified allowing biomolecule binding which leads to toxicity (red outline).

Understanding the adverse outcome pathway of a drug can prove critical in formulating for toxicity reduction but understandably this may not be a luxury afforded to a formulator working with NCEs. The pathway for acetaminophen described in Figure 2.10 provides an ideal platform for understanding the mechanism of action of antidote formulations. The red outline of Figure 2.10 shows the pathway undertaken during acetaminophen induced liver toxicity. Antidote formulations that interrupt this pathway or promote the non-toxic pathway (green outline) could therefore act as potential hepatoprotectants.

Several studies have shown this principle to work in practice. Lee et al. demonstrated the use of the drug chlormethiazole, an active inhibitor of the P450 enzyme CYP2E1 which was

able to completely attenuate the toxicity of acetaminophen. Pretreatment of chlormethiazole via i.p. administration 30 min before an i.v. dose of acetaminophen into mice was able to maintain serum AST and ALT levels, comparable to that of control animals. Furthermore chlormethiazole was also effective when administered 2 hours after acetaminophen, resulting in less elevated AST and ALT levels compared to that of the non-formulated drug [219]. The drug cimetidine is reported to show similar hepatoprotective properties via the same mechanism of action of preventing P450 acetaminophen oxidation [220]. Dimethylsulfoxide (DMSO) is known to promote the activity of CYP2E1 and would potentially not be a suitable protectant for acetaminophen. However studies by Yoon et al. showed that although CYP2E1 activity is upregulated through DMSO administration, the solvent is also a substrate for CYP2E1 and actively competes for the enzyme thereby preventing the formation of NAPQI when both are administered i.p. into mice [69].

The current therapeutic treatment for acetaminophen overdose is through the administration of N-acetylcysteine (NAC). NAC is an L-cysteine analogue that serves as a precursor to GSH [221] the main endogenous defence against the reactive acetaminophen metabolite NAPQI (see Figure 2.10). To provide cellular protection during times of acetaminophen toxicity, levels of GSH or other suitable antioxidant must be maintained. Several studies in the literature report co-administrative efforts that provide antioxidant benefits during acetaminophen toxicity. Sesame oil administered orally to rats was able to maintain GSH levels, reduce ROS generation and inhibit lipid peroxidation, preventing increases in serum transaminase levels without affecting acetaminophen oral absorption. Histological analysis showed sesame oil prevented centrilobular necrosis within liver tissue possibly due to the various antioxidants contained within it [222]. Many plant based extracts are known to offer hepatic protection against acetaminophen toxicity. The literature reports on extracts of garlic, *Phyllantus niruri* and the fruit of *Cornus officinalis* as hepatoprotectants of acetaminophen injury [66, 102, 223].

2.8 Building Toxicity Protection Pathways

From the data discovered through the literature that detail the pharmacokinetic and pharmacodynamic modulating formulations for toxicity reduction, it is possible to construct toxicity protection pathways. These pathways show how different formulation techniques are effective at reducing the toxicity of a drug at specific points of an AOP profile. The pathways are partly built on the AOP framework but provide more detail at the cellular level of toxicity to accommodate the pharmacodynamics modulating formulations.

Consider the aminoglycoside antibiotic gentamicin, used to treat gram-negative bacterial infections. As with all aminoglycosides, gentamicin's use is limited by its nephrotoxicity. Pharmacokinetic studies show that gentamicin is principally excreted via the kidneys in a largely un-metabolised form [224]. This serves as the first step of the drug's AOP profile with high kidney exposure *in vivo*. Once within the kidneys the drug accumulates within the epithelial cells of the proximal tubule [99] largely by megalin/cubilin mediated endocytosis [225, 226] and governed by the electrostatic interaction of gentamicin to negatively charged membrane phospholipids [101, 225, 227], resulting in the loss of brush border which leads to tubular necrosis. Intracellular, gentamicin mostly accumulates within lysosomes but also interacts with the Golgi complex and endoplasmic reticulum (ER). Within lysosomes, gentamicin is responsible for membrane destabilisation, lysosomal aggregation, alterations in lipid metabolism and phospholipidosis [100, 228]. ER stress inflicted by gentamicin activates the unfolded protein response, which, depending on intensity drives the cell towards apoptosis [229]. Free cytosolic gentamicin is then able to attack mitochondria inhibit ATP production, amplifying oxidative stress and initiating the intrinsic apoptosis pathway [230, 231] through mitochondrial permeability transition pore opening and cytochrome c release into the cytosol. Activation of the extracellular calcium sensing receptor (CaSR) by gentamicin has also been suggested as a cause of tubular cell death [232]. Subsequently the glomerular mesangial cells are affected which manifests as a reduction in the glomerular filtration rate (GFR) and renal blood flow (RBF) contributing to proteinuria augmented by gentamicin-induced tubular damage [132, 233].

Understanding the basic principles of gentamicin's AOP and integrating the data collected from the literature allows a toxicity protection pathway to be built. Figure 2.11 conceptualises the toxicity protection pathway for gentamicin, which is briefly discussed below.

Liposomal formulations of gentamicin as discussed in section 2.7.1.1 have been shown to reduce the volume of distribution of the drug, thereby retaining the drug within the central compartment which prevents kidney accumulation and toxicity [124, 135, 137]. Liposomal formulations act on the pharmacokinetic profile of the drug. There are many formulations which act on the pharmacodynamics of the drug to prevent toxicity many of which are shown in Figure 2.11. Receptor mediated endocytosis of the drug can be prevented using cytochrome c fragments which act as an antagonist to the receptor megalin [225]. Cyclodextrin sulphates have been shown to prevent lysosomal degradation by displacing

gentamicin from the negatively charged lysosomal membranes [234, 235]. Promoting the exocytosis of gentamicin is another effective toxicity reduction approach which has been demonstrated by Beauchamp et al. (1997) [236]. Using the bactericidal drug fleroxacin, co-administered with gentamicin the group were able to show a significant reduction in nephrotoxicity in rats. Morales et al. (2010) [231] used the drug metformin to stabilise mitochondrial permeability and respiration, normally disrupted through the use of gentamicin. The administration of metformin was able to maintain mitochondrial homeostasis, reduce lipid peroxidation and maintain cellular antioxidant mechanisms all of which prevented cell necrosis. Metformin has also been shown to prevent the elevation of BUN and creatinine in rats [237]. Lastly the use of various agents that behave as antioxidants are frequently reported in the literature, gentamicin toxicity results in an increase of reaction oxygen species and reactive nitrogen species, furthermore the depletion of endogenous defences such as GSH, Superoxide dismutase and catalase are frequently reported [238]. Antioxidant agents such as vitamin C & E [238], melatonin [239], trapidil [240], lycopene [241], curcumin [242, 243], probucol [244], eugenol [243, 245] and sesame oil [246] have all been demonstrated to ameliorate gentamicin nephrotoxicity.

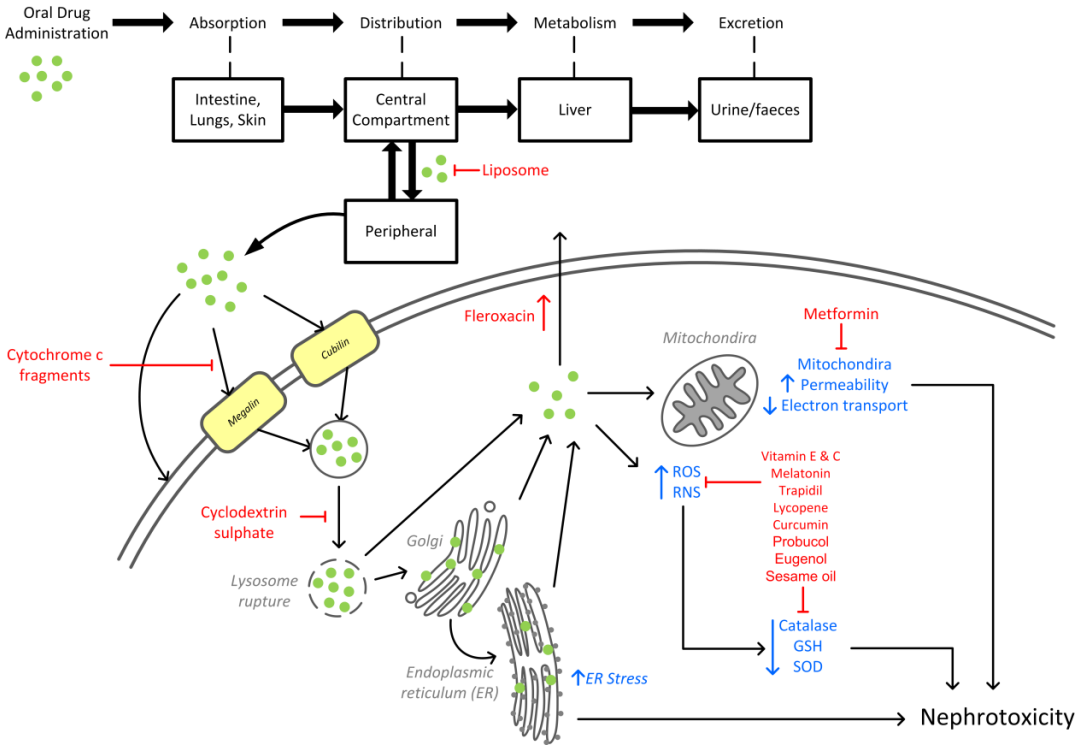


Figure 2.11. Toxicity protection pathway for gentamicin. Formulations that prevent a particular toxicity pathway are shown in red.

The toxicity pathway for gentamicin shown in Figure 2.11 is by no means a complete toxicity profile of the drug. Rather it only includes the parts for which formulations that reduced its toxicity were discovered from the literature. Furthermore when new information becomes available it can also be incorporated into the toxicity protection pathway. If the toxicity pathways for other drugs are known then it may be possible to construct networks for which common pathways overlap. Thus a formulation successfully tried with one drug may be successful with an untested drug if there is some overlap in their toxicity pathways. For example, the literature reports that the nephrotoxicity of cisplatin can also be ameliorated using the vitamins C and E [33, 247]. This suggests that both gentamicin and cisplatin deplete endogenous antioxidant defences as part of their toxicity pathways. Moreover the other antioxidants used with gentamicin such as, lycopene, sesame oil or melatonin could be tried with cisplatin also.

Discovering pathways for drugs however is a labour intensive process, most frequently achieved by piecing together individual bits of information extracted from the literature. Formulating for NCEs may prove to be extremely difficult as very little is often known about their AOPs.

2.9 Summary

This chapter has provided an overview of the scientific aspects of drug formulation for toxicity reduction. To fully understand a drugs toxic pathway, it is important to understand the major aspects of its pharmacokinetic and pharmacodynamic profile. Several different types of formulations have been discussed that are able to affect a drug's pharmacokinetic profile as well as interfere with a drug's pharmacodynamic profile. Although the current technologies which affect a drug's pharmacokinetic profile are limited, they are effective in achieving significant toxicity reductions *in vivo*. The formulations available for modifying a drug's pharmacodynamic response are more varied but frequently tend to be, other drugs, vitamins, oils or antioxidants. Collectively the pharmacokinetic and pharmacodynamic modulation formulations achieve their toxicity reduction in one of four proposed strategies discussed in section 2.7 of this chapter.

Using information extracted from the literature, it is possible to construct toxicity protection pathways based on an AOP framework which maps formulations to specific points of an AOP pathway. If two or more drugs overlap in their toxicity pathway then potentially the formulations along that pathway can be effectively used for both drugs.

Although this has not been demonstrated through experimentation it may be possible to discover examples from the literature that reinforce this theory. Literature searches revealed that the most frequently formulated drugs for toxicity reduction are the cytotoxics and anti-infectives, making them an ideal choice to begin building toxicity protection pathways for.

3

Machine Learning Approaches in Predictive Toxicology

This chapter provides an informative discussion of how experimental toxicity data are processed and prepared for use with machine learning techniques in predictive toxicology. More specifically, details are provided of the procedures required to successfully build classification models and the pitfalls encountered with such work. Information on data gathering, data curation, chemical structure handling and descriptor generation are discussed, as well as the various machine learning techniques currently available, with examples provided of their application in toxicity prediction problems reported in the literature. The chapter ends with a discussion of how predictive classification models are validated and assessed for performance.

3.1 Predictive Toxicology

Predictive toxicology is concerned with the development of *in silico* non-experimental models that make use of existing data for the assessment of chemical toxicity. Whilst a shift away from experimentation sounds ideal, in reality this is not the case. Rather, a welcomed change from assessing adverse effects at high doses in experimental animals has been replaced by a requirement to understand the underlying principal mechanisms of toxicants on organisms at the molecular, cellular and organ level. This type of data can in principle be

gathered using *in vitro* experiments negating or reducing the requirement of animal experimentation and thereby continuing to support the ongoing efforts of the 3Rs.

Predictive toxicology can therefore be considered as the application of mathematical interpretations and computational approaches to predict the toxic potential of a new compound. More significantly these approaches are based on information processing (data gathering, curating and mining) to arrange data in an appropriate manner for computational approaches to be used effectively. Although this thesis is not concerned with the prediction of compound toxicities it does make effective use of many of the approaches used in predictive toxicology to achieve the objectives set out in chapter 1. It is therefore useful to provide a brief overview of some of the techniques and approaches commonly used in this field of study.

3.1.1 *In silico* Approaches to Predictive Toxicology

In silico predictive toxicology can be considered as one alternative to animal testing for which several different approaches have been developed. There are many advantages associated with the use of *in silico* approaches which include the reduction of animal use, cost-effectiveness and fast turn-around of results. These *in silico* toxicity approaches are based on the premise that the toxic manifestation of a drug is related to its chemical structure. Therefore if the chemical structure of a new drug is known, it should in principle be possible to predict a toxic outcome, by relating its structure to a similar structure of known toxicity [248]. Such relationships are termed (quantitative) structure-activity relationships ((Q)SARs) [249]. Whilst this definition of QSAR fits with the description of QSARs recently presented by Tropsha et al. (2010) [250], some authors specifically refer to the term QSAR when SAR relationships become quantified [251]. In this thesis the term QSAR is used to refer to chemical structures whose activity (toxicity) can be reported as continuous (e.g. LD_{50}) or categorical (e.g. toxic and non-toxic) values.

Generally there are two main strategies to building predictive toxicology models; Expert systems and data-driven QSAR methods;

Expert systems are based on human reasoning approaches which rely on an underlying knowledge base containing rules from which toxicity endpoints are derived. The knowledge base is unlikely to contain the exact answer to a particular new problem, but instead the expert system must utilise this knowledge and infer new information from it. Rules are compiled from experimental data of compounds that result in one or more toxic

manifestations within a biological system [249, 252]. DEREK is one such commercially available expert system developed by Lhasa Limited [253]. It utilises a knowledge base derived from human expertise, with structural alerts associated with toxicological effects. Some of the endpoint predictions DEREK is able to provide include, mutagenicity, carcinogenicity and skin sensitisation. OncoLogic [254] and HazardExpert Pro [255] are other examples of expert systems that contain human derived knowledge [256]. A significant advantage of these systems is that they are highly mechanistically based through a thorough analysis of the information, which means the reliability of the knowledge derived rules can be assured [249]. Other expert systems derive their rules for the association of a fragment structure with toxicity automatically. TOPKAT [257] uses statistical methods such as linear multiple regression equations, whilst CASE [258] uses a Bayesian approach to present a probable toxicity endpoint with confidence levels. These systems have the ability to analyse large datasets but provide no mechanistic information [249].

Data-driven QSAR methods develop predictive models from experimental data. Such methods utilise statistical and machine learning approaches to build models from the input data to make data-driven predictions or decisions, without any prior knowledge of the underlying mechanisms (toxicity mechanisms). Machine learning approaches can be split into supervised and unsupervised approaches. Supervised learning approaches include Naïve Bayes, neural networks, decision trees and random forests which are used with labelled data (for example, toxic or non-toxic) to train the models. In supervised learning the data contains an input (X) or set of inputs and an output (Y) label. The task of machine learning approaches is to learn and predict the output for a given set of inputs. Unsupervised approaches include techniques such as clustering or principal component analysis that use unlabelled data to fit models. Since the data used are unlabelled, there is no error or prediction evaluation provided. The aim of unsupervised learning is to assess data distribution, reduce high dimensionality or find patterns hidden amongst the data.

The remainder of this chapter will focus on supervised machine learning techniques that derive QSAR data-driven predictions of toxicology endpoints. Such models use labelled data containing a set of inputs (derived descriptors) and related them to an output (toxicity endpoint). An overview of how such models for toxicity predictions are built and examples reported in the literature is provided.

3.2 Developing Machine Learning Approaches for Toxicity Predictions

One of the main challenges of building predictive toxicity models is information reuse. The aim of such *in silico* approaches is to make use of the vast numbers of experimentally generated data currently available to learn and develop knowledge from. The following sections of this chapter discuss some of the main requirements for building predictive QSAR models using machine learning techniques. Several key steps are discussed, which include; establishing toxicity endpoints, gathering data, curating data, chemical data representation, identifying appropriate machine learning techniques, model validation and performance measurements.

Although other significant factors in developing *in silico* classification models can be mentioned, these key points are relevant to the work in this thesis and are collectively discussed in the sections below.

3.2.1 Toxicity Endpoint and Data Collection

To begin building predictive QSAR models a toxicity endpoint must first be established. Chapter 2 discusses several toxicity types that are often reported in experimental studies, although this is not an exhaustive list it provides some insights into the kinds of toxicity endpoints which may be considered. Although picking an endpoint first may appear rather counter intuitive, it helps determine the nature of the data required for any supervised machine learning techniques to be applied.

Data can be gathered through a variety of sources that pertain to the toxicity endpoint to be modelled. In-house sources may be available to laboratories that generate such data, alternatively publicly available resources can also be considered. Several online databases have been developed for such purposes, many of which are discussed by Toropov et al. (2014) [259]. Searching for specific compounds or toxicity endpoints can also be considered. Gathering individual pieces of data is extremely labour intensive although anomalies can be easier to identify. Using in-house data is also sometimes advantageous if they represent the chemical space for which predictions are sought. Where specific data collections result in scarce data it may be necessary to broaden the search criteria, for example collecting data on one specific species may result in few data, which can be resolved by collecting data on species from the same taxa [260]. Alternatively where data

are plentiful it may be appropriate to reduce the data size, for instance filtering data pertaining to a single sex of an animal for *in vivo* data.

Acquiring data from multiple data sources will inevitably require some element of data integration, in particular when the acquired data are held in multiple formats. Integration techniques will vary according to the complexity of the data being gathered. The many different integration approaches are beyond the scope of this thesis but the literature contains several published reports on the current techniques available [261, 262]. Once suitable data sources are found, capturing these data into a suitable format is critical, for many cases a text or Excel file will suffice and enables modelling software to easily read the data.

The data must then be assessed for suitability, if the information contained does not allow for machine learning approaches to be applied then can any additional data be obtained and added. For instance if data are gathered for the aquatic toxicity to *Daphnia* but it only contains chemical names and no chemical structures, it is possible to obtain structures from alternative sources which can then be integrated. If sufficient structures cannot be obtained, then this will limit the amount of data with which QSAR models can be built. A large data size does not necessarily result in better models; rather the size, scope and chemical space of the data play a significant role in defining the applicability domain of the model i.e. the constraints to which is model is considered reliable [263].

It is worthwhile mentioning here that, for a dataset gathered in a table format (e.g. Excel) the layout of the data in its simplest form should contain chemical information, i.e. a chemical structure and a toxicity endpoint (referred to as the class for classification modelling) corresponding to each chemical structure. For example if the toxicity endpoint being considered was nephrotoxicity, then the class of each chemical could be considered as toxic or non-toxic. For toxicity endpoints such as LD_{50}, the endpoint information may be a continuous value. It is also important to remember that whilst data on compounds which are nephrotoxic may be relatively easy to come by, data reported on the non-nephrotoxic potential of a compound are sometimes more difficult to obtain. For classification models it is imperative that compounds are known to be toxic or non-toxic and not just assumed so because there are no data to suggest otherwise [17].

3.2.2 Data Curation

Whilst a large number of data are desirable, it may contain noise that prevents effective learning. It is imperative to deal with such instances as part of a data curation step. For example if outliers within the data are observed or identical instances resulting in two very different outcomes, for instance one compound labelled as non-toxic whist in a separate record the same compound is labelled as toxic then some form of pre-processing to deal with such occurrences is vital for effective model learning. For classification purposes it is important that two identical instances are not classified differently. Not recognising and dealing with such situations can result in poor predictive performance with results that become unexplainable. Several methods have been developed to assess the quality of the data being used and frequently look at factors such as; the reliability of the data to represent the toxicity endpoint, the consistency of the data to ensure low statistical variance and the reproducibility of the data to ensure the data are independent of any bias [264, 265].

Finally, the chemical structures contained within the data are linked to an activity (toxicity) presented in the data. If a chemical structure is available, but a toxicity endpoint for that chemical is not established then that datum point becomes unusable for modelling. The aim is to establish a link between any structural and/or physico-chemical properties of a chemical and an effect. If mechanistic knowledge is available about structural features then this can be used as a basis for parameter selections that are likely to influence toxicity endpoint.

3.2.3 Chemical Data Representation

Data representation of a toxicity endpoint is relatively straightforward and has been mentioned above and in chapter 2, so is not discussed again in this section. This section rather focuses on the representation of chemical data before QSAR models can be developed.

Whilst there are many chemical structure representation models (Kekulé structure, Lewis structure and condensed structure) with widely established nomenclature processes that provide systematic names such as that created and developed by the International Union of Pure and Applied Chemistry (IUPAC) [266], for computational machine learning approaches a more sophisticated machine readable representation is required. There are essentially two classes of chemical properties that are described from the constituent parts

of a molecule's atoms and bonds; explicit and derived properties. Explicit properties are those required to specify the mathematical graph of a molecule; that is the ordering of the set of atoms (nodes) and bonds (edges) that make up a molecule's structure. Derived properties are those which can then be computed from the graph.

For computational modelling the chemical representation must therefore be such that mathematical calculations can be made to develop descriptors (derived properties) that are able to describe the molecular entity in some way. A descriptor may simply be the molecular weight or atom counts of a compound which can be calculated from the condensed representation of a molecule, for example C_2H_6O, ethanol, whose molecular weight would be calculated as 46.07 g/mol is made up of 2 carbon atoms, 6 hydrogen atoms and 1 oxygen atom. Constitutional descriptors as such are useful but do not provide information about the molecular geometry or atom connectivity. For more complex descriptor calculations a better chemical representation in electronic format is therefore required.

Many chemical file formats have been developed for computational use, which relate a molecule's chemical structure to enable detailed descriptors to be calculated. Most of these formats are based on the physical valence model, which represents molecules based on mathematical graph theory. The usefulness of each format varies with limitations associated with each type. Three of the most commonly used formats, SMILES, InChi and Molfiles are described below:

3.2.3.1 SMILES

SMILES (simplified molecular-input line-entry system) are line notation strings that represent molecules. Typically a number of number of variations can be used to write SMILES which are all equally valid, for example ethanol can be represented as CCO, OCC, C(O)C and C(C)O. Canonicalization algorithms can be used to generate unique SMILES amongst all valid possibilities. Atoms in SMILES are represented by the standard chemical element abbreviation in square brackets, such as [Ag] for silver. Hydrogen atoms are assumed to fill the valence spaces with adjacent atoms connected by single bonds, double and triple bonds are represented as an equal sign (=) and number sign (#) respectively. SMILES carry the same information as connection tables but its advantage is that it is linguistically constructed which can be human-interpreted rather than a computer data

structure. One final advantage of a SMILES string is that it is digitally small and so does not require vast amount of hard disk space to store multiple structures [267-269].

3.2.3.2 InChi

InChi (IUPAC International Chemical Identifier) is a unique line notation string representation of a chemical structure, derived from the structural graph representation of a structure that is independent of the way the structure was drawn. A single molecule will therefore always produce the same InChi, which is a unique identifier for a particular compound. InChi uses hierarchical layers (encoded by a forward slash, /) based on a core parent structure to represent various levels of chemical complexity that can include bond connectivity, isotope information, tautomeric information, stereochemistry and electrical charge information. Not all layers are necessarily required, for example the stereo information layer can be omitted if that information is not relevant to the application. Omitting this information means the same InChi will be produced for the source structure and all of its stereo isomers [266, 270, 271]. The InChi string for ethanol is represented by InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3.

3.2.3.3 Molfile

Molfiles are based on connection tables (CTAB) that contain a table detailing the atoms of a molecule and another table detailing the bonded atom connections and bond types. Connection tables represent a single molecular structure that is divided into several blocks that carry information pertaining to the molecule. The header block carries the name of the molecule, the atom and bond counts and 3D coordinates (x, y, z). The CTAB for ethanol is shown in Figure 3.1. SDF file formats (structure-data files) wrap molfiles and multiple compounds are delimited by four dollar ($$$$) signs. SDF files have the ability to include associated data [272].

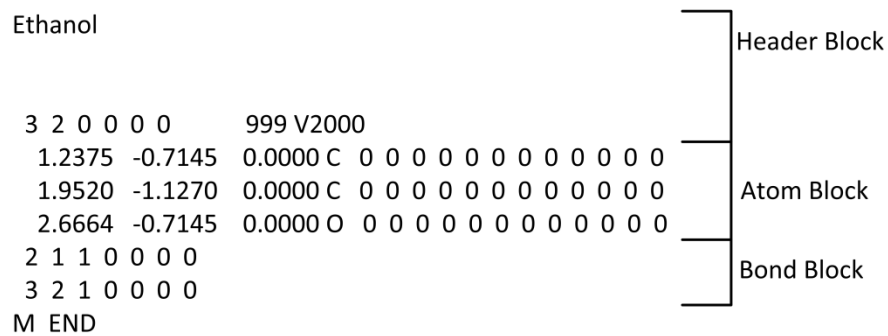


Figure 3.1. Molfile CTAB for ethanol.

The atom block specifies the atomic symbol with mass differences, charge, stereochemistry and any associated hydrogens for each atom. Finally the bond block contains one line per bond and specifies which two atoms are connected by a bond, the bond type along with bond stereochemistry and topology [272].

3.2.3.4 Molecular Descriptors and Fingerprints

Producing informative data from molecular structures/representations is important in computational applications and in particular for QSAR models. Molecular descriptors and fingerprints are alternative ways to encode and represent chemical data; their purpose is to capture the salient aspects of a chemical structure for application with machine learning approaches. The purpose of producing molecular descriptors and fingerprints (derived properties) is to derive a relationship between one or more properties of a compound and the toxicity endpoint being assessed. Thus if a new chemical possesses similar or identical properties then predictions about its toxicity can be made.

A critical step in building predictive QSAR models is to determine which properties are linked causally to the toxicity endpoint being modelled. Some properties will show no correlation, whilst it is possible that some may coincidently result in a superficial relationship, which require identifying and removing if the models are to be correctly interpreted later. For successful modelling it is important to use the properties for which a rationale to the toxicity endpoint can be justified. In many cases however a mechanistic understanding of this relationship is poorly understood which may therefore require the inclusion of chemical properties which are not easily explainable [260].

Molecular descriptors are calculated from the chemical representations discussed above. Various physicochemical properties of a chemical compound are output as numerical values that can be correlated against a compounds activity/toxicity. It is estimated that thousands of descriptors now exist many of which have been extensively reviewed and reported by several authors [273-275]. Most descriptors can be calculated using commercially available software such as CODESSA [276] or DRAGON [277] which also offers a web based service, E-DRAGON [278]. Of the many descriptors available they can be categorised into four main categories as, constitutional, topological, electrostatic and geometrical.

Constitutional descriptors provide relatively simple information of the chemical structure, such as the molecular weight, number of atoms, number of carbon atoms, number of

double bonds to name a few. Such descriptors are useful when the activities of a series of compounds are thought to be correlated to the size of the molecule [279].

Topological descriptors are single valued descriptors that are calculated from the 2D graph of a molecule, and characterise the size, shape or degree of branching of a structure. The Wiener Index is one such example of a topological descriptor which sums the lengths of the shortest paths between pairs of atoms of all the heavy atoms within the graph structure. The Randić index, Zagreb index and Chi connectivity indices are other topological descriptors [280].

Electrostatic descriptors provide information about the molecular charge distribution. Electrical charges in compounds dictate electrostatic interactions that are fundamental to biological interactions. The topological polar surface area (TPSA) descriptor for instance, is a measure of a compound's molecular surface area. The octanol-water partition coefficient (ClogP) is another example of a descriptor with an electrostatic component as are atomic charge and net orbital charge descriptors [275].

Finally geometrical descriptors are derived from the x, y, z spatial coordinates of a structure and describe the size and shape of a structure. Since geometric descriptors require prior knowledge of the atoms in 3D space they can provide more detailed information and discriminating potential over topological descriptors. However the drawbacks of such calculations are that geometry optimisation of chemical structures is required and for flexible molecules, several conformations may exist. Exploring the relationship between a molecular structure and biological activity is often effective through the use of geometrical descriptors since they are often better at describing the 3D interactions of a ligand to a receptor. Some examples of geometrical descriptors include, moments of inertia, molecular surface area, shadow indices, WHIM descriptors and 3D MoRSE descriptors [275].

Molecular fingerprints provide an alternative means of describing the structure of a chemical compound. They work on the basis of encoding the structure of a chemical compound as a binary bit string with each bit representing the presence (1) or absence (0) of a particular group, atom or other structural fragment. Fingerprints are therefore based around a dictionary of existing chemical fragments that are designed to pick out features which may be of interest in chemical structures. There are many different fingerprint generation tools, of which the MACCS keys fingerprints is frequently used [281]. MACCS fingerprints code 166 structural keys into a bit vector. An example of some of the MACCS

fingerprints generated for the drug diazepam is shown in Figure 3.2, which has been adapted from [282].

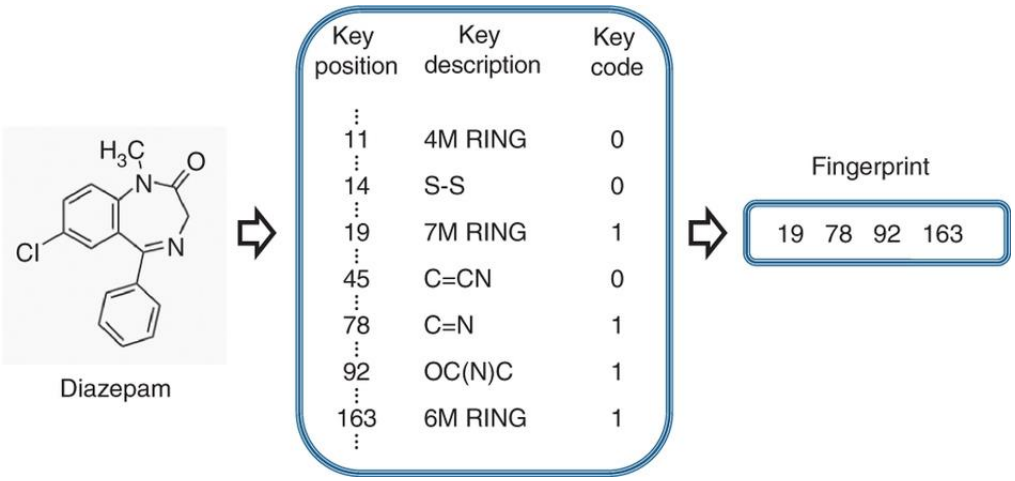


Figure 3.2. MACCS keys fingerprint for the drug diazepam. Image adapted from [282]

The sample fingerprint generated in Figure 3.2 shows the MACCS key positions, the key descriptions and the key code generated for diazepam. Positions in the key code that contain a 1, indicate the presence of that substructure, whereas a 0 indicates absence. For diazepam, the key code shows that a 7-membered ring (7M RING) a C=N bond, a OC(N)C group and a 6-membered ring (6M RING) are present. Whereas a 4-membered ring (4M RING) an S-S bond and a C=CN group are absent. It is important to note that diazepam contains two 6-membered rings, however using a bit vector fingerprint will only indicate the presence or absence and not quantify.

Comparing the bit string of two compounds can therefore provide a measure of "similarity" between their chemical structures. To compare the similarity there are various coefficient metrics which have been developed; examples include the Cosine coefficient, the Dice coefficient, the Hamming distance and the Euclidean distance, which are all extensively reported in the literature [283, 284]. One of the most frequently used similarity measures for binary fingerprints is the Tanimoto coefficient, which is defined as:

$$T_{AB} = \frac{c}{a + b - c} \tag{1}$$

where a is the number of bits set to 1 in molecule A, b is the number of bits set to 1 in molecule B and c is the number of bits common to both. The output range is 0 to 1, although a value of 1 does not necessarily mean the molecules are the same. Another issue with using fingerprints is that some compounds may result in a fingerprint with very little

information or possibly none at all, if any of their sub-structural properties do not occur within the dictionary.

3.2.4 Model Selection

Through prior knowledge of the type of data available an appropriate machine learning modelling technique can then be chosen. It is important to consider that some machine learning techniques such as neural networks are difficult to interpret with limited transparency whilst techniques such as decision trees are considered more transparent given a pathway to a prediction can be logically followed.

To remain within the scope of this thesis a brief overview with examples is provided for the various machine learning techniques available that have been used to develop QSAR classification models for toxicity predictions and reported in the literature. For the purpose of categorical toxicity endpoints (i.e. toxic or non-toxic) classification algorithms tend to be the obvious choice of machine learning techniques used. Whilst there are a host of different techniques available this section focuses on those most commonly used in predictive toxicology modelling.

3.2.4.1 Artificial Neural Networks

Artificial neural networks (ANN) are models based on the architecture of biological neural networks. They consist of three or more layer which consists of an input layer, a hidden layer or layers and an output layer (Figure 3.3). When training, ANNs carry a weight between each neuron which are varied as the network learns how to form a relationship between the input and output layers. When constructing models the number of number of hidden layers and elements can be varied to optimise the model [285]. Although inspired by the structure of biological brains, an ANN is not capable of simulating higher brain functions. In actual fact a typical ANN is smaller than the brain of a 302 neuron *Caenorhabditis elegans* [286].

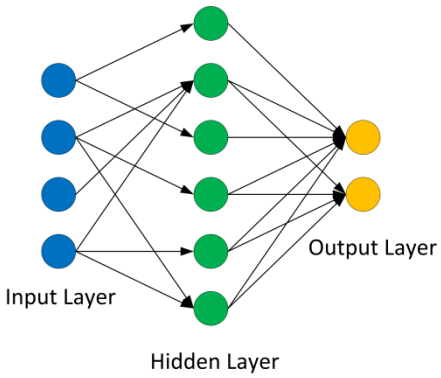


Figure 3.3. Artificial neural network, consisting of an input layer (blue), one hidden layer (green) and an output layer (yellow).

ANNs have been applied to a wide range of problems which include predicting bioactivity of oestrogen receptor agonists [287], and kinase inhibitors [288]. In relation to toxicity predictions, ANNs have been applied in aquatic toxicity studies [289, 290], hERG blockade studies [251] and for the prediction of toxicity of benzene analogues administered as an oral dose to rats [291]. Interestingly the authors of the benzene analogue study reported that the performance of the model was dependent on the molecular representation of their chemical structures and the choice of training set [291]. This supports the theory that ANNs suffer from overfitting and are susceptible to learning noise that is present in the training set, making them less able to generalise for the test data [292].

3.2.4.2 Decision Trees and Random Forests

Decision trees (DT) form a model of decisions which result in a tree like structure. At each node of the tree a branch is split based on the actual values within the training set applied sequentially. A DT begins at the root node, which contains all the training data, as it splits into intermediate nodes each of these nodes contain the sum of the samples connected below it, until a terminal node is reached (Figure 3.4). Intermediate nodes carry the information of the training set whilst a terminal node results in a “decision” (e.g. toxic or non-toxic). Once a DT has been established a query compound is channelled to a terminal node through decisions based on the properties of its structure.

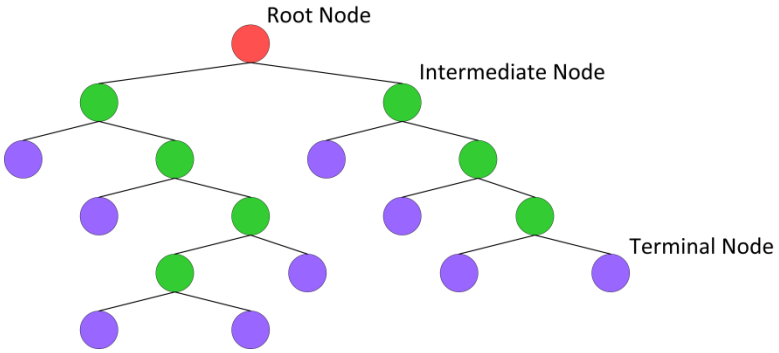


Figure 3.4. Decision tree layout, showing the root node (red), intermediate nodes (green) and the terminal nodes (purple).

There are many DT algorithms which have been developed, of which the CART [293] and C4.5 [294] are particularly popular [295].

Random forest (RF) is an ensemble method based on a forest of decision trees [296]. It works on the principle that many decision trees collectively, will predict better than a single tree. For each tree built, RFs continually split the training data into two randomly chosen sets, a local set to grow trees with and an out-of-bag (OOB) set to assess the goodness of fit, this process is known as the bootstrap phase. The OOB set may contain approximately 37 % of the training set data selected randomly [285]. Independent decision trees are then constructed from the random selection of variables (compound properties) sampled independently from a subset of variables. As the tree grows each node becomes increasingly more homogenous until a terminal node is reached. The Gini criterion is used to ensure each node of a tree is partitioned in the best possible way [297]. Each tree is grown to the largest possible size with no pruning. The output of a RF is the mode of the prediction of the individual trees. The number of trees that a random forest builds is a parameter set by the user, with Svetnik et al. (2003) [298] demonstrating that 500 trees is usually sufficient for many models. It is useful for binary classification purposes that an odd number of trees are chosen so the mode prediction cannot be evenly split. In any case if an even number of trees is chosen for a binary class a random selection between the two predictions is made. RFs are generally better adapted to overfitting and are considered more robust to parameter changes, unlike that seen with ANNs [298].

DTs and RFs have been successfully used in many predictive toxicity studies. Using molecular descriptors of phenols Ren et al. (2003) [299] used DT models to predict the aquatic toxicity mechanism of this group of compounds, whilst Martin et al. (2013) [300] demonstrated the ability of RF models to predict aquatic toxicity mechanisms of action

using a dataset of 924 compounds for which they achieved overall accuracies of 84.5 to 87.7 %. Cheng et al. (2003) [17] utilised an ensemble recursive partitioning approach to predict dose-dependent human hepatotoxicity, using only 2D structural information they achieved accuracies of greater than 80 %. RF models have also been successfully validated against external datasets for the prediction of the aquatic toxicity of *Tetrahymena pyriformis* [301]. The nephrotoxicity data of 41 well-characterized compounds that are toxic or non-toxic to human primary renal proximal tubular cell (PTCs) were used to build RF predictive classifiers. This study correlated the expression of the biomarkers interleukin-6 and 8 to the toxicity of a compound. They found that their RF model produced a balanced accuracy of 87.8 %, whilst an SVM model resulted in a balanced accuracy of 81.6 %. Interestingly the authors reported interleukin-8 alone to result in higher accuracies than interleukin-6 alone [19]. A range of other toxicities have also been investigated using RF models which include mutagenicity [302] and skin sensitisation [303].

3.2.4.3 Support Vector Machines

Support vector machines (SVM) work on the basis of identifying a separation line of some hyperplane which separates molecules of a dataset represented as an instance in the "feature space" of their descriptors. SVMs aim to map the data into this dimensional space using a linear or non-linear function, such that each instance of a binary class lies on opposite sides of the hyperplane. Many linear hyperplanes are potentially possible, but an SVM will attempt to maximise the margin between two classes. A good SVM model is one that achieves the maximum margin between the closest data points (known as support vectors) and the hyperplane (Figure 3.5).

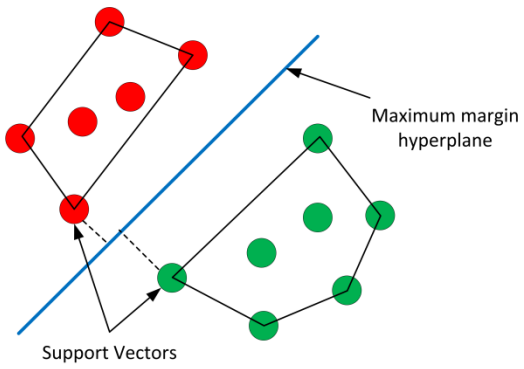


Figure 3.5. Support vector machine, showing maximum margin hyperplane and support vectors for a binary class.

The benefit of finding the maximum hyperplane margin means the model can offer the best generalisation possible allowing plentiful room for the correct classification of a test

compound [295]. Misclassified compounds of an SVM model will most probably lie close to the hyperplane [263]. SVMs can be adapted to multiclass problems or regression [304, 305].

For prediction classification problems SVMs have been used for a range of different toxicity endpoints. Lowe et al. (2010) [306] demonstrated a superior predictive accuracy of SVMs when compared to an RF model for the prediction of phospholipidosis. They also reported on the improved performance of their model through the use of circular fingerprints (representing structures by atom neighbourhoods [307]) when compared to descriptors generated by the E-DRAGON software. Cardiotoxicity studies that predicted hERG inhibition have also been successful using SVM models. Li et al. (2007) [308] showed their SVM binary classification models were able to distinguish between hERG channel blockers and non-blockers up to an accuracy of 94 %, with a similar study reporting accuracies of 80 % using an external validation test set [309]. Mutagenicity is another toxicity endpoint extensively studied using predictive QSAR classification models, Liao et al. (2007) [310] developed SVM classifiers and tested their models on two publicly available datasets and an in-house agrochemical dataset for which they reported predictive accuracies of 80 % or greater for all datasets. Interestingly a study reported by Low et al. (2011) [18] replaced chemical descriptors with toxicogenomic biological descriptors and reported predictions for rat hepatotoxicity, using SVMs and RFs, which produced better predictions than could be achieved using conventional chemical descriptors. Furthermore, through this approach they were able to identify 85 transcripts relevant to the mechanism of drug-induced liver injury which were then linked to chemical structural alerts [18].

3.2.4.4 k-Nearest Neighbour

The k-nearest neighbour (kNN) algorithm is one of simplicity; its principle being that a prediction (classification) is made based on the number (k) of its nearest neighbours within a specified feature space. In a QSAR model the training compound descriptors described as position vectors would constitute the feature space, which can be of a high dimensionality. Neighbours are therefore measured within this feature space using a distance measure, typically the Euclidean distance, although alternative metrics can be considered. Since kNNs do not relate the descriptors directly to an activity (toxicity) there is relatively low risk of under- and overfitting. Rather kNNs work on the hypothesis that similar chemical structures will result in similar toxicity profiles. Under this premise however they can be

prone to active cliffs, for which similar structures do not constitute to similar toxicity profiles [311].

