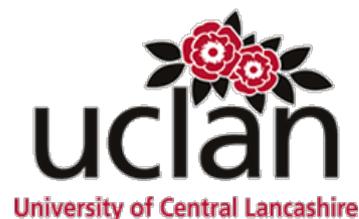# A Multidisciplinary Investigation into the Design, Synthesis and Evaluation of a Novel Class of Anti-Glioblastoma Drug Fragments

Christopher Sherer

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire

January 2017

STUDENT DECLARATION FORM

Concurrent registration for two or more academic awards

Either

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

Or

~~*I declare that while registered for the research degree, I was with the University's specific permission, a *registered candidate/*enrolled student for the following award:~~

_____

Material submitted for another award

Either

*I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work

Or

~~*I declare that the following material contained in the thesis formed part of a submission for the award of~~

_____

delete as appropriate

Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

Signature of Candidate _____

Type of Award   Doctor of Philosophy

School   School of Pharmacy and Biomedical Sciences

ABSTRACT

Cancer is the second biggest global killer,[1,2] with cancers of the brain and central nervous system accounting for a disproportionately high number of deaths.[3] The most prolific cancer of the central nervous system is glioblastoma, for which prognosis is still very poor. In this project, analogues of two lead compounds with known activity against glioblastoma cell lines (compounds **4** and **5**, Figure 1) were produced in order to develop structure-activity relationships and discover compounds with superior activities against glioblastoma.
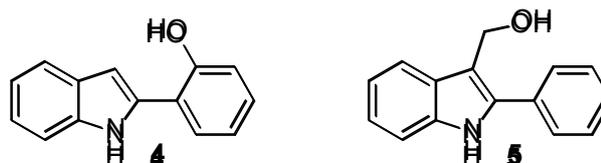

*Figure 1 - The structures of the lead compounds **4** and **5***

Analogues of compound **4** were the result of a rigorous similarity search of the ZINC database,[4,5] as well as using chemical intuition to identify potential analogues. A scaffold-hopping approach was undertaken, through which two new compound classes were identified as potentially superior lead compounds for future work (Figure 2).
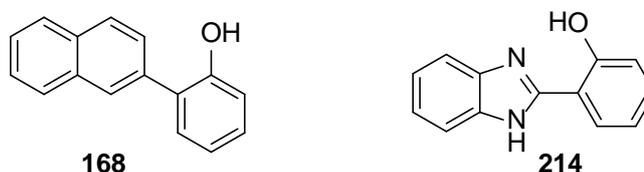

*Figure 2 - The structures of two analogues of compound **4** with different scaffolds and superior activity, compounds **168** and **214***

Compound **4** is known to induce cell death through the induction of elevated levels of cellular reactive oxygen species (ROS),[6] which may be formed *via* the radical form of compound **4** and its analogues. The connection between the anticancer activity of **4** and its analogues with the propensity of these compounds to form radicals was also investigated. Enthalpic values relevant to radical formation (BDE, AIP, PDE, PA and ETE) were calculated using a density functional theory (DFT) approach. Although no strong correlation was found for the whole series of compounds, the data indicates that correlations may exist within certain structural classes.

The anticancer activity of compound **5**, a prodrug, was compared against 11 analogues of both the prodrug and active form of the compound (Scheme 1). It was found that compound **9** has superior activity to that of the prodrug **5**. Substitutions at the *N*-position of **5** were also found

to have a significant effect on activity, with an *N*-tosyl analogue having significantly improved activity against glioblastoma cell lines and short term cultures. The results obtained suggest that future work on this series should therefore be based around compound **9**, a subclass of indoles that have wide ranging anticancer activity, but have not yet been reported against glioblastoma.
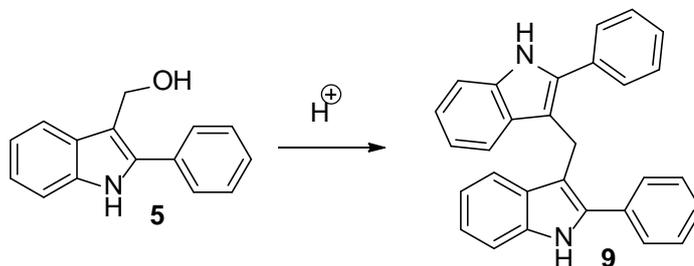
*Scheme 1 - The degradation of **5** into its suspected active form (**9**)*

In conclusion, analogues were discovered within this project which improved upon the anticancer activity of both compounds **4** and **5**. For compound **4**, two alternative scaffolds were identified as superior and novel lead compounds against glioblastoma, and there is some indication that there may be a correlation between radical formation and anticancer activity within specific structural classes of this functional class of compounds.   For prodrug **5**, substituents at the *N*-positon were found to have a significant effect on activity, and the activity of the active form (**9**) was found to be superior to the activity of the prodrug.

CONTENTS

LISTS OF ILLUSTRATIVE MATERIALS

List of figures

List of schemes

List of tables

List of equations

List of graphs

ACKNOWLEDGEMENTS

ABBREVIATIONS, ACRONYMS AND INITIALISMS

ACN – Acetonitrile

ACS – The American chemical society

AIC – 5-Aminoimidazole-4-carboxamide

AIP – Adiabatic ionisation potential

APCI – Atmospheric pressure chemical ionisation

BBB – Blood-brain barrier

BDE – Bond dissociation enthalpy

CAS – Chemical abstract service

CES1 – Carboxylesterase 1

DCM – Dichloromethane

DEPT – Distortionless Enhancement by Polarisation Transfer

DFT – Density functional theory

DIM – 3,3'-Diindolylmethane

DMF – Dimethylformamide

DMFDMA – N,N-dimethylformamide dimethyl acetal

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

EC_{50} – Half maximal effective concentration

EI – Electron ionisation

EPSRC – The Engineering and Physical Sciences Research Council

ESI – Electrospray ionisation

ET – Electron transfer

ETE – Electron transfer enthalpy

EtOAc – Ethyl acetate

EWG – Electron withdrawing group

FDA – The United States Food and Drug Administration

GBM – Glioblastoma multiforme

GC-MS – Tandem gas chromatography-mass spectrometry

GGA – Generalised gradient approximation

HAT – Hydrogen atom transfer

HF – Hartree-Fock

HF-SCF – Hartree-Fock self-consistent field

HMBC – Heteronuclear multiple-bond correlation spectroscopy

HRMS – High resolution mass spectrometry

HSQC – Heteronuclear single quantum coherence spectroscopy

I3C – Indole-3-carbinol

IC_{50} – Half maximal inhibitory concentration

IR – Infrared

LBDD – Ligand-based drug design

LCAO – Linear correlation of atomic orbitals

LC-MS – Tandem liquid chromatography-mass spectrometry

LDA – Local density approximation

LogP – Partition coefficient

m.p. – Melting point

MEM – Minimum essential medium

Mesyl – Methanesulfonyl

MGMT - O^6-methylguanine-DNA methyltransferase

MITC – 3-Methyl-(triazen-1-yl)imidazole-4-carboxamide

MS – Mass spectrometry

MTS – 3-(4,5-Dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium

MTT – 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide

NADH – Nicotinamide adenine dinucleotide

NADPH – Nicotinamide adenine dinucleotide phosphate

NEAA – Non-essential amino acids

NMR – Nuclear magnetic resonance

PA – Proton affinity

PCA – Principle component analysis

PDE – Proton dissociation enthalpy

PE – Petroleum ether (40 to 60 ^{o}C)

PL – Proton loss

PMS – Phenazine methosulfate

PT – Proton transfer

QSAR – Quantitative structure-activity relationship

R_f – Retardation factor

RNS – Reactive nitrogen species

ROS – Reactive oxygen species

ROS – Reactive oxygen species

SAR – Structure-activity relationship

SBDD – Structure-based drug design

SCF – Self-consistent field

SET – Single electron transfer

SET-PT – Sequential electron transfer proton transfer

SPLET – Sequential proton loss electron transfer

SRB – Sulforhodamine B

STAB – Sodium triacetoxyborohydride

TBS or TBDMS – tert-butyldimethylsilyl

TCA – Trichloroacetic acid

TCT – cyanuric chloride

TEA – Triethyl amine

TF – Thomas-Fermi

THF – Tetrahydrofuran

TLC – Thin layer chromatography

Tosic – *p*-Toluenesulfonic

TsOH – *p*-Toluenesulfonic acid

UEG – Uniform electron gas

UV – Ultraviolet

WHO – World Health Organisation

XC – Exchange-correlation

XRD – X-ray diffraction

CHAPTER 1 – INTRODUCTION

1.1   Cancer

Cancer is one of the biggest global killers, responsible for 8.2 million deaths in 2012,[1] the most recent year for which reliable data is available. This makes cancer the second biggest global killer after cardiovascular disease.[2] Reproductive, dietary, metabolic, hormonal and age-based factors are expected to drastically increase cancer incidences between now and 2030.[7]

The most common types of cancer internationally are lung (13%), breast (12%) and colorectal (9.8%).[8] Although relatively uncommon, cancers of the brain and central nervous system (1.7%) have disproportionately high mortality rates. According to Cancer Research UK, cancers of the central nervous system accounted for 9,365 of the 331,487 cases (2.83%) in the UK in 2011 (with 2011 being the most recent year for which such data is available), while accounting for 5,187 of the 161,823 deaths (3.21%). The ten-year net survival of malignant brain tumours for patients diagnosed in 2011 is expected to be around 13.5%, comparing poorly with the predicted value of 49.8% for all cancer patients.[3] As age is a major risk factor for many types of cancer including many cancers of the central nervous system,[9] incidence rates are expected to continue their upward trend as life expectancy continues to increase.

Tumours found in the brain may or may not have originated there, and one main way to classify tumours is based on where the tumour originated. A primary tumour is one which originated at the site, and a secondary tumour is one which originated elsewhere in the body, then metastasised to the brain. Metastatic tumours have the properties of the original tumour, for example a tumour of the lung which metastasised to the brain would be made up of mutated lung cells, not of mutated brain cells.

There are over 150 types and subtypes of tumours of the central nervous system,[10] which can be classified in a number of ways. A common way of classifying tumours is based on its "tumour grade", and brain tumours are often graded on a four-tier scale (going from Grade I to Grade IV). Grading is done by inspecting the cancerous cells under a microscope for similarity to the healthy cell, and the more dissimilar the cancerous cell is to a healthy cell (referred to as a "poorly differentiated" or "anaplastic" cell), the higher the grade it is given. The more similar a cancerous cell looks to its healthy counterpart, the slower the bulk tumour is likely to develop. The more abnormal the cancerous cell looks, the more likely the tumour is to grow aggressively. High-grade tumours (often called "malignant" tumours) are relatively fast growing, have a larger chance of recurrence (more likely to come back after removal), are more likely to metastasise, are more difficult to treat and are therefore more dangerous to the patient. Low-grade tumours

(often called "benign" tumours) are relatively slow growing, are less likely to recur, are less likely to metastasise, are typically easier to treat and therefore pose much less of a danger to the patient.

1.1.1   Glioblastoma

Of the 155 types and subtypes of tumours of the central nervous system, the most prolific is glioblastoma. Glioblastomas begin in the brain as mutated astrocytes, a type of glial cell named after their star-like appearance. They were first described as glioblastoma multiforme (GBM) in 1926,[11] with the name alluding to the belief that the tumour originated from a glioblast, an undifferentiated glial cell. The "multiforme" component of the name was due to the variation in appearance between cells within the same tumour, making them appear to have multiple forms.

Glioblastomas are a WHO Grade IV tumour,[10] the most aggressive grade. Coupled with the difficulty in treating these tumours and their sensitive location, the prognosis for patients diagnosed with glioblastoma is often very poor. Prognoses have improved steadily over recent years due to improvements in surgical methods, radiotherapy and the creation and development of better chemotherapies, but are still very poor.

The standard treatment procedure is currently maximal safe surgical resection followed by simultaneous radiotherapy to the resection site and chemotherapy with temozolomide, in what is sometimes known as the "Stupp protocol" after Roger Stupp, who first described this method.[12] The Stupp protocol leads to a mean survival of 14.6 months and a two year survival of 26.5%.[12] Glioblastomas are however largely considered to be currently incurable.
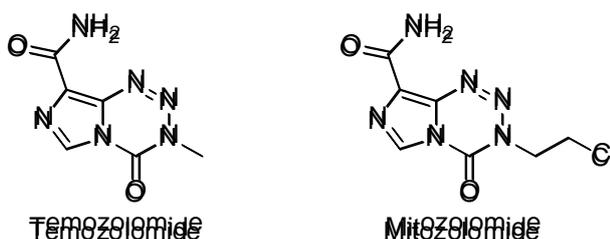
Figure 3 - The structures of temozolomide (left) and mitozolomide (right), two imidazotetrazines that were shown to have broad-spectrum antitumour activity

Temozolomide, the current first line therapy for glioblastomas, was first published by the Stevens group at the University of Aston in Birmingham (now simply called "Aston University") in 1984 alongside a series of other imidazotetrazines that showed broad-spectrum antitumour activity (Figure 3)[13]. Temozolomide was actually the second compound from this publication

to make it to clinical trials, but after the similarly structured mitozolomide showed unacceptable toxicity in phase II trials,[14–16] temozolomide was entered into phase I trials in 1992.[17] FDA approval for the use of temozolomide was originally given in 1999 for use on anaplastic astrocytomas, and in 2005 for use on newly diagnosed glioblastomas.
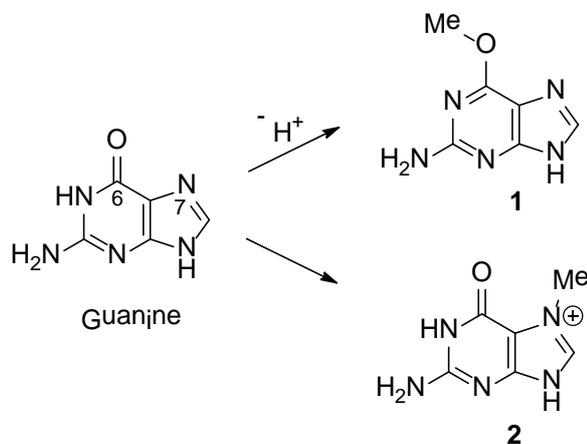
*Scheme 2 - Guanine (left) and the products of O^6 methylation (**1**) and N^7 methylation (**2**)*

Temozolomide is a DNA alkylating agent, meaning it can damage DNA by adding alkyl groups to some residues. In the case of temozolomide, it is known to add a methyl group typically at the O^6 or N^7 position of guanine residues of DNA (Scheme 2). This is achieved because temozolomide is a prodrug, and degrades under physiological conditions to form 3-methyl-(triazen-1-yl)imidazole-4-carboxamide (MTIC). MTIC itself degrades into 5-aminoimidazole-4-carboxamide (AIC) and the methyldiazonium ion (Scheme 3).
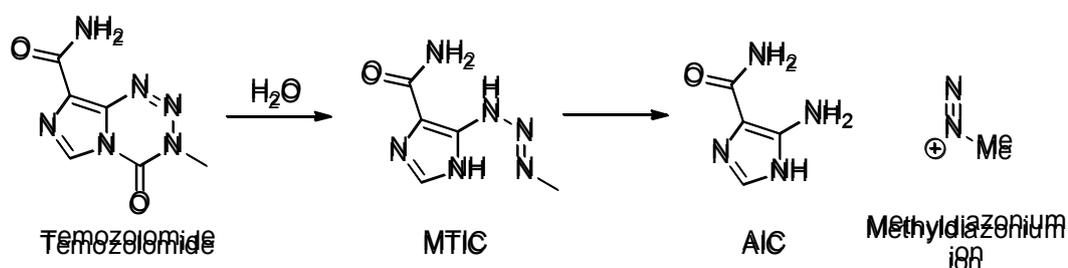
Scheme 3 - The degradation of temozolomide to MTIC and finally to AIC and the methyldiazonium ion

This methyldiazonium ion is the species which is responsible for the actual methylation of DNA, as shown in Scheme 4.
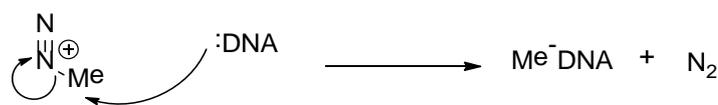
Scheme 4 - Methylation of DNA by the methyldiazonium ion

1.1.2   The Blood-Brain Barrier

Aside from the obviously sensitive nature of the surroundings of a glioblastoma tumour, a further issue to consider when designing chemotherapies is the blood-brain barrier (BBB, shown in Figure 4). The BBB is a semi-permeable membrane separating brain cells (and by extension brain tumours) and the extracellular fluid from the blood stream, and exists solely in the central nervous system. The BBB is predominantly composed of endothelial cells, much like the rest of the circulatory system, but in the brain, the endothelial cells are bound to one another with "tight junctions", meaning that compounds can't pass between the endothelial cells, as they can elsewhere in the circulatory system. The BBB also includes pericytes, which cover around a third of the surface of the capillary, and have roles in the formation and maintenance of the BBB. Finally, the capillaries are kept structurally sound and held in place by the foot processes of astrocytes.
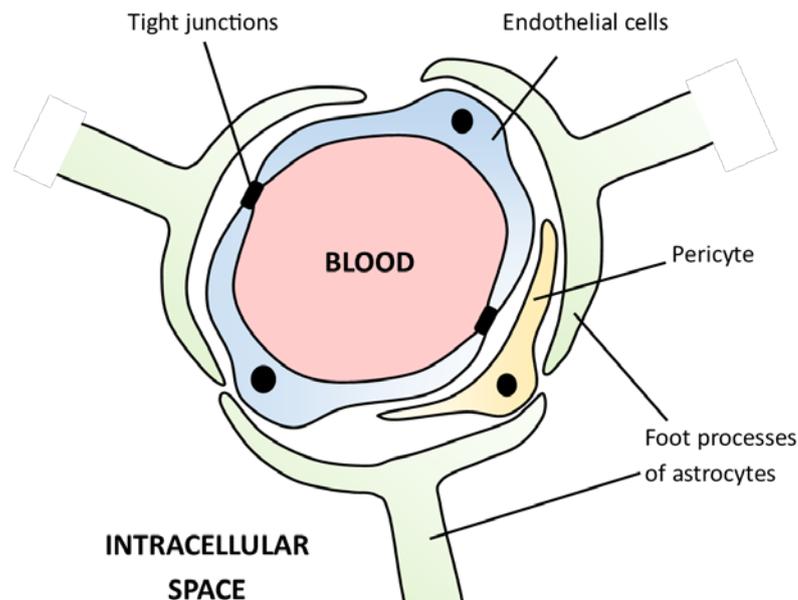
Figure 4 - The structure of the blood-brain barrier [*]

The task of the BBB is to prevent chemicals harmful to the brain (often called "neurotoxins") from making it to the brain, while allowing water, certain gasses, sugars, amino acids and some lipid-soluble compounds including hormones through. It is therefore highly selective, and poses a significant problem when attempting to administer drugs *via* a route that utilises the blood stream.

---

[*] Original graphic created solely for this thesis

Due to its small size and lipophilicity,[18] temozolomide readily permeates the blood-brain barrier.[19] The same cannot be said for another glioblastoma chemotherapy carmustine (Figure 5), although a workaround has been developed which bypasses the blood-brain barrier completely. Carmustine, as part of a degradable polymer wafer, is applied directly to the resection site during surgery. As the polymer breaks down, the drug is released in a controlled manner directly to the remaining cancer cells.[20]

Temozolomide   Carmustine
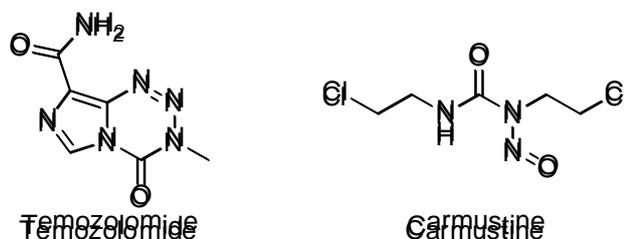
Figure 5 - The structures of two glioblastoma chemotherapeutic drugs, temozolomide and carmustine

1.1.3   Chemoresistance

A further problem with chemotherapies that is especially pervasive with respect to treating glioblastomas is the tendency for cancer cells to become resistant to anticancer drugs, in a process known as chemoresistance. Due to the apparent propensity for glioblastomas to become chemoresistant, a phenomenon that is still not fully understood,[21] more chemotherapies will always be required.

The primary form of chemoresistance with the current first line chemotherapy temozolomide is that the DNA methylation shown in Scheme 4 can be undone by the protein O^6-methylguanine-DNA methyltransferase (MGMT). The MGMT protein is essential in humans for maintaining genomic stability, as it can repair the naturally occurring O^6-methylguanine back to guanine, thus preventing transcription errors. However, as mentioned earlier, temozolomide's mode of action is to induce this mutation in order to damage cancerous cells. There are two potential solutions to this form of chemoresistance.
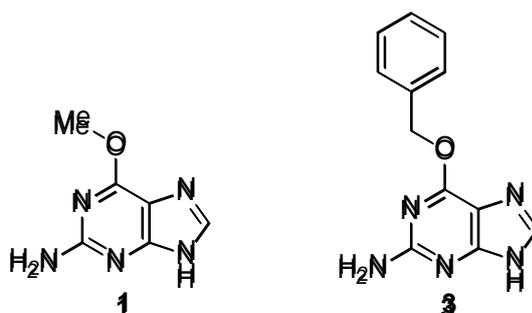
*Figure 6 - O^6-Methylguanine (**1**) and the inactivator of MGMT O^6-benzylguanine (**3**)*

The first potential solution is to prevent the action of MGMT. This is an active area of research, and the most promising method discovered thus far is using O^6-benzylguanine (Figure 6) to inactivate MGMT by irreversibly binding to it, thus meaning that guanine residues alkylated at the O^6 position remains alkylated.[22] By coadministering O^6-benzylguanine with an alkylating agent, chemoresistance due to MGMT can be prevented,[23] and this method has made it all the way to clinical testing with promising results.[24,25]

The second potential solution to chemoresistance is simply to have more and varied therapies. So far, alkylating agents are the best therapy for glioblastomas, but in the cases where MGMT-based chemoresistance is problematic, there is currently very little that can be done. By having diverse chemotherapies available, the potential for running out of options is decreased. This is the ultimate justification for this thesis.

Although the mechanisms of action for the prevailing standard chemotherapies for glioblastomas (temozolomide and carmustine) are as DNA alkylators, a growing number of active compounds are found to induce cell death *via* oxidative stress, by inducing the generation of reactive oxygen species.[26]

1.2 Reactive Oxygen Species

The term reactive oxygen species (ROS), though poorly technically defined, is often taken as a collective term for small oxygen-containing radicals such as superoxide ($O_2^{\bullet-}$), the hydroxide radical ($HO^\bullet$), alkoxyl radicals ($RO^\bullet$) or peroxyl radicals ($ROO^\bullet$). Small, non-radical oxygen-containing compounds that are either oxidisers, or can easily be converted into radicals, are often also considered to be reactive oxygen species, such as ozone (O_3), singlet oxygen ($^1O_2$) and hypochlorous acid (HOCl).[27,28] Analogous nitrogen-containing radicals such as the nitric oxide radical ($^\bullet NO$) are referred to as reactive nitrogen species (RNS), but due to RNS being so physiologically similar to ROS, RNS are often implicitly included as ROS.

Physiologically, ROS are essential for healthy cell functions and maintaining homeostasis,[29,30] as well as being implicated in many diseases associated with the ageing process,[31–33] Alzheimer's disease [34–36] and cancer.[37–40] It is this paradox that makes ROS-based therapies both promising and challenging.

Within the cell, ROS can be generated under either physiological or pathological conditions. Major physiological sources for ROS generation include the electron transport chain during oxidative phosphorylation in mitochondria, as a by-product of protein oxidation in the endoplasmic reticulum, and as a by-product from some oxidases in peroxizomes.[41] Major pathological sources of ROS include oxidising water to a range of ROS *via* ionising radiation,[42,43] in inflamed tissues during a respiratory burst (also known as an oxidative burst),[44] or due to unbound metals in the body *via* the Fenton reaction (Scheme 5).[45]

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + HO^\bullet + HO^-$$

Scheme 5 - The Fenton reaction, generating the hydroxyl reactive oxygen species

Within the cell, a major deleterious effect of ROS is the autoxidation of unsaturated lipids (denoted as RH in Scheme 6), which is a natural process whereby a ROS initiates a radical chain reaction that results in the degradation of unsaturated lipids in a cell membrane, causing damage to the cell. The initiation step of this chain reaction rarely happens spontaneously, but the hydrogen radical can be abstracted by a ROS or other radical species (labelled as "X•" in Scheme 6.**I**). This chain reaction is so destructive because the termination steps (Scheme 6.**IV**, Scheme 6.**V** and Scheme 6.**VI**) rarely happen due to the high relative concentration of RH, meaning that the step shown in Scheme 6.**III** happens many times before the chain reaction is terminated.

Initiation	RH + X• → R• + XH	(**I**)
Propagation	R• + O_2 → ROO•	(**II**)
	ROO• + RH → R• + ROOH	(**III**)
Termination	R• + R• → RR	(**IV**)
	R• + ROO• → ROOR	(**V**)
	ROO• + ROO• → ROOR + O_2	(**VI**)
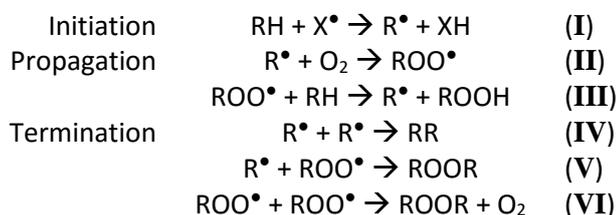
Scheme 6 - Steps in the radical chain reaction for the autoxidation of lipids

1.2.1 Using ROS to kill cancer cells

Cells are highly adaptive to their environment. Although there is a "sweet spot" with regard to ROS levels within a cell, small deviations from these levels lead to low levels of oxidative stress. This oxidative stress can be redressed by the cell in the form of an adaption such as increased

proliferation[46] or the synthesis of mutant proteins beneficial for survival under the new conditions (also known as adaptive translation).[47] More significant deviations in ROS levels leads to higher levels of oxidative stress, which can lead to cell senescence,[48] apoptosis,[49,50] necrosis [51] or autophagic cell death.[52,53] It has been shown that ROS-based therapies can be more effective, or even selective, in cancerous cells that have naturally elevated ROS levels or naturally decreased antioxidant levels.[28]

ROS-based therapies are far from novel, and have been in use since as early as 1969 when procarbazine was approved by the FDA.[54] There have since been many more examples of such drugs, including drugs that cause ROS-induced apoptosis (adenosine, doxorubicin, imexon),[55–57] ROS-induced complement-dependant cytotoxicity (rituximab),[58] ROS-induced senescence (cisplatin)[59,60] and ROS-induced autophagy (gemcitabine/cannabinoid combination).[61]

1.2.2   Antioxidants

Antioxidants are chemicals that prevent oxidation reactions. Although the term is typically associated with dietary or endogenous antioxidants, antioxidants have roles as diverse as preventing fats in packaged food going rancid,[62] stabilising fuels[63,64] and preventing the degradation of rubbers.[65] In each of these roles, some sort of oxidation reaction (either a loss of electrons, loss of hydrogen or gain of oxygen) is prevented or preferentially occurs on a sacrificial antioxidant.

As previously mentioned, ROS homeostasis (and by extension oxidative stress) is maintained both by regulating ROS generation and regulating antioxidant generation; higher levels of ROS production can be counteracted with higher antioxidant levels and vice versa. Such antioxidants can act in three different ways, and are classified based on this.

The first way in which antioxidants can act is as oxygen scavengers. In food preservation, a modified atmosphere of reduced oxygen levels or an atmosphere that is absent of oxygen altogether leads to prevention of food rancidity and reduction in the levels of many bacteria and fungi.[66] By employing a modified atmosphere within the packaging, the shelf life of the food can be drastically increased. Although limiting oxygen levels in packaging is common practice within the packaged food industry, creating oxygen-free conditions *in vitro* would not be a beneficial property of a physiological antioxidant, so this mechanism of antioxidant activity is not relevant to this work.

The second way in which antioxidants can act is by chelating to unbound metals in the body. Free metals in the body, particularly free iron or free copper, can be responsible for generating ROS via the Fenton reaction,[67] as has been previously mentioned. Conditions such as haemochromatosis or Wilson's disease are caused by elevated levels of iron or copper respectively in the body, which leads to elevated ROS levels and tissue damage. Compounds that can chelate these metal ions can act as preventative antioxidants by preventing the metals from generating ROS in the first place, and such therapies for conditions such as haemochromatosis and Wilson's disease are known.[68–71]

The final and most physiologically relevant way in which antioxidants can act is by acting as a free radical terminator. Free radical terminators are sacrificial antioxidants, meaning a single antioxidant molecule can terminate a single radical chain reaction. In the radical chain reaction shown in Scheme 6, this would mean quenching one of the two radical species (i.e. $R^\bullet$ or $ROO^\bullet$), and the steps involved with radical chain reaction termination by an antioxidant (AH) is shown in Scheme 7.

$$ROO^\bullet + AH \rightarrow ROOH + A^\bullet \qquad \textbf{(I)}$$
$$RO^\bullet + AH \rightarrow ROH + A^\bullet \qquad \textbf{(II)}$$
$$ROO^\bullet + A^\bullet \rightarrow ROOA \qquad \textbf{(III)}$$
$$RO^\bullet + A^\bullet \rightarrow ROA \qquad \textbf{(IV)}$$
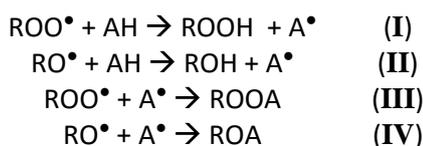
Scheme 7 - Termination of radicals by antioxidant AH

As can be seen in Scheme 7.**III** and Scheme 7.**IV**, the radical form of the antioxidant ($A^\bullet$) appears as the quenching species. If the antioxidant radical is stable enough, $A^\bullet$ can also play the role of $X^\bullet$ in Scheme 6.**I**. In other words, these compounds can act as both pro-oxidants and antioxidants in a concentration-dependant manner.[72,73]

Phenolic compounds can therefore act as pro-oxidants by forming their respective radical species. This could hypothetically occur *via* a number of potential pathways, as shown in Scheme 8. The simplest way that this could occur is by homolytic cleavage of the O-H bond, and subsequent loss of a hydrogen radical in a process known as "Hydrogen Atom Transfer" (HAT). This process therefore yields two compounds that could either terminate or initiate a radical chain reaction. The next process, along the top of Scheme 8, is the sequential loss of a proton then an electron, and is known as "Sequential Proton Loss Electron Transfer" (SPLET). It doesn't produce a hydrogen radical, but it does produce a lone electron, which could generate a ROS (for example, the electron may partake in the partial reduction of O_2 to superoxide, a known ROS). The final pathway for phenolic radical formation is the loss of an electron followed by the loss of a proton, in a process known as "Sequential Electron Transfer Proton Transfer" (SET-PT).

Like with SPLET, it produces a phenolic radical and an electron capable of forming a ROS, but has a different intermediate.
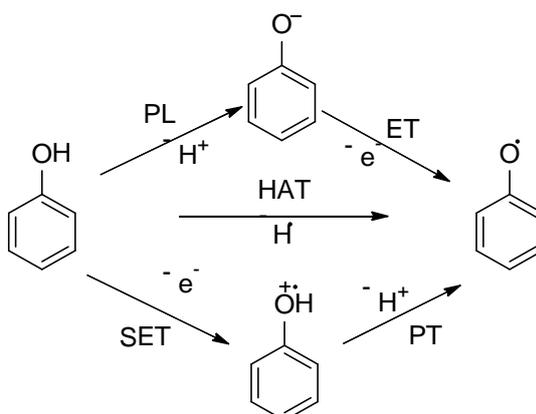
Scheme 8 - Possible ways in which a phenolic compound could form its respective radical species. HAT = Hydrogen Atom Transfer. PL = Proton Loss. ET = Electron Transfer. SET = Single Electron Transfer. PT = Proton Transfer

When studying these mechanistic steps, the enthalpy of each step has a specific name; bond dissociation enthalpy (BDE) for HAT, proton affinity (PA) for PL, electron transfer enthalpy (ETE) for ET, adiabatic ionisation potential (AIP) for SET and proton dissociation enthalpy (PDE) for PT. These enthalpic values are often determined computationally, and are plotted against antioxidant activity in an attempt to find correlations. BDE is often found to correlate with antioxidant activity,[74,75] and BDE and IP are occasionally found to correlate anticancer activity.[76]

This project includes work on a lead compound which is known to promote the formation of ROS (which will be discussed later). An attempt will therefore be made to correlate the activity of this lead compound and its analogues with the enthalpic values previously described in order to investigate if the class of compounds under investigation can be considered to be directly responsible for the generation of ROS.

1.3   The Lead Compounds

Indoles have long since been known to have anticancer properties through a variety of mechanisms including as topoisomerase inhibitors,[77–88] PKC inhibitors,[89–96] PDGF signal transduction inhibitors,[97] methuosis initiators,[98–100] G2/M abrogators,[101–104] Trk inhibitors,[105–108] angiogenesis inhibitors,[109,110] PARP inhibitors[111,112] and casein kinase 2 inhibitors[113]. This project was born out of the work of Prabhu *et al*,[6] who found that the indole-based compounds **4** and **5** (Figure 7) had reasonable activity against the glioblastoma cell line U87. These structures

themselves were designed based on a review authored by Lal and Snape, which described the variety of biological properties that 2-phenylindoles have been implicated in.[114] Due to their range of activities and their relatively simple structures, 2-phenylindoles make appealing lead structures for a medicinal chemistry research project.
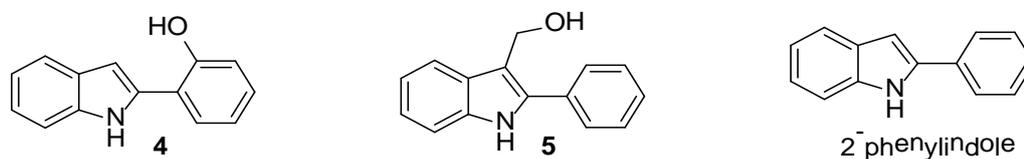
*Figure 7 - The structures of compound **4** (left), compound **5** (centre) and 2-phenylindole (right)*

Prabhu *et al.* originally considered compounds **4** and **5** to be of the same pharmacological class of compounds, and indeed these two compounds do have structural similarities (Figure 8). Both compounds contain a 2-phenylindole skeleton (shown in blue) and a hydroxide group that can be close in space (shown in red), combining into what would form a believable pharmacophore. However, to understand why this is now thought unlikely to be the case, the background of **5** needs to be examined.
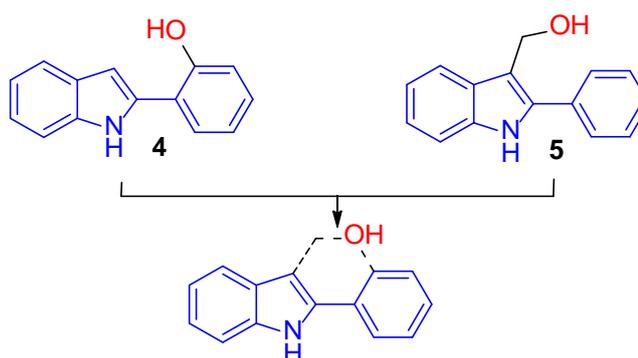
*Figure 8 - Commonalities in the structures of **4** (top left) and **5** (top right)*

1.3.1   The Lead Compound 5
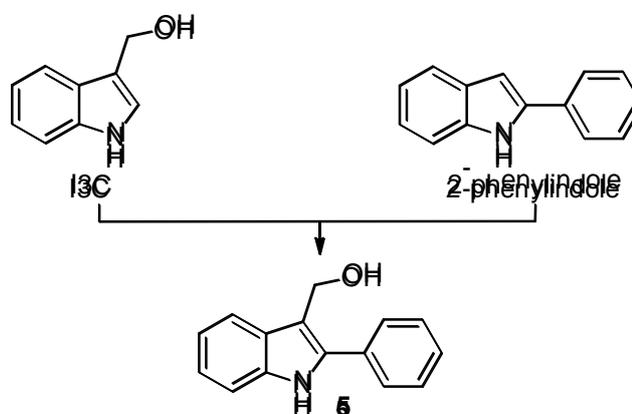
*Figure 9 - Compound **5** (below) was created as a hybrid structure of I3C (left) and 2-phenylindole (right)*

Compound **5** was created as a hybrid structure between the naturally occurring compound indole-3-carbinol (I3C) and the known privileged structure 2-phenylindole (Figure 9). I3C is found naturally in particularly high concentrations in cruciferous vegetables such as cabbage, kale, cauliflower and sprouts. I3C is by far the most studied component of cruciferous vegetables, a class of vegetable that appears to contain many anticancer agents.[115] *In vitro*, I3C forms many degradation products (Scheme 9),[116] which have been an active area of study for over 30 years,[117,118] during which time I3C has been shown to be active against a wide variety of cancers including cancer of the breast,[119] colon,[120–122] prostate[123–125] and, most recently, against glioblastomas of the brain.[6] Many of the I3C metabolites have also been shown to have wide ranging effects in biological assays.[123–126]
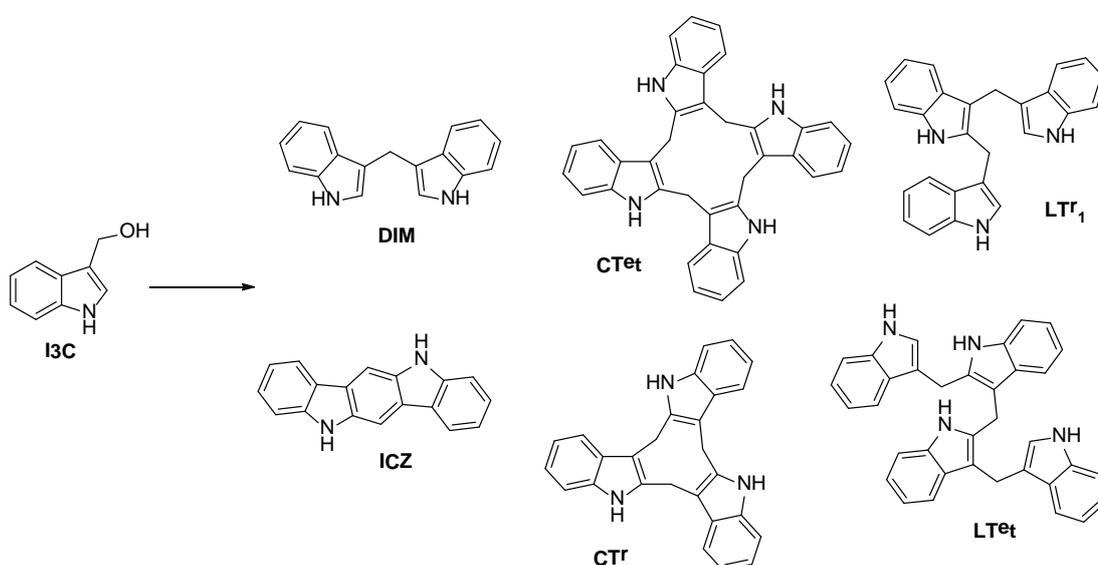
Scheme 9 - Indole-3-carbinol (I3C) and its metabolites 3,3'-diindolylmethane (DIM), a cyclic tetramer (CTet), the first linear trimer (LTr1), indolo[3,2-b]carbazole (ICZ), a cyclic trimer (CTr) and a linear tetramer (Ltet)

The simplest and mostly widely studied I3C metabolite is 3,3'-diindolylmethane (DIM). The mechanism of the conversion of I3C to DIM is well known, and proceeds under even slightly acidic conditions (even $CDCl_3$ in an NMR sample tube) *via* a 3-methyleneindolinium intermediate (**6**),[117] as shown in Scheme 10. DIM, which is known to be an active form of the prodrug I3C,[127] is formed in a reaction that appears to occur within 60 minutes under the conditions of the stomach.[128] Under normal physiological conditions, it is thought that I3C is converted almost exclusively into DIM,[118] whereas under the conditions of the stomach, DIM, CTr and LTr are the major products, and are thought to be produced in roughly equal quantities.[128]
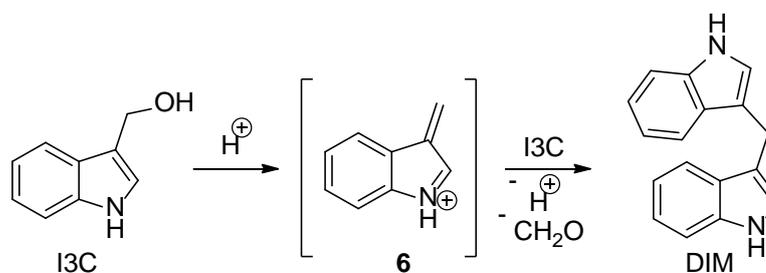
Scheme 10 - Generation of DIM from I3C via a 3-methyleneindolinium intermediate[116]

Although the ultimate fate of I3C is known, it is thought that there may be other chemical changes that occur while I3C is in the body. One such example of this, as described by Bjeldanes *et al*, is that I3C appears to form adducts with cellular thiols such as glutathione, cysteine and other such proteins. This has been shown to occur in cell culture media both in the presence and absence of cells, indicating that the conversion is not enzymatically controlled or facilitated.[129] The formation of such adducts, as was the case with the formation of DIM, is believed to proceed *via* the 3-methyleneindolinium ion (**6**). An example of such a formation is shown in Scheme 11 between I3C and the representative and common thiol cysteine to form the metabolite **7**.
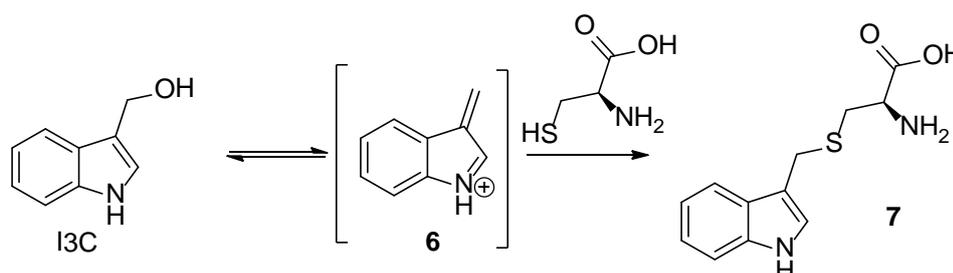
Scheme 11 - Proposed pathway for the formation of I3C-cysteine adduct

In the same paper by Bjeldanes *et al*,[129] the path that I3C can take through cultured MCF-7 breast cancer cells was investigated. The concentrations of I3C, the I3C-cysteine adduct **7** (shown in Scheme 11) and DIM were determined in the free media, in the cytosol of the cells and in the nuclei of the cells. It was found that the concentration of I3C in the media gradually decreased while the concentrations of its metabolites barely increased, indicating that either I3C or its metabolites were being taken up by the cells faster than I3C could be metabolised. I3C or DIM was not detected in the cytosol in significant quantities, but the concentration of adduct **7** was shown to steadily increase. Within the nucleus, I3C again wasn't detected in significant quantities, whereas the adduct **7** was detected initially, but was seemingly steadily converted to DIM. These observations suggested that I3C was converted to the adduct **7** in the media, adduct **7** was taken up by the cell, then passed into the nucleus where it was converted to DIM. The

fact that I3C doesn't seem to exist in significant concentrations inside the cells further suggests that I3C it not directly responsible for any biological effect, and instead it is a metabolite of I3C that is the active species.

Due to considering I3C as a prodrug, and based on the mechanism of formation of DIM (as shown in Scheme 10), compound **5** (Figure 7) might be expected to metabolise in much the same way, thus forming an analogous pharmacologically active metabolite. Scheme 12 suggests a formation of the 2-phenylindole analogue of DIM, compound **9**, which proceeds *via* the 3-methyleneindolinium intermediate **8**.
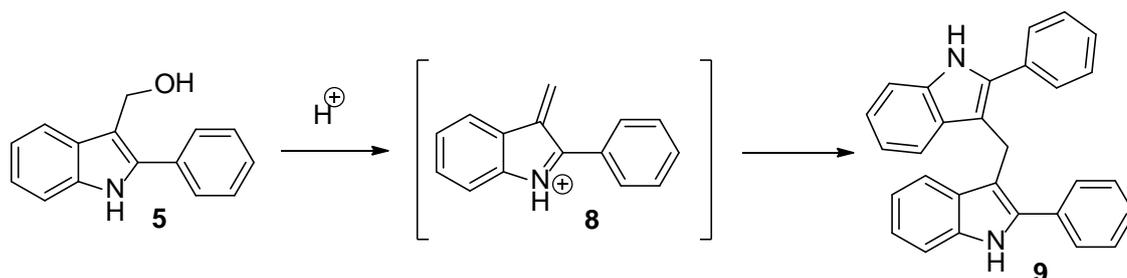
*Scheme 12 - Hypothesised analogous mechanism of the degradation of compound **5***

Returning to Figure 8, it is now obvious that compounds **3** and **5** cannot be part of the same pharmacological class in this instance, as comparing the structure of the prodrug **5** to the structure of compound **4** is not comparing like for like. What would be a more fair comparison would be to compare the structure of **4** with compound **9** or DIM, which obviously share very few structural features, and would therefore be unlikely to be considered as two members of the same chemical class.

Compound **5** is a novel compound and therefore a novel anticancer compound, and has been shown to be a marked improvement over its lead compound of I3C. Work within the group has found that against the U87 glioblastoma cell line and using the MTS assay, I3C has an IC_{50} of 526 µM, whereas compound **5** has an IC_{50} of 176 µM.[6]

Prior to this project, the compound **9** (Scheme 12) had yet to be tested as an anticancer compound. It has previously been investigated along with a series of DIM analogues as antibacterial compounds, but was found to be inferior to DIM in this respect, and was not taken forwards for further development.[130]

Overall, investigating the potential of both compound **5** and its metabolite compound **9** as glioblastoma therapies appears to be a novel area of research. This seems unusual considering

how much attention has been directed at I3C and at 2-phenylindoles, but this work will hopefully lead to the development of a more potent subclass of an already well studied class of anticancer compounds.

1.3.2   The Lead Compound 4

The second lead compound, compound **4** (Figure 10), was also found by Prabhu *et al*. to have anticancer activity against the U87 glioblastoma cell line, albeit to a lesser extent (IC_{50} of compound **5** was 176 µM, compared to an IC_{50} of 379 µM for compound **4**).[6] Compound **4** was tested due to it being a 2-phenylindole, a subclass of indoles which have many biological effects including reported anticancer properties.[114] Compound **4** appears to be a novel anticancer agent against glioblastoma cell lines, whilst its parent compound, 2-phenylindole, has been shown by Prabhu *et al*. to be inactive. This suggests that compound **4** may be a novel fragment with respect to antiglioblastoma compounds, and may be the simplest example of this class of compounds.
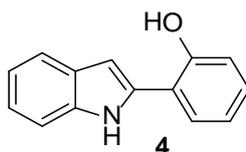
*Figure 10 - The structure of compound **4***

The mode of activity of compound **4** has also been briefly investigated. Using the Image-iT™ LIVE Green Reactive Oxygen Species Detection Kit, it was found that the addition of compound **4** to cells causes the levels of cellular reactive oxygen species (ROS) to increase (Figure 11). The Image-iT™ LIVE Green Reactive Oxygen Species Detection Kit determines quantities of cellular ROS fluorometrically; a ROS oxidises the molecular probe, and ROS levels can therefore be quantised using a plate reader. Because elevated ROS levels are typically associated with oxidative stress,[131] is was therefore hypothesised that compound **4** induces cellular death by a mechanism which includes elevated ROS levels. This hypothesis has literature precedents, as many phenol derivatives and phenol-containing compounds are known to act in this way.[132–135]
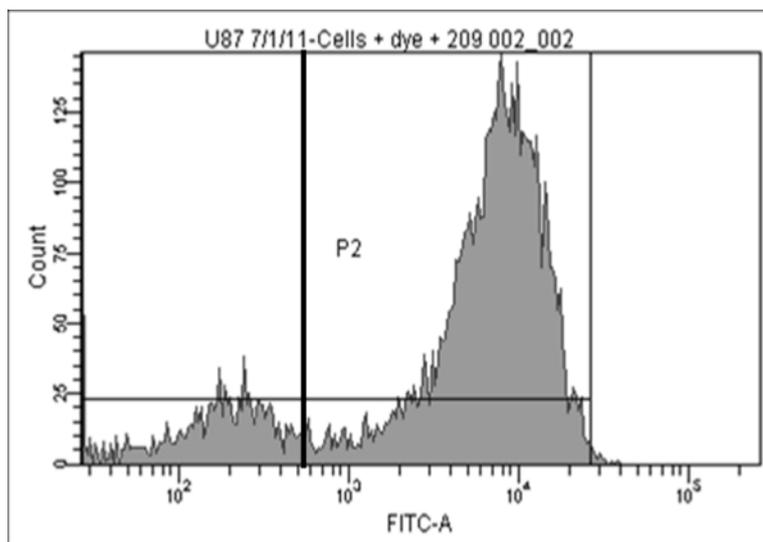
*Figure 11 - Detected ROS levels in U87 cells upon addition of compound **4**. Black vertical bar indicates baseline ROS level, and a peak shift to the right indicates elevated ROS levels*

Compound **4** has also been shown to induce apoptosis *in vitro*. Using the Annexin V-FITC apoptosis detection kit, 1321N1 and U87 cells were shown to have entered the early stages of apoptosis 24 hours after addition of compound **4**, and were shown to have entered the late stages of apoptosis 48 hours after addition of compound **4**.
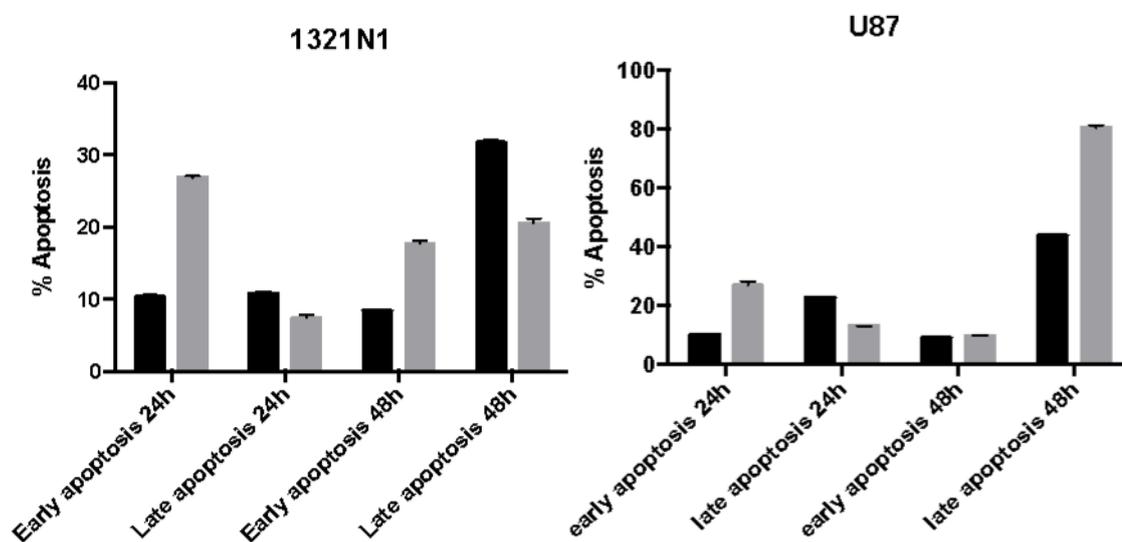
*Graph 1 - Percentage of cells in early and late apoptosis in 1321N1 and U87 cells 24 and 48 hours after addition of compound **4** (200 μM, darker bars) and the positive control cisplatin (100 μM on 1321N1 cell line, 200 μM on U87 cell line, lighter bars)*

1.4   Project Structure

In this project, a multifaceted approach to drug development is taken, whereby multiple different approaches to designing analogues of known active lead compounds are undertaken. First, the traditional approach of synthesising and testing analogues is carried out, whereby chemically relevant analogues are produced using nothing more than chemical intuition and literature precedents.   The second approach to determining potential functional analogues is to carry out a computational similarity search to discover a wider breadth of structural analogues. A similarity search is a way of exploring otherwise unconsidered or unknown areas of chemical space, hopefully leading to some functional analogues which will help elucidate otherwise neglected aspects of a structure-activity relationship (SAR).

Once the structural analogues have been determined and either procured or synthesised, a series of enthalpic values about their propensities to form radicals will be calculated. These values will be calculated computationally based on energies determined using density functional theory (DFT).

The anticancer activities of these structural analogues will also be established against glioblastoma cell lines and short term cultures. In the case of compound **4**, because current evidence suggests that the anticancer activity of this lead compound is related to ROS, the relationship between anticancer activities and the energies of the aforementioned ROS-forming steps for this series of analogues will be investigated.

CHAPTER 2 – PROJECT BACKGROUND

2.1 The Indole Privileged Structure

Although nature tends to produce biologically active compounds that make full use of all three dimensions, the state of chemical synthesis currently strongly favours the synthesis of rather planar compounds.[136] One outcome of this is the use of "scaffolds", i.e. the use of a molecular framework at the centre of a drug molecule around which biologically relevant substituents can be added in order to add activity or affect properties such as solubility or cell permeability. A 2009 paper based on the ZINC database found that over 241,000 unique scaffolds are currently known,[137] with the most common and versatile of these scaffolds being known as "privileged structures" or "privileged scaffolds".
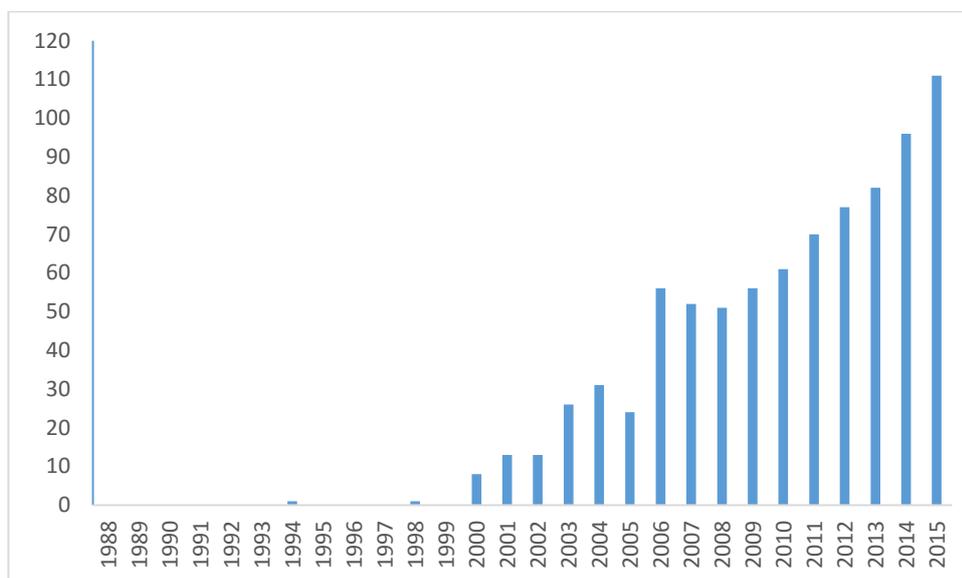
Graph 2 - Occurrence of the terms "privileged structure", "privileged structures", "privileged scaffold" or "privileged scaffolds" within the title or abstract of publications between 1988 and 2015, as found using Web of Science [†]

The terms "privileged structures" and "privileged scaffolds" have been in use since the term was first mentioned in 1988,[138] and have rapidly gone from a niche term to part of the standard medicinal chemistry lexicon (Graph 2). The specific definition has changed slightly over time, but perhaps the most widely accepted and succinct definition was given by Duarte *et al*, who described privileged structures as "molecular frameworks which are [capable] of providing useful ligands for more than one type of receptor or enzyme target by judicious structural modifications".[139] Privileged structures include 2-imidazoline,[140] quinoline,[141] indole,[142]

[†] Original graphic created solely for this thesis

isatin[143] and chromone[144] (Figure 12) all five of which are highly rigid, planar compounds (although these are not features which are necessarily inherent to being a privileged structure). This in part explains why much of the explored drug-like chemical space involves largely planar compounds.
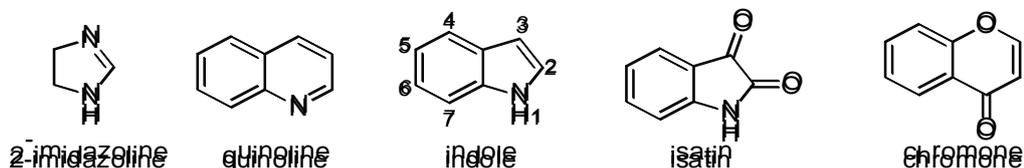
Figure 12 - The privileged structures 2-imidazoline, quinolone, indole, isatin and chromone. The numbering system for indole is shown

The indole privileged structure is of particular importance, and is believed to be the most ubiquitous privileged structure known.[114,145] The indole structure contains seven positions where derivatisation can occur (described in Figure 12), is perfectly planar, and contains an NH group at the 1-position which is not capable of acting as a hydrogen bond acceptor, and is known to act as a poor hydrogen bond donor. The classification of being a privileged structure is certainly earned in the case of indole, which makes up the structural core of compounds with incredible diversity in activity, including the psychedelic drug LSD, the antimigraine drug rizatriptan, the "happy" hormone serotonin and the amino acid tryptophan (Figure 13).
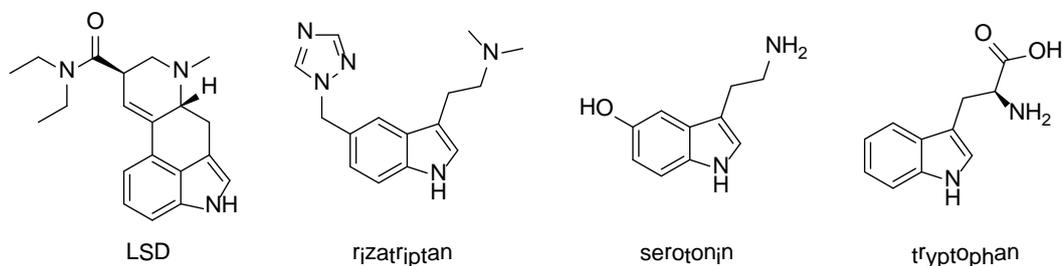
Figure 13 - From left to right, the indole containing compounds LSD, rizatriptan, serotonin and tryptophan

2.2   Indole Syntheses

Due in part to their wide-ranging uses, the development of indole syntheses has been ongoing for many years, and much progress has been made in our ability to synthesise increasingly complex indoles under more mild and generic conditions.

2.2.1 The Fischer Indole Synthesis
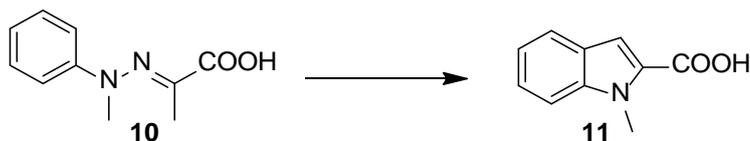
Scheme 13 - The original Fischer indole synthesis

The most widely researched method of synthesising indoles is the Fischer indole synthesis, which was first carried out by Herman Emil Fischer in 1883.[146] Fischer discovered this synthesis while synthesising 1-methylindole-2-carboxylic acid (**11**) from the phenylhydrazone **10** using this procedure, shown in Scheme 13 (although the exact product of this reaction wasn't identified until 1884)[147]. The Fischer indole synthesis involves cyclising a phenylhydrazone (**13**) to an indole (**17**) under acidic conditions, and has since been developed to allow for the synthesis of indoles with substituents at any of the seven substitutable positions.
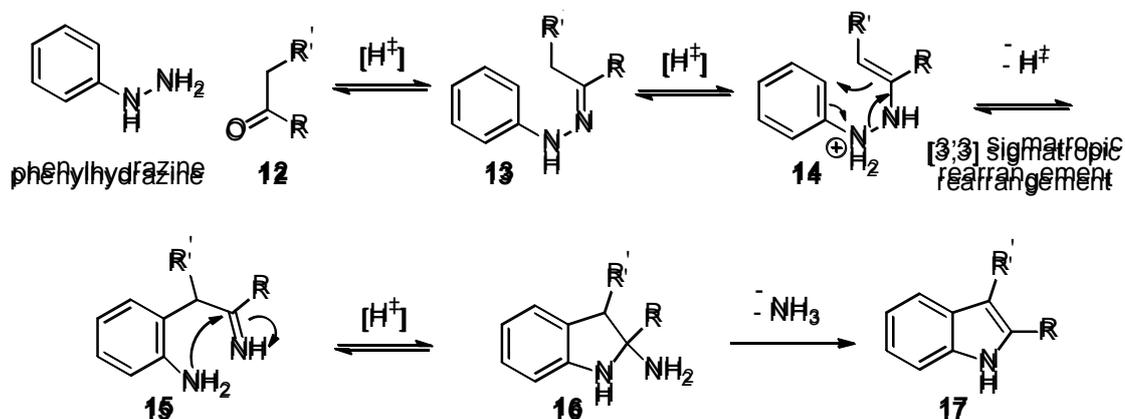
Scheme 14 - The mechanism of the Fischer indole synthesis

A major benefit of the Fischer indole synthesis is that, under the same reaction conditions, the hydrazone can be generated *in situ* from a hydrazine and a ketone, both of which are classes of chemicals with wide commercial availability. The actual mechanism of this tandem reaction, shown in Scheme 14, has since been elucidated by Allen and Wilson in 1943,[148] and was based on a suggestion first postulated by Robinson and Robinson in 1918.[149] Starting from the aforementioned phenylhydrazine, the mechanism involves a condensation reaction between the hydrazine and an aldehyde to form the hydrazone, followed by a [3, 3] sigmatropic rearrangement, cyclisation, and subsequent loss of ammonia to yield the product indole.
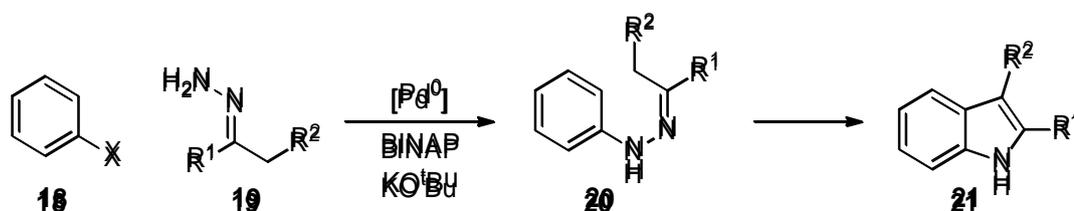
Scheme 15 - The Buchwald modification for generation of N-*arylhydrazones, followed by cyclisation to the indole under Fischer conditions*

If the hydrazone can't be synthesised under these classic conditions, an alternative and increasingly popular entry into the Fischer indole synthesis is the palladium-catalysed cross coupling of an aryl bromide (**18**) and a hydrazone (**19**), forming the *N*-arylhydrazone (**20**). This so called "Buchwald modification" of the Fischer indole synthesis, shown in Scheme 15, was first published by Stephen Buchwald *et al* in 1998,[150] and helped confirm the proposed role of the hydrazone intermediate in Scheme 14.
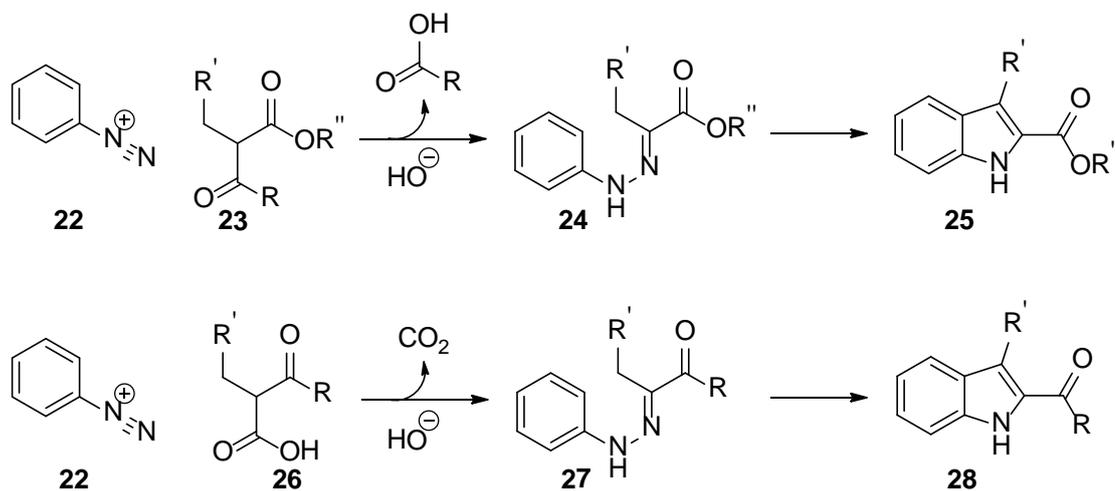
Scheme 16 - The Japp-Klingemann synthesis of indoles from aryldiazonium salts

Aside from substitutions at the *N*-position, the main limitation of the Buchwald modification is the stability of aryl hydrazines, making it unreliable for large-scale industrial use. However, aryl hydrazines can be avoided altogether by employing the Japp-Klingemann synthesis of aryl hydrazones directly from an aryldiazonium salt (**22**) and a β-keto ester (**23**) or β-keto acid (**26**), as shown in Scheme 16. The Japp-Klingemann synthesis was first reported in 1887 by Francis Japp and Felix Klingemann,[151] when a small range of hydrazones (compounds **24** and **27**) were synthesised. These hydrazones could then be cyclised to their respective indoles (compounds **25** and **28**) using classic Fischer conditions. The commercial potential of this reaction was proven in 1999 by Reddy *et al*, when the Japp-Klingemann synthesis was included in the multistep

synthesis of the potential anticancer agent melatonin, which was synthesised on a 5-10kg scale (Scheme 17).[152]
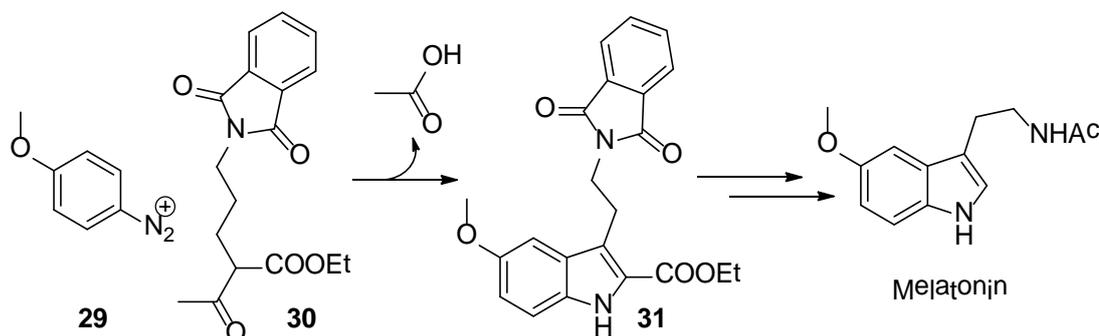
Scheme 17 - Reddy's multi-kilogram scale synthesis of melatonin, employing the Japp-Klingemann hydrazone synthesis and the Fischer indole synthesis

The Fischer synthesis is by far the most developed reaction for synthesising indoles, and as such has the broadest scope. This reaction is particularly compatible with a wide range of substitutions on all possible carbons due to the wide range of conditions that have been developed,[153–157] and the new C-C and C-N bonds are produced with a reasonably high degree of atom economy. Substitutions on the *N*-position are however much more difficult and less developed, and the procedure is obviously unsuitable for reagents or products that are acid sensitive.

