In Silico and *In Vitro* Investigation into the Next Generation of New Psychoactive Substances

Michelle Jennifer Botha

Submitted to the University of Hertfordshire in Partial Fulfilment of the requirements of the degree of Doctor of Philosophy

October 2018

In loving Memory of my Mother

Sharon Botha

Acknowledgements

I would like to express my gratitude to my PhD supervisory team: Dr Stewart B. Kirton, Dr Jacqueline L. Stair, Dr Suzanne Fergus and Dr David Y. S. Chau. A special thank you to Dr Kirton for being an incredible supervisor without your support none of this would have been possible and I will always be grateful for your endless supply of words of wisdom.

I would like to extend thanks to the University of Hertfordshire for funding this project and all the staff members within the Department of Clinical and Pharmaceutical Sciences who always provided me with support, with a special thank you to Emeritus Prof. Andrew Hutt for support and educating me on what is expected from a researcher and to Dr Ute Gerhard and Dr Val Mullane for the kindness they have shown me over the years.

To the wonderful Technical staff in the chemistry department namely Mrs Sarah Collyer, Mrs Judith Brooks, Mr Paul Turner and Dr Mark Scott your support both as colleagues and friends made research and teaching an absolute pleasure.

To the amazing research students at UH both past and present a massive thank you for all the love and laughter we shared. I would like to thank: Dr Rama Camara, Dr Satyajit Shetage, Dr Hassan Farah, Dr Jesus Castro-Calvo and Dr Amira Guirguis who has shown me nothing but support and friendship for the past four years. I would also like to thank Dr Laxmi Kerai, Ms Ilaria Passarini, Ms Mubinah Beebeejean, Ms Soumaya Hakim, Ms Rhamiya Mahendran and last but by no means least Ms Abigail Martin your friendship has been invaluable to me. A special thank you to Mr Peter Haddow and Ms Taleen Shakouri for always having the ability to make me smile.

I would like to dedicate this thesis to my parents Mark and Rose who have always supported me I would not be where I am today without you both. Finally I would like to extend my deepest gratitude to my closest friends: Dimitri, Faye and Marco for the decades of friendship and support I am truly grateful.

Abstract

New Psychoactive Substances (NPS) were designed to be legal alternatives to existing established recreational drugs. They have fast become a very popular and up until 2016, NPS were legal, cheap and freely accessible via the internet and high street "head shops". The rapid expansion in the number of these drugs has reached epidemic proportions, whereby hundreds of NPS have been developed and sold within the last five-year period. As NPS are synthesized in clandestine laboratories there is little to no control in the manufacture, dosage and packaging of these drugs.

The public health risks posed by these drugs are therefore far-reaching. Fatalities and severe adverse reactions associated with these compounds have become an ongoing challenge to healthcare services, primarily because these drugs have not previously been abused and therefore there is little pharmacological information available regarding NPS.

There are a number of different biological receptors that are implicated in the effects of NPS and the mechanism of action for the majority of these drugs is still largely unknown. It is of great importance to try and establish an understanding of how various classes of NPS interact on a molecular level. In this thesis, structure-based and ligand-based *in Silico* methodologies were employed to gain a better understanding of how NPS may interact with monoamine transporters (MAT). Key findings included both molecular docking studies and a number of robust and predictive QSAR models for the dopamine and serotonin transporters provided insight into how promiscuity of NPS between the different MAT isoforms could arise.

In addition, pharmacophore models were generated to identify chemical entities that were structurally dissimilar to known existing NPS that had the potential to interact with the cannabinoid 1 receptor (CB1) and hence were hypothesised could elicit similar biological responses to known potent synthetic cannabinoids.

Thirteen of these compounds were identified and carried forward for *in vitro* and *ex vivo* analyses, where preliminary results have shown that two compounds activate the CB1 receptor. Further optimisation of these compounds could yield a novel SC scaffold that was previously unseen. Additionally, the compounds identified and the methodology employed in the generation of these new chemical scaffolds could be used to guide Early Warning Systems (EWS) and facilitate law enforcement with respect to emergent NPS.

List of Figures

List of Tables

Abbreviations

3D	Three-dimensional
BBB	Blood Brain Barrier
CADD	Computer Aided Drug Design
cAMP	Cyclic adenosine monophosphate
CB1	Cannabinoid 1 receptor
CHO	Chinese Hamster Ovary cell
CNR1	Cannabinoid Receptor 1
DA	Dopamine
DAT	Dopamine Transporter
EFS	Electrical Field Stimulation
ELISA	Enzyme Linked ImmunoSorbent Assay
Hz	Hertz
IBMX	3-Isobutyl-1-methylxanthine
LDH	Lactate Dehydrogenase
MOE	Molecular Operating Environment
MTS	(3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)
NE	Norepinephrine
NET	Norepinephrine Transporter
Nm	Nanometers
NPS	Novel/New Psychoactive Substances
PLIF	Protein-Ligand Interaction Fingerprint
Pg	Picogram
QSAR	Quantitative Structure Activity Relationship
SC	Synthetic cannabinoid
SF	Scoring function
SER	Serotonin
SERT	Serotonin Transporter
T_c	Tanimoto Coefficient
VDW	van der Waals

Contents

Chapter 1

General Introduction

1.1 Traditional drugs of abuse

Traditional drugs of abuse include cannabis, heroin, cocaine and ecstasy[1], the consumption of these substances has been extensively documented throughout the last century.

Smoking heroin was known as "chasing the dragon" and was initially popularised in Shanghai in the 1920s[2]. This practice spread globally until the 1960s, when intravenous use of heroin became more common[2]. Currently heroin is the 2nd most popular narcotic in Europe[3] and the global amount of heroin seized reached a record high of 91 tonnes in 2016[4].

Cocaine use in the 1980s was mainly associated with the wealthier members of society[5], which ultimately gave rise to the cheaper smokable freebase alternative "crack"[5]. Both forms of this drug have become a global health burden[6]. Cannabis plants have also been cultivated by man for centuries and used for textile purposes (hemp) and the medicinal purposes of cannabis have been reported throughout history. The last century has seen the popularity of cannabis rise to a point where it has become the most widely abused drug globally[3,4].

Psychoactive/Psychotropic substances derived from plant based origins such as opium, khat, cannabis, coca leaves and magic mushrooms have also been abused for centuries[7] . In 1961, when opium smoking had become a serious public health threat, legislation was brought into effect to make the supply and use of narcotic substances illegal. Thus, the 1961 single convention on narcotic drugs[8] was introduced, which prohibited the production and supply of traditional drugs of abuse such as heroin, cocaine and opium. This began the process of criminalising drugs, drug-use and psychoactive substances in general. A number of subsequent drug laws[9,10] were brought in and enforced following the 1961 single convention on narcotic drugs, in order to guard against the risks to health and society at large that these substances posed. Despite the increase in legislation and its enforcement, suppliers of illicit narcotics continued to exploit the vulnerable and addicted, and in an effort to maximize their profits began to 'cut' their narcotics with other substances[11–15]. For example, in the late 90's there was a decrease in purity of traditional drugs such as cocaine[16] and it is believed that this was one factor that accelerated the growth of an emergent category of compounds called "Legal Highs" or New Psychoactive Substances, which users, often referred to in the literature as "psychonauts[17]", pursued as legal, and somewhat erroneously presumed safer replacements to the traditional drugs of abuse.

1.1.1 New Psychoactive Substances (NPS)

A New Psychoactive Substance (NPS) is defined by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) as 'a new narcotic or psychotropic drug, in pure form or in preparation, that is

not controlled by the United Nations drug conventions, but which may pose a public health threat comparable to that posed by substances listed in these conventions'[18].

Historically NPS were highly popularized in the mid-2000s and commonly referred to as "herbal highs", "designer drugs", "research chemicals" and "legal highs"[16] aptly named due to their legal status in the UK prior to the 2016 Psychoactive Substance Act[19]. "Legal highs" were easily purchased from high street "head shops" and online retailers at a much lower cost than more typical recreational drugs such as cocaine and MDMA[20].

1.1.2 NPS Classification

Categorisation of NPS varies greatly. Some literature, particularly that aimed at devising appropriate clinical intervention to treat NPS intoxication, will classify NPS according to their pharmacological effects using, for example, the following categories: stimulants, cannabinoids, hallucinogens and depressants[21].

The United Nations office on Drugs and Crime (UNODC) uses a semi-systematic classification system which divided NPS into nine categories based on their drug class i.e. aminoindane, tryptamines , synthetic cathinones, synthetic cannabinoids, plant-based substances, piperazines, phenethylamines, other substances and phencyclidine-type substances[22].

The European Monitoring Centre for Drugs and Drug Abuse (EMCDDA) expands on this and classifies NPS according to 13 types, some of which overlap with the UNODC classification system. These groups are: Aminoindanes, Arylalkylamines, Arylcyclohexylamines, Benzodiazepines, Cannabinoids, Cathinones, Indolalkylamines (i.e. tryptamines), Opioids, Others, Phenethylamines, Piperazine derivatives, Piperidines & Pyrrolidines and Plants & Extracts[23,24]. Similar to the classification system employed by the UNODC, the groups used by the EMCDDA are not systematic groups and can relate to the psychoactive effect of a compound e.g. cannabinoids, the chemical structure of the compounds e.g. cathinones or the source from which the substances have been extracted (e.g. plant material). Differences in classification and nomenclature of NPS add to the complexity of identifying and ameliorating the risks associated with this already diverse group of compounds[22].

Over the last 8 years, 560 previously unseen NPS have been identified in Europe[25] (Figure 1.1), these include compounds from the EMCDDA classes of Piperazines, Benzodiazepines Arylamines, Tryptamines, Opioids, Phenethylamines, Others, Synthetic cathinones and Synthetic cannabinoids[26] (SC).
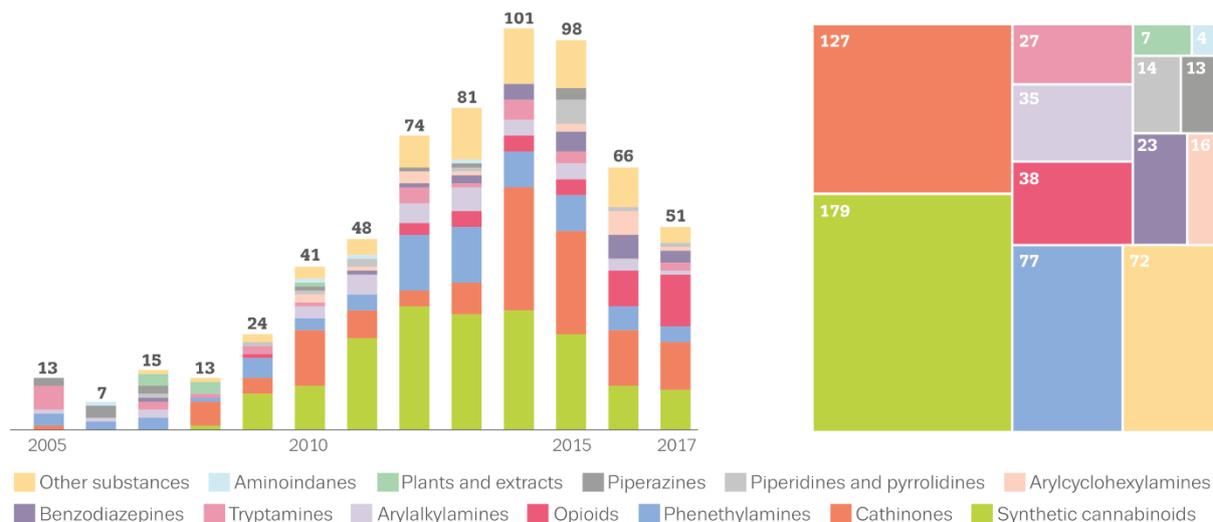
Figure 1.1: Graph showing the increasing number and type of NPS[26] being reported by the EMCDDA over a 12 year period (left), and the total number per category (right).

1.2 NPS Pharmacology

A large number of NPS are believed to interact in the body with monoamine transporter (MAT) proteins. MATs structurally consist of 12 transmembrane helices and play an active role in the release and re-uptake inhibition of the monoamine neurotransmitters (MNt); dopamine (DA), norepinephrine (NE) and serotonin (SER). The biological roles DA, NE and SER play within the body are varied and include mood stabilisation, appetite and aggression regulation, sexual arousal, cognition, mental wellbeing and decision making[27]. Neurotransmitter modulation can be carried out by their counterpart transporter proteins i.e. for dopamine DAT, for norepinephrine NET and for serotonin the SERT transporter. As MNt all play extensive roles in cognitive and emotional processes, the desire to alter their levels within the brain to elicit some form of psychoactive response has paved the way for recreational drug abuse.

The diverse biological roles of these compounds have led to the development of synthetic analogues and mimetics i.e. NPS, which are abused in order to elicit the desired recreational effects. As an example, dopamine mediates the sensation of pleasure in the brain. Unsurprisingly, given their structural similarity to dopamine, drugs such as cocaine, amphetamine, methamphetamine and methylphenidate mimic the effect of dopamine, by either promoting its release or preventing its reuptake, this means that these drugs are routinely abused.

NE transmission is associated with the flight or fight response and arousal. Clinical studies have demonstrated that low levels of NE in the brain are linked to depression[28,29]. As higher levels are

associated with heightened levels of arousal and/or feelings of wellbeing, chemically inducing an increase of NE in the brain to simulate arousal and/or euphoria have been achieved via the use of narcotics such as cocaine and amphetamines. As these compounds are illegal and subject to controls, NPS were initially developed to achieve the same results whilst circumventing the legislation. For example, cathinone derivatives such as 3, 4-methylenedioxypyrovalerone (MDPV) or "bath salts" have previously been marketed as legal alternatives to 3, 4-methylenedioxymethamphetamine (MDMA), amphetamine and cocaine[30].

1.2.1 Monoamine Transporters

All MATs consist of 12 transmembrane helices connected via both intra and extracellular loops located in monoaminergic neurons (Figure 1.2). These proteins are responsible for re-uptake of monoamine neurotransmitters using sodium and chloride ion gradient systems to transport the monoamines[31]. All 3 MAT have been studied extensively with relation to depression and addiction. It has been well-established that the transporters are responsible for the reuptake of monoamines which is indicated by the process being inhibited by antidepressants drugs. However, the mechanism of action for this inhibition of re-uptake is not fully understood[32].
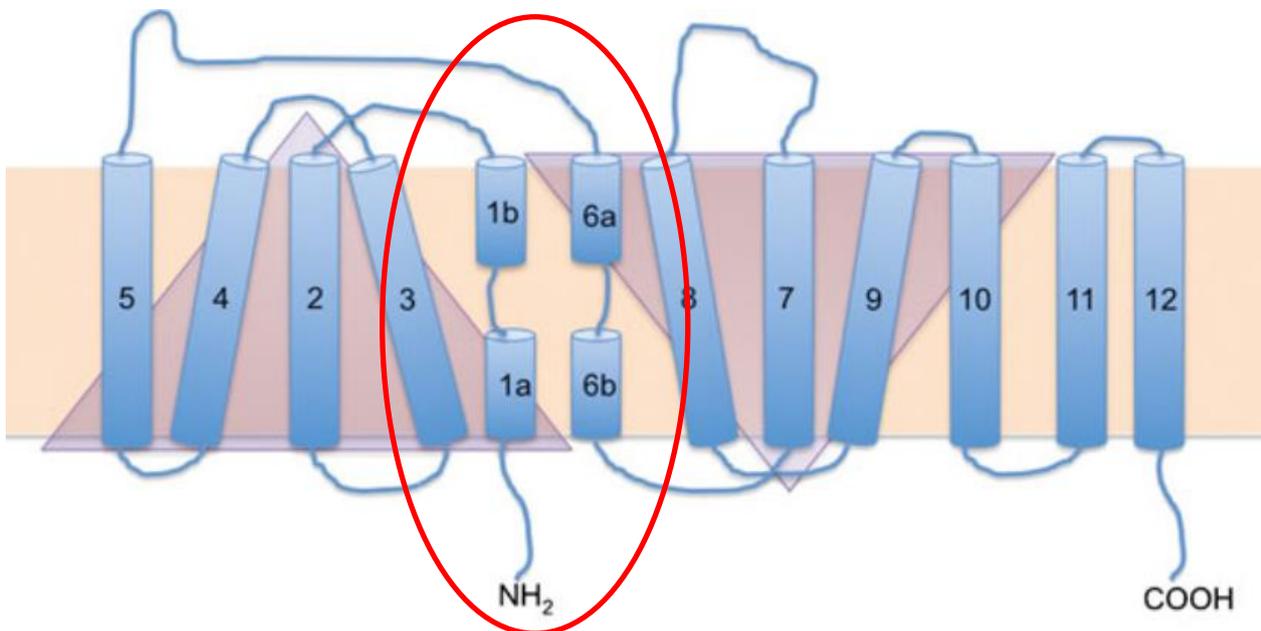
Figure 1. 2: Figure 1.2: 2D representation of the MAT[33] the 12 transmembrane helices are illustrated and the binding site circle in red.

There is a degree of promiscuity between the monoamines and MAT receptors as a consequence of their chemical structure similarities (Figure 1.3). This promiscuity of binding can also be seen with a variety of other molecules. For example, a number of antidepressants and appetite suppressant drugs (sertraline, fluoxetine, nisoxetine and Mazindol), have been found to have affinities with all three MAT [34],[35].
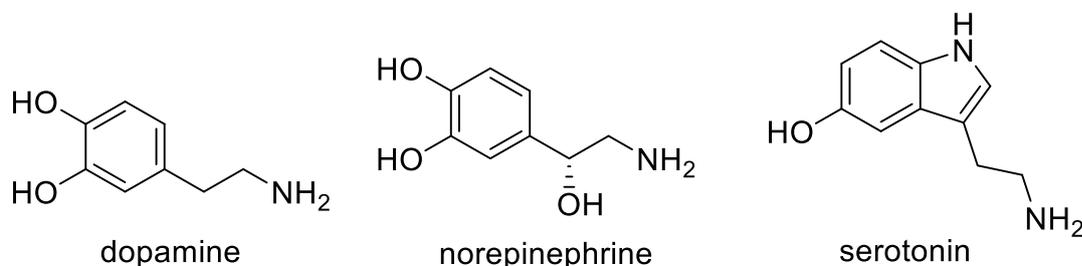
Figure 1. 3: Chemical structures of dopamine, norepinephrine and serotonin

The rapid emergence of NPS caused a number of NPS related poisonings and fatalities, and as these compounds had not previously been abused, the pharmacological effects of many of them were largely unknown[36]. The mechanism of action for NPS is not yet fully understood. It is known that the there is a high level of promiscuity between some NPS and the MAT isoforms. Given that each MAT is responsible for a large number of cognitive processes, and that such processes are altered with the use of NPS, compounds with the potential to interact with all three isoforms pose a potent potential threat to public health.

1.2.2 Dopamine Active Transporter – DAT

Dopamine concentrations in the brain are modulated by DAT. Ligands that interact with the protein will either inhibit reuptake of dopamine (cocaine) which will result in feelings of euphoria, or stimulate the release of synaptic dopamine (amphetamine) which often produces increased levels of confidence and energy[37]. Uncontrolled levels of dopamine in the brain are reported to be responsible for many neurological disorders such as depression, bipolar, Parkinson's disease and attention deficit hyperactivity disorder (ADHD)[38]. The mechanism of action for ADHD is associated with low levels of dopamine, drugs such a methylphenidate are known to stimulate the release of dopamine into the extracellular space to help manage this condition[39].

1.2.3 Norepinephrine Active Transporter – NET

NET recycles NE as well as DA and SER from the synapse to presynaptic neurons. There are fewer NET-selective ligands that have been investigated in comparison to DAT and SERT[34]. It has been suggested

that NET requires a secondary extracellular site for norepinephrine transport[40], as this feature is not common among all three MATs and NET has a high binding affinity to DA it could indicate that there is a similar transport mechanism between DAT and NET[41].

1.2.4 Serotonin Active Transporter – SERT

SERT is responsible for maintaining normal concentrations of serotonin in the brain. When the reuptake mechanism doesn't function properly unregulated concentration of this monoamine can result in depression, anxiety, stress, appetite deregulation and impaired cognition[42]. Disorders caused by the imbalance of serotonin have given rise to a group of drugs known as selective serotonin reuptake inhibitors (SSRI). Antidepressant drugs such as fluoxetine and citalopram work by blocking the reuptake of serotonin back into the serotonergic neurons, as it is believed that increasing the level of serotonin at the synaptic cleft is responsible for antidepressant activity[43].

Understanding structural differences between the MAT isoforms will provide insight into the neuropharmacological effects of psychostimulants. Identification of crucial protein-ligand interactions of NPS to a specific MAT isoform may highlight structural features that are required for selectivity between DAT, NAT and SERT. This thesis will set out to establish what the structural differences between the MAT isoforms are, and how those differences can be exploited to understand the selectivity of DAT, NET and SERT for a number of NPS

1.3 Factors influencing the rise in popularity of NPS: Perceived legality of NPS

Many drug users have revealed that it was their impression that the term "legal highs" implied that these compounds were safe[44] and the idea of "consequence-free" drug use has been one of the driving forces behind the popularity of this group of compounds. Up until 2016 these substances were not considered controlled substances, despite being based on the chemical structures of known illicit compounds. The NPS differed from controlled, illegal compounds as a consequence of small modifications to the illicit compound's chemical structure, which resulted in the circumvention of the law and afforded the new molecule "legal" status.

An example of this was the compound mephedrone, which is structurally similar to cathinone a naturally occurring stimulant found in the khat plant (*Catha edulis*)[19]. Figure 1.4 shows the chemical similarity between cathinone and mephedrone. The minor modification to the chemical structure of mephedrone, in comparison to cathinone, circumvented the then existing legislation, making the provision and consumption of that compound legal as it was not prohibited by the 1971 Misuse of Drugs Act[9].

Cathinone   Mephedrone
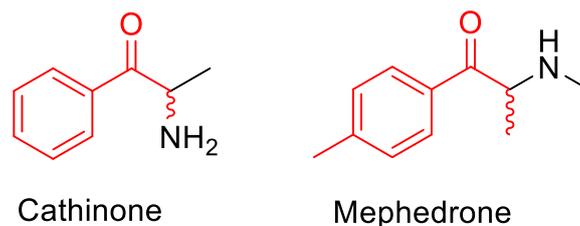
Figure 1. 4: The chemical structures of Cathinone and mephedrone, identical sections of the compounds are highlighted in red.

Mephedrone gained notoriety in 2008-9 and was a highly sought after party-drug as its physiological effects were likened to those of cocaine and amphetamines[45]. Figure 1.5 highlights the high degree of chemical similarity between mephedrone, Methylone and methylenedioxypyrovalerone (MDPV), and as such it is clear to see why these compounds would elicit similar physiological responses.

Mephedrone   Methylone   MDPV
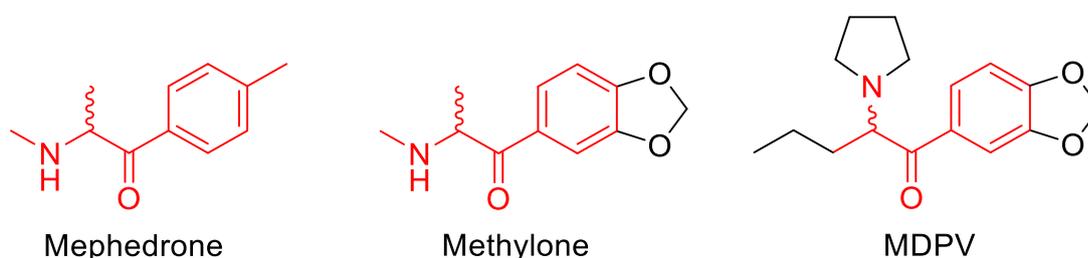
Figure 1. 5: The chemical structures of the most widely abused synthetic cathinones reported in 2014[46], mephedrone (4-methyl methcathinone), Methylone (3,4-methylenedioxymethcathinone) and MDPV (3,4-methylenedioxyprovalerone) The part of the molecules that are identical between the structures are highlighted in red

The popularity of mephedrone, and other NPS, was enhanced by their potency, easy access and legal status[45]. NPS have seen a rapid emergence onto the market[47] and the continued growth in numbers of these compounds has caused detrimental effects that have been felt across the globe[48]. For example, a number of NPS (from classes such as synthetic cannabinoids, cathinones[30] and phenethylamines) have been attributed as a causative agent in a number of reported fatalities from around the world [4,5,6]. The discovery of mephedrone gave rise to over 130 synthetic cathinones[46], However the authorities were quick to recognize the potential danger of these compounds, and mephedrone is now illegal in the UK. It became one of the first NPS to be banned under a modification to the 1971 Misuse of Drugs Act when on April 16 2010 it, and other substituted cathinones were made "Category B" substances[49].

1.3.1 Enhanced Potency and reduced costs of NPS compared to traditional drugs of abuse.

The demand for customer satisfaction, market competition and the battle to circumvent increasing amounts of legislation has fuelled a continual supply of new chemically diverse NPS[50]. Customer satisfaction, in this instance, is defined as ease of access to NPS, low cost and high potency. For example, the NPS N-Benzylpiperazin (BZP) is similar in terms of its chemical structure to amphetamine, but is marketed at lower costs (BZP approx. £2 per tablet in comparison to approx. £8-12 per gram of amphetamine according to the EMCDDA[51] and until recently circumvented the UK legislation on controlled psychoactive substances, despite being reported to have 10 times greater potency when compared to amphetamine[52]. Synthetic cannabinoids are 2-100[53] times more potent than traditional cannabis, and users of the popular SC cocktail "Spice" require a smaller quantity of the drug to achieve a more intense high[54] enhancing the appeal of NPS to drug users.

1.3.2 The role of the internet in access to NPS

The growing popularity of NPS was thought to be due to several factors including marketing strategies e.g. buy one get one free and colourful packaging[55]. The packaging associated with NPS has been cleverly designed, using names that are synonymous with known illicit drugs, such as "Gogaine" a previously legal alternative to cocaine which is known to contain ethylphenidate an analogue of the prescription drug methylphenidate (Ritalin)[56]. Online retailers of NPS would commonly employ a marketing strategy to name products after movies, social media sites and celebrities that were socially relevant at the time of distribution (e.g. Facebook, Charlie Sheen and Black Widow)[57].

The online distribution of NPS has facilitated the ease of access for both new and experienced drug abusers, which enhances the public health risks associated with these substances[51]. Online studies carried out stated that NPS users found the ease of acquisition, legality of sale[57] and variety of NPS via the internet as one of the greatest benefits of NPS use[58]. Online NPS forums and social media have provided a community whereby NPS users can candidly discuss their opinions and experiences of NPS[59], this coupled with home delivery via internet purchasing has facilitated the growth and interest in NPS[60].

The availability of NPS via online retail has reduced since the Psychoactive Substances Act[61], however through the use of cryptocurrency and the "Dark Net" NPS are still available[62]. The Dark Net /Deep Net/ Cryptomarket is an intentionally hidden marketplace for a diverse range of illegal activity including the sale of illicit drugs. The digital footprint of users of the Dark Net is untraceable and obfuscated as it is only accessible through specially configured browsers[63], which mask IP addresses and other digital identifiers. The technology used to hide users' activity also includes the anonymous

method of buying and selling items using a digital currency i.e. Bitcoin[64]. The use of this technology to almost eliminate the chance of detection means that despite the introduction of legislation to ban NPS, the popularity and use of these compounds continues almost unabated[65], with a reported drop in use for those aged 16-59 of only 0.3%. Although this is a reduction, there is evidence that some of those using NPS prior to the introduction of the ban having returned to using traditional drugs of abuse[66].

1.3.3 Public Health Concerns regarding NPS

The ease of availability, coupled with the diverse range of NPS being offered, has culminated in a global health concern. The number of fatalities and drug poisonings attributed to NPS abuse is on the rise[23,67,68] The latest trend in NPS relates to the abuse of the fentanyl family of compounds[69] (opioid analgesic) with nearly 6000 deaths being reported from 2012-2014[70] in the USA. Although the fentanyl epidemic appears to be localized to America, there have been reports of overdoses in the UK, where the reported number of fentanyl-related deaths stands at 58, and opioid overdoses have doubled (597 to 1209) in a four year period (2012-16)[71]. Concerns are growing around Europe that the prevalence of fentanyl is increasing, and there is an urgent need for effective harm reduction strategies to be put in place[72]. However, fatalities and drug overdoses are not limited to opiates, and have also been linked to the following NPS categories; Cathinones[73], Phenethylamine[67], Synthetic Cannabinoids[74] and Benzodiazepines[45].

Another key public health issue for NPS is the lack of information available about what type and amount of NPS are being abused, and the risks associated with the consumption of these new compounds. Many NPS have not previously been widely used and appear only briefly on the market[58,59,62,75]. Therefore, the clinical and pharmacological data associated with the risks for these compounds are not readily available for use by clinicians in the treatment of NPS overdoses[6].

1.3.4 Psychoactive Substances Act

In an effort to combat the burgeoning public health crisis posed by the advent of so-called "legal highs", the UK introduced a blanket ban of all NPS as of April 2016. The Psychoactive Substance Act 2016[19] prohibits the sale, possession and possession with intent to sell of psychoactive compounds, including the previously legal NPS, but makes provision for research institutions to be exempt from the act. The ability to effectively enforce this act has raised concerns and the ability of the blanket ban on these compounds to address public health concerns raised by NPS has been subject to criticism. Firstly, it has been argued that as a result of the elimination of online retailers, the NPS market may merge with already established traditional drug trading market and will focus the sale of these drugs

towards the most vulnerable groups and individuals[76]. Consequently, although this may reduce NPS availability it is argued that it will not reduce overall drug-related harm. Secondly, many academics view the act to be legally, and scientifically, flawed and any possible prosecutions will be costly. In addition, proving the psychoactive nature of seized contraband, given the number and chemical diversity of new compounds entering the market, may be problematic[77].

1.4 Synthetic Cannabinoids

As the diversity and availability of NPS has increased, the number of seizures associated with common drugs of abuse such as cocaine and heroin has become less frequent[25]. However, the use of cannabis does not seem to have been impacted to the same extent.

Cannabis is the most commonly abused recreational drug amongst young adults aged 12 and over, according to a national health survey conducted by the U.S Government[78]. Similarly, UK authorities have stated that 29% of people aged 16-59 have used the drug during their lifetime[79]. In 2013 over 670,000 seizures of cannabis/cannabis products were reported[25] with approximately 360 tonnes of cannabis (plants, resin and herbal cannabis) being seized in the UK in 2016[3]. As such, it is not surprising that the synthetic cannabinoids have amassed so much popularity[3], constitute the greatest number of new NPS entering the market over the last twelve years, and comprise a significant proportion of those NPS most recently seized (2.5 tonnes in 2015) . Synthetic Cannabinoid Receptor Agonists (SCRA) interact with the CB1 and CB2 cannabinoid receptors and elicit psychoactive effects designed to mimic the effect of THC, which is the primary psychoactive component of cannabis.

A report that has recently been published by the EMCDDA stated that there are currently 179 synthetic cannabinoids being monitored by the EU early warning system[80]. The increase in the use of SC is thought to be due to the ease of access, affordability, a more potent cannabis like psychoactive effect and the difficulty associated with detection of the compounds in urine samples[81]. It is likely that there will continue to be a substantial increase in this category of NPS, as recent articles have shown that both cannabis and synthetic cannabinoids continue to be the most popularly abused recreational drugs[3,82,83]. Again, the relatively mainstream nature of cannabis, lulls users of the synthetic cannabinoids into a false sense of security regarding the harmful nature of these compounds. Figure 1.6 highlights the increase in the number of SC poisonings in the US over a four-year period (2011-2015), and compares this to the trend in reported poisonings from other NPS and traditional drugs of abuse, reiterating that abuse of SC is a genuine global public health risk.
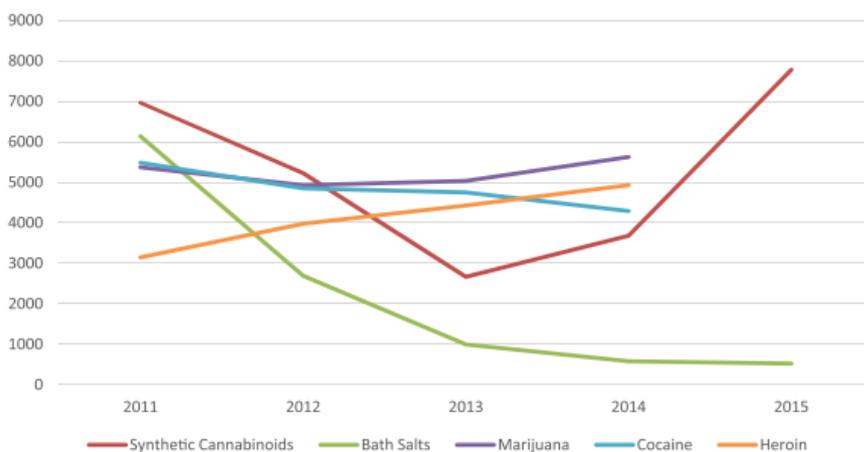
Figure 1. 6: line graph showing the number of reports to US poison control centres for SC (red), "bath salts" (cathinones) (green), marijuana(purple), cocaine (blue) and heroin (orange)[84].

1.4.1 Rise of Spice in UK prisons

In 2008 the EMCDDA first detected a new NPS called Spice[85], commonly sold as an all-natural smokable herbal mixture that was being advertised as a marijuana substitute[85]. Spice is also sold under brand names such as K2[86] and Black Mamba[53] often containing one ,or a mixture of synthetic cannabinoids, mimicking the effect of Δ^9-THC. The commonly detected constituents of Spice include but are not limited to JWH-018, JWH-073, HU-210 and CP-47,497[86], see figure 1.7 for chemical structures.
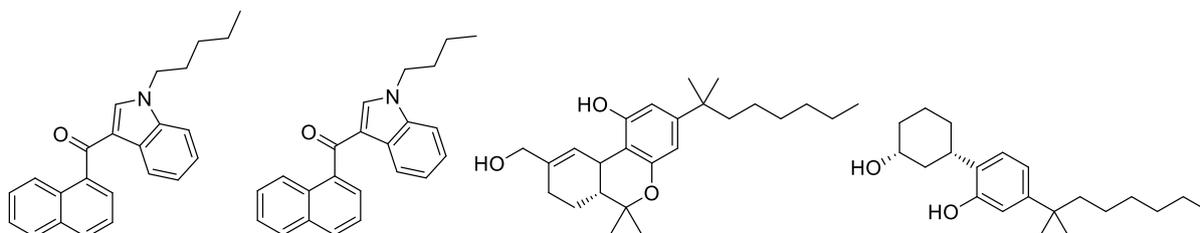
Figure 1. 7: Chemical structures of JWH-018, JWH-073, HU-210 and CP-47,497

There has been a rise in the use of Spice in UK prisons and a move away from using traditional cannabis products. It is commonly thought that this trend is due to a number of factors including SC being substituted for, e.g. marijuana, to facilitate the avoidance of drug detection, as SC have a much less detectable smell than traditional cannabis products, and do not show up in traditional drug screening such as urine tests. The compounds also tend to have a higher potency than plant-based cannabinoids[87]. Staff members within English prisons have claimed that the use of SC is endemic, causing significant problems for staff and prisoners[54] and that quantity of SC seized in English and Welsh prisons is now far greater than cannabis and heroin[88]. The far-reaching effect of SC is causing

20

chaos globally and although there have been attempts by government policies to eradicate NPS the use of SC is still prevalent.

Trends in the abuse of NPS are continuously changing, and it is therefore imperative that research communities investigate the compounds that could be exploited as the next generation of NPS by investigating the mechanisms by which molecules interact with the endocannabinoid system.

1.4.2 Endocannabinoid system (ES)

The endocannabinoid system is highly complex, consisting of endocannabinoids (endogenous cannabinoids), enzymes associated with the synthesis and degradation of endocannabinoids and two cannabinoid receptors (CB1 and CB2)[89]. The ES is one of the most important biological systems for mammals for homeostatic control of mood regulation, response to pain[90], energy expenditure, temperature regulation, regulation of the immune system and neural transmission[91]. The work presented in this thesis will focus solely on CB1 receptors, however it should be noted that in addition to CB1 and CB2 two other receptors are known to be implicated in the mediation of endocannabinoids these include the transient receptor potential (TRP) and the peroxisome proliferator activated receptors (PPAR's)[92]. The cannabinoids receptor (CB1 and CB2) are G-Protein coupled receptors (GPCR), which primarily couple to G proteins of the G_i and G_o classes[92]. Under specific conditions CB receptors have also been known to couple to both G_s and G_q G-proteins[93]. The activation of these receptors result in inhibition of adenylate cyclase activity causing a cascade of biochemical pathways being activated[94]. The CB2 receptor has an overall sequence identity of 44% (68% within the transmembrane reigions[90]) with CB1. Structural similarity between these two receptors means that cannabinoid-like molecules will likely bind to both CB1 and CB2 receptors, albeit with varying affinities, and as such many research groups[90,91,95] have searched for CB1 or CB2 selective ligands for different pharmacological reasons.

1.4.3 CB1 receptors

CB1 receptors are mainly located in the central nervous system and are involved in the psychoactive effect of Δ^9-tetrahydrocannabinol. They are the most abundant GPCRs in the brain[89]. The regions of the brain that are associated with cognitive function, movement, and sensory functions are known to contain densely populated regions of CB1 receptors[96], and hence molecules that bind to these receptors, such as the SC, can have marked effects on these functions. In 2016, the crystal structure for the CB1 receptor was elucidated[97] (see Figure 1.8). The publication of this structure, in the context of the work presented in Chapter 5 and 6 has provided reassuring results when compared to results

obtained from computational models of the receptor, carried out before the experimental structure was available.
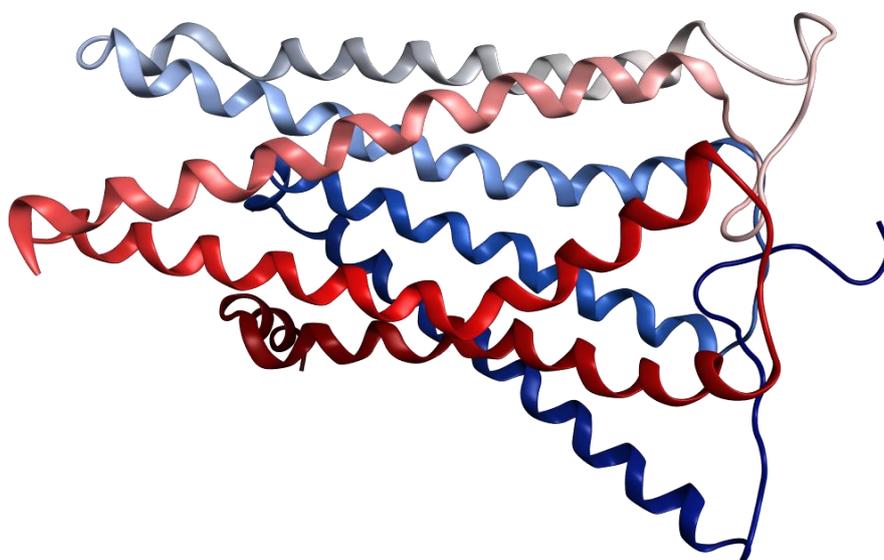
Figure 1. 8: Ribbon diagram depicting the secondary structural elements of the X-ray crystal structure of the human CB1 receptor (PDB accession 5TGZ), (image produced in MOE[98])

1.4.4 CB2 receptors

CB2 receptors are found in peripheral tissues and are implicated in the immune system[96], they are highly expressed in primary immune cells (leukocytes, macrophages, neutrophils and both B and T lymphocytes[94]). Due to the location and expression of CB2 receptors, drug discovery projects based on this receptor focus mainly on inflammation and pain treatment[90]. However due to the structural similarity between the two cannabinoid receptors, attempts in finding novel drugs that can treat a wide range of diseases without the psychoactive side effects have proved challenging[90].

1.4.5 Endocannabinoids

In 1964 the chemical structure of Δ^9-THC (Figure 1.9) was elucidated and identified as the main psychoactive constituent of marijuana[99]. This gave rise to the research into the numerous different compounds present in a cannabis plant and subsequently led to the discovery of endocannabinoids.
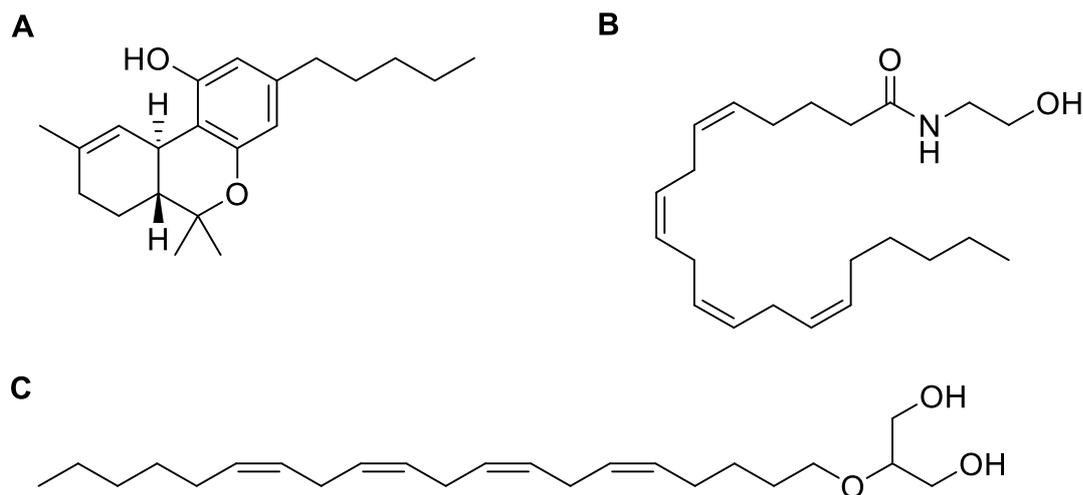
Figure 1. 9: Chemical Structure of Δ9-THC (A) and the endogenous cannabinoids anandamide (B) and 2-arachidonyl glycerol (C).

Endogenous cannabinoid ligands (endocannabinoids) are a family of intercellular signalling molecules[100]. In 1992, the first endogenous cannabinoid anandamide (AEA, Figure 1.8) was discovered. This compound was shown to have equivalent binding activity as Δ^9-THC at the CB1 receptor, which inspired research into a selective CB2 ligand[101]. The endogenous agonist 2-arachidonyl glycerol (2-AG, Figure 1.8) was sought out as CB2 specific. However, it was found that this particular molecule had similar binding affinities at both CB1 and CB2, again equivalent to THC[102]. Therefore, it is clear to see that scientists have previously carried out studies into identifying new chemical compounds with the potential to bind to the CB receptors, for therapeutic benefits. One tool that could be exploited by the research communities to facilitate their efforts is computer-aided drug-design.

1.5 Computer Aided Drug Design (CADD)

The last five decades has seen the implementation of CADD in the pharmaceutical industries going from strength to strength and it has played an important role in the discovery of novel, therapeutically active molecules. Two common approaches to CADD are structure-based drug design (SBDD) and ligand-based drug design (LBDD), Figure 1.10 summarises typical methodologies employed in CADD.
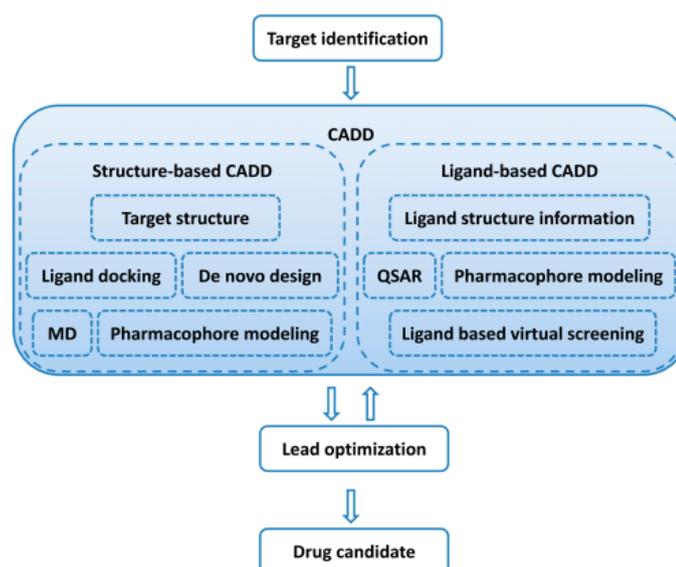
Figure 1. 10: A typical workflow employed in CADD pipeline adapted from Sliwoski et al[103].

A number of structure-based, and ligand-based CADD methodologies used in this thesis will be briefly introduced below.

1.5.1 Molecular Docking

Molecular docking uses computational representations of the protein and small molecule drug candidate structures to investigate how these small molecules may interact in the binding site of a target protein. The two key aspects of docking are the placement of a ligand within a binding site and a scoring or evaluation step in terms of the overall protein-ligand complex.

This process requires accurate information regarding the size, shape and composition of the binding sites of the receptors being investigated, and as such relies heavily on the availability of high-quality experimental structures of the receptors of interest. These structures are most commonly elucidated via means of X-ray crystallography or solution nuclear magnetic resonance (NMR) imaging. Molecular docking studies using high-quality crystal structures provide valuable information regarding protein ligand interactions and have been commonly employed in CADD.

Molecular docking uses an algorithm to form a protein ligand (P-L) complex, and the reliability of a docking algorithm lies in its ability to predict valid P-L complexes. Firstly, the search algorithm needs to comprehensively evaluate the potential energy landscape to establish the lowest energy conformation between both protein and ligand[104] and secondly the scoring function needs to accurately distinguish likely protein-ligand conformations from those that are unlikely

In this context, scoring functions are a set of mathematical models used to predict the binding affinity of a small molecule within an active site of a protein[105] and commonly estimate the energies associated with protein-ligand binding. An ideal scoring function would be able to predict absolute binding affinities and be able to appropriately rank different P-L conformations in order of binding free energy values[106]. However, this is not currently achievable with the state-of-the-art algorithms available. This is because scoring functions are calculated using approximations of the protein-ligand binding process and, as such, there are some associated limitations[106]. A fundamental limitation of a scoring function is the accuracy and availability of experimental data. Many scoring functions are calibrated on known and experimentally derived protein-ligand complexes and common P-L interactions. As it is not possible to account for all known P-L conformations in the development of a scoring function, there are inherent limitations of the functions ability to accurately predict the binding of chemical structures that have not been used in its development and testing. However, the accuracy of a scoring functions to rank molecules in order of their known binding affinities has shown to be improved if the study utilises a consensus model[66]. This method used a combination of scoring functions[107], in an effort to overcome inherent bias and/or limitations in a single model. This consensus scoring methodology has been employed in Chapter 3 of the thesis.

1.5.1.1 Experimental Receptor Structures

Numerous experimentally derived structures are available via publicly accessible repositories, such as the Protein Data Bank (PDB). At the time of writing , there are 143,840 structures available for download[108] and as previously mentioned the most common techniques used to elucidate high-resolution experimental structures are x-ray crystallography (89.5% of structures in the PDB) and solution NMR (8.5% of structures in the PDB), although no NMR-derived structures were investigated as part of these studies.

One way in which the quality of an x-ray crystal structure can be assessed is by looking at its resolution (usually measured in Å). Crystal resolution is a measure of the level of detail that can be accurately ascertained from the diffraction pattern obtained from the crystallographic experiment. The lower the number, the more detail in the structure can be resolved. A high resolution structure (of 1 Å or less) is detailed enough to show the presence of each atom in the electron density map[109], but such levels of precision are difficult to achieve, and it is possible to garner useful information about ligand-protein interactions from lower resolution structures. A common threshold that has been used for the development of docking algorithms and scoring functions is structures with resolution values of 2.5Å or lower.

1.5.1.2 Homology models

It is not trivial to obtain experimental receptor structures for all receptors of interest, especially those receptors that are membrane bound. Therefore, in the absence of available experimentally-derived structures, homology models of these proteins are often substituted for experimental structures in molecular docking studies.

A homology model is constructed by taking the known amino acid sequence of a protein and aligning the sequence to an experimentally derived 3D structure of a similar protein. This is called the template, and the homology model is referred to as the target. Homology models that have been refined and carefully validated can have a wide range of applications in drug discovery[110].

Online homology model libraries (e.g. SWISS-MODEL[111] and Protein Model Portal[112]) contain a number of freely accessible models that can be used in docking experiments, and these repositories were utilised during these studies. However, although homology models are useful tools in CADD they must be used with care as, with all models, they are interpreted and not derived directly from empirical evidence, and therefore may be subject to errors. For example, an error in the sequence alignment used to generate a homology model, can result in changes to the putative binding site for a modelled protein, and give misleading information when attempting to understand how small molecules may bind to that protein[113].

Given that protein structure is crucial for its function, and therefore the basic sequence alignment must be correct in order to produce a valid homology model. It is important to identify any shortcoming in homology models prior to using them in experiments. Assessing model quality should be conducted using a variety of independent and complementary techniques that examine the overall quality of the protein fold from a number of perspectives. Examples of protein validation methodologies that were used as part of this thesis are Ramachandran Plots[114], which assess the quality of the overall fold of a protein, Verify 3D[115] which looks at the quality of the immediate environment an amino acid residue resides in, and Errat[116] which examines non-bonded distances between pairs of atoms which gives an indication of the electronic quality of the protein (See Chapter 2 for details on how each of these methodologies works). All these protein validation programmes are available via free-to-use online servers which can result in a cost-effective way to ensure the quality of homology models prior to embarking on docking studies.

1.5.2 Quantitative structure-activity relationships (QSAR)

In the absence of either experimentally derived structures, or high-quality homology models, the research scientist can turn to ligand-based drug design (LBDD). For example, the use of Quantitative

Structure-Activity Relationship (QSAR) models to rationalize the differences between experimentally measured biological activities for a series of analogues as a function of their differing physicochemical properties was first conceptualized by Corwin Hansch in 1962 and since then has been widely used in industry and academia[117]. QSAR has been used in this thesis to try and understand what drives selectivity between the MAT isoforms, DAT, NET and SERT (Chapter 4).

QSAR is a statistical method of correlating the biological response (quantitative) of a series of analogue molecules to their physicochemical properties via a set of molecular descriptors[118]. Typically, QSAR studies start with a database of biologically active compounds with experimentally measure biological activities. These compounds are divided into a training set (typically 80% of compounds) which are used to build the model, and a test set (20% of total compounds) which are used to assess the generalizability of the model generated by the training set.

A series of molecular descriptors that capture the physicochemical properties of the compounds in the dataset such as their steric, electronic and lipophilic characteristics, are calculated, and then using techniques such as multiple linear regression (MLR)[119] the descriptors that are responsible for explaining the variation in biological activity are identified.

Hence, a simple equation is generated based-on the relationship between the descriptors identified as important. This equation is then used to predict the activity of the compounds in the test set. The QSAR model is assessed according to the correlation between predicted and experimentally derived activity values[120] for the test and training sets, where a correlation coefficient of 1 indicates a perfect fit between experimental and predicted results.

1.5.3 Pharmacophores

According to the IUPAC definition a pharmacophore can be defined as "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interaction with a specific biological target and to trigger (or block) its biological response"[121]. Pharmacophore modelling can either be structure-based whereby the target (active site of a protein) is analysed so that complementary chemical features (between ligand and complex) can be identified or ligand-based which identifies common chemical features from a set of 3D ligands that are deemed important for biological activity[121]. Pharmacophores have proven to be exceptionally useful for efficient virtual screening of large chemical databases. Chapter 5 focuses on the development of pharmacophore models which were used to search the ZINC database which contains over 17 million chemical structures[122]. The aim of this study was to identify a library of chemically dissimilar, wholly novel compounds that displayed the potential to interact with the CB1 receptor.

1.5.4 Virtual screening and High Throughput Screening

Virtual screening is widely used in drug discovery, whereby large, virtual libraries of compounds are systematically and quickly assessed against a predetermined criterion, such as the degree of fit to a pharmacophore as described above without the need to purchase or synthesize the compounds[123].The aim of virtual screening is in identifying novel compounds that are likely to bind to a specific drug target. Virtual screening is an alternative to High-throughput screening (HTS), especially in projects where resources are limited. HTS is a technique that experimentally assesses the biological activity of up-to millions of compounds by conducting automated individual biochemical assays[124]. This method is highly sophisticated and well established[125], and commonly employed after molecular docking ,pharmacophore and virtual screening development studies. High-throughput can often be expensive and time consuming[126] so the concept of "cherry picking"[127] the most ideal compounds is a sophisticated way of producing a library of the most suitable candidates possible that can be taken forward for biological evaluation.

The research presented in this thesis has used virtual screening in conjunction with pharmacophores to search a diverse library of commercially available compounds, with the additional step of filtering virtual hits resulting in a library of optimized compounds (Chapter 5).

Having identified those compounds which are predicted to bind to the receptor of choice, it is important to conduct experimental biological evaluations as the final measure of the predictive ability of these models.

1.6 Biological Evaluation

When a novel compound with a predicted potential to bind is identified the next stage in the drug discovery pipeline is the evaluation of biological activity, this is conducted by the use of biological assay. Commonly employed assays include functional assays which assess the inhibition of a substrate's uptake[128]and is an example of functional assay. Radioligand binding assays (example of a binding assay which use a radioactively labelled molecules which can be used to measure the rate of binding and binding affinity[129].

There is a lack of pharmacological data associated with NPS in general[130,131], and developing pharmacological profiles on NPS can be useful in generating vital information with the public health risks this class of drugs pose[36]. There are a wide range of receptors involved in forming interactions with NPS, these include, monoamine transporters, Cannabinoid 1 and 2 receptors, GABA and opioid receptors[128]. As more pharmacological data becomes available, the mechanisms of action associated with NPS are being better understood and the toxicology[132] and epidemiology of addiction[133] can be

treated more effectively. Functional assays were used here to investigate the binding of virtual screening hits to the CB1 receptor (Chapter 6). Commonly employed pharmacological tests that are routinely used to evaluate biological evaluation of cannabinoids include *in vitro* competitive binding affinity studies. This method uses radiolabelled cannabinoid molecules which are used to identify binding affinity (K_i) at the CB receptors[53]. Other methods include the use of behavioural studies whereby Tetrad tests are used to investigate the effect cannabinoids have on locomotive suppression, hypothermia, analgesia and catalepsy[81]. Organ bath studies are also used to identify the effect of cannabinoids on isolated segments on smooth muscle tissue[134].

1.6.1 Aims

There has been a massive expansion in the number and type of novel NPS that have been report in the past decade, although much research has gone into investigating already established NPS there is a gap in the market for methodologies in identifying novel NPS that have yet to be exploited.

A number of NPS are known to mimic the effect of known illicit drugs (cocaine and amphetamines) and the mechanism of action of these drugs has been well documented in the literature, therefore *in Silico* techniques can be applied to gain a better understanding of how NPS are acting within known receptors.

Therefore, the broad aims of the research laid out in this thesis is to:

- Establish if *in Silico* methodologies can be used to identify what physicochemical properties are required to convey selectivity for one receptor over another for NPS.

As stated by the literature there is a large degree of promiscuity of NPS with MAT, identifying structural differences between NPS may provide insight into why. The use of molecular modelling and homology models will provide insight into how reliable this methodology is terms of replicating the available biological data. The data obtained from docking studies will hopefully explain what gives rise to selectivity and differences in the affinity of small molecules between DAT, NET and SERT.

- Build predictive models that can be applied to a range of molecules that could potentially elicit psychoactivity.

The development of predictive models to determine a molecules ability to interact with one or more of the MAT is a methodology that can be applied to large virtual libraries. The identification of potential psychoactive chemicals may prove useful in determining which direction the NPS market could take. The number of NPS has increased rapidly over the last decade, the diversity of chemicals that have already been exploited as NPS is staggering. Identification of potentially psychoactive novel

scaffolds or novel chemical structures, as a pre-emptive measure could be used to advise early warning systems already in place. The information obtained from newly identified chemicals could be used to guide research into the mechanism of action of NPS or molecules that elicit psychoactivity.

- Identify at least one novel chemical scaffold that will interact with the CB1 receptor.

Synthetic cannabinoids are currently the most abundant NPS that have been reported, the aim of identifying novel SC-like molecules will provide information on chemical structures that have not been exploited as cannabinoids. The biological evaluation conducted on the novel virtual hits identified can provide information on the mechanism of action of SC. The molecules evaluated will be designed to be as structurally distinct from known SC, this will establish new chemical scaffolds that interact with the CB1 receptor. Any molecules found to demonstrate an affinity for the CB1 receptor can be further investigated to highlight pharmacological properties that are associated with SC. In addition, identifying previously unseen potential psychoactive chemical entities can be used to provide law enforcement and global organizations such as the EMCDDA an early warning system of potential drug structures that could be used to elicit psychoactivity.

Chapter 2

In Silico Methodologies

This chapter will focus on providing an overview of the different computational techniques employed throughout this thesis. A summary of the methodologies used is presented in Figure 2.1
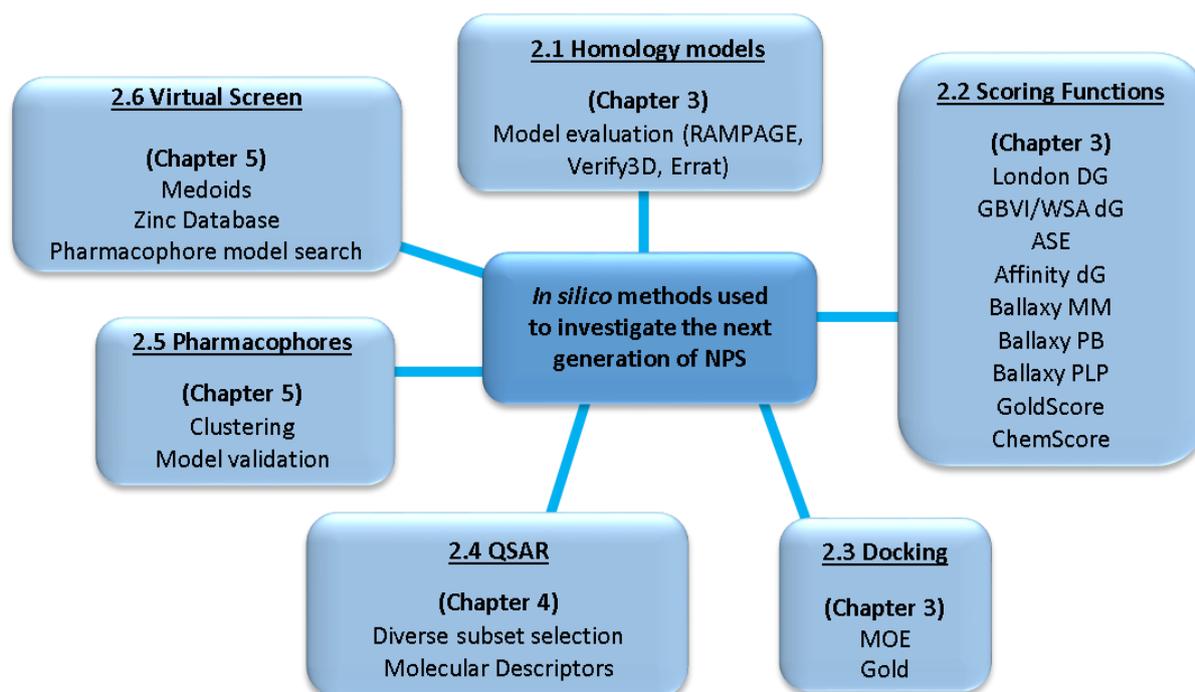
Figure 2. 1: An overview of the methodologies employed in the *in Silico* studies presented in this thesis.

2.1 Homology Models: Evaluation

The use of homology models is common practice in the absence of an available experimentally determined structure. Quality assessment of these models is a very important step in ensuring the reliability of model before they are used to test hypotheses. A number of complimentary, yet independent techniques were used in this thesis to assess the quality of homology models.

2.1.1 Ramachandran Plot (RAMPAGE, Chapter 3)

Evaluation of the stereochemical quality of the polypeptide backbone was carried out using a Ramachandran plot obtained from Rampage[135] (see Figure 2.2). Residues are placed into allowed and disallowed regions on the Ramachandran plot, based on the dihedral angles within the peptide bonds of the simulated structure. The torsion angles around the bonds between N-Cα denoted by φ (Phi) and the bond between Cα-C' denoted by ψ (Psi) are responsible for the flexibility required for the protein to adopt its characteristic secondary structure. The third torsion angle is denoted by ω (Omega) and accounts for the Cβ-N bond which is fixed to 180 degrees due to the partial double bond

character of the peptide bond. This plot allows for easy visual representation of the distribution of amino acid residues in a protein structure according to where you would expect to find the residue within a secondary structure (i.e. α helices or β-sheets). Atoms are treated as simple impenetrable spheres with dimensions corresponding to their van der Waals radii. Therefore, phi and psi angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone. Protein models that have 90% or more residues in the allowed region are largely accepted as being reliable models[136].
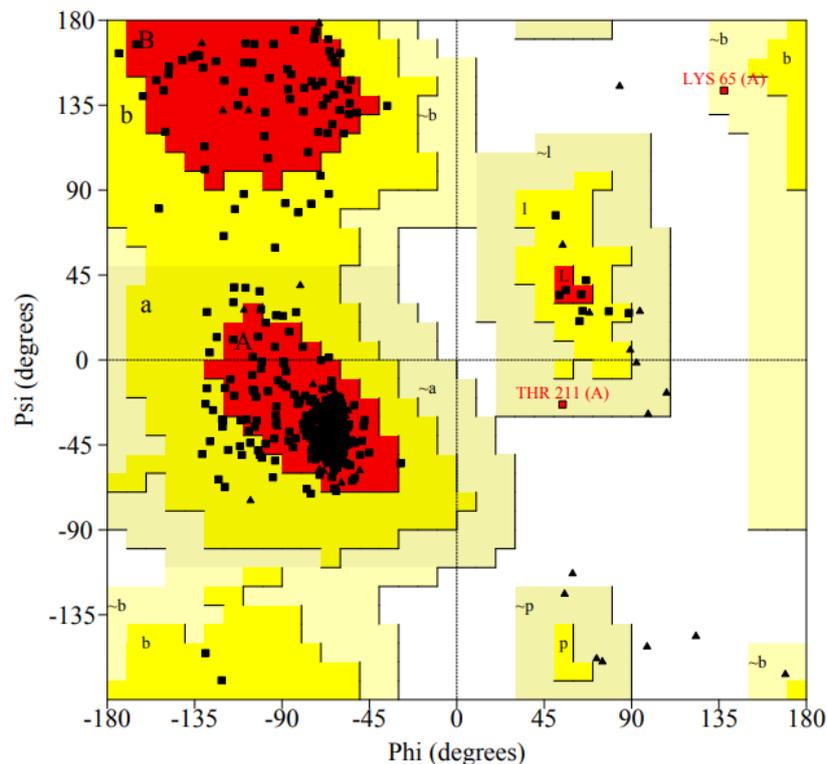
Figure 2. 2: A Ramachandran plot for the homology model for DAT (accession code Q01959[112]), Ramachandran analysis plots the torsional angles (Φ, x-axis andΨ, y-axis) of all the residues. The plot is split into areas according to secondary structure (regions labelled B/b for β-sheets, A/a for α-helices and L/l for left handed helices). The conformations of residues are categorised into four groups: most favoured (which are found in the red sections), allowed residues (plotted in the yellow sections of the graph) additionally allowed conformations (cream) and disallowed conformations found in the white sections of the graph.

2.1.2 Verify3D (Chapter 3)

Verify3D[137] establishes the compatibility of a protein's amino acid sequence with a known 3D structure by assigning the amino acid to a common protein structural classes i.e. α- helices, β-sheets, based on its environment and the surrounding residues. Verify3D works by reducing the 3D environment of

each individual residue to one dimension and assigning an average score derived from a 21-residue sliding window based on solvent accessibility, the polarizability of the residues and the preference for protein secondary structure. Amino acid residue type (polar, non-polar, hydrophobic etc.) are giving a score with respect to solvent accessibility e.g. polar residues will be assigned a large positive score if they are solvent exposed. However, if this same type of residue is found to be buried in a deep hydrophobic pocket the residue will be assigned a large negative score. It is the sum of the scores for each of the residues that describes the overall quality of the protein. Regions of poor model quality will have a score less than 0.2. Using data for the comparison of 1D and 3D structures Verify 3D is able to give information about the quality of the overall model at the residue level. Figure 2.3 is an example of a Verify3D plot for the NET homology model (accession code P23975) the open source software Structure Analysis and Verification Server (SAVES) metaserver (http://nihserver.mbi.ucla.edu/SAVS/) was used in order to produce Verify3D[138] plots.
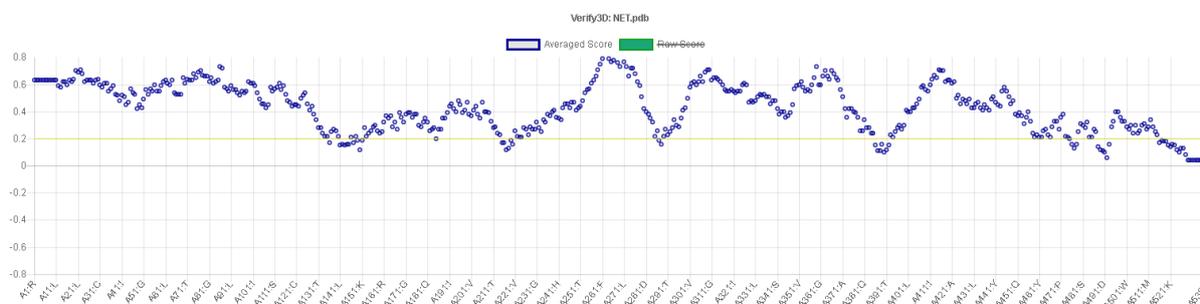
Figure 2. 3: Verify 3D plot for the Net homology model (accession code P23975[112]). The blue dots denote the average score for each amino acid residue, a value of 0.2 or greater indicates the residue is in a favourable amino acid environment.

2.1.3 Errat (Chapter 3)

ERRAT[139] is a program that uses error values in the distances between pairs of non-bonded atoms as an indication of model quality. It assesses the distribution of different types of atoms with respect to one another in the protein models, after having categorised them as either C, O or N and thus defining six possible pairwise interactions (CC, CO, CN, OO, ON, NN). Statistical analysis of the non-bonded atom-atom interactions is plotted on a graph that highlights residues that can be rejected (i.e. those that have not adopted an expected conformation) at the 95% and the 99% confidence levels. ERRAT is the most sensitive method of protein model validation in comparison to the Ramachandran plot, which is a more coarse grained approach examining overall fold and Verify3D[138], which focuses on assessing the quality of amino acid environment. In combination all three provide a powerful way of

identifying likely areas of protein misfolding in homology models. Figure 2.4 is the Errat plot output for the SERT homology model (accession code P31652[112]).
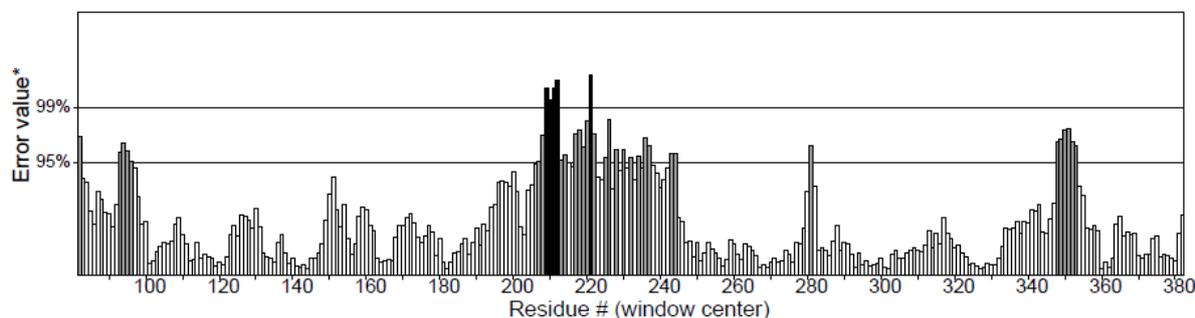
Figure 2. 4: Errat plot obtained for the SERT homology model (accession code P31652[112]). ERRAT plots assess the distribution of different atom types. Values are plotted as a function of the position of a sliding 9-residue window. The 95% and 99% error lines on the graph show at what confidence level the residues can be rejected for not having the correct distribution between atom types.

2.2 Scoring Functions (Chapter 3)

Scoring functions are used to evaluate protein-ligand interactions and are typically classified into three categories; force-field-based, empirical and knowledge-based[140].

Force-field based scoring functions are based on physical atomic interactions[141] and attempt to calculate the atomic interaction energies of a protein-ligand interaction. A force-field scoring function is calculated using both experimental data and *ab initio* quantum mechanical calculations[106].

Empirical scoring functions use weighted energy terms to estimate the binding affinity of a protein ligand complex, the energy terms considered for this scoring function are VDW energy, electrostatics, hydrogen bonds, desolvation, entropy and hydrophobicity[106].

Conversely, knowledge based scorning functions are derived from crystal structures and are calculated by taking the sum of pairwise statistical potentials between protein and ligand[140].

The utilisation and success of each category of scoring function in docking studies and structure based drug design is well established[105]. What follows is a closer examination of the scoring functions that were used as part of the studies that comprise this thesis.

2.2.1 London dG (Chapter 3 and Chapter 5)

The London dG[98] is an empirical scoring function, which estimates the free energy of binding of the ligand from a given pose. The functional form of the scoring function is a sum of terms as described in Equation 2.1 and Table 2.1:

$$\Delta G = c + E_{flex} + \sum_{h\text{-bonds}} c\text{HB}f\text{HB} + \sum_{m\text{-lig}} c\text{M}f\text{M} + \sum_{atoms\ i} \Delta Di$$

Equation 2.1: One of the two equations used in the calculation of London ΔG scoring[98].

Table 2. 1: The composite terms used to calculate the overall London dG scoring function values.

Subscript	Description
c	the average change of rotational and translational entropy
E_{flex}	energy loss of ligand flexibility
^fHB	measure of geometric imperfections of hydrogen bonds
^cHB	energy of an ideal hydrogen bond
^fM	measure of geometric imperfection of metal ligations
^cM	energy of an ideal metal ligation
D_I	the desolvation energy of an atom

The difference in desolvation energies, ΔD_i (Equation 2.2) is calculated according to the following formula, the terms of which are explained in Table 2.2.

$$\Delta DI = c_i R3_i \left\{ \iiint_{u \notin A \cup B} |u|^{-6} du - \iiint_{u \notin B} |u|^{-6} du \right\}$$

Equation 2.2: the equation used to calculate the difference in desolvation energies for the London dG scoring function[98].

Table 2. 2: The terms used to calculate the difference in desolvation energies which is used to calculate the London dG scoring function.

Subscript	Description
A, B	protein and or ligand volumes with I belonging to volume B
R_i	is the solvation radius of the atom I (taken as the OPLS-AA VDW sigma parameter plus 0.5 Å)
c_i	is the desolvation coefficient of atom i
u	is an energy term relating to the desolvation energy of the atom

2.2.2 GBVI/WSA ΔG (Chapter 3 and Chapter 5)

The GBVI/WSA ΔG[98] (Equation 2.3) is a force-field-based scoring function, which estimates the free energy of binding of the ligand from a given pose. It has been trained using the MMFF94x and AMBER99 force-field on the 99 protein-ligand complexes of the SIE training set[142]. The functional form is a sum of terms:

$$\Delta G = c + \alpha [\tfrac{2}{3}(\Delta E_{coul} + \Delta_{EvdW} + \beta\ \Delta SA_{weighted}$$

Equation 2.3: The equation used in the calculation of GBVI/WSA ΔG scoring function, the terms for this equation are defined in Table 2.3.

Table 2. 3: the terms used to calculate the GBVI/WSA ΔG scoring function (see equation 2.3)[98]

Subscript	Description
c	Represents the average gain/loss of rotational and translational entropy.
α, β	Are constants which were determined during training (along with c and, are force-field dependent). If not using an AMBER force-field, the parameters will be set by default to the MMFF trained parameters.
E_{Coul}	Is the coulombic electrostatic term which is calculated using currently loaded charges, using a constant dielectric of $\varepsilon_i=1$.
E_{sol}	Is the solvation electrostatic term which is calculated using the GB/VI solvation model.
E_{vdW}	Is the van der Waals contribution to binding.
$SA_{weighted}$	Is the surface area, weighted by exposure. This weighting scheme penalizes exposed surface area.

2.2.3 ASE (Alpha Spheres and Excluded Volume, Chapter 3)

The Alpha Spheres and Excluded Volume (ASE) scoring function[143] Is a knowledge-based, shape-based scoring function where the score is "proportional to the sum of the Gaussians $R_1R_2e^{-d2/2}$ over all ligand atom–receptor atom pairs and ligand atom–alpha sphere pairs. R_1 and R_2 are the radii of the atoms in Å, or are 1.85 Å for alpha spheres, d is the distance between the pair in Å. The proportionality constant has a default value of 0.035 kcal/mol[98]".

2.2.4 Affinity ΔG[98] (Chapter 3)

This knowledge-based scoring function estimates the enthalpic contribution to the free energy of binding using a linear function (Equation 2.4):

$$\Delta G = C_{hb}f_{hb} + C_{ion}f_{ion} + C_{mlig}f_{mlig} + C_{hh}f_{hh} + C_{hp}f_{hp} + C_{aa}f_{aa}$$

Equation 2.4: The equation used in the calculation of Affinity ΔG scoring[98].

Where the f terms fractionally count atomic contacts of specific types and the C's are coefficients that weight the term contributions to the affinity estimate (Table 2.4).

Table 2. 4: A table of the terms used to calculate the Affinity ΔG scoring function (see equation 2.4).

Subscript	Description
hb	Interactions between hydrogen bond donor-acceptor pairs. An optimistic view is taken; for example, two hydroxyl groups are assumed to interact in the most favourable way.
ion	Ionic interactions. A Coulomb-like term is used to evaluate the interactions between charged groups. This can contribute to or detract from binding affinity.
mlig	Metal ligation. Interactions between nitrogen/Sulphur and transition metals are assumed to be metal ligation interactions.
hh	Hydrophobic interactions, for example, between alkane carbons. These interactions are generally favourable.
hp	Interactions between hydrophobic and polar atoms. These interactions are generally unfavourable.
aa	An interaction between any two atoms. This interaction is weak and generally favourable.

2.2.5 BALLaxy[144] (Chapter 3)

Biochemical Algorithms Library (BALL) is an open-access software that provides a service for rescoring docking conformation generated using other software, using three different scoring functions. The first scoring function used is MM[145], an AMBER-based (Assisted Model Building with Energy Refinement) scoring function which is of the molecular mechanics type (Equation 2.5).

$$E_{binding} = E_{complex} - (E_{receptor} + E_{ligand})$$

Equation 2.5: the equation used to calculate values for the BALLaxy MM scoring function. $E_{binding}$ is the calculated binding energy of a protein-ligand interaction devised from the energy terms associated with the protein-ligand complex ($E_{complex}$), the energy term associated with the protein/receptor ($E_{receptor}$) and the energy term associated with the ligand (E_{ligand}).