If k=1 then the prediction is simply the single nearest instance to the test compound. Generally the value of k in kNN models is set low, although this makes them susceptible to noise. If the k value is set too large however, then instances from other classes may be included into its neighbourhood [295]. Consider the example shown in **Figure 3.6**. For a test compound (purple), if k is set to 1 it is simply predicted within the same class to which the yellow compounds belong. If k is set to 2 then it will still be predicted as the same class to which the yellow compounds belong given it is out voted 2:1 (yellow:green). If however k is set to 3, then the test compound is now near 3 green compound and 2 yellow compounds it is therefore out voted 3:2 (green:yellow) and will be classified into the same class as what the green compounds belong to. For binary classification problems it is sometimes useful to select an odd number for k, to avoid the chance of tied votes.

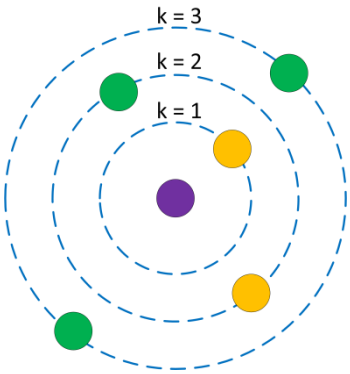


Figure 3.6. k-nearest neighbour model. Classification of the purple compound will depend on the k value set by the user.

kNNs work well with localised data which form distinct clusters within the feature space, with global coverage possible if the feature space is evenly distributed. Since kNNs rely on a distance to a neighbour(s), in principle this value could be used as a measure of prediction confidence.

Due to their simplicity kNN models are frequently used with classification toxicity problems. A kNN QSAR model predicting the algal toxicity of 91 compounds against *Chlorella vulgaris* revealed three descriptors related to the hydrophobicity, electrophilicity and a function of molecular size corrected for the presence of heteroatoms resulted in an effective model, whilst a multiple linear regression approach was unable to model the data

[312]. A hybrid SVM-kNN model for kinase inhibition was shown to produce effective predictions on an in-house dataset used by Briem et al. (2005) [288].

Using a relatively small dataset of 26 compounds, Honorio et al. (2005) were able to model the psychoactivity of cannabinoid compounds, successfully validating the model with an external validation set of 8 cannabinoid compounds with known activity [313]. This study demonstrates the effective use of kNNs with small datasets of localised data, although it may be limited when applied to larger or more diverse datasets. Furthermore, Gadaleta et al. (2014) [314] developed kNN models using repeated dose toxicity data of rats for which their model was only able to predict correctly a small number of compounds, showing the limited capability of kNNs when the feature space complexity is high.

3.2.4.5 Naïve Bayes

Naïve Bayes classifiers use a branch of mathematics known as probability theory to find the most likely classification. An estimation of the *posterior* probabilities of class is determined based on feature information. It assumes that the features, f_i, are conditionally independent of each other given the output classification, *c*, which results in the formula:

$$P(c|f_1, f_2, \dots f_n) = \alpha P(c) \prod_{i=1}^{n} P(f_i|c) \quad (1)$$

A prediction results in a class being assigned with the highest estimated probability. For QSAR models however, the assumption of the conditional independence of the features (descriptors) is not always valid, but can be obtained through feature selection procedures. Naïve Bayes are prone to false predictions if over-weighted or irrelevant features are selected and more so if the training set contains a different distribution of toxic/non-toxic compounds to the test set.

Nevertheless many successful classification models for toxicity predictions have been demonstrated using naïve Bayes classifiers. A study [315] reported on the use of naïve Bayes and RF classifiers to map *in vitro* toxicity results of pesticide compounds into an *in vivo* effect. Although many different classifiers were tested the most effective results were produced by the naïve Bayes and RF models, interestingly they also demonstrated that using a lower number of trees in their RF models improved their results. Whilst the authors were generally able to predict toxicity they were unable to successfully predict a specific endpoint, determined by the various *in vitro* toxicity assays that the pesticides were tested

against. Unfortunately the study did not take into account a descriptor reduction step which resulted in a sparse feature space and may be the cause of their RF models producing better results with a lower number of trees.

3.2.5 Model Validation

Models that fit data are only useful if they are able to accurately predict for the general case. A model must therefore be able to generalise from the data used to train it and make predictions for an unknown instance(s). To assess a trained model's ability to do this it must be validated. The most common approach used is to randomly split the originally collected dataset into two sets; a training set and a test set. The split is typically 80:20 or 70:30, training:test, although other splits can be used. This method is useful since it does not require a second round of data collection and curation which is a lengthy time consuming process. The training set is then used to train the model to learn from the information contained within it. The classification properties of the training set are known to the model. Once trained, the model is then run against the test set, for which the classification values are hidden and not known to the model. The concordance of these predicted values, against the hidden actual values is measured to assess the accuracy of the model. This form of validation is considered an "external validation," provided any feature selection or model parameter settings are not based on the test set and it is solely used to assess model predictions once training is complete. Some authors argue that a true external validation set is compiled of data from an alternative source to the original dataset [316-318]. Data contained in the originally collected dataset may contain compounds within the same chemical space, in particular if in-house data is used. Splitting this data into a training set and test set is therefore likely to result in good predictions given the test set is similar to the training set.

A counter argument however, can also be made against using a test set that is chemically very diverse to the training set. For effective predictions to be made the test set must fall within the applicability domain of the training set [263]. A test instance that has no similar instance to it within the training set is less likely to be predictable, resulting in poor model performance.

More efficient ways to split the dataset into a training set and test set have been developed which arguably allow better training to occur. Of these the cross validation method is frequently used. An *n*-fold cross validation involves splitting the data randomly or in a

stratified way to result in *n* different folds. One fold is used as the test set while the remaining *n*-1 folds are used as the training set. Each fold is then cyclically permutated so that every fold is used as the test set once with the remaining *n*-1 folds used to train the model. Using this method, n models are created which sequentially train on all the data, whilst also providing predictions for each of the instances within the dataset [319].

Careful consideration must be taken when using a cross validation process; any feature selection techniques or model parameters used should be independent of the *n* test sets and any information regarding the test set must not find its way into the model building process. Typically a 5- or 10-fold cross validation technique is employed although a special case of this validation, termed Leave-One-Out (LOO) cross validation is sometime used. LOO, as the name suggests involves leaving one instance for model validation while training on the remaining data set. This process is recursively looped until all instances within the dataset are predicted for.

Another approach to testing the reliability of a model is known as y-randomisation or y-scrambling. In this approach the class (or y values) are randomly assigned to instances within the dataset and the models are run according to the settings and parameters previously tested. Randomising the class, disrupts any link between property descriptors (features) and the class so no possible correlation remains. If the models are still able to produce equivalent accuracy measurements then the models are considered unreliable as they are most likely modelling noise rather than signal [318].

3.2.5.1 Measuring Classification Model Performance

For QSAR classification models, the predictive accuracy of a model can be measured by the number of correct predictions divided by the total number of predictions made. For example, if a model was asked to predict 10 compounds for which it was correct 7 times, the accuracy could be considered to be 70 %. For classification problems however this is an inadequate measure of model performance.

Consider a binary classification model for which the classes are disproportionately split, say, 90:10. A model which favoured the major class, every time it made a prediction, would inadvertently seem to be highly predictive. However in reality due to the unbalanced dataset the model is unable to discriminate between the classes. An alternative measure is therefore required.

Considering a single class of a binary classification model, predictions can be considered as correctly recognised class examples (true positives (TP)), correctly recognised examples that do not belong to the considered class (true negatives (TN)), examples incorrectly assigned to the class (false positive (FP)), and examples not recognised as class examples (false negatives (FN)), satisfying; TP+FP+TN+FN = N, with N being the total size of the test set [320]. A confusion matrix of the observed against the predicted would look like (Table 3.1):

Table 3.1. Confusion matrix for a binary classifier.

		Observed	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

From the confusion matrix shown in Table 3.1 the accuracy discussed above can be represented as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

The precision of a model can be defined as a measure of the accuracy of the number of elements in a specific class, regardless of whether they should be present or not. It is given by:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Using the confusion matrix it is possible to calculate the sensitivity, sometimes referred to as the recall and is considered as the true positive rate, i.e. the number of true predictions divided by the total number that should be true. For a binary classification this can be considered as the correct classification rate of one of the two classes and is given by:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

Furthermore the correct classification rate of the other class is given by the specificity, which is a measure of the true negative rate, i.e. the number of negative predictions divided by the total number that should be negative. This is given by:

$$Specificity = \frac{TN}{FP + TN} \tag{5}$$

For an unbalanced binary dataset, for which the accuracy is not a true reflection of model performance, a mean of the correct classification rates of the two classes can be calculated to give the balanced accuracy of a model. The balanced accuracy is able to avoid inflated accuracy rates on unbalanced datasets, it is given by:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{6}$$

Similarly, the Matthews Correlation Coefficient (MCC) can be used to assess the overall discriminative ability of a binary classifier. As with the balanced accuracy, the MCC takes into account all the values from the confusion matrix and is particularly suited with highly unbalanced datasets [321, 322]. The MCC is given by:

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP + FP)\ (TP + FN)\ (TN + FP)\ (TN + FN)}} \tag{7}$$

The MMC value is always between -1 and +1. A value of -1 indicates total disagreement (i.e. all predictions were incorrect) and a value of +1 which indicates total agreement (i.e. all prediction were correct). A value of 0 (zero) suggests predictions are made randomly [322].

For binary classification models, the performance measurements described above are the ones most frequently reported in the literature. Whilst other statistical measures can be explored [320], they are not discussed here as this thesis is concerned with classification models only.

3.2.5.2 Applicability Domain

The concept of an applicability domain is considered as the chemical space in which a model makes reliable accurate predictions. Therefore the output of the model is only valid if the compound being predicted lies within the applicability domain [323]. This domain can be defined as a theoretical multidimensional chemical space [318], where each dimension

can represent the structural, biological, or physico-chemical properties of a compound [324]. Since the training set used to train the model represents a small proportion of the chemical space, it seems reasonable to set the domain from the compounds used to train the model. Most methods reported in the literature provide a numerical value such as a similarity measure between the test and training set [324]. This value can then be used as a threshold for further predictions. Predictions made outside of the domain can be high but considered unreliable, on the other hand predictions made within the domain can be low but deemed reliable [318, 323].

3.3 Summary

In this chapter the development of QSAR models using machine learning techniques for classification problems has been discussed. A stepwise approach which begins with establishing a toxicity endpoint prior to data collection and curation is explained. Information regarding chemical structure handling and chemical representation in terms of descriptor and fingerprint generation is explained. An understanding of the importance of descriptor/fingerprint generation and potential correlation against activity (toxicity) is one of the key concepts of predictive toxicity models. An overview has been provided of some of the popular machine learning techniques used with classification problems and their relative abilities have been discussed. Furthermore, examples of ANNs, DTs, RFs, SVMs, kNNs and Naïve Bayes models reported in the literature and their success rates as reported in the literature have been discussed. Reasons for model validation and the approaches used have been described and the statistical methods used to assess classification model performances have been discussed. Finally the applicability domain of a model has been discussed with a brief overview of its definition and how the domain parameters can be set.

Whilst a comprehensive in-depth detailed discussion of all the methods used in computational toxicology problems is beyond the scope of this thesis, this chapter aims to provide an informative discussion of the methods used and explored in subsequent chapters.

4

Graphs for Visualising and Extracting Data

This chapter is concerned with the application of graph theory for scientific data representation, visualisation and the quick identification of drug-vehicle relationships, which show significant changes in patterns of toxicity. A novel bipartite graph preparation methodology is proposed and discussed for the rapid extraction of interesting relationships within large datasets that require minimal data curation.

Furthermore, the toxicity differences afforded by two different vehicles for a specific drug or group of drugs can be arranged into a Venn diagram to identify multiple vehicle relationships and identify the number of data points that are available for these relationships. This methodology therefore allows the rapid visualisation and identification of data a scientist may consider interesting to work with. A practical example of how this methodology can be applied to real scientific data is demonstrated in this chapter.

4.1 Graph Theory

Graph theory is an area of discrete mathematics that uses complex networks to solve practical problems. In the domain of science alone, graph theory has been used for brain networks, gene networks, protein side chain predictions, molecular descriptor calculations and cellular functional organisations [275, 325-328].

A graph *G* comprises a pair of sets (*V*, *E*) where *V* is the set of vertices or nodes ($v_1 \ldots v_n$) and *E* is the set of edges or arcs formed between pairs of vertices. *E* is considered a multiset given that its element can occur more than once. Any two vertices in *G* are said to be linked if there exists an edge that connect vertices v_i to v_j given that both v_i and v_j are elements of *V* ($v_i \in$ V and $v_j \in$ V). The edge connecting vertices v_i and v_j is denoted as (v_i, v_j). The total number of vertices within a graph is denoted as $|V|$ and the set of vertices adjacent to vertex v_i, are referred to as the neighbours of v_i. The number of edges connecting to the vertex v_i, is referred to as the degree of v_i. Whereas the number of vertices connected to edge e_i, is referred to as the degree of e_i. Edges can be weighted or carry additional information which may represent the relationship of the two vertices it connects. For a subgraph *G'* to be formed from graph *G* it must satisfy the relationships of $V' \subseteq V$ and $E' \subseteq E$. A graph is considered undirected if there is no start or endpoint between two connected vertices. Similarly a directed graph or digraph is therefore a graph whose vertices are connected via directed edges.

A bipartite graph is a graph whose vertices' set *V* can be divided into two disjoint sets *X* and *Y*, with each edge $e \in E$, connected at one end to a vertex in *X* and at the other end to a vertex in *Y*. Vertices that are elements of *X* or *Y* cannot be connected to one another by an edge for bipartite to hold true. A directed bipartite graph would therefore contain directed edges which connect the two disjoint sets (see Figure 4.1).

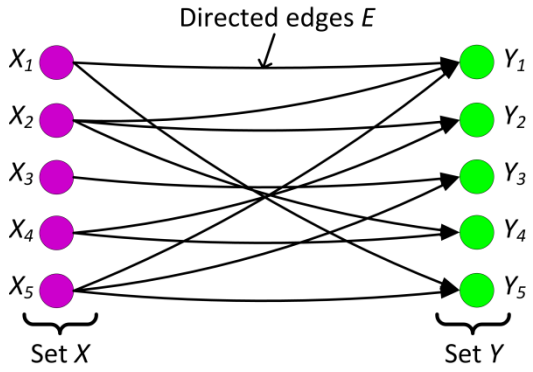


Figure 4.1. Directed bipartite graph showing two disjointed sets of *V* (X and Y) connected by directed edges *E*.

If all the vertices of a graph *V* are ordered ($v_1 \ldots v_n$), then the edge connection between any two vertices (v_i and v_j) can be give given by the adjacency matrix *A* of *G*:

$$a_{ij} = \begin{cases} 1 & if \quad v_i v_j \in E \\ 0 & if \quad v_i v_j \notin E \end{cases}$$

4.2 Data for Graph Representation Methodology

The dataset used for the work in this chapter (NIH dataset) was obtained from the National Institutes of Health's Developmental Therapeutics Program (DTP) [20]. The dataset is publicly available and has been collected over a period of approximately 30 years, beginning in the 1950's up until the 1980's. The dataset contains experimental records from their in vivo screening program against animals inoculated with cancer cell lines to assess the effectiveness of the compounds against the developing tumour. The dataset contains 2,724,199 records detailing the in vivo experimental screening results of 227,093 unique compound identification numbers (NSC). Each compound was tested by one or more laboratories (screener) at various doses against different cancer cell lines in various animal models. Experiments were most commonly conducted in mice, rats and hamsters using different dosing schedules and administration routes. Test compounds were administered using a range of different vehicles, which facilitates the administration of the compound. There are 39 different vehicles used throughout the dataset for which some vehicles had been used more frequently than others (Figure 4.2)

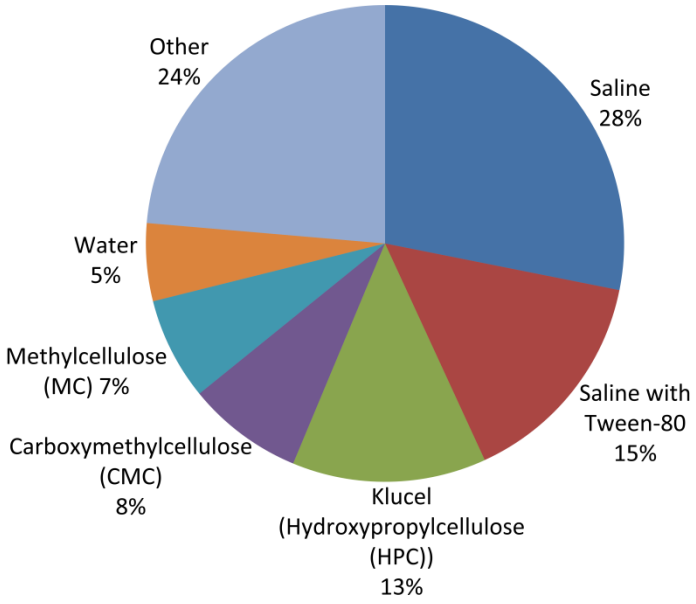


Figure 4.2. Vehicle distribution frequency of the NIH dataset. A total of 39 different vehicles are present, of which some vehicles have been used more frequently than others. The six most frequently occurring vehicles in the dataset are shown in the pie chart with the remaining grouped into the 'Other' category.

Although the primary objective of the DTP was to discover antineoplastic compounds, the work in this chapter aims to visualise these data through the methodology proposed below, to demonstrate how large numbers of data can be rapidly processed and interesting

relationships extracted for scientific purposes. For each experiment detailed in the dataset the survival rate on a specified day was recorded as a measure of toxicity. This information is used to construct bipartite graphs which allow for rapid identification of interesting drug-vehicle relationships that show significant changes in patterns of toxicity.

The NIH dataset is made available in appendix A as an electronic file.

4.3 Data Pre-Processing

The methodology proposed in this chapter requires a minimal amount of curation for interesting drug-vehicle relationships to be visualised and extracted. All data curation and visualisation was handled using a Knime workflow [329]. The NIH dataset was downloaded as a 63 x 2,724,199 data table, and interpreted using a document referred to as "instruction_14," available for download from the DTP website [20].

For each of these 2,724,199 experiments within the NIH dataset, a number of animals, typically 6, 8 or 10 were used per dosing schedule. At a specified experimental endpoint (TOXDAY) the toxicity outcome of these experiments in terms of animal survival was reported. The dataset therefore contains a record of the animals at the start of an experiment (SURVIVOR_NUMBER_START) and also a record of the animals remaining at the end of an experiment (SURVIVOR_NUMBER_TOXDAY).

To compare two individual experiments that used identical experimental conditions but different starting animal counts, the % survival of each experiment was calculated as:

$$\% \text{ survival} = \frac{\text{SURVIVOR\_NUMBER\_TOXDAY}}{\text{SURVIVOR\_NUMBER\_START}} \text{ x } 100 \quad (1)$$

Any records which resulted in a % survival rate of > 100 % were removed from the dataset. These are clearly errors as the surviving animals at the end of an experiment cannot exceed the number of animals at the start of an experiment. To improve computational processing times, columns carrying information deemed unnecessary for the purpose of this work were removed, although this is not considered necessary as part of this methodology. For example the column labelled TREATMENT_TIME, recorded the time at which dosing began. This information was considered not to have an influence on experimental outcome (toxicity) and so was entirely removed from the dataset. Other columns that were removed included, but were not limited to, the date of the experiment (TEST_DT) and the laboratory (SCREENER) in which the experiments were conducted. Attempts to identify any correlation

between the columns removed, for example, SCREENER, and the survival outcome were unsuccessful and so removal of these columns was considered acceptable. Other columns were removed based on the limited information that they carried. For example a column that recorded the sex (SEX_CD) of an animal was present. However of 2,724,199 records available, 463,450 were recorded as male, 448,483 were recorded as female, 259 as mixed sex and 1,812,007 fields were missing. If the records that were missing were entirely removed from the dataset, this would significantly reduce the number of data available to build graphs and extract interesting relationships from. For this reason it was decided to aggregate the data in this column, i.e. experiments were not separated by sex.

It is important to note at this point however that an inoculated tumour growing on an animal host can affect the survival outcome of the animal, if a tumour grew rapidly then there is a possibility it will lead to lethality. The decision to aggregate on the tumour code column (TUMOR_CD) was taken since the survival outcomes of animals were recorded on day 0 (zero) or day 5 (TOXDAY) after the first dose was administered and as such the toxicity outcome would be largely influenced by the drug administered rather than the developing tumour.

The columns which remained were considered important to experimental outcome and are listed in Table 4.1 below:

Table 4.1. List of columns retained after data curation of the NIH dataset for data visualisation. The columns listed were considered influential on experimental toxicity outcome i.e. survival rate.

Original column heading	Description of column
NSC	Identification number of compound tested (NSC- National Service Number)
HOST_GROUP_CD	Type of animal
HOST_CD	Species/strain of animal
ADMIN_ROUTE	Route of administration of drug
INTERVAL	Time between treatment (dosing) in terms of interval unit
INTERVAL_UNIT	Minutes (M), hours (H) or days (D)
VEHICLE_CD	Vehicle used
NUMBER_INJECTIONS	Total number of injections administered
FIRST_INJECTION_DAY	Day of first administered injection relative to day zero (inoculation)
REPETITION	Repetition of injection cycle
RESTART_DAYS	Days on which repetition occurs
DOSE_AMOUNT	Dose in mg/kg unless otherwise stated
TOXDAY	Day toxicity is measured
SURVIVOR_NUMBER_START	Animal count at start of experiment
SURVIVOR_NUMBER_TOXDAY	Animal count on toxday

Once the % survival for each of the 2,724,199 individual experiments is calculated a % mean survival value is calculated for any replicate experiments that are present. To do this all records that contained identical experimental conditions were grouped together and a mean of their % survival values is calculated. The experimental conditions that were aggregated include: NSC, HOST_GROUP_CD, HOST_CD, ADMIN_ROUTE, VEHICLE_CD, NUMBER_INJECTIONS, FIRST_INJECTION_DAY, INTERVAL_UNIT, INTERVAL, DOSE_AMOUNT_UNITS, DOSE_AMOUNT, TOXDAY, REPETITION and RESTART_DAYS.

4.4 Methodology of Graph Construction

To build graphs for the visualisation of interesting relationships and data extraction, the challenge is therefore to arrange the NSC, DOSE_AMOUNT, VEHICLE_CD and % mean survival into a graph's vertices such that the toxicity differences offered by different vehicles can be easily visualised and relevant information readily extracted. A further aspect of graph representation in this study was to show that minimal data pre-processing is required from which interesting data relationships can be extracted.

A rather discernible observation is that the drug (NSC), dose (DOSE_AMOUNT) and vehicle (VEHICLE_CD) are considered the causality of % mean survival. The opposite (i.e. % mean survival causing drug, dose amount and vehicle) cannot hold true, therefore the first point to be made is that the graph to be created must be a directed graph (digraph) which has a direction from the drug, dose amount and vehicle leading to the % mean survival.

Initially a directed graph (G_1) was constructed with the NSC, DOSE_AMOUNT, VEHICLE_CD and % mean survival represented as vertices with directed edges connecting one vertex to another. The digraph G_1 can therefore be represented as; $G_1 = (V_1, E_1)$, where set V_1 consists of the four elements shown in Table 4.2 below.

Table 4.2. Elements of set V_1 for digraph $G_1 = (V_1, E_1)$.

Vertices (V_1)	Elements of V_1	Connecting to element of V_1
NSC (Drug)	*N*	*D*
DOSE_AMOUNT	*D*	*VE*
VEHICLE_CD	*VE*	*S*
% mean survival	*S*	-

The digraph G_1 is therefore a directed graph which forms a set of edges (E_1) with the elements of set V_1 in the order $N \rightarrow D \rightarrow VE \rightarrow S$. A mock example of graph G_1 is shown in

Figure 4.3 below which is produced from the mock data shown in Table 4.3. For illustrative purposes actual values from the NIH dataset have been replaced with mock values, although the heading titles of each field have been retained.

Table 4.3. Data table for digraph G_1. Although the data headings are retained from the NIH dataset, the values for each instance are arbitrary.

Row Number	NSC	DOSE_AMOUNT	VEHICLE_CD	% mean survival
1	Drug X	10	Vehicle 1	100 %
2	Drug X	10	Vehicle 2	80 %
3	Drug Y	10	Vehicle 1	80 %
4	Drug Y	20	Vehicle 1	100 %
5	Drug Y	20	Vehicle 2	80 %

Digraph G_1 (Figure 4.3) provides a simple visualisation of the mock data showing the relationship between drug, dose, vehicle, and survival represented by the elements *N*, *D*, *VE* and *S* respectively of set V_1. The digraph is arranged so that it shows a drug used at a specific dose and vehicle leading to a toxicity outcome. For example rows 1 and 2 of Table 4.3 show Drug X at a dose of 10 mg/kg being used with Vehicle 1 and Vehicle 2 resulting in a % mean survival of 100 and 80 % respectively. These two relationships can be visualised from graph G_1 shown in Figure 4.3.

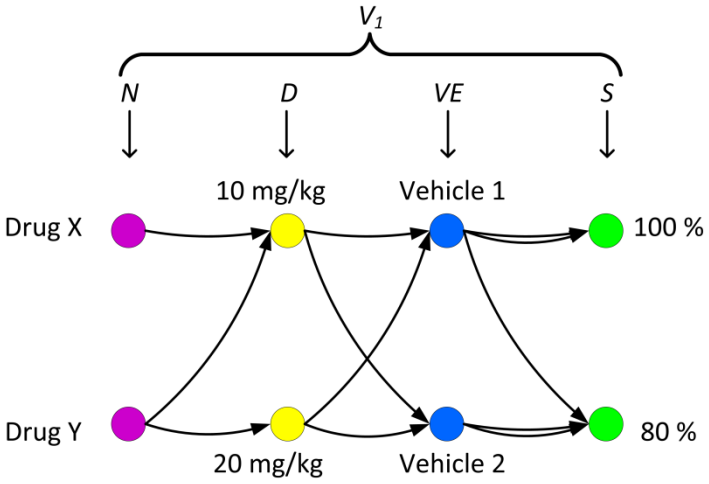


Figure 4.3. Digraph G_1 showing vertices V_1 connected by edge set E_1 (arrows). V_1 contains the elements *N*, *D*, *VE* and *S* shown as purple, yellow, blue and green vertices respectively.

A major concern related to this form of graph layout however is that the individual relationships can be difficult to follow without further information being included into the graph or by referring back to the original data table. For instance, at the vertex representing Vehicle 1 there are 3 directed out-degree edges that leave this vertex. Two

edges lead to the vertex 100 % and one edge leading to the vertex 80 %. Since the in-degree edges of the vertex Vehicle 1 can come from three possible sources (Drug X/10 mg/kg, Drug Y/10 mg/kg or Drug Y/20 mg/kg) it is difficult to ascertain from the graph which out-degree edge belongs to which of the three possible in-degree permutations.

Given the issues associated with graph G_1 an alternative solution was required. To maintain the simplicity of graph representation it was not ideal to add more information to the graph making it overtly complex to visualise and analyse.

An alternative graph representation could therefore remove element *S* of set V_1 from graph G_1 and add this information as a weighted edge label. The proposed graph G_2, can therefore be represented as $G_2 = (V_2, E_2)$ where set V_2 contains the elements *N*, *D* and *VE* ordered as $N \rightarrow D \rightarrow VE$ and set E_2 now contains the element *S* ($S \in E_2$). Edges that connect vertex *D* to vertex *VE* can now be labelled with element *S* of set E_2. Using the data presented in Table 4.3 graph G_2 can be visualised as shown in Figure 4.4 below.

Graph G_2 shows the set V_2 containing the elements *N*, *D* and *VE* and edge set E_2 containing element *S* (red text). Using the % mean survival (*S*) as a weighted edge that connects vertex *D* to vertex *VE* makes visualising the data much easier. For example it is much easier to see that Drug Y at a dose of 20 mg/kg has been tested with two different vehicles, Vehicle 1 and Vehicle 2 which resulted in the % mean survival rates of 100 and 80 % respectively. Such information is much less apparent from graph G_1.

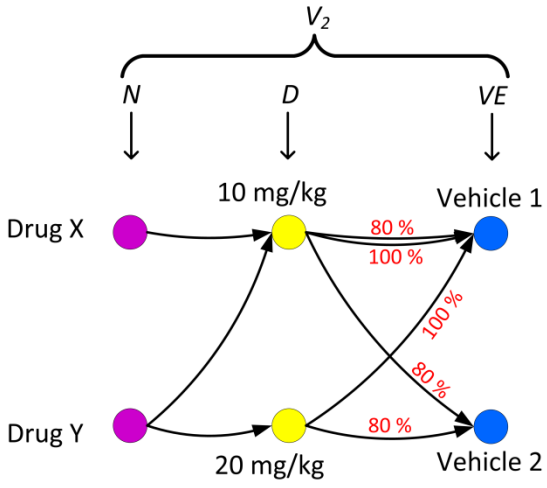


Figure 4.4. Weighted digraph G_2 showing set V_2 connected by edge set E_2 which contains element *S* (red text). V_2 contains the elements *N*, *D* and *VE* shown as purple, yellow and blue vertices respectively.

Graph G_2 does however still carry some inherent problems. For example, at the vertex 10 mg/kg, the number of out-degree edges is 3 and the number of in-degree edges is 2. This suggests that at a dose of 10 mg/kg, Drug X or Drug Y has only been tested with one of the vehicles, as confirmed by the data in Table 4.3, however from the layout of graph G_2 it is impossible to determine this and therefore this weighted graph is also not an optimal solution.

Multiple elements contained within the set *V*, result in graphs that are difficult to visualise and follow. To overcome these issues a directed bipartite graph may be the optimal solution. A bipartite graph contains two disjointed elements within its vertices set. The data in Table 4.3 would therefore require further processing to create two elements that can be represented by the vertices set *V*. Removing one or more of the elements (data instances) would result in losing critical information that enables differences in toxicity to be visualised and extracted. For example if the dose amount column was removed from Table 4.3 it would seem as if two records (rows 3 and 4) produced two different toxicity outcomes for the same drug and vehicle combination, when in actual fact the differences are related to the dose of the drug. Similar misleading information would result from the removal of the other columns carrying vital information listed in Table 4.3.

Since the removal of critical data or the inclusion of additional data, are not ideal solutions, the approach taken in this study was to combine data columns and use non-numerical values as edge weights or labels.

Since the drug (NSC) and dose amount (DOSE_AMOUNT) are important factors of toxicity which need to be kept constant when comparing two or more different vehicles it was decided that these values could be combined into a single data value which would make up one element (*N-D*) of the vertices set *V*. The other element would simply be the % mean survival value (*S*). Graph G_3 can therefore be represented as $G_3 = (V_3, E_3)$, where V_3 contains the elements *N-D* and *S* ordered as *N-D* $\rightarrow$ *S* and E_3 now contains the element *VE* ($VE \in E_3$). The NSC and DOSE_AMOUNT columns in Table 4.3 must therefore be combined before G_3 can be constructed; this processed data for graph G_3 is shown in Table 4.4 below.

With the drug (NSC) and dose (DOSE_AMOUNT) now combined as a single value (NSC-Dose) it is possible to construct a directed bipartite graph (G_3) as shown in Figure 4.5 below.

Table 4.4. Data table for graph G_3 showing drug (NSC) and drug dose (DOSE_AMOUNT) combined into a single value.

Row Number	NSC-Dose (*N-D*)	VEHICLE_CD (*VE*)	% mean survival (*S*)
1	Drug X-10	Vehicle 1	100 %
2	Drug X-10	Vehicle 2	80 %
3	Drug Y-10	Vehicle 1	80 %
4	Drug Y-20	Vehicle 1	100 %
5	Drug Y-20	Vehicle 2	80 %

From graph G_3 it becomes immediately apparent that arranging a graph in this manner allows quick visualisation of data from which interesting relationships can be easily singled out. From this graph we can see that Drug X at 10 mg/kg and Drug Y at 20 mg/kg have both been tested with Vehicles 1 and Vehicles 2 and that a difference in survival outcome related to the difference in vehicles used is easily observed from the graph (G_3). The graph also enables data that is not considered useful to be rapidly removed. For example graph G_3 shows that Drug Y at a dose of 10 mg/kg was only tested using one vehicle (Vehicle 1) and so a relationship to another vehicle cannot be made. This dataum value therefore does not present any interesting toxicity comparisons and can be entirely removed from the graph.

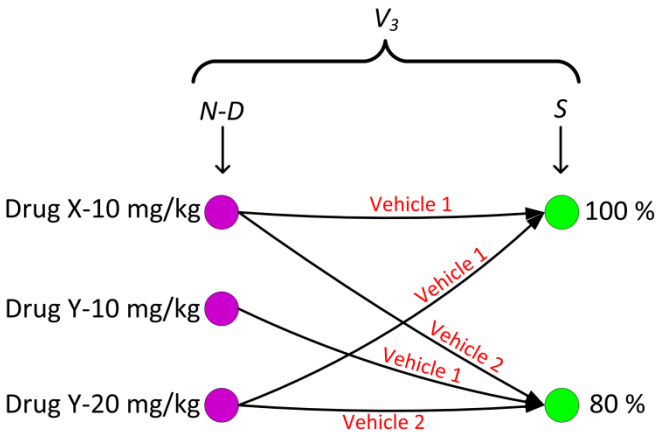


Figure 4.5. Directed bipartite graph G_3 showing set V_3 connected by edge set E_3 which contains element *VE* (red text). V_3 contains the elements *N-D* and *S* shown as purple and green vertices respectively.

To quickly filter graph G_3 so that any drug-dose (NSC-Dose) combinations that have only been tested with a single vehicle are removed from the graph, we apply an out-degree filter to the *N-D* element of set V_3. Filtering all the vertices of element *N-D* which possess an out-degree of <2 removes all instances of a drug-dose combinations (NSC-Dose) for

which only one vehicle has been used. Applying this filter produces the graph shown in Figure 4.6 below.

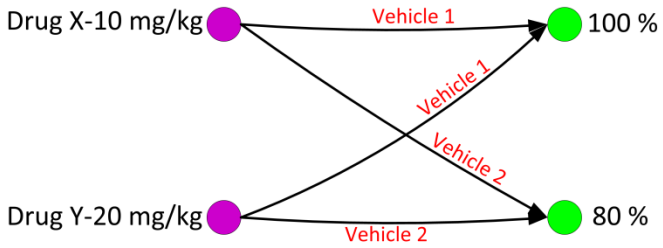


Figure 4.6. Directed bipartite graph G_3 filtered for drug-dose relationships that have more than one vehicle associated with them.

Arranging the data as a directed bipartite graph (G_3) where vertex V_3, contains the elements drug-dose (NSC-Dose (*N-D*)) and % mean survival (*S*) in the arrangement *N-D* → *S* with the edges, E_3, containing the element *VE*, provides an elegant and unique way of easily visualising the data for interesting toxicity relationships. The graph quickly helps identify the two drug-dose combinations for which two different vehicles have been used both resulting in a difference in toxicity outcome. Furthermore single data values, for which interesting relationships are not present, can be quickly removed through vertex filtering criteria.

Using graph extraction techniques it may also be possible to extract subgraphs showing only the connecting elements of a specific drug-dose, or all the connections made by a specific vehicle or pairs of vehicles. This arrangement of bipartite graph representation makes the extraction of several different data arrangements possible.

4.5 Graphs for Data Visualisation

For data visualisation, graphs can be arranged and organised in several different ways. The type of data that is chosen to be displayed in a graph can have a significant impact on the way a graph is then mined for knowledge extraction. For the NIH dataset that has been pre-processed for graph visualisation, several possibilities were explored and are discussed in the methodology section above.

So that only records of identical experimental conditions were compared the largest subset of identical experimental conditions was filtered for from the pre-processed NIH dataset. To do this the following experimental conditions were grouped; HOST_GROUP_CD,

HOST_CD, ADMIN_ROUTE, NUMBER_INJECTIONS, FIRST_INJECTION DAY, INTERVAL_UNIT, INTERVAL, TOXDAY, REPETITION and RESTART_DAYS.

This resulted in the grouping together of many experimental records containing different drugs, dose amounts, vehicles and % mean survival rates. The largest subset of identical experimental records that resulted from this grouping contained 387,662 records with 63,783 unique drug numbers (NSC number). This single subset was chosen to demonstrate the methodology of this chapter in practice. All the records within any one grouping are considered to be identical in experimental protocol and therefore can be directly compared for toxicity outcome, provided the same drug and dose were used with two or more different vehicles. The experimental conditions can thus be removed, leaving behind a dataset which contains the following columns; NSC, DOSE_AMOUNT, VEHICLE_CD, and % mean survival.

The application of the methodology provided in section 4.4 was then used to extract interesting drug-vehicle toxicity relationships from the pre-processed NIH dataset. For the illustrative purposes of this thesis some minor changes to the above methodology were made to make reporting of the data easier to demonstrate and visualise. These changes will be discussed below as they become relevant.

The NIH dataset recorded dose amounts (DOSE_AMOUNT) as a seven digit figure. A dose of 250 mg/kg would thus be recorded as 0000250. To make visualisations less cluttered the leading 0 (zeros) of the dose amounts were removed. The drug number (NSC) and dose amounts (DOSE_AMOUNT) were then combined as discussed in the methodology (section 4.4) to produce one single value separated by a hyphen (-), so for the NSC number, 740, dosed at 0000330 mg/kg, the combined value would look like; 740-330. This was done for all drug-dose combinations in the pre-processed dataset.

Since the % mean survival rates have been calculated from replicate experiments, this resulted in a range of continuous values within the dataset. To simplify matters a decision was taken to bin the % mean survival values into 10 bins as follows (Table 4.5);

Binning is performed for the purpose of ease of visualisation although the % mean survival rate continuous values could equally have been used. The dataset now contains the following three columns; NSC-Dose, VEHICLE_CD and Bin.

Table 4.5. Bin values for % mean survival rates.

% mean survival range	Bin
0 – 10 %	10 %
11 – 20 %	20 %
21 – 30 %	30 %
31 – 40 %	40 %
41 – 50 %	50 %
51 – 60 %	60 %
61 – 70 %	70 %
71 – 80 %	80 %
81 – 90 %	90 %
91- 100 %	100 %

A directed bipartite graph can then be created from this dataset. The vertices set *V* therefore contains the two disjointed elements, NSC-Dose (*N-D*) and % mean survival rates which are now represented by the bin values. For continuity purposes, the bins will be referred to as the % survival set and this element can therefore be labelled as *S*. Edges in the graph are represented by the edge set *E* which contain information regarding the vehicles (*VE*) used with a particular drug and dose amount (*N-D*).

A graph representing the 387,662 records of identical experimental conditions is not visually useful as this graph contains all possible associations of data i.e. interesting and non-interesting relationships. To demonstrate this, Figure 4.7 shows the graph $G_4 = (V_4, E_4)$, produced from a subset of the data and contains 10,000 of the 387,662 records. Elements of *N-D* are shown in purple, and elements of *S* are shown in green (which are obscured by the purple *N-D* elements). These two disjoined elements of set V_4 are connected by the edge set E_4 which contains the element *VE* (which again is obscured by the purple of the *N-D* elements).

The purpose of graph representation is to visualise drug-vehicle relationships where the use of two different vehicles, results in a change in toxicity outcome. Elements of *N-D* can therefore be filtered using an out degree filter. A filter setting of >1 would mean each element of *N-D* must have two or more elements of *VE* connected to it. This would mean two or more different vehicles are associated with that vertex (*N-D*). If this is not the case those vertices will be removed from the graph.

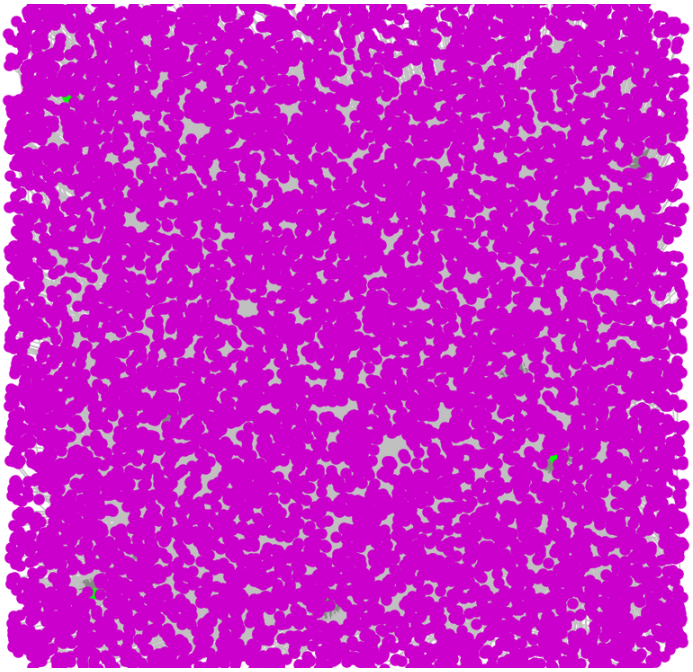


Figure 4.7. Graph, G_4, representing a subset of 10,000 records. Elements of the vertices set V_4, are shown in purple (*N-D*) and green (*S*), which are connected by edge set E_4 which contains the element *VE*.

The filter also serves a second important purpose. As mentioned earlier, this methodology requires minimal data curation. As part of this curation, missing records were not dealt with in any way and not removed from the dataset. Consider the data in Table 4.6, which contains 4 mock records, for which the vehicle and % mean survival, have not been recorded for the 4th instance of this table.

Table 4.6. Mock graph data containing missing vehicle (*VE*) and % mean survival (*S*) information.

NSC-Dose (*N-D*)	Vehicle (*VE*)	% mean survival (*S*)
740-250	Saline	80 %
740-250	HPC	100 %
14762-50	Saline	60 %
178-25	-	-

A graph produced from the information in Table 4.6 would result in the vertex, 178-25 (*N-D*), containing no out degree connections since no vehicle (*VE*) information is available to connect it to a % mean survival rate (*S*) which is also not available. This vertex represented as a graph would contain a circular reference, i.e. an edge connecting the vertex to itself. This vertex therefore contains no out degrees as it does not connect to any elements of the set *S*. The out degree filter of >1, used to filter out elements of *N-D* with only a single vehicle associated with it will also therefore filter out elements of *N-D* for which data were

missing from the dataset. This is considered a unique advantage of this type of data representation and methodology as datasets are most often curated thoroughly before any kind of processing occurs.

Once all elements of *N-D* that contain only one out degree have been removed, the graph now only contains drug-doses (*N-D*) that have two or more different vehicles tested with them. This is still however 23,793 records (each containing pairs or more of data) of the 387,662 initial records of the pre-processed dataset.

Another filter is therefore applied to the data of the graph. This time a filter for the element *VE* of edge set E_4 is used. *VE* represents the different vehicle types contained within the dataset. A subset of the edge set E_4 can therefore be extracted with all connecting vertices attached. As this chapter focuses on visualisation as well as data extraction it was decided to extract the subgraph containing the three most commonly occurring vehicles of the NIH dataset. These are; Saline, Saline with Tween-80 and hydroxypropylcellulose (HPC) (see Figure 4.2). Filtering the graph for these three vehicles results in 16,911 records of data for which each vertex of element *N-D* contained 2 or more of the three different vehicles selected, connected to it.

