PUBLICATIONS

1. Tehan, B.G., Lloyd, E.J., Wong, M.G. Atypical Antipsychotics: Modelling, QSAR and Database Searching. in QSAR 2000. Dusseldorf: Prous Publishers, **2000**

2. Tehan, B. G.; Lloyd, E. J. and Wong, M. G. Molecular Field Analysis of Clozapine Analogues in the Development of a Pharmacophore Model of Antipsychotic Drug Action. *Journal of Molecular Graphics and Modelling*, **2001**, *19*, 417-426

3. Tehan, B. G.; Lloyd, E. J.; Wong, M. G. and Chalmers, D.K. Analysis of Agonism by Dopamine at the Dopaminergic D_2 G-Protein Coupled Receptor based on Comparative Modelling of Rhodopsin. *Molecular Simulations*, **2002**, 28, 865-888

4. Tehan, B.G.; Lloyd, E.J.; Wong, M.G.; Pitt, W.R.; Montana, J.G.; Manallack, D.T. and Gancia, E. Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids. *Quantitative Structure-Activity Relationships*, **2002**, *21*, 457-472

5. Tehan, B.G.; Lloyd, E.J.; Wong, M.G.; Pitt, W.R.; Manallack, D.T. and Gancia, E. Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 2: Application to Amines, Anilines and Various Nitrogen Containing Heterocyclic Compounds. *Quantitative Structure-Activity Relationships*, **2002**, *21*, 473-485.

6. Manallack, D.T., Tehan, B.G., Gancia, E., Ford, M.G., Livingstone, D.J., Whitley, D.C. and Pitt, W.R. A Consensus Neural Network-Based Technique for Discriminating Soluble and Poorly Soluble Compounds. *Journal of Chemical Information and Computer Sciences*, **2003**, 43, 674-679.

7. Wong, M. G.; Tehan, B. G. and Lloyd, E. J. Molecular Mapping in the CNS. *Current Pharmaceutical Design* **2002**, *8*, 1547-1570.

Errata

p19 last para, sentence 2: "units" for "untis"
p 53 caption to Figure 2.2: "ring plane through the centroid" for "ring centroid"
p 65 Table 2.4, heading of first column: "Conformer" for "Ligands"
p 69 para 3, last sentence: "Two compounds from the Philips dataset were not used" for "Two compounds were not used"
p 100 Figure 2.23: "N" for "O"
p 109 caption to Figure 2.30: "G and G'" for "D and D'"
p 114 para 1, sentence 1: "spiperone (see Figure 3.13, page 180)" for "spiperone"
p 118 last para, sentence 1: "Figure 2.14" for "Figure14"
p 119 para 1, last sentence: "in the" for "the in"
p 174 para 1, sentence 1: "chemists" for "chemist"
p 174 last para, sentence 6: "too" for "to"
p 176 para 1, sentence 3: "as it contained" for "as contained"
p 193 last para, sentence 3: "quite" for "quiet"
p 211 last para, sentence 1: "show" for "shows"
p 221 reference 13: "2002, 28, 865-888" for "2001, in press"
p 228 para 1, sentence 1: "processes" for "process"
p 229 para 2, sentence 1: "for" for "of"
p 243 after equation (4): Omit "model"
p 246 last para, sentence 1: "for the reduced set" for "reduced set"
p 255 para 1, sentence 2: "rotatable" for "rotational"
p 257 last para, last sentence: "2,3,5,6-tetrachloro" for "2,3,5,6-dichloro"
p 262 para after equation (51), last sentence: "borne" for "bourn"
p 272 para 2, sentence 4: "were" for "was"
p 282 para 1, sentence 3: "conformers" for "conformer"

Addendum

p 232 para 1, sentence 4: Insert new sentence after "excluded." "An example of a toxic functionality excluded from the analysis was 1,2-phenylenediamine."
p 232 para 2, after sentence 4 ending "within the datasets." Insert: "The ambiguous compounds identified in Table 1, p 240 were identified initially as outliers in the generation of equations. All outliers, compounds with a pK_a more than 1-2 log units outside the value predicted by the derived equation, were checked against the literature where available."
p288 para 3, after sentence 3 ending "of spiperone." Insert: "The apparent contradiction between the volumes at the active sites of the D_2 and D_4 models is due to the different orientation of the space surrounding the active site in the two receptors. In the D_2 receptor model this space is above and below the p-fluorophenyl group of spiperone, whereas in the D_4 receptor model it is adjacent to the fluoro atom of that same group. This difference has potential to be used in the design of molecules with specificity for one or other of these receptor sub-types."

COMPUTATIONAL DESIGN OF NOVEL ANTIPSYCHOTICS

Submitted in total fulfilment for the Degree of
Doctor of Philosophy

by

Benjamin Tehan
B. AppSc (Swinburne University of Technology)

Victorian College of Pharmacy (Monash University)
Department of Medicinal Chemistry
381 Royal Parade, Parkville 3052
Victoria, Australia

July, 2003

ABSTRACT

The primary aim of this work has been to develop D_4 selective antipsychotics, using a number of complementary computational techniques, specifically ligand and structure-based design, database mining and bioavailability assessment.

In chapter 1 a review is presented of the development of antipsychotics and the various hypotheses proposed to explain their efficacy in controlling schizophrenia. One enduring hypothesis is the involvement of dopamine receptor subtypes, mainly D_2, and this thesis explores the selective involvement of D_4 relative to D_2 receptors. An overview is then presented of modern computational chemistry techniques in drug design and development.

In chapter 2 tricyclic and extended pharmacophores were developed from conformational analysis of high affinity antipsychotics and D_2 and D_4 active compounds. These pharmacophore models were then used as the basis for developing predictive CoMFA models for generating potentially D_4 selective ligands.

Chapter 3 involved building homology models of the D_2 and D_4 receptors based on rhodopsin with clozapine, and alternatively, spiperone bound in the proposed active site. As part of the model building process, mutagenesis experiments were examined to assess the likely binding orientation of the ligands within the receptors. Models were then built of D_2 and D_4 receptors using the proposed binding orientations and conformations deduced from the pharmacophores of chapter 2. All constructed models were analysed and grouped into families and assessed against mutagenesis data. The best resultant models were then used as templates for the superimposition of the CoMFA models generated in chapter 2. The additional information from these analyses aided in the generation of ideas for D_4 selective compounds.

In order to provide some assessment of bioavailability, a comprehensive set of QSPR models were developed in chapter 4. Compounds with experimentally determined pK_a values were extracted from the PHYSPROP© database and filtered so as to remove

compounds that were not drug-like. Semi-empirical frontier electron theory descriptors were calculated for all compounds and linear regression models constructed correlating the descriptors and experimentally determined pK_a values. The benefits and limitations of this technique were then discussed in relation to published techniques and predictive pK_a methods. The derived models were used to predict the pK_a's of compounds proposed in chapter 5.
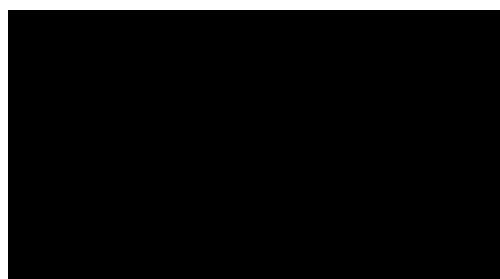
Chapter 5 involved using the information gained from chapters 2 and 3 in combination with database searching using the pharmacophores derived from chapter 2, to propose several series of compounds. The CoMFA models derived in chapter 2 were used to predict the affinities of proposed compounds, showing that they are D_4 selective, and their pK_a values were predicted using the linear regression models developed in chapter 4. The pK_a's of all compounds were predicted to be within a suitable range to enable them to be unionised for crossing the blood brain barrier, yet appropriately ionised for interaction at target receptors.

DECLARATION

I declare to the best of my knowledge that the material presented in this thesis has not been previously submitted by me or any other person for a degree at this or any other university.

The studies contained in this thesis represent my own original work except where established methodology has been acknowledged within the text.

This thesis is less than 100,000 words in length, excluding Tables, Figure legends, Appendices, References and Cover Pages.

Benjamin Tehan   July 2003

TABLE OF CONTENTS

Chapter 1 - Introduction

Chapter 2 - Ligand-Based Design

Chapter 3 - Structure-Based Design

Chapter 4 - pK$_a$ Calculation

Chapter 5 - Compound Design

ABBREVIATIONS

APD	antipsychotic drug
CNS	central nervous system
DA	dopamine
5HT	5-hydroxy-tryptamine (serotonin)
GPCR	G protein-coupled receptor
EPS	extrapyramidal side-effects
SNC	substantia nigra par compacta
VTA	ventral tegmental area
cAMP	cyclic adenosine monophosphate
GABA	γ-amino butyric acid
Glu	glutamate
DRN	dorsal raphe nucleus
MRN	median raphe nucleus
NMDA	*N*-methyl-D-aspartate
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazole propionate
PCP	phencyclidine
LSD	lysergic acid diethylamide
ADME/T	adsorption, distribution, metabolism, excretion and toxicity
AAA	active analogue approach
GASP	genetic algorithm similarity program
RMS	root mean square
QSAR	quantitative structure activity relationship
QSPR	quantitative structure property relationship
SEAL	steric and electrostatic alignment
GA	genetic algorithm
FBSS	field-based similarity searching
MLR	multiple linear regression
ANN	artificial neural network
CoMFA	comparative molecular field analysis
CoMSIA	comparative molecular similarity indices analysis

PCA	principle component analysis
PLS	partial least squares
LOO	leave one out
3D	three dimensional
DNA	deoxyribonucleic acid
GOLPE	generating optimal linear PLS estimations
FFD	fractional factorial design
SRD	smart region definition
HQSAR	hologram quantitative structure-activity relationships
vHTS	virtual high throughput screening
ARD	automatic relevance determination
VLA	volume learning algorithm
SOM	self organising maps
BPNN	forward-feed back-propagation neural networks
NMR	nuclear magnetic resonance
BBB	blood brain barrier
HOMO	highest occupied molecular orbital
LUMO	lowest occupied molecular orbital

PUBLICATIONS

1. Tehan, B.G., Lloyd, E.J., Wong, M.G. Atypical Antipsychotics: Modelling, QSAR and Database Searching. in QSAR 2000. Dusseldorf: Prous Publishers, **2000**

2. Tehan, B. G.; Lloyd, E. J. and Wong, M. G. Molecular Field Analysis of Clozapine Analogues in the Development of a Pharmacophore Model of Antipsychotic Drug Action. *Journal of Molecular Graphics and Modelling,* **2001**, *19,* 417-426

3. Tehan, B. G.; Lloyd, E. J.; Wong, M. G. and Chalmers, D.K. Analysis of Agonism by Dopamine at the Dopaminergic D_2 G-Protein Coupled Receptor based on Comparative Modelling of Rhodopsin. *Molecular Simulations,* **2002**, 28, 865-888

4. Tehan, B.G.; Lloyd, E.J.; Wong, M.G.; Pitt, W.R.; Montana, J.G.; Manallack, D.T. and Gancia, E. Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids. *Quantitative Structure-Activity Relationships,* **2002**, *21,* 457-472

5. Tehan, B.G.; Lloyd, E.J.; Wong, M.G.; Pitt, W.R.; Manallack, D.T. and Gancia, E. Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 2: Application to Amines, Anilines and Various Nitrogen Containing Heterocyclic Compounds. *Quantitative Structure-Activity Relationships,* **2002**, *21,* 473-485.

6. Manallack, D.T., Tehan, B.G., Gancia, E., Ford, M.G., Livingstone, D.J., Whitley, D.C. and Pitt, W.R. A Consensus Neural Network-Based Technique for Discriminating Soluble and Poorly Soluble Compounds. *Journal of Chemical Information and Computer Sciences,* **2003**, 43, 674-679.

7. Wong, M. G.; Tehan, B. G. and Lloyd, E. J. Molecular Mapping in the CNS. *Current Pharmaceutical Design* **2002**, *8,* 1547-1570.

ACKNOWLEDGEMENTS

DEDICATION

For my Bunny and my Mother and Father.

Chapter 1

Introduction

1.1 SCHIZOPHRENIA

Many people do not have a clear understanding of schizophrenia, which is a physical illness, just like diabetes or asthma. Schizophrenia has typical signs and symptoms that are recognisable in patients with this illness, and like all illnesses, the symptoms vary from person to person. The symptoms of schizophrenia were initially classified by Emil Kræplin and later expanded upon by Eugene Bleuler, whereby the term *schizophrenia* (Gr: 'split mind') was introduced. From this work and from contributions from many other eminent psychiatrists the disease state of schizophrenia has been categorised into five distinct subtypes[1]; catatonic, paranoid, disorganised, simple and residual schizophrenia.

Schizophrenia affects approximately 1% of the population with the onset of the illness occurring between the ages of 15 to 30 and persisting in many cases for the lifetime of the patient[2]. There appears to be no distinct cultural or socio-economic boundaries for the prevalence of schizophrenia and both sexes are equally affected, although onset usually occurs earlier in men. Genetic factors play some part in developing schizophrenia: an identical twin of a schizophrenic will have a 48% chance of developing schizophrenia[3]; a child of a schizophrenic will have a 13% chance and a grandchild of a schizophrenic will have a 5% chance. The symptoms of the disease are split into two areas, positive symptoms and negative symptoms, consistent with the hypothesised areas of dysfunction within the brain. Positive symptoms, which are the most recognisable symptoms of schizophrenia, are mental experiences that most people don't have[4] and take the form of hallucinations, delusions, disorganised speech and disorganised behaviour. Negative symptoms are behaviours that are absent from a person's regular behaviour[4]. They are the most common symptoms of schizophrenia and include a lack of motivation, social withdrawal, poverty of speech, difficulty concentrating and difficulty in showing or feeling emotion. Cognitive impairments are also classified under the negative symptoms but many psychiatrists believe that they should be classified separately[4]. All these symptoms are governed be various pathways of the brain, which are discussed at further length in the next section 'Receptor Hypotheses'. The devastating toll this disease takes on the quality of life of an individual is reflected in the high rate of suicide amongst schizophrenics.

Individuals with schizophrenia have a rate of suicide between 4 and 13%, and between 25 and 50% of all schizophrenics attempt suicide[5] at some stage of their life.

Despite decades of research the causes of schizophrenia are still unknown. Although there is a strong genetic component to the inheritance of schizophrenia other non-genetic factors implicated include: season of birth, place of birth, influenza infection during pregnancy and obstetrical complications[6]. Although the causes of schizophrenia remain unknown it is thought that the symptoms are due to a disturbance or imbalance of the subject's brain chemistry.

1.1.1 Receptor Hypotheses

Neural transmission in the brain takes place through a variety of small molecules and peptides, or neurotransmitters, which interact with specific receptor sites to elicit a response. Any abnormalities in neurotransmitter levels and binding can cause various neurological problems, including schizophrenia. All the major neurotransmitters in the brain have been linked in some way with schizophrenia. However the major neurotransmitters associated with schizophrenia appear to be dopamine and serotonin.

1.1.1.1 Dopamine (DA) and its Receptors

Dopamine (see Figure 1.1) is one of the major catecholamines in the central nervous system (CNS). It is involved in the regulation of a variety of functions, including locomotor activity, emotion and effect, and neuroendocrine secretion.

Figure 1.1. Dopamine
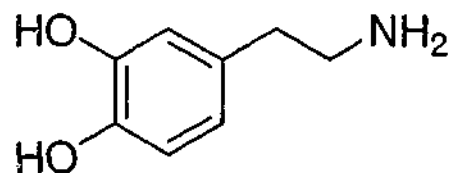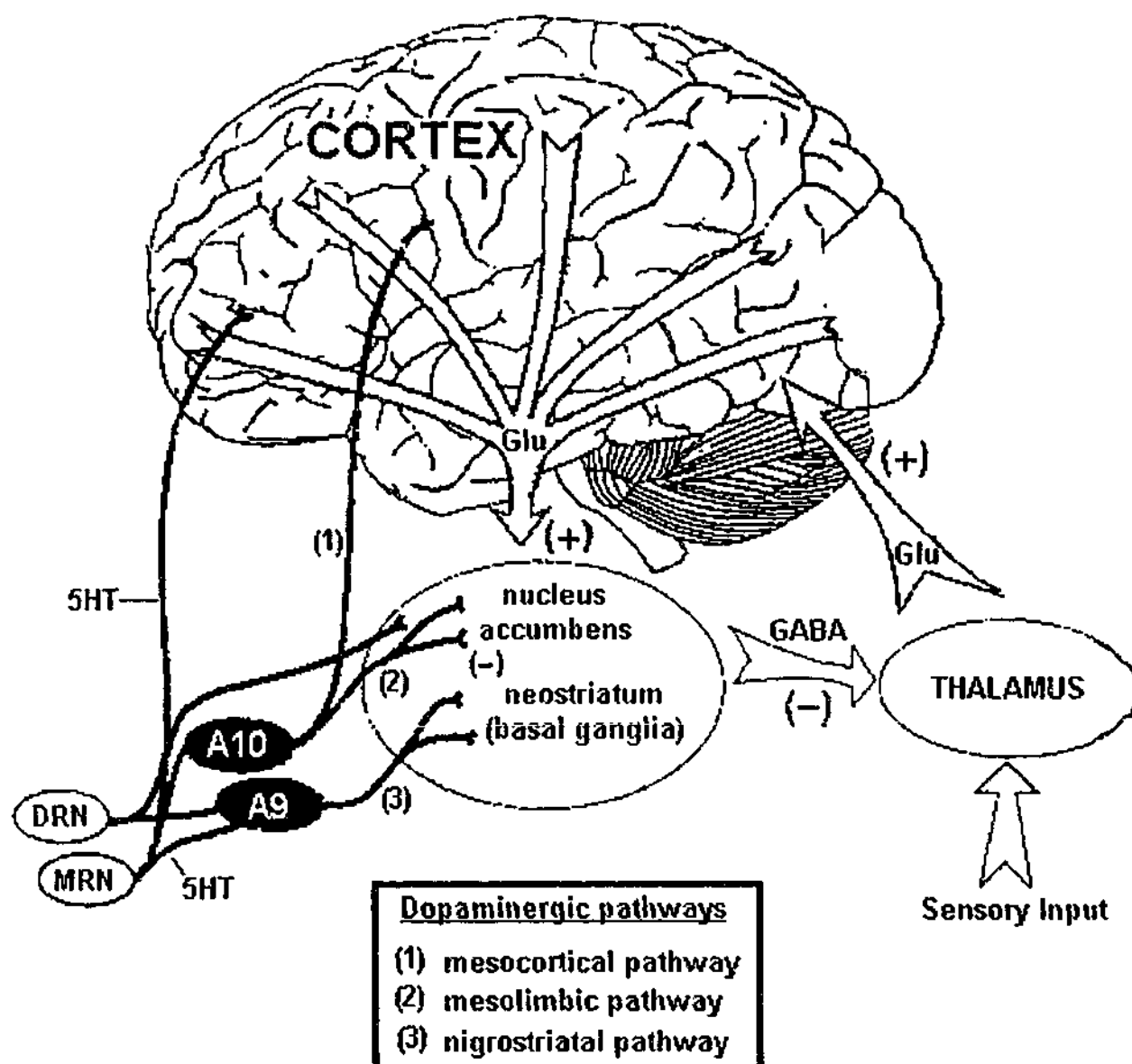
Figure 1.2. Brain Diagram with various neurotransmitter pathways, adapted from Schmidt et al.[7]

Pathways[8]

In relation to schizophrenia, the significant dopaminergic pathways in the brain (Figure 1.2) emanate from groups of cells in the midbrain (A9 and A10) and hypothalamus. Dopaminergic neurons, designated A9, ascend to the neostriatum and this pathway is called the *nigrostriatal pathway* (see (3) in Figure 1.2). Dopaminergic neurons, designated A10, innervate mesolimbic regions (nucleus accumbens, olfactory tubercle, septum), and are called *the mesolimbic pathway* (see (2) in Figure 1.2). In addition, A10 dopaminergic neurons innervate mesocortical sites (cingulate, entorhinal, prefrontal, and pyriform cortices), and this is referred to as *the mesocortical pathway* (see (1) in Figure 1.2). In the hypothalamus, groups of dopaminergic cell bodies project to the pituitary and modulate neuroendocrine regulation of prolactin secretion; this

projection is referred to as *the tuberoinfundibular pathway*. Another group of neurons, A8, are also located in the midbrain and contribute to the dopaminergic innervation of striatal and mesolimbic, but not mesocortical, sites. Together, the A8 and A9 neurons make up 70% of brain dopamine and are involved in the modulation of motor behaviour.

The mesolimbic and mesocortical pathways are thought to play a prominer.. role in psychiatric illness. The mesolimbic pathway plays a role in emotions, memory and reward systems and therefore hyperdopaminergic activity of this pathway is associated with psychosis and the positive symptomatology of schizophrenia. The mesocortical pathway is involved in cognition, motivation, emotional control and reasoning, so that hypodopaminergic activity in this pathway leads to the negative symptomatology of schizophrenia. In the treatment of schizophrenia blockade of the mesolimbic pathway with antipsychotic drugs ameliorates positive symptoms. However concomitant blockade of the nigrostriatal and tuberoinfundibular pathways results in movement disorders and troublesome neuroendocrine side-effects (breast enlargement, excessive lactation and menstrual irregularities), respectively.

Receptors[9]

There are five main subtypes of dopaminergic receptor having a relative abundance in the order $D_1 > D_2 > D_3 > D_5 > D_4$. They are all G protein-coupled receptors (GPCRs), in which signal transduction involves guanosine tri-phosphate, and are grouped into two main families, D_1-like and D_2-like, with the major difference being the ability to either up-regulate or down-regulate cyclic adenosine mono-phosphate (cAMP).

The D_1-like group stimulate cAMP formation and consist of the receptors D_1 and D_5. The D_1 receptor is the most common dopamine receptor and is expressed in higher levels than any other dopamine receptor. It is located in areas of dopaminergic control such as the striatum, nucleus accumbens, limbic system and hypothalamus. The D_5 receptor is expressed at a much lower level and shows a distribution restricted to the hippocampus and hypothalamus.

The D_2-like group inhibit cAMP formation and consist of the receptors D_{2LONG}, D_{2SHORT}, D_3 and D_4 with various splice subtypes. The D_2 receptor is mainly found in the striatum, nucleus accumbens and olfactory tubercle where it is co-expressed with γ-aminobutyric acid (GABA) neurons. It is also found in the substantia nigra pars compacta, A9, and the ventral tegmental area, A10, and in the pituitary where it regulates the production and secretion of prolactin. The expression pattern of the long and short D_2 receptor isoforms appears to vary over regions specified, with the longer isoform predominating in the striatum and nucleus accumbens and the shorter isoform predominating in the dopaminergic cell bodies and axons. This strongly suggests that the D_{2SHORT} isoform is a dopamine autoreceptor, whereas the D_{2LONG} isoform is likely to be a postsynaptic receptor[10]. The D_3 receptor is specifically distributed in limbic areas like the shell of the nucleus accumbens, olfactory tubercle and the islands of Calleja, and hence has received much attention. The D_4 receptor, although the least abundant, is also of considerable interest due to its distribution within the limbic and cortical regions and relatively low abundance in the striatum.

1.1.1.2 Dopamine Hypothesis

Carlson and Lindqvist first postulated in 1963 the hypothesis that the dopaminergic system is overactive in schizophrenics. This was based on the finding that the administration of antipsychotics, haloperidol and chlorpromazine, lead to an increase in dopamine metabolites in mouse brain[11]. This hypothesis was further strengthened by the fact that psychostimulants, such as amphetamine, increase dopaminergic transmission by increasing release of dopamine, inducing psychotic states resembling those observed in the positive symptoms of schizophrenia[12]. This hypothesis was further substantiated based on the findings of a correlation between the potency of antipsychotics as dopamine D_2 antagonists and their clinically effective doses[13].

Based on these findings it had been proposed that the symptoms of schizophrenia are due to hyperdopaminergic activity within the brain. However a few observations challenge the dopamine hypothesis of schizophrenia, namely the slow onset of effect of antipsychotics and the fact that some patients do not respond to conventional antipsychotic medication. In addition to this, negative symptoms are usually not

improved but worsened by typical antipsychotic drugs, indicating that these symptoms are associated with a deficit of dopaminergic function.

A revised dopamine hypothesis was based on findings by Pycock *et al*[14], where they were able to show that lesions of rat prefrontal cortex dopamine-containing neurons resulted in a dopamine deficiency in the mesocortical pathway and enhanced dopaminergic activity in the mesolimbic pathway. This work led researchers to hypothesise that schizophrenia may result from damage to the frontal lobe during development, causing dopamine deficiencies in the mesocortical pathways and the observation of negative symptoms. The negative feedback from this mesocortical dopamine deficiency to the ventral tegmental area results in hyperdopaminergic activity of the mesolimbic system and observation of the positive symptoms of schizophrenia.

1.1.1.3 Dopamine D_4 Hypothesis

The dopamine D_4 hypothesis had its origins in the early nineties with the discovery of the D_4 receptor[15]. The two main driving forces of the D_4 hypothesis were the pattern of distribution of the D_4 receptors within the brain and the D_4 selectivity exhibited by the atypical antipsychotic drug clozapine. In addition to this a significant elevation of D_4 receptor density in post-mortem schizophrenic brain tissue had been reported by Seeman *et al*[16]. However other groups have failed to reproduce these findings[17], although recently Stefanis *et al*[18] have shown a three fold increase in D_4 receptor *m*RNA in the frontal cortex of schizophrenics compared to controls.

As mentioned earlier, the D_4 receptors are located mainly in the mesocortical and mesolimbic regions of the brain and antagonism of these pathways was thought to give rise to the amelioration of the symptoms of schizophrenia without causing significant side-effects through blockade of the nigrostriatal pathway. This train of thought resulted in many companies putting significant resources into the development of D_4 selective antagonists[19-32]. A number of these compounds, (L-745870, U-101958 and fananserin) made it to phase II trials. However L-745870 and U-101958 were later shown to be partial agonists at the D_4 receptor[33] and thus trials were discontinued. Fananserin trials have been stopped as no statistical difference was seen between this

compound and placebo[34]. Obviously the two partial agonists, L-745870 and U-101958, are not ideal compounds with which to test the D_4 antagonism hypothesis. Fananserin, although D_4 selective with respect to D_2, is also one order of magnitude more $5HT_{2A}$ selective over D_4[35]. This $5HT_{2A}$ selectivity over D_4 also may not be ideal to test the D_4 hypothesis alone. Interestingly another $5HT_{2A}$ and D_4 selective compound, belaperidone, has shown some promise[36] with phase II trials being completed and results from phase III trials pending. The results from further trials of antagonists selective only for D_4 will serve to determine whether or not this artefact of clozapine is responsible for its favourable side-effect profile.

1.1.1.4 Serotonin (5HT) and its Receptors

It is widely known that the serotonergic pathways contribute to the regulation of a variety of psychological and biological functions. Mood, anxiety, arousal, attention, aggression, suicidality, and cognition are among the former, and sleep-wake cycle, appetite, pain sensation, and brain maturation are among the latter.

Figure 1.3. Serotonin
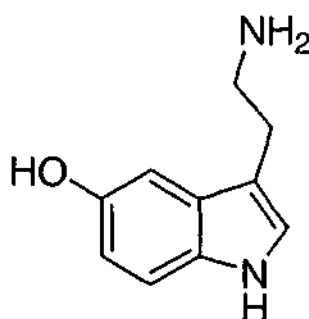
Pathways

The serotonergic pathways (see Figure 1.2) primarily arise from the Raphe nuclei, which are divided into two main areas, the dorsal (DRN) and median (MRN) Raphe nuclei. The DRN project mainly to the substantia nigra pars compacta (SNC) (A9), the striatum and the nucleus accumbens, whilst the MRN project mainly to the SNC (A9), ventral tegmental area (A10) and prefrontal cortex.

Receptors[37]

There are at least 14 separate types of 5HT receptors existing in seven major families $5HT_{1-7}$. They are all members of the GPCR superfamily, except $5HT_3$ which is an ion-channel receptor belonging to the GABA receptor superfamily.

Receptors belonging to the $5HT_1$ family are adenylate cyclase-linked GPCRs, sharing 40% amino acid sequence homology. These receptors are characterised by a high affinity for serotonin and their ability to down regulate adenylate cyclase.

$5HT_2$ receptors are coupled to G proteins, linked to phospholipase C, and stimulate phosphoinositol turnover. They share 60% amino acid sequence homology and a low affinity for serotonin. $5HT_{2A}$ is often referred to as the classic $5HT_2$ receptor and is present in high concentration in the olfactory bulb, hippocampus, frontal cortex, and piriform and entorhinal cortices. $5HT_{2B}$ is located primarily in peripheral tissue such as the stomach or lungs. $5HT_{2C}$, previously known as $5HT_{1C}$, has been reclassified on the basis of its amino acid sequence and second messenger characteristics. It is present in the choroid plexus, anterior olfactory nucleus, piriform and entorhinal cortices, striatum, and amygdala.

$5HT_3$ receptors are present in low density in limbic and striatal structures. They facilitate the release of serotonin, dopamine, and GABA, and they inhibit acetylcholine and noradrenaline release.

Other 5HT GPCRs, linked to adenylate cyclase, include $5HT_4$, $5HT_{5A}$, $5HT_{5B}$, $5HT_6$ and $5HT_7$. The role of these receptors and their pharmacological characterisation is still under investigation[38-40]. However, clozapine was found to have a high affinity for both $5HT_6$ and $5HT_7$, thus suggesting a potential role for these receptors in the future development of atypical antipsychotics.

1.1.1.5 Serotonergic Hypothesis

The observation of an LSD-induced psychosis in healthy subjects was the first indication of a potential relationship between serotonin function and schizophrenia. LSD-induced psychosis was later refined to be a potential model for some aspects of schizophrenia (paranoid delusions, hallucinations, positive symptoms), but not for other aspects (such as disorganisation and negative symptoms)[41]. The serotonergic hypothesis is also strengthened by the effectiveness of $5HT_2$ antagonists, such as ritanserin, against the negative symptoms of schizophrenia[42]. In addition a genetic association has been observed between a polymorphism at position 102 in the coding region of $5HT_{2A}$ and an increased risk of schizophrenia[43]. Therefore the serotonergic receptors, and in particular $5HT_{2A}$, still remain a prominent biological target

Serotonin / Dopamine interaction:

Evidence indicates that 5HT fibers innervate those DA pathways thought to play a role in both positive (mesolimbic) and negative (mesocortical) symptoms, as well as movement disorders (nigrostriatal). For example more recently developed antipsychotics have a greater potency at the 5HT receptors than the D_2 receptors, and this approximately 10 fold greater potency combined with the effects on the DA and 5HT pathways has been proposed as the basis for their reduced side-effects[44]. Ziprasidone and sertindole are good examples of this new type of atypical antipsychotic.

1.1.1.6 Other Hypotheses

There are many other hypotheses for the basis of schizophrenia, a few of which are discussed below.

Glutamate Hypothesis[45,46]

The glutamate hypothesis of schizophrenia has been developed based on the observation that psychotic symptoms induced by phencyclidine (PCP) and related agents, which are antagonists of the NMDA glutamate receptor, closely resemble both the positive and

negative symptoms of schizophrenia. Many studies have reported a dysfunction of the glutamatergic system in schizophrenia, which is hypothesised to alter dopaminergic neurotransmission as follows:

- hyperdopaminergic activity may be caused by hypofunctioning NMDA receptors which are present on DA-containing nerve terminals where they act as heteroreceptors mediating an inhibitory regulation of DA-release.

- dysfunction of the glutamatergic neurons projecting to the striatum and limbic regions may cause the dopaminergic hyperactivity seen in schizophrenic brains, see Figure 1.2.

Acetylcholine Hypothesis

Early studies by several investigators attributed the psychotic symptoms of schizophrenia to dopaminergic-cholinergic imbalances[47,48]. This is based on the following observations: infusion of physostigmine into normal patients leads to behaviour similar to the negative symptoms of schizophrenia[47]. Anticholinergic agents have occasionally been reported to be effective in treating negative symptoms of schizophrenia, though it is not clear if this was merely due to a reduction in the side-effects of antipsychotics[48]. Antipsychotic drugs with high anticholinergic activity (e.g., clozapine, fluperlapine, and zotepine) have reduced side-effect profiles and also reduce the magnitude of negative symptoms.

Norepinephrine (Noradrenaline) Hypothesis

Upon discontinuation of antipsychotic therapy, the rate of relapse correlates well with the activation of the noradrenergic system, and a worsening of psychotic symptoms[49]. Studies with various noradrenergic agonists and antagonists reveal several findings of interest: β-adrenoceptor antagonists such as propranolol are ineffective alone but are clearly beneficial for akathisia[50], α_1-adrenoceptor antagonists are generally devoid of antipsychotic activity as demonstrated by negative studies using prazosin, and α_2-adrenoceptor agonists such as clonidine may have moderate therapeutic effects[51] while α_2-adrenoceptor antagonists such as yohimbine generally worsen psychotic

symptoms in acutely ill schizophrenics. Overall, a unifying hypothesis for the involvement of adrenoceptors in schizophrenia remains elusive.

γ-Aminobutyric Acid (GABA) Hypothesis

Interestingly, activation of $GABA_A$ receptors by isoguvacine or muscimol can produce psychosis in humans. The GABA hypothesis of schizophrenia postulates that decreased GABA inhibition of DA neurotransmission is responsible for schizophrenic symptoms[52]. This is seen in Figure 1.2 where decreased GABAergic inhibition will lead to increased glutamatergic activity and hence increased dopaminergic activity. However, clinical trials of GABAergic medications have been disappointing with a general lack of positive results and some reports indicating a worsening of psychotic symptoms[53].

1.1.2 Antipsychotic Drugs

Antipsychotic drugs may be defined as medications that alleviate delusions, hallucinations and some aspects of thought disorder that occur in a variety of illnesses, most notably schizophrenia. The mechanism of action of these drugs has focused on their antagonistic interaction with the central nervous system (CNS) neurotransmitter dopamine[13]; however recent work strongly implicates the differing dopamine subtypes and various other neurotransmitters as further targets of action.

In animals, antipsychotics were found to reduce or inhibit spontaneous movement, exploratory behaviour, operant conditioning behaviour, and conditioned avoidance behaviour. Also, amphetamine-induced hyperactivity, apomorphine induced aggression, and dopamine agonist-induced stereotypies were all blocked by antipsychotic medications. As with patients, antipsychotic medications were observed to cause indifference in animals and to induce cataleptic immobility while not affecting spinal reflexes. Each one of these effects along with many others were then investigated in an attempt to identify which behavioural patterns would be most specific with respect to predicting clinical efficacy[54].

The results of tests studying the ability of antipsychotics to block conditioned avoidance in both rats and monkeys demonstrated a high correlation between the potencies of the medication and antipsychotics in man[55-57]. Additionally, the ability of antipsychotic drugs to reduce amphetamine induced stereotypies, hyperactivity and aggression in animal models was also found to be specific for predicting antipsychotic efficacy in humans[54]. Thus, the expectation developed that new antipsychotic agents should block conditioned avoidance, and reduce or inhibit amphetamine-induced stereotypies, hyperactivity, or aggression in animals. Another interesting note is that originally catalepsy was also used as an indication of clinical efficacy, and drugs which exhibited low incidence of catalepsy were labelled as weak APD's, as is the case for melperone. However, with the advent of clozapine in the 1970's and its favourable side-effect profile, a further classification of atypical and typical APDs emerged.

Typical APD by definition are those that are most likely to induce so-called extrapyramidal side effects (EPS); such APDs are termed neuroleptics. By contrast, one of the major properties of atypical APDs is that they don't produce EPS. Initially atypical antipsychotics were classified as a group of APD's that would produce little or no catalepsy in doses that are effective in animal models at predicting antipsychotic activity. This later changed to their reduced liability to produce EPS at clinically effective doses. However as the scope of reduced EPS covers many areas theoretically emanating from many different receptor interactions, it is possible for APD's with very different receptor profiles to be assigned as atypical. Many people believe this is flawed because it is believed possible to find some common biological features underlying the effects of so-called atypical and typical APDs[58].

Atypical antipsychotics may be further classified into type A or type B[59]. Type A atypical drugs produce neither EPS nor elevations in prolactin levels; many of these compounds have been found to be potent $5HT_2$ receptor antagonists as well as D_2 antagonists. Type B atypical drugs produce mild EPS (at doses above those required for efficacy) and transient increases in serum prolactin. Clozapine and fluperlapine are examples of type A atypicals, while remoxipride and amisulpride may be classified as type B atypical antipsychotics.

Thus the term 'atypical' has evolved to now mean those APDs that exhibit the following clinical properties: a lessening of the positive and negative symptoms of schizophrenia; alleviation of neurocognitive deficits; devoid of extra pyramidal side-effects (discussed next); cause little or no tardive dyskinesia; and do not produce a sustained elevation of prolactin levels.

1.1.2.1 Extra Pyramidal Side-effects (EPS)

EPS are one major class of side effects of antipsychotic drugs and their metabolites and come in many forms, due to interactions with a range of receptors in the brain and other parts of the body. EPS include:

- Parkinsonism is the general term for Parkinson's disease-like symptoms which include general slowness of movement, rigidity, unsteadiness and tremor. Reduction of the dose and use of anticholinergics reverse the symptoms.

- Dystonia consists of irregular spasms of facial, neck, and trunk muscles and sustained abnormal postures. These symptoms can occur early in the treatment and they respond to anticholinergics.

- Akathisia is experienced by the patient as an irresistible urge to move within minutes from any sitting, lying, or standing position. Akathisia is subjectively very distressing and responds satisfactorily to dose reduction and propranolol.

- Tardive Dyskinesia consists of oro-facio-lingual daytime movements, which may be associated with larger movements of trunk and extremities. These can be induced by long-term treatment with antipsychotics, but dyskinesia is also a feature of chronic psychosis by itself. Dose adjustment can prevent or slowly reverse the dyskinesia.

- Neuroleptic malignant syndrome is a rare and potentially fatal adverse affect with symptoms of hyperthermia, rigidity, tachycardia, hypertension, stupor, and leukocytosis. Antipsychotic treatment should be stopped and maximal supportive care instituted.

1.1.2.2 Other Side-effects

Many other side-effects have been reported to occur to varying degrees for different APDs and these side-effects undoubtably result from the ability of these drugs to bind to a whole range of receptors. Examples of these many side-effects are weight gain, sexual dysfunction, hypersalivation, insomnia and tachycardia, to mention a few. Two of the more serious side-effects, which have led to drugs being withdrawn from the market, are prolongation of the QT interval, causing potentially fatal ventricular arrhythmia, and agranulocytosis. Agranulocytosis is a drug induced decrease in white blood cell count, stringent weekly monitoring of which is necessary to avoid the possibility of lethal infection.

1.1.2.3 Chemical Classification of Antipsychotic Drugs

Antipsychotic drugs have long been classified according to their chemical structure. This was originally devised to determine if certain structural features were related to particular patterns of clinical efficacy or side effects[60].

Phenothiazines

Phenothiazines were the first major class of antipsychotics to be developed, with Chlorpromazine (see Figure 1.4) introduced in France in 1951, still widely in use. There are three main groups of this class, (1) aliphatics, (2) piperidines and (3) piperazines, based on differences in the composition of the side chain linked to the nitrogen atom in the phenothiazine tricyclic nucleus. For clinical activity, a rule of thumb suggests there must be three carbon atoms between the nitrogen in the central ring and the tertiary amine, and that substitution of an electron withdrawing substituent at R_1 increases the potency. Thioxanthenes are closely related to the phenothiazines, differing by the substitution of the nitrogen by carbon in the central ring, resulting in slightly less potency.
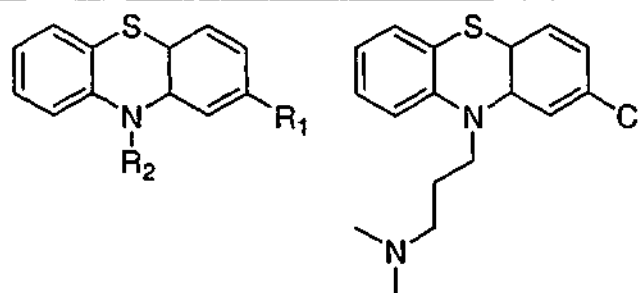
Figure 1.4. Phenothiazine substructure and chlorpromazine.

Butyrophenones

The general structure of a butyrophenone includes a tertiary amine containing at least one aromatic ring linked to the amine nitrogen by a keto group attached to an intermediate chain of three carbons. Substitution of an electron withdrawing substituent at R_1 increases the potency, and reduction of the keto group to an alcohol decreases potency. Since its introduction into the United States in 1967, haloperidol (see Figure 1.5) has remained one of the most widely used front line therapies for schizophrenia. It is classed as a typical APD since long term use results in EPS. Diphenylbutylpiperidines (for example fluspirilene) are closely related to butyrophenones except that the keto group is replaced by another substituted phenyl ring.
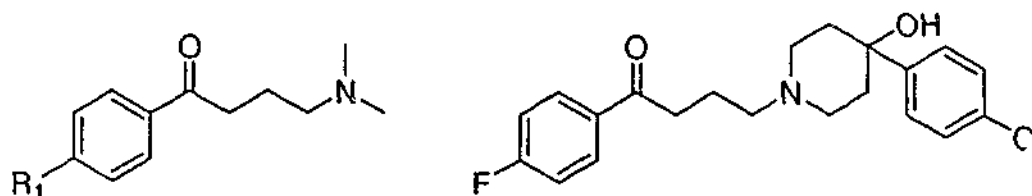
Figure 1.5. Butyrophenone substructure and haloperidol.

Benzamides

Benzamides (Figure 1.6) are also closely related to butyrophenones with an aromatic ring, in most cases extensively substituted, linked to the tertiary amine nitrogen through an amide group attached to a chain of two carbons. Some of this series of compounds are D_2 selective (remoxipride), and cause little EPS at clinically effective doses[61].
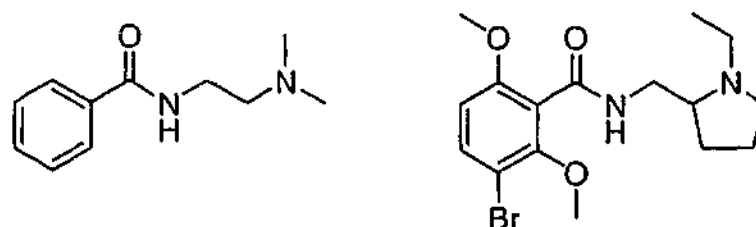
Figure 1.6. Benzamide substructure and remoxipride.

Dibenzepines

The dibenzepines have the general structure shown in Figure 1.7 in which R_1 is carbon, oxygen, sulfur or nitrogen, and invariably the side chain is a piperazine with a tertiary distal nitrogen. Examples of these compounds are clozapine, $R_1 = N$; loxapine, $R_1 = O$ and clothiapine, $R_1 = S$. Substitution of an electron withdrawing substituent at R_2 increases the potency of the antipsychotic.
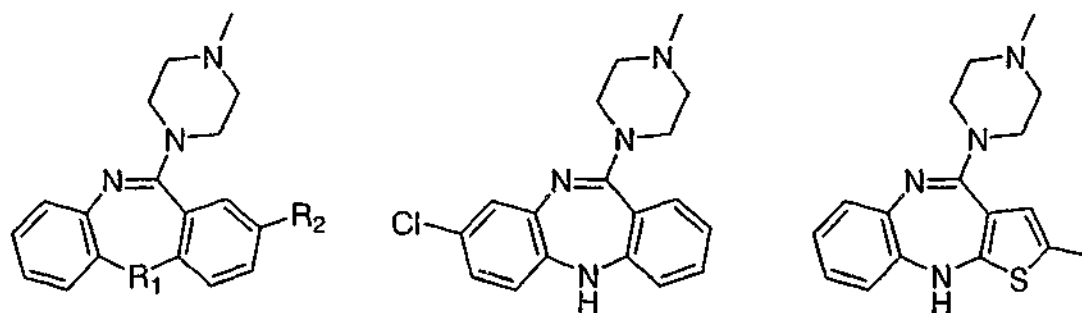
Figure 1.7. Dibenzepine substructure, clozapine (centre) and olanzapine (right).

The major member of this class, the dibenzodiazepine clozapine (see Figure 1.7), was first introduced in the early 1970's but was subsequently withdrawn due to a high incidence of agranulocytosis. It has recently been reintroduced as techniques for monitoring white blood cell count have become possible. It is the most effective APD for treatment-resistant or refractory schizophrenics[62], that is patients who do not satisfactorily respond to first line treatment with typical APDs such as haloperidol.

Extensions of this class include the thienobenzodiazepines, like olanzapine (see Figure 1.7), which is a type B atypical as it induces dose-dependent EPS.

Benzisoxazole

These compounds are a recent type of antipsychotic to be brought onto the market with some success. They are similar to the piperidine subclass of phenothiazines in some respects, with an aromatic ring linked to the tertiary nitrogen incorporated into a piperidine ring (see Figure 1.8). Substitution of an electron withdrawing substituent at R_1 increases the potency of the antipsychotic. Although this series are proposed to be atypical APDs, studies of risperidone (Figure 1.8) have shown that EPS is linearly related to daily dose[63].

Replacement of the oxygen in the indole ring with sulphur results in compounds called benzisothiazoles, e.g. ziprasidone.
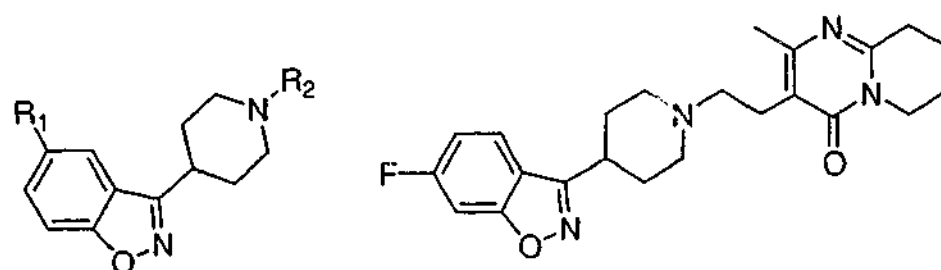
Figure 1.8. Benzisoxazole substructure and risperidone

Miscellaneous

Many indole derivatives have been synthesised and tested for antipsychotic activity, the reduced indole molindone (see Figure 1.9) being one of the early more successful compounds from this series. A number of newer types of antipsychotics are emerging with differing substructural untis, the most significant of these being the APD aripiprazole. Aripiprazole (Figure 1.9) has been shown to have high postsynaptic dopaminergic antagonist activity and high dopaminergic autoreceptor agonist activity[64] and has just been given FDA approval in the United States for the treatment of schizophrenia.
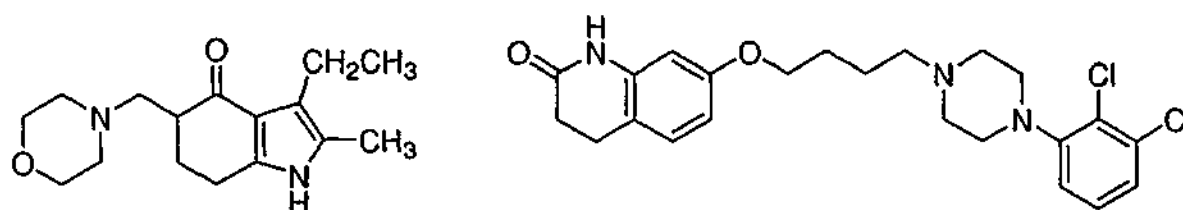
Figure 1.9. Molindone (left) and Aripiprazole (right).

1.1.2.4 Structure-Activity Relationships (SARs) of Antipsychotic Drugs

With many of the early series of APDs, which were analogues of intensively investigated parent molecules, SARs were readily devised from the various potency measures within the series. However, with the multitude of novel substructural units as the basis for new APDs, together with binding data at receptor subtypes, SAR analysis has become more difficult. There have been many attempts to rationalise APD SAR[65], with animal models being used to predict clinical effects of APDs as far back as 1966[66-68]. In 1974 Gschwend[69] published an excellent review on the development of tricyclic, butyrophenone, diarylbutylamine, and reserpine-like neuroleptic drugs. There have also been many excellent SARs on the tricyclic substructural compounds[70-72], in one of which Schmutz[73] derives correlations between the Hansch constant π and the relative partition coefficients; and the Hammett parameter (σ_m) with both pK_a, and apomorphine antagonism. Rognan *et al.*[74] used optically active benzamides to map the binding site of the D_2 receptor and determine key features for this series of compounds. One of the most recent SAR on antipsychotics and related compounds is the excellent paper by Boström *et al*[75] which incorporates the tricyclics, benzamides and a number of other different substructural units into a D_4 pharmacophore. Surprisingly this D_4 pharmacophore established by Boström *et al*[75] has many similarities to the general antipsychotics model published by Gschwend[69], which serves to highlight how modern computational techniques are now making pharmacophore mapping possible where only rudimentary SAR could once be derived. This effect of technology is not only seen in pharmacophore mapping but is also evident in all aspects of drug design from crystallography to biotechnology and beyond.

1.2 DRUG DESIGN

The process of drug design and development has shown a paradigm shift since the days when sole reliance was placed on bench chemists making as many analogues as possible for both developing and testing a theoretical model. This change has mainly been driven by the need to ensure that as many lead compounds are generated in the shortest possible time. Economic rationalisation in combination with advances in computational technology have seen the field of computational chemistry become an essential part of the drug design and discovery process. There are a number of ways to go about the drug design and discovery process, with the most common methods being ligand-based and structure-based design. Adjunct to these compound design methods are ADME/T (adsorption, distribution, metabolism, excretion and toxicity) prediction and database mining.

1.2.1 Ligand-Based Design

In traditional ligand-based design, a pharmacophore is constructed from a set of ligands, the ligands are then superimposed and some type of statistical analysis is performed to obtain a predictive model. The first step in construction of the pharmacophore model is selection of a set of high affinity ligands with structural diversity and low flexibility. Next, functional groups considered essential for biological activity (pharmacophoric elements) are identified. The most important step is identification of the putative receptor-bound conformation, usually done through a combination of manual and automated conformational analysis. Once this is accomplished the ligands are superimposed so that pharmacophoric elements overlap and a common template of key binding groups, a pharmacophore model, is identified.

The original ligand-based design method for proposing a pharmacophore, the active analogue approach[76] (AAA) was developed by Marshall. The global minimum energy conformation of the tightest binding ligand was the proposed receptor-bound conformation and used as a template for subsequent superimpositions and volume comparisons. The pharmacophoric elements are the atoms of interest, for example a benzene ring (hydrophobic region) and oxygen or nitrogen atoms (H-bond donors or

acceptors). The AAA for compound alignment was later modified to alignment of hypothetical points of interaction[77]. Here pharmacophoric elements are not only atoms but also points of likely interaction with a hypothetical receptor. For example commonly used distances used to model points of interaction are 2.8 angstrom along the vector of the proposed H-bond donor or acceptor, and 3.5 angstrom above and below the plane of the aromatic rings (indicating a π-π stacking arrangement). These traditional methods of alignment are rapidly being replaced by more novel methods of automated alignment. One of these methods is the *field fit*[78] method implemented in the Sybyl Tripos[79] software, in which compounds are fitted into the steric and electrostatic fields previously calculated from the tightest binding ligand or ligands. This method has been used in a number of papers[80-83] with noticeable improvements in the predictive abilities of these models. All these methods use manual conformational analysis and only differ in the methods of compound alignment. Many other methods for compound alignment that have been devised by several groups are covered in an excellent review by Lemmen and Lengauer[84].

There are a number of methods of automated conformational analysis. One of these is the alignment tool available within the Sybyl software suite, called GASP[85,86] (genetic algorithm similarity program). GASP uses a genetic algorithm to explore conformational space and discover the correspondence between functional groups in different molecules and the alignment of these groups in a common geometry. It is able to automatically identify rotatable bonds and pharmacophore features such as rings and potential hydrogen-bonding sites. Results have shown that GASP is able to predict pharmacophores similar to those derived by conventional conformational analysis[87]. Another popular automated conformational analysis method is Catalyst, contained within the InsightII suite[88]. Catalyst develops 3D pharmacophore models from a collection of molecules possessing a range of diversity in both structures and activities. These pharmacophores can be used to retrieve structures from 3D databases fitting the hypotheses, or as models to predict the activities of novel compounds. Apex-3D, available as an add-on to InsightII[88] also uses an automated conformational analysis method for identification of biophores. These biophores, which represent the set of 3D complementary binding sites present in the macromolecule, can be used for building qualitative activity prediction rules and for creating search queries to identify new leads in a 3D-database. Identified biophores can be used as starting points for constructing

3D-QSAR models when good quantitative data are available. Another alignment tool that explores conformational space is SEAL[89,90] (steric and electrostatic alignment), which optimises the overlap of molecular volumes as well as matching groups, assigned on criteria such as hydrophobic interaction or hydrogen bond acceptor/donor capacity. A very recent publication by Jewell et al.[91] uses a modified version of a field-based similarity searching (FBSS) method, originally developed for 3D searching in chemical structure databases, to align compounds. The FBSS method calculates molecular field values in a 3D grid surrounding compounds, which are then fed through a genetic algorithm (GA) to maximise their similarity index and thus align the compounds. The method allows for conformational flexibility by encoding torsional rotations[92] although this may result in 'highly strained structures'. The authors have investigated inclusion of an energy calculation in the GA's fitness function to account for strained structures, with no great benefit, and are currently investigating other means of sampling conformational space such as using multiple conformers. Pitman et al.[93] also used the field-based similarity search method. The FBSS method is implemented in a program called FLASHFLOOD that is able to deal with the alignment of flexible compounds by representing them in terms of fragments and torsional angles. An assembly algorithm is used to reconstruct the compound from fragments. However one major difference from the FBSS method presented by Jewell et al.[91] is that FLASHFLOOD is not confined to a particular field, but enables derivation of a continuum of simple descriptive through to quantum mechanical fields for use in searches. The group of Labute and Williams[94] have also devised a method that searches all conformational space for the compounds in question. They use a Monte Carlo related method to assign random dihedral angles to all rotatable bonds which are then minimised and scored, hence making it suitable only for small molecules.

Other novel methods of non-manual ligand alignment that have proved successful for particular situations involve using the ligand alignment obtained from docking simulations as the basis for a comparative field analysis. Sippl[95] has shown that improved predictive ability of their model was obtained using a docking alignment rather than traditional ligand-based alignment methods. Predetermined alignment[96,97] is also on the increase, due to the emergence of more and more crystal structures of proteins with ligands bound in the active site. These give us a much better idea of the

binding conformation of a drug, though of course there may be more than one way for a ligand to bind into a particular active site.

Once an alignment has been generated there are many ways in which subsequent quantitative structure-activity relationships (QSAR) can be derived. Originally QSAR was based on using standard statistical analyses[98] in conjunction with multiple linear regression (MLR) using physicochemical descriptors[99]. Later developments in this field applied advanced statistical methods of analysis, with the partial least squares (PLS)[100,101] method gaining recognition in the early eighties.

In the late eighties one of the most significant methods developed for 3D-QSAR was introduced, comparative molecular field analysis (CoMFA)[102]. Here compounds are placed into a 3D grid box in which electrostatic and steric fields are calculated at each point of the grid resulting from different interactions between a variety of probes and the compounds under study. The fields are tabulated and the result data are analysed using principle component analysis (PCA) and PLS to give the final predictive model. Although, in principle, CoMFA sounds simple enough to implement there are a number of pitfalls that could lead to an incorrect model and conclusions, which may be avoided by heeding the sound advice given in references[103-105]. CoMFA fields have also been supplemented by additional fields, such as the hydrophobic interactions (HINT)[106] field, which can give further insights into receptor interactions. Another later addition to the PLS 3D-QSAR field was comparative molecular similarity indices analysis (CoMSIA)[107]. The main differences between CoMFA and CoMSIA are that CoMSIA maps highlight those regions within the grid box occupied by ligand skeleta requiring a particular physicochemical property important for activity. In addition to this, CoMSIA has a number of additional fields: hydrogen bond acceptor, hydrogen bond donor and hydrophobic fields. The GRID[108] program developed in the mid eighties by Goodford also uses probes to find information about interaction capabilities. In GRID there are a large number of probes available for the user to investigate a structure. For example a protein that interacts with DNA could be analysed with a phosphate probe to determine possible sites where DNA could bind. In the early stages CoMFA-like studies could not be performed from GRID data as no simple PLS algorithm for generating predictive models had been developed. This was overcome with the introduction of generating optimal linear PLS estimations (GOLPE)[109]. GOLPE was originally developed in the early nineties to enhance the quality of PLS models. GOLPE achieved this by using

fractional factorial design (FFD) to select combinations of variables for which a PLS model was derived. Only variables that significantly affected the predictive ability of a model were kept. GOLPE was designed to work with CoMFA or GRID-based fields and generate predictive models. Many papers have used GOLPE[95,110,111], which has been further improved using an algorithm called smart region definition (SRD)[111]. SRD groups descriptors into regions of neighbouring variables bearing the same chemical and statistical information, so that the predictive ability of the model is increased by reducing multicolinearity in 3D-QSAR.

One of the more recent novel 3D-QSAR techniques is GRID-independent descriptors (GRIND)[112]. GRIND generates novel molecular descriptors from interaction fields computed from programs such as GRID which are then reduced and encoded into variables. The variables can be displayed as "correlograms", and used in chemometric analysis. GRIND provides a fast and simple way of obtaining structure-activity relationships without the need for compound alignment. A limitation is that the result is still determined by the conformation of the structure. Another alignment independent method is the hologram quantitative structure-activity relationships (HQSAR)[113] method developed by Tripos. Here molecules are described in 3D encoded fragments (fingerprints) that contain explicit stereochemical information. These descriptors are then used to build quantitative models that relate a biological activity or property to chemical structure. The model generated shows which fragments make positive or negative contributions to the model by appropriate colour coding. This has the benefit that it can be applied to very large datasets of compounds, but the resultant models are not as informative as those produced by CoMFA or CoMSIA. Alternatively HQSAR can be used for other purposes such as testing combinatorial libraries for suitability by performing virtual high throughput screening (vHTS).

Recently pharmacophore fingerprints have been developed and used in the construction of molecular descriptors for similarity and diversity applications such as virtual screening, library design[114] and QSAR[115]. The reader is referred to the comprehensive review by Mason *et al.*[116] for further insights.

Apart from the traditional linear methods of QSAR analysis, various non-linear methods have emerged. With the advent of artificial intelligence methods, including neural

network (non-linear) techniques, applied to statistical analysis it is possible to predict the activity of new CNS compounds in cases where a simple linear relationship does not exist. Some of these non-linear methods are forward-feed back-propagation, Bayesian, Kohonen self organising maps (SOM) and Hopfield recurrent neural networks (NN), with the most popular being the simple forward-feed NN.

Artificial neural networks (ANNs) were first developed in the 1940's but did not get much attention until the 1980's when Hopfield addressed some of the initial problems associated with their use[117]. Even then, when forward-feed back-propagation neural networks (BPNN) were first used for quantitative structure activity relationship (QSAR) purposes by Aoyama et al.[118], many problems were still evident. The main problems associated with ANN use were network architecture, problems of chance effects[119] and overfitting and overtraining. Various groups have devised clever methods for the elimination of these problems which involve techniques from clustering to genetic algorithms[120,121], to separating data into training, validation and test sets[122]. These techniques and further extensions are well covered in greater detail with appropriate references in the review by Manallack and Livingstone[123].

In the last few years since that review we have seen a number of new developments dealing with these problems. The group of Burden et al.[124] have coupled automatic relevance determination (ARD) of network weights with their Bayesian regularised artificial NN (BRANN)[125]. ARD, first developed in the mid 90's[126,127], effectively turns off irrelevant variables by reducing their weights thus allowing all variables to be included in the analysis without ill effect. This method has been applied to compounds active at the benzodiazepine and muscarinic receptors as well as to analysis of some toxicological data[124]. Tetko et al.[128] have developed a new QSAR method utilising ANNs and CoMFA, called the volume learning algorithm (VLA). The VLA method uses a Kohonen SOM to define cluster zones in space around the molecule, then the mean values of these clusters are used to train a BPNN. This has the benefit of reducing the number of input parameters whilst preserving the spatial and structural information of the molecule. There are a number of advantages of this approach in that non-linear relationships can be established and the predictive abilities of these models are higher as there are fewer problems associated with multicolinearity of input variables. The

method has been successfully applied to a series of cannabinoid CB_1 receptor agonists[128].

1.2.2 Structure-Based Design

The second main method of drug design and discovery mentioned earlier is structure-based design. In structure based design the proposed receptor binding site (biophore) is used for docking experiments or for building up compounds in a *de novo* approach. The proposed receptor binding site can be obtained from nuclear magnetic resonance (NMR), or X-ray analysis of the protein, or built using homology modelling techniques based on crystal structures of structurally related proteins.

Both X-ray diffraction and NMR spectroscopy are well established methods of protein structure prediction. There are currently 18,282 and 3,290 structures solved using X-ray diffraction and NMR spectroscopy, respectively, deposited in the Protein Data Bank[129] (as of 1st July 2003). Although X-ray diffraction and NMR spectroscopy have made much progress in the field of protein structure determination, large classes of proteins still cannot be investigated with these methods. Either the proteins cannot be crystallised to a sufficient crystal size for X-ray diffraction, or they cannot be brought into a sufficiently concentrated solution for NMR spectroscopy. Therefore models are constructed *in silico* based on known crystal structures of homologous proteins; this process is termed homology modelling. This was first done on an insulin-like growth factor in 1978[130]. Nowadays there are a number of programs capable of automatically producing homology models of proteins. Once a protein with a proposed receptor binding site is obtained it can be used for *de novo* building of ligands or as the basis for docking experiments.

There are two main approaches to *de novo* drug design, the "inside out" method and the "outside in" method. The "inside out" method[131] works by placing a fragment inside the receptor cavity and then the ligand is built up from this fragment to interact with the maximum number of possible interaction sites within the receptor. The "outside in" method[131] works in the opposite manner, where interacting fragments are placed onto the interaction sites within the receptor and the ligand is constructed by joining these

fragments together inside the receptor cavity. The CONCERTS[132] program by Pearlman and Murcko and the LUDI[133] program by Bohm are good examples of this type of *de novo* design.

Probably the most popular structure based design method is that of docking. Docking works by taking the ligand and placing or building it inside the receptor cavity and then applying some scoring function to asses the quality of the fit. There are a number of programs that have been constructed to perform such tasks: DOCK[134], FlexX[135] and Autodock[136] are but a few. These programs can give variable results, with different programs and scoring functions being best suited to differing situations; the reader is referred to the comprehensive reviews by Bissantz *et al*[137] and Stahl *et al*[138] for details.

Bissantz *et al*[139] have also published an excellent paper on the limitations and methods for docking compounds into GPCRs built using homology modelling techniques. This paper discusses the limitations of docking agonists into GPCR homology models derived from the inactive state of rhodopsin, which is thought to be closer to an "antagonist bound" state. Results from the docking of antagonists into Bissantz's GPCR models indicated that the inactive state of rhodopsin is a suitable basis for these models.

1.2.2.1 GPCRs

GPCRs constitute the largest and most diverse group of transmembrane proteins involved in signal transduction. They are activated by a wide range of extracellular ligands including small biogenic amines, large protein hormones, neuropeptides and chemokines. GPCRs are also fundamental receptors for the sensory perception of light, taste and smell. They all share a common characteristic, the seven distinct hydrophobic regions, each of which is 20 to 30 amino acids in length within their sequence[140]. These hydrophobic regions form the seven transmembrane domains of these receptors, hence the alternative name for this family, 7TM receptors.

GPCRs are a primary agent by which cells sense and respond to their external environment. Information from an extracellularly occurring receptor-ligand recognition

event is transferred through conformational rearrangements within the receptor protein to the intracellular compartment. The ligand induced activation of the receptor, leads to an association of an intracellular G-protein that in turn triggers the catalysis of GDP–GTP exchange on the G-protein. This leads to a multistep cascade of signal transduction events. This multistep cascade involving intracellular effectors and subsequent generation of second messengers is specific to the ligand type and the distinct member of the GPCR family involved in the process. Typically G-proteins either stimulate or inhibit the production of second messengers. All of this eventually leads to the physiological response of the cell to the stimulus. The enormous diversity of receptors, G-proteins and effectors, together with the widespread distribution of receptors across many tissues, reflects the important role that this superfamily plays in regulating physiological and pathophysiological processes.

1.2.3 Database Mining

Database Mining is a technique that is usually used in combination with ligand-based and structure-based design. Database mining can be performed in a variety of ways. Pharmacophores[87,141], shape filters[142] and/or physicochemical characteristics[143] are but a few ways to reduce the number of compounds being analysed or to aid in the identification of new substructures for drug design. Typically in most pharmaceutical docking analyses the compounds to be docked are filtered through Lipinski's 'rule of five'[143] to remove compounds with poor drug-like characteristics.

1.2.4 ADME/T Prediction

In silico prediction of a compound's physicochemical characteristics plays an important role in the drug discovery process, as this enables compounds with poor ADME/T characteristics to be removed, saving valuable time and money. Poor ADME/T characteristics are one of the major reason for the high attrition rates of compounds in development, where more than 90% of all compounds fail[144].

There are a number of ways of going about *in silico* ADME/T prediction, with probably the simplest way being to derive a quantitative structure property relationship (QSPR)

between the property of interest and a set of descriptors. Semi-empirical[145-148] or *ab initio*[149-152] descriptors are typically used, but fragment based approaches[153] and a variety of other approaches[154] are possible. As with QSAR analyses, linear regression[147,148], non-linear regression[150,154], genetic algorithms and classification methods can be used to derive the relationship.

Some of the typical properties that are calculated are solubility, pK_a, logP, blood brain barrier (BBB) penetration and Caco-2 cell permeability. These calculated properties are usually indicative of various aspects of a compound including:

- calculated Caco-2 cell permeability is useful as an estimation of oral intestinal drug absorption[155]

- predicted BBB penetration indicates likely CNS activity

- calculated relative solubility suggests the type of formulation necessary for optimum adsorption and distribution

- calculated pK_a of a compound will determine its percentage protonation at a particular pH which may in turn reflect its ability to diffuse across a membrane

There is an extensive suite of off-the-shelf ADME prediction tools with pre-derived QSPR relationships in various modules from Tripos, Accelrys and Schrodinger, to name a few. There are also toxicological and metabolite databases[156,157] able to predict adverse effects and metabolites of a compound. The reader is referred to the recent review by Butina *et al*[158] for additional information on *in silico* ADME prediction.

1.3 THE SCOPE OF THIS THESIS

The primary aim of this work has been to develop D_4 selective antipsychotics, using a number of differing complementary computational techniques. The background and methodologies are fully described in the following chapters of this thesis, with a brief introductory summary of the studies performed given below.

1.3.1 Ligand-Based Design (Chapter 2)

In this chapter a tricyclic pharmacophore was developed from conformational analysis of high affinity tricyclic antipsychotics and D_2 and D_4 active compounds. This pharmacophore was used as the initial basis for the construction of CoMFA models. Differing charges, charge states, alignments, and alignments within the CoMFA grid were then considered in the generation of optimum models. Analysis of the tricyclic CoMFA model aided in the generation of ideas for the construction of D_4 selective tricyclic ligands. An extended pharmacophore was then developed from conformational analysis of high-affinity extended antipsychotics and D_2 and D_4 active compounds. This extended pharmacophore was then compared to a pharmacophore generated automatically from GASP[85,86]. The two pharmacophores, tricyclic and extended, were then combined and refined using a published model from Boström et al.[75]. This combined pharmacophore model was then used as the basis for further CoMFA models. The optimum alignment method, charges and charged state, as determined earlier, were used in the generation of the final CoMFA models. Analysis of these CoMFA models aided in the generation of ideas for differing D_4 selective compounds.

1.3.2 Structure-Based Design (Chapter 3)

Part A of chapter 3 involved the generation of homology models of the dopamine D_2 receptor using Modeller[159,160]. Initially threader database files of bacteriorhodopsin[161] and rhodopsin[162] were created to perform a threaded alignment of the D_2 sequence onto these sequences. The aligned sequences were then checked against previous alignments[163] and against highly conserved residues within all GPCRs[164]. Homology models of dopamine D_2 were then built with and without dopamine bound in the active

site. The bound and unbound bacteriorhodopsin-based models were then compared against each other and against the published D_2 model based on bacteriorhodopsin, 1115[163]. Rhodopsin-based D_2 models were compared to one another and to the bacteriorhodopsin-based D_2 models. All models were analysed with respect to proline kinks, solvent accessible residues, proposed sodium binding sites, proposed ligand binding sites and overall stereochemical quality.

Part B of chapter 3 involved building models of the D_2 and D_4 receptors based on rhodopsin with clozapine and spiperone bound in the proposed active site. Before these models were built, mutagenesis experiments[165-177] were examined to assess the likely binding orientation of the ligands within the receptors. Models were then built of the D_2 and D_4 receptors using these proposed binding orientations and conformations deduced from the pharmacophores of chapter 2. All constructed models were analysed and grouped into families and assessed against mutagenesis data. The best resultant models were then used as templates for the superimposition of the CoMFA models generated in chapter 2. The additional information from these analyses aided in the generation of ideas for D_4 selective compounds.

1.3.3 pK$_a$ Prediction (Chapter 4)

Compounds with experimentally determined pK$_a$ values were extracted from the PHYSPROP© database[178] and filtered so as to remove compounds that were not drug-like. Semi-empirical frontier electron theory descriptors[179] were then calculated for all of the compounds. Linear regression models were then constructed between the descriptors and their experimentally determined pK$_a$ values. The benefits and limitations of this technique were then discussed in relation to a number of other papers[145,180,181] and techniques[182] on predictive pK$_a$ methods. The predictive models derived were used to predict the pK$_a$'s of compounds proposed in chapter 5.

1.3.4 Compound Design (Chapter 5)

Using the information gained from chapters 2 and 3 in combination with database searching using the pharmacophores derived from chapter 2, a number of series of

compounds were proposed. The CoMFA models derived in chapter 2 were used to predict the affinities of proposed compounds, and the pK_a values were predicted using the linear regression models developed in chapter 4.

Chapter 2

Ligand-Based Design

2.1 TRICYCLIC MODEL INTRODUCTION

Many structural classes of antipsychotics have been developed and prominent among these classes is the dibenzepine type, in which the central seven membered ring is substituted with oxygen, nitrogen, sulphur or carbon[1]. A significant number of these dibenzepines also contain a piperidine or piperazine ring attached to the seven membered ring. One recently re-introduced atypical antipsychotic drug of the dibenzepine type, clozapine (Figure 2.1), has shown the most promise, particularly in treatment-resistant schizophrenia[2], due to its low incidence of EPS[3]. The exact basis for the activity of clozapine is unknown, but the fact that it acts at a range of CNS receptors and their subtypes (H_1, $5HT_{2A}$, $5HT_{1C}$, α_1, M_1, D_2, D_4,)[4], suggests a complex blend of interactions is required for its atypical antipsychotic drug efficacy. One consequence of clozapine's affinity to such a large variety of receptors is the production of several, clinically limiting side effects such as sedation (due to potent H_1 antagonism), hypersalivation (due to M_1 antagonism) and tachycardia[5]. It is thought that clozapine's atypical antipsychotic profile is due to its relatively low affinity for the D_2 receptors and high affinity for D_4 and $5HT_{2A}$ receptors. Meltzer et al.[6] proposed that atypicality arises when an APD's $5HT_{2A}$ activity is over ten fold greater than that of D_2. It has also been proposed that clozapine's preference for D_4 receptors and their location mainly in the mesolimbic regions[7] gives rise to its atypicality, as there is less blockade of the nigrostriatal pathway which is associated with movement disorders. In an effort to try and better understand these complex interactions, we have focused on the dopaminergic receptors most often implicated in the aetiology of this disease: the dopaminergic receptors D_2 and D_4. The serotonergic receptor $5HT_{2A}$ is also investigated in relation to the tricyclic dibenzepine compounds.

2.2 METHOD DEVELOPMENT

2.2.1 Tricyclic Pharmacophore

The pharmacophore model was based on (R)- and (S)-octoclothepin[8], as representative molecules of the dibenzepine type, and tefludazine[8] (Figure 2.1). These molecules have high potency at dopaminergic and serotonergic receptors and their differing stereoselective potencies have potential in establishing a binding pattern[9]. The octoclothepin isomers also provide a similar overall framework to clozapine, to which clozapine should presumably conform closely in its binding at the D_2, D_4 and $5HT_{2A}$ receptors. The structurally dissimilar tefludazine, whose potency resides in just one diastereomer (1R,3S), provides a flexible template for reducing the wide range of pharmacophore models produced by the octoclothepin conformers.
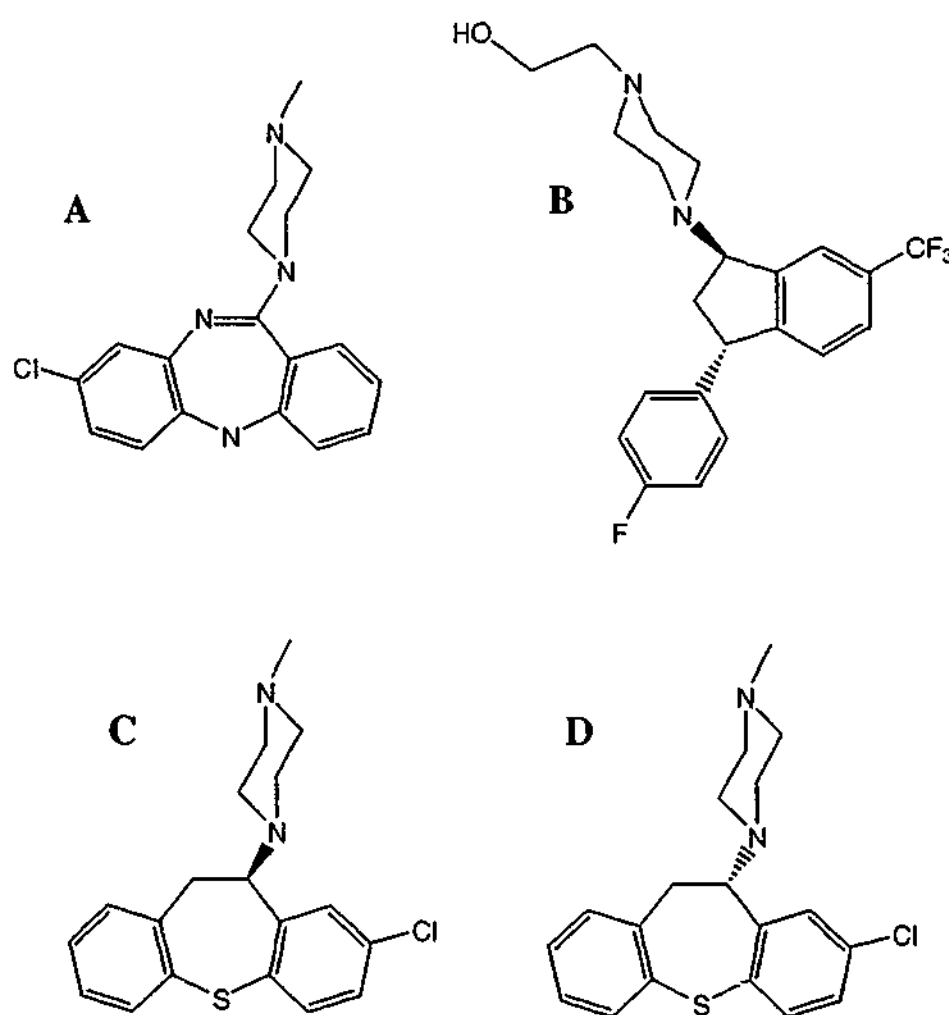
Figure 2.1. *A clozapine, B (1R,3S)-tefludazine, C (R)-octoclothepin, D (S)-octoclothepin.*

2.2.1.1 Conformational Analysis

Structures for (R)- and (S)-octoclothepin[10] enantiomers were obtained from the Cambridge Crystallographic Database[11] (CCDC). Conformational analysis of the central tricyclic ring was simulated with a seven-membered ring using the subroutine *Find Transition State* in the program *CAChe*[12] version 3.1. Simulated annealing for ten cycles to 1000 degrees using default parameters for base temperature and cooling rate, was undertaken with Sybyl[13] version 6.5 to reveal other possible conformations of the tricyclic structure.

Using ten different local minimum arrangements , a systematic search was performed using the Tripos force field[14] in Sybyl. Van der Waals radius scale factors were set at 0.65 with charges assigned by the Gasteiger and Marsili method[15]. The search was performed in 1° increments on all conformations and configurations of octoclothepin around the bond to the piperazine ring. Torsion angle versus energy plots were then produced showing the location of the local and global minima.

These local and global minima were further analysed using MacroModel[16] version 5 and Amsol[17] version 6.6. The MacroModel parameters used were the OPLS force field and water as solvent. The Amsol parameters used were the AM1 semi-empirical method with water as solvent, and heptane as solvent. This was undertaken to compare and contrast the conformational energies in various solutions.

2.2.1.2 Superimposition: Defining the Pharmacophore

To develop a pharmacophore for D_2 antagonist interaction, the different conformations and configurations generated from the previous analysis were compared by superimposition. Dummy atoms were used to represent points of likely interaction[18] at the D_2 receptor, as shown in Figure 2.2 for (S)-octoclothepin. To mimic π-π interactions, dummy atoms were built 3.5 Å above and below the planes of each aromatic ring[19]. An additional dummy atom was built 2.8 Å from the distal nitrogen of the piperazine ring, along the vector of the ammonium hydrogen, representing a proton donor/acceptor interaction[19]. The pK$_a$ of the distal nitrogen was calculated to be 8.4 ±

0.7, using ACD[20] version 3 pK$_a$ calculator, and octoclothepin isomers are therefore likely to consist of an equilibrium mixture containing approximately 90% protonated molecules at a physiological pH of 7.4. All conformations were then superimposed onto selected differing configurations, using the root mean square (RMS) fit procedure within Sybyl. A weighting of 1 was assigned to the dummy atom pairs from the distal nitrogen, and weightings of 0.3 were assigned to the dummy atom pairs above and below the planes of each aromatic ring. This was done so that the tricyclic ring of the enantiomers, with four dummy superimposition points, had approximately the same priority as the dummy atom emanating from the distal nitrogen.
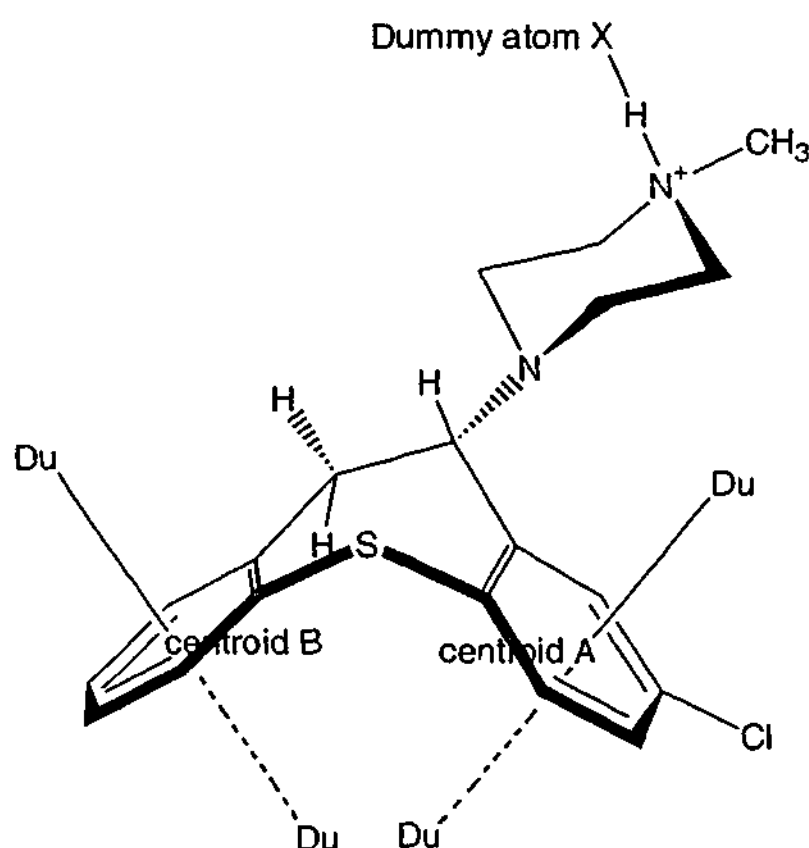
Figure 2.2. *(S)-octoclothepin showing dummy atoms. X and Du are 2.8 angstrom and 3.5 angstrom from the nitrogen atom and perpendicular to the benzene ring centroid, respectively.*

The active (1R,3S) tefludazine diastereomer[21] was built in Sybyl using standard geometries, then had dummy atoms assigned to its points of likely interaction. Superimpositioning using similar weightings of dummy atoms, of its low energy conformers upon the resulting octoclothepin pharmacophore models was undertaken. A search of the CCDC was performed to confirm that conformations proposed as

biologically active were also energetically feasible. From this a pharmacophore model was defined.

All different combinations and conformations of key functional groups were considered in the superimposition process, with the aim of developing a consensus pharmacophore. However it is recognised that less likely alternatives may exist for isolated compounds.

2.2.2 Tricyclic Molecular Field Analysis

In this section, comparative molecular field analysis (CoMFA) is used to develop a pseudoreceptor model for D_2, D_4 and $5HT_{2A}$ antagonist activity using published data from a series of clozapine analogues[22-24], in an attempt to elucidate some aspects of clozapine's favourable receptor binding profile.

CoMFA was carried out on the series of clozapine analogues given in Tables 2.1, 2.2 and 2.3 using their published receptor binding affinities[22-24].

Table 2.1. Receptor binding affinities (K_i, nM) of piperazine analogues from Phillips *et al.*[22,23]
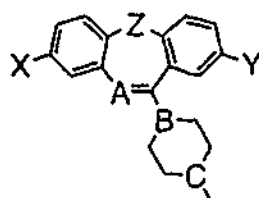
Compound	X	Y	Z[†]	A	B	C	D_2	D_4	$5HT_{2A}$
1A clozapine	Cl	H	NH	N	N	N	220	21	8
1B isoclozapine	H	Cl	NH	N	N	N	47	16	7
1C	H	H	NH	N	N	N	2500	420	39
1D	Cl	H	NCH_3	N	N	N	1100	120	140
1E	Cl	H	O	N	N	N	150	23	11
1F loxapine	H	Cl	O	N	N	N	21	4.9	N/A
2A	Cl	H	CH_2	CH	CH	N	560	2700	3
3A	Cl	H	CH_2	CH	N	N	520	2900	3
3B	H	Cl	CH_2	CH	N	N	1	1	2
3C	H	H	$CHCH_3$	CH	N	N	94	12	47
3D	H	H	$C=CH_2$	CH	N	N	57	47	11
3F	Cl	H	$C=C(CH_3)_2$	CH	N	N	690	930	230
3G	H	H	$C=C(CH_3)_2$	CH	N	N	290	2700	250
3H	Cl	H	O	CH	N	N	8.7	0.9	3
3I	H	Cl	O	CH	N	N	2.5	0.54	3
3J	H	H	O	CH	N	N	21	2	4
P8*	H	H	O	CH	CH	N	82	9.5	3

* indicates compound from Phillips *et al* 1995, numbered P1-P8 in order of new compounds presented

† bold atom represents the atom in ring (Z)

Table 2.2. Receptor binding affinities (K_i, nM) of piperidine analogues from Phillips *et al.*[22,23]
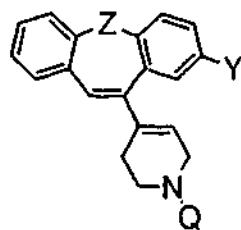
Compound	Y	Z[†]	Q	D_2	D_4	$5HT_{2A}$
4A	H	**C**=CH$_2$	CH$_3$	250	330	30
4B(r)	H	**CH**CH$_3$	CH$_3$	640	190	74
4B(s)	H	**CH**CH$_3$	CH$_3$	730	400	200
4C(pR_apS_b)	H	**C**=C(CH$_3$)$_2$	CH$_3$	520	150	2200
P1[*]	H	**CH**$_2$	CH$_3$	89	12	6
P2[*]	Cl	**C**=C(CH$_3$)$_2$	CH$_3$	440	100	760
P3[*]	H	**CH**$_2$	CH$_2$CH$_3$	44	28	4.5
P4[*]	H	**CH**$_2$	CH$_2$CH=CH$_2$	83	60	20
P5[*]	H	**O**	CH$_3$	230	29	6
P6[*]	H	**O**	CH$_2$CH$_3$	61	26	6
P7[*]	H	**O**	CH$_2$CH=CH$_2$	110	37	22

* indicates compound from Phillips *et al* 1995, numbered P1-P8 in order of new compounds presented

† bold atom represents the atom in ring (Z)

Table 2.3. Receptor binding affinities (K_i, nM) of analogues from Liégeois et al.[24]

Compound	Z	X	Y	D_2**	$5HT_{2A}$**
8	O	H	H	10965	22387
9	O	H	Cl	363	45
10	S	H	H	6607	1148
11	S	H	Cl	263	45
12	S	H	H	2291	282
13	O	H	H	14791	1413
14	O	Cl	H	1202	65
15	O	CH₃	H	1230	98
16	S	H	H	4074	513
clozapine				45.7	3.9
isoclozapine				12.6	1.8
clothiapine				4.5	0.59
isoclothiapine				30.2	4.2
fluperlapine				132	3.9

*NMP denotes N-methylpiperazine

**Binding affinities calculated from pK_i values

2.2.2.1 Molecular Field Analysis

Quantitative structure activity relationships (QSAR) were created for all the data sets developed in the following sections using the method outlined here. The values for CoMFA used were: a grid spacing of 1.5Å, an sp^3 carbon probe with a +1 charge, box smoothing and standard CoMFA scaling. CoMFA electrostatic and steric fields were generated with a grid spacing of 1.5Å to ensure no atoms of interest were missed, as the approximate carbon-carbon bond length is around 1.5Å. When a grid spacing of 1Å was used for the CoMFA electrostatic and steric fields large quantities of smaller areas of supposed interaction were generated, and this only served to obscure key points in the model. A grid spacing of 2Å was also tried for the CoMFA electrostatic and steric fields. This produced models with similar statistics to 1.5Å spacing but their predictive abilities were not as good, due to the larger spacing between the grid points. Partial

least squares (PLS) analysis was carried out on each database. Leave one out (LOO) cross-validation was used to select the number of principal components for cross-validated statistics with column filtering set at 2.0 kcal/mol. The final CoMFA model was generated using no cross-validation and the number of components suggested by the LOO validation run[25].