2.2.2   Synthesis of Indoles From 2-Alkynyl Anilines

2-Alkynyl anilines (**34**) can be intramolecularly cyclised by a number of methods to form indoles, and are themselves commonly synthesised *via* Sonogashira coupling between 2-haloanilines (**32**) and a terminal alkyne (**33**), as shown in Scheme 18. Methods for the cyclisation step are mostly differentiated based on the choice of catalyst.
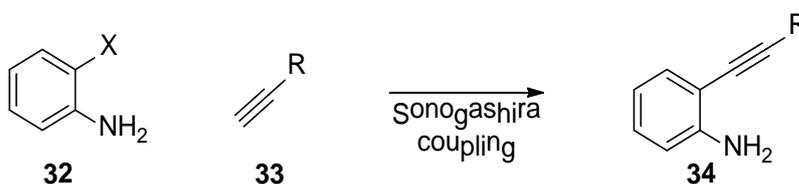
*Scheme 18 - Example synthesis of a 2-alkynyl aniline (**34**) via the Sonogashira reaction*

Alkoxides such as potassium *tert*-butoxide or sodium ethoxide can be used to cyclise 2-alkynyl anilines or their carbamate derivatives (**35**), although the carbamate group is cleaved off under reaction conditions. This reaction is typically carried out in a protic solvent (often the corresponding alcohol of the alkoxide).
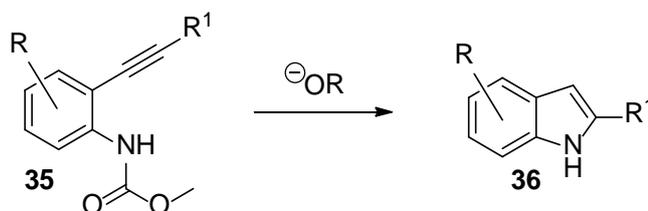
Scheme 19 - Alkoxide-mediated cyclisation of the carbamate derivative of a generic 2-alkynyl aniline (35)

to form an indole (36)

An alkoxide-mediated cyclisation of a 2-alkynyl aniline has been successfully used in the synthesis of an analogue of rebeccamycin (**39**), a potent antitumour agent, as shown in Scheme 20.
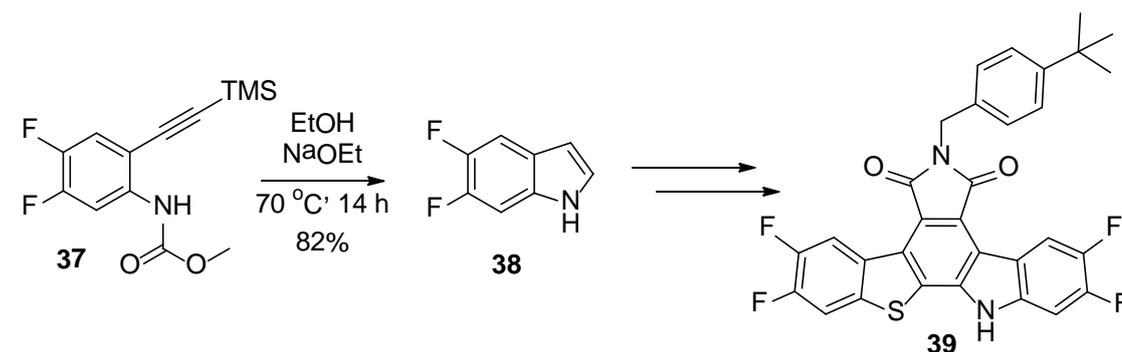
Scheme 20 - The use of alkoxide-mediated cyclisation conditions in the synthesis of a rebeccamycin

analogue

Inorganic and organometallic catalysts can also be employed to cyclise 2-alkynyl anilines. Copper-mediated cyclisations, such as the Castro synthesis, are widely reported. The Castro synthesis, first reported by Castro *et al* in 1963[158] and shown in Scheme 21, utilises an excess of copper (I) salts as a catalyst, and yields indoles with very high atom economy.
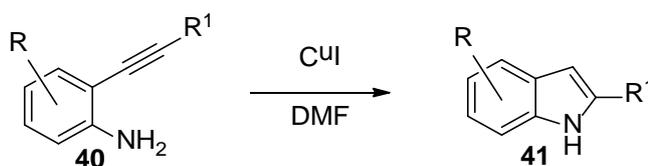
Scheme 21 - The standard Castro indole synthesis

The standard Castro reaction begins with a 2-alkynyl aniline (**40**), however, a one-pot reaction beginning with carboxylated *o*-iodoanilines was developed by Cacchi *et al* to generate 2-aryl indoles (Scheme 22), potentially saving a reaction step in generating the 2-alkyl aniline.[159] These reactions were found to proceed with moderate to excellent yields for the synthesis of 2-aryl indoles, and in poor yields for 2-alkyl indoles.
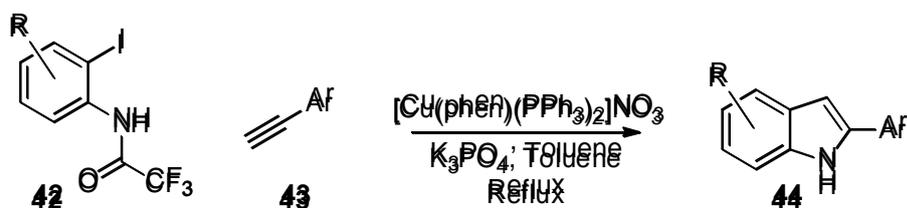
Scheme 22 - Cacchi's one-pot synthesis of 2-aryl indoles

Although a mechanism for the Castro synthesis is yet to be conclusively proven, a reasonable proposal is shown in Scheme 23.
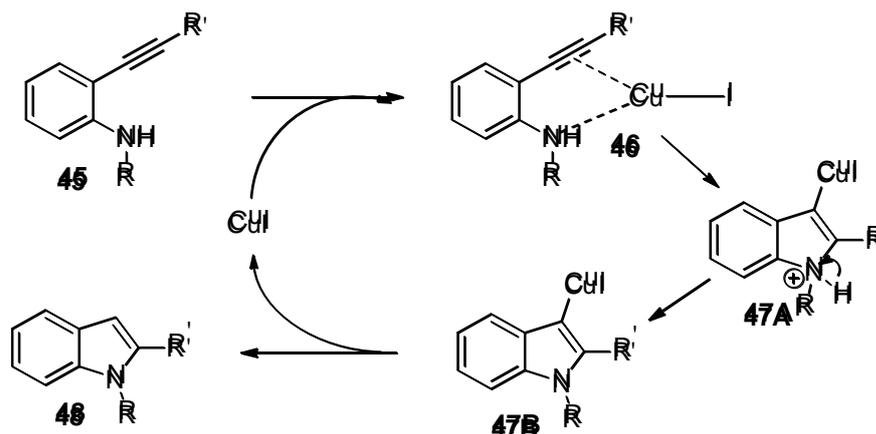
Scheme 23 - Proposed mechanism for the Castro synthesis

The Castro synthesis has found use in complex multi-step syntheses, such as in the total synthesis of α-C-mannosyltrytophan,[160] a naturally occurring C-glycosylamino acid. The step employing the Castro reaction, shown in Scheme 24, produces the indole **50** from the 2-alkynyl aniline **49**, and was just one of 10 steps in this difficult synthetic project.
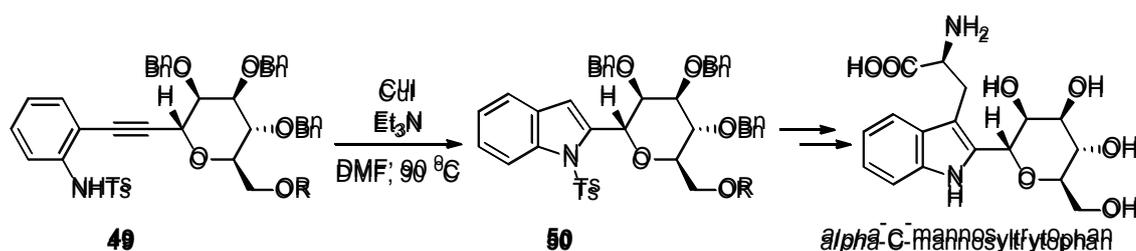
Scheme 24 - The Castro step in the total synthesis of α-C-mannosyltrytophan

As well as copper-mediated cyclisations, palladium-mediated cyclisations of 2-alkynyl anilines have also been reported. These reactions typically begin with 2-alkyl anilines, although some one-pot indole syntheses carried out under Sonogashira conditions and beginning at the *o*-iodoaniline have been developed. Although typically less common due to the higher cost of palladium and the higher temperatures that these reactions typically require, considerable work has gone into the development of this class of indole syntheses.

58

The conditions developed by Larock *et al* are a good example of the potential of a one-pot, two step, palladium catalysed indole synthesis, although these conditions are not to be confused with the Larock indole synthesis, which is described later. The first step is a coupling between an *o*-iodoaniline (**51**) and terminal alkyne (**52**) under solvent-free Sonogashira conditions and under gentle microwave heating. The second step adds in an organoiodide (**54**) and some solvent, producing the target indole (**54**) in good to excellent yields.[161]
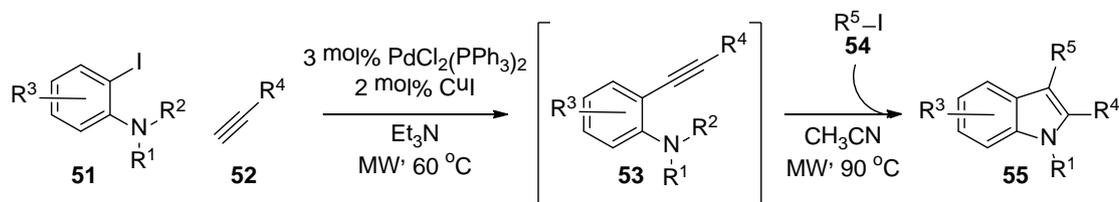
Scheme 25 - Larock's One-pot, two-step palladium-catalysed indole synthesis

A catalytic cycle for this reaction was proposed (Scheme 26), which includes the palladium(0) catalyst oxidatively inserting into the oranoiodide bond of compound **54** and coordinating to the alkyne to form complex **56**, followed by the aniline lone pair attacking the weakened alkyne bond and subsequent formation of a carbon-palladium bond to form compound **57**, and finally the elimination and regeneration of the palladium(0) catalyst.
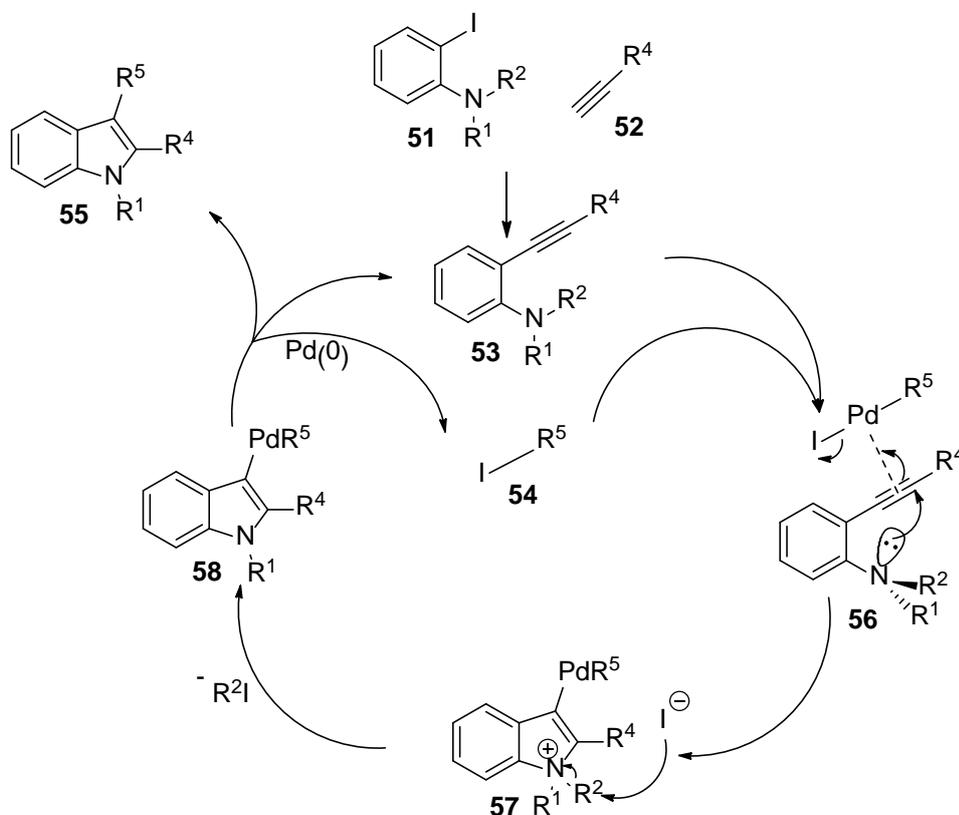
Scheme 26 - Proposed catalytic cycle for Larock's cyclisation of a 2-alkyl aniline

As well as many other palladium-catalysed heteroannulations of 2-alkyne anilines, there are also a number of less developed cyclisations mediated by other metals and metal complexes such as indium,[162] mercury,[163] silver[164] and iodine[165,166] compounds.

2.2.3   Larock Indole Synthesis

A major drawback of using 2-alkynyl anilines is that substituents at the 3- position are not possible. A similar method to those mentioned above, and developed by the same group that developed the one-pot microwave synthesis of indoles from *o*-iodoanilines and terminal alkynes, is the so-called Larock synthesis. This reaction was first described by Richard Larock in 1991 (Scheme 27),[167] who made a range of indoles with substitutions at the 1-, 2- and 3-positions (**61**) out of internal alkynes (**60**).
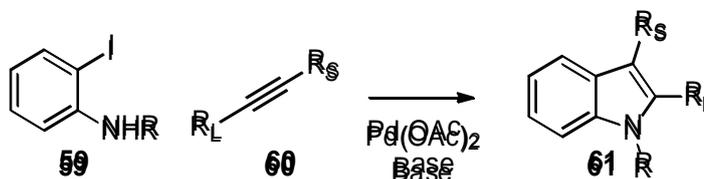
Scheme 27 - The Larock indole synthesis. R_L is more sterically bulky than R_S

The mechanism is thought to proceed via insertion of palladium into the C-I bond to form complex **62**, followed by coordination of the alkyne to the palladium to form complex **63**, and subsequent formation of the indole **61**, as shown in Scheme 28.[168] The regioselectivity of this reaction can also be described by this catalytic cycle; in the reductive elimination step which forms complex **64**, the R_L group is directed adjacent to the palladium due to the relatively long Pd-C bond lowering the steric clash felt by the larger R group.
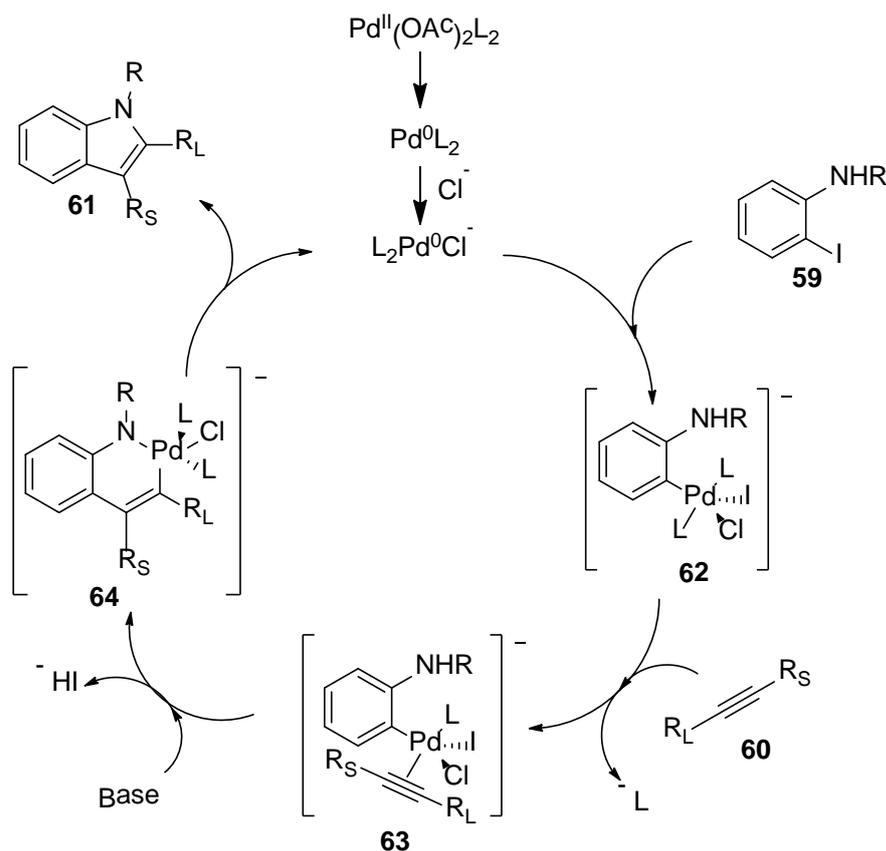
Scheme 28 - Catalytic cycle of the Larock indole synthesis. R_L is more sterically bulky than R_S

As can be seen in the catalytic cycle, a major drawback of the Larock indole synthesis is that an iodine in the *ortho*-position of the aniline is essential. This can be a tight limitation, especially during convergent syntheses of large molecules. A further reason for the Larock indole synthesis having poor uptake industrially is that an excess of at least two equivalents of the alkyne is often required, severely reducing the efficiency of the reaction.

2.2.4   Reductive Cyclisations

Reductive cyclisations of *ortho*-substituted nitrophenyls have received much interest from those involved in synthesising indoles on a large scale due to how reliably these reactions can be scaled up. *o*-Nitrotoluenes, common precursors to *ortho*-substituted nitrophenyls, also have the additional benefit of often being widely commercially available.

The Leimgruber-Batcho synthesis of indoles, peculiarly, was widely used industrially before being published in any scientific journals. Amongst the first publications on the topic was a review by Leimgruber and Batcho in 1985,[169] whereas the original patent for the reaction was filed in 1970.[170]   This synthesis converts an *o*-nitrotoluene (**65**) to an enamine (**67**) using N,N,-dimethylformamide dimethyl acetal (DMFDMA) and an optional amine (**66**), often

pyrrolidine. This enamine is then reductively cyclised to the 1,2-unsubstituted indole (**68**), as shown in Scheme 29.
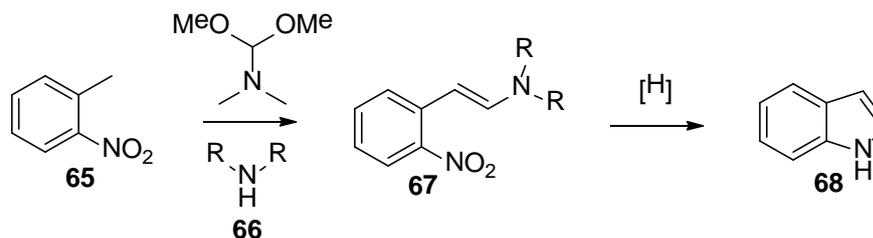
Scheme 29 - The Leimgruber-Batcho indole synthesis

The mechanism for this reaction occurs in two steps. The first step, shown in Scheme 30, shows the role of the optional amine (pyrrolidine in this example) in the formation of a more reactive reagent (**69**).[171]
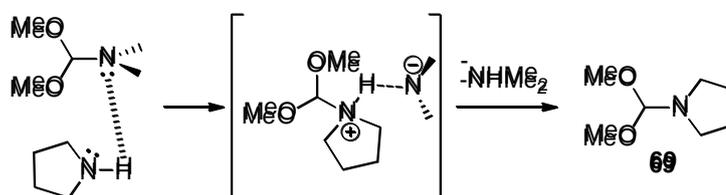
Scheme 30 – The formation of a more reactive reagent, with pyrrolidine used as an example amine

After the *in situ* formation of this reagent, the deprotonated nitrotoluene (**71**) goes on to form an enamine (**67**), the nitrophenyl group of which can be reduced under a number of conditions to form an aniline (**73**), which subsequently cyclises to form the respective indole (**68**), as shown in Scheme 31. The reaction is typically run at high temperatures for extended periods of time.
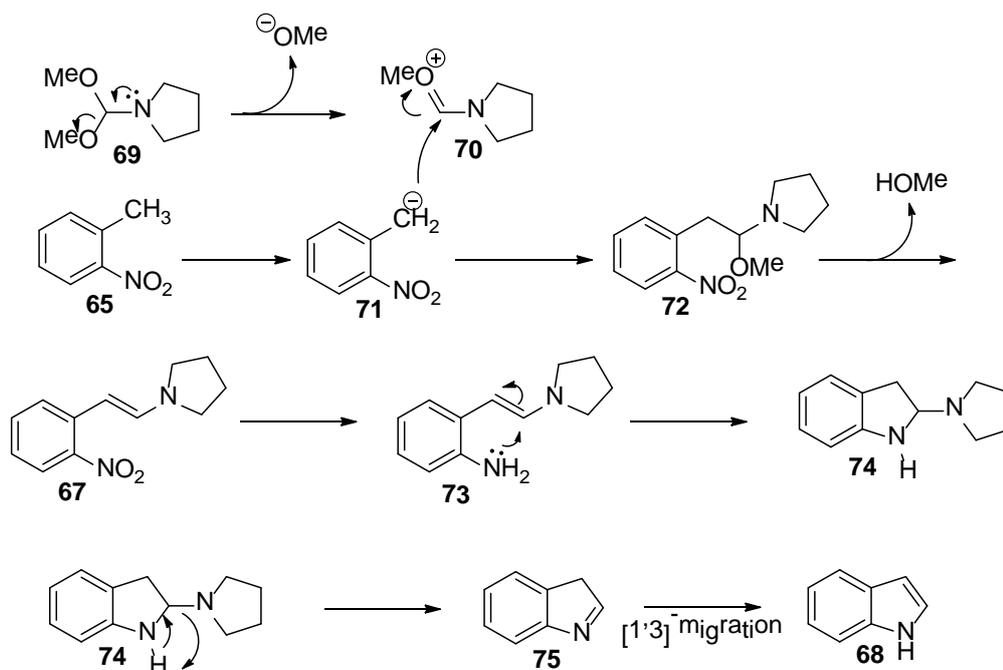
Scheme 31 - The mechanism of the Leimgruber-Batcho synthesis

This synthesis has retained popularity industrially due to the availability of substituted *o*-nitrotoluenes, which the reaction is generally tolerant towards.[169] Substitutions at the 2- and 3-positions are less developed but possible. Substitutions on the nitrogen are not possible within the Leimgruber-Batcho method without employing a separate step.

This process has been employed in the synthesis of a number of biologically relevant indoles, including the synthesis of a series of indolopyrrolocarbazole alkaloids by Ohkubo *et al* (Scheme 32).[172]
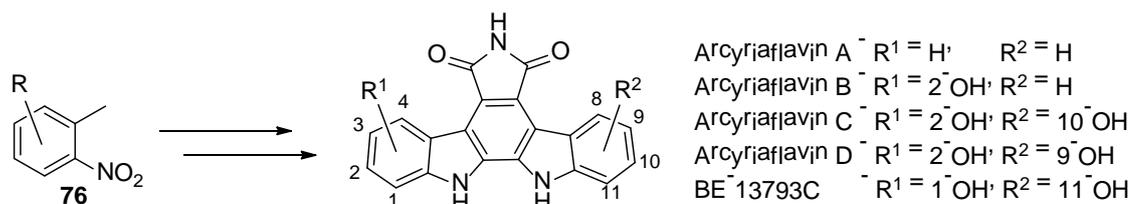
Scheme 32 - Ohkubo's synthesis of a series of a series of arcyriaflavins and an analogue

Overall, the Leimgruber-Batcho indole synthesis is fairly well developed, but development and academic interest appears to be waning. There are however a number of other reductive cyclisation options, such as the reductive cyclisation of *o*-nitrobenzylcarbonyls. *o*-Nitrobenzylcarbonyls, which can be synthesised from *o*-nitrotoluenes, can be reductively cyclised in a process known as the Reissert synthesis. This synthesis can be traced back to 1897 when Reissert cyclised ethyl *o*-nitrophenylpyruvate (**79**) to form indole-2-carboxylic acid (**80**), as

shown in Scheme 33.[173] The Reissert synthesis involves the condensation of an *o*-nitrotoluene (**77**) with an oxalic ester (for instance diethyl oxalate, **78**, as shown in Scheme 33) to form an *o*-nitrobenzylcarbonyl (ethyl *o*-nitrophenylpyruvate, **79**, in Scheme 33). Like the Leimgruber-Batcho synthesis, this synthesis is tolerant of substitutions on the *o*-nitrotoluene starting material, but is less capable of producing indoles with substituents on the 3- position of the indole. The product indole will always contain a carboxylic acid on the 2-position, which can be easily decarboxylated upon heating.
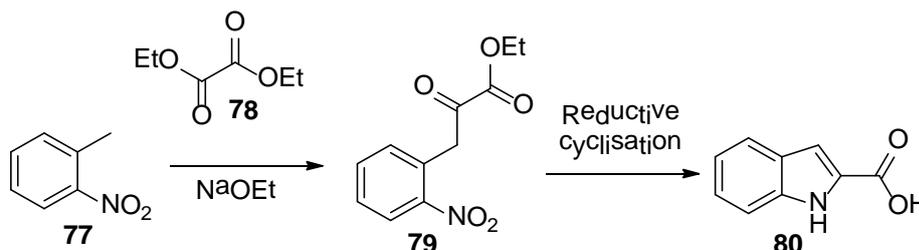
Scheme 33 - The original Reissert synthesis of an indole

A much more versatile method of synthesising *o*-nitrobenzylcarbonyls is by employing the Truce-Smiles rearrangement. The Truce-Smiles rearrangement (Scheme 34),[174] a variant on the Smiles rearrangement,[175] is a particularly powerful C-C bond forming reaction, and lends itself well to the synthesis of *o*-nitrobenzylcarbonyls. The Truce-Smiles rearrangement has recently been shown to be useful as the first step in a tandem synthesis of indoles.[176,177]
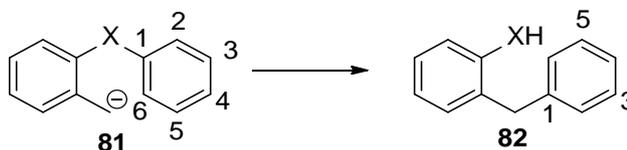
Scheme 34 - The Truce-Smiles rearrangement, where X is an electronegative group or atom

The Truce-Smiles rearrangement requires the incoming nucleophile to be a carbanion, as seen in Scheme 34. This can be achieved through the use of deprotonated conjugated functional groups such as acetyl groups or through the use of organolithium compounds. This rearrangement has been utilised by the Snape group to form a range of *o*-nitrobenzylcarbonyls. The antitumour agent 2-(2'-hydroxyphenyl)indolen (**4**) was synthesised from one such *o*-nitrobenzylcarbonyl compound (**86**) formed using the Truce-Smiles method (Scheme 35).[176]
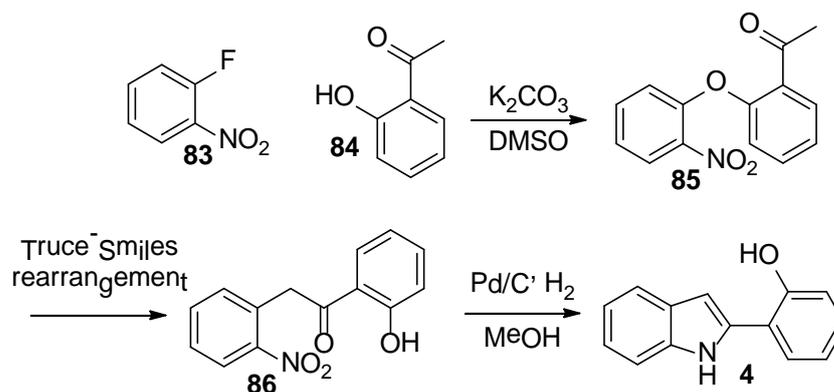
Scheme 35 - A Truce-Smiles rearrangement and subsequent cyclisation to form the anticancer compound 2-(2'-hydroxyphenyl)indole (4)

The Truce-Smiles rearrangement may soon also find use in industrial processes, as it was recently utilised in the multi-step synthesis of the antiviral compound **89**, which was successfully synthesised on a 50 kg scale (Scheme 36).[178]
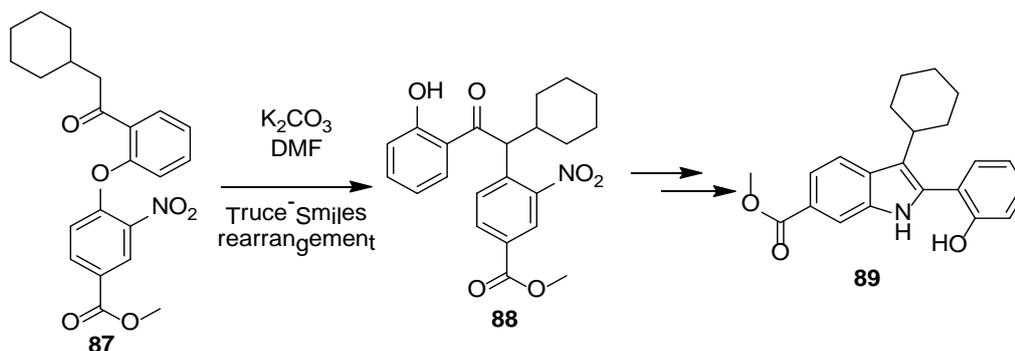
Scheme 36 - A Truce-Smiles rearrangement on a 50 kg scale

2.2.5 The Bartoli Indole Synthesis

The Bartoli indole synthesis (often simply called the Bartoli reaction) was first described by the Bartoli group in 1989,[179] and provides a simple way of synthesising 2- and 3-substituted indoles from commercially available *o*-nitrophenyl compounds (**90**) and a vinyl Grignard reagent (**91**), as shown in Scheme 37. It does, however, require at least three equivalents of the Grignard reagent, so has poor atom economy, as well as being inefficient to carry out on scale.
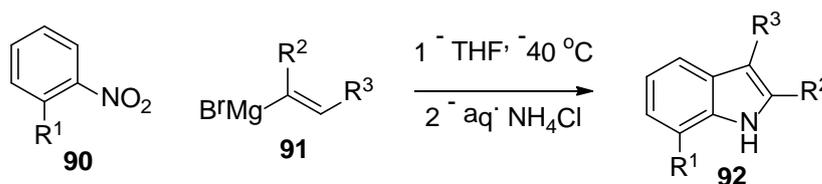
Scheme 37 - The Bartoli indole synthesis

The mechanism of the Bartoli indole synthesis (Scheme 38) begins with one equivalent of the Grignard reagent **91** to convert the nitro compound **90** to the nitroso compound **94**, a further equivalent of the Grignard reagent **91** to add an alkene to the nitroso group to form compound **95**, a Claisen rearrangement (**96**), a subsequent cyclisation (**97**) and a rearrangement to form the indoline **98**, which reacts with the final equivalent of the Grignard reagent **91**. This *N*-substituted indoline **99**, on an acid work-up, forms the product indole **92**.
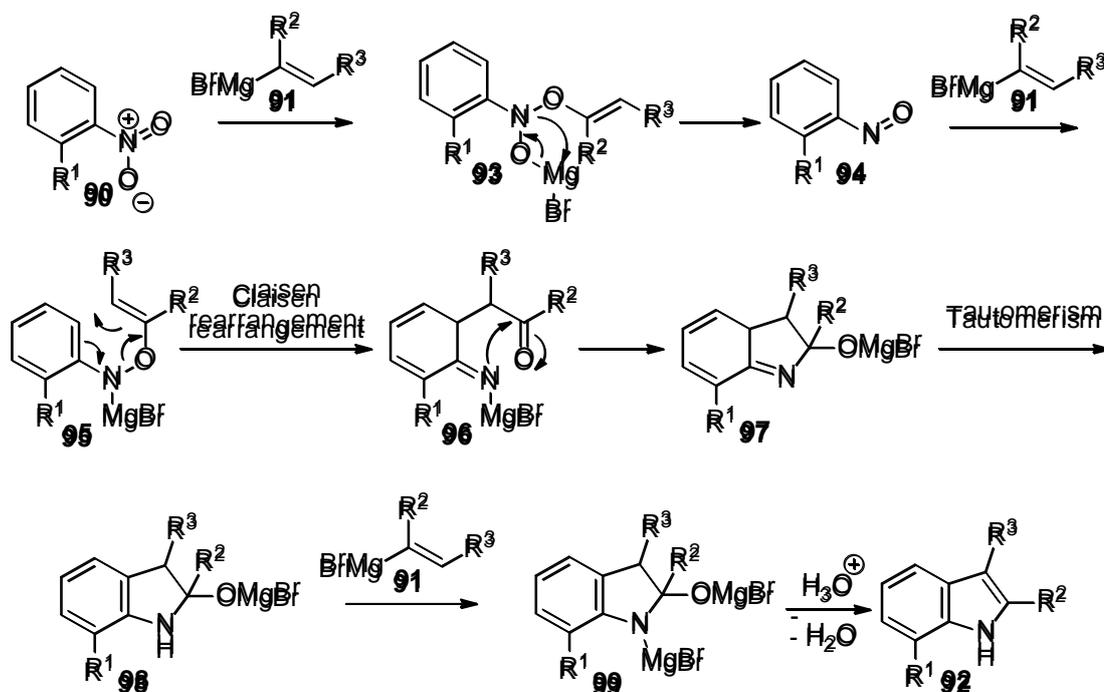
Scheme 38 - Mechanism of the Bartoli indole synthesis

2.2.6   Summary of Indole Syntheses

A large variety of syntheses have been described, and a summary of these syntheses is presented in Table 1.

SYNTHESIS	STRENGTHS	WEAKNESSES
FISCHER INDOLE SYNTHESIS	Most highly developed method Substituents at all positions are possible	Unsuitable for acid-sensitive reagents or products
FROM 2-ALKYNYL ANILINES	Reagents can be easily synthesised Wide variety of conditions have been developed	Substituents at the 3-position are difficult Substituents at the 1-position are often difficult
LAROCK INDOLE SYNTHESIS	Substituents at the 2- and 3-positions are well tolerated	Two equivalents of the alkyne are often required Stereocontrol of this reaction limits the use of this reaction
REDUCTIVE CYCLISATIONS	These reactions scale well A wide variety of conditions have been developed	Generally poor atom economy Substituents at the 2- positions are less developed Substituents at the 3- position are very difficult
BARTOLI INDOLE SYNTHESIS	Reagents are easily synthesised Substituents at the 2- and 3-positions are particularly well tolerated	A substituent at the 7-position is essential Three equivalents of Grignard reagent are required

Table 1 - Summary of indole syntheses

These reactions allow for a large area of indolic chemical space to be accessed. Although being able to access as wide an area of chemical space as possible is important, knowing which areas of chemical space are available and accessible or knowing which areas of chemical space are of synthetic interest is just as important.

2.3 Developing Analogues

The role of the medicinal chemist at the earliest stages of drug development is to establish a correlation between structural changes to a known lead compound and the effect that these changes have on the compound's activity, developing what is known as a structure-activity relationship (SAR). Although this general concept has remained consistent throughout the history of drug development, the specific approach to developing an SAR has changed considerably. This is partially down to improvements in identifying and understanding how compounds interact with target sites, and partially due to the development of computational techniques.

The traditional approach of finding active analogues, known as "analogue-based drug design", can be broadly split into two classes. The first approach is known as "structure-based drug design" (SBDD),[180,181] which requires the structure of the active site to be known (often determined *via* X-ray diffraction or NMR). Drugs that have the potential to interact strongly with this active site can then be synthesised and tested. This approach is heavily favoured if the structure of the active site is known, as it allows for a higher number of active compounds to be tested. With the SBDD approach, the potential strength of a drug-receptor interaction is now almost exclusively determined computationally. This is due to the recent improvements in docking software and the reduction in the cost of computational power.

The second approach is useful in the cases where the active site is unknown or is difficult to structurally determine. In these cases, an approach called "ligand-based drug design" (LBDD) can be employed. LBDD begins with making structural analogues of a lead compound, each of which tends to add in or remove a single structural feature that is to be investigated. This approach allows the medicinal chemist to highlight which structural features of the drug are required for activity, while also exploring the immediate chemical space of the lead compound. Based on how the SAR develops, this process can be continued iteratively, with finer detail about the structural requirements of the drug being determined.

LBDD can lead to the determination of a pharmacophore, which is a description of the common structural properties of a class of drugs that hit a specific target with good efficacy. Qualities of

the pharmacophore may include hydrogen bond acceptors in certain positions, structural rigidity in specific areas, planar areas, 3D areas etc. It is conceptually similar to the SBDD method whereby appropriate fragments are assembled into compounds of interest, but LBDD takes much more trial and error to build up the relationship between structure and activity.

On balance, when the target site of a compound is known, the SBDD approach is almost universally preferred. As well as having a lower chance of producing inactive analogues due to the increased knowledge about the target site, the SBDD approach allows investigation of a wider breadth of chemical space due to not being limited to direct analogues of known active compounds.

2.3.1   Chemical Space

The purpose of both the SBDD and LBDD approaches is ultimately to test analogues of a known active compound with the intention of determining which chemical changes result in a change in activity, an approach often referred to as exploring chemical space.

Chemical space is the term used to describe the hyperdimensional virtual space within which all possible chemical species exists, with each possible molecular property as an axis of this virtual space. A more widely accepted and rigorous definition, published in 2015 by von Lilienfeld was given as "the combinatorial set of all compounds that can be isolated and constructed from possible combinations and configurations of N_l atoms and N_e electrons in real space."[182] The aim of drug discovery is ultimately to explore as much chemical space as possible with the aim of finding pharmacologically active areas, and as a result building up a structure-activity relationship. Due to the exploration of chemical space that is inherent within the field of drug discovery, the majority of the literature on the topic of chemical space is relevant to the medicinal chemist.

Background on Chemical Space

Due to its inherent hyperdimensional nature, visualisation of chemical space is an inherently difficult endeavour. Each property of a molecule (e.g. mass, LogP, colour, number of rotatable bonds etc.) can be visualised as a separate axis of chemical space, meaning that a complete representation of chemical space will have countless dimensions. A visual representation of these dimensions is obviously impossible, but there are a number of techniques that chemical cartographers can employ to circumvent this issue. One such technique is to plot only the most important parameters for the particular situation, thus drastically reducing the number of necessary axes, in a process known as Principle Component Analysis (PCA).[183] An example of

this is in the work of Ruddigkeit *et al.*, who were trying to determine, amongst other things, how the 3D shapes of known drug-like molecules differed from the 3D shapes that would be expected to be found in general drug-like chemical space, and found that synthetic compounds are typically much flatter and rod-like than the more spherical compounds one would expect to find (Figure 14).[136]
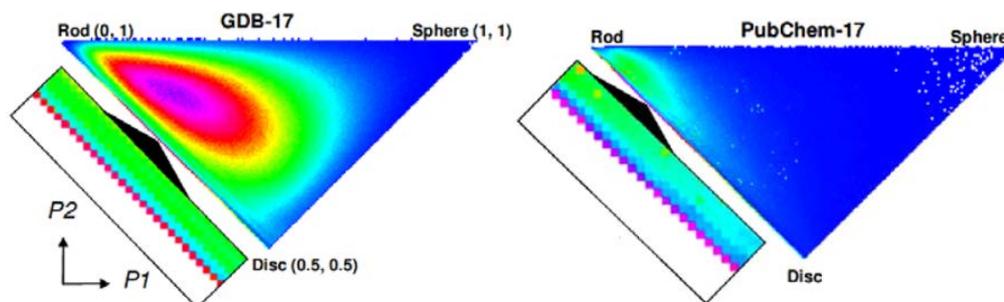
Figure 14 - Comparison of the 3D shapes of compounds found in general drug-like chemical space (left) and known drug-like compounds (right). Taken from the work of Ruddigkeit et al.*[136]*

Paraphrasing Douglas Adams, the famed chemistry blogger Derek Lowe once said "Chemical space is big. Really big … You just won't believe how vastly, hugely, mindboggling big it is."[184] An estimate of the total size of chemical space isn't particularly useful from a drug discovery perspective; clusters of f-block elements are unlikely to be useful as drugs. Estimating the size of chemical space of compounds currently known to be drug-like is a more meaningful endeavour, but values vary wildly.[185] Bohacek *et al.* estimates this value to be 10^{60} compounds (≤30 heavy atoms, with heavy atoms being C, N, O or S),[186] whereas Ertl *et al*. suggested a value of 10^{23} compounds (≤36 heavy atoms, with heavy atoms being C, N, O, S, P, Si, Se or a halogen),[187] but the most up to date estimate put the value at around 10^{33} drug-like compounds (≤ 36 heavy atoms, with heavy atoms being C, N, O, S or a halogen, a molecular mass of ≤500 Da and includes stable compounds only).[185] Even the smallest of these values is beyond the realms of human comprehension. The largest chemical database available, the CAS database, contains information not just on known drug-like compounds but on any known organic compound, as well as known alloys, minerals, polymers, salts and coordination complexes. Despite this lack of limitation, it still only contains 113 million entries, 12 orders of magnitude fewer than the lowest estimate of drug-like chemical space.

The entirety of chemical space couldn't possibly be fully investigated on any realistic timeframe, let alone within the timeframe required for a drug discovery project. Therefore, efficient ways

of investigating pharmacologically relevant areas of chemical space are of utmost importance to the modern medicinal chemist.

Searching Chemical Space – Analogue-Based Drug Design

As mentioned earlier, the traditional way of exploring chemical space is to simply make analogues of a lead compound. This process is low-throughput, arduous and expensive, but due to its ease, it is still commonly carried out. Representative examples of the two approaches to analogue-based drug design, SBDD and LBDD, are shown below.

One approach to SBDD, known as molecular docking, involves taking a database of compounds and sequentially virtually fitting them into the known active site, and reporting their potential as active compounds based on the strength of the drug-receptor interaction.[188] In this process, a known active site has a compound virtually inserted into a known active site (i.e. "docked"), and the drug-receptor interaction is calculated under the assumption that a strong interaction is an indicator for potential activity.
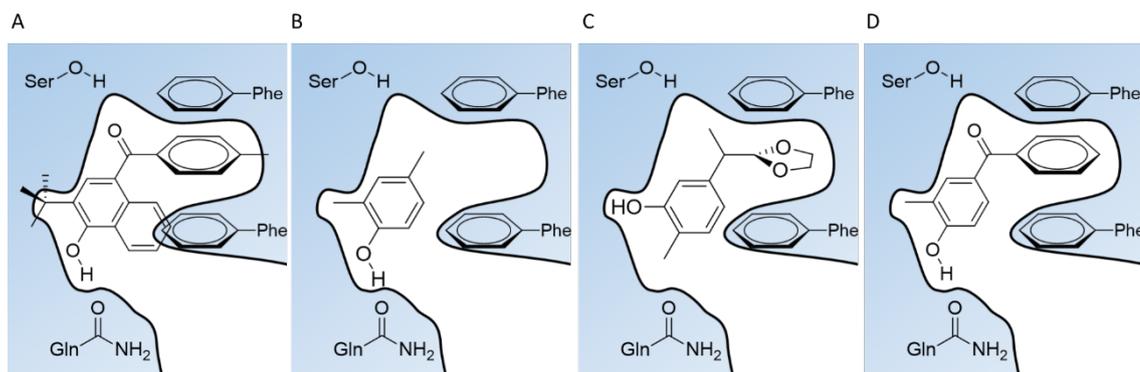
Figure 15 - An example of molecular docking, showing four example compounds docked to an active site. (A) - compound has a poor fit due to being too large, (B) - compound has a poor fit due to being too small, (C) - compound has a poor fit due to interacting poorly with adjacent residues, (D) - compound has a good fit due to being an appropriate size and shape and interacting well with nearby residues‡

Another SBDD method, which approaches the problem of identifying compounds that interact strongly with an active site in a different way, is to generate a compound based on combining functional groups that interact well with key residues of the active site into a singular compound, which is often known as fragment-based drug design. A hypothetical example of this process is

---

‡ Original graphic created solely for this thesis

shown in Figure 16, and shows the process broken down into four steps. The steps are as follows: (1) determination of the active site, (2) the identification of key interacting residues such as serines, glutamines and phenylalinines, (3) determining what fragments will interact with these residues (e.g. by hydrogen bonding, pi stacking, hydrophobic interactions etc.), and (4) constructing or assembling these fragments into a drug that can be synthesised.
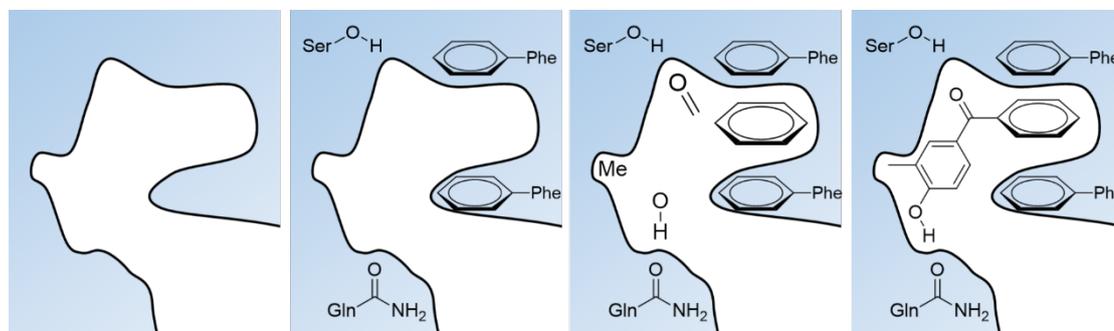
Figure 16 - An approach to SBDD - Determine active site, identify interacting residues, choose fragments that interact with these residues, combine fragments into drugs[§]

As SBDD approaches such as those described above require knowledge of the active site. If the active site is unknown or uncharacterised, an LBDD-based approach is required. In this approach, it is typical for series of analogues of a known active compound to be produced in order to investigate the role of certain structural features. This procedure can be roughly split into two separate approaches: depth first or breadth first.

A depth-first approach focuses on making small changes to a compound which is typically known to already have high activity. It may focus on improving qualities such as toxicology, solubility or blood-brain barrier permeability, or may just focus on further improving the activity of the lead compound. This approach typically investigates a more local area of chemical space, but will investigate that area in depth. The work is also often carried out in an iterative manner; the best compound from the first series of analogues will go on to be the lead compound for the next series.

A good example of this approach is the work of Wasik *et al*., who made analogues of 3,4-methylenedioxymethamphetamine, a psychoactive drug with known anti-lymphoma properties.[189] Wasik *et al*. iteratively produced two series of analogues in order to reduce the

---

[§] Original graphic created solely for this thesis

psychoactive qualities of the drug in order to make this class of compounds a viable therapy. The analogues produced, shown in Figure 17, show that substitutions at only one position of the lead compound was investigated. In the first series, it was found that the phenyl analogue **106** was the best candidate of the first series (based on anti-lymphoma activity and psychoactive levels). The phenyl analogue was therefore used as the lead compound for the second series, which showed the bulkier aromatic substituents in analogues **116**, **117** and **118** made for the best drug candidates.
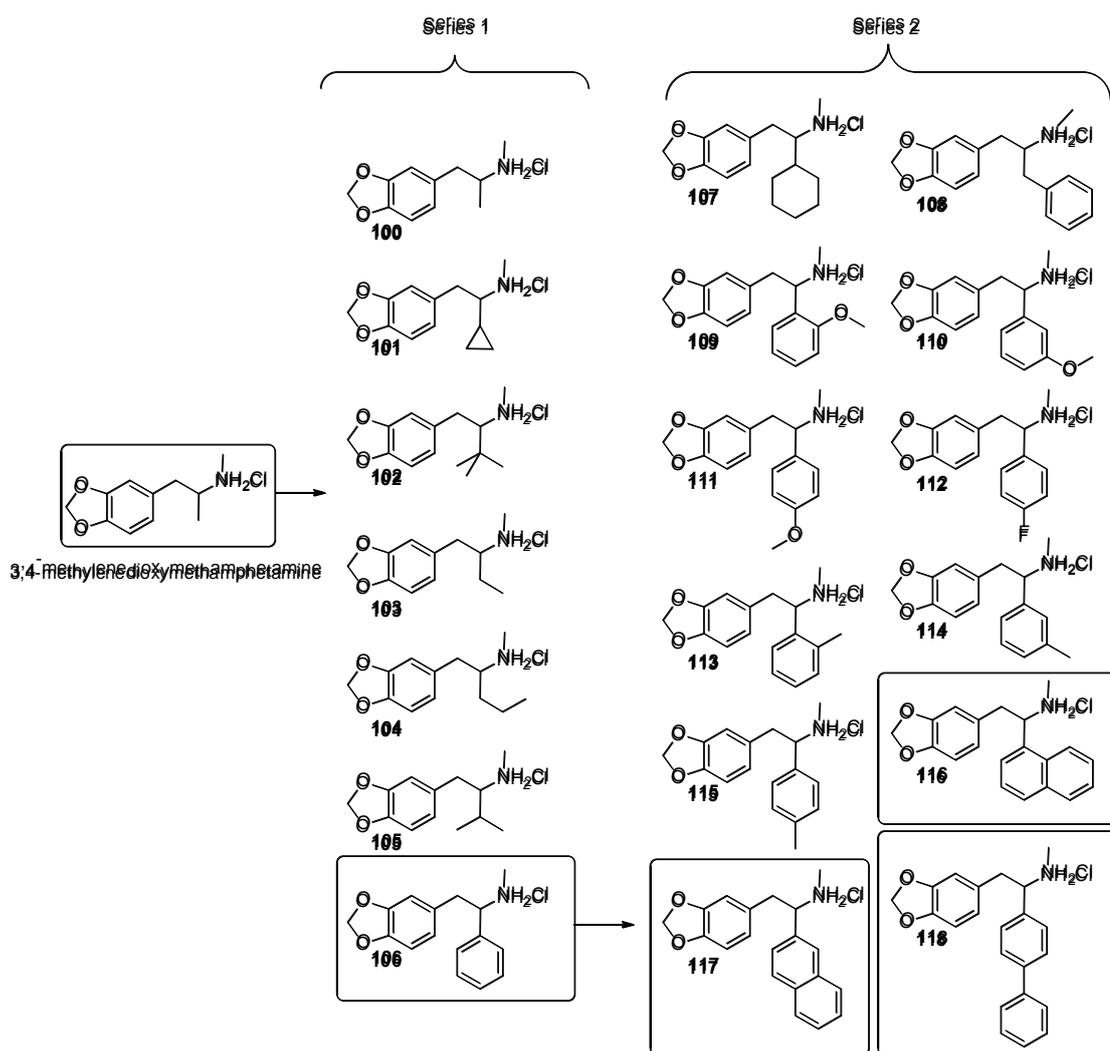
Figure 17 - The two series of compounds produced by Wasik et al.,[189] showing the most active compound(s) of each series in a box

An alternative approach to a depth-first analogue search is a breadth-first analogue search, in which a much broader area of chemical space is investigated, albeit investigated less thoroughly

than in a depth-first approach. This often means that the effects of more structural features are investigated.

One example of a breadth-first approach to finding functional analogues is to change the very core of the molecule, in an approach known as scaffold hopping. In this approach, the core (or "scaffold") of the molecule is changed, but an attempt is made to retain much of the surrounding derivatisation. In small-molecule drug discovery, the scaffold of a molecule can simply be substituted for a similar scaffold (e.g. replacing an indole with a naphthalene), by opening or closing a ring (e.g. by breaking a C-C bond in a cyclohexane ring), or by replacing the scaffold with something that gives the same overall topology.[190] A classic and very early example of successful scaffold hopping was the development of the pain relief medication tramadol from the opioid morphine. Tramadol was developed by breaking three of the five rings of morphine, allowing for a more flexible molecule which could better fit the opioid receptors (Figure 18).
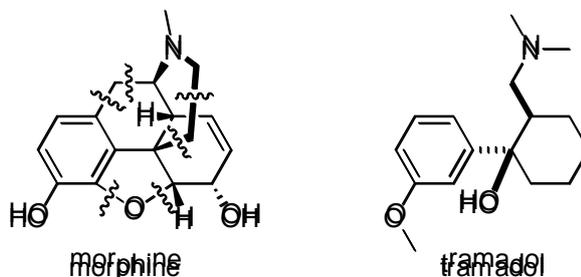
Figure 18 - Morphine (left) and its scaffold-hopped analogue tramadol (right)

Typically, a breadth-first approach is carried out when the lead compound has reasonable but inadequate activity. Qualities of the analogues of the lead compound such as solubility or toxicity are typically not considered at this stage, as these issues can be addressed later on in the drug development process.

The depth-first and breadth-first approach are different ways of addressing the problem of searching chemical space. Within a drug development programme, time is a finite resource, so only a finite number of analogues can be tested. The depth-first approach focuses on the more immediate surrounding chemical space, potentially investigating as little as a single substituent of a drug. The breadth-first approach allows a much wider area of chemical space to be investigated, but the level of detail that each structural feature is investigated to is inherently much lower.

Searching Chemical Space – Similarity Searching

It is apparent that we are never going to run out of chemical space to explore, and we are never going to attain a comprehensive library of known drug-like compounds. The method of drug discovery can therefore shift to finding efficient ways to virtually explore chemical space, which means moving away from the traditional method of synthesising compounds the chemist believes to be similar and moving towards testing compounds that some software considers to be similar. This is of particular interest when one considers that humans are notoriously poor and inconsistent at identifying potential lead or drug-like compounds.[191,192]

However, where the traditional chemist has chemical intuition and experience from where to draw inspiration when designing analogues, software needs to draw from databases. There are many chemical databases which are differentiated based on their sizes and slightly different focuses.

The smallest of the major databases is DrugBank,[193–195] which currently contains 8206 entries. This database contains over 2000 FDA-approved drugs (the vast majority of which are small molecules), some nutraceuticals, and over 6000 experimental drugs. It is well curated, and each entry (known as a DrugCard) includes detailed information on both the physical properties of the drug, pharmacokinetic profiling and information on the drug's target(s). All compounds within the database are therefore known to have some biological activity. The database is free for non-commercial use, and can be interrogated either through the browser, or downloaded and interrogated offline. However, with only 8206 entries, it is a very small database that covers a tiny portion of chemical space.

Next in order of size is BindingDB,[196–200] with over half a million entries for drug like molecules, as well as information on over 5000 proteins and over 1.2 million small molecule/protein interactions. As the name suggests, it is a database of the interactions between small, drug-like molecules and druggable protein targets. The database is free to use and is downloadable, and because the focus is on drug-protein interaction, it can be of particular use in projects that focus on such interactions.

Considerably outside the realm of what could be tackled by a human, ChEMBL[201,202] contains over 1.5 million compounds (as well as data on over 11 thousand targets), along with data on calculated LogP values, polar surface area, hydrogen bond donors and acceptors and number of rotatable bonds. It is free to use and the database is downloadable. There are, however, no subsets of the database available if one wishes to search for, say, drug fragments rather than

drug-like molecules. The result is that interrogating the database can be more computationally expensive, and that there are fewer relevant compounds available in the database.

The ZINC database[4,5] contains over 22 million compounds, but can be split into database subsets such as "lead-like", "fragment-like", "drug-like" etc. If one is carrying out a similarity search, this allows for a much more efficient methodology, as compounds of significantly different mass do not need to be included in the search. In the case of this project, the lead compound is considered fragment-like due to having a molecular weight below 250 g mol^{-1} and fewer than five rotatable bonds,[203] so the fragment-like subset of over 800 thousand compounds would be used. Furthermore, ZINC's novelty is that all compounds within the database are purchasable, meaning that in theory, compounds that are similar could be purchased and tested, drastically increasing throughput.

The largest free database of known compounds is the PubChem database,[204,205] with over 89 million compounds. The focus of the PubChem database is to aid medical research, and to that end it includes the activities of the included compounds against biological assays through its BioAssay database,[206–209] which itself includes well over a million entries. Although a limit of 1000 atoms and 1000 bonds exists for entry into the database, this is more than sufficient for small drug development. It is free to use and download, and can be interrogated online through a browser or offline.

The Chemical Abstracts Service (CAS), run by the American Chemical Society (ACS), has a proprietary chemical database called CAS REGISTRY, currently containing over 114 million entries. Although large, the database is not specific to drug-like compounds and also includes information on salts, polymers, alloys, minerals and other inorganic compounds. The database is also not free to use, and cannot be downloaded, so has to be interrogated through other services such as SciFinder.[210] The major advantage of CAS REGISTRY is that it is generated from literature mentions, so all compounds can be found in the scientific literature. This often means that synthetic procedures for compounds are easy to find.

By far the largest chemical database is the GDB-17 database,[136,211] which contains over 116 billion entries, making it orders of magnitude larger than any other chemical database. The GDB-17 database was generated by enumerating molecules of up to 17 atoms of carbon, oxygen, nitrogen, sulfur and halogens, and then applying restrictions on chemical stability and synthetic feasibility, so therefore predominantly contains compounds that have never been characterised. The result is that the GDB-17 database is the most conclusive mapping of the chemical space

relevant to drug discovery. This database does, however, have a number of drawbacks. Primarily, it is far too large to be conclusively interrogated using easily accessible computational hardware. Secondly, the vast majority of the database is of previously uncharted chemical space. This means that although the compounds in these regions may be synthetically accessible in theory, much of the database is practically inaccessible using current synthetic methods. Therefore, the GDB-17 database would be very likely to yield novel compounds when carrying out a similarity search, but the trade-off would be that the synthetic throughput would be very low and difficult.

CHAPTER 3 – INTRODUCTION TO QUANTUM CHEMICAL METHODS FOR ENERGETIC CALCULATIONS OF MOLECULAR SPECIES

3.1   Introduction

As mentioned earlier, the generation of ROS, and therefore the propensity of a species to form a radical, may be related to the anticancer activity of a species. The mechanistic steps involved in converting a neutral species to form a radical *via* various pathways are defined in Scheme 8 (Section 1.2.2) as the bond dissociation enthalpy (BDE), proton affinity (PA), electron transfer enthalpy (ETE), adiabatic ionisation potential (AIP) and proton dissociation enthalpy (PDE). To calculate the values of the BDE, PA, ETE, AIP and PDE, the neutral, radical, anionic and radical cationic species must first be conformationally optimised, and their energies calculated. This is a routine computational problem that can be tackled by a large range of commercially and freely available software packages.

3.2   The Schrödinger Equation

The aim of geometry optimisation and energy calculations is to generate an approximate solution to the time-independent Schrödinger equation, an equation that relates the wavefunction of a system to its total energy. The general form of the time-independent Schrödinger equation is given in Equation 1, where Ψ is the wavefunction, E is the total system energy and $\hat{H}$ is the total Hamiltonian.

$$\hat{H}\Psi = E\Psi$$

Equation 1 - General form of the time-independent Schrödinger equation

The equation for the Hamiltonian $\hat{H}$, is given in Equation 2, where M and N refer to the nuclei and electrons respectively, μ refers to the mass of a particle, $\hbar$ is the reduced Planck's constant, $\vec{R}$ and $\vec{r}$ refer to the position of a nucleus and electron respectively, Z is the nuclear charge and ε_0 is the vacuum permittivity.

$$\hat{H} = -\sum_{A}^{M} \frac{\hbar^2}{2\mu_A}\nabla_{\vec{R}_i}^2 - \sum_{i}^{N} \frac{\hbar^2}{2\mu_i}\nabla_{\vec{r}_i}^2 + \sum_{A}^{M}\sum_{B}^{M} \frac{Z_A Z_B e^2}{4\pi\varepsilon_0|\vec{R}_A - \vec{R}_B|} + \sum_{i}^{N}\sum_{j}^{N} \frac{e^2}{4\pi\varepsilon_0|\vec{r}_i - \vec{r}_j|}$$
$$- \sum_{i}^{N}\sum_{A}^{M} \frac{Z_A e^2}{4\pi\varepsilon_0|\vec{R}_A - \vec{r}_i|}$$

Equation 2 - The Hamiltonian component of the Schrödinger equation

Within this equation, the kinetic energy of the nuclei (T_{nuc}) is given by the first term, the kinetic energy of the electrons (T_{el}) is given by the second term, the potential energy of repulsion between nuclei ($V_{nuc-nuc}$) is given by the third term, the potential energy of repulsion between electrons (V_{el-el}) is given by the fourth term, and the potential energy of attraction between

each electron and nuclei (V_{nuc-el}) is given by the fifth term. The Hamiltonian operator can therefore be written in a more simple form, as shown in Equation 3.

$$\hat{H} = T_{nuc} + T_{el} + V_{nuc-nuc} + V_{el-el} + V_{nuc-el}$$

Equation 3 - The simplified form of the Hamiltonian operator

Since exact solutions to the Schrödinger equation are not available for systems consisting of more than two particles, only approximate solutions may be obtained. The only option is to approximate a solution to the Schrödinger equation, and there are a number of potential methods for doing this. The upcoming section will describe these methods, and explain the justifications for selecting the methodology employed.

3.3   The Born-Oppenheimer Approximation

The first major approximation employed is the Born-Oppenheimer approximation. This states that due to the large difference in mass and mobility between the nuclei and electrons within a molecule, the nuclei can be considered stationary. The justification for this is that a nucleus is often several thousand times greater in mass than an electron. As a result, the nuclei moves much slower and much less than the electrons. The motion of the nuclei can therefore be neglected when considering the motion of electrons.

When using the Born-Oppenheimer approximation, we can consider the nuclei stationary (often referred to as "clamped nuclei") with the electrons moving around them. The outcome of this approximation is that the Hamiltonian operator of the Schrödinger equation can also be separated out into a Hamiltonian for the electrons ($\hat{H}_{el}$) and Hamiltonian for the nuclei ($\hat{H}_N$), as shown in Equation 4.

$$\hat{H}_{el} = T_{el} + V_{el-el} + V_{nuc-el}$$
$$= -\sum_i^N \frac{\hbar^2}{2\mu_i}\nabla^2_{\vec{r}_i} + \sum_i^N \sum_j^N \frac{e^2}{4\pi\varepsilon_0|\vec{r}_i - \vec{r}_j|} - \sum_i^N \sum_A^M \frac{Z_A e^2}{4\pi\varepsilon_0|\vec{R}_A - \vec{r}_i|}$$

$$\hat{H}_{nuc} = T_{nuc} + V_{nuc-nuc} = -\sum_A^M \frac{\hbar^2}{2\mu_A}\nabla^2_{\vec{R}_i} + \sum_A^M \sum_B^M \frac{Z_A Z_B e^2}{4\pi\varepsilon_0|\vec{R}_A - \vec{R}_B|}$$

Equation 4 - The electronic Hamiltonian (top) and nucleic Hamiltonian (bottom), the separation of which are possible as a result of the Born-Oppenheimer approximation

Due to the approximation that the nuclei are stationary, this makes their kinetic energy (T_{nuc}) zero. The nuclear Hamiltonian therefore simply becomes the repulsive nuclear-nuclear

interaction ($V_{nuc-nuc}$), which when the nuclei are stationary is a constant, so is a computationally cheap and therefore short calculation.

3.4   Hartree-Fock theory

Once the Born-Oppenheimer approximation has been employed, the computational bottleneck is calculating the repulsive electron-electron interaction (V_{el-el}). Simplifying this term can be achieved by employing Hartree-Fock theory. Although largely superseded by density functional theory (DFT), Hartree-Fock theory forms the basis for DFT as well as almost all quantum chemistry methods.

3.4.1   Hartree approximation

The spatial molecular orbitals of a species can be roughly approximated as a linear expansion of its constituent spatial atomic orbitals, in what is known as the linear combination of atomic orbitals (LCAO) method. As shown in Equation 5, the spatial molecular orbital of electron i (φ_i) can be defined in terms of a set of K basis functions (ϕ_μ) and a set of expansion coefficients ($C_{\mu i}$).

$$\varphi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu$$

Equation 5 - LCAO approach to constructing molecular orbitals from basis functions

The Hartree approximation uses this approach to simplify the electronic wavefunction in terms of one electron molecular orbitals. The molecular wavefunction is the wavefunction of all electrons within a molecule, and can then be approximated as the product of all of its constituent one electron spatial wavefunctions as seen in Equation 6, where $\varphi_i(\vec{r}_i)$ is the one-electron spatial wavefunction of the ith electron.

$$\Psi_{el} = \Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N) \approx \varphi_1(\vec{r}_1)\varphi_2(\vec{r}_2)\varphi_3(\vec{r}_3)\dots\varphi_N(\vec{r}_N)$$

Equation 6 - LCAO method of constructing molecular orbitals, as described in terms of electronic wavefunctions

As it stands, this wavefunction breaks the Pauli exclusion principle. Pairs of electrons are described as occupying the quantum state due to the spin states of the electrons not being included in the spatial orbitals. To amend this, spin orbitals must be used in place of spatial orbitals. Spin orbitals are generated as the product of a spatial orbital and a spin state (either α or β), which can be either spin up (ω = +½) or spin down (ω = -½), as shown in Equation 7. In a

closed shell molecule, there will be as many occupied spin up atomic orbitals as occupied spin
down atomic orbitals. In a ground state molecule, the lowest energy orbitals will be occupied
and the higher energy orbitals (known as "virtual orbitals") will be unoccupied.

$$\chi(\vec{x}) = \chi(\vec{r}, s) = \begin{Bmatrix} \varphi(\vec{r})\alpha(\omega) \\ or \\ \varphi(\vec{r})\beta(\omega) \end{Bmatrix}$$

Equation 7 - Generation of spin orbitals from spatial orbitals

The electronic wavefunction, Ψ_{el}, can now be approximated as the product of all one electron
spin orbitals, as shown in Equation 8.

$$\Psi_{el} = \Psi(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N) \approx \chi_1(\vec{x}_1)\chi_2(\vec{x}_2)\chi_3(\vec{x}_3) \dots \chi_N(\vec{x}_N)$$

Equation 8 - Approximating the electronic wavefunction as the product of all one electron spin orbitals

3.4.2   Slater determinants

Equation 8 still does not fully satisfy the Pauli exclusion principle, which also states that a many-
electron system must have the property of antisymmetry, requiring that if two electrons are
interchanged, there must be a change of sign of the wavefunction. For example, in a two
electron system, an appropriate equation that satisfies the requirement for antisymmetry is
shown in Equation 9.

$$\Psi(\vec{x}_1, \vec{x}_2) = -\Psi(\vec{x}_2, \vec{x}_1)$$

Equation 9 - A two electron wavefunction that satisfies the requirement for antisymmetry

With a two electron system, this equation is easy to construct, but with many electrons, the
number of interchangeable electrons grows rapidly. One way to assure that the rules of
antisymmetry dictated by the Pauli exclusion principle are adhered to is to use a Slater
determinant. A Slater determinant utilises the mathematical properties of an antisymmetric
matrix which ensures that any multielectronic wavefunction will have the property of
antisymmetry, and the general form of the equation is shown in Equation 10

$$(\vec{x}_1, \vec{x}_2, \dots \vec{x}_N) = \Phi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & \cdots & \chi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & \cdots & \chi_N(\vec{x}_N) \end{vmatrix}$$

Equation 10 - The Slater Determinant for a system of N electrons

Equation 10 describes the Slater determinant (Φ) as a matrix of $N!$ terms, which is then normalised by the factor $1/\sqrt{N!}$. Interchanging any two electrons can be described mathematically by exchanging two rows in the matrix. The result of this will be a change in the sign of the wavefunction, therefore the Slater determinant (and by association the wavefunction) is antisymmetric.

3.4.3   Variation principle

The variation principle, on which Hartree-Fock theory is based, tells us that the calculated value of the Hamiltonian operator, $\hat{H}$, will always be higher in energy (i.e. is an upper bound) to the exact ground state energy, E_{exact}. Therefore, when calculating the energy of a series of trial wavefunctions for a molecule, the best wavefunction is that which has the lowest calculated energy.

Looking back to Equation 5, trial wavefunctions are generated in part by selecting values for the expansion coefficients ($C_{\mu i}$). Due to the variation principle, energy minimisation is therefore achieved by varying these values in order to lower the calculated energy.

3.4.4   The Hartree-Fock equation

Much of the Schrödinger equation has now been simplified, but as mentioned earlier, the computational bottleneck is calculating the repulsive electron-electron interaction (V_{el-el}), the only remaining many-body component of the equation.

The V_{el-el} component is constructed from the sum of all two-electron repulsive interactions, and is shown again in full in Equation 11.

$$V_{el-el} = \sum_{i}^{N}\sum_{j}^{N}\frac{e^2}{4\pi\varepsilon_0\left|\vec{r}_i - \vec{r}_j\right|}$$

Equation 11 - electron-electron interaction term of the Schrödinger equation

As mentioned earlier, the Hartree-Fock approximation is a method for reducing this many body problem into a series of one-electron problems. This is achieved by considering each electron not as interacting with every other electron separately, but as being in an average field caused by all other electrons. This averaged field (i.e. averaged potential) is known as the Hartree-Fock potential, with the Hartree-Fock potential of the ith electron given the symbol $v_{HF}(i)$. The electronic Hamiltonian operator for the ith electron ($\hat{H}_{el}(i)$) is now simply the sum of the one-electron components and the Hartree-Fock potential felt by the electron (Equation 12). Within

Hartree-Fock theory, this approximation of the one-electron electronic Hamiltonian is called the Fock operator (f_i), and is described in terms of the kinetic energy of the ith electron ($T(i)$), the interaction between the Ath nucleus and ith electron (V_{Ai}) and the Hartree-Fock potential.

$$\hat{H}_{el}(i) \approx f_i = T(i) + \sum_{A=1}^{M} V_{Ai} + v_{HF}(i)$$

Equation 12 - The form of the one electron Hamiltonian operator for electron i

Therefore the full electronic Schrödinger equation can be approximated as Equation 13.

$$\sum_{i}^{N} \hat{H}_{el}(i) \, \Psi(\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_N) = E\Psi(\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_N)$$

Equation 13 - Approximation for the Schrödinger equation after applying the Hartree-Fock approximation

This equation can then be broken down into separate equations for each electron (Equation 14) to yield the Hartree-Fock equations. The eigenvalues (ε_i) in this equation are equivalent to the molecular orbital energies.

$$f_i \chi_i(\vec{x}_i) = \varepsilon_i \chi_i(\vec{x}_i)$$

Equation 14 - The Hartree-Fock equation

3.4.5 Self-Consistent Field Method

The Schrödinger equation would now be in a computationally solvable state if a calculation could begin. However, in order to calculate the eigenvalue for the ith electron (ε_i), one must know the Hartree-Fock potential ($v_{HF}(i)$), but in order to know the value of the Hartree-Fock potential, the form of the orbitals ($\chi_i(\vec{x}_i)$) must be known. The orbitals can't be known until the Hartree-Fock equation is solved, so the equation is therefore unsolvable analytically.

A workaround for this problem is rather than attempt to analytically solve the Hartree-Fock equation, to instead take an initial guess of the orbitals (which includes some adjustable parameters), and use this to construct a Fock operator. This Fock operator can then be used to make a better approximation of the orbitals, which can then be used to construct an improved Fock operator. This iterative process continues until the improvement in the precision of the Fock operator falls to within predefined acceptable levels. This method of iteratively improving the solution to such an equation is known as the self-consistent field (SCF) method, and when applying this method to the Hartree-Fock equation it is known as the Hartree-Fock self-consistent field (HF-SCF) method.

3.5   Basis sets

So far, as shown in Equation 5, molecular orbitals have been described in terms of their basis functions (ϕ, also known as atomic orbital functions) and sets of expansion coefficients (C), but the nature of the basis functions has not yet been described.

Within computational chemistry, a basis function is a mathematical descriptor of an atomic orbital, and each orbital of each element has its own basis function or functions. Because molecular orbitals are considered as a linear combination of atomic orbitals (i.e. the LCAO method), they are calculated as a linear combination of basis functions.

Series of basis functions for a series of elements are collated into basis sets, with almost all basis sets only describing a subset of the periodic table. Basis sets however do not perfectly describe orbitals, and there are hundreds of basis sets to choose from, with different basis sets being more appropriate for certain tasks. Selecting the right basis set is important in order to get data which more closely matches experiment.

3.5.1   Minimal Basis Sets

As the name suggests, minimal basis sets are the smallest and therefore computationally cheapest basis sets available. They are defined by having only one basis function for each occupied atomic orbital, so hydrogen is described by only one basis function (for the 1s orbital) and carbon is described by five basis functions (for the 1s, 2s, $2p_x$, $2p_y$ and $2p_z$ orbitals).

Although calculations that use minimal basis sets are computationally inexpensive, these basis sets typically describe orbitals poorly, so are rarely used when calculating the properties of small molecules.

3.5.2   Split Valence Basis Sets

Single basis functions struggle to correctly describe the electron density within an orbital over a wide range of radii. The solution to this is to include multiple basis functions that are designed to describe the electron density of an orbital at different radii, i.e. different basis functions will describe orbitals of different sizes which, when overlapped, will more accurately describe realistic orbitals properties. The downside is that more basis functions means an increase in computational cost.

When including split valence orbitals in a basis set, priority is given to splitting the valence orbitals as these orbitals have the biggest impact on the properties of a molecule. If more

computational time or power is available, core orbitals can also be split. When all orbitals are split into two basis functions, this is known as a "double-zeta" (or double- ζ) basis set, and when all orbitals are split into three basis functions, this is known as a "triple-zeta" (or triple- ζ) basis set.

3.5.3   Extended Basis Sets

In order to further improve the accuracy of basis sets, there are two more additions that can be included: polarisation functions and diffuse basis functions.

Polarisation functions allow previously inflexible basis functions to be distorted by nearby nuclei, much like how orbitals are polarised.
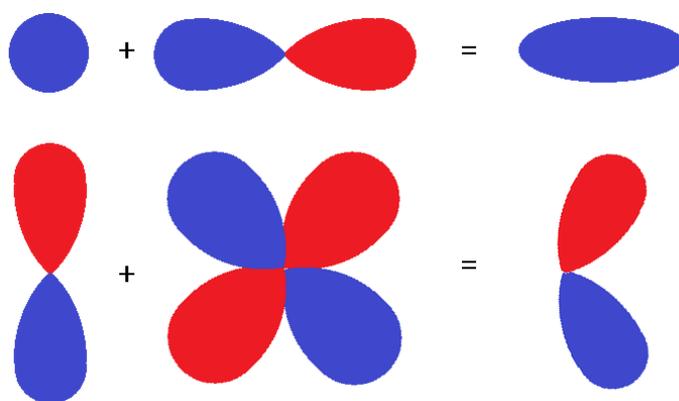
Figure 19 - Impact of adding a d *polarisation function to an* s *orbital (top) and of adding a* p *polarisation function to a* d *orbital (bottom)*[**]

Polarisation functions are typically added first to heavy atoms (i.e. all atoms except hydrogen), and basis sets which have these polarisation functions are denoted as such with an asterisk (or "star") in their name, for example the 6-311G* basis set is the 6-311G basis set with such added polarisation functions. If more computational time or power is available, polarisation functions are added to the hydrogens, and this is denoted by two asterisks, for example the 6-311G** basis set is the 6-311G basis set with added polarisation functions on all atoms.

Diffuse basis functions are a further potential addition to basis sets. Diffuse basis functions model the portions of an orbital that are far from the nucleus, which are necessary to more accurately model anions that typically have a spread out charge. Diffuse basis functions on heavy atoms are denoted by the addition of a plus (or in some cases the term "aug") within the

---

[**] Original graphic created solely for this thesis

name of the basis set, such as 6-311+G or aug-cc-pVDZ. Additionally, having diffuse basis functions on hydrogen atoms is denoted by the addition of two plusses within the name of the basis set, such as 6-311++G.