The MM score estimates the binding energy of a protein-ligand interaction by optimising the atom-based relationships via the amber force-field. Energy values are calculated for the following

1. Atom types
2. Bond and angle parameters
3. Dihedral parameters
4. van der Waal parameters
5. Electrostatic energies

The second scoring function in BALLaxy is PB which is a Poisson-Boltzmann scoring function, i.e. another molecular mechanics-based model. This scoring function estimates the free energy binding ΔG using the following equation (Equation 2.6). For each protein-ligand complex, the ligand is subjected to up to 1000 steps of Cartesian coordinates minimization within the fixed protein structure using the Szybki minimizer and the MMFF94s force field[146].

$$\Delta G_{bind} = \Delta H_{elec}^{sol} + \Delta H_{vdW} - T\Delta S_{HPhob} - T\Delta S_{RotB}$$

Equation 2.6 where by ΔH_{elec}^{sol} represents the Poisson-Boltzmann electrostatics calculated by ZAP (a multilevel, multigrid solver)[147] (OpenEye, Santa Fe, NM, 2006) with Bond radii. ΔH_{vdW} is the van der Waals energy[146].

$T\Delta S_{HPhob}$ is calculated using Equation 2.7 and the temperature-dependent change in entropy due to the rotatable bonds in the complex becoming fixed upon complexation of the protein and ligand $T\Delta S_{RotB}$ is calculated using Equation 2.8[146]

$$T\Delta S_{HPhob} = \text{(surface area buried upon complex formation)} \times 0.006 \text{ kcal/mol Å}^2$$

Equation 2.7 - where the surface area term is computed with ZAP (OpenEye, Santa Fe, NM, 2006). The coefficient 0.006 kcal/mol Å^2 accounts for the partitioning of solute molecules between aqueous and organic phases[146].

$$T\Delta S_{RotB} = \text{number of rotatable bonds} \times 0.7 \text{ kcal/mol}$$

Equation 2.8 a rotatable bond penalty term, from the BALLaxy PB scoring function included to account for the loss of binding energy due to the freezing of the internal degrees of freedom of the ligand, upon binding to the protein to form a protein-ligand complex[146].

The third scoring function used is PLP which uses pair/piece wise linear potentials and is empirical[148,144,149]. This scoring function is a pairwise sum over all ligand and protein heavy atoms that estimates the binding potential of a molecule. The parameters used in the PLP scoring function are based on four different ligand atom types (donor, acceptor, both and nonpolar), which interact via steric and hydrogen bond potentials with the protein atom type[150]. Each pair of atoms has only one type of interaction, primary and secondary amines are classed as donors, oxygen and nitrogen atoms (without hydrogens) are classed as acceptors, OH groups, and water molecules are classed as both and carbon and nonpolar[150]. Figure 2.5 illustrates the how the pairwise linear potentials are derived.

Interaction Type	A	B	C	D	E	F
Steric	3.4	3.6	4.5	5.5	-0.4	20.0
Hydrogen bond	2.3	2.6	3.1	3.4	-2.0	20.0
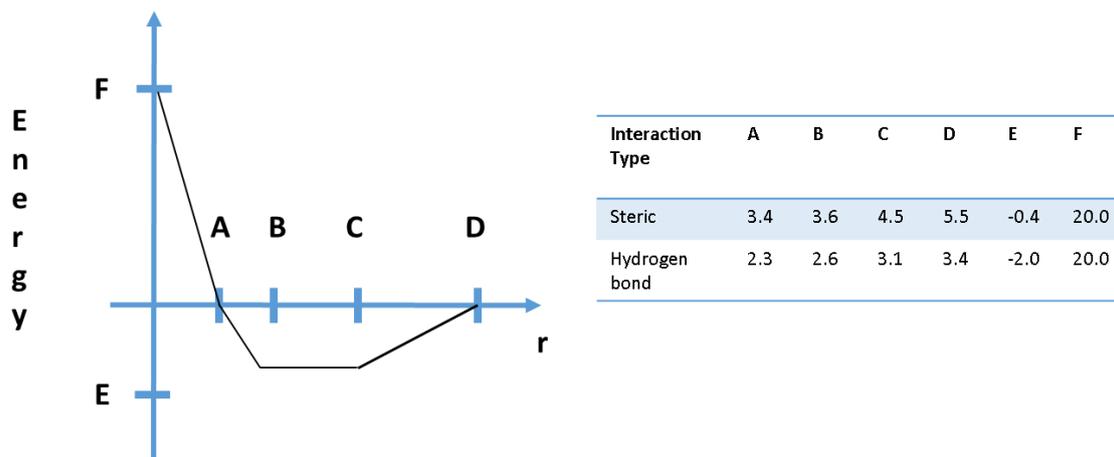
Figure 2. 5: The piecewise linear pairwise potential functions used for the protein-ligand interaction energy and a table to show the parameters of the atomic pairwise protein-ligand potentials. Values for A, B, C and D are distances (in Å) for two different interaction types (steric and hydrogen bond). The values for E and F are arbitrary energy units (for both steric and hydrogen bond interactions) and the letter r denotes interatomic distance. The diagram was adapted from Gehlhaal *et al*[150]

2.2.6 GoldScore & ChemScore (Chapter 3)

The GoldScore fitness function is the original scoring function provided with GOLD[151], and is the one selected by default for GOLD versions 5.0 and earlier. It has been optimised for the prediction of ligand binding positions and takes into account factors such as H-bonding energy, van der Waals energy, metal interaction and ligand torsion strain.

The GoldScore function is made up of four components[152]:

1. protein-ligand hydrogen-bond energy
2. protein-ligand van der Waals energy
3. ligand internal van der Waals energy
4. ligand torsional strain energy

The GoldScore fitness score is derived from the following equation

$$\text{GOLD Fitness} = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int}$$

Equation 2.9: the equation used to calculate a GOLD fitness score where S_{hb_ext} is the protein-ligand hydrogen bond score, S_{vdw_ext} is the protein-ligand van der Waals score, S_{hb_int} is the contribution to the fitness due to the intramolecular hydrogen bonds in the ligand and S_{vdw_int} is the contribution due to the intramolecular strain in the ligand[153].

The ChemScore fitness function, which is an alternative scoring function implemented in the GOLD software, estimates a term, ΔG that represents the total free energy change that occurs on ligand binding, and was trained by regression against binding affinity data for 82 complexes[154]. The ChemScore fitness function also incorporates a protein-ligand atom clash term and an internal energy term. ChemScore takes account of hydrophobic-hydrophobic contact area, hydrogen bonding, ligand flexibility and metal interaction. Although partly derived using binding affinity data, ChemScore values should not be used explicitly as values for binding energy or binding affinity as the data set the scoring function is based on is limited to 82 complexes and does not account for all possible protein-ligand interactions.

The ChemScore function was defined as

$$\Delta G_{GOLD\text{-}ChemScore} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot}$$

Equation 2.10: Where by ΔG_0 is the free energy change of reaction. ΔG_{hbond} is the hydrogen bond contribution to overall binding, ΔG_{metal} and ΔG_{lipo} are metal-ligand and lipophilic binding contributions (respectively) and ΔG_{rot} is a term that penalises flexibility.

2.3 Docking (Chapter 3 and Chapter 5)

2.3.1 Identification of putative binding sites in protein models in MOE (Chapter 3 and Chapter 5)

MOE's Site Finder uses the relative positions and accessibility of the receptor atoms to identify potential binding sites (areas of accessible tightly packed atomic regions with minimal solvent exposure). This is done by identifying hydrophilic or hydrophobic alpha spheres using two different probe radii: probe radius 1 is the radius of a hypothetical hydrophilic hydrogen bonding atom and probe radius 2 is the radius of a hypothetical hydrophobic atom. Individual alpha spheres are collected into separate sites by a double-linkage clustering algorithm.

2.3.2 Docking using MOE (Chapter 3 and Chapter 5)

Figure 2.6 is an overview of the docking methodology, for MOE, employed in both Chapter 3 and Chapter 5.
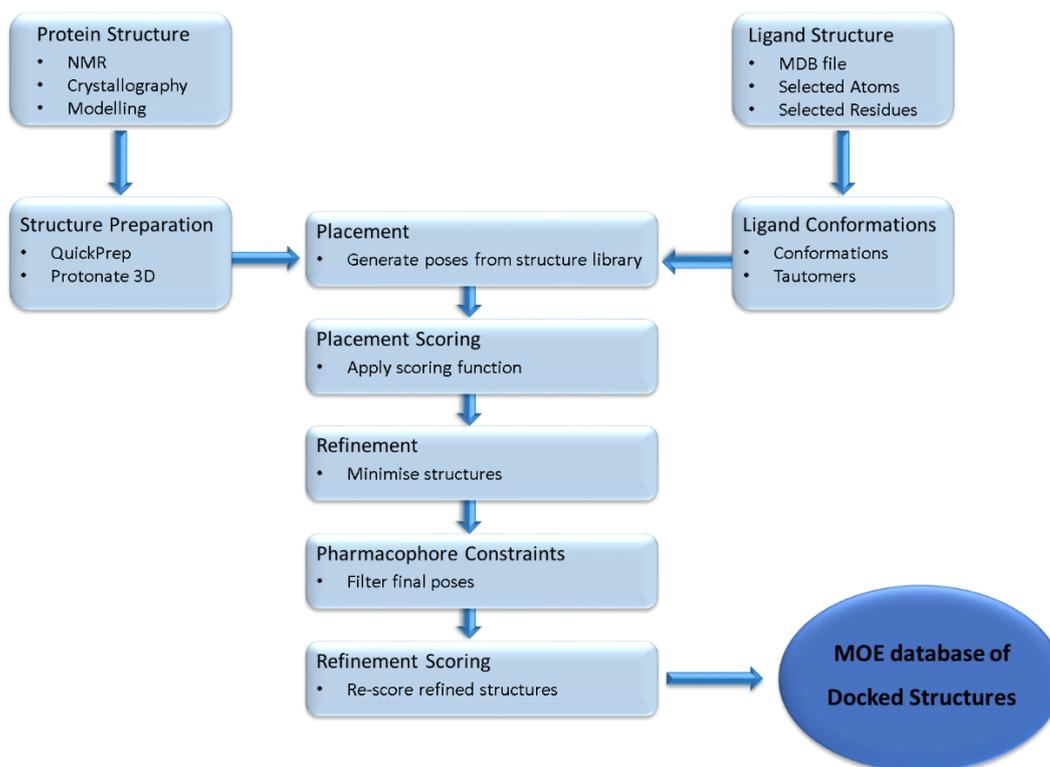
Figure 2. 6: an overview of the stages of a docking study adapted from MOE[98]

The MOE docking methodology employed in Chapter 3 and chapter 5 uses the following steps;

Placement - the placement stage of a docking study is conducted using the triangle matcher application, poses are generated by aligning ligand triplets of atoms on triplets of alpha spheres. Figure 2.7 illustrates the triangle placement method. Each of the generated poses is then assigned a score.
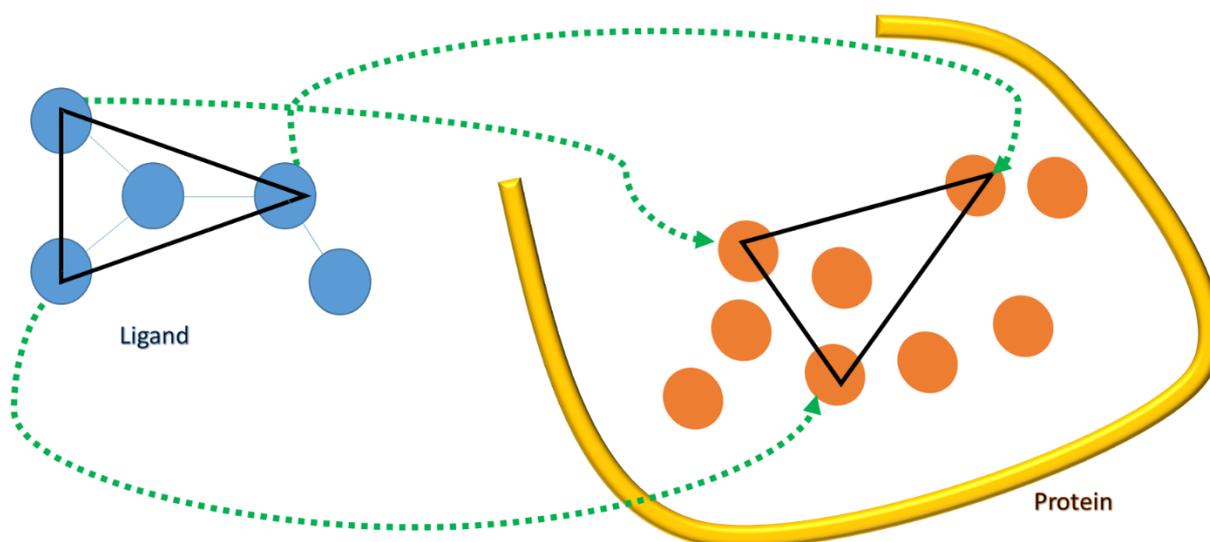
Figure 2. 7: Illustration of how the triangle placement method works. A triangle is formed based on a triplet of atoms within a ligand (represented by the blue circles), the triangle is then placed into the

binding site (illustrated by the yellow boundary), when the triangle matches with a triplet of atoms formed by the residues in the binding site (orange circles) a docked posed is established.

Initial Scoring - poses generated by the placement methodology can be rescored using one of the available methods. Typically, scoring functions emphasize favourable hydrophobic, ionic and hydrogen bond contacts.

Refinement - poses resulting from the placement stage can be refined using either the explicit molecular mechanics force-field method (eleven force-field parameters are available in MOE) or a grid-based energetics method.

Final Scoring - the final poses can be rescored using one of scoring schemes available in MOE.

2.3.3 Docking using Genetic Optimisation for Ligand Docking (GOLD) (Chapter 3)

GOLD uses a genetic algorithm (GA) to generate docked protein-ligand conformations. This type of algorithm uses the principles of biological evolution to discover optimal solutions to problems associated with molecular modelling and docking[151].

Identification of docked poses is carried out in five main steps:

The process begins with the development of a population, in the case of docking each docked pose is considered an individual and this particular pose (individual) is characterised by a set of variables termed "genes" which are used to form a "chromosome".

A fitness function is then assigned to each individual, the function equates to how well the individual competes with others in the population.

Based on the fitness function, the best individuals (docked pose) are selected to pass on their genes to the next generation of docked poses.

The fitness function is evaluated in six stages

1. Conformation of both ligand and binding site is generated
2. Least square fitting procedure
3. A hydrogen-bonding energy term is evaluated for the complex. This is the sum of individual bond energies for all ligand and protein donor hydrogen and acceptor atoms across the entire complex.
4. A pairwise steric interaction energy for all of the protein ligand atoms is calculated using a softened 4-8 Lennard Jones Potential (Equation 2.10)

$$E_{ij} = A/r_{ij}^8 - B/r_{ij}^4$$

Equation 2.11: 4-8 Lennard Jones Potential, where E_{ij}= the interaction energy between the atom pair, and r is the distance between the atoms. Adjustments are made for atoms involved in a hydrogen bond ($E_{ij} = 0$, r is scaled by a factor of 1.43), and all pairwise interactions across the complex are summed to give an overall 'complex energy'.

5. A value for the internal energy of the ligand in the ligand receptor complex is then calculated using a 6-12 Lennard-Jones potential (Equation 2.11) and the Tripos force-field torsional (Equation 2.12).

$$E_{ij} = C/r_{ij}^{12} - D/r_{ij}^6$$

Equation 2.12: 6-12 Lenard Jones potential equation, where E_{ij} is the interaction energy between the atom pair and r is the distance between atoms. C and D represent constants dependent upon the identities of the atoms involved.

$$E_{ijki} = 1/2 \, V_{ijki} \, [1 + (n_{ijki}/|n_{ijki}|).\cos(n_{ijki}.\omega_{ijki})]$$

Equation 2.13: Tripos force-field torsional, where E_{ijki} is the torsional energy associated with four consecutively bonded atoms, ω is the torsional angle, n is the periodicity and V is the barrier to rotation.

6. The three calculated energy terms (Equations 2.10, 2.11 and 2.12) are added together to give an overall fitness

A pair of parent individuals are mated and a crossover point is chosen at random and "offspring" poses are created from the parent poses.

Mutation is introduced to certain new offspring as a way of maintaining diversity within the population. Finally, when the genetic algorithm can no longer produce offspring that are significantly different to the established population a termination step is introduced, and the docked pose is returned

2.3.4 Spearman's Rank (Chapter 3)

Spearman's rank (ρ) is a statistical method employed to assess correlation between two ranked datasets (see Equation 2.13[155]) and is used in this thesis to compare the rankings of conformations generated for a dataset of small molecules in different docking experiments, to their relatively ranked, experimentally determined biological activities against a given receptor.

$$\rho = 1 - 6 \sum d^2 / (n^3 - n)$$

Equation 2.14: equation used to calculate Spearman rank correlation coefficients (ρ), where d is the difference between the ranked values across the 2 ranked datasets and n represents the number of individual members of the dataset.

2.4 QSAR (Chapter 4)

2.4.1 Identification of Diverse subsets

In order to identify diverse subsets to exploit in the building and testing of QSAR models, the diverse subset algorithm in MOE was used. The MOE Diverse Subset application assigns a diversity ranking order to all entries in a database. The N reference set entries (molecules included in the database) are all given a rank of 1. The diverse subset entries are given ranks 2 (highest rank) to (M+1), where M is the number of diverse entries to be ranked.

Hence, the diverse subset is selected by choosing entries that are the farthest from both the N reference entries and the currently-ranked entries.

The farthest entries were determined using FP:MACCS representations of the molecules (see Section 2.6.3) as a discriminant. Distance was computed using the FP:MACCS calculated from both the N reference entries and the entries that had yet to be ranked. The diverse subsets were then identified using the following steps;

1. The minimum of its distances between the ranked entries and the reference set are computed
2. The entry with the largest such minimum distance is deemed to be the farthest from the reference entries.

2.4.2 Similarity coefficients (Chapter 4 and 5)

Similarity coefficients can be calculated, and average coefficients used as a measure of how similar a molecule is to other molecules in a dataset. Tanimoto coefficients (T_c) were calculated using the following equation (Equation 2.15):

$$T_c = \frac{N_{ab}}{N_a + N_b - N_{ab}}$$

Equation 2.15: The equation used to calculate T_c where N_a is the number of bits on a set in molecule A, N_b the number of bits on a set in molecule B and N_{ab} the number of bits set on common to both molecules[156].

2.4.3 Building and evaluating QSAR models (Chapter 4)

A QSAR model is a numerical formula that is developed using a predetermined training set and evaluated for its predictive ability be using a test set of molecules, representative of those in the training set, but critically which have not been used to derive the QSAR model. The MOE QuaSAR suite produces linear regression models, which can be used to predict experimental results and compared to experimentally derived data. In its simplest form, a linear model specifies the relationship between a dependent (response) variable Y, and a set of predictor variables (X).

A correlation coefficient (r^2 value) between experimentally derived data and predicted values is generated in addition to a cross-validated correlation coefficient (q^2 value) and both are used as measures of model quality. The parameters of the linear models were determined using the method of Partial Least Squared (PLS).

2.4.3.1 Partial least squares

Partial least squares is a statistical method used for developing predictive models when there are many factors involved and these factors are highly collinear. Development of QSAR models can include the use of hundreds of different molecular descriptors and it is the aim of model development to identify which descriptors are important for predictivity. In PLS, the descriptors are described as latent factors/variables.

The identification of latent variables important for predictivity is achieved by establishing a relationship between two data sets. Latent variables are chosen in such a way as to provide maximum correlation with the dependent variable, e.g. biological activity. Figure 2.4 illustrates a hypothetical data set with two variables (X_1 and X_2) from graph **a** it can be seen that the variables are highly collinear. From this, two latent (orthogonal) variables can be applied (t_1 and t_2) that are a linear combination of the original variables. This produces a model that relates the activity to the first latent variable t_1.
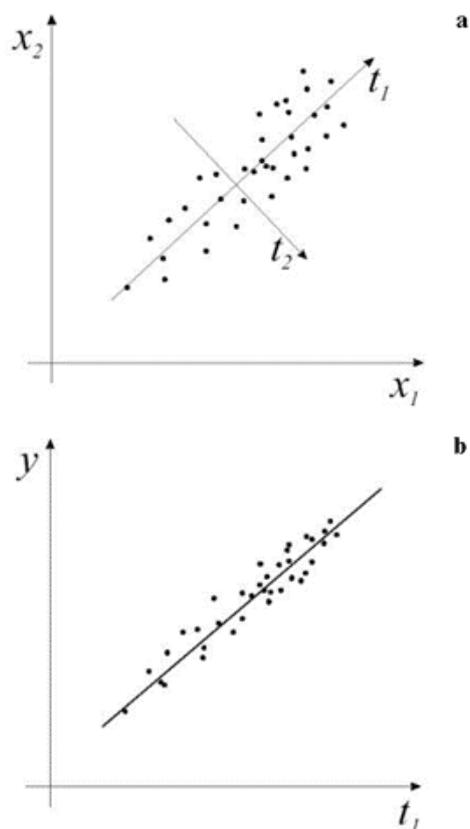
Figure 2. 8: Change of original descriptors to latent variables (a) and construction of activity model containing one PLS factor (b)[157].

2.4.3.2 Leave One Out Cross validation

Leave one out Cross-validation is a procedure by which portions of the training set are systematically "left out" of the model building and subsequently used to simulate "new" models, the predictive ability of which are compared to the original. This is conducted for each of the variables (data points) within the training set and the differences in values obtained from each leave one out cross validation are used to calculate the root mean square of error of cross validation (RMSECV) and the revised correlation coefficient. This latter value is often referred to at the q^2 value[119].

Cross validation in this manner gives an indication of how sensitive the models are to the training set used to generate them, and gives an indication via the q^2 value how robust the model is.

2.5 Pharmacophores (Chapter 5)

2.5.1 Clustering systems

JChemAxon was employed as a clustering method, using the Maximum Common Substructure (MCS) approach in ChemAxon[158]. In essence this means that structurally similar molecules are clustered with one another in a hierarchical manner.

This method treats all molecules as initial separate clusters. All clusters are evaluated for similarity and molecules that have similarities are then merged to form a second layer of clusters. This process will be repeated until all clusters that can be merged are. The forming of new clusters causes the dataset to split into levels forming a Dendrogram (see Figure 2.8). The top level of the dendrogram contains the smallest common substructure containing at most 9 atoms (as defined by the default settings). The bottom level contains all the individual compounds in the dataset which are referred to as the "leaves" of the dendrogram.
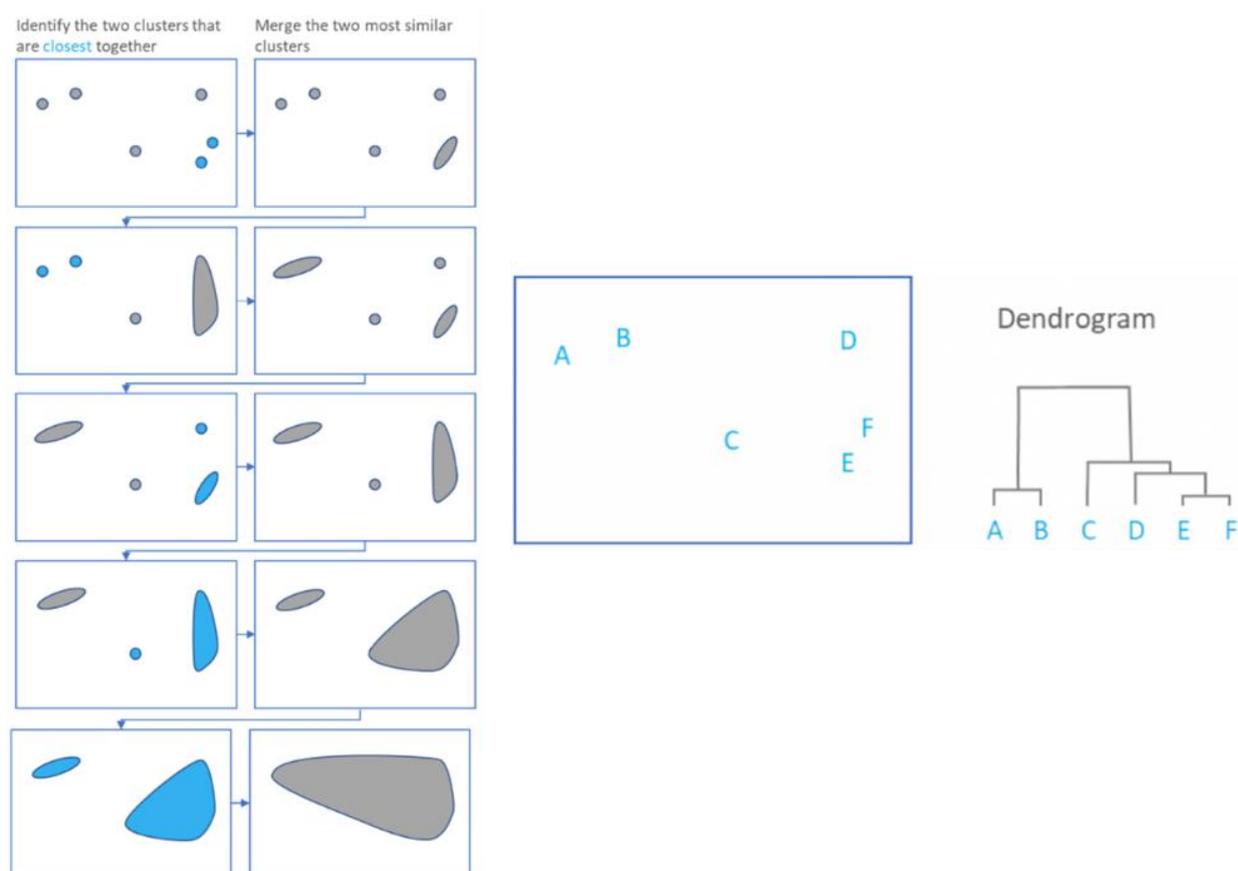
Figure 2. 9: Diagram to illustrate the method of hierarchical clustering[159].

2.5.2 Pharmacophore Model generation and validation (Chapter 5)

2.5.2.1 Flexible alignment

In order to generate a ligand-based pharmacophore, it is necessary to flexibly align the compounds in a dataset to one another. The flexible alignment application in MOE produces flexibly aligned motifs based on a collection of small molecules. Each motif is given a score that quantifies the quality of the alignment in terms of both internal strain and overlap of molecular features[160].

Alignment motifs are calculated in MOE using the internal energy of the molecules and Gaussian feature densities. There are eight Gaussian feature densities used to develop an alignment score. These are:

1. Volume
2. Aromatic
3. Donor
4. Acceptor
5. Hydrophobe
6. logP
7. Molar refractivity
8. Surface exposure

The overlap of these features will contribute to the final alignment score.

Alignments are "sampled" using a RIPS-style (Random Incremental Pulse Search) procedure that incorporates the quantitative measure of goodness of an alignment defined above into the optimization stage. Such a procedure happens via the following three stages:

1. Generation of a conformation for each molecule by randomly rotating bonds and inverting unconstrained chiral centres followed by performance of a rigid-body optimization of the similarity function.
2. Minimization of the grand alignment score S with respect to the coordinates of all of the atoms. S is simply a combination of the similarity measure (see above), and the average strain energy of the molecules (denoted by the symbol U).

3. If the new configuration has not been seen before (RMSD greater than a predetermined threshold, up to topological symmetry), then a value k is set to otherwise, k = k + 1. If k is greater than a predefined threshold amount the search is terminated.

2.5.3 Enrichment Factors (Chapter 5)

Enrichment factors were calculated as a measure of pharmacophore model ability to extract expected compounds preferentially from a virtual screening experiment. The E_f were used to investigate how modified pharmacophore models compared to the initial models generated, and if the models performed better than random with regards to extracting the expected molecules from a database. E_f were calculated at three different stages of the pharmacophore search, to ascertain how well the model was able to retrieve selected compounds in the early stages of the pharmacophore search.

Enrichment rate $_{x\%}$ =

$$\frac{(\% \text{ of relevant Synthetic Cannabinoids }_{\text{at } x\%} \div \% \text{ of New Psychoactive Substances database }_{\text{at } x\%})}{(\text{Total number of Synthetic Cannabinoids} \div \text{total number of New Psychoactive Substances})}$$

Equation 2.16: The equation[161] used to determine the enrichment rate at different stages of the pharmacophore search, x% is the percentage of the database screened (2, 5 and 10%).

The equation in 2.16 was used to calculate E_f for the first 2, 5 and 10% of database screened.

2.6 Virtual screening

2.6.1 Virtual screen of database using pharmacophore model (Chapter 5)

The Pharmacophore search application in MOE was used to facilitate a virtual screen. The main characteristics of a MOE pharmacophore search are described in Table 2.5.

Table 2. 5: A description of the characteristic used in the MOE pharmacophore search application

Search Characteristic	Description
Boolean Expression	assigns features as donor or not cationic using Boolean expressions
Substructure Matching	uses SMARTS patters to search for functional groups
Volume and Shape Filtering	restrictions can be applied to the shape of the matched ligands
Partial Matching	pharmacophore query features can be marked as essential and must be matched for a hit to be identified. Other features may be left unmatched and considered a "partial match"
Group Constraints	can be applied to match a group of selected query features
Existing Conformations	uses external conformation databases
Systematic Matching	All possible matches of a ligand conformation and the query are systematically examined

Virtual hit molecules were identified using the default systematic matching setting and the hits were ranked using RMSDx values (weighted RMSD values plus an applied penalty for each missing feature).

The MOE pharmacophore search compares each molecule conformation from the provided database to the designated pharmacophore. The algorithm used for generating matches works as follows:

1. A single molecule is screened for matches against the query features for the generated pharmacophore.
2. If the molecule does not satisfy the query features it is not reported as a hit.
3. If the distances between the molecule annotation points do not agree with the distances between the matched query features (after discounting the radial tolerances), the molecule is rejected.
4. If the molecule is retained, the molecule annotation points and the matched query features are superposed against one another using rigid-body superposition. If the molecule features fail to fall within the specified tolerance radii of the matched query features, the mapping is

rejected. During the superposition, the relative weights of the matched query features are inversely proportional to their radii.

5. If any of the molecule atoms violate any of the volume restrictions in the query, the molecule conformation is rejected.

6. If the molecule has successfully passed all of the above criteria it is retained as a "hit" against the designated pharmacophore. All reported hits are compiled into an output database.

2.6.2 FP: MACCS structural keys (Chapter 5)

MACCS structural keys were developed for the purpose of assigning numerical representations to label substructures of molecules (hereafter referred to as keys), and then linearly combining these keys together to provide an information-rich one-dimensional representation of a three-dimensional chemical structure. Each key describes a small substructure consisting of approximately one to ten non-hydrogen atoms. FP:MACCS keys are derived from 166 public MDL MACCS structural keys[162]. The 166 different descriptive keys used encode for atom types, bond types and aromatic elements[162]. Figure 2.9 illustrates how a four number MACCS key is derived for the molecule diazepam, this highlights how informative a 1D representation of a molecule can be. This simplistic way of representing 3D information using MACCS keys allows for efficient manipulation of molecular databases, and is used in this thesis to differentiate between structurally similar and structurally distinct molecules in the derivation of test and training sets for QSAR models, and to ensure that maximum structural diversity was ensured in the selection of molecules from the virtual screening experiments
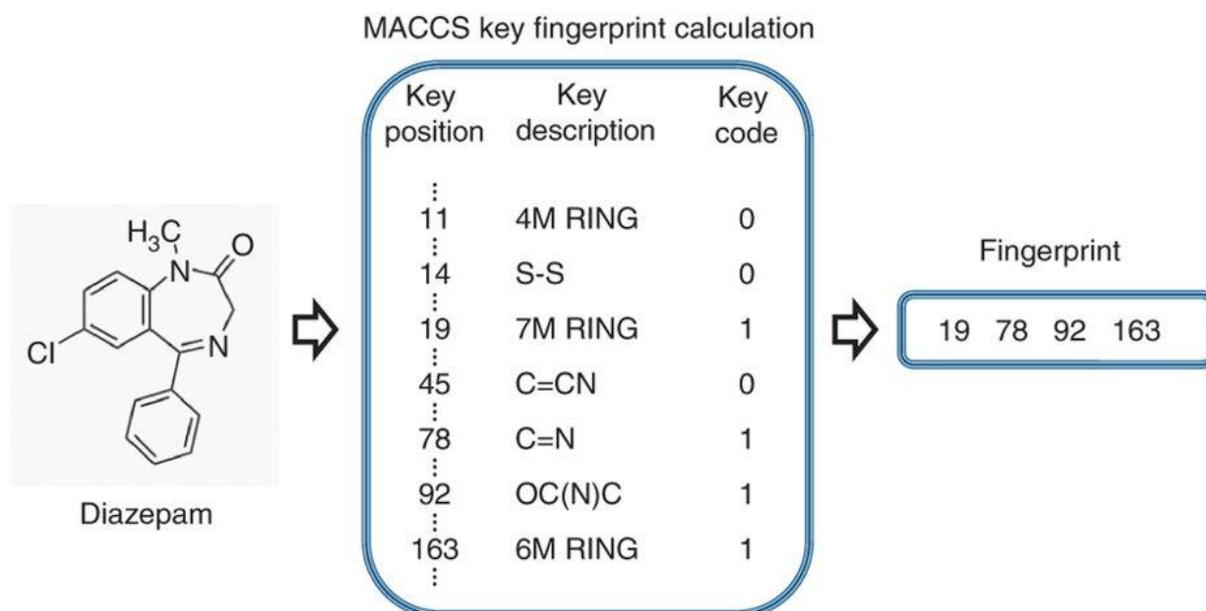
Figure 2. 10: Diagram to illustrate how a MACCS key fingerprint for diazepam is calculated[163].

2.7 Biological screening (Chapter 6)

Compounds that have been identified as being potential CB1 agonists were evaluated for their biological activity. A cAMP assay was carried out to determine the accumulation of cAMP generated by the selected compounds. This method was employed as changes in intracellular cAMP has been shown to indicate the activation of the CB1 receptor.

The second method employed was the use of an electrical field stimulation (EFS) study. An organ bath which contained sections (approximately 1 cm in length) of rat ileum submerged in Krebs buffer was used to carry out this investigation. Electrical impulses were applied to the tissue and simultaneously a potential CB1 agonist was added to the organ bath. This study was used to identify any potential activity the compounds of interest had on the contractive force of the tissue sample. The details of both these studies are discussed in Chapter 6.

Chapter 3

In Silico investigations into selectivity between the monoamine transporters (MAT) Dopamine Transporter (DAT), Norepinephrine Transporter (NET) and Serotonin Transporter (SERT).

3.1 Introduction

Monoamine transporters (DAT, NET and SERT), are located in the plasma membranes of the monoaminergic neurons. They consist of twelve transmembrane helices and are responsible for the release or reuptake of the monoamines dopamine (DA), norepinephrine (NE) and serotonin (SER)[31,35,164–166]. As monoamines are vital in the role of normal brain function, the regulation of these chemicals is critically important[167].

New psychoactive substances (NPS) act in a similar manner to known classic illicit substances[24]. This is due to the structural similarity between NPS and known illegal recreational drugs such as ecstasy, amphetamines and cocaine[168], (see Figure 3.1).
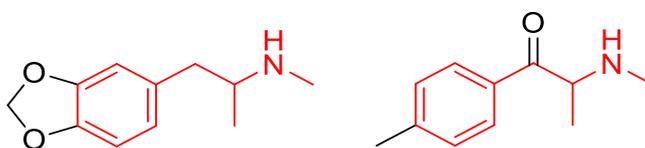
Figure 3. 1: The chemical structure of MDMA (ecstasy) on the left and mephedrone (bath salts or meow meow) a popular NPS on the right. Highlighted in red are the portions of carbon skeleton of the two molecules that are identical.

Due to similarities in chemical structure with known psychoactive compounds, a large number of NPS will act on DAT, NET or SERT[24,67,68,169] and in many cases any given NPS will demonstrate affinity, of different magnitudes, with all three MAT[48,43,52]. It is well documented that there is a large degree of ligand promiscuity that occurs between the three MAT[3,170,27,171] i.e. drug molecules that have a high affinity for one MAT are likely to bind to all three MAT, and extensive research has been conducted into trying to develop selective monoamine reuptake inhibitors such as the selective serotonin reuptake inhibitors (SSRIs)[27,37,171–176]

In order to understand how these molecules are forming interactions with the transporter proteins, and hence how differences in affinity for a compound between the MAT isoforms may arise, molecular docking studies were carried out.

When the present study was initially conducted (in 2015), there were no experimental structures (e.g. X-ray crystal structures) for the human monoamine transporters, and so comparative/homology models of human DAT, NET and SERT were used. Homology models are routinely used in docking studies when no experimentally derived structures have been elucidated. Although there are some limitations to this method, with the use of protein validation techniques, robust results can be obtained[177,178].

Since the initial study was completed in 2015 a number of dDAT crystal structures complexed with a variety of ligands have been published e.g. the model with the PDB accession code 4XP9 which is the dDAT crystal structure complexed with D-amphetamine (at 2.8 Å)[179]. In early 2016 the human serotonin transporter X-ray structure complexed with the antidepressant S-citalopram was also elucidated[180]. Consequently, complementary molecular docking studies have been carried out on these emergent experimental structures and results compared to those from the initial homology modelling studies. The results from the original docking experiments, and the comparison between these studies, and the experiments carried out on the crystal structures is discussed in detail in this chapter. Conclusions are drawn as to whether or not it is possible to explain the difference in experimental affinity between DAT, NET and SERT based on an examination of how the protein and ligands interact with each of the isoforms.

The comparative DAT, NET an SERT models used in the study were derived from the template 4M48 (*Drosophila melanogaster* dopamine transporter, dDAT) and accessed from the proteinmodelportal.org (2015) server. These models were rigorously validated using complementary, yet independent methodologies prior to docking studies beginning, and deemed viable for use in molecular docking studies.

As such, the first aim of this study was to establish if MAT homology models provided a sufficient representation of the human protein structure in terms of their ability to replicate trends in observed experimental activities via their docking scores. The second aim was to examine whether it was possible to use molecular docking as a methodology to help explain what gives rise to selectivity between DAT, NET and SERT. The third aim was to compare results obtained from the MAT homology to newly available crystal structures.

3.2 Methods

3.2.1 Homology models

Three comparative models of the sodium symporter neurotransmitters (NSS), DAT (accession number Q01959), NET (accession number P23975) and SERT (accession number P31645) were obtained as protein database (PDB) files from www.proteinmodelportal.org.

3.2.2 Protein model validations

The DAT, NET and SERT homology models were uploaded to the Rampage[135] (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) server to generate Ramachandran plots. The

Structure Analysis and Verification Server (SAVES) metaserver (http://nihserver.mbi.ucla.edu/SAVS/) was used in order to produce ERRAT[139] and Verify3D[138] plots for the models. The data obtained from Rampage and SAVES was analysed for each of the NSS models and all residues that violated one or more of the three validation criteria were recorded.

As all three models were based on the same template *Drosophila melanogaster* dopamine transporter (PDB accession code 4M48), this structure was also analysed using the Rampage, and SAVES software to provide a benchmark from which to compare the quality of the homology models.

3.2.3 Protein model sequence and identity.

All three MATs and the template protein were superimposed and aligned using the "Align/Superimpose" application in MOE. This generated pairwise percentage sequence and percentage identity values between the homology models and the template.

3.2.4 New Psychoactive Substances dataset

The dataset used for the docking studies to the homology models and crystal structures was obtained from the study conducted by Iversen *et al*[181]. It comprises 21 NPS each of which have experimentally measured biological activity data (Ki values) for each of the isoforms DAT, NET and SERT.

3.2.5 Identification of MAT binding sites.

Binding sites for all three MAT were elucidated in MOE and then cross referenced with available literature to ensure that all residues that had been identified as important to binding were contained within the putative binding sites defined. The default settings in MOE were used to define the cavities i.e. probe radius 1: 1.4 Å probe radius 2: 1.8 Å, connection distance 2.5 Å and a minimum site size of 3 residues. The residues identified from the literature as being important for protein ligand interaction in the MAT that were not identified by MOE were manually added to the binding site composition prior to docking studies taking place.

3.2.6 Docking in MOE

Protein models were prepared using the "quick prep" application in MOE, and the binding site was defined using the "site finder" application. The database molecules were protonated and compiled into MDB files for the docking study. Induced fit docking studies were carried out in MOE, using the MMFF94 force field. For comparative purposes, a secondary study using the AMBER10: EHT force field was also conducted.

In the docking application within MOE the default setting was selected. Parameters used are the triangle matcher placement, rigid receptor refinement, London dG & Generalised-born volume integral/weighted surface area (GWVI/WSA). dG scoring functions were applied. The number of docked poses generated was terminated after a maximum of 30 poses had been established or until the conformation of the ligand poses reached a default RMSD cut-off of 3.0 Å.

The 21 molecules from the NPS dataset were docked into each of the MAT models in turn. The docked poses for each ligand were analysed based on the scoring function value obtained, ligand interaction diagrams and the 3D conformation of the docked ligand.

3.2.6 Docking in Genetic Optimisation for Ligand Docking (GOLD)

Genetic Optimisation for Ligand Docking (GOLD) is an alternative piece of docking software that uses a genetic algorithm to position the ligands in the binding cavity. This type of algorithm is based on the principles of biological evolution[182,183]. PDB files for DAT (accession number Q01959), NET (accession number P23975) and SERT (accession number P31645) were prepared in MOE. The native ligand (DA for DAT, NE for NET and SER for SERT) was then docked into the corresponding MAT homology model. The protein-ligand complex which had the highest S value was saved as a PDB file and uploaded to the GOLD visualizer HERMES.

To prepare the protein structures for docking the "add hydrogens" application was used. To define the binding site, a known binding site residue (see Table 3.1) was isolated for each MAT isoform and then using the Cartesian coordinates of the residue (obtained by interrogating the PDB file) a binding site cavity was constructed by using the default setting in GOLD which defines the binding site as all residues within a 15 Å radius of the selected residue.

Two scoring functions were implemented during the docking study using GOLD; GOLDscore (an empirical scoring function) and CHEMscore (a molecular mechanics-based scoring function, see Chapter 2 for details). The Genetic Algorithm default settings were applied with a population size 100, selection pressure 1.1, number of operations 100,000, number of islands 5, niche size 2, migrate 10, mutate 95, and crossover 95. Once the scoring function and rescore scoring function was selected the Iversen dataset was docked and ranked according to the scoring function results.

3.2.7 Spearman's Rank (ρ)

To determine the extent of the correlation between the relative rankings of the experimentally observed values and the relative rankings of the computationally generated scoring function results, Spearman's rank correlation coefficients (Eq 3.1) were calculated.

These pairwise comparison values were calculated for the experimental data with respect to all scoring functions used in the docking studies and the individual components of the S scoring function used in MOE (i.e. rmsd_refine, E_conf, E_place and E_refine). Correlation coefficients were checked for statistical significance using critical value tables[184].

$$\rho = 1 - 6\textstyle\sum d^2 / (n^3 - n)$$

Equation 3.1: Formula used to calculate Spearman's rank correlation coefficient (ρ) between relatively ranked experimental and relatively ranked computationally generated scoring function data.

Where

d = the difference in value between experimental and computational rank value,

n = the number of samples in dataset

3.2.8 Consensus scoring

In order to remove potential bias from the docking results, a number of different scoring functions were used to recalculate scores for the poses generated by the docking studies. The openly accessible server Galaxy/Ballaxy[149] (https://ballaxy.bioinf.uni-sb.de/) was used to re-score the docked poses created in MOE. The scoring functions used to achieve this were (1) Molecular Mechanics (MM), (2) Pairwise Linear Potential (PLP) and (3) Poisson-Boltzmann (PB)[148,185,186].

The ranked data for all the different scoring functions were then compared to the ranking of the experimental data using Spearman's rank. A Spearman's rank was calculated based on the ranking trend of nine scoring functions.

Subsequently, a Spearman's rank value for experimental relative ranking of highest to lowest affinity vs predicted relative ranking of highest to lowest affinity was calculated based on all the docked data for each of the MAT isoforms.

3.3 Results

3.3.1 Homology models

At the time of the initial study (2015) there were no human crystal structures available for DAT, NET or SERT. As such, homology models for human DAT, NET and SERT isoforms were obtained from proteinmodelportal.org. The homology models were validated, with regards to their structural properties, before being used in subsequent docking studies. Homology models are built on

experimental structures (such as x-ray crystals) that share a similar primary structure[177] to the desired protein, whereby these experimental "templates" are used to predict the likely 3D conformation (homology model) for a "target" structure (homology model). For the DAT, NET and SERT homology models the template used was the crystal structure 4M48[187]. The homology models are available from the SWISS-MODEL repository[188].

The homology models for DAT, NET and SERT have sequence identities with the template ranging from 53-55% to the 4M48 template. It is generally considered that models that have over 50% sequence identity to its template are suitable to be taken forward drug discovery investigation[110].

From Figure 3.2 it can be seen that there is a high level of structural conservation between the secondary structural elements in all three MAT's and the template structure, the difference between the structures arises in the orientation of the extracellular loops, which can easily be seen.
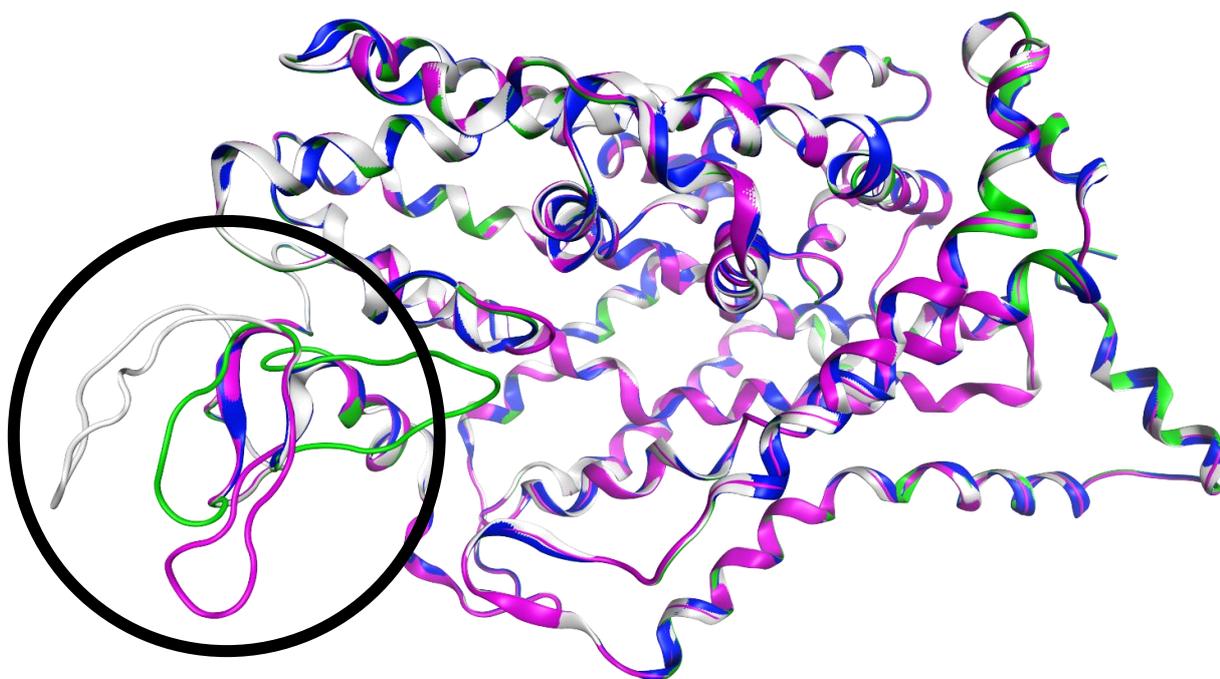
Figure 3. 2: Visual representation of the superimposed template 4m48 (blue) and homology models of DAT (Q01959, green), NET (P23975, white) and SERT (P31645, magenta) showing a largely conserved secondary structure between the template and the MAT models but variation in loop regions (circled in black).

Figure 3.3 is a matrix of the residue identity percentages between the three homology models and the template crystal structure. The bigger the percentage, the greater the similarity between two protein structures[189]. As stated by Chothia and Lesk[190] homology models that have a sequence identity of >50% with the template will provide a close general model. Figure 3.3 shows that the template crystal

structure appears to be most similar to the NET homology model but all models have a pairwise sequence identity with the template and each of the other isoforms of greater than 50% and as such can be taken forwards into docking studies, subject to protein validation.

	4M48	DAT	NET	SERT
4M48		54.8	58.8	53.1
DAT	54.8		69.9	50.7
NET	58.8	69.9		52.7
SERT	53.1	50.7	52.7	

Figure 3. 3: Matrix of residue identity percentage values for 4M48, DAT, NET and SERT.

3.3.2 Homology model validation

The three homology models and the template structure were analysed for structural validity using three protein validation tests: Ramachandran (RC) plot analysis[135], Verify3D[115] analysis and ERRAT[139] analysis.

These validation methods were used as a quality sense check of the models and to identify any residues implicated in structural errors[190], as these could, in turn, produce errors when using the models in docking studies. Ramachandran analysis looks at the overall quality of the backbone of the protein structure, Verify3D analysis looks at compatibility of the 3D structure of the model with the amino acid sequence and ERRAT analysis looks at the distance between pairs of non-bonded atoms (CC, CN, CO, NN, NO and OO)[116]. All techniques are independent, yet complementary to one another and are in combination are useful to identify regions of error in a protein model.

3.3.2.1 Ramachandran Analysis

Ramachandran plots are a geometric validation method which analyse the three main components of a proteins structure[135]. The components are the backbone conformation, the side chain conformation and the Cα geometry[135]. As protein structure and function are heavily related, it is vital that the protein backbone has a structurally appropriate conformation, and a homology model which was unable to demonstrate a structurally appropriate conformation would be of limited use in docking studies.

Figure 3.4 is a Ramachandran plot generated for the template structure 4M48, and the homology models of DAT, NET and SERT using RAMPAGE[135].
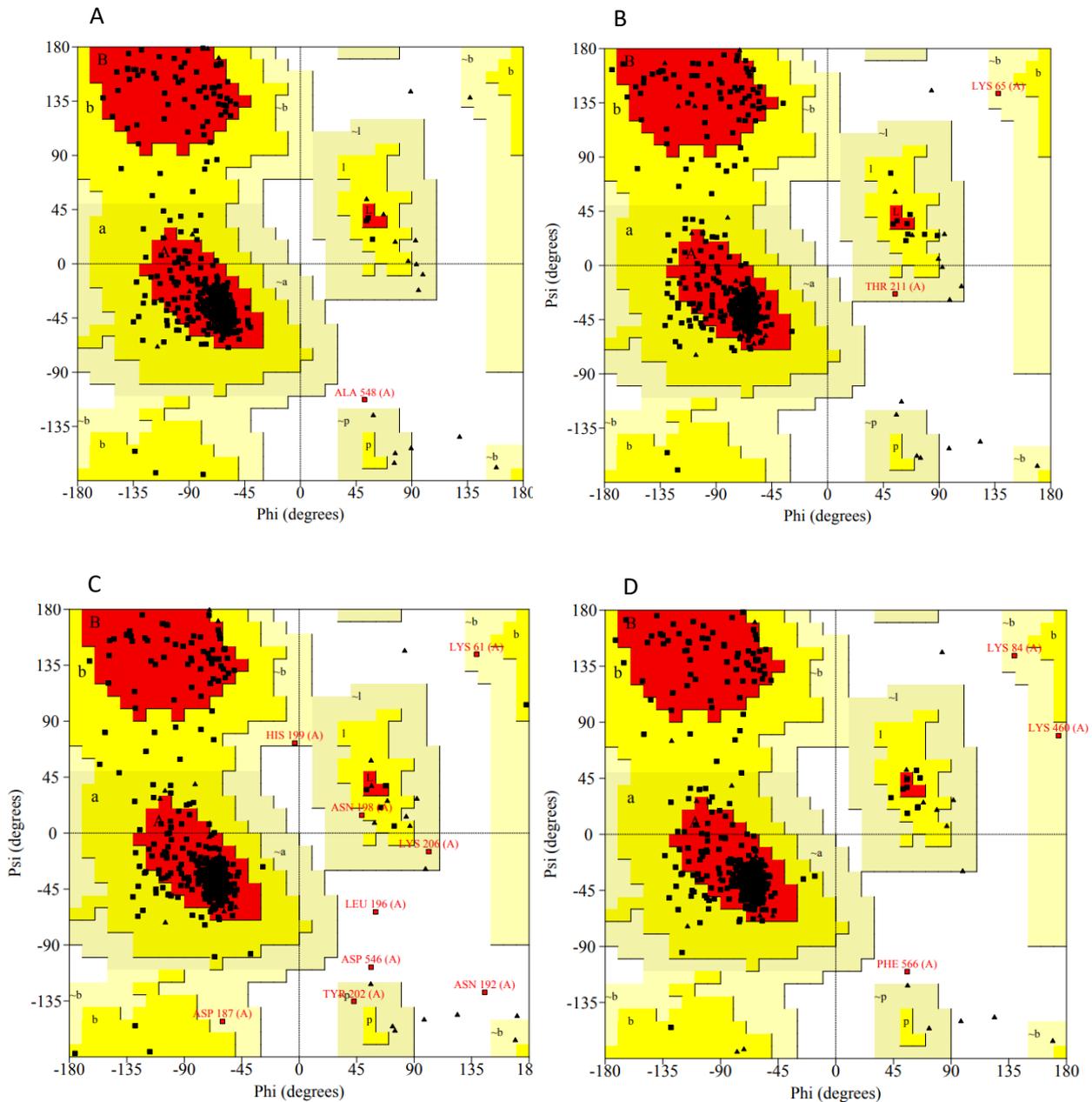
Figure 3. 4: Ramachandran plots for 4M48 (**A**), DAT (**B**), NET (**C**) and SERT (**D**). RC analysis plots the torsional angles (Φ, x-axis and Ψ, y-axis) of all the residues. The plot is split into areas according to secondary structure (regions labelled B/b for β-sheets, A/a for α-helices and L/l for left handed helices). The conformations of residues are categorised into four groups: most favoured (which are found in the red sections), allowed residues (plotted in the yellow sections of the graph) additionally allowed conformations (cream) and disallowed conformations found in the white sections of the graph.

The number of residues that were in the disallowed region of the RC plot was less than 2% for all of the homology models and less than 1% for the template structure (see Table 3.2 for results). Further analysis showed that none of the amino acid residues that were found to be in the disallowed region on the Ramachandran plots were part of the putative MAT binding sites. This gives confidence that the residues that make up the binding site will not produce erroneous docking conformational poses, as a consequence of an inappropriate overall protein fold.

For DAT, NET and SERT the residues that were found in the outlier region of the RC plot were all residues that were found in loops on the models. This was not unexpected as the majority of reported errors found in homology models are due to poorly predicted loop structures within a protein[191].

3.3.2.2 Verify 3D Analysis

Verify3D analyses the compatibility of a proteins 3D structure with its 1D structure (amino acid sequence). Each residues is categorised based on its environment and location and is used to evaluate the overall fitness of the amino acid sequence in relation to its 3D environment[115].The results obtained for DAT, NET and SERT for Verify3D (Table 3.1 and Figure 3.5) show that between 84-94% of residues obtained a positive value which indicated a favourable environment (conversely a negative value for a residue indicates it is in an unfavourable environment).
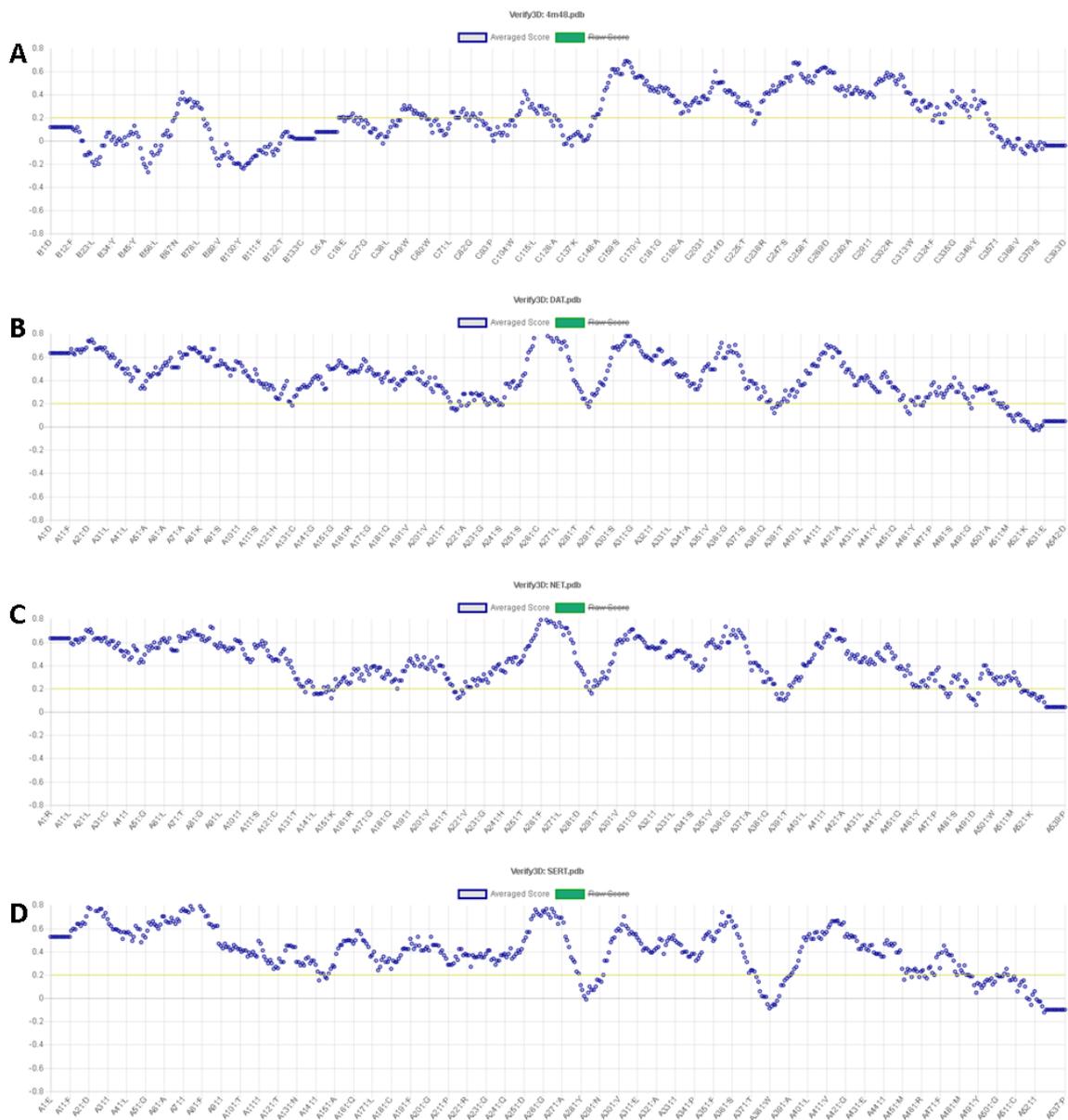
Figure 3. 5: Verify3D plots for 4M48 (A), DAT (B), NET (C) and SERT (D). The blue dots represent the average score based on a 21-residue sliding window. A negative average score indicates a residue that does not have a favourable 3D-1D compatibility, a positive average score indicates a good 3D-1D compatibility.

3.3.2.3 Errat Analysis

ERRAT analysis (Figure 3.6) looks at the non-bonded interactions between carbons, nitrogen and oxygen atoms within a protein. The results obtained for DAT, NET and SERT for ERRAT showed that all models had an overall quality factor of >85%, this means that at least 85% of the non-bonded interaction between carbon, nitrogen and oxygen within the model were at acceptable/expected

distances[139]to one another, and this is an indication that overall the model was of sufficiently high quality to be used in subsequent docking studies.
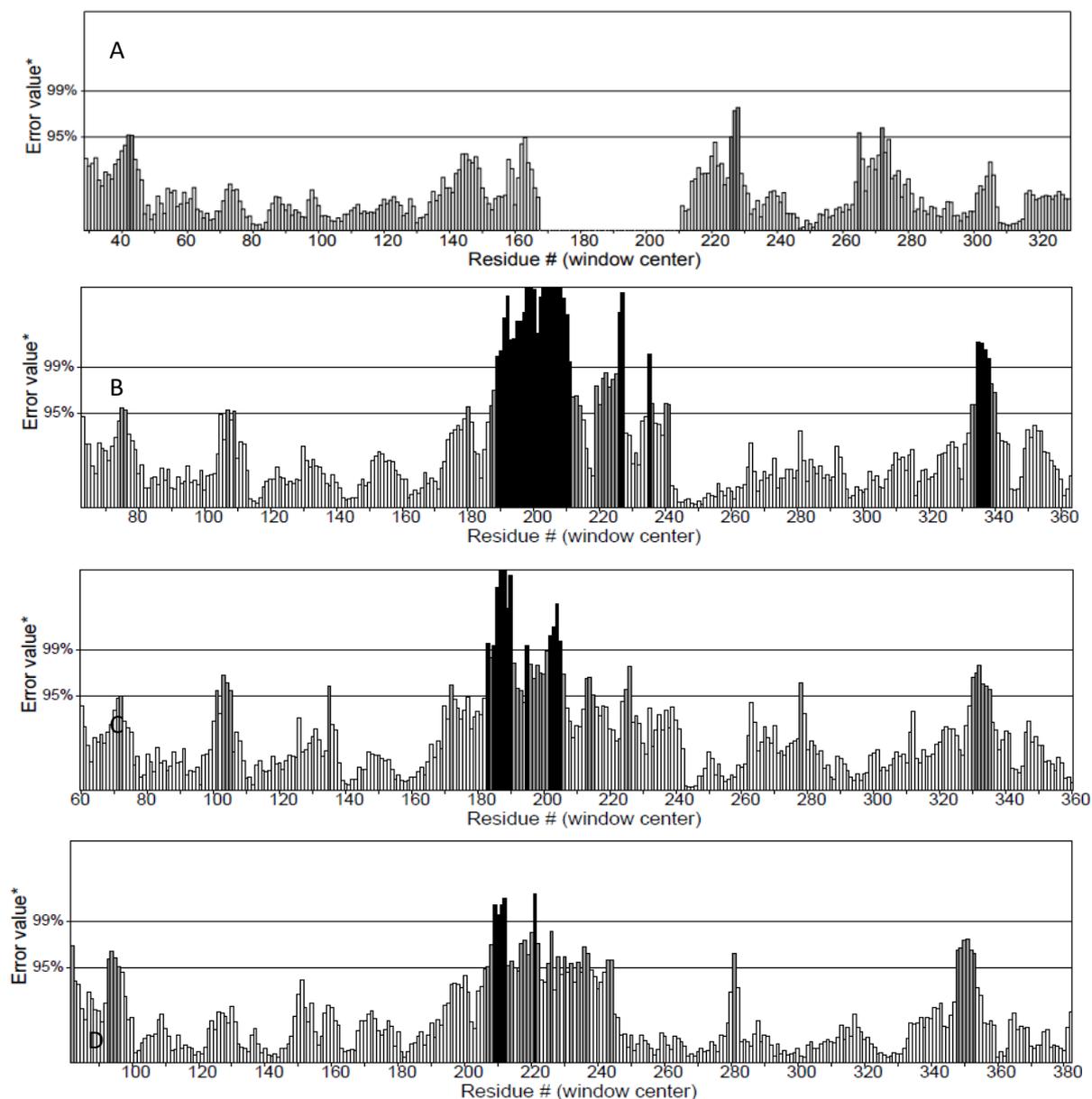
Figure 3. 6: ERRAT plots for 4M48 (A), DAT (B), NET (C) and SERT (D) based on the first 320-380 residues. ERRAT plots assess the distribution of different atom types. Values are plotted as a function of the position of a sliding 9-residue window. The 95% and 99% error lines on the graph show at what confidence level the residues can be rejected for not have the correct distribution between atom types. The gap shown in image A is due to a gap in the amino acid sequence which is a loop that had not been present in the crystal structure.

The results obtained from the protein validation analysis for all three models and the template 4M48 showed that the protein structures were of an appropriate quality to be used in docking studies. 4M48 outperforms all of the models in each of the tests, which is a reassuring indication that the models were not over fitted. Hence, the models could be used in experiments to ascertain what the structural differences between the isoforms were that could account for the differences in experimental affinity observed for the 21 compounds of the Iversen.

During the lifetime of this project, a number of experimental structures of the MAT isoforms were placed into the public domain, and as such the docking experiments were repeated using the experimental structures. Subsequent to these experiments taking place, the experimental structures were validated in the same manner as the homology models. The results of these validation studies on the experimental structures are summarised in Table 3.1.

Table 3. 1: A list of the protein models/ crystal structures and the associated protein validation results. Results are displayed as the percentage of amino acids that have appropriate scores/values for Ramachandran plot analysis, ERRAT and Verify3D protein structure validation analysis

Experimental structure/Model	RC (%) of amino acids in the favoured/allowed region of the RC plot	Verify (%) of residues with an average score of > 0.2	Errat (%) Overall Quality factor
4XP9 (DAT)	100	91.78	92.16
4XPA.1A (NET)	99.8	92.35	91.98
5I75 (hSERT)	100	99.53	87.67
4M48	100	94.42	93.51
DAT	100	89.67	88.36
NET	99.7	89.24	85.77
SERT	99.8	84.17	89.45

All experimental structures in Table 3.1 show improvements with respect to the quality of structure for ERRAT and Verify3D in comparison to the homology models used initially. The overall protein backbone quality is unchanged. Again, none of the residues that failed the validation tests were part of the binding site.

3.3.3 Docking studies

3.3.3.1 Putative binding site identification.

Using the Site Finder application in MOE putative binding cavities were elucidated for each of the homology models. These were validated by cross-referencing with literature to identify residues that were known to form protein-ligand interactions implicated in biological responses[24,192,193]. For each isoform, the putative binding site containing the greatest number of residues that had previously been reported as implicated in binding in the literature was selected from the list produced in MOE as the preferred binding site for docking studies.

The MAT isoforms are listed in Table 3.2 along with the volume of the preferred binding site and a Propensity for Ligand Binding (PLB) [194] score calculated by MOE. The largest cavities, with the highest PLB scores were selected to take forward into docking studies for DAT and SERT.

Table 3. 2: a list of the MAT homology models with associated binding site volume of the largest cavity identified and PLB value calculated using the MOE software.

	PLB	volume ($Å^3$)
Q01959 (DAT)	3.80	270
P23975 (NET)	0.53	87
P31645 (SERT)	3.50	249

The putative binding site cavity for NET was much smaller than DAT and SERT and NET had a PLB value which was low in comparison to the other two isoforms. The lower PLB score was necessarily a consequence of the smaller putative binding site for NET, as it follows that a constrained binding site, such as that identified for NET, would be less amenable to binding ligands than the more accessible cavities identified for DAT and SERT. It should be noted that the binding cavity elucidated by MOE for NET, was the second binding site available out of a list of 10 possible binding sites. This was chosen because this second cavity contained an aspartate (Asp75) residue that has been shown to be crucial for forming protein-ligand interactions in NET[176].

3.3.3.2 DAT

Table 3. 3: A list of all the DAT binding site residues that were reported in the literature[35,132] and binding site residues that were identified by MOE. The tables highlight any residues that failed the quality control tests during model validation, and which test(s) were failed.

DAT			
Residues reported in the literature[165,195]	Faults with residue	Type of fault	Residue present in putative binding site identified by MOE
D-79	no	-	Yes
L-80	no	-	No
A-81	yes	V3D (-0.25)	yes
V-152	yes	V3D (-0.09)	yes
Y-156	no	-	yes
F-320	no	-	yes
S-321	no	-	yes
F-326	no	-	yes
S-422	yes	Errat (95%)	yes
G-426	no	-	no

3.3.3.3 NET

Table 3. 4: A list of all the NET binding site residues that were reported in the literature and binding site residues that were identified by MOE. The tables highlight any residues that failed the quality control tests, and which test(s) were failed.

NET			
Residues in the literature[176,196]	Faults with residue	Type of fault	Residue present in MOE binding site
F-72	yes	Errat (95%)	yes
D-75	no	-	yes
A-145	no	-	yes
V-148	yes	V3D (-0.09)	yes
Y-152	no	-	yes
F-317	no	-	yes
F-323	no	-	yes

3.3.3.4 SERT

Table 3. 5: A list of all the SERT binding site residues that were reported in the literature and binding site residues that were identified by MOE. The tables highlight any residues that failed the quality control tests, and which test(s) they failed.

SERT			
Residues in the literature[197,198]	**Faults with residue**	**Type of fault**	**Residue present in MOE binding site**
Y-95	yes	Errat (95%)	yes
A-96	yes	Errat (95%)	Yes
D-98	no	-	Yes
L-99	no	-	No
G-100	no	-	Yes
W-103	no	-	Yes
R-104	no	-	Yes
Y107	no	-	Yes
I-172	yes	V3D (-0.54)	Yes
A-173	no	-	No
Y-175	no	-	Yes
Y-176	no	-	Yes
I-179	no	-	Yes
F-335	no	-	Yes
S-336	no	-	Yes
F-341	no	-	Yes
V343	no	-	No
K-399	no	-	No
D-400	no	-	No
S-438	yes	Errat (95%)	Yes
T-439	yes	Errat (95%)	Yes
G-442	no	-	Yes
E-493	no	-	Yes

3.3.4 Docking of native substrates

Dopamine, norepinephrine and serotonin were docked into all 3 MAT homology models to gain insight into where the molecular docking algorithm would place the native substrates, and how it would score them with respect to relative binding energies (S values) in each of the MAT isoforms. This provided information on how well MOE was able to dock endogenous ligands into the relevant MAT. From the literature there are a number of residues that have been established as playing a key role in forming protein ligand interactions within the three MAT. D-amphetamine, citalopram and paroxetine are

known form hydrogen bonds, to the Asp residues in DAT[187,199]. Docked conformations obtained from the MAT isoforms and the native substrates highlight the same well-established residues as being important in the formation of protein ligand interactions. These findings offer reassurance that the docking algorithm would be able to identify feasible docking conformations for the 21 NPS of the Iversen dataset and would allow conclusive results to be obtained.
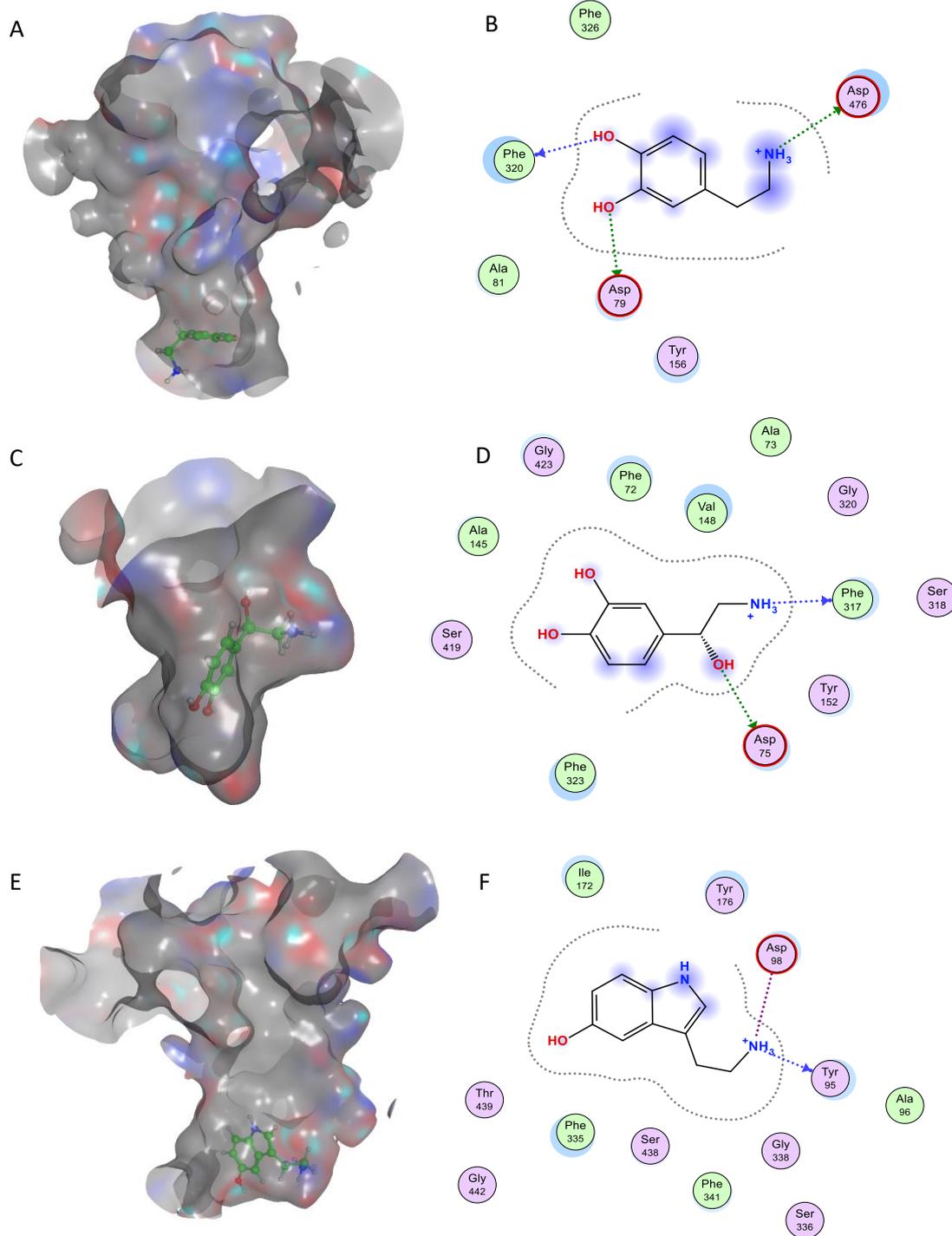
Figure 3. 7: Illustration of the MOE-defined putative binding cavity for A) DAT, C) NET and E) SERT showing the highest-ranked docked pose for A) dopamine, C) norepinephrine and E) serotonin (green ligand). Images B, D and F are the Protein Ligand Interaction Fingerprint (PLIF) for the respective docked pose. Potential hydrogen bonds being formed between ligand and protein are shown with dotted green arrows. B) is the docked dopamine/DAT pose with interactions between dopamine, Asp79[200], Phe320 and Asp476[200] with an S score of -4.6602 kcal/mol. D) is the docked norepinephrine/NET pose displaying interactions between the ligand and Asp75[176] and Phe317[176] and an S score of -4.9853 kcal/mol. F) is the docked serotonin/SERT pose with interaction between serotonin, Asp 98[198] and Tyr 95[198] and an S score of -5.3776 kcal/mol.

