DEVELOPMENT OF ANALYTICAL METHODS TO DERIVE HYDROPHOBICITY PARAMETERS FOR USE AS DESCRIPTORS FOR THE PREDICTION OF THE ENVIRONMENTAL AND HUMAN HEALTH RISK OF CHEMICALS

MOIRA RUTH LEDBETTER

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

This research programme was carried out in collaboration with
Safety and Environmental Assurance Centre
Unilever Research

April 2012

LJMU RD10 Decl (1992/V1)

LIVERPOOL JOHN MOORES UNIVERSITY

Candidate's declaration form
(This form must be typed)

Note: This form must be submitted to the University with the candidate's thesis.

Name of candidate: Moira Ruth Ledbetter

School: Pharmacy and Biomolecular Science

Degree for which thesis is submitted: PhD

1. Statement of related studies undertaken in connection with the programme of research (see Regulations G4.1 and 4.4)

Attended lectures and workshops relating to QSAR, computational modeling, computation chemistry and predictive toxicology, additionally have attended faculty research seminars.

Attended the H4033A Agilent HPLC ChemStation and Instrument Operation Training course run by Agilent, April 2010

2. Concurrent registration for two or more academic awards (see Regulation G4.7)

I declare that while registered as a candidate for the University's research degree, I have not been a registered candidate or enrolled student for another award of the LJMU or other academic or professional institution

3. Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award

Signed.....Date 27th April 2012
(Candidate)

Acknowledgements

Firstly I would like to thank my director of studies, Prof. Mark T. D. Cronin, for his continual support and guidance over the past three years. Similar thanks go to my other supervisors Dr Judy C. Madden, Dr Steve Gutsell and Dr Geoff Hodges. Special thanks must also go to Prof. John Dearden and Dr David Roberts for their thoughts.

I would also like to thank Dr Steve Gutsell and Dr Geoff Hodges (Unilever Ltd.) who contributed many of the initial ideas for this research, and I appreciate their guidance throughout. Other employees at Unilever to which I owe my gratitude are Jayne Roberts, Sean O'Connor, Chris Sparham, Louise Pope, Alyn Parry and David Thorley.

I gratefully acknowledge the financial support provided by Unilever Ltd., without which this research would not have been possible.

I would like to acknowledge the help and support of the QSAR research group which includes Dr Steve Enoch, Dr Mark Hewitt, Dr Claire Ellison, Dr Katarzyna Przybylak and Przemyslaw Piechota. I acknowledge Dr Phil Rowe for his help and knowledge in statistics. I finally acknowledge the support of the technical staff at LJMU for their help and support in keeping the HPLC system running, this includes Dr Stan Lambert and Mr Phil Salmon.

In memory of Dr. Stan Lambert

Abstract

There is a requirement to assess chemical safety to both 'man' and the environment. Traditionally this was determined through animal testing. Due to ethical and legislative reasons there is a requirement to develop alternatives. *In silico* methods are one such alternative approach and include (Quantitative) Structure-Activity Relationships ((Q)SARs) which formalise the relationships between toxicity for a series of chemicals and their physico-chemical and structural properties. Many toxicity QSARs are based on hydrophobicity which has often been characterised by the logarithm of the octanol-water partition coefficient (log P). An alternative method of characterisation that may be biologically more relevant is the use of Immobilised Artificial Membrane High Performance Liquid Chromatography (IAM-HPLC). In this thesis published IAM retention index (k_{IAM}) values, including details of the experimental procedure, have been collated into a database. The effect of variability of experimental procedure on reported values was investigated. Key experimental parameters were identified that ensure comparable log k_{IAM} values. An experimental IAM-HPLC method was optimised covering a range of hydrophobicities (log P of -1.35 to 6.03) including compounds both unionised and ionised under the conditions of analysis. Method robustness was demonstrated across system of analysis, column and stationary phase batch. Approaches to predict log $k_{IAM\ (pH\ 7.4)}$ were investigated. A theoretical structural fragment and factor approach, and a 'classical' descriptor based QSAR approach, were applied to measured (in this work) and literature log $k_{IAM\ (pH\ 7.4)}$ values. The use of log k_{IAM} and log P (to allow comparison) as descriptors in QSARs to predict skin absorption was investigated. Log k_{IAM} as a descriptor is comparable to log P and an improvement is seen when the degree of ionisation is taken into account. Given the physiological pH of skin, the ionisation of compounds is an important factor in determining skin absorption. The use of log k_{IAM} to predict aquatic toxicity was investigated for a number of species and endpoints with log k_{IAM} being found to be a useful descriptor for narcotic mechanisms of action. The use of IAM-HPLC was also successfully demonstrated for surfactants.

Contents

Abbreviations

ANOVA - ANalysis Of VAriance

aq - aqueous

CAS. - Chemical Abstract Service

CMC - Critical Micelle Concentration

D - distribution co-efficient of concentrations of all species in octanol/water

EC_{50} - Effect concentration observed affecting 50% of the population tested

EU - European Union

F - Fisher statistic

FATS - Fish Acute Toxicity Syndrome

GC – Gas Chromatography

GLM - General Linear Model

HPLC – High Performance Liquid Chromatography

IAM – Immobilised artificial membrane

IAM.PC - Column stationary phase consisting of a single chain of phosphatidylcholine bonded to aminopropyl silica with no endcapping

IAM.PC.DD - Column stationary phase consisting of a single chain of phosphatidylcholine bonded to aminopropyl silica which has been endcapped with C_3 and C_{10} alkyl chains

IAM.PC.DD2 – Column stationary phase consisting of double chain of phosphatidylcholine bonded to aminopropyl silica which has been endcapped with C_3 and C_{10} alkyl chains

IAM.PC.MG - Column stationary phase consisting of double chain of phosphatidylcholine bonded to aminopropyl silica which has been endcapped with methylglycolate

IGC_{50} – Inhibition of growth concentration affecting 50% of the population

K_{ow} - octanol-water partition coefficient

K_P - permeability coefficient

LC_{50} – Lethal concentration affecting 50% of the population tested

LC/MS – Liquid Chromatography/Mass Spectrometry

log – logarithm base 10

Log k_{IAM} – Logarithmic values for the relative retention factor measured using an immobilised artificial membrane column and high performance liquid chromatography

Log $k_{IAM\ (pH\ 7.4)}$ - Logarithmic values for the relative retention factor measured using an immobilised artificial membrane column and high performance liquid chromatography determined at a pH of 7.4

Log $P_{(exp)}$ – Experimentally determined logarithmic value for the octanol-water partition coefficient

LOO - Leave One Out

MeOH - Methanol

MLR - Multiple Linear Regression

MS – Mass Spectrometry

n – number of observations

NSRDS - National Standard Reference Data System

OECD - Organisation for Economic Cooperation and Development

P - octanol-water partition coefficient

P_{ow} - octanol-water partition coefficient

PBS - 10mM phosphate buffered saline

PC – phosphatidylcholine

PRESS - Predicted REsidual Sum of Sqaures

Q^2 – cross validated square of the correlation coefficient

(Q)SAR - (Quantitative) Structure-Activity Relationship

$r^2_{(adj)}$ - the square of the correlation coefficient adjusted for degrees of freedom

REACH - Registration, Evaluation, Authorisation and restriction of CHemical substances

RI – Refractive Index

RP - Reverse Phase

RRF - Relative Retention Factors

s - standard deviation

SC - stratum corneum

SMILES - Simplified Molecular Input Line Entry Specification/Systems

SS_T - Total Sum of Squares

t_0 – retention time of the unretained compound

TGD - Technical Guidance Document

t_r – retention time of the sample

UKTAG - United Kingdom Technical Advisory Group

UV – Ultra Violet

VCCLAB - Virtual Computational Chemistry LABoratory

1 Introduction

1.1 Chemical safety

Man and the environment are exposed to many chemicals of anthropogenic source on a daily basis[1]. For man these can be from various sources in everyday life including food, personal care products and pharmaceuticals[2]. In addition, workers are often exposed to various chemicals in their occupation[3]. The environment receives chemicals from a variety of routes, including the disposal of chemicals in wastewater from domestic sources (e.g. via sewage treatment plants), industry and agriculture. The range of concentrations to which 'man' and the environment are exposed is large. The exposure also depends upon how the chemicals are used, the exposure pathways, chemical properties and any control measures (in factories or sewage treatments plants etc.)[4]. All chemicals have an intrinsic hazard, which varies from relatively benign to extremely harmful[2]. The potential for exposure to chemicals is highly variable. However, exposure alone is insufficient to cause an adverse effect on the body or the environment[5].

In order to determine the safety, or otherwise, of exposure to a chemical, the relative risk must be determined. Risk in this context, is the probability of occurrence of an adverse effect in the body or the environment resulting from a given exposure to a chemical or mixture. Hazard is the inherent capacity of a chemical or mixture to cause adverse effects to the body or the environment[6]. Exposure to a chemical includes the duration and / or the frequency of contact with the chemical and the concentration at which the chemical is present[4]. There is a fundamental relationship linking risk, hazard and exposure in terms of the toxicology:

$$R = f\,H, E \qquad (1.1)$$

Where:

R is risk

H is hazard

E is exposure

The hypothesis that the level of exposure to a compound is important in determining the harm a chemical may induce is well established and a fundamental concept in toxicology. As far back as the 16th century Paracelsus stated:

> *"All substances are poisons: there is none which is not a poison. The right dose differentiates a poison and a remedy"* [7]

Thus, in order to understand the potential of a chemical to harm human health or the environment, knowledge is required of the inherent toxicity of the chemical and the exposure to that chemical. Exposure can be short term, causing acute toxicity or long term, causing chronic toxicity. The assessment of acute or chronic effects requires knowledge of the effects of chemicals in humans or, more commonly, surrogate animals or other test systems[8]. Acute exposure is characterised by a single administration often at a relatively high dose. This is representative of a single use of a chemical, or an accidental exposure. Chronic toxicity is more commonly at a lower dose (sub-lethal) and for either a longer period of time or with repeated doses, normally for a significant proportion of the generation time for a species, with some tests spanning several generations to determine the effect of exposure to a chemical on reproductive capacity[9]. Such an exposure scenario is typical of normal (daily) use of pharmaceuticals, personal care products and food ingredients. Personal care products include deodorants, toothpastes, shampoos, soaps, fragrances and lotions etc..

1.2 Toxicology

Toxicology is the study of poisons. It is a scientific discipline built up from centuries of human curiosity to understand their environment. It is now formalised into a process to determine the hazard of chemicals[10]. This is not only to ensure the safety of consumers, workers and the environment, but also to comply with the need to satisfy regulatory requirements. Toxicology attempts to determine the hazards associated with exposure not only to humans, but also to environmental species[5]. As the testing paradigms differ e.g. testing on humans is very limited and hence surrogate species are used, whereas for the environment representative species can be considered, they are considered separately for the purposes of this introductory chapter and the thesis as a whole.

1.2.1 Human health

For a compound to harm human health, exposure at a concentration at or above the toxicity threshold is required. Relevant routes of exposure include inhalation, oral and dermal exposure[10]. Exposure itself may be through occupational use, deliberate application, or accidental. Occupational exposure includes obvious occupations, for example those involved in manufacture of chemicals and use of these chemicals in production processes, and the less obvious occupations, for example exposure to contaminated water through lifeguarding and commercial fishing[3]. Accidental exposure includes activities occurring where the person is exposed to chemicals following unintentional spills or release into the environment[11, 12]. Deliberate exposure includes the application of personal care products and pharmaceuticals to the body, the use of cleaning products[13] and compounds contained within food and drink. Exposure can be to solids, liquids or gases through the dermal layers of the skin, via inhalation, or through the diet. With regard to personal care products, exposure can be as a result of two main uses i.e. the application of leave-on products (such as creams, deodorants and perfumes etc.) and wash-off products (such as shampoos and soaps) to the skin. The safety and exposure assessment of these products takes into account (amongst other factors) the method of use, quantity of exposure and frequency of exposure.

1.2.2 Environmental toxicity

Interest in the effects of chemicals on the environment has been both as a response to protect environmental species and to ensure the quality of drinking water for human use. The post-industrialised world has come to appreciate environmental quality, and the effect of human activities on species in rivers, the soil and in the air. In environmental risk assessment the environmental impact (or risk) of chemical exposure is generally assessed based on equation (1.2).

$$Risk\ characterisation\ ratio = \frac{PEC}{PNEC} \qquad (1.2)^{14}$$

Where:

PEC is the predicted environmental concentration

PNEC is the predicted no-effect concentration

It was in the 1960s that man-made effects on the environment (and hence humans) came more to the attention of the public. One of the many "disasters" credited with raising our attention to environmental effects was the Love Canal[15, 16] disaster (Niagara Falls, New York). This saw schools and housing built on top of the Love Canal that had previously been filled with chemical waste. Increased occurrences of birth defects and health problems were noted in those living and working in the area. The issues were also brought to the world's attention through the publication in 1962 of Silent Spring[17] by Rachel Carson. This brought to public attention the effects of pesticides and pollution on the environment. It is widely credited with launching the environmental movement and facilitated the ban on the pesticide DDT in the US in 1972[18].

In order to be able to understand the effects of a chemical on the environment (and hence control them) information on species covering a range of taxa and relevant environmental compartments of concern is required (some common tests are discussed later). More fundamentally, for a chemical to have an effect on the environment, exposure to that chemical must occur. Environmental routes of exposure include waste from manufacturing processes, agriculture and water treatment plants; accidental release into the environment as well as normal use and disposal of the end-product i.e. normal use of home and personal care products ultimately end up down the drain. Although many ingredients that are released into the environment are biodegradable, due to the volume and wide dispersed use of chemicals in down the drain products, there is a potential for exposure to the environment.

Environmental toxicity is assessed using representative species. For instance, for new compounds to be registered in the European Union toxicity values are determined for a "base set" of species. Algae, invertebrates and fish are the minimum required[19].

1.3 Determination of toxicity

The toxicity of a chemical has traditionally been determined by *in vivo* testing. More modern approaches include the use of validated (or other accepted) *in vitro* experiments.

> *In vivo* (literally translated means "in life") experiments are performed within the living animal.

In vivo tests are often costly and time consuming. In addition, there are ethical issues and legislation surrounding the use of animals in the assessment of chemical safety. Therefore, there has been a move towards the use of less invasive tests, the use of fewer animals and alternatives to whole animal testing.

> *In vitro* (literally translated means "in glass") experiments are performed using isolated organs, tissues, cells or biochemical systems.

There is considerable literature on relevant tests and techniques to assess effects to environmental species and human health[5, 20]. The purpose of this thesis is not to provide a detailed review of these methods, but to show how information arising from the tests can be used for modelling. One of the purposes of a toxicity test is to develop dose-response information to provide an indication of the maximum dose where a chemical does not cause an effect (such as a No Observed Effect Concentration, NOEC or the No Observed Effect, NOEL), or a concentration which may induce a 50% response (e.g. the concentration causing 50% lethality within a population, LC_{50})[9].

The toxicity of a chemical substance can also be obtained by non-test or computational techniques. These are often referred to as *in silico* techniques and can be defined as follows:

> *In silico* experiments are performed using computational methods; Quantitative Structure-Activity Relationships (QSARs), read across and category formation are examples of *in silico* methods[21].

In this context, a (Q)SAR is a formalisation of the relationship of the effects (e.g. toxicity) of a series of chemicals and their physico-chemical and structural properties (e.g. hydrophobic, electronic and steric parameters). A Quantitative Structure-Property Relationship (QSPR) is the formalisation of a relationship between a

compounds property (e.g. aqueous solubility) and descriptors. Quantitative Structure-Retention Relationship (QSRR) is the formalisation of a relationship between chemical structure and the chromatographic retention time. In this thesis, QSAR, QSPR and QSRR relationships will be referred to as QSARs. QSARs in their wider sense are models that predict toxicity endpoints, properties and drug potency etc.. Descriptors used in a QSAR can be experimentally determined or calculated.

(Q)SARs have been developed that show reasonable predictions for various toxicological endpoints[22]. Many of these models use physico-chemical and structural properties as descriptors. The most commonly used descriptors are those which are a measure of hydrophobicity[23].

1.4 Regulations

Due to the problems associated with environmental exposure of chemicals, in addition to historical controversy over drugs (such as thalidomide), strong regulatory frameworks have been developed to ensure the safe use of chemicals. Within Europe, the European Commission is ultimately responsible for chemical regulation. For instance, the new Registration, Evaluation, Authorisation and restriction of Chemical substances (REACH) regulation, which covers the European Union, aims to promote safe use of chemicals with respect to human health and the environment, close knowledge gaps and promote non-animal testing[24]. REACH legislation specifically promotes the 3Rs for animal testing: Replacement, Refinement and Reduction[25]. Which are defined as follows:

> "Replacement: Methods which avoid or replace the use of animals in research that has the potential to cause them harm.
>
> Reduction: Methods which minimise animal use and enable researchers to obtain comparable levels of information from fewer animals or to obtain more information from the same number of animals.
>
> Refinement: Improvements to husbandry and procedures which minimise pain, suffering, distress or lasting harm and/or improve animal welfare in situations where the use of animals is unavoidable." [26]

REACH has an emphasis on limiting the requirement for testing on animals to determine chemical safety. Therefore, the requirement for alternative methods to determine accurate predictions of toxicity has increased. Specifically, REACH promotes the use of non-animal test data through the development and application of *in silico* models such as (Q)SARs and *in vitro* testing. It encourages the sharing of existing and future chemical toxicity and fate data[27]. It should be noted that the regulation of pesticides and pharmaceuticals is often different to other chemicals and is not the subject, specifically, of this thesis.

The regulatory process is supported by international efforts for human and environmental chemical safety. These have been introduced by international institutions such as the Organisation for Economic Cooperation and Development (OECD), the World Health Organisation (WHO) and the European Commission (EC). The OECD introduced the Environment, Health and Safety Programme (EHS)[28], the WHO introduced the International Programme on Chemical Safety (IPCS)[29], the WHO Pesticide Evaluation Scheme (WHOPES)[30] and the EC have introduced the Cosmetics Regulation[31].

It is appropriate to assess the safety of chemicals as part of social and corporate responsibility, and fulfils requirements of regulations. The determination of toxicity is termed "toxicity testing". Information from these tests will be used in various ways in this thesis, it is introduced very briefly below and more detail is given as required in individual chapters.

1.5 The importance of membrane partitioning in determining toxicity

Exposure to a chemical is not sufficient alone to cause harm. To cause harm, the chemical has to reach a target site of action in a sufficient concentration to produce a toxic response. The ability of a chemical to reach a site of action is related to its ability to partition across various biological membranes. There are many biological membranes relevant to toxicity. For dermal exposure the membranes of interest are within the stratum corneum. For oral exposure the membranes of interest are within the gastrointestinal tract. For inhalation the membrane of interest is within the lung.

In terms of environmental exposure, partitioning into fish or other organisms or partitioning into soil, sediment or the air are important processes. All these routes of exposure have a partitioning process as a common factor. In this sense partitioning across a membrane usually involves a compound leaving an aqueous environment, such as the water in the river, or blood etc. (an exception being direct application to the skin) and partitioning into and through the membrane. Assuming this is a passive diffusion process (as opposed to active or facilitated transport) means that given sufficient time and other factors (such as infinite concentration) the process will reach steady state. This is comparable, in theory, to the equilibrium in the determination of the octanol-water partition coefficient (log P) (assuming the same hydrophobic forces control both membrane partitioning and log P). Therefore, it has long been considered that log P, or other measures of hydrophobicity, can provide useful information regarding potential internal exposure to a chemical.

Direct study of the system or animal of interest potentially allows the determination of the partitioning of interest. More commonly a surrogate partitioning system is studied; this is especially true when considering toxicity to human health.

1.6 Relationship of hydrophobicity to partitioning through a biological membrane

For well over a century it has been acknowledged that decreasing aqueous solubility (and hence increasing hydrophobicity) has been associated with increasing biological activity and toxicity[32, 33, 34]. This very general relationship holds when the biological effect is brought about by a non-specific phenomenon (such as non-polar narcosis), or the rate-limiting step is controlled by hydrophobicity. The quantitative relationship between log P and biological effects is, of course, a cornerstone of QSAR.

The reason for the strong dependence of non-specific toxicity and transport on hydrophobicity is the structure of biological membranes. A biological membrane is a selective barrier between two environments[35], consisting predominantly of a lipid bilayer (with the hydrophobic tails on the interior of the bilayer) and proteins (most membranes are 50% (by weight) lipid[36]). These lipids consist of a hydrophobic

hydrocarbon tail and a hydrophilic polar headgroup[37]. The major phospholipid in cell membranes is phosphatidylcholine.

1.7 Hydrophobicity

Hydrophobicity is an implicit property of a chemical and literally means 'water fearing'. Thus hydrophobicity describes the way in which a compound will behave in an aqueous solution[38]. A highly hydrophilic (low hydrophobicity) compound will be highly soluble in water (a polar solvent), but less soluble in a non-polar solvent (e.g. octanol). Conversely a highly hydrophobic compound will be more soluble in a non-polar solvent than water. Thus the hydrophobicity of a compound is highly correlated with its ability to partition across membranes, bind to proteins, be metabolised and elicit toxicity[39]. Hydrophobicity also describes the capacity of a compound to partition between two phases and is an important descriptor for the prediction of toxicity[40].

1.7.1 The octanol-water partition coefficient (log P)

It is not possible to derive an absolute measure of hydrophobicity, since all measured values are relative to the system of analysis. Hydrophobicity can, however, be characterised and parameterised by measurements such as the partition coefficient (P) of a compound between two immiscible phases (polar and non-polar). The vast majority of measurements of partitioning have been made using water as the polar phase and octanol (despite its expense, toxicity and other drawbacks) as the non-polar phase. As such the logarithm of the octanol-water partition coefficient (log K_{ow}), has become the standard measure of hydrophobicity. Log K_{OW} is often referred to as log P and so it will be referred to as this throughout this thesis. Log P is defined mathematically by equation (1.3).

$$\log P = \log[x]_{oct} - \log[x]_{aq} \qquad (1.3)$$

Where:

P is the octanol-water partition coefficient

$[x]_{oct}$ is the concentration of sample in the octanol phase at equilibrium

$[x]_{aq}$ is the concentration of sample in the water phase at equilibrium

There are many experimental and predictive methods available to determine log P[41, 42] some of which are outlined below and a summary of the techniques is provided in Table 1.

Log P is applicable to a compound that is unionised under the conditions of analysis. If under the conditions of interest or analysis the compound is ionised the distribution co-efficient (log D) can be used. The pH of determination is reported when the compound is ionised under the conditions of analysis. Log D is defined by equation (1.4).

$$\log D = \log[x]_{oct} - \left(\log[x]_{aq}^{ionised} + \log[x]_{aq}^{unionised}\right) \quad (1.4)$$

Where:

D is the distribution coefficient

$[x]_{oct}$ is the concentration of sample in the octanol phase at equilibrium

$[x]_{aq}^{ionised}$ is the concentration of sample in the ionised form in the water phase at equilibrium

$[x]_{aq}^{unionised}$ is the concentration of sample in the unionised form in the water phase at equilibrium

1.7.2 Experimental methods to determine log P

1.7.2.1 Shake-flask method

Traditionally the log P of a compound has been determined using the shake-flask method. The method involves the mutual pre-saturation of water and octanol. A stock solution of the test compound, of known concentration is prepared in octanol. The two solvents and the test compound are mixed by shaking for 24 hours. Following shaking the phases are separated by centrifugation and the concentration of the test compound in each phase is determined using compound-specific methods. Techniques for the analysis of the aqueous, or octanol layer (ideally both phases are analysed for the determination of sample concentration) include gas chromatography (GC), high performance liquid chromatography (HPLC) and ultraviolet/visible (UV/vis) spectrometry. The shake flask method is suitable for analysis over the log P range -2 to 4. A standardised method to determine the octanol-water partition

coefficient of a compound using the shake flask method is available as OECD guideline 107[43].

For compounds with a high log P, the slow-stir method has been developed, this is a variation on the shake flask method. The slow-stir method stirs the solvents during pre-saturation and equilibrium; this method takes 2-3 days of stirring to reach steady state[44, 45].

The shake-flask and slow-stir variation are both methods that are accurate and repeatable. However, the techniques are time-consuming per sample and the solvents need to be pre-saturated prior to analysis. In addition, for extremely hydrophilic or hydrophobic compounds it is difficult to determine the concentration of the sample in one of the phases due to the low concentration that would be present. The OECD guideline for the determination of log P using the shake-flask method specifically excludes surface active agents[46]. It has been shown by Short *et. al.*[47] that the determination of log P for some surfactants is possible; however, the determination is not straight forward.

1.7.2.2 **HPLC method**

HPLC is defined and described in Section 1.8. For the determination of log P a Reverse Phase-High Performance Liquid Chromatography (RP-HPLC) method has been standardised and is available as OECD guideline 117[48]. The method is suitable for determining log P in the range 0 to 6. It is capable of determining log P for neutral compounds and ionisable compounds in their unionised free form. It is not suitable for the analysis of strong acids or bases, metal complexes or surface active agents. The RP-HPLC method requires the analysis of reference compounds to construct a calibration graph, of which there are 60 listed in the OECD guideline. The test compound's log P is then determined by interpolation. The RP-HPLC method of determining log P is faster than the shake flask method. Also due to the nature of HPLC, sample purity is less of an issue and mixtures can be analysed.

1.7.2.3 **Potentiometric titrations**

The log P of ionisable compounds can be determined experimentally using acid-base titrations. The titration is first performed under aqueous conditions to determine the

pKa. The titration is then repeated with n-octanol added to the aqueous phase to determine an apparent pKa value. From these two values the log P can be calculated using equations (1.5) and (1.6) for monoprotic acids and bases respectively.

$$P = (10^{p_oKa - pKa} - 1)\frac{V_{H_2O}}{V_{org}} \qquad (1.5)^{49}$$

$$P = (10^{(pKa - p_oKa)} - 1)\frac{V_{H_2O}}{V_{org}} \qquad (1.6)^{49}$$

Where:

p_oKa is the apparent pKa in the presence of organic solvent

pKa is the pKa in water

V_{H2O} is the volume of water

V_{org} is the volume of organic solvent

This method is relatively quick for individual samples. However, the method requires high purity samples and is not suitable for surface active compounds, volatile or light sensitive compounds. The method of determining log P for ionised compounds using potentiometric titrations has been standardised and is available as OECD guideline 122[50].

There are other methods to allow for the determination of log P for example solid phase micro extraction[51], electrochemical determination[52] and filter probe technique[53]. The methods detailed above and summarised in Table 1 are all covered by OECD guidelines, the methods have been standardised and both recommendations and exclusions for the suitability of the methods for the analysis of classes of compounds are included.

Technique	Range of log P	Compounds suitable for analysis according to OECD guidelines	Compounds excluded from OECD guidelines
Shake flask	-2 to 4	Neutral and charged compounds	Surface active agents
RP-HPLC	0 to 6	Neutral and ionisable compounds in their non-ionised form	Strong acids and bases, metal complexes, substances that react with the eluent or surface active agents
Potentiometric titration	-2 to 7	Ionised compounds	Surface active agents, volatile or light sensitive compounds

Table 1 - Method of determining log P, along with the range of the methods and classes of compounds both suitable and unsuitable for determination according to the OECD guideline[43, 48, 50]

1.7.3 Predictive methods of determination

As well as measuring log P, it is desirable to predict the value because there are a large number of compounds for which experimental values are unavailable. Additionally, predicted values are useful as a tool in product development as well as the derivation of QSARs. Log P can be calculated based on the structure of the compound. The ability to calculate and predict hydrophobicity based on the structure of a compound is possible due to the additive nature of hydrophobicity. There are many methods to calculate log P based on different approaches.

The reductionist[54] approach is a statistical method that calculates fragment values, based on experimental log P values and the structure of the compounds. The constructionist[55] approach builds a molecule up from experimental values for fundamental fragments (e.g. CH_4, C_2H_6 and H_2). Neural networks[56] are another statistical approach for determining log P; however, the approach is less transparent.

Many of the methods to calculate log P have been computerised. KOWWIN[57] is an example of an atom and fragment approach that is freely available as part of the

EPISuite programme[57]. Some methods are available as standalone programmes whilst others are available on-line, Virtual Computational Chemistry Laboratory (VCCLAB)[58,59] is an example of a web-based platform that calculates log P using a number of different methods.

1.7.4 Limitations of log P

1.7.4.1 Determination of log P

Based on the methods to determine log P described above, it is clear that each method has a domain of applicability and there are advantages and disadvantages to each method (refer to Table 1). For example, RP-HPLC and shake-flask methods are both accurate and reliable. However, neither method is recommended for the determination of log P for surface active agents. Despite many methods being available to determine log P, many compounds are outside the domain of a standardised method including surfactants and compounds with extreme log P values (less than -2 and greater than 6). Additionally there are numerous studies that indicate that results obtained from the various available log P methods can vary widely for a single compound[42, 60]. Therefore, consistency is required for these methods to provide accurate analytical values.

Surfactants are organic molecules and have the general form of a hydrophilic head group and a long hydrophobic tail. Surfactants align themselves at the interface of hydrophilic/hydrophobic phases, i.e. at the water/oil or water/air interface. One reason surfactants are outside the domain of most methods to determine log P is the difficulties involved in measuring properties of surfactants. This is due to their partitioning between phases. At low concentration this causes the surfactant to become ordered at the interface. At higher concentrations i.e. above a certain critical concentration, referred to as the critical micelle concentration (CMC) they form micelles (CMC is temperature dependent). Micelles are aggregations of surface active molecules. In polar solvents the hydrophilic heads remain in contact with the polar solvent and the hydrophobic tails are protected at the centre of the micelle. Upon the formation of micelles the thermodynamic and physical properties of the system change. This makes accurate measurement of partition coefficients particularly difficult, as it is recommended that measurements are obtained in the presence of surfactant in its free form i.e. below the CMC. Surface active agents

partition at the interface of the aqueous and octanol layers. As the effect of this "between phases" partitioning occurs at the interface of the aqueous and octanol layers, the effect is relative to the ratio of the solvent surface area and the bulk solvent. Therefore, the effect of between phases partitioning is small and quantifying the concentration of surfactant in each phase of the bulk solvents should be a reasonable measure of hydrophobicity. Additionally Roberts[61] has reported that the formation of micelles does not affect the measured log P value determined because hydrophobicity is the ratio of sample concentration in water and octanol at equilibrium; this *ratio* is not affected by the presence of micelles.

1.7.5 Use of hydrophobicity as a surrogate to membrane permeability

There are many examples, including those given in this thesis, of hydrophobicity being used as a useful surrogate for membrane permeability. However, in addition to the experimental limitations in the determination of log P itself, there are also limitations in the use of log P as a descriptor for the prediction of toxicity (to both human health and the environment).

A large number of QSARs for toxicological endpoints are based on hydrophobicity as a descriptor. As stated above, this is because it is often assumed that hydrophobicity is important in such models as it provides an estimate for the chemical's passage through, and accumulation in, biological membranes. It can also provide a measure of receptor binding. Whilst log P is a good surrogate for hydrophobicity, accurate determination of log P does not necessarily lead to accurate *in silico* predictions. This is due, in part, to the differences between octanol-water partitioning and the partitioning of a chemical across biological membranes[39]. Octanol-water partitioning mimics a single partitioning across one membrane[62]. However, in biological systems partitioning occurs across multiple membranes. These limitations in octanol-water partitioning led to the development of different methods to determine partitioning[63].

Leahy *et al.*[64] argued that it is improbable that all membranes possess the same physical characteristics. To overcome this they proposed a 'critical quartet' of solvent-water systems. The four partitioning systems proposed use different solvents

for the organic phase; these are alkane (inert), octanol (amphiprotic), chloroform (hydrogen bond donor), propylene glycol dipelargonate (PGDP) (hydrogen bond acceptor). Table 2 shows the partitioning values across the four solvent systems for selected aromatic compounds. Partitioning of a compound in different solvent systems can be relatively more or less hydrophobic. For example phenol in the alkane/water system has a partition coefficient of -0.87, however in the octanol-water system the value is 1.47. This illustrates that in an alkane-water system, phenol is relatively hydrophilic in nature, but in the octanol-water system the phenol appears more hydrophobic in nature.

Compound	**Log P_{alk}**	**Log P_{oct}**	**Log P_{CHCl3}**	**Log P_{PGDP}**
Benzene	2.24	2.13	2.80	2.36
Phenol	-0.87	1.47	0.36	1.17
Aniline	-0.04	0.90	1.42	0.95
Toluene	2.89	2.73	3.41	2.89
Benzoic acid	-0.84	1.87	0.46	1.15
Nitrobenzene	1.44	1.85	2.93	2.16
Anisole	2.06	2.11	3.12	2.41

Table 2 - Compound and partition coefficient determined for water and alkane, octanol, chloroform and propylene glycol dipelargonate (PGDP)[64]

There are other solvent systems such as cyclohexane/water[65, 66] and liposome/water[67] which have their own specific advantages and disadvantages. In addition, various systems, using artificial membranes have been developed, including Immobilised Artificial Membranes (IAM) described in Section 1.9 below.

1.7.6 Use of hydrophobicity in modelling human health and environmental endpoints

Due to hydrophobicity being of fundamental importance in producing toxicological effects and enabling penetration through membranes, many QSARs are derived around descriptors for hydrophobicity. Log P is the most commonly applied descriptor and both measured and calculated values are used. The role of hydrophobicity in QSARs for skin penetration and aquatic toxicity is introduced below and developed further in Chapters 6 and 7 respectively. In addition,

possibilities for the use of another, ideally more biologically relevant, measure of hydrophobicity are described.

1.7.6.1 Skin absorption

One of the indicators of toxicity towards human health is skin absorption. Skin absorption can be described by penetration, which is the ability of a compound to pass through the stratum corneum. Potts and Guy[68] amongst others have developed QSARs to predict skin penetration based on molecular weight and hydrophobicity. The Flynn[3] dataset has been modelled extensively by various groups, Moss and Cronin[69] reported the QSAR (1.7) using both log P and molecular weight as descriptors using this dataset. Hydrophobicity models the partitioning of the compound through the membrane.

$$\log K_{p\,(cm/s)} = 0.74 \log P - 0.0091\, MW - 2.39 \qquad (1.7)$$

n = 116, s = 0.42, $r^2_{(adj)}$ = 0.82, F = 266

Where:

K_p is skin permeability coefficient

MW is the molecular weight

n is the number of values

s is the standard deviation

$r^2_{(adj)}$ is the square of the correlation coefficient adjusted for degrees of freedom

F is the Fisher statistic

Many QSARs have been developed to predict skin penetration using a variety of descriptors including solubility in water and octanol[70], melting point[71], H-bond acceptor and donor ability[72] and linear free-energy relationship descriptors[73]. However, the majority of QSARs also include a descriptor based on hydrophobicity.

Bouwman *et al.*[74] collated 33 publicly available QSARs for skin penetration, of the 33 QSARs presented only four were found to be suitable for use by a non-QSAR expert and would pass the OECD QSAR validation criteria[75]. The OECD has developed five validation principles for (Q)SARs[75].

The five principles are -:

- A defined endpoint
- An unambiguous algorithm
- A defined domain of applicability
- Appropriate measures of goodness of fit, robustness and predictivity
- A mechanism of interpretation (If possible)

Only one (reported by Magnusson *et al.*[76]) of the four QSARs had good predictive performance for the external test set tested by Bouwman *et al.*[74], although the r^2 derived from its training set was lower than the other three QSARs. It was concluded by Bouwman *et al.*[74] that QSARs should currently only be used for predicting skin penetration, as part of a weight of evidence approach to substantiate existing data. This improves the predictability and quality over using these QSARs as standalone methods.

1.7.6.2 **Aquatic toxicity**

Aquatic toxicity endpoints of interest depend upon the species of interest. A number of endpoints can be obtained, e.g. for acute toxicity to fish the endpoint commonly reported is LC_{50} lethality to 50% of a population. For chronic toxicity the No Observable Effect Concentration (NOEC) is a common endpoint of interest. For *Daphnia magna* the LC_{50}, EC_{50} and NOEC are commonly measured, whereas for algae it is the EC_{50} and NOEC. Many QSARs have been developed to predict these endpoints, hydrophobicity being a commonly used descriptor[77].

For prediction of aquatic toxicity, the mode and mechanism of action should be taken into account. Here mode of action refers to the way the toxic endpoint is expressed e.g. narcosis, hypersensitivity etc. Mechanism of action can be biological or chemical, and describes the known biological or chemical process by which the toxic effect is produced e.g. receptor binding, disruption of membrane[78] etc..

For example van Leeuwen *et al.*[79], assessed the early life stage toxicity of compounds to zebra fish. Compounds were separated into two classes based on their

mode of action, namely non-polar narcosis (class 1) and polar narcosis (class 2) (according to the Verhaar classification[80]). The QSARs developed reported by van Leeuwen *et al.* are shown in equations (1.8) and (1.9) for compounds with a non-polar narcosis and polar narcosis mode of action respectively.

$$\log 1/NOEC = 1.06 \log P - 4.57 \qquad (1.8)$$

$r^2_{(adj)} = 0.97$, s = 0.17

$$\log 1/NOEC = 0.66 \log P - 2.05 \qquad (1.9)$$

$r^2_{(adj)} = 0.98$, s = 0.16

Generally compounds with non-polar narcosis (class 1) or polar narcosis (class 2) mode of actions are predicted well by QSARs with hydrophobicity included as a descriptor.

1.8 High Performance Liquid Chromatography (HPLC)

The main focus of this thesis is the use of a novel analytical technique (Immobilised Artificial Membrane High Performance Liquid Chromatography (IAM-HPLC)) to provide information regarding the hydrophobicity of chemicals to develop QSAR models. Therefore, this section describes the basis of this analytical technique.

Analytical chemistry provides information on chemical substances to help isolate, purify, characterise and quantify components. Chromatography is one of the most common analytical techniques[81]. It can be used to separate components of a mixture and hence, purify a compound. This technique is also a preparative technique, used on a relatively large scale. Chromatography as a separation technique is based on a stationary phase and a mobile phase. There are many types of chromatography that have different types of mobile or stationary phase[82]. Chromatography is a historic technique which literally means "colour writing". It was invented by M. S. Tswett in 1903 for separating the pigments from plants. He used a column containing calcium carbonate and alumina as a solid stationary phase and a solvent as the liquid mobile phase to separate six pigments from plants[83].

The type of chromatography considered in the experimental work in this thesis is High Performance Liquid Chromatography (HPLC). HPLC consists of a liquid mobile phase and a solid stationary phase. The stationary phase is commonly a column packed with silica, to which a stationary phase has been chemically bonded (bare silica can also be used as the stationary phase). Columns can also be packed with polymer, this is required for analysis performed at extreme pH. The separation is based on the different affinities of the components of the sample mixture to the mobile and stationary phase. Normal phase (NP)-HPLC, consists of a polar stationary phase and a non-polar mobile phase. Reverse phase (RP)-HPLC is more commonly used and consists of a non-polar stationary phase and a polar mobile phase. Figure 1 illustrates the separation of two components in a mixture and the resulting chromatogram.

There are several types of detectors used in HPLC, the most common are Ultra Violet (UV) and diode array detectors (DAD). The UV detector consists of a UV lamp that shines through a flow cell, a fraction of the eluent from the HPLC column enters the flow cell and a change in intensity of light is measured by a photodiode. The DAD detector operates over a range of wavelengths; light passes through a grating that splits the light into individual wavelengths that are targeted at diodes accepting a narrow band of wavelengths[84]. Both these detectors require the molecule of interest to contain a chromophore. In this thesis a refractive index detector is predominantly used. The refractive index detector consists of a reference cell (which is prefilled with the mobile phase), the eluent flows through a second cell (the sample cell). Light is passed through the two cells. If the two cells differ in refractive index the photosensitive detector causes a variation in the output signal. The refractive index detector is a universal detector, widely applicable. The detector is temperature dependant, is not suitable for gradient analysis (due to the reference cell being pre filled with mobile phase) and is less sensitive than more specific detectors[81].

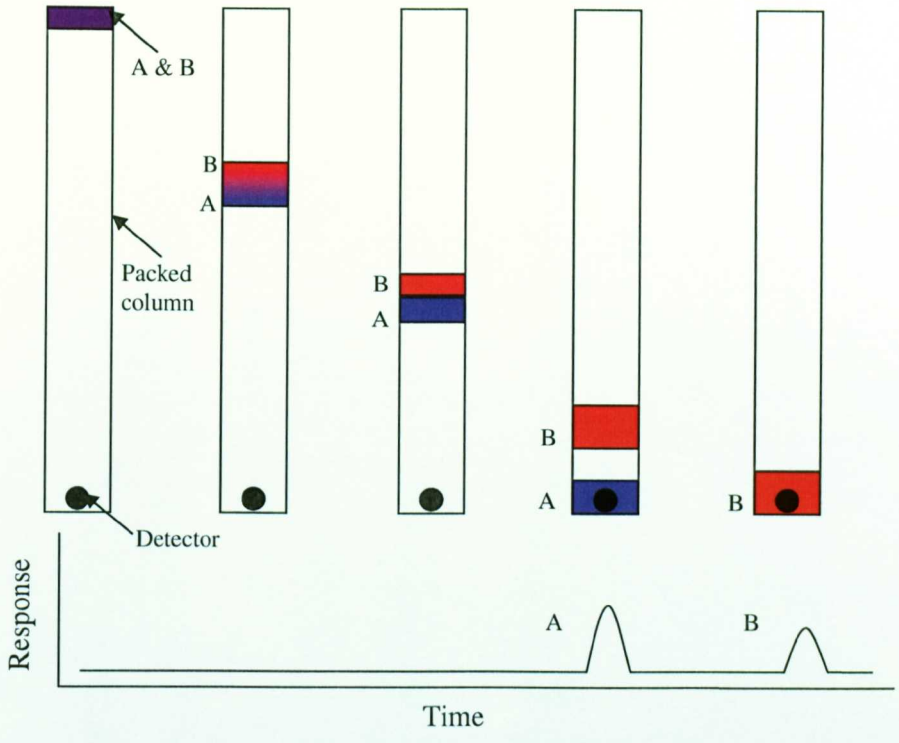


Figure 1 – Separation of two components in a mixture using chromatography and the resulting chromatogram

The output from chromatography is a retention time, to allow comparison of analysis between results obtained under similar conditions, the relative retention factor (RRF) or log k is calculated using equation (1.10).

$$\log k = log\left[\frac{(t_r - t_0)}{t_0}\right] \qquad (1.10)$$

Where:

t_r is the retention time of the sample

t_0 is the retention time of the unretained compound

In this thesis, a chromatography technique, using an Immobilised Artificial Membrane (IAM) column has been investigated. IAM-HPLC is introduced briefly below and in more detail in the relevant chapters in this thesis.

1.9 Immobilised artificial membrane (IAM) HPLC

Immobilised Artificial Membrane High Performance Liquid Chromatography (IAM-HPLC) was developed in the 1980s to determine partitioning behaviour of chemicals using a technique that mimics biological membranes more realistically than traditional techniques (octanol-water partitioning in particular)[85]. IAM-HPLC attempts to simulate the more complex partitioning across biological membrane(s), including hydrophobic and hydrophilic contributions involved in the partitioning. The IAM-HPLC column mimics both these contributions through the compound's interaction with the mobile and stationary phases. The IAM column combines hydrophobic, ion pairing and hydrogen bonding interactions in the partitioning process.

There are four IAM-HPLC stationary phases available commercially in columns of various lengths (10 to 150mm). All of the available stationary phases have a phosphatidylcholine (PC) backbone bonded to aminopropyl silica; phosphatidylcholine is the major phospholipid in cell membranes.

IAM.PC	Contains a double chain of PC with no end capping
IAM.PC.MG	Contains a double chain of PC, the residual amine groups are endcapped with methylglycolate
IAM.PC.DD	Contains a single chain PC, the residual amine groups are endcapped with C_3 and C_{10} alkyl chains
IAM.PC.DD2	Contains a double chain of PC, the residual amine groups are endcapped with C_3 and C_{10} alkyl chains[86, 87]

End capping involves deactivating the residual amine groups by converting them to alkylamide groups to improve column stability. The structures of the commercially available IAM columns are illustrated in Figure 2.

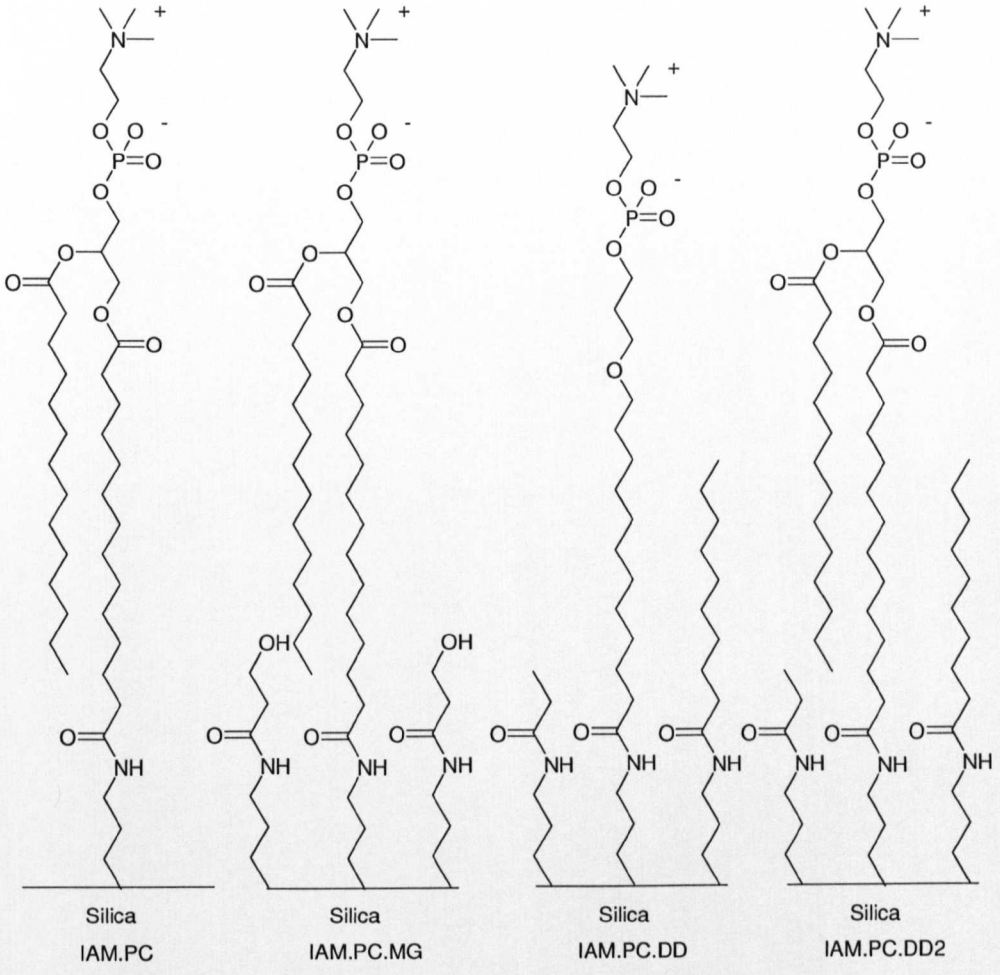


Figure 2 - Structure of stationary phase contained in the four different commercially available IAM-HPLC columns

Due to the IAM-HPLC columns containing PC, it is hypothesised that the partitioning process will be more biologically relevant than the octanol-water partitioning. The octanol-water partitioning system is traditionally accepted as an informative model for membrane partitioning, mimicking many of the intermolecular forces involved in membrane partitioning including the hydrophobic interactions, van der Waals forces, hydrogen bonding and ion-dipole bonds (Table 3)[88]. However, the octanol-water system fails to account for all of the intermolecular forces involved in membrane partitioning, the most important deficiency being the inability to account for ionic bonds[89]. Partitioning within an IAM column accounts for contributions from ion pairing and hydrogen bonding, whereas, octanol-water partitioning does not. Both techniques account for hydrophobicity. It should be noted

that although different partitioning systems may account for the same intermolecular forces, they may not do so equally.

<table>
<tr><th colspan="2">Intermolecular forces involved in membrane partitioning</th><th>Octanol-Water partitioning</th><th>IAM-Water partitioning</th></tr>
<tr><td colspan="2">Charge transfer and acryl/aryl stacking interactions</td><td colspan="2" rowspan="2"></td></tr>
<tr><td colspan="2">Ionic bonds</td></tr>
<tr><td colspan="2">Ion-dipole bonds</td><td rowspan="4">Polar interactions</td><td rowspan="3">Polar interactions</td></tr>
<tr><td colspan="2">H-bonds</td></tr>
<tr><td rowspan="3">Van der Waals forces</td><td>Orientation forces: Permanent dipole – permanent dipole</td></tr>
<tr><td>Induction forces: Permanent dipole – induced dipole</td></tr>
<tr><td>Dispersion forces: Instantaneous dipole – induced dipole</td><td>Hydrophobic interactions</td><td>Hydrophobic interactions</td></tr>
<tr><td colspan="2">Hydrophobic interactions</td></tr>
</table>

Table 3 – Intermolecular forces involved in membrane partitioning and a comparison of the interactions measured by the octanol-water and IAM-water partitioning systems[88]

It has been demonstrated that compounds which are weakly acidic or basic have low skin absorption capabilities when in the ionised form. It is possible to increase skin absorption of these compounds by the careful use of vehicles that minimise the degree of ionisation[90]. An alternative method to increase skin absorption of ionisable compounds is the use of ion pairing agents[91]. It has been suggested that it may be important to account for the intramolecular forces within ionised compounds when considering partitioning. In particular, as mentioned above, ionic bonds are not accounted for by log P, therefore, alternative measures of hydrophobicity are of interest.

IAMs are used as an HPLC stationary phase based on phosphatidylcholine, which is the major phosphlipid in cell membranes. Therefore, partitioning determined using IAM-HPLC should account more fully for the intermolecular forces involved in

membrane partitioning[89]. It is clear that hydrophobicity is relevant to toxicity, and understanding hydrophobicity can provide an insight into toxicity, as discussed in Section 1.5. In addition, partitioning into the PC stationary phase may take into account molecular size as well as hydrophobicity due to the similarities between the stationary phase and biological membranes. It is suggested that due to the charges on the chromatographic support material not being sufficiently shielded, IAM columns may be limited for the determination of log k_{IAM} values for ionic compounds[92]. However, the use of IAM columns to determine partitioning into membranes could be an improvement over the octanol-water partitioning system since this does not account for ionic bonds.

1.9.1 IAM-HPLC retention indices as a descriptor in QSARs

Good correlations have been observed between IAM-HPLC retention indices (log k_{IAM}) and penetration through biological membranes such as the blood-brain barrier[93, 94], skin[95, 96] and monocultures of CaCo-2 cells[97]. The penetration of these biological membranes affects the ability of a compound to reach the target site of action and therefore its toxicity.

Considering ecotoxicity, there have only been limited studies considering the application of IAM HPLC to determine hydrophobicity parameters used in predicting environmental toxicity endpoints. Ward *et. al*[98]. reported a good correlation of log k_{IAM} with the 48h EC_{50} toxicity data of a set of closely related homogeneous quaternary alkylammonium sulfobetaine surfactants to *Daphnia magna.*

Nasal *et al.*[96] demonstrated that for skin penetration log k_{IAM} shows greater correlation to skin permeability (log P_m) than log P, for steroids and compounds known to permeate the skin in an ionised form. However, log P was shown to have greater correlation to skin permeability of the human epidermis for phenolic compounds. Despite these successes there have been no published models for the prediction of skin penetration of surfactants. In addition, currently there are no approved, validated QSARs available for predicting skin penetration of general chemicals.

Surfactants are one of the most common groups of chemicals that enter the environment from domestic use. Despite this there are only limited data publicly available for the toxicity of surfactants. With regard to non-test data and models, there is only limited literature available correlating the log k_{IAM} of surfactants to ecotoxicity. The limited IAMs data relating to surfactants is a result of the difficulties involved in the analysis of surfactants as discussed above.

1.10 **Research aims**

There is a great need to develop QSAR models to predict toxicity to assist in the safety risk assessment of chemicals. Due to the known limitations of using log P in QSAR modelling as described above, there is considerable interest in using alternative more biologically relevant measurements of hydrophobicity. However, the common alternatives to log P (e.g IAM-HPLC) lack robust methodology and have been insufficiently evaluated in this regard. Therefore, the aim of research described herein was to develop an IAM-HPLC method to derive hydrophobicity parameters for use as descriptors for the prediction of environmental and human health endpoints.

Specific objectives to achieve that aim were:

- To collate the existing published experimental log k_{IAM} values available in the literature into a database
- To investigate the effect of experimental variability on the reported log k_{IAM} values
- To optimise a robust IAM-HPLC assay to allow the experimental determination of log k_{IAM}
- To develop methods, using the log k_{IAM} values determined experimentally, to predict IAM hydrophobicity values
- To model the ability of a molecule to cross the human skin barrier using the IAM hydrophobicity measurements in the database and those determined experimentally
- To model toxicity to organisms relevant to the aquatic environment using the IAM hydrophobicity measurements from the database and those determined experimentally
- To determine if IAM-HPLC is a suitable technique to determine the hydrophobicity of surfactants

1.11 References

[1] Timbrell J.A. (2009) *Principles of Biochemical Toxicology* 4th edition, New York, Informa Healthcare, pp 1-5.

[2] Gallo M.A. (2001) History and Scope of Toxicology. In Klassen C.D. ed, *Casarett & Doull's Toxicology The Basic Science of Poisons*, 6th edition, New York, McGraw Hill Medical, pp 3-10.

[3] Flynn G. (1990) Physicochemical Determinants of Skin Absorption. In Gerrity T.R., Henry C.J. eds, *Principles of Route-to-Route Extrapolation for Risk Assessment,* New York, Elsevier, pp 93-127.

[4] Musch A. (1996) Exposure: Qualitative and Quantitative Aspects. In Niesink J.M., de Vries J. Hollinger M.A. eds, *Toxicology Principles and Applications*, Boca Raton, CRC Press, pp 17-40.

[5] Loomis T.A., Wallace Hayes A. (1996) *Loomis's Essentials of Toxicology*, 1st edition, San Diego, Academic Press, pp 1-15.

[6] van Leeuwen C.J. (2007) General Introduction. In van Leeuwen C.J., Hermens J.L.M. eds, *Risk Assessment of Chemicals: An Introduction*, 2nd edition, London, Kluwar Academic Publishers, pp 1-36.

[7] Temkin C.L., Rosen G., Zilboorg G., Sigerist H.E. transl. (1996) ed, John Hopkins, *Paracelsus Four Treatises of Theophrastus von Hohenheim Called Paracelusus* Sigerist. H., Baltimore, University Press.

[8] Kroes R. (2004) Toxicity Testing and Human Health. In van Leeuwen C.J., Hermens J.L.M. eds, *Risk Assessment of Chemicals: An Introduction*, London, Kluwar Academic Publishers, pp 147-174.

[9] Eaton D.L., Gilbert S.G. (2008) Principles of Toxicology. In Klaasen C.D. ed, *Casarett & Doull's Toxicology The Basic Science of Poisons*, 7th edition, New York, McGraw Hill Medical, pp 11-43.

[10] Duffus J.H. (1996) Introduction to Toxicology. In: Duffus J.H., Worth H.G.J. eds, *Fundamental Toxicology for Chemists*, Cambridge, Royal Society of Chemistry, pp 1-16.

[11] Krausmann E., Renni E., Campedel M., Cozzani V. (2011) Industrial Accidents Triggered by Earthquakes, Floods and Lightning: Lessons Learned From a Database Analysis *Nat. Hazards* 59: 285-300.

[12] Giger W. (2009) The Rhine Red, The Fish Dead – The 1986 Schweizerhalle Disaster, A Retrospect and Long-Term Impact Assessment *Environ. Sci. Pollut. R.* 16: S98-111.

[13] Scow K., Wechsler A.E., Stevens J., Wood M., Callahan M.A. (1979) *Identification and Evaluation of Waterborne Routes of Exposure from Other than Food and Drinking Water*, EPA-440/4-79-016, Washington, US EPA.

[14] Karman C.C., Reerink H.G. (1998) Dynamic Assessment of the Ecological Risk of the Discharge of Produced Water from Oil and Gas Producing Platforms, *J. Hazard Matter.* 61: 43-51.

[15] Engelhaupt E. (2008) Happy Birthday, Love Canal, *Environ. Sci. Technol.* 42: 8179-8186.

[16] Paigen B. (1982) Controversy at Love Canal *Hastings Centre Report* 3: 29-37 available from < http://www.jstor.org/stable/3561826> [Accessed 3rd October 2011]

[17] Carson R. (1962) *Silent Spring*, New York, Mariner Book.

[18] Schnoor J.L. (2004) Top 10 Environmental Success Stories *Environ. Sci. Technol.* 38: 319A-319A.

[19] Hutchinson T.H., Barrett M.B., Constable D., Hartmann A., Hayes E., Huggett D., Laenge R., Lillicrap A.D., Straub J.O., Thompson R.S. (2003) A Strategy to Reduce the Numbers of Fish Used in Acute Ecotoxicity Testing of Pharmaceuticals *Environ. Toxicol. Chem.* 22: 3031-3036.

[20] Loomis T.A., Wallace Hayes A. (1996) *Loomis's Essentials of Toxicology*, 1st edition, San Diego, Academic Press, pp 208-248.

[21] Schaafsma G., Kroese E.D., Tielemans E.L.J.P., van de Sandt J.J.M., van Leeuwen C.J. (2009) REACH, Non-Testing Approaches and the Urgent Need for a Change in Mind Set *Regul. Toxicol. Pharmacol.* 53:70-80.

[22] Madden J.C. (2010) Introduction to QSAR and Other In Silico Methods to Predict Toxicology. In: Cronin M.T.D., Madden J.C. eds, *In Silico Toxicology: Principles and Applications*, Cambridge, Royal Society Chemistry, pp 11-29.

[23] Cronin M.T.D. (2006) The Role of Hydrophobicity in Toxicity Prediction *Curr. Comput. Aided Drug Des.* 2:405-413.

[24] European Commission (2006) *Off. J. Eur. Un.*, L 396/1 of 30.12.2006.

[25] European Commission (2010) *Off. J. Eur. Un.*, L 276/33 of 20.10.2010.

[26] Holmes A.M., Creton S., Chapman K. (2010) Working in Partnership to Advance the 3Rs in Toxicity Testing *Toxicology* 267: 14-19.

[27] European Commission Environment Directorate General (2007) *REACH in brief*, available from <http://ec.europa.eu/environment/chemicals/reach/pdf/2007_02_reach_in_brief.pdf> [Accessed 5th November 2010]

[28] The Organisation for Economic Cooperation and Development OECD (1998) *Savings to Governments and Industry Resulting From the Environmental Health and Safety Programme*, OECD Report ENV/EPOC/MIN (98) 5, Organisation for Economic Cooperation and Development

[29] International Programme on Chemical Safety (IPCS) (2004) The IPCS Harmonization Project: 2004 Stocktake Analysis Summary. International Programme on Chemical Safety, World Health Organization.

[30] Wright J.W. (1971) The WHO Programme for the Evaluation and Testing of New Insecticides *Bull. Wld. Hlth. Org.* 44: 11-22.

[31] European Commission (2009) *Off. J. Eur. Un.*, L 342/59 of 22.12.2009.

[32] Overton E. (1897) Osmotic Properties of Cells in the bearing on Toxicity and pharmacology *Z. Physik. Chem.* 22: 189-209.

[33] Meyer H. (1899) On the Theory of Alcohol Narcosis 1. Which Property of Anesthetics Gives Them Their Narcotic Activity? *Acrh. Exper. Pathol. Pharmarkol.* 42: 109-118.

[34] Richet C. (1893) On the Relationship Between the Toxicity and the Physical Properties of Substances *Comp. Rend. Soc. Biol.* 9: 775-776.

[35] Latash M.L. (2008) *Neurophysiological Basis of Movement*, 2nd edition, Leeds, Human Kinetics, pp 8.

[36] Yang C.Y., Cai S.J., Liu H., Pigeon C. (1996) Immobilized Artificial Membranes – Screens for Drug Membrane Interactions *Adv. Drug Deliver. Rev.* 23: 229-256.

[37] Hendry B. (1981) Membrane Physiology and Cell Excitation, Springer, Guilford, pp 14-25.

[38] Florence A.T., Attwood D. (2011), *Physicochemical Principles of Pharmacy*, 5th edition, Cornwall, Royal Pharmaceutical Society, pp 5.

[39] Florence A.T., Attwood D. (2011), *Physicochemical Principles of Pharmacy*, 5th edition, Cornwall, Royal Pharmaceutical Society, pp 141-184.

[40] Cronin M.T.D., Livingstone D.J. (2004) Calculation of Physicochemical Properties. In Cronin M.T.D., Livingstone D.L. eds, *Predicting Chemical Toxicity and Fate*, Florida, CRC Press, pp 31-40.
[41] Caron G., Ermondi G., Scherrer R.A. (2007) Lipophilicity, Polarity, and Hydrophobicity. In: Taylor JB, Triggle DJ. eds. *Comprehensive Medicinal Chemistry II*, Oxford, Elsevier, pp 425-448.
[42] Tetko I.V., Livingstone D.J (2007) Rule-Based Systems to Predict Lipophilicity In: Taylor J.B., Triggle D.J. eds., *Comprehensive Medicinal Chemistry II*, 1st edition, Oxford, Elsevier, pp 649-668.
[43] The Organisation for Economic Cooperation and Development OECD (1995) *OECD Guidelines for the Testing of Chemicals, No. 107: Partition coefficient (n-octanol/water): Shake flask method*, Paris, Organisation for Economic Cooperation and Development.
[44] Brooke D.N., Dobbs A.J., Williams N. (1986) Octanol:Water Partition Coefficients (P): Measurement, Estimation, and Interpretation, Particularly for Chemicals with $P > 10^5$ *Ecotoxicol. Environ. Saf.* 11: 251-260.
[45] de Bruijn J., Busser F., Seinen W., Hermens J. (1989) Determination of Octanol/Water Partition Coefficients for Hydrophibic Organic Chemicals with the "Slow-Stirring" Method *Environ. Toxicol. Chem.* 8: 499-512.
[46] McWilliams P, Payne G. (2002) Bioaccumulation Potential of Surfactants: A Review In: Balson, T., Craddock H.A., Dunlop J., Frampton H., Payne G., Reid P. eds. Chemistry *in the Oil Industry VII Performance in a Challenging Environment*, Cambridge, Royal Society of Chemistry, pp 44-55.
[47] Short. J, Roberts J., Roberts D.W., Hodges G., Gutsell S., Ward R.S. (2010) Partical Methods for the Measurement of Log P for Surfactants *Ecotox. Environ. Safe.* 73: 1484-1489.
[48] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-Octanol/Water), High Performance Liquid Chromatography (HPLC) Method*, Paris, Organisation for Economic Cooperation and Development.
[49] Giaginis C., Tsantili-Kakoulidou A. (2008) Alternative Measures of Lipophilicity: From Octanol-Water Partitioning to IAM Retention *J. Pharm. Sci.* 97: 2484-3004.
[50] The Organisation for Economic Cooperation and Development OECD (2000) *OECD Guidelines for the Testing of Chemicals, Proposal for a New Guideline No. 122: Partition coefficient (n-octanol/water): pH-Metric Method for Ionisable Substances*, Paris, Organisation for Economic Cooperation and Development.
[51] Dean J.R., Tomlinson W.R., Makovskaya V., Cumming R., Hetheridge M., Comber M. (1996) Solid-Phase Microextraction as a Method for Estimating the Octanol-Water Partition Coefficient *Anal. Chem.* 68: 130-133.
[52] Kontturi K., Murtomäki L. (1992) Ekectrochemical Determination of Partition Coefficients of Drugs *J. Pharm. Sci.* 81: 970-975.
[53] Barbett S.P., Hill A.P., Livingstone D.J., Wood J. (2008) A new Method for the Calculation of Partition Coefficients from Experimental Data for Both Mixtures and Pure Compounds *Quant. Struct.-Act. Relat.* 11: 505-509.
[54] Rekker R.F. (1977) *The Hydrophobic Fragmental Constant*, New York, Elsevier.
[55] Hansch C., Leo A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, New York, Wiley-Interscience Publication.
[56] Tetko I.V., Yu V.Y., Villa A.E.P. (2001) Prediction of n-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices *J. Chem. Inf. Comput. Sci.* 41: 1407-1421.

[57] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.
[58] Tetko I.V., Gasteiger J.,Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., Prokopenko V.V. (2005) Virtual Computational Chemistry Laboratory – Design and Description *J. Comput. Aid. Mol. Des.* 19: 453-63.
[59] VCCLAB, Virtual Computational Chemistry Laboratory(2005) available from http://www.vcclab.org
[60] Kerns E.H., Di Li. (2008) *Drug-like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization*, 1st edition, San Diego, Academic Press, pp 260.
[61] Roberts D.W. (2000) Aquatic Toxicity – Are Surfactant Properties Relevant? *J. Surfactants Deterg.* 3: 309-315.
[62] Dearden J.C. (1985) Partitioning and Lipophilicity in Quantitative Structure-Activity Relationships *Environ. Health Perspect.* 61: 203-228.
[63] Sangster J. (1997) *Octanol-Water Partition Co-Efficients: Fundamentals and Physical Chemistry*, New York, Wiley, pp 57-74.
[64] Leahy D. E., Morris J. J., Taylor P. J., Wait A. R. (1992) Model Solvent Systems for QSAR 2. Fragment Values (F-Values) for the Critical Quartet *J. Chem. Soc. Perk. T. 2.* 4: 723-731.
[65] Abraham M.H., Chadha H.S., Leitao R.A.E., Mitchell R.C., Lambert W.J., Kaliszan R., Nasal A., Haber P. (1997) Determination of Solute Lipopholicity, as Log P(Octanol) and Log P(Alkane) using Poly(stytene-divinylbenzene) and Immobilised Artificial Membrane Stationary Phases in Reversed-Phase High-Performance Liquid Chromatography *J. Chromatogr. A* 766: 35-47.
[66] Abraham M.H., Acree Jr. W.E., Leo A.J., Hoekman D., Cavanaugh J.E. (2010) Water-Solvent Partition Coefficients and ΔLog P Values as Predictors for Blood-Brain Distribution; Application of the Akaike Information Criterion *J. Pharm. Sci.* 99: 2492-2501.
[67] Escher B.I., Schwarzenbach R.P. (2000) Evaluation of Liposome-Water Partitioning of Organic Acids and Bases. 2. Comparison of Experimental Determination Methods *Environ. Sci. Technol.* 34: 3962-3968.
[68] Potts R.O., Guy R.H (1992) Predicting Skin Permeability *Pharm. Res.* 9: 663-669.
[69] Moss G.P., Cronin M.T.D. (2002) Quantitative Structure-Permeability Relationships for Percutaneous Absorption: Re-analysis of Steroid Data *Int. J. Pharm.* 238: 105-109.
[70] Thomas J., Majumdar S., Wasdo S., Majumdar A., Sloan K.B. (2007) The Effect of Water Solubility of Solutes on Their Flux Through Human Skin *In Vitro*: An Extended Flynn Database Fitted to the Roberts-Sloan Equation *Int. J. Phram.* 339: 157-167.
[71] Barratt M.D. (1995) Quantitative Structure-Activity Relationships for Skin Permeability *Toxic. In Vitro* 9: 27-37.
[72] Dearden J.C., Cronin M.T.D., Broohm J.K (2010) A New QSAR Model for Human Skin Permeability *J. Pharm. Pharmacol.* 61: A51-A52.
[73] Abraham M.H., Martins F. (2004) Human Skin Permeation and Partition: General Linear Free-Energy Relationship Analyses *J. Pharm. Sci.* 93: 1508-1523.

[74] Bouwman T., Cronin M.T.D., Bessems J.G.M., van de Sandt J.J.M. (2008) Eurotox Article: Improving the Applicability of (Q)SARs for Precutaneous Penetration in Regulatory Risk Assessment *Hum. Exp. Toxicol.* 27: 269-276.
[75] The Organisation for Economic Cooperation and Development OECD (2007) *OECD Environment Health and Safety Publications, Series on Testing and Assessment No.69 Guidance document on the validation of (Q)uantitative Structure-Activity Relationship [(Q)SAR] Models*, Paris, The Organisation for Economic Cooperation and Development
[76] Magnusson B.M., Anissimov Y.G., Cross S.E., Roberts M.S. (2004) Molecular Size as the Main Determinant of Solute Maximum Flux Across the Skin *J. Invest. Dermatol.* 122: 993-999.
[77] European Union (1995) Overview of Structure-Activity Relationships for environmental Endpoints Part 1: General outline and procedure, Report of the EU-DG-XII Project "QSAR for Predicting Fate and Effects of Chemicals in the Environment (contract ~ EV5V-CT92-0211).
[78] Greim H., Synder R. (2009) Introduction to the Discipline of Toxicology. In Greim H., Snyder R. eds, Toxicology and Risk Assessment a Comprehensive Introduction, Chichester, Wiley, pp 1-18.
[79] van Leeuwen C.J., Adema D.M.M., Hermens J. (1990) Quantitative Structure-Activity Relationships for Fish Early Stage Toxicity *Aquat. Toxicol.* 16: 321-334.
[80] Verhaar H.J.M., van Leeuwen C.J., Hermens J.L.M. (1992) Classifying Environmental Pollutants. 1: Structure-Activity Relationships for Prediction of Aquatic Toxicity *Chemosphere* 25: 471-491.
[81] Skoog D.A., Leary J.J. (1992) Principles of Instrument Analysis, 4th edition, Fort Worth, HBJ Saunders, pp. 628-642.
[82] Sewell P.A., Clarke B. (1987) *Chromatographic Separations*, Chichester, Wiley, pp. 1-38.
[83] Robards R., Haddad P., Jackson P. (1994) *Principles and Practice of Modern Chromatographic Method*, New York, Elsevier Science, pp 2.
[84] Lindsay S. (1987) *High Performance Liquid Chromatography*, Wiley, Chichester, pp 9-83.
[85] Pidgeon C., Venkatarum U.V. (1989) Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids *Anal. Biochem.* 76: 36-47.
[86] Lepont C., Poole C.F. (2002) Retention Characteristics of an Immobilized Artificial Membrane column in Reversed-Phase Liquid Chromatography *J. Chromatogr. A* 946: 107-124.
[87] Taillardat-Bertschinger A., Barbato F., Quercia M.T., Carrupt P.A., Reist M., La Rotonda M.I., Testa B. (2002) Structural Properties Governing Retention Mechanisms on Immobilized Artificial Membrane (IAM) HPLC Columns *Helv. Chim. Acta.* 85: 519-532.
[88] Testa B., Crivori P., Reist M., Carrupt P. (2000) The Influence of Lipophilicity on the Pharmacokinetic Behavior of Drugs: Concepts and Examples *Perspect. Drug Discov.* 19: 179-211.
[89] Taillardat-Bertschinger A., Carrupt P., Barbato F., Testa B. (2003) Immobilized Artificial Membrane HPLC in Drug Research *J. Med. Chem.* 46: 655-665.
[90] Singh I., Sri P. (2010) Precutaneous Penetration Enhancement in Transdermal Drug Delivery *Asian J. Pharm.* 4: 92-101.

[91] Takahashi K., Rytting J.H. (2000) Novel Approach to Improve Permeation of Ondansetron Across Shed Snake Skin as a Model Membrane *J. Pharm. Pharmacol* 53: 789-794.
[92] Escher BI, Schwarzenbach RP, Westall JC. (2000) Evaluation of Liposome-Water Partitioning of Organic Acids and Bases. 2. Comparison of Experimental Determination Methods *Environ. Sci. Technol.* 34: 3962-3968.
[93] Reichel A., Begley D.J. (1998) Potential of Immobilized Artificial Membranes for Predicting Drug Penetration Across the Blood-Brain Barrier *Pharm. Res.* 15: 1270-1274.
[94] Salminen T., Pulli A., Taskinen J. (1997) Relationship Between Immobilised Artificial Membrane Chromatographic Retention and the Brain Penetration of Structurally Diverse Drugs *J. Pharm. Biomed. Anal.* 15: 469-477.
[95] Barbato F., Cappello B., Miro A., La Rotonda M.I., Quaglia F. (1998) Chromatographic Indices on Immobilized Artificial Membranes for the Prediction of Transdermal Transport of Drugs *Il Farmaco* 53: 655-661.
[96] Nasal A., Sznitowska M., Buciński A., Kaliszan R. (1995) Hydrophobicity Parameter from High-Performance Liquid Chromatography on an Immobilized Artificial Membrane Column and its Relationship to Bioactivity *J. Chromatogr. A* 692: 83-89.
[97] Chan E.C.Y., Tan W.L., Ho P.C., Fang L.J. (2005). Modeling Caco-2 Permeability of Drugs Using Immobilized Artificial Membrane Chromatography and Physicochemical Descriptors *J. Chromatogr. A* 1072: 159-168.
[98] Ward R.S., Davies J., Hodges G., Roberts D.W. (2003) Applications of Immobilised Artificial Membrane Chromatography to Quaternary Alkylammonium Sulfobetaines and Comparison of Chromatographic Methods for Estimating the Octanol-Water Partition coefficient *J. Chromatogr. A* 1007: 67-75.

2 Database of literature values for IAM retention indices and experimental conditions

2.1 Introduction

Databases are available for a wide range of experimental physico-chemical parameters, e.g. the PhysProp database supporting KOWWIN and MPBPWIN[1]. KOWWIN has access to measured log P values for 13,000 compounds[2] and MPBPWIN has access to boiling points, melting points and vapour pressure values for more than 3,000 compounds[1]. However, there is no equivalent database for the logarithm of the IAM-HPLC retention index (log k_{IAM}) despite it becoming an increasingly utilised physico-chemical property. The availability of a database of log k_{IAM} values will allow ready access to log k_{IAM} values which can then be easily incorporated as a descriptor into Quantitative Structure Activity-Relationships (QSARs), to predict biological endpoints of interest. A database also allows for easier identification of the impact of experimental conditions on reported log k_{IAM} values.

2.2 Aim of the chapter

The aim of this chapter was to develop a database of log k_{IAM} values. Specifically, the objectives were to:

1. Collate the currently available IAM literature data
2. Assess the effect of experimental conditions on the variability in the reported log k_{IAM} values
3. Determine optimum experimental conditions to allow for the determination of comparable log k_{IAM} values for additional compounds

2.3 Method

2.3.1 Collation of log k_{IAM} values into a database

Data relating to IAM-HPLC were collated from the peer-reviewed scientific literature[3-55]. Specifically, log k_{IAM} results were collated along with the key documented experimental conditions under which the k_{IAM} values were obtained. Only values obtained on commercially available IAM HPLC columns were included in the database in order to allow a satisfactory comparison of results. In addition, values obtained by gradient HPLC analysis were excluded from analysis in this

thesis, but are still available within the database. These values are excluded from further analysis as they are not directly comparable with those obtained from isocratic analysis[56].