3.6 Density Functional Theory

A major deficiency of Hartree-Fock theory is the treatment of electron-electron interactions in an averaged way *via* the use of the Hartree-Fock potential, v_{HF}, which results in instantaneous electron-electron interactions (*i.e.* electron correlation) being ignored. Although there are methods for accounting for electron correlation, they are typically either inaccurate or computationally very expensive. A further limitation of Hartree-Fock theory is that it doesn't scale well for use on large molecules. The computational expense of Hartree-Fock calculations scales as N^4, where N is the number of basis functions. This computational cost can be reduced somewhat by ignoring the interaction of orbitals which are a considerable distance from one another, but the Hartree-Fock method is still impractical for use with large molecules or large basis sets.

Due to the poor scaling of the Hartree-Fock method, an alternative method is required. One option is to employ semi-empirical methods, whereby computational complexity is reduced by replacing some of the intensive calculations with experimentally derived values. Although these methods are computationally cheaper, there is often a trade-off with accuracy. Alternatively, improved *ab initio* methods can be employed. *Ab initio*, which translates to "from the beginning", describes techniques for determining properties using only methods derived from first principles, i.e. no experimental or empirical parameterisation is employed. Although not a true *ab initio* method,[††] density functional theory (DFT) is now widely considered as the predominant method for calculating structures, energies and properties of small molecules. It is an expansion of HF theory, retaining many similarities, but DFT methods have been shown to additionally account very well for electron correlation

Rather than considering the wavefunction as the fundamental unknown, DFT instead focuses on the electron probability density, which can be described as the probability of finding an electron

---

[††] Some of the functionals used in this thesis (described later) are parameterised, therefore the methods employed are technically classed as semi-empirical. However, the vast majority of the method is derived from first principals.

in a specific area. The functionals in DFT therefore refer to functionals of electron density, whereas in the HF method the functionals describe orbitals.

One of the models on which DFT is based is the Thomas-Fermi (TF) model, which actually predates the Hartree-Fock method. The Thomas-Fermi model equates a value for energy (E_{TF}) as a summation of the electronic kinetic energy (T_{el}), the electron-nucleus interactions (U_{el-nuc}) and the electron-electron interaction (U_{el-el}), as shown in Equation 15. This yields an equation that is strikingly similar to the electronic Hamiltonian used in the Hartree-Fock method after application of the Born-Oppenheimer approximation (Equation 4).

$$E_{TF} = T_{el} + U_{el-nuc} + U_{el-el}$$

Equation 15 - Thomas-Fermi method for calculating the energy of a system

Although this was an important first step, values obtained for the kinetic energy of the electrons (T_{el}) were a large source of error in the energetic calculation for a system. Although reasonable values could be obtained for atomic systems, the method could not handle molecular systems.

3.6.1 Hohenberg-Kohn Theory

The basis of DFT first emerged over 30 years later in the form of the two Hohenberg-Kohn theorems, as developed by Hohenberg and Kohn.

The first theorem states that the ground state electron density of a system, termed $\rho(r)$, uniquely determines the external potential ($V_{ext}(r)$, the term used in DFT to describe the interactions between electrons and nuclei and previously termed V_{nuc-el}), which therefore determines the Hamiltonian operator, which therefore defines the entire system. This was proven by showing that two different external potentials couldn't possibly result in the same electron density.

The second theorem states that the form of $\rho(r)$ that results in the lowest total energy of the system is the ground state density of the system, and that this total energy is the ground state energy of the system. It also states that, as in the HF method, these value can be obtained variationally.

Hohenberg and Kohn were therefore able to show that the exact electronic energy functional ($E_{el}[\rho]$) for the ground state of a molecule could be determined by the electron probability density (as stated by their first theorem), and that it was equal to the sum of the electronic kinetic energies, the electron-nucleus interactions and the classical electron-electron

interactions (collectively termed $E_{classical}[\rho]$), plus the exchange-correlation energy ($E_{XC}[\rho]$) as shown in Equation 16.

$$E_{el}[\rho] = E_{classical}[\rho] + E_{XC}[\rho]$$
$$E_{classical}[\rho] = T_{el}(r) + V_{el-el}(r) + V_{ext}(r)$$

Equation 16 - One of the findings of Hohenberg and Kohn

The equation for $E_{classical}[\rho]$ is again very similar to those seen in the Hartree-Fock method (Equation 4) and the Thomas-Fermi method (Equation 15). The $E_{XC}[\rho]$ term includes a correction to T_{el} to account for electron-electron interaction as well as accounting for the non-classical electron-electron interaction energy

Again, although Hohenberg-Kohn theory gives information on the form of these equations, it gives no way of determining its value. It can, however, be tackled using the variational principle, as only the ground state electron density will yield the ground state electronic energy functional. In other words, the lower the calculated electronic energy of a system, the closer the electron density is to the real solution. We can therefore calculate the electronic energy by finding the electron density that minimises E_{el}.

3.6.2   Kohn-Sham Theory

Although the Hohenberg-Kohn theorems are a very important proof of concept, they don't help in determining solutions to these equations. Kohn-Sham theory attempts to identify the Hohenberg-Kohn functional through the use of two approximations.

The first approximation is that the electrons are considered to be non-interacting, and moving through a field of constant potential generated by the other electrons, yet with the same electron density as there would be with interacting particles. This fictitious system is called a "Kohn-Sham system". This system is then computationally inexpensive to solve in a similar fashion to within Hartree-Fock theory, by linearly combining the one-electron orbitals to give the total density of the system. The Kohn-Sham system can be described by Kohn-Sham equations, a common form of which is given in Equation 17, with $v_{eff}(\vec{r})$ representing the external potential through which the electrons move (i.e. the "Kohn-Sham potential"), $\phi_i(\vec{r})$ is known as the Kohn-Sham orbital, and ε_i is the orbital energy of said Kohn-Sham orbital.

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}(\vec{r})\right)\phi_i(\vec{r}) = \varepsilon_i\phi_i(\vec{r})$$

Equation 17 - A common representation of the Kohn-Sham equations

The Kohn-Sham potential, $v_{eff}(\vec{r})$, is made up of the electron-electron interaction, the external potential acting on the system (v_{ext}) and the exchange-correlation potential (v_{XC}), as shown in Equation 18.

$$v_{eff}(\vec{r}) = \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' + v_{ext}(\vec{r}) + v_{XC}(\vec{r})$$

Equation 18 - Constituent factors in the Kohn-Sham potential

The electron density of the system can be described in terms of Kohn-Sham orbitals (Equation 19).

$$\rho(r) = \sum_i^N |\phi_i(\vec{r})|^2$$

Equation 19 - The electron density in terms of Kohn-Sham orbitals

Because the Kohn-Sham potential depends on the electron density, a solution can be determined by using the self-consistent field (SCF) method.

The second approximation is to construct the so-called Hohenberg-Kohn functional, $F[\rho]$, to group all unknown variables into a single component. In a similar notation as given previously for Hartree-Fock theory, the functional can be broken into the sum of all single electron kinetic energies ($T_{el}[\rho]$), the coulombic electron-electron interaction energy ($J[\rho]$, considered as an electron interacting with a mean field generated by the other electrons), and the exchange and correlation energy ($E_{XC}[\rho]$), as shown in Equation 20.

$$F[\rho] = T_{el}[\rho] + J[\rho] + E_{XC}[\rho]$$

Equation 20 - Rearranged form of the Hohenberg-Kohn functional

The exchange-correlation energy component ($E_{XC}[\rho]$) is therefore made up of a corrective kinetic energy term (T_c) and a non-classical contribution to electron-electron interaction energy (E_{ncl}). The equation for $E_{XC}[\rho]$ is shown in Equation 21.

$$E_{XC}[\rho] = T_c + E_{ncl}$$

Equation 21 - Equation for the exchange-correlation energy component of the Kohn-Sham equation

3.6.3   The Exchange-Correlation Functional

The exchange-correlation functional, $E_{XC}[\rho]$, cannot be determined systematically, so it must be approximated. There are a number of methods for doing this, but this section will only focus on those methods that are relevant to this project.

The most common method for determining $E_{XC}[\rho]$ is called the local density approximation (LDA). Within this method, electrons are treated as a uniform electron gas (UEG), whereby they are approximated as being averaged evenly over the space of the molecule. The LDA form of E_{XC} is given in Equation 22, which includes $\varepsilon_{XC}(\rho(\vec{r}))$, the exchange-correlation potential per particle in a UEG of density $\rho(\vec{r})$.

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r})\varepsilon_{XC}(\rho(\vec{r})) \, d\vec{r}$$

Equation 22 - The local density approximation method for approximating the exchange-correlation potential

The LDA approach relies on the species under investigation being "soft", i.e. species which are relatively large, have low charge states and are highly polarisable, such as salts. This typically doesn't hold true for drug-like molecules, which are typically only weakly polarisable.

An alternative to the LDA approach is the generalised gradient approximation (GGA). The GGA approach includes both the local electron density and gradient of the electron density, as opposed to the UEG approximation employed within the LDA approach. As a method of calculating exchange-correlation energy, there are many different exchange-correlation functionals for use with GGA. Including the popular functionals B3P86 and B3LYP. Both B3P86 and B3LYP employ the B3 exchange functional, but have different correlation functionals (P86 as opposed to LYP).

B3 functionals such as B3P86 or B3LYP are also known as "hybrid functionals", meaning that they include input on the exchange-correlation energy from multiple methods. In the case of B3 functionals, input from the three methods (hence the "3" part of the functional name) LDA, GGA and Hartree-Fock are included.

3.6.4   Summary of DFT

In conclusion, modern molecular modelling software that utilises the DFT method will attempt to find an approximation to the Schrödinger equation using a variety of simplifications and approximations to calculate the electron probability density of a system. The primary approximation is the Born-Oppenheimer approximation, which considers the nuclei and electrons of a system separately due to their large differences in mass and speed. The Hohenberg-Kohn theorems next tell us firstly that a system can be uniquely described by the ground state electron density of said system and secondly that the lowest calculated energy of a system will be closest in energy to the true energy of the system, and that therefore the energy

of a system can be determined variationally. The next approximation is included within Kohn-Sham theory, and approximates all electrons as non-interacting particles moving through a field of constant potential generated by the other electrons, which drastically simplifies the computational complexity required. The final significant approximation involves the use of a functional to approximate the exchange-correlation energy. Bringing all of this together allows for software to calculate the energies of compounds, and therefore allows the medicinal chemist access to related energetic values such as bond dissociation enthalpies.

DFT, when complimenting empirical evidence, is a powerful tool for giving insight into reaction mechanisms and molecular structures that would be difficult if not impossible to determine experimentally. DFT is well suited to determining relative energies of species such as the relative energy of a radical to its neutral counterpart, and is therefore an ideal method for calculating values such as of BDE, PA, ETE, AIP and PDE. The accuracy and precision of the values obtained will depend largely on the methodological choices made, such as using a relevant functional and basis set, as well as reasonable input geometries. These choices can often be reduced down to a trade-off between computational cost and quality of data obtained, so a solid methodological foundation is essential.

CHAPTER 4 – SYNTHETIC METHODS

4.1   Introduction

As discussed in the introduction chapter, the traditional approach to drug development has typically been to simply synthesise various analogues of a known active compound based on chemical intuition and literature precedent.   This approach is undertaken in this chapter, whereby analogues of both lead compounds are produced with the intention of developing a structure-activity relationship (SAR).

4.2   Design and Synthesis of Analogues of I3C and Compound 5

As was shown earlier, the anticancer activity of indole-3-carbinol (I3C) is thought to be due to the formation of the metabolic active product 3,3'-diindolylmethane (DIM), which is formed *via* a 3-methyleneindolinium intermediate,[116] as shown in Scheme 39.   The activity of the analogous compound **5** is therefore thought to occur due to the formation of an analogous metabolic product, as shown in Scheme 40.
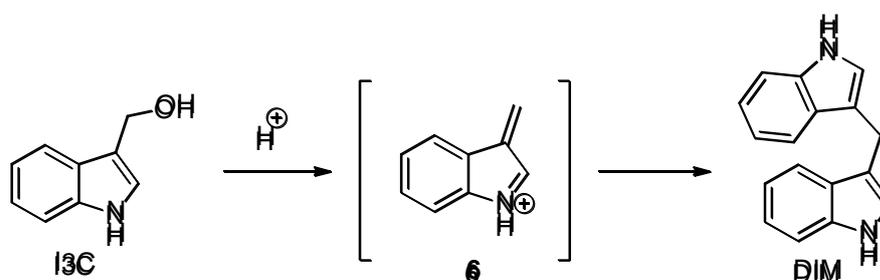
Scheme 39 - Generation of DIM from I3C via a 3-methyleneindolinium intermediate [116]
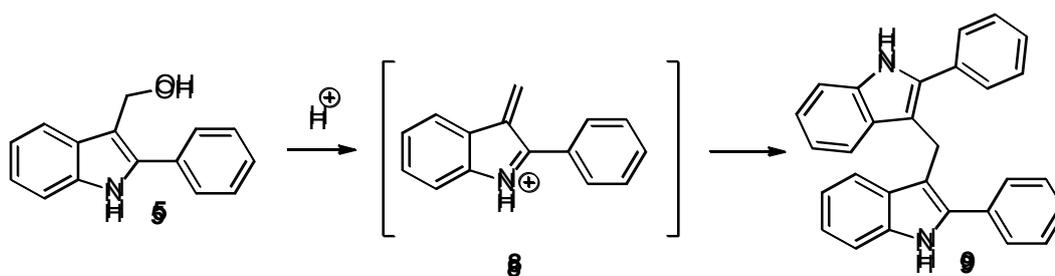
Scheme 40 - Hypothesised analogous mechanism of the degradation of compound 5

I3C (and, by extension, compound **5**) can therefore be considered as prodrugs. This leads to two options when considering the design of analogues; analogues of both the prodrug and the active form can be produced and tested.   In this section, focus will initially be on analogues of the prodrug, and focus on analogues of the active form will follow on.

4.2.1   Analogues of the prodrug form

With prodrugs, a significant quality that can often be controlled is the rate of conversion to the active drug. In the common case of prodrugs that contain a carrier (with some well-known examples including aspirin,[212] progabide[213] and alprazolam[214]), this is often done by changing the carrier group to affect the rate of metabolism, thus the active form of the drug is structurally unaffected. In the case of I3C and compound **5**, the metabolism of the prodrug is more complicated, so designing superior prodrugs while leaving the active form unaffected is difficult.

To more fully understand how structural changes to I3C may affect the metabolism, a more complete metabolic mechanism is required. Such a mechanism, shown in Scheme 41, shows that the structurally most important features of I3C are the hydroxymethyl group at the 3-position, conjugation to delocalise the resultant cation and the nitrogen at the 1-position. The hydroxymethyl group is important as it goes on to produce two leaving groups (H_2O and CH_2O), the conjugation allows the cationic charge to be more delocalised, and the nitrogen is important as it gives the positive charge on an electronegative atom on which to sit, thus lowering the energy of the intermediate cation.
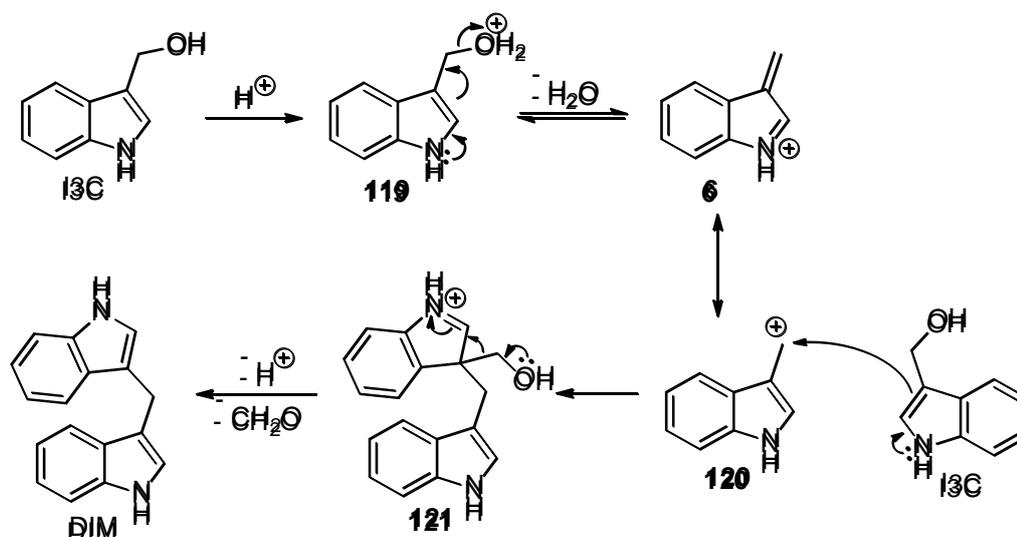
Scheme 41 - A more complete mechanism for the acid degradation of I3C to DIM

The Role of Conjugation

The conversion of I3C to DIM is only possible due to the relatively low energy of the intermediate cation. A significant amount of this stability comes from the delocalisation of charge. By adding more conjugated centres, such as the addition of the 2-phenyl group in compound **5**, the charge can be more delocalised, and therefore will have a lower barrier to formation. In the example

of I3C, the cation can be delocalised over two centres, whereas the cation of compound **5** can be delocalised over five centres (Figure 20)
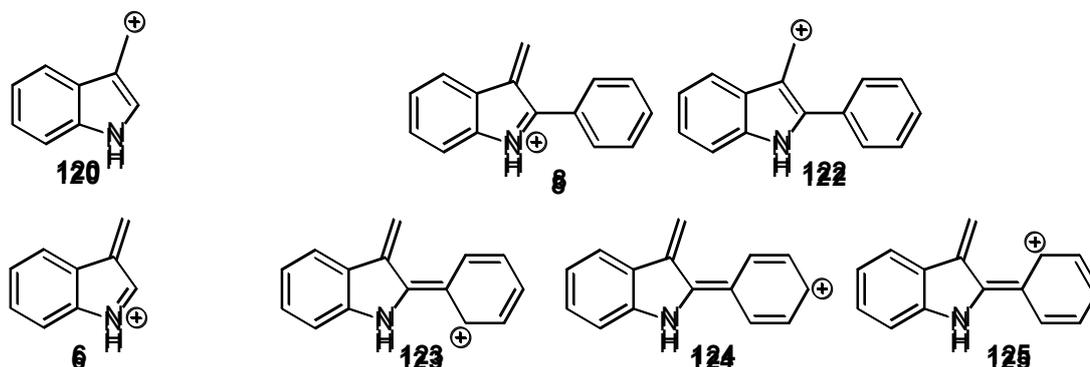
*Figure 20 - The two resonance forms of the I3C cation (left) and the five resonance forms of the cation for compound **5** (right)*

As previously discussed in the introduction, the EC_{50} values for I3C and compound **5** have been reported as 526 µM and 176 µM respectively, so there is a considerable improvement upon the addition of the 2-phenyl ring. To try and elucidate if this is coincidental, an attempt was made to test analogues both with and without a 2-phenyl substituent.

The Role of the Carbinol Group

The carbinol group of I3C (more commonly referred to as a hydroxymethyl group within the wider field of chemistry) is mechanistically involved in the metabolism of I3C. It would therefore be likely to have a significant effect on the metabolism of I3C. Based on Scheme 41, it would appear that having the hydroxymethyl group conjugated with an electronegative atom on which the positive charge can rest is important to the ability of the molecule to form the intermediate cation. By extending the carbon chain of the hydroxymethyl group to a hydroxyethyl group, the conjugation is disrupted, and this should prevent the formation of the active form of the drug. This disruption to conjugation is shown in Scheme 42.
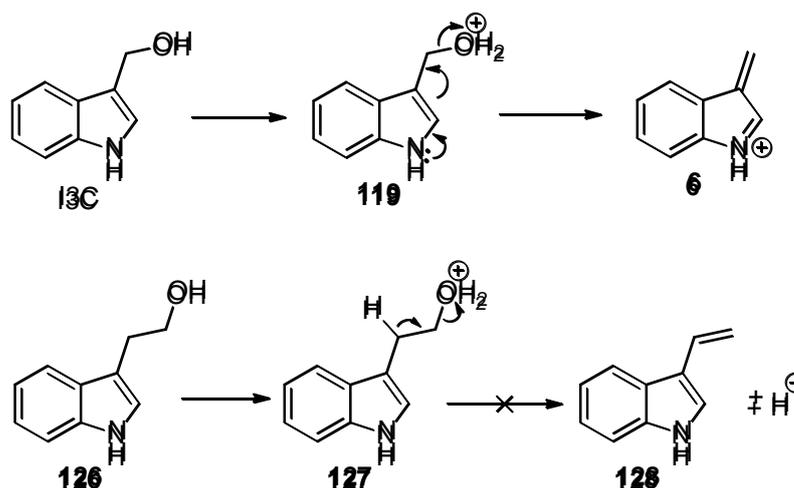
Scheme 42 - The formation of the intermediate cation is possible with a 3-hydroxymethyl group (top) but is not possible with a 3-hydroxyethyl group (bottom)

The hydroxyethyl analogue of I3C (compound **126**) was therefore purchased for testing, and the hydroxyethyl analogue of compound **5** (compound **129**) was synthesised for testing. Their structures, with comparisons to their hydroxymethyl analogues, are shown in Figure 21. The synthesis of compound and **129** is described in Section 8.2.4.
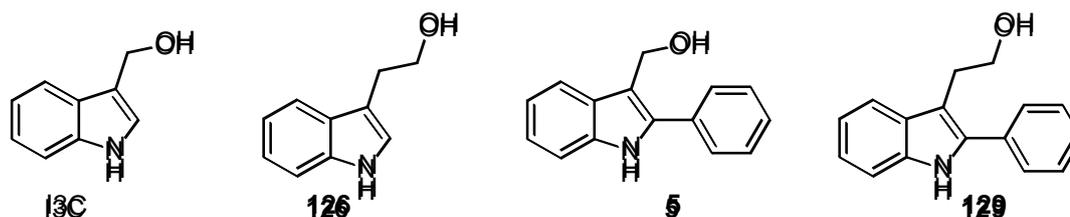
*Figure 21 - From left to right, I3C, compound **126**, compound **5** and compound **129***

An alternative and simpler way of determining the necessity of the 3-hydroxymethyl group is simply to remove it. This approach of stripping a molecule back until it loses its activity is a common approach in determining the structurally vital components of a compound. The dehydroxymethylated compounds indole and 2-phenylindole (shown in Figure 22) were therefore procured for testing.
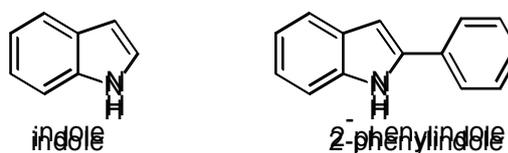
Figure 22 - The structures of indole (left) and 2-phenylindole (right)

The effect of N-*substituents*

Looking back to the 3-methyleneindolinium intermediate (**6**) shown in Scheme 41, the third important structural feature of I3C is the electronegative nitrogen atom on which a cationic charge can sit. By adding substituents that transiently affect the electronics of this nitrogen, or by replacing the nitrogen with a different element, the ease with which their respective metabolites are formed may be affected.

The first step was to synthesise some *N*-substituted analogues. By adding a substituent at the *N*-position, this would affect not just the structure of the prodrug, but also the structure of the active form. It was decided therefore that a hydrolysable substituent should be tested.

One such example of a group that can be hydrolysed off an indole is an acetyl group. *N*-acetylindoles are known to be hydrolysed under both basic[215] and acidic[216] conditions at mild temperatures, so would be expected to be hydrolysed in the body, especially in the presence of numerous cellular proteases. The rate of hydrolysis compared to the rate of dimerisation would determine in which order these two processes would be likely to occur. For example, if the acetyl group is removed too easily, then I3C would be readily formed and the rate of dimerisation would likely be the same as the rate of dimerisation of I3C, so no change in activity might be expected. However, if the acetyl group is stable enough, it might have an effect on the rate of formation of the dimer, which may result in a different activity. These competing pathways are shown in Scheme 43.
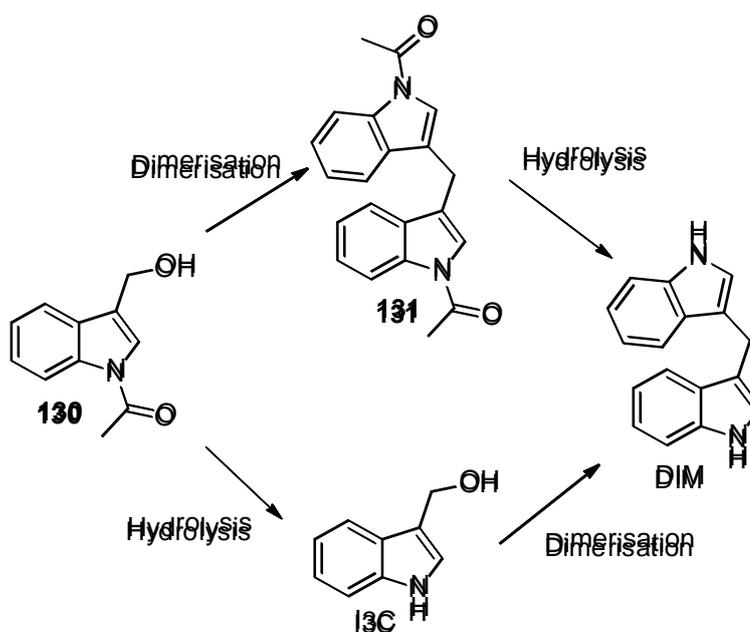
*Scheme 43 - The two potential pathways to DIM from compound **130***

The *N*-acetyl analogue of I3C (compound **130**) and the *N*-acetyl analogue of compound **5** (compound 132) were synthesised as described in sections 8.2.5 and 8.2.6 respectively. Their structures are shown in Figure 23.
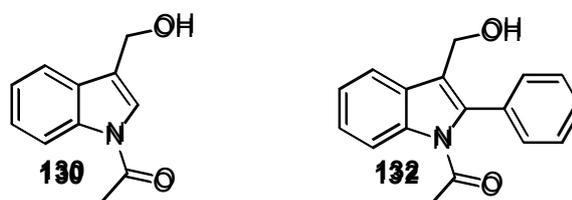
*Figure 23 - The two N-acetyl analogues, compound **130** (left) and compound **132** (right)*

To test if the inclusion of a hydrolysable group is significant, an analogue containing a non-hydrolysable *N*-substituent should be included. This non-hydrolysable group should be as chemically similar as possible to the acetyl group in order to keep the analogue as otherwise similar as possible to the *N*-acetyl analogues. Therefore, the group should have electron withdrawing mesomeric properties, be relatively small, and be a hydrogen bond acceptor. A substituent that fulfils all of these criteria is the tosyl group (also known as the *p*-toluenesulfonyl group, commonly abbreviated as Ts). The *N*-tosyl analogue of I3C (compound **133**, shown in Figure 24) was synthesised (as described in Section 8.2.7) to be tested in direct comparison to I3C and its *N*-acetyl analogue.
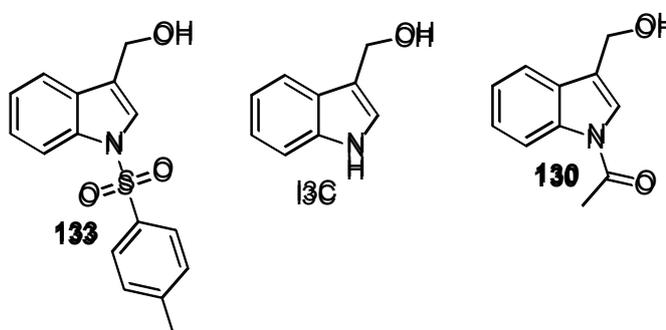
*Figure 24 - Compound **133** (left) alongside I3C (centre) and compound **130** (right)*

Alternative heteroatoms

A second way of affecting the electronic environment at the 1-positon of I3C is to change the heteroatom completely. Analogous heteroaromatic systems to indole include the benzofuran (an oxygen-containing heteroaromatic system) and benzothiophene (a sulfur-containing heteroaromatic system), the structures of which are shown in Figure 25.

Figure 25 - The structures of indole (left), benzofuran (centre) and benzothiophene (right)

Another difference between indole and the two analogues in Figure 25 is the lack of a hydrogen at the 1-position. This may be significant, as the potential of this hydrogen to be abstracted may lead to a significant stabilisation of the 3-methyleneindolinium intermediate during the metabolism of I3C to DIM. Other than this effect, benzofuran contains an oxygen in place of a nitrogen, which is a more electronegative element and may therefore lead to a less readily-formed cationic intermediate. Conversely, sulfur is less electronegative than nitrogen, so the formation of a cationic intermediate from a benzothiophene analogue may be less difficult. As a result of these potentially complex factors, the choice between testing a benzofuran or benzothiophene analogue became fairly arbitrary, and the decision to synthesise the benzothiophene analogue **134** for testing (Figure 26), the synthesis of which is described in Section 8.2.8, was ultimately down to commercial availability of the required starting material.
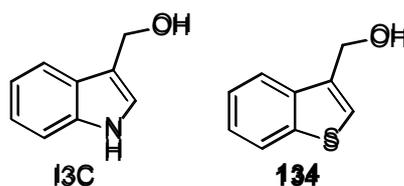
Figure 26 - I3C (left) and its benzothiophene analogue (right)

4.2.2 Analogues of the active compound DIM

The final comparison within this series of compounds is to investigate the necessity of using the prodrug at all. It is known that the metabolism of I3C to DIM doesn't occur with quantitative yields, so one might assume that by testing the isolated active compound, activity would increase along with the increase in quantity of active compound. However, this may be an oversimplification. As discussed in the introduction chapter, the work of Staub[129] and Bjeldanes[116] suggested that in the case of I3C, DIM is generated in the nucleus of the cells. By introducing DIM to cells, it is possible that the compound is unable to make it to the nucleus, where it appears to be acting.

A further issue is solubility. I3C is only sparingly soluble in aqueous media, and by losing the OH group, DIM becomes a highly hydrophobic molecule. Of course, solubility can always be

addressed later in the drug development process through the addition of solubilising groups, so this hydrophobicity need not pose a significant problem at this stage.

To investigate the necessity of using the prodrugs, the active forms of both I3C and compound **5** (DIM and the compound **9**) were purchased and synthesised respectively for testing. Their structures are shown in Figure 27, and the synthesis for compound **9** is described in Section 8.2.3.
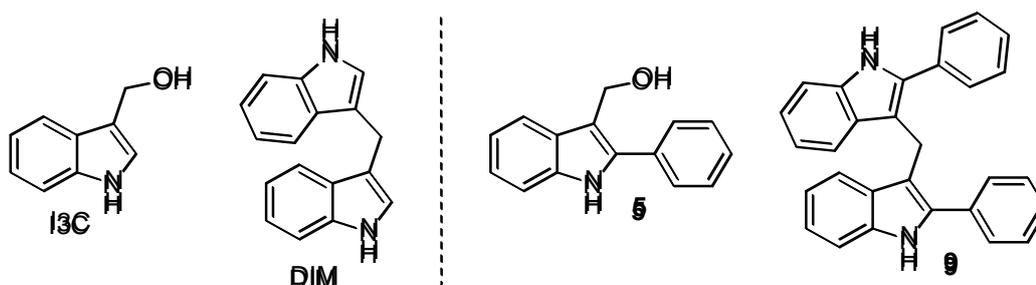
*Figure 27 - The prodrug I3C and its active form DIM (left) and the prodrug compound **5** alongside its active form compound **9** (right)*

4.3   Design and Synthesis of Analogues of Compound 4

As yet, the mechanism of action for compound **4** is unknown (the structure of which is shown in Figure 28), making the design of analogues much more difficult. Because the target is unknown, and it is unknown what form the compound is in when it is delivering its activity, the structural features that are important are unknown. As a result, the only approach is to investigate as many structural features as resources will allow.
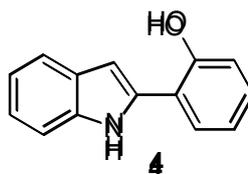
*Figure 28 - The structure of compound **4***

The first striking structural feature is that the compound is simply an indole connected to a phenol. This raises the question as to the necessity of either the indole fragment or the phenol fragment; is anticancer activity observed when either fragment is not present? This is an easy question to address. Indole and a series of eight phenols were purchased in order to determine if one fragment of the molecule was important or if there was a synergistic effect in play between the two fragments. A representation of separating compound **4** into its constituent fragments is shown in Figure 29.

101
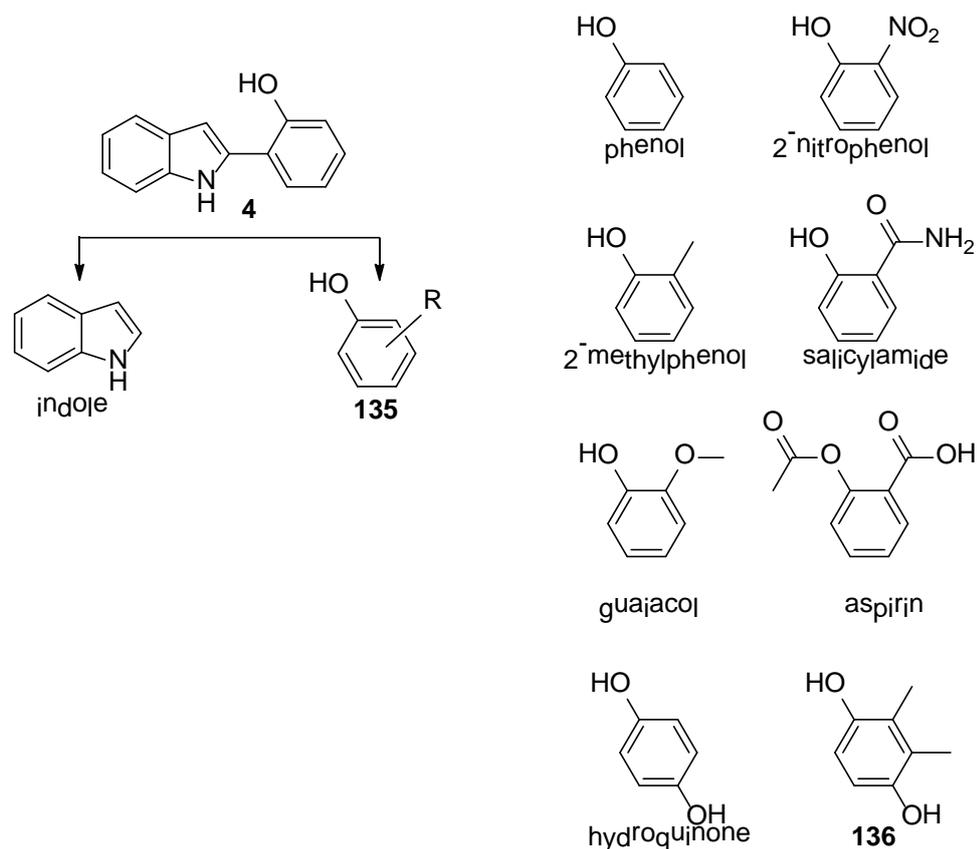
*Figure 29 - A representation of separating compound **4** into its two constituent fragments in order to test the fragments separately, and the eight derivatised phenols that were included in this work*

The next structural feature of interest is that the indole and phenol fragments are conjugated. Current evidence suggests that reactive oxygen species (ROS) are involved in the mechanism of action for compound **4**,[6] so it is possible that the ability of compound **4** to form a radical is relevant to its activity. If phenol was to form the phenolic radical, that radical can be delocalised across four centres, whereas if compound **4** was to form a phenolic radical, it could be delocalised across eight centres. This greater level of delocalisation will lead to a more stable and more readily formed radical, which may be relevant to the activity of this compound.
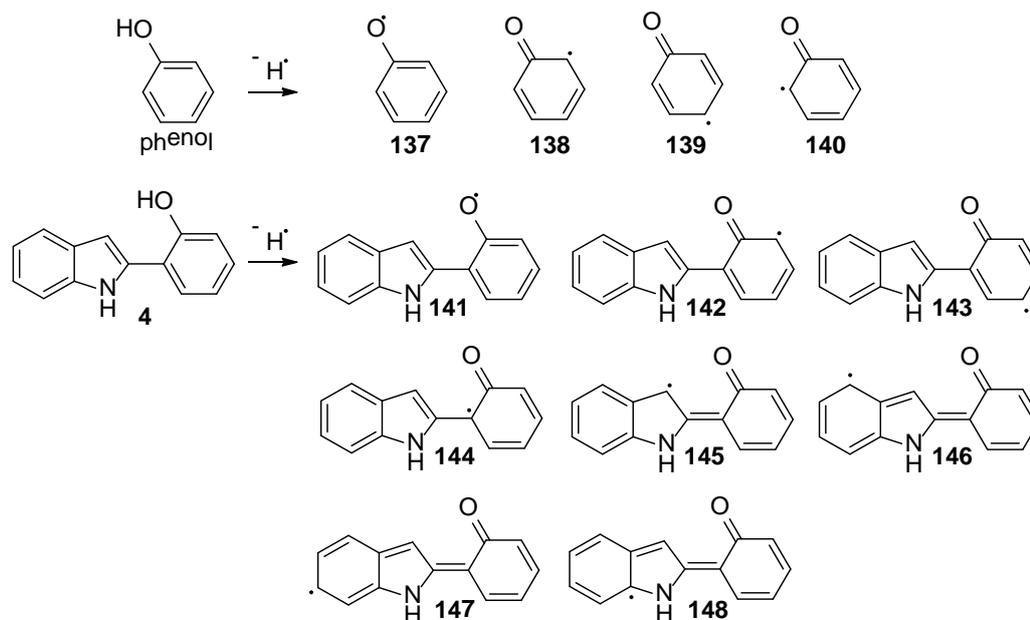
*Figure 30 - The four resonance structures of the phenol radical and the eight resonance structures of the radical form of compound **4***

In order to determine if this conjugation is important, an analogous unconjugated compound would be necessary. There are a number of possible structures for such an analogue, each with its own drawback. The structures of three such possible compounds are shown in Figure 31.
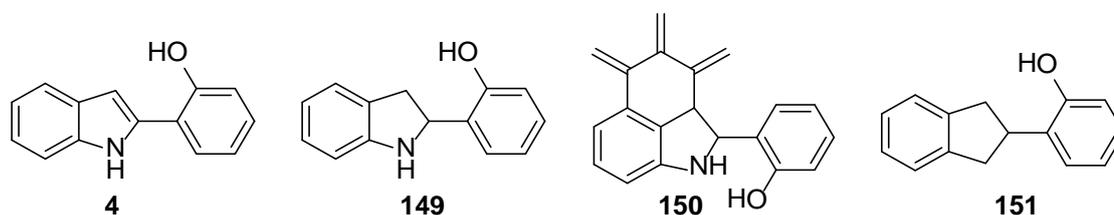
*Figure 31 - Compound **4** (far left) alongside three compounds with reduced conjugation: two indolines (centre left and centre right) and an indane (far right)*

Compound **149**, shown in Figure 31, is an indoline that is simply the reduced form of compound **4**, but differs from compound **4** in two significant ways. The first is that whereas the 3D structure of compound **4** is largely planar, the five membered ring of compound **149** would assume an envelope conformation. The carbon at the 2-position is now also sp^3 hybridised, significantly affecting the physical shape of the molecule. The second significant difference is that the NH group in compound **149** now acts as an amine, significantly increasing the potential for acting as both a hydrogen bond acceptor and donor. This may have an effect on intermolecular interactions with a potential binding site, as well as intramolecular interactions with the phenolic group.

The second analogue, compound **150** attempts to address some of the issues with planarity by trying to fix the 3-position more in plane with the fused aromatic ring. This would however be a significant synthetic challenge due to the large strain that this would introduce to the molecule, as well as significantly affecting both the size, shape and hydrophobicity of the molecule. There may also be steric clashes between the OH group and the new vinyl group, affecting relative position of the phenol ring.

The third analogue, compound **151**, attempts to address the addition of an amine in the first analogue. This indane would obtain a broadly similar overall conformation to compound **149**, yet would be incapable of acting as either a hydrogen bond donor or acceptor. Indoles are known to participate in weak hydrogen bonding,[217–219] so by removing the nitrogen altogether, the potential for an important interaction may be removed.

The choice comes down to either the simpler indoline (compound **149,** which is likely to over-represent the extent of hydrogen bonding) or the indane (compound **151**, which will under-represent the extent of hydrogen bonding). The choice was made to synthesise compound **149**, as described in Section 8.2.9, on account of it having a dipole that more accurately matched the dipole of compound **4**.

Another structural feature to investigate is the necessity of the phenolic OH group. As mentioned earlier, the radical of **4** may be involved in the production of ROS, so a loss of activity would be expected if the OH group or the entire phenolic group was to be removed. To investigate this, indole and 2-phenylindole (the structures of which are shown in Figure 32) were purchased in order to be tested alongside compound **4**.
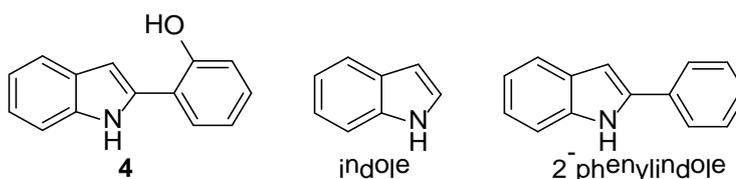
*Figure 32 - The structures of compound **4** (left), indole (centre) and 2-phenylindole (right)*

Another way of investigating the necessity of the phenolic system is to investigate if a similar system has similar activity. The most similar systems to phenolic systems are anilinic (with an NH$_2$ in place of the OH) or thiophenolic (with an SH in place of the OH) systems. The structures of the anilinic analogue (compound **152**) and thiophenolic analogue (compound **153**) of compound **4** are shown in Figure 33.
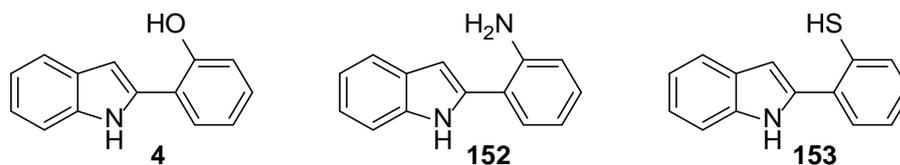
*Figure 33 - Compound **4** alongside its anilinic and thiophenolic analogues*

Due largely to the wider abundance of anilines compared to thiophenols (and therefore a higher potential for ease of synthesis and purchase of future analogues), it was decided that aniline **152** would be included in this work.

An alternative approach to determining the role of the phenolic group of compound **4** is to add protecting groups. If the phenolic OH group is essential for activity, masking this OH with a group that can't be removed by hydrolysis should reduce the anticancer activity, whereas masking this OH with a group that can be removed by hydrolysis should maintain the same anticancer activity, albeit with a delay, while at the same time providing a prodrug approach.

To investigate this hypothesis, six O-substituted analogues of compound **4** were synthesised, including a benzyl, *tert*-butyldimethylsilyl (also known as TBDMS, or simply TBS), benzoyl, heptanoyl, methanesulfonyl (also known as a mesyl) and acetyl analogue. The structures of compound **4** and these six analogues are shown in Figure 34, and their syntheses are described in sections 8.2.10 to 8.2.15..
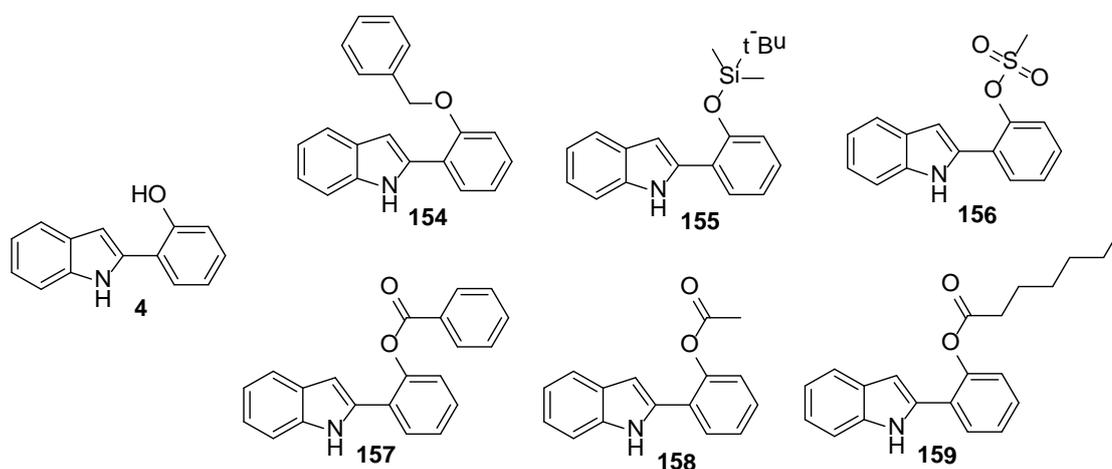
*Figure 34 – The structures of compound **4** alongside some O-substituted derivatives: the benzyl analogue **154**, the TBS analogue **155**, the mesyl analogue **156**, the benzoyl analogue **157**, the acetyl analogue **158** and the heptanoyl analogue **159***

Of the six analogues shown in Figure 34, the three esters (the heptanoyl **159**, the acetyl **158** and the benzoyl **157**) would be expected to undergo enzyme-catalysed hydrolysis, albeit at different rates. Protecting alcohols (or phenols in this case) with acyl groups to form esters is the most

common strategy employed when designing prodrugs,[220] and is employed in drugs as diverse as the chemotherapeutic irinotecan to the anti-inflammatory aspirin (Scheme 44)
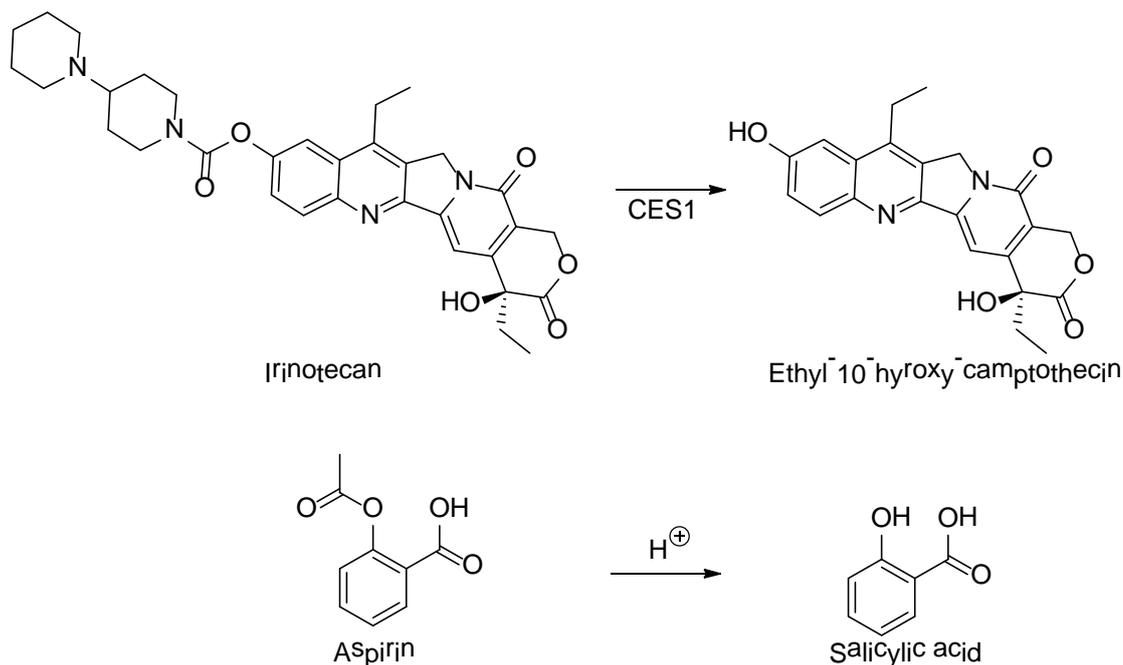
Scheme 44 - The prodrug irinotecan being metabolised by carboxylesterase 1 (CES1) to its active form (top) and aspirin metabolysing to salicylic acid under acidic conditions (bottom)

The other three derivatives shown in Figure 34 (i.e. the benzyl **154**, the TBS **155** and the mesyl **156**) would not be expected to be removed easily, so might be expected to show no activity. It is however always possible that these groups could be cleaved by an enzyme (either hydrolysed or *via* redox mechanisms) to produce the active compound **4**. For example, enzymatic cleavage of Si-O bonds is known,[221,222] although typically these have only been reported for more hydrophilic residues.

Up until now, all the focus has been on compounds that closely resemble the lead structure of compound **4** in shape. Investigating an analogue that is electronically similar to compound **4** but has a different shape may yield information on any potential active site that the drug may be interacting with. An isomer of compound **4** whereby the phenol group is simply at a different position on the indole ring would have a different shape but otherwise very similar electronics.
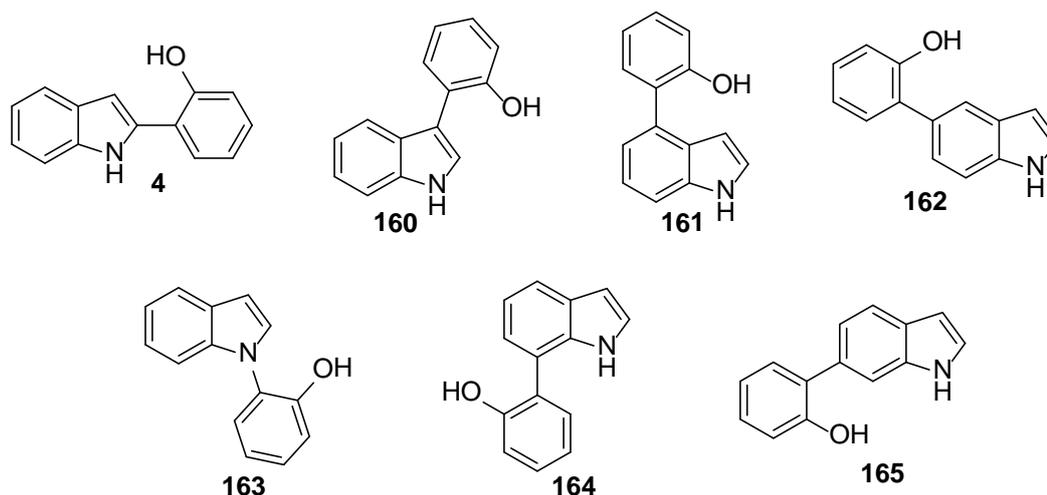
Figure 35 - The seven isomers of 2'-hydroxyphenylindole

The seven isomers of the lead compound **4** are shown in Figure 35. In the interest of keeping analogues as similar as possible, only isomers with the phenol ring as a substituent on the pyrrole ring should be considered. A further constraint should be that the NH group is preserved in case it is involved in some weak hydrogen bonding. This leaves only the 3-substituted isomer, compound **160**, as shown in Figure 36. This compound was therefore synthesised, as described in Section 8.2.16.
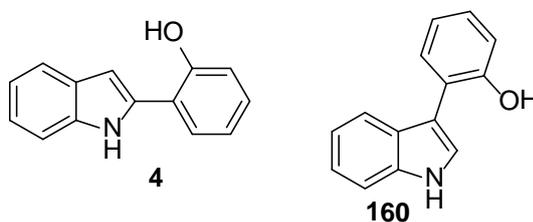
*Figure 36 - Compound **4** (left) and compound **160** (right)*

The analogue shown in Figure 36 differs in a few subtle ways from the lead compound. Primarily, and most importantly is that the basic shape of the molecule is slightly different. Secondly, because the substituent is on an adjacent carbon, the resonance structures of the respective radicals are slightly different. These resonance structures, depicted in Figure 37, show that the radical can be delocalised onto different carbons between the two isomers, however it is unknown what effect this would have on activity.
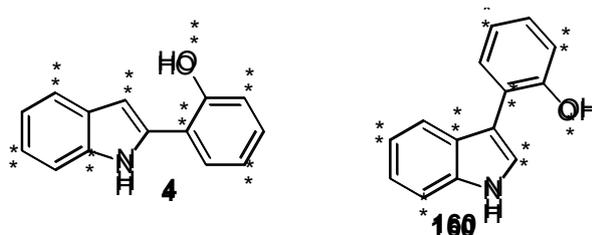
*Figure 37 - The possible places that a phenolic radical could become delocalised to on compound **4** and compound **160***

The final structural feature of compound **4** that necessitates investigation is the indole core. This is perhaps the most important structural change available, as the development that can be done on a compound further along the drug development process is severely limited by the choice of drug core. A series of analogues (which were by no means exhaustive) were devised that could begin to explore whether the anticancer activity of compound **4** is dependent on the inclusion of an indole core, the structures of which are shown in Figure 38.
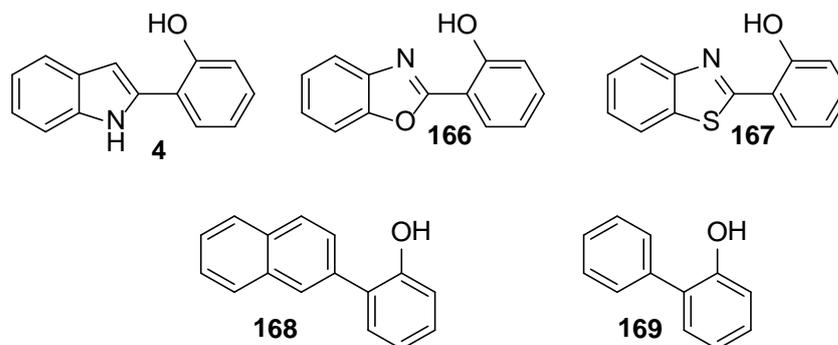
*Figure 38 - Compound **4** alongside analogues that include a non-indole core: a benzoxazole (compound **166**), a benzothiazole (compound **167**), a naphthyl (compound **168**) and a phenyl (compound **169**)*

The benzoxazole and benzothiazole analogues (compounds **166** and **167**) were included due to having a very similar shape, being aromatic, and containing an aromatic nitrogen. These cores differ from the lead compound in that they include two moderately strong hydrogen bond acceptors but no hydrogen bond donors, whereas the indole core contains a weak hydrogen bond donor and a weak hydrogen bond acceptor.

The naphthalene analogue **168** has a slightly different shape to compound **4** and contains no hydrogen bond donor or acceptor, but is isoelectronic. This compound may yield important information about the necessity of the NH group in compound **4**, so was synthesised (as described in Section 8.2.17) for use later in the project.

The phenyl analogue **169** is a stripped back version of the naphthalene analogue, and will hopefully bridge the gap in chemical space between the highly conjugated analogues investigated thus far and the much simpler phenols described earlier.

4.3.1   Conclusion on Analogues of Compound 4

This section has described the design of analogues that can investigate the importance of the structural features of compound **4**. The choices of analogues discussed within this section, along with the structural features under investigation, are brought together in Figure 39.

*Figure 39 - A summation of the structural features and analogues of compound **4***

4.4   Miscellaneous Synthetic Challenges and Methods

In the course of this work, a number of synthetic and analytical challenges presented themselves which in some cases limited the intended scope of the work being undertaken. In some cases, existing synthetic methods were incapable of producing the required compounds, therefore certain structural features and derivatisations could not be realised despite much experimentation changing standard reaction parameters. In other cases, an unexpected product was formed, resulting in the inclusion of an unplanned analogue in the work. In one particular case, a rearrangement which had gone unreported for over 40 years was responsible for the formation of a compound that proved very difficult to characterise.

4.4.1 Structural features that were unable to be investigated

For both the lead compounds **4** and **5**, there were analogues and structural features which could not be investigated. Some analogues could not be produced due to synthetic limitations, whereas some analogues could not be tested due to being too unstable.

One such analogue of I3C that was intended to be investigated was an *N*-benzyl analogue (compound **170**), which was intended to be tested alongside compound **133,** the *N*-tosyl analogue (the structures of which are shown in Figure 40). This compound was found to be too unstable for use as a drug, as it decomposed fully over a period of just a few days. This compound would therefore give no useful information with regards to an SAR, as any potential pharmacological effects may be due to the effects of a degradation product and not the synthesised compound.
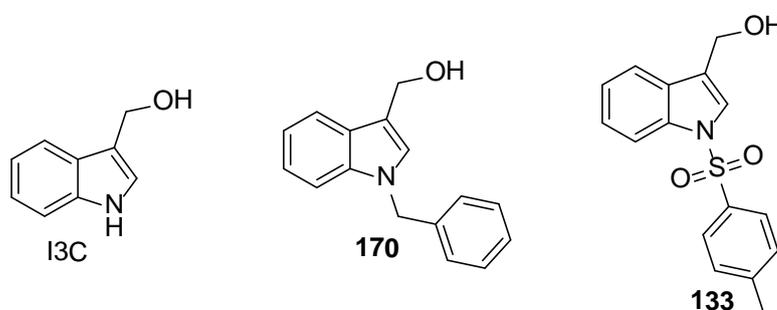
Figure 40 - The structures of I3C, its N-*benzyl analogue (**170**) and its* N-*tosyl analogue (**133**)*

It can be confidently asserted that compound **170** itself was successfully synthesised from indole-3-carboxaldehyde (compound **171**), as shown in Scheme 45. The degradation product is, however, more difficult to assess, and due to time constraints was not fully investigated.
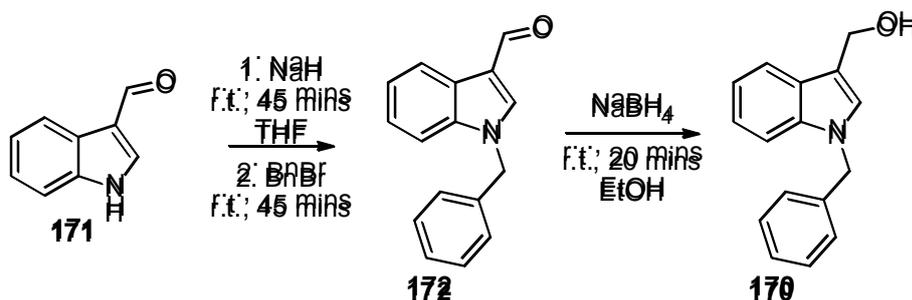
*Scheme 45 - The synthesis of compound **170***

These *N*-substituted I3C analogues shown in Figure 40 might be expected to degrade in much the same way as I3C does into a series of metabolites, as discussed in the introduction, and there is some evidence that this occurs. Whereas I3C degrades into a number of metabolites including 3,3'-diindolylmethane (DIM), the *N*-benzyl analogue **170** might be expected to form the

analogous metabolite **174** (Scheme 46). The characteristic peak of DIM, the bridging CH_2 group, has a ^{1}H NMR shift of 4.26 ppm (singlet) and a ^{13}C NMR shift of 21.35 ppm. Crude NMR spectra of the degradation products of compound 133 shows a ^{1}H NMR shift at 4.29 ppm (singlet) and a ^{13}C NMR shift at 21.17 ppm, suggesting that the analogous degradation product **174** may have been formed.
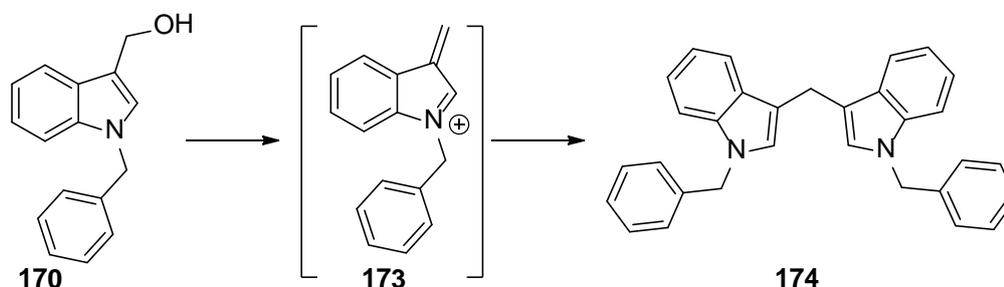
*Scheme 46 - Suspected degradation product of compound **170** into the DIM analogue **174***

It may seem unusual that compound **170** would be significantly less stable than compound **133** or I3C. However, it may be the case that the inductive effect of the benzyl group stabilises the cationic intermediate (shown in Scheme 46 as compound **173**), making the degradation more favoured. The *N*-tosyl group of compound **133** on the other hand would destabilise a nitrogen-based cation, whilst also reducing the propensity for the nitrogen to donate its lone pair of electrons, which would result in a disfavoured degradation pathway.

There were also structural features of compound **4** that could not be investigated due to synthetic limitations. One such feature was to be the rotation around the indole-phenol bond. By rotating about this bond, as shown in Figure 41, the OH group can be located in a different area of space, potentially affecting any drug-protein interactions that may occur. By adding a second OH group at the 6'-position, as in compound **175**, it may be possible to improve this interaction.
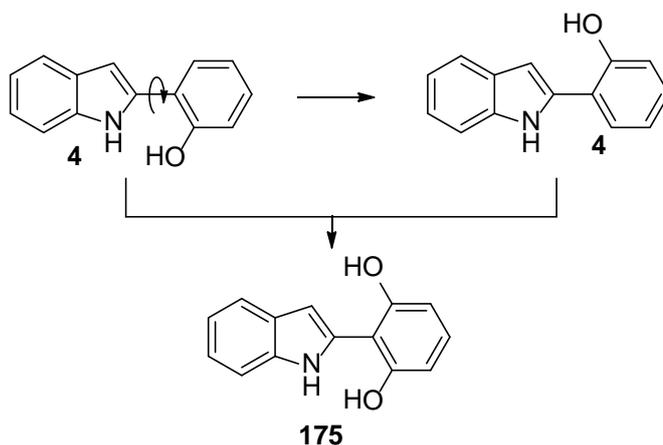
*Figure 41 - An illustration of the rotation around the indole-phenol bond in compound **4**, and the hybrid structure of these two rotamers (compound **175**)*

Many attempts to synthesise the dihydroxylated analogue **175** under a range of Fischer conditions were made. The acid catalyst utilised ranged from polyphosphoric acid[223] (both without and with either ethanol or toluene as a diluant), *p*-toluenesulfonic acid,[224] Eaton's reagent (formed *in situ* from phosphorus pentoxide and methanesulfonic acid)[225,226] and cyanuric chloride (also known as TCT).[227]
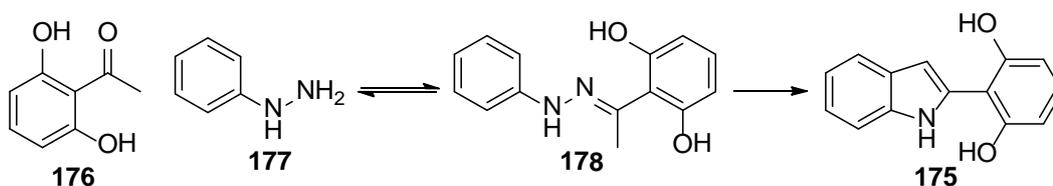
*Scheme 47 - The two-step Fischer synthesis of compound **175***

Under Fischer conditions, the formation of an indole is a two-step reaction which goes *via* a hydrazone (Scheme 47). However, because the second step occurs under similar conditions to the first step (albeit at higher temperatures), this synthesis is often carried out in one pot. It was found that when mild conditions were employed, the intermediate hydrazone, compound **178**, could be isolated without issue. However, when attempting to cyclise the hydrazone using a series of acid catalysts, the dihydoxyacetophenone starting material (compound **176**) seemed to be reformed.

The reversibility of the hydrazone formation is known. The formation of the hydrazone results in the loss of water, and the reformation of the starting materials therefore requires water. In order to prevent this, the cyclisation step was again retried using dried toluene, Dean-Stark apparatus (in order to remove any water), and molecular sieves on a larger scale, but only traces of product were ever detected.

The synthesis was also attempted using compound **179** as a starting material (Scheme 48), which is an *O*-methoxy protected form of compound **176**. These methoxy groups are routinely used to protect OH groups during syntheses, and can be removed relatively easily. Upon failure to make this synthesis work, the Fischer synthesis was abandoned in favour of an alternative approach.
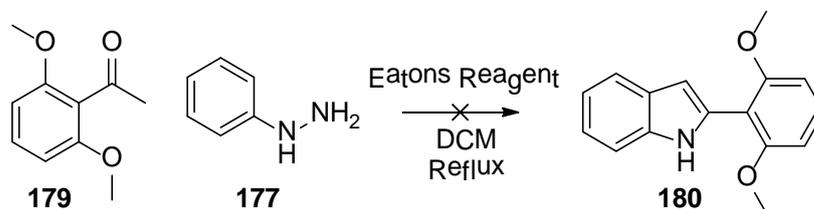
Scheme 48 - Failed Fischer synthesis using dimethoxyacetophenone and phenylhydrazine

The Truce-Smiles rearrangement was attempted as an approach to synthesise compound **175**.[177]   The first step of this reaction, which couples the acetophenone **176** and fluoronitrobenzene (**181**) under basic conditions (Scheme 49), was repeatedly shown not to occur.
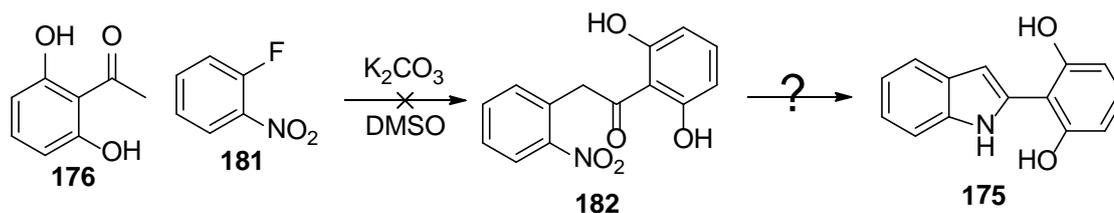
Scheme 49 - The attempted synthesis of compound 175 via *the Truce-Smiles rearrangement*

Although the target product was not formed, mass spectroscopy and a ^{1}H NMR spectrum suggest the formation of compound **183**, as shown in Scheme 50. This is also chemically reasonable when considering the pKa values of an *o*-hydroxyacetophenone and the conjugate acid of potassium carbonate, and potassium carbonate should be able do deprotonate these phenolic groups. At this point it was decided in the interest of pursuing other targets, this target compound would no longer be pursued.
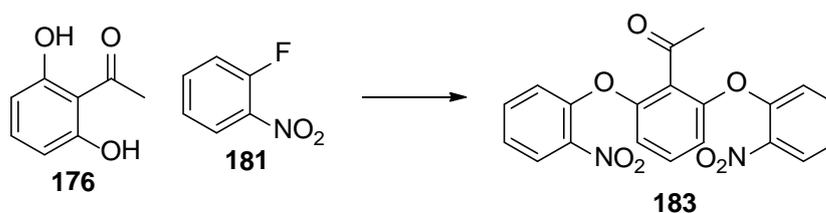
*Scheme 50 - Suspected product formed from reaction between 2,6-dihydroxyacetophenone (**176**) and*
*o-fluoronitrobenzene (**181**)*

Another area of chemical space that was intended to be explored was the activity of a hydroquinone analogue of compound **4**, such as compound **184** shown in Scheme 51. Hydroquinones were of interest because they are known to be commonly involved in redox cycling, whereby they can readily form reactive oxygen species (ROS).[228,229]
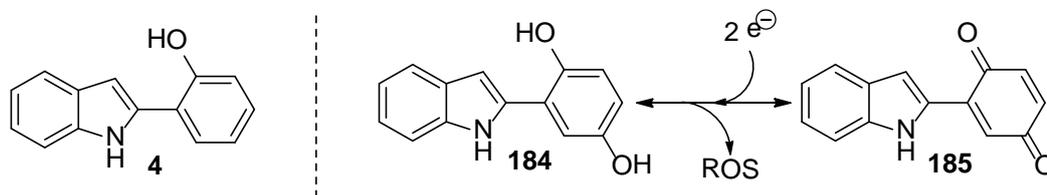
*Scheme 51 - Compound **4** (left) and the conversion between compound **184** and its benzoquinone **185** (right)*

After the challenges encountered with compound **175**, less effort was put into compound **184**. Synthesis of this analogue was attempted a number of times under Fischer conditions using polyphosphoric acid and Eaton's reagent. However, after these attempts were unsuccessful, it was decided not to pursue this target any longer in the interest of pursuing other analogues.

As will be discussed in Chapter 6, in the course of the project the 3-methyl analogue of compound **4**, referred to throughout this thesis as compound **186**, became a compound of interest. This compound was synthesised, as described in Section 8.2.18, but it was observed that on standing over a period of a few days, compound **186** (a thick yellow oil) had crystallised into a pale yellow solid, and on doing so had changed by ^{1}H NMR.
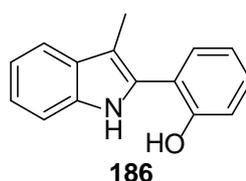
Figure 42 - The structure of the desired compound 186

The resultant product was purified by recrystallization (ethyl acetate and petroleum ether) to form large amber crystals. A series of NMR experiments, including a ^{1}H, ^{13}C, DEPT135, ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC, as well as a low-resolution mass spectrometry experiment were carried out. A summation of this data is shown in Table 2.