Figure 3.7 A and B shows that the dopamine molecule was bound at the bottom of the putative cavity. This docked conformation had the highest ranked s-value when the London dG and GBVI/WSA dG scoring functions were used and demonstrates interaction with residues that have been shown by experiment to be involved in protein ligand interactions in DAT[37,195]. This provided confidence that the docking study was able to generate credible, well-established interactions between DAT and its native ligand.

Figure 3.7 C and D shows norepinephrine docked in the lower part of the NET cavity, which was more constrained than both DAT and SERT. The docking study indicated that NE is forming the key protein-ligand interactions reported by Schlessinger *et al*, which highlights Asp75, Phe72, Tyr152 and Phe317[176], playing important roles in norepinephrine binding in NET. This was supported by the PLIF for NET and norepinephrine which suggest that the binding site elucidated by MOE was valid and could be carried forwards into docking studies.

Figure 3.7 E and F shows that the highest ranked docked pose for serotonin, much like dopamine's optimal docked conformation, is at the bottom of the cavity, and is shown to interact with residues that have been experimentally determined to be important in protein ligand binding[199]. This gives confidence, once more, that the model can be used in the docking of NPS and is likely to generate credible conformations for protein-ligand interactions which can help elucidate the differences in binding affinity between the MAT isoforms.

The values obtained from the docking study for the MATs and their native ligands are shown in table 3.6. This showed dopamine (DA) as having better S values (and by inference binding affinity) with NET and SERT than their native ligands (NE and SER respectively). Hence, the algorithm was unable to determine the preferential binding of dopamine with DAT (both SER and NE were predicted to bind more strongly) or norepinephrine with NET (SER was predicted to bind more strongly). This could be a consequence of limitations in the docking algorithm, the scoring function used, the putative binding

site elucidated by MOE, or a combination of all three of these factors. Alternatively, it may be evidence to support that the relative promiscuity observed for ligands which is seen between the three MAT isoforms is a consequence of the fact that binding sites are inherently non-selective for the majority of substrates. To test these hypotheses a series of further experiments, using relative rankings rather than absolute S scores, were conducted.

Table 3. 6: The S scores (kcal/mol) obtained for the highest ranked poses when the native substrates of each of the MATs DA, NE and SER were docked into the putative binding cavities of the DAT, NET and SERT homology models.

	Dopamine (DA)	Norepinephrine (NE)	Serotonin (SER)
DAT	-4.6682	-4.7047	-4.7057
NET	-4.7191	-4.9835	-5.4317
SERT	-5.0198	-4.9049	-5.3776

3.3.5 Docking studies using MOE

Table 3. 7: Table to show the relative rankings, where 1 indicates the molecule with the best predicted affinity, and 21 the molecule with the worst predicted affinity, for the best (i.e. most highly scored) docked poses of the 21 NPS of the Iversen dataset in the putative binding sites of the monoamine transporters DAT, NET and SERT. These are compared to the relative ranking of observed biological activities. Studies have been carried out using i) the Amber10: EHT force field and the London ΔG scoring function in MOE and ii) the MMFF94x force field and the default scoring function in MOE.

Ligand	Biological activity ranking (DAT/NET/SERT)	DAT		NET		SERT	
		Docking ranking Using AMBER10	Docking ranking Using MMFF94x	Docking ranking Using AMBER10	Docking ranking Using MMFF94x	Docking ranking Using AMBER10	Docking ranking Using MMFF94x
Desoxypipradrol	1/7/17	7	16	7	11	7	9
1-Naphyrone	2/12/4	12	20	10	6	14	3
Naphyrone	3/6/3	1	5	2	4	1	8
Nomifensine	4/1/13	8	17	8	10	13	2
Amphetamine	5/2/15	5	4	5	20	9	1
6-APB	6/3/12	18	1	17	7	18	19
Cocaine	7/18/6	3	14	3	18	3	17
5-APB	8/5/5	2	2	1	9	2	21
Mephedrone	9/11/14	13	3	14	5	16	4
Methylethcathinone	10/21/10	19	19	21	12	19	5
Methiopropamine	11/9/19	17	8	15	16	17	6
Methylenedioxy-N-benzylcathinone	12/20/21	10	7	12	13	8	7
(S/+)-MDMA	13/10/8	20	18	19	19	21	16
Benzedrone	14/17/16	9	21	9	8	4	20
5-iodo-2-aminoindane	15/8/7	16	11	13	1	10	12
(R/-)- MDMA	16/15/11	11	15	11	15	15	10
Fluoxetine	17/19/1	4	10	4	14	5	13
Methylenedioxy-aminotetralin	18/16/9	6	9	6	21	6	14
Amitriptyline	19/4/2	14	16	16	2	12	15
Methylenedioxy-aminoindane	20/13/20	15	20	18	17	11	18
Dimethylamylamine	21/14/18	21	5	20	3	20	11
r^2 value		0.146	0.006	0.017	0.078	0.082	0.033

A relative ranked score has been selected for comparing the data as there is no direct relationship between the scoring function S value of the docking algorithm and the biologically determined activity value (r^2 ranging from 0.006 to 0.146). The ligands that form the best interactions with the MAT will have better S values and can be used to indicate a ligand's ability to interact with a receptor and elicit a response.

Table 3.7 shows the ligands in ranked order of their biological activity for DAT NET and SERT in the second column. The remainder of the table shows the ligands numbered according to the relative ranking of their S values (or docked scores) generated by MOE in the docking experiments that were carried out.

The pairwise correlation between the relative rankings according to biological activity and docked score give very low r^2 values regardless of the force field employed (AMBER10 or MMFF94x). A possible reason as to why the correlation was so low may be due to the docking algorithm not being able to successfully distinguish between the ligand interactions with the MAT isoforms when forming docked poses, as a consequence of the similarity of the binding cavities of the MATs (Figure 3.8), and therefore not being able to effectively rank the selectivity of the ligands in the dataset.
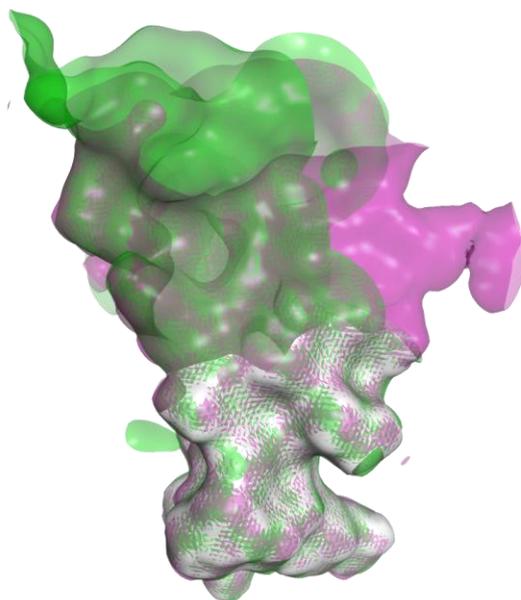
Figure 3.8: The overlaid putative binding sites of the DAT (green), NET (white) and SERT (magenta) homology models, elucidated using the SiteFinder module in MOE. This highlights

similarities in the shapes of the cavities, which could partially explain the observed promiscuity of binding between the three MAT isoforms and the ligands in the Iversen dataset.

As the binding site of NET was a lot smaller in volume in comparison to DAT and SERT (Figure 3.8), a second docking study was carried out in MOE. This study used a composite putative cavity that resulted from combining the top two binding sites elucidated by the software (see Figure 3.8). The modified binding site had a volume of 341 Å^3 and PLB score of 4.25. When the dataset was docked into this composite NET cavity an r^2 value of 0.034 was obtained. This value is comparable to the values achieved for the docking studies on the initial cavity, and give confidence that the poor performance for NET was not a consequence of the constrained nature of the putative binding cavity.
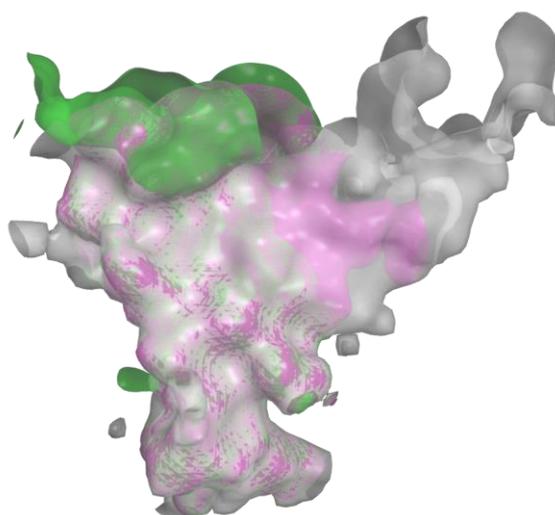
Figure 3.9: The overlaid putative binding sites of the DAT (green), composite NET (white) and SERT (magenta) homology models, identified using the SiteFinder module in MOE. This highlights similarities in the size and shapes of the cavities.

The composite binding site (Figure 3.9) of NET is much more similar to DAT and SERT overall, however the opening to the cavity is much broader with NET. Despite this modification to the NET putative binding site, there is no significant improvement in the docking results. This, again, speaks to the similarities between the binding sites of the isoforms making it difficult for the docking algorithm/scoring function to effectively distinguish between the ligands in the Iversen dataset.

3.3.6 Docking studies using Genetic Optimisation for Ligand Docking (GOLD)

The initial docking results obtained from MOE did not provide results that could help explain selectivity of ligands between the different MAT isoforms. A second docking study was carried out using the computational software GOLD to rule out this result arising as a limitation with the docking algorithm/scoring function used in MOE. GOLD uses a genetic algorithm to guide docking which uses the idea of evolution to develop conformational poses. The 21 compounds of the Iversen dataset were docked into the MAT models using GOLD (Table 3.8) and the docked data was ranked and compared to the biologically ranked data to see if there was a correlation (in the same way as for the MOE docking algorithm).

Table 3. 8: r^2 values for the correlation between the ranked biological data and ranked docking score values for DAT, NET and SERT using the scoring functions GOLDscore and CHEMscore

MAT	GOLDscore r^2	CHEMscore r^2
DAT	0.021	0.017
NET	0.017	0.000
SERT	0.152	0.000

Table 3.8 shows that very little correlation was obtained between ranked biological data and ranked docking data when using GOLD. This is consistent with the results obtained from the MOE docking studies and indicates that the lack of correlation between experimental and computational data was not likely to be as a result of limitations from a specific docking algorithm.

3.3.7 Spearman's Rank (ρ)

The docking studies carried out in MOE and GOLD produced very low r^2 values which indicates very little correlation between the experimentally derived data obtained from Iversen *et al* and the data produced from the docking algorithms in MOE and GOLD. Due to the complex nature of protein ligand interactions the ability for a single scoring function to correctly distinguish between these interactions is limited[154].

Spearman's rank (Table 3.9) was employed to see if there was correlation between the way the results had been ranked in terms of biologically activity and ranked docked poses (S-values). Many docking studies have employed Spearman's rank to evaluate the overall accuracy of

ranked data[201,] given the acknowledged lack of correlation between scoring functions and experimentally measured biological activity data.

To measure the significance of the ρ coefficient values and the strength of the correlation between results critical tables were used to highlight ρ values that were statistically significant (above 90%[184]). ρ values above 0.37 are significant at 90% and ρ values above 0.44 are significant at 95%[184] for a dataset containing 21 compounds.

Overall the best performing scoring function used in terms of ρ coefficient was the Affinity ΔG function. This produced the highest overall ρ value for DAT (significant at 90%). However only the docking studies in DAT generated statistically significant results for the Spearman ranking experiments, with additional ρ values for the London ΔG, αHB and GOLDscore achieved for DAT also statistically significant at 90%[202].

The scoring functions in MOE are the sum of different components based on different individual energy terms (Chapter 2). The scores of the two components from the London ΔG scoring function were ranked for the 21 docked compounds and ρ values calculated by comparing these values to their relatively ranked biological activities. This experiment was carried out to see if isolated properties of a scoring function could be used to better rank the docked posed obtained. The results obtained for the scoring functions in the GOLD software, ChemScore and GoldScore, were also analysed in this way (Table 3.9).

Table 3. 9: ρ coefficient values obtained for the docked ranking study conducted in MOE using four different scoring functions: London ΔG, Affinity, ASE and α HB (see Chapter 2 for details on scoring functions). ρ coefficient values calculated when only individual components (E_place and E_conf) of the London ΔG were used to rank the dataset. The table also includes docked ranking study conducted in GOLD using the two scoring functions GOLDscore and CHEMscore. Results that are statistically significant at 90% and above are shown in bold

ρ coefficient values

	London ΔG	Affinity ΔG	ASE	α HB	E_place	E_conf	GoldScore	ChemScore
DAT	**0.38**	**0.42**	0.34	**0.40**	0.22	0.25	**0.40**	0.28
NET	0.13	0.30	0.05	0.08	0.03	-0.02	0.04	0.19
SERT	0.30	0.28	0.13	0.28	-0.03	-0.12	0.19	0.34

None of the **ρ** values calculated using the individual components of the London ΔG scoring functions were significant at 90% for any of the isoforms. Table 3.9 also shows there was no overall improvement in **ρ** values when using a different docking algorithm (i.e. the GOLD algorithm).

In an effort to investigate whether inherent biases in scoring function were responsible for the lack of significant correlation, a rescoring exercise was carried out. Ballaxy[149] is a simple rescoring application which rescores docked poses generated in either the MOE or GOLD docking algorithms. The three scoring functions available via Ballaxy are MM- an AMBER[203] based scoring function, PB – a Poisson-Boltzmann scoring functions and PLP a pair wise linear potential scoring function[185,186,204].

Table 3. 10: The **ρ** coefficient values calculated when docked poses were taken from MOE and then rescored using the rescoring software Ballaxy[149]. Results that are statistically significant at 95 % are shown in bold

	ρ coefficient values		
	MM	PLP	PB
DAT	-0.22	**0.41**	0.27
NET	0.04	0.07	0.12
SERT	0.23	0.23	0.11

3.3.8 Consensus Scoring

In a final effort to try and improve correlation between computational and experimental results consensus scoring was employed (see Table 3.12). The aim here was to see if combining multiple scoring functions (consensus score[,205,206]) would improve how well the ranked docked data would best match the ranked biological data, and is based on the premise that implicit bias in a single scoring function is reduced by comparing and combining the results of several independently derived instruments. Scoring functions from MOE and GOLD were used in combination with 3 scoring functions from the software Ballaxy[148] (Table 3.11) to generate consensus rankings for each of the 21 ligands in the Iversen dataset.

Table 3. 11: A list of all the scoring functions used in the consensus of scoring functions.

MOE scoring functions Amber10:EHT[98]	Ballaxy Scoring functions[144]	GOLD[151]
Top scored S value (London ΔG)	MM	GoldScore
Average of all S values (London ΔG)	PLP	ChemScore
Average top 5 values (London ΔG)	PB	
Affinity DG		
ASE scoring function		
Alpha HB		

Table 3. 12: The **ρ** coefficient values calculated when a consensus of scoring functions was used. Statistically significant results (90% and above confidence) are shown in bold.

MAT	ρ coefficient for consensus scoring
DAT	**0.38**
NET	0.05
SERT	**0.40**

Table 3.12 shows the results obtained for the consensus scored ρ coefficients based on nine different scoring functions. Both DAT and SERT produced values were statistically significant at 90% [202] confidence.

As there is no standardized scoring function that can be applied to molecular docking studies, it was expected that the individual scoring functions did not produce strong correlation coefficients; this is why a number of different scoring functions were employed. To gain a better prediction of binding affinities, different docking and rescoring programs were used in attempt to combat the problems associated with protein flexibility and the number of different ligand conformations. Studies have shown that consensus scoring can be used to improve ranked data correlation however these studies were conducted using crystal structures[207]. Using homology models may have limited how well consensus scoring was able to rank the docked data. Another limitation of the study may be due to the small dataset available. Many consensus score studies

use between 100[208,209]-1000 ligands[205], as the correlation of ranked data was being analysed having a small dataset may not be suitable when looking at trends between ranked datasets.

3.3.9 Experimental structures of the MATs.

In 2016, the human SERT crystal structure[180] bound to paroxetine was published. This crystal structure showed paroxetine had the ability to form potential hydrogen-bonding interactions with both Tyr95 and Asp98, as predicted by the docking studies above and that Tyr 176, Ser 438, Ile 172 and Phe335 also defined the binding cavity – again consistent with the computer modelling. This gives further confidence that molecules which bind to DAT and SERT with a high affinity are likely to be found towards the bottom of the cavity, as predicted by the models.

The binding of ligands towards the bottom of the cavity is likely due to key interactions that are formed with the aspartate residue (DAT Asp 79, NET Asp75 and SERT Asp98). This highly conserved residue plays a key role in the recognition of biogenic amine DA, NE and SER[39,40,174,176,179,187,193,210–212]. It has also been reported as forming an interaction with virtually every known inhibitor of DAT, NET and SERT[39], and hence is likely to play a key role in the binding of the NPS to the MATs.

Also, after the initial computational modelling studies were conducted a series of dDAT crystal structures (May, 2015) co-crystallised with the ligands; dopamine, 3, 4-dichlorophenethylamine, D-amphetamine, methamphetamine, cocaine, β-CFT and RTI-55 were released[179].

In light of these advances, it was deemed pertinent to repeat the docking experiments carried out on the new crystal structures, to investigate any differences that arose between them and the initial studies carried out on the comparative models of the human isoform. The crystal structure 4XP9 complexed with D-amphetamine[187], was selected as a representative from the 4X series as this structure had the best resolution. The crystal structure was prepared for docking as described previously, and the Iversen dataset was docked into the crystal structure. Preliminary dockings studies indicated that these crystal structures improved the Spearman's correlation coefficient for DAT (0.38 to 0.42, significant at 90%) and SERT (0.30 to 0.48) showing that the results for SERT are now significant at a 95% confidence interval.

A newer homology model for NET (February, 2018) has also been released based on the dDAT crystal structure 4XPA[213]. Protein structure validation of this model indicates that it is of a high enough quality to be used in future docking studies, but this experiment was not conducted as

part of this study due to time constraints. The new homology model may provide results that have a better correlation with the biological data available.

3.3.10 5I75 human serotonin transporter crystal structure

In April 2016 the human serotonin transporter structure at 3.15 Å resolution was elucidated by X-ray crystallography by Coleman *et al*[180]. This structure was subjected to the same protein validation methods as the homology models. As expected, given this is a refined experimental structure, the crystal structure had little to no structural issues (see Table 3.1). However, when a docking study using the crystal structure was carried out, and compared to the results achieved with docking to the SERT homology model, there was no significant increase in correlation between relative rankings of computational and experimental data for the crystal structure compared to the homology model. Figure 3.11 shows a high level of conservation of the helices and a protein backbone RMSD value of 2.65 Å between the crystal structure and the homology model which is an acceptable value for the homology model to be used in docking studies[214]. However, this value is slightly higher than would be anticipated given the conserved nature of the secondary structural elements, and its magnitude is due to the variation in the orientation of the extracellular loops between the homology model and the crystal structure.
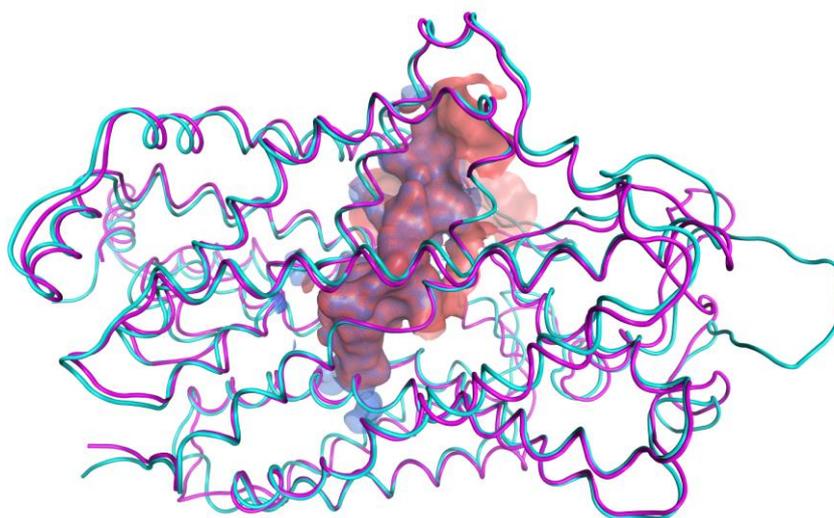
Figure 3. 10: The superimposed alignment of the SERT homology model (P31645, magenta) with the MOE putative binding site in red and the overlaid human SERT crystal structure (PDB Accession code: 5I75, cyan) with the binding site in blue. Backbone RMSD between model and crystal structure is 2.65 Å.

Figure 3.11 below shows the difference in the position of a docked molecule (fluoxetine) in the SERT homology model and the position of the co-crystallised molecule (citalopram) in the SERT crystal structure (PDB Accession Code 5I75). The crystal structure complex is shown in greater detail in Figure 3.12 and may provide additional information over the homology model/docking structures regarding where a known SERT selective[197,215] ligand is likely to bind.
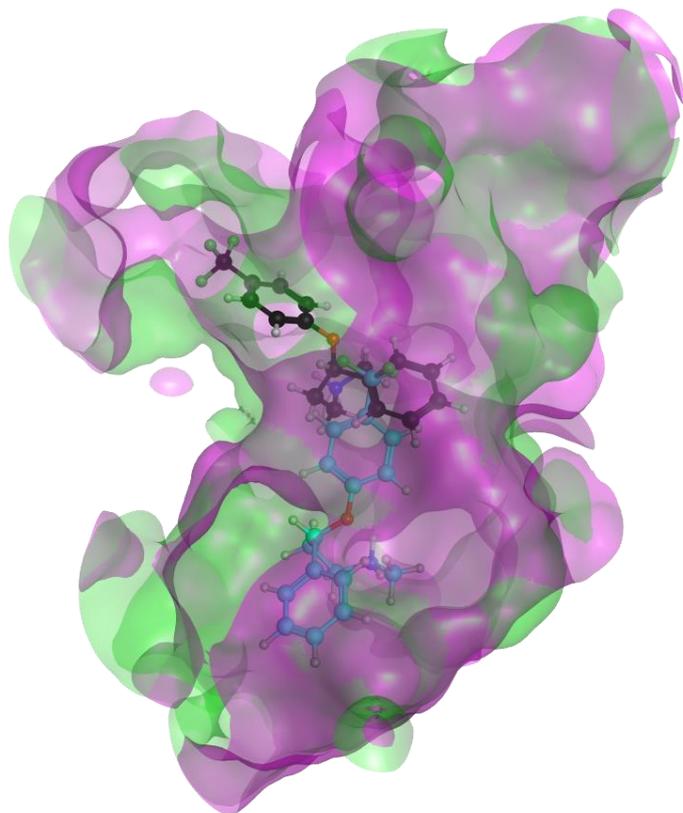
Figure 3. 11: Overlaid images of: the binding site of the SERT homology model (green) with the highest ranked docked conformation of fluoxetine (black ball and stick) and the binding site of the x-ray crystal structure of human SERT (5I75, magenta) with the docked conformation of fluoxetine (cyan ball and stick ligand) in 5I75. The difference in bound position of fluoxetine in comparison to citalopram may be a result of the docking algorithm used in MOE.
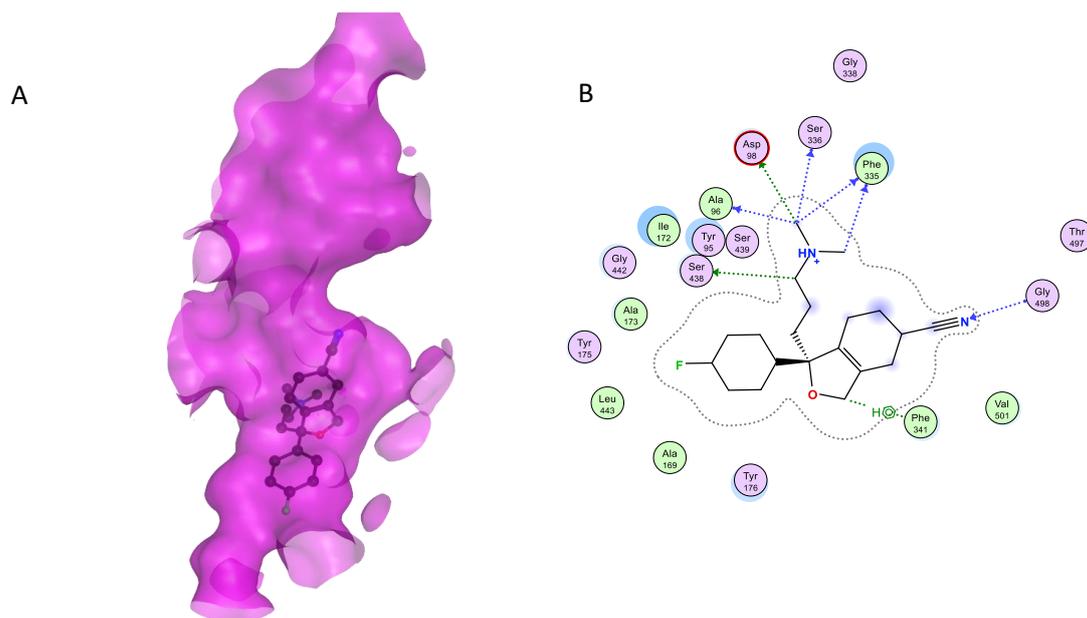
Figure 3. 12: **A**) is the binding site (magenta) of 5I75 with the co-crystallised ligand, citalopram (black ball and stick figure). **B**) PLIF image of the bound citalopram in 5I75. Protein-ligand interactions are illustrated with green dotted lines to show interactions between side chains of residues and blue dotted lines to show interactions between the ligand and the protein backbone.

In Figure 3.12 it can be seen that citalopram is bound at the bottom of the cavity. 5I75 has a narrower entrance to the lower section of the binding site (approximately 6 Å in diameter) in comparison to the SERT homology model (9.6 Å at the narrowest point and 13.6 Å at the widest point). The difference is shape of binding site could explain the increase in significant Spearman ranking results for the experiments carried out on the crystal structures when compared to the experiments carried out with the homology models. The initial SERT homology model docking experiments had a ρ value of 0.30 which is not significant at 90%. However, when the Iversen dataset was re-docked into 5I75, the crystal structure experiments obtained a ρ value of 0.48 which is significant at 95%. This may be due to the fact that the narrow binding site entrance in the crystal structure provided a steric "block" against docked conformations less representative of what is expected *in vivo* i.e. by restricting the molecules to the bottom of the binding site.

Figure 3.13: shows the overlaid binding site for 5I75 and the SERT homology model. With the exception of Asp98 there appears to be variation in position of side chain between model and crystal structure. The differences in residue orientation is quite likely the cause of an altered size and shape of binding site and ultimately account for the differences in docking data obtained.
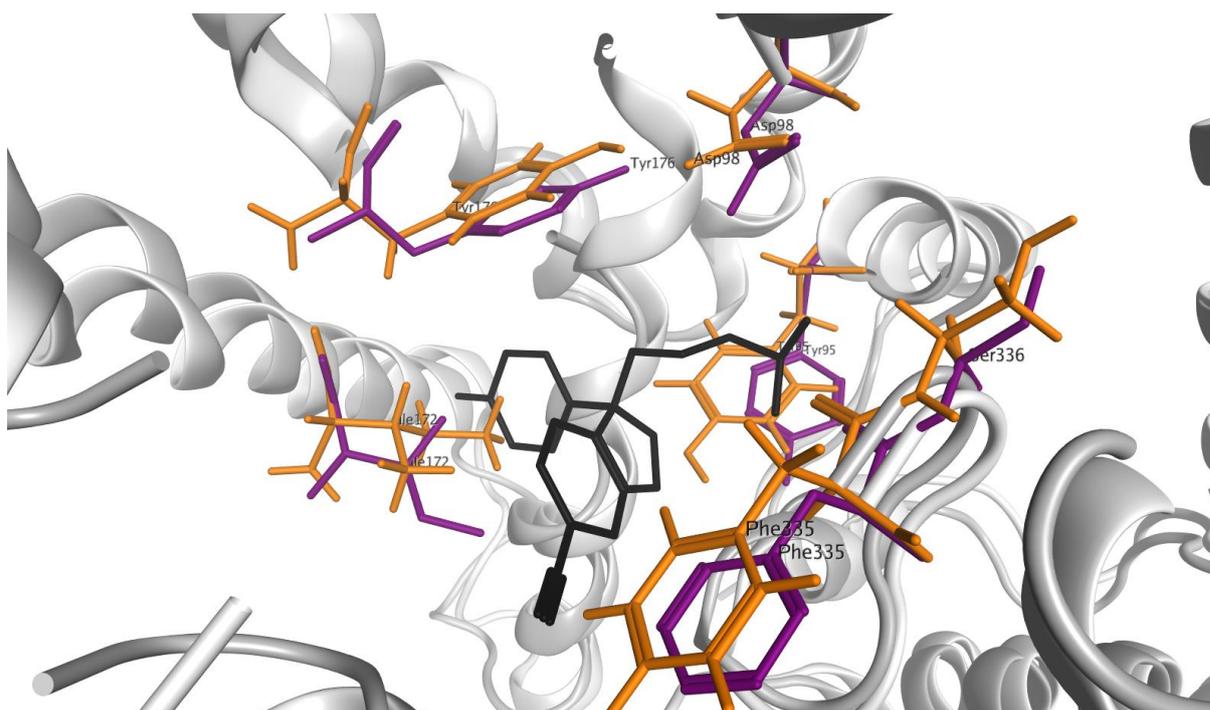
Figure 3. 13: the overlaid binding site for the SERT homology model (residues in orange) and the SERT crystal structure 5I75 (residues in purple), the residues highlighted include Asp98, Tyr98, Tyr175, Ser336, I172, Ala96, and Phe335 with the complexed ligand Citalopram indicated with the black stick model.

The binding site residues for the SERT homology model and the SERT 5I75 crystal structure are not aligned which indicates a shift in protein backbone, and a concomitant alteration in the shape of the binding cavity. This is illustrated in Figure 3.13. The carboxylate group on the Asp98 residue differs in orientation with a maximum distance between oxygen atoms between model and experimental structures of 1.45 Å and the β carbons of the Ile172 residues are separated by a distance of 1.12 Å. The phenol ring of the Tyr95 is also not aligned between model and crystal structure with a maximum distance of 2.11 Å between the atoms of this residue. These conformational differences between binding site residues could explain why different results were obtained from the docking study using the crystal structure and the homology model.

3.3.11 Investigating the Emergent DAT Crystal Structures

Using one of the newer dDAT crystal structures (the dDAT 4XP9 crystal structure at 2.8 Å bound to D-amphetamine[179] published in May 2015) a further "re-docking" study was carried out in attempt to improve ρ values for DAT, with the intention of gaining further insight into what differences there are between the MATs which could explain selectivity.,

When this crystal structure was used to dock the ligands from the Iversen paper the ρ values did improve (from 0.38 to 0.42), but this is not a significant difference to the results obtained previously. This is not surprising given that the superimposition between 4XP9 and the 4M48-derived homology models (Figure 3.14) shows that the experimental structure and the homology model are very similar.
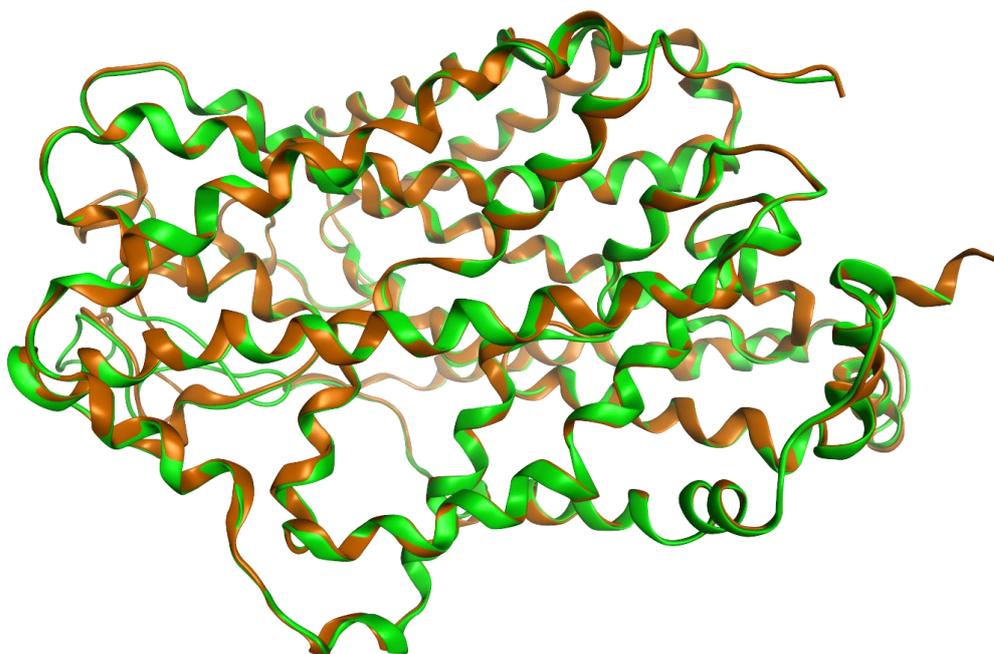
Figure 3. 14: the superimposed alignment of the DAT homology model (based on the template 4M48, green) and crystal structure 4XP9 (bronze). A protein backbone RMSD value of 0.729Å was calculated in MOE.

The docked D-amphetamine ligand (Figure 3.15) has a conformation very similar to the complexed D-amphetamine ligand that was present in the structure 4XP9. The similarities between the experimental ligand position and the docked conformation gives confidence that the docking algorithm used in MOE is able to reproduce very likely protein ligand interactions, and that the homology model is an appropriate surrogate for the dDAT crystal structures in the docking experiments.
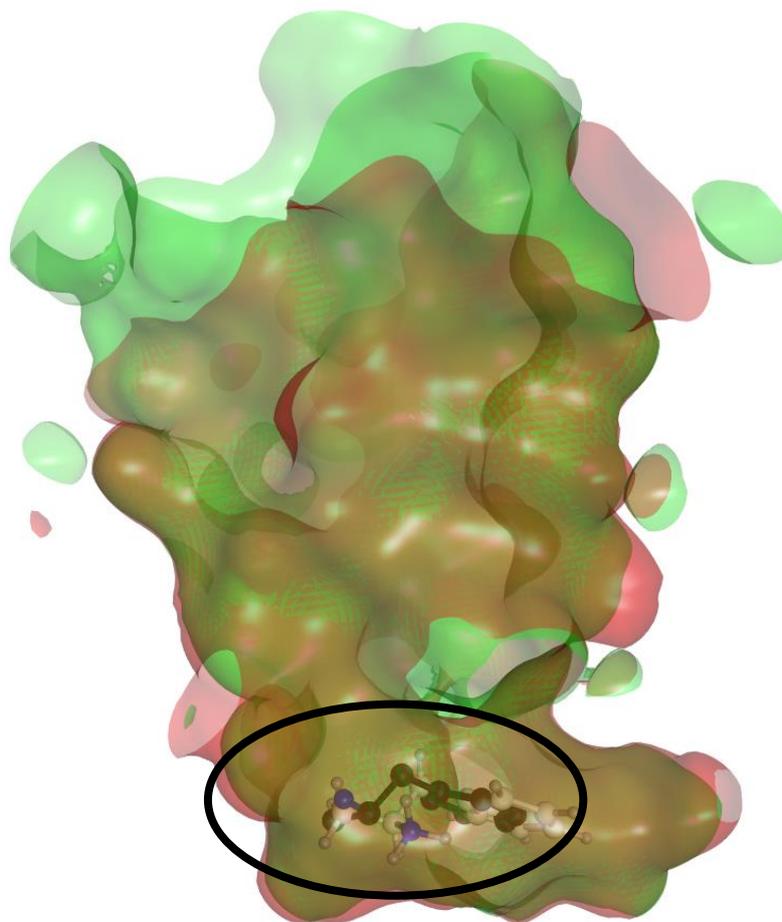
Figure 3.15: The overlaid binding sites of the DAT homology model (green) and the dDAT crystal structure 4XP9 (red). The black ball and stick figure is D-amphetamine docked into the homology model and the white ball and stick figure circled in black is the complexed D-amphetamine found in 4XP9.

3.3.12 Investigating the Emergent NET models

A preliminary docking study was carried out using the newly published NET homology model which is based on the template 4XP4[179] a dDAT transporter in complex with cocaine. The putative binding site of the new homology model was larger in volume (174 $Å^3$) in comparison to the original homology model (which had a binding site volume of 87 $Å^3$ see Figure 3.16). Using the 4XP4A-based NET model a ρ value of 0.21 was obtained for the docking of the 21 ligands. Similar to the previous studies, a composite binding site was created for NET which combined the two largest cavities identified by MOE was carried out. The composite binding site had a site volume of 341 $Å^3$, and the docking study carried out using this active site returned a ρ value of 0.24. This is only slighter higher than the smaller 4XP4A-NET binding site and not significant at

90% confidence. When analysing the docked conformations of the ligands in the composite binding site it was noted that docked ligands were consistently placed in the bottom of the cavity, an area that was also available in the first iteration of the experiment. As such, the small difference between the two experiments is not unexpected.
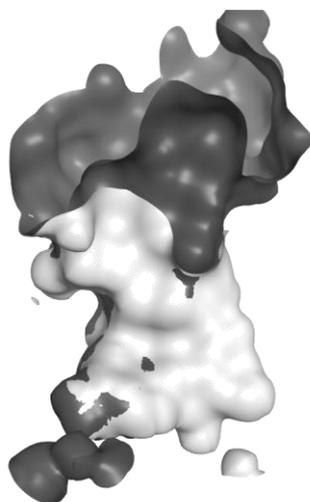
Figure 3. 16: the overlaid putative binding sites elucidated in MOE, the original homology models binding site is shown in white and the NET homology model based on 4XP4A binding site is shown in dark grey.

3.4 Discussion

3.4.1 Homology models

The homology models available were able to produce valid docked conformational poses for the ligands based on the literature available regarding the native ligands (DA, NA and SER). However, as the models were very similar in topology and sequence this may have limited how structurally different the binding sites are, and hence how able docking experiments would be to discriminate between NPS binding. Sequence identify and similarity amongst all three models at the binding site were analysed, (Figure 3.17). The lack of variation in amino acid composition between the binding sites is evident, and this may have reduced the ability to identify selectivity amongst the MAT via docking studies.

Figure 3. 17: Alignment of amino acid sequences for the homology models DAT, NET and SERT. The highly conserved aspartate residue is highlighted by the black box. The residues highlighted in green show that all three models have identical residues at this position between the aligned sequences and the areas highlighted in red show that one or more of the residues are not identical.

The residue alignment in Figure 3.17 show that there is a high level of sequence identity amongst the residues that are in close proximity to the conserved aspartate residues and therefore the binding site compositions are similar. Analysis of the binding site residues show that although they are not identical many of the residues are similar e.g. the three closest highlighted red columns to the conserved aspartate (in Figure 3.17) show a mixture of similar hydrophobic residues (G, A, Y, F, I and V) proving binding site similarity. Figure 3.18 shows the binding site residues that are conserved amongst all three MAT and are implicated in the formation of protein ligand interactions. There appears to be no conformational differences of the side chains, highlighting the structural similarity of binding sites for DAT, NET and SERT.
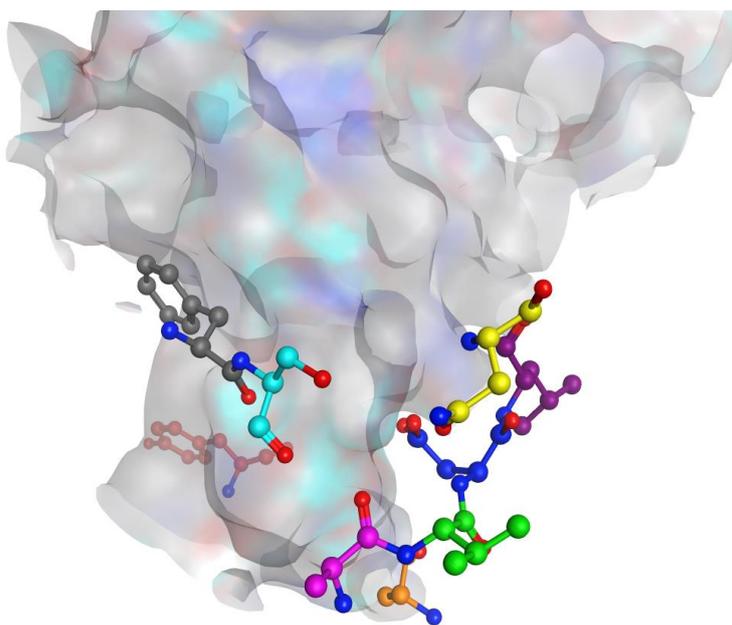
Figure 3. 18: The overlaid binding site for DAT, NET and SERT, the conserved aspartate (dark blue), phenylalanine (grey), tyrosine (red), alanine (magenta), serine (cyan and orange), valine (green), leucine (purple) and asparagine (yellow) residues have been highlighted to show there are little differences in side chain orientation of the residues known to be important for protein-ligand interactions in the MAT homology models.

3.4.2 Docking

All of the models produced docked conformational poses that formed protein ligand interactions with residues that were known to be important, and were later confirmed as such by the publication of x-ray crystal structures. For the majority of results obtained for DAT, it is evident that the residue Asp79 forms an important role in protein-ligand interactions. This highly conserved aspartate in DAT, NET and SERT forms an interaction between the protonated amine of the biogenic amines. From a model alignment it is clear that the three aspartate residues are equivalent see figure 3.19.
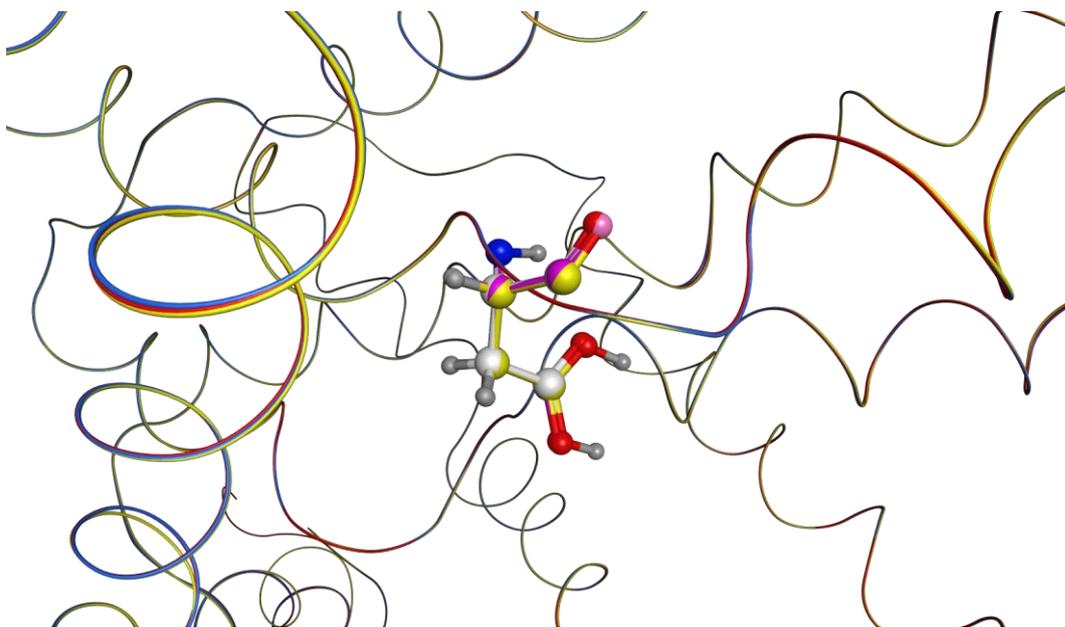
Figure 3. 19: shows the aspartate residue Asp79 in the homology models; DAT (white), Asp75 in NET (magenta) and Asp98 in SERT (yellow) overlaid. This image shows that this particular residue is highly conserved in all three MAT. The three MAT models have been aligned and superimposed, with DAT being coloured yellow, NET being coloured red and SERT being coloured red.

The results obtained from the docking study in both MOE and GOLD showed that both computational algorithms were able to dock the ligands and form protein ligand interactions with all of the residues highlighted in Tables 3.3-3.5. This indicates that the models used were of a high enough standard this gives confidence that they can be used to provide valid results.

The models used were all structurally valid as demonstrated by the protein validation methods, all of the binding site amino acid residues were structurally sound and did not violate any of the

validation tests. Newer published crystal structures show an improvement in structural quality and as future work these structures should be used for carrying out further docking studies.

3.5 Conclusion

The aim of this study was to see if the homology models available for DAT, NET and SERT were of a suitable structural quality, which would allow for molecular docking studies to be conducted in order to provide information that corresponded to the literature.

The homology models were found to be structurally acceptable based on the data obtained from the protein structure validations tests carried out. The binding site application in MOE was able to identify valid binding site cavities that corresponded to literature findings based on studies conducted on MAT, using both homology models and experimental data. Residues that had been reported in the literature as playing an integral role in the recognition of MAT inhibitors were found to be in the binding sites the software had elucidated. Using the homology models, differences regarding the binding sites for the MAT were observed. The selected binding site for NET had a much smaller binding cavity in comparison to DAT and SERT, therefore to see if an increase in cavity size would improve correlation a composite binding site was created. The composite binding site for NET was created by combining the initial binding site with a second site that had been elucidated in MOE. The second site sat directly on top of the initial site however when the volume of NETs binding site was increased it did not significantly improve correlation between the ranked datasets (ρ value increased from $0.21 - 0.24$).

There doesn't appear to be an obvious explanation as to why the NET homology model and ranked docked data achieved such low ρ values despite a number of measures taken to try and optimise the binding site and docking parameters. More information regarding the NET binding site and the mechanism in which a protein ligand interaction is formed is required.

Table 3.14 shows that the consensus of scoring functions improved the ρ values for DAT and SERT but there was no improvement with the NET ρ values. This indicates that there are issues with the NET homology model being able to produce ranked docking conformations that are in agreement with experimentally derived data. Inspection of the biological data will be conducted in the next chapter to see if there are any associated issues that could cause erroneous results.

The overall amino acid sequence for all the MAT homology models was over 50% identical, this indicate that that the models may be too similar to one another to and therefore information

regarding selectivity cannot be extrapolated. When focussing on the binding sites of each of the MAT homology models there is a percentage identity of 75%, this reinforces the idea that the high similarity of the models limits how well selective docking data can be obtained.

This study was a structure-based approach to start the investigation into identifying a novel NPS. The homology models used for DAT and SERT were able to produce statistically significant correlation between experimentally derived data and the docking study results, but did not provide insight into the reasons for the differences in affinity for molecules between the isoforms. The DAT, NET and SERT homology models were found to be of a high enough quality to be used for docking studies, however the NET homology model did not produce docked data that correlated significantly to the literature values. Due to limited information being obtained from the MAT the next step will be to employ a ligand-based approach to investigate selectivity between the isoforms. Identifying physicochemical properties for chemical structures that are selective for one MAT over another may provide more information to help the design of a novel NPS.

Chapter 4

Development of Quantitative Structure Activity Relationship Models to Identify Key Physicochemical Properties Required for Selectivity between DAT, NET and SERT

4.1 Introduction

This Chapter focuses on a ligand-based approach to identify key differences between compounds that will convey selectivity for one monoamine transporters (MAT) over another. At the time of the study, there were no experimental structures available for any of the human MAT isoforms, so employing a ligand-based approach removed the need for high resolution crystal structures. From Chapter 3, it was highlighted that limited information regarding selectivity was gathered from a structure-based approach, so using a ligand-based approach has been employed as a complimentary technique.

Ligand-based approaches can be used to analyse a database of compounds, identify key physicochemical properties and generate a predictive model of potential biological activity[216,217]. In particular, Quantitative Structure Activity Relationship (QSAR) modelling is a method widely employed in research and industry to achieve this[117]. This approach relates chemical structure, and calculated physicochemical properties to experimentally observed biological activity using mathematical regression models which can then be applied to a virtual library of compounds in an effort to identify novel compounds with potential to bind to the receptor of interest[117].

As previously discussed in Chapter 3, there is a large degree of promiscuity with respect to new psychoactive substances (NPS) and their binding to the human MAT isoforms. Compounds that bind to the MATs will, in general, preferentially bind to one of the isoforms (either DAT, NET or SERT), but will also demonstrate affinities with the other two MAT isoforms. This chapter will focus on a ligand-based approach to building QSAR models to identify which physicochemical properties, if any, a molecule requires for it to preferentially bind to one MAT isoform over another.

4.2 Methods

4.2.1 Data set preparation

Two independent datasets[35,218] were identified from the literature and combined to construct the dataset for the following studies[35,181]. In total, 31 compounds (Table 4.1) representing a range of different classes of NPS, and other psychoactive compounds, with experimental biological activity measurements for each of the monoamine transporters DAT, NET and SERT were used in the development of the QSAR models.

The compounds in the dataset were "washed" using the wash compound application in Molecular Operating Environment (MOE) at physiological pH (pH 7.0) in order to identify likely physiological protonation. The classification system presented in Table 4.1 was adapted from the EMCDDA classification[219], and known narcotics in the dataset that fell outside of the EMCDDA classification were classed as "others".

Table 4. 1: The name, experimentally determined pK_i values[218] for each MAT isoform, chemical structure and classification of the psychoactive compounds used in the *in Silico* studies on DAT, NET and SERT.

| Compound | pK$_i$ | | | Structure | Classification |
	DAT	NET	SERT		
1. Mephedrone	6.08	5.88	4.99		Cathinone
2. Methiopropamine	6.05	6.09	4.14		Phenethylamine
3. Methylenedioxy-N-benzylcathinone (MNB-cathinone)	6.01	5.37	4.56		Cathinone
4. 5-APB (1-(benzofuran-5-yl)propan-2-amine)	6.30	6.33	5.78		Phenethylamine
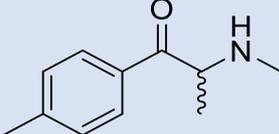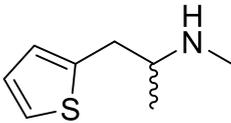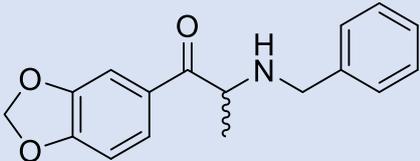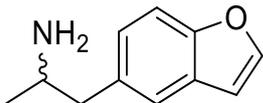
Compound	pK$_i$			Structure	Classification
	DAT	NET	SERT		
5. 6-APB (1-(benzofuran-6-yl)propan-2-amine)	6.63	6.52	5.26		Phenethylamine
6. Desoxypipradrol	7.30	6.26	4.27		Piperidines & Pyrrolidines
7. 5-Iodo-2-aminoindane (5-IAI)	5.61	6.09	5.75		Aminoindane
8. Benzedrone	5.64	5.50	4.75		Cathinone
9. Dimethylamylamine	4.74	5.77	3.75		Others
10. Methylenedioxy-Aminoindane (MDAI)	5.12	5.78	4.93		Aminoindane
11. Methylenedioxy-Aminotetralin (MDAT)	5.20	5.64	5.65		Phenethylamine
12. Naphyrone	7.28	6.70	6.63		Cathinone
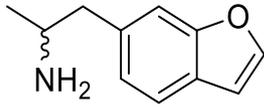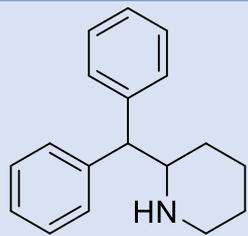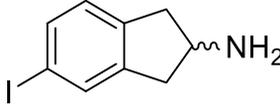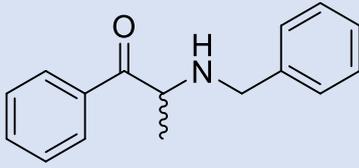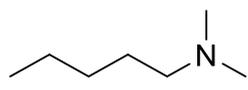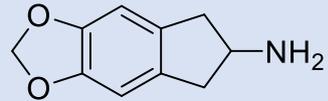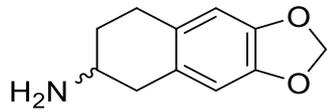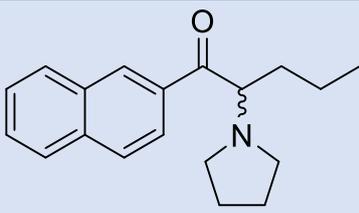
Compound	pK$_i$			Structure	Classification
	DAT	NET	SERT		
13. 1-Naphyrone	**7.32**	**6.27**	**6.45**		Cathinone
14. Methylethcathinone	**6.08**	**5.36**	**5.43**		Cathinone
15. Amitriptyline	**4.61**	**7.00**	**7.83**		Others
16. Nomifensine	**6.90**	**7.07**	**5.56**		Others
17. Cocaine	**6.19**	**5.20**	**5.62**		Others
18. *R*-MDMA	**5.21**	**5.50**	**5.17**		Phenethylamine
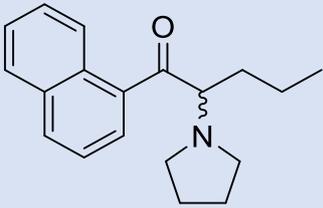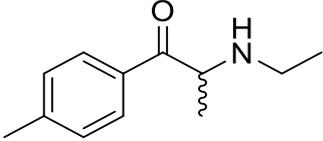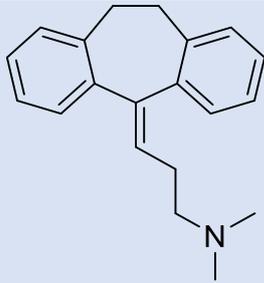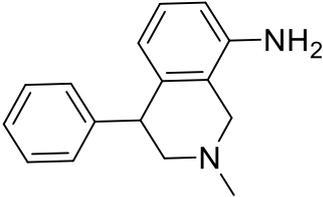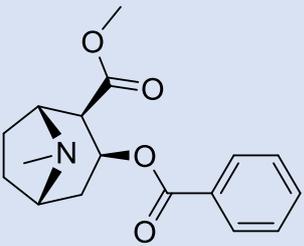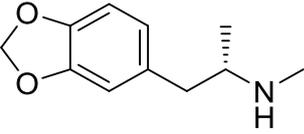
Compound	pK_i			Structure	Classification
	DAT	NET	SERT		
19. *S*-MDMA	5.62	5.70	5.56		Phenethylamine
20. *S*-Amphetamine	6.54	6.20	4.78		Phenethylamine
21. Fluoxetine	5.04	5.20	6.55		Others
22. RTI-55	8.49	8.60	9.31		Others
23. WIN 35428	7.58	7.50	6.90		Others
24. GBR 12935	7.67	6.65	5.19		Piperazine
25. Bupropion	5.56	5.86	4.35		Cathinone
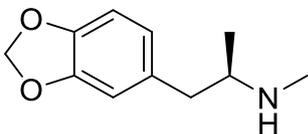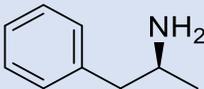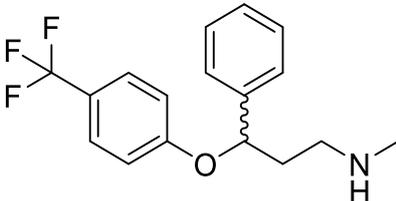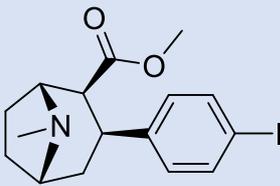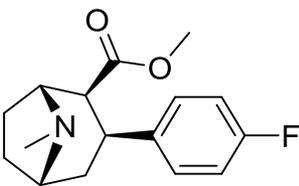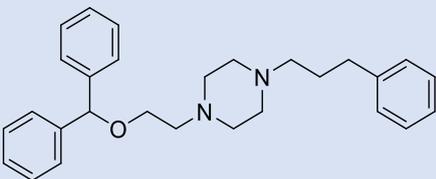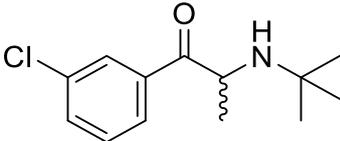
Compound	pK$_i$			Structure	Classification
	DAT	NET	SERT		
26. Nisoxetine	6.32	8.29	6.42		Others
27. Desipramine	4.10	8.40	7.21		Others
28. Nortriptyline	4.86	8.47	6.79		Others
29. Mazindol	7.56	8.49	6.82		Others
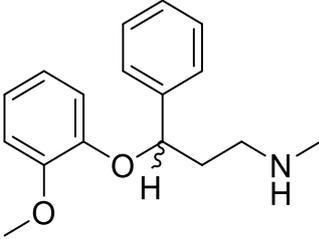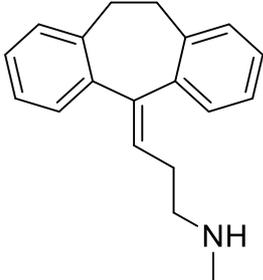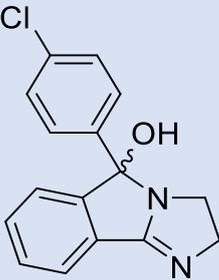
Compound	pK$_i$			Structure	Classification
	DAT	NET	SERT		
30. Imipramine	**5.01**	**7.17**	**8.11**		Others
31. Citalopram	**5.00**	**6.00**	**8.27**		Others
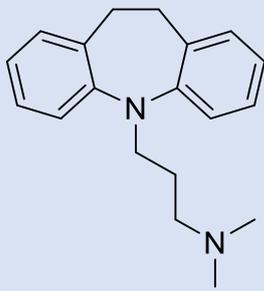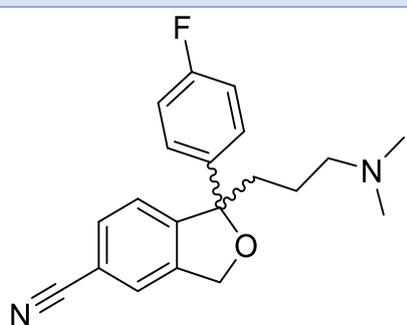
4.2.2 Identification of a diverse subset

The NPS dataset (31 compounds) was uploaded into the MOE[98] software. Three training sets were identified, one for each of the MAT isoforms (DAT, NET and SERT), using FP:MACCS structural keys as the diversity metric.

Each dataset was divided into a diverse training set (80% of the dataset – 25 compounds) and a test set (20% of the dataset - 6 compounds), using the diverse subset application in MOE. The training sets for each MAT isoform were then used to build and validate QSAR models. As the pKi values of individual compounds varied between DAT, NET and SERT this resulted in the training and test sets for each of the isoforms being comprised of different compounds.

A second method for identification of diverse subsets to create training and test sets for each of the MAT isoforms was also used for comparative purposes. This time, the training sets were compiled by looking at Tanimoto coefficients (T$_c$)[220] across the dataset as a whole. This approach

was used to ensure that the training and test sets each contained compounds that represented the whole range of structural diversity of the dataset, as demonstrated by their T_c values.

To this end, a similarity coefficient matrix based on structural similarity of the entire dataset, was produced using the open access software OpenBabel[221] which calculates pairwise T_c between compounds. The average T_c for an individual molecule was then calculated based on the pairwise values obtained for all compounds in the dataset. Any compounds which were shown to have an average Tc of less than 0.2 were deemed significantly different from the remainder of the dataset, and were removed from the study.

Having calculated average Tc values for each of the compounds, the molecules in the dataset were then placed into 1 log unit "bins" according to their experimental pKi value range (i.e. 4-5, 5-6, 6-7 etc.). For each bin, the molecules that had the highest and lowest pKi values in the bin were placed in the training set. The remaining compounds in the bin were then analysed and selected for either the test or training set based on their Tc scores i.e. molecules that had the highest average T_c value, and hence were most like other molecules in the dataset, were placed in the test set. It is important that the training and test sets both represent the dataset as a whole in order to be able to generate and appropriately evaluate QSAR models[222].

As before, 80% of the compounds were placed in the training set, and 20% of the compounds were placed in the test set. The distribution of the T_c and pK_i values were examined to ensure that the training and test sets were selected to mirror the dataset as a whole. The training and test sets that resulted for each of the MAT isoforms were evaluated for normality using the Shapiro-Wilks test[223].

4.2.3 Descriptor Selection for QSAR Models

A molecular database (MDB) for each of the MAT isoforms was curated in MOE. This MDB file contained only the training set compounds identified for a specific MAT isoform and the biological data pertaining to that MAT isoform. A total of 435 molecular descriptors (see Appendix A) were calculated for each molecule using the descriptor calculator application in MOE. The descriptor values were scaled relative to their maximum reported values to obtain values ranging between -1 and 1 for all descriptors and were then correlated to the pKi values of the compounds i.e. a correlation matrix was produced.

Correlation coefficients between pK_i and descriptor values that were above 0.7 (positively correlated) or below -0.7 (negatively correlated) were identified. The molecular descriptors were ranked (in descending order) in terms of absolute r^2 values. The top 20% of the descriptors were then used as the initial set of descriptors in the construction of the QSAR models. Using the correlation matrix, these descriptors were systematically checked for cross-correlation against other descriptors in the list. If one descriptor correlated to another descriptor with a correlation coefficient above 0.7 or below -0.7 then the descriptor shown to correlate least with biological activity was removed from the list. This was to avoid over-representation of any single type of physicochemical property in the final QSAR models.

4.2.4 Building and evaluating QSAR models

QSAR models were built for each MAT isoform using both the estimated linear model (ELM) protocol and the Estimated Normalized Linear Model (ENLM) protocols of the QuaSAR module in MOE. A correlation coefficient (r^2 value) and a cross-validated correlation coefficient (q^2 value) were generated and used as measures of the initial model quality. From these initial models, the molecular descriptor shown to be contributing least to explaining the variance in the data was removed and a new QSAR model was generated using the remaining descriptors. This descriptor removal process was carried out iteratively until the r^2 and q^2 values were in close proximity to one another, and the model had the highest r^2 values possible using the least number of molecular descriptors. The resultant best model for each MAT isoform was then used to predict the biological activity of the compounds in the corresponding test set. The correlation between the biological activity and predicted activity of the test set compounds (r^2) was used to evaluate the predictive ability of each of the models. The extreme studentized deviate test[224] was employed for outlier detection, this was applied to the test set predicted pKi values obtained for the optimised DAT, NET and SERT QSAR models.

4.3 Results and Discussion

4.3.1 Identification of the diverse subset

Figure 4.1 shows the distribution of molecules in the FP:MACCS and Tanimoto trainings sets, with reference to the experimentally derived pKi, and the distribution of the activity of the molecules across the dataset as a whole. When comparing the distributions, it can be seen that the Tanimoto training set is a better representation of the distribution of molecules across the

dataset as a whole than the FP:MACCS-derived training set. Specifically, this is because the distribution of molecules in the most heavily populated area of the dataset (between pKi 5 – 7) for the Tanimoto-derived training set mirrors the general trend across the dataset as a whole, which is not the case for the FPMACCS-derived training set, which places a disproportionately greater number of compounds from the pKi 5-6 range into the training set, when compared with the dataset as a whole.
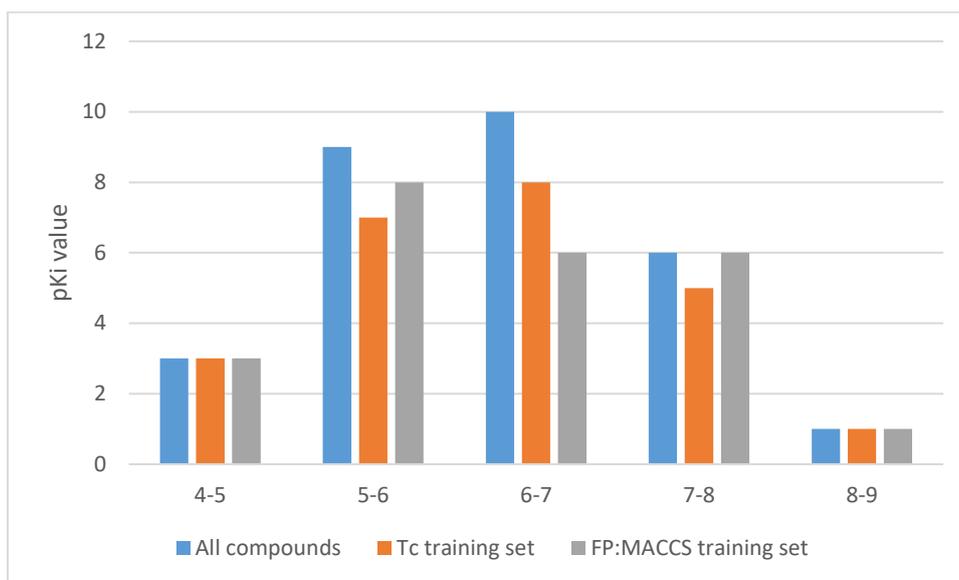
Figure 4. 1: The distribution of all compounds classified by pKi value for all compounds in the dataset (blue), the compounds in the Tanimoto training set (orange) and the compounds in the FP:MACCS training set (grey).

Both training sets identified (FP:MACCS and Tc) were used to construct QSPR models for the DAT isoform. As the training sets were different, the test sets used to assess the predictive ability of the models produced were also different. The composition of the two different test sets is given in Table 4.2.

Table 4. 2: **A**, The FP:MACCS DAT test set compounds, associated pKi values and average pairwise T$_c$. **B**, the Tanimoto DAT test set compounds, associated pKi values and average pairwise T$_c$.

A **FP:MACCS** derived test set	pKi value	Average pairwise T$_c$	**B** **Tanimoto** derived test set	pKi value	Average pairwise T$_c$
Mazindol	7.56	0.21	Mazindol	7.56	0.21
Mephedrone	6.08	0.42	Mephedrone	6.08	0.42
MDAI	5.12	0.35	MDAI	5.12	0.35
Nortriptyline	4.87	0.33	Nortriptyline	4.87	0.33
Dimethylamylamine	4.74	0.17	6-APB	6.54	0.31
Cocaine	6.19	0.27	1-Naphyrone	7.28	0.42

Given the overlap between the test set molecules, it was expected at this stage that the methodologies for generating the QSAR equations would be broadly equivalent, and as such it would be expected that models of similar robustness and predictive ability would be produced, regardless of the training and test set used to build and validated them.

4.3.1.2 Initial QSAR Model for DAT – FPMACCS Derived training set

Initial QSAR models (see Equations 4.1) for DAT were built and validated using the FP:MACCS training/test set and their predictivity and robustness interrogated (Table 4.2).

5-descriptor model for DAT (FP:MACCS training set)

pKi / SD (pKi) = 11.4844 -0.36724 * E_sol / SD(E_sol) -0.33315 * npr2 / SD (npr2) +0.19984 * PEOE_VSA-0 / SD(PEOE_VSA-0) -0.19889 * SlogP_VSA5 / SD(SlogP_VSA5) -0.30776 * vsurf_HB1 / SD (vsurf_HB1)

4-descriptor model for DAT (FP:MACCS training set)

pKi / SD (pKi) = 11.6867 -0.34777 * E_sol / SD(E_sol)-0.32113 * npr2 / SD (npr2)-0.41605 * SlogP_VSA5 / SD(SlogP_VSA5)-0.21456 * vsurf_HB1 / SD (vsurf_HB1)

3-descriptor model for DAT (FP:MACCS training set)

pKi / SD (pKi) = 11.2047 -0.44253 * E_sol / SD(E_sol)-0.35306 * npr2 / SD (npr2)-0.39015 * SlogP_VSA5 / SD(SlogP_VSA5)

Equations 4.1: 5, 4 and 3-descriptor QSAR estimated normalized linear models built using the DAT training set derived using the FP:MACCS structural keys in conjunction with the MOE diverse subset tool.

To evaluate how well the models predicted the activity of unseen compounds the MACCS-derived DAT QSAR models were applied to the test set (as shown in Table 4.2). The correlation between predicted pKi and the experimental pKi was obtained, and used as a measure of the predictivity of the models. The robustness of the models was assessed by comparing the r^2 and q^2 values obtained from training the model. The results of this investigation are given in Table 4.3.

Table 4. 3: Summary of the r^2 and q^2 values for the three best-performing different estimated normalized linear QSAR models (Equations 4.1) developed for DAT using a training set identified via the diverse subset application in MOE.

Model	FP:MACCS training set		FP:MACCS test set
	r^2	q^2	r^2
5 - descriptor	0.73	0.52	0.19
4 - descriptor	0.70	0.54	0.21
3 - descriptor	0.65	0.50	0.22

The FP:MACCS derived 5-descriptor model had the highest r^2 value and a relatively high q^2 value for the training set, which indicated that the model fitted the training set data well and appeared to be robust. However, when the model was applied to the FPMACCS-DAT test set an r^2 value of 0.19 was obtained, which indicates poor predictive ability. The 3 and 4-descriptor models produced comparable r^2 and q^2 values (Table 4.3) to the 5-descriptor model. As a rule of thumb, models with fewer descriptors tend to be more generalizable[117]. As such, these models were also applied to the FPMACCS-DAT test set to see if a model with fewer descriptors would have better predictive ability. Marginal, but not significant, increases were observed with test set r^2 values of no greater than 0.22 obtained. This indicated that all three FPMACCS-derived models were not predictive and suggests that the model could be over-fitted.

Over-fitting is a phenomenon whereby a model is able to predict, to an acceptable level, the activities of the compounds found in its training set, but is unable to mirror this predictivity when applied to an unseen test set of molecules i.e. it is not generalizable. Often this is the result of a physicochemical property being over represented and/or the model including more descriptors

than is necessary[225], but given that care has been taken to remove cross-correlated descriptors, and the fact that the maximum number of descriptors examined in this study is five (giving a maximum ratio of five records to each descriptor) these reasons for poor model performance are unlikely.

Over-fitting can also be a consequence of constructing the initial model using an inappropriate or unrepresentative training and/or test sets, which seems the most likely scenario for this experiment. If we examine the test set, it is noticeable that two of the compounds (nortriptyline and Dimethylamylamine) come from the molecules demonstrating a pKi between 4 and 5, the lowest end of the experimental activity range. Previous studies[222] have shown that QSAR models can be sensitive to experimental data at the extremes of the data range, and having a test set containing two molecules which appear at the lowest end of the activity range could be one of the reasons that the model does not appear to be generalisable.

It is also unreasonable to expect a descriptor-based QSAR model to be able to accurately predict the activity of a molecule that has a significantly different in terms of its chemical structure when compared to the molecules in the training set used to derive the QSAR model. Experiments showed that dimethylamylamine had an average pairwise Tanimoto similarity, when compared to all other molecules in the dataset of 0.17. That means, this molecule was significantly structurally different to every other molecule in the dataset, and hence expecting a model to be able to accurately predict the activity of such a disparate molecule is unreasonable.

To test the impact of inclusion of this structure on the perceived predictive ability of the FPMACCS-derived DAT model, it was removed from the test set. The r^2 value for the DAT test set (for the 3-descriptor model) increased from 0.22 to 0.63. This highlights the importance of the construction of appropriate test and training sets during QSAR construction and validation, and highlights the potential limitations of automated tools for generating diverse test and training sets, especially where the size of the dataset is limited.

Although such *post hoc* rationalisation of results is able to explain the poor performance of a QSAR test set, it is of interest to identify a methodology for determining a training and test set which would ideally eliminate the need to carry out such analyses. As such, an alternative method for identifying representative training and test sets for constructing the QSAR models was investigated based on calculating the pairwise similarity coefficients of molecules and using these, alongside the known biological activity of the molecules, to manually assign test and training sets for the MAT isoforms.

4.4.1 Similarity Coefficients

Having identified dimethylamylamine as being structurally unrelated to the other compounds in the NPS dataset used to construct the QSAR models, average pairwise Tanimoto coefficients were calculated to establish if any other compounds were distinctly dissimilar to the rest of the molecules in the dataset, as this had already demonstrated the potential to impact on the apparent quality of any models developed.

For example, if a structurally dissimilar compound was placed in the test set without having its chemistry represented in the training set, it would be impossible for any model generated to accurately predict the activity of this dissimilar compound (any predicted values would be artefactual) and hence this could give an inaccurate representation of the quality of the model, as was the case for the FP:MACCS-derived model (Section 4.3.1.2). As such, compounds that are distinctly dissimilar from all others should be removed from the dataset prior to test and training sets being defined and models being constructed. For the purposes of this experiment an average Tc value of less than 0.2 was the threshold for a molecule to be considered significantly dissimilar to the others in the dataset.