For each log k_{IAM} value the following information was recorded: the column stationary phase; length and internal diameter of the column; mobile phase; detector type (and setting if applicable); temperature; pH of the mobile phase; flow rate; injection volume; analyte diluent and concentration; the unretained compound (a compound that does not interact with the column, required to calculate RRF values (equation (1.10)); the standard compound (if applicable) and a log P value (if applicable). In addition to information extracted from the scientific papers, the Chemical Abstract Service (CAS) number, Simplified Molecular Input Line Entry Specification/Systems (SMILES) string, molecular weight were all noted. In addition to the log P value reported in the source paper, both the experimental (where available) and estimated log P values were extracted from KOWWIN v4.10[1] and included in the database.

OECD guideline 117[57] specifies 60 reference compounds with accurate log P values; these reference compounds are highlighted in the database which was compiled as a Microsoft Excel Spreadsheet.

2.3.2 Log P output from KOWWIN

When interpreting the predictions of log P from KOWWIN[1], and considering the relationship between log P and log k_{IAM} values, it is critical to ensure that there is confidence in the predictions of log P. As part of this it is essential that any predicted values fall within the domain of applicability for the model by which it is calculated[58]. Currently, there is no universally accepted applicability domain for KOWWIN for the prediction of log P[2]. However, there are several conditions to be considered when interpreting the output from KOWWIN. For the estimated log P of a molecule from KOWWIN to be considered appropriate, it must meet the following criteria[1,2]:

1. The molecular weight is within the range of the training set (18.02 to 719.92 g/mol).
2. All the fragments of the molecule being considered are in the training set and the frequency of each fragment is within the range of occurrence within the training set.
3. The log P predicted is within the range of the training set (log P -4.57 to 8.19).

If any of these conditions were not met, the compound was considered to be outside the domain of KOWWIN. Only four compounds were considered to be outside the domain of KOWWIN and were removed from further analysis, but remain in the database. The presence of the relevant fragments in KOWWIN was assessed. However, the frequency of any fragment within a compound, being greater than in the KOWWIN training set was not assessed. The four compounds that have been removed from further analysis are detailed in Table 4 along with the reason they are outside the domain of the KOWWIN training set.

Compound	**log P**		**log k_{IAM}**	**Rationale for exclusion from log k_{IAM} database**
	Experimental	**Predicted**		
Lucifer yellow (77944-88-8)	No value available	-6.79	-0.680	Predicted log P of -6.79, below log P range for the training set
Amiodarone (1951-25-3)	7.51	8.81	1.85	Predicted log P of 8.81, above log P range for the training set
Rifampin (13292-46-1)	No value available	3.90	2.881	Molecular weight of 822, above molecular weight range for the training set
Vinblastine (865-21-4)	3.7	3.42	2.555	Molecular weight of 810, above molecular weight range for the training set

Table 4 - Compounds (CAS no.) excluded from further analysis of log k_{IAM} database, experimental and predicted log P values, log k_{IAM} value and the rationale for their exclusion[1]

2.3.3 Statistical analysis of database log k_{IAM} results

The log k_{IAM} values were analysed statistically to investigate the effect of experimental variability. In order to assess the trends within the data, and provide an assessment of experimental variability, the relationship between both log k_{IAM} obtained under two different conditions and log k_{IAM} and log P was investigated by linear regression analysis using Minitab version 15.1.1.0[59]. Comparison between two log k_{IAM} under different conditions of analysis allows for direct interpretation of the effect of experimental variability, comparison of log k_{IAM} and log P compares both experimental variability and differences in partitioning between the two systems. The following statistical information was recorded for the relationships between log k_{IAM} and log P: the number of values (n); the square of the correlation coefficient adjusted for degrees of freedom ($r^2_{(adj)}$); the standard deviation (s); the Fisher statistic (F) and the Fisher statistic for a given confidence level and correct for both the number of descriptors and degrees of freedom ($F\alpha$).

The database was split into subsets, allowing variability in experimental parameters to be investigated, whilst maintaining consistent, comparable experimental conditions for the remaining parameters. Analysis of the database allowed the affects of the variability of experimental conditions on the experimental log k_{IAM} values reported to be evaluated. The effects of column stationary phase, column length, pH, temperature, flow rate and mobile phase were all investigated. The range of these experimental variables is stated in Table 5.

Experimental condition	Variables considered
Column stationary phase	IAM.PC, IAM.PC.MG, IAM.PC.DD & IAM.PC.DD2
Column length	100 & 150mm columns considered for IAM.PC.MG & IAM.PC.DD2 stationary phases
pH	3, 4.5, 5.4, 7 & 7.4 were compared for acids 3, 5, 5.4, 7 & 7.4 were compared for bases
Temperature	Combinations of temperature at 25°C & 35°C and 30°C & 45°C were considered for IAM.PC.DD2 column stationary phase Temperatures of 22°C & 25°C were considered for IAM.PC.DD column station phase
Flow rate	0.5 & 1.0 mL/min
Mobile phase	100mM phosphate buffered saline (PBS) & Acetonitrile (MeCN), 10mM ammonium acetate buffer, 50mM ammonium acetate buffer & MeCN, 10mM PBS, 10mM phosphate buffer & MeCN

Table 5- The experimental conditions investigated and the range of variables considered for each condition for log k_{IAM}

2.4 Results and Discussion

2.4.1 Overview of database

The literature search returned 1910 experimental log k_{IAM} values for 647 compounds. These are predominantly polar organic compounds and include drug molecules and surfactants. The compounds are acidic, basic and neutral, and are both ionised and unionised under the conditions of analysis. Of these, 1,686 experimental log k_{IAM} values for 555 compounds were obtained using isocratic IAM-HPLC methods with the column utilised in more than one paper. All values were included in the database. Only the isocratic log k_{IAM} values were investigated for experimental variability. Inevitably there were multiple values for the same compound, obtained under different conditions of analysis. As one of the aims of this investigation was to establish the effect of experimental conditions, and subsequent variability in the reported log k_{IAM} values, all values were considered and averages were not taken. An overview of the database is summarised in Table 6. Of the 555 compounds, 400 have experimental log P values available from the PhysProp database available in

KOWWIN[1]. For compounds where experimental log P values were not available, predicted log P values were derived from KOWWIN[1], Figure 3 shows a good correlation between experimental and predicated log P values from KOWWIN. The database of log k_{IAM} values and associated information is available in the electronic supplementary material (Chapter 2 – log k_{IAM} database.xls), included with this thesis.

Total number of experimental log k_{IAM} values	1,910
Number of log k_{IAM} values obtained under isocratic conditions	1,686
Number of compounds analysed under isocratic conditions	555
Types of compound included	Predominantly polar organic compounds and include drug molecules and surfactants
Range of compounds forms	Acidic, basic and neutral Ionised and unionised
Molecular weight range	32 to 822 g/mol (32 to 645 g/mol for analyses of the effects of experimental variability)
Log P range	-6.79 to 8.81 (-4.49 to 7.62 for analyses of the effects of experimental variability)

Table 6- Summary of the properties for compounds included in the database (not all compounds included in the database were included in the analysis of experimental variability, refer to Table 4)

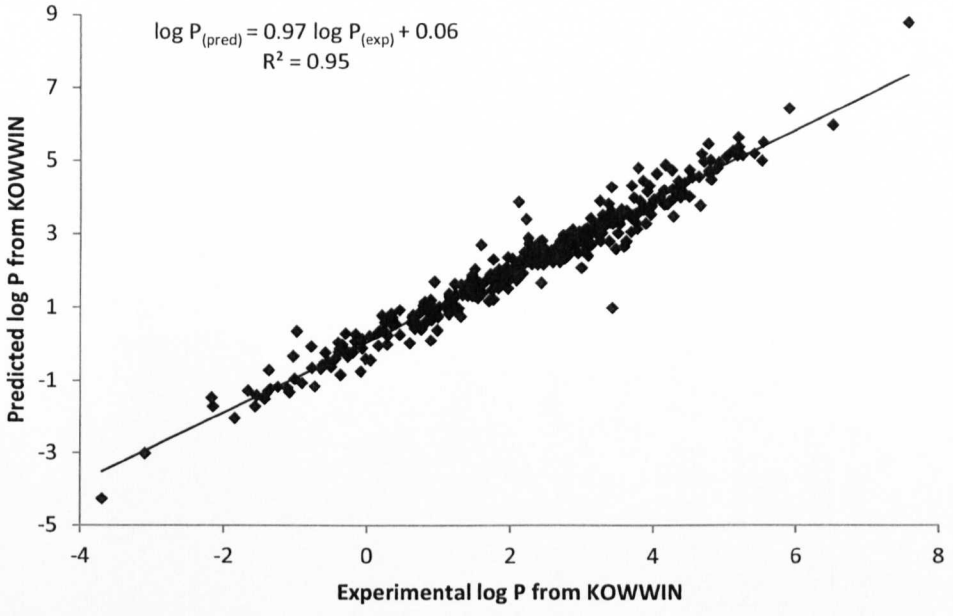


Figure 3 - Evaluation of predicted and experimental log P values from KOWWIN[1]

2.4.2 Initial analysis of database benchmarking against log P

In order to explore the effect of experimental variability on log k_{IAM} values, the data were compared with log P values. As discussed in Section 1.9 differences exist between log k_{IAM} and log P. However, they are both descriptors of partitioning. Using a fixed set of log P values allowed the experimental variability to be explored without the need to resort to multivariate statistics. Experimental log P values from KOWWIN were used. The relationship between experimental log P values from KOWWIN and experimental log k_{IAM} values is shown in Figure 4, (note that this includes multiple values for the same compounds which will influence the statistics, therefore, the associated statistics are of little interpretative value).

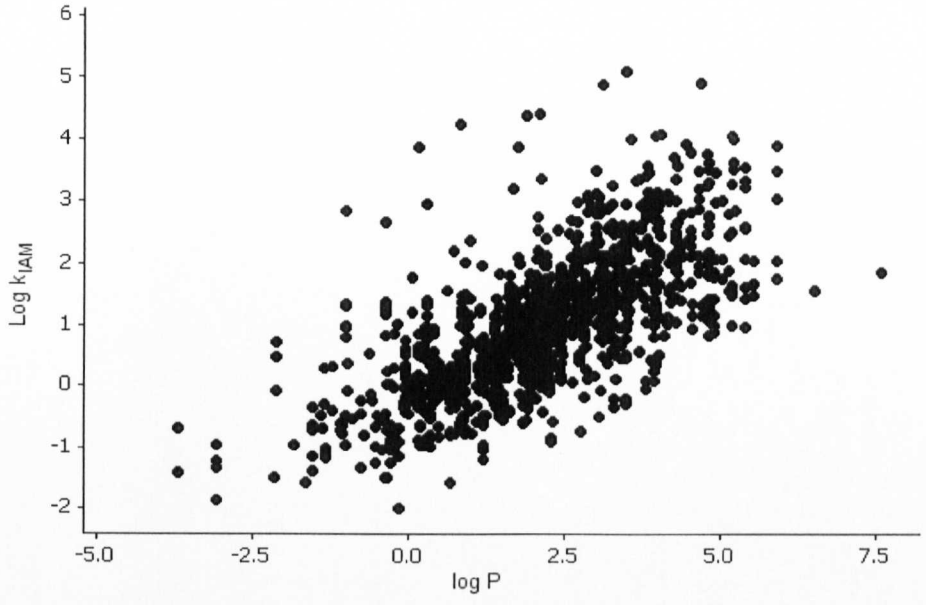


Figure 4 - Plot of experimental log P against experimental log k_{IAM} for all isocratic results taken from the database.

Whilst there is an apparent trend between increasing log k_{IAM} and log P, there is a high degree of scatter. A degree of scatter is expected, as the correlation between log P and log k_{IAM} is not perfect due to differences between the partitioning of octanol-water and membrane-water as discussed in Section 1.9. Initial linear regression analysis between these two sets of data gives the following relationship:

$$\log k_{IAM} = 0.476 \log P_{experimental} - 0.0233 \qquad (2.1)$$

n = 1444, $r^2_{adj} = 0.484$, s = 0.766, F = 1360, $F_{1, 1442}$ α, 0.001 = 1.04

2.4.3 Consideration of log k_{IAM} values available for the OECD reference compounds

The variability in the reported experimental values for log k_{IAM} is well illustrated by considering those values available for the OECD reference compounds for HPLC log P determination[57]. Figure 5 shows the plot of log k_{IAM} for the OECD reference compounds measured using the IAM.PC.DD2 column, for which log k_{IAM} has been determined multiple times. IAM.PC.DD2 log k_{IAM} values are illustrated here as they

represent the majority of values (824 of 1686 values) in the database. The range of log k_{IAM} values for some individual compounds in Figure 5 illustrates the existence of experimental variability and the effect on reported log k_{IAM} values even for well-characterised compounds (with regards to log P).

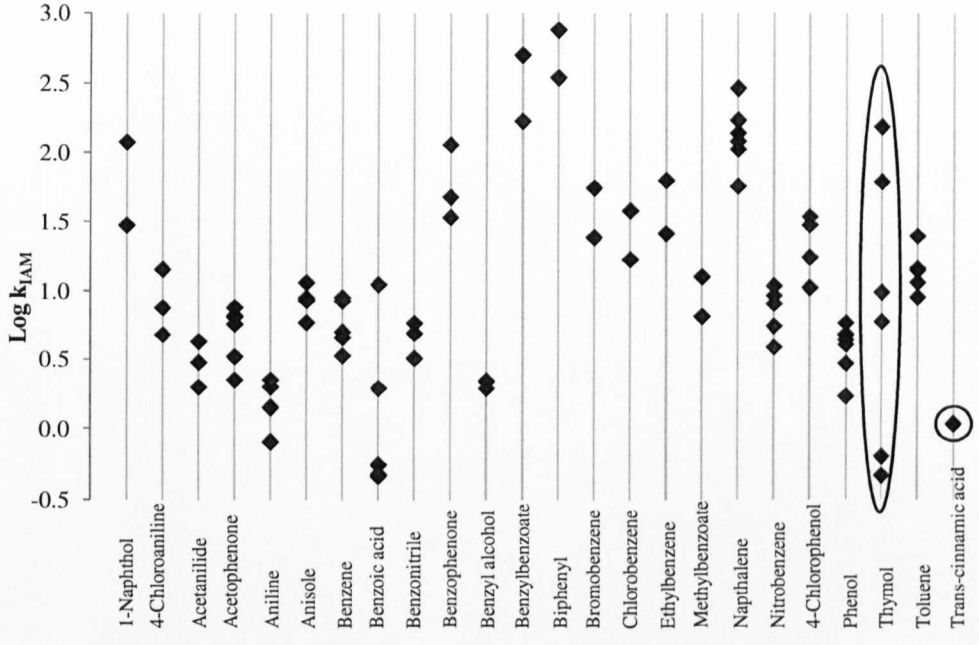


Figure 5 - Plot of log k_{IAM} measured using the IAM.PC.DD2 column for the OECD reference compounds, with multiple log k_{IAM} values reported in the database.

Log k_{IAM} values for two compounds are highlighted in Figure 5 and show the extremes of variability in the data. Thymol (log P of 3.3) has six measured log k_{IAM} values which range from -0.33 to 2.19; whereas trans-cinnamic acid (log P of 2.1) has five concordant values of log k_{IAM} at 0.04. Four of the six values for thymol were reported by Reiner[41]; for these values, two variables, temperature and organic modifier in the mobile phase, were varied. For trans-cimmanic acid the only condition that varied was the additive in the mobile phase[19]. Thus, given the concordance of the values for trans-cimmanic acid, it suggests that the additive has no effect on the measured log k_{IAM} value obtained (for this compound) (refer to 2.4.4.6 for the effect of mobile phase on log k_{IAM}). These two examples illustrate the need for consistent experimental procedures for log k_{IAM} values to be reliable, reproducible and comparable.

2.4.4 Analysis of experimental variability with regard to log k_{IAM}

The effect of the experimental variables under which the log k_{IAM} values were obtained was investigated using the log k_{IAM} values in the database.

2.4.4.1 Column stationary phase

There are four commercially available IAM HPLC columns, containing similar types of column stationary phase, as shown in Figure 2. The main commercial supplier of IAM columns is Regis Chemical Company of Morton Grove, IL, USA. They supply four main types of column in a range of column lengths. The different IAM columns and their main intended uses are as follows:

1. IAM.PC / IAM.PC.MG – Designed for protein purification
2. IAM.PC.DD – Designed for membrane permeability predictions relating to optimising bioavailability of drugs. (N.B. the IAM.PC.DD column is no longer commercially available as it has been superseded by the IAM.PC.DD2 column).
3. IAM.PC.DD2 – Designed for membrane permeability prediction in drug discovery, this column interacts more significantly with compounds that are hydrophobic in nature and so are not well retained by the IAM.PC.DD column.
4. IAM Fast-Screen Mini Column – Designed for high throughput estimation of drug permeability.

These columns all contain a phosphatidylcholine (PC) backbone bonded to aminopropyl silica. Each type of stationary phase is end-capped with a different group, which affects both the columns' separation ability and the stability of the column.

Ideally the relationship of log k_{IAM} obtained on different column stationary phases would be investigated by comparing data for common compounds against each other. However, due to the low overlap of common compounds this was not possible. Therefore, to investigate whether the column stationary phase has an effect on log k_{IAM}, the relationship between log k_{IAM} and log P was considered for each stationary phase separately. Generally, with the exception of the IAM.PC.DD stationary phase, there is an improvement in the relationship when stationary phases

are analysed individually compared to consideration of all stationary phases together. Figure 6 (a-d) shows the plot of log P experimental against log k_{IAM} for the different stationary phases. Plots a, b and c in Figure 6 show reduced scatter and an increased correlation for log k_{IAM} against log P compared to Figure 4 where all the stationary phases were considered together. Regression analysis between log k_{IAM} for each stationary phase and log P gives:

$$\log k_{IAM\,(IAM.PC)} = 0.434 \log P - 0.713 \qquad (2.2)$$

$n = 48$, $r^2_{(adj)} = 0.810$, $s = 0.382$, $F = 201$, $F_{1,46}\ \alpha, 0.001 = 12.4$

$$\log k_{IAM\,(IAM.PC.DD2)} = 0.510 \log P + 0.0591 \qquad (2.3)$$

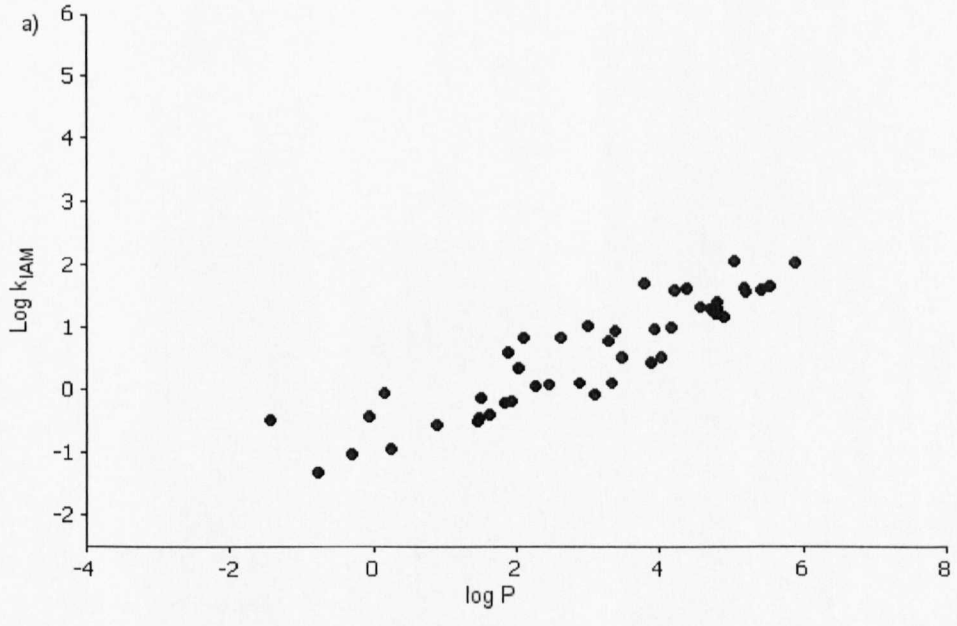
$n = 824$, $r^2_{(adj)} = 0.575$, $s = 0.685$, $F = 1120$, $F_{1,822}\ \alpha, 0.001 = 11.0$

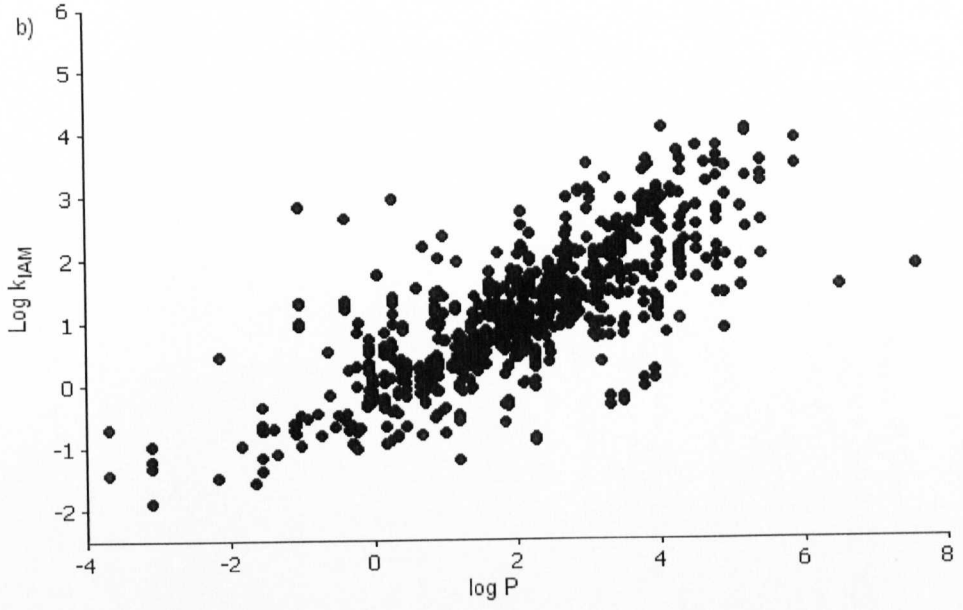
$$\log k_{IAM\,(IAM.PC.MG)} = 0.423 \log P - 0.0322 \qquad (2.4)$$

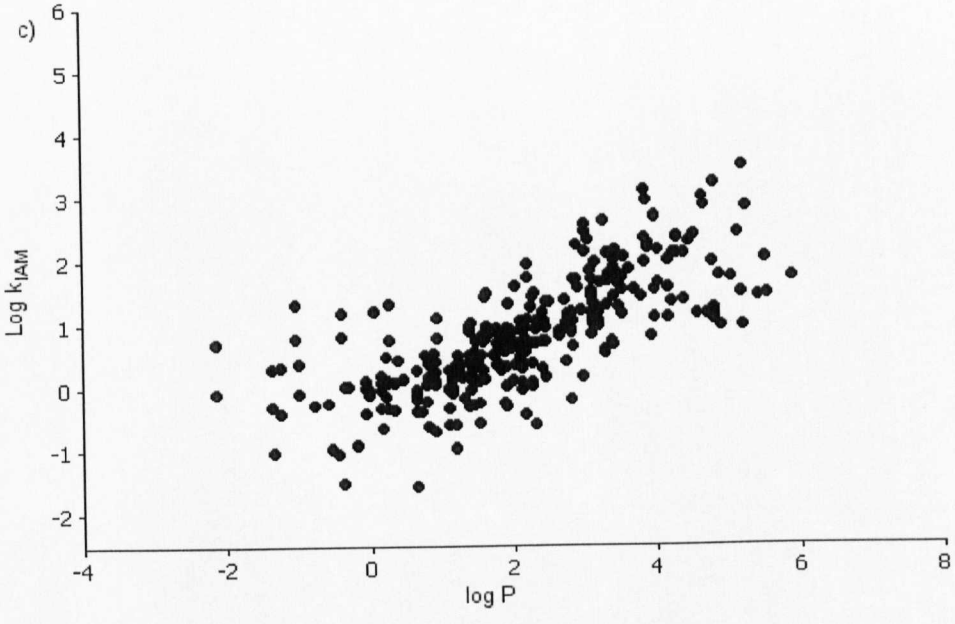
$n = 355$, $r^2_{(adj)} = 0.532$, $s = 0.607$, $F = 403$, $F_{1,353}\ \alpha, 0.001 = 11.2$

$$\log k_{IAM\,(IAM.PC.DD)} = 0.606 \log P - 0.484 \qquad (2.5)$$

$n = 215$, $r^2_{(adj)} = 0.379$, $s = 1.073$, $F = 131$, $F_{1,213}\ \alpha, 0.001 = 11.2$




b)
Log k_{IAM}
log P


c)
Log k_{IAM}
log P

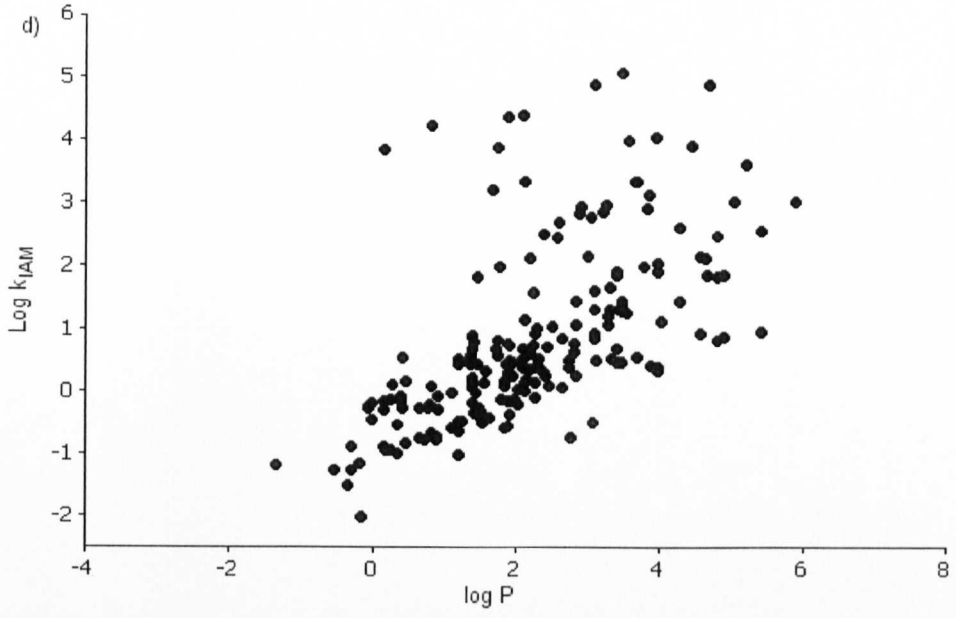


Figure 6 - Plot of experimental log P against experimental log k_{IAM} for the four commercial stationary phases, a) IAM.PC, b) IAM.PC.DD2, c) IAM.PC.MG and d) IAM.PC.DD.

The increased fit in equations (2.2), (2.3) and (2.4) over equation (2.1) indicates that the different column stationary phases, although based on the same PC backbone, interact with compounds differently depending on the columns' endcapping. Therefore, for the purpose of comparing log k_{IAM} values and considering consistent datasets, log k_{IAM} values obtained using different stationary phases should be considered separately. If log k_{IAM} was being considered as a predictor for log P, the IAM.PC column would be the column of choice.

2.4.4.2 **Temperature**

Temperature is known to be a key variable affecting HPLC elution[56]. Therefore, it should be a key parameter affecting any measured log k_{IAM} values. Temperature is also known to be a key parameter for the determination of log P[60]. K is related to temperature through the Van't Hoff equation (2.6)[61].

$$\log K = \frac{\Delta G^{\circ}}{-2.303\,RT} \qquad (2.6)$$

Where:

log K is the thermodynamic equilibrium constant

ΔG° is the free energy change (in joules per mole)

R is the Molar gas constant with a value of 8.31 J/(K·mol)

T is the temperature in kelvin

Analysis of the database indicates that the temperatures under which log k_{IAM} values were determined varied from 22-45°C. To investigate the specific effect of temperature on the k_{IAM} values, measurements for compounds analysed at different temperatures, where all other experimental conditions were the same or similar, were retrieved from the database (refer to Appendix 1.1, Table 1 for details).

Log k_{IAM} values were available for four compounds analysed at two temperatures (22 and 25°C) using the IAM.PC.DD stationary phase and a constant pH. For the IAM.PC.DD2 stationary phase, 16 pairs of results were considered for 11 compounds which were obtained using either an aqueous mobile phase or an organically modified mobile phase at two temperatures (25 and 37°C and 30 and 45°C). Figure 7 illustrates the relationship of log k_{IAM} between two temperatures for both the IAM.PC.DD and the IAM.PC.DD2 columns where all other experimental conditions were the same or similar. This shows the effect of temperature on log k_{IAM} values over the range analysed is negligible and log k_{IAM} values obtained using different temperatures are, within the temperature range considered, comparable. It should be noted that these columns are only stable up to 60°C[62]. Although temperature is important in HPLC when it is used as a separation technique, it has been shown to be negligible in this application, as it is Relative Retention Factor (RRF) values that are of interest (calculated using equation (1.10)). The temperature range considered for the IAM-HPLC column is 20-60°C. Therefore, given equation (2.6) the effect of temperature on K is small. Additionally, RRF values are considered here and the effect of temperature would affect both the unretained compound and compound of interest.

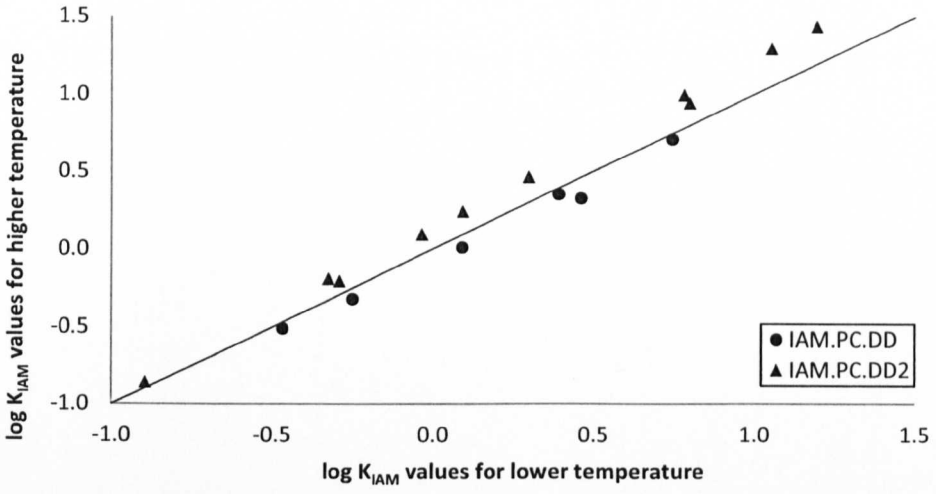


Figure 7 - Comparison of log k_{IAM} values obtained using both the IAM.PC.DD and the IAM.PC.DD2 columns at two different temperatures, and a constant pH, also plotted is the line of unity

2.4.4.3 **pH**

The pH of the mobile phase affects the degree of dissociation of ionisable compounds and hence partitioning[5, 44]. For compounds unionised under the conditions of study, log k_{IAM} is independent of the effect of pH.

To investigate the effect of pH on log k_{IAM} values, compounds with multiple results obtained on a single stationary phase at various pHs and published in a single study were investigated. The data obtained meeting these criteria were refined further to contain only compounds that had a published pKa value. Consistent log k_{IAM} values meeting these criteria were available for 18 compounds of which three were acids and 15 bases (refer to Appendix 1.1, Tables 2 & 3 for details).

To determine the effect of ionisation, log k_{IAM} values were plotted against (pKa-pH) for acids and (pH-pKa) for bases (where pKa is specific to each compound and pH relates to the mobile phase, reported in the source data and the database); this relationship is shown in Figure 8 for acids and Figure 9 for bases. The terms were derived from the Henderson-Hasselbalch equation[63], (pKa-pH), equation (2.7) for acids and (pH-pKa), equation (2.8) for bases. This provides an indication of the degree of ionisation. The compound is 50% ionised when (pKa-pH) or (pH-pKa) is 0.

$$\% \, ionised = \frac{100}{1+10^{(charge\,(pKa-pH)}} \tag{2.7}$$

$$\% \, ionised = \frac{100}{1+10^{(charge\,(pH-pKa)}} \tag{2.8}$$

All acidic compounds show a similar positive relationship between log k_{IAM} and (pKa-pH); however, the relationship is compound-dependent. All basic compounds show a positive rectilinear trend between log k_{IAM} and (pH-pKa). However, there are two obvious outliers, highlighted in Figure 9. These outlying points are for 2-phenylamine, both values were obtained using lower concentrations of phosphate buffer as the mobile phase, which may explain the differences (the effect of mobile phase on reported log k_{IAM} values is reported below).

For both acids and bases the trend between log k_{IAM} and (pKa-pH) and (pH-pKa) respectively is rectilinear and positive, however, the gradient is compound-dependent. As the degree of ionisation increases, the value of log k_{IAM} decreases. The pH of analysis is important and needs to be both considered and reported in the determination of log k_{IAM}. Log k_{IAM} results obtained at different pHs cannot be compared unless the speciation of the compound does not change significantly with the conditions of analysis.

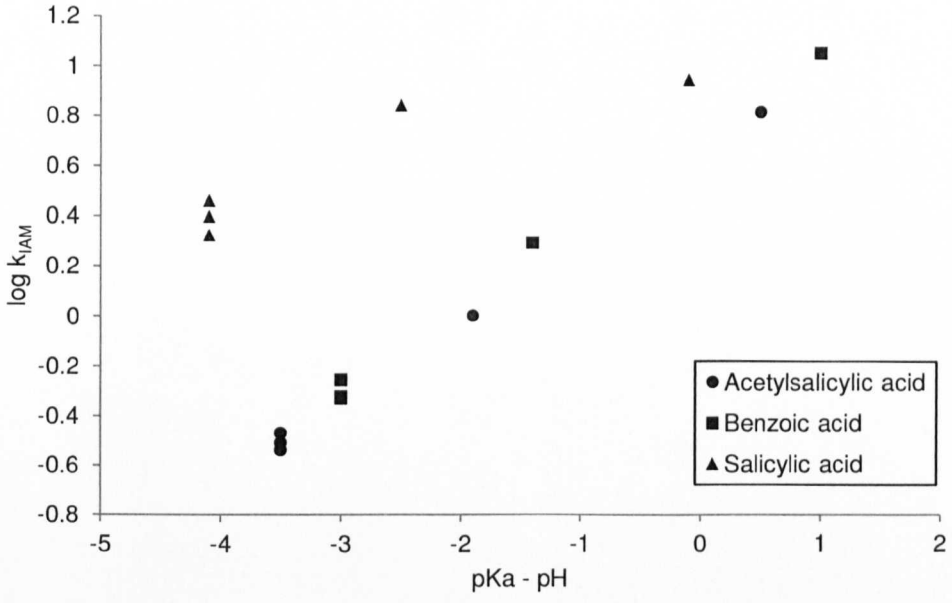


Figure 8 - Plot of log k_{IAM} against (pKa-pH) for selected acids which have experimental results obtained under similar conditions reported in a single study.

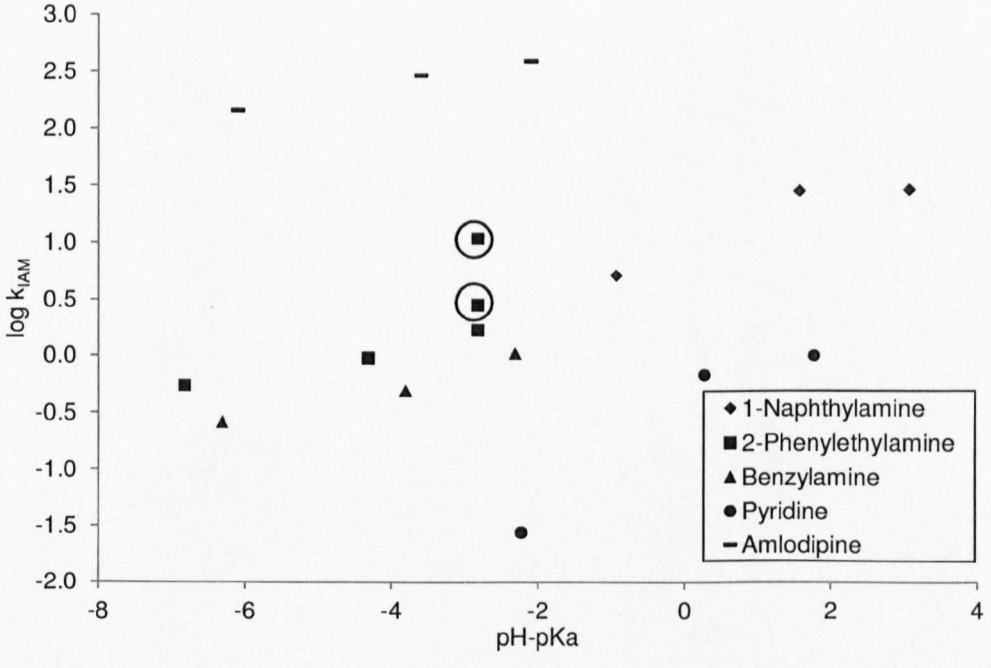


Figure 9 - Plot of log k_{IAM} against (pH-pKa) for selected bases which have experimental results obtained under similar conditions reported in a single study. Circled are outliers for 2-Phenylethylamine.

2.4.4.4 Flow rate

Log k_{IAM} values are RRF values, as described by equation (1.10). A lower or higher flow rate will affect both the unretained compound and the compound of interest to the same extent; therefore, flow rate should not affect the log k_{IAM} value reported.

Analysis of the values within the database shows that the flow rates for log k_{IAM}, when specifically stated, ranged from 0.2 mL/min to 2 mL/min. Kotecha *et al.*[27] and Lázaro *et al.*[28] analysed eight compounds under similar conditions (refer to Appendix 1.1, Table 4 for details of the compounds and log k_{IAM} values, experimental conditions are described below in Table 7). The pH used in the two studies differed by 0.4 log units; however, since the pH of determination was significantly different from the pK_a, it is not anticipated to have any effect. Log k_{IAM} values for the compounds measured at different flow rates are plotted in Figure 10. This indicates that there is no significant difference in the log k_{IAM} values at the different flow rates demonstrating that flow rate does not affect the RRF based on the limited data available.

Experimental condition	**Data from Kotecha *et al.*[27]**	**Data from Lázaro *et al.*[28]**
Column stationary phase	IAM.PC.DD2	
Mobile phase	Phosphate buffer and acetonitrile as organic modifier	
pH	7.4	7.0
Flow rate at 25°C	0.5 mL/min	1.0 mL/min

Table 7 - Experimental conditions for flow rate analysis of log k_{IAM}

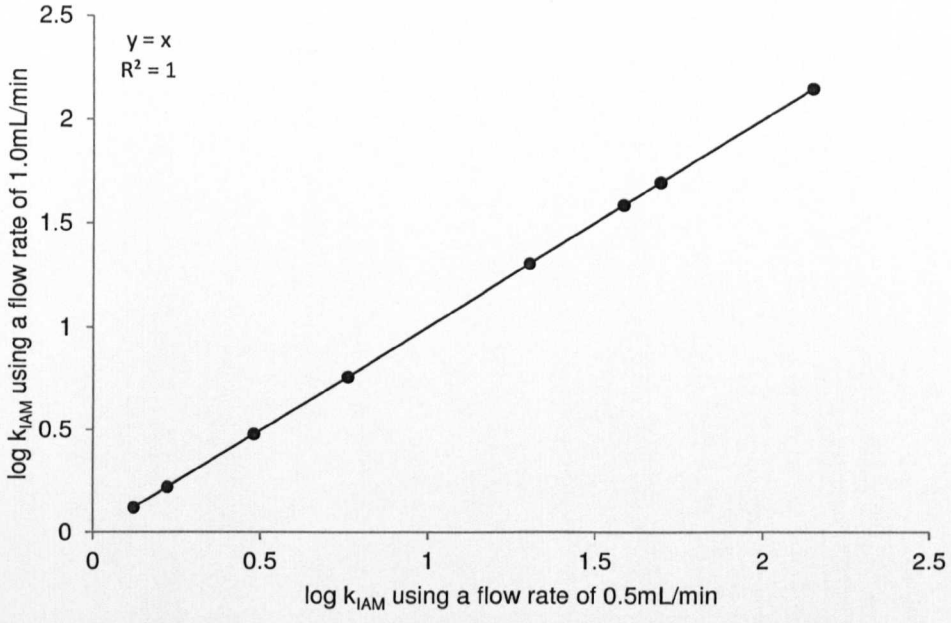


Figure 10 - Comparison of log k_{IAM} for compounds measured at a flow rate of 0.5mL/min and 1.0mL/min, data from references 27 and 28, and detailed in Appendix 1.1, Table 4.

2.4.4.5 Column length

Due to the nature of HPLC methods, there should be no effect of column length on log k_{IAM} values obtained. Provided all other experimental parameters are comparable. Since log k_{IAM} is a relative retention factor, equation (1.10). Therefore, a shorter or longer column will affect both the unretained compound and the compound of interest to the same extent. The database was searched for compounds obtained using columns of different length, under otherwise, comparable experimental conditions. For the IAM.PC.MG column eight compounds were analysed on a 150column and 100mm column (refer to Appendix 1.1, Table 5 for details of specific compounds and the papers the log k_{IAM} values were obtained from). Log k_{IAM} values obtained using columns of different length is shown in Figure 11, a 1:1 correlation is observed, indicating that log k_{IAM} values obtained on columns of different lengths are comparable.

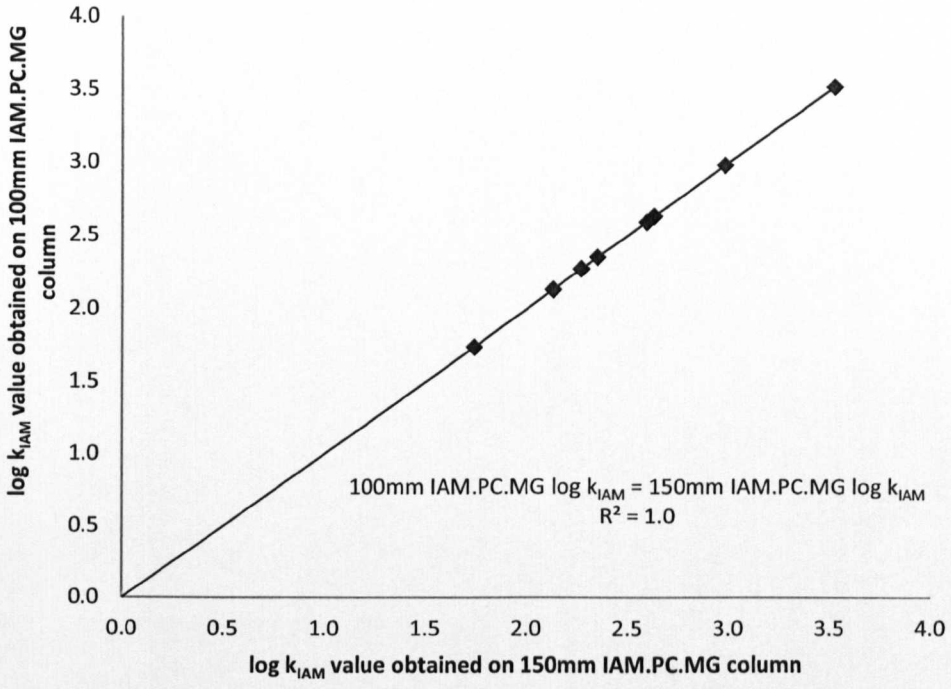


Figure 11 - Comparison of log k_{IAM} for compounds measured using a 100mm and 150mm IAM.PC.MG column, data detailed in Appendix 1.1, Table 5

2.4.4.6 Mobile phase

Ideally the effect of mobile phase on log k_{IAM} values would be investigated by comparing log k_{IAM} values obtained with different mobile phases. However, this requires common compounds to be analysed using different mobile phases, but under otherwise comparable experimental parameters. There were insufficient compounds fulfilling this criterion in the database. Therefore, in order to investigate the effect of mobile phase, log k_{IAM} values were compared to log P. Log k_{IAM} values were collated from four papers[7, 25, 31, 32]. The data were obtained using the IAM.PC.DD2 column stationary phase, pH at 7.0 or 7.4 and at either room temperature or 30°C (refer to Appendix 1.1, Table 6 for details of specific compounds and the papers the log k_{IAM} values were obtained from). These four studies included data from three different mobile phases (100mM phosphate buffer with acetonitrile, 10mM ammonium acetate buffer and 50mM ammonium acetate buffer with acetonitrile). Figure 12 shows the relationship between log P and log k_{IAM} for the data from the three different mobile phases. The gradient of log k_{IAM} against log P is different for each mobile phase. The difference in gradient illustrates that there is an effect on log k_{IAM} values of mobile phase. Therefore, it may be

concluded that log k_{IAM} values obtained using different mobile phases are not comparable directly. It may be possible to compare different mobile phases, using a correction factor, provided there are enough common compounds analysed to determine a correction factor. This may be useful in the development of QSARs, where large datasets are desirable. The application of correction factors to compare log k_{IAM} values obtained using different mobile phases was not investigated here.

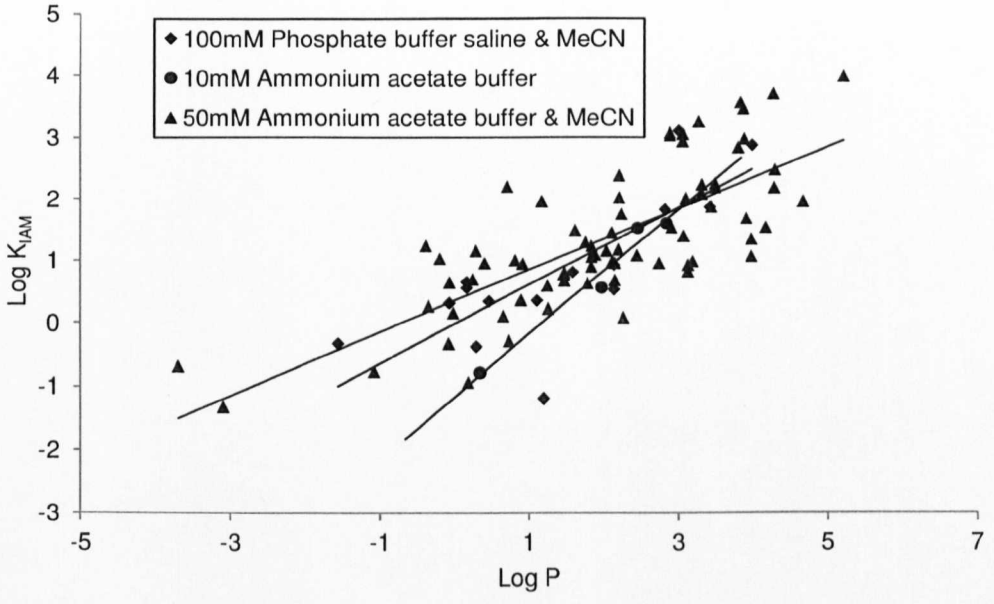


Figure 12 - Plot of log P against log k_{IAM} for three different mobile phases and the trendline for each mobile phase

2.4.5 Development of a reduced dataset of log k_{IAM} values

Figure 4 demonstrates that there is a significant positive trend between log k_{IAM} and log P, however, with a high degree of scatter. By reducing or eliminating experimental variability and considering log k_{IAM} values from different stationary phases separately (Figure 6, a-d), the data for log k_{IAM} are more consistent and comparable. In order to investigate the structural basis of k_{IAM} more thoroughly, a reduced dataset was formed for which column stationary phase, mobile phase and pH are consistent (values reported in Table 8). The IAM.PC.DD2 column was chosen due to the number of published results within the database (Of 1686 experimental log k_{IAM} values in the database 824 were obtained using the IAM.PC.DD2 column), of the columns considered only the IAM.PC column

provided an improved correlation with log P. The large number of existing experimental log k_{IAM} values obtained using the IAM.PC.DD2 column is also advantageous for the development of QSARs to model toxicity and permeability of compounds.

Variable	**Condition**
Column	IAM.PC.DD2
Column length	N/A
Mobile phase	Phosphate buffer with organic modifier as required
Flow rate	N/A
pH	7.4
Temperature	N/A (below 60°C due to column stability)

Table 8 – Conditions that were standardised in the reduced dataset

As the effects of temperature, column length and flow rate are considered to be negligible, these conditions were not standardised. The reduced dataset contains 125 log k_{IAM} values for 105 compounds. As performed previously, the log k_{IAM} values for the reduced dataset were plotted against log P (Figure 13). Linear regression analysis using Minitab gave:

$$\log k_{IAM} = 0.398 \log P + 0.381 \qquad (2.9)$$

$n = 89$, $r^2_{(adj)} = 0.532$, $s = 0.672$, $F = 100$, $F_{1,87}\ \alpha, 0.001 = 11.7$

The improvement of equation (2.9) over (2.1) is due to the reduction in variability in the experimental procedure, i.e. a standardisation, of the conditions for which log k_{IAM} values are compared. The improvement is relatively minor, as shown by the slight reduction in the degree of scatter in Figure 13 compared to Figure 4. This is at least partly due to the removal of multiple log k_{IAM} values for individual compounds obtained under different conditions of analysis, multiple measurements are more likely to be available for more common compounds with well determined log P

values and compounds used as reference materials, during the analysis of new compounds.

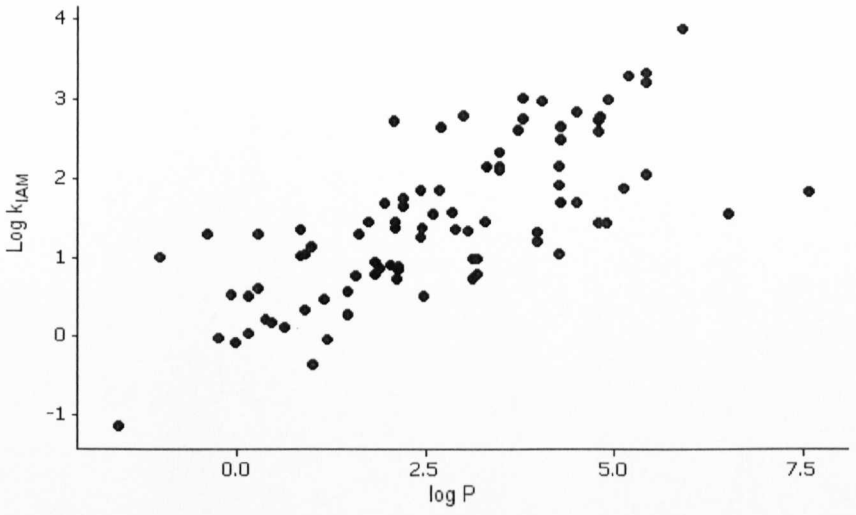


Figure 13 - Plot of experimental log P values against experimental log k_{IAM} values for the reduced dataset (variables selected are IAM.PC.DD2 column, pH 7.4 and phosphate buffer (with organic modifier as required) as the mobile phase).

The reduced dataset does still contain multiple log k_{IAM} for some compounds. Taking the mean values for the log k_{IAM} replicates in the reduced database has little effect on the regression analysis, regression analysis of the mean values gave:

$$\text{Log } k_{IAM} = 0.389 \log P + 0.444 \qquad (2.10)$$

$n=70$, $r^2_{adj}=0.543$, $s=0.662$, $F=82$, $F_{1,68}\ \alpha, 0.001 = 12.0$

Both reduced datasets are available in supplementary material (Chapter 2 – log k_{IAM} database.xls).

2.5 Conclusions

This study aimed to compile literature values for log k_{IAM} and assess the effects of experimental variability. The purpose of analysing log k_{IAM} values and their comparison with log P was not to develop the relationship with log P, but to compare and benchmark log k_{IAM} values with a similar, well characterised and accepted measure of Hydrophobicity. The effect of experimental variability on log k_{IAM} values was investigated through the comparison of log k_{IAM} values obtained under

consistent experimental procedures, with the exception of the variable of interest. Where this was not possible, due to insufficient results obtained under comparable experimental conditions, log k_{IAM} values were compared to log P. It should be noted that this comparison was not ideal, as variability between the two measures of hydrophobicity could be caused by either of or both the experimental variability in the procedure and partitioning differences between octanol/water and membrane/water.

A total of 1910 experimental log k_{IAM} values for 647 compounds have been collated into a database, of which 1686 were isocratic experimental literature log k_{IAM} values for 555 compounds. Analysis of the isocratic values in the database shows considerable variation in experimental parameters for determining log k_{IAM}. The effect of column stationary phase, mobile phase, temperature, pH, flow rate and column length were investigated. The results of this investigation showed that values obtained using columns of different stationary phase and/or mobile phases need to be considered separately to obtain consistent log k_{IAM} values. In addition, the data also showed that, provided experimental parameters are similar, values obtained using different column lengths, flow rates and temperature over the range studied (22-45°C) can be compared directly. The effect of pH on log k_{IAM} was demonstrated to be compound dependent; analysis carried out at different pHs can be compared only if the compounds are unionised at the pH of analysis.

Based on the analysis of the data in this study, it is likely that a standardised method for the determination of log k_{IAM} would improve confidence in the measurement of hydrophobicity. The investigation has demonstrated that due to the range of experimental conditions reported in the literature, caution is recommended when combining log k_{IAM} values from different sources for purposes such as developing QSARs. Many QSARs include hydrophobicity as a descriptor. If log k_{IAM} describes biological partitioning more realistically than log P, the dataset utilised should be measured under consistent experimental conditions, to increase confidence in the output from the QSAR.

2.6 References

[1] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.

[2] Nikolova-Jeliazkova N., Jaworska J. (2005) An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KowWin. *Altern. Lab. Anim.* 33: 461-470.

[3] Abraham M.H., Chadha H.S., Leitao R.A.E., Mitchell R.C., Lambert W.J., Kaliszan R., Nasal A., Haber P. (1997) Determination of Solute Lipopholicity, as Log P(Octanol) and Log P(Alkane) using Poly(stytene-divinylbenzene) and Immobilised Artificial Membrane Stationary Phases in Reversed-Phase High-Performance Liquid Chromatography *J. Chromatogr. A* 766: 35-47.

[4] Alvarez F.M., Bottom C.B., Chikhale P., Pidgeon C. (1993) Immobilised Artificial Membrane Chromatography. Prediction of Drug Transport across Biological Barriers In Ngo NT, eds. *Molecular Interactions in Bioseparations*, New York, Plenum Press, pp 151-167.

[5] Amato M., Barbato F., Morrica P., Qiaglia F., La Rotonda M.I. (2000) Interactions Between Amines and Phosphlipids: A Chromatographic Study of Immobilized Artificial Membrane (IAM) Stationary Phases at Various pH Values *Helv. Chim. Acta.* 83: 2836-2847.

[6] Barbato F., Cappello B., Miro A., La Rotonda M.I., Quaglia F. (1998) Chromatographic Indices on Immobilized Artificial Membranes for the Prediction of Transdermal Transport of Drugs *Il Farmaco* 53: 655-661.

[7] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2005) Can Protonated β-Blockers Interact with Biomembranes Stronger than Neutral Isolipophilic Compounds? A Chromatographic Study on Three Different Phospholipid Stationary Phases (IAM-HPLC) *Eur. J. Pharm. Sci.* 25: 379-386.

[8] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2004) Prediction of Drug-Membrane Interactions by IAM-HPLC: Effects of Different Phospholipid Stationary Phases on the Partition of Bases *Eur. J. Pharm. Sci.* 22: 261-269.

[9] Barbato F., Cirocco V., Grumetto L., La Rotonda M.I. (2007) Comparison Between Immobilized Artificial Membrane (IAM) HPLC Data and Lipophilicity in n-Octanol for Quinolone Antibacterial Agents *Eur. J. Pharm. Sci.* 31: 288-297.

[10] Barbato F., La Rotonda M.I., Quaglia F. (1997) Chromatographic Indexes on Immobilized Artificial Membranes for Local Anesthetics: Relationships with Activity Data on Closed Sodium Channels *Pharm. Res.* 14: 1699-1705.

[11] Barbato F., La Rotonda M.I., Quaglia F. (1996) Chromatographic Indices Determined on an Immobilized Artificial Membrane (IAM) Column as Descriptors of Lipophilic and Polar Interactions of 4-Phenyldihydropyridine Calcium-Channel Blockers with Biomembranes *Eur. J. Med. Chem.* 31: 311-318.

[12] Barbato F., La Rotonda M.I., Quaglia F. (1997) Interaction of Nonsteroidal Antiinflammatory Drugs with Phospholipids: Comparison Between Octanol/Water Partition Coefficients and Chromatographic Indexes on Immobilized Artificial Membranes *J. Pharm. Sci.* 86: 225-229.

[13] Barton P., Davis A.M., McCarthy D.J., Webborn P.J.H. (1997) Drug-Phospholipid. 2. Predicting the Sites of Drug Distribution Using n-Octanol/Water Distribution Coefficients *J. Pharm. Sci.* 86: 1034-1039.

[14] Caldwell G.W., Masucci J.A., Evangelisto M., White R. (1998) Evaluation of the Immobilized Artificial Membrane Phosphatidylcholine Drug Discovery Column for

High-Performance Liquid Chromatographic Screening of Drug-Membrane Interactions *J. Chromatogr. A* 800:161-169.

[15] Chan E.C.Y., Tan W.L., Ho P.C., Fang L.J. (2005). Modeling Caco-2 Permeability of Drugs Using Immobilized Artificial Membrane Chromatography and Physicochemical Descriptors *J. Chromatogr. A* 1072: 159-168.

[16] Cimpean D.M., Poole C.F. (2002) Systematic Search for Surrogate Chromatographic Models of Biopartitioning Processes *Analyst* 127: 724-729.

[17] Darrouzain F., Dallet P., Dubost J., Ismaili L., Pehourcq F., Bannwarth B., Matoga M., Guillaume Y.C. (2006) Molecular Lipophilicity Determination of a Huperzine Series by HPLC: Comparison of C18 and IAM Stationary Phases *J. Pharm. Biomed. Anal.* 41: 228-232.

[18] Demare S., Roy D., Legerdre J.Y. (1999) Factors Governing the Retention of Solutes on Chromatographic Immobilized Artificial Membranes: Application to Anti-Inflammatory and Analgesic Drugs *J. Liq. Chromatogr. Relat. Technol.* 22: 2675-2688.

[19] Escher B.I., Schwarzenbach R.P., Westall J.C. (2000) Evaluation of Liposome-Water Partitioning of Organic Acids and Bases. 2. Comparison of Experimental Determination Methods *Environ. Sci. Technol.* 34: 3962-3968.

[20] Genty M., González G., Lere C., Desangle-Gouty V., Legendre J. (2001) Determination of the Passive Absorption Through the Rat Intestine Using Chromatographic Indices and Molar Volume *Eur. J. Pharm. Sci.* 12: 223-229.

[21] Hoest J., Christensen I.T., Jørgensen F.S., Hovgaard L., Frokjaer S. (2007) Computational Prediction of Solubilizer' Effect on Partitioning *Int. J. Pharm.* 329: 46-52.

[22] Hollósy F., Valkó K., Hersey A., Nunhuck S., Kéri G., Bevan C. (2006) Estimation of Volume of Distribution in Humans from High Throughput HPLC-Based Measurements of Human Serum Albumin Binding and Immobilized Artificial Membrane Partitioning *J. Med. Chem.* 49: 6958-6971.

[23] Kaliszan R., Kaliszan A., Wainer I.W. (1993) Deactivated Hydrocarbonaceous Silica and Immobilized Artificial Membrane Stationary Phases in High-Performance Liquid Chromatographic Determination of Hydrophobicities of Organic Bases: Relationship to Log P and CLOGP *J. Pharm. Biomed. Anal.* 11: 505-511.

[24] Kaliszan R., Nasal A., Buciński A. (1994) Chromatographic Hydrophobicity Parameter Determined on an Immobilized Artificial Membrane Column: Relationships to Standard Measures of Hydrophobicity and Bioactivity *Eur. J. Med. Chem.* 29: 163-170.

[25] Kangas H., Kotiaho T., Salminen T., Kostiainen R. (2001) N-in-one Determination of Retention Factors for Drugs by Immobilized Artificial Membrane Chromatography Coupled to Atmospheric Pressure Chemical Ionization Mass Spectrometry *Rapid Commun. Mass Spectrom.* 15: 1501-1505.

[26] Kępczyńska E., Bojarski J., Haber P., Kaliszan R. (2000). Retention of Barbituric Acid Derivatives on Immobilized Artificial Membrane Stationary Phase and its Correlation with Biological Activity *Biomed. Chromatogr.* 14: 256-260.

[27] Kotecha J., Shah S., Rathod I., Subbaiah G. (2008) Prediction of Oral Absorption in Humans by Experimental Immobilized Artificial Membrane Chromatography Indices and Physicochemical Descriptors *Int. J. Pharm.* 360: 96-106.

[28] Lázaro E., Ráfols C., Rosés M. (2005) Characterization of Immobilized Artificial Membrane (IAM) and XTerra Columns by Means of Chromatographic Models *J. Chromatogr. A* 1081: 163-173.

[29] Lázaro E., Ráfols C., Abraham M.H., Rosés M. (2006) Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns *J. Med. Chem.* 49: 4861-4870.

[30] Lepont C., Poole C.F. (2002) Retention Characteristics of an Immobilized Artificial Membrane Column in Reversed-Phase Liquid Chromatography *J. Chromatogr. A* 946: 107-124.

[31] Li J., Cui S., He Z. (2006) Quantitative Structure-Retention Relationship Studies Using Immobilized Artificial Membrane Chromatography I: Amended Linear Solvation Energy Relationships with the Introduction of a Molecular Electronic Factor *J. Chromatogr. A* 1132: 174-182.

[32] Li J., Sun J., He Z. (2007) Quantitative Structure-Retention Relationship Studies with Immobilized Artificial Membrane Chromatography II: Partial Least Squares Regression *J. Chromatogr. A* 1140: 174-179.

[33] Luco J.M., Salinas A.P., Torriero A.A.J., Vázquez R.N., Raba J., Marchevsky E. (2003) Immobilized Artificial Membrane Chromatography: Quantitative Structure-Retention Relationships of Structurally Diverse Drugs *J. Chem. Inf. Comput. Sci.* 43: 2129-2136.

[34] Luo H., Zheng C., Cheng Y.K. (2007) The Retention Properties of Nucleobases in Alkyl C_8-/C_{18}- and IAM- Chromatographic Systems in Relation to Log P_{OW} *J. Chromatogr. B* 847: 245-261.

[35] Masucci J.A., Caldwell G.W., Foley J.P. (1998) Comparison of the Retention Behavior of β-Blockers Using Immobilized Artificial Membrane Chromatography and Lysophospholipid Micellar Electrokinetic Chromatography *J. Chromatogr. A* 810: 95-103.

[36] Matysiak J. (2008) QSAR of Antiproliferative Activity of N-Substituted 2-Amino-5-(2,4-dihydroxyphenyl)-1,3,4-thiadiazoles in Various Human Cancer Cells *QSAR Comb. Sci.* 27: 607-617.

[37] Nasal A., Sznitowska M., Buciński A., Kaliszan R. (1995) Hydrophobicity Parameter from High-Performance Liquid Chromatography on an Immobilized Artificial Membrane Column and its Relationship to Bioactivity *J. Chromatogr. A* 692: 83-89.

[38] Ong S., Pidgeon C. (1995) Thermodynamics of Solute Partitioning into Immobilized Artificial Membranes *Anal. Chem.* 67: 2119-2128.

[39] Ottiger C., Wunderli-Allenspach H. (1999) Immobilized Artificial Membrane (IAM)-HPLC for Partition Studies of Neutral and Ionized Acids and Bases in Comparison with the Liposomal Partition System *Pharm. Res.* 16: 643-650.

[40] Pehourcq F., Jarry C., Bannwarth B. (2003) Potential of Immobilized Artificial Membrane Chromatography for Lipophilicity Determination of Arylpropionic Acid Non-Steroidal Anti-inflammatory Drugs *Pharm. Res.* 33: 137-144.

[41] Reichel A., Begley D.J. (1998) Potential of Immobilized Artificial Membranes for Predicting Drug Penetration Across the Blood-Brain Barrier *Pharm. Res.* 15: 1270-1274.

[42] Reiner N.G., Labuckas D.O., García D.A. (2009) Liphophilicity of Some GABAerfic Phenols and Related Compounds Determined by HPLC and Partition Coefficients in Different Systems *J. Pharm. Biomed. Anal.* 49: 686-691.

[43] Ross B.P., Braddy A.C., McGeary R.P., Blanchfield J.T. (2004). Micellar Aggregation and Membrane Partitioning of Bile Salts, Fatty Acids, Sodium Dodecyl Sulfate, and Sugar-Conjugated Fatty Acids: Correlation with Hemolytic Potency and Implications for Drug Delivery *Mol. Pharm.* 1: 233-245.

[44] Salminen T., Pulli A., Taskinen J. (1997) Relationship Between Immobilised Artificial Membrane Chromatographic Retention and the Brain Penetration of Structurally Diverse Drugs *J. Pharm. Biomed. Anal.* 15: 469-477.

[45] Sarr F.S., André C., Guillaume Y.C. (2008) Statins (HMG-coenzyme A Reductase Inhibitors)-Biomimetic Membrane Binding Mechanism Investigated by Molecular Chromatography *J. Chromatogr. B* 868: 20-27.

[46] Sprunger L., Blake-Taylor B.H., Wairegi A., Acree Jr. W.E., Abraham M.H. (2007) Characterization of the Retention Behavior of Organic and Pharmaceutical Drug Molecules on an Immobilized Artificial Membrane Column with the Abraham Model *J. Chromatogr. A* 1160: 235-245.

[47] Di Stefano A., Sozio P., Iannitelli A., Cerasa L.S., Fonana A., Di Biase G., D'Amico G., Di Giulio M., Carpentiero C., Grumetto L., Barbato F. (2008) Characterization of Alkanoyl-10-Ominocyclines in Micellar Dispersions as Potential Agents for Treatment of Human Neurodegenerative Disorders *Eur. J. Pharm. Sci.* 34: 118-128.

[48] Taillardat-Bertschinger A., Marca Martinet C.A., Carrupt P.A., Reist M., Caron G., Fruttero R., Testa B. (2002) Molecular Factors Influencing Retention on Immobilized Artificial Membranes (IAM) Compared to Partitioning in Liposomes and n-Octanol *Pharm. Res.* 19: 729-737.

[49] Taillardat-Bertschinger A., Galland A., Carrupt P.A., Testa B. (2002) Immobilized Artificial Membrane Liquid Chromatography: Proposed Guideline for Technical Optimization of Retention Measurements *J. Chromatogr. A* 953: 39-53.

[50] Taillardat-Bertschinger A., Barbato F., Quercia M.T., Carrupt P.A., Reist M., La Rotonda M.I., Testa B. (2002) Structural Properties Governing Retention Mechanisms on Immobilized Artificial Membrane (IAM) HPLC Columns *Helv. Chim. Acta.* 85: 519-532.

[51] Uekusa Y., Takeshita Y., Ishii T., Nakayama T. (2008) Partition Coefficients of Polyphenols for Phosphatidylcholine Investigated by HPLC with an Immobilized Artificial Membrane Column *Biosci. Biotechnol. Biochem.* 72: 3289-3292.

[52] Valkó K., Du C.M., Bevan C., Reynolds D.P., Abraham M.H. (2000) Rapid-Gradient HPLC Method for Measuring Drug Interactions with Immobilized Artificial Membrane: Comparison with Other Lipophilicity Measures *J. Pharm. Sci.* 89: 1085-1096.

[53] Vrakas D., Giaginis C., Tsantili-Kakoulidou A. (2008) Electrostatic Interactions and Ionization Effect in Immobilized Artificial Membrane Retention a Comparative Study with Octanol-Water Partitioning. *J. Chromatogr. A* 1187: 67-78.

[54] Yen T.E., Agastonovic-Kustrin S., Evans A.M., Nation R.L., Ryand J. (2005) Prediction of Drug Absorption Based on Immobilized Artificial Membrane (IAM) Chromatography and Calculated Molecular Descriptors. *J. Pharm. Biomed. Anal.* 38: 472-478.

[55] Yoon C.H., Kim S.J., Shin B.S., Lee K.C., Yoo S.D. (2006) Rapid Screening of Blood-Brain Barrier Penetration of Drugs Using the Immobilized Artificial Membrane Phosphatidylcholine Column Chromatography. *J. Biomol. Screen.* 11: 13-20.

[56] Dong M.W. (2006) *Modern HPLC for Practicing Scientists*, New Jersey, Wiley, pp 34.

[57] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-*

octanol/water), High Performance Liquid Chromatography (HPLC) Method, Paris, The Organisation for Economic Cooperation and Development.

[58] Netzeva T.I., Worth A.P., Aldenberg T., Benigni R., Cronin M.T.D., Gramatica P., Jaworska J.S., Kahn S., Klopman G., Marchant C.A., Myatt G., Nikolova-Jeliazkova N., Patlewicz G.Y., Perkins R., Roberts D.W., Schultz T.W., Stanton D.T., van de Sandt J.J.M., Tong W., Veith G., Yang C. (2005) Current Status of Method for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The Report and Recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* 33: 155-173.

[59] Minitab Inc., Minitab® Statistical Software version 15, Coventry: Minitab Inc., 2007.

[60] Dearden J.C., Bresnen G.M. (1988) The Measurement of Partition Coefficients and Lipophilicity *Quant. Struct.-Act. Relat.* 7: 133-144.

[61] Atkins P. (2001) *Atkin's The Elements of Physical Chemistry*, 3rd edition, Bath, Oxford University Press, pp 160.

[62] Personal communication with Ted Szczerba, Regis Chemical Company, Morton Grove, IL, USA.

[63] Ebbing D.D., Gammon S.D. (2009) *General Chemistry*, 9th edition, Belmont, Houghton Mifflin, pp 681.

3 Optimisation and robustness of an IAM-HPLC assay

3.1 Introduction

Chapter 2 described the collation and investigation of a database of 1910 literature log k_{IAM} values for 647 compounds (1686 log k_{IAM} values for 555 compounds obtained under isocratic HPLC conditions) obtained under a variety of experimental conditions. The experimental parameters that were recorded in the database included column stationary phase, mobile phase composition, pH of the mobile phase, flow rate of the mobile phase and column temperature. Section 2.4.4.1 demonstrated that splitting the log k_{IAM} database into subsets, on the basis of column stationary phase, gave greater consistency than when all data were compared together (with the exception of the IAM.PC.DD stationary phase). A subset of the database, where the reported log k_{IAM} values were obtained under consistent experimental conditions (i.e. analysis performed on the IAM.PC.DD2 column using a phosphate-buffered mobile phase at pH 7.4, with organic modifier as required (methanol and acetonitrile are most commonly used organic modifiers used)), indicated that still greater consistency in log k_{IAM} values could be obtained when experimental parameters were standardised, in addition to standardising the column.

It can, therefore, be concluded that greater standardisation of the method from which log k_{IAM} values are obtained will allow for the comparison of the values obtained from different laboratories. Additionally, results obtained under standardised conditions also increase the confidence in the values used in subsequent QSAR analyses and therefore, the output from the QSAR. The factors affecting experimental variability are described in detail below.

3.1.1 IAM HPLC columns

As illustrated in Figure 2, and described in Section 1.9 there are four commercially available IAM-HPLC columns. A total of 53 scientific papers contributed to the values in the k_{IAM} database and the data therein were used to investigate the effect of experimental variability. Of these 53 studies 1686 values were obtained using the IAM.PC, IAM.PC.DD2, IAM.PC.MG and IAM.PC.DD2 columns. Two additional studies (not included in the analysis of experimental variability) used a different IAM HPLC column (RexChrom IAM PC2 s-12-300-IAM-PC)[1,2]. One of these two

studies reported log k_{IAM} values obtained from a gradient IAM-HPLC method. The log k_{IAM} values from this column were not included in the investigation of experimental variability due to lack of data available for comparison. No published experimental log k_{IAM} values were available using the IAM fast-screen mini columns.

The IAM HPLC assay was optimised in this investigation using the IAM.PC.DD2 column (4.6x100mm) fitted with a guard cartridge. The IAM.PC.DD2 column was chosen following analysis of experimental variability within the k_{IAM} database and selection of experimental conditions form a reduced dataset (refer to section 2.4.5). Log k_{IAM} values obtained under the consistent experimental procedure that are comparable to the IAM HPLC assay developed in this investigation, can be modelled together, aiding in the development of methods to predict log k_{IAM}. The use of the IAM.PC.DD2 column has a number of advantages over other columns including 1) there are a greater number of published results obtained using this column; 2) it is available in a shorter length than comparable columns, leading to shorter analysis times; 3) the column is endcapped with C_{10} and C_3 alkyl chains improving column stability.

3.1.2 **Experimental parameters for HPLC**

HPLC analysis can be either Normal Phase (NP) where the stationary phase is polar and the mobile phase is non-polar, or Reverse Phase (RP) where the stationary phase is non-polar and the mobile phase is polar (Refer to Section 1.8 for a more detailed description of HPLC). The majority of HPLC analysis, including IAM-HPLC, is RP-HPLC, so a polar mobile phase is required. With regard to the mobile phase of the log k_{IAM} values in the log k_{IAM} database, the majority of studies did not exceed 60% organic modifier. Of the studies using high concentrations of organic modifier, many comment on short column lifetime. Indeed, the manufacturers do not recommend that more than 30% organic modifier is added to the mobile phase[3].

Since one of the aims of IAM-HPLC is to model the ability of compounds to penetrate biological membranes, using a biologically relevant buffer at a physiologically important pH seems appropriate. This means the log k_{IAM} values generated may have greater relevance to understanding biological data and may,

therefore, be of more use to predict effects such as toxicity. However, what would be considered a biologically relevant pH depends upon the system of interest. For instance the pH of surface river water varies. In the United Kingdom the Technical Advisory Group (UKTAG) existing standards for the pH of river water are within the range of pH 6 to 9[4]. The generally accepted pH of the human skin surface is 5.5[5], with the range of measured pH values for human skin being pH 4.2 to 5.9[6]. In addition the OECD guidelines for RP-HPLC determination of n-octanol-water partition coefficient state that:

> "If the log P_{ow} value is determined for the use in environmental hazard classification or in environmental risk assessment, the test should be performed in the pH range relevant for the natural environment, i.e. in the pH range 5.0-9."[7]

As pH can affect the extent of ionisation, the pH at which the analysis is carried out is important for compounds that are ionisable (refer to Section 2.4.4.3). A phosphate buffered saline solution at a pH of 7.4 and a concentration of 10mM was used as the mobile phase and sample diluent (for samples of low hydrophobicity) to maintain a constant pH when analysing compounds in their ionised form. In addition, the buffered saline maintained a constant pH when analysing compounds where extrapolation was required e.g. for highly hydrophobic compounds.

A pH of 7.4 is within the advisable ranges from the OECD guideline[7]; this is also the pH value for many toxicity test systems e.g. short term inhibition of growth assay to *Tetrahymena pyriformis*[8] and is an osmotic and physiological match to cells[9]. In addition, the column manufacturer recommends that the column be used within the pH range of 2.5 to 7.5, with pHs above 7.5 being detrimental to the stationary phase of the column[3].

As discussed in Section 2.4.4.2 the effect of temperature on log k_{IAM} measurement is negligible. However, it is recognised that the column manufacturers recommend the temperature of the column is maintained at 60°C or lower to ensure column stability[3].