All elements of *N-D* therefore potentially carry interesting relationships which may show a difference in toxicity outcome with the use of different vehicles. This information can now be easily visualised for any of the drug names (NSC) in the dataset which remains. For example, consider the NSC number 180259. The elements of vertex *N-D* can be searched for all instances of this drug number using a wildcard pattern search to account for the hyphenated dose that is combined to the drug number. Searching the elements of *N-D* for 180259 produces the following subgraph, G_{4a} (Figure 4.8).

Subgraph $G_{4a} = (V_{4a}, E_{4a})$ shows the data extracted for drug number (NSC) 180259 from the parent graph G_4, for the various doses shown. The graph is a directed bipartite graph, since elements of the vertices set V_{4a}, *N-D* (purple) and *S* (green) are not connected to themselves. Elements of the edge set E_{4a}, *VE*, have been coloured red and blue which represent the vehicles saline and saline with Tween-80 respectively.

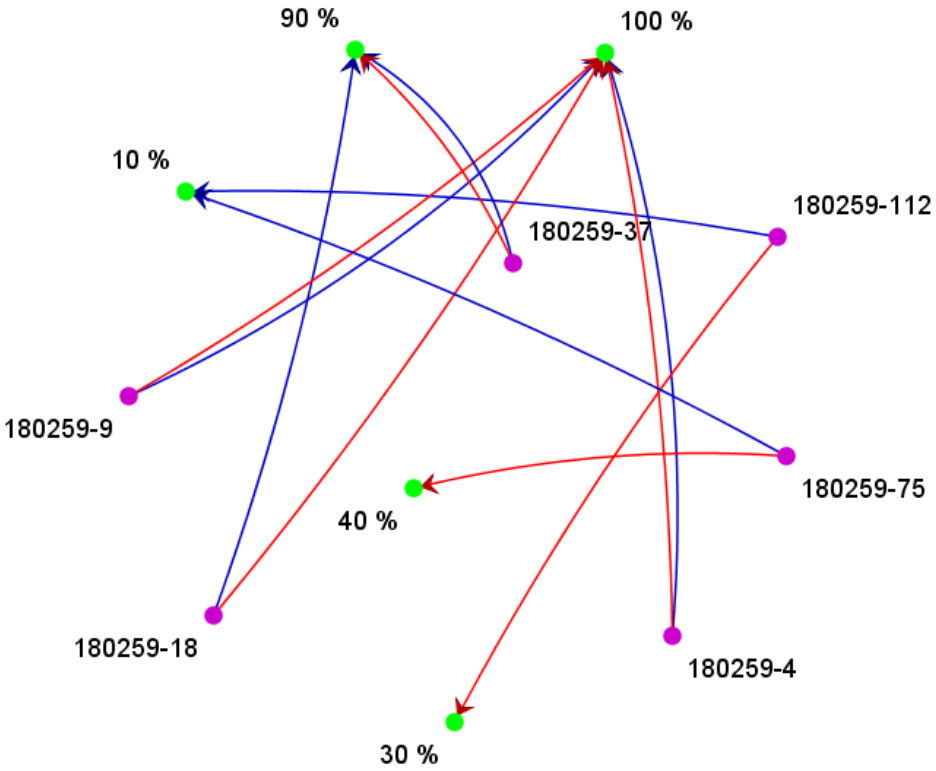


Figure 4.8. Subgraph G_{4a} for drug number (NSC) 180259, showing elements of the vertices set V_{4a}, *N-D* (purple) and *S* (green), with elements of the edge set E_{4a}, *VE*, shown in red (Saline) and blue (Saline with Tween-80).

If the subgraph G_{4a} is rearranged, so that the two elements of set V_{4a}, *N-D* and *S*, are aligned opposite each other, the visualisation and relationships become more clear (Figure 4.9).

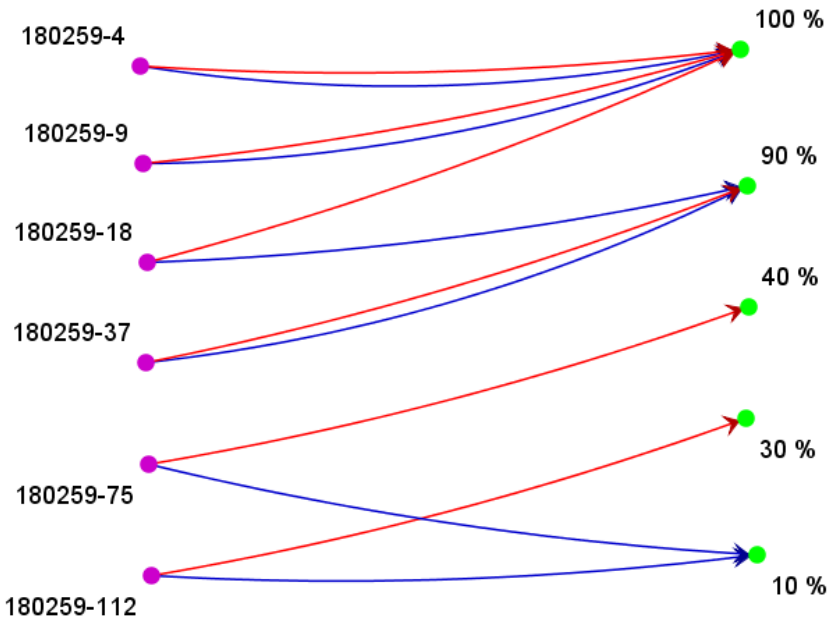


Figure 4.9. Subgraph G_{4a} with elements rearranged for ease of visualisation.

From Figure 4.9, the elements of *N-D* shown in purple all pertain to the drug number (NSC) 180259 for which each vertex of *N-D* (purple) has a dose attached to it. For example the value 180259-4 represents drug number (NSC) 180259 at a dose of 4 mg/kg. The elements of *N-D* in Figure 4.9 have intentionally been arranged with an increasing dose amount as you move from top to bottom. Similarly the survival element, *S* (which represents the binned survival values) are shown in green, they too have been arranged so the % mean survival values decrease as you move from top to bottom. The edge set E_{4a} which contains the element *VE* contains coloured edges to represent the vehicles saline (red) and saline with Tween-80 (blue). Arranging the graph in this way makes visualisation and exploring the graph much easier.

From Figure 4.9 the first thing that becomes apparent is that dose amounts for drug number (NSC) 180259, range from 4 to 112 mg/kg, roughly doubling each time except for the very highest dose. What is also apparent from this graph representation is that for every drug-dose combination the vehicles saline (red edges) and saline with Tween-80 (blue edges) have both been used to administer the drug. For the two lowest drug doses of 4 and 8 mg/kg (180259-4 and 180259-8), the survival outcome for both vehicles is identical (100 %). However from the visualisation it is easy to see that at a dose of 18 mg/kg (180259-18), the vehicle saline is more protective than the vehicle saline with Tween-80, resulting in survival outcomes of 100 and 90 % respectively. At the 37 mg/kg dose (180259-37), the survival rate observed for both vehicles is identical, at 90 %. For the two highest doses however, of 75 and 112 mg/kg (180259-75 and 180259-112), there is a difference in the survival outcome between the two different vehicles, with saline (red) always offering more toxicity protection than saline with Tween-80 (blue). The same information can also be extracted from the graph in Figure 4.8, only it is not as easy to visualise.

Similarly other subgraphs can be extracted from graph G_4. A search for the drug number (NSC) 143015, resulted in the extraction of subgraph $G_{4b} = (V_{4b}, E_{4b})$, which is shown in Figure 4.10 below.

As with subgraph G_{4a} shown in Figure 4.9, the subgraph G_{4b} shown in Figure 4.10 and extracted from the parent graph G_4, shows that drug 143015 has been tested at three different doses (purple) with the vehicles saline (red) and saline with Tween-80 (blue). For all three doses the visualisation shows that the vehicle saline always results in better toxicity protection than the vehicle saline with Tween-80.

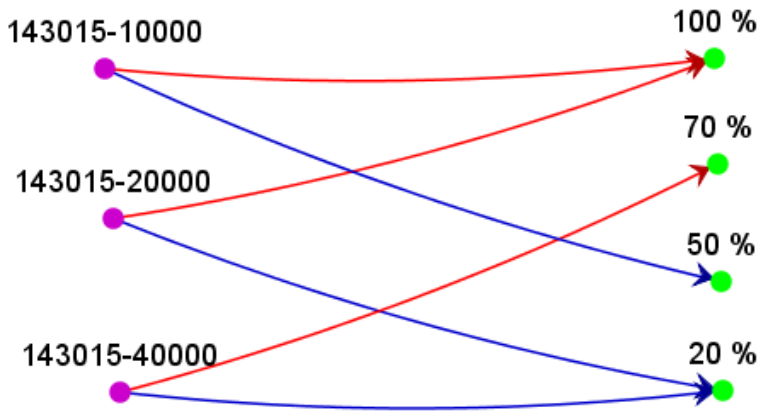


Figure 4.10. Subgraph G_{4b} for drug number (NSC) 143015, showing elements of the vertices set V_{4b}, *N-D* (purple) and *S* (green), with elements of edge set E_{4b}, *VE*, shown in red (Saline) and blue (Saline with Tween-80).

As a final example of data visualisation, consider the subgraph $G_{4c} = (V_{4c}, E_{4c})$, shown in Figure 4.11, which has been extracted from the parent graph G_4.

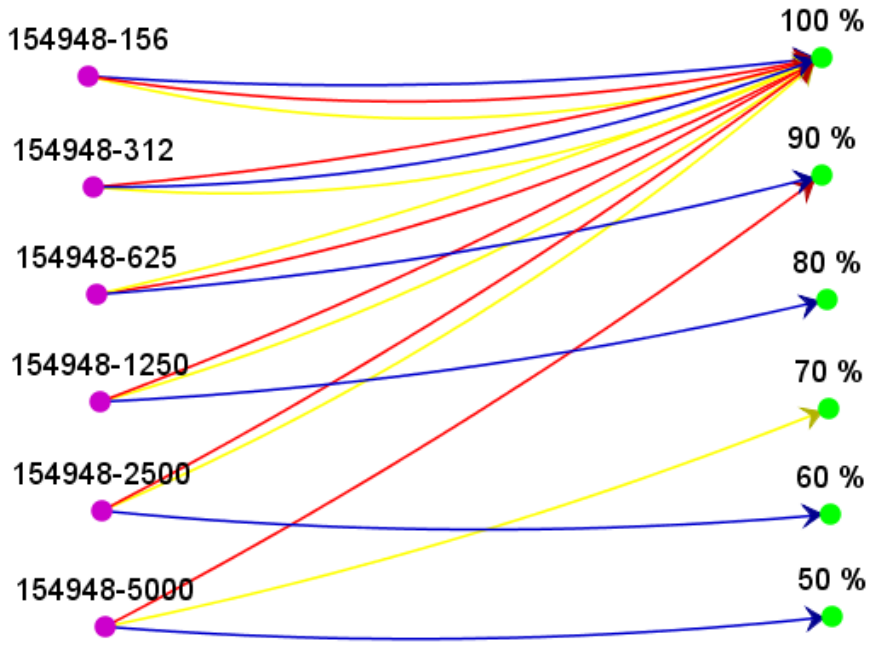


Figure 4.11. Subgraph G_{4c} for drug number (NSC) 154948, showing elements of vertices set V_{4c}, *N-D* (purple) and *S* (green), with elements of edge set E_{4c}, *VE*, shown in red (saline), blue (saline with Tween-80) and yellow (HPC).

This graph shows the drug number (NSC) 154948 having been tested at 6 different doses (156, 312, 625, 1250, 2500 and 5000 mg/kg). Furthermore, at all doses the drug was administered using 3 different vehicles, saline (red), saline with Tween-80 (blue) and HPC (yellow). From the subgraph it is quick to visualise the differences in toxicity protection that the each vehicle provides. At the two lowest doses (154948-156 and 154948-312) all three

vehicles provide the same level of toxicity protection resulting in an animal survival rate of 100 %. At a dose of 625, 1250 and 2500 mg/kg (154948-625, 154948-1250 and 154948-2500) however, the vehicles saline (red) and HPC (yellow) result in survival rates of 100 %, but the vehicle saline with Tween-80 (blue) consistently results in lower survival rates, thus offering lower toxicity protection. At the highest dose of 5000 mg/kg (154948-5000) the three vehicles all offer different levels of toxicity protection. Saline (red) resulted in the best survival rate of 90 %, followed by HPC (yellow) with a survival rate of 70 % and lastly saline with Tween-80 (blue) which resulted in a survival rate of 50 %.

4.6 Graphs for Extracting Toxicity Relationships

Visualising the data using subgraph extraction techniques allows patterns in toxicity related to the use of different vehicles for a specific drug to be easily visualised from a large dataset with minimal pre-processing. However it would be useful if the drug numbers (NSC) within the dataset could be grouped together depending on how their toxicity profiles are influenced by different vehicles. For example from the subgraphs G_{4a}, G_{4b} and G_{4c} (see section 4.5) the drugs shown in the graphs all seem to result in better toxicity profiles when used with the vehicle saline rather than the vehicle saline with Tween-80. Furthermore, for the drug number (NSC) 154948 shown in subgraph G_{4c}, HPC also seems to offer better toxicity protection than saline with Tween-80, and on one occasion at a dose of 5000 mg/kg it is also apparent from subgraph G_{4c} that the vehicle saline can be considered more protective than the vehicle HPC. Therefore from the subgraph G_{4c}, for drug number (NSC) 154948 the following relationships can be deduced; 1) saline is more protective than saline with Tween-80, 2) saline is more protective than HPC and 3) HPC is more protective than saline with Tween-80.

If sufficient numbers of these relationships for different drugs can be extracted then it would be useful to see how many data (i.e. number of compounds) exists for each group, and more specifically if any of the drugs overlap between groups. For example if a set of drugs show that the use of saline provides better toxicity protection than saline with Tween-80 then do these same drugs also show that HPC offers better toxicity protection than saline with Tween-80? This will not be possible to demonstrate for all the drugs as some drugs will not have been tested with some of the vehicles, as is the case with the data shown in subgraphs G_{4a}, G_{4b} and G_{4c}. It would nevertheless be useful if as many comparisons as possible can be made and demonstrated.

4.6.1 Mining Graphs for Data Relationships

To extract groups of drugs for which one vehicle offers more protection than another the graph G_4 was mined for such relationships. This graph contains data relating to 3 different vehicles (saline, saline with Tween-80 and HPC) as discussed in section 4.5.

These vehicles were first treated in a pairwise manner, so that the following comparisons could initially be made;

1) Saline is better than saline with Tween-80

2) Saline with Tween-80 is better than saline

3) Saline is better than HPC

4) HPC is better than saline

5) Saline with Tween-80 is better than HPC

6) HPC is better than saline with Tween-80

For every vertex of the element *N-D* of graph G_4 (see section 4.5) it is therefore possible to extract the difference in toxicity outcome for each of the vehicles connected to it. At the very least this can be done in a pairwise manner since each vertex of the element *N-D* contains at least two different vehicles connected to it.

If the outcome for two or more vehicles is equivalent, these points are discarded as they do not show any differences in toxicity. Such instances are most likely to occur at very low doses where the survival rates may be 100 % or at very high doses where survival rates may be 0 % regardless of vehicles used.

Consider the mock plot shown in Figure 4.12, for a particular drug at a range of doses used with two different vehicles (saline and saline with Tween-80). A difference in toxicity outcome is observed for the two vehicles, however at a dose of 10 and 20 mg/kg (low doses) the toxicity outcome is the same (100 %) and at the very high dose ranges, of 90 and 100 mg/kg again the toxicity differences between the two vehicles is the same (0 %). These data points do not show any interesting relationships (toxicity differences) and so are not considered. For the purpose of forming drug groupings, only the data points where there is a toxicity difference, is of interest (data points between vertical green lines) and only those data points are considered in the data extraction process.

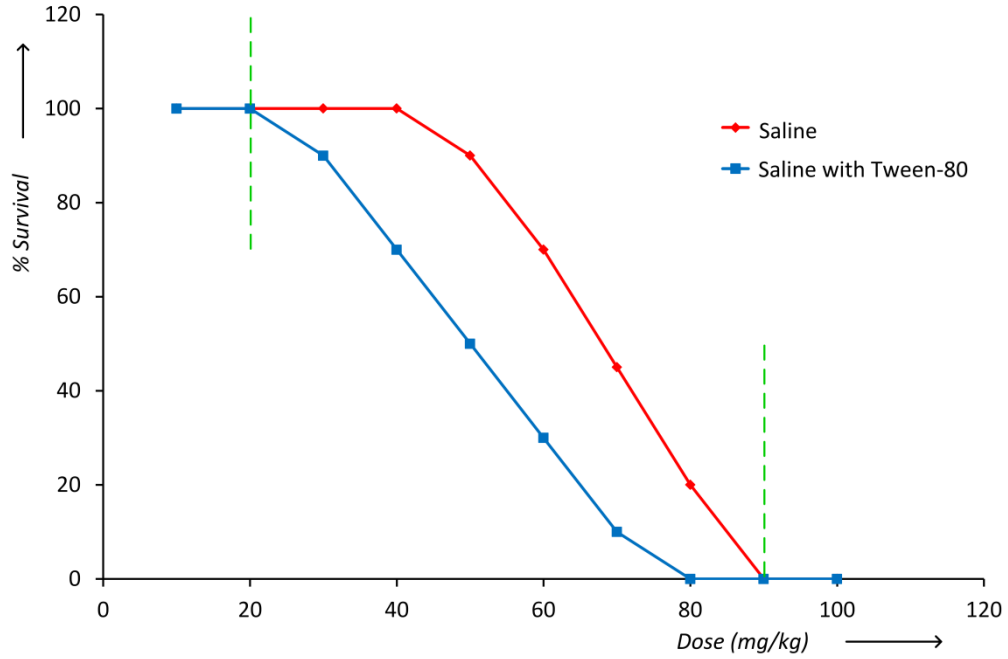


Figure 4.12. Mock plot of vehicles saline (red) and saline with Tween-80 (blue) showing differences in toxicity outcome for a given drug. Data points between the vertical green lines are of interest.

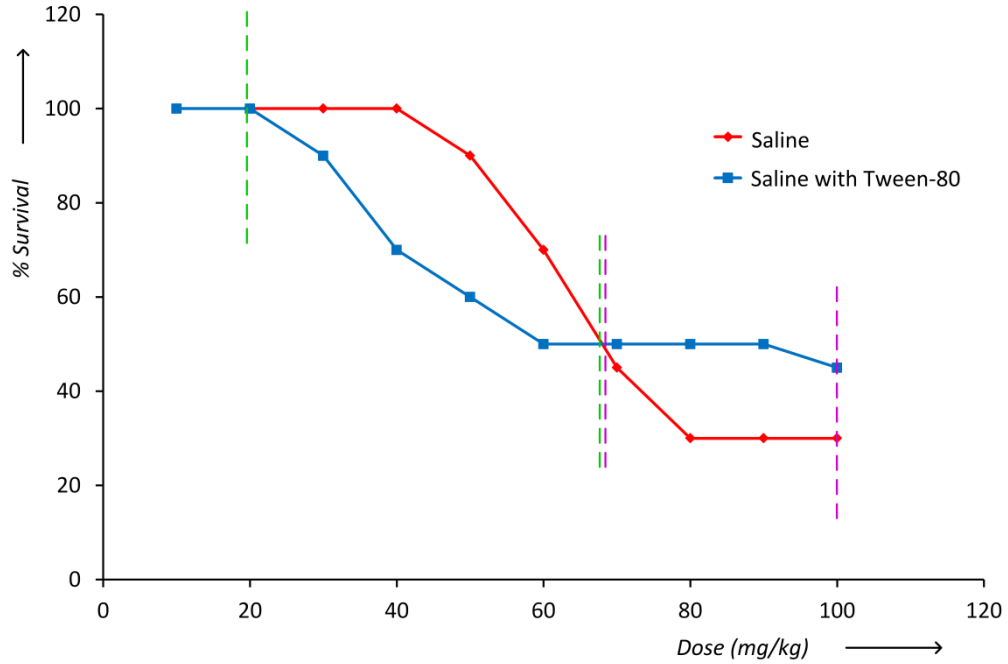


Figure 4.13. Mock plot of saline (red) and saline with Tween-80 (blue) showing conflicting differences in toxicity outcome for a given drug. Data points between the vertical green lines suggest the vehicle saline is offering better toxicity protection than saline with Tween-80, however the opposite relationship is suggested for the data points between the vertical purple lines.

Furthermore, one assumption is made for the purpose of this work; if a specific vehicle used with a particular drug shows a better toxicity outcome than another vehicle used with

the same drug, then this should be the case at all the dose ranges tested where a difference is measurable. If for a specific drug the vehicles cross over in toxicity protection over a range of doses then this drug is removed completely from any groups as it suggests conflicting toxicity outcomes.

For example consider the mock plot shown in Figure 4.13, which shows the toxicity profiles of the vehicles saline (red) and saline with Tween-80 (blue) for a particular drug. The plot suggests that the vehicle saline offers better toxicity protection between the doses of 20 and 70 mg/kg (vertical green lines), whereas between the doses of 70 and 100 mg/kg the vehicle saline with Tween-80 seems to offer the better toxicity protection (vertical green lines).

The toxicity pattern of the two vehicles is conflicting for this particular drug and as such this kind of data extracted from the graphs will be removed as an uncertainty arises as to which group this drug would belong to.

To begin with, all the drug-doses (elements of vertex *N-D* of graph G_4) showing the relationship of saline being better than saline with Tween-80 were extracted and any conflicting or equivalent data points regardless of dose for a specific drug removed. The binned % survival values were converted to integers so toxicity differences could be established numerically. This leaves behind a list of drug numbers (hyphenated doses removed) for which the vehicle saline offers better toxicity protection than the vehicle saline with Tween-80, for which no conflicting data for the drugs in this list existed in the graph. Only unique drug numbers are retained in the list i.e. only one instance of any replicate drug numbers is retained.

Next all the drug numbers which showed the opposite relationship, saline with Tween-80 is better than saline, were extracted with any conflicting or equivalent data points removed.

This resulted in the extraction of two groups, with each group containing a list of drug numbers (NSC) without any hyphenated doses attached. Both these lists should in theory contain no overlapping drug numbers.

This pairwise extraction process was performed for the other 4 relationships shown in the numbered list at the beginning of this section. The number of unique drugs extracted from the graph for each of the 6 groups is shown in Table 4.7 below.

Table 4.7. Total number of compounds per group for the various vehicle pairings of saline, saline with Tween-80 and HPC.

Row Number	Group	Number of drugs
1	Saline is better than saline with Tween-80	212
2	Saline with Tween-80 is better than saline	238
3	Saline is better than HPC	214
4	HPC is better than saline	163
5	Saline with Tween-80 is better than HPC	220
6	HPC is better than saline with Tween-80	157

From these 6 lists, which have conflicting data removed, for any two vehicles which form a pairing, say saline and saline with Tween-80, which is shown in the first two rows of Table 4.7, the list of drugs contained in the two opposing groups should be unique, i.e. the list of drugs in the lists of rows 1 & 2 (opposing groups) of Table 4.7 should be unique. Similarly the lists of row 3 & 4 and 5 & 6 should be unique.

This can be demonstrated by producing Venn diagrams for each opposing vehicle pairing (i.e. rows 1 & 2, 3 & 4 and 5 & 6) resulting in the three Venn diagrams shown below in Figure 4.14.

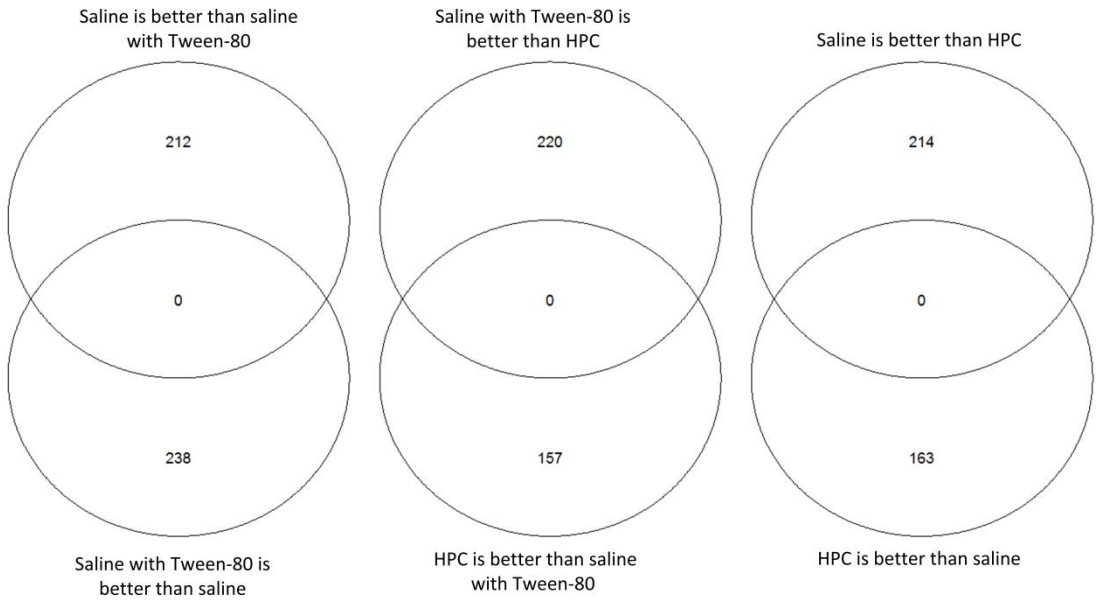


Figure 4.14. Venn diagrams showing no overlap between any two opposing vehicle pairings.

For all three Venn diagrams the overlapping segment contains zero data points, therefore the drug lists isolated from the graph G_4 are unique to each opposing group with no conflicts.

Using the Venn diagrams, it is now possible to isolate the number of interesting relationships that are shown across more than two groups to identify the number of data points that may be available for any future work considered. So for example, consider the relationship saline is better than saline with Tween-80 (row 1, Table 4.7) and saline with Tween-80 is better than HPC (row 5, Table 4.7) A Venn diagram of these two lists is shown in Figure 4.15.

The Venn diagram shown in Figure 4.15 shows the relationship of the number of drugs for which the vehicle saline is better than saline with Tween-80 and for saline with Tween-80 is better than HPC.

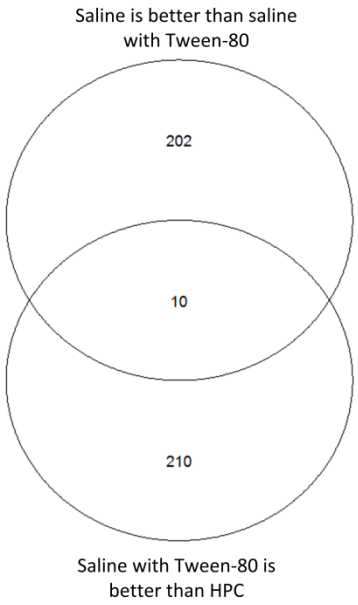


Figure 4.15. Venn diagram of saline is better than saline with Tween-80 and saline with Tween 80 is better than HPC.

From the Venn diagram we can see that there are 10 drugs that form an overlap, therefore for those 10 drugs the following two statements hold true;

1) Saline is better than saline with Tween-80 and

2) Saline with Tween-80 is better than HPC

Given these two statements, can the presumption that saline is better than HPC be made for those 10 drugs?

To find out how many of the 10 drugs this holds true for we simply add the data from the list containing drugs for saline is better than HPC (row 3, Table 4.7) to see if any overlap occurs. This is shown in the Venn diagram shown in Figure 4.16 below.

Figure 4.16. Venn diagram showing the overlap of saline is better than saline with Tween-80, saline with Tween-80 is better than HPC and saline is better than HPC.

This Venn diagram shows that of the 10 drugs from Figure 4.15 which show an overlap and for which the deduction that saline is better than HPC could questionably be made. In actual fact that deduction holds true for 8 of those 10 drugs, as shown by the central overlap of the Venn diagram in Figure 4.16. The remaining two drugs would therefore have either showed an equivalence in toxicity protection between saline and HPC and were removed from the data or were not tested with the vehicle HPC.

The Venn diagram can be further enhanced by adding the other vehicle relationships of Table 4.7, from which, other overlaps of data can be seen. Due to the constraints of the Venn diagram library (gplots) used with the R package in Knime, a maximum of 5 lists can be added to form a Venn diagram. For this reason the grouping with the lowest number of drugs in Table 4.7 (row 6) has been omitted from the Venn diagram shown in Figure 4.17.

From the Venn diagram of Figure 4.17 the overlaps of 5 different vehicle pairings which offer better toxicity protection over the other are shown. The overlaps of the Venn diagram help to quickly pick out the relationships for which there is data available. For example, from Figure 4.17 the overlap which shows the relationship saline is better than saline with Tween-80, saline with Tween-80 is better than HPC and saline is better that HPC is shown by the green circle in the Venn diagram. This is the same relationship that was shown in Figure 4.16. From the Venn diagram of Figure 4.17 it is also possible to extract the relationship saline with Tween-80 is better than HPC, HPC is better than saline and saline with Tween-80 is better than saline. This relationship is circled red in the Venn diagram and is made up of 6 different drugs.

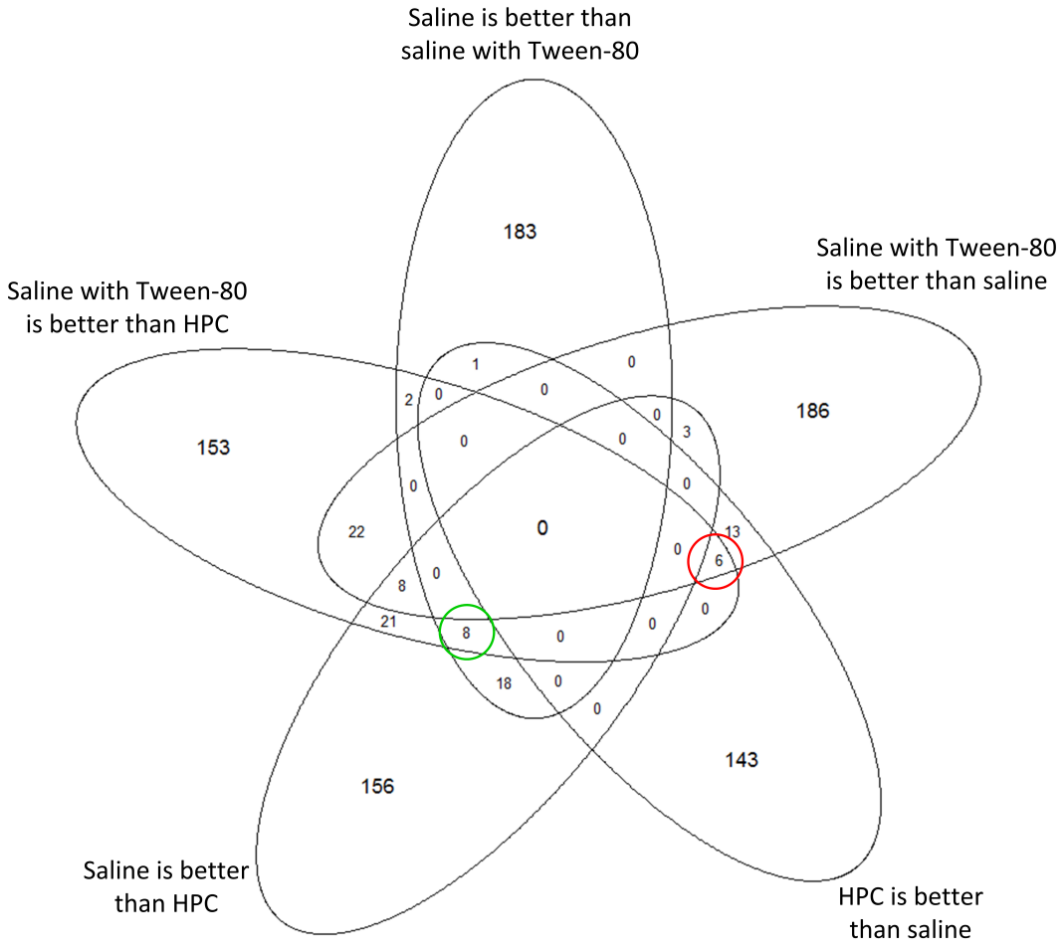


Figure 4.17. Venn diagram showing the overlay of 5 different vehicle pairings.

The numbers in the overlaps of the Venn diagram of Figure 4.17, show the number of unique drugs in the lists extracted from the graphs. Where crossovers suggest a conflict of vehicle relationships these all contain zeros (0) in the Venn diagram. If it was possible to add more lists to the diagram other relationships could be readily extracted.

The red and green circles of the Venn diagram in Figure 4.17 represent the drugs for which patterns in toxicity relating to three different vehicles is observable and for each of those drugs in the list the pattern is identical. This makes those drugs interesting to study from a drug-vehicle toxicity reduction perspective.

4.7 Summary

The work presented in this chapter has demonstrated the application of graphs for data visualisation and extraction of useful information. A methodology has been proposed and developed for the arrangement of data into graphs which allows drug-vehicle relationships to be quickly visualised with minimal data curation. Several graph representations were

explored before a directed bipartite graph was considered to be the optimal solution. Furthermore, to represent data on the vertices of the graph some data values (drug and dose) had to be combined. The methodology also uses labelled edge data since the common practice of using weighted edges were not feasible for the intended application.

The methodology has been successfully demonstrated using the NIH dataset from which interesting drug-vehicle relationships which show differences in toxicity were visualised. Extracting these drug-vehicle relationships from the graphs resulted in lists of drugs generated for which a particular vehicle offered better toxicity protection than another. These lists were then plotted using Venn diagrams from which multiple vehicle relationships could be deduced. This helped identify the number of data points (unique drugs) that would be available to study should any further analysis be desired. Due to the limitations of the Venn diagram library of the R package, larger Venn diagrams containing more lists could not be created. However substituting one list for another may still offer further insights into the data.

5

Modelling Drug-Vehicle Relationships

This chapter is concerned with the work based on building drug-vehicle relationships for the prediction of a drug vehicle for toxicity reduction. It demonstrates the use of data mining techniques to extract, curate and organise scientific information into a framework that can be utilised for knowledge gain. From these principles successful predictive models are built based on machine learning techniques (decision trees and random forests). The work in this chapter provides a unique example of how experimental scientific data are transitioned from *in vivo* experimental results to useful *in silico* information.

The main focus of this chapter is to describe the methodology that has been developed, based on the area under a curve concept to validate and build predictive models, the principles of which can be easily applied to other scientific data to produce similar models. The methodology discussed, provides a means of making pairwise comparisons between two functional variables, in this case two vehicles, which offer different levels of toxicity protection that can then be successfully modelled. The results are reported in the subsequent chapter.

Finally the concepts of “big data” and “data repurposing” are explored with respect to the dataset used and the challenges encountered are discussed.

5.1 Dataset

The dataset used for the work in this chapter is the NIH dataset discussed in chapter 4. Whilst a description of the dataset has previously been provided there are significant differences in the methodologies used and so some information may be repeated during the data curation and extraction information provided in section 5.4.

The work in this chapter aims to repurpose these data for an alternative goal. The aim is to mine the dataset to establish drug-vehicle relationships with respect to toxicity. These relationships are then explored to assess whether a vehicle has any influence on the compound's toxicity

5.2 Principles of Big Data

Although the NIH dataset used in this chapter does not meet all the requirements that are generally accepted to describe big data [330] it does meet some of the criteria and therefore the challenges that come with it. Big data is not simply a large volume of data or a substantial sized database with millions of records. There are currently several proposed definitions of big data, of which the three V's (Volume, Velocity & Variety) seem to be the most prominent [330], they are discussed below with reference to the NIH dataset:

Volume – refers to the size or volume of data being large. Although no minimum size limit exists, the data frequently come from multiple autonomous sources and are considered heterogeneous in nature. The size of the NIH dataset is certainly large with approximately 2.7 million records and these data have been compiled from different sources, given that various laboratories (known as screeners) contributed to the screening program. This brings about its own challenges given that the complexity of the data gathered from multiple sources will carry different error rates which are difficult to measure. Cross comparisons of the same experiment from different laboratories also becomes difficult, in particular if conflicting outcomes have been reported. The term veracity is sometimes associated with big data [331]; it refers to the quality or trustworthiness of the data. Techniques that handle the veracity of data help filter out "noise" to leave behind reliable trustworthy data that can be used with some element of confidence.

Velocity – is related to the speed at which data are generated, making datasets dynamic due to the absorption of additional complementary data. Although the NIH dataset is large, it is not changing dynamically or increasing in size since complementary data are not being

added. However all datasets can be added to if additional suitable data become available. The data can be compiled and structured accordingly before some form of integration is applied. The NIH dataset used here is as it was when acquired from the source [20] and no additional data were added. Most of the data have been compiled from experimentation that covers an era spanning some time from the 1950's up until the 1980's. It covers a period where computers were not common place in industry and so a substantial effort has been made retrospectively to make the data electronically available. This is something to bear in mind when considering the reliability of the data since errors in transcription can be made, which eventually translate into any models that are built.

Variety – expresses that data is gathered in different forms or via different means. This stands true to some extent for the NIH dataset, since different laboratories (screeners) contributed to the data that have been compiled into this dataset. The form and structure of the data contributed however have remained constant across all screeners. A constant structure from multiple sources is beneficial for the ease of manipulating and handling data through any data mining techniques. The ease of working with structured data saves considerable time and effort and should be considered beforehand when generating or gathering big data.

Since the three V concept for big data was first proposed [330] a range of other Vs have since been suggested by experts working in the big data analytics field. Two additions Vs of particular interest to the work in this chapter are Value and Visualisation.

Value – pertains to the knowledge that can be extracted from the data and the usefulness of this information, the ultimate goal being the discovery of something new, novel and of value. For scientific data, that useful information may lead to new discoveries or scientific principles that were previously unknown. For predictive analysis models it allows informed decisions to be made based on large volumes of scientific data that would ordinarily be unprocessable by the human mind.

Visualisation – can be considered one of the end points of big data mining, although as demonstrated in chapter 4 it can be used in parallel to quickly extract useful information also. As an endpoint, visualisation deals with the concept of being able to visualise the data that have been extracted in a format that is easily comprehensible. This can be in the form of charts and graphs or summarised data tables that present some statistical relevance of the data extracted.

5.3 Methodology

The methodology developed as part of this work was to use the area under a curve approach to compare the output of two variables, in this case two vehicles. This approach is widely applicable for comparing the functional relationship of two variables. Chapter 2 discussed some of the methods used experimentally to measure toxicity. The NIH dataset reported toxicity as the survival count of animals per dose administered. Typically groups of 6 or 10 animals were used per dosing schedule from which a % survival rate outcome could be calculated. It is therefore possible to extract survival versus dose relationships for compounds that were administered using different vehicles. The methodology of this proposed approach is discussed in this section.

For any given drug administered separately over a dosage range into animal subjects using any given vehicle, V^k, the dosage range and survival outcome can be represented by two arrays D^k and S^k, which are of equal size, *n* (*n*>1) where:

$$D^k = [d_1^k \dots d_n^k] \text{ (dosage amount array)}$$

$$S^k = [s_1^k \dots s_n^k] \text{ (survival rate array)}$$

and *n* is the number of doses given in the experiment.

The relationship of the vehicle V^k to the toxicity of the drug is therefore represented by the two arrays as:

$$V^k = [D^k, S^k]$$

Elements of the array D^k are ordered ascendingly so that $d_1^k < d_n^k$ where *n*>1. Given this, s_1^k represents the % survival rate for the smallest dose amount of drug administered using vehicle V^k, and value s_n^k would then represent the % survival rate for the largest dose amount of drug administered using vehicle V^k.

For making pairwise comparisons of two different vehicles (V^1 and V^2) given the same drug, a plot of arrays S^1 against D^1 would produce a survival versus dose curve for vehicle V^1 and similarly a plot of the arrays S^2 against D^2 would produce a survival versus dose curve for vehicle V^2. A mock survival versus dose curve for vehicle V^1 and V^2 is shown in Figure 5.1. The survival outcome difference between that offered by vehicle V^1 in comparison to that offered by vehicle V^2 can then be determined by calculating their

respective area under the curve (AUC). The differences between their AUCs provides a measurement that distinguishes between the toxicity reductions afforded by the vehicles. From Figure 5.1 it is apparent that the toxicity of the drug administered with vehicle V^2 is lower, due to a higher survival rate than for that offered by vehicle V^1. The AUC of vehicle V^2 would therefore be larger than the AUC of vehicle V^1.

Although the elements in the arrays for V^1 may not be equivalent in length or value to that of the elements in the arrays for V^2 it is important that the maximum number of data points which span the largest area possible under the survival versus dose curve are used to compare the toxicity differences between the two vehicles. Using this approach ensures comparisons across as many doses as possible thus providing a utility framework that enables predictive models to be built from as much of the available data as possible. The advantage of this approach is that it reduces the chance of experimental variability influencing the predictive models, since experimental variability is more influential over a single data point compared to data spanning a range of doses.

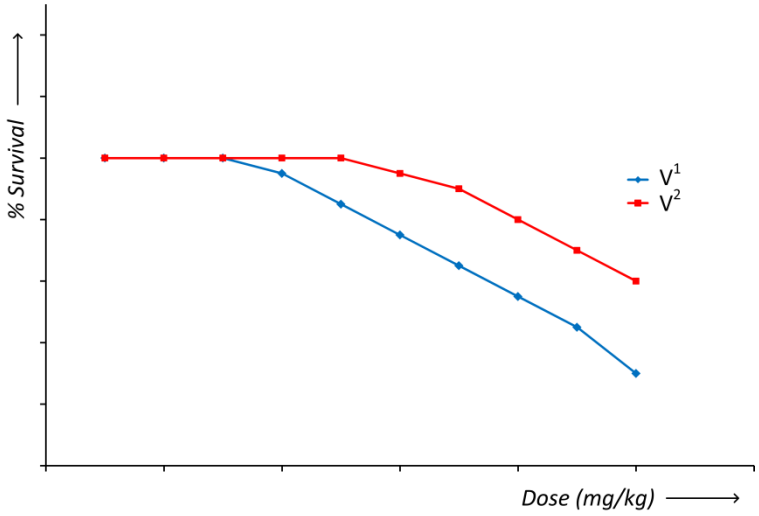


Figure 5.1. Mock survival versus dose curves for vehicles V^1 and V^2 showing the survival outcome differences as dose increases. The survival curve of vehicle V^1 at lower doses is obscured by the curve of vehicle V^2.

An additional problem is encountered when comparing two vehicles used for the same drug which is dosed at different concentrations. These comparisons require further processing steps that ensure the maximum number of data points is used; these approaches are discussed below.