2.2.2.2 Compound Alignments

Several alternative alignments of the analogues were developed using RMS superimpositioning and field fitting of the ligands. The RMS fit procedure within Sybyl was used to align the ligands to the pharmacophore developed from the conformational analysis and superimpositions described above. Additional alignments were developed using the *field fit* method[26] from Sybyl, where molecules are minimised with respect to the electronic and steric environment of the tightest binding ligand.

In addition to looking at the alignment of the compounds on one another, the effect of charge on the calculation of the electrostatic fields, and the effect this had on the resultant CoMFA models was examined. The effect of rotating the database of compounds within the CoMFA grid space was also analysed, to see whether or not any significant improvements could be made in the predictive nature of the model.

Template Alignment

All compounds were built within Sybyl using standard geometries, and minimised using the Powell minimisation method for 1000 iterations or until they reached a RMS-displacement termination criteria of 0.01Å with MMFF94S[27,28] force field and MM94FF charges.

In this model, dummy atoms to mimic π-π interactions dummy atoms were built 3.5 Å above and below the planes of each aromatic ring[19]. An additional dummy atom was built 2.8 Å from the distal nitrogen of the piperazine ring, along the vector of the ammonium hydrogen, for a proton donor/acceptor interaction[19]. These dummy atoms, Du and dummy atom X, shown in Figure 2.2 were used for superimpositioning. Whilst

this method worked extremely well for the tricyclic set of compounds, it was discovered later that for compounds without the tricyclic substructure, this method placed too much emphasis on the orientation of the planes of these aromatic groups. Therefore for the final superimpositions analysed in this work only the centroids of the aromatic groups, centroids A and B, and point of interaction for the protonated nitrogen were considered for superimposition. This method is justified by past analysis[29] of ligand binding orientations within the protein data bank, showing that in many cases π-π interactions of aromatic groups are not parallel to one another but may take a variety of orientations. However the planarity of the phenyl rings is important for other charge/charge interactions such as those of aromatic hydrogens and oxygen atoms. Thus, attempts were made to ensure some planarity of the superimposed phenyl rings, as long as the energy penalties in doing so were not too high. This early knowledge-based analysis of ligand binding orientations has been expanded on and is now implemented in such programs as Iso-star[30].

Field Fitting Alignment

For the D_2 and $5HT_{2A}$ models, compound 3B Table 2.1 was used for the field fitting alignment whereas for the D_4 model, compound 3I Table 2.1 was used for the field fitting alignment. Minimisations were performed using *MAXIMIN2* with *field fit* on and the Tripos force field using the Powell minimisation method for 500 iterations or until compounds reached an RMS-displacement termination criterion of 0.01Å. A further quick minimisation for 50 iterations with the Powell minimisation method, MMFF94S force field and MMFF94 charges was undertaken to ensure reasonable geometries.

During a minimisation involving field fitting, the derivatives of the field fit penalty with respect to the co-ordinates of each individual atom are added to the other force field derivatives. This results in the compound being minimised to adopt the steric and electronic environment it is being fitted to, in order to minimise the total energy. The field fit steric energy penalty is calculated by the following expression, with the electrostatic field fit penalty being of a similar form[13].

$$E_{steric} = \frac{SFF}{\sum\limits_{p} W(p)} \sum\limits_{p} \left[W(p) * (\sum\limits_{i} V(i,p) - T(p))^2 \right]$$

where:

- p is a lattice point

- T(p) is the value of the steric template field at lattice point p

- V(i,p) is the steric energy of interaction between the ith atom in the target molecule and a probe atom at lattice point p

- $\sum\limits_{i}$ is the sum over all atoms in the target molecule

- W(p) is a user supplied weight for lattice point p

- $\sum\limits_{p}$ is the sum over all points in the lattice

- SFF is a user-supplied overall weight for the field fit term

The user supplied overall weight for the field fit term (SFF), is the weight of the field fit penalty compared with the energy and derivatives found using the Tripos force field[13]. If the value given to overall weight term (SFF) during a field fitting minimisation is too large, the compounds being field fitted may be unreasonably distorted to accommodate the large value from the field fit penalty. If the overall weight term (SFF) is too low there will be little influence from the procedure. Thus a number of different values for SFF were analysed. The resulting databases were also subjected to CoMFA to see if any improvements in their squared correlation coefficients and predictive abilities could be seen.

Analyses of Different Charges and Charged States

In addition to MMFF94 charges that were on the compounds from earlier minimisation, Gasteiger and Marsili, AM1 and ESP charges were analysed within CoMFA. Non-scaled semi empirical ESP charges were added to the compounds using the Mopac 6 module within Sybyl, with the keywords AM1, ESP, MMOK and the compounds corresponding charge. Non-scaled ESP charges were also calculated using the PM3 Hamiltonian for comparison. Mopac AM1 charges were calculated with the keywords

AM1, NOINTER, LET, 1SCF, MMOK.   This analysis of different charges was undertaken whilst the compounds were in both the neutral and protonated form.

Rotations within the CoMFA grid space

As the molecules being analysed within the CoMFA module contain a fair degree of planarity due to the presence of aromatic rings, there is always a chance that during an analysis the compounds in question may not be interacting in an optimum fashion with the CoMFA grid points. Therefore in order to assess whether or not this is a significant factor, all the molecules in the database were rotated 360 degrees in 10 degree increments around the x y and z axes, whilst calculating the q^2 or predictive capabilities of the model at each point.

2.3 RESULTS AND DISCUSSION

2.3.1 Tricyclic Pharmacophore

2.3.1.1 Conformational Analysis

All conformations of (R)- and (S)-octoclothepin were examined because, although often considered to be rigid, the octoclothepin structures exhibit flexibility in several regions. Firstly the piperazine ring is able to rotate around the connecting bond that attaches it to the central seven membered ring of the tricyclic structure (Figure 2.3). Concomitantly, the tricyclic structure is also able to undergo a ring inversion[31] resulting in A-fold and B-fold types (Figure 2.3) as shown from the simulated annealing process. Irurre[32] *et al.* experimentally determined the conformational energy inversion barrier for dibenzo[b,e]azepine derivatives to be around 12 kcal/mol. This suggests that this ring inversion will occur under physiological conditions. This is evidenced by both atropisomers of clozapine being found in the crystal structure[33,34].
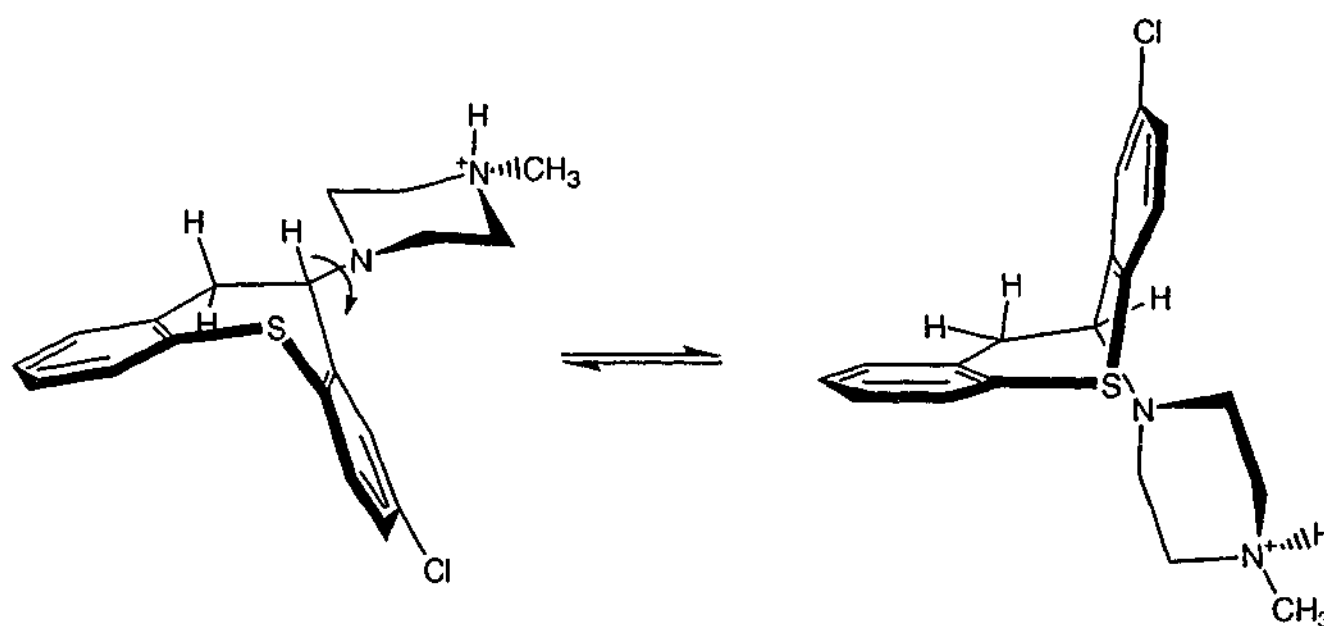
Figure 2.3. *Piperazine ring rotation and tricyclic ring system inversion.*

Secondly, the piperazine ring is able to adopt either a *pseudo*-axial or *pseudo*-equatorial state with respect to the tricyclic structure, shown in Figure 2.4. This was calculated to require an energy input of approximately 9 kcal/mole using the *find transition state*

process in *CAChe*. This compares favourably with results obtained previously by Hendrickson[31], who calculate that the chair/twist-chair and boat/twist-boat families in cycloheptane can be interconverted by passing over a barrier of 8.5 kcal/mol.
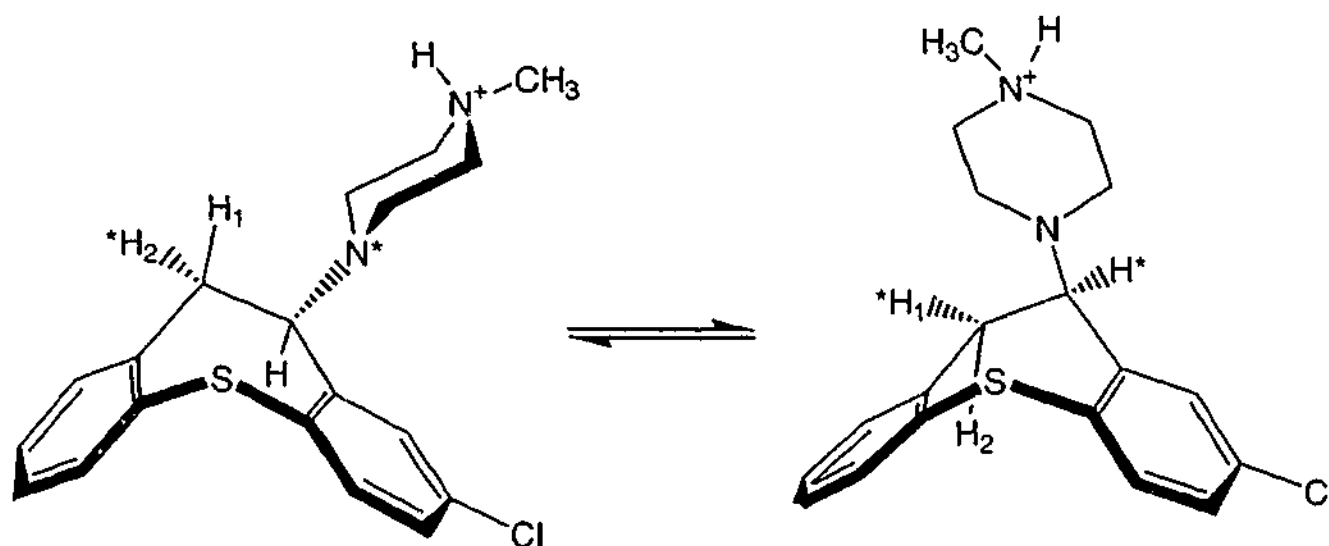
Figure 2.4.   A-fold conformer of (R)-octoclothepin showing alternative pseudo-*equatorial and* pseudo- *axial orientations of the piperazine ring.   H* or N* denotes equatorial position.*

The conformational analysis gave ten low energy conformations for the enantiomers, shown in Figure 2.5.   The relative energies of the differing conformations using different methods of calculation and different solvents are shown in Table 2.3.
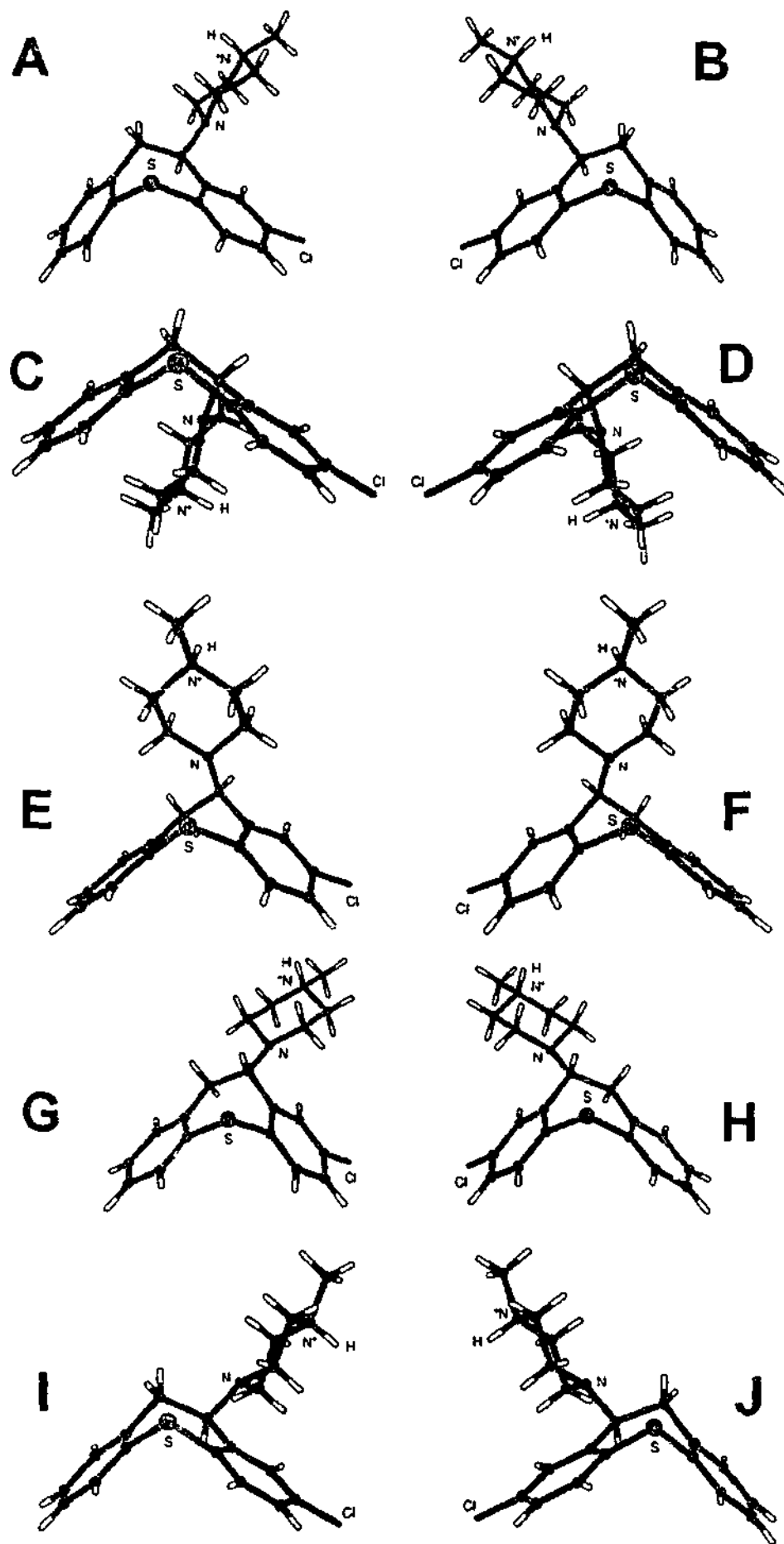
Figure 2.5. *Ten main local and global minimum energy conformers of (R)- and (S)-octoclothepin A. (S)- A-fold equatorial; B. (R)- B-fold equatorial; C. (S)- A-fold axial; D. (R)- B-fold axial; E. (R)- A-fold equatorial; F. (S)- B-fold equatorial; G. (R)- A-fold 2 equatorial; H. (S)- B-fold 2 equatorial; I. (R)- A-fold axial; J. (S)- B-fold axial.*

The first column of Table 2.4 shows the heat of formation *in vacuo* calculated using the Tripos force field with Gasteiger and Marsili charges, indicating the global minimum of (R)- and (S)-octoclothepin are the A-fold *pseudo*-equatorial conformer (A) and the B-fold *pseudo*-equatorial conformer (B), respectively.

Table 2.4. Relative Energies in kcal./mol using Differing Methods and Solvents

Ligands	Tripos (SS) Vacuum	Amsol Vacuum	Amsol Water	Amsol Heptane	MacroModel Vacuum	MacroModel Water
A & B	0	3.0	0	0	0	0
C & D	0.3	0.4	3.3	1.7	5.4	4.2
E & F	0.8	0	2.7	1.8	2.7	3.1
G & H	2.1	3.4	0.6	0.2	3.4	3.0
I & J	5.0	3.9	0.8	0.4	2.4	3.0

(SS) denotes energies from systematic searches

The results for the final heat of formation *in vacuo* from Amsol were in a different order of increasing energy compared to those from the Tripos force field, although there was only a 3.9 kcal./mol difference between the lowest and highest energy conformations. The results for the conformational energies in solution (water) from Amsol gave a similar order to MacroModel (water), with a 3.7 kcal./mol range between the lowest and highest energy conformations, compared to a 4.2 kcal./mol range from MacroModel. The results for conformational energies in solution (heptane) from Amsol also showed a range of less than 5 kcal./mol between lowest and highest energy conformations, and gave the same global minimum as most other methods.

The results obtained from MacroModel for the conformational energies *in vacuo* were different from those obtained with the Tripos force field with respect to the order of increasing energy. However the global minima were the same for both (R)- and (S)-octoclothepin.

These semi-empirical and molecular mechanics calculations are in good agreement with one another, with all having a range of less than 5 kcal./mol between all conformations and five out of six methods giving the same global minimum.

The global minimum of (S)-octoclothepin, conformer (B), was consistent with the crystal structure obtained from the Cambridge Crystallographic Database[11].

2.3.1.2 Superimposition

As (R)- and (S)-octoclothepin are both active at the dopaminergic and serotonergic receptors, it is assumed here that they are acting in a conformationally similar manner. The binding affinities in nM of the (R)- and (S)-octoclothepin enantiomers for the D_2 receptor[35] are 7.1 and 4.5, for the D_4 receptor[35] are 2.8 and 1.5, and for the $5HT_{2A}$ receptor[36] are 0.33 and 0.14, respectively. Therefore superimpositions were performed and the measured RMS fit of the R conformers on to the S conformers was examined to find pairs of conformers that were capable of interacting with common sites in the receptor.

The best RMS fit, with a value of 0.45 Å, resulted from the superimposition of all five dummy atoms of the A-fold (S)-octoclothepin *pseudo*-equatorial, conformer **G**, on to the global minimum A-fold (R)-octoclothepin *pseudo*-equatorial, conformer **A**, and the mirror image B-fold (R)-octoclothepin *pseudo*-equatorial, conformer **H**, on to the global minima B-fold (S)-octoclothepin, conformer **B** (see Figure 2.5). This indicates that these are strong candidates for the relevant conformation mediating antipsychotic activity.
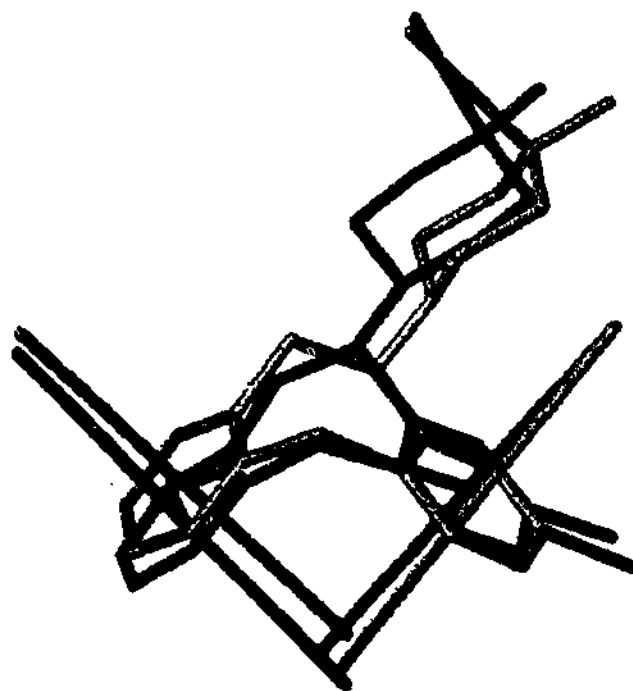
Figure 2.6. *Superimposition of A-fold (R)-, light shading, and (S)-, dark shading, octoclothepin showing dummy atoms.*

2.3.1.3 Defining the Pharmacophore

To differentiate between the mirror image pharmacophore models generated above, we superimposed low energy conformations of the D_2 and D_4 active compound (1R,3S)-tefludazine onto the pharmacophores. D_2, and D_4 binding affinities[35] for tefludazine are 4.6 and 15 nM, respectively. This gave a better fit with the A-fold conformation, suggesting that this conformation is more likely to represent the biologically active one. Figure 2.7 shows the resulting superimposition of the R and S enantiomers of octoclothepin and (1R,3S)-tefludazine.
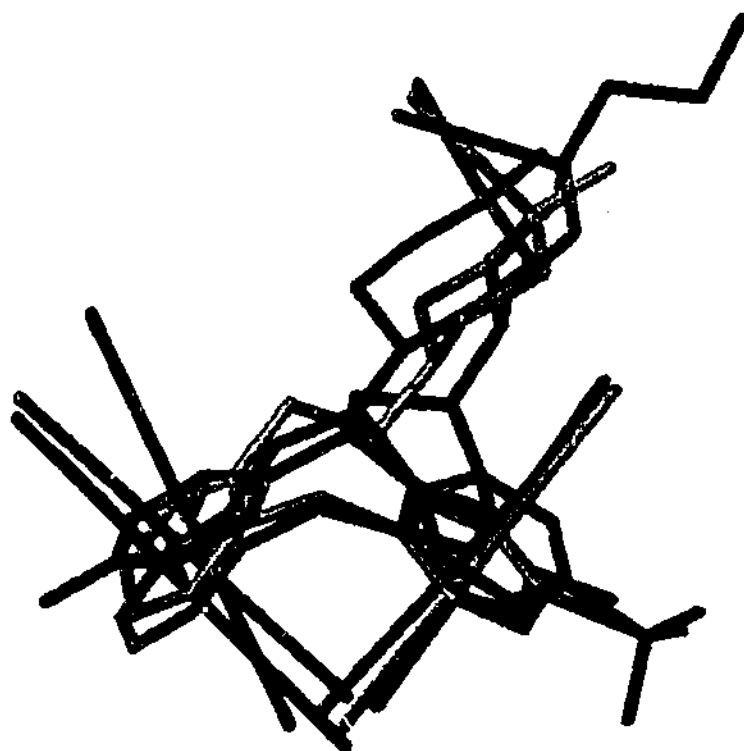
Figure 2.7.   Tefludazine, black, superimposed on A-fold (R)-octoclothepin, light shading, and (S)-octoclothepin, dark shading.

A search of the Cambridge Crystallographic Database revealed one conjugated pyridine, 1,2,3,6-tetrahydro-1-Q- of the type in the Phillips[22] dataset (Table 2.2), with an A-fold conformation matching that deduced in this work. All other structures of conjugated dienes[37-41] similar to the type in Table 2.2, that were neither sterically hindered nor co-crystallised with counter ions, crystallised in a *pseudo*-trans arrangement about their single bond.

2.3.1.4 Comparison with other D₂, D₄ and 5HT₂ₐ Antagonist Pharmacophores

Figure 2.8 shows the distances between the likely points of interaction in the derived D_2, D_4 and $5HT_{2A}$ pharmacophore.
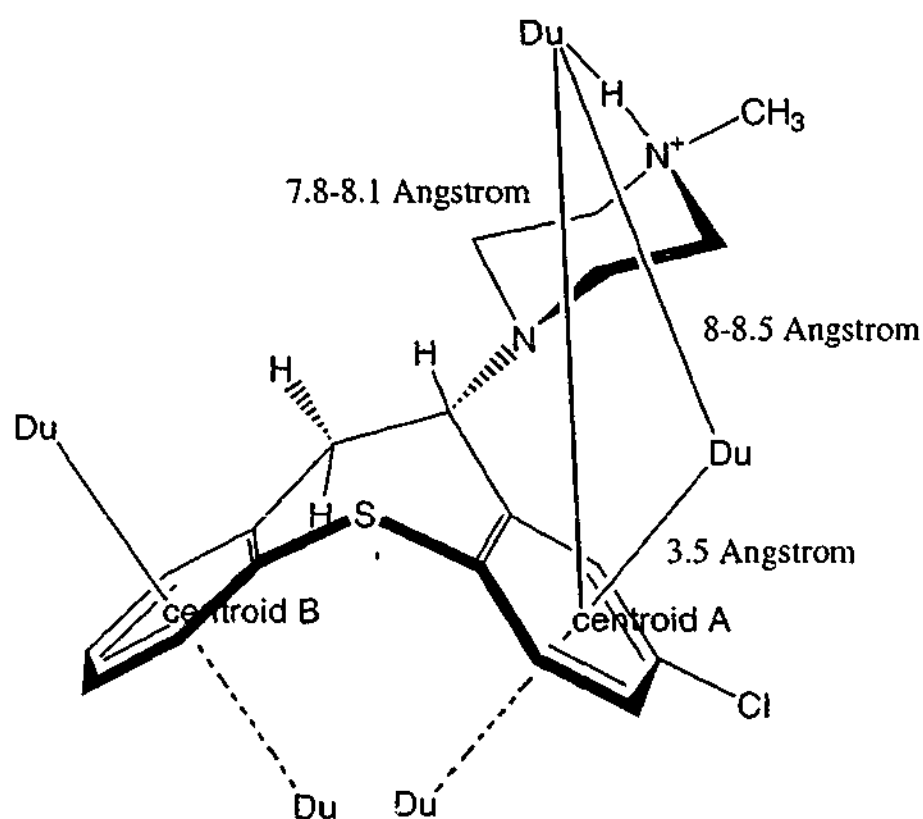
Figure 2.8. Proposed pharmacophore.

This model is consistent with the model developed by Bøgesø[8] *et al.* However the global minimum energy conformation of (R)-octoclothepin was used with dummy atoms representing likely points of interaction for superimpositions. Differing weightings of the superimposition points were also used in the RMS alignment of the ligands. In addition, the position of the distal nitrogen in the model generated here covers a circular area of approximately 2.0Å diameter, resulting in distances from the centre of the aromatic ring, centroid A, to the nitrogen of between 5.4 and 6.7Å. These distances are in agreement with some of the older models for antipsychotic activity that have been generated, such as the generalised model for antipsychotics of Lloyd and Andrews[42], which has a distance of 5.2Å from nitrogen to aromatic centroid. The D_4 model is also consistent with that of Boström et al[35], although the range of compounds examined in Boström's paper far exceeds the range looked at in this section of the chapter and is discussed in greater detail later in "Extended Pharmacophore: Comparison with other D_2 and D_4 antagonist pharmacophores".

2.3.2 Tricyclic Molecular Field Analysis

2.3.2.1 Data Selection

Compounds were chosen on the basis of binding affinity and close structural similarity. The main difficulties in the development of reliable alignments is not only ligand flexibility but also the ambiguity of postulated receptor-ligand interactions and the possibility of more than one binding site or binding mode. Therefore we preferred to obtain a model based on a set of congeners that acted in a similar fashion, rather than trying to fit compounds that were too structurally dissimilar.

The three datasets used in the tricyclic model development were taken from Phillips et al.[22,23] and Liégeois et al.[24].

Tables 2.1 and 2.2 contain the 5H-dibenzo[b,e][1,4]diazepine, dibenz[b,f]oxepin and chiral 5H-dibenzo[a,d]cycloheptene analogues from Phillips et al.[22,23]. The binding affinities for the D_2 and D_4 receptors were determined by inhibition of [3H]spiperone from COS-7 cells transfected with a gene expressing the human D_2(long) and D_4 receptors, in the presence of sodium chloride. The binding affinities for the $5HT_{2A}$ receptor were determined by inhibition of ^{125}I-labelled lysergic acid diethylamide binding to membranes containing the cloned rat $5HT_{2A}$ receptor. The compounds from these two papers were considered as ideal candidates for modelling together, due to the fact they were from the same laboratory and had identical procedures for determination of their binding affinities. Two compounds were not used in the CoMFA models generated: 2B, as it did not possess a distal nitrogen, (pR_apS_b)-(-)-4c, as it was in the B-fold conformation and thus the incorrect atropisomeric form for the pharmacophore.

Table 2.3 contains a series of pyridobenzoxazepine and pyridobenzothiazepine derivatives as well as a number of well known antipsychotic drugs and their iso-counterparts from the paper of Liégeois et al.[24]. Their binding affinities for the $5HT_{2A}$ receptor were determined by inhibition of [3H]-labelled ketanserin binding to pretreated prefrontal cortex brain tissue from female Wistar rats[43]. The binding affinities for the D_2 receptor were determined.

These compounds were considered as good candidates for modelling with the previous datasets of Phillips *et al.*[22,23], as they also contained the 6-7-6 tricyclic substructure. The inclusion of a heteroatom into this substructure by Liégeois *et al.*[24], should also shed more light on the electrostatic interactions required for potency. However when comparing the binding affinities of identical compounds, clozapine and isoclozapine, from the Liégeois and Phillips datasets, it can be seen that the binding affinities reported by Liégeois *et al.*[24] are similar for the $5HT_{2A}$ receptor and marginally tighter for the D_2 receptor. This must be kept in mind when analysing the CoMFA predictions of these compounds.

2.3.2.2 Compound Alignments

Template Alignment

When the compounds were minimised using the MMFF94S force field and MM94FF charges and then superimposed onto the proposed pharmacophore (see Figure 2.8), it could be seen that compounds with a similar substructural unit all superimposed on one another in 3D space (see Figure 2.9). These substructural units could be roughly divided into four main groups depending on the seven membered ring and the group attached: diazepine attached to piperazine; cycloheptene attached to piperidine; cycloheptene attached to piperazine and cycloheptene attached to pyridine, 1,2,3,6-tetrahydro.

The CoMFA analysis on these compounds was performed using the MMFF94 charges and a grid spacing of 1.5 angstrom. The resultant model had four components giving q^2 of 0.472 and r^2 of 0.807; the relative contributions to the model were 43% and 57% from the steric and electrostatic fields, respectively.
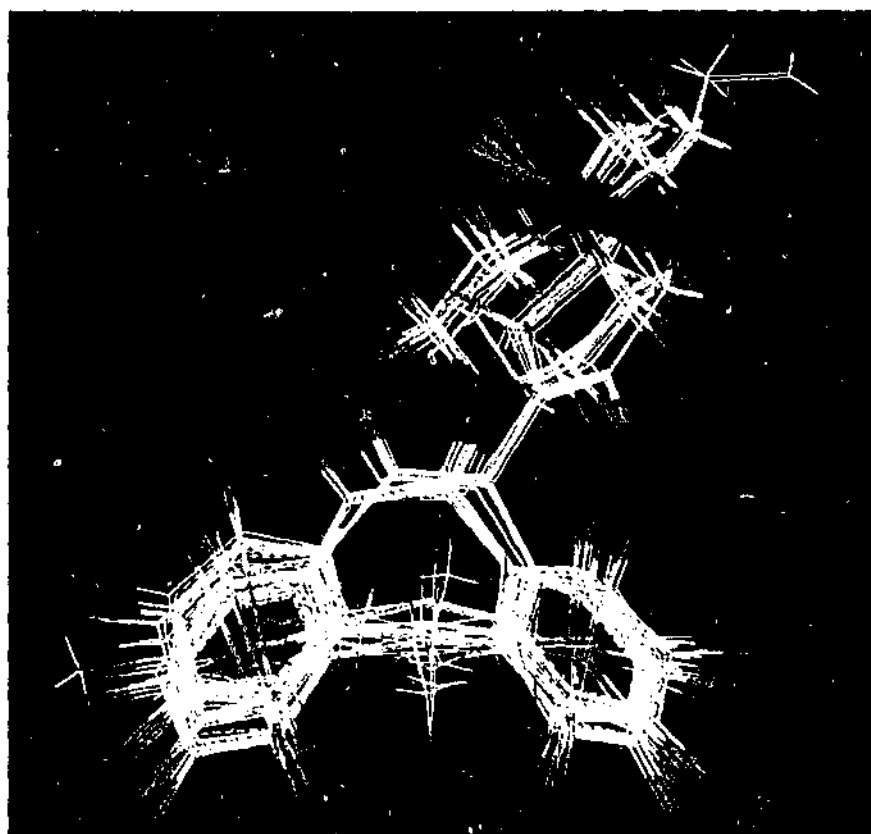
Figure 2.9. *Template based alignment of the tricyclic compounds showing the grouping of the substructural units.*

The two cycloheptene attached to piperidine compounds, 2A and P8, were both outliers, which was not surprising as they were the only two representative compounds of this substructure and were not well aligned with the other compounds. The compound that was most poorly predicted was 3A with a residual of 1.3 log units. This was initially surprising, however upon closer inspection of the source of the binding affinity for this compound[44] it was shown that a different method involving rat caudate nuclei was used for compounds 3A and 3B which may explain some discrepancies.

Field Fitting

In the field fitting procedure within Sybyl, molecules are minimised with respect to the electronic and steric environment of the tightest binding ligand (see Figure 2.10).

Figure 2.10. *Steric and electrostatic fields, respectively, of compound 3B Table 2.1, from which the D_2 and $5HT_{2A}$ alignments were generated. Favoured region for the steric field is shown in yellow. Favoured regions for positive and negative charge for the electrostatic fields are shown in red and green, respectively.*

As the field fitting procedure caused the ligands being analysed to adopt high-energy conformations, they subsequently had to be reminimised to obtain reasonable geometries. It was found that insufficient minimisation produced conformations that were 15 kcal/mole or more above their local minimum. A 10 kcal/mole difference between a conformer and its local minima is considered to be a reasonable cutoff[45] and these conformers were not considered reasonable. Alternatively, when the molecules in the databases were minimised for too many iterations, the piperazine ring often rotated relative to the seven-membered ring so that the dummy atom from the distal nitrogen no longer fitted within the constraints of the pharmacophore. After several trials the best results came from a quick minimisation of 50 iterations. In addition, one must also consider the effect of altering the weight of the field fit penalty in the field fitting procedure. If the user supplied overall weight - SFF value - is too high, the compounds being minimised may become significantly distorted and if the SFF value is too low there is little influence from the procedure. Thus it was necessary to find a complementary balance between the SFF value and a quick minimisation performed post field fitting.

Table 2.5. Average energy (kcal/mol) above global minima before and after field fitting.

	SFF = 1	SFF = 5	SFF = 10	SFF = 20	SFF = 50	SFF = 100
without minimisation*	22.42	42.65	65.80	101.83	198.62	347.72
with minimisation*	0.76	1.19	1.64	1.25	1.45	1.92
distance**	7.7	7.8	7.8	7.8	7.9	7.9

* alone and after fieldfitting and minimisation for 50 iterations, whilst varying the SFF factor for all the compounds from the Tricyclic dataset with the D_2 binding affinities.

** average distance between dummy atoms X and A

Table 2.5, which shows the average energies after minimisation, indicates that the compounds all return to a conformation that is realistic, without deviating too far from the constraints of the pharmacophore. Table 2.5 also shows that setting the SFF factor to 1 causes the least structural deformations. As the average energies of this dataset were lower, thus less movement of the compounds into the steric and electrostatic fields of the tightest binding compounds was seen in comparison to datasets with higher SFF values. Consequently there was really little or no influence from the procedure as is evident when comparing the q^2 values obtained for no field fitting with field fitting and SFF set to 1 (see Table 2.6).

Table 2.6. Adjusting the SFF variable in the field fitting algorithm and calculating the q^2 and r^2 with default settings for the tricyclic dataset with D_2 binding affinities.

	no field fitting	SFF=1	SFF = 5	SFF = 10	SFF = 20	SFF = 50	SFF = 100
q^2	0.430	0.499	0.558	0.559	0.512	0.423	0.442
n	3	3	3	3	6	2	3
r^2	0.699	0.800	0.817	0.831	0.912	0.687	0.783

From these analyses it is difficult to see which SFF value offers the most improvement in the CoMFA analyses as SFF values of 5, 10 and 20 all perform well, although there are twice as many components in the SFF 20 model. Therefore in all subsequent field fitting procedures a SFF value of 10 in combination with a post field-fitting minimisation of 50 iterations was used, as this method provided the most predictive model as indicated by the q^2 value.

The field fitting method offered significant improvements in all areas of the CoMFA model generation when compared to the template alignment method; thus this method was used for all subsequent analyses.

Charge Analysis

The tricyclic dataset was analysed within CoMFA using four differing methods of calculating charge on the compounds to assess which method worked best. Gasteiger and Marsili charges were used because much of my early CoMFA work was done using these charges. With the release of Sybyl 6.4 and the addition of the Merck molecular Force Field and its corresponding charges, these MMFF94 charges were also analysed. Mopac AM1 charges were also calculated to assess whether or not semiempirical charges were significantly better. The ESP semiempirical charges were also assessed to see if the charges calculated from a distribution over the compounds surface affected results to any degree.

The difference in charge distribution between the methods is quite evident (see Figure 2.11), indicating how the CoMFA electrostatic fields differ.
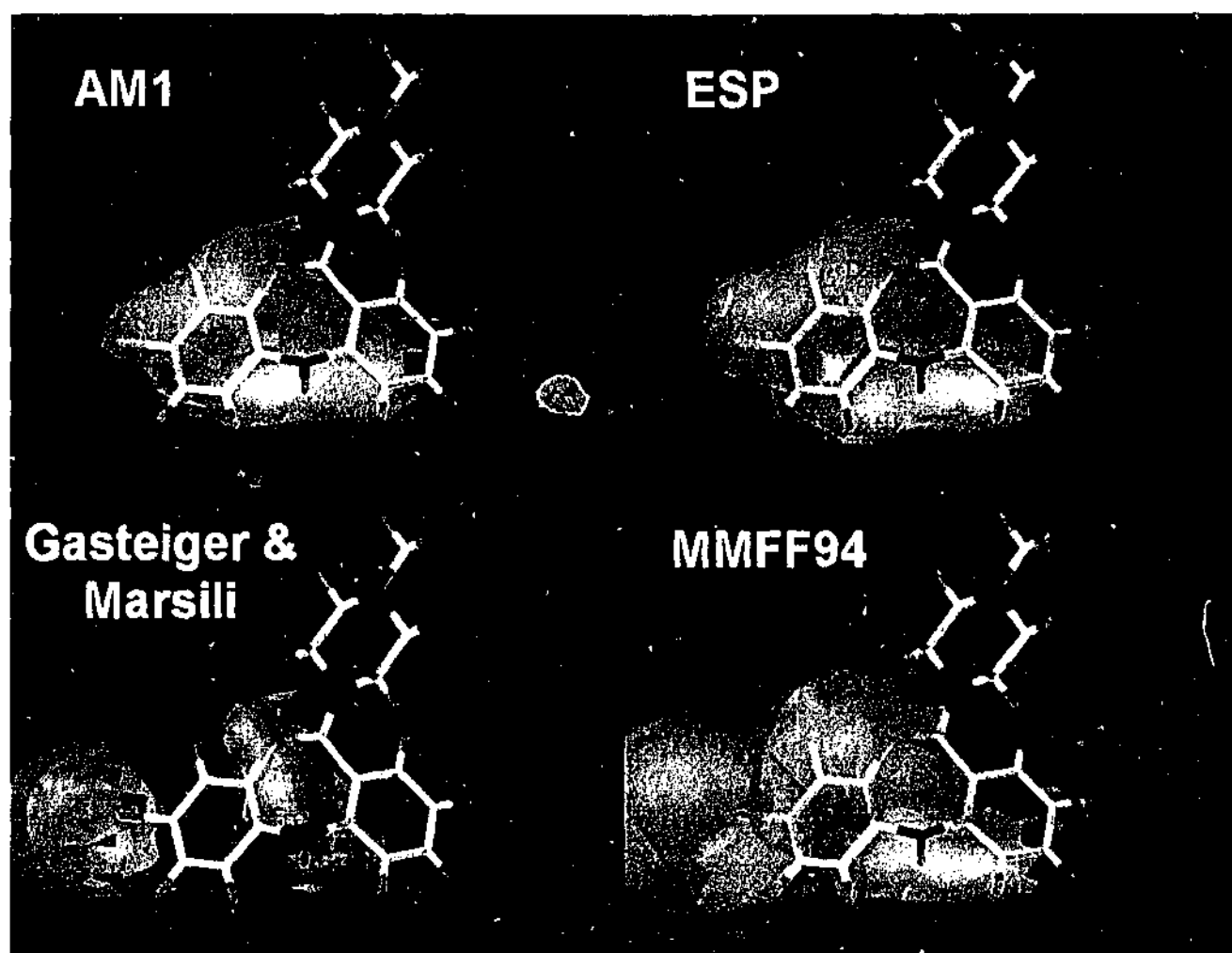
Figure 2.11. *Isopotential surface calculation of clozapine using AM1, ESP, MMFF94 and Gasteiger and Marsili charges. Blue and red indicate negative and positive charge, respectively.*

The correlation between the charges from the two MM methods is $r = 0.482$, resulting in the distinctly different isopotential surfaces. The correlation between the charges from the two semi-empirical methods is $r = 0.852$, which may explain why the isopotential surfaces look similar. There are subtle differences between the two semi-empirical methods specifically in some of the atoms of interest such as the distal nitrogen, the methyl carbon attached to the distal nitrogen as well as many atoms in the tricyclic substructure. It can also be seen in Figure 2.11 that the positively charged distal nitrogen has a dominant effect and effectively shields the negative regions of the molecule.

The CoMFA electrostatic fields did differ, although the only significant difference in the contributions from these fields was from the AM1 charges. In the AM1 CoMFA analysis the relative contributions from the two fields were 56% and 44% for steric and electrostatic respectively, whereas for all other analyses the contributions were

approximately 48% and 52% for steric and electrostatic, respectively. Interestingly the fields derived using Mopac AM1 charges gave the poorest correlation to binding affinity when used in CoMFA followed by the two molecular mechanics methods of charge calculation (see Table 2.7). In addition to this there also seemed to be no correlation between the magnitude of the charges on the atoms and the predictive ability obtained by the models, indicating that charge distribution over the compound was the most important feature in calculation of the electrostatic fields. This is why it was reassuring to see that ESP charges performed best, as point charges calculated to represent the charge distribution over the surface of the compound should be expected to perform better than charges calculated from simplistic atom types. ESP[46] calculates the expectation values of the electrostatic potential of a molecule on a uniform distribution of points. The resultant ESP surface is then fitted to atom centred charges that best reproduce the distribution, in a least squares sense.

Table 2.7. Analysis of CoMFA using differing charges on the field fitted tricyclic dataset with SFF set to 10 and D_2 binding affinities.

Charge method	components (n)	q^2	r^2
MMFF94	4	0.597	0.887
Gasteiger & Marsili	7	0.617	0.952
AM1	6	0.580	0.933
ESP$_{AM1}$	4	0.685	0.927
ESP$_{PM3}$	4	0.674	0.908

As the positively charged distal nitrogen has a dominant effect, the charge analysis was also performed for the same datasets in a neutral form. The dominant effect of the positively charged nitrogen is easily evident when viewing the isopotential surfaces of the charged and uncharged clozapine compounds (see Figures 2.11 and 2.12, respectively). This dominant effect of the positively charged nitrogen is also evident when looking at the charge distribution from the semi-empirical AM1 method. In Figure 2.11 the positively charged nitrogen effectively shields the chlorine substituent resulting in a net charge of zero, whereas for the uncharged compound the charge on the chlorine is -0.026. This results in an isopotential surface for the uncharged AM1 clozapine compound that is more like the MM methods of charge distribution, indeed

even the correlation between AM1 charges and Gasteiger & Marsili charges increases from r=0.64 to r=0.79.

A similar order of prediction was seen when looking at the uncharged CoMFA models, with the fields derived from the AM1 method giving the poorest correlation to binding affinity, the two MM methods next and the ESP model being the most predictive model (see Table 2.8). However the relative contributions from the fields did change noticeably from approximately 50% each to a 33% steric and 66% electrostatic field contribution, although when viewing the resulting CoMFA fields no additional information could be gained in comparison to the CoMFA fields generated from the charged compounds.
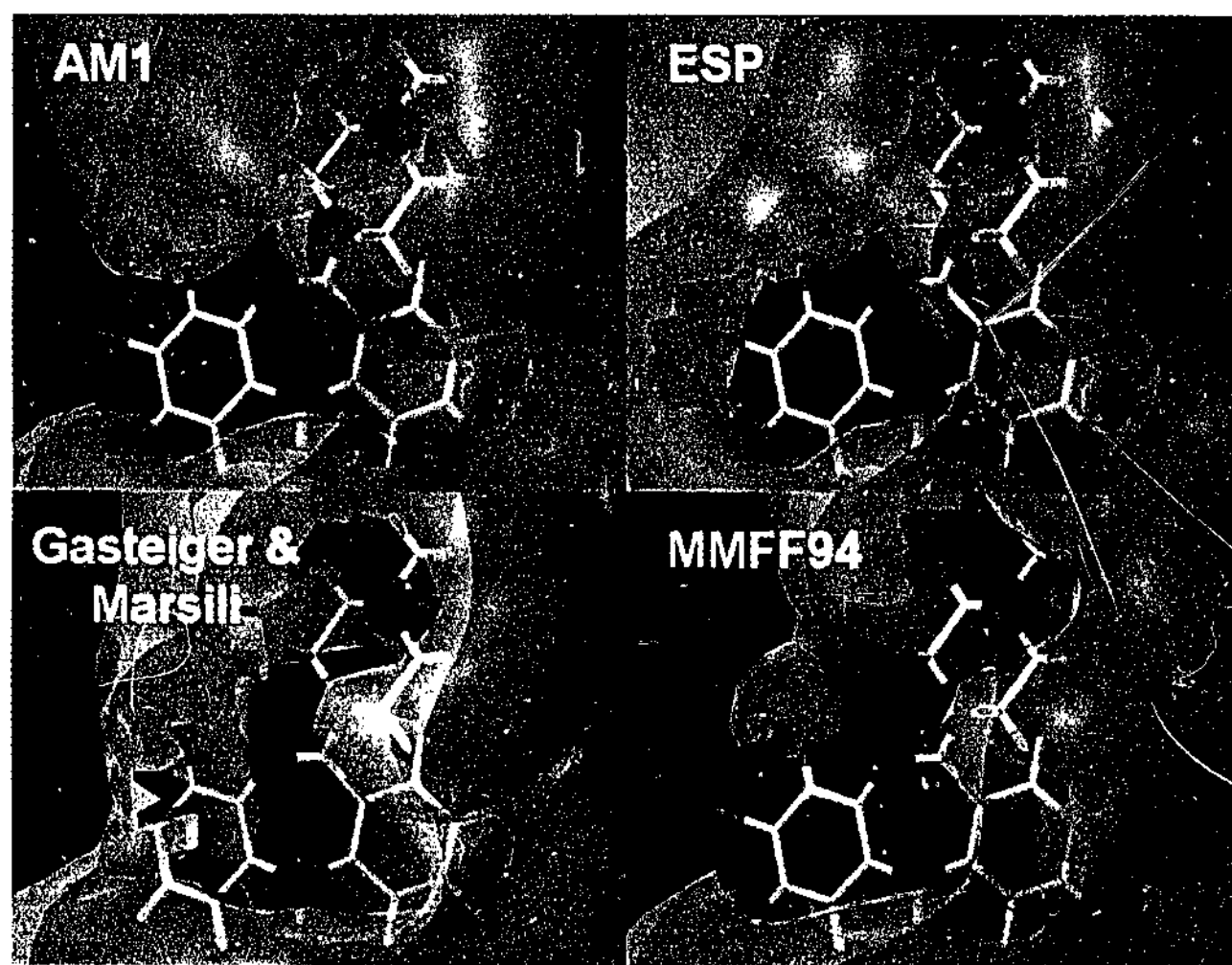
Figure 2.12. Isopotential surface calculation of clozapine, with a net zero charge, using AM1, ESP, MMFF94 and Gasteiger & Marsili charges. Blue and red indicate negative and positive charge, respectively.

Table 2.8. Analysis of CoMFA using differing charges on the field fitted tricyclic dataset with SFF set to 10 and D_2 binding affinities.

Charge method	components (n)	q^2	r^2
MMFF94	10	0.692	0.986
Gasteiger & Marsili	10	0.664	0.983
AM1	4	0.565	0.880
ESP_{AM1}	4	0.712	0.921

So although analysis of both the charged and uncharged datasets was interesting to see, in the difference that charge can have on the electrostatic characteristics of a compound, it did not highlight any additional regions of interest in the CoMFA models. As the compounds are expected to be in a protonated state under physiological conditions, charged or protonated compounds were used in all following datasets.

It was reassuring to see some continuity in the CoMFA models generated using both the charged and uncharged compounds, for the different methods of charge calculation. And as the ESP method of charge calculation performed best in both situations it was used in the generation of all subsequent CoMFA models.

Rotations within the CoMFA grid space

Rotating the Tricyclic template database model around the x, y and z-axes in 10° increments did show there was the possibility of improving the model obtained in terms of its predictive nature (see Figure 2.13). The database used in this analysis was the field fitted database with the SFF value set at 10.
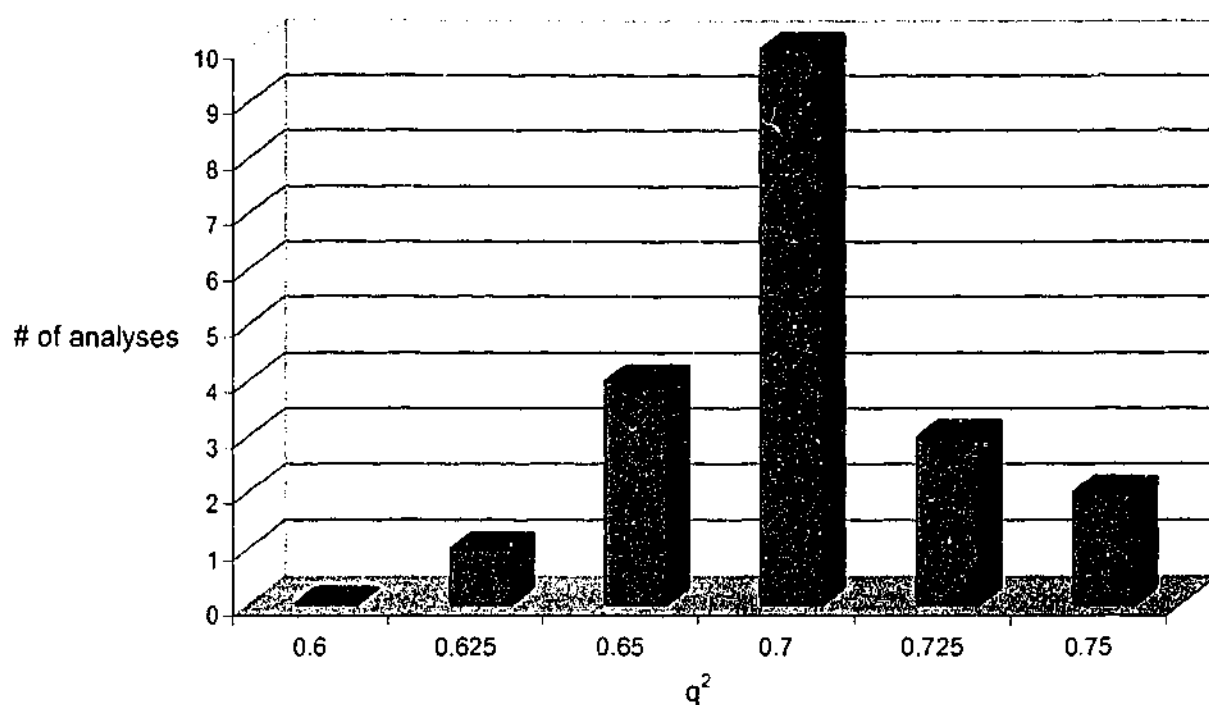
Figure 2.13. *Histogram of q^2 values obtained from ~~ ..g the field-fitted tricyclic database model around the x, y and z axes in 10° increments. Note: only q^2 values obtained with the same number of components (4) were chosen.*

An improvement from the initial q^2 value of 0.685 to 0.75 was gained by rotating the molecules in the database within the CoMFA grid. The predictive ability of the various orientations of the molecules within the CoMFA grid shows a normal distribution. The initial q^2 obtained could easily have been at either end of this distribution demonstrating the importance of doing such an analysis. Thus such an analysis was performed for all final models developed, with the most predictive model being chosen for inspection of its electrostatic and steric fields.

2.3.2.3 CoMFA Field Interpretations

All fields shown in this section were generated using the method outlined immediately above. The contour levels for visualisation of the CoMFA fields were set to represent 50% of the relative contribution of each factor involved in correlating the pharmacophore model to activity. This must be done, because if the distribution of the factors involved in the whole field signal are not symmetrical, the default contour settings will give a display that may lead to a wrong interpretation of the results[47].

D$_2$ Dopaminergic model

When the database of compounds was rotated -30° around the z-axis relative to the starting orientation, the best model with the highest q^2, r^2 and the lowest standard error of prediction was found. The resultant model had 4 components giving a q^2 of 0.75 and a r^2 of 0.927 with a standard error of estimate of 0.268; the relative contributions to the model were 4 5% a nd 5 5% f rom t he s teric a nd electrostatic f ields, r espectively. T he predicted activities of all compounds were within ±0.75 pK$_i$ log units of their measured values.
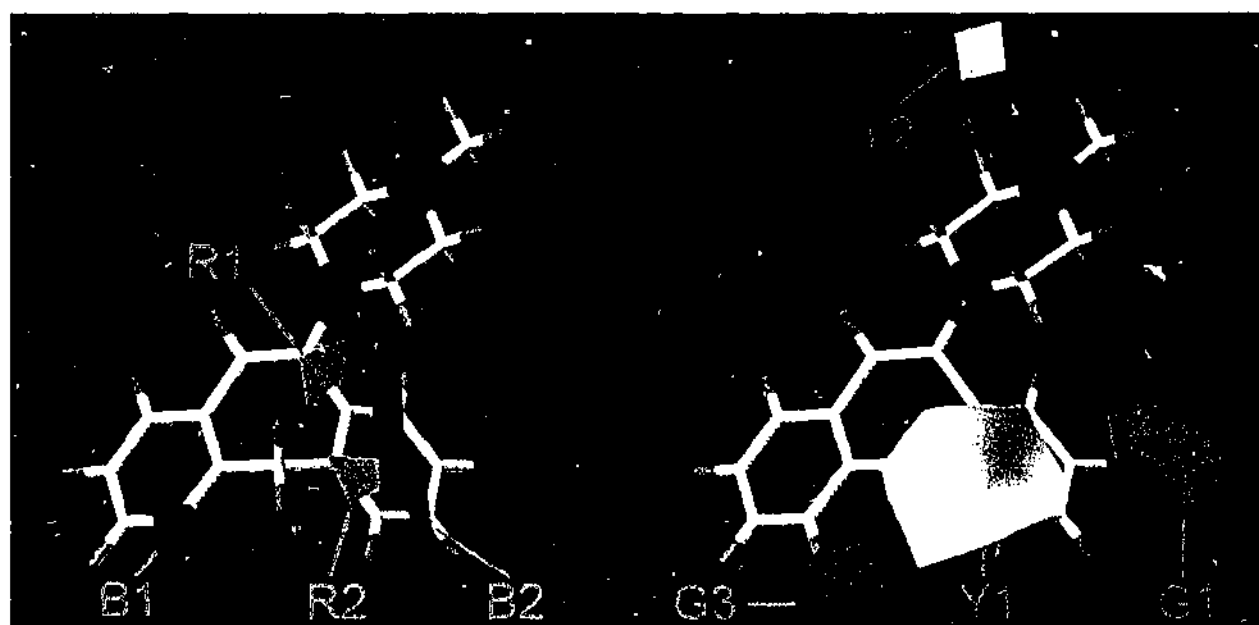
Figure 2.14. *CoMFA fields showing electrostatic (left) and steric (right) contributions from analysis of the D$_2$ binding affinities.*

The electrostatic map, Figure 2.14, shows two regions where partial negative charge enhances activity (red), R1 and R2. The region at position R1 shows the detrimental effect that a nitrogen in position A or B can have on D$_2$ binding affinity, as the nitrogen atom will draw electrons from the carbon at R1 leaving it with a partial positive charge. This detrimental effect can be clearly seen when comparing the binding affinities of compounds 1E and 1F to compounds 3H and 3I from Table 2.1. This region at R1 may also indicate that piperidine rings, compared to piperazine rings, are preferential for D$_2$ binding, however there are not enough compounds with a piperidine ring to draw a solid conclusion. The second region where partial negative charge enhances activity is at R2; this region is due to the presence of a nitrogen atom in position 6 on the compounds 13-16 in Table 2.3. The nitrogen is withdrawing electrons from the neighbouring carbon

and imparting a partial positive charge, which is detrimental to D_2 binding affinity. This can be more clearly seen when comparing compound 1E to compound 14, resulting in binding affinities that differ by almost an order of magnitude.

The electrostatic map, Figure 2.14, also shows two regions where a partial positive charge enhances activity at the D_2 receptor. The region at B1 is easily explained by comparing compounds 1F and 9, where the presence of hydrogen with a partial positive charge or absence of nitrogen in ring B increases binding affinity. The second region where partial positive charge enhances activity is on ring A labelled B2. The explanation of such a region is not immediately evident until the ESP charge distribution on ring A is examined. As ESP charges are calculated from the expectation values of the electrostatic potential on the surface of a molecule, the final point charges assigned to the atoms may differ significantly from the AM1 hamiltonian whence they were derived initially and its corresponding charges, see Figure 2.15. This method of charge calculation ends up assigning a partial positive charge to the chlorine substituent and the carbon atoms at positions 1 and 3 of ring A. These partial positive charges at positions 1 and 3 are caused by the presence of a chlorine substituent at position 2 showing the benefits of having a chlorine substituent for enhancing D_2 activity.
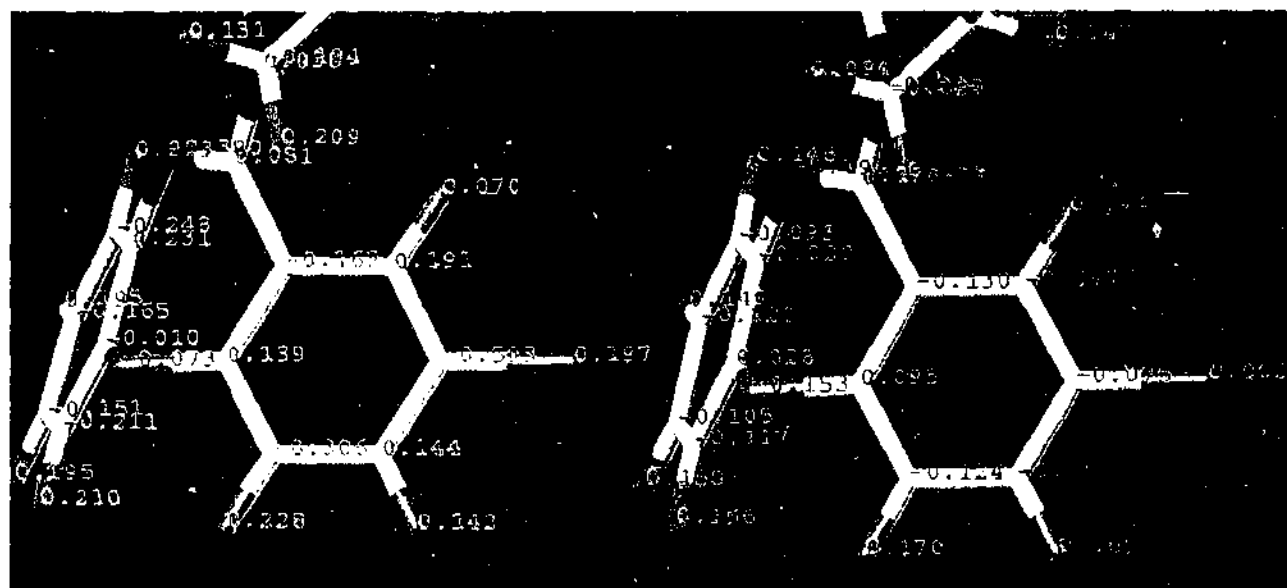
Figure 2.15. *Diagram showing the differing charge distributions calculated by the ESP charge calculation method (left) and the MOPAC AM1 method (right). The ESP partial positive charges on carbons in position 1 and 3 are 0.191 and 0.144, respectively.*

The steric map (Figure 2.14) shows a yellow region indicating areas where steric hindrance occurs at Y1, indicative of the unfavourable interactions that are occurring with bulky alkyl groups. This can easily be seen when comparing compounds 4B(R) and 4B(S) with compound P1, where the activity drops almost ten fold with the addition of a methyl group at position Z. The other region where additional bulk is unfavourable is at Y2: this region emanates from the fieldfitting of compounds with a large substituent at position Q into the steric field of compound 3B. Whilst fieldfitting these larger substituents into this smaller steric field the compounds are moved across to the left to occupy this region, Y2. The poorer affinity of compounds with a large substituent at position Q is seen when looking at Table 2.2.

The steric map (Figure 2.14) also shows two regions where extra bulk is favourable at G1 and G3. The region at G1 is indicative of the benefits of a chlorine substituent occupying this region for D_2 binding affinity. The region G3 is caused by hydrogen occupying this region and the absence of nitrogen in ring B, the detrimental effect of nitrogen in ring B to D_2 binding affinity being obvious when looking at compounds 8 to 11 in Table 2.3.

D_4 Dopaminergic model

As there was no D_4 dopaminergic binding data for the Liégeois dataset, these compounds were not included in the D_4 tricyclic model. The best model was found when the database of compounds was rotated 60° around the x-axis relative to the starting orientation. The best resultant model had 2 components giving a q^2 of 0.351 and a r^2 of 0.721; the relative contributions to the model were 56% and 44% from the steric and electrostatic fields, respectively. Further improvement of the model was gained by reanalysing the QSAR model by region focusing using default values and the weight by discriminant power set to 0.3. This significantly improved the predictive ability of the model, increasing to 0.401 and a r^2 0.723 with 2 components, the relative contributions of the model changing slightly to 55% and 45% from the steric and electrostatic fields, respectively. The predicted activities of all compounds were within ± 1 pK_i log units of their measured values with the exception of compound 3H, which had a residual of 1.3 log units. Interestingly the next two poorly predicted compounds

were 1A and 2A with residuals of 0.9 and -0.9 log units, respectively, which indicates that the model was not predicting 8-chloro substituted compounds well. Even though the model did not show very good predictive statistical results it is still of some interest, as i t h as b een s hown t hat C oMFA m odels s howing a q^2 h igher t han 0 .3 h ave a l ow probability of chance correlation[48].

Further attempts to improve the model by altering the grid size spacing gave minor improvements, although for consistency these were omitted and only the CoMFA procedure of region focussing was employed. Region focusing is an iterative procedure that refines a model by increasing the weight for those lattice points that are most pertinent to the model. Region focusing did significantly improve the m odel on this occasion although this was not always the case. In the case for the optimum D_2 binding affinity dataset the model was only marginally improved from a q^2 of 0.75 to 0.753 with the further complication of introducing two additional components. This was not seen as an improvement of the model, only a complication. In addition, region focusing did not prove to be as successful in increasing the predictive ability of the models in comparison to rotation of the database around the xyz axes. As when region focusing was applied to the original D_2 binding dataset which had a q^2 of 0.685, an improvement to only 0.733 was seen with this technique. This improvement fell short of the improvement to 0.75 gained by dataset rotation, therefore region focusing was only used as a final measure and only mentioned if the gains were significant.
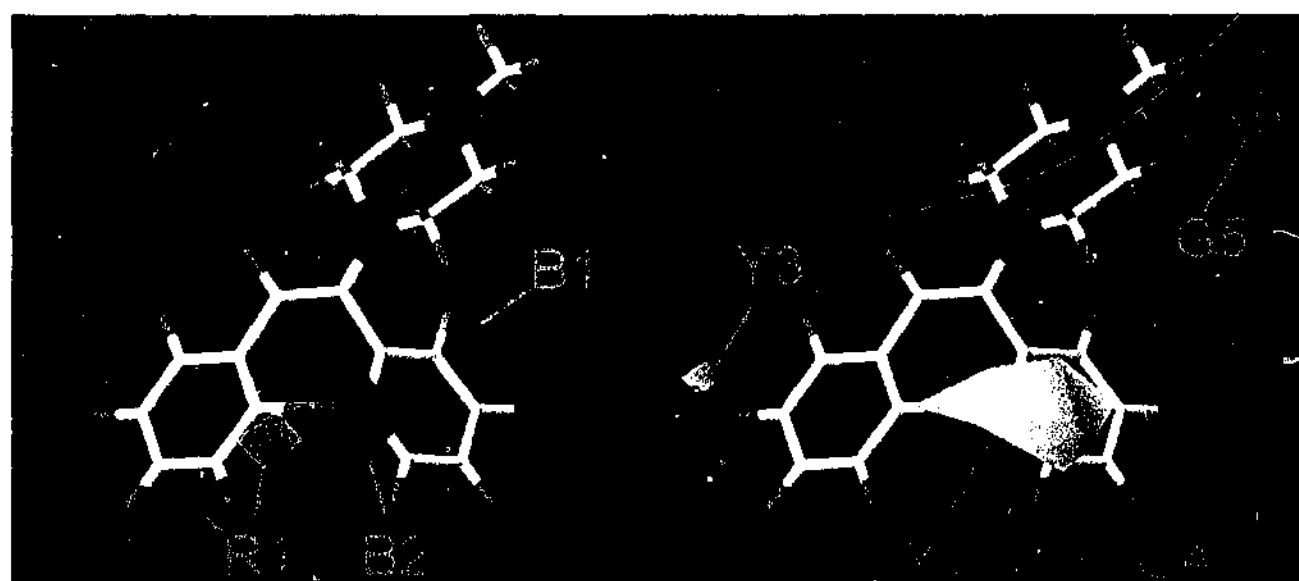
Figure 2.16. CoMFA fields showing electrostatic (left) and steric (right) contributions from analysis of the D_4 binding affinities.

The electrostatic map, Figure 2.16, also shows two regions where a partial positive charge enhances activity at the D_4 receptor at B1 and B2. Both these areas arise from the presence of a chlorine substituent in position 2 of ring A. The chlorine substituent in position 2 has the effect of altering the charge distribution around ring A, resulting in partial positive charges on carbons in regions B1 and B2. This was shown earlier when looking at the D_2 electrostatic fields and is easily evident when viewing Figure 2.15. The electrostatic map, Figure 2.16, shows two regions where partial negative charge enhances activity (red), both labelled R1. These regions also seem to be an artefact of the chlorine in position 8 altering charge distribution on ring B, indicating that chlorine here is detrimental to D_4 activity. However when comparing compounds 1A and 1C in Table 2.1 the opposite is seen, although the predicted affinities from the CoMFA model do indeed predict 1C to have greater activity than 1A. This indicates that D_4 activity predictions for compounds containing electron withdrawing substituents at position 8 on ring B may be prone to error.

The steric map, Figure 2.16, shows two yellow regions indicating areas where steric bulk hinders binding affinity at region Y1 and Y3. Region Y1 is due to unfavourable interactions that are occurring with bulky alkyl groups. Region Y3 is indicating that large substituents present, such as chlorine, are reducing binding affinity. This region at Y3 is similar in priciple to the electrostatic region R1, in that our D_4 model is predicting compounds with a chloro substituent at position 8 of ring B poorly. Indeed the model is predicting 1C and 3J to have greater activity than 1A and 3H, respectively, when this is not the case. Thus it appears as though predictions for compounds containing large substituents at position 8 on ring B are prone to error.

The steric map, Figure 2.16, also shows two regions (green) where extra bulk is favourable at G4 and G5. The region G5 shows the benefits of an ethyl as opposed to a methyl substituent (see compound P5 compared to P6 in Table 2.2), however it should be noted that further extensions to propylene derivatives do marginally reduce binding affinities. The second area of favourable steric interaction is in the plane of the tricyclic structure labelled G4. The favourable steric interactions here are more evident when comparing compounds 4B(r) and 4B(s), where the methyl group is in the plane of the tricyclic structure for 4B(r) resulting in 2-fold greater affinity.

The high degree of similarity between the interpretations of the steric and electrostatic fields for D_4 and D_2 is not surprising, as the correlation between the binding affinities for the two sets of compounds is $r = 0.85$.

5HT$_{2A}$ Serotonergic model

The starting orientation of the database of compounds gave the best model with the highest q^2, r^2 and the lowest standard error of prediction. The resultant model had 5 components giving a q^2 of 0.70 and a r^2 of 0.952; the relative contributions to the model were 51% and 49% from the steric and electrostatic fields, respectively. The predicted activities of all compounds were within ± 0.6 pK$_i$ log units of their measured values.
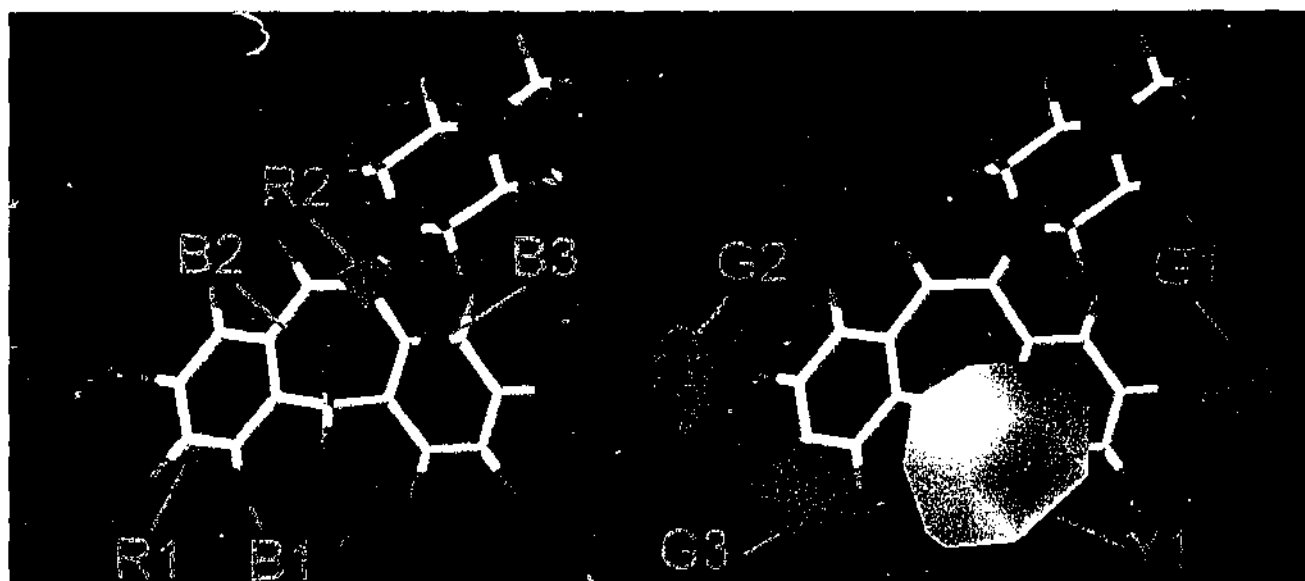
Figure 2.17. *CoMFA fields showing electrostatic (left) and steric (right) contributions from analysis of the 5HT$_{2A}$ binding affinities.*

The electrostatic map, Figure 2.17, shows three regions where a partial positive charge enhances activity at the 5HT$_{2A}$ receptor. In region B1, demonstrating the benefits of hydrogen in this position, see compound 3J compared to 8. At B2 there is a small region where positive charge enhances activity which is due to modifications in the charge distribution on ring B from chlorine in position 8 having a beneficial effect to binding. Similarly there is a small region at R1, where partial negative charge enhances activity, also due to an 8-chloro substituent modifying charge distribution. The third region where a partial positive charge enhances activity is B3 located on ring A. Region B3 is due to the effect of nitrogen in ring A, as in compounds 13-16, changing the

charge distribution around the ring and imparting a greater negative charge in this area, which is detrimental to binding affinity. This modification of charge distribution by nitrogen is also partly the cause for the favourable electronegative region where the piperidine ring joins the tricyclic substructure. Another contribution to this favourable electronegative region is the absence of a nitrogen in position A, as a nitrogen would withdraw electrons imparting a partial positive charge which is detrimental to activity (see compound 1E compared to 3H).

The steric map (see Figure 2.17), shows three distinct regions where activity is enhanced by larger substituents. In regions G1 and G2, showing the benefits of a chloro substituent in these positions, and in region G3 where the hydrogen interacts. The region G3, similar to B1, is due to the effect of replacing the carbon with nitrogen, which reduces binding affinity at the $5HT_{2A}$ receptor. The detrimental effects of bulky alkyl substituents in region Y1 are again evident.

2.4 CONCLUSIONS: TRICYCLIC MODEL

We developed a pharmacophore model that shows how (R)- and (S)-octoclothepin can fit together using superimpositioning of the points of interaction of the compounds. This was then used to develop a CoMFA model. A number of parameters were explored in the construction of CoMFA models. It was shown that by using the field fitting method for compound alignment, models with a higher predictive ability could be gained than models obtained by the template alignment method. The weightings associated with the field fit method were also explored, and it was shown that an overall weight term of 10 resulted in the most predictive model. Analysis of different charged states and methods for charge generation of our compounds showed that ESP charges gave the most predictive resultant CoMFA model, which was consistent for both charged states. No additional insights were gained by performing CoMFA on neutral compounds, thus the protonation state most likely under physiological condition, +1, was used. Rotating the compounds within the CoMFA grid space gave a normal distribution for the leave one out (LOO) cross-validation, demonstrating that the predictive ability can be improved by doing such an analysis. By exploring such parameters in the generation of CoMFA models, optimum models were generated that provide an excellent framework to aid in the design of new antipsychotics.

2.4.1 Designing Compounds

Extensions of the themes produced by the D_2 model could be undertaken to increase or decrease binding affinity so as to attain a binding profile similar to clozapine. Adding larger more electronegative substituents in region $G1_{D2}$, ensuring there are no heteroatoms in aromatic ring B could increase binding affinity. In turn D_2 binding affinity could be reduced by bulky alkyl substituents occupying $Y1_{D2}$ or the presence of nitrogen in position A. More novel methods of altering the binding affinity could also . be achieved by modifying the type of aromatic ring present as ring A, to change the electronic environment. This modification of ring A has already been carried out by Liégeois et al[49] with compounds showing appreciable D_2 affinity and excellent CoMFA models being generated from the series. Unfortunately the CoMFA models were generated using IC_{50} values and could not be used in our model. This

modification of ring A has also been carried out with much success resulting in the atypical antipsychotic olanzapine.

Extensions of the themes produced by the D_4 model are similar to those of the D_2 model, although no compounds from Table 2.3 were included in this analysis. Therefore the effect on D_4 binding affinity upon removal or addition of heteroatoms in rings A and B is unknown. A number of additional steric themes are present in the D_4 model, which suggest slightly larger substituents may be added at position 4 of ring A to occupy region $G4_{D4}$. Indeed if a chlorine or similarly large electronegative substituent were added at position 4, a similar electronic distribution around ring A to substitution at position 2 would be maintained whilst occupying this additional area. There is also scope for increasing D_4 binding affinity by adding slightly larger substituents at position Q to occupy region $G5_{D4}$, indeed one could expand upon this theme to bicyclodecane compounds.

Again there are similarities from the themes extending from the $5HT_{2A}$ model to those shown in the D_2 model, with the addition of larger more electronegative substituents occupying region $G1_{5HT2A}$ and the removal of heteroatoms in the aromatic rings increasing binding affinity. However the $5HT_{2A}$ model introduces a number of additional themes with the presence of larger more electronegative substituents at position 8 of ring B occupying region $G2_{5HT2A}$ and also increases binding affinity.

The results of these CoMFA studies also suggest that a nitrogen substituent in position Z is not essential for a comparable profile to that of clozapine. A consequence of removal of this nitrogen could be the potential to diminish clozapine's major clinical side effect, since there is evidence that formation of a nitrenium ion at this position is involved in the mechanism of agranulocytosis[50].

Clozapine medications usually require a high dose regimen[51] compared with typical antipsychotic drugs. By manipulating the balance between electronic and steric requirements at different positions in accordance with this CoMFA analysis, it may be possible to develop a clozapine analogue with optimal bioavailability at reduced dose. In turn this may have the advantage of reducing side effects through increased binding

with the correct receptors and decreased binding with those receptors causing side effects.

Overall, this work will ideally lead to the development of novel antipsychotics that have greater efficacy against treatment-resistant schizophrenia and would be devoid of some of clozapine's clinically limiting side effects.

2.5 EXTENDED MODEL INTRODUCTION

The other main groups of antipsychotics are the elongated or extended structures. These APDs usually consist of an aromatic group connected to a protonated nitrogen and followed by a further aromatic group (see Figure 2.18). There are many structural classes of extended APDs ranging from butyrophenones and benzamides to heterocyclylpiperidines.
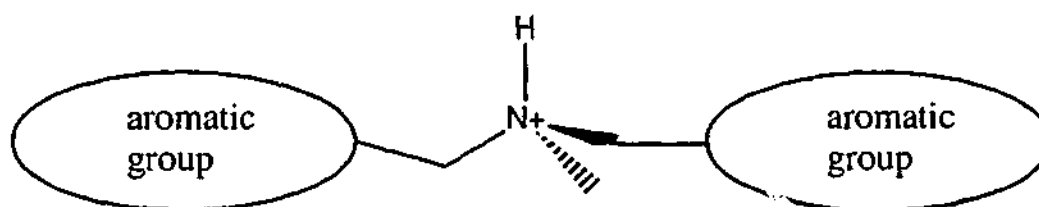
Figure 2.18. *Simplified representation of extended or elongated antipsychotic structure.*

These types of compounds have received much attention during the last decade, as they provide additional areas in which to explore conformational space in the hope of designing compounds with greater selectivity and specificity towards various receptors implicated in schizophrenia. Much of the impetus behind this work has stemmed from the reintroduction of the antipsychotic drug clozapine that exhibits a low incidence of EPS[3], which has been proposed to be due to its high selectivity for the D_4 receptor compared to the D_2 receptor. Therefore an antipsychotic targeting the D_4 receptor should theoretically show efficacy and be devoid of EPS.

Many drug companies have explored this approach during the 1990's, shown by the recent introduction and development of several structure activity relationships (SARs) for D_4 antagonists[52-67].

2.6 METHOD

2.6.1 Extended Pharmacophore

At the commencement of this investigation we were unable to find a D_4 pharmacophore model which would amass all this data. Therefore we developed a pharmacophore based on published data sets in several papers on D_4 antagonism[52,53,55,56]. These were chosen on the basis of binding affinity and close structural similarity. The main difficulties in the development of reliable alignments is not only ligand flexibility but also the ambiguity of postulated receptor-ligand interactions and the possibility of more than one binding site or binding mode. Therefore we preferred to obtain a model based on a set of congeners that acted in a similar fashion, rather than trying to fit compounds that were too structurally dissimilar.

The pharmacophore was designed using tight binding D_2 and D_4 receptor ligands as well as ligands exhibiting intramolecular hydrogen bonding or that were conformationally constrained (see Table 2.9 and Figure 2.19).

Table 2.9. Binding affinities of compounds used in pharmacophore elucidation.

Compound	D_4 binding affinity (nM)	D_2 binding affinity (nM)
1 (Moore #4)	0.39	6.1
2 (nemonapride)	0.21	0.16
3 (Unangst #5)	1.5	436
4 (Rowley #28)	930	690
5 (Rowley #29)	44	160
6 (Rowley #30)	3.6	130
7 (Carling #8)	0.6	56
8 (Rowley #32)	1.6	200
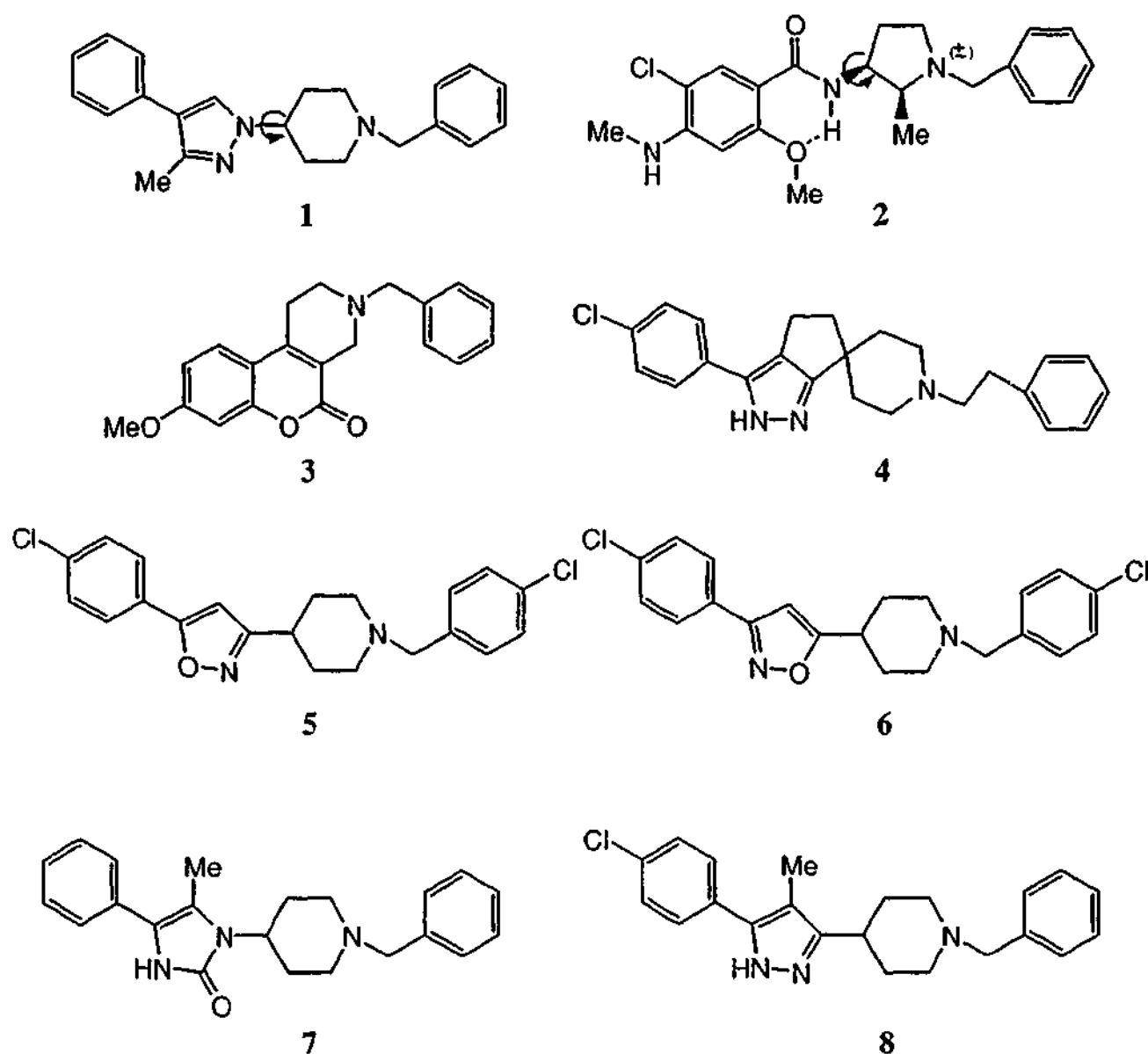
Figure 2.19. *Ligand structures used to determine pharmacophore.*

2.6.1.1 Conformational Analysis

As most compounds were not available in the Cambridge Crystallographic Database[11] (CCDC) they were built in Sybyl[13] by combining and altering standard fragments using standard geometries. Compounds were initially built in an extended conformation, then minimised using the MMFF94s force field, MMFF94 charges and the Powell minimisation method for 1000 iterations, or until a termination gradient of 0.01 kcal/(molÅ) was reached.