3.2 Aim of the Chapter

The aim of this chapter was to standardise and optimise an IAM HPLC assay for the analysis of both unionised and ionised compounds (under the conditions of analysis) covering a range of hydrophobicities. The optimisation was performed using the experimental parameters as described in Section 3.3.4. The method was standardised using compounds for which reliable experimental log P values were available, i.e. standard compounds for the OECD RP-HPLC log P assay[7]. Following optimisation of the IAM HPLC method, the robustness of the method was investigated. A robust method increases confidence in use of the method and the log k_{IAM} values generated using the method.

3.3 Method

3.3.1 Materials

Methanol (HPLC gradient grade), NaCl, KCl, $Na_2HPO_4 \cdot 7(H_2O)$ and KH_2PO_4 were purchased from Fisher Scientific (Loughborough UK). Water was de-ionised using a Triple red system to 18.2mΩ. All samples were (unless otherwise stated) obtained from commercial sources and were of 98% purity or greater and used without further purification

3.3.2 HPLC Instrumentation

Multiple HPLC systems were used; all were Agilent 1100 (or 1200 where stated) systems. The first system consisted of a G1379A degasser, G1311A quaternary pump, G1313A autosampler, G1316A column heater and G1362A refractive index detector (1200). The second consisted of G1310A isocratic pump, G1367A autosampler, G1362A refractive index detector and a Jones chromatography 7955 external column heater. A third consisted of a G1379A degasser, G1311A quaternary pump, G1313A autosampler, G1316A column heater and G1315B diode array detector. The columns were all IAM.PC.DD2 (4.6 x 100 mm; Regis Chemical Company, Morton Grove, IL, USA). The chromatograms were recorded by Agilent Technologies ChemStation for LC systems, version B.03.01-SR1.

3.3.3 Data analysis of log $k_{IAM\ (pH\ 7.4)}$ results

The chromatograms were integrated and the retention times recorded. The retention time was adjusted by subtracting the system dead-time i.e. the elution time for an

unretained compound (t_o) (refer to Section 3.3.4.8 for details). The t_o time was determined as the shortest time taken for three injections of water to travel through the system and column for each analysis. Equation (1.10) was applied to each adjusted retention time to determine the Relative Retention Factor (RRF), also referred to as k_{IAM}.

3.3.4 Method optimisation

Based on the conditions used to determine the experimental log k_{IAM} values collated into the database (refer to Chapter 2 for details) and the manufacturer's recommendations for use of IAM columns with regard to column stability, the IAM-HPLC assay was optimised. Details of the method development are given below and the finalised method is detailed in Section 3.3.4.11.

3.3.4.1 Choice of IAM-HPLC column

The IAM-HPLC assay was optimised using the IAM.PC.DD2 column (4.6x100mm) for the reasons detailed in section 2.4.2 and section 3.1.1.

3.3.4.2 Preparation of mobile phase

The mobile phase was 10mM phosphate buffered saline (PBS)[10] (which contains 2.7 mM KCl, 1.5 mM KH_2PO_4, 137 mM NaCl, and 8.1 mM $Na_2HPO_4 \cdot 7H_2O$) or mixtures of PBS and methanol (ratios of PBS:MeOH ranging from 70:30 to 40:60), depending on the solubility of the sample. For compounds with high hydrophobicity (i.e. long retention times), analysis was performed in 5% increments of methanol and a calibration curve constructed (of log $k_{IAM\ (pH\ 7.4)}$ against % methanol in the mobile phase). The value for log $k_{IAM\ (pH\ 7.4)}$ was then determined by extrapolation to $100\%_{aq}$. All organically modified mobile phases were premixed due to using the refractive index detector. All mobile phases were filtered through 0.45µm nylon filters, adjusted to pH 7.4 and degassed prior to analysis. Table 9 shows a summary of mobile phases used in the construction of the calibration graphs. The samples were prepared in triplicate leading to increased confidence in the extrapolated log $k_{IAM\ (pH\ 7.4)}$ values determined.

Range of log P	Methanol:10 mM PBS ratio used as the mobile phase composition
log P < 1.5	0:100 (100% PBS)
1.5 < log P < 3.7	30:70, 35:65, 40:60, 45:55, 50:50, 55:45 and 60:40
3.7< log P < 4.8	40:60, 45:55, 50:50, 55:45 and 60:40
4.8< log P < 5.7	50:50, 55:45 and 60:40

Table 9 - Mobile phase used in the log k_{IAM} analysis for the construction of calibration graphs based on the compounds hydrophobicity

3.3.4.3 Choice of temperature

The column was maintained at 25°C using a Peltier heating block or external column oven. This was to reduce the effect of day-to-day variability of using ambient temperature.

3.3.4.4 Choice of external standard

An external standard is used as part of the analyses to ensure that the system is running as expected and to act as a reference sample when analysing new compounds. Only three of the 51 studies in the k_{IAM} database mention use of an internal standard. Two papers[11, 12] report the use of 3-nitroaniline, the third[13] used 1,3-dimethyl-5-fluorouracil as the internal standard. An external standard was used in this current investigation instead of an internal standard used in previous papers[11-13]. This is because an internal standard is recommended for HPLC only when the sample is unstable or sample preparation requires many steps; in these cases an internal standard allows the chromatography to remain quantitative in nature[14]. The IAM HPLC method optimised here is qualitative in nature, samples are stable under experimental conditions and sample preparation is a single step, therefore, an external standard is sufficient for the purposes of this investigation. 3-Nitroaniline was used in this study as the external standard during method optimisation, sample analysis and robustness testing. This material was chosen as it had also been used as a sample in other studies[11, 15, 16, 17, 18] from which log k_{IAM} values were collated in the k_{IAM} database.

3.3.4.5 Choice of unretained compound for analysis of compounds by IAM HPLC

The relative retention factor in an IAM column (log k_{IAM}) is defined in equation (1.10), where t_r and t_0 are the adjusted retention times for the sample and an unretained compound, respectively. The unretained compound recommended by the

column manufacturer is citric acid. However, Demare *et al.*[15] reported that the use of citric acid led to irreproducible results. Additionally, initial analysis within this study, using citric acid, indicated a slight inconsistent retention leading to irreproducible results. As a result, water was used as the unretained compound and was included with each run in this investigation; this also means the unretained compound is different from any mobile phase used, as the mobile phase is 10mM PBS with varying percentages of methanol.

3.3.4.6 **Preparation of samples for analysis by IAM HPLC**

Due to low solubility in water, the external standard, 3-nitroaniline, was prepared in water at a concentration of $5x10^{-3}$M. All other samples were prepared in methanol or water (depending on their solubility) at a concentration of $1x10^{-2}$M. All samples were prepared in triplicate and filtered prior to analysis through a 0.45μm nylon filter. A 10μL aliquot was injected (all samples were injected in triplicate).

3.3.4.7 **Solubility of the samples for IAM HPLC analysis**

To ensure that compounds did not precipitate during the analysis, the solubility of the compounds was assessed. Phenol (log P_{exp} 1.5 (experimental log P value from KOWWIN[19])), naphthalene (log P_{exp} 3.6), bibenzyl (log P_{exp} 4.6) and triphenylamine (log P_{exp} 5.7) were checked for precipitation out of solution, across the range of methanol ratios in the mobile phase. Samples were prepared in methanol at a concentration of $1x10^{-2}$M and in 10mM PBS for phenol. 100μL of the sample was added to vials and appropriate volumes of PBS and MeOH added so the total volume per vial was 1mL. The following ratios of methanol in the mobile phase were used: MeOH:10mM PBS 0:100 (phenol only), 30:70, 40:60, 50:50, 60:40. The vials were shaken and left for 30 minutes. The samples were then assessed for precipitation. For some samples the assessment of precipitation was visual, as the precipitation was obvious. For samples where the assessment was not obvious, additional techniques were applied e.g. cross polarisation microscopy and Mie scattering.

Mie scattering is the scattering of light where the particle causing the scattering is similar in size to the wavelength of light. The intensity of the scattering depends on both the wavelength and shape of the particle. Mie scattering reduces to Rayleigh scattering when the particles are smaller, where Rayleigh scattering is defined as the

incoherent scattering of light by particles all of whose dimension are much smaller than the wavelength of light[20].

Phenol was found to be soluble in all percentages of methanol considered as mobile phase conditions. Therefore, samples with a log P $\leq$ 1.5 were analysed using 100% PBS as the mobile phase (aqueous analysis). It is acknowledged that this assumes an inverse proportionality of log P with log S[21].

The naphthalene samples were observed using cross polarisation microscopy. Naphthalene is an isotropic crystal which means that the optical axes are non-equivalent. When polarised light enters a non-equivalent axis it is refracted into two rays; this is called birefringence (i.e. light is observed when a sample is viewed between the cross polarisers). For the naphthalene samples, 20μL of the sample was taken from each vial (containing a range of MeOH:PBS ratios) and viewed between a cross polariser. The microscope magnification was set at 10x, this was connected to a camera eyepiece. The cross polariser was set to extinction i.e. when no birefringent sample is present there is an absence of light. When a biorefringent sample is placed on the slide bed between the cross polarisers light is observed. For solutions at, and below, 20% methanol, naphthalene was found to precipitate out of solution. For solutions of 30% methanol, and above there was no evidence of precipitation. To account for the issue of sample solubility, analysis of naphthalene and compounds with a log P greater than 1.5 and less than 3.6 was carried out using mobile phases containing between 30% and 60% methanol.

Both bibenzyl and triphenylamine do not show birefringence, so were analysed using UV absorbance spectrometry and the results analysed for Mie scattering. Mie scattering was not apparent for the samples analysed. Bibenzyl appears not to precipitate when the ratio of MeOH:PBS is 40:60 or greater. Triphenylamine appears not to precipitate when the ratio of MeOH:PBS is 50:50 or greater. Therefore the mobile phase constitution used were 40:60 – 60:40 for 3.7 < log P $\leq$ 4.8 and 50:50 – 60:40 for 4.8 < log P $\leq$ 5.7. The mobile phase compositions used are summarised in Table 9.

3.3.4.8 **Determination of extra system time**

For each HPLC system used, the extra system time was determined by injecting both water (the unretained compound) and a sample into the system with the column both in line (column in the flow path of the system) and out of line (column not in the flow path of the system). The time taken for a sample to reach the detector when the column is out of line is termed the extra system time. For all analyses, the extra system time was subtracted from the recorded retention time, thus any time recorded was solely the on column time for this system. The process of determination of extra system time is shown diagrammatically in Figure 14. The recorded retention time is the reported output of the chromatography system, the extra system time is calculated for each system and subtracted from the recorded retention time, this leaves the on column time. The log k_{IAM} value was calculated using the on column time for both the sample and the unretained compound using equation (1.10).

The adjusted retention time, rather than the recorded retention time, takes into account changes within the system during analysis and between systems and thus helps to reduce variability due to the instrument. The adjusted retention time was used in analyses and modelling. The extra system time was recalculated for each system used or when the flow path was changed.

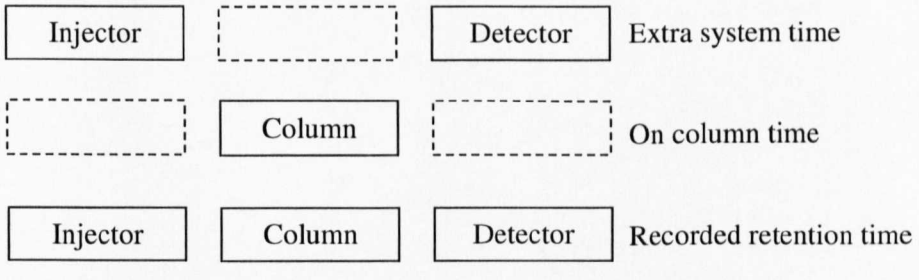


Figure 14 - Illustration of extra system time, on column time and the recorded retention time

3.3.4.9 **Selection of compounds for analysis**

Compounds were chosen initially from the list of 60 OECD reference compounds for RP-HPLC determination of log P[7]. 21 reference compounds were analysed covering a range of log P_{exp} values from 0.3 to 6.5. These compounds were unionised at the pH of analysis i.e. pH 7.4. The degree of ionisation was determined using the pKa values from the PhysProp database[19]. There are no reference compounds in the

OECD guideline with a negative log P value. Therefore, in addition to compounds from the reference list, four polar solvents were also chosen as test compounds. The solvents analysed were methanol (log P -0.77), acetonitrile (log P -0.34), acetone (log P -0.24) and dimethyl sulfoxide (log P -1.35) as these were readily available and are well characterised with regard to log P_{exp}. The range of log P_{exp} values covered by the OECD reference compounds and the additional four solvents is therefore -1.35 to 5.7.

The well characterised OECD reference compounds and four solvents were used to optimise the IAM-HPLC method developed here. Following method optimisation additional compounds were chosen to extend the domain of the IAM HPLC method. This included analysing compounds with high log P values (log P values were based on predicted and experimental values as specified in Table 16), compounds ionised under the conditions of analysis, surface active compounds and compounds that extended the range of functional groups analysed.

3.3.4.10 **Sample analysis**

All samples were prepared in triplicate on different days and each preparation was injected in triplicate, to account for the potential impact of day-to-day variability on the system. Each run had the following order of analysis: PBS (blank); water (unretained compound); 3-nitroaniline (external standard); samples followed by a single injection of the external standard (for runs with analysis time of 60 minutes or longer, the external standard was injected following each sample injection). A blank was run so that any background noise or peaks could be identified as being system related.

Run times per injection for aqueous analysis were 20 minutes for the external standard and samples per injection. When only PBS and water were analysed the run time was reduced to 5 minutes. When organic modifier was added to the mobile phase run times per injection for the standard were reduced to between 5 and 15 minutes depending on the %v/v methanol in the mobile phase. Run times for samples requiring a calibration graph to be produced over a range of ratios of MeOH:10mM PBS in the mobile phase were between 20 and 90 minutes.

3.3.4.11 **Optimised method**

The optimised method conditions are summarised in Table 10 and the procedure to obtain results is summarised in Table 11.

Parameter	**Condition**
Column	IAM.PC.DD2 (4.6x100mm)
Mobile phase	10mM phosphate buffered saline (10mM PBS) or 10mM PBS & 30-60% MeOH as required
Flow rate	1mL/min
Column temperature	25°C
pH of mobile phase	7.4
Injection volume	10µL
Detector	Refractive index at 40°C (DAD used for surfactants)
External standard	$5x10^{-3}$M 3-nitroaniline
Unretained compound	Water

Table 10 –Summary of IAM-HPLC chromatographic conditions used in this study as the optimised IAM HPLC assay

Sample selection	• Compounds diluent and mobile phase were selected based on the criteria discussed above. • The pKa_{exp} and log P_{exp} of compounds selected for analysis were obtained using EPISuite 4.1 (refer to Section 2.3.2 for details of the domain of KOWWIN)[19].
Preparation of mobile phase	• 10mM Phosphate buffered saline (PBS) was prepared, the pH was adjusted to pH 7.4 and filtered through 0.45µm nylon filters prior to use. • If calibration was required, methanol and PBS were independently measured and premixed to the required composition (30-60% v/v methanol). The pH of the mobile phase was adjusted to pH 7.4 and filtered, through 0.45µm nylon filters prior to use.
Priming of the HPLC system	• The column was stored in aceotonitrile and the HPLC system was stored in 30% v/v methanol:water. • The required lines were primed using the following procedure with water. The required lines were flushed with the purge valve open with water for 5 minutes at 5mL/min. Flow rate was then reduced to 1mL/min and purge valve shut. • The system, column and refractive index reference cell were primed using the following procedure with water - The system was run for 30 minutes at a flow rate of 1mL/min with the reference cell purging. • The lines were then primed with the mobile phase of analysis. • The system, column and refractive index reference cell were primed with mobile phase.
Preparation of samples	• Samples were prepared at 10^{-2}M or a concentration below the CMC value for surfactants, in either methanol or PBS depending on the sample solubility. • Samples were filtered through 0.45µm nylon filters. • Aliquots of the samples were transferred to vials. • Samples were prepared in triplicate (on different days) and injected in triplicate to reduce the effect of day-to-day variability.
Analysis of samples	• Samples were injected in triplicate. • The system conditions were as shown in Table 10. • Samples were injected in the order described in Section 3.3.4.10.
System shut down	• The lines used were primed with water. • The system, column and reference cell were purged with water. • The lines used were primed with acetonitrile. • The system, column, and reference cell was purged with acetonitrile. • If the system was not being used again the column was removed. • The system was shut down.

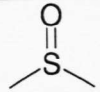
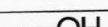
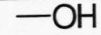
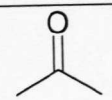
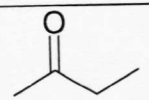
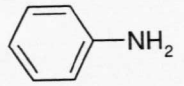
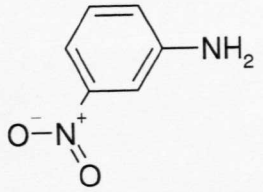
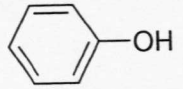
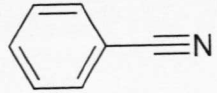
Table 11 - Summary of the IAM-HPLC assay optimised

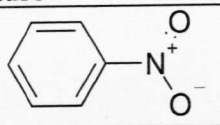
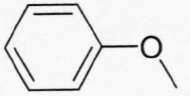
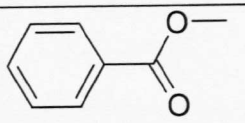
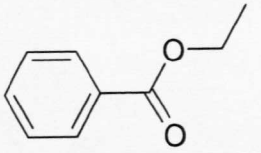
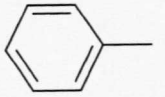
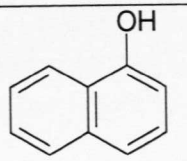
3.4 Results and discussion

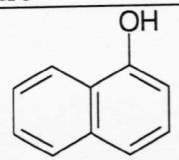
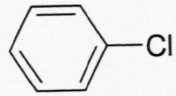
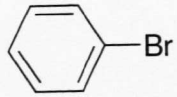
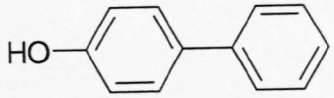
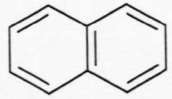
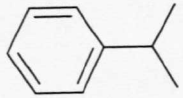
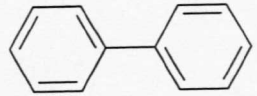
Initial analysis considered unionised OECD reference compounds and four solvents. The log $k_{IAM\ (pH\ 7.4)}$ values were related to log P_{exp} to investigate the relationship of log $k_{IAM\ (pH\ 7.4)}$ with a well characterised and accepted descriptor of hydrophobicity.

3.4.1 OECD reference materials and solvents

The optimised IAM-HPLC method was used to determine the log $k_{IAM\ (pH\ 7.4)}$ values for 21 of the RP-HPLC reference compounds, the four polar solvents and 3-nitroaniline as the external standard. All samples were prepared as detailed in Section 3.3.4.6, using the eluent and mobile phase(s) detailed in Table 12.

Sample	CAS number	Structure	Log P[7]*	Sample eluent	Mobile phase(s)
Dimethyl sulfoxide	67-68-5		-1.35[19]	PBS	PBS
Methanol	67-56-1		-0.77[19]		
Acetonitrile	75-05-8		-0.34[19]		
Acetone	67-64-1		-0.24[19]		
Butanone	78-93-3		0.3		
Aniline	62-53-3		0.9		
3-Nitroaniline	99-09-2		1.34[19]		PBS and 30-60% methanol
Phenol	108-95-2		1.5		PBS
Benzonitrile	100-47-0		1.6	methanol	30-60% methanol

Sample	CAS number	Structure	Log P*	Sample eluent	Mobile phase(s)
Nitrobenzene	98-95-3		1.9	methanol	30-60% methanol
Anisole	100-66-3		2.1		
Methyl benzoate	93-58-3		2.1		
Benzene	71-73-2		2.1		
Ethyl benzoate	93-89-0		2.6		
Toluene	108-88-3		2.7		
1-Naphthol	90-15-3		2.7		

Sample	CAS number	Structure	Log P*	Sample eluent	Mobile phase(s)
1-Naphthol	90-15-3		2.7		30-60% methanol
Chlorobenzene	108-90-7		2.8	methanol	
Bromobenzene	108-86-1		3.0		
Biphenyl-4-ol	92-69-3		3.2		
Naphthalene	91-20-3		3.6		
Cumene	98-82-8		3.7		
Biphenyl	92-52-4		4.0		40-60% methanol

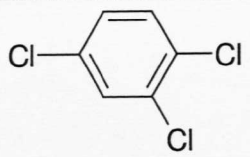
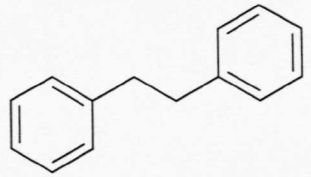
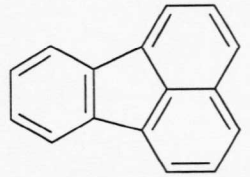
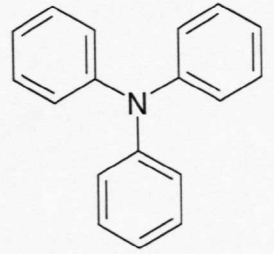
Sample	CAS number	Structure	Log P*	Sample eluent	Mobile phase(s)
1,2,4-Trichlorobenzene	120-82-1		4.2	methanol	40-60% methanol
Bibenzyl	103-29-7		4.8		
Fluoranthene	206-44-0		5.1		50-60% methanol
Triphenylamine	603-34-9		5.7		

Table 12 – Compound name, CAS number, log P_{exp} values, sample eluent and mobile phase(s) for compounds analysed using the IAM-HPLC assay developed (log P values are from reference 7 unless specified) * Experimental log P values from KOWWIN are reported to 2 d.p (as reported) and log P values from the OECD guideline are reported to 1 d.p (as reported).

Figure 15 shows the calibration graph used to determine an aqueous log $k_{IAM\ (pH\ 7.4)}$ value for the external standard, 3-nitroaniline, from the experimental log $k_{IAM\ (pH\ 7.4)}$ values obtained over the mobile phase range of 30:70 – 60:40 MeOH:10mM PBS. The full results are in Appendix 1.2, Table 7 and Table 8.

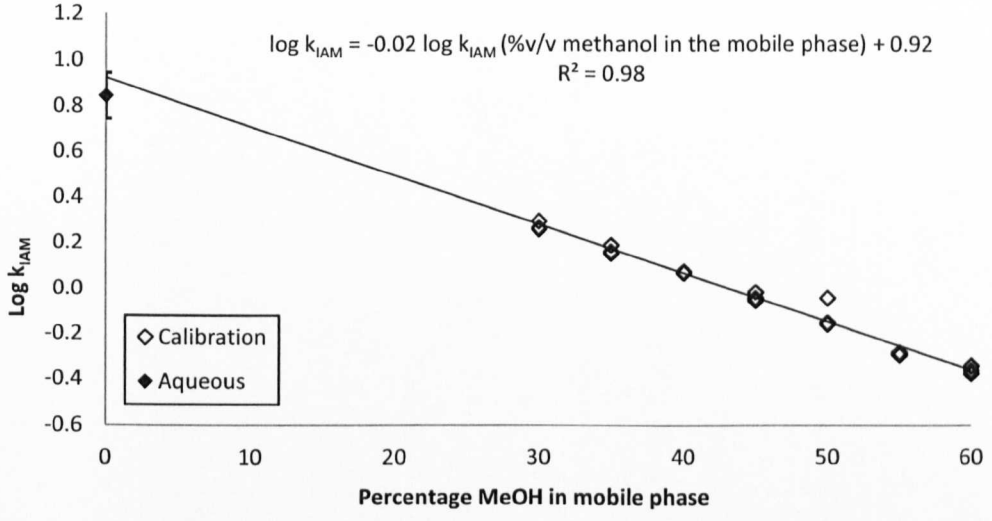


Figure 15 - Calibration graph for 30-60% methanol in the mobile phase for the external standard, 3-nitroaniline, with extrapolation to $100\%_{aq}$ and the $100\%_{aq}$ log $k_{IAM\ (pH\ 7.4)}$ value. ± 0.1 log units shown is the OECD repeatability acceptance criteria

The results illustrated in Figure 15 indicate that the extrapolated log $k_{IAM\ (pH\ 7.4)}$ value is within ±0.1 log units of the $100\%_{aq}$ value determined experimentally for 3-nitroaniline. This is in-line with the quality criteria (refer to Table 13) stated in OECD Guideline no. 117 for the determination of log P using RP-HPLC[7], which assesses the robustness of a reported result. For a log P determination to be considered reliable it must be determined in duplicate, under identical conditions and the quality criteria must be met for the result to be considered reliable and reproducible[7]. It is not unreasonable to apply the same criterion for repeatability to log k_{IAM} values also. Thus, the extrapolated log k_{IAM} value for 3-nitroaniline is within the acceptable range of repeatability. This provides confidence in the extrapolation method used for hydrophobic compounds.

Quality Criteria for the determination of log P by RP-HPLC[7]	
Repeatability	The value of log P_{ow} derived from repeated measurements made under identical conditions and using the same set of reference compounds should fall within a range of ±0.1 log units
Reproducibility	If the measurements are repeated with a different set of reference substances, results may differ. Typically, the correlation coefficient for the relationship between log K and log P_{ow} for a set of test substances is around 0.9, corresponding to an octanol-water partition coefficient of log P_{ow} ± 0.5 log units

Table 13 - The OECD guideline for RP-HPLC determination of n-octanol-water partition coefficient quality criteria to increase confidence in the experimental values determined covering repeatability and reproducibility[7]

The mean log $k_{IAM\ (pH\ 7.4)}$ values for all compounds analysed are reported in Table 14. The calibration graphs for all compounds analysed using extrapolation, except 3-nitroaniline are shown in Appendix 1.2 Figures 1-5. The equations for the extrapolations are reported in Appendix 1.2 Table 10. The intercepts from these graphs are the reported log $k_{IAM\ (pH\ 7.4)}$ values. The quality criteria shown in Table 13 were applied to each calibration graph. Extrapolation to $100\%_{aq}$ was required for all compounds with a log P greater than 1.5; extrapolations were performed using varying percentages of methanol in the mobile phase as detailed in Table 12. For full details of the log $k_{IAM\ (pH\ 7.4)}$ values obtained from triplicate injection and the standard deviation for all compounds analysed refer to Appendix 1.2, Table 7 & 8.

The relationship of log $k_{IAM\ (pH\ 7.4)}$ and log P_{exp} was investigated and is shown in Figure 16. There is a strong relationship between the two parameters with $r^2 = 0.97$. This significant correlation between the two measures of hydrophobicity is to be expected as the compounds were chosen on the basis of being stable and predominantly non-ionisable.

Sample	Log $k_{IAM\ (pH\ 7.4)}$ (2 d.p)
Methanol*	-1.92
Dimethyl sulfoxide*	-1.46
Acetonitrile*	-1.04
Acetone*	-0.89
Butanone*	-0.48
Aniline*	0.21
Phenol*	0.65
Benzonitrile	0.75
3-Nitroaniline*	0.85
Benzene	0.95
Nitrobenzene	1.00
Anisole	1.06
Methyl benzoate	1.36
Toluene	1.44
Chlorobenzene	1.63
Ethyl benzoate	1.78
Bromobenzene	1.80
Cumene	2.22
1-Naphthol	2.25
Naphthalene	2.48
Biphenyl-4-ol	2.77
1,2,4-Trichlorobenzene	2.97
Biphenyl	3.13
Bibenzyl	3.63
Fluoranthene	4.25
Triphenylamine	4.53

* Aqueous determination of log $k_{IAM\ (pH\ 7.4)}$, extrapolation not required

Table 14 - Summary of log $k_{IAM\ (pH\ 7.4)}$ values extrapolated to 100%$_{aq}$ conditions for OECD reference compounds analysed in this study

A good correlation between log $k_{IAM\ (pH\ 7.4)}$ and log P_{exp} was observed for compounds with log P_{exp} in the range of -0.34 to 5.7. Analysis of the detailed log $k_{IAM\ (pH\ 7.4)}$ results in Appendix 1.2, Table 7 & 8 indicates that all injections for any individual compound, at a single mobile phase composition, are within the range of ±0.1 log units and the standard deviation for repeat measurements is low, except for methanol, for which the precision of the results is discussed below. The quality criteria for reproducibility are, therefore, met for all compounds except methanol.

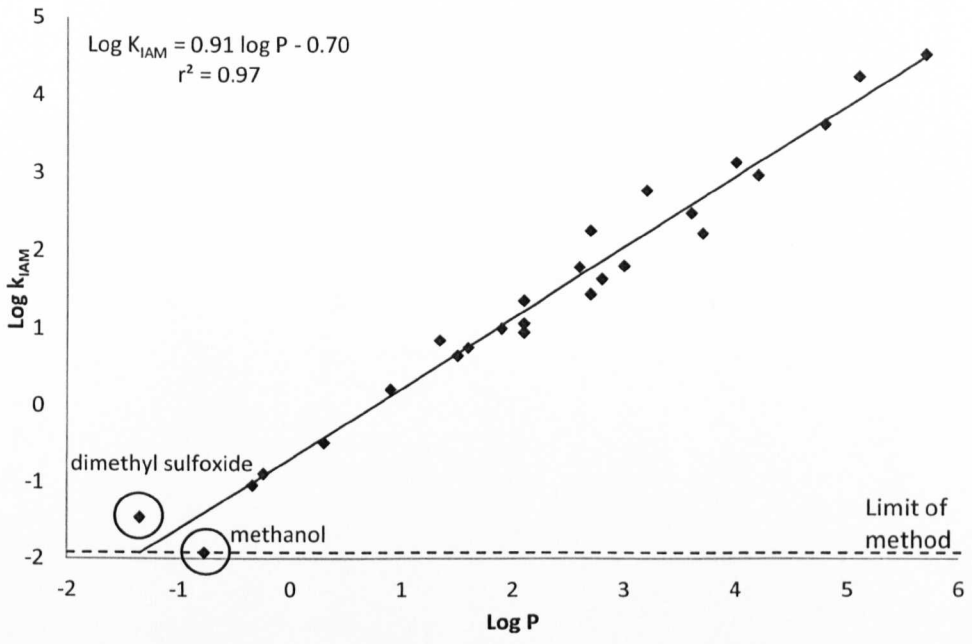


Figure 16 - Plot showing the relationship of log P (experimental) and log $k_{IAM\ (pH\ 7.4)}$ values for the selected OECD reference compounds and solvents analysed in this study (methanol and dimethyl sulfoxide have been highlighted, these are discussed in the text below).

Two compounds, methanol (log P_{exp} of 0.77) and dimethylsulfoxide (log P_{exp} of -1.35), highlighted in Figure 16, are outliers to the relationship. This is due to the low log $k_{IAM\ (pH\ 7.4)}$ values of these compounds, which are near the limit of the method (the limit of the method is defined by the unretained compound, water). Any compound that is less hydrophobic than methanol (log $k_{IAM\ (pH\ 7.4)}$ of -1.92) may be outside the domain of the experimental method. For compounds with low log P_{exp} values approaching -1.38 (log P_{exp} of water), the question of whether the peaks observed are retained or unretained should be considered. The aim of this analysis is not to demonstrate a one-to-one correlation with log P_{exp}. Compounds of particular interest are those for which this correlation breaks down. Therefore, these compounds will be included in the development of predictive methods, in later chapters.

The upper limit of the method is determined by both the solubility of samples and the percentage of methanol in the mobile phase the column can withstand. Currently, the limit of the method is a log P_{exp} of 5.7; this is due to the solubility of the sample (in the mobile phase) and the requirement for analysis that requires extrapolation to

be performed at three different mobile phase compositions, to allow a calibration graph to be plotted and the results extrapolated to $100\%_{aq}$.

The lower limit of the IAM-HPLC method was determined as being a log $k_{IAM\ (pH\ 7.4)}$ value of -1.92. This concurs with the literature log k_{IAM} values collated into the database (refer to Chapter 2 and electronic supplementary material (Chapter 2 – log k_{IAM} database.xls). The lowest log k_{IAM} value reported in the database is -2.00 for cefuroxime reported by Yoon *et al.*[22]. However, this result was obtained using the IAM.PC.DD column, so is not comparable with the experimental method currently being considered here which used the IAM.PC.DD2 column.

Several papers[1, 13 & 23-31] report log k_{IAM} values for compounds that have a log P_{exp} value less than -1.38 (log P_{exp} of water, the unretained compound). Considering log k_{IAM} values in the database that were obtained under comparable experimental procedures (IAM.CP.DD2 column, phosphate buffer mobile phase, pH 7.4 or 7.0, for compounds unionised under the conditions of analysis), there are five log k_{IAM} values relating to four compounds with very low log P values. These compounds all have a higher than expected log k_{IAM} response given the log P_{exp} values. These structures and both the log P and log k_{IAM} values are listed in Table 15. The compounds are all highly hydrophilic and, given their experimental log P value, would not be expected to be retained by the column. However, it must be noted that the stationary phase of the IAM.PC.DD2 column is a double chain of phosphatidylcholine with an ester bond. The IAM.PC.DD2 stationary phase has increased hydrogen bonding capability compared to the IAM.PC.DD column, which increases the retention of highly hydrophilic compounds. Therefore, these compounds may be slightly retained on the column.

Log k_{IAM} and log P are both measures of hydrophobicity. Since log k_{IAM} is an alternative measure of hydrophobicity, it is the compounds for which the response does not fit the correlation, that are of interest. This also explains the scatter of data with a general linear trend. The differences between partitioning of octanol-water and membrane-water were discussed in Sections 1.7 and 1.9.

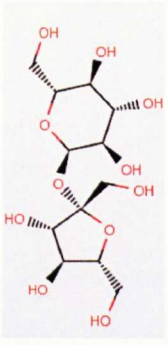
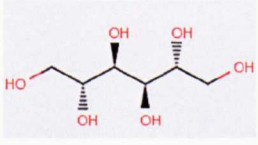
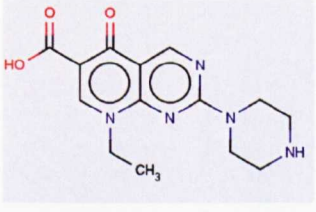
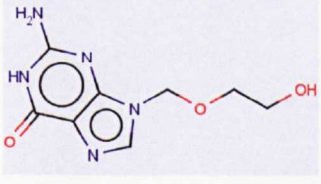
Compound	Structure	pKa	Log P[19]	Log k_{IAM} from database	Reference for log k_{IAM} value
Sucrose		12.62	-3.70	-1.42	23
Mannitol		13.57	-3.10	-1.87 & -1.21	23 & 24
Pipemidic acid			-2.15	0.48	25
Acyclovir			-1.56	-1.15	26

Table 15 - Compound name, structure,pKa, experimental log P and log k_{IAM} for compounds with a higher than expected log k_{IAM} values

All the compounds identified in Table 15 as having a higher log k_{IAM} value than expected from their log P values have an high capability to hydrogen bond (these compounds contain both multiple hydrogen bond donors and acceptors). The compounds are able to hydrogen bond with the stationary phase, meaning the retention of the compound is increased.

Analysis of the detailed log $k_{IAM\ (pH\ 7.4)}$ results in Appendix 1.2, Table 7 & 8 indicates that the OECD quality criteria for repeatability (refer to Table 13) for log P would have been passed by all compounds studied here for log $k_{IAM\ (pH\ 7.4)}$, under all mobile phase compositions and under triplicate determinations, except for methanol. As noted above, the analysis of methanol is near to the limit of the method which has decreased the precision of the measurement and hence there is reduced confidence in this log $k_{IAM\ (pH\ 7.4)}$ value.

3.4.2 Extending the domain of the IAM-HPLC assay

The optimised IAM HPLC method is able to determine log $k_{IAM\ (pH\ 7.4)}$ values for unionised compounds between -1.92 and 4.53. The lower limit of the method is a log $k_{IAM\ (pH\ 7.4)}$ value of -1.92. It is not expected that the optimised method will retain compounds that are less hydrophobic than this. To extend the upper range of the assay, compounds with a reliable log P higher than 5.7 were selected and analysed. Additionally compounds that are ionised under the conditions of analysis were also selected and analysed to investigate the effect of ionisation on log $k_{IAM\ (pH\ 7.4)}$ retention.

3.4.2.1 High log P compounds

Although the OECD list of reference compounds includes compounds up to a log P_{exp} of 6.5, it was only possible to analyse compounds up to 5.7 (triphenylamine). The only other reference compound with a higher log P_{exp} is dichlorodiphenyltrichloroethane (DDT) (log P_{exp} 6.5). Due to legislation surrounding the safe use of this compound, analysis was not possible. Therefore, to extend the upper limit of the method further sources of reliable log P values were investigated.

The PhysProp database[19] contains log P values (mostly experimental values or predicted by KOWWIN) for highly structurally heterogeneous compounds. The PhysProp database was searched for compounds with high log P values and which were also commercially available in high purity. Four compounds were chosen to extend the upper log P range of the IAM-HPLC assay from 5.7, potentially up to 6.8 (refer to Table 16 for details). Samples of these compounds were all prepared in methanol and analysed as per the optimised method using mobile phases containing 50 – 60% methanol as an organic modifier. Log $k_{IAM\ (pH\ 7.4)}$ values determined for the

high log P compounds analysed are reported in Table 16. These were determined using the calibration graphs in Appendix 1.2 Figure 6. The log $k_{IAM\ (pH\ 7.4)}$ values determined for 2-terphenyl and 4-terphenyl fit the relationship with log P determined for the lower log P compounds analysed previously (Figure 17). However, the log $k_{IAM\ (pH\ 7.4)}$ values determined for both N,N'-dicyclohexylcarbodiimide and ethoxylated tetrabromobisphenol A are significantly lower than would be expected from the literature log P values taken from the PhysProp database[19]. There are at least two possible reasons to account for these compounds being apparent outliers; the log P values from PhysProp database[19] (refer to Table 16) could be incorrect (i.e. too high); or the relationship between log P and log $k_{IAM\ (pH\ 7.4)}$ deviates from linearity for compounds with log P greater than 6.

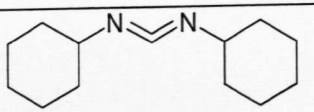
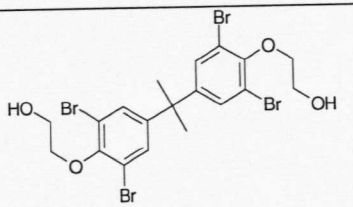
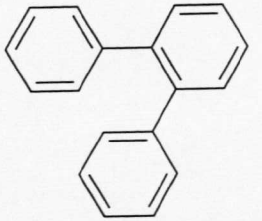
Compound name	CAS No.	Structure	Log P from PhysProp database[19]	Log $k_{IAM\ (pH\ 7.4)}$ (2 d.p)
N,N'-Dicyclohexylcarbodiimide	538-75-0		6.83 Estimated	3.15
Ethoxylated tetrabromobisphenol A	4162-45-2		6.78 Estimated	4.30
2-Terphenyl	84-15-1		5.52 Estimated	4.27
4-Terphenyl	92-94-4		6.03 Experimental	4.93

Table 16 – Name, CAS number, structure, log P and k_{IAM} values for the highly hydrophobic compounds analysed using the optimised IAM-HPLC assay and the log k_{IAM} values extrapolated to 100%$_{aq}$

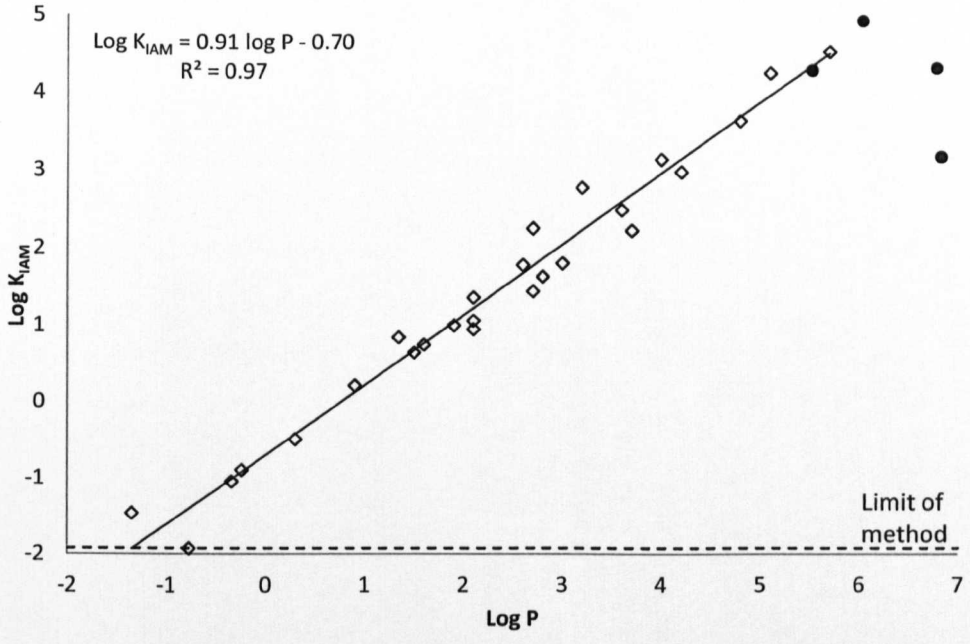


Figure 17 - Plot showing relationship of log P against log $k_{IAM\ (pH\ 7.4)}$ for the OECD reference compounds, solvents and the four hydrophobic compounds selected to extend the range of the log P relationship (shown by the filled circle) (with log P taken from the PhysProp database[19]) and reported in Table 16

The OECD reference compounds previously analysed (Section 3.4.1) are well characterised and there is confidence in the log P_{exp} values of these compounds. For the high log P compounds analysed to extend the range of the method, three of the four log P values in Table 16 are predicted from KOWWIN.

To investigate the variability of calculated log P values, the compounds were analysed with the Virtual Computational Chemistry Laboratory (VCCLAB)[32, 33], which predicts log P using a range of algorithms[34] (refer to Table 17 for details of the algorithms used). The relationship between log $k_{IAM\ (pH\ 7.4)}$ and log P_{exp} for all the compounds previously analysed is shown in Figure 18. In addition, the range of log P values (Table 17) for the high log P compounds and phenol are illustrated. Phenol is included in the analysis by way of a benchmark, as phenol is a well characterised compound with regard to log P.

For phenol the range of log P values determined by the various calculation methods are reasonably concordant, with a range of less than 0.3 log units. The OECD

guideline of ±0.1 log units is for repeatability within an experiment i.e. between duplicate measurements. For inter experimental variability the acceptable range increases to 0.3 log units[35], i.e. the determination of log P between experiments using different techniques, standards or laboratories. However, for the high log P compounds the range of predicted log P values is much broader. Therefore, it is much more difficult to select the correct value of log P based on the values predicted.

Due to the difficulty in determining the correct log P value, it was decided to determine the log P values of ethoxylated tetrabromobisphenol A and N,N'-dicyclohexylcarbodiimine experimentally using the OECD RP-HPLC method[7].

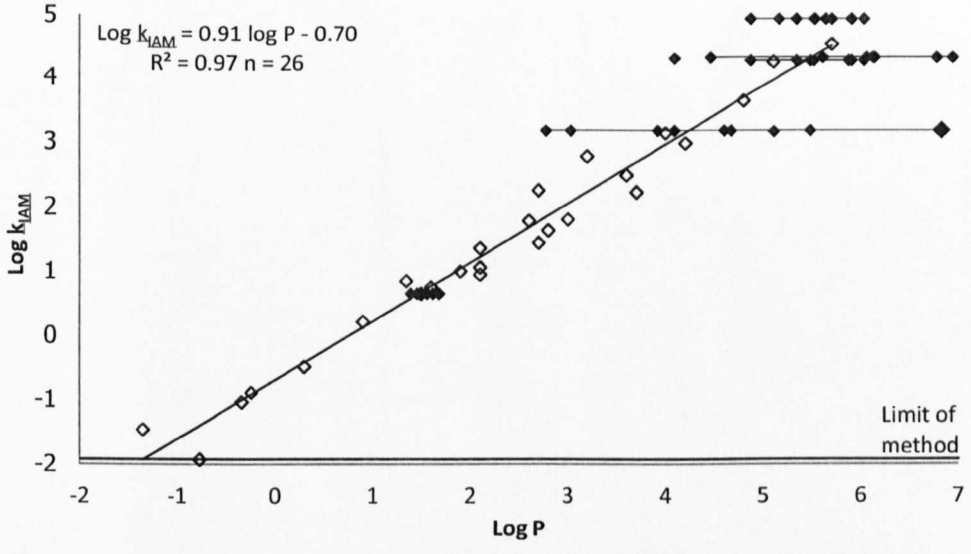


Figure 18 - Plot of log k_{IAM} against log P for all compounds analysed showing the range of calculated log P values for 2-terphenyl, 4-terphenyl, ethoxylated tetrabromobisphenol A, N,N'-dicyclohexylcarbodiimine and phenol (log P values reported in Table 17)

Compound	Log P										
	ALogPS[a]	AC log P[b]	XLOGP2[c]	XLOGP3[c]	MiLogP[d]	KOWWIN[e]	ALOGP[f]	MLOGP[b]	Experimental	Range	Std deviation
2-terphenyl	5.87	5.34	5.9	6.03	5.48	5.52	4.87	5.16		4.87 – 6.03	0.40
4-terphenyl	5.7	5.34	5.9	5.64	5.53	5.52	4.87	5.16	6.03	4.87 – 6.03	0.36
ethoxylated tetrabromobisphenol A	6.13	5.61	6.94	6.06	6.12	6.78	6.14	4.46		4.46 – 6.94	0.76
N,N'-dicyclohexylcarbodiimine	4.6	2.78	5.48	4.67	3.92	6.83	4.09	3.03		2.78 – 6.83	1.31
phenol	1.39	1.68	1.62	1.46	1.46	1.51	1.56	1.51	1.46	1.39 – 1.68	0.09

[a] – Developed using associative neural networks method[36]
[b] - Fragment based algorithm[37]
[c] - Atom contribution method with correction factors[37]
[d] – Counts of atoms, bonds, fragments and functional groups[37]
[e] - Atom and fragment based method for correction factors[38]
[f] – E-state Indices based associative neural networks method[37]

Table 17 - Log P values for 2-terphenyl, 4-terphenyl, ethoxylated tetrabromobisphenol A, N,N'-dicyclohexylcarbodiimine and phenol using different specified methods of determination[32, 33, 34]

3.4.2.2 Determination of log P for highly hydrophobic compounds using the OECD RP-HPLC method

To measure the log P experimentally for the highly hydrophobic compounds discussed in Section 3.4.2.1, OECD guideline number 117[7] for the determination of n-octanol/water partition coefficient using RP-HPLC was followed. The method used is detailed in Table 18. OECD guideline 117 states

> *"a calibration graph using at least six points has to be established. [...] The reference compounds should normally have log P_{ow} values which encompass the log P_{ow} of the test substance, i.e. at least one reference compound should have a P_{ow} above that of the test substance, and another a P_{ow} below that of the test substance. Extrapolation should only be used in exception cases. It is preferable that these reference substances should be structurally related to the test substance."*

Parameter	Condition
Mobile phase	25:75 Water:Methanol at pH 7.2
Flow rate	1mL/min
Temperature	30°C
Method of detection	RI detector at 40°C
Column	Agilent eclipse XDB-C18 (4.6x150mm)
Reference compounds used (reference log P)	Biphenyl-4-ol (3.2) Naphthalene (3.6) Biphenyl (4.0) 1,2,4 Trichlorobenzene (4.2) Bibenzyl (4.8) Fluoranthene (5.1) Triphenylamine (5.7)

Table 18 - Overview of OECD RP-HPLC method used to determine log P for highly hydrophobic compounds

A selection of reference compounds was chosen; these encompassed the expected log P values for the test compounds (based on the experimental log k_{IAM} value determined). The reference compounds were separated into two test mixtures (test

mixture 1 contained biphenyl-4-ol; biphenyl and fluoranthene. Test mixture 2 contained naphthalene; 1,2,4-trichlorobenzene; bibenzyl and triphenylamine) and were analysed at the start and the end of the analysis. Water was analysed as the unretained compound. The two test compounds (N,N-dicyclohexylcarbodiimide and ethoxylated tetrabromobisphenol A) were prepared in mobile phase at 10^{-2}M as individual samples and were prepared in duplicate. All samples were injected in duplicate.

Figure 19 shows the calibration curve constructed from the RRF values from this RP-HPLC study (refer to Appendix 1.2, Table 11) and the reference log P values. The OECD guideline quality criteria were followed to ensure the experimental log P values determined were both repeatable and reproducible; these criteria are detailed in Table 13. Analysis of the individual log P values (refer to Appendix 1.2, Table 11) for all the reference compounds and the test compounds shows good repeatability within ±0.1 log units, complying with the quality criteria required by the OECD guideline[7].

Figure 19 indicates that there is a strong relationship between the experimental RRF values and reference log P values (r = 0.97). Using the measured RRF values for the test compounds and the calibration graph, the log P values were interpolated following the OECD guideline[7]. The log P values for ethoxylated tetrabromobisphenol A and N,N'-dicyclohexylcarbodiimide were determined to be 4.09 and 5.11 respectively (Figure 19).

The experimentally determined values of log P for both ethoxylated tetrabromobisphenol A and N,N'-dicyclohexylcarbodiimide are lower than those reported in the PhysProp database[19]. When these experimental log P values are used to assess the relationship with log k_{IAM} (Figure 20), they are not such significant outliers. As a result of using more accurate data, the upper log P limit of the method was extended up to a log P of 6.03 giving a log k_{IAM} response of 4.97. Figure 20 confirms that over the range of log P (-1.35 to 6.03) there is a linear relationship with log k_{IAM}. As a result of these corrections, the upper limit of the method has not been established, therefore, further work is required. It is advised that both the log P value

and log $k_{IAM\ (pH\ 7.4)}$ value are determined experimentally, for highly hydrophobic compounds, whilst investigating the upper limit of the method.

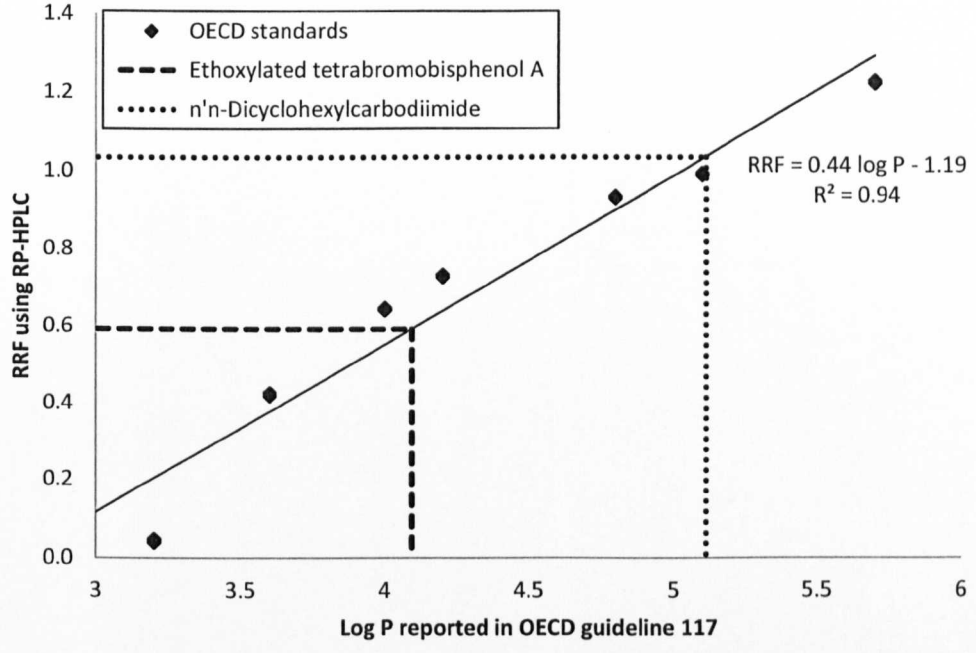


Figure 19 – Calibration graph of RRF values determined using the OECD RP-HPLC[7] method against reference log P values.

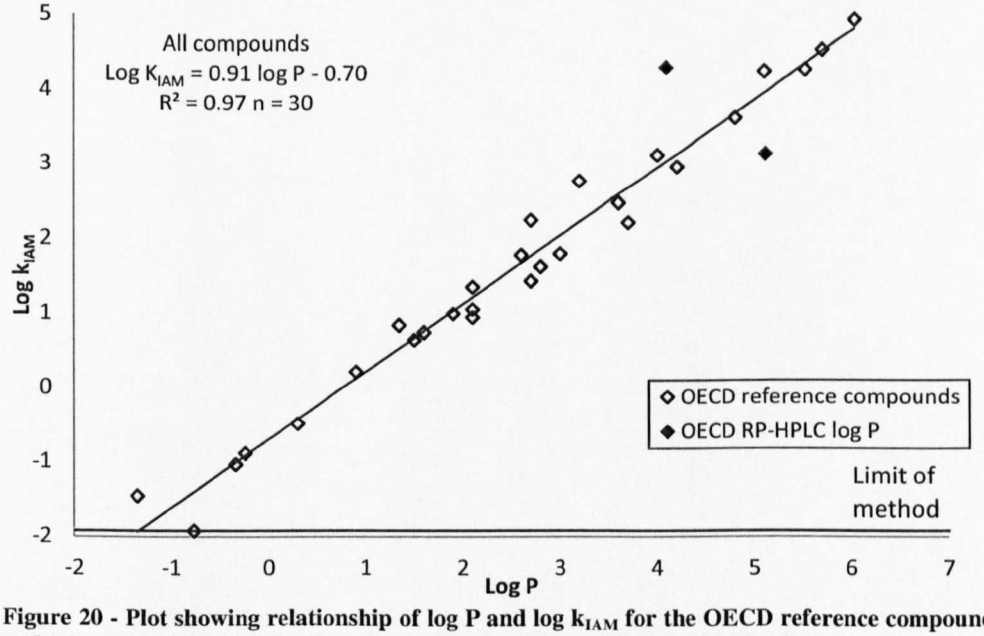


Figure 20 - Plot showing relationship of log P and log k_{IAM} for the OECD reference compounds and solvents and the high log P compounds using the experimental log P values

3.4.2.3 Compounds ionised under the conditions of analysis

To extend the domain of the optimised IAM-HPLC method to include compounds ionised under the assay conditions, several compounds were selected on the basis that they are ionised under the conditions of analysis (refer to Table 19 for pKa values and percentages ionised). These compounds were benzoic acid, 2-nitrobenzoic acid, 3-iodobenzoic acid and 3-aminobenzoic acid. Log k_{IAM} was measured for these compounds using aqueous analysis at pH 7.4. The log $k_{IAM\ (pH\ 7.4)}$ of pentanoic and hexanoic acid were measured using organically modified mobile phases and the log $k_{IAM\ (pH\ 7.4)}$ values extrapolated to aqueous conditions. Log $k_{IAM\ (pH\ 7.4)}$ refers to the measured experimental log k_{IAM} value determined at a pH of 7.4, the pH is specified because at this pH the compound is ionised. An alternative description for this is log $D_{IAM\ (pH\ 7.4)}$, indicating the experimental value has been determined for the compound in an ionised state.

The pKa values for each of these compounds analysed were taken from the PhysProp database[19]. The percentage of each compound ionised was calculated using equation (3.1). Equation (3.1) is a rearrangement of the Henderson-Hasselbach equation, (3.2)[39].

$$\%\ ionised = \frac{100}{1+10^{(charge\ (pH-pKa)}} \qquad (3.1)$$

Where

Charge = +1 for bases and -1 for acids

$$pH = pKa + \log\frac{[base]}{[acid]} \qquad (3.2)$$

A summary of the experimentally determined log $k_{IAM\ (pH\ 7.4)\ (Ion)}$ values and calculated log $k_{IAM\ (Neu)}$ values (for the neutral, unionised species) (calculated using equations (3.3) for the acids, equation (3.4) would be used for the analysis of bases), for compounds analysed in the ionised form, is reported in Table 19. Full experimental values and the calibration graphs for both pentanoic acid and hexanoic acid are in Appendix 2.1, Tables 12 and 13 and Figure 7.

$$\log k_{IAM\,(Neu)} = \log k_{IAM\,(pH\,7.4)(Ion)} - \log\left(1 + 10^{(7.4 - pKa)}\right)^{40} \quad (3.3)$$

$$\log k_{IAM\,(Neu)} = \log k_{IAM\,(pH\,7.4)(Ion)} - \log\left(1 + 10^{(pKa - 7.4)}\right)^{40} \quad (3.4)$$

Compound	Ionised log $k_{IAM\,(pH7.4)}$	Neutral log k_{IAM}	Log P	pKa	Percentage ionised at pH 7.4
Benzoic acid	-0.80	2.41	1.87	4.19	99.9
2-Nitrobenzoic acid	-1.11	3.82	1.46	2.47	100
3-Iodobenzoic acid	0.41	3.96	3.13	3.85	100
3-Aminobenzoic acid	-1.46	2.87	0.65	3.07	100
Pentanoic acid	-0.06	2.50	1.39	4.84	99.7
Hexanoic acid	-0.06	2.46	1.92	4.88	99.7

Table 19 - Summary of log $k_{IAM\,(pH\,7.4)}$, log k_{IAM}, log P, pKa and percentage ionisation for compounds ionised under the conditions of analysis.

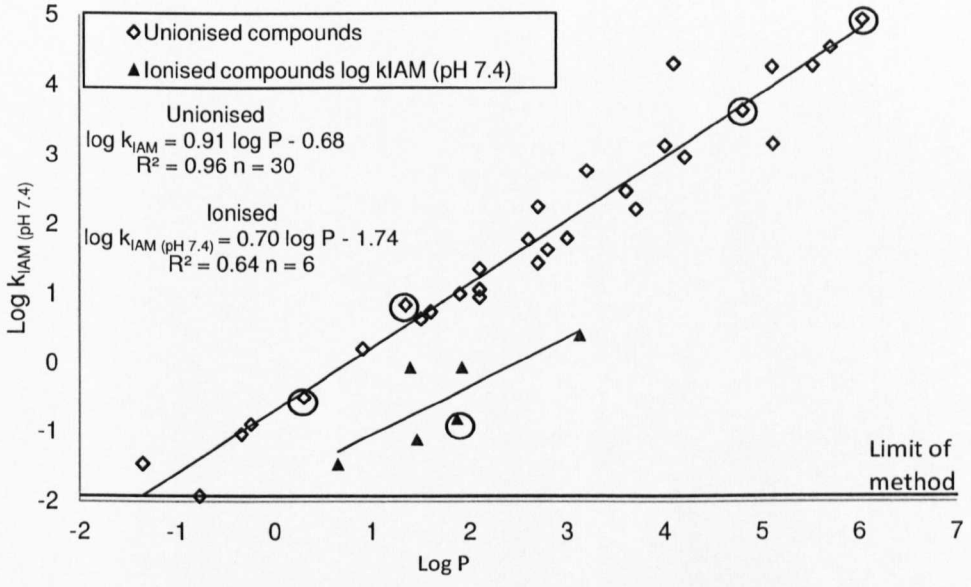


Figure 21 - Plot of log $k_{IAM\,(pH\,7.4)}$ against log P_{exp} for both ionised and unionised compounds under the conditions of analysis (the compounds circled are five compounds that were analysed further in Section 3.4.3 to determine the robustness of the method).

Figure 21 shows the relationship between log P (for the unionised species) and log $k_{IAM\,(pH\,7.4)}$ for all the unionised compounds previously analysed. Figure 21 clearly illustrates that the log $k_{IAM\,(pH\,7.4)}$ for ionised compounds (as expected) is lower than the log $k_{IAM\,(pH\,7.4)}$ value for unionised compounds with a similar log P value, by approximately one log unit. This indicates that these log P values are likely to be for the unionised species.

A linear relationship is observed (as shown in Figure 21) between log $k_{IAM\ (pH\ 7.4)}$ for ionised compounds and log P. However, the correlation (r = 0.8) is not as good as for unionised compounds.

3.4.3 Robustness testing for optimised IAM-HPLC assay

The optimised IAM-HPLC assay, described in section 3.3.4, was able to determine log $k_{IAM\ (pH\ 7.4)}$ for compounds with log P in the range -1.35 to 6.03 that are both the ionised and unionised state under the conditions of analysis. For the IAM-HPLC assay to be a reliable alternative for determining hydrophobicity, the robustness of the method needs to be assessed. Log $k_{IAM\ (pH\ 7.4)}$ values were determined for five compounds varying from very hydrophilic to hydrophobic (log P ranging from 0.3 to 6.03); these are detailed in Table 20. The compounds chosen for analysis covered a wide range of log P values consistent with many materials of interest to industry. The robustness of the method was assessed using five different IAM.PC.DD2 columns, three batches of stationary phase and two HPLC systems.

Three of the compounds chosen were OECD reference materials for the determination of log P by RP-HPLC[7]. These are benzoic acid, butanone and bibenzyl. They are well characterised and there is high confidence in the reported log P values. To ensure the method was robust for ionisable compounds benzoic acid was included. Since benzoic acid is 99.9% ionised under the experimental conditions of the optimised IAM-HPLC assay and the log P_{exp} value reported in KOWWIN[19] is for the unionised species, it is unrealistic for the log P and log $k_{IAM\ (pH\ 7.4)}$ to fit the existing correlation for neutral species. However, this analysis investigated repeatability across columns, batches of stationary phase and systems of analysis which are not affected by this. The remaining compounds analysed were 3-nitroaniline which is the external standard used in the IAM-HPLC method and 4-terphenyl which is a highly hydrophobic compound.

3.4.3.1 Method to test the robustness of the optimised IAM-HPLC assay

3.4.3.1.1 IAM-HPLC method to test the robustness of assay

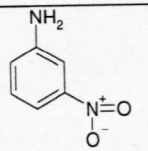
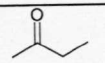
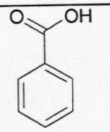
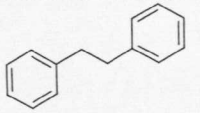
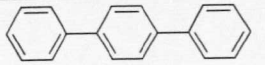
The optimised IAM-HPLC method described in section 3.3.4.11, and summarised in Table 10, was investigated for robustness. The compounds used to assess method

robustness, the diluent and the mobile phase conditions are detailed in Table 20. The robustness analysis was performed on five columns (two columns contained stationary phase from a single batch from the supplier, a further two columns contained stationary phase from a second batch and the final column contained stationary phase from a third batch). Analysis was performed on one column on three separate days to establish the effect of day-to-day variability. One of the five columns was analysed on a second system to establish the effect of system variability.

3.4.3.1.2 **Statistical analysis of robustness testing results**

The Minitab[41] statistical package (version 15.1.1.0) was used for the analysis of data. Two-way ANalysis Of VAriance (ANOVA) was used to test for similarity where the data were balanced and the General Linear Model (GLM) was used to perform the same analysis when the data were unbalanced. Balanced and unbalanced refers to the number of data points compared for each condition (conditions investigated were the effect of system, batch and column on measured log $k_{IAM\ (pH\ 7.4)}$). If the same number of data are considered for each condition the data are balanced; if a different number of data are considered for each condition the data are unbalanced. Two-way ANOVA was used as the five substances analysed cover a range of log P and log $k_{IAM\ (pH\ 7.4)}$ values. The first variable in the Two-way ANOVA was the compound, the second variable was either the day, column, batch or system to compare for relative effects.

In addition, the Tukey test, which is a *post-hoc* test was applied to the output from the ANOVA test. This reports an upper and lower limit for difference which has to be within a predetermined range. Overall the Tukey test has a test wide confidence interval of 95%[42].

Compound	CAS No.	Structure	Log P			Sample diluent	Mobile phase conditions
			Experimental[a]	KOWWIN Estimated[a]	OECD[b] recommended		
3-Nitroaniline	99-09-2		1.37	1.47	-	PBS	PBS & 40-60% MeOH
2-Butanone	78-93-3		0.29	0.26	0.3	PBS	PBS
Benzoic acid	65-85-0		1.87*	1.87	1.9	PBS	PBS
Bibenzyl	103-29-7		4.79	4.74	4.8	MeOH	40-60% MeOH
4-Terphenyl	92-94-4		6.03	5.52	-	MeOH	40-60% MeOH

[a] From KOWWIN[19] [b] from the OECD Guideline for determining log P by RP-HPLC[7] (* corrected for ionisation)

Table 20- Name, CAS number, structure, experimental and calculated log P values, sample diluent and mobile phase conditions for compounds used to assess the robustness of IAM-HPLC method

For the determination of log P by RP-HPLC, the OECD guideline reports repeatability criteria of ±0.1 log units (for repeat measurements under identical conditions using the same reference compounds). If different reference compounds are used in the analysis, the acceptable range increases to ±0.5 log units for the reproducibility criteria[7] (refer to Table 13 for the quality criteria).

The acceptable variation in log $k_{IAM\ (pH\ 7.4)}$ values was considered to be 0.3 log units (i.e ±0.15), this is the same experimental variability considered acceptable for experimental determination of log P[35]. The OECD guideline repeatability criterion of ±0.1 log units was also applied to the log $k_{IAM\ (pH\ 7.4)}$ results. It is noted here that the more stringent criterion considers the analysis of six reference compounds during the determination of log P. However, the robustness testing considers five compounds independently, without the use of reference compounds, this criterion is not directly applicable to the analysis to determine method robustness. The OECD criterion for reproducibility, allows a variation of ±0.5 log units and again includes six reference compounds. Thus the range of ±0.15 and ±0.1 log units in the measured log $k_{IAM\ (pH\ 7.4)}$ values was used to determine their robustness (or otherwise) and assess the effect of the various experimental conditions.

3.4.3.2 **Results and discussion for robustness of IAM-HPLC assay**

The arithmetic means of log $k_{IAM\ (pH\ 7.4)}$ values obtained for all compounds across all columns are reported in Table 21. The full results and standard deviations are reported in Appendix 1.2 Tables 14 and 15. The standard deviation for each compound is small, which shows that the results obtained for each individual compound are similar in response across the five columns. For each compound considered the range of log $k_{IAM\ (pH\ 7.4)}$ values obtained across all columns are within the allowable range of ±0.15 log units.

Column	Batch	system	Aqueous plate count (max & min)	40-60% plate count (max & min)	Arithmetic mean log $k_{IAM\ (pH\ 7.4)}$				
					3-nitro aniline	2-butanone	benzoic acid	bibenzyl	4-terphenyl
1	1	1	1126-1156	677 - 1164	0.800	-0.390	-1.060	3.52	4.79
2	2	1	1297-1350	581-1682	0.785	-0.465	-0.945	3.63	4.93
3	3	2	614-663	599-1029	0.785	-0.363	-0.880	3.49	- *
4	3	1	714-768	495 - 1092	0.820	-0.330	-0.840	3.60	5.06
5	1	1	1978-2184	671-1827	0.846	-0.484	-0.804	3.63	4.93
Arithmetic mean (standard deviation) across all data					0.807 (0.026)	-0.406 (0.066)	-0.906 (0.101)	3.57 (0.07)	4.93 (0.11)
Log $k_{IAM\ (pH\ 7.4)}$ range					0.061	0.154	0.256	0.140	0.270

*No value reported. Column 3 produced a poorly defined peak at one of the required mobile phase concentrations therefore extrapolation of data was not performed. Refer to text for further details.

Batch 1, 2 and 3 refer to the batch of stationary phase the column is packed with.

Table 21 - Arithmetic mean and experimental log $k_{IAM\ (pH\ 7.4)}$ values obtained for the five samples, using five IAM-HPLC columns, containing three batches of stationary phase and analysed on one of two systems

No log $k_{IAM\ (pH\ 7.4)}$ value was reported for 4-terphenyl using column 3 (Refer to Table 21). This is because the value for log $k_{IAM\ (pH\ 7.4)}$ for 4-terphenyl was calculated from extrapolation of log $k_{IAM\ (pH\ 7.4)}$ values measured using 50, 55 and 60% methanol in the mobile phase. For column 3 a peak was identified using both 55 and 60% methanol. When using a mobile phase of 50% methanol, a peak for 4-terphenyl can be identified above the background noise for column 3. However, without overlaying previous chromatograms obtained using the other columns, it would not be possible to recognise this peak for column 3 confidently. Extrapolation of the two values obtained for 55 and 60% methanol gives a log $k_{IAM\ (pH\ 7.4)}$ value within the range obtained for the remaining four columns. However, this value has not been used in the analysis or reported as extrapolation from two points could be unreliable.

Columns 3 and 4 have a low plate count compared to all other columns. The supplier of the column, Regis Technologies, distribute columns with a plate count above 500 (standard conditions 80/20 0.1M potassium dihydrogen phosphate/acetonitrile), with the columns usually running at about 1,000 plate counts. The performance of the IAM columns is stated not to be based on theoretical plate count but on retention time. Therefore, a low plate count was not considered to be of concern[3]. Whilst this may be true for moderately hydrophobic compounds, for highly hydrophobic compounds with long retention times, peak shape is broader and resolution can become limiting.

3.4.3.2.1 **Day-to-day variability**

The precision of the IAM-HPLC method was determined by repeat injection (intra-day variability) and repeat preparation (inter-day variability). Both intra-day and inter-day variability were determined for all five compounds using column 5. Intra-day variability was determined by injecting a single preparation of each sample three times. For each analysis, the three chromatograms generated from triplicate injections were comparable when overlaid (Figure 22 illustrates three chromatograms generated for the triplicate injection of 3-nitroaniline, using column 1).

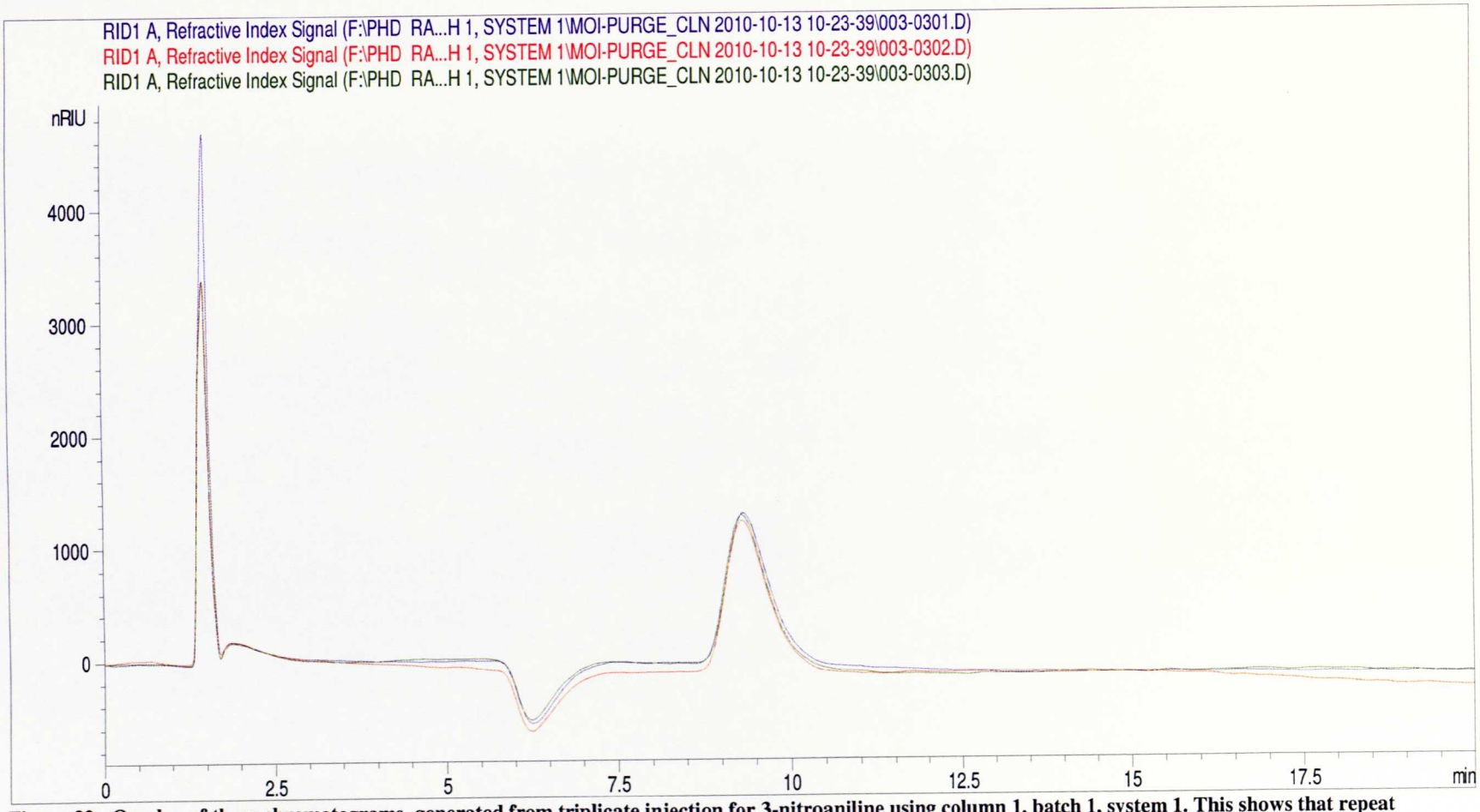


Figure 22 - Overlay of three chromatograms, generated from triplicate injection for 3-nitroaniline using column 1, batch 1, system 1. This shows that repeat injections in triplicate are comparable and that intra-day variability is minimal

The five compounds were analysed on three different days. Two-way ANOVA analysis of the log $k_{IAM\ (pH\ 7.4)}$ values obtained was performed to assess inter-day variability. Inter-compound variability is expected and is, therefore, discounted. From the ANOVA analysis the effects of substance, day and the interaction between substance and day were indicated as being significant. As expected, there is a significant difference between the log $k_{IAM\ (pH\ 7.4)}$ values for different substances. The Tukey test was applied for day–to-day variability, comparing values measured on a separate day against all other days. All Tukey test confidence intervals show there to be no significant difference between days of analysis (refer to Appendix 1.2, Table 16 for the Tukey test upper and lower limit for difference results for day-to-day variability). Thus, analysis of the data indicates the IAM-HPLC method to be robust to intra-day and inter-day variability.

3.4.3.2.2 Analysis of variance of log k_{IAM} values obtained on different HPLC systems, columns and batch of stationary phase

Table 21 summarises the log $k_{IAM\ (pH\ 7.4)}$ values obtained for five compounds using five different columns. The data from the day-to-day analyses, which were carried out in triplicate, were used along with single preparations analysed on four columns. The column and batch of stationary phase are recorded along with the HPLC system from which the results were obtained. The data were analysed for variance caused by the column, stationary phase batch and system.

3.4.3.2.2.1 System-to-system variability

Five compounds were analysed on two systems, refer to Table 21 for full details. Due to the data being unbalanced (single preparation, injected in triplicate and triplicate preparation, injected in triplicate), the general linear model (GLM) was applied to the data for analysis of variance. The variables were substance (five levels, five compounds) and system (two levels, two systems). The Tukey test was applied to assess the effect of the system, the upper and lower limits for difference were -0.06 and 0.037 respectively. This confirms that the system used did not affect the log $k_{IAM\ (pH\ 7.4)}$ values. This is unsurprising since in this experimental design the system time was taken into account and only the on-column time was analysed.