STRUCTURAL FEATURE	EVIDENCE	IMPLIED STRUCTURE
8 AROMATIC CH PROTONS	^{1}H NMR DEPT135	
5 AROMATIC QUATERNARY CARBONS	^{13}C NMR DEPT 135	Fewer aromatic carbons, so some aromaticity has been lost
TWO HYDROGEN-BONDING HYDROGENS (Δ = 13.59, 2.78)	^{1}H NMR	Phenol has been retained? Due to loss of aromaticity, NH may be hydrogen bonding? New hydrogen bonding group?
CH$_3$ GROUP WITH NO ADJACENT HYDROGENS	^{1}H NMR (3H, singlet)	3-Methyl group retained?
AN M/Z OF 239	Mass spectrometry	Unlikely to have dimerised – Nitrogen rule Gain of oxygen?
VERY DESHIELDED PROTON (Δ = 13.59)	^{1}H NMR	One of the hydrogen-bonding groups involved in strong intramolecular hydrogen bonding
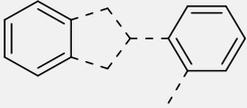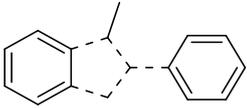
*Table 2 - Spectroscopic evidence for structural features of the degradation product of compound **186***

Based on the evidence in Table 2, a number of possible structures (some of which are shown in Figure 43) were derived, but none of these structures satisfied all of the evidence. Many of the compounds were still based on 2-phenylindoles, which don't match the aromaticity suggested by NMR. Finding a way for a proton with a ^{1}H shift of 13.59 was also a challenge.
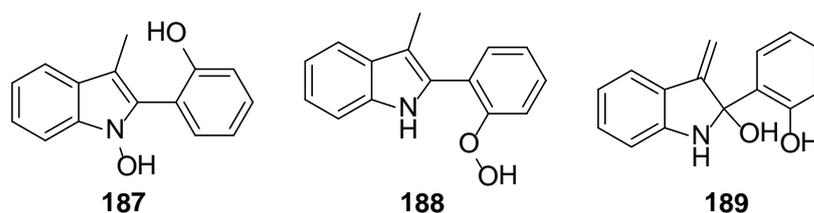
Figure 43 - Possible structures derived from evidence shown in Table 2

An alternative approach was to try and determine how compound **186** could degrade. As discussed in Section 6.3.2, this analogue was designed in order to investigate the effect of stabilising the phenolic radical, as shown in Figure 44.
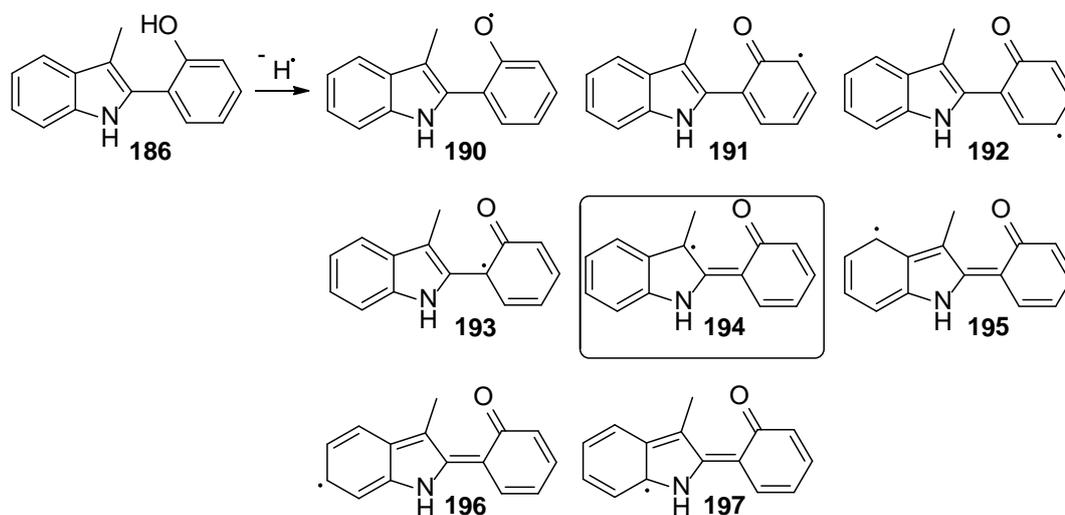
*Figure 44 - The delocalisation of the phenolic radical of compound **186**. The resonance form stabilised by the 3-methyl group is identified with a box.*

Because the 3-methyl group would be expected to stabilise a radical that could be delocalised to the 3-position of the indole ring, it stands to reason that the radical would formally spend more time on this lower energy position. It might therefore be expected that another radical might react at this positon. A number of possible radicals were considered, and their respective radical termination products are shown in Scheme 52.
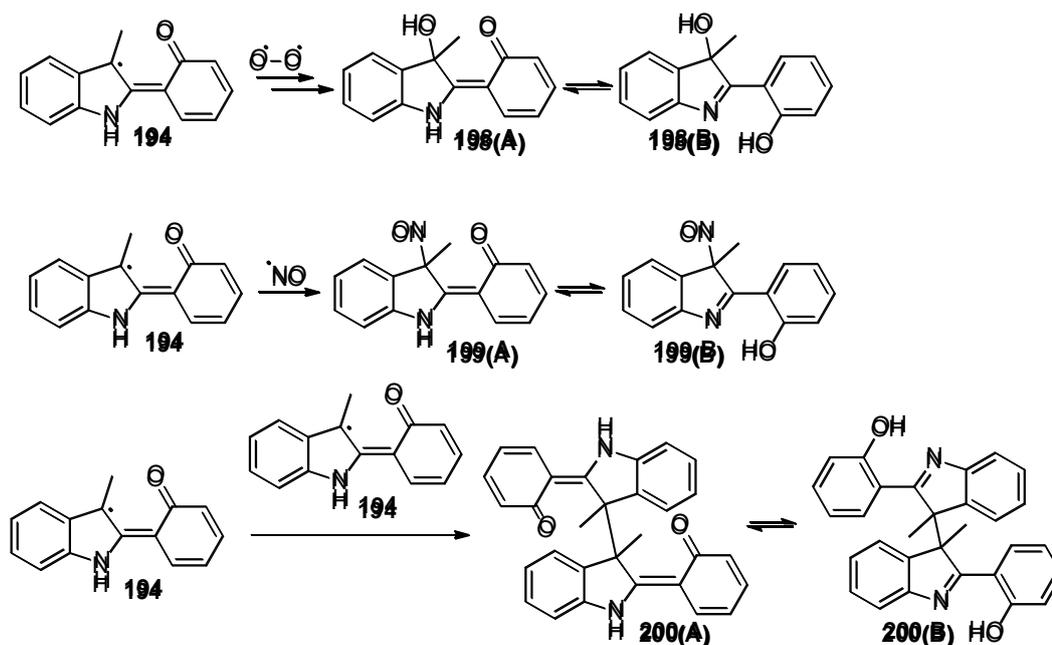
Scheme 52 - Potential products of a number of radical termination products

Of the radical termination products shown in Scheme 52, compound **198** fits all of the spectroscopic evidence shown in Table 2. It has a mass of 239, has reduced aromaticity, has two exchangeable protons, and the phenolic proton which is hydrogen-bonded to the newly formed imine could have a very high ^{1}H NMR shift.

In order for this assignment to be plausible, formation of the product needs to be plausible. Aerial oxidations of 2,3-disubstituted indoles are also known,[230] and a proposed mechanism of formation is shown in Scheme 53.
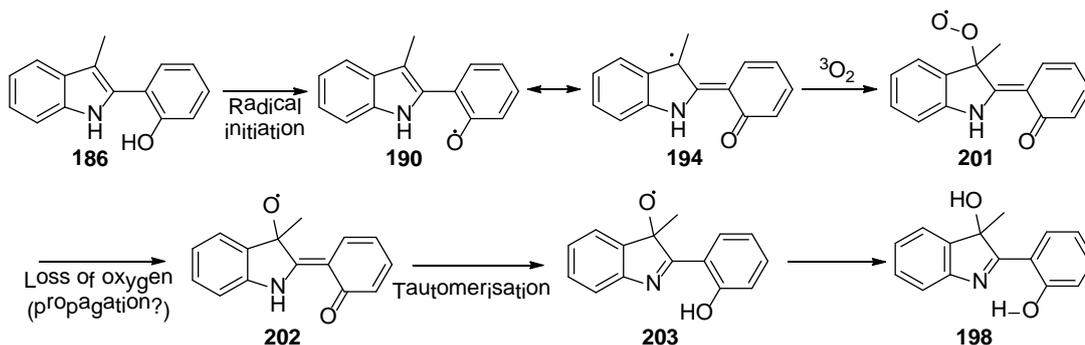
*Scheme 53 - Proposed mechanism for aerial oxidation of compound **186** to compound **198***

In order to confirm this assignment, the crystal structure of the product was obtained by the EPSRC UK National Crystallography Service at the University of Southampton.[231] An image of the crystal structure is shown in Figure 45. This assignment is in agreement with a previously reported formation of this product.[230]
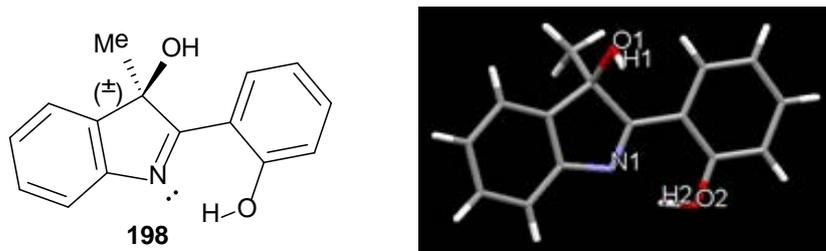
Figure 45 - The skeletal structure of compound 198 (left) alongside its XRD structure (right)

The most notable structural feature in this crystal structure is the bridging hydrogen (labelled "H2" in Figure 45). The N1···H2 interaction length is 1.63 Å, hinting at a potential resonance form (Scheme 54). This would lead to a highly deshielded proton, which would result in a very high ^{1}H NMR shift, therefore explaining the unusually downfield peak at $\delta = 13.5$.
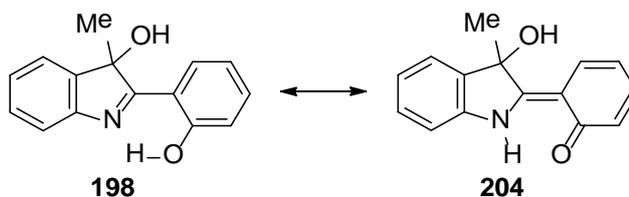
*Scheme 54 - Potential resonance form of compound **198***

Although compound **198** was an unexpected product, the anticancer activity of this compound was also established, and is discussed in Chapter 7. Further information on its synthetic pathway is given in Section 8.2.18.

4.4.2 A novel approach to 3,3'-diindolylmethanes

As has been established in the main body of work, 3,3'-diindolylmethanes are promising anticancer compounds. 3,3'-Diindolylmethanes, such as the simplest example of DIM, shown in Scheme 55, are known to be formed from indole-3-carbinols both under laboratory[232] and physiological[116] conditions.
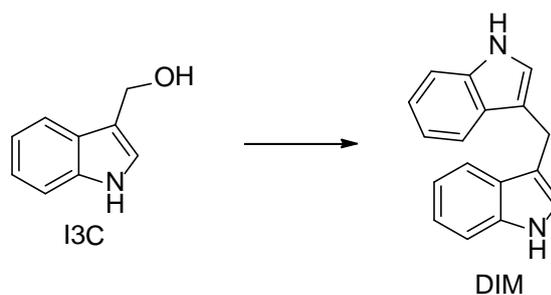
Scheme 55 - Degradation of I3C to DIM

Although it is possible to synthesise DIM analogues from their respective indole-3-carbinols, the instability of these reagents makes this synthesis disfavoured. The most common synthesis for DIM analogues is to synthesise them from their 3-unsubstituted indoles and formaldehyde, however this limits the products to symmetrical 3,3'-diindolylmethanes, and requires the use of carcinogenic formaldehyde. A representative reaction for the formation of DIM from indole is shown in Scheme 56.
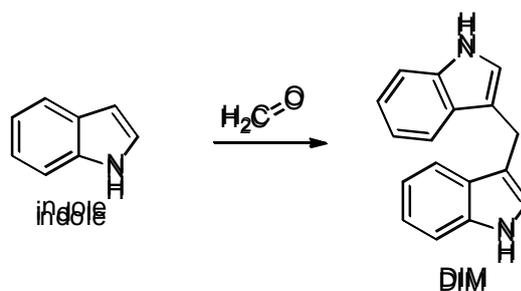
Scheme 56 - Synthesis of DIM from indole and formaldehyde

In the course of this work, a novel synthesis of 3,3'-diindolylmethanes was discovered, which uses safer reagents, and should allow for unsymmetrical 3,3'-diindolylmethanes to be synthesised. Although only compound **9** was produced using this method (Scheme 57), this proof of concept can be developed in the future.
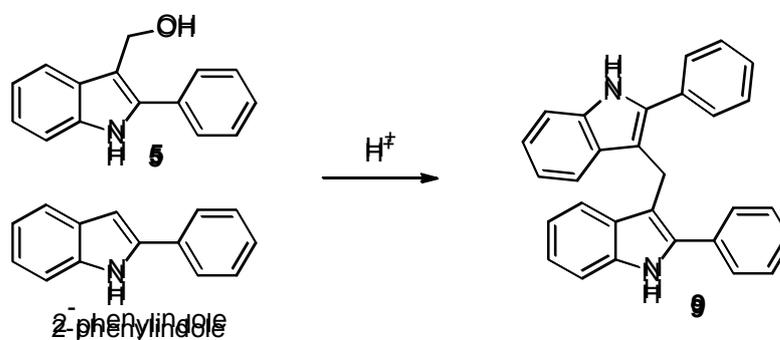
Scheme 57 - Novel synthesis of a DIM analogue

Briefly, 2-phenylindole was dissolved in ethanol and acetic acid. Compound **5**, dissolved in ethanol, was added to the reaction gradually, and the reaction was stirred for 1 minute until complete by TLC. The solvent was then removed *in vacuo*, and the product was isolated and purified by preparative liquid chromatography to yield compound **9** (62 mg, 58% yield). A full characterisation and synthetic method for this product is given in Section 8.2.3.

The potential of this reaction as a method for synthesising unsymmetrical DIM analogues is clear; by employing an excess of the 3-unsubstituted indole and/or introducing the 3-substituted indole gradually, the 3-substituted indole should tend to react with the 3-unsubstituted indole.

4.4.3   A microwave mediated Fischer indole synthesis

In the process of finding efficient syntheses for indoles, a publication by Creencia *et al.*, was found which allegedly reports on a one-pot, one-step microwave assisted Fischer indole synthesis.[233] This paper however fails to report the temperatures at which the reactions were

carried out, and the results were therefore unrepeatable. Efforts therefore began to develop generic conditions for a microwave assisted Fischer indole synthesis.

As it was a compound of interest at the time, the methodological development of these conditions was carried out on the synthesis of compound **186**, the reaction for which is shown in Scheme 58.
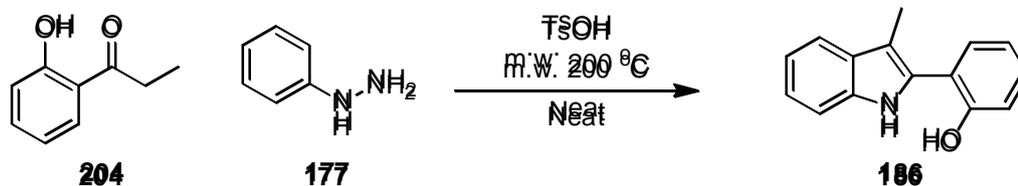
*Scheme 58 - Microwave-mediated synthesis of compound **186***

During a preliminary reaction, tosic acid was trialled as the acid catalyst. The reaction was carried out at 200 °C for 5 minutes, and product was detected by NMR. Therefore, tosic acid, which was also the acid catalyst reported to have been used by Creencia *et al*., was utilised as the acid catalyst for this set of condition screening.

 The yield of the reaction would be determined by GC-MS, relative to the external standard 2-phenylindole. The retention times of the two reagents, the product and the standard were determined, and their relative peak areas were calculated.

Compound	Retention time (R$_t$ / minutes)	Relative intensity
	6.31	0.50 ± 0.076
	6.85	0.80 ± 0.12
	12.58	0.87 ± 0.13
	11.71	1
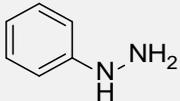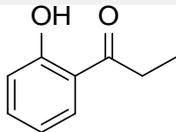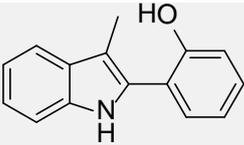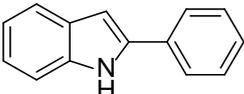
Table 3 - Determination of retention time (R$_t$ / minutes) and peak intensity relative to 2-phenylindole.

Values are an average of five runs

As can be seen in Table 3, 2-phenylindole makes a suitable internal standard for determining the concentration (and therefore yield) of the product using GC-MS.

The next step would be to determine an appropriate temperature and time for the reaction to run for. Once the reaction had reached the specified temperature, reactions were run at both 160 °C and 200 °C for 1, 2, 5, 10 and 20 minutes (with reactions run at 200 °C also being run for 30, 40 and 50 minutes). The experimental procedure for these reactions was as follows:

Phenylhydrazine (compound **177**, 108 mg, 1 mmol), 2-hydroxypropiophenone (compound **204**, 225 mg, 1.5 mmol) and tosic acid (19 mg, 0.1 mmol) were added to a microwave reaction vessel and were heated under microwave radiation at the specified temperature for the specified period of time. At the end of the reaction time, the crude reaction was cooled to room temperature, and one equivalent of 2-phenylindole was added as a standard. The reaction was dissolved in dichloromethane (6 mL) and the tosic acid was filtered off through cotton wool. The reaction was then diluted in dichloromethane (54 mL), and the relative quantities of all components was determined *via* GC-MS.

T = 160 ºC	
Reaction time (minutes)	Yield (%)
1	3
2	5
5	11
10	17
20	31

T = 200 ºC	
Reaction time (minutes)	Yield (%)
1	49
2	65
5	73
10	78
20	86
30	89
40	87
50	93

Table 4 - Table of yields for the reaction shown in Scheme 58 at two temperatures and a range of reaction times

The yield values for this methodology work are shown in Table 4, with the graphs for this data shown in Graph 3. Further information on this methodological development is given in Appendix 1. This data shows that the higher temperature reaches a good yield at a significantly quicker rate than the lower temperature without any apparent degradation of the product, as shown by the plateau in yield. As there appears to be very little benefit in heating beyond 20 minutes at 200 ºC, this reaction time and temperature was used for all subsequent microwave-mediated Fischer indole syntheses.
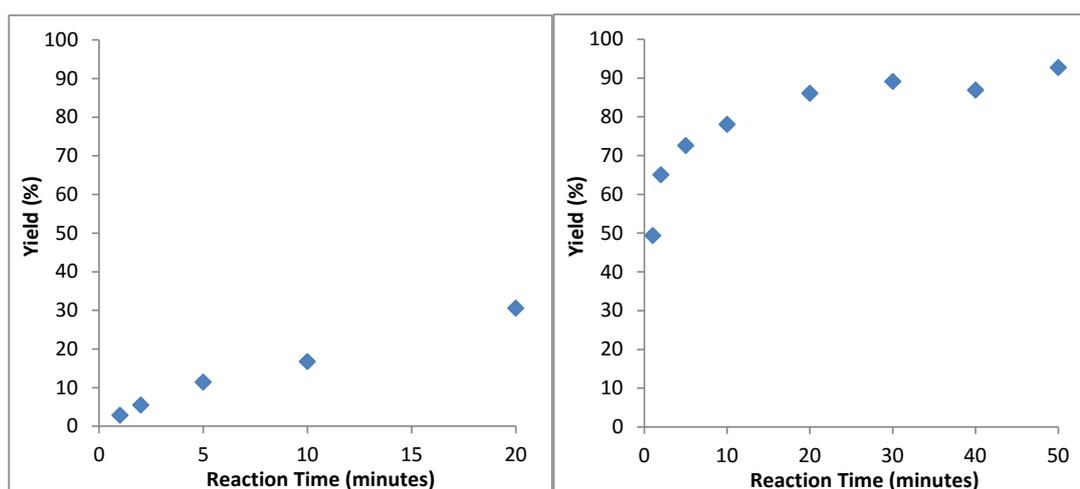
Graph 3 - Product yield against time for the formation of compound 186 (Scheme 58) at 160 ºC (left) and 200 ºC (right)

This chapter has described the justification and methods for synthesising the compounds of interest mentioned thus far, which were all designed using chemical intuition. The next chapter will look at ways of computationally exploring a broader area of chemical space, and how that can be utilised as part of this project.

CHAPTER 5 – SIMILARITY SEARCHING

5.1   Introduction

The previous chapter describes the process of intuitively designing compounds to investigate the significance of certain structural features, but this intuitive process will inherently be limited to areas of chemical space that the medicinal chemist carrying out the process is familiar with. By using this intuitive approach, the likelihood of finding truly novel active analogues is drastically reduced.

Carrying out a computational similarity search against a database expands the area of chemical space being explored from what is familiar to the single medicinal chemist to that which is documented in a curated database. Of course, the choice of database is an important one, and can strongly influence the results of a similarity search. For example, similarity searches on databases focused on compounds with known biological activity are more likely to return active compounds but less likely to return novel compounds.

This section details a similarity search that was carried out in order to find otherwise unconsidered structural analogues of compound **4**. A similarity search was not carried out for I3C or compound **5** due to the majority of the analogues of these compounds described in Chapter 4 being analogues of a prodrug, and structural analogues of these compounds would be unlikely to form the active form of the drug in the same way that I3C and compound **5** do.

5.2   Methodological Choices

In order to investigate a broader area of chemical space, it was decided that a similarity search would be carried out to find structurally similar analogues of the lead compound **4**. Therefore, decisions need to be made on what database of compounds to interrogate, and what software to use to do the interrogation. As this project is a predominantly depth-first approach to discovering active analogues as opposed to a breadth first approach, any methodological choices should reflect this.

5.2.1   Choice of database

As described in the introduction, there are a number of available databases that can be used in a similarity search, and the most commonly used databases are summarised in Table 5.

Database (Number of Entries)	Pros	Cons
DrugBank[193–195] (8206)	All compounds have known biological activity. Many compounds are already FDA-approved.	Very small, therefore less likely to generate any hits.
BindingDB[196–200] (541 thousand)	Most compounds are known to interact with a protein(s).	Relatively small.
ChEMBL[201,202] (1.5 million)	Reasonable number of compounds. Many physiological properties listed.	Large variety of molecular weights included, meaning relatively few compounds will have a good size match with the lead compound.
ZINC[4,5] (22 million)	Well curated, with a subsets of fragment-like drugs. Links to vendors where most compounds can be purchased.	Lots of the vendor links are outdated.
PubChem[204,205] (89 million)	Large area of chemical space covered. Biological data is included in the database.	No subdirectories, meaning database interrogation would be prohibitively arduous.
CAS REGISTRY (114 million)	Most extensive database of known chemicals.	Impractically large. Contains many non-drug like compounds. Not free to use.
GDB-17[136,211] (116 billion)	Conclusive mapping of chemical space of compounds of 17 heavy atoms or fewer.	Limited to compounds of 17 atoms or fewer. Impractically large. Most compounds are synthetically difficult.

Table 5 - Comparison of the most commonly used available chemical databases

On balance, the most suitable database for this project is the ZINC database.[4,5] It is of particular interest due to how the data is curated into subsets, which increase computational efficiency by removing compounds that are very dissimilar before interrogation begins. The full ZINC database can be downloaded as subsets such as shards (molecular weight $< 190\,\mathrm{g\,mol^{-1}}$),

fragment-like (molecular weight $\leq$ 250, logP $\leq$ 3.5, number of rotatable bonds $\leq$ 5),[203] lead-like (molecular weight $\leq$ 350, logP $\leq$ 3.5, number of rotatable bonds $\leq$ 7)[234] and drug-like (molecular weight of 100-500, logP $\leq$ 5, number of rotatable bonds $\leq$ 7, polar surface area $<$ 150, hydrogen bond donors $\leq$ 5 and hydrogen bond acceptors $\leq$ 10).[235] The database can be further cut down by filtering based on commercial availability. In this work, the subset of data that has been opted for is the fragment-like subset of compounds which are advertised as currently in stock, a subset that includes 700,000 entries. The fragment-like subset was chosen because the filtering criteria was most appropriate for the lead compounds being used. The choice to only include commercially available compounds that were advertised as in stock was to hopefully increase throughput of compounds by being able to simply buy in any compound of interest and send it off for testing.

5.2.2 Similarity searching software

The aim of this subset of the project is to search an appropriate database of over 700,000 compounds for structures that are similar to that of the lead compound **4**. This therefore requires a software that can generate a similarity index between a lead compound and a large database.

There are many softwares available for similarity searching,[236] but one such software that fills these requirements and has been successfully utilised within the group in the past is ShaEP.[237] ShaEP, standing for **Sha**pe and **E**lectrostatic **P**otential, is a software which calculates the shape and electrostatic potential similarity between two rigid bodies (i.e. between two 3D structures which can't be conformationally adjusted). ShaEP works by overlaying two molecules in such a way that the volume overlap of the two molecules are maximised, before generating a structural similarity score and an electrostatic similarity score and outputting an overall similarity score as an average of these two values. It was designed and built with the purpose of handling large datasets, and each comparison is typically calculated on a subsecond timescale. Because ShaEP employs rigid body molecules (i.e. the software does not manipulate the geometries of molecules in order to enhance similarities), multiple conformations of each compound are accepted, and the similarity score reported is the highest value achieved between each pair of conformations.

5.2.3 Preliminary work

Before undertaking such a large area of work, some preliminary work on smaller datasets was performed in order to gain an appreciation of the similarity values outputted by ShaEP. The first

step to this preliminary work was to determine the similarity scores between three compounds that are clear to a chemist as highly similar.

Using the freeware Balloon,[238] five conformers of the lead compound **4** and two of its very close analogues were produced, as shown in Figure 46. These compounds are clear analogues in the eyes of a medicinal chemist, as phenols and anilines share similar chemistry,[239,240] and the indole and naphthalene scaffold are both isoelectronic. ShaEP was then used to determine their similarity scores, and the values for their shape similarity and electrostatic potential similarity along with an average of the two values (i.e. the overall similarity score) is displayed in Figure 46.
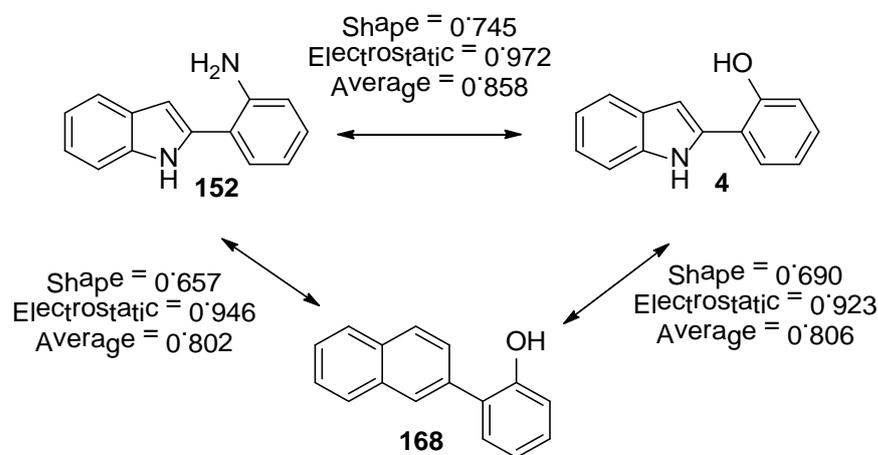
Figure 46 - Similarity between three known analogues, as determined by ShaEP

As might be expected, compound **4** and its aniline analogue compound **152** are considered to be the most similar with regards shape and electrostatic potential. The naphthalene analogue **168** is considered to have a closer shape similarity to compound **4**, but a closer electronic similarity to compound **152**. Based on these results, what a medicinal chemist might call "very similar", ShaEP would consider to have a similarity of around 0.8.

The next step was to determine the necessity of generating conformers. ShaEP, which only deals with rigid bodies, tolerates multiple input conformations for each compound to account for the inherent flexibility of many molecules. However, by including structural conformers, this leads to more pairs of compounds to calculate similarity indices for, therefore increasing computational cost. It would therefore be important to determine the effect of including structural conformers.

Arbitrarily, the first 686 compounds from the "fragments now" subset of the ZINC database were extracted as a trial database, and up to three conformers of each compound in this trial database

were generated using Balloon. Of the 686 compounds in this trial database, 1210 conformers were generated (at an average of 1.76 conformers per compound). These two databases (one including 686 compounds and one including the 686 compounds its 524 conformers) were interrogated using the five conformers of compound **4** generated previously. The results of these searches are shown in Table 6.

Database		Similarity score		
		0.6	0.7	0.8
686 compounds (no conformers)	Interrogation time	90 s	98 s	97 s
	Hits	50	0	0
686 compounds plus conformers	Interrogation time	187 s	187 s	188 s
	Hits	117	3	0

Table 6 - Comparison of interrogating a database without and with generated conformers

Perhaps unsurprisingly, by almost doubling the number of conformers to calculate similarity scores for, the computational time almost doubles too. There is, however, a large increase in the number of hits found. Because compounds with a similarity score of over 0.8 will be of particular interest (as discussed above), and because time is not a limiting factor during this section of work, the increase in chances of finding hits of a higher similarity score outweighs the inconsequential increase in computational time.

5.2.4 Method

Based on the findings from the preliminary work, the work proper could begin. As was decided based on the preliminary work, the whole 700,000 compounds of the "fragments now" subset of the ZINC database was used, and had up to three conformations per compound generated using Balloon, taking around a month of computational time on a standard desktop computer. This conformer-containing database was then interrogated using the five previously generated conformers of **4**, which again took around a month of computational time. A table of the most similar hits found through this method is given in Appendix 2.

5.3 Results

Following the similarity screening of the ZINC "fragments now" database, the 105 compounds with the highest overall similarity value were taken forward as being considered "similar" by ShaEP, with the lowest similarity score being carried forward being 0.77. 105 compounds of

interest is within the realms of what can be analysed by eye, and upon inspection it became apparent that these compounds could be grouped based on their similarity to the lead compound and its presumed mechanism of activity being *via* a phenolic radical.

The first and largest group comprised of compounds that had no way of generating a phenolic (or analogous) radical, for example, compounds that had no phenol or aniline ring. Some examples of these compounds are shown in Figure 47, and includes compounds with no group that can lose a hydrogen radical (such as compounds **205** and **208**), unconjugated alcohols (such as compound **5**) and protected alcohols and alcohol analogues (such as compounds **206** and **207**), as well as some other non-phenolic compounds. This class of compounds contains 59 entries, and as these compound don't fit with the current evidence on how the lead compound acts, they were not included in the work beyond this point. By excluding this class of compounds, this leaves 46 compound of interest.
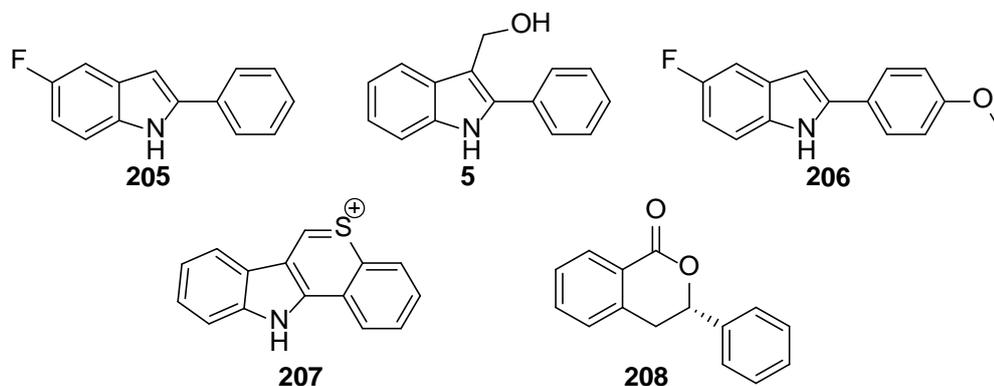
Figure 47 - Examples of compounds considered to have no functional groups capable of forming a phenolic (or analogous) radical

The next class of compounds, of which there are 15 examples, are isomers or pseudoisomers of compound **4**. This class of compounds, some examples of which are shown in Figure 48, include an aromatic alcohol or amine located on a 2-phenylindole core (such as compounds **152**, **209** and **210**), as well as some compounds which might be expected to hydrolyse to produce these compounds. Unsurprisingly, compounds in this class typically had the highest similarity score, with the potential prodrugs such as the amide analogue shown in Figure 48 having a slightly lower similarity score.
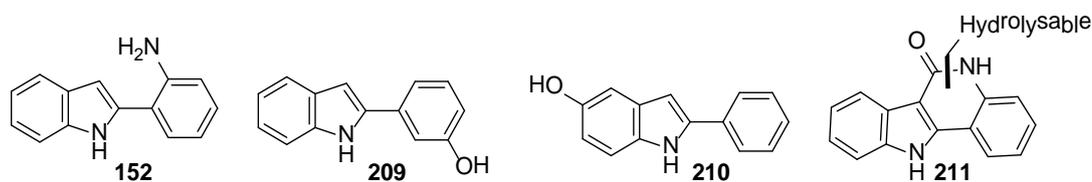
*Figure 48 - Examples of isomers of the lead compound **4** and its aniline analogue*

This class of compounds are the most similar to the lead compound **4**, and as a result had largely already been considered as potential analogues prior to beginning this section of work. Because the purpose of this section of work was to encounter analogues which would otherwise not have been considered, focus was given to those compounds that appear to be more dissimilar. As a result, compounds in this class will also not be considered further.

The third class of compounds are compounds which may be able to form a conjugated radical, but have a core other than 2-phenylindole. A total of 26 compounds were assigned to this class, with some examples shown in Figure 49. This class of compounds also includes compounds such as the pyridone-containing compound **213**, which has the resonance form of compound **216** whereby a proton lies on the oxygen, therefore allowing for a phenolic-like radical to be formed. As previously mentioned, this area of work was intended to suggest analogues that would have otherwise not have been considered, so this class of compounds are of particular interest due to their potential for scaffold-hopping.[190,241]
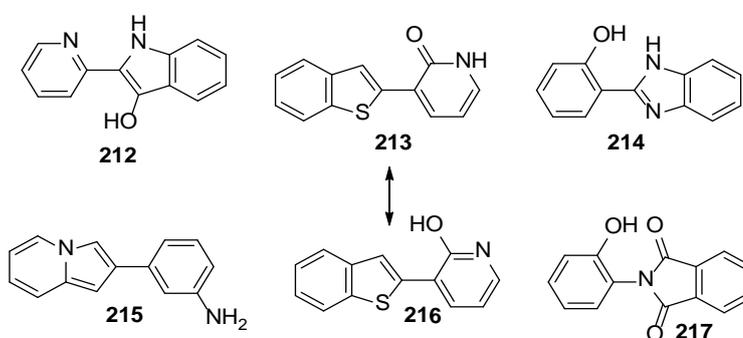
*Figure 49 - Examples of analogues of compound **4** that have a different core*

Within this class, a total of 14 different cores (not including the 2-phenylindole core) were identified. Of the 14 different cores found, all of which are displayed in Figure 50, there is one core that occurs far more often than any other.

11 Occurrences
Benzimidazole

3 Occurrences
2-Phenylbenzofuran

2 Occurrences
2-(3-(2-Pyridinyl))benzofuran

1 Occurrence
2-Phenylisoindoline

1 Occurrence
2-(2-Pyridyl)indole

1 Occurrence
2-Phenyl-4-azaindole

1 Occurrence
2-(2-Pyridyl)benzothiophene

1 Occurrence
2-Phenylindolizine

1 Occurrence
2-Phenyltetrahydroisoquinoline

1 Occurrence
2-(3-(2-Pyridinyl))naphthalene

1 Occurrence
2-(6-(2-Pyridinyl))naphthalene

1 Occurrence
2-Phenylphthalimide

1 Occurrence
2-(6-(2-Pyridinyl))benzothiophene

1 Occurrence
3-Phenyl-2(1H)-quinolinone

Figure 50 - The 14 different cores identified during the similarity search

The benzimidazole core appears 11 times in the best 105 hits, accounting for over a third of the examples in this class of compounds. Based on this data, it is difficult to say whether this apparent abundance of benzimidazoles is because they are particularly similar to the 2-phenylindole lead structure, or if this scaffold is particularly highly represented within the ZINC database (or a combination of both factors). What is known is that the benzimidazole scaffold has previously been described as a privileged structure,[242,243] suggesting that it would probably be well represented within databases of known compounds. Regardless, it seems only sensible that a benzimidazole analogue of compound **4** should be included in future work. Discussion as to which benzimidazole compound to include will be discussed later.

The final class of compounds are those that don't seem to be close analogues of the lead compound at all in the eye of the medicinal chemist, and appear to be considerably different. Perhaps unsurprisingly, this class of compounds was the smallest class with only five entries, three of which are shown in Figure 51. These compounds are all simply *ortho*-substituted phenols, but it is difficult as a medicinal chemist to otherwise see the structural similarity.
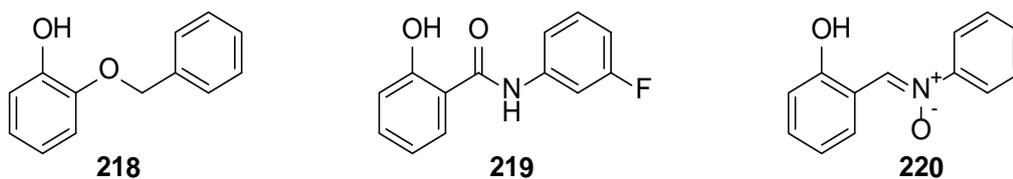
Figure 51 - Examples of those compounds considered to be considerably different to the lead compound, yet may still be capable of forming a phenolic radical

The reason for carrying out this similarity search is, as mentioned earlier, to probe otherwise unconsidered areas of chemical space. Due to constraints both on time and resources, the 46 compounds that make up the "isomers and pseudoisomers", "different core" and "considerably different" classes is still more compounds than can be tested, so the question now turns to choosing the most appropriate compounds to put forward to biological testing. As this area of work is attempting to probe novel areas of chemical space, perhaps the first compounds that should be disregarded are those that are particularly obvious to the medicinal chemists, i.e. the 15 isomers and pseudoisomers of compound **4**. These 15 compounds represent a narrow breadth of chemical space, and to spend the limited time and resources available exploring such a small area would be inefficient. By disregarding these compounds, 31 compounds of interest in the "different core" and "considerably different" classes remain.

When building up a structure-activity relationship, comparisons are made between two or more compounds to see the difference that structural changes have on activity. It is therefore ideal if there is only one difference in structure when comparing two compounds, so that the change in activity can be confidently attributed to a specific structural change. For example, one could not compare an *ortho*-phenol with a *meta*-aniline, as one would not be able to tell if any subsequent change in activity was due to changing the functional group or due to changing the position of the group. With this in mind, and considering the previously discussed justifications for wanting to include a benzimidazole compound in the future work, the benzimidazole that is as otherwise identical to the lead compound should be chosen.
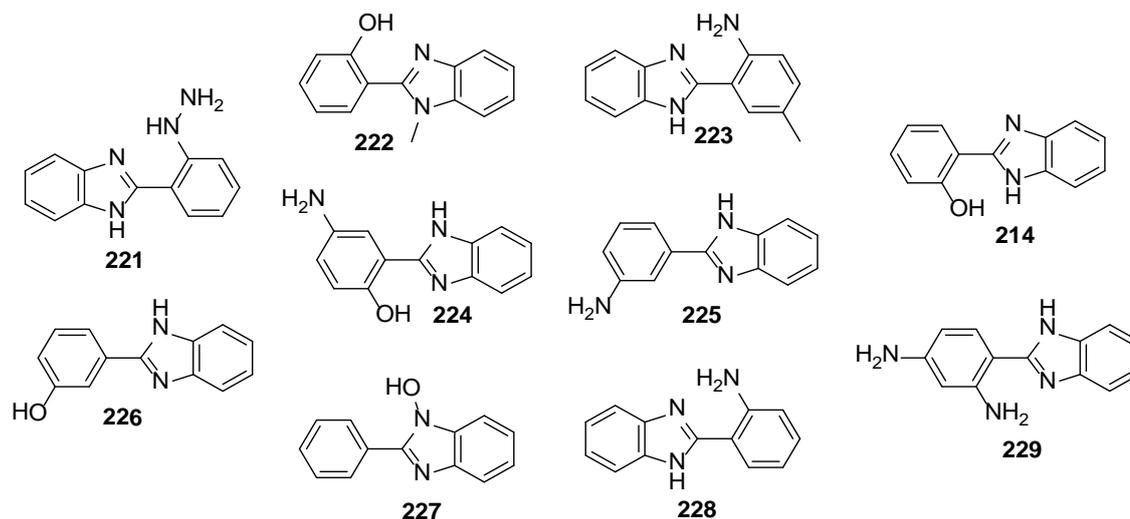
Figure 52 - The 10 benzimidazoles found during the similarity search

Of the 10 benzimidazoles that appeared within the 105 best hits (shown in Figure 52), only compound **214** differs to compound **4** in only one way, the structure of which is shown in Figure 53 alongside the lead compound **4**. All other options can be disregarded based on having NH_2 groups in place of the OH group (such as with compound **228**), moving the position of the OH group about the scaffold (such as with compound **226**), the addition of methyl groups about the scaffold (such as compound **222**), or a combination of these reasons.
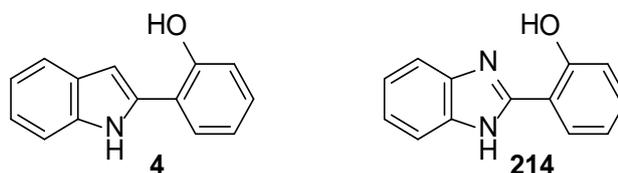
*Figure 53 - The lead compound **4** (left) with its most closely resembling benzimidazole compound **214** (right)*

Aside from compound **214**, there are a number of other compounds found during this similarity search that differed from the lead compound in only the structure of the scaffold, which are shown in Figure 54. Of these are three pyridones (compounds **230**, **231** and **213**, which have a phenolic-like resonance form), a 3-hydroxypyridine (compound **232**) and a phthalimide (compound **217**). Of these five compounds, the compound that is different in only one aspect from the lead compound is compound **217**, so this compound was purchased and would be included in future work.
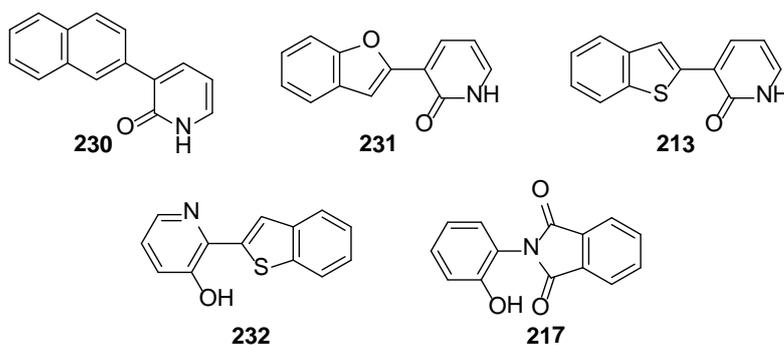
*Figure 54 - Direct analogues of compound **4**: three pyridones (top), a 2-hydroxypyridine (bottom left) and a phthalimide (bottom right)*

As well as testing some different scaffolds, the five "considerably different" compounds shown in Figure 55 may also be of interest. If the largely inflexible lead compound **4** is interacting with an active site, then compounds which can be orientated into a similar shape and have a similar electrostatic potential, may also be able to interact with the same active site in the same way.
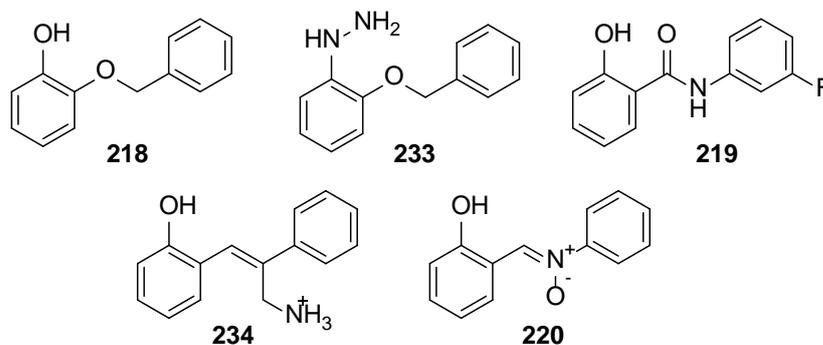
*Figure 55 - Compounds considered similar by ShaEP, but have a considerably different structure to the lead compound **4***

The commercial availability of the five "considerably different" compounds, shown in Figure 55, was investigated. Although all of the compounds in the ZINC "fragments now" database are technically in stock, many of the vendors are based in the USA or China, so lead times in excess of weeks is not uncommon. Along with the extra costs of importing chemicals, this makes many of the compounds in the ZINC database prohibitively expensive. Of the five compounds shown in Figure 55, only compound **218** was available in the UK at a reasonable price. Fortunately, this compound was the simplest example of this class of compounds, so is likely to have been the compound of choice regardless of availability.

5.4   Conclusion

This section of work described the similarity search carried out by the ShaEP software on the ZINC "fragments now" subset of commercially available fragment-like compounds, which had up to three conformations for each entry generated by Balloon. All software used was freeware, and the hardware on which it was all carried out was a standard desktop computer. The methodology described in this section could easily be carried out by any research group with the inclination to carry out their own structural similarity search.
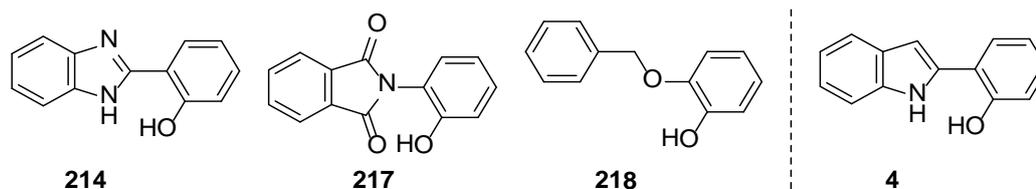
*Figure 56 - The three compounds taken forward to the next section of work, with the lead compound **4** shown on the far right*

The search yielded three compound of interest, shown in Figure 56, which were taken forward to the proceeding sections of work. The first compound, compound **214**, was chosen as a chemically appropriate representative of the many benzimidazoles that ShaEP considered as similar to the lead compound. The second compound, compound **217**, was chosen as another chemically appropriate analogue that contains a previously unconsidered scaffold at its core. The third compound, compound **218**, was chosen to investigate the possibility of a much less rigid molecule being able to interact with an active site in the same way as the structurally dissimilar lead compound (despite being identified as similar by ShaEP).

One constraint of this method was in generating conformers of the ZINC database. Although the time taken to generate said conformations could be reduced with the use of more specialised or advanced hardware, this significant amount of time would be a one-off, and subsequent similarity searches could use the same database of conformations. Even as the ZINC database updates, as it often does, only the newest entries to the database would need to have their conformations generated, meaning that the database of conformations could be routinely kept up to date at a negligible computational cost.

Aside from the one-time cost of generating conformations for the interrogatable database, the major computational cost is interrogating the database. In this work, that computational cost took around a month, which would be too long for many research or industrial groups. This time could be significantly reduced to perhaps a week or less using a number of methods.

One way in which the time taken to interrogate a database could be decreased would be to use fewer conformations of the lead structure. In this work, five conformations of the lead structure were used, which may seem excessive for a structure with only two rotatable bonds, but a decrease in this number of conformations would likely result in lower similarity scores for some compounds. Investigating the connection between conformations of input structures and magnitude of similarity scores may lead to either a more efficient or more thorough method.

Another way of potentially decreasing the time taken to interrogate the database would be to reduce the size of the database through the use of molecular fingerprints. Molecular fingerprints are used in databases where interrogation needs to be done in seconds or less, such as catalogues of chemicals from chemical retailers. By reducing a compound down to a binary file of features that the compound contains, searching a database of millions of molecular fingerprints can occur on a timescale of seconds. Compounds in this work were only considered if they contained a phenolic or anilinic system, so by removing all compounds that don't contain these moieties, the size of the database would decrease considerably. Because molecular fingerprints are stored as very small binary files, generating a database of appropriate compounds could be done rapidly at little computational cost, and could reduce the overall computational time by a large factor.

The choice of database also has a large impact on the interrogation time. The "fragments now" subset of the ZINC database, which includes fragment-like compounds which are advertised as currently in stock, was used in this work. This subset of the ZINC database is ideal if one plans on simply buying chemicals off the shelf for testing regardless of cost of lead time. If one wishes to simply explore chemical space and is willing to synthesise compounds, the larger subset of "clean fragments" would likely be more appropriate. It is over twice the size, and excludes compounds such as thiols and aldehydes, which are notorious for often giving false positives in many assays.

On the other hand, if one does not plan on simply buying compounds off the shelf, then the novelty of the ZINC database is irrelevant. Looking back to Table 5, the PubChem database may be useful for such a project, assuming that the entries in the database could be molecularly fingerprinted, thus reducing interrogation times. The PubChem database is much larger, is also free to use and download, and could be cut down to include compounds of a similar size to those found in the ZINC "fragments now" subset.

CHAPTER 6 – MOLECULAR MODELLING RESULTS

6.1   Introduction

Current evidence suggests that the anticancer activity of compound **4** is related to an increase in levels of reactive oxygen species (ROS). As it is known that under certain conditions phenols can generate ROS non-enzymatically,[244] it is therefore worth investigating if there is a connection between the ability of compound **4** to form a radical and its anticancer activity.

As discussed in the introduction, there are a number of mechanisms by which a phenolic radical can be formed from a phenol: hydrogen atom transfer (HAT), sequential proton loss-electron transfer (SPLET), single electron transfer (SET) or sequential electron transfer-proton transfer (SET-PT), as shown in Scheme 59.
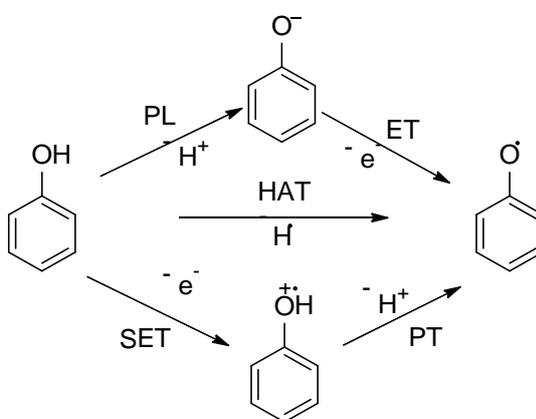
Scheme 59 - Possible ways in which a phenolic compound could form its respective radical species. HAT = Hydrogen Atom Transfer. PL = Proton Loss. ET = Electron Transfer. SET = Single Electron Transfer. PT = Proton Transfer

The propensity of a compound to undergo any of the mechanistic steps shown in Scheme 59 can be determined by measurable values associated with said mechanistic steps; bond dissociation enthalpy (BDE) for HAT, proton affinity (PA) for PL, electron transfer enthalpy (ETE) for ET, adiabatic ionisation potential (AIP) for SET and proton dissociation enthalpy (PDE) for PT. The smaller the enthalpic value for a specific step, the more favoured that pathway becomes.

It is unknown which of the mechanistic pathways mentioned above, if indeed any, are undergone by compound **4** and its analogues. It is therefore essential to determine the enthalpic values for all of the mechanistic steps in order to be able to conclusively detect any correlations.

6.2   Computational Details

6.2.1   Software

In this section of work, all calculations were carried out by software included in the Schrödinger software suite.[245]   The Schrödinger software suite was used because it is widely used and trusted, is easy to learn due to its good documentation, and was available to use in-house. A full description of the theoretical basis of the quantum chemical calculations is given in Chapter 3. A simplified process flowchart demonstrating the molecular modelling calculations performed, descriptions of which are given in further detail below, is shown in Figure 57.
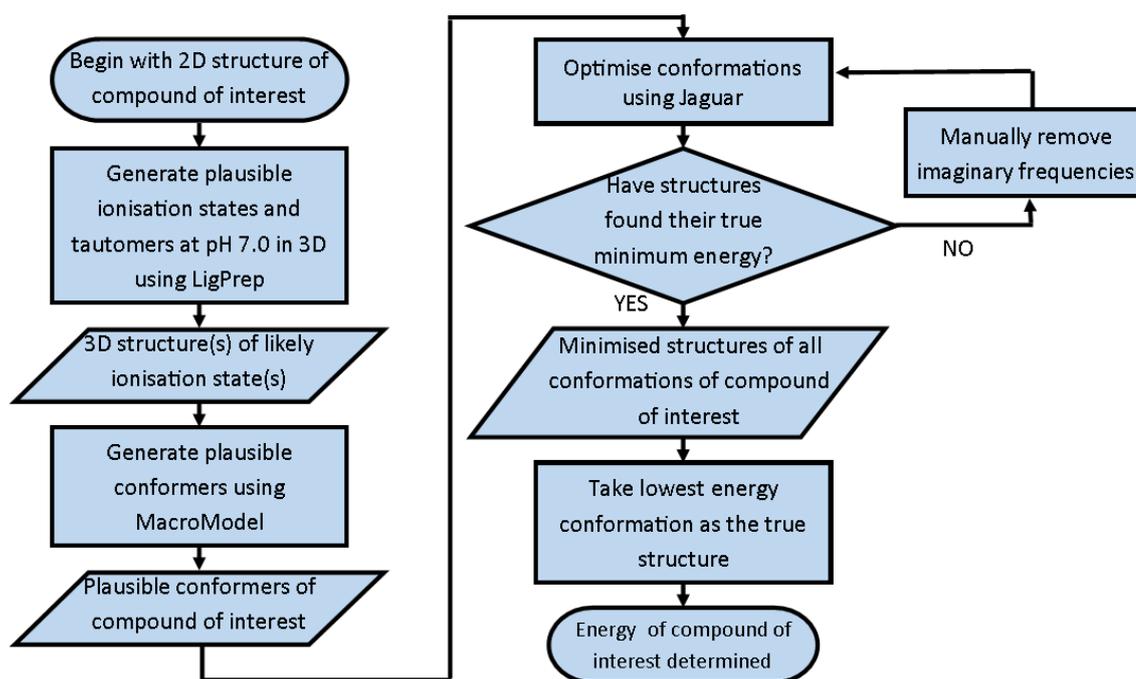
Figure 57 - A process flowchart for the molecular modelling section of this work

6.2.2   Ligand Preparation

Using 2D input structures, 3D structures and plausible ionisation states and tautomers at pH 7.0 (physiological conditions) were generated with LigPrep v3.6. The OPLS3 force field,[246] which has been parameterised for use with drug-like small molecules, is also now well parameterised for halogen bonding and halogen interactions. The Epik program, which does the bulk of the calculations within LigPrep, was used due to its ability to calculate energetic penalties for the states it predicts.

Once the favourable ionisation states and tautomers were determined, MacroModel v11.0[245] conformation searches were used to generate potential conformers. The Monte Carlo Multiple Minima (MCMM) torsional sampling method[245] was performed using the OPLS3 force field.[246]

A Monte Carlo approach is a stochastic method that generates a large number of potential conformers by the random adjustment of select dihedral angles. Energy minimisation of the generated conformations was carried out using up to 500 iterations of the truncated-Newton conjugate-gradient (TNCG) method, set to converge on a gradient threshold of 0.05 units. Duplicate conformations were removed by route mean squared distance (RMSD) comparisons. Conformations within 10 kcal mol^{-1} of the lowest energy conformation were taken forward.

6.2.3 DFT Calculations

Optimisation of the Monte Carlo geometries was carried out using Jaguar v9.0[245,247], a software which can use DFT to precisely determine a more accurate structure and therefore energy of a system. As this is the vital step in the process of our molecular modelling protocol, significant effort went into designing the methodology used through Jaguar, which is documented in the coming section. A maximum of 200 optimisation steps was allowed, symmetry was turned off, the medium density SCF grid was used and self-consistent field spin was restricted. True minima were determined by no imaginary frequencies. Zero-point energies were calculated at 298.15 K.

6.3 Results and Discussion

6.3.1 Benchmarking Studies – Method and Basis Set Selection

When using DFT, the major methodological factor is to determine the most appropriate level of theory to use, i.e. what basis set and functional to employ. The ideal way of doing this would be to compare relevant calculated properties of chemically relevant trial structures calculated with all combinations of basis sets and functionals against empirical data, and then to choose the level of theory that gives the best result. Obviously, there are far too many basis sets and functionals for this to be practical, so a more intuitive approach needs to be considered.

What is typically done in these situations is to look into what levels of theory are common in the literature for calculating similar properties for similar chemical classes. Because the activity of compound **4** is believed to stem from its phenolic ring, and we are particularly interested in how readily compound **4** forms a radical, methods for determining BDE values of phenolics were searched for in the literature. A representative sample is shown in Table 7.

Compounds investigated	Functional	Basis Set	Reference
Sesamol derivatives	B3LYP	6-31G(d,p) – Optimisation 6-311++G(2d,2p) – SPE	[248]
Catechol derivatives	B3LYP	6-31G(d,p)	[249]
Magnolol derivatives	B3LYP	6-31G(d,p) – Optimisation 6-311++G(2d,2p) – SPE	[250]
Chroman-6-ol derivatives	B3LYP	6-31G(d,p) – Optimisation 6-311++G(2d,2p) – SPE	[251]
p- and m-substituted phenols	B3LYP	6-311++G**	[252]
Natural polyphenols	B3P86	6-31+G(d,p)	[253]

Table 7 - Levels of DFT theory used in a representative sample of the literature for calculating BDE values of phenols and phenolic derivatives

As can be seen in Table 7, the B3LYP functional[254–256] is by far the most popular functional for this type of work, so shall be considered.[248–252] The B3P86 functional[256,257] has also been shown to give good data,[253] so shall also be included. With regards to basis sets, split-valence Pople-type basis sets seem popular, in particular the double-zeta basis set 6-31G[258–267] and the slightly larger triple-zeta basis set 6-311G,[265,266,268–273] with varying diffuse and polarisable functions. The smaller basis sets are typically used for the optimisation step due to their computational efficiency, but due to improvements in desktop computing power, computational expense is now less of an issue. Therefore, the larger 6-311G basis set was tested here with varying diffuse and polarisable functions.

The selection of trial compounds was also important. Trial compounds should be chemically similar to those that are being investigated, while also having empirical data that the calculated values can be compared against. In this work, the trial compounds must be chemically similar to compound **4**, and must have published experimental values for radical formation.
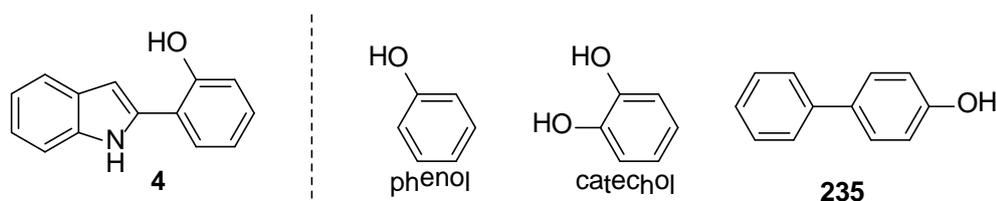
*Figure 58 - Compound **4** (left) with the three trial compounds phenol, catechol and 4-phenylphenol*

The trial compounds used in this work, shown in Figure 58, are phenol (due to being the simplest phenolic), catechol (due to being a well characterised *ortho*-substituted phenol) and 4-phenylphenol (due to the hydroxyl group being conjugated with a second aromatic system, such as in compound **4**). The accepted mechanism of formation of their radicals is hydrogen atom transfer (HAT), for which there are experimental BDE values available for comparisons with our calculated equivalents.[274–276]

Compound	Literature BDE Value / kcal mol^{-1}	B3LYP/ 6-311G*+	B3LYP/ 6-311G**	B3LYP/ 6-311G**+	B3LYP/ 6-311G**++
Phenol	88.7[274]	79.2	83.6	83.7	83.6
Catechol	79.3[275]	69.8	73.0	74.1	74.0
4-Phenylphenol	84.8[276]	76.6	81.1	81.0	80.9

Compound	Literature BDE Value / kcal mol^{-1}	B3P86/ 6-311G*+	B3P86/ 6-311G**	B3P86/ 6-311G**+	B3P86/ 6-311G**++
Phenol	88.7[274]	83.1	87.7	87.7	87.7
Catechol	79.3[275]	73.0	76.4	77.4	77.3
4-Phenylphenol	84.8[276]	80.5	85.2	85.1	85.0

Table 8 - Comparison of empirical BDE values against those calculated at eight different levels of theory for the three trial compounds shown in Figure 58, with B3LYP results shown on top and B3P86 results shown underneath. BDE values are given in kcal mol^{-1}

The data for the calculated BDE values for the three trial compounds at eight levels of theory is shown in Table 8. Surprisingly, using the more popular B3LYP functional significantly underestimates BDE values. All methods tested reproduced the relative trends in BDE values, with catechol having the lowest BDE value and phenol having the highest BDE value. However, the B3P86 functional gives much more accurate absolute values, especially when using more polarisable and diffuse functions. The best levels of theory tested are the B3P86/6-311G**+ and B3P86/6-311G**++ levels of theory, which gave largely equal values. Due to the slight computational benefit of using fewer diffuse functions, the B3P86/6-311G**+ level of theory, this level of theory was chosen to be used in the rest of this work.

6.3.2   Preliminary Work – Singly Substituted Analogues of Compound 4

In order to get familiarised with the method and to get an appreciation for the magnitudes of BDE value differences when considering compound **4** and its analogues, a series of compounds as similar to one another as possible but with differing BDEs would be required. Since activity via a phenolic radical was suspected, a way to affect the BDE value would be to affect the stability of the radical species formed.
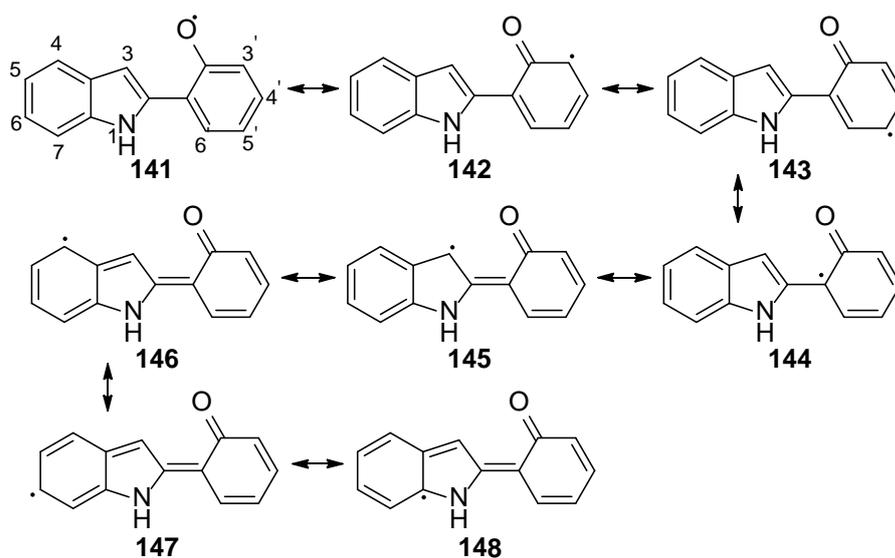
Scheme 60 - The resonance forms of the radical of compound 4

Typically, tertiary radicals are more stable than secondary radicals, which are more stable than primary radicals, and in the radical form of compound **4** the radical can become delocalised across many centres of the molecule (Scheme 60). By adding a sterically small, non-hydrogen bonding electron-donating group (e.g. a methyl group) to a position where the radical can be formally localised, this should lower the relative energy of the radical, therefore reducing the BDE. Conversely, a small, relatively non-hydrogen bonding[277] electron-withdrawing group (e.g. a fluorine) might be expected to have the opposite effect, and therefore increase the BDE. Therefore, BDE values for a series of singly- substituted methylated and fluorinated analogues of compound **4** were calculated, and are shown in Table 9.

HO

4   3   3'

5   4'

6   6'

7   N   5'
    H

4
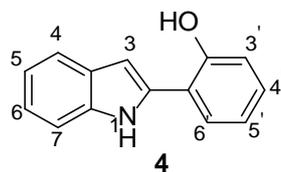
Me-substituted				F-substituted			
Substituent	Compound	BDE	Difference	Substituent	Compound	BDE	Difference
1-Me	236	86.3	9.1	1-F	245	84.5	7.4
6'-Me	237	79.4	2.3	3'-F	246	78.7	1.6
3-Me	186	77.8	0.7	7-F	247	78.3	1.2
-	4	77.1	0.0	6'-F	248	77.8	0.7
5-Me	238	77	-0.1	5-F	249	77.6	0.4
7-Me	239	76.9	-0.2	4-F	250	77.4	0.3
4-Me	240	76.8	-0.3	-	4	77.1	0.0
6-Me	241	76.5	-0.7	6-F	251	76.7	-0.4
4'-Me	242	76.3	-0.8	4'-F	252	76.6	-0.5
3'-Me	243	75.9	-1.2	3-F	253	76.2	-0.9
5'-Me	244	75.6	-1.5	5'-F	254	75.8	-1.3

*Table 9 - Calculated BDE values in kcal mol^{-1} for a series of methyl-substituted and fluoro-substituted derivatives of compound **4**, ordered from highest to lowest BDE value*

The most dramatic changes in the calculated BDE values are found in compounds **236** and **245**, which are derivatised at the 1-position (i.e. *N*-substituents). Substituents at this position have a much higher relative BDE, which can be explained by comparing the structure of the 1-substituted radical species with that of the radical of the parent compound **4**. Although the neutral species for this class of compounds are far from planar (compound **4** has an N-C^2-C$^{1'}$-C$^{2'}$ dihedral angle of -137.8^O), its radical form is planar (N-C^2-C$^{1'}$-C$^{2'}$ dihedral angle for compound **4** = 0.2°, shown in Figure 59), and there appears to be an interaction between the proton of the NH and the oxygen. This interaction considerably lowers the energy of the radical, reducing the size of the BDE. However, in the 1-methyl analogue **236** (shown in Figure 59) and 1-fluoro analogue **245** (not pictured), this interaction is not possible, planarity is lost (1-Me N-C^2-C1'-C$^{2'}$ dihedral angle = 39.4°, 1-F N-C^2-C$^{1'}$-C$^{2'}$ dihedral angle = 171.2°), the relative energy of the radical is raised, and therefore the BDE increases.
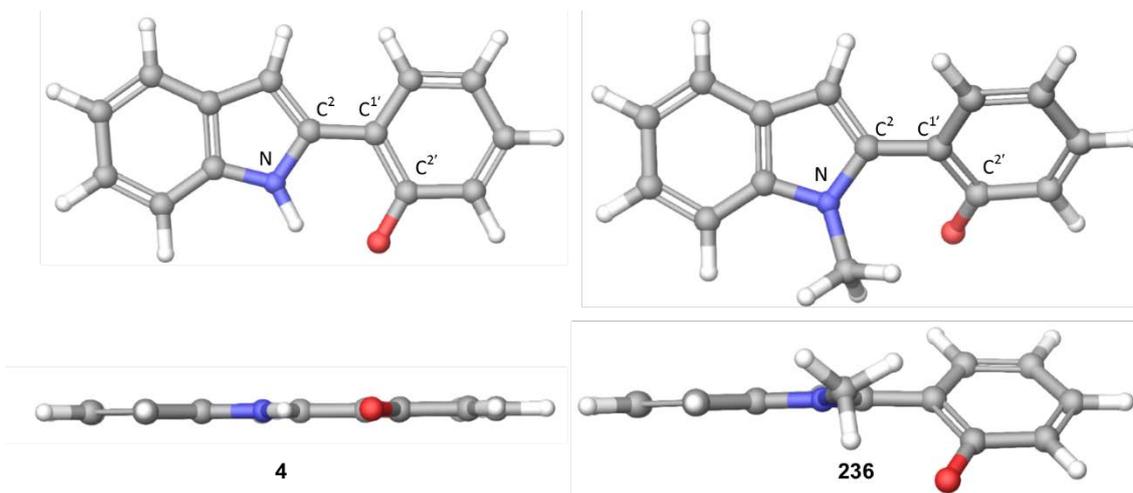
*Figure 59 - The radical forms of compound **4** (left) and its 1-methyl analogue **236** (right), shown from face on (top) and from underneath (bottom), showing how planarity is affected when there are substitutions on the* N-*position*

Steric effects for the 3- and 6'-substituted compounds might be expected to reduce the planarity of their radicals in a similar manner, thus weakening the NH···O interaction. However, on examination of their energetically minimised structures, it would appear that planarity is indeed not lost, with all 3- and 6'-substituted analogues having N-C^2-$C^{1'}$-$C^{2'}$ dihedral angles $< 1^{\circ}$. This suggests that the stabilising NH···O interaction should therefore still be present, strengthening the hypothesis that such an interaction is important. However, as can be seen in Table 9, the 3- and 6'-substituted compounds give uncharacteristically high BDE values, indicating a unfavourable interaction. This unfavourable interaction is more apparent when measuring the N-C^2-$C^{1'}$ and C^2-$C^{1'}$-$C^{2'}$ angles, which shows that in the 3- and 6'-substituted compounds, the linearity of the molecules is affected and that these angles become strained. A comparison of these differences is shown between compound **4** and its 6'-methyl derivative, compound **237**, in Figure 60.

*Figure 60 - The difference in N-C^2-$C^{1'}$ and C^2-$C^{1'}$-$C^{2'}$ angles between compound **4** and its 6'-Me derivative compound **237***

146

Excluding the 1-, 3- and 6'-substituted derivatives for the steric reasons mentioned above, Figure 61 shows data for relative difference in BDE from compound **4** (Table 9) against the position of the substituent, with data for the methylated derivatives on the left and the fluorinated derivatives on the right, as well as a structure in the centre which shows the positions that the radical can be formally placed *via* resonance (as demonstrated fully in Scheme 60). Clear from the figure is that substituents at the positions where the radical will be delocalised leads to greater stabilisation of the species, which is reflected in a lower BDE.

*Figure 61 - BDE values (relative to the BDE of compound **4**) of singly-substituted methyl (left) and fluorine (right) derivatives of compound **4** at the specified positions. The centre structure shows which carbons the radical can formally be located on (shown with an asterisk), as depicted in Scheme 60*

In Table 10 we see that the methylated derivatives tend to decrease BDE values more, by around 0.5 kcal mol^{-1}, when located on positions where the radical can be delocalised to. This was as expected, and is likely to be due to tertiary radicals being more stable than secondary radicals.

Substituents	Radical BDE	Non-radical BDE
Methylated	-0.9	-0.4
Fluorinated	-0.5	0.4

Table 10 - Average calculated BDE values for derivatives located on a potential radical site ("Radical BDE") and not on a potential radical site ("Non-radical BDE")

As can be seen in Figure 61, the trend for the fluorinated derivatives is considerably skewed by the 3'-substituent of compound **246**. This is due to a particularly low energy neutral species due to the hydrogen bonding between the OH group and the 3'-F group, which is shown in Figure 62. This stabilisation makes the loss of the hydrogen radical considerably disfavoured, therefore increasing the BDE value by 1.6 kcal mol^{-1} relative to compound **4**.

*Figure 62 - The structure of the neutral species of compound **246**, which shows the hydrogen bonding between the OH group and the 3'-F group*

By also excluding the 3'- position from the analysis of fluorine-substituted derivatives (along with the sterically hindering substituents at the 1-, 3- and 6'-positions), a trend emerges whereby the fluorinated derivatives tend to decrease BDE values by an average of 0.5 kcal mol^{-1} when located on a position where the radical can be delocalised to, and increase the BDE values by an average of 0.4 kcal mol^{-1} when located on a position where the radical can't be delocalised to (Table 10). This is explained by the fact that fluorine can act as an electron withdrawing group due to its electronegativity, as well as an electron donating group by its mesomeric effect due to its lone pair(s). The data in Table 10 therefore suggests that when the fluorine substituent is located on a position where the radical can be delocalised to, such as the 5'-position, the fluorine substituent stabilises the radical *via* mesomeric effects, whereas when located on a position where the radical can't be delocalised to, such as the 4'-position, the fluorine substituent destabilises the radical by drawing electron density out due to its electronegativity. These examples are depicted in Figure 63.
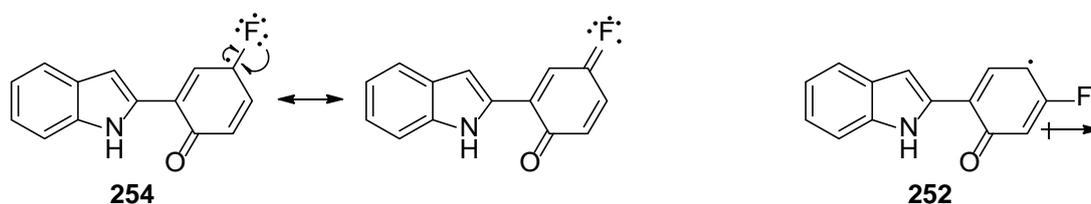
Figure 63 - The stabilising effects of fluorine via *its mesomeric effect (left) and the destabilising effect of fluorine* via *its inductive effect (right)*

6.3.3 Calculated BDE Values

A worked example of how the enthalpic values were determined can be found in Appendix 3, and a full table of calculated BDE, AIP, PDE, PA and ETE data is included in Appendix 4. Appendix 4 includes absolute values for all compounds, but as relative values are often of more use than absolute values, the values are also given relative to compound **4**.

The enthalpic values calculated, a selection of which are shown in Table 11, will be compared against anticancer activity in Chapter 7.