Most of the compounds containing nitrogens are expected to be protonated at physiological pH and as this can influence conformation, they were modelled in the protonated state. The pK_a value of each ligand used in defining the pharmacophore was

calculated using ACD pK$_a$ calculator[20], and then their percent protonation at physiological pH was determined (Table 2.11).

Some compounds show a strong probability for intramolecular hydrogen bonding. Examples are all compounds in the Ohmori set and some of the Carling data set. The influence of intramolecular hydrogen bonding must be considered when conformational analysis is undertaken. Therefore a search of the CCDC for compounds with intramolecular hydrogen bonding similar to that of the ligands used was undertaken to obtain probable conformations.

1 and 2 (see Figure 2.18) were subjected to conformational analysis using a systematic search around the bonds shown in Figure 2.18. The Tripos force field[68] was used with van der Waals radius scale factors set at 0.65, charges were assigned by the Gasteiger and Marselli method and two degree increments were used to locate positions of local and global minima. The lowest energy conformations within each of the local minima of the potential energy surface were identified as potential binding conformations. Mopac[69,70] calculations were done in five-degree increments around the same bonds used in the systematic searches of 1 to verify energy differences or similarities from those obtained by the molecular mechanics systematic search.

Simulated annealing for ten cycles, starting at 1000 degrees using default parameters for the base temperature (200 °C) and cooling rate (exponential decay), was used to generate a number of structures. These were minimised within Sybyl to reveal the full range of conformations of the protonated pyrrole ring of 2 and other possible conformations of the semi-rigid chromenone, 3.

The large number of possibilities that the benzyl substituents attached to the tertiary nitrogen of piperidine or pyrrole (see Figure 2.19) could adopt was explored via a conformational energy map constructed to identify any minima. The conformational minima were compared to relevant crystal structures from the CCDC, and those shown in Figure 2 of the paper by Rowley *et al*.

A search of the CCDC for compounds having a phenyl ring joined to a pyrazole ring was undertaken to check the relative orientation of the aromatic rings. The range of

conformations generated from conformational searching for each compound was reduced further by comparing them to compounds **3** to **6**. These latter compounds provide a filter for the conformational space due to the presence of additional structural moieties (fused tricycle, **3**; spiro ring system, **4**; isoxazoles in **5** and **6**).

2.6.1.2 Superimposition: Defining the Pharmacophore

To develop a possible pharmacophore for D_4 antagonist interaction, the different conformations generated from the above analyses, were compared by superimposition. Dummy atoms, used to represent points of likely receptor interaction (as shown in Figure 2.20 for **1**) were used to study superimpositions. Superimpositions were undertaken using the centroids of the aromatic rings in question. An additional dummy atom was built 2.8 Å from the distal nitrogen of the piperidine ring, along the vector of the ammonium hydrogen, to represent a proton donor or acceptor interaction[19]. Another additional dummy atom was built 2.8 Å from the nitrogen of the pyrazole ring, along the vector of the lone pair for appropriate ligands. For ligands with a 2-methoxybenzamide ring structure, the dummy atom was built 2.8Å from the oxygen along a vector angle of 128°[71], eight degrees away from the direction of the idealised sp^2 lone pair. If only one hydrogen bonding interaction with oxygen is envisaged, it may still be further away from the direction of the idealised sp^2 lone pair[72]. Therefore additional dummy atoms were built along a vector in further 10° increments (see Figure 2.21), to assess whether the pharmacophore could be refined. The dummy atom for the chromen-2-one moiety of **3** was built 2.8Å along the vector of the lone pair, from the sp^3 oxygen involved in the ring structure. All dummy atoms used in the superimpositioning are postulated to represent proton donor or acceptor interactions.
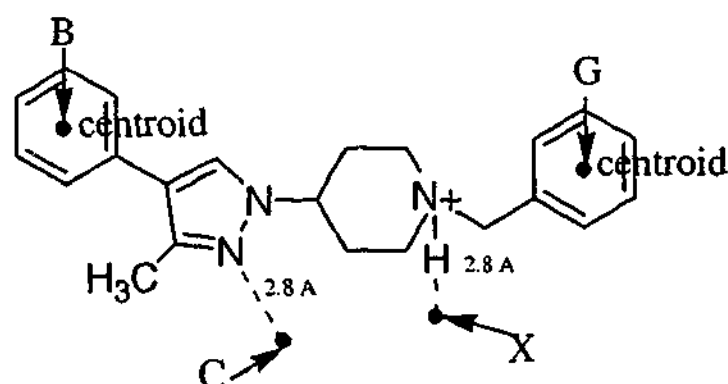
Figure 2.20. Dummy atoms, B, C, X and G used in the superimposition studies, are represented by solid circles. Dummy atoms B and G are at the benzene ring centroids.
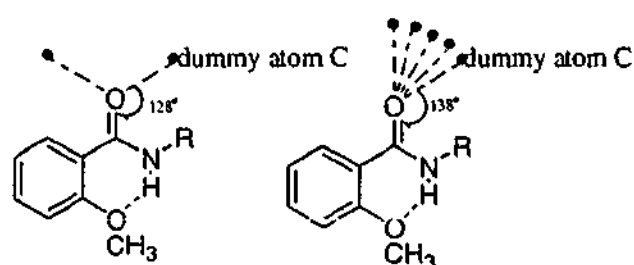
Figure 2.21. Angular variation of dummy atoms at oxygen illustrated for 2.

The weightings of the dummy atoms for superimpositioning were adjusted so that greatest weight was given to dummy atoms attached to the ring nitrogens, in particular dummy atom X that is essential for specific ligand binding. This was done to emphasise these interactions that are more site specific compared with those representing hydrophobic interactions of aromatic rings (see Table 2.10).

Table 2.10. Dummy atom weightings.

Ligand	dummy B	dummy C	dummy X	dummy G
1	0.3	0.7	1	0.3
2	0.3	0.7	1	0.3
3	**	0.5	1	0.5

** indicates that there is no ring near this position to create a centroid.

2.6.1.3 Pharmacophore Validation using GASP

Genetic algorithm similarity program (GASP)[73,74] is a module available within the Sybyl software suite that uses a genetic algorithm to discover the correspondence between functional groups in different molecules and the alignment of these groups in a

common geometry. GASP was run on the D_4 antagonists **1, 2, 7**, and **8** in both ring flexible and rigid modes to assess the possible binding conformations of these compounds.

2.6.2 Extended Molecular Field Analysis

In this section Comparative Molecular Field Analysis (CoMFA) is used to develop a pseudoreceptor model for D_2, and D_4 antagonist activity using published data from a series of extended structures[52,53,55,56], in an attempt to elucidate some aspects of selectivity and specificity within the D_2 and D_4 receptors.

CoMFA was carried out on the series of extended analogues given in Tables Moore, Rowley, Carling, Ohmori, Kesten, Sanner, Belliotti and Thurkauf (Appendix) using the published receptor binding affinities.

2.6.2.1 Molecular Field Analysis

QSAR were created for all the data sets developed in the following sections using the method outlined earlier.

The values for CoMFA used were: a grid spacing of 1.5Å, an sp^3 carbon probe with a +1 charge, box smoothing and standard CoMFA scaling. CoMFA electrostatic and steric fields were generated with a grid spacing of 1.5Å. Partial Least Squares (PLS) analysis was carried out on each database. Leave one out (LOO) cross-validation was used to select the number of principal components for cross-validated statistics with column filtering set at 2.0 kcal/mol. The final CoMFA model was generated using no cross-validation and the number of components suggested by the LOO validation run[25].

Compounds were aligned using the *fieldfit* method, which minimises the compound in question into the steric and electrostatic fields of the tightest binding ligand. The overall weight term, SFF value, for this procedure was set at 10. This was followed by a quick minimisation for 50 iterations to obtain the final alignment.

Earlier analysis of different charged states and methods for charge generation of our compounds, showed that ESP charges gave the best resultant CoMFA models for both charged states. No additional insights were gained by performing CoMFA on neutral compounds, thus the protonation state most likely under physiological condition, +1, was used. Rotation of the compounds within the CoMFA grid was also performed to obtain optimum final models.

2.7 RESULTS AND DISCUSSION

2.7.1 Extended Pharmacophore

2.7.1.1 Conformational Analysis

The results for the computed protonated states of the ligands used in the development are shown in Table 2.11, which emphasises the importance of modelling ligands in a protonated state as five of the six ligands are predominantly protonated at physiological pH. The exception is 3 whose pK_a is dubious, particularly since the pK_a of mono-unsaturated piperidines, a part of 3, have pK_a values of the order of 7.75, giving approximately 70% protonation at pH 7.4.

LogD (Table 2.11), the octanol/water partition coefficient at a selected pH which is a measure of a compound's permeability or lipophilicity[75], was calculated using the logD calculator from ACD. LogD gives an indication as to absorption or permeation of the blood brain barrier (BBB), with compounds having a log D greater than five expecting poor absorption or permeation[76].

Table 2.11. Characteristics of Ligands Used for Defining Proposed Pharmacophore

Ligand	K_i(nM)	pK_a (pyrrole/piperidine)	%protonation at pH=7.4	logP	logD (pH=7.4)
1	0.39	8.06	82	4.96	4.21
2	0.21	8.3	89	3.66	2.72
3	1.5	5.91	3	4.30	4.28
4	930	8.75	96	6.65	4.92
5	44	7.69	66	5.58	5.11
6	3.6	7.83	73	5.27	4.70

pK_a, logP and logD calculated using ACD labs version 3

A survey of the CCDC for compounds containing the 2-methoxybenzamide structure[77-90] without additional substituents, showed an almost planar conformation of the hydrogen bonded six membered rings in 100% of these compounds. Therefore planar conformations were used in the generation of conformers for further work. However it

should be noted that 2-methoxybenzamides may not always be planar in solution and this must be kept in mind during the analysis.

The molecular mechanics systematic search conducted on **1** generated three main families: *gauche -ve, gauche +ve* and *antiperiplanar*, shown in Figure 2.22. Mopac systematic search calculations on **1** gave a similar profile although only relative energies are shown in Figure 2.22. The global minimum from Mopac calculations was found to be located at a torsion angle of 240 degrees, corresponding to the *periplanar conformer*. This gave four main starting conformations considered as possible binding modes, shown in Figure 2.23.
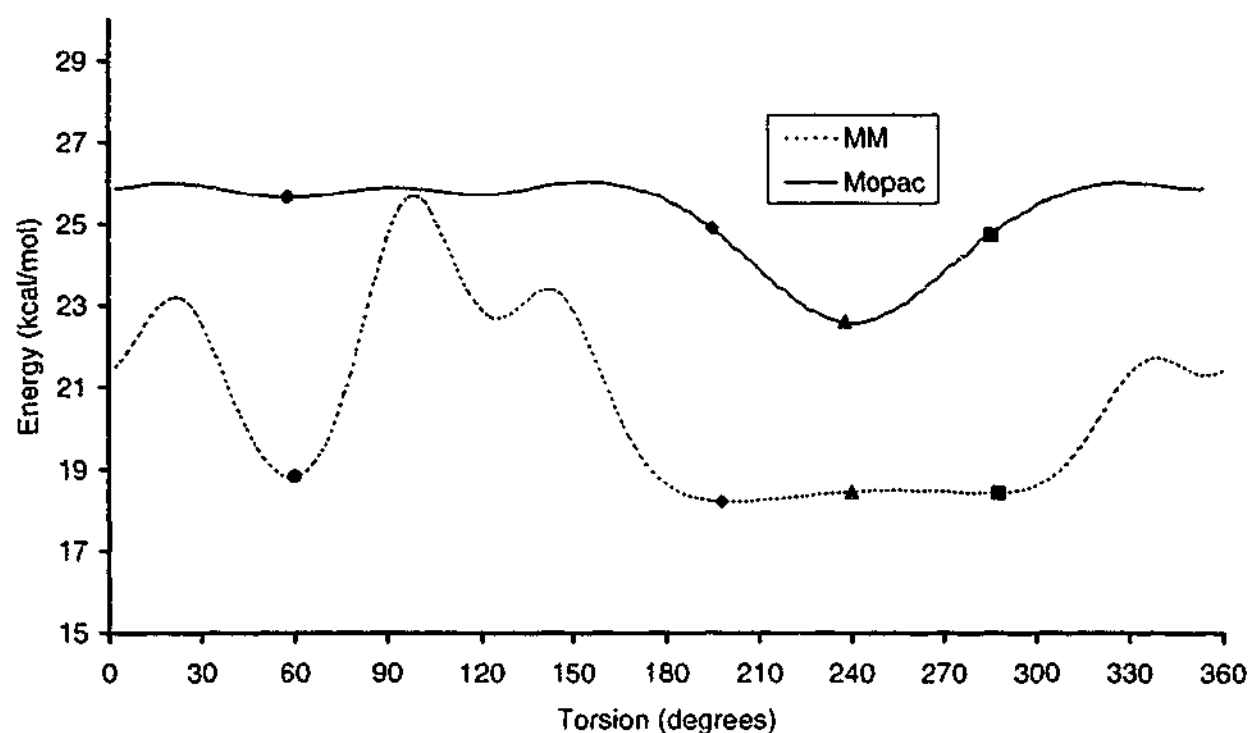
Figure 2.22. Systematic search of 1. Circles represent the antiperiplanar conformer, diamonds represent the gauche -ve conformer, triangles represent the periplanar conformer and squares represents the gauche +ve conformer.

dummy X   dummy X   dummy X   dummy X
                 dummy C

dummy C   dummy C

gauche +ve   periplanar   antiperiplanar   gauche -ve
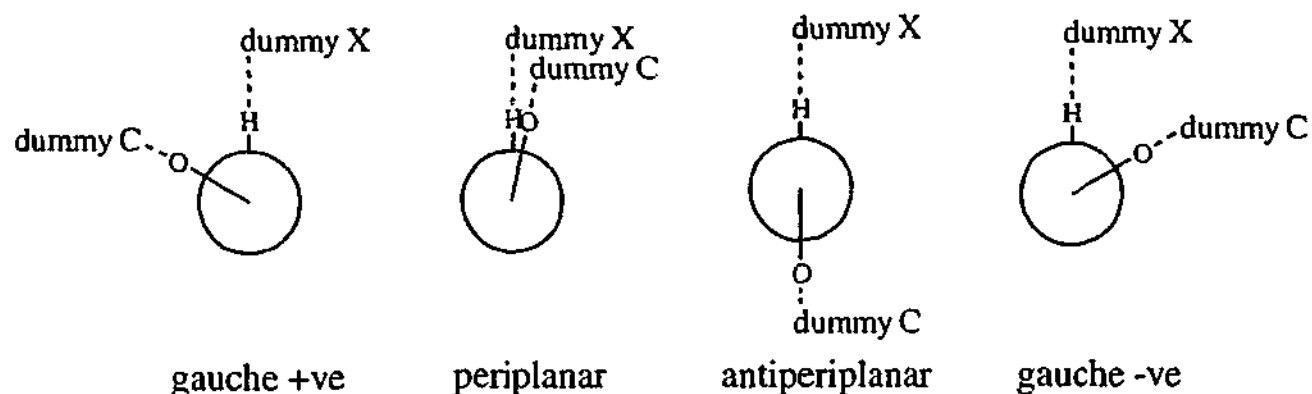
Figure 2.23. Newman-like projections of four initial binding modes.

Table 2.12. Relative energies of **1** conformations with different force fields and solvents

Force Field (solvent)	ΔE	Conformations	(kcal/mol)	
	gauche -ve	gauche +ve	periplanar	antiperiplanar
Tripos Systematic Search	0	0.21	0.25	0.75
MM94S	2.07*	2.21*	0	3.45
Amsol (vacuum)	0[†]	0.01[†]	0.02	2.76
Amsol (water)	2.16[†]	2.14[†]	2.13	0
Amsol (alkane)	0.51[†]	0.51[†]	0.51	0

* indicates that the protonated nitrogen and the pyrrole ring nitrogen were held as aggregates due to significant conformational changes,

[†] indicates that the compounds adopted a periplanar conformation due to minimisation.

A study of the relative energies of these four starting conformers of **1**, Table 2.12, was undertaken. The different force fields and solvents show that all possible binding modes are energetically feasible. In the presence of a solvent, the antiperiplanar conformer appears energetically more favourable as a possible binding mode; however the energy differences between all conformations are minimal. The gauche -ve and gauche +ve conformations, despite being minimised into the periplanar conformation by Amsol, are less than 3kcal/mol higher in energy than this conformation.

For the simulated annealing of the protonated pyrrole ring of **2** we used the four structures shown in Figure 2.24 as separate starting conformations This is because the simulated annealing process within Sybyl is known to have problems in modelling the inversion of the nitrogen atom[91].
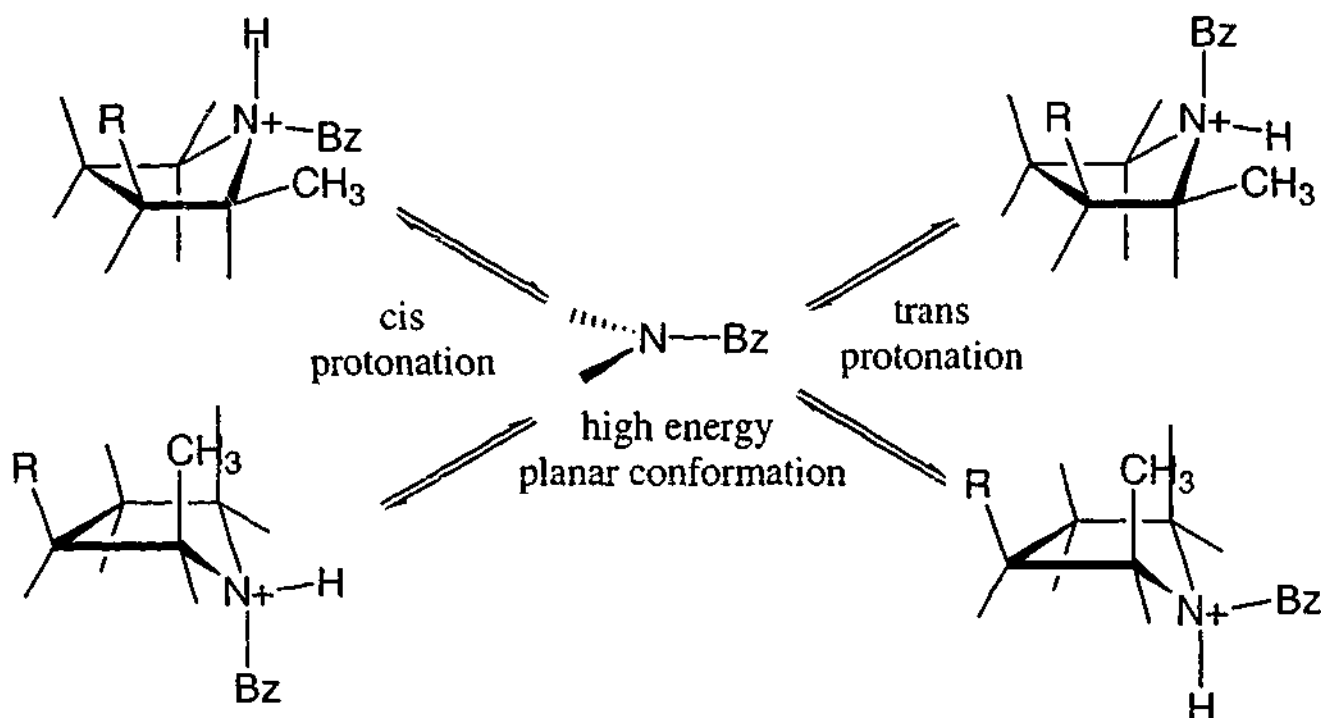
Figure 2.24. *Starting conformations for simulated annealing. In each case R represents 2-methoxy benzamide.*

When the nitrogen in the pyrrole ring is protonated the hydrogen can attack from either the *cis* or *trans* side (see Figure 2.24), with respect to the methyl substituent at position 2 on the ring. Although attack is expected from both sides of the ring, attack from the *cis* side would be a lot more unfavourable due to steric interaction between both the methyl and amide substituents present in positions 2 and 3 respectively.

Simulated annealing of a fragment of **2** containing the protonated pyrrole ring and benzyl, methyl and amide substituents followed by minimization, gave five significantly different envelope conformations, with the benzyl substituent in an equatorial position. All conformations generated with the benzyl group in an axial position were over five kcal/mol higher in energy than the highest energy equatorial benzyl conformation that was considered, thus they were not considered for further conformational analyses.

The fragment with the lowest energy of all the fragments considered, by about four kcal/mol, due to another intramolecular hydrogen bond forming a seven membered hydrogen bonded ring is shown in the extended form in Figure 2.25. This conformation of 2 was not considered as a possible binding mode as the protonated nitrogen is inaccessible for binding at a postulated receptor binding site, and also didn't fit with any of the existing binding modes from 1.
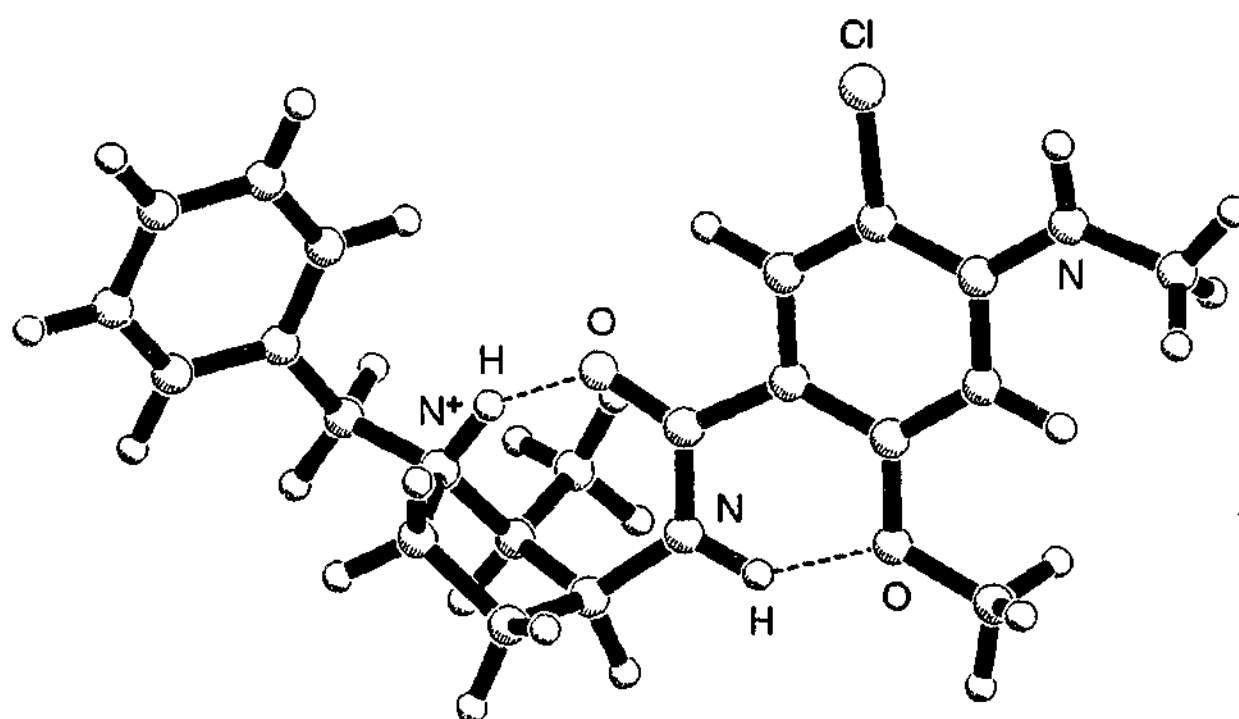
Figure 2.25.   Lowest MM energy conformation of 2.

As 2 is such a tight binding ligand it is expected to adopt a low energy conformation, which would have to be an extension of one of the remaining four low energy fragments elucidated from simulated annealing.

Comparing distances between important pharmacophoric elements from systematic searches of the four remaining low energy fragments of 2 with those from 1, revealed that only one of the fragments could account for all possible binding conformations. Therefore the extension of this fragment, Figure 2.26, was used in further conformational analyses.
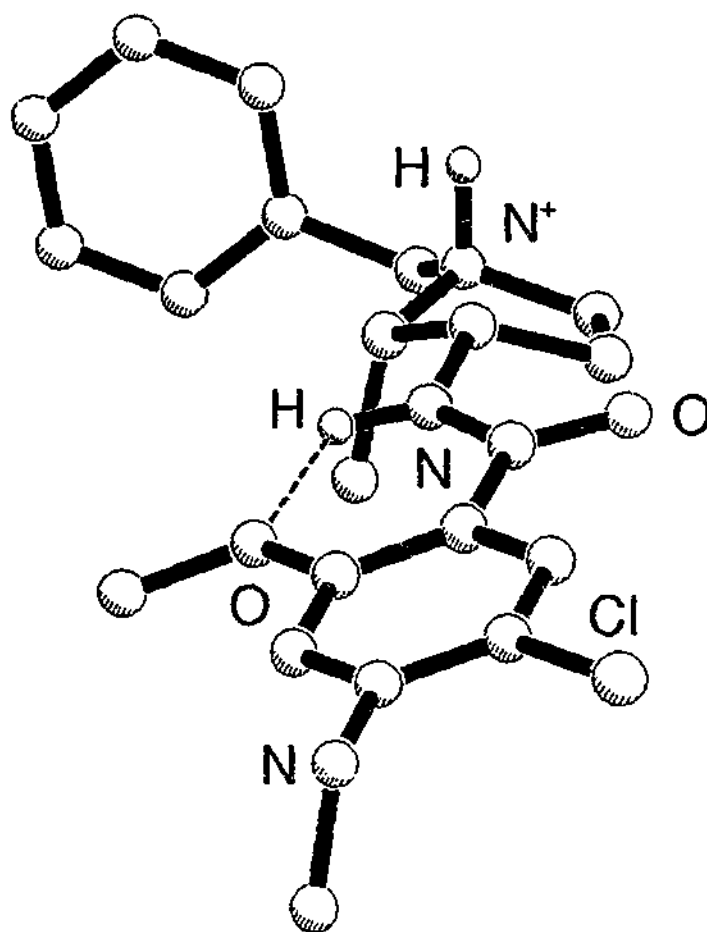
Figure 2.26. *3-D structure of 2, showing relevant hydrogens only.*

Table 2.13. Relative energies of 2 conformations with different force fields and solvents

Force Field (solvent)	ΔE	Conformations	(kcal/mol)	
	gauche -ve	gauche +ve	periplanar	antiperiplanar
Tripos Systematic Search	0.67	0	2.97	11.67
MM94S	5.23	0	7.01*	5.19*
Amsol (vacuum)	13.76	10.08	5.22	0**
Amsol (water)	0	2.52	3.79	11.95
Amsol (alkane)	0	4.55	4.71	7.43

* indicates that protonated nitrogen and amide group were held as aggregates due to significant conformational changes,

** indicates significantly different conformation due to minimisation.

Again a systematic study of the relative energies of the four starting conformers of 2, Table 2.13, was undertaken. The different force fields and solvents show that all possible conformations generated are energetically feasible, with the *gauche -ve* and *gauche +ve* seeming to be the best candidates as binding modes in four out of five cases.

A conformational potential energy surface was constructed for the benzyl substituent on 1, to give an idea as to its likely position.
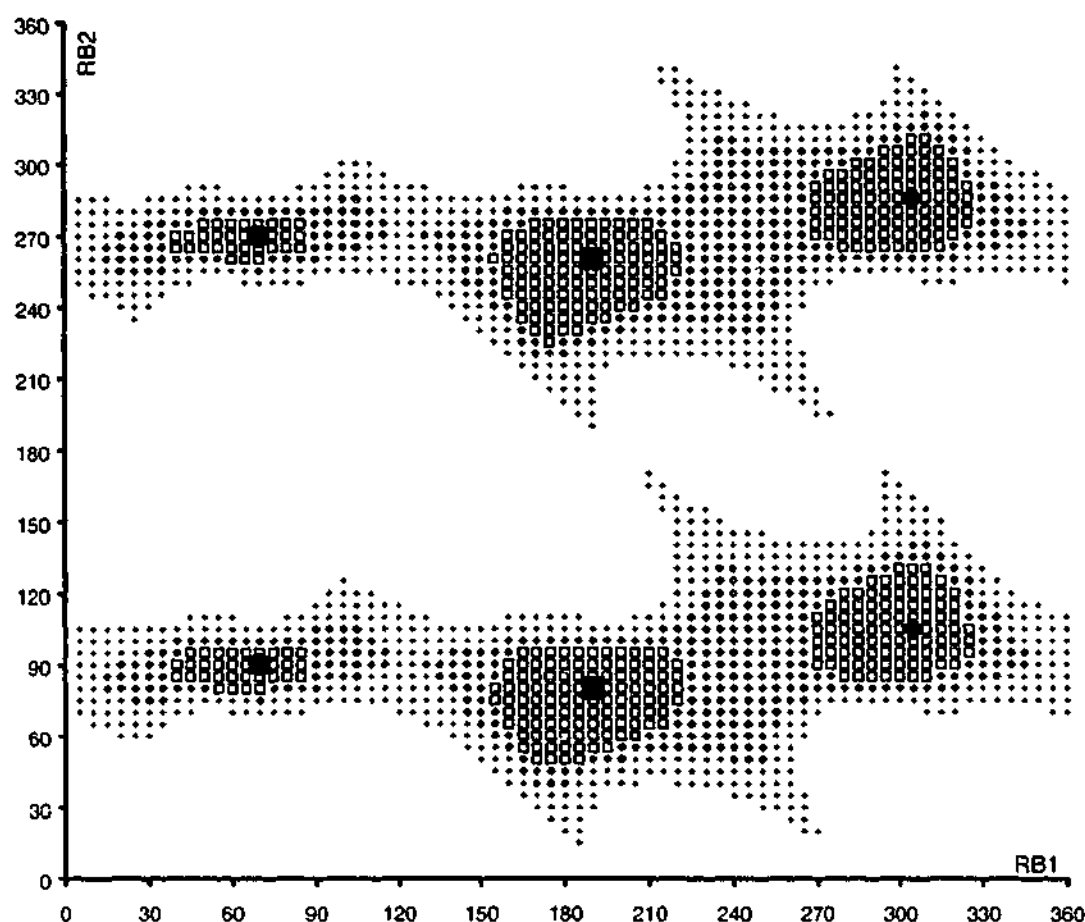
Figure 2.27. *Conformational potential energy surface. Squares, diamonds and circles indicate the position of the global and local minima, respectively, of the benzyl substituent of 1. RB1 and RB2 represent the two rotatable bonds joining the benzyl substituent.*

The conformational potential energy surface map (Figure 2.27) shows that there are six main minima. However because a 180° rotation of the phenyl ring gives the same result, there are only three distinct minima. The conformers at (190°, 80°/260°) representing the global minima are 0.33 and 0.96 kcal/mol lower in energy than the conformers at (305°, 105°/285°) and (70°, 90°/270°) respectively.

A similar conformational potential energy surface map for the benzyl substituent of 2 (not shown) gave two minima, corresponding to those at (190°, 80°/260°) and (305°, 105°/285°) in Figure 2.27. The conformation represented by (190°, 80°/260°) in Figure 2.27 is in agreement with the crystal structures of conformations presented in the paper by Rowley *et al* and the crystal conformers[88,92-95] from the CCDC. The conformation represented by (305°, 105°/285°) in Figure 2.27 is in agreement with the crystal conformers[89,95,96] from the CCDC. Therefore both low energy arrangements of the

benzyl ring seem possible as a binding mode. The global minimum conformation represented by (190°, 80°/260°) was used in further analyses.

A search of the CCDC for compounds with a phenyl pyrrole moiety, showed that these aromatic moieties were not coplanar. A number of structures[97-103] were found with the torsion angle ABCD (see Figure 2.28) varying between ±(13 to 77) degrees, with an absolute average of 48°. Therefore the most energetically favourable positive and negative torsion angles obtained from Mopac minimisations of ligands with the phenyl pyrrole moiety were chosen as the two initial conformations.
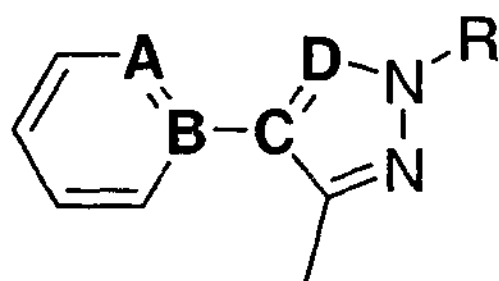
Figure 2.28. Phenyl/Pyrrole Torsion Angle

2.7.1.2 Superimposition: Defining the Pharmacophore

In order to reduce the number of possible conformations the phenyl ring could adopt, 1 and 2 were compared. With its intramolecularly constrained conformation, 2 gave some idea as to the position of the plane of the phenyl ring. The best fit with the 2 conformer was seen when the torsion angle ABCD (see Figure 2.28) was in the negative orientation for the *gauche +ve conformer*, with a RMS fit of 0.691Å, and in the positive orientation for the *gauche -ve conformer*, with a RMS fit of 0.653Å. Therefore these torsions were used for further modelling. The fits for the *periplanar* and *antiperiplanar conformers* were not improved by altering this torsion angle, thus taking away some significance from these as possible binding modes.

The *periplanar conformer* already shown to be less likely as a binding mode due to the displacement of the phenyl planes from differing compounds was eliminated by comparing 4 to the existing *periplanar* model. Compound 4 (see Figure 2.19) which is a conformationally constrained molecule due to its spiro ring system, has the nitrogen

lone pair pointing directly upwards parallel with the protonated piperidine hydrogen, as in the *periplanar conformation*. The binding affinity of 930 nM for **4** indicates that this is not a favourable conformation for binding.

Therefore in order to accommodate all the ligands analysed so far, only the *gauche -ve*, *gauche +ve* and *antiperiplanar* conformations could be possible.

To further reduce the conformational space, conformational searching was performed on the chromenone system. Simulated annealing of **3** gave two possible conformations for binding that fitted with the existing models. The two possible conformations were the *gauche +ve* (see Figure 2.29) and *gauche -ve conformer* models. The *antiperiplanar conformer* model was unable to satisfactorily accommodate any conformation of **3** without the chromenone plane being too far displaced from the existing planes. Thus if compound **3** is binding in the same site as compounds **1** and **2** it is unlikely that the *antiperiplanar conformer* is a binding mode. Therefore from analyses so far the *gauche +ve* and *gauche -ve conformer* models seem most viable from the original possibilities. However when superimposing compound **3** onto these models the lone pair from the oxygen included in the ring of **3** is not pointing to the same location in space as that of the existing *gauche +ve* and *gauche -ve* models and is separated by a distance of 1.25Å from the lone pairs of the model. Therefore a hydrogen bond from this region would not be as strong unless the binding site is capable of some movement. However when **3** is translated to the left or right, so as to align the protonated nitrogen, the distance between the dummy atoms from the nitrogen on the pyrrole ring and the ring oxygen in the chromenone was found to be approximately 2.1Å (see Figure 2.11). This distance of 2.1Å is consistent with rotation of a carboxylic acid group on the receptor with which the oxygen included in the ring of **3** could interact.
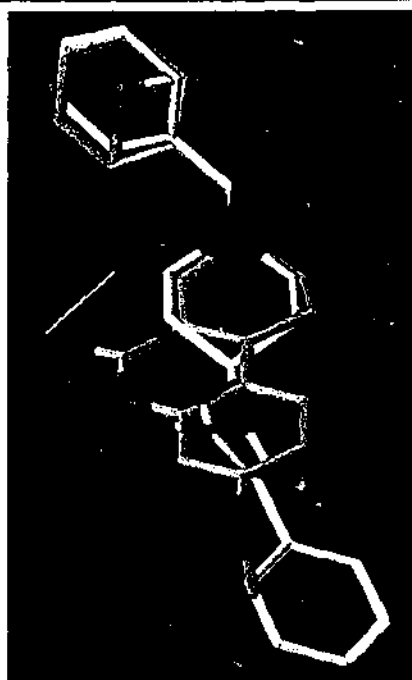
Figure 2.29. *3 (orange) superimposed on 1 (white) in the gauche +ve conformation showing the placement of a possible carboxylic acid interaction site.*

The previous analyses established that if the molecules studied were binding in the same binding site, the most likely binding modes were *gauche -ve* and *gauche +ve conformations*. In order to distinguish which conformer is a more likely candidate for a possible binding mechanism, **5** and **6** were examined. The crystal structures of these compounds[55] show that **5** adopts the *gauche +ve conformer* orientation and **6** an orientation midway between an *antiperiplanar* and *gauche -ve conformer* orientation. As **6** has a binding affinity 10 times greater than that of **5** it can be said that this ligand does not need to undergo such a large energy rotation penalty in comparison to **5**, to achieve the correct orientation. Therefore the *gauche -ve conformer* orientation that it is closer to is more likely to be the biologically active conformer. However, this increased affinity could also be due to a more favourable logD, as **5**'s logD is greater than five indicating poor absorption or permeation of the BBB. **6**'s increased affinity could also be due to better directionality of the lone pairs from the participating oxygen atom or even due to overly strong hydrogen bonding. Although when compared to similar compounds from the same paper (Rowley #35 and Rowley #36) where there is also an oxygen atom present, the binding affinity is only 0.5-fold greater. This suggests that the oxygen atom alone does not account for the 10-fold greater activity exhibited by **6** compared to **5** and that some other properties are at play. In addition to this the most stable hydrogen bond to the isoxazole ring is with the nitrogen, which is consistent with the CSD scatterplot[30] that shows that the majority of H-bonds in crystal structures go

to the ring N, not the ring O. Thus we can tentatively say there is a greater probability that the *gauche -ve conformer* orientation, which is similar to the crystal structure of **6**, is a more likely candidate for binding, unless other factors are at play.

Figure 2.30 shows he models generated here accommodate all the tight binding ligands from the papers mentioned throughout. **1** fits extremely well into the model and has a distance of 3.90Å between dummy atoms B and C, the dummy atom from the pyrrole ring and protonated nitrogen respectively. The distance between dummy atoms B and C for **2** is 3.77Å and the placement of its plane from the phenyl ring and hydrogen bonded six membered ring is marginally askew compared to **1**. This minor discrepancy in plane placement might explain the higher binding affinity for **2** compared to the other compounds analysed

The only compound that did not initially fit into the model was **3**. However after translation to align the protonated nitrogens it was shown to fit well into the model. The directionality of the lone pair from the oxygen atom was roughly parallel with the lone pairs from the other ligands (see Figure 2.31).
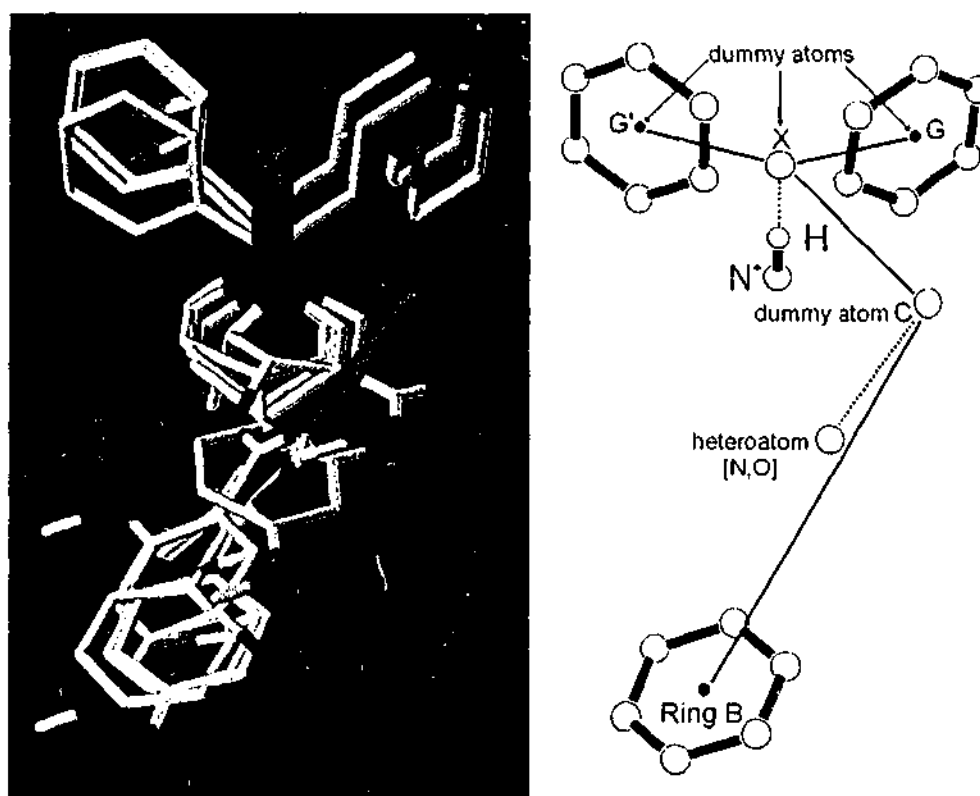
Figure 2.30. *The gauche -ve orientation model, with dummy atoms D and D' corresponding to benzyl conformations square and diamond from Figure 2.27,*

respectively. Distances for proposed pharmacophore: BC = 6.4 ± 0.5Å, BX = 9.4 ± 0.5Å, CX = 3.8 ± 0.1Å, XG = XG' = 4.2 ± 0.3Å
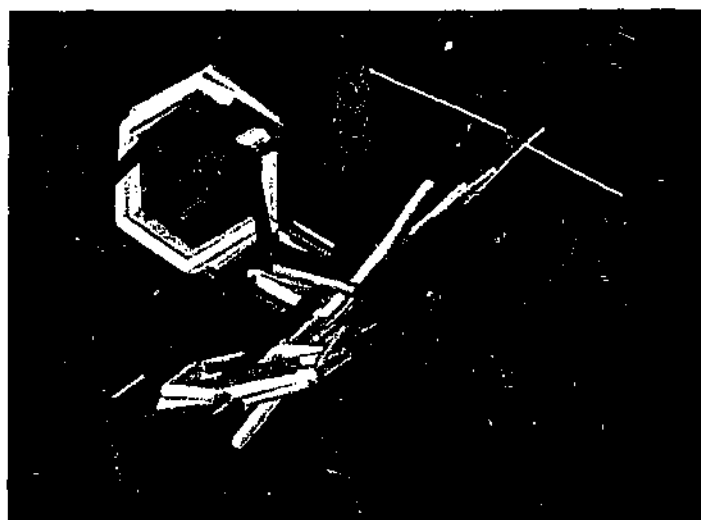
***Figure 2.31.** gauche -ve orientation model, looking along the plane of the rings.*

The model constructed here is subject to some limitations, as all ligands were initially constructed in a low energy extended conformation and this may not be the case *in-vivo*, although the conformational analysis was used to counteract this problem. All modelling was done *in vacuo*, which may not be the case under physiological conditions. However the energies we calculated in different solvents, (Tables 2.12 and 2.13), demonstrate that the proposed conformations are realistic for different media. Use of intramolecular hydrogen bonding in the modelling assumes a constraint that may not exist under biological conditions. In addition the benzyl substituents of the ligands could not be definitively placed into a particular region; however these findings were supported by a survey of the CCDC and comparing conformations to crystal structures and global minima, respectively.

2.7.1.3 Pharmacophore Validation using GASP

GASP was run on the antagonists **1, 2, 7,** and **8** in both ring flexible and rigid modes to assess the possible binding conformations of these compounds. The highest scoring pharmacophore with obvious similarities to the model constructed via conformational analysis is shown below in Figure 2.32. The distances between ring B and position X, and ring G and position X are 9.37Å and 4.21Å, respectively. These distances compare

favourably with the distances derived from conformational analysis model with BX = 9.4 ± 0.5 and GX = 4.2 ± 0.3 (see Figure 2.30).
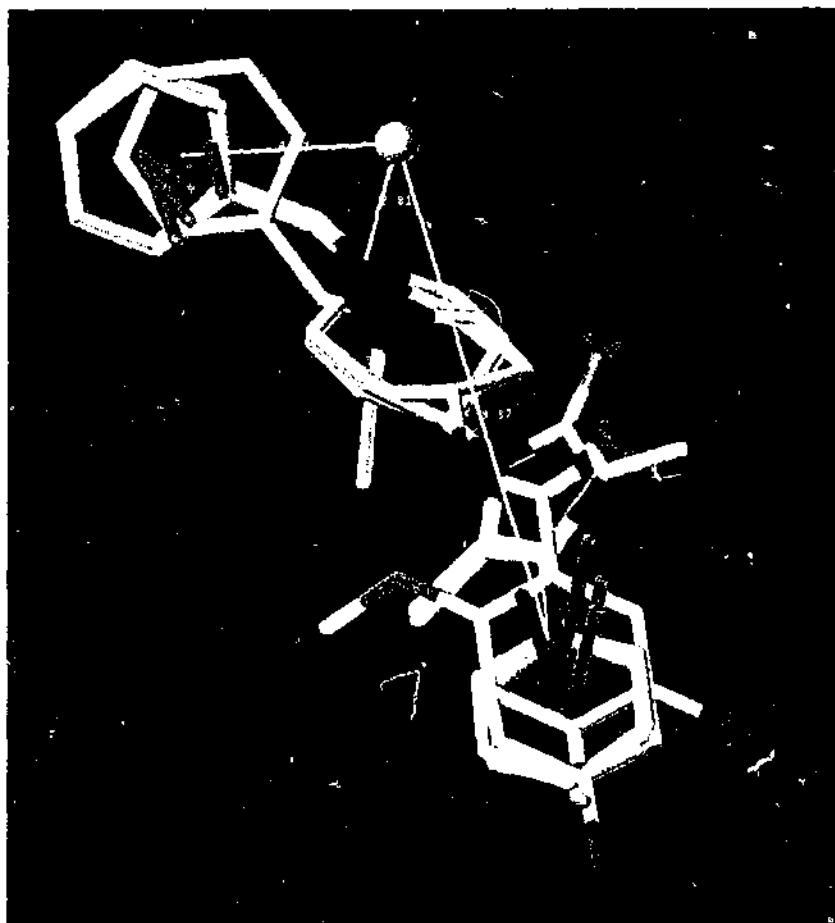
Figure 2.32. *The highest scoring resultant pharmacophore from the GASP analysis of compounds 1, 2, 7 and 8.*

2.7.1.4 Comparison with other D_2 and D_4 Antagonist Pharmacophores

At the commencement of this work there was little published material on D_4 pharmacophores, apart from the tricycle model, derived from octoclothepin and tefludazine, proposed by Boström *et al*[104]. However at the same time that this model was first shown at the 13th European Symposium on QSAR, Jonas Boström presented a model combining both the tricyclic and elongated pharmacophores and this work was later published in JCAMD in November 2000. The model developed here has some similarities to the combined model developed by Boström *et al.*[35], although different ligands were used in its development and only regions B and G were explored, see Figure 2.30. The model proposed by Boström *et al.* aids in the refinement of the model presented here allowing for differing placements of the chromenone moiety within the

model. Indeed it would seem more likely that the chromenone occupies regions A and G of the Boström *et al.* model[35], see Figure 2.33.
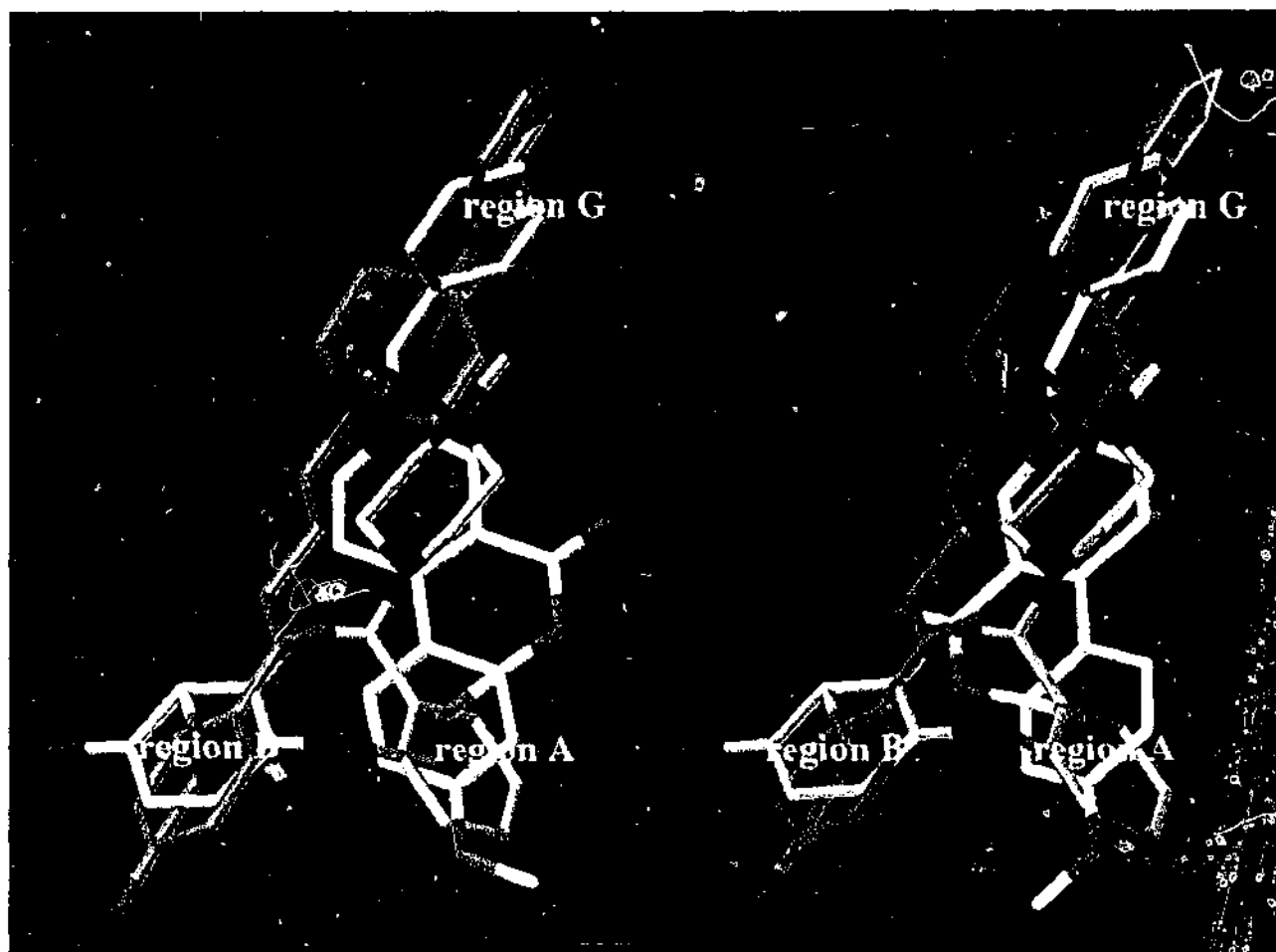
Figure 2.33. *Different possible binding modes of chromenone (white) shown with spiperone (magenta) and clozapine (orange) superimposed on pharmacophore.*

Given that the chromenone moiety is preferentially located in regions A and G, the reasoning for the removal of the *antiperiplanar* orientation is flawed. In addition to this when compound 1 used in the pharmacophore elucidation is placed in a *gauche -ve* orientation onto the CoMFA fields derived in the tricyclic analysis, unfavourable interactions can be seen with sterically prohibited areas, see Figure 2.34.

Figure 2.34. *Compound 1 placed onto the D_2 CoMFA steric fields from the tricyclic analysis, showing the unfavourable interaction between the methyl group from the pyrazole ring and the sterically prohibited areas.*

Therefore although the pharmacophore proposed may be a possible binding mode for compounds without large substituents from position 3 on the pyrazole ring, it appears energetically more favourable for compounds similar in structure to compound 1 to adopt an *antiperiplanar* orientation. This preference for an *antiperiplanar* orientation is seen when looking at compound 1 whilst solvated (see Table 2.12) "relative energies of 1". However compound 2 with its methoxy-benzamide structure still fits best into the pharmacophore when in the *gauche -ve* orientation (see Figure 2.35), and no unfavourable steric interactions are observed. In fact, in contrast to compound 1 unfavourable steric interactions are seen when compound 2 is in an *antiperiplanar* orientation. Moreover, compound 2 is preferentially orientated in a *gauche -ve* orientation whilst solvated (see Table 2.13) "relative energies of 2". Therefore changes were made to the orientation of compounds similar in structure to 1, thus compounds 1, 5, 6, 7 and 8 were modified to adopt an *antiperiplanar* orientation for their subsequent CoMFA analyses.

Figure 2.35. LEFT: *Compounds 1 (orange) and 2 (magenta) superimposed, with 1 in an antiperiplanar orientation and 2 in a gauche -ve orientation, only relevant hydrogens are shown. RIGHT: Final pharmacophore model with regions A, B, G, and the hypothetical point of interaction, X, shown.*

The model presented by Boström *et al.* also allowed for further refinement of the model generated here in regards to the placement of the benzyl substituent, by comparison to the semi-rigid compound, spiperone with its spiro ring system (see Figure 2.33). The excellent work by Boström *et al.* and the range of compounds covered in this paper was the impetus for extending the range of compounds analysed in the following section *Extended Molecular Field Analysis.*

2.7.2 Extended Molecular Field Analysis

Initially only compounds covering areas B and G of the pharmacophore from the datasets of Carling *et al.*[54], Rowley *et al.*[55], Moore *et al.*[53] and Ohmori *et al.*[52] were analysed. With the expansion of the pharmacophore model to include the tricyclic pharmacophore developed earlier, we were able to analyse the compounds mentioned in

the tricyclic section as well as a number of additional datasets from Sanner[60], Kesten[63], Belliotti[66] and Thurkauf[67]. This increased the area being analysed within the CoMFA model to include compounds covering regions A and B, and regions A and G in addition to regions B and G (see Figure 2.35). This brought the total number of compounds being analysed to 178 and 187 for the D_4 and D_2 binding activities, respectively.

2.7.2.1 Data Selection

The compounds used in the final CoMFA model are from the data sets of Phillips *et al.*[22,23], Liégeois *et al.*[24], Carling *et al.*[54], Rowley *et al.*[55], Moore *et al.*[53], Ohmori *et al.*[52] Sanner *et al.*[60], Kesten *et al.*[63], Belliotti *et al.*[66] and Thurkauf *et al.*[67]. The datasets of Phillips *et al.*[22,23] and Liégeois *et al.*[24] cover regions A and B. the datasets of Carling *et al.*[54], Rowley *et al.*[55], Moore *et al.*[53] and Ohmori *et al.*[52] cover regions B and G and the datasets of Sanner *et al.*[60], Kesten *et al.*[63], Belliotti *et al.*[66] and Thurkauf *et al.*[67] cover regions A and G.

The first three new datasets of Carling *et al.*[54], Rowley *et al.*[55] and Moore *et al.*[53] are from the same laboratory, Merck Sharp and Dohme U.K., and contain ligands with intramolecular hydrogen bonding and sub nanomolar binding affinities.

The paper of Carling *et al.* contains a series of 4-ureido-N-benzylpiperidines with some interesting intramolecular hydrogen bonding possibilities as well some Imidazol-2-one-N-benzylpiperidines with sub-nanomolar binding affinities. The binding affinities were obtained by displacement of [³H]spiperone from cloned human D_4 receptors expressed stably in HEK-393 cells. All 30 new compounds presented in this paper were modelled.

The paper of Rowley *et al.* contains a series of heterocyclylpiperidines consisting mainly of pyrazole- and isoxazole-piperidines moieties. The binding affinities were obtained by displacement of [³H]spiperone from cloned human D_4 receptors expressed stably in HEK-393 cells. Of the 35 new compounds presented in this paper 32 were modelled, compounds 20, 21 & 26 were not included in the model because of

incompatible structural arrangements and poor binding affinities. In compounds 20 and 21 the different placement of the nitrogen in the piperidine ring gave differing directional vectors to the other compounds and thus differing distances between the pharmacophoric elements of interest. For compound 26 the addition of a methyl group into the piperidine ring forced the diketone to an axial conformation. This is similar to compound 27 however it is stated for compound 27 that, "there is little energy difference between the two ring flip conformations, and the pyrazole-equatorial conformation required for binding is energetically accessible"[55]. Therefore the pyrazole-equatorial conformation of compound 27 was added to the model.

The paper of Moore *et al.* contains a series of 4-N-*linked*-heterocyclic piperidine compounds very similar to those in the paper by Rowley *et al.* All 16 new compounds presented in the paper were modelled. The binding affinities again were obtained by displacement of [^{3}H]spiperone from cloned human D_4 receptors expressed stably in HEK-393 cells.

The Ohmori data set is from the Yamanouchi Pharmaceutical Company, which manufactures the D_4 radioligand nemonapride (YM-09151-2). The Ohmori *et al.* paper contains the racemic radioligand nemonapride as well as 22 new N-(1-Benzyl-3-pyrrolidinyl)-5-chloro-4-[(cyclopropylcarbonyl)amino]-2-methoxybenamide derivatives and reference compounds. The binding affinities were obtained by displacement of [^{3}H]nemonapride from cloned human $D_{4.7}$ receptors expressed stably in CHO cells. The reference compound haloperidol shows comparable binding affinities at D_4 and D_2. However D_4 and D_2 binding affinities for clozapine are approximately 4-fold higher indicating that the binding affinities in the high nanomolar range, as is the case for clozapine, may actually be tighter than those reported compared to the testing by the Merck Sharp and Dohme laboratories. All new ligands reported and nemonapride were modelled.

The Sanner dataset is from Pfizer Central Research in Groton U.S.A. and presents 15 new pyrido[1,2-a]pyrazine analogues. The binding affinities were obtained by displacement of quinpirole from cloned human D_2 and D_4 receptors expressed stably in CHO cells. The binding affinities of the reference compounds haloperidol and

clozapine were similar to those in the Merck Sharp and Dohme papers. All compounds were modelled with the exception of the optical isomers ($\pm$)3a and ($\pm$)10a.

The Belliotti and Kesten datasets are both from the same laboratories, Parke-Davis Pharmaceutical Research Michigan U.S.A. The binding affinities for the Belliotti dataset were obtained by displacement of haloperidol from cloned human D_{2L} and $D_{4.2}$ receptors expressed stably in CHO cells. All new compounds presented in the Belliotti paper were modelled. The binding affinities for the Kesten dataset were obtained by displacement of [^{3}H]spiperone from cloned human D_{2L} and $D_{4.2}$ receptors expressed in CHO-K1 cells. All new compounds with the exception of compounds 22 to 31 were modelled where binding data was available. Compounds 22 to 31 were not modelled due to their highly flexible nature.

The Thurkauf dataset is from the Neurogen Corporation and contains 8 new 2-Phenyl-4-(piperazine-1-ylmethyl)imidazole analogues, all of which were added to our analyses. The binding affinities for the Thurkauf dataset were obtained by displacement of [^{3}H]YM 09151, nemonapride, from cloned human D_2 and D_4 receptors expressed in CHO cells. The binding of the reference compounds clozapine and haloperidol are comparable to those of the Phillips datasets and marginally higher than those presented by the Merck Sharp and Dohme group.

2.7.2.2 CoMFA Field Interpretations

All fields shown in this section were generated using the method outlined in Molecular Field Analysis. The contour levels for visualisation of the CoMFA fields were set to represent 50% of the relative contribution of each factor involved in correlating the pharmacophore model to activity.

D_2 Dopaminergic model

When the database of compounds was rotated 30° around the y-axis relative to the starting orientation, the best model with the highest q^2, r^2 and the lowest standard error of prediction was found. The resultant model built from 187 compounds had 8

components giving a q^2 of 0.54 and a r^2 of 0.859 with a standard error of estimate of 0.336; the relative contributions to the model were 43% and 57% from the steric and electrostatic fields, respectively. The 187 compounds spanned a pK_i range of almost 5 log units. The predicted activities of all compounds were within ±0.85 pK_i log units of their measured values. An improvement in the predictive ability of the model to 0.56 was seen when region focussing was applied to the analysis, although this was with the inclusion of an additional component. In actual fact the F statistic for the region focussed model was lower than that of the original analysis, indicating that region focussing did not improve the model.
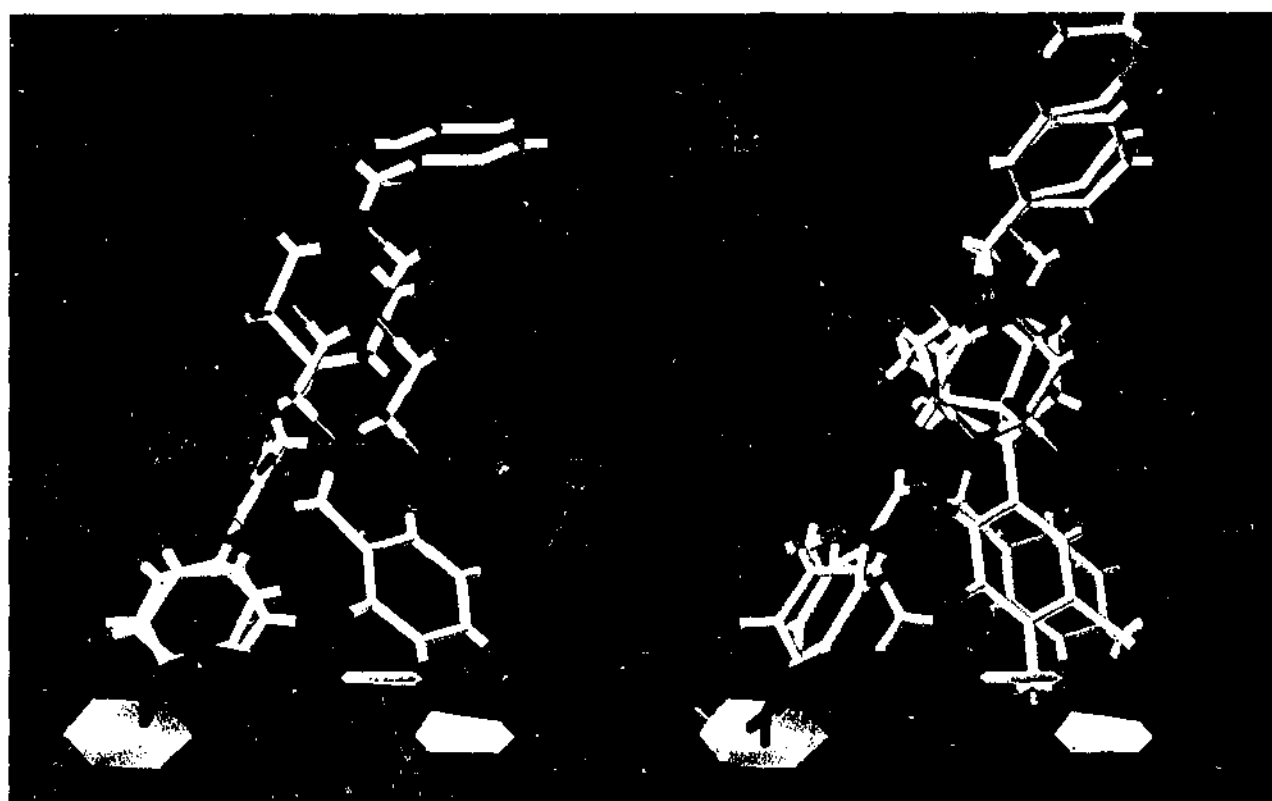
Figure 2.36. *CoMFA fields showing steric contributions from analysis of all the D_2 binding affinities for the 187 compounds. The diagram on the left shows clozapine with Carling #15, the diagram on the right shows clozapine (orange), nemonapride (left) and Belliotti #16. Green areas indicate where additional steric bulk is favourable, and yellow areas where it is not favourable for D_2 binding.*

There are a number of similarities between the previously developed tricyclic model (see Figure14), and this complete analysis (see Figure 2.36). Region G1 where additional bulk is provided by the chlorine substituent in the tricyclic model is again present in this complete model. Interestingly the detrimental effect of additional bulk from the bulky alkyl substituents of the tricyclic model seems to be tolerated in this

complete model which may be due to the additional range of compounds not having this feature in common yet also exhibiting poor D_2 binding affinities. Steric bulk is also shown as favourable for increasing binding in two regions above ring B, G3 and G4. These regions are a result of the tightly binding compounds from the Ohmori dataset occupying this area due to a different planarity of their hydrophobic ring system. The methoxy substituent of compounds from the Ohmori dataset occupies region G2. The Ohmori compounds are again implicated in the region Y2 where additional steric bulk is unfavourable to D_2 binding. The logic behind this is easily seen when looking at the Ohmori dataset with the binding affinity of compounds decreasing with increasing size of substituent in this area. Another feature of the steric contours shown in Figure 2.36 is the detrimental effect of bulky substituents in region Y1; this position corresponds to positions 3 and 4 of ring A, for the compounds in the Belliotti, Thurkauf and Sanner datasets. This is easily seen the in right hand side of Figure 2.36 with the bulky 3,4-di-methyl substituent of Belliotti #16 reducing its binding affinity in comparison to compounds from the same dataset with only a para methyl substituent.
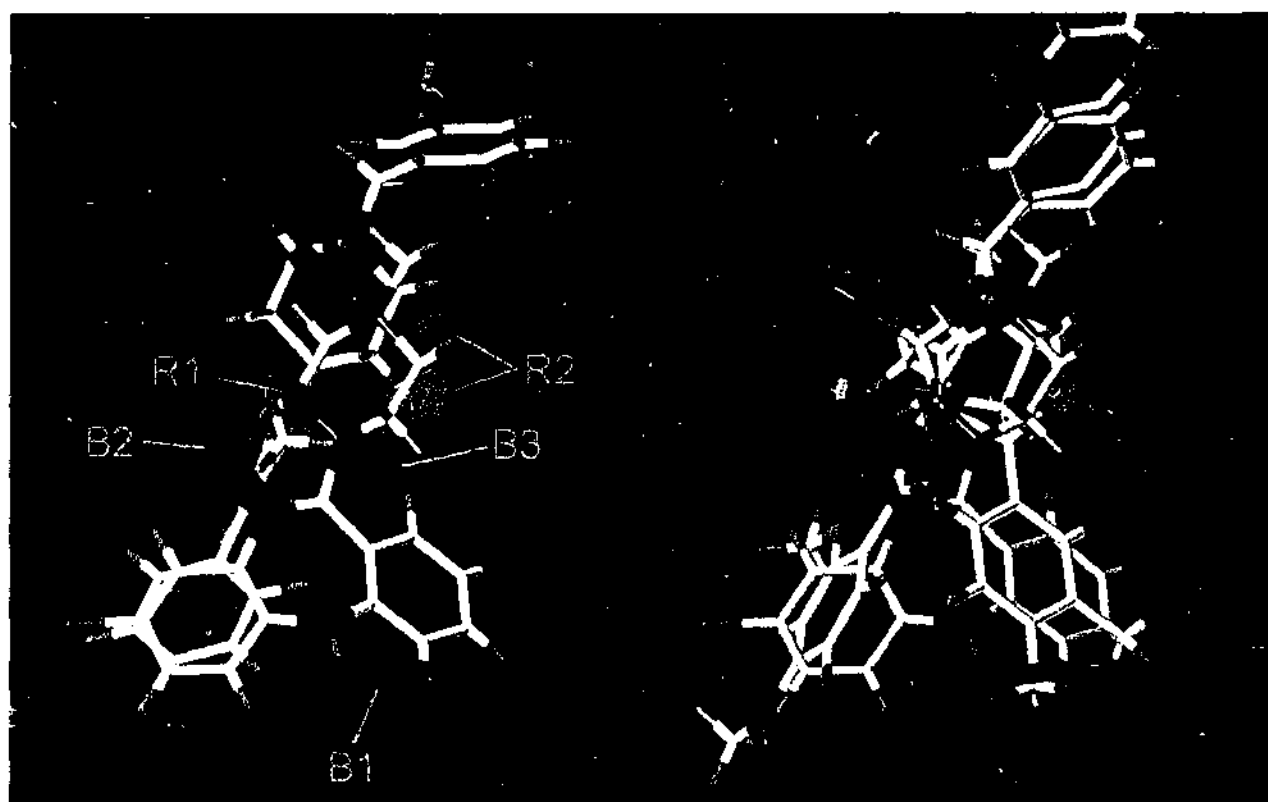
Figure 2.37. *CoMFA fields showing electrostatic contributions from analysis of all the D_2 binding affinities for the 187 compounds. The diagram on the left shows clozapine with Carling #15, the diagram on the right shows clozapine (orange), nemonapride (left) and Belliotti #16. Blue regions indicate where partial positive charge increases affinity and red regions indicate where partial negative charge increases affinity.*

The electrostatic fields, shown in Figure 2.37, show a number of regions where partial positive charge enhances activity. Again the detrimental effect that a nitrogen in position A. region B2, can have on D_2 binding affinity, as shown in the tricyclic model, is represented in this complete model. In addition to this a nitrogen present in region B3, position B of the piperazine ring, is also shown to be detrimental to D_2 binding affinity. This position also corresponds to the position of the nitrogen in piperazine ring for compounds that bind in regions A to G, such as those compounds from the Belliotti, Sanner, Thurkauf and Kesten datasets. The detrimental effect of the nitrogen can be seen when comparing Belliotti #16 to Belliotti #20, (Appendix Table Belliotti), where the presence of nitrogen in region B3 reduces binding 4-fold. The detrimental effect of electronegative substituents in region B1, the para position of ring A, is evident in this model. This is easily seen when looking at compounds Sanner #10j and #10l compared to Sanner #10c and #10i, respectively, with their electronegative substituents occupying region B1 and reducing binding affinity (Appendix Table Sanner).

There are a number of regions where partial negative charge enhances activity, located around the piperazine ring substructure. These regions, R1 and R2, around the piperazine ring appear to be due to the charge distribution of this substructure. This emanates from the different electronic environments of the piperazine and piperidine substructures. The piperazine ring with its additional nitrogen withdraws electrons from the neighbouring carbons, thus imparting on them a partial positive charge relative to the same carbons on a piperidine ring. This indicates that a piperidine ring is preferential for increasing D_2 binding affinity, and this is evident when comparing compounds Belliotti #16 to Belliotti #20 and Kesten #19 to Kesten #5 (Appendix Table Kesten). Region R1 may also be due in part to the compounds from the Ohmori and Carling datasets that both have oxygen located in this vicinity and are in general tighter binding compounds than those present in many other datasets (Appendix Tables Ohmori and Carling).

D_4 Dopaminergic model

When the database of compounds was rotated 60° around the x-axis relative to the starting orientation, the best model with the highest q^2, r^2 and the lowest standard error

of prediction was found. The resultant model had 8 components giving a q^2 of 0.52 and a r^2 of 0.862 with a standard error of estimate of 0.366; the relative contributions to the model were 4 8% a nd 5 2% f rom t he s teric a nd electrostatic f ields, r espectively. T he predicted activities of all compounds were within ±0.85 pK_i log units of their measured values, with the exception of compounds 3A and P1 that had residuals of -1.28 and 1.16 log units respectively. An improvement in the predictive ability of the model to 0.534 was seen when region focussing was applied to the analysis, although this was with the inclusion of two additional components. In actual fact the F statistic for the region focussed model was lower than that of the original analysis, indicating that region focussing did not improve the model.
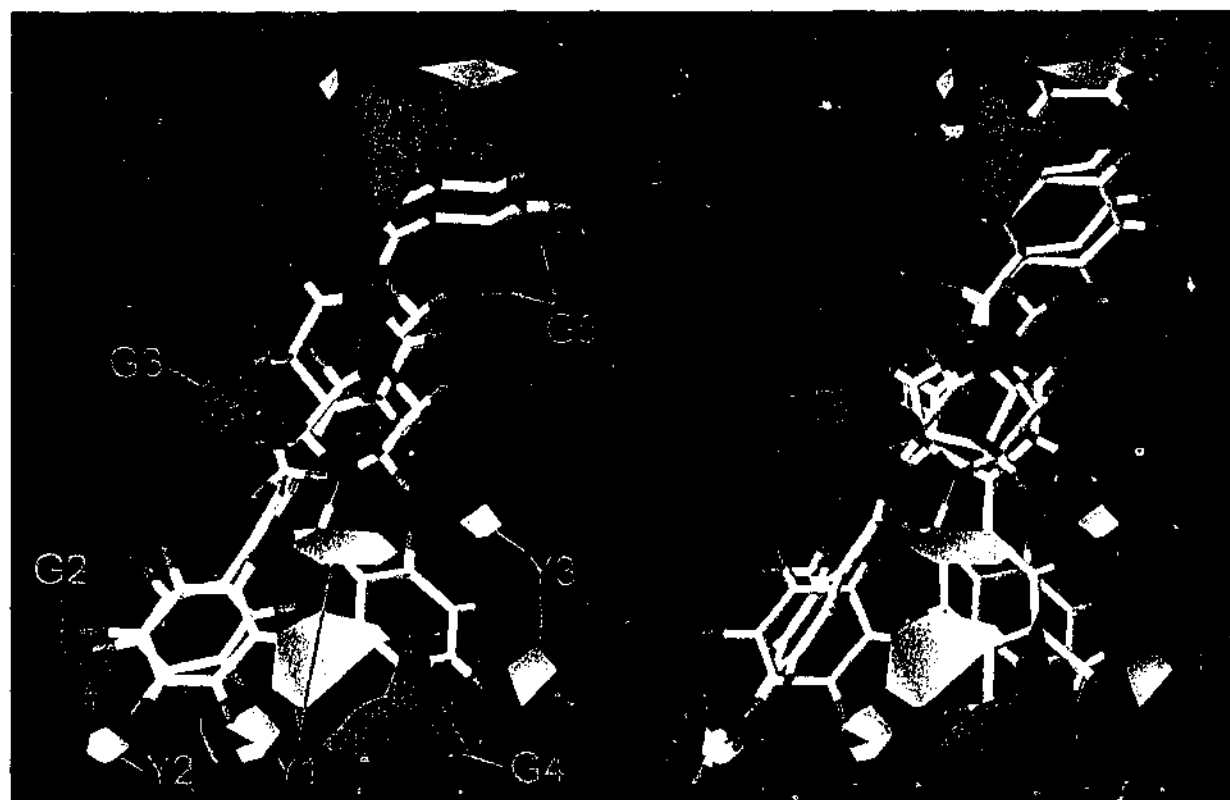
Figure 2.38. *CoMFA fields showing steric contributions from analysis of all the D_4 binding affinities for the 178 compounds overlayed. The diagram on the left shows clozapine with Carling #15. the diagram on the right shows clozapine (orange), nemonapride (left) and Belliotti #16. Green areas indicate where additional steric bulk is favourable, and yellow areas where it is not favourable for D_4 binding.*

The steric fields from the complete CoMFA model show a multitude of smaller areas, due to the large number of components used in the model. There are some similarities to the tricyclic model developed earlier, in particular the region where large alkyl group groups were reducing binding affinity in region Y1 from the tricyclic model. The

region where additional bulk is favourable to binding affinity in region G4 is again present in this complete D_4 model. There are a number of other regions where additional steric bulk is favourable to binding affinity, for example in region G5 where the extra hydrophobic group sits. Additional bulk in region G3, corresponding to the methyl substituent from the imidazolone ring of Carling #15, also increases binding affinity. This can be seen when comparing compounds Carling #10 to Carling #11, where the addition of an ethyl substituent increases binding affinity over one order of magnitude. Not unlike the D_2 steric model the Ohmori compounds are implicated in the region Y2 where additional steric bulk is unfavourable to D_2 binding. It also appears as though large substituents on ring A occupying region Y3 are not as well tolerated in this D_4 model compared to the D_2 model.
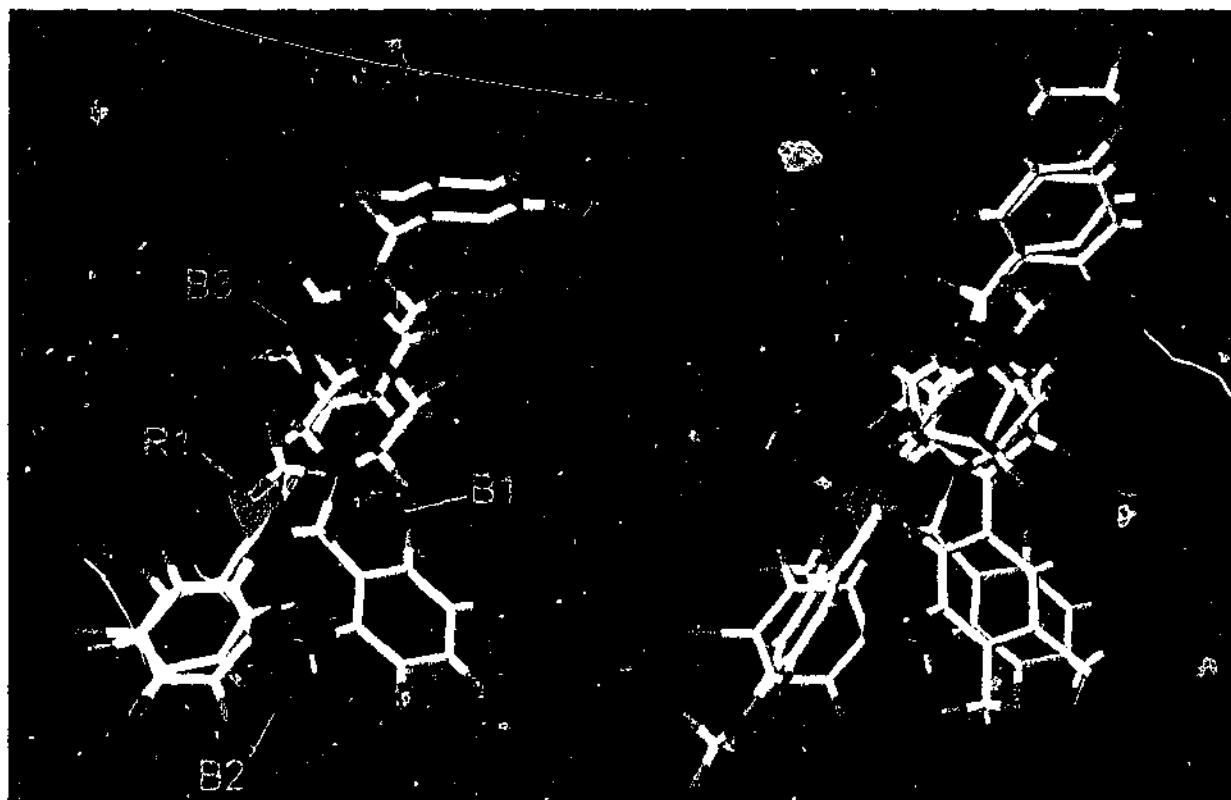
Figure 2.39. *CoMFA fields showing electrostatic contributions from the complete D_4 model. The diagram on the left shows clozapine with Carling #15, the diagram on the right shows clozapine (orange), nemonapride (left) and Belliotti #16 (right). Blue regions indicate where partial positive charge increases affinity and red regions indicate where partial negative charge increases affinity.*

The electrostatic contributions for D_4 from the complete model differ slightly to those obtained from the tricyclic model, with the exception of the region of partial positive charge B1. However in this model the partial positive charge extends from near

position 1 on ring A back underneath the carbon linked to the piperazine ring. This additional region where partial positive charge enhances activity is due to the beneficial effects of the proximal nitrogen in the piperazine ring substructure withdrawing electrons and imparting a partial positive charge in this area. This indicates that the D_4 model favours piperazine rings over piperidine rings when attached to the tricyclic substructural unit. This can be seen when comparing compounds 3J and P8 from Table 2.1. Interestingly this does not hold true for all compounds, as when looking at the compounds from the Belliotti dataset a preference for piperidine rings can be seen, see compound Belliotti #24 compared to Belliotti #31. However this serves to highlight the fact that these two differing families of compounds, those that bind in regions A and G and those that bind in regions A and B, do indeed bind in slightly different orientations with respect to their piperidine and piperazine rings. The second region where partial positive charge enhances activity is region B2. This region could be due to electronegative heteroatoms at position Z imparting a partial positive charge in this region. Although when looking at compounds 3a and 3b compared to 3h and 3i Table 2.1, respectively, in one instance we see a significant drop in affinity when the CH_2 is substituted for O and in the other instance only a small change, indicating that other factors are also at play. Upon closer inspection of these compounds it appears as though the chloro substituent in position 8 in conjunction with the heteroatom is having a significant effect on the electronic environment of ring B, and aiding in binding affinity.

The only area where partial negative charge enhances activity is region R1, located around the nitrogen of the amino linkage for compounds from the Ohmori dataset. This region is due to the combined effects of the electronegative amino linkage and the electronegative pyrazole, imidazolone and isoxazole linkages present in the Carling, Rowley and Moore datasets, and their relatively tight binding affinities.

2.8 CONCLUSION: EXTENDED MODEL

We developed a D_2 and D_4 pharmacophore model that was able to accommodate all the aforementioned ligands. After analysis of the model presented by Boström *et al.* and subsequent refinement of the placement of the chromenone it was shown however that compounds similar in structure to **1** would adopt an *antiperiplanar* conformation whilst methoxy-benzamide compounds would bind in the *gauche -ve* conformation. The main requirements for the pharmacophore are protonated tertiary nitrogen at position X and two of the possible three areas of hydrophobic interaction around dummy atoms A, B and G.

2.8.1 Designing Compounds

Extensions of the themes produced by the D_2 model could be undertaken to increase or decrease binding affinity at the D_2 receptor or to modify the binding profile of a compound so as to attain greater selectivity between the two dopaminergic receptors. The additional bulk in region $G1_{D2}$ present in this analysis could be further explored with other halides, methyl or even triflouromethyl substituents, as has been shown by Liao et al[105,106] with their series of 2- or 8-trifluoromethylsulfonyloxy (TfO) substituted compounds. The other areas where additional bulk increases affinity are regions $G3_{D2}$ and $G4_{D2}$ above ring B, due to the different planarity of the Ohmori compounds. This area could be occupied by designing compounds with a similar framework to that of the Ohmori dataset, by using a pyrrolidine ring instead of a piperidine or piperazine ring as the ring that contains the basic nitrogen. This preferred planarity of this aromatic ring could be partially realised in the tricyclic compounds by altering the angle between the two planes of these compounds. This would be best achieved by placing a sulphur atom in position X between the two ring planes, as compounds with sulphur have a smaller angle between the two planes compared to oxygen and nitrogen[33,107,108]. This smaller angle would place ring B in a theoretically preferential plane for binding and increased binding affinity. Further extensions of the theme to attain greater selectivity between the two receptors could be achieved by ensuring that compounds were designed with substituents that occupied region $Y1_{D2}$ or $G4_{D4}$ off ring A. It was shown again that substituents are not well tolerated here in the D_2 model, region $Y1_{D2}$, however they are

shown to be beneficial in the D_4 model, region $G4_{D4}$, making this region an ideal area to exploit for increasing D_4 selectivity. The electrostatic fields show that either the presence or absence of an electronegative substituent at position A can also control D_2 binding. Indeed this region where partial positive charge enhances activity, region $B2_{D2}$, is in direct contrast to what we see in the electrostatic fields generated from D_4 binding affinity data, region $R1_{D4}$. This area is thus an ideal area to exploit for designing compounds with greater selectivity between these two dopaminergic receptors. Another area that is ideal to exploit in regards to increasing selectivity is the ring containing the basic nitrogen. Most compounds with a piperidine ring bind better at the D_2 receptor compared to compounds with a piperazine ring and vice versa for the D_4 receptor. This gives an indication that there are differing residues with different electronic characteristics interacting with this area of the ligands for the two receptors. This could be exploited by designing compounds with different electronic characteristics in this region, by using 1,3-morpholine or 1,3-thiomorpholine derivatives, however this may come at the expense of reducing the basicity of the distal nitrogen which may reduce binding affinities and thus would have to be carefully monitored.

Extensions of the themes produced by the D_4 receptors could be explored to tailor D_4 binding affinity or to increase selectivity between the D_4 and D_2 receptors. Region $Y1_{D4}$ where bulky alkyl groups were reducing binding affinities could be used to increase selectivity between the two receptors. Additional space here seems to be tolerated better in the D_2 model and this could be exploited to increase D_2 selectivity if desired. Further areas that might possibly be exploited to increase selectivity would be the region $G3_{D4}$, corresponding to the methyl substituent from the imidazolone ring of Carling #15, and also region $G5_{D4}$, corresponding to the area around ring G. These areas have been shown to increase binding affinity for the D_4 model and are not present to the same extent in the steric fields produced from the D_2 model.

Chapter 3

Structure-Based Design

3.1 GPCR MODELS: INTRODUCTION

GPCRs represent one of the most important families of drug targets for the pharmaceutical industry. A survey done in 1995 by GlaxoWellcome states that of the top 100 best selling prescription drugs, more than 20% exert their therapeutic effect by targeting GPCRs[1], whilst over 50% of all modern drugs are targeted at GPCRs[2]. For every GPCR that has been purified, expressed and subsequently had an active ligand found for it, this ligand has been of great pharmaceutical benefit. Building theoretical models for GPCRs, which represent one of the most prominent classes of validated drug targets, is important for modern drug discovery.