3.4.3.2.2.2 Column-to-column variability

Column-to-column variability was assessed using the GLM (due to the data being unbalanced), the two variables were substance (five levels, five compounds) and column (five levels, five columns. The system used was shown not to affect the IAM-HPLC results obtained in Section 3.4.3.2.2.1.

The Tukey test was applied to investigate the effects of column-to-column variability. Some comparisons showed significant variability, however, the range of the upper and lower limits for difference was within the predetermined allowable range of ±0.1 log units, (refer to Appendix 1.2, Table 17 for the Tukey test upper and lower limit for difference results for column-to-column variability). The five columns used in the analysis contain stationary phase from three different batches, and for this reason the effect of batch-to-batch variability was investigated.

3.4.3.2.2.3 Batch-to-batch variability

GLM analysis followed by the Tukey test demonstrates the maximum variance is 0.11.This is a variance of 2.6% for the range of log P considered and is well within the 0.3 log units for acceptable inter experimental variability (refer to Appendix 1.2, Table 18 for the Tukey test upper and lower limit for difference results for batch-to-batch variability). Therefore, batch-to-batch variability falls within the allowable range of experimental error and is, therefore, not significant.

3.4.3.3 Overview of method robustness testing results

The optimised IAM-HPLC method to determine log $k_{IAM\ (pH\ 7.4)}$ has been demonstrated to provide precise and accurate results for compounds covering a wide range of hydrophobicity. The effects of HPLC system, column and stationary phase batch on log $k_{IAM\ (pH\ 7.4)}$ values have been investigated and the method has been demonstrated to be robust under the conditions investigated. Although the batch of stationary phase shows some variability, the range of difference for log $k_{IAM\ (pH\ 7.4)}$ is within the acceptable range of 0.3 log units for inter experimental-variability and so is not considered significant. An IAM-HPLC method that is robust has increased confidence in the log $k_{IAM\ (pH\ 7.4)}$ values obtained for the compounds analysed using this method.

3.5 Conclusions

An IAM-HPLC assay has been optimised using a range of compounds covering a wide range of hydrophobicites (log P of -1.35 to 6.03 and a log $k_{IAM\ (pH\ 7.4)}$ of -1.92 to 4.93) that includes compounds both neutral and ionised under the conditions of analysis. The lower limit of the method has been demonstrated, however the upper limit has not conclusively been determined. It is recommended that both log $k_{IAM\ (pH\ 7.4)}$ and log P be determined experimentally whilst analysing highly hydrophobic compounds for determination of the upper limit of the log k_{IAM} method.

The method has been demonstrated to be robust across system of analysis, column and stationary phase batch. Given this, there is confidence in the log k_{IAM} values determined for a range of 36 aliphatic and aromatic compounds. A robust method for determining log k_{IAM} values increases the confidence of log k_{IAM} values obtained. Therefore, the use of log k_{IAM} as an alternative descriptor to log P in QSARs predicting toxicity endpoints does not adversely affect confidence in the output from any QSARs developed. In addition, the robust assay is applicable to compounds covering a wide range of hydrophobicities and covers both ionised and unionised compounds. Extending the domain of compounds considered would further increase the applicability of the robust assay optimised here.

Using experimental log k_{IAM} values determined here, methods to predict log $k_{IAM\ (pH\ 7.4)}$ will be investigated (Chapters 4 and 5). The ability to predict log $k_{IAM\ (pH\ 7.4)}$ extends the applicability of log k_{IAM} as a descriptor due to the descriptor becoming more widely accessible, to predict toxicity endpoints (Chapter 6 and 7). Additionally the ability to predict log k_{IAM} aids in the design of experiments and reduces the cost, time and resources required to determine log k_{IAM} values.

3.6 References

[1] Hollósy F., Valkó K., Hersey A., Nunhuck S., Kéri G., Bevan C. (2006) Estimation of Volume of Distribution in Humans from High Throughput HPLC-Based Measurements of Human Serum Albumin Binding and Immobilized Artificial Membrane Partitioning *J. Med. Chem.* 49: 6958-6971.

[2] Valkó K., Du C.M., Bevan C., Reynolds D.P., Abraham M.H., (2000) Rapid-Gradient HPLC Method for Measuring Drug Interactions with Immobilized

Artificial Membrane: Comparison with Other Lipophilicity Measures *J. Pharm. Sci.* 89: 1085-1096.

[3] Personal communication with Ted Szczerba, Regis Chemical Company, Morton Grove, IL, USA.

[4] UK Environmental Standards and Conditions (Phase 1) Final report (2008) UK Technical Advisory Group on the Water Framework Directive, available from <http://www.wfduk.org/UK_Environmental_Standards/LibraryPublicDocs/UKTAG%20ReportAug%202006UKEnvironmentalStandardsandConditionsFinalReport> [Accessed 11th August 2009]

[5] Ananthanpadmanabhan K.P., Moore D.J., Subramanyan K. (2004) Cleansing Without Compromise: The Impact of Cleansers on the Skin Barrier and the Technology of Mild Cleansing *Dematologic Therapy* 17: 16-25.

[6] Ehlers C., Ivens U.I., Møller M.L., Senderovitz T., Serup J. (2001) Females Have Lower Skin Surface pH Than Men *Skin Res. Technol.* 7: 90-94.

[7] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-Octanol/Water), High Performance Liquid Chromatography (HPLC) Method*, Paris, The Organisation for Economic Cooperation and Development OECD.

[8] Schultz T.W. (1997) Tetratox: *Tetrahymena pyriformis* Population Growth Impairment Endpoint – A Surrogate for Fish Lethality *Toxicol. Methods* 7: 289-309.

[9] <http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/Product_Information_Sheet/1/d5773pis.Par.0001.File.tmp/d5773pis.pdf> [Accessed 13 May 2011]

[10] Dulbecco R., Vogt M.J. (1954) Plaque Formation and Isolation Lines with Poliomyelitis Viruses *J. Exp. Med.* 99: 167-182.

[11] Caldwell G.W., Masucci J.A., Evangelisto M., White R. (1998) Evaluation of the Immobilized Artificial Membrane Phosphatidylcholine Drug Discovery Column for High-Performance Liquid Chromatographic Screening of Drug-Membrane Interactions *J. Chromatogr. A* 800: 161-169.

[12] Masucci J.A., Caldwell G.W., Foley J.P. (1998) Comparison of the Retention Behavior of β-Blockers Using Immobilized Artificial Membrane Chromatography and Lysophospholipid Micellar Electrokinetic Chromatography *J. Chromatogr. A* 810: 95-103.

[13] Luo H., Zheng C., Cheng Y.K. (2007) The Retention Properties of Nucleobases in Alkyl C_8-/C_{18}- and IAM- Chromatographic Systems in Relation to Log P_{OW} *J. Chromatogr. B* 847: 245-261.

[14] Synder L.R., Kirkland J.J., Glajh J.L. (1997) *Practical Guide to HPLC Method Development*, 2nd edition, New York, John Wiley & Sons, pp 655-660.

[15] Demare S., Roy D., Legerdre J.Y. (1999) Factors Governing the Retention of Solutes on Chromatographic Immobilized Artificial Membranes: Application to Anti-Inflammatory and Analgesic Drugs *J. Liq. Chromatogr. Relat. Technol.* 22: 2675-2688.

[16] Lázaro E., Ráfols C., Rosés M. (2005) Characterization of Immobilized Artificial Membrane (IAM) and XTerra Columns by Means of Chromatographic Models *J. Chromatogr. A* 1081: 163-173.

[17] Lázaro E., Ráfols C., Abraham M.H., Rosés M. (2006) Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns *J. Med. Chem.* 49: 4861-4870.

[18] Sprunger L., Blake-Taylor B.H., Wairegi A., Acree Jr. W.E., Abraham M.H. (2007) Characterization of the Retention Behavior of Organic and Pharmaceutical

Drug Molecules on an Immobilized Artificial Membrane Column with the Abraham Model *J. Chromatogr. A* 1160: 235-245.

[19] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.

[20] Lerner R.G., Trigg G.L. (1990) *Encyclopaedia of Physics*, 2nd edition, Weinheim, Wiley-VCH Verlag, pp 1041.

[21] Hewitt M., Cronin M.T.D., Enoch S.J., Madden J.C., Roberts D.W., Dearden J.C. (2009) In Silico Prediction of Aqueous Solubility: The Solubility Challenge *J. Chem. Inf. Model.* 49: 2572-2587.

[22] Yoon C.H., Kim S.J., Shin B.S., Lee K.C., Yoo S.D. (2006) Rapid Screening of Blood-Brain Barrier Penetration of Drugs Using the Immobilized Artificial Membrane Phosphatidylcholine Column Chromatography. *J. Biomol. Screen.* 11: 13-20.

[23] Genty M., González G., Lere C., Desangle-Gouty V., Legendre J. (2001) Determination of the Passive Absorption Through the Rat Intestine using Chromatographic Indices and Molar Volume *Eur. J. Pharm. Sci.* 12: 223-229.

[24] Chan E.C.Y., Tan W.L., Ho P.C., Fang L.J. (2005). Modeling Caco-2 Permeability of Drugs Using Immobilized Artificial Membrane Chromatography and Physicochemical Descriptors *J. Chromatogr. A* 1072: 159-168

[25] Barbato F., Cirocco V., Grumetto L., La Rotonda M.I. (2007) Comparison Between Immobilized Artificial Membrane (IAM) HPLC Data and Lipophilicity in n-Octanol for Quinolone Antibacterial Agents *Eur. J. Pharm. Sci.* 31: 288-297.

[26] Vrakas D., Giaginis C., Tsantili-Kakoulidou A. (2008) Electrostatic Interactions and Ionization Effect in Immobilized Artificial Membrane Retention A Comparative Study with Octanol-Water Partitioning. *J. Chromatogr. A* 1187: 67-78.

[27] Alvarez F.M., Bottom C.B., Chikhale P., Pidgeon C. (1993) Immobilised Artificial Membrane Chromatography. Prediction of Drug Transport across Biological Barriers In Ngo N.T. ed. *Molecular Interactions in Bioseparations*, New York, Plenum Press, pp 151-167.

[28] Hoest J., Christensen I.T., Jørgensen F.S., Hovgaard L., Frokjaer S. (2007) Computational Prediction of Solubilizer' Effect on Partitioning *Int. J. Pharm.* 329: 46-52.

[29] Kotecha J., Shah S., Rathod I., Subbaiah G. (2008) Prediction of Oral Absorption in Humans by Experimental Immobilized Artificial Membrane Chromatography Indices and Physicochemical Descriptors *Int. J. Pharm.* 360: 96-106.

[30] Li J., Cui S., He Z. (2006) Quantitative Structure-Retention Relationship Studies Using Immobilized Artificial Membrane Chromatography I: Amended Linear Solvation Energy Relationships with the Introduction of a Molecular Electronic Factor *J. Chromatogr. A* 1132: 174-182.

[31] Li J., Sun J., He Z. (2007) Quantitative Structure-Retention Relationship Studies with Immobilized Artificial Membrane Chromatography II: Partial Least Squares Regression *J. Chromatogr. A* 1140: 174-179.

[32] Tetko I.V., Gasteiger J.,Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., Prokopenko V.V. (2005) Virtual Computational Chemistry Laboratory – Design and Description *J. Comput. Aid. Mol. Des.* 19: 453-63.

[33] VCCLAB, Virtual Computational Chemistry Laboratory(2005) available from <http://www.vcclab.org>

[34] Mannhold R., Poda G.I., Ostermann C., Tetko I.V. (2009) Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds *J. Pharm. Sci.* 98: 861-893.
[35] Dearden J.C., Bresnen G.M. (1988) The Measurement of Partition Coefficients and Lipophilicity *Quant. Struct.-Act. Relat.* 7: 133-144.
[36] Tetko I.V., Bruneau P. (2004) Application of ALOGPS to Predict 1-Octanol/Water Distribution Coefficients, Log P, and Log D, of AstraZeneca In-House Database *J. Pharm. Sci.* 93: 3101-3110.
[37] Yu Y., Yang W., Gao Z., Lam M.H.W., Liu X., Wang L., Yu H. (2008) RP-HPLC Measurement and Quantitative Structure-Property Relationship Analysis of the n-Octanol-Water Partitioning Coefficients of Selected Metabolites of Polybrominated Diphenyl Ethers *Environ. Chem.* 5: 332-339.
[38] Meylan W.M., Howard P.H. (1995) Atom/Fragment Contribution Method for Estimation Octanol-Water Partitioning Coefficients *J. Pharm. Sci.* 84: 83-91.
[39] Ebbing D.D., Gammon S.D. (2009) *General Chemistry*, 9th edition, Belmont, Houghton Mifflin, pp 681.
[40] Pehourcq F., Jarry C., Bannwarth B. (2003) Potential of Immobilized Artificial Membrane Chromatography for Lipophilicity Determination of Arylpropionic Acid Non-Steroidal Anti-Inflammatory Drugs *Pharm. Res.* 33: 137-144.
[41] Minitab Inc. (2007) Minitab® Statistical Software version 15, Coventry: Minitab Inc..
[42] Rowe P.H. (2007) *Essential Statistics for the Pharmaceutical Sciences*, Chichester, Wiley.

4 Prediction of log $k_{IAM\ (pH\ 7.4)}$ using structural fragments and correction factors

4.1 Introduction

Hydrophobicity is an intrinsic chemical property that relates to a chemical's ability to partition between a polar and non-polar phase. Traditionally, hydrophobicity has been described by the logarithm of the octanol-water partition coefficient (log P or log K_{ow})[1]. Log P has numerous applications, including being a common descriptor in Quantitative Structure-Activity Relationships (QSARs) to model the biological activity of chemicals from their physico-chemical properties[2].

Interest in predicting log P has been on-going for many decades and the principle is well established. For instance, Cohen and Edsal[3] were the first to notice the additive effect of lipophilicity for amino acids in 1943. Fujita[4] and co-workers published the initial findings on the additive constituent nature of the lipophilicity of small organic molecules in 1964. There are at least three main approaches to predict hydrophobicity from the contribution of fragments and molecular sub-structures (as opposed to whole molecule effect). The first is based on the substitution of a parent molecule to determine hydrophobic substitution constants (π values)[5]. A second approach involves building the structure of a molecule from first principles based on fundamental fragments, so that the molecule is constructed from a set of fragments and assigning hydrophobic values to each fragment. The third approach also involves fragments, however, molecules are broken down into fragments and values assigned based on a statistical approach e.g. such as multiple regression analysis[6].

Hansch and Leo[5] determined π values for aromatic substituents, based on the substitution of a parent compound, and generated π substituent values using the relationship shown in equation (4.1). If the hydrophobicity parameter e.g. log P, of a parent molecule is known, it is possible to determine the relative effect of the substituent, on the hydrophobicity of the substituted molecule.

$$\pi_x = \log P_x - \log P_{parent} \quad (4.1)$$

where:

π_x is the hydrophobic substitution constant of the substituent X;

$\log P_x$ is the hydrophobicity of a substituted compound;

$\log P_{parent}$ is the hydrophobicity of the unsubstituted parent compound.

In the above relationship a positive value for π_x means, relative to the parent compound, the substituent increases the hydrophobicity of the compound. Conversely a negative value for π_x means, relative to the parent compound, the substituent reduces the hydrophobicity of the compound. This relationship has been shown to be useful in correlation analysis particularly for aromatic compounds[5].

Analysis of fragments and factor values provides fundamental information regarding their relative hydrophobicity. For instance, Davis[7, 8] showed a significant difference between the effects on the partition co-efficients of CH_3 and CH_2 and that there is a difference in CH_3 values, depending upon whether it is attached to a ring or is the terminal carbon in an aliphatic chain. However, the substitution of a parent molecule, as discussed above, to determine π values, assigns a partition co-efficient of zero to a hydrogen atom. If CH_3 and CH_2 have different partition co-efficients then hydrogen cannot have a value of zero. This shows the fundamental error in the π substitution value approach for predicting log P[6].

For aliphatic compounds Hansch and Leo developed fragments and correction factors based on a constructionist approach. Correction factors include, but are not limited to, multiple halogenations on the same carbon atom, chain branching and Proximity Effects of H-polar groups (P.E 1, P.E 2 and P.E 3, where the number indicates the separation between polar groups, illustrated in Figure 23). The constructionist approach developed from simple fragments, i.e. CH_3, CH_2, CH, C and H using the log P values for hydrogen, methane and ethane[9]. Factor values were developed to account for additional structural elements that affect the partitioning equilibrium, such as double and triple bonds, ring closure and the proximity effect of H-polar fragments[5].

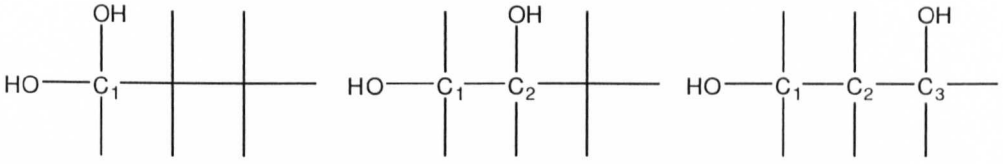


Figure 23 - Proximity effect with one, two and three carbons separation

A reductionist approach was adopted by Nys and Rekker[6] to generate fragments to calculate log P. They used statistical analysis of available log P data for a large number of structures to calculate values for CH_3, CH_2 and CH and other structural fragments as well as correction factors, that included Proximity Effect (P.E. 1 and P.E. 2), ring closure, cross conjugation and chain folding[6]. This reductionist fragment approach was updated by Mannhold *et al.*[10] to improve the fragment values for halogens based on mono halogenated alkanes and to improve the correction factors i.e. condensed aromatics, H attached to electronegative fragments and proximity effects etc. It should be noted the many correction factors are common to both the reductionist and constructionist approach, however the values are different.

Meylan and Howard[11] developed an atom/fragment contribution method. Their approach is a reductionist approach, developed using multiple linear regression analysis of reliable experimental log P values. In the initial regression analysis atom and fragment values were calculated, determining the regression equation for log P. A second regression analysis was used to calculate correction factors. Log P values can be calculated by summing the counts of atom/fragments and correction factors based on the structure of a compound. This approach is 2-D and hence is not able to differentiate between isomers. Electronic interactions of multiple atom/fragments are corrected for by factor values.

A limitation of the fragment approach is the presence of unknown fragments in a molecule. If a fragment is not included in the training set, a compound containing this fragment would be outside the domain of the method. This is overcome in the atom contribution approach, because all atoms are considered in the training set, therefore, all combinations of atoms, in fragments are possible and thus in the domain. However, this approach can oversimplify 'unknown' fragments, for example if they require correction factors, these would not be included, leading to incorrect predictions. The atom contribution approach developed by Meylan and

Howard[11] has been incorporated into the KOWWIN[12] software for the estimation of log P values.

Log k_{IAM} is an alternative measure of hydrophobicity to log P. As mentioned in Section 1.9, IAM-HPLC has the potential to be a biologically more realistic surrogate for determining hydrophobicity for biological membranes[13]. For IAM-HPLC to be a useful indicator of hydrophobicity, for example as a descriptor in QSAR analyses, the ability to predict log $k_{IAM\ (pH\ 7.4)}$ is desirable.

4.2 Aim of the Chapter

The aim of this Chapter was to investigate the use of structural fragments and correction factors to predict log $k_{IAM\ (pH\ 7.4)}$. To achieve this, log $k_{IAM\ (pH\ 7.4)}$ values determined for 36 compounds in Chapter 3 were used. In addition the log $k_{IAM\ (pH\ 7.4)}$ values of 30 new compounds were determined.

The ability to predict log k_{IAM} values has a number of advantages, including:

1. The ability to design experiments for the determination of log k_{IAM} based on a predicted value.
2. A reduced requirement to perform the IAM-HPLC analysis, which reduces both time and cost for the determination of log k_{IAM}.
3. Wider use and, therefore, greater acceptance of log k_{IAM} as a descriptor for hydrophobicity.
4. QSARs developed using log k_{IAM} as a descriptor that can be determined either experimentally or predicted, are more broadly applicable.

Analysis of the fragment and factor values allowed for the investigation of the differences between octanol-water partitioning and IAM partitioning based on the π substitution values determined for the two systems. Overall, this allowed for the determination of fragment and factor values to predict log $k_{IAM\ (pH\ 7.4)}$ values without the requirement for experimental determination.

4.3 Method

4.3.1 IAM-HPLC log $k_{IAM\ (pH\ 7.4)}$ values

In total log $k_{IAM\ (pH\ 7.4)}$ was determined for 66 compounds; the values for 36 compounds were reported in Chapter 3 and those for a further 30 compounds are reported in this chapter. The compounds analysed include 23 aliphatic compounds and 43 aromatic compounds and cover a range of hydrophobicities. The new compounds were selected for analysis to determine log $k_{IAM\ (pH\ 7.4)}$ based on the follow criteria:

- The compound contains additional fragments for which a log $k_{IAM\ (pH\ 7.4)}$ contribution can be assessed.
- The compounds for which log $k_{IAM\ (pH\ 7.4)}$ was measured contained a fragment considered within other compounds - adding confidence to fragment values generated
- The analysis allowed investigation of basic interactions between groups

Unionised and ionised compounds were included in the analyses to predict log $k_{IAM\ (pH\ 7.4)}$ using structural features and correction factors. All compounds that were ionised at pH 7.4 were 99% or greater in the ionised form. Ionised and unionised compounds were considered together. Therefore, additional ionised fragments were included to account for ionisation i.e. OH and O^-. The 30 further compounds analysed in this Chapter were of high purity (98% or greater) and used without further purification (with the exception of cyclopentadiene which was fractionally distilled at a temperature of 40-42°C prior to use to ensure analysis was performed on the monomer, the distillate was kept in an ice bath, the sample was prepared and analysed by IAM-HPLC immediately, triplicate preparation occurred on one day for this compound.) and commercially available as detailed in Chapter 3. Compounds analysed by IAM-HPLC are detailed in Table 22, in addition the structures and details of each compounds contribution to the analysis is detailed in Appendix 1.3 Table 22. Compounds from the reduced literature dataset were not included in the determination of fragment and factor values. Many of the compounds within the reduced dataset are complex molecules, containing multiple additional fragments, and potentially additional factor values per compound. In addition there is the

potential for additional, or more complex interactions between groups. Using experimental values obtained from a consistent experimental procedure that has been demonstrated to be robust increases confidence in the log $k_{IAM\ (pH\ 7.4)}$ values used to determine fragment and factor values. Although the reduced dataset contained experimental values obtained under broadly similar conditions, not all conditions were reported and variability could still be caused by differences in experimental parameters.

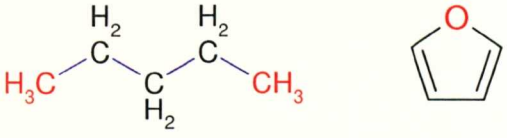
4.3.1.1 **Substituent π values**

In order to develop $\pi_{log\ k_{IAM\ (pH\ 7.4)}}$ substitutent values and to determine potential differences between octanol-water and IAM-water partitioning, experimental log $k_{IAM\ (pH\ 7.4)}$ values were determined for mono-substituted benzenes. The log $k_{IAM\ (pH\ 7.4)}$ values for the substituted compound were subtracted from that for the parent compound, benzene, to give the π value for the substituent based on equation (4.2),[5] adapted from the relationship for log P, equation (4.1).

$$\pi_x = \log k_{IAM\ (pH\ 7.4)\ x\text{-benzene}} - \log k_{IAM\ (pH\ 7.4)\ benzene} \qquad (4.2)$$

4.3.1.2 **Fragment values and factor values for log k_{IAM}**

The structure of each compound for which log $k_{IAM\ (pH\ 7.4)}$ was determined was broken down into theoretical structural constituents i.e. fundamental fragments based on the compound's structure e.g. CH_{2ali} , CH_{ar}, OH_{ali}, OH_{ar} (where $_{ali}$ and $_{ar}$ refer to aliphatic and aromatic fragments respectively). The frequency of fragments i.e. how often each fragment occurred in the molecule was recorded. These theoretical structural constituents are referred to as fundamental fragments (note that functional groups are not broken down into constituent atoms). In addition to the fundamental fragments, factor values for each compound were also calculated (i.e. ring closure, bond factors (the total number of bonds between fragments within a molecule minus one indicating molecular flexibility), aromatic-aromatic conjugation). It should be noted that aliphatic and aromatic environments were treated as distinct and different i.e. the OH group in pentanol is different to OH group in phenol.



Pentane Furan

Pentane = 2 (CH_3) + 3 (CH_2) + 3 (bond factors)*

Furan = 4 ($CH_{aromatic}$) + 1 ($O_{aromatic}$)

* Bond factor is the total number of bonds between fragments minus one, indicating molecular flexibility

Figure 24 - Theoretical structural fragments and factors for pentane and furan

Figure 24 shows the structures for both pentane and furan and how these molecules were broken down into fundamental fragments and factors. The same approach was applied to all molecules. Additional fragments were identified as compounds containing functional groups not considered previously were analysed. The frequency and occurrence of the fragments and factors were calculated manually.

4.3.1.3 **Statistical analysis**

Multiple linear regression (MLR) analysis was performed to determine the co-efficients for the theoretical fragments and factors identified for each structure analysed. MLR used the counts of the frequency of the theoretical fragments and structural features as the independent variable and the experimental log $k_{IAM\ (pH\ 7.4)}$ as the dependent variable. The analysis was performed on the log $k_{IAM\ (pH\ 7.4)}$ values for all compounds analysed with the exception of ethoxylated tetrabromobisphenol A and N,N'-dicyclohexylcarbodiimide. Due to the uniqueness of their structures compared to the other compounds analysed, inclusion of these compounds would require the calculation of multiple new fragments per compound which would generate a number describing the combined fragments. MLR analysis was performed using SPSS v15.0[14]. From the resulting regression equation the specific fragment and factor values were taken from its regression coefficients (refer to Table 24).

For all analyses an additional 'compound zero' was added. This is a zero response for the absence of structural fragments or factors. The inclusion of this structural information into the analysis encourages a small constant term and generates fragments that are chemically sensible in terms of the additive/subtractive nature of

the fragments and factors. Additionally a single CH_{ar} value (CH_{ar} = 0.158) was included for the regression analysis of aromatic compounds; this is determined from the log $k_{IAM\ (pH\ 7.4)}$ value for benzene i.e. 0.95/6 = 0.158. This fundamental calculation is derived from Hansch and Leo[5] who classified benzene as a well characterised compound for the determination of log P. They used this well characterised value in preference to predicted values generated using the fragment approach[5].

The following statistical information was recorded for the MLR analysis to determine fragment and factor values: n, $r^2_{(adj)}$, s, F and the Total Sum of Squares (SS_T). For the purposes of cross validating the fragment and factor values determined the following additional statistical information was also recorded Predicted REsidual Sum of Sqaures (PRESS) and the cross validated square of the correlation coefficient (Q^2)[15].

4.3.1.3.1 Cross Validation

The fragment and factor values obtained were evaluated using leave-one-out (LOO) analysis and K-fold cross validation. An external test set of compounds, for which log $k_{IAM\ (pH\ 7.4)}$ values were predicted and then determined experimentally by a second analyst in an external laboratory, were also investigated.

The Minitab[16] statistical software (version 15.1.1.0) was unable to determine PRESS or Q^2 for LOO regression analysis because some of the aliphatic and aromatic compounds analysed contained unique fragments. Therefore, leave-one-out analysis was performed manually. Only the compounds that did not contain unique fragments were omitted from the LOO analysis. In this approach, each compound was removed independently and the regression analysis re-performed, this generated new coefficients which were used to predict log $k_{IAM\ (pH\ 7.4)}$ for the LOO compound.

K-fold cross validation was performed as a second validation process, again this process was performed manually. All compounds containing unique fragments were excluded from the test set. The remaining compounds were given a unique identifier and were placed in numerical order. On each iteration, 25% of compounds were removed and the model generated using MLR analysis. The coefficients generated

were used to predict log $k_{IAM\ (pH\ 7.4)}$ for the compounds left out. This process was repeated so each of the four groups was left out once, resulting in four-fold cross validation.

In order to analyse an external test set, log $k_{IAM\ (pH\ 7.4)}$ was predicted for six compounds (three aliphatic and three aromatic compounds). The log $k_{IAM\ (pH\ 7.4)}$ values were determined experimentally using the same methodology as the training set. The predicted log $k_{IAM\ (pH\ 7.4)}$ values were compared to the experimental log $k_{IAM\ (pH\ 7.4)}$ values.

4.4 Results and discussion

4.4.1 Log $k_{IAM\ (pH\ 7.4)}$ values

Log $k_{IAM\ (pH\ 7.4)}$ values have been determined for 66 compounds, the results are summarised in Table 22. The results for compounds shaded in grey are reported in Section 3.4 with full experimental results available for reference in Appendix 1.2, Tables 7-9, 12 and 13. For compounds not previously analysed the full results are available in Appendix 1.3, Table 19 and 20. The plots for compounds where calibration to extrapolate to aqueous conditions was required, due to high hydrophobicity are in Appendix 1.3, Figures 8 to 12 for reference. Structures of all compounds analysed are shown in Appendix 1.3, Table 22 for reference.

4.4.2 π values for mono-substituted aromatic compounds

Log $k_{IAM\ (pH\ 7.4)}$ values were determined for 12 simple aromatic compounds and benzene and are summarised in Table 22. Applying the approach developed by Leo and Hansch[5] for log P to log $k_{IAM\ (pH\ 7.4)}$, the relative hydrophobicity parameter of a substituent was determined using knowledge of the parent molecule by applying the adapted relationship in equation (4.2).

Twelve fundamental π values were determined for the mono-substituted aromatic compounds. For example, the value for an aromatic methyl group is calculated using log $k_{IAM\ (pH\ 7.4)}$ for toluene and benzene as illustrated in Figure 25, using equation (4.2).

Compound	CAS No.	log $k_{IAM\ (pH\ 7.4)}$	log k_{IAM}	log P^{12}	pKa	% ionised at pH 7.4[a]
Aliphatic compounds						
Methanol	67-56-1	-1.92	-1.92	-0.77		
Dimethyl sulfoxide	67-68-5	-1.46	-1.46	-1.35		
Acetonitrile	75-05-8	-1.04	-1.03	-0.34		
Acetone	67-64-1	-0.89	-0.89	-0.24		
Butanone	78-93-3	-0.48	-0.48	0.29		
Pentanoic acid	109-52-4	-0.06	2.50	1.39	4.84	99.7
Hexanoic acid	142-62-1	-0.06	2.46	1.92	4.88	99.7
2-Pentanone	107-87-9	-0.03	-0.03	0.91		
Cyclohexanone	108-94-1	-0.00	0.00	0.81		
Pentylamine	110-58-7	0.24	0.24	1.49		
Pentanol	71-41-0	0.25	0.25	1.51		
Cyclopentadiene	542-92-7	0.64	0.64	2.25		
Cyclopentene	142-29-0	1.11	1.11	2.46		
Pentene	25377-72-4	1.50	1.50	2.66		
Cyclohexene	110-83-8	1.54	1.54	2.86		
Cyclopentane	287-92-3	1.55	1.55	3.00		
Dibutylether	142-96-1	1.78	1.78	3.21		
Pentane	109-66-0	2.00	2.00	3.39		
Cyclohexane	110-82-7	2.05	2.05	3.44		
Hex-1-ene	592-41-6	2.08	2.08	3.39		
Hex-2-ene	4050-45-7	2.10	2.10	3.00		
Hexane	110-54-3	2.70	2.70	3.90		

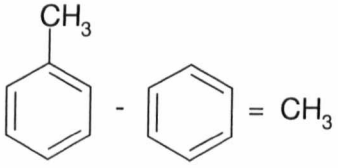
Compound	CAS No.	log $k_{IAM\ (pH\ 7.4)}$	log k_{IAM}	log P^{12}	pKa	% ionised at pH 7.4[a]
n,n'-Dicyclohexylcarbodiimide	538-75-0	3.15	3.15	5.11		
Mono-substituted benzenes						
Benzene (parent)	71-43-2	0.95	0.95	2.13		
Benzoic acid	65-85-0	-0.80	2.41	1.87	4.19	99.9
Aniline	62-53-3	0.21	0.21	0.90		
Phenol	108-95-2	0.65	0.65	1.46		
Benzonitrile	100-47-0	0.75	0.75	1.60		
Nitrobenzene	98-95-3	1.00	1.00	1.90		
Anisole	100-66-3	1.06	1.06	2.10		
Methylbenzoate	93-58-3	1.36	1.36	2.10		
Toluene	108-88-3	1.44	1.44	2.73		
Chlorobenzene	108-90-7	1.63	1.63	2.80		
Ethylbenzoate	93-89-0	1.78	1.78	2.64		
Bromobenzene	108-86-1	1.80	1.80	3.00		
Cumene	98-82-8	2.22	2.22	3.66		
Heterocyclic aromatics						
Pyrrole	109-97-7	-0.130	-0.130	0.75		
Furran	110-00-9	-0.05	-0.05	1.34		
Thiophene	110-02-1	0.36	0.36	1.81		
Di-substituted benzenes						
3-Aminobenzoic acid	99-05-8	-1.46	2.87	0.65	3.07	99.9
2-Nitrobenzoic acid	552-16-9	-1.11	3.82	1.46	2.47	99.9
4-Aminophenol	123-30-8	-0.44	1.48	0.04	5.48	98.8

Compound	CAS No.	log $k_{IAM\ (pH\ 7.4)}$	log k_{IAM}	log P^{12}	pKa	% ionised at pH 7.4[a]
Hydroquninone	123-31-9	-0.17	-0.167	0.59		
1,2-Dihydroxybenzene	120-80-9	0.10	0.10	0.88		
4-Aminobenzoic acid	150-13-0	0.38	5.40	0.83	2.38	99.9
3-Iodobenzoic acid	618-51-9	0.41	3.96	3.13	3.85	99.9
4′-Hydroxyacetophenone	99-93-4	0.72	0.72	1.35		
3-Nitroaniline	99-09-2	0.85	0.85	1.37		
3-Hydroxyacetophenone	121-71-1	0.87	0.87	1.39		
Resorcinol	108-46-3	0.10	0.103	0.80		
2'-Hydroxyacetophenone	582-24-1	1.09	1.09	1.92		
4-Cresol	106-44-5	1.21	1.21	1.94		
2-Chlorophenol	95-57-8	1.35	1.35	2.15		
4-Chlorophenol	106-48-9	1.74	1.74	2.39		
3-Chlorophenol	108-43-0	1.81	1.81	2.50		
Tri-substituted benzenes						
1,2,4-Trichlorobenzene	120-82-1	2.97	2.97	4.02		
Conjugated Aromatic compounds						
1-Naphthol	90-15-3	2.25	2.25	2.85		
Naphthalene	91-20-3	2.48	2.48	3.30		
Biphenyl-4-ol	92-69-3	2.77	2.77	3.20		
Biphenyl	92-52-4	3.13	3.13	3.98		
Bibenzyl	103-29-7	3.63	3.63	4.79		
Fluoranthene	206-44-0	4.26	4.26	5.16		
2-Terphenyl	84-15-1	4.29	4.29	5.52		

Compound	CAS No.	log $k_{IAM\ (pH\ 7.4)}$	log k_{IAM}	log P[12]	pKa	% ionised at pH 7.4[a]
Ethoxylated tetrabromobisphenol A	4162-45-2	4.30	4.30	4.09		
Triphenylamine	603-34-9	4.54	4.54	5.74		
4-Terphenyl	92-94-4	4.93	4.93	6.03		

[a] Percentage ionised calculated from substitution of Henderson-Hasselbach equation[17], equation (3.1)

Table 22 – Name, CAS number, experimental log k_{IAM} value and experimental log P values for samples analysed. In addition for compounds ionised under the conditions of analysis (pH 7.4) log $k_{IAM\ (pH\ 7.4)}$, pKa and percentage ionised are also reported (compounds shaded have been analysed previously and discussed in Chapter 3)



Log $k_{IAM\ (pH\ 7.4)}$ 1.44 - 0.95 = 0.49

Figure 25 – Toluene and benzene allow the calculation of the π substituent value for the CH_3 fragment

Figure 26 illustrates the mono-substituted aromatic compounds analysed and the π values derived using equation (4.2). Table 23 summarises the experimental log $k_{IAM\ (pH\ 7.4)}$ π values determined and the equivalent π values reported by Hansch and Leo[5] for log P. For comparison compounds for which π values were not reported by Hansch and Leo[5], were calculated in this study by applying the same approach using log P values from KOWWIN[12].

The relationship between the log P π and log $k_{IAM\ (pH\ 7.4)}$ π is shown in Figure 27. There is a strong positive correlation with $r^2 = 0.96$. There is one clear outlier (circled) for the carboxylic acid group in the ionised form $C(=O)O^-$. This compound was not included in the determination of r^2. The $C(=O)O^-$ is an expected outlier due to the fragment's ionised state and the differences between IAM-HPLC and octanol-water partitioning discussed in Section 1.7.5 and Section 1.9. This confirms the log P π value for the carboxylic acid fragment is in the unionised form.

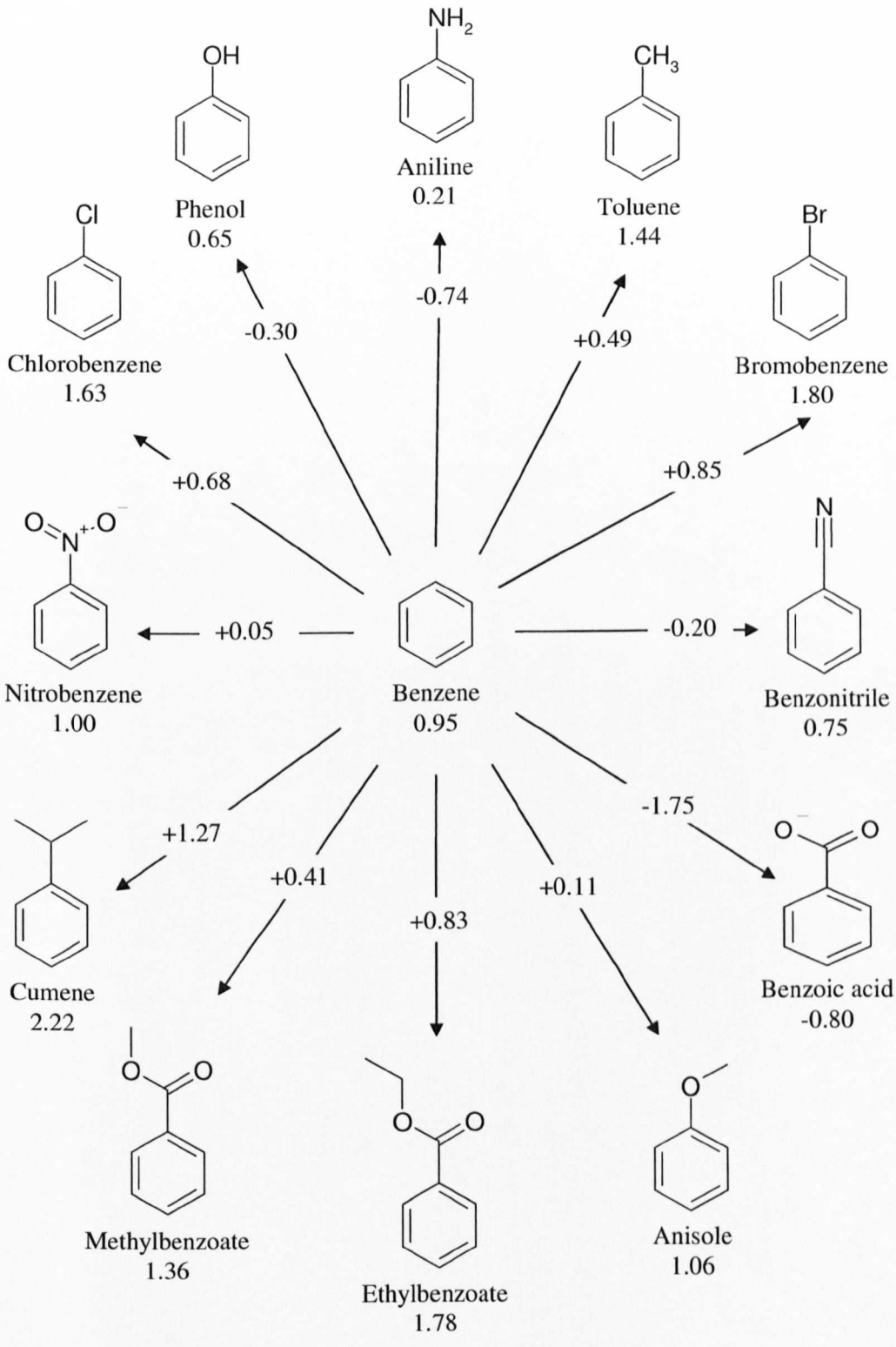


Figure 26 – Mono-substituted aromatic compounds, for which experimental log $k_{IAM\ (pH\ 7.4)}$ values are available, and π values derived from equation (4.2) for the substituent.

Fragment	SMILES	Log $k_{IAM\ (pH\ 7.4)\ \pi\ value}$	Log $P_{\pi\ value}$ [5]
NH_2	N	-0.74	-1.23
OH	O	-0.30	-0.67
OCH_3	O(C)	0.11	-0.02
NO_2	[N+](=O)[O-]	0.05	-0.28
Cl	Cl	0.68	0.71
CN	C#N	-0.20	-0.57
CH_3	C	0.49	0.56
Br	Br	0.85	0.86
COO^-	$C(=O)O^-$	-1.75	-0.32
$COOCH_3$	C(=O)OC	0.41	-0.01*
$COOCH_2CH_3$	C(=O)OCC	0.83	0.51*
$CHCH_3CH_3$	C(C)(C)	1.27	1.53*

* log $P_{\pi\ value}$ calculated using experimental log P values from the PhysProp database available in KOWWIN[12]

Table 23 - List of Fragments and SMILES notation for fragment values generated from the substitution of benzene for log $k_{IAM\ (pH\ 7.4)}$ and the equivalent fragment determined by Hansch and Leo, or calculated using the same relationship for log P

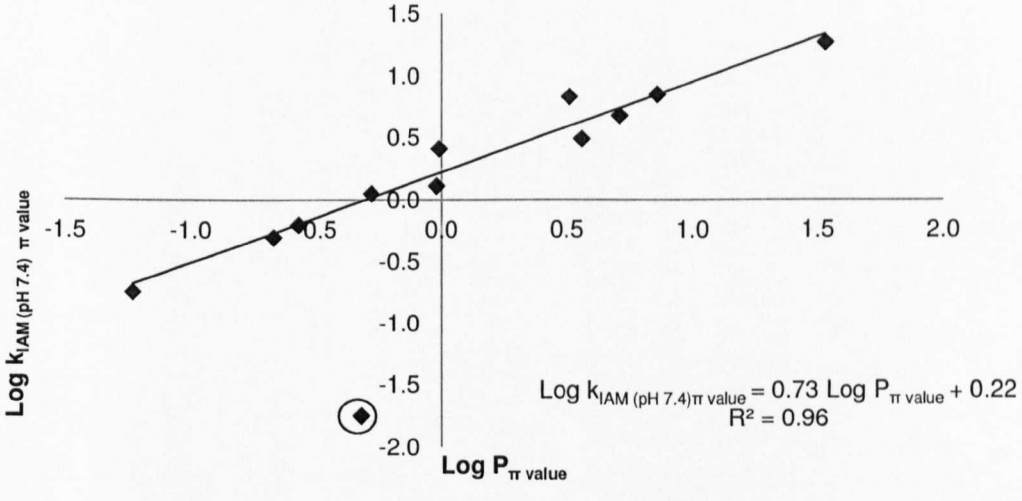


Figure 27 - Relationship between π values derived from values for log $k_{IAM\ (pH\ 7.4)}$ and Hansch and Leo[5] π values for log P. Circled fragment is for the carboxylic acid group in the ionised form $C(=O)O^-$.

As can be seen from Table 23 and Figure 27, the majority of π values for log P and log $k_{IAM\ (pH\ 7.4)}$ are proportional. This is to be expected as the same method of calculation was applied and since both log P and log $k_{IAM\ (pH\ 7.4)}$ are measures of hydrophobicity.

Analysis of the π substituent values determined for the methoxy group O(C), the nitro group [N+](=O)[O-] and the acetyl group C(=O)OC differ in sign for log P and log $k_{IAM\ (pH\ 7.4)}$ method of determination, however, all values are close to zero. Whilst there is a difference in sign for these π substitution values, this does not indicate that these compounds are outliers. Instead it may indicate a difference in partitioning between octanol-water and membrane-water. The identification of such differences is similar to the differences in the partition coefficients for each of the four systems described by the 'critical quartet'[18] of partitioning systems which can have values of different signs, as detailed in Chapter 1, section 1.7.5 and Table 2.

The investigation of substitution of a parent molecule to determine π values for both log P or log $k_{IAM\ (pH\ 7.4)}$ implicitly assigns a partitioning value of zero to hydrogen, as discussed above. This is a fundamental error in this approach that allows for comparison of simple substitutions only. Therefore, the ability to predict log $k_{IAM\ (pH\ 7.4)}$ based on fragment values was also investigated.

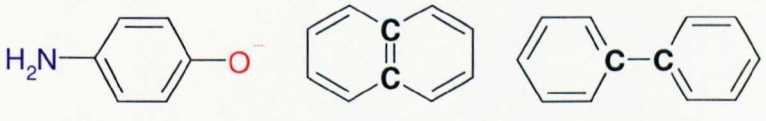
4.4.3 **Fragments**

Two fragment approaches have been developed for log P, a constructionist approach using fundamental fragments and a reductionist approach based on multiple regression analysis. Both approaches are described in section 4.1. The reductionist approach has been applied in this investigation to the log $k_{IAM\ (pH\ 7.4)}$ values determined experimentally.

The investigation utilised the log $k_{IAM\ (pH\ 7.4)}$ values determined experimentally for the 66 compounds, described previously. The structure of each compound analysed was broken down into theoretical fragments and structural features. Counts, or the frequency of occurrence of fragments and structural features were recorded.

As discussed above, Hansch and Leo[5] used the experimental log P value for benzene to calculate the CH_{ar} fragment in preference to calculated values from their analyses, determining the fragment for CH_{ar} as 0.35 (2.13/6 = 0.35), for log P. Applying this approach to log $k_{IAM\ (pH\ 7.4)}$ values, CH_{ar} was determined to be 0.16. Calculated values of CH_{ar} using MLR analysis were similar to this value. Therefore, further adjustments were not made.

Considering all 66 compounds analysed in this investigation, the compounds were broken down into theoretical structural fragments and features. Two compounds were excluded from further analysis due to the unique nature of the structure, these were N,N'-dicyclohexylcarbodiimide and ethoxylated tetrabromobisphenol A. For the remaining 64 compounds, 36 fragments and factors were required to describe the chemical structure of all compounds. Full fragment and factor counts are listed in Appendix 1.3 Table 23 and 24. The fragments and factor counts were identified using the method in Figure 24. It is important to note that all determinations of log k_{IAM} were performed at pH 7.4, therefore some compounds and therefore, theoretical fragments are ionised at this pH i.e. 4-aminophenol is ionised under the conditions of analysis, therefore, the fragment considered is also ionised. The fragments (OH_{ar}) and (O^-_{ar}) are taken to be distinct and separate fragments. Additionally, for conjugated aromatic systems there are three distinct aromatic carbon environments, these are CH_{ar}, C_{ar} and $C_{ar\ con}$. Where CH_{ar} are unsubstituted aromatic carbons, C_{ar} is an aromatic carbon with a substituent attached and $C_{ar\ con}$ accounts for aromatic conjugation within a structure as illustrated in Figure 28. The factor proximity effect 2 (P.E. 2) is the separation of electronegative groups based on two carbons separation (Figure 23). Given the struxtures of the compounds, for which fragment and factors were calculated, proximity effects for other carbon separation distances were not required.



4-Aminophenol Naphthalene Biphenyl

4-Aminophenol = 4 (CH_{ar}) + 2 (C_{ar}) + 1 (O^-_{ar}) + 1 ($NH_{2\ ar}$) +1 (bond factors)

Naphthalene = 8 (CH_{ar}) + 2 ($C_{ar\ con}$)

Biphenyl = 10 (CH_{ar}) + 2 (C_{ar}) + 1 ($C_{ar\ con}$)

Figure 28 - Theoretical structural fragments and factors for 4-aminophenol, naphthalene and biphenyl

In this investigation 64 compounds contribute 36 fragments, giving a compound/descriptor ratio of 1.78. MLR analysis was performed using the log $k_{IAM\ (pH\ 7.4)}$ values (reported in Table 22) as the dependent variable and the fragment and

factor counts (reported in Appendix 1.3 Table 23 and Table 24) as the independent variables. Methanol was included in the analysis, despite the variability in the log $k_{IAM\ (pH\ 7.4)}$ discussed in Section 3.4.1. The fragment and factor values determined (the coefficients from the MLR regression analysis, full equation not shown) are reported in Table 24. MLR analysis gives the following statistics:

$n = 64$, $r^2_{(adj)} = 0.956$, $s = 0.380$, $F = 40$, $F_{36, 27}\ \alpha, 0.001 = 3.36$, $SS_T = 214$

Factor		**Log $k_{IAM\ (pH\ 7.4)}$ Value**	**Log P value[6]**
Bond		0.14	-0.12
Ring closure		-0.76	-0.09
Proximity effect (2)		-0.27	
Fragment		**Log $k_{IAM\ (pH\ 7.4)}$ Value**	**Log P value[6]**
-O-	Heteroaromatic in ring substituent	-0.77	-0.08
-NH-		-0.85	-0.65
-S-		-0.36	0.36
CH_{ar}	Aromatic fragments	0.18	0.35
C_{ar}		0.30	0.22
$C_{ar\ con}$		0.44	0.44
OH		-0.55	-0.44
O^-		-1.47	
NH_2		-0.44	-1.00
COO^-		-1.78	-4.13
NO_2		-0.30	-0.03
Cl		0.51	0.94
CN		-0.45	-0.34
Br		0.60	1.09
I		1.00	1.35
CH_3		0.27	
CH_2		0.47	
C=O		-0.89	-1.09
C(=O)O		-0.33	-0.56
CH		-0.20	

Fragment		Log $k_{IAM\ (pH\ 7.4)}$ Value	Log P value[6]
N	Aromatic fragments	0.66	
ar-O-al		-0.59	
CH_3	Aliphatic fragments	0.31	0.89
CH_2		0.35	0.66
$CH_{2\ unsat}$		-0.01	
CH_{unsat}		0.12	
OH		-2.13	-1.64
C=O		-1.73	-1.90
S=O		-2.23	-3.01
NH_2		-2.04	-1.54
CN		-1.35	-1.27
COO^-		-2.09	-5.19
ali-O-ali		-1.94	-0.61

Table 24 - Fragment and factor values determined by MLR analysis for all compounds, equivalent fragment values for log P are shown for comparison

The fragment and factor values generated by MLR analysis for all compounds are of a similar magnitude to the equivalent log P fragment values determined by Rekker[6]. The results show that for log $k_{IAM\ (pH\ 7.4)}$ the aliphatic CH_3 fragment is slightly less hydrophobic than the aliphatic CH_2 fragment. In addition the bond factor increases the hydrophobicity of a compound. This is opposite to the trend observed by both Rekker[6] and Hansch and Leo[5] and does not agree with knowledge of physical chemistry. It is also appreciated that for the validity or otherwise of the fragments, the recommendation of the European Union Technical Guidance Document (TDG)[2] should be borne in mind. This states that:

> "for all fragment/group contribution models the compound/descriptor ratio must be larger than 2"[2]

In this analysis 64 compounds have been analysed, for the determination of 36 fragments and factors. This is a compound/descriptor ratio of 1.78. It is also noted that the bond factor count and CH_2 count are highly correlated. Together these factors may account for the discrepancies in the values determined. To increase

confidence in the CH_2, CH_3 and bond factor, log $k_{IAM\ (pH\ 7.4)}$ should be determined for a larger dataset to improve the accuracy of the values generated. This would need to be undertaken without the introduction of new fragments and factors. A further suggestion would be the analysis of compounds containing CH and C (fully substituted) fragments. This may indicate if one fragment value is anomalous which has a disproportionate affect on the fragment and factor values.

Table 25 lists all compounds used to investigate fragment and factor values, the experimental log $k_{IAM\ (pH\ 7.4)}$ values reported previously in Table 22. The predicted log $k_{IAM\ (pH\ 7.4)}$ values were calculated using the fragment and factor counts in Appendix 1.3 Table 23 and Table 24 and the fragment and factor coefficients reported in Table 24. For example

Pentane = 2 (CH_3) + 3 (CH_2) + 3 (bond factors)
Pentane = 2(0.31) + 3 (0.35) + 3 (0.14) = 2.09 predicted
= 2.00 experimental

Furan = 4 ($CH_{aromatic}$) + 1 ($O_{aromatic}$)
Furan = 4 (0.18) + 1 (-0.77) = -0.05 predicted
= -0.05 experimental

4-Aminophenol = 4 (CH_{ar}) + 2 (C_{ar}) + 1 (O^-_{ar}) + 1 ($NH_{2\ ar}$) +1 (bond factors)
4-Aminophenol = 4 (0.18) + 2 (0.30) + 1 (-1.47) + 1 (-0.44) +1 (0.14) = -0.45 predicted
= -0.44 experimental

Naphthalene = 8 (CH_{ar}) + 2 ($C_{ar\ con}$)
Naphthalene = 8 (0.18) + 2 (0.44) = 2.32 predicted
= 2.48 experimental

Biphenyl = 10 (CH_{ar}) + 2 (C_{ar}) + 1 ($C_{ar\ con}$)
Biphenyl = 10 (0.18) + 2 (0.30) + 1 (0.44) = 2.84
= 3.13 experimental

1,2-Dihydroxybenzene = 4 (CH_{ar}) + 2 (C_{ar}) + 2(OH) + 1 (bond factors) + 1 (P.E.2)

1,2-Dihydroxybenzene = 4 (0.18) + 2 (0.30) + 2(-0.55) + 1 (0.14) + 1 (-0.27) = 0.09 predicted

= 0.10 experimental

Nineteen compounds were identified and are highlighted as the residual is outside of the range of ±0.2 log units, which, for the purposes of this investigation, was considered to be acceptable variation in predicted log $k_{IAM\ (pH\ 7.4)}$ values. 0.2 log units is an arbitrary value and double that of the OECD guideline for the experimental determination of log P by RP-HPLC[19]. The OECD guideline of ±0.1 log units refers to the experimental determination of log P whereas, here we are making predictions based on experimental measurements. As such, any inherent error within the experimental values propagates through to the subsequent predictions. Therefore the allowable experimental error is increased.

Compound	Identifier for cross validation	Experimental log $k_{IAM\ (pH\ 7.4)}$	Predicted log $k_{IAM\ (pH\ 7.4)}$	Residual	Unique Fragment
Zero	.	0.00	0.00	0.00	
CH_{ar}	.	0.16	0.18	-0.02	
Pyrrole	.	-0.13	-0.13	0.00	-NH-
Furan	.	-0.05	-0.05	0.00	-O-
Thiophene	.	0.36	0.36	0.00	-S-
Anisole	.	1.06	1.06	0.00	$_{ar}$-O-$_{ali}$
Benzonitrile	.	0.75	0.75	0.00	C#N
Bromobenzene	.	1.80	1.80	0.00	Br
Bibenzyl	.	3.63	3.62	0.01	CH_2
Triphenylamine	.	4.54	4.54	0.00	N
Cumene	.	2.22	2.21	0.01	CH
3-Iodobenzoic acid	.	0.41	0.41	0.00	I
4-Aminophenol	.	-0.44	-0.45	0.01	O
Pentylamine	.	0.24	0.23	0.01	NH2
Dimethylsulfoxide	.	-1.46	-1.47	0.01	S=O
Dibutylether	.	1.78	1.76	0.02	-O-
Acetonitrile	.	-1.04	-1.04	0.00	C#N
Trans-2-hexene	1	2.10	2.12	-0.02	
Naphthalene	2	2.48	2.32	0.16	
Hexanoic acid	3	-0.06	0.18	-0.24	
Resorcinol	4	0.10	0.36	-0.26	
1,2,4-trichlorobenzene	5	2.97	3.25	-0.28	
2 Nitrobenzoic acid	6	-1.11	-0.89	-0.22	

Compound	Identifier for cross validation	Experimental log $k_{IAM\ (pH\ 7.4)}$	Predicted log $k_{IAM\ (pH\ 7.4)}$	Residual	Unique Fragment
Chlorobenzene	7	1.63	1.71	-0.08	
3-Nitroaniline	8	0.85	0.72	0.13	
Fluoranthene	9	4.26	4.44	-0.18	
3 Aminobenzoic acid	10	-1.46	-0.76	-0.70	
1,2-Dihydroxybenzene	11	0.10	0.09	0.01	
1-Naphthol	12	2.25	1.89	0.36	
3-Chlorophenol	13	1.81	1.42	0.39	
2-Chlorophenol	14	1.35	1.15	0.20	
Pentane	15	2.00	2.09	-0.09	
Pentanoic acid	16	-0.06	-0.31	0.25	
Cyclopentane	17	1.55	1.55	0.00	
Methanol	18	-1.92	-1.82	-0.10	
4-Cresol	19	1.21	1.18	0.03	
Pentanol	20	0.25	0.14	0.11	
2'-Hydroxyacetophenone	21	-0.17	0.47	-0.64	
Aniline	22	0.21	0.76	-0.55	
Benzene	23	0.95	1.08	-0.13	
Cyclopentene	24	1.11	1.09	0.02	
Cyclohexane	25	2.05	2.04	0.01	
Biphenyl	26	3.13	2.84	0.29	
4-Terphenyl	27	4.93	4.74	0.19	
4'-Hydroxyacetophenone	28	0.72	0.47	0.25	
Cyclohexanone	29	0.00	0.02	-0.02	
Pentene	30	1.50	1.54	-0.04	
Hex1ene	31	2.08	2.03	0.05	
4-Aminobenzoic acid	32	0.38	-0.76	1.14	
2-Pentanone	33	-0.03	0.01	-0.04	
4-Chlorophenol	34	1.74	1.42	0.32	
Acetone	35	-0.89	-0.97	0.08	
Cyclopentadiene	36	0.64	0.63	0.01	
Biphenyl-4-ol	37	2.77	2.55	0.22	
Ethylbenzoate	38	1.78	1.81	-0.03	
Hexane	39	2.70	2.58	0.12	
3'-Hydroxyacetophenone	40	0.87	0.47	0.40	
2-Terphenyl	41	4.29	4.74	-0.45	
Methylbenzoate	42	1.36	1.32	0.04	
Hydroquninone	43	-0.17	0.36	-0.53	
Phenol	44	0.65	0.65	0.00	
Cyclohexene	45	1.54	1.58	-0.04	
Toluene	46	1.44	1.47	-0.03	
Butanone	47	-0.48	-0.48	0.00	
Nitrobenzene	48	1.00	0.90	0.10	
Benzoic acid	49	-0.80	-0.58	-0.22	

Table 25 - Experimental and predicted log $k_{IAM\ (pH\ 7.4)}$ values (from values in Table 22 and Appendix 1.3 Tables 23 and 24) for all compounds, compounds containing unique fragments (only present in a single compound) were identified and random numbers assigned to the remaining compounds for cross validation

Figure 29 shows the relationship between experimental log $k_{IAM\ (pH\ 7.4)}$ values and the predicted log $k_{IAM\ (pH\ 7.4)}$ values from Table 25 and Appendix 1.3, Tables 23 and 24 for all compounds considered.

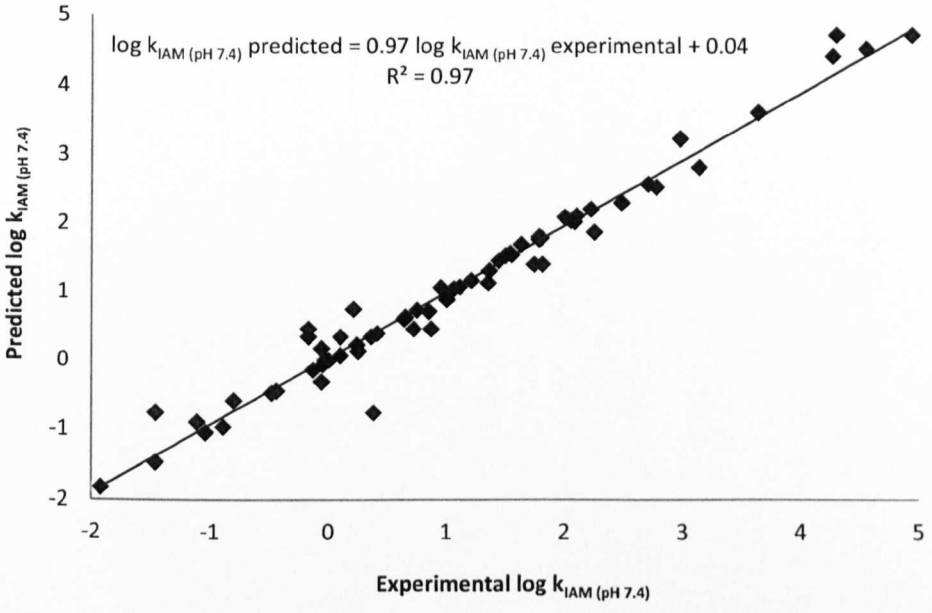


Figure 29 - Plot of experimental log $k_{IAM\ (pH\ 7.4)}$ against predicted log $k_{IAM\ (pH\ 7.4)}$ for all compounds, results reported in Table 25

It is clear from the strong rectilinear correlation in Figure 29 and the corresponding residuals from Table 25 that the fragment values determined provide a good estimate of the experimental log $k_{IAM\ (pH\ 7.4)}$ value. Cross validation of the fragment and factor values was undertaken using two approaches, leave-one-out (LOO) and K-fold (here K = 4). In addition, an external set of validation compounds was analysed.

For LOO cross validation, MLR analysis was performed using the 49 compounds that do not contain unique fragments. The coefficients for the fragment and factor values calculated from the regression analysis performed whilst leaving one compound out each time are reported in Table 26 to Table 28 (the red text highlights fragment and factor values not falling within 20% of the values determined for all compounds).

Fragment and factor values		Factor			Hetroaromatic in ring substituent			Aromatic fragment							
		bond	ring	P.E. 2	O	NH	S	CH_{ar}	C_{ar}	$C_{ar\ con}$	OH	O^-	NH_2	COO^-	NO_2
All compounds		0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
Compound removed	1	0.14	-0.77	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	2	0.13	-0.73	-0.28	-0.73	-0.81	-0.32	0.17	0.33	0.45	-0.56	-1.47	-0.44	-1.79	-0.30
	3	0.16	-0.83	-0.27	-0.78	-0.86	-0.37	0.18	0.29	0.44	-0.56	-1.47	-0.44	-1.79	-0.300
	4	0.15	-0.77	-0.33	-0.76	-0.84	-0.35	0.18	0.30	0.44	-0.50	-1.46	-0.45	-1.77	-0.28
	5	0.16	-0.80	-0.36	-0.71	-0.79	-0.30	0.17	0.33	0.47	-0.60	-1.47	-0.46	-1.77	-0.27
	6	0.15	-0.77	-0.13	-0.75	-0.83	-0.34	0.18	0.31	0.45	-0.58	-1.35	-0.56	-1.63	-0.13
	7	0.14	-0.75	-0.27	-0.78	-0.86	-0.37	0.18	0.30	0.44	-0.55	-1.46	-0.43	-1.78	-0.30
	8	0.14	-0.75	-0.25	-0.78	-0.86	-0.37	0.18	0.30	0.44	-0.55	-1.38	-0.52	-1.72	-0.40
	9	0.14	-0.75	-0.26	-0.70	-0.78	-0.29	0.16	0.27	0.63	-0.50	-1.41	-0.36	-1.71	-0.22
	10	0.15	-0.78	-0.32	-0.74	-0.82	-0.33	0.17	0.32	0.45	-0.55	-1.68	-0.24	-1.56	-0.43
	11	0.14	-0.76	-0.28	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.29
	12	0.16	-0.80	-0.24	-0.76	-0.84	-0.35	0.18	0.31	0.42	-0.60	-1.49	-0.44	-1.81	-0.32
	13	0.14	-0.76	-0.23	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.5	-1.480	-0.43	-1.80	-0.31
	14	0.15	-0.77	-0.43	-0.76	-0.84	-0.35	0.18	0.31	0.44	-0.55	-1.44	-0.47	-1.74	-0.25
	15	0.12	-0.73	-0.27	-0.76	-0.84	-0.35	0.18	0.31	0.44	-0.55	-1.46	-0.43	-1.78	-0.30
	16	0.16	-0.83	-0.27	-0.78	-0.86	-0.37	0.18	0.29	0.44	-0.56	-1.47	-0.44	-1.79	-0.30
	17	0.14	-0.75	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	18	0.07	-0.51	-0.27	-0.74	-0.82	-0.33	0.17	0.34	0.45	-0.54	-1.45	-0.42	-1.77	-0.29
	19	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	20	0.07	-0.51	-0.27	-0.74	-0.82	-0.33	0.17	0.34	0.45	-0.54	-1.45	-0.42	-1.77	-0.29
	21	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	22	0.11	-0.69	-0.14	-0.85	-0.93	-0.44	0.20	0.25	0.40	-0.56	-1.81	-0.05	-1.96	-0.40
	23	0.16	-0.79	-0.27	-0.81	-0.89	-0.40	0.19	0.27	0.42	-0.55	-1.46	-0.43	-1.79	-0.30
	24	0.14	-0.77	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	25	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	26	0.15	-0.78	-0.28	-0.74	-0.82	-0.33	0.17	0.30	0.46	-0.54	-1.46	-0.42	-1.77	-0.28
	27	0.14	-0.76	-0.28	-0.77	-0.85	-0.36	0.18	0.27	0.44	-0.52	-1.43	-0.40	-1.75	-0.26

Fragment and factor values		Factor			Hetroaromatic in ring substituent			Aromatic fragment							
		bond	ring	P.E. 2	O	NH	S	CH_{ar}	C_{ar}	$C_{ar\ con}$	OH	O^-	NH_2	COO^-	NO_2
All compounds		0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
Compound removed	28	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	29	0.05	-0.90	-0.27	-0.74	-0.82	-0.33	0.17	0.34	0.45	-0.54	-1.45	-0.42	-1.77	-0.29
	30	0.14	-0.75	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	31	0.14	-0.75	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	32	0.13	-0.72	-0.20	-0.81	-0.89	-0.40	0.19	0.28	0.42	-0.56	-1.12	-0.76	-2.14	-0.09
	33	0.15	-0.78	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.79	-0.30
	34	0.14	-0.76	-0.23	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.58	-1.48	-0.43	-1.80	-0.31
	35	0.15	-0.80	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	36	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	37	0.14	-0.75	-0.25	-0.77	-0.85	-0.36	0.18	0.29	0.44	-0.56	-1.46	-0.42	-1.78	-0.30
	38	0.15	-0.77	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.79	-0.30
	39	0.15	-0.74	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.79	-0.30
	40	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	41	0.15	-0.77	-0.26	-0.76	-0.84	-0.35	0.18	0.38	0.45	-0.64	-1.55	-0.51	-1.86	-0.38
	42	0.15	-0.77	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	43	0.15	-0.78	-0.40	-0.76	-0.84	-0.35	0.18	0.30	0.44	-0.45	-1.44	-0.47	-1.75	-0.26
	44	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	45	0.15	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	46	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-030
	47	0.14	-0.76	-0.27	-0.77	-0.85	-0.36	0.18	0.30	0.44	-0.55	-1.47	-0.44	-1.78	-0.30
	48	0.15	-0.77	-0.24	-0.76	-0.84	-0.35	0.18	0.31	0.45	-0.56	-1.49	-0.42	-1.78	-0.38
	49	0.13	-0.74	-0.32	-0.80	-0.88	-0.39	0.19	0.28	0.43	-0.54	-1.39	-0.50	-1.63	-0.30

Table 26 – Leave-one-out results for all compounds, red text indicates values not falling within ±20% of the fragment and factor values for all compounds

Fragment and factor values		Aromatic fragment										$_{ar}$-O-$_{al}$
		Cl	CN	Br	I	CH_3	CH_2	C=O	C(=O)O	CH	N	
All compounds		0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.33	-0.20	0.66	-0.59
Compound removed	1	0.51	-0.45	0.60	1.00	0.27	0.48	-0.895	-0.33	-0.21	0.66	-0.60
	2	0.50	-0.44	0.62	1.00	0.27	0.50	-0.905	-0.32	-0.21	0.73	-0.58
	3	0.50	-0.45	0.60	1.00	0.27	0.45	-0.887	-0.33	-0.15	0.62	-0.58
	4	0.50	-0.44	0.61	1.05	0.25	0.48	-0.942	-0.33	-0.19	0.67	-0.59
	5	0.76	-0.41	0.64	1.04	0.30	0.49	-0.875	-0.30	-0.14	0.73	-0.55
	6	0.49	-0.44	0.61	0.69	0.29	0.48	-0.874	-0.32	-0.18	0.68	-0.58
	7	0.52	-0.46	0.60	1.04	0.27	0.47	-0.888	-0.34	-0.22	0.65	-0.60
	8	0.51	-0.46	0.59	0.93	0.27	0.47	-0.886	-0.34	-0.22	0.65	-0.60
	9	0.57	-0.33	0.72	1.05	0.37	0.60	-0.808	-0.21	-0.09	1.02	-0.48
	10	0.50	-0.43	0.62	0.81	0.27	0.48	-0.905	-0.31	-0.17	0.70	-0.57
	11	0.51	-0.45	0.60	1.01	0.27	0.47	-0.890	-0.33	-0.20	0.66	-0.59
	12	0.49	-0.45	0.60	0.96	0.27	0.45	-0.878	-0.34	-0.17	0.61	-0.59
	13	0.48	-0.45	0.60	0.97	0.28	0.48	-0.866	-0.33	-0.20	0.66	-0.59
	14	0.49	-0.44	0.61	1.12	0.27	0.48	-0.903	-0.33	-0.19	0.67	-0.59
	15	0.51	-0.45	0.61	1.01	0.27	0.49	-0.911	-0.35	-0.29	0.70	-0.62
	16	0.50	-0.45	0.60	1.00	0.27	0.45	-0.887	-0.33	-0.15	0.62	-0.58
	17	0.51	-0.45	0.60	1.00	0.27	0.48	-0.892	-0.33	-0.20	0.66	-0.59
	18	0.53	-0.45	0.60	1.02	0.28	0.55	-0.908	-0.34	-0.38	0.80	-0.63
	19	0.51	-0.45	0.60	1.00	0.24	0.47	-0.888	-0.33	-0.21	0.66	-0.60
	20	0.53	-0.45	0.60	1.02	0.28	0.55	-0.908	-0.34	-0.38	0.80	-0.63
	21	0.51	-0.45	0.60	1.00	0.27	0.47	-0.570	-0.33	-0.20	0.66	-0.60
	22	0.54	-0.51	0.55	1.10	0.27	0.45	-0.850	-0.38	-0.31	0.56	-0.67
	23	0.52	-0.47	0.58	1.01	0.26	0.44	-0.882	-0.35	-0.20	0.56	-0.61
	24	0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.33	-0.20	0.66	-0.59
	25	0.51	-0.45	0.60	1.00	0.27	0.48	-0.891	-0.33	-0.20	0.66	-0.59
	26	0.51	-0.41	0.64	1.02	0.30	0.50	-0.885	-0.30	-0.15	0.75	-0.55
	27	0.54	-0.42	0.63	1.04	0.30	0.51	-0.862	-0.30	-0.18	0.75	-0.57

Fragment and factor values		Aromatic fragment										$_{ar}$-O-$_{al}$
		Cl	CN	Br	I	CH_3	CH_2	C=O	C(=O)O	CH	N	
All compounds		0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.33	-0.20	0.66	-0.59
Compound removed	28	0.51	-0.45	0.60	1.00	0.27	0.47	-1.015	-0.33	-0.20	0.66	-0.60
	29	0.53	-0.45	0.60	1.02	0.28	0.56	-0.760	-0.23	0.03	0.83	-0.49
	30	0.51	-0.45	0.60	1.01	0.27	0.48	-0.892	-0.33	-0.21	0.67	-0.60
	31	0.51	-0.45	0.60	1.01	0.27	0.48	-0.892	-0.33	-0.21	0.67	-0.60
	32	0.52	-0.48	0.57	1.32	0.27	0.46	-0.869	-0.36	-0.26	0.61	-0.63
	33	0.51	-0.45	0.60	1.00	0.27	0.46	-0.896	-0.33	-0.20	0.64	-0.60
	34	0.48	-0.45	0.60	0.98	0.28	0.48	-0.870	-0.33	-0.20	0.66	-0.60
	35	0.51	-0.45	0.60	1.00	0.27	0.47	-0.881	-0.32	-0.16	0.65	-0.58
	36	0.51	-0.45	0.60	1.00	0.27	0.47	-0.893	-0.33	-0.21	0.66	-0.60
	37	0.53	-0.44	0.61	1.00	0.29	0.49	-0.861	-0.32	-0.20	0.69	-0.59
	38	0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.29	-0.20	0.66	-0.59
	39	0.51	-0.45	0.60	1.00	0.27	0.47	-0.874	-0.31	-0.14	0.65	-0.58
	40	0.51	-0.45	0.60	1.00	0.27	0.47	-1.090	-0.33	-0.20	0.66	-0.59
	41	0.43	-0.52	0.53	0.92	0.20	0.40	-0.960	-0.40	-0.26	0.45	-0.66
	42	0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.37	-0.19	0.66	-0.59
	43	0.49	-0.44	0.61	1.09	0.22	0.48	-0.996	-0.32	-0.17	0.68	-0.58
	44	0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.33	-0.20	0.66	-0.59
	45	0.51	-0.45	0.60	1.00	0.27	0.47	-0.891	-0.33	-0.19	0.66	-0.59
	46	0.51	-0.45	0.60	1.00	0.30	0.47	-0.888	-0.33	-0.21	0.66	-0.60
	47	0.51	-0.45	0.60	1.00	0.27	0.47	-0.892	-0.33	-0.20	0.66	-0.59
	48	0.50	-0.44	0.61	0.97	0.28	0.48	-0.890	-0.32	-0.19	0.68	-0.59
	49	0.52	-0.47	0.58	0.92	0.26	0.47	-0.892	-0.35	-0.24	0.63	-0.62

Table 27 – Leave-one-out results for all compounds, red text indicates values not falling within ±20% of the fragment and factor values for all compounds

Fragment and factor values		Aliphatic										
		CH_3	CH_2	$CH_{2\ unsat}$	CH_{unsat}	OH	C=O	S=O	NH_2	CN	COO^-	$_{ali}O_{ali}$
All compounds		0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
Compound removed	1	0.32	0.35	-0.02	0.12	-2.14	-1.74	-2.24	-2.05	-1.36	-2.10	-1.95
	2	0.33	0.36	0.00	0.12	-2.13	-1.75	-2.25	-2.03	-1.370	-2.08	-1.91
	3	0.28	0.35	-0.04	0.11	-2.14	-1.71	-2.19	-2.08	-1.32	-1.87	-2.00
	4	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.95
	5	0.29	0.34	-0.02	0.11	-2.13	-1.71	-2.20	-2.06	-1.33	-2.11	-1.97
	6	0.31	0.35	-0.01	0.11	-2.13	-1.72	-2.22	-2.05	-1.35	-2.10	-1.95
	7	0.32	0.35	-0.01	0.12	-2.13	-1.74	-2.23	-2.03	-1.36	-2.09	-1.93
	8	0.32	0.35	-0.01	0.12	-2.13	-1.74	-2.23	-2.03	-1.36	-2.09	-1.93
	9	0.32	0.35	-0.01	0.12	-2.13	-1.74	-2.23	-2.03	-1.36	-2.09	-1.93
	10	0.30	0.34	-0.02	0.11	-2.13	-1.72	-2.21	-2.05	-1.34	-2.10	-1.96
	11	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.03	-1.35	-2.09	-1.94
	12	0.29	0.34	-0.03	0.11	-2.13	-1.71	-2.20	-2.06	-1.33	-2.11	-1.98
	13	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	14	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.94
	15	0.36	0.36	-0.02	0.12	-2.16	-1.79	-2.29	-2.05	-1.40	-2.11	-1.96
	16	0.28	0.35	-0.04	0.11	-2.13	-1.71	-2.19	-2.08	-1.32	-2.38	-2.00
	17	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	18	0.42	0.36	0.10	0.13	-1.90	-1.81	-2.37	-1.91	-1.46	-1.99	-1.72
	19	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.93
	20	0.42	0.36	0.10	0.13	-2.34	-1.81	-2.37	-1.91	-1.46	-1.99	-1.72
	21	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	22	0.35	0.36	0.02	0.13	-2.13	-1.77	-2.28	-2.01	-1.39	-2.07	-1.87
	23	0.29	0.34	-0.02	0.11	-2.13	-1.71	-2.21	-2.05	-1.33	-2.10	-1.97
	24	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	25	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	26	0.30	0.34	-0.02	0.11	-2.13	-1.72	-2.21	-2.05	-1.34	-2.10	-1.96
	27	0.32	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.36	-2.09	-1.93

Fragment and factor values		Aliphatic										
		CH_3	CH_2	$CH_{2\ unsat}$	CH_{unsat}	OH	C=O	S=O	NH_2	CN	COO^-	$_{ali}O_{ali}$
All compounds		0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
Compound removed	28	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	29	0.30	0.45	-0.02	0.22	-2.13	-1.61	-2.10	-2.05	-1.34	-2.10	-1.86
	30	0.32	0.35	0.04	0.12	-2.13	-1.74	-2.24	-2.03	-1.36	-2.09	-1.93
	31	0.32	0.35	-0.05	0.12	-2.13	-1.74	-2.24	-2.03	-1.36	-2.09	-1.93
	32	0.34	0.36	0.01	0.12	-2.13	-1.75	-2.25	-2.02	-1.37	-2.08	-1.90
	33	0.30	0.34	-0.02	0.11	-2.13	-1.70	-2.22	-2.05	-1.34	-2.10	-1.96
	34	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	35	0.30	0.35	-0.02	0.12	-2.13	-1.75	-2.20	-2.06	-1.34	-2.11	-1.96
	36	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.23	-2.04	-1.35	-2.10	-1.95
	37	0.32	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.36	-2.09	-1.93
	38	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.95
	39	0.29	0.34	0.02	0.11	-2.10	-1.68	-2.19	-2.00	-1.33	-2.05	-1.87
	40	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	41	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.95
	42	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.95
	43	0.30	0.35	-0.02	0.11	-2.13	-1.72	-2.21	-2.05	-1.34	-2.10	-1.96
	44	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	45	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.94
	46	0.31	0.35	-0.01	0.12	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.93
	47	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.23	-2.04	-1.35	-2.09	-1.94
	48	0.31	0.35	-0.01	0.11	-2.13	-1.73	-2.22	-2.04	-1.35	-2.10	-1.95
	49	0.33	0.35	0.00	0.12	-2.13	-1.74	-2.24	-2.03	-1.37	-2.09	-1.92

Table 28 – Leave-one-out results for all compounds, red text indicates values not falling within ±20% of the fragment and factor values for all compounds

Leave-one-out statistical analysis of predicted and experimental log $k_{IAM\ (pH\ 7.4)}$ values gives a PRESS = 14.2 and $Q^2 = 0.93$. The PRESS value is relatively low and the Q^2 value is high. This indicates that the fragments and factors determined for aliphatic and aromatic compounds are a good predictor of log $k_{IAM\ (pH\ 7.4)}$.