5.3.1 Extrapolation

Extrapolation of data points was considered beyond the maximum and/or below the minimum dose for each vehicle array. Consider the following set of arrays for vehicles V^1 and V^2 below:

Dose amount (mg/kg) for V^1 (D^1): [15, 20, 30, 40, 50, 60]

% survival for V^1 (S^1): [100, 100, 100, 70, 30, 0]

Dose amount (mg/kg) for V^2 (D^2): [10, 20, 30, 40, 50, 60, 70]

% survival for V^2 (S^2): [100, 100, 100, 80, 50, 30, 10]

Extrapolation of the survival versus dose curve below the minimum recorded dose amount is considered if the minimum dose amount for V^1 is greater than the minimum dose amount for V^2 and the survival related to the minimum dose amount, s_1^1, is 100 %. Similarly if the maximum dose amount of V^1 is less than the maximum dose amount for V^2 and the % survival rate at the maximum dose, s_{n}^1, is 0 % then extrapolation was also a possibility. Given that the minimum dose amount for V^1 (15 mg/kg) is greater than the minimum dose for V^2 (10 mg/kg) and the % survival rate for V^1 at its minimum dose is 100 % then it is reasonable to assume that the % survival rate of V^1 at a lower, extrapolated dose of 10 mg/kg would also result in a 100 % survival, since a lower dose would naturally correspond to a lower toxicity. At the high dose end of the array a similar assumption can be made. With the maximum dose of V^1 (60mg/kg) being less than the maximum dose for V^2 (70 mg/kg) and the survival rate for V^1 at its maximum dose being 0 % it is therefore reasonable to assume that the survival rate of V^1 at a higher extrapolated dose of 70 mg/kg would also result in a 0 % survival rate. The extrapolated arrays for V^1 would then look like (extrapolated values underlined);

Extrapolated dose amount (mg/kg) for V^1 (D^1): [<u>10</u>, 15, 20, 30, 40, 50, 60, <u>70</u>]

Extrapolated % survival for V^1 (S^1): [<u>100</u>, 100, 100, 100, 70, 30, 0, <u>0</u>]

In the example above, if extrapolation is not used then the 15 mg/kg dose for V^1 and the 10 and 70 mg/kg doses for V^2 and their respective % survival rates would have to be discarded since they cannot be directly compared. Extrapolation is possible at both ends of an array, one end of the array or neither end of the array depending on the dose amounts and only if the % survival rate values for the minimum and maximum doses corresponded

to 100 and 0 % respectively. The pre- and post-extrapolation curves for V^1 and V^2 are shown in Figure 5.2 and Figure 5.3 respectively.

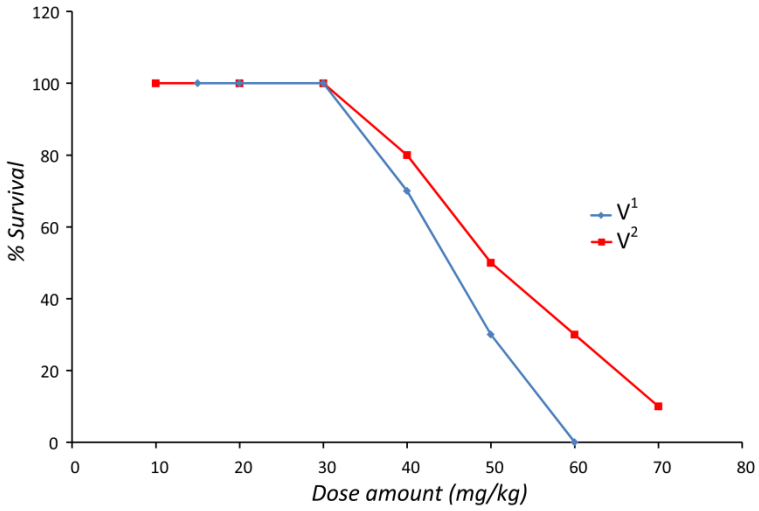


Figure 5.2. Pre-extrapolated curve for vehicles V^1 and V^2.

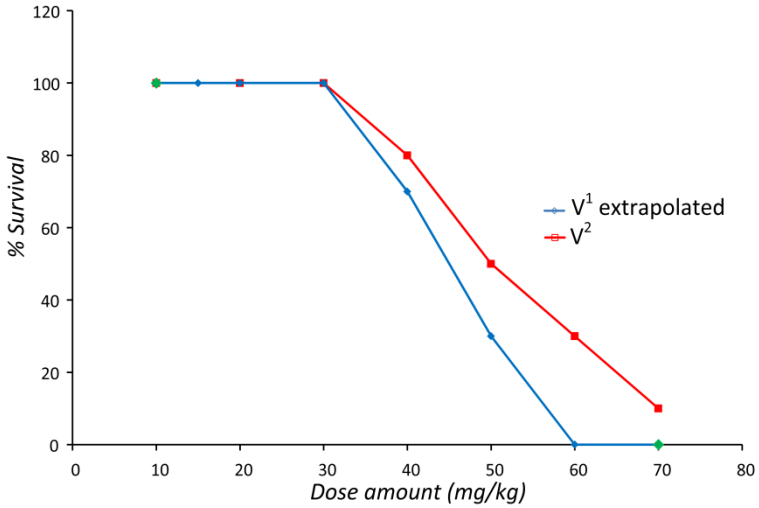


Figure 5.3. Post-extrapolated curve for vehicles V^1 and V^2. Extrapolated data points for vehicle V^1 are shown in green.

Figure 5.3 shows the extrapolated survival values for V^1 at the doses of 10 and 70 mg/kg. The AUCs of the two curves can now be compared across a wider range of doses, which provides more reassurance that the differences between the AUCs of the two vehicles (V^1 and V^2) are real and not due to experimental noise. Using the extrapolation method the percentage differences between the two AUCs will change, which plays a significant role when classifying the data (see section 5.5).

5.3.2 Interpolation

Interpolation of data is considered if extrapolation is not possible. Interpolation of data is therefore used when the minimum and maximum doses of an array of one vehicle result in a <100 or >0 % survival rate respectively. With such arrays the assumptions made during extrapolation are not possible (see section 5.3.1). Therefore to maximise the number of datum points over which AUCs can be calculated, interpolation of data is employed.

Consider the following set of arrays for vehicles V^3 and V^4 below, and shown graphically in Figure 5.4:

Dose amount (mg/kg) for V^3 (D^3): [5, 20, 30, 40, 50, 80]

% survival for V^3 (S^3): [90, 80, 70, 60, 30, 10]

Dose amount (mg/kg) for V^4 (D^4): [10, 20, 30, 40, 50, 60, 70]

% survival for V^4 (S^4): [100, 100, 80, 70, 40, 30, 25]

The arrays for vehicles V^3 and V^4 are comparable over a common dose range of 20 to 50 mg/kg, with extrapolation not a possibility, this would result in the loss of two datum points from V^3 (5 and 80 mg/kg and their respective survival rates) and the loss of three datum points from V^4 (10, 60 and 70 mg/kg and their respective survival rates). To overcome some of these issues linear interpolation between datum points can be used. Given that the minimum dose for V^4 is 10 mg/kg, linear interpolation can be used to generate a datum point at a mock dose of 10 mg/kg for V^3 using the V^3 data points of 5 and 20 mg/kg and their respective % survival rates.

To interpolate a % survival rate for V^3 (s_{i}^3) at a mock dose of 10 mg/kg (d_{i}^3), the two dose amounts of 5 (d_1^3) and 20 mg/kg (d_2^3) and their respective % survival rates (s_1^3 and s_2^3) are used. The value of s_{i}^3 along the straight line is given by:

$$s_i^3 = s_1^3 + (s_2^3 - s_1^3)\frac{d_i^3 - d_1^3}{d_2^3 - d_1^3} \quad (1)$$

where d_1^3 = 5 mg/kg, d_2^3 = 20 mg/kg, s_1^3 = 90 %, s_2^3 = 80 %, d_i^3 = 10mg/kg, then s_i^3 = 86.6 %.

Similarly if extrapolation at the high dose end of V^3 is not possible then an interpolated mock datum point for a common dose value contained within V^4 can be calculated. In the

example above, a mock dose value of 70 mg/kg can be added between the 50 and 80 mg/kg values present in the array V^3 with its respective % survival rate interpolated and added into the array S^3. The interpolated arrays for V^3 would then look like (interpolated values underlined):

Interpolated dose amount (mg/kg) for V^3 (D^3): [5, 10, 20, 30, 40, 50, 70, 80]

Interpolated % survival for V^3 (S^3): [90, 86.6, 80, 70, 60, 30, 16.6, 10]

The pre- and post-interpolated curves for V^3 and V^4 are shown in Figure 5.4 and Figure 5.5 below.

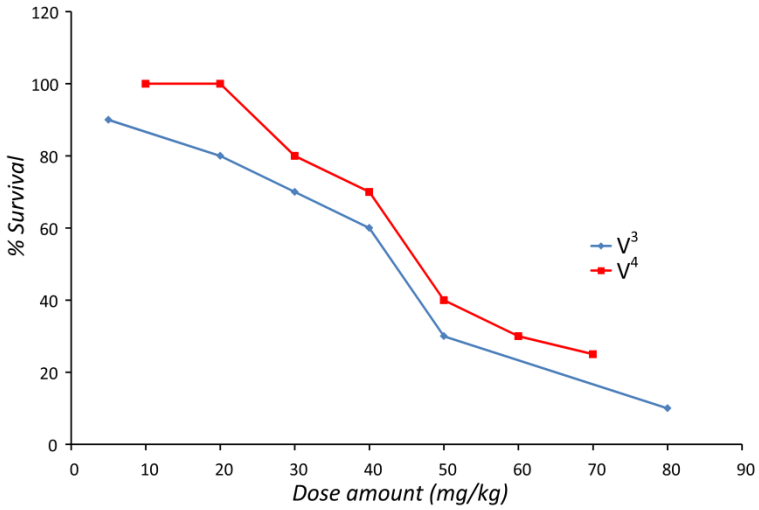


Figure 5.4. Pre-interpolated curve for vehicles V^3 and V^4.

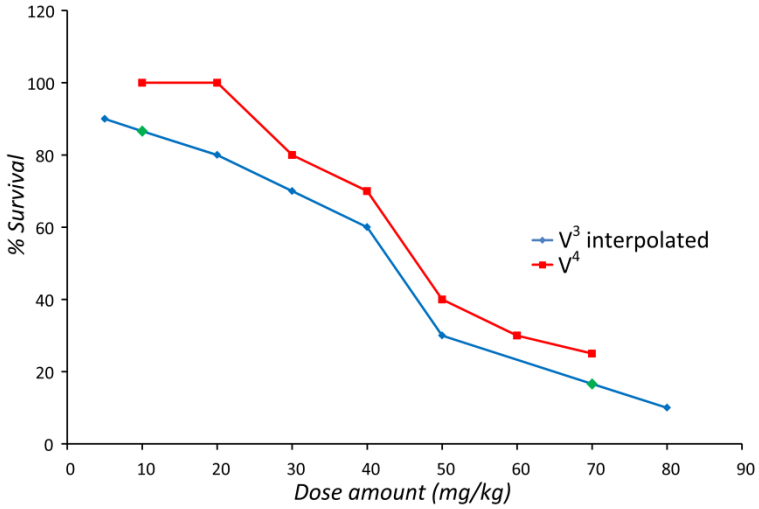


Figure 5.5. Post-interpolated curve for vehicles V^3 and V^4. Interpolated data points for vehicle V^3 are shown in green.

Using the extrapolation and interpolation methods maximises the common dose range over which vehicles can be compared and so utilises as many of the data available as possible. This technique therefore provides a method of comparing data when not all experimental parameters are consistent.

5.3.3 AUC Calculation

Once extrapolation and interpolation procedures are complete the area under the % survival rate versus dose curve (AUC) for the two vehicles can be calculated, over the largest common dose range possible. Given a dose range of $[d_1^k \ldots d_n^k]$ that corresponds to a % survival rate array of $[s_1^k \ldots s_n^k]$, the AUC is calculated as follows:

$$AUC = \sum_{x=1}^{n-1} (d_{x+1}^k - d_x^k) \left(\frac{s_x^k - s_{x+1}^k}{2}\right) \quad (2)$$

The AUCs for two different vehicles used for the administration of the same drug can then be compared to see which offers the greater toxicity protection. The greater the difference in AUCs between two different vehicles, the greater in the confidence that the difference in toxicity is real. Using this methodology enables a threshold to be set where an AUC difference above the threshold is considered significant and AUCs which differ below the threshold can be deemed equivalent. For example if the AUC difference threshold was set to 10 %, then any AUC differences of ≥10 % are deemed significant and AUCs with a difference of <10 % are considered equivalent.

The threshold is set according to how reliable the data are deemed to be. In particular the signal-to-noise ratio of the data is important to consider, if the data being processed contain a large amount of noise a larger AUC threshold may be considered. If however the data are considered very reliable then a lower AUC threshold may be used. The consequence of setting a high threshold however is fewer data available to model.

5.4 Data Curation and Extraction

Before survival versus dose plots can be extracted, the NIH dataset required pre-processing and curating which was accomplished using a Knime workflow [329]. The NIH dataset was downloaded as a 63 x 2,724,199 data table and interpreted using a document referred to as “Instruction_14,” available for download from the DTP website [20]. Initial data curation involved removing entire columns from the dataset that were irrelevant to the purpose of

this work. For example the column labelled TREATMENT_TIME, was the time at which dosing began. The information in this column was considered not to have an influence on animal survival rate and so was entirely removed from the data set. Furthermore this column only contained data for 1661 of the 2,724,199 instances and so was incomplete. Other such columns that were deleted included, but were not limited to, the date of the experiment (TEST_DT) or the laboratory (SCREENER) in which the experiments were conducted. Attempts to identify a difference in toxicity outcome by comparison of SCREENER or TEST_DATE were unsuccessful and therefore aggregation was considered acceptable. Several columns carried information about the inoculated cell lines that animals carried. Since the dataset comes from an *in vivo* screening program, compounds were tested for their ability to reduce tumours. These columns were removed from the data since they were not relevant to the drug-vehicle relationships being modelled. It is important to note at this point however that an inoculated tumour growing on an animal host can affect the survival outcome of the animal, if a tumour grew rapidly then there is a possibility it will lead to lethality. The decision to aggregate on such columns was taken since the survival outcomes of animals were recorded on day 0 (zero) or day 5 (TOXDAY) after the first dose was administered and as such the toxicity outcome would be largely influenced by the drug administered rather than the developing tumour.

When repurposing data, which was initially intended for a particular use, the challenges faced are complex and deciding what information to use and what to discard is an interesting problem. Large datasets such as the NIH dataset can rapidly shrink in size after pre-processing and curation. To minimise this impact careful consideration must be given to how the data are handled. For the NIH dataset these issues have been solved by logical reasoning. Whilst the removal of columns was a relatively straightforward process, deciding which columns could be aggregated on was more challenging. For example, a column containing information about the sex of the animals in each experiment was reported in the NIH dataset. Of the 2,724,199 records, 463,450 were recorded as male, 448,483 were recorded as female, 259 as mixed sex and 1,812,007 fields were missing. If each of these were treated separately, then extracting any relevant data sufficient enough to build models would be extremely difficult. Furthermore the 1,812,007 records that were missing would have to be discarded since there is no way of satisfactorily assigning a value to them. The impact of this would mean 66.5 % of the data is immediately lost. For this reason it was decided to aggregate the data in this column, i.e. individual experiments were not separated by sex.

The columns that were retained were considered relevant enough to affect the toxicity outcome of the experiments. These key columns are listed in Table 5.1 with their original heading name and a brief description of their value.

Table 5.1. List of columns retained after data curation of the NIH dataset for building predictive models. The columns listed were considered influential on experimental toxicity outcome i.e. survival rate.

Original column heading	Description of column
NSC	Identification number of compound tested (NSC- National Service Number)
HOST_GROUP_CD	Type of animal
HOST_CD	Species/strain of animal
ADMIN_ROUTE	Route of administration of drug
INTERVAL	Time between treatment (dosing) in terms of interval unit
INTERVAL_UNIT	Minutes (M), hours (H) or days (D)
VEHICLE_CD	Vehicle used
NUMBER_INJECTIONS	Total number of injections administered
FIRST_INJECTION_DAY	Day of first administered injection relative to day zero (inoculation)
REPETITION	Repetition of injection cycle
RESTART_DAYS	Days on which repetition occurs
DOSE_AMOUNT	Dose in mg/kg unless otherwise stated
TOXDAY	Day toxicity is measured
SURVIVOR_NUMBER_START	Animal count at start of experiment
SURVIVOR_NUMBER_TOXDAY	Animal count on toxday

Once all unnecessary columns are removed, individual records within the dataset must be curated. Records which are considered erroneous are removed from the dataset. For example, the dataset recorded the toxicity outcome of an experiment as the number of animals surviving at the end of an experiment (SURVIVOR_NUMBER_TOXDAY). If this number was greater than the number of animals at the start of an experiment (SURVIVOR_NUMBER_START) this is clearly an error in the recording of the data and without any way of confirming the information the only option was to remove the entire record from the dataset. Other individual instances that were missing from important fields were also entirely removed from the dataset. Fields that were considered important include the vehicle field (VEHICLE_CD) or the dose amount (DOSE_AMOUNT). This initial curation reduced the dataset from 2,724,199 records to 2,297,845.

The *in vivo* experiments recorded in this dataset largely used 6 or 10 animals per dosing experiment. To compare the experimental outcome of an experiment using 6 animals with those using 10, a percentage survival rate (% survival) was calculated as follows:

$$\% \text{ survival} = \frac{\text{SURVIVOR\_NUMBER\_TOXDAY}}{\text{SURVIVOR\_NUMBER\_START}} \text{ x } 100 \quad (3)$$

Since the dataset contained some records of replicate experiments which would naturally on occasions result in differences in experimental outcome it was necessary to calculate a mean of the % survival values across all replicate experiments (% mean survival). Not all the records contained replicate experiments and the difference in the number of replicate experiments between compounds was sometimes large. For example some compounds contained several hundred replicate experiments whilst others only reported a single experiment.

Dealing with multiple replicate experiments by meaning the outcome value is a relatively common statistical approach which can help minimise the error of the measured values. The variance and standard deviation of these multiple data points could also be reported although this was not done since some experiments only carried a single datum point.

Having multiple datum records from replicate experiments additionally carries the complexity of dealing with any outlier values that may have been produced. An outlier is considered an observation of the multiple measurements taken for a particular experiment that results in a value or values that deviates greatly from the other observed observations. There are many literature publications on outlier detection in datasets, with many methods proposed to deal with such cases [332-338]. Such methods are certainly useful but often specific to a given problem, methods that deal well with continuous data may not perform as well with categorical data. A simple yet effective approach here would be to remove any identified outliers from the dataset which would minimise the noise. However other literature reports argue that what may seem as apparent outliers are in fact part of real data and should not be handled in any special way since it may introduce bias into the data [339, 340]. Furthermore, Gottmann et al. (2001) [341] reported on the reproducibility of rodent data for carcinogenicity SAR studies and found a concordance of only 57 % between the data from two sources reporting on the same compounds tested. Their results suggest that the low concordance of *in vivo* rodent studies is in fact part of the real data. Considering these findings together with the NIH dataset which contained some experiments with multiple records and others with only a single record reported, it was decided for this study that no processing of outliers would be performed and all replicate experimental outcome values (regardless of suspected outliers) would be used to calculate

an average % survival. The mean % survival was used in preference to the median to give equal weight to all the data points therefore not treating any data points as outliers.

To construct % mean survival versus dose curves from the processed dataset, experiments were grouped on identical experimental conditions, that is: NSC, HOST_GROUP_CD, HOST_CD, ADMIN_ROUTE, INTERVAL_UNIT, INTERVAL, NUMBER_INJECTIONS, FIRST_INJECTION_DAY, REPETITION, RESTART_DAYS and TOXDAY. By grouping experiments on these fields, array lists could be constructed of the DOSE_AMOUNT, % mean survival and VEHICLE_CD fields. These lists could then be used to plot % mean survival versus dose curves for two different vehicles that were used with a particular compound. An example of the array lists extracted from the NIH dataset for compound NSC 63479 for two different vehicles, saline and carboxymethylcellulose (CMC) is shown below:

DOSE_AMOUNT (mg/kg): [300, 1000, 3300, 10000, 300, 100, 3330, 10000]

% mean survival: [100, 100, 33, 0, 100, 100, 100, 67]

VEHICLE_CD: [Saline. Saline, Saline, Saline, CMC, CMC, CMC, CMC]

Once these array lists were extracted from the NIH dataset they can be used for the pairwise comparison of the two vehicles. The array lists in the example above are split into their corresponding vehicle arrays as detailed below:

DOSE_AMOUNT (mg/kg) (saline): [300, 1000, 3300, 10000]

% mean survival (saline): [100, 100, 33, 0]

DOSE_AMOUNT (mg/kg) (CMC): [300, 100, 3330, 10000]

% mean survival (CMC): [100, 100, 100, 67]

The spilt arrays for the two different vehicles used with compound NSC 63479, at the various dose amounts can then be used to produce % mean survival versus dose plots as shown in Figure 5.6.

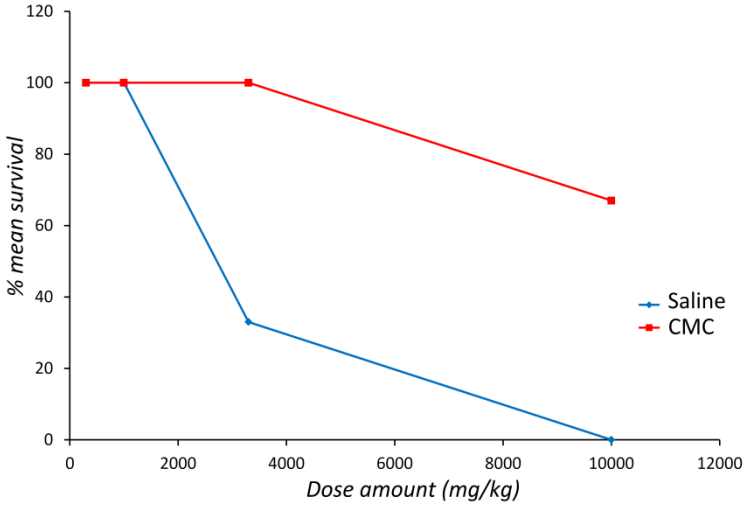


Figure 5.6. % mean survival versus dose plots for compound NSC 63479 used with the vehicles saline and carboxymethylcellulose (CMC).

The % mean survival versus dose plot for compound NSC 63479 shows that at the two lower doses of 300 and 1000 mg/kg , there is no observable difference between the vehicles, with both producing a 100 % mean survival. As the dose increases however to 3300 and 10,000 mg/kg, it becomes clear that the vehicle CMC provides better toxicity protection against compound NSC 63479 than the vehicle saline.

5.5 Classification Through Pairwise Comparisons

Using the calculated AUC values for two different vehicles used with the same compound, over a common dose range, allows a pairwise comparison to be made that distinguishes whether one vehicle is more protective against the drug's toxicity over another vehicle. Whilst the example in Figure 5.6 clearly shows a large difference between the two curves of saline and CMC, not all differences are as clear to see. For this reason a threshold of 30 % or greater was set. For a vehicle to be more protective over another, it would have to result in an AUC value of 30 % or greater than the vehicle it was being compared against. Given the significant challenges faced with the NIH dataset, which was un-curated and contained noise that was difficult to measure, a large threshold of 30 % is used to ensure the differences observed are related to the vehicles and not due to noise or experimental variation.

Using the AUC data and the 30 % threshold value, compounds in the dataset were classified as follows. For any given compound used with two different vehicles, for instance saline and CMC, if saline resulted in an AUC value that was 30 % or greater than the AUC for CMC, this compound would be assigned to the class termed saline>CMC. Likewise if CMC had

resulted in an AUC of 30 % or greater, then that compound would be assigned to the class termed CMC>saline. For any AUC comparisons that produced difference of less than 30 %, these vehicles were considered equivalent and not assigned to a class.

In this study, AUC thresholds of 40 and 60 % were also considered to see how this could influence model performance.

During compound classification, there were occasions which resulted in a compound being assigned to more than one class. This occurred when one set of experiments suggested that the vehicle saline for instance was more protective than CMC but for the same compound under a different set of experiments the opposite was found to be true. To ensure compounds are only assigned to a single class, any compounds that resulted in conflicting information were removed entirely from the data. This naturally reduced the number of compounds that could be modelled but this final curation step was critical for building successful predictive models. The classification assigned to the data (compounds) for each of the models is available in Appendix A as an electronic file.

5.6 Chemical Structure Curation

For QSAR models to be built chemical structure information is required for interpretation of the compounds within the dataset. Several chemical structure files (.sdf) were obtained from the DTP website [20]. These files were prepared by the NIH at various points over a time period in which the main *in vivo* dataset was made electronically available. These various files were compiled together to produce a large chemical structure dataset which could be related to the NIH dataset experiments. Since some files contained duplicated compound information which on occasion conflicted with other files, it was considered that the most recently dated file contained the most accurate information and so structures from this file were taken in the first instance. Any remaining unidentified structures were then obtained from the next most recently dated file and so on until as many compound structures as possible were identified. Of the entire NIH dataset that contains 227,093 unique compound numbers (NSC), all but 5,510 of the compounds were identified with a structure relating to them.

Curated structures are available in Appendix A as an electronic file.

5.7 Molecular Descriptor and Fingerprint Generation

For the compounds successfully classified, a set of descriptors were calculated that numerically encode the structural features of the compounds so they can be processed by the models. 24 molecular property descriptors based on the Indigo tool kit [342] were calculated using the Molecule Properties node in Knime (version 1.1.4.201308021053). Similarly MACCS key fingerprints were also generated using the CDK toolkit fingerprints Knime node (version 1.5.2.201409032225). This node generates 166 substructure patterns denoted as binary vectors which indicate the presence or absence of a particular substructure within the parent structure. A description of each of the MACCS keys fingerprints can be found here [281].

5.8 Descriptor Curation

The descriptor pool generated is vast and provides a large diversity of chemical properties for the compounds being modelled. Since the data processing steps outlined above will naturally reduce the number of compounds which can be compared, the models built typically contained compound numbers in the low hundreds, this is discussed in more detail in the subsequent chapter. Due to the low number of compounds being modelled it is necessary to objectively screen the descriptor pool so that successful models can be built on a few descriptors allowing the models to generalise during the learning stage.

Firstly descriptors that resulted in a constant value for all compounds were removed from the descriptor pool. Models are unable to distinguish between classes if a constant value for a descriptor is used and would hinder effective model learning. All but one of any two or more descriptors that perfectly correlated were also removed from the descriptor pool. Only one descriptor needs to be retained in such cases since the models do not gain any more information from retaining more. A mutual descriptor correlation coefficient, R^2, value of 1 was used to determine which descriptors could be removed, although other studies report using values of 0.9 or above [343-345].

Lastly the binary class of the compounds were correlated with the remaining descriptor pool. The highest correlating descriptors at a 10:1 ratio of data points to descriptors were selected. For the decision tree models, $n/10$ descriptors were selected when n data points existed. Whilst for the random forest models, $n/10$ descriptors were selected from a pool of $2n/10$ when n data points existed. For the models built using interpolation only at the 60 % AUC threshold (see Table 6.3), this translates to 7 descriptors ($n/10$) given 70 data points

(*n*) for the decision trees and 7 descriptors (*n*/10) from a pool of 14 (2*n*/10) given 70 data points (*n*) for the random forest models.

5.9 Decision Tree and Random Forest Models

Two machine learning techniques were employed to build classification prediction models: the C.45 decision tree (DT) and random forest (RF).

Briefly, the C.45 decision tree is an algorithm developed by Quinlan [346]. Attributes (in this case descriptors) are chosen that most effectively split the tree into their respective classes. The attribute that exhibits the highest normalised information gain becomes the chosen split criterion. To split using the binary MACCS fingerprints, this will simply be a presence (1) or absence (0) criterion. For the molecular properties descriptors this will be a cut off value in the fashion of a "greater/less than or equal to" type. C.45 then repeats on the smaller sub-lists. The Gini index quality measure was used for the decision tree models with no pruning.

Random forest (RF) models are an ensemble learning method developed by Breiman [296]. RFs construct a multitude of underlying decision trees and generate an output prediction based on the mode of the classes of the individual trees. Put simply, RFs work on the principle that a "strong learner" (random forest) can be developed by combining a group of "weak learners" (decision trees). A subset of the training set (local set) is chosen, from which individual trees are built, the remaining samples are then used to estimate the goodness of fit. Trees are grown by splitting the local set at each node according to the value of a random variable (descriptor) sampled independently from a subset of variables. In this study the number of trees generated in the random forest was set to 100. No significant improvements in predictions were observed with greater numbers.

The classified dataset which contained the binary class, e.g. Saline>CMC or CMC>Saline was processed for vehicle prediction using a 10-fold cross validation training method. This approach utilises 90 % of the data from each class to train the models, whilst providing a prediction for the remaining 10 %. The process is then iterated over for 10 repetitions to ensure predictions are made for all the compounds in the binary classified dataset.

Using the NIH dataset several binary classification models were built which compared in a pairwise fashion the difference in toxicity resulting from the use of two different vehicles.

The results of the RF and DT models produced from this dataset are presented and discussed in chapter 6.

5.10 Methodology Flowchart

The AUC methodology providing pairwise comparisons described in this chapter is presented as a flowchart in Figure 5.7.

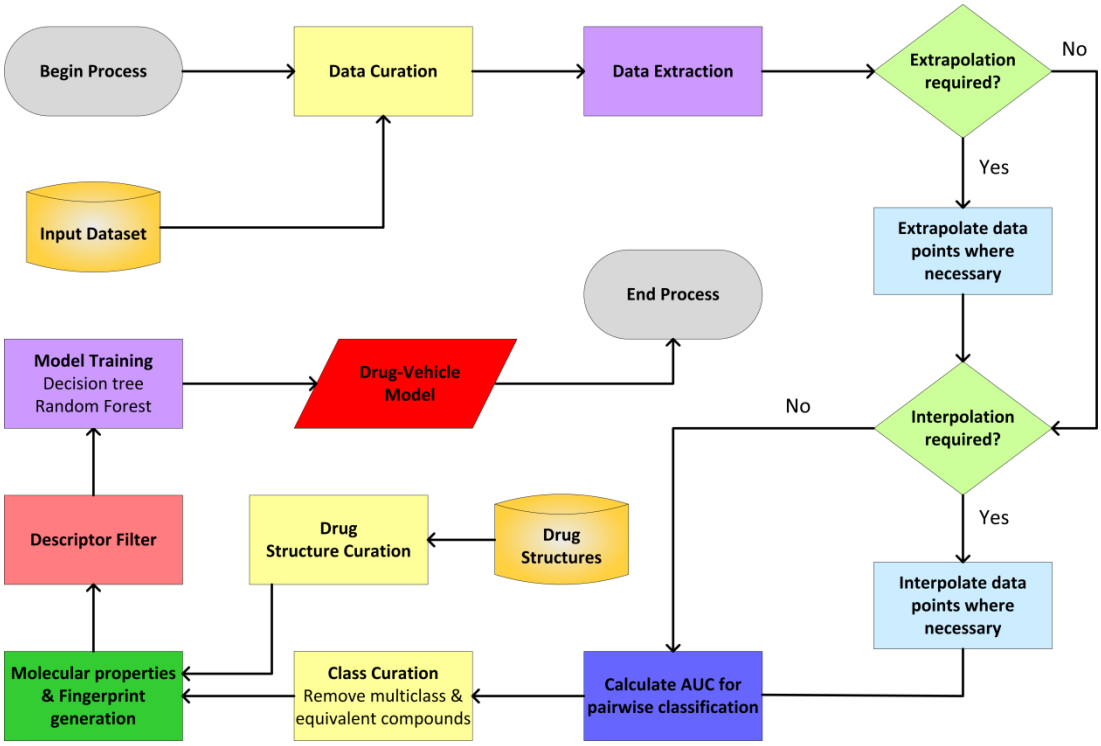


Figure 5.7. Flowchart of the AUC pairwise comparison methodology for binary classification and vehicle prediction.

The flowchart provides a useful overview of the methodology discussed which can in theory be applied to other scientific data of a similar nature. The inputs of the method are the Input Dataset and the Drug Structures (orange) which are discussed in sections 5.1 and 5.6 respectively. Following this, the Data Curation, Class Curation and Drug Structure Curation steps (yellow) which describe the main curation approaches undertaken are presented in sections 5.4, 5.5 and 5.6 respectively. The Data Extraction and Model Training processes (purple) are discussed in sections 5.4 and 5.9 respectively. The AUC Methodology approach which utilises Extrapolation and Interpolation (pastel blue) to calculate the pairwise AUC differences (blue) for classification is discussed in section 5.3. Section 5.7 describes the Molecular properties and Fingerprints that were generates, with section 5.8

explaining how descriptors were objectively screened. Finally the flowchart ends with the Drug-Vehicle models that were produced which are reported in the subsequent chapter.

5.11 Conclusion

The methodology provided in this chapter has been successfully used to build predictive models for the selection of drug vehicles that reduce toxicity. The method proposes an AUC methodology using interpolation and extrapolation techniques, which can in theory be applied to any scientific data where the relationship of two functional variables can be investigated over a range of datum points. As part of the work in this thesis this proposed methodology has been successfully used with the NIH dataset to successfully extract data and build pairwise classification models.

6

Drug-Vehicle Model Results

This chapter reports the results of the decision tree and random forest binary classification models developed with the methodology proposed in chapter 5. The results of each pairwise classification model are reported with data extracted from the decision tree models providing some interpretation of the results obtained. For each model an account is provided of the size of the data relating to the number of compounds that were available before and after data curation. The tabulated results for each model processed show exactly how many compounds were modelled and the number of compounds within each class (class split). The class accuracy and an overall balanced accuracy are provided as statistical measures of model performance.

6.1 Model Overview

Initially decision tree and random forest models were built using a 30 % AUC threshold, which dictates that the AUC difference between the survival versus dose curves of vehicle 1 (V^1) and vehicle 2 (V^2) must have a percentage difference of 30 % or more for one vehicle to be considered more protective than the other. If V^1 resulted in an AUC of 30 % or more greater than V^2 for a particular compound (NSC number), that compound was assigned the class V^1>V^2 (read, vehicle 1 is better than vehicle 2) if the reverse was true than that compound would be classed as V^2>V^1 (read, vehicle 2 is better than vehicle 1) (see chapter 5, section 5.5). Where the difference was less than 30 % the vehicles were considered

equivalent and those particular data points removed from the dataset. Furthermore, any compounds which presented conflicting trends (i.e. $V^1>V^2$ and $V^2>V^1$) arising from different experimental conditions were also removed from the dataset. Two other thresholds of 40 and 60 % were also considered to assess the impact on the accuracies of the models. The extrapolation and interpolation methods described in chapter 5, section 5.3 were also assessed for their impact on model performance. Thus at each of the three different threshold levels, 30, 40 and 60 %, models were interpolated only with no extrapolation (Interpolation only), interpolated with only high dose extrapolation (since toxicity is correlated with higher doses) (Extrapolation high), and interpolated with extrapolation at the high and low doses (Extrapolation high-low). The classification assigned to the data (compounds) for each of the models is available in Appendix A as an electronic file.

6.2 Vehicles Available to Model

Six vehicles were identified as the most frequently used within the NIH dataset (see chapter 4, Figure 4.2). These vehicles are: Saline, Saline with Tween-80, Hydroxypropylcellulose (HPC), Carboxymethylcellulose (CMC), Methylcellulose (MC) and Water.

Table 6.1. Vehicle pairs with sufficient data available to model.

Vehicle Pair	Sufficient data available to model
Saline & CMC	Yes
Saline & HPC	Yes
Saline & Saline with Tween-80	Yes
Saline & MC	Yes
Saline & Water	Yes
CMC & HPC	Yes
CMC & Saline with Tween-80	No
CMC & MC	Yes
CMC & Water	No
HPC & Saline with Tween-80	Yes
HPC & MC	No
HPC & Water	No
Saline with Tween-80 & MC	No
Saline with Tween-80 & Water	No
MC & Water	No

Combining these vehicles pairwise produces the 15 pairs shown in Table 6.1, for which sufficient data were available for 8 of these pairs to build models; sufficient data are considered as ≥10 compounds in each class.

The vehicle pairs for which sufficient data were available to model are described in the subsequent sections of this chapter. It is important to note that using the various thresholds and extrapolation/interpolation methods, potentially, 18 models are produced for each of the vehicle pairings. Some of these models are produced from insufficient datum points, which is considered as any class that has < 10 compounds within it. These models are considered unreliable and are annotated in the data tables presented for each of the vehicle pairings.

6.2.1 Other Vehicles Considered for Modelling

Initially six of the most occurring vehicles within the NIH dataset, for which sufficient data were available, were considered for pairwise comparisons to build reliable models. Other vehicles contained within the NIH dataset which may not occur as frequently but for which sufficient data for pairwise comparisons were available were also considered. It was possible to construct four further models which involved the following pairwise comparison; Distilled water + alcohol & Saline, Distilled water + alcohol & MC, Distilled water + alcohol & CMC and Acetone & Saline.

These models were generated using the same parameters discussed briefly in section 6.1 for which a more detailed explanation can be found in chapter 5.

6.3 Model Data

As the data from the NIH dataset are curated, the total amount of data will shrink as erroneous data points are removed. This will subsequently have an effect on the number of data points which can be extracted to build each of the models discussed in section 6.2. It is therefore useful to understand this impact for each of the models built. Table 6.2 provides some informative details pertaining to the vehicle occurrence, the number of records, and number of compounds for each of the models reported in this chapter.

Table 6.2 presents some key values for each of the models built. As an example, consider the first pairing in the table which describes the data for the models of the vehicles saline and CMC. Of the entire NIH dataset where vehicles were reported as being used, saline occurs in 28 % of these records whereas CMC occurs in 8 % (See vehicle occurrence, Table

6.2). After data curation, as discussed in chapter 5, section 5.4, the vehicles saline and CMC corresponded to a total of 747,305 individual experimental records (Total individual records, Table 6.2), which is accounted for by 75,504 unique NSC numbers (see No. of unique NSC numbers (for individual records), Table 6.2). For binary pairwise classification models however, both saline and CMC are required to have been used with the same drug (NSC number) under the same experimental conditions that were considered to be influential on toxicity outcome (see chapter 5, section 5.4). Aggregating the curated data to achieve this resulted in 2,221 records (see Total paired records, Table 6.2) corresponding to 1,454 unique NSC numbers (see No. of unique NSC numbers (for paired records), Table 6.2). For survival versus dose curves to be generated a minimum number of two doses (datum points) were necessary (i.e. n>1, see chapter 5 section 5.3). Furthermore any NSC numbers with sufficient datum points, but no chemical structure available, were also removed from the dataset. The remaining data could then be modelled using the pairwise classification method developed in chapter 5, section 5.5. The exact number of compounds from which each of the models is built is reported in the Model results section (section 6.4).

Table 6.2. Number of records and unique NSC numbers per model.

Model	Vehicle occurrence (%)	Total individual records	No. of unique NSC numbers (for individual records)	Total paired records	No. of unique NSC numbers (for paired records)
Saline & CMC	28 & 8	747,305	75,504	2,221	1,454
Saline & Saline with Tween-80	28 & 15	896,078	94,096	4,057	3,092
Saline & HPC	28 & 13	856,496	85,767	2,475	1,743
Saline & Water	28 & 5	693,557	58,777	1,535	589
Saline & MC	28 & 7	729,170	85,867	1,229	1,039
CMC & HPC	8 & 13	435,763	56,903	421	347
CMC & MC	8 & 7	308,437	55,663	1,731	1,512
HPC & Saline with Tween-80	13 & 15	584,536	77,098	2,226	1,646
Distilled water + alcohol & Saline	3 & 28	638,257	65,033	554	454
Distilled water + alcohol & MC	3 & 7	199,389	42,489	2,100	1,864
Distilled water + alcohol & CMC	3 & 8	217,524	36,099	553	399
Acetone & Saline	0.36 & 28	591,495	55,803	606	526

6.4 Model Results

This section discusses the results produced from each of the pairwise comparisons of the models detailed in Table 6.2. For each vehicle pairing a discussion of the results is provided in relation to the different models built and the outcome of their predictivity. The most accurate model for each pairing is described and the findings of the DT models showing which MACCS keys fingerprints (described here [281]) and molecular property descriptors were used are reported. The model resulting in the most accurate prediction for each of the pairings has been highlighted in yellow in the results tables and any results shaded in grey are the result of unreliable models due to insufficient datum points.

For each of the models built, the number of compounds available and the number of descriptors used for both the DT and RF models are reported. Descriptors were selected by the correlation against class and the number of descriptors used in each model is determined as $1/10^{th}$ the number of datum points available, for RF the number of descriptors in the pool was twice the number to be used in the model. The number of trees in each RF model was 100.

Each of the results reported describing the models also records the data bias in the dataset modelled, i.e. the excess of compounds in one class over the other. In the majority of cases the data bias was <2:1 and therefore no correction of the bias was deemed necessary in the training set/test set selections.

6.4.1 Saline & CMC Model Results

Table 6.3 shows the results of the DT and RF models for the vehicles saline and CMC at the three different thresholds with the different interpolation/extrapolation techniques applied. The output of the models show that the RFs tend to outperform the DTs, which is expected considering RFs are an ensemble method of "weak learners" (see chapter 5, section 5.9) [347, 348]. As the % AUC threshold is increased, the accuracies of both the DTs and RFs improve. This can be seen graphically in Figure 6.1, where for each model the 40 or 60 % thresholds are the most accurate.

It is important to note that as the % AUC threshold is increased more compounds fall into the equivalent group and are removed from the dataset, leaving behind fewer compounds for training and predicting. For this pairing extrapolation/interpolation changes do not

seem to affect the trends in model performance, although the number of compounds included in the model is consistently greater when extrapolating at the high end only.

Table 6.3. Decision tree and random forest model outputs for the vehicles Saline and CMC. The most accurate model is highlighted in yellow.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Saline>CMC/ CMC>Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Saline>CMC (%)	Correct CMC>Saline (%)	Balanced Accuracy (%)
DT	Interpolation only	30	148	97/51	1.9	15	60.8	70.1	43.1	56.6
DT	Extrapolation high	30	170	105/65	1.6	17	58.2	63.8	49.2	56.5
DT	Extrapolation high-low	30	114	67/47	1.4	11	64.9	67.2	61.7	64.4
RF	Interpolation only	30	148	97/51	1.9	15(30)	67.6	77.3	49.0	63.2
RF	Extrapolation high	30	170	105/65	1.6	17(34)	62.9	74.3	44.6	59.5
RF	Extrapolation high-low	30	114	67/47	1.4	11(23)	65.8	73.1	55.3	64.2
DT	Interpolation only	40	112	70/42	1.7	11	64.3	77.1	42.9	60.0
DT	Extrapolation high	40	136	80/56	1.4	14	68.4	71.3	64.3	67.8
DT	Extrapolation high-low	40	89	49/40	1.2	9	73.0	73.5	72.5	73.0
RF	Interpolation only	40	112	70/42	1.7	11(22)	70.5	77.1	59.5	68.3
RF	Extrapolation high	40	136	80/56	1.4	14(27)	59.6	66.3	50.0	58.1
RF	Extrapolation high-low	40	89	49/40	1.2	9(18)	67.4	77.6	55.0	66.3
DT	Interpolation only	60	70	45/25	1.8	7	71.4	84.4	48.0	66.2
DT	Extrapolation high	60	99	61/38	1.6	10	73.7	82.0	60.5	71.2
DT	Extrapolation high-low	60	52	31/21	1.5	5	65.4	67.7	61.9	64.8
RF	Interpolation only	60	70	45/25	1.8	7(14)	68.6	80.0	48.0	64.0
RF	Extrapolation high	60	99	61/38	1.6	10(20)	74.7	78.7	68.4	73.6
RF	Extrapolation high-low	60	52	31/21	1.5	5(10)	80.8	80.6	81.0	80.8

The most accurate prediction, resulted from the RF model using a 60 % threshold with data extrapolated at both high and low dose ranges (Extrapolation high-low), which produces a balanced accuracy of 80.8 %. This model also produced individual prediction accuracies of 80.6 and 81.0 % for saline>CMC and CMC>saline respectively, showing that the individual predictive outcomes for the two vehicles is evenly distributed. 52 compounds were available for modelling at this threshold.