Only one compound in the dataset had an average Tc value of less than 0.2. Dimethylamylamine had the lowest average similarity coefficient of 0.17 (± 0.15) (see Figure 4.2 for all T_c data). It was removed from the dataset prior to the construction of the models based on the Tc-derived training sets and test sets.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Average
1	1	0.36	0.40	0.42	0.40	0.49	0.42	0.74	0.19	0.29	0.27	0.66	0.66	0.91	0.28	0.29	0.28	0.37	0.37	0.63	0.33	0.30	0.31	0.33	0.71	0.35	0.28	0.28	0.19	0.28	0.26	0.42
2	0.36	1	0.21	0.27	0.25	0.32	0.27	0.29	0.16	0.21	0.19	0.31	0.31	0.34	0.20	0.19	0.15	0.27	0.27	0.36	0.21	0.20	0.20	0.21	0.32	0.22	0.20	0.20	0.11	0.20	0.16	0.26
3	0.40	0.21	1	0.23	0.22	0.30	0.28	0.56	0.09	0.64	0.59	0.40	0.40	0.44	0.18	0.30	0.21	0.74	0.74	0.28	0.33	0.24	0.24	0.23	0.42	0.33	0.32	0.18	0.22	0.19	0.20	0.36
4	0.42	0.27	0.23	1	0.82	0.38	0.39	0.34	0.11	0.29	0.27	0.36	0.36	0.40	0.23	0.22	0.15	0.29	0.29	0.56	0.26	0.22	0.22	0.26	0.37	0.27	0.25	0.23	0.13	0.25	0.20	0.32
5	0.40	0.25	0.22	0.80	1	0.36	0.37	0.33	0.10	0.28	0.26	0.34	0.34	0.38	0.22	0.21	0.15	0.28	0.28	0.52	0.24	0.21	0.21	0.25	0.35	0.26	0.24	0.22	0.13	0.24	0.19	0.31
6	0.49	0.32	0.30	0.38	0.36	1	0.44	0.45	0.25	0.34	0.35	0.65	0.65	0.53	0.35	0.42	0.31	0.37	0.37	0.53	0.33	0.44	0.45	0.43	0.47	0.34	0.25	0.35	0.21	0.25	0.33	0.41
7	0.42	0.27	0.28	0.39	0.37	0.44	1	0.35	0.14	0.46	0.36	0.40	0.40	0.40	0.28	0.29	0.18	0.36	0.36	0.53	0.29	0.37	0.29	0.28	0.37	0.30	0.24	0.28	0.16	0.24	0.29	0.35
8	0.74	0.29	0.56	0.34	0.33	0.45	0.35	1	0.14	0.28	0.25	0.61	0.61	0.81	0.24	0.42	0.27	0.34	0.34	0.47	0.31	0.29	0.29	0.31	0.68	0.31	0.24	0.24	0.27	0.24	0.23	0.39
9	0.19	0.16	0.09	0.11	0.10	0.25	0.14	0.14	1	0.08	0.09	0.20	0.20	0.17	0.11	0.12	0.16	0.12	0.12	0.17	0.15	0.16	0.17	0.15	0.15	0.16	0.12	0.11	0.07	0.12	0.14	0.17
10	0.29	0.21	0.64	0.29	0.28	0.34	0.46	0.28	0.08	1	0.81	0.31	0.31	0.31	0.20	0.23	0.16	0.84	0.84	0.36	0.34	0.25	0.26	0.24	0.28	0.35	0.21	0.20	0.14	0.21	0.25	0.35
11	0.27	0.19	0.59	0.27	0.26	0.35	0.36	0.25	0.09	0.81	1	0.32	0.32	0.28	0.21	0.22	0.19	0.75	0.75	0.32	0.37	0.29	0.29	0.27	0.26	0.39	0.21	0.21	0.14	0.21	0.27	0.35
12	0.66	0.31	0.40	0.36	0.34	0.65	0.40	0.61	0.20	0.31	0.32	1	1.00	0.72	0.26	0.31	0.36	0.34	0.34	0.47	0.31	0.44	0.45	0.34	0.65	0.32	0.24	0.27	0.21	0.24	0.27	0.42
13	0.66	0.31	0.40	0.36	0.34	0.65	0.40	0.61	0.20	0.31	0.32	1.00	1	0.72	0.26	0.31	0.36	0.34	0.34	0.47	0.31	0.44	0.45	0.34	0.65	0.32	0.24	0.27	0.21	0.24	0.27	0.42
14	0.91	0.34	0.44	0.40	0.38	0.53	0.40	0.81	0.17	0.31	0.28	0.72	0.72	1	0.27	0.32	0.30	0.39	0.39	0.57	0.33	0.32	0.32	0.35	0.78	0.35	0.26	0.27	0.21	0.26	0.25	0.43
15	0.28	0.20	0.18	0.23	0.22	0.35	0.28	0.24	0.11	0.20	0.21	0.27	0.27	0.27	1	0.26	0.16	0.22	0.22	0.29	0.25	0.25	0.23	0.26	0.26	0.25	0.26	1.00	0.19	0.26	0.28	0.29
16	0.29	0.19	0.30	0.22	0.21	0.42	0.29	0.42	0.12	0.23	0.22	0.31	0.31	0.32	0.26	1	0.18	0.24	0.24	0.28	0.23	0.22	0.23	0.26	0.29	0.21	0.26	0.26	0.31	0.26	0.25	0.28
17	0.28	0.15	0.21	0.15	0.15	0.31	0.18	0.27	0.16	0.16	0.19	0.36	0.36	0.30	0.16	0.18	1	0.19	0.19	0.18	0.34	0.56	0.57	0.34	0.26	0.35	0.16	0.16	0.17	0.16	0.34	0.27
18	0.37	0.27	0.74	0.29	0.28	0.37	0.36	0.34	0.12	0.84	0.75	0.34	0.34	0.39	0.22	0.24	0.19	1	1.00	0.38	0.40	0.26	0.26	0.28	0.35	0.41	0.23	0.22	0.15	0.23	0.24	0.38
19	0.37	0.27	0.74	0.29	0.28	0.37	0.36	0.34	0.12	0.84	0.75	0.34	0.34	0.39	0.22	0.24	0.19	1.00	1	0.38	0.40	0.26	0.26	0.28	0.35	0.41	0.23	0.22	0.15	0.23	0.24	0.38
20	0.63	0.36	0.28	0.56	0.52	0.53	0.53	0.47	0.17	0.36	0.32	0.47	0.47	0.57	0.29	0.28	0.18	3.80	3.80	1	0.33	0.27	0.28	0.33	0.48	0.35	0.28	0.29	0.15	0.28	0.24	0.61
21	0.33	0.21	0.33	0.26	0.24	0.33	0.29	0.31	0.15	0.34	0.37	0.31	0.31	0.33	0.25	0.23	0.34	0.40	0.40	0.33	1	0.37	0.38	0.45	0.29	0.84	0.22	0.25	0.21	0.22	0.36	0.34
22	0.30	0.20	0.24	0.22	0.21	0.44	0.37	0.29	0.16	0.25	0.29	0.44	0.44	0.32	0.25	0.22	0.56	0.26	0.26	0.27	0.37	1	0.85	0.32	0.29	0.38	0.23	0.25	0.17	0.23	0.30	0.33
23	0.31	0.20	0.24	0.22	0.21	0.45	0.29	0.29	0.17	0.26	0.29	0.45	0.45	0.32	0.23	0.23	0.57	0.26	0.26	0.28	0.38	0.85	1	0.34	0.30	0.39	0.23	0.23	0.17	0.23	0.36	0.34
24	0.33	0.21	0.23	0.26	0.25	0.43	0.28	0.31	0.15	0.24	0.27	0.34	0.34	0.35	0.26	0.26	0.34	0.28	0.28	0.33	0.45	0.32	0.34	1	0.30	0.47	0.22	0.26	0.25	0.22	0.39	0.32
25	0.71	0.32	0.42	0.26	0.35	0.47	0.37	0.68	0.15	0.28	0.26	0.65	0.65	0.78	0.25	0.28	0.26	0.35	0.35	0.48	0.29	0.29	0.30	0.30	1	0.30	0.24	0.26	0.23	0.24	0.22	0.39
26	0.35	0.22	0.33	0.37	0.26	0.34	0.30	0.31	0.16	0.35	0.39	0.32	0.32	0.35	0.25	0.21	0.35	0.41	0.41	0.35	0.84	0.38	0.39	0.47	0.30	1	0.22	0.25	0.20	0.22	0.38	0.35
27	0.28	0.20	0.19	0.27	0.24	0.25	0.24	0.24	0.12	0.21	0.21	0.24	0.24	0.26	0.26	0.26	0.16	0.23	0.23	0.28	0.22	0.23	0.23	0.22	0.24	0.22	1	0.26	0.15	1.00	0.21	0.28
28	0.28	0.20	0.18	0.25	0.22	0.35	0.28	0.24	0.11	0.20	0.21	0.27	0.27	0.27	1.00	0.26	0.16	0.22	0.22	0.29	0.25	0.25	0.23	0.26	0.26	0.25	0.26	1	0.19	0.26	0.28	0.29
29	0.19	0.11	0.22	0.13	0.13	0.21	0.16	0.27	0.07	0.14	0.14	0.21	0.21	0.21	0.19	0.31	0.17	0.15	0.15	0.15	0.21	0.17	0.17	0.25	0.23	0.20	0.15	0.19	1	0.15	0.23	0.21
30	0.28	0.20	0.19	0.25	0.24	0.25	0.24	0.24	0.12	0.21	0.21	0.24	0.24	0.26	0.26	0.26	0.16	0.23	0.23	0.28	0.22	0.23	0.23	0.22	0.24	0.22	1.00	0.26	0.15	1	0.21	0.28
31	0.26	0.16	0.20	0.20	0.19	0.33	0.29	0.23	0.14	0.26	0.27	0.27	0.27	0.25	0.28	0.25	0.34	0.24	0.24	0.24	0.36	0.30	0.36	0.39	0.22	0.38	0.21	0.28	0.23	0.21	1	0.29

Figure 4. 2: Correlation matrix of the pairwise similarity coefficients for the 31 compounds in the dataset (see table 4.1 for compound ID). The more similar a compound is to another, the higher the value of the coefficient. All pairwise values less than 0.20 were highlighted in red to give a visual representation of which compounds are most dissimilar to others in the dataset, and the final column of the correlation matrix is the average value of similarity coefficients for a compound in relation to the dataset as a whole. Compound 9 (Dimethylamylamine) has a much lower average T_c (0.17) in comparison to the rest of the dataset which indicates this molecule is structurally dissimilar to the rest of the dataset and should not be used to construct and validate models

4.4.2 Diverse Tanimoto Subsets

The Tanimoto-derived training and test sets for DAT, NET and SERT were identified using the pK_i values as the discriminant variable. Figure 4.3 shows the distribution of pKi for the molecules for all three MAT isoforms, DAT NET and SERT across the dataset as a whole. It should be noted that SERT has the largest range of experimental pK_i values (7 log units) for the molecules in the NPS dataset. The range of experimental pKi values for DAT and NET is smaller (5 log units for both).

Normal distribution of the experimental values is observed for both DAT and SERT (99% confidence using the Shapiro-Wilks test). However, this is not true for NET. When the same statistical test was applied to NET, and based on the *p* value at the 90% confidence limit, it was determined that the distribution of NET is not normal (i.e. it is skewed). If data is not normally distributed, the model will not be able to accurately predict pKi values that fall outside of the normal distribution range. This potentially places inherent limitations on any predictive models developed for the NET isoform, which will need to be ameliorated by the careful selection of appropriate test and training sets for this MAT.
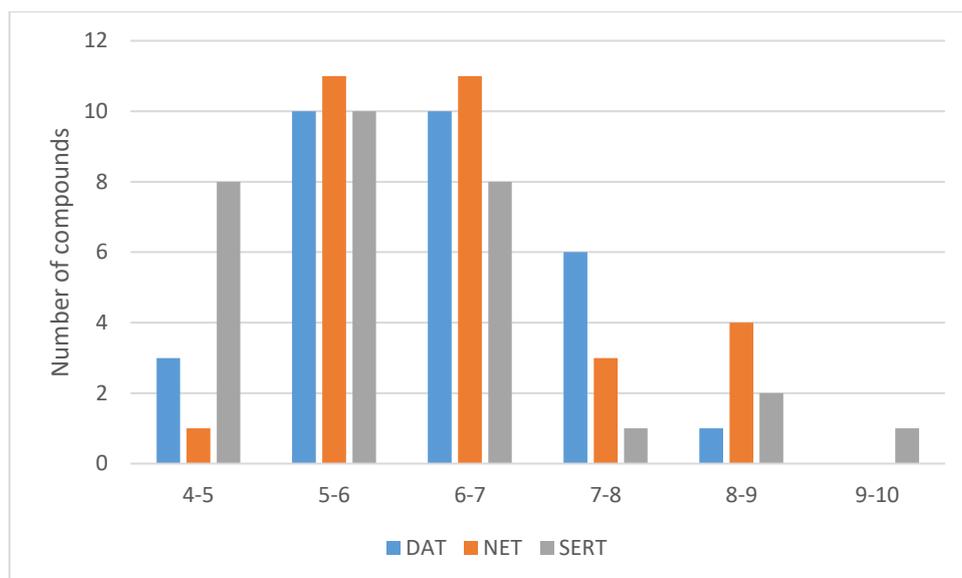
Figure 4. 3: The distribution of pKi values, represented in one log unit divisions, across the NPS in the dataset as a whole for DAT (blue), NET (orange) and SERT (grey).

To ensure the training sets for each of the MAT isoforms were as structurally diverse as possible, similarity coefficients (see Figure 4.2) were used to guide whether a molecule was placed into the training or test set, as described in the methods section above. This was carried out to increase the degree of dissimilarity between individual members in the training set, whilst ensuring the training and test sets were still representative of the overall dataset, with an overall ambition to achieve the most generalizable QSAR models possible.

The resultant training and test set for each of the MAT isoforms contained compounds that represented the spread of pK_i values across the dataset as a whole (which can be seen in Figures 4.4, 4.5 and 4.6).
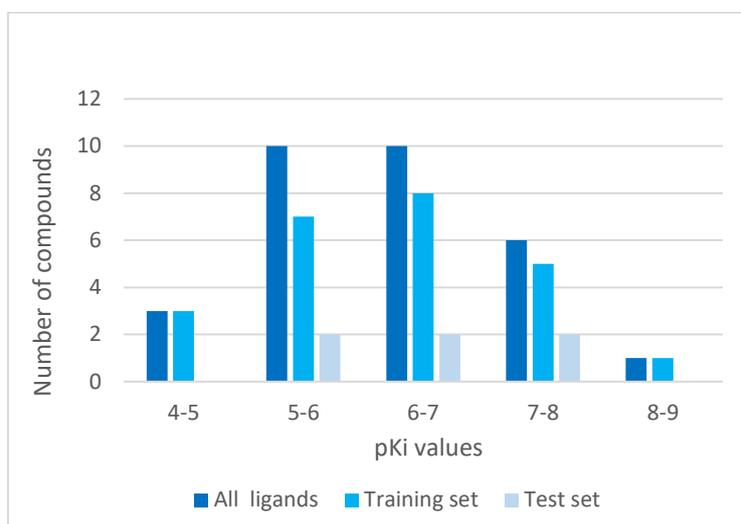
Figure 4. 4: The distribution of pKi values for the dataset compounds in the NPS dataset for DAT (dark blue) compared to the relative distributions in the training (blue) and test (light blue) sets used to build and validate the Tc-derived DAT QSAR model.

The distribution of pKi values for both test and training set in DAT were normal (p=0.902, W=0.967, H_0 is accepted). The pKi ranges with the largest number of molecules were pKi 5-6 and pKi 6-7. Both the test and training sets reflected this overall distribution. Figure 4.4 shows that the trend of pKi values for the training set was mirrored by the test set. However, as there was only one compound (for the entire DAT dataset) that had a pKi of above 8, there was no capacity to have a test set compound in this pKi value range.
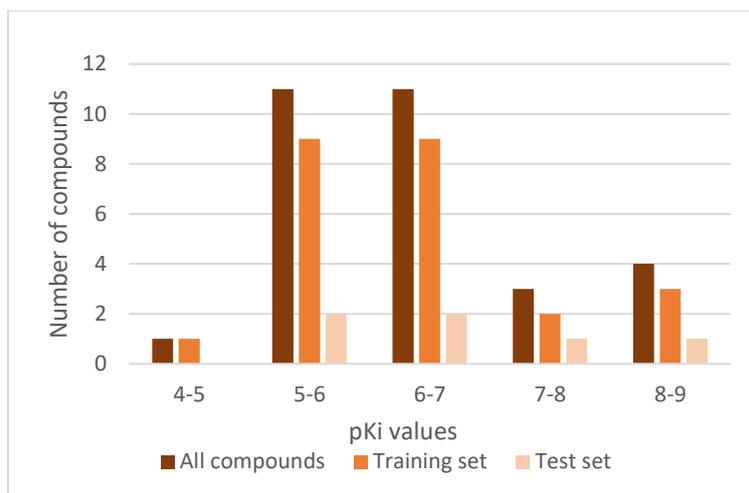
Figure 4. 5: The distribution of pKi values for the dataset compounds in the NPS dataset for NET (dark orange) compared to the relative distributions in the training (orange) and test (light orange) sets used to build and validate the NET QSAR model.

The distribution of pKi values for NET did not have a normal distribution from pKi range 4-9 (p=0.902, W=0.876, H_0 was rejected). There was an uneven distribution of pKi values with the compounds associated with NET. Twenty-three compounds out of the total dataset had a pK_i value between 5-7, with only 1 compound (nortriptyline) having a pKi value in the pKi 4-5 range. Three compounds fell into the 7-8 pKi range and 4 compounds were located in the 8-9 pKi range. Despite the skew of pKi values for the NET isoform, training and test sets were identified for NET that mirror the distribution of pKis across the dataset as a whole. This gives the best chance of identifying a QSAR model for the isoform which is robust and predictive.
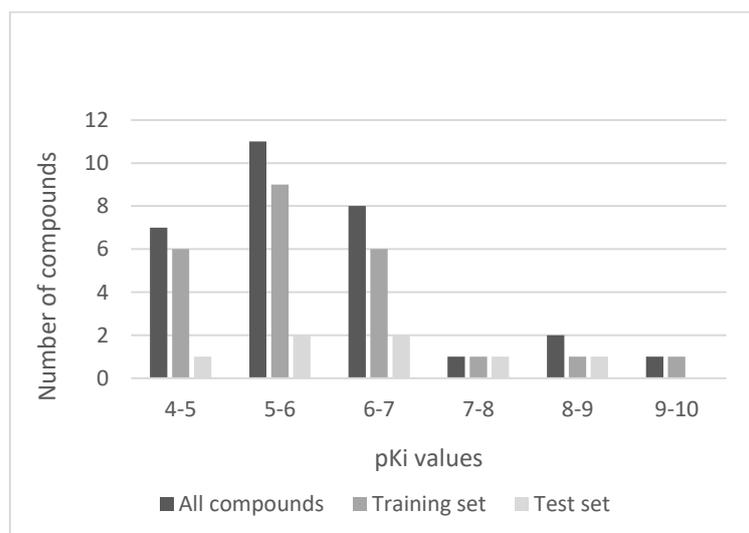
Figure 4. 6: The distribution of pKi values for the dataset compounds in the NPS dataset for SERT (dark grey) compared to the relative distributions of the compounds in the training (grey) and test (light grey) sets used to build and validate the SERT QSAR.

The pKi values for the compounds associated with SERT had a broader range in comparison to DAT and NET. The data for SERT were normally distributed (p=0.902, W=0.957, H_0 is accepted). The SERT pKi values ranged from 3-10 for the dataset in comparison to DAT and NET which had a range of 4-9. The data presented in Figure 4.5 shows that the both training and test sets mirror the distribution of pKi across the whole dataset thereby showing training and test sets that are clearly representative of the dataset as a whole.

4.4.3 Descriptor Selection

As the pKi values for the compounds in the dataset were different for each of the MAT isoforms the compounds within the training set and test set for DAT, NET and SERT were different. The differences in training sets, unsurprisingly, gave rise to different molecular descriptors emerging as being important for describing the observed variance in pKi for each of the MAT isoforms. Therefore, the QSAR equations obtained for DAT, NET and SERT are different. Table 4.4 details the differences between the training and test sets for each of the MAT.

4.4.5 Building and evaluating QSAR models

Table 4. 4: The names of the compounds that formed the training and test sets for the DAT, NET and SERT QSAR models.

Name of compound in each training and test set					
DAT Model		**NET Model**		**SERT Model**	
Training set	**Test set**	Training set	**Test set**	Training set	**Test set**
5-APB	**6-APB**	1-Naphyrone	**5-IAI**	1-Naphyrone	**Amitriptyline**
5-Iodo-aminoindane	**Nortriptyline**	5-APB	**Bupropion**	5-APB	**Bupropion**
Benzedrone	**Mazindol**	6-APB	**Citalopram**	5-Iodo-aminoindane	**Cocaine**
Bupropion	**MDAI**	Amitriptyline	**Fluoxetine**	6-APB	**Imipramine**
Citalopram	**Mephedrone**	Benzedrone	**Mazindol**	Benzedrone	**Naphyrone**
Cocaine	**1-Naphyrone**	Cocaine	**WIN 35428**	Citalopram	**Nomifensine**
Desipramine		Desipramine		Desipramine	
Desoxypipradrol		Desoxypipradrol		Desoxypipradrol	
Fluoxetine		GBR 12935		Fluoxetine	
GBR 12935		Imipramine		GBR 12935	
Imipramine		MDAD		Mazindol	
MDAT		MDAT		MDAI	
Methiopropamine		Mephedrone		MDAT	
Methylethcathinone		Methiopropamine		Mephedrone	
MNB-cathinone		Methylethcathinone		Methiopropamine	
Naphyrone		MNB-cathinone		Methylethcathinone	
Nisoxetine		Naphyrone		MNB-cathinone	
Nomifensine		Nisoxetine		Nisoxetine	
Amitriptyline		Nomifensine		Nortriptyline	
R-MDMA		Nortriptyline		RTI-55	
RTI-55		R-MDMA		R-MDMA	
S-Amphetamine		RTI-55		S-Amphetamine	
S-MDMA		S-amphetamine		S-MDMA	
WIN 35428		S-MDMA		WIN 35428	

4.4.5.1 Development of DAT QSAR models

In an effort to develop a robust system to identify appropriate training and test sets, and to improve upon the r^2 values obtained for the FPMACCS-derived models outlined above, an investigation was carried out to determine if altering the composition of the test and training sets, to ensure that they are representative of the dataset as a whole, could improve the predictivity of the models.

As such, the training and test sets for the DAT isoform werr re-distributed according to structural diversity, ensuring that the range of pKis identified for experimental studies were represented in both of the sets. This was achieved by dividing the dataset into one log unit bins in terms of pKi values and using average T_c values as a guide to placing molecules into either a test or training set, thus ensuring proportionate representation of chemical structure and observed biological activity in both the test and training sets (see section 4.2.2). The Tc-derived training set for DAT was then used to construct a new QSAR model.

The Tc-derived DAT training set unsurprisingly identified a different set of descriptors as being important in predicting the variance in the dataset when compared to the FPMACCS-derived model. The iterative process described earlier was employed when optimising the Tc-derived QSAR model i.e. the molecular descriptors in the QSAR model that were ranked lowest in terms of relative importance for their contribution to explaining the variance in the dataset were removed one by one until an appropriate model, with good r^2 and q^2 values and a minimum number of molecular descriptors was identified.

Table 4.5 shows the r^2 and q^2 values for the estimated normalized linear Tc-derived DAT QSAR model. The model built using 3 descriptors (Equations 4.2) was identified as the most appropriate as this has the highest r^2 and q^2 values which is indicative of a predictive and robust model, and comprises a minimum number of uncorrelated molecular descriptors. Hence, the model should be generalisable.

When the 3-descriptor model was applied to the Tc-derived test set, an improvement in the correlation between the predicted and experimental pKi values over the FPMACCS-derived model was observed (Figure 4.7). As such, the methodology for identifying test and training sets by examining their experimental pKi and calculated average Tc values was used to construct test and training sets for the NET and SERT isoforms.

Table 4. 5: the r^2 and q^2 values for the three different estimated normalized QSAR models developed for DAT using the 3, 4 and 5 descriptors most highly correlated with biological activity. These models were built using T_c-derived training and test set.

DAT Model	DAT Tanimoto-derived training set results		DAT Tanimoto-derived test set results
	r^2	q^2	r^2
5 - descriptor	0.76	0.48	0.57
4 - descriptor	0.71	0.40	0.59
3 - descriptor	0.68	0.51	0.63

5-descriptor model for DAT (Tanimoto-derived training set)

pKi / SD (pKi) = 6.59630 - 0.75775 * b_max1len / SD(b_max1len) + 0.48262 * FASA_H / SD(FASA_H) - 0.38144 * opr_leadlike / SD (opr_leadlike) -0.34355 * PEOE_VSA+0 / SD(PEOE_VSA+0) - 0.26472 * vsurf_IW6 / SD (vsurf_IW6)

4-descriptor model for DAT (Tanimoto-derived training set)

pKi / SD (pKi) = 5.449 - 0.74460 * b_max1len / SD(b_max1len) + 0.52065 * FASA_H / SD(FASA_H) - 0.43807 * opr_leadlike / SD (opr_leadlike) -0.27305 * PEOE_VSA+0 / SD(PEOE_VSA+0)

3-descriptor model for DAT (Tanimoto-derived training set)

pKi / SD (pKi) = 5.27760 -0.70255 * b_max1len / SD(b_max1len) +0.38911 * FASA_H / SD(FASA_H) - 0.29130 * opr_leadlike / SD (opr_leadlike)

Equations 4.2: 5, 4 and 3-descriptor QSAR DAT estimated normalized linear QSAR equations. The 3-descriptor model is the most robust, and is built using the descriptors b_max1len, FASA_H and opr_leadlike.
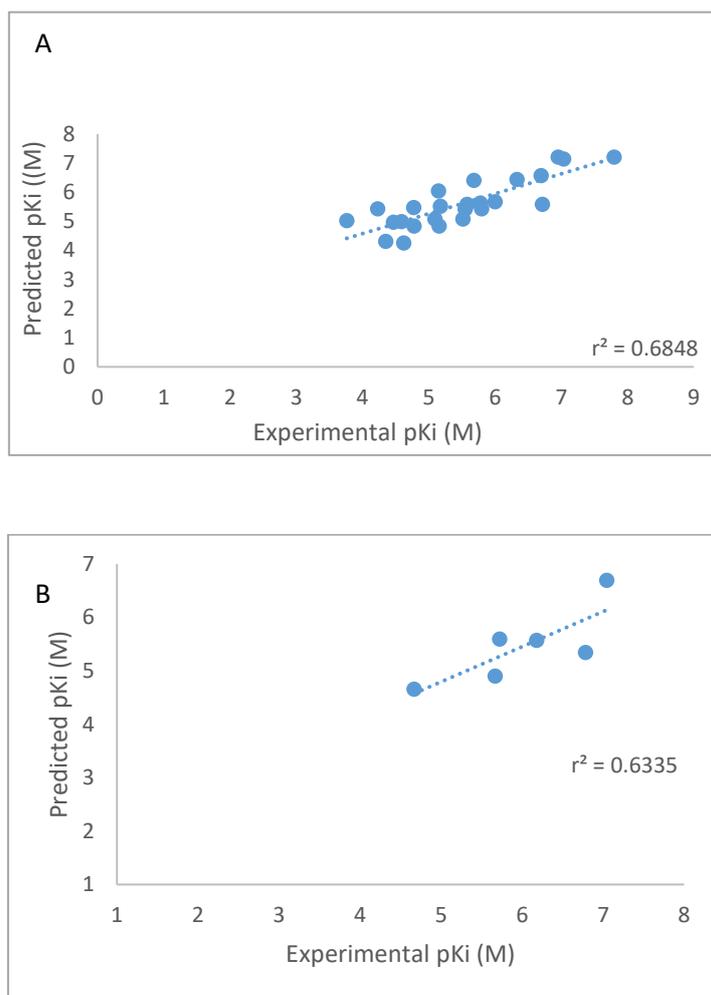
Figure 4.7: Correlation graphs representing the 3-descriptor QSAR model performance for the T_c-derived training and test set for DAT. (A) Plot of the predicted vs experimental values of pKi for the training set according to the QSAR model (r^2 =0.68). (B) Plot of the predicted vs experimental values of pKi for the test set according to the QSAR model (r^2 =0.63), no outliers were detected (P = 0.01)

4.5.1 Interpreting the 3-descriptor T_c-derived DAT QSAR model

The most predictive DAT QSAR model was the 3-descriptor model built using the Tc-derived training set. A table illustrating the relative importance of the descriptors, as indicated by the QuaSAR module in MOE is give below (Table 4.6).

Table 4. 6: The molecular descriptors used to build the T_c-derived 3-descriptor DAT QSAR model and their relative importance for describing the observed variance in the experimental data, as determined by MOE.

Order of Importance	Molecular descriptor	Descriptor description
1.000000	b_max1len	Maximum single bond chain length
0.790002	FASA_H	Fractional hydrophobic Surface Area
0.386015	opr_leadlike	The Oprea rules lead-like assessment

If we consider the QSAR model generated in more detail we observe the following. Both b_max1len and opr_leadlike have negative signs before the coefficients, and hence can both be considered to be penalty terms in the overall prediction of biological affinity.

B_max1len is an integer describing the length of the longest consecutive single bond chain in a molecule. The negative coefficient for b_max1len indicates that a shorter single bond chain length will predict a higher pKi value. Information regarding the MAT binding sites were taken from Chapter 3 in an effort to contextualise this observation.

The presence of long single bond chains may allow for a high degree of flexibility within the compound which may prevent strong protein-compound interaction being formed, in comparison to more rigid and compact compounds (e.g. Desoxypipradrol). A possible argument for this could be based on entropy, i.e. rigid molecules will have fewer entropic penalties associated with binding and result in a stronger protein-ligand interaction, and hence are predicted to have higher pKis. Again, this is evident when comparing RTI-55 (pKi = 8.49) and Desipramine (pKi = 4.10). This could indicate that less flexible compounds bind better to DAT, and hence there are steric factors involved in the binding of small molecules in the DAT binding site.

Analysis of the DAT binding site (using the homology model, accession number Q01959 [226,227]), shows that there is a narrowing in the cavity towards the bottom (Figure 8), which is where the literature and the docking studies of Chapter 3 suggest the majority of compounds bind. The opening to the cavity has an approximate diameter of 10 Å and the bottom of the cavity has an approximate diameter of 7 Å. Compact and rigid molecules may therefore be able to better access, and bind to the bottom

of the cavity than more flexible ones, and be subject to fewer entropic penalties upon binding. Hence, this could explain why the b_maxlen descriptor is highlighted as being important in explaining the variance in the 3-descriptor Tc-derived DAT model.
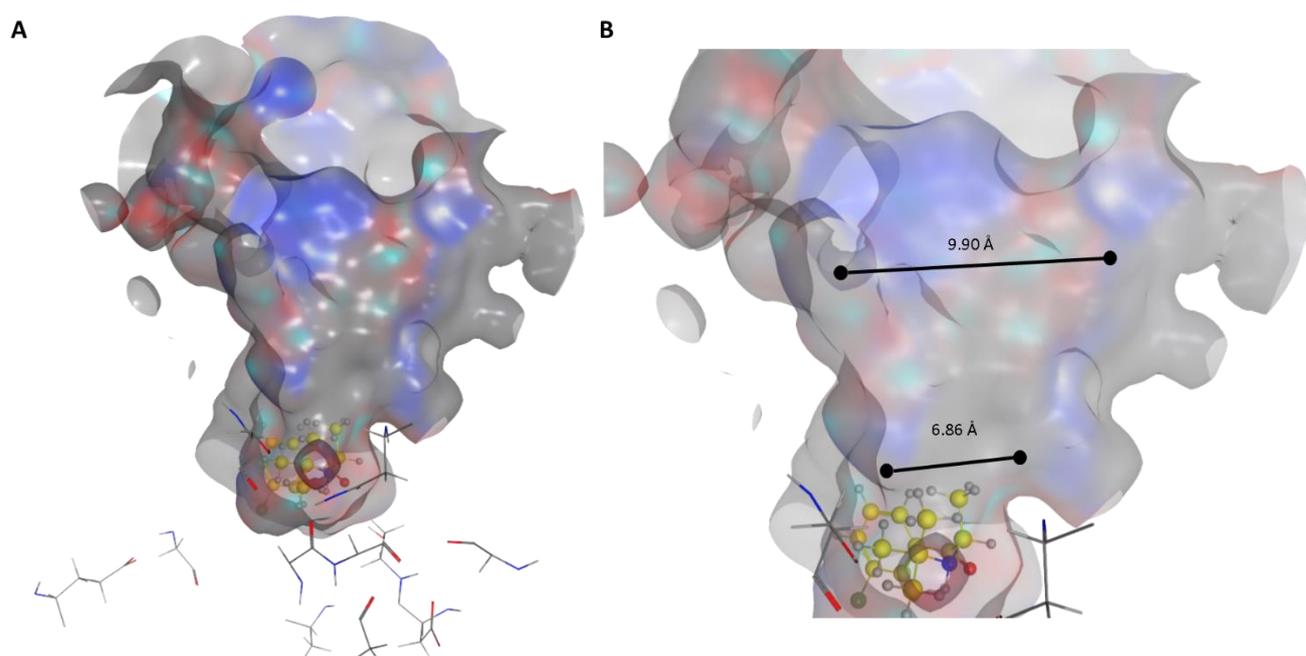
Figure 4. 8: **A** – overview of the DAT binding site (homology model, accession number Q01959) containing the docked ligand RTI-55 (yellow stick model). **B** – Zoomed in section of the DAT binding site, with approximate diameters (in Å) at two different regions of the binding site illustrating the narrowing of the binding site.

Figure 4.8 shows RTI-55 docked into the DAT binding site. This ligand has the highest experimentally derived pKi value (pK_i – 8.50) for DAT. RTI-55 also has the small b_max1len value of 2 (the largest being 7) of all the molecules in the dataset. As such, it is unsurprising that the 3-descriptor QSAR equation predicts a high binding affinity for RTI-55 (pK_i = 7.21) which is consistent with its experimentally determined value.

FASA_H is a term describing the water accessible surface area of all hydrophobic atoms of a compound. As there is a positive coefficient for FASA_H in the 3-descriptor DAT QSAR equation this implies that a greater accessible surface area (of hydrophobic atoms) equates to a higher pKi.

The results obtained from Chapter 3 indicate that hydrophobic interactions are important binding in the DAT active site, as a large number of the residues that make up the binding site in DAT, and hence are involved in protein-ligand contact, are hydrophobic (e.g. L80, A81, V152, F320 and F326). The FASA_H descriptor looks at the water accessible surface area (ASA) of all hydrophobic atoms in a

compound, and it follows that if a hydrophobic small molecule was able to displace the water from the binding cavity, it would be able to form interactions with the resultant hydrophobic surface area of the protein.

One hypothesis for why this descriptor is important is as follows. Given the hydrophobic nature of the binding cavity, it could be energetically favourable to displace the water-hydrophobic surface contacts and replace these with hydrophobic protein-ligand contacts. This is supported by the positive coefficient in the equation which indicates the more ability there is to form hydrophobic protein-ligand contacts i.e. the greater the water accessible surface area, the higher the pKi i.e. a compound that can fill the bottom of the binding cavity will have a higher pKi than one which only partially fills it (see Figure 4.8). This interpretation provides context as to why this descriptor was considered important for determining biological activity, this is supported by the high affinity observed, and predicted for RTI-55 which fills the DAT cavity.

The values that a molecule can achieve with respect to the opr_leadlike descriptor can be, 1 which denotes that 2 or fewer violations to the Oprea[228] drug like criteria have been incurred and 0 which indicates that the compound has 3 or more violations to the Oprea criteria. The criteria include having between 0-2 hydrogen bond donors, between 2 and 9 hydrogen bond acceptors, no more than 8 rotatable bonds and between 1 and 4 rigid bonds. As the coefficient associated with this descriptor is negative this indicates that any compound that had a value for opr_leadlike greater than 0 would incur a penalty (reduction in calculated pK_i value), this indicates that compounds should not adhere the Oprea drug-like criteria.

All of the compounds in the DAT training set had a calculated opr_leadlike value of 1 (except GBR 12935) which signifies that the compounds have two or less violations of the Oprea drug-like criteria. The opr_leadlike descriptor is a composite of a variety of physicochemical properties, but opr_leadlike values obtained give no information on which composite part of the descriptor has been violated. Therefore, little information can be extracted from this particular descriptor with respect to its importance in the QSAR equation. The opr_leadlike had the lowest ranked importance of the descriptors included in the model which means the values for these descriptors contribute least to the model.

Consequently, it is likely that the descriptor acts as a "correction factor" for the predicted biological activities. Removing the descriptor and generating a 2-descriptor model shows the training set r^2 and q^2 values remain relatively high ($r^2 = 0.67$, $q^2 = 0.50$) but the test r^2 values decreases from $0.63 - 0.35$. Generating a simple 2-descriptor model, and seeing that it performs less well than the 3-descriptor model, demonstrates that the opr_leadlike descriptor plays a crucial role as a correction factor in

predictivity. To further test the hypothesis of the opr_leadlike descriptor behaving as a correction factor a further experiment using a modified 2-descriptor equation was applied (Equation 4.3). Assuming that the opr_leadlike was a correction factor, and hence would be a constant (-0.29310), the value at the beginning of the 3-descriptor DAT equation was modified (5.27760 -0.29310 = 4.9845)

$$pKi / SD (pKi) = 4.9845 -0.70255 * b\_max1len / SD(b\_max1len) +0.38911 * FASA\_H / SD(FASA\_H)$$

Equation 4.3: 2-descriptor QSAR DAT estimated normalized linear QSAR equations built using the descriptors b_max1len and FASA_H. A value of -0.29310 was subtracted from the correction factor at the beginning of the 3-descriptor equation.

When equation 4.3 was applied to the DAT training (minus GBR-12935) and test sets, having removed the small number of compounds for which the opr_leadlike value was zero, the r^2 and q^2 values remained the same. This proved that the presence of the opr_leadlike descriptor in the model was a simple correction factor to prevent over prediction.

As mentioned earlier, an optimal model should have the highest r^2 value (which indicates the predictive ability of the model) and a similar cross-validated q^2 value (which accounts for how robust the model is). When building a QSAR, based on current practice the maximum number of descriptors to be used is a ratio of 1 descriptor for every 5 compounds[117]. For this rule to be obeyed, a maximum of 4 descriptors should have been used to build the QSAR models for DAT, NET and SERT. However, in the case of the DAT model, use of fewer descriptors produced more accurately predicted pKi values when applied to the test set. Regarding QSAR development, the principle of "Occam's razor"[117] is commonly mentioned which suggests that a reasonable QSAR model will have a small number of simple descriptors and the simpler the model the better the predictive ability. Given these principles seem to have been borne out for the 3-descriptor T_c-derived DAT model, they were applied to constructing QSAR models for the NET and SERT isoforms.

4.3.6 Development of NET QSAR models

NET QSAR models were built using the T_c-derived training set. Equations 4.4 show the form of 3, 4 and 5-descriptor NET QSAR models obtained.

5-descriptor model for NET (Tanimoto-derived training set)

pK_i / SD (pKi) = 4.06975 +0.63056 * PEOE_VSA-0 / SD(PEOE_VSA-0) + 0.52082 * PM3_HF / SD(PM3_HF) -0.08051 * Q_VSA_FPNEG / SD(Q_VSA_FPNEG) +0.41258 * vsurf_CW4 / SD (vsurf_CW4) -0.56082 * PEOE_VSA+3 / SD(PEOE_VSA+3)

4-descriptor model for NET (Tanimoto-derived training set)

pK_i/ SD (pK_i) = 0.4.92208 + 0.49132 * PEOE_VSA-0/SD (PEOE_VSA-0) + 0.36602 * PM3_HF/SD (PM3_HF)– 0.36034 * Q_VSA_FPNEG/SD (Q_VSA_FPNEG) + 0.42240 *vsurf_CW4/SD (vsurf_CW4)

3-descriptor model for NET (Tanimoto-derived training set)

pK_i / SD (pKi) = 6.02884 + 0.40054 * PEOE_VSA-0 / SD(PEOE_VSA-0) - 0.39877 * PEOE_VSA+3 / SD(PEOE_VSA+3) - 0.42037 * Q_VSA_FPNEG / SD(Q_VSA_FPNEG)

Equation 4.4: 5, 4 and 3-descriptor QSAR estimated normalized linear models built using the T_c-derived NET training set.

The performances of the 3, 4 and 5-descriptor Tc-derived NET QSAR models with respect to their predictivity and robustness are summarised below (Table 4.7).

Table 4. 7: Summary of the r^2 and q^2 values for the three best-performing estimated normalized linear QSAR models (Equations 4.4) for the T_c-derived NET training and test sets.

NET Model	NET Tanimoto training set results		NET Tanimoto test set results
	r^2	q^2	r^2
5 - descriptor	0.66	0.44	0.11
4 - descriptor	0.62	0.39	0.13
3 - descriptor	0.60	0.39	0.10
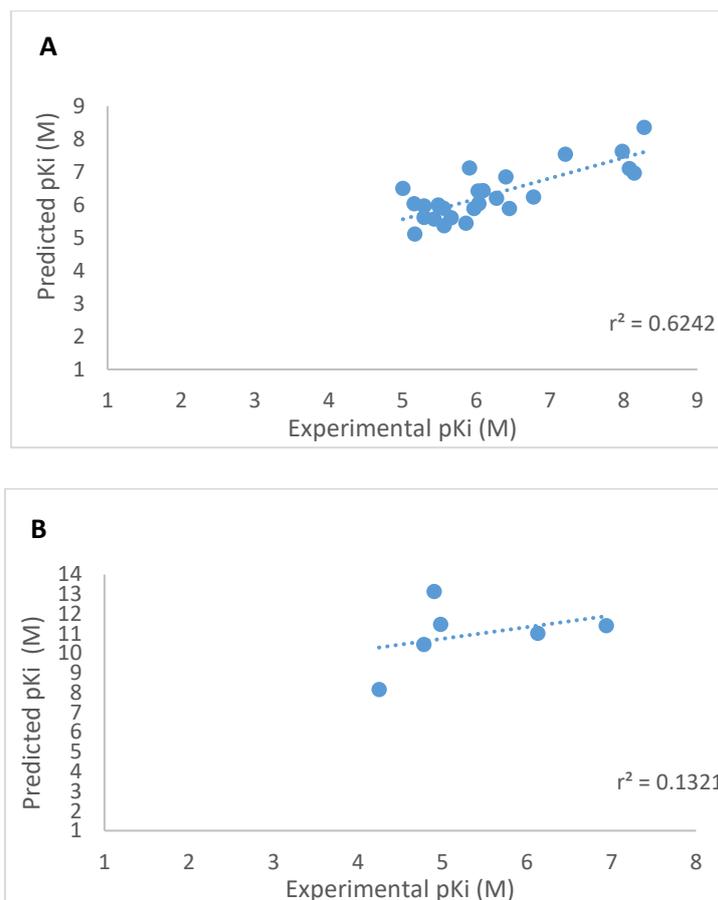
Figure 4. 9: Plots to show the performance for the NET QSAR model (Tc-derived training and test set) (A) Plot of the predicted vs experimental values of pKi for the training set according to the 4-descriptor Tc-derived NET QSAR model (r^2 = 0.62), no outliers were detected (P = 0.01). (B) Plot of the predicted vs experimental values of pKi for the test set according to the 3-descriptor Tc-derived NET QSAR model (r^2 = 0.13), no outliers were detected (P = 0.01)

4.5.2 Evaluation of NET **QSAR** models

During the descriptor selection process, it was noted that only electronic descriptors were identified as being important for explaining the observed variance in the biological activities of the Tc-derived training set for the NET QSAR models. Table 4.8 details the molecular descriptors used to build the NET QSAR models.

The best performing model was the T_c-derived 4-descriptor NET QSAR model, as this produced the highest r^2 for the test set, and the r^2 and q^2 values obtained for the training set were similar across the 3, 4 and 5-descriptor models. The 4-descriptor NET model returns an r^2 value of 0.62 for the training set, but for the test set the r^2 value achieved, was 0.13.

124

These results indicate that the QSAR models built for NET are not generalizable, as they are unable to successfully predict the biological activity of the test set compounds to an acceptable level.

Given that care has been taken to ensure that the test and training sets were representative of the dataset as a whole, and that the descriptors carried forwards to construct the NET models do not show significant cross-correlation, the reason that the model is not generalizable is not immediately clear. As such it was important to examine the descriptors in more detail (Table 4.8), as well as the NET dataset as a whole.

Table 4. 8: The molecular descriptors and their relative importance used to build the T_c-derived 4-descriptor NET QSAR model.

Order of Importance	Molecular descriptor	Descriptor description
1.000000	PEOE_VSA-0	Total negative 0 van der Waals surface area
0.542010	PM3_HF	Heat of Formation (kcal)
0.446303	vsurf_CW4	Capacity factor at -2.0
0.153069	Q_VSA_FPNEG	Fractional polar negative VDW surface area

Given that these descriptors were each designed to capture electronic characteristics, they were checked again to ensure that they were not cross-correlated with one another (Table 4.9). None of the descriptors were correlated to another with an absolute correlation value of 0.7 or greater, and hence fulfilled the inclusion criteria outlined in the methods section above.

Table 4. 9: Correlation matrix of the electronic descriptors used to build the 4-descriptor NET QSAR model. The values show that none of these descriptors correlate with one another i.e. have values greater than 0.7, or less than -0.7.

	PEOE_VSA-0	PM3_HF	Q_VSA_FPNEG	vsurf_CW4
PEOE_VSA-0	1	-0.03379	-0.256334588	-0.289223744
PM3_HF	-0.033790726	1	-0.602620891	-0.689260691
Q_VSA_FPNEG	-0.256334588	-0.60262	1	0.58275374
vsurf_CW4	-0.289223744	-0.68926	0.58275374	1

In an effort to understand the poor performance of the NET models, each of the descriptors used in the construction of the 4-descriptor model was examined.

PM3_HF is a molecular descriptor that calculates the enthalpy of formation, with the overall model indicating molecules that have a higher calculated enthalpy of formation will result in prediction of a greater pKi value. There was no obvious relationship between the importance of this descriptor and the composition of the binding site of NET. A possible explanation for this may be because the molecular descriptor PM3_HF is related to a diverse range of physico-chemical properties and lacks specificity.

Vsurf_CW4 is a capacity factor that calculates the ratio of hydrophilic surface to the total molecular surface. The positive coefficient associated with this descriptor indicates that molecules with a greater relative hydrophilic area will equate to a higher pKi values. Similarities can be drawn between this and the FASA_H descriptor highlighted as being important for the DAT QSAR model, in that the descriptor is essentially rewarding those molecules that can form hydrophobic contacts with the hydrophobic residues in the putative NET binding cavity and is emphasising the importance of a molecule being able to fill the cavity.

PEOE_VSA-0 is the most important descriptor identified for the 4-descriptor NET QSAR model. It is an electronic descriptor which accounts for the partial charge on the van der Waals surface area, in this case looking at the sum of the van der Waals surface area that has a partial charge between -0,05 and 0 i.e. a surface that is very weakly negatively charged, or neutral The positive coefficient associated with this descriptor indicates that compounds that have a large weakly negative/neutral VSA value will give rise to a greater pKi value. However, due to the nature of the NET binding site it is expected that molecules that have positively charged regions would interact better with NET, due to the dominance of the Asp 75 residue on the electronic characteristics of the binding cavity (Figure 4.9).

Q_VSA_FPNEG describes the fractional negative polar VSA, such that the partial charge is less than -0.2 divided by the total surface area, and is calculated based on the atomic partial charges stored in the MDB file for a molecule. The negative coefficient associated with this descriptor implies that any values obtained other than 0 would result in a penalty being applied and a reduction in predicted pKi value. Analysis of the NET binding site conducted in Chapter 3 did not allude to the importance regarding polarity of the binding site residues; to the contrary 85% of the binding site was composed of hydrophobic amino acid residues. Again, it is possible that this penalty arises due to the presence of the Asp 75 residue, and its role in determining the electronic properties of the NET binding cavity. The electronegative environment created by the deprotonated aspartate would repel molecules with an overall average negative charge – a feasible reason as to why this is a negative term in the best-performing QSAR equation. However, these hypotheses are merely that, and no firm conclusions can

be drawn from these suppositions. This prevents any meaningful data being taken from the descriptors associated with the NET QSAR model.

The almost exclusive interactions by compounds, demonstrated in the docking studies, at the negatively charged region of the NET binding cavity (Figure 4.10) could be a reason as to why only electronic descriptors were suggested by MOE. Analysis of the NET binding site highlights a large negatively charged region (Figure 4.10).
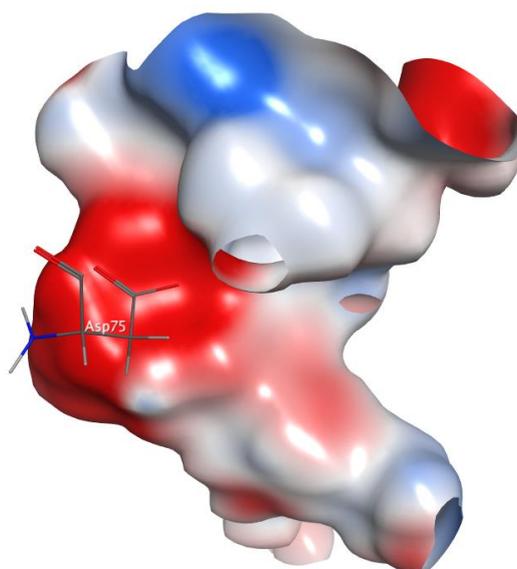
Figure 4. 10: The putative NET binding site (based on the homology model, accession number P23975). The binding site is coloured according to negatively (red) and positively (blue) charged regions Areas with no charge are coloured white. The Figure illustrates an aspartate residue (Asp75) which is responsible for the large negatively charged region highlighted.
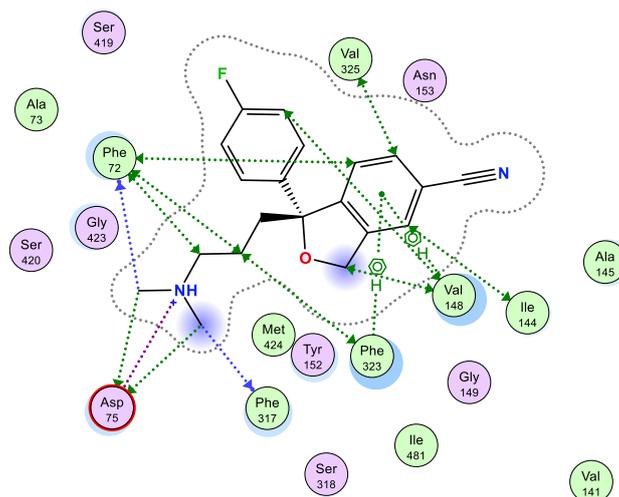
Figure 4. 11: A PLIF diagram to show citalopram bound to the NET binding site. The PLIF shows the ligand tightly bound in the active site which supports the PEOE descriptor being identified as important in the 4-descriptor NET QSAR model.

Given that there is little insight gained from analysis of the QSAR as to why the NET models were not generalizable, attention was turned to examining the dataset itself. The dataset could be the reason poor QSAR models were generated, as there is a narrow range of pKi values associated with NET which could the reason the model is not generalizable.

The QSAR model produced for NET could potentially be limited by the uneven distribution of pKi values which inherently restricted how well the test set could be represented by the training set. According to the Shapiro-Wilks test conducted on the pKi values for NET it showed that the distribution of data was not normal[229] and the majority of the compounds in the dataset (23/31) have a pKi between 5 and 7. This compression of the activity range may limit how well the training set could be defined and therefore how well the QSAR model is able to predict activity. This was identified at the onset, and in-depth analysis of the NET training set was conducted to try and produce a training set that accurately represented the dataset as a whole. Representative test and training sets were identified, however despite this the experiment was unsuccessful and it must be concluded that the most probable cause is the small dataset[119] with its limited range of pKi values.

4.6.1 Development of SERT QSAR models

SERT QSAR models were built using the Tanimoto-derived training set. Equations 4.5 show 3, 4 and 5-descriptor T_c-derived SERT QSAR models.

5-descriptor model for SERT (Tanimoto-derived training set)

pKi / SD (pKi) = 0.86008 -0.32863 * a_don / SD(a_don) +0.27612 * E_tor / SD(E_tor) +0.58373 * PEOE_VSA-0 / SD(PEOE_VSA-0) +0.12663 * BCUT_SLOGP_3 / SD(BCUT_SLOGP_3) +0.10300 * FASA+ / SD(FASA+)

4-descriptor model for SERT (Tanimoto-derived training set)

pKi / SD (pKi) = 2.10886 -0.31514 * a_don / SD(a_don) +0.25569 * E_tor / SD(E_tor) +0.57078 * PEOE_VSA-0 / SD(PEOE_VSA-0) +0.07779 * BCUT_SLOGP_3 / SD(BCUT_SLOGP_3)

3-descriptor model for SERT (Tanimoto-derived training set)

pK_i/ SD (pK_i) = 3.285274 – 0.36519* a_don/SD (a_don) + 0.25467 * E_tor/SD (E_tor) + 0.58228 * PEOE_VSA-0 / SD (PEOE_VSA-0)

Equations 4.5: 5, 4 and 3-descriptor QSAR estimated normalized linear models built using the SERT training set devised using diverse subset derived from T_c.

The performances of the T_c-derived SERT QSAR models in terms of their predictivity and robustness are summarised below (Table 4.10).

Table 4. 10: Summary of the r^2 and q^2 values for the three best-performing estimated normalized linear QSAR models (Equations 4.5) developed for SERT using a training set identified via the T_c-derived training set

SERT Model	SERT Tanimoto training set Results		SERT Tanimoto test set results
	r^2	q^2	r^2
5 - descriptor	0.88	0.82	0.73
4 - descriptor	0.86	0.79	0.70
3 - descriptor	0.87	0.80	0.74
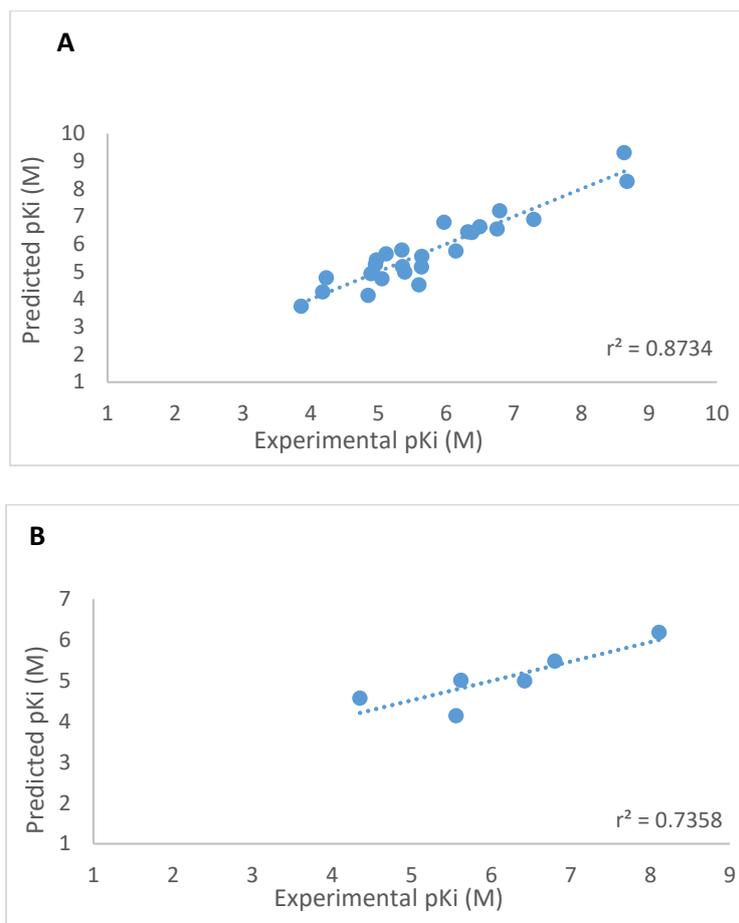
Figure 4. 12: Graphs representing the 3-descriptor model performance for the Tanimoto training and test set for SERT. (A) Plot of the predicted vs experimental values of pKi for the training set according to the QSAR model (r^2 =0.87), no outliers were detected (P = 0.01). (B) Plot of the predicted vs experimental values of pKi for the test set according to the QSAR model (r^2 =0.74), no outliers were detected (P = 0.01).

4.6.2 Evaluation and **Interpretation** of SERT QSAR models

The 5, 4 and 3-descriptor Tc-derived SERT QSAR models seemed to perform equally well in terms of their predictivity and robustness. The near identical values obtained for the training set results for all models indicates that the fourth and fifth descriptors identified by MOE contribute little/nothing to helping further explain the variance in the experimental pKis for the SERT training set, and hence should not be considered further. The remaining 3-descriptor model taken forward proved to be generalisable as evidenced by the high r^2 value obtained for the T_c-derived SERT test set. Table 4.11 gives the descriptors used in the T_c-derived 3-descriptor SERT QSAR model, in order of importance with respect to their overall contribution in explaining the variance in the T_c-derived SERT training set.

130

Table 4. 11: The molecular descriptors, relative importance and a description of what they represent, used in building the Tc-derived 3-descriptor SERT QSAR model.

Order of Importance	Molecular descriptor	Descriptor description
1.000000	PEOE_VSA-0	Total negative 0 van der Waals surface area
0.627169	a_don	Number of Hydrogen bond donor atoms
0.437357	E_tor	Torsion energy

PEOE_VSA-0 (described in section 4.3.6.1) is the most important descriptor identified for the SERT QSAR model. The positive coefficient associated with this descriptor indicates that compounds that have a large weakly negative/neutral VSA value will give rise to a greater pKi value. Again, as for DAT, this suggests, that largely hydrophobic interactions are going to be the most important factor for determining activity in the SERT binding cavity.

The descriptor a_don calculates the number of hydrogen bond donor atoms (not counting basic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH). The coefficient for a_don is negative which indicates that molecules containing fewer hydrogen bond donor atoms, excluding groups like protonated amines, a motif common across all of the NPS in the dataset, would result in higher calculated pKi values. The equation suggests the ideal is to have no hydrogen bond donor atoms present other than the protonated amines.

The information obtained from Chapter 3 regarding the SERT binding site has five residues with side chains able to act as hydrogen bond acceptors, (D98, Y175, Y176, S336 and T439, 5 out of 23 residues) although the interaction between the deprotonated Asp 98 residue and any hydrogen bond donor is going to be significantly stronger than the interactions between the HBD and the alcoholic groups on the tyrosine, serine and threonine sidechains. This is evident from the docking studies, which exclusively show high-ranked poses demonstrating interaction between Asp 98 and the protonated amine groups of the NPS.

The requirement for small molecules to have a low number of hydrogen bond donors (with the exeption of the biogenic amine) also fits with the first descriptor, PEOE_VSA-0, which shows, by inference, that hydrophobic molecules with a high degree of neutral/minimally negatively charged surface area will be favoured to bind in the SERT cavity via their interaction with the hydrophobic residues therein.

Examination of the dataset shows that a low a_don score is consistent with compounds that have higher pKi values (Figure 4.12) associated with SERT e.g. RTI-55 (pk$_i$ = 9.31), Citalopram (pk$_i$ = 8.27) and

Imipramine (pk$_i$ = 8.11) which all have a calculated a_don value of 0. Figure 4.13 shows the chemical structures of the compounds that have the highest 3 pk$_i$ values associated with SERT and the lowest 3 pk$_i$ values associated with SERT.

Citalopram
pK$_i$ = 8.27
a_don = 0

RTI-55
pK$_i$ = 9.31
a_don = 0

Imipramine
pKi = 8.11
a_don = 0

Methiopropamine
pK$_i$ = 4.14
a_don = 1

Desoxypipradrol
pK$_i$ = 4.27
a_don = 1

Buproprion
pK$_i$ = 4.35
a_don = 1
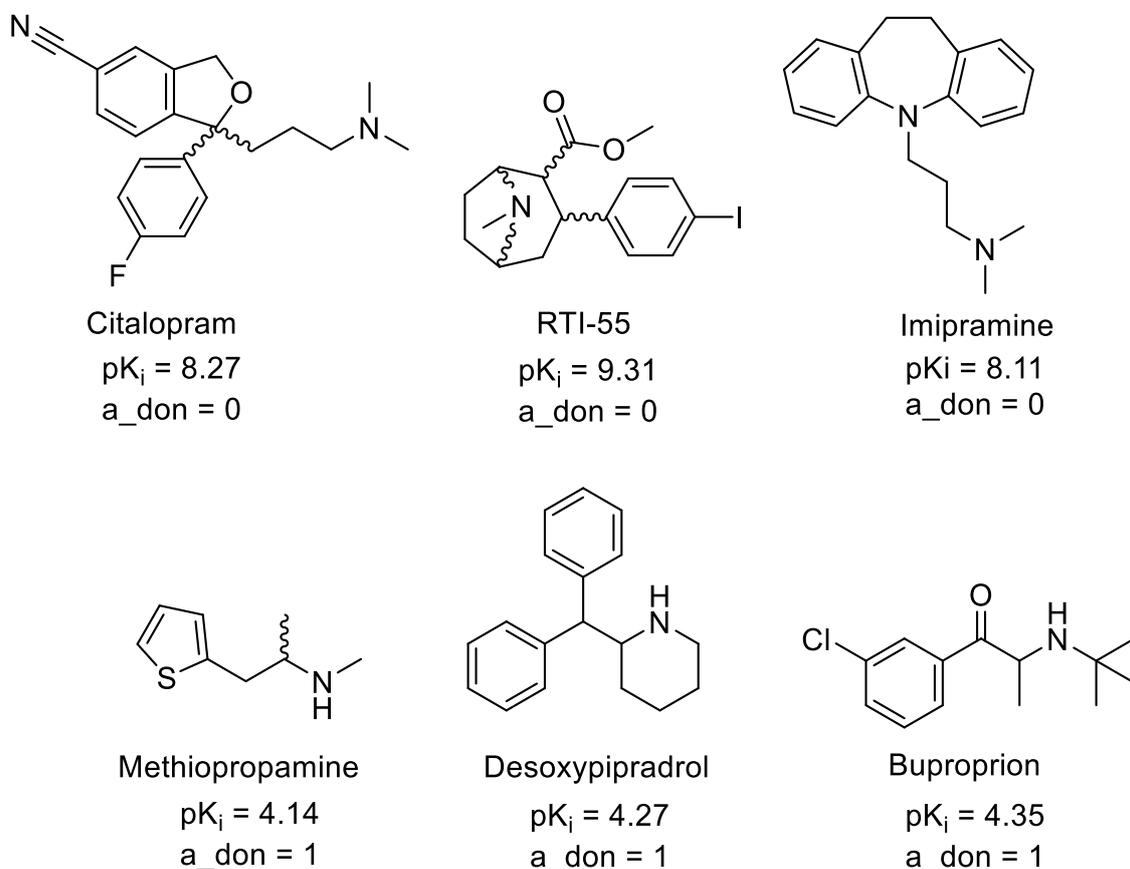
Figure 4. 13: chemical structure, associated SERT pK$_i$ values and a_don values for citalopram, RTI-55, Imipramine, Methiopropamine, Desoxypipradrol and Bupropion

E_tor are positive values that calculate the potential torsional energy of a compound. E_tor values are associated with the overall flexibility of a molecule and a more flexible molecule will have a smaller E_tor values. RTI-55 has the greatest SERT associated pKi value and also the largest E_tor value, in contrast Methiopropamine has a relatively low pKi and E_tor value which indicates that molecular rigidity plays an important role in affinity for the SERT binding site.

Methiopropamine
E_tor = 3.373
pKi= 4.14

RTI-55
E_tor = 18.379
pKi= 9.31
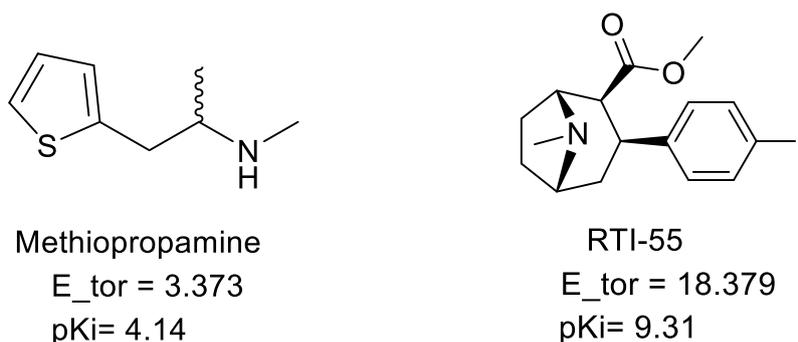
Figure 4. 14: the chemical structures, associate SERT pKi values and E_tor values for RTI-55 and Methiopropamine.

In order for a protein-ligand interaction to occur in SERT, more rigid molecules are favored. This is similar to DAT as SERT also appear to have a subtle narrowing of the binding site towards the bottom of the cavity. Figure 4.15 is a docked conformation of citalopram, the ligand fills the bottom of the cavity. From docking studies carried out in Chapter 3, ligands that are rigid are able to better access the the binding site and similar to DAT docking results rigid molecules may incur fewer entropic pentalties upon binding.
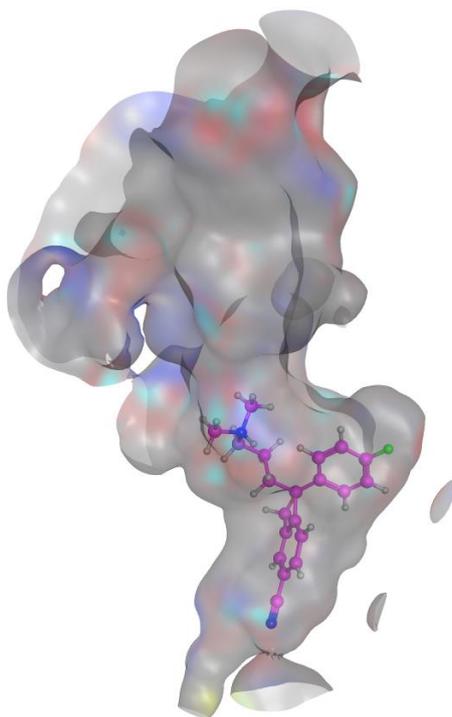
Figure 4. 15: the MOE elcuidated binding site for SERT (accession code 5I75) with a docked conformatio of citalopram (magenta stick model)

SERT produced the best QSAR models (highest r^2 = 0.87, q^2 = 0.80, using 3 descriptors) of the three MAT isoforms. This could be due to the fact that the SERT dataset had the broadest range of pKi values, which implicitly facilitates the construction of a more generalizable QSAR model[117]. In addition to this, 54% of the data set compounds had higher affinities with SERT in comparison to DAT and NET, with 20% of the compounds categorised as Selective Serotonin Reuptake Inhibitors (SSRI)[3535], therefore providing more information regarding the physicochemical differences that result in the different pKi values.

4.7 Conclusions

QSAR models have given insight into the properties required of NPS with respect to interaction with the individual MATs, DAT, NET and SERT. It was possible using molecular descriptors to highlight how subtle structural differences may enhance a compounds affinity for one monoamine transporter over another.

Robust and generalizable QSAR models were built for both DAT (training set $r^2 - 0.68$, test set q^2 0.42) and SERT (training set $r^2 - 0.87$, test set $q^2 - 0.80$), which could be used to predict the biological activity of new psychoactive substances. The best performing QSAR models were those associated with SERT, which is unsurprising given that the dataset for this isoform displayed the widest range of pKi values (7 log unit span).

From the QSAR models built, it can be seen that there are physicochemical differences that may convey a degree of selectivity between monoamine transporters. DAT selective compounds should be rigid, with shorter chains of single bonds and should contain have surface areas that are relatively hydrophobic to facilitate interaction between the small molecule and the DAT binding site.

Similarly, SERT selective compounds appear to favour inflexible, charged moieties and no hydrogen bond donor atoms present on the compound. These key findings should be considered when developing DAT or SERT selective compounds. The similarity of molecular descriptors for DAT and SERT may go to explain why there is promiscuity between the two MAT and the subtle differences in descriptors between DAT and SERT should be exploited further when attempting to design novel compounds that are MAT selective.

Issues arose when trying to identify different types of physicochemical features, (other than electronic) that correlate compounds with a large pKi value for NET. From the QSAR models produced it was evident that electronic descriptors were of most importance for NET, however the compression of pKi values could possibly be a limit in producing a generalizable model to be built for NET.

The results obtained from Chapter 3 and 4 have shown that the methods employed are applicable in identifying what physicochemical properties are required to convey selectivity between different MAT isoforms, and hence this gives confidence that NPS and their receptors can be exploited and interrogated successfully using computational methods. Therefore, the next stage of this research will focus on a particular category of NPS, the synthetic cannabinoids, and will use computational methodologies with the aim of identifying a novel library of synthetic cannabinoids that can be evaluated for biological activity using *in vitro* and *in vivo* techniques.

Chapter 5

Identification of Novel Compounds with Predicted Affinity for the Cannabinoid 1 (CB1) Receptor.

5.1 Introduction

The vast and rapid influx of NPS onto the market has exacerbated the overall detrimental effects of this group of drugs and it has become almost impossible to keep up-to-date records of known NPS. A review article from Drug Wise UK[230] states that the rapid discovery and generation of novel NPS may be due to the rapid growth of the internet and the easy access of disregarded patents from pharmaceutical companies[230].

Having demonstrated that it is possible to gain insight into the key physicochemical properties that are implicated in ligand interactions with the monoamine transporters, and hence demonstrating that *in Silico* methodologies can be applied to new psychoactive compounds, the next stage of this project was extended to investigate the NPS class of synthetic cannabinoids (SC). This class of NPS dominates the market[231] and there are recent notifications of SC overdoses[232–234] being reported post the introduction of the Psychoactive Substances act (2016). From the extensive published data on the number and type of synthetic cannabinoids, and the popularity amongst users for this illicit group of substances, it was decided to focus the next stage of the investigations on this particular class of NPS.

At the time of this study there were 256 SC listed by the UNODC[232], making SC the most predominant class of NPS globally. The arrival of Spice onto the market, an umbrella term for a substances containing a mix of shredded plant-derived and man-made hallucinogens, typically including a cocktail of SC, has presented a formidable public health challenge that affects a number of different societal groups including the homeless[235], the incarcerated[87] and adolescents[54]. Consequently, the SCs pose one of the biggest public health risks in terms of continued illicit drug use, and as such research into the next generation of chemical scaffolds with the potential to be exploited as SC is vital.

However, synthetic cannabinoids are by no means novel. The chemists Alexandros Makriyannis, Raphael Mechoulam and John W. Huffman have been investigating the medicinal properties of tetrahydrocannabinol (THC)-like compounds, the main psychoactive component of cannabis, for decades. The legitimate research conducted by Huffman, Makriyannis and Mechoulam gave rise to the JWH, AM and HU (respectively) synthetic cannabinoid series, which in recent years have been "hijacked" by the illegal drug trade community, mass produced and sold[236]. The compound JWH-018 was one of the first generation synthetic cannabinoids[237] reportedly abused, and approximately 130[219] novel SC molecules have been inspired by this compound. The rapid generation of new SC have become an urgent matter of public health concern, due to the lack of data available on the mechanism of action within the body[234].

As previously mentioned, one of the biggest concerns is around the abuse of Spice[86,232,234,238] also referred to as K2, which is sold as a herbal incense product. This product, which is actually a number

of independently-produced products each of which results in similar psychoactive effects, has been reported by the EMCDDA to be emerging at and unprecedented rate[85]. Although each sample of Spice is different in its exact composition, it is common for all Spice products to contain a number of different SC including JWH-018, JWH-073 and CP-47,497[239,240]. Numerous global reports of "zombie-like" behaviour[235,83] as a result of spice consumption have been published, and the devastating effects of these drugs reinforces the potential lethality of SC.

Therefore, the following piece of research aims to identify molecules that are as chemically distinct from known synthetic cannabinoids, but which align to pharmacophores that suggest they will bind to the cannabinoid receptor 1 (CB1). This is in an attempt to identify completely novel ligands that *in Silico* modelling suggests have the potential to act as SC.

To this end, a library of 162 known synthetic cannabinoids were used in a ligand-based drug design approach to develop pharmacophore models that were then used to identify novel compounds with potential to bind to the CB1 receptor (and hence give rise to psychoactivity). This database of 162 compounds contains analogues from the well-established JWH, AM and HU SC series as well as other previously identified SC compounds that have been reported as being abused.

This chapter will detail how the development of four pharmacophore models based on the structures of known SCs were used to virtually screen a molecular database consisting of approximately 17 million compounds to produce a feasible library of molecules that had the potential to interact with the CB1 receptor. The compounds that were identified as potential CB1 binders were specifically selected so as to be structurally dissimilar to the known SC but with potential to elicit similar pharmacological effects

Ultimately, thirteen molecules were identified as potential CB1 agonists based on a number of criteria including likely propensity for binding to the CB1 receptor, likelihood of the compound crossing the blood brain barrier, solubility and availability for commercial purchase.

5.2 Method

Figure 5.1 gives an overview of the methodology used to identify novel compounds that had potential to interact with the CB1 receptor.
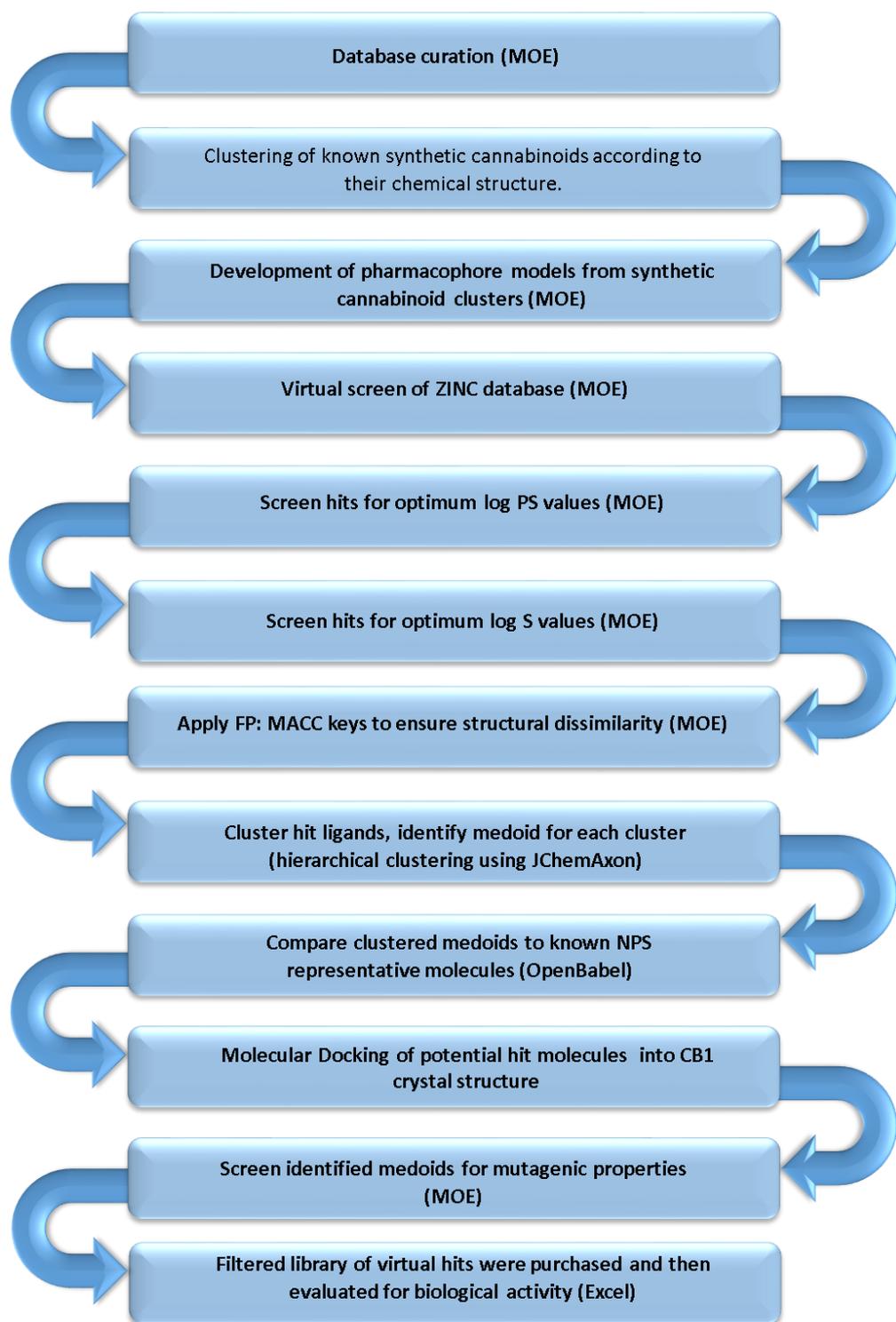
Figure 5. 1: The general strategy employed in this chapter in the identification of compounds that were chemically distinct from known SC molecules but were predicted to have the potential to bind to the CB1 receptor.