The K-fold cross validation approach was also used to investigate the fragment and factor values determined for all compounds. The compounds that do not contain unique fragments (Identified in Table 25) were assigned an identifier. These compounds were split into four groups of equal size (compound numbers 1-12, 13-24, 25-36 and 37-49). Each group was removed from the training set independently and the MLR analysis repeated. The coefficients for the fragments and factors determined in 4-fold cross validation are reported in Table 29.

Table 29 shows the fragment and factor values from both the initial analysis and each of the 4-fold cross validation results. It is clear that all cross validation models exclude a fragment or factor value from the analysis (relative to the model derived from the full dataset), this in effect reduces the domain of the models. In addition the removal of a fragment or factor value from each cross validation model will change the remaining values due to the high interdependency of the fragment and factor values.

Due to the relatively low number of compounds analysed (n=64), the low compound/descriptor ratio (64/36 = 1.78) and the high interdependency of fragments and factors, there is variation observed during both leave-one-out cross-validation, and the more stringent 4-fold cross-validation.

To meet the TGD guideline of a compound/descriptor ratio greater than 2, the analysis of more compounds is required. However, these need to be selected so as not to increase the number of fragment and factors required to fully describe the hydrophobicity of the compounds. Despite this drawback, the correlation between predicted log $k_{IAM\ (pH\ 7.4)}$ and experimental log $k_{IAM\ (pH\ 7.4)}$ (Figure 29) shows the concept of applying fragment and factor values to predict log $k_{IAM\ (pH\ 7.4)}$ is valid.

Fragment and factor values		All	1 - 12	13 - 24	25 - 36	37 - 49
Factor	bond	0.14	0.21	0.02	-0.01	0.16
	ring	-0.76	-0.95	-0.37	-2.14	-0.76
	P.E. 2	-0.27	-0.43			-0.44
Hetroaromatic in ring substituent	O	-0.77	-0.58	-0.92	-0.74	-0.78
	NH	-0.85	-0.66	-1.00	-0.82	-0.86
	S	-0.36	-0.17	-0.51	-0.33	-0.37
Aromatic fragment	CH_{ar}	0.18	0.13	0.22	0.17	0.18
	C_{ar}	0.30		0.23	0.30	0.36
	$C_{ar\ con}$	0.44	1.36	0.37	0.46	0.43
	OH	-0.55	-0.31	-0.58	-0.49	-0.51
	O^-	-1.47	-1.25	-1.82	-1.04	-1.45
	NH_2	-0.44	0.07	0.03	-0.67	-0.60
	COO^-	-1.78	-0.95	-1.98	-2.07	-1.66
	NO_2	-0.30	0.34	-0.42	-0.01	-0.39
	Cl	0.51	1.34	0.55	0.58	0.42
	CN	-0.45	0.09	-0.57	-0.40	-0.52
	Br	0.60	1.14	0.48	0.65	0.53
	I	1.00	1.04	1.12	1.38	0.90
	CH_3	0.27	0.78	0.12	0.36	0.11
	CH_2	0.47	0.94	0.48	0.67	0.39
	C=O	-0.89	-0.40	-0.57	-0.70	-1.26
	$C(=O)O^-$	-0.33	0.20	-0.49	-0.11	
	CH	-0.20	0.45	-0.69	0.26	-0.19
	N	0.66	2.13	0.55	1.10	0.42
$_{ar}$-O-$_{al}$		-0.59	-0.04	-0.80	-0.36	-0.64
Aliphatic fragment	CH_3	0.31	0.23	0.52	0.28	0.28
	CH_2	0.35	0.33	0.39	0.52	0.34
	$CH_{2\ unsat}$	-0.01	-0.09	0.11	0.28	0.01
	CH_{unsat}	0.12	0.10	0.13	-1.01	0.11
	OH	-2.13	-2.14	-1.93	-1.54	-2.10
	C=O	-1.73	-1.66	-2.51	-2.01	-1.67
	S=O	-2.23	-2.13	-1.91	-2.08	-2.17
	NH_2	-2.04	-2.14	-1.56	-1.32	-2.01
	CN	-1.35	-1.27	-2.21	-2.13	-1.32
	COO^-	-2.09	-1.90	-1.73	-1.83	-2.06
	$_{ali}$-O-$_{ali}$	-1.94	-2.11	-0.08	-0.17	-1.89

Table 29 - Fragment and factor values determined initially and determined using four-fold cross validation

4.4.4 External validation of fragment and factor values for both aliphatic and aromatic compounds

Three aliphatic and three aromatic compounds were chosen to form the external test set. These compounds were chosen on the basis of being within the domain of the fragment and factor values already determined and being commercially available. The test set compounds selected are detailed in Table 30, along with compounds that have been previously analysed and are similar in terms of their theoretical fragment and factors that contribute to determine their hydrophobicity. The predicted log $k_{IAM\ (pH\ 7.4)}$ values were determined using the fragment and factor values determined here and reported in Table 24. The counts of fragment and factors were calculated manually as detailed in Section 4.3.1.2 and reported in Appendix 1.3 Table 25.

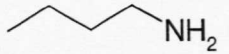
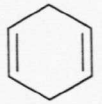
Four of the test set compounds were chosen with the expectation that they will be well predicted. Two compounds were chosen for being more difficult to predict. These were 2-nitrophenol (due to the compound being only partially ionised at pH 7.4) and butylamine (due to the compound being 99% ionised under the conditions of analysis. No fragment has been determined for NH_3^+; however, in Chapter 3 equations (4.3) and (4.4) were determined for unionised and ionised compounds respectively, therefore, an ionisation factor of -1.06 has been used to calculate the effect of ionisation). Predictions were made for both butylamine and 2-nitrophenol for both the ionised and unionised forms for comparison, additionally for 2-nitrophenol a weighted average approach[20] was also attempted to calculated log $k_{IAM\ (pH7.4)}$ for a partially ionised compound.

$$\log k_{IAM\ (pH\ 7.4)(unionised)} = 0.91 \log P - 0.68 \qquad (4.3)$$

$n = 30,\ r^2_{(adj)} = 0.96$

$$\log k_{IAM\ (pH\ 7.4)(ionised)} = 0.70 \log P - 1.74 \qquad (4.4)$$

$n = 6,\ r^2_{(adj)} = 0.64$

Compound	Similar compounds with log $k_{IAM\ (pH\ 7.4)}$ values determined in this study	Structure	Experimental log P[12]	Predicted log $k_{IAM\ (pH\ 7.4)}$	Experimental log $k_{IAM\ (pH\ 7.4)}$	pKa	% ionised at pH 7.4
1-Butanol	Methanol, pentanol		0.88	-0.35	-0.35	16.10	Unionised
1-Butylamine	Pentylamine		0.97	-1.32 ionised -0.26 unionised -1.06 ionised (equation (4.4))	-1.44 Sec-Butylamine	10.78	99.9
1,4-Cyclohexadiene	Cyclopentene, cyclohexene, pentene, hex-1-ene, hex-2-ene, cyclopentane, cyclohexane, pentane, hexane, cyclopentadiene		2.3	0.60	1.46		

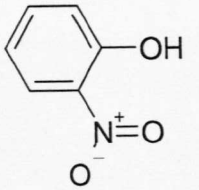
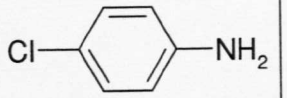
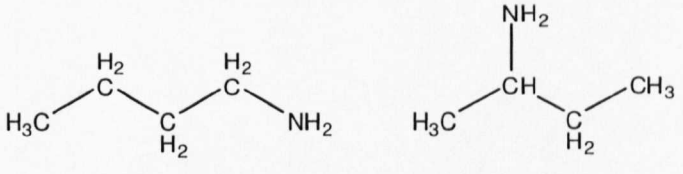
Compound	Similar compounds with log $k_{IAM\ (pH\ 7.4)}$ values determined in this study	Structure	Experimental log P[12]	Predicted log $k_{IAM\ (pH\ 7.4)}$	Experimental log $k_{IAM\ (pH\ 7.4)}$	pKa	% ionised at pH 7.4
2-Nitrophenol	Nitrobenzene, phenol		1.79	-0.64 ionised 0.28 unionised -0.27 weighted average	0.43	7.23	60.0
4-Chloroaniline	Chlorobenzene, aniline		1.83	0.93	1.39	3.98	0.04
1,4-Dichlorobenzene	Chlorobenzene		3.44	1.88	2.38		

Table 30 - External validation compounds analysed, along with similar compounds analysed and the predicted log $k_{IAM\ (pH\ 7.4)}$ value for each validation compound

The experimental log $k_{IAM\ (pH\ 7.4)}$ values for the external test set were obtained by Miss Siiri Latvala at a Unilever research laboratory using one of the IAM-HPLC columns used in the robustness testing of the method. The log $k_{IAM\ (pH\ 7.4)}$ results were obtained following the experimental method detailed in Section 3.3.4.11 and the results are detailed in Table 30.

It should be noted that for the external test set, the prediction was for n-butylamine, which is within the domain of the training set. However, the compound analysed was sec-butylamine. The difference in structure and therefore fragments is shown in Figure 30. The training set does not contain a fragment for $CH_{aliphatic}$ so sec-butylamine is outside the training set and a prediction should not be made.

Although, the difference in structure between n-butylamine and sec-butylamine is small, that two CH_2 groups in n-butylamine are a CH and CH_3 group in sec-butylamine. The change in geometric shape and therefore molecular volume has a significant impact upon hydrophobicity. Therefore the prediction for sec-butylamine has been recalculated based on the log P value from KOWWIN and equation (4.4), determined in chapter 3 for ionised compounds. The prediction of log $k_{IAM\ (pH\ 7.4)}$ is therefore -1.06 log $k_{IAM\ (pH7.4)}$ (calculated using a log P value of 0.97).



n-Butylamine sec-Butylamine

n-Butlyamine = 1 (CH_3) + 3 (CH_2) + 1 (NH_2) + 3 (bond factors)

sec-Butylamine = 2 (CH_3) + 1 (CH_2) + 1 (CH)+ 1 (NH_2) + 3 (bond factors)

Figure 30 - Structure and fragment and factor counts for n-butylamine and sec-butylamine

Figure 31 illustrates the relationship between experimental and predicted log $k_{IAM\ (pH\ 7.4)}$ values for the external test set. The log $k_{IAM\ (pH\ 7.4)}$ values for the five compounds, for which predictions were based on the fragment and factor values are well predicted. Highlighted on Figure 31 is the line of unity expected for well predicted/measured values. Circled is the ionised value for sec-butylamine

(calculated based on the log P value and eqation (4.4) for ionised compounds), which is inline with all other predictions, indicating that both methods of predicting log $k_{IAM\ (pH\ 7.4)}$ are reasonable approaches. Additionally circled, is the weighted average value for 2-nitrophenol, which is partially ionised at pH 7.4. This indicates that applying weighted averages of the unionised and ionsed log k_{IAM} values based on the percentage ionised may provide a good indication of the log $k_{IAM\ (pH\ 7.4)}$ value for partially ionised species. However, this would need to be investigated further for a range of compounds to validate this proposition.

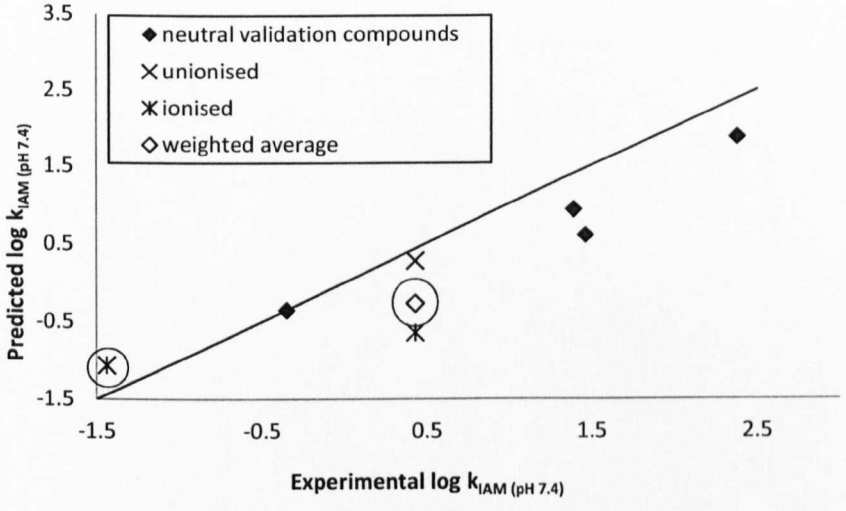


Figure 31 - Plot of experimental log $k_{IAM\ (pH\ 7.4)}$ against predicted log $k_{IAM\ (pH\ 7.4)}$ for the external validation test compounds

4.5 Conclusions

Breaking compounds into theoretical structural fragments and factors is a well accepted method for the prediction of log P. It has been shown that the technique is applicable to log $k_{IAM\ (pH\ 7.4)}$ values as well. For predictions of log $k_{IAM\ (pH\ 7.4)}$ to be made based on structural fragment and features there is a requirement for the theoretical fragment and features to have been considered by the training set. The application of fragments to predict log $k_{IAM\ (pH\ 7.4)}$ requires that the environment of the fragment is considered i.e. aliphatic and aromatic environments and ionised/unionised groups be considered as distinct groups and, therefore, distinct fragments. It should be noted that the method was trained using ionisable compounds that were either the fully ionised or unionised species at pH 7.4. The effect of partially ionised compounds has not been investigated here, except for 2-nitrophenol

as an external validation compound. It is reasonable to expect predictions made applying weighted averages to the relevant fragment values would provide an accurate prediction of log $k_{IAM\ (pH\ 7.4)}$. However, this would need to be investigated prior to predictions based on weighted averages being applied.

The fragement and factor values determined using the training set were both interally and externally validated. The fragment and factor values determined here have good predictive capabilities (Q^2 =0.95) determined using the leave-one-out analysis. In addition the predicted log $k_{IAM\ (pH\ 7.4)}$ values for the external validation compounds compared to the experimental log $k_{IAM\ (pH\ 7.4)}$ have an r^2 of 0.95 (based on the predictions corrected for ionisation for the external validation compounds).

The dataset to determine fragment and factor values is currently small, with a low compound/descriptor ratio (1.78). It has also been identified that CH_3 should be more hydrophobic than the CH_2 and additional bonds should reduce the hydrophobicty of a compound. However, for the limited dataset considered here this was found not to be the case, this could be due to the relatively low compound/descriptor ratio and the high inter-correlation between fragments and factors (i.e. number of CH_2 fragments and number of bonds are highly correlated). To improve the fragment and factor values determined, the compound/descriptor value needs to be increased (increasing confidence in the fragment and factor values determined), along with the structural diversity of the compounds considered (extending the domain of applicability). Initially as the number of compounds increases, and/or new fragments and factors are introduced, the fragment and factor values will change, however, as this process is repeated, the variability in the fragment and factor values determined will be reduced.

There is an observable difference between the partitioning of a compound in the octanol-water system and partitioning in the IAM-HPLC system. This is illustrated by the change in sign for some π substituents (Figure 27). The fragments that show a change in relative hydrophobicity between log P and log $k_{IAM\ (pH\ 7.4)}$ are the nitro group, the methoxy group and the acetyl group. These all are described as more hydrophobic by the IAM system than the octanol-water system, these are groups that are able to form hydrogen bond donor/acceptor bonds. It is these interactions that the

IAM partitioning system describes that are not described by the octanol-water system.

Despite the use of fragment and factors showing good predictive abilities in relation to log $k_{IAM\ (pH\ 7.4)}$ there is still a requirement for new compounds to be within the applicability domain of the model, which is currently relatively small. Therefore, methods to determine log $k_{IAM\ (pH\ 7.4)}$ using a traditional QSAR descriptor approach will be considered in Chapter 5.

4.6 References

[1] Dearden J.C. (1985) Partitioning and Lipophilicity in Quantitative Structure-Activity Relationships *Environ. Health Perspect.* 61: 203-228.

[2] European Union. (1995) *Overview of Structure-Activity Relationships for Environmental Endpoints Part 1: General Outline and Procedure, Report of the EU-DG-XII Project "QSAR for Predicting Fate and Effects of Chemicals in the* Environment (Contract ~ EV5V-CT92-0211).

[3] Cohen E., Edsal J. (1943) *Proteins, Amino Acids and Peptides,* New York, Reinhold, pp 200.

[4] Fujita T., Iwasa J., Hansch C. (1964) A New Substituent Constant, π, Derived from Partition Coefficients *J. Amer. Chem. Soc.* 86: 5175-5180.

[5] Hansch C., Leo A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology,* New York, Wiley-Interscience Publication.

[6] Rekker R.F. (1977) *The Hydrophobic Fragmental Constant,* New York, Elsevier.

[7] Davis S. S. (1973) Determination of the Thermodynamics of the Methyl Group in Solutions of Drug Molecules *J. Pharm. Pharmac.* 25: 1-12.

[8] Davis S. S. (1973) Use of Substituent Constants in Structure-Activity Relations and the Importance of the Choice of Standard State *J. Pharm. Pharmac.* 25: 293-296.

[9] Leo A., Jow P.Y.C, Silipo C., Hansch C. (1975) Calculation of Hydrophobic Constant (Log P) from π and f Constants *J. Med. Chem.* 18: 865-868.

[10] Mannhold R., Rekker R.F., Dross K., Bijloo G., de Vries G. (1998) The Lipophilic Behaviour of Organic Compounds: 1. An Updating of the Hydrophobic Fragmental Constant Approach *Quant. Struct.-Act. Relat.* 17: 517-536.

[11] Meylan W.M., Howard P.H. (1995) Atom/Fragment Contribution Method for Estimation Octanol-Water Partitioning Coefficients *J. Pharm. Sci.* 84: 83-91.

[12] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.

[13] Pidgeon C., Venkatarum U.V. (1989) Immobilized Artificial Membrane Chromatography: Supports Composed of Membrane Lipids *Anal. Biochem.* 76: 36-47.

[14] SPSS Inc. software products (2006) SPSS® version 15.0.1, Chicago, SPSS Inc. software products

[15] Leach A.R., Gillet V.J. (2003) *An Introduction to Cheminformatics,* Dordrecht, Kluwer Academic Publishers, pp 81-83.

[16] Minitab Inc. (2007) Minitab® Statistical Software version 15, Coventry: Minitab Inc..
[17] Ebbing D.D., Gammon S.D. (2009) *General Chemistry*, 9th edition, Belmont, Houghton Mifflin, pp 681.
[18] Leahy D.E., Morris J.J., Taylor P.J., Wait A.R. (1992) Model Solvent Systems for QSAR 2. Fragment Values (F-Values) for the Critical Quartet *J. Chem. Soc. Perk. Trans. 2.* 4: 723-731.
[19] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-Octanol/Water), High Performance Liquid Chromatography (HPLC) Method*, Paris, The Organisation for Economic Cooperation and Development OECD.
[20] Roberts D.W. (1991) QSAR Issues in Aquatic Toxicity of Surfactants *Sci. Total Environ.* 109/110: 557 – 568.

5 QSAR - Predicting log k_{IAM} from physico-chemical properties and structural descriptors

5.1 Introduction

Descriptors for hydrophobicity and log P in particular are commonly used in QSARs to model and estimate the biological activity of chemicals from their physico-chemical properties[1]. There have been many QSARs published in the literature that utilise descriptors, including log P, that have to be determined experimentally[2-7]. Use of these QSARs becomes restricted unless the descriptor can be readily determined (includes experimental determinations and calculations) for new compounds. This is clearly a limiting factor. However, such QSARs can be applied to a larger chemical domain if a reliable method of calculating the descriptor is available.

Log k_{IAM} is considered to be an alternative measure of hydrophobicity to log P. The optimised IAM HPLC assay that was assessed for robustness in Chapter 3 allows for the experimental determination of log $k_{IAM\ (pH\ 7.4)}$ values for compounds of interest. The optimised assay has been found to be suitable to assess a range of compounds from hydrophilic to highly hydrophobic. It was shown in Section 3.4.2.3 that for ionised compounds log P and log $k_{IAM\ (pH\ 7.4)}$ are different due to differences between a compounds partitioning between the IAM stationary phase-mobile phase and partitioning between octanol-water.

Use of QSARs based on log k_{IAM} will be extended by the availability of a reliable method to predict log k_{IAM}. However, there are currently no standardised and widely applicable methods to predict log k_{IAM} from chemical structure.

As discussed in Chapter 4, log $k_{IAM\ (pH\ 7.4)}$ can be predicted from structural fragments and correction factors[8]. The sum of these fragment and factors provides a good indication of log $k_{IAM\ (pH\ 7.4)}$. The fragment and factor values determined in Chapter 4 showed good predictive capabilities using the leave-one-out cross-validation method. In addition, the external test set chemicals not seen *a priori* by the algorithm were well predicted. Predicting log $k_{IAM\ (pH\ 7.4)}$ using the sum of the fragments and correction factors present in a compound requires that all fragments and factors within that compound are represented in the training set. If any fragment or factor is

not contained within the training set, the compound would be considered outside the domain of the method and a prediction should not be made. This was emphasised whilst making predictions of log k_{IAM} for the external test set chemicals. A prediction was made for n-butylamine (inside the domain); however, analysis was performed on sec-butylamine (outside the domain). The domain is small (currently), due to the limited training set used to develop the model.

An alternative to the fragment and correction factor method to predict log $k_{IAM\ (pH\ 7.4)}$ is the use of structural or non-empirical descriptors that are calculated based on the chemical's structure (e.g. atom counts, ring counts, hydrogen bond acceptors and donor counts, log S, log P etc.). This, to some extent, removes the necessity for specific fragments to be present in the training set, but is still be limited in terms of the applicability domain. There are numerous examples of this type of Quantitative Structure-Property Relationship (QSPR)/QSAR developed e.g. for log P[9, 10] and melting point[2] etc.. Therefore, there is a strong and rational basis for this approach to predict physico-chemical properties. It is acknowledged that log k_{IAM} is a property of a compound and, therefore, the term QSPR could be used to describe the relationships generated. However, the term QSAR will be used in its wider context throughout the thesis.

The literature was searched for QSARs currently available to predict log k_{IAM} and log $k_{IAM\ (pH\ 7.4)}$. Considering the sources of data used to compile the log k_{IAM} database, 29 papers[4, 11-38] reported a total of 97 QSARs that predict log k_{IAM} within certain classes of chemistry. In addition to publications supporting the log k_{IAM} database, another two QSARs were published in a further two papers[5, 39].

The individual QSARs are available in Appendix 1.4 equations (A.1.1) to (A.1.97). It is acknowledged that although published in peer-reviewed literature, not all the QSARs investigated are valid. QSARs are statistically invalid if the compound/descriptor ratio is less than 5 and if the F statistic is less than the relevant F_α value for the relationship determined. Additionally, QSARs with low $r^2_{(adj)}$ values will have poor predictive capabilities[40]. Not all QSARs report sufficient statistics to determine their validity. Therefore, when investigating the number and types of

descriptors in these QSARs, the validity or otherwise of the QSARs was not considered.

The QSARs were analysed in terms of the number and types of descriptors contained within them. The number of descriptors in each of the 97 QSARs is summarised in Table 31. This indicates that the majority of QSARs have only a single descriptor, whilst others require up to seven descriptors to model log k_{IAM}.

Number of descriptors	Number of QSARs	Range of r^2_{adj}
1	56	0.35 – 0.99
2	6	0.89 – 0.97
3	5	0.79 – 0.97
4	5	0.89 - 0.94
5	22	0.73 – 0.98
6	2	0.91 – 0.96
7	1	0.95

Table 31 - Number of descriptors per model for calculating log k_{IAM}

The types of descriptors used in the QSARs considered are detailed in Table 32, along with the number of occurrences of each descriptor. Of the QSARs to predict log k_{IAM} published to date, the majority use only one descriptor to predict log k_{IAM} and predominantly this is a measure of partitioning. Alternatively, other approaches have used the Abraham solvatochromatic descriptors[12, 20, 25, 26, 32, 35, 39]. These descriptors are well characterised in the form of linear free energy relationships (LFERs) which are commonly used for modelling partitioning. Whilst there are a large number of QSARs (refer to Appendix 1.4 equations (A.1.1) to (A.1.97)) to predict log k_{IAM}, all are for specific and limited chemical classes.

Descriptor		Number of occurrences	Classes of compounds QSARs apply to
log P	Partitioning	55	Structurally diverse neutral and ionised compounds including drugs, also specific chemical classes with small datasets i.e. β-blockers
log D		11	
log P_{SOL} aqueous/multilamellar		1	Local anaesthetics
Log P_{PS} aqueous/multilamellar		1	
Excess molar refraction	Solvatochromatic parameters	24	Structurally diverse drugs and Structurally diverse neutral compounds
Polarisability		28	
Effective hydrogen bond acidity		28	
Effective hydrogen bond basicity		28	
Characteristic volume		28	
Number of rotational bonds		1	Structurally diverse drugs
Number of rings		1	
Molecular weight		1	
Total surface area		1	
Refraction index		4	Carboxylic drugs
Geometrical index		2	
Fraction of positively charged species		7	Structurally diverse neutral and basic drugs
Fraction of negatively charged species		2	
Carboxylic acids	Indicator parameter	6	Structurally diverse neutral and basic drugs
Amine		1	
Fluoroquinolones		2	

Table 32 - Descriptors used in the QSARs and the number of occurrences of each descriptor

Valkó *et al.*[41, 42] reported that for both Chromatographic Hydrophobicity Index $_{Acetonitrile}$ (CHI_{MeCN}) and the Chromatographic Hydrophobicity Index values referring to IAM chromatography (CHI_{IAM}). According to Valkó *et al.* there is a requirement for a hydrogen bond acidity term to account for the effect of ionised compounds[41, 42]. The CHI value was obtained from analysis using a rapid gradient HPLC method. The CHI values reported using this method include a second parameter in addition to the relative retention factor. The second factor is a slope value based on the percentage organic modifier in the mobile phase. Therefore, the CHI values are not from an aqueous solution but are based on an optimum organic phase concentration in the mobile phase[43]. This indicates that other factors, in addition to log P, may be relevant to modelling log k_{IAM}.

5.2 Aim of the Chapter

There are no global approaches to predict log k_{IAM} which are analogous to methods available for the prediction of log P e.g. KOWWIN[44], VCCLAB[9, 10] etc. Therefore, the aim of this investigation was to develop a predictive model for log k_{IAM} on the basis of "classical" QSAR descriptors using physico-chemical properties and structural descriptors. Here classical refers to descriptors that allow mechanistic interpretation of the models.

5.3 Method

5.3.1 Datasets of log k_{IAM} values

The log k_{IAM} values from the reduced dataset (averaged) from the database (n = 105) reported in Chapter 2 and the experimentally determined log k_{IAM} values (n = 66) reported in Chapter 3 and Chapter 4 were combined into a single dataset for QSAR analysis. The datasets were not combined for the determination of fragment and factor values, due to the number of descriptors being calculated, any variation in experimental procedure affecting log $k_{IAM\ (pH\ 7.4)}$ would have propagated through a large number of variable. Whilst variability in log $k_{IAM\ (pH\ 7.4)}$ could affect the QSARs developed, as the number of descriptors is fewer, the affect will be far less.

Two compounds appear in both the reduced dataset and the compounds analysed experimentally; these are detailed in Table 33. Both values for chlorobenzene were within ±0.1 log units specified in the OECD guidelines for acceptable repeatability[45].

The two values for naphthalene were further apart (0.33 log units). However, the two values are similar in magnitude and within reasonable experimental error[46], given that the experimental procedure is not fully described in the literature. Direct comparison of the two results is inappropriate. For all further analyses, the experimental log k_{IAM} values determined in this thesis were used in preference to the experimental values from the literature (included in the database), due to consistency within the experimental procedure.

Compound	Experimental log k_{IAM} value determined in thesis	Experimental log k_{IAM} value from the literature
Chlorobenzene	1.58	1.63
Naphthalene	2.48	2.15

Table 33 - Compounds with experimental log k_{IAM} value and a reduced dataset log k_{IAM} value

This resulted in a dataset containing log $k_{IAM\ (pH\ 7.4)}$ values for 169 compounds (dataset A) that were obtained under consistent and comparable experimental conditions (refer to Chapter 2 for full details). Additionally, the experimental log $k_{IAM\ (pH\ 7.4)}$ values determined in this thesis were used as a second smaller dataset (n=66) (dataset B). The datasets are available as electronic supplementary material (Chapter 5 – Predicting log k_{IAM} from descriptors (datasets).xls).

5.3.2 Generating descriptors

For each compound included in the dataset the 3D structure was generated from the Simplified Molecular Input Line Entry Specification/Systems strings (SMILES) using TSAR version 3.3[47] (available from http://accelrys.com/). Around 90 physico-chemical and structural descriptors were calculated for the dataset using TSAR version 3.3[47] and EPISuite version 4.1[44], these are detailed in Table 34. Descriptors that were non-suitable (i.e. all values being zero or identical) were removed. Some compounds contained missing descriptors, i.e. descriptors could not be calculated for some compounds.

Software	Calculated descriptor
TSAR version 3.3[47]	Molecular mass; molecular surface area; molecular volume; inertia moments; ellipsoidal volume; total dipole; dipole moments; lipophilicity; molecular refractivity; simple and valence-corrected molecular connectivity indices: zero order, 2nd to 6th order path, 3rd to 4th order cluster, 3rd to 6th order ring; shape flexibility; rotatable bonds (Σ rot); number of : halogen atoms (Σ Hal), H-bond donor centres (Σ_{HBD}), H-bond acceptor centres (Σ_{HBA}), heteroatoms, rings; total energy; electronic energy; surface area; mean polarisability; ionisation potential; energy of the lowest unoccupied molecular orbital (E_{LUMO}); energy of the highest occupied molecular orbital (E_{HOMO})
Episuite[44]	Logarithm of the octanol-water partition coefficient (log P), logarithm of the aqueous solubility (log S)

Table 34 - Physico-chemical and structural descriptors calculated and their source

5.3.3 Statistical analysis of the data

Stepwise regression analysis of the calculated descriptors was performed using Minitab[48] (version 15.1.1.0) to identify the significant descriptors relating to log $k_{IAM\ (pH\ 7.4)}$ values. Multiple-linear regression was used to develop QSARs using the significant descriptors identified in stepwise regression.

The following statistical information was recorded for the MLR analysis in the development of QSARs for the prediction of log $k_{IAM\ (pH\ 7.4)}$ from descriptors: n, $r^2_{(adj)}$, s and F values.

5.4 Results and discussion

This analysis aimed to develop QSARs based on "classical" physico-chemical properties and structural descriptors to predict log $k_{IAM\ (pH\ 7.4)}$. A database of 169 compounds (dataset A) and a dataset of 66 compounds (dataset B) were used to develop the QSARs. Dataset A covered aliphatic and aromatic structures which are predominantly polar organic compounds and includes drug molecules. The molecular weight in the dataset ranged from 32 to 645 g/mol, the log P (experimental values from KOWWIN[44]) values ranged from -1.56 to 7.57. Dataset B covered aliphatic and aromatic compounds that were predominantly polar organic molecules,

the molecular weight ranged from 32 to 631 g/mol, the log $P_{(exp)}$[44] ranged from -1.35 to 6.83. The datasets are available as electronic supplementary material (Chapter 5 – Predicting log k_{IAM} from descriptors (datasets).xls).

5.4.1 Comparison of database and experimental log k_{IAM} values vs log $P_{(exp)}$ for dataset A

There is a clear relationship between log $k_{IAM\ (pH\ 7.4)}$ and log P (Refer to Chapter 2 and Chapter 3, Figure 21). Therefore, to gain an understanding of the distribution of data within the dataset, log k_{IAM} was plotted against log P (Figure 32). Figure 32 indicates that the dataset covers a wide and representative range of log P. For neutral compounds there is a clear relationship between log P and log k_{IAM}. However, the relationship for potentially ionised compounds is more complex. As discussed in section 3.4.2.3 the observed log $k_{IAM\ (pH\ 7.4)}$ response for ionised compounds is lower than the log $k_{(IAM)\ pH\ 7.4)}$ value for unionised compounds with a similar log P value by approximately one log unit.

Figure 32 illustrates the relationship between log $k_{IAM\ (pH\ 7.4)}$ and log P for three groups of compounds: unionised experimental log $k_{IAM\ (pH\ 7.4)}$ values, equation (4.3), ionised experimental log $k_{IAM\ (pH\ 7.4)}$ values, equation (4.4) and database log $k_{IAM\ (pH\ 7.4)}$ values, equation (2.10).

Of the compounds highlighted in Figure 32 as being potential outliers from dataset A, four have been identified as being ionised under the conditions of determination for log k_{IAM} (compounds shaded in Table 35). These ionised compounds were separated into a subset of ionised database log k_{IAM} values (filled black diamonds Figure 32).

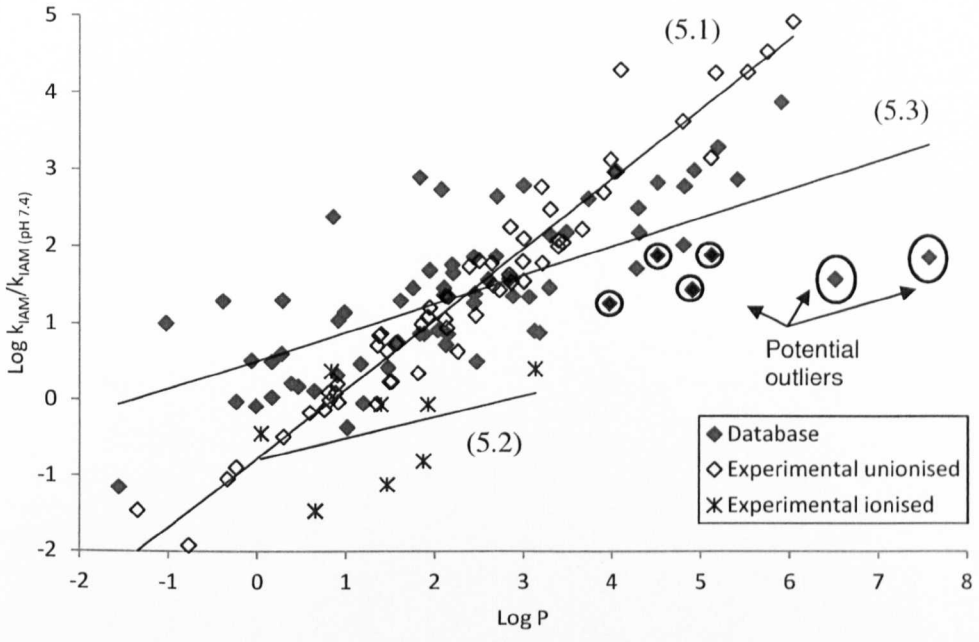


Figure 32 - Plot of experimental log $k_{IAM\ (pH\ 7.4)}$ (dataset A) plotted against log P
5.1 shows the trend between log P and log $k_{IAM\ (pH.7.4)}$ for the compounds unionised under the conditions of analysis, determined experimentally in Chapter 3 (included in dataset A and B) (equation (4.2))
5.2 shows the trend between log P and log $k_{IAM\ (pH.7.4)}$ for compounds fully ionised under the conditions of analysis, determined experimentally (included in dataset A and B) (equation (4.3))
5.3 shows the trend between log P and log $k_{IAM\ (pH.7.4)}$ for the compounds from the database (included in dataset A) (equation (2.12))
Circled compounds identify results from the database that may be potential outliers

The rectilinear trend for ionised compounds from the database is similar to that seen for compounds determined experimentally in this thesis. This is also true for unionised compounds (Figure 32). It is clear that two outliers remain (outliers circled that have a log P greater than 6, Figure 32). These points refer to 1,1-bis(*p*-chlorophenyl)-2,2-dichloroethylene and amiodarone; neither compound is ionisable. Despite there being no reason for their removal based on the chemistry or properties of these compounds, due to the high leverage both these values give to the relationship they have been removed as outliers from further QSAR analysis. It should be noted that neither the log $k_{IAM\ (pH\ 7.4)}$ values, nor the log P values reported in the literature for these two compounds have been verified in this study. As shown in Section 3.4.2, the log P values predicted for highly hydrophobic compounds were considerably higher than values determined experimentally in this study. In this case the reliability of both the log P value and the log k_{IAM} values reported are unknown and they have been removed from further analysis.

Compound	CAS no.	Reference	log k_{IAM}	log P[44]	pKa[44]
Amiodarone	1951-25-3	49	1.85	7.57	-
1,1-Bis(*p*-chlorophenyl)-2,2-dichloroethylene	72-55-9	49	1.57	6.51	-
Mefenamic acid	61-68-7	38	1.88	5.12	4.2
Desipramine	50-47-5	49	1.44	4.90	10.4
Diclofenac	15307-86-5	24	1.88	4.51	4.15
Ibuprofen	15687-27-1	24, 50	1.27	3.97	4.91

Table 35 - Database compounds identified as potential outliers in Figure , with the CAS no., reference, log k_{IAM}, log P value and pKa value

Linear regression analysis between log P and log $k_{IAM\ (pH\ 7.4)}$ for all unionised compounds in dataset A gives the following relationship:

$$\log k_{IAM\ (pH\ 7.4)\ (Neu)} = 0.641 \log P - 0.091 \qquad (5.1)$$

n = 126, r^2_{adj} = 0.715, s = 0.662, F = 315, $F_{1,\ 124}$ α, 0.001 = 11.5

All compounds fully ionised under the conditions of analysis gives the following relationship for between log P and log $k_{IAM\ (pH\ 7.4)}$ for dataset A:

$$\log k_{IAM\ (pH\ 7.4)(Ion)} = 0.548 \log P - 1.10 \qquad (5.2)$$

n = 12, r^2_{adj} = 0.722, s = 0.591, F = 29.64, $F_{1,\ 10}$ α, 0.001 = 21.0

Figure 33 shows the relationships for both ionised and unionised compounds with log P. Both equation (5.1) for unionised compounds (log $k_{IAM\ (pH\ 7.4)\ (Neu)}$) and equation (5.2) for ionised compounds (log $k_{IAM\ (pH\ 7.4)\ (Ion)}$) show a good correlation with log P.

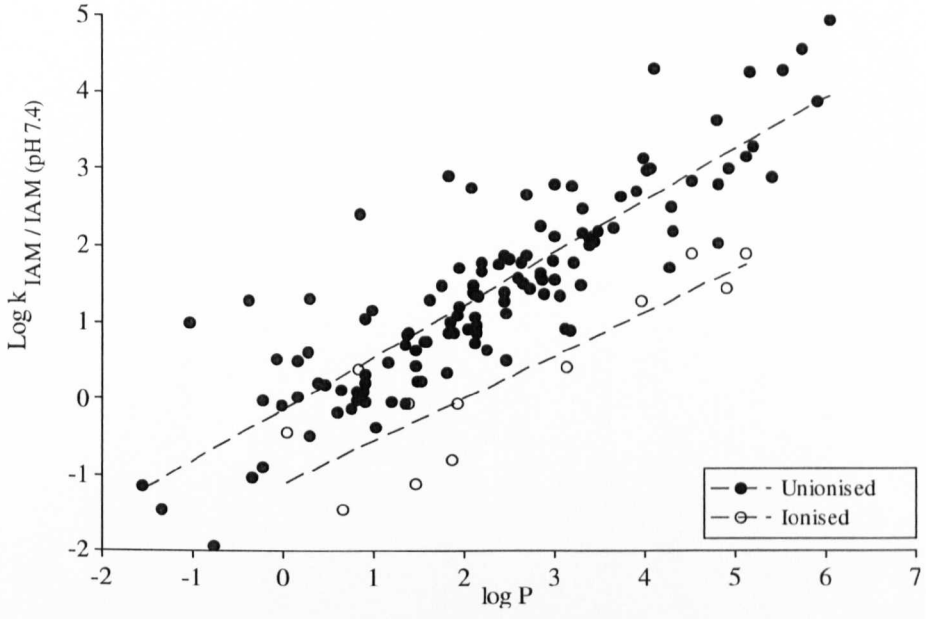


Figure 33 – Plot of log P against log $k_{IAM\ (pH\ 7.4)}$ for dataset A

5.4.2 QSARs for experimental log $k_{IAM\ (pH\ 7.4)}$ values only – Dataset B

Stepwise regression analysis was used to identify the most significant descriptors for log $k_{IAM\ (pH\ 7.4)}$ values for the compounds analysed experimentally in this study (dataset B) (ionised and unionised compounds are considered together). Initial stepwise regression identified the octanol-water partition coefficient log P as the most significant descriptor. Stepwise regression was repeated with the removal of log P, the aqueous solubility (log S) was then identified as the most significant descriptor. It is noted that log S is highly inversely correlated with log P (with a correlation coefficient of -0.93.). Given the high correlation between log P and log S, stepwise regression on log $k_{(IAM)\ pH\ 7.4)}$ values was performed with the removal of both log P and log S as a descriptor, surface area (SA) was identified as the most significant descriptor. The significant descriptors identified across the three stepwise regression were log P, log S, and SA, the number of H-bond donors (Σ_{HBD}), the number of hydrogen bond acceptors (Σ_{HBA}), total molecular energy (E_T), number of rotatable bonds (Σ_{Rot}) and molecular shape flexibility (flex) and molecular volume (vol). The significant descriptors for each stepwise regression were investigated for there degree of inter-correlation (Table 36 - Table 38).

	Log P	Σ_{HBD}	SA	E_T	Σ_{Rot}	flex
Σ_{HBD}	-0.32					
SA	0.79	0.09				
E_T	0.55	-0.42	0.35			
Σ_{Rot}	0.49	0.01	0.67	0.12		
Flex	0.49	0.02	0.62	0.11	0.96	

Table 36 - Correlation matrix for significant descriptors identified for the interpretation of log k_{IAM} using step wise regression for dataset B, with the inclusion of log P as a descriptor

	Log S	E_T
E_T	-0.47	
SA	-0.71	0.35

Table 37 - Correlation matrix for significant descriptors identified for the prediction of log k_{IAM} (pH 7.4) using stepwise regression, with the exclusion of log P from the analysis

	SA	Σ_{HBA}	E_T
Σ_{HBA}	0.278		
E_T	0.351	-0.307	
Vol	0.983	0.270	0.386

Table 38 - Correlation matrix for significant descriptors identified for the prediction of log k_{IAM} (pH 7.4) using stepwise regression, with both log P and log S excluded from the analysis

For log P the highest correlation observed was between Σ_{Rot} and flex with a correlation coefficient of 0.96. For log S (log P excluded) the highest correlation observed was between log S and SA with a correlation co-efficient of -0.71. For SA (log P and log S excluded), the highest correlation observed was between SA and molecular volume with a correlation efficient of 0.98. Including descriptors into a QSAR that are highly inter-correlated may lead to instability in the regression equation[40]. Hence only one of the two inter-correlated descriptors should be used. The remaining correlation coefficients were less than 0.5. This indicates that these descriptors are relatively independent of each other and model different contributions of the compounds. Regression analysis resulted in the QSARs detailed in Table 39.

	No. of descriptors		
	1	2	3
Log P	$\log k_{IAM\ (pH\ 7.4)} = 0.842 \log P - 0.778$ (5.3) $n = 66\ r^2_{adj} = 0.865,\ s = 0.561,$ $F = 418,\ F_{2,63}\ \alpha,\ 0.001 = 7.77$	$\log k_{IAM\ (pH\ 7.4)} = 0.827 \log P - 0.113\ \Sigma_{HBD} - 0.684$ (5.4) $n = 66\ r^2_{adj} = 0.865,\ s = 0.631,$ $F = 209,\ F_{2,63}\ \alpha,\ 0.001 = 7.77$	Not reported – Two descriptor QSAR showed no improvement over one descriptor QSAR
Log S	$\log k_{IAM\ (pH\ 7.4)} = -0.748 \log S + 350$ (5.5) $n = 60\ r^2_{adj} = 0.753,\ s = 0.723,$ $F = 180,\ F_{2,58}\ \alpha,\ 0.001 = 7.96$	$\log k_{IAM\ (pH\ 7.4)} = -0.665 \log S + 0.028\ E_T + 3.20$ (5.6) $n = 60\ r^2_{adj} = 0.782,\ s = 0.679,$ $F = 106,\ F_{3,57}\ \alpha,\ 0.001 = 6.17$	$\log k_{IAM\ (pH\ 7.4)} = -0.472 \log S + 0.035\ E_T + 0.012\ SA$ (5.7) $n = 60\ r^2_{adj} = 0.819,\ s = 0.619,$ $F = 89,\ F_{4,56}\ \alpha,\ 0.001 = 5.31$
SA*	$\log k_{IAM\ (pH\ 7.4)} = 0.02\ SA - 2.18$ (5.8) $n = 66\ r^2_{adj} = 0.524,\ s = 1.06,$ $F = 72\ F_{1,64}\ \alpha,\ 0.001 = 12.0$	$\log k_{IAM\ (pH\ 7.4)} = 0.03\ SA - 0.95\ \Sigma_{HBA} - 2.04$ (5.9) $n = 66\ r^2_{adj} = 0.857,\ s = 0.577,$ $F = 196\ F_{2,63}\ \alpha,\ 0.001 = 7.77$	$\log k_{IAM\ (pH\ 7.4)} = 0.03\ SA - 0.84\ \Sigma_{HBA} + 0.022\ E_T - 1.84$ (5.10) $n = 66\ r^2_{adj} = 0.872,\ s = 0.546,$ $F = 149\ F_{3,62}\ \alpha,\ 0.001 = 6.17$

Table 39 - QSARs developed to predict log $k_{IAM\ (pH\ 7.4)}$ based on the predominant descriptor from stepwise regression analysis
*** Log S values not available for N,N-dicyclohexylcarbodiimine; 2-chlorophenol; 1-naphthol; hexene, 2-nitrobenzoic acid and 1,2 dihydroxybenzene, therefore not included in these analyses.**

For log P the two descriptor QSAR, equation (5.4) is less significant than the 1 descriptor QSAR, equation (5.3). This is indicated by the far lower F value. Additionally, there is no improvement in the $r^2_{(adj)}$ value. Therefore, the inclusion of a single additional descriptor does not improve the QSAR developed. This was also seen with the inclusion of further descriptors (data not reported).

Given log P and log $k_{IAM\ (pH\ 7.4)}$ are both descriptors of partitioning, it is unsurprising that log P is the predominant descriptor to predict log $k_{IAM\ (pH\ 7.4)}$ accounting for 87% of the variance. However, it is the 13% variance that is unaccounted for by log P that is of most interest. For these compounds IAM stationary phase-PBS mobile phase partitioning is different to octanol-water partitioning.

For log S as the number of descriptors increases the the $r^2_{(adj)}$ value increases, equations (5.5) to (5.7), however, the F value decreases. In addition, it is clear that QSARs developed using log S with multiple descriptors do not offer improvement over the QSAR developed for log P as a single descriptor. As noted above, log S and log P are highly inversely correlated and log P is the predominant descriptor in many QSARs to determine solubility[51].

For SA the QSARs developed improve between one and two descriptors, equations (5.8) and (5.9), the further addition of a third descriptor does not improve that statistical significance of the QSAR, equation (5.10).

QSARs were developed to predict log $k_{IAM\ (pH\ 7.4)}$ using physico-chemical descriptors. Plotting $r^2_{(adj)}$ against the number of descriptors for the QSARs developed using log P, log S and SA as the predominant descriptors (Figure 34) illustrates that for both log P and log S, the $r^2_{(adj)}$ values plateau at one descriptor, whereas QSARs using SA as the first descriptor plateau at two descriptors. Therefore, it may be concluded that log P as the sole descriptor is the most appropriate QSAR to predict log $k_{IAM\ (pH\ 7.4)}$, equation (5.3).

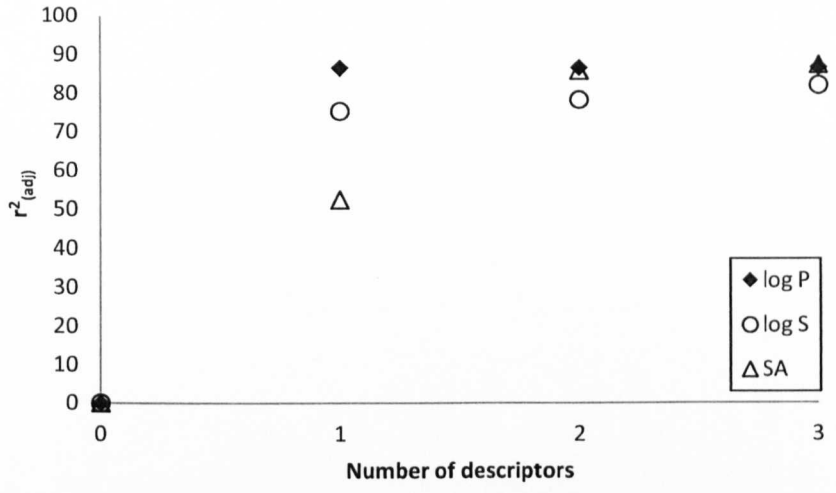


Figure 34 - Plot of $r^2_{(adj)}$ values for QSARs developed with log P and log S and SA as descriptors against the number of descriptors in the QSAR models

Log $k_{IAM\ (pH\ 7.4)}$ and log P are both measures of partitioning and the dominance of log P is not unexpected in predicting log $k_{IAM\ (pH\ 7.4)}$. Due to the high correlation between log P and log S, it is anticipated that the QSARs containing either log P, or log S are modelling partitioning through the two descriptors.

SA as a descriptor accounts for the size of the molecule whereas Σ_{HBD} and Σ_{HBA} account for the ability of the molecules to interact with hydrogen bond donors/acceptors. The latter descriptors (Σ_{HBD} and Σ_{HBA}) are more representative of the complex interactions that occur in the IAM-HPLC partitioning process. E_T is the total energy of the molecule and provides an indication of a molecules size and geometry which affects partitioning between two phases. Therefore, the descriptors identified as significant in predicting log $k_{IAM\ (pH\ 7.4)}$ are mechanistically interpretable and provide confirmation not only of the model, but also of the reliability and meaning of the log $k_{IAM\ (pH\ 7.4)}$ data themselves.

There are a number of ways to define the domain of a QSAR[52]. If the domain is defined by the structures of the compounds used to develop the model i.e. structural domain[53], then the domain of these QSARs are the same as the domain for the fragment and factor model determined in Chapter 4. This is because the compounds used to develop the models are identical. However, if the domain is defined by a physico-chemical property[54] i.e. range of log P, log S etc. then the domain with

respect to functionality may be increased i.e. chemical space within the domain is larger. However, compounds vastly different from those used to generate the model may be poorly predicted. Consensus has not been reached on how to define the domain of a QSAR.

If we use a structural definition of the domain for both the QSARs developed here and the fragment and factor model of Chapter 4 the domains are identical. However, the predictive capabilities of the fragment and factor model are greater than the QSAR models. It is also noted that the requirements to retrain the descriptor based QSAR models, to extend the domain are less intensive than to retrain the fragment and factor model.

5.5 Conclusions

QSARs have been developed to determine log k_{IAM} for the combined dataset of 134 compounds (dataset A). QSARs were developed considering ionised and neutral compounds separately, both QSARs showed good predictive capabilities (r^2_{adj} = 0.72 and 0.71 respectively, equations (5.1) and (5.2)). Considering just the experimentally determined log $k_{IAM\ (pH\ 7.4)}$ values (dataset B) the most significant descriptor was indentified as log P. This is consistent with existing knowledge that hydrophobicity is important in modelling log k_{IAM} and is supported by the number of QSARs reported in the literature utilising log P or log D as the sole descriptor in the QSAR.

QSARs were developed both with the inclusion and exclusion of log P as a descriptor, excluding log P allowed for the elucidation of potential descriptors that, due to the significance of log P in describing log $k_{IAM\ (pH\ 7.4)}$ were lost within the initial analysis.

For both log P and log S, the use of more than one descriptor did not significantly increase the statistical fit of the QSARs. Whereas for SA the $r^2_{(adj)}$ values plateau at two descriptors and the statistical fit of the QSAR improves between one and two descriptors. Additionally, when log P is excluded as a potential descriptor the predominant descriptor becomes log S for solubility, which itself is highly correlated to log P.

Using traditional descriptors to develop QSARs to predict log k_{IAM} generates QSARs that have far lower $r^2_{(adj)}$ values than those determined from the fragment and factor value approach (Chapter 4). The use of traditional QSAR descriptors (log P or alternative structural or non-empirical descriptors) reduces the potential for a compound being outside the predictive domain of the QSAR, which in turn increases the applicability of the QSAR. Therefore, QSARs, such equation (5.3) provide a more global model than the use of fragments, albeit with a reduced precision in the prediction. However given the current limited training set, the requirement to retrain the models as more compounds are analysed (containing unknown structural fragments and features to the models), for both approaches is great. Given the complexity of the fragment and factor model approach and the intensive approach adopted to cross validate the model; the descriptor based QSAR approach is far quicker and easier to retrain.

Given the current training set from which both the fragment and factor model and the traditional QSAR models were developed; it is recommended that the domain for both methods is currently defined by the structures within the training set. Although this reduces the structural diversity of the domain, confidence is increased in predictions obtained from both model approaches considered here. If the diversity of the compounds increased and number of compounds used to train the QSAR models increased greatly, the domains for both approaches would change appropriately. It could be argued that with a sufficient number of compounds the domains of the two approaches could and perhaps should be defined separately.

5.6 References

[1] European Union. (1995) *Overview of Structure-Activity Relationships for Environmental Endpoints Part 1: General Outline and Procedure, Report of the EU-DG-XII Project "QSAR for Predicting Fate and Effects of Chemicals in the Environment (Contract ~ EV5V-CT92-0211).*

[2] Madden J.C. (2010) Introduction to QSAR and Other In Silico Methods to Predict Toxicology. In: Cronin M.T.D., Madden J.C. eds, *In Silico Toxicology: Principles and Applications*, Cambridge, Royal Society Chemistry, pp 11-29.

[3] Modarresi H., Dearden J.C., Modarress H. (2006) QSPR Correlation of Melting Point for Drug Compounds Based on Different Sources of Molecular Descriptors *J. Chem. Inf. Model.* 46: 930-936.

[4] Nasal A., Sznitowska M., Buciński A., Kaliszan R. (1995) Hydrophobicity Parameter from High-Performance Liquid Chromatography on an Immobilized

Artificial Membrane Column and its Relationship to Bioactivity *J. Chromatogr. A* 692: 83-89.
[5] Ward R.S., Davies J., Hodges G., Roberts D.W. (2003) Applications of Immobilised Artificial Membrane Chromatography to Quaternary Alkylammonium Sulfobetaines and Comparison of Chromatographic Methods for Estimating the Octanol-Water Partition Coefficient *J. Chromatogr. A* 1007: 67-75.
[6] Moss G.P., Cronin M.T.D. (2002) Quantitative Structure-Permeability Relationships for Percutaneous Absorption: Re-analysis of Steroid Data *Int. J. Pharm.* 238: 105-109.
[7] Potts R.O., Guy R.H (1992) Predicting Skin Permeability *Pharm. Res.* 9: 663-669.
[8] Meylan W.M., Howard P.H. (1995) Atom/Fragment Contribution Method for Estimation Octanol-Water Partitioning Coefficients *J. Pharm. Sci.* 84: 83-91.
[9] Tetko I.V., Gasteiger J., Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., Prokopenko V.V. (2005) Virtual Computational Chemistry Laboratory – Design and Description *J. Comput. Aid. Mol. Des.* 19: 453-63.
[10] VCCLAB, Virtual Computational Chemistry Laboratory(2005) available from http://www.vcclab.org
[11] Salminen T., Pulli A., Taskinen J. (1997) Relationship Between Immobilised Artificial Membrane Chromatographic Retention and the Brain Penetration of Structurally Diverse Drugs *J. Pharm. Biomed. Anal.* 15: 469-477.
[12] Abraham M.H., Chadha H.S., Leitao R.A.E., Mitchell R.C., Lambert W.J., Kaliszan R., Nasal A., Haber P. (1997) Determination of Solute Lipophilicity, as Log P(Octanol) and Log P(Alkane) using Poly(stytene-divinylbenzene) and Immobilised Artificial Membrane Stationary Phases in Reversed-Phase High-Performance Liquid Chromatography *J. Chromatogr. A* 766: 35-47.
[13] Amato M., Barbato F., Morrica P., Qiaglia F., La Rotonda M.I. (2000) Interactions Between Amines and Phospholipids: A Chromatographic Study of Immobilized Artificial Membrane (IAM) Stationary Phases at Various pH Values *Helv. Chim. Acta.* 83: 2836-2847.
[14] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2005) Can Protonated β-Blockers Interact with Biomembranes Stronger than Neutral Isolipophilic Compounds? A Chromatographic Study on Three different Phospholipid Stationary Phases (IAM-HPLC) *Eur. J. Pharm. Sci.* 25: 379-386.
[15] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2004) Prediction of Drug-Membrane Interactions by IAM-HPLC: Effects of Different Phospholipid Stationary Phases on the Partition of Bases *Eur. J. Pharm. Sci.* 22: 261-269.
[16] Barbato F., Cirocco V., Grumetto L., La Rotonda M.I. (2007) Comparison Between Immobilized Artificial Membrane (IAM) HPLC Data and Lipophilicity in n-Octanol for Quinolone Antibacterial Agents *Eur. J. Pharm. Sci.* 31: 288-297.
[17] Barbato F., La Rotonda M.I., Quaglia F. (1997) Chromatographic Indexes on Immobilized Artificial Membranes for Local Anesthetics: Relationships with Activity Data on Closed Sodium Channels *Pharm. Res.* 14: 1699-1705.
[18] Barbato F., La Rotonda M.I., Quaglia F. (1996) Chromatographic Indices Determined on an Immobilized Artificial Membrane (IAM) Column as Descriptors of Lipophilic and Polar Interactions of 4-Phenyldihydropyridine Calcium-Channel Blockers with Biomembranes *Eur. J. Med. Chem.* 31: 311-318.
[19] Barbato F., La Rotonda M.I., Quaglia F. (1997) Interaction of Nonsteroidal Antiinflammatory Drugs with Phospholipids: Comparison Between Octanol/Water

Partition Coefficients and Chromatographic Indexes on Immobilized Artificial Membranes *J. Pharm. Sci.* 86: 225-229.
[20] Cimpean D.M., Poole C.F. (2002) Systematic Search for Surrogate Chromatographic Models of Biopartitioning Processes *Analyst* 127: 724-729.
[21] Demare S., Roy D., Legerdre J.Y. (1999) Factors Governing the Retention of Solutes on Chromatographic Immobilized Artificial Membranes: Application to Anti-Inflammatory and Analgesic Drugs *J. Liq. Chromatogr. Relat. Technol.* 22: 2675-2688.
[22] Kaliszan R., Kaliszan A., Wainer I.W. (1993) Deactivated Hydrocarbonaceous Silica and Immobilized Artificial Membrane Stationary Phases in High-Performance Liquid Chromatographic Determination of Hydrophobicities of Organic Bases: Relationship to Log P and CLOGP *J. Pharm. Biomed. Anal.* 11: 505-511.
[23] Kępczyńska E., Bojarski J., Haber P., Kaliszan R. (2000). Retention of Barbituric Acid Derivatives on Immobilized Artificial Membrane Stationary Phase and its Correlation with Biological Activity *Biomed. Chromatogr.* 14: 256-260.
[24] Kotecha J., Shah S., Rathod I., Subbaiah G. (2008) Prediction of Oral Absorption in Humans by Experimental Immobilized Artificial Membrane Chromatography Indices and Physicochemical Descriptors *Int. J. Pharm.* 360: 96-106.
[25] Lázaro E., Ráfols C., Rosés M. (2005) Characterization of Immobilized Artificial Membrane (IAM) and XTerra Columns by Means of Chromatographic Models *J. Chromatogr. A* 1081: 163-173.
[26] Lepont C., Poole C.F. (2002) Retention Characteristics of an Immobilized Artificial Membrane Column in Reversed-Phase Liquid Chromatography *J. Chromatogr. A* 946: 107-124.
[27] Li J., Cui S., He Z. (2006) Quantitative Structure-Retention relationship Studies Using Immobilized Artificial Membrane Chromatography I: Amended Linear Solvation Energy Relationships with the Introduction of a Molecular Electronic Factor *J. Chromatogr. A* 1132: 174-182.
[28] Li J., Sun J., He Z. (2007) Quantitative Structure-Retention Relationship Studies with Immobilized Artificial Membrane Chromatography II: Partial Least Squares Regression *J. Chromatogr. A* 1140: 174-179.
[29] Luco J.M., Salinas A.P., Torriero A.A.J., Vázquez R.N., Raba J., Marchevsky E. (2003) Immobilized Artificial Membrane Chromatography: Quantitative Structure-Retention Relationships of Structurally Diverse Drugs *J. Chem. Inf. Comput. Sci.* 43: 2129-2136.
[30] Pehourcq F., Jarry C., Bannwarth B. (2003) Potential of Immobilized Artificial Membrane Chromatography for Lipophilicity Determination of Arylpropionic Acid Non-Steroidal Anti-inflammatory Drugs *Pharm. Res.* 33: 137-144.
[31] Sarr F.S., André C., Guillaume Y.C. (2008) Statins (HMG-coenzyme A Reductase Inhibitors)-Biomimetic Membrane Binding Mechanism Investigated by Molecular Chromatography *J. Chromatogr. B* 868: 20-27.
[32] Sprunger L., Blake-Taylor B.H., Wairegi A., Acree Jr. W.E., Abraham M.H. (2007) Characterization of the Retention Behavior of Organic and Pharmaceutical Drug Molecules on an Immobilized Artificial Membrane Column with the Abraham Model *J. Chromatogr. A* 1160: 235-245.
[33] Di Stefano A., Sozio P., Iannitelli A., Cerasa L.S., Fonana A., Di Biase G., D'Amico G., Di Giulio M., Carpentiero C., Grumetto L., Barbato F. (2008) Characterization of Alkanoyl-10-Ominocyclines in Micellar Dispersions as Potential

Agents for Treatment of Human Neurodegenerative Disorders *Eur. J. Pharm. Sci.* 34: 118-128.

[34] Taillardat-Bertschinger A., Barbato F., Quercia M.T., Carrupt P.A., Reist M., La Rotonda M.I., Testa B. (2002) Structural Properties Governing Retention Mechanisms on Immobilized Artificial Membrane (IAM) HPLC Columns *Helv. Chim. Acta.* 85: 519-532.

[35] Valkó K., Plass M., Bevan C., Reynolds D., Abraham M.H. (1998) Relationships Between the Chromatographic Hydrophobicity Indices and Solute Descriptors Obtained by Using Several Reversed-Phase, Diol, Nitrile, Cyclodextrin and Immobilised Artificial Membrane-Bonded High-Performance Liquid Chromatography Columns *J. Chromatogr. A* 797: 41-55.

[36] Lázaro E., Ráfols C., Abraham M.H., Rosés M. (2006) Chromatographic Estimation of Drug Disposition Properties by Means of Immobilized Artificial Membranes (IAM) and C18 Columns *J. Med. Chem.* 49: 4861-4870.

[37] Vrakas D., Giaginis C., Tsantili-Kakoulidou A. (2006) Different Retention Behavior of Structurally Diverse Basic and Neutral Drugs in Immobilized Artificial Membrane and Reversed-Phase High Performance Liquid Chromatography: Comparison with Octanol-Water Partitioning. *J. Chromatogr. A* 1116: 158-164.

[38] Vrakas D., Giaginis C., Tsantili-Kakoulidou A. (2008) Electrostatic Interactions and Ionization Effect in Immobilized Artificial Membrane Retention A Comparative Study with Octanol-Water Partitioning. *J. Chromatogr. A* 1187: 67-78.

[39] Plass M., Valko K., Abraham M.H. (1998) Determination of Solute Descriptors of Tripeptide Derivatives Based on High-Throughput Gradient High-Performance Liquid Chromatography Retention Data *J. Chromatogr. A* 803: 51-60.

[40] Dearden J.C., Cronin M.T.D., Kaiser K.L.E. (2009) How Not to Develop a Quantitative Structure-Activity or Structure-Property Relationship (QSAR/QSPR) *SAR QSAR Environ. Res.* 20: 241-266.

[41] Valkó K., Du C.M., Bevan C., Reynolds D.P., Abraham M.H. (2000) Rapid-Gradient HPLC Method for Measuring Drug Interactions with Immobilized Artificial Membrane: Comparison with Other Lipophilicity Measures *J. Pharm. Sci.* 89: 1085-1096.

[42] Valkó K., Du C.M., Bevan C., Reynolds D.P., Abraham M.H. (2001) Rapid Method for the Estimation of Octanol/Water Partitioning Coefficient (log P_{Oct}) from Gradient RP-HPLC Retention and a Hydrogen Bond Acidity Term (Sigma-Alpha H_2) *Curr. Med. Chem.* 8: 1137-1146.

[43] Valkó K., Bevan C., Reynolds D.P. (1997) Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log P/log D *Anal. Chem.* 69: 2022-2029.

[44] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.

[45] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition coefficient (n-octanol/water), High performance liquid chromatography (HPLC) method*, Paris, Organisation for Economic Cooperation and Development.

[46] Dearden J.C., Bresuen G.M. (1988) The Measurement of Partition Coefficients and Lipophilicity *Quant. Struct.-Act. Relat.* 7: 133-144.

[47] Accelrys Software Inc. (2010) TSAR version 3.3, San Diego: Accelrys Software Inc..

[48] Minitab Inc. (2007) Minitab® Statistical Software version 15, Coventry: Minitab Inc..
[49] Barton P., Davis A.M., McCarthy D.J., Webborn P.J.H. (1997) Drug-Phospholipid. 2. Predicting the Sites of Drug Distribution Using n-Octanol/Water Distribution Coefficients *J. Pharm. Sci.* 86: 1034-1039.
[50] Kangas H., Kotiaho T., Salminen T., Kostiainen R. (2001) N-in-one Determination of Retention Factors for Drugs by Immobilized Artificial Membrane Chromatography Coupled to Atmospheric Pressure Chemical Ionization Mass Spectrometry *Rapid Commun. Mass Spectrom.* 15: 1501-1505.
[51] Hewitt M., Cronin M.T.D., Enoch S.J., Madden J.C., Roberts D.W., Dearden J.C. (2009) In Silico Prediction of Aqueous Solubility: The Solubility Challenge *J. Chem. Inf. Model.* 49: 2572-2587.
[52] Tropsha A., Gramatica P., Gombar V.K. (2003) The Importance of Being Earnest: Validation in the Absolute Essential for Successful Application and Interpretation of QSPR Models *QSAR Comb. Sci.* 22: 69-77.
[53] Dimitrov S., Dimitrova G., Pavlov T., Dimitrova N., Patlewicz G., Niemela J., Mekenyan O. (2005) A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models *J. Chem. Inf. Model.* 45: 839-849.
[54] Netzeva T.I., Worth A.P., Aldenberg T., Benigni R., Cronin M.T.D., Gramatica P., Jaworska J.S., Kahn S., Klopman G., Marchant C.A., Myatt G., Nikolova-Jeliazkova N., Patlewicz G.Y., Perkins R., Roberts D.W., Schultz T.W., Stanton D.T., van de Sandt J.J.M., Tong W., Veith G., Yang C. (2005) Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships *Altern. Lab. Anim.* 33: 1-19.

6 QSAR – Using log k_{IAM} as a descriptor to predict skin absorption

6.1 Introduction

6.1.1 Skin absorption

The skin is the largest organ of the body[1]. It forms a barrier, predominantly to prevent dehydration, but also to prevent the entry of xenobiotics into the body[2]. Additionally, the skin has metabolic capability[3]. The ability of chemicals to penetrate the skin depends on the chemical's properties (i.e. hydrophobicity, size etc.), concentration and other factors including temperature, damage to the skin and age of the subject[4]. Understanding the absorption of molecules through human skin is important to the design of formulations for the delivery of drugs applied topically[5]. It is important in the design of personal care products (where absorption may be less desirable); and generally in the risk assessment of dermal exposure of chemicals[6]. In order to assess the absorption of chemicals through the skin, it is important to understand the physiology of the skin and the factors which affect skin absorption[7].

The definitions with respect to dermal absorption are diverse. The definitions detailed below have been proposed by the World Health Organisation (WHO)[8] and are used by the Europeans Commission's Scientific Committee on Consumer Safety (SCCS). These definitions will be used throughout the thesis.

> "**Dermal absorption**: Process is a global term which describes the passage of compounds across the skin"[8]

In *in vivo* analysis absorption would constitute the quantity that has reached systemic circulation. For *in vitro* experiments this would be the quantity that reached the receptor cell.

> "**Penetration**: Which is the entry of a substance into a particular layer or structure such as the entrance of a compound into the stratum corneum" [8]

> "**Permeation**: Which is the penetration through one layer into another, which is both functionally and structurally different from the first layer" [8]

6.1.2 **Anatomy of the skin barrier**

A cross section of human skin is shown in Figure 35. The skin is composed of two main layers; these are the epidermis and dermis[9]. The outer layer of the epidermis is referred to as the stratum corneum. The stratum corneum is the main barrier to chemicals crossing the skin barrier[10]. The stratum corneum predominantly comprises of sphingolipids, which contain phosphatidylcholine (PC) attached to a sphingosine backbone[11]. The IAM stationary phase is the same PC structure endcapped with different groups. This makes IAM-HPLC a potential source of useful information for the prediction of skin absorption of chemicals across the stratum corneum.

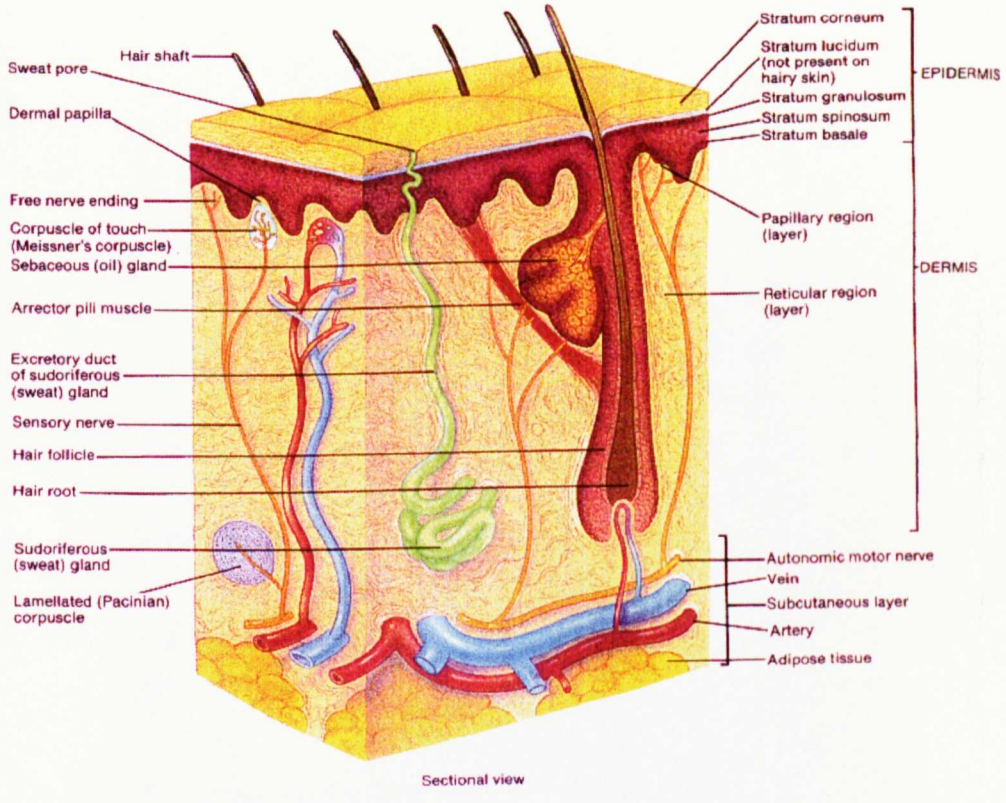


Figure 35 - Cross section of human skin[12]

Skin absorption involves a compound partitioning into, and through the stratum corneum. Size[6], hydrophobicity[13] and degree of ionisation[10, 14] of chemicals are important properties that determine whether or not a compound crosses the skin barrier. Additionally, the skin contains hair follicles and sweat glands that provide a route through the epidermis that by-passes the stratum corneum[9]. However, this route has not been shown to contribute significantly to the absorption of chemicals (a

poor correlation is obtained between follicle density and absorption, when site of application is compared[9]). Despite this, the potential for compounds to by-pass the stratum corneum through hair follicles or sweat glands is a potential source of difference between predictions based on measures of hydrophobicity and experimentally determined skin absorption values.