There are a number of ways of building theoretical GPCR models. Initially these models were built using ideal helices of the sequence in question[3,4], using electron microscopy structural data from bacteriorhodopsin to align the helices. Later models were built by using the X-ray structure of bacteriorhodopsin[5-7], and even more recently rhodopsin[8,9], as a template. Using these templates, residues are mutated to the aligned sequence residues resulting in a model. Clearly the main drawback of using a template in any model construction is that the model is biased towards the conformation of the model it is based on. This problem can be overcome by careful consideration of the various factors in play[7].

Another method used for GPCR model development is comparative modelling[10,11]. This method was originally developed in the late eighties by Sali et al.[10] in a program called Modeller (versions 1-4) and due to its many successes[11-13] has recently been incorporated into the InsightII suite by Accelrys under the name Modeler. This method enables a model to be built from the sequence alignment of a structure with one or more related structures. This is done by satisfying both local molecular geometry and spatial restraints, derived empirically from a database[14] of protein structures. This is the method employed in construction of GPCR models in this chapter, and is discussed in more detail later.

Another method for GPCR model construction developed recently by Shacham et al.[15], called PREDICT, employs a concept similar to *ab initio* protein folding approaches.

This method is similar to *ab initio* techniques in that it takes into account both internal protein-protein interactions (i.e., the amino acid sequence) and the protein-membrane interactions. This method has been validated by construction of the rhodopsin structure from amino-acid sequence alone, resulting in a model with a RMS fit of 3.87Å compared to the original. This technique has the benefits that GPCR models that may not be similar to rhodopsin can be constructed with reasonable accuracy.

Apart from theoretical modelling techniques that are available for GPCR model construction there is another practical method developed by Yeagle *et al.*[16]. In this method the tertiary structure of rhodopsin is derived using solution structures of peptide fragments to define the secondary structure and long range experimental distance constraints from the intact protein to pack the secondary structures into a full tertiary structure. This method has also been applied to the cytoplasmic face of the activated (metarhodopsin II) receptor, to show differences between the two structures and suggest how the receptor is activated to couple with transducin.

Constructed GPCR models created are able to be validated using data from mutagenesis studies[5-7,9,17-25] and substituted cysteine accessibility method (SCAM) studies[19-22,25-29], to ensure that they comply with experimental data. The SCAM method[30] identifies water accessible residues, that are proposed to exist on the interior of the receptor.

The wide variety of techniques available for GPCR model construction and validation are likely to increase in the future since drugs targeted at GPCRs are currently a large source of income for pharmaceutical companies. Additionally with the multitude of new receptors being found due to the unravelling of the human genome, the need for modelling tertiary structures of these new sequences and subsequent structure based design from these models will become even more important.

3A.1 HOMOLOGY MODELLING

It is well known that high levels of amino acid sequence homology leads to similar 3D structure. Model building of a target structure based on homologous sequences with known structures is termed comparative modelling. All current comparative modelling methods consist of four steps: template selection; template-target alignment; model building; and model evaluation.

Originally, searches of homologous sequences to the target were done with alignment programs such as FASTA[31] or BLAST[32]. However in the last decade other approaches, such as Threader[33], have been developed to recognise similar protein folds and sequences. Once a suitable template has been found the target and template are aligned using one of the above mentioned methods. The target and template are likely to be correctly aligned if sharing more than 40% sequence identity however they may need to be realigned if they share less than 30% identity[34].

When an alignment is thought to be correct the model building begins, this can be done with a number of programs such as COMPOSER[35], JIG-SAW[36] or Modeller[10,14]. COMPOSER and JIG-SAW work by first identifying structurally conserved regions, regions where the target and template align, and building these. Next structurally variable regions, regions were there is no alignment with the target, are built using either systematic conformational searches, molecular dynamics, genetic algorithms or even by searching through loop databases and identifying homologous loops that fit within the constraints of the model. Side chains are then added to the model so as minimise the energy of the system and to conform with allowed side chain angles found in rotamer libraries. The model building process of Modeller is discussed in greater detail later.

Model evaluation is done with programs such as PROCHECK[37] or WHATCHECK[38], which examine the stereochemical integrity of the constructed models. Further refinement of the models can be performed via minimisation or molecular dynamics using programs like CHARMM[39] , NAMD[40] . X-PLOR or various Tripos or Accelrys modules. However this final refinement step on the modelling process has to be done

cautiously so that the resultant structures are of good quality. For example, poor quality structures may result if the wrong environment, such as the absence of counter ions or the incorrect solvent is used.

Comparative modelling using Modeller[10,11], is different from programs such as COMPOSER and JIG-SAW as it involves satisfaction of spatial restraints. Initially the modelling procedure begins with alignment of the target sequence onto a known, related 3-D structure. Then many distance and dihedral restraints are derived from the alignment of the target sequence on the template 3-D structure. The form of these restraints is obtained from the statistical analysis of the relationships between many pairs of homologous structures. Modeller relies on a database of 105 family alignments that includes 416 proteins of known 3-D structure[14]. A spatial restraint or probability density function (pdf) on a given $C\alpha_1$-$C\alpha_2$ distance, d, is obtained by least-squares fitting a sum of a number of Gaussian functions to a histogram of $C\alpha_1$-$C\alpha_2$ distances derived from the database of protein alignments. In practice more complicated spatial restraints are used that depend upon additional information derived from such things as similarity between the proteins, solvent accessibility and distance from a gap in the alignment. For example, the likelihood of a particular mainchain dihedral will depend on the type of residue, the dihedral of its equivalent residue in the sequence and the sequence similarity between the two proteins. These spatial restraints are combined with CHARMM[39] energy terms enforcing correct stereochemistry to make up the objective function[10]. Finally optimisation of the objective function in Cartesian space produces the 3-D model. The optimisation initially uses the variable target function method[41] employing conjugate gradient methods then switches to molecular dynamics with simulated annealing to refine the model.

The Modeller approach does have some limitations when dealing with GPCRs, since the database that the restraints are derived from does not include any GPCRs, although this is not surprising, as the first 3-D structure of a true GPCR was determined in 2000[42]. The benefits of such a method are that a number of different structures can be used to derive the initial distance and dihedral spatial restraints. However this benefit was not realised, as only one structure was available for use in these analyses. Other benefits of Modeller are that alignments over areas such as proline kinks do not force the generated

protein to adopt this conformation, as is the case in using a template. Additionally, sidechains are positioned to minimise the objective function, which is further complemented by molecular dynamics and simulated annealing methods. Furthermore, models can be constructed with relative ease with ligands bound in the active site and without the need for additional docking programs. The disadvantages of this additional feature are that the model may be biased by the initial placement of the ligand and additional constraints must be added to the construction of the model. Therefore one must be careful not to place too many constraints on the model construction and hence bias this process. In addition careful visualisation of the ligand and the surrounding residues in the proposed active site must be undertaken to check their stereochemical integrity.

Taking these pitfalls into account, Modeller was used to construct D_2 models with bacteriorhodopsin as the template. Models were built (a) without any ligands bound in the active site and (b) with the endogenous agonist, dopamine, bound in the active site for comparison to the recently released bacteriorhodopsin based D_2 model (1I15)[7]. Unfortunately the bacteriorhodopsin structure (1BRD)[43] that the theoretical D_2 model (1I15) of Teeter et al[7]. was based on was of poor resolution, 3.5Å in the X and Y directions and 10Å in the Z direction. Therefore the bacteriorhodopsin structure (1C3W)[44] at 1.55Å resolution was used. This structure was exactly the same strain of bacteriorhodopsin as 1BRD and therefore had 100% residue identity and 100% similarity of the alignment[45]. Additionally D_2 models were constructed using bovine rhodopsin[42], 1F88 at 2.8Å resolution, as the basis. Models were constructed (a) without any ligands bound in the active site and (b) with dopamine bound in the active site for comparison to the bacteriorhodopsin-based D_2 models.

3A.2 METHOD: HOMOLOGY MODELLING

3A.2.1 Method Overview

Figure 3.1. Schematic of method overview
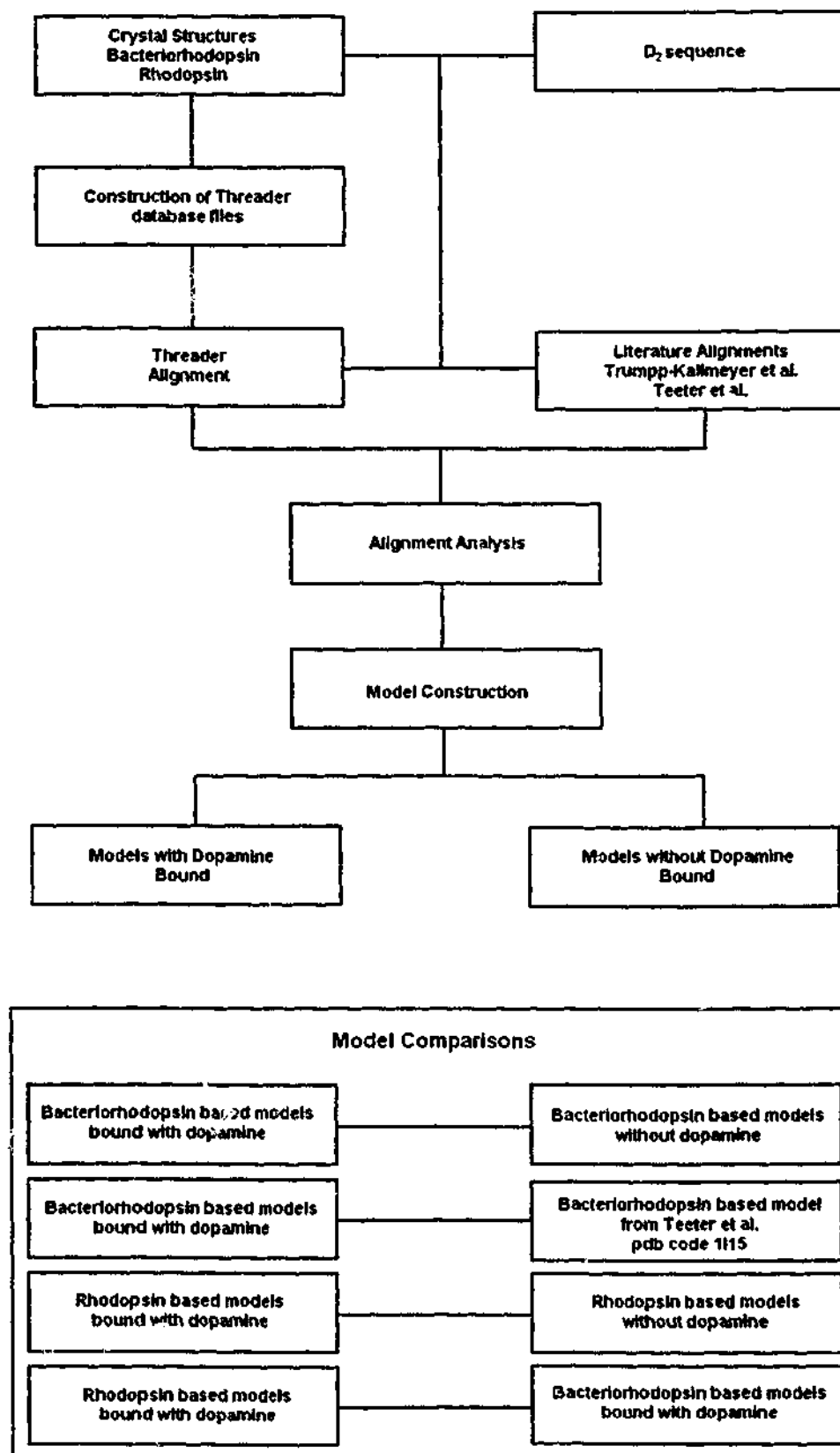
Figure 3.1 shows a schematic of the method overview that is discussed below. Dopaminergic receptor models were constructed using crystal structures of bacteriorhodopsin and rhodopsin as starting points. Models were constructed by initially creating a threader database file from the ground state wild type bacteriorhodopsin (1C3W)[44] and bovine rhodopsin (1F88)[42], by running Procheck[37,46], naccess[47] and strsum[33] on the structure. Threader[33] was then used to align the sequence of D_2 onto the sequence of bacteriorhodopsin and rhodopsin. The alignment file of bacteriorhodopsin was altered so as to comply with the alignment presented by Teeter et al.[7] and mostly comply with the alignment of GPCR's according to Trump-Kallmeyer[6]. The alignment file of rhodopsin was altered so as to comply with Trump-Kallmeyer. Modeller[10,11] was then run on the alignment files to create tertiary structures of the D_2 receptor. The resulting structures were then minimised within Sybyl 6.7[48] using Kollman all-atom charges and force field[49]. The models created were analysed within Sybyl to check their compliance with mutagenesis[23] and substituted cysteine accessibility method (SCAM) studies[19-22,25,27-29] and their stereochemical quality was checked within Procheck[37,46]. The bound and unbound models based on the differing crystal structures were then compared against one another and the published theoretical D_2 model, 1115, from Teeter et al.[7] to assess their suitability. A complete table showing the alignment of D_2, bacteriorhodopsin and rhodopsin sequences with their numbering is shown in Table 3.1. Where the Ballesteros and Weinstein numbering system does not account for residues, in cytoplasmic and extracellular loops, the numbering scheme based on that of the D_2 long sequence (P14416) from Swiss-Prot[50] is used.

3A.2.2 Generating the Alignment of D_2 onto Bacteriorhodopsin and Rhodopsin

3A.2.2.1 Construction of the Threader database file

The threader database file created contains a numerical representation of the bacteriorhodopsin and rhodopsin crystal structure from which it was derived. This gives an indication as to the type, volume and characteristics of the residues within the sequence of the protein. This information is crucial for the threading of the D_2 sequence

onto that of rhodopsin or bacteriorhodopsin. However before this can be done a number of files must be created within Procheck.

Procheck checks the stereochemical quality of a protein structure, producing a number of postscript plots analysing its overall and residue-by-residue geometry. It also produces two files needed to create a threader database file. One file holds the 'cleaned-up' version of the original PDB file, with any wrong atom-labels corrected in accordance with the IUPAC naming conventions. The other file contains residue information used by the phi/psi plotting programs. Procheck also produces a number of log files that give an indication as to which residues are causing problems in construction of the database file.

Naccess, a stand alone program that calculates the atomic accessible surface defined by rolling a probe of given size around a van der Waals surface, was then run to create a residue accessibility file. The residue accessibility file contains summed atomic accessible surface areas over each protein or nucleic acid residue; this file is necessary for creating a threader database file.

Strsum, which is short for *structure summariser*, is used to generate the tdb (Threader database) files so that users can include their own structures in the threader fold library. Strsum combines the three output files to generate the threader database file.

The resulting tdb file is then placed into the fold library directory and Threader is ready to be run on the Threader database file created.

3A.2.2.2 Running Threader

Threader was run with the depth flag set to 1000 to obtain the best result. The print alignment in modeller format was set to obtain an output for putting into Modeller. The sequence similarity (-S flag) was adjusted to 100 so that the alignment of the sequences was not based purely on threading, giving a 50-50 mixture of threading and sequence alignment.

The modeller alignment file created in Threader is checked within Seaview[51] to see if there is a good correlation between positive, negative, hydrophobic and hydrophilic residues for the two sequences. The alignment file is also checked against the alignments presented by Teeter[7] and Trump-Kallmeyer[6] to see if further adjustments need to be made. The alignment is then ready for input into Modeller.

3A.2.3 Three-dimensional D_2 Model Generation

In a process termed comparative modelling, Modeller calculates a 3-D full-atom model from an alignment of two sequences and one corresponding known structure. Modeller does this by satisfying specific spatial restraints. In this case restraints are partially derived from the bacteriorhodopsin or rhodopsin crystal structure and the protein database mentioned earlier to create a 3-D structure of the dopaminergic D_2 receptor.

Modeller requires an alignment file, a restraints file and an input file in order to run, as well as the pdb structure on which the model is to be based. The alignment file obtained from threader must be further altered to include breaks within the sequence alignments. The input file contains pointers to the alignment file, the pdb structure onto which the model is based, various subroutines to be called and additional data such as the number of models to be constructed. The deviation parameter in the input file was set high so as to ensure a high degree of randomisation in the models created. The resulting output files of modeller that are of interest are in protein data bank (pdb) format.

3A.2.4 Minimising the Potential Energy of the Protein

All pdb files were minimised in three steps in order to lower the proteins chances of being caught in local minima and to remove drastic and local clashes:
- minimise only hydrogens, holding the rest of molecule fixed,
- minimise only sidechains, holding the backbone fixed,
- minimise the whole molecule.

Each minimisation step was for 1000 iterations or until a RMS gradient of 0.01 kcal/mol was reached using Kollman all atom charges and Kollman force field. The endogenous

ligand dopamine had Gasteiger and Marsili charges and Kollman atom types assigned to it before minimisation. Additional constraints were added in the minimisation of the bound receptor. The m-hydroxy and nitrogen of dopamine were constrained to be within hydrogen bonding distance, 2.8 angstrom, of $Ser^{5.46}$ and $Asp^{3.32}$ respectively[7,23], to mimic the important interactions of these groups at the receptor.

3A.2.5 Analysing the dopamine bound and unbound receptor states of bacteriorhodopsin based models relative to 1I15.

The two groups of ten resulting pdb files, five without the agonist dopamine bound and five with dopamine bound, were then compared amongst each group and to the model $1I15^{[7]}$. This was done using the following criteria:

- Analysis of the modeller objective function, which gives an indication as to the quality of the model produced.

- The RMS fit for all atoms and for C_α carbons was calculated for the differing groups. These was done using 'fit monomer' and by selecting all residues that were in common with the D_2 model 1I15, as the residues of interest.

- The energies (kcal/mol) of the two groups before and after minimisation were examined using Kollman all atom force field and Kollman charges.

- The Ramachandran plot statistics, from Procheck, of the two groups and the D_2 model 1I15 were examined.

- Areas of possible incorrect model building, such as removal and building proline kinks, were examined in detail.

- Models had their secondary structure predicted, by 'assign secondary structure' within Sybyl, and this was compared to theoretical predictions.

- The active site and sodium-binding site of the models were compared to the proposed active site and sodium-binding site of 1I15.

3A.2.6 Analysing the bound and unbound receptor states of rhodopsin based models relative to bacteriorhodopsin based models.

The groups of ten resulting pdb files, five without the agonist dopamine bound and five with dopamine bound, were then compared amongst one another and to the bacteriorhodopsin based models. This was done using the following criteria.

- Analysis of the modeller objective function, which gives an indication as to the quality of the model produced.

- The Ramachandran plot statistics, from Procheck, of the two groups and rhodopsin were examined.

- Models had their secondary structure predicted, by 'assign secondary structure' within Sybyl, and this was compared to theoretical predictions.

- The active site and sodium-binding site of the models were compared to the proposed active site and sodium-binding site from bacteriorhodopsin based models.

- The rhodopsin and bacteriorhodopsin models were also visualised with SCAM data highlighted, from a number of papers[19-22,25,27-29], to see if they all complied with these practical analyses and to see if there were any significant differences.

Table 3.1. Numbering scheme of the bacteriorhodopsin, rhodopsin and D_2 dopaminergic receptor.

TM1

	First	1.30	1.31	1.32	1.33	1.34	1.35	1.36	1.37	1.38	1.39	1.40	1.41	1.42	1.43	1.44	1.45	1.46	1.47	1.48	1.49	1.50	1.51	1.52	1.53	1.54	1.55	1.56	1.57	1.58	1.59	1.60	1.61	1.62	1.63	1.64	Last
Rho	35	W	Q	F	S	M	L	A	A	Y	M	F	L	L	I	M	L	G	F	P	I	**_N_**	F	L	T	L	Y	V	T	V	Q	H	K	K	L	R	69
Bac	9			E	W	I	W	L	A	L	G	T	A	L	M	G	L	G	T	L	Y	F	L	V	K	G	M	G									33
D_2	32	P	H	Y	N	Y	Y	A	T	L	L	T	L	L	I	A	V	I	V	F	G	**_N_**	V	L	V	C	M	A	V	S	R	E	K	A	L	Q	66

TM2

	First	2.38	2.39	2.40	2.41	2.42	2.43	2.44	2.45	2.46	2.47	2.48	2.49	2.50	2.51	2.52	2.53	2.54	2.55	2.56	2.57	2.58	2.59	2.60	2.61	2.62	2.63	2.64	2.65	2.66	2.67	2.68	Last
Rho	71	P	L	N	Y	I	L	L	N	L	A	V	A	**_D_**	L	F	M	V	F	G	G	F	T	T	T	L	Y	T	S	L	H	G	101
Bac	38					D	A	K	K	F	Y	A	I	T	T	L	V	P	A	I	A	F	T	M	Y	L	S	M	L	L	G		63
D_2	68	T	T	N	Y	L	I	V	S	L	A	V	A	**_D_**	L	L	V	A	T	L	V	M	P	W	V	V	Y	L	E	V	V	G	98

TM3

	First	3.22	3.23	3.24	3.25	3.26	3.27	3.28	3.29	3.30	3.31	3.32	3.33	3.34	3.35	3.36	3.37	3.38	3.39	3.40	3.41	3.42	3.43	3.44	3.45	3.46	3.47	3.48	3.49	3.50	3.51	3.52	3.53	3.54	3.55	3.56	Last
Rho	107	P	T	G	C	N	L	E	G	F	F	A	T	L	G	G	E	I	A	L	W	S	L	V	V	L	A	I	E	**_R_**	Y	V	V	V	C	K	141
Bac	79			Y	W	A	R	Y	A	D	W	L	F	T	T	P	L	L	L	L	D	L	A	L	L	V	D	A	D	Q							105
D_2	104	R	I	H	C	D	I	F	V	T	L	D	V	M	M	C	T	A	S	I	L	N	L	C	A	I	S	I	D	**_R_**	Y	T	A	V	A	M	138

TM4

	First	4.38	4.39	4.40	4.41	4.42	4.43	4.44	4.45	4.46	4.47	4.48	4.49	4.50	4.51	4.52	4.53	4.54	4.55	4.56	4.57	4.58	4.59	4.60	4.61	4.62	4.63	Last
Rho	149	G	E	N	H	A	I	M	G	V	A	F	T	**_W_**	V	M	A	L	A	C	A	A	P	P	L	V	G	174
Bac	108						I	L	A	L	V	G	A	D	G	I	M	I	G	T	G	L	V	G	A	L		127
D_2	148	S	K	R	R	V	T	V	M	I	S	I	V	**_W_**	V	L	S	F	T	I	S	C	P	L	L	F	G	173

TM5

	First	5.35	5.36	5.37	5.38	5.39	5.40	5.41	5.42	5.43	5.44	5.45	5.46	5.47	5.48	5.49	5.50	5.51	5.52	5.53	5.54	5.55	5.56	5.57	5.58	5.59	5.60	5.61	5.62	Last
Rho	200	N	E	S	F	V	I	Y	M	F	V	V	H	F	I	I	**_P_**	L	I	V	I	F	F	C	Y	G	Q	L	V	227
Bac	137					W	W	A	I	S	T	A	A	M	L	Y	I	L	Y	V	L	F	F	G	F	T				157
D_2	186	N	P	A	F	V	V	Y	S	S	I	V	S	F	Y	V	**_P_**	F	I	V	T	L	L	V	Y	I	K	I	Y	213

TM6

	First	6.28	6.29	6.30	6.31	6.32	6.33	6.34	6.35	6.36	6.37	6.38	6.39	6.40	6.41	6.42	6.43	6.44	6.45	6.46	6.47	6.48	6.49	6.50	6.51	6.52	6.53	6.54	6.55	6.56	6.57	6.58	6.59	6.60	Last
Rho	245	K	A	E	K	E	V	T	R	M	V	I	I	M	V	I	A	F	L	I	C	W	L	**_P_**	Y	A	G	V	A	F	Y	I	F	T	277
Bac	167						V	A	S	T	F	K	V	L	R	N	V	T	V	V	L	W	S	A	Y	P	V	V	W	L	I	G			192
D_2	366	Q	K	E	K	K	A	T	Q	M	L	A	I	V	L	G	V	F	I	I	C	W	L	**_P_**	F	F	I	T	H	I	L	N	I	H	398

TM7

	First	7.32	7.33	7.34	7.35	7.36	7.37	7.38	7.39	7.40	7.41	7.42	7.43	7.44	7.45	7.46	7.47	7.48	7.49	7.50	7.51	7.52	7.53	7.54	7.55	7.56	7.57	Last
Rho	285	P	I	F	M	T	I	P	A	F	F	A	K	T	S	A	V	Y	N	**_P_**	V	I	Y	I	M	M	N	310
Bac	205	T	L	L	F	M	V	L	D	V	S	A	K	V	G	F	G	L	I	L	L	R	S					226
D_2	405	P	V	L	Y	S	A	F	T	W	L	G	Y	V	N	S	A	V	N	**_P_**	I	I	Y	T	T	F	N	430

*residues highlighted in bold italics are highly conserved across all GPCRs, they mark the X.50 point (where X is the transmembrane number).

3A.3.1 Analysis of Bacteriorhodopsin and Rhodopsin

The statistics of the phi and psi region plots for the bacteriorhodopsin structure indicate a very good stereochemical quality of the protein structure, see Table 3.2, where a good quality structure has over 90% of residues in most favoured regions[37,46]. The procheck statistics of the bacteriorhodopsin structure are clearly much better than those of rhodopsin. However this is expected, as the resolution of this crystal structure is 1.55Å compared to 2.8Å for rhodopsin. Although the statistics of the rhodopsin model are not as good, it was the only X-ray crystal structure of a true GPCR, when this work was started. Therefore both structures were acceptable starting points for further analyses.

Table 3.2. Ramachandran plot statistics of residues from the bacteriorhodopsin structure 1C3W and rhodopsin structure 1F88 from the Protein DataBase.

	% in most favoured regions	% in additional allowed regions	% in generously allowed regions	% in disallowed regions
bacteriorhodopsin	97.9	2.1	0.0	0.0
rhodopsin	77.7	20.3	1.7	0.3

3A.3.2 Generating the Alignment of D_2 onto Bacteriorhodopsin and Rhodopsin

3A.3.2.1 Construction of the Threader database file

Analyses of the crystal structure of rhodopsin revealed two chain breaks between

- A235 and S240, missing residues QQQE.
- P327 and S334, missing residues LGDDEA.

Break 1 is located eight residues after H5 in the cytoplasmic loop C3. The complete C3 loop, although not included in the original model, is said not to fold over the helical region. Break 2 is located 4 residues after the cysteine linkages at the end of helix loop H8. This missing section is said to cover H8 from the solvent region[42]. This was taken into consideration when remodelling these sections of the protein.

Additional to the two chain breaks present in the crystal structure, there were a number of incorrect residues present in the structure. T335 was missing the sidechain of this residue and S334 was missing the hydroxy group of its sidechain. These were added according to allowed chi angles of the sidechain and steric constraints of the protein. The resulting structure was then subjected to local minimisations in the areas where the structure was modified and then run through procheck again.

The RMS fit of the modified and original rhodopsin models, see Table 3.3, show an extremely close match over the 333 residues that the structures had in common, the main discrepancies occurring near the chain breaks where the additional residues were placed.

Table 3.3. Modified structure of rhodopsin compared to the original crystal structure.

	RMS fit (Å^2) of Cα carbons	RMS fit (Å^2) of whole structure
Modified structure	0.0975	0.1580

3A.3.2.2 Running Threader

Usually when analysing the statistical output from Threader the weighted pairwise energy Z-score, the ((energy - mean) / standard_deviation), is used as the primary field for selecting a correct match, and a value greater than two would indicate a possible match. The Z-score is a standardised score with predefined limits set from analyses using the complete fold library. Therefore when using a small library of folds, as in this case, it is stated the Z-scores from pairwise energy may not be a useful measure of significance and checking the core-shuffled Z-scores is advised[33].

Both the mean and minimum core-shuffled Z-scores have the implication that a large positive value indicates a correct match and a negative value indicates probably an incorrect match. The native core-shuffled Z-score denotes how low the model energy is in comparison to the native energy of the template structure, where a value greater than one indicates a probably incorrect match. From analysis of the core-shuffled Z-scores obtained, see Table 3.4, 1F88 is a better match than that of 1C3W, and that the latter may be incorrect. 1F8800 is the threader database file created from the crystal structure

of rhodopsin[42]. The poor alignment of the sequence of D_2 onto bacteriorhodopsin is most likely due to the low sequence homology, ~6-11%[5], between these two whereas the alignment of D_2 onto rhodopsin seems more plausible due to the >25% sequence homology within transmembrane segments.

Table 3.4. Statistical output from Threader where the sequence of D_2 is threaded onto the threader database file of 1C3W, created in the procedure outlined in the method.

pairwise energy	Z-score (native) core-shuffled	Z-score (mean) core-shuffled	Z-score (minimum) core-shuffled	Z-score (mean) pairwise energy	Z-score (weighted) pairwise energy	% of structure aligned	% of sequence aligned	Pdb code
-490.73	-0.84	2.72	0.64	0.86	0.87	85.9	67.5	1f8800
-332.53	-4.29	2.24	-0.52	-9.99	-9.99	98.7	53.0	1c3w00

* Only 9 of the 17 columns are shown

A recent practical evaluation of comparative modelling and threading methods[12] states that where there is low homology between the target and template sequences, threading performs much better than fully automated comparative modelling. The example used in the paper[12] having 7% sequence homology resulted in a model with a RMS deviation of 8Å using standard threader methods. Fully automated comparative modelling involves using Modeller to generate a tertiary structure without the aid of an alignment file. As the bacteriorhodopsin based model under construction was also of low sequence homology, the alignment that was returned from threader was carefully examined in comparison to previous alignments obtained by other groups[6,7]. The threader alignment of D_2 onto 1C3W was not at all similar to other predicted alignments and the statistics indicated an incorrect alignment, therefore this alignment was discarded. The alignment presented in the paper by Teeter et al.[7] was used as a starting point and was compared to those proposed by Trump-Kallmeyer et al.[6]. Most transmembrane helices were very similar with the exception of helices three and four. The alignment of Teeter differed by four residues in TM3 and by seven residues in TM4, compared to that of Trump-Kallmeyer. As comparisons were being made to the theoretical model 1I15, constructed by Teeter et al., the alignment presented by Teeter et al. was used in conjunction with comparative modelling, not the fully automated comparative modelling procedure of Modeller.

The alignment of D_2 onto rhodopsin returned from Threader was of remarkable similarity to the alignment proposed by Trump-Kallmeyer. The only changes to the alignment were that TM4 had to be shifted two residues so as to align the highly conserved tryptophan ($Trp^{4.50}$), and TM6 had to be shifted one residue so as to align the highly conserved tryptophan ($Trp^{6.48}$) and proline ($Pro^{6.50}$) residues. Importantly, the cysteine in the second extracellular loop (Cys^{E2}) had to be adjusted to align with the corresponding cysteine (Cys^{E2}) of rhodopsin that is involved in the disulfide linkage with the highly conserved TM3 cysteine ($Cys^{3.25}$). This disulfide linkage constrains the second extracellular loop (E2) of rhodopsin and D_2. In the crystal structure of rhodopsin this loop folds deeply into the centre of rhodopsin and places this loop in contact with retinal. This is thought to be the case for D_2 and most other GPCRs, with this loop being implicated as being proximal to the binding site[42].

The resultant alignment of D_2 onto rhodopsin (Figure 3.2), shown in single amino acid nomenclature, shows a high degree of hydrophobicity in the transmembrane regions, as would be expected for membrane bound receptors.
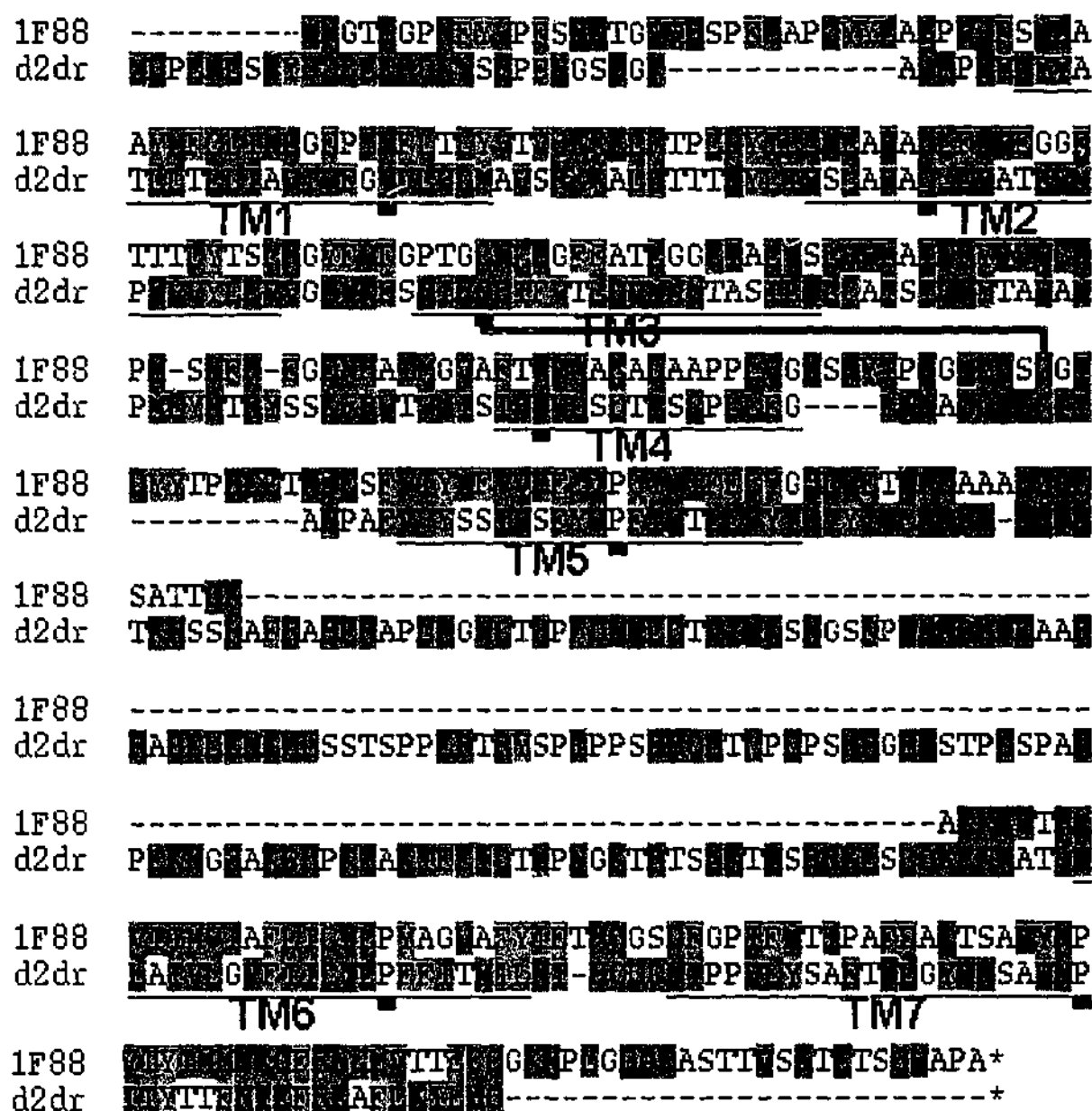
Figure 3.2. *Final alignment of dopamine D_2 sequence with rhodopsin 1F88 after adjusting the alignment output from Threader. The highly conserved residues used in the alignment of the sequences are marked with a small black box beneath them, these residues correspond to X.50, where X is the helix number. A blue line connecting the corresponding cysteines shows the highly conserved disulfide linkage between extracellular loop E2 and TM3. The transmembrane regions are shown sequentially with a thin black line below them. Gaps within the sequence alignments are represented by a -.*

3A.3.3 Three-dimensional D_2 Model Generation

As fully automated comparative modelling does not result in an accurate structure where there is low sequence homology[112], comparative modelling using the previously

aligned sequence of Teeter for bacteriorhodopsin and the modified alignment of rhodopsin was undertaken to generate the tertiary structures.

3A.3.3.1 Adjusting the alignment file

The alignment files from rhodopsin and bacteriorhodopsin must be further altered to include breaks within the sequence alignments, where no alignment takes place. In the alignment file construction over 150 non-aligned residues between helices 5 and 6 were removed. This was because Modeller is unable to deal with more than 100 non-aligned residues and the residues were of no significance to the models being constructed.

3A.3.3.2 Additional Constraints

The disulfide linkage between $C399_{D2}$ and $C401_{D2}$, that has been shown to exist in D_2 dopaminergic receptor[19], was added as a constraint. Further constraints were added in construction of the model with dopamine bound in the active site these were obtained from the D_2 mutagenesis paper of Mansour et al.[23] and the modelling paper by Teeter et al.[7]. The m-hydroxy and nitrogen of dopamine were constrained to be within hydrogen bonding distance of $Ser^{5.46}$ and $Asp^{3.32}$ respectively.

3A.3.4 Analysing the bound and unbound receptor states of the bacteriorhodopsin based models in comparison to 1I15

The modeller function initially serves as a comparison within each set of bound and unbound models produced, thus if a model with an objective function a few orders of magnitude higher is produced, it could quickly be seen incorrect. The quality of the models produced, judged by examination of the modeller objective function, does vary slightly amongst each set of models, see Table 3.5. However there are no orders of magnitude difference in each set of bound and unbound models produced. The quality of the bound receptor models does appear to be much lower, as indicated by a large objective function. However the inclusion of the ligand, dopamine, is responsible for this deviation, as the modeller objective function tries to accommodate dopamine into spatial restraints where none exist.

Table 3.5. Modeller objective function of bacteriorhodopsin-based models for both the bound and unbound states

Model	unbound receptor	bound receptor
1	1382	132530
2	1819	94988
3	1476	132087
4	1582	66328
5	1797	117040
mean ± S.D.	1611 ± 193	108594 . 28141

Table 3.6. RMS fit for bacteriorhodopsin based models onto the theoretical D_2 model 1115, prior to minimisation.

Model	RMS fit for unbound receptor		RMS fit for bound receptor	
	C_α carbons	All atoms	C_α carbons	All atoms
1	1.7503	2.8172	1.6795	2.5499
2	1.7195	2.7283	1.6494	2.5241
3	1.6447	2.5327	1.6853	2.5790
4	1.7181	2.6565	1.7225	2.6343
5	1.7543	2.6327	1.7208	2.6286
average	1.7174	2.6735	1.6915	2.5832

Table 3.7. RMS fit for minimised bacteriorhodopsin based models onto the theoretical D_2 model 1115

Model	RMS fit for unbound receptor		RMS fit for bound receptor	
	C_α carbons	All atoms	C_α carbons	All atoms
1	2.0136	3.1180	1.7223	2.5788
2	1.7135	2.7440	1.8394	2.7274
3	1.8523	2.6244	1.6877	2.6127
4	1.9448	2.8082	1.9801	2.7409
5	1.9561	2.6791	1.7865	2.7002
average	1.8961	2.7947	1.8032	2.6720

In their original comparative protein modelling paper, Sali and Blundell[111] state that minimisation does not significantly improve the overall quality of the model produced. Tables 6 and 7 are consistent with this observation as the RMS fit of the models decreases after minimisation. The slight decrease in RMS fit of the models, ~0.2Å on average, is a small price to pay for the dramatic improvement in energy of the models in

question. The energies of the models can improve by up to 5 orders of magnitude, see Table 3.8. In unbound model 1 there is approximately a 90,000 kcal/mol improvement from minimisation of the hydrogens, over 2000 kcal/mol improvement in sidechain minimisation and over 1000 kcal/mol improvement in the final step of the minimisation where no atoms were constrained. These data conform to previous observations by Sali and Blundell[11] where they indicate that most of the improvement in energy of the model is due to relaxation of the structures that were strained in some positions by the addition of the hydrogen atoms. The difference in energy of the bound and unbound models originates from the fact that some hydrogen are added to the unbound models in a strange fashion. For example hydrogens added to the histidine residues are not added at correct bond angles or bond lengths, resulting in unusually high-energy structures.

Table 3.8. Calculated energies (kcal/mol) of models using Kollman all atom charges and Kollman force field before minimisation and after minimisation.

Models	Unbound Pre-minimisation	receptor Post-minimisation	Bound Pre-minimisation	Receptor Post-minimisation
1	95240	-2340	18173	-2494
2	51195	-2379	14578	-2424
3	76681	-2372	15605	-2403
4	73937	-2367	10620	-2433
5	79231	-2316	12795	-2433

Tables 6 and 7, are in agreement with previous observations by Schoonman[12] who state that a more accurate model is obtained if the model is built with a ligand in the active site. This method of building a model is said to be better as it stops the unrealistic filling of active site areas. This artefact of Modeller is seen when comparing the RMS deviation between the bound and unbound receptor states, with an improvement of ~0.1 Å^2 over bound compared to unbound models.

3A.3.4.1 Procheck stereochemical quality check

All the models constructed have a good stereochemical quality with over 90% of residues in most favoured regions (Figure 3.3). The residues implicated in the

disallowed regions are located at the ends of the helices and are not part of either the proposed binding site or the sodium-binding site[7].

phi/psi region plot statistics
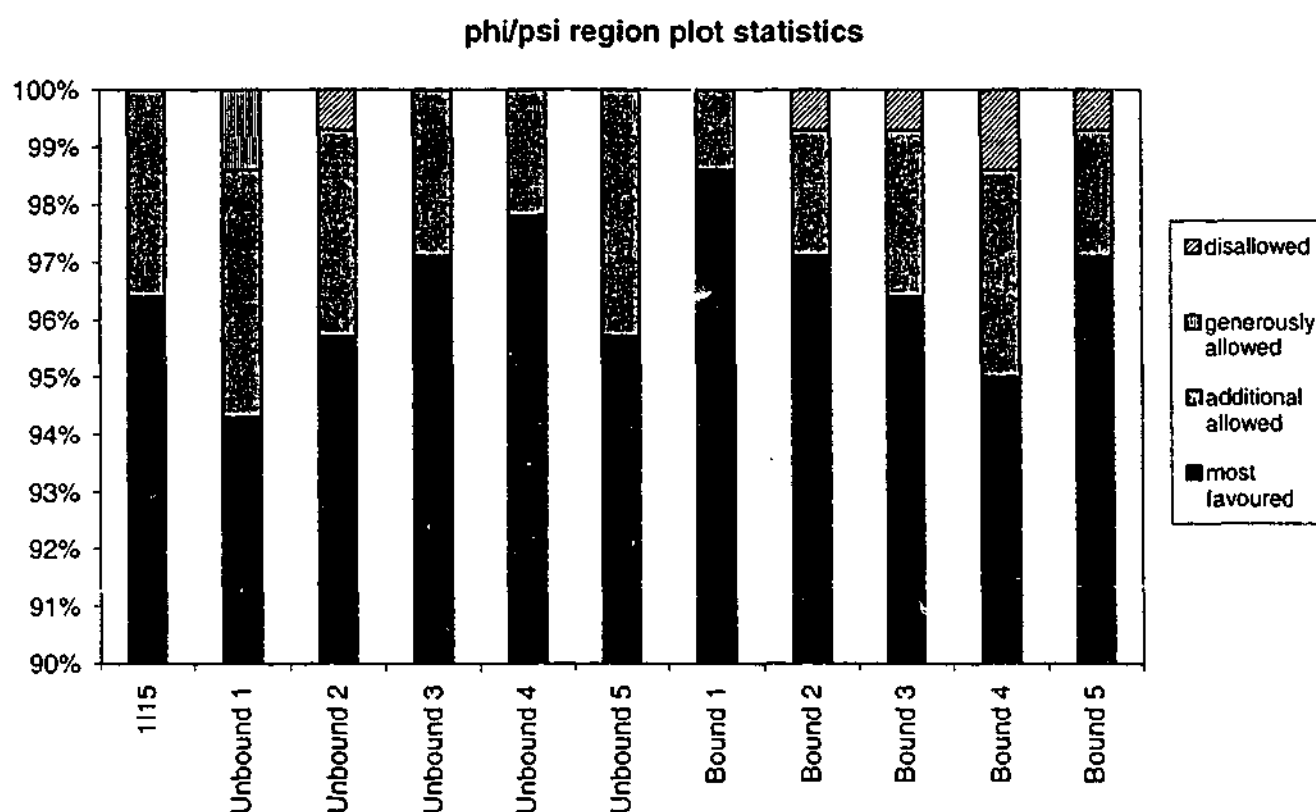
Figure 3.3. Procheck phi and psi region plot statistics of residues of the five bound and unbound bacteriorhodopsin based D_2 models and the theoretical model 1I15.

3A.3.4.2 Proline Kinks

Two of the major drawbacks of using a template in the development of a GPCR model are the alignment of the sequence over helices where there is a kink due to the presence of a proline residue and also trying to force a proline kink where there is none in the template model. Proline kinks were generated in 1I15 by designing a template for helical proline residues from the bacteriorhodopsin helices that contained proline kinks. Areas around the proline residues were then minimised in a local minimisation procedure resulting in removal of most close contacts and a proline kink. Removal of proline kinks was done in a similar fashion using a local minimisation procedure on residues that had been aligned onto bacteriorhodopsin proline kinks resulting in a straightened helix[7]. One of the pitfalls of using automated modelling packages is that they use ideal helix angles[7], which would result in an elongated helix. The method used by Modeller for establishing and removing proline kinks comes from satisfaction

of spatial and empirical restraints derived from the initial alignment, a database of protein structures as well as various other factors. Thus the procedure used by Modeller should result in similar lengths of helices and similar distances between proline i and $i\text{-}4$ residues as were obtained by Teeter *et al.*, as Modeller does not primarily use ideal helical angles. Tables 9 and 10 show that the method employed by Modeller in deriving and removing proline kinks does agree very well with the method employed by Teeter *et al.* The one exception to this, see Table 3.9, is the removal of the proline kink at Phe$^{6.52}$. Although the kink is removed, the distance between i and $i\text{-}4$ still remains relatively large which is indicative of a kink. Not surprisingly there is a proline residue, Pro$^{6.50}$, involved in a kink just two residues away from Phe$^{6.52}$ which is forcing the residues further apart than expected. Interestingly the distances between i and $i\text{-}4$ residues for the placement of proline residues do increase after minimisation and, vice versa, decrease after minimisation for i and i-4 residues in areas of removal of the proline residues. These data indicate that some minimisation after generation of the initial model is advisable, even though most initial i and $i\text{-}4$ values are similar to those obtained by Teeter.

Table 3.9. Average distance (Å) between C_α carbons of i and i-4 residues corresponding to either the insertion or removal of a proline residue for all bound models.

Residue	1I15 (Teeter model)	pre-minimisation	post-minimisation
Pro$^{2.59}$	7.420	6.945	7.681
Pro$^{5.50}$	6.185	6.590	6.969
Pro$^{6.50}$	6.997	7.003	7.422
Ala$^{2.54}$	7.362	7.342	6.043
Thr$^{3.37}$	6.474	6.949	6.558
Phe$^{6.52}$	5.832	7.720	7.877

Table 3.10. Average length (Å) from first to last C_α carbon of helices containing proline kinks from all bound models.

Helix	1I15	Constructed Models pre-minimisation	post-minimisation
2	38.1	37.7	37.8
5	30.4	28.7	28.9
6	37.1	36.6	36.4

3A.3.4.3 Secondary Structure Prediction

The secondary structure of the models was predicted using "assign secondary structure" within Sybyl, a method that uses standard Kabasch and Sander assignment criteria. Of the five bound models only one model had all seven helices predicted correctly, while the other 4 models had at least one break within one helix. In the four models with breaks within helices, every model had a break of alpha helical periodicity at $Tyr^{7.43}$ in TM7. This break in alpha helical nature is located just above the sodium-binding site, which may be what is causing these phenomena, as a number of electronegative residues in the area are not counterbalanced by a sodium ion or water molecules during minimisation. Another break in alpha helical nature occurs in two models in TM6 at residues $Leu^{6.49}$ - $Pro^{6.50}$. The presence of proline here is the cause of this break.

3A.3.4.4 Proposed Binding Site

In all the D_2 models constructed without any ligand in the active site, a number of residues are incorrectly placed. The tryptophan ($Trp^{4.50}$) has a different *chi* angle to 1I15 resulting in the sidechain facing out into the lipid bilayer and thus not matching the proposed model. SCAM data indicates that this residue, $Trp^{4.50}$, is in a water accessible environment and not facing out into the lipid bilayer[29]. This differing placement of $Trp^{4.50}$ in the unbound models also enables other residues to encroach further into the binding site area resulting in a few more slight differences between these models. This poor arrangement of the binding site residues in the unbound models is expected, as it has been proposed earlier that the model should be built with a ligand in the active site in order to stop unrealistic filling of the binding site[12]. When this is the case and models are constructed with the ligand in the active site, the resulting models are of remarkable similarity to each other (Figure 3.4).
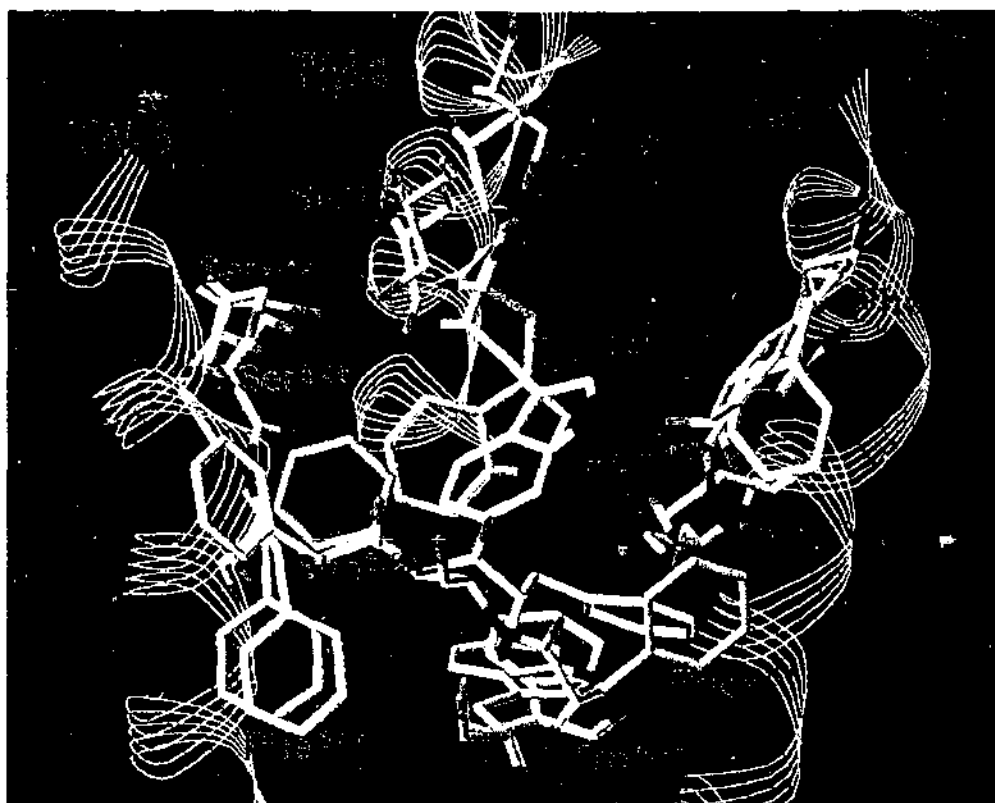
Figure 3.4. *Residues implicated in binding site from Teeter et al. Orange residues are from 1115 (Teeter et al.) and the white residues are from a model generated with dopamine bound in the proposed active site. For clarity only helices 3, 4 and 5 are shown from the model constructed.*

The RMS fit for the 12 residues implicated in the binding site is 1.80Å for all atoms present. The two residues that deviate most from the 1115 model, are $Phe^{6.52}$ and $Trp^{4.50}$, and it can be seen from Figure 3.4 that they only differ in the orientation of their respective aromatic planes. This deviation is not to such a great extent as to cause problems in the packing of the other residues, as was the case for the active site of the unbound models.

3A.3.4.5 Proposed Sodium-Binding Site

G protein coupled receptors interact with the G proteins, $G\alpha_i$ and $G\alpha_o$, to modulate several intracellular effectors and subsequent second messenger systems. An important feature of $G\alpha_i$-coupled receptors, such as the D_2 dopaminergic receptor, is that they are sensitive to pH and sodium. Sodium has been shown to enhance the affinity of some antagonists and reduce the binding affinity of agonists and the benzamide family of antagonists[52,53] at the D_2 receptor. Central to the D_2 dopaminergic receptors sensitivity

to sodium is $Asp^{2.50}$, it has been shown that mutation of this residue abolishes the receptors sensitivity to sodium[53]. Therefore the putative sodium-binding site as identified by Teeter *et al.* was examined in relation to the models constructed.

The RMS fit to the Teeter *et al.* model for the residues implicated in the sodium-binding site is 1.144 Å for all atoms present. Modifying the chi angles on a number of residues (Figure 3.5) could easily further reduce the deviation between the models produced and the theoretical structure 1115. The two residues $Asn^{7.49}$ and $Asp^{2.50}$ could easily have their side chain angles modified to fit well with the proposed model at little energy cost. The pyramidal sodium-binding site is defined by $Asp^{2.50}$, $Ser^{3.39}$, $Asn^{3.42}$, $Ser^{7.46}$ and $Asn^{7.49}$ at each vertex of the site.
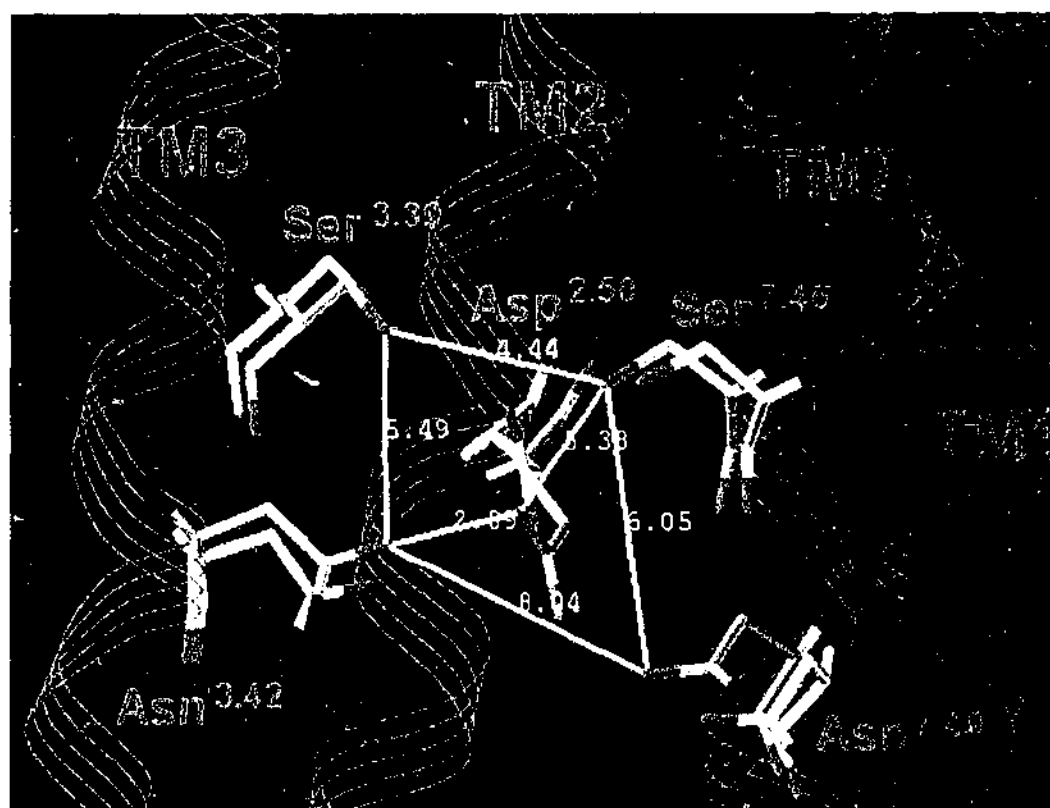
Figure 3.5. *Lateral view of superimposition of proposed sodium-binding site of bacteriorhodopsin based D_2 model, residues shown in white and helices shown in magenta line trace, and of 1115 (residues shown in orange). Sodium and helices 4, 5 and 6 have been removed for clarity.*

3A.3.5 Analysing the bound and unbound receptor states of rhodopsin based models relative to bacteriorhodopsin based models.

The quality of the rhodopsin based models produced, judged by examination of the modeller objective function, see Table 3.11, is not as high for each set as those obtained for the bacteriorhodopsin models. The higher objective function for each set is due to the fact that the D_2 models based on rhodopsin are being aligned to a greater number of residues. The average of the objective function is 2370 for unbound rhodopsin models compared to 1611 for unbound bacteriorhodopsin models. The average of the objective function for the bound rhodopsin models is 176946 compared to 108594 for bound bacteriorhodopsin models. Again the higher objective function for bound models compared to unbound models comes from trying to accommodate dopamine into the spatial restraints where none exist.

Table 3.11. Modeller objective function of rhodopsin based models for both the bound and unbound states

Model	unbound receptor	bound receptor
1	2770	133917
2	2336	187197
3	1829	192452
4	2959	176946
5	1954	163930
mean ± S.D.	2370 ± 490	176946 ± 23361

3A.3.5.1 Procheck stereochemical quality check

All the rhodopsin based models constructed have a fair stereochemical quality with over 90% of residues in most favoured and allowed regions (Figure 3.6). Most residues implicated in the disallowed regions are located in the intra and extracellular loops as well as at the ends of the helices and are not part of either the proposed binding site. A few models constructed, unbound model 1 and bound model 3, did have residues in disallowed regions that were in close vicinity to the proposed sodium-binding site and proposed active site respectively.

phi/psi region plot statistics
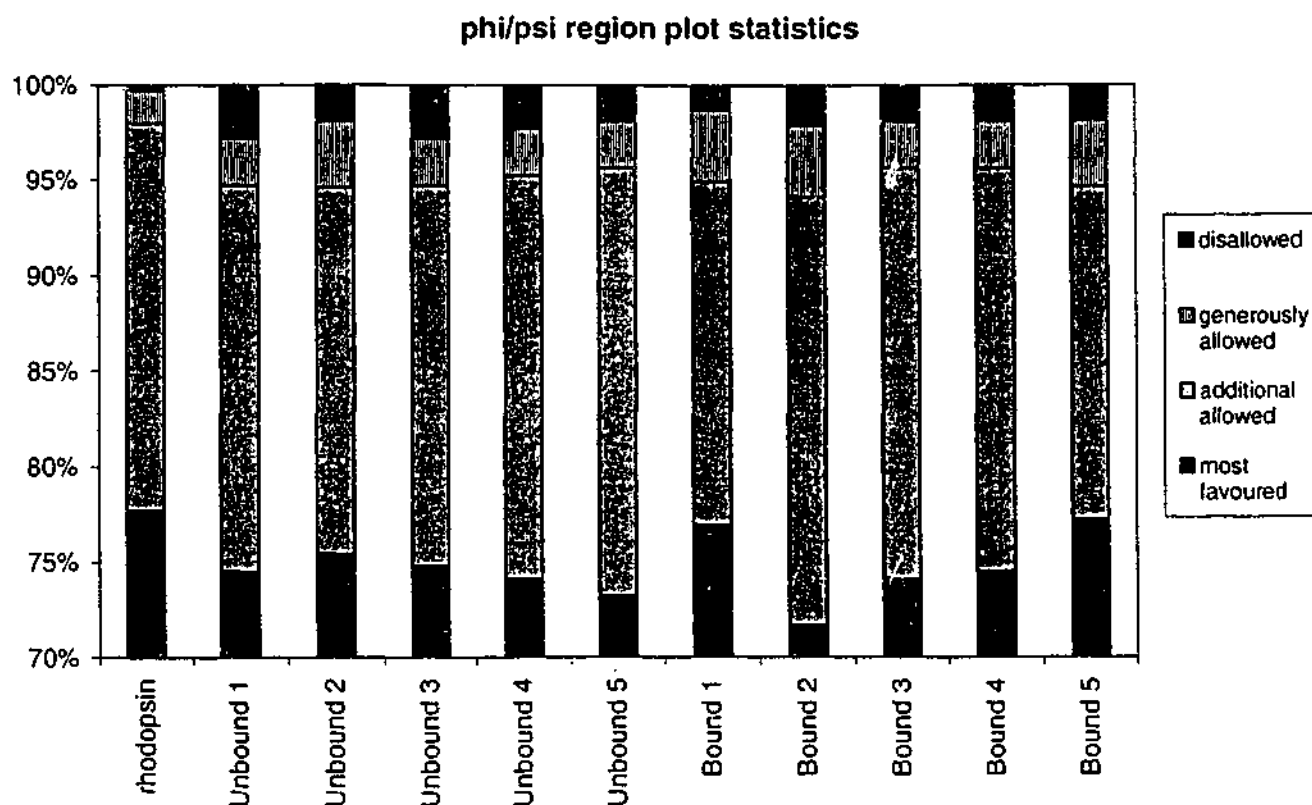
Figure 3.6. *Ramachandran plot statistics from the five bound and unbound rhodopsin based D_2 models and the rhodopsin crystal structure they are based upon.*

3A.3.5.2 Secondary Structure Prediction

The secondary structure of the models was predicted using "assign secondary structure" within Sybyl, a method that uses standard Kabasch and Sander assignment criterion. Out of the five bound models not one model had all seven helices predicted as complete helices. Every model produced had a break in alpha helical periodicity in TM2 and TM5 at residues $Leu^{2.56}$ - $Val^{2.57}$ and $Ser^{5.46}$ - $Phe^{5.47}$, respectively. Both of these breaks in alpha helical nature correspond to *i-4* regions where a proline is present in the sequence being aligned, in addition to this the two breaks both correspond to a π-helical segment in the rhodopsin structure[54]. The π-helical segment of TM5 in rhodopsin is due to the presence of a proline in the *i-4* region of the structure. The π-helical segment of TM2 in rhodopsin is due to the presence of two sequential glycines, $Gly^{5.46}$ and $Gly^{5.47}$, in the sequence. Thus it is not unusual that these two breaks in alpha helical nature of the receptors being constructed are present and actually correspond to π-helical segments. Another area where in two of the five models constructed there is a break in alpha helical nature is in TM7 located around $Gly^{7.42}$ - $Val^{7.44}$. This break is probably due to Modeller trying to align the D_2 sequence over an extended "eclipsed" or

"Pauling" 3_{10}-helix that is present in the rhodopsin model[54] from $Pro^{7.38}$ - $Val^{7.47}$. Additionally this break in alpha helical nature in TM7 is located just above the sodium-binding site, which could be causing this phenomenon, as the large number of electronegative residues in the area, even though being counterbalanced by a sodium ion, could still be causing problems. Additionally kinks in helices 2, 6 and 7 force the top half of their helices to bend into the plane of the page in a clockwise fashion. Helices 3 and 5 are also entirely tilted into the plane of the page in a clockwise fashion (Figure 3.7). The models produced also show that the antiparallel beta sheets present in extracellular loop E2 of the rhodopsin model are also present in the models constructed. This extracellular loop penetrates deep into the helices and is within six angstrom of the aspartate residue that is implicated in binding compounds, which is consistent with mutagenesis studies implicating this loop as being in close proximity to the binding site[29]. The loop is held in place by a highly conserved disulfide bond that is between the top of TM3 and the extracellular loop E2, (Figure 3.7).

Figure 3.7. Cartoon display of secondary structure of rhodopsin based model, using "assign secondary structure" from within Sybyl. $Asp^{3.32}$, shown in space filling mode, and the highly conserved disulfide linkage between loop E2 and TM3 is also shown in capped stick mode. Breaks in α-helices two, five and seven can be seen at $Leu^{2.56}$ - $Val^{2.57}$, $Ser^{5.46}$ - $Phe^{5.47}$, and $Gly^{7.42}$ - $Val^{7.44}$, respectively.

3A.3.5.3 Proposed Binding Site

In all the D_2 models constructed, with and without any ligand, there was a very similar arrangement of residues in the proposed binding site. There were not the large differences seen between the bound and unbound bacteriorhodopsin based models, due to unrealistic filling of the active site. However in the unbound models we found that the hydrophobic arms of two isoleucine residues, from loop E2, orientated themselves into the hydrophobic binding cavity and hence displaced $His^{6.55}$ relative to the bound models. When dopamine was present $His^{6.55}$ was able to penetrate further into the binding cavity and be within binding range of a suitable ligand. $His^{6.55}$ has been proposed to form a hydrogen bond with the 5-sulfonamide of the antagonist sulpiride[28].

The main differences between the rhodopsin active site and the bacteriorhodopsin active site (Figure 3.8) were the placement of TM4 and its residues and also the placement of $Trp^{6.48}$ on TM6. However in D_2 models constructed with antagonists in the binding site we have seen $Trp^{6.48}$ on TM6 adopt a similar conformation to that of bacteriorhodopsin-based models, forming the base of the binding cavity. When fitting the active sites together a RMS fit of 2.217Å for backbone atoms was obtained when excluding the residues on TM4. The residues on TM4 of the rhodopsin model are located approximately 1-2 turns further down the helix and in the case of $Trp^{4.50}$ on rhodopsin are pointing into the lipid bilayer. Additionally the axis of helices three, five and partially six of the rhodopsin model are tilted clockwise into the page when viewing them from the extracellular side, whereas the axes of helices on the bacteriorhodopsin model are not tilted at all.
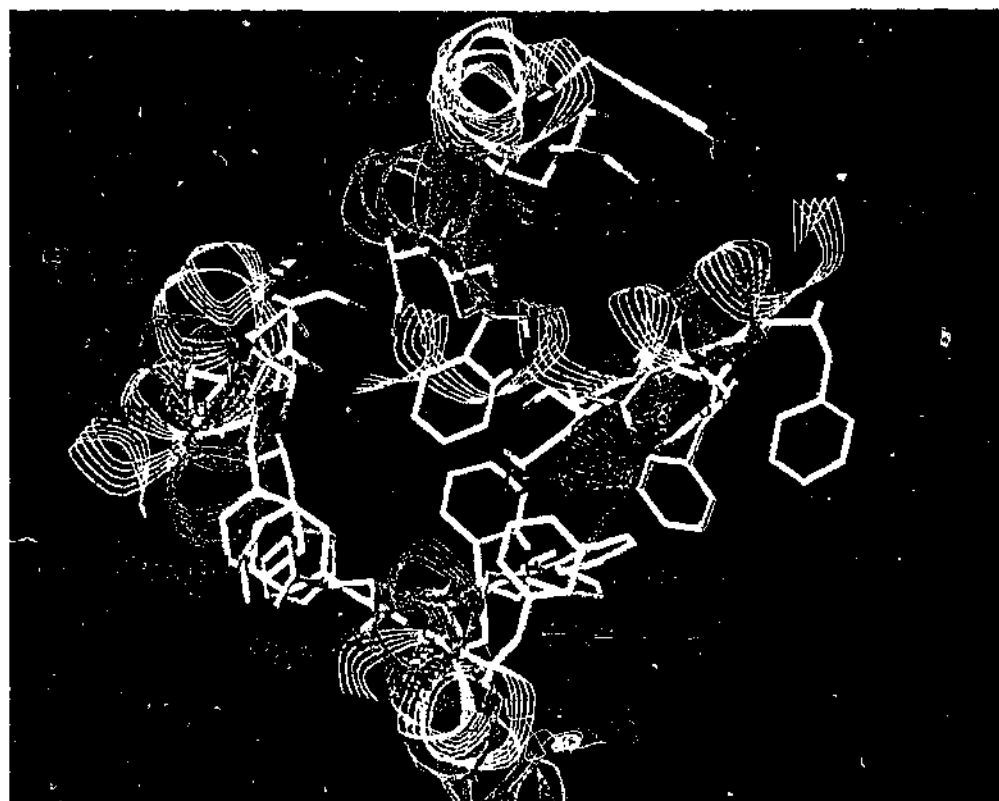
Figure 3.8. *Extracellular view of the active site of bacteriorhodopsin-based D_2 model, residues shown in orange and helices shown in pink line traces, superimposed over the active site of rhodopsin-based D_2 model, residues shown in white and helices shown in yellow line traces. For clarity helices one, two and seven are omitted.*

3A.3.5.4 Proposed Sodium-Binding Site

The D_2 dopaminergic receptor proposed sodium binding site and in particular $Asp^{2.50}$ was examined in relation to previous bacteriorhodopsin model. The RMS fit for the residues implicated in the sodium-binding site is 2.815Å for backbone atoms. The deviation between the rhodopsin and bacteriorhodopsin based D_2 models produced, (Figure 3.9) is mainly due to $Asp^{2.50}$ on the rhodopsin-based model being approximately a turn higher on TM2. As the $Asp^{2.50}$ is a turn higher on TM2 it makes the distance between this residue and $Asn^{3.42}$ too large for the $Asn^{3.42}$ to be involved in the sodium-binding site. This data in consistent with recent modelling of the D_2 sodium binding site were it is proposed that $Ser^{3.39}$, $Ser^{7.46}$ and $Asp^{2.50}$ are the crucial residues involved in the binding of sodium[55]. However the two models still show a remarkable similarity considering the different basis used to produce both models (Figure 3.9).

Figure 3.9. *Superimposition of sodium binding site of rhodopsin based D_2 model,*
residues shown in white and helices shown in magenta line trace, onto sodium binding
site of bacteriorhodopsin based D_2 model 1115, residues shown in orange. Helices 4, 5
and 6 have been removed for clarity. Oxygen and nitrogen are shown in red and blue,
respectively.

3A.3.5.5 Substituted Cysteine Accessibility Method (SCAM)

The SCAM method (Figure 3.10) identifies water accessible residues within the binding
site cavity by assessing the binding of methanthiosulfonate to the ionised form of
cysteine. This binding to the ionised form of cysteine takes place approximately one
billion times faster than to the unionised form. As cysteine is more likely to be ionised
in an aqueous environment, this method indicates on the cysteine substituted residues
where this reaction takes place, the likely location of an aqueous environment.
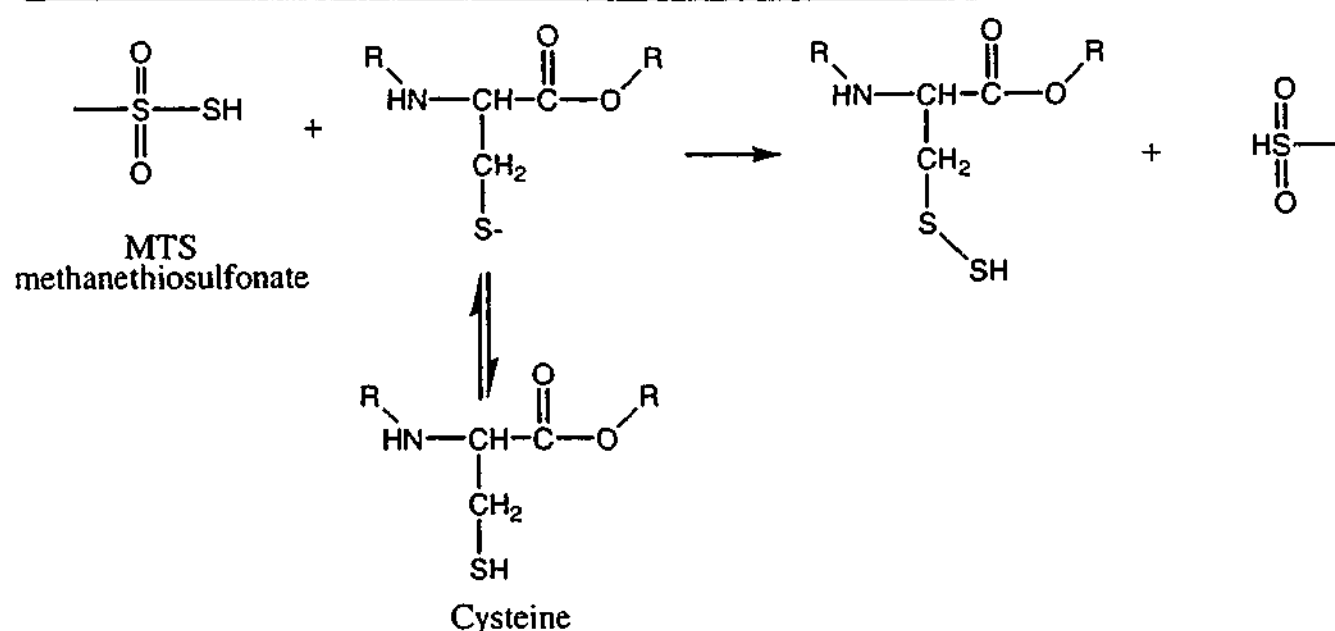
Figure 3.10. Reaction of MTS to ionised cysteine.

Of the *circa* 30 residues identified as being in an aqueous environment[19-22,25,27-29], about twelve seem not to be facing into the centre of the receptor as expected, but out into the lipid bilayer (Figure 3.11). Of these twelve residues most are located on the lipid side of TM5 that is thought to unravel during activation of the receptor[20]. The other residues that do not comply with the data collected from this technique are mainly located on the intracellular half of TM7, $Ala^{7.37}$, $Asn^{7.45}$, $Asn^{7.49}$, facing the lipid bilayer along with $Pro^{2.59}$ from TM2 and $Ile^{6.39}$ from TM6 that are also facing the lipid bilayer. The data from the SCAM studies was exactly the same for 1I15 as for the bacteriorhodopsin-based models when the data were applied to this theoretical model.

The non-compliance of a number of residues could be for a number of reasons, these being that the cysteine substitution could alter the secondary structure of the protein making the substituted cysteine water accessible whereas in the wild type protein this may not be the case. The more likely reason for this discrepancy in the models constructed here is that the bacteriorhodopsin structure these models are based on may not be the ideal choice.

Figure 3.11. *Tube depiction of helical wheels of bacteriorhodopsin based D_2 model showing SCAM affected residues coloured in red and unaffected residues in white, viewed from the extracellular side of the receptor.*

The rhodopsin-based model has fewer SCAM residues facing out into the lipid bilayer (Figure 3.12) than the bacteriorhodopsin-based model. Again the residues on the top of TM5, that is proposed to unravel, are all water accessible. The significant differences between the two models are that the residues on TM7 in the rhodopsin model that are water accessible are predominantly facing into the receptor whereas this is not the case for bacteriorhodopsin based models. Also another major difference is that $Trp^{4.50}$ on the rhodopsin model is facing out into the lipid bilayer whereas in the case of bacteriorhodopsin this is not the case as SCAM data suggests. However preliminary studies whereby $Trp^{4.50}$ is substituted for Cys and minimised inside a bilayer show that the cysteine residue does indeed orientate itself inwards and away from the lipid bilayer, thus making it solvent accessible.
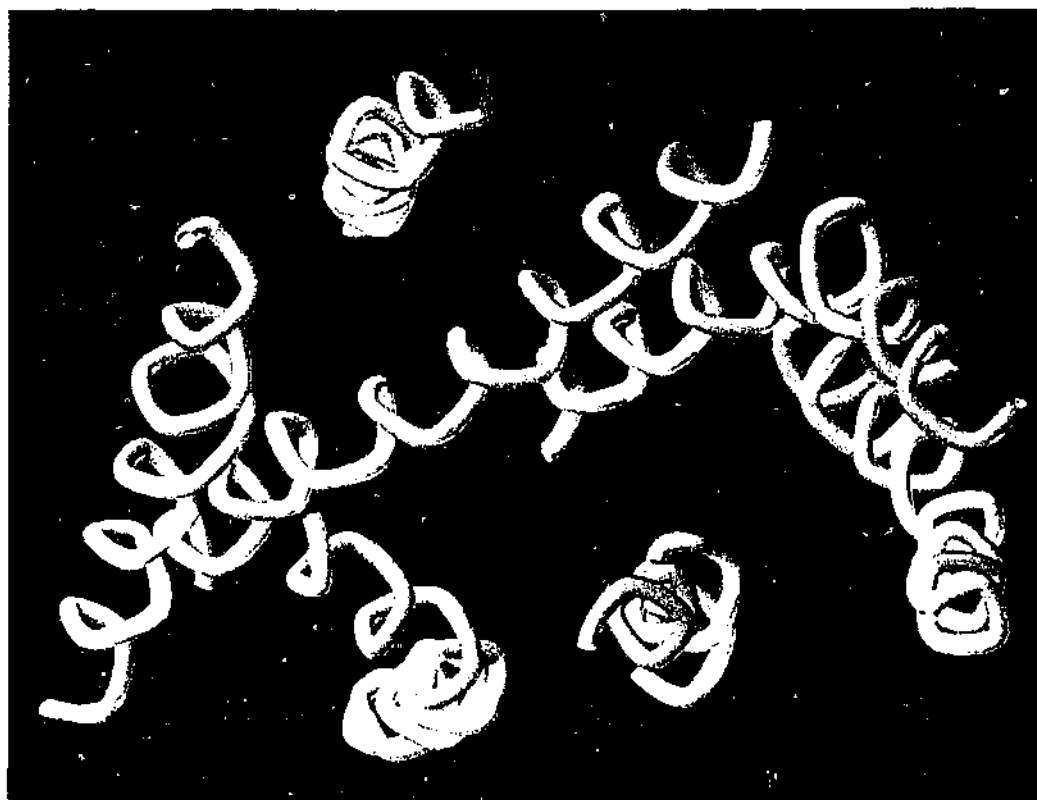
Figure 3.12. *Tube depiction of helical wheels of rhodopsin based D_2 model showing SCAM affected residues coloured in orange and unaffected residues in white, viewed from the extracellular side of the receptor.*

3A.4 CONCLUSION: HOMOLOGY MODELLING

From the analyses conducted it has been shown that the use of comparative modelling techniques to build GPCRs is a viable technique. One downfall of the technique is that models constructed without any ligand in the active site are not as good as models built with the ligand in the active site, due to unrealistic filling of the active site. This unrealistic filling of the active site in unbound bacteriorhodopsin-based models resulted in structures that had a higher RMS deviation and had different arrangements of residues implicated in binding compared to bound models. All bacteriorhodopsin-based bound models constructed compared very favourably with 1I15, with similar residue arrangements around the ligand and sodium binding sites. The bound and unbound models also had similar proline kinks within their helices that were of comparable lengths to 1I15. It should be noted that the agonist dopamine is not the most suitable compound to be built into the inactive state of bovine rhodopsin and bacteriorhodopsin, which are both thought to resemble the "antagonist bound" state. However the presence of dopamine in the active site does stop the unrealistic filling of the site.

The bacteriorhodopsin and rhodopsin-based models compared well to one another, with remarkable similarity around important areas such as ligand and sodium binding sites. The rhodopsin based models, however, were the better set of models of the two types as they contained extra information such as extracellular loops that are implicated in ligand selectivity, and binding of ligands in the active site. In addition to this the rhodopsin models conformed to more of the data from the SCAM technique, and areas where non alpha helical architecture of the rhodopsin crystal structure were present were well accounted for by the presence of proline residues in the D_2 sequence being aligned. The general statistics of the rhodopsin-based models are of lower quality in comparison to the bacteriorhodopsin-based models. However this seems to be an artefact of the starting structures they were based upon and the fact that more residues are being accommodated in the model construction. In the future as higher quality X-ray structures of rhodopsin are produced it is likely that the statistics of models based upon these structures will approach those of models based upon bacteriorhodopsin.

With the multitude of sequences for orphan receptors being currently deduced, it is surprising that this technique is not being used to a greater extent, as comparative modelling using Modeller performs as well as, and if not better than, traditional template model generation techniques.

Further work has already been done using this technique on the D_2 and D_4 GPCRs with rhodopsin[42] as the basis of the models and varying the types in ligands in the active site, this is discussed in the following section.

3B.1 STRUCTURE-BASED DESIGN

Structure Based Design along with Ligand Based Design are the two principal methods of drug design and discovery used by computational chemist. In structure based design the proposed receptor binding site (biophore) is used for docking experiments or for building up compounds in a *de novo* approach.

The method of ligand docking that is explored in this chapter is not conventional docking whereby a ligand is fitted inside a receptor cavity. The docking procedure of Modeller involves building a homology model of your receptor with your ligand inside. This method is more similar to an induced fit method whereby induced flexibility of protein side chains is modelled by generating a protein model with a ligand bound whilst trying to minimise the objective function. Unlike some induced fit docking methods however Modeller does not take into account conformational changes of the ligand in order to minimise the energy of the system created.

3B.1.1 Method Overview: Structure-Based Design

The docking procedure of Modeller has two major problems associated with it, as it does not start with a predefined structure into which your compounds are docked. The "brute force" way in which these receptor-ligand complexes are built, in that the homology model is created with the ligand bound in it, can cause problems if the initial ligand placement is incorrect. The problems associated with doing such a "brute force" docking procedure are that the ligand is biased by its original placement inside the template receptor. This can be overcome by initially placing the ligand into areas where it has been proposed to bind as shown by mutagenesis experiments. The next step is to place constraints on the model generation, such as indicating which atoms should be within hydrogen bonding distance of one another. However one must be careful not to place to many constraints in the model generation procedure otherwise this will most definitely bias the procedure. The conformations of all ligands used in this section of work are all directly from the conformations used in the pharmacophore modelling and CoMFA models of Chapter 2.

The second major problem associated with this technique is that these receptor ligand complexes are in actual fact homology models in the first place, which immediately lends them to problems of variability. The problem of model variability can be overcome by building numerous models and inspecting all the models together. Once all the models are created, visual analysis of all the models together usually reveals one or two dominant families of ligand placement within the receptor. To assess their validity ligand placements were compared to mutagenesis data indicating which residues should be protected by the bound ligand. Finally the ligand-receptor models were then compared with the CoMFA models generated in chapter 2. From this we were able to postulate which residues were responsible for various interactions in our CoMFA models, and identify new areas that could be explored within the cavity of our binding site.

3B.2 METHOD: STRUCTURE-BASED DESIGN

3B.2.1 Placement of ligands within the receptor

Various papers looking at the mutagenesis of residues in the D_2 and D_4 receptor were examined in detail to gain some insight into the placement of clozapine and spiperone within the receptors. These papers included one of the original D_2 mutagenesis papers by Mansour *et al.*, which examined the tricyclics clozapine and chlorpromazine as well as the high affinity D_2 compounds spiperone and raclopride. The paper by Simpson *et al.* which examined residues implicated in D_2 and D_4 selectivity was also examined in detail as contained a wealth of mutagenesis data for both *N*-methylspiperone and clozapine. The excellent papers by Javitch and co-workers were also examined in detail as they contained further information which gave indication as to which residues were protected by or in proximity to the antagonist binding site. Analysis of this data allowed a number of assumptions to be made as to the likely binding sites of these compounds.

3B.2.2 Model variability

The problem of model variability was overcome by generating ten models for each orientation of the ligand in each receptor in question. Models were constructed with Modeller using bovine rhodopsin (1F88) as the basis and the alignment shown in Table 3.12, the only constraint added during the generation of the models was that the protonated nitrogen of the ligand be within hydrogen bonding distance of the essential aspartate residue $Asp^{3.32}$. All models were then minimised according to the procedure set out in "3A.2.4 Minimising the Potential Energy of the Protein" with an additional constraint being that the ligand was fixed in the conformation used in the CoMFA models of Chapter 2. Each group of ten models was then superimposed onto one another using the 'Align Homologous Structures' algorithm within Sybyl. Using the aligned receptor models as a basis, ligands within each model were then compared to one another and ligands with a RMS difference of less than 3Å^2 were classified into a single family. The frequency of family types and number of contributing conformations

were then examined in conjunction with the mutagenesis data from all papers to asses the validity of each model.

3B.2.3 Comparisons to CoMFA models

The model that conformed best to all the data available was chosen as the basis for superimposition of the CoMFA models. The D_2 and D_4 tricyclic CoMFA models were then superimposed, using clozapine as the basis for superimposition, onto the best resultant D_2 and D_4 receptor models containing clozapine, respectively. And the extended D_2 and D_4 CoMFA models were superimposed, using spiperone as the basis for superimposition, onto the best resultant D_2 and D_4 receptor models containing spiperone, respectively. The extended CoMFA models were superimposed onto receptor models containing spiperone because the largest amount of mutational data for the D_2 and D_4 receptors was from mutations analysed using spiperone and N-methylspiperone.

Table 3.12. Numbering scheme of rhodopsin and the D_2 and D_4 dopaminergic receptors.