5.2.1   Database curation

A dataset of 162 SC was compiled into a Molecular Database (MDB) using the MOE software. The molecules in the database were converted into SMILES strings and screened for duplicates. These compounds were taken from the European Database of New Drugs (EDND)[241]. A second database of 655 known Novel Psychoactive Substances (NPS) (including the 162 SC taken from the EDND[241]) was collated and used in the development and optimisation of pharmacophore models.

5.2.2 Clustering systems

Initially, the SC were clustered using the Maximum Common Substructure (MCS) application in ChemAxon[158]. The cluster default parameters were used i.e. normal mode and the 'jcsearch' algorithm with the default settings (MCS mode = fast, minimal MCS size = 9, matching atom types = true, bond type = true, charge = true, keep rings = true, required cluster count = 1, maximal level count = 10)

In addition, a second clustering system was employed using core chemical scaffolds as the criteria for clustering. For this approach, the scaffold classification system described by Uchiyama *et al*[242] was used and the synthetic cannabinoids were grouped into 15 categories (benzoylindoles, carboxyamide derivatives, carboxyindoles, classical cannabinoids, cyclopropylindoles, cyclohexylphenols, naphthoylbenzimidazoles, naphthoylnaphthalenes, naphthoylindazoles, napthoylindoles, naphthoylpyrroles, pentylindoles, phenylacetylindoles, others and quinolinyl carboxylates).

5.2.3 Pharmacophore development

The pharmacophore models were built from clusters that contained 16 or more molecules. Although there is no direct rule for the minimum number of compounds, a leading computational software manufacturer (Schrödinger[243]) suggest that no less than 16 ligands should be used in the development of pharmacophore models

All the compounds in an individual cluster were flexibly aligned with respect to one another using the flexible alignment module in MOE. Default settings were applied with respect to the alignment i.e. Iteration limit of 200 and Failure limit of 20. Flexibly aligned motifs were generated and assigned a value. The value corresponds to a score that quantifies the quality of the alignment for both internal strain and overlap of molecular features with the structural features present on the aligned motif. The lower the score, the better aligned all the molecules are.

Using the top ranked/lowest scored flexibly aligned motif for each group of SCs a set of consensus features were assigned to the alignment to produce an initial pharmacophore model. The initial model

was used to search the 655 compound NPS database and "hit" ligands were identified. A hit ligand is described as compound that has defined features that achieve a minimum "match" size with the consensus pharmacophore (two features need to be matched). A buffer of 0.5 Å from the centre of a pharmacophoric feature is applied for the direct overlay of each feature and the hit ligands were ranked by lowest RMSD values, which is calculated between the query features and their matching ligand annotation points.

The initial consensus pharmacophore models generated were validated by examining how well the model was able to identify SC from the same class used to develop the pharmacophore in the early stages (i.e. first ten percent of compounds in the database) in a virtual screen search, using the 655 NPS database established previously.

The top 10% of hits from the virtual screen, ranked according to how well they fit the pharmacophore model, were used to calculate an enrichment factor - a metric that is commonly used to quantify the performance of pharmacophore models.

Enrichment factors (E_f – Equation 5.1) were calculated at 2, 5 and 10% of the NPS database search to establish how many SCs from the class of compounds used to generate the pharmacophore the model was able to retrieve at those intervals.

Enrichment factor $_{10\%}$ = <u>(% of relevant SC $_{at\ 10\%}$ ÷ % of NPS database $_{at\ 10\%}$)</u>

Number of SC ÷ number of NPS

Equation 5.1: The equation[161] used to determine the enrichment rate for first 10% of the NPS database screened.

The pharmacophore model was then interrogated using a Leave One Out (LOO) approach, whereby a consensus feature was removed and the modified pharmacophore model was used search the 655 compound NPS database. Enrichment factors at 2,5 and 10% were calculated for the modified pharmacophore, and compared to the initial model to see how the enrichment rate was affected by the removal of a particular feature. This iterative process was carried out until the highest enrichment rate was obtained using a model with the fewest possible features.

To give an example, a generalisable pharmacophore model that is selective for a given class of compounds (e.g. the carboxyamides) will contain a minimum number of pharmacophoric features and highly rank molecules that display the pharmacophoric features of the carboxyamides during the virtual screening process. This will result in the proportion of the molecules from the carboxyamide cluster found in the first 10% of the virtual screen (i.e. the first 10% of the database ranked according

to how well it fits the pharmacophore model) being significantly greater than random and displaying an enrichment factor greater than 1 at 2, 5 and 10% of the database screened.

5.2.4 Medoid selection

A medoid was selected for each of the fifteen SC clusters identified. The medoid was obtained by creating a correlation matrix of pairwise Tanimoto coefficients (T_c) for each molecule in the cluster. The sum of T_c for each cluster member was calculated and the molecule that had the greatest T_c total value was selected as the medoid for that cluster.

5.2.5 Database selection

The standard drug-like library containing 17,900,742 compounds (at the time of access, November 2014) was downloaded from the ZINC database[244]. The standard drug-like subset from this database was used in the virtual screening experiments that follow.

5.2.6 Virtual screen of database using pharmacophore model

The best-performing pharmacophore models generated for each of the classes of SC investigated were used to search the drug-like subset from the ZINC database. The Pharmacophore search application in MOE[98] was employed (using the same parameters stated in section 5.2.3).

The initial library of virtual hit molecules retrieved was refined to ensure the hits had similar physicochemical properties to known synthetic cannabinoids, and by inference had good potential to bind to the CB1 receptor *in vitro/in vivo*. This was achieved by applying filters for physicochemical properties that would remove molecules that did not align within the range of those properties calculated for known SC structures.

5.2.7 Log PS

The first filter applied was the molecular descriptor log PS (Equation 5.2) which was calculated as follows:

log PS = - 2.19 * 0.262 * log D + 0.00683 * vsa_base − 0.009*TPSA[245]

Equation 5.2: log PS is the log of the permeability-surface area coefficient[246], log D is the solubility (at pH 7.4) calculated by ChemAxon, vsa_base is the van der Waals' surface area due to basic atoms, and TPSA is the topological polar surface area. The latter two descriptors were calculated by MOE[98].

The log PS values were calculated for 188 known cannabinoids and synthetic cannabinoids i.e. the 162 SC identified from the EDND report and 26 endogenous cannabinoids (e.g. cannabidiol, anandamide & 2-arachiodonoylglycerol) identified from the literature[92,247]. The range of log PS values based on these 188 cannabinoids was recorded.

5.2.8 Log S

Log S values were calculated in MOE using the molecular descriptor application for the 188 cannabinoids/synthetic cannabinoid ligands. The range of log S values identified was recorded.

5.2.9 Chemical similarity

The virtual hit molecules that remained after the first two filters had been applied were assessed for chemical structure similarity to the SC medoids associated with the pharmacophore model used for the virtual screen. Two different similarity metrics were used, the first process employed was MACCS structural keys.

The virtual hits retrieved for a specific pharmacophore were assigned MACCS structural keys[248] which were compared to the MACCS structural key of the medoid for the SC cluster used to generate the pharmacophore. A cut off value of 25% similarity was applied and any virtual hit that exceeded this threshold was discarded. This method was employed as calculating pairwise T_c values for large databases (between 4000-100,000 molecules) is time consuming. Virtual hits that were less than 25% similar to the cluster medoid were then analysed using a second similarity metric. Pairwise T_c values were calculated for the virtual hits based on the SC medoid, any ligand that had a calculated pairwise T_c of greater than 0.6 was removed.

5.2.10 Clustering & Cluster Medoid Selection

The remaining hits from the virtual screen were clustered in ChemAxon JKlustor Library MCS (see section 5.2.2 for parameters). A representative molecule (medoid[158]) was identified as described previously and taken forward for each of the clusters identified. Single ligands that were not part of a larger cluster (i.e. singletons) were treated as medoids in their own right. This step was conducted to ensure that the minimum number of the most structurally diverse compounds (termed virtual hit representatives) were carried forwards into the next stage of the experiment.

5.2.11 Comparison to NPS representatives

The virtual hit representatives identified were interrogated to ensure they were structurally distinct from known SC. Any virtual hit representative that had a pairwise T_c of greater than 0.6 with one of the fifteen SC medoids previously identified was removed.

The virtual hit representatives were also screened for any similarities towards other classes of known NPS. This was carried out to ensure the representative molecules carried forwards were novel with respect to their chemical scaffolds. Pairwise T_c were calculated, for the virtual hit representative molecules and 11 NPS representative ligands taken from a study conducted by Zloh *et al*[249]. Once again, any compound that had a pairwise Tc of greater than 0.6 when compared to an NPS medoid was removed.

Finally, pairwise T_c values were calculated for the virtual hit representative molecules and the initial 162 SC included the molecules that were not included in developing the pharmacophore models. Again, any which demonstrated pairwise Tc values of greater than 0.6 were removed from the study.

5.2.12 Molecular Docking of potential hit molecules

The remaining virtual hits were docked into the CB1 crystal structure (PDB accession code 5GTZ[250]). The structure was prepared using the "quickprep" application in MOE. Using the site finder application in MOE, a binding site was elucidated based on residues implicated in protein-ligand interaction at the CB1 binding site. The residues in the binding cavity include Met103, Gly166, Ser167, Phe170, Val196, Trp356, Phe379, Ser383, Cys386 and Leu387[97] . Each ligand was docked into the binding site with a cut-off of 15 conformational poses. The default scoring functions (London dG and GBVI/WSA dG) in MOE were applied.

In addition, the SC medoid ligands were docked into the CB1 receptor. As each of the SC medoids is known to elicit psychoactivity, the data obtained from this docking study can be used as a benchmark for the virtual screening compounds docking results. The virtual screening hits were ranked in order of their best-performing docking score. Virtual screening compounds that had a score of less than -6.55 kcalmol^{-1} were discarded. This cut-off value was selected as none of the benchmark medoid SC compounds had a docking score of less than -6.55 kcalmol^{-1}.

5.2.13 Final selection of possible synthetic cannabinoid candidates

The resultant library of virtual hits was assessed for possible mutagenic properties. Potentially toxic groups were assigned using the Kazius set[251]. Any molecules that had toxic groups were discarded.

Remaining molecules were then purchased and taken forward for *in vitro/ex vivo* screening to assess their potential to bind to the CB1 receptor.

5.3 Results and Discussion

5.3.1 Database collation

The synthetic cannabinoid ligands were collated using databases from the European Database of New Drugs (EDND), EMCDDA[252,253] and the UNODC[254].

5.3.2 Clustering

A total of 162 SC were available at the time of this study.

The initial clustering approach using JChemAxon produced 54 clusters from the 162 molecules. The maximum number of ligands in any of the clusters generated was four. A pharmacophore model generated using a small number of structurally similar ligands will produce a highly selective model, which would be of little use in identifying potential new scaffolds for exploitation. As such, basing a pharmacophore on any of the JChemAxon clusters identified is likely to limit the number of structurally diverse "hits" being retrieved.

Therefore, an alternative, manual, clustering approach was employed based on common chemical scaffolds observed for the SC.

The EMCDDA classifies SC into seven categories and the UNODC uses six categories (see table 5.1)

Table 5. 1: The major structural synthetic cannabinoid groups according to the EMCDDA[255]and the UNODC[256].

Seven structural SC groups according to the EMCCDA	six structural SC groups according to the UNODC
Naphthoylindoles	Classical cannabinoids
Naphthylmethylindoles	Nonclassical cannabinoids
Naphthoylpyrroles	Hybrid cannabinoids
Naphthylmethylindenes	Aminoalkylindoles
Phenylacetylindoles	Eicosanoids
Cyclohexylphenols	Others
Classical cannabinoids	

The two classification systems described in Table 5.3 could not be applied to the 162 SC used in this study as many of the compounds could not be assigned to one of these categories. A finer-grained approach was required to identify more specific structural classifications in order to produce predictive pharmacophores, so the SC were grouped using the clusters described by Uchiyama *et al*[13] (see Table 5.2)*.

Table 5. 2: The 15 different categories of synthetic cannabinoids as defined by Uchiyama et al[242] and the number of molecules within each cluster.

Name of common structural scaffold	Number of molecules in cluster
Benzoylindoles	6
Carboxyamide derivatives	54
Carboxyindoles	18
Classical cannabinoids	3
Cyclopropylindoles	10
Cyclohexylphenols	3
Naphthoylbenzimidazoles	2
Naphthoylnaphthalenes	2
Naphthoylindazoles	2
Naphthoylindoles	48
Naphthoylpyrrols	1
Pentylindoles	2
Phenylacetylindoles	1
Others	8
Quinolinyl carboxylates	2

Three SC clusters identified from the Uchiyama classification had at least 16 ligands (carboxyamide derivatives, carboxyindoles and naphthoylindoles). These were taken forward into pharmacophore development studies.

5.3.3 Pharmacophore development

The flexibly aligned molecules from the three SC clusters were used to suggest a consensus set of pharmacophore features. These are formed based on annotation points assigned on the aligned atoms or centroid annotations which are located at geometric centric of a subset of aligned atoms. The annotation points also describe the atom types in terms of aromatic, hydrophobic, hydrogen bond

donor atoms or hydrogen bond acceptor atoms. The features form spheres of different radii which are determined by the aligned molecules and it is the different combination of feature types and spatial arrangement of each sphere that give rise to a unique pharmacophore model. The individual models were then used to screen the ZINC drug-like subset to match any potential hit ligands.

All pharmacophore models were developed and optimised using the same Leave One Out (LOO) method. Enrichment plots were used to show how the removal of a feature impacts on how efficiently the model is able to retrieve expected hit ligands (i.e. molecules belonging to the same class of compounds used to derive the pharmacophore). The performance of each model was assessed by plotting the number of hits retrieved at 2, 5 and 10% of the total database scanned. The virtual hits identified by each of the pharmacophore models were ranked by RMSD values, whereby a SC-specific pharmacophore model would match the majority, if not all, of the expected (specific) SC hits first before matching with any other ligands within the NPS database.

The initial pharmacophore models for the three different SC classes generated without modification was used as a comparison for all subsequent pharmacophores in the enrichment plots. Additionally, each plot includes a random line. This line is used to assess whether the pharmacophore models are able to identify expected hit compounds better than random.

5.3.3.1 Carboxyamide derivatives (CD) pharmacophore

The CD pharmacophore model generated from the aligned data based on the 54 CD molecules, produced a seven-feature pharmacophore (see Figure 5.2) which, when tested, retrieved only three of the possible 54 CD molecules from the entire 655 ligand NPS database.
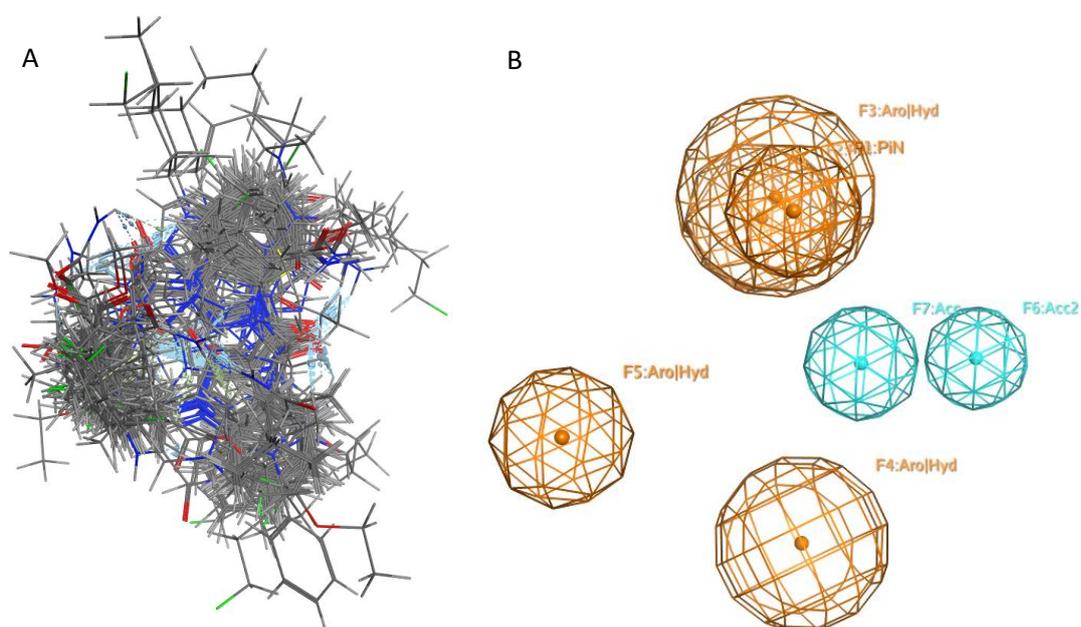
Figure 5. 2: (**A**) a diagram showing all 54 carboxyamide derived synthetic cannabinoids flexibly aligned and superimposed with respect to one another and (**B**) The seven-feature pharmacophore model generated from the flexible alignment of the carboxyamide derivative ligands. The Aro|Hyd (3, 4 and 5 orange) features represent aromatic/hydrophobic moieties, the PiN (1 and 2 orange features) represent features located along implicit lone pair of electrons or implicit hydrogen associated with ring structures and the cyan features represent hydrogen bond acceptors (6 and 7 cyan features).

From the enrichment plot (Figure 5.3) it can be seen that pharmacophore model generated based on all the carboxyamide derivatives was unsuccessful at retrieving expected hits from the NPS dataset. This indicated that despite being constructed using known carboxyamide compounds, the pharmacophore wasn't discriminative for carboxyamides. An Investigation into why this might have arisen was carried out.
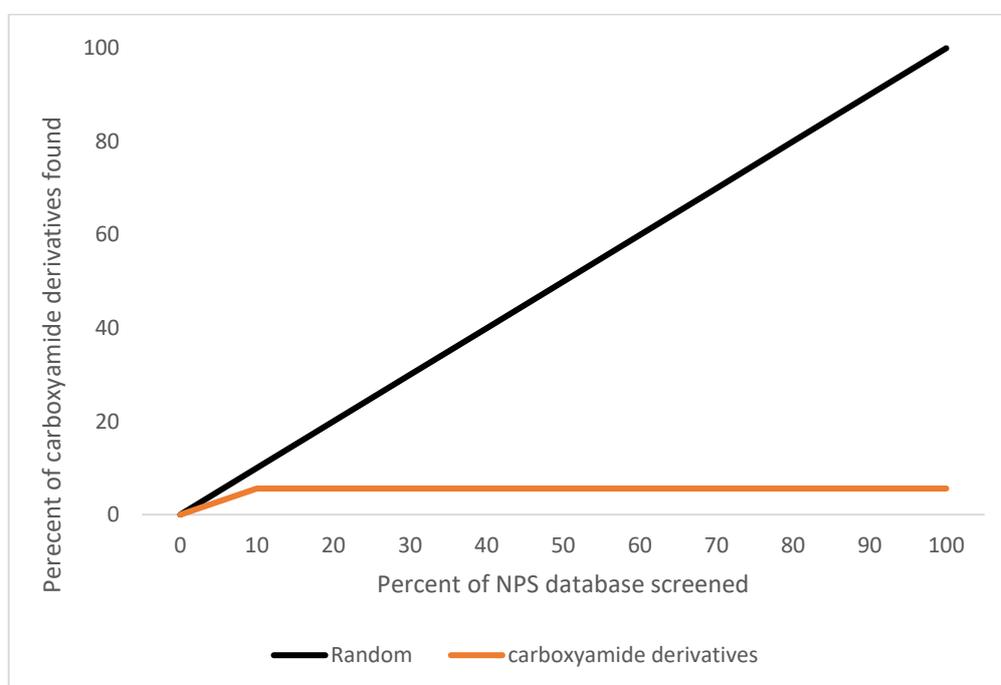
Figure 5. 3: An enrichment plot to show how well the carboxyamide derivative pharmacophore model for the carboxyamide derivatives (orange line) was able to identify the carboxyamide SC in the 655 compound NPS database. The black line indicates how a pharmacophore model would perform if carboxyamide SC were matched at random. The graph shows the pharmacophore is unable to identify carboxyamide SCs from the NPS database.

5.3.3.1.1 Clustering of carboxyamide derivatives

When the original 7-feature pharmacophore model for carboxyamide derivatives was used to screen the 655 compound NPS database, only 3 of the expected 54 compounds were identified.

Investigations showed that the carboxyamide derivatives, arbitrarily categorised by the scaffold-based categorisation of Uchiyama *et al,* were comprised of two structurally distinct groups based on the pairwise T_c calculated. This inherent diversity in the group could therefore be the reason behind the initial pharmacophore being non-discriminant, and as such the group was divided into two separate clusters based on the T_c values Cluster 1 (1-ethyl-1*H*-indazole-3-carbaldehydes) and Cluster 2 (2-ethyl - 3, 4-dihydro - 2*H* – pyrrole - 5 -carboxamides) (Figure 5.4). Each new cluster was flexibly aligned and new pharmacophore models were generated

Table 5.3 details the compounds that make up Cluster 1 and Cluster 2. Flexible alignment scores for the two new clusters (Table 5.4) are lower in value than the initial flexible alignment of the 54 compounds. This suggests a greater structural similarity, and hence better consensus, in the two new carboxyamide clusters when compared to the original alignment.
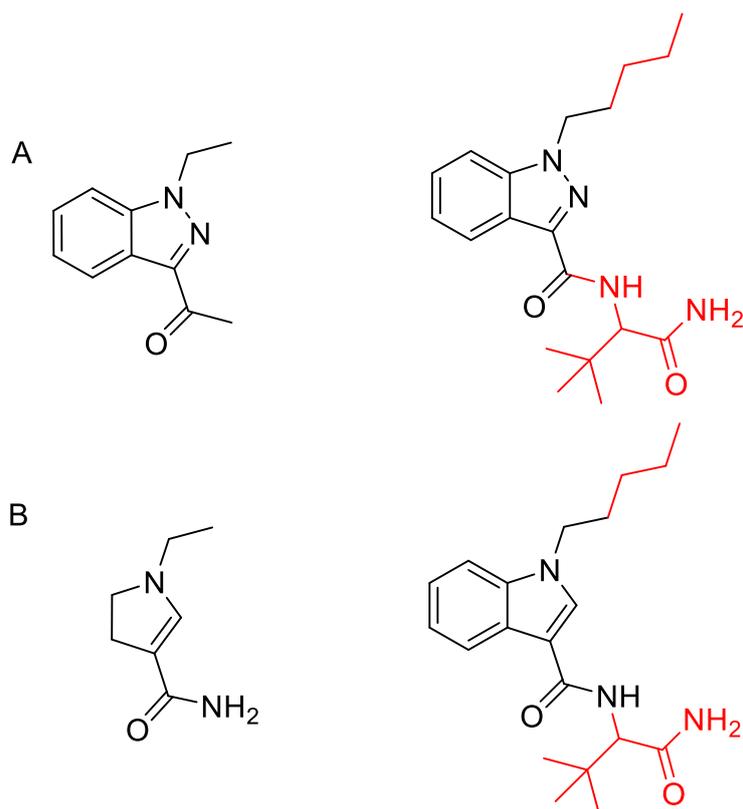
Figure 5. 4: Molecule A is the common substructure found in Cluster 1 (1-ethyl-1*H*-indazole-3-carbaldehyde), which contained 29 carboxyamide derivatives. Molecule B is the common substructure found in Cluster 2 (2-ethyl - 3, 4-dihydro - 2*H* − pyrrole - 5 -carboxamide) which contained 16 carboxyamide derivative ligands. An example of a synthetic cannabinoid with cluster A and B has been shown. Sections of the molecule that are different to the cluster have been coloured red.

The remaining 10 ligands identified from the initial carboxyamide group formed singletons and were not included in the generation of new pharmacophore models.

Table 5. 3: The names of the SC carboxyamide derivatives that were categorised into Cluster 1 and Cluster 2.

Carboxyamide derivatives contained in Cluster 1		Carboxyamide derivatives contained in Cluster 2
5C-AKB48	ADB-FUBINACA	5F-ADBICA
5F-AB-PINACA	ADB-PINACA	5F-AMBICA
5F-ADB-PINACA	AKB-57	5F-AMB-PICA - MMB-2201
5F-AKB48	AMB-FUBINACA	5F-APP-PICA - PX-1
5F-AMB / 5F-AMB-PINACA	Apinaca /AKB-47	5F-PY-PICA
5F-APP-PINACA - PX-2	APP-CHMINACA / PX-3	ADAMANTYL-THPINACA
5F-EMB-PINACA - 5F-AEB	APP-FUBINACA	ADB-CHMICA
5F-MDMB-PINACA - 5F-ADB	CUMYL-5FPINACA	ADBICA
5F-PY-PINACA	CUMYL-PINACA	AMB-CHMICA
AB-CHMINACA	CUMYL-THPINACA - SGT-42	CUMYL-5F-P7AICA
AB-FUBINACA	EMB-FUBINACA	CUMYL-5FPICA
AB-FUBINACA 2-fluorobenzyl isomer	FUB-AKB48	CUMYL-BICA
AB-PINACA	MDMB-FUBINACA	CUMYL-PICA
AB-PINACA N-(2-fluoropentyl) isomer		MDMB-CHMICA
ADAMANTYL-THPINACA		MDMB-FUBICA
ADB-CHMINACA		STS-135

Table 5. 4: The alignment scores for the 54 CD and the two identified CD clusters (C1 and C2), more negative values indicate a better flexible ligand alignment.

	All Carboxyamide derivatives	Carboxyamide derivatives cluster 1	Carboxyamide derivatives cluster 2
Scored value	-39.851	-58.446	-82.197

5.3.3.1.2 Pharmacophore development of 1-ethyl-1H-indazole-3-carbaldehyde carboxyamide derivatives (Cluster 1_Carboxyamide).

Alignment of the 29 ligands from Cluster 1 yielded a pharmacophore model with 6 features. This model identified 46 CD hits from the NPS database when used as a filter in a virtual screening experiment. All 29 of the carboxyamide derivatives used to generate the pharmacophore were matched by the pharmacophore when 100% of the NPS database had been screened.
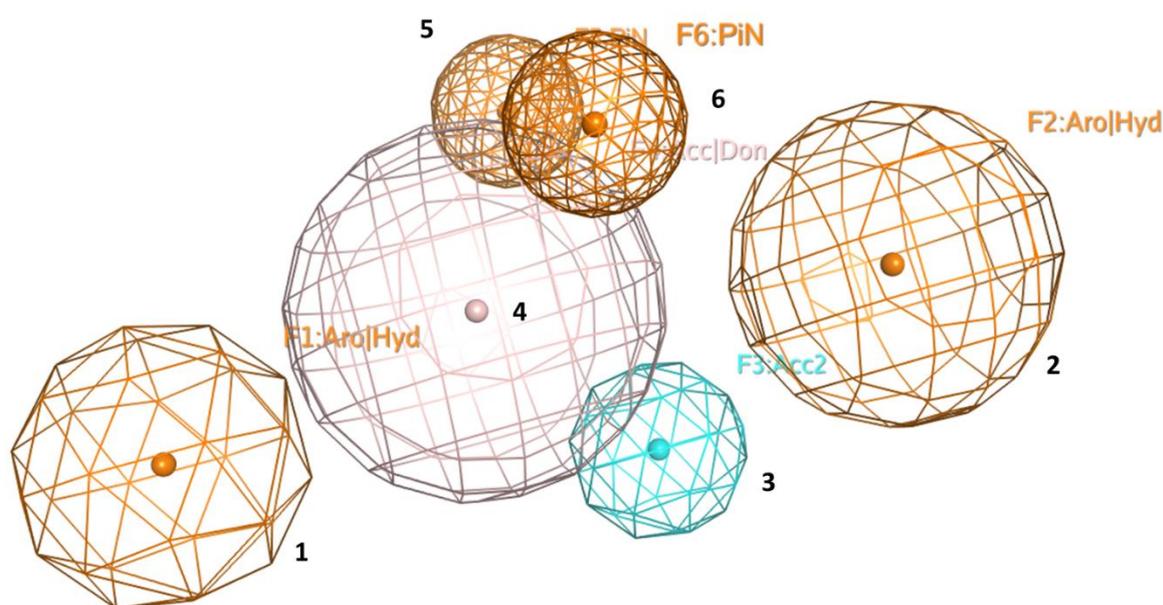
Figure 5. 5: The pharmacophore model based on the flexible alignment of the 29 carboxyamide derivative ligands from Cluster 1 (Cluster1_Carboxyamide). The Aro|Hyd (1 and 2, orange) features represent aromatic/hydrophobic moieties, the PiN (5 and 6 orange) represent features located along implicit lone pair of electrons or implicit hydrogen associated with ring structures, the light pink feature (4) represents a hydrogen bond acceptor/donor region and the cyan feature (3) represents a hydrogen bond acceptor feature.

An iterative LOO process was then carried out. This process involved the removal of a one feature at a time from the Cluster 1 pharmacophore shown in Figure 5.5. The modified pharmacophore models were then used to re-screen the NPS database to see how many of the carboxyamide derivative synthetic cannabinoids are retrieved at various stages of the virtual screen. Where the search performed better than random, an E_f was calculated at 2, 5 and 10% of the search (Table 5.5).

This process was repeated until a pharmacophore was generated that contained the fewest number of features and retrieved the highest percentage of the expected synthetic cannabinoids within the first 10% of the NPS database virtual screen. Figure 5.6 is an enrichment plot of the first LOO experiment for the Cluster 1 pharmacophore. From the plot, it can be seen that removal of any of the features reduced the number of carboxyamide Cluster 1 compounds the pharmacophore model was able to detect from the total NPS database at 2, 5 and 10%, indicating that the original pharmacophore generated was the best-performing.
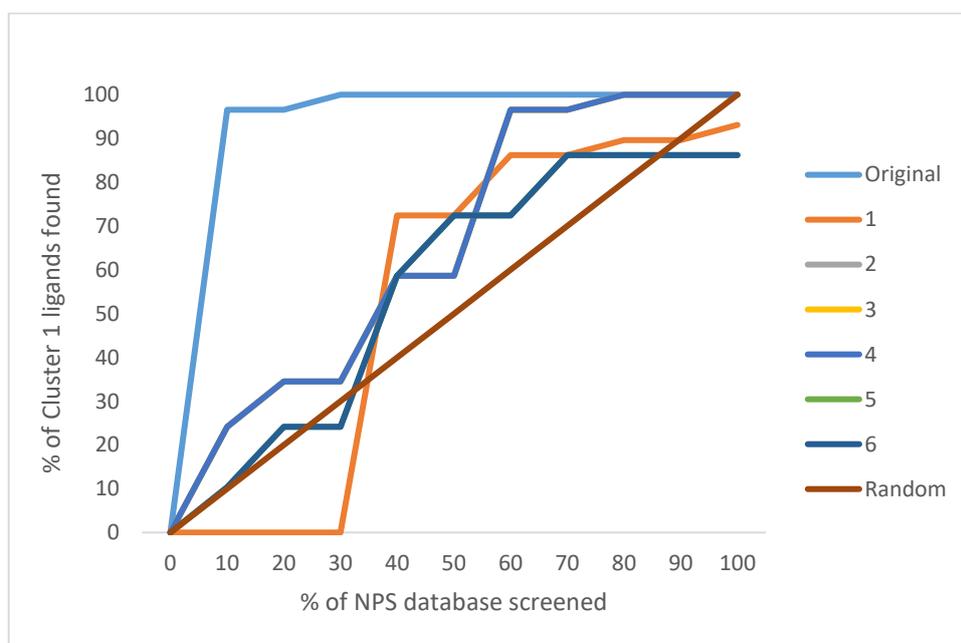
Figure 5. 6: Enrichment plot of the Carboxyamide cluster 1 pharmacophore models, the numbered lines correspond to the feature that was removed from the original pharmacophore and how this affected the model's ability to match with Cluster 1 ligands. The graph also includes a brown line that represents a random hit rate.

Table 5. 5: The calculated enrichment factor (E_f) for the Carboxyamide cluster 1 pharmacophore model. The table shows E_f for the original consensus pharmacophore. The E_f is given at different stages of the virtual screen (i.e. at 2, 5 and 10% of molecules considered).

Pharmacophore	Enrichment factors		
	at 2%	at 5%	at 10%
original	12.1	9.6	8.3
1	0.0	0.0	0.0
2	5.2	2.1	2.4
3	5.2	2.1	2.4
4	5.2	2.1	2.4
5	5.2	2.1	2.4
6	5.2	2.1	2.4

The enrichment plot in Figure 5.6 and the E_f value in Table 5.5 shows the initial carboxyamide cluster 1 pharmacophore model as the best at retrieving Cluster 1 ligands in the early stages of a virtual screening search. The highest E_f is obtained when 2% of the NPS database has been searched where the E_f value of 12.1 indicates that the original pharmacophore model is 12.1 times better at selectively identifying Cluster 1 ligands in comparison to a model finding them at random. As such the initial six - feature carboxyamide pharmacophore model was taken forward to use to identify potential novel SC scaffolds in the virtual screen of the ZINC database.

5.3.3.1.3 Pharmacophore development of 2-ethyl - 3, 4-dihydro - 2*H* − pyrrole - 5 - carboxamide carboxyamide derivatives (Cluster 2).

The aligned Cluster 2 carboxyamide ligands generated a six-feature pharmacophore model (see figure 5.7)
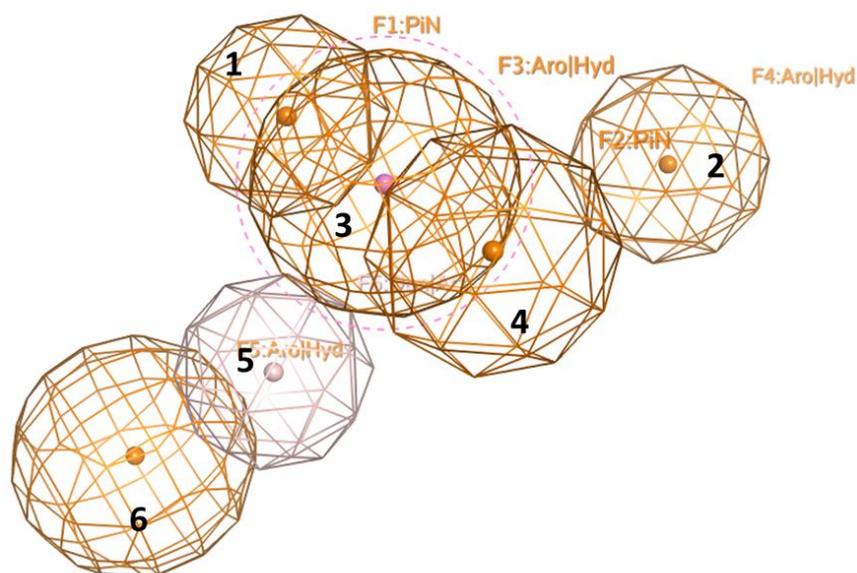
Figure 5. 7: The pharmacophore model with based on the flexible alignment of the 16 carboxyamide derivative ligands from Cluster 2. The Aro|Hyd (2, 3, 4 and 6, orange) features represent aromatic/hydrophobic moieties, the PiN (1, orange) represent a feature located along implicit lone pair of electrons or an implicit hydrogen associated with ring structures and the light pink feature (5) represents a hydrogen bond acceptor/donor region.

Figure 5.8 shows the enrichment plot obtained using the initial carboxyamide_cluster2 pharmacophore model, and the iterative LOO pharmacophore models, whereby one feature was sequentially removed, from the initial model and the resultant modified pharmacophore models used to search the 655 compound NPS database, to identify which model performed the best at identifying carboxyamides.
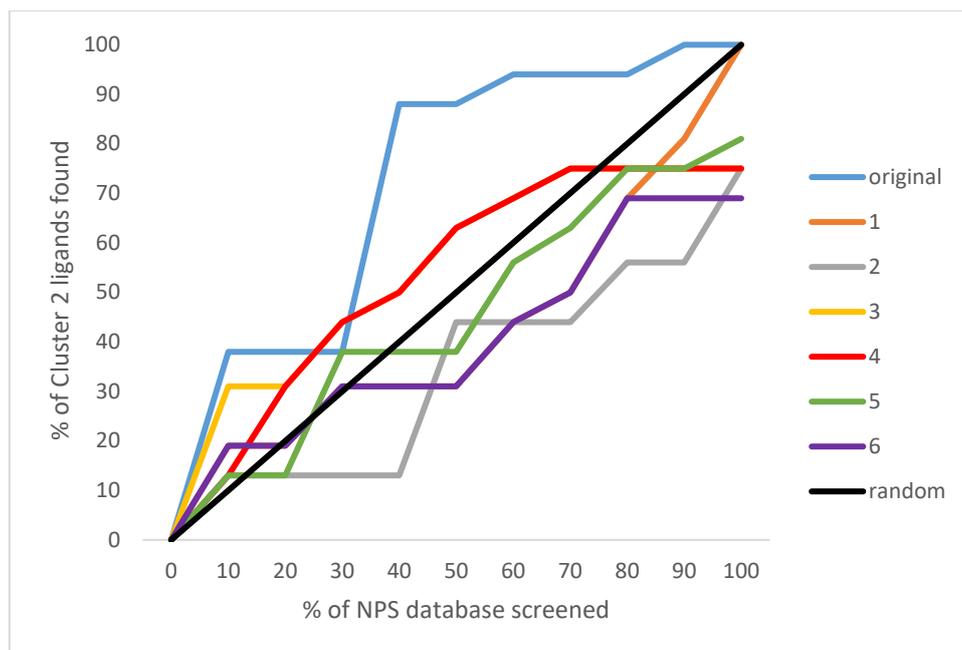
Figure 5. 8: Enrichment plot obtained from the carboxyamide cluster 2 pharmacophore searches. The lines show what percentage of Cluster 2 ligands have been retrieved from the NPS database as a function of the total database searched. Each coloured line represents a different model. The light blue line represents the original consensus pharmacophore model. The numbered lines represent the results of a search after the removal of the numbered pharmacophore feature (see Figure 5.7). The graph also includes a black line that represents a random hit rate.

The enrichment plot and enrichment factors calculated (see Table 5.6) indicate that the initial 6-feature pharmacophore model is the most successful model with respect to retrieving the greatest number of carboxyamide cluster 2 ligands within the first 10% of the NPS database searched. As such this model was taken forward and used to screen the ZINC database for potential novel SC scaffolds.

Table 5. 6: The calculated enrichment factors (E$_f$) for the Cluster 2 pharmacophore model, the table shows E$_f$ for the original consensus pharmacophore and E$_f$ when a feature has been removed. The E$_f$ are given at different percentages which indicate how many carboxyamide derivatives had been found at 2, 5 and 10% of the NPS database search.

	Enrichment factors		
	at 2%	at 5%	at 10%
original	3.1	3.8	3.8
feature 1 removed	0	0	1.9
feature 2 removed	0	1.3	1.3
feature 3 removed	0	1.3	2.5
feature 4 removed	0	0	1.3
feature 5 removed	0	0	1.3
feature 6 removed	0	1.3	1.9

5.4 Carboxyindole pharmacophore

The carboxyindoles group identified by the Uchiyama[257] strategy for clustering the SCs contained 18 molecules. The initial carboxyindole pharmacophore model generated consisted of 9 features (see Figure 5.9).
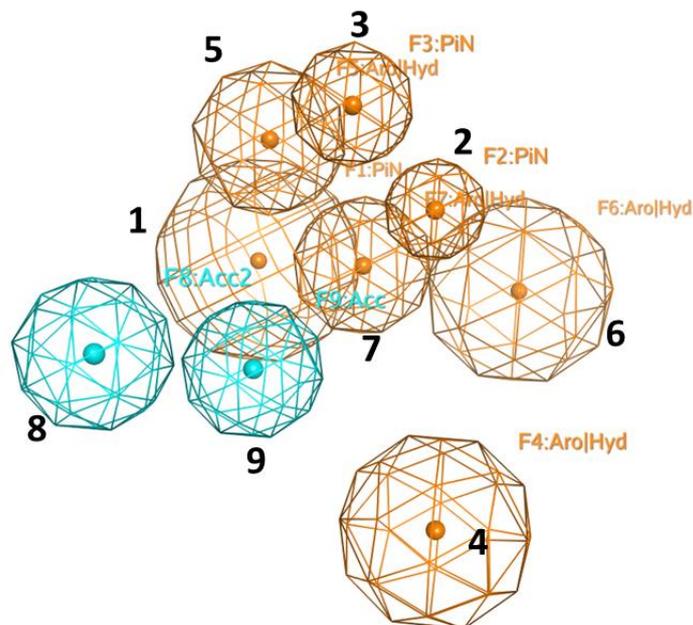
Figure 5. 9: The consensus pharmacophore model based on the flexible alignment of the 18 carboxyindole ligands. The Aro|Hyd (4, 5, 6 and 7 orange) features represent aromatic/hydrophobic features, the PiN (1, 2 and 3 orange) represent features located along implicit lone pair of electrons or implicit hydrogen associated with ring structures and the cyan (8 and 9) features represents hydrogen bond acceptors.
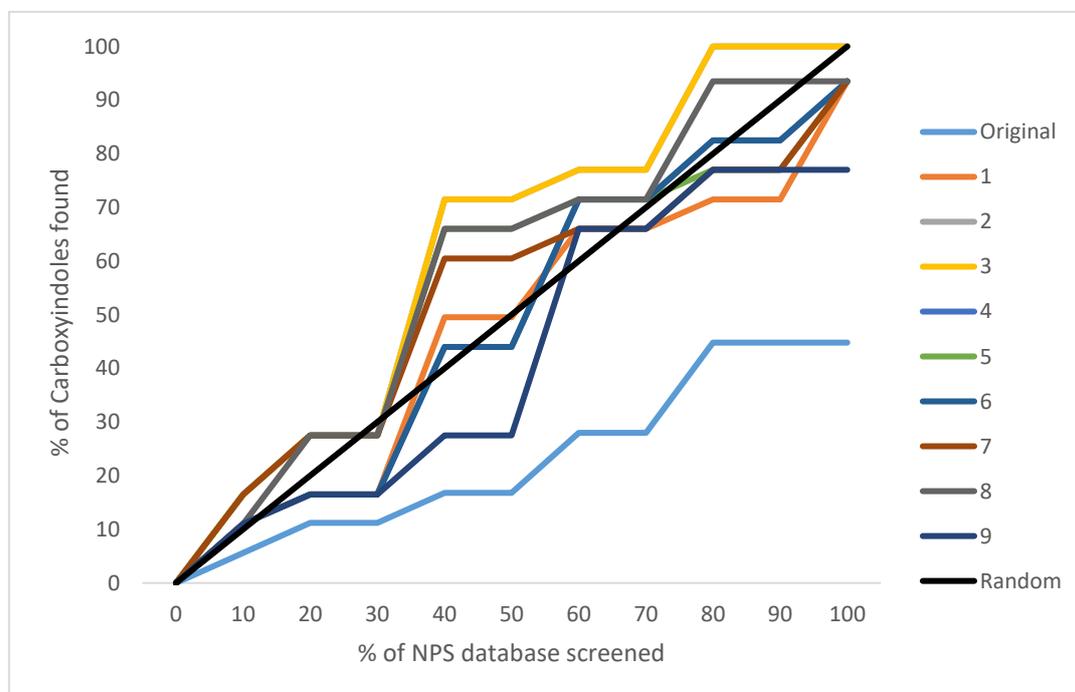
Figure 5. 10: Enrichment plot obtained from the carboxyindole pharmacophore model search, the lines show what percentage of carboxyindole ligands have been retrieved from the NPS database. Each coloured line represents a different model (it should be noted there is some overlap of data

158

points which has resulted in lines 1 and 2 not being visible). The light blue line represents the original consensus pharmacophore model. The different numbered coloured lines represent the results of a search after the removal of that pharmacophoric feature from the original pharmacophore (Figure 5.9). The graph also includes a black line that represents a random hit rate.

The LOO process was able to eliminate 3 features (Feature 2, Feature 3 and Feature 4) and the refined 6-feature model (Figure 5.11) was able to retrieve 100% of the carboxyindole ligands in the dataset in comparison to 45% of the total compounds retrieved by the initial model. The removal of features was achieved using an iterative process to ensure the highest E_f was obtained using the fewest number of features. Removal of features 2, 3 and 4 individually resulted in an improved hit rate (see Table 5.7). The next step was to remove both features 2 and 3 to see if an improved E_f was obtained. An E_f value of 3.3 was obtained when features 2 and 3 were removed compared to 2.8 when only one feature was removed. Removal of features 3 and 4 provided the same E_f of 3.3 as was obtained for the removal of features 2 and 3. Therefore in an attempt to improve the enrichment factor further features 2, 3 and 4 were removed this resulted in an E_f of 5.5 which demonstrates that the best carboxyindole model contains only six features (Figure 5.12).
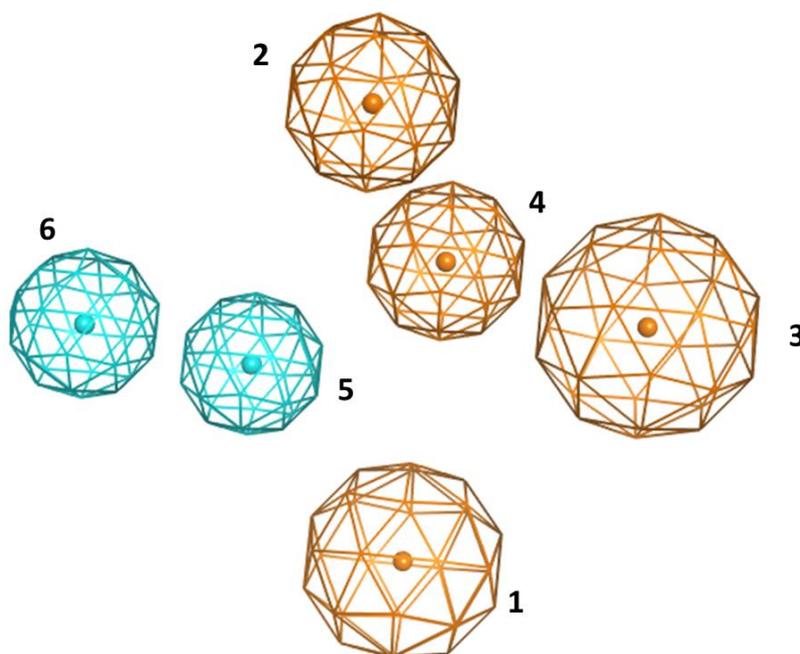
Figure 5. 11: The refined 6-feature carboxyindole pharmacophore model arrived at via the LOO methodology. The Aro|Hyd (1, 2, 3 and 4 orange) features represent aromatic/hydrophobic moieties, and the cyan features (5 and 6) represent hydrogen bond acceptor feature.
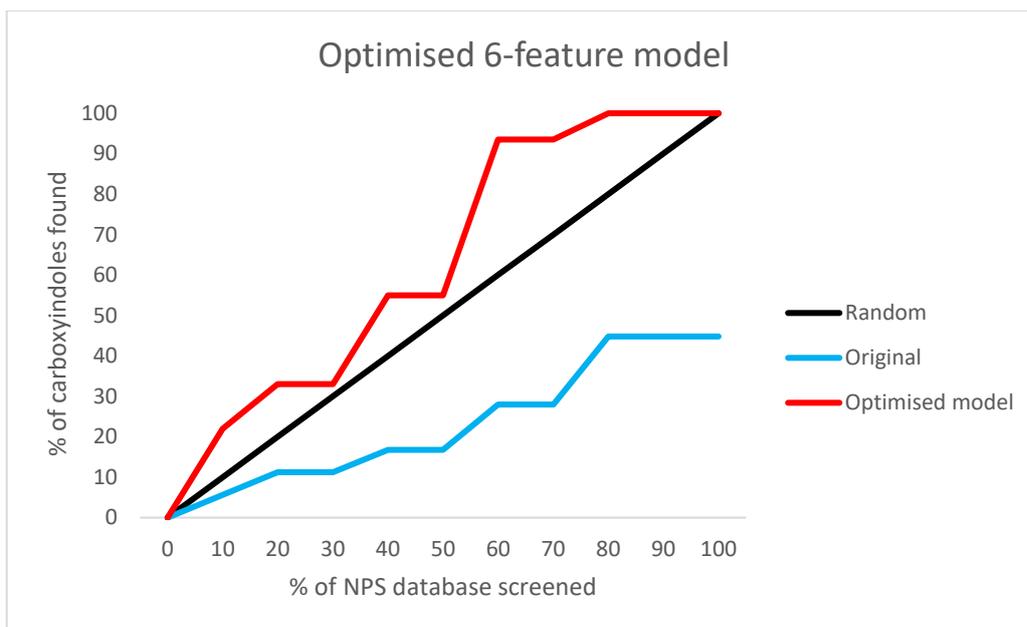
Figure 5. 12: Enrichment plot that compares the performance of the original (blue) 9-feature carboxyindole pharmacophore model to the 6-feature model (red) generated by the iterative LOO method, (see Figure 5.11). The graph also includes a black line that represents a random hit rate.

Table 5. 7: The calculated enrichment factors (E_f) for the carboxyindole pharmacophore model. The table shows E_f for the initial 9-feature carboxyindole pharmacophore model and E_f when a feature has been removed. The E_f are given at different stages of the search which indicate how many carboxyindole ligands had been retrieved within the top 2, 5 and 10% of the NPS database search.

Pharmacophore	Enrichment factors		
	at 2%	at 5%	at 10%
Initial	0	0	0.6
feature 1 removed	2.8	3.3	1.0
feature 2 removed	2.8	3.3	1.7
feature 3 removed	2.8	3.3	1.7
feature 4 removed	0	0	1.0
feature 5 removed	0	0	1.7
feature 6 removed	0	0	1.1
feature 7 removed	0	0	1.7
feature 8 removed	0	0	1.1
feature 9 removed	0	0	1.1
features 1 & 2 removed	8.3	5.8	5.6
features 2 & 3 removed	8.3	5.8	5.6
features 1, 2 & 3 removed	8.3	5.8	5.6

The optimised six-feature pharmacophore model was then carried forward and used to screen the ZINC database for potential novel SC scaffolds.

Analysis of the flexible alignment of the carboxyindoles and the initial consensus pharmacophore show that features 1, 2 and 3 are PiN centroid projections based on the π-system formed from the aligned aromatic rings. The PiN projections generated, place features at a 90° angle to the aligned benzene rings and the presence of these features will restrict the number of hits found in the screen as only molecules that have a benzene ring in the same orientation, to that of the original pharmacophore will be detected. Figure 5.13 shows two benzene rings with the same apparent orientation with respect to the aromatic feature (2) of the pharmacophore, but one benzene ring has been rotated by 180° relative to the other. Although the benzene ring itself has not changed position the associated PiN feature has also rotated by 180°, and hence the overlap of the second orientation with the pharmacophore is considered poor when compared to the first. The PiN features therefore impart a high degree of selectivity in terms of absolute orientation and ultimately place restrictions

on the number of ligands retrieved from a virtual screen when they are present. Therefore, it makes sense that the enrichment factors increase upon the removal of these features.
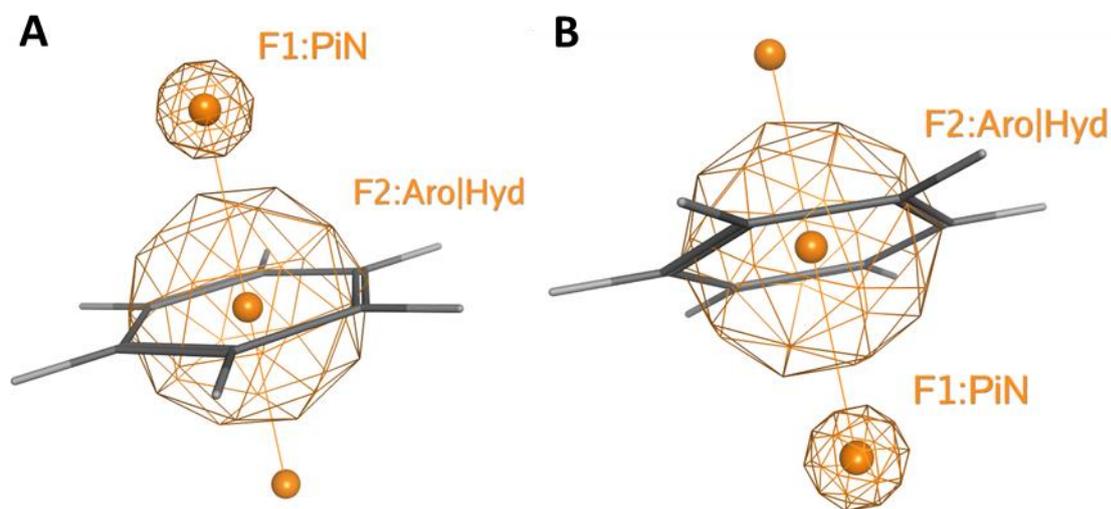
Figure 5. 13: Two benzene rings that have been annotated with PiN features, the benzene ring depicted as B shows the direction of the PiN feature when the benzene ring is rotated by 180°.

5.3.3.2 Naphthoylindole pharmacophore

The original pharmacophore model generated for the naphthoylindole cluster a model with 8 features shown in Figure 5.14
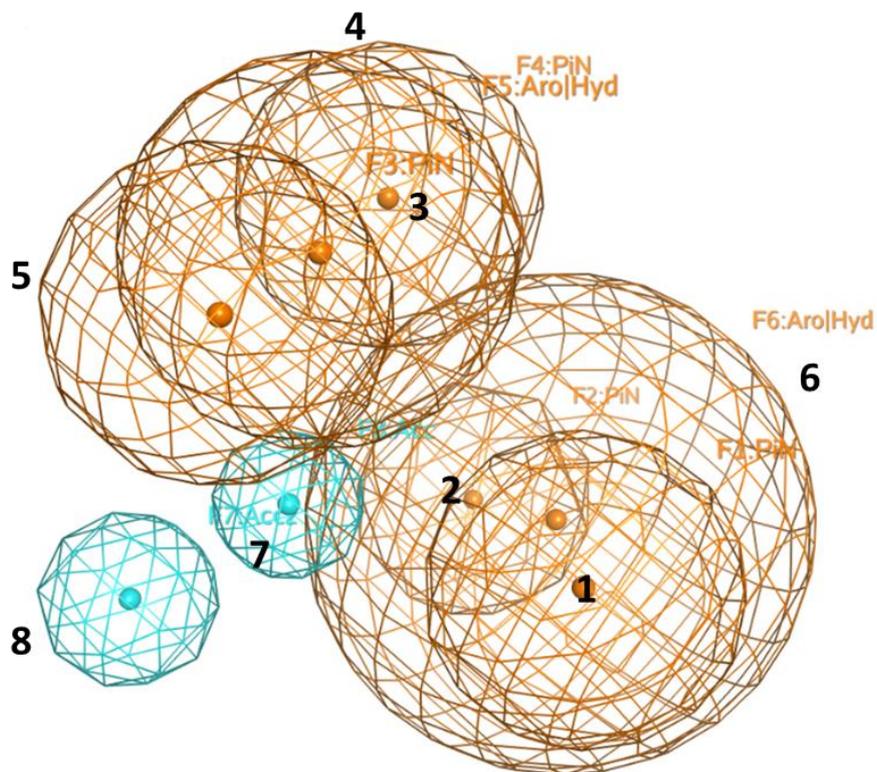
Figure 5. 14: The consensus pharmacophore model based on the flexible alignment of the 48 naphthoylindole derivative ligands. The Aro|Hyd (5 and 6 orange) features represent aromatic/hydrophobic features, the PiN (1, 2, 3 and 4 orange features) represent features located along implicit lone pair of electrons or implicit hydrogen associated with ring structures, and the cyan feature represents a hydrogen bond acceptor feature (7 and 8).
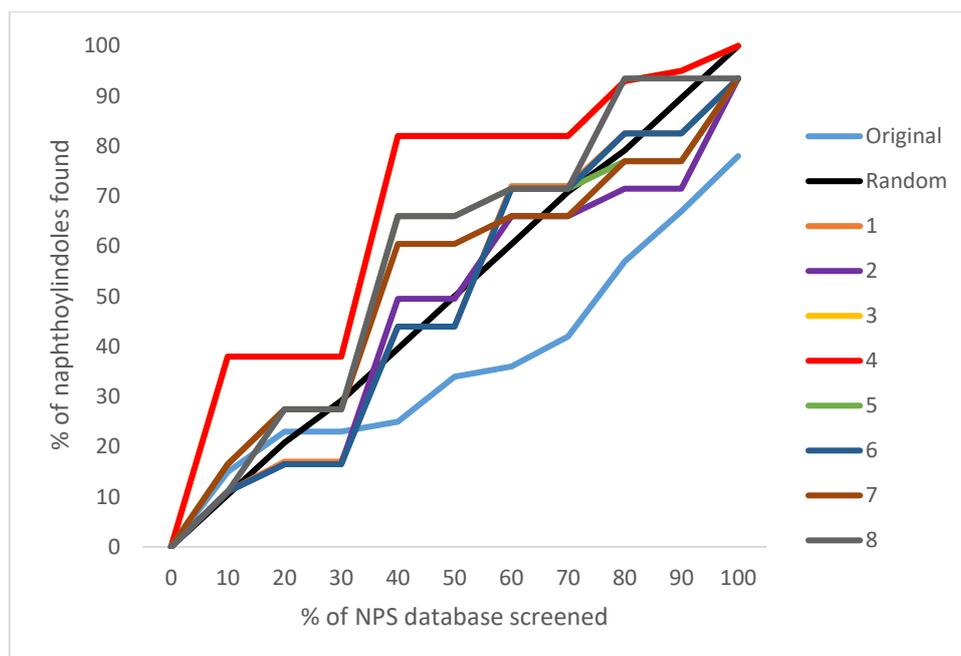
Figure 5. 15: Enrichment plot obtained from the naphthoylindole pharmacophore model search. The lines show what percentage of naphthoylindole ligands have been retrieved from the NPS database at different stages of the virtual screen. Each coloured line represents a different model, it should be noted that lines 1, 3 and 5 are present but are obscured by lines 4 and 6. The light blue line represents the original consensus pharmacophore model. The different numbered coloured lines represent models that result from the removal of that particular pharmacophoric feature (see Figure 5.14) via a LOO process i.e. the results of a search after the removal of a pharmacophore feature. The graph also includes a black line that represents a random hit rate.

Table 5. 8: The calculated enrichment factors (E_f) for the naphthoylindole pharmacophore model. The table shows E_f for the initial 8 feature naphthoylindole pharmacophore and E_f when a feature has been removed. The E_f are given at different stages of the virtual screen which indicates how many naphthoylindole ligands had been retrieved at 2, 5 and 10% of the NPS database search

	Enrichment factors		
	at 2%	at 5%	at 10%
Initial	2.1	1.5	3.1
feature 1 removed	2.1	1.5	3.1
feature 2 removed	4.2	9.8	7.9
feature 3 removed	4.2	4.7	7.9
feature 4 removed	2.1	1.5	3.1
feature 5 removed	2.1	9.8	3.5
feature 6 removed	2.1	1.5	2.3
feature 7 removed	2.1	1.5	3.5
feature 8 removed	2.1	1.5	2.3
features 2 & 3 removed	13.5	9.8	7.9

The optimised naphthoylindole pharmacophore model was reduced from eight features down to six (features 2 and 3 were removed both were PiN features), the removal of these features resulted in the refined 6-feature naphthoylindole pharmacophore model, (see figure 5.16) having an E_f of 6.3 when the first 2% of the NPS database search.

When analysing the flexible alignment of the naphthoylindoles, there were two regions that were heavily populated by aromatic regions. The removed PiN features (2 and 3) account for the collective

projection features generated from the large number of aromatic rings in that particular region. The presence of these features imparts a selectivity with respect to the absolute orientation of aromatic rings that becomes too restrictive to the virtual screening approach as explained above and explains the reduced selectivity of the initial pharmacophore model, compared to the model that has these features removed, thus enabling the full range of naphthoylindoles to be detected by the optimised 6-feature model.
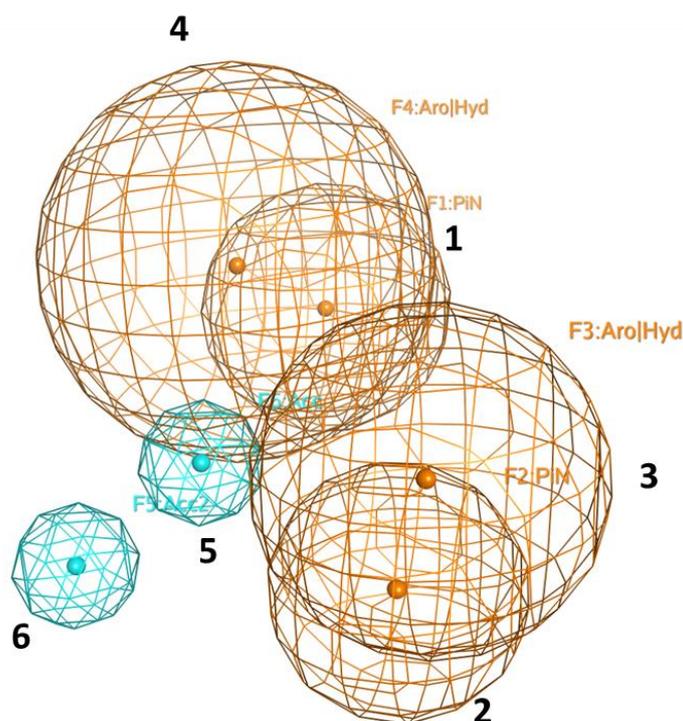
Figure 5. 16: Refined 6 feature naphthoylindole pharmacophore model, features are numbered in black. The Aro|Hyd (3 and 4 orange) features represent aromatic/hydrophobic features, the PiN (1 and 2 orange features) represent features located along implicit lone pair of electrons or implicit hydrogen associated with ring structures, and the cyan feature represents a hydrogen acceptor feature (5 and 6).

5.5 Medoid selection

From each of the four SC clusters a medoid was identified and used as a benchmark to ensure the virtual hits selected from the pharmacophore search were structurally dissimilar to the medoids of the clusters of the known synthetic cannabinoids used to develop the pharmacophore models. This would help ensure novel scaffolds for potential SCs could be identified.

Table 5. 9: shows the common name and chemical structure of the SC medoids from each synthetic cannabinoid cluster used to develop the pharmacophores above. The portions of the first two molecules that have been coloured red highlight the identical moiety between the cluster 1 and cluster 2 carboxyamide medoids.

Synthetic cannabinoid cluster	Medoid Name	Medoid Chemical Structure
Carboxyamide cluster 1	ADB-PINACA	
Carboxyamide cluster 2	ADBICA	
Carboxyindole	AM-694 methyl substituted for Iodine	
Naphthoylindole	JWH-018	
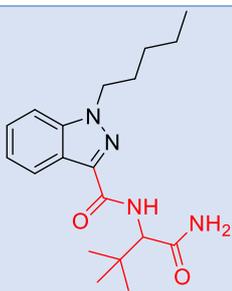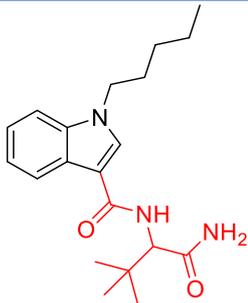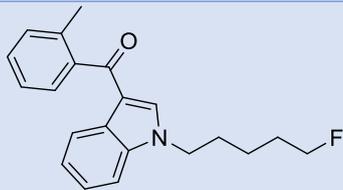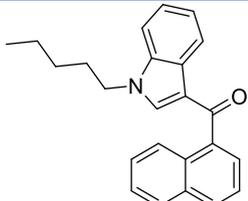
The carboxyamide medoids differ as follows. The Cluster 1 medoid contains an indazole ring whereas the Cluster 2 medoid contains an indole ring. The similarity of these ligands is not unexpected, as the medoids have been generated from the same parent SC cluster (the cluster fragment is coloured red

in Table 5.9). As the majority of the naphthoylindole molecules are analogues of JWH-018, it is also not unexpected that this ligand is the medoid for that naphthoylindole group.

5.5.1 Virtual library selection

The four optimised pharmacophore models were used to screen a set of drug-like molecules in an attempt to identify potential unexploited chemical scaffolds for synthetic cannabinoids. There are a number of different types of libraries available from zinc.docking.org[244], these databases include options such as "clean" datasets where stricter filtering rules (e.g. compounds that contain aldehydes and thiol groups have been removed) have been applied. This is to remove groups that may have unfavourable pharmacokinetic properties[258]. The standard drug-like[259] subset was selected for the virtual screen (at the time of access November 2016), there were 17,900,742 molecules in the drug-like subset.

5.5.2 Virtual screen of database using pharmacophore model

The four refined pharmacophore models identified in Sections 5.3.3.1- 5.3.3.3 were used to search the ZINC "drug-like" database. A cut-off RMSDx value of 1.5 Å was applied for a molecule to be considered a hit. Any compound that fell outside of this range was removed from the screening process. The initial virtual search conducted using the four models retrieved a total of approximately 7.4 million hits from database.

5.5.3 Filtration of virtual hits

The initial virtual hit library of 7.4 million compounds was reduced in size by applying filters that would provide a final library of molecules that were structurally different to known SC but would have similar physicochemical properties. The protocol for achieving this is summarised in Figure 5.17.
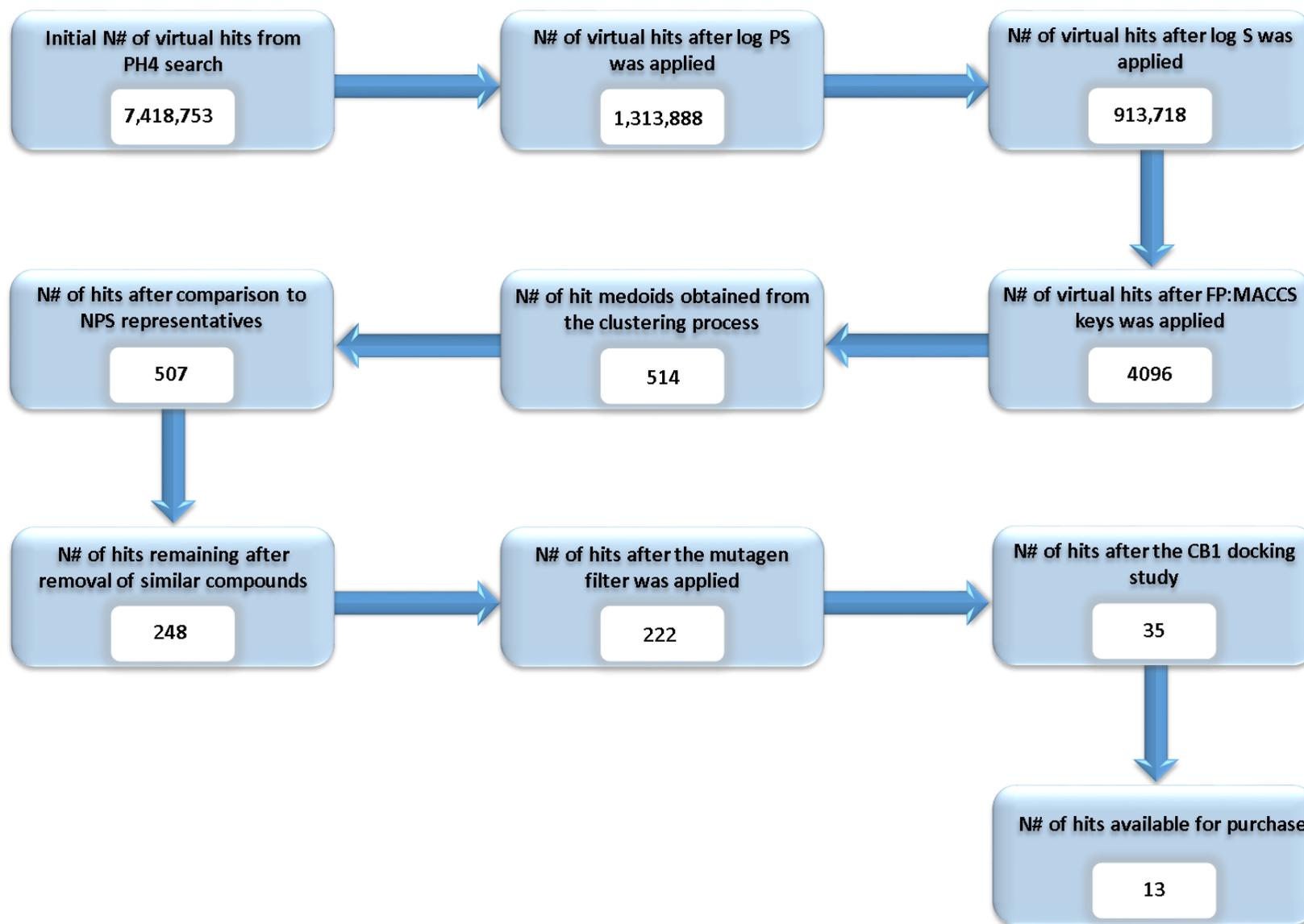
Figure 5. 17: A diagram to show how the number of virtual hits were filtered. The numbers stated are the sum of all the virtual hits found for the four pharmacophore models (carboxyamide cluster 1 and 2, carboxyindole and naphthoylindole).

168

5.5.4 Filtration of hits

The number of molecules retrieved had to be reduced to curate a manageable database of molecules that could be taken further for biological evaluation.

5.5.5 Log PS

The first filter applied to the database was log PS which is the log of permeability surface area product. This filter was applied as it is strongly correlated to the values obtained when calculating the log of the Blood Brain Barrier BBB concentration[245]. Experimental log_{BBB} values were unavailable for the virtual hits, and could not be calculated due to the prohibitive cost of the software required. However, an estimation of the ability for a compound to permeate the blood-brain barrier was deemed important, as it is necessary for a SC to cross the blood brain barrier to elicit a psychoactive effect. For each SC cluster (carboxyindoles 1 and 2, carboxyamides and naphthoylindoles) a minimum and maximum log PS value was established based on the values calculated for all the known SC in that particular cluster. The range of Log PS values was -2.50 to -0.22, and any compounds from the virtual screen that fell outside of this range were rejected.

5.5.6 Log S

The second filter applied was a molecular descriptor that describes aqueous solubility. This filter was applied as it is necessary for virtual hits to be soluble in aqueous media as virtual hits identified will be evaluated for possible biological activity. The experiments to establish this will be conducted using a variety of different aqueous based media, and therefore it is necessary for screening hits to be soluble to some extent in water. The range of log S values identified was -8.93 to -2.88, and hit compounds that fell outside of this range were rejected.

5.5.7 Structural similarity

To minimise the structural similarity between the four identified SC medoids, the NPS representative ligands (see Table 5.11) and the virtual hits retrieved, two different similarity metrics were used. Firstly FP: MACCS structural keys were generated and assigned to both the virtual hits and the identified medoids. Any virtual hit that had a structural similarity greater than 25% to any of the medoids or NPS representatives were rejected. This is to ensure that the virtual hits identified were as structurally distinct from all known NPS as possible allowing for completely novel ligands to be identified.

Secondly, pairwise T_c were calculated for the virtual hits and the initial SC database (containing 162 SC), this was carried out to identify if any of the virtual hits were similar to any known SC not just the SC molecules used to develop the different pharmacophore models. The highest T_c value obtained was 0.46 between AMB-CHMICA and compound **7**, this ensures that the identified virtual hits are structurally different to all other categories of SC and NPS.

Table 5. 10: A list of the NPS representatives used in the comparison between NPS classes and hit molecules.

Name of NPS	NPS classification as described by Zloh *et al*
N-Me-2-AI	Aminoindane
2-MAPB	Arylalkylamines
Methoxetamine (MXT)	Arylcyclohexylamines
Flubromazepam	Benzodiazepines
Mephedrone (4-MMC)	Cathinones
DPT	Indolealkylamines
Fentanyl butanamide analogue	Opioids
Afloqualone	Others
25H-NBOMe	Phenethylamines
DBZP	Piperazines Derivatives
HDMP-28 (methylnaphthidate)	Piperidines & Pyrrolidines

5.5.8 Clustering & Cluster Medoid Selection

Clustering was employed to identify groups of structurally similar molecules from the resulting virtual hits. From each identified cluster a medoid compound was identified and taken forward for further investigation. Medoids were identified from clusters that contained 3 or more compounds. All molecules in clusters with two or fewer compounds were carried forward for further investigation (i.e. no medoids were identified for these clusters). This approach was taken to reduce the number of virtual hits retrieved and would produce a library of optimised compounds which could be evaluated for their ability to elicit psychoactivity.

Three of the virtual hits shared structural similarity with an NPS representative with a T_c of greater than 0.6 and were removed from the study, table 5.11 show the chemical structure of the virtual hits and the corresponding similar representative ligand.

Table 5. 11: shows the chemical structures of the three medoids ligands that were removed from the virtual hit library. The compounds were removed as they had a Tc of greater than 0.6 with the known NPS representatives. This was carried out to ensure the medoids used were structurally distinct from known synthetic cannabinoids and known NPS representatives.

Chemical structure of hit medoid compound	Name and structure of NPS
3-(3,4-dichlorophenyl)-4-methylquinazolin-2(3H)-one	Afloqualone
7-chloro-5-(2-chlorophenyl)-3-methoxy-1-methyl-1,3-dihydro-2H-benzo[e][1,4]diazepin-2-one	Flubromazepam
8-chloro-6-(2-fluorophenyl)-1-methylene-2,4-dihydro-1H-benzo[f]imidazo[1,5-a][1,4]diazepine	
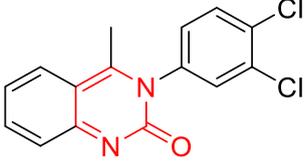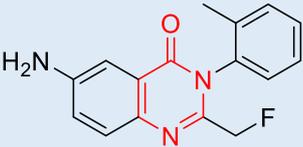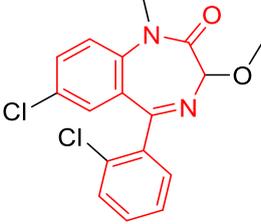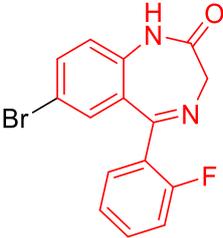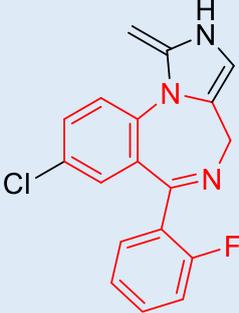
The presence of 2,3,6,7,-tetrahydro-1*H*-1,4-diazepine group attached to two halogenated benzene rings was found in 2 different hit medoid compounds which accounts for the T_c value of 0.6 with Flubromazepam. Similarly, the 5,6-dihydropyrimidine-4 (3*H*)-one group that is found in Afloqulone was also present in at two of the hit medoid compounds and again explains the T_c value of 0.64.