6.1.3 Determination of skin absorption

Traditionally dermal absorption of chemicals has been determined through *in vivo*[15] or *in vitro*[16] experiments. The ability of a compound to cross the skin barrier can be measured experimentally and can be described by the permeability coefficient (K_p) this is expressed mathematically as in equation (6.1). An alternative description of the ability of a compound to cross the skin barrier is the flux (J) of a compound, equation (6.2), or maximum flux (J_{max}). More recently QSARs[17] have been developed that allow for the prediction of skin absorption.

$$K_P(cm/hr) = \frac{K_m D_m}{L} \qquad (6.1)\ ^{10}$$

$$J\ (mol/cm^2\ hr) = K_P \Delta C \qquad (6.2)\ ^{18}$$

Where

K_P is the permeability coefficient

K_m is the partition coefficient of the compound into the skin membrane

D_m is the membrane diffusion coefficient

L is the path length of the membrane

ΔC is the difference in concentration between the two sides of the membrane

6.1.3.1 *In vivo* method of determining skin absorption

A standardised method to determine the skin absorption of a chemical using an *in vivo* technique is described in OECD guideline 427[15]. *In vivo* testing involves the application of the substance, generally radiolabelled, to clipped skin (studies are usually performed on rats, although other species can be used). The application area is covered to prevent ingestion of the substance. Following exposure (tests usually last between 6 and 24 hours) the area of application is cleaned. The excreta and expired air are monitored for radioactive metabolites. The test involves several

groups of animals; one group is killed at the end of the exposure time. Other groups are killed at scheduled time points. Following the exposure period the animals are observed for visible signs of irritation and the animals are subsequently killed. Blood is collected, the application site is removed and the carcass is analysed for any unexcreted material. In addition, the application site is analysed for any residual material[15]. *In vivo* testing is expensive, time consuming and there are ethical issues surrounding the use of animals. Therefore, the use of alternative methods to *in vivo* testing is preferable where possible.

6.1.3.2 *In vitro* method of determining skin absorption

A standardised *in vitro* method to determine skin absorption of a chemical is described in OECD guideline 428[16]. This recognises both static methods, i.e. Franz-type static diffusion cell, and flow through methods, i.e. Bronaugh-type flow through diffusion cell (refer to Figure 36). The static and flow through diffusion cell methods are similar. The skin sample is placed between two chambers (the upper chamber is the donor chamber and the lower chamber is the receptor chamber). The donor chamber contains the test sample. This can be applied neat, dilute, or in a formulation. The application and duration of exposure mimics the end use, i.e. the specific expected exposure. The use of a physiologically relevant fluid in the receptor cell is preferred[19].

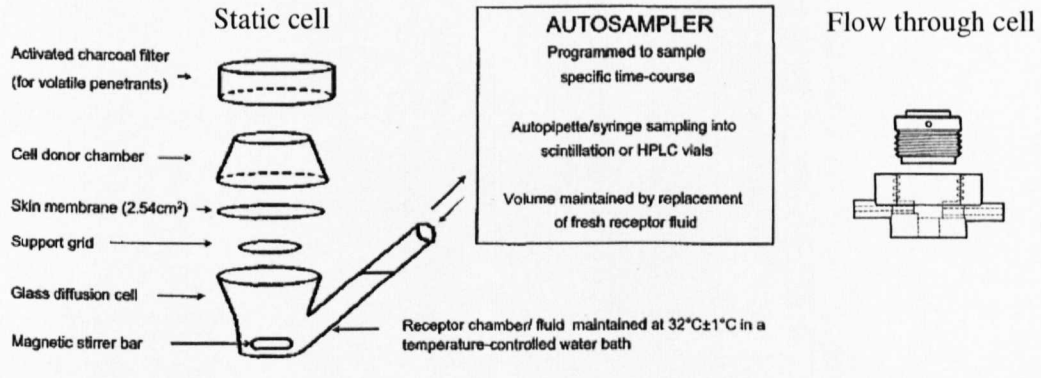


Figure 36 - Static and flow through cells for *in-vitro* dermal absorption[19]

In the static set-up the receptor cell is continuously stirred and sampled at regular intervals. In the flow through set-up the receptor chamber fluid is continuously refreshed. However, the rate of flow must not hinder diffusion of the sample into the

receptor fluid. Samples are collected from the receptor cell for up to 24 hours. Once the test has been completed all components of the system are analysed to determine recovery. The amount of sample absorbed through the skin is then determined. An alternative method is the time-course method; this is similar to both the static and flow-through methods. The initial dose is applied to a number of sample groups. Then at various time points (generally t = 0.5, 1, 2, 4, 8 and 24 hours), groups of the cells are analysed and absorption determined across the time points.

The *in vitro* method of determining skin absorption, although standardised, has many variables. These include the species from which the skin is obtained (which includes rat, pig and human)[11], the type of skin used (which includes full thickness dermatomed skin, epidermal membrane (heat, enzymically or chemically separated), or split thickness skin prepared with a dermatome), test set-up used (which includes static diffusion cell, flow through methods and time-course methods), test substance preparation (which includes pure compounds, diluted solutions, use of a vehicle or formulation to apply the material of interest) and the duration of exposure. Therefore, whilst many *in vitro* absorption data have been published, there is considerable variability in the experimental procedures used, and therefore, result obtained.

6.1.3.3 Sources of skin absorption data

There are several datasets of skin absorption values. The Flynn dataset[9] contains previously published *in vitro* data from studies using human excised skin to determine permeability coefficients for 93 compounds. This dataset has been modelled by numerous researchers, using a range of descriptors. Potts and Guy[6] found log P and molecular weight to be the most relevant descriptors describing 67% of the variability in the data (30% experimental variation in skin permeability data is not uncommon[20]). Analysis by Thomas *et al.*[21], of an edited and extended Flynn dataset (n=114) (the edited Flynn dataset was reported by Majumdar *et al.*[22] (n= 62), the extension included nine contributions[23-32] (n=52) from eight laboratories) fitted the data to the Robert-Sloan[33] equation, equation (6.3). This analysis described 93% of the variability within the dataset, indicating the model is over-fitted.

$$\log J_{M\,Aq} = x + y \log S_{Oct} + (1 - y) \log S_{Aq} - z\,MW \quad (6.3)$$

Where:

$J_{M\,Aq}$ is the maximum flux from an aqueous vehicle

S_{Oct} is the solubility in octanol

S_{Aq} is the solubility in water

MW is the molecular weight

The analysis by Barratt[7] of the Flynn dataset (n=91 (compounds were removed either due to duplication, or due to an experimental log P value being unavailable)) identified the most relevant descriptors as log P, molecular volume and melting point. The resulting model described 76% of the variability within the dataset. Abraham *et al.*[34] modelled an extended Flynn[9] dataset (n=119), (The additional data were obtained from the following references 24, 29, 35-56) using the general linear free-energy relationship descriptors (solute excess refractivity, solute dipolarity/polarisability, overall solute hydrogen bond acidity and basicity and the McGowan characteristic volume). These descriptors described 83% of the variability within the dataset. Dearden *et al.*[57] modelled an extended Flynn[9] dataset published by Abraham and Martins[34] (n=107). Dearden *et al.* found H-bond acceptor and donor ability, 2nd and 4th order valence molecular connectivity (these descriptors describe unique paths[58] (of different lengths) through a molecule; the molecule is hydrogen suppressed but the number of valence electrons is taken into account, refer to Figure 37 for details), ZZX plane VAMP octupole moment and aqueous solubility to be relevant descriptors. The majority of these models include a descriptor of hydrophobicity and a descriptor of molecular size.

Johnson *et al.*[35] investigated the primary references for the Flynn dataset and found that the Scheuplein[59] values, relating to steroids, were substantially different to those values reported by other sources for the same compounds. The corrected values for steroids along with the remaining Flynn[9] dataset have been incorporated into the Moss[13] dataset which contains measured and calculated K_P values for 119 compounds. In this case the parameters which were found to describe 82% of the variability were log P and molecular weight.

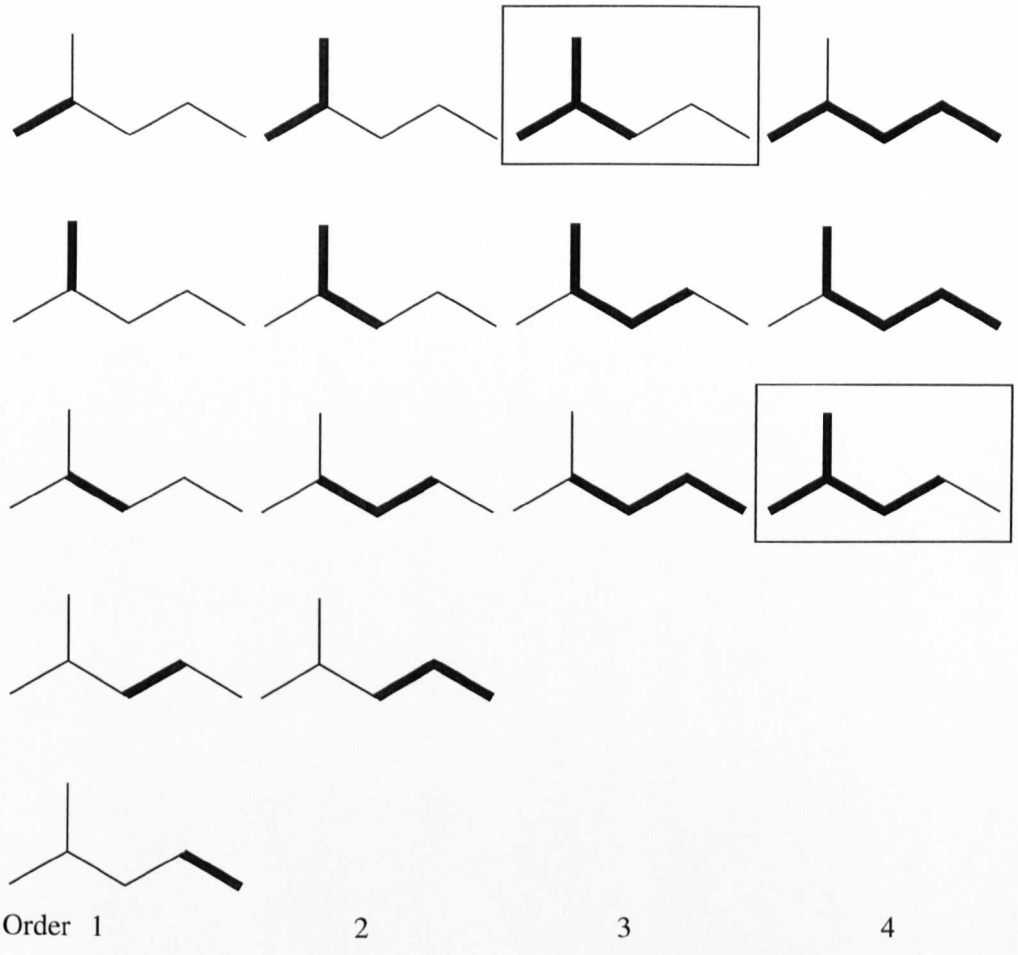


Figure 37 – Connectivity index paths of orders 1, 2, 3 and 4 for 2-methyl pentane (in the boxes are a cluster of order 3 and a path cluster of order 4)[58]

Evaluations and predictions of DErmal absorption of TOXic chemicals (EDETOX) is an electronic database[60] (available from http://edetox.ncl.ac.uk/), which contains both *in vivo* and *in vitro* skin absorption results for more than 300 compounds. The database contains the experimental conditions as well as a citation for the original source of the data. EDETOX includes data from both the Flynn[9] and Moss[13] datasets.

6.1.3.4 Use of QSARs in determining skin absorption

With the introduction of REACH[61], which emphasises limiting the requirement for testing on animals and promotes the use of non-animal test data, alternatives to the traditional methods of determining skin penetration are required. Although the *in vitro* method is classed as a non-animal test under the REACH legislation it is still expensive and time consuming. Therefore, the development of QSARs is beneficial

in providing a quick and cheap indication of a chemical's ability to penetrate the skin. This also allows for targeted *in vitro* testing and would aid with experimental design.

Skin permeability can be considered as a series of partitioning processes i.e. vehicle to skin, lipid to aqueous, aqueous to lipid. Therefore, it appears logical that QSARs for predicting skin penetration are based on measures of hydrophobicity and molecular size[6, 13, 39, 62]. Predominantly the descriptor of hydrophobicity is log P and the descriptor for molecular size is molecular mass. An indicator of hydrophobicity represents the partitioning of the compound through the membrane (given as K_m in equation (6.1)) and an indication of molecular size is used to model the membrane diffusion coefficient (given as D_m in equation (6.1)).

The QSAR reported by Moss and Cronin[13], equation (6.4), supports the conclusions of Johnson *et al.* that the Scheuplein *et al.*[59] K_P data (part of the Flynn dataset) may be erroneous. This was demonstrated using alternative sources of K_P values (reported by Johnson *et al.*[35] and Degim *et al.*[52]) for steroids. Using the updated values, steroids were no longer outliers to the model developed.

$$\log K_P = -2.29 + 0.74 \log P - 0.0091\, MW \qquad (6.4)$$

$n = 116\ r^2_{adj} = 0.82,\ s = 0.42,\ F = 266,\ F_{2,\,113}\ \alpha,\ 0.001 = 7.41$

In addition to hydrophobicity and molecular size, many other descriptors have been found to be important in modelling skin penetration e.g. solubility[21], melting point[7] and H-bond acceptor and donor capability[57].

6.1.3.5 Use of IAM in determining skin absorption

The rate limiting step in skin absorption of chemicals is the partitioning of chemicals across the stratum corneum[9]. The predominant lipid present in the stratum corneum is phosphatidylcholine (PC), this is also the stationary phase that the IAM-HPLC column contains. Therefore, IAM-HPLC could be a potential surrogate to log P in predicting skin absorption of chemicals. The use of IAM-HPLC values to model skin absorption through the development of QSARs takes into account both hydrophilic and hydrophobic interactions during partitioning into and across lipids membranes[11].

Investigations into predicting skin absorption using IAM-HPLC have found that log k_{IAM} is a better descriptor than log P for short chain alcohols[11] and compounds known to absorb through the skin in an ionised form[14]. In addition, Barbato *et al.* analysed 12 drugs (comprising acidic, basic and non-ionisable compounds) and found log k_{IAM} to be a comparable descriptor to log P in describing skin absorption[10].

6.2 Aim of the chapter

The aim of the chapter was to investigate the use of log $k_{IAM\ (pH\ 7.4)}$ as an alternative descriptor of hydrophobicity in QSARs to predict skin absorption. Additionally, given that many QSARs to predict skin absorption contain two descriptors (one for hydrophobicity and a second for molecular size) and IAM-HPLC accounts for more intra/inter-molecular interactions than log P, the requirement of a second descriptor was also investigated.

6.3 Method

6.3.1 Skin absorption datasets

Skin absorption data were obtained from the literature; specifically the EDETOX[60] electronic database and both the Moss[62] and Flynn[9] datasets were used. These datasets were cross-referenced with the IAM data from both the literature database (refer to Chapter 2, for full details) and those determined experimentally within this thesis (refer to Chapter 3 and Chapter 4 for log $k_{IAM\ (pH\ 7.4)}$ results). It was necessary to divide the resulting data into subsets containing data obtained under comparable experimental methods for both the skin absorption and log k_{IAM} determination. Whilst it was noted that the EDETOX database contains both the Moss and Flynn datasets, both these datasets been modelled extensively by many research groups to develop QSARs to predict skin absorption. Therefore, QSARs were developed for the Moss, Flynn and EDETOX datasets separately, using log k_{IAM} as a descriptor for hydrophobicity and an additional descriptor for molecular size (where the dataset size allowed). QSARs were also developed using log P_{exp}[63] to allow comparison between the hydrophobicity descriptors.

6.3.2 Log $k_{IAM\ (pH\ 7.4)}$ values

Log $k_{IAM\ (pH\ 7.4)}$ values have been determined using a robust IAM-HPLC assay for 66 diverse compounds (Chapters 3 and 4). These values along with consistent and

comparable log k_{IAM} values from the log k_{IAM} database (Chapter 2) were used to develop QSARs for the prediction of skin absorption. Log k_{IAM} values determined experimentally in this thesis will be referred to as experimental values, whereas experimental values collated in the log k_{IAM} database will be referred to as literature values.

6.3.3 Statistical analysis of the data

Regression analysis was performed to generate the QSARs for each dataset. This was performed using log P_{exp} (from KOWWIN[63]), log k_{IAM} and log $k_{IAM\ (pH\ 7.4)}$ (log $k_{IAM\ (pH\ 7.4)}$ was determined experimentally in Chapter 3, log k_{IAM} were values from the database where the pH was specified) as descriptors of hydrophobicity. Molecular weight[63] was used as a descriptor for molecular size. The linear regression analyses were performed using Minitab[64] (version 15.1.1.0). The following statistical information was recorded: n, $r^2_{(adj)}$, s and F values.

6.4 Results and discussion

6.4.1 EDETOX

6.4.1.1 Experimental log $k_{IAM\ (pH\ 7.4)}$ values

The EDETOX database was searched for compounds for which log $k_{IAM\ (pH\ 7.4)}$ had been determined. EDETOX returned 166 *in vitro* results for ten compounds which had been determined experimentally in this thesis (Chapters 3 and 4). K_p or flux values obtained under consistent and comparable experimental conditions were required for modelling skin penetration. The experimental procedures reported in the EDETOX database were preformed in such diverse ways that it was not possible to collate datasets containing five compounds[65] that were preformed under comparable conditions. Experimental differences within the EDETOX data included species, diffusion cell type, membrane type, vehicle, exposure area and duration of exposure. Despite the high number of values obtained, many entries do not contain flux or K_P values, instead percentage recovered, or percentage absorbed are reported. The later measures are not comparable.

6.4.1.2 Database log k_{IAM} values

EDETOX returned 685 *in vitro* results for 80 compounds from the log k_{IAM} values within the database (Chapter 2). Five subsets of skin absorption data were collated

for compounds which have consistent and comparable experimental conditions for the IAM measurement (consistent column stationary phase, mobile phase composition and pH) and K_P determination (species, membrane type, cell type and vehicle were standardised). Averages of log k_{IAM} and K_P were taken, for compounds with multiple results. The same issues were encountered with variability in experimental procedure and reporting of Kp values, as mentioned above. However, the effects were less pronounced due to an increased number of compounds and results available.

The log k_{IAM} values from the database were taken from several papers[66-81]. The five subsets of data are detailed in Appendices 1.5, Tables 26 & 27.

For each data subset regression analysis was performed using Minitab. Analysis was performed with both log k_{IAM} and log P_{exp} for comparison. The small number of compounds in all but one of the subsets means only one descriptor should be used[65]. Subset one contains 10 compounds meaning, the inclusion of two descriptors is possible, regression analysis performed using molecular weight as an additional descriptor to both measures of hydrophobicity.

All regression analyses were poor, with all having either an $r^2_{(adj)} < 0.20$ or an F value less than the relevant F_α value (refer to Appendix 1.5, Table 28 for a summary of subsets). It is clear from the analysis that the datasets are too small to develop statistically valid QSARs especially as outliers cannot be identified or removed. In addition, although the five subsets of data were, as far as possible, collated based on consistent data, there is still variability in both the IAM and EDETOX data, due to subtle variations in methods, as these are not consistently reported.

6.4.1.3 Combined consistent experimental and database log k_{IAM} (pH 7.4) values

Combining these small datasets (whilst maintaining consistent experimental procedures) was not investigated. Although the size of the dataset could potentially increase (which is beneficial), this increase is small (less than 5) and would be outweighed by the increase in variability in experimental procedure for both log k_{IAM} and skin absorption determinations.

6.4.2 Flynn and Moss datasets

6.4.2.1 Experimental log $k_{IAM\ (pH\ 7.4)}$ values

Of the 66 compounds (includes 8 compounds ionised at pH 7.4) for which log $k_{IAM\ (pH\ 7.4)}$ has been determined experimentally (Table 22), the Flynn dataset contains K_p values for 11 compounds (including 2 compounds ionised at pH 7.4) and the Moss dataset contains K_p values for 19 compounds (including 3 compounds ionised at pH 7.4) (refer to Appendixes 1.5, table 29).

Regression analysis was applied to both the Flynn and Moss datasets with either log $k_{IAM\ (pH\ 7.4)}$, or log P_{exp} as descriptors for hydrophobicity and molecular weight as a descriptor for molecular size. Regression analysis produced the QSARs reported in Table 40.

For the Flynn dataset, log $k_{IAM\ (pH\ 7.4)}$ as a descriptor in a two descriptor QSAR does not result a valid QSAR (equation (6.5)), this is a chance relationship demonstrated by the low F value which is less than the relevant F_α value. Additionally all one descriptor QSARs (using log P or log $k_{IAM\ (pH\ 7.4)}$) are invalid due to the low F value compared to the relevant F_α value (equations (6.9) and (6.10)).

For the Moss datasets the use of log $k_{IAM\ (pH\ 7.4)}$ is not an improvement over the use of log P, as a descriptor of hydrophobicity for either the one or two descriptor QSARs. Whilst both one descriptor QSARs developed from the Moss dataset are valid, the resulting QSARs are statistically similar to the two descriptor QSARs.

Neutral and ionised compounds	Flynn		Moss	
	Log $k_{IAM\ (pH\ 7.4)}$	Log P	Log $k_{IAM\ (pH\ 7.4)}$	Log P
Two descriptor Hydrophobicity and molecular weight	Log K_P = -0.32 + 1.13 log $k_{IAM\ (pH\ 7.4)}$ – 0.0227 MW (6.5) n = 11 r^2_{adj} = 0.57, s = 0.631, F = 7.7, $F_{2,8}$ α, 0.001 = 18.5	Log K_P = -1.31 + 1.39 log P – 0.0285 MW (6.6) n = 11 r^2_{adj} = 0.86, s = 0.354, F = 33, $F_{2,8}$ α, 0.001 = 18.5	Log K_P = -3.26 + 0.420 log $k_{IAM\ (pH\ 7.4)}$ + 0.00907 MW (6.7) n = 19 r^2_{adj} = 0.70, s = 0.407, F = 22, $F_{2,16}$ α, 0.001 = 11.0	Log K_P = -3.00 + 0.693 log P – 0.0202 MW (6.8) n = 19 r^2_{adj} = 0.94, s = 0.189, F = 130, $F_{2,16}$ α, 0.001 = 11.0
	Invalid QSAR - F value is less than the relevant F_α value			
One descriptor Hydrophobicity	Log K_P = -2.35 + 0.65 log $k_{IAM\ (pH\ 7.4)}$ (6.9) n = 11 r^2_{adj} = 0.45, s = 0.713, F = 9.3, $F_{1,9}$ α, 0.001 = 22.8	Log K_P = -3. 02 + 0.767 log P (6.10) n = 11 r^2_{adj} = 0.61, s = 0.603, F = 16.6, $F_{1,9}$ α, 0.001 = 22.8	Log K_P = -2.44 + 0.592 log $k_{IAM\ (pH\ 7.4)}$ (6.11) n = 19 r^2_{adj} = 0.66, s = 0.434, F = 36.3, $F_{1,17}$ α, 0.001 = 15.7	Log K_P = -3.14 + 0.653 log P (6.12) n = 19 r^2_{adj} = 0.94, s = 0.187, F = 268, $F_{1,17}$ α, 0.001 = 15.7
	Invalid QSAR - F value is less than the relevant F_α value			

Table 40 - QSARs developed to predict skin absorption using experimental log $k_{IAM\ (pH\ 7.4)}$ values

Neutral compounds	Flynn		Moss	
	Log $k_{IAM\ (pH\ 7.4)}$	Log P	Log $k_{IAM\ (pH\ 7.4)}$	Log P
Two descriptor Hydrophobicity and molecular weight	Log K_P = 1.75 + 1.80 log $k_{IAM\ (pH\ 7.4)}$ – 0.0493 MW (6.13) n = 9 r^2_{adj} = 0.92, s = 0.302, F = 45.3, $F_{2,6}$ α, 0.001 = 27.0	Log K_P = -1.31 + 1.37 log P – 0.0281 MW (6.14) n = 9 r^2_{adj} = 0.87, s = 0.387, F = 26.6, $F_{2,6}$ α, 0.001 = 27.0	Log K_P = -1.72 + 0.909 log $k_{IAM\ (pH\ 7.4)}$ – 0.0101 MW (6.15) n = 16 r^2_{adj} = 0.95, s = 0.180, F = 133, $F_{2,13}$ α, 0.001 = 12.3	Log K_P = -2.93 + 0.698 log P – 0.00286 MW (6.16) n = 16 r^2_{adj} = 0.96, s = 0.149, F = 199, $F_{2,13}$ α, 0.001 = 12.3
		Invalid QSAR - F value is less than the relevant F_α value		
One descriptor Hydrophobicity	Log K_P = -2.38 + 0.661 log $k_{IAM\ (pH\ 7.4)}$ (6.17) n = 9 r^2_{adj} = 0.46, s = 0.773, F = 7.8, $F_{1,7}$ α, 0.001 = 29.3	Log K_P = -3.13 + 0.767 log P (6.18) n = 9 r^2_{adj} = 0.62, s = 0.653, F = 13.9, $F_{1,7}$ α, 0.001 = 29.3	Log K_P = -2.57 + 0.680 log $k_{IAM\ (pH\ 7.4)}$ (6.19) n = 16 r^2_{adj} = 0.92, s = 0.227, F = 163, $F_{1,14}$ α, 0.001 = 17.8	Log K_P = -3.13 + 0.643 log P (6.20) n = 16 r^2_{adj} = 0.96, s = 0.152, F = 380, $F_{1,14}$ α, 0.001 = 17.8
	Invalid QSAR - F value is less than the relevant F_α value			

Table 41 - QSARs developed to predict skin absorption for neutral compounds using experimental log $k_{IAM\ (pH\ 7.4)}$ values

It is noted that two compounds in the Flynn dataset (hexanoic acid and pentanoic acid) and three compounds in the Moss dataset (hexanoic acid, pentanoic acid and benzoic acid) are 99% ionised under the conditions of log $k_{IAM\ (pH\ 7.4)}$ determination. However, at the generally accepted pH of the stratum corneum surface, pH of 5.5[82] (with measured values ranging from pH 4.2 to 5.9[83]), the degree of ionisation is between 80 and 95% for these compounds. The potential of these compounds being outliers (due to the effect of ionisation) was therefore investigated.

For the Flynn dataset hexanoic acid has a large residual error and for the Moss dataset benzoic acid has a large residual error. Additionally the 3-D scatter plot (Figure 38) indicates that hexanoic acid and benzoic acid (circled) are potentially outliers for the Moss dataset. Therefore, all ionised compounds were removed from the datasets and the regression analysis repeated, the QSARs generated are reported in Table 41.

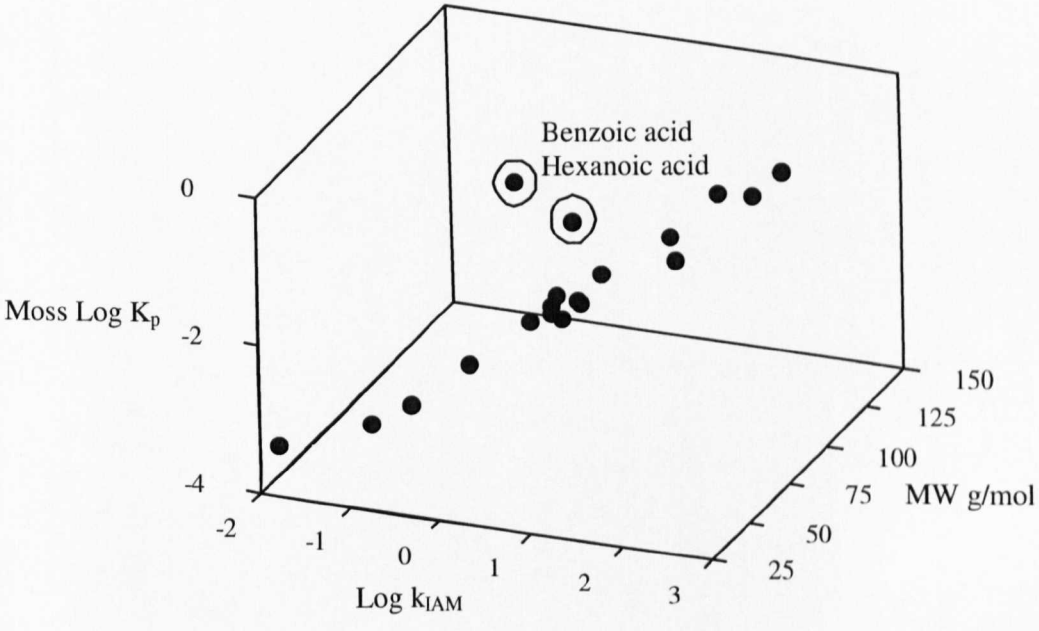


Figure 38 - Plot of Moss log K_P data against log k_{IAM} and molecular weight

For the Flynn dataset QSARs based on a single descriptor of hydrophobicity remain invalid (equations (6.17) and (6.18)). Considering the two descriptor QSARs, log P resulted in a valid QSAR, with the inclusion of ionised compounds (equation (6.6)), however, when the ionised compounds are removed the equivalent QSAR is invalid

(equation (6.14)). For log $k_{IAM\ (pH\ 7.4)}$ the QSAR resulting from the removal of ionised compounds ($r^2_{(adj)}$ =0.92) (equation (6.13)), is a considerable improvement from the previously invalid QSAR (equation (6.5)).

For the Moss dataset both one and two descriptor QSARs (for both descriptors of hydrophobicity) are valid. The two descriptor QSAR using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor considerably improves (increased $r^2_{(adj)}$ values and F values) following the removal of the ionised compounds (equation (6.15)). Upon removal of ionised compounds, both two descriptor QSARs (using log P or log $k_{IAM\ (pH\ 7.4)}$ as descriptors) are strong and comparable QSARs (equations (6.15) and (6.16)).

When developing QSAR models it is important to use the highest quality data available to train the model. The Moss dataset contains updated and corrected skin absorption values for seven of the nine compounds considered here. It is therefore, reasonable to place more credibility in the QSARs developed using this dataset.

For the Moss dataset considered here, one descriptor QSARs are stronger than the two descriptor equivalent QSARs. This observation is based on the relative simplicity of models using a single descriptor and the overall improved statistics (comparable $r^2_{(adj)}$, slight increase in s and considerable increase in F value). It has been demonstrated that when the speciation of a compound is taken into account the use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor to predict Kp is a genuine alternative to the use of log P. Given the pH of the skin, and therefore, the expected conditions of exposure (i.e. a pH of 5.5[82, 83]), QSARs developed using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor that require the speciation of the compound to be considered are more relevant than QSARs using log P without the requirement for the speciation of the compound to be taken into account.

The relationship reported by Moss[62] for the full dataset of 116 compounds, equation (6.4) has a far lower $r^2_{(adj)}$ value (0.82, compared to 0.96), higher s value (0.42 compared to 0.15) and a higher F value (266 compared to 199), compared to both comparable two descriptor QSARs developed here (equations (6.15) and (6.16)). This indicates that additional compounds included in the complete Moss QSAR are potentially compounds which are more difficult to accurately predict.

6.4.2.2 Database log k_{IAM} values

Of the compounds included in the Moss[62] and Flynn[9] papers 51 compounds are represented in the log k_{IAM} database (refer to Chapter 2). A total of 259 log k_{IAM} values are available for these compounds. The data were split into subsets based on consistent log k_{IAM} experimental procedure (overview reported in Table 42, full datasets in Appendix 1.5 Tables 30-35).

Dataset	IAM column	pH	Mobile phase	No. of compounds with K_p data	
				Moss[62]	**Flynn[9]**
6	IAM.PC.DD2	7.0	Phosphate buffer with organic modifier as required	27	17
7	IAM.PC.MG			13	6
8	IAM.PC.DD			11	5
9	IAM.PC.DD2	7.4		8	5
10	IAM.PC.MG			8	6
11	IAM.PC.DD			6	5

Table 42 - Summary of experimental conditions for datasets of IAM data for analysis against the Moss and Flynn skin penetration data

Regression analysis was performed on each dataset independently, with the Moss and Flynn data considered separately. Log k_{IAM} and log P were the first descriptor in all QSARs, molecular weight was an additional descriptor where the number of compounds in the dataset was sufficient to statistically allow the use of two descriptors ($n \geq 10$). The QSARs generated are reported in Table 43.

For datasets 7, 9 and 11 both the Moss and the Flynn data resulted in invalid QSARs for both the one descriptor and two descriptor QSARs (using either log k_{IAM} or log P and molecular weight (where dataset size allowed)). This was demonstrated by the low F coefficient being less than the relevant F_α value. The failure of these datasets to produce valid models may be a result of their small size and a higher proportion of these compounds being ionised under the conditions of analysis.

Three datasets produced valid QSARs. For dataset 6 all two descriptor QSARs for both the Moss and Flynn data are valid. For the Flynn dataset the use of log $k_{IAM\ (pH\ 7.4)}$ or log P as a descriptor produces equivalent QSARs (equations (6.23) and (6.24)). However, for the Moss data, the two descriptor QSAR using log P, equation (6.22), is more significant than the comparable log $k_{IAM\ (pH\ 7.0)}$ QSAR, equation (6.21).

For dataset 8 only one and two descriptor QSARs using log P as a descriptor are valid (equations (6.25) and (6.26)). However the one descriptor QSAR is significantly more significant than the two descriptor QSAR based on the lower F value and higher s value. This indicates that for this dataset the additional descriptor is not required.

For dataset 10 only one descriptor QSARs were developed (due to the dataset size), only one valid QSAR (equation (6.27)) was developed, this was for the Moss dataset using log P.

Given that the experimental procedure for the determination of log $k_{IAM\ (pH\ 7.4)}$ in this thesis and that used to determine log $k_{IAM\ (pH\ 7.4)}$ values for subset 9, are similar these dataset were combined (subset 12, refer to Appendix 1.5 Table 36) and the analysis repeated.

<table>
<tr><th rowspan="2">Dataset</th><th rowspan="2">Data source</th><th colspan="2">One descriptor QSAR</th><th colspan="2">Two descriptor QSAR</th></tr>
<tr><th>Log $k_{IAM\ (pH\ 7.4)}$</th><th>Log P</th><th>Log $k_{IAM\ (pH\ 7.4)}$</th><th>Log P</th></tr>
<tr><td rowspan="2">6</td><td>Moss</td><td colspan="2" rowspan="2">All QSARs invalid - F value is less than the relevant F_{α} value</td><td>Log $K_{P\ (Moss)}$ = -1.59 + 0.648 log $k_{IAM\ (pH\ 7.0)}$ – 0.00863 MW (6.21)
n = 27 r^2_{adj} = 0.582, s = 0.711,
F = 19, $F_{2,24}$ α, 0.001 = 9.3</td><td>Log $K_{P\ (Moss)}$ = -2.11 + 0.435 log P – 0.00693 MW (6.22)
n = 27 r^2_{adj} = 0.705, s = 0.597,
F = 32, $F_{2,24}$ α, 0.001 = 9.3</td></tr>
<tr><td>Flynn</td><td>Log $K_{P\ (Flynn)}$ = -0.977 + 1.12 log $k_{IAM\ (pH\ 7.0)}$ – 0.0162 MW (6.23)
n = 17 r^2_{adj} = 0.786, s = 0.728,
F = 30, $F_{2,14}$ α, 0.001 = 11.8</td><td>Log $K_{P\ (Flynn)}$ = -1.61 + 0.695 log P – 0.0134 MW (6.24)
n = 17 r^2_{adj} = 0.774, s = 0.748,
F = 28, $F_{2,14}$ α, 0.001 = 11.8</td></tr>
<tr><td rowspan="2">7</td><td>Moss</td><td colspan="2" rowspan="2">All QSARs invalid - F value is less than the relevant F_{α} value</td><td colspan="2">All QSARs invalid - F value is less than the relevant F_{α} value</td></tr>
<tr><td>Flynn</td><td></td><td></td></tr>
<tr><td rowspan="2">8</td><td>Moss</td><td>QSAR invalid - F value is less than the relevant F_{α} value</td><td>Log $K_{P\ (Moss)}$ = -2.98 + 0.597 log P (6.25)
n = 11 r^2_{adj} = 0.931, s = 0.164,
F = 134, $F_{1,9}$ α, 0.001 = 22.9</td><td>QSAR invalid - F value is less than the relevant F_{α} value</td><td>Log $K_{P\ (Moss)}$ = -3.04 + 0.578 log P – 0.0079 MW (6.26)
n = 11 r^2_{adj} = 0.924, s = 0.172,
F = 61, $F_{2,8}$ α, 0.001 = 18.5</td></tr>
<tr><td>Flynn</td><td colspan="2">All QSARs invalid - F value is less than the relevant F_{α} value</td><td></td><td></td></tr>
</table>

Dataset	Data source	One descriptor QSAR		Two descriptor QSAR	
		Log $k_{IAM\ (pH\ 7.4)}$	Log P	Log $k_{IAM\ (pH\ 7.4)}$	Log P
9	Moss	All QSARs invalid - F value is less than the relevant F_α value			
	Flynn				
10	Moss	QSAR invalid - F value is less than the relevant F_α value	Log $K_{P\ (Moss)}$ = -4.77 + 0.722 log P (6.27) $n = 8$, $r^2_{adj} = 0.861$, $s = 0.374$, $F = 134$, $F_{1,6}\ \alpha, 0.001 = 35.5$		
	Flynn	All QSARs invalid - F value is less than the relevant F_α value			
11	Moss	All QSARs invalid - F value is less than the relevant F_α value			
	Flynn				

Table 43 - QSARs developed to predict skin absorption for compounds with log k_{IAM} values from the database

6.4.2.3 Experimental and literature (subset 9) log k_{IAM} values combined (subset 12)

It was identified when modelling the experimental log $k_{IAM\ (pH\ 7.4)}$ values that ionised compounds were outliers. Therefore compounds in subset 12 were investigated for their degree of ionisation. It was possible to create subsets based on the degree of ionisation as detailed in Table 44. The QSARs generated are reported in Table 45.

	Number of compounds in subset 12		
	Total	Unionised	Ionised
Flynn dataset	16	12	4
Moss dataset	27	19	8

Table 44 - Number of compounds unionised, ionised and total for both the Moss and Flynn datasets for consistent log k_{IAM} conditions (IAM.PC.DD2, pH 7.4 and PBS mobile phase)

Considering all compounds (neutral and ionised combined) one descriptor QSARs for both the Moss and Flynn dataset are invalid due to low F values compared to the relevant Fα values. For the Flynn data both two descriptor QSARs are equivalent (similar $r^2_{(adj)}$, S and F values)(equations (6.30) and (6.31)). For the Moss data the two descriptor QSAR using log P (equation (6.29)), is a stronger QSAR (higher $r^2_{(adj)}$, lower S and higher F value) than the equivalent QSAR using log $k_{IAM\ (pH\ 7.4)}$ (equation (6.28)).

Considering solely the compounds that are ionised at pH 7.4 all QSARs developed were invalid due to low Fα values. This is not surprising for the Flynn dataset, as this contains only 4 compounds, and five compounds are recommended for a one descriptor QSAR.

Considering the neutral compounds for the Flynn dataset, only the two descriptor QSARs are valid, due to low Fα values for the one descriptor QSARs. However, both two descriptor QSARs, equations (6.35) and (6.36) (using log P or log $k_{IAM\ (pH\ 7.4)}$ are comparable and equivalent. For the Moss data both two descriptor QSARs are valid. Both measures of hydrophobicity are strong and comparable, equations (6.33) and (6.34).

Dataset	Data source	One descriptor QSAR		Two descriptor QSAR	
		Log $k_{IAM\ (pH\ 7.4)}$	Log P	Log $k_{IAM\ (pH\ 7.4)}$	Log P
12 (Dataset 9 + experimental)	Moss	All QSARs invalid - F value is less than the relevant F_α value		Log $K_{P\ (Moss)}$ = -1.69 + 0.699 log $k_{IAM\ (pH\ 7.4)}$ – 0.0078 MW (6.28) n = 27 r^2_{adj} = 0.538, s = 0.573 F = 16, $F_{2,24}$ α, 0.001 = 9.34	Log $K_{P\ (Moss)}$ = -2.42 + 0.612 log P – 0.007 MW (6.29) n = 27 r^2_{adj} = 0.758, s = 0.414 F = 41, $F_{2,24}$ α, 0.001 = 9.34
	Flynn			Log $K_{P\ (Flynn)}$ = -1.01 + 0.944 log $k_{IAM\ (pH\ 7.4)}$ – 0.0145 MW (6.30) n = 16 r^2_{adj} = 0.825, s = 0.560 F = 36, $F_{2,13}$ α, 0.001 = 12.31	Log $K_{P\ (Flynn)}$ = -2.15 + 0.813 log P – 0.0117 MW (6.31) n = 16 r^2_{adj} = 0.813, s = 0.580 F = 34, $F_{2,13}$ α, 0.001 = 12.31
12 (Neutral)	Moss	QSAR invalid - F value is less than the relevant F_α value	Log $K_{P\ (Moss)}$ = -3.25 + 0.559 log P (6.32) n = 19 r^2_{adj} = 0.525, s = 0.573, F = 20, $F_{2,17}$ α, 0.001 = 15.72	Log $K_{P\ (Moss)}$ = -1.85 + 0.864 log $k_{IAM\ (pH\ 7.4)}$ – 0.00845 MW (6.33) n = 19 r^2_{adj} = 0.954, s = 0.178 F = 188, $F_{2,16}$ α, 0.001 = 10.97	Log $K_{P\ (Moss)}$ = -2.75 + 0.762 log P – 0.0064 MW (6.34) n = 19 r^2_{adj} = 0.959, s = 0.168 F = 213, $F_{2,16}$ α, 0.001 = 10.97
	Flynn	All QSARs invalid - F value is less than the relevant F_α value		Log $K_{P\ (Flynn)}$ = -1.09 + 0.991 log $k_{IAM\ (pH\ 7.4)}$ – 0.0146 MW (6.35) n = 12 r^2_{adj} = 0.834, s = 0.592, F = 29, $F_{2,9}$ α, 0.001 = 16.39	Log $K_{P\ (Flynn)}$ = -2.30 + 1.05 log P – 0.0125 MW (6.36) n = 12 r^2_{adj} = 0.897, s = 0.465, F = 49, $F_{2,9}$ α, 0.001 = 16.39
12 (Ionised)	Moss	All QSARs invalid - F value is less than the relevant F_α value			
	Flynn				

Table 45 - QSARs developed to predict skin absorption for compounds with log k_{IAM} values for dataset 12

The QSARs developed for the unionised compounds are more significant than those developed for the combined dataset, this is evidenced by the considerably larger F value obtained.

6.5 Conclusions

Due to the large variability in experimental conditions used for obtaining log k_{IAM} values in the database (refer to Chapter 2) and the variability in experimental methods for determining K_P values, the resulting datasets of comparable data are small. However, for the Moss skin absorption data, modelled using log $k_{IAM\ (pH\ 7.4)}$ (determined in this thesis), a good QSAR using two descriptors has been developed, equation (6.15), with an $r^2_{(adj)}$ value of 0.95 and is equivalent to the two descriptor QSAR using log P.

The EDETOX database contains a large number of K_P values covering a wider range of chemicals than either the Moss, or Flynn datasets. However, the range of experimental variability these results were obtained under is also greatly increased. This means subsets of compounds with consistent Kp and log k_{IAM} procedures are small. The small subsets of consistent data mean that many of the QSARs developed are invalid (due to the F value being less than the relevant Fα value), additionally the identification of outliers is difficult and their removal from analysis has high leverage on the subsequent QSARs developed.

Where the subsets of data are large enough to allow the development of two descriptor QSARs, the two descriptor QSAR is an improvement over the equivalent one descriptor QSAR for log $k_{IAM\ (pH\ 7.4)}$ (Table 41). This supports the requirement of a descriptor of molecular size with the use of log k_{IAM} to predict K_p.

For datasets which produce valid QSARs, it has been demonstrated that on the whole, the use of log k_{IAM}, or the use of log P produce equivalent QSARs. The exceptions are subset 8 and 10 where log k_{IAM} produce invalid QSARs whereas log P produces good QSARs. A potential reason for this could be due to variability in procedure for the determination of log k_{IAM}. Despite the procedure being standardised as far as possible within a subset some variability still exists, due to

published literature not fully recording methods. This supports the requirement for consistent standardised methodology for the determination of log k_{IAM} values.

For subset 12 (combined consistent dataset of experimental and database log $k_{IAM\ (pH\ 7.4)}$ values) the two descriptor QSARS developed are more significant than the equivalent one descriptor QSARs, additionally the use of log $k_{IAM\ (pH\ 7.4)}$ is equivalent to the use of log P (equations (6.33) to (6.36)).

It was also identified that removal of the ionised compounds produces more significant QSARs with considerably higher F values (QSAR developed for neutral compounds). No valid QSARs were developed for ionised compounds due to the small number of compounds within both the Moss and Flynn datasets. It should be noted that although the Moss dataset for ionised compounds was large enough to allow models using one descriptor, compounds analysed in the thesis were mono- or di- functional whereas the compounds from the database are poly-functional, which given the small number of compounds considered may contribute to the validity of QSARs developed.

Log k_{IAM} is an alternative descriptor to log P in the development of QSARs for the prediction of skin absorption. Although QSARs using log P initially appear stronger, when ionisation is taken into account the QSARs using log P do not alter. However, for QSARs using log k_{IAM} as a descriptor, when the speciation of the compound is taken into account the QSARs become significantly stronger and equivalent to the QARS using log P. Given the biological pH of the skin[82, 83], the speciation of compounds is important, and is a key factor affecting skin absorption of compounds.

It is recommended that the degree of ionisation of compounds needs to be considered when developing QSARs to predict the absorption of chemicals across human skin. It is, therefore, also recommended that the determination of log k_{IAM} and/or log P is performed at pH 5.5 for compounds that are partially ionised at either pH 5.5 (pH of skin), or 7.4 (pH of IAM-HPLC analysis) before the development of QSARs to predict skin absorption is undertaken.

6.6 References

[1] Cohen D.E., Rice R.H. (2001) Toxic Responses of the Skin In Klassen C.D. ed, *Casarett & Doull's Toxicology The Basic Science of Poisons*, 6th edition, New York, McGraw Hill Medical, 653-671.

[2] Tortora G.J., Anagnostakos N.P., Grabowski S.R. (1993) *Principles of Anatomy and Physiology*, New York, Harper Colins, pp 126-130.

[3] Kielhorn J., Melching-Kollmuβ S., Mangelsdorf I. (2006) *Environmental Health Criteria 235 Dermal Absorption*, Geneva, World Health Organisation, pp 32-37.

[4] Hughes M.F., Fisher H.L., Birnbaum L.S., Hall L.L. (1994) Effect of Age on the *In Vitro* Percutaneous Absorption of Phenols in Mice *Toxic. In Vitro* 8: 221-227.

[5] Chen L., Lain G., Han L. (2008) Use of "Bricks and Mortar" Model to Predict Transdermal Permeation: Model Development and Initial Validation *Ind. Eng. Chem. Res.* 47: 6465-6472.

[6] Potts R.O., Guy R.H (1992) Predicting Skin Permeability *Pharm. Res.* 9: 663-669.

[7] Barratt M.D. (1995) Quantitative Structure-Activity Relationships for Skin Permeability *Toxic. In Vitro* 9: 27-37.

[8] World Health Organisation (WHO) (2005) Dermal Absorption. WHO/IPCS Environmental Health Criteria, available from <http://www.who.int/ipcs/methods/dermal_absorption/en/> [Accessed 13th December 2011]

[9] Flynn G. (1990) Physicochemical Determinants of Skin Absorption. In Gerrity T.R., Henry C.J. eds., *Principles of Route-to-Route Extrapolation for Risk Assessment*, New York, Elsevier, pp.93-127.

[10] Barbato F., Cappello B., Miro A., La Rotonda M.I., Quaglia F. (1998) Chromatographic Indices on Immobilized Artificial Membranes for the Prediction of Transdermal Transport of Drugs *Il Farmaco* 53: 655-661.

[11] Alvarez F.M., Bottom C.B., Chikhale P., Pidgeon C. (1993) Immobilised Artificial Membrane Chromatography. Prediction of Drug Transport across Biological Barriers In Ngo NT, eds. *Molecular Interactions in Bioseparations*, New York, Plenum Press, pp 151-167.

[12] Tortora G.J., Anagnostakos N.P., Grabowski S.R. (1993) *Principles of Anatomy and Physiology*, New York, Harper Colins, pp 128.

[13] Moss G.P., Cronin M.T.D. (2002) Quantitative Structure-Permeability Relationships for Percutaneous Absorption: Re-analysis of Steroid Data *Int. J. Pharm.* 238: 105-109.

[14] Nasal A., Sznitowska M., Buciński A., Kaliszan R. (1995) Hydrophobicity Parameter from High-Performance Liquid Chromatography on an Immobilized Artificial Membrane Column and its Relationship to Bioactivity *J. Chromatogr. A* 692: 83-89.

[15]The Organisation for Economic Cooperation and Development OECD (2004) *OECD Guidelines for the Testing of Chemicals, No. 427: Skin Absorption: in vivo Method*, Paris, Organisation for Economic Cooperation and Development.

[16] The Organisation for Economic Cooperation and Development OECD (2004) *OECD Guidelines for the Testing of Chemicals, No. 428: Skin Absorption: in vitro Method*, Paris, Organisation for Economic Cooperation and Development.

[17] Lian G., Chen L., Han C. (2008) An Evaluation of Mathematical Models for Predicting Skin Permeability *J. Pharm. Sci.* 97: 584-598.

[18] Blank I.H., McAuliffe B.S. (1985) Penetration of Benzene Through Human Skin *J. Invest. Dermatol.* 85: 522-526.

[19] Kielhorn J., Melching-Kollmuβ S., Mangelsdorf I. (2006) *Environmental Health Criteria 235 Dermal Absorption*, Geneva, World Health Organisation, pp 38-59.
[20] Guy R.H., Hadgraft J. (1998) Physicochemical Aspects of Percutaneous Penetration and Its Enhancement *Pharm. Res.* 5: 753-758.
[21] Thomas J., Majumdar S., Wasdo S., Majumdar A., Sloan K.B. (2007) The Effect of Water Solubility of Solutes on Their Flux Through Human Skin *In Vitro*: An Extended Flynn Database Fitted to the Roberts-Sloan Equation *Int. J. Phram.* 339: 157-167.
[22] Majumder S., Thomas J., Wasdo S.C., Sloan S.E., Roberts M.S. (2004) Molecular Size as the Main Determinant of Solute Maximum Flux Across the Skin *J. Invest. Dermatol.* 122: 993-999.
[23] Cordero J.A., Alarcon L., Escrobano E., Obach R., Domenech J. (1997) A Comparative Study of the Transdermal Penetration of a Series of Nonsteroidal Antiinflammatory Drugs *J. Pharm. Sci.* 86: 503-508.
[24] Dal Pozzo A., Donzelli G., Liggeri E., Roderiguez L. (1991) Percutaneous Absorption of Nicotinic Acid Derivatives *in vitro*. *J. Pharm. Sci.* 80: 54-57.
[25] Dal Pozzo A., Pastori N. (1996) Percutaneous Absorption of Parabens from Cosmetic Formulations *Int. J. Cosmet. Sci.* 18: 57-66.
[26] Goosen C., Laing T.J., du Plessis J., Goosen T.C. Lu G.W., Glynn G.L. (2002) Percutaneous Delivery of Thalidomide and its N-alkyl Analogs *Pharm. Res.* 19: 434-439.
[27] Gyurosiova L., Laitinen L., Raiman J., Cizmarik J., Sedlarova E., Hirvonin J. (2002) Permeability Profiles of M-alkoxysubstituted Pryyolidinoethyl Esters of Phenylcarbamic Acid Across Caco-2 Monolayers and Human Skin *Pharm. Res.* 19: 162-168.
[28] Modamio P., Lastra C.F., Marino E.L (1998) Transdermal Absorption of Celoprolol and Bisoprolol in Human Skin *In Vitro Int. J. Pharm.* 173: 141-148.
[29] Johnson M.E., Blankschtein D., Langer R. (1997) Evaluation of Solute Permeation Through the Stratum Corneum: Lateral Bilayer Diffusion as the Primary Transport Mechanism *J. Phram. Sci.* 86: 1162-1172.
[30] Modamio P., Lastra C.F., Marino E.L (2000) A Comparative *In Vitro* Study of Precutaneous Penetration of β-blockers in Human Skin *Int. J. Pharm.* 194: 249-259.
[31] Morimoto Y., Hatanaka T., Sugibayashi K., Omoya H. (1992) Prediction of Skin Permeability of Drugs: Comparison of Human and Hairless Rat Skin *J. Pharm. Pharmacol.* 44: 634-639.
[32] Roy S.D., Fujiki J., Fleitman J.S. (1993) Permeabilities of Alkyl p-Aminobenzoates Through Living Skin Equivalents and Cadaver Skin *J. Pharm. Sci.* 82: 1266-1268.
[33] Robert W.J., Sloan K.B. (1999) Correlation of Aqueous and Lipid Solubilities with Flux for Prodrugs of 5-Fluorouracil, Theophylline and 6-Mercaptopurine: A Potts-Guy Approach *J. Pharm. Sci.* 88: 515-522.
[34] Abraham M.H., Martins F. (2004) Human Skin Permeation and Partition: General Linear Free-Energy Relationship Analyses *J. Pharm. Sci.* 93: 1508-1523.
[35] Johnson M.E., Blankschtein D., Langer R. (1995) Permeation of Steroids Through Human Skin *J. Pharm. Sci*, 84: 1144-1146.
[36] Potts R.O., Guy R.H. (1995) A Predictive Algorithm for Skin Permeability: The Effects of Molecular Size and Hydrogen Bond Activity *Pharm. Res.* 12: 1628-1633.

[37] Wilschut A., ten Berge W.F., Robinson P.J., McKone T.E. (1995) Estimating Skin Permeation: The Validation of Five Mathematical Skin Permeation Models *Chromosphere* 30: 1275-1296.
[38] Scheuplein R.J. (1965) Mechanism of Percutaneous Adsorption. . Routes of Penetration and the Influence of Solubility *J. Invest. Dermatol.* 45: 334-346.
[39] Blank I.H., McAuliffe B.S. (1985) Penetration of Benzene Through Human Skin *J. Invest. Dermatol.* 85: 522-526.
[40] Singh P., Roberts M.S. (1994) Skin Permeability and Local Tissue Concentration of Nonsterodial Antiinflammatory Drugs After Topical Application *J. Pharm. Exp. Ther.* 268: 144-151.
[41] Abraham M.H., Martins F., Mitchell R.C. (1997) Algorithms for Skin Permeability Using Hydrogen Bond Descriptors: The Problem of Steroids *J. Pharm. Pharmacol.* 49: 858 – 865.
[42] Peck K.D., Ghanem A-H., Higuchi W.I. (1995) The Effect of Temperature on the Permeation of Polar and Ionic Solutes Through Human Epidermal Membrane *J. Pharm. Sci.* 84: 975-982.
[43] Johnson M.E., Mitragotri S., Patel A., Blankschtein D., Langer R. (1996) Synergistic Effects of Chemical Enhancers and Therapeutic Ultrasound on Transdermal Drug Delivery *J. Pharm. Sci.* 85: 670-679.
[44] Mitragotri S. (2000) *In Situ* Determination of Partition and Diffusion Coefficients in the Lipid Bilayers of Stratum Corneum *J. Pharm. Res.* 17: 1026-1029.
[45] Anderson B.D., Higuchi W.I., Rayker P.V. (1998) Heterogeneity Effects on Permeability-Partition Coefficient Relationships in Human Stratum Corneum *Pharm. Res.* 5: 566-573.
[46] Scheuplein R.J., Blank R.J. (1971) Permeability of the Skin *Physiol. Rev.* 51: 702-747.
[47] Bond J.R., Barry B.W. (1998) Limitation of Hairless Mouse Skin as a Model for *In Vitro* Permeation Studies Through Human Skin: Hydration Damage *J. Invest. Dermatol.* 90: 486-489.
[48] Mitragotri S., Edwards D.A., Blankschtein D., Langer S. (1995) A Mechanistic Study of Ultrasonically Enhanced Transdermal Drug Delivery *J. Pharm. Sci.* 84: 697-712.
[49] Roberts M.S., Anderson R.A., Swarbrick J. (1977) Permeability of Human Epidermis to Phenolic Compounds *J. Pharm. Pharmacol.* 26: 677-683.
[50] Roberts M.S. (1976) *Percutaneous Absorption of Phenolic Species* [PhD thesis] University of Sydney.
[51] Anderson B.D., Rayker P.V. (1989) Solute Structure-Permeability Relationships in Human Stratum Corneum *J. Invest. Dermatol.* 93: 280-286.
[52] Degim I.T., Pugh W.J., Hadgraft J. (1998) Skin Permeability Data: Anomalous Results *Int. J. Pharm.* 170: 129-133.
[53] Hirvonen J., Rytting J.H., Paronen P., Urtti A. (1991) Dodecyl-N,N-dimethylamine Acetate and Azone Enhance Drug Penetration Across Human Skin and Rabbit Skin *Pharm. Res.* 8: 933-937.
[54] Hadgraft J., Ridout G. (1987) Development of Model Membranes for Percutaneous Absorption Measurements 1. Isoproply Myristate *Int. J. Pharm.* 39: 148-156.
[55] Watkinson A.C., Brain K.R., Waters K.A. (1993) The Penetration of Ibuprofen Through Human Skin *In Vitro*: Vehicle, Enhancer, and pH Effects. In: Brain K.R.,

James V.J., Walters K.A., eds. *Prediction of Percutaneous Penetration*, SDS Publishing, Cardiff, pp. 335-341.

[56] Anigbogu A.N.C., Williams A.C., Barry B.W. (1996) Permeation Characteristics of 8-Methoxypsoralen Through Human Skin: Relevance to Clinical Treatment *J. Pharm. Pharmacol.* 48: 357-366.

[57] Dearden J.C., Cronin M.T.D., Broohm J.K (2009) A New QSAR Model for Human Skin Permeability *J. Pharm. Pharmacol.* 61: A51-A52.

[58] Netzeva T.I (2004) Whole Molecule and Atom-Based Topological Descriptors. In Cronin M.T.D., Livingstone D.L. eds, *Predicting Chemical Toxicity and Fate*, Florida, CRC Press, pp 61-83.

[59] Scheuplein R.J., Blank I.H., Brauner G.I., MacFarlane D.J. (1969) Precutaneous Absorption of Steroids *J. Invest. Dermatol.* 52: 63-70.

[60] EDETOX database available from <http://edetox.ncl.ac.uk/> [Accessed 4th April 2011]

[61] European Commission, *Off. J. Eur. Un.*, L 396/1 of 30.12.2006 (2006).

[62] Bouwman T., Cronin M.T.D., Bessems J.G.M., van de Sandt J.J.M. (2008) Eurotox Article: Improving the Applicability of (Q)SARs for Percutaneous Penetration in Regulatory Risk Assessment *Hum. Exp. Toxicol.* 27: 269-276.

[63] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.

[64] Minitab Inc. (2007) Minitab® Statistical Software version 15, Coventry: Minitab Inc..

[65] Dearden J.C., Cronin M.T.D., Kaiser K.L.E (2009) How Not to Develop a Quantitative Structure-Activity or Structure-Property Relationship (QSAR/QSPR) *SAR QSAR Environ. Res.* 20: 241-266.

[66] Chan E.C.Y., Tan W.L., Ho P.C., Fang L.J. (2005). Modeling Caco-2 Permeability of Drugs Using Immobilized Artificial Membrane Chromatography and Physicochemical Descriptors *J. Chromatogr. A* 1072: 159-168.

[67] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2005) Can Protonated β-Blockers Interact with Biomembranes Stronger than Neutral Isolipophilic Compounds? A Chromatographic Study on Three different Phospholipid Stationary Phases (IAM-HPLC) *Eur. J. Pharm. Sci.* 25: 379-386.

[68] Barbato F., di Martino G., Grumetto L., La Rotonda M.I. (2004) Prediction of Drug-Membrane Interactions by IAM-HPLC: Effects of Different Phopholipid Stationary Phases on the Partition of Bases *Eur. J. Pharm. Sci.* 22: 261-269.

[69] Barbato F., Cirocco V., Grumetto L., La Rotonda M.I. (2007) Comparison Between Immobilized Artificial Membrane (IAM) HPLC Data and Lipophilicity in n-Octanol for Quinolone Antibacterial Agents *Eur. J. Pharm. Sci.* 31: 288-297.

[70] Barbato F., La Rotonda M.I., Quaglia F. (1997) Chromatographic Indexes on Immobilized Artificial Membranes for Local Anaesthetics: Relationships with Activity Data on Closed Sodium Channels *Pharm. Res.* 14: 1699-1705.

[71] Caldwell G.W., Masucci J.A., Evangelisto M., White R. (1998) Evaluation of the Immobilized Artificial Membrane Phosphatidylcholine Drug Discovery Column for High-Performance Liquid Chromatographic Screening of Drug-Membrane Interactions *J. Chromatogr. A* 800:161-169.

[72] Demare S., Roy D., Legerdre J.Y. (1999) Factors Governing the Retention of Solutes on Chromatographic Immobilized Artificial Membranes: Application to

Anti-Inflammatory and Analgesic Drugs *J. Liq. Chromatogr. Relat. Technol.* 22: 2675-2688.

[73] Genty M., González G., Lere C., Desangle-Gouty V., Legendre J. (2001) Determination of the Passive Absorption Through the Rat Intestine using chromatographic Indices and Molar Volume *Eur. J. Pharm. Sci.* 12: 223-229.

[74] Kaliszan R., Nasal A., Buciński A. (1994) Chromatographic Hydrophobicity Parameter Determined on an Immobilized Artificial Membrane Column: Relationships to Standard Measures of Hydrophobicity and Bioactivity *Eur. J. Med. Chem.* 29: 163-170.

[75] Kangas H., Kotiaho T., Salminen T., Kostiainen R. (2001) N-in-one Determination of Retention Factors for Drugs by Immobilized Artificial Membrane Chromatography Coupled to Atmospheric Pressure Chemical Ionization Mass Spectrometry *Rapid Commun. Mass Spectrom.* 15: 1501-1505.

[76] Kotecha J., Shah S., Rathod I., Subbaiah G. (2008) Prediction of Oral Absorption in Humans by Experimental Immobilized Artificial Membrane Chromatography Indices and Physicochemical Descriptors *Int. J. Pharm.* 360: 96-106.

[77] Lázaro E., Ráfols C., Rosés M. (2005) Characterization of Immobilized Artificial Membrane (IAM) and XTerra Columns by Means of Chromatographic Models *J. Chromatogr. A* 1081: 163-173.

[78] Taillardat-Bertschinger A., Marca Martinet C.A., Carrupt P.A., Reist M., Caron G., Fruttero R., Testa B. (2002) Molecular Factors Influencing Retention on Immobilized Artificial Membranes (IAM) Compared to Partitioning in Liposomes and n-Octanol *Pharm. Res.* 19: 729-737.

[79] Taillardat-Bertschinger A., Galland A., Carrupt P.A., Testa B. (2002) Immobilized Artificial Membrane Liquid Chromatography: Proposed Guideline for Technical Optimization of Retention Measurements *J. Chromatogr. A* 953: 39-53.

[80] Taillardat-Bertschinger A., Barbato F., Quercia M.T., Carrupt P.A., Reist M., La Rotonda M.I., Testa B. (2002) Structural Properties Governing Retention Mechanisms on Immobilized Artificial Membrane (IAM) HPLC Columns *Helv. Chim. Acta.* 85: 519-532.

[81] Vrakas D., Giaginis C., Tsantili-Kakoulidou A. (2008) Electrostatic Interactions and Ionization Effect in Immobilized Artificial Membrane Retention A Comparative Study with Octanol-Water Partitioning. *J. Chromatogr. A* 1187: 67-78.

[82] Ananthanpadmanabhan K.P., Moore D.J., Subramanyan K. (2004) Cleansing Without Compromise: The Impact of Cleansers on the Skin Barrier and the Technology of Mild Cleansing, *Dematologic Therapy* 17: 16-25.

[83] Ehlers C., Ivens U.I., Møller M.L., Senderovitz T., Serup J. (2001) Females have Lower Skin Surface pH than Men *Skin Research and Technology* 7: 90-94.

7 QSAR – Using log k_{IAM} to predict acute aquatic toxicity

7.1 Introduction

The aquatic environment is exposed to a wide range of chemicals from a variety of sources. Chemicals arise from both domestic and commercial use as well as from their disposal. They may enter into various compartments of the environment including water, soil, and the air. Many widely dispersed, commonly used chemicals are disposed of primarily 'down the drain' such as those in home and personal care products. Other common chemicals to enter the environment include agrochemicals (including herbicides and pesticides), pharmaceuticals[1] and industrial chemicals (including by-products of industrial activity). One of the key environmental compartments to consider in safety assessment is the aquatic environment, as safety of chemicals to the environment is assessed across both the aquatic and terrestrial environments[2]. To understand the toxicity of compounds on the environment, toxicity should be considered across all trophic levels of the environmental ecosystem, from primary producers to secondary and top level consumers[3, 4]. There are over 1.5 million taxonomically classified species in the world, and it is not feasible, or necessary to test them all. Therefore, the testing that is performed covers a range of different taxa and trophic levels which are as representative as possible of the environment being assessed. From a regulatory perspective "considered" does not necessarily mean that animal testing is required; the use of read across or *in silico* predictions can be used or alternatively tests can be waived, provided this can be justified.

7.1.1 Regulatory Assays

Under the REACH legislation all chemicals with significant exposure have to be assessed for their environmental toxicity. This testing covers both acute and chronic endpoints[5], although the required level of testing depends on the production/imported tonnage of the chemical being assessed. The 'base set' of tests includes algae, invertebrates and fish[6] and covers both acute and chronic endpoints. Acute toxicity is generally expressed in terms of a concentration of a chemical which is lethal for, or causes an adverse effect to, 50% of the test organisms, or leads to a 50% reduction in test organism compared to the control. These are often reported as LC_{50}, EC_{50} or IC_{50} values. Chronic toxicity tests are longer in duration and may

include lower and/or repeated doses and observation of sublethal effects (such as fecundity, growth rate etc.). Such chronic effects are usually expressed in terms of a No Observed Effect Concentration (NOEC), or an equivalent EC_x. Chronic testing also covers tests involving multiple generations and during sensitive life stages[7].

7.1.1.1 Algae

Algae are representative of primary producers in the aquatic environment and are included as part of the regulatory 'base set' of test species. The OECD guideline supporting this study (Guideline number 201) describes a standardised method for the determination of EC_{50} values and its applicability to a range of standard recognised species e.g. *Pseudokirchneriella subcapitata*[8]. There are a number of effects that can be measured. Commonly the test endpoint considered is the inhibition of growth rate (often 50%) for an exposure period of 72 hours. This is usually recorded as the EC_{50} value.

7.1.1.2 Invertebrates

Toxicity values are reported for a wide range of inveterate species. However, *Daphnia magna* is commonly used as the test species. Commonly the effect observed is immobilisation (EC_{50}) for an exposure period of 48 hours. One standard and commonly followed method of determining EC_{50} values to *Daphnia* is detailed in an OECD guideline 202[9].

7.1.1.3 Vertebrates

The regulatory 'base set' toxicity test for vertebrates is for fish. The endpoint of interest is generally lethality. The OECD guideline number 203[10] describes the 96 hour standard method, in which there are a number of recommended freshwater species including *Danio rerio* (zebra fish), *Poecilia reticulata* (guppy), *Pimephales promelas* (fathead minnow) and *Oncorhynchus mykiss* (rainbow trout)[11].

7.1.2 Known mechanisms of action

Chemicals cause toxicity which is expressed by a mode of action and brought about by a mechanism of action. As discussed in Chapter 1, a mechanism of action describes a known biological, or chemical process by which a toxic effect is produced (e.g. receptor binding, disruption of membrane)[12], whereas a mode of

action relates to the way the endpoint is expressed (e.g. narcosis, hypersensitivity)[13]. Mechanisms of action are more difficult to define and confirm compared to modes of action. However, when a mechanism of action is identified, toxicity tends to be predicted well if a model can be developed for compounds acting by the same mechanism of action. Mode of action is more easily defined than mechanism of action because it is based on direct observations and measurements[14]. Observations include behavioural symptoms i.e. swimming activity, startle response, body movements, body coloration, measurements include respiratory pattern and hemorrhage)[15].

Verhaar (1992)[16] defined a classification scheme to assign compounds into one of four modes of action. These rules have been incorporated into Toxtree (available from http://toxtree.sourceforge.net/) and the OECD QSAR Toolbox (available from http://www.oecd.org/document/54/0,3746,en_2649_34379_42923638_1_1_1_1,00.html).

The four modes of action as defined by Verhaar are outlined below. If a compound cannot be assigned to one of these modes of action, it is classified as class 5 (Not possible to classify according to these rules by Toxtree).

7.1.2.1 **Non-polar narcosis (baseline toxicity, class 1)**

Non-polar narcosis is brought about by accumulation of compounds in membranes, causing general anaesthesia. Non-polar narcotics are inert chemicals i.e. they are unreactive, where the mechanism of action is non-specific. The observed toxicity is driven by the hydrophobicity of the chemical. Hence, there is a strong relationship between toxicity and log P[17]. This mechanism of action is also referred to as 'baseline' toxicity because it is the minimum toxicity a chemical can elicit. The Fish Acute Toxicity Syndrome (FATS) whole animal response observed in rainbow trout for compounds acting by the non-polar narcosis mechanism of action is hypoactivity with an under-reaction to external stimuli, whilst the respiration is rapid and shallow[14]. The effect of non-polar narcosis is reversible when fish are exposed to clean water[18].

7.1.2.2 **Polar narcosis (class 2)**

The polar narcotic mechanism of action is again associated with non-reactive inert chemicals. Their toxicity is increased compared to non-polar narcotics. Polar-narcotics are characterised by a hydrogen bond donor group on an aromatic molecule (i.e. phenols and anilines). The FATS response in rainbow trout for the polar narcotics mechanism is muscular activity associated with seizures followed by cardiovascular-respiratory collapse. Similarly to non-polar narcosis, the effect is reversible when fish are exposed to freshwater[18].

7.1.2.3 **Unspecific reactive chemicals (Class 3)**

Reactive chemicals either react unselectively with common structures of biomolecules, or are metabolised into more reactive species within the organism of interest[16]. Many reactions involve the formation of covalent bonds with the target site, which are irreversible. Effects include membrane irritancy, enzyme inhibition and mutagenicity[19]. Examples of reactive chemicals include electrophiles reacting with nuclophilic sites within biological molecules. Electrophiles include aldehydes, epoxides, α,β unsaturated carbonyl compounds[20] and the nucleophilic sites include those on peptides, proteins and DNA. Epoxides are reactive chemicals and react via alkylation of macromolecules. Unspecific reactive compounds exhibit significantly higher toxicity than predicted from baseline narcosis and additionally toxicity increases with reactivity up to a certain level[21].

7.1.2.4 **Specific mode of action chemicals (Class 4)**

Compounds are classified into class 4 based on specific knowledge about the mechanism of action, for example, compounds that react with specific receptors in a non-covalent manner[22]. DDT, organotin compounds and organophosphates are class 4 compounds. Organotin compounds act through an indirect mechanism which inhibits cytochrome P-450-dependent monooxygenase which oxidises testosterone to estradiol[23]. Organophosphates mechanism of action is the irreversible acetylcholinesterase inhibition at the cholinergic synapses. The accumulation of acetylcholinesterase causes desensitisation of the cholinergic receptor sites due to over stimulation of the cholinergic pathways[24]. Such specific mechanisms can be species specific and result in toxicity elevated significantly above baseline narcosis.

7.1.3 Russom classification scheme for mode of action

An alternative classification scheme developed by Russom *et al.* (1997)[25] contains more categories than the Verhaar scheme. This scheme relates features of chemicals to one of eight modes of action, the classifications include:

Non-polar narcosis, or narcosis I (characterised by depressed activity, underactivity and mortality of the fish during first 24 hrs of exposure) (additive with octanol)

Polar narcosis, or narcosis II (characterised by hyperactivity, overreaction to outside stimuli and delayed mortality) (additive with phenol)

Ester narcosis, or narcosis III (characterised by spontaneous locomotor activity, convulsions, spasms, tetany, scoliosis, lordosis and/or haemorrhaging in the vertebral column)

Oxidative phosphorylation uncoupling (additive with 2,4-dinitrophenol)

Acetylcholinesterase (AChE) inhibition

Respiratory inhibition (additive with cyanide), respiratory inhibitors act through a variety of receptors

Electrophile/proelectrophile reactivity

Several mechanisms of CNS seizure responses, CNS seizure agents act through a variety of receptors.

7.1.4 Alternative methods to determine acute aquatic toxicity

The generation of acute toxicity data is expensive and time consuming. Both of these factors increase for chronic toxicity testing. Additionally, the introduction of REACH[26] requires the implementation of the 3Rs with regards to toxicity testing on animals. The 3Rs are Reduction, Refinement and Replacement of animal use. Reduction includes reducing the total number of animals used in the assay,

refinement includes improving methods i.e using algae and *Daphnia* toxicity data to determine the threshold concentration for acute toxicity testing to fish, determining if fish are more or less sensitive than algae and *Daphnia* as well as more efficient experimental design. The implementation of the threshold concentration approach to determine acute toxicity to fish is detailed in an OECD guideline number 126[27]. The REACH legislation also emphasises the requirement for non-animal test data. As discussed in Chapter 1, accepted non-animal test data include those from category formation, read-across and the use of Quantitative Structure-Activity Relationships (QSARs). Given the expense, and both the high time and animal use costs of performing animals tests the move in emphasis towards non-animal test data is beneficial as the desire to understand the impact of chemicals on the environment continues.