TM1 — First **1.30**, Last **1.64**

	First	1.30	1.31	1.32	1.33	1.34	1.35	1.36	1.37	1.38	1.39	1.40	1.41	1.42	1.43	1.44	1.45	1.46	1.47	1.48	1.49	1.50	1.51	1.52	1.53	1.54	1.55	1.56	1.57	1.58	1.59	1.60	1.61	1.62	1.63	1.64	Last
Rho	35	W	Q	F	S	M	L	A	A	Y	M	F	L	L	I	M	L	G	F	P	I	*N*	F	L	T	L	Y	V	T	V	Q	H	K	K	L	R	69
D_2	32	P	H	Y	N	Y	Y	A	T	L	L	T	L	L	I	A	V	I	V	F	G	*N*	V	L	V	C	M	A	V	S	R	E	K	A	L	Q	66
D_4	32	G	Q	G	A	A	A	L	V	G	G	V	L	L	I	G	A	V	L	A	G	*N*	S	L	V	C	V	S	V	A	T	E	R	A	L	Q	66

TM2 — First **2.38**, Last **2.68**

	First	2.38	2.39	2.40	2.41	2.42	2.43	2.44	2.45	2.46	2.47	2.48	2.49	2.50	2.51	2.52	2.53	2.54	2.55	2.56	2.57	2.58	2.59	2.60	2.61	2.62	2.63	2.64	2.65	2.66	2.67	2.68	Last
Rho	71	P	L	N	Y	I	L	L	N	L	A	V	A	*D*	L	F	M	V	F	G	G	F	T	T	T	L	Y	T	S	L	H	G	101
D_2	68	T	T	N	Y	L	I	V	S	L	A	V	A	*D*	L	L	V	A	T	L	V	M	P	W	V	V	Y	L	E	V	V	G	98
D_4	68	P	T	N	S	F	I	V	S	L	A	A	A	*D*	L	L	L	A	L	L	V	L	P	L	F	V	Y	S	E	V	Q	G	98

TM3 — First **3.22**, Last **3.56**

	First	3.22	3.23	3.24	3.25	3.26	3.27	3.28	3.29	3.30	3.31	3.32	3.33	3.34	3.35	3.36	3.37	3.38	3.39	3.40	3.41	3.42	3.43	3.44	3.45	3.46	3.47	3.48	3.49	3.50	3.51	3.52	3.53	3.54	3.55	3.56	Last
Rho	107	P	T	G	C	N	L	E	G	F	F	A	T	L	G	G	E	I	A	L	W	S	L	V	V	L	A	I	E	*R*	Y	V	V	V	C	K	141
D_2	104	R	I	H	C	D	I	F	V	T	L	D	V	M	M	C	T	A	S	I	L	N	L	C	A	I	S	I	D	*R*	Y	T	A	V	A	M	138
D_4	105	P	R	L	C	D	A	L	M	A	M	D	V	M	L	C	T	A	S	I	F	N	L	C	A	I	S	V	D	*R*	F	V	A	V	A	V	139

TM4 — First **4.38**, Last **4.63**

	First	4.38	4.39	4.40	4.41	4.42	4.43	4.44	4.45	4.46	4.47	4.48	4.49	4.50	4.51	4.52	4.53	4.54	4.55	4.56	4.57	4.58	4.59	4.60	4.61	4.62	4.63	Last
Rho	149	G	E	N	H	A	I	M	G	V	A	F	T	*W*	V	M	A	L	A	C	A	A	P	P	L	V	G	174
D_2	148	S	K	R	R	V	T	V	M	I	S	I	V	*W*	V	L	S	F	T	I	S	C	P	L	L	F	G	173
D_4	148	G	S	R	R	Q	L	L	L	I	G	A	T	*W*	L	L	S	A	A	V	A	A	P	V	L	C	G	173

E2

	First												Last
Rho	179	I	P	E	G	M	Q	C	S	C	G	I	189
D_2	174	L	N	N	A	D	Q	N	E	C	I	I	184
D_4	177	V	R	G	R	D	P	A	V	C	R	L	187

TM5 — First **5.35**, Last **5.62**

	First	5.35	5.36	5.37	5.38	5.39	5.40	5.41	5.42	5.43	5.44	5.45	5.46	5.47	5.48	5.49	5.50	5.51	5.52	5.53	5.54	5.55	5.56	5.57	5.58	5.59	5.60	5.61	5.62	Last
Rho	200	N	E	S	F	V	I	Y	M	F	V	V	H	F	i	I	*P*	L	I	V	I	F	F	C	Y	G	Q	L	V	227
D_2	185	N	P	A	F	V	V	Y	S	S	I	V	S	F	Y	V	*P*	F	I	V	T	L	L	V	Y	I	K	I	Y	213
D_4	189	D	R	D	Y	V	V	Y	S	S	V	C	S	F	F	L	*P*	C	P	L	M	L	L	L	Y	W	A	T	F	216

TM6 — First **6.28**, Last **6.60**

	First	6.28	6.29	6.30	6.31	6.32	6.33	6.34	6.35	6.36	6.37	6.38	6.39	6.40	6.41	6.42	6.43	6.44	6.45	6.46	6.47	6.48	6.49	6.50	6.51	6.52	6.53	6.54	6.55	6.56	6.57	6.58	6.59	6.60	Last
Rho	245	K	A	E	K	E	V	T	R	M	V	I	I	M	V	I	A	F	L	I	C	W	L	*P*	Y	A	G	V	A	F	Y	I	F	T	277
D_2	366	Q	K	E	K	K	A	T	Q	M	L	A	I	V	L	G	V	F	I	I	C	W	L	*P*	F	F	I	T	H	I	L	N	I	H	398
D_4	387	G	R	E	R	K	A	M	R	V	L	P	V	V	V	G	A	F	L	L	C	W	T	*P*	F	F	V	V	H	I	T	Q	A	L	419

TM7 — First **7.32**, Last **7.47**

	First	7.32	7.33	7.34	7.35	7.36	7.37	7.38	7.39	7.40	7.41	7.42	7.43	7.44	7.45	7.46	7.47	7.48	7.49	7.50	7.51	7.52	7.53	7.54	7.55	7.56	7.57	Last
Rho	285	P	I	F	M	T	I	P	A	F	F	A	K	T	S	A	V	Y	N	*P*	V	I	Y	I	M	M	N	310
D_2	405	P	V	L	Y	S	A	F	T	W	L	G	Y	V	N	S	A	V	N	*P*	I	I	Y	T	T	F	N	430
D_4	427	P	R	L	V	S	A	V	T	W	L	G	Y	V	N	S	A	L	N	*P*	V	I	Y	T	V	F	N	452

*residues highlighted in bold italisc are highly conserved across all GPCRs, they mark the X.50 point (where X is the transmembrane number).

*residues highlighted in bold are cysteines involved in the highly conserved disulfide linkage between TM3 and the extracellular loop E2

3B.3 RESULTS AND DISCUSSION: STRUCTURE-BASED DESIGN

Table 3.12. Numbering scheme of the rhodopsin and the D_2 and D_4 dopaminergic receptors.

<u>FOLD OUT</u>

3B.3.1 Binding Modes of Antagonists

3B.3.1.1 Mansour *et al.*

The paper by Mansour *et al.*[23] investigated the effects on ligand binding at the D_2 receptor for mutations on $Asp^{3.32}$ to Asn and Gly; $Met^{3.35}$ to Cys and Gly; $Ser^{5.43}$ to Ala; and $Ser^{5.46}$ to Ala. The ligands studied in this analysis relate to the antagonists examined in Chapter 2, which included chlorpromazine, raclopride, clozapine and spiperone. Chlorpromazine is structurally similar to the tricyclics, raclopride is structurally similar to the benzamides, while clozapine and spiperone were used in the pharmacophore elucidation (see Chapter 2).

The paper by Mansour *et al.*[23] shows that $Asp^{3.32}$ is critical for binding for both agonists and antagonists, and that $Met^{3.35}$ mutations did not effect the binding of raclopride (data not shown).

Table 3.13. K_i values (nM ± S.D.) of dopamine receptor antagonists competing with [^{3}H]raclopride bound to various D_2 receptor mutants.

	$Ser^{5.43}$ to Ala	$Ser^{5.46}$ to Ala	Wild type
raclopride*	0.235±0.020	0.714±0.044	0.166±0.015
clozapine	75.6±6.8	27.7±2.1	41.2±5.8
chlorpromazine	0.788±0.143	0.227±0.040	0.472±0.028
spiperone	0.079±0.008	0.298±0.01	0.096±0.046

*K_d values (nM ± S.D.) of [^{3}H]raclopride

From Table 3.13 it can be seen that mutations of $Ser^{5.46}$ significantly affects the binding affinities of both raclopride and spiperone, indicating that there is some interaction between this residue and the ligands in question. From these data and from computer modelling of the D_2 receptor, Mansour *et al.*[23] have proposed that $Ser^{5.46}$ interacts with the carbonyl oxygen of raclopride and the lactam proton of spiperone. However it should be noted that these suggestions are only speculative and that this modelling was done in 1992 when the basis of these receptor models was bacteriorhodopsin.
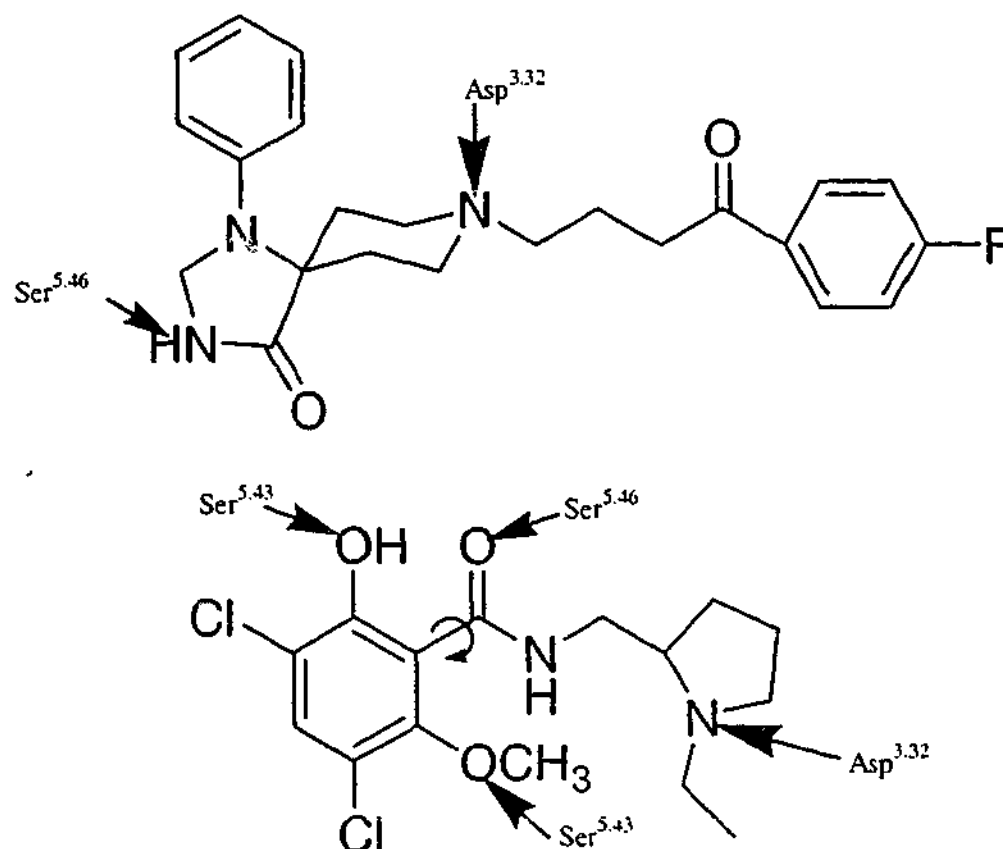
Figure 3.13. *Binding modes as proposed by Mansour et al.*

Table 3.13 also gives some indication as to interactions between $Ser^{5.43}$ and the tricyclic compounds, as when $Ser^{5.43}$ is mutated to Ala we see a two-fold decrease in binding affinity. In conjunction with the moderate increase in binding affinity when $Ser^{5.46}$ is mutated to Ala, this may indicate that the tricyclic substructure is around the same height within the helix as that of $Ser^{5.43}$. If the tricyclic substructure is around the same height as $Ser^{5.43}$ then there would be favourable interactions between this residue and the heteroatoms contained within the tricyclic substructure. Additionally mutation of $Ser^{5.46}$ to Ala would create a slightly more hydrophobic region below the plane of the tricyclic substructure and aid in its binding.

From the analysis of this paper and analysis of the D_2 models created earlier we are able to tentatively say that spiperone binds in the conformation shown in Figure 3.13. However we are also not able to discount the fact that the carbonyl oxygen of spiperone would be able to bind with $Ser^{5.46}$ and would in fact have spiperone bound in the reverse orientation to that proposed by Mansour *et al.* When looking at the binding mode of raclopride we see from previously created D_2 models that the distance between $Ser^{5.46}$ and $Asp^{3.32}$ varies between 8 - 11 Å which does not fit with the distance between the amide oxygen and tertiary nitrogen of raclopride. Instead this larger distance fits better

with the distance between the hydroxyl oxygen and tertiary nitrogen of raclopride which is approximately 8 Å. For the binding mode of clozapine we see the distance between the distal nitrogen and tricyclic substructure heteroatom of clozapine fits well with the distance between $Ser^{5.43}$ and $Asp^{7.32}$. Therefore we can tentatively say that clozapine binds with its tricyclic substructure in approximately the same plane as $Ser^{5.43}$ in the D_2 receptor.

3B.3.1.2 Simpson *et al.*

The paper by Simpson *et al.*[25] looks at D_4/D_2 receptor selectivity by mutating the solvent accessible residues of the D_2 and D_4 receptors that differ from one another and then assessing these changes by examining the binding affinities of different compounds. The compounds used to investigate D_2/D_4 selectivity included N-methylspiperone, CPPMA (L-745870), clozapine and three other Roche in house compounds. The investigation of clozapine in this paper is of significant interest and the N-methylspiperone compound has obvious similarities to the compounds under investigation here. The D_4 selective compound CPPMA is of limited interest because it was shown to display no clinical efficacy[56], possibly due to its partial agonist properties[57], and thus was not considered further.

Table 3.14. Comparison of affinities of antagonists for wild-type D_2 and D_4 receptors and mutant D_2 receptors. Mutant D_2 receptors have their corresponding D_4 amino acid present in the mutation.

Mutants	N-methylspiperone		Clozapine	
	K_D (pM) ± S.D.	$K_D D_2/K_D$mut	K_i (nM) ± S.D.	$K_i D_2/K_i$mut
D_4	360±70	0.2	12.4±0.6	23
D_2	79±8	1	280±30	1
$D_2 W^{2.60}L$	150±7	0.5	106±2	2.5
$D_2 V^{2.61}F$	28±17	2.8	5.7±0.9	49
$D_2 L^{2.64}S$	300±10	0.3	780±20	0.4
$D_2 F^{3.28}L$	230±10	0.3	560±110	0.5
$D_2 V^{3.29}M$	910±40	0.1	370±90	0.8
$D_2 F^{4.54}A/S^{4.57}A$	112±12	0.7	430±90	0.7
$D_2 L^{4.60}V/F^{4.62}C$	110±13	0.7	1650±1250	0.2
$D_2 F^{5.38}Y$	33±23	2.4	51±1	5.5
$D_2 V^{5.45}C$	110±10	0.7	370±210	0.8
$D_2 T^{6.54}V$	110±10	0.7	280±180	1
$D_2 Y^{7.35}V$	310±7	0.3	60±4	4.7
$D_2 F^{7.38}V$	102±11	0.8	200±80	1.4

The mutational data for clozapine (Table 3.14) is of much interest as we see much larger increases in mutant D_2 affinity for single amino acid mutations. The most significant of the clozapine mutational data is that of D_2 $V^{2.61}F$, where we see almost a fifty fold increase in binding affinity. This large change would most certainly indicate that clozapine is interacting with this residue in the D_4 receptor, and that the aromatic stacking that this mutation would provide is interacting with the aromatic tricyclic substructure of clozapine. The other two interesting mutations from Table 3.14 are $F^{5.38}Y$ and $Y^{7.35}V$, where we see in both cases close to a five-fold increase in binding affinity. The first of these two mutations, $F^{5.38}Y$, may be explained in that the electron withdrawing hydroxyl substituent will enhance the π-π interactions with adjacent aromatic rings. However when we look at this aromatic interaction, $F^{5.38}Y$, and that of $V^{2.61}F$ we see that both residues are on opposite sides of the receptor, indicating that there may be more than one possible binding mode for clozapine at the D_4 receptor. The $Y^{7.35}V$ mutation is interesting because there is an opposite effect to that for spiperone which may be due to the fact that clozapine is a larger, more globular ligand and benefits from the additional space that this mutation would provide. So although

this mutation, $Y^{7.35}V$, may not directly affect clozapine's binding it would most likely affect the packing of residues in its vicinity. There is one significant dual mutation of the D_2 receptor that reduces the affinity of clozapine, namely $L^{4.60}V/F^{4.62}C$. These two residues are at the top of TM4 and although the single mutation $L^{4.60}V$ is a conservative mutation the $F^{4.62}C$ mutation is rather significant. Loss of aromatic binding through the $F^{4.62}C$ mutation would indicate that some portion of the aromatic tricyclic substructure is indeed interacting with $F^{4.62}$ in the D_2 receptor.

When looking at the mutational binding data for spiperone in Table 3.14 we can see that two of the mutations $V^{2.61}F$ and $F^{5.38}Y$ actually increase spiperones binding affinity at these mutant D_2 receptors. This would indicate that the increase in aromatic character of the two substituents aids in binding aromatic portions of the ligand, spiperone. The $F^{5.38}Y$ mutation, although an aromatic mutation, can be rationalised by the electron withdrawing effect of the hydroxyl substituent which will enhance the stacking interactions with adjacent aromatic rings. These mutations give an indication as to where the aromatic portions of spiperone may reside in the D_4 receptor. Interestingly both these mutations lie on opposite sides of the receptor, indicating that more than one possible binding mode may exist for spiperone at the D_4 receptor or that the ligand is binding in an extended conformation. There are three significant mutations of the D_2 receptor that reduce the affinity of spiperone, $L^{2.64}S$, $V^{3.29}M$ and $Y^{7.35}V$. All these mutations result in a net loss of hydrophobicity to the corresponding mutant receptor and in the first two mutations a change in electronic character of the amino acid. This would indicate that some hydrophobic portion of spiperone is binding in this region of the D_2 receptor. These three residues, $L^{2.64}$, $V^{3.29}$ and $Y^{7.35}$ are actually in close proximity to one another on helices 2, 3 and 7, respectively. The unfavourable interaction these corresponding D_4 counterparts $S^{2.64}$, $M^{3.29}$ and $V^{7.35}$ have with N-methylspiperone also explains why the binding of N-methylspiperone is less at the D_4 receptor even though the $V^{2.61}F$ and $F^{5.38}Y$ mutations increase binding affinity.

Table 3.15. Comparison of affinities of N-methylspiperone for wild-type D_4 and D_2 receptors and mutant D_4 receptors. Mutant D_4 receptors have their corresponding D_2 amino acid present in the mutation.

Mutants	N-methylspiperone	
	K_D (pM)	K_DD_2/K_Dmut
D_4	360±70	0.22
D_2	79±8	1
D_4 $V^{7.35}Y$	650±80	0.12
D_4 $L^{3.28}F/M^{3.29}V$	178±20	0.45
D_4 $L^{2.60}W/F^{2.61}V/S^{2.64}L$	220±30	0.35
D_4 $L^{2.60}W/F^{2.61}V/S^{2.64}L/L^{3.28}F/M^{3.29}V$	790±164	0.1
D_4 $L^{2.60}W/F^{2.61}V/S^{2.64}L/L^{3.28}F/M^{3.29}V/V^{7.35}Y$	1070±160	0.1

The mutations at the D_4 receptor were only examined with two ligands N-methylspiperone and CPPMA, but as CPPMA is a partial agonist it was not considered. When looking at the mutational binding data in Table 3.15 we see that both the double mutation $L^{3.28}F/M^{3.29}V$ and the triple mutation $L^{2.60}W/F^{2.61}V/S^{2.64}L$ gave tighter binding affinities for N-methylspiperone compared to the wild-type D_4 receptor. This was similar to that observed earlier for N-methylspiperone when we saw the D_2 mutations $V^{3.29}M$ and $F^{3.28}L$ decrease N-methylspiperone's binding affinity due to the change in electronic character and decrease in hydrophobicity, respectively, of the binding site. On this occasion with the removal of this electronic character and increase in hydrophobicity we see the binding affinity of spiperone increase, indicating that spiperone is interacting with these residue in both the D_2 and D_4 receptors. The triple mutation, $L^{2.60}W/F^{2.61}V/S^{2.64}L$, is rather complicated as earlier we saw the $V^{2.61}F$ mutation increase the binding affinity of N-methylspiperone on the D_2 mutant whereas the $W^{2.60}L$ and $L^{2.64}S$ mutations both decreased binding affinity. Although we see a moderate increase in N-methylspiperone's binding affinity it would be difficult to say which mutation this is due to, as the increase in hydrophobicity from the $S^{2.64}L$ and $L^{2.60}W$ mutations should both increase N-methylspiperone's binding affinity whereas the $F^{2.61}V$ mutation should decrease N-methylspiperone's binding affinity. The further mutations involving five and six amino acids also are a little complicated to interpret with any great degree of confidence. These multiple mutations could cause significant changes in the packing arrangement of the helices and even the backbone bends of the helices and thus further interpretation would be contentious.

Overall the paper by Simpson *et al.* does explain some selectivity exhibited by the D_2 and D_4 receptors for extended compounds, due to the aromatic microdomain between helices two, three and seven. However the binding mode and selectivity for D_4 exhibited by clozapine was not examined in any great detail in this paper, although it would appear that there are multiple or at least different binding modes for clozapine at the D_4 and D_2 receptors. It appears that the high affinity binding site for clozapine at the D_4 receptor is in the region between helices two, three and seven, whereas for the D_2 receptor appears to be between helices three, four, five and six.

From these two papers and from the analysis of a number of other modelling papers[19-22,26-29] it is possible to propose the binding modes of clozapine, shown in Figure 3.14.
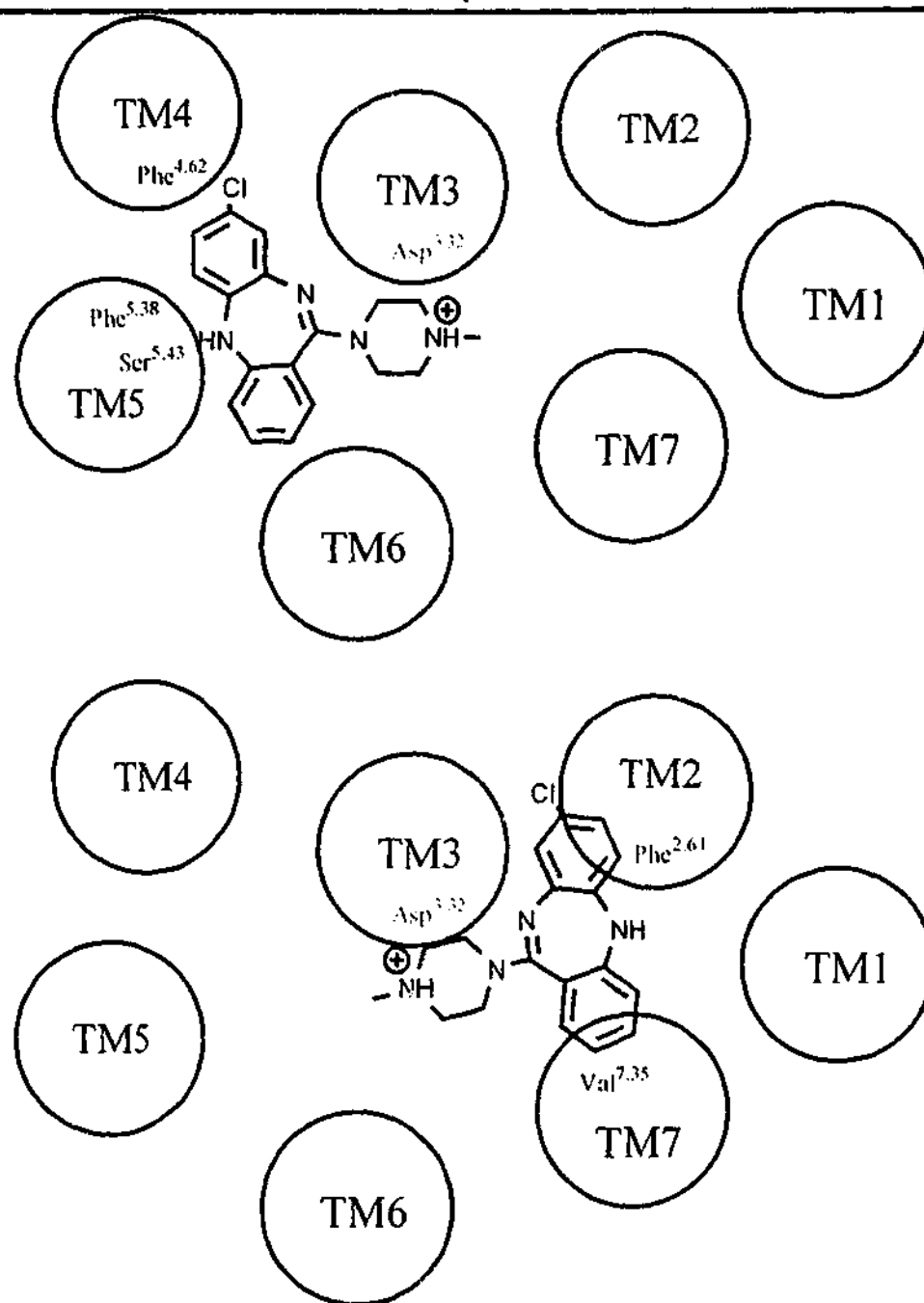
Figure 3.14. *Possible binding modes of clozapine at amino acids in the D_2 (top) and D_4 (bottom) receptors. The essential amino acid $Asp^{3.32}$ is shown in red, the other residues implicated in binding clozapine are shown in blue.*

The possible binding modes for spiperone are slightly more difficult to deduce due to the flexible nature of spiperone and the number of hydrogen bonds it is capable of forming with the receptor. The proposal put forward by Mansour, that the lactam proton of spiperone interacts with $Ser^{5.46}$, would indicate that methylation of nitrogen would give a drop in affinity from loss of a hydrogen bond. However from the data of Simpson *et al.* this is not the case where the binding affinity of *N*-methylspiperone, 79 ± 8 pM[25], is tighter than that of spiperone, 96 ± 46 pM[23]. Actually the binding affinity of spiperone reported by Mansour *et al.* may be artificially low compared to that of *N*-methylspiperone because there is a noticeable disparity in the binding affinity of

clozapine from the two groups. The clozapine affinity reported by Mansour *et al.* (41.2±5.8 nM) is four times tighter than that of Simpson *et al.* (280±30 nM). On this basis, then, it is proposed that it is not the lactam proton of spiperone that is interacting with $Ser^{5.46}$. However, this does not exclude a water-mediated hydrogen bond with the carbonyl oxygen of the lactam or even just electronic stabilisation of the amide linkage by $Ser^{5.46}$. In addition it still cannot be ruled out the that the carbonyl oxygen of spiperone would be able to bind with $Ser^{5.46}$ and would in fact have spiperone bound in the reverse orientation to that proposed by Mansour *et al.* Thus it is difficult to say which end of spiperone binds closest to $Ser^{5.46}$, as either orientation would satisfy the data available. The possible binding modes for spiperone at the D_2 and D_4 receptors are shown in Figure 3.15.
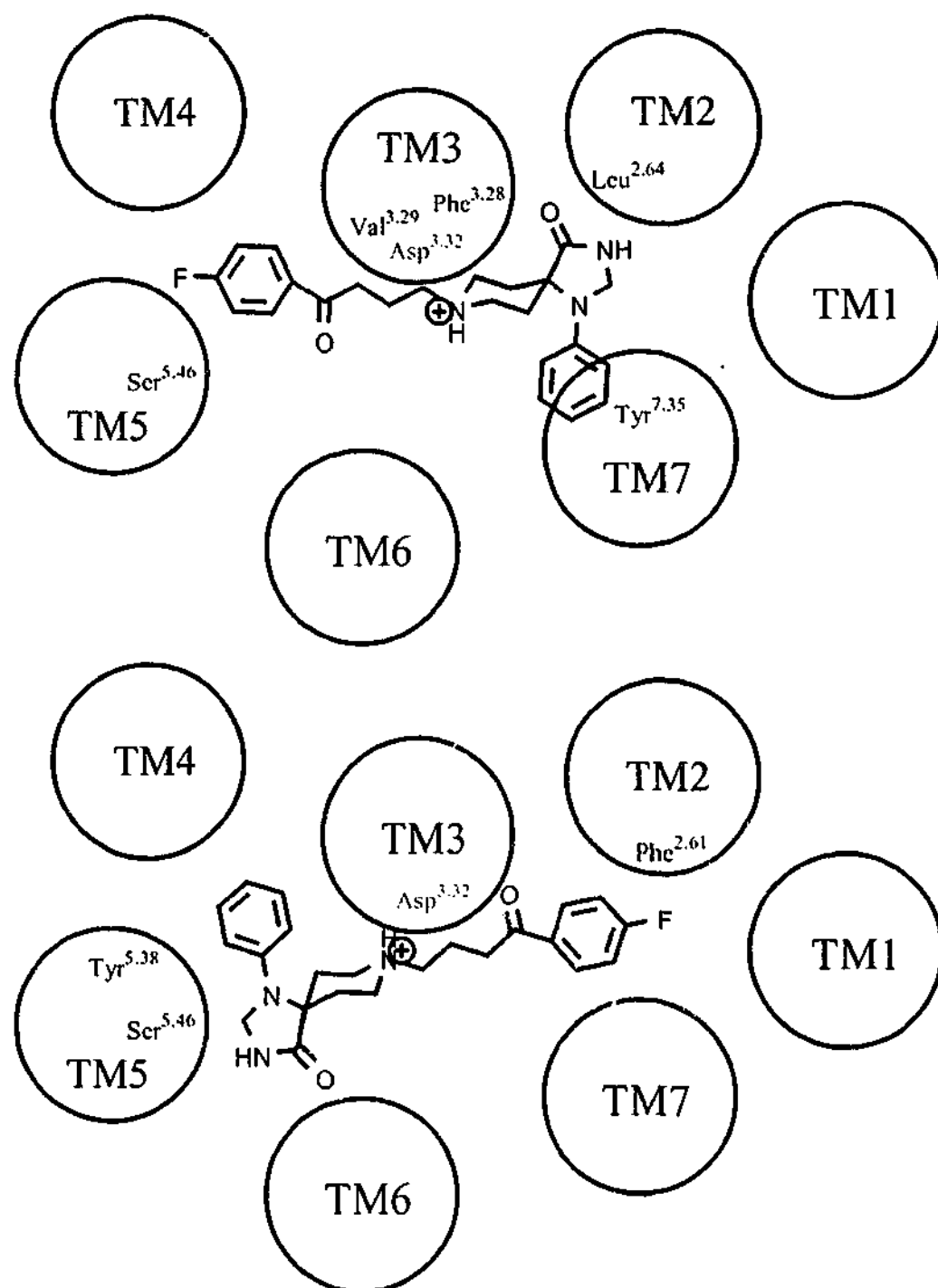
Figure 3.15. Possible binding modes of spiperone at amino acids of the D_2 (top) and D_4 (bottom) receptors, the essential amino acid Asp$^{3.32}$ is shown in red, the other residues implicated in binding spiperone are shown in blue.

3B.3.1.3 Sulpiride Protection of Residues

In addition to the two previous papers looking at various mutations within the binding site of the D_2 and D_4 receptors there have been a number of excellent papers[19-22,26-29] looking at identifying solvent accessible residues within the D_2 receptor. Apart from only identifying residues that were solvent accessible, a number of other experiments were carried out to shed more light on the binding site of the D_2 receptor. In this experiment all residues along each helix were in turn mutated to cysteine and then the antagonist sulpiride was added. Methanthiosulfonate (MTS) was then added and any cysteine residues still able to react with MTS (Figure 3.10), and thus not protected by the binding of the antagonist, were identified (Table 3.16). *Protection of a substituted cysteine is most simply explained by its proximity to the sulpiride and [^3H]-N-methylspiperone binding site*[27]. Not all residues identified are in proximity to the sulpiride-binding site as residues further inside the receptor are protected from the extracellular route that MTS is taking.

Table 3.16. Sulpiride protection of residues in the D_2 receptor.

TM1	TM2	TM3	TM4	TM5	TM6	TM7
$Asn^{1.50}$	$Asp^{2.50}$	$Asp^{3.26}$	$Trp^{4.50}$	$Phe^{5.38}$	$Val^{6.40}$	$Leu^{7.34}$
	$Leu^{2.51}$	$Ile^{3.27}$	$Ser^{4.53}$	$Val^{5.39}$	$Phe^{6.44}$	$Phe^{7.38}$
	$Val^{2.53}$	$Phe^{3.28}$	$Phe^{4.54}$	$Val^{5.40}$	$Pro^{6.50}$	$Thr^{7.39}$
	$Val^{2.57}$	$Val^{3.29}$	$Ile^{4.56}$	$Tyr^{5.41}$	$Phe^{6.52}$	$Trp^{7.40}$
	$Pro^{2.59}$	$Val^{3.33}$	$Ser^{4.57}$	$Ser^{5.42}$	$Thr^{6.54}$	$Tyr^{7.43}$
	$Trp^{2.60}$	$Cys^{3.36}$	$Leu^{4.60}$	$Ser^{5.43}$	$Ile^{6.56}$	$Asn^{7.45}$
	$Val^{2.61}$	$Ser^{3.39}$	$Leu^{4.61}$	$Ile^{5.44}$		$Asn^{7.49}$
	$Leu^{2.64}$	$Ile^{3.40}$	$Gly^{4.63}$	$Val^{5.45}$		$Pro^{7.50}$
	$Glu^{2.65}$	$Leu^{3.43}$	$Asn^{4.65}$	$Ser^{5.46}$		$Tyr^{7.53}$
		$Ser^{3.47}$		$Phe^{5.47}$		
				$Val^{5.49}$		
				$Pro^{5.50}$		
				$Tyr^{5.58}$		
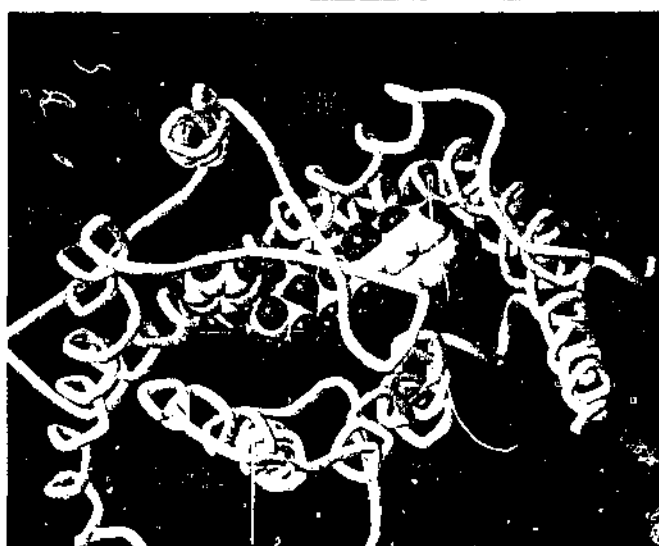
Figure 3.16. *Sulpiride protected residues from Table 3.16 shown in orange on a D₂ receptor model with a proposed bound orientation of spiperone shown in spacefill rendering.*

From Figure 3.16 it can be seen that spiperone binds in an extended fashion from between helices two, three and seven across to between helices four, five and six. This data is consistent with the previous mutational data of Mansour *et al.* and Simpson *et al.* The larger number of residues being protected at the top of the receptor would indicate that sulpiride and *N*-methylspiperone bind relatively close to the extracellular surface of the receptor and almost perpendicular to the helices.

From this information we are able to confidently place the ligands clozapine and spiperone into the binding site of the D_2 and D_4 receptors. In turn, this placement of these ligands can be used as a template for placement of the pharmacophore and CoMFA models derived previously.

3B.3.2 Model Generation

Over one hundred models were created of ligands bound in the D_2 and D_4 receptors, with the only constraint used in all model generation being that the protonated distal nitrogen was placed within hydrogen bonding distance of the essential Asp$^{3.32}$. The two starting orientations of clozapine (Figure 3.14) and spiperone (Figure 3.15) were used for the initial placement of all ligands in the template. Once all the models were created for each receptor and its differing starting placement of ligand, the models were overlayed to see if a particular arrangement of ligand and contacting residues inside the

receptor was favoured. The orientations were also checked against the initial data that determined the placement of the ligand.

3B.3.2.1 Receptor Models with Clozapine

Clozapine bound in region TM4, TM5 and TM6 of the D_2 dopaminergic model

From the ten models created with clozapine initially placed in the region between helices four, five and six, there was one main family of ligand placements derived from six of the ten models. The other four models placed clozapine into differing orientations with the exception of two models that had clozapine's tricyclic substructure parallel with the helices of the models produced. These two families of models were assessed to see if the placement of clozapine conformed to the mutagenesis data available. The smaller family of models with clozapine parallel to the helices did not conform to the mutagenesis data and thus was excluded. The larger family of models with clozapine almost perpendicular to the axes of the helices (Figure 3.17) conformed well to the mutagenesis data available.
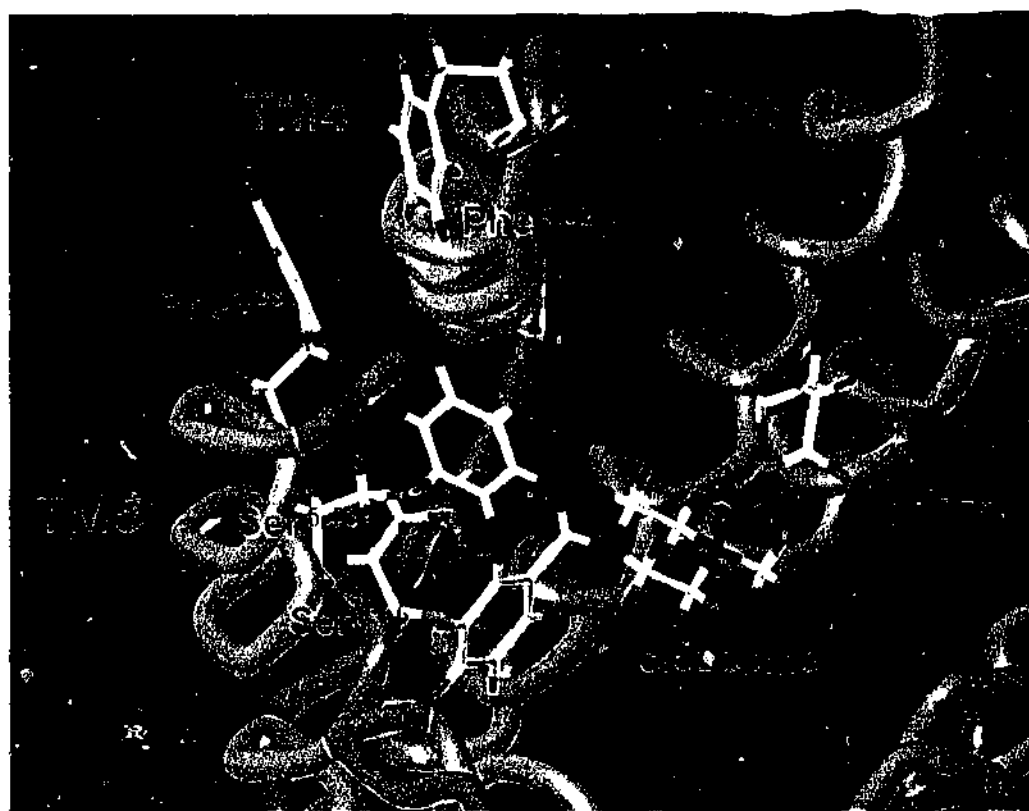
Figure 3.17. Optimum placement of clozapine in the region between helices four, five and six, as derived from six of the ten models produced. Residues implicated in the binding of clozapine from mutagenesis experiments are shown.

When comparing the proposed binding site with mutagenesis data it can be seen from Figure 3.17 that the tricyclic substructure of clozapine is positioned at the same height as $Ser^{5.43}$ in the receptor. It can also be seen how the mutation of $Ser^{5.46}$ to Ala would increase hydrophobicity in the binding site and thus aid in the binding of clozapine. The mutation of $Phe^{4.62}$ to Cys and the consequent loss of affinity at the D_2 receptor can be rationalised by the $Phe^{4.62}$ residue stabilising the $Phe^{5.38}$ on TM5 that interacts with the tricyclic substructure. The removal of $Asp^{3.32}$ would also result in a total loss of clozapine's binding affinity as there would no longer be a hydrogen bond to the protonated distal nitrogen. The complete set of interactions for clozapine in this orientation in the binding site can be seen in the ligplot-generated diagram (Figure 3.18) which shows a two-dimensional representation of all residues within 4 angstrom of the ligand. The Ligplot generated diagram shows the hydrophobic contacts that the tricyclic substructure is making with $Phe^{6.52}$, $Phe^{5.38}$ and $Ile184^{loop\ E2}$. The interaction of $Ser^{5.43}$ with tricyclic substructure atoms, C10 and C11, of clozapine can also be seen, indicating that the tricyclic substructure is almost in the same plane as this residue. The piperazine ring can be seen to be interacting with a number of hydrophobic residues, $Phe^{6.51}$, $Trp^{6.48}$ and $Val^{3.33}$. The additional hydrogen bond between $His^{6.55}$ and the N4 nitrogen of the tricyclic substructure can also be seen. It has been proposed that $His^{6.55}$ forms a hydrogen bond with the 5-sulfonamide of sulpiride, as mutation of $His^{6.55}$ to Cys results in a decrease in the binding affinity of sulpiride. Unfortunately this same experiment has not been carried out with clozapine, and it would be a good test of this proposed binding site to see if a decrease in clozapine's affinity was seen with the mutation of $His^{6.55}$ to Cys.

Figure 3.18. *Ligplot analysis of the binding site interactions of clozapine bound in the region between helices four, five and six of the D_2 receptor model. The red fan-like representations show the hydrophobic contacts between residues and corresponding ligand atoms.*

Clozapine bound in region TM2, TM3 and TM7 of the D_2 dopaminergic model

There were two families derived from the ten models produced of clozapine bound in the region between helices two, three and seven. One family contained four models with a similar placement of clozapine within the TM2, TM3 and TM7 region. This family (Figure 3.19) had an orientation of clozapine that interacted quiet significantly with the backbone of TM7 and caused this helix to rotate a further third of a turn anticlockwise, thus leaving the $Tyr^{7.35}$ facing out into the lipid bilayer. $Tyr^{7.35}$ had been proposed to interact with the packing arrangement of the binding site as shown by the increase in binding affinity of clozapine when this residue was mutated to valine. Since this proposed binding mode is not explained by mutational data it was discarded.

Figure 3.19. *Placement of clozapine in the region between helices two, three and seven, as derived from four of the ten D_2 models produced.*

The second family of clozapine's placement within the binding site was seen in three of the ten models created. Although less models were produced with clozapine in this orientation, the binding profile (Figure 3.20) fits better with the mutagenesis data. From looking at Figure 3.20 it can be seen how the mutation of $Val^{2.61}$ to Phe would increase the binding of clozapine through a favourable π-π interacton arrangement with the tricyclic substructure. In addition to this, the mutation of $Tyr^{7.35}$ to Val would mean that the aromatic π-π interaction arrangement of $Tyr^{7.35}$ with $Phe^{6.51}$ would be lost giving the binding site more flexibility and space, which would aid in binding of clozapine. As stated earlier, $Tyr^{7.35}$ had been proposed to interact with the packing arrangement of the binding site.

Figure 3.20. *Placement of clozapine in the region between helices two, three and seven, as derived from three of the ten models produced. Residues implicated in the binding of clozapine from mutagenesis experiments are shown.*

The complete set of interactions of clozapine in this binding orientation are shown in the ligplot-generated diagram (Figure 3.21). The interactions of $Trp^{6.48}$ and $Phe^{6.51}$, which appear to be partly controlled by $Tyr^{7.35}$, can be seen with the piperazine ring of clozapine. $Tyr^{7.35}$ which is located above $Phe^{6.51}$ and $Trp^{6.48}$ and hence not seen in Figure 3.21 are in a triple π-π interaction arrangement, which is creating a relatively rigid barrier and limiting the size of the binding site. Thus any disturbance to this triple π-π interaction arrangement should aid in the binding by giving more flexibility where it is needed. Apart from the hydrophobic interaction of $Val^{2.61}$ with the tricyclic substructure we can also see hydrophobic interactions from $Leu^{1.29}$, $Val^{2.57}$, $Trp^{7.40}$, $Tyr^{7.43}$ and $Phe^{3.28}$.

Figure 3.21. *Ligplot analysis of the binding site interactions of clozapine bound in the region between helices two, three and seven of the* D_2 *receptor model.*

When looking at the two proposed binding mechanisms of clozapine in the D_2 receptor we see two mechanisms that are both in part supported by mutagenesis data. This would indicate that there are definitely two clozapine-binding sites in the D_2 receptor. However it appears as though they are not of equal affinity. The rather crowded binding site between helices two, three and seven appears to be the lower affinity site due to steric considerations and the lack of an ideally placed residue through which a π-π interaction arrangement can be made with the tricyclic substructure.

Clozapine bound in region TM4, TM5 and TM6 of the D_4 dopaminergic model

There was one main family derived from seven of the ten models produced of clozapine bound in the region between helices four, five and six of the D_4 receptor. Another small family made up of two models with clozapine bound parallel to the helices was also present but this model did not conform with the mutagenesis data and was excluded. The main family orientation (Figure 3.22) shows clozapine in a similar orientation as

for the D_2 model (Figure 3.17). Clozapine is again almost perpendicular to the helices in approximately the same plane as $Ser^{5.43}$. The mutational data that indicates a favourable interaction between clozapine and $Tyr^{5.38}$ can be seen in Figure 3.22. However, for the D_4 model, we do not have the stabilising interaction on $Tyr^{5.38}$ from $Phe^{4.62}$ as is the case for the D_2 receptor model. In the D_4 receptor model the $Phe^{4.62}$ is a Cys which would not stabilise $Tyr^{5.38}$. With no stabilising interaction the binding of clozapine in the model may not be as tight in this region as it is in the D_2 model. This may indicate that this region between helices four, five and six is the lower-affinity binding site for clozapine in the D_4 receptor model.
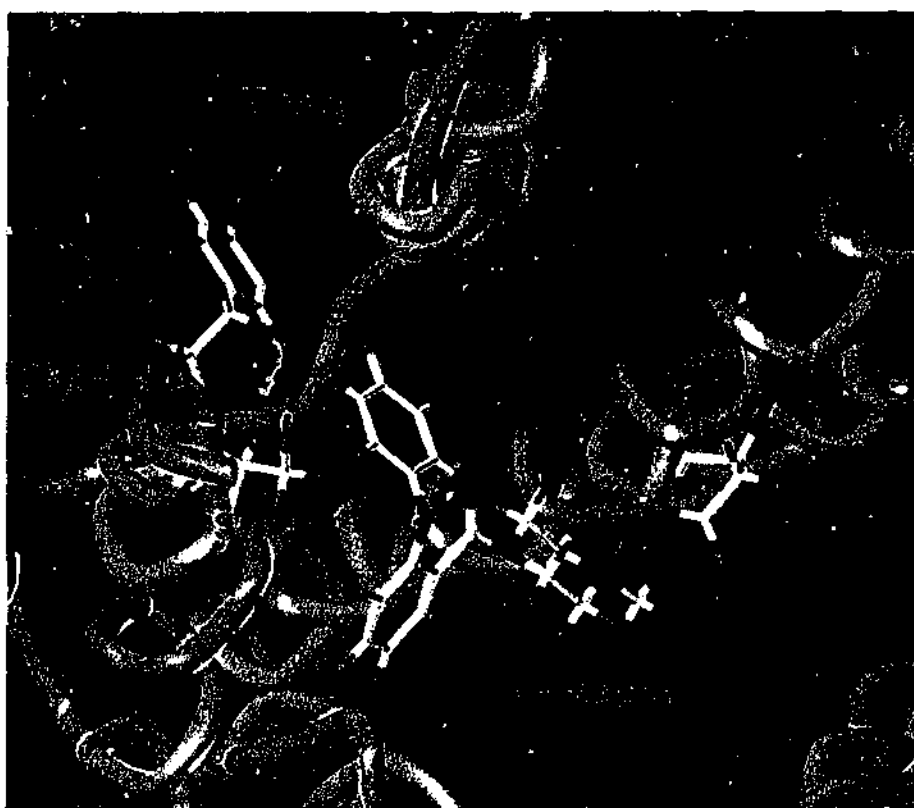
Figure 3.22. Clozapine bound in the region between helices four, five and six, as derived from seven of the ten D_4 models produced. Residues implicated in the binding of clozapine from mutagenesis experiments are shown.

The ligplot-generated diagram (Figure 3.23) shows the complete set of interactions for clozapine in the D_4 receptor model. The interactions are almost identical between the D_4 and D_2 receptor models, see Figure 3.18 compared to Figure 3.23, although the corresponding D_4 residues are implicated. The only two differences are that the interactions with residues in the extracellular loop E2 differ by one place, see Table 3.12. The D_2 model shows clozapine interacting with $Ile184^{loop\ E2}$ which corresponds to $Arg186^{loop\ E2}$ on the D_4 model, whereas the D_4 model is interacting with $Leu187^{loop\ E2}$.

The only other difference between the two models is that the D_2 model is interacting with one more residue, $Phe^{6.52}$. The ligplot-generated diagram shows the hydrophobic contacts that the tricyclic substructure is making with $Tyr^{5.38}$ and $Leu187^{loop\ E2}$. The interaction of $Ser^{5.43}$ with tricyclic substructure atoms, C9 and C10, of clozapine can be seen, indicating that the tricyclic substructure is almost in the same plane as this residue. The piperazine ring can be seen to be interacting with a number of hydrophobic residues, $Phe^{6.51}$, $Trp^{6.48}$ and $Val^{3.33}$. The additional hydrogen bond between $His^{6.55}$ and the proximal nitrogen of the tricyclic substructure can also be seen.
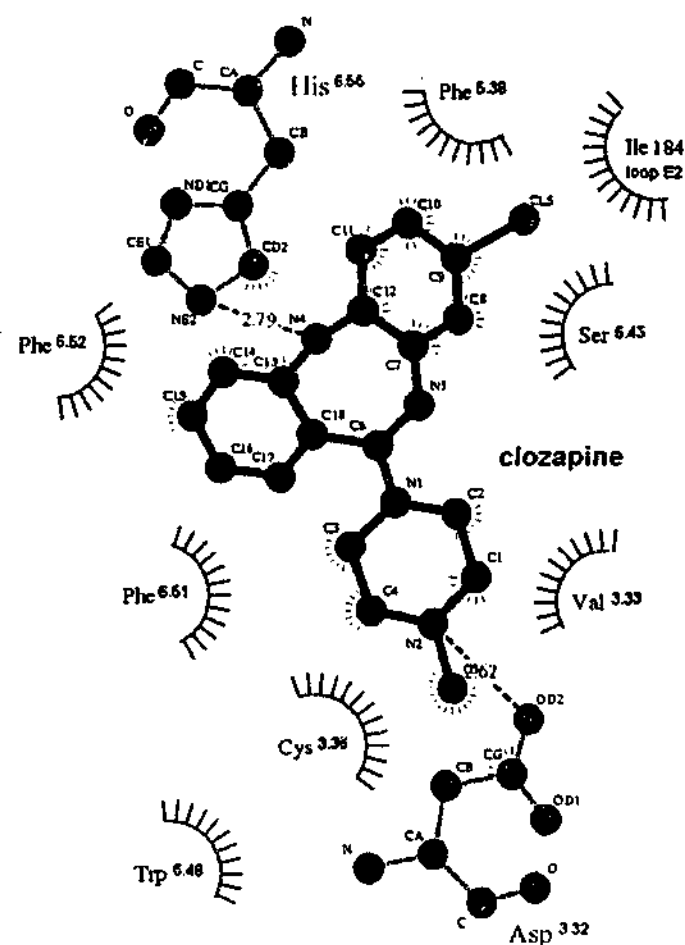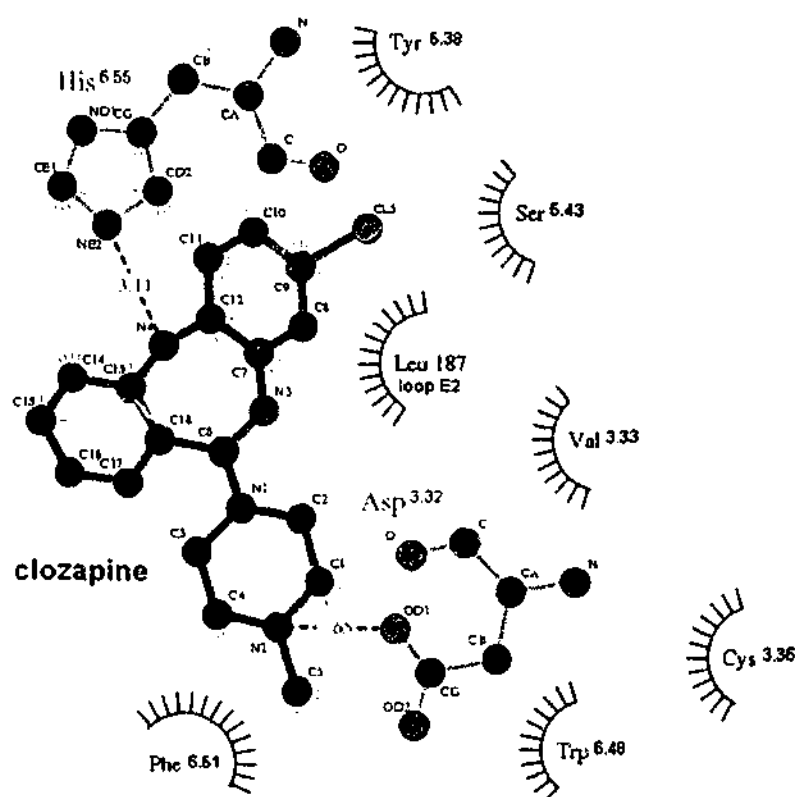

Figure 3.23. Ligplot analysis of the binding site interactions of clozapine bound in the region between helices four, five and six of the D_4 receptor model.

Clozapine bound in region TM2, TM3 and TM7 of the D_4 dopaminergic model

In nine of the ten models created a very similar arrangement of clozapine in the region between helices two, three and seven of the D_4 receptor models was seen (Figure 3.24). This high percentage of similar clozapine orientations within the models is in stark contrast to what was seen in the D_2 receptor models created with clozapine bound

between helices two, three and seven. This would indicate that the orientation of clozapine in this region of the D_4 model is a much more stable orientation compared to the D_2 models. This added stability of the D_4 models is most likely due to the favourable π-π interactions between $Phe^{2.61}$ and the aromatic tricyclic substructure of clozapine. This favourable interaction is supported by the mutational data that shows the binding affinity of clozapine increase when $Val^{2.61}$ of the D_2 model is mutated to phenylalanine. In addition to this favourable interaction the extra space that is available in this region of the D_4 model due to the $Val^{7.38}$ allows clozapine to fit neatly within the binding site without some of the steric clashes that are seen in the D_2 model.
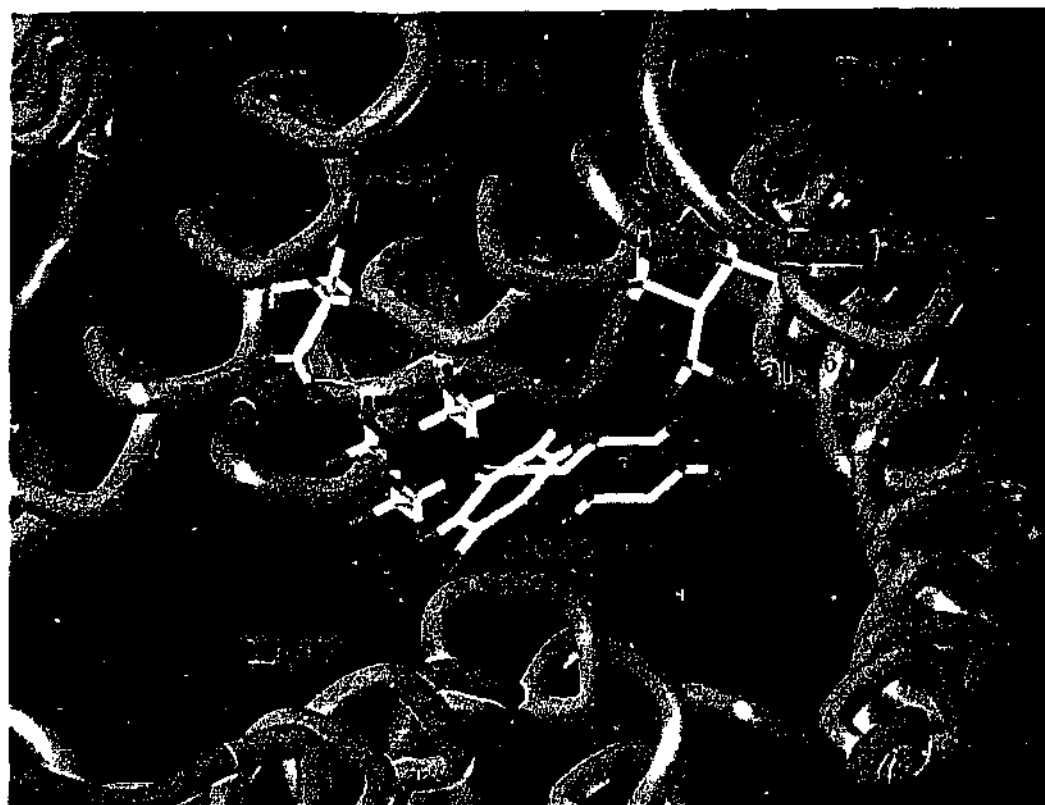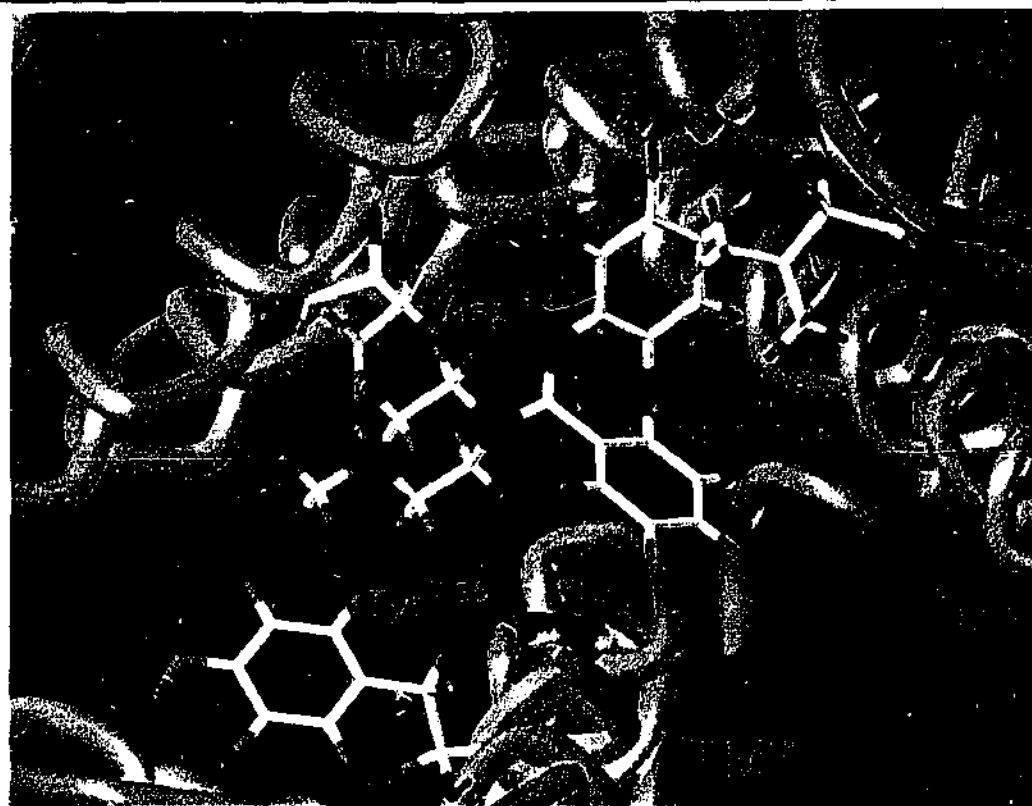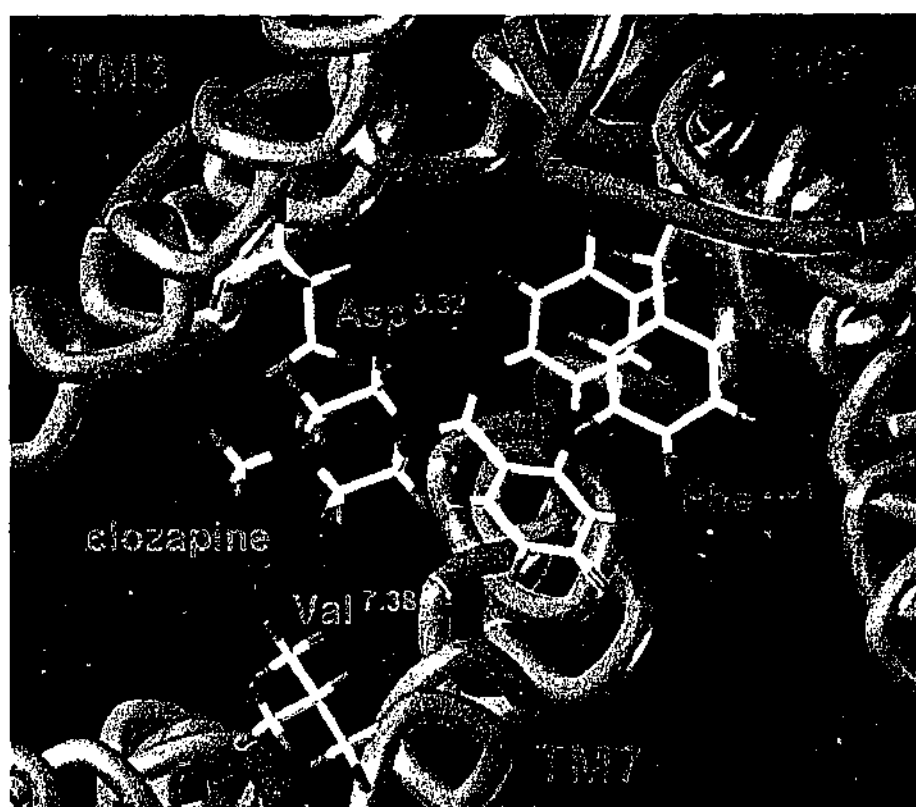

Figure 3.24. *Clozapine in the region between helices two, three and seven, as derived from nine of the ten D_4 models produced. Residues implicated in the binding of clozapine from mutagenesis experiments are shown*

The full set of interactions for clozapine bound in this orientation as shown in the ligplot-generated diagram (Figure 3.25) show a remarkable similarity to those of the D_2 model (Figure 3.21). The hydrophobic interactions with the tricyclic substructure from residues $Trp^{7.40}$ and $Val^{2.57}$ are seen in both diagrams. However there are also a number of different interactions with the tricyclic substructure. In the D_4 model we see that the $Leu^{2.58}$ corresponds to a $Met^{2.58}$ in the D_2 model and this might provide an area to explore in order to gain selectivity between the two receptors. In addition to this we see

that the tricyclic substructure in the D_2 model is virtually hemmed in the binding site by Leu[1.39], which corresponds to a Gly[1.39] in the D_4 model. This is more evident when comparing Figure 3.21 with Figure 3.25, where there are no interactions with amino acids from TM1. The much smaller size of Gly compared to Leu would provide additional space to explore in the D_4 model that should aid in developing a D_4 selective compound.
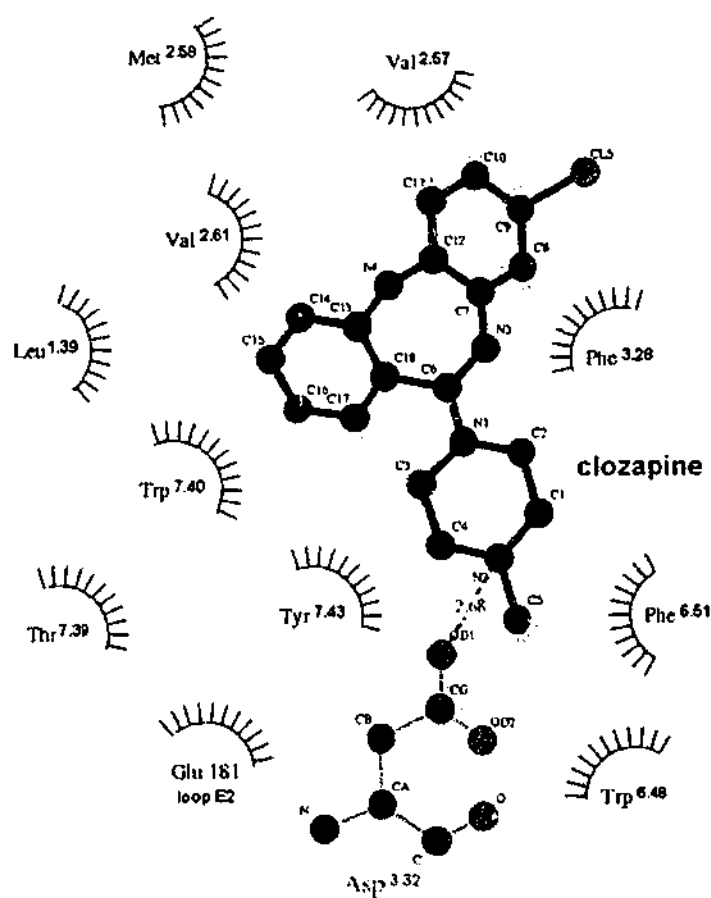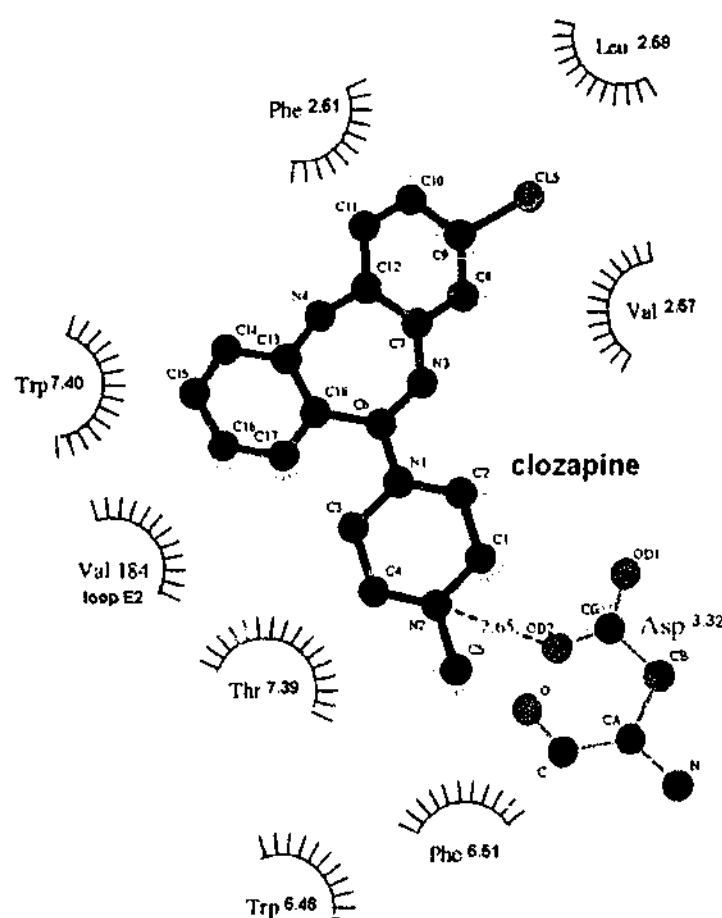

Figure 3.25. *Ligplot analysis of the binding site interactions of clozapine bound in the region between helices two, three and seven of the D_4 receptor model.*

Other differences between the two binding sites include the interacting residues from the extracellular loop E2. In the D_2 model (Figure 3.21) we see a negatively charged Glu181[loop E2] interacting with the piperazine ring of clozapine which is not present in the D_4 model. From the previous CoMFA result that showed piperidine compared to piperazine rings were preferential for D_2 binding indicating, that the less negative piperidine ring may be better tolerated in the D_2 receptor, and thus aid in selectivity.

3B.3.2.2 Receptor Models with Spiperone

Spiperone bound in the D₂ dopaminergic model

In the initial placement of spiperone in the D_2 dopaminergic binding site, spiperone was orientated so that the *para*-fluoro end of spiperone was in the region between helices four, five and six. Spiperone may be placed with either the *spiro* ring system closer to the extracellular or intracellular sides and both of these starting orientations were investigated. In seven o f the ten models g enerated with spiperone bound, where the starting orientation had the *spiro* ring system closer to the intracellular side, a similar final orientation was seen (Figure 3.26).

Figure 3.26. *Spiperone bound in the D₂ dopaminergic receptor, with the para-fluoro end i n t he re gion b etween h elices f our, f ive a nd s ix. T he re sidues i mplicated i n t he binding region are shown labelled.*

In nine of the ten models produced with the initial placement of the *spiro*-ring system facing out to the extracellular side, spiperone was completely reorientated so that the *spiro*-ring system was again closer to the intracellular side of the receptor. This indicated that spiperone did not preferentially bind with the *spiro*-ring system orientated towards the extracellular side. In fact the nine inverted orientations found here compared very favourably with the seven orientations found concurrently in which the

starting orientation had the *spiro* ring system closer to the intracellular side. The full set of interactions for spiperone bound in this orientation are shown in Figure 3.27, generated using Ligplot. The interactions implicated in mutagenesis experiments can be seen with the fluoro substituent and Ser$^{5.43}$ interacting, the hydrophobic interaction between spiperone and Phe$^{3.28}$ and the electrostatic interaction between spiperone and Asp$^{3.32}$. An additional hydrogen bond is also seen between the lactam oxygen of spiperone and Asn$^{7.45}$. There are a number of other hydrophobic contacts with spiperone and Val$^{2.61}$ and Tyr$^{7.43}$ in the region between helices two, three and seven and with Val$^{3.33}$, Phe$^{6.51}$ and Ile184$^{loop E2}$ in the region between helices four, five and six. The interactions between spiperone and Val$^{3.29}$ and Tyr$^{7.35}$, as implicated from mutagenesis experiments, are not seen. However these residues may be involved in secondary packing around spiperone as both Val$^{3.29}$ and Tyr$^{7.35}$ are in close proximity to residues interacting with spiperone.
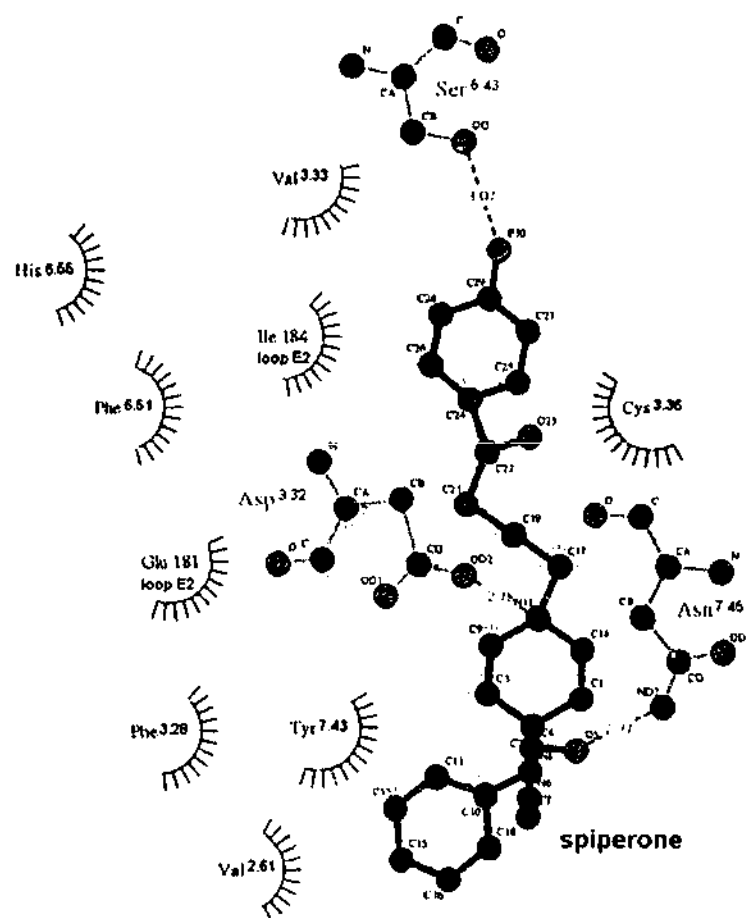
Figure 3.27. Ligplot analysis of the binding site interactions of spiperone bound with the para-fluoro end in the region between helices four, five and six of the D$_2$ receptor model.

Alternative orientation of Spiperone bound in the D_2 dopaminergic model

In an alternative placement of spiperone in the D_2 dopaminergic binding site, spiperone was orientated so that the *para*-fluoro end of spiperone was in the region between helices two, three and seven (Figure 3.28). Both starting orientations of spiperone, with the *spiro*-ring system facing inwards and outwards were also investigated. The orientation of spiperone with the *spiro*-ring system closer to the intracellular side is the same as that proposed by Mansour et al.[231], see Figure 3.13. Of the models generated with the *spiro*-ring system facing inwards and outwards a similar orientation was seen six and seven times, respectively, for each set. In none of these models was the inverted spiperone seen, in which the *spiro*-ring system was facing in the opposite direction to which it was initially placed. This result was not surprising, as the bulky end of spiperone was not in the crowded region between helices two, three and seven.
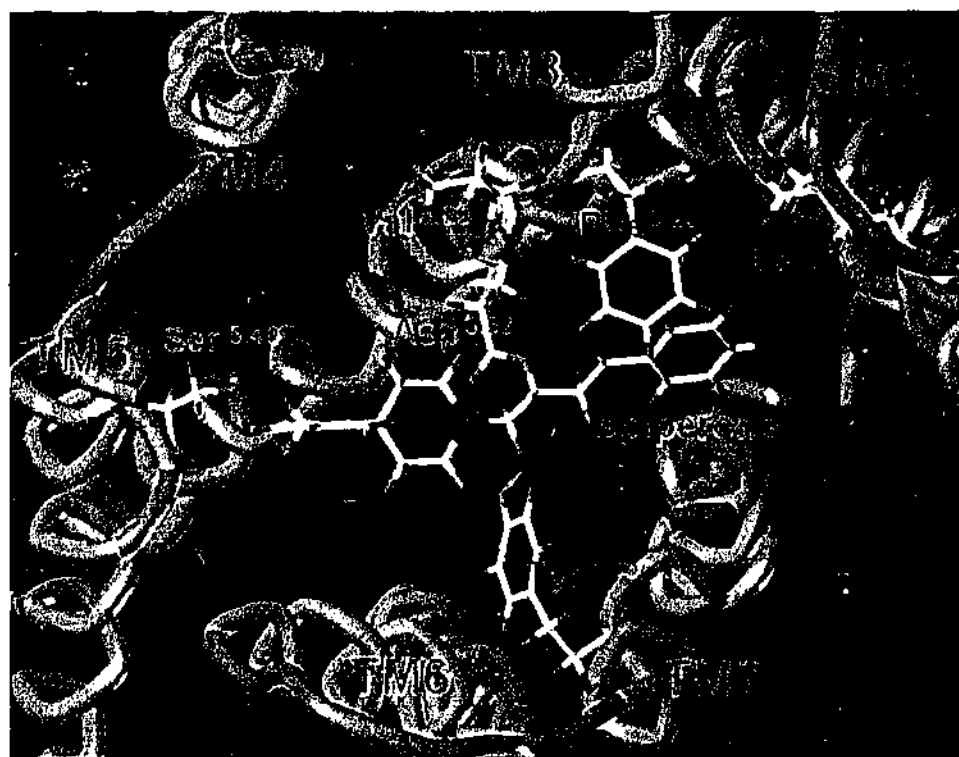
Figure 3.28. *Spiperone bound in the D_2 dopaminergic receptor, with the para-fluoro end in the region between helices two, three and seven. The residues implicated in the binding region are shown labelled.*

Each preferred orientation of spiperone also conformed well to the residues implicated in mutagenesis studies, see Figure 3.29. Both orientations showed interactions with $Asp^{3.32}$, $Phe^{3.28}$ and $Ser^{5.46}$, and the spiperone conformation with the intracellular *spiro*-ring had an additional hydrogen bond between the lactam nitrogen and $Ser^{5.46}$.