5.5.9 Similarity of virtual hits

A library of 507 virtual hits remained. The final stage of optimising the hits to be carried forwards into testing was to ensure the database of hits selected were as structurally diverse as possible as this would ensure that the breadth of the filtered structures was captured and limit the number compounds being taken forward. This was achieved by calculating pairwise T_c for all the resulting hits and removing any ligands that had T_c of over 0.6 (the molecule with the lowest T_c value was retained) which produced a database of 248 ligands.

5.5.10 Evaluation of drug likeness and toxicity

A database of 248 ligands resulted after the application of the filters listed in section 5.4.5 - 5.4.11. These compounds were then analysed to ensure they all followed the drug like criteria as described by Lipinski[259] and Oprea[228]. Any ligands that violated the criteria were removed. Finally, using the mutagen descriptor option in MOE, ligands that had known mutagenic subgroups were removed from the database leaving 222 potential virtual hits.

5.5.11 Molecular Docking of potential hit molecules

222 ligands were docked using MOE into the CB1 crystal structure (accession code 5TGZ, Resolution: 2.8 Å). Each conformation generated by the software was analysed for the number and type of interactions made within the CB1 binding site. Amino acid residues that had been reported in the literature as being key in the interaction of ligands with the CB1 binding site[97] were used, alongside the reported S-score values, as a benchmark for ranking the docked poses of the ligands. Amino acid residues of interest include Met103, Gly166, Ser167, Phe170, Val196, Trp356, Phe379, Ser383, Cys386 and Leu387[97]. An example of a protein-ligand interaction fingerprint (PLIF) for a docked screening compound is given in Figure 5.18. From this ligand interaction diagram, it can be seen that all of the amino acid residues previously reported as playing a role in binding of small molecules to the CB1 receptor (except Gly166) are either making an interaction with the ligand or are in close proximity to the ligand. This indicates that the likelihood of this compound being able to form an appropriate interaction with the CB1 receptor, and hence elicit psychoactivity, is feasible and warrants further experimental investigation.
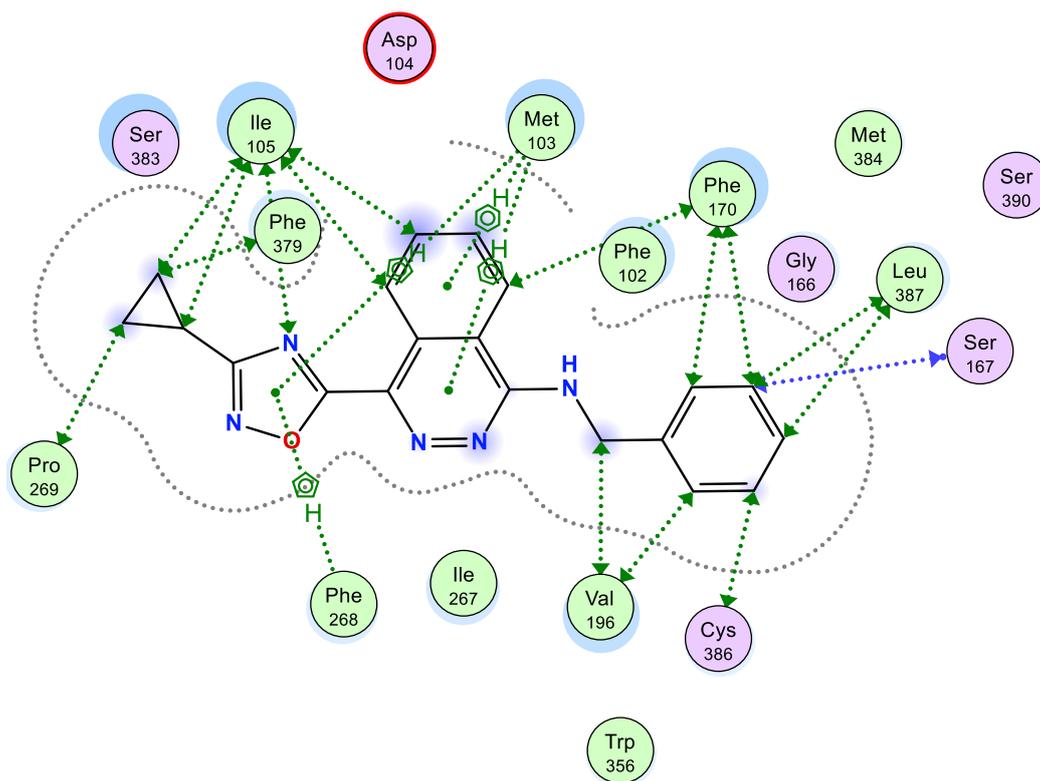
Figure 5.18: PLIF, depicting the ligand N-benzyl-4-(3-cyclopropyl-1,2,4-oxadiazol-5-yl)phthalazin-1-amine docked in the binding site of the CB1 receptor (5TGZ[97]). This docked conformation produced a score of -8.31 Kcal mol^{-1}, in comparison to JWH-018, a known CB1 binder, which produced a score of -7.52 Kcal mol^{-1}.

Docked ligands that had a scored value of -6.55 Kcalmol^{-1} or lower were highlighted as the most likely candidates to successfully bind to CB1 *in vitro*/*in vivo*. This cut-off value was used as the SC JWH-018 had an average docked value of -6.55 (± 0.38) Kcalmol^{-1} based on 15 docked conformations. Thirty-five ligands scored values of -6.55 Kcalmol^{-1} or less, and from the 35 ligands selected 13 compounds were identified (due to financial limitations). To ensure the compounds were structurally dissimilar a correlation matrix of T$_c$ was constructed, see Figure 5.19.

The 13 compounds selected fulfilled all the criteria required to ensure the compounds would interact with the CB1 receptor and have physicochemical properties that were similar to known CB1 agonists therefore providing a reassuring evidence that these compounds would have the ability to elicit psychoactivity.

	M1	M2	M3	M4	1	2	3	4	5	6	7	8	9	10	11	12	13
M1		0.50	0.33	0.31	0.30	0.23	0.22	0.18	0.14	0.13	0.16	0.17	0.27	0.26	0.21	0.24	0.25
M2	0.50		0.54	0.55	0.24	0.24	0.24	0.20	0.11	0.14	0.11	0.16	0.27	0.29	0.30	0.22	0.29
M3	0.33	0.54		0.88	0.19	0.23	0.26	0.15	0.15	0.12	0.15	0.14	0.28	0.30	0.25	0.22	0.32
M4	0.31	0.55	0.88		0.18	0.23	0.24	0.13	0.14	0.12	0.16	0.13	0.28	0.29	0.26	0.20	0.30
1	0.30	0.24	0.19	0.18		0.41	0.23	0.17	0.11	0.16	0.13	0.15	0.20	0.22	0.21	0.17	0.23
2	0.23	0.24	0.23	0.23	0.41		0.26	0.19	0.11	0.11	0.12	0.18	0.18	0.25	0.23	0.19	0.26
3	0.22	0.24	0.26	0.24	0.23	0.26		0.23	0.13	0.12	0.14	0.15	0.20	0.24	0.39	0.15	0.38
4	0.18	0.20	0.15	0.13	0.17	0.19	0.23		0.08	0.10	0.11	0.12	0.14	0.18	0.21	0.11	0.21
5	0.14	0.11	0.15	0.14	0.11	0.11	0.13	0.08		0.09	0.09	0.18	0.13	0.13	0.09	0.22	0.12
6	0.13	0.14	0.12	0.12	0.16	0.11	0.12	0.10	0.09		0.26	0.13	0.15	0.14	0.11	0.12	0.09
7	0.16	0.11	0.15	0.16	0.13	0.12	0.14	0.11	0.09	0.26		0.07	0.13	0.15	0.12	0.09	0.16
8	0.17	0.16	0.14	0.13	0.15	0.18	0.15	0.12	0.18	0.13	0.07		0.13	0.12	0.19	0.27	0.19
9	0.27	0.27	0.28	0.28	0.20	0.18	0.20	0.14	0.13	0.15	0.13	0.13		0.51	0.16	0.19	0.19
10	0.26	0.29	0.30	0.29	0.22	0.25	0.24	0.18	0.13	0.14	0.15	0.12	0.51		0.20	0.20	0.20
11	0.21	0.30	0.25	0.26	0.21	0.23	0.39	0.21	0.09	0.11	0.12	0.19	0.16	0.20		0.15	0.49
12	0.24	0.22	0.22	0.20	0.17	0.19	0.15	0.11	0.22	0.12	0.09	0.27	0.19	0.20	0.15	>	0.18
13	0.25	0.29	0.32	0.30	0.23	0.26	0.38	0.21	0.12	0.09	0.16	0.19	0.19	0.20	0.49	0.18	

Figure 5. 19: correlation matrix of the pair wise T_c values for the four SC medoids (where M1, M2, M3 and M4 represent ADB-PINACA, ADBICA, AM-694 methyl substituted for iodine and JWH-018 respectively) and the 13 compounds selected for biological evaluation, none of the compounds have a pairwise T_c value of above 0.6 indicating that all 13 compounds are structurally dissimilar.

Figure 5.20 highlights the pairwise T_c values for the selected and the SC medoid, this gives confidence that the compounds purchased were structurally distinct from one another.

5.4.12 Final selection of possible synthetic cannabinoid candidates

Table 5.12 shows the structures of the ligands that from the filtering and docking process were predicted most likely to bind to CB1, and hence had an opportunity to elicit psychoactive effects. These compounds were also chosen due to their availability for purchase. The compounds were purchased from MolPort.

Table 5. 12: The chemical structures and systematic names of the final selected 13 virtual hits. The hits were assigned numbers which were then used for the biological evaluation.

Chemical structure and systematic name of the 13 compounds purchased and taken forward for biological evaluation
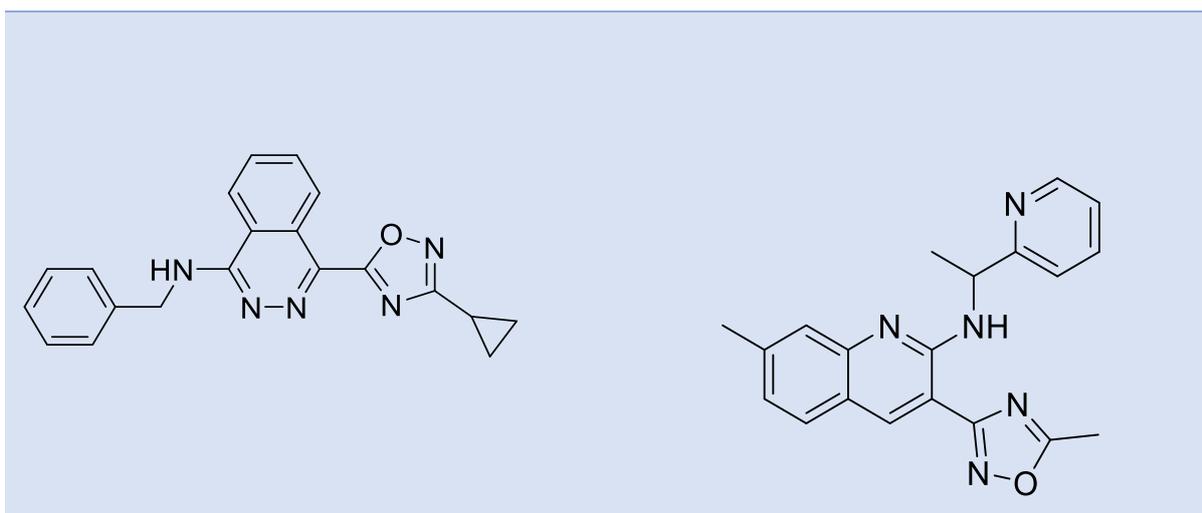
Compound 1

N-benzyl-4-(3-cyclopropyl-1,2,4-oxadiazol-5-yl) phthalazin-1-amine

Compound 2

7-methyl-3-(5-methyl-1,2,4-oxadiazol-3-yl)-N-(1-(pyridin-2-yl)ethyl)quinolin-2-amine
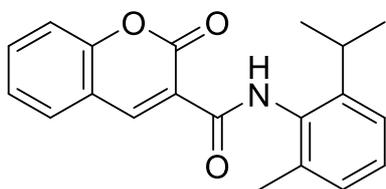
Compound 3

N-(2-isopropyl-6-methylphenyl)-2-oxo-2H-chromene-3-carboxamide
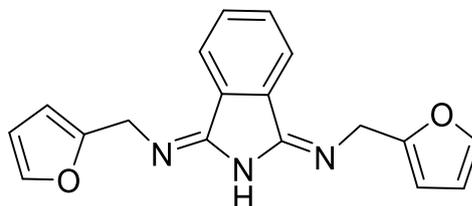
Compound 4

(1E,3E)-N1,N3-bis(furan-2-ylmethyl)isoindoline-1,3-diimine
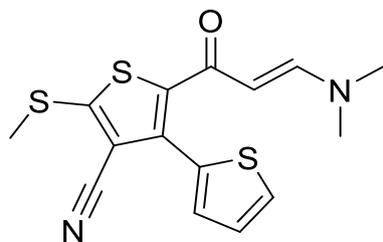
Compound 5

(E)-2'-(3-(dimethylamino)acryloyl)-5'-(methylthio)-[2,3'-bithiophene]-4'-carbonitrile
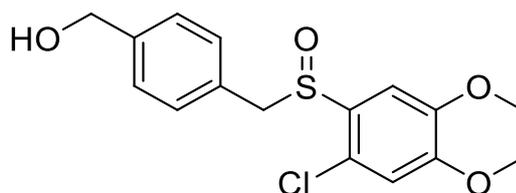
Compound 6

(4-(((7-chloro-2,3-dihydrobenzo[b][1,4]dioxin-6-yl)sulfinyl)methyl)phenyl)methanol
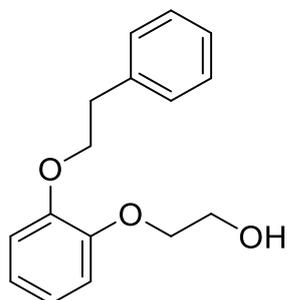
Compound 7

2-(2-phenethoxyphenoxy)ethan-1-ol
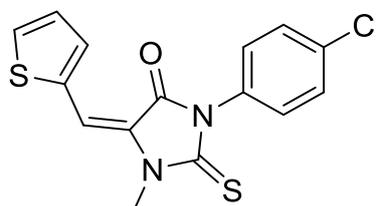
Compound 8

(E)-3-(4-chlorophenyl)-1-methyl-5-(thiophen-2-ylmethylene)-2-thioxoimidazolidin-4-one

176

Chemical structure and systematic name of the 13 compounds purchased and taken forward for biological evaluation

Compound 9

8-chloro-3-methyl-10-phenylpyrimido[4,5-b]quinoline-2,4(3H,10H)-dione
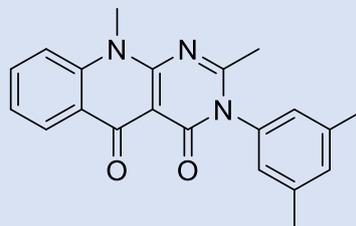
Compound 10

3-(3,5-dimethylphenyl)-2,10-dimethylpyrimido[4,5-b]quinoline-4,5(3H,10H)-dione
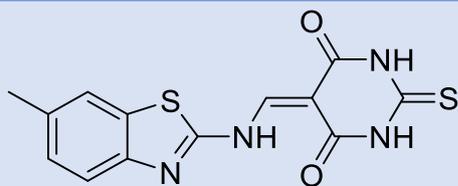
Compound 11

4-bromo-N-(3-ethynylphenyl)benzamide
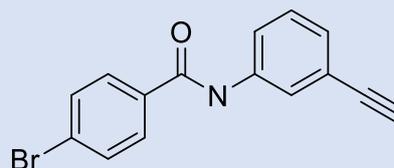
Compound 12

5-[[(6-methyl-1,3-benzothiazol-2-yl)amino]methylidene]-2-sulfanylidene-1,3-diazinane-4,6-dione

Compound 13

N-(2,3-dihydro-1H-inden-5-yl)-2,4-difluorobenzamide

5.6 Conclusions

A library of known synthetic cannabinoids was used to develop pharmacophores, which in turn were used to identify compounds that would be predicted to have the same propensity for binding to the CB1 receptor as known SC, but crucially would be significantly structurally different from known SC and other known NPS.

The class of synthetic cannabinoids that contained the most compounds was the carboxyamide derivatives. From this group, two different pharmacophore models were generated. The cluster 2 pharmacophore model was able to retrieve 40% of the desired synthetic cannabinoid within the first 10% of the total NPS database and at 2% of the total database screened had an E_f of 3.1 during virtual screening experiments. Pharmacophore models for the napthoylindole and carboxyindole synthetic cannabinoid clusters were also generated with E_f of 13.5 and 8.3 respectively at 2% of the total database screened. Both of these models were able to retrieve 22-25% of the desired synthetic cannabinoid ligands within the first 10% of the NPS database, and were a significant improvement over random, which indicates that the pharmacophore models developed were successful at identifying appropriate compounds from a database of known NPS.

The initial virtual screen conducted using the four pharmacophore models generated, retrieved between 1.3-2.3 million hit molecules per pharmacophore model which is approximately 7% of the total ZINC drug-like database. The log PS and log S filters reduced the number of potential molecules in each SC sub cluster down to between 0.3 and 4.7% of the ZINC database. The filter that excluded the most ligands was the FP: MACCS similarity filter. This filter served a dual purpose by first reducing the number of ligands and secondly ensuring the ligands carried further into the study had no more than 25% structural similarity[103] to the SC medoids i.e. there was more opportunity to identify a chemical scaffold that was truly novel/unexploited.

The docking studies conducted using the virtual hits indicate that an interaction between these ligands and the CB1 receptor is feasible and could potentially lead to the identification of novel compounds that could bind to the CB1 receptor. The 13 compounds identified for progression into *in vitro* studies all had appropriate physicochemical properties that were determined from known SC to suggest they had the potential to bind to the CB1 receptor. The virtual hit compounds also showed similar interactions in the CB1 receptor binding site to known SC such as JWH-018 as evidenced by the docking studies. A T_c pairwise comparison showed that the 13 virtual hits were structurally different to known SC as the highest calculated Tc between the virtual hits and the medoid ligands was 0.32 (Figure 5.19).

The next stage of this research is to screen the compounds *in vitro* to establish any possibility that the molecules identified produce a biological response.

Chapter 6 *In Vitro* **and** *Ex Vivo* **evaluation of potential CB1 receptor ligands.**

6.1 Introduction

The endocannabinoid system is considered complex and has yet to be fully understood. Cannabinoid receptors belong to the family of G Protein Coupled Receptors (GPCRs), activation of these receptors initiates a biochemical cascade which ultimately alters normal neurotransmitter release[260].

Thirteen potential synthetic cannabinoid (SC) like compounds were identified as likely "hit" candidates from the virtual screen conducted out in Chapter 5. "Hit" candidates were molecules that were identified using pharmacophores used to screen a virtual library, and selected based on a number of physicochemical filters. The compounds were selected based on their structural dissimilarity to SC and purchased accordingly. It is the aim of this chapter to illustrate the approach used to identify the biological activity of these compounds and investigate the activation of the CB1 receptor, which can help indicate the potential of these compounds to elicit a psychoactive effect. The potential biological activity of these compounds were assessed using mammalian cell lines and immunosorbent technique ELISA[1,2–7]. A preliminary study involving electrical field stimulations (EFS), was also employed in an attempt to further validate these molecules as potential cannabinoid receptor agonists[266–269].

To ascertain whether the compounds of interest interacted with the CB1 receptor, a cAMP ELISA was carried out to assess whether there is an increase in cAMP when the cell is exposed to a compound. This particular assay has been extensively employed[264,270,271,263,272], and a number of studies report that SC increase cAMP accumulation in cells located in the brain when experiments are conducted without the presence of the commonly used stimulant forskolin[271,273,274]. Forskolin directly activates adenylyl cyclase and increases cAMP concentrations[275]. The commercially available CHO-CNR1 cell-line that expresses the CB1 receptor was purchased and a batch growth curve was generated alongside the use of the MTS and LDH assays to identify growth characteristics required to optimise the ELISA experiment, as well as to establish initial cell-based toxicological responses to the compounds. Cell culture was also employed as a standalone technique to generate information about the cell growth profile. A study carried out by Sibaev *et* al, (2013) showed that the endogenous cannabinoid anandamide (see Figure 6.1) activates the CB1 receptors located in the ileum of rodents. The most commonly used rodents for this type of experiment are rat[276] and guinea pig[277] . As such, to obtain further validation on the biological activity of the screening compounds, an EFS study using rat ileum was carried out as evidence exists to suggest that SC depresses or has an active role in gut motility[266,267,268].
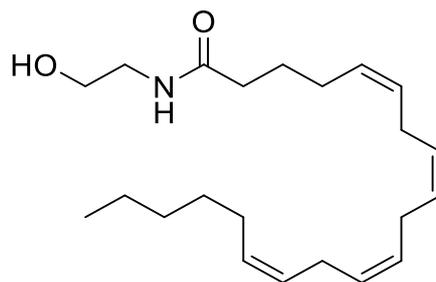
Figure 6. 1: The chemical structure of the endogenous cannabinoid anandamide (*N*-arachidonoylethanolamine, AEA).

Based on the premise that the compounds selected from the virtual screen would interact with the CB1 receptor, the EFS study was conducted to further confirm these molecules a valid CB1 receptor ligand.

This chapter will detail the methodology employed to identify possible biological activity exerted by the compounds of interest. The studies carried out suggest that compound 1 and compound 12 (Figure 6.2) interact with the CB1 receptor and this could be indicative of the ability to elicit psychoactivity.

Compound **1** also significantly increased the amount of cAMP in *in vitro* studies in comparison to MAM-2201 which is a known potent SC.
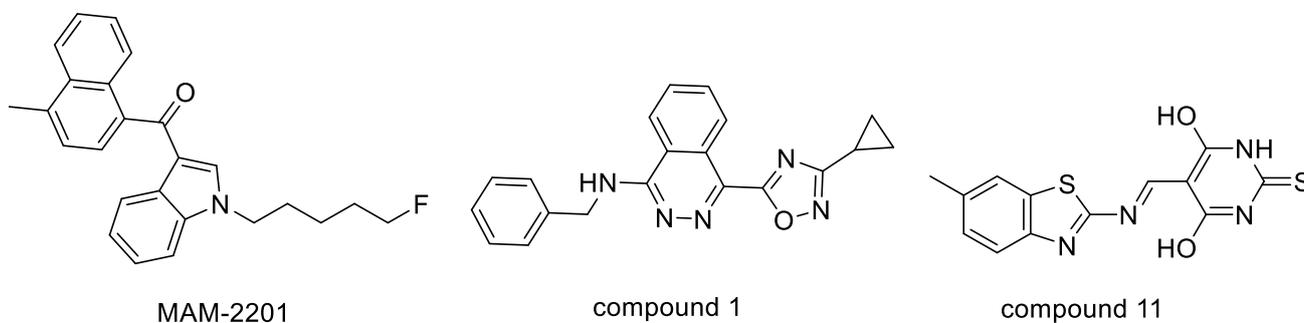
Figure 6. 2: The chemical structure of MAM-2201 (the positive control used) compound **1** and compound **11**, the two compounds that showed increased levels of accumulated cAMP in the cAMP ELISA.

6.2 Materials

6.2.1 Equipment

Fluorescent samples were analysed using the Promega Luminometer Glomax (Southampton, UK), colorimetric assays were read using the LabSystems Multiskan Ascent plate reader (Agilent Technologies Inc, California, USA). Electrical field stimulation tests were carried using Harvard Apparatus 30 mL organ baths (Cambridge, UK)

6.2.2 Cell culture reagents

Chinese hamster ovary epithelial cells (CHO-CNR1, Gene I.D 1268, passage number: 4, ECACC: 12110601 lot number: 13A006) were purchased from Culture Collections, Public Health England (Salisbury, UK). Cell culture reagents for CHO-CNR1 cells included Ham's F12 nutrient mixture, trypsin and phosphate buffered saline (PBS) were purchased from Sigma-Aldrich, Dorset, UK. L-glutamine, penicillin-streptomycin and foetal bovine serum (FBS, pH 7.4) were all purchased from ThermoFisher Scientific (Loughborough, UK). Sterile preparation of stock solutions and chemicals were performed either by filtration through a 0.22 μm Whatmann sterile filter and/or autoclaving at 121°C at 1 bar for 1 h.

6.2.3 Selected "hit" compounds

Thirteen compounds (see Table 6.1) were purchased from MolPort (Riga, Latvia). The positive control MAM-2201 was provided by an in-house library of NPS held at the University of Hertfordshire (purchased from Chiron Surrey, UK). The positive control was prepared following the protocol stated by the Home Office licence.

Table 6. 1: Compounds that were evaluated for biological activity.

Compound	Name	Structure
1	N-benzyl-4-(3-cyclopropyl-1,2,4-oxadiazol-5-yl)phthalazin-1-amine	
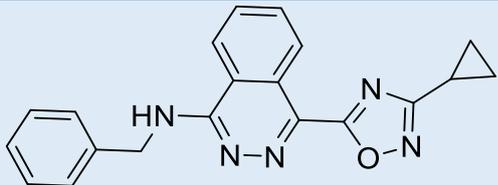
2	7-methyl-3-(5-methyl-1,2,4-oxadiazol-3-yl)-N-(1-(pyridin-2-yl)ethyl)quinolin-2-amine	
3	N-(2,3-dihydro-1H-inden-5-yl)-2,4-difluorobenzamide	
4	4-bromo-N-(3-ethynylphenyl)benzamide	
5	2-(2-phenethoxyphenoxy)ethan-1-ol	
6	(4-(((7-chloro-2,3-dihydrobenzo[b][1,4]dioxin-6-yl)sulfinyl)methyl)phenyl)methanol	
7	(E)-3-(4-chlorophenyl)-1-methyl-5-(thiophen-2-ylmethylene)-2-thioxoimidazolidin-4-one	
8	1-(furan-2-yl)-N-(3-{[(furan-2-yl)methyl]imino}-2,3-dihydro-1H-isoindol-1-ylidene)methanamine	

9	(E)-2'-(3-(dimethylamino)acryloyl)-5'-(methylthio)-[2,3'-bithiophene]-4'-carbonitrile	
10	N-(2-isopropyl-6-methylphenyl)-2-oxo-2H-chromene-3-carboxamide	
11	(E)-4,6-dihydroxy-5-(((6-methylbenzo[d]thiazol-2-yl)imino)methyl)pyrimidine-2(1H)-thione	
12	8-chloro-3-methyl-10-phenylpyrimido[4,5-b]quinoline-2,4(3H,10H)-dione	
13	3-(3,5-dimethylphenyl)-2,10-dimethylpyrimido[4,5-b]quinoline-4,5(3H,10H)-dione	
Positive Control	MAM-2201 ([1-(5-fluoropentyl)-1H-indol-3-yl](4-methyl-1-naphthalenyl)-methanone) [265,278]	
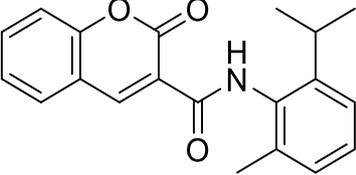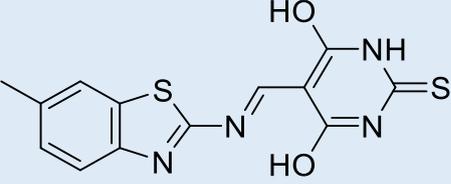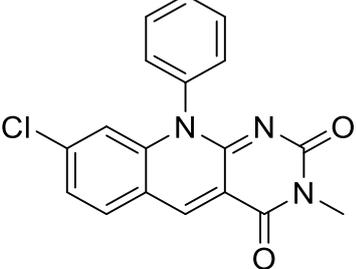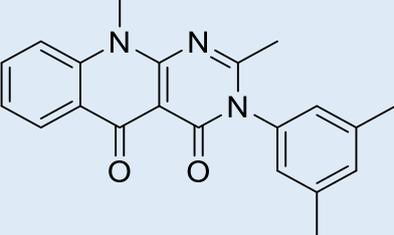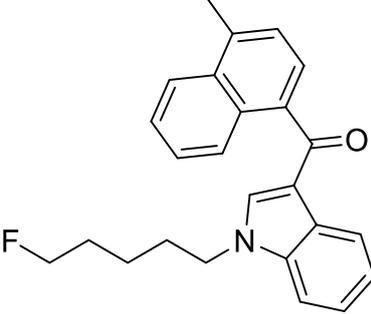
6.2.5 Enzyme-linked immunosorbent assay (ELISA)

The cyclic AMP multispecies competitive ELISA kit was purchased from ThermoFisher scientific (Loughborough, UK). 3-Isobutyl-1-methylxanthine (IBMX) purchased from Sigma-Aldrich (Dorset, UK) was used in the cAMP assay study. The positive control used was the known synthetic cannabinoid [1-(5-fluoropentyl)-1H-indol-3-yl](4-methyl-1-naphthalenyl)-methanone (MAM-2201 see table 6.1).

6.2.6 Electrical Field Stimulation (EFS) study

Tissue preparation was conducted on site (University of Hertfordshire, UK) by fully trained personnel.

- Wistar Rats, from Charles River, UK.
- Each Rat is given: Aspen Wood Chew, Water, and Lab 3 Diet
- Schedule one Procedure, Euthanized through exposure to increasing CO_2 levels and followed by cervical dislocation.
- Weight: 250-350g
- Sex: Male

Data was recorded using Labscribe iWorx (iWorx Systems, Inc Dover, UK) which was connected to the EFS apparatus and included the Harvard 6002 stimulator and a Harvard Variable cycle timer (Harvard apparatus, Massachusetts, US).

6.3 Methods

6.3.1 CHO-CNR1 cell revival

The cryovial of CHO-CNR1 (passage number 4) was removed from liquid nitrogen storage and rapidly brought to 37 ^{o}C by immersing in a 37 ^{o}C. Thawed cells were dispersed in 3 mL of pre-warmed complete media (37 ^{o}C) and then transferred to a 15 mL centrifuge tube with an additional 7 mL of pre-warmed media. The cell suspension was centrifuged at 250 g for 5 min at room temperature (~19 ^{o}C) to form a cell pellet. The old media was aspirated, and the cell pellet was then re-suspended in 10 mL of fresh, pre-warmed complete media, and transferred to a, media-pre-wetted, T75 flask and left to incubate (37°C, 5% v/v CO_2 humidified incubator) for 12h. After the incubation period, 15 mL of fresh complete media was added.

6.3.2 CHO-CNR1 cell culture

Cells were cultured in a Ham's F12 nutrient media that was supplemented with 8mM L-glutamine, 1% v/v penicillin-streptomycin and 10% v/v foetal bovine serum (FBS) and dispersed within T75 culture flasks. Cells were passaged when cell confluency reached approximately 90% and were maintained in a humidified incubator at 37 ^{o}C and 5% (v/v) CO_2. As CHO-CNR1 cells are adherent, trypsin was required to detach cells from the bottom of the T75 flask when passaging. (Trypsinization exploits the enzyme trypsin to break down the proteins associated with cell adherence, after this process has occurred the supernatant is neutralised with complete media[279]. Trypsin is deactivated by protease inhibitors present in the added FBS[280]). Briefly, old media was aspirated away and, 7 mL of trypsin solution was added to the cells and left to incubate. Cell detachment was observed under the microscope, cells that were detached could be seen floating on the surface of the flask (~10 minutes). The trypsinized supernatant was neutralised with 7 mL of fresh complete media and transferred to a centrifuge tube. The supernatant was centrifuged at 250 G for 5 min at room temperature 19 ^{o}C to form a cell pellet. The cell pellet was then re-suspended in 10 mL, or as required in terms of dilution/cells per mL of fresh pre-warmed complete media, and transferred to a, media-pre-wetted, T75 flask and left to incubate (37°C, 5% v/v CO2 humidified incubator).

6.3.3 Trypan blue cell viability exclusion assay

Absolute cell counts, and estimated cell viability was carried out using the Trypan blue exclusion technique. 10 μL of a 0.4% (w/v) Trypan blue solution was mixed with an equal volume of cell suspension before being introduced into an improved Neubauer haemocytometer chamber. The four outermost corner squares which contained cells that had not taken up the trypan blue stain were

counted as being viable. Cells that had not taken up the stain remained dark blue and were deemed as dead cells. This is due to Trypan blue being absorbed through cell membranes that have been compromised. Total cell number was calculated as number of cells per millilitre of cell suspension.

6.3.4 Cell metabolic activity assay/MTS assay

The CellTiter 96 Aqueous One solution cell proliferation assay (Promega, Southampton, UK) was used to determine mitochondrial activity of CHO-CNR1 cells via the conversion of MTS to a formazan product accomplished by NADPH or NADH, which is produced by enzymes in metabolically active cells[281]. The methodology was followed according to the manufacturer's instructions with the following amendments. A 10µl aliquot of the Aqueous One Solution reagent was added to each well containing 50µl cell suspension on a 96 well plate. The plate was then covered in aluminium foil and incubated for 90 mins in 37°C, 5% v/v CO_2 humidified incubator. Absorbance was measured at 490nm using Multiskan ascent plate reader. MTS assay was used to determine % cell viability (respective of control). Complete F12 media and Triton X-100 (lysis solution) were used as background and negative controls, respectively.

6.3.5 CytoTox-ONE membrane integrity assay

The CytoTox-ONE membrane integrity assay kit (Promega, Southampton, UK) was used to determine LDH release (due to a compromised cell membrane) from CHO-CNR1 cells according to the manufacturer's instruction with the following amendments. 50 µl of the cell supernatant was transferred into each well of a black 96-well plate. Triton X-100 (lysis solution), 1% v/v DMSO and media were used as controls. 25 µl of the CytoTox-ONE reagent was added to each well, covered with foil and then left for 20 min at room temperature (~19C). Next, 25µl of stop solution (provided with the Promega CytoTox-ONE assay kit) was added to the wells. Fluorescence was measured using Glomax multi detection system plate reader at 560 excitation and 490nm emission. Data obtained was used to determine percentage viability of CHO-CNR1 cells using the following equation;

Percent cytotoxicity = <u>100 x (Experimental – Culture medium back ground)</u>

(Maximum LDH Release – Culture Medium Background)

Equation 6.1: Calculation used to determine percent cytotoxicity of CHO-CNR1 cells using the CytoTox ONE membrane integrity assay [18,281]. The maximum LDH release value was taken from cells that had been lysed with Triton X, the culture media background value was taken using complete media.

6.3.6 cAMP ELISA

The cAMP ELISA was carried out following the manufacturer's instructions. The compounds were dissolved in 1 mL DMSO which was used as the stock solution. This was diluted with complete media so that the final DMSO concentration within each well was 1 %. Prior to conducting the ELISA assay, cells were incubated with a buffer mixture of complete media and 0.5 mM of IBMX for 45 mins in order to stabilise any cAMP produced[282]. The IBMX buffer was aspirated away and cells were washed with wash buffer (from cAMP ELISA kit) and replaced with a 200 µL mixture of Ham's F12 complete media (180 µL) and drug mixture (20 µL) at a final concentration of 1×10^{-5} mM. Cells were lysed by treatment of the sample with lysing reagent (200 µL of 0.1 M hydrochloric acid and 0.1 (v/v) % of Triton X) for 10 mins. Cell lysis was monitored by visual inspection using a microscope during this process, lysed cells do not have the atypical shape of CHO cells but tend to be more elliptical. The 96 well plate was then centrifuged (600 G at room temperature for 10 mins) and the resulting cell supernatant was directly used for the cAMP ELISA. The assay was carried out in triplicate. The solution (200 µL) containing the compounds of interest were added to the wells of the cAMP ELISA. 50 µL of cAMP-AP conjugate was added to the well. To the same wells 50 µL of cAMP antibody was added, the plate was sealed using the plastic cover provided and left to incubate at room temperature on a plate shaker (500 rpm) for 2 h. After the incubation period, the contents of the well were emptied and washed in triplicate using the 1X wash buffer provided. To the wells 200 µL of pNpp (*p*-nitrophenyl phosphate) substrate solution was added and left to incubate at room temperature for 1 h (no shaking). After the 1 h incubation period, 50 µL of stop solution was added to each of the wells and the optical density was recorded at 405 nm.

6.3.7 Cell Staining

Cell staining was performed using Shandon™ Kwif-Diff™ stains purchased from ThermoFisher Scientific (Loughborough, UK). A 96 well plate containing approximated 100,000 cells per well were fixed using 3.7 % (w/v) paraformaldehyde (PFA, in PBS pH 7.4). Cells were washed in PBS before fixing, this was done to remove dead cells, debris and FBS from the sample. PFA was added and left to stand at room temperature for 20 min. The 96 well plate was then washed three times with PBS. The cells were stained following the Shandon™ Kwif-Diff™ staining procedure which consisted of adding 50 µl of reagent 1 (methanol) to each well. The reagent was left to stain the cells for 20 sec, removed and the washed with PBS. The same step was conducted using reagent 2 (eosin) and once more with reagent 3 (methylene blue). After the three stains had been applied to the wells, and washed with PBS, the cells were finally washed with deionised water and left to air dry for 1 h.

6.3.8 Electrical field stimulation (EFS)

Approximately 1 cm segments of full-thickness rat ileum were prepared and cleaned whilst submerged in Krebs buffer (see Table 6.2 for composition). The ileum tissue was cleaned by using a pipette filled with fresh Krebs buffer to flush the ileum tissue. This was done until only clean buffer could be seen flowing through the ileum. The tissue was placed between two metal prongs which were attached to an electrode inside a 30 cm^3 organ bath (Harvard Apparatus, Cambridge, UK). The organ bath was filled with 15 mL of Krebs buffer which was kept at 37°C using a heated water bath and constantly oxygenated with Carbogen gas (95% oxygen & 5% CO_2).

Table 6. 2: The chemical composition of the Krebs Buffer solution (pH 7.2) prepared in double distilled water and used throughout the EFS study.

	moles	g/L
NaCl	119	6.96
KCl	4.7	0.35
NaH$_2$PO$_4$	1	0.12
MgSO$_4$	1.2	0.14
NaHCO$_3$	25	2.1
Glucose	11	1.98
CaCl$_2$	2.5	0.28

Labscribe (iWorx, version 3.011400) was used to calibrate the transducer, determine the baseline tension and record all data generated during the experiment(s). Viability of the tissue was determined using carbachol (10^{-4} M)[266] and then thoroughly cleaned by washing the tissue four times with fresh Krebs buffer. The tissue was left to rest at 37°C in Krebs buffer to equilibrate for a minimum of 10 minutes in order to reach a stable resting basal tone. The EFS study was carried out as a single experiment.

Two looped sections of cotton thread (Marvel Fabrics, Hatfield, UK) were sewn into the freshly prepared tissue: one to the top and one to the bottom of the tissue section sample. The looped cotton threads were attached close to the edges of the dissected tissue (2-3 mm). One loop was attached to the bottom of the organ bath to anchor the tissue which was maintained at a tension of approximately 1 g, the second loop was attached to a transducer which was connected to the Labscribe iworks software to measure tissue contraction. The tension was maintained at a constant value throughout the experiment.

A frequency response curve (1 Hz to 60 Hz) was constructed, at supramaximal voltage of 50 V, pulse width, of 0.5 s, for 2 seconds every 3 minutes. 150 µL of each compound at a concentration of 10^{-6} M was added to the organ bath using a pipette immediately prior to stimulation, the change in tension (g) was then recorded at 1, 10, 20, 40 and 60 Hz.

6.4 Results

6.4.1 CHO-CNR1 cell revival and culture

Following the revival process, cell number reached 80-90% confluency (see figure 6.3) within approximately 48 hours. Cells were passaged, on average, every two to three days or until the T75 flask reached 80-90% confluency.
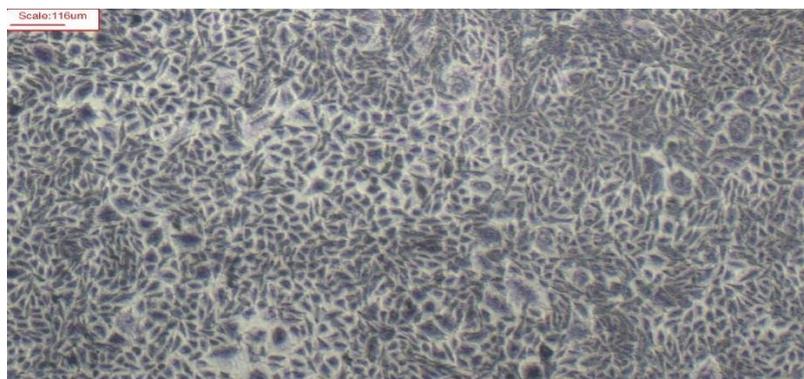
Figure 6. 3: Representative image of CHO-CNR1 cells at 90-100% confluency. Cells are adhered to a T75 flask after 48 hrs. Images taken with a GXCAM-9 digital microscope C-mount camera (GT Vision Suffolk, UK) mounted on an Olympus CKX41 microscope at x4 magnification.
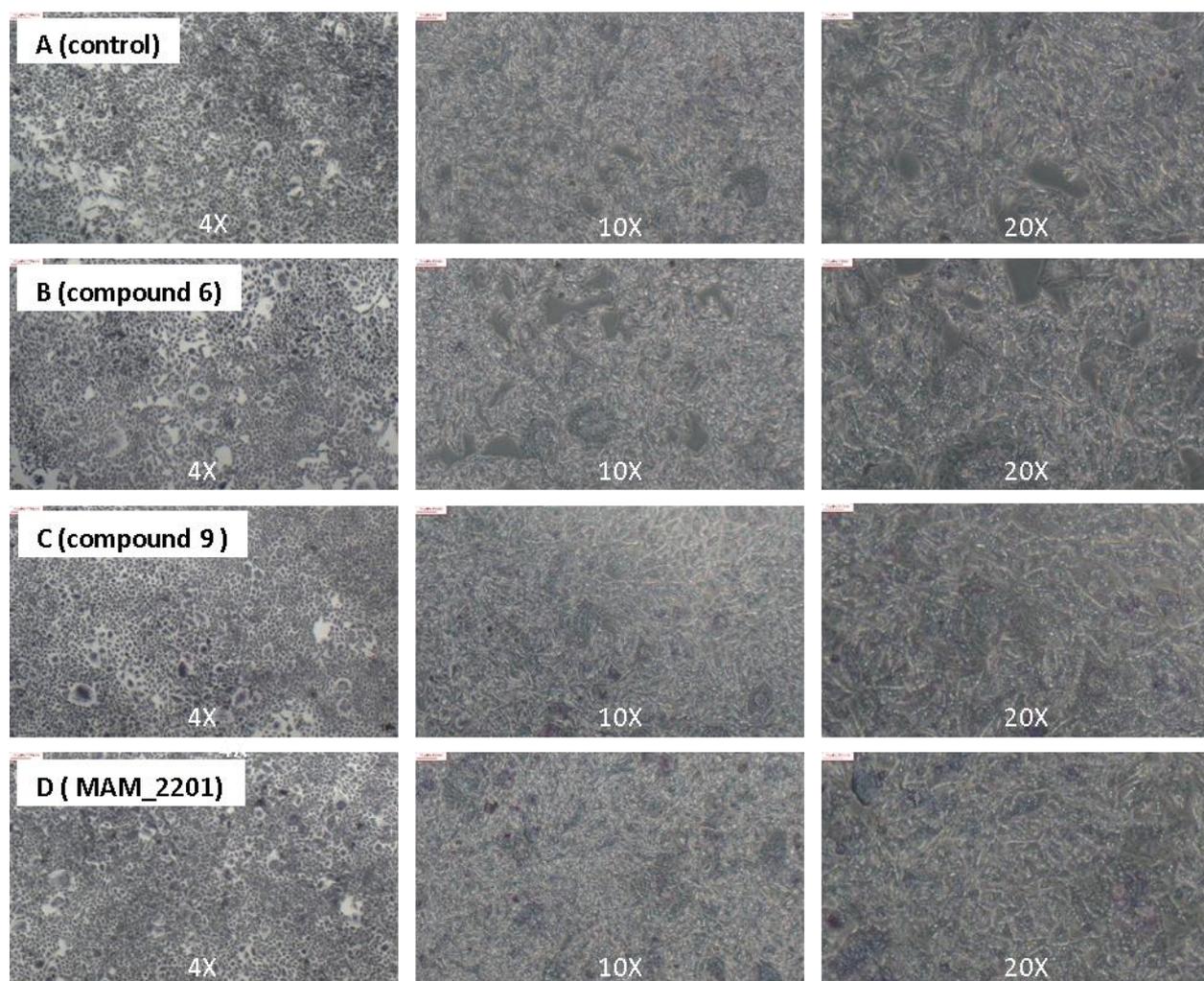
Figure 6. 4: Representative images of CHO cells A: at 4x, 10x and 20x, B: 4x, 10x and 20x, C: at 4x, 10x and 20x and D at 4x, 10x and 20x. Images taken with a GXCAM-9 digital microscope C-mount camera (GT Vision Suffolk, UK) mounted on an Olympus CKX41 microscope. Image is a randomly selected representative sample from 3 experiments.

6.4.2 CHO-CNR1 cell batch growth curve

A growth profile was generated for three different cell passage numbers (6, 8 and 11) using the trypan blue exclusion assay and used to elucidate the exponential growth phase. From Figure 6.5 it was deduced that this phase occurs between days 3-4. This information was used to best assess when to carry out the cAMP ELISA plate assay. For all three assays 96 well plates were used and were seeded with 10,000 cells per well.
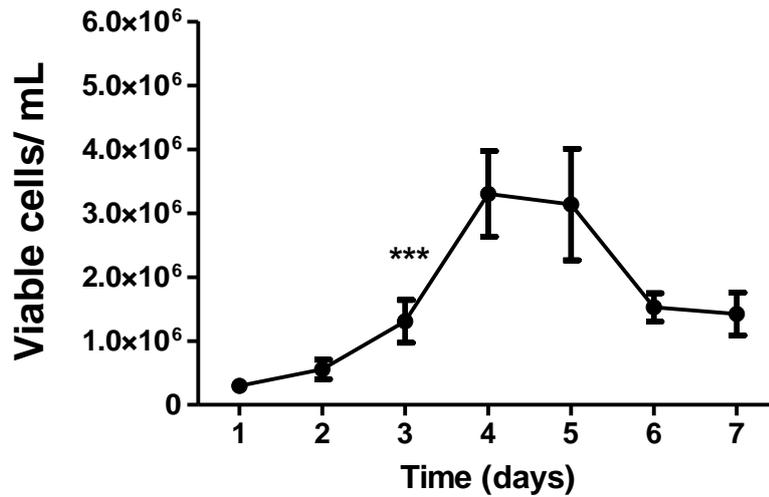
Figure 6. 5: The growth curve of CHO-CNR1 cells over a seven-day period, using 3 different passage number (3, 8 and 11). For each passage number, four trypan blue counts were conducted giving a total n=12. Cells were grown in F12 Hams complete media and cell counts was conducted using trypan blue exclusion assay. Results are expressed as a mean +/- SD of three independent experiments with n =12. P value of 0.0001 (***) was determined relative to day 1 and day 3 using 1-way ANOVA. Post-test comparisons were made using Dunnett's test at 95% confidence interval.

6.4.3 MTS cell viability assay

The MTS assay (Figure 6.6) was used to characterise the metabolic activity of the cells over a 7-day period.
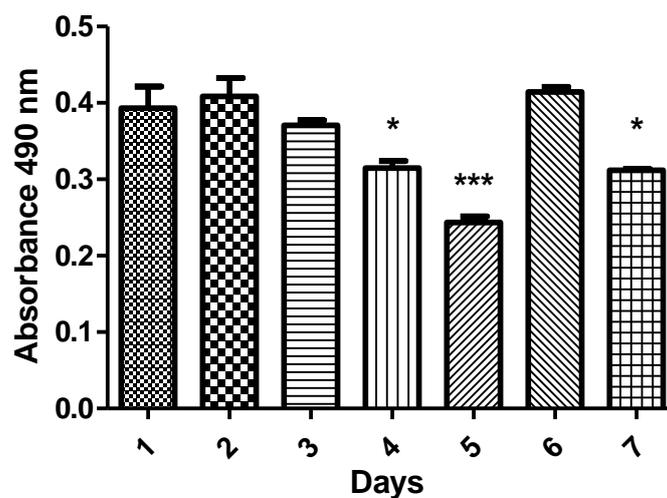
Figure 6. 6: The absorbance values (in nanometres) obtained from the MTS assay that was carried out over a seven-day period. The data points with standard deviation bars are calculated based on three

different passages (3, 8 and 11) that were assayed in triplicate. Results are expressed as a mean +/- SD of three independent experiments with n =12. P value of 0.0001 (***) was determined relative to day 1 using 1-way ANOVA. Post-test comparisons were made using Tukeys test at 95% confidence interval.

6.4.4 LDH cytotoxicity assay

The LDH assay is used as a measure of the amount of lactate dehydrogenase present. LDH is released upon cell death and can be used to calculate the percentage of dead cells. Figure 6.7 shows that the maximum amount of LDH present is 25% after a 7-day period, this value is compared to the total cell death which is calculated as 100% presence of LDH.
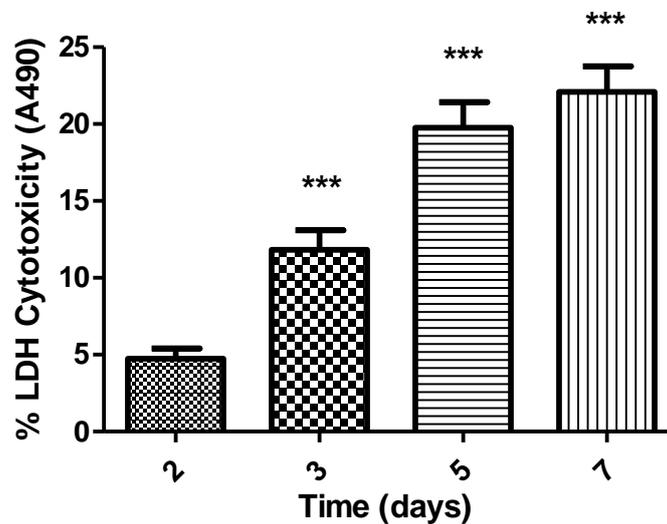
Figure 6. 7: The percentage of dead CHO-CNR1 cells over a seven-day period. The percentage values were calculated using the % cytotoxicity equation (Section 6.3.5, Methods) Results are expressed as a mean + SD of three independent experiments with n =12. P value of 0.0001 (***) was determined relative to day using 1-way ANOVA. Post-test comparisons were made using Turkeys test at 95% confidence interval.

The data obtained from these three assays conducted indicated that biological evaluation should be carried out 3-4 days after passaging, this is due to the percentage of LDH being relatively low and cells appear to be viable and able to proliferate (from data obtained from MTS and Trypan blue assays)

6.4.5 CHO-CNR1 cytotoxicity screening

To ensure the compounds of interest were not cytotoxic, an MTS assay was carried out at a concentration of 10^{-5} M, this concentration was selected as it has been reported in the literature[283,265,284] as being appropriate for the positive control MAM-2201.

A 96 well plate was seeded with 100,000 cells per well and incubated with the compounds for 90 mins. Figure 6.7 details the results of the MTS assay, the graph shows that Compound 6 has significantly (P = 0.0001) increased the mitochondrial activity in comparison to the negative control (cells only) indicating no cytotoxic effect of the compounds.
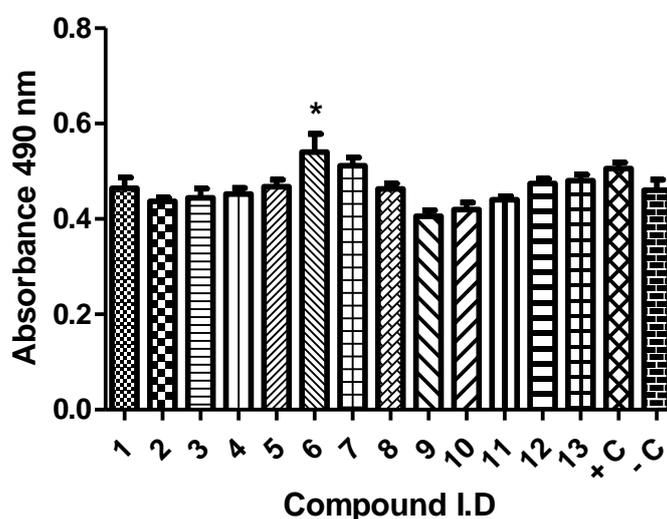
Figure 6. 8: results (n = 6) of the MTS assay carried out after 90-minute incubation period of the compounds with the CHO-CNR1 cells. Results are expressed as a mean + SD of two independent experiments with n =6. P value of 0.0001 (*) was determined relative to negative control using 1-way ANOVA. Post-test comparisons were made using Dunnett's test at 95% confidence interval. The positive control (C+) is the results obtained when MAM-2201 was added to the cells and the negative control (C-) are cells that have had no addition chemicals added.

6.4.6 cAMP assay

The commercially available cAMP ELISA assay includes a cAMP standard that are used to construct an internal calibration curve (see Figure 6.9). The calibration curve was used to quantify the cAMP present in the supernatant samples collected after incubation with the compounds. All results obtained were within the range of the internal standard curve produced.
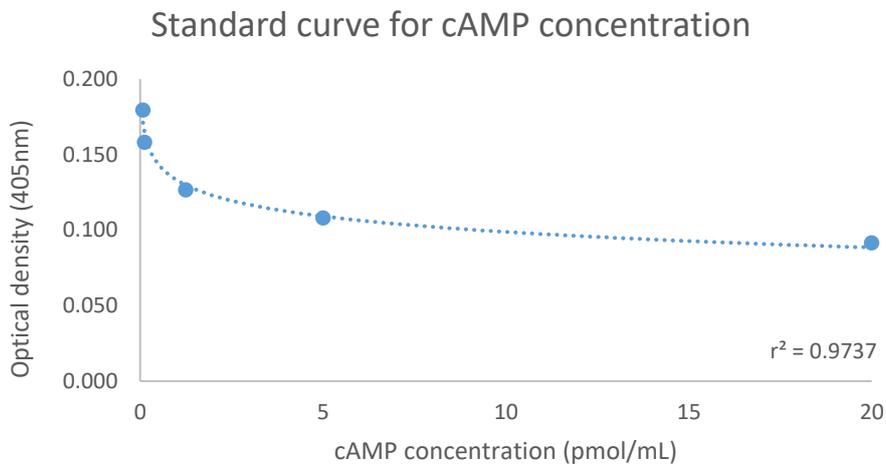
Figure 6. 9: The standard curve used to determine cAMP concentrations for assayed samples.

A number of studies conducted have shown there is a correlation between cannabinoid molecules and an increase in the accumulation of intracellular cAMP levels[285,271,286], this is indicative of binding affinity for the CB1 agonists and the CB1 receptor.
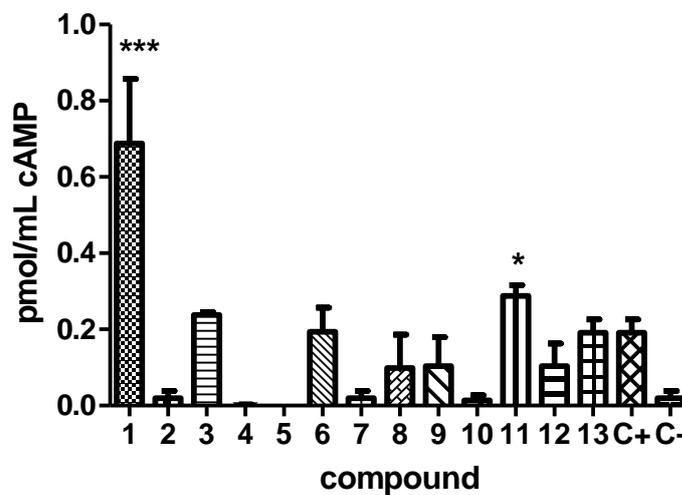
Figure 6. 10: histogram summarising the cAMP production in treated samples, the calculated amount (pmol/mL) of cAMP that was present in each well for the compounds of interest (n=6), MAM-2201 (C+, positive control) and cells that had been assayed with only the addition of IBMX (C-, negative control). Results are expressed as a mean + SD of three independent experiments with n = 9. P value of <0.001 (***) was determined relative to positive control (MAM-2201), using 1-way ANOVA. Post-test comparisons were made using Dunnett's test at 95% confidence interval.

The ELISA assay was carried out in triplicate using three different passage numbers and three different ELISA plates and, from figure 6.10 it can be seen that compound **1** (0.69 pmol/mL), compound **3** (0.24 pmol/mL) **and** compound **11** (0.39 pmol/mL) had the highest values of cAMP present. Compound 1, **3** and **11** produced more cAMP than the positive control which produced 0.24 pmol/milk This could indicate that these molecules have a higher efficacy in activating the CB1 receptor, or that the compounds of interest are causing a response that is responsible for the increase in cAMP.

6.4.7 Electrical field stimulation

Tissue viability was assessed before the compounds were screened, this was carried out by treating the tissue with 1 µM of carbachol; the addition of this compound induces muscle contraction and therefore proves the tissue was viable. Figure 6.11 shows the magnitude of the contraction after the addition of Carbachol, this viability test was conducted on all tissue samples used.

Figure 6.11 is a diagram of an evoked contraction caused by the addition of carbachol, this is commonly used to test tissue viability[268,269,274]. After the tissue was deemed viable EFS was conducted with only Krebs buffer in the organ bath. This was used as a negative control. Once the tissue had returned to a resting tension the positive control was added and contraction tension at the different frequencies (1-60 Hz) were recorded (see Table 6.3). After the addition of each compound, and once all frequencies had been applied the tissue was thoroughly washed out and left to return resting tension before another EFS study was carried out.
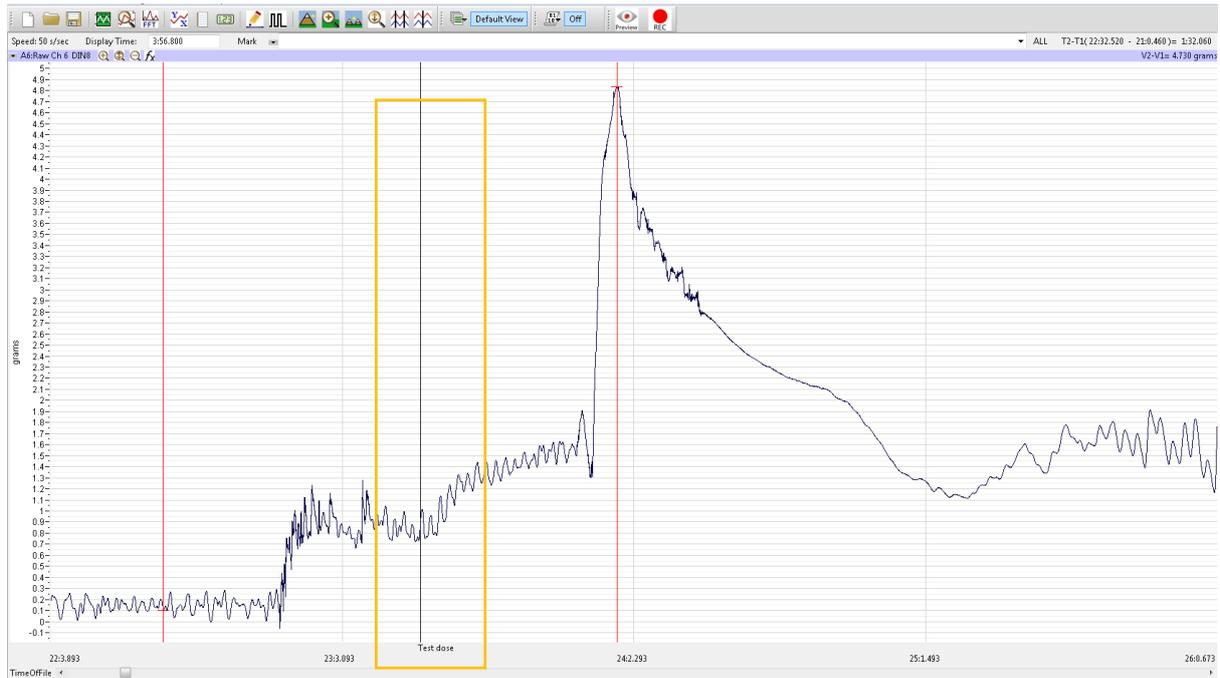
Figure 6. 11: EFS trace of an induced muscle contraction. The line that indicates test dose (highlighted in yellow) shows the point at which the carbachol was added to the organ bath. The peak shows that the tissue has contracted, the quantitative data is obtained by subtracting the maximum tension values (V2) from the resting tension value (V1). These values were obtained by placing red cursor lines at the base line tension and a second red cursor line at the peak of the contraction. The values for V1 and V2 can be seen in top right-hand corner of figure 6.10 i.e. 4.730 g.
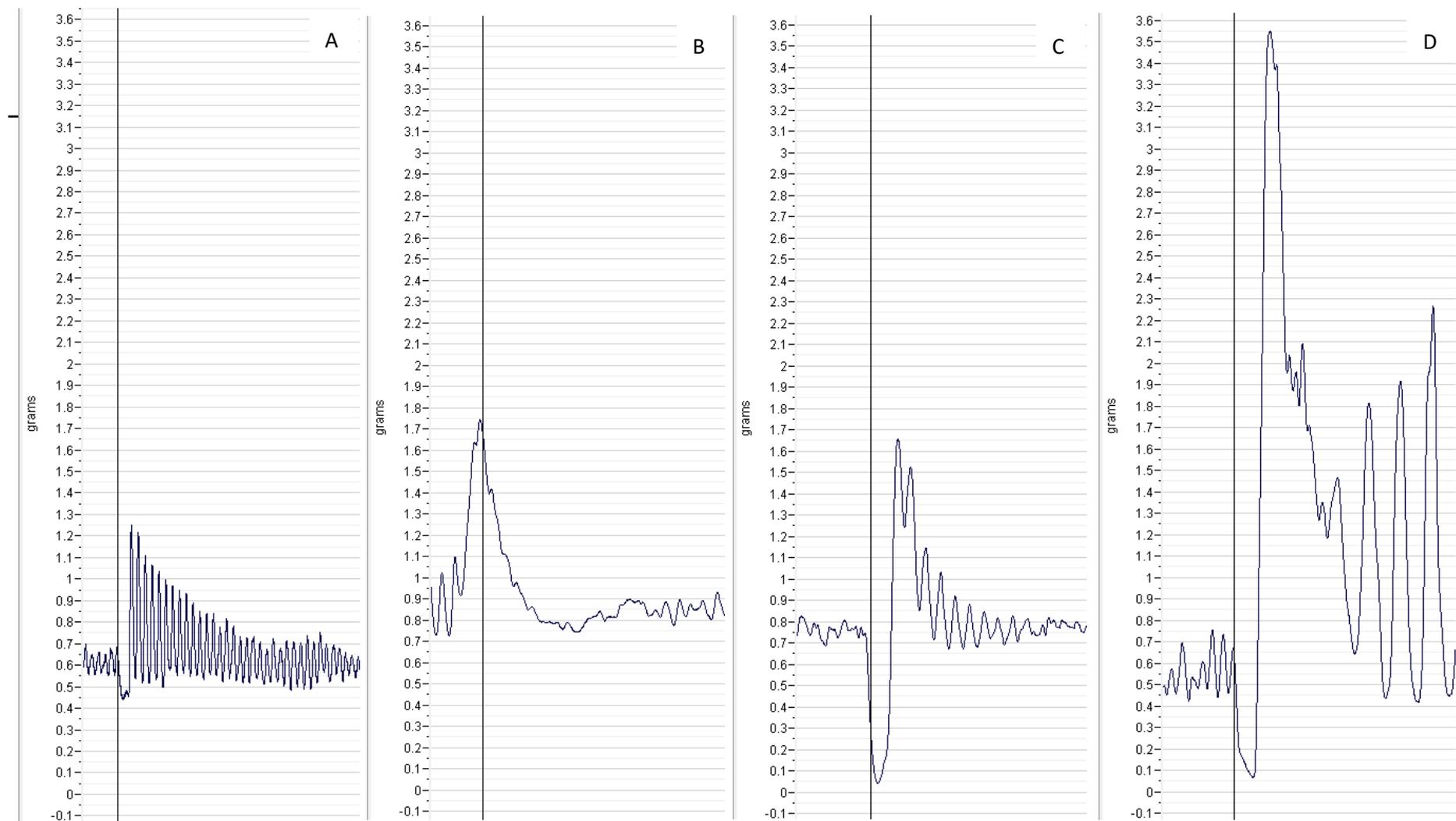
Figure 6. 12: Representative traces of the EFS-evoked an EFS pulse of 10 Hz is applied to A – Compound 1, B – compound 11, C – positive control (MAM-2201) and D – negative control (tissue and Kerbs buffer only). Compound 1 exhibited a tension change of 0.604 g, compound 11 exhibited a tension change of 0.703 g and the positive control exhibited a relaxation wave (0.679 g) and then a contraction of 0.930 g. The data for compounds 1 and 11 have been shown as these compounds exerted the greatest change in tension to comparison to all the other compounds screened.

Figure 6.12 shows the changes in tension at 10 Hz in the presence of compound 1, 12, the negative control and the positive control. At 10 Hz compound **1** and MAM-2201 had very similar contraction patterns, both had a small delay after the addition of the compound and then a large peak which was recorded after the EFS was applied. Compound **11** appeared to have a different effect on the tissue, a peak was recorded after the pulse was applied however the tissue seemed to relax and reach a resting basal tone much quicker in comparison to other compounds.

Table 6. 3: The percentage increase in tension values (g) observed after the EFS had been applied to the tissue sample (tissue samples taken from: Wistar Rats, from Charles River, UK) at the different frequencies in the presence of the positive control (MAM-2201), the 13 compounds of interest and a negative control (Krebs buffer only) n = 1. See appendix for all EFS traces

	1 Hz	10 Hz	20 Hz	40 Hz	60Hz
Percentage Increase					
Negative Control	25.1	100.0	100.0	87.7	77.2
Positive Control - MAM-2201	22.6	25.5	51.2	49.4	40.0
compound 1	24.1	27.3	14.0	13.3	11.9
compound 2	6.1	44.5	66.0	54.6	44.0
compound 3	23.2	49.4	50.6	43.4	37.1
compound 4	20.9	56.1	59.7	49.1	39.7
compound 5	17.6	33.2	53.1	51.6	42.3
compound 6	21.1	30.5	59.2	57.1	52.2
compound 7	9.9	33.3	36.6	34.1	29.3
compound 8	13.0	27.7	36.4	32.8	25.5
compound 9	20.1	31.9	33.1	31.7	23.3
compound 10	10.4	30.4	35.5	26.3	20.5
compound 11	34.8	28.3	30.2	27.9	33.1
compound 12	16.9	30.1	30.4	54.3	52.9
compound 13	25.0	17.8	23.6	22.3	18.1

The results in Table 6.3 show that after the initial evoked contraction at 1 Hz Compound 1, 3, 11 and 13 had a percentage increase in tension value of greater than the value obtained for the positive control MAM-2201 (22.6%). When the frequency pulse is increase to 10 Hz Compound **1** still depressed the contraction - more so than the positive control, Compound **11** also displayed inhibition but not to the same extent as MAM-2201. Although the results do suggest that some of these compounds have the ability to inhibit evoked contractions at varying frequencies, more investigation is required for conclusive evidence to prove these compounds as being CB1 receptor ligands.

6.5 Discussion:

To validate the results obtained from Chapter 5 *in vitro* and *ex vivo* methods were carried out to establish if any of the compounds identified interacted with the CB1 receptor. *In vitro* and *ex vivo* tests are routinely used in the characterisation of novel drug entities. The therapeutic applications of CB1 ligands have been extensively investigated for their uses in analgesia, appetite regulation[287], nausea, and emesis[277]. The work presented in this chapter highlights how *in Silico* methods can be coupled with biological evaluation in the identification of novel CB1 receptor ligands.

CHO-CNR1 cells

CHO cells were used in this study as this cell line was purchased with expression of the CB1 receptor and has been well documented in the literature for investigations conducted using synthetic cannabinoids[,273]. A review article published in 2002 highlights how the molecules Δ^9-THC, CP55940, L-759656, HU-308 and JWH-133 were able to elicit a response from CHO cells that have been transfected with CB1 and CB2 receptors[96]. One proposed mechanism of action is that the activation of the CB1 receptor results in changes in the amount of cellular cAMP[288] which has been recorded using CHO cells and cAMP assays.

Growth Curve

A batch growth curve was produced to establish the different phases within the life cycle of this particular cell line. This was done to identify when the cells were actively growing which is important as this phase is when the cells are the most sensitive to changes in their environment[289], as NPS have been shown to stimulate cellular responses in a wide range of cells[290–292]. Figure 6.5 show the four phases of cell growth, the lag phase where growth is very slow was seen between 24-48 hr after passaging as the cells are adapting to their new environment[293]. During the exponential phase, cells are known to be more active this is the ideal time to conduct any *in vitro* cytotoxicity testing this is

seen between day 3 and 4. The stationary phase can be seen at days 5-6 and the death phase is seen at day 7, this is typical of CHO cell growth as seen by Petrov and Tsupkina (2013)[294].

Trypan blue exclusion assay

The determination of cell viability is crucial for cell culture when evaluating the properties of a compound in a biological assay[295]. In addition to cell viability, the number of cells within a certain volume of media is also important as this can be used to determine the growth rate of a particular cell line. Trypan blue exclusion assay was used as this method provides rapid estimation of cell number, cells are stained with the blue dye the viable cells do not absorb the dye and remain white in colour. Non-viable cells are stained blue due to a lack of integrity in the cell membrane which allows cells to absorb the dye[296]. This assay, in combination with MTS and LDH, is routinely used to determine overall cell health[297]. From the results obtained the cell growth profile is in accordance to what is reported in the literature[298].

MTS assay

The MTS assay is commonly used in measuring cell viability or drug toxicity by measurement of mitochondrial metabolic rate[299]. Other cell viability assays include MTT and XTT which involves the enzyme mitochondrial dehydrogenase reducing the MTT/XTT to a measurable formazan product. This conversion is widely believed to happen in viable cells[300]. MTS was selected over MTT/XTT as this is a one-step convenient, rapid process for assessing cell viability which is sensitive and does not require a termination step like MTT/XTT[299]. The results obtained for the MTS assay shows very little decrease in mitochondrial activity over a 7-day period. As the cells begin decreasing in number after day 6 it would be expected to see a decrease in the absorbance values for the MTS from day 6 onwards. To gain a better understanding of the mitochondrial activity profile of this cell line the MTS assay should be continued post day 7. The results obtained from the MTS assay (Figure 6.6) conducted show that the mitochondrial activity does start to decline after day 5 which is concomitant with the results obtained from the trypan blue assay. However, after day 6 there is an increase in activity which is unexpected. To overcome this, the MTS assay could be carried out over a longer period of time to see if there is a more obvious trend in decline in mitochondrial activity.

LDH assay

The use of LDH assays in determination of cell death is widely established[301]. This assay measures the amount of lactate dehydrogenase in the supernatant. When the plasma membrane is damaged LDH is released into the extracellular space and used as an indication of percentage of cell death[301]. Other

types of cytotoxic assays include Neutral Red which has been reported as being less sensitive in comparison to LDH[301] and ATP content, which is known to be a very sensitive assay but fluctuations in temperature and small changes during the incubation period can lead to misleading results[301].

Cytotoxicity screening

To ensure the selected test compounds were not cytotoxic an MTS assay was conducted. A study carried out by Funada[302] showed that, in a concentration dependent manner, MAM-2201 was cytotoxic and caused cell apoptosis[302]. Apoptosis differs to necrosis as the former is generally viewed as a passive event whereas necrosis is associated with acute toxic injury[303]. Figure 6.8 shows that the addition of the test compounds (with the exception of compound **6**) to the CHO cells had no significant negative effect on mitochondrial activity. This indicates that over a short period of time these compounds are not cytotoxic. However, to further validate this an LDH assay should be conducted. Compound **6** did show significant increased mitochondrial activity in comparison to the negative control (cells only), this could indicate that the presence of this compound is a causing a response, but further investigation will need to be conducted to validate this. When comparing the results of the LDH and MTS assays from day 1-3 the absorbance values for MTS are relatively high and are not significantly decreasing, the percentage of LDH release is increasing which suggest the cells are growing. After day 6 the absorbance values for MTS start to increase and the percentage of LDH release is significantly higher which indicated the cells are actively dying.

cAMP ELISA

Quantification of cAMP is a widely used method to explore the functionality of synthetic cannabinoid agonists[304], as the compounds screened were selected based on physicochemical similarities to known SC. It is assumed the compounds 1-13 would act as SC agonists. Many *in vitro* assays conducted look at the inhibition of forskolin induced cAMP production[285,305-41], as the activation of the CB1 receptor has been implicated in the accumulation of cAMP.

The compounds identified in Chapter 5 were done so on the basis that the chemical structure would interact with the CB1. However, information regarding the nature in which the ligand would bind was unknown. The majority of psychoactive CB1 ligands are full agonists[97,306,307] however there are a small number of selective psychoactive CB1 ligands that have been identified as antagonists (SR141716A[308] and LY320135[305]) and the role these drugs play in cAMP production is different to the CB1 agonists, which could indicate although these compounds were identified as CB1 agonists they may behave as CB1 antagonists

Figure 6.9 shows that compound **1** significantly increases the production of cAMP. As this study was not conducted using forskolin induced cAMP production, this may indicate that activation of the CB1 receptor is occurring[273]. Additionally, some studies have shown that CB1 inverse agonists are known to increase cAMP production[308,309]. The significant increase in cAMP due to compound **1** does indicate that there is an interaction occurring at the CB1 receptor but the compound's mechanism of action will need to be further evaluated with binding assay studies.

EFS study

Activation of the CB1 receptor has been widely established in the reduction of gastrointestinal motility in vivo[276,134]. Compounds such as AM251, AM258, HU210 and THC have been reported in the literature as inhibitors of smooth muscle contraction[134]. The aim of this study was to see if any of the compounds identified in Chapter 5 activated the CB1 receptors located in rat ileum. Results from this would further validate the cAMP assay results and indicate that the selected compounds interacted with the CB1 receptor and could also potentially elicit psychoactivity.