Structure	Compound	Enthalpic values / kcal mol^{-1}				
		BDE	AIP	PDE	PA	ETE
	4	77.1	173.4	231.3	327.0	77.7
	149	90.2	177.4	240.4	345.4	72.4
	214	87.1	187.3	227.5	330.6	84.2
	Hydroquinone	82.1	193.1	216.6	350.1	59.6
	168	86.6	184.9	229.3	339.7	74.5
	152	84.8	172.8	239.6	344.4	67.9
	166	98.0	193.6	232.1	345.1	80.5
	2-Nitrophenol	101.1	223.1	205.5	337.4	91.3
	Phenol	87.7	209.1	206.3	349.3	66.0
	169	87.0	193.9	220.7	342.4	72.3
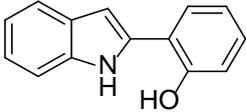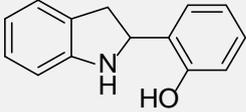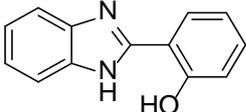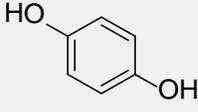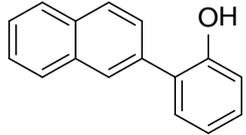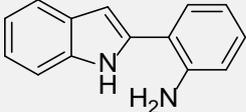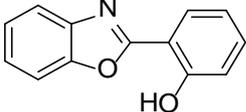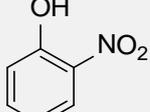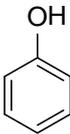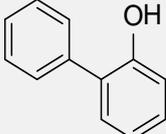
Table 11 - A selection of compounds included in this work, with their calculated BDE, AIP, PDE, PA and ETE values shown. Full data is included in Appendix 4

These enthalpic values that have been calculated are gas phase values, and any solvation or enzymatic effects on radical formation are not considered. However, in lieu of not knowing what solvation or enzymatic effects are occurring in a complex cellular environment, these gas phase values are the only data from which we can attempt to draw conclusions. From a search of the literature, it would appear that relative trends for the enthalpic values within a chemical class and mechanistic pathway are unaffected by solvation, although the favoured mechanistic pathway may be solvent dependant.[248,250,251,278–280]

As discussed in the introduction to this chapter, there are four possible mechanistic pathways in which a phenolic radical can be formed from a phenol. These pathways are hydrogen atom transfer (HAT), sequential proton loss-electron transfer (SPLET), single electron transfer (SET) or sequential electron transfer-proton transfer (SET-PT). Each mechanistic step has an enthalpic value associated with it; bond dissociation enthalpy (BDE) for HAT, proton affinity (PA) for PL, electron transfer enthalpy (ETE) for ET, adiabatic ionisation potential (AIP) for SET and proton dissociation enthalpy (PDE) for PT. A smaller enthalpic value indicates a more favoured mechanistic step. In the case of sequential mechanistic steps, the highest enthalpic value of the multiple steps is likely to act as the rate determining step.

As can be seen by the representative data in Table 11, in the gas phase the most probable radical forming mechanism is HAT, due to the relatively low BDE values. This appears to be in agreement with the considerable number of publications in the literature which appear to focus much more on BDE than any other enthalpic value.[248,250,251,253,281]

As predicted relative BDE values hold far greater accuracy when comparisons are made within structurally similar series of compounds, we follow suit in our analysis herein.
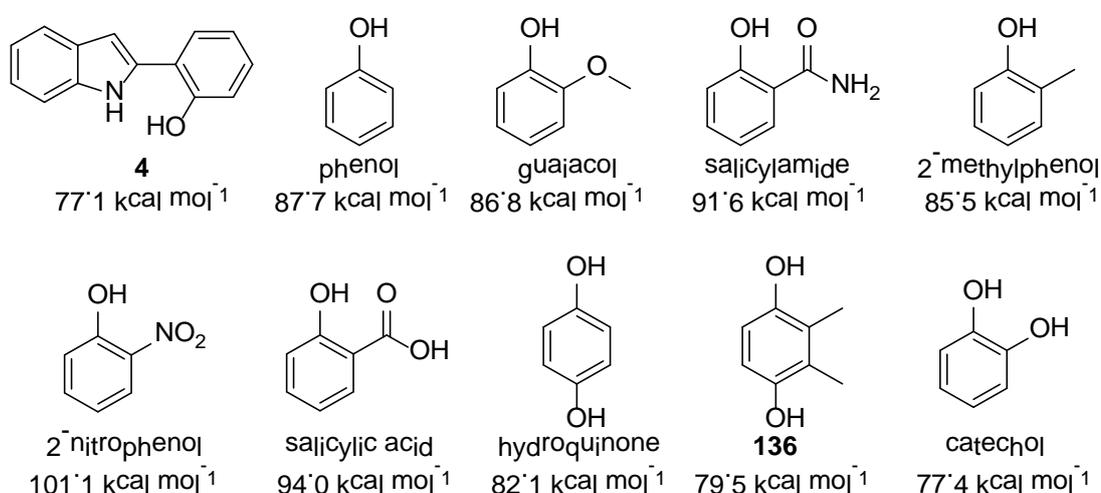
Figure 64 - The BDE values of compound **4** and all simple phenols

The first class of compounds are simple phenols (Figure 64), most of which are *ortho*-substituted phenols. The trend in BDE values amongst this class of compounds can be well described by Hammett constants,[282] which is understandable considering that Hammett describe the extent to which substituents donate or withdraw electron density from a conjugated system, which is a major way in which phenolic radicals can be stabilised. Although the majority of the compounds in this class are *ortho*-substituted phenols, Hammett constants are only available for *para*-substituents in order to focus exclusively on the electronic effects and therefore exclude steric effects. Due to the *ortho-* and *para-* positions being conjugated and that Hammett constants exclusively describe electronic effects, using *para-* constants is considered valid.
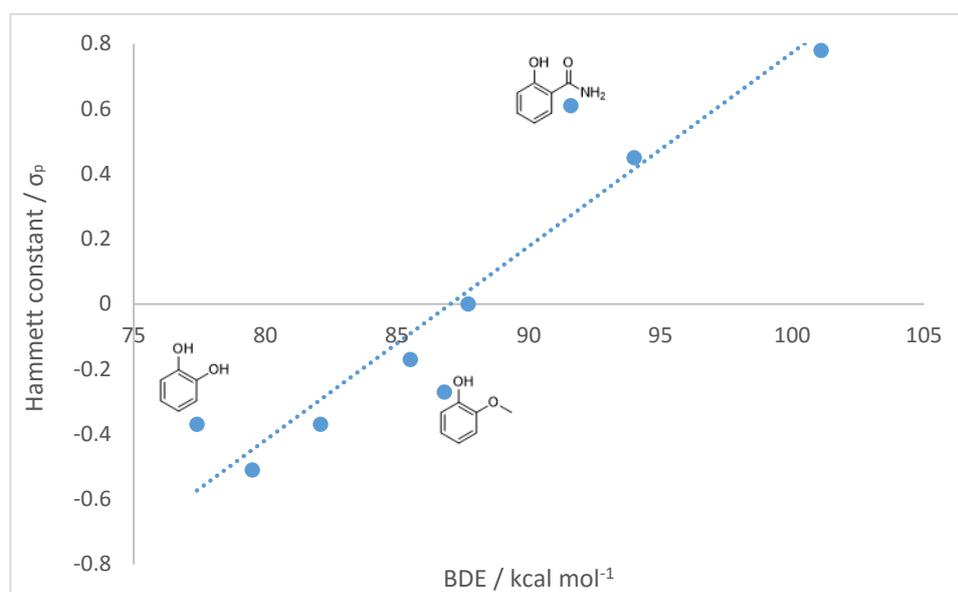
Graph 4 - The effect of Hammett constants on BDE values for the series of nine simple phenols shown in Figure 64

Graph 4 shows a good correlation between BDE values plotted against Hammett constants (R^2 = 0.866). There are three obvious outliers that break a close to perfect linear trend. The two outliers above the trendline (the *o*-$CONH_2$ of salicylamide and *o*-OH of catechol) have lower BDE values than would be expected by their Hammett values due to hydrogen bonding between the substituent and the radical, therefore lowering the energy of the radical. The outlier below the line is the *o*-OMe derivative, guaiacol, which has a higher BDE value than expected due to the hydrogen bonding present in the neutral species, resulting in a lowering of energy of the neutral species and therefore a higher BDE value.
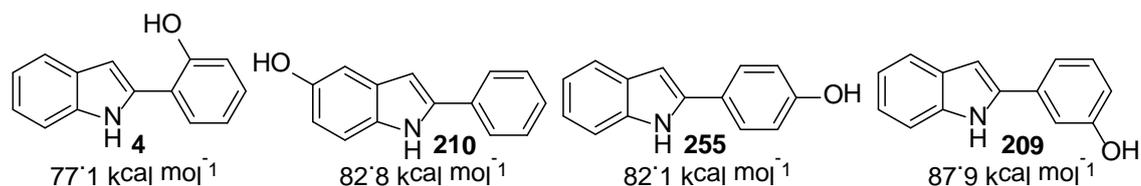
*Figure 65 - Compound **4** and some of its 2-phenylindole isomers*

The next class of compounds are isomers of compound **4** that are based around the 2-phenylindole scaffold, the structures of which are shown in Figure 65. Of this class, compound **4** has the lowest BDE value (77.1 kcal mol^{-1}), largely down to the stabilising intramolecular interaction between the NH group and the phenolic radical, as discussed in Section 6.3.2. As would be expected due to their conjugated positions, the 5-OH isomer (compound **210**) and 4'-OH isomer (compound **255**) have very similar BDE values (82.8 kcal mol^{-1} and 82.1 kcal mol^{-1} respectively), differing by only 0.7 kcal mol^{-1}. Trying to determine why the 3'-OH isomer (compound **209**) has a considerably higher BDE value (87.9 kcal mol^{-1}) than compounds **210** and **255** is more difficult. They all have the same number of canonical radical forms (Figure 66), and there is no intramolecular interactions in any of the compounds. It does, however, seem that the spin density of compound **209** is less delocalised, with 95% of the unpaired electron being localised to the phenolic ring in compound **209**, compared to 72% for compound **255**. With less delocalisation for the radical of compound **209**, the radical is relatively less stable and has a higher BDE value.
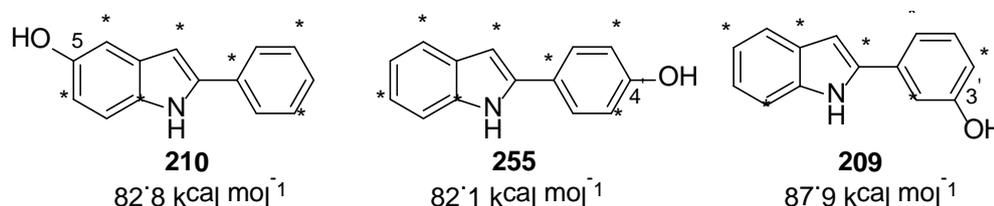
*Figure 66 - Canonical radical forms of compounds **210**, **255** and **209**, with BDE values for each structure shown underneath*

The next class of compounds, shown in Figure 67, are compounds which contain a pyridone as part of their structure.

256
81·7 kcal mol⁻¹

257
73·2 kcal mol⁻¹

213
93·4 kcal mol⁻¹

258
95·6 kcal mol⁻¹

259
97·1 kcal mol⁻¹

260
95·1 kcal mol⁻¹

4
77·1 kcal mol⁻¹
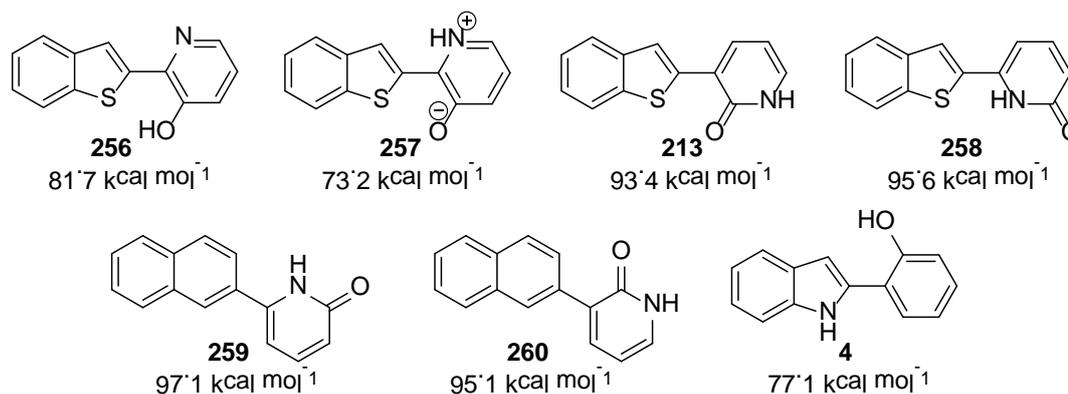
Figure 67 - Compound 4 alongside the compounds which contain a pyridone or hydroxypyridine as part of their structure. BDE values are shown below each structure

The BDE values shown in Figure 67 show that the pyridone compounds **213**, **258**, **259** and **260** have consistently higher BDE values. These compounds are, however, considerably different to compound **4** due to the BDE value being based on the loss of a hydrogen from an NH group rather than a phenolic group.

The 3-hydroxypyridine analogues shown in Figure 67 (compound **256**, for which the BDE of the OH group was determined, and compound **257**, for which the BDE of the NH group was determined) were both considered plausible under physiological conditions by LigPrep. They do, however, have vastly different BDE values. It would be probable that these two compounds can easily tautomerise, meaning that the lower BDE value of compound **257** would be the more appropriate value to focus on. However, compound **257** is not a phenol analogue, so a direct comparison to compound **4** is difficult to make. Nonetheless, compound **257** is one of the few compounds analysed that has a lower BDE value than compound **4**, presumably because of the weak H-N⁺ interaction.

The next class of compounds, shown in Figure 68, is made up of analogues of compound **4** which have different scaffolds (i.e. scaffold-hopping analogues). The effect of hydrogen-bonding in the neutral species is immediately apparently, with all the compounds that have been found to participate in intramolecular hydrogen bonding *via* molecular modelling having consistently higher BDE values than compound **4**. The OH group compound **217** is seen to hydrogen bond to an oxygen, whereas the OH groups in compounds **166**, **167** and **214** are seen to hydrogen bond to the lone pair of the nitrogen. The loss of the stabilising NH···O interaction in the radical form of compounds **160**, **169** and **168** is also met with an increase in BDE values. Most noticeably, by moving the phenol from the 2-position of the indole (as in compound **4**, BDE = 77.1 kcal mol⁻¹)

to the 3-position (as in compound **160**, BDE = 82.9 kcal mol^{-1}) results in a BDE increase of 5.8 kcal mol^{-1}.

4
77.1 kcal mol^{-1}

160
82.9 kcal mol^{-1}

217
90.0 kcal mol^{-1}

169
87.0 kcal mol^{-1}

168
86.6 kcal mol^{-1}

166
98.0 kcal mol^{-1}

167
92.5 kcal mol^{-1}

261
83.2 kcal mol^{-1}

214
83.2 kcal mol^{-1}
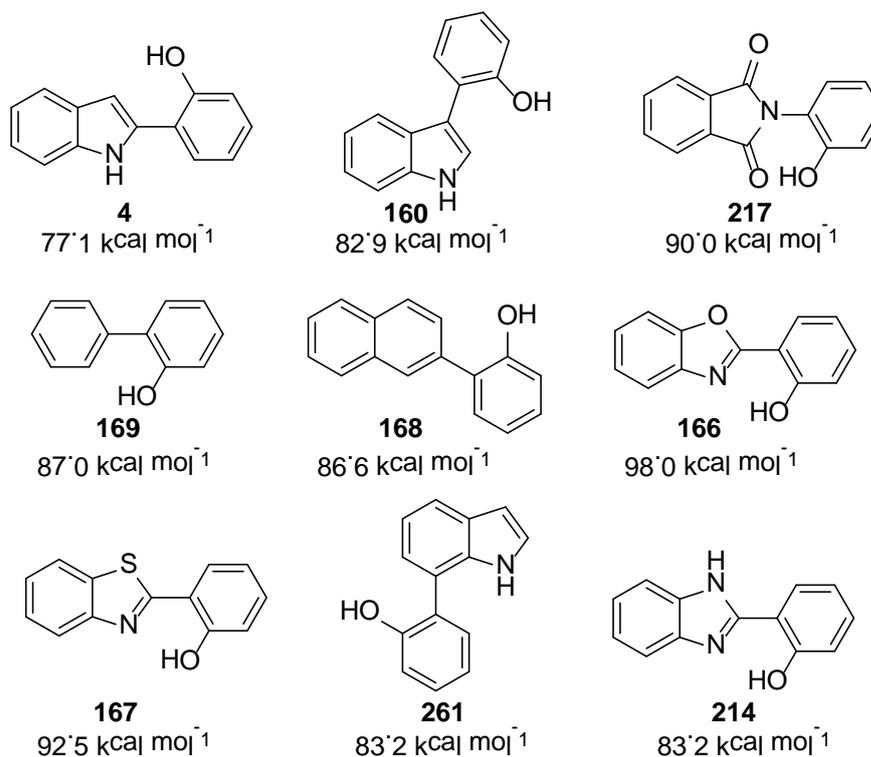
*Figure 68 - Scaffold-hopping analogues of compound **4**. BDE values are shown below the structures*

Comparing the isomers of compound **4** (compound **4**, compound **160** and compound **261**), an unexpected trend emerges. Compound **261**, which still has the stabilising NH···O interaction in its radical form (albeit to a lesser extent) has a higher BDE value than even compound **160**, which has no stabilising interaction at all. This is despite all three compounds having the same number of canonical radical structures. This unexpectedly high BDE for compound **261** may, however, just be due to the considerable strain that the radical form of the compound is under, as shown in Figure 69; the radical species is both out of plain and has both the C-O and N-H groups forced away from their natural position
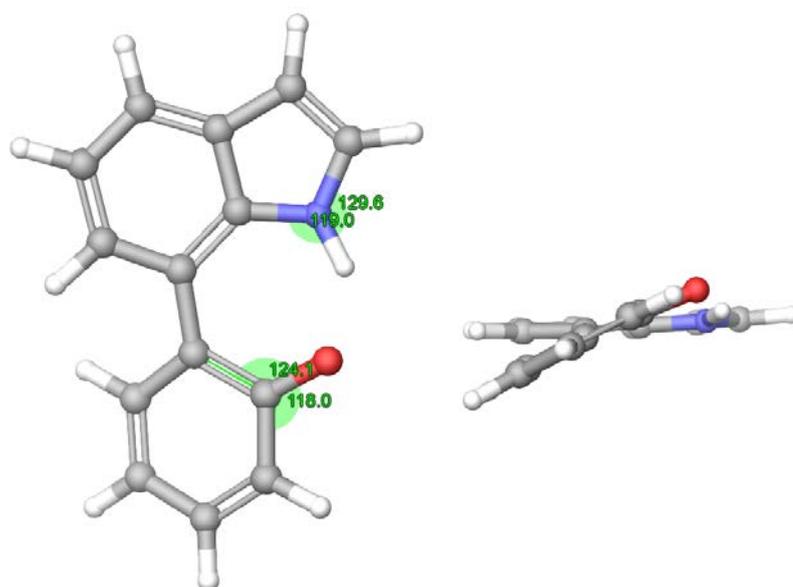
*Figure 69 - Bond angles of the radical form of compound **261** (left) and a view that shows the extent of planarity in this species (right)*

The next class of compounds is much smaller, and includes anilinic analogues of compound **4**, the structures of which are shown in Figure 70. The phenolic compound **4** being 7.7 kcal mol^{-1} lower than the anilinic compound **152** is as expected, as the BDE values of phenols are known to be lower than those of anilines.[74,283] The comparison of compound **152** to compound **228** is equivalent to the comparison of compound **4** with compound **214** in the previous series, and the anilinic analogues do indeed maintain the same trend, with the reasonably strong intramolecular hydrogen bond making the abstraction of a hydrogen atom less favoured, resulting in the BDE of compound **228** being 12.5 kcal mol^{-1} higher than that of compound **152**.
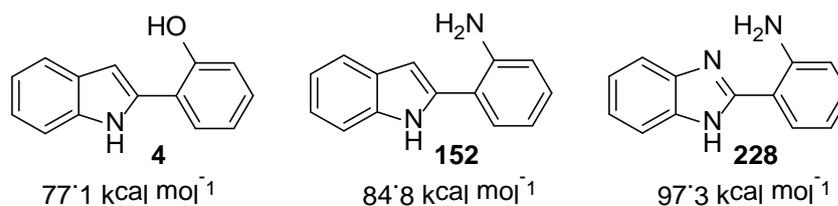
*Figure 70 - Compound **4** alongside two anilinic analogues. BDE values are shown below the structures*

The final compounds to be discussed are compounds which don't fit in any other class, and are shown in Figure 71.
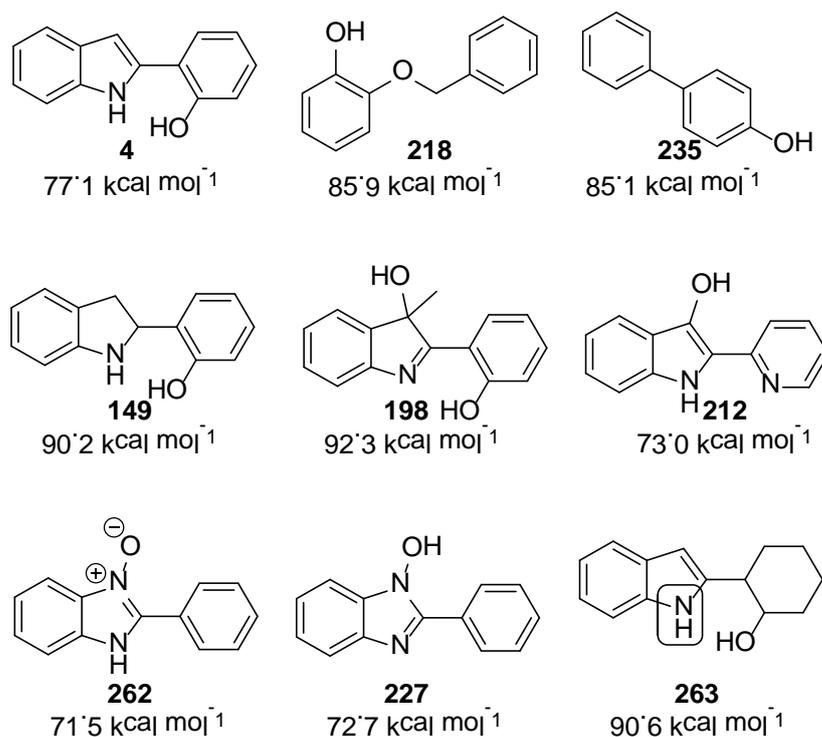
*Figure 71 - Compound **4** alongside a number of analogues which don't belong to any other class discussed thus far. BDE values are shown below structures*

So far, the majority of compounds have had a higher BDE value than compound **4**, so the stand-out compounds in Figure 71 are those that have a lower BDE value than that of compound **4** (BDE = 77.1 kcal mol^{-1}), which are compound **212** (BDE = 73.0 kcal mol^{-1}) and the tautomeric compounds **262** (BDE = 71.5 kcal mol^{-1}) and **227** (BDE = 72.7 kcal mol^{-1}).

Compound **212**, which has a BDE value 4.1 kcal mol^{-1} lower than that of compound **4**, is the only compound within this work that has the 2-(2-pyridyl)indole scaffold. This indicates that this scaffold may be of interest if a compound with lower BDE value was sought. An obvious compound of interest would therefore be an analogue which would have the radical-stabilising NH···O interaction that compound **4** has, such as compound **264** shown in Figure 72.
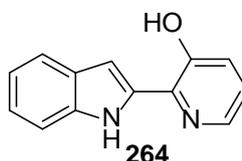
Figure 72 - Possible compound of interest if a compound with a lower BDE value was sought

The tautomers **262** and **227** also have particularly low BDE values of 71.5 kcal mol^{-1} and 72.7 kcal mol^{-1}, compared to a value of 77.1 kcal mol^{-1} for compound **4**. These compounds also have considerably lower BDE values than the isomeric compound **214** (Figure 68), which has a

BDE value of 83.2 kcal mol^{-1} due to the hydrogen bonding in the neutral species making the hydrogen atom more difficult to abstract. An analogue of this compound which allows for the stabilising radical NH···O interaction that compound **4** has is shown as compound **265** in Figure 73.

265
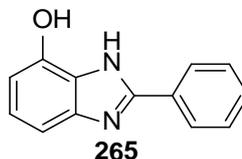
*Figure 73 - Possible analogue of compounds **214** and **227** which allows for a stabilising NH···O radical interaction*

6.4   Concluding statements

This chapter has shown the method used to determine the enthalpic values for the possible mechanisms involved in forming a phenolic radical from a phenol. It has been shown that in the gas phase, the most probable mechanism for this formation hydrogen atom transfer (HAT), the enthalpic value for which is bond dissociation enthalpy (BDE).

Section 6.3.3 discussed the effects on BDE brought on by structural changes, and found that the lead compound **4** has a reasonably low BDE of 77.1 kcal mol^{-1}, with only a small number of analogues being shown to have a BDE value lower than this (Figure 74).

| **262** | **227** | **212** | **257** |
| 71.5 kcal mol^{-1} | 72.7 kcal mol^{-1} | 73.0 kcal mol^{-1} | 73.2 kcal mol^{-1} |
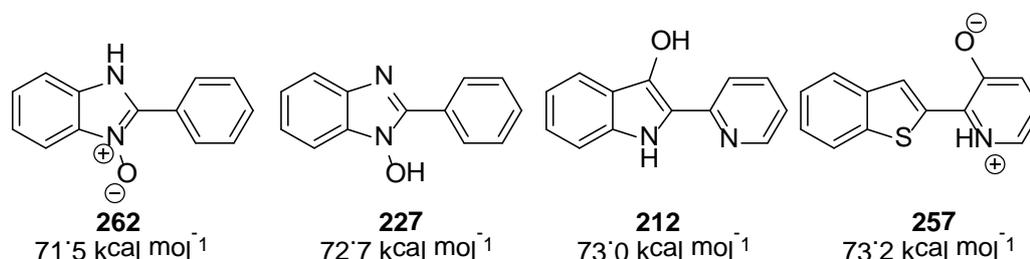
*Figure 74 - The compounds with a lower BDE value than compound **4**, with BDE values shown below each structure*

In the next chapter, the anticancer activity of the compounds described in Chapter 4 and Chapter 5 is established, and the hypothesis that the anticancer activity of compound **4** and its analogues is related to the ease with which they can form their phenolic or anilinic radicals is investigated.

CHAPTER 7 – ANTICANCER ASSAY RESULTS AND DISCUSSION

7.1   Introduction

Throughout this project, the analogues of compound **4** and of I3C and compound **5** have been considered as entirely separate series.   There are a total of 12 compounds in the I3C and compound **5** series and 30 compounds in the compound **4** series.   This chapter describes how the anticancer activity of these compounds were determined, as well as building up a basic structure-activity relationship (SAR) based on these activities.

7.1.1   Anticancer Assays

As it would be unethical and impractical to test prospective anticancer compounds on animals or humans at such an early stage in the drug development process, the efficacies of compounds in this work are established on cultured mammalian cells.   The more active compounds will obviously kill a larger number of cultured cancerous cells, but cell death is difficult to directly measure. The workaround for this is to measure the remaining healthy cells using an established assay, and deduce the efficacy from that.

By far the most well-known, established and accepted assay is the MTT assay. This colourimetric assay determines cellular metabolism by detecting NADH and NADPH, sometimes collectively referred to NAD(P)H.   NAD(P)H reduces the straw-yellow MTT into a ring-opened formazan, which has a bright purple colour. Conversion of MTT to its formazan is therefore correlated to the number of viable cells.
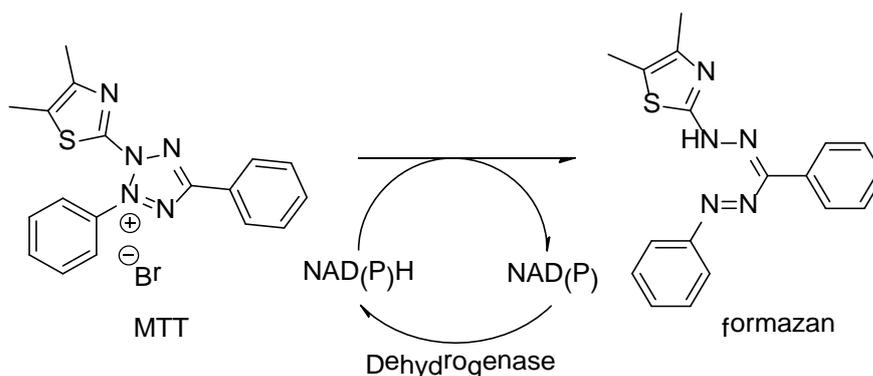
Scheme 61 - Single-electron reduction of MTT to a formazan by NAD(P)H

There are many variants on the MTT assay, such as the MTS assay, which has a simpler protocol and a slightly different tetrazolium salt. The MTS assay, like the MTT assay, is a colourimetric assay whereby MTS undergoes single-electron reduction. Unlike in the MTT assay, MTS cannot permeate the cell, so requires an intermediate electron carrier such as phenazine methosulfate (PMS). A representative reaction cycle is shown in Scheme 62.
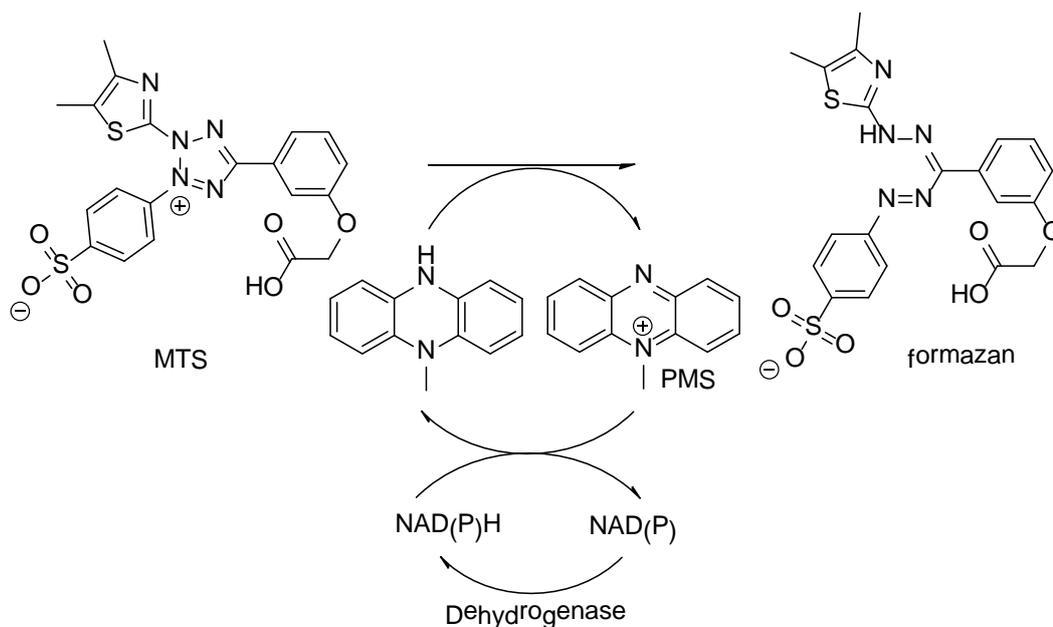
Scheme 62 - Single-electron reduction of MTS to a formazan by NAD(P)H, utilising the intermediate electron carrier PMS, as occurs in the MTS assay

There are, however, a number of downsides to these methods. Primarily, the formazan products are often toxic, resulting in a possible harm to researchers, as well as killing the cells, meaning that only one measurement may be made at a time. The protocol is also highly time-constrained; once the protocol has been begun, there is no point at which it can be paused, somewhat limiting its use.

An alternative to NAD(P)H assays such as the MTT and MTS assays is the Sulforhodamine B (SRB) assay, which, although being less common, has been commonly used since its original publication in 1990.[284] The process involves first fixing the cells with trichloroacetic acid (TCA), then applying the SRB dye to cells. The assay works by the SRB dye binding to basic amino acids on the surface of the fixed cells. Unbound dye is then washed off, and the quantity of the bright pink dye is determined colourimetrically. The assay therefore measures cellular protein content, which is proportional to the number of healthy cells. Because the cells are fixed, the end-point measurement is not time critical, which can make for a preferred procedure. However, as with the MTT/MTS assay, the cells are killed in the process, meaning that only one measurement may be made at a time.

A relatively modern assay that is rapidly becoming very popular is the PrestoBlue assay. The active agent in PrestoBlue is resazurin, which is reduced to resorufin by NAD(P)H (Scheme 63). Much like the MTT and MTS assays, the PrestoBlue assay therefore detects metabolically healthy cells.
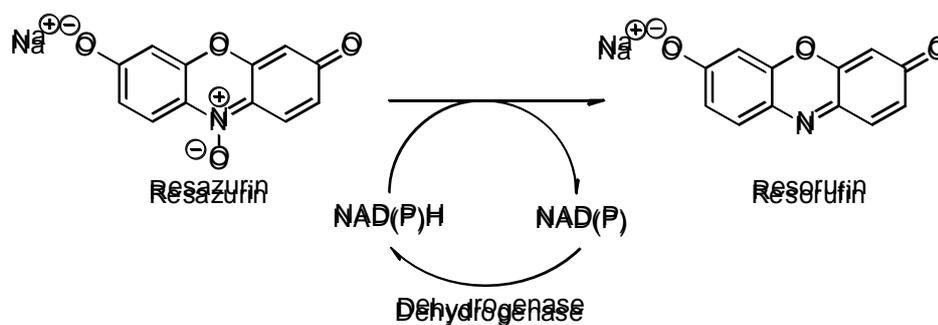
Scheme 63 - Single-electron reduction of resazurin to resorufin by NAD(P)H, as occurs in the PrestoBlue cell viability assay

Although the PrestoBlue assay is more expensive than the MTT, MTS or SRB assay, it is gaining widespread usage due to being non-toxic, allowing for the dye to be removed after measurements are carried out, thus allowing the cells to continue growing. This significantly reduces the workload associated with time course experiments, as well as being safer for those carrying out the experiments.

7.2   Replicating Past Work

This work is a continuation of the work done previously within the Snape group,[6] in which a small number of analogues were tested on the two cell lines 1321N1, an astrocytic cell line, and U87, a glioblastoma cell line. These IC_{50} values, which were determined using the MTS assay, are displayed in Table 12.

Structure	Compound	MTS assay IC$_{50}$ value / μM	
		132N1	U87
	I3C	309 ± 2	526 ± 2
	5	111 ± 1	176 ± 1
	2-phenylindole	---	---
	266	---	---
	4	445 ± 4	379 ± 4
	267	---	---
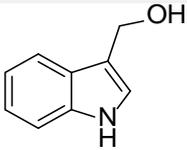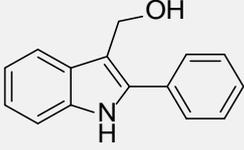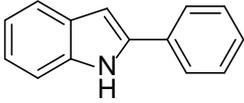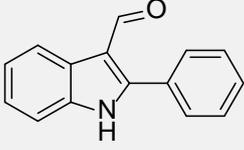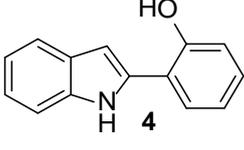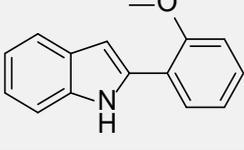
*Table 12 - IC$_{50}$ values for some analogues of I3C and compound **4**, as determined by Prabhu.[6] Dashes indicate that IC$_{50}$ values were not reached*

To confirm that the trends shown by Prabhu are repeatable, it would be necessary to retest the active compounds previously tested using a different assay (Figure 75). Whereas Prabhu *et al.* had used the MTS assay, the retesting was done here using the PrestoBlue assay (the procedure for which can be found in Appendix 5). The PrestoBlue assay was chosen due to evidence suggesting that the MTS assay is unreliable when working with antioxidant molecules such as phenols, a limitation that the PrestoBlue assay does not have. These three active compounds were also retested by Dr Farzana Rowther and Dr Ibrahim Tolaymat at the University of Wolverhampton using the sulforhodamine B (SRB) assay (the procedure for which can be found in Appendix 6).
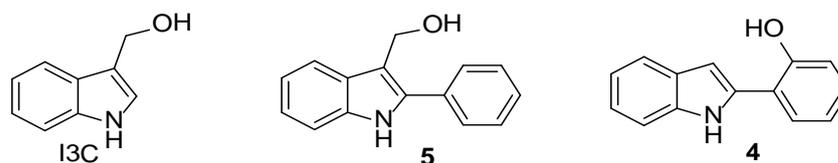
Figure 75 - The active compounds found by Prabhu et al.*, I3C (left), compound 5 (centre) and compound 4 (right)*

The results for the retested compounds can be seen in Table 14 and Table 14, where "high" indicates that at the highest concentration tested a maximal inhibition had still not been attained and "n/a" indicates that the compound was not tested on that cell line. In both instances the compounds were tested against the glioblastoma cell line U87, but Prabhu *et al.* used the less aggressive astrocytoma cell line 1321N1 as a control cell line,[6] whereas this work used the non-cancerous astroglial cell line SVG p12 as a control cell line. Rowther and Tolaymat used no control cell line, but instead included a second glioblastoma cell line U251, along with testing the two more active compounds on three short-term glioblastoma cell cultures. Although the values differ significantly between the assays and methodologies, compound **5** is shown in each case to be the most active, and I3C is shown to be the least active, so the relative trends are the same throughout. From these data, it can be confidently confirmed that these classes of compounds respond well to a wide variety of assay conditions, and that the activities reported are representative of the true activities of the compounds and are not skewed by inappropriate assay conditions.

Compound	MTS		PrestoBlue	
	U87	1321N1	U87	SVG p12
I3C	526	309	374	348
5	176	111	113	88.5
4	379	445	207	86.1

Table 13 - EC$_{50}$ values (μM) for compounds tested by Prabhu et al. using the MTS assay and this work using the PrestoBlue assay

Compound	SRB				
	U87	U251	IN1472	IN1528	IN1760
I3C	High	High	n/a	n/a	n/a
5	134	279	300	411	552
4	233	338	332	497	466

Table 14 - EC$_{50}$ values (μM) for compounds tested by Rowther and Tolaymat using the SRB assay

7.3   Methodology

Aside from the work done in replicating prior results, all anticancer testing was done by Dr Farzana Rowther and Dr Ibrahim Tolaymat at the University of Wolverhampton.   Their methodology, which is described in more detail in Appendix 6, includes the use of the SRB assay[284] to determine EC$_{50}$ values of all trial compounds.

Rowther and Tolaymat initially tested all compounds against two cell lines and three short-term cultures: U87, U251, IN1472, IN1528 and IN1760. U87 and U251 are both established cell lines derived from glioblastomas, and are both commonly used as glioblastoma models. IN1472, IN1528 and IN1760 are all short term cultures. In the interest of improving throughput, it was decided that compounds tested in the later series would be tested against fewer cultures. For these tests, the U251 cell line was chosen over the U87 cell line due to evidence that it acts and reacts more similarly to a true glioblastoma.[285] With respect to the short term cultures, there is much less empirical evidence as to which ones act as an appropriate model, so choices had to be made based on personal experience. The IN1760 cell line was included due to its observed chemoresistance, and may therefore be able to identify chemotherapies that may be active against chemoresistant tumours. Between the IN1472 and IN1528 cultures, IN1472 had been shown to give less reproducible data than IN1528, so IN1472 was dropped from the panel of cultures against which the later series of compounds were tested. This panel of cell cultures was therefore made up of U251, IN1528 and IN1760.

7.4   Results and Discussion

In this section, the significance and effects of the structural features investigated will be discussed in terms of their effects on anticancer activity, thus building up a structure-activity relationship (SAR). As has been the case throughout this work, the I3C and compound **4** series of compounds will be discussed independently of one another.

7.4.1 The Structure-Activity Relationship of the I3C and Compound 5 Series

As discussed in Chapter 1, there are a number of structural features of compound **5** that were to be investigated. These features included the role of the 2-phenyl group, the carbinol group, *N*-substituents, the role of the heteroatom and comparing the prodrug to the active form. These structural features and the analogues designed to probe them are summarised in Figure 76.
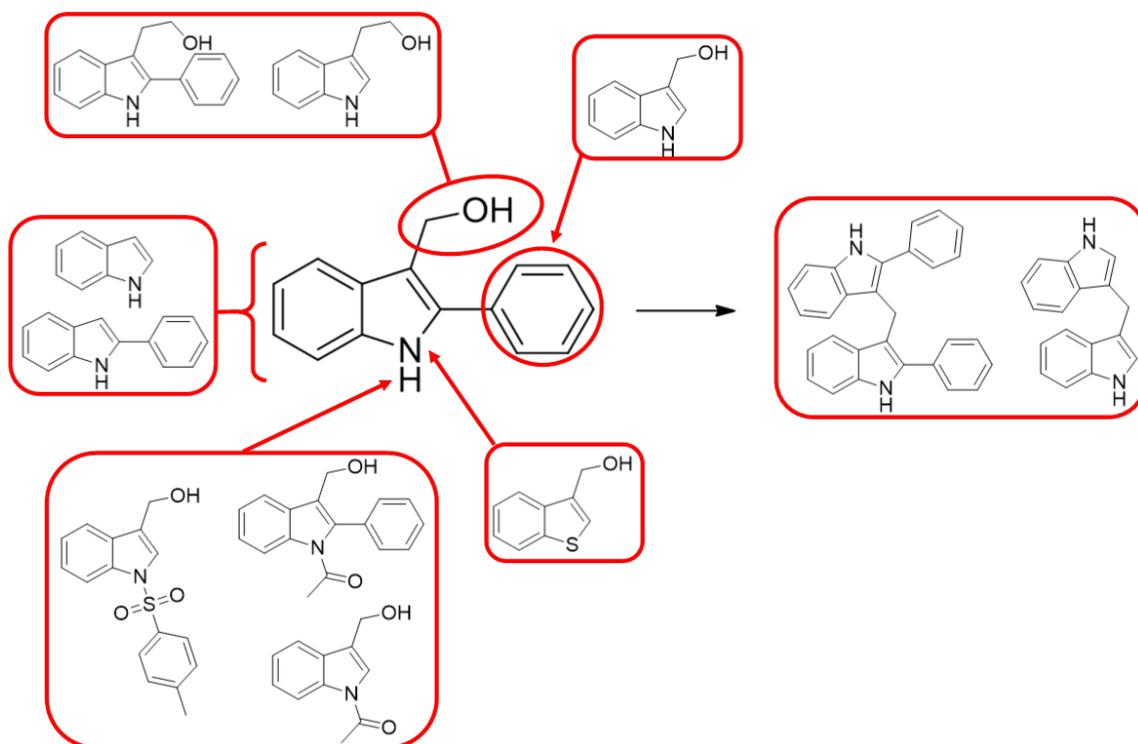
*Figure 76 - The structural features of compound **5**, and the analogues tested to investigate these structural features*

7.4.1.1 The role of the 3-carbinol group
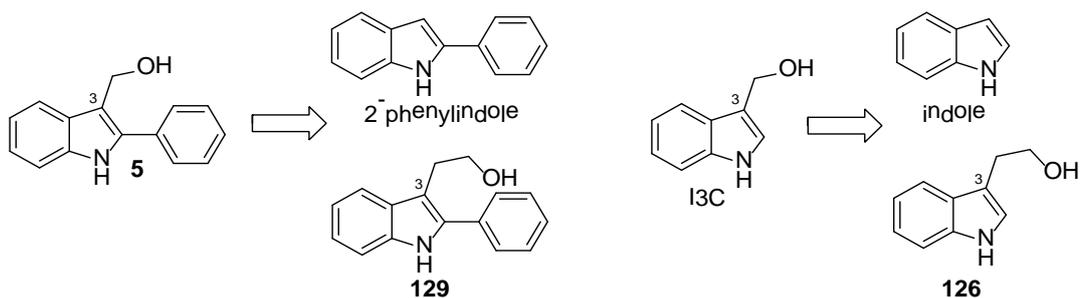
Figure 77 - Analogues designed to probe the significance of the 3-carbinol group

I3C is considered a prodrug to the active form of the drug 3,3-'diindolylmethane (DIM). The 3-carbinol group is considered important due to its role in the metabolism of I3C to DIM, as shown in Scheme 64. The mechanism involves the loss of the OH group (which is lost as water), which relies on the resultant carbocation being stabilised *via* conjugation. By extending the

length of the carbon chain from a carbinol (i.e. a CH_2OH, as in I3C and compound **5**) to hydroxyethane (i.e. a CH_2CH_2OH as in compounds **129** and **126**), the carbocation formed *via* loss of the OH can't become delocalised, and the diindolylmethane cannot be formed. Likewise, by having no 3-substituent at all (as with indole and 2-phenylindole), a diindolylmethane metabolic product cannot be formed.
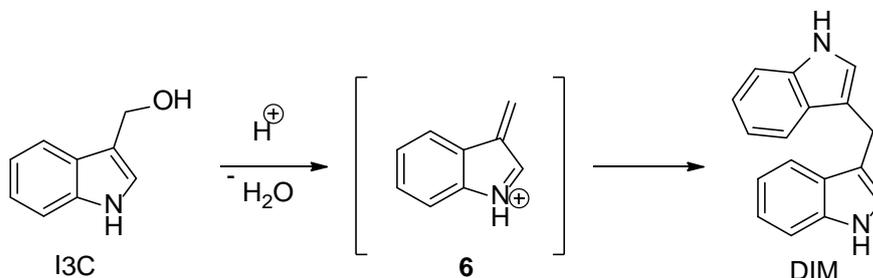
Scheme 64 - The degradation of I3C to DIM is reliant upon a conjugated 3-carbinol group

Initially, the anticancer activities of I3C and compound **5** were compared against those of indole and 2-phenylindole respectively (Table 15). Against the five varieties of cultured cells against which the compounds were tested, both I3C and indole had EC_{50} values that were too high to calculate based on the concentration ranges used, so no conclusions can be drawn from this. However, compound **5** was shown to have EC_{50} values in the range of 133 – 552 µM, whereas 2-phenylindole was shown to have EC_{50} values that were too high to calculate, showing a large reduction in activity on removal of the 3-carbinol group.

Compound	U87	U251	IN1472	IN1528	IN1760
I3C	-	-	-	-	-
5	133.4	279.2	299.9	410.5	551.8
indole	-	-	-	-	-
2-phenylindole	-	-	-	-	-

Table 15 - Comparison of EC_{50} values (µM) upon removal of the 3-carbinol group

Exchanging the 3-carbinol group of I3C and compound **5** with a 3-ethan-2-ol group in compounds **126** and **129** yields less conclusive results (Table 16). Comparing I3C with its 3-ethan-2-ol analogue **126** shows that I3C has consistently higher activity across all five cultures on which they were tested. However, the 2-phenyl analogues show much less consistency, with compound **5** having the higher activity against the two established cell lines U87 and U251, whereas its 3-ethan-2-ol analogue, compound **129,** has the higher activity against the three

short-term cultures IN1472, IN1528 and IN1760. These trends were confirmed upon retesting at a later date on the cultures U251, IN1528 and IN1760.

Compound	U87	U251	IN1472	IN1528	IN1760
I3C	290	390	950	119	1580
5	30	130	480	10	1430
126	580	1220	1950	1740	4570
129	110	250	350	4	330

Compound	U251	IN1528	IN1760
I3C	1040	2740	35940
5	390	2220	-
126	5440	44810	-
129	600	1080	8690

Table 16 - Comparison of EC$_{50}$ values (μM) upon changing the 3-carbinol group to a 3-ethan-2-ol group

7.4.1.2 The effect of N-substituents
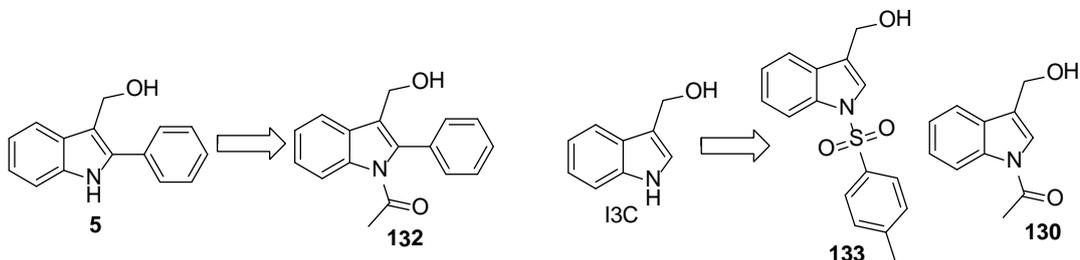
Figure 78 - Analogues designed to probe the effect of adding N-substituents

As shown in Scheme 64, a cationic intermediate is involved in the conversion of the prodrug I3C to the active form of DIM, and it is thought that the stability of this intermediate impacts the quantity of the diindolylmethane product produced. By increasing the stability of the intermediate, yield of the diindolylmethane should increase, and the observed anticancer activity should therefore improve. Conversely, destabilising the cation should result in lower anticancer activity.

Figure 78 shows the three analogues that were produced in order to probe this hypothesis. The two *N*-substituents chosen are both electron withdrawing groups of different strengths (the acetyl group is much less electron withdrawing than the tosyl group).

When tested against three glioblastoma cell cultures, the effect of the acetyl shows a clear negative effect on anticancer activity, with both I3C and compound **5** having better activity than their *N*-acetyl analogues **130** and **132** (Table 17). This implies that the *N*-acetyl group is not effectively removed either by hydrolysis or by enzymatic activity.

Compound	U251	IN1528	IN1760
I3C	1040	2740	35940
5	390	2220	-
130	1760	5820	31720
132	600	19800	76050

*Table 17 - Comparison of EC$_{50}$ values (μM) between I3C and compound **5** and their* N-*acetyl analogues*

Conversely, an improvement in activity is noticed between I3C and its *N*-tosyl analogue **133** (Table 18). This was unexpected, as an electron withdrawing group at the *N*-position would be expected to destabilise the intermediate cation, therefore reduce the quantity of the resultant diindolylmethane. It was also unexpected since the *N*-acetyl group (which is also electron-withdrawing) shows a decrease in activity. These findings do however agree with a literature precedent, that shows that *N*-sulfonyl analogues of I3C have a significant improvement in activity compared to I3C itself when studied against cancerous cell cultures.[286]

Compound	U251	IN1528	IN1760
I3C	380	700	3370
133	140	440	650

Table 18 - Comparison of EC$_{50}$ values (μM) between I3C and its N-*tosyl analogue*

After testing two different electron-withdrawing *N*-substituents, it seems apparent that *N*-substitution is not simply an electronic change, and that the specific substituent may have a more complicated effect. This effect may be related to the ability of the prodrug to form the active compound, the interaction of the *N*-substituted active compound with a protein, or the solubility or uptake of either the prodrug or active compound. The fact that the range of activities for the two *N*-substituents is so large suggests that the *N*-position would be a viable position to investigate derivatising.

7.4.1.3 The effect of a different heteroatom in the heteroaromatic system
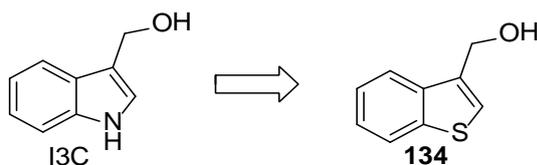

Figure 79 - The benzothiophene analogue designed to probe the effect of changing the heteroatom

As with the *N*-substituted analogues, the benzothiophene analogue **134** differs from I3C in that the intermediate cation will have a different stability. This is due to two reasons: the changing of the most electronegative atom and the reduced aromaticity in the five-membered ring compared to indole.[287]

Compound	U251	IN1528	IN1760
I3C	380	700	3370
134	1240	3990	1080

Table 19 - Comparison of EC$_{50}$ values (μM) between I3C and its benzothiophene analogue

Comparing the EC$_{50}$ values of I3C and compound **134**, there is no clear superior compound, with I3C having better activity against the U251 and IN1528 cell lines and compound **134** having better activity against the IN1760 culture.   If one was searching for a new primary chemotherapy, this data would obviously be inconclusive as to which is the superior chemotherapy. However, as IN1760 is considered by groups that work with such short-term cultures to be a particularly chemoresistant cell line, the benzothiophene analogue may turn out to be a useful fragment against chemoresistant tumours after further optimisation.

7.4.1.4 The effect of administering the isolated diindolylmethane
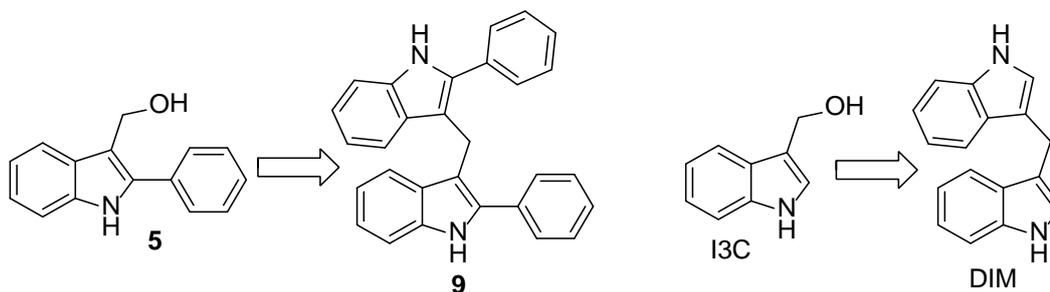
*Figure 80 - The diindolylmethane metabolites of compound **5** and I3C*

As the *in vitro* conversion of I3C to DIM is thought to be far below 100%,[128] by testing the isolated active form, it can be ensured that there is a much higher bioavailability of the active

compound. Assuming the active compound can be taken up by the cells, this should result in a higher observed anticancer activity.

Compound	U251	IN1528	IN1760
I3C	380	700	3370
9	50	270	50
DIM	280	440	230

Table 20 - Comparison of EC$_{50}$ values (μM) between I3C and some diindolylmethane analogues

The EC$_{50}$ values for I3C, DIM and compound **9** in Table 20 show some convincing trends. Despite the significantly reduced solubility of DIM compared to I3C and the suspected mechanism of uptake of I3C discussed in the introduction, the isolated active compound has significantly higher anticancer activity than its prodrug. It is worth noting that at the highest concentrations tested, solubility issues were encountered, however solubility issues can be rectified later on in the drug development process.

What is particularly interesting is that compound **9** has higher activity than even DIM. It was initially assumed that the role of the 2-phenyl group in compound **5** was to aid in the formation of the diindolylmethane metabolic product. However, because compound **9** has higher activity than DIM itself, this shows that the 2-phenyl group has a beneficial effect on the active form of the drug as well.

What is also noteworthy is the considerable improvement of both diindolylmethanes (but especially compound **9**) against the chemoresistant IN1760 cell line. This is of particular interest as it suggests that this class of compounds not only have reasonable activity against the established cell line U251, they may also have a broad spectrum of activity against otherwise chemoresistant cells.

It is apparent that due to the higher activity and improved stability of diindolylmethanes compared to their prodrug forms, future work in this area should focus exclusively on the active diindolylmethane form of the drug, and that further work into the prodrug form would be redundant.

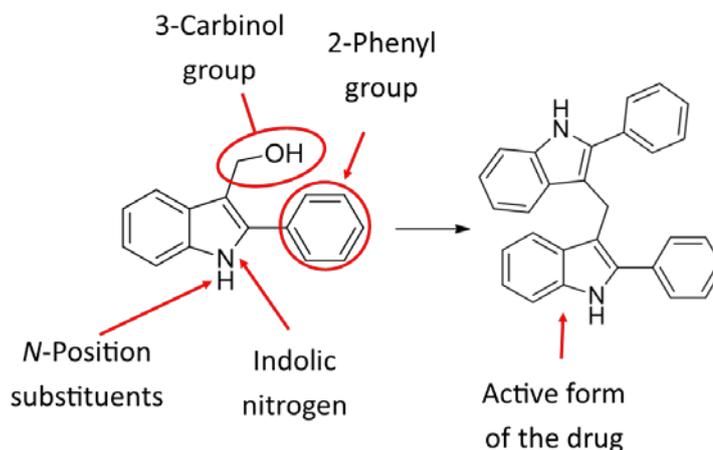*7.4.1.5 Concluding statements on the SAR of the I3C and compound 5 series*

*Figure 81 - The structural features of compound **5** that were investigated as part of this work*

A total of five structural features were investigated as part of this work, which are summarised in Figure 81. Unsurprisingly, the most vital structural feature of the prodrug form was the 3-carbinol group, without which the active form of the drug could not form. The nitrogen of the indolic scaffold also appears to be important, as replacing the indole scaffold with a benzothiophene scaffold, as in compound **134**, drastically reduced the activity of the compound against certain cell cultures. *N*-substituted analogues of the prodrug appear to be a viable way of affecting a significant change in activity, with the *N*-tosyl substituent (compound **133**) appearing to improve activity, as has also been shown by Weng *et al.*[286]

Comparing the activity of prodrug **5** with that of the active form **9** shows no obvious benefit to using a prodrug in future, and indeed showed significantly higher activity when administering the isolated metabolite. The prodrugs are also difficult to store and difficult to synthesise due to their inherent instability.

The 2-phenyl group consistently showed an improvement in activity both in the prodrug and active form of the drug. This structural feature should be included as standard in any future work.

Overall, the most promising compound of this series is compound **9**. If further work is to go into this series of compounds, this compound should be used as a lead compound. As *N*-substituents have been shown to have a significant effect on the activity of I3C and compound **5**, *N*-substituted analogues of this diindolylmethane should be investigated, and a potential unsymmetric synthesis of such compounds is outlined in Section 4.4.2.

7.4.2   The Structure-Activity Relationship of the Compound 4 Series

Structural analogues for compound **4**, which are summarised in Figure 82 alongside the major structural features of compound **4**, were considered partially due to chemical intuition and partially from the findings of the similarity search carried out in Chapter 5.   The structural features under investigation include structural isomers, the necessity of conjugation, the role of the phenolic ring, the role of the indole and the necessity of the 2-phenylindole core.   Many of the structural features addressed in this work are of interest because of the suspected role of the phenolic radical in the activity of compound **4**.

Figure 82 - The structural features and analogues of compound 4

7.4.2.1 Comparing the anticancer activity of compound 4 to that of a range of phenols
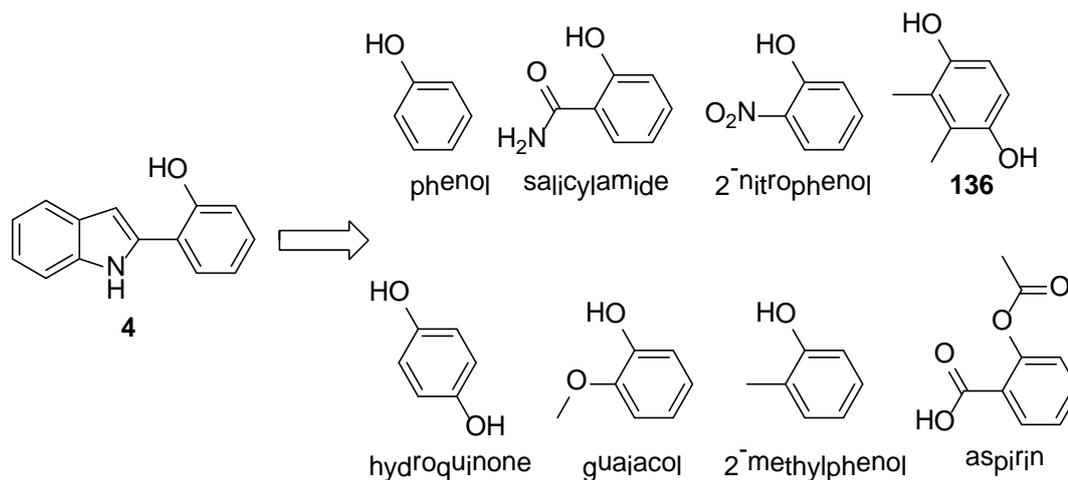

Figure 83 - The series of phenols tested in order to establish the importance of the indole group

Compound **4** can be considered as an *ortho*-substituted phenol, with an indole group at the *o*-position. To investigate the necessity of such a large substituent, the anticancer activities of a series of phenols (Figure 83) were determined.

Compound	U87	U251	IN1472	IN1528	IN1760
phenol	930	2420	1990	1000	7350
salicylamide	850	2420	1430	1190	8660
2-nitrophenol	950	2270	190	1500	6560
2-methylphenol	740	1960	1530	1040	3800
guaiacol	570	4230	1520	1170	1660
aspirin	780	3240	1640	1320	2910

Table 21 - Comparison of EC$_{50}$ values (μM) between a series of simple o-substituted phenols

Initially, the anticancer activities of a series of simple *ortho*-substituted phenols was determined, and this data is shown in Table 21. Comparing the activities of individual compounds against individual cell cultures generally shows that there isn't much variation in EC$_{50}$ values (with the exception of 2-nitrophenol against the IN1472 cell line), however 2-methylphenol and guaiacol tend to have the generally higher activities.

Compound	U251	IN1528	IN1760
4	300	760	100
phenol	5110	5170	1170
2-methylphenol	3110	1550	4500
aspirin	1110	5970	1200
hydroquinone	220	67540	1310
136	-	-	-

*Table 22 - Comparison of EC_{50} values (μM) between compound **4**, some simple o-substituted phenols and two hydroquinones*

A subset of those compounds tested in Table 21 were then retested, as well as two hydroquinones (although compound **136** was found to be insoluble under testing conditions). This data is shown in Table 22 alongside data for compound **4**.

Comparing the phenols with hydroquinone, we see that hydroquinone has far higher activity against the U251 cell line, comparable activity against the IN1760 cell line, and almost no activity against the IN1528 cell line. This inconclusive data, coupled with the inconclusive data for the insoluble compound **136**, means that no conclusion can be drawn for the relative activities of phenols and hydroquinones. As mentioned in Chapter 4, however, hydroquinones are known to be involved in redox cycling,[228,229] so should still be considered in future work.

Comparing the activity of the phenols with that of compound **4**, it is obvious that against all three cell cultures, compound **4** has significantly higher anticancer activity. The magnitude of the improvement in activity suggests that either the *ortho*-indole group is an excellent substituent to the phenol group, or that there is a unique and perhaps synergistic effect between the indole and phenol moieties.

Literature precedents for simple phenols with anticancer activity are difficult to come by. While there are many examples of more complex phenols, polyphenols and hydroquinones having reasonable anticancer activity,[288–292] the anticancer activities of simple phenols seem underreported, suggesting that they have generally been found to have poor activity, and the data has therefore gone unpublished. As is in agreement with what has been found in this project, there is some evidence that simple phenols can have some anticancer activity, although increasing the complexity of the phenols seems to significantly improve activity.[293,294]
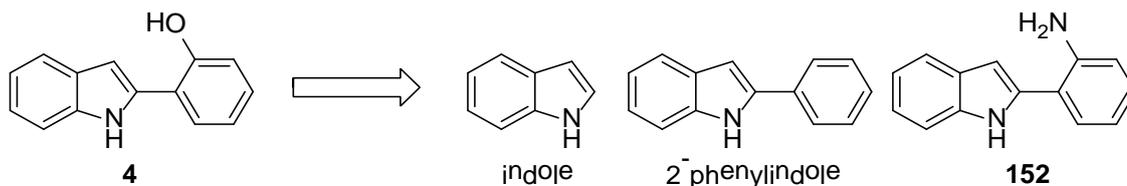
7.4.2.2 Determining the importance of the phenolic group in compound 4

*Figure 84 - The analogues used to probe the importance of the phenolic group*

In order to determine the significance of the phenolic group of compound **4**, a series of close analogues were developed. These analogues included an analogue without the phenol group (indole), an analogue without the phenolic OH group (2-phenylindole) and an aniline analogue (compound **152**).

Compound	U87	U251	IN1472	IN1528	IN1760
4	233.1	337.7	332.4	497.1	465.6
152	354.6	537.5	527.1	690.9	4791.1
indole	-	-	-	-	-
2-phenylindole	-	-	-	-	-

Table 23 - Comparison of EC$_{50}$ values (μM) between compound 4 and some analogues probing the significance of the phenolic group.

Upon removal of either the phenolic OH group or the whole phenolic substituent, the EC$_{50}$ value rises to above what can be determined under the assay conditions employed. Although the extent of activity loss cannot be determined, it can be confidently asserted that the phenolic system is essential for anticancer activity.

Comparing compound **4** with its aniline analogue **152** across all five cell cultures that they were tested on shows compound **4** having consistently higher activity, although both compounds have reasonable activity. Normally, this would result in other similar analogues such as thiophenols or phenylphosphines (Figure 85) becoming compounds of interest, however, such compounds are uncommon pharmaceutical agents.
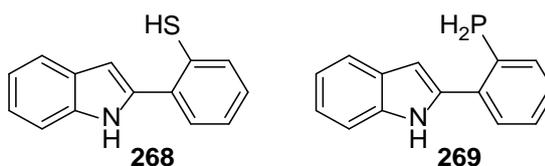
Figure 85 - Two possible future compounds of interest, based on the activity of compound 4 and compound 152

7.4.2.3 Determining the importance of a specific scaffold via scaffold hopping
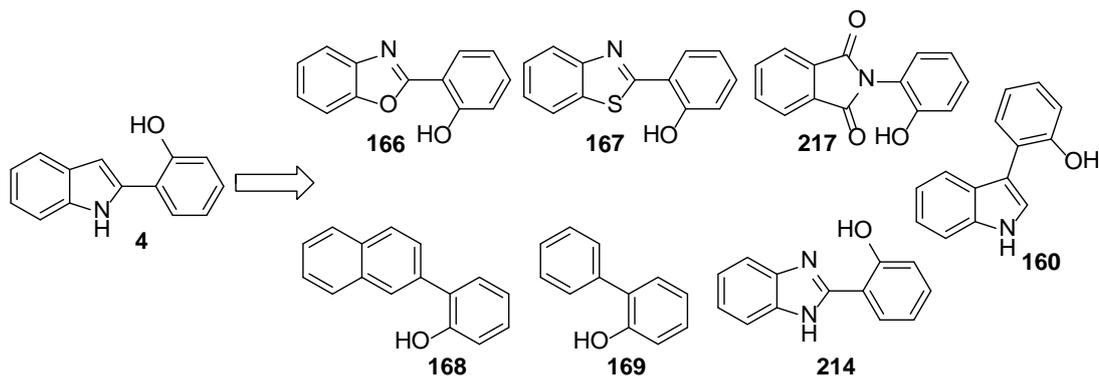
*Figure 86 - The analogues of compound **4** with different scaffolds*

The five analogues of compound **4** developed by scaffold-hopping, shown in Figure 86, were included in this work in order to potentially discover new novel drug fragments for future study within the Snape group.

Compound	U87	U251	IN1472	IN1528	IN1760
4	110	180	1360	650	680
166	140	1670	-	-	1670
167	860	4210	640	3880	1110
168	20	70	2390	20	340

Compound	U251	IN1528	IN1760
4	60	270	60
169	350	210	370
160	40	300	90
217	680	530	1110
214	50	20	20

*Table 24 - Comparison of EC_{50} values (μM) between compound **4** and a series of analogues with different scaffolds*

Compound **4** was tested against a naphthalene analogue (compound **168**), a benzoxazole analogue (compound **166**) and a benzothiazole analogue (compound **167**) as one series, and against a phenyl analogue (compound **169**) a 3-phenylindole analogue (compound **160**), a 2-phenylbenzimidazole analogue (compound **214**) and a 1-phenylphthalimide analogue

(compound **217**) in a separate series (Figure 86). The EC$_{50}$ values for all of these comparisons are shown in Table 24.

Compared against the benzoxazole analogue **166**, compound **4** has higher activity across all five cell cultures. Compared against the benzothiazole analogue **167**, compound **4** has higher activity against four of the five cell cultures, with the benzothiazole having a lower EC$_{50}$ against IN1472 than compound **4**. This seems to indicate that the 2-phenylindole core is better than the highly similar benzoxazole and benzothiazole cores in this class of antiglioblastoma compounds.

Comparing compound **4** against its naphthalene analogue (compound 168) shows the naphthalene analogue having higher activity against four of the five cell cultures, with compound **4** having higher activity against only IN1472, indicating that the naphthalene analogue appears to be the more active compound. This is very telling, as it indicates that the weakly hydrogen bonding NH group of compound **4** is not essential for anticancer activity, and that the role of the scaffold may be limited to such effects as sterics and/or π-stacking.

Compound **4** was compared against compounds **169** and **160** on three cell cultures. Comparing compound **4** with compound **169**, compound **4** was found to have superior activity against the U251 and IN1760 cell cultures, whereas compound **169** was found to have superior (albeit close) activity against the IN1528 cell line.

Compound **4** and compound **160** were compared as these isomers may give information about the importance of shape. Their electronics are expected to be fairly similar, as they have the same number of canonical radical forms, as shown in Figure 87. These isomers were found to have very similar activities against all three cell cultures on which they were tested, with compound **4** being slightly more active against the U251 and IN1760 lines, and compound **160** being more active against the IN1528 line.
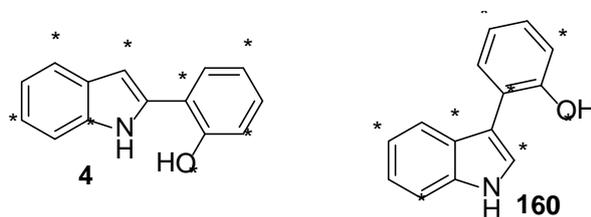

*Figure 87 - The possible places that a phenolic radical could become delocalised to, identified by an asterisk (*) on compounds **4** and **160***

Compounds **217** and **214** were both found through the similarity search in Chapter 5, however, they show drastically different activities. Compound **217** has drastically reduced activity against

all three cell cultures, whereas compound **214** has the highest activity of all compounds it was tested alongside. This may indicate that specific scaffolds can have activity against cultured glioblastoma cells. Even if activity is not occurring *via* the same mechanism, the process of scaffold hopping, if it yields compounds of higher activity than the lead compound, will still have been worthwhile.

Considering all of this data together, it would seem that the activity of compound **4** is not highly sensitive to 3D shape, as shown by the significantly different shapes of compounds **4** and **160**. This conclusion is derived from the chemically dissimilar compounds such as compound **214** and compound **168** which have higher activity than compound **4**.

7.4.2.4 Reducing the conjugation of compound 4
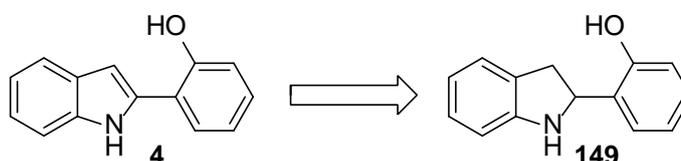
*Figure 88 - The analogue of compound **4** with reduced conjugation*

In order to investigate the necessity of conjugation between the proposed phenol radical and the indole moiety, an analogue without the conjugation between the two ring systems, compound **149** (shown in Figure 88), was tested. As discussed in Chapter 4, a suspected issue with this analogue is that its 3D shape is somewhat different to that of compound **4** due to the loss of planarity inherent in reducing the conjugation of this lead compound. However, as has been shown by the analogues in Figure 86, anticancer activity within this class of compounds is not highly sensitive to 3D shape, so this issue may be alleviated.

Compound	U251	IN1528	IN1760
4	60	270	60
149	7	420	150

*Table 25 - Comparison of EC$_{50}$ values (μM) between compound **4** and its reduced analogue **149***

The EC$_{50}$ values in Table 25 shows that compound **4** has the higher activity against both short-term cultures, whereas compound **149** has higher activity against the established U251 cell line. Therefore, determining the superior antiglioblastoma agent is difficult based solely on the data in Table 25. What can be stated is that the high level of conjugation in compound **4**, which was originally considered to be essential for its activity, is perhaps not the only factor involved in promoting anticancer activity. This lack of conjugation in compound **149** is also reflected in its

BDE value, which is higher than that of phenol (BDE of compound **4** is 77.1 kcal mol^{-1}, of phenol is 87.7 kcal mol^{-1}, and of compound **149** is 90.2 kcal mol^{-1}). Again, this suggests that activity is not directly related to the stability of the radical or the ease with which the radical is formed (i.e. the BDE value).
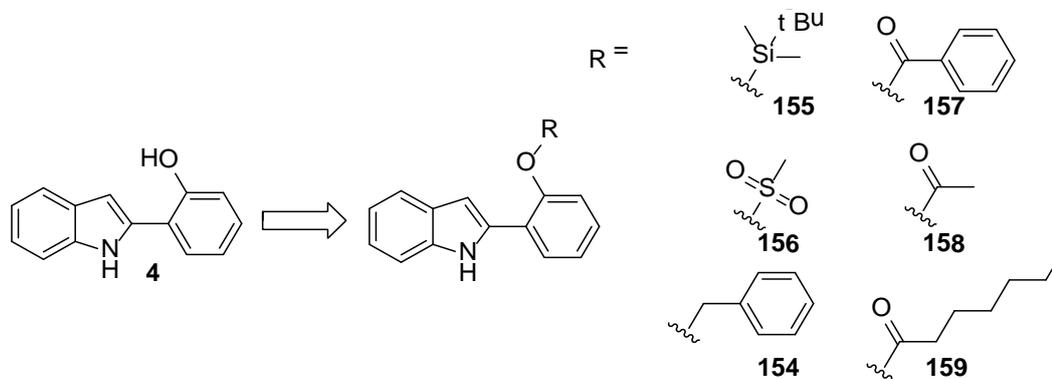
7.4.2.5 The effect of O-protected analogues

*Figure 89 - O-protected analogues of compound **4***

As discussed in the introduction, compound **4** is known to act rapidly. Compound **4** is also thought to act *via* a phenolic radical. The series of compounds shown in Figure 89 are intended to investigate these two factors.