The resultant orientations also had a number of hydrophobic and polar interactions in common, with spiperone interacting with $Tyr^{7.43}$, $Phe^{6.51}$, $Val^{3.33}$ and $Cys^{3.36}$, $His^{6.55}$, respectively. The similarity between the two alignments made it difficult to work out which was the more likely orientation of spiperone in the binding site of the D_2 models created.
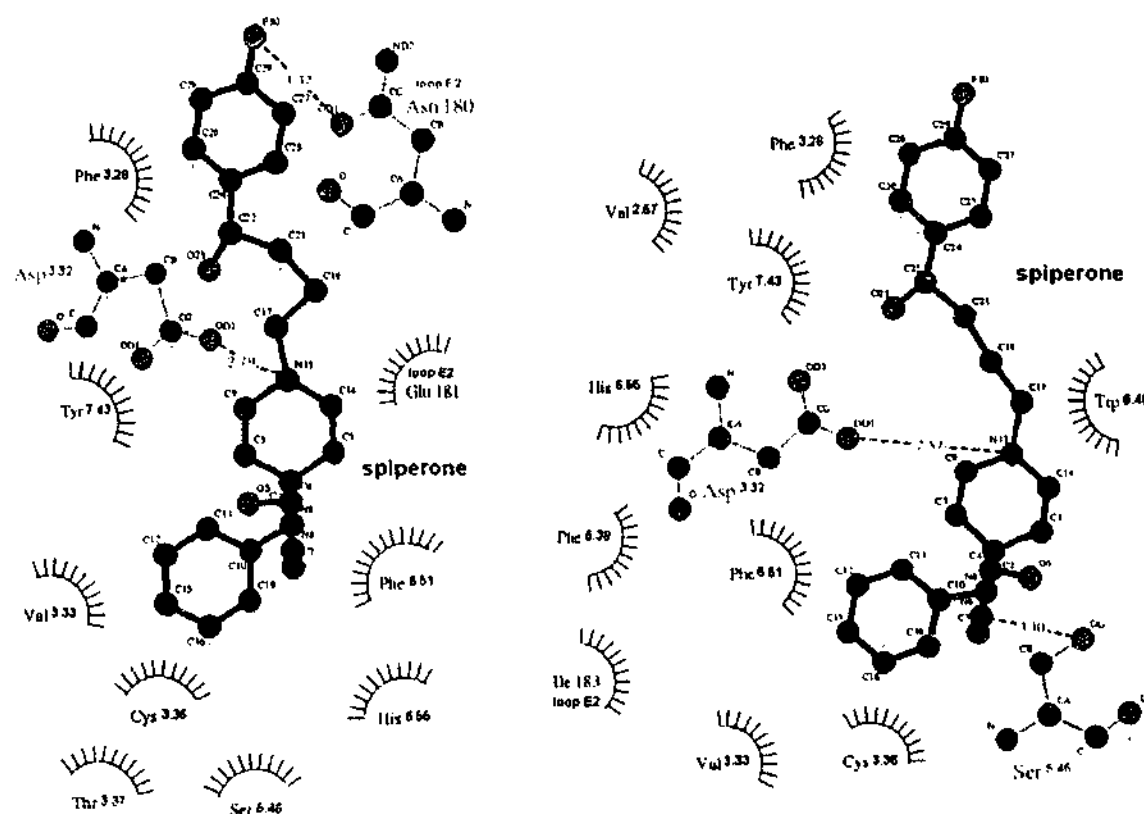
Figure 3.29. *Diagrams generated from Ligplot analysis of the binding site interactions of spiperone bound with the para-fluoro end in the region between helices two, three and seven of the D_2 receptor model. The diagram on the left was from the placement of spiperone with the spiro-ring system closer to the extracellular surface, and the diagram on the right with the spiro-ring system closer to the intracellular side.*

As it was difficult to decide on the orientation of spiperone in the binding site the placement of spiperone in relation to all the residues implicated in the mutagenesis experiments was examined in greater detail. From this analysis it could be seen that the spiperone conformation with the *spiro*-ring system closer to the extracellular surface was in much closer contact with all residues implicated in mutagenesis. Figure 3.30 shows spiperone in much closer contact to the $Leu^{2.64}$ residue when the *spiro*-ring system is closer to the extracellular side. When the intracellular orientation is observed it can be seen that spiperone is almost two turns further down TM2 than $Leu^{2.64}$. In

addition to this it was seen that the extracellular *spiro*-ring orientation was able to protect all the residues that were proposed to be in close proximity to the sulpiride or *N*-methylspiperone binding site.
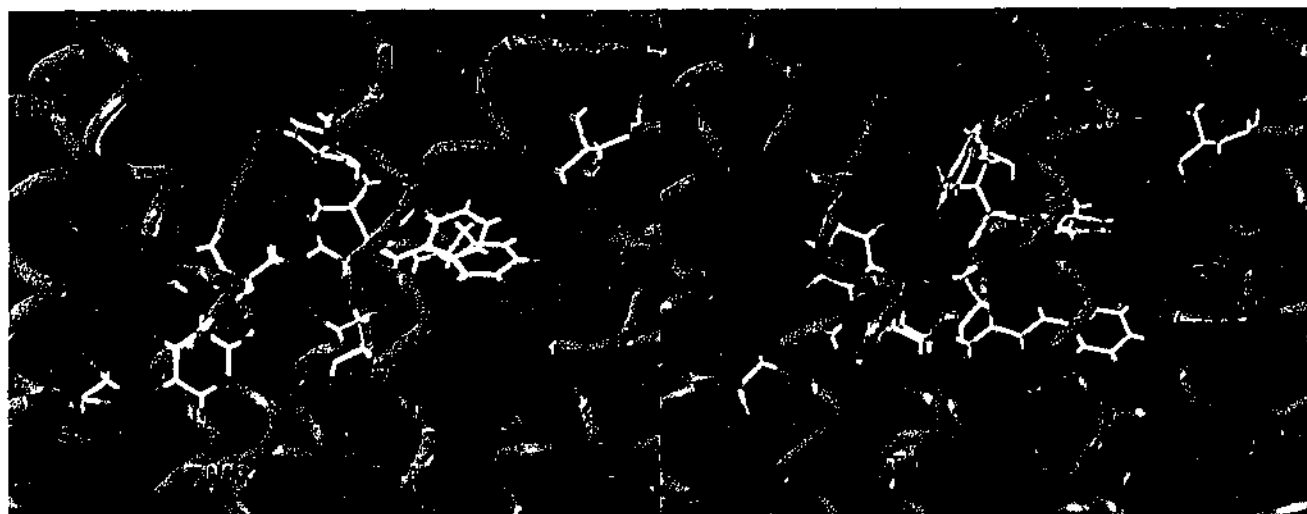
Figure 3.30. *Placement of spiperone in the D$_2$ models with the spiro-ring system closer to the extracellular side (left) and closer to the intracellular side (right). The orientation on the right is similar to that proposed by Mansour et al.*

Therefore for the final placement of spiperone in the D$_2$ receptor models there were only two possibilities, shown in Figures 3.26 and 3.30 (left). However when the initial model (Figure 3.26) was examined in relation to the sulpiride protection of residues, it could be seen that this orientation would not be able to protect all the residues implicated in sulpiride protection, as it was too far inside the receptor, see Figure 3.31. Therefore the orientation of spiperone that was used as the basis for extended pharmacophore superimposition was the orientation shown in Figure 3.30 (left), as it was in closer proximity to sulpiride protected residues.
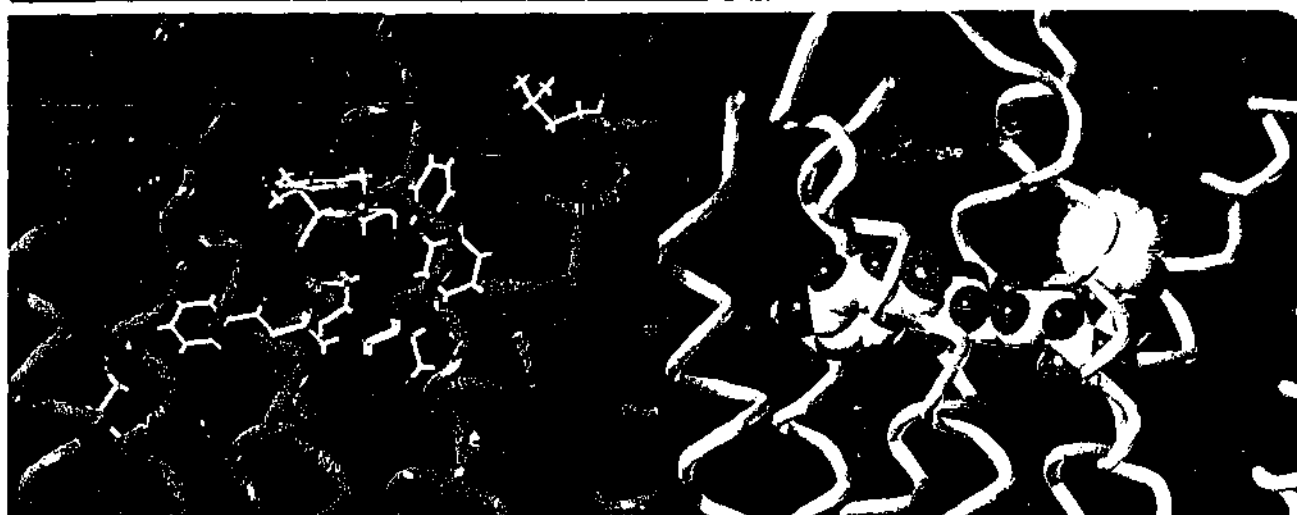
Figure 3.31. *Placements of spiperone in the D2 model with the para-fluoro end in the region between helices four, five and six, and the spiro-ring system closer to the intracellular side. The diagram on the right shows spiperone in spacefill mode and the residues identified to be protected by the binding of sulpiride or N-methylspiperone coloured in orange.*

Spiperone bound in the D_4 dopaminergic model

For the initial placement of spiperone in the D_4 receptor models, spiperone was orientated so that the *para*-fluoro end of spiperone was in the region between helices four, five and six. Both intra and extracellular orientations of the *spiro*-ring system were also investigated. As was seen in the D_2 analysis with spiperone bound in this orientation, there was again only one significant family of orientations (Figure 3.32). As when models were built with the *spiro*-ring system of spiperone closer to the extracellular surface we again saw the inversion of spiperone in four of the ten models created. In four other models created the phenyl-*spiro* end of spiperone was orientated to be perpendicular the axis of the helices, resulting in severe steric clashes with helices two and seven. The extra space in the region between helices two, three and seven in the D_4 models compared to the D_2 models was not able to accommodate the phenyl-*spiro* end of spiperone satisfactorily. As was seen with the D_2 models, spiperone was thought to be too far inside the receptor to protect residues implicated in proximity to the sulpiride or N-methylspiperone binding site.
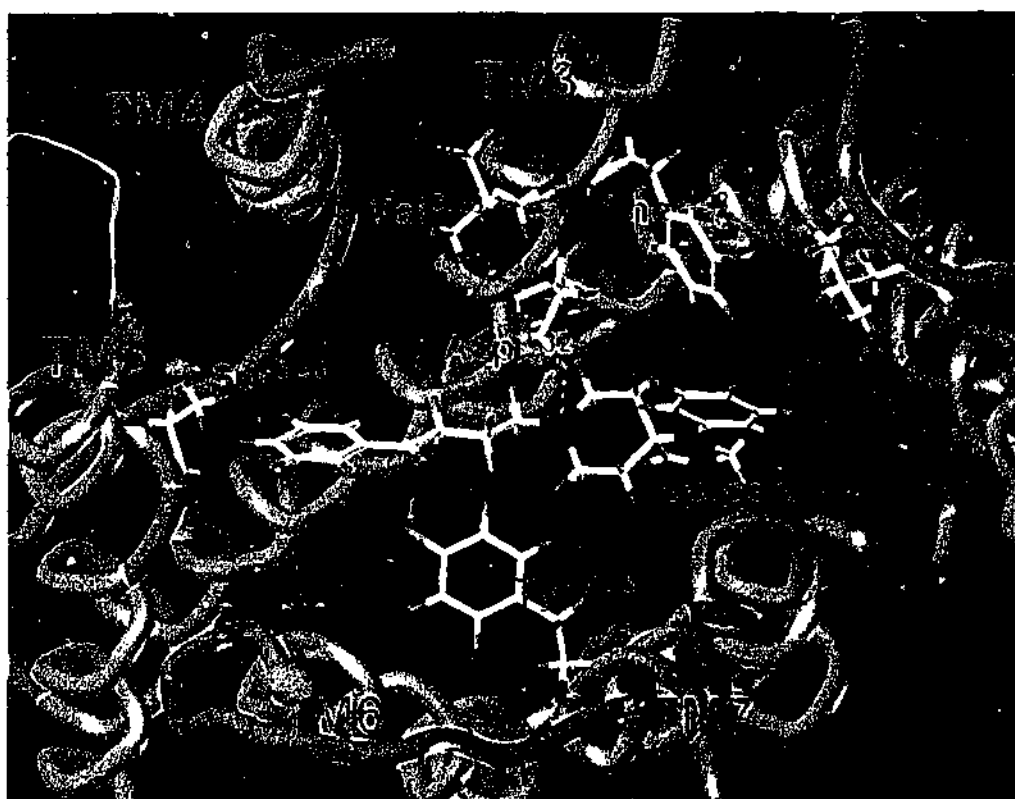
Figure 3.32. *Spiperone bound in the D_4 receptor model, with the para-fluoro end in the region between helices four, five and six. The residues implicated in the binding region are shown labelled.*

Alternative orientation of Spiperone bound in the D_4 dopaminergic model

In this placement of spiperone in the D_4 dopaminergic-binding site, spiperone was orientated so that the *para*-fluoro end of spiperone was in the region between helices two, three and seven. Both starting orientations of spiperone, with the *spiro*-ring system facing inwards and outwards were also investigated. The most significant families of orientations from the analyses conformed well to the residues implicated in mutagenesis studies. Both orientations showed interactions with $Asp^{3.32}$ and were in close proximity to $Ser^{5.46}$, $Phe^{2.61}$ and $Tyr^{5.38}$, see Figure 3.33. The spiperone conformation with the extracellular *spiro*-ring orientation had a number of hydrophobic contacts with $Trp^{7.40}$, $Tyr^{7.43}$, $Leu^{3.28}$ and $Val^{3.33}$ and an additional hydrogen bond between the keto oxygen and $Tyr^{7.39}$. The resultant orientations also had a number of hydrophobic and polar interactions in common, with spiperone interacting with $Phe^{6.51}$ and $Cys^{3.36}$, respectively (Figure 3.33)
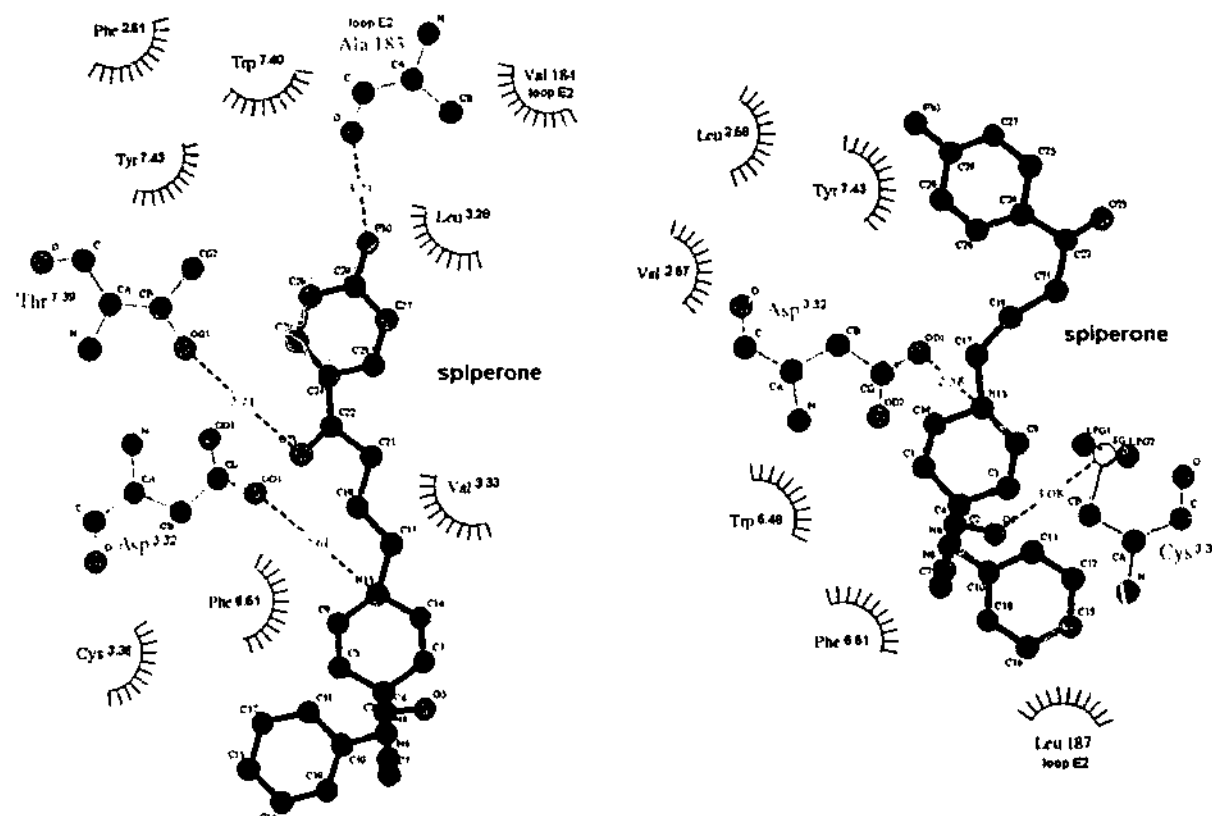
Figure 3.33. D_4 *ligplot analysis diagrams. Diagram on left is from the placement of spiperone with the spiro-ring system closer to the extracellular surface, and the diagram on the right with the spiro-ring system closer to the intracellular side.*

However again a similar theme to what was seen in the D_2 models produced, was also evident here. That is, when the *spiro*-ring system was closer to the intracellular side it appeared as though spiperone was too far inside the receptor to adequately protect residues t hought t o b e n ear t he s ulpiride o r *N*-methyl s piperone b inding site, (Figure 3.34)
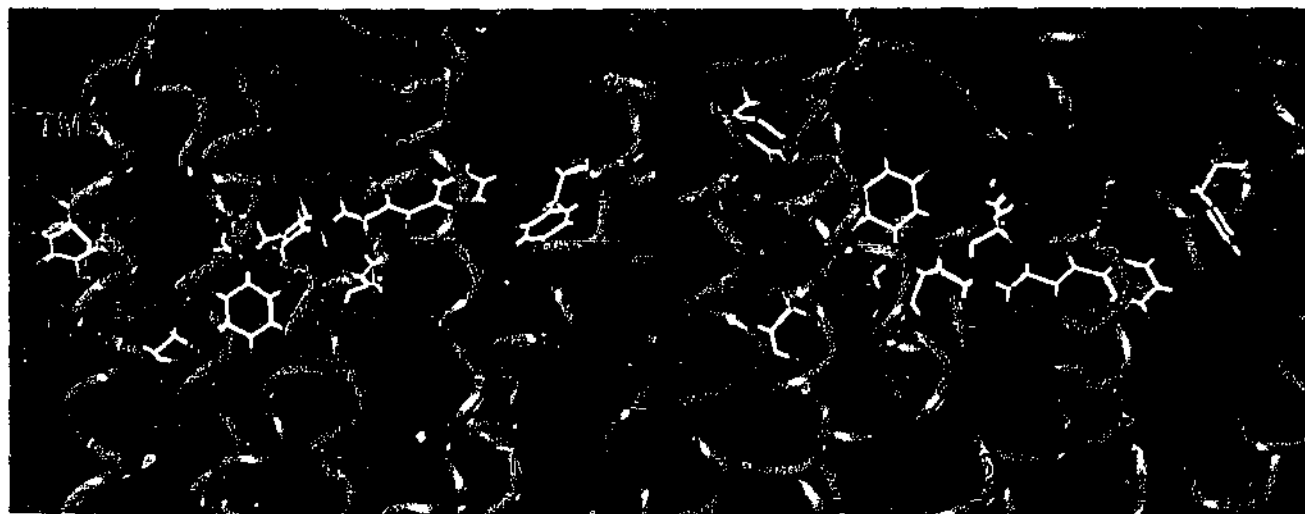
Figure 3.34. S piperone b ound w ith the s piro-ring s ystem c loser t o t he extracellular surface (left) and closer to the intracellular surface (right.*

Therefore a similar orientation of spiperone in the D_4 binding site, as compared to the D_2 binding site, was obtained.   This orientation was used as the basis for superimposition of the extended D_4 CoMFA model onto the D_4 receptor models.

3B.3.3 Analysis of the CoMFA Models in the Receptors

3B.3.3.1 Tricyclic CoMFA model

Clozapine bound in the D_2 dopaminergic receptor model

From the analysis of the D_2 models created it appears as though clozapine could bind in either orientation deduced. However the orientation between helices four, five and six is the proposed high affinity binding site as it conforms to the majority of the published mutagenesis data.   Therefore this orientation was used as the basis for the superimposition of the D_2 tricyclic CoMFA model developed in chapter 2, and the resultant model with steric and electrostatic fields is shown in Figure 3.35.
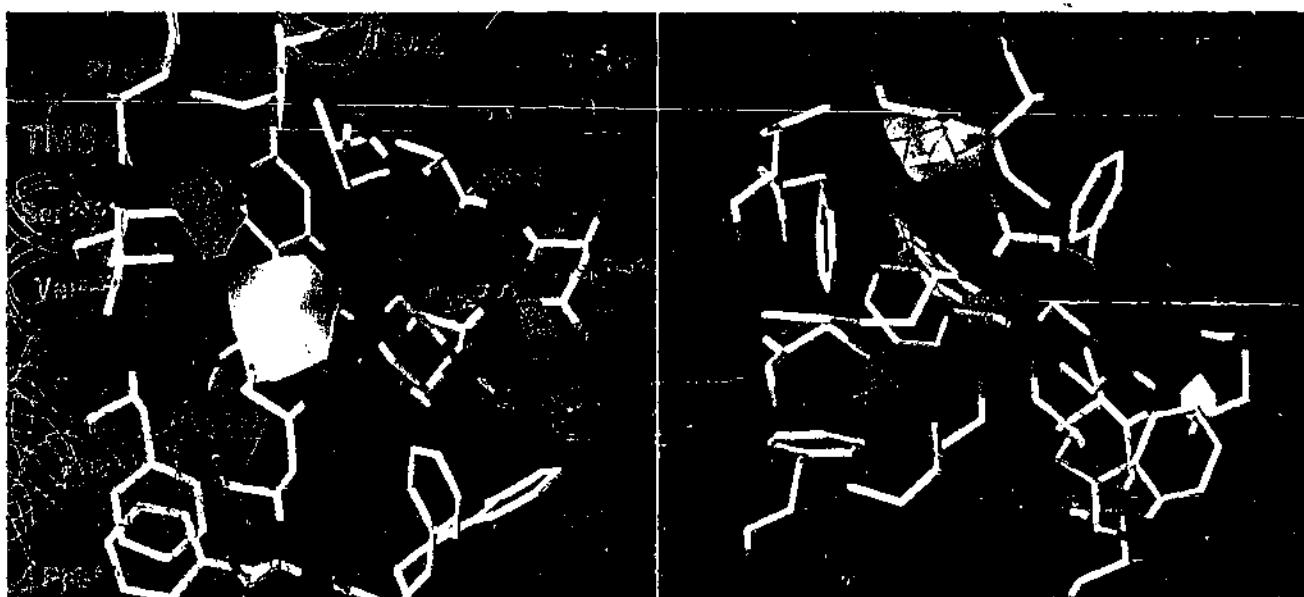
Figure 3.35. D_2 CoMFA model superimposed into the D_2 model receptor. The diagram on the left shows the extracellular view of the CoMFA model with residues within 3Å of clozapine, shown with carbons in orange, depicted. The helices of the receptor are shown in magenta line trace. The diagram on the right is the same depiction rotated 90° into the plane of the page.

The steric hindrance predicted in the model is easily seen with the large area above clozapine being due to the presence of His[6.55] and the smaller area past the distal

The steric hindrance predicted in the model is easily seen with the large area above clozapine being due to the presence of $His^{6.55}$ and the smaller area past the distal nitrogen being due to an unfavourable interaction with the backbone of TM3. The areas where additional bulk would be favourable, shown in green, are also seen not to be overlapping with any of the residues within 3Å of clozapine. The concordance of the steric fields from the CoMFA model with the receptor model indicates that this is a likely orientation of clozapine in the D_2 receptor.

The electrostatic fields from the CoMFA model are more difficult to explain within the confines of the receptor model. Although the regions where partial positive charge increases affinity (blue areas) on the tricyclic substructure indicate that carbon atoms are favoured here as opposed to nitrogen, as was seen in some compounds from Table 2.3 chapter 2. Carbon atoms within the tricyclic substructure would be preferentially tolerated within the hydrophobic cavity that clozapine is binding in. The red region on the hydrogenated nitrogen of the tricyclic substructure is a good indication that an electronegative heteroatom here would be favourable for binding, which is reiterated by the hydrogen bond placement between the nitrogen and $His^{6.55}$. (Figure 3.23).

This orientation of clozapine, and all tricyclic substructural compounds, in the D_2 receptor model aids in the generation of ideas for exploration of the groups that could be tolerated within the cavity of the model.

Clozapine bound in the D_4 dopaminergic receptor model

The analysis of the D_4 models created revealed that clozapine could bind in both orientations presented, although the highly favourable interaction of clozapine with $Phe^{2.61}$, as shown from mutagenesis data (Table 3.14) indicated that the bound orientation between helices two, three and seven was the high affinity binding site. The orientation of clozapine between helices two, three and seven was used as the template for the superimposition of the D_4 tricyclic CoMFA model generated in chapter 2 with the resultant model being shown in Figure 3.36.

Figure 3.36. D_4 CoMFA model superimposed onto clozapine in the D_4 receptor model. The diagram on the left shows the extracellular view depicted with residues within 3Å of clozapine, clozapine shown with carbons in orange. The helices of the receptor are shown in magenta line trace. The diagram on the right is the same depiction rotated 90° into the plane of the page.

Figure 3.36 reveals two unfavourable steric regions from the CoMFA model, shown in yellow. The largest of these regions is located on top of $Tyr^{7.43}$ just off TM7 and the smaller region located between helices two and three near $Leu^{3.28}$. Both unfavourable steric areas are in inaccessible areas of the receptor. The two regions where additional bulk is favourable to the binding affinity of compounds are shown in green. There is one area at the back of the tricyclic substructure surrounded by $Tyr^{7.43}$ and $Trp^{7.40}$, which would be ideal to explore in order to gain D_4 selectivity. The second region where steric bulk would be favourable is between TM6 and TM7 near $Trp^{6.48}$ and $Phe^{6.51}$. Both of the regions where additional bulk is favourable do not lie on top of any of the residues within 3Å of clozapine.

The electrostatic regions of the CoMFA model shows that partial positive charge is favourable on the hydrogenated nitrogen of the tricyclic substructure, which is an indication that a carbon would be preferential for increasing binding affinity. This is supported by the fact that there is no hydrogen bond to this atom, unlike what was seen in the D_2 model, and by the fact that this atom is sitting directly between two hydrophobic residues, $Tyr^{7.43}$ and $Phe^{2.61}$. This extremely hydrophobic area would indicate that a carbon would be preferable in this position.

3B.3.3.2 Extended CoMFA model

Spiperone bound in the D₂ dopaminergic receptor model

Analysis of the D_2 models created with spiperone bound revealed one dominant orientation that adhered to available mutagenesis data, (Figure 3.30 (left)). This orientation was used as the basis for the superimposition of the extended D_2 CoMFA model from chapter 2, giving the combined model shown in Figure 3.37.

Figure 3.37. *Extended D_2 CoMFA model superimposed onto spiperone in the D_2 receptor model. The diagram on the left shows the extracellular view depicted with residues within 3Å of spiperone. The helices of the receptor are shown in magenta line trace. The diagram on the right is the same depiction rotated 90° into the plane of the page.*

Figure 3.37 shows that the steric fields conform to the homology model presented however the multitude of small areas of favourable and unfavourable steric interaction make it difficult to gain any real understanding from the model. These are due to the large number of components used in the generation of the model initially. However it appears as though unfavourable steric interactions (yellow regions) between helices two, three and seven are due to clashes with Asn[180] and the backbone of TM7. The two areas of favourable steric bulk, which are devoid of overlapping residues, are near Phe[3.28] and in the vicinity of the *spiro* ring system. The electrostatic regions from the extended CoMFA model are also difficult to interpret, again due to the multitude of small areas of interaction. In general it appears as though the CoMFA fields generated from the extended CoMFA model are not particularly suited to superimpositioning within the D₂ receptor models created. Therefore a CoMFA model generated from a

smaller set of compounds that only occupied regions B and G of the pharmacophore model was superimposed onto the placement of spiperone in the D_2 receptor model. A CoMFA model that only covers regions B and G was generated, as spiperone is only proposed to reside in these areas of the pharmacophore model. This CoMFA model, generated from 94 compounds that only occupied regions B and G, had q^2 of 0.486, r^2 of 0.893 and was generated from five components. The smaller number of components made it easier to interpret, as there were not so many small areas of postulated interaction. Figure 3.38 shows the reduced CoMFA model superimposed within the D_2 receptor model. A similar arrangement of steric contributions can be seen in Figure 3.38 compared to Figure 3.37 however there is a large area of favourable steric bulk that overlaps with Glu[181] on the extracellular loop E2. This anomaly with Glu[181] is due to the reduced number of compounds being analysed within the model. In addition to this the steric contributions (Figure 3.38) occupy much larger areas and the unfavourable interaction at the *para*-fluoro end of spiperone now can be seen to be encompassing Leu[2.64] as well as Asn[180] from loop E2. The electrostatic contributions (Figure 3.38) also appear easier to interpret as the large area of favourable partial positive charge above spiperone can now be seen to interact with the Glu[181]. The negative charge on the glutamic acid would interact favourably with any partial positive charge in this area.
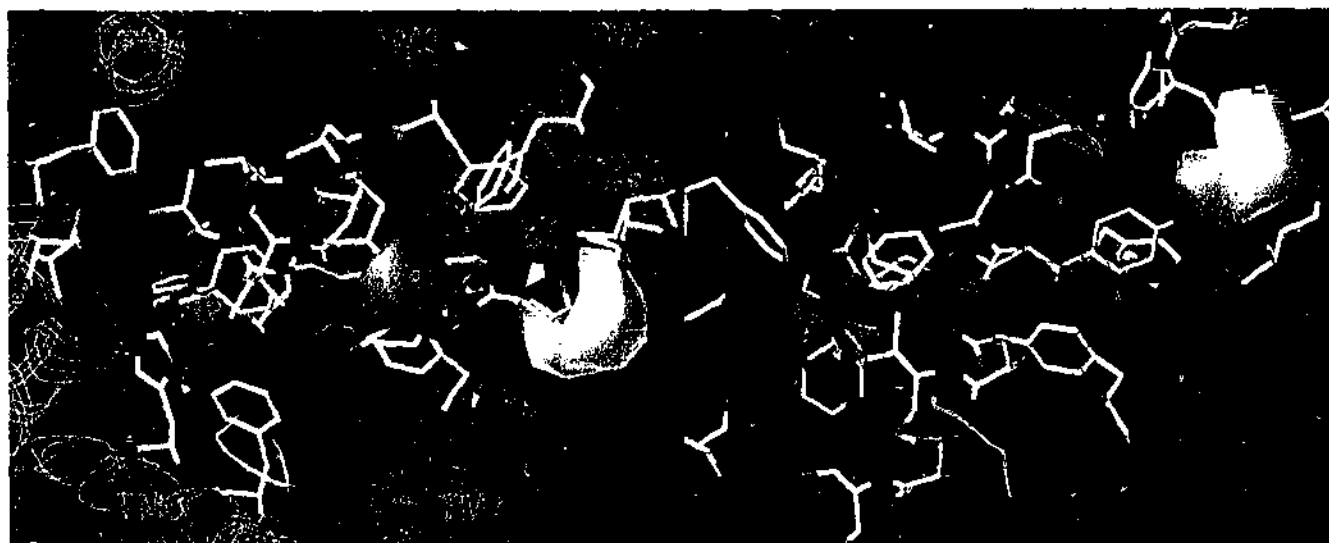
Figure 3.38. D_2 *CoMFA model generated from compounds residing in areas B and G of the pharmacophore superimposed onto spiperone in the D_2 receptor model. The diagram on the left shows the extracellular view depicted with residues within 3Å of spiperone. The helices of the receptor are shown in magenta line trace. The diagram on the right is the same depiction rotated 90° into the plane of the page. Residues are*

not labelled, as they are the same as those depicted in Figure 3.37, with the exception of Leu$^{2.64}$, which is labelled.

Despite the relative conformity of the reduced B/G CoMFA model within the receptor, when clozapine is placed into this spiperone bound D_2 receptor model in the same orientation as that seen in the pharmacophore relative to spiperone, a significant steric clash is seen between clozapine and TM7 (Figure 3.39). This would indicate that superimposition of the extended pharmacophore model using only spiperone as the basis may give spurious results. The inconsistency of the pharmacophore model, and in particular region A, may be due to a number of reasons, one being that when building homology models with Modeller it has a tendency to unrealistically fill the active site of a model when a ligand is not present.
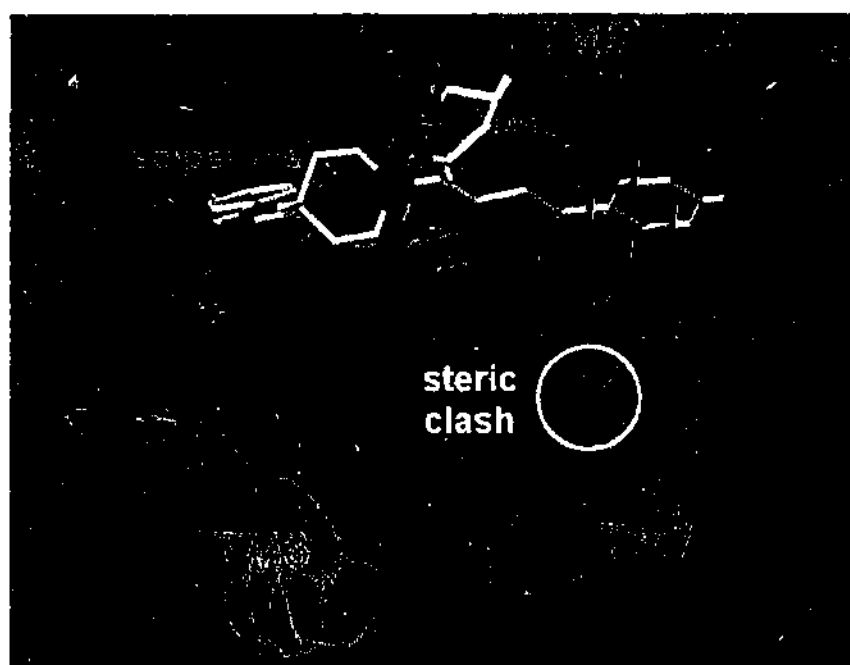
Figure 3.39.   Clozapine superimposed onto spiperone, as from the pharmacophore model of chapter 2, with spiperone in the orientation deduced from the homology models.

Another more likely reason that region A does not fit strictly within the region defined by the pharmacophore is that regions A and B of the pharmacophore model are not strictly defined areas in which this portion of the ligand must reside. As regions A and B of the pharmacophore model are both hydrophobic regions the interactions there are likely to be amorphous in nature and not defined to a particular geometry as a hydrogen bond may be. It is likely that their interactions within the receptor will move about to a

likely to be amorphous in nature and not defined to a particular geometry as a hydrogen bond may be. It is likely that their interactions within the receptor will move about to a moderate extent in order to find the lowest possible energy well for them and the receptor to reside in. This is what appears to have happened in the construction of spiperone in the D_2 receptor model. In fact, this non-conformity of the pharmacophore within the D_2 model is not that surprising, as we saw earlier that in the binding of clozapine within TM2, TM3 and TM7 the region here was extremely crowded. This crowded region would cause any compound not occupying both regions A and B to adopt a position that would be somewhere between the two. Thus when the pharmacophore from chapter 2 is superimposed back onto a docked compound not occupying both regions A and B we would expect to see some overlap with the surrounding residues. As a continuation of this logic we might expect the extended pharmacophore to superimpose well onto the models constructed with clozapine bound in the region between helices two, three and seven. Unfortunately this is not the case as we have the problem of the unrealistic filling of region G in the models constructed here. Therefore to accurately superimpose the extended CoMFA model into a D_2 or D_4 receptor model, models should be built with a ligand that occupies all regions A, B and G of the pharmacophore. There are a number of types of antipsychotic ligands that fit within this criterion, namely some of the diphenylbutylpiperidines such as fluspirilene or some of the older dihydrodibenzo[b,f]azepines such as carpipramine. However these were not looked at in this body of work, although these compounds do give insights into the design of new types of compounds that may be able to occupy regions A, B and G.

Thus although the extended CoMFA model does not fit ideally into the receptor model created, we can see that by also comparing the B/G CoMFA model we are able to gain some further insight. These additional insights aid in deciding which areas may be explored in the design of new compounds.

Spiperone bound in the D_4 dopaminergic receptor model

Analysis of the D_4 models created with spiperone bound revealed one dominant orientation that conformed with the mutagenesis data available (Figure 3.34 (left)). This orientation was used as the basis for the superimposition of the extended D_4

occupying regions B and G only, giving the combined models shown in Figures 3.40 and 3.41.

The D_4 CoMFA model for regions B and G generated from 93 compounds, had 5 components giving q^2 of 0.52 and r^2 of 0.86. The predicted activities of all compounds were within ±0.9 pK$_i$ log units of their measured values
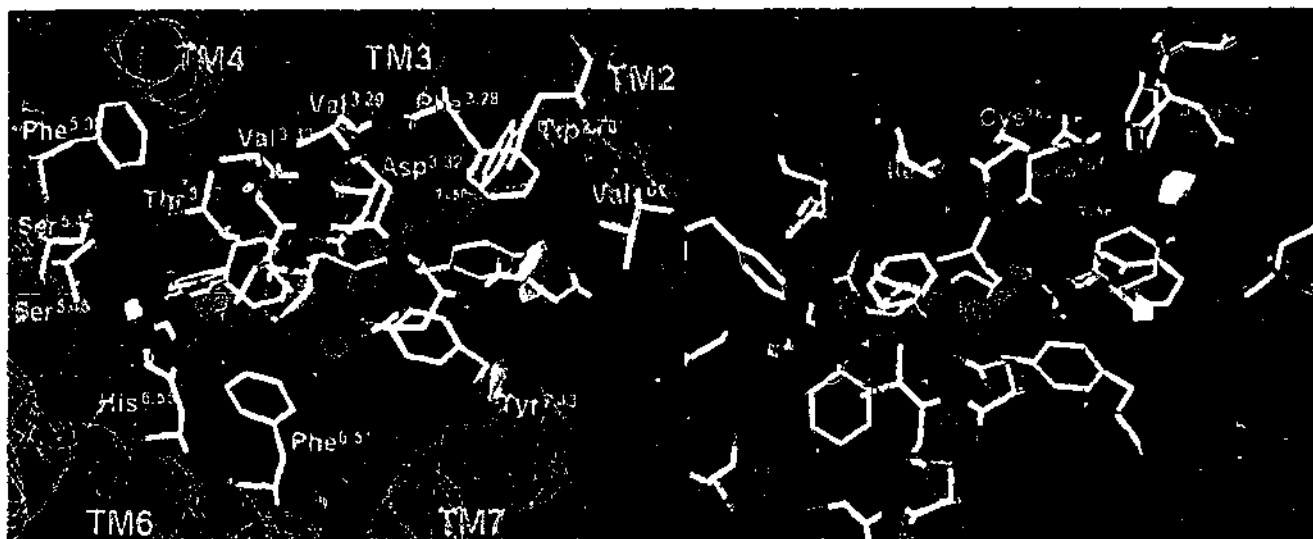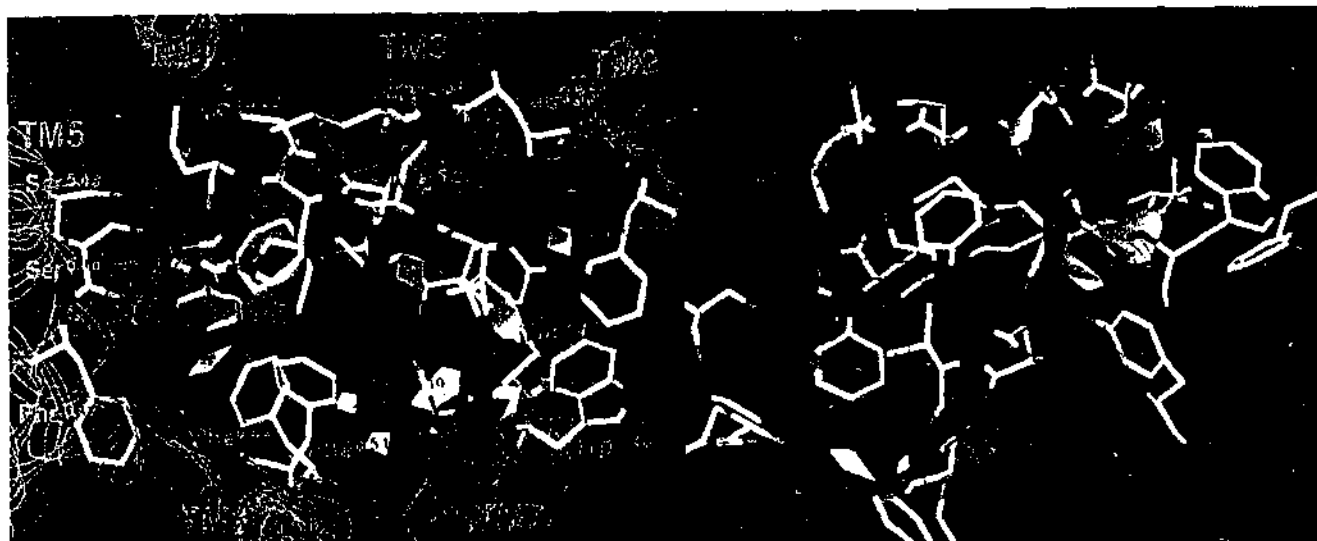

Figure 3.40.   Extended D_4 CoMFA model superimposed onto spiperone in the D_4 receptor model.   The diagram on the left shows the extracellular view depicted with residues within 3Å of spiperone.   The helices of the receptor are shown in magenta line trace.   The diagram on the right is the same depiction rotated 90° into the page.
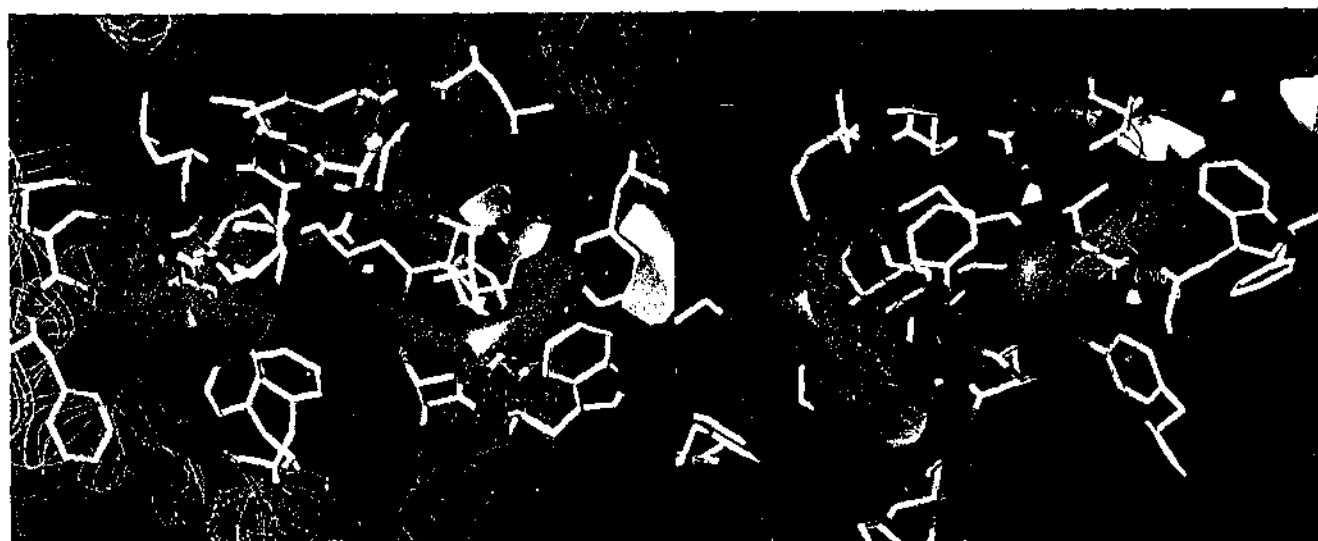
Figure 3.41. D_4 CoMFA model generated from compounds residing in areas B and G of the pharmacophore superimposed onto spiperone in the D_4 receptor model, showing steric fields only.   The depictions and residue labels are the same Figure 3.40.

A similar placement of steric fields between the D_4 and D_2 B/G CoMFA models is seen when comparing Figures 3.38 and 3.41, although it appears as though greater steric bulk is more favourable for region G near helices three and four for the D_4 B/G CoMFA model. The greater availability of space within this region for the D_4 receptor model would be due to the presence of two valine residues, $Val^{4.56}$ and $Val^{4.60}$, which correspond to $Ile^{4.56}$ and $Leu^{4.60}$, respectively, in the D_2 model. In addition to this difference the region where additional steric bulk is favourable near TM7 is much greater in the D_4 B/G CoMFA model compared to the D_2 model. The reasoning behind this is not due to any residues in TM7, as the TM7 interacting residues are identical between D_4 and D_2, but due to $Leu^{1.39}$ in the D_2 model occupying a much greater space compared to the $Gly^{1.39}$ on the D_4 model. The $Leu^{1.39}$ on the D_2 receptor not only occupies a much greater area but also affects the packing of residues in its vicinity, including $Tyr^{7.43}$, which encroaches further into the binding site. Another favourable steric region in common between the two B/G CoMFA models is the region interacting with the extracellular loop E2. This anomaly is seen in both the B/G CoMFA models but not in both the extended CoMFA models, see Figures 3.37 and 3.40, indicating the compounds left out of the analysis are important for validating fields seen in the B/G CoMFA model. The electrostatic fields, see Figure 3.40, indicate that some of the atoms in the backbone of the extracellular loop E2 are interacting favourably with the bound ligand. However the multitude of small regions make it difficult to assign these regions to any particular residue. Some of the electrostatic differences between the D_4 and D_2 receptors in this region could be due to the presence of $Met^{3.35}$ in the D_4 receptor model, which corresponds to a $Leu^{3.35}$ in the D_2 receptor. $Met^{3.35}$ can be seen to be interacting with the top of the ligand spiperone when bound in the D_4 receptor.

A similar problem was seen as for the D_2 receptor model, with the steric clash between clozapine and TM7 when clozapine is placed into this spiperone bound D_4 receptor model in the same orientation as that seen in the pharmacophore relative to spiperone. This is due to the same reason as was mentioned for the D_2 receptor, that the hydrophobic interactions are amorphous in nature and not defined to a particular point.

Thus although the extended CoMFA model does not fit ideally into the receptor model created we again see that the B/G CoMFA model supports both the position of

spiperone in the receptor and the smaller fields from the extended CoMFA model. This

will aid in the design of new compounds and their fine-tuning for the D_4 receptor.

3B.4 CONCLUSION: STRUCTURE-BASED DESIGN

3B.4.1 Model Generation

3B.4.1.1 Tricyclic model

The placement of clozapine in the D_2 and D_4 receptor models differed quite significantly. The most significant D_2 model as seen in six of the ten models created, had clozapine placed in the region between helices four, five and six (Figure 3.17); this model was also consistent with the majority of the mutational data.

The most significant D_4 model, as seen in nine of the ten models created, had clozapine bound between helices two, three and seven (Figure 3.24). Although in seven of the ten models we saw a similar placement of clozapine between helices four, five and six, this model did not conform to most of the mutational data. In addition to this when we saw the $Phe^{4.62}Cys$ D_2 mutation (Table 3.14) we saw a 10 fold drop in the affinity of clozapine at the mutant D_2 receptor. The reason behind this is that the $Phe^{4.62}$ stabilises the placement of $F^{5.38}$ in the D_2 receptor that interacts with clozapine. As $Phe^{4.62}$ is $Cys^{4.62}$ in the D_4 receptor we do not have this additional stabilisation, making this site a lower affinity binding site at the D_4 receptor.

3B.4.1.2 Extended Model

The final placement of spiperone in the D_2 (Figure 3.30, left) and D_4 (Figure 3.34, left) receptor models did not differ in the manner that was seen for the binding of clozapine. Both final models with spiperone bound conformed well to the mutational data and were able to protect all the residues implicated in the 'protection by sulpiride'.

However it should be noted that the conformation of spiperone bound in the models produced here was that used in the pharmacophore elucidation. It is highly likely that other possible conformations of spiperone may bind in the D_4 and D_2 receptors. Indeed when looking at one of the excellent papers[26] by Javitchs' group they have spiperone

bound in a different conformation to that proposed to fit within the constraints of the pharmacophore developed in Chapter 2. Even the orientation of spiperone bound in the model presented by Shi *et al.*[26] is different from that proposed here.

Unfortunately not all the possible conformations of spiperone can be looked at and built into receptor models in the body of this work.

3B.4.2 Analysis of the CoMFA Models in the Receptors

3B.4.2.1 Tricyclic Model

The tricyclic CoMFA model conformed well to the two different placements of clozapine in the receptors. The areas of additional space present in the D_4 CoMFA model were seen in the placement of clozapine in the D_4 receptor.

3B.4.2.2 Extended Model

The extended CoMFA models conformed reasonably well to the receptor models created but the overlapping of some areas with TM7 indicated these were not ideal. The subsequent generation of CoMFA models occupying regions B and G conformed well in regards to the lack of overlap with TM regions. However the overlap of other steric regions with the extracellular loop E2 indicated that B/G CoMFA models alone did not generate ideal fields. Overall when looking at the regions the two differing models had in common, a good agreement could be seen between the receptor and CoMFA models.

Chapter 4

pK_a Prediction

4.1 pK$_a$ INTRODUCTION

It has long been recognised that the pK$_a$, see inset box, of a compound plays a key role in affecting many process involved in its pharmacokinetics and pharmacodynamics[1]. Indeed, in order for a drug to get to a site of action it may need to cross many membranes and/or barriers. The ability of a drug to cross a membrane by lipid diffusion will depend on its partition coefficient between the aqueous phase and the lipid phase of a membrane. In general the higher the partition coefficient, the higher will be the affinity for the lipid membrane and the more rapidly the drug is absorbed. In reality there are many other processes such as active transport that also

> What is pK$_a$?
>
> According to the Henderson-Hasselbalch equation, the relationship between pH, pKa, and the relative concentrations of an acid and its salt is as follows:
>
> $$pH = pK_a + \log\left(\frac{[A^-]}{[HA]}\right)$$
>
> where [A$^-$] is the molar concentration of the salt (dissociated species) and [HA] is the concentration of the undissociated acid. Therefore when the concentrations of the salt and acid are equal, the pH of the system equals the pKa of the acid.

affect this process. However, as the non-ionised form of the drug is of main concern in the equilibrium between the lipid and aqueous phases, the pH of the aqueous phase will affect the overall partition coefficient of an ionisable substance.

The relationship between the observed overall partition coefficient (P$_O$) and the partition coefficient for the non-ionised form (P$_N$) is given by the equation[1].

$$\log\left(\frac{P_N}{P_O} - 1\right) = pH - pK_a$$

Of key importance to the pharmaceutical industry are membranes that affect the bioavailability of a compound (i.e. in the gut), partitioning within the body (e.g. blood brain barrier (BBB)) and those associated with various ADME (adsorption, distribution, metabolism and excretion) characteristics. These membrane crossing properties are extremely important when designing effective antipsychotics (APDs), because in order to cross the BBB an APD should preferentially be in an unionised state, and to act on

the target receptor an ionic state is most often required. Therefore the pK_a window that an APD should possess is rather limited, effectively 1-2 pK_a log units, and thus this must be carefully controlled.

Therefore as pK_a prediction is so important in the modern drug discovery process, it seems only a natural extension of the computational chemist to be able to predict these properties of their virtual compounds and also to advise ways of fine tuning a compound for optimum bioavailability. This is a very difficult process for APDs as many are amines and heterocyclic compounds, however a feasible method is discussed later.

There are a number of ways to measure and predict the pK_a of a compound. Well established titrimetric methods[2] of measuring pK_a are available, but these all require samples of the compound itself. Current trends are such that it is becoming more important to determine the pK_a and other characteristics associated with a drug, before synthetic work is undertaken with the aim of avoiding making compounds that are predicted to have poor pharmacokinetics. A variety of theoretical prediction methods have been established based on semi-empirical and *ab initio* quantum mechanical calculations as well as linear free energy relationships (LFER).

Programs such as ACD Labs[3] pK_a predictor and Pallas' pKalc[4] use LFER, which although very fast, requires prederivation of a vast number of fragment constants and correction factors to be used in the estimation. Additionally LFER can be prone to large errors if compound fragments are not contained within the fragment library used to calculate the pK_a.

Programs such as Jaguar[5] utilise *ab initio* quantum mechanical calculations employing self consistent reaction field (SCRF) continuum treatment of solvation and systematic corrections to predict pK_a. Whilst these methods are very reliable they are very expensive in the computational time required for such predictions, with drug-sized compounds taking approximately half a day depending on their size and the speed of the computer.

Other methods that have also been explored have used classically derived semi-empirical quantum chemistry descriptors. The first of these studies, by Grüber and Buss, emerged in the late 1980's and used the energies of the HOMO and LUMO as well as atomic charges in combination with differences in free energy of the neutral and anionic states[6] for pK_a prediction. More recently Citra[7] from the Syracuse organisation used Coulson charges and bond orders to predict pK_a. These methods are almost as fast as fragment based methods with compounds taking approximately a few minutes per pK_a prediction. These studies will be outlined in full in the discussion.

We also have used semi-empirical methods in our prediction of pK_a, as they are very fast and reasonably accurate. although our method differs from those of Grüber and Buss[6] and Citra[7] in so far as we have used descriptors derived from *frontier electron theory* and not classical semi-empirical methods.

Frontier electron theory[8] is based on the assumption that a chemical reaction should occur at the position of largest density of the electrons in the frontier orbitals. The frontier orbitals are defined according to the type of the reaction, the highest occupied molecular orbital in the electrophilic reaction, the lowest unoccupied molecular orbital in the nucleophilic reaction, and both of the above in a radical reaction. The theory is based on splitting the energy of the whole system into its σ- and π- components. During the process of a chemical reaction the σ-part of the energy of the system reaches a maximum, at which point a transition state is defined. Whilst in this transition state, hyperconjugation takes place between the initial π-electron system and the pseudo-π-orbital located at the reaction centre (Figure 4.1).
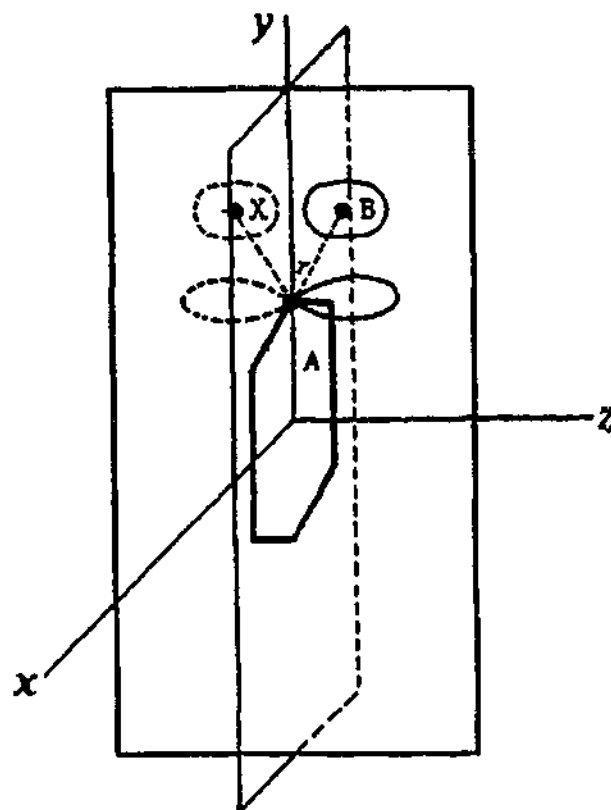
Figure 4.1.   *The transition state of the conjugated compound showing the pseudo-π-orbitals, X and B, located at the reaction centre.*

By means of a perturbation method, the π-part of the energy of the system involving the pseudo-π-orbital is obtained.   The *frontier electron theory* is then derived from an approximate expression of this perturbation energy.   This enables us to understand many of the features of the transition state and the many dynamic processes that involve transition states.   One of the dynamic processes that has already been looked at is hydrogen bonding[9], where a correlation between a number of *frontier electron theory* descriptors, (electrophilic superdelocalisability and atom self polarisability) was found with hydrogen bond strength.   From this we envisaged that it should be possible to look at other dynamic processes such as disassociation constants with *frontier electron theory* descriptors.

4.2 METHOD

4.2.1 Datasets

The experimental pK_a datasets were extracted from the PHYSPROP© database[10] (April 2001) which contains physical and chemical properties for over 24,950 compounds, of which 1847 compounds have pK_a values. As our interest was in predicting pK_a of molecules (or molecular fragments) with relevance to the pharmaceutical industry, a variety of filters were applied to the 1847 compounds to reduce the dataset in size. This included a molecular weight cut-off of 700, and that most drugs fall below a molecular weight of 500[11]. Those compounds which contained toxic functionality were also excluded. In addition, all the mixtures of compounds and salts were removed. It was also decided that only compounds whose pK_a was experimentally determined between a 10°C – 30°C temperature range should be considered for the basis of our analyses.

The remaining 1,671 compounds were imported into a Unity[13] database to allow for substructure searching. Our initial efforts were aimed at exploring the functional groups studied by Grüber and Buss[6]. This included phenols, benzoic acids and aliphatic acids. Later work included anilines, amines and heterocyclic compounds to expand the range of functional groups investigated. Simple substructure searches were conducted to extract the compounds of interest and these searches were refined to ensure no ambiguity within the datasets. For example compounds that contained more than one functional group from those described above (e.g. anilinophenols) were analysed to ensure the pK_a value given was for the functional group of interest (see Table 1). If the pK_a value could not be definitively assigned to the group under study the compound was removed. Additionally if more than one identical functional group of interest existed on the compound in question, they were removed, as the pK_a value could not be reliably assigned to one particular group, except in cases where an identical electronic environment existed for both groups.

Grüber and Buss[6] also demonstrated that further splitting of the datasets resulted in improved prediction of pKa values. Therefore the substituted aromatic datasets (phenols, benzoic acids and anilines) were split into two main groups, ortho substituted

compounds and para/meta substituted compounds. The ortho substituted compounds of the phenols and benzoic acids were again divided into three smaller subsets of intramolecularly hydrogen bonded compounds, sterically constrained compounds and the remaining ortho substituted compounds.

Criteria for inclusion into the intramolecularly hydrogen bonded subset involved the substituent of interest being next to a suitable heteroatom, see Figure 4.2. This excluded ortho-nitro substituted groups as they are poor hydrogen bond acceptors[14].
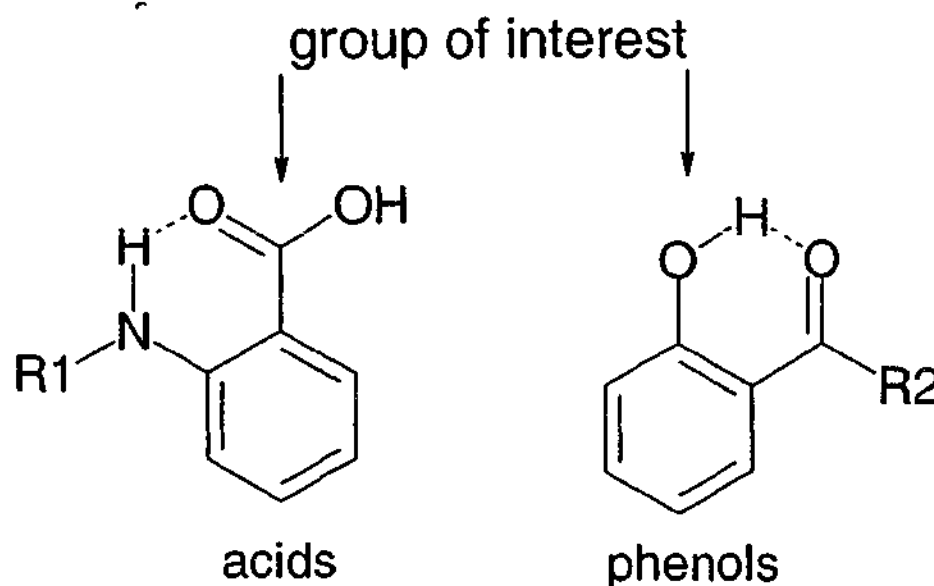
Figure 4.2. Examples of aromatic intramolecularly hydrogen bonded compounds.

Criteria for inclusion into the sterically constrained subset of ortho substituted compounds required that 2,6 substitution be present with substituents of a sufficient size to hinder the group of interest. This excluded 2,6-diflouro substituted compounds but included 2,6-dichloro substituted compounds and substituents of this size or greater, see Figure 4.3. Visual analysis shows that 2,6 substituents with a van der Waals radius of $1.75 \text{Å}^{[15]}$ or greater demonstrated a steric clash between the phenolic or anilinic hydrogen and the ortho substituent. Only 2,6-dimethyl groups or larger were considered for the sterically constrained subset of the aromatic carboxylic acids.
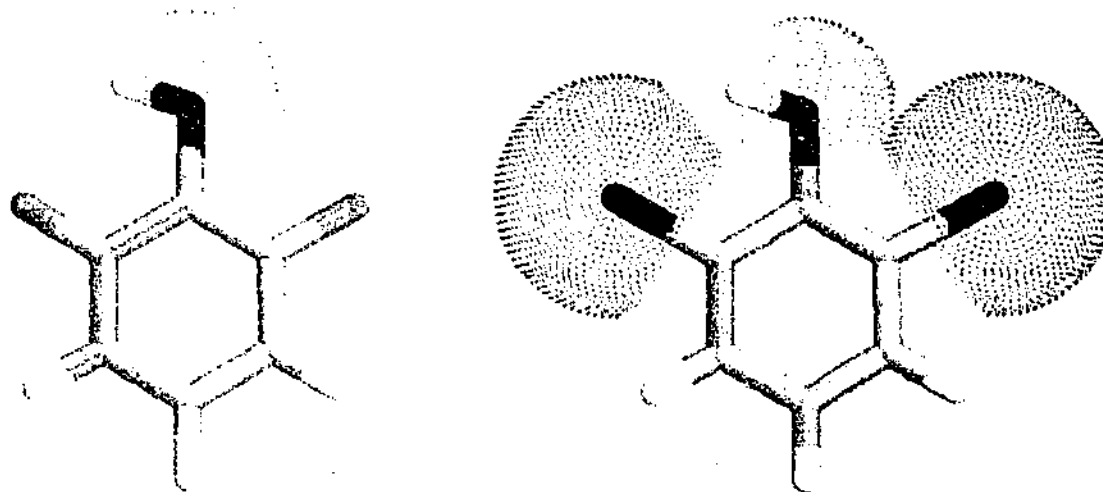
Figure 4.3. *2,6-diflourophenol and 2,6-dichlorophenol with van der Waals radius dot surfaces shown. The larger van der Waals radius of chlorine, 1.75Å, is seen to be sterically hindering the hydroxy group.*

Subdivision of the aliphatic carboxylic acids into amino acids and remaining carboxylic acids along with subdivision of the amines into primary, secondary and tertiary groups was also undertaken to shed more light on these diverse sets.

The heterocyclic compounds were divided into groups depending on the main ring system within the compound. This resulted in five main groups of heterocyclic pyridines, pyrimidines, imidazoles, benzimidazoles and quinolines. The pyridines were further split into ortho substituted and meta, para substituted subsets. Unfortunately there were not large enough groups of other heterocyclic ring systems, complying to our selection criteria, contained within the database for regression analyses. A number of other heterocyclic ring systems were also considered for study, however, as these often contained only a few examples they were not examined to maintain good statistical practices.

The six main sets comprised over 700 compounds with 175 phenols, 99 benzoic acids, 185 aliphatic acids, 55 anilines, 77 amines and 150 nitrogen containing heterocyclic compounds.

These compounds and their further subdivisions are available as supplementary material in the appendix.

4.2.2 Software

All the structures were initially extracted as 2D MOL SD files and converted into 3D models using Corina[16]. The QM calculations were carried out with a modified version of Mopac 6.01 (Peter Bladon, Interchem Chemical Services, Glasgow), able to calculate additional atomic properties (keyword: PROPER)[17]. Both the Sybyl[13] and Tsar[18] software packages were used for data manipulation and statistical analyses, primarily using the default parameters.

The modified version of Mopac 6.01 has been interfaced to Sybyl through an in-house SPL (Sybyl Programming Language) script. The script works on a Sybyl molecular database, creating the Mopac input file for each molecule in the databases and running Mopac. Once the QM calculation is completed for all the structures, a Sybyl spreadsheet is created and the relevant properties are imported. Subsequent QSPR (quantitative structure-property relationship) analyses are relatively straightforward.

4.2.3 Theoretical Descriptors

All the structures were fully optimised (EF routine, PRECISE), prior to any parameter calculation using the AM1[19] Hamiltonian. The PM3 Hamiltonian is also available within Mopac, although previous work examining hydrogen bonding strength[9] showed that AM1 and PM3 were very similar, with AM1 performing slightly better at times. Therefore PM3 was not considered any further.

Beside the charge (Q), as defined in Mopac, some additional atomic properties[8] were computed from the eigenvectors $c_{\alpha j}$ and the eigenvalues λ_j where α refers to the atomic orbital (i.e. s, p_x, p_y and p_z) and j to the molecular orbital. Given a molecule with N molecular orbitals, whose levels from 1 to m are occupied, and an atom p with q atomic orbitals, these properties are defined as follows.

Electrophilic frontier electron density (FE)

$$FE(p) = \sum_{\alpha=1,q} c_{\alpha m}^2$$

is the sum of all the squared eigenvectors of p on the HOMO.

Nucleophilic frontier electron density (FN)

$$FN(p) = \sum_{\alpha=1,q} c^2_{\alpha(m+1)}$$

is the sum of all the squared eigenvectors of p on the LUMO.

Electrophilic superdelocalisability (SE)

$$SE(p) = 2 * \sum_{j=1,m} \sum_{\alpha=1,q} \left(c^2_{\alpha j} / \lambda_j \right)$$

where the sum is over all the atomic orbitals of p and all the occupied molecular orbitals.

Nucleophilic superdelocalisability (SN)

$$SN(p) = 2 * \sum_{n+1,N} \sum_{\alpha=1,q} \left(c^2_{\alpha j} / - \lambda_j \right)$$

where the sum is over all the atomic orbitals of p and all the unoccupied molecular orbitals.

Radical superdelocalisability (SR)

$$SR(p) = \sum_{j=1,m} \sum_{\alpha=1,q} \left(c^2_{\alpha j} / \lambda_j \right) + \sum_{j=m+1,N} \sum_{\alpha=1,q} \left(c^2_{\alpha j} / - \lambda_j \right)$$

Atom self-polarisability (ALP)

$$ALP(p) = -4 * \sum_{j=1,m} \sum_{\alpha=1,q} \sum_{k=m+1,N} \left(c^2_{\alpha j} * c^2_{\alpha k} / \lambda_i - \lambda_k \right)$$

where the sum is over all the atomic orbitals of p and all the occupied and unoccupied molecular orbitals.

All the above properties have been calculated on and around all atoms in the groups of interest, see Figure 4.4.

The following molecular properties were also taken into account:

Energy of the highest occupied molecular orbital (E_{HOMO})

Energy of the lowest unoccupied molecular orbital (E_{LUMO}).

phenols

aromatic
carboxylic acids

aliphatic
carboxylic acids

anilines

amines

pyridines

pyrimidines

imidazoles

benzimidazoles

quinolines

barbituates
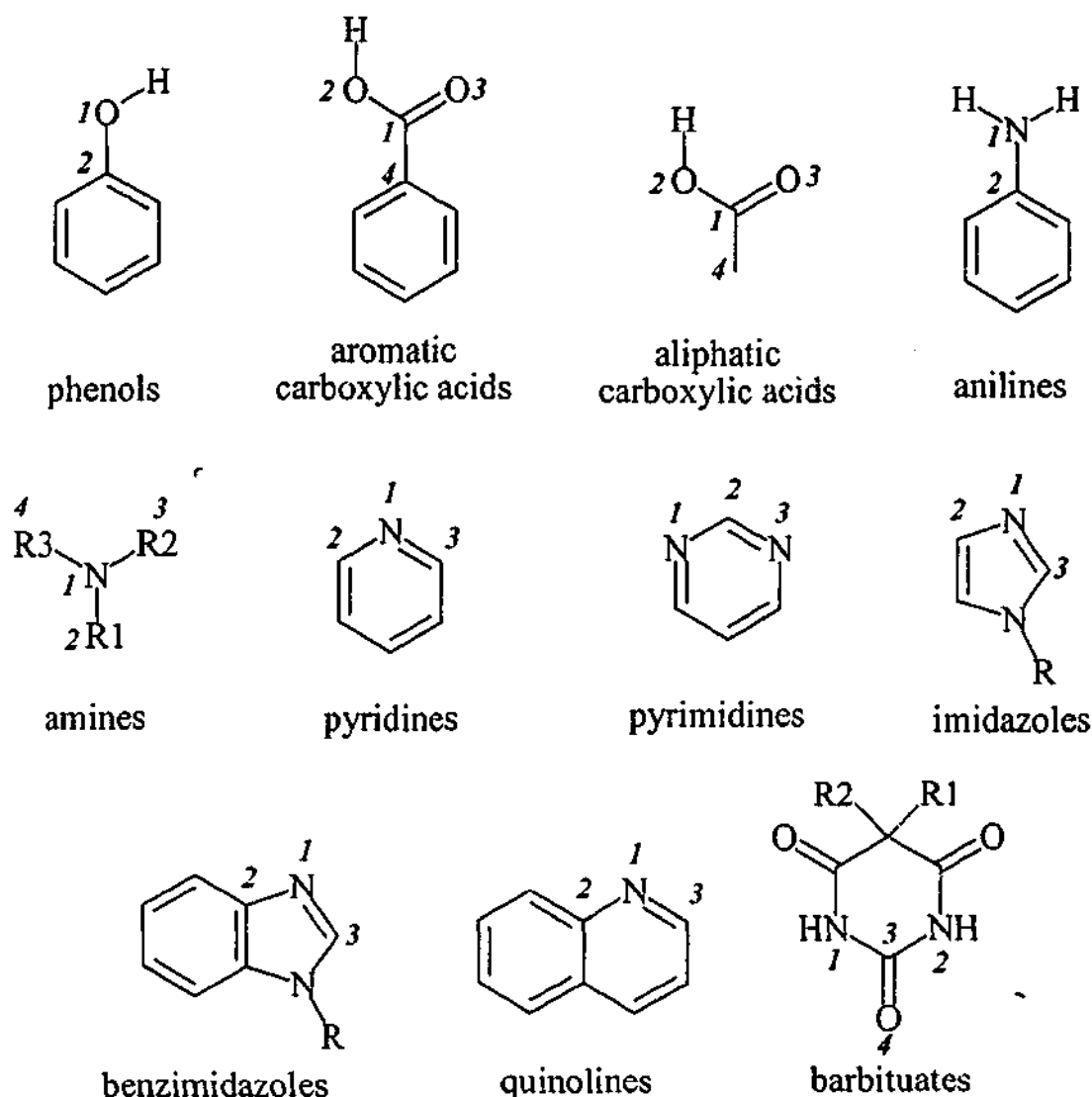
Figure 4.4. Numbering of atoms contained on the groups of interest.

4.2.4 Method application

The six main datasets of over 700 compounds were analysed in the following manner. Correlation matrices between experimentally determined pK_a values and all the QM parameters were initially calculated and viewed within Tsar. Using this information we were able to determine the most highly correlated descriptors. Scatter plots of experimental pK_a values versus the most highly correlated properties were then analysed and regression equations were derived for these single descriptors. Finally by applying m ultiple l inear r egression (MLR) o n a ll o f t he d escriptors f urther e quations were derived containing up to four descriptors.

MLR, as implemented in Tsar, was used to derive the equations. The default parameters were used in most cases although both forward and backward stepwise

regression analyses were also examined by adjusting the 'F to enter' and 'F to leave' parameters. In some cases this approach was adopted to include a term involving electrophilic superdelocalisability which consistently appeared to be correlated with pK_a. Moreover we were able to ensure that no more than four terms were included in our regression models. Our objective was to obtain statistically significant models, while keeping the number of parameters as low as possible. The cross-validated result (r_{cv}^2) refers to the leave-one-out protocol.

Additionally any compounds whose residual was greater than 2 pK_a log units following the derivation of the regression equations were further scrutinised. Typically this occurred in only one to two compounds in a limited number of equations. Where it was felt justified to leave these compounds out of the analyses, further regression equations were developed.

4.3 RESULTS

Histogram of pKa temperatures
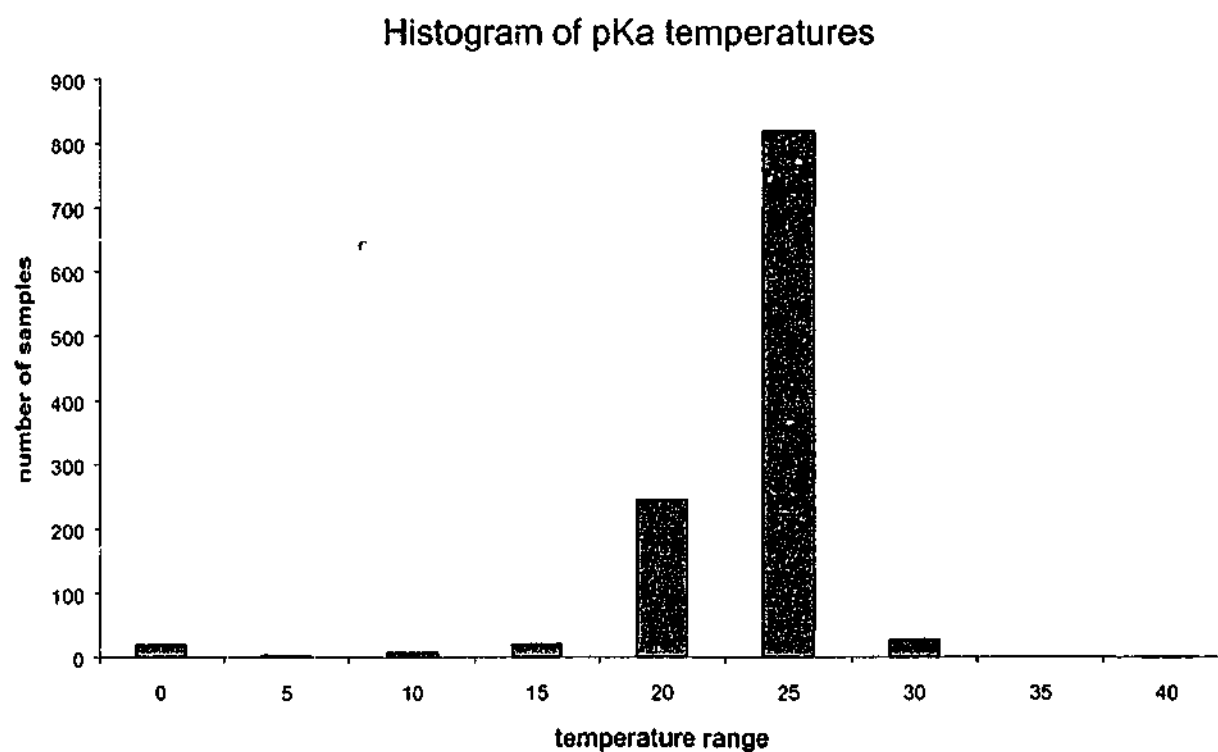
Figure 4.5. Histogram of the temperatures at which the pK_a's were determined. -

On examination, the majority of the compounds studied had the pK_a determined at 20°C to 25°C (see Figure 4.5). Moreover, pK_a does not appear to vary considerably over this temperature range of 10°C – 30°C (e.g. 6-aminopurine 0.2 units[12], leucine 0.05 units[12], piperazine 0.31 units[12]).

Table 1. Ambiguities between pK_a values from the Syracuse[10] database and literature.

Name	CAS Code	Syracuse pK_a	pK_1	pK_2	pK_3
Fenoprofen	031879-05-7	7.3	4.5[20]		
Asparagine	000070-47-3	8.82	3.22[20]	8.22[20]	
Tryptophan	000073-22-3	7.38	-	9.28[20]	16.82[20]
3-Chloro-4-Hydroxybenzoic Acid	003964-58-7	7.52	-	7.52[20]	
2-Aminobenzoic acid	000118-92-3	2.14	2.11[20]	4.95[20]	
3-Aminobenzoic acid	000099-05-8	3.07	3.12[20]	4.74[20]	
4-Aminobenzoic acid	000150-13-0	2.38	2.41[20]	4.85[20]	
4-Aminosalicylic acid	000065-49-6	2.05	2.05[20]	3.66[20]	13.74[20]
2-Aminophenol	000095-55-6	4.84	4.84[12]	10.45[20]	
3-Aminophenol	000591-27-5	4.37	4.37[12]	9.86[20]	
4-Aminophenol	000123-30-8	5.48	5.48[12]	9.75[20]	
4-nitro-aminophenol	000099-57-0	3.1	3.12[12]	7.62[12]	
Nicotine	000054-11-5	3.1	3.55[12]	8.13[12]	
Anabasine	000494-52-0	11	3.21[12]	11[12]	
N,N-Dimethyl-3-Pyridylmethylamine	002055-21-2	8	4.31[12]	8.86[12]	
Nicotinic Acid	000059-67-6	4.75	2.07[12]	4.73[12]	
Isonicotinic Acid	000055-22-1	4.9	1.7[12]	4.89[12]	
Picolinic Acid	000098-98-6	5.39	1.06[12]	5.37[12]	
3-Hydroxypyridine	000109-00-2	8.72	4.81[12]	8.74[12]	
Papaverine	000058-74-2	8.07	6.41[12]	8.01[12]	

- indicates the value was not found

After the changes to the pK_a values for the compounds listed in Table 1, the analyses were carried out and the most significant findings are reported as follows.

4.3.1 Phenols

Following the removal of compounds that were unsuitable for our analyses, the final phenol dataset consisted of 175 compounds. The range of substituents on the phenols included; aldehydes, alkyl chains, amides, amines, carboxylic acids, halides, esters, ethers, nitrates, oximes, sulphonamides and sulphonates, with a pK_a range of 12 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. A number of calculated properties showed a good correlation to pK_a.

4.3.1.1 All phenols

- model with one descriptor SE_1

(SE_1 is the electrophilic superdelocalisability for atom 1, where 1 refers to the atom number in Figure 4.4 corresponding to the appropriate analysis)

$$pK_a = -7.46*SE_1 - 61.65 \tag{1}$$

$n=175, r=0.90, r^2=0.81, r_{cv}^2=0.81, F=736.12, s=1.003$

From this result we were indeed encouraged to see the good correlation between the electrophilic superdelocalisability on atom 1 (SE_1) with pK_a using a single term equation (1).

- model with four descriptors

$$pK_a = -45.20*AQ_1 + 9.77*FE_1 + 3.39*ALP_2 - 6.34*SE_2 + 27.21 \tag{2}$$

$n=175, r=0.97, r^2=0.93, r_{cv}^2=0.93, F=594.74, s=0.599$

There was an increase in both the squared correlation coefficient (r^2) and the cross-validated squared correlation coefficient (r_{cv}^2) when going from one to four descriptors for the entire dataset of phenols, c.f. equations (1) and (2). It was interesting to note that SE_1 was replaced by SE_2, but this was entirely feasible as the two terms were within 0.01 units of each other in respect to their correlations with pK_a. The plot of pK_a observed versus predicted, according to equation (2) is reported in Figure 4.6. The following two compounds had residuals greater than 2 pK_a log units: 3,5-dichloro-2'-methyl-4'-nitrosalicylanilide (090426-03-2) and 2-amino-4,6-nitrophenol (000096-91-3)

It is interesting to speculate on why these compounds were not predicted as well as the others and it may be that our semi-qualitative method is unable to model complex

phenols with multiple substituents. After removing these compounds and reanalysing the dataset, the increase in r^2 and r_{cv}^2 was only 0.02 units.

pKa (observed) vs pKa (predicted) using four descriptors

$R^2 = 0.93$
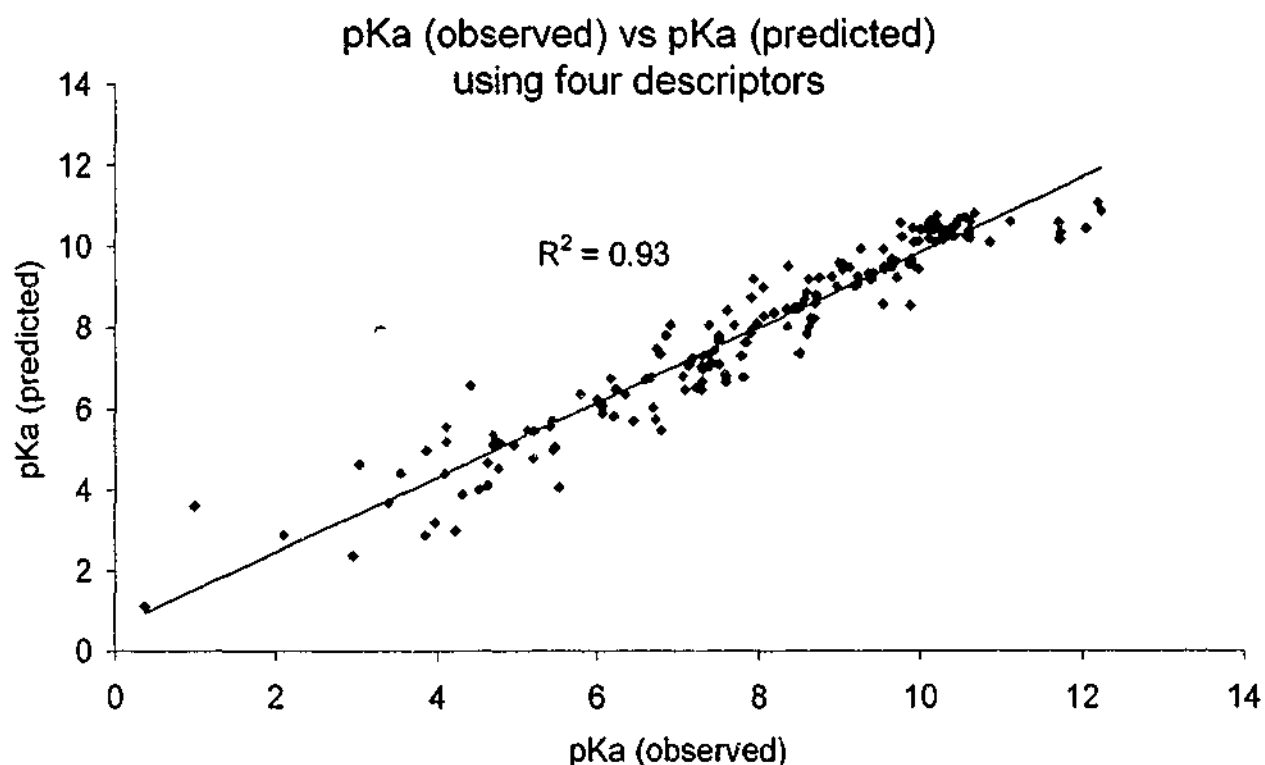
Figure 4.6. Predicted against observed pK_a for all phenols using equation (2) with four descriptors.

As previous work[6] using quantum chemical descriptors to explain pK_a's had shown that improvements could be made in dividing the data into smaller subsets, we adopted a similar approach. The phenol dataset was divided into a meta/para substituted dataset of 58 compounds and an ortho substituted dataset of 118 compounds. The ortho subset was further split using the method outlined previously, into three smaller classes: intramolecularly hydrogen bonded compounds, sterically constrained compounds and the remaining ortho compounds giving 26, 34 and 58 compounds in the three classes, respectively.

The correlation between each electronic property and the experimental pK_a values was calculated for each single class. When considering the whole data set, a number of calculated properties had shown a good correlation to pK_a and on separately analysing the subsets improved statistics were found in most cases.

4.3.1.2 Phenol subsets

meta/para substituted phenols

- model with one descriptor SE_1

$$pK_a = -5.21*SE_1 - 39.87 \tag{3}$$

$n=58$, $r=0.96$, $r^2=0.92$, $r_{cv}^2=0.91$, $F=644.74$, $s=0.315$

The prediction of pK_a's for meta/para substituted phenols did not improve by including additional descriptors. There is an obvious improvement over the one term equation for all phenols, with the r^2 going from 0.80 to 0.92, showing the benefits of placing the data into subsets.

ortho substituted phenols

- intramolecularly hydrogen bonded

  - model with one descriptor SE_1

$$pK_a = -11.33*SE_1 - 99.31 \tag{4}$$

$n=26$, $r=0.91$, $r^2=0.83$, $r_{cv}^2=0.81$, $F=119.92$, $s=0.688$

A three term equation was also generated model with an improved r^2. However the predictive ability as judged by the r_{cv}^2 did not improve (result not shown). Removing one compound from the dataset, methyl salicylate (000119-36-8), did give a significant improvement.

- model with one descriptor SE_1

$$pK_a = -10.37*SE_1 - 90.48 \tag{5}$$

$n=25$, $r=0.93$, $r^2=0.87$, $r_{cv}^2=0.86$, $F=158.80$, $s=0.537$

The inconsistencies with prediction of methyl salicylate may result from the fact it is able to form hydrogen bonds with either oxygen on the ester substituent, while this is not the case with any other compound. These results for the hydrogen bonded subset make significant improvements over the dataset containing all the phenols, especially when looking at equation (1) with only one descriptor.

sterically constrained

- model with one descriptor SE_1

$$pK_a = -8.38 * SE_1 - 70.30 \tag{6}$$

n=34, r=0.96, r^2=0.91, r_{cv}^2=0.90, F=343.50, s=0.965

The improvement in the one term equation, (6), in comparison to that for all phenols, equation (1), is easily evident. There is a slight additional improvement in introducing an extra term into equation (6) (results not shown), although this is at the expense of simplicity of interpretation of the model.

remaining ortho

- model with one descriptor SE_1

$$pK_a = -6.48 * SE_1 - 52.25 \tag{7}$$

n=58, r=0.95, r^2=0.91, r_{cv}^2=0.90, F=551.43, s=0.575

Only minor improvements are seen when adding two extra terms to equation (7) (results not shown). The results from equation (7) are better than those obtained for all phenols with one descriptor, equation (1), and comparable to those obtained using four descriptors, equation (2).

4.3.2 Benzoic Acids

The benzoic acid dataset consisted of 99 compounds. The range of substituents on the benzoic acids included; alcohols, aldehydes, alkyl chains, amides, amines, halides, esters, ethers, nitrates, sulphonamides and sulphonates, with a pK_a range of 4.5 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. Again electrophilic superdelocalisability on atoms 1, 2 and 3 showed a good correlation to pK_a.

4.3.2.1 All benzoic acids

- model with one descriptor SE_1

$$pK_a = -9.60*SE_1 - 41.74 \tag{8}$$

n=99, r=0.81, r^2=0.66, r_{cv}^2=0.64, F=185.08, s=0.567

- model with four descriptors

$$pK_a = -0.86*LUMO - 7.73*SE_1 + 3.04*ALP_4 - 2.33*SE_4 + 16.35 \tag{9}$$

n=99, r=0.93, r^2=0.87, r_{cv}^2=0.85, F=153.99, s=0.357

It was interesting to observe that electrophilic superdelocalisability was the single descriptor with the highest correlation to experimental pK_a values. However, while there was a good correlation with the oxygen atom (atom 2), see Figure 4.4, the best correlation was with the carbonyl carbon (atom 1). This may be due to the electronic effects of substituents on the ring affecting the carbonyl carbon to a greater extent than the oxygen (atom 2). Interestingly the order of correlation with pK_a was $SE_1 > SE_3 > SE_2$ with values going from $0.81 > 0.78 > 0.72$, respectively.

There was a substantial improvement in the regression equation when adding three further descriptors to give equation (9).

4.3.2.2 Benzoic acid subsets

The aromatic carboxylic acids were divided into a meta/para subset and an ortho subset giving 46 and 53 compounds, respectively. The ortho subset was further divided into the intramolecularly hydrogen bonded group and the remaining ortho compounds, giving 31 and 22 compounds per group, respectively. Only two compounds (2,6-dinitrobenzoic acid and 2,6-ditertbutylbenzoic acid) were classified as sterically constrained and these were studied within the 'remaining ortho' compounds subset.

meta/para substituted compounds

- model with one descriptor SE_3

$$pK_a = -1.98*SE_3 - 16.69 \tag{10}$$

n=46, r=0.86, r^2=0.75, r_{cv}^2=0.72, F=131.61, s=0.228

without 3,4-diaminobenzoic acid (000619-05-6)

$$pK_a = -2.14*SE_3 - 18.27 \hspace{3cm} (11)$$

$n=45$, $r=0.93$, $r^2=0.86$, $r_{cv}^2=0.85$, $F=274.94$, $s=0.167$

Equation (11) was obtained by omitting 3,4-diaminobenzoic acid (000619-05-6). This result is difficult to explain and we may speculate that problems in predicting this compound may be due to factors which can not be modelled accurately using this approach such as intermolecular hydrogen bonding. There was a large improvement when the 3,4-diaminobenzoic acid was removed from the regression equation, see equations (10) and (11). The descriptor with the greatest correlation was found to be self-atom polarisability on atom 3 (ALP_3) followed by SE_4, SE_3, SE_2 and SE_1 all with correlations greater than $r^2 = 0.82$. Improvements were not seen in r_{cv}^2 of the regression equation with multiple descriptors for the meta/para subset.

ortho substituted compounds

- intramolecularly hydrogen bonded
  - model with one descriptor SE_1

$$pK_a = - 9.52*SE_1 - 41.51 \hspace{3cm} (12)$$

$n=31$, $r=0.90$, $r^2=0.80$, $r_{cv}^2=0.76$, $F=118.19$, $s=0.447$

Inclusion of additional descriptors did not improve the predictive capabilities of equation (12) (results not shown). Removal of the only compound able to form hydrogen bonds with both 2,6-ortho substituents, 2,6-dihydroxybenzoicacid (000303-07-1), improved the predictive abilities of the model, see equation (13)

- model with one descriptor SE_1

$$pK_a = - 9.04*SE_1 - 39.20 \hspace{3cm} (13)$$

$n=30$, $r=0.91$, $r^2=0.83$, $r_{cv}^2=0.82$, $F=140.02$, $s=0.383$

There is a significant improvement in the statistics of the single descriptor equation (13) reduced set of hydrogen bonded benzoic acids compared to the single descriptor equation (8) complete set. Also equation (13) is comparable to equation (9) and much easier to interpret.

- remaining ortho substituted compounds
  - model with one descriptor SE_1

$$pK_a = -8.76*SE_1 - 38.43 \qquad (14)$$

n=22, r=0.87, r^2=0.76, r_{cv}^2=0.74, F=63.72, s=0.429

  - model with three descriptors

$$pK_a = -6.74*SE_1 - 5.34*AQ_4 + 3.13*ALP_4 + 38.20 \qquad (15)$$

n=22, r=0.95, r^2=0.90, r_{cv}^2=0.81, F=53.78, s=0.293

There is a reasonable improvement in both r^2 and r_{cv}^2 when going from equation (14) to (15). The inclusion of the charge (AQ_4) and atom self-polarisability (ALP_4) on atom 4 improves the predictive ability of the model and gives results that are comparable to those for all benzoic acids, see equation (9), although both models are somewhat more difficult to interpret than single term superdelocalisability equations.

4.3.3 Aliphatic Carboxylic Acids

The aliphatic carboxylic acid dataset consisted of 186 compounds. The range of substituents on the aliphatic carboxylic acids included; alcohols, aldehydes, alkyl chains, amides, amines, aromatic ring systems, carboxylic acids, halides, esters, ethers, nitrates, sulphonamides, sulphinates, sulphonates and others, with a pK_a range of just over 5 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. No correlation was found between any of the descriptors and pK_a for all of the atoms investigated.

4.3.3.1 All aliphatic carboxylic acids

The correlation matrix showed there were only poor correlations between pK_a and any of the descriptors present. There were tenuous correlations with both SE_2 and SE_3 with a value just over 0.5.

- model with four descriptors

$$pK_a = 4.29*ALP_1 - 41.77*AQ_2 - 30.03*AQ_3 + 0.71*FE_3 + 56.06 \qquad (16)$$

$n=185, r=0.83, r^2=0.69, r_{cv}^2=0.67, F=101.28, s=0.564$

Equation (16) was the best that could be obtained between pK_a and the descriptors, indicating that the set may be too diverse for analysis on its own and may benefit from further subdivision.