A number of pharmacological studies conducted using EFS and rat ileum have been reported when validating the efficacy of known CB1 receptor ligands[269,268]. The mechanism in which CB1 agonists reduce smooth muscle contractility has been debated. One mechanism that has been proposed is related to reduction of acetylcholine release [268]. The results obtained were done so using only one concentration of compound. Therefore, to further validate this particular study a dose dependant study should be carried out. The study should also be repeated. However, due to the available amount of compound and tissue samples available this was not possible. This should be considered for future work. When comparing the results obtained from the MTS cytotoxicity study to the EFS results, there does not appear to be a link between significantly different MTS values and EFS values. To further validate this observation more data needs to be obtained from both the EFS and MTS cytotoxicity studies.

A study conducted by Izzo et al[267] showed that the cannabinoid agonists WIN55,212-2 and CP55,940 decreased muscle reflex contraction. Both compounds are known to be selective full CB1 agonists[306,307]. The EFS study conducted was a preliminary investigation to see if any of the hit compounds would have an effect on isolated smooth muscle tissue. A negative control test using only the tissue sample and EFS was used as a bench mark to establish to what extent the tissue would contract without the addition of any chemicals. The positive control, MAM-2201, was used to see if this synthetic cannabinoid inhibited contraction.

The negative control showed a contraction of 2.942 g at 10 Hz. Figure 6.12 (**D**) shows the magnitude of the contraction and, that directly after the application of EFS, the tissue was contracting and relaxing which indicates the EFS pulse was strong enough to cause residual contractions.

MAM-2201 did show a reduction in contraction in comparison to the negative control, after the EFS was applied at 10 Hz. Figure 6.12 (**C**) shows that the contraction at this frequency exerted a contractile force of 0.762 g which is considerably smaller than the negative control. This indicates that MAM-2201 inhibits gut wall muscle contraction.

Compound **1** inhibited the initial relaxation phase after direct application of EFS and the contractile force at 10 Hz was 0.636 g which is slightly less than MAM-2201. This may indicate that compound **1** is able to inhibit contractions to a greater extent in comparison to the positive control. Compound **11** also showed signs of contractile inhibition with a contraction force of 0.846 g. However, after the application of EFS the muscle tissue did not have the same contraction pattern to the positive control or compound **1** (see Figure 6.12, **B**). This could indicate that the relaxation phase after the application of EFS is completely inhibited by compound **11**. Both compounds display signs of contraction inhibition which indicates activation of the CB1 receptor. Compound 13 appeared to initially inhibit contraction after the application of EFS, but as the contraction values in grams do not vary substantially when the frequency is increased this indicates that the tissue may not have been viable at this stage, again reinforcing the need to further investigate this particular study.

Compound **1** and **11** were docked into the same binding site in CB1 receptor, the docked configurations were analysed and compared to MAM-2201. There are a number of amino acid residues (Met103, Gly166, Phe170, and Val196)[97] that are integral to CB1 binding. Protein Ligand Interaction Fingerprints (PLIF) were constructed (see figure 6.13) for the docked poses and these showed that compound **1** and **11** are forming interactions with residues that have been reported in the literature as being fundamental to protein-ligand interaction at the CB1 binding site[97]. Although this does imply that a protein ligand interaction is likely to occur between the CB1receptor and compounds **1** and **11** in order to unequivocally confirm this a co-crystallised model of the protein and the ligands will need to be obtained. These compounds were identified using pharmacophore models based on known SC, they were selected on being structurally different to known SC however they do share common chemical functionalities with ligands known to bind to the CB1 receptor. Due to the systematic identification of these compounds the expected result of the cAMP assay was that the test compounds would interact with the CB1 receptor.
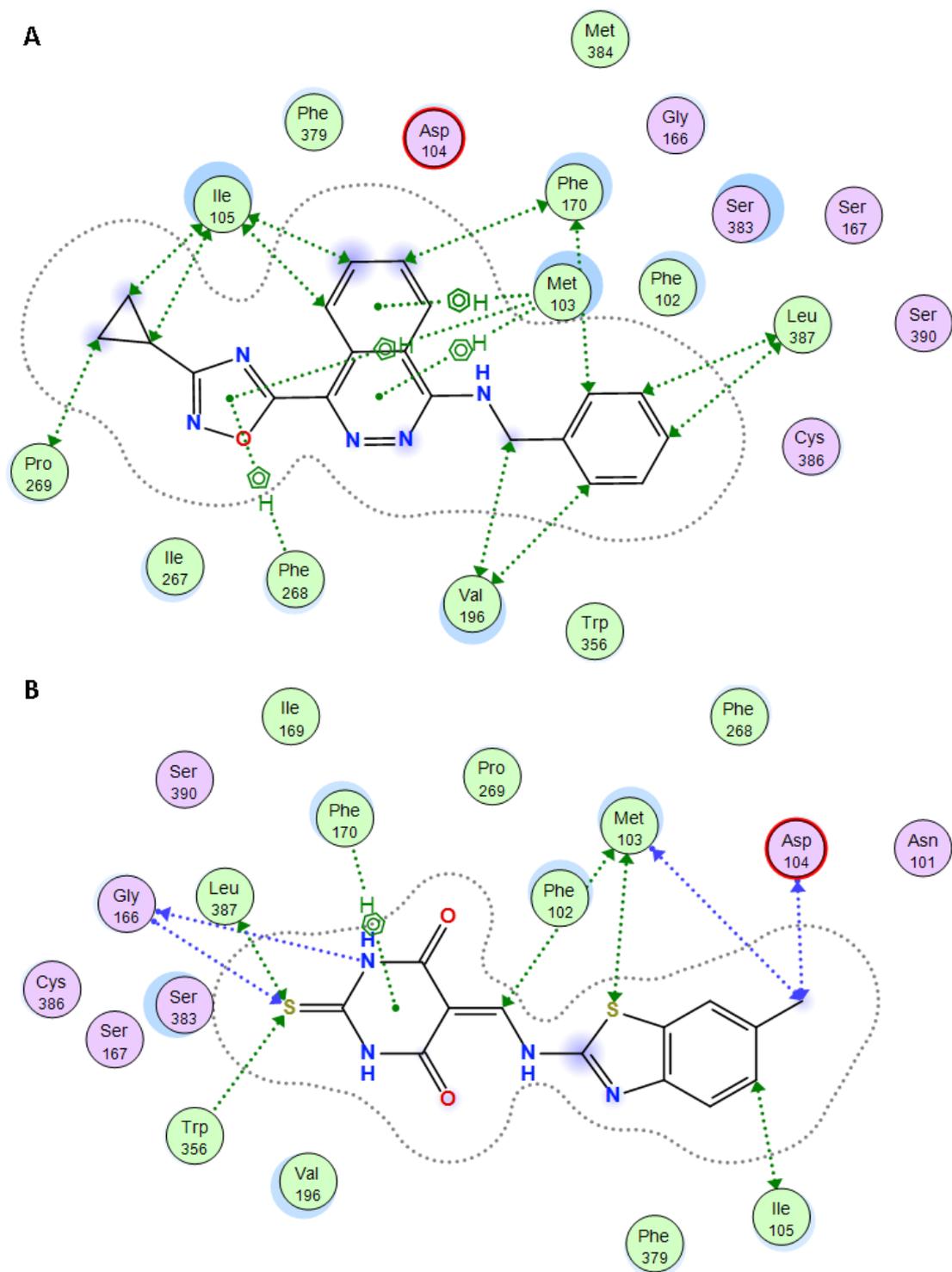
Figure 6. 13: Protein Ligand Interaction Fingerprints (PLIF) for Compound 1 (**A**) and Compound 11 (**B**). The compounds were docked into the CB1 receptor (accession code 5TGZ) using the London ΔG and GBVI/WSA ΔG scoring functions, Compound 1 had a scored value of -8.29 Kcalmol^{-1} and Compound 11 had a scored value of -6.72 Kcalmol^{-1}.

6.6 Conclusions

The aim of this study was to identify if any of the compounds from the virtual screen in Chapter 5 would exhibit similar biological activity as known synthetic cannabinoids. Using CHO-CNR1 cells that had expression of the CB1 receptor a cAMP ELISA was conducted and the results indicate that compound **1** was able to increase the production of cAMP in comparison to the positive control (MAM-2201). An increase in cAMP would indicate that the compounds identified are activating the CB1 receptor in a similar mechanism to known SC and this could indicate that these compounds have the ability to cause a psychoactive effect. Initially the compounds identified were thought to act as CB1 agonists. From the docking study conducted on these compounds they appear to bind to the same region as the known SC antagonist AM6538[97], but also form interactions with residues that are associated with the SC agonist JWH-018. Further work would be required to identify the mechanism of action for both Compound **1** and **11.**

It should be noted that both *in vitro* studies carried out were preliminary and would need to be further investigated before confirming these compounds for an interaction with the CB1 receptor.

From previous docking studies conducted on the known synthetic cannabinoids compound **1** forms interactions with residues in the CB1 receptor agonists that are common to known potent synthetic cannabinoids (JWH-018 and MAM-2201). To ascertain where compound **1** would elicit psychoactivity *in vivo* studies would need to be carried out.

The preliminary EFS study conducted did indicate that there was inhibition of smooth muscle contraction which is consistent with a physiological effect displayed by a known CB1 agonist.

Compound **1** is displaying the type of behaviour that is consistent for known SC, this indicates promising results that this compound is a truly novel SC. However, to identify the compound as either agonist or an antagonist a more extensive experiment would need to be conducted in order to further validate the results.

Chapter 7

General Conclusions and Future Work.

7.1 General Conclusions

At the outset, the intention of the studies presented in this thesis was to determine whether or not *in Silico* methodologies, which had been successfully used in other areas to understand the subtleties and complexities of protein-ligand interaction, were transferable to studies on New Psychoactive Substances, and their associated receptors in order to increase understanding in this area.

For example, in this project, the use of molecular docking was employed to establish whether *in Silico* methodologies could successfully rank a library of NPS in such a way that the computational results would mirror the biological potency values established through *in vitro* binding assays. Through the investigation of a number of different scoring functions and the implementation of consensus scoring, this was achieved for two of the three MAT isoforms investigated.

It is well established that a number of monoamine transporter inhibitors have varying affinities for all three MAT isoforms, DAT, NET and SERT[310]. A wide range of NPS are also known to be MAT inhibitors, and as such they exhibit a large degree of promiscuity between DAT, NET and SERT. An investigation of the MATs binding site composition was conducted (using homology models as no experimental structures were available at the time) to identify differences in structure between the isoforms, particularly with reference to their putative binding cavities, that would explain how selectivity of a given NPS arose from one isoform to the next. Initial docking studies conducted to compare biologically derived data to computational data yielded statistically significant results for DAT only. This lead to the utilisation of a wide-range of different scoring functions and ultimately employment of a consensus of scoring functions, derived from two different docking algorithms and the use of an algorithm that rescored poses generated by other scoring functions, in order to improve the initial results by virtue of ameliorating the inherent bias of a single scoring function, which arises as a result of approximations in force-fields and/or the necessarily limited nature of datasets used to derive them. Following the iterative refinements, significant results were obtained for both DAT and SERT using consensus scoring.

However, docking studies using the NET homology model and the application of different scoring functions/consensus approaches were unable to generate results that showed significant correlation between biological measurements and computational predictions. This was hypothesised to be most likely due to the small dataset and the skew of the distribution of biological activities from normal, for the NET compounds.

The underlying reason for observed promiscuity of NPS between the MATs was also investigated. Results from Chapter 3 have shown that there is a high degree of structural and topological similarity between the MAT homology models. This lack of structural difference present between the MAT is

likely to have been one of the factors that made it difficult to isolate any substantial structural differences between the MAT which could then have been rationalised to explain the selectivity of NPS between one MAT isoform and the next.

The homology models were all built using the same template, and the inherent limitations with this is such that a difference in the overall fold of the isoforms would not be expected given that they were built on the same template. This means that subtle differences in topology that occur naturally between the MATs would not necessarily be apparent in the models. Complementary, yet independent protein validation techniques were used to probe any issues of this nature, and showed that in the absence of experimental structures each of the models were valid to use for the investigations i.e. there were no major structural defects identified which would be cause for worry in docking experiments. However, as all models are approximations, it would be naïve to assume that there were no intrinsic limitations in using models to conduct these investigations rather than high-quality experimentally determined structures.

However, during the course of these studies, newly published crystal structures (for DAT, PDB Accession Code 4XP9 and SERT, PDB Accession Code 5I75) became available and a rigorous comparative analysis was conducted to try and identify structural differences between these crystal structures and the homology models that could better explain the selectivity profiles of the NPS investigated for these isoforms. The crystal structures were subjected to the same structural quality checks as the homology models, and unsurprisingly performed to a higher standard. Hence, all crystal structures were deemed appropriate for docking studies. Comparison of the results from the docking studies using the homology models with the docking studies using the crystal structures revealed differences in the positioning of highly ranked docked poses in the crystal structures which gave rise to an improvement in the correlation between observed biological activity and predicted computational ranking for the crystal structure experiments compared to the homology model docking experiments.

Comparison of the SERT structure and model highlighted a shift in protein backbone which gave rise to a binding site that was a different shape to that seen in the homology model. The cavity in the crystal structure was more restrictive, due to a narrower opening to the site, which placed the majority of the highly ranked docking poses in the bottom of the SERT cavity for the crystal structure – whereas some highly ranked poses in the homology model showed interaction of the small molecule with residues that define the entrance to the cavity. An identical comparative study was conducted on the newly published DAT crystal structure a slight improvement in docking results, for similar reasons (i.e. a more restricted binding cavity) over the homology modelling studies. However, the crystal structures

showed there were no obvious structural differences identified in the binding cavities between the DAT and SERT isoforms that could be used to explain the differences in affinity for a given NPS.

There was no crystal structure for NET published. However, a newer NET homology model (4XP4a) was published, and this was used to highlight differences between it and the NET homology model used in the initial experiments. Although this model (4xP4a) gave rise to a larger binding site it did not significantly improve the docking results.

These studies showed the MATS were very similar in topology and structure, which may be the reason as to why a docking algorithm was unable to definitively identify the subtle differences that could result in selectivity. Even high-quality crystal structure did provide conclusive insight into the key protein-ligand interactions being formed between NPS and the MATs to explain the reasons behind isoform selectivity, although useful insights into the key interactions between NPS and the MAT isoforms were identified. Similarities in the DAT and SERT binding cavities highlighted a plausible reason as to why there is such a high level of promiscuity amongst MATs. As such, alternative approaches to investigate selectivity were needed.

Where Chapter 3 detailed a structure-based approach to understanding selectivity, the inherent limitations of scoring functions, homology models and the structural similarity between MATs afforded modest results in trying to understand what gives rises to selectivity amongst MAT.

Therefore, the next step was to prosecute a ligand-based approach, which would switch the focus to the small molecules that bound to the MATS and would utilise an understanding of the physicochemical properties of the NPS to gain insight into selectivity. Predictive computational models were successfully generated to identify the key physicochemical properties required to explain the differences in affinity for known NPS binding to the three different monoamine transporters. This use of QSAR models provided further insight into the properties required of NPS with respect to forming protein-ligand interaction with DAT, NET or SERT.

QSAR models have been exploited for specific groups of drugs such as benzodiazepines[311], cathinones[312] and tricyclic antidepressants[313] however up until the work described in this thesis there was no study conducted on specific monoamine transporter isoforms. A number of QSAR models for DAT, NET and SERT that incorporated models with varying numbers of molecular descriptors were investigated as part of these studies, and the models that yielded the highest r^2 and q^2 values for each isoform were analysed for generalizability using NPS test sets.

A 3-descriptor DAT QSAR model was built using the descriptors b_max1len, FASA_H and opr_leadlike, which account for steric, hydrophobic and drug-like physicochemical properties respectively. The

highest-performing model developed for SERT also included three molecular descriptors PEOE_VSA-0, a_don and E_tor, which describe hydrophobic, electronic and steric properties respectively. Although the molecular descriptors for DAT and SERT are different, similarities between the two sets of molecular descriptors can be extrapolated. The steric descriptors used in the DAT and SERT QSAR models both described a preference for relatively inflexible molecules. Additionally, both equations identified hydrophobic interactions being integral for determining activity for DAT and SERT. The differences in descriptors for DAT and SERT that did arise included the fact that DAT selective molecules should not adhere to opr_leadlike criteria. Investigating this in greater detail suggested that this term was in fact a simple correction factor. As with the structure-based approach, although the exact descriptors in each of the QSAR models were different, the properties they were describing were comparable, yet further evidence to support the observed promiscuous binding of the NPS between these two MAT isoforms. The DAT and SERT pKi values were normally distributed and from this robust and generalisable models were built for these two isoforms.

The four molecular descriptors associated with the NET QSAR model only accounted for electronic properties and did not produce a robust or generalisable QSAR. A potential reason for this could be the dataset used in the development of the NET model. The biological data available for this isoform was not normally distributed, unlike DAT and SERT, and this may have limited the ability to generate any generalisable QSAR model built based on the NET pK_i values, despite the steps that were taken to ensure that the training and test sets were representative of the dataset as a whole.

Data obtained from Chapter 3 highlighted that the MAT binding cavities are not significantly different so it may be expected that all three isoforms would produce very similar QSAR equations. For the two isoforms that produced predictive models (DAT and SERT) it would appear on first inspection that this was not the case given that the investigations showed that different descriptors were identified for the QSAR models for each isoform. However, when these descriptors were examined in greater depth, it was found that the properties they are describing are closely related – hence the QSAR models are similar. The information obtained from Chapter 4 provided a deeper insight into the type of interactions formed between NPS and MAT, and helped to improve the understanding around why there is a degree of promiscuity associated with NPS and the MAT isoforms.

Having identified that molecular modelling techniques were applicable to aiding understanding of the NPS, attention was then focussed on using them to identify previously unexploited chemical scaffolds with the potential to be developed into the next generation of NPS. The development of pharmacophore models based on known synthetic cannabinoids in combinations with a systematic filtering process allowed for the identification of a library of completely novel chemical structures that

were predicted to have the ability to interact with the CB1 receptor and possibly elicit psychoactive properties.

A ligand-based approach was utilised in Chapter 5 to identify these pharmacophores and subsequently potential new scaffolds for SC. The methodology employed was developed to ensure that novel molecules identified using virtual high throughput screening were structurally dissimilar to known SC, but retained both the predicted ability to bind to the CB1 receptor, and a range of physicochemical properties similar to known SC, so as to maximise the possibility of the compound binding to the CB1 receptor. To this end, the systematic methodology developed in Chapter 5 outlines a thorough workflow that can be applied to maximise the possibility of identifying active compounds against the CB1 receptor whilst maintaining complete novelty in chemical structure.

A limitation of this study was the number of molecules available per class of structurally dissimilar synthetic cannabinoid. Pharmacophore models could only be built for three different SC classes, which meant that the broad range of structural diversity that is present within the SC was not exploited to its full potential. This is likely to have reduced the number, and the structural diversity, of virtual hits returned by the screening process. However, a large number of structurally diverse molecules were returned from the screens with the pharmacophores that had been generated, and as such the experiment was successful.

Shortly after this study was conducted the human CB1 crystal structure was published. Docking of the virtual hits into the binding cavity of CB1 was used as an additional filter, and to prioritise which compounds were purchased for the biological testing experiments. Moving forwards, the availability of an experimentally derived structure allows for a structure-based approach to be conducted, this is an avenue that should be further explored for novel CB1 agonists/antagonists.

The successful outcome of the virtual screening experiments led to the purchase of a number of compounds with the potential to bind to the CB1 receptor. Both *in vitro* and *ex vivo* techniques were conducted to evaluate the biological activity of the novel compounds identified. The use of cAMP biological assays provided quantitative data that showed the majority of the "hit" compounds identified did have an effect on the cAMP production of CHO cells that expressed the CB1 receptor. This could be an indication of an interaction with the cannabinoid receptor being formed. However, it should be noted that the experiments conducted in Chapter 6 do not categorically show that the CB1 receptor has been activated by the compounds screened.

The cAMP assay conducted was done so without the presence of Forskolin which has previously been employed in many CB1 activity-based studies, Forskolin was not employed due to financial and time

constraints. This study was conducted as preliminary investigation to see if the hits screened showed any interaction with the CB1 receptor by altering the amount of cAMP present.

To provide a definitive answer on the biological activity of these compounds a more sophisticated binding assay would be required. Future work could include the use of a radiolabelled binding assay to provide quantitative data for the hit compounds that are shown to impact on cAMP production.

As there were limitations associated with the cell-based assay employed, a secondary biological experiment was conducted using whole tissues samples. This approach was used as there are established methodologies that have been published using cannabinoids and ileum tissue, and to provide a complementary study for the non-definitive cell-based assay.

The preliminary Electrical Field Stimulations (EFS) did show that in the presence of compound **1** and compound **12** the tissue sample demonstrated an increased contraction period, which is consistent with the activation of the CB1 receptors commonly found in ileum tissue. Again, the EFS study conducted does not explicitly prove that the virtual hits are interacting with the CB1 receptor but does act as additional evidence, alongside the results from the cell-based assay that these hits are causing a biological response that could be due to an interaction of the virtual hits with the CB1 receptor.

Therefore, the preliminary data has shown that two of the thirteen compounds screened show greater affinity to the CB1 receptor in comparison to a known potent SC. Traditional high-throughput virtual screening (HTVS) projects will achieve approximate hit rates of 0.1%[314]. The *in Silico* methodology employed in this study has showed a potential hit rate of 15.4% which is substantially greater than typical HTVS experiments, and demonstrates the power of the tool as an agent for identifying bioactive molecules.

Overall, this research has shown that *in Silico* techniques are applicable in the rationalisation of observed biological activity for the NPS and in the development of wholly novel potential NPS., given that computational techniques made it possible to systematically search a library of over 17 million compounds and produce a small library of compounds some of which displayed promising results with respect to biological activity in two independent, yet complementary, studies.

7.2 Future Work

The biological experimental data obtained highlighted two compounds that show a likely interaction with the CB1 receptor. Future work will be focused on the development of a complete pharmacological profile for each of the compounds. The utilisation of a newly developed G-protein coupled receptor (GPCR) activation assay based on NanoLuc binary technology[4] can be applied to profile biological activity of the molecules identified in Chapter 5. In addition to this, a more extensive study using rat ileum and known CB1 antagonist would reveal the mechanism of action of identified hits, and would establish whether the molecules are in fact CB1 agonists.

Future work conducted on compound 1 and compound 12, will produce a series of synthetic analogues that can then be investigated for their affinity with the CB1 receptor. Molecules of interest will be investigated for viable synthetic routes and the compounds will be synthesised and characterised using standard analytical techniques such as nuclear magnetic resonance (NMR) and liquid chromatography mass spectroscopy (LCMS). The characterised molecules will then be examined for biological activity using cAMP ELISA using the same methodology as outlined in Chapter 6. Structure activity relationships (SAR) can be elucidated that will provide information on how to produce molecules with an increased potency, and potentially selectivity, for the CB1 receptor. *De novo* design strategies can also be employed alongside the virtual screening results to identify further novel scaffolds that have the potential to be developed into SC.

Chapter 5 highlighted 35 compounds that when docked into the CB1 receptor had a greater binding score than a known potent SC (MAM-2201). Due to financial and time constraints these compounds could not be acquired, so future work will also look into investigating these compounds. If funds can be identified, it would be of interest to repeat the initial cell-based assays in conjunction with Forskolin, as the use of this compound has been well documented in the literature in determining the activity of potential CB1 agonists. In depth EFS studies will also be conducted on the synthesised molecules to provide more evidence that an interaction at the CB1 receptor is occurring. Radioligand binding assays can also be conducted to provide a full biological profile and prove definitively that the synthesized molecules are biologically active. The most potent molecule can then be subjected to further development to produce a library of analogues. Which can then be docked into the crystal structure of the CB1 receptor to provide information on which analogues are most likely to form and interaction and therefore be an indicator of potential psychoactivity.

Finally, another avenue of research that can be investigated is structure-based drug design, utilising the recently published CB1 crystal structure and all of the known SC to generate a consensus pharmacophore based on the interactions of all of the different categories of SC and not just those

with large enough numbers to be exploited via ligand-based strategies. This approach for developing novel SC could lead to the identification of compounds with distinct chemical scaffolds additional to those already identified in this thesis, and provide further routes for investigation into the next generation of new psychoactive substances.

References

(1)   Degenhardt, L.; Whiteford, H. A.; Ferrari, A. J.; Baxter, A. J.; Charlson, F. J.; Hall, W. D.; Freedman, G.; Burstein, R.; Johns, N.; Engell, R. E.; et al. Global Burden of Disease Attributable to Illicit Drug Use and Dependence: Findings from the Global Burden of Disease Study 2010. *Lancet* **2013**, *382*, 1564–1574.

(2)   Strang, J.; Griffiths, P.; Gossop, M. Heroin Smoking by "Chasing the Dragon": Origins and History. *Addiction* **1997**, *92*, 673–683.

(3)   E.M.C.D.D.A., European Monitoring Centre for Drugs and Drug Addiction. (2018). Europeasn Drug Report: Office for Official Publications of the European Communities,

(4)   United Nations Office on Drugs and Crime. *ANALYSIS OF DRUG MARKETS Opiates, Cocaine, Cannabis, Synthetic Drugs*; Vienna, 2018.

(5)   Jekel, J. F.; Allen, D. F.; H, M. P. Trends in Drug Abuse in the Mid-1980s. *Yale J. Biol. Med.* **1987**, *60*, 45–52.

(6)   Sordo, L.; Indave, B. I.; Barrio, G.; Degenhardt, L.; Fuente, L. De; Bravo, M. J. Cocaine Use and Risk of Stroke : A Systematic Review �. *Drug Alcohol Depend.* **2014**, *142*, 1–13.

(7)   Feng, L.; Battulga, A.; Han, E.; Chung, H. New Psychoactive Substances of Natural Origin : A Brief Review. *J. Food Drug Anal.* **2017**, *25* (3), 461–471.

(8)   United Nations Office on Drugs and Crime. *SINGLE CONVENTION ON NARCOTIC DRUGS , 1961*; 1972.

(9)   Misuse of Drugs Act 1979, c. 38. Available at: https://www.legislation.gov.uk/ukpga/1971/38 (Accessed: 30 January 2018).

(10)   Drugs (Prevention of Misuse), Act and Subsidiary Legislation 2002, c.4.07. Available at: http://agc.gov.ms/wp-content/uploads/2010/02/drugs_prevention_of_misuse_act.pdf (Accessed: 13 March 2019).

(11)   Coomber, R. How Often Does the Dilution of Heroin Actually Occur. *Int. J. Drug Policy* **1997**, *8* (4), 178–189.

(12)   Coomber, R. The Adulteration of Drugs: What Dealers Do to Illicit Drugs, and What They Think Is Done to Them. *Addict. Res. Theory* **1997**, *5*, 297–306.

(13)   Coomber, R. Vim In The Veins - Fantasy or Fact: The Adulteration of Illicit Drugs. *Addict. Res.* **1997**, *5* (3), 195–212.

(14)   Coomber, R. Adulteration of Drugs : The Discovery of a Myth. *Contemp. Drug Probl.* **1997**, *24*, 239–271.

(15)   Preble, E.; Casey, J. J. Taking Care of Business—The Heroin User's Life on the Street. *Int. J. Addict.* **1969**, *4* (1), 1–24.

(16)   Davidson, C. New Psychoactive Substances. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2012**, *39* (2), 219–220.

(17)   Deluca, P.; Davey, Z.; Corazza, O.; Di, L.; Farre, M.; Holmefjord, L.; Scherbaum, N.; Siemann, H.; Skutle, A.; Torrens, M.; et al. Identifying Emerging Trends in Recreational Drug Use; Outcomes from the Psychonaut Web Mapping Project. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2012**, *39* (2), 221–226.

(18)   European Monitoring Centre for Drugs and Drug Addiction. *Action on new drugs* http://www.emcdda.europa.eu/activities/action-on-new-drugs.

(19)   *Psychoactive Substances Act 2016*, c.2. Available at: http://www.legislation.gov.uk/ukpga/2016/2/contents/enacted (Accessed 13 March 2019).

(20)   Norman, C.; Lee, G. A. A Snapshot of Novel Psychoactive Substances (Legal Highs) Use in London. *Int. Emerg. Nurs.* **2017**, *35*, 56–58.

(21)   Tracy, D. K.; Wood, D. M.; Baumeister, D. Novel Psychoactive Substances : Types , Mechanisms of Action , and Effects. **2017**, *6848* (January), 1–8.

(22)   United Nations Office on Drugs and Crime. *MARKET ANALYSIS OF SYNTHETIC DRUGS Amphetamine-Type Stimulants, New Psychoactive Substances DRUG*; 2017.

(23)   Elliott, S.; Evans, J. A 3-Year Review of New Psychoactive Substances in Casework. *Forensic Sci. Int.* **2014**, *243*, 55–60.

(24)   Liechti, M. E. Novel Psychoactive Substances (Designer Drugs): Overview and Pharmacology of Modulators of Monoamine Signalling. **2015**, No. January, 1–12.

(25)   EMCDDA. European Drug Report Trends and Developments. *Eur. Monit. Drugs Drugs Addict.* **2015**, 1–86.

(26)   EMCDDA. New Psychoactive Substances in Europe. *Eur. Monit. Drugs Drugs Addict.* **2015**, *1*, 1–10.

(27)   Bröer, S.; Gether, U. The Solute Carrier 6 Family of Transporters. *Br. J. Pharmacol.* **2012**, *167*

(2), 256–278.

(28)   Paczkowski, F. A.; Sharpe, I. A.; Dutertre, S.; Lewis, R. J. X-Conotoxin and Tricyclic Antidepressant Interactions At the Norepinephrine Transporter Define a New Transporter Model. *J. Biol. Chem.* **2007**, *282* (24), 17837–17844.

(29)   Zhou, J. Norepinephrine Transporter Inhibitors and Their Therapeutic Potential. *Drugs Futur.* **2004**, *29* (12), 1235–1244.

(30)   Baumann, M. H.; Partilla, J. S.; Lehner, K. R.; Thorndike, E. B.; Hoffman, A. F.; Holy, M.; Rothman, R. B.; Goldberg, S. R.; Lupica, C. R.; Sitte, H. H.; et al. Powerful Cocaine-like Actions of 3,4-Methylenedioxypyrovalerone (MDPV), a Principal Constituent of Psychoactive "bath Salts" Products. *Neuropsychopharmacology* **2013**, *38* (4), 552–562.

(31)   Lin, Z.; Canales, J. J.; Björgvinsson, T.; Thomsen, M. M.; Liu, Q.; Torres, G. E.; Caine, S. B. *Monoamine Transporters: Vulnerable and Vital Door Keepers*; 2012.

(32)   Solich, J.; Kolasa, M.; Kusmider, M.; Faron-gorecka, A.; Pabian, P.; Zurawek, D.; Szafran-pilch, K.; Dziedzicka-wasylewska, M. Norepinephrine Transporter Knock-out Alters Expression of the Genes Connected with Antidepressant Drugs Action. *Brain Res.* **2015**, *1594*, 284–292.

(33)   Manepalli, S.; Surratt, C. K.; Madura, J. D.; Nolan, T. L. Monoamine Transporter Structure, Function, Dynamics, and Drug Discovery: A Computational Perspective. *AAPS J.* **2012**, *14* (4), 820–831.

(34)   Zhou, Z.; Zhen, J.; Karpowich, N. K.; Law, C. J.; Reith, M. E. a; Wang, D.-N. Antidepressant Specificity of Serotonin Transporter Suggested by Three LeuT-SSRI Structures. *Nat. Struct. Mol. Biol.* **2009**, *16* (6), 652–657.

(35)   Torres, G. E.; Gainetdinov, R. R.; Caron, M. G. Plasma Membrane Monoamine Transporters: Structure, Regulation and Function. *Nat. Rev. Neurosci.* **2003**, *4* (1), 13–25.

(36)   Hondebrink, L.; Zwartsen, A.; Westerink, R. H. S. Effect Fingerprinting of New Psychoactive Substances (NPS): What Can We Learn from in Vitro Data ? *Pharmacol. Ther.* **2018**, *182* (October 2017), 193–224.

(37)   Reith, M. E. a.; Blough, B. E.; Hong, W. C.; Jones, K. T.; Schmitt, K. C.; Baumann, M. H.; Partilla, J. S.; Rothman, R. B.; Katz, J. L. Behavioral, Biological, and Chemical Perspectives on Atypical Agents Targeting the Dopamine Transporter. *Drug Alcohol Depend.* **2015**, *147*, 1–19.

(38)   Vaughan, R. a.; Foster, J. D. Mechanisms of Dopamine Transporter Regulation in Normal and

Disease States. *Trends Pharmacol. Sci.* **2013**, *34* (9), 489–496.

(39)   Kristensen, A. S.; Andersen, J.; Jørgensen, T. N.; Sørensen, L.; Eriksen, J.; Loland, C. J.; Strømgaard, K. SLC6 Neurotransmitter Transporters : Structure , Function , and Regulation. **2011**, *63* (3), 585–640.

(40)   Wang, C.-I. a; Shaikh, N. H.; Ramu, S.; Lewis, R. J. A Second Extracellular Site Is Required for Norepinephrine Transport by the Human Norepinephrine Transporter. *Mol. Pharmacol.* **2012**, *82* (5), 898–909.

(41)   Borgkvist, A.; Malmlöf, T.; Feltmann, K.; Lindskog, M.; Schilström, B. Dopamine in the Hippocampus Is Cleared by the Norepinephrine Transporter. *Int. J. Neuropsychopharmacol.* **2011**, 1–10.

(42)   Olivier, B. Serotonin: A Never-Ending Story. *Eur. J. Pharmacol.* **2015**, *753*, 2–18.

(43)   Sitte, H. H.; Freissmuth, M. Amphetamines , New Psychoactive Drugs and the Monoamine Transporter Cycle. *Trends Pharmacol. Sci.* **2015**, *36* (1), 41–50.

(44)   Corazza, O.; Demetrovics, Z.; Brink, W. Van Den; Schifano, F. ' Legal Highs ' an Inappropriate Term for ' Novel Psychoactive Drugs ' in Drug Prevention and Scientific Debate. *Int. J. Drug Policy* **2013**, *24* (1), 82–83.

(45)   Drugs, P. O. N. The Misuse of Benzodiazepines among High-Risk Opioid Users in Europe.

(46)   German, C. L.; Fleckenstein, A. E.; Hanson, G. R. Bath Salts and Synthetic Cathinones : An Emerging Designer Drug Phenomenon. *Life Sci.* **2014**, *97* (1), 2–8.

(47)   Coulon, P.; Gorji, A. Tightrope or Slackline? The Neuroscience of Psychoactive Substances. *Trends Pharmacol. Sci.* **2016**, *37* (7), 1–11.

(48)   Beharry, S.; Gibbons, S. An Overview of Emerging and New Psychoactive Substances in the United Kingdom. *Forensic Sci. Int.* **2016**, *267* (May), 25–34.

(49)   *The Misuse of Drugs Act 1971 (Amendment) Order 2010*; 2010, c.38. Available at: http://www.legislation.gov.uk/uksi/2010/1207/introduction/made (Accesses 13th March 2019).

(50)   Zawilska, J. B.; Andrzejczak, D. Next Generation of Novel Psychoactive Substances on the Horizon - A Complex Problem to Face. *Drug Alcohol Depend.* **2015**, *157*, 1–17.

(51)   E.M.C.D.D.A., European Monitoring Centre for Drugs and Drug Addiction. (**2015**). European Drug Report: Office for Official Publications of the European Communities.

(52)   Miliano, C.; Serpelloni, G.; Rimondo, C.; Mereu, M.; Marti, M.; De Luca, M. A. Neuropharmacology of New Psychoactive Substances (NPS): Focus on the Rewarding and Reinforcing Properties of Cannabimimetics and Amphetamine-Like Stimulants. *Front. Neurosci.* **2016**, *10* (April), 153.

(53)   Castaneto, M. S.; Gorelick, D. A.; Desrosiers, N. A.; Hartman, R. L.; Pirard, S.; Huestis, M. A. Synthetic Cannabinoids: Epidemiology, Pharmacodynamics, and Clinical Implications. *Drug Alcohol Depend.* **2014**, *144*, 12–41.

(54)   Blackman, S.; Bradley, R. From Niche to Stigma — Headshops to Prison : Exploring the Rise and Fall of Synthetic Cannabinoid Use among Young Adults. *Int. J. Drug Policy* **2017**, *40*, 70–77.

(55)   Zawilska, J. B.; Andrzejczak, D. Next Generation of Novel Psychoactive Substances on the Horizon - A Complex Problem to Face. *Drug and Alcohol Dependence*. 2015.

(56)   Maskell, P. D.; Smith, P. R.; Cole, R.; Hikin, L.; Morley, S. R. Seven Fatalities Associated with Ethylphenidate. *Forensic Sci. Int.* **2016**, *265*, 70–74.

(57)   Corazza, O.; Valeriani, G.; Bersani, F. S.; Corkery, J.; Martinotti, G.; Bersani, G.; Schifano, F. " Spice ," " Kryptonite ," " Black Mamba ": An Overview of Brand Names and Marketing Strategies of Novel Psychoactive Substances. *J. Psychoactive Drugs* **2014**, *46* (4), 37–41.

(58)   Sunkel, C. Trends in New Psychoactive Substances from Surface and " Dark " Net Monitoring. *The Lancet*. 2017, pp 2015–2017.

(59)   Al-imam, A.; Abdulmajeed, B. A. The NPS Phenomenon and the Deep Web : Internet Snapshots of the Darknet and Potentials of Data Mining. *Glob. J. Health Sci.* **2017**, *9* (11), 86–101.

(60)   Soussan, C.; Andersson, M.; Kjellgren, A. The Diverse Reasons for Using Novel Psychoactive Substances - A Qualitative Study of the Users ' Own Perspectives. *Int. J. Drug Policy* **2018**, *52*, 71–78.

(61)   Haden, M.; Wood, D. M.; Dargan, P. I. The Impact of the Psychoactive Substances Act 2016 on the Online Availability of MDMB-CHMICA. **2018**, No. February 2017, 619–622.

(62)   Wadsworth, E.; Drummond, C.; Deluca, P. The Dynamic Environment of Crypto Markets : The Lifespan of New Psychoactive Substances (NPS) and Vendors Selling NPS. *Brain Sci.* **2018**, No. May 2016.

(63)   Markus, T.; Marie, A.; Nefau, T.; Martinez, M.; Lahaie, E.; Malzcewski, A.; Pazitny, M.; Belackova, V.; Brandt, S. D. Online Test Purchased New Psychoactive Substances in 5 Different European Countries : A Snapshot Study of Chemical Composition and Price. *Int. J. Drug Policy* **2017**, *44*, 105–114.

(64)   For, I.; Road, S.; Road, S. Everything You Always Wanted to Know about Drug Cryptomarkets * (* but Were Afraid to Ask). *Int. J. Drug Policy* **2016**, *35*, 1–6.

(65)   Dan Broadfield. *Drug Misuse: Findings from the 2016/17 Crime Survey for England and Wales Further Information*; 2017.

(66)   The Young People's Health Partnership. *NOVEL PSYCHOACTIVE SUBSTANCES INSIGHT REPORT: The View from Young People*; 2017.

(67)   Corkery, J. M.; Durkin, E.; Elliott, S.; Schifano, F.; Ghodse, a. H. The Recreational Tryptamine 5-MeO-DALT (N,N-Diallyl-5-Methoxytryptamine): A Brief Review. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **2012**, *39* (2), 259–262.

(68)   Adamowicz, P.; Zuba, D.; Byrska, B. Fatal Intoxication with 3-Methyl-N-Methylcathinone (3-MMC) and 5-(2-Aminopropyl)Benzofuran (5-APB). *Forensic Sci. Int.* **2014**, *245*, 126–132.

(69)   Raffa, R. B.; Jr, J. V. P.; Ba, J. A. L.; Jr, R. T. The Fentanyl Family : A Distinguished Medical History Tainted by Abuse. **2017**, No. September 2017, 154–158.

(70)   Richard, F.; Pollack, H. Addressing the Fentanyl Threat to Public Health. *N. Engl. J. Med.* **2017**, *7* (376), 2017–2019.

(71)   Torjesen, I. Fentanyl Misuse in the UK : Will We See a Surge in Deaths ? *BMJ* **2018**, *1564* (April), 9–10.

(72)   Grzonkowski, P.; Kacprzak, Ł.; Zawilska, J. B. Abuse of Fentanyl : An Emerging Problem to Face. **2018**, *289*, 207–214.

(73)   Davey, Z.; Corazza, O.; Schifano, F.; Deluca, P. Mass-Information: Mephedrone, Myths, and the New Generation of Legal Highs. *Drugs and Alcohol Today* **2010**, *10* (3), 24–28.

(74)   Shanks, K. G.; Winston, D.; Heidingsfelder, J.; Behonick, G. Case Reports of Synthetic Cannabinoid XLR-11 Associated Fatalities. *Forensic Sci. Int.* **2015**, *252*, e6-9.

(75)   Hout, M. C. Van; Hearne, E. New Psychoactive Substances (NPS) on Cryptomarket Fora : An Exploratory Study of Characteristics of Forum Activity between NPS Buyers and Vendors. *Int. J. Drug Policy* **2017**, *40*, 102–110.

(76)   Gross, M. Drugs: Blanket Ban or Harm Reduction? *Curr. Biol.* **2015**, *25* (13), R523–R525.

(77)   Reuter, P.; Pardo, B. Can New Psychoactive Substances Be Regulated Effectively? An Assessment of the British Psychoactive Substances Bill. *Addiction* **2016**, *10* (11), 1–7.

(78)   Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings. *US Pat. 1,870,942* **2013**, 1–63.

(79)   Vetulani, J. DRUG ADDICTION . PART I . PSYCHOACTIVE SUBSTANCES IN THE PAST AND PRESENCE Psychoactive Substances of the Old World. *Pol. J. Pharmacol.* **2001**, *53*, 201–214.

(80)   EMCDDA. *Perspectives on Drugs Synthetic Cannibinoids in Europe*; Luxembourg, 2015.

(81)   Tai, S.; Fantegrossi, W. E. Synthetic Cannabinoids: Pharmacology, Behavioral Effects, and Abuse Potential. *Curr. Addict. Reports* **2014**, *1* (2), 129–136.

(82)   E.M.C.D.D.A., European Monitoring Centre for Drugs and Drug Addiction. (**2018**). *Fentanils and Synthetic Cannabinoids : Driving Greater Complexity into the Drug Situation*:Office for Official Publications of the European Communities.0

(83)   Warren, K. E.; Phil, M.; Tay, S.; Wen, L. S.; Sc, M. The Role of Public Health in Combatting Synthetic Cannabinoid Use in Adolescents. *J. Adolesc. Heal.* **2017**, *60* (5), 483–486.

(84)   Reuter, P.; Pardo, B.; If, T. D. New Psychoactive Substances: Are There Any Good Options for Regulating New Psychoactive Substances ? *Int. J. Drug Policy* **2017**, *40*, 117–122.

(85)   Brents, L. K.; Prather, P. L. The K2/Spice Phenomenon: Emergence, Identification, Legislation and Metabolic Characterization of Synthetic Cannabinoids in Herbal Incense Products. *Drug Metab. Rev.* **2014**, *46* (1), 72–85.

(86)   Zawilska, J. B.; Wojcieszak, J. Spice/K2 Drugs--More than Innocent Substitutes for Marijuana. *Int. J. Neuropsychopharmacol.* **2014**, *17* (3), 509–525.

(87)   Ralphs, R.; Williams, L.; Askew, R.; Norton, A. Adding Spice to the Porridge 1 : The Development of a Synthetic Cannabinoid Market in an English Prison. *Int. J. Drug Policy* **2017**, *40*, 57–69.

(88)   Langford, A. M.; Bolton, J. R. Synthetic Cannabinoids : Variety Is Definitely Not the Spice of Life. *J. Forensic Leg. Med.* **2018**, *59* (July), 36–38.

(89)   Mackie, K. Cannabinoid Receptors: Where They Are and What They Do. *J. Neuroendocrinol.* **2008**, *20* (SUPPL. 1), 10–14.

(90)   Ogawa, L. M.; Burford, N. T.; Liao, Y.; Scott, C. E.; Hine, A. M.; Dowling, C.; Chin, J.; Power, M.;

Jr, E. J. H.; Emerick, V. L.; et al. Discovery of Selective Cannabinoid CB 2 Receptor Agonists by High-Throughput Screening. *Soc. Lab. Autom. Screen.* **2018**, *23* (4), 373–383.

(91) Martínez-pinilla, E.; Varani, K.; Reyes-resina, I.; Angelats, E.; Vincenzi, F.; Ferreiro-vera, C.; Oyarzabal, J.; Canela, E. I.; Lanciego, J. L.; Nadal, X.; et al. Binding and Signaling Studies Disclose a Potential Allosteric Site for Cannabidiol in Cannabinoid CB 2 Receptors. *Front. Pharmacol.* **2017**, *8*, 1–10.

(92) L, H.-C.; Mackie, K. An Introduction to the Endogeous Cannabinoids System. *Biol. Psychiatry* **2017**, *79* (7), 516–525.

(93) Mallipeddi, S.; Janero, D. R.; Zvonok, N.; Makriyannis, A. Functional Selectivity at G-Protein Coupled Receptors : Advancing Cannabinoid Receptors as Drug Targets. *Biochem. Pharmacol.* **2017**, *128*, 1–11.

(94) Kendall, D. A.; Yudowski, G. A. Cannabinoid Receptors in the Central Nervous System : Their Signaling and Roles in Disease. *Front. Cell. Neurosci.* **2017**, *10*, 1–10.

(95) Pertwee, R. G. Targeting the Endocannabinoid System with Cannabinoid Receptor Agonists: Pharmacological Strategies and Therapeutic Possibilities. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367* (1607), 3353–3363.

(96) Howlett, A. C. International Union of Pharmacology. XXVII. Classification of Cannabinoid Receptors. *Pharmacol. Rev.* **2002**, *54* (2), 161–202.

(97) Hua, T.; Vemuri, K.; Pu, M.; Qu, L.; Han, G. W.; Wu, Y.; Zhao, S.; Shui, W.; Li, S.; Korde, A.; et al. Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **2016**, *167* (3), 750–762.e14.

(98) Chemical Computing Group. Molecular Operating Environment. 1010 Sherbrook St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 2015.

(99) Mechoulam, R.; Gaoni, Y. The Absolute Configuration of Delta 1 -Tetrahydrocannabinol, The Major Active Constituent of Hashish. *Tetrahedron Lett.* **1967**, *8* (12), 1109–1111.

(100) Sharkey, K. A. Endocannabinoids: Biology, Mechanism of Action and Functions. *Int. J. Obes.* **2006**, *30* (SUPPL. 1), 4–6.

(101) Mechoulam, R.; Hanuš, L. O.; Pertwee, R.; Howlett, A. C. Early Phytocannabinoid Chemistry to Endocannabinoids and Beyond. *Nat. Rev. Neurosci.* **2014**, *15* (11), 757–764.

(102) Mechoulam, R.; Parker, L. Towards a Better Cannabis Drug. *Br. J. Pharmacol.* **2013**, *170* (7), 1363–1364.

(103) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334–395.

(104) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins Struct. Funct. Genet.* **2002**, *47* (4), 409–443.

(105) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. *Phys. Chem. Chem. Phys.* **2010**, *12* (40), 12899–12908.

(106) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring Functions and Their Evaluation Methods for Protein–ligand Docking: Recent Advances and Future Directions. *Phys. Chem. Chem. Phys.* **2010**, *12* (40), 12899.

(107) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing Scoring Functions for Protein-Ligand Interactions. *J. Med. Chem.* **2004**, *47* (12), 3032–3047.

(108) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The Worldwide Protein Data Bank (WwPDB): Ensuring a Single , Uniform Archive of PDB Data. **2018**, *35* (November 2006), 2006–2008.

(109) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M.; Physics, B. Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures. *FEBS Lett.* **2015**, *275* (1), 1–21.

(110) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discov. Today* **2009**, *14* (13–14), 676–683.

(111) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, *22* (2), 195–201.

(112) Haas, J.; Roth, S.; Arnold, K.; Kiefer, F.; Schmidt, T.; Bordoli, L.; Schwede, T. The Protein Model Portal - A Comprehensive Resource for Protein Structure and Model Information. *Database* **2013**, *31*, 1–8.

(113) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided. Drug Des.* **2011**, *7* (2), 146–157.

(114) Hollingsworth, S. A.; Karplus, P. A. A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins. *Biomolecular Concepts*. **2010**, 1 (3-4), 271-283,

(115) Eisenberg, D.; Luthy, R.; Bowie, J. U. VERIFY3D Assessment of Protein Model with 3Dprofiles. *Methods Enzymol.* **1997**, *277* (20), 396–404.

(116) Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**.

(117) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (12), 4977–5010.

(118) Roy, K.; Ambure, P.; Aher, R. B. How Important Is to Detect Systematic Error in Predictions and Understand. *Chemom. Intell. Lab. Syst.* **2017**, *162* (January), 44–54.

(119) Kiralj, R.; Ferreira, M. M. C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.* **2009**, *20* (4), 770–787.

(120) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2013**, (66), 334-395.

(121) Yang, S.-Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discov. Today* **2010**, (15), 444-450.

(122) Koes, D. R.; Camacho, C. J. ZINCPharmer: Pharmacophore Search of the ZINC Database. *Nucleic Acids Res.* **2012**, *40* (W1), 409–414.

(123) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432* (7019), 862–865.

(124) Subramaniam, S.; Mehrotra, M.; Gupta, D. Bioinformation Virtual High Throughput Screening (VHTS) - A Bioinformation. **2008**.

(125) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; et al. Impact of High-Throughput Screening in Biomedical Research. **2011**.

(126) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discovery Today*. **2009**, (14), 676-672.

(127) Bleicher, K. H.; Böhm, H.-J.; Müller, K.; Alanine, A. I. A Guide to Drug Discovery: Hit and Lead Generation: Beyond High-Throughput Screening. *Nat. Rev. Drug Discov.* **2003**, *2* (5), 369–378.

(128) Iversen, L.; Gibbons, S.; Treble, R.; Setola, V.; Huang, X.; Roth, B. L. Neurochemical Profiles of Some Novel Psychoactive Substances. *Eur. J. Pharmacol.* **2013**, *700*, 147–151.

(129) Rickli, A.; Hoener, M. C.; Liechti, M. E. Monoamine Transporter and Receptor Interaction Profiles of Novel Psychoactive Substances: Para-Halogenated Amphetamines and Pyrovalerone Cathinones. *Eur. Neuropsychopharmacol.* **2015**, *25* (3), 365–376.

(130) Luethi, D.; Kolaczynska, K. E.; Docci, L.; Krähenbühl, S.; Hoener, M. C.; Liechti, M. E. Pharmacological Profile of Mephedrone Analogs and Related New Psychoactive Substances. *Neuropharmacology* **2018**, *134*, 4–12.

(131) Tang, M. H.; Hung, L.; Lai, C.; Ching, C.; Mak, T. W. L. 9-Year Review of New Psychoactive Substance Use in Hong Kong: A Clinical Laboratory Perspective. *Hong Kong J. Emerg. Med.* **2018**, 1–7.

(132) Liechti, M. E. Novel Psychoactive Substances (Designer Drugs): Overview and Pharmacology of Modulators of Monoamine Signalling. *Swiss Medical Weekly*. **2015**, 145, 1-12.

(133) Mayer, F. P.; Burchardt, N. V.; Decker, A. M.; Partilla, J. S.; Li, Y.; McLaughlin, G.; Kavanagh, P. V.; Sandtner, W.; Blough, B. E.; Brandt, S. D.; et al. Fluorinated Phenmetrazine "Legal Highs" Act as Substrates for High-Affinity Monoamine Transporters of the SLC6 Family. *Neuropharmacology* **2018**, *134*, 149–157.

(134) Makwana, R.; Molleman, A.; Parsons, M. Pharmacological Characterization of Cannabinoid Receptor Activity in the Rat-Isolated Ileum Myenteric Plexus-Longitudinal Muscle Preparation. *Br. J. Pharmacol.* **2010**, *159* (8), 1608–1622.

(135) Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by C Alpha Geometry: Phi,Psi and C Beta Deviation. *Proteins-Structure Funct. Genet.* **2003**, *50* (3), 437–450.

(136) Hooda, V.; Gundala, P. B.; Chinthala, P. Sequence Analysis and Homology Modeling of Peroxidase from Medicago Sativa. *Bioinformation* **2012**, *8* (20), 974–979.

(137) Bowie, J. U.; Lüthy, R.; Eisenberg, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science* **1991**, *253* (5016), 164–170.

(138) Eisenberg, D.; Luthy, R.; Bowie, J. U. Verify3D: Assessment of Protein Models with Three-Dimensional Profiles. *Methods Enzymol.* **1997**, *277* (8), 396–404.

(139) Colovos, C.; Yeates, T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**, *2* (9), 1511–1519.

(140) Liu, J.; Wang, R. Classification of Current Scoring Functions. *J. Chem. Inf. Model.* **2015**, *55* (3),

475–482.

(141) Huang, N.; Kalyanaraman, C.; Irwin, J. J.; Jacobson, M. P. Physics-Based Scoring of Protein - Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46* (1), 243–253.

(142) Naïm, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; et al. Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space. *J. Chem. Inf. Model.* **2007**, *47* (1), 122–133.

(143) Goto, J.; Kataoka, R.; Muta, H.; Hirayama, N. ASEDock-Docking Based on Alpha Spheres and Excluded Volumes. *J. Chem. Inf. Model.* **2008**, *48* (3), 583–590.

(144) Hildebrandt, a. K.; Stockel, D.; Fischer, N. M.; de la Garza, L.; Kruger, J.; Nickels, S.; Rottig, M.; Scharfe, C.; Schumann, M.; Thiel, P.; et al. Ballaxy: Web Services for Structural Bioinformatics. *Bioinformatics* **2014**, *31* (1), 121–122.

(145) Graves, A. P.; Shivakumar, D. M.; Boyce, S. E.; Jacobson, M. P.; Case, D. A.; Shoichet, B. K. Rescoring Docking Hit Lists for Model Cavity Sites : Predictions and Experimental Testing. **2008**, 914–934.

(146) Thompson, D. C.; Humblet, C.; Joseph-mccarthy, D. Investigation of MM-PBSA Rescoring of Docking Poses. **2008**, 1081–1091.

(147) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22* (6), 608–640.

(148) Shin, W.-H.; Lee, G. R.; Seok, C. Evaluation of GalaxyDock Based on the Community Structure–Activity Resource 2013 and 2014 Benchmark Studies. *J. Chem. Inf. Model.* **2015**, (10), 1-8.

(149) Hildebrandt Andreas, Katharina Dehof Anna, Rurainski Alexander, Bertsch Andreas, Schumann Marcel, C Toussaint Nora, Moll Andreas, Stockel Daniel, Nickels Stefan, C Mueller Sabine, Hans-Peter Lenhof, K. O. Galaxy/Ballaxy https://ballaxy.bioinf.uni-sb.de/ (accessed Jul 2, 2015).

(150) Rejtol, P. A.; Gehlhaarl, D. K.; Verkhivkerl, G. M.; Sherman, I.; Fogel, D. B.; Freer, S. T. Molecular Recognition of the Inhibitor AC-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Curr. Biol.* **1995**, *3*, 317–324.

(151) Jones, G.; Willett, P.; Glen, R.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.

(152) Cambridge Crystallographic Data Centre. *GOLD User Guide A Component of the GOLD Suite*;

2018.

(153) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins Struct. Funct. Genet.* **2003**.

(154) Yang, J. M.; Frank Hsu, D. Consensus Scoring Criteria in Structure-Based Virtual Screening. *Emerg. Inf. Technol. Conf. 2005* **2005**, *2005*, 165–167.

(155) Mukaka, M. M. Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient in Medical Research. *Malawi Med. J.* **2012**, *24* (3), 69–71.

(156) Godden, J. W.; Xue, L.; Bajorath, J. Combinatorial Preferences Affect Molecular Similarity/Diversity Calculations Using Binary Fingerprints and Tanimoto Coefficients. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (1), 163–166.

(157) Makarenko, A. S.; Mauri, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Ertl, P.; Tetko, I. V.; Radchenko, E. V.; Zefirov, N. S.; Gasteiger, J.; Palyulin, V. A.; Livingstone, D. Virtual Computational Chemistry Laboratory – Design and Description. *J. Comput. Aided. Mol. Des.* **2005**, 19 (6), 453–463.

(158) JChem. ChemAxon. 2015. Available at: *https://chemaxon.com/*(accessed on 13th March 2019).

(159) Bock, T. DISPLAYR https://www.displayr.com/what-is-hierarchical-clustering/ (accessed Oct 15, 2018).

(160) Labute, P.; Williams, C.; Feher, M.; Sourial, E.; Schmidt, J. M. Flexible Alignment of Small Molecules. *J. Med. Chem.* **2001**, *44* (10), 1483–1490.

(161) Shin, W.-H.; Zhu, X.; Bures, M.; Kihara, D. Three-Dimensional Compound Comparison Methods and Their Application in Drug Discovery. *Molecules* **2015**, *2* (2), 147–185.

(162) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280.

(163) Vilar, S.; Uriarte, E.; Santana, L.; Lorberbaum, T.; Hripcsak, G.; Friedman, C.; Tatonetti, N. P. Similarity-Based Modeling in Large-Scale Prediction of Drug-Drug Interactions. *Nat. Protoc.* **2014**, *9*, 2147–2163.

(164) Rothman, R. B.; Baumann, M. H. Monoamine Transporters and Psychostimulant Addiction. *Biochem. Pharmacol.* **2008**, *75* (1), 196–217.

(165) Kristensen, A. S.; Andersen, J.; Jørgensen, T. N.; Sørensen, L.; Eriksen, J.; Loland, C. J.;

Strømgaard, K. SLC6 Neurotransmitter Transporters : Structure , Function , and Regulation. *Pharmacol. Rev.* **2011**, *63* (3), 585–640.

(166)   Reith, M. E. a. *Neurotransmitter Transporters Structure, Function and Regulation*, 2nd editio.; Reith, M. E. a., Ed.; Humana Press: Illinois, 2002.

(167)   Howell, L. L.; Kimmel, H. L. Monoamine Transporters and Psychostimulant Addiction. *Biochem. Pharmacol.* **2008**, *75* (1), 196–217.

(168)   Iversen, L.; White, M.; Treble, R. Designer Psychostimulants: Pharmacology and Differences. *Neuropharmacology* **2014**, *87C*, 59–65.

(169)   Freudenmann, R. W.; Öxler, F.; Bernschneider-Reif, S. The Origin of MDMA (Ecstasy) Revisited: The True Story Reconstructed from the Original Documents. *Addiction* **2006**, *101* (9), 1241–1245.

(170)   Daws, C., L. Unfaithful Neurotransmitter Transporters: Focus on Serotonin Uptake and Implications for Antidepressant Efficacy. *Pharmacol. Ther.* **2013**, *31* (9), 1713–1723.

(171)   Larsen, M. B.; Sonders, M. S.; Mortensen, O. V.; Larson, G. A.; Zahniser, N. R.; Amara, S. G. Dopamine Transport by the Serotonin Transporter : A Mechanistically Distinct Mode of Substrate Translocation. **2011**, *31* (17), 6605–6615.

(172)   Berger, M.; Gray, J. a; Roth, B. L. The Expanded Biology of Serotonin. *Annu. Rev. Med.* **2009**, *60*, 355–366.

(173)   Lane, R. M. Antidepressant Drug Development: Focus on Triple Monoamine Reuptake Inhibition. *J. Psychopharmacol.* **2015**, *29* (5), 526–544.

(174)   Bisgaard, H.; Larsen, M. A. B.; Mazier, S.; Beuming, T.; Newman, A. H.; Weinstein, H.; Shi, L.; Loland, C. J.; Gether, U. The Binding Sites for Benztropines and Dopamine in the Dopamine Transporter Overlap. *Neuropharmacology* **2011**, 60, 182-190.

(175)   Marks, D. M.; Pae, C.-U.; Patkar, A. a. Triple Reuptake Inhibitors: The next Generation of Antidepressants. *Curr. Neuropharmacol.* **2008**, *6* (4), 338–343.

(176)   Schlessinger, A.; Geier, E.; Fan, H.; Irwin, J. J.; Shoichet, B. K.; Giacomini, K. M.; Sali, A. Structure-Based Discovery of Prescription Drugs That Interact with the Norepinephrine Transporter, NET. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (38), 15810–15815.

(177)   Wallner, B.; Elofsson, A. All Are Not Equal: A Benchmark of Different Homology Modeling Programs. *Protein Sci.* **2005**, *14* (5), 1315–1327.

(178) Bonvin, M. J. J.; Weng, Z.; Schueler-furman, O.; Kozakov, D. And Protein Docking. **2018**, *85* (1), 10–16.

(179) Wang, K. H.; Penmatsa, A.; Gouaux, E. Neurotransmitter and Psychostimulant Recognition by the Dopamine Transporter. *Nature* **2015**, 521, 321-326

(180) Coleman, J. A.; Green, E. M.; Gouaux, E.; Health, O.; Health, O. X-Ray Structures and Mechanism of the Human Serotonin Transporter. *Nature* **2016**, *532* (7599), 334–339.

(181) Iversen, L.; Gibbons, S.; Treble, R.; Setola, V.; Huang, X. P.; Roth, B. L. Neurochemical Profiles of Some Novel Psychoactive Substances. *Eur. J. Pharmacol.* **2013**, *700* (1–3), 147–151.

(182) Leach, R., A. *Molecular Modelling, Principles and Applications*, second edi.; Pearson Education: Dorset, 2001.

(183) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.

(184) Ramsey, P. H. Critical Values for Spearman's Rank Order Correlation. *Source J. Educ. Stat.* **1989**, *14* (3), 245–253.

(185) Li, Y.; Liu, Z.; Wang, R. Test MM-PB / SA on True Conformational Ensembles of Protein - Ligand Complexes. **2010**, 1682–1692.

(186) Lee, M. C.; Yang, R.; Duan, Y. Comparison between Generalized-Born and Poisson-Boltzmann Methods in Physics-Based Scoring Functions for Protein Structure Prediction. *J. Mol. Model.* **2005**, *12* (1), 101–110.

(187) Wang, K. H.; Penmatsa, A.; Gouaux, E. Neurotransmitter and Psychostimulant Recognition by the Dopamine Transporter. *Nature* **2015**, *521*, 322–327.

(188) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T. SWISS-MODEL: Modelling Protein Tertiary and Quarternary Structure Using Evolutionary Infomation. *Nucleic Acids Res.* **2014**, *42*, 252–258.

(189) Kufareva, I.; Abagyan, R. Homology Modeling. **2012**, *857*, 231–257.

(190) Chothia, C.; Lesk, A. M. The Relation between the Divergence of Sequence and Structure in Proteins. *EMBO J.* **1986**, *5* (4), 823–826.

(191) Fiser, A.; Do, R. K.; Sali, A.; Fiser, A.; Kinh, R.; Do, G.; Andrej, S. Modeling of Loops in Protein

Structures [In Process Citation] Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9* (09), 1753–1773.

(192) Rickli, A.; Hoener, M. C.; Liechti, M. E. Monoamine Transporter and Receptor Interaction Profiles of Novel Psychoactive Substances: Para-Halogenated Amphetamines and Pyrovalerone Cathinones. *Eur. Neuropsychopharmacol.* **2015**.

(193) Koldsø, H.; Christiansen, A. B.; Sinning, S.; Schiøtt, B. Comparative Modeling of the Human Monoamine Transporters: Similarities in Substrate Binding. *ACS Chem. Neurosci.* **2013**, *4* (2), 295–309.

(194) Soga, S.; Shirai, H.; Koborv, M.; Hirayama, N. Use of Amino Acid Composition to Predict Ligand-Binding Sites. *J. Chem. Inf. Model.* **2007**, *47* (2), 400–406.

(195) Penmatsa, A.; Wang, K. H.; Gouaux, E. X-Ray Structure of Dopamine Transporter Elucidates Antidepressant Mechanism. *Nature* **2013**, *503* (7474), 85–90.

(196) Lynagh, T.; Khamu, T. S.; Bryan-Lluka, L. J. Extracellular Loop 3 of the Noradrenaline Transporter Contributes to Substrate and Inhibitor Selectivity. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **2014**, *387* (1), 95–107.

(197) Andersen, J.; Taboureau, O.; Hansen, K. B.; Olsen, L.; Egebjerg, J.; Strøomgaard, K.; Kristenen, A. S. Location of the Antidepressant Binding Site in the Serotonin Transporter: Importance of Ser-438 in Recognition of Citalopram and Tricyclic Antidepressants. *J. Biol. Chem.* **2009**, *284* (15), 10276–10284.

(198) Gabrielsen, M.; Kurczab, R.; Ravna, A. W.; Kufareva, I.; Abagyan, R.; Chilmonczyk, Z.; Bojarski, A. J.; Sylte, I. Molecular Mechanism of Serotonin Transporter Inhibition Elucidated by a New Flexible Docking Protocol. *Eur. J. Med. Chem.* **2012**, *47* (1), 24–37.

(199) Coleman, J. A.; Green, E. M.; Gouaux, E. X-Ray Structures and Mechanism of the Human Serotonin Transporter. *Nature* **2016**, *532* (7599), 334–339.

(200) Huang, X.; Zhan, C. G. How Dopamine Transporter Interacts with Dopamine: Insights from Molecular Modeling and Simulation. *Biophys. J.* **2007**, *93* (10), 3627–3639.

(201) Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. Receptor-Ligand Molecular Docking. *Biophys. Rev.* **2014**, *6* (1), 75–87.

(202) Ramsey, P. H. Critical Values for Spearman's Rank Order Correlation. *J. Educ. Behav. Stat.* **1989**, *14* (3), 245–253.

(203) Hsieh, M. J.; Luo, R. Physical Scoring Function Based on AMBER Force Field and Poisson-Boltzmann Implicit Solvent for Protein Structure Prediction. *Proteins Struct. Funct. Genet.* **2004**, *56* (3), 475–486.

(204) Andelman, D. Electrostatic Properties of Membranes : The Poisson – Boltzmann Theory. *Handb. Biol. Phys.* **1995**, *1*, 603–641.

(205) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (5), 1422–1426.

(206) Liu, S.; Fu, R.; Zhou, L. H.; Chen, S. P. Application of Consensus Scoring and Principal Component Analysis for Virtual Screening against β-Secretase (BACE-1). *PLoS One* **2012**, *7* (6).

(207) Feher, M. Consensus Scoring for Protein-Ligand Interactions. *Drug Discov. Today* **2006**, *11* (9–10), 421–428.

(208) Paul, N.; Rognan, D. ConsDock: A New Program for the Consensus Analysis of Protein-Ligand Interactions. *Proteins Struct. Funct. Genet.* **2002**, *47* (4), 521–533.

(209) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42* (25), 5100–5109.

(210) Beuming, T.; Kniazeff, J.; Bergmann, M. L.; Shi, L.; Gracia, L.; Newman, A. H.; Javitch, J. a; Weinstein, H.; Loland, C. J. Transporter Overlap. *Physiology* **2009**, *11* (7), 780–789.

(211) Loland, C. J. The Use of LeuT as a Model in Elucidating Binding Sites for Substrates and Inhibitors in Neurotransmitter Transporters. *Biochim. Biophys. Acta* **2014**, *1850* (3), 500–510.

(212) Kristensen, A. S.; Andersen, J.; Jorgensen, T. N.; Sorensen, L.; Eriksen, J.; Loland, C. J.; Stromgaard, K.; Gether, U. SLC6 Neurotransmitter Transporters: Structure, Function, and Regulation. *Pharmacol. Rev.* **2011**.

(213) Penmatsa, A.; Wang, K. H.; Gouaux, E. X-Ray Structures of Drosophila Dopamine Transporter in Complex with Nisoxetine and Reboxetine. *Nat. Struct. Mol. Biol.* **2015**.

(214) Levoin, N.; Calmels, T.; Krief, S.; Danvy, D.; Berrebi-Bertrand, I.; Lecomte, J. M.; Schwartz, J. C.; Capet, M. Homology Model versus X-Ray Structure in Receptor-Based Drug Design: A Retrospective Analysis with the Dopamine D3 Receptor. *ACS Med. Chem. Lett.* **2011**, *2* (4), 293–297.

(215) Ravna, A. W.; Sylte, I.; Dahl, S. G. Molecular Mechanism of Citalopram and Cocaine Interactions

with Neurotransmitter Transporters. *J. Pharmacol. Exp. Ther.* **2003**, *307* (1), 34–41.

(216) Manepallia, Sanker Laura M. Geffertb, Christopher K. Surrattb, and J. D. M. Discovery of Novel Selective Serotonin Reuptake Inhibitors Throuhg Development of a Protein-Based Pharmacophore. *J. Chem. Inf. Model.* **2015**, *25* (8), 713–724.

(217) Giaginis, C.; Tsantili-Kakoulidou, A.; Theocharis, S. Applying Quantitative Structure-Activity Relationship (QSAR) Methodology for Modeling Postmortem Redistribution of Benzodiazepines and Tricyclic Antidepressants. *J. Anal. Toxicol.* **2014**, *38* (5), 242–248.

(218) Iversen, L.; Gibbons, S.; Treble, R.; Setola, V.; Huang, X. P.; Roth, B. L. Neurochemical Profiles of Some Novel Psychoactive Substances. *Eur. J. Pharmacol.* **2013**.

(219) European Monitoring Centre for Drugs and Drug Addiction. New Psychoactive Substances in Europe. *EU early Warn. Syst.* **2015**, No. March, 12.

(220) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**.

(221) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3* (10), 1–14.

(222) Ali, J.; Camilleri, P.; Brown, M. B.; Hutt, A. J.; Kirton, S. B. Revisiting the General Solubility Equation: In Silico Prediction of Aqueous Solubility Incorporating the Effect of Topographical Polar Surface Area. *J. Chem. Inf. Model.* **2012**, *52* (2), 420–428.

(223) Shapiro, S. S.; Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). **1965**, *52* (3), 591-598.

(224) Motulsky, H. GraftPad Software ps://www.graphpad.com/quickcalcs/Grubbs1.cfm (accessed Aug 11, 2018).

(225) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (1), 1–12.

(226) Haas, Juergen., Roth S., A. No Title http://www.proteinmodelportal.org/query/up/Q01959 (accessed Sep 7, 2016).

(227) Haas, J.; Roth, S.; Arnold, K.; Kiefer, F.; Schmidt, T.; Bordoli, L.; Schwede, T. The Protein Model Portal - A Comprehensive Resource for Protein Structure and Model Information. *Database* **2013**, *2013* (8), 1–8.

(228) Zamora, I.; Oprea, T.; Cruciani, G.; Pastor, M.; Ungell, A. L. Surface Descriptors for Protein-

Ligand Affinity Prediction. *J. Med. Chem.* **2003**, *46* (1), 25–33.

(229) Maggiora, G. M. On Outliers and Activity Cliffs - Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46* (4), 1535.

(230) Shapiro, H. *NPS Come of Age: A UK Overview*; 2016. Available at: https://www.drugwise.org.uk/wp-content/uploads/NPSComeofAge.pdf (Accessed at 13th March 2019)

(231) Tai, S. and W. F. Synthetic Cannabinoids: Pharmacology, Behavioral Effects and Abuse Potential. **2016**, *8* (5), 583–592.

(232) Ernst, L.; Brandhorst, K.; Papke, U.; Altrogge, A.; Zodel, S.; Langer, N.; Beuerle, T. Identification and Quantification of Synthetic Cannabinoids in 'Spice-like' Herbal Mixtures: Update of the German Situation in Early 2017. *Forensic Sci. Int.* **2017**, *277*, 51–58.

(233) Gerace, E.; Petrarulo, M.; Bison, F.; Salomone, A.; Vincenti, M. Toxicological Findings in a Fatal Multidrug Intoxication Involving Mephedrone. *Forensic Sci. Int.* **2014**, *243*, 68–73.

(234) Seely, K. A.; Lapoint, J.; Moran, J. H.; Fattore, L. Spice Drugs Are More than Harmless Herbal Blends: A Review of the Pharmacology and Toxicology of Synthetic Cannabinoids. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **2012**, *39* (2), 234–243.

(235) Adams, A. J.; Banister, S. D.; Irizarry, L.; Trecki, J.; Schwartz, M.; Gerona, R. "Zombie" Outbreak Caused by the Synthetic Cannabinoid AMB-FUBINACA in New York. *N. Engl. J. Med.* **2017**, *376* (3), 235–242.

(236) Wiley, J. L.; Marusich, J. A.; Huffman, J. W.; Balster, R. L.; Thomas, B. F. Hijacking of Basic Research: The Case of Synthetic Cannabinoids. *Methods Rep. RTI. Press.* **2011**, *2011*.

(237) Wiley, J. L.; Lefever, T. W.; Marusich, J. A.; Grabenauer, M.; Moore, K. N.; Huffman, J. W.; Thomas, B. F. Evaluation of First Generation Synthetic Cannabinoids on Binding at Non-Cannabinoid Receptors and in a Battery of in Vivo Assays in Mice. *Neuropharmacology* **2016**, *110*, 143–153.

(238) Schneir, A. B.; Cullen, J.; Ly, B. T. "Spice" Girls: Synthetic Cannabinoid Intoxication. *J. Emerg. Med.* **2011**, *40* (3), 296–299.

(239) Tai, S.; Fantegrossi, W. E. Synthetic Cannabinoids: Pharmacology, Behavioral Effects, and Abuse Potential. *Curr. Addict. Reports* **2014**, *1*, 129–136.

(240) Shevyrin, V.; Melkozerov, V.; Endres, G. W.; Shafran, Y.; Morzherin, Y. On a New Cannabinoid

Classification System: A Sight on the Illegal Market of Novel Psychoactive Substances. *Cannabis Cannabinoid Res.* **2016**, *1* (1), 186–194.

(241) EMCDDA. EDND - European Infomation system and database on new drugs https://ednd-cma.emcdda.europa.eu/.

(242) Uchiyama, N.; Kikura-Hanajiri, R.; Hakamatsuka, T. Evaluation of an On-Site Drug-Testing Device for the Detection of Synthetic Cannabinoids in Illegal Herbal Products. *Pharm. Soc. Japan* **2015**, *135* (4), 535–541.

(243) Schrodinger Press. Phase 4.3 User Manual. In *Phase User Manual Copyright*; 2015; pp 1–230.

(244) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182.

(245) Goodwin, J. T.; Clark, D. E. In Silico Predictions of Blood-Brain Barrier Penetration : Considerations to " Keep in Mind ." *J. Pharmacol. Exp. Ther.* **2005**, *315* (2), 477–483.

(246) Goodwin, J. T. In Silico Predictions of Blood-Brain Barrier Penetration: Considerations to "Keep in Mind." *J. Pharmacol. Exp. Ther.* **2005**, *315* (2), 477–483.

(247) Matias, I.; Bisogno, T.; Marzo, V. Di. Endogenous Cannabinoids in the Brain and Peripheral Tissues : Regulation of Their Levels and Control of Food Intake. *Int. J. Obes.* **2006**, *30*, 7–12.