Figure 89 includes three compounds that would not be expected to be easily hydrolysed under physiological conditions (the TBS analogue **155**, the mesyl analogue **156** and the benzyl analogue **154**), which will probe the activity of analogues that cannot readily form a phenolic radical, and three analogues which would be expected to be hydrolysed under physiological conditions (the benzoyl analogue **157**, the acetyl analogue **158** and heptanoyl analogue **159**), which are intended to modulate the activity of compound **4**.

Compound	U251	IN1528	IN1760
4	60	70	30
154	670	250	6670
155	40	60	10470
157	50	40	440
159	40	100	310
156	60	80	30
158	30	120	10

Table 26 - Comparison of EC_{50} values (μM) between compound **4** and a series of O-substituted derivatives

The activity data for the compounds shown in Figure 89 is shown in Table 26. This data shows that the three analogues that were expected to have a loss of activity due to suspected irremovable substituents, compounds **154**, **155**, and **156**, do not have appear to act as expected. While the O-benzyl analogue **154** has significantly lower activity against all three cultures on which these compounds were tested, the O-TBS analogue **155** shows similar activity to compound **4** on the U251 and IN1528 cultures, however it shows a loss of activity against the IN1760 culture, and the O-mesyl analogue **156** shows comparable activity to compound **4** against all three cell cultures.

As discussed in Section 4.3, enzymatic cleavage of Si-O bonds is uncommon but known,[221,222] so the enzymatic removal of TBS group of compound **155** is possible. An alternative possibility is that compound **155** is an active agent in its own right. Silicon-based derivatives of existing drugs have been shown to retain and in some instances improve activity,[295] potentially due to the high lipophilicity (and therefore improved membrane permittivity) of silicon residues.[296,297]

Of the three O-substituted analogues that were expected to be hydrolysable, compounds **157**, **158** and **159**, all three have comparable activity to compound **4** against the U251 and IN1528 cultures. The O-acetyl analogue **158** also has comparable activity against IN1760, with the O-benzoyl analogue **157** and the O-heptanoyl analogue **159** having lower activity against IN1760.

The comparable activity of the O-acyl analogues **157**, **158** and **159** to compound **4** was expected due to deacylation being a common drug biotransformation.[298] What is more difficult to justify is the activities of the other three analogues. There are known enzymes that cleave benzyl groups,[299] and it stands to reason that there are potential metabolic pathways that involve the cleaving of TBS or mesyl groups. The alternative explanations are either that compounds **154**,

155, and **156** are acting *via* a different mechanism, or that by derivatising at the *O*-position does not necessarily form an inactive compound. The latter of these potential explanations would considerably alter the basic assumptions on the mechanistic process of compound **4**. A further possibility is that these *O*-substituted compounds, which may be expected to act as prodrugs, are simply being metabolised differently in different cell cultures. Genetic mutations in tumours are now known to significantly affect cell signalling, resulting in drastically different metabolic potential between tumours,[300] so these unexpected trends may simply be a result of different cell cultures being able to metabolise these potential prodrugs differently.

The comparative activity of compound **4** against its *O*-methyl derivative (Figure 90) has been previously carried out by Prabhu *et al.*,[6] who found that the *O*-methyl derivative was inactive against the U87 cell line.
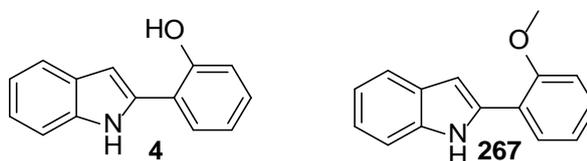
*Figure 90 - Compound **4** (left) and its* O*-methyl derivative **267** (right), as tested by Prabhu et al.*[6]

To investigate the possibility of *O*-protected analogues of compound **4** being used as prodrugs for compound **4**, time course assays for both compound **4** and its *O*-heptanoyl analogue **159** were carried out within the research group by Saurabh Prabhu using the MTS assay. The results from these assays, which were carried out on the primary glioblastoma culture BTNW911 and the short-term human glioma cell culture IN859, are shown in Graph 5. These data shows that compound **4** has rapid initial activity against both cultures, whereas compound **159** reduced cell viability much more gradually. Importantly, both compounds tend towards the same activity over time. This would agree with the hypothesis that the heptanoyl group can be cleaved off to produce compound **4**, as the activity of compound **159** is equal to that of compound **4**, yet it takes time for the enzymatic cleavage to occur.
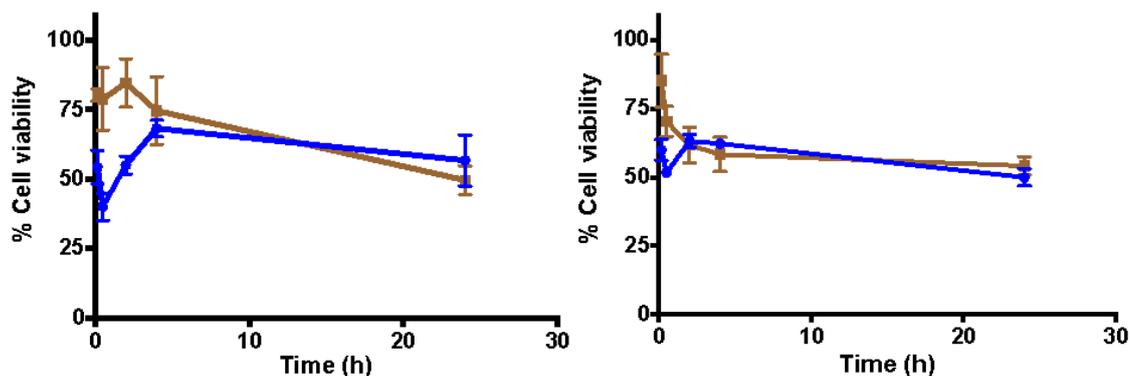
Graph 5 - Time course graphs comparing the activity of compound 4 (blue) against compound 159

(brown) against BTNW911 (left) and IN859 (right)

7.4.2.6 Miscellaneous analogues
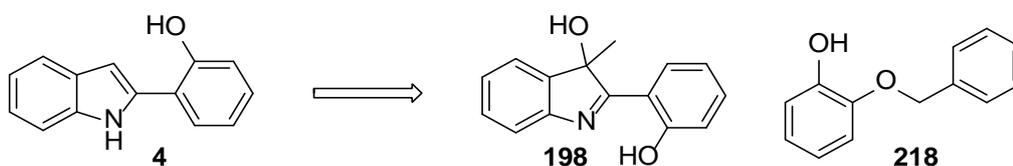
Figure 91 - Miscellaneous analogues of compound 4

Finally, comparisons between compound **4** and its two miscellaneous analogues can be drawn. Compound **198** was produced as a degradation product of a compound of interest, as described in Chapter 4, and is chemically dissimilar to compound **4** despite being structurally similar. Compound **218** was found during the similarity search in Chapter 5, and is both chemically and structurally dissimilar, yet was considered by the ShaEP software to be highly similar in terms of both shape and electronics of the system.

Compound	U251	IN1528	IN1760
4	60	270	60
198	110	120	50
218	6	70	340

Table 27 - Comparison of EC$_{50}$ values (μM) between compound 4 and two miscellaneous analogues

Interestingly, both of these analogues show some reasonable activity against at least one of the cultures on which they were tested, although clear trends are difficult to elucidate. Compound **198** has higher activity than compound **4** against the IN1528 short-term cultures, comparable activity against the IN1760 short-term culture and worse activity against the U251 cell line. Compound **218** has drastically higher activity against both the U251 and IN1528 cultures, yet

drastically lower activity against the IN1760 culture. Based solely on this data, it is difficult to rank these three compounds due to their ambiguous trends.

Both of these analogues are, to the eye of a chemist, considerably different to the lead compound **4**. Compound **198**, despite being fairly similar in terms of shape, is chemically far removed from the lead compound, whereas compound **218** is little more than a substituted simple phenol. Regardless, both of these compounds showed reasonable activity against at least one culture on which they were tested. This suggests one of two things; either these active compounds have comparable activity due to their ability to assume a comparable 3D shape, thus being able interact with a protein or enzyme in the same way in which compound **4** does, or that these active compounds are not functional analogues of one another and that their activities is due to different mechanisms of action.

7.4.2.7 Concluding statements on the SAR of compound 4
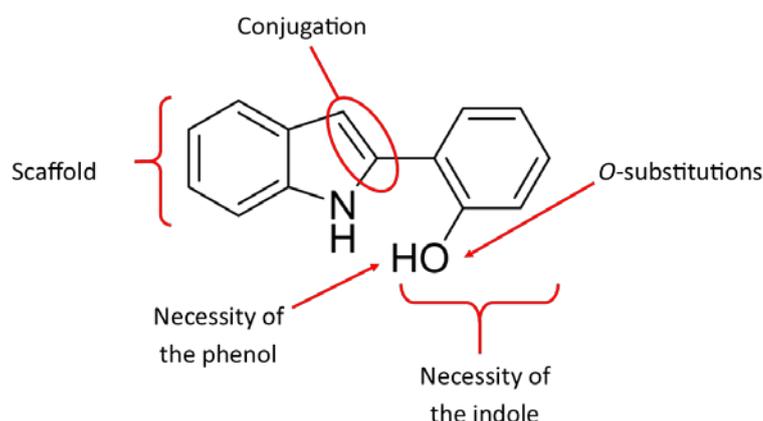
*Figure 92 - The structural features of compound **4** that were investigated as part of this work*

There were five structural features of compound **4** that were investigated as part of this work, which are summarised in Figure 92. It was found that a phenolic (or analogous) group was required for activity, as removal of this group resulted in a loss of activity. It was also shown that simple *ortho*-substituted phenols had much lower activity than compound **4**, suggesting that the indole moiety has significant importance, as suggested in Section 2.1 where indole was described as a privileged structure.

Investigating different scaffolds yielded some interesting findings. It was found that the 2-phenylnaphthalene scaffold of compound **168** improved activity, indicating that the NH group of the indole is not necessary for activity. Other scaffolds were also shown to have activity, although the majority of the scaffolds tested had little activity.

The effect of *O*-substituted analogues showed lower activity for some compounds and comparable activity for others (with the possible exception of the *O*-acetyl derivative **158**). The meaning of these findings are inconclusive, as it may indicate that the substituents are being cleaved *in vitro* or that the OH group of compound **4** is not as essential for activity as once thought.

This work has shown through scaffold hopping that compounds **168** and **214** (shown in Figure 93) are the superior fragments on which to base the design of new drugs around, and future work should therefore utilise these compounds as lead compounds. A significant breadth of chemical space has been explored so far within this work, however, the local chemical space of any of the active functional analogues of compound **4** has not been probed to any appreciable depth. Future work should therefore focus on refining and developing the structures of compounds **168** and **214** in order to further improve their activities.
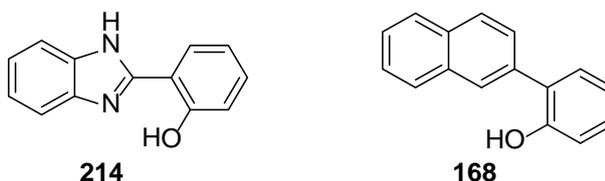
*Figure 93 - The structures of compounds **214** and **168***

Initially, it should be confirmed that compounds **168** and **214** act in the same way as compound **4**, in that they induce an increase in ROS levels in a cell, followed by apoptosis. If their mode of activity is found to be the same as compound **4**, the next step should be to investigate induceing elevated ROS levels, such as interacting with proteins or enzymes that can upregulate ROS production or downregulate antioxidant production.

7.5 Comparing anticancer activity to BDE

As was discussed in Chapter 6, the BDE values (as well as AIP, PDE, PA and ETE values) were calculated for a number of analogues of compound **4**. Because compound **4** is known to induce the rapid generation of ROS followed by cellular death,[6] the intention behind calculating these enthalpic values was to determine if there might be a connection between the ease of formation of these phenolic (and analogous) radicals and their anticancer activities. As phenols are known to be oxidised by molecular oxygen, thus generating ROS,[301,302] it was suspected that this reaction may be responsible for the increase in ROS observed upon addition of compound **4** to cultured cells.

In order to compare EC_{50} values to BDE values across a number of batches of anticancer tests, EC_{50} values were compared against BDE values within each batch of testing, and the trend lines within each batch were compared for similarity. These comparisons are given in Appendix 7, and information on which batches contain which compounds can be found within this appendix. A comparison between EC_{50} values and BDE values against the U251 cell line is given as a representative example in Graph 6. As can be seen from this graph, there is very poor agreement between the trend lines for each batch of compounds, and even the extent to which each batch of compounds conforms to its trendline is poor. This lack of agreement is also true for the other cell cultures on which the compounds were tested.
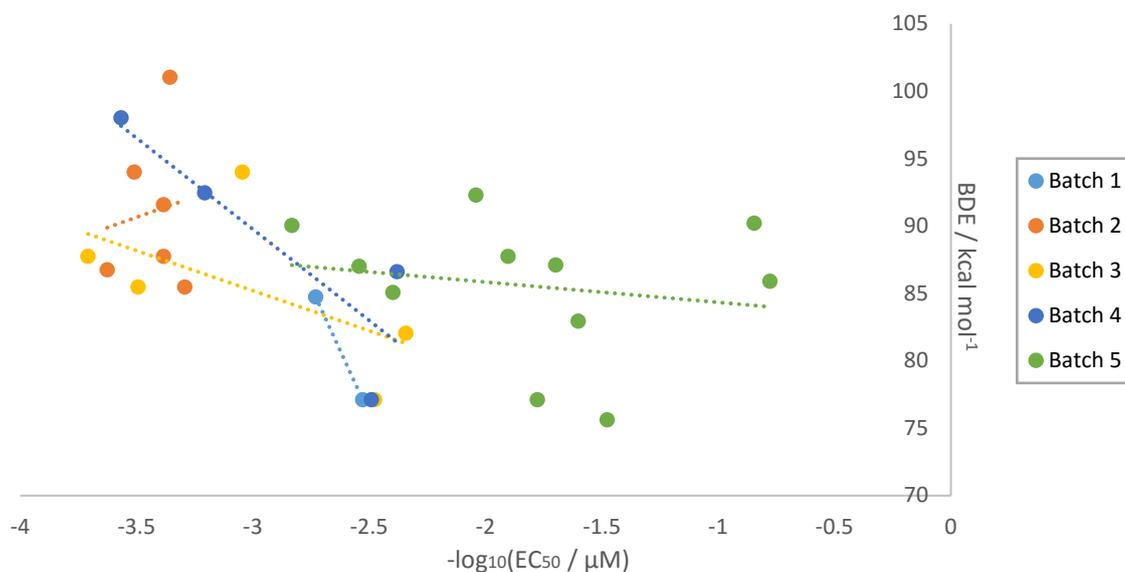
Graph 6 - Combined data for all six batches of compounds, comparing EC_{50} values for the U251 cell line (μM) against BDE values (kcal mol^{-1})

In the interest of completeness, comparisons were also made between EC_{50} values and the other four enthalpic values calculated in Chapter 6 (AIP, PDE, PA and ETE), but no significant correlation was found in any of these cases. A final attempt was made to find a correlation between anticancer activity and one of the calculated enthalpic values by normalising all EC_{50} values to that of compound **4**, then plotting this data as one series against each enthalpic value. The strongest correlation was found between anticancer activity and BDE against the IN1528 culture (Graph 7), although with an R^2 value of just 0.45, even this is not a strong enough correlation to be considered significant.
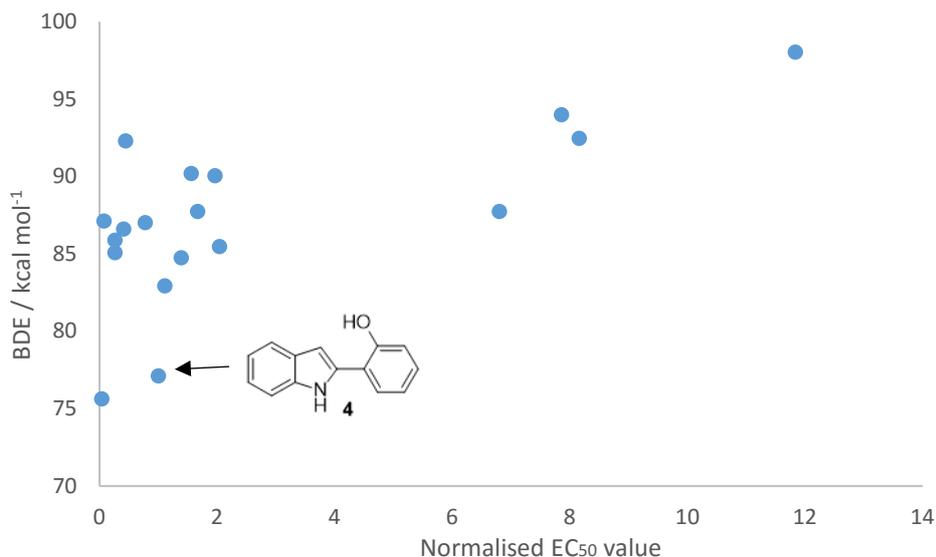
Graph 7 - Normalised dosage data for the IN1528 short-term culture plotted against calculated BDE values (in kcal mol^{-1})

These data shows that this is a poor overall correlation between the observed anticancer activity and the calculated enthalpic values for the entire data set, but there are some other revealing trends. On examination of the normalised data, it is apparent that particularly low BDE, AIP and PA values against the U251 and IN1528 cultures are consistently correlated with a higher activity, although no such trends are observed for the chemoresistant IN1760 culture. These observations are for both compounds with a BDE below 80 kcal mol^{-1} (compounds **4** and **244**), the five compounds with an AIP below 180 kcal mol^{-1} (compounds **4**, **152**, **244**, **160** and **149**) and both compounds with a PA of below 80 kcal mol^{-1} (compounds **4** and **244**), the structures of which are shown in Figure 94. It may, therefore, be possible that with a larger data set, a similar trend may emerge whereby particularly low enthalpic values can be used as a predictor of reasonable anticancer activity, although this approach would inherently miss a large number of active compounds with higher enthalpic values such as compound **168** (although false positives or negatives are inherent with every *in silico* drug design approach). At the moment, this trend is too weak to be considered significant, and the compounds in Figure 94 may simply have been identified due to being part of a highly similar chemical class.
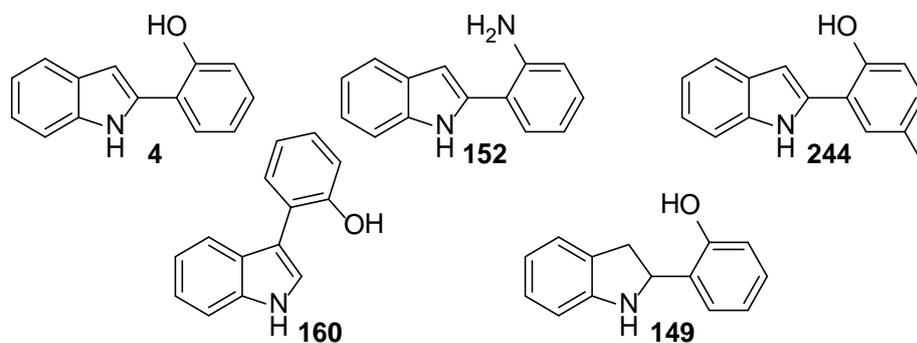
Figure 94 - The structures of compounds which appeared below the "threshold level" of at least one calculated enthalpic value, as described in the text

Therefore, within this project, a strong and clear correlation between the propensity of this class of compounds to form radicals and anticancer activity has not been identified. As with all inconclusive data, this can indicate one of two things. The first thing that this could indicate is that the hypothesis was simply incorrect. This would therefore indicate that the phenolic radicals formed by the class of compounds investigated are unlikely to form ROS by direct means (such as by directly reducing molecular oxygen to superoxide). An alternative hypothesis would therefore be that compounds such as compound **4** either upregulate the production of ROS or downregulate the production of antioxidants.

The second thing that this inconclusive data could indicate is that the hypothesis was not suitably investigated due to insufficient data. If it is a case of insufficient data, then more experimental data is required in order to establish whether a calculated enthalpic value such as BDE is truly useful predictor of anticancer activity, however, the possibility of a more restricted hypothesis (e.g. particularly low values of BDE, AIP and/or PA indicate reasonable activity within a class of compounds) may still be plausible. On the other hand, a limit of this investigation could be that the analogues investigated are too structurally dissimilar to have their activities directly compared, and therefore a series of more structurally similar compounds based around the same scaffold (for example the 2-phenylindole or 2-phenylnaphthalene scaffold) would be better suited to investigating this hypothesis.

Considering this, the data was then revisited with the intention of analysing a "single chemical class", and investigating if any trends between activity and enthalpic values was present. The largest chemical class that was investigated as part of this work were the six simple *ortho*-substituted phenols, so their activities (i.e. their $-\log_{10}(EC_{50})$ values) were compared against the five calculated enthalpic values (BDE, AIP, PDE, PA and ETE), with the trendline equations and R^2 correlation value for each comparison shown in Table 28. Excluding the U251 cell line

($R^2 = 0.394$), reasonable to good correlations were obtained between certain enthalpic values and anticancer activity. This data shows that the PDE values correlate best with $-\log_{10}(EC_{50})$ values against the U87 cell line ($R^2 = 0.677$), the BDE values correlate best with $-\log_{10}(EC_{50})$ values against the IN1472 ($R^2 = 0.700$) and IN1528 ($R^2 = 0.844$) culture, and the PA values correlate best with $-\log_{10}(EC_{50})$ values against the IN1760 culture ($R^2 = 0.548$). This is in agreement with the findings from Chapter 6, which suggested that PDE values are more likely to be larger (and therefore the rate-limiting factor) than AIP values in the SET-PT radical formation pathway, and that PA values are more likely to be higher (and therefore the rate-limiting factor) than ETE values in the SPLET radical formation pathway, as shown in Scheme 8 in Section 1.2.2.

		U87	U251	IN1472	IN1528	IN1760
BDE	Trendline	y = -40.0x - 24.8	y = 6.10x +112.0	**y = 12.7x +123.0**	**y = -81.8x - 160.6**	y = -7.34x +64.3
	R^2	0.315	0.016	**0.700**	**0.844**	0.121
AIP	Trendline	y = -100.0x - 83.4	y = 37.0x + 333.1	y = 19.8x +266.9	y = -94.9x - 85.3	y = -19.3x +136.1
	R^2	0.659	0.198	0.564	0.379	0.281
PDE	Trendline	**y = 60.1x +386.3**	**y = -30.8x +106.5**	y = -7.04x +190.7	y = 13.0x +252.3	y = 12.0x +255.9
	R^2	**0.677**	**0.394**	0.204	0.020	0.308
PA	Trendline	y = 56.2x + 507.9	y = -30.6x +240.3	y = -10.5x +312.8	y = 57.1x +520.5	**y = 19.2x +414.9**
	R^2	0.412	0.269	0.317	0.271	**0.548**
ETE	Trendline	y = -96.2x - 205.1	y = 36.7x +199.5	y = 23.3x +144.8	y = -138.9x - 353.5	y = -26.5x - 22.9
	R^2	0.438	0.141	0.561	0.583	0.380

Table 28 - Trendlines and R^2 correlation values for the comparisons of $-\log_{10}(EC_{50})$ values against calculated enthalpic values. The data in red bold text indicates the highest R^2 value for that particular cell culture

The gradient values shown in Table 28, may also be of significance. Considering again the data shown in bold in Table 28 (U251 data excepted), against the U87, IN1472 and IN1760 cultures, the gradients are positive, indicating that as the phenolic radicals become easier to form, their activity increases, and *vice versa* for the IN1528 short-term culture, which has a negative

gradient. This may be indicative of the previously mentioned "threshold theory", whereby the U87, IN1472 and IN1760 cultures may have mutated to have elevated ROS levels, and are therefore easier to kill by inducing even higher levels of ROS. Conversely, the U251 and IN1528 cultures may be more resilient to ROS if these cells have mutated to have lowered ROS levels.

In conclusion, this data suggests that a correlation between radical formation and activity may still in fact be present, but only within a more specific chemical class. In order to improve on the activity observed in this series of simple *ortho*-substituted phenols, future work could include a similar series of close analogues, and attempt to replicate the observed potential correlations shown in Table 28, but based on Compounds **168** or **214** (Figure 95), which would be ideal lead compounds for such a series.
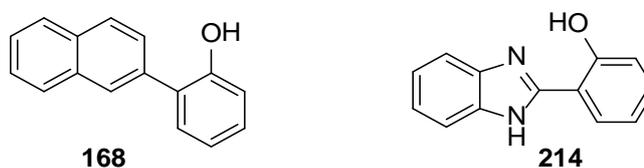
*Figure 95 - The structures of compound **168** and **214***

CHAPTER 8 – EXPERIMENTAL

8.1   Overview of Materials and Methods

Reactions were followed by analytical thin layer chromatography (TLC) using plastic-backed TLC plates coated in silica G/UV$_{254}$, run in a variety of solvent systems and visualised with a UV light at 254 nm, *p*-anisaldehyde stain and/or potassium permanganate stain. Commercially available solvents and reagents were purchased from Fisher, Sigma Aldrich, TCI and Fluorochem and were used without further purification unless specified in the syntheses.   Flash column chromatography was carried out on Davisil silica 60 Å (40 – 63 μm) under bellows pressure. High resolution mass spectra were obtained at the EPSRC UK National Mass Spectrometry Facility in Swansea University's College of Medicine using a LTQ Orbitrap XL™ Hybrid Ion Trap-Orbitrap Mass Spectrometer coupled to a TriVersa NanoMate® ESI source. Low resolution mass spectra were obtained on a Thermo Finnigan LCQ Advantage MAX using electrospray ionisation (ESI) or atmospheric pressure chemical ionisation (APCI). Crystal structures were obtained at the EPSRC National Crystallography Service at the University of Southampton using a Rigaku AFC12 goniometer equipped with an enhanced sensitivity (HG) Saturn724+ detector mounted at the window of an FR-E+ SuperBright molybdenum rotating anode generator with VHF Varimax optics (70μm focus). ^{1}H and ^{13}C NMR were carried out on a Bruker Fourier 300 (300 MHz) or a Bruker Advance III 400 (400 MHz) with broad band decoupling, and all chemical shifts (δ) quoted in parts per million (ppm) relative to the residual solvent peaks of C*H*Cl$_3$ (δ$_H$ 7.26, δ$_c$ 77.16) or *d$_5$*-DMSO (δ$_H$ 2.50, δ$_C$ 39.52). *J* values are given in Hertz (Hz). Infrared spectra were recorded on a solid sample using a Thermo Nicolet IR-200 FT-IR. Melting points are uncorrected, and were recorded using a Stuart SMP10. Preparative liquid chromatography was carried out on a Teledyne Isco CombiFlash® Rf 200. Elemental analysis was carried out using a Thermo Scientific™ FLASH 2000 CHNS/O Analyser. Petroleum ether refers to the fraction that boils between 40-60 °C. Assignments of NMR spectra was aided with the use of DEPT-135, and in some cases HSQC and HMBC.

8.2   Syntheses

8.2.1   Procedure for the synthesis of 2-(1*H*-indol-2-yl)phenol (4)[6]
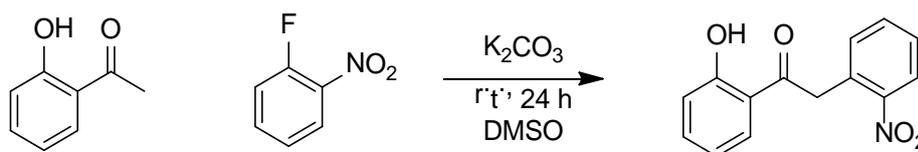
To a solution of 2'-hydroxyacetophenone (5 g, 36.7 mmol) and 1-fluoro-2-nitrobenzene (3.9 mL, 36.7 mmol) in DMSO (37 mL) was added potassium carbonate (12.7 g, 91.8 mmol). The reaction

was stirred at room temperature for 24 hours. Upon completion as indicated by TLC analysis, the reaction mixture was diluted with water (300 mL) and acidified to pH ca. 3 using 1 M HCl. The crude reaction was extracted with ethyl acetate (2 × 300 mL), and the organic layers were washed with water (2 × 300 mL) and brine (1 × 300 mL), dried (MgSO$_4$), and the solvent removed *in vacuo*. The crude product was purified by recrystallisation (ethyl acetate) to give 1-(2-hydroxyphenyl)-2-(2-nitrophenyl) as a dark green solid (3.60 g, 47% yield). R$_f$ = 0.35 (5:3 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 11.83 (1 H, br s, OH), 8.20 (1 H, dd, *J* = 8.0, 1.0 Hz, Ar), 7.91 (1 H, dd, *J* = 8.0, 1.5 Hz, Ar), 7.65 (1 H, td, *J* = 7.5, 1.5 Hz, Ar), 7.58-7.49 (2 H, m, Ar), 7.36 (1 H, dd, *J* = 7.5, 1.0 Hz, Ar), 7.04 – 6.94 (2 H, m, Ar), 4.80 (2 H, s, CH$_2$).
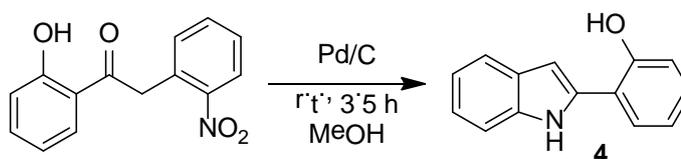
Methanol (50 mL) was charged into a flask containing Pd/C (232 mg, 18 wt%), followed by the addition of 1-(2-hydroxyphenyl)-2-(2-nitrophenyl)ethanone (1.26 g, 6.02 mmol). The flask was evacuated and backfilled with hydrogen, and the reaction was stirred at room temperature for 3.5 hours. Upon consumption of the starting material by TLC, the reaction was filtered through Celite® and washed with methanol (20 mL). The filtrate was concentrated *in vacuo*, and was purified *via* flash column chromatography on silica gel (10:3 petroleum ether: ethyl acetate) to afford the title compound as a pale yellow solid (0.70 g, 55% yield). R$_f$ = 0.18 (10:3 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, *d6*-DMSO): δ_H = 11.13 (1 H, br s, NH), 10.15 (1 H, br s, OH), 7.75 (1 H, dd, *J* = 7.5, 1.0 Hz, Ar), 7.51 (1 H, d, *J* = 7.5 Hz, Ar), 7.45 (1 H, d, *J* = 8.0 Hz, Ar), 7.14 (1 H, td, *J* = 7.5, 1.0 Hz, Ar), 7.05 (1 H, t, *J* = 7.5 Hz, Ar) 7.02 – 6.86 (4 H, m, Ar); MS (ESI): *m/z* 210 ([M+H]$^+$).

8.2.2 Procedure for the synthesis of 2(phenyl-1*H*-indol-3-yl)methanol (5)[6]
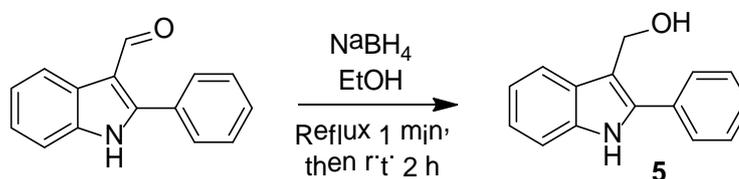
2-Phenylindole-3-carboxaldehyde (1.00 g, 4.52 mmol) and NaBH$_4$ (0.34 g, 9.04 mmol) were added to ethanol (34 mL) and heated at reflux for 1 minutes followed by stirring at room temperature for 2 hours. NaOH (1%, 35 mL) was added, and the solvent was removed *in vacuo*. The reaction was extracted with Et$_2$O (3 × 30 mL), dried (MgSO$_4$), filtered and the solvent

removed *in vacuo*. Due to the title compound being degraded on silica, the reaction was purified by recrystallization (EtOAc, few drops of PE) to yield the product as a colourless solid (187 mg, 19% yield). R_f = 0.21 (5:2 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, DMSO): δ_H = 11.35 (1 H, s, NH), 7.79 (2H, m, Ar), 7.66 (1 H, d, J = 7.5 Hz, Ar), 7.52 (2 H, t, J = 7.5 Hz, Ar), 7.35-7.44 (2 H, m, Ar), 7.12 (1 H, td, J = 7.5 Hz, 1.0 Hz, Ar), 7.03 (1 H, td, J = 7.5, 1.0 Hz, Ar), 4.96 (1 H, t, J = 5.0 Hz, OH), 4.67 (2 H, d, J = 5.0 Hz, CH$_2$); ^{13}C NMR (75 MHz, DMSO): δ_C = 135.94 (Ar, C$_q$), 135.82 (Ar, C$_q$), 132.41 (Ar, C$_q$), 128.74 (Ar, C$_q$), 128.71 (Ar, CH), 127.97 (Ar, CH), 127.57 (Ar, CH), 121.62 (Ar, CH), 118.94 (Ar, CH), 112.35 (Ar, C$_q$), 111.13 (Ar, CH), 53.86 (CH$_2$); MS (ESI): *m/z* 206 ([M-OH]$^+$).

8.2.3 Procedure for the synthesis of bis(2-phenyl-1*H*-indol-3-yl)methane (9)[303]
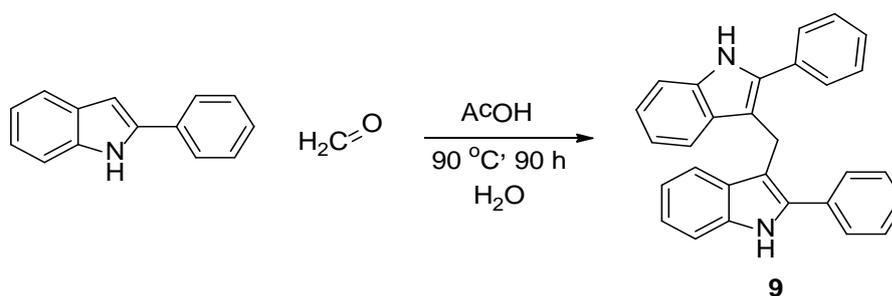
Under a nitrogen atmosphere and in the dark, 2-phenylindole (338 mg, 1.75 mmol), acetic acid (50.1 µL, 0.88 mmol) and formaldehyde solution (37 wt%, 70.1 µL) were stirred in water (40 mL) at 90 °C for 6 days. The crude reaction was extracted with ethyl acetate (2 × 15 mL), and the combined organic phases were dried (MgSO$_4$), filtered and concentrated *in vacuo*. The crude product was purified by preparative liquid chromatography (19:1 to 9:1 petroleum ether: ethyl acetate) to yield the title compound as a pale red solid (195 mg, 56%). R_f = 0.37 (9:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.01 (2H, br s, NH), 7.60 – 7.52 (4H, m, Ar), 7.45 – 7.18 (10H, m, Ar), 7.08 (2H, td, J = 7.5 Hz, 1.0 Hz, Ar) 6.86 (2H, td, J = 7.5, 1.0 Hz, Ar), 4.57 (2H, s, CH$_2$); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 135.99 (Ar, C$_q$), 134.60 (Ar, C$_q$), 133.41 (Ar, C$_q$), 129.46 (Ar, C$_q$), 128.79 (Ar, CH), 128.44 (Ar, CH), 127.61 (Ar, CH), 122.06 (Ar, CH), 120.12 (Ar, CH), 119.53 (Ar, CH), 112.14 (Ar, C$_q$), 110.63 (Ar, CH), 21.42 (CH$_2$); MS (APCI): *m/z* 397 ([M-H$_2$+H]$^+$); HRMS found: [M+H]$^+$ 399.1859, C$_{29}$H$_{22}$N$_2$+H$^+$ requires 399.1856.

8.2.4 Procedure for the synthesis of 2-(2-phenyl-1*H*-indol-3-yl)ethanol (129)[304]
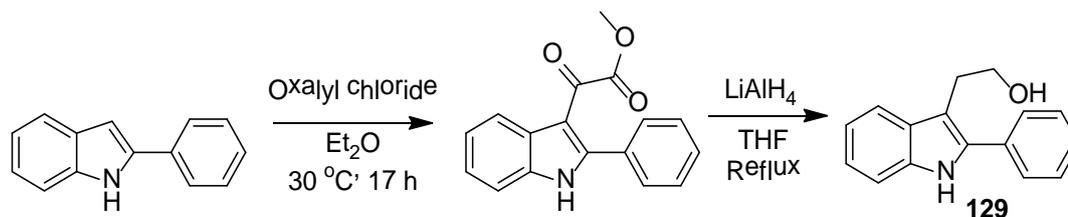
Oxalyl chloride (2.7 mL, 31.5 mmol) in dry Et$_2$O (10 mL) was added to 2-phenylindole (1.93 g, 10 mmol) in dry Et$_2$O (40 mL) dropwise at 0 °C under an atmosphere of nitrogen. The reaction was stirred at 30 °C for 17 h before being cooled to 0 °C and quenched with MeOH (2 mL). The reaction was concentrated *in vacuo* to yield methyl 2-oxo-2-(2-phenyl-1*H*-indol-3-yl)acetate, which was used in the next step of the reaction without further purification. To this crude product was added dry THF (15 mL) along with 4 Å molecular sieves (1 g). The supernatant was added dropwise to a suspension of LiAlH$_4$ (1.57 g, 41.4 mmol) in dry THF (40 mL) at 0 °C. The reaction was heated at reflux for 6 h before being quenched at 0 °C first by water (1.5 mL), then NaOH (10%, 3 mL) and finally water (4.5 mL). The reaction was filtered through Celite® and the solids were washed with MeOH. The filtrates were dried (MgSO$_4$), filtered and concentrated *in vacuo*. The product was purified *via* flash column chromatography on silica gel (5:3 → 1:1 petroleum ether: ethyl acetate) to afford a white solid (1.76 g, 74% yield). M.p. 89 – 91 °C; R$_f$ = 0.52 (5:3 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.15 (1 H, br s, NH), 7.70-7.60 (3 H, m, Ar), 7.48 (2 H, t, *J* = 7.5 Hz, Ar), 7.43-7.34 (2 H, m, Ar), 7.28-7.12 (2 H, m, Ar), 3.98 (2 H, t, *J* = 6.5 Hz, CH$_2$), 3.19 (2 H, t, *J* = 6.5 Hz, CH$_2$), 1.56 (1 H, br s, OH); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 135.92 (Ar), 135.88 (Ar), 132.84 (Ar), 129.16 (Ar), 128.95 (Ar), 128.17 (Ar), 127.91 (Ar), 122.50 (Ar), 119.86 (Ar), 119.16 (Ar), 110.93 (Ar), 108.87 (Ar), 63.02 (CH$_2$), 28.06 (CH$_2$); IR (neat, cm^{-1}) ν = 3396 (N-H stretch), 1450 (aromatic C-C stretch), 1037 (C-O stretch); MS (ESI): *m/z* 238 ([M+H]$^+$); HRMS found: [M+H]$^+$ 238.1227, C$_{16}$H$_{15}$NO+H$^+$ requires 238.1226.

8.2.5 Procedure for the synthesis of 1-(3-(hydroxymethyl)-1*H*-indol-1-yl)ethanone (130)[305]
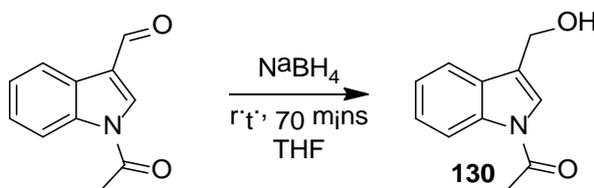
1-Acetyl-1*H*-indole-3-carbaldehyde (0.5 g, 1.9 mmol) and NaBH$_4$ (0.14 g, 3.7 mmol) were added to dry THF (20 mL), and the reaction was stirred at room temperature under a nitrogen atmosphere for 70 mins. Upon completion by TLC, the reaction was quenched with saturated

NH$_4$Cl (5 mL), and the solvent was removed *in vacuo*. The reaction was extracted with ethyl acetate (2 × 5 mL), and the combined organic layers were washed with water (10 mL). The organic layer was then dried (MgSO$_4$), filtered, and the solvent was removed *in vacuo*. The product was purified by flash column chromatography on silica gel (1:5 petroleum ether: ethyl acetate) to yield the title compound as a light orange-brown powder (0.18 g, 36% yield). M.p. 134 – 135 °C; R$_f$ = 0.50 (1:5 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, *d*6-DMSO): δ_H = 8.31 (1 H, d, *J* = 8.0 Hz, Ar), 7.71 (1 H, s, Ar), 7.69 - 7.62 (1 H, m, Ar), 7.36-7.22 (2 H, m, Ar), 5.16 (1 H, t, *J* = 5.5 Hz, OH), 4.65 (2 H, dd, *J* = 5.5, 1.0 Hz, CH$_2$), 2.62 (3 H, s, CH$_3$); ^{13}C NMR (75 MHz, *d*6-DMSO): δ_C = 169.37 (C=O), 135.42 (Ar, C$_q$), 129.40 (Ar, C$_q$), 124.77 (Ar, CH), 123.93 (Ar, CH), 123.17 (Ar, CH), 122.64 (Ar, C$_q$), 119.70 (Ar, CH), 115.92 (Ar, CH), 55.20 (CH$_2$), 23.86 (CH$_3$); IR (neat, cm^{-1}) *v* = 3501 (free O-H stretch, sharp), 1684 (C=O stretch), 1007 (C-O stretch); MS (EI): *m/z* 130 ([M-OH-Ac]$^-$, 100%), 147 ([M-Ac]$^-$, 43%); HRMS found: [M+H]$^+$ 190.0862, C$_{11}$H$_{11}$NO$_2$+H$^+$ requires 190.0863.

8.2.6 Procedure for the synthesis of 1-(3-(hydroxymethyl)-2-phenyl-1H-indol-1-yl)ethanone (132)
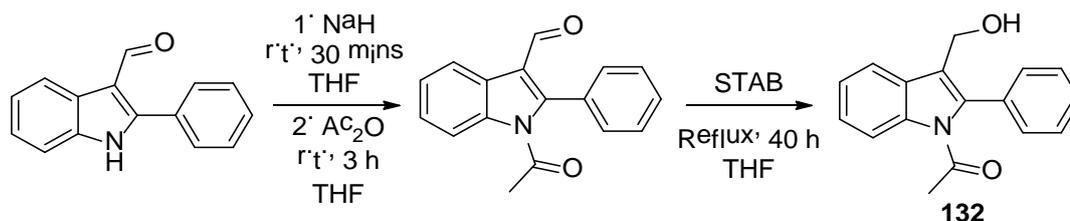
A solution of 2-phenylindole-3-carboxyaldehyde (1 g, 4.52 mmol) with sodium hydride (60% in oil, 0.36 g, 9 mmol) in dry THF (20 mL) was stirred at room temperature for 30 mins. Acetic anhydride (1.27 mL, 13.5 mmol) was added and stirred for a further 3 hours until the reaction was complete by TLC. The crude reaction was quenched with the addition of water (1 mL) dropwise. The reaction was diluted with water (30 mL), and was extracted with ethyl acetate (2 × 15 mL). The organic phase was dried (MgSO$_4$), filtered and concentrated *in vacuo*, and purified by flash column chromatography on silica gel (10:1 petroleum ether: ethyl acetate) to yield 1-acetyl-2-phenyl-1*H*-indole-3-carbaldehyde (1.07 g, 90% yield). R$_f$ = 0.57 (2:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 9.76 (1H, s, CHO), 8.44 - 8.37 (1H, m, Ar), 8.33 – 8.25 (1H, m, Ar), 7.63 – 7.52 (5H, m, Ar), 7.50 – 7.38 (2H, m, Ar), 2.03 (3H, s, CH$_3$).

A solution of 1-acetyl-2-phenyl-1*H*-indole-3-carbaldehyde (388 mg, 1.47 mmol) and sodium triacetoxyborohydride (STAB, 375 mg, 1.77 mmol) in dry THF (35 mL) was heated at reflux for 40 hours. The solvent was then removed *in vacuo*, the crude reaction was redissolved in CH$_2$Cl$_2$, washed with saturated sodium carbonate, dried (MgSO$_4$), filtered and concentrated *in vacuo*.

The reaction was purified by flash column chromatography on silica gel (9:1 petroleum ether: ethyl acetate with 1% TEA → 2:1 petroleum ether: ethyl acetate with 1% TEA) to yield a mixture of the title product and 2-phenylindole-3-carboxyaldehyde. This mixture was then dissolved in a small quantity of CH_2Cl_2 and filtered to yield the title product as a yellow gum (203 mg, 52% yield). R_f = 0.37 (2:1 petroleum ether: ethyl acetate); ^1H NMR (400 MHz, $d6$-DMSO): δ_H = 8.26 (1H, d, J = 8.0 Hz, Ar), 7.78 (1H, m, J = 7.5 Hz, Ar), 7.56 – 7.49 (5H, m, Ar), 7.39 – 7.28 (2H, m, Ar), 4.98 (1H, br s, OH), 4.39 (2H, s, CH_2), 1.94 (3H, s, CH_3); ^{13}C NMR (100 MHz, $d6$-DMSO): δ_C = 170.85 (C=O), 136.15 (Ar, C_q), 136.04 (Ar, C_q), 132.15 (Ar, C_q), 130.24 (Ar, CH), 128.86 (Ar, C_q), 128.78 (Ar, CH), 128.54 (Ar, CH), 124.92 (Ar, CH), 123.23 (Ar, CH), 121.45 (Ar, C_q), 119.86 (Ar, CH), 115.49 (Ar, CH), 53.97 (CH_2), 27.34 (CH_3); IR (neat, cm^{-1}) v = 3397 (O-H stretch, broad), 1701 (C=O stretch), 1016 (C-O stretch); MS (EI): m/z 265 (M); HRMS found: [M+H]$^+$ 266.1180, $C_{17}H_{15}NO_2$+H$^+$ requires 266.1176.

8.2.7 Procedure for the synthesis of (1-tosyl-1H-indol-3-yl)methanol (133)
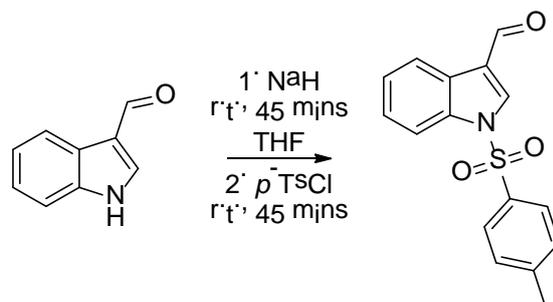
Sodium hydride (60% in oil, 328 mg, 8.20 mmol) in dry THF (15 mL) was added slowly to indole-3-carboxaldehyde (596 mg, 4.10 mmol) in dry THF (20 mL) at 0 °C. The reaction was then stirred at room temperature for 45 minutes. p-Toluenesulfonyl chloride (938 mg, 4.92 mmol) in dry THF (15 mL) was added dropwise to the reaction, and it was stirred at room temperature for 45 minutes until completion by TLC. The solvent was removed in $vacuo$, and the crude product was used in the next step of the reaction without further purification. ^1H NMR (300 MHz, CDCl$_3$): δ_H = 10.09 (1H, s, CHO), 7.28 – 7.22 (2H, m, Ar), 7.95 (1H, d, J = 7.5 Hz, Ar), 7.85 (2H, d, J = 8.5 Hz, Ar), 7.39 (2H, pd, J = 7.5, 1.5 Hz, Ar), 7.29 (2H, d, J = 8.0 Hz, Ar), 2.38 (3H, s, CH_3); ^{13}C NMR (75 MHz, CDCl$_3$) δ_C = 185.51 (CHO), 146.30 (Ar), 136.38 (Ar), 135.52 (Ar), 134.42 (Ar), 130.46 (Ar), 127.37 (Ar), 126.44 (Ar), 126.39 (Ar), 125.18 (Ar), 122.73 (Ar), 122.46 (Ar), 113.37 (Ar), 21.83 (CH_3).
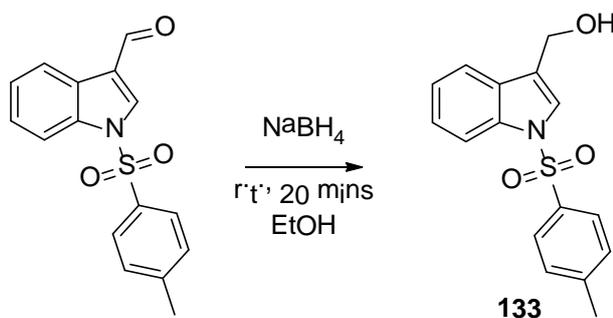
Dry ethanol (30 mL) and sodium borohydride (310 mg, 8.20 mmol) was added to the crude reaction from the previous step, and the reaction was stirred at room temperature for 20 minutes. Upon completion by TLC, the solvent was removed *in vacuo*, the excess sodium borohydride was quenched with sodium hydroxide solution (1%, 40 mL), and the reaction was extracted with ethyl acetate (3 × 20 mL) before being dried (MgSO$_4$), filtered and purified by preparative liquid chromatography (19:1 → 2:1 petroleum ether: ethyl acetate) to yield the title product as a yellow gummy solid (600 mg, 49% yield over two steps). R_f = 0.28 (9:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 7.99 (1H, d, J = 8.5 Hz, Ar), 7.77 (2H, d, J = 8.5 Hz, Ar), 7.61 (1H, d, J = 7.5 Hz, Ar), 7.55 (1H, s, Ar), 7.34 (1H, t, J = 7.5 Hz, Ar), 7.29 – 7.18 (3H, m, Ar), 4.82 (2H, d, J = 4.5 Hz, CH$_2$), 2.24 (3H, s, CH$_3$), 1.62 (1H, t, J = 5 Hz, OH); ^{13}C NMR (75 MHz, CDCl$_3$) δ_C = 145.15 (Ar, C$_q$), 135.53 (Ar, C$_q$), 135.31 (Ar, C$_q$), 130.05 (Ar, CH), 129.54 (Ar, C$_q$), 126.99 (Ar, CH), 125.13 (Ar, CH), 123.92 (Ar, CH), 123.43 (Ar, CH), 122.34 (Ar, C$_q$), 120.00 (Ar, CH), 113.83 (Ar, CH), 57.32 (CH$_2$), 21.72 (CH$_3$).

8.2.8 Procedure for the synthesis of benzo[*b*]thiophen-3-ylmethanol (134)
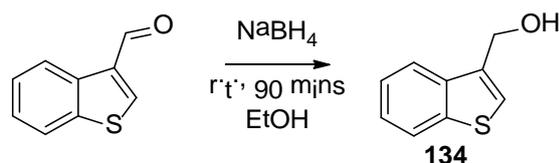
To dry ethanol (20 mL) was added benzo[*b*]thiophene-3-carbaldehyde (246 mg, 1.52 mmol) and sodium borohydride (115 mg, 3.03 mmol), and the reaction was stirred at room temperature for 90 mins. When the reaction was complete by TLC, excess sodium borohydride was quenched with sodium hydroxide solution (1%, 20 mL), and the reaction was extracted with ethyl acetate (3 × 20 mL) before being dried (MgSO$_4$), filtered and concentrated *in vacuo* to yield the title product in quantitative yield (249 mg). R_f = 0.40 (5:2 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 7.90 – 7.83 (2H, m, Ar), 7.44 – 7.34 (3H, m, Ar), 4.92 (2H, s, CH$_2$), 1.95 (1H, br s, OH); ^{13}C NMR (75 MHz, CDCl$_3$) δ_C = 140.87 (Ar), 137.75 (Ar), 136.04 (Ar), 124.69 (Ar), 124.32 (Ar), 123.95 (Ar), 122.99 (Ar), 122.05 (Ar), 59.88 (CH$_2$); MS (EI): *m/z* 164 (M).

8.2.9   Procedure for the synthesis of 2-(indolin-2-yl)phenol (149)
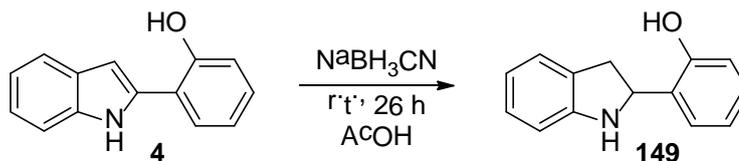
To a flask containing glacial acetic acid (10 mL) was added compound **4**(1.0 mmol, 209 mg) and sodium cyanoborohydride (12.0 mmol, 754 mg), and the reaction was stirred at room temperature for 26 hours. The crude reaction was then quenched by careful addition of water (100 mL), before adding solid sodium hydroxide to pH ~12. The crude reaction was extracted with ethyl acetate (3 × 20 mL), dried over $MgSO_4$ and filtered. The product was purified *via* flash column chromatography on silica gel (19:1 petroleum ether: ethyl acetate) to afford the title compound as a pale orange solid (92 mg, 43% yield). R_f =0.60 (3:1 petroleum ether: ethyl acetate), 0.16 (19:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 9.75 (1H, br s, OH), 7.28 – 7.10 (3H, m, Ar), 7.06 (1H, d, *J* = 7.5 Hz, Ar), 6.98 – 6.80 (4H, m, Ar), 4.92 (1H, dd, *J* = 12.5, 8.5 Hz, CH), 4.34 (1H, br s, NH), 3.38 – 3.04 (2H, m, CH$_2$); ^{13}C NMR (75 MHz, CDCl$_3$); δ_C = 156.88 (C$_q$, Ar), 148.70 (C$_q$, Ar), 130.80 (C$_q$, Ar), 129.12 (CH, Ar), 128.46 (CH, Ar), 127.67 (CH, Ar), 124.93 (C$_q$, Ar), 124.80 (CH, Ar), 121.46 (CH, Ar), 119.54 (CH, Ar), 117.64 (CH, Ar), 111.95 (CH, Ar), 65.59 (CH), 38.42 (CH$_2$); MS (ESI): *m/z* 212 ([M+H]$^+$).

8.2.10 Procedure for the synthesis of 2-(2-(benzyloxy)phenyl)-1*H*-indole (154)
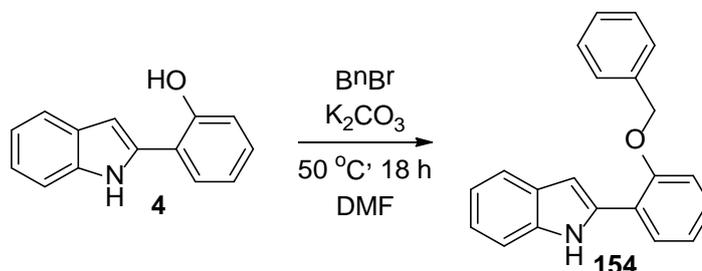
To a flask containing DMF (15 mL), compound **4** (0.62 mmol, 130 mg) and potassium carbonate (172 mg, 1.24 mmol) was added benzyl bromide (88 μL, 0.74 mmol), and the reaction was stirred at 50°C overnight. Upon completion by TLC, the crude reaction was diluted with water (90 mL), and was extracted with ethyl acetate (3 × 20 mL). The organic layer was washed with water (10 mL) and brine (10 mL), and was purified by preparative liquid chromatography (19:1 petroleum ether: ethyl acetate) to afford the title compound as a pale yellow solid (160 mg, 86% yield). R_f = 0.47 (9:1 petroleum ether: ethyl acetate), 0.28 (19:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H =9.78 (1H, br s, NH), 7.91 (1H, d, *J* = 7.5 Hz, Ar), 7.65 (1H, d, *J* = 7.0 Hz, Ar), 7.58 – 7.41 (5H, m, Ar), 7.34 – 7.05 (6H, m, Ar), 6.95 (1H, s, Ar), 5.24 (2H, s, CH$_2$); ^{13}C NMR (75

MHz, CDCl$_3$): δ_C = 155.02 (Ar, C$_q$), 136.47 (Ar, C$_q$), 136.14 (Ar, C$_q$), 135.92 (Ar, C$_q$), 129.07 (Ar, CH), 128.70 (Ar, CH), 128.61 (Ar, CH), 128.38 (Ar, CH), 128.03 (Ar, C$_q$), 127.88 (Ar, CH), 122.07 (Ar, CH), 121.88 (Ar, CH), 121.06 (Ar, C$_q$), 120.33 (Ar, CH), 119.87 (Ar, CH), 113.61 (Ar, CH), 110.94 (Ar, CH), 99.85 (Ar, CH), 71.38 (CH$_2$); IR (neat, cm^{-1}) v = 3045.2 (Aromatic C-H stretch), 1226.7 (C-O stretch), 1007.5 (C-O stretch), 737.5 (Aromatic C-H out-of-plane bend); MS (APCI): m/z 300 ([M+H]$^+$).

8.2.11 Procedure for the synthesis of 2-(2-((*tert*-butyldimethylsilyl)oxy)phenyl)-1*H*-indole (155)
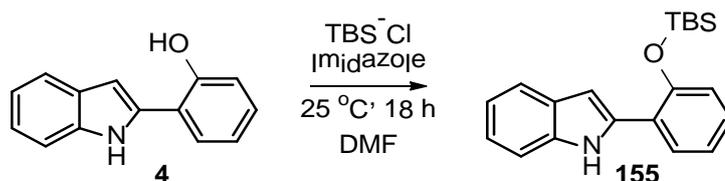
To a flask containing compound **4** (0.65 mmol, 135 mg) and imidazole (3.9 mmol, 266 mg) in DMF (25 mL) was added *tert*-butyldimethylsilyl chloride (3.2 mmol, 482 mg), and the reaction was stirred at 25 °C for 18 hours. The reaction was quenched with water (75 mL), extracted with ethyl acetate (3 × 25 mL) and purified *via* preparative liquid chromatography (petroleum ether) to yield the title compound as a white solid (179 mg, 85% yield). M.p. 92 - 94 °C; R$_f$ = 0.25 (petroleum ether); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 9.54 (1H, br s, NH), 7.78 (1H, dd, J = 8.0, 1.5 Hz, Ar), 7.65 (1H, d, J = 7.5, Ar), 7.35 (1H, d, J = 7.5 Hz, Ar), 7.25 – 7.03 (4H, m, Ar), 6.96 (1H, dd, J = 8.0, 1.0 Hz, Ar), 6.82 (1H, s, Ar), 1.05 (9H, s, ^{t}Bu), 0.22 (6H, s, 2 × Me); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 152.02 (Ar, C$_q$), 136.37 (Ar, C$_q$), 135.99 (Ar, C$_q$) , 129.12 (Ar, CH), 128.46 (Ar, CH) , 128.32 (Ar, C$_q$), 123.31 (Ar, C$_q$), 122.35 (Ar, CH), 121.90 (Ar, CH), 120.74 (Ar, CH), 120.43 (Ar, CH), 119.88 (Ar, CH), 110.70 (Ar, CH), 99.81 (Ar, CH), 26.07 (C(CH$_3$)$_3$), 18.49 (3 × C(CH$_3$)$_3$), -4.11 (2 × Me); IR (liquid film, cm^{-1}) v = 1242 (SiO-C); MS (EI): m/z 266 (M-^{t}Bu, 100%), 323 (M, 22%); HRMS found: [M+H]$^+$ 324.1779, C$_{20}$H$_{25}$NOSi+H$^+$ requires 324.1778.

8.2.12 Procedure for the synthesis of 2-(1*H*-indol-2-yl)phenyl methanesulfonate (156)
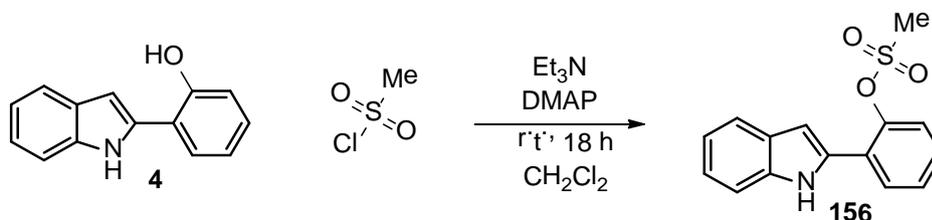
Compound **4** (0.97 mmol, 204 mg), methanesulfonyl chloride (1.16 mmol, 90 μL), triethylamine (1.46 mmol, 205 μL) and DMAP (0.1 mmol, 12 mg) was stirred in CH$_2$Cl$_2$ (10 mL) at room

temperature for 18 hours. Upon completion by TLC, the crude reaction was washed with 1M HCl (2 × 10 mL), water (10 mL) and brine (10 mL) to afford the title product as an orange solid (281 mg, 100% yield). M.p. 111 - 114°C (from CH$_2$Cl$_2$); R$_f$ = 0.27 (3:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.95 (1H, br s, NH), 7.79-7.72 (1H, m, Ar), 7.66 (1H, d, J = 8.0 Hz, Ar), 7.56 – 7.50 (1H, m, Ar)7.49 – 7.39 (3H, m, Ar), 7.25 (1H, td, J = 7.5, 1.0 Hz, Ar), 7.16 (1H, td, J = 7.5, 1.0 Hz, Ar), 6.87 (1H, dd, J = 2.0, 1.0 Hz, Ar), 2.79 (3H, s, CH$_3$); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 145.73 (Ar, C$_q$), 136.89 (Ar, C$_q$), 133.07 (Ar, C$_q$), 130.20 (Ar, CH), 129.36 (Ar, CH), 128.28 (Ar, C$_q$), 128.23 (Ar, CH), 126.12 (Ar, C$_q$), 124.26 (Ar, CH), 123.17 (Ar, CH), 120.96 (Ar, CH), 120.58 (Ar, CH), 111.38 (Ar, CH), 103.38 (Ar, CH), 37.60 (CH$_3$); IR (neat, cm^{-1}) v = 1358 (S=O stretch), 1154 (S=O stretch); MS (ESI): m/z 288 ([M+H]$^+$); HRMS found: [M+H]$^+$ 288.0691, C$_{15}$H$_{13}$NO$_3$S+H$^+$ requires 288.0689.

8.2.13 Procedure for the synthesis of 2-(1H-indol-2-yl)phenyl benzoate (157)
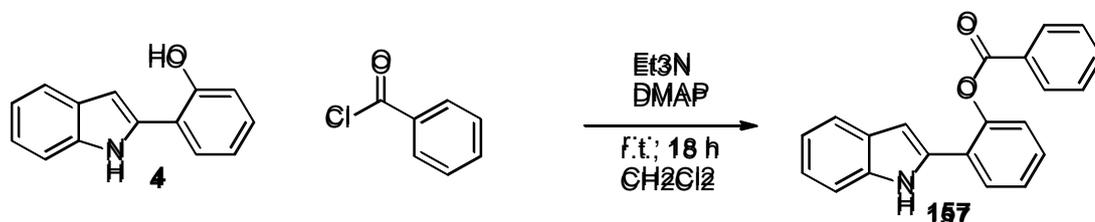
Compound **4** (0.97 mmol, 202 mg), benzoyl chloride (1.16 mmol, 135 µL), triethylamine (1.46 mmol, 203 µL) and DMAP (0.1 mmol, 12 mg) was stirred in CH$_2$Cl$_2$ (10 mL) at room temperature for 18 hours. Upon completion by TLC, the crude reaction was washed with 1M HCl (2 × 10 mL), water (10 mL) and brine (10 mL). The crude reaction was purified *via* preparative liquid chromatography on silica gel (19:1 to 9:1 petroleum ether: ethyl acetate) to afford the title product as a yellow solid (141 mg, 47% yield). M.p. 148 - 168 °C (decomposes); R$_f$ = 0.18 (9:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.60 (1H, s, NH), 8.23 (2H, d, J = 7.5 Hz, Ar), 7.79 – 7.03 (11H, m, Ar), 6.83 (1H, s, Ar); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 165.24 (C=O), 147.66 (Ar, C$_q$), 136.54 (Ar, C$_q$), 134.09 (Ar, CH), 133.73 (Ar, C$_q$), 130.29 (Ar, CH), 129.27 (Ar, C$_q$), 129.08 (Ar, CH), 128.96 (Ar, CH), 128.80 (Ar, CH), 128.61 (Ar, C$_q$), 126.80 (Ar, CH), 125.85 (Ar, C$_q$), 123.78 (Ar, CH), 122.50 (Ar, CH), 120.79 (Ar, CH), 120.14 (Ar, CH), 110.96 (Ar, CH), 102.80 (Ar, CH); IR (neat, cm^{-1}) v = 1725 (C=O stretch), 1260 (C-O stretch); MS (ESI): m/z 314 ([M+H]$^+$); HRMS found: [M+H]$^+$ 314.1184, C$_{21}$H$_{15}$NO$_2$+H requires 314.1176.

8.2.14 Procedure for the synthesis of 2-(1*H*-indol-2-yl)phenyl acetate (158)
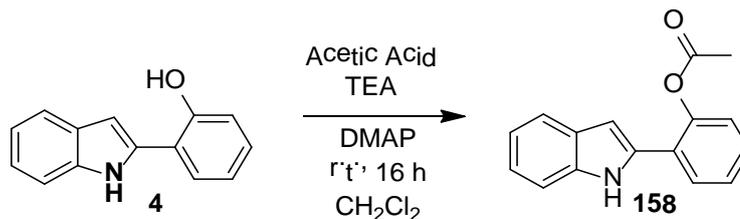
2-(1*H*-Indol-2-yl)phenol (101.7 mg, 0.59 mmol), acetic acid (0.06 mL, 0.54 mmol), triethylamine (54 μL) and DMAP (6 mg, 0.05 mmol) were added to CH_2Cl_2 (5 mL) and stirred at room temperature for 16 h. The reaction was washed with 1 M HCl (3 × 5 mL), water (1 × 5 mL) and brine (1 × 5 mL), dried ($MgSO_4$), filtered and the solvent was removed *in vacuo*. The reaction was purified by flash column chromatography on silica gel (5:1 petroleum ether: ethyl acetate) to yield the title product as an off-white solid (84.3 mg, 71% yield). M.p. 112-127 °C; R_f = 0.23 (5:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$): δ_H = 8.50 (1H, s, NH), 7.70-7.62 (2H, m, Ar), 7.43-7.32 (3H, m, Ar), 7.26-7.09 (3H, m, Ar), 6.78-6.82 (1H, m, Ar), 2.31 (3H, s, CH_3); ^{13}C NMR (75 MHz, $CDCl_3$): δ_C = 169.42 (C=O), 147.58 (Ar), 136.59 (Ar), 133.92 (Ar), 129.28 (Ar), 128.88 (Ar), 128.71 (Ar), 126.81 (Ar), 125.86 (Ar), 123.67 (Ar), 122.65 (Ar), 120.88 (Ar), 120.32 (Ar), 111.00 (Ar), 102.73 (Ar), 21.40 (CH_3); IR (neat, cm^{-1}) v = 3357 (N-H stretch), 1735 (C=O stretch), 1368 (aromatic C-C stretch), 1220 (C-O stretch); MS (ESI): *m/z* 252 ([M+H]$^+$); HRMS found: [M+H]$^+$ 252.1021, $C_{16}H_{13}NO_2$+H$^+$ requires 252.1019.

8.2.15 Procedure for the synthesis of 2-(1*H*-indol-2-yl)phenyl heptanoate (159)
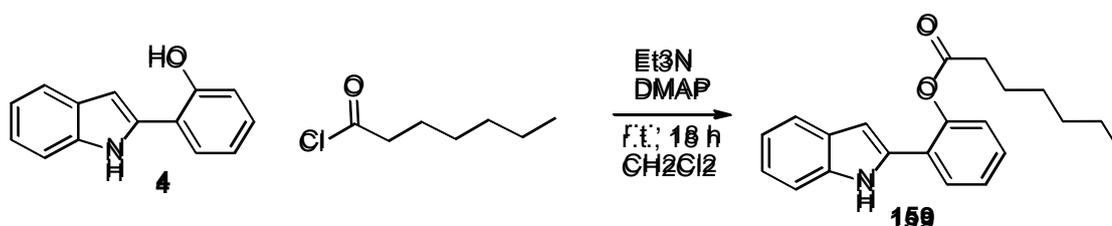
Compound **4** (0.92 mmol, 192 mg), heptanoyl chloride (1.10 mmol, 171 μL), triethylamine (1.38 mmol, 192 μL) and DMAP (0.1 mmol, 12 mg) was stirred in CH_2Cl_2 (10 mL) at room temperature for 18 hours. Upon completion by TLC, the crude reaction was washed with 1M HCl (2 × 10 mL), water (10 mL) and brine (10 mL). The crude reaction was purified *via* preparative liquid chromatography on silica gel (19:1 to 9:1 petroleum ether: ethyl acetate) to afford the title product as a yellow solid (225 mg, 76% yield). M.p. 51 - 54 °C; R_f = 0.15 (9:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, $CDCl_3$): δ_H = 8.53 (1H, br s, NH), 7.70 – 7.61 (2H, m, Ar), 7.41 – 7.27 (3H, m, Ar), 7.25 – 7.10 (3H, m, Ar), 6.81 – 6.76 (1H, m, Ar), 2.59 (2H, t, *J* = 7.5 Hz, CH_2), 1.70 (2H, pent, *J* = 7.5 Hz, CH_2), 1.39 – 1.19 (6H, m, 3 × CH_2), 0.87 (3H, t, *J* = 6.5 Hz, CH_3); ^{13}C NMR

(75 MHz, CDCl$_3$): δ_C = 172.36 (C=O), 147.64 (Ar, C$_q$), 136.54 (Ar, C$_q$), 133.96 (Ar, C$_q$), 129.35 (Ar, CH), 128.82 (Ar, CH), 128.65 (Ar, C$_q$), 126.65 (Ar, CH), 125.89 (Ar, C$_q$), 123.65 (Ar, CH), 122.55 (Ar, CH), 120.81 (Ar, CH), 120.23 (Ar, CH), 110.97 (Ar, CH), 102.64 (Ar, CH), 34.66 (CH$_2$), 31.53 (CH$_2$), 28.88 (CH$_2$), 24.90 (CH$_2$), 22.54 (CH$_2$), 14.14 (CH$_3$); IR (neat, cm^{-1}) v = 1145 (C-O stretch), 1746 (C=O stretch), 2928 (C-H stretch); MS (ESI): m/z 322 ([M+H]$^+$); HRMS found: [M+H]$^+$ 322.1805, C$_{21}$H$_{23}$NO$_2$+H requires 322.1802.