7.1.4.1 QSARs

The REACH[26] legislation specifically includes the use of QSARs as part of the approach to reduce, refine and replace animals in the assessment of a chemicals safety. The European Chemicals Agency (EChA) has published a guidance document on the implementation of non-testing approaches, which includes category formation and the development of QSARs[28], as well as the requirements to validate these methods.

Many QSARs have been developed to predict various acute aquatic toxicity endpoints. In 1995 the European Union evaluated the use of QSARs to predict environmental endpoints and produced a Technical Guidance Document (TGD)[29]. This guidance document has subsequently been updated to reflect and support the requirements of REACH[30]. This includes applying weight of evidence to existing data and the testing requirements if no existing data are available.

The TGD (1995) evaluated 271 QSARs, published between 1980 and 1993 with regard to their quality. Quality considerations included the QSARs being transparent, interpretable and reproducible, outliers needed to be explained, the statistics of the QSARs reported, descriptors needed to be readily available and the ease of use of the QSAR and its interpretation[29]. QSARs covered 19 different endpoints for aquatic toxicity. The majority of the QSARs considered contained only one descriptor (169

of 271 QSARs). The descriptors used in these QSARs can be split into three broad categories (physico-chemical descriptors, structural descriptors and endpoints i.e species-species extrapolation). From the 271 QSARs considered, 382 descriptors were used of which 170 were log P (The TGD grouped similar descriptors together). The TGD developed a strategy for applying QSARs in the prediction of aquatic toxicity. The strategy recommends the application of Ecotox models 1-10 for non-polar narcotics (class 1 compounds) and Ecotox models 11-14 for polar narcotics (class 2 compounds)[29]. Ecotox models 1-14 are reported in Appendix 1.6, Equations A.1.98 to A.1.111.

7.1.4.1.1 Predicting aquatic toxicity using knowledge of the mechanism of action

The observed physiological response in an aquatic organism is different between compounds that elicit toxicity by the non-polar narcosis and polar narcosis mechanisms of action[14,18]. Despite these observed differences, attempts have been made to model these mechanisms together. Abraham and Rafols[31] found polarisability and hydrogen-bond basicity to be important descriptors in modelling tadpole narcosis for non-polar and polar narcotic compounds, equation (7.1). However, this QSAR required the use of four descriptors to model these mechanisms in a single equation, whereas many QSARs modelling polar and non-polar narcotics separately use a single descriptor, which is commonly a term for hydrophobicity[29].

$$\log \frac{1}{[narcosis](mol\ dm^{-3})_{Tadpole}} = 0.579 + 0.842R_2 - 0.334\pi_2^H - 2.871\Sigma\beta_2^\circ + 3.097\,V_X \qquad (7.1)$$

n = 84, r^2 = 0.973. S = 0.246

Where

R_2 is the solute excess molar refraction

π_2^H is the solute dipolarity/polarisability

$\Sigma\beta_2^\circ$ is the hydrogen-bond basicity

V_X is the solute volume

Vaes *et al.*[32] modelled guppy LC_{50} data for non-polar narcotic and polar narcotic compounds with both log P and log membrane (L-α-dimyristoyl phosphatidyl

choline)-water partition coefficient (K_{DMPC}). Using log P as a descriptor there was a difference between the two modes of action for non-polar and polar narcotic compounds and two QSARs are required. However, for log K_{DMPC}, non-polar and polar narcotic compounds can be modelled together. Vaes *et al.* did not report the relationships discussed above as equations, only in graphical and tabular form. It should be noted that the findings of Vaes *et al.*[32] are based on a total dataset of 18 compounds. This consisted of eight non-polar narcotics and ten polar narcotic compounds, which is a relatively small dataset when considering non-polar and polar narcotics separately.

7.1.4.2 Application of IAM-HPLC to QSARs for aquatic toxicity

IAM-HPLC provides a measure of hydrophobicity that has the potential to be more biologically relevant than octanol-water partitioning[33]. Given the most common descriptors in QSARs for aquatic toxicity relate to hydrophobicity, it should be of interest whether log k_{IAM} could improve on correlations using log P. For instance, Ward *et al.*[33] found a good correlation of log k_{IAM} with the 48h LC_{50} toxicity data of a set of homogeneous quaternary alkylammonium sulfobetaine surfactants to *Daphnia magna*. Vaes *et al.*[32] demonstrated a better correlation between partition coefficients measured using L-α-dimyristoylphosphatidylcholine (DMPC)-water (log K_{DMPC}), and guppy LC_{50} toxicity than that obtained using log P. However, the dataset considered class 1 and class 2 compounds together and the number of compounds considered was relatively small. Despite the relative success of using IAM values, it has been used only in a very limited number of situations. Therefore, to evaluate its use more thoroughly, further studies should be undertaken.

7.2 Aim of the chapter

The aim of this chapter was to investigate the use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor in QSARs to predict acute aquatic toxicity. To help achieve this aim for *Daphnia magna*, EC_{50} values were determined experimentally. This increased the number of compounds for which both *Daphnia magna* EC_{50} and log $k_{IAM\ (pH\ 7.4)}$ values were available for modelling. The actual measured concentrations, as opposed to nominal concentrations, of the *Daphnia* test solutions were determined through chemical quantification. QSARs were developed based on literature and experimental toxicity values determined as part of this thesis, using log k_{IAM} as a descriptor. Additionally,

the approach proposed by Vaes *et al.*[32] that non-polar and polar narcotics can be modelled together when phosphatidylcholine membrane/water partitioning is used as a descriptor instead of the more commonly used, log P, was investigated.

7.3 Method

7.3.1 Datasets

The Technical Guidance Document for QSARs predicting the fate and effects of chemicals in the environment (TGD)[29] and the OECD QSAR Toolbox[34] version 2.2.1.1120 (available from http://www.qsartoolbox.org/) were searched for acute aquatic toxicity values. Endpoints of interest included 48hr *Daphnia magna* EC_{50}, 96hr *Pimephales promelas* (fathead minnow) LC_{50}, 96hr and 14 day *Poecilia reticulta* (guppy) LC_{50}, 48hr *Tetrahymena pyriformis* IGC_{50} and 72hr *Pseudokirchneriella subcapitata* EC_{50}. The TGD and QSAR Toolbox were searched for compounds for which log k_{IAM} values have been determined, either reported in the log k_{IAM} database (refer to Chapter 2 for full details), or determined experimentally within this thesis (refer to Chapter 3 and Chapter 4 for log $k_{IAM\ (pH\ 7.4)}$ values). The TGD[27] and OECD QSAR Toolbox[34] data were considered to be high quality due to the application of data quality assessment in their collation.

The values reported for each endpoint considered were converted to log 1/endpoint (mol/L) from the units reported. Where multiple values were reported, averages were taken, provided the values did not cover multiple orders of magnitude. If the range of reported values covered multiple orders of magnitude the compound was excluded from the dataset.

All compounds for which log k_{IAM} values were available were classified by the Verhaar classification scheme according to their mechanism of action using Toxtree version 1.6 (available from http://toxtree.sourceforge.net/download.html).

7.3.2 Experimental determination of *Daphnia magna* EC_{50} values

7.3.2.1 Compound selection

Compounds were chosen to include both well-characterised non-polar narcotic (class 1) and well-characterised polar narcotics (class 2). These were classified according to the Verhaar rules[16] using Toxtree version 1.6. Compounds for which log k_{IAM} had

been determined experimentally (results reported in Chapters 3 and 4), and for which *Daphnia magna* EC_{50} values were not available were identified as possibilities for testing. In addition, the following considerations were also taken into account when considering the suitability of determining *Daphnia magna* EC_{50} values experimentally.

- Log P within range of log 0 to 5
- Non-volatile
- Stable to water and light (if possible)
- To include non-polar and polar narcotic compounds
- Commercially available
- Unionised under the testing conditions of the *Daphnia magna* test
- Toxicity less than 1g/L (i.e. values greater than 1g/L)

Substances were excluded as potential test compounds if the log P was outside the range 0 to 5 (due to solubility and expectations of reduced bioavailability in the environment/test system), the compound was ionised at conditions of analysis (for either IAM-HPLC determination, or EC_{50} *Daphnia magna* determination), or if the predicted toxicity was greater than 1g/L (toxicity greater than 1g/L is unlikely to be of environmental concern and also solubility potentially becomes limiting) according to ECOSAR (a module of EPISuite version 4.1)[35]. 3-Nitroaniline, the external standard used in the optimised IAM-HPLC assay (refer to Chapter 3 for full details), was excluded due to being light-sensitive. The five compounds for which EC_{50} to *Daphnia magna* were determined are detailed in Table 46.

7.3.2.2 *Daphnia magna* EC_{50} toxicity determination

To determine 48hr *Daphnia magna* EC_{50} toxicity values experimentally, OECD guideline number 202[9] was followed. *Daphnia magna* neonates (<24hr old) were exposed to a range of concentrations of each test compound, prepared in Elendt M7 medium (M4 and M7 are water containing set concentrations of salt, a standardised hard water environment for the *Daphnia*; the preparation of M7 is detailed in Appendix 1.6, Table 37), for 48 hours under semi-static conditions. The five concentrations the *Daphnia* are exposed to are determined based on the log P of the

compound and its predicted toxicity (Table 46), due to the requirement for interpolation. The five concentrations cover one order of magnitude and have an interval of 0.25 log units between the concentrations. Using this set up it is expected that the lowest concentration is a NOEC, the top concentration is EC_{100} and EC_{50} is the middle concentration.

At 24 and 48 hours, the number of immobilised *Daphnia* in each test vessel was recorded. At 24 hours the test media were refreshed. The 0hr, 24hr old, 24hr new and 48hr test media were retained for subsequent quantification of the test compound concentration. A 100 mL sample of each test solution was taken at 0, 24 (old and new) and 48 hrs and preserved with 3% formalin. These samples were refrigerated until the analysis was performed. Table 47 provides an overview of the method followed to determine *Daphnia magna* EC_{50}.

Compound	Experimental log P[35]	Molecular Mass	*Daphnia pulex.* 96hr LC_{50}[36] (mg/L)	Predicted LC_{50} using equation A.1.101 and A.1.111[29] *Daphnia magna* LC_{50} (mg/L)	Verhaar Classification[16]	Solubility[35] (mg/L)	Log $k_{IAM\ (pH\ 7.4)}$	Predicted *Daphnia magna* EC_{50} (mg/L) ECOSAR[35]
Pentanol	1.51	88.15	342.9	160	Non-polar narcotic	22000	0.245	134
1,2,4-Trichlorobenzene	4.00	181.45	2.23	1.50	Non-polar narcotic	49	2.97	2.12
Phenol	1.46	94.11		23.6	Polar narcotic	82800	0.647	9.30
Biphenyl-4-ol	3.20	170.21		4.62	Polar narcotic	56.2	2.77	1.68
Methylbenzoate	2.12	136.15		65.9	Not possible to classify according to these rules	2100	1.36	48.8

Equation A.1.101 was used for non-polar narcotics (Methylbenzoate, although not a non-polar narcotic was included to determine a predicted LC_{50} value, based on a baseline prediction), equation A.1.111 was used for polar narcotics

Table 46 – Compounds selected for the determination of EC_{50} toxicity to *Daphnia magna*, with the log P value from KOWWIN[35]experimental *Daphnia pulex* 96hr LC_{50} and predicted *Daphnia magna* LC_{50} values[29],Verhaar classification from Toxtree, experimental solubility from KOWWIN[35], experimental log $k_{IAM\ (pH\ 7.4)}$ and the predicted EC_{50} value in mg/L from ECOSAR[35].

Test media were prepared at different concentrations for each test material as detailed in Table 48. Dilutions of the stock solutions were prepared using Elendt M7 at the concentrations specified in Table 48. Each exposure vessel (120ml glass crystallising dishes) contained about 100mL test medium and five *Daphnia*. The test media were renewed at 24 hours. Water qualities were determined at specified time points as detailed in Table 49.

Experimental parameter	**Condition**
No. of test concentrations	5 concentrations (4 concentrations covering the expected range and a blank)
No. of test vessels per concentration	4
No. of *Daphnia magna* neonates per test vessel	5
Test duration	48hr
Time points *Daphnia magna* checked for immobilisation	0hr – Test started 24hr – No. of immobile neonates counted and Elendt media changed 48hr – No. of immobile neonates counted
Time points at which Elendt test media was collected and preserved for quantification post *Daphnia magna* analysis	0hr, 24hr old, 24hr new, 48hr

Table 47 - Overview of the *Daphnia magna* EC_{50} toxicity test method

Compound	Nominal stock solution concentration	Stock solution diluent	Nominal test solution concentrations	Test solution diluent
Pentanol	4000mg/L	Elendt M7	0, 56, 100,180, 320, 560 mg/L	Elendt M7
1,2,4-Trichlorobenzene	10mg/L	Millipore Water	0, 1.0, 1.8, 3.2, 5.6, 10 mg/L	Elendt M7
Phenol	500mg/L	Millipore Water	0, 1.8, 3.2, 5.6, 10, 18 mg/L	Elendt M7
Biphenol-4-ol	56mg/L	Millipore Water	0, 0.56, 1.0, 1.8, 3.2, 5.6 mg/L	Elendt M7
Methylbenzoate	1800mg/L	Millipore Water	0, 18, 32, 56, 100, 180 mg/L	Elendt M7

Table 48 - Stock concentration, diluent and test solution concentrations and diluent prepared for each compound analysed for *Daphnia magna* EC_{50} toxicity test

All test vessels were assigned a random number, which related to the vessels test location. The temperature of five test vessels were checked along with the water qualities at each time point as detailed in Table 49.

Time (hrs)	New/old media	pH and dissolved oxygen	Total water hardness
0	New	All concentrations	0.0 & highest
24	Old	0.0 & highest remaining concentration	0.0 & highest remaining concentration
24	New	All concentrations	0.0 & highest
48	old	All concentrations	0.0 & highest remaining concentration

Table 49 - Water quality tests performed at specified time points during *Daphnia magna* EC_{50} toxicity determination

7.3.3 Chemical analysis methods

Test solutions of chemical concentration in Elendt media from 48hr *Daphnia* EC_{50} toxicity analysis were collected at 0hr, 24hr old, 24hr new and 48hr for quantification of the test compound in solution. The method of quantification was different for each test compound, as detailed in Appendix 1.6, Tables 38 - 42. For all methods, unless specified the *Daphnia* Elendt test solutions were filtered prior to

analysis, for headspace GC analysis the samples were used in the assay as stored. For the analysis of biphenyl-4-ol the *Daphnia* Elendt test solutions were diluted to 50% using methanol and filtered prior to analysis.

7.3.4 Determination of *Daphnia magna* acute toxicity

The number of immobilised Daphnia for each concentration (the actual concentrations obtained from chemical analysis of the *Daphnia* test solutions), at both 24 and 48 hours, was entered into BMPDIN[37] (the program generates, where possible, three statistical analyses of the data using the moving average method, the probit method and the binomial method) to determine the EC_{50} values for the compounds analysed. Where BMPDIN is able to calculate more than one value, the order of preference is Probit, moving average, non-linear interpolation. All methods provide a 95% confidence interval for the value; however, due to the accuracy of the methods the size of the interval is smaller for the Probit method. The statistical method used is compound-dependent and is specified in the results.

7.3.5 Statistical analysis

QSARs were developed using linear regression in Minitab[38] (version 15.1.1.0). QSARs were endpoint- and species- specific. Log P and log $k_{IAM\ (pH\ 7.4)}$ were used as the descriptors of hydrophobicity. The following statistical information was recorded: n, $r^2_{(adj)}$, s and F values.

7.4 Results and discussion

7.4.1 Datasets

The TGD[29] contains toxicity data relating to the following endpoints of interest: 48hr LC_{50} *Daphnia*, 96hr LC_{50} fathead minnow and 14 day LC_{50} guppy. Three subsets of data have been collated, these are detailed in Table 50. The number of compounds for each endpoint from the TGD is low, because of the small degree of overlap between toxicity data and log $k_{IAM\ (pH\ 7.4)}$ values.

Subset	Subset endpoint details	Total no. of compounds	No. of compounds per Verhaar classification	
			Non-polar narcosis	Polar narcosis
1	48hr *Daphnia magna* LC_{50} (*Daphnia*)	20	13	7
2	96hr *Pimephales promelas* LC_{50} (fathead minnow)	16	10	5
3	14 day *Poecilia reticulta* LC_{50} (guppy)	12	6	6

Table 50 - Subsets of data from the TGD including endpoint details and the number of compounds considered according to their class

The OECD QSAR Toolbox returned 7073 values, across a range of aquatic toxicity endpoints, for the dataset of 134 compounds (consisting of 66 experimental log $k_{IAM\ (pH\ 7.4)}$ values determined in this study and 70 database log $k_{IAM\ (pH\ 7.4)}$ values determined under consistent, comparable experimental conditions (less the two compounds reported in both sources). Subsets of the data were collated based on consistent species, endpoint and duration of experiment. Where multiple values were obtained for a single compound, an average was taken, unless the values covered multiple orders of magnitude; where this was the case the compound was excluded from analysis. The subsets of data collated for the development of QSARs are detailed in Table 51.

Subset	Subset endpoint details	Total no. of compounds	No. of compounds per Verhaar classification				
			Non-polar narcosis	Polar narcosis	Unspecific reactive	Specific	Not possible to classify according to these rules
4	48hr *Daphnia magna* EC_{50} (*Daphnia*)	32	15	11	1	0	5
5	96hr *Poecilia reticulta* LC_{50} (guppy)	24	15	8	1	0	0
6	96hr *Pimephales promelas* LC_{50} (fathead minnow)	40	21	11	3	0	5
7	48hr *Tetrahymena pyriformis* IGC_{50}	44	16	14	8	0	6
8	72hr *Pseudokirchneriella subcapitata* EC_{50}	17	8	6	0	0	3

Table 51 - Subsets of toxicity data from the OECD QSAR Toolbox, including details of the endpoint and the number of compounds according to their class, for which experimental or literature log $k_{IAM\ (pH\ 7.4)}$ values were available

7.4.2 *Daphnia magna* EC_{50} 48hr toxicity test results

The *Daphnia magna* test vessels were placed on the test square detailed in Appendix 1.6, Tables 43 to 47. The mean water qualities for each compound are detailed in Appendix 1.6, Table 48. The experimental results for the number of *Daphnia magna* immobilised at each time point, as well as the 24hr and 48hr percentage immobile, are detailed in Appendix 1.6, Tables 49 to 53 for the range of concentrations assessed. The quantification of the test solutions was as detailed in Appendix 1.6, Tables 38 to 42. The EC_{50} values are reported later following the quantification of the test solutions.

For 1,2,4-trichlorobenzene the initial analysis was performed using an open system. However, quantification demonstrated significant loss of 1,2,4-trichlorobenzene during the *Daphnia* toxicity test (results not shown), this is possibly due to the high volatility and low solubility of 1,2,4-trichlorobenzene[39]. The analysis was repeated using a closed system and the test solutions quantified following analysis without intermediate storage. A closed system of analysis required the use of screw cap vessels for the analysis.

7.4.3 Quantification of Elendt media from *Daphnia* toxicity testing

Calibration samples, quality control samples and *Daphnia* Elendt solutions were prepared for each compound as detailed in Table 48. The quantification of each compound was performed as detailed in 7.3.3 and Appendix 1.6, Tables 38-42.

The calibration graphs for all compounds are illustrated in Appendix 1.6, Figures 13-17. All calibration graphs show a strong rectilinear trend with the exception of biphenul-4-ol which shows a strong power trend. The $r^2_{(adj)}$ for each compounds are reported in Table 52.

For pentanol the quality control samples are similar in response to the calibration standards. From the calibration graph, actual concentrations were determined as detailed in Appendix 1.6, Table 54.

For 1,2,4-trichlorobenzene the quality control samples are similar in response to the calibration standards. From the calibrations graph actual concentrations were determined as detailed in Appendix 1.6, Table 55. The EC_{50} value for 1,2,4-trichlorobenzene based on measured concentration could not be determined due to problems with the calculation software. However, all concentrations were within ± 20% of nominal concentrations. Therefore, nominal concentrations have been used to calculate the EC_{50} value for this compound using Probit analysis.

For phenol the quality control samples are similar in response to the calibration standards. From the calibration graph, actual concentrations of the phenol solutions used in the *Daphnia* EC_{50} 48hr toxicity testing were determined as detailed in Appendix 1.6, Table 56.

For biphenyl-4-ol the quality control samples are similar in response to the calibration standards. From the equation of the calibration graph, actual concentrations of the biphenyl-4-ol solutions used in the *Daphnia* EC_{50} 48hr toxicity testing were determined as detailed in Appendix 1.6, Table 57.

For methylbenzoate the quality control samples are similar in response to the calibration standards. There is a marked difference in response for samples prepared in Elendt and samples prepared in water. The time weighted average actual concentration, for each nominal concentration was used to determine the 48hr *Daphnia* magna EC_{50} value from the equation of the calibration graph, actual concentrations of the methyl benzoate solutions used in the 48hr *Daphnia* EC_{50} toxicity testing were determined as detailed in Appendix 1.6, Table 58.

The average actual concentration for each nominal concentration for each compound was used to determine the 48hr *Daphnia magna* EC_{50} value (reported in Table 52, along with the statistical method of determination).

7.4.4 Determination of *Daphnia magna* acute toxicity

The EC_{50} values were calculated using the method in Section 7.3.4 and both the quantified actual concentrations for each test solution and the number of immobilised *Daphnia magna* at that concentration (for 1,2,4-trichlorobenzene

nominal concentrations were used as mentioned above). The number of immobilised *Daphnia* was as detailed in Appendix 1.6, Tables 49 to 53, and the actual concentrations of each test solution were as detailed in Appendix 1.6, Tables 54 to 58. The 48hr *Daphnia magna* acute aquatic toxicity values are reported in Table 52.

Compound	$r^2_{(adj)}$	**Method**	**48hr *Daphnia magna* EC_{50} from measured concentrations**	
			(mg/L)[b]	**log (mol/L)**
Phenol	1.00	Probit	8.6 (95% confidence intervals are 7.1 & 11.0)	-4.04
Pentanol	1.00	Non-linear interpolation	350 (interpolation between 291.94 & 464.96)	-2.40
Methylbenzoate	1.00	Non-linear interpolation	49 (interpolation between 52.0 & 99.1)	-3.45
Biphenyl-4-ol	0.99	Probit	4.0 (95% confidence intervals are 3.0 & 4.9)	-4.63
1,2,4-Trichlorobenzene[a]	0.99	Probit	1.7 (95% confidence intervals are 1.45 & 1.95)	-5.03

[a] - Based on nominal concentrations
[b] – Reported to 2 sig. fig.

Table 52 – Compound, method of determination EC_{50} 48hr *Daphnia* (in both mg/L and mol/L) values

7.4.5 QSARS for predicting acute aquatic toxicity using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor

Datasets of toxicity values for compounds for which log $k_{IAM\ (pH\ 7.4)}$ had been determined experimentally were collated for a variety of acute aquatic endpoints. Toxicity data from the TGD (1995)[29] and OECD QSAR Toolbox[34] were kept separate, due to the inclusion of data from the TGD within the OECD QSAR Toolbox. The TGD toxicity data were used as an independent source since the QSARs reported therein were considered high quality. Log $k_{IAM\ (pH\ 7.4)}$ values were

not available for all compounds used to develop the high quality QSARs. Therefore equivalent QSARs, using log P as a descriptor, were developed to allow comparison of the QSARs using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor. An overview of both the TDG[29] and OECD QSAR Toolbox[34] datasets is shown in Table 50 and Table 51. The individual datasets are reported in Appendix 1.6, Tables 59 to 66. QSARs were developed for the following endpoints, 48hr *Daphnia magna* EC_{50}, 96hr *Pimephales promelas* (fathead minnow) LC_{50}, 96hr and 14 day *Poecilia reticulta* (guppy) LC_{50}, 48hr *Tetrahymena pyriformis* IGC_{50} and 72hr *Pseudokirchneriella subcapitata* EC_{50}. The QSARs developed are reported in Table 53.

The development of QSARs was attempted for non-polar narcosis, polar narcosis and general narcosis (non-polar and polar narcotics modelled together) to assess the applicability of Vaes *et al.*[32] findings to the IAM phosphatidylcholine stationary phase (membrane-buffer partitioning). Vaes *et al.*[32] demonstrated, for fathead minnow LC_{50} toxicity values, that there is a distinction between non-polar and polar narcotics, and that compounds need to be separated based on their mechanism of action for modelling purposes when the hydrophobicity parameter considered is octanol-water partitioning. However, the same authors demonstrated that this distinction between mechanisms of action was not required when membrane (L-α-dimyristoyl phosphatidyl choline)-water partitioning is considered as the descriptor for hydrophobicity.

Dataset / Endpoint	Descriptor	Non-polar narcosis	Polar narcosis	General narcosis
(1) *48hr* Daphnia magna EC_{50}	log $k_{IAM\ (pH\ 7.4)}$	Log $1/EC_{50} = 2.21 + 0.980 \log k_{IAM\ (pH\ 7.4)}$ (7.2) $n = 13\ r^2_{adj} = 0.887$, $s = 0.392$, $F = 95$, $F_{1, 11}\ \alpha, 0.001 = 19.7$	Invalid QSAR - F value is less than the relevant F_α value	Log $1/EC_{50} = 2.64 + 0.810 \log k_{IAM\ (pH\ 7.4)}$ (7.3) $n = 20\ r^2_{adj} = 0.752$, $s = 0.474$, $F = 59$, $F_{1, 18}\ \alpha, 0.001 = 15.4$
	log P	Log $1/EC_{50} = 1.15 + 0.942 \log P$ (7.4) $n = 13\ r^2_{adj} = 0.920$, $s = 0.331$ $F = 138$, $F_{1, 11}\ \alpha, 0.001 = 19.7$		Log $1/EC_{50} = 2.07 + 0.714 \log P$ (7.5) $n = 20\ r^2_{adj} = 0.683$, $s = 0.536$, $F = 41$, $F_{1, 18}\ \alpha, 0.001 = 15.4$
(4) *48hr* Daphnia magna EC_{50}	log $k_{IAM\ (pH\ 7.4)}$	Log $1/EC_{50} = 2.21 + 1.01 \log k_{IAM\ (pH\ 7.4)}$ (7.6) $n = 15\ r^2_{adj} = 0.840$, $s = 0.659$, $F = 75$, $F_{1, 13}\ \alpha, 0.001 = 17.8$	Invalid QSAR - F value is less than the relevant F_α value	Log $1/EC_{50} = 2.85 + 0.822 \log k_{IAM\ (pH\ 7.4)}$ (7.7) $n = 26\ r^2_{adj} = 0.618$, $s = 0.813$, $F = 41$, $F_{1, 24}\ \alpha, 0.001 = 14.0$
	log P	Log $1/EC_{50} = 1.20 + 0.970 \log P$ (7.8) $n = 15\ r^2_{adj} = 0.848$ $s = 0.644$, $F = 77$, $F_{1, 13}\ \alpha, 0.001 = 17.8$		Log $1/EC_{50} = 2.28 + 0.736 \log P$ (7.9) $n = 26\ r^2_{adj} = 0.549$, $s = 0.883$, $F = 31$, $F_{1, 24}\ \alpha, 0.001 = 14.0$

Dataset / Endpoint	Descriptor	Non-polar narcosis	Polar narcosis	General narcosis
(4) *48hr* Daphnia magna EC_{50} (aniline removed as a compound with a high degree of leverage)	log $k_{IAM\ (pH\ 7.4)}$	Analysis not performed as removed compound is a polar narcosis,	Invalid QSAR - F value is less than the relevant F_{α} value	$\log 1/EC_{50} = 2.59 + 0.915 \log k_{IAM\ (pH\ 7.4)}$ (7.10) $n = 25\ r^2_{adj} = 0.765$, $s = 0.638$, $F = 79$, $F_{1, 23}\ \alpha, 0.001 = 14.4$
	log P			$\log 1/EC_{50} = 1.91 + 0.839 \log P$ (7.11) $n = 25\ r^2_{adj} = 0.707$, $s = 0.713$, $F = 59$, $F_{1, 23}\ \alpha, 0.001 = 14.4$
(5) *96hr guppy* LC_{50}	log $k_{IAM\ (pH\ 7.4)}$	$\log 1/LC_{50} = 2.09 + 0.886 \log k_{IAM\ (pH\ 7.4)}$ (7.12) $n = 15\ r^2_{adj} = 0.872$, $s = 0.481$, $F = 96$, $F_{1, 13}\ \alpha, 0.001 = 17.8$	Invalid QSAR - F value is less than the relevant F_{α} value	$\log 1/LC_{50} = 2.30 + 0.804 \log k_{IAM\ (pH\ 7.4)}$ (7.13) $n = 23\ r^2_{adj} = 0.800$, $s = 0.529$, $F = 89$, $F_{1, 21}\ \alpha, 0.001 = 14.8$
	log P	$\log 1/LC_{50} = 1.32 + 0.804 \log P$ (7.14) $n = 15\ r^2_{adj} = 0.886$, $s = 0.453$, $F = 110$, $F_{1, 13}\ \alpha, 0.001 = 17.8$		$\log 1/LC_{50} = 1.61 + 0.799 \log P$ (7.15) $n = 23\ r^2_{adj} = 0.771$, $s = 0.566$, $F = 75$, $F_{1, 21}\ \alpha, 0.001 = 14.8$

Dataset / Endpoint	Descriptor	Non-polar narcosis	Polar narcosis	General narcosis
(3) 14 day guppy *LC_{50}*	$\log k_{IAM\ (pH\ 7.4)}$	$\text{Log } 1/LC_{50} = 1.93 + 1.04 \log k_{IAM\ (pH\ 7.4)}$ (7.16) $n = 6, r^2_{adj} = 0.976, s = 0.217,$ $F = 205, F_{1,4}\ \alpha, 0.001 = 47.1$	$\text{Log } 1/LC_{50} = 2.77 + 0.834 \log k_{IAM\ (pH\ 7.4)}$ (7.17) $n = 6, r^2_{adj} = 0.977, s = 0.082,$ $F = 210, F_{1,4}\ \alpha, 0.001 = 47.1$	$\text{Log } 1/LC_{50} = 2.35 + 0.951 \log k_{IAM\ (pH\ 7.4)}$ (7.18) $n = 12, r^2_{adj} = 0.873, s = 0.365,$ $F = 76, F_{1,10}\ \alpha, 0.001 = 21.0$
	log P	$\text{Log } 1/LC_{50} = 1.08 + 0.921 \log P$ (7.19) $n = 6\ r^2_{adj} = 0.952, s = 0.308,$ $F = 99, F_{1,4}\ \alpha, 0.001 = 47.1$	$\text{Log } 1/LC_{50} = 2.20 + 0.829 \log P$ (7.20) $n = 6\ r^2_{adj} = 0.986, s = 0.064,$ $F = 350, F_{1,4}\ \alpha, 0.001 = 47.1$	$\text{Log } 1/LC_{50} = 1.89 + 0.769 \log P$ (7.21) $n = 12\ r^2_{adj} = 0.734, s = 0.528,$ $F = 31, F_{1,10}\ \alpha, 0.001 = 21.0$
(2) *96hr* fathead minnow *LC_{50}*	$\log k_{IAM\ (pH\ 7.4)}$	$\text{Log } 1/LC_{50} = 2.22 + 0.909 \log k_{IAM\ (pH\ 7.4)}$ (7.22) $n = 10, r^2_{adj} = 0.895, s = 0.506,$ $F = 77, F_{1,8}\ \alpha, 0.001 = 25.4$	Invalid QSAR - F value is less than the relevant F_α value	$\text{Log } 1/LC_{50} = 2.35 + 0.937 \log k_{IAM\ (pH\ 7.4)}$ (7.23) $n = 15, r^2_{adj} = 0.891, s = 0.465,$ $F = 89, F_{1,13}\ \alpha, 0.001 = 17.8$
	log P	$\text{Log } 1/LC_{50} = 1.14 + 0.935 \log P$ (7.24) $n = 10\ r^2_{adj} = 0.937, s = 0.391,$ $F = 135, F_{1,8}\ \alpha, 0.001 = 25.4$		$\text{Log } 1/LC_{50} = 1.33 + 0.963 \log P$ (7.25) $n = 15\ r^2_{adj} = 0.870, s = 0.508,$ $F = 93, F_{1,13}\ \alpha, 0.001 = 17.8$

Dataset / Endpoint	Descriptor	Non-polar narcosis	Polar narcosis	General narcosis
(6) *96hr* fathead minnow LC_{50}	log $k_{IAM\ (pH\ 7.4)}$	Log $1/LC_{50} = 2.19 + 0.904 \log k_{IAM\ (pH\ 7.4)}$ (7.26) $n = 21\ r^2_{adj} = 0.916$, $s = 0.450$, $F = 219$, $F_{1, 19}\ \alpha, 0.001 = 15.1$	Invalid QSAR - F value is less than the relevant F_α value	Log $1/LC_{50} = 2.42 + 0.864 \log k_{IAM\ (pH\ 7.4)}$ (7.27) $n = 32$, $r^2_{adj} = 0.829$, $s = 0.556$, $F = 82$, $F_{1, 30}\ \alpha, 0.001 = 13.3$
	log P	Log $1/LC_{50} = 1.83 + 0.683 \log P$ (7.28) $n = 21\ r^2_{adj} = 0.815$, $s = 0.667$, $F = 89$, $F_{1, 19}\ \alpha, 0.001 = 15.1$		Log $1/LC_{50} = 2.06 + 0.679 \log P$ (7.29) $n = 32\ r^2_{adj} = 0.743$, $s = 0.682$, $F = 91$, $F_{1, 30}\ \alpha, 0.001 = 13.3$
(7) 48hr *Tetrahymena pyriformis* IGC_{50}	log $k_{IAM\ (pH\ 7.4)}$	Log $1/ICG_{50} = 1.65 + 0.746 \log k_{IAM\ (pH\ 7.4)}$ (7.30) $n = 16\ r^2_{adj} = 0.958$ $s = 0.255$, $F = 342$, $F_{1, 14}\ \alpha, 0.001 = 17.1$	Log $1/ICG_{50} = 2.49 + 0.637 \log k_{IAM\ (pH\ 7.4)}$ (7.31) $n = 14\ r^2_{adj} = 0.748$, $s = 0.371$, $F = 39$, $F_{1, 12}\ \alpha, 0.001 = 18.6$	Log $1/ICG_{50} = 2.00 + 0.741 \log k_{IAM\ (pH\ 7.4)}$ (7. 32) $n = 30\ r^2_{adj} = 0.880$ $s = 0.490$, $F = 122$, $F_{1, 28}\ \alpha, 0.001 = 13.5$
	log P	Log $1/ICG_{50} = 1.11 + 0.675 \log P$ (7.33) $n = 16\ r^2_{adj} = 0.931$, $s = 0.327$, $F = 202\ F_{1, 14}\ \alpha, 0.001 = 17.1$	Invalid QSAR - F value is less than the relevant F_α value	Log $1/ICG_{50} = 1.44 + 0.690 \log P$ (7.34) $n = 30\ r^2_{adj} = 0.714$, $s = 0.598$, $F = 73\ F_{1, 28}\ \alpha, 0.001 = 13.5$

Dataset / Endpoint	Descriptor	Non-polar narcosis	Polar narcosis	General narcosis
(8) 72hr *Pseudokirchneriella subcapitata* EC_{50}	$\log k_{IAM\ (pH\ 7.4)}$	Invalid QSAR - F value is less than the relevant F_{α} value		$\text{Log } 1/EC_{50} = 2.85 + 0.556 \log k_{IAM\ (pH\ 7.4)}$ (7.35) $n = 14\ r^2_{adj} = 0.793\ s = 0.407,$ $F = 51,\ F_{1,\ 12}\ \alpha,\ 0.001 = 18.6$
	log P			$\text{Log } 1/EC_{50} = 2.47 + 0.504 \log P$ (7.36) $n = 14\ r^2_{adj} = 0.767,\ s = 0.433,$ $F = 44\ F_{1,\ 12}\ \alpha,\ 0.001 = 18.6$

Table 53 – QSARs developed to predict acute aquatic toxicity for the following endpoints: 48hr *Daphnia magna* EC_{50}, 96hr *Pimephales promelas* (fathead minnow) LC_{50}, 96hr and 14 day *Poecilia reticulta* (guppy) LC_{50}, 48hr *Tetrahymena pyriformis* IGC_{50} and 72hr *Pseudokirchneriella subcapitata* EC_{50}

7.4.5.1 QSARs to predict 48hr Daphnia magna EC_{50}

Both the TGD and OECD QSAR Toolbox provide *Daphnia* EC_{50} values for the development of QSARs, dataset 1 (Appendix 1.6, Table 59) and dataset 4 (Appendix 1.6, Table 60) respectively. The two datasets are of similar size (n= 20 and 26) and have a similar distribution for the two mechanisms of action.

Considering the non-polar narcotics, both datasets produce valid strong QSARs, using either log k_{IAM} or log P as a descriptor. For dataset 1, log P is the significantly better QSAR (higher r^2_{adj} and F values and lower s value). Whereas for dataset 4 the QSARs are equivalent (similar s, F and r^2_{adj}) for log k_{IAM} and log P (QSARs are reported in Table 53).

Considering the polar narcotics, neither dataset using either log k_{IAM} or log P produce valid QSARs due to the F values being lower than the relevant Fα value. It is noted when plotting hydrophobicity (either log k_{IAM} or log P) against toxicity (Appendix 1.6, Figure 18 and Figure 19) that the range of both hydrophobicity and toxicity considered for polar narcotic compounds is considerably narrower than for non-polar narcotics.

Considering the QSARs for general narcosis for both datasets the use of log k_{IAM} as a descriptor produces a significantly better QSAR than the use of log P (higher r^2_{adj} and F values and lower s value) (Table 53, equations (7.3), (7.5), (7.7) and (7.9)), however, the QSARs produced for non-polar narcosis is significantly more significant than the general narcosis QSARs. Figure 18 and Figure 19 in Appendix 1.6, clearly illustrate that the distinction between non-polar and polar narcotics is less pronounced using log $k_{IAM\ (pH\ 7.4)}$ as a predictor of toxicity than log P. This supports log k_{IAM} producing the stronger QSAR for general narcosis.

There is a clear outlier in dataset 4 for a polar narcotic (aniline, circled in Appendix 1.6, Figure 19), which has a high degree of leverage on the QSARs developed for polar narcotics. This supports the findings of Ramos *et al.*[40, 41] and Kühn *et al.*[42] that *Daphnia* show increased sensitivity and therefore, excess toxicity to aromatic amines, indicating a specific mechanism of action, however the mechanism involved is not yet known.

Therefore, aniline was removed from dataset 4 and the analysis repeated. The QSARs for polar narcosis remain invalid due to the F value being less than the relevant Fα value. However, the general narcosis QSARs improve considerably for both descriptors of hydrophobicity (Table 53, equations (7.10) and (7.11)). It is noted that the QSARs for non-polar narcosis remain significantly better than those for general narcosis.

7.4.5.2 QSARs to predict 96hr guppy LC_{50}

The OECD QSAR Toolbox provides 96hr guppy LC_{50} values for the development of QSARs, dataset 5 (Appendix 1.6, Table 61). Considering the non-polar narcotics, use of log P produces a slightly statistically stronger QSAR (Table 53, equation (7.12)) (higher r^2_{adj}, lower S and higher F values). Whereas for the polar narcotics neither measure of hydrophobicity produces a valid QSAR, due to the F values being lower than the relevant Fα value, Figure 20 in Appendix 1.6 (plot of hydrophobicity against toxicity) illustrates the narrow range of hydrophobicity and toxicity the polar narcotic compounds cover. Figure 20 in Appendix 1.6 also illustrates the distinction between mechanism of action for both log k_{IAM} and log P is small.

The QSARs using log k_{IAM} for general narcosis is slightly stronger than that using log P (Table 53, equations (7.13) and (7.15) respectively). It is noted that the QSARs for non-polar narcosis is stronger than for general narcosis (higher r^2_{adj}, lower S and higher F values).

7.4.5.3 QSARs to predict 14 day guppy LC_{50}

The TGD provided 14 day guppy LC_{50} values for the development of QSARs, dataset 3 (Appendix 1.6, Table 62). Both measures of hydrophobicity for both individual and general mechanisms of narcosis are valid (Table 53, equations (7.16) to (7.21)). Considering non-polar narcotics use of log k_{IAM} produces a significantly better QSAR (equation (7.16)) (equivalent r^2_{adj}, lower s and higher F values). For polar narcosis the use of either measure of hydrophobicity produces strong equivalent QSARs. QSARs for general narcosis and valid, however, are significantly weaker than those for separate mechanisms of action. This is supported by Appendix

1.6, Figure 21, which illustrates a distinction between the two mechanisms of action for log P and, to a lesser extent for log $k_{IAM\ (pH\ 7.4)}$.

7.4.5.4 QSARs to predict 96hr fathead minnow LC_{50}

The TGD and OECD QSAR Toolbox provided 96hr fathead minnow LC_{50} values for the development of QSARs, dataset 2 and dataset 6 respectively (Appendix 1.6, Table 63 and Table 64). Considering the non-polar narcotics both log k_{IAM} and log P produce valid QSARs, for dataset 2 the use of log P results in a better QSAR, equation (7.24) (higher r^2_{adj}, lower s and higher F values) whereas for dataset 6 the use of log k_{IAM} results in the better QSAR (equation 7.26). For the polar narcotics, for both datasets neither log P, nor log $k_{IAM\ (pH\ 7.4)}$ as descriptors produced valid QSARs. Again the range of hydrophobicities considered for modelling polar narcotics is narrow.

Considering general narcosis, both log k_{IAM} and log P produce valid QSARs, with log k_{IAM} producing the statistically stronger QSARs (equations (7.23) and (7.27)) (higher r^2_{adj}, lower s and equivalent F values) for both datasets. The QSARs for non-polar narcosis are better than the QSARs for general narcosis. This is supported by Appendix 1.6, Figure 22 and Figure 23, which shows the distinction between mechanisms to be far less pronounced for log $k_{IAM\ (pH\ 7.4)}$ than for log P.

7.4.5.5 QSARs to predict 48hr *Tetrahymena pyriformis* IGC_{50}

The OECD QSAR Toolbox provided 48hr *Tetrahymena pyriformis* IGC_{50} values for the development of QSARs, dataset 7 (Appendix 1.6, Table 65). All QSARs for both measures of hydrophobicity for both individual and general mechanisms of narcosis are valid; the exception is log P to model polar narcosis. Considering both non-polar and polar narcosis as separate datasets the use of log k_{IAM} produces a significantly better QSAR (equation (7.30) and (7.32)) (equivalent or higher r^2_{adj}, lower s and higher F values). Figure 24 in Appendix 1.6, shows no significant distinction between the mechanisms of action for either descriptors of hydrophobicity.

The QSAR for non-polar narcosis is more significant than the QSAR for general narcosis. However, the QSAR, for general narcosis is more significant than for polar narcosis.

7.4.5.6 QSARs to predict 72hr *Pseudokirchneriella subcapitata* EC_{50}

The OECD QSAR Toolbox provided 72hr *Pseudokirchneriella subcapitata* EC_{50} values for the development of QSARs, dataset 8 (Appendix 1.6, Table 66). Neither measure of hydrophobicity (log k_{IAM} or log P) produced valid QSARs for the individual mechanisms of action (non-polar and polar narcosis). For all QSARs the F values are smaller than the relevant Fα value, indicating a high probability of a chance correlation, it is noted that the number of compounds considered in the datasets is very small. For general narcosis both log k_{IAM} and log P produce valid QSARs (equations (7.36) and (7.35) respectively), use of log $k_{IAM\ (pH\ 7.4)}$ produces a slightly statistically stronger QSAR. Figure 25 in Appendix 1.6, shows no significant distinction between the mechanisms of action for either descriptors of hydrophobicity. It does illustrate a narrow range of both toxicity and hydrophobicity for polar narcotic compounds.

7.4.5.7 Summary discussion

This chapter has described the determination of actual (test media quantified, rather than nominal concentration) 48hr *Daphnia magna* EC_{50} data relating to five compounds covering a range of mechanisms of action. The use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor to predict aquatic toxicity endpoints has also been investigated. QSARs have been developed using both log P and log $k_{IAM\ (pH\ 7.4)}$ as descriptors to allow comparison of equivalent QSARs. The toxicity endpoints investigated were: 48hr *Daphnia magna* EC_{50}, 96hr *Poecilia reticulta* (guppy) LC_{50}, 14 day *Poecilia reticulta* (guppy) LC_{50}, 96hr *Pimephales promelas* (fathead minnow) LC_{50}, 48hr *Tetrahymena pyriformis* IGC_{50} and 72hr *Pseudokirchneriella subcapitata* EC_{50}.

QSARs were developed for non-polar narcosis and polar narcosis as separate mechanisms of action as well as for general narcosis. General narcosis was modelled to determine whether the findings of Vaes *et al.*[32], that for membrane/water partitioning the narcosis mechanisms of action can be modelled together, is applicable to phosphatidylcholine membrane/water partitioning (IAM).

For the majority of endpoints, no valid QSARs were developed for polar narcotics. However, from Figure to 18 Figure 25 in Appendix 1.6, it is clear the range of both

hydrophobicity and toxicity covered by polar narcotic compounds is considerably narrower for all endpoints considered. The range of hydrophobicity covered by the datasets is 2-3 log units for the polar narcotic compounds, compared to 6 log units for the non-polar narcotic compounds. Valid QSARs were developed for 14 day guppy LC_{50}, where log P and log $k_{IAM\ (pH\ 7.4)}$ were found to both produce strong QSARs (equations (7.17) and (7.20)) and 48hr *Tetrahymena pyriformis* IGC_{50} where log $k_{IAM\ (pH\ 7.4)}$ produced a valid QSAR and the use of log P did not (equation (7.31)). To increase the confidence in the QSAR models developed the number of compounds considered for each endpoint with regards to log $k_{IAM\ (pH\ 7.4)}$ values needs to be increased, as the existing datasets considering log P are considerably larger.

Plotting toxicity against hydrophobicity for both log P and log $k_{IAM\ (pH\ 7.4)}$ for the various endpoints illustrates that the distinction between the mechanisms of action for *Daphnia*, fathead minnow and guppy is less pronounced when modelled using log $k_{IAM\ (pH\ 7.4)}$ compared with log P. For both the 48hr *Tetrahymena pyriformis* IGC_{50} and 72hr *Pseudokirchneriella subcapitata* EC_{50} there is little distinction between mechanisms of action using either log P or log $k_{IAM\ (pH\ 7.4)}$. In addition, for all endpoints the QSARs developed for the combined mechanisms of action, the use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor produces stronger QSARs than those using log P.

It is clear that QSARs for individual mechanisms of action are stronger. It is evident that modelling non-polar and polar narcotics together is possible. Moreover log $k_{IAM\ (pH\ 7.4)}$ is a better descriptor for these combined models. Where possible the use of QSARs developed for specific mechanisms of action is recommended, as not only are these QSARs generally better, but the observed and measured physiological and behavioural responses[14] indicate a difference in mechanism of action.

The use of combined models may be useful if the specific mechanism of action is unclear. The QSARs developed for non polar narcotics are significantly more significant, than those developed for the combined mechanisms of action. For 14 day guppy LC_{50} toxicity data (dataset 3, the only dataset to produce six valid QSARs) all QSARs for individual mechanisms of action are statistically more significant than those for the combined mechanism of action. This is contrary to the findings of Vaes *et al.*[32]. However, for the polar narcotics considered by Vaes *et al.*[32] the range of log

P considered covers only 3.5 log units, whereas the range considered here is extended in both directions to cover between 5 and 6 log units (depending on the dataset). To improve the models developed for modelling the combined narcotic mechanisms of action together a wider range of compounds covering both classes equally needs to be considered, ideally the range of hydrophobicities for polar narcotics would also be extended to reduce the effect of clusters on the resulting QSARs.

7.5 Conclusions

It has been demonstrated that log $k_{IAM\ (pH\ 7.4)}$ is a suitable descriptor for describing various acute aquatic toxicity endpoints. Log $k_{IAM\ (pH\ 7.4)}$ is a complementary descriptor to log P for modelling narcosis. To improve the predictive capabilities and the domain of these QSARs systematic analysis of the compounds used to develop the high quality QSARs using log P as a descriptor is recommended.

7.6 References

[1] Sarmah, A.K., Meyer M.T., Boxall A.B.A. (2006) A Global Perspective on the Use, Sales, Exposure Pathways, Occurrence, Fate and Effects of Veterinary Antibiotics (Vas) in the Environment *Chemosphere* 65: 725-759.

[2] Kooijman B. (1996) Toxic effects as process perturbations In: Kooijman S.A.L.M., Bedaux J.J.M. eds, *The Analysis of Aquatic Toxicity Data*, Amsterdam, VJ University Press, pp 9-13.

[3] Calow P., Forbes V.E. (2009) Ecotoxicology: Not Just Wildlife Toxicology. In Greim H., Snyder R. eds, Toxicology and Risk Assessment a Comprehensive Introduction, Chichester, Wiley, pp 194-203.

[4] Wilkinson M. (1996) Ecotoxicity. In: Duffus J.H., Worth H.G.J. eds, *Fundamental Toxicology for Chemists*, Cambridge, Royal Society of Chemistry, pp 181-195.

[5] Grindon C., Combes R., Cronin M.T.D., Roberts D.W., Garrod J.F. (2008) Integrated Decision-Tree Testing Strategies for Environmental Toxicity with Respect to the Requirements of the EU REACH Legislation *Altern. Lab. Anim.* 36: 29-42.

[6] Hutchinson T.H., Barrett M.B., Constable D., Hartmann A., Hayes E., Huggett D., Laenge R., Lillicrap A.D., Straub J.O., Thompson R.S. (2003) A Strategy to Reduce the Numbers of Fish Used in Acute Ecotoxicity Testing of Pharmaceuticals *Environ. Toxicol. Chem.* 22: 3031-3036.

[7] van Leeuwen C.J. (2004) Ecotoxicological Effects In van Leeuwen C.J., Hermens J.L.M. eds, *Risk Assessment of Chemicals: An Introduction*, London, Kluwar Academic Publishers, pp 175-237.

[8] The Organisation for Economic Cooperation and Development OECD (2011) *OECD Guidelines for the Testing of Chemicals, No. 201: Freshwater Alga and*

Cyanobacteria, Growth Inhibition Test, Paris, Organisation for Economic Cooperation and Development OECD.

[9] The Organisation for Economic Cooperation and Development OECD (2004) *OECD Guidelines for the Testing of Chemicals, No. 202: Daphnia sp., Acute Immobilisation Test*, Paris, Organisation for Economic Cooperation and Development OECD.

[10] The Organisation for Economic Cooperation and Development OECD (1992) *OECD Guidelines for the Testing of Chemicals, No. 203: Fish, Acute Toxicity Test*, Paris, Organisation for Economic Cooperation and Development OECD.

[11] Boxall A., Barrett K. (2001) *Higher Tier Laboratory Aquatic Toxicity Testing, Cranfield Centre for EcoChemistry Reseach Report No. JA4317E for DETR*

[12] Greim H., Synder R. (2009) Introduction to the Discipline of Toxicology. In Greim H., Snyder R. eds, *Toxicology and Risk Assessment a Comprehensive Introduction*, Chichester, Wiley, pp 1-18.

[13] Roberts D.W. (2010) Mechanisms of Toxic Action in *In Silico* Toxicology. In. Cronin M.T.D., Madden J.C. eds, *In Silico Toxicology: Principles and Applications*, Cambridge, Royal Society Chemistry, pp 334-345.

[14] McKim J.M., Bradbury S.P., Niemi G.J. (1987) Fish Acute Toxicity Syndromes and Their Use in the QSAR Approach to Hazard Assessment *Environ. Health Perspect.* 71: 171-186.

[15] Drummond R.A., Russom C.L., Geiger D.L., Defor D.L. (1986) Behavioral and Morphological Changes in Fathead Minnow (*Pimphales promelas*) as Diagnostic Endpoints for Screening Chemicals According to Mode of Action. In: Poston T.M., Purdy R. eds. *Aquatic Toxicology and Environmental Fate 9th Aquatic Toxicity Symposium*, Baltimore, American Society for Testing and Materials, pp 415-435.

[16] Verhaar H.J.M., van Leeuwen C.J., Hermens J.L.M. (1992) Classifying Environmental Pollutants. 1: Structure-Activity Relationships for Prediction of Aquatic Toxicity *Chemosphere* 25: 471-491.

[17] van Wezel A., Punte S.S., Opperhuizen A. (1995) Lethal Body Burdens of Polar Narcotics: Chlorophenols *Environ. Toxicol. Chem.* 14: 1579-1585.

[18] Bradbury S.P., Henry T.R., Niemi G.J., Carlson R.W., Snarski V.M. (1989) Use of Respiratory-Cardiovascular Responses of Rainbow Trout (*Salmo gairdneri*) in Identifying Acute Toxicity Syndromes in Fish: Part 3. Polar Narcotics *Environ. Toxicol. Chem.* 8: 247-261.

[19] Hermens J. (1995) Prediction of Environmental Toxicity Based on Structure-Activity Relationships Using Mechanistic Information *Sci. Total Environ.* 171: 235-242.

[20] Escher B.I., Hermens J.L.M. (2002) Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects *Environ. Sci. Technol.* 36: 4201-4217.

[21] Deneer J.W., Sinnige T.L., Seinen W., Hermens J.L.M. (1988) A Quantitative Structure-Activity Relationship for the Acute Toxicity of some Epoxy Compounds to the Guppy *Aquat. Toxicol.* 13: 195-204.

[22] Enoch S.J., Hewitt M., Cronin M.T.D. Azam S., Madden J.C. (2008) Classification of Chemicals According to Mechanism of Aquatic Toxicity: An Evaluation of the Implementation of the Verhaar Scheme in Toxtree *Chemosphere* 73: 243-248.

[23] Fent K. (1995) Endocrinically Active Substances in the Environment: State of the Art, Published presentation available from [http://www.epa.gov/edrlupvx/Pubs/uba3_96.pdf] Accessed 12th October 2011.
[24] Bajgar J. (2004) Organophosphates/Nerve Agent Poisoning: Mechanism of Action, Diagnosis, Prophylaxis, and Treatment *Adv. Clin. Chem.* 38: 151-216.
[25] Russom C.L., Bradbury S.P., Broderius S.J., Hammermeister D.E., Drummond R.A. (1997) Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (*Pimephales promelas*) *Environ. Toxicol. Chem.* 16: 948-967.
[26] European Commission, *Off. J. Eur. Un.*, L 396/1 of 30.12.2006 (2006).
[27] The Organisation for Economic Cooperation and Development OECD (2010) *OECD Series on Testing and Assessment, No. 126: Short Guidance on the Threshold Approach for Acute Fish Toxicity*, Paris, Organisation for Economic Cooperation and Development OECD.
[28] European Chemicals Agency (2008) *Guidance on Information Requirements and Chemical Safety Assessment Chapter R.6: QSARs and Grouping of Chemicals*, available from [http://www.echa.europa.eu/documents/10162/17224/information_requirements_r6_en.pdf]
[29] European Union. (1995) *Overview of Structure-Activity Relationships for Environmental Endpoints Part 1: General Outline and Procedure, Report of the EU-DG-XII Project "QSAR for Predicting Fate and Effects of Chemicals in the Environment (Contract ~ EV5V-CT92-0211).*
[30] European Chemicals Agency (2008) *Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7b: Endpoint specific guidance*, available from [http://www.echa.europa.eu/documents/10162/17224/information_requirements_r7b_en.pdf]
[31] Abraham M.H., Rafols C. (1995) Factors that Influence Tadpole Narcosis. An LFER Analysis *J. Chem. Perkin Trans. 2* 10: 1843-1851.
[32] Vaes W.H.J., Ramos E.U., Verhaar H.J.M., Hermens J.L.M. (1998) Acute Toxicity of Nonpolar Versus Polar Narcosis: Is There a Difference? *Environ. Toxicol. Chem.* 17: 1380-1384.
[33] Ward R.S., Davies J., Hodges G., Roberts D.W. (2003) Applications of Immobilised Artificial Membrane Chromatography to Quaternary Alkylammonium Sulfobetaines and Comparison of Chromatographic Methods for Estimating the Octanol-Water Partition Coefficient *J. Chromatogr. A* 1007: 67-75.
[34] Organisation for Economic Cooperation and Development OECD (2010) The OECD QSAR Toolbox version 2.2.1.1120, Organisation for Economic Cooperation and Development OECD, available from <http://www.qsartoolbox.org/>.
[35] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.
[36] Ikemoto Y., Motoba K., Suzuki T., Uchida M. (1992) Quantitative Structure-Activity Relationships of Nonspecific and Specific Toxicants in Several Organism Species *Environ. Toxicol. Chem.* 11: 931-939.
[37] Stephen C.E. (1977) Methods for Calculating an LC_{50} In Mayer F.L., Hamelink J.L. eds, *Aquatic Toxicology and Hazard Evaluation*, Baltimore, American Society for the Testing and Materials, pp 65-84.
[38] Minitab Inc. (2007) Minitab® Statistical Software version 15, Coventry: Minitab Inc..

[39] van Wijk D., Cohet E., Gard A., Caspers N., van Ginkel C., Thompson R., de Rooij, C., Garny V., Lecloux A. (2006) 1,2,4-Trichlorobenzene Risk Assessment with Special Emphasis on the Osparcom Region North Sea *Chemosphere* 62: 1294-1310.

[40] Ramos E.U., Verneer C., Vaes W.H.J., Hermens J.L.M. (1998) Acute Toxicity of Polar Narcotics to Three Aquatic Species (*Daphnia magna, petcilia reticulate and Lymnaea stagnalis*) and its Relation to Hydrophobicity *Chemosphere* 4: 633-650.

[41] Ramos E.U., Vaal M.A., Hermens J.L.M. (2002) Interspecies Sensitivity to the Aquatic Toxicity of Aromatic Amines *Environ. Toxicol. Phar.* 3-4: 149-158

[42] Kühn R., Pattard M., Pernak K., Winter A. (1988) Results of the Harmful Effects of Selcted Water Pollutants (Aniline, Phenols, Aliphatic Compounds) to *Daphnia magna Wat.Res.* 23: 495-499.

8 Determining the hydrophobicity of surfactants using IAM-HPLC

8.1 Introduction

Surfactants (surface-active molecules) are one of the most common and important groups of chemicals that enter the environment from domestic, commercial and industrial sources[1]. Surfactants are found in a diverse range of products including detergents[2], paints[3], personal care products[4] (including shampoos, conditioners and toothpaste) and agrochemicals[5] (including some herbicides and pesticides). The range of products containing surfactants is diverse due to the range of properties surfactants can possess, i.e. they may act as foaming, anti-foaming, wetting emulsifying, or dispersing agents.

In 2009 the global market for surfactants continued to grow with the value increasing to US $24.33 billion[6], with an expected production of 14.2 million metric tons forecast globally for 2010[7]. Anionic and non-ionic surfactants are the dominant classes of surfactant produced[1].

Surfactants are organic molecules and have the general form of a hydrophilic head group and a long hydrophobic tail. It is the nature of the hydrophilic head group that is responsible for the classification of the surfactants into one of four categories, these being non-ionic, cationic, anionic and amphoteric surfactants. The ionised head group of a surfactant is usually balanced by a counter-ion e.g. sodium for anionic surfactants, and chlorine or bromine for cationic surfactants. The general form for these classes is shown in Figure 39.

Due to the structure of surfactants, specifically that they contain a hydrophilic head and a hydrophobic tail, they align themselves at the interface of hydrophilic/hydrophobic phases, i.e. at the water/oil or water/air interface. Above a certain concentration, known as the critical micelle concentration (CMC), surfactants form micelles. Micelles are aggregations of surface active molecules. In polar solvents the hydrophilic heads remain in contact with the polar solvent and the hydrophobic tails are protected at the centre of the micelle (Figure 40). In a non-polar solvent the inverse is the case i.e. it is the hydrophilic heads that form the core

of the micelle being protected from the non-polar environment. This is less favourable and less stable, due to the interaction and localisation of head group charges. The CMC value is compound specific and can be measured by various methods described in Section 8.1.1.

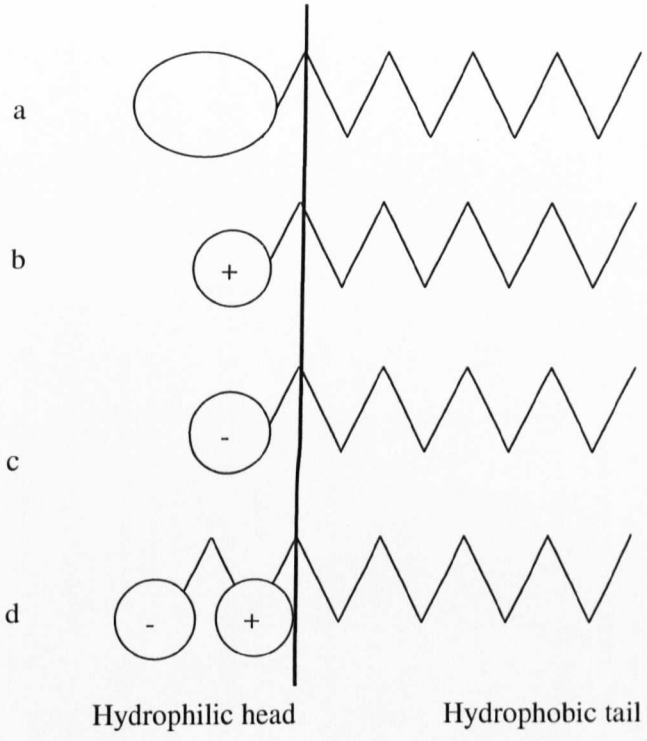


Figure 39 – General form for the four surfactant classes [a] non-ionic, [b] cationic, [c] anionic and [d]amphoteric

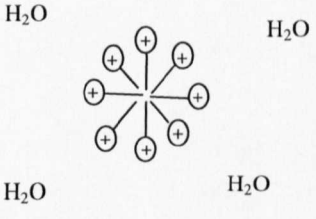


Figure 40 - Example of a micelle in a polar solvent

Above the CMC value the physical and thermodynamic properties of solutions of surfactants change[8]. It is the aggregation of surfactants at the octanol-water interface that makes the determination of log P difficult using traditional methods. Measurement of the partition coefficient of surfactants is particularly difficult due to the widely reported requirement for the compound to be in the free form i.e. to exist as unassociated molecules without the presence of micelles[9].

A proportion of the surface-active agent is between-phases at the aqueous/lipid interface. Roberts[10] has reported that the formation of micelles does not affect the measured log P value because hydrophobicity is the ratio of sample concentration in water and octanol at equilibrium. This ratio is not affected by the presence of micelles. It should be noted that as the effect of this "between-phases" partitioning occurs at the interface of the aqueous and lipid layers the effect is relative to the surface area of the interface. Therefore, traditionally the set-up and sampling techniques used in the determination of log P of surfactants has been thought to affect the result obtained[11].

In general, the assessment of the toxicity of surfactants at concentrations above the CMC, i.e. at a concentration where micelles have formed, is not as useful as analysis of toxicity is predominantly for the free form. When chemicals enter the environment their relative concentration decreases as the compound disperses. In addition, for the toxic effects to occur, the compound needs to partition across a biological membrane. Micelles comprise aggregations of a number of molecules and hence are large by their nature, which makes partitioning across membranes more difficult. Therefore, it is the toxicity of surfactants in the free form that is of interest[12]. It is noted that surfactants can enhance the penetration (and therefore, possibly the toxicity) of other compounds across membranes i.e. co-administration of drugs with surfactants to increase membrane permeability compared to the drug alone[13] and the use of surfactants as a co-substrate enhancing solubilisation[14] and facilitating transport leading to increased toxicity or biodegradation[15].

8.1.1 Methods for measuring CMC

There are several experimental methods to determine CMC values for a surface active agent. These include the surface tension method, electrical conductivity and optical and spectroscopic methods. Brief experimental details of each of these methods are provided below.

8.1.1.1 Surface tension method

Surface tension is the ability of a liquid to resist an external force. When added to a liquid a surfactant reduces the surface tension. As the concentration of surfactant increases the surface tension will decrease until the CMC is reached. Once the

concentration of the surfactant exceeds the CMC, the surface tension of the liquid remains constant. Consequently the CMC value of a surfactant can be determined by measuring the surface tension of a solvent at various concentrations of surfactant. A number of methods are available to measure surface tension including the maximum pull / de Nouy ring method, the Wilhelmy plate detachment method and the drop weight method[16].

The maximum pull / de Nouy method of determining surface tension (and by inference CMC) measures the maximum force required as a probe is withdrawn from a test solution of the surfactant. The experimental set-up means the surface contact angle is near 0°[17] and the effect of the contact angle is negligible. The force acting on the probe is, therefore, proportional to the meniscus of liquid adhering to the probe. This is related directly to the surface tension by the following relationship:

$$\gamma = F_{max}/2\pi r_p \quad [18] \qquad (8.1)$$

Where

γ is the surface tension

r_p is the perimeter of the probe

F_{max} is the maximum force applied to the probe until surface tension is broken

To determine the CMC value the surface tension is plotted against concentration. The CMC is the point where an increased concentration does not lower the surface tension, i.e. the intersection of two lines of different gradients. The specifics of the maximum pull method used in this thesis to determine the CMC value of surfactants analysed are detailed in section 8.3.2.

8.1.1.2 **Electrical conductivity**

Electrical resistance of solutions of the surfactant at various concentrations are measured and interpreted in terms of the specific conductivity. The conductivity is plotted against the surfactant concentration; the CMC value is the concentration where there is a sharp change in gradient of a plot of electrical conductivity against concentration[16, 19].

8.1.1.3 Optical and spectroscopic methods

To determine the CMC value of a surfactant using optical or spectroscopic techniques a solution of the surfactant is prepared (where the concentration of the surfactant is above the CMC) and a dye is added. The intensity of light is measured at an angle 90° from the incident beam. As the surfactant is diluted the intensity of the dye decreases; at the CMC value the intensity will drop abruptly. A similar method measures the change in refractive index with concentration to determine the CMC value. For some surfactants the absorption spectrum is different for the free and micellar form allowing for the determination of the CMC[16].

8.1.2 Datasets of experimental CMC values

CMC values have been collated into a National Standard Reference Data System (NSRDS). Specifically CMC values of aqueous surfactant systems[16] were collated from 87 publications along with a guide as to the quality of the values reported. The NSRDS for CMC values were collated from literature values from 1926 to 1966. Although there are more recent CMC values reported in the literature, this is the most comprehensive source of CMC values available, which has not subsequently been updated. Additionally, an assessment of the quality of the data collated has been reported in the NSRDS. The NSRDS contains multiple CMC values for many compounds. The CMC values are classed as "recommended" or simply "reported". For example, dioctylsulfosuccinate sodium salt (surfactant 14) has one recommended value and three further values reported in the complete table (Table 54).

Temperature °C	CMC value	Quality assessment		Method	Reference
		Surfactant material	Method of determination		
25	$2.5x10^{-3}$ M*	Very pure material, may still contain impurities	Precise to about 3% Accurate to 5%	Surface tension log plot	20
25	$5.55x10^{-3}$ M	Commercial surfactant material	Accurate to an order of magnitude	Unspecified conductance	21
29.9	$5.65x10^{-3}$ M	Very pure material, may still contain impurities	Accurate to an order of magnitude	Equivalent conductance graph	22
29.9	$6.15x10^{-3}$ M	Very pure material, may still contain impurities	Accurate to an order of magnitude	Specific conductance graph	22

*** Recommended**

Table 54 - CMC values for dioctylsulfosuccinate sodium salt from NSRDS, including experimental conditions reported, quality assessment of material and method of determination and the reference for the data[16].

8.1.3 *Toxicity of surfactants*

Aquatic toxicity of surfactants is commonly assessed using the same species and tests as discussed in Sections 1.2.2 and 7.1 i.e. algae, *Daphnia* and fish. Surfactants tend to be present in products as mixtures of different isomers and / or chain lengths. For this reason many of the available toxicity data relate to these mixtures[23]. If the toxicity of individual chain lengths is understood, the toxicity of mixtures can be inferred.

8.1.4 QSARs for surfactants

Roberts has developed QSPRs that can predict properties of surfactants including their hydrophobicity (log P)[23] and their CMC value[24]. Roberts has also developed QSARs to predict acute lethal toxicity of anionic and non-ionic surfactants to both *Daphnia* and *Gammarus*[23] and biodegradation of linear alkylbenzene sulphonate (LAS) surfactants[25]. Many others[26, 27] have developed QSARs predicting the toxicity of surfactants to these and other species. QSARs for the prediction of log P for select classes of surfactants[28] have been developed, as well as additional fragments and factors[23, 29] to allow the prediction of log P for surfactants using the Hansch & Leo[30] approach.

8.1.5 The application of IAMs to surfactants

Ward *et al.*[31] reported a good correlation of log k_{IAM} with 48hr LC_{50} toxicity data to the water flea (*Daphnia magna)* for a set of closely related alkylammonium sulfobetaines, a class of amphoteric surfactants. Their analysis was obtained using the IAM.PC.DD2 column. A good correlation was reported between log P and log k_{IAM} and between log k_{IAM} and log $1/LC_{50}$ for *Daphnia magna.*

8.2 Aim of the Chapter

The aim of this chapter was to assess the suitability of IAM-HPLC as a method to determine the hydrophobicity of surfactants covering a range of surfactant classes. To achieve this aim the CMC value for the surfactants analysed also needed to be determined to ensure that the IAM-HPLC analysis was performed with the surfactant in the free form. Therefore, CMC values were obtained from the NSRDS database, or experimentally determined. The log $k_{IAM\ (pH\ 7.4)}$ values obtained for the surfactants

analysed were compared to log P values and trends (in log P, or log $k_{IAM\ (pH\ 7.4)}$) within the homologous series were investigated.

8.3 Method

8.3.1 Determination of CMC values

All surfactants for which both CMC values and hydrophobicity (using IAM-HPLC) were determined are listed in Table 55, along with an identifier, the structure and the surfactant class.

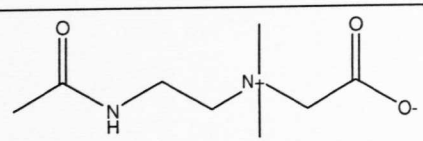
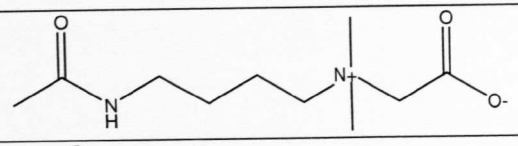
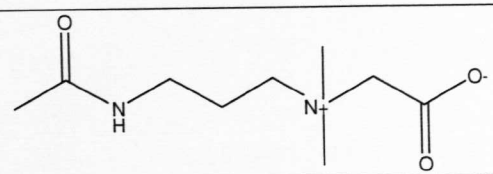
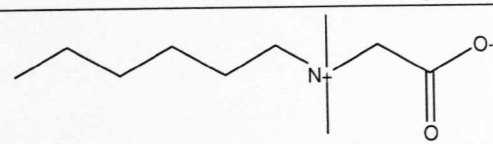
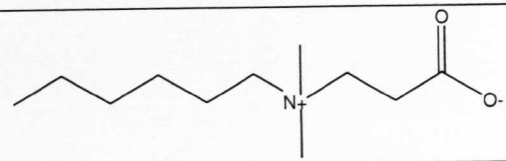
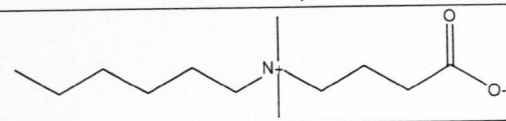
Surfactants 1-8 were provided by Unilever, surfactants 9-14 were sourced commercially. Surfactants were selected for analysis on the basis of their availability as single chain lengths and to cover a range of surfactant classes.

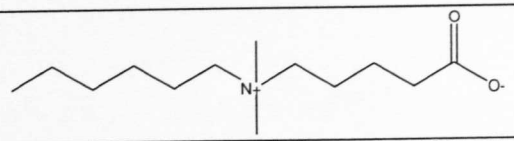
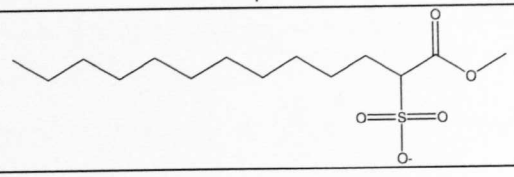
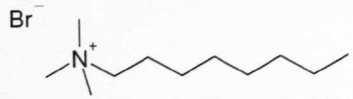
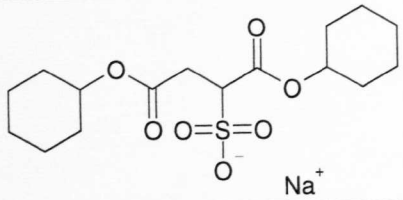
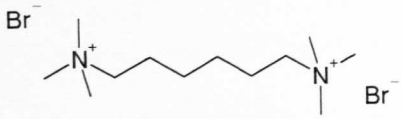
For the surfactants that were commercially available in single chain lengths, the NSRDS[16] was checked for the corresponding CMC values. For surfactants which did not have CMC values listed in the NSRDS, the CMC values were determined experimentally in this project as detailed in Section 8.3.2.

8.3.2 Experimental determination of CMC using the maximum pull / de Nouy method

8.3.2.1 Materials

Methanol (HPLC gradient grade), NaCl, KCl, $Na_2HPO_4 \cdot 7(H_2O)$ and KH_2PO_4 were purchased from Fisher Scientific (Loughborough UK). Water was demineralised using reverse osmosis and then passed through an ion exchange unit. All samples (unless specified) were obtained from commercial sources and were a single chain length of 98% purity or greater. These were used without further purification. Dodecylbenzenesulfonic acid (CAS no. 85536-14-7) was used as the external standard in the analysis.