(248) Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular Fingerprint Similarity Search in Virtual Screening. *Methods* **2015**, *71* (C), 58–63.

(249) Zloh, M.; Samaras, E. G.; Calvo-Castro, J.; Guirguis, A.; Stair, J. L.; Kirton, S. B. Drowning in Diversity? A Systematic Way of Clustering and Selecting a Representative Set of New Psychoactive Substances. *RSC Adv.* **2017**, *7* (84), 53181–53191.

(250) Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P. E. B. The Protein Data Bank H.M. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(251) Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J. Med. Chem.* **2005**, *48* (1), 312–320.

(252) E.M.C.D.D.A., European Monitoring Centre for Drugs and Drug Addiction. (**2011**). Europena Drug Reprot:Office for Official Publications of the European Communities.

(253) E.M.C.D.D.A., European Monitoring Centre for Drugs and Drug Addiction. (**2014**). Synthetic Cannabinoids in Europe:Office for Official Publications of the European Communities.

(254)   Wong, Y. L. *Synthetic Cannabinoids in Herbal Products*; **2011**.

(255)   Comley, R. a; Salinas, C. a; Slifstein, M.; Petrone, M.; Marzano, C.; Bennacef, I.; Shotbolt, P.; Van der Aart, J.; Neve, M.; Iavarone, L.; et al. Monoamine Transporter Occupancy of a Novel Triple Reuptake Inhibitor in Baboons and Humans Using Positron Emission Tomography. *J. Pharmacol. Exp. Ther.* **2013**, *346* (2), 311–317.

(256)   United Nations Office on Drugs and Crime. *Synthetic Cannabinoids in Herbal Products*; 2011.

(257)   Uchiyama, N.; Kikura-Hanajiri, R.; Hakamatsuka, T. Evaluation of an On-Site Drug-Testing Device for the Detection of Synthetic Cannabinoids in Illegal Herbal Products. *Yakugaku Zasshi* **2015**, *135* (3), 535–541.

(258)   Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem* **2008**, *3* (3), 435–444.

(259)   Lipinski, C. A. Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44* (1), 235–249.

(260)   Singh, R.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D. L.; Reggio, P. H.; Guarnieri, F. Activation of the Cannabinoid CB1 Receptor May Involve a W6 48/F3 36 Rotamer Toggle Switch. *J. Pept. Res. Off. J. Am. Pept. Soc.* **2002**, *60* (6), 357–370.

(261)   Gurney, S. M. R.; Scott, K. S.; Kacinko, S. L.; Presley, B. C.; Logan, B. K. Pharmacology, Toxicology, and Adverse Effects of Synthetic Cannabinoid Drugs. *Forensic Sci. Rev.* **2014**, *26* (1), 54–76.

(262)   Morales, P.; Goya, P.; Jagerovic, N.; Hernandez-Folgado, L. Allosteric Modulators of the CB $_1$ Cannabinoid Receptor: A Structural Update Review. *Cannabis Cannabinoid Res.* **2016**.

(263)   Fattore, L.; Fratta, W. Beyond THC: The New Generation of Cannabinoid Designer Drugs. *Front. Behav. Neurosci.* **2011**, *5* (September), 1–12.

(264)   Bilici, R. Synthetic Cannabinoids. *North. Clin. Istanbul* **2014**, *1* (2), 121–126.

(265)   Irie, T.; Kikura-Hanajiri, R.; Usami, M.; Uchiyama, N.; Goda, Y.; Sekino, Y. MAM-2201, a Synthetic Cannabinoid Drug of Abuse, Suppresses the Synaptic Input to Cerebellar Purkinje Cells via Activation of Presynaptic CB1 Receptors. *Neuropharmacology* **2015**, *95*, 479–491.

(266)   Sibaev, A.; Yuece, B.; Allescher, H. D.; Saur, D.; Storr, M.; Kurjak, M. The Endocannabinoid Anandamide Regulates the Peristaltic Reflex by Reducing Neuro-Neuronal and Neuro-Muscular Neurotransmission in Ascending Myenteric Reflex Pathways in Rats. *Pharmacol. Reports* **2014**,

66 (2), 256–263.

(267) Izzo, a a; Mascolo, N.; Tonini, M.; Capasso, F. Modulation of Peristalsis by Cannabinoid CB(1) Ligands in the Isolated Guinea-Pig Ileum. *Br. J. Pharmacol.* **2000**, *129* (5), 984–990.

(268) Aviello, G.; Romano, B.; Izzo, A. A. Cannabinoids and Gastrointestinal Motility: Animal and Human Studies. *Eur. Rev. Med. Pharmacol. Sci.* **2008**, *12* (SUPPL. 1), 81–93.

(269) Manara, L.; Croci, T.; Maffrand, J.; Mukenge, S. Functional Assessment in the Muscular Layers of Neuronal Cannabinoid Receptors of Human Ileum and Colon. **2002**, 262–269.

(270) Ford, B. M.; Tai, S.; Fantegrossi, W. E.; Prather, P. L. Synthetic Pot: Not Your Grandfather's Marijuana. *Trends Pharmacol. Sci.* **2017**, *38* (3), 257–276.

(271) Yang, G.; Sau, C.; Lai, W.; Cichon, J.; Li, W. CB1 Canabinoid Receptor-Mediated Increases in Cyclic AMP Accumulation Are Correlated with Reduced Gi/o Function. **2015**, *344* (6188), 1173–1178.

(272) Pirard, S.; Huestis, M. A. Synthetic Cannabinoids: Epidemiology, Pharmacodynamics, and Clinical Implications. **2015**, 12–41.

(273) Maneuf, Y. P.; Brotchie, J. M. Paradoxical Action of the Cannabinoid WIN 55,212-2 in Stimulated and Basal Cyclic AMP Accumulation in Rat Globus Pallidus Slices. *Br.J.Pharmacol.* **1997**, *120*, 1397–1398.

(274) Cadas, H.; Gaillet, S.; Beltramo, M.; Venance, L.; Piomelli, D. Biosynthesis of an Endogenous Cannabinoid Precursor in Neurons and Its Control by Calcium and CAMP. *J.Neurosci.* **1996**, *16* (12), 3934–3942.

(275) Insel, P. A.; Ostrom, R. S. Forskolin as a Tool for Examining Adenylyl Cyclase Expression, Regulation, and G Protein Signaling. *Cell Mol Neurobiol* **2003**, *23* (3), 305–314.

(276) Storr, M. A.; Bashashati, M.; Hirota, C.; Vemuri, V. K.; Keenan, C. M.; Duncan, M.; Lutz, B.; MacKie, K.; Makriyannis, A.; MacNaughton, W. K.; et al. Differential Effects of CB1 Neutral Antagonists and Inverse Agonists on Gastrointestinal Motility in Mice. *Neurogastroenterol. Motil.* **2010**, *22* (7), 787–797.

(277) Izzo, A. A.; Sharkey, K. A. Cannabinoids and the Gut: New Developments and Emerging Concepts. *Pharmacol. Ther.* **2010**, *126* (1), 21–38.

(278) Costain, W. J.; Tauskela, J. S.; Rasquinha, I.; Comas, T.; Hewitt, M.; Marleau, V.; Soo, E. C. Pharmacological Characterization of Emerging Synthetic Cannabinoids in HEK293T Cells and

Hippocampal Neurons. *Eur. J. Pharmacol.* **2016**, *786*, 234–245.

(279) Gori, G. B. Trypsinization of Animal Tissues for Cell Culture: Theoretical Considerations and Automatic Apparatus. *Appl. Microbiol.* **1964**, *12* (2), 115–121.

(280) R. Ian Freshney. *Culture of Animal Cells*, 7th ed.; Davies-Capes, A., Gregory, C., Przyborski, S., Eds.; John Wiley & Sons: New Jersey, 2016.

(281) Promega. CytoTox-ONE ™ Homogeneous Membrane Integrity Assay. *Promega* **2009**, 1–14.

(282) Nørskov-Lauritsen, L.; Thomsen, A. R. B.; Bräuner-Osborne, H. G Protein-Coupled Receptor Signaling Analysis Using Homogenous Time-Resolved Förster Resonance Energy Transfer (HTRF®) Technology. *Int. J. Mol. Sci.* **2014**, *15* (2), 2554–2572.

(283) Wiley, J. L.; Marusich, J. A.; Lefever, T. W.; Antonazzo, K. R.; Wallgren, M. T.; Cortes, R. A.; Patel, P. R.; Grabenauer, M.; Moore, K. N.; Thomas, B. F. AB-CHMINACA, AB-PINACA, and FUBIMINA: Affinity and Potency of Novel Synthetic Cannabinoids in Producing Delta9-Tetrahydrocannabinol-Like Effects in Mice. *J Pharmacol Exp Ther* **2015**, No. September, 328–339.

(284) Ford, B. M.; Franks, L. N.; Tai, S.; Fantegrossi, W. E.; Stahl, E. L.; Berquist, M. D.; Cabanlong, C. V; Wilson, C. D.; Penthala, N. R.; Crooks, P. A.; et al. Characterization of Structurally Novel G Protein Biased CB 1 Agonists : Implications for Drug Development. *Pharmacol. Res.* **2017**, *125*, 161–177.

(285) Rubino, T.; Vigano', D.; Massi, P.; Spinello, M.; Zagato, E.; Giagnoni, G.; Parolaro, D. Chronic Delta-9-Tetrahydrocannabinol Treatment Increases CAMP Levels and CAMP-Dependent Protein Kinase Activity in Some Rat Brain Regions. *Neuropharmacology* **2000**, *39* (7), 1331–1336.

(286) Howlett, A. C.; Blume, L. C.; Dalton, G. D. CB(1) Cannabinoid Receptors and Their Associated Proteins. *Curr. Med. Chem.* **2010**, *17* (14), 1382–1393.

(287) Howlett, A. C. The Cannabinoid Receptors. *Prostaglandins Other Lipid Mediat.* **2002**, *68–69*, 619–631.

(288) Marisol, &; Castaneto, S.; Gorelick, D. A.; Desrosiers, N. A.; Hartman, R. L.; Pirard, S.; Huestis, M. A. Synthetic Cannabinoids: Epidemiology, Pharmacodynamics, and Clinical Implications. *Drug Alcohol Depend.* **2014**, *144*, 12–41.

(289) Twentyman, P., R.; Bleehen, N., M. The Sensitivity of Cells in Exponential and Stationary Phases

of Growth to Bleomycin and to 1,3-Bis(2-Chloroethyl)-1-Nitrosourea. *Br. J. Cancer* **1973**, *28*, 500–507.

(290) Hondebrink, L.; Kasteel, E. E. J.; Tukker, A. M.; Wijnolts, F. M. J.; Verboven, A. H. A.; Westerink, R. H. S. Neuropharmacological Characterization of the New Psychoactive Substance Methoxetamine. *Neuropharmacology* **2017**, *123*, 1–9.

(291) Richter, L. H. J.; Flockerzi, V.; Maurer, H. H.; Meyer, M. R. Pooled Human Liver Preparations, HepaRG, or HepG2 Cell Lines for Metabolism Studies of New Psychoactive Substances? A Study Using MDMA, MDBD, Butylone, MDPPP, MDPV, MDPB, 5-MAPB, and 5-API as Examples. *J. Pharm. Biomed. Anal.* **2017**, *143*, 32–42.

(292) Curran, N. M.; Griffin, B. D.; O'Toole, D.; Brady, K. J.; Fitzgerald, S. N.; Moynagh, P. N. The Synthetic Cannabinoid R(+)WIN 55,212-2 Inhibits the Interleukin-1 Signaling Pathway in Human Astrocytes in a Cannabinoid Receptor-Independent Manner. *J. Biol. Chem.* **2005**, *280* (43), 35797–35806.

(293) Madar, D.; Dekel, E.; Bren, A.; Zimmer, A.; Porat, Z.; Alon, U. Promoter Activity Dynamics in the Lag Phase of Escherichia Coli. *BMC Syst. Biol.* **2013**, *7* (1), 1–13.

(294) Yu. P. Petrov and N. V. Tsupkina. Growth Characterisitics of CHO Cells in Culture. *Cell tissue biol.* **2013**, *7* (1), 72–78.

(295) Altman, S. A.; Randers, L.; Rao, G. Comparison of Trypan Blue Dye Exclusion and Fluorometric Assays for Mammalian Cell Viability Determinations. *Biotechnol. Prog.* **1993**, *9* (6), 671–674.

(296) Hess, C.; Schoeder, C. T.; Pillaiyar, T.; Madea, B.; Müller, C. E. Pharmacological Evaluation of Synthetic Cannabinoids Identified as Constituents of Spice. *Forensic Toxicol.* **2016**, *34* (2), 329–343.

(297) Park, J. G.; Choi, S. S.; Park, T. H. Enhancement of Cell Growth and Viability of CHO Cells in Serum-Free Media by 30Kc6 Gene Expression. *Process Biochem.* **2007**, *42* (1), 8–15.

(298) Han, Y.; Liu, X.-M.; Liu, H.; Li, S.-C.; Wu, B.-C.; Ye, L.-L.; Wang, Q.-W.; Chen, Z.-L. Cultivation of Recombinant Chinese Hamster Ovary Cells Grown as Suspended Aggregates in Stirred Vessels. *J. Biosci. Bioeng.* **2006**, *102* (5), 430–435.

(299) Wang, P.; Henning, S. M.; Heber, D. Limitations of MTT and MTS-Based Assays for Measurement of Antiproliferative Activity of Green Tea Polyphenols. *PLoS One* **2010**, *5* (4).

(300) Wang, S.; Yu, H.; Wickliffe, J. K. Limitation of the MTT and XTT Assays for Measuring Cell

Viability Due to Superoxide Formation Induced by Nano-Scale TiO2. *Toxicol. Vitr.* **2011**, *25* (8), 2147–2151.

(301) Niles, A. L.; Moravec, R. A.; Riss, T. L. In Vitro Viability and Cytotoxicity Testing and Same-Well Multi-Parametric Combinations for High Throughput Screening. *Curr. Chem. Genomics* **2009**, *3*, 33–41.

(302) Tomiyama, K. ichi; Funada, M. Cytotoxicity of Synthetic Cannabinoids on Primary Neuronal Cells of the Forebrain: The Involvement of Cannabinoid CB1receptors and Apoptotic Cell Death. *Toxicol. Appl. Pharmacol.* **2014**, *274* (1), 17–23.

(303) Manning, F.; Zuzel, K. Comparison of Types of Cell Death: Apoptosis and Necrosis. *J. Biol. Educ.* **2003**, *37* (3), 141–145.

(304) Muccioli, G. G.; Lambert, D. M. Current Knowledge on the Antagonists and Inverse Agonists of Cannabinoid Receptors. *Curr. Med. Chem.* **2005**, *12* (12), 1361–1394.

(305) Felder, C. C.; Joyce, K. E.; Briley, E. M.; Glass, M.; Mackie, K. P.; Fahey, K. J.; Cullinan, G. J.; Hunden, D. C.; Johnson, D. W.; Chaney, M. O.; et al. LY320135, a Novel Cannabinoid CB1 Receptor Antagonist, Unmasks Coupling of the CB1 Receptor to Stimulation of CAMP Accumulation. *J. Pharmacol. Exp. Ther.* **1998**, *284* (1), 291–297.

(306) Sheng, W. S.; Hu, S.; Min, X.; Cabral, G. A.; Lokensgard, J. R.; Peterson, P. K. Synthetic Cannabinoid WIN55,212-2 Inhibits Generation of Inflammatory Mediators by IL-1β-Stimulated Human Astrocytes. *Glia* **2005**, *49* (2), 211–219.

(307) Jan, C. R.; Lu, Y. C.; Jiann, B. P.; Chang, H. T.; Su, W.; Chen, W. C.; Huang, J. K. Novel Effect of CP55,940, a CB1/CB2 Cannabinoid Receptor Agonist, on Intracellular Free Ca2+ Levels in Bladder Cancer Cells. *Chinese Journal of Physiology*. 2002, pp 33–39.

(308) Mato, S.; Pazos, A.; Valdizán, E. M. Cannabinoid Receptor Antagonism and Inverse Agonism in Response to SR141716A on CAMP Production in Human and Rat Brain. *Eur. J. Pharmacol.* **2002**, *443* (1–3), 43–46.

(309) Meng, T.; Wang, J.; Peng, H.; Fang, G.; Li, M.; Xiong, B.; Xie, X.; Zhang, Y.; Wang, X.; Shen, J. Discovery of Benzhydrylpiperazine Derivatives as CB1receptor Inverse Agonists via Privileged Structure-Based Approach. *Eur. J. Med. Chem.* **2010**, *45* (3), 1133–1139.

(310) Lin, Z.; Canales, J. J.; Björgvinsson, T.; Thomsen, M.; Qu, H.; Liu, Q. R.; Torres, G. E.; Caine, S. B. Monoamine Transporters: Vulnerable and Vital Doorkeepers. *Prog. Mol. Biol. Transl. Sci.* **2011**.

(311) Marder, M.; Estiú, G.; Blanch, L. B.; Viola, H.; Wasowski, C.; Medina, J. H.; Paladini, A. C. Molecular Modeling and QSAR Analysis of the Interaction of Flavone Derivatives with the Benzodiazepine Binding Site of the GABAAreceptor Complex. *Bioorganic Med. Chem.* **2001**, *9* (2), 323–335.

(312) Glennon, R. A.; Dukat, M. Annals of Forensic Research and Analysis Synthetic Cathinones: A Brief Overview of Overviews with Applications to the Forensic Sciences. *Ann. Forensic Res. Anal.* **2017**, *4* (2), 1–7.

(313) Avram, S.; Buiu, C.; Duda-Seiman, D. M.; Duda-Seiman, C.; Mihailescu, D. 3D-QSAR Design of New Escitalopram Derivatives for the Treatment of Major Depressive Disorders. *Sci. Pharm.* **2010**, *78* (2), 233–248.

(314) Zhu, T.; Cao, S.; Pin-Chih, S.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M.; Hevener, K. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based upon a Critical Literature Analysis. *J. Med. Chem.* **2013**, *12* (56), 6560–6572.

Electrical field stimulation trace in the presence of Carbachol
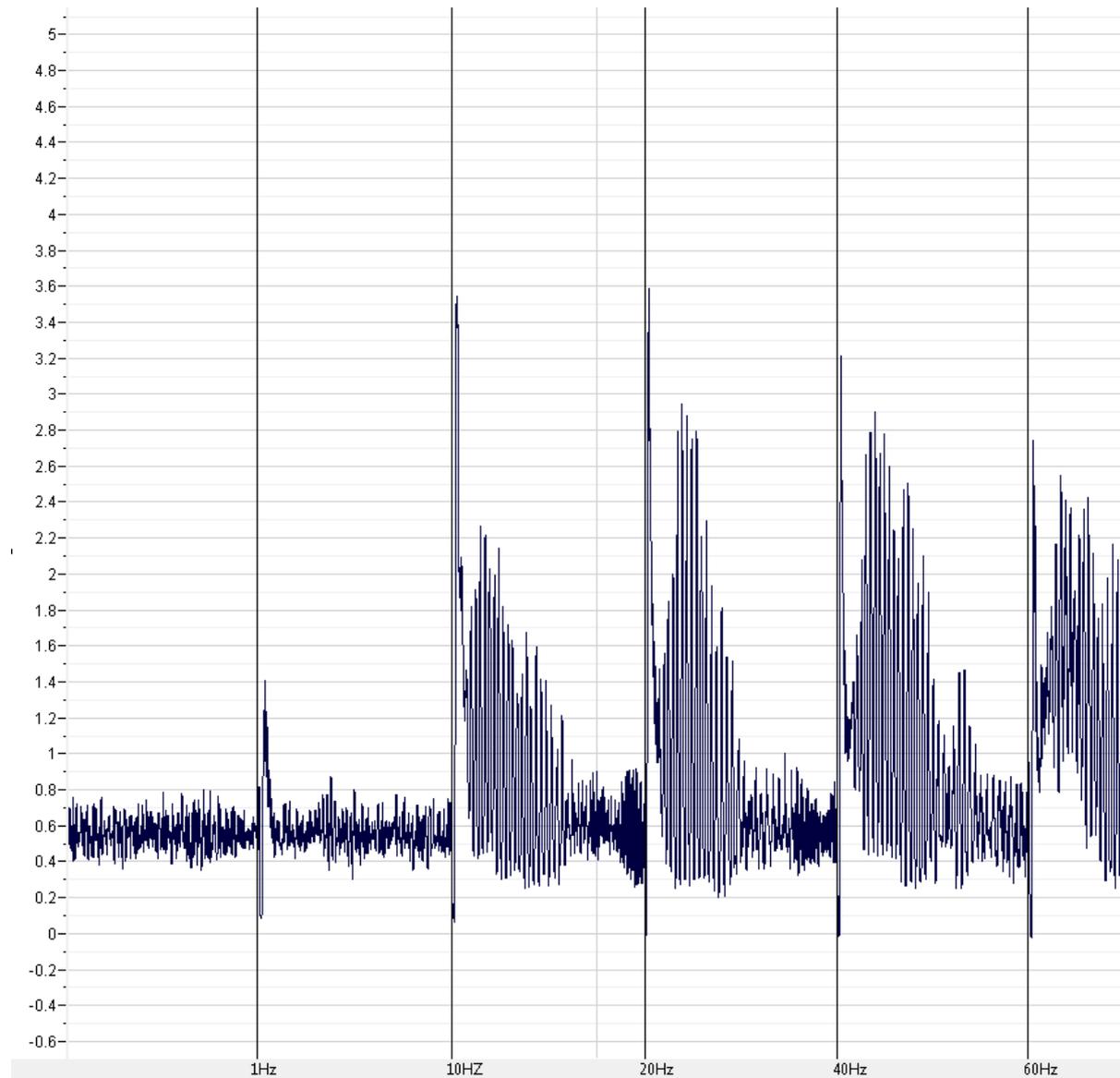
Electrical field stimulation trace in the presence of the positive control MAM-2201
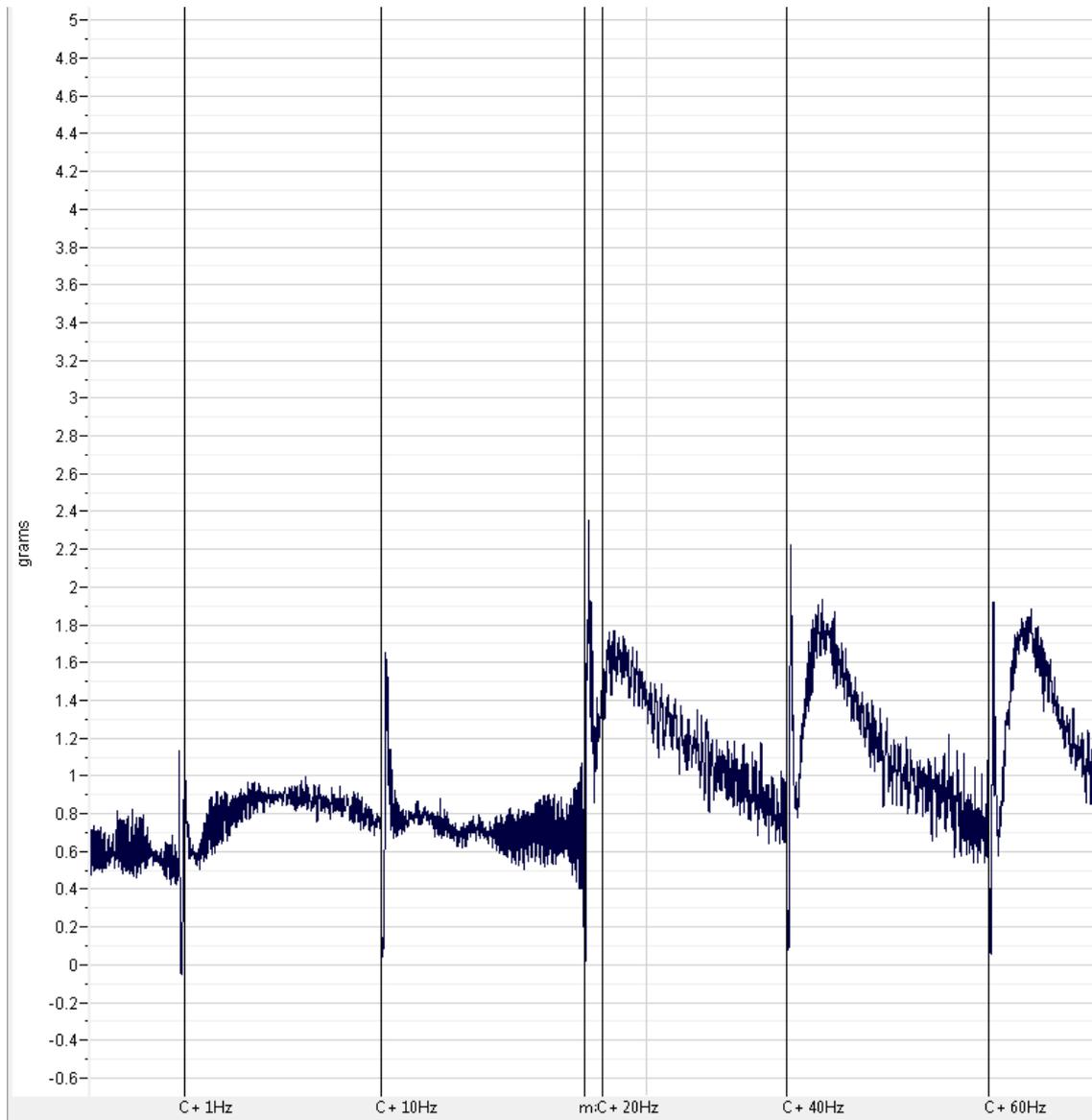
Electrical field stimulation trace in the presence of compound 1
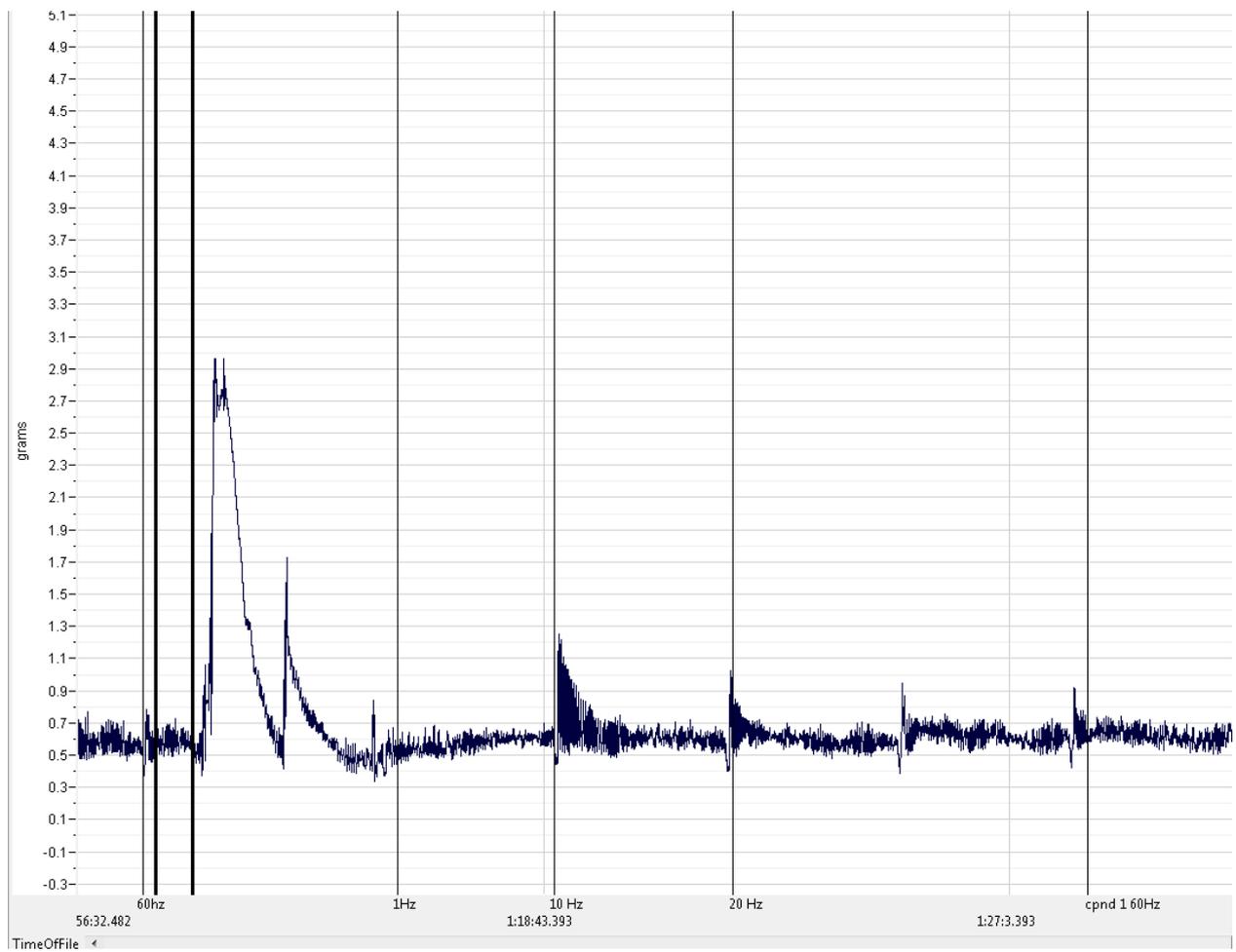
Electrical field stimulation trace in the presence of compound 2
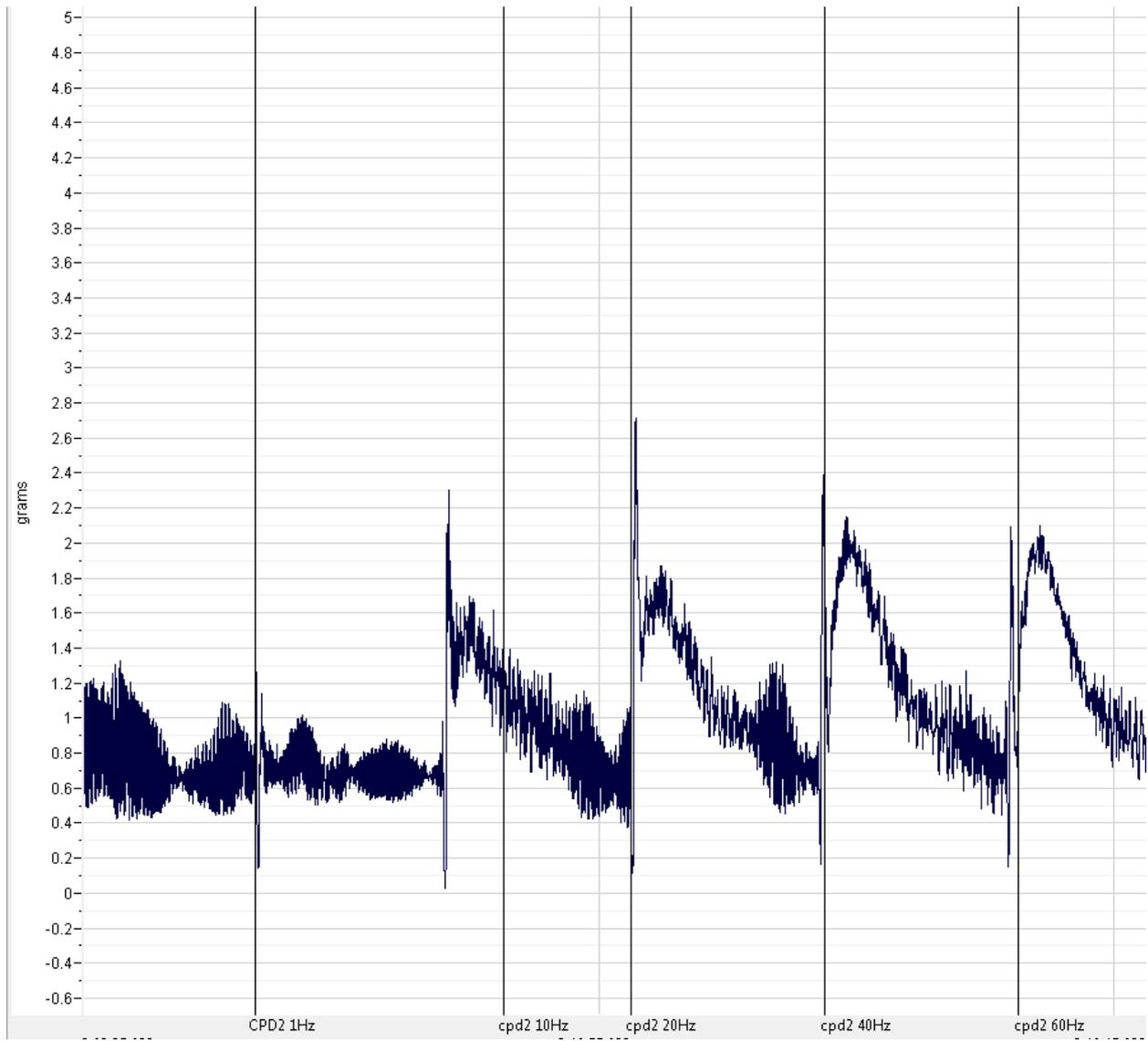
Electrical field stimulation trace in the presence of compound 3
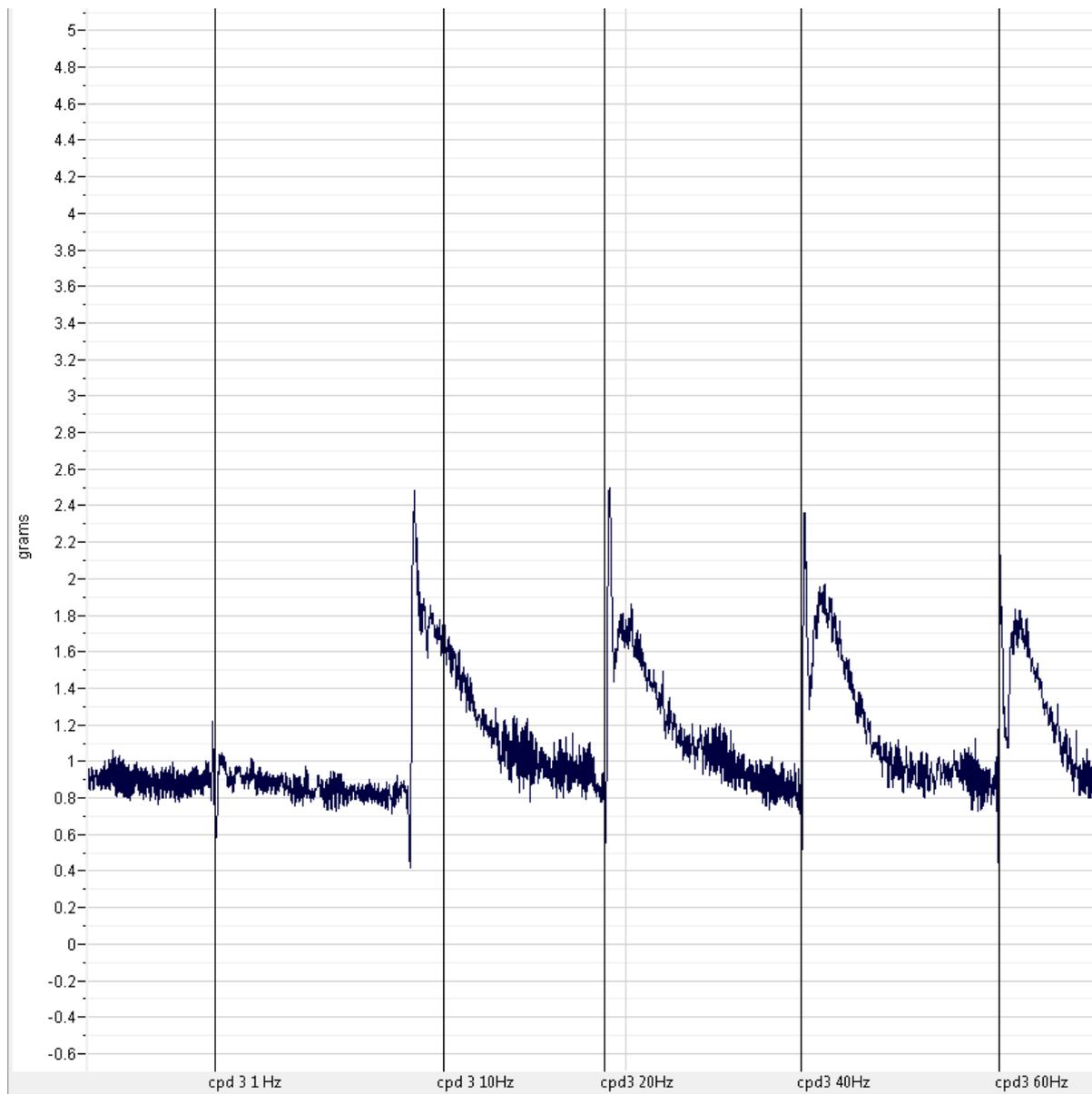
Electrical field stimulation trace in the presence of compound 4
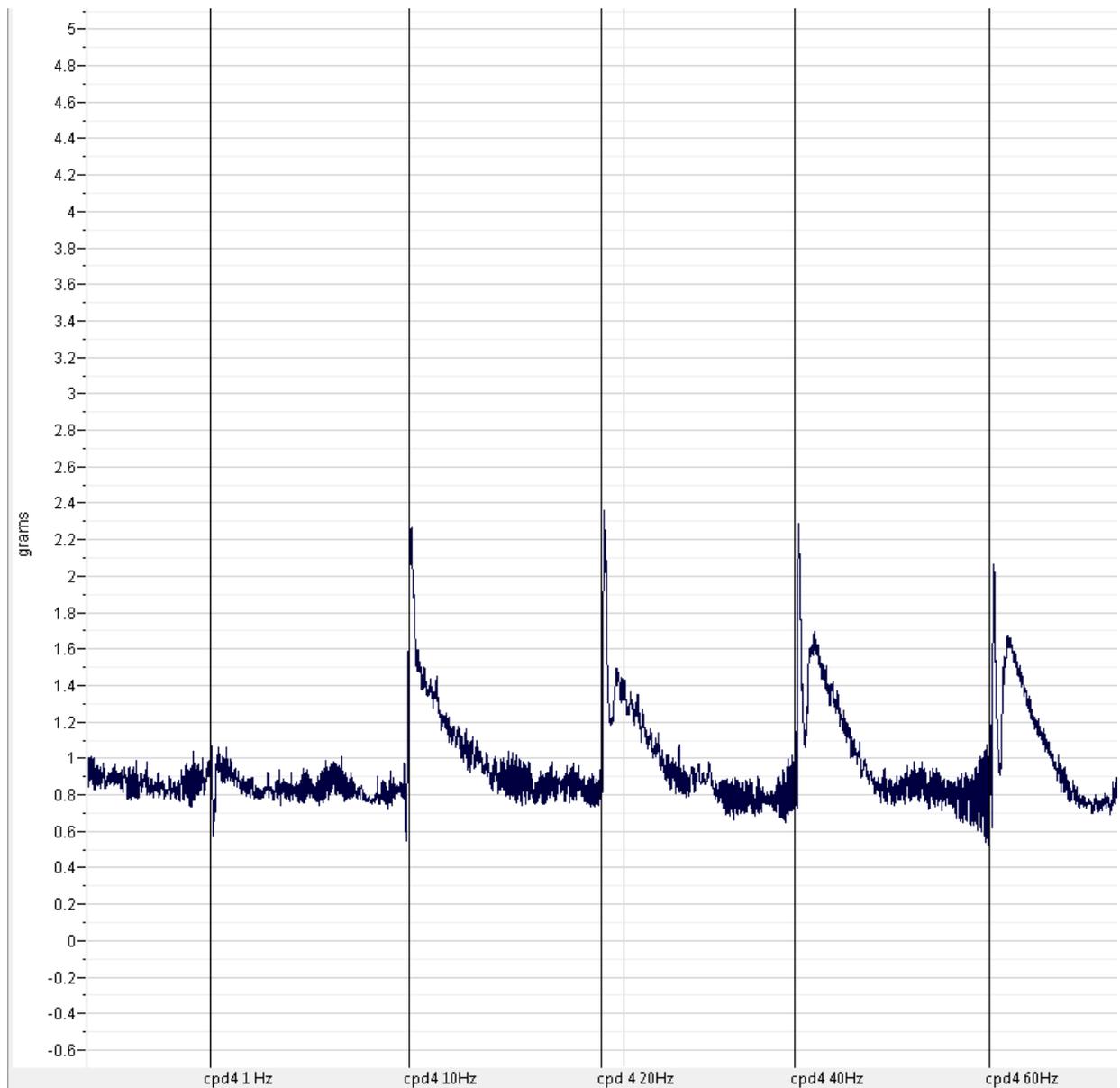
Electrical field stimulation trace in the presence of compound 5
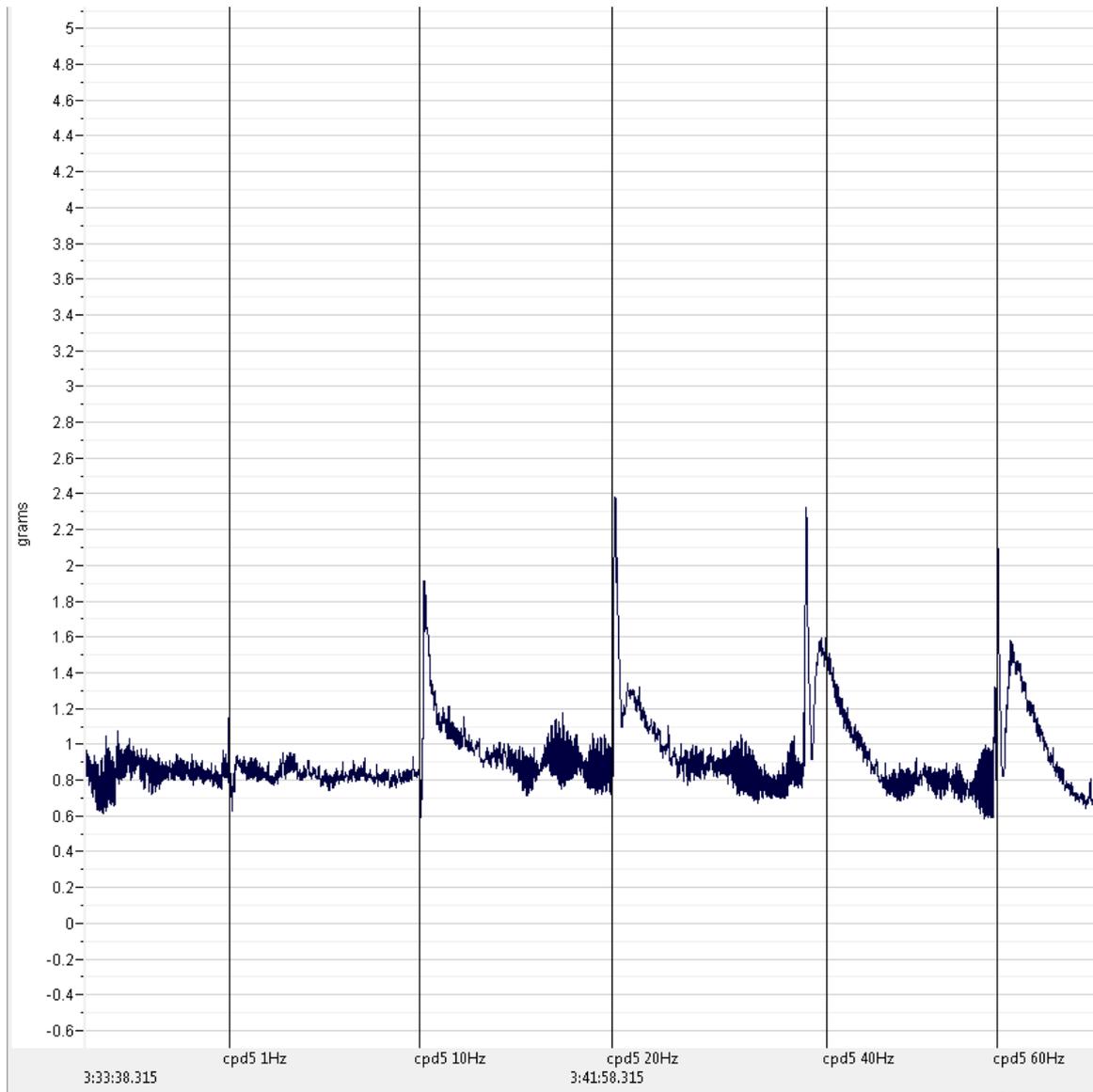
Electrical field stimulation trace in the presence of compound 6
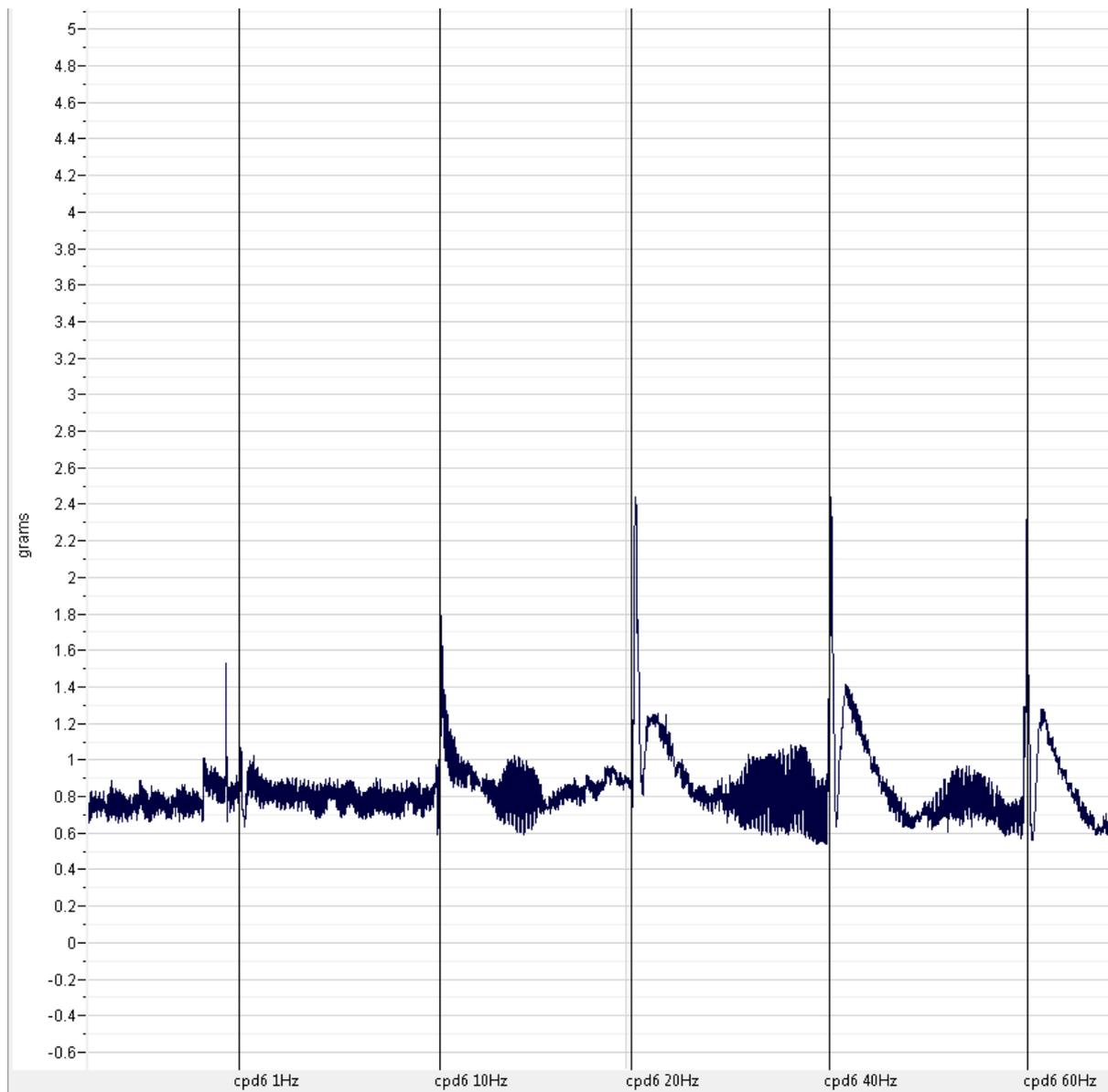
Electrical field stimulation trace in the presence of compound 7
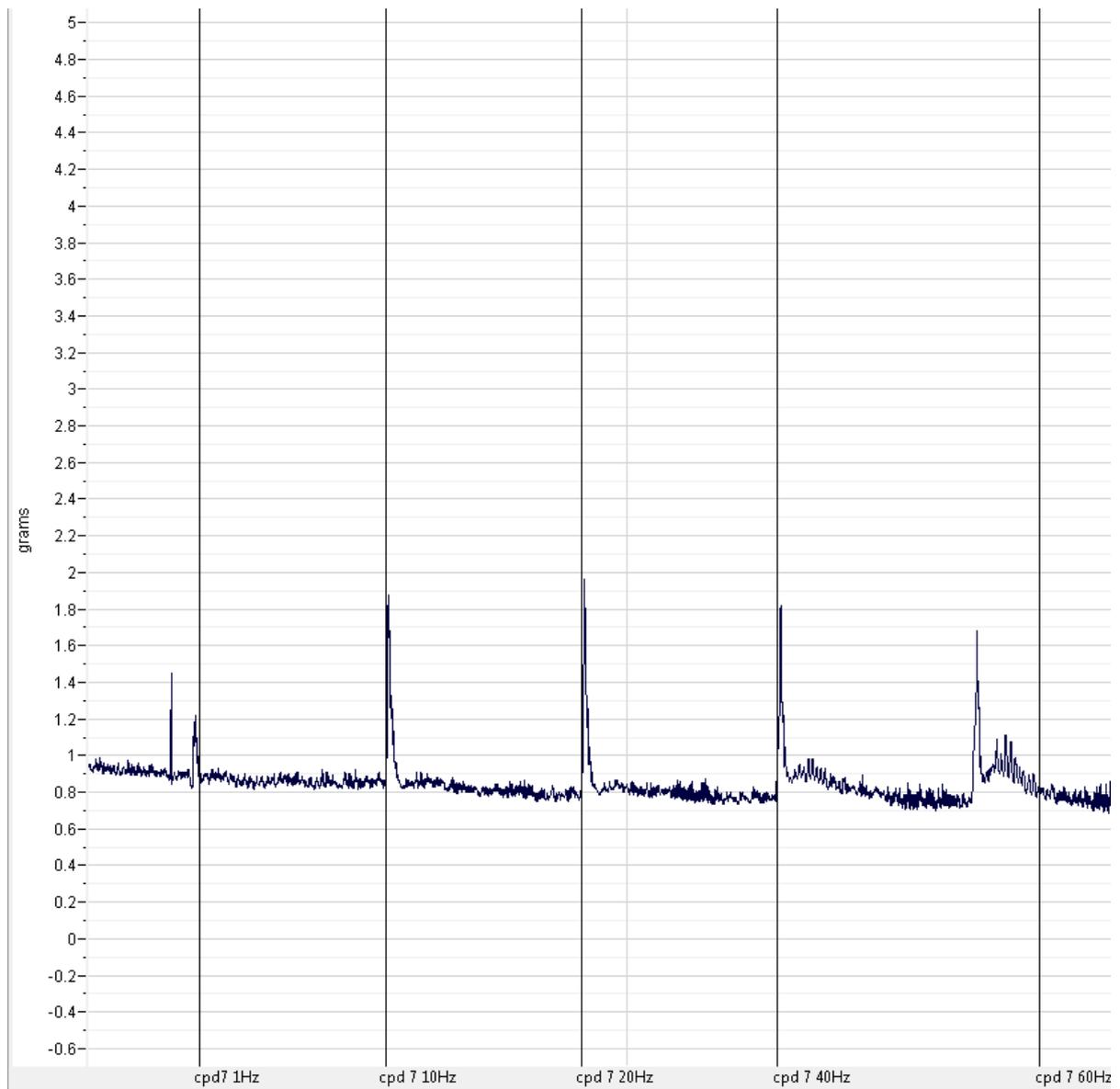
Electrical field stimulation trace in the presence of compound 8
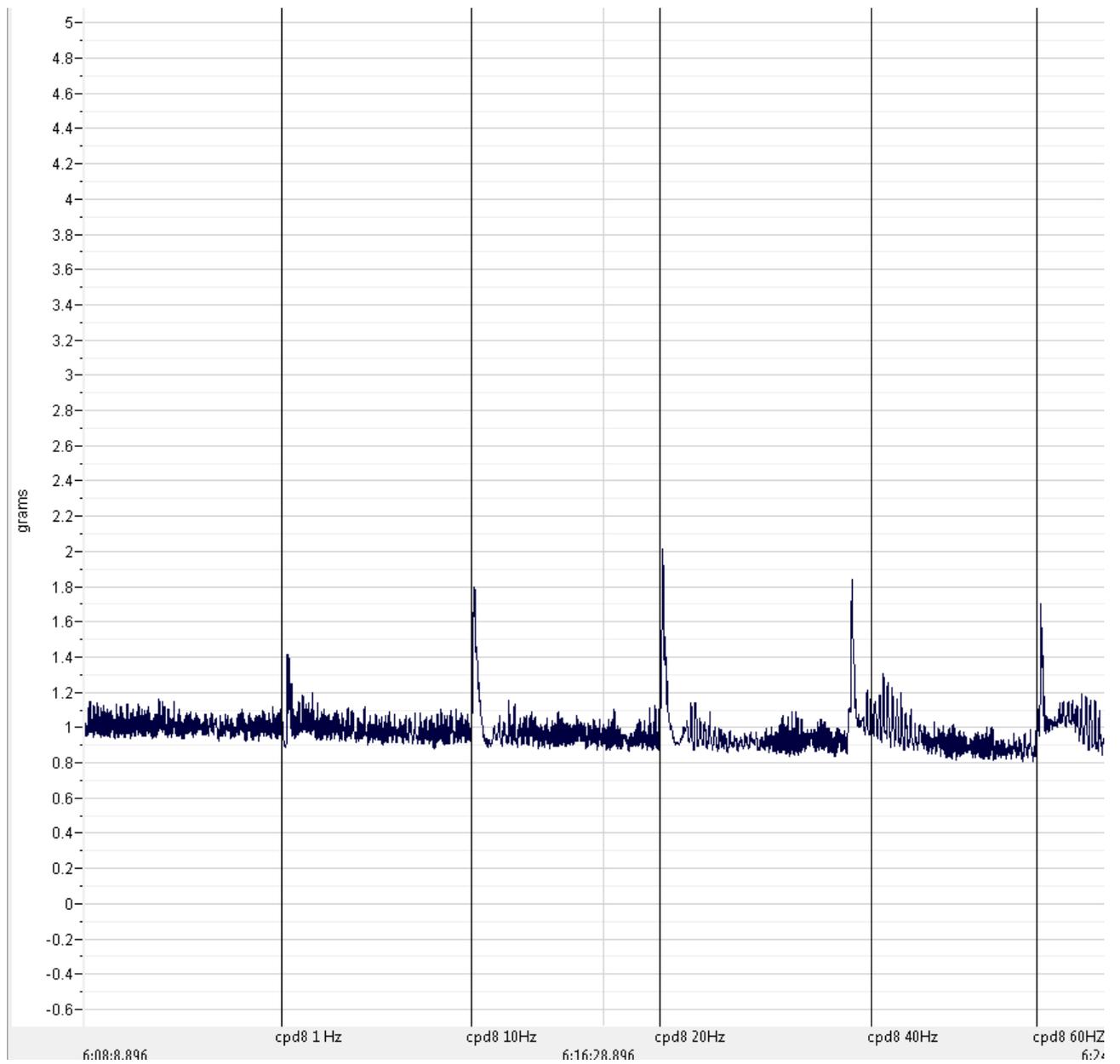
Electrical field stimulation trace in the presence of compound 9
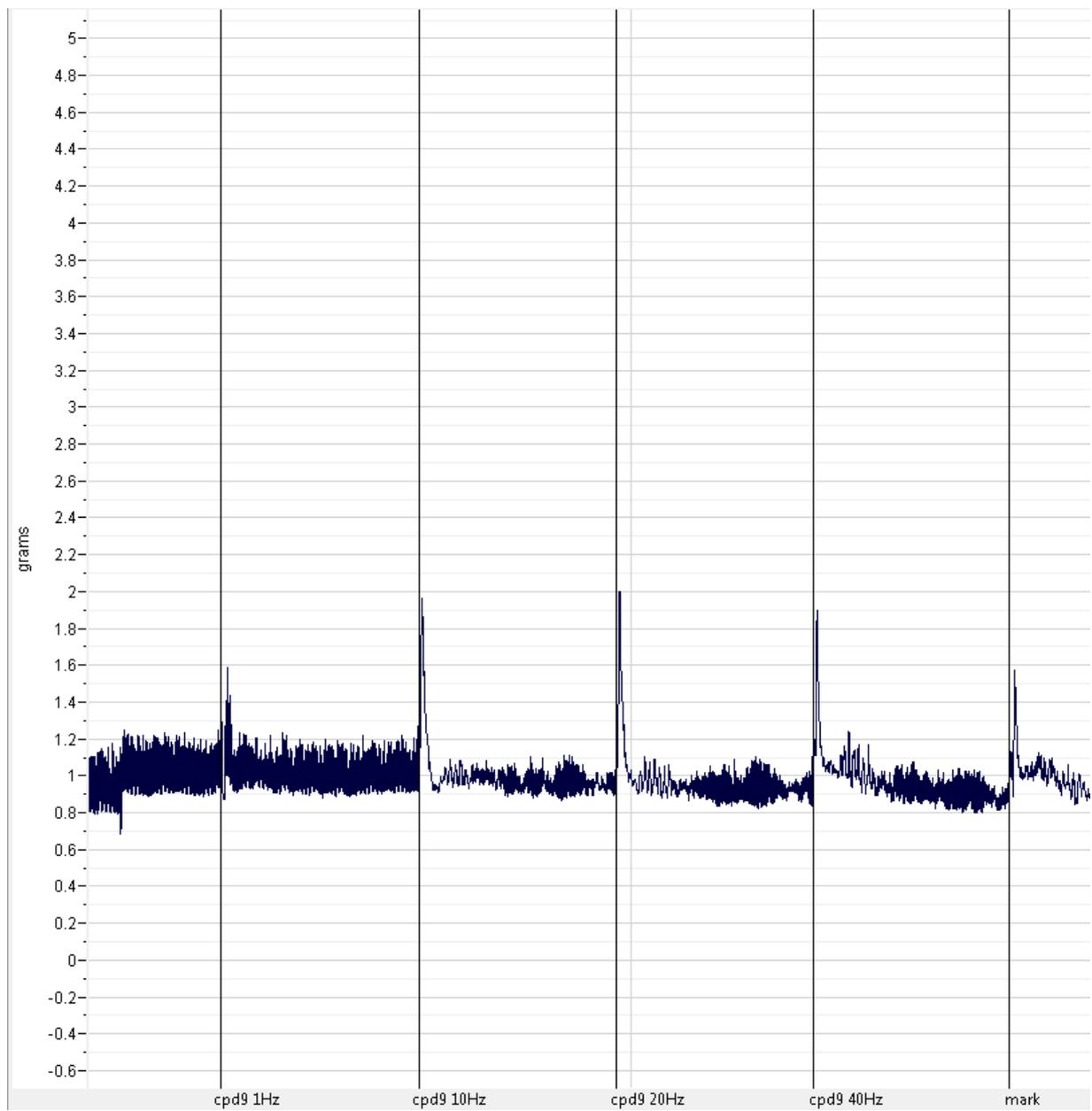
Electrical field stimulation trace in the presence of compound 10
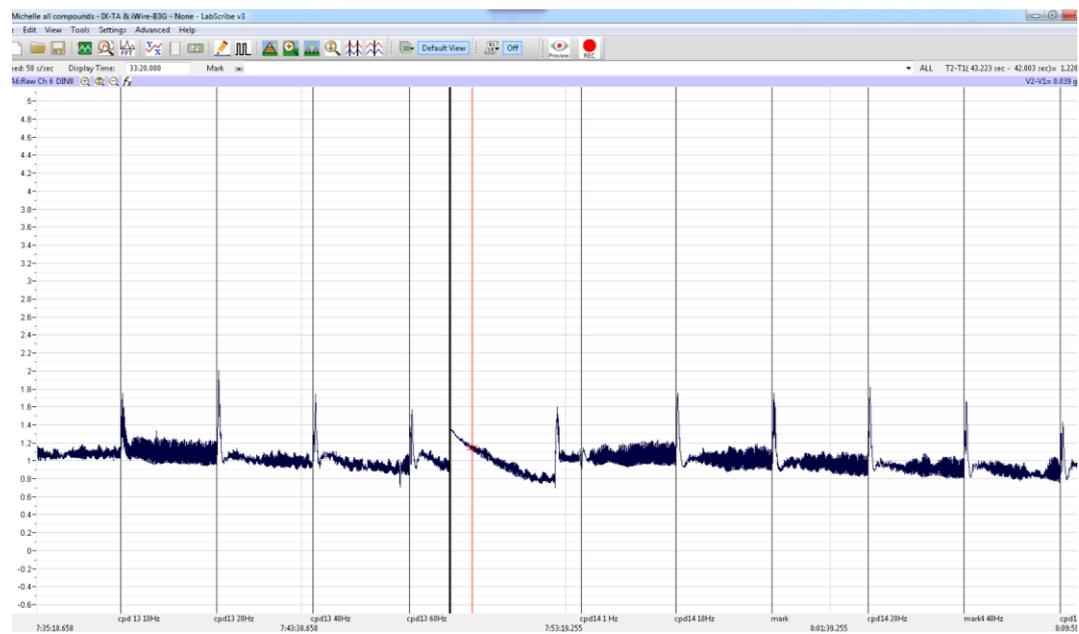
Electrical field stimulation trace in the presence of compound 11
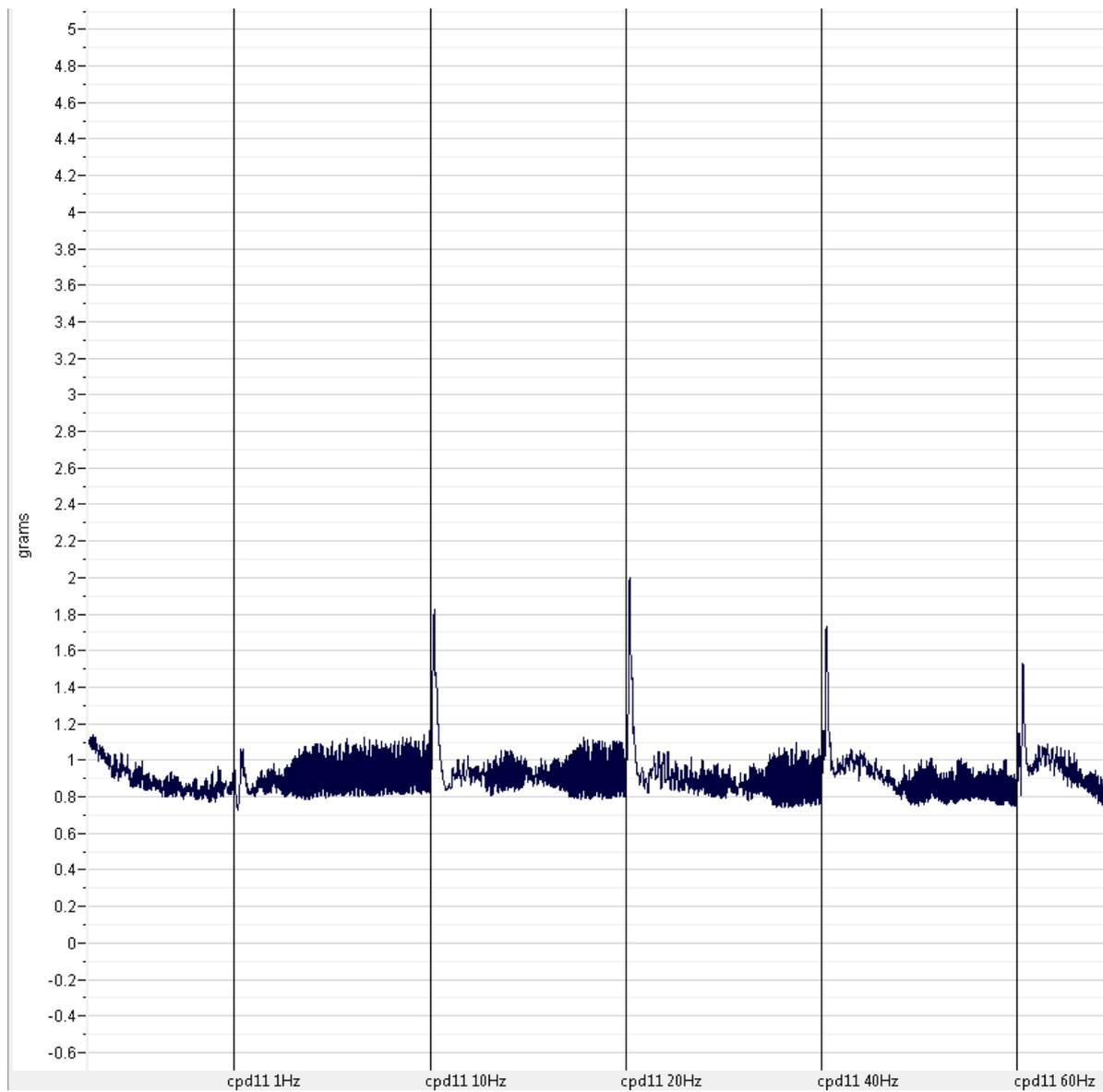
Electrical field stimulation trace in the presence of compound 12
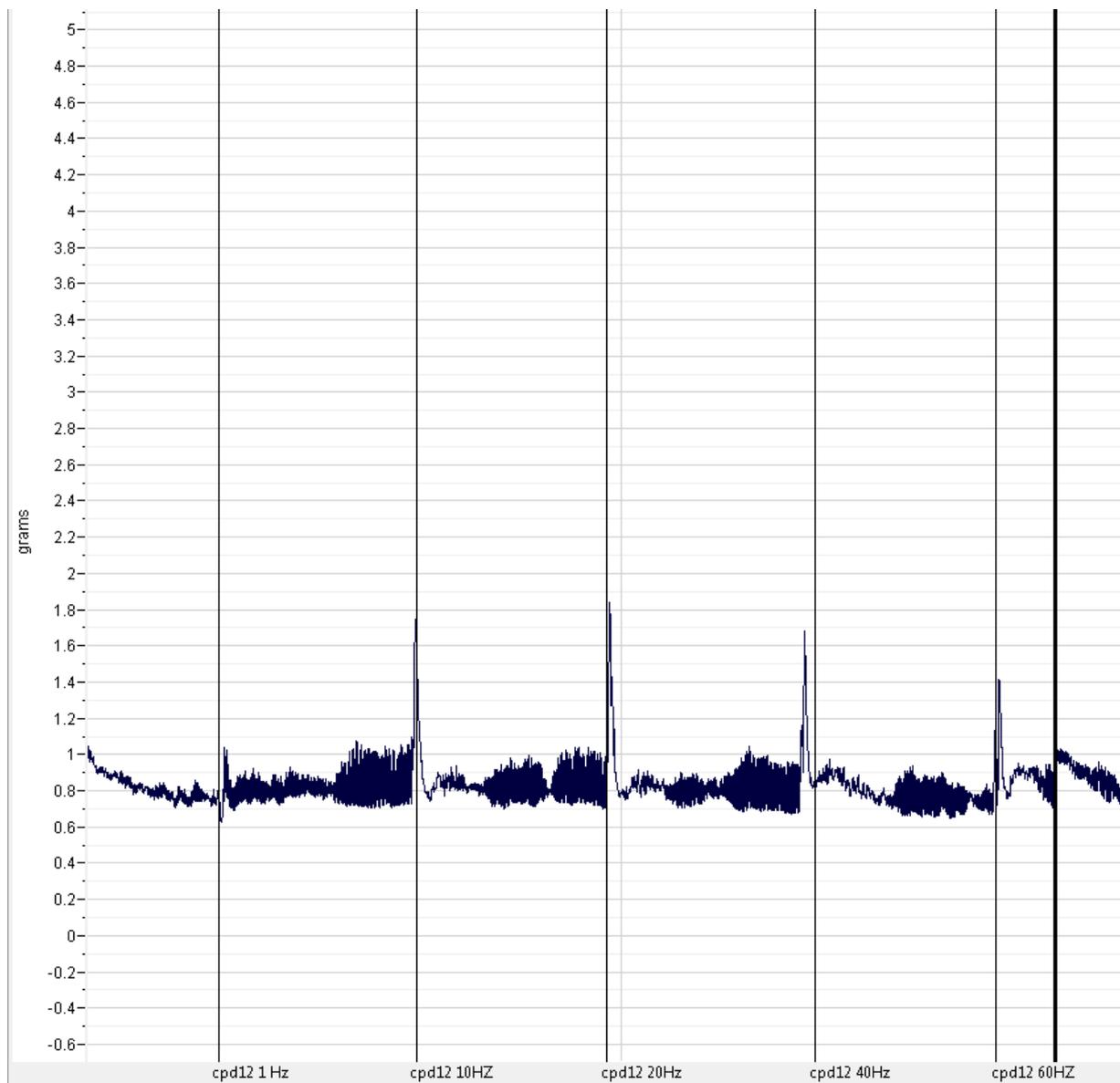
Electrical field stimulation trace in the presence of compound 13
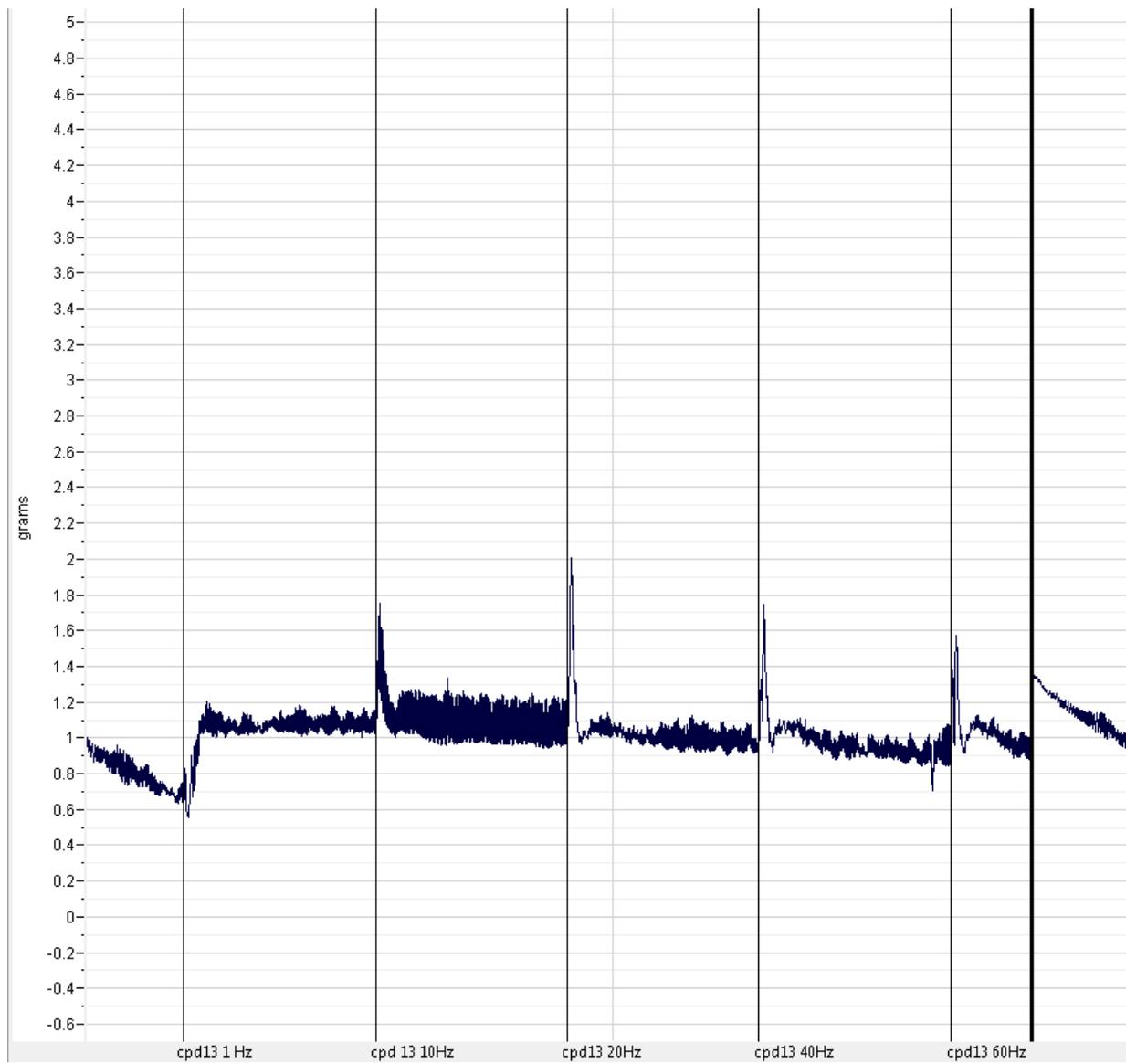
Posters

1. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "Investigations into the Next Generation of New Psychoactive Substances (NPS)". 1st School of Life and Medical Science Research Conference. University of Hertfordshire, UK, April 2015.

2. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "How QSAR models can be used to predict the biological activity of New Psychoactive Substances (NPS)". UK QSAR/PCF Forum meeting. GSK, Stevenage, UK, March 2016.

3. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "Quantitative Structure Activity Relationship (QSAR) models to predict biological activity of New Psychoactive Substances (NPS) in monoamine transporters". 1st School of Life and Medical Science Research Conference. University of Hertfordshire, UK, April 2016.

4. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "Quantitative Structure Activity Relationship (QSAR) models to predict biological activity of New Psychoactive Substances (NPS) in monoamine transporters" The IV international Conference on Novel Psychoactive substances (NPS) 2016, Budapest May 2016.

5. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "In Silico studies on the identification of novel synthetic cannabinoid structures". 2nd School of Life and Medical Science Research Conference. University of Hertfordshire, UK, April 2017.

6. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "*In Silico* studies into the next generation of New Psychoactive Substances (NPS)". The V international Conference on Novel Psychoactive substances (NPS) 2017, Vienna October 2017.

7. M. Botha, J. L. Stair, S. Fergus, S. B. Kirton. "Identification and Biological Evaluation of Potential Novel Synthetic Cannabinoid Molecules" The pharmaceutical analysis postgraduate research awards and careers symposium Royal Society of Chemistry, London, November 2017.