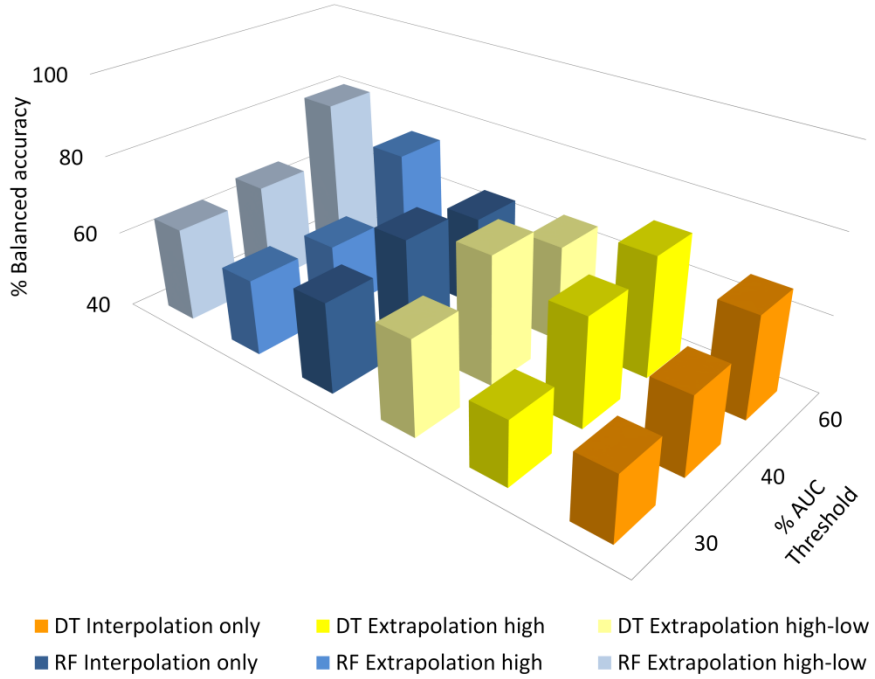


Figure 6.1. Decision tree and random forest balanced accuracies for the vehicles Saline and CMC.

Whilst the RF models tend to outperform the DT models, the DTs do provide some insightful scientific information pertaining to the descriptors that the branches split on. From the nine different decision trees built (each iterated 10 times), six MACCS fingerprints and one molecular property descriptor were most frequently used to split branches on. These fingerprints include (MACCS position-description); 120-HETEROCYCLIC ATOM >1, 83-QAAAA@1, 53-QHAAAQH 109-ACH2O, 115-CH3ACH2A and 137-HETEROCYCLE. The Number of heteroatoms in aliphatic rings was the most frequently used molecular property descriptor.

6.4.2 Saline & Saline with Tween-80 Model Results

The information presented in **Table 6.4** shows the results of the DT and RF models for the vehicles saline and saline with Tween-80 at the various thresholds and different interpolation/extrapolation techniques used. As with the saline & CMC models discussed in section 6.4.1 the RF models tend to outperform the DTs with the exception of the 40 % threshold, where the DTs consistently perform better.

Table 6.4. Decision tree and random forest model outputs for the vehicles Saline and Saline with Tween-80. The most accurate model is highlighted in yellow.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Saline>Saline with Tween-80/Saline with Tween-80>Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Saline>Saline with Tween-80 (%)	Correct Saline with Tween-80>Saline (%)	Balanced Accuracy (%)
DT	Interpolation only	30	153	86/67	1.3	15	58.2	74.4	37.3	55.9
DT	Extrapolation high	30	204	107/97	1.1	20	52.5	47.7	57.7	52.7
DT	Extrapolation high-low	30	125	68/57	1.2	13	52.8	60.3	43.9	52.1
RF	Interpolation only	30	153	86/67	1.3	15(31)	61.4	67.4	53.7	60.6
RF	Extrapolation high	30	204	107/97	1.1	20(41)	52.9	47.7	58.8	53.2
RF	Extrapolation high-low	30	125	68/57	1.2	13(25)	53.6	60.3	45.6	53.0
DT	Interpolation only	40	109	61/48	1.3	11	67.0	72.1	60.4	66.3
DT	Extrapolation high	40	158	79/79	1.0	16	65.8	64.6	67.1	65.8
DT	Extrapolation high-low	40	84	46/38	1.2	8	65.5	58.7	73.7	66.2
RF	Interpolation only	40	109	61/48	1.3	11(22)	58.7	60.7	56.3	58.5
RF	Extrapolation high	40	158	79/79	1.0	16(32)	55.7	54.4	57.0	55.7
RF	Extrapolation high-low	40	84	46/38	1.2	8(17)	65.5	71.7	57.9	64.8
DT	Interpolation only	60	52	29/23	1.3	5	63.5	75.9	47.8	61.8
DT	Extrapolation high	60	95	47/48	1.0	10	55.8	55.3	56.3	55.8
DT	Extrapolation high-low	60	35	22/13	1.7	4	62.9	77.3	38.5	57.9
RF	Interpolation only	60	52	29/23	1.3	5(10)	71.2	69.0	73.9	71.4
RF	Extrapolation high	60	95	47/48	1.0	10(19)	57.9	61.7	54.2	57.9
RF	Extrapolation high-low	60	35	22/13	1.7	4(7)	71.4	72.7	69.2	71.0

For the DTs, better balanced accuracies are observed at the 40 % threshold than at the 30 or 60 % thresholds (Figure 6.2). Interestingly the best accuracies for all model permutations are observed using interpolation only, with the exception of the RF model at a threshold of 40 % which produces the greatest accuracy using the extrapolation at the high and low dosage range (Extrapolation high-low). All models that employ extrapolation at the high dosage ranges only (Extrapolation high) produce the greatest number of total compounds that can be classified and modelled although this particular permutation may not result in the greatest accuracies achieved.

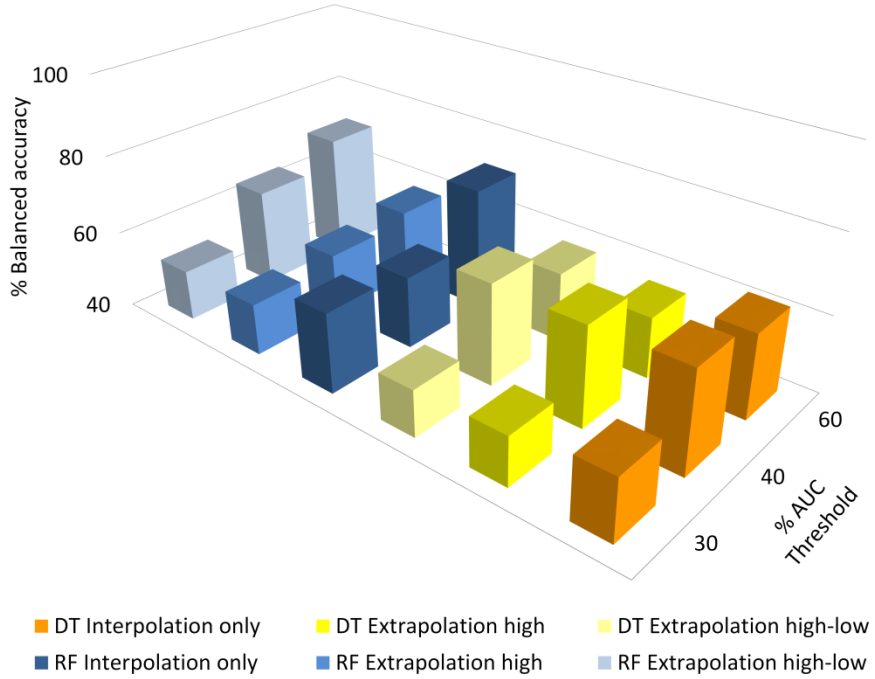


Figure 6.2. Decision tree and random forest balanced accuracies for the vehicles Saline and Saline with Tween-80.

The most accurate prediction (balanced accuracy) resulted from the RF model using a 60 % threshold with data interpolated only (Interpolation only), which produces a balanced accuracy of 71.4 %. The individual class accuracies at these settings for saline>saline with Tween-80 and saline with Tween-80>saline were 69.0 and 73.9 % respectively which suggests the model is not overly biased in favour of one class. The total number of compounds for this model was 52, with a class split of 29/23 for saline>saline with Tween-80/saline with Tween-80>saline. The RF model at a 60 % threshold with extrapolation at the high and low dose ranges (Extrapolation high-low) produced similar results albeit with fewer compounds.

Information extracted from the DT data show that five MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 93-QCH3, 116-CH3AACH2A, 95-NAAO, 97-NAAAO and 99-C=C.

6.4.3 Saline & HPC Model Results

The information presented in Table 6.5 shows the results of the DT and RF models for the vehicles saline and HPC at the various thresholds and different interpolation/extrapolation techniques used. The accuracies of the DTs and RFs do not differ much at all the thresholds tested, although for both models the accuracies tend to improve as the threshold is

increased. The number of compounds available decreases as the threshold is increased and as with the previous models the highest number of compounds is always achieved when extrapolation occurs at the high dose end only (Extrapolation high). For the Interpolation only and Extrapolation high-low models at all thresholds the best balanced accuracies are observed at the 60 % threshold, for both DTs and RFs. However, for the Extrapolation high models, the best balanced accuracies are observed at the 30 % threshold (see Figure 6.3), although for both the DT and RF models the prediction accuracies of the major class are far greater than the minor class (see Table 6.5).

Table 6.5. Decision tree and random forest model outputs for the vehicles Saline and HPC. The most accurate model is highlighted in yellow.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Saline>HPC/ HPC>Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Saline>HPC (%)	Correct HPC>Saline (%)	Balanced Accuracy (%)
DT	Interpolation only	30	124	66/58	1.1	12	53.2	60.6	44.8	52.7
DT	Extrapolation high	30	144	80/64	1.3	14	61.1	72.5	46.9	59.7
DT	Extrapolation high-low	30	90	43/47	0.9	9	62.2	44.2	78.7	61.5
RF	Interpolation only	30	124	66/58	1.1	12(25)	55.6	60.6	50.0	55.3
RF	Extrapolation high	30	144	80/64	1.3	14(29)	59.0	68.8	46.9	57.8
RF	Extrapolation high-low	30	90	43/47	0.9	9(18)	54.4	46.5	61.7	54.1
DT	Interpolation only	40	78	42/36	1.2	8	44.9	47.6	41.7	44.6
DT	Extrapolation high	40	105	59/46	1.3	11	53.3	59.3	45.7	52.5
DT	Extrapolation high-low	40	52	25/27	0.9	5	59.6	60.0	59.3	59.6
RF	Interpolation only	40	78	42/36	1.2	8(16)	65.4	73.8	55.6	64.7
RF	Extrapolation high	40	105	59/46	1.3	11(21)	54.3	66.1	39.1	52.6
RF	Extrapolation high-low	40	52	25/27	0.9	5(10)	59.6	52.0	66.7	59.3
DT	Interpolation only	60	39	18/21	0.9	4	69.2	66.7	71.4	69.0
DT	Extrapolation high	60	68	38/30	1.3	7	54.4	65.8	40.0	52.9
DT	Extrapolation high-low	60	22	10/12	0.8	2	72.7	60.0	83.3	71.7
RF	Interpolation only	60	39	18/21	0.9	4(8)	71.8	72.2	71.4	71.8
RF	Extrapolation high	60	68	38/30	1.3	7(14)	52.9	60.5	43.3	51.9
RF	Extrapolation high-low	60	22	10/12	0.8	2(4)	77.3	60.0	91.7	75.8

The most accurate prediction is achieved using the RF model at the 60 % threshold with extrapolation at the high and low dose ranges (Extrapolation high-low), this model resulted in a balanced accuracy of 75.8 %. For this model, 22 compounds were available which

resulted in a class split of 10/12 for saline>HPC/HPC>saline. The individual accuracies of this model for saline and HPC were 60.0 and 91.7 % respectively, which shows that this model learnt better from the major class compounds. The RF model at the 60 % threshold using interpolation only produced a similar result (71.8 %) but with a class split of 18/21 for saline>HPC/HPC>saline and individual accuracies of 72.2 and 71.4 % respectively, which suggests that this model was trained more evenly across both classes.

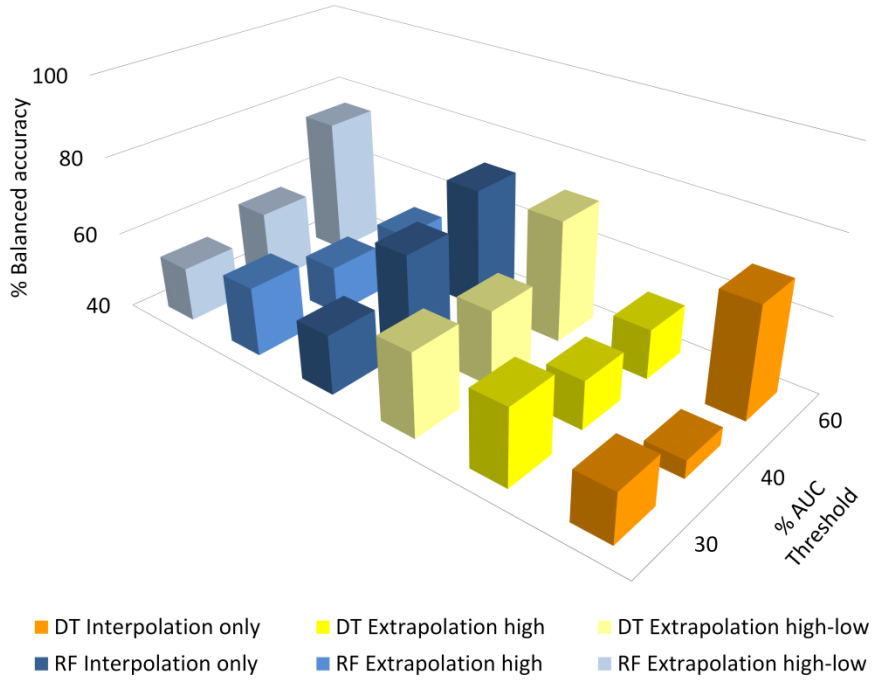


Figure 6.3. Decision tree and random forest balanced accuracies for the vehicles Saline and HPC.

Information extracted from the DT data shows that four MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 102-QO, 131-QH > 1, 148-AQ(A)A and 160-CH3.

6.4.4 Saline & Water Model Results

The information presented in Table 6.6 shows the results of the DT and RF models for the vehicles saline and water at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. At the 30 % threshold the RF models outperform all the DT models although this trend is not seen at the 40 % threshold for which the interpolation only (Interpolation only) and extrapolation at the high dose range

(Extrapolation high) models only are considered reliable. At the 60 % threshold, only data from the extrapolation at high dose range (Extrapolation high) are considered reliable, for which the RF model outperforms the DT.

The accuracies of both the DTs and RFs generally improve as the threshold is increased from 30 to 40 % although this is difficult to assess at the 60 % threshold since 4 out of 6 models are excluded due to insufficient datum points (see Figure 6.4). As with all previous models the number of compounds available per model is reduced as the threshold is increased leaving behind fewer compounds to build models with.

Table 6.6. Decision tree and random forest model outputs for the vehicles Saline and Water. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Saline>Water/ Water>Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Saline>Water (%)	Correct Water>Saline (%)	Balanced Accuracy (%)
DT	Interpolation only	30	42	25/17	1.5	4	52.4	68.0	29.4	48.7
DT	Extrapolation high	30	51	30/21	1.4	5	60.8	80.0	33.3	56.7
DT	Extrapolation high-low	30	33	15/18	0.8	3	60.6	46.7	72.2	59.4
RF	Interpolation only	30	42	25/17	1.5	4(8)	69.0	76.0	58.8	67.4
RF	Extrapolation high	30	51	30/21	1.4	5(10)	74.5	73.3	76.2	74.8
RF	Extrapolation high-low	30	33	15/18	0.8	3(7)	60.6	53.3	66.7	60.0
DT	Interpolation only	40	29	15/14	1.1	3	69.0	66.7	71.4	69.0
DT	Extrapolation high	40	34	18/16	1.1	3	82.4	94.4	68.8	81.6
DT	Extrapolation high-low	40	23	9/14	0.6	2	82.6	66.7	92.9	79.8
RF	Interpolation only	40	29	15/14	1.1	3(6)	89.7	86.7	92.9	89.8
RF	Extrapolation high	40	34	18/16	1.1	3(7)	76.5	83.3	68.8	76.0
RF	Extrapolation high-low	40	23	9/14	0.6	2(5)	91.3	77.8	100.0	88.9
DT	Interpolation only	60	18	10/8	1.3	2	83.3	80.0	87.5	83.8
DT	Extrapolation high	60	24	14/10	1.4	2	75.0	85.7	60.0	72.9
DT	Extrapolation high-low	60	13	7/6	1.2	1	53.8	100.0	0.0	50.0
RF	Interpolation only	60	18	10/8	1.3	2(4)	83.3	80.0	87.5	83.8
RF	Extrapolation high	60	24	14/10	1.4	2(5)	83.3	85.7	80.0	82.9
RF	Extrapolation high-low	60	13	7/6	1.2	1(3)	61.5	57.1	66.7	61.9

The most accurate prediction resulted from the RF model using a 40 % threshold with data interpolated only (Interpolation only), which produces a balanced accuracy of 89.8 %. The

individual accuracies at these settings for saline and water were 86.7 and 92.9 % respectively. The total number of compounds for this model was 29, with a class split of 15/14 for saline>water/water>saline. A DT model using the 40 % threshold and a RF model using the 60 % threshold both with extrapolation at the high dose range only (Extrapolation high) also produced balanced accuracies that were in the 80 % region.

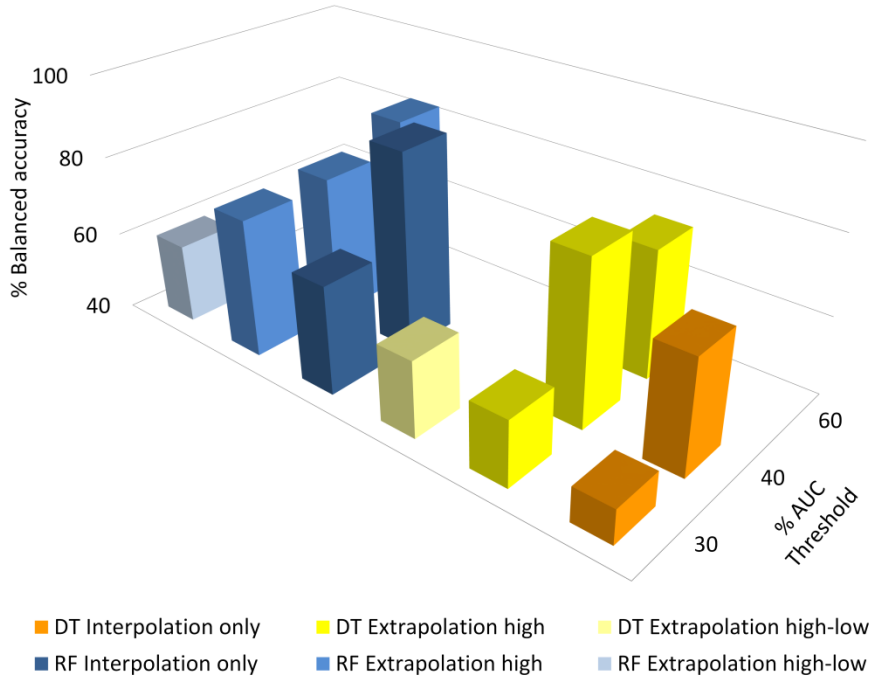


Figure 6.4. Decision tree and random forest balanced accuracies for the vehicles Saline and Water. Models with insufficient data have been excluded from the chart.

Information provided from the DT data shows that one MACCS fingerprint and one molecular property descriptor were most commonly used to split branches on, these were; 117-NAO (MACCS position-description) and the Number of heteroatoms in aliphatic rings molecular descriptor.

6.4.5 Saline & MC Model Results

The information presented in Table 6.7 shows the results of the DT and RF models for the vehicles saline and MC at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. At the 30 % threshold all RF models outperform the DT models although this pattern is less apparent at the 40 % threshold. The two RF models deemed reliable at the 60 % threshold also outperform the two DT models considered

reliable at the 60 % threshold (see Figure 6.5). The accuracies for both DTs and RFs generally improve as the threshold is increased which correlates with a decrease in the number of compounds available to model.

Table 6.7. Decision tree and random forest model outputs for the vehicles Saline and MC. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Saline>MC/ MC>Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Saline>MC (%)	Correct MC>saline (%)	Balanced Accuracy (%)
DT	Interpolation only	30	66	41/25	1.6	7	56.1	63.4	44.0	53.7
DT	Extrapolation high	30	85	52/33	1.6	9	52.9	55.8	48.5	52.1
DT	Extrapolation high-low	30	56	40/16	2.5	6	66.1	85.0	18.8	51.9
RF	Interpolation only	30	66	41/25	1.6	7(13)	65.2	68.3	60.0	64.1
RF	Extrapolation high	30	85	52/33	1.6	9(17)	61.2	69.2	48.5	58.9
RF	Extrapolation high-low	30	56	40/16	2.5	6(11)	76.8	90.0	43.8	66.9
DT	Interpolation only	40	47	31/16	1.9	5	68.1	83.9	37.5	60.7
DT	Extrapolation high	40	68	45/23	2.0	7	63.2	77.8	34.8	56.3
DT	Extrapolation high-low	40	38	28/10	2.8	4	78.9	92.9	40.0	66.4
RF	Interpolation only	40	47	31/16	1.9	5(9)	70.2	80.6	50.0	65.3
RF	Extrapolation high	40	68	45/23	2.0	7(14)	58.8	66.7	43.5	55.1
RF	Extrapolation high-low	40	38	28/10	2.8	4(8)	78.9	89.3	50.0	69.6
DT	Interpolation only	60	27	17/10	1.7	3	74.1	82.4	60.0	71.2
DT	Extrapolation high	60	51	32/19	1.7	5	60.8	81.3	26.3	53.8
DT	Extrapolation high-low	60	20	13/7	1.9	2	85.0	84.6	85.7	85.2
RF	Interpolation only	60	27	17/10	1.7	3(6)	77.8	82.4	70.0	76.2
RF	Extrapolation high	60	51	32/19	1.7	5(10)	72.5	75.0	68.4	71.7
RF	Extrapolation high-low	60	20	13/7	1.9	2(4)	90.0	100.0	71.4	85.7

The most accurate prediction resulted from the RF model using a 60 % threshold with data interpolated only (Interpolation only), which produces a balanced accuracy of 76.2 %. The individual accuracies for this model for saline>MC and MC>saline were 82.4 and 70.0 % respectively. The total number of compounds available for this model was 27, with a class split of 17/10 for saline>MC/MC>saline.

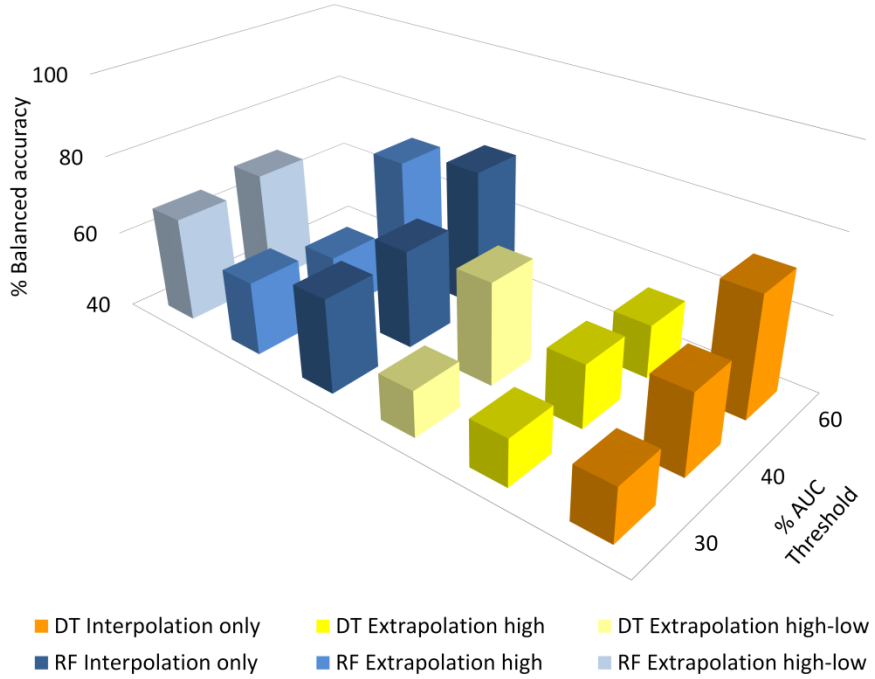


Figure 6.5. Decision tree and random forest balanced accuracies for the vehicles Saline and MC. Models with insufficient data have been excluded from the chart.

Information provided from the DT data shows that three MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 160-CH3 and 114-CH3CH2A and 144-Anot%A%Anot%A.

6.4.6 CMC & HPC Model Results

The information presented in Table 6.8 shows the results of the DT and RF models for the vehicles CMC and HPC at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. Due to the number of insufficient data available, all models at the 30 % threshold and only two models at the 40 % threshold were considered to have sufficient datum points for the models output to be acceptable. At the 30 % threshold, RF models marginally outperform the DT models, although at the 40 % threshold with extrapolation at the high dose range (Extrapolation high) the DT model outperforms the RF (see Figure 6.6).

Table 6.8. Decision tree and random forest model outputs for the vehicles CMC and HPC. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (CMC>HPC/ HPC>CMC)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct CMC>HPC (%)	Correct HPC>CMC (%)	Balanced Accuracy (%)
DT	Interpolation only	30	34	18/16	1.1	3	58.8	50.0	68.8	59.4
DT	Extrapolation high	30	38	20/18	1.1	4	57.9	65.0	50.0	57.5
DT	Extrapolation high-low	30	25	11/14	0.8	3	60.0	54.5	64.3	59.4
RF	Interpolation only	30	34	18/16	1.1	3(7)	61.8	66.7	56.3	61.5
RF	Extrapolation high	30	38	20/18	1.1	4(8)	57.9	55.0	61.1	58.1
RF	Extrapolation high-low	30	25	11/14	0.8	3(5)	64.0	54.5	71.4	63.0
DT	Interpolation only	40	19	12/7	1.7	2	73.7	83.3	57.1	70.2
DT	Extrapolation high	40	28	15/13	1.2	3	75.0	86.7	61.5	74.1
DT	Extrapolation high-low	40	11	6/5	1.2	1	54.5	100.0	0.0	50.0
RF	Interpolation only	40	19	12/7	1.7	2(4)	78.9	91.7	57.1	74.4
RF	Extrapolation high	40	28	15/13	1.2	3(6)	60.7	60.0	61.5	60.8
RF	Extrapolation high-low	40	11	6/5	1.2	1(2)	81.8	66.7	100.0	83.3
DT	Interpolation only	60	10	6/4	1.5	1	60.0	100.0	0.0	50.0
DT	Extrapolation high	60	19	11/8	1.4	2	78.9	81.8	75.0	78.4
DT	Extrapolation high-low	60	6	3/3	1.0	1	50.0	100.0	0.0	50.0
RF	Interpolation only	60	10	6/4	1.5	1(2)	90.0	100.0	75.0	87.5
RF	Extrapolation high	60	19	11/8	1.4	2(4)	78.9	81.8	75.0	78.4
RF	Extrapolation high-low	60	6	3/3	1.0	1(1)	66.7	66.7	66.7	66.7

The most accurate prediction resulted from the DT model using a 40 % threshold with data extrapolated at the high dose range (Extrapolation high), which produces a balanced accuracy of 74.1 %. The individual accuracies for this model for CMC>HPC and HPC>CMC were 86.7 and 61.5 % respectively. The total number of compounds available for this model was 28, with a class split of 15/13 for CMC>HPC/HPC>CMC. This DT model therefore predicts the major class better than the minor class, although the major class was only marginally more populated.

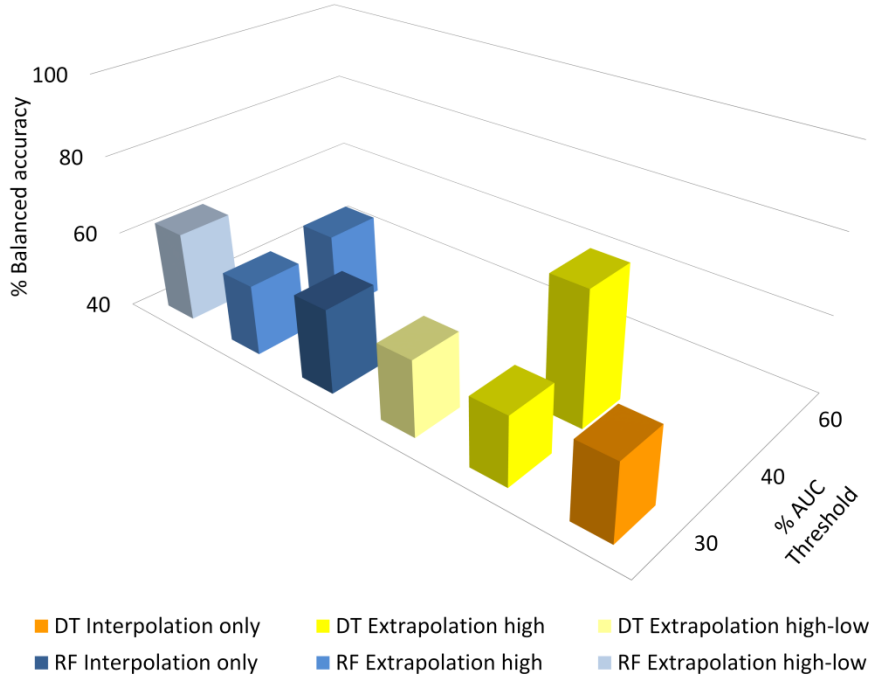


Figure 6.6. Decision tree and random forest balanced accuracies for the vehicles CMC and HPC. Models with insufficient data have been excluded from the chart.

Information provided from the DT data shows that three MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 103-CL, 152-OC(C)C and 155-A!CH2!A.

6.4.7 CMC & MC Model Results

The information presented in Table 6.9 show the results of the DT and RF models for the vehicles CMC and MC at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. The balanced accuracy of the DT models at the 30 and 40 % thresholds do not differ greatly. The difference in the balanced accuracy for the RF models at the 30 and 40 % thresholds is more apparent, with one model at the 30 % threshold (Extrapolation high) performing better than at the 40 % threshold (see Figure 6.7). Interestingly the DT and RF model at the 60 % threshold using Interpolation only produced identical outputs.

Table 6.9. Decision tree and random forest model outputs for the vehicles CMC and MC. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (CMC>MC/ MC>CMC)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct CMC>MC (%)	Correct MC>CMC (%)	Balanced Accuracy (%)
DT	Interpolation only	30	69	41/28	1.5	7	65.2	78.0	46.4	62.2
DT	Extrapolation high	30	80	49/31	1.6	8	61.3	69.4	48.4	58.9
DT	Extrapolation high-low	30	59	39/20	2.0	6	72.9	82.1	55.0	68.5
RF	Interpolation only	30	69	41/28	1.5	7(14)	59.4	70.7	42.9	56.8
RF	Extrapolation high	30	80	49/31	1.6	8(16)	66.3	75.5	51.6	63.6
RF	Extrapolation high-low	30	59	39/20	2.0	6(12)	72.9	79.5	60.0	69.7
DT	Interpolation only	40	50	29/21	1.4	5	64.0	75.9	47.6	61.7
DT	Extrapolation high	40	62	37/25	1.5	6	58.1	67.6	44.0	55.8
DT	Extrapolation high-low	40	39	26/13	2.0	4	71.8	76.9	61.5	69.2
RF	Interpolation only	40	50	29/21	1.4	5(10)	60.0	69.0	47.6	58.3
RF	Extrapolation high	40	62	37/25	1.5	6(12)	54.8	64.9	40.0	52.4
RF	Extrapolation high-low	40	39	26/13	2.0	4(8)	74.4	76.9	69.2	73.1
DT	Interpolation only	60	25	13/12	1.1	3	72.0	69.2	75.0	72.1
DT	Extrapolation high	60	40	20/20	1.0	4	75.0	55.0	95.0	75.0
DT	Extrapolation high-low	60	21	12/9	1.3	2	57.1	100.0	0.0	50.0
RF	Interpolation only	60	25	13/12	1.1	3(5)	72.0	69.2	75.0	72.1
RF	Extrapolation high	60	40	20/20	1.0	4(8)	57.5	60.0	55.0	57.5
RF	Extrapolation high-low	60	21	12/9	1.3	2(4)	76.2	83.3	66.7	75.0

The most accurate prediction resulted from the DT model using a 60 % threshold with data extrapolated at the high dose range only (Extrapolation high), which produces a balanced accuracy of 75.0 %. The individual accuracies for this model for CMC>MC and MC>CMC were 55.0 and 95.0 % respectively. The total number of compounds for this model was 40, with an even 20/20 class split, which shows the model was much better at predicting the MC>CMC class than the CMC>MC class.

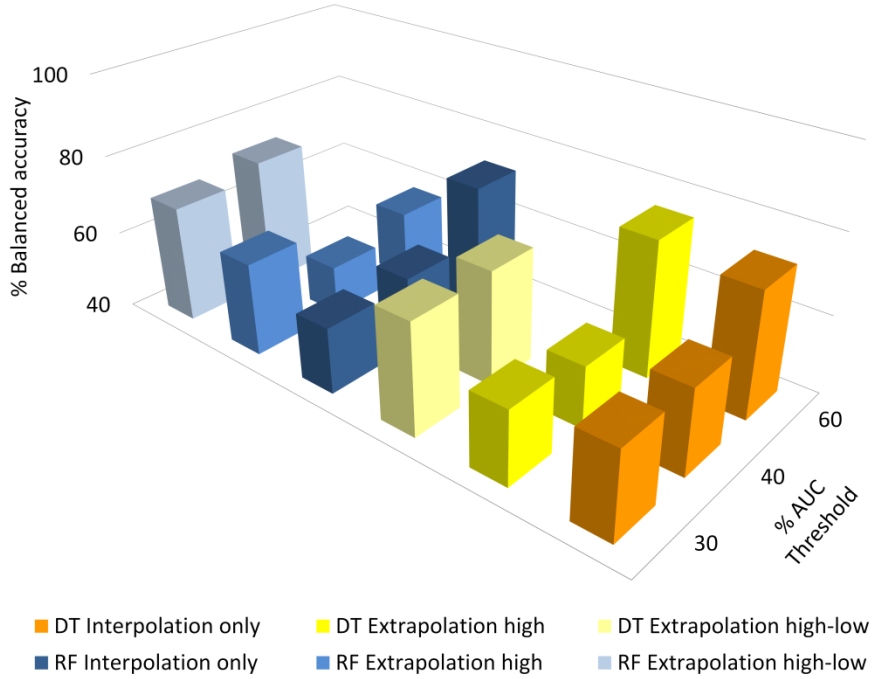


Figure 6.7. Decision tree and random forest balanced accuracies for the vehicles CMC and MC. Models with insufficient data have been excluded from the chart.

Information provided from the DT data shows that four MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 105-A$A($A)$A, 125-AROMATIC RING > 1, 151-NH and 75-A!N$A.

6.4.8 HPC & Saline with Tween-80 Model Results

The information presented in Table 6.10 show the results of the DT and RF models for the vehicles HPC and saline with Tween-80 at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered.

The accuracies of both DTs and RFs are generally better at the 40 and 60 % thresholds compared to the 30 % threshold however some models perform better at 40 % than at 60 % (see Figure 6.8). The number of compounds available decreases as the threshold is increased and as with the previous models the highest number of compounds is always achieved when extrapolation occurs at the high dose range only (Extrapolation high). Both the DT and RF models employing extrapolation at the high and low dose range (Extrapolation high-low) resulted in insufficient data for accurate models to be built.

Table 6.10. Decision tree and random forest model outputs for the vehicles HPC and Saline with Tween-80. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class split (HPC>Saline with Tween-80/ Saline with Tween-80>HPC)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct HPC>Saline with Tween-80 (%)	Correct Saline with Tween-80>HPC (%)	Balanced Accuracy (%)
DT	Interpolation only	30	106	55/51	1.1	11	53.8	69.1	37.3	53.2
DT	Extrapolation high	30	124	65/59	1.1	12	60.5	52.3	69.5	60.9
DT	Extrapolation high-low	30	89	43/46	0.9	9	69.7	62.8	76.1	69.4
RF	Interpolation only	30	106	55/51	1.1	11(21)	52.8	56.4	49.0	52.7
RF	Extrapolation high	30	124	65/59	1.1	12(25)	66.1	69.2	62.7	66.0
RF	Extrapolation high-low	30	89	43/46	0.9	9(18)	55.1	46.5	63.0	54.8
DT	Interpolation only	40	68	30/38	0.8	7	64.7	70.0	60.5	65.3
DT	Extrapolation high	40	86	41/45	0.9	9	64.0	68.3	60.0	64.1
DT	Extrapolation high-low	40	49	22/27	0.8	5	71.4	54.5	85.2	69.9
RF	Interpolation only	40	68	30/38	0.8	7(14)	76.5	80.0	73.7	76.8
RF	Extrapolation high	40	86	41/45	0.9	9(17)	60.5	58.5	62.2	60.4
RF	Extrapolation high-low	40	49	22/27	0.8	5(10)	67.3	63.6	70.4	67.0
DT	Interpolation only	60	32	19/13	1.5	3	65.6	78.9	46.2	62.6
DT	Extrapolation high	60	47	27/20	1.4	5	72.3	85.2	55.0	70.1
DT	Extrapolation high-low	60	20	14/6	2.3	2	70.0	100.0	0.0	50.0
RF	Interpolation only	60	32	19/13	1.5	3(6)	65.6	63.2	69.2	66.2
RF	Extrapolation high	60	47	27/20	1.4	5(9)	76.6	88.9	60.0	74.4
RF	Extrapolation high-low	60	20	14/6	2.3	2(4)	80.0	100.0	33.3	66.7

Interestingly the most accurate prediction of 76.8 % was seen at the 40 % threshold for the RF model using the Interpolation only technique. For this model, 68 compounds were available which resulted in a class split of 30/38 for HPC>saline with Tween-80/saline with Tween-80>HPC. The individual accuracies for this model for HPC>saline with Tween-80 and saline with Tween-80>HPC were 80.0 and 73.7 % respectively, which shows this model to predict the minor class better than the major class.

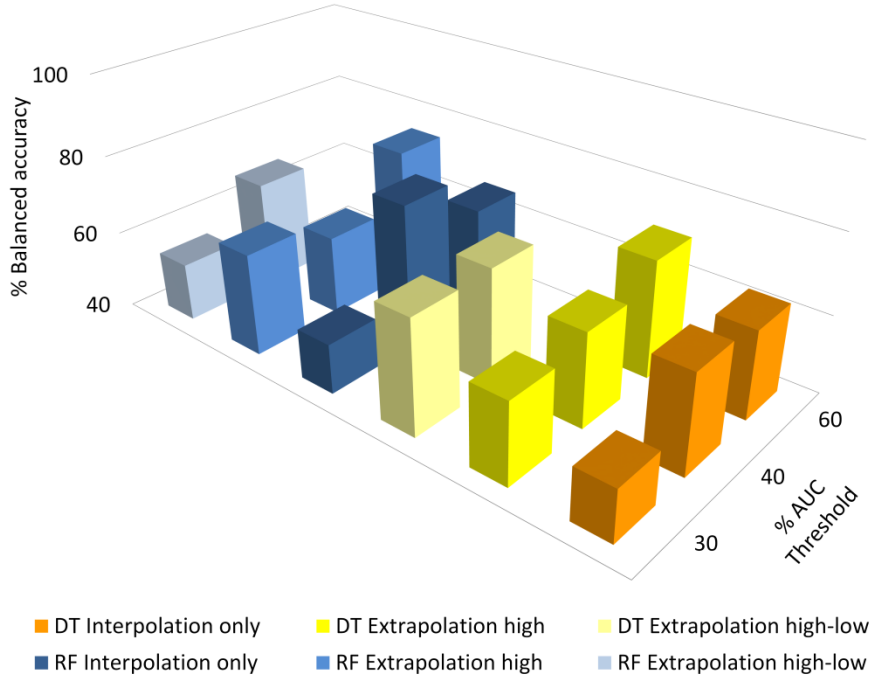


Figure 6.8. Decision tree and random forest balanced accuracies for the vehicles HPC and Saline with Tween-80. Models with insufficient data have been excluded from the chart.

Information extracted from the decision tree models shows that only three MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 114-CH3CH2A, 150-A!A$A!A and 163-6M RING.

6.4.9 Distilled Water + Alcohol & Saline Model Results

The information presented in Table 6.11 show the results of the DT and RF models for the vehicles distilled water + alcohol and saline at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. Thus all models at the 60 % threshold are considered unreliable since insufficient data were available at that threshold. At the 30 % threshold most models produce balanced accuracies in the 50-60 % region which generally improves when the threshold is 40 % (see Figure 6.9). The DT models at the 30 % and 40 % threshold using extrapolation at the high and low dose range (Extrapolation high-low) resulted in balanced accuracies of <50 %.