Visual analysis of the graph of predicted against observed pK_a from equation (16), see Figure 4.7, shows the poor predictive ability of this model. This is reiterated by comparing the differing slopes of the lines of best fit for the different subsets, discussed below.
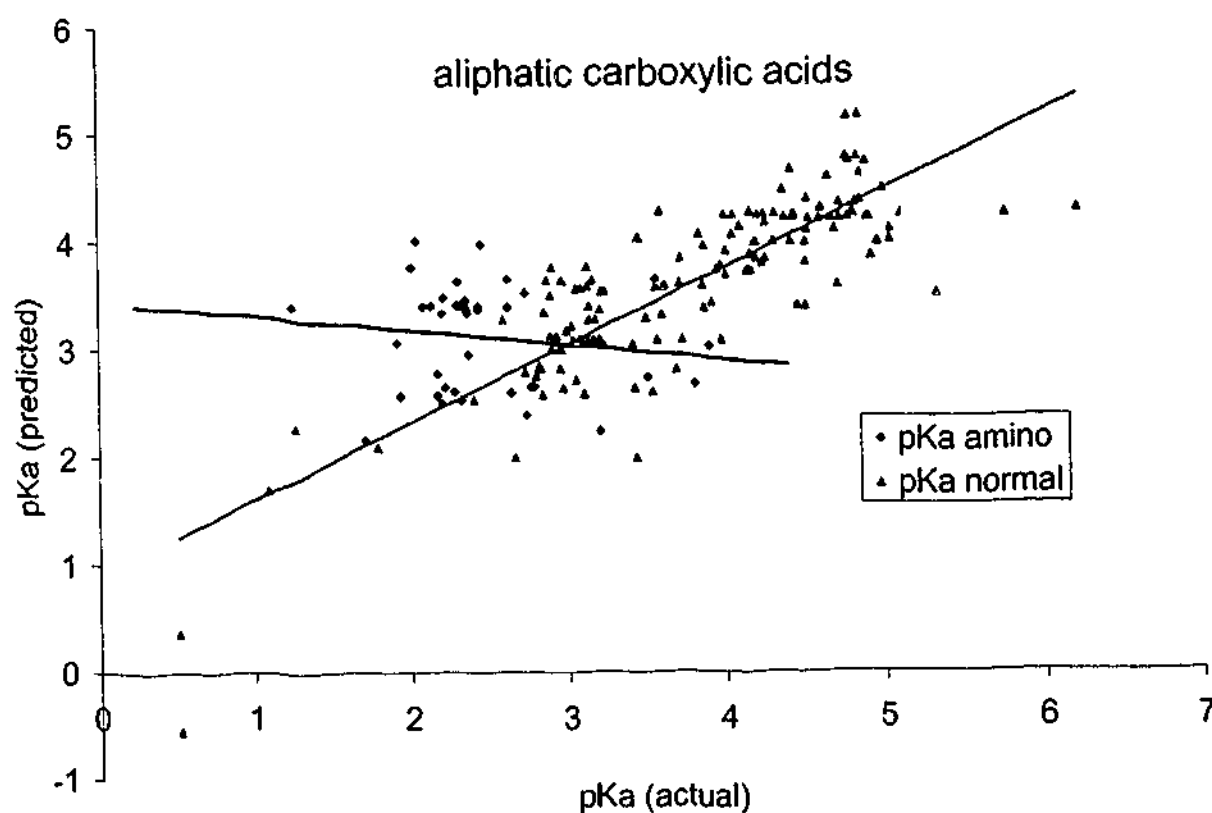
Figure 4.7. pK_a actual versus predicted from equation (16) for all aliphatic carboxylic acids with the different subsets coloured accordingly.

4.3.3.2 Aliphatic carboxylic acid subsets

The aliphatic acids were split into two smaller subsets in order to gain more information. The 185 aliphatic acids were split into a subset of 42 amino acids and the remaining 143 carboxylic acids. The amino acids were separated to mimic the hydrogen bonded sets of the previously analysed aromatic compounds. However no simple way could be found to classify highly branched or flexible compounds that could mimic the sterically constrained subset.

remaining aliphatic compounds

The correlation matrix again showed there were only poor correlations between pK_a and any of the descriptors present, therefore a two term equation was developed.

- model with two descriptors

$$pK_a = 3.24*ALP_3 - 2.80 *SE_3 - 19.43 \tag{17}$$
$$n=143, r=0.84, r^2=0.70, r_{cv}^2=0.69, F=165.80, s=0.510$$

With the removal of two significant outliers, MCPB (000094-81-5) and trinexapac (104273-73-6), we were able to improve the statistics somewhat, see equation (18).

- model with two descriptors

$$pK_a = 3.53*ALP_3 - 2.65*SE_3 + 25.01 \tag{18}$$
$$n=141, r=0.90, r^2=0.80, r_{cv}^2=0.80, F=282.85, s=0.394$$

It was interesting to note that the outlier (MCPB) has a pK_a value 0.5 log units higher than any of the other compounds in the dataset. The unusually high pK_a of 6.2 may be compounded by the highly flexible nature of this compound causing it to be modelled incorrectly. Trinexapac also has a high pK_a and other heteroatoms, such as hydroxy and multiple carbonyl groups, contained within its structure which may cause intermolecular hydrogen bonding to occur that cannot be accounted for in this method.

amino acids

The correlation matrix showed there were no reasonable correlations between pK_a and any of the descriptors present, the highest correlation being $r = 0.5$ for SE_3.

While we tried to find correlations between pK_a and all the descriptors, the best equation, (20), was unacceptably poor.

- model with three descriptors

$$pK_a = -28.75*AQ_3 + 12.17*SE_3 - 2.87*AQ_3 + 2.46 \tag{20}$$

$n=42$, $r=0.68$, $r^2=0.46$, $r_{cv}^2=0.34$, $F=10.24$, $s=0.411$

This indicates that these descriptors alone do not appear to be able to model the pK_a of amino acids. This may be due to the amphoteric nature of these compounds.

4.3.4 Anilines

The aniline dataset comprised 55 compounds. The range of substituents on the anilines included; alcohols, aldehydes, alkyl chains, aromatic ring systems, carboxylic acids, halides, esters, ethers, nitrates and sulphonates, covering a pK_a range of almost 11 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. Electrophilic superdelocalisability on atoms 1 and 2 showed the best correlation to pK_a.

4.3.4.1 All anilines

- model with one descriptor SE_1

$$pK_a = -7.83*SE_1 - 66.64 \tag{21}$$

$n=55$, $r=0.88$, $r^2=0.77$, $r_{cv}^2=0.75$, $F=178.88$, $s=0.958$

Again the addition of one more descriptor to equation (21), AQ_2, only improved the predictive aspects (r_{cv}^2) of the model by 0.04 units (results not shown).

4.3.4.2 Aniline subsets

The anilines were divided into a meta/para subset and an ortho subset, which gave 26 and 30 compounds, respectively.

meta/para substituted anilines

- model with one descriptor

$$pK_a = -3.04*SE_1 - 50.43 \qquad (22)$$

$n=26$, $r=0.85$, $r^2=0.72$, $r_{cv}^2=0.70$, $F=62.46$, $s=0.641$

- model with two descriptors

$$pK_a = 1.99*LUMO - 9.76*AQ_2 + 4.13 \qquad (23)$$

$n=26$, $r=0.91$, $r^2=0.83$, $r_{cv}^2=0.75$, $F=54.93$, $s=0.517$

The changing of descriptors to LUMO and AQ_2 improved the regression equation statistics for the meta/para anilines, see equations (22) and (23). Additionally when the compound 2-nitro-p-toluidine (000089-62-3) was removed from equation (23) the statistics improved even further.

- model with two descriptors

$$pK_a = 1.74*LUMO - 13.08*AQ_2 + 4.47 \qquad (24)$$

$n=25$, $r=0.94$, $r^2=0.88$, $r_{cv}^2=0.85$, $F=81.08$, $s=0.401$

Subdivision into the meta/para set did therefore improve the statistics of the model produced when comparing this model to the equation for all aniline compounds, (compare equations (24) to (21)). However, this was at the expense of ease of interpretation.

ortho substituted anilines

The ortho subset was not further divided into the intramolecularly hydrogen bonded and sterically constrained subsets, as there were only eight cases of these all together. Suitable and statistically relevant models could not be derived and thus the ortho compounds were studied as a single group.

The three hydrogen bonded cases contained within ortho subset were ethyl-anthranilate (000134-20-3), o-aminobenzoic acid ethyl ester (000087-25-2) and 2-aminobenzoic acid (000118-92-3).

The five sterically constrained cases contained within ortho subset were 2,6-dimethyl-4-nitroaniline (016947-63-0), 2,6-dichloro-4-nitroaniline (000099-30-9), 2,6-dichloroaniline (000608-31-1), 2,3,5,6-tetrametyl-4-nitroaniline (013171-61-4) and 2,4,6-trichloroaniline (000636-30-6).

- model with one descriptor

$$pK_a = -7.67*SE_1 - 65.55 \tag{25}$$

$n=30$, $r=0.88$, $r^2=0.78$, $r_{cv}^2=0.75$, $F=100.90$, $s=1.060$

- model with two descriptors

$$pK_a = -6.38*AQ_2 - 9.98*SE_1 - 53.09 \tag{26}$$

$n=30$, $r=0.91$, $r^2=0.83$, $r_{cv}^2=0.80$, $F=68.06$, $s=0.942$

An improvement in the statistics of regression equation (25) compared to (26) was seen when two descriptors were used to model the pK_a. Including another descriptor, charge on atom 2 (AQ_2), improves both the squared correlation and the predictive ability. Additionally, when the outlier 3-nitro-4-toluidine (000119-32-4) was removed from the analyses a noticeable improvement was seen, see equation (27).

- model with two descriptors

$$pK_a = -6.22*AQ_2 - 12.35*SE_1 - 51.59 \tag{27}$$

$n=29$, $r=0.95$, $r^2=0.90$, $r_{cv}^2=0.90$, $F=118.632$, $s=0.733$

The predictive capability of equation (27) exceeds that for the entire anilines set (equation (21)). This is even when the eight intramolecularly hydrogen bonded and sterically constrained compounds are included in the ortho analysis.

It is interesting to note that the charge residing on the carbon joined to the anilinic nitrogen plays an important role in all equations.

4.3.5 Amines

The amine dataset comprised 77 compounds. The range of substituents on the amines included; alkyl chains, aromatic ring systems, aliphatic ring systems, morpholine ring systems, halides, esters and ethers, with a pK_a range of almost 6 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. Electrophilic superdelocalisability on atom 1 showed the best correlation to pK_a.

4.3.5.1 All amines

- model with one descriptor SE_1

$$pK_a = -8.01*SE_1 - 65.78 \qquad (28)$$

$n=77$, $r=0.85$, $r^2=0.72$, $r_{cv}^2=0.72$, $F=193.80$, $s=0.650$

The predictive ability of the regression equations increased when the two compounds, azetidine (000503-29-7) and pyrrolidine (000123-75-1), were removed from the original equation (28) resulting in equation (29). These compounds may not be well predicted due to the strain on the bond angles of the four and five membered ring systems.

- model with one descriptor SE_1

$$pK_a = -7.93*SE_1 - 65.07 \qquad (29)$$

$n=75$, $r=0.88$, $r^2=0.77$, $r_{cv}^2=0.76$, $F=252.28$, $s=0.567$

4.3.5.2 Amine subsets

The 77 amines were divided into primary, secondary and tertiary amine subsets possessing 23, 23 and 31 compounds, respectively. Attempts were not made to divide these subsets any further, as this would have resulted in subsets too small to analyse.

primary amines

- model with one descriptor SE_1

$$pK_a = -8.05*SE_1 - 66.56 \qquad (30)$$

$n=23$, $r=0.93$, $r^2=0.86$, $r_{cv}^2=0.85$, $F=136.84$, $s=0.404$

The addition of two further descriptors in equation (30) improves the predictive ability of the regression model by a modest amount (results not shown). The regression model obtained for the primary amines, equation (30), was noticeably better than the final model obtained for all amines, equation (29).

secondary amines

- model with one descriptor FN_1

$$pK_a = 5.77*FN_1 + 9.48 \tag{31}$$

$n=23$, $r=0.76$, $r^2=0.58$, $r_{cv}^2=0.55$, $F=28.76$, $s=0.593$

- model with three descriptors

$$pK_a = -4.94*HOMO - 6.86*SE_1 + 0.76*SN_1 - 89.14 \tag{32}$$

$n=23$, $r=0.88$, $r^2=0.77$, $r_{cv}^2=0.67$, $F=21.12$, $s=0.461$

The regression model obtained using three descriptors, equation (32), has a higher r^2 and r_{cv}^2 than that of equation (31). An inspection of the plot of predicted against observed pK_a values using equation (31) demonstrates that the compounds fell into two distinct regions with the line of best fit passing between each region. In this case FN_1 was suspected of having a non-normal distribution and has resulted in a model of poor predictive ability. The non-normal distribution was made up of two distinct classes, purely aliphatic compounds and those containing at least one or more double bonds. These two classes were separated and analysed again however no correlation could be found with any of the descriptors. Equation (32) however, has a normal distribution of values over the 2 pK_a log units. Equation (32) was also improved substantially by removal of the outlier morpholine (000110-91-8).

- model with three descriptors

$$pK_a = -5.19*HOMO - 5.23*SE_1 + 0.82*SN_1 - 74.96 \tag{33}$$

$n=22$, $r=0.91$, $r^2=0.84$, $r_{cv}^2=0.75$, $F=30.37$, $s=0.353$

It is interesting that morpholine is an outlier, as in the original analysis containing all amines, morpholine and its derivatives are well predicted. Thus it seems the outlier status of morpholine may be an artefact of the descriptors used in equation (33) compared to (29), or all morpholines may need to be put into a further subset.

tertiary amines

- model with one descriptor SE_1

$$pK_a = -9.72*SE_1 - 81.71 \tag{34}$$

n=31, r=0.89, r^2=0.79, r_{cv}^2=0.77, F=112.04, s=0.559

Only a slight improvement was seen with the addition of descriptors, FN_1 and ALP_1, to the original equation (34) (results not shown), for the tertiary amines, although removal of fenpropimorph (067564-91-4) gave a larger improvement, see equation (35), with a much less complicated interpretation. Fenpropimorph may not be well predicted due to the eight rotational bonds and morpholine subunit, all of which may be involved in additional intermolecular interactions that are unaccounted for in this method.

- model with one descriptor SE_1

$$pK_a = -9.32*SE_1 - 77.91 \tag{35}$$

n=30, r=0.92, r^2=0.84, r_{cv}^2=0.82, F=150.08, s=0.462
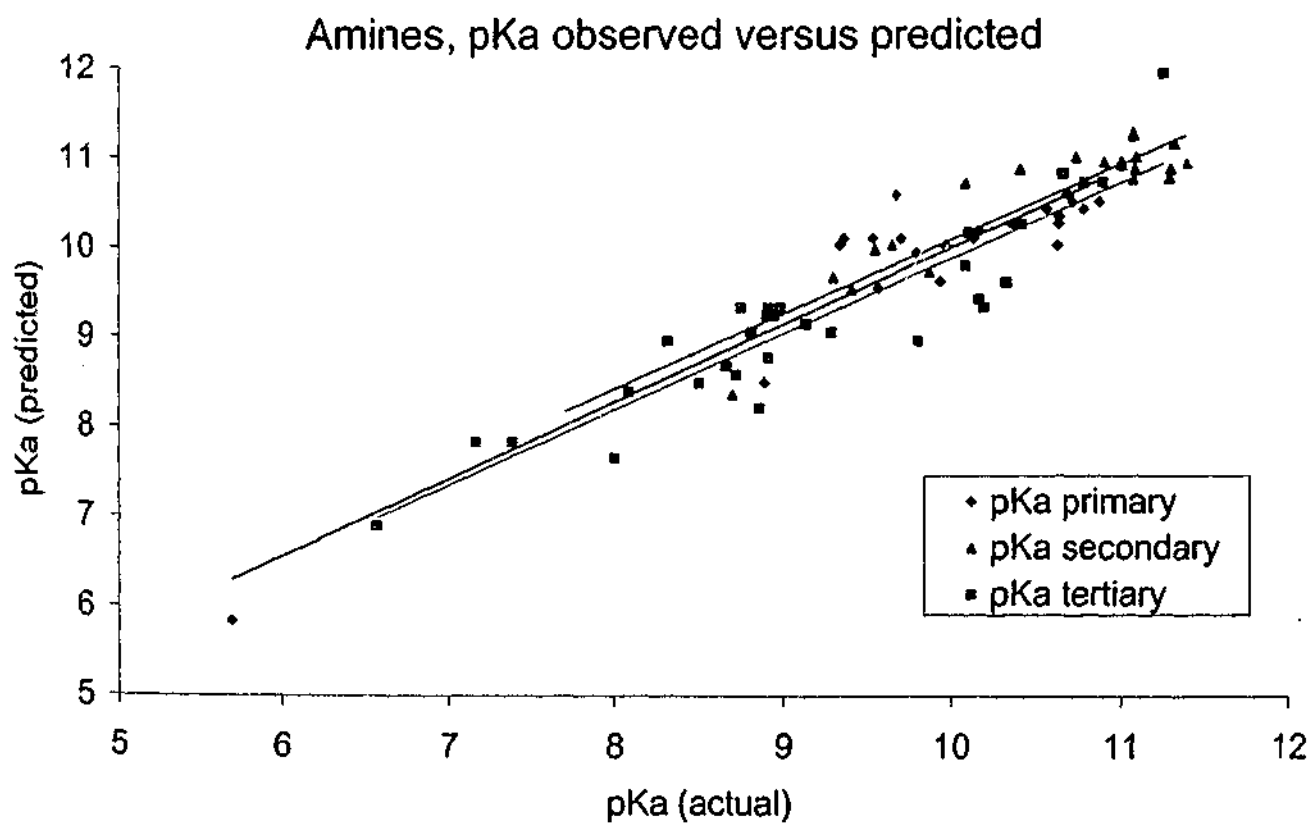
Figure 4.8 pK_a *observed versus predicted for primary secondary and tertiary amines using equations (30), (33) and (35), respectively.*

4.3.6 Heterocyclic compounds

The heterocyclic dataset consisted of 150 compounds. Some of the substituents on the heterocycles included; alcohols, aldehydes, alkyl chains, amides, amines, aromatic ring systems, aliphatic ring systems, carboxylic acids, cyanos, esters, ethers, halides, ketones, nitrates and thioethers, with a pK_a range of over 11 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. Electrophilic superdelocalisability on the nitrogen atom, (atom 1) of the four heterocycles, showed the best correlation to pK_a.

4.3.6.1 All heterocycles

- model with one descriptor SE_1

$$pK_a = -6.61*SE_1 - 54.35 \tag{36}$$

$n=150$, $r=0.74$, $r^2=0.55$, $r_{cv}^2=0.54$, $F=182.51$, $s=1.460$

- model with three descriptors

$$pK_a = 32.58*AQ_1 + 3.61*ALP_1 - 8.59*SE_1 + 5.17 \tag{37}$$

$n=150$, $r=0.85$, $r^2=0.72$, $r_{cv}^2=0.69$, $F=123.41$, $s=1.168$

There were improvements in using additional descriptors to predict pK_a for the heterocyclic compounds; however there was a large spread of predicted values, reiterated in the high standard deviation shown in both equations (36) and (37). The high standard deviation of prediction may result from trying to simultaneously analyse four heterocyclic systems with subtle differences in their electronic nature (pyridines, pyrimidines, imidazoles and benzimidazoles).

The 150 heterocyclic compounds were then separated into five main groups, according to Figure 4.3, giving 82 pyridines, 14 pyrimidines, 16 imidazoles, 10 benzimidazoles and 28 quinolines.

4.3.6.2 All pyridines

The 82 pyridines contained the same range of substituents as the complete heterocyclic set, and covered a pK_a range of 11 log units.

- model with one descriptor SE_1

$$pK_a = -8.02*SE_1 - 66.52 \qquad (38)$$

n=82, r=0.76, r^2=0.57, r_{cv}^2=0.55, F=107.65, s=1.456

- model with three descriptors

$$pK_a = 52.71*AQ_1 + 6.45*ALP_1 - 10.26*SE_1 + 49.84 \qquad (39)$$

n=82, r=0.88, r^2=0.78, r_{cv}^2=0.77, F=92.62, s=1.053

Equation (39) shows an improvement over equation (38) with the inclusion of three descriptors and there is a noticeable improvement over the analysis of all the heterocycles, equation (37). Worringly, the standard deviation of equation (39) is still over 1 log unit, which is largely due to the three outliers detected in the analyses.

The three outliers: 2,6-dichloropyridine (002402-78-0), nicotine (000054-11-5) and anabasine (000494-52-0) were removed and further regression models were constructed. Outliers 2 and 3 both contain a meta substituted aliphatic heterocyclic ring system, and are the only two compounds of this type in the data set, indicating that these compounds are not predicted well.

- model with four descriptors

$$pK_a = -1.84*LUMO + 57.93*AQ_1 + 6.76*ALP_1 - 13.88*SE_1 + 24.70 \qquad (40)$$

n=79, r=0.92, r^2=0.84, r_{cv}^2=0.82, F=100.48, s=0.851

After removal of the three outliers the predictive ability of the regression equation improves substantially and an extra term, LUMO, was included in the descriptor selection process. All other statistics parameters from equation (40) have been improved compared to equation (39). However it is interesting to note that in the new regression equation the compound with the greatest residual is 2,3,5,6-dichloropyridine (002402-79-1), which would indicate that both regression equations, (39) and (40), have

problems in predicting 2,6-dichloro substituted pyridines. This may be due to steric considerations with the bulky chlorine groups hindering the pyridyl nitrogen.

4.3.6.3 Pyridine subsets

The pyridines were further subdivided into a meta/para subset and an ortho subset, giving 48 and 34 compounds, respectively.

meta/para substitution

- model with one descriptor SE_1

$$pK_a = - 7.83 \cdot SE_1 - 64.317 \tag{41}$$

$n=48$, $r=0.86$, $r^2=0.74$, $r_{cv}^2=0.72$, $F=128.84$, $s=0.790$

No improvement was seen with the inclusion of additional descriptors. However the same two meta/para substituted outliers that were present in equation (39) are also present in this analysis.

The two outliers: nicotine (000054-11-5) and anabasine (000494-52-0) were removed.

- model with one descriptor SE_1

$$pK_a = - 8.63 \cdot SE_1 - 71.20 \tag{42}$$

$n=46$, $r=0.93$, $r^2=0.86$, $r_{cv}^2=0.85$, $F=279.39$, $s=0.581$

The removal of the two outliers from equation (41) provided a significant improvement to give equation (42). Additionally, there is an improvement over equation (40) which models the complete pyridine set using four descriptors thus, demonstrating the benefits of subdividing this set.

ortho substitution

- model with one descriptor SE_1

$$pK_a = - 9.19 \cdot SE_1 - 77.51 \tag{43}$$

$n=34$. $r=0.78$, $r^2=0.61$, $r_{cv}^2=0.56$, $F=50.49$, $s=1.849$

- model with two descriptors

$$pK_a = 55.60*AQ_1 - 14.69*SE_1 - 118.32 \tag{44}$$

$n=34$, $r=0.93$, $r^2=0.87$, $r_{cv}^2=0.82$, $F=101.71$, $s=1.097$

A marked improvement was seen over equation (43), with the addition of the descriptor (AQ_1) to give equation (44). There were three outliers present here, two of which were mentioned earlier in relation to equations (39) and (40). The three outliers: 2,6-dichloropyridine (002402-78-0), 2,3,4,5,6-pentachloropyridine (002176-62-7) and 2,6-ditertbutylpyridine (000585-48-8) were removed.

- model with three descriptors

$$pK_a = 56.27* AQ_1 - 14.81*SE_1 - 119.95 \tag{45}$$

$n=31$, $r=0.96$, $r^2=0.93$, $r_{cv}^2=0.86$, $F=115.26$, $s=0.747$

After removal of the 2,6-dichloro and 2,6-ditert butyl compounds an excellent correlation was achieved, demonstrating that steric hindrance appears to affect the pK_a of these pyridine compounds. Unfortunately, there were not enough sterically hindered compounds to construct a statistically significant model.

4.3.6.4 Pyrimidines

There were 14 pyrimidine compounds present in the heterocyclic dataset. Substituents on the pyrimidines included alkyls, amines, aromatic ring systems, ethers, esters, halides and nitrates, spanning a pK_a range of 9 log units. The initial correlation matrix showed that there was a greater correlation with one nitrogen (atom number 3) than the other (atom number 1) see Figure 4.4. This was investigated further as both nitrogens were expected to be equivalent. Analysis of the AM1 minimised structures revealed that the greatest correlation was indeed with the nitrogen that was not sterically hindered by substituents attached to atom 2, see Figure 4.9. As only a number of compounds had substituents of this nature, these were removed and when the correlation matrix was recalculated it showed that both nitrogens were almost identical in an electronic sense. This demonstrated the importance of molecular conformation in pursuing these studies.

2-ethoxypyrimidine   pyrimethanil
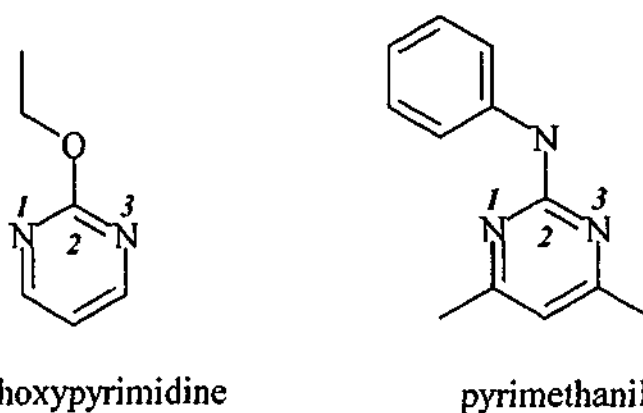
Figure 4.9. Examples of substituents from atom 2 that were causing steric hindrance of nitrogen atom 1, 2-ethoxypyrimidine and pyrimethanil.

The highest correlation from the correlation matrix was shown to be the charge on atom 3, AQ_3, followed closely by the electrophilic superdelocalisability, SE_3.

- model with one descriptor AQ_3

$$pK_a = -39.79*AQ_3 - 5.14 \tag{46}$$

$n=14$, $r=0.88$, $r^2=0.78$, $r_{cv}^2=0.66$, $F=42.09$, $s=1.100$

No improvement in r_{cv}^2 was seen with additional descriptors, although after the removal of the outlier, fenclorim (003740-92-9), a marked improvement was seen, see equation (47).

- model with one descriptor AQ_3

$$pK_a = -41.11*AQ_3 - 5.59 \tag{47}$$

$n=13$, $r=0.94$, $r^2=0.88$, $r_{cv}^2=0.84$, $F=83.17$, $s=0.805$

The discrepancy in fenclorim's prediction may be due to the phenyl substituent attached to atom 2 sterically interfering with both nitrogens. Indeed, steric and conformational considerations have already been shown to affect pK_a in both the pyridine and pyrimidine datasets.

4.3.6.5 Imidazoles

The imidazole dataset comprised 16 compounds. Some of the substituents on the imidazoles included: alcohols, aldehydes, alkyl chains, amines, aromatic ring systems, aliphatic ring systems, carboxylic acids, esters, ethers, halides, ketones, thioethers and nitrates, with a pK_a range of almost 9 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for the entire dataset. Electrophilic superdelocalisability on the nitrogen atom, atom 1 (see Figure 4.4), showed the best correlation to pK_a.

- model with one descriptor SE_1

$$pK_a = -8.21*SE_1 - 69.05 \tag{48}$$

$n=16$, $r=0.95$, $r^2=0.91$, $r_{cv}^2=0.90$, $F=155.49$, $s=0.864$

- model with two descriptors

$$pK_a = 1.04* AQ_1 - 6.72*SE_1 - 55.55 \tag{49}$$

$n=16$, $r=0.98$, $r^2=0.95$, $r_{cv}^2=0.94$, $F=137.36$, $s=0.664$

There was a slight improvement on what was already a good correlation when comparing equations (48) and (49). The introduction of the descriptor, AQ_1, reduced the standard deviation by almost a third and the F statistic marginally. This would enable us to make predictions with much greater confidence even though the dataset was of limited size.

4.3.6.6 Benzimidazoles

The benzimidazole dataset was of a limited size with only 10 compounds satisfying the selection criteria. The substituents present in the dataset include; alkyl chains, amines, amides, aromatic ring systems, ethers, halides and nitrates, with a pK_a range of just over 3 log units.

The correlation between each electronic property and the experimental pK_a values was calculated for all the benzimidazoles. Electrophilic superdelocalisability on the nitrogen

atom, atom 1 (see Figure 4.4), showed the best correlation to pK_a, followed closely by energy of the highest occupied molecular orbital, HOMO.

- model with one descriptor SE_1

$pK_a = -4.33*SE_1 - 33.33$

$n=10$, $r=0.77$, $r^2=0.59$, $r_{cv}^2=0.55$, $F=13.12$, $s=0.748$ (50)

- model with three descriptors

$pK_a = 16.59*FN_1 - 4.82*SE_1 + 0.02*SN_1 - 38.13$ (51)

$n=10$, $r=0.95$, $r^2=0.91$, $r_{cv}^2=0.83$, $F=23.41$, $s=0.400$

There is improvement in the predictive ability of equation (51) when compared to equation (50). The F statistic, while statistically significant, is considerably smaller than in our other analyses but remains adequate for the number of compounds being analysed. Of course, for this analysis 9 descriptors were considered for these 10 compounds and it must be bourn in mind that the relationship may have been derived by chance[21].

To try and account for the fact that our benzimidazole equation may have been derived by chance, a single term equation was derived using both the imidazoles and benzimidazoles.

- model with one descriptor SE_1

$pK_a = -7.79*SE_1 - 65.09$

$n=26$, $r=0.93$, $r^2=0.87$, $r_{cv}^2=0.85$, $F=161.04$, $s=0.805$ (52)

The similarities of this single term equation (52) to that derived for the imidazoles alone, equation (48), are immediately evident however there is a significant difference between equation (52) and equation (50) derived from the benzimidazoles alone. The difference between equations (52) and (50) may arise from the fact that only a range of 3 log units of pK_a are available for the benzimidazoles. Indeed it would be of great interest to see what similarities there may be if a greater range of compound pK_a's were available.
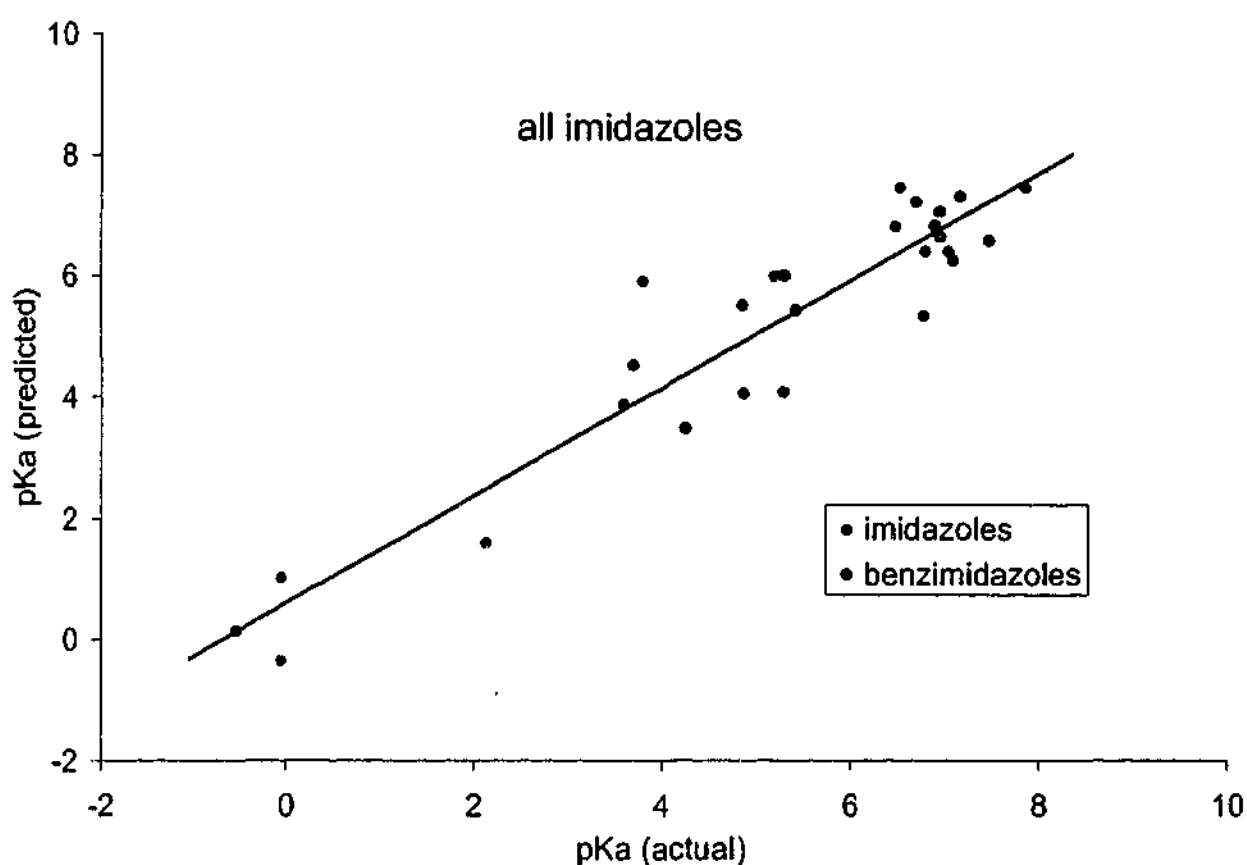
Figure 4.10. Plot of actual versus predicted for both benzimidazoles and imidazoles

4.3.6.7 Quinolines

The quinoline dataset consisted of 28 compounds with substituents ranging from; alkyl chains, alcohols, ethers, halides to nitrates, with a pK_a range of just over 3 log units. The correlation between each electronic property and the experimental pK_a values was calculated for all the quinolines. LUMO on the nitrogen atom, atom 1 (see Figure 4.4), showed the best correlation to pK_a, followed by electrophilic superdelocalisability.

- model with one descriptor LUMO

$pK_a = 5.71*LUMO - 7.86$

$n=28$, $r=0.89$, $r^2=0.80$, $r_{cv}^2=0.77$, $F=104.81$, $s=0.432$ (53)

No improvement in the predictive capabilities of equation (53) was seen with the addition of further descriptors (results not shown).

4.3.6.8 Barbiturates

An analysis was also attempted on a series of 14 di-substituted barbiturates, unfortunately no significant correlation was found with any of the descriptors present. There was a general trend with the HOMO energy and the pK_a for the series, shown in equation (54)

- model with one descriptor HOMO

$$pK_a = -0.29*HOMO - 4.70 \tag{54}$$

n=14, r=0.74, r^2=0.55, r_{cv}^2=0.34, F=14.95, s=0.164

The poor result of this analysis is probably due to the fact that the series of compounds only covers a pK_a range of 1 log unit. No improvement was found in regression equation (54) when additional descriptors were used in this analysis.

4.4 DISCUSSION

Current trends are such that it is becoming more important to determine the pK_a and other characteristics associated with a drug, before synthetic work is undertaken with the aim of avoiding making compounds that are predicted to have poor pharmacokinetics. Being able to do this accurately and in a reasonable amount of time is also of extreme importance so that these filters can be applied to lead compounds, virtual libraries and database collections to give an indication as to a compounds qualities.

Predicting the pK_a of a compound is a very important characteristic, as it gives an indication as to a compound's membrane and/or barrier permeability. A fast and reasonably accurate way of doing this is using semi-empirical quantum mechanical methods. This has been done before by a number of groups[6,7,22], although only carboxylic acids, alcohols and phenols have been studied. We have studied these groups and have extended our work to investigate anilines, amines and other heterocyclic compounds, with a novel set of semi-empirically derived *frontier electron theory* descriptors.

4.4.1 Grüber and Buss

The first group to use semi-empirical methods for pK_a determination were Grüber and Buss[6]. The MNDO method was used as it was the most widespread method of the time; AM1 was released not long after they finished all their MNDO calculations. In total 183 compounds were analysed by Grüber and Buss[6] consisting of 99 phenols, 52 benzoic acids and 32 aliphatic carboxylic acids. The descriptors calculated were HOMO energies of the anion and neutral species, atomic charges at all atoms and the energy difference between the anion and neutral species.

4.4.1.1 Phenols

Grüber and Buss[6] obtained good results for all phenols in a six term equation using atomic charges and HOMO energy. Their equation with an r^2 of 0.94 for 99 compounds

was comparable to our equation (2) with an r^2 of 0.92 for 180 compounds using a four term equation, see Figure 4.1. Unfortunately no r_{cv}^2 is reported in the paper of Grüber and Buss[6], so it is hard to judge the predictive ability of their model in comparison to ours; however the F statistics indicate both models are of high quality.

Results for the meta/para substituted phenols for Grüber and Buss[6] again were similar to our findings. While our r^2 was slightly lower, equation (3), we were encouraged by the results as our study has been applied to twice the number of compounds. Additionally, our result was with the use of a single descriptor, electrophilic superdelocalisability of the phenolic oxygen.

The results of the three ortho subsets (equations (5), (6) and (7)) are comparable to those of Grüber and Buss[6] and in most cases much easier to interpret as only one descriptor is needed, electrophilic superdelocalisability. Grüber and Buss[6] however did not split their ortho substituted phenols into smaller subsets thus making it difficult to compare the results.

4.4.1.2 Benzoic Acids

Reasonable results were obtained by Grüber and Buss[6] for a series of substituted benzoic acids and these were improved with the removal of four outliers. This gave a 5 term equation, consisting of atomic charges, resulting in an $r^2 = 0.88$ for the 48 compounds. Our results compared favourably to those obtained by Grüber and Buss[6] as we were able to examine 99 molecules under consideration giving us a dataset twice as large, see equation (9).

The meta/para subset was predicted well by Grüber and Buss[6] with a 4 term equation, consisting of HOMO energy and atomic charges, with an $r^2 = 0.93$ for 25 compounds. The results of our study here are not as good as those of Grüber and Buss[6] for this subset, although removal of the 3,4-diaminobenzoic acid gave a reasonable one term equation, (11), with atom self-polarisability as the descriptor.

The removal of two outliers from the ortho set of Grüber and Buss[6] gave a 4 term equation, consisting of atomic charges and anion/neutral state energy differences, with an $r^2 = 0.92$ for 25 compounds. Our results were comparable, see equation (13), although again it is difficult to compare as Grüber and Buss[6] have not split their compounds into further subsets.

Again as no r_{cv}^2 is reported in the paper of Grüber and Buss[6], it is hard to judge the predictive ability of their model in comparison to ours. However the F statistics indicate both models are of high quality, although in most cases our F statistic is significantly higher due to the much larger number of compounds being analysed.

4.4.1.3 Aliphatic Carboxylic Acids

Grüber and Buss[6] managed to get reasonable results for the aliphatic acids, with a three term equation which was improved with the removal of two outliers to give a four term equation with an $r^2 = 0.86$. This was a good result as there are a lot of problems associated with the flexibility of these compounds. Our normal subset of the aliphatic carboxylic acids, see equation (18), can be directly compared with those of Grüber and Buss[6], as their group of aliphatic carboxylic acids also did not contain amino acids or α-hydroxy carboxylic acids. Our findings do compare favourable with those of Grüber and Buss[6] even though our r^2 value is not as high. We have analysed over three times as many compounds which also provides a significant model that is very simple to interpret as it has a single descriptor, electrophilic superdelocalisability.

Electrophilic superdelocalisability features prominently in all equations, and is usually the descriptor with the highest correlation to pK_a. This is not that surprising, as SE is proportional to the sum of the squared coefficients of the occupied orbitals and thus this parameter is proportional to the electron density over the atom. In other words, the lower the electron density on the atom bonded to the hydrogen, the more easily the hydrogen will be abstracted and thus the more acidic the compound. This is reiterated in the negative slope for all equations of SE versus pK_a and can be seen in Figure 4.11 for the subsets of the phenols.
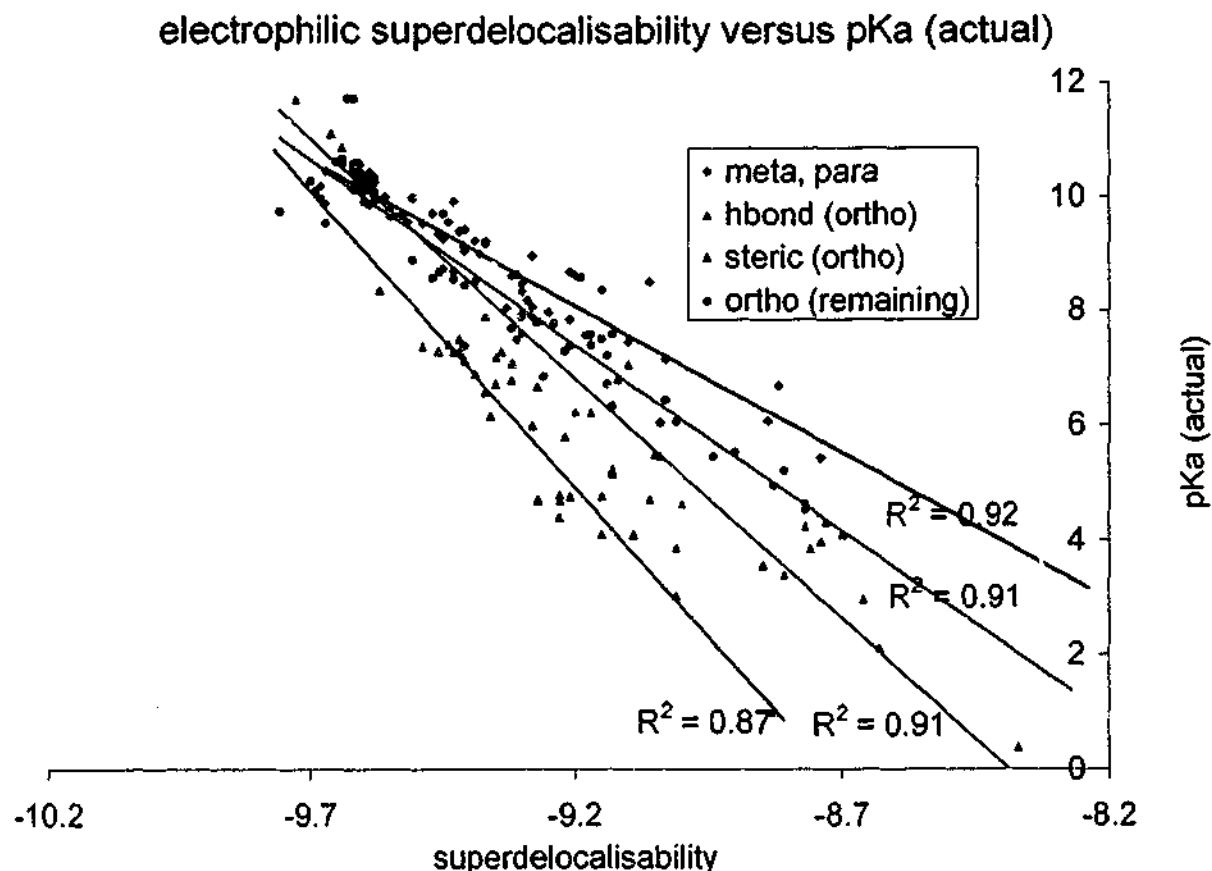
Figure 4.11. *Electrophilic superdelocalisability of the phenolic oxygen atom versus measured pKa, using equations (3), (5), (6) and (7).*

4.4.2 Citra

The second study exploiting semi-empirical quantum mechanical descriptors to predict pK_a was conducted by Citra[7] from the Syracuse research corporation. Rather than focus on the differences in energy of the anion and neutral acid, his work focused on descriptors associated with the neutral acid. The descriptors chosen by Citra[7] were bond order of the O-H bond as well as partial atomic charges on the oxygen and hydrogen atoms. Adjunct to this he also performed multiple geometry optimisations for which the above descriptors were calculated for each geometry. The values of the descriptors used in the analysis are the weighted average of properties computed for all optimised conformations.

4.4.2.1 Phenols

For the 100 phenols analysed from the Syracuse database a three term equation with an r^2 of 0.96 was found. This was using the charge on the hydrogen and oxygen and their

For the 100 phenols analysed from the Syracuse database a three term equation with an r^2 of 0.96 was found. This was using the charge on the hydrogen and oxygen and their bond order. Although the three terms are probably unnecessary here as one term the Coulson charge on the oxygen is shown to have an $r^2 = 0.95$.

We attempted to reproduce (in part) the work done here by Citra[7] on our group of 180 phenols, extracted from the Syracuse database, although without doing any conformational analysis. Unfortunately we were not able to achieve similar results to Citra[7]: for our 180 phenols we obtained an r^2 of 0.66 for Coulson charges on the oxygen alone and r^2 of 0.88 using Coulson charges and bond orders. This did not compare as well as our single and multiple term equations, see equations (1) and (2), for the same dataset; however no conformational analysis had been performed. We thought that some of the larger phenol derivatives, such as salicylanilides and thiophenesulfonamides, could be causing additional problems here, so we decided only to look at the compounds that both our dataset and those of Citra[7] had in common. This resulted in a dataset of 66 compounds that were also analysed in the same manner as had been done by Citra[7], although with no conformational analyses. Although the r^2 for Coulson charges did improve from 0.66 to 0.82, see Figure 4.12, and the result for two descriptors improved from 0.88 to 0.90, this was still less than the r^2 of 0.92, see Figure 4.11, obtained by just electrophilic superdelocalisability on the oxygen alone.

pKa observed versus predicted
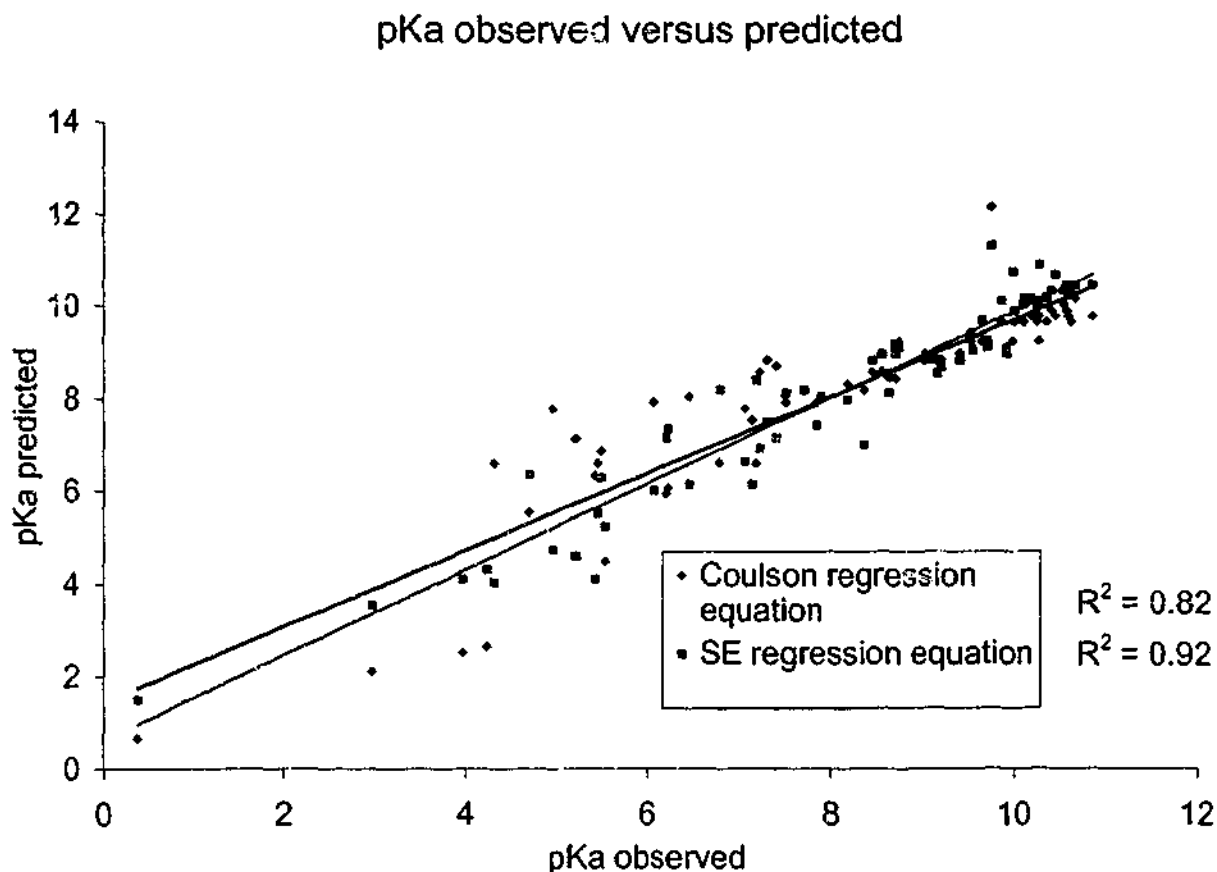
Figure 4.12.   Analysis of the dataset that we had in common with Citra[7] using the single best descriptor from each analyses, electrophilic superdelocalisability and Coulson charges from the oxygen.

This data indicated that doing conformational analysis on the series of compounds that you have is a very good idea, although this may come at the expense of additional time required.

4.4.2.2 Benzoic Acids

Good results were obtained by Citra[7] for the analyses of the 31 benzoic acids in their dataset, giving a 3 term equation with an $r^2 = 0.89$ using the charge on the hydrogen and oxygen and their bond order. Our results compared favourably to those obtained by Citra[7], with a four term equation giving an $r^2 = 0.86$ for three times as many compounds, see equation (9). Using descriptors with multiple conformation weighted values does not make a large improvement in this case. Unfortunately further splitting of this data was not performed, therefore it is difficult to compare further equations.

4.4.2.3 Aliphatic Carboxylic Acids

It appears again that using weighted conformational values for the descriptors does give an improvement in the regression equation, with an $r^2 = 0.84$ obtained by Citra[7] compared to our $r^2 = 0.79$, see equation (18). It is not surprising that descriptor values derived from multiple conformations make an improvement, as there are a lot of problems associated with the flexibility of these compounds.

Interestingly it was found that Coulson charge, the most significant descriptor from the analyses of Citra[7], has a very high correlation ($r>0.99$) with atom self-polarisability (ALP) for identical atoms.

The correlation between Coulson charges and atom self-polarisability (ALP) is not surprising as both descriptors are calculated by summing up contributions from atomic and molecular orbitals.

Coulson charge distribution[23] is calculated by determining the charge distribution for each orbital separately, and then summing distributions for all the occupied molecular orbitals. In a particular molecular orbital

$$\Psi = c_1 \Psi_1 + ... + c_n \Psi_n$$

the normalisation condition is

$$\sum c_r^2 = 1$$

and the familiar interpretation of ψ^2 as a density shows us that we may associate fractional charges c_1^2, c_2^2,..., c_n^2 of an electron with the nuclei 1, 2,..., n. The normalisation condition ensures that the complete electron is thus accounted for. By summing c_1^2 for each occupied orbital, we obtain the total electronic charge on nucleus 1.

4.4.3 Schüürmann

Schüürmann[22] estimated pK_a using semi-empirical quantum mechanical gas-phase and solution-phase calculations, for a series of 16 carboxylic acids and 15 phenols. A method employing Gibbs free energy for one-molar solution at 298 K was used for estimation. It was noted that semi-empirical continuum-solvation models predict pK_a in qualitative agreement with experiment; however they were not accurate enough to yield absolute values, although scaling through linear regression for specific compound classes did result in models that were highly predictive. Squared correlation coefficients greater than 0.9 were achieved for both datasets using pK_a versus ΔH^{COSMO} and ΔG^{SM2}, calculated from the continuum-solvation models AM1-COSMO and AM1-SM2, respectively. The datasets chosen are of a limited size although they show the obvious benefits of linear scaling for specific compound classes in pK_a prediction.

As the descriptors studied here vary from gas-phase to continuum-solvation to small-cluster models, it is difficult to compare all these directly to the frontier electron theory descriptors. However the gas-phase models can be directly compared to the frontier electron theory descriptors which were also calculated in a gaseous state. Our results for carboxylic acids, see equation (18), is comparable to results obtained by Schüürmann[6,7,22] with six times as many compounds being analysed. The results obtained from the phenol dataset, see equation (2), was comparable with over ten times as many compounds being analysed. Additionally our splitting of the phenol dataset into intramolecularly hydrogen bonded compounds and other subsets, showed there was a difference between the compounds in differing subsets (see Figure 4.7). This was not easily evident in this smaller analysis of Schüürmann[6,7,22] where they stated, "there is almost no difference between the conformations with and without intramolecular hydrogen bonding". This statement is most likely an artefact of the small number of compounds that were analysed by Schüürmann[6,7,22], and not the descriptors used.

The extension by Schüürmann[6,7,22] into continuum-solvation models is a good logical progression that leads to significant improvements. The results obtained here are excellent for both phenols and carboxylic acids, and it would be interesting to see how frontier electron theory descriptors calculated from the output of AMSOL (using AM1-

COSMO or AM1-SM2) compared to the heats of formation (ΔH^{COSMO}) and Gibbs free energy (ΔG^{SM2}) equations obtained already.

4.4.4 Gross

Prediction of pK_a values for nitrogen containing compounds has not been attempted using semi-empirical methods although there have been a number of excellent papers from the group of Gross et al.[24,25] using ab initio techniques to predic pK_a values of anilines. Here natural charge (Q_n), relative proton-transfer enthalpy (ΔH_{prot}), minimum electrostatic potential (V_{min}) and minimum local ionisation energy on the molecular surface ($I_{S,min}$) all correlated well with experimental pK_a values with the squared correlation coefficient being greater than 0.85 in all models. This paper[24] also showed noticeable improvements when the complete set of compounds was split into further subsets.

The results obtained by Gross et al.[24] are noticeably better than our results, especially when looking at the results using only one descriptor, see equations (21), (22) and (25). With the inclusion of additional descriptors in our models we were able to achieve results with only marginally lower squared correlation coefficients after having analysed almost twice as many compounds, see equations (24) and (27). An interesting point raised in the discussion by Gross et al.[24] suggests that for p-hydroxyaniline, a hydrogen bond between the hydroxyl hydrogen and a water molecule enhances the substituents electron donating character. As such phenomena cannot be accounted for using semi-empirical techniques it may well explain some discrepancies associated with hydroxylated compounds.

Surprisingly no ortho-substituted anilines were present in the analyses of Gross et al.[24], this may have been because they wanted to compare their results with those obtained using Hammett σ values. Therefore we attempted a similar exercise whereby we summed Hammett parameters, assuming they were additive and where they were available, for 20 meta/para substituted anilines.

The following equation was derived:

$$pK_a = -3.1\sigma + 4.3 \tag{55}$$

$$n = 20, r = 0.87, r^2 = 0.75, r^2_{cv} = 0.74, F = 55.37, s = 0.684$$

This equation was similar to that obtained by Gross *et al.*[24], shown below:

$$pK_a = -3.03(\pm0.13)^*\sigma + 4.46(\pm0.06)$$

$$n = 36, r^2 = 0.94, F = 530, s = 0.310$$

The difference in the statistics of the two equations may be due to the fact that our substituents also included a sulfone group and iodine groups.

Equation (55) derived from the Hammett parameters was directly compared to an analysis of the 20 meta/para substituted anilines using SE_1 which gave an r^2 of 0.735. This indicated to us that our technique is comparable to using Hammett parameters; however it has the added benefit of being applicable to not only meta/para substituted compounds.

4.5 CONCLUSION

This work has shown that descriptors derived from *frontier electron theory* are able to predict the pK$_a$ of a variety of compound classes. This method can be used as an alternative to LFER methods, where unknown fragments are present in the compound, or to more computationally intensive *ab initio* methods, where time is of the essence. This method also has the advantage that it is comparable to Hammett parameters with the added benefit of being applicable to ortho substituted compounds.

There are a number of ways that this method could be improved, by calculating the descriptors as an energy weighted average of properties computed for all of the optimised conformations, as was done by Citra[7]. However weighted contributions would also come at the expense of additional computational time. Another way to improve this method may be in the calculation of *frontier electron theory* descriptors initially. It would be interesting to see how frontier electron theory descriptors calculated from the output of AMSOL, using AM1-COSMO or AM1-SM2, and not traditional AM1 would compare to the results obtained here.

Chapter 5

Compound Design

5.1 COMPOUND DESIGN

Rational compound design probably first began in the late 19th century with John Langley and Paul Ehrlich. Around this time Langley first proposed that drugs interact with specific receptors and Ehrlich proposed the concept of receptors, as being the "binding group of the protoplasmic molecule to which a foreign newly introduced group binds"[1]. Although drugs had been around for a lot longer than this, these concepts marked the beginning of the end for trial and error as the only approach to drug design.

Now at the turn of the millennium, although trial and error still exists in the drug design process, intelligent drug design strategies aim to minimise detrimental outcomes. An integral part of intelligent drug design strategies are the *in silico* processes of ligand-based design, structure-based design, ADME/T (adsorption, distribution, metabolism, excretion and toxicity) prediction and database mining. Although not all aspects of the drug design process were able to be explored in the course of this work a number were utilised to design compounds that are D_4/D_2 selective with the potential to act as atypical antipsychotic drugs.

5.2 METHOD

To best design D_4-selective antipsychotic drugs, clozapine and spiperone and the pharmacophores deduced from these compounds (Chapter 2) in conjunction with the receptors built with these compounds were used. Clozapine and spiperone bind potently at the D_4 receptor, with clozapine probably being the most significant atypical antipsychotic available to date. In addition to this, there is a wealth of binding data for both these compounds at mutant D_2 and D_4 receptors, which enables more tangible conclusions to be drawn from what is a rather complicated picture. Database mining using the pharmacophores from chapter 2 aided in the generation of ideas for compound design along with a number of papers that discuss the most common chemical substituent replacements for producing drug-like compounds[2.3].

Each compound was built within Sybyl[4] using standard geometries and angles; compounds were then minimised for 10,000 iterations or until the RMS gradient fell below 0.01kcal using the MMFF94s force field[5.6] and MMFF94 charges[5.6]. The MMFF94 charges where then replaced with ESP charges on the minimised compound, as these were shown to perform best in CoMFA analyses. Compounds were then aligned to the pharmacophore using the RMS fit procedure within Sybyl. The affinities at the differing receptors for each compound were predicted via input into the CoMFA models created. The pK_a of each compound was predicted by calculating the frontier electron theory descriptors[7], using the AM1 Hamiltonian[8], for each compound which were then input into the linear regression models developed. For all tertiary amines the regression equation (35) from chapter 4 was used for calculating pK_a values.

Equation (35) $pK_a = -9.32*SE_l - 77.91$

5.3 RESULTS AND DISCUSSION

5.3.1 Database Searching

To aid in the generation of ideas for the design of new compounds, the pharmacophores generated in chapter 2 were used to search the ISIS mddr3d database. All searches of the mddr3d database were done in a fully flexible search mode, so all rotatable bonds of compounds being searched were fully explored. Although a fully flexible search mode requires additional time, this is necessary as a compound's conformer in the database are generated automatically b y Concord and these various conformers may not fit within the constraints of the proposed search.

5.3.1.1 Tricyclic Pharmacophore

Both the simple and advanced search models in Figure 5.1 revealed all the compounds used in the elucidation of the tricyclic pharmacophore. The simple model retrieved 1276 antipsychotic compounds including a number of extended antipsychotic structures such as haloperidol. As the 1276 compounds from the simple search model were too numerous to analyse in detail, the simple search model was refined using the detailed model shown in Figure 5.1.
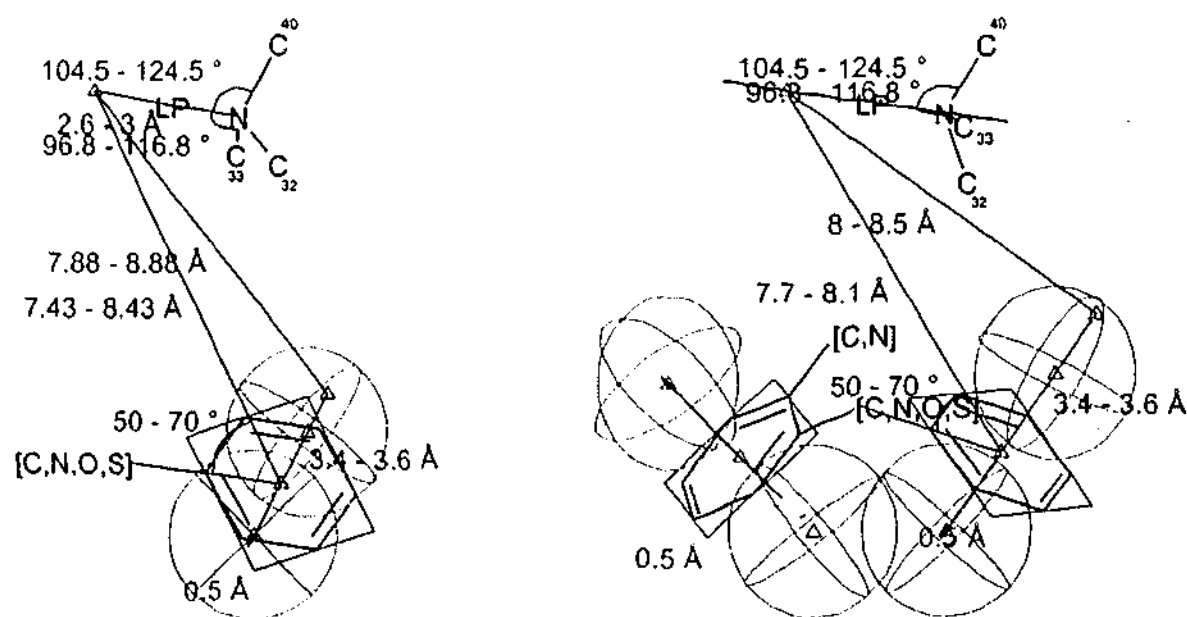
Figure 5.1. Tricyclic pharmacophore queries used in the searching of the ISIS mddr3D database. Simple query (left), refined query (right), based on Figure 2.8 (Chapter 2).

These searches revealed a number of new substructures that could be used in the construction of novel antipsychotics. In particular a number of interesting bicyclo compounds were retrieved including a series of 1,4-diazabicyclo[4.4.0]decane and 1,4-diazabicyclo[4.3.0]nonane compounds (see Figure 5.2) bound to an oxazepine tricyclic substructure.
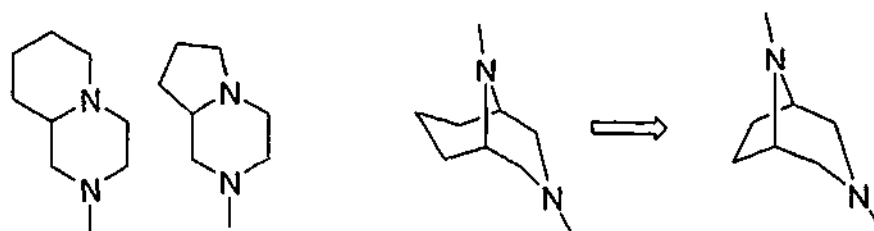
Figure 5.2. Interesting substructures from database search and possible variations

5.3.1.2 Extended Pharmacophore
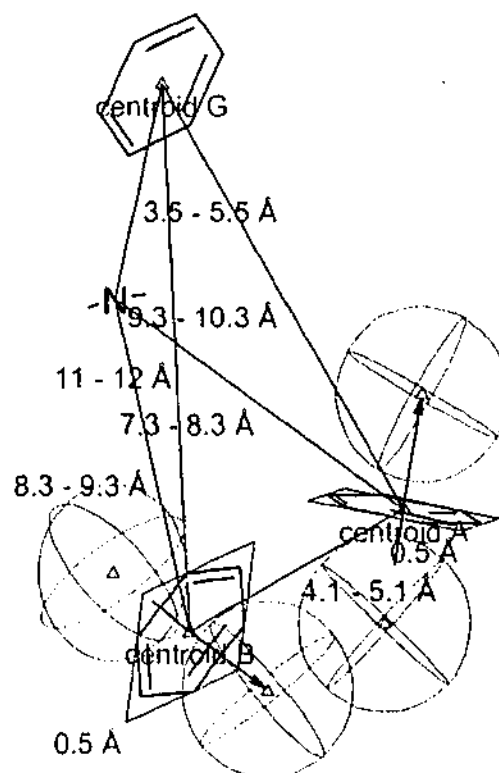
Figure 5.3. Extended pharmacophore search model used to search the ISIS mddr3d database, based on Figure 2.35 (Chapter 2).

The extended query shown in Figure 5.3 was used to mine the ISIS mddr3d database, and retrieved 606 compounds of which 54 were classified as being antipsychotics. The search revealed a number of older typical antipsychotics such as fluspirilene and pimozide as well as a number of interesting D_4-selective clozapine analogues (see

Figure 5.4). It is proposed that by modifying the typical antipsychotics so as to achieve greater D_4 selectivity we may see reduced side effects in such a series. Using the compounds in Figure 5.4 as the basis for designing new ligands that occupy the A, B and G regions of the pharmacophore, a series of compounds was proposed for synthesis.

fluspirilene

pimozide

11-[4-(1.3-Benzodioxol-5-ylmethyl)piperazin-1-yl]-8-chlorodibenz[b,f][1,4]oxazepine
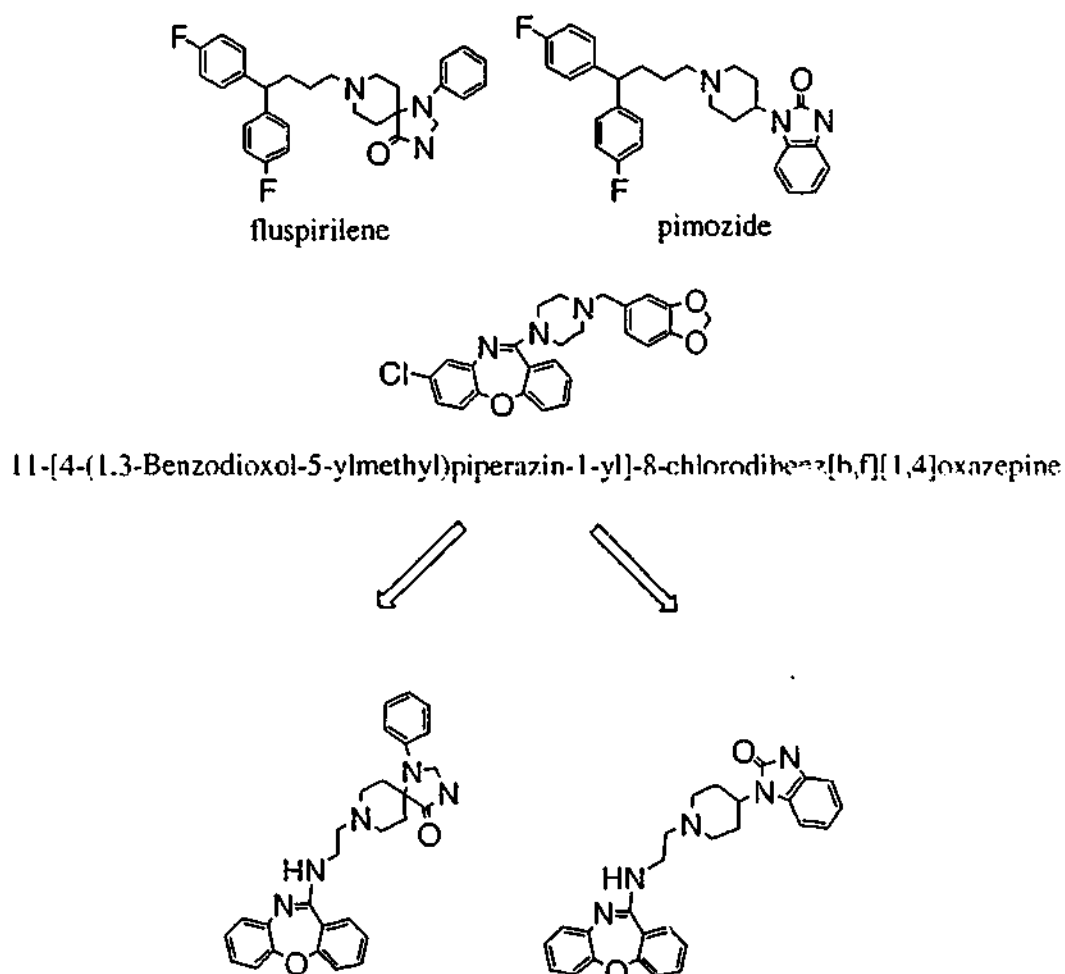
Figure 5.4. Compounds retrieved from the extended pharmacophore search and two compounds similar to the proposed ABG series.

The main concern with the extended pharmacophore search is that all the requirements of the search must be met in order for a compound to be retrieved. As most of the compounds used in the extended CoMFA analyses only occupied two of the three regions, they were not retrieved in this search. It should be noted however that many of these compounds were retrieved in the simple tricyclic pharmacophore search. Therefore additional searches were made of the ISIS database using only queries that consisted of regions A and G and regions B and G. The search using the B and G region query identified all the Merck Sharpe and Dohme compounds used earlier in chapter 2 for the pharmacophore and CoMFA modelling. Although the A and G region search did identify over 50 known antipsychotics, none of the compounds used in our

CoMFA model were identified. Upon closer inspection of the database none of the compounds used in our CoMFA model occupying regions A & G were present in the database.

5.3.2 Ligand-Based Design: CoMFA models

5.3.2.1 Tricyclic CoMFA Overview

The tricyclic CoMFA analyses from chapter 2 revealed that the addition of larger, more electronegative substituents at Y (Figure 5.5) w ould increase D_2 binding a ffinity. In turn, D_2 binding affinity could be reduced by having bulky alkyl substituents occupying position Z or by the presence of nitrogen in position A. More novel methods of altering the binding affinity could also be achieved by modifying the type of aromatic ring present as ring A, to change the electronic environment. Steric regions produced by the D_4 CoMFA model from chapter 2 were similar to those of the D_2 model; however a number of additional favourable steric regions were present in the D_4 model, which suggested slightly larger substituents may be added at position W to increase affinity. There is also scope for increasing D_4 binding affinity by adding slightly larger substituents at position Q, which as stated earlier one could exploit using bicyclo compounds.
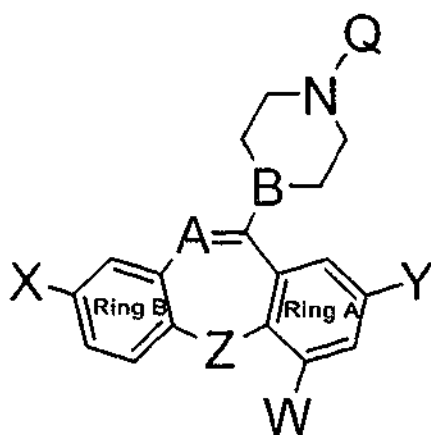
Figure 5.5. Tricyclic structure showing sites of possible variation for compound design.

5.3.2.2 Extended CoMFA Overview

The extended CoMFA analyses from chapter 2 revealed greater D_4 selectivity could be achieved by ensuring that compounds were designed with substituents that occupied position W (Figure 5.5). It was shown that substituents are beneficial in the D_4 model, region $G4_{D4}$ (Figure 2.38), and not well tolerated here in the D_2 model, region $Y1_{D2}$ (Figure 2.36). The D_2 electrostatic fields also show that either the presence or absence of an electronegative substituent at position A (Figure 5.5) can modify D_2 binding. This region where partial positive charge enhances activity is in direct contrast to what we see in the electrostatic fields generated from D_4 binding affinity data. Another worthwhile area to exploit for increasing selectivity is the ring containing the basic nitrogen. Most compounds with a piperidine ring bind better at the D_2 receptor compared to compounds with a piperazine ring and vice versa for the D_4 receptor. This could be exploited by designing compounds with a piperazine ring or substituent with similar electronic characteristics. Morpholine or thiomorpholine ring systems could be used but this may come at the expense of reducing the basicity of the distal nitrogen, which may reduce binding affinities and thus would have to be carefully monitored. Further areas that have been shown to increase binding affinity for the D_4 model which are not present to the same extent in the steric fields produced from the D_2 model are; region $G3_{D4}$ (Figure 2.38) which is above and out from position A (Figure 5.5), and region $G5_{D4}$ (Figure 2.38), corresponding to the area around ring G (Figure 5.3). In addition D_4 selectivity could be managed via the position Z (Region $Y1_{D4}$ (Figure 2.38)) where bulky alkyl groups reduce D_4 binding affinities. Additional bulk here seems to be tolerated better in the D_2 model and this could be exploited to modulate D_4 selectivity if desired.

5.3.3 Structure-Based Design: Receptor Models

5.3.3.1 Tricyclic models

The placement of clozapine differed quite significantly between the D_2 and D_4 receptor models created. The most significant D_2 model had clozapine placed in the region between helices four, five and six. This model was also consistent with the majority of

the mutational data, whereas the most likely position for the binding of clozapine within the D$_4$ receptor model, deduced from the receptor models and mutational data, is between helices two, three and seven. Surface plots of clozapine bound in the different receptors were then created (see Figures 5.6 and 5.7) to assess further differences that may be exploited in the design of D$_4$ selective atypical antipsychotics.
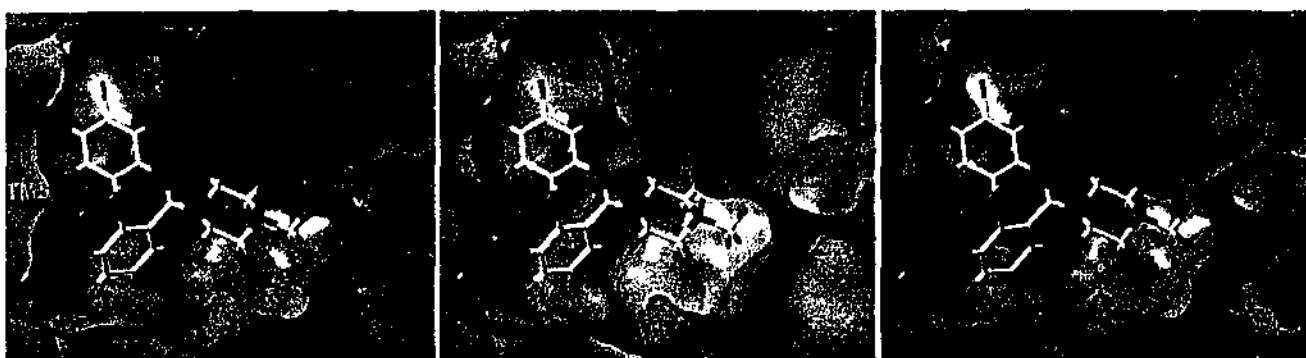
Figure 5.6. Surfaces generated from the extracellular view of the D$_2$ receptor with clozapine bound between helices four, five and six (labelled in red).
Lipophilic (green → brown increasing hydrophobicity) left,
Electrostatic (orange → blue increasing electronegativity) centre,
Hydrogen-bonding (blue acceptor, red donor) right.
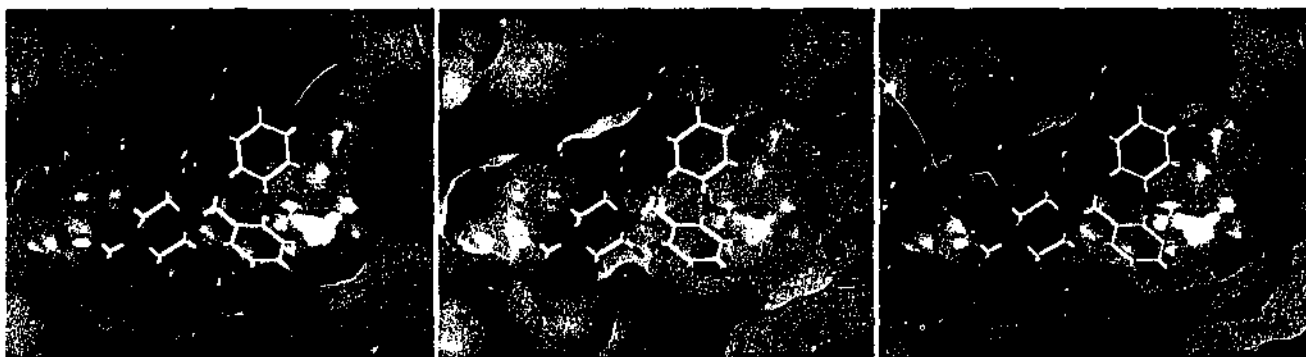
Figure 5.7. Surfaces generated from the extracellular view of the D$_4$ receptor with clozapine bound between helices two, three and seven (labelled in red).
Lipophilic (green → brown increasing hydrophobicity) left,
Electrostatic (orange → blue increasing electronegativity) centre,
Hydrogen-bonding (blue acceptor, red donor) right.

The differing surface plots of clozapine bound at the D$_2$ receptor (see Figure 5.6) show that there is little room for additional substituents at position W (Figure 5.5) whereas the plot of clozapine in the D$_4$ receptor (see Figure 5.7) shows there is room here for additional substituents. The D$_2$ and D$_4$ diagrams also show differences in

hydrophobicity at position Y and around ring A. The D_4 lipophilic surface plot indicates that different types of aromatic rings may be beneficial for D_4 selectivity. The region around the piperazine ring of clozapine also differs in regards to hydrophobicity, showing that a less hydrophobic ring would aid D_4 selectivity. In both the D_2 and D_4 diagrams we see similar features around position X indicating that selectivity would probably not be gained by varying substituents at this position. The region corresponding to position Q (Figure 5.5) however, does differ between the D_2 and D_4 diagrams, showing that there is more space in this area for the D_4 receptor. It appears as though bicyclo compounds would be preferentially tolerated here in the D_4 receptor compared to the D_2 receptor.

5.3.3.2 Extended Models

The final placement of spiperone in the D_2 and D_4 receptor models did not differ in the manner that was seen for the binding of clozapine; however differing interactions were still seen when looking at the models.

The two diagrams (Figures 5.8 and 5.9) show that the binding site of the D_2 receptor is overall more hydrophobic in nature. In addition there is more room around the *p*-fluoro end of spiperone in the D_2 model which is consistent with previous hypotheses that this area is involved in D_4/D_2 selectivity[9]. Although there is less room in the D_4 receptor between helices two, three and seven, it appears as though slightly longer compounds may still be preferentially accommodated here, due to the additional space near the fluoro substituent of spiperone. This observation corresponds well to the different amino acids located at 1.39, where the larger $Leu^{1.39}$ of the D_2 model occupies additional space here and modifies the packing of adjacent residues. One significant difference between the two electrostatic surfaces generated is the presence of a partial positive region located near the phenyl portion of ring G in the D_4 diagram. This may indicate that electronegative substituents, such as cyano groups, attached to the phenyl ring or even different aromatic rings, such as indole, may aid in D_4 selectivity. This exploitation of the partial positive charge has already been carried out with the addition of methoxy and cyano groups to phenyl ring G, (compounds #30 to #32, Table Carling (Appendix)) reducing D_2 affinity substantially whilst only slightly reducing D_4 affinity.

Figure 5.8. Surfaces generated from the extracellular view of the D_2 receptor with spiperone bound, approximate position of helices labelled in red.

Lipophilic (green → brown increasing hydrophobicity) left,

Electrostatic (orange → blue increasing electronegativity) centre,

Hydrogen-bonding (blue acceptor, red donor) right.
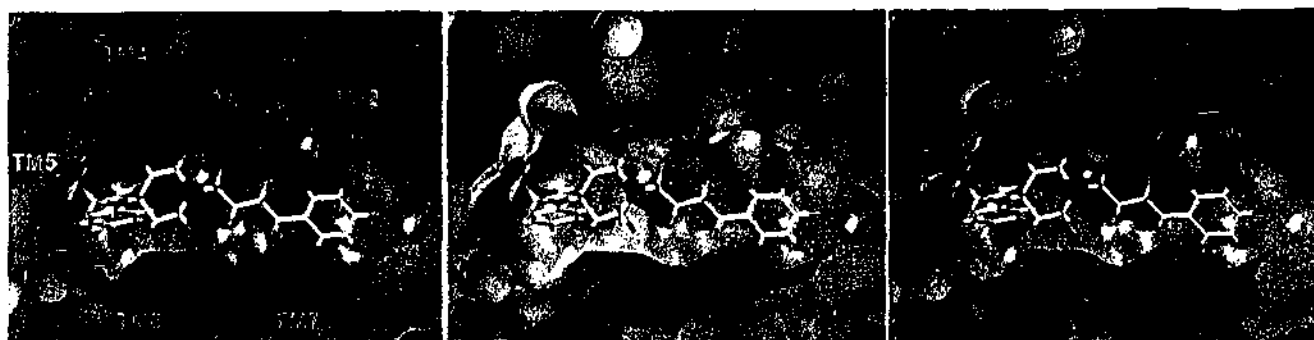
Figure 5.9. Surfaces generated from the extracellular view of the D_4 receptor with spiperone bound, approximate position of helices labelled in red.

Lipophilic (green → brown increasing hydrophobicity) left,

Electrostatic (orange → blue increasing electronegativity) centre,

Hydrogen-bonding (blue acceptor, red donor) right.

5.3.4 Proposed Compounds

Taking into account all the information from the different *in silico* methods, several series of compounds have been proposed for synthesis. The three main series of compounds were designed to occupy regions A/B, regions B/G and regions A/B/G.

5.3.4.1 A/B Series

The A/B series or tricyclic compounds were designed whilst keeping in mind the significant differences between the binding of clozapine at the D_2 and D_4 receptors and

the different substructural variations seen from the database searching. The tricyclic CoMFA analyses also gave further insight into variations that would be preferentially tolerated in D_4 binding.

Table 5.1. Compounds designed to occupy the A and B regions of the pharmacophore with their predicted D_2 and D_4 affinities (pK_i) and pK_a's.
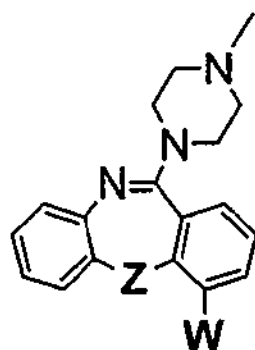
Compound	Z	W	Predicted		
			pK_i (D_2)	pK_i (D_4)	pK_a
AB_1	CH_2	methyl	5.96	7.48	8.27
AB_2	CH_2	chloro	6.58	7.85	7.87
AB_3	CH_2	fluoro	6.18	7.79	7.77
AB_4	CH_2	cyano	5.71	7.89	7.35
AB_5	NH	methyl	6.16	7.65	8.66
AB_6	NH	chloro	6.44	7.82	8.24
AB_7	NH	fluoro	6.35	7.73	8.15
AB_8	NH	cyano	5.93	7.86	7.75
AB_9	O	methyl	6.08	7.74	8.18
AB_10	O	chloro	6.43	8.16	7.78
AB_11	O	fluoro	6.17	7.93	7.69
AB_12	O	cyano	5.64	8.11	7.30

Only positions W and Z were varied in the design of new compounds as it had already been shown that electronegative atoms at positions A and B aid in D_4 selectivity, thus nitrogen was kept in positions A and B. Position X was not varied as this position had been shown to affect both D_2 and D_4 affinities in the same manner. Position Y also was not varied as substitution at position W will have the same electronic effects on ring A. The positions not varied here (A, B, X and Y), had been altered and shown to increase D_2 and D_4 affinity in a similar manner; however as D_4 selectivity was the primary goal this is not included. The series of compounds (Table 5.1) show that the bulkier electronegative substituent, cyano compared to fluoro, at position W gives high affinity

and the greatest selectivity at the D_4 receptor. However it can also be seen from looking at the series that the presence of a cyano group also reduces the pK_a of the compound the most, and therefore this must be carefully monitored. There is a general trend in regards to the prediction of pK_a values with the more electron withdrawing groups reducing the pK_a values the most. However predicted pK_a values are still in a favourable range to enable the APD to be unionised for crossing the BBB and ionised to act on the receptor. The most selective of the cyano substituted compounds is the oxazepine derivative, and this compound was used as the basis for further structural modifications.

Table 5.2. Compounds designed to occupy the A and B regions of the pharmacophore with their predicted D_2 and D_4 affinities (pK_i) and pK_a's.

compounds AB_12 #1,2   compound AB_12 #6   compound AB_12 #7

compounds AB_12 #3-5   compound AB_12 #8   compound AB_12 #9

| Compound | R | Predicted | | |
		pK_i (D_2)	pK_i (D_4)	pK_a
AB_12_1	O	5.27	8.08	7.29
AB_12_2	S	4.93	8.15	6.36
AB_12_3	NH	5.19	8.05	7.47
AB_12_4	O	5.44	8.03	6.84
AB_12_5	S	5.48	8.23	7.25
AB_12_6		5.27	8.20	8.34
AB_12_7		5.44	8.37	8.31
AB_12_8		5.05	8.19	7.17
AB_12_9		5.08	8.18	6.40

The modifications to compound AB_12 (Table 5.1) show that selectivity between the two receptors can be increased by further modifications. The morpholine and thiomorpholine compounds (AB_12 #1,2) increase selectivity marginally by reducing D_2 affinity, although the pK_a's of these compounds is reduced. The pK_a of the morpholine compound is thought to be too high as a similar compound in the regression analysis was predicted to be too high from equation (34), chapter 4, and was removed as an outlier. This reduction of pK_a of these compounds would not appear to be significant enough to effect their value as potential APDs; however the synthesis of such compounds could be difficult. The modification of ring A of compound AB_12 did increase selectivity marginally through a reduction in D_2 affinity and in the case of the thiazole compound (AB_12_5, Table 5.2) an increase in D_4 affinity. The thiazole ring system has already shown much promise with the APD olanzapine. D_4 selectivity was also increased with the incorporation of an additional ring system onto the piperazine ring, with a reduction in D_2 affinity and an increase in D_4 affinity for all compounds. In reality compounds AB_12_6 and AB_12_7 are identical but placement of the ring system in the CoMFA models was investigated by looking at the differing orientations it could take. The pK_a of compound AB_12_9 is most likely predicted slightly lower than would be expected, since pyrrolidine compounds were shown to be predicted lower than expected from equation (35); see compounds 020173-28-8 and 000120-94-5, Table 10, Appendix.