Identifier	Surfactant name	Structure	Surfactant class	pKa[32]
1	2-((acetamidoethyl)dimethylammonio)acetate		Amphoteric	
2	2-(4-acetamidobutyl)dimethylammonio)acetate		Amphoteric	
3	2-(3-acetamidopropyl)dimethylammonio)acetate		Amphoteric	
4	2-(hexyldimethylammonio)acetate		Amphoteric	
5	3-(hexyldimethylammonio)propanoate		Amphoteric	
6	4-(hexyldimethylammonio)butanoate		Amphoteric	

Identifier	Surfactant name	Structure	Surfactant class	pKa[32]
7	5-(Hexyldimethylammonio)pentanoate		Amphoteric	
8	C13 Methyl Fatty Acid Ester Sulphonate (FAES)		Anionic	-5.21
9	Trimethyloctylammonium bromide		Cationic	
10	Dicyclohexylsulfosuccinate sodium salt		Anionic	-5.89 -5.51 0.31
11	Hexamethonium bromide		Cationic	

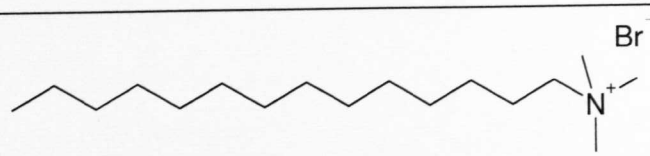
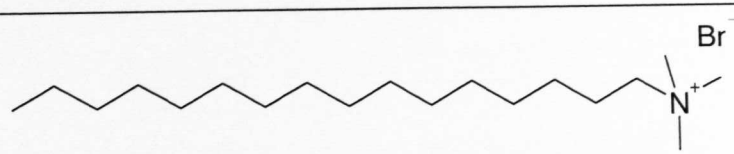
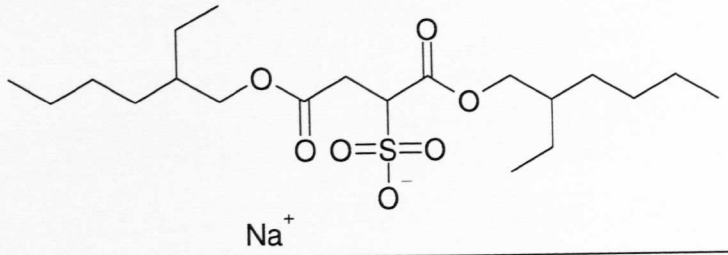
Identifier	Surfactant name	Structure	Surfactant class	pKa[32]
12	Myristyltrimethyl ammonium bromide		Cationic	
13	Hexadecyltrimethyl ammonnium bromide		Cationic	
14	Dioctylsulfosuccinate sodium salt		Anionic	-5.69 -6.21

Table 55 – Identifier, surfactant name, structure and surfactant class for a number of surfactants considered for analysis by IAM-HPLC

8.3.2.2 Instrumentation

A Delta-8 multichannel mircotensionmeter was used to determine the surface tension of samples. The range of this system is 10-100 mN m^{-1} with a resolution of 0.01 mN m^{-1}. This employs the maximum pull method of determining CMC values. The tension was measured over a range of concentrations (0.005-10000 mg/L, with each sample being half as concentrate, as that measured previously) and the results plotted on a log-log scale. The surface tension of water at 25°C is 72 mN m^{-1} and 27 mN m^{-1} for 60:40 methanol:water at 25°C, as reported by Vázquez *et. al.*[33]. These values were used as blank values; any deviation from these values is due to the salts from the 10mM phosphate buffer or the surfactant. If a surfactant solution has the same surface tension as the solvent across the range of concentrations considered, the CMC value for this compound is above the range of concentrations considered here.

8.3.2.3 Sample preparation

All samples were prepared at a nominal concentration of 20mg/mL for the determination of CMC. Samples were prepared in water, three of the surfactants were additionally prepared in 10mM phosphate buffered saline (PBS) and 60:40 MeOH/10mM PBS (preparation of PBS as detailed in Section 3.3.4.2. Dodecylbenzenesulfonic acid was prepared at a concentration of 20mg/mL in water and analysed as an external standard.

8.3.3 IAM-HPLC analysis

The log $k_{IAM\ (pH\ 7.4)}$ value for the surfactants was determined using the IAM-HPLC method detailed in Chapter 3. However, samples were not prepared at a concentration of 10^{-2}M (as used in previous analysis); instead the samples were all prepared at a concentration below the CMC value. This is compound specific and the CMC values are reported in Table 56. The concentration at which the surfactants were prepared for analysis using IAM-HPLC are reported in Table 57. Additionally, some samples (samples 1-7) were detected using ultra violet (UV) detection at 210nm instead of using a refractive index detector as specified in Chapter 3.

8.3.4 Interpretation of log k_{IAM} values determined for surfactants

For the surfactants analysed using IAM-HPLC, SMILES strings for the structures were entered into KOWWIN[34] and VCCLAB[35, 36] to obtain experimental (where

available) and predicted log P values. The pKa for each surfactant, and hence the degree of ionisation was determined by entering the SMILES strings into SPARC v4.6[32] (SPARC Performs Automated Reasoning in Chemistry)[37]. This also allowed the calculation of log P using the relationship between log $k_{IAM\ (pH\ 7.4)}$ and log P, reported in Chapter 3 (equations (4.3) and (4.4), for unionised and ionised compounds respectively).

The experimental log k_{IAM} values determined for the surfactants were compared to the log P values obtained from VCCLAB[35, 36]. Additionally, the combined fragment value for CH_2 and the correction factor for the number of bonds (determined in Chapter 4) were investigated by analysis of log $k_{IAM\ (pH\ 7.4)}$ values obtained for a short homologous series.

8.4 Results and discussion

8.4.1 CMC values from the National Standard Reference Data System

The NSRDS of CMC values[16] was searched for the surfactants listed in Table 55. The CMC values from the NSRDS are reported in Table 56; both the recommended CMC value and the range of CMC values from the complete table were recorded where available. For the surfactants that did not have a CMC value reported in the NSRDS[16], CMC values were determined experimentally

Identifier	CMC from National Standard Reference Data System[16]					Experimental CMC value (M)		
	Recommended value	Reference	Complete table	Reference	NSRDS entry no.	Water	PBS	60:40 MeOH PBS
1						$> 5.2x10^{-2}$		
2						$>4.8x10^{-2}$		
3						$> 4.8x10^{-2}$		
4						$> 5.2x10^{-2}$		
5						$> 4.8x10^{-2}$		
6						$> 4.6x10^{-2}$		
7						$> 4.3x10^{-2}$	$2.2x10^{-2}$	Not surface active
8						$2.02x10^{-3}$	$5.1x10^{-4}$	Not surface active
9			$1.3 - 2.8x10^{-1}M$	39, 40	93			
10						$> 2.6x10^{-2}$	$> 2.6x10^{-2}$	Not surface active
11						$> 2.8x10^{-2}$	$6.9x10^{-3}$	
12			$1.3x10^{-3} - 4.2x10^{-4}M$	41, 42	98	1.71x10-3		
13	$9.8x10^{-4}M$	38	$1.0x10^{-3} – 9.8x10^{-4}$ M	38, 43	99			
14	$2.6x10^{-3}M$	20	$2.5 – 6.1x10^{-3}M$	20, 22	262	$2.83x10^{-3}$		

Table 56 - Recommended and range of CMC values and reference number for surfactants analysed using IAM-HPLC from National Standard Reference Data System of CMC values[16] and the experimental CMC values

8.4.2 Experimental determination of CMC values

Using the method detailed in Section 8.3.2, aqueous CMC values were determined for 12 surfactants (surfactants indentified as 1-8, 10-12 & 14, in Table 55). Additionally, CMC values for surfactants 7, 8 & 10 were also determined in 10mM PBS and 60:40 MeOH:10 mM PBS. This was to allow comparison between the aqueous CMC value and the CMC value determined using the HPLC solvents. CMC values for surfactants 12 and 14 were determined experimentally despite literature values being available. This was to ensure that the values obtained using the maximum pull method were consistent with the values reported in the National Standard Reference Data System of CMC values[16].

CMC values were determined from log-log plots of surface tension against concentration. The CMC value is the point of intersection of the two distinct gradients, i.e. the "elbow" where the surface tension levels off despite an increase in concentration of the surfactant. The data for surfactant 14 (dioctylsulfosuccinate sodium salt) is shown in Figure 41 as an example. The graphs for all other surfactants are available in Appendix 1.7, Figures 26-43. The CMC values determined are reported in Table 56.

It can be seen from Table 56 that for the surfactants prepared in 60:40 methanol:10mM PBS, no CMC value was determined. The surfactants are not surface-active across this range of concentrations. This is apparent as the surface tension of the sample and solvent is equal to the surface tension of the solvent at all concentrations investigated. The surfactant does not lower the surface tension of solvent across the range of concentrations assessed. For the surfactants analysed in both water and 10mM PBS, the CMC value is lower for 10mM PBS compared to water. This is in line with the findings of many studies into the effects of electrolytes on micellisation[44-47]. Baloch *et al.*[45] demonstrated that CMC values decrease as ionic strength increases in the order of $Na^+ > K^+ > Li^+$. This effect is observed due to the salts increasing the induced dipoles within the water molecules, thus leading to reduced solubility of the hydrocarbon chains in the salt solution, causing a reduction in the CMC value (i.e. the surfactant molecules aggregate at a lower concentration)[48].

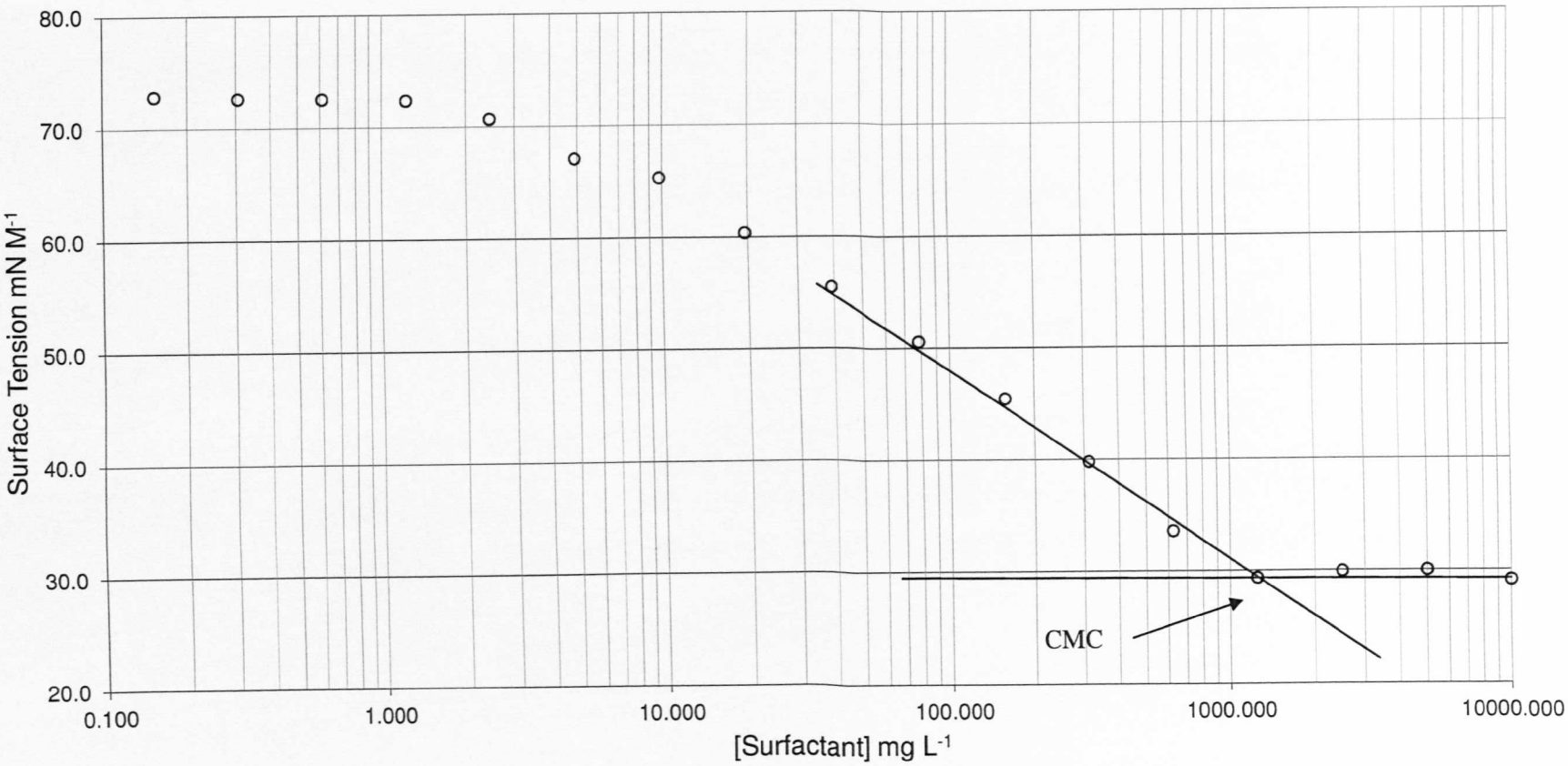


Figure 41 - Example result for the determination of aqueous CMC using the maximum pull method for surfactant 14 (dioctylsulfosuccinate sodium salt)

For the surfactants for which CMC values were determined experimentally and for which literature values were also reported in the NSRDS[16] the two CMC values were compared. For both surfactants (12 and 14) the CMC values are comparable and consistent with the literature values available (Table 56).

8.4.3 IAM-HPLC analysis of surfactants

Surfactants 1-14 were analysed using IAM-HPLC, with either RI or UV detection. A summary of the log $k_{IAM\ (pH\ 7.4)}$ values is detailed in Table 57. The full results are available in Appendix 1.7 Table 67 and 68 and Figure 44. The log $k_{IAM\ (pH\ 7.4)}$ values determined are repeatable and reproducible when measured in triplicate. This increases confidence in the values obtained and indicates that this method of determining log $k_{IAM\ (pH\ 7.4)}$ is a potentially useful method to determine, experimentally, the hydrophobicity of surfactants.

Identifier	Surfactant concentration used for IAM-HPLC analysis	Log $k_{IAM\ (pH\ 7.4)}$	Method of detection
1	10^{-2}M	-1.02	UV
2	10^{-2}M	-0.75	UV
3	10^{-2}M	-0.90	UV
4	10^{-2}M	-0.15	UV
5	10^{-2}M	-	UV
6	10^{-2}M	-1.04	UV
7	10^{-2}M	-0.02	UV
8	$5x10^{-4}$M	N/D	N/A
9	10^{-2}M	1.59	RI
10	$6.4x10^{-3}$M	1.81	RI
11	$7x10^{-3}$M	N/D	N/A
12	$4x10^{-4}$M	N/D	N/A
13	$2.4x10^{-4}$M	N/D	N/A
14	$7x10^{-4}$M	N/D	N/A

Table 57 - Experimental log $k_{IAM\ (pH\ 7.4)}$ values determined, and method of detection for the surfactants analysed, N/D not determined, N/A not applicable

Surfactant 5 has no log $k_{IAM\ (pH\ 7.4)}$ value reported. For each analysis of this surfactant a peak was observed. For preparation 1 the elution time was greater than for the unretained compound (water), whilst for preparations 2 and 3 the elution time was shorter than that of the water. As can be seen from the full data in Appendix 1.7 Table 67 for surfactant 5, the reported log $k_{IAM\ (pH\ 7.4)}$ value is lower than -2, which is the lower limit of the method as discussed in Section 3.4.1. Thus this value is considered to be unreliable and is not reported.

Log $k_{IAM\ (pH\ 7.4)}$ values were not determined for surfactants 8 and 11 to 14. These surfactants have lower CMC values relative to the other surfactants considered here. No peak was detected from the chromatograms; it is suspected that for these surfactants the detectors (RI or UV) were not sufficiently sensitive to detect the low concentrations of these surfactants.

There are no experimental log P values available in the literature for the surfactants considered here. Therefore, log P values were predicted from equation (4.4) for ionised compounds and equation (4.3) for unionised compounds (equations were determined from the analysis of compounds in Chapter 3, Figure 21). It is noted that these were determined using non-surfactant materials, therefore, these may not be as appropriate for the charged surfactants considered here. The log $k_{IAM\ (pH\ 7.4)}$ values used were those reported in Table 57.

Additionally, the SMILES strings for the surfactants were entered in the VCCLAB software[35, 36] to determine predicted log P values using a range of models. All results are reported in Table 58. It should be noted that the predictions from VCCLAB were anticipated to be poor, due to the surfactants being ionic compounds predominantly with counterion salts. In addition no method used to predict log P has been identified as being suitable for analysis of surfactants. Therefore, care is required when interpreting the log P values reported in Table 58.

Identifier	Log k_{IAM} (pH 7.4)	Predicted log P Equation 4.3 and 4.4	Log P									% ionised
			AlogPs[c]	AC log P[d]	XlogP2[e]	XlogP3[e]	AB/logP[52]	miLogP[f]	KOWWIN[g]	ALogP[h]	MLogP[d]	
1	-1.02	-0.37[b]	-1.55	-0.59	-2.81	-0.81	-1.04	-5.6	-2.72	-2.85	-3.76	N/A
2	-0.75	-0.08[b]	-1.69	0.34	-2.09	-0.1	-0.59	-5.49	-1.73	-1.69	-3.14	N/A
3	-0.90	-0.24[b]	-1.7	-0.12	-2.45	-0.45	-1.1	-5.55	-2.22	-2.79	-3.44	N/A
4	-0.15	0.48[b]	-1.31	2.2	0.18	2.21	2.01	-4.51	-0.47	0.2	-1.84	N/A
6	-1.04	-0.40[b]	-1.36	3.13	0.7	2.47	3.29	-4.11	0.51	0.55	-1.27	N/A
7	-0.02	0.60[b]	-1	3.59	1.05	2.83	3.27	-3.61	1.00	1.01	-1.00	N/A
9	1.59	1.35[b]	-1.98	4.05	4.18	4.71	-	-	-	-	-	N/A
10	1.81	5.07[a]	2.68	0.65	2.49	2.51	-	-0.17	1.76	-	-	99.99

[a] – Calculated using equation 4.4
[b] – Calculated using equation 4.3
[c] – Developed using associative neural networks method[49]
[d] – Fragment based algorithm[50]
[e] – Atom contribution method with correction factors[50]
[f] - Counts of atoms, bonds, fragments and functional groups[50]
[g] – Atom and fragment based method for correction factors[51]
[h] – e-state Indices based associative neural networks method[50]

Table 58- Predicted log P values for the surfactants analysed using IAM-HPLC, determined using equations 4.3 and 4.4 or using VCCLAB[35, 36, 52]

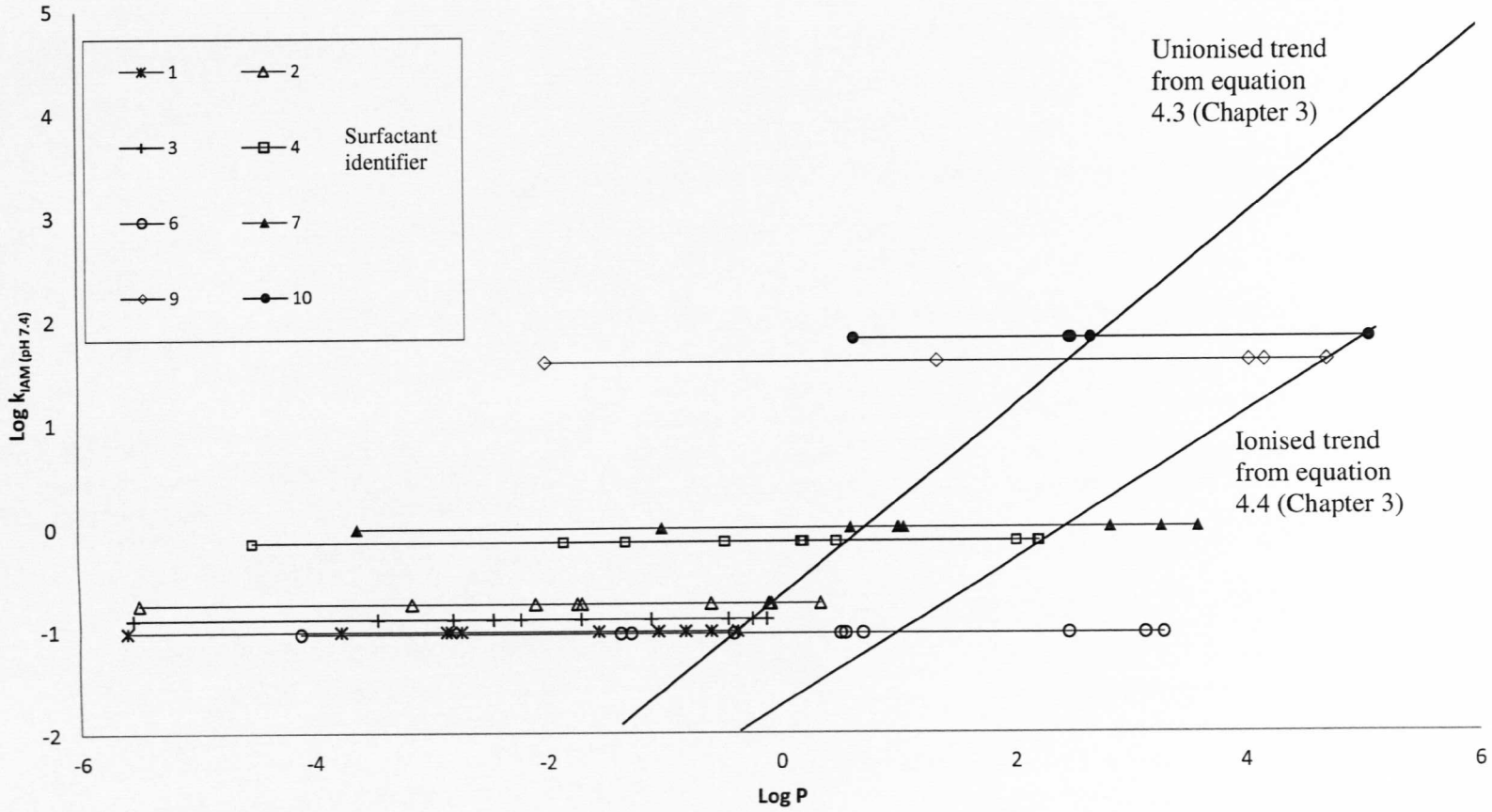


Figure 42 - Plot of log k_{IAM} against log P for all compounds showing the range of calculated log P values from Table 58, for the surfactants analysed

Figure 42 illustrates the relationship between log P and log $k_{IAM\ (pH\ 7.4)}$ determined in Chapter 3 for both ionised and unionised compounds. Additionally, it shows the range of log P values predicted using VCCLAB and reported in Table 58. This clearly shows the wide range of predicted log P values determined for surfactants using a range of methods. The spread of log P values obtained indicates that the current methods for predicting log P of surfactants can lead to spurious predictions. This is a reflection of both the difficulty in experimentally determining log P for surfactants and the more complicated interactions within surfactants that predictive methods need to account for i.e. chain folding, proximity effects and less common fragments. It has been demonstrated that the experimental determine of log $k_{IAM\ (pH\ 7.4)}$ for surfactants is possible, the results obtained are repeatable and reproducible. It is noted that the determination of log P using HPLC for surfactants is not recommended[53]. Given the relationships between log P and log $k_{IAM\ (pH\ 7.4)}$ (derived in Chapter 3 and Chapter 5) it should be possible to infer log P based on experimental log $k_{IAM\ (pH\ 7.4)}$ values.

Given it is possible to predict log P based on a compound's structure (and it was shown in Chapter 4, that this approach was applicable to log $k_{IAM\ (pH\ 7.4)}$), the applicability of this approach to surfactants was investigated. To achieve this, the structures of the surfactants analysed were compared for similarity. Surfactants 1, 2 and 3 form a small homologous series; the surfactants have the same backbone illustrated in Figure 43 with different numbers of carbons in the spacer unit (between the amide and quaternary amine group). Table 59 shows that as the number of carbons in the chain increases the experimental log $k_{IAM\ (pH\ 7.4)}$ value increases linearly. This is expected given the hydrophobic nature of the CH_2, discussed in Chapter 4. The fragment value for CH_2 and a bond is calculated as 0.12 (when comparing surfactants 3 and 1) and 0.15 (when comparing surfactants 2 and 3). The fragment values were calculated by subtracting the log $k_{IAM\ (pH\ 7.4)}$ value for the longer surfactant from the shorter surfactant. The values are similar to each other; however they are smaller than the $CH_{2\ aliphatic}$ fragment and combined bond value of 0.49 determined in Chapter 4 (0.35 (CH_2 fragment) + 0.14 (bond factor)).

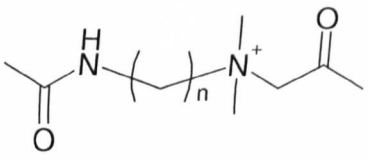


Figure 43 – Structural backbone of carboxybetaines for a short homologous series analysed using IAM-HPLC

Identifier	No of carbons in spacer unit	Log $k_{IAM\ (pH\ 7.4)}$
1	2	-1.02
3	3	-0.90
2	4	-0.75

Table 59 - Identifier, number of carbons in spacer unit and experimental log $k_{IAM\ (pH\ 7.4)}$ values for a short homologous series of carboxybetaines

There are a number of contributions (fragment and correction factor values) to the hydrophobicity of a surfactant. The addition of CH_2 to a molecule increases the hydrophobicity; this should be a constant value. This appears not to be the case when determining the value of the CH_2 fragment in molecules where large polar moieties are present. However, the CH_2 fragment value has been determined in this case without the consideration of proximity. If one takes CH_2 to be a constant and consider the additional factor of proximity, this would lead to an apparent smaller CH_2 value as (observed above). This effect is likely to account for the differences between the two surfactant fragment (CH_2 and bond) values and the aliphatic value, although without additional measurements it is not possible to quantify the affect of proximity. The difference between the values, (CH_2 and a bond) indicates that there are additional interactions involved in the hydrophobicity which impact on the partitioning of surfactants that are not present in the partitioning of simple organic chemicals.

Full analysis of the surfactants for fragment and factor values was not performed (either in Chapter 4 or here) due to the relatively small number of surfactants analysed. Additionally, for the surfactants analysed there are multiple new fragments and different factor values that would need to be introduced. Including:

N^+ quaternary ammonium fragment

Br^- counter ion

SO_3^-

Chain folding

It should be noted that when determining the hydrophobicity of surfactants the inclusion of the N^+ fragment into a molecule, not only affects the hydrophobicity of the molecule but also the geometry of the molecule changes. Therefore, the quaternary N^+ fragment, that currently does not have a specific fragment value, cannot be obtained by linear extrapolation of primary, secondary and tertiary amine fragments[30]. In addition for compounds, including surfactants, containing branched alkyl chains, the chains have the potential to fold (Figure 44). Folding reduces the size of the cavity within the solvent that the molecule occupies. Folding also increases "sharing" of water molecules as the number of close groups within the molecule increases. The smaller the cavity, the lower the energy required to solvate the molecule (the folded conformation is lower in energy in solution, contrary to the low energy conformation in a vacuum). Water sharing is also the lower energy interaction (due to steric effects). Together these interactions reduce the hydrophobicity of the surfactant[24, 54]. Folding is particularly important for gemini surfactants which are surfactants that contain more than one hydrophobic tail and more than one hydrophilic head group[55].

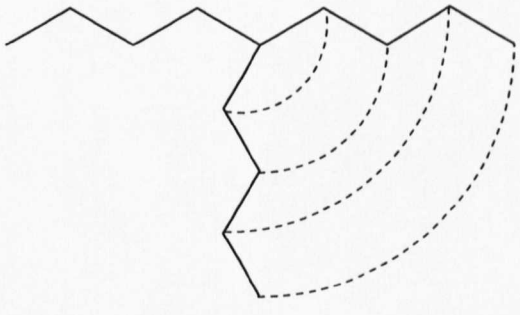


Figure 44 - Chain folding in a branched molecule

To determine fragment and factor values for the prediction of log $k_{IAM\ (pH\ 7.4)}$ for surfactants, the number of surfactants analysed needs to be increased, in a systematic manner such that the effect of branching, chain length and folding, proximity effects, counter ions and new fragments can all be considered.

8.5 Conclusions

It has previously been demonstrated by Ward *et. al.*[31] that IAM-HPLC is a technique suitable for determining a measure of hydrophobicity for quaternary alkylammonium sulfobetaines. The analysis here has extended the range of surfactant classes suitable for analysis using IAM-HPLC to include cationic and anionic surfactants. Additionally, the range of amphoteric surfactants analysed has been extended to include carboxybetaines. For some of the surfactants analysed a log $k_{IAM\ (pH\ 7.4)}$ value was not determined. For these surfactants a lower limit of detection is required than that provided by UV or RI given the requirement to perform the analysis below the CMC for each surfactant.

The log k_{IAM} values for the surfactants were determined from preparation in triplicate, and injected in triplicate. The results obtained showed good repeatability and reproducibility under the conditions of analysis. Given the broad range of log P values obtained using a range of predictive methods, IAM-HPLC may be a more appropriate method for determining hydrophobicity. It has also been shown that log P can be calculated from log $k_{IAM\ (pH\ 7.4)}$ values. This is useful given the prevalence of log P as a descriptor in QSARs.

The comparison of log $k_{IAM\ (pH\ 7.4)}$ values for three surfactants that form a short homologous series resulted in a similar fragment value for the CH_2 and a bond unit. This indicates that the fragment and factor values method of predicting log $k_{IAM\ (pH\ 7.4)}$ should be applicable to surfactants. To extend the fragment and factor values method determined in Chapter 4, a systematic approach to the log $k_{IAM\ (pH\ 7.4)}$ analysis of surfactants needs to be undertaken. This would allow the calculation of fragment and correction factor values that are required i.e. the quaternary ammonium fragment, proximity effect etc..

The use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor to predict both the ability of surfactants to absorb across the skin barrier and to predict ecological toxicity endpoints, such as *Daphnia* EC_{50} values, should be investigated further.

8.6 References

[1] Lara-Martín P.A., Gómez-Parra A., González-Mazo E. (2008) Reactivity and Fate of Synthetic Surfactants in Aquatic Environments *Trends* 27: 648-695.

[2] Khandal R.K., Kaushik S., Seshadri G., Khandal D. (2009) Production of Solvents for Detergent Industry In: Zoller U., Sosis P. eds. *Handbook of Detergents Part F: Production Surfactant Science Series:142*, New York, CRC press, pp 491-530.

[3] Croll S. (2007) Overview of Developments in the Paint Industry Since 1930 In: *Modern Paints Uncovered: Proceedings From the Modern Paints Uncovered Symposium*, Los Angeles, Getty Conservation Institute, pp 17-29.

[4] Goddard E.D. (1999) Polymer/Surfactant Interaction in Applied Systems In: Goddard E.D., Gruber J.V. eds, *Principles of Polymer Science and Technology in Cosmetics and Personal Care Cosmetic Science and Technology Series/Volume 22*, New York, Marcel Dekker, pp 181-216.

[5] Tadros T.F. (1995) Surfactants in Agrochemicals In: Zoller U., Sosis P. eds. *Surfactant Science Series :54*, New York, CRC press, pp 1-6.

[6] Acmite Market Intelligence (2010) Market Report: World Surfactant Market 1: 38, available from < http://www.acmite.com/market-reports/chemicals/world-surfactant-market.html> [Accessed 25th September 2011]

[7] Storck W.J. (2003) Surfactant Makers are Suffering *Chem. Eng. News* 81: 21-22.

[8] Fainerman V.B., Möbius D., Miller R. (2001) *Surfactants: Chemistry, Interfacial Properties, Applications*, Amsterdam, Elsevier Science, pp 75.

[9] Short J., Roberts J., Roberts D.W., Hodges G., Gutsell S, Ward R.S. (2010) Practical Methods for the Measurement of Log P for Surfactants *Ecotox. Environ. Safe.* 73: 1484-1489.

[10] Roberts D.W. (2000) Aquatic Toxicity – Are Surfactant Properties Relevant? *J. Surfactants Deterg.* 3: 309-315.

[11] Dearden J.C., Bresnen G.M. (1988) The Measurement of Partition Coefficients and Lipophilicity *Quant. Struct.-Act. Relat.* 7: 133-144.

[12] Singer M.M, George S, Tjeerdema R.S. (1995) Relationship of Some Physical-Chemical-Properties of Oil Dispersants and their Toxicity to Marine Organisms *Arch. Environ. Con. Tox.* 29: 33-38.

[13] Ross B.P., Braddy A.C., McGeary R.P., Blanchfield J.T. (2004). Micellar Aggregation and Membrane Partitioning of Bile Salts, Fatty Acids, Sodium Dodecyl Sulfate, and Sugar-Conjugated Fatty Acids: Correlation with Hemolytic Potency and Implications for Drug Delivery *Mol. Pharm.* 1: 233-245.

[14] Volkering F., Breure A.M., van Andel J.G., Rulkens W.H. (1995) Influence of Nonionic Surfactants on Bioavailability and Biodegradation of Polycyclic Aromatic Hydrocarbons *Appl. Environ. Microbiol.* 61:1699-1705.

[15] Volkering F., Breure A.M., Rulkens W.H. (1998) Microbiological Aspects of Surfactant use for Biological Soil Remediation *Biodegradation* 8:401-417.

[16] Mukerjee P., Mysels K.J. (1971) *Critical Micelle Concentrations of Aqueous Surfactant Systems*, Washington, National Standard Reference Data System-National Bureau of Standards.

[17] Johans C., Suomalanien P. (2003) *Delta-8 Multichannel Microtensionmeter*, Version 0.1, Instrument Instruction Manual, Kibron Inc.

[18] Fell J.T. (2007) Surface and Interfacial Phenomena. In: Aulton M.E ed, *Aulton's Pharmaceutics The Design and Manufacture of Medicines*, 3rd edition, New York, Churchill Livingstone Elsevier, pp 59-69.

[19] Dukhin A.S., Goetz P.J. (2006) How Non-Ionic "Electrically Neutral" Surfactants Enhance Electrical Conductivity and Ion Stability in Non-Polar Liquids *J. Electroanal. Chem.* 588: 44-50.
[20] Williams E.F., Woodbury N., Dixon J.K. (1957) Purification and Surface Tension Properties of Alkyl Sodium Sulfosuccinates *J. Colloid. Sci.* 12:452-459.
[21] Haffner F.D., Piccione G.A., Rosenblum C. (1942) Conductances of Solutions of Several Alkyl Sulfates and Sulfosuccinates *J. Phys. Chem.* 46: 662-670.
[22] Miller M.L., Dixon J.K. (1958) Conductance of Dialkyl Sodium Sulfosuccinate Surface-Active Agents *J. Colloid. Sci.* 13: 411-417.
[23] Roberts D.W. (1991) QSAR Issues in Aquatic Toxicity of Surfactants *Sci. Total Environ.* 109 110: 557-568.
[24] Roberts D.W. (2002) Application of Octanol/Water Partition Coefficients in Surfactant Science: A Quantitative Structure-Property Relationship for Micellization of Anionic Surfactants *Langmuir* 18: 345-352.
[25] Roberts D.W. (1991) Application of QSAR to Biodegradation of Linear Alkylbenzene Sulphonates (LAS) Isomers and Homologues *Sci. Total Environ.* 109-110: 301-306.
[26] Wong D.C., Dorn P.B., Chai E.Y. (1997) Acute Toxicity and Structure-Activity Relationships of Nine Alcohol Ethoxylate Surfactants to Fathead Minnow and *Daphnia magna Environ. Toxicol. Chem. 16:* 1970-1976.
[27] Lindgren Å., Sjöström M., Wold S. (1996) QSAR Modelling of the Toxicity of Some Technical Non-Ionic Surfactants towards Fairy Shrimps, *Quant. Struct.-Act. Rel.* 15: 208-218.
[28] Ahel M., Giger W. (1993) Partitioning of Alkylphenols and Alkylphenol Polyethoxylates Between Water and Orgranic Solvents *Chemosphere* 26: 1471-1478.
[29] Hodges G., Roberts D.W., Marshall S.J., Dearden J.C. (2006) The Aquatic Toxicity of Anionic Surfactants to *Daphnia magna* – A Comparative QSAR Study of Linear Alkylbenzene Sulophonates and Ester Sulphonates *Chemosphere* 63: 1443-1450.
[30] Hansch C., Leo A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, New York, Wiley-Interscience Publication.
[31] Ward R.S., Davies J., Hodges G., Roberts D.W. (2003) Applications of Immobilised Artificial Membrane Chromatography to Quaternary Alkylammonium Sulfobetaines and Comparison of Chromatographic Methods for Estimating the Octanol-Water Partition coefficient *J. Chromatogr. A* 1007: 67-75.
[32] SPARC (SPARC Performs Automated Reasoning in Chemistry) version 4.6 <http://ibmlc2.chem.uga.edu/sparc> [Accessed 13th December 2011]
[33] Vázquez G., Alvarez E., Navaza J.M. (1995) Surface Tension of Alcohol + Water from 20 to 50°C *J. Chem. Eng. Data* 40: 611-614.
[34] U.S. Environmental Protection Agency (2010) Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.10. United States Environmental Protection Agency, Washington, DC, USA.
[35] Tetko I.V., Gasteiger J.,Todeschini R., Mauri A., Livingstone D.J., Ertl P., Palyulin V.A., Radchenko E.V., Zefirov N.S., Makarenko A.S., Tanchuk V.Y., Prokopenko V.V. (2005) Vitual Computational Chemistry Laboratory – Design and Description *J. Comput. Aid. Mol. Des.* 19: 453-63.
[36] VCCLAB, Virtual Computational Chemistry Laboratory(2005) available from <http://www.vcclab.org>

[37] Carreira A., Hilal S.H., Karickhoff S.W. (1994) Estimation of Chemical Reactivity Parameters and Physical Properties of Organic Molecules Using SPAC. In Politzer P., Murray J.S. eds, *Theroetical and Computational Chemistry, Quantitative Treatment of Solute/Solvent Interactions*, Amsterdam, Elsevier, pp 291-336.
[38] Hartley G.S., Samis C.S. (1936) Transport Number of Paraffin-Chain Salts in Aqueous Solution. Part I. Measurement of Transport Numbers of Cetylpyridinium in Terms of Micelle Formation, With Some Data Also for Cetane Sulphonic Acid *Trans. Faraday Soc.* 32: 795-815.
[39] Scott A.B, Tartar H.V. (1943) Electrolytic Properties of Solutions of Parraffin-Chain Quaternary Ammonium Salts *J. Am. Chem. Soc* 65: 692-698.
[40] Haydon D.A., Taylor F.H. (1962) Adsorption of Sodium Octyl and Decyl Sulphates and Octyl and Decyl Trimethylammonium Bromides at the Decane-Water Interface *Trans. Faraday Soc.* 58: 1233-1250.
[41] Trap H.J.L, Hermans J.J. (1955) Light Scattering by Solutions of Some Kationic Soaps (Diffusion de la Lumière pas des Solutions de Quelques Savons Cationques) *Kkl. Nederl. Akad. Wetensch., Ser.* B 58: 97-108.
[42] Venable R.L., Nauman R.V. (1964) Miccellar Weights of and Solubilization of Benzene by a Series of Tetradecylammonium Bromides. The Effect of the Size of the Charged Head *J. Phys. Chem.* 68: 3498-3503.
[43] Klevens H.B. (1948) Critical Micelle Concentrations as Determined by Refraction *J. Phys. Colloid Chem.* 52: 130-148.
[44] Moreira L., Firoozabadi A. (2010) Molecular Thermodynamic Modeling of Specific Ion Effects on Micellization of Ionic Surfactants *Langmuir* 26: 15177-15191.
[45] Baloch M.K., Hameed G., Bano A. (2002) Effect of Electrolyte Concentration and Temperature on CMC of Surfactants *J. Chem. Soc. Pakistan* 2: 77-86.
[46] Chattopadhyay A., Harikumar K.G. (1996) Dependence of Critical Micelle Concentration of a Zwitterionic Detergent on Ionic Strength: Implications in Receptor Solubilization *FEBS Lett.* 391: 199-202.
[47] Mukerjee P. (1965) Salt Effects on Non-Ionic Association Colloids *J. Phys. Chem.* 69: 4038-4040.
[48] Personal communication with David Roberts, Liverpool John Moores University, Liverpool, UK
[49] Tetko I.V., Bruneau P. (2004) Application of ALOGPS to Predict 1-Octanol/Water Distribution Coefficients, Log P, and Log D, of AstraZeneca In-House Database *J. Pharm. Sci.* 93: 3101-3110.
[50] Yu Y., Yang W., Gao Z., Lam M.H.W., Liu X., Wang L., Yu H. (2008) RP-HPLC Measurement and Quantitative Structure-Property Relationship Analysis of the n-Octanol-Water Partitioning coefficients of Selected Metabolites of Polybrominated Diphenyl Ethers *Environ. Chem.* 5: 332-339.
[51] Meylan W.M., Howard P.H. (1995) Atom/Fragment Contribution Method for Estimation Octanol-Water Partitioning Coefficients *J. Pharm. Sci.* 84: 83-91.
[52] Mannhold R., Poda G.I., Ostermann C., Tetko I.V. (2009) Calculation of Molecular Lipohilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds *J. Pharm. Sci.* 98: 861-893.
[53] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-Octanol/Water), High Performance Liquid Chromatography (HPLC) Method*, Paris, Organisation for Economic Cooperation and Development OECD.

[54] Roberts J.F., Marshall S.J., Roberts D.W. (2007) Aquatic Toxicity of Ethoxylated and Propoxylated Alcohols to *Daphnia magna Environ. Toxicol. Chem.* 26: 68-72
[55] Hait S.K., Moulik S.P. (2002) Gemini Surfactants: A Distinct Class of Self-Assembling Molecules *Curr. Sci.* 82: 1101-1111

9 Discussion

This chapter summarises the key conclusions from Chapters 2-8, placing them in the context of the current regulatory position of chemical risk assessment, as well as discussing where future work could be directed to extend and further develop the applicability of IAM-HPLC as a surrogate for membrane partitioning in the determination of toxicity.

9.1 Summary

Traditionally the safety assessment of chemicals has been determined experimentally using the animal of interest, or through extrapolation from one species to another e.g. toxicity determined in rat and extrapolated to human toxicity. Due to many factors including, but not limited to, the high cost of experiments, ethical issues surrounding the use of animals and both national and international legislation, the use of animals in toxicity testing is becoming less favoured. For instance, the European Union's REACH legislation[1,2] and the Cosmetics Regulation[3] specifically promote the use of non-animal test data and alternatives to animal testing to assess a chemical's safety.

Alternatives to animal testing include *in vitro* and *in silico* methods. One of the most forward thinking initiatives in the REACH legislation is the positive promotion of *in silico* approaches. These include read across, category formation and the formation of (Quantitative) Structure-Activity Relationships ((Q)SARs). The impact of these *in silico* techniques has been highlighted in a recent European Chemicals Agency (EChA) report. This indicated that, following an analysis of the REACH dossiers for the high tonnage volume chemicals, there has been considerable use of what was termed "read-across". This accounted for between 20-30% of the information contained with the dossiers across all endpoints. The wide scale uptake of read-across was not anticipated and is, at least in part, due to the success of freely available software such as the OECD QSAR Toolbox[4, 5]. It has been postulated that if the use of non-animal testing (which includes the use of QSARs, read-across and test waiving) is maximised, opposed to the worst case where use of these tools is minimal, the use of animals in testing, could be reduced from 4.1 million to 1.6 million for mammalian animals and from 0.51 to 0.13 million non-mammalian animals under the REACH legislation. This is a considerable reduction in the use of animals, and would additionally lead to substantial savings of both cost and time[6].

There have been many QSARs developed to calculate physico-chemical properties and toxicity endpoints of interest for chemicals relating to both human health and the environment. The ability to calculate physico-chemical properties is important, as it increases the applicability of QSARs to chemicals for which the physico-chemical descriptor has not been previously determined experimentally, allowing for the prediction of toxicity using calculated properties.

Many QSARs previously developed and published in the peer reviewed scientific literature, particularly for the endpoints discussed in Chapters 6 and 7 (skin absorption and acute aquatic toxicity respectively) include a descriptor for hydrophobicity. Most commonly this is a measure of partitioning, specifically the logarithm of the octanol-water partition coefficient (log P), although other descriptors of hydrophobicity are also used, such as the logarithm of the distribution coefficient (log D) at a specified pH[7].

Hydrophobicity is a well characterised property of many chemicals, in part due to the systematic analyses of log P by Hansch and Leo[8]. In addition, there have been many advances in computational capability. These advances have led to many predictive methods to determine log P. Many have built on initial models by Hansch and Leo[8] using a constructionist approach and by Rekker[9] using a reductionist approach. Alternative approaches include the application of the Abraham solvatochromatic parameters[10, 11] and the development of associative neural networks [12, 13]. The ease of calculating log P as compared to experimental analysis extends the use of QSARs with log P as a descriptor. Therefore, the computational models can be used to estimate toxicity from chemical structure.

The use of hydrophobicity as a descriptor in QSARs predicting toxicity endpoints often assumes that log P is an appropriate surrogate for biological membrane partitioning. As discussed in Chapter 1, there are differences between octanol-water partitioning and partitioning across biological membranes. Not least, octanol-water partitioning describes a single partition, whereas partitioning in biological systems can be across multiple membranes. The use of log P in QSARs may, therefore, be a gross over-simplification of the complications of biological systems. Immobilised

Artificial Membrane (IAM) partitioning using High Performance Liquid Chromatography (HPLC) provides the possibility to describe both the hydrophobic and hydrophilic contributions involved in the transport across biological membrane(s). The ion pairing and hydrogen bonding contributions are not modelled with the use of octanol-water partitioning. IAM has the advantage of containing phosphatidylcholine, which is the primary phospholipid in many cells. Therefore, partitioning in the IAM columns relates to a more biologically relevant partitioning process.

9.1.1 Database of literature log k_{IAM} values and the effect of experimental variability on reported values

In order to assess the state of the art of IAM-HPLC, the existing published literature values for log k_{IAM} were collated into a database. The experimental parameters under which the log k_{IAM} values were obtained were recorded along with the citation. The database contains 1910 experimental log k_{IAM} values for 647 compounds (1686 isocratic IAM-HPLC log k_{IAM} values for 555 compounds) from 53 papers. There were obviously multiple values for some of the compounds. These compounds were investigated for comparability of the log k_{IAM} values reported. Due to the large range of log k_{IAM} values for individual compounds analysed under varying experimental conditions, the effect of experimental variability on reported log k_{IAM} values was investigated. The key parameters that should be consistent when comparing experimental log k_{IAM} values were determined to be column stationary phase, mobile phase and pH of the mobile phase. Due to the effect of experimental parameters on reported log k_{IAM} values identified, caution is required in combining and comparing log k_{IAM} datasets. In addition, the variability in reported log k_{IAM} values demonstrates the requirement for a standardised IAM-HPLC procedure. This would allow for the collation and comparison of log k_{IAM} values and aid in the development of QSARs to calculate toxicity endpoints

Investigation of the log k_{IAM} database determined that of 1686 log k_{IAM} values 824 were obtained using the IAM.PC.DD2 column. In addition to the IAM.PC.DD2 being the most commonly used column, when the database was split into subsets of data based on stationary phase, only the IAM.PC column demonstrated an improved correlation with log P compared to the IAM.PC.DD2 column. Despite the high

correlation of the IAM.PC column data and log P, the column is not end-capped. This may be disadvantageous because the end-capping process is used to improve the column's stability and separation capabilities. Non end-capped columns may result in shorter column lifetimes and poorer resolution for highly hydrophobic compounds. Additionally, the aim of the investigation was not to achieve a perfect correlation between log P and log k_{IAM}. It is the compounds for which the correlation breaks down that are of interest, as well as compounds for which the use of traditional methods for determining log P is either not recommended, or is difficult to determine accurately.

9.1.2 Optimisation and robustness testing of IAM-HPLC assay

In order to overcome the problem of experimental variability, an IAM-HPLC assay has been standardised using experimental conditions consistent with existing literature; specifically the reduced dataset of the log k_{IAM} database. The assay was shown to be capable of assessing compounds with a wide range of hydrophobicities, covering log $k_{IAM\ (pH\ 7.4)}$ values from -1.92 to 4.53 (equivalent to log P of -1.35 to 6.03). The assay initially analysed a selection of the OECD reference materials[14] for the determination of log P (RP-HPLC method) to investigate the relationship between the two measures of partitioning, for well characterised compounds. In this thesis the domain of the IAM-HPLC assay has been extended to include surfactants and compounds ionised under the conditions of analysis. There is increased confidence in the log $k_{IAM\ (pH\ 7.4)}$ values obtained using this standardised assay following robustness testing of the assay across five columns (which included three batches of stationary phase), two HPLC systems and five compounds covering the full range of hydrophobicities. The IAM-HPLC assay standardised in Chapter 3 has been demonstrated to be robust to changes in column, stationary phase batch and HPLC system.

The use of IAM-HPLC to determine log $k_{IAM\ (pH\ 7.4)}$ values experimentally has been demonstrated in Chapter 3 for a wide range of hydrophobicities. Analysis of the experimental log $k_{IAM\ (pH\ 7.4)}$ values illustrates that IAM-HPLC partitioning (in a similar fashion to log P) can differentiate between isomers (Figure 45) and confirms that IAM-HPLC partitioning is 3-D in nature as it accounts for steric attributes of a molecule.

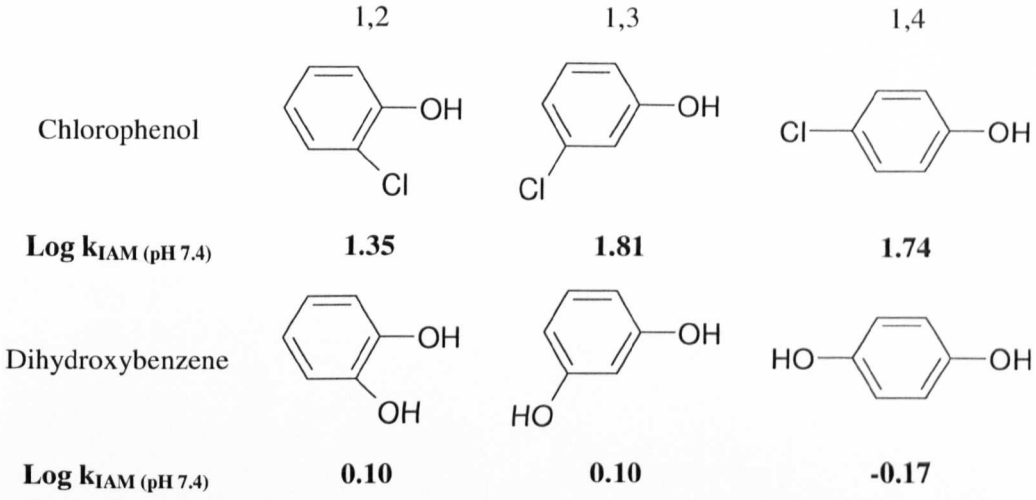


Figure 45 - The effect of substitution position on experimental log $k_{IAM\ (pH\ 7.4)}$ for chlorophenols and dihydroxybenzenes

9.1.3 Methods to predict log k_{IAM}

In order to extend the use of log $k_{IAM\ (pH\ 7.4)}$ as a measure of hydrophobicity, methods to predict log $k_{IAM\ (pH\ 7.4)}$ were investigated. Two approaches were applied, these were predictions based on chemical structure using fragments, and a traditional QSAR approach using physico-chemical and structural descriptors. The ability to predict log $k_{IAM\ (pH\ 7.4)}$ has a number of advantages. The applicability of log $k_{IAM\ (pH\ 7.4)}$ as a potential descriptor in QSARs is increased due to ease of access of the log $k_{IAM\ (pH\ 7.4)}$ values. Predicted log $k_{IAM\ (pH\ 7.4)}$ values aid the design of experiments to determine log $k_{IAM\ (pH\ 7.4)}$ values (i.e. determining which mobile phases are recommended for use etc.). If reliable predictions can be made for log $k_{IAM\ (pH\ 7.4)}$ the need to perform the IAM-HPLC analysis is negated, reducing time, resources and cost of determining log $k_{IAM\ (pH\ 7.4)}$. Additionally, as the use of log $k_{IAM\ (pH\ 7.4)}$ increases, its acceptance as a descriptor of hydrophobicity also increases, leading to wider domains of applicability of QSARs that contain log $k_{IAM\ (pH\ 7.4)}$ as a descriptor. The ability to predict log $k_{IAM\ (pH\ 7.4)}$ potentially extends the use of these QSARs to compounds for which log $k_{IAM\ (pH\ 7.4)}$ has not been determined experimentally.

The π substitution values were determined for log $k_{IAM\ (pH\ 7.4)}$ values from IAM-HPLC and compared to log P π values determined by Hansch and Leo[8]. A good correlation was identified between log P π and log $k_{IAM\ (pH\ 7.4)}$ π values. However,

there were three fragments for which the relative hydrophobicity in the system was different. The nitro group, acetyl group and methoxy groups are relatively more hydrophobic within the IAM system, compared to the octanol-water system. This indicates that fragments partition differently in the IAM system and octanol-water systems. The IAM system investigated appears to be able to interact with hydrogen bond donors/acceptors. These interactions act to increase the affinity of the analyte to the stationary phase leading to an increased retention time and hence log $k_{IAM\ (pH\ 7.4)}$ values.

In Chapter 4 the reductionist approach developed by Rekker[9] for predicting log P based on the chemical structure was applied to the compounds for which log $k_{IAM\ (pH\ 7.4)}$ was determined. The structures were broken down into theoretical structural fragments. The frequency of occurrence of each fragment was recorded along with structural features. Using multiple linear regression analysis, co-efficients were determined for each fragment and correction factor value considered. The applicability of the reductionist approach using fragment and correction factor values to predict hydrophobicity based on a chemical structure has been demonstrated for log $k_{IAM\ (pH\ 7.4)}$ as it is to log P. For the fragment and factor values determined for log $k_{IAM\ (pH\ 7.4)}$, CH_3 has a smaller co-efficient than CH_2 and is therefore described as less hydrophobic. This is opposed to the trend observed by both Hansch and Leo[8] and Rekker[9] and from that expected from principles of organic chemistry. To improve on the fragment and factor values determined, a larger data set is required. However, in addition, the compound/descriptor ratio also needs to increase. Increasing the number of compounds considered would increase the predictive applicability of log $k_{IAM\ (pH\ 7.4)}$ values whilst increasing the compound/descriptor ratio would reduce the high inter-dependency of fragment and factor values. Despite there being errors in the fragment values for CH_3 and CH_2, the fragment and factor values determined have good predictive abilities, considering the leave-one-out (LOO) cross validation approach. In addition, values for compounds not seen *a priori* by the method were well predicted by the fragment and correction factor values determined using the training set.

The use of fragment and factor values to predict log k_{IAM} values has the potential problem that novel compounds might contain fragment and factors that were not

included in the training set. These compounds would, therefore, be outside the domain of the method. Due to the number and type of compounds considered so far in the determination of log $k_{IAM\ (pH\ 7.4)}$ values, the domain is currently limited. Over time this would be overcome with further analysis of compounds using the standardised and robust IAM-HPLC assay optimised in Chapter 3.

The use of descriptors in a traditional QSAR approach to predict log $k_{IAM\ (pH\ 7.4)}$ was investigated in Chapter 5. Such an approach would potentially allow the prediction of log $k_{IAM\ (pH\ 7.4)}$ values for compounds which contain fragments or factors outside the domain of the fragment method developed in Chapter 4, provided they are within the domain of the QSAR. A large number of 2-D and 3-D descriptors were calculated for the compounds with log $k_{IAM\ (pH\ 7.4)}$ values either determined experimentally in this thesis, or from the reduced dataset of the log k_{IAM} database. Using stepwise regression the significant descriptors were identified, log P was identified as the most significant descriptor, the inclusion of additional descriptors in the QSARs did not improve the models developed.

If log P is excluded from the stepwise regression, the logarithm of the aqueous solubility (log S) becomes the most significant descriptor. However, use of log S and additional descriptors is significantly less predictive than the QSAR developed with log P as the sole descriptor. Additionally, log S is highly correlated with log P. The three additional descriptors identified by stepwise regression as being significant are the number of H-bond acceptors, molecular surface area and an estimate of total energy. The order of significance of these descriptors changes depending upon the inclusion of log P, log S, or no other descriptors in the QSAR. However, log P as the sole descriptor produces the most significant QSAR.

As discussed in Chapter 5 the applicability domain of the models to predict log $k_{IAM\ (pH\ 7.4)}$ for both the fragment and correction factor values, and the descriptor based QSAR approach can be defined in different ways, leading to different domains of applicability for each method. However, it is recommended that both methods to predict log $k_{IAM\ (pH\ 7.4)}$ use an applicability domain based on structure. Alternative definitions of the domain (e.g. based on the range of a descriptor for example log P, log S, etc.), although potentially increasing the chemical space, may also increase the

likelihood of poor predictions due to the potential introduction of novel functionality. If the number of compounds and the structural diversity of the training set increases, the method of determining the applicability domain of the descriptor based QSAR could be reassessed in future.

9.1.4 QSARs using log k_{IAM} as a descriptor to predict skin absorption

As discussed above, IAM-HPLC measurements account for partitioning interactions, such as hydrogen bond donor/acceptor interactions and ionic bonds, that are not adequately described by the octanol-water system. Hydrogen bond donor/acceptor interactions are important in skin absorption. Therefore, the use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor in QSARs to calculate the ability of a chemical to cross the skin barrier was investigated. Both the log $k_{IAM\ (pH\ 7.4)}$ values determined experimentally in this thesis and the existing literature log k_{IAM} values were considered, along with databases of skin absorption values. The skin absorption data were obtained from the EDETOX database[15], and both the Moss[16] and Flynn[17] datasets, that are also included within the EDETOX database. Both the Moss and Flynn datasets were used as independent sources as these authors report QSARs based on consistent data. In contrast, the EDETOX database is simply a compilation of all experimental results with methodologies published in the literature.

Subsets of log k_{IAM} and skin absorption values were collated based on consistent and comparable experimental parameters (as far as possible) for both log k_{IAM} and skin absorption. Despite there being a large volume of data available for both parameters, the overlap in compounds analysed with consistent and comparable data was small. The small overlap means the identification of outliers and potential removal of these compounds was not possible. It was found that for the subsets of data where valid QSARs could be developed, the use of two descriptors, generally, produced better QSARs than the use of a single descriptor for hydrophobicity. The use of either log k_{IAM}, or log P as a descriptor of hydrophobicity and molecular weight as a second descriptor for size, produced better QSARs. This indicates that partitioning in IAM does not sufficiently account for molecular size, which describes the membrane diffusion coefficient. This means that QSARs developed to predict skin absorption, using either log P, or log $k_{IAM\ (pH\ 7.4)}$ require a second descriptor to account for the membrane diffusion coefficient. Developing subsets for modelling skin absorption

data requires consistent and comparable data for both parameters. It was also noted that the QSARs developed where the compounds were unionised at both pH 5.5 and 7.4 (pH of human skin and HPLC analysis respectively) were statistically better than those developed considering both unionised and ionised compounds (Table 6 chapter 45). For ionised compounds the datasets did not produce valid QSARs; however, it is noted that the size of these datasets were small (n=4 and 8).

Due to the limited size of the datasets, it was not possible statistically to determine the predictive capabilities of the QSARs developed. To investigate the capability of IAM to predict skin absorption, and to determine the IAM conditions most suitable for skin penetration predictions, a larger dataset of consistent skin penetration data and log k_{IAM} values determined for each of the potential conditions is required. However, for the limited data available, QSARs developed using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor generally performed as well as, and in some cases better than those using log P as a descriptor.

9.1.5 QSARs using log k_{IAM} as a descriptor to predict acute aquatic toxicity

The use of log $k_{IAM\ (pH\ 7.4)}$ as a descriptor in QSARs for the prediction of the 'base set'[18] of aquatic test species was investigated. The publically available aquatic toxicity data are more consistent in methodology than data relating to skin absorption. However, similar issues regarding consistent experimental procedures for both aquatic toxicity and log $k_{IAM\ (pH\ 7.4)}$ were encountered. The requirement for data relating to both IAM partitioning and aquatic toxicity to be consistent and comparable led to small dataset for the development of QSARs. Due to both the small dataset size and the narrow range of hydrophobicities considered for polar narcotic compounds, many QSARs developed for the single mechanism of action were invalid (due to the F value being less than the relevant Fα value). For non-polar narcotic compounds, for all endpoints considered, statistically significant QSARs were developed. Log $k_{IAM\ (pH\ 7.4)}$ as a descriptor of hydrophobicity was either significantly better or marginally poorer than those using log P for the non-polar narcotic compounds depending on the species in question. It was also noted that non-polar narcotics were more common in the datasets and covered a wider range of hydrophobicity (5 to 6 log units), than for the polar narcotics (2 to 3 log units),

which may explain the development of valid QSARs for non-polar narcotics and invalid QSARs for polar narcotics.

Through the development of QSARs for both non-polar and polar narcosis as independent datasets based on the mechanism of action, and as a combined set for general narcosis, it was observed that modelling polar and non-polar narcotics separately leads to better and statistically more significant QSARs, compared to combining the mechanisms of action. This is contrary to the finding of Vaes *et al.*[19] who determined log K_{DMPC}, using a shake-flask procedure, rather than a HPLC procedure. The findings in this thesis were perhaps to be expected due to the observed difference in physiological responses between the two mechanisms of action[20, 21]. The non-polar narcosis mechanism of action causes general anaesthesia whereas, the polar narcosis mechanism of action causes muscular activity associated with seizures followed by cardiovascular-respiratory collapse. It is noted though that when toxicity is plotted against hydrophobicity the distinction between the narcosis mechanisms of action is less pronounced for log $k_{IAM\ (pH\ 7.4)}$ than it is for log P. It is useful to have a general model for all narcotics because is allows predictions to be made where the specific mechanism of action is unclear, a general narcosis model has a well defined and large applicability domain.

9.1.6 **Determination of hydrophobicity of surfactants using IAM-HPLC**

One of the classes of compounds that are specifically excluded from the traditional standardised methods for the determination of log P is surfactants (although not recommended[22], the log P of surfactants can be determined using these methods). The analysis of surfactants poses several problems. At concentrations above the critical micelle concentration (CMC) surfactants form micelles. The determination of log P and log $k_{IAM\ (pH\ 7.4)}$ needs to be assessed on the free form rather than the micellular. This is a more relevant value considering the expected environmental concentrations. It is toxicity to the environment that is of interest, which is intrinsically linked to the predicted environmental concentration. Additionally, the surfactant will order itself at the octanol-water interface. This effect is relative to the volume of solvents and the surface area of the interface, and although this does not affect the hydrophobicity value determined, the experimental set-up and procedure requires consideration for log P determination. A similar ordering to that observed in

octanol-water may be found in the environment at the lipid/aqueous interface in many biological species. Therefore, the potential for IAM-HPLC to determine the hydrophobicity of surfactants was investigated. For the determination of log $k_{IAM\ (pH\ 7.4)}$ of surfactants the CMC value needs to be known in advance to ensure analysis is performed below the CMC. Ward *et al.*[23] demonstrated that log k_{IAM} was a suitable measure of hydrophobicity for a series of closely related alkylammonium sulfobetaines (amphoteric surfactants). In this thesis log $k_{IAM\ (pH\ 7.4)}$ has been shown to be applicable to carboxybetaines, extending the domain of amphoteric surfactants. The domain has also been extended to include both cationic and anionic surfactants (quaternary ammonium bromides and sulfosuccinate sodium salts). No attempt to determine log $k_{IAM\ (pH\ 7.4)}$ of non-ionic surfactants has been made. Therefore, there is no knowledge on the applicability of log $k_{IAM\ (pH\ 7.4)}$ to non-ionic surfactants.

The analysis of single chain lengths is preferable to mixtures, because the more complicated interactions involved in the partitioning of surfactants (i.e. proximity of polar groups, which is group dependent, can complicate the additive nature of hydrophobicity). There are limited experimental log P values and toxicity data publically available for the compounds investigated here. Therefore, modelling to allow either the prediction of log $k_{IAM\ (pH\ 7.4)}$, or toxicity endpoints of interest was not possible. However, analysis of a short homologous series demonstrated that the fragment and correction factor method to predict log $k_{IAM\ (pH\ 7.4)}$ (developed on chapter 4) could be successfully applied to surfactants. To achieve this it is recommended that surfactants are analysed systematically to allow for the inclusion of surfactants into the fragment and factors approach to predict log $k_{IAM\ (pH\ 7.4)}$ discussed above and in Chapter 4. The inclusion of surfactants into the fragment and correction factor method requires the determination of new fragments (i.e. SO_3^-, N^+, etc.) and correction factors (i.e. proximity effect, branching and chain folding etc.)

9.2 Overview of findings

Although the requirement of consistent and comparable data for both log $k_{IAM\ (pH\ 7.4)}$ and either skin absorption, or environmental toxicity endpoints of interest resulted in small datasets, a good QSAR for modelling the absorption of chemicals across the skin barrier was developed (equation (6.33)). The domain of this QSAR has not been explicitly determined. However, the QSAR was developed for unionised

compounds. No valid QSARs could be developed for ionised compounds due to the small number of compounds considered.

For acute aquatic toxicity good QSARs were developed for the domain of non-polar narcotics, using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor, for the following endpoints: *Daphnia magna* EC_{50} (equations (7.2) and (7.6)), 96hr guppy LC_{50} (equations (7.12)), 14 day guppy LC_{50} (equations (7.16)), 96hr fathead minnow LC_{50} (equations (7.26)) and 48hr *Tetrahymena pyriformis* IGC_{50} (equations (7.30)). For polar narcotics one endpoint produced a good QSAR, 14 day guppy LC_{50} (equations (7.17)). For the majority of endpoints for polar narcotics a narrow range of hydrophobicity was considered, resulting in clusters of data. For 72hr *Pseudokirchneriella subcapitata* EC_{50} a QSAR was developed for the general narcosis mechanism of action (non-polar and polar narcosis combined) (equations (7.35)), the distinction between mechanisms of action was not observed for this endpoint. This was despite the fact that valid QSARs were not possible for either non-polar or polar narcotics individually (for all other endpoints considered, QSARs for individual mechanisms of action were significantly better than those for the combined narcosis mechanisms of action). General narcosis models were good for lots of the species considered here. General narcosis models are useful when the specific mechanism of action is unclear. Additionally, the combined mechanism of action QSAR also has a clear and large domain of applicability. For the combined mechanism of action log $k_{IAM\ (pH\ 7.4)}$ produced better QSARs than those using log P as a descriptor.

Chapter 6, 7 and 8 illustrate the potential of log $k_{IAM\ (pH\ 7.4)}$ to provide an equivalent or improved descriptor in QSARs. QSARs containing log $k_{IAM\ (pH\ 7.4)}$ as a descriptor that are equivalent to QSARs developed containing log P as a descriptor are especially useful if the QSARs cover a different domain of chemicals. This is especially true given that many methods of determining hydrophobicity either experimentally, or by calculation, specifically exclude surfactants from the domain. Therefore, the systematic analysis of surfactants and subsequent development of QSARs is recommended. This would considerably extend the classes of compounds for which QSARs are available for predictions. The development of QSARs using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor which is more biologically relevant than log P, may also result in more biologically relevant QSARs.

Together these applications help to fulfil a larger requirement of the REACH legislation, to protect human health, the environment and to promote non-animal testing. This is achieved through the development of a robust IAM-HPLC method and the development of QSARs based on log $k_{IAM\ (pH\ 7.4)}$ values, both of which are non-animal test methods. The use of IAM-HPLC and the subsequent QSARs to predict log $k_{IAM\ (pH\ 7.4)}$ has the potential to close knowledge gaps due to the advantages of this method (namely savings of time and cost) over more traditional methods of determining hydrophobicity. The combination of IAM-HPLC and QSAR models, allows for the prediction toxicity endpoints, additionally this saves time, cost and reduces and/or replaces the requirement for animal testing (for the endpoint of interest).

The novel aspects of this research include the development of a database of log k_{IAM} values (and subsequent investigation into the importance of consistent experimental procedures for the determination of log $k_{IAM\ (pH\ 7.4)}$). The development of a robust optimised IAM-HPLC assay and the use of log $k_{IAM\ (pH\ 7.4)}$ values to develop fragment and factor values that allow the prediction of log $k_{IAM\ (pH\ 7.4)}$ based solely upon a chemicals structure.

The industrially relevant findings and hence potential applications of this project include developing a standard operating procedure (SOP) for IAM HPLC for industrial chemicals, which has been shown to be robust across column, stationary phase batch and system of analysis, allowing for consistent analysis of chemicals using IAM-HPLC. The development of both a structure based and QSAR descriptor based method for the prediction of log $k_{IAM\ (pH\ 7.4)}$ has extended accessibility to log $k_{IAM\ (pH\ 7.4)}$ as a descriptor. This ability to predict log $k_{IAM\ (pH\ 7.4)}$ increases the usability of the QSARs developed for skin absorption and acute aquatic toxicity.

As the world moves away from the use of animals in the testing of chemical safety, the science needs to catch up. The use of IAM-HPLC will not independently replace the use of animals in the safety testing of chemicals. However, IAM-HPLC along with other methods (e.g. read-across) will reduce the use of animal testing initially and as the science develops may help in the replacement of animal testing in the

future. Given the ultimate aim is chemical safety there needs to be confidence in the replacement methods adopted. In the short term this will undoubtedly require the use of both non-animal and animal test data, until the replacements are accepted as a reliable alternative. Additionally, the replacement of animals in the testing of chemicals is only a benefit provided it does not put human health and the environment at an increased (and therefore, unacceptable) risk. Where the replacement of animal testing would increase the risk to man and the environment to an unacceptable level, the emphasis should be on reduction and refinement rather than replacement of animals in assessing chemical safety.

9.3 Future work

Given the potential benefits and applications of this project, there is much future work that could be undertaken to ensure the techniques progress and are applied. The future work includes the continual updating of published log k_{IAM} values in the database, using measurements obtained under consistent experimental conditions. There is a requirement for the systematic, experimental analysis of additional compounds. This would allow the extension of the fragment and factor values method developed in Chapter 4, through the inclusion of additional fragment and correction factors. Additionally, an increase in the compound/descriptor ratio would increase confidence in the fragments and factors determined. The systematic analysis should also consider the availability of toxicity data for both human health and ecological toxicity endpoints. This would allow QSARs to be developed with the ability to identify and remove outliers (as required), and allow for the definition of the applicability domain for skin absorption QSARs. Together this would increase confidence in the QSARs developed. There are many toxicity endpoints not considered in this thesis, for which QSARs could be developed using log $k_{IAM\ (pH\ 7.4)}$ as a descriptor. Log $k_{IAM\ (pH\ 7.4)}$ should be considered as an alternative descriptor for all endpoints for which log P has been used in modelling.

9.4 References

[1] European Commission (2006) *Off. J. Eur. Un.*, L 396/1 of 30.12.2006.
[2] European Commission Environment Directorate General (2007) REACH in brief, available from <http://ec.europa.eu/environment/chemicals/reach/pdf/2007_02_reach_in_brief.pdf> [Accessed 5th November 2010]
[3] European Commission (2003) *Off. J. Eur. Un.*, L 66/26 of 11.3.2003.
[4] Spielmann H., Sauer U.G., Mekenyan O. (2011) A Critical Evaluation of the 2011 ECHA Reports on Compliance with the REACH and CPL Regulations and on the Use of Alternatives to Testing on Animals for Compliance with the REACH Regulation *Altern. Lab. Anim.* 39: 481-493.
[5] European Chemicals Agency (2011) The Use of Alternatives to Testing on Animals for the REACH Regulation, available from < http://echa.europa.eu/doc/117reports/alternatives_test_animals_2011_en.pdf > [Accessed 3rd November 2011]
[6] European Commission eds. Van Der Jagt K., Munn S., Tørsløv J., de Bruijn J. (2004) *Alternative Approaches Can Reduce the Use of Test Animals under REACH Addendum to the report: Assessment of Additional Testing Needs Under REACH Effects of (Q)SARS Risk Based Testing and Voluntary Industry Initiatives*, available from < http://publications.jrc.ec.europa.eu/repository/handle/111111111/8790> [Accessed 15th December 2011]
[7] Cronin M.T.D., Aptula A.O., Duffy J.C., Netzeva T.I., Rowe P.H., Valkova I.V., Schultz T.W. (2002) Comparative Assessment of Method to Develop QSARs for the Prediction of the Toxicity of Phenols to *Tetrahymena pyriformis Chemosphere* 49: 1201-1221.
[8] Hansch C., Leo A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*, New York, Wiley-Interscience Publication.
[9] Rekker R.F. (1977) *The Hydrophobic Fragmental Constant*, New York, Elsevier.
[10] Plass M., Valko K., Abraham M.H. (1998) Determination of Solute Descriptors of Tripeptide Derivatives Based on High-Throughput Gradient High-Performance Liquid Chromatography Retention Data *J. Chromatogr. A* 803: 51-60.
[11] Sprunger L., Blake-Taylor B.H., Wairegi A., Acree Jr. W.E., Abraham M.H. (2007) Characterization of the Retention Behavior of Organic and Pharmaceutical Drug Molecules on an Immobilized Artificial Membrane Column with the Abraham Model *J. Chromatogr. A* 1160: 235-245.
[12] Tetko I.V., Bruneau P. (2004) Application of ALOGPS to Predict 1-Octanol/Water Distribution Coefficients, Log P, and Log D, of AstraZeneca In-House Database *J. Pharm. Sci.* 93: 3101-3110.
[13] Yu Y., Yang W., Gao Z., Lam M.H.W., Liu X., Wang L., Yu H. (2008) RP-HPLC Measurement and Quantitative Structure-Property Relationship Analysis of the n-Octanol-Water Partitioning coefficients of Selected Metabolites of Polybrominated Diphenyl Ethers *Environ. Chem.* 5: 332-339.
[14] The Organisation for Economic Cooperation and Development OECD (1994) *OECD Guidelines for the Testing of Chemicals, No. 117: Partition Coefficient (n-Octanol/Water), High Performance Liquid Chromatography (HPLC) Method*, Paris, Organisation for Economic Cooperation and Development OECD.
[15] EDETOX database available from <http://edetox.ncl.ac.uk/> [Accessed 4th April 2011]

[16] Moss G.P., Cronin M.T.D. (2002) Quantitative Structure-Permeability Relationships for Percutaneous Absorption: Re-analysis of Steroid Data *Int. J. Pharm.* 238: 105-109.

[17] Flynn G. (1990) Physicochemical Determinants of Skin Absorption. In Gerrity T.R., Henry C.J. eds., *Principles of Route-to-Route Extrapolation for Risk Assessment,* New York, Elsevier, pp.93-127.

[18] Hutchinson T.H., Barrett M.B., Constable D., Hartmann A., Hayes E., Huggett D., Laenge R., Lillicrap A.D., Straub J.O., Thompson R.S. (2003) A Strategy to Reduce the Numbers of Fish Used in Acute Ecotoxicity Testing of Pharmaceuticals *Environ. Toxicol. Chem.* 22: 3031-3036.

[19] Vaes W.H.J., Ramos E.U., Verhaar H.J.M., Hermens J.L.M. (1998) Acute Toxicity of Nonpolar Versus Polar Narcosis: Is There a Difference? *Environ. Toxicol. Chem.* 17: 1380-1384.

[20] McKim J.M., Bradbury S.P., Niemi G.J. (1987) Fish Acute Toxicity Syndromes and Their Use in the QSAR Approach to Hazard Assessment *Environ. Health Perspect.* 71: 171-186.

[21] Bradbury S.P., Henry T.R., Niemi G.J., Carlson R.W., Snarski V.M. (1989) Use of Respiratory-Cardiovascular Responses of Rainbow Trout (*Salmo gairdneri*) in Identifying Acute Toxicity Syndromes in Fish: Part 3. Polar Narcotics *Environ. Toxicol. Chem.* 8: 247-261.

[22] Short J., Roberts J., Roberts D.W., Hodges G., Gutsell S, Ward R.S. (2010) Practical Methods for the Measurement of Log P for Surfactants *Ecotox. Environ. Safe.* 73: 1484-1489.

[23] Ward R.S., Davies J., Hodges G., Roberts D.W. (2003) Applications of Immobilised Artificial Membrane Chromatography to Quaternary Alkylammonium Sulfobetaines and Comparison of Chromatographic Methods for Estimating the Octanol-Water Partition coefficient *J. Chromatogr. A* 1007: 67-75.