Table 6.11. Decision tree and random forest model outputs for the vehicles Distilled water + alcohol and Saline. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Distilled water + alcohol>Saline/ Saline>Distilled water + alcohol)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Distilled water + alcohol>Saline (%)	Correct Saline>Distilled water + alcohol (%)	Balanced Accuracy (%)
DT	Interpolation only	30	41	17/24	0.7	4	68.3	41.2	87.5	64.3
DT	Extrapolation high	30	47	21/26	0.8	5	63.8	47.6	76.9	62.3
DT	Extrapolation high-low	30	36	13/23	0.6	4	55.6	15.4	78.3	46.8
RF	Interpolation only	30	41	17/24	0.7	4(8)	56.1	29.4	75.0	52.2
RF	Extrapolation high	30	47	21/26	0.8	5(9)	55.3	38.1	69.2	53.7
RF	Extrapolation high-low	30	36	13/23	0.6	4(8)	75.0	46.2	91.3	68.7
DT	Interpolation only	40	32	14/18	0.8	3	75.0	57.1	88.9	73.0
DT	Extrapolation high	40	38	17/21	0.8	4	57.9	47.1	66.7	56.9
DT	Extrapolation high-low	40	29	10/19	0.5	3	55.2	20.0	73.7	46.8
RF	Interpolation only	40	32	14/18	0.8	3(6)	71.9	57.1	83.3	70.2
RF	Extrapolation high	40	38	17/21	0.8	4(8)	68.4	58.8	76.2	67.5
RF	Extrapolation high-low	40	29	10/19	0.5	3(6)	62.1	40.0	73.7	56.8
DT	Interpolation only	60	19	8/11	0.7	2	78.9	100.0	63.6	81.8
DT	Extrapolation high	60	27	9/18	0.5	3	63.0	55.6	66.7	61.1
DT	Extrapolation high-low	60	17	5/12	0.4	2	70.6	0.0	100.0	50.0
RF	Interpolation only	60	19	8/11	0.7	2(4)	63.2	62.5	63.6	63.1
RF	Extrapolation high	60	27	9/18	0.5	3(5)	55.6	44.4	61.1	52.8
RF	Extrapolation high-low	60	17	5/12	0.4	2(3)	82.4	60.0	91.7	75.8

The best prediction accuracy, of 73.0 % is seen using Interpolation only for the DT model at the 40 % threshold, which is the highest threshold available for these models. For this model, 32 compounds were available which resulted in a class split of 14/18 for distilled water + alcohol>saline/saline>distilled water + alcohol. The individual accuracies for this model for distilled water + alcohol>saline and saline>distilled water + alcohol were 57.1 and 88.9 % respectively, showing the major class is predicted much more accurately than the minor class. Interestingly the RF model at the 40 % threshold using Interpolation only produces a balanced accuracy of 70.2 % but also predicts the major class significantly better than the minor class (see Table 6.11).

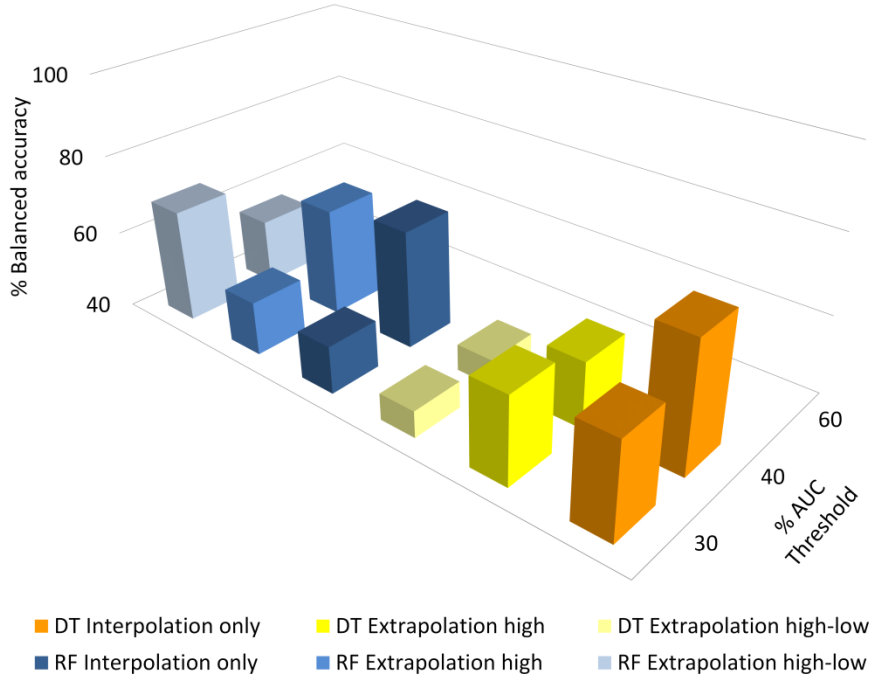


Figure 6.9. Decision tree and random forest balanced accuracies for the vehicles Distilled water + alcohol and Saline. Models with insufficient data have been excluded from the chart.

Information extracted from the decision tree models shows that three MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 98-QAAAAA@1, 166-FRAGMENTS and 94-QN.

6.4.10 Distilled Water + Alcohol & MC Model Results

The information presented in **Table 6.12** show the results of the DT and RF models for the vehicles distilled water + alcohol and MC at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. The greatest accuracies for this model were achieved at the 30 % threshold. Only the Extrapolation high models at the 60 % threshold produced sufficient data, but both the DT and RF produced balanced accuracies in the 50 % region (see Figure 6.10). Of the four models available at the 40 % threshold, the RF models were either equivalent or better than the DT models. It is important to note that the DT model at the 40 % threshold using Interpolation only predicted the major class with 100 % accuracy but predicted the minor class with 0 % accuracy indicating that the compounds in these two classes cannot be successfully modelled using the descriptors that were used.

Table 6.12. Decision tree and random forest model outputs for the vehicles Distilled water + alcohol and MC. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Distilled water + alcohol>MC/ MC>Distilled water + alcohol)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Distilled water + alcohol>MC (%)	Correct MC>Distilled water + alcohol (%)	Balanced Accuracy (%)
DT	Interpolation only	30	50	32/18	1.8	5	66.0	59.4	77.8	68.6
DT	Extrapolation high	30	78	43/35	1.2	8	57.7	67.4	45.7	56.6
DT	Extrapolation high-low	30	31	19/12	1.6	3	74.2	57.9	100.0	78.9
RF	Interpolation only	30	50	32/18	1.8	5(10)	74.0	75.0	72.2	73.6
RF	Extrapolation high	30	78	43/35	1.2	8(16)	65.4	72.1	57.1	64.6
RF	Extrapolation high-low	30	31	19/12	1.6	3(6)	77.4	78.9	75.0	77.0
DT	Interpolation only	40	41	28/13	2.2	4	68.3	100.0	0.0	50.0
DT	Extrapolation high	40	68	39/29	1.3	7	61.8	69.2	51.7	60.5
DT	Extrapolation high-low	40	22	16/6	2.7	2	72.7	100.0	0.0	50.0
RF	Interpolation only	40	41	28/13	2.2	4(8)	75.6	89.3	46.2	67.7
RF	Extrapolation high	40	68	39/29	1.3	7(14)	60.3	64.1	55.2	59.6
RF	Extrapolation high-low	40	22	16/6	2.7	2(4)	81.8	100.0	33.3	66.7
DT	Interpolation only	60	25	16/9	1.8	3	60.0	87.5	11.1	49.3
DT	Extrapolation high	60	53	28/25	1.1	5	56.6	57.1	56.0	56.6
DT	Extrapolation high-low	60	11	8/3	2.7	1	72.7	100.0	0.0	50.0
RF	Interpolation only	60	25	16/9	1.8	3(5)	80.0	68.8	100.0	84.4
RF	Extrapolation high	60	53	28/25	1.1	5(11)	56.6	57.1	56.0	56.6
RF	Extrapolation high-low	60	11	8/3	2.7	1(2)	100.0	100.0	100.0	100.0

The greatest prediction accuracy of 78.9 % is achieved with a DT model at the 30 % threshold using extrapolation at the high and low dose range (Extrapolation high-low). For this model, 31 compounds were available which resulted in a class split of 19/12 for distilled water + alcohol>MC/MC>distilled water + alcohol. The individual accuracies for this model for distilled water + alcohol and MC were 57.9 and 100.0 % respectively, which shows the minor class is better predicted than the major class.

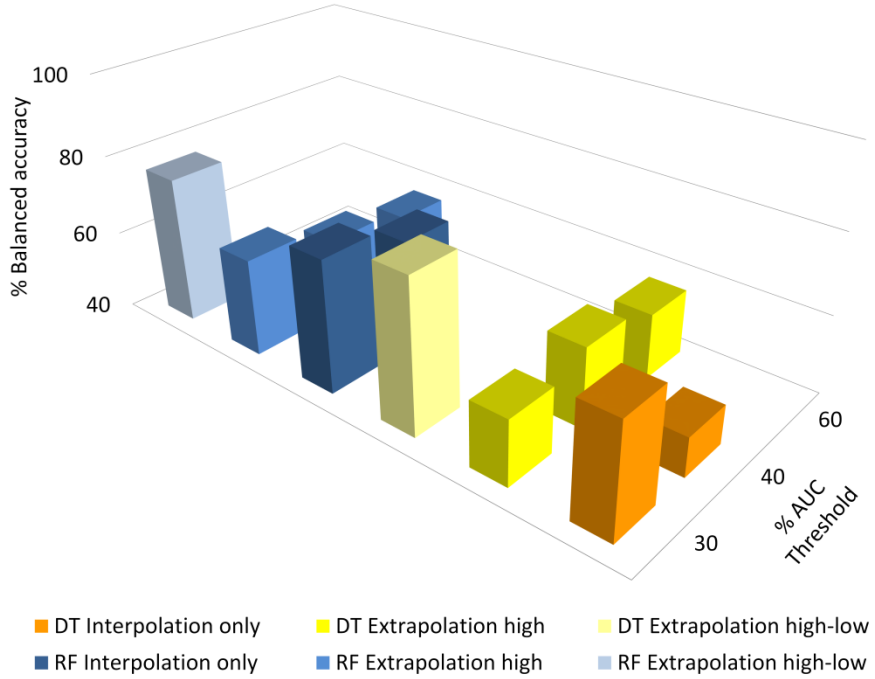


Figure 6.10. Decision tree and random forest balanced accuracies for the vehicles Distilled water + alcohol and MC. Models with insufficient data have been excluded from the chart.

Information extracted from the decision tree models shows that three MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 112-AA(A)(A)A, 72-OAAO and 79-NAAN.

6.4.11 Distilled Water + Alcohol & CMC Model Results

The information presented in Table 6.13 show the results of the DT and RF models for the vehicles distilled water + alcohol and CMC at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. At the 60 % threshold only the Extrapolation high DT and RF models were considered to have sufficient data for the models to be considered reliable (see Figure 6.11). Interestingly the models in the 30 % threshold generally resulted in better predictions than at the 40 or 60 % thresholds.

Table 6.13. Decision tree and random forest model outputs for the vehicles Distilled water + alcohol and CMC. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Distilled water + alcohol>CMC/ CMC>Distilled water + alcohol)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Distilled water + alcohol>CMC (%)	Correct CMC>Distilled water + alcohol (%)	Balanced Accuracy (%)
DT	Interpolation only	30	42	21/21	1.0	4	73.8	81.0	66.7	73.8
DT	Extrapolation high	30	51	27/24	1.1	5	74.5	70.4	79.2	74.8
DT	Extrapolation high-low	30	35	15/20	0.8	4	62.9	33.3	85.0	59.2
RF	Interpolation only	30	42	21/21	1.0	4(8)	78.6	81.0	76.2	78.6
RF	Extrapolation high	30	51	27/24	1.1	5(10)	76.5	77.8	75.0	76.4
RF	Extrapolation high-low	30	35	15/20	0.8	4(7)	65.7	46.7	80.0	63.3
DT	Interpolation only	40	29	16/13	1.2	3	55.2	56.3	53.8	55.0
DT	Extrapolation high	40	39	20/19	1.1	4	61.5	50.0	73.7	61.8
DT	Extrapolation high-low	40	25	11/14	0.8	3	60.0	81.8	42.9	62.3
RF	Interpolation only	40	29	16/13	1.2	3(6)	75.9	68.8	84.6	76.7
RF	Extrapolation high	40	39	20/19	1.1	4(8)	69.2	65.0	73.7	69.3
RF	Extrapolation high-low	40	25	11/14	0.8	3(5)	56.0	45.5	64.3	54.9
DT	Interpolation only	60	17	10/7	1.4	2	100.0	100.0	100.0	100.0
DT	Extrapolation high	60	27	14/13	1.1	3	66.7	64.3	69.2	66.8
DT	Extrapolation high-low	60	14	7/7	1.0	1	100.0	100.0	100.0	100.0
RF	Interpolation only	60	17	10/7	1.4	2(3)	100.0	100.0	100.0	100.0
RF	Extrapolation high	60	27	14/13	1.1	3(5)	70.4	71.4	69.2	70.3
RF	Extrapolation high-low	60	14	7/7	1.0	1(3)	92.9	85.7	100.0	92.9

The greatest prediction accuracy was observed to be 78.6 % for a RF model at the 30 % threshold using Interpolation only. For this model there were 42 compounds available with a class split of 21/21 for distilled water + alcohol>CMC/CMC>distilled water + alcohol. The individual accuracies of this model were 81.0 and 76.2 for distilled water + alcohol>CMC and CMC>distilled water + alcohol respectively, showing that this model was marginally better at predicting the distilled water + alcohol>CMC class despite the even class split.

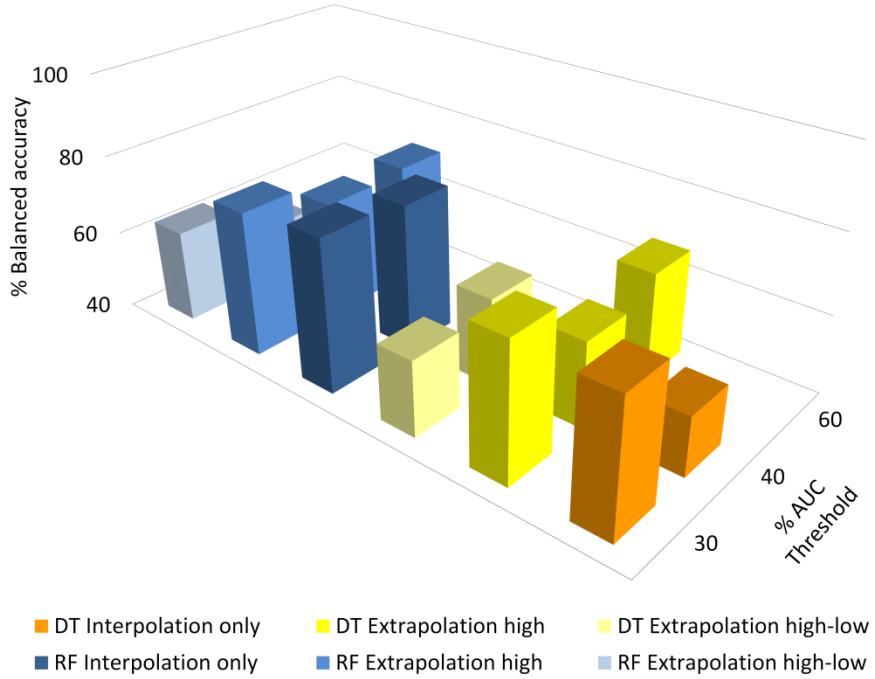


Figure 6.11. Decision tree and random forest balanced accuracies for the vehicles Distilled water + alcohol and CMC. Models with insufficient data have been excluded from the chart.

Information extracted from the decision tree models shows that four MACCS fingerprints were most commonly used to split branches on, these were (MACCS position-description); 156-NA(A)A, 16-QAA@1, 109-ACH2O and 94-QN.

6.4.12 Acetone & Saline Model Results

The information presented in Table 6.14 show the results of the DT and RF models for the vehicles acetone and saline at the various thresholds and different interpolation/extrapolation techniques used. The data shaded grey are considered to be insufficient (i.e. <10 compounds in a class) for reliable models to be built and the accuracies associated with them should not be considered. It is interesting to note that all the models extrapolated at the high and low dose range (Extrapolation high-low) were the models that resulted in insufficient data (see Figure 6.12). Of the remaining models, at each of the thresholds the RF models tend to outperform the DT models with the exception at the 60 % threshold using Interpolation only. Both DT models at the 40 and 60 % threshold with extrapolation at the high dose range (Extrapolation high) resulted in a 0 % accuracy for predicting the minor class which always happens to be acetone.

Table 6.14. Decision tree and random forest model outputs for the vehicles Acetone and Saline. The most accurate model is highlighted in yellow. Models considered unreliable due to insufficient data are shaded in grey.

Model	Interpolation/ Extrapolation	% AUC Threshold	No. of Compounds	Class Split (Acetone/Saline)	Data Bias	No. of descriptors (pool of descriptors)	Accuracy (%)	Correct Acetone>Saline (%)	Correct Saline>Acetone (%)	Balanced Accuracy (%)
DT	Interpolation only	30	42	22/20	1.1	4	57.1	59.1	55.0	57.0
DT	Extrapolation high	30	54	17/37	0.5	5	72.2	17.6	97.3	57.5
DT	Extrapolation high-low	30	22	8/14	0.6	2	63.6	0.0	100.0	50.0
RF	Interpolation only	30	42	22/20	1.1	4(8)	83.3	72.7	95.0	83.9
RF	Extrapolation high	30	54	17/37	0.5	5(11)	70.4	23.5	91.9	57.7
RF	Extrapolation high-low	30	22	8/14	0.6	2(4)	81.8	50.0	100.0	75.0
DT	Interpolation only	40	35	19/16	1.2	4	54.3	52.6	56.3	54.4
DT	Extrapolation high	40	49	14/35	0.4	5	71.4	0.0	100.0	50.0
DT	Extrapolation high-low	40	17	6/11	0.5	2	64.7	0.0	100.0	50.0
RF	Interpolation only	40	35	19/16	1.2	4(7)	68.6	57.9	81.3	69.6
RF	Extrapolation high	40	49	14/35	0.4	5(10)	77.6	28.6	97.1	62.9
RF	Extrapolation high-low	40	17	6/11	0.5	2(3)	88.2	66.7	100.0	83.3
DT	Interpolation only	60	27	16/11	1.5	3	77.8	93.8	54.5	74.1
DT	Extrapolation high	60	38	11/27	0.4	4	71.1	0.0	100.0	50.0
DT	Extrapolation high-low	60	12	4/8	0.5	1	66.7	75.0	62.5	68.8
RF	Interpolation only	60	27	16/11	1.5	3(5)	66.7	75.0	54.5	64.8
RF	Extrapolation high	60	38	11/27	0.4	4(8)	81.6	36.4	100.0	68.2
RF	Extrapolation high-low	60	12	4/8	0.5	1(2)	66.7	50.0	75.0	62.5

The best prediction is observed for the RF model at the 30 % threshold using Interpolation only which results in a balanced accuracy of 83.9 %. For this model 42 compounds were available which resulted in a class split of 22/20 for acetone>saline/saline>acetone. The individual accuracies for this model for acetone>saline and saline>acetone were 72.7 and 95.0 % respectively, which shows that this model is better at predicting the minor class albeit marginally minor. Interestingly the DT models at the 40 and 60 % threshold using extrapolation at the high dose range (Extrapolation high) result in a 0 % prediction accuracy for the minor class (acetone>saline) and 100 % accuracy for the major class (saline>acetone), which indicates the data available cannot be suitably modelled with the descriptors that were used.

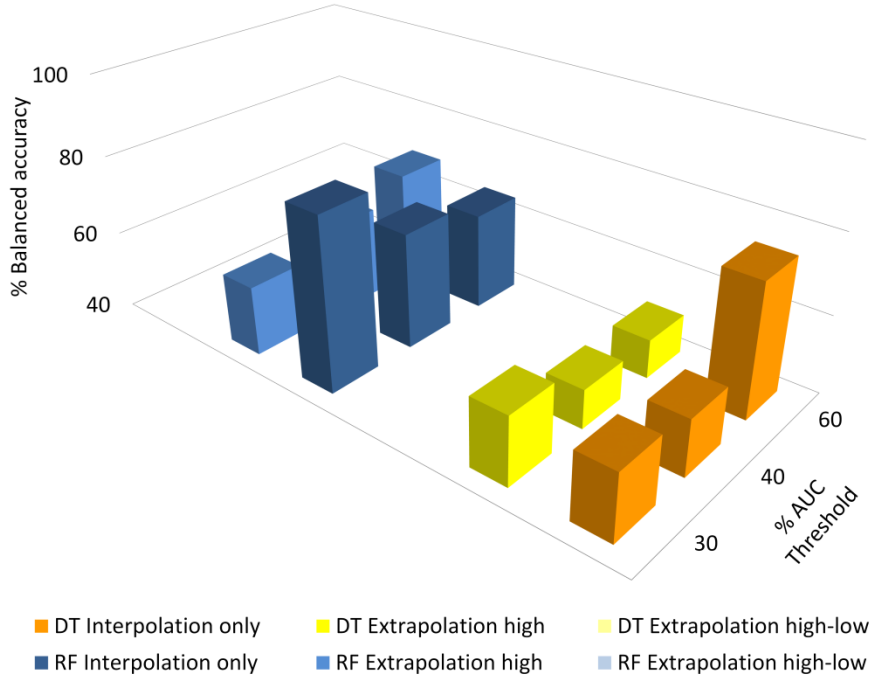


Figure 6.12. Decision tree and random forest balanced accuracies for the vehicles Acetone and Saline. Models with insufficient data have been excluded from the chart.

Information extracted from the decision tree models shows that one MACCS fingerprint and one molecular property descriptor were most commonly used to split branches on. This MACCS fingerprint was, 38-NC(C)N (MACCS position-description) and the molecular property descriptor was the Number of carbons.

6.5 Knowledge Extraction From Predictive Models

Information extracted from the DT models pertaining to the fingerprints and molecular descriptors used to split trees on provides a useful insight into how each model makes decisions prior to providing a prediction.

If any descriptor(s) used, correlate highly with the binary class of a model then these descriptors may be useful in turning this information into valuable scientific knowledge. More specifically if any descriptors discovered are to perfectly correlate with the class of a model, this would provide invaluable information about the type of compounds that are suitable for toxicity reduction and generalisations can be formed from this information.

To discover such information, all the DT models can be assessed for descriptor correlation. Models that were deemed unreliable for predictions due to insufficient data can also be used to gather descriptor information. To reiterate, insufficient datum points are

considered as <10 compounds in any one class. These models are considered unreliable for predictions due to the 10-fold cross validation method used in training the models (see chapter 5 section 5.9).

For the models reported in this chapter using the AUC methodology, the 60 % threshold is the point at which the most signal is observed, i.e. there is less noise in the data. Models at the 60 % threshold therefore become the point at which any knowledge discovery would be most apparent, despite the number of datum points being low.

An analysis of the DT data for all vehicle pairings using all interpolation/extrapolation techniques was conducted at the 60 % threshold to assess any strong correlations which may occur between a descriptor and the binary class.

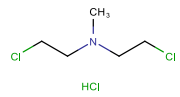
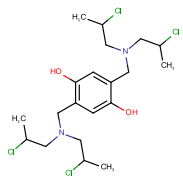
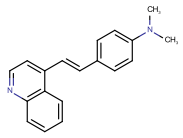
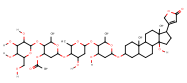
6.5.1 Distilled Water + Alcohol & CMC Knowledge Extraction

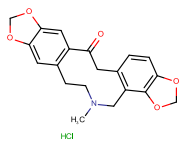
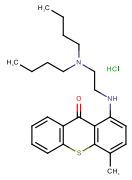
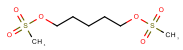
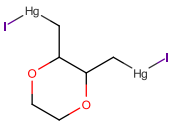
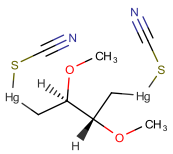
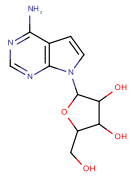
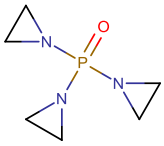
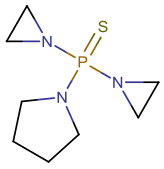
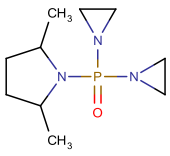
The models for the vehicle pairing distilled water + alcohol & CMC are reported in section 6.4.11 of this chapter. At the 60 % threshold only 2 of the 6 possible models contained sufficient datum points to be considered reliable. For this vehicle pairing at the 60 % threshold the descriptors used by the DT models for Interpolation only, Extrapolation high and Extrapolation high-low were extracted regardless of model reliability. It is not effective predictions but rather compound descriptions that are sought after here. Models considered unreliable may still contain this information from the descriptor generation step, prior to descriptor filtering. Furthermore it is often the models at the 60 % threshold that are able to eliminate noise from the data.

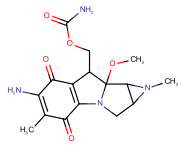
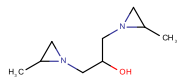
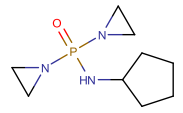
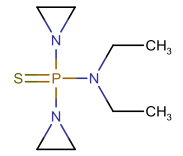
For two of the DT models at the 60 % threshold; Interpolation only and Extrapolation high-low, two MACCS fingerprints perfectly correlate with the binary class. These fingerprints are (MACCS position-description); 16-QAA@1 and 22-3M RING, which implies that compound structures containing a 3-membered ring, are solely classified into one class. Upon inspection of the compounds contained within the two models (Interpolation only and Extrapolation high-low, both at the 60 % threshold) it was identified that the 3-membered ring contained within these compounds was an aziridine ring structure. For both the Interpolation only and Extrapolation high-low models all compounds which contained the aziridine ring were classified as CMC>distilled water + alcohol, which implies that the use of the vehicle CMC with structures containing the aziridine functional group are better suited to a toxicity reduction compared to the vehicle distilled water + alcohol.

To demonstrate this, consider the information presented in Table 6.15, which shows the structures of the compounds that were available for the Interpolation only model at the 60 % threshold. The table shows the class that each of these compounds were assigned to prior to developing predictive models. For predictive purposes this particular model is considered unreliable due to insufficient data points being present in the CMC>Distilled water + alcohol class (i.e. <10 compounds). The MACCS fingerprint information shown in the table however provides some interesting scientific interpretations. The fingerprint 16-QAA@1 (MACCS position-description) relates to a structure containing a 3-membered ring with a heteroatom present, this is indicated by the presence (1) or absence (0) in the table. When comparing this fingerprint column to the class column it becomes apparent that for all compounds with the 3-membered ring present, belong to one class (CMC>distilled water + alcohol), whilst the compounds for which the 3-membered ring is not present, belong to the other class (distilled water + alcohol>CMC). An identical observation is seen with the Extrapolation high-low model at the 60 % threshold, only with fewer compounds.

Table 6.15. Compound structures for the Distilled water + alcohol & CMC model. Information extracted from the Interpolation only DT model using a 60 % threshold.

Structure	NSC	16-QAA@1	Class
	762	0	Distilled water + alcohol>CMC
	3434	0	Distilled water + alcohol>CMC
	4236	0	Distilled water + alcohol>CMC
	7532	0	Distilled water + alcohol>CMC

	11440	0	Distilled water + alcohol>CMC
	15788	0	Distilled water + alcohol>CMC
	17019	0	Distilled water + alcohol>CMC
	21934	0	Distilled water + alcohol>CMC
	30911	0	Distilled water + alcohol>CMC
	56408	0	Distilled water + alcohol>CMC
	9717	1	CMC>Distilled water + alcohol
	55740	1	CMC>Distilled water + alcohol
	56296	1	CMC>Distilled water + alcohol

	56410	1	CMC>Distilled water + alcohol
	57561	1	CMC>Distilled water + alcohol
	58944	1	CMC>Distilled water + alcohol
	60096	1	CMC>Distilled water + alcohol

The correlations observed for this model can thus be used to form generalisations about the type of compounds for which the vehicle CMC may offer better toxicity protection than distilled water + alcohol. In this specific example it can be argued that any compounds containing the 3-membered aziridine ring can have their toxicity reduced relative to distilled water + alcohol through the use of the vehicle CMC.

6.6 Constructing Partially Ordered Sets from Vehicle Pairings

With a strong correlation observed for aziridine containing compounds and their respective class for the distilled water + alcohol & CMC models, an analysis of the classification of all aziridine containing structures for all other models developed in the chapter was conducted.

For each of the vehicle pairings using Interpolation only at the 60% threshold the classification divide of all aziridine containing compounds was extracted. The aim was to assess the class divide for all the other models and determine whether aziridine containing compounds predominantly favour one class for these other models also. If this can be demonstrated then the data collected can be used to form a partially ordered set which forms an ordering of vehicles which provide progressively better toxicity reduction for this group of compounds.

The class divide extracted from all possible models for all aziridine containing compounds is shown in **Table 6.16**. The table shows the vehicle pairings represented as X & Y and the number of aziridine containing compounds that are classified into each binary class (X>Y or Y>X). For example, consider the first vehicle pairing in **Table 6.16** (Row ID 1), Saline & CMC. This vehicle pairing is represented as X & Y for saline & CMC respectively. The two binary classes formed are therefore saline>CMC represented as X>Y and CMC>saline represented as Y>X. The numbers in the table represent the number of aziridine containing compounds that fall into each class using Interpolation only at the 60 % threshold for all the models discussed in this chapter. Vehicle pairings for which no aziridine containing compounds are detected contain no data in the table and are shaded grey.

Table 6.16. Classification divide of aziridine containing compounds for all vehicle pairings. Data shaded in grey did not have any aziridine containing compounds.

	Vehicle Pairing		Number of aziridine containing compounds per class	
Row ID	**X**	**Y**	**X>Y**	**Y>X**
1	Saline	CMC	5	1
2	Distilled water + alcohol	CMC	0	7
3	Distilled water + alcohol	MC	2	0
4	Distilled water + alcohol	Saline	1	8
5	Saline	MC	4	0
6	CMC	MC	2	1
7	Saline	Saline with Tween-80	3	0
8	Saline	HPC	0	2
9	Saline	Water	-	-
10	CMC	HPC	-	-
11	HPC	Saline with Tween-80	-	-
12	Acetone	Saline	-	-

From the data presented in Table 6.16 it is apparent that the number of aziridine containing compounds for each of the vehicle pairings is relatively low, however for many of the vehicle pairings it is easy to see that one class is predominantly favoured over the other. It is therefore possibly to arrange an ordering of these vehicles to show how the

toxicity of aziridine containing compounds can be reduced when the vehicles are compared to one another.

To begin with consider the first vehicle pairing of saline & CMC (Row ID 1) in the table. The data shows that 5 aziridine containing compounds are assigned to the class saline>CMC and only 1 aziridine compound to the class CMC>saline. With the class saline>CMC being predominantly favoured by these compounds we can create the following ordering:

$$\text{Saline} \longrightarrow \text{CMC} \tag{1}$$

This ordering can be read as "saline is better than CMC" and is the first step of the partially ordered set. Now consider Row ID 2 of Table 6.16, which shows the vehicle CMC to be completely favoured over the vehicle distilled water + alcohol at a ratio of 7 to 0. For aziridine containing compounds CMC can therefore be considered as better than distilled water + alcohol and can be used to create the following ordering:

$$\text{Saline} \longrightarrow \text{CMC} \longrightarrow \text{Distilled water + alcohol} \tag{2}$$

Taking the data from Row ID 3 of Table 6.16 and applying these same ordering principles shows the vehicle distilled water + alcohol to be predominantly favoured over the vehicle MC (i.e. distilled water + alcohol is better than MC), the ordering can then be represented as:

$$\text{Saline} \longrightarrow \text{CMC} \longrightarrow \text{Distilled water + alcohol} \longrightarrow \text{MC} \tag{3}$$

This ordering (3) has been constructed from the data contained in Table 6.16 and can be cumbersomely read as "saline is better than CMC which is better than distilled water + alcohol which is better than MC". This ordering of vehicles that reduce the toxicity of aziridine containing compounds shows that the vehicle saline is best suited at reducing toxicity, whilst MC is the least suited.

Although each of the individual orderings shown holds true, as seen from Row IDs 1, 2 and 3 of Table 6.16, arranging this information into such an ordered set must satisfy the other relationships that can be made from this ordering. More specifically, from the ordering shown in (3), the following three relationships can be deduced: 1) saline is better than

distilled water + alcohol, 2) saline is better than MC and 3) CMC is better than MC. These deduced relationships must hold true for the ordered set shown in (3) to be valid.

As proof of concept, the data from the other vehicle pairings shown in Table 6.16 must agree with these deduced relationships.

Row ID 4 of Table 6.16 shows the first deduced relationship to hold true. That is, for the vehicle pairings distilled water + alcohol & saline, 8 aziridine containing compounds were classified as saline>distilled water + alcohol whilst only 1 aziridine containing compound was classified as distilled water + alcohol>saline. The class saline>distilled water + alcohol is therefore predominantly favoured and agrees with the first deduced relationship. This is shown as a red dotted arrow in the ordering below:

(4)

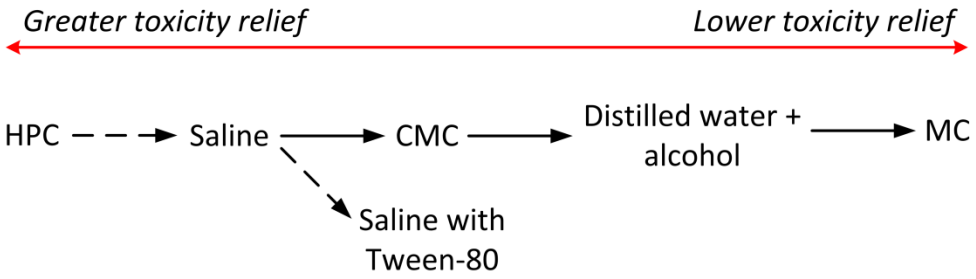
The second deduced relationship can be satisfied from the data in Row ID 5 of Table 6.16, which shows the class saline>MC is completely favoured over the class MC>saline. This relationship therefore satisfies the second deduced relationship and is shown as a green dotted arrow in the ordering below:

(5)

Lastly, the third deduction made can be shown to hold true from the data in Row ID 6 of Table 6.16, which shows the class CMC>MC to be favoured over the class MC>CMC. Although this preference is marginal (i.e. 2:1), it agrees with the third deduced relationship and from the data available does not suggest a counter argument. This relationship can therefore be considered true and is shown as a blue dotted arrow in the ordering below:

(6)

To complete this ordering and form a partially ordered set, two other relationships shown in Row IDs 7 and 8 of Table 6.16 are to be included. Row ID 7 shows the class saline>saline with Tween-80 is completely favoured over the class saline with Tween-80>saline. Similarly, Row ID 8 shows the class HPC>saline to be completely favoured over the class saline>HPC. Including these two relationships into the ordered set results in the partially ordered set shown below:



(7)

This completed partially ordered set (7) shows that the vehicle HPC is better than the vehicle saline for aziridine containing compounds. It also shows the vehicle saline is better than the vehicle saline with Tween-80 for aziridine containing compounds. The vehicle HPC has been connected to saline via a dashed arrow, which indicates that although the relationship between HPC and saline is shown to be true from the data in Table 6.16, the other relationships cannot be confirmed through any available data (i.e. HPC is better than CMC, HPC is better than distilled water + alcohol or HPC is better than MC).

The vehicle saline in the partially ordered set (7) has also been connected to the vehicle saline with Tween-80, showing that saline is better than saline with Tween-80 for aziridine containing compounds. The vehicle saline with Tween-80 however has been placed off the main branch connecting the other vehicles. It has been arranged as such since from the data available it is impossible to determine where in the main branch of this partially ordered set saline with Tween-80 should sit. Although it is known that saline is better than saline with Tween-80, it is not known if CMC, distilled water + alcohol or MC is better or worse than saline with Tween-80 and therefore saline with Tweet-80 cannot be integrated into the main branch. Furthermore, the vehicle saline with Tween-80 has been connected to the vehicle saline with a dashed line to show the potential relationship that reads HPC is better than saline with Tween-80 is untested due to insufficient data.

The partially ordered set (7) shows the ordering of the vehicles which can be used for aziridine containing compounds, it is considered a partial order since not every element is related (or has insufficient data to make a relation). The ordering from left to right shows

progressively less toxicity protection offered by the vehicles for aziridine containing compounds in relation to each other.

6.7 Conclusion

This chapter has presented twelve classification models that have been developed using the AUC methodology proposed in chapter 5. The methodology has been successfully demonstrated to show how classifiers can be built through the repurposing of *in vivo* scientific data for knowledge extraction. An additional feature of this methodology is the use of the interpolation/extrapolation techniques which increases the size of the AUC which can be compared, whilst possibly reducing the noise in the data.

The pairwise vehicle relationships that were possible tend to be between the most commonly occurring vehicles within the NIH dataset, although vehicles such as acetone and distilled water + alcohol which have an occurrence of 0.36 and 3 % respectively were successfully paired with other vehicles to build models from.

For any given vehicle pairing, as the % AUC threshold is increased a decrease in the number of compounds available to model is observed. This is explained by the larger threshold resulting in more vehicles falling into the equivalent class and therefore being removed from the dataset. With the vehicle saline being the most commonly occurring vehicle within the NIH dataset it is unsurprising that the models with the largest data points included the vehicle saline paired with another vehicle.

Figure 6.13 provides a graphic for the number of data points that were available for the various models developed using different thresholds and interpolation/extrapolation techniques. The dotted lines represent models at the 30 % threshold, the dashed lines represent the models at the 40 % threshold and the solid lines represent the models at the 60 % threshold. The colours blue, red and yellow represent the Interpolation only, Extrapolation high and Extrapolation high-low techniques respectively. For each vehicle pairing reported at a given % AUC threshold, the Extrapolation high procedure always results in the highest number of compounds, followed by the Interpolation only, with Extrapolation high-low resulting in the lowest number of compounds.

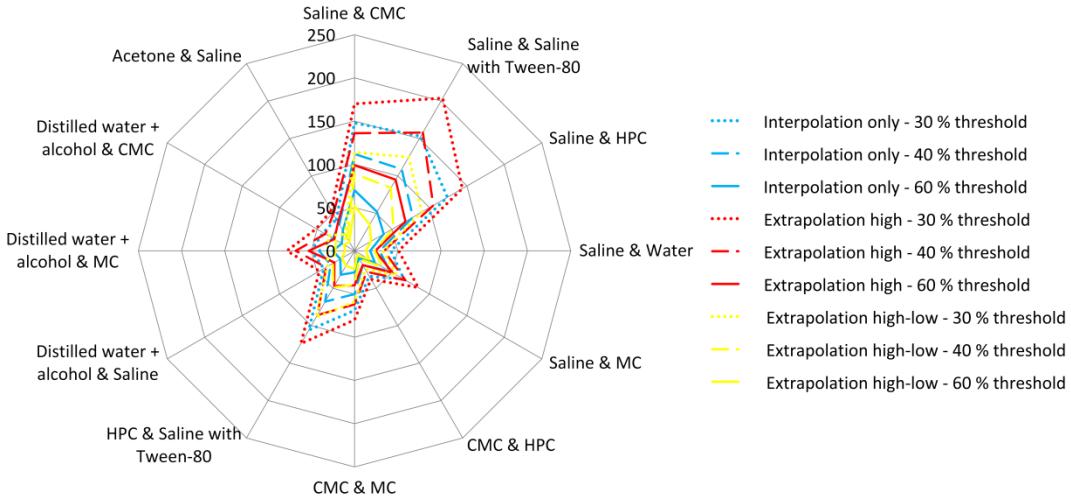


Figure 6.13. Data points per model using different thresholds and interpolation/extrapolation techniques.

Of the different thresholds and interpolation/extrapolation techniques used, a trend is not observed when correlated against the predictive accuracies of each model. Rather different models achieve their best predictive outcomes at different thresholds which use different interpolation/extrapolation techniques.

Of the 12 vehicle pairing models reported, RFs produce better balanced accuracies for 8 of these models with the remaining 4 models producing better balanced accuracies from the DT models, this is generally expected given RFs are an ensemble method. Of the different % AUC thresholds tested, 5 of the 12 models produce better accuracies at the 60 % threshold, 4 at the 40 % threshold and 3 at the 30 % threshold. This indicates that better predictive outcomes are made when the % AUC threshold is set to 60 % which filters out more noise from the data. It is important to note that some models, for example the CMC & HPC model and the distilled water + alcohol & saline model, did not produce sufficient data points at the 60 % threshold and as a result produced their best accuracies at a 40 % threshold. Had sufficient data been available to build models at the 60 % threshold then those models may also have resulted in better predictive outcomes at the 60 % threshold. This would provide further reassurance that more signal within the data is observed at a higher % AUC difference. Of the interpolation/extrapolation techniques tested, Interpolation only results in the best model for 7 out of the 12 vehicle pairings, with Extrapolation high-low and Extrapolation high representing 3 and 2 models respectively. The Extrapolation high technique results in lower accuracies despite having the most datum points, suggesting this technique is ineffective with the NIH dataset used in this study.

The overall balanced accuracies of the different models varies, with balanced accuracies exceeding 80 % for some models. For all vehicle pairings, the best balanced accuracy achieved is always above 70 %. These results suggest that the methodology proposed and tested is a reliable way to build predictive classifiers. More specifically these classifiers were built from *in vivo* real scientific data which is prone to high variability [341]. For all the models developed the data bias of the two binary classes is never greater than 3 fold and for the models which produced the best balanced accuracies for each of the vehicle pairings this bias is never greater than 2 fold between the two classes. The low data bias seen with all these models meant any corrective measures to balance the two classes was not necessary. This is reflected with some models producing better individual accuracies for the minor class which is possible given the class imbalance is not significant.

Information extracted from the DT tree models is presented for each vehicle pairings in Table 6.17. The MACCS keys fingerprints are more frequently used across all models than the molecular property descriptors. Many of the fingerprints used by the models contain oxygen and/or nitrogen atoms, for example, 72-OAAO, 79-NAAN, 95-NAAO, 97-NAAAO, 117-NAO and 151-NH, which may infer some element of hydrogen bonding occurring between the compound and a vehicle used. Fingerprints such as 53-QHAAAQH can also be implicated in hydrogen bonding if the atom symbol Q in this fingerprint is a nitrogen (N) or oxygen (O) atom. In Table 6.17, fingerprints which are potentially implicated in hydrogen bonding are highlighted in yellow, whilst the fingerprints not implicated in hydrogen bonding are highlighted in blue. Furthermore, vehicles such as HPC, CMC and MC are polymers and have the potential to form a chelate with the compound, supported by hydrogen bonding which can lower the entropy of the system resulting in less disorder. The effects of hydrogen bonding may aid the solubilisation and/or permeability of a compound and as discussed in chapter 2, an improved solubility/permeability can confer to a reduction in toxicity. Interestingly, many fewer fingerprints which represent branching are used by the DT models, which can potentially hinder the chelating ability of a vehicle towards a compound, reducing the potential of forming hydrogen bonds.

Of the 12 models produced, four MACCS fingerprints and one molecular property descriptor occur in 2 different models. The MACCS fingerprint 109-ACH2O is present in the saline & CMC models and the distilled water + alcohol & CMC models, both of which contain CMC as one of the vehicles being modelled. Similarly 160-CH3 is present in the saline & HPC and saline & MC models, both of which contain saline as one of the vehicles.

Table 6.17. Frequently used fingerprints and descriptors extracted form DT models. Descriptors highlighted in yellow can potentially form hydrogen bonds whilst descriptors in blue are not implicated in hydrogen bonding.