8.2.16 Procedure for the synthesis of 2-(1H-indol-3-yl)phenol (160)
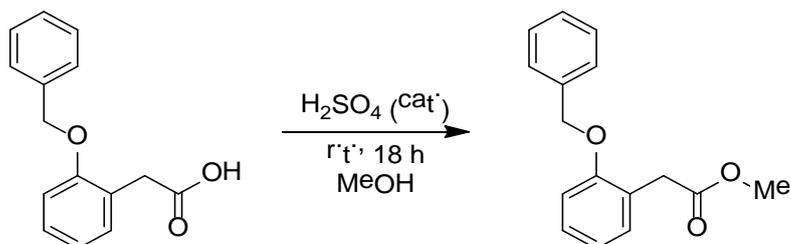
2-Benzyloxyphenylacetic acid (983 mg, 4.06 mmol) was stirred in methanol (45 mL) and sulfuric acid (2 drops) at room temperature for 72 hours to yield methyl 2-(2-(benzyloxy)phenyl)acetate as a pale red viscous oil. R$_f$ = 0.60 (3:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 7.45 – 7.29 (5H, m, Ar), 7.29 – 7.18 (2H, m, Ar), 6.94 (2H, m, Ar), 5.08 (2H, s, O-C$\underline{H}_2$-Ar), 3.70 (2H, s, Ar-C$\underline{H}_2$-COOMe), 3.64 (3H, s, CH$_3$).
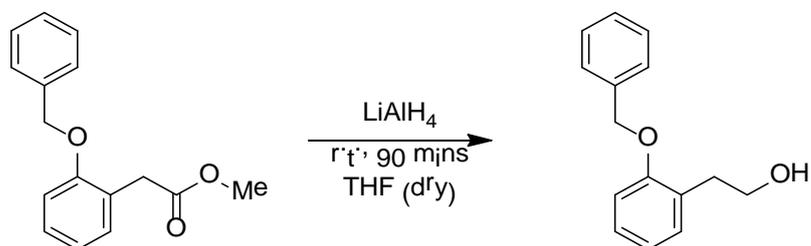
The crude methyl ester was reduced to the alcohol by stirring in THF (dry, 15 mL) at 0°C and adding LiAlH$_4$ (308 mg, 8.11 mmol) under an atmosphere of nitrogen. The ice bath was then removed, and the reaction was stirred at room temperature for 90 minutes. The reaction was quenched with NaOH$_{(aq)}$ (10 mL, 0.1 M). The crude product was extracted with ethyl acetate (3 × 10 mL), and the combined organic layers were dried over MgSO$_4$ and filtered before being concentrated *in vacuo* to yield crude 2-(2-(benzyloxy)phenyl)ethanol as a pale yellow oil. R$_f$ = 0.22 (3:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 7.49 – 7.32 (5H, m, Ar), 7.28 – 7.19 (2H, m, Ar), 7.00 - 6.91 (2H, m, Ar), 5.10 (2H, s, Ar-C$\underline{H}_2$-OAr), 3.87 (2H, t, J = 6.5, C$\underline{H}_2$-OH), 2.98 (2H, t, J = 6.5 Hz, Ar-C$\underline{H}_2$-CH$_2$OH), 1.95 (1H, br s, OH).
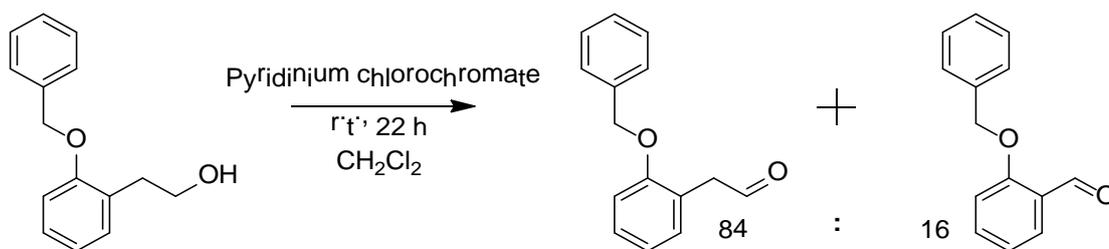
The crude alcohol was oxidised to the aldehyde by stirring with pyridinium chlorochromate (1.31 g, 6.08 mmol) in CH_2Cl_2 (40 mL) in the presence of crushed molecular sieves (650 mg) at room temperature for 22 hours. The solvent was removed *in vacuo*, and the crude reaction was dissolved in diethyl ether (40 mL), before being filtered through Celite®. The crude reaction was purified by flash column chromatography (19:1 petroleum ether: ethyl acetate) to yield 0.70 g of an 84:16 ratio of products,[306] which included 468 mg (51% yield over 3 steps) of the major product 2-(2-(benzyloxy)phenyl)acetaldehyde as a pale yellow oil. R_f = 0.61 (3:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 9.72 (1H, t, *J* = 2.0 Hz, CHO), 7.45 – 7.24 (6H, m, Ar), 7.20 - 7.14 (1H, m, Ar), 7.00 - 6.93 (2H, m, Ar), 5.08 (2H, s, O-CH$_2$-Ar), 3.70 (2H, d, *J* = 2.0 Hz, CH$_2$-CHO).
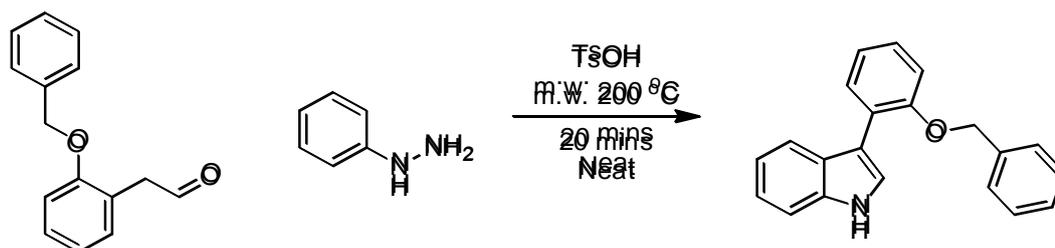
The mixture of aldehydes (0.70g, of which 468 mg, 2.07 mmol was the pictured aldehyde) was added to a microwave vial with phenyl hydrazine (336 mg, 3.11 mmol) and *p*-toluenesulfonic acid (40 mg, 0.21 mmol) and heated at 200 °C for 20 minutes. The crude product was purified by flash column chromatography on silica gel (19:1 → 9:1 petroleum ether: ethyl acetate) to yield 3-(2-(benzyloxy)phenyl)-1*H*-indole (200 mg, 32% yield). ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.12 (1H, br s, NH), 7.95 (1H, d, *J* = 7.5 Hz, Ar), 7.79 (1H, dd, *J* = 7.5, 1.5 Hz, Ar), 7.53 (1H, d, *J* = 2.5 Hz, Ar), 7.42 – 7.11 (11H, m, Ar), 5.14 (2H, s, CH$_2$).
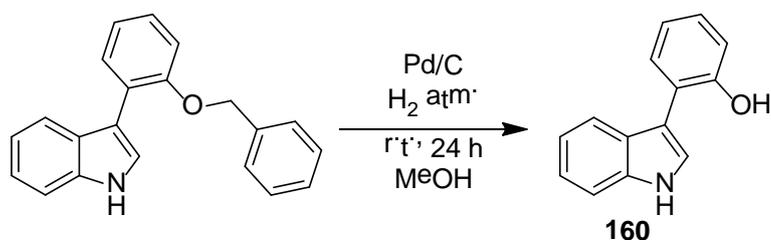
Methanol (10 mL) was added to a two-necked flask containing 3-(2-(benzyloxy)phenyl)-1*H*-indole (200 mg, 0.67 mmol) and palladium on carbon (20 mg, 10 wt%). The flask was evacuated and backfilled with nitrogen twice, before being evacuated and backfilled with hydrogen twice. The reaction was then stirred at room temperature under an atmosphere of hydrogen for 24 hours, and the crude product was filtered through Celite® and purified by flash column chromatography on silica gel (4:1 → 3:1 petroleum ether: ethyl acetate) to yield 2-(1H-indol-3-yl)phenol (119 mg, 85% yield) as a thick orange oil. R_f = 0.29 (3:1 petroleum ether: ethyl acetate), 0.12 (4:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.31 (1H, br s, NH), 7.73 (1H, d, *J* = 8 Hz, Ar), 7.50 (1H, dd, *J* = 7.5, 1.5 Hz, Ar), 7.45 – 7.31 (3H, m, Ar), 7.31 – 7.20 (2H, m, Ar), 7.19 – 7.08 (2H, m, Ar), 5.69 (1H, br s, OH); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 153.22 (Ar, C$_q$), 136.34 (Ar, C$_q$), 130.89 (Ar, CH), 128.57 (Ar, CH), 126.21 (Ar, C$_q$), 123.43 (Ar, CH), 122.91 (Ar, CH), 121.09 (Ar, C$_q$), 120.74 (Ar, CH), 120.54 (Ar, CH), 119.74 (Ar, CH), 115.45 (Ar, CH), 111.66 (Ar, CH); MS (ESI): *m/z* 208 ([M-H]$^-$); HRMS found: [M+H]$^+$ 210.0913, C$_{14}$H$_{11}$NO+H$^+$ requires 210.0913.

8.2.17 Procedure for the synthesis of 2-(naphthelen-7-yl) phenol (168)[307]
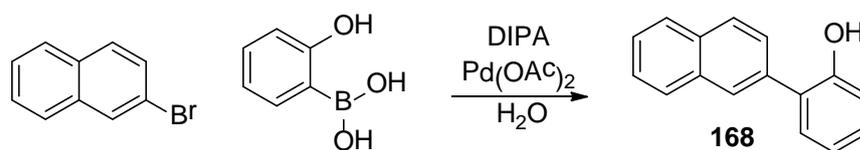
To a flask containing water (5 mL), was added 2-bromonaphthalene (331 mg, 1.6 mmol), 2-hydroxyphenylboronic acid (500 mg, 2.4 mmol), diisopropylamine (DIPA, 0.45 mL, 3.2 mmol) and palladium (II) acetate (12.3 mg, 0.25 mol%). The reaction was heated to reflux for 3.5 hours, and was allowed to cool to room temperature before being diluted with brine (40 mL) and extracted with ethyl acetate (5 × 20 mL). The reaction was filtered through Celite®, and the filtrate dried (MgSO$_4$) and filtered. The crude reaction was purified *via* flash column chromatography on silica gel (5:1 petroleum ether: ethyl acetate) to afford the title compound as a pale yellow solid (288 mg, 82% yield). M.p. 93-96 °C; R_f = 0.37 (5:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.03-7.97 (2H, m, Ar), 7.96-7.88 (2H, m, Ar), 7.66-7.54 (3H, m, Ar), 7.43-7.30 (2H, m, Ar), 7.13 (2H, m, Ar), 5.42 (1H, s, OH); ^{13}C NMR (75 MHz, CDCl$_3$):

δ_C = 152.66 (Ar, C_q), 134.57 (Ar, C_q), 133.66 (Ar, C_q), 132.77 (Ar, C_q), 130.25 (Ar, CH), 129.35 (Ar, CH), 129.20 (Ar, CH), 128.15 (Ar, C_q), 128.13 (Ar, CH), 127.94 (Ar, CH), 127.88 (Ar, CH), 127.23 (Ar, CH), 126.72 (Ar, CH), 126.52 (Ar, CH), 121.07 (Ar, CH), 116.02 (Ar, CH); IR (neat, cm^{-1}) v = 3520 (OH stretch), 750 (Aromatic C-C bend); MS (EI): m/z 220 (M); HRMS found: [M+H]$^+$ 221.0960, $C_{16}H_{12}O$+H$^+$ requires 221.0961.

8.2.18 Procedure for the synthesis of 2-(2-hydroxyphenyl)-3-methyl-3H-indol-3-ol (198)
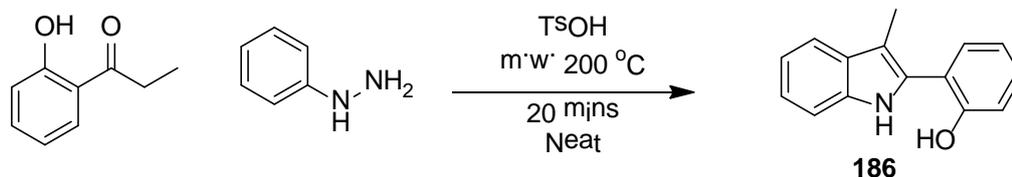
Method 1:

2′-Hydroxypropiophenone (93.7 mg, 0.87 mmol), phenylhydrazine (200.8 mg, 1.34 mmol) and *p*-toluenesulfonic acid (18.6 mg, 0.10 mmol) were heated under microwave irradiation at 200°C for 20 minutes. The crude reaction product was purified *via* preparative liquid chromatography (petroleum ether → 5:1 petroleum ether: ethyl acetate) to yield 2-(3-methyl-1*H*-indol-2-yl)phenol as a highly viscous yellow oil (143 mg, 74% yield).

Method 2:[308]

2′-Hydroxypropiophenone (10 mmol, 1.50 g), phenylhydrazine (10 mmol, 1.08 g) and acetic acid (4 drops) were stirred in ethanol (10 mL) at reflux for 2 hours. The reaction was then cooled to room temperature, and the ethanol was removed *in vacuo*. Polyphosphoric acid (5.5 g) was then added, and the reaction was heated at 170°C for 30 mins. The reaction was cooled to room temperature, and diluted with water (100 mL), and neutralised with sodium hydroxide (1 M). The crude reaction was extracted with CH_2CH_2 (2 × 50 mL), dried over $MgSO_4$ and filtered. The crude reaction was purified *via* preparative liquid chromatography (4:1 petroleum ether: ethyl acetate) to afford 2-(3-methyl-1*H*-indol-2-yl)phenol as a highly viscous yellow oil (1.61g, 72% yield).

^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 8.06 (1H, br s, NH), 7.64 (1H, d, *J* = 7.5 Hz, Ar), 7.41 – 7.30 (3H, m, Ar), 7.26 (1H, td, *J* = 7.5, 1.5 Hz, Ar), 7.19 (1H, td, *J* = 7.5, 1.0 Hz, Ar), 7.08 – 7.00 (2H, m, Ar), 5.55 (1H, br s, OH), 2.33 (3H, s, CH$_3$); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 153.45 (C_q, Ar), 136.56 (C_q Ar), 130.52 (CH, Ar), 130.15 (CH, Ar), 129.47 (C_q, Ar), 129.40 (C_q, Ar), 122.84 (CH, Ar), 120.81 (CH, Ar), 119.91 (CH, Ar), 119.23 (C_q, Ar), 119.16 (CH, Ar), 115.95 (CH, Ar), 111.09 (CH, Ar), 110.70 (C_q, Ar), 9.48 (CH$_3$); IR (neat, cm^{-1}) v = 3399 (OH stretch), 744 (Ar C-H bend); MS (EI): m/z 223 ([M]$^+$).

186 → **198**
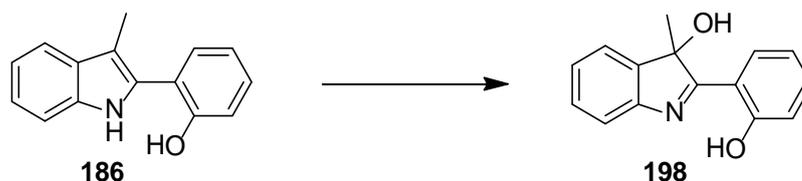
The title product was formed by the degradation of 2-(3-methyl-1H-indol-2-yl)phenol in air, and was purified by recrystallization (ethyl acetate and petroleum ether). M.p. 135-136 °C; R_f = 0.28 (7:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 13.59 (1H, br s, Ar-OH), 8.20 (1H, dd, J = 8.5, 1.5 Hz, Ar), 7.52 (1H, d, J = 7Hz, Ar), 7.47 – 7.28 (4H, m, Ar), 7.01 – 6.93 (2H, m, Ar), 2.78 (1H, br s, OH), 1.78 (3H, s, CH$_3$); ^{13}C NMR (75 MHz, CDCl$_3$): δ_C = 181.85 (Ar, C$_q$), 161.65 (Ar, C$_q$), 149.95 (Ar, C$_q$), 141.11 (Ar, C$_q$), 133.47 (Ar, CH), 129.94 (Ar, CH), 129.76 (Ar, CH), 126.85 (Ar, CH), 121.97 (Ar, CH), 120.33 (Ar, CH), 118.87 (Ar, CH), 117.83 (Ar, CH), 114.89 (C-PhOH), 84.73 (C-CH$_3$), 26.25 (CH$_3$); IR (neat, cm^{-1}) v = 3414 (O-H stretch), 1090 (C-OH stretch), 741 (C-H bend); MS (EI): m/z 239 ([M]$^+$); HRMS found: [M-H$^+$] 238.0875, C$_{15}$H$_{13}$NO$_2$-H$^+$ requires 238.0874; Anal. Calcd for C$_{15}$H$_{13}$NO$_2$: C, 75.3 %; H, 5.5 %; N, 5.8 %. Found: C, 75.0 %; H, 5.5 %; N, 5.8 %. XRD data is shown in Appendix 8.

8.2.19 Procedure for the synthesis of 2-(1H-indol-2-yl)-4-methylphenol (244)

244
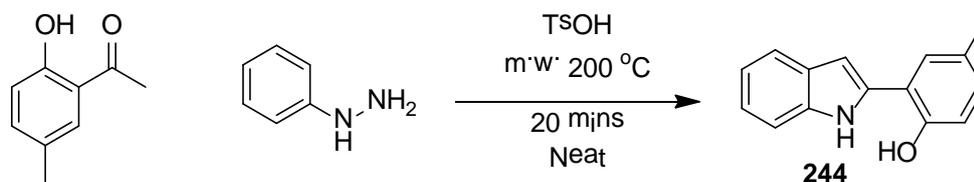
Phenylhydrazine (108 mg, 1.00 mmol), 2-hydroxy-5-methylacetophenone (216 mg, 1.44 mmol) and *p*-toluenesulfonic acid (20 mg, 0.10 mmol) were heated under microwave irradiation at 200°C for 20 minutes. The crude reaction was purified *via* flash column chromatography on silica gel (5:1 petroleum ether: ethyl acetate) to afford the title compound as a dark solid (68 mg, 31% yield). R_f = 0.09 (5:1 petroleum ether: ethyl acetate); ^{1}H NMR (300 MHz, CDCl$_3$): δ_H = 9.28 (1H, br s, NH), 7.65 (1H, d, J = 8.0 Hz, Ar), 7.51 (1H, d, J = 1.5 Hz, Ar), 7.42 (1H, d, J = 8.0 Hz, Ar), 7.20 (1H, td, J = 7.5, 1.0 Hz, Ar), 7.13 (1H, td, J = 7.5, 1.0 Hz, Ar), 7.01 (1H, dd, J – 8.0, 2.0 Hz, Ar), 6.78 – 6.88 (2H, m, Ar), 5.43 (1H, br s, OH), 2.25 (3H, s, CH3); ^{13}C NMR (100 MHz, CDCl$_3$): δ_C = 150.10 (C$_q$, Ar), 136.62 (C$_q$, Ar), 135.27 (C$_q$, Ar), 130.87 (C$_q$, Ar), 129.61 (CH, Ar), 128.74 (CH, Ar), 128.65 (C$_q$, Ar), 122.27 (CH, Ar), 120.55 (CH, Ar), 120.22 (CH, Ar), 119.01 (C$_q$, Ar), 116.64 (CH, Ar), 111.13 (CH, Ar), 100.23 (CH, Ar), 20.70 (CH$_3$); MS (EI): m/z 224 ([M+H]$^+$); HRMS found: [M-H]$^-$ 222.0924, C$_{15}$H$_{13}$NO-H$^+$ requires 222.0924.

REFERENCES

References

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, *CA. Cancer J. Clin.* **2015**, *65*, 87–108.

[2] World Health Organisation, Global health observatory data repository, "Global health observatory data repository.," can be found under http://apps.who.int/gho/data/node.main.CODWORLD, **2016**.

[3] Cancer Research UK, "Cancer Stats Reports - Cancer Research UK," can be found under http://publications.cancerresearchuk.org/cancerstats/, **2015**.

[4] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182.

[5] J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.

[6] S. Prabhu, Z. Akbar, F. Harris, K. Karakoula, R. Lea, F. Rowther, T. Warr, T. Snape, *Bioorg. Med. Chem.* **2013**, *21*, 1918–1924.

[7] F. Bray, A. Jemal, N. Grey, J. Ferlay, D. Forman, *Lancet Oncol.* **2012**, *13*, 790–801.

[8] "International Agency for Research on Cancer. GLOBOCAN 2012 v1.0," can be found under globocan.iarc.fr., **n.d.**

[9] Office for National Statistics, "Cancer Statistics Registrations, England (Series MB1), No. 43, 2012," can be found under http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-352128, **2015**.

[10] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, D. W. Ellison, *Acta Neuropathol. (Berl.)* **2016**, *131*, 803–820.

[11] P. Bailey, H. Cushing, *A Classification of the Tumors of the Glioma Group on a Histogenetic Basis With a Correlated Study of Prognosis*, J. B. Lippincott Company, Philadelphia, **1926**.

[12] R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. B. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, et al., *N. Engl. J. Med.* **2005**, *352*, 987–996.

[13] M. F. Stevens, J. A. Hickman, R. Stone, N. W. Gibson, G. U. Baig, E. Lunt, C. G. Newton, *J. Med. Chem.* **1984**, *27*, 196–201.

[14] J. P. Neijt, M. E. L. van der Burg, J. P. Guastalla, M. George, M. Piccart, J. Vermorken, F. Carnino, N. Rotmensz, *Acta Oncol.* **1989**, *28*, 663–665.

[15] J. F. Smyth, S. Gundersen, J. Renard, H. M. Pinedo, E. E. C. T. Group, others, *Eur. J. Cancer Clin. Oncol.* **1989**, *25*, 755–757.

[16] M. Harding, V. Docherty, R. Machie, A. Dorward, S. Kaye, *Eur. J. Cancer Clin. Oncol.* **1989**, *25*, 785–788.

[17] E. S. Newlands, G. R. Blackledge, J. A. Slack, G. J. Rustin, D. B. Smith, N. S. Stuart, C. P. Quarterman, R. Hoffman, M. F. Stevens, M. H. Brampton, *Br. J. Cancer* **1992**, *65*, 287.

[18] S. S. Agarwala, J. M. Kirkwood, *The oncologist* **2000**, *5*, 144–151.

References

[19] M. Patel, C. McCully, K. Godwin, F. M. Balis, *J. Neurooncol.* **2003**, *61*, 203–207.

[20] S. H. Lin, L. R. Kleinberg, *Expert Rev. Anticancer Ther.* **2008**, *8*, 343–359.

[21] D. Beier, J. B. Schulz, C. P. Beier, *Mol Cancer* **2011**, *10*, 128–139.

[22] A. E. Pegg, M. Boosalis, L. Samson, R. C. Moschel, T. L. Byers, K. Swenn, M. E. Dolan, *Biochemistry (Mosc.)* **1993**, *32*, 11998–12006.

[23] M. Esteller, J. Garcia-Foncillas, E. Andion, S. N. Goodman, O. F. Hidalgo, V. Vanaclocha, S. B. Baylin, J. G. Herman, *N. Engl. J. Med.* **2000**, *343*.

[24] J. A. Quinn, *J. Clin. Oncol.* **2005**, *23*, 7178–7187.

[25] J. A. Quinn, S. X. Jiang, D. A. Reardon, A. Desjardins, J. J. Vredenburgh, J. N. Rich, S. Gururangan, A. H. Friedman, D. D. Bigner, J. H. Sampson, et al., *J. Clin. Oncol.* **2009**, *27*, 1262–1267.

[26] D. G. Deavall, E. A. Martin, J. M. Horner, R. Roberts, *J. Toxicol.* **2012**, *2012*, 1–13.

[27] R. Apak, M. Özyürek, K. Güçlü, E. Çapanoğlu, *J. Agric. Food Chem.* **2016**, *64*, 1046–1070.

[28] J. Wang, J. Yi, *Cancer Biol. Ther.* **2008**, *7*, 1875–1884.

[29] B. D'Autréaux, M. B. Toledano, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 813–824.

[30] J. Chiu, I. W. Dawes, *Trends Cell Biol.* **2012**, *22*, 592–601.

[31] D. Harman, *Mutat. Res.* **1992**, *275*, 257–266.

[32] D. Harman, *J. Gerontol.* **2002**, *11*, 298–300.

[33] A. Ayer, C. W. Gourlay, I. W. Dawes, *FEMS Yeast Res.* **2014**, *14*, 60–72.

[34] G. H. Kim, J. E. Kim, S. J. Rhie, S. Yoon, *Exp. Neurobiol.* **2015**, *24*, 325.

[35] H. W. Querfurth, F. M. LaFerla, *N Engl J Med* **2010**, *362*, 329–344.

[36] V. Chiurchiù, A. Orlacchio, M. Maccarrone, *Oxid. Med. Cell. Longev.* **2016**, *2016*, 1–11.

[37] Q. Wu, X. Ni, *Curr. Drug Targets* **2015**, *16*, 13–19.

[38] D. Ziech, R. Franco, A. Pappa, M. I. Panayiotidis, *Mutat. Res. Mol. Mech. Mutagen.* **2011**, *711*, 167–173.

[39] K. W. Lee, H. J. Lee, *Mech. Ageing Dev.* **2006**, *127*, 424–431.

[40] J. K. Strzelczyk, A. Wiczkowski, *Współczesna Onkol.* **2012**, *3*, 230–233.

[41] G. C. Brown, V. Borutaite, *Mitochondrion* **2012**, *12*, 1–4.

[42] P. A. Riley, *Int. J. Radiat. Biol.* **1994**, *65*, 27–33.

References

[43] T. Yamamori, H. Yasui, M. Yamazumi, Y. Wada, Y. Nakamura, H. Nakamura, O. Inanami, *Free Radic. Biol. Med.* **2012**, *53*, 260–270.

[44] T. Guzik, R. Korbut, T. Adamek-Guzik, *J. Physiol. Pharmacol.* **2003**, *54*, 469–487.

[45] B. Halliwell, J. Gutteridge, *FEBS Lett.* **1992**, *307*, 108–112.

[46] M. R. Ramsey, N. E. Sharpless, *Nat. Cell Biol.* **2006**, *8*, 1213–1215.

[47] T. Pan, *Annu. Rev. Genet.* **2013**, *47*, 121–137.

[48] A. Takahashi, N. Ohtani, K. Yamakoshi, S. Iida, H. Tahara, K. Nakayama, K. I. Nakayama, T. Ide, H. Saya, E. Hara, *Nat. Cell Biol.* **2006**, *8*, 1291–1297.

[49] B. M. Kim, H. W. Chung, *Biochem. Biophys. Res. Commun.* **2007**, *363*, 745–750.

[50] T. Ozben, *J. Pharm. Sci.* **2007**, *96*, 2181–2196.

[51] Y.-S. Kim, M. J. Morgan, S. Choksi, Z. Liu, *Mol. Cell* **2007**, *26*, 675–687.

[52] R. Scherz-Shouval, E. Shvets, E. Fass, H. Shorer, L. Gil, Z. Elazar, *EMBO J.* **2007**, *26*, 1749–1760.

[53] R. Scherz-Shouval, Z. Elazar, *Trends Cell Biol.* **2007**, *17*, 422–427.

[54] Roche Laboratories, *Clin. Parmacology Ther.* **1970**, *11*, 617–620.

[55] Y. Ma, J. Zhang, Q. Zhang, P. Chen, J. Song, S. Yu, H. Liu, F. Liu, C. Song, D. Yang, et al., *Biochem. Biophys. Res. Commun.* **2014**, *448*, 8–14.

[56] W. . Tsang, S. P. . Chau, S. . Kong, K. . Fung, T. . Kwok, *Life Sci.* **2003**, *73*, 2047–2058.

[57] K. Dvorakova, C. M. Payne, M. E. Tome, M. M. Briehl, T. McClure, R. T. Dorr, *Biochem. Pharmacol.* **2000**, *60*, 749–758.

[58] B. Bellosillo, N. Villamor, A. López-Guillermo, S. Marcé, J. Esteve, E. Campo, D. Colomer, E. Montserrat, *Blood* **2001**, *98*, 2771–2777.

[59] K. Qu, T. Lin, Z. Wang, S. Liu, H. Chang, X. Xu, F. Meng, L. Zhou, J. Wei, M. Tai, et al., *Front. Med.* **2014**, *8*, 227–235.

[60] X. Wang, S. C. Wong, J. Pan, S. W. Tsao, K. H. Fung, D. L. Kwong, J. S. Sham, J. M. Nicholls, *Cancer Res.* **1998**, *58*, 5019–5022.

[61] M. Donadelli, I. Dando, T. Zaniboni, C. Costanzo, E. Dalla Pozza, M. T. Scupoli, A. Scarpa, S. Zappavigna, M. Marra, A. Abbruzzese, et al., *Cell Death Dis.* **2011**, *2*, e152.

[62] K. Robards, A. F. Kerr, E. Patsalides, *Analyst* **1988**, *113*, 213–224.

[63] S. Jain, M. P. Sharma, *Renew. Sustain. Energy Rev.* **2010**, *14*, 667–678.

[64] H. K. Imdadul, H. H. Masjuki, M. A. Kalam, N. W. M. Zulkifli, M. M. Rashed, H. K. Rashedul, I. M. Monirul, M. H. Mosarof, *RSC Adv* **2015**, *5*, 67541–67567.

References

[65] R. N. Datta, N. M. Huntink, S. Datta, A. G. Talma, *Rubber Chem. Technol.* **2007**, *80*, 436–480.

[66] D. A. Pereira de Abreu, J. M. Cruz, P. Paseiro Losada, *Food Rev. Int.* **2012**, *28*, 146–187.

[67] B. Halliwell, Jm. Gutteridge, *Biochem. J.* **1984**, *219*, 1.

[68] T. P. Flaten, J. Aaseth, O. Andersen, G. J. Kontoghiorghes, *J. Trace Elem. Med. Biol.* **2012**, *26*, 127–130.

[69] P. C. Adams, J. C. Barton, *The Lancet* **2007**, *370*, 1855–1860.

[70] M. Wiggelinkhuizen, M. E. C. Tilanus, C. W. Bollen, R. H. J. Houwen, *Aliment. Pharmacol. Ther.* **2009**, *29*, 947–958.

[71] K. F. Murray, D. Lam, K. V. Kowdley, *Expert Opin. Pharmacother.* **2003**, *4*, 2239–2251.

[72] L. R. Fukumoto, G. Mazza, *J. Agric. Food Chem.* **2000**, *48*, 3597–3604.

[73] B. Tian, Y. Hua, *Food Chem.* **2005**, *91*, 413–418.

[74] J. S. Wright, E. R. Johnson, G. A. DiLabio, *J. Am. Chem. Soc.* **2001**, *123*, 1173–1183.

[75] M. Leopoldini, T. Marino, N. Russo, M. Toscano, *J. Phys. Chem. A* **2004**, *108*, 4916–4922.

[76] D. Farmanzadeh, M. Najafi, *Struct. Chem.* **2015**, *26*, 831–844.

[77] K. Kojiri, H. Kondo, T. Yoshinari, H. Arakawa, S. Nakajima, F. Satoh, K. Kawamura, A. Okura, H. Suda, M. Okanishi, *J. Antibiot. (Tokyo)* **1991**, *44*, 723–729.

[78] H. Nakano, S. Ōmura, *J. Antibiot. (Tokyo)* **2009**, *62*, 17–26.

[79] S. Tanaka, M. Ohkubo, K. Kojiri, H. Suda, *J. Antibiot. (Tokyo)* **1992**, *45*, 1797–1798.

[80] T. Yoshinari, A. Yamada, D. Uemura, K. Nomura, H. Arakawa, K. Kojiri, E. Yoshida, H. Suda, A. Okura, *Cancer Res.* **1993**, *53*, 490–494.

[81] T. Yoshinari, M. Matsumoto, H. Arakawa, H. Okada, K. Noguchi, H. Suda, A. Okura, S. Nishimura, *Cancer Res.* **1995**, *55*, 1310–1315.

[82] T. Yoshinari, M. Ohkubo, K. Fukasawa, S. Egashira, Y. Hara, M. Matsumoto, K. Nakai, H. Arakawa, H. Morishima, S. Nishimura, *Cancer Res.* **1999**, *59*, 4271–4275.

[83] H. Arakawa, M. Morita, T. Kodera, A. Okura, M. Ohkubo, H. Morishima, S. Nishimura, *Cancer Sci.* **1999**, *90*, 1163–1170.

[84] T. Kaneko, H. Wong, J. Utzig, J. Schurig, T. W. Doyle, *J. Antibiot. (Tokyo)* **1990**, *43*, 125–127.

[85] Y. Yamashita, N. Fujii, C. Murakata, T. Ashizawa, M. Okabe, H. Nakano, *Biochemistry (Mosc.)* **1992**, *31*, 12069–12075.

[86] D. E. Nettleton, T. W. Doyle, B. Krishnan, G. K. Matsumoto, J. Clardy, *Tetrahedron Lett.* **1985**, *26*, 4011–4014.

References

[87] S. Léonce, V. Perez, M.-R. Casabianca-Pignede, M. Anstett, E. Bisagni, A. Pierré, G. Atassi, *Invest. New Drugs* **1996**, *14*, 169–180.

[88] S. Le Mée, A. Pierré, J. Markovits, G. Atassi, A. Jacquemin-Sablon, J.-M. Saucier, *Mol. Pharmacol.* **1998**, *53*, 213–220.

[89] P. D. Davis, C. H. Hill, E. Keech, G. Lawton, J. S. Nixon, A. D. Sedgwick, J. Wadsworth, D. Westmacott, S. E. Wikinson, *FEBS Lett.* **1989**, *259*, 61–63.

[90] D. Toullec, P. Pianetti, H. Coste, P. Bellevergue, T. Grand-Perret, M. Ajakane, V. Baudet, P. Soissin, E. Boursier, F. Loriolle, et al., *J. Bilogical Chem.* **1991**, *266*, 15771–15781.

[91] M. M. Behrens, J. L. Martinez, C. Moratilla, J. Renart, *Cell Growth Differ.* **1995**, *6*, 1375–1380.

[92] M. M. Faul, J. R. Gillig, M. R. Jirousek, L. M. Ballas, T. Schotten, A. Kahl, M. Mohr, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1857–1859.

[93] E. Galanis, J. C. Buckner, *J. Clin. Oncol.* **2010**, *28*, 1097–1098.

[94] Alliance for Clinical Trials in Oncology, *Daunorubicin, Cytarabine, and Midostaurin in Treating Patients With Newly Diagnosed Acute Myeloid Leukemia*, ClinicalTrials.gov, Bethesda (MD): National Library of Medicine (US), **2014**.

[95] W. Luo, T. R. Sharif, P. J. Houghton, M. Sharif, *Cell Growth Differ.* **1997**, *8*, 1225–1240.

[96] W. Lin, Y. Wang, S. Lin, C. Li, C. Zhou, S. Wang, H. Huang, P. Liu, G. Ye, X. Shen, *Eur. J. Med. Chem.* **2012**, *47*, 214–220.

[97] L. S. Chin, S. F. Murray, K. M. Zitnay, B. Rami, *Clin. Cancer Res.* **1997**, *3*, 771–776.

[98] J. H. Overmeyer, A. M. Young, H. Bhanot, W. A. Maltese, *Mol Cancer* **2011**, *10*, 69–85.

[99] M. W. Robinson, J. H. Overmeyer, A. M. Young, P. W. Erhardt, W. A. Maltese, *J. Med. Chem.* **2012**, *55*, 1940–1956.

[100] C. J. Trabbic, H. M. Dietsch, E. M. Alexander, P. I. Nagy, M. W. Robinson, J. H. Overmeyer, W. A. Maltese, P. W. Erhardt, *ACS Med. Chem. Lett.* **2014**, *5*, 73–77.

[101] X. Jiang, B. Zhao, R. Britton, *Mol. Cancer Ther.* **2004**, *3*, 1221–1227.

[102] R. Boutros, V. Lobjois, B. Ducommun, *Nat. Rev. Cancer* **2007**, *7*, 495–507.

[103] S. Deslandes, S. Chassaing, E. Delfourne, *Mar. Drugs* **2009**, *7*, 754–786.

[104] S. Mahboobi, H. Pongratz, H. Hufsky, J. Hockemeyer, M. Frieser, A. Lyssenko, D. H. Paper, J. Bürgermeister, F.-D. Böhmer, H.-H. Fiebig, *J. Med. Chem.* **2001**, *44*, 4535–4553.

[105] M. Ohmichi, S. J. Decker, L. Pang, A. R. Saltiel, *Biochemistry (Mosc.)* **1992**, *31*, 4034–4039.

[106] A. M. Camoratto, J. P. Jani, T. S. Angeles, A. C. Maroney, C. Y. Sanders, C. Murakata, N. T. Neef, L. L. Vaught, J. T. Isaacs, C. A. Dionne, *Int. J. Cancer* **1997**, *72*, 673–679.

References

[107]   S. J. Miknyoczki, H. Chang, A. Klein-Szanto, C. A. Dionne, B. A. Ruggeri, *Clin. Cancer Res.* **1999**, *5*, 2205–2212.

[108]   A. E. Evans, K. D. Kisselbach, X. Liu, A. Eggert, N. Ikegaki, A. M. Camoratto, C. A. Dionne, G. M. Brodeur, *Med. Pediatr. Oncol.* **2001**, *36*, 181–184.

[109]   B. Ruggeri, J. Singh, D. Gingrich, T. Angeles, M. Albom, H. Chang, C. Robinson, K. Hunter, P. Dobrzanski, S. Jones-Bolin, et al., *Cancer Res.* **2003**, *63*, 5978–5991.

[110]   R. L. Hudkins, N. C. Becknell, A. L. Zulli, T. L. Underiner, T. S. Angeles, L. D. Aimone, M. S. Albom, H. Chang, S. J. Miknyoczki, K. Hunter, et al., *J. Med. Chem.* **2012**, *55*, 903–913.

[111]   S. Miknyoczki, H. Chang, J. Grobelny, S. Pritchard, C. Worrell, N. McGann, M. Ator, J. Husten, J. Deibold, R. Hudkins, et al., *Mol. Cancer Ther.* **2007**, *6*, 2290–2302.

[112]   J. Deibold, R. L. Hudkins, S. J. Miknyoczki, B. Ruggeri, (Caphalon, Inc.), *Method of Radio-Sensitizing Tumors Using A Radio-Sensitizing Agent*, **2007**, WO2008/63644 A1.

[113]   R. Prudent, V. Moucadel, C.-H. Nguyen, C. Barette, F. Schmidt, J.-C. Florent, L. Lafanechere, C. F. Sautel, E. Duchemin-Pelletier, E. Spreux, et al., *Cancer Res.* **2010**, *70*, 9865–9874.

[114]   S. Lal, T. J. Snape, *Curr. Med. Chem.* **2012**, *19*, 4828–4837.

[115]   G. Murillo, R. G. Mehta, *Nutr. Cancer* **2001**, *41*, 17–28.

[116]   K. R. Grose, L. F. Bjeldanes, *Chem. Res. Toxicol.* **1992**, *5*, 188–193.

[117]   J.-R. Weng, C.-H. Tsai, S. K. Kulp, C.-S. Chen, *Cancer Lett.* **2008**, *262*, 153–163.

[118]   H. L. Bradlow, *In Vivo* **2008**, *22*, 441–445.

[119]   L. M. Howells, B. Gallacher-Horley, C. E. Houghton, M. M. Manson, E. A. Hudson, *Mol. Cancer Ther.* **2002**, *1*, 1161–1172.

[120]   H. R. Frydoonfar, D. R. McGrath, A. D. Spigelman, *Colorectal Dis.* **2002**, *4*, 205–207.

[121]   E. A. Hudson, L. M. Howells, B. Gallacher-Horley, L. H. Fox, A. Gescher, M. M. Manson, *BMC Cancer* **2003**, *3*, 2.

[122]   Q. Zheng, Y. Hirose, N. Yoshimi, A. Murakami, K. Koshimizu, H. Ohigashi, K. Sakata, Y. Matsumoto, Y. Sayama, H. Mori, *J. Cancer Res. Clin. Oncol.* **2002**, *128*, 539–546.

[123]   M. Nachshon-Kedmi, S. Yannai, A. Haj, F. A. Fares, *Food Chem. Toxicol.* **2003**, *41*, 745–752.

[124]   S. R. Chinni, Y. Li, S. Upadhyay, P. K. Koppolu, F. H. Sarkar, *Oncogene* **2001**, *20*, 2927–2936.

[125]   H. R. Frydoonfar, D. R. McGrath, A. D. Spigelman, *ANZ J. Surg.* **2003**, *73*, 154–156.

[126]   D. T. Verhoeven, H. Verhagen, A. R. Goldbohm, P. A. van den Brandt, G. van Poppel, *Chem. Biol. Interact.* **1997**, *103*, 79–129.

References

[127] P. H. Jellinck, P. G. Forkert, D. S. Riddick, A. B. Okey, J. J. Michnovicz, H. L. Bradlow, *Biochem. Pharmacol.* **1993**, *45*, 1129–1136.

[128] C. A. De Kruif, J. W. Marsman, J. C. Venekamp, H. E. Falke, J. Noordhoek, B. J. Blaauboer, H. M. Wortelboer, *Chem. Biol. Interact.* **1991**, *80*, 303–315.

[129] R. E. Staub, C. Feng, B. Onisko, G. S. Bailey, G. L. Firestone, L. F. Bjeldanes, *Chem. Res. Toxicol.* **2002**, *15*, 101–109.

[130] S. Roy, R. Gajbhiye, M. Mandal, C. Pal, A. Meyyapan, J. Mukherjee, P. Jaisankar, *Med. Chem. Res.* **2014**, *23*, 1371–1377.

[131] M. Schieber, N. S. Chandel, *Curr. Biol.* **2014**, *24*, R453–R462.

[132] Y. Iwasaki, T. Hirasawa, Y. Maruyama, Y. Ishii, R. Ito, K. Saito, T. Umemura, A. Nishikawa, H. Nakazawa, *Toxicol. In Vitro* **2011**, *25*, 1320–1327.

[133] G. Galati, O. Sabzevari, J. X. Wilson, P. J. O'Brien, *Toxicology* **2002**, *177*, 91–104.

[134] S. Fujisawa, T. Atsumi, Y. Murakami, Y. Kadoma, *Arch. Immunol. Ther. Exp. (Warsz.)* **2005**, *53*, 28–38.

[135] V. Bindhumol, K. C. Chitra, P. P. Mathur, *Toxicology* **2003**, *188*, 117–124.

[136] L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.

[137] B. Yan, M. Xue, B. Xiong, K. Liu, D. Hu, J. Shen, *Acta Pharmacol. Sin.* **2009**, *30*, 251–258.

[138] B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, *J. Med. Chem.* **1988**, *31*, 2235–2246.

[139] C. D. Duarte, E. J. Barreiro, C. A. Fraga, *Mini Rev. Med. Chem.* **2007**, *7*, 1108–1119.

[140] M. Krasavin, *Eur. J. Med. Chem.* **2015**, *97*, 525–537.

[141] S. Bongarzone, M. L. Bolognesi, *Expert Opin. Drug Discov.* **2011**, *6*, 251–268.

[142] C. Sherer, T. J. Snape, *Eur. J. Med. Chem.* **2015**, *97*, 552–560.

[143] G. S. Singh, Z. Y. Desta, *Chem. Rev.* **2012**, *112*, 6104–6155.

[144] R. S. Keri, S. Budagumpi, R. K. Pai, R. G. Balakrishna, *Eur. J. Med. Chem.* **2014**, *78*, 340–374.

[145] D. A. Horton, G. T. Bourne, M. L. Smythe, *Chem. Rev.* **2003**, *103*, 893–930.

[146] E. Fischer, F. Jourdan, *Berichte Dtsch. Chem. Ges.* **1883**, *16*, 2241–2245.

[147] E. Fischer, O. Hess, *Berichte Dtsch. Chem. Ges.* **1884**, *17*, 559–568.

[148] C. F. H. Allen, C. V. Wilson, *J. Am. Chem. Soc.* **1943**, *65*, 611–612.

[149] G. M. Robinson, R. Robinson, *J. Chem. Soc. Trans.* **1918**, *113*, 639–645.

References

[150]   S. Wagaw, B. H. Yang, S. L. Buchwald, *J. Am. Chem. Soc.* **1998**, *120*, 6621–6622.

[151]   F. R. Japp, F. Klingemann, *Berichte Dtsch. Chem. Ges.* **1887**, *20*, 2942–2844.

[152]   C. Prabhakar, N. V. Kumar, M. R. Reddy, M. R. Sarma, G. O. Reddy, *Org. Process Res. Dev.* **1999**, *3*, 155–160.

[153]   P. R. Brodfuehrer, B.-C. Chen, T. R. Sattelberg, P. R. Smith, J. P. Reddy, D. R. Stark, S. L. Quinlan, J. G. Reid, J. K. Thottathil, S. Wang, *J. Org. Chem.* **1997**, *62*, 9192–9202.

[154]   E. Yasui, M. Wada, N. Takamura, *Tetrahedron Lett.* **2006**, *47*, 743–746.

[155]   K. Maruoka, M. Oishi, H. Yamamoto, *J. Org. Chem.* **1993**, *58*, 7638–7639.

[156]   C. Cao, Y. Shi, A. L. Odom, *Org. Lett.* **2002**, *4*, 2853–2856.

[157]   S. Wagaw, B. H. Yang, S. L. Buchwald, *J. Am. Chem. Soc.* **1999**, *121*, 10251–10263.

[158]   R. D. Stephens, C. E. Castro, *J. Org. Chem.* **1963**, *28*, 3313–3315.

[159]   S. Cacchi, G. Fabrizi, L. M. Parisi, *Org. Lett.* **2003**, *5*, 3843–3846.

[160]   T. Nishikawa, Y. Koide, S. Kajii, K. Wada, M. Ishikawa, M. Isobe, *Org Biomol Chem* **2005**, *3*, 687–700.

[161]   Y. Chen, N. A. Markina, R. C. Larock, *Tetrahedron* **2009**, *65*, 8908–8915.

[162]   N. Sakai, K. Annaka, T. Konakahara, *Org. Lett.* **2004**, *6*, 1527–1530.

[163]   T. Kurisaki, T. Naniwa, H. Yamamoto, H. Imagawa, M. Nishizawa, *Tetrahedron Lett.* **2007**, *48*, 1871–1874.

[164]   B. C. J. van Esseveldt, F. L. van Delft, J. M. M. Smits, R. de Gelder, H. E. Schoemaker, F. P. J. T. Rutjes, *Adv. Synth. Catal.* **2004**, *346*, 823–834.

[165]   M. Amjad, D. W. Knight, *Tetrahedron Lett.* **2004**, *45*, 539–541.

[166]   D. Yue, R. C. Larock, *Org. Lett.* **2004**, *6*, 1037–1040.

[167]   R. C. Larock, E. K. Yum, *J. Am. Chem. Soc.* **1991**, *113*, 6689–6690.

[168]   R. C. Larock, E. K. Yum, M. D. Refvik, *J. Org. Chem.* **1998**, *63*, 7652–7662.

[169]   A. D. Batcho, W. Leimgruber, *Org. Synth.* **1985**, *63*, 214.

[170]   A. D. Batcho, W. Leimgruber, *Process and Intermediates for the Preparation of Indoles from Ortho-Nitrotoluenes*, **1973**, 3,732,245.

[171]   I. Wawer, J. Osek, *J. Chem. Soc. Perkin Trans. 2* **1988**, 993–996.

[172]   M. Ohkubo, T. Nishimura, H. Jona, T. Honma, H. Morishima, *Tetrahedron* **1996**, *52*, 8099–8112.

[173]   A. Reissert, *Berichte Dtsch. Chem. Ges.* **1897**, *30*, 1030–1053.

References

[174]   W. Truce, W. J. Ray, O. L. Norman, D. B. Eickermeyer, *J. Am. Chem. Soc.* **1958**, *80*, 3625–3629.

[175]   A. A. Levy, H. C. Rains, S. Smiles, *J. Chem. Soc. Resumed* **1931**, 3264–3269.

[176]   T. J. Snape, *Chem. Soc. Rev.* **2008**, *37*, 2452–2458.

[177]   T. Snape, *Synlett* **2008**, 2689–2691.

[178]   A. D. Alorati, A. D. Gibb, P. R. Mullens, G. W. Stewart, *Org. Process Res. Dev.* **2012**, *16*, 1947–1952.

[179]   G. Bartoli, G. Palmieri, *Tetrahedron Lett.* **1989**, *30*, 2129–2132.

[180]   S. Kalyaanamoorthy, Y.-P. P. Chen, *Drug Discov. Today* **2011**, *16*, 831–839.

[181]   A. C. Anderson, *Chem. Biol.* **2003**, *10*, 787–797.

[182]   O. A. von Lilienfeld, *Int. J. Quantum Chem.* **2013**, *113*, 1676–1689.

[183]   D. I. Osolodkin, E. V. Radchenko, A. A. Orlov, A. E. Voronkov, V. A. Palyulin, N. S. Zefirov, *Expert Opin. Drug Discov.* **2015**, *10*, 959–973.

[184]   D. Lowe, *Med Chem Commun* **2015**, *6*, 12–12.

[185]   P. G. Polishchuk, T. I. Madzhidov, A. Varnek, *J. Comput. Aided Mol. Des.* **2013**, *27*, 675–679.

[186]   R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* **1996**, *16*, 3–50.

[187]   P. Ertl, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380.

[188]   X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157.

[189]   A. M. Wasik, M. N. Gandy, M. McIldowie, M. J. Holder, A. Chamba, A. Challa, K. D. Lewis, S. P. Young, D. Scheel-Toellner, M. J. Dyer, et al., *Invest. New Drugs* **2012**, *30*, 1471–1483.

[190]   H. Sun, G. Tawa, A. Wallqvist, *Drug Discov. Today* **2012**, *17*, 310–324.

[191]   M. S. Lajiness, G. M. Maggiora, V. Shanmugasundaram, *J. Med. Chem.* **2004**, *47*, 4891–4896.

[192]   P. S. Kutchukian, N. Y. Vasilyeva, J. Xu, M. K. Lindvall, M. P. Dillon, M. Glick, J. D. Coley, N. Brooijmans, *PLoS ONE* **2012**, *7*, e48476.

[193]   D. S. Wishart, *Nucleic Acids Res.* **2006**, *34*, D668–D672.

[194]   D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, *Nucleic Acids Res.* **2007**, *36*, D901–D906.

[195]   C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, et al., *Nucleic Acids Res.* **2011**, *39*, D1035–D1041.

References

[196]   X. Chen, Y. Lin, M. K. Gilson, *Biopolymers* **2001**, *61*, 127–141.

[197]   X. Chen, Y. Lin, M. Liu, M. K. Gilson, *Bioinformatics* **2002**, *18*, 130–139.

[198]   X. Chen, M. Liu, M. K. Gilson, *Comb. Chem. High Throughput Screen.* **2001**, *4*, 719–725.

[199]   M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, *Nucleic Acids Res.* **2016**, *44*, D1045–D1053.

[200]   T. Liu, Y. Lin, X. Wen, R. N. Jorissen, M. K. Gilson, *Nucleic Acids Res.* **2007**, *35*, D198–D201.

[201]   A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, et al., *Nucleic Acids Res.* **2014**, *42*, D1083–D1090.

[202]   A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, et al., *Nucleic Acids Res.* **2012**, *40*, D1100–D1107.

[203]   R. A. Carr, M. Congreve, C. W. Murray, D. C. Rees, *Drug Discov. Today* **2005**, *10*, 987–992.

[204]   S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, et al., *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

[205]   Q. Li, T. Cheng, Y. Wang, S. H. Bryant, *Drug Discov. Today* **2010**, *15*, 1052–1057.

[206]   Y. Wang, T. Suzek, J. Zhang, J. Wang, S. He, T. Cheng, B. A. Shoemaker, A. Gindulyte, S. H. Bryant, *Nucleic Acids Res.* **2014**, *42*, D1075–D1082.

[207]   Y. Wang, E. Bolton, S. Dracheva, K. Karapetyan, B. A. Shoemaker, T. O. Suzek, J. Wang, J. Xiao, J. Zhang, S. H. Bryant, *Nucleic Acids Res.* **2010**, *38*, D255–D266.

[208]   Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, Z. Zhou, L. Han, K. Karapetyan, S. Dracheva, B. A. Shoemaker, et al., *Nucleic Acids Res.* **2012**, *40*, D400–D412.

[209]   Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, S. H. Bryant, *Nucleic Acids Res.* **2009**, *37*, W623–W633.

[210]   J. R. Garritano, *Sci. Technol. Libr.* **2013**, *32*, 346–371.

[211]   L. Ruddigkeit, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 56–65.

[212]   P. K. Halen, P. R. Murumkar, R. Giridhar, M. R. Yadav, *Mini Rev. Med. Chem.* **2009**, *9*, 124–139.

[213]   K. J. Bergmann, *Clin. Neuropharmacol.* **1985**, *8*, 13–26.

[214]   G. W. Dawson, S. G. Jue, R. N. Brogden, *Drugs* **1984**, *27*, 132–147.

[215]   P. Linda, A. Stener, A. Cipiciani, G. Savelli, *J. Heterocycl. Chem.* **1983**, *20*, 247.

[216]   S. Sekiguchi, J. F. Bunnett, *J. Am. Chem. Soc.* **1981**, *103*, 4871–4874.

References

[217]   A. Chaudhuri, S. Haldar, H. Sun, R. E. Koeppe, A. Chattopadhyay, *BBA - Biomembr.* **2014**, *1838*, 419–428.

[218]   Y. Yamamoto, A. Wakisaka, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 1405–1408.

[219]   T. Montoro, C. Jouvet, A. Lopez-Campillo, B. Soep, *J. Phys. Chem.* **1983**, *87*, 3582–3584.

[220]   K. M. Huttunen, H. Raunio, J. Rautio, *Pharmacol. Rev.* **2011**, *63*, 750–771.

[221]   M. B. Frampton, R. Simionescu, T. Dudding, P. M. Zelisko, *J. Mol. Catal. B Enzym.* **2010**, *66*, 105–112.

[222]   A. Maraite, M. B. Ansorge-Schumacher, B. Ganchegui, W. Leitner, G. Grogan, *J. Mol. Catal. B Enzym.* **2009**, *56*, 24–28.

[223]   Y. Liu, W. W. McWhorter, *J. Am. Chem. Soc.* **2003**, *125*, 4240–4252.

[224]   Y.-H. Kwak, S.-H. Hwang, Y.-K. Kim, H. Jung, J. Lee, J.-O. Lim, H.-J. Song, *Organic Light Emitting Device Containing Polycyclic Heteroaryl Compounds*, **2012**, EP 2 298 739 B1.

[225]   P. E. Eaton, G. R. Carlson, J. T. Lee, *J. Org. Chem.* **1973**, *38*, 4071–4073.

[226]   D. Zhao, D. L. Hughes, D. R. Bender, A. M. DeMarco, P. J. Reider, *J. Org. Chem.* **1991**, *56*, 3001–3006.

[227]   E. Siddalingamurthy, K. M. Mahadevan, J. N. Masagalli, H. N. Harishkumar, *Tetrahedron Lett.* **2013**, *54*, 5591–5596.

[228]   A. Rybina, B. Thaler, R. Krämer, D.-P. Herten, *Phys. Chem. Chem. Phys.* **2014**, *16*, 19550.

[229]   C. Jiang, S. Garg, T. D. Waite, *Environ. Sci. Technol.* **2015**, *49*, 14076–14084.

[230]   B. Robinson, M. Uppal Zubair, *J. Chem. Soc. C* **1971**, 976–977.

[231]   S. J. Coles, P. A. Gale, *Chem Sci* **2012**, *3*, 683–689.

[232]   W.-R. Chao, D. Yean, K. Amin, C. Green, L. Jong, *J. Med. Chem.* **2007**, *50*, 3412–3415.

[233]   E. C. Creencia, M. Tsukamoto, T. Horaguchi, *J. Heterocycl. Chem.* **2011**, *48*, 1095–1102.

[234]   S. J. Teague, A. M. Davis, P. D. Leeson, T. Oprea, *Angew. Chem. Int. Ed.* **1999**, *38*, 3743–3748.

[235]   C. A. Lipinski, *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249.

[236]   Y.-C. Chen, *Trends Pharmacol. Sci.* **2015**, *36*, 78–95.

[237]   M. J. Vainio, J. S. Puranen, M. S. Johnson, *J. Chem. Inf. Model.* **2009**, *49*, 492–502.

[238]   M. J. Vainio, M. S. Johnson, *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.

[239]   W.-T. Wu, L. Zhang, S.-L. You, *Chem Soc Rev* **2016**, *45*, 1570–1580.

[240]   T. Kano, Y. Ohyabu, S. Saito, H. Yamamoto, *J. Am. Chem. Soc.* **2002**, *124*, 5365–5373.

References

[241]   H.-J. Böhm, A. Flohr, M. Stahl, *Drug Discov. Today Technol.* **2004**, *1*, 217–224.

[242]   M. Gaba, S. Singh, C. Mohan, *Eur. J. Med. Chem.* **2014**, *76*, 494–505.

[243]   P. Zhan, D. Li, J. Li, X. Chen, X. Liu, *Mini-Rev. Org. Chem.* **2012**, *9*, 397–410.

[244]   Y. Iwasaki, M. Oda, Y. Tsukuda, Y. Nagamori, H. Nakazawa, R. Ito, K. Saito, *J. Food Hyg. Soc. Jpn.* **2014**, *55*, 167–176.

[245]   *Schrödinger*, LLC, New York, NY, **2016**.

[246]   E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, et al., *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

[247]   A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, R. A. Friesner, *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142.

[248]   M. Najafi, M. Najafi, H. Najafi, *Can. J. Chem.* **2012**, *90*, 915–926.

[249]   H.-Y. Zhang, Y.-M. Sun, X.-L. Wang, *Chem. Eur. J.* **2003**, *9*, 502–508.

[250]   M. Najafi, M. Najafi, H. Najafi, *Bull. Korean Chem. Soc.* **2012**, *33*, 3607–3617.

[251]   M. Najafi, E. Nazarparvar, K. H. Mood, M. Zahedi, E. Klein, *Comput. Theor. Chem.* **2011**, *965*, 114–122.

[252]   E. Klein, V. Lukeš, *J. Mol. Struct. THEOCHEM* **2006**, *767*, 43–50.

[253]   P. Košinová, F. Di Meo, E. H. Anouar, J.-L. Duroux, P. Trouillas, *Int. J. Quantum Chem.* **2011**, *111*, 1131–1142.

[254]   A. D. Becke, *Phys. Rev. B* **1988**, *38*, 3098–3100.

[255]   C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.

[256]   A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.

[257]   J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.

[258]   P. C. Hariharan, J. A. Pople, *Mol. Phys.* **1974**, *27*, 209–214.

[259]   W. J. Hehre, *J. Chem. Phys.* **1972**, *56*, 2257.

[260]   R. Ditchfield, *J. Chem. Phys.* **1971**, *54*, 724.

[261]   M. S. Gordon, *Chem. Phys. Lett.* **1980**, *76*, 163–168.

[262]   P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta Berl.* **1973**, *28*, 213–222.

[263]   V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, L. A. Curtiss, *J. Comput. Chem.* **2001**, *22*, 976–984.

[264]   V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, *J. Chem. Phys.* **1998**, *109*, 1223.

References

[265]   J.-P. Blaudeau, M. P. McGrath, L. A. Curtiss, L. Radom, *J. Chem. Phys.* **1997**, *107*, 5016.

[266]   R. C. Binning, L. A. Curtiss, *J. Comput. Chem.* **1990**, *11*, 1206–1216.

[267]   M. M. Francl, *J. Chem. Phys.* **1982**, *77*, 3654.

[268]   A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639.

[269]   M. P. McGrath, L. Radom, *J. Chem. Phys.* **1991**, *94*, 511.

[270]   K. Raghavachari, G. W. Trucks, *J. Chem. Phys.* **1989**, *91*, 1062.

[271]   P. J. Hay, *J. Chem. Phys.* **1977**, *66*, 4377.

[272]   A. J. H. Wachters, *J. Chem. Phys.* **1970**, *52*, 1033.

[273]   R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

[274]   R. M. Borges dos Santos, J. A. Martinho Simões, *J. Phys. Chem. Ref. Data* **1998**, *27*, 707.

[275]   G. Brigati, M. Lucarini, V. Mugnaini, G. F. Pedulli, *J. Org. Chem.* **2002**, *67*, 4828–4832.

[276]   L. R. Mahoney, M. A. DaRooge, *J. Am. Chem. Soc.* **1975**, *97*, 4722–4731.

[277]   J. A. K. Howard, V. J. Hoy, D. O'Hagan, G. T. Smith, *Tetrahedron* **1996**, *52*, 12613–12622.

[278]   M. Najafi, M. Zahedi, E. Klein, *Comput. Theor. Chem.* **2011**, *978*, 16–28.

[279]   J. Đorović, J. M. D. Marković, V. Stepanić, N. Begović, D. Amić, Z. Marković, *J. Mol. Model.* **2014**, *20*, DOI 10.1007/s00894-014-2345-y.

[280]   Z. S. Marković, J. M. Dimitrić Marković, D. Milenković, N. Filipović, *J. Mol. Model.* **2011**, *17*, 2575–2584.

[281]   P. Trouillas, P. Marsal, D. Siri, R. Lazzaroni, J.-L. Duroux, *Food Chem.* **2006**, *97*, 679–688.

[282]   C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **1991**, *91*, 165–195.

[283]   E. Bendary, R. R. Francis, H. M. G. Ali, M. I. Sarwat, S. El Hady, *Ann. Agric. Sci.* **2013**, *58*, 173–181.

[284]   P. Skehan, R. Storeng, D. Scudiero, A. Monks, J. McMahon, D. Vistica, J. T. Warren, B. Heidi, S. Kenney, M. R. Boyd, *J. Natl. Cancer Inst.* **1990**, *82*, 1107–1112.

[285]   V. L. Jacobs, P. A. Valdes, W. F. Hickey, J. A. De Leo, *ASN NEURO* **2011**, *3*, 171–181.

[286]   J.-R. Weng, C.-H. Tsai, S. K. Kulp, D. Wang, C.-H. Lin, H.-C. Yang, Y. Ma, A. Sargeant, C.-F. Chiu, M.-H. Tsai, *Cancer Res.* **2007**, *67*, 7815–7824.

[287]   A. Martinez, M.-V. Vazquez, J. L. Carreon-Macedo, L. E. Sansores, R. Salcedo, *ChemInform* **2003**, *34*, DOI 10.1002/chin.200348033.

[288]   L. H. Yao, Y. M. Jiang, J. SHI, F. A. Tomas-Barberan, N. Datta, R. Singanusong, S. S. Chen, *Plant Foods Hum. Nutr.* **2004**, *59*, 113–122.

References

[289]   J. Dai, R. J. Mumper, *Molecules* **2010**, *15*, 7313–7352.

[290]   E. I. Parkinson, P. J. Hergenrother, *Acc. Chem. Res.* **2015**, *48*, 2715–2723.

[291]   P. J. O'brien, *Chem. Biol. Interact.* **1991**, *80*, 1–41.

[292]   J. D. Groarke, A. Nohria, *Circulation* **2015**, *131*, 946–1949.

[293]   R. W. Owen, A. Giacosa, W. E. Hull, R. Haubner, B. Spiegelhalder, H. Bartsch, *Eur. J. Cancer* **2000**, *36*, 1235–1247.

[294]   C. A. Gomes, T. Girão da Cruz, J. L. Andrade, N. Milhazes, F. Borges, M. P. M. Marques, *J. Med. Chem.* **2003**, *46*, 5395–5401.

[295]   S. Gately, R. West, *Drug Dev. Res.* **2007**, *68*, 156–163.

[296]   G. A. Showell, J. S. Mills, *Drug Discov. Today* **2003**, *8*, 551–556.

[297]   J. S. Mills, G. A. Showell, *Expert Opin. Investig. Drugs* **2004**, *13*, 1149–1157.

[298]   T. Satoh, M. Hosokawa, *Annu. Rev. Pharmacol. Toxicol.* **1998**, *38*, 257–288.

[299]   E. Barbayianni, I. Fotakopoulou, M. Schmidt, V. Constantinou-Kokotou, U. T. Bornscheuer, G. Kokotos, *J. Org. Chem.* **2005**, *70*, 8730–8733.

[300]   R. A. Cairns, I. S. Harris, T. W. Mak, *Nat. Rev. Cancer* **2011**, *11*, 85–95.

[301]   J. Lee, J.-I. Choe, S. Jung, S. W. Ham, *Bull. Korean Chem. Soc.* **2006**, *27*, 33–34.

[302]   G. Cohen, R. E. Heikkila, *J. Biol. Chem.* **1974**, *249*, 2447–2452.

[303]   A. H. Jackson, N. Prasitpan, P. V. R. Shannon, A. C. Tinker, *J. Chem. Soc. [Perkin 1]* **1987**, 2543.

[304]   C. Liu, W. Zhang, L.-X. Dai, S.-L. You, *Org. Lett.* **2012**, *14*, 4525–4527.

[305]   D. E. Ames, R. E. Bowman, D. D. Evans, W. A. Jones, *J. Chem. Soc. Resumed* **1956**, 1984–1989.

[306]   R. A. Fernandes, P. Kumar, *Tetrahedron Lett.* **2003**, *44*, 1275–1278.

[307]   C. Liu, Y. Zhang, N. Liu, J. Qiu, *Green Chem.* **2012**, *14*, 2999.

[308]   X.-H. Xu, M. Taniguchi, A. Azuma, G. K. Liu, E. Tokunaga, N. Shibata, *Org. Lett.* **2013**, *15*, 686–689.

[309]   J. E. Bartmess, *J. Phys. Chem.* **1994**, *98*, 6420–6424.

[310]   G. M. Lewandowicz, B. Harding, W. Harkness, R. Hayward, D. G. T. Thomas, J. L. Darling, *Eur. J. Cancer* **2000**, *36*, 1955–1964.

[311]   W. Voigt, in *Chemosensitivity*, Springer, **2005**, pp. 39–48.

[312]   L. Palatinus, G. Chapuis, *J. Appl. Crystallogr.* **2007**, *40*, 786–790.

References

[313]   G. M. Sheldrick, *Acta Crystallogr. A* **2008**, *64*, 112–122.

APPENDICIES

Appendix 1 – Raw data and calculations used in the determination of appropriate microwave irradiation time in the microwave-assisted Fischer indole synthesis of 2-arylindoles

As discussed in Section 4.4.3, a method for the microwave-assisted Fischer indole synthesis was required. The methodology was developed while synthesising compound **186** (Figure 96).
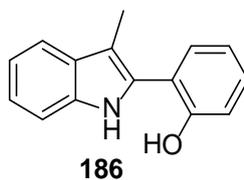
*Figure 96 - Compound **186***

Reactions were run at 160 °C and 200 °C for between one and 50 minutes, and the product yield was determined by GC-MS, using 2-phenylindole as an internal standard. The relative peak area of compound **186** to 2-phenylindole was found to be 1.34. The raw data for the methodological work on this reaction is shown in Table 29 and Table 30.