Vehicle Pair	MACCS Key Fingerprints	Molecular Descriptors
Saline & CMC	120-HETEROCYCLIC ATOM >1	Number of heteroatoms in aliphatic rings
	83-QAAAA@1	-
	53-QHAAAQH	-
	109-ACH2O	-
	115-CH3ACH2A	-
	137-HETEROCYCLE	-
Saline & Saline with Tween-80	93-QCH3	-
	116-CH3AACH2A	-
	95-NAAO	-
	97-NAAAO	-
	99-C=C	-
Saline & HPC	102-QO	-
	131-QH > 1	-
	148-AQ(A)A	-
	160-CH3	-
Saline & Water	117-NAO	Number of heteroatoms in aliphatic rings
Saline & MC	160-CH3	-
	114-CH3CH2A	-
	144-Anot%A%Anot%A	-
CMC & HPC	103-CL	-
	152-OC(C)C	-
	155-A!CH2!A	-
CMC & MC	105-A$A($A)$A	-
	125-AROMATIC RING > 1	-
	151-NH	-
	75-A!N$A.	-
HPC & Saline with Tween-80	114-CH3CH2A	-
	150-A!A$A!A	-
	163-6M RING	-
Distilled water + alcohol & Saline	98-QAAAAA@1	-
	166-FRAGMENTS	-
	94-QN	-
Distilled water + alcohol & MC	112-AA(A)(A)A	-
	72-OAAO	-
	79-NAAN	-
Distilled water + alcohol & CMC	156-NA(A)A	-
	16-QAA@1	-
	109-ACH2O	-
	94-QN	-
Acetone & Saline	38-NC(C)N	Number of carbons

94-QN occurs in the distilled water + alcohol & MC and distilled water + alcohol & CMC models, both of which contain distilled water + alcohol as one of the vehicles. Lastly the fingerprint 114-CH3CH2A, occurs in the saline & MC and HPC & saline with Tween-80, however neither of those two models contain a common vehicle. The one molecular property descriptor which occurs in 2 different models is the Number of heteroatoms in aliphatic rings. This descriptor is present in the saline & CMC and saline & water models for which saline is the common vehicle between the two different models.

From the DT information extracted, it was possible to identify descriptors which strongly correlated with the class of a model. This was initially done with one of the distilled water + alcohol & CMC models, which led to the discovery of compounds containing the 3-membered ring aziridine which correlated strongly with the class of the model data. From this information it was clearly apparent that compounds containing the aziridine ring structure were better suited to the vehicle CMC than distilled water + alcohol. This builds on the predictive model work discussed earlier and demonstrates how such information extracted from *in vivo* data can be turned into knowledge.

Furthermore, an analysis of all the other vehicle pairings identified similar strong correlations between one class of a particular model and any aziridine containing compounds. Using this information it was possible to construct a partially ordered set of the vehicles according to the toxicity reduction afforded by them for any aziridine containing compounds. The data extracted also acted as proof of concept for the main branch of the partially ordered set. Although the partially ordered set was constructed from low numbers of data, none of the data resulted in counter arguments to the ordering of the vehicles. Whilst the polymer HPC seems to be the most protective vehicle against the toxicity of aziridine containing compounds, MC was deemed the least protective overall. The vehicle saline appeared to offer better protection than saline with Tween-80, which suggests that the surfactant Tween-80 may reduce the ability of saline to offer any kind of toxicity relief to aziridine containing compounds. The vehicle distilled water + alcohol also rated low in toxicity protection, potentially due to the alcohol content within this vehicle. Interestingly the vehicles which appear more frequently within the entire NIH dataset (HPC, saline, and saline with Tween-80) also seem to be the vehicles which offer the better toxicity relief against aziridine containing compounds. Whilst the vehicles which appear less often within the entire NIH dataset seem to be the ones offering the least toxicity protection against aziridine containing compounds.

7

Thesis Conclusion

The work presented in this thesis is a multidisciplinary effort which involves the use of computational machine learning techniques with real scientific data, for the purpose of data extraction, data visualisation, data modelling, and knowledge discovery.

The main objective of this thesis, as set out in chapter 1, was to demonstrate that drug formulations can be modelled for toxicity reduction. This has been shown in the original contributions made by chapters 5 and 6 of this thesis. Furthermore, at the time of writing, there is no known published work in the public domain of similar credit.

7.1 Research Contributions

Formulating drugs to mitigate toxicity is extremely useful to the pharmaceutical industry and allows NCEs that would originally be dropped due to unacceptable toxicities, to continue through the developmental pipeline. This translates to a saving in cost and time allowing better medicines to reach the end user. Knowing which formulation will provide toxicity relief for a specific drug is an interesting challenge but often requires the input of an expert with many years of experience.

The objective of this thesis was therefore to develop models capable of predicting which formulation would be suited to a particular drug.

To begin with an extensive literature review was required which covered all aspects of drug toxicology relating to formulation work. Chapter 2 reports the information gathered from this review. An overview of the different types of drug toxicities are discussed as well as the

various ways of studying toxicity and taking measurements, such as dose related measurements, biomarker measurements and histopathological measurements. The chapter also reports on the various technologies currently available for toxicity reduction through drug formulation, surprisingly there are relatively few different types.

More importantly it is imperative to understand the underlying pharmacokinetic and pharmacodynamic principles of drug toxicity when considering drug formulations. All drugs will exhibit an underlying pharmacokinetic and pharmacodynamic based toxic response. Understanding these responses allows formulations to counter these effects thereby reducing toxicity. The mechanism(s) by which a formulations reduces a drug's toxicity can be grouped into the following four strategies; 1) active organ protection, 2) avoid susceptible organs, 3) active targeting and 4) availability enhancement.

The majority of the drugs reported in the literature that have been formulated for toxicity reduction were the cytotoxics and anti-infectives. This resulted in a small diversity of drugs which would prove difficult to model.

Nevertheless, if a sufficient number of data can be gathered it is possible to build a toxicity protection pathway based on the AOP framework which plots the upstream and downstream events of a drug's toxicity pathway. An example for the drug gentamicin was provided in section 2.8. Along this pathway, formulations known to prevent specific toxic mechanisms are added. If two different drugs shared a common pathway, then they may in principle benefit from the same formulation to reduce toxicity.

The work presented in chapter 2 provides a good scientific overview of drug formulations and toxicity which has relevance to all subsequent chapters.

Chapter 3 of this thesis presents an overview of how predictive models are built and the processes undertaken, from data collection through to model validation. This chapter discusses some of the pitfalls encountered with this type of work and provides useful information with regards to data curation and more importantly chemical data representation. The chapter goes on to discuss some of the more popular machine learning techniques used for binary classification problems (ANN, DT, RF, SVM, kNN and Naïve Bayes) in reference to toxicity studies, which will also support the soft computing approaches explored later in the thesis. References to the literature are provided where necessary. The chapter ends by reporting on the different ways models can be statistically analysed using accuracy and precision metrics, with the balanced accuracy or Matthews

Correlation Coefficient (MCC) favoured with unbalanced datasets. The work presented in this chapter is directly relevant to chapters 5 and 6 of this thesis.

Chapter 4 proposes a methodology based on graph theory and Venn diagram representation for the extraction of interesting data relationships that can contribute to a knowledge-based approach of modelling toxicity in drug-vehicle relationships. The methodology presented can in theory be used for other types of data where the relationship between two variables is sought. Using the NIH dataset, which was largely un-curated, a novel bipartite graph arrangement using labelled edges was shown to provide the optimal solution to the problem of visualising and extracting interesting drug-vehicle relationships. Subgraphs were extracted to demonstrate the capabilities of visualising such data using this arrangement. From the graph paired comparisons between vehicles were made and each drug grouped into one or more list. After removing conflicting or equivalent data these lists were plotted onto a Venn diagram from which multiple vehicle relationships can now be identified for drugs in the overlapping segments of the Venn diagram. This methodology helps to quickly identify the number of interesting datum points a particular dataset may contain.

Chapter 5 proposes a novel methodology based on the area under a curve principle that uses interpolation and extrapolation methods to classify data based on multiple data points. The methodology is applied to the NIH dataset for the pairwise classification of drug-vehicle relationships, from which binary classification models are built (RF and DT). To investing the effectiveness of this methodology, three % AUC thresholds were used with three different interpolation/extrapolation methods. This methodology along with some of the results from chapter 6, contributes to original work in this field and has been published in Springer, Soft Computing (doi:10.1007/s00500-015-1925-9) see publications section for details.

Chapter 6 presents the results of the various models created from the 12 vehicle pairings extracted from the NIH dataset using the methodology proposed in chapter 5. Potentially 9 RF and 9 DT models could be produced for each vehicle pairing, although this is dependent on sufficient data being available. Models that used a 60 % AUC threshold with interpolation only generally seemed to perform better, suggesting a 60 % threshold is required for any signal to be observed amongst the noise in the dataset.

These findings provide some reassurance with regards to the methodology developed and shows which parameters produce the better results for this particular dataset. For each of the vehicle pairings the best accuracy produced is always above 70 % with some models resulting in accuracies of 80 % or more.

Descriptor information extracted from the DT models shows that the decision trees were using fingerprints that have the potential to form hydrogen bonds much more frequently than fingerprints that do not. Interestingly vehicles such as HPC, CMC and MC have the ability to form chelates with compounds which are supported by hydrogen bonding, which lowers the entropy of the system making a drug-vehicle relationship more stable.

Of all the models developed, one vehicle pairing, distilled water + alcohol & CMC, resulted in the class of the data strongly correlating with a single descriptor. This descriptor was identified as a 3-membered ring descriptor. Inspection of the structures revealed this 3-membered ring to be aziridine. Extracting data from the other 11 vehicle pairing models revealed a similar pattern, with aziridine containing compound generally correlating strongly with one class of the model.

The vehicle pairings of these models could thus be arranged into a sequence of partially ordered sets which provided an ordering of vehicles which progressively provide better toxicity protection for aziridine containing compounds.

Thus using the methodology proposed in chapter 5 it was possible to build successful classification models from which interesting relationships could be formed and previously unknown knowledge, pertaining to aziridine containing compounds extracted.

To summarise, the possibility of building toxicity protection pathways as shown for gentamicin in chapter 2 for the prediction of drug formulations is an interesting one and constitutes a knowledge based approach, which can be time consuming but also lacking in sufficient data. Alternatively it is also possible to build predictive models using a data driven approach as demonstrated by the work in chapters 5 and 6 of this thesis. From this it has also been possible to derive rules for groups of compounds.

7.2 Future work

Given the novelty of some of the work presented in this thesis, there are many future directions that are possible and worth considering. From the work presented in chapter 2 of this thesis there is the potential to build toxicity protection pathways for other drugs and

assess if any integration of pathways are possible. If so this may provide another unique direction into what is considered to be a relatively unexplored domain. The potential to integrate information from publicly available data such as the AOP Wiki page [349] is also something to be considered. The literature can also be searched for more drug formulation data that have become available since the initial collection of data.

In chapter 4 of this thesis, bipartite graph representation was introduced and a methodology discussed for the quick visualisation and extraction of interesting relationships. At present these relationships were extracted in the form of '"better than" according to the binning values. If however bins are not used but replaced with continuous values then the comparisons become quantifiable. Comparisons are not simply better or worse; rather they can be quantified to give a measure of how much better something is. Using this approach allows ordering of data which adds another dimensionality to the methodology.

Of the drug-vehicle classification models presented in chapters 5 and 6 of this thesis, there is scope for model refinement as well as expert model interpretations. For instance would replacing the MACCS fingerprints in the model with a large number of alternative descriptors that are objectively screened improve the results; or would using alternative models such as SVMs or ANN produce better predictions? The current approach of classification may also be extended to non-linear dataset preparations which may potentially generate more data points to model with or describe the datum more accurately.

Model refinement is a worthwhile approach; however, alternative ways to handle noisy data with low concordance may be a better approach. Given the models developed in chapters 5 and 6 resulted in their most accurate predictions at the 60 % AUC threshold, this is suggestive of a noisy dataset with low concordance and ways of handling such data could be a better starting point. Many articles in the literature report on methods of handling noisy data for linear and non-linear data [350, 351]. An alternative solution could be based on the use of fuzzy numbers to represent the data in order to explore a robust approach to the noise invariably present in such data collected from multiple experiments.

To assess the robustness of the AUC methodology proposed an alternative dataset could be used. Moreover since the methodology is not specific to toxicity data it could be tried with

other types of data where two functional variables are to be compared, microarray data or cosmetics data are two such alternatives.

From a scientific knowledge perspective it would be interesting to explore the models developed to see if other drug structures with specific functional groups show strong correlations with any descriptors. This was demonstrated for the aziridine containing compounds in the work reported in chapter 6, from which it was possible to create partially ordered sets and derive rules.

Lastly, the findings of the DT models, showed a preference for hydrogen bonding fingerprints overwhelmingly over non-hydrogen bonding fingerprints. Discussing these data and the aziridine ring structure data with a formulation expert, may provide interpretations that cannot be derived from the model descriptors. Predictive models tend to reveal correlation but not causation unless the correlation can be scientifically interpreted. For example, it is understood that aziridines are reactive alkylating agents that may react by ring opening [352]. Could it be possible therefore that the vehicles which reduce the toxicity of aziridine ring containing compounds do so by attenuating ring opening?

References

1. Kola, I., and Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? Nature Reviews Drug Discovery *3*, 711-715.

2. Basavaraj, S., and Betageri, G.V. (2014). Can formulation and drug delivery reduce attrition during drug discovery and development—review of feasibility benefits and challenges. Acta Pharmaceutica Sinica B *4*, 3 - 17.

3. Arrowsmith, J., and Miller, P. (2013). Trial Watch: Phase II and Phase III attrition rates 2011-2012. *12*, 569.

4. DiMasi, J.A., Hansen, R.W., and Grabowski, H.G. (2003). The price of innovation: new estimates of drug development costs. Journal of Health Economics *22*, 151-185.

5. Biswa Mohan, S., Dinda S.C., Kumar B.V.V., R., and J.R., P. (2012). Computational Approaches for Drug Design and Discovery Process. Journal of Current Pharma Research *2*, 600-611.

6. Light, D.W., and Warburton, R. (2011). Demythologizing the high costs of pharmaceutical research. Biosocieties *6*, 34-50.

7. E.N Bharath, S.N Manjula, and Vijaychand, A. (2011). In Silico Drug Design-Tool For Overcoming The Innovation Deficit In the Drug Discovery Process. International Journal of Pharmacy and Pharmaceutical Sciences *3*, 8-12.

8. Szejtli, J. (2005). Cyclodextrin complexed generic drugs are generally not bio-equivalent with the reference products: Therefore the increase in number of marketed drug/cyclodextrin formulations is so slow. Journal of Inclusion Phenomena and Macrocyclic Chemistry *52*, 1-11.

9. Barenholz, Y. (2012). Doxil (R) - The first FDA-approved nano-drug: Lessons learned. Journal of Controlled Release *160*, 117-134.

10. Gupta, S., Kesarla, R., and Omri, A. (2013). Formulation strategies to improve the bioavailability of poorly absorbed drugs with special emphasis on self-emulsifying systems. ISRN pharmaceutics *2013*, 848043-848043.

11. Leuner, C., and Dressman, J. (2000). Improving drug solubility for oral delivery using solid dispersions. European Journal of Pharmaceutics and Biopharmaceutics *50*, 47-60.

12. Mennini, N., Furlanetto, S., Bragagni, M., Ghelardini, C., Mannelli, L.D.C., and Mura, P. (2014). Development of a chitosan-derivative micellar formulation to improve celecoxib solubility and bioavailability. Drug Development and Industrial Pharmacy *40*, 1494-1502.

13. Wasan, E.K., Bartlett, K., Gershkovich, P., Sivak, O., Banno, B., Wong, Z., Gagnon, J., Gates, B., Leon, C.G., and Wasan, K.M. (2009). Development and characterization of oral lipid-based Amphotericin B formulations with enhanced drug solubility, stability and antifungal activity in rats infected with Aspergillus fumigatus or Candida albicans. International Journal of Pharmaceutics *372*, 76-84.

14. Savjani, K.T., Gajjar, A.K., and Savjani, J.K. (2012). Drug solubility: importance and enhancement techniques. ISRN pharmaceutics *2012*, 195727-195727.

15. Zhou, S., Li, G.-B., Huang, L.-Y., Xie, H.-Z., Zhao, Y.-L., Chen, Y.-Z., Li, L.-L., and Yang, S.-Y. (2014). A prediction model of drug-induced ototoxicity developed by an optimal support vector machine (SVM) method. Computers in Biology and Medicine *51*, 122-127.

16. Greene, N., Fisk, L., Naven, R.T., Note, R.R., Patel, M.L., and Pelletier, D.J. (2010). Developing Structure-Activity Relationships for the Prediction of Hepatotoxicity. Chemical Research in Toxicology *23*, 1215-1222.

17. Cheng, A., and Dixon, S.L. (2003). In silico models for the prediction of dose-dependent human hepatotoxicity. Journal of Computer-Aided Molecular Design *17*, 811-823.

18. Low, Y., Uehara, T., Minowa, Y., Yamada, H., Ohno, Y., Urushidani, T., Sedykh, A., Muratov, E., Kuz'min, V., Fourches, D., et al. (2011). Predicting Drug-Induced

Hepatotoxicity Using QSAR and Toxicogenomics Approaches. Chemical Research in Toxicology *24*, 1251-1262.

19. Su, R., Li, Y., Zink, D., and Loo, L.-H. (2014). Supervised prediction of drug-induced nephrotoxicity based on interleukin-6 and-8 expression levels. Bmc Bioinformatics *15*.

20. Development Therapeutics Program (DTP) http://dtp.nci.nih.gov/.

21. NC3Rs. (https://www.nc3rs.org.uk/).

22. Eisenbrand, G., Pool-Zobel, B., Baker, V., Balls, M., Blaauboer, B.J., Boobis, A., Carere, A., Kevekordes, S., Lhuguenot, J.C., Pieters, R., et al. (2002). Methods of in vitro toxicology. Food and Chemical Toxicology *40*, 193-236.

23. Pramyothin, P., Ngamtin, C., Poungshompoo, S., and Chaichantipyuth, C. (2007). Hepatoprotective activity of Phyllanthus amarus Schum. et. Thonn. extract in ethanol treated rats: In vitro and in vivo studies. Journal of Ethnopharmacology *114*, 169-173.

24. Lincopan, N., Mamizuka, E.M., and Carmona-Ribeiro, A.M. (2005). Low nephrotoxicity of an effective amphotericin B formulation with cationic bilayer fragments. Journal of Antimicrobial Chemotherapy *55*, 727-734.

25. Ekwall, B., Silano, V., Paganuzzi-Stammati, A., and Zucco, F. (1990). Short-term Toxicity Tests for Non-genotoxic Effects.

26. Vandermeulen, G., Rouxhet, L., Arien, A., Brewster, M.E., and Preat, V. (2006). Encapsulation of amphotericin B in poly(ethylene glycol)-block-poly(epsilon-caprolactone-co-trimethylenecarbonate) polymeric micelles. International Journal of Pharmaceutics *309*, 234-240.

27. Xiong, Y., Guo, D., Wang, L., Zheng, X., Zhang, Y., and Chen, J. (2009). Development of nobiliside A loaded liposomal formulation using response surface methodology. International Journal of Pharmaceutics *371*, 197-203.

28. Gupta, U., Dwivedi, S.K.D., Bid, H.K., Konwar, R., and Jain, N.K. (2010). Ligand anchored dendrimers based nanoconstructs for effective targeting to cancer cells. International Journal of Pharmaceutics *393*, 185-196.

29. Jung, S.H., Lim, D.H., Jung, S.H., Lee, J.E., Jeong, K.-S., Seong, H., and Shin, B.C. (2009). Amphotericin B-entrapping lipid nanoparticles and their in vitro and in vivo characteristics. European Journal of Pharmaceutical Sciences *37*, 313-320.

30. Pourgholami, M.H., Wangoo, K.T., and Morris, D.L. (2008). Albendazole-Cyclodextrin Complex: Enhanced Cytotoxicity in Ovarian Cancer Cells. Anticancer Research *28*, 2775-2779.

31. Khan, M., Varadharaj, S., Shobha, J.C., Naidu, M.U., Parinandi, N.L., Kutala, V.K., and Kuppusamy, P. (2006). C-phycocyanin ameliorates doxorubicin-induced oxidative stress and apoptosis in adult rat cardiomyocytes. Journal of Cardiovascular Pharmacology *47*, 9-20.

32. S. Palani, S.R., S.Karthi, Selvi Archana, B. Senthil Kumar (2010). In vivo analysis of nephro & hepato protective effects and antioxidant activity of Madhuca longifolia against acetaminophen-induced toxicity & oxidative stress. JPR: BioMedRx: An International Journal *3*, 9 - 16.

33. Maliakel, D.M., Kagiya, T.V., and Nair, C.K.K. (2008). Prevention of cisplatin-induced nephrotoxicity by glucosides of ascorbic acid and alpha-tocopherol. Experimental and Toxicologic Pathology *60*, 521-527.

34. Pereverzeva, E., Treschalin, I., Bodyagin, D., Maksimenko, O., Langer, K., Dreis, S., Asmussen, B., Kreuter, J., and Gelperina, S. (2007). Influence of the formulation on the tolerance profile of nanoparticle-bound doxorubicin in healthy rats: Focus on cardio- and testicular toxicity. International Journal of Pharmaceutics *337*, 346-356.

35. Garg, M., Dutta, T., and Jain, N.K. (2007). Reduced hepatic toxicity, enhanced cellular uptake and altered pharmacokinetics of stavudine loaded galactosylated liposomes. European Journal of Pharmaceutics and Biopharmaceutics *67*, 76-85.

36. Sheikh, S., Ali, S.M., Ahmad, M.U., Ahmad, A., Mushtaq, M., Paithankar, M., Mandal, J., Saptarishi, D., Sehgal, A., Maheshwari, K., et al. (2010). Nanosomal

Amphotericin B is an efficacious alternative to Ambisome (R) for fungal therapy. International Journal of Pharmaceutics *397*, 103-108.

37. Lauer, B., Tuschl, G., Kling, M., and Mueller, S.O. (2009). Species-specific toxicity of diclofenac and troglitazone in primary human and rat hepatocytes. Chemico-Biological Interactions *179*, 17-24.

38. Singh, B., and Gupta, R.S. (1985). Species-specific differences in the toxicity and mutagenicity of the anticancer drugs mithramycin, chromomycin A3, and olivomycin. Cancer Research *45*, 2813-2820.

39. Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. Regulatory Toxicology and Pharmacology *32*, 56-67.

40. ICH (2009). Guidance on nonclinical safety for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3(R2). (http://www.ich.org/home.html).

41. OECD (2008). Test Guideline 407. Repeated Dose 28-day Oral Toxicity Study in Rodents. http://www.oecd.org/.

42. OECD (1998). Test Guideline 408. Repeated Dose 90-day Oral Toxicity Study in Rodents. http://www.oecd.org/.

43. OECD (1998). Test Guideline 409. Repeated Dose 90-day Oral Toxicity Study in Non-Rodents. http://www.oecd.org/.

44. OECD (2001). Test Guideline 420. Acute Oral Toxicity – Fixed Dose Procedure. http://www.oecd.org/.

45. OECD (2001). Test Guideline 423. Acute Oral Toxicity – Acute Toxic Class Method. http://www.oecd.org/.

46. OECD (2001). Test Guideline 425. Acute Oral Toxicity – Up-and-Down Procedure. http://www.oecd.org/.

47. OECD (2009). Test Guideline 452. Chronic Toxicity Studies. http://www.oecd.org/.

48. OECD (2009). Test Guidline 412. Subacute Inhalation Toxicity 28-day Study. http://www.oecd.org/.

49. OECD (2009). Test Guidline 413. Subchronic Inhalation Toxicity 90-day Study. http://www.oecd.org/.

50. OECD (2010). Test Guidline 417. Toxicokinetics. http://www.oecd.org/.

51. Vanetten, E.W.M., Vanvianen, W., Tijhuis, R.H.G., Storm, G., and Bakkerwoudenberg, I. (1995). Sterically stabilized amphotericin B-liposomes: toxicity and biodistribution in mice. Journal of Controlled Release *37*, 123-129.

52. Kaminskas, L.M., McLeod, V.M., Kelly, B.D., Cullinane, C., Sberna, G., Williamson, M., Boyd, B.J., Owen, D.J., and Porter, C.J.H. (2012). Doxorubicin-Conjugated PEGylated Dendrimers Show Similar Tumoricidal Activity but Lower Systemic Toxicity When Compared to PEGylated Liposome and Solution Formulations in Mouse and Rat Tumor Models. Molecular Pharmaceutics *9*, 422-432.

53. Krasko, M.Y., Golenser, J., Nyska, A., Nyska, M., Brin, Y.S., and Domb, A.J. (2007). Gentamicin extended release from an injectable polymeric implant. Journal of Controlled Release *117*, 90-96.

54. Xiong, R., Lu, W., Li, J., Wang, P., Xu, R., and Chen, T. (2008). Preparation and characterization of intravenously injectable nimodipine nanosuspension. International Journal of Pharmaceutics *350*, 338-343.

55. Nakamagoe, M., Tabuchi, K., Uemaetomari, I., Nishimura, B., and Hara, A. (2010). Estradiol protects the cochlea against gentamicin ototoxicity through inhibition of the JNK pathway. Hearing Research *261*, 67-74.

56. Lynch, E.D., Gu, R.D., Pierce, C., and Kil, J. (2005). Reduction of acute cisplatin ototoxicity and nephrotoxicity in rats by oral administration of allopurinol and ebselen. Hearing Research *201*, 81-89.

57. Yao, L., Zhao, X., Li, Q., Zu, Y., Fu, Y., Zu, B., Meng, X., and Liu, C. (2012). In vitro and in vivo evaluation of camptothecin nanosuspension: A novel formulation with high antitumor efficacy and low toxicity. International Journal of Pharmaceutics *423*, 586-588.

58. Srdjenovic, B., Milic-Torres, V., Grujic, N., Stankov, K., Djordjevic, A., and Vasovic, V. (2010). Antioxidant properties of fullerenol C-60(OH)(24) in rat kidneys, testes, and lungs treated with doxorubicin. Toxicology Mechanisms and Methods *20*, 298-305.

59. Pereverzeva, E., Treschalin, I., Bodyagin, D., Maksimenko, O., Kreuter, J., and Gelperina, S. (2008). Intravenous tolerance of a nanoparticle-based formulation of doxorubicin in healthy rats. Toxicology Letters *178*, 9-19.

60. Zou, Y.Y., Priebe, W., Stephens, L.C., and PerezSoler, R. (1995). Preclinical toxicity of liposome-incorporated annamycin: Selective bone marrow toxicity with lack of cardiotoxicity. Clinical Cancer Research *1*, 1369-1374.

61. Soyez, H., Schacht, E., Jelinkova, M., and Rihova, B. (1997). Biological evaluation of mitomycin C bound to a biodegradable polymeric carrier. Journal of Controlled Release *47*, 71-80.

62. Mimoso, I.M., Francisco, A.P.G., and Cruz, M.E.M. (1997). Liposomal formulation of netilmicin. International Journal of Pharmaceutics *147*, 109-117.

63. Anadon, A., Castellano, V., and Rosa Martinez-Larranaga, M. (2014). Biomarkers of drug toxicity.

64. Das, S., Roy, P., Auddy, R.G., and Mukherjee, A. (2011). Silymarin nanoparticle prevents paracetamol-induced hepatotoxicity. International Journal of Nanomedicine *6*, 1291-1301.

65. Bhattacharyya, S., and Mehta, P. (2012). The hepatoprotective potential of Spirulina and vitamin C supplemention in cisplatin toxicity. Food & Function *3*, 164-169.

66. Bhattacharjee, R., and Sil, P.C. (2006). The protein fraction of Phyllanthus niruri plays a protective role against acetaminophen induced hepatic disorder via its antioxidant properties. Phytotherapy Research *20*, 595-601.

67. Cremers, H.F.M., Verrijk, R., Bayon, L.G., Wesseling, M.M., Wondergem, J., Heuff, G., Meijer, S., Kwon, G.S., Bae, Y.H., Kim, S.W., et al. (1995). Improved distribution and reduced toxicity of adriamycin bound to albumin-heparin microspheres. International Journal of Pharmaceutics *120*, 51-61.

68. Amin, Z.A., Bilgen, M., Alshawsh, M.A., Ali, H.M., Hadi, A.H.A., and Abdulla, M.A. (2012). Protective Role of Phyllanthus niruri Extract against Thioacetamide-Induced Liver Cirrhosis in Rat Model. Evidence-Based Complementary and Alternative Medicine.

69. Yoon, M.Y., Kim, S.J., Lee, B.-H., Chung, J.-H., and Kim, Y.C. (2006). Effects of dimethylsulfoxide on metabolism and toxicity of acetaminophen in mice. Biological & Pharmaceutical Bulletin *29*, 1618-1624.

70. Bonventre, J.V., and Yang, L. (2011). Cellular pathophysiology of ischemic acute kidney injury. Journal of Clinical Investigation *121*, 4210-4221.

71. Campion, S., Aubrecht, J., Boekelheide, K., Brewster, D.W., Vaidya, V.S., Anderson, L., Burt, D., Dere, E., Hwang, K., Pacheco, S., et al. (2013). The current status of biomarkers for predicting toxicity. Expert Opinion on Drug Metabolism & Toxicology *9*, 1391-1408.

72. Khan, M., Shobha, J.C., Mohan, I.K., Naidu, M.U.R., Prayag, A., and Kutala, V.K. (2006). Spirulina attenuates cyclosporine-induced nephrotoxicity in rats. Journal of Applied Toxicology *26*, 444-451.

73. Italia, J.L., Bhatt, D.K., Bhardwaj, V., Tikoo, K., and Kumar, M.N.V.R. (2007). PLGA nanoparticles for oral delivery of cyclosporine: Nephrotoxicity and pharmacokinetic studies in comparison to Sandimmune Neoral (R). Journal of Controlled Release *119*, 197-206.

74. Sirota, J.C., Klawitter, J., and Edelstein, C.L. (2011). Biomarkers of acute kidney injury. Journal of toxicology *2011*, 328120-328120.

75. Pedraza-Chaverri, J., Gonzalez-Orozco, A.E., Maldonado, P.D., Barrera, D., Medina-Campos, O.N., and Hernandez-Pando, R. (2003). Diallyl disulfide ameliorates gentamicin-induced oxidative stress and nephropathy in rats. European Journal of Pharmacology *473*, 71-78.

76. Maldonado, P.D., Barrera, D., Rivero, I., Mata, R., Medina-Campos, O.N., Hernandez-Pando, R., and Pedraza-Chaverri, J. (2003). Antioxidant S-allylcysteine prevents gentamicin-induced oxidative stress and renal damage. Free Radical Biology and Medicine *35*, 317-324.

77. Nassar, I., Pasupati, T., Judson, J.P., and Segarra, I. (2010). Histopathological study of the hepatic and renal toxicity associated with the co-administration of imatinib and acetaminophen in a preclinical mouse model. The Malaysian journal of pathology *32*, 1-11.

78. Welling, P.G. (1995). Differences between pharmacokinetics and toxicokinetics. Toxicologic Pathology *23*, 143-147.

79. Lin, J.H., and Lu, A.Y.H. (1997). Role of pharmacokinetics and metabolism in drug discovery and development. Pharmacological Reviews *49*, 403-449.

80. Kim, S.C., Kim, D.W., Shim, Y.H., Bang, J.S., Oh, H.S., Kim, S.W., and Seo, M.H. (2001). In vivo evaluation of polymeric micellar paclitaxel formulation: toxicity and efficacy. Journal of Controlled Release *72*, 191-202.

81. Le Garrec, D., Gori, S., Luo, L., Lessard, D., Smith, D.C., Yessine, M.A., Ranger, M., and Leroux, J.C. (2004). Poly(N-vinylpyrrolidone)-block-poly(D,L-lactide) as a new polymeric solubilizer for hydrophobic anticancer drugs: in vitro and in vivo evaluation. Journal of Controlled Release *99*, 83-101.

82. Ploemen, J.-P.H.T.M., Kramer, H., Krajnc, E.I., and Martin, I. (2007). The use of toxicokinetic data in preclinical safety assessment: A toxicologic pathologist perspective. Toxicologic Pathology *35*, 834-837.

83. Gabizon, A., Shiota, R., and Papahadjopoulos, D. (1989). Pharmacokinetics and Tissue Distribution of Doxorubicin Encapsulated in Stable Liposomes With Long Circulation Times. Journal of the National Cancer Institute *81*, 1484-1488.

84. Cao, N., and Feng, S.-S. (2008). Doxorubicin conjugated to D-alpha-tocopheryl polyethylene glycol 1000 succinate (TPGS): Conjugation chemistry, characterization, in vitro and in vivo evaluation. Biomaterials *29*, 3856-3865.

85. Chaturvedi, P.R., Decker, C.J., and Odinecs, A. (2001). Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. Current Opinion in Chemical Biology *5*, 452-463.

86. Jambhekar, S.S., and Breen, P.J. (2009). Basic Pharmacokinetics, (Pharmaceutical Press).

87. James, L.P., Mayeux, P.R., and Hinson, J.A. (2003). Acetaminophen-Induced Hepatotoxicity. Drug Metabolism and Disposition *31*, 1499-1506.

88. Hinson, J.A., Roberts, D.W., and James, L.P. (2010). Mechanisms of acetaminophen-induced liver necrosis. Handbook of experimental pharmacology, 369-405.

89. Thummel, K.E., Lee, C.A., Kunze, K.L., Nelson, S.D., and Slattery, J.T. (1993). Oxidation of acetaminophen to N-acetyl-p-aminobenzoquinone imine by Human CYP3A4. Biochemical Pharmacology *45*, 1563-1569.

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

91. OECD (2013). Guidance document and a template for developing and assessing adverse outcome pathways. http://www.oecd.org/.

92. Allen, T.E.H., Goodman, J.M., Gutsell, S., and Russell, P.J. (2014). Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. Chemical Research in Toxicology *27*, 2100-2112.

93. Groh, K.J., Carvalho, R.N., Chipman, J.K., Denslow, N.D., Halder, M., Murphy, C.A., Roelofs, D., Rolaki, A., Schirmer, K., and Watanabe, K.H. (2015). Development and application of the adverse outcome pathway framework for understanding and predicting chronic toxicity: I. Challenges and research needs in ecotoxicology. Chemosphere *120*, 764-777.

94. Miller, R.P., Tadagavadi, R.K., Ramesh, G., and Reeves, W.B. (2010). Mechanisms of Cisplatin Nephrotoxicity. Toxins *2*, 2490-2518.

95. Naughton, C.A. (2008). Drug-induced nephrotoxicity. American Family Physician *78*, 743-750.

96. Pabla, N., and Dong, Z. (2008). Cisplatin nephrotoxicity: Mechanisms and renoprotective strategies. Kidney International *73*, 994-1007.

97. Rybak, L.P., Whitworth, C.A., Mukherjea, D., and Rarakumar, V. (2007). Mechanisms of cisplatin-induced ototoxicity and prevention. Hearing Research *226*, 157-167.

98. Grill, M.F., and Maganti, R.K. (2011). Neurotoxic effects associated with antibiotic use: management considerations. British Journal of Clinical Pharmacology *72*, 381-393.

99. Luft, F.C., Yum, M.N., Walker, P.D., and Kleit, S.A. (1977). Gentamicin Gradient Patterns and Morphological Changes in Human Kidneys. Nephron *18*, 167-174.

100. Sundin, D.P., Sandoval, R., and Molitoris, B.A. (2001). Gentamicin inhibits renal protein and phospholipid metabolism in rats: Implications involving intracellular trafficking. Journal of the American Society of Nephrology *12*, 114-123.

101. Lipsky, J.J., Cheng, L., Sacktor, B., and Lietman, P.S. (1980). Gentamicin uptake by renal tubule brush border membrane vesicles. Journal of Pharmacology and Experimental Therapeutics *215*, 390-393.

102. Lee, N.-H., Seo, C.-S., Lee, H.-Y., Jung, D.-Y., Lee, J.-K., Lee, J.-A., Song, K.Y., Shin, H.-K., Lee, M.-Y., Seo, Y.B., et al. (2012). Hepatoprotective and Antioxidative Activities of Cornus officinalis against Acetaminophen-Induced Hepatotoxicity in Mice. Evidence-Based Complementary and Alternative Medicine.

103. Piao, H., Kamiya, N., Watanabe, J., Yokoyama, H., Hirata, A., Fujii, T., Shimizu, I., Ito, S., and Goto, M. (2006). Oral delivery of diclofenac sodium using a novel solid-in-oil suspension. International Journal of Pharmaceutics *313*, 159-162.

104. Dhana Lekshmi, U.M., Poovi, G., Kishore, N., and Reddy, P.N. (2010). In vitro characterization and invivo toxicity study of repaglinide loaded poly (methyl methacrylate) nanoparticles. International journal of pharmaceutics *396*, 194-203.

105. Bangham, A.D., and Horne, R.W. (1964). Negative staining of phospholipids and their structural modification by surface-active agents as observed in the electron microscope. Journal of molecular biology *8*, 660-668.

106. Mufamadi, M.S., Pillay, V., Choonara, Y.E., Du Toit, L.C., Modi, G., Naidoo, D., and Ndesendo, V.M.K. (2011). A review on composite liposomal technologies for specialized drug delivery. Journal of drug delivery *2011*, 939851-939851.

107. Ulrich, A.S. (2002). Biophysical aspects of using liposomes as delivery vehicles. Bioscience Reports *22*, 129-150.

108. Akbarzadeh, A., Rezaei-Sadabady, R., Davaran, S., Joo, S.W., Zarghami, N., Hanifehpour, Y., Samiei, M., Kouhi, M., and Nejati-Koshki, K. (2013). Liposome: classification, preparation, and applications. Nanoscale Research Letters *8*.

109. Immordino, M.L., Dosio, F., and Cattel, L. (2006). Stealth liposomes: review of the basic science, rationale, and clinical applications, existing and potential. International Journal of Nanomedicine *1*, 297-315.

110. Patil, Y.P., and Jadhav, S. (2014). Novel methods for liposome preparation. Chemistry and Physics of Lipids *177*, 8-18.

111. Sharma, A., and Sharma, U.S. (1997). Liposomes in drug delivery: progress and limitations. International Journal of Pharmaceutics *154*, 123-140.

112. Tyagi, S., Sharma, P.K., and Malviya, R. (2015). Advancement and Patents on Liposomal Drug Delivery. Global Journal of Pharmacology *9*, 116 - 173.

113. Torchilin, V.P. (2005). Recent advances with liposomes as pharmaceutical carriers. Nature Reviews Drug Discovery *4*, 145-160.

114. Lo, Y.L., Tsai, J.C., and Kuo, J.H. (2004). Liposomes and disaccharides as carriers in spray-dried powder formulations of superoxide dismutase. Journal of Controlled Release *94*, 259-272.

115. Iwanaga, K., Ono, S., Narioka, K., Kakemi, M., Morimoto, K., Yamashita, S., Namba, Y., and Oku, N. (1999). Application of surface coated liposomes for oral delivery of peptide: Effects of coating the liposome's surface on the GI transit of insulin. Journal of Pharmaceutical Sciences *88*, 248-252.

116. Gaspar, M.M., PerezSoler, R., and Cruz, M.E.M. (1996). Biological characterization of L-asparaginase liposomal formulations. Cancer Chemotherapy and Pharmacology *38*, 373-377.

117. Cevc, G. (2004). Lipid vesicles and other colloids as drug carriers on the skin. Advanced Drug Delivery Reviews *56*, 675-711.

118. Felgner, J.H., Kumar, R., Sridhar, C.N., Wheeler, C.J., Tsai, Y.J., Border, R., Ramsey, P., Martin, M., and Felgner, P.L. (1994). Enhanced gene delivery and mechanism

studies with a novel series of cationic lipid formulations. Journal of Biological Chemistry *269*, 2550-2561.

119. Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., and Danielsen, M. (1987). Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proceedings of the National Academy of Sciences of the United States of America *84*, 7413-7417.

120. Liu, Y., Liggitt, D., Zhong, W., Tu, G.H., Gaensler, K., and Debs, R. (1995). Cationic Liposome-mediated Intravenous Gene Delivery. Journal of Biological Chemistry *270*, 24864-24870.

121. Wagner, A., and Vorauer-Uhl, K. (2011). Liposome technology for industrial purposes. Journal of drug delivery *2011*, 591325-591325.

122. Szoka, F., and Papahadjopoulos, D. (1978). Procedure for preparation of liposomes with large internal aqueous space and high capture by reverse-phase evaporation. Proceedings of the National Academy of Sciences of the United States of America *75*, 4194-4198.

123. Kuang, Y., Liu, J., Liu, Z., and Zhuo, R. (2012). Cholesterol-based anionic long-circulating cisplatin liposomes with reduced renal toxicity. Biomaterials *33*, 1596-1606.

124. Yukihara, M., Ito, K., Tanoue, O., Goto, K., Matsushita, T., Matsumoto, Y., Masuda, M., Kimura, S., and Ueoka, R. (2011). Effective Drug Delivery System for Duchenne Muscular Dystrophy Using Hybrid Liposomes Including Gentamicin along with Reduced Toxicity. Biological & Pharmaceutical Bulletin *34*, 712-716.

125. Allen, T.M., Hansen, C., Martin, F., Redemann, C., and Yauyoung, A. (1991). Liposomes containing synthetic lipid derivatives of poly(ethylene glycol) show prolonged circulation half-lives in vivo. Biochimica Et Biophysica Acta *1066*, 29-36.

126. Drummond, D.C., Meyer, O., Hong, K.L., Kirpotin, D.B., and Papahadjopoulos, D. (1999). Optimizing liposomes for delivery of chemotherapeutic agents to solid tumors. Pharmacological Reviews *51*, 691-743.

127. Li, X., Ding, L., Xu, Y., Wang, Y., and Ping, Q. (2009). Targeted delivery of doxorubicin using stealth liposomes modified with transferrin. International Journal of Pharmaceutics *373*, 116-123.

128. Yoo, H.S., and Park, T.G. (2004). Folate receptor targeted biodegradable polymeric doxorubicin micelles. Journal of Controlled Release *96*, 273-283.

129. Wang, H., Zhao, P., Liang, X., Gong, X., Song, T., Niu, R., and Chang, J. (2010). Folate-PEG coated cationic modified chitosan - Cholesterol liposomes for tumor-targeted drug delivery. Biomaterials *31*, 4129-4138.

130. Luft, F.C., Bloch, R., Sloan, R.S., Yum, M.N., Costello, R., and Maxwell, D.R. (1978). Comparative nephrotoxicity of aminoglycoside antibiotics in rats. Journal of Infectious Diseases *138*, 541-545.

131. Hottendorf, G.H., and Williams, P.D. (1986). Aminoglycoside Nephrotoxicity. Toxicologic Pathology *14*, 66-72.

132. Martinez-Salgado, C., Lopez-Hernandez, F.J., and Lopez-Novoa, J.M. (2007). Glomerular nephrotoxicity of aminoglycosides. Toxicology and Applied Pharmacology *223*, 86-98.

133. Mingeot-Leclercq, M.P., and Tulkens, P.M. (1999). Aminoglycosides: Nephrotoxicity. Antimicrobial Agents and Chemotherapy *43*, 1003-1012.

134. Sastrasinh, M., Knauss, T.C., Weinberg, J.M., and Humes, H.D. (1982). Identification of the aminoglycoside binding site in rat renal brush border membranes. Journal of Pharmacology and Experimental Therapeutics *222*, 350-358.

135. Swenson, C.E., Stewart, K.A., Hammett, J.L., Fitzsimmons, W.E., and Ginsberg, R.S. (1990). Pharmacokinetics and in vivo activity of liposome-encapsulated gentamicin. Antimicrobial Agents and Chemotherapy *34*, 235-240.