T = 160 °C	Peak Area		Normalisation Factor	Normalised Area		
Time (mins)	Standard	Product		Standard	Product	Yield (%)
1	1.39×10^9	3.93×10^7	1.34	1.39×10^9	5.27×10^7	3
2	1.15×10^9	6.31×10^7	1.34	1.15×10^9	8.46×10^7	5
5	1.24×10^9	1.42×10^8	1.34	1.24×10^9	1.90×10^7	11
10	1.22×10^9	2.05×10^8	1.34	1.22×10^9	2.75×10^8	17
20	1.39×10^9	4.25×10^8	1.34	1.39×10^9	5.70×10^8	31

Table 29 - Table of peak areas from GC-MS, and the yield calculation at 160 °C

T = 200 °C Time (mins)	Peak Area		Normalisation Factor	Normalised Area		Yield (%)
	Standard	Product		Standard	Product	
1	2.36×10^9	8.70×10^8	1.34	2.36×10^9	1.17×10^9	49
2	2.73×10^9	1.33×10^9	1.34	2.73×10^9	1.78×10^9	65
5	2.99×10^9	1.62×10^9	1.34	2.99×10^9	2.17×10^9	73
10	3.40×10^9	1.98×10^9	1.34	3.40×10^9	2.65×10^9	78
20	2.06×10^9	1.33×10^9	1.34	2.06×10^9	1.78×10^9	86
30	7.76×10^8	5.18×10^8	1.34	7.76×10^8	6.94×10^8	90
40	9.17×10^8	5.94×10^8	1.34	9.17×10^8	7.96×10^8	87
50	1.07×10^8	7.40×10^8	1.34	1.07×10^9	9.92×10^8	93

Table 30 - Table of peak areas from GC-MS, and the yield calculation at 200 °C

Appendix 2 – Analogues of compound 4 found as part of the similarity search
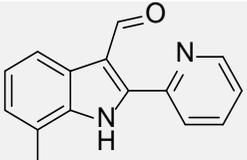
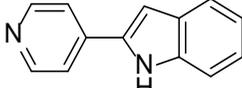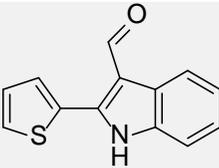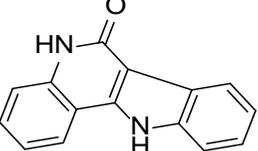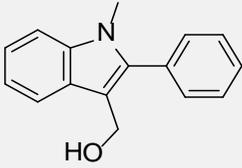| Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC45235565	0.765	0.633	0.898				×
	ZINC00159358	0.766	0.660	0.873			×	
	ZINC01025941	0.766	0.607	0.925				×
	ZINC02144604	0.766	0.607	0.926				×
	ZINC37401152	0.766	0.603	0.930		×		
	ZINC00144326	0.767	0.639	0.895		×		
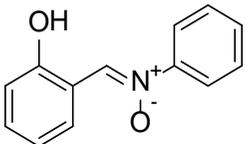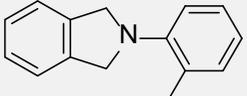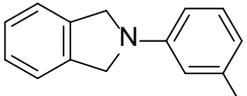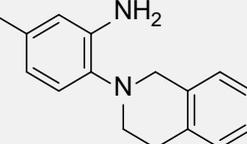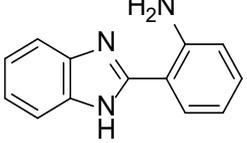
A5
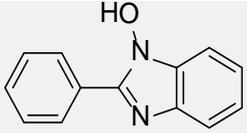
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC01391331	0.767	0.617	0.918		×		
	ZINC00397633	0.767	0.595	0.940				×
	ZINC00236875	0.768	0.611	0.924		×		
	ZINC02538818	0.768	0.617	0.919		×		
	ZINC71711321	0.768	0.617	0.919				×
	ZINC13403054	0.768	0.645	0.890				×
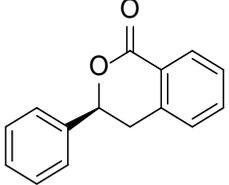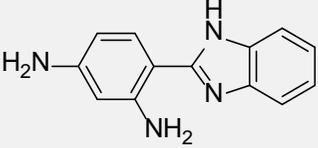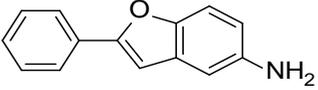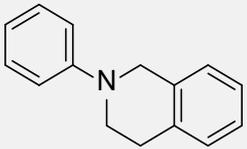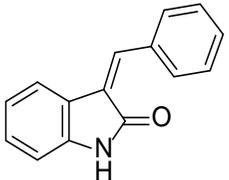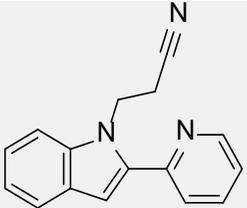
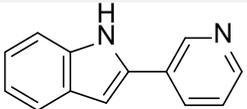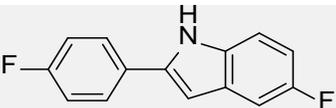| Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC13194625	0.768	0.638	0.898				×
	ZINC00507751	0.769	0.616	0.921				×
	ZINC00122913	0.769	0.618	0.919		×		
	ZINC03847971	0.769	0.591	0.947				×
	ZINC00478097	0.769	0.601	0.937				×
	ZINC65339301	0.769	0.600	0.938		×		
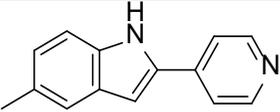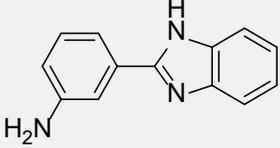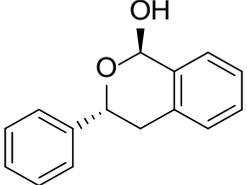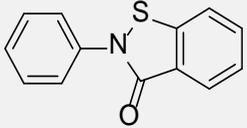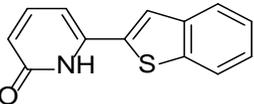
A7
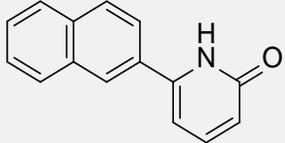
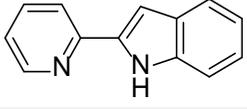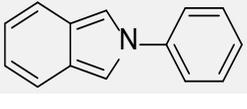| Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC65339268	0.769	0.624	0.914		×		
	ZINC46085466	0.770	0.598	0.941				×
	ZINC32099770	0.770	0.641	0.899		×		
	ZINC00246170	0.770	0.618	0.921		×		
	ZINC01716698	0.770	0.670	0.870			×	
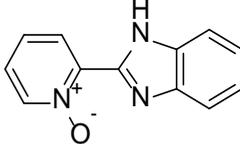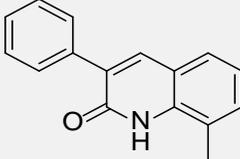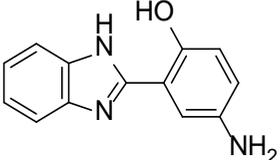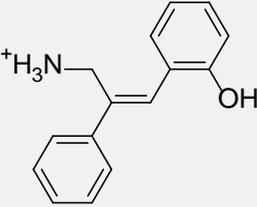
A8
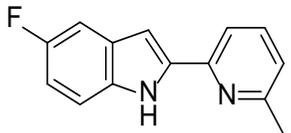
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC95835619	0.770	0.638	0.902				×
	ZINC79387317	0.770	0.618	0.923				×
	ZINC23625305	0.771	0.599	0.942				×
	ZINC65339264	0.771	0.600	0.942		×		
	ZINC40450030	0.771	0.618	0.924		×		
	ZINC00056346	0.771	0.593	0.949		×		
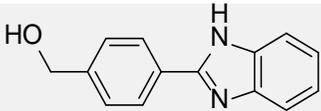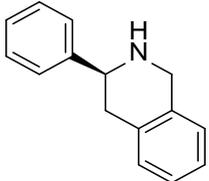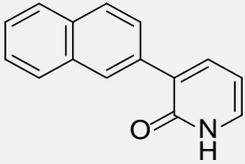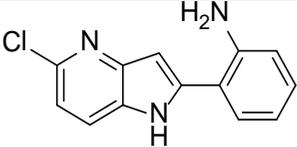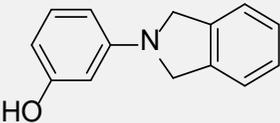
A9
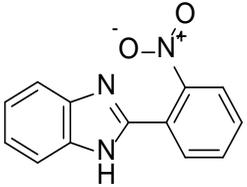
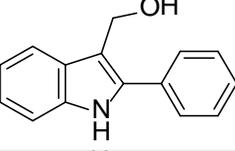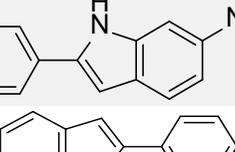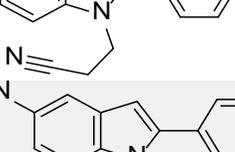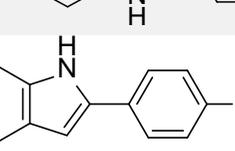| Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC00188622	0.771	0.645	0.897				×
	ZINC65490232	0.771	0.601	0.941		×		
	ZINC06567825	0.771	0.575	0.968			×	
	ZINC04824058	0.772	0.622	0.921				×
	ZINC66091588	0.772	0.636	0.908				×
	ZINC03881190	0.772	0.623	0.920		×		
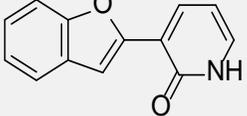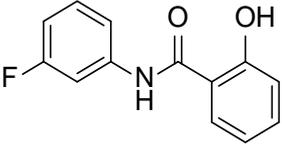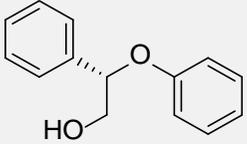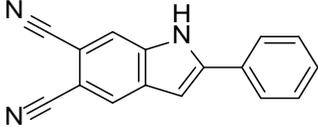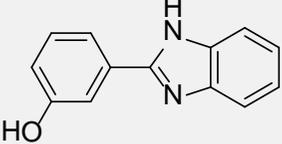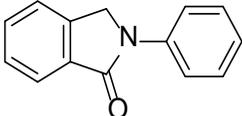
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC00364176	0.772	0.606	0.939				×
	ZINC00329431	0.772	0.607	0.937				×
	ZINC00080737	0.772	0.682	0.863				×
	ZINC13284363	0.773	0.628	0.918				×
	ZINC00148440	0.775	0.609	0.941		×		
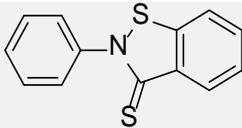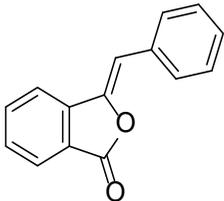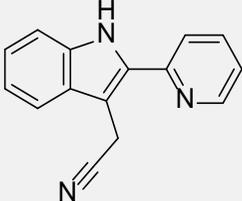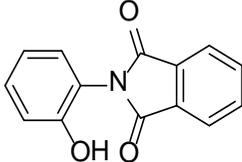
A10
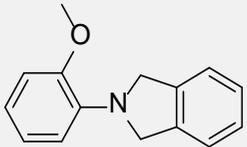
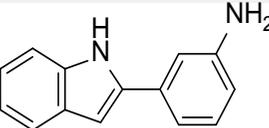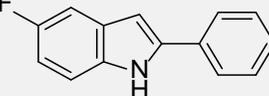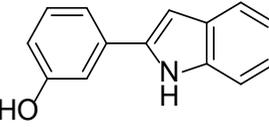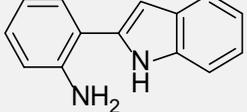| Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC00559647	0.776	0.624	0.928				×
	ZINC12350724	0.777	0.611	0.943		×		
	ZINC00393231	0.777	0.665	0.889				×
	ZINC01705340	0.778	0.646	0.909				×
	ZINC14001405	0.778	0.676	0.881				×
	ZINC05480960	0.778	0.609	0.947				×
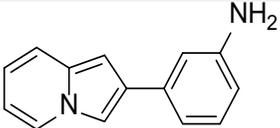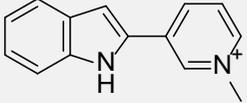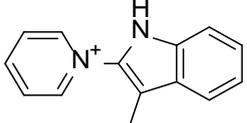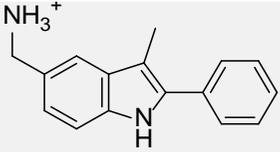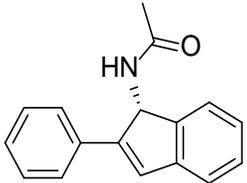
A11

A12
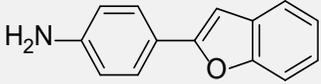
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC00122320	0.778	0.638	0.918		×		
	ZINC34268568	0.778	0.635	0.922				×
	ZINC15084532	0.779	0.616	0.941				×
	ZINC39232949	0.779	0.650	0.908				×
	ZINC04086135	0.779	0.658	0.901	×			
	ZINC00166610	0.779	0.665	0.893		×		
	ZINC05775431	0.779	0.695	0.864				×
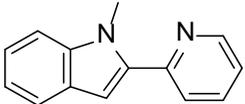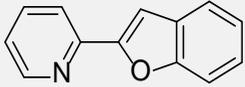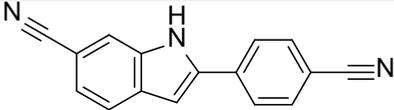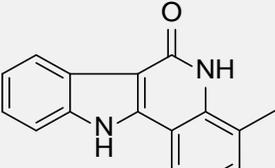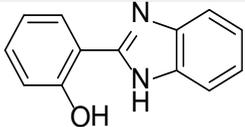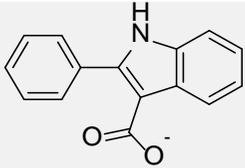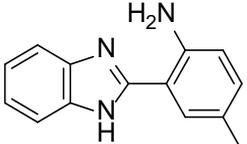
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC19486558	0.780	0.615	0.945		×		
	ZINC00050693	0.781	0.642	0.920				×
	ZINC38918044	0.782	0.597	0.967		×		
	ZINC02505847	0.783	0.623	0.943			×	
	ZINC05837089	0.783	0.666	0.901		×		
	ZINC02534051	0.784	0.653	0.915				×
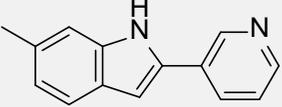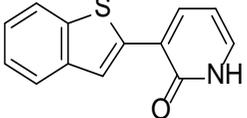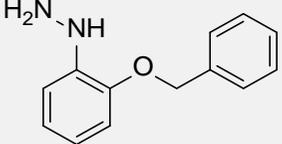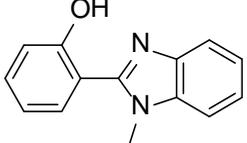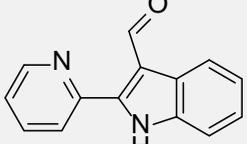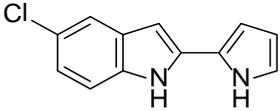
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC66055026	0.784	0.625	0.942				×
	ZINC22015465	0.784	0.629	0.940				×
	ZINC00406909	0.785	0.649	0.920			×	
	ZINC05105345	0.785	0.674	0.895				×
	ZINC00103501	0.786	0.706	0.866				×
	ZINC65339297	0.786	0.603	0.970		×		
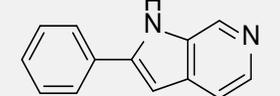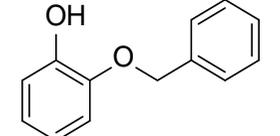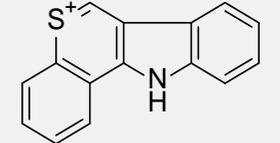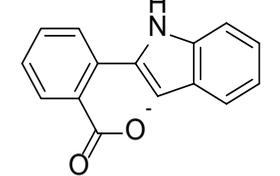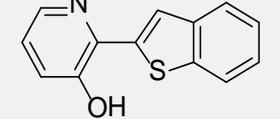
A14
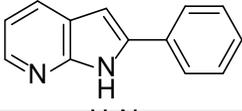
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC32233020	0.787	0.635	0.940				×
	ZINC39119558	0.788	0.650	0.925		×		
	ZINC33884628	0.788	0.608	0.967				×
	ZINC05480972	0.789	0.629	0.948				×
	ZINC22015465	0.789	0.645	0.933				×
	ZINC01518176	0.789	0.652	0.926				×
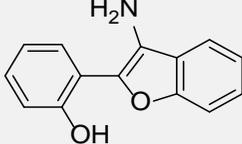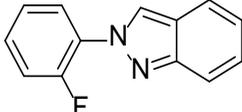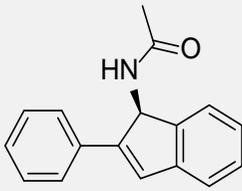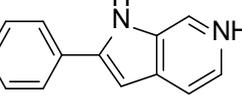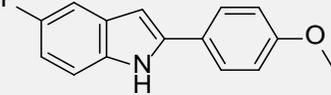
A15
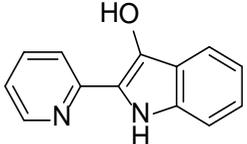
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC03845012	0.790	0.635	0.946		×		
	ZINC05117514	0.792	0.593	0.991				×
	ZINC00501252	0.792	0.681	0.903	×			
	ZINC13285077	0.792	0.639	0.945		×		
	ZINC15445928	0.793	0.693	0.893				×
	ZINC40448330	0.793	0.645	0.942				×
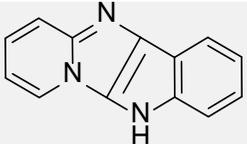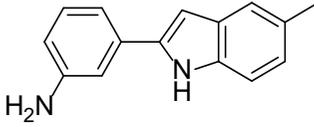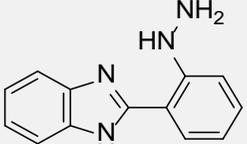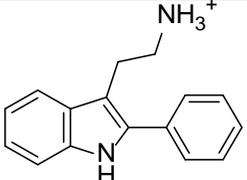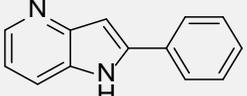
A17
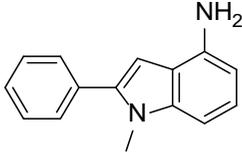
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC35325036	0.794	0.684	0.903	×			
	ZINC05777096	0.794	0.642	0.947				×
	ZINC01519035	0.794	0.646	0.942				×
	ZINC19724840	0.795	0.647	0.942				×
	ZINC03944854	0.796	0.674	0.918	×			
	ZINC08707999	0.797	0.689	0.904				×
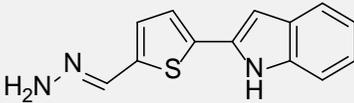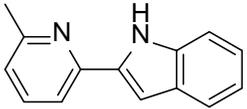
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC95831280	0.797	0.671	0.923				×
	ZINC40448330	0.798	0.663	0.932				×
	ZINC01516739	0.798	0.673	0.923				×
	ZINC98233994	0.799	0.658	0.939				×
	ZINC00050695	0.799	0.657	0.941				×
	ZINC01518212	0.800	0.654	0.946				×
	ZINC00050716	0.801	0.659	0.942				×
	ZINC39241206	0.806	0.616	0.996				×
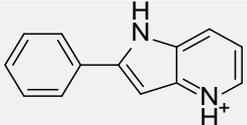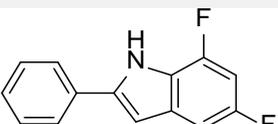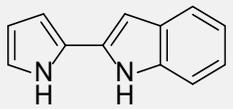
A18
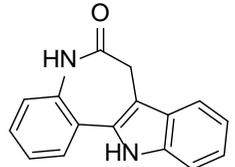
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC01626613	0.808	0.718	0.898	×			
	ZINC03844886	0.810	0.699	0.921	×			
	ZINC08752894	0.812	0.706	0.919				×
	ZINC34543816	0.818	0.691	0.945	×			
	ZINC02495043	0.820	0.693	0.946				×
	ZINC33978742	0.820	0.672	0.968	×			
	ZINC37632465	0.821	0.698	0.944	×			
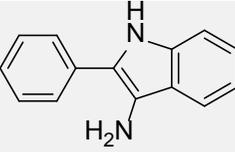
A20
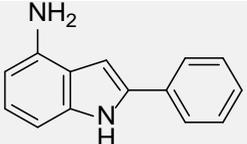
Structure	ZINC Identifier	Similarity	Shape	Electronic	Isomers and pseudoisomers	Different core	Very different	No relevant functional group
	ZINC35325033	0.822	0.699	0.945	×			
	ZINC89203406	0.823	0.701	0.945	×			
	ZINC01708901	0.825	0.703	0.946	×			
	ZINC01663940	0.826	0.708	0.945	×			
	ZINC01518159	0.828	0.713	0.944				×
	ZINC12356961	0.835	0.725	0.944	×			
	ZINC00154770	0.874	0.752	0.996	×			
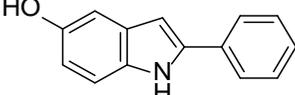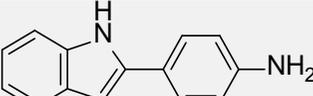
Appendix 3 – Determination of enthalpic values: a worked example

This appendix describes the full process for the determination of enthalpic values used in this thesis, using compound **4** as a worked example.

Generating 3D structures of likely ionisation states

To begin with, the likely ionisation states of each compound needed to be determined, and an approximate 3D structure for each of these ionisation states would be required as a starting point for the following step. To do this, 2D structures were imported into LigPrep from ChemDraw as .cdx files, which would then output probable ionisation states at a specified pH of 7.0 for each compound as 3D structures (Figure 97).
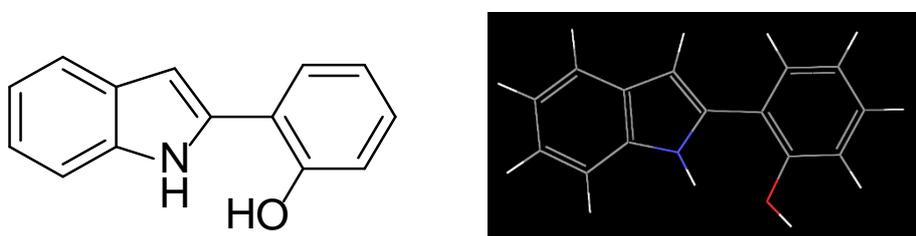
Figure 97 - The process of generating the 3D structures of likely ionisation state(s). Construction of 2D structure in Chemdraw (left) and the outputted 3D structure of the likely ionisation state(s) (right)

In the case of multiple ionisation states being generated, all potential ionisation states within 20 kcal mol^{-1} of the lowest energy state (as determined by LigPrep) were taken forward.

Generating plausible conformers as a starting point for the optimisation step

Now that the most probable ionisation states of each compound are known, the process of optimising their geometry can begin. The first part of this process is producing a number of reasonable conformers from which the geometry optimisation can begin. Multiple conformers were used as starting points in order to improve the chances of finding the global minima.

MacroModel was used to generate these conformers from the 3D structures outputted by LigPrep, as shown in Figure 98, and any conformer within 15 kcal mol^{-1} of the lowest energy conformer (as determined by MacroModel) were taken forward as starting points in the geometry optimisation steps. Typically, between one and five conformers taken forward to the next step. In the case of compound **4**, four conformers were found, three of which were within 15 kcal mol^{-1} of the lowest energy conformation. Therefore, these three conformers were used as starting points in the geometry optimisation steps.
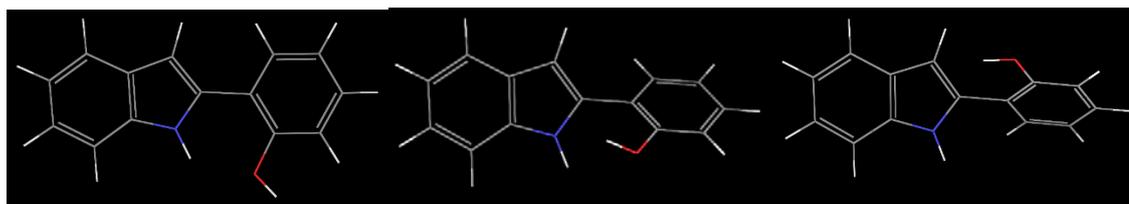
*Figure 98 - The three conformers of compound **4** taken forward as starting points in the geometry optimisation steps*

These conformers were used as starting points for the neutral and radical cationic species. For the radical and anionic species, the relevant proton was manually removed.

Geometry optimisation

The geometries of the neutral, radical cationic, anionic and radical species from the previous step were optimised using Jaguar. Whether true minima had been found was determined by the absence of imaginary frequencies. If imaginary frequencies were found, they were manually removed, and the optimisation was restarted. The energy and geometry of each species was taken to be that of the lowest energy conformer (Figure 99).

Neutral

Radical

Anionic
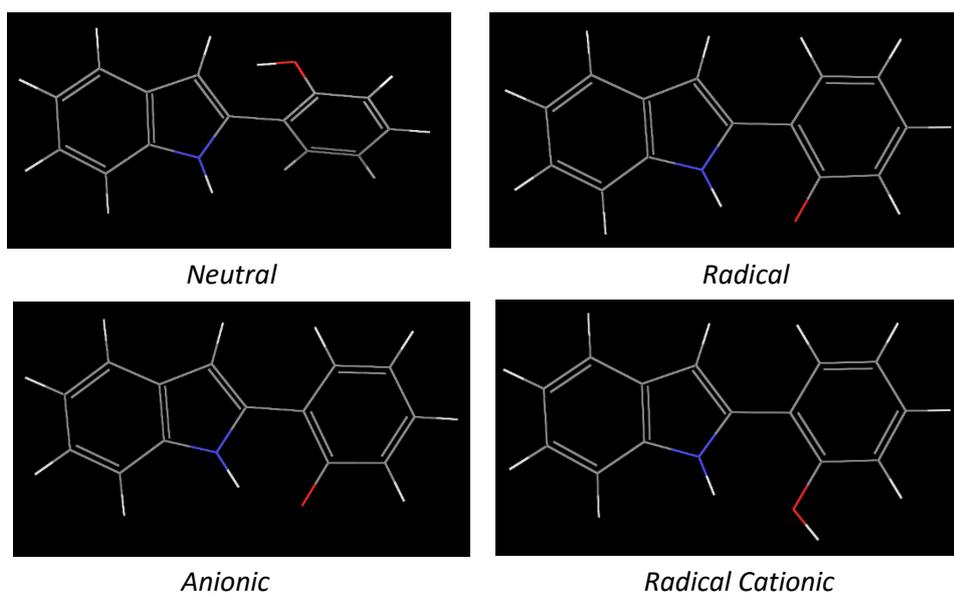
Radical Cationic

*Figure 99 - The optimised geometries of the neutral, radical, anionic and radical cationic forms of compound **4***

Calculation of enthalpic values

The enthalpic values were calculated as follows:

$$\text{BDE} = H(\text{R}^\bullet) + H(\text{H}^\bullet) - H(\text{RH})$$
$$\text{AIP} = H(\text{RH}^{+\bullet}) + H(\text{e}^-) - H(\text{RH})$$
$$\text{PDE} = H(\text{R}^\bullet) + H(\text{H}^+) - H(\text{RH}^{+\bullet})$$
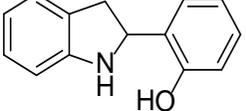$$\text{PA} = H(\text{R}^-) + H(\text{H}^+) - H(\text{RH})$$
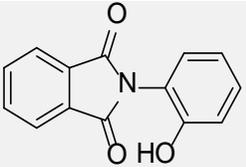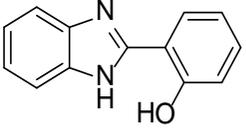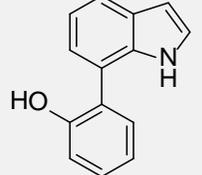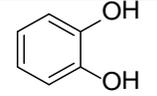$$\text{ETE} = H(\text{R}^\bullet) + H(\text{e}^-) - H(\text{R}^-)$$
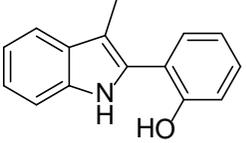
where "H" indicates the enthalpy of a species. The enthalpy of the hydrogen radical was calculated as -325.4 kcal mol^{-1}, and literature values for the enthalpies of the proton and electron were used.[309]
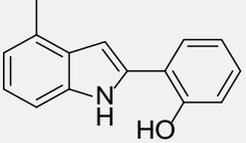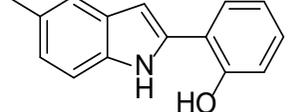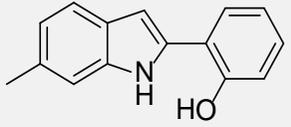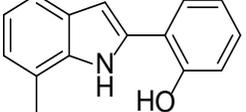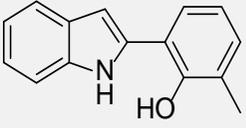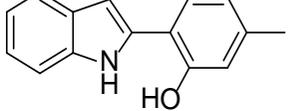
Appendix 4 – Total BDE, AIP, PDE, PA and ETE data
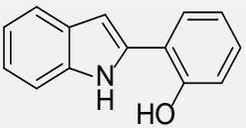
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	4	77.1	173.4	231.3	327.0	77.7	0.0	0.0	0.0	0.0	0.0
	228	97.3	176.5	248.4	344.0	80.9	20.1	3.1	17.1	16.9	3.2
	262	71.5	175.5	223.6	326.6	72.5	-5.6	2.1	-7.7	-0.4	-5.2
	227	72.7	189.9	210.4	327.8	72.5	-4.4	16.5	-21.0	0.8	-5.2
	258	95.6	187.9	235.3	338.8	84.4	18.5	14.5	4.0	11.8	6.7
	259	97.1	187.9	236.7	342.4	82.3	19.9	14.5	5.4	15.4	4.6
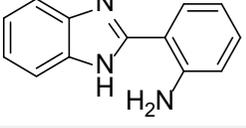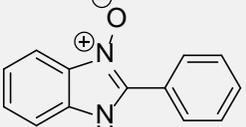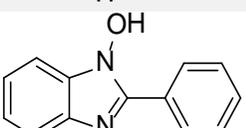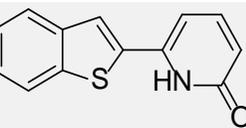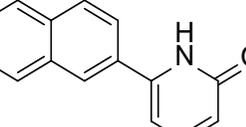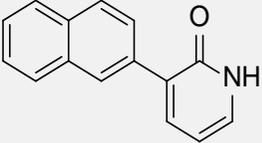
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	260	95.1	180.6	242.1	340.5	82.2	18.0	7.2	10.8	13.5	4.5
	213	93.4	179.3	241.7	336.5	84.5	16.3	5.9	10.4	9.5	6.8
	257	73.2	176.5	224.3	319.6	81.2	-4.0	3.1	-7.0	-7.4	3.5
	256	81.7	182.6	226.8	328.2	81.2	4.6	9.2	-4.6	1.1	3.5
	212	73.0	169.8	230.8	338.5	62.1	-4.1	-3.6	-0.5	11.5	-15.6
	211	102.1	184.9	244.8	345.2	84.5	24.9	11.5	13.5	18.2	6.8
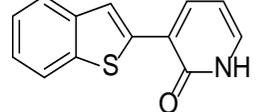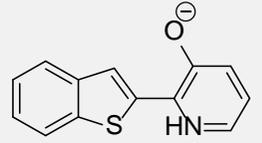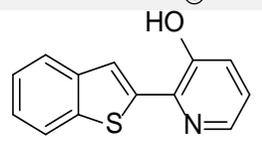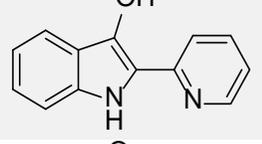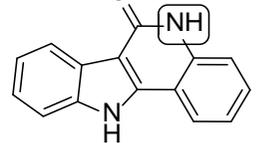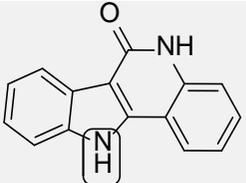
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	211	95.5	---	---	333.2	89.9	18.4	---	---	6.2	12.2
	255	82.1	173.1	236.6	335.5	74.2	5.0	-0.3	5.3	8.5	-3.5
	210	82.8	177.5	232.9	348.9	61.4	5.6	4.1	1.5	21.9	-16.2
	209	87.9	177.7	237.8	339.4	76.1	10.7	4.3	6.5	12.4	-1.6
	198	92.3	188.9	231.0	331.3	88.6	15.2	15.5	-0.3	4.3	10.9
	149	90.2	177.4	240.4	345.4	72.4	13.1	4.0	9.1	18.3	-5.2
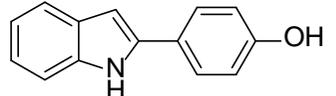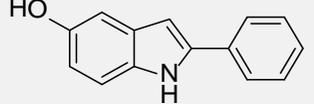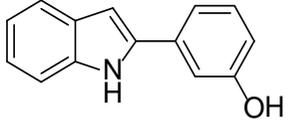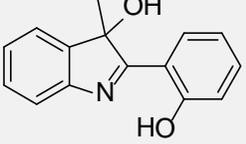
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	217	90.0	193.2	224.4	340.4	77.2	12.9	19.8	-6.9	13.4	-0.5
	214	87.1	187.3	227.5	330.6	84.2	10.0	13.9	-3.8	3.6	6.5
	261	83.2	181.5	229.3	331.9	78.9	6.1	8.1	-2.0	4.8	1.2
	Catechol	77.4	200.6	204.4	339.5	65.4	0.3	27.2	-26.9	12.5	-12.2
	236	86.3	180.7	233.2	338.7	75.2	9.1	7.3	1.8	11.6	-2.5
	186	77.8	172.0	233.4	329.4	76.1	0.7	-1.4	2.1	2.3	-1.6
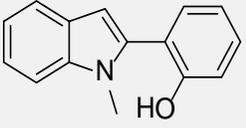
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	240	76.8	170.6	233.8	327.5	77.0	-0.3	-2.8	2.5	0.5	-0.7
	238	77.0	171.5	233.1	327.8	76.8	-0.1	-1.9	1.8	0.8	-0.9
	241	76.5	169.4	234.7	328.0	76.1	-0.7	-4.0	3.4	1.0	-1.6
	239	76.9	171.7	232.8	327.4	77.1	-0.2	-1.7	1.5	0.4	-0.6
	243	75.9	172.1	231.5	327.0	76.5	-1.2	-1.3	0.1	0.0	-1.2
	242	76.3	170.2	233.7	327.8	76.1	-0.8	-3.2	2.4	0.8	-1.6
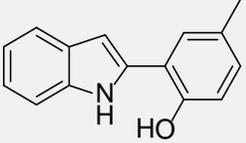
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	244	75.6	171.4	231.8	328.1	75.1	-1.5	-2.0	0.5	1.1	-2.6
	237	79.4	175.9	231.1	329.4	77.6	2.3	2.5	-0.2	2.4	-0.1
	245	84.5	187.3	224.8	335.1	77.1	7.4	13.9	-6.6	8.0	-0.6
	253	76.2	173.9	229.9	325.0	78.8	-0.9	0.5	-1.4	-2.0	1.1
	250	77.4	175.7	229.3	324.3	80.7	0.3	2.3	-2.0	-2.8	3.0
	249	77.6	177.4	227.8	324.1	81.1	0.4	4.0	-3.6	-2.9	3.4
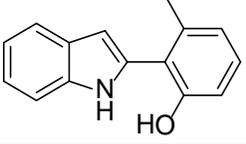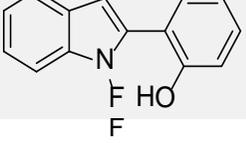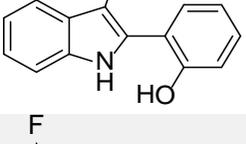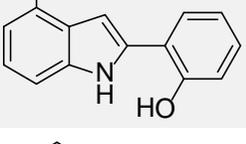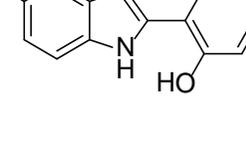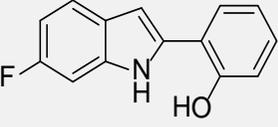
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	251	76.7	174.3	230.0	324.7	79.6	-0.4	0.9	-1.4	-2.3	1.9
	247	78.3	177.0	228.9	325.3	80.5	1.2	3.6	-2.4	-1.7	2.9
	246	78.7	176.8	229.5	325.2	81.0	1.6	3.4	-1.8	-1.8	3.3
	252	76.6	175.5	228.7	322.5	81.6	-0.5	2.1	-2.6	-4.5	3.9
	254	75.8	176.8	226.7	324.3	79.1	-1.3	3.4	-4.7	-2.7	1.4
	248	77.8	176.7	228.7	323.9	81.4	0.7	3.3	-2.6	-3.1	3.7
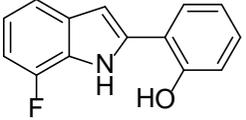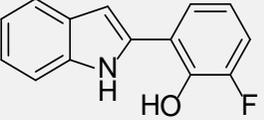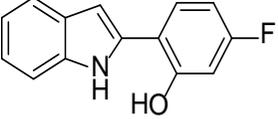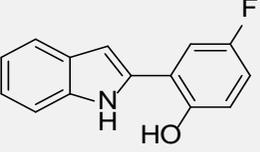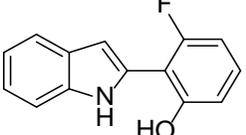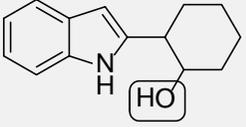
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	263	91.2	176.9	242.0	339.9	78.9	14.1	3.5	10.6	12.8	1.2
	263	90.6	176.6	241.7	339.6	78.6	13.5	3.2	10.4	12.6	0.9
	218	85.9	189.1	224.4	344.6	68.9	8.8	15.7	-6.9	17.5	-8.8
	Hydroquinone	82.1	193.1	216.6	350.1	59.6	4.9	19.7	-14.7	23.0	-18.1
	168	86.6	184.9	229.3	339.7	74.5	9.5	11.5	-2.0	12.7	-3.2
	152	84.8	172.8	239.6	344.4	67.9	7.6	-0.6	8.2	17.4	-9.8
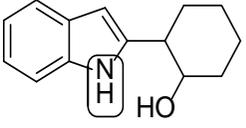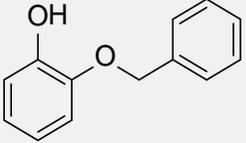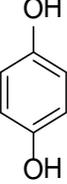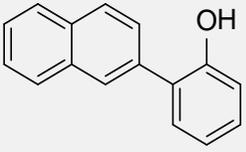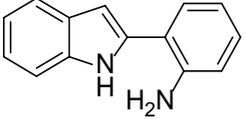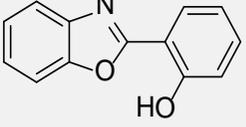
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	166	98.0	193.6	232.1	345.1	80.5	20.9	20.2	0.7	18.1	2.8
	167	92.5	191.3	228.8	336.2	83.9	15.3	17.9	-2.5	9.2	6.2
	Salicylic acid	94.0	210.0	211.7	347.0	74.6	16.9	36.6	-19.7	20.0	-3.1
	2-Nitrophenol	101.1	223.1	205.5	337.4	91.3	23.9	49.7	-25.8	10.3	13.6
	Salicylamide	91.6	202.2	217.0	335.0	84.2	14.5	28.8	-14.3	8.0	6.5
	2-Methylphenol	85.5	201.5	211.6	347.6	65.5	8.4	28.1	-19.7	20.6	-12.2
	Phenol	87.7	209.1	206.3	349.3	66.0	10.6	35.7	-25.0	22.3	-11.7
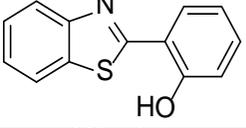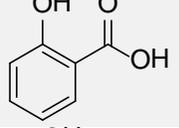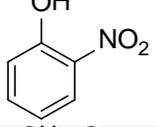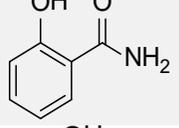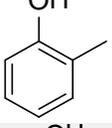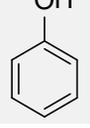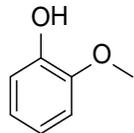
Structure	Compound	Absolute					Relative to Compound 4				
		BDE	AIP	PDE	PA	ETE	BDE	AIP	PDE	PA	ETE
	Guaiacol	86.8	193.5	220.8	353.3	61.1	9.6	20.1	-10.5	26.2	-16.6
	136	79.5	185.3	221.8	350.2	56.9	2.4	11.9	-9.5	23.1	-20.8
	169	87.0	193.9	220.7	342.4	72.3	9.9	20.5	-10.6	15.3	-5.4
	160	82.9	175.4	235.1	343.8	66.8	5.8	2.0	3.8	16.7	-10.9
	235	85.1	191.0	221.7	343.1	69.5	8.0	17.6	-9.6	16.1	-8.2
	Methanol	102.0	258.6	171.0	381.2	48.5	24.9	85.2	-60.3	54.1	-29.2
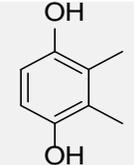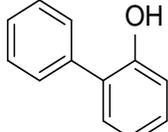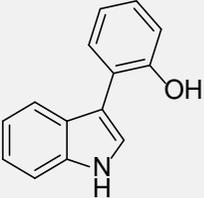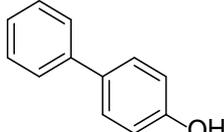
Appendix 5 – The PrestoBlue assay procedure

The U87 and SVG p12 cells (obtained from ECACC) were maintained in Minimum Essential Medium (MEM) supplemented with 10% (v/v) foetal bovine serum (FBS), 2mM L-glutamine, 1 mM sodium pyruvate and 1% (v/v) non-essential amino acids (NEAA) (Lonza) in a 37 $^{\circ}$C humidified incubator supplied with 5% CO_2. The cells were trypsinised before being seeded at a density of 10^3 cells/well (81 μL) into 96-well plates and incubated in growth media for 24 hours, before being treated with a solution of the compound (9 μL, preloaded in and containing ≤0.5% absolute DMSO) in triplicate and left for incubation for a further 24 hours. At this point, 10 μL of the pre-warmed PrestoBlue solution was added to each well, and incubated for 60 minutes. After 60 minutes, the absorbance was recorded at 612 nm using an XFluor4 GENios Pro plate reader. Half of the maximal effective concentration (EC_{50}) compared to untreated vehicle controls containing ≤0.5% absolute DMSO was calculated by nonlinear regression analysis.

Appendix 6 – SRB Assay Procedure

Short-term cell cultures (IN1472, IN1528, IN1760) were prepared from approximately 10mg of adult GBM biopsy tissue and maintained in Hams F10 nutrient mix [Invitrogen, Paisley UK] containing 10% foetal calf serum in a 37 °C non-CO_2 incubator as previously described.[310] Passages of 10 to 14 were employed for the current study. In addition, we also employed U251 and U87 which are established glioblastoma cell lines cultured under similar conditions.

Treated cells were assessed for their capacity to proliferate following treatment with compounds using sulphorhodamine B (SRB) assay.[311] Briefly, 3000 cells were seeded per well in a 96 well plate and allowed to reach exponential growth (48 hours). Compounds were dissolved in DMSO [Sigma Aldrich; UK] and cells were treated for 72 hours with serial dilutions of the drug.   The culture medium was removed and the cells fixed in 10% trichloroacetic acid [Sigma Aldrich; UK] on ice for 30 min followed by washing in water and air-drying. Cells were stained with 0.4% sulforhodamine B [Sigma Aldrich; UK] prepared in 1% acetic acid for 15-20 mins, washed in 1% acetic acid and air-dried. The dye was solubilized in 100 µL of 10mM Tris (not buffered) and read at 560nm [Dynatech MR5000] for quantification.

Analysis- Sigmoidal Dose Response (Variable Slope) (Non-Linear Fit)

Appendix 7 – Comparisons between EC_{50} values and radical-forming enthalpic values

This appendix contains the comparisons between EC_{50} values of compounds against the radical-forming enthalpic values BDE, AIP, PDE, PA and ETE. All enthalpic values are given in kcal mol^{-1}, and all EC_{50} values are given in μM. The data is arranged as one graph per cell culture.

Raw data

Batch 1	EC$_{50}$ values / μM				
	U87	U251	IN1472	IN1528	IN1760
4	233.1	337.7	332.4	497.1	465.6
152	354.6	537.5	527.1	690.9	4791.1

Batch 2	EC$_{50}$ values / μM				
	U87	U251	IN1472	IN1528	IN1760
Phenol	930	2420	1990	1000	7350
Salicylamide	850	2420	1430	1190	8660
2-Nitrophenol	950	2270	190	1500	6560
2-Methylphenol	740	1960	1530	1040	3800
Guaiacol	570	4230	1520	1170	1660
Aspirin	780	3240	1640	1320	2910

Batch 3	EC$_{50}$ values / μM		
	U251	IN1528	IN1760
4	300	760	100
Phenol	5110	5170	1170
2-Methylphenol	3110	1550	4500
Aspirin[7]	1110	5970	1200
Hydroquinone	220	67540	1310

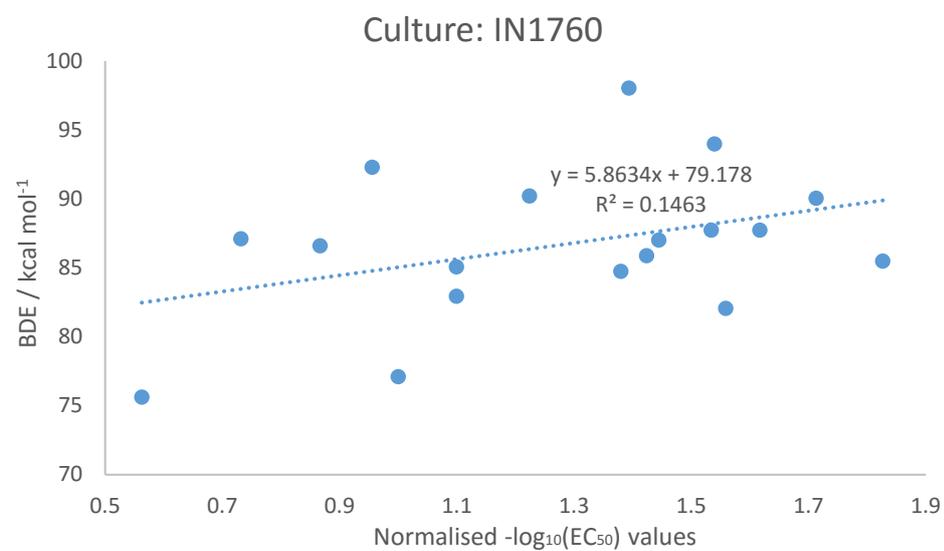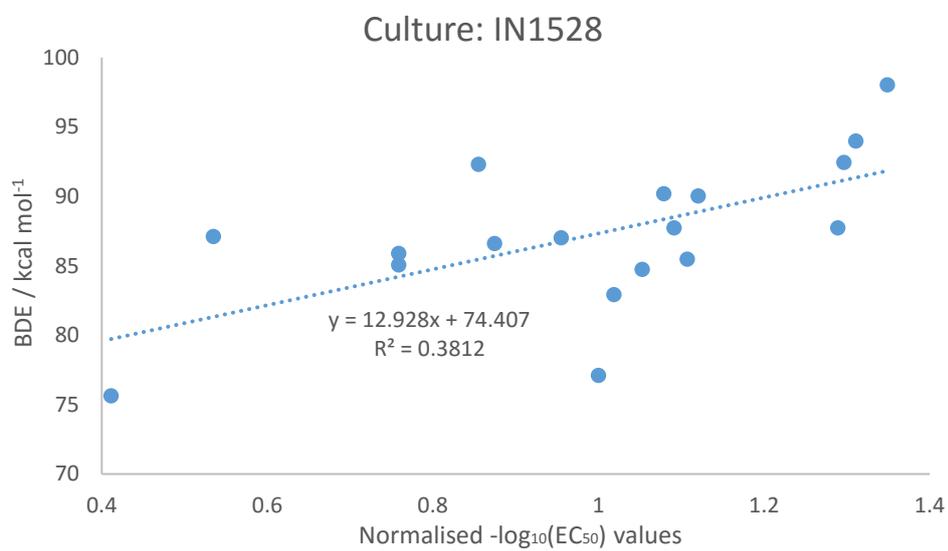| Batch 4	EC$_{50}$ values / μM		
	U251	IN1528	IN1760
4	310	1190	18430
166	3690	14080	879800[8]
167	1610	9710	-
168	240	490	4970

---

[7] Although EC$_{50}$ data is for aspirin, the enthalpic values used were calculated for deacetylated active agent (salicylic acid)

[8] Anomalous data – not included in analysis

Batch 5	EC$_{50}$ values / μM		
	U251	IN1528	IN1760
4	60	270	60
Phenol	80	450	750
198	110	120	50
244	30	10	10
169	350	210	370
217	680	530	1110
214	50	20	20
235	250	70	90
160	40	300	90
149	7	420	150
218	6	70	340

EC$_{50}$ values / μM

Comparisons between EC$_{50}$ values and BDE values

Culture: U251

$y = 6.0123x + 80.311$
$R^2 = 0.1262$

BDE / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values

Culture: IN1528

$y = 12.928x + 74.407$
$R^2 = 0.3812$

BDE / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values

Culture: IN1760

$y = 5.8634x + 79.178$
$R^2 = 0.1463$

BDE / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values

A39
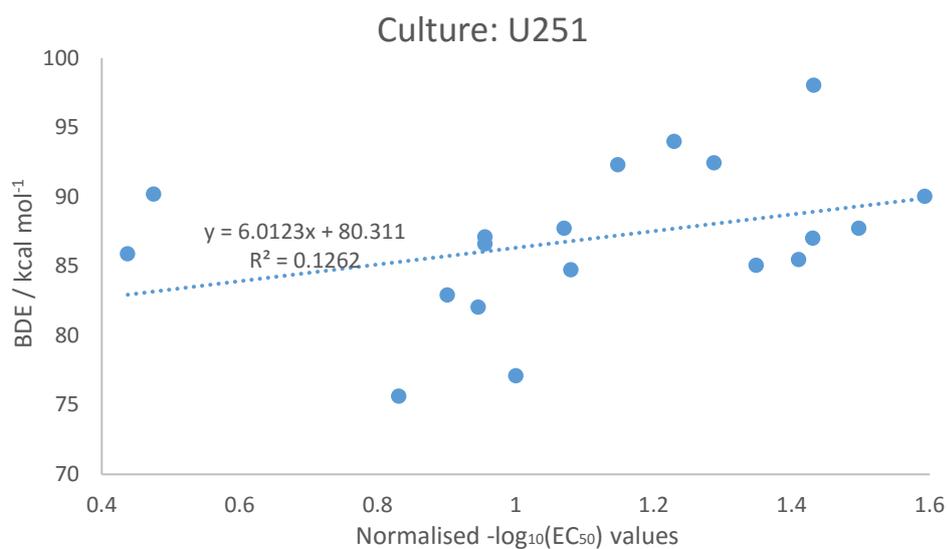
Comparisons between EC_{50} values and AIP values

Culture: U251

$y = 19.854x + 167.83$
$R^2 = 0.2773$

AIP / kcal mol^{-1} (y-axis)
Normalised -log$_{10}$(EC$_{50}$) values (x-axis)

Culture: IN1528

$y = 23.708x + 166.09$
$R^2 = 0.247$

AIP / kcal mol^{-1} (y-axis)
Normalised -log$_{10}$(EC$_{50}$) values (x-axis)

Culture: IN1760

$y = 23.254x + 160.05$
$R^2 = 0.4361$

AIP / kcal mol^{-1} (y-axis)
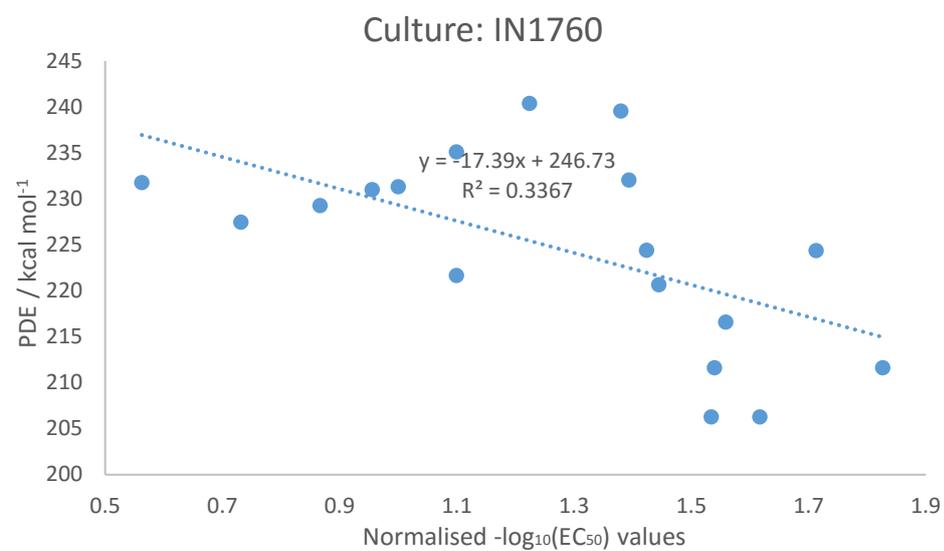Normalised -log$_{10}$(EC$_{50}$) values (x-axis)

A40
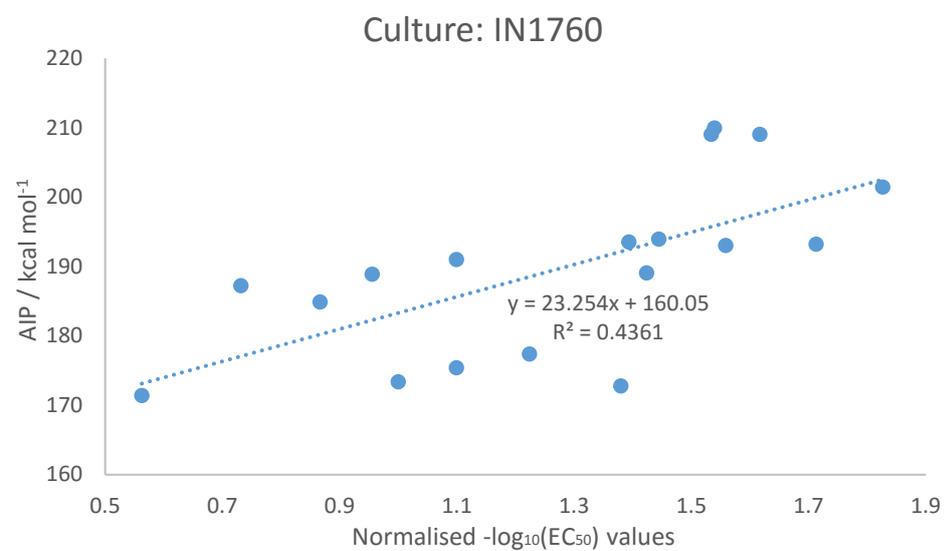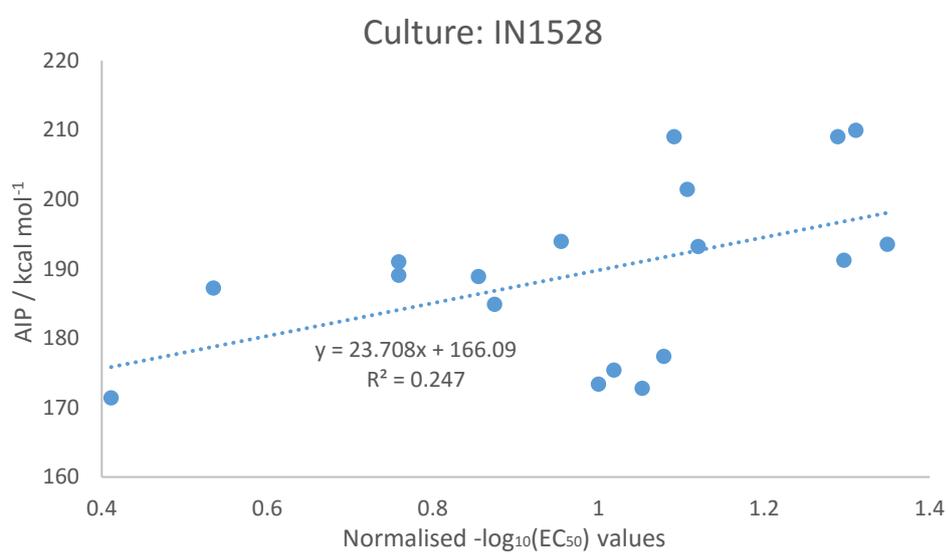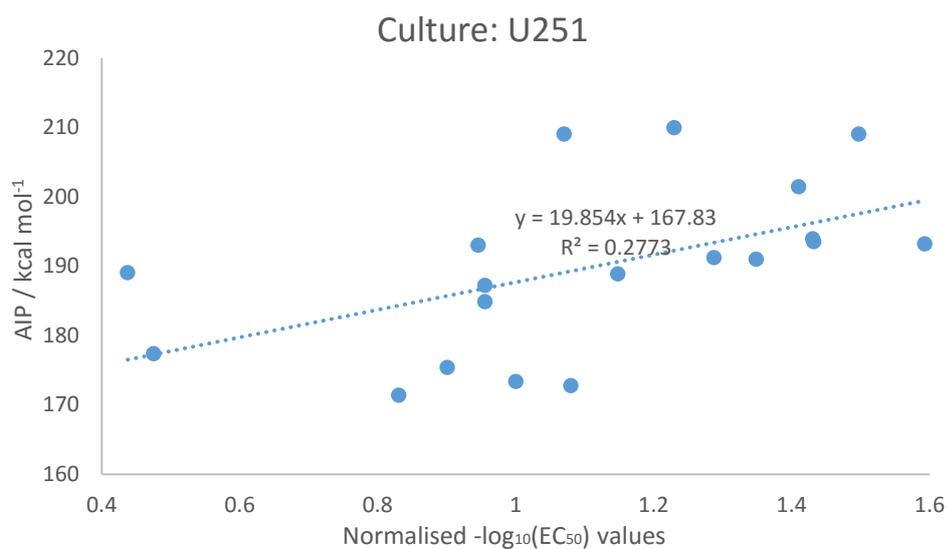
Comparisons between EC_{50} values and PDE values

Culture: U251

Equation on plot: $y = -13.842x + 240.08$, $R^2 = 0.1846$. X-axis: Normalised $-\log_{10}(EC_{50})$ values. Y-axis: PDE / kcal mol^{-1}.

Culture: IN1528

Equation on plot: $y = -10.78x + 235.92$, $R^2 = 0.0723$. X-axis: Normalised $-\log_{10}(EC_{50})$ values. Y-axis: PDE / kcal mol^{-1}.

Culture: IN1760

Equation on plot: $y = -17.39x + 246.73$, $R^2 = 0.3367$. X-axis: Normalised $-\log_{10}(EC_{50})$ values. Y-axis: PDE / kcal mol^{-1}.
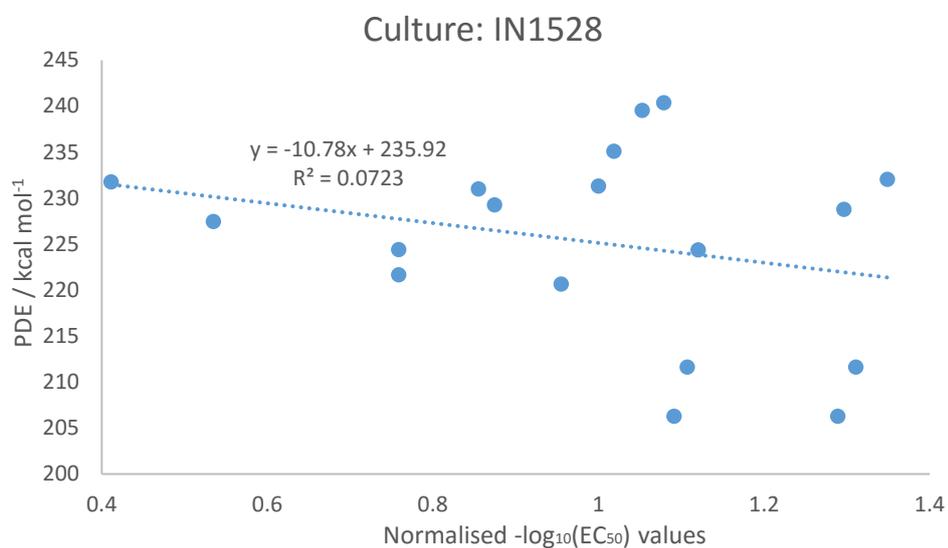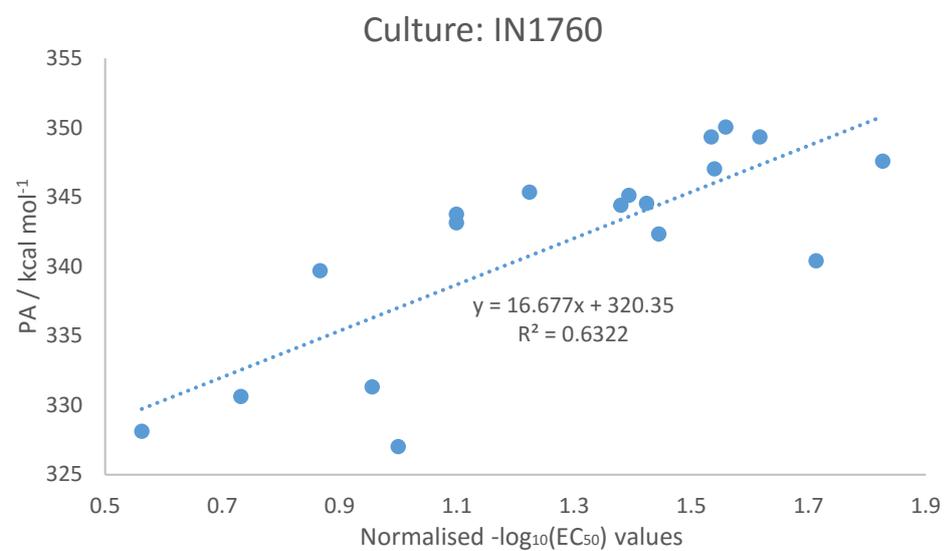
A41
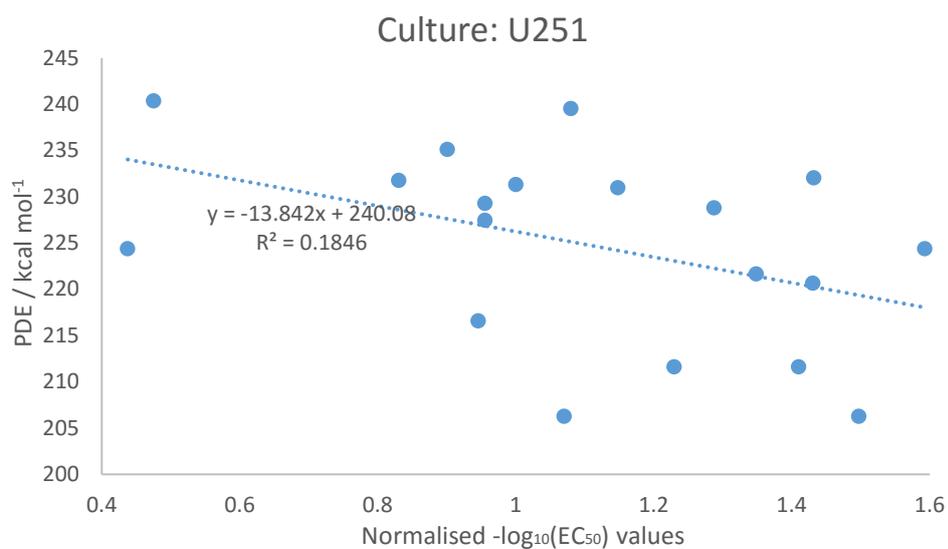
Comparisons between EC$_{50}$ values and PA values

Culture: U251

y = 3.2383x + 337.76
R² = 0.0202

PA / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values

Culture: IN1528

y = 16.33x + 324.66
R² = 0.3486

PA / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values

Culture: IN1760

y = 16.677x + 320.35
R² = 0.6322

PA / kcal mol^{-1}

Normalised -log$_{10}$(EC$_{50}$) values
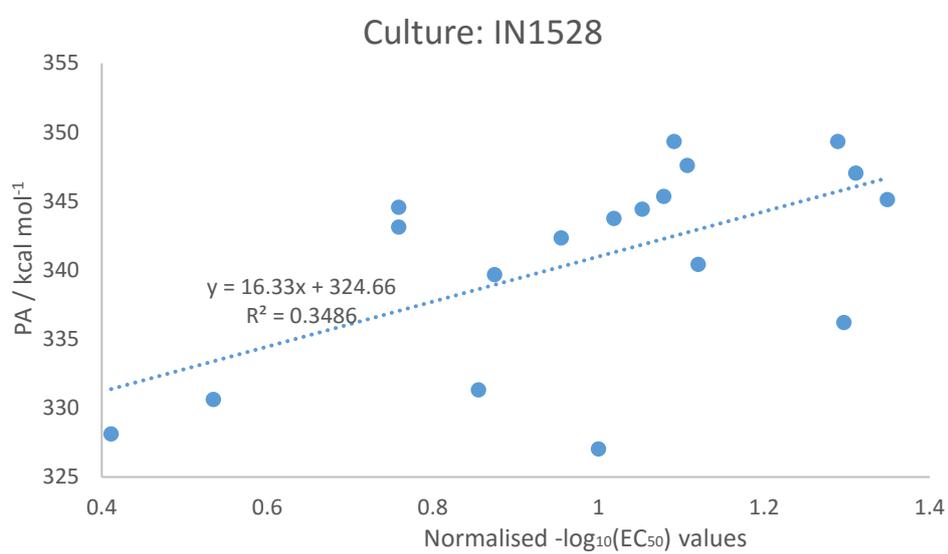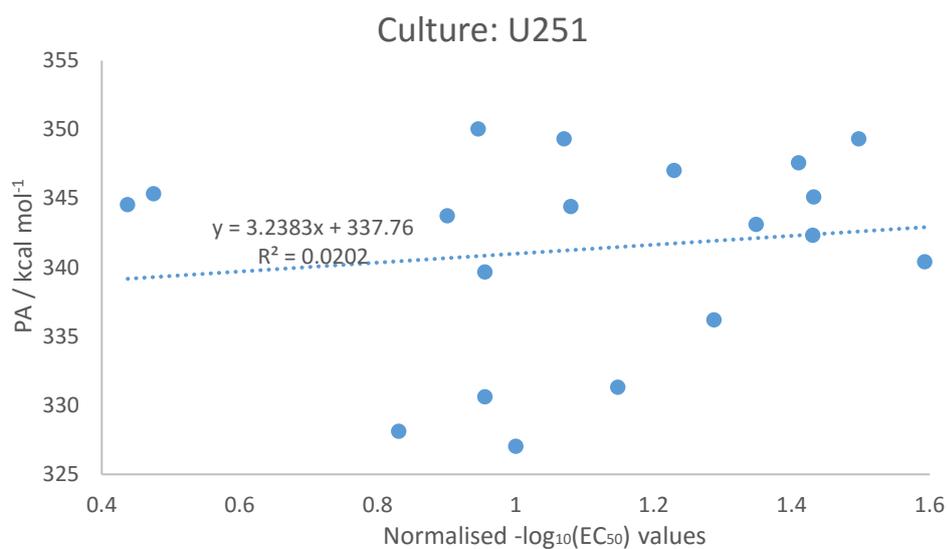
Comparisons between EC_{50} values and ETE values

Culture: U251

$$y = 2.774x + 70.151$$
$$R^2 = 0.014$$

Culture: IN1528

$$y = -3.4019x + 77.353$$
$$R^2 = 0.0162$$

Culture: IN1760

$$y = -10.814x + 86.427$$
$$R^2 = 0.2764$$
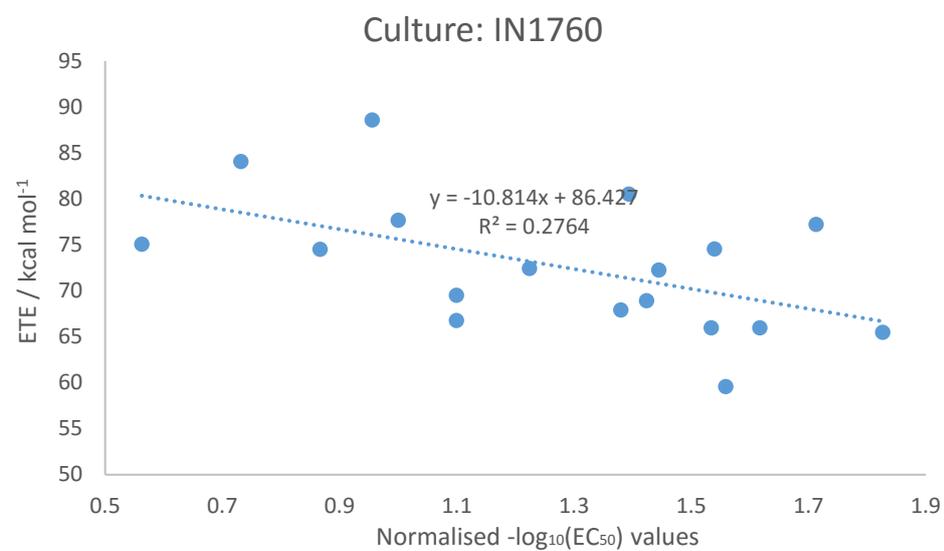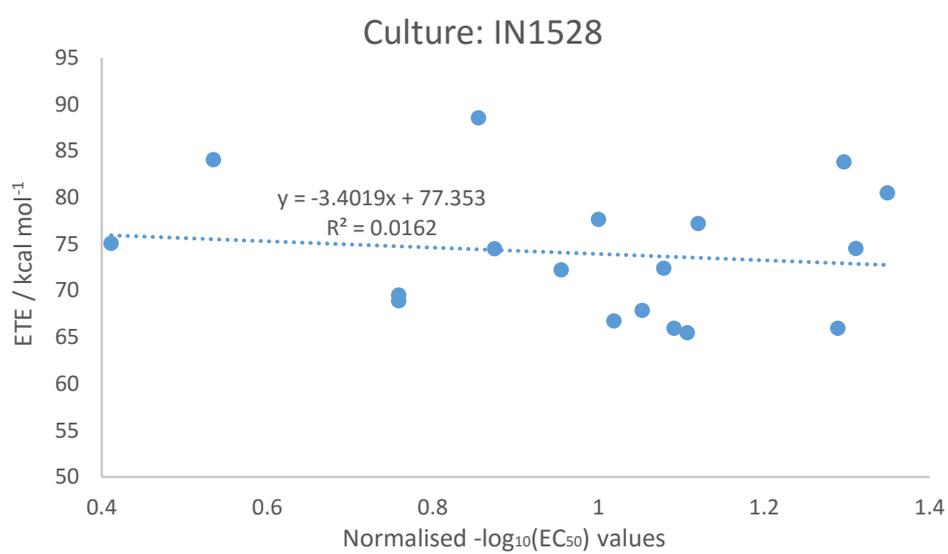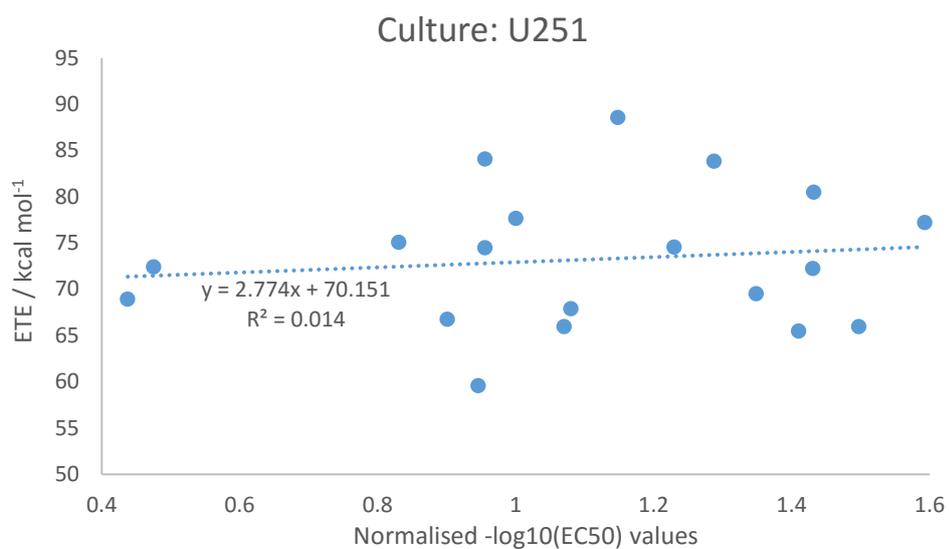
Comparisons between the activity of simple *ortho*-substituted phenols and their calculated enthalpic values

		U87	U251	IN1472	IN1528	IN1760
BDE	Trendline	y = -40.0x - 24.8	y = 6.10x +112.0	y = 12.7x +123.0	y = -81.8x - 160.6	y = -7.34x +64.3
	R^2	0.315	0.016	0.700	0.844	0.121
AIP	Trendline	y = -100.0x - 83.4	y = 37.0x + 333.1	y = 19.8x +266.9	y = -94.9x - 85.3	y = -19.3x +136.1
	R^2	0.659	0.198	0.564	0.379	0.281
PDE	Trendline	y = 60.1x +386.3	y = -30.8x +106.5	y = -7.04x +190.7	y = 13.0x +252.3	y = 12.0x +255.9
	R^2	0.677	0.394	0.204	0.020	0.308
PA	Trendline	y = 56.2x + 507.9	y = -30.6x +240.3	y = -10.5x +312.8	y = 57.1x +520.5	y = 19.2x +414.9
	R^2	0.412	0.269	0.317	0.271	0.548
ETE	Trendline	y = -96.2x - 205.1	y = 36.7x +199.5	y = 23.3x +144.8	y = -138.9x - 353.5	y = -26.5x - 22.9
	R^2	0.438	0.141	0.561	0.583	0.380

Appendix 8 – XRD data for decomposition of 2-(3-methyl-1*H*-indol-2-yl)phenol

It was observed that on standing over a period of a few days, compound **186**, a thick yellow oil, had crystallised into a pale yellow solid, and on doing so had changed by ^{1}H NMR. The resultant product was purified by recrystallization (ethyl acetate and petroleum ether) to form large amber crystals. The crystal structure of the product was obtained by the EPSRC UK National Crystallography Service at the University of Southampton.[231]
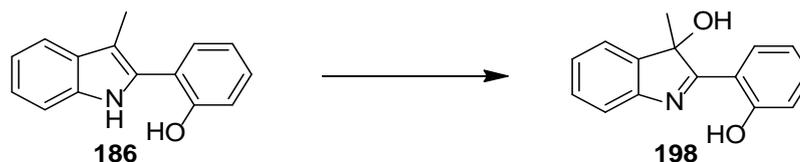
Scheme 65 - Formation of compound 198

Formula: $C_{15}H_{13}NO_2$; M_r = 239.26; crystal dimensions: 0.090 × 0.090 × 0.030 mm; crystal system; triclinic; space group: *P*-1; *a* = 7.4707(3) Å, *b* = 8.5536(3) Å, *c* = 9.2611(3) Å, α = 96.251(3)°, β = 97.624(3)°, γ = 98.134(3)°; Volume = 575,73(4) Å^3; *Z* = 2; ρ_{calcd} = 1.380 Mg/m^3; μ = 0.092 mm^{-1}; Mo *K*α radiation, λ = 0.71073 Å; T = 100(2) K; $2\Theta_{max}$ = 54.95°; 10399/2622 measured/independent reflections; R_{int}: 0.0216; *R* = 0.0392, *wR* = 0.0997; $\Delta\rho_{max}$ = 0.378 eÅ^{-3}, $\Delta\rho_{min}$ = −0.206 eÅ^{-3}. Orange block crystals gave good diffraction. The data were collected on a Rigaku AFC12 goniometer equipped with an enhanced sensitivity (HG) Saturn 724+ detector mounted at the window of an FR-E+ SuperBright molybdenum rotating anode generator with VHF Varimax optics (70 μm focus). Cell determination and data collection was carried out using CrystalClear-SM Expert 3.1 b27 (Rigaku, 2013). Data reduction, cell refinement and absorption correction was carried out using CrysAlisPro 1.171 (Oxford Diffraction, 2015). Structure solution was carried out using SUPERFLIP.[312] Structure refinement was carried out using SHELXL-2014.[313] Graphics were generated using enCIFer (International Union of Crystallography, 2014).
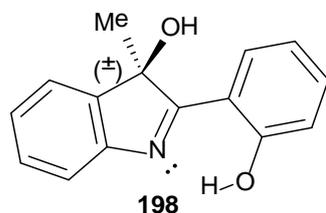

Figure 100 - Skeletal structure of compound 198 (left) alongside its crystal structure (right)