136. Bermudez, L.E., Yauyoung, A.O., Lin, J.P., Cogger, J., and Young, L.S. (1990). Treatment of disseminated Mycobacterium avium complex infection of beige mice with liposome-encapsulated aminoglycosides. Journal of Infectious Diseases *161*, 1262-1268.

137. Morgan, J.R., and Williams, K.E. (1980). Preparation and properties of liposome-associated gentamicin. Antimicrobial Agents and Chemotherapy *17*, 544-548.

138. Parekh, A.C., and Dave, C.V. (1976). Hamycin: Separation of toxicity and antifungal activity in mice. Life Sciences *19*, 1737-1741.

139. Mehta, R.T. (1989). Liposomes as drug carriers for polyene antibiotics. Advanced Drug Delivery Reviews *3*, 283-306.

140. Proffitt, R.T., Satorius, A., Chiang, S.M., Sullivan, L., and Adlermoore, J.P. (1991). Pharmacology and toxicology of a liposomal formulation of amphotericin B (AmBisome) in rodents. Journal of Antimicrobial Chemotherapy *28*, 49-61.

141. Szoka, F.C., Milholland, D., and Barza, M. (1987). Effect of lipid composition and liposome size on toxicity and in vitro fungicidal activity of liposome-intercalated amphotericin B. Antimicrobial Agents and Chemotherapy *31*, 421-429.

142. Warnock, D.W. (1991). Amphotericin B: an introduction. Journal of Antimicrobial Chemotherapy *28*, 27-38.

143. Andreoli, T.E., and Monahan, M. (1968). The interaction of polyene antibiotics with thin lipid membranes. The Journal of general physiology *52*, 300-325.

144. Mehta, R.T., McQueen, T.J., Keyhani, A., and Lopezberestein, G. (1991). Liposomal Hamycin: Reduced Toxicity and Improved Antifungal Efficacy in vitro and in vivo. Journal of Infectious Diseases *164*, 1003-1006.

145. Mehta, R.T., Hopfer, R.L., Gunner, L.A., Juliano, R.L., and Lopezberestein, G. (1987). Formulation, toxicity, and antifungal activity in vitro of liposome-encapsulated nystatin as therapeutic agent for systemic candidiasis. Antimicrobial Agents and Chemotherapy *31*, 1897-1900.

146. Mehta, R.T., Hopfer, R.L., McQueen, T., Juliano, R.L., and Lopezberestein, G. (1987). Toxicity and therapeutic effects in mice of liposome-encapsulated nystatin for systemic fungal infections. Antimicrobial Agents and Chemotherapy *31*, 1901-1903.

147. Octavia, Y., Tocchetti, C.G., Gabrielson, K.L., Janssens, S., Crijns, H.J., and Moens, A.L. (2012). Doxorubicin-induced cardiomyopathy: From molecular mechanisms to therapeutic strategies. Journal of Molecular and Cellular Cardiology *52*, 1213-1225.

148. Herman, E.H., Rahman, A., Ferrans, V.J., Vick, J.A., and Schein, P.S. (1983). Prevention of Chronic Doxorubicin Cardiotoxicity in Beagles by Liposomal Encapsulation. Cancer Research *43*, 5427-5432.

149. Vanhoesel, Q., Steerenberg, P.A., Crommelin, D.J.A., Vandijk, A., Vanoort, W., Klein, S., Douze, J.M.C., Dewildt, D.J., and Hillen, F.C. (1984). Reduced Cardiotoxicity and Nephrotoxicity with Preservation of Antitumor Activity of Doxorubicin Entrapped in Stable Liposomes in the LOU/M Wsl Rat. Cancer Research *44*, 3698-3705.

150. Gabizon, A.A. (1992). Selective Tumor Localization and Improved Therapeutic Index of Anthracyclines Encapsulated in Long-Circulating Liposomes. Cancer Research *52*, 891-896.

151. Owen, S., Chan, D., and Shoichet, M. (2012). Polymeric micelle stability. Nano Today *7*, 53-65.

152. Schramm, L.L. (2000). Surfactants: Fundamentals and Applications in the Petroleum Industry, (Cambridge University Press).

153. Muller, P. (1994). Glossary of terms used in physical organic chemistry (IUPAC Recommendations 1994). Pure and Applied Chemistry *66*, 1077-1184.

154. Xu, W., Ling, P., and Zhang, T. (2013). Polymeric micelles, a promising drug delivery system to enhance bioavailability of poorly water-soluble drugs. Journal of drug delivery *2013*, 340315-340315.

155. Lavasanifar, A., Samuel, J., and Kwon, G.S. (2002). Poly(ethylene oxide)-block-poly(L-amino acid) micelles for drug delivery. Advanced Drug Delivery Reviews *54*, 169-190.

156. Allen, C., Maysinger, D., and Eisenberg, A. (1999). Nano-engineering block copolymer aggregates for drug delivery. Colloids and Surfaces B-Biointerfaces *16*, 3-27.

157. Yu, B.G., Okano, T., Kataoka, K., and Kwon, G. (1998). Polymeric micelles for drug delivery: solubilization and haemolytic activity of amphotericin B. Journal of Controlled Release *53*, 131-136.

158. Torchilin, V.P. (2001). Structure and design of polymeric surfactant-based drug delivery systems. Journal of Controlled Release *73*, 137-172.

159. Mondon, K., Gurny, R., and Moeller, M. (2008). Colloidal Drug Delivery Systems - Recent Advances With Polymeric Micelles. Chimia *62*, 832-840.

160. Seo, K., Chung, S.W., Byun, Y., and Kim, D. (2012). Paclitaxel loaded nano-aggregates based on pH sensitive polyaspartamide amphiphilic graft copolymers. International Journal of Pharmaceutics *424*, 26-32.

161. Shin, H.-C., Alani, A.W.G., Rao, D.A., Rockich, N.C., and Kwon, G.S. (2009). Multi-drug loaded polymeric micelles for simultaneous delivery of poorly soluble anticancer drugs. Journal of Controlled Release *140*, 294-300.

162. Lee, E.S., Na, K., and Bae, Y.H. (2003). Polymeric micelle for tumor pH and folate-mediated targeting. Journal of Controlled Release *91*, 103-113.

163. Mourya, V.K., Inamdar, N., Nawale, R.B., and Kulthe, S.S. (2011). Polymeric Micelles: General Considerations and their Applications. Indian Journal of Pharmaceutical Education and Research *45*, 128-138.

164. Kataoka, K., Harada, A., and Nagasaki, Y. (2001). Block copolymer micelles for drug delivery: design, characterization and biological significance. Advanced Drug Delivery Reviews *47*, 113-131.

165. Kedar, U., Phutane, P., Shidhaye, S., and Kadam, V. (2010). Advances in polymeric micelles for drug delivery and tumor targeting. Nanomedicine-Nanotechnology Biology and Medicine *6*, 714-729.

166. van Zuylen, L., Verweij, J., and Sparreboom, A. (2001). Role of formulation vehicles in taxane pharmacology. Investigational New Drugs *19*, 125-141.

167. Huo, M., Zhang, Y., Zhou, J., Zou, A., Yu, D., Wu, Y., Li, J., and Li, H. (2010). Synthesis and characterization of low-toxic amphiphilic chitosan derivatives and their

application as micelle carrier for antitumor drug. International Journal of Pharmaceutics *394*, 162-173.

168. Barwicz, J., Christian, S., and Gruda, I. (1992). Effects of the aggregation state of amphotericin B on its toxicity to mice. Antimicrobial Agents and Chemotherapy *36*, 2310-2315.

169. Yu, B.G., Okano, T., Kataoka, K., Sardari, S., and Kwon, G.S. (1998). In vitro dissociation of antifungal efficacy and toxicity for amphotericin B-loaded poly(ethylene oxide)-block-poly(beta-benzyl-L-aspartate) micelles. Journal of Controlled Release *56*, 285-291.

170. Mizumura, Y., Matsumura, Y., Hamaguchi, T., Nishiyama, N., Kataoka, K., Kawaguchi, T., Hrushesky, W.J.M., Moriyasu, F., and Kakizoe, T. (2001). Cisplatin-incorporated polymeric micelles eliminate nephrotoxicity, while maintaining antitumor activity. Japanese Journal of Cancer Research *92*, 328-336.

171. Szejtli, J. (1998). Introduction and general overview of cyclodextrin chemistry. Chemical Reviews *98*, 1743-1753.

172. Del Valle, E.M.M. (2004). Cyclodextrins and their uses: a review. Process Biochemistry *39*, 1033-1046.

173. Davis, M.E., and Brewster, M.E. (2004). Cyclodextrin-based pharmaceutics: Past, present and future. Nature Reviews Drug Discovery *3*, 1023-1035.

174. Schneiderman, E., and Stalcup, A.M. (2000). Cyclodextrins: a versatile tool in separation science. Journal of Chromatography B *745*, 83-102.

175. Arima, H., Yunomae, K., Miyake, K., Irie, T., Hirayama, F., and Uekama, K. (2001). Comparative studies of the enhancing effects of cyclodextrins on the solubility and oral bioavailability of tacrolimus in rats. Journal of Pharmaceutical Sciences *90*, 690-701.

176. Loftsson, T., and Duchene, D. (2007). Cyclodextrins and their pharmaceutical applications. International Journal of Pharmaceutics *329*, 1-11.

177. Arikan, S., Yigit, A.A., and Zengin, N. (2004). Effect of pre-haemolytic concentrations of alpha, beta and gamma cyclodextrins on the osmotic fragility of dog erythrocytes. Revue De Medecine Veterinaire *155*, 500-503.

178. Hanumegowda, U.M., Wu, Y., and Adams, S.P. (2014). Potential Impact of Cyclodextrin-Containing Formulations in Toxicity Evaluation of Novel Compounds in Early Drug Discovery. Journal of Pharmaceutics & Pharmacology *2*, 1-5.

179. Irie, T., and Uekama, K. (1997). Pharmaceutical applications of cyclodextrins .3. Toxicological issues and safety evaluation. Journal of Pharmaceutical Sciences *86*, 147-162.

180. Loftsson, T., and Brewster, M.E. (1996). Pharmaceutical applications of cyclodextrins .1. Drug solubilization and stabilization. Journal of Pharmaceutical Sciences *85*, 1017-1025.

181. Rajewski, R.A., and Stella, V.J. (1996). Pharmaceutical applications of cyclodextrins .2. In vivo drug delivery. Journal of Pharmaceutical Sciences *85*, 1142-1169.

182. Challa, R., Ahuja, A., Ali, J., and Khar, R.K. (2005). Cyclodextrins in drug delivery: an updated review. AAPS PharmSciTech *6*, E329-357.

183. Masson, M., Loftsson, T., Masson, G., and Stefansson, E. (1999). Cyclodextrins as permeation enhancers: some theoretical evaluations and in vitro testing. Journal of Controlled Release *59*, 107-118.

184. Uekama, K., Irie, T., Sunada, M., Otagiri, M., and Tsubaki, K. (1981). Protective effects of cyclodextrins on drug-induced hemolysis in vitro. Journal of Pharmacobio-Dynamics *4*, 142-144.

185. Irie, T., Kuwahara, S., Otagiri, M., Uekama, K., and Iwamasa, T. (1983). Reduction in the local tissue toxicity of chlorpromazine by beta-cyclodextrin complexation. Journal of Pharmacobio-Dynamics *6*, 790-792.

186. Puglisi, G., Ventura, C.A., Spadaro, A., Campana, G., and Spampinato, S. (1995). Differential effects of modified beta-cyclodextrins on pharmacological activity and bioavailability of 4-biphenylacetic acid in rats after oral administration. Journal of Pharmacy and Pharmacology *47*, 120-123.

187. Wu, Z., Tucker, I.G., Razzak, M., Yang, L., McSporran, K., and Medlicott, N.J. (2010). Absorption and tissue tolerance of ricobendazole in the presence of hydroxypropyl-beta-cyclodextrin following subcutaneous injection in sheep. International Journal of Pharmaceutics *397*, 96-102.

188. Yoshida, A., Yamamoto, M., Itoh, T., Irie, T., Hirayama, F., and Uekama, K. (1990). Utility of 2-hydroxypropyl-beta-cyclodextrin in an intramuscular injectable preparation of nimodipine. Chemical & Pharmaceutical Bulletin *38*, 176-179.

189. Tanaka, A., Araki, H., Komoike, Y., Hase, S., and Takeuchi, K. (2001). Inhibition of both COX-1 and COX-2 is required for development of gastric damage in response to nonsteroidal antiinflammatory drugs. Journal of Physiology-Paris *95*, 21-27.

190. Takeuchi, K. (2012). Pathogenesis of NSAID-induced gastric damage: Importance of cyclooxygenase inhibition and gastric hypermotility. World Journal of Gastroenterology *18*, 2147-2160.

191. Lin, S.Z., Wouessidjewe, D., Poelman, M.C., and Duchene, D. (1994). In vivo evaluation of indomethacin/cyclodextrin complexes - gastrointestinal tolerance and dermal anti-inflammatory activity. International Journal of Pharmaceutics *106*, 63-67.

192. Cappello, B., di Maio, C., Iervolino, M., Miro, A., and Calignano, A. (2009). Etodolac/cyclodextrin formulations: physicochemical characterization and in vivo pharmacological studies. Drug Development and Industrial Pharmacy *35*, 877-886.

193. Espinar, F.J.O., Igea, S.A., Mendez, J.B., and Jato, J.L.V. (1991). Reduction in the ulcerogenicity of naproxen by complexation with β-cyclodextrin. International Journal of Pharmaceutics *70*, 35-41.

194. Baboota, S., Dhaliwal, M., and Kohli, K. (2005). Physicochemical characterization, in vitro dissolution behavior, and pharmacodynamic studies of rofecoxib-cyclodextrin inclusion compounds. preparation and properties of rofecoxib hydroxypropyl beta-cyclodextrin inclusion complex: a technical note. AAPS PharmSciTech *6*, E83-90.

195. Govindarajan, R., and Nagarsenker, M.S. (2005). Formulation studies and in vivo evaluation of a flurbiprofen-hydroxypropyl beta-cyclodextrin system. Pharmaceutical Development and Technology *10*, 105-114.

196. Zhao, C., Wang, Y., Su, Y., Zhang, H., Ding, L., Yan, X., Zhao, D., Shao, N., Ye, X., and Cheng, Y. (2011). Inclusion complexes of isoflavones with two commercially available dendrimers: Solubility, stability, structures, release behaviors, cytotoxicity, and anti-oxidant activities. International Journal of Pharmaceutics *421*, 301-309.

197. Albertazzi, L., Gherardini, L., Brondi, M., Sato, S.S., Bifone, A., Pizzorusso, T., Ratto, G.M., and Bardi, G. (2013). In Vivo Distribution and Toxicity of PAMAM Dendrimers in the Central Nervous System Depend on Their Surface Chemistry. Molecular Pharmaceutics *10*, 249-260.

198. Esfand, R., and Tomalia, D.A. (2001). Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications. Drug Discovery Today *6*, 427-436.

199. Medina, S.H., and El-Sayed, M.E.H. (2009). Dendrimers as Carriers for Delivery of Chemotherapeutic Agents. Chemical Reviews *109*, 3141-3157.

200. Neerman, M.F., Chen, H.-T., Parrish, A.R., and Simanek, E.E. (2004). Reduction of drug toxicity using dendrimers based on melamine. Molecular Pharmaceutics *1*, 390-393.

201. Anton, N., and Vandamme, T.F. (2011). Nano-emulsions and Micro-emulsions: Clarifications of the Critical Differences. Pharmaceutical Research *28*, 978-985.

202. McClements, D.J. (2012). Nanoemulsions versus microemulsions: terminology, differences, and similarities. Soft Matter *8*, 1719-1729.

203. Anton, N., Benoit, J.-P., and Saulnier, P. (2008). Design and production of nanoparticles formulated from nano-emulsion templates - A review. Journal of Controlled Release *128*, 185-199.

204. Mason, T.G., Wilking, J.N., Meleson, K., Chang, C.B., and Graves, S.M. (2006). Nanoemulsions: formation, structure, and physical properties. Journal of Physics-Condensed Matter *18*, R635-R666.

205. Ricci, M., Blasi, P., Giovagnoli, S., Perioli, L., Vescovi, C., and Rossi, C. (2004). Leucinostatin-A loaded nanospheres: characterization and in vivo toxicity and efficacy evaluation. International Journal of Pharmaceutics *275*, 61-72.

206. Ankola, D.D., Battisti, A., Solaro, R., and Kumar, M.N.V.R. (2010). Nanoparticles made of multi-block copolymer of lactic acid and ethylene glycol containing periodic side-chain carboxyl groups for oral delivery of cyclosporine A. Journal of the Royal Society Interface *7*, S475-S481.

207. Liu, F., Park, J.-Y., Zhang, Y., Conwell, C., Liu, Y., Bathula, S.R., and Huang, L. (2010). Targeted Cancer Therapy With Novel High Drug-Loading Nanocrystals. Journal of Pharmaceutical Sciences *99*, 3542-3551.

208. Moreno, D., Zalba, S., Navarro, I., Tros de Ilarduya, C., and Garrido, M.J. (2010). Pharmacodynamics of cisplatin-loaded PLGA nanoparticles administered to tumor-bearing mice. European Journal of Pharmaceutics and Biopharmaceutics *74*, 265-274.

209. Brime, B., Frutos, P., Bringas, P., Nieto, A., Ballesteros, M.P., and Frutos, G. (2003). Comparative pharmacokinetics and safety of a novel lyophilized amphotericin B lecithin-based oil-water microemulsion and amphotericin B deoxycholate in animal models. Journal of Antimicrobial Chemotherapy *52*, 103-109.

210. Brime, B., Molero, G., Frutos, P., and Frutos, G. (2004). Comparative therapeutic efficacy of a novel lyophilized amphotericin B lecithin-based oil-water microemulsion and deoxycholate-amphotericin B in immunocompetent and neutropenic mice infected with Candida albicans. European Journal of Pharmaceutical Sciences *22*, 451-458.

211. Luftensteiner, C.P., Schwendenwein, I., Eichler, H.G., Paul, B., Wolfl, G., and Viernstein, H. (1999). Toxicity of a particulate formulation for the intraperitoneal application of mitoxantrone. International Journal of Pharmaceutics *180*, 251-260.

212. Luftensteiner, C.P., Schwendenwein, I., Paul, B., Eichler, H.G., and Viernstein, H. (1999). Evaluation of mitoxantrone-loaded albumin microspheres following intraperitoneal administration to rats. Journal of Controlled Release *57*, 35-44.

213. Whittem, T., Parton, K., and Turner, K. (1996). Effect of polyaspartic acid on pharmacokinetics of gentamicin after single intravenous dose in the dog. Antimicrobial Agents and Chemotherapy *40*, 1237-1241.

214. Raucy, J.L., Lasker, J.M., Lieber, C.S., and Black, M. (1989). Acetaminophen activation by human liver cytochromes P450IIE1 and P450IA2. Archives of Biochemistry and Biophysics *271*, 27-83.

215. Rumack, B.H. (2002). Acetaminophen hepatotoxicity: The first 35 years. Journal of Toxicology-Clinical Toxicology *40*, 3-20.

216. Hodgman, M.J., and Garrard, A.R. (2012). A Review of Acetaminophen Poisoning. Critical Care Clinics *28*, 499 - 516.

217. Nakagawa, H., Maeda, S., Hikiba, Y., Ohmae, T., Shibata, W., Yanai, A., Sakamoto, K., Ogura, K., Noguchi, T., Karin, M., et al. (2008). Deletion of apoptosis signal-regulating kinase 1 attenuates acetaminophen-induced liver injury by inhibiting c-jun N-terminal kinase activation. Gastroenterology *135*, 1311-1321.

218. Jones, D.P., Lemasters, J.J., Han, D., Boelsterli, U.A., and Kaplowitz, N. (2010). Mechanisms of Pathogenesis in Drug Hepatotoxicity Putting the Stress on Mitochondria. Molecular Interventions *10*, 98-111.

219. Lee, H.C., Jung, S.A., Jung, H.K., Yi, S.Y., Kim, D.Y., Moon, I.H., and Park, S.S. (1999). Protective effect of chlormethiazole, a sedative, against acetaminophen-induced liver injury in mice. Korean Journal of Internal Medicine *14*, 27-33.

220. AlMustafa, Z.H., AlAli, A.K., Qaw, F.S., and AbdulCader, Z. (1997). Cimetidine enhances the hepatoprotective action of N-acetylcysteine in mice treated with toxic doses of paracetamol. Toxicology *121*, 223-228.

221. Acharya, M., and Lau-Cam, C.A. (2010). Comparison of the protective actions of N-acetylcysteine, hypotaurine and taurine against acetaminophen-induced hepatotoxicity in the rat. Journal of Biomedical Science *17*.

222. Chandrasekaran, V.R.M., Wan, C.-H., Liu, L.-L., Hsu, D.-Z., and Liu, M.-Y. (2008). Effect of sesame oil against acetaminophen-induced acute oxidative hepatic damage in rats. Shock *30*, 217-221.

223. Hattori, A., Yamada, N., Nishikawa, T., Fukuda, H., and Fujino, T. (2001). Protective effect of ajoene on acetaminophen-induced hepatic injury in mice. Bioscience Biotechnology and Biochemistry *65*, 2555-2557.

224. Gyselync.Am, Forrey, A., and Cutler, R. (1971). Pharmacokinetics of gentamicin: distribution and plasma and renal clearance. Journal of Infectious Diseases *124*, S70-&.

225. Nagai, J., Saito, M., Adachi, Y., Yumoto, R., and Takano, M. (2006). Inhibition of gentamicin binding to rat renal brush-border membrane by megalin ligands and basic peptides. Journal of Controlled Release *112*, 43-50.

226. Schmitz, C., Hilpert, J., Jacobsen, C., Boensch, C., Christensen, E.I., Luft, F.C., and Willnow, T.E. (2002). Megalin deficiency offers protection from renal aminoglycoside accumulation. Journal of Biological Chemistry *277*, 618-622.

227. Frommer, J.P., Senekjian, H.O., Babino, H., and Weinman, E.J. (1983). Intratubular microinjection study of gentamicin transport in the rat. Mineral and Electrolyte Metabolism *9*, 108-112.

228. Debroe, M.E., Paulus, G.J., Verpooten, G.A., Roels, F., Buyssens, N., Wedeen, R., Vanhoof, F., and Tulkens, P.M. (1984). Early effects of gentamicin, tobramycin, and amikacin on the human kidney. Kidney International *25*, 643-652.

229. Fribley, A., Zhang, K., and Kaufman, R.J. (2009). Regulation of Apoptosis by the Unfolded Protein Response. Apoptosis: Methods and Protocols, Second Edition *559*, 191-204.

230. Mather, M., and Rottenberg, H. (2001). Polycations induce the release of soluble intermembrane mitochondrial proteins. Biochimica Et Biophysica Acta-Bioenergetics *1503*, 357-368.

231. Morales, A.I., Detaille, D., Prieto, M., Puente, A., Briones, E., Arevalo, M., Leverve, X., Lopez-Novoa, J.M., and El-Mir, M.-Y. (2010). Metformin prevents experimental gentamicin-induced nephropathy by a mitochondria-dependent pathway. Kidney International *77*, 861-869.

232. Gibbons, C.E., Maldonado-Perez, D., Shah, A.N., Riccardi, D., and Ward, D.T. (2008). Calcium-sensing receptor antagonism or lithium treatment ameliorates aminoglycoside-induced cell death in renal epithelial cells. Biochimica Et Biophysica Acta-Molecular Basis of Disease *1782*, 188-195.

233. Luft, F.C., Aronoff, G.R., Evan, A.P., and Connors, B.A. (1981). The effect of aminoglycosides on glomerular endothelium: a comparative study. Research Communications in Chemical Pathology and Pharmacology *34*, 89-95.

234. Shiotani, K., Irie, T., Uekama, K., and Ishimaru, Y. (1995). Cyclodextrin sulfates in parenteral use: protection against gentamicin nephrotoxicity in the rat. European Journal of Pharmaceutical Sciences *3*, 139-151.

235. Uekama, K., Shiotani, K., Irie, T., Ishimaru, Y., and Pitha, J. (1993). Protective effects of cyclodextrin sulphates against gentamicin-induced nephrotoxicity in the rat. Journal of Pharmacy and Pharmacology *45*, 745-747.

236. Beauchamp, D., Laurent, G., Grenier, L., Gourde, P., Zanen, J., HeusonStiennon, J.A., and Bergeron, M.G. (1997). Attenuation of gentamicin-induced nephrotoxicity in rats by fleroxacin. Antimicrobial Agents and Chemotherapy *41*, 1237-1245.

237. Amini, F.G., Rafieian-Kopaei, M., Nematbakhsh, M., Baradaran, A., and Nasri, H. (2012). Ameliorative effects of metformin on renal histologic and biochemical alterations of gentamicin-induced renal toxicity in Wistar rats. Journal of Research in Medical Sciences *17*, 621-625.

238. Kadkhodaee, M., Khastar, H., Arab, H.A., Ghaznavi, R., Zahmatkesh, M., and Mahdavi-Mazdeh, M. (2007). Antioxidant vitamins preserve superoxide dismutase activities in gentamicin-induced nephrotoxicity. Transplantation Proceedings *39*, 864-865.

239. Reiter, R.J., Tan, D.X., Sainz, R.M., Mayo, J.C., and Lopez-Burillo, S. (2002). Melatonin: reducing the toxicity and increasing the efficacy of drugs. Journal of Pharmacy and Pharmacology *54*, 1299-1321.

240. Buyukafsar, K., Yazar, A., Dusmez, D., Ozturk, H., Polat, G., and Levent, A. (2001). Effect of trapidil, an antiplatelet and vasodilator agent on gentamicin-induced nephrotoxicity in rats. Pharmacological Research *44*, 321-328.

241. Karahan, I., Atessahin, A., Yilmaz, S., Ceribasi, A.O., and Sakin, F. (2005). Protective effect of lycopene on gentamicin-induced oxidative stress and nephrotoxicity in rats. Toxicology *215*, 198-204.

242. Ali, B.H., Al-Wabel, N., Mahmoud, O., Mousa, H.M., and Hashad, M. (2005). Curcumin has a palliative action on gentamicin-induced nephrotoxicity in rats. Fundamental & Clinical Pharmacology *19*, 473-477.

243. Reddy, A.C.P., and Lokesh, B.R. (1996). Effect of curcumin and eugenol on iron-induced hepatic toxicity in rats. Toxicology *107*, 39-45.

244. Kumar, K.V., Naidu, M.U.R., Shifow, A.A., and Ratnakar, K.S. (2000). Probucol protects against gentamicin-induced nephrotoxicity in rats. Indian Journal of Pharmacology *32*, 108-113.

245. Said, M.M. (2011). The protective effect of eugenol against gentamicin-induced nephrotoxicity and oxidative damage in rat kidney. Fundamental & Clinical Pharmacology *25*, 708-716.

246. Hsu, D.-Z., Liu, C.-T., Li, Y.-H., Chu, P.-Y., and Liu, M.-Y. (2010). Protective effect of daily sesame oil supplement on gentamicin-induced renal injury in rats. Shock *33*, 88-92.

247. Ajith, T.A., Usha, S., and Nivitha, V. (2007). Ascorbic acid and alpha-tocopherol protect anticancer drug cisplatin induced nephrotoxicity in mice: a comparative study. Clinica Chimica Acta *375*, 82-86.

248. Nikolova, N., and Jaworska, J. (2004). Approaches to measure chemical similarity - A review. Qsar & Combinatorial Science *22*, 1006-1026.

249. Cronin, M.T.D. (2001). Prediction of drug toxicity. Farmaco *56*, 149-151.

250. Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics *29*, 476-488.

251. Thai, K.-M., and Ecker, G.F. (2008). Classification models for hERG inhibitors by counter-propagation neural networks. Chemical Biology & Drug Design *72*, 279-289.

252. Benfenati, E., and Gini, G. (1997). Computational predictive programs (expert systems) in toxicology. Toxicology *119*, 213-225.

253. Lhasa Limited http://www.lhasalimited.org/.

254. Oncologic http://www.epa.gov/tsca-screening-tools/oncologictm-computer-system-evaluate-carcinogenic-potential-chemicals.

255. HazardExpert Pro http://www.compudrug.com/hazardexpertpro.

256. Greene, N. (2002). Computer systems for the prediction of toxicity: an update. Advanced Drug Delivery Reviews *54*, 417-431.

257. TOPKAT http://accelrys.com/.

258. MultiCASE http://www.multicase.com/.

259. Toropov, A.A., Toropova, A.P., Raska, I., Jr., Leszczynska, D., and Leszczynski, J. (2014). Comprehension of drug toxicity: Software and databases. Computers in Biology and Medicine *45*, 20-25.

260. Cronin, M.T.D., and Madden, J.C. (2010). In Silico Toxicology Principles and Applications, (Royal Society of Chemistry).

261. Lapatas, V., Stefanidakis, M., Jimenez, R.C., Via, A., and Schneider, M.V. (2015). Data integration in biological research: an overview. Journal of Biological Research-Thessaloniki *22*, 1-16.

262. Lenzerini, M. (2002). Data integration: a theoretical perspective. Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 233-246.

263. Varnek, A., and Baskin, I. (2012). Machine Learning Methods for Property Prediction in Chemoinformatics: Quo Vadis? Journal of Chemical Information and Modeling *52*, 1413-1437.

264. Schneider, K., Schwarz, M., Burkholder, I., Kopp-Schneider, A., Edler, L., Kinsner-Ovaskainen, A., Hartung, T., and Hoffmann, S. (2009). ToxRTool, a new tool to assess the reliability of toxicological data. Toxicology Letters *189*, 138-144.

265. Klimisch, H.J., Andreae, M., and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology *25*, 1-5.

266. IUPAC http://www.iupac.org/.

267. Daylight Theory Manual http://www.daylight.com/dayhtml/doc/theory/.

268. Openbabel http://openbabel.org/docs/dev/Cheminf101/represent.html.

269. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences *28*, 31-36.

270. Stein, S.E., Heller, S.R., and Tchekhovskoi, D.V. (2011). IUPAC International Chemical Identifier (InChI).

271. Heller, S.R., McNaught, A., Pletnev, I., Stein, S., and Tchekhovskoi, D. (2015). InChI, the IUPAC International Chemical Identifier. Journal of Cheminformatics *7*.

272. Symyx CTfile Formats.

273. Karelson, M. (2000). Molecular Descriptors in QSAR/QSPR, (Wiley-Blackwell).

274. Devillers, J., and Balaban, A.T. (2000). Topological Indices and Related Descriptors in QSAR and QSPR, (CRC Press).

275. Todeschini, R., and Consonni, V. (2000). Handbook of Molecular Descriptors. (Wiley VCH).

276. Katritzky, A.R. CODESSA PRO, http://www.codessa-pro.com/index.htm.

277. Talete DRAGON, http://www.talete.mi.it/index.htm.

278. Virtual Computational Chemistry Laboratory E-DRAGON, http://www.vcclab.org/lab/edragon/.

279. Ekinci, D. (2012). Medicinal Chemistry and Drug Design. (InTech).

280. Randic, M. (1993). Novel molecular descriptor for structure-property studies. Chemical Physics Letters *211*, 478-483.

281. Maya Chem Tools http://www.mayachemtools.org/docs/modules/html/MACCSKeys.html.

282. Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., and Tatonetti, N.P. (2014). Similarity-based modeling in large-scale prediction of drug-drug interactions. Nature Protocols *9*, 2147-2163.

283. Willett, P., Barnard, J.M., and Downs, G.M. (1998). Chemical Similarity Searching. Journal of Chemical Information and Computer Sciences - *38*, 983–996.

284. Holliday, J.D., Hu, C.Y., and Willett, P. (2002). Grouping of coefficients for the calculation of inter-molecular similarity and dissimilarity using 2D fragment bit-strings. Combinatorial Chemistry & High Throughput Screening *5*, 155-166.

285. Mitchell, J.B.O. (2014). Machine learning methods in chemoinformatics. Wiley Interdisciplinary Reviews-Computational Molecular Science *4*, 468-481.

286. Connors, B.W., and Long, M.A. (2004). Electrical synapses in the mammalian brain. Annual Review of Neuroscience *27*, 393-418.

287. Li, H., Ung, C.Y., Yap, C.W., Xue, Y., Li, Z.R., and Chen, Y.Z. (2006). Prediction of estrogen receptor agonists and characterization of associated molecular descriptors by statistical learning methods. Journal of Molecular Graphics & Modelling *25*, 313-323.

288. Briem, H., and Gunther, J. (2005). Classifying "kinase inhibitor-likeness" by using machine-learning methods. Chembiochem *6*, 558-566.

289. Basak, S.C., Grunwald, G.D., Gute, B.D., Balasubramanian, K., and Opitz, D. (2000). Use of statistical and neural net approaches in predicting toxicity of chemicals. Journal of Chemical Information and Computer Sciences *40*, 885-890.

290. Kaiser, K.L.E., and Niculescu, S.P. (1999). Using probabilistic neural networks to model the toxicity of chemicals to the fathead minnow (Pimephales promelas): A study based on 865 compounds. Chemosphere *38*, 3237-3245.

291. Vracko, M., Novic, M., and Zupan, J. (1999). Study of structure-toxicity relationship by a counterpropagation neural network. Analytica Chimica Acta *384*, 319-332.

292. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. (http://arxiv.org/abs/1207.0580).

293. Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A. (1984). Classification and Regression Trees, (CRC Press).

294. Salzberg, S. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan, Volume - 16, (Morgan Kaufmann Publishers, Inc., 1993).

295. Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., et al. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems *14*, 1-37.

296. Breiman, L. (2001). Random Forests. Machine Learning *45*, 5-32.

297. Raileanu, L.E., and Stoffel, K. (2004). Theoretical comparison between the Gini Index and Information Gain criteria. Annals of Mathematics and Artificial Intelligence *41*, 77-93.

298. Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., and Feuston, B.P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences *43*, 1947-1958.

299. Ren, S.J. (2003). Phenol mechanism of toxic action classification and prediction: a decision tree approach. Toxicology Letters *144*, 313-323.

300. Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., and Barron, M.G. (2013). Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. Journal of Chemical Information and Modeling *53*, 2229-2239.

301. Polishchuk, P.G., Muratov, E.N., Artemenko, A.G., Kolumbin, O.G., Muratov, N.N., and Kuz'min, V.E. (2009). Application of Random Forest Approach to QSAR

Prediction of Aquatic Toxicity. Journal of Chemical Information and Modeling *49*, 2481-2488.

302. McCarren, P., Springer, C., and Whitehead, L. (2011). An investigation into pharmaceutically relevant mutagenicity data and the influence on Ames predictive potential. Journal of Cheminformatics *3*.

303. Li, S.Q., Fedorowicz, A., Singh, H., and Soderholm, S.C. (2005). Application of the random forest method in studies of local lymph node assay based skin sensitization data. Journal of Chemical Information and Modeling *45*, 952-964.

304. Joachims, T., Finley, T., and Yu, C.-N.J. (2009). Cutting-plane training of structural SVMs. Machine Learning *77*, 27-59.

305. Hsu, C.W., and Lin, C.J. (2002). A comparison of methods for multiclass support vector machines. Ieee Transactions on Neural Networks *13*, 415-425.

306. Lowe, R., Glen, R.C., and Mitchell, J.B.O. (2010). Predicting Phospholipidosis Using Machine Learning. Molecular Pharmaceutics *7*, 1708-1714.

307. Glen, R.C., Bender, A., Arnby, C.H., Carlsson, L., Boyer, S., and Smith, J. (2006). Circular fingerprints: Flexible molecular descriptors with applications from physical chemistry to ADME. Idrugs *9*, 199-204.

308. Li, Q., Jorgensen, F.S., Oprea, T., Brunak, S., and Taboureau, O. (2008). hERG classification model based on a combination of support vector machine method and GRIND descriptors. Molecular Pharmaceutics *5*, 117-127.

309. Doddareddy, M.R., Klaasse, E.C., Shagufta, Ijzerman, A.P., and Bender, A. (2010). Prospective Validation of a Comprehensive In silico hERG Model and its Applications to Commercial Compound and Drug Databases. Chemmedchem *5*, 716-729.

310. Liao, Q., Yao, J., and Yuan, S. (2007). Prediction of mutagenic toxicity by combination of recursive partitioning and support vector machines. Molecular Diversity *11*, 59-72.

311. Maggiora, G.M. (2006). On outliers and activity cliffs - Why QSAR often disappoints. Journal of Chemical Information and Modeling *46*, 1535-1535.

312. Cronin, M.T.D., Netzeva, T.I., Dearden, J.C., Edwards, R., and Worgan, A.D.P. (2004). Assessment and modeling of the toxicity of organic chemicals to Chlorella vulgaris: Development of a novel database. Chemical Research in Toxicology *17*, 545-554.

313. Honorio, K.M., and da Silva, A.B.F. (2005). A study on the influence of molecular properties in the psychoactivity of cannabinoid compounds. Journal of Molecular Modeling *11*, 200-209.

314. Gadaleta, D., Pizzo, F., Lombardo, A., Carotti, A., Escher, S.E., Nicolotti, O., and Benfenati, E. (2014). A k-NN Algorithm for Predicting Oral Sub-Chronic Toxicity in the Rat. Altex-Alternatives to Animal Experimentation *31*, 423-432.

315. Mishra, M., Fei, H., and Huan, J. (2013). Computational prediction of toxicity. International Journal of Data Mining and Bioinformatics *8*, 338-348.

316. Zefirov, N.S., and Palyulin, V.A. (2001). QSAR for boiling points of "small" sulfides. Are the "high-quality structure-property-activity regressions" the real high quality QSAR models? Journal of Chemical Information and Computer Sciences *41*, 1022-1027.

317. Kubinyi, H., Hamprecht, F.A., and Mietzner, T. (1998). Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. Journal of Medicinal Chemistry *41*, 2553-2564.

318. Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C.P., and Agrawal, R.K. (2011). Validation of QSAR Models - Strategies and Importance. International Journal of Drug Design and Discovery *2*, 511-519.

319. Tropsha, A., Gramatica, P., and Gombar, V.K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. Qsar & Combinatorial Science *22*, 69-77.

320. Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management *45*, 427-437.

321. Plewczynski, D., Spieser, S.A.H., and Koch, U. (2006). Assessing different classification methods for virtual screening. Journal of Chemical Information and Modeling *46*, 1098-1106.

322. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics *16*, 412-424.

323. Sushko, I., Novotarskyi, S., Koerner, R., Pandey, A.K., Cherkasov, A., Lo, J., Gramatica, P., Hansen, K., Schroeter, T., Mueller, K.-R., et al. (2010). Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. Journal of Chemical Information and Modeling *50*, 2094-2111.

324. Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., et al. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM Workshop 52. Atla-Alternatives to Laboratory Animals *33*, 155-173.

325. Bullmore, E.T., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience *10*, 186-198.

326. Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L. (2003). A graph-theory algorithm for rapid protein side-chain prediction. Protein Science *12*, 2001-2014.

327. Estrada, E., Patlewicz, G., and Uriarte, E. (2003). From molecular graphs to drugs. A review on the use of topological indices in drug design and discovery. Indian Journal of Chemistry Section a-Inorganic Bio-Inorganic Physical Theoretical & Analytical Chemistry *42*, 1315-1329.

328. Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: Understanding the cell's functional organization. Nature Reviews Genetics *5*, 101-U115.

329. Knime http://www.knime.org/.

330. Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. (Meta Group).

331. Mircea, R.T., and Ivan, M.L. (2014). Big Data: present and future. Database Systems Journal *5*, 32-41.

332. Daneshpazhouh, A., and Sami, A. (2014). Entropy-based outlier detection using semi-supervised approach with few positive examples. Pattern Recognition Letters *49*, 77-84.

333. Otey, M.E., Ghoting, A., and Parthasarathy, S. (2006). Fast distributed outlier detection in mixed-attribute data sets. Data Mining and Knowledge Discovery *12*, 203-228.

334. Liu, H.C., Shah, S., and Jiang, W. (2004). On-line outlier detection and data cleaning. Computers & Chemical Engineering *28*, 1635-1647.

335. Zhang, J., Tao, X., and Wang, H. (2014). Outlier detection from large distributed databases. World Wide Web-Internet and Web Information Systems *17*, 539-568.

336. Angiulli, F., and Pizzuti, C. (2005). Outlier mining in large high-dimensional data sets. Ieee Transactions on Knowledge and Data Engineering *17*, 203-215.

337. Ren, C.-X., Dai, D.-Q., He, X., and Yan, H. (2015). Sample Weighting: An Inherent Approach for Outlier Suppressing Discriminant Analysis. Ieee Transactions on Knowledge and Data Engineering *27*, 3070-3083.

338. Steinmetz, F.P., Madden, J.C., and Cronin, M.T.D. (2015). Data Quality in the Human and Environmental Health Sciences: Using Statistical Confidence Scoring to Improve QSAR/QSPR Modeling. Journal of Chemical Information and Modeling *55*, 1739-1746.

339. Bedson, P., and Farrant, T.J.D. (2009). Practical Statistics for the Analytical Scientist, (Royal Society of Chemistry).

340. Gemperline, P. (2006). Practical Guide To Chemometrics, (CRC Press).

341. Gottmann, E., Kramer, S., Pfahringer, B., and Helma, C. (2001). Data quality in predictive toxicology: Reproducibility of rodent carcinogenicity experiments. Environmental Health Perspectives *109*, 509-514.

342. Indigo knime node tool kit http://sourceforge.net/projects/cdk/.

343. Eldred, D.V., and Jurs, P.C. (1999). Prediction of acute mammalian toxicity of organophosphorus pesticide compounds from molecular structure. Sar and Qsar in Environmental Research *10*, 75-99.

344. Eldred, D.V., Weikel, C.L., Jurs, P.C., and Kaiser, K.L.E. (1999). Prediction of fathead minnow acute toxicity of organic compounds from molecular structure. Chemical Research in Toxicology *12*, 670-678.

345. Rodgers, A.D., Zhu, H., Fourches, D., Rusyn, I., and Tropsha, A. (2010). Modeling Liver-Related Adverse Effects of Drugs Using kNearest Neighbor Quantitative Structure Activity Relationship Method. Chemical Research in Toxicology *23*, 724-732.

346. Shafer, J., Agrawal, R., and Mehta, M. (1996). SPRINT: A scalable parallel classifier for data mining. Proceedings of the International Conference on Very Large Data Bases, 544-555.

347. Caruana, R., and Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning. (Pittsburgh, Pennsylvania, USA: ACM), pp. 161-168.

348. Diaz-Uriarte, R., and de Andres, S.A. (2006). Gene selection and classification of microarray data using random forest. Bmc Bioinformatics *7*.

349. AOP Wiki https://aopkb.org/aopwiki/index.php/Main_Page.

350. Craven, P., and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. Numerische Mathematik *31*, 377-403.

351. Wahba, G. (1977). Practical Approximate Solutions to Linear Operator Equations When the Data are Noisy. Siam Journal on Numerical Analysis *14*, 651-667.

352. Benigni, R., and Bossa, C. (2011). Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology. Chemical Reviews *111*, 2507-2536.

Appendix A

Appendix A is an electronic zip file titled:

Pritesh Mistry PhD Thesis 2015

Appendix A contains the following files:

1) NIH dataset

2) Curated drug structures (sdf file)

3) Compound classification tables for 12 vehicle pairings