5.3.4.2 B/G Series

The B/G series were designed whilst keeping in mind the significant differences between the binding of spiperone at the D_2 and D_4 receptors. The additional length available within the D_4 receptor in conjunction with the partial positive charge near region G were both utilised to design D_4 selective compounds.

Table 5.3. Compounds designed to occupy the B and G regions of the pharmacophore
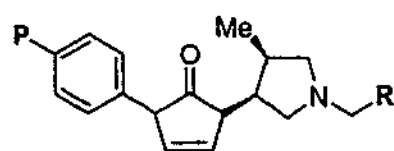
compounds #1-6
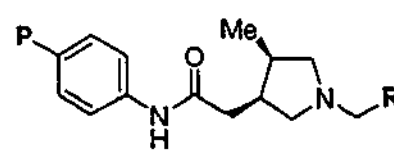
compounds #7-12
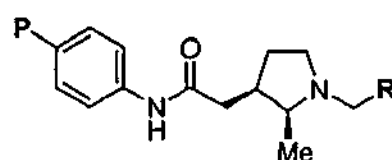
compounds #13-18

with their predicted D_2 and D_4 affinities (pK_i) and pK_a's.

| Compound | P | R | Predicted | | |
			$pK_i (D_2)$	$pK_i (D_4)$	pK_a
BG_1	H	phenyl	7.88	7.99	7.38
BG_2	H	2-cyanophenyl	7.21	8.51	5.97
BG_3	Cl	phenyl	7.75	8.15	6.99
BG_4	Cl	2-cyanophenyl	7.09	8.67	5.60
BG_5	CL	indole	6.94	8.83	7.52
BG_6	Cl	isoindole	6.95	9.03	8.38
BG_7	H	phenyl	7.50	7.56	8.20
BG_8	H	2-cyanophenyl	6.80	8.02	6.73
BG_9	Cl	phenyl	7.23	7.73	7.84
BG_10	Cl	2-cyanophenyl	6.53	8.20	6.39
BG_11	CL	indole	6.53	8.52	8.31
BG_12	Cl	isoindole	6.51	8.83	9.18
BG_13	H	phenyl	8.54	7.71	8.30
BG_14	H	2-cyanophenyl	7.92	7.82	6.83
BG_15	Cl	phenyl	8.47	7.83	7.92
BG_16	Cl	2-cyanophenyl	7.85	7.93	6.47
BG_17	CL	indole	7.90	8.86	8.38
BG_18	Cl	isoindole	7.82	9.00	9.25

The three substructures from the B/G series show that by careful consideration of the factors i nvolved i n l igand b inding, c ompounds c an b e t ailored t o a s pecific r eceptor. With the three B/G substructures we see that the additional length in the form of the p-chloro s ubstituent a nd t he i nteraction w ith t he p artial p ositive c harge by t he c yano group both increase D_4 affinity and reduce D_2 affinity. The theme of the interaction with the partial positive near ring G is explored further by the modification of the

phenyl ring to indole and *iso*indole. These modifications again increase D_4 selectivity. The *iso*indole compared to the indole compound, when visualised is best positioned for optimal interaction with the partial positive charge, and this is reiterated by its higher predicted affinity. These modifications were also able to take a D_2 selective compound (BG_13) and turn it into a D_4 selective compound (BG_18). The final compounds from these structural modifications (BG_6, BG_12 and BG_18) are all D_4 selective compounds, and in the first two instances by over two orders of magnitude. A similar theme is seen in regards to the predicted pK_a values seen here compared to the pK_a predictions from the AB series: that is that the electronegative cyano substituent significantly reduces the predicted pK_a value. The predicted pK_a value for these five-membered ring systems varies almost three log units for each substructural series. Whether or not this large pK_a range would be realised is uncertain, as no tertiary amines with *iso*indole substituents were present in the original analysis from which the regression equation was derived. In addition to this it is likely the predicted pK_a values for five membered tertiary amines may be artificially low (see compounds 020173-28-8 and 000120-94-5, Table 10, Appendix).

5.3.4.3 A/B/G series

The A/B/G series of compounds were based on an extension of the compounds proposed in Figure 5.4. The combination of the D_4 selective oxazepine with the typical antipsychotic fluspirilene should give a D_4 selective compounds with antipsychotic properties. Interestingly in the development of this series of compounds it was found that the attachment of the tricyclic substructure with an A-fold (Figure 2.3) resulted in a compound that did not fit within the constraints of the pharmacophore. The B-fold compound, however, did fit well into the model although this adds some confusion to the interpretation of the various positions and rings identified earlier. Figure 5.10 serves to clarify this problem. Only positions A', B', Z and W' were varied in the construction of the compounds. Position X was not varied, as it was not seen to be appreciably different in the analysis of the receptor models. Position Y was also seen to give a similar electronic environment around ring A when compared to position W' and thus was also not varied. As positions A and Z had been shown to give rise to selectivity, they were varied. However as position A did not exactly conform to position A', both

A' and B' were varied to judge their effect.   Only heteroatoms were substituted at position B' to reproduce the effect of a piperazine ring.

Figure 5.10.   Structural basis of compounds in the A/B/G series(right) and the corresponding positions shown on the tricyclic substructure(left).

The table of compounds (Table 5.4) exhibits compounds which, in many cases show over three orders of magnitude in selectivity towards the D_4 receptor.   The reason for this is the B-fold conformation that is adopted and the relatively different attachment of the tricyclic portion.   The differing attachment and B-fold cause the tricyclic substructure to rotate in a slight clockwise manner compared to clozapine.   This in turn causes ring A to partly occupy the region around position W.   This also is why we do not see any significant improvement in D_4 affinity when the larger chloro substituent is added at position W', as ring A already occupies this region.   When a cyano substituent was added at position W' the additional bulk of this group compared with hydrogen significantly reduced D_4 binding affinity, and for this reason this series of compounds was removed.   As the region around position W was shown to be detrimental to D_2 binding affinity and the proposed compounds occupy this region all subsequent D_2 binding affinities are poor.   Whether or not this is what would happen in the receptor is not entirely certain, as one might expect some flexing of both the receptor and the compounds in question to accommodate them further.   Unfortunately we are not able to model these additional factors into our predictions of the D_2 binding affinities.   It would be very interesting to see if indeed these proposed detrimental D_2 interactions are realised in the receptor binding assays.

Table 5.4. Compounds designed to occupy the A, B and G regions of the pharmacophore with their predicted D_2 and D_4 affinities (pK_i) and pK_a's.
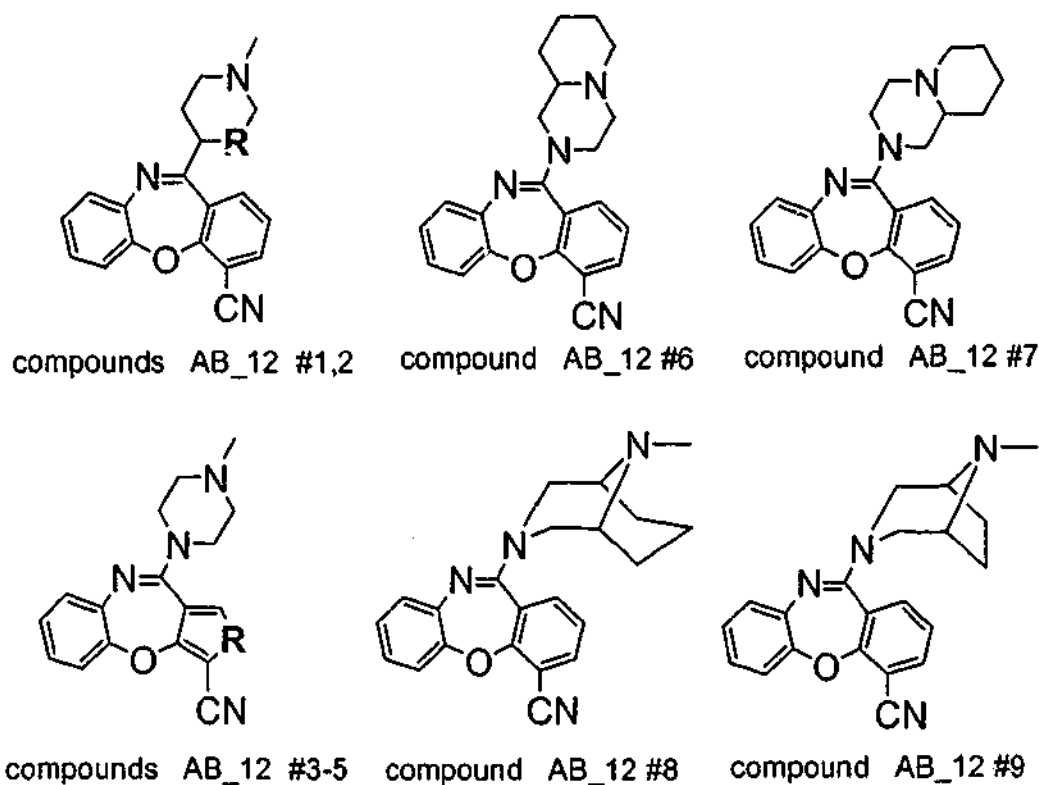
Compound	Z	A'	B'	W'	Predicted		
					D_2	D_4	pK_a
ABG_1	CH_2	CH	NH	H	5.22	8.71	8.08
ABG_2	CH_2	CH	NH	Cl	4.86	8.75	7.75
ABG_3	CH_2	CH	O	H	5.51	8.91	8.00
ABG_4	CH_2	CH	O	Cl	5.06	8.78	7.68
ABG_5	CH_2	N	NH	H	5.53	8.56	8.48
ABG_6	CH_2	N	NH	Cl	4.98	8.52	8.15
ABG_7	CH_2	N	O	H	5.73	8.65	8.51
ABG_8	CH_2	N	O	Cl	5.28	8.38	8.21
ABG_9	NH	CH	NH	H	5.27	8.45	8.45
ABG_10	NH	CH	NH	Cl	5.23	8.57	8.16
ABG_11	NH	CH	O	H	5.67	8.35	8.37
ABG_12	NH	CH	O	Cl	5.54	8.22	8.08
ABG_13	NH	N	NH	H	5.79	8.37	8.78
ABG_14	NH	N	NH	Cl	5.63	8.49	8.45
ABG_15	NH	N	O	H	5.76	8.15	8.72
ABG_16	NH	N	O	Cl	5.65	8.01	8.41
ABG_17	O	CH	NH	H	5.55	9.16	8.05
ABG_18	O	CH	NH	Cl	5.13	8.83	7.78
ABG_19	O	CH	O	H	6.06	9.33	7.98
ABG_20	O	CH	O	Cl	5.71	8.50	7.72
ABG_21	O	N	NH	H	6.20	9.12	8.40
ABG_22	O	N	NH	Cl	5.72	9.01	8.10
ABG_23	O	N	O	H	6.54	9.06	8.34
ABG_24	O	N	O	Cl	6.10	8.95	8.06

The compounds with the greatest D_4 selectivity in this series are the compounds with a CH_2 moiety in position Z and a CH moiety in position A'. The chloro substituent at position W' is shown to increase D_4 selectivity in most cases by reducing the affinity at the D_2 receptor by more than it does at the D_4 receptor. However the full potential of this substituent is not realised due to the rotation of the tricyclic substructure. All predicted pK_a values are in a favourable range so as enable the APD to be unionised to cross the BBB, and ionised to act on the receptor.

5.4 CONCLUSION

Utilising the information gained from the ligand and structure based design strategies reported in this thesis, in conjunction with database mining, a number of compounds have been proposed and predicted to be D_4 selective. The pK_a's of all compounds are predicted to be within a suitable range to enable them to be unionised for crossing the BBB, yet appropriately ionised for interaction at target receptors.

The overall approach has been both comprehensive and novel since it incorporates and integrates deep consideration of SAR studies of known ligands with knowledge of the SAR characteristics of the receptor subtypes, D_2 and D_4, with which they selectively interact. The novelty lies in the two-fold consideration given to the complementary binding of ligands within the GPCR models for D_2 and D_4 receptor subtypes to explore selectivity; and consideration of bioavailability as expressed by the extent of ionisation of the basic nitrogen groups. In addition, the proposed molecules, having been constructed from moieties within drug databases, and used as replacements in drug-like compounds, have a strong likelihood of possessing other drug-like properties, an extremely important requirement for new chemical entities.

However, further exploration of this overall drug design approach is dependent on investigation of the synthetic feasibility and pharmacological properties of the proposed compounds.

APPENDIX

Table Carling. Receptor binding affinities (K_i, nM) of analogues from Carling *et al.*
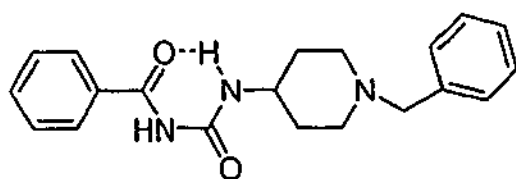
Carling, compound #3
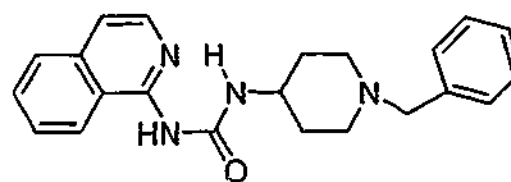
Carling, compound #7
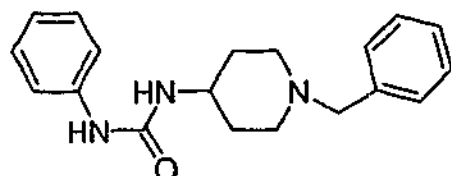
Carling, compound #4
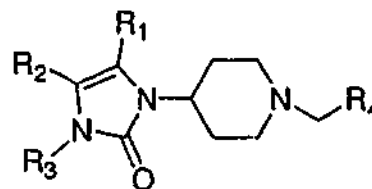
Carling, compounds #8-22,24
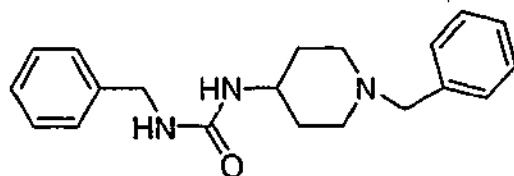
Carling, compound #5
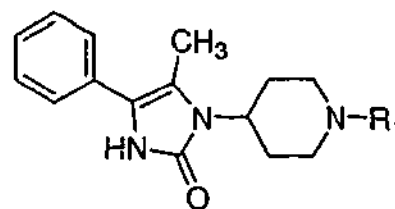
Carling, compound #6

Carling, compounds #23,25-32
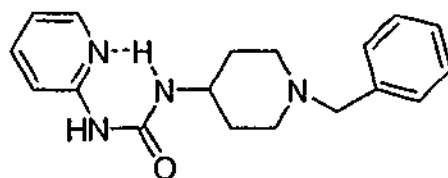
Compound	R_1	R_2	R_3	R_4	D_2	D_4
3					1700	34
4					86	390
5					1600	2700
6					190	150
7					150	5.5
8	methyl	phenyl	H	phenyl	56	0.60
9	phenyl	methyl	H	phenyl	280	71
10	H	phenyl	H	phenyl	82	18
11	ethyl	phenyl	H	phenyl	26	0.86
12	H	phenyl	methyl	phenyl	11	0.84
13	methyl	phenyl	methyl	phenyl	15	0.75
14	methyl	phenyl	phenyl	phenyl	8.8	4.5
15	methyl	4-chlorophenyl	H	phenyl	170	0.71
16	methyl	3-chlorophenyl	H	phenyl	73	5.2
17	methyl	2-chlorophenyl	H	phenyl	38	1.1
18	methyl	4-methoxyphenyl	H	phenyl	110	12
19	methyl	4-fluorophenyl	H	phenyl	290	1.7
20	methyl	4-methylphenyl	H	phenyl	52	1.1

21	methyl	2-thienyl	H	phenyl	7.8	1.0
22	methyl	2-pyridinyl	H	phenyl	270	4.5
23	methyl	phenyl	H	methyl	1600	2500
24	methyl	phenyl	H	cyclohexyl	83	2.0
25	methyl	phenyl	H	3-phenylpropyl	220	0.56
26	methyl	phenyl	H	4-phenylbutyl	330	5.5
27	methyl	phenyl	H	2-(2-chlorophenyl)ethyl	1500	2.1
28	methyl	phenyl	H	2-(3-chlorophenyl)ethyl	1300	0.46
29	methyl	phenyl	H	2-(4-chlorophenyl)ethyl	60	1.3
30	methyl	phenyl	H	2-(3-methylphenyl)ethyl	95	0.33
31	methyl	phenyl	H	2-(3-methoxyphenyl)ethyl	900	0.64
32	methyl	phenyl	H	2-(3-cyanophenyl)ethyl	1900	0.96

Table Rowley. Receptor binding affinities (K_i, nM) of analogues from Rowley *et al.*
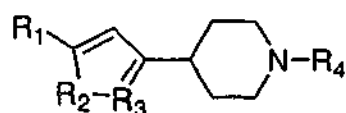
Rowley, compounds #3-19, 29, 32-34, 36
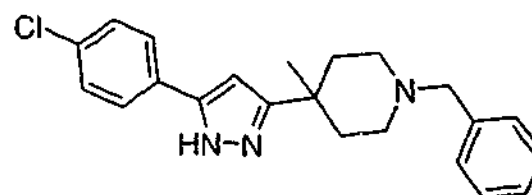
Rowley, compound #27
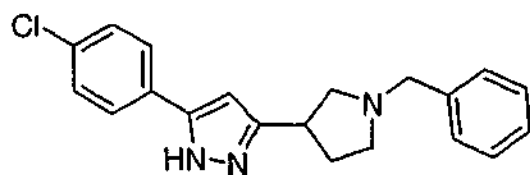
Rowley, compound #22
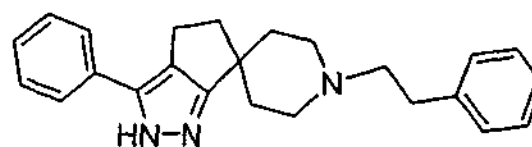
Rowley, compound #28
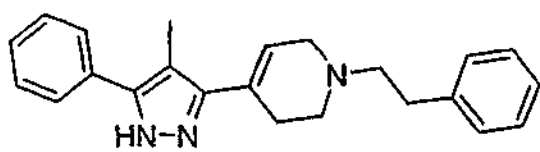
Rowley, compound #23
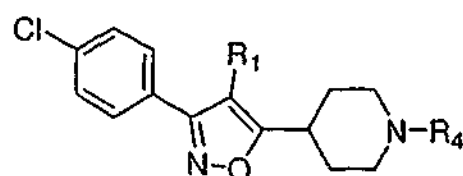
Rowley, compound #24

Rowley, compounds #30,35
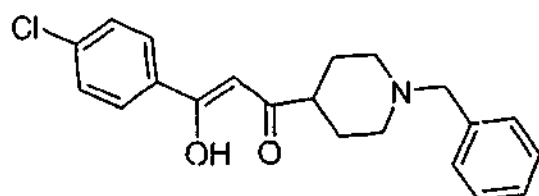
Rowley, compound #25
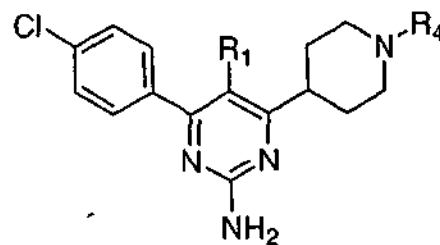
Rowley, compounds #31,37

Compound	R_1	R_2	R_3	R_4	D_2	D_4
3	4-chlorophenyl	NH	N	2-(4-chlorophenyl)ethyl	240	61
4	4-chlorophenyl	NH	N	H	1800	3300
5	4-chlorophenyl	NH	N	methyl	590	1200
6	4-chlorophenyl	NH	N	cyclohexylethyl	120	13
7	4-chlorophenyl	NH	N	2-(3-chlorophenyl)ethyl	570	60
8	4-chlorophenyl	NH	N	2-(2-chlorophenyl)ethyl	700	20
9	4-chlorophenyl	NH	N	2-phenylethyl	125	9.1
10	4-chlorophenyl	NH	N	3-phenylpropyl	520	5.5
11	4-chlorophenyl	NH	N	4-phenylbutyl	560	110
12	methyl	NH	N	benzyl	650	470
13	phenyl	NH	N	benzyl	66	11
14	benzyl	NH	N	benzyl	310	160
15	cyclohexyl	NH	N	2-phenylethyl	61	10
16	2-thienyl	NH	N	2-phenylethyl	65	19
17	2-pyridinyl	NH	N	2-phenylethyl	860	22
18	3-pyridinyl	NH	N	2-phenylethyl	900	23
19	4-pyridinyl	NH	N	2-phenylethyl	1300	52
22					120	110
23					90	3.1
24					1900	23
25					120	12
27					240	47
28					690	930
29	4-chlorophenyl	O	N	4-chlorobenzyl	160	44
30	H			4-chlorobenzyl	130	3.6
31	H			benzyl	8.6	1.0
32	4-chlorophenyl	NH	N	benzyl	200	1.6
33	4-chlorophenyl	NH	N	benzyl	35	1.5
34	4-chlorophenyl	NH	N	2-phenylethyl	290	1.2
35	methyl			2-phenylethyl	1700	2.5
36	4-chlorophenyl	NH	N	2-phenylethyl	1700	3.5
37	methyl				1500	4.7

.

Table Moore. Receptor binding affinities (K_i, nM) of analogues from Moore *et al.*
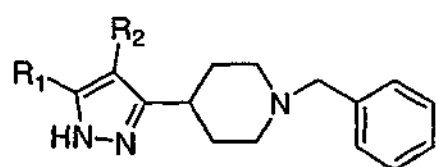
Moore compounds #1,2
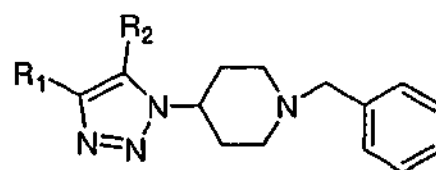
Moore compounds #8-10
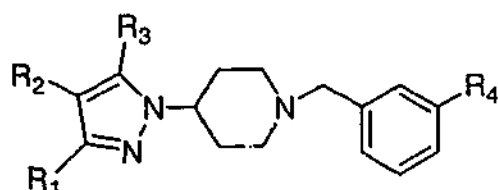
Moore compounds #3-6,13-16
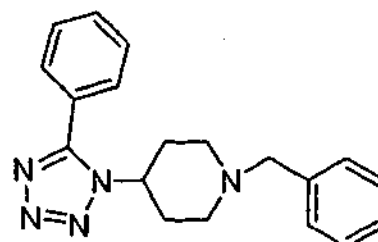
Moore compound #11
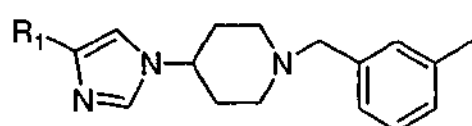
Moore compound #7
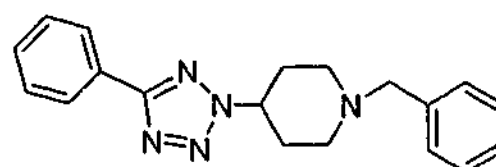
Moore compound #12

Compound	R_1	R_2	R_3	R_4	D_2	D_4
1	phenyl	H			66	11
2	phenyl	methyl			90	4.7
3	H	phenyl	methyl	H	140	2.5
4	methyl	phenyl	H	H	6.1	0.39
5	methyl	H	phenyl	H	310	140
6	H	H	phenyl	H	510	15
7	phenyl				450	40
8	methyl	phenyl			1200	18
9	phenyl	methyl			1600	17
10	phenyl	H			460	32
11					1800	1200
12					1100	27
13	H	phenyl	methyl	cyano	1700	5.2
14	H	phenyl	methyl	chloro	370	3.1
15	methyl	phenyl	H	cyano	15	1.2
16	methyl	phenyl	H	chloro	15	3.1

Table Ohmori. Receptor binding affinities (K_i, nM) of analogues from Ohmori *et al*.
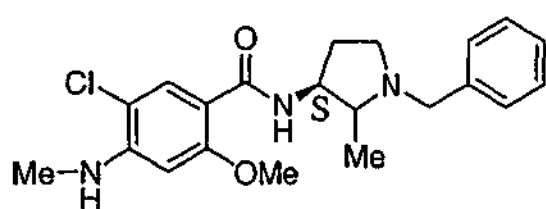
nemonapride
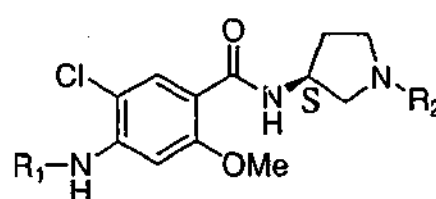
Ohmori compounds   #1,2a-d,5a-l,7

Compound	R_1	R_2	D_2	D_4
1	H	benzyl	0.98	1.1
(R)-1	H	benzyl	1.2	0.55
2a	trifluoromethylcarbonyl	benzyl	12	1.9
2b	ethenylcarbonyl	benzyl	27	1.3
2c	cyclohexylcarbonyl	benzyl	1200	110
2d	phenylcarbonyl	benzyl	630	87
5a	ethylcarbonyl	benzyl	32	1.5
5b	isopropylcarbonyl	benzyl	50	3.0
5c	cyclopropylcarbonyl	benzyl	220	2.1
(R)-5c	cyclopropylcarbonyl	benzyl	190	5.6
5d	cyclobutylcarbonyl	benzyl	170	5.2
(R)-5d	cyclobutylcarbonyl	benzyl	240	7.0
5e	cyclopentylcarbonyl	benzyl	740	20
(R)-5e	cyclopentylcarbonyl	benzyl	690	18
5f	cyclopropylcarbonyl	cyclohexyl	84	2.1
5g	cyclopropylcarbonyl	cycloheptyl	20	2.7
5h	cyclopropylcarbonyl	BCN*	140	3.8
5i	cyclopropylcarbonyl	2-adamantyl	110	4.4
5j	cyclopropylcarbonyl	1-(1-methylcyclohexyl)	32	3.3
5k	cyclopropylcarbonyl	2-phenylethyl	82	3.4
5l	methyl	BCN*	0.81	0.67
7	cyclopropylcarbonyl	H	14000	2200
nemonapride			0.16	0.21

*BCN bicyclo[3.3.1]non-9-yl

Table Sanner. Receptor binding affinities (K_i, nM) of analogues from Sanner *et al.*
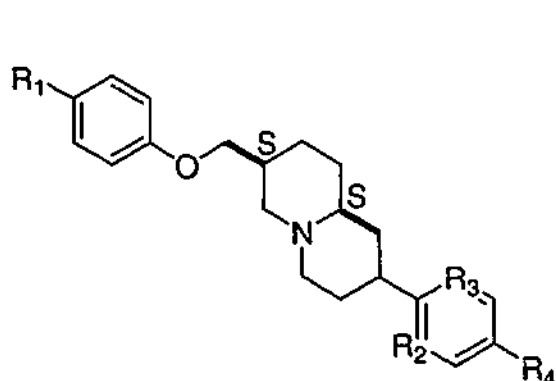
Sanner compounds #3b,c
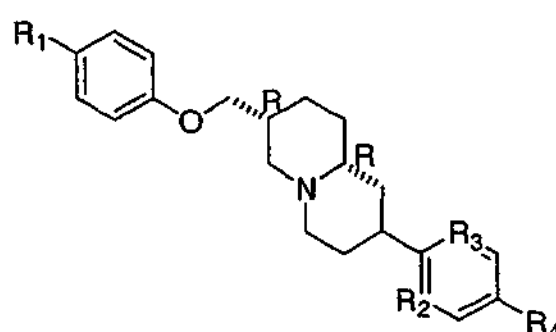
Sanner compound #3d
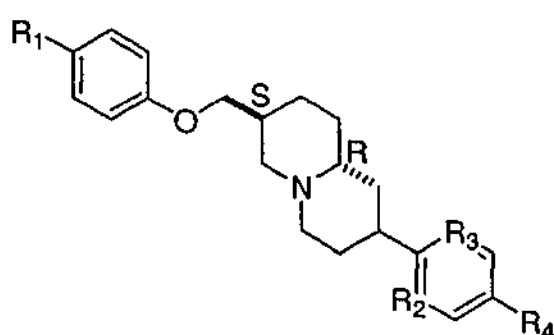
Sanner compounds #10m
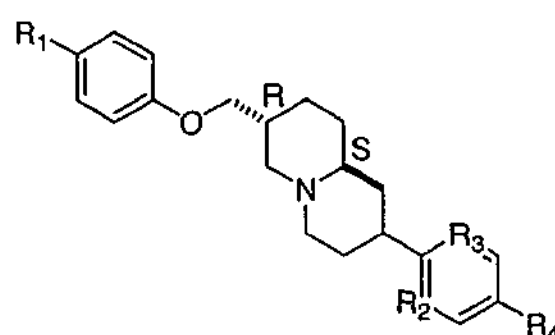
Sanner compounds #10a-l

Compound	R_1	R_2	R_3	R_4	D_2	D_4
3b	H	N	N	H	38	1.7
3c	fluoro	N	N	fluoro	68	2.0
3d	fluoro	N	N	fluoro	980	39
10b	H	N	N	H	1140	3.5
10c	fluoro	N	N	H	1196	2.8
10d	methoxy	N	N	H	918	18
10e	tert-butyl	N	N	H	1720	20
10f	methoxycarbonyl	N	N	H	254	127
10g	*N*-acetyl	N	N	H	948	316
10h	H	CH	N	H	187	1.7
10i	H	CH	CH	H	44	1.6
10j	fluoro	N	N	fluoro	3310	3.4
10k	fluoro	CH	N	chloro	1880	3.2
10l	fluoro	CH	CH	fluoro	206	5.3
10m	fluoro	N	N	fluoro	195	106

Table Belliotti. Receptor binding affinities (K_i, nM) of analogues from Belliotti *et al*.
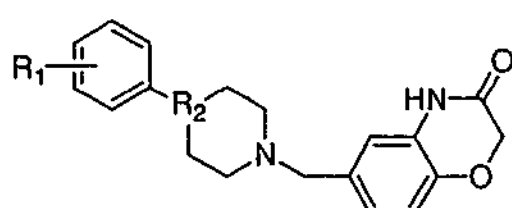
Belliotti compounds #16-23
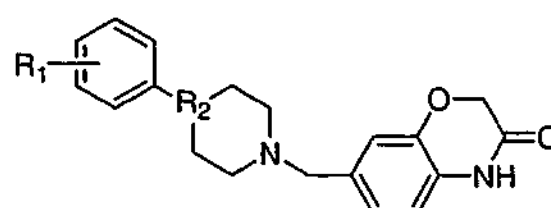
Belliotti compounds #24-31

Compound	R_1	R_2	D_2	D_4
16	3,4-dimethyl	CH	572	1.8
17	4-methyl	CH	52.5	2.8
18	4-methoxy	CH	698	3.8
19	4-methyl	N	493	4.3
20	3,4-dimethyl	N	2340	5.1
21	4-methoxy	N	5590	12.3
22	H	N	2200	18.3
23	4-chloro	N	5880	62.4
24	4-methyl	CH	287	2.6
25	3,4-dimethyl	CH	610	2.6
26	4-chloro	N	413	4.3
27	4-methoxy	N	5880	5.0
28	H	N	129	6.1
29	4-methoxy	CH	1130	6.7
30	3,4-dimethyl	N	346	6.9
31	4-methyl	N	2980	10 7

Table Kesten. Receptor binding affinities (K_i, nM) of analogues from Kesten *et al.*
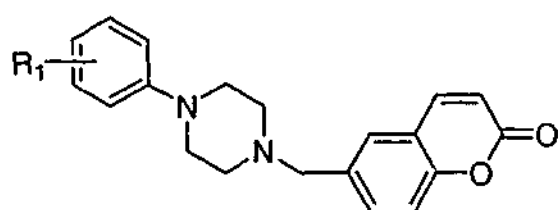
Kesten compounds   #6,8,10,12
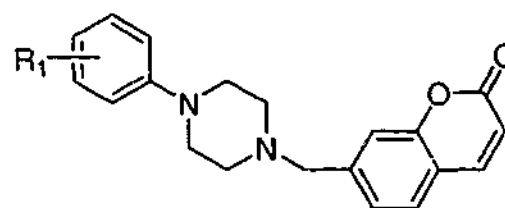
Kesten compounds   #5,7,9,11,13-18
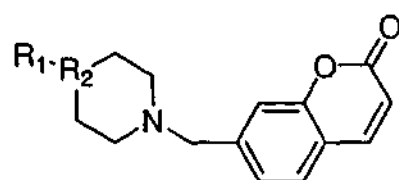
Kesten compounds   #19,21
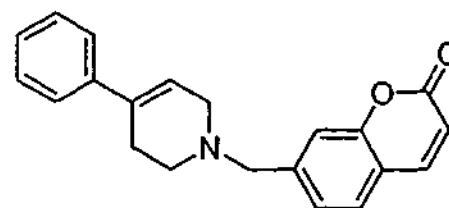
Kesten compound   #20

Compound	R_1	R_2	D_2	D_4
5	H		5300	27
6	H		2170	30
7	4-methyl		5700	17
8	4-methyl		5836	11
9	2-methyl		1940	16
10	2-methyl		1212	63
11	3-methyl		4400	32
12	3-methyl		NT	74
13	3-chloro,4-methyl		5882	6.9
14	3,4-dimethyl		433	14
15	4-fluoro		NT	59
16	4-chloro		NT	93
17	4-methoxy		NT	75
18	t-butyl		NT	227
19	CH	phenyl	91	19
20			121	4.4
21	N	benzyl	NT	2700

NT not tested

Table Thurkauf. Receptor binding affinities (K_i, nM) of analogues from Thurkauf *et al.*
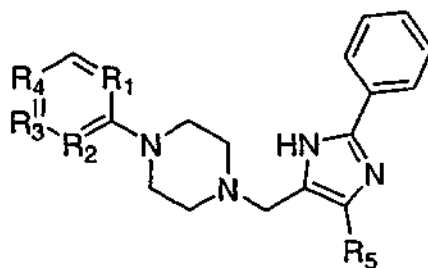
Compound	R_1	R_2	R_3	R_4	R_5	D_2	D_4
3a	CH	CH	CH	CH	H	254	5.2
3b	CH	N	CH	CH	H	2190	8.5
3c	CH	CH	N	CH	H	296	73
3d	CH	CH	CH	N	H	3500	7300
3e	N	CH	N	CH	H	4130	50
3f	CH	N	CH	N	H	4130	1400
3g	N	N	CH	CH	H	2250	3.8
3h	N	N	CH	CH	methyl	1440	6.0

Table 1. Meta/para-substituted phenols

Observed and calculated pK_a values for the meta/para-substituted phenols dataset using eq. (3).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000051-67-2	4-(2-aminoethyl)phenol[†]	9.77	9.88
000059-50-7	3-methyl-4-chlorophenol	9.20*	9.31
000080-46-6	4-tert-amylphenol	10.43	10.09
000088-30-2	3-trifluoromethyl-4-nitrophenol	6.07	6.18
000088-04-0	4-chloro-3,5-dimethyl phenol	9.70	9.47
000092-69-3	4-phenylphenol	9.55	9.75
000094-26-8	4-hydroxy butyl benzoate	8.47	8.58
000095-77-2	3,4-dichlorophenol	8.63	8.63
000095-65-8	3,4-dimethylphenol	10.36	10.14
000098-54-4	4-t-butylphenol	10.39	10.09
000098-17-9	3-trifluoromethylphenol	8.95	8.48
000099-93-4	4-hydroxyacetophenone	8.05	8.74
000099-89-8	4-isopropylphenol	10.24	10.04
000100-83-4	3-hydroxybenzaldehyde	8.98	9.00
000100-02-7	4-nitrophenol	7.15	7.17
000103-90-2	n-(4-hydroxyphenyl)acetamide	9.38	9.20
000106-48-9	4-chlorophenol	9.41	9.15
000106-44-5	4-cresol	10.26	10.04
000106-41-2	4-bromophenol	9.17	8.94
000108-95-2	phenol	9.99	9.93
000108-68-9	3,5-dimethylphenol	10.19	10.14
000108-43-0	3-chlorophenol	9.12	9.15
000108-39-4	3-cresol	10.09	10.04
000120-47-8	4-hydroxybenzoic acid,ethyl ester	8.34	8.58
000121-71-1	3-hydroxyacetophenone	9.25	9.36
000123-30-8	4-aminophenol	10.45	10.51
000123-08-0	4-hydroxybenzaldehyde	7.61	8.58
000123-07-9	4-ethylphenol	10.00	10.04
000150-76-5	4-methoxyphenol	10.10	10.14
000150-19-6	3-methoxyphenol	9.65	9.78
000371-41-5	4-fluorophenol	9.91	9.26
000372-20-3	3-fluorophenol	9.21	9.05
000402-45-9	4-trifluoromethylphenol	8.68	8.11
000500-99-2	3,5-dimethoxyphenol	9.34	9.41
000540-38-5	4-iodophenol	9.21	8.94

000554-84-7	3-nitrophenol	8.36	7.80
000577-71-9	3,4-dinitrophenol	5.42	5.66
000580-51-8	3-phenylphenol	9.64	9.88
000585-34-2	3-(1,1-dimethylethyl)-phenol	10.12	10.14
000586-11-8	3,5-dinitrophenol	6.69	6.08
000591-35-5	3,5-dichlorophenol	8.18	8.53
000591-27-5	3-aminophenol	9.86	10.09
000591-20-8	3-bromophenol	9.03	9.15
000609-19-8	3,4,5-trichlorophenol	7.84	8.11
000618-45-1	3-isopropylphenol	10.16	10.09
000620-17-7	3-ethylphenol	9.90	10.04
000621-34-1	3-ethoxyphenol	9.65	9.83
000622-62-8	4-ethoxyphenol	10.13	10.25
000626-41-5	3,5-dibromophenol	8.06	8.48
000626-02-8	3-iodophenol	9.03	9.15
000645-56-7	4-propylphenol	10.34	10.04
000698-71-5	3-ethyl-5-methylphenol	10.10	10.14
000767-00-0	4-cyanophenol	7.97	8.32
000873-62-1	3-cyanophenol	8.61	8.68
001073-72-9	4-methiophenol	9.53	9.57
001470-94-6	5-indanol	10.32	10.19
002042-14-0	3-nitro-4-cresol	8.62	8.06
007339-87-9	hydroxyacetophenone	8.05	8.74

[†] This compound may be partially in a zwitterionic form. However, it has been modelled in the neutral form.

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

Table 2. Ortho-substituted phenols: hydrogen bonded

Observed and calculated pK_a values for the ortho-substituted phenols dataset (capable of forming internal hydrogen bonds) using eq. (4).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000065-45-2	2-hydroxybenzamide	8.89*	9.10
000087-17-2	salicylanilide	7.40	8.19
000119-36-8	methyl salicylate	9.87	7.74
000148-53-8	2-vanillin	7.91	6.83
001151-51-5	3,5,4'-trichloro salicylanilide	4.70	5.25
001697-18-3	2'-chloro salicylanilide	7.31	7.85

CAS no.	Chemical name	Exp pKa	Calc pKa
002389-37-9	5-nitro salicylanilide	3.03	2.76
002627-77-2	4'-bromo salicylanilide	7.31	7.40
003679-64-9	4'- chloro-5-bromo salicylanilide	6.00	5.81
003679-63-8	4'-chloro salicylanilide	7.30	7.51
004214-48-6	3,5-dichloro salicylanilide	4.70	5.70
004638-48-6	5-chlorosalicylanilide	6.17	6.72
025933-30-6	5- chloro-2'-methyl salicylanilide	6.60	6.83
037183-28-1	2',4'-dichloro salicylanilide	7.14	7.29
037183-26-9	2'-nitro salicylanilide	6.91	7.06
037399-40-9	2'-nitro-4'-chloro salicylanilide	6.74	6.61
040912-87-6	5-bromo-2-hydroxy-n,3-dimethyl-benzamide	7.52	7.40
054850-02-1	3,5-dichloro-4'-fluoro salicylanilide	4.80	5.25
072699-09-3	3,5-dibromo-2'-nitro-4'-chloro salicylanilide	4.11	3.66
077067-91-5	2'-methyl-4'-chloro salicylanilide	7.43	7.63
077068-04-3	5-fluoro-2'-methyl-4'-bromo salicyanilide	7.10	6.27
077068-02-1	5-fluoro-2'-methyl-4'-chloro salicyanilide	7.30	6.49
079402-07-6	3,5-dibromo-2',4'-difluoro salicylanilide	4.77	4.34
080033-99-4	3,5-dichloro-2',4'-difluoro salicylanilide	4.77	5.02
090426-05-4	3,5,-4'-trichloro-4'-nitro salicylanilide	4.11	4.34
090426-03-2	3,5-dichloro-2'-methyl-4'-nitro salicylanilide	4.41	5.25

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

Table 3. Ortho-substituted phenols: no hydrogen bonds

Observed and calculated pK_a values for the ortho-substituted phenols dataset (excluding those forming intra-molecular hydrogen bonds) using eq. (7).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000051-28-5	2,4-dinitrophenol	4.09	3.30
000058-90-2	2,3,4,6-tetrachlorophenol	5.22	6.56
000066-56-8	2,3-dinitrophenol	4.96	4.28
000087-86-5	pentachlorophenol	4.70	6.03
000087-65-0	2,6-dichlorophenol	6.79	7.99
000087-64-9	2-methyl-6-chlorophenol	8.69	9.05
000088-89-1	2,4,6-trinitrophenol	0.38	0.80
000088-87-9	4-chloro-2,6-dinitrophenol	2.96	3.00
000088-85-7	2-sec-butyl-4,6-dinitrophenol	4.62	3.83
000088-75-5	2-nitrophenol	7.23	6.63
000088-69-7	2-isopropylphenol	10.47	10.26

000088-18-6	2-t-butylphenol	10.28	10.11
000088-06-2	2,4,6-trichlorophenol	6.23	7.09
000089-83-8	thymol	10.62	10.49
000089-68-9	chlorothymol	9.98	9.43
000089-64-5	4-chloro-2-nitrophenol	6.46	5.80
000090-43-7	2-phenylphenol	9.92	10.11
000090-05-1	2-methoxyphenol	9.98	10.72
000090-00-6	2-ethylphenol	10.20	10.19
000093-51-6	4-methyl-2-methoxyphenol	10.28	10.87
000094-71-3	2-ethoxyphenol	10.11	10.79
000095-95-4	2,4,5-trichlorophenol	7.40	7.16
000095-87-4	2,5-dimethylphenol	10.41	10.26
000095-57-8	2-chlorophenol	8.56	8.83
000095-56-7	2-bromophenol	8.45	8.67
000095-55-6	2-aminophenol	9.75	11.32
000095-48-7	2-cresol	10.28	10.04
000096-76-4	2,4-di-t-butylphenol	11.72	10.34
000097-54-1	2-methoxy-4-(1-propenyl)phenol	9.88	10.64
000097-53-0	eugenol	10.19	10.72
000098-28-2	4-(tert-butyl)-2-chlorophenol	8.58	9.13
000098-27-1	4-(t-butyl)-2-cresol	10.59	10.26
000099-57-0	2-amino-4-nitrophenol	7.60	6.56
000099-28-5	2,6-dibromo-4-nitrophenol	3.39	4.13
000105-67-9	2,4-dimethylphenol	10.60	10.19
000118-79-6	2,4,6-tribromophenol	6.80	6.48
000119-34-6	phenol, 4-amino-2-nitro-	7.81	7.62
000119-33-5	4-methyl-2-nitrophenol	7.40*	6.86
000120-83-2	2,4-dichlorophenol	7.89	7.84
000121-33-5	vanillin	7.40	8.67
000128-39-2	2,6-di-t-butylphenol	11.70	11.10
000128-37-0	2,6-di-t-butyl-4-methylphenol (bht)	12.23	11.25
000131-89-5	2-cyclohexyl-4,6-dinitrophenol	4.52	3.83
000329-71-5	2,5-dinitrophenol	5.21	4.13
000367-12-4	2-fluorophenol	8.70	8.83
000446-36-6	5-fluoro-2-nitrophenol	6.07	5.65
000496-78-6	2,4,5-trimethylphenol	10.57	10.42
000526-75-0	2,3-dimethylphenol	10.54	10.26
000527-60-6	2,4,6-trimethylphenol	10.86	10.42
000533-58-4	2-iodophenol	8.51	8.52
000534-52-1	4,6-dinitro-o-cresol	4.31	3.53

000554-52-9	2-methyldopamine[†]	9.54	10.64
000573-56-8	2,6-dinitrophenol	3.97	3.60
000576-26-1	2,6-dimethylphenol	10.62	10.19
000576-24-9	2,3-dichlorophenol	7.70	7.99
000583-78-8	2,5-dichlorophenol	7.51	7.92
000603-86-1	6-chloro-2-nitrophenol	5.48	5.95
000608-71-9	pentabromophenol	4.62	5.57
000608-33-3	2,6-dibromophenol	6.67	7.62
000609-93-8	2,6-dinitro-p-cresol	4.23	3.83
000611-20-1	2-cyanophenol	6.86	7.54
000611-07-4	5-chloro-2-nitrophenol	6.05	5.87
000615-58-7	2,4-dibromophenol	7.79	7.39
000618-80-4	2,6-dichloro-4-nitrophenol	3.55	4.44
000619-08-9	2-chloro-4-nitrophenol	5.45	5.12
000621-59-0	isovanillin	8.89	9.43
000644-35-9	2-propylphenol	10.47	10.19
000697-82-5	2,3,5-trimethylphenol	10.67	10.42
000700-38-9	5-methyl-2-nitrophenol	7.41	6.86
000731-92-0	2,4-dinitro-6-phenylphenol	3.85	3.75
000732-26-3	2,4,6-tri(tert-butyl)phenol	12.19	11.32
000771-61-9	pentafluorophenol	5.53	4.81
000885-82-5	4-phenyl-2-nitrophenol	6.73	6.63
000933-75-5	2,3,6-trichlorophenol	5.80	7.24
000935-95-5	2,3,5,6-tetrachlorophenol	5.14	6.56
000946-31-6	6-chloro-2,4-dinitrophenol	2.10	2.77
001568-70-3	4-methoxy-2-nitrophenol	7.31	7.24
001570-64-5	2-methyl-4-chlorophenol	9.71	8.98
001689-84-5	bromoxynil	3.86	5.65
001879-09-0	2-(1,1-dimethylethyl)-4,6-dimethylphenol	12.04	10.57
002078-54-8	phenol, 2,6-bis(1-methylethyl)-	11.10	10.57
002409-55-4	2-(tert-butyl)-4-methylphenol	11.72	10.26
002423-71-4	2,6-dimethyl-4-nitrophenol	7.07	6.33
002432-12-4	4-methyl-2,6-dichlorophenol	7.19	8.22
003217-15-0	4-bromo-2,6-dichlorophenol	6.21	6.86
003555-18-8	4-(sec-butyl)-2-nitrophenol	7.59	6.93
003964-58-7	3-chloro-4-hydroxybenzoic acid	7.52	6.71
004901-51-3	2,3,4,5-tetrachlorophenol	6.35	6.56
005428-54-6	phenol, 2-methyl-5-nitro-	8.59	7.01
006640-27-3	2-methyl-4-chlorophenol	8.74	8.98
013181-17-4	bromofenoxim	5.46	5.87

† This compound may be partially in a zwitterionic form. However, it has been modelled in the neutral form.

* Value taken from Rapoport, M., Kinney Hancock, C., and Meyers, E.A. *J. Am. Chem. Soc.* 83, 3489-3494 (1961).

Table 4. Meta/para-substituted benzoic acids

Observed and calculated pK_a values for the meta/para-substituted benzoic acids dataset using eq. (11).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000051-44-5	3,4-dichlorobenzoic acid	3.64	3.79
000051-36-5	3,5-dichlorobenzoic acid	3.54	3.75
000057-66-9	probenecid	3.40	3.65
000062-23-7	p-nitrobenzoic acid	3.44	3.33
000065-85-0	benzoic acid	4.19	4.21
000074-11-3	4-chlorobenzoic acid	3.98	3.99
000093-09-4	2-naphthoic acid	4.17	4.23
000093-07-2	3,4-dimethoxybenzoic acid	4.36	4.30
000098-73-7	4-(tert-butyl)- benzoic acid	4.40	4.32
000099-96-7	p-hydroxybenzoic acid	4.54	4.28
000099-94-5	p-toluic acid	4.37	4.30
000099-50-3	3,4-dihydroxybenzoic acid	4.48*	4.15
000099-34-3	3,5-dinitrobenzoic acid	2.82	2.81
000099-10-5	3,5-dihydroxybenzoic acid	4.04	3.97
000099-05-8	3-aminobenzoic acid†	4.74	4.28
000099-04-7	m-toluic acid	4.27	4.25
000100-09-4	p-methoxybenzoic acid	4.47	4.36
000121-92-6	m-nitrobenzoic acid	3.46	3.43
000121-34-6	4-hydroxy-3-methoxybenzoic acid	4.51	4.23
000149-91-7	3,4,5-trihydroxybenzoic acid	4.21**	3.93
000150-13-0	4-aminobenzoic acid†	4.85	4.62
000455-38-9	m-fluorobenzoic acid	3.86	3.91
000456-22-4	p-fluorobenzoic acid	4.14	4.01
000528-45-0	3,4-dinitrobenzoic acid	2.82	2.77
000530-57-4	4-OH-3,5-dimethoxybenzioc acid	4.34	4.15
000535-80-8	m-ch'orobenzoic acid	3.81	3.97
000536-66-3	cumic acid	4.35	4.30
000585-76-2	m-bromobenzoic acid	3.81	3.95
000586-89-0	p-acetylbenzoic acid	3.70	3.87

000586-76-5	p-bromobenzoic acid	4.00	3.93
000586-38-9	m-methoxybenzoic acid	4.09	4.19
000618-51-9	3-iodobenzoic acid	3.85	3.97
000619-86-3	p-ethoxybenzoic acid	4.45***	4.40
000619-66-9	4-formylbenzoic acid	3.77	3.85
000619-65-8	p-cyanobenzoic acid	3.55	3.71
000619-64-7	4-ethylbenzoic acid	4.35	4.28
000619-58-9	4-iodobenzoic acid	4.00	3.91
000619-21-6	3-formylbenzoic acid	3.84	3.85
000619-05-6	3,4-diamino-benzoic acid[†][++]	3.49	4.50
001132-21-4	3,5-dimethoxybenzoic acid	3.97	4.19
001877-72-1	m-cyanobenzoic acid	3.60	3.75
002215-77-2	p-phenoxybenzoic acid	4.52	4.28
003739-38-6	m-phenoxybenzoic acid	3.92	4.09
004052-30-6	p-methylsulfonylbenzoic acid	3.64	3.49
005438-19-7	4-propoxybenzoic acid	4.46***	4.40
007496-53-9	4-[(acetylamino)amino]-benzoic acid	4.20	4.52

[†] This compound may be partially in a zwitterionic form. However, it has been modelled in the neutral form.

[++] pK_a predicted using equation (10)

* Value taken from Shorter, J., and Stubbs, F.J. *J. Chem. Soc.* 1180 (1949).

** Value taken from http://www.sirius-analytical.com.

*** Value taken from Brown, H.C. *et al.*, in *Determination of Organic Structures by Physical Methods*; E.A. Braude, F.C. Nachod (Eds.); Academic Press: New York, 1995; Cavill, G.W.K., Gibson, N.A., and Nyholm, R.S. *J. Chem. Soc.* 2466 (1949).

Table 5. Ortho-substituted benzoic acids

Observed and calculated pK_a values for the ortho-substituted benzoic acids dataset using eq. (14).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000050-85-1	4-methylsalicylic acid	3.40	3.21
000050-84-0	2,4-dichlorobenzoic acid	2.68	2.85
000050-79-3	2,5-dichlorobenzoic acid	2.47	2.82
000050-78-2	acetylsalicylic acid	3.49	3.31
000050-31-7	2,3,6-trichlorobenzoic acid	1.50	2.33
000050-30-6	2,6-dichlorobenzoic acid	1.59	2.64
000059-07-4	2-ethoxy-4-aminobenzoic acid[†]	5.09	4.96
000061-68-7	mefenamic acid	4.20	3.87

000065-49-6	p-aminosalicylic acid[t]	3.66	3.84
000069-72-7	salicylic acid	2.97	3.07
000083-40-9	3-methylsalicylic acid	2.95	3.10
000088-67-5	2-iodobenzoic acid	2.93	2.68
000088-65-3	o-bromobenzoic acid	2.88	3.10
000089-86-1	2,4-dihydroxybenzoic acid	3.11	3.21
000089-56-5	5-methylsalicylic acid	3.15	3.14
000089-55-4	5-bromosalicylic acid	2.66	2.19
000089-52-1	n-acetyl o-aminobenzoic acid	3.40	3.00
000091-52-1	2,4-dimethoxybenzoic acid	4.36	4.40
000091-40-7	n-phenyl o-aminobenzoic acid	3.99	3.77
000092-70-6	2-naphthalenecarboxylic acid, 3-hydroxy-	2.79	2.82
000096-97-9	5-nitrosalicylic acid	2.12	1.70
000099-60-5	2-chloro-4-nitro-benzoic acid	2.14	1.73
000118-92-3	2-aminobenzoic acid	4.95	4.08
000118-91-2	2-chlorobenzoic acid	2.89	3.21
000118-90-1	o-toluic acid	3.98	3.63
000119-30-2	2-hydroxy-5-iodo-benzoic acid	2.62	2.61
000129-66-8	2,4,6-trinitrobenzoic acid	0.65	0.75
000133-91-5	3,5-diiodosalicylic acid	2.30	2.22
000133-90-4	3-amino-2,5-dichlorobenzoic acid[t]	3.40	2.78
000303-38-8	2,3-dihydroxybenzoic acid	2.91	2.71
000303-07-1	2,6-dihydroxybenzoic acid	1.05	2.50
000321-14-2	5-chlorosalicylic acid	2.65	2.64
000445-29-4	2-fluorobenzoic acid	3.27	3.28
000490-79-9	2,5-dihydroxybenzoic acid	2.95	2.89
000552-16-9	2-nitrobenzoic acid	2.17*	2.26
000577-56-0	o-acetylbenzoic acid	4.13	3.00
000579-75-9	o-methoxybenzoic acid	3.90	4.15
000609-99-4	2-hydroxy-3,5-dinitro-benzoic acid	0.70	0.60
000610-30-0	2,4-dinitrobenzoic acid	1.42	1.09
000632-46-2	2,6-dimethylbenzoic acid	3.35	3.59
000652-32-4	2,3,5,6-tetrafluoro-4-methyl-benzoic acid	2.00	2.33
000947-84-2	[1,1'-biphenyl]-2-carboxylic acid	3.46	3.21
001466-76-8	2,6-dimethoxybenzoic acid	3.44	4.19
001521-38-6	2,3-dimethoxybenzoic acid	3.98	3.52
001918-00-9	3,6-dichloro-2-methoxybenzoic acid	1.97	2.64
002243-42-7	o-phenoxybenzoic acid	3.53	3.84
002438-04-2	o-isopropylbenzoic acid	3.63	3.42
002516-96-3	2-chloro-5-nitrobenzoic acid	2.17	1.80

003970-35-2	2-chloro-3-nitrobenzoic acid	2.02**	1.77
004727-29-1	n-phenylphthalamic acid	2.50	2.12
005344-49-0	2-chloro-6-nitro-benzoic acid	1.34	2.26
021327-86-6	2-chloro-6-methyl-benzoic acid	2.75	2.96
src003-73-5	2-[(acetylamino)amino]-benzoic acid	4.20	4.05

† This compound may be partially in a zwitterionic form. However, it has been modelled in the neutral form.

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

** Value from Dippy, J.F.J., and Hughes, S.R.H. *Tetrahedron, 19*, 1527 (1963).

Table 6. Aliphatic carboxylic acids: excluding amino acids

Observed and calculated pK_a values for the aliphatic carboxylic acids dataset (excluding amino acids) using eq. (18).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000050-21-5	lactic acid	3.86	3.89
000053-86-1	indomethacin	4.50	4.70
000061-78-9	glycine, n-(4-aminobenzoyl)-	3.80	3.78
000061-33-6	benzylpenicillin	2.74	2.52
000061-32-5	methicillin	2.77	3.14
000064-69-7	iodo-acetic acid	3.15	3.17
000064-19-7	acetic acid	4.76	4.43
000068-11-1	mercaptoacetic acid	3.55	3.62
000075-99-0	2,2-dichloro-proprionic acid	1.79	1.89
000075-98-9	2,2-dimethyl propanoic acid	5.03	4.60
000076-93-7	benzilic acid	3.05	3.94
000076-05-1	trifluoroacetic acid	0.52	0.04
000076-03-9	trichloroacetic acid	0.51	1.01
000077-06-5	gibberellic acid	4.00	3.76
000079-43-6	dichloroacetic acid	1.26	1.68
000079-31-2	isobutyric acid	4.84	4.54
000079-14-1	hydroxyacetic acid	3.83	3.72
000079-11-8	chloroacetic acid	2.87	2.84
000079-09-4	propionic acid	4.88	4.46
000079-08-3	bromoacetic acid	2.89	2.81
000081-25-4	cholic acid	4.98	4.74
000085-34-7	(2,3,6-trichlorophenyl)acetic acid	3.70	3.85
000086-87-3	naphthaleneacetic acid	4.23	4.18

000087-51-4	indole-3-acetic acid	4.75	4.87
000087-08-1	phenoxymethylpenicillin	2.79	2.73
000088-09-5	2-ethylbutyric acid	4.71	4.72
000090-64-2	a-hydroxyphenylacetic acid	3.41	3.70
000093-76-5	2,4,5-trichlorophenoxyacetic acid	2.83	3.11
000093-72-1	2-(2,4,5-trichlorophenoxy)propionic acid	2.84	3.10
000094-82-6	4-(2,4-dichlorophenoxy)butyric acid	4.95	3.84
000094-81-5	4-(4-chloro-o-tolyloxy)butyric acid (MCPB)[++]	6.20	3.90
000094-75-7	2,4-dichlorophenoxyacetic acid	2.73	3.32
000094-74-6	2-methyl-4-chlorophenoxyacetic acid	3.13	3.42
000097-61-0	2-methyl- pentanoic acid	4.79	4.57
000098-89-5	cyclohexanecarboxylic acid	4.90	4.57
000099-66-1	valproic acid	4.60	4.69
000102-32-9	3,4-dihydroxyphenylacetic acid	4.25	4.31
000103-82-2	phenylacetic acid	4.31	4.44
000104-03-0	p-nitrophenylacetic acid	3.85	3.08
000104-01-8	p-methoxyphenylacetic acid	4.36	4.56
000107-94-8	3-chloropropionic acid	3.99	3.77
000107-92-6	butyric acid	4.82	4.48
000116-53-0	2-methyl- butanoic acid	4.81	4.57
000117-34-0	diphenylacetic acid	3.94	4.51
000120-36-5	2-(2,4-dichlorophenoxy)propanoic acid	3.10	3.27
000122-88-3	p-chlorophenoxyacetic acid	3.10*	2.83
000122-59-8	phenoxyacetic acid	3.17	3.06
000123-76-2	levulinic acid	4.64	4.33
000141-82-2	malonic acid	2 85	2.74
000141-76-4	3-iodopropionic acid	4.09	3.85
000144-49-0	fluoroacetic acid	2.59	2.78
000300-85-6	β-hydroxybutyric acid	4.41	4.33
000305-03-3	chlorambucil	5.75	4.39
000306-08-1	4-OH-3-methoxy- benzeneacetic acid	4.41	4.43
000327-97-9	chlorogenic acid	2.66	3.38
000331-25-9	m-fluorophenylacetic acid	4.13	4.09
000348-10-7	o-fluorophenoxyacetic acid	3.08	3.49
000372-09-8	cyanoacetic acid	2.45	2.59
000404-98-8	m-fluorophenoxyacetic acid	3.13	3.61
000405-79-8	p-fluorophenoxyacetic acid	3.13	3.61
000405-50-5	p-fluorophenylacetic acid	4.24	4.15
000462-60-2	n-(aminocarbonyl)glycine	3.89	2.94
000467-69-6	flurenol	1.09	4.18

000473-81-4	glyceric acid	3.55	3.35
000501-52-0	beta-phenylpropionic acid	4.66	4.40
000503-74-2	isovaleric acid	4.77	4.57
000503-66-2	hydracrylic acid	4.51	3.99
000515-30-0	α-hydroxy-α-methyl benzeneacetic acid	3.53	3.94
000516-05-2	methyl malonic acid	3.12	2.86
000539-35-5	mycobacidin	5.10	4.17
000581-96-4	2-naphthaleneacetic acid	4.25	4.38
000588-32-9	m-chlorophenoxyacetic acid	3.07	2.92
000588-22-7	3,4-dichlorophenoxyacetic acid	2.92	2.65
000594-61-6	α-hydroxy-i-butyric acid	3.61	4.07
000595-46-0	dimethylmalonic acid	3.15	3.22
000595-37-9	2,2-dimethyl butyric acid	5.03	4.66
000598-78-7	2-chloropropionic acid	2.80	3.17
000598-72-1	α-bromopropionic acid	2.97	3.17
000601-75-2	ethylmalonic acid	2.96	3.33
000614-61-9	o-chlorophenoxyacetic acid	3.05	3.55
000616-62-6	propylpropanedioic acid	2.99	3.22
000617-31-2	2-hydroxy- pentanoic acid	3.89**	3.95
000622-47-9	p-methylphenylacetic acid	4.37	4.49
000646-07-1	4-methylpentanoic acid	4.84	4.54
000689-13-4	hadacidin	3.50	3.32
000940-64-7	p-methylphenoxyacetic acid	3.21	3.86
001643-15-8	m-methylphenoxyacetic	3.20	3.08
001759-53-1	cyclopropanecarboxylic acid	4.83	4.63
001798-99-8	m-bromophenoxyacetic acid	3.09	3.57
001798-11-4	p-nitrophenoxyacetic acid	2.89	2.16
001798-06-7	p-iodophenylacetic acid	4.18	4.06
001821-12-1	4-phenylbutyric acid	4.76	4.39
001877-75-4	p-methoxyphenoxyacetic acid	3.21	3.89
001877-73-2	m-nitrophenylacetic acid	3.97	3.56
001878-94-0	p-iodophenoxyacetic acid	3.16	2.74
001878-93-9	m-iodophenoxyacetic acid	3.13	3.57
001878-92-8	o-iodophenoxyacetic acid	3.17	3.42
001878-91-7	p-bromophenoxyacetic acid	3.13	2.69
001878-88-2	m-nitrophenoxyacetic acid	2.95	3.05
001878-87-1	o-nitrophenoxyacetic acid	2.90	3.01
001878-85-9	o-methoxy phenoxyacetic acid	3.23	4.01
001878-82-6	p-cyanophenoxyacetic acid	2.93	2.54
001878-69-9	m-iodophenylacetic acid	4.16	4.12

001878-68-8	p-bromophenylacetic acid	4.19	4.06
001878-66-6	p-chlorophenylacetic acid	4.19	4.12
001878-65-5	m-chlorophenylacetic acid	4.14	4.15
001878-49-5	o-methylphenoxyacetic acid	3.23	3.67
001879-58-9	m-cyanophenoxyacetic acid	3.03	3.41
001879-56-7	o-bromophenoxyacetic acid	3.13	3.52
002088-24-6	m-methoxyphenoxyacetic acid	3.14	3.05
002270-20-4	5-phenylpentanoic acid	4.88	4.43
002976-75-2	(1-naphthalenyloxy)-acetic acid	3.20	3.00
003813-05-6	4-chloro-2-oxo-3(2H)-benzothiazoleacetic acid	3.04	3.50
005292-21-7	cyclohexylacetic acid	4.80	4.57
006324-11-4	o-hydroxyphenoxyacetic acid	3.02	3.58
010502-44-0	2-hydroxy-2-(4-methoxyphenyl)acetic acid	3.42	3.79
014387-10-1	4-ethyl- benzeneacetic acid	4.37	4.49
015307-86-5	diclofenac	4.15	4.16
015687-27-1	ibuprofen	4.45***	4.61
016484-77-8	(R)-2-(4-chloro-o-tolyloxy)propionic acid	3.68	3.53
016563-41-0	3-(1-naphthalenyloxy)-propanoic acid	4.00	3.95
018046-21-4	fentiazac	3.60	4.02
018698-96-9	2-iodo- benzeneacetic acid	4.04	4.24
020225-24-5	2-ethylpentanoic acid	4.71	4.66
022071-15-4	ketoprofen	4.45	4.38
022131-79-9	alcofenac	4.29****	4.30
022204-53-1	naprosyn	4.15	4.52
031879-05-7	fenoprofen	4.50	4.38
032857-63-9	4-(1,1-dimethylethyl)-benzeneacetic acid	4.42	4.52
036330-85-5	fenbufen	4.51	4.42
038194-50-2	sulindac	4.70	4.30
040828-46-4	suprofen	3.91	4.11
040843-25-2	2-[4-(2,4-dichlorophenoxy)phenoxy]propanoic acid	3.43	2.96
053808-88-1	lonazolac	4.30	4.27
055335-06-3	3,5,6-trichloro-2-pyridyloxyacetic acid	2.68*****	2.83
055863-26-8	tiopinac	3.71	4.20
058667-63-3	N-benzoyl-N-(3-chloro-4-fluorophenyl)-DL-alanine	3.72	3.55
069335-91-7	2-[4-[[5-(trifluoromethyl)-2-pyridinyl]oxy]phenoxy]propanoic acid	3.12	3.66
069806-34-4	2-[4-[[3-chloro-5-(trifluoromethyl)-2-pyridinyl]oxy]phenoxy]propanoic acid	2.90	2.90
074103-06-3	ketorolac	3.49	3.84
089894-13-3	phenoxyacetic acid, 4-chloro-3-nitro	2.96	2.77

| 104273-73-6 | 4-(cyclopropylhydroxymethylene)-3,5-dioxocyclohexanecarboxylic acid[++] | 5.32 | 3.79 |
| - | 4-methyl-umbelliferyl β-d-glucuronide | 2.82 | 2.75 |

* Value taken from CRC Handbook of Chemistry and Physics; 81st ed.; Lide, D.R., and Lide, Jr. (Eds.), Chapman and Hall: London, 2001.

** Value from H. C. Brown et al. E.A. Braude, F.C. Nachod (Eds.), Determination of Organic Structures by Physical Methods, Academic Press, N.Y., 1955, 567-662.

*** Value from Avdeef A., Box, K.J., Comer, J.E., Hibbert, C., and Tam, K.Y. *Pharm. Res. 15*, 209-215 (1998); Balon, K., Riebesehl, B.U., and Muller, B.W. *Pharm. Res. 16*, 882-888 (1999).

**** Value from Hansch, C. (Ed.), *Comprehensive Medicinal Chemistry*; Vol. 6.; Pergamon Press: Oxford, 1990.

***** Value from *Martindale - The Extra Pharmacopoeia*; 31st ed.; The Pharmaceutical Press: London, 1996.

[++] value predicted using eq. (17)

Table 7. Anilines

Observed and calculated pK$_a$ values for the aniline dataset using eq. (21).

CAS no.	Chemical name	Exp pK$_a$	Calc pK$_a$
000099-05-8	3-aminobenzoic acid	3.07*	3.76
000150-13-0	4-aminobenzoic acid	2.38	2.45
000065-49-6	p-aminosalicylic acid	2.05	2.02
000123-30-8	4-amino-phenol	5.48	5.72
000591-27-5	3-amino-phenol	4.37	4.85
000618-87-1	3,5-dinitroaniline	0.30	-2.12
000626-43-7	3,5-dichloroaniline	2.51	2.89
000092-67-1	4-aminobiphenyl	4.35	5.18
000611-05-2	3-methyl-4-nitroaniline	1.64	0.39
001137-41-3	4-benzoylaniline	2.24	3.44
034761-82-5	3,5-dimethyl-4-nitrobenzenamine	2.54	0.82
000089-62-3	2-nitro-p-toluidine	0.40	1.91
005470-49-5	4-methylsulfonylaniline	1.35	0.49
000635-22-3	4-chloro-3-nitro-benzenamine	1.90	0.28
000455-14-1	p-trifluoromethylaniline	2.45	2.02
000619-45-4	methyl p-aminobenzoate	2.47	2.89
000094-25-7	butyl-p-aminobenzoate	2.47	3.00
000094-12-2	propyl-p-aminobenzoate	2.49	3.00
000094-09-7	p-aminobenzoic acid, ethyl ester	2.51	3.00
000095-76-1	3,4-dichloroaniline	2.97	3.11

000098-16-8	3-trifluoromethylaniline	3.49	2.78
000591-19-5	m-bromoaniline	3.58	4.20
000626-01-7	3-iodo-benzenamine	3.61	5.94
000540-37-4	4-iodo-benzenamine	3.78	3.76
000106-40-1	p-bromoaniline	3.86	3.87
006933-10-4	3-methyl-4-bromoaniline	4.05	4.20
000118-92-3	2-aminobenzoic acid†	2.14	3.76
000095-55-6	o-aminophenol	4.84	4.52
000099-57-0	2-amino-4-nitrophenol	3.10	2.02
000615-57-6	2,4-dibromoaniline	2.30	3.11
000119-32-4	3-nitro-4-toluidine	3.03	0.71
000090-41-5	2-aminobiphenyl	3.83	5.94
000099-30-9	2,6-dichloro-4-nitroaniline	-2.55	-2.45
016947-63-0	2,6-dimethyl-4-nitrobenzenamine	0.98	0.49
000095-82-9	2,5-dichloroaniline	2.05	3.22
000097-02-9	2,4-dinitroaniline	-4.25	-5.49
000089-63-4	4-chloro-2-nitroaniline	-1.02	-1.03
000121-87-9	2-chloro-4-nitroaniline	-0.94	-1.14
000771-60-8	2,3,4,5,6-pentafluoroaniline	-0.28	-0.49
000608-31-1	2,6-dichloroaniline	0.42	3.44
000096-96-8	4-methoxy-2-nitro-benzenamine	0.77	1.26
000099-52-5	4-nitro-2-toluidine	1.04	0.49
000608-27-5	2,3-dichloroaniline	1.76	3.33
000554-00-7	2,4-dichloroaniline	2.00	3.11
000087-25-2	o-aminobenzoic acid, ethyl ester	2.18	4.42
000134-20-3	methyl anthranilate	2.23	4.20
000099-55-8	5-nitro-2-toluidine	2.35	1.69
013171-61-4	2,3,5,6-tetramethyl-4-nitrobenzenamine	2.36	1.91
000099-59-2	2-methoxy-5-nitroaniline	2.49	2.45
000615-36-1	o-bromoaniline	2.53	4.31
000615-43-0	2-iodoaniline	2.60	4.31
000606-22-4	2,6-dinitroaniline	-5.00	-4.95
000634-93-5	2,4,6-trichloroaniline	-0.03	2.02
000636-30-6	2,4,5-trichloroaniline	1.09	2.02
000102-56-7	2,5-dimethoxyaniline	3.93	5.18

* Macro pK_a (Schulman, S. G., Rosenberg, L. S., and Sturgeon, R. J., *J. Pharm. Sci.* 67, 334-337 (1978)).

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.

Table 8. Primary amines

Observed and calculated pK_a values for the primary amine dataset using eq. (30).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
002954-50-9	2-naphthalenamine, 1,2,3,4-tetrahydro-	9.93	9.86
000064-04-0	2-phenylethylamine	9.96	10.40
000156-34-3	benzeneethanamine, α-methyl-, (R)-	10.13	10.51
000107-10-8	propylamine	10.71	11.06
000109-73-9	n-butylamine	10.78	10.95
000075-04-7	ethylamine	10.87	11.06
000108-91-8	cyclohexanamine	10.63	10.73
000100-46-9	benzylamine	9.33	10.40
000104-84-7	benzenemethanamine, 4-methyl-	9.36	10.51
000064-13-1	p-methoxyamphetamine	9.53	10.51
004764-17-4	3,4-methylenedioxyamphetamine	9.67	11.17
022374-89-6	benzenepropanamine, α-methyl-	9.79	10.30
002038-57-5	3-phenyl propylamine	10.16	10.62
013214-66-9	4-phenylbutylamine	10.36	10.73
000078-81-9	isobutylamine	10.68	11.17
000617-89-0	2-aminomethylfuran	8.89	8.34
000753-90-2	2,2,2-trifluoroethylamine	5.70	4.74
000054-04-6	mescaline	9.56	9.75
000107-11-9	allylamine	9.70	10.51
000513-49-5	sec-butylamine	10.56	10.95
000074-89-5	methylamine	10.62	10.40
000075-31-0	isopropylamine	10.63	10.84
000075-64-9	t-butylamine	10.68	11.17

Table 9. Secondary amines

Observed and calculated pK_a values for the secondary amine dataset using eq. (29)

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000589-08-2	benzeneethanamine, n-methyl-	10.08	9.32
000109-89-7	diethylamine	11.09	10.51
000110-89-4	piperidine	11.28	10.19
000111-92-2	dibutylamine	11.39	10.40
000110-91-8	morpholine	8.49	7.90
000124-02-7	diallylamine	9.29	9.42
000103-67-3	n-methylbenzylamine	9.54	9.21

CAS no.	Chemical name	Exp pK	Calc pK
014321-27-8	benzenemethanamine, n-ethyl-	9.64	9.86
000093-88-9	benzeneethanamine, n,β-dimethyl-	9.87	9.64
000537-46-2	methamphetamine	9.87	9.64
000110-68-9	methylbutylamine	10.90	9.75
000142-84-7	dipropylamine	11.00	10.40
000111-49-9	hexamethyleneimine	11.07	11.28
000108-18-9	diisopropylamine	11.07	11.38
000768-66-1	piperidine, 2,2,6,6-tetramethyl-	11.72*	12.04
000109-05-7	2-methylpiperidine	11.08	10.73
000503-29-7	azetidine	11.29	8.99
000123-75-1	pyrrolidine	11.31	8.99
000494-52-0	anabasine	8.70	8.88
001006-64-0	pyrrolidine, 2-phenyl-	9.40	9.21
000124-40-3	dimethylamine	10.73	9.21
000458-88-8	2-propylpiperidine	11.00	10.84
000101-83-7	dicyclohexylamine	10.40	11.38

* Value from Sosnovsky, G., *Life Sci.* 62, 639-648 (1989).

Table 10. Tertiary amines

Observed and calculated pK_a values for the tertiary amine dataset using eq (35)

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000100-74-3	4-ethylmorpholine	7.67*	7.68
067564-91-4	fenpropimorph	6.98	8.01
000598-56-1	ethyl dimethylamine	10.16	8.88
000927-62-8	dimethylbutylamine	10.19	8.77
000109-02-4	n-methylmorpholine	7.38	7.03
002055-21-2	n,n-dimethyl-3-pyridylmethylamine	8.00	6.81
000102-70-5	2-propen-1-amine, n,n-di-2-propenyl-	8.31	8.34
006304-27-4	2-pyridineethanamine, n,n-dimethyl-	8.75	8.77
013450-66-3	3-pyridylethyl-2-n-piperidine	8.81	8.44
020173-26-6	n,n-dimethyl-2-(3-pyridyl)ethylamine	8.86	7.46
000103-83-3	n,n-dimethylbenzylamine	8.91	8.12
000076-99-3	methadone	8.94	8.66
000132-22-9	chlorpheniramine	9.13	8.55
020173-28-8	3-pyridylethyl-2-(n-pyrrolidine)	9.28	8.44
000626-67-5	n-methylpiperidine	10.08	9.32
000120-94-5	1-methyl-pyrrolidine	10.32	9.10
000121-44-8	triethylamine	10.78	10.40

005470-02-0	1-propylpiperidine	10.41	9.86
000079-55-0	piperidine, 1,2,2,6,6-pentamethyl-	11.25	11.82
000054-11-5	nicotine	8.18**	7.79
000538-07-8	bis(2-chloroethyl)ethylamine	6.57	5.94
000063-75-2	arecoline	7.16	7.03
000091-46-3	ethanamine, n,n-dimethyl-2-[5-methyl-2-(1-methyl)]	8.66	8.01
000054-32-0	moxisylyte	8.72	7.90
003478-94-2	piperalin	8.90	8.66
000083-98-7	orphenadrine	8.91	8.77
000058-73-1	diphenhydramine	8.98	8.77
000075-50-3	trimethylamine	9.80	8.34
067306-00-7	fenpropidin	10.10	9.75
000102-69-2	tripropylamine	10.65	10.51
000102-82-9	tri n-butylamine	10.89	10.40

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

** Value from Chamberlain, K., Evans, A. A. and Bromilow, R. H., *Pestic. Sci. 43*, 167-169 (1995).

Table 11. Meta/para pyridines

Observed and calculated pK$_a$ values for the pyridine (meta/para) dataset using eq. (42).

CAS no.	Chemical name	Exp pK$_a$	Calc pK$_a$
001122-54-9	4-acetylpyridine	3.59	3.24
000626-55-1	3-bromopyridine	2.91	3.57
000614-18-6	nicotinic acid, ethyl ester	3.35	3.53
001570-45-2	isonicotinic acid, ethyl ester	3.45	2.91
001120-87-2	4-bromopyridine	3.78	3.49
000500-22-1	3-formylpyridine	3.80	3.53
000626-61-9	4-chloropyridine	3.84	3.93
000872-85-5	4-formylpyridine	4.77	2.93
000553-26-4	4,4'-dipyridyl	4.82	4.19
000093-60-7	nicotinic acid, methyl ester	3.13	3.41
002459-09-8	i-nicotinic acid, methyl ester	3.26	2.79
002457-47-8	3,5-dichloropyridine	0.67	2.12
000499-81-0	dinicotinic acid†	1.10	1.29
000626-60-8	3-chloropyridine	2.84	3.56
000372-47-4	3-fluoropyridine	2.97	2.99
000350-03-8	3-acetylpyridine	3.18	3.65
001120-90-7	3-iodopyridine	3.25	3.72

000059-67-6	nicotinic acid	2.07*	3.08
000100-55-0	3-pyridinemethanol	4.90	4.49
000586-95-8	4-pyridinemethanol	5.33	5.54
002859-67-8	3-pyridinepropanol	5.47	5.18
000536-78-7	3-ethylpyridine	5.56	5.35
002116-65-6	4-benzylpyridine	5.59	5.63
005344-27-4	4-pyridineethanol	5.60	5.12
000100-43-6	4-vinylpyridine	5.62	5.37
000108-99-6	3-methylpyridine	5.63	5.30
002629-72-3	4-pyridinepropanol	5.84	5.45
000536-75-4	4-ethylpyridine	5.87	5.65
000108-89-4	4-methylpyridine	5.98	5.67
003978-81-2	4-(tert-butyl)pyridine	5.99	5.78
001122-81-2	4-propylpyridine	6.05	5.65
000591-22-0	3,5-dimethylpyridine	6.15	5.42
000583-58-4	3,4-dimethylpyridine	6.46	5.78
000620-08-6	4-methoxypyridine	6.47	6.15
020173-26-6	n,n-dimethyl-2-(3-pyridyl)ethylamine†	4.30	5.23
000494-52-0	anabasine† ⁺⁺	3.21	5.50
000100-48-1	4-cyanopyridine	1.90	2.41
000054-11-5	nicotine† ⁺⁺	3.10	5.86
000059-26-7	nikethamide	3.50	3.89
109151-40-8	(E)-3-nicotinoylacrylic acid†	3.82	3.00
001008-88-4	3-phenylpyridine	4.80	4.95
000055-22-1	isonicotinic acid	1.70*	2.44
000110-86-1	pyridine	5.23	5.18
000939-23-1	4-phenylpyridine	5.55	5.28
003731-52-0	3-pyridinemethaneamine†	5.96	5.24
002530-26-9	3-nitropyridine	1.18	0.71
000109-00-2	3-hydroxypyridine	4.80**	3.79
007295-76-3	3-methyoxypyridine	4.91	4.31

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.

⁺⁺ pK_a predicted using equation (41)

* Macro pK_a (Halle, J.-C., Lelievre, J., and Terrior, F., *Can. J. Chem.* 74, 613-616 (1996)).

** Macro pK_a (Metzler, D.E., and Snell, E. E., *J. Am. Chem. Soc.* 77, 2431-2437 (1955)).

Table 12. Ortho pyridines

Observed and calculated pK_a values for the pyridine (ortho) dataset using eq. (44).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000104-90-5	5-ethyl-2-methylpyridine	6.51	5.99
003748-84-3	2,3,5,6-tetramethylpyridine	7.90	7.59
000372-48-5	2-fluoropyridine	-0.44	1.11
000109-09-1	2-chloropyridine	0.49	2.14
000109-04-6	2-bromopyridine	0.90	2.20
000100-69-6	2-vinylpyridine	4.98	5.64
000101-82-6	2-benzylpyridine	5.13	5.14
000103-74-2	2-pyridineethanol	5.31	4.45
002859-68-9	2-pyridinepropanol	5.61	5.28
000140-76-1	2-methyl-5-vinylpyridine	5.67	5.26
005944-41-2	2-t-butylpyridine	5.76	6.07
000100-71-0	2-ethylpyridine	5.8°	5.69
000589-93-5	2,5-dimethylpyridine	6.40	5.96
000583-61-9	2,3-dimethylpyridine	6.57	6.02
000108-48-5	2,6-lutidine	6.60	6.89
000108-47-4	2,4-dimethylpyridine	6.99	6.31
000108-75-8	2,4,6-collidine	7.43	7.57
002402-78-0	2,6-dichloropyridine	-2.86	0.27
002176-62-7	2,3,4,5,6-pentachlorpyridine	-1.00	-3.60
002402-77-9	2,3-dichloropyridine	-0.85	0.35
002402-79-1	2,3,5,6-tetrachlorpyridine	-0.80	-2.71
001008-89-5	2-phenylpyridine	4.48	5.22
000109-06-8	2-methylpyridine	6.00	5.62
000366-18-7	2,2-bipyridine	4.33	4.35
C00586-98-1	2-pyridinemethanol	4.86	3.84
001603-41-4	2-amino-5-methylpyridine	7.22	5.98
018438-38-5	2-methylthiopyridine	3.59	4.30
000142-08-5	2-hydroxypyridine	0.75*	1.52
001628-89-3	2-methoxypyridine	3.06	2.98
006231-18-1	pyridine, 2,6-dimethoxy-	1.60	1.31
001122-62-9	2-acetylpyridine	2.73	1.92
001121-60-4	2-pyridinecarboxyaldehyde	3.80	3.22
002459-07-6	picolinic acid, methyl ester	2.21	2.32
000098-98-6	picolinic acid	1.06**	1.67

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

Table 13. Pyrimidines

Observed and calculated pK$_a$ values for the pyrimidine dataset using eq. (47)

CAS no.	Chemical name	Exp pK$_a$	Calc pK$_a$
004595-60-2	pyrimidine, 2-bromo-	-1.63	0.59
034253-03-7	2-pyrimidinecarboxylic acid, methyl ester	-0.68	0.22
000823-09-6	pyrimidine, 2-(methylthio)-	0.59	2.20
014080-32-1	pyrimidine, 5-nitro-	0.72	-0.88
000289-95-2	pyrimidine	1.23	1.50
003739-82-0	pyrimidine, 2-ethoxy-	1.27	2.37
003438-46-8	4-methylpyrimidine	1.91	1.58
014001-64-0	2-SMe-4,6-dimethylpyrimidine	2.12	2.98
001558-17-4	4,6-dimethylpyrimidine	2.70	2.38
005469-70-5	2-aminopyrimidine	3.45	3.31
053112-28-0	pyrimethanil	3.52	3.63
003740-92-9	fenclorim**	4.23	1.02
000767-15-7	2-amino-4,6-dimethylpyrimidine	4.82	4.05
001004-38-2	2,4,6-pyrimidinetriamine	6.81	6.10

** using equation (46)

Table 14. Imidazole and benzimidazoles

Observed and calculated pK$_a$ values for the imidazole and benzimidazole dataset using eq. (52).

CAS no.	Chemical name	Exp pK$_a$	Calc pK$_a$
Imidazoles			
003034-41-1	1-methyl-4-nitro-1H-imidazole	-0.53	0.38
003034-38-6	4-nitroimidazole	-0.05	-0.05
003034-42-2	1-methyl-5-nitroimidazole	2.13	1.82
002466-76-4	1H-imidazole, 1-acetyl-	3.60	3.95
068694-11-1	triflumizole	3.70	4.59
000670-96-2	1H-imidazole, 2-phenyl-	6.48	6.77
000616-47-7	1H-imidazole, 1-methyl-	6.95	7.01
000288-32-4	imidazole	6.95	6.66
000693-98-1	1H-imidazole, 2-methyl-	7.85	7.39

CAS no.	Chemical name	Exp pK	Calc pK
000584-85-0	anserine†	7.04	6.40
067747-09-5	prochloraz	3.80	5.92
053910-25-1	pentostatin	5.20	6.00
004238-71-5	1H-imidazole, 1-(phenylmethyl)-	6.70	7.15
000092-13-7	pilocarpol	6.78	5.37
051481-61-9	cimetidine	6.80	6.42
035554-44-0	imazalil base	6.53	7.41
Benzimidazoles			
000051-17-2	benzimidazole	5.53*	6.00
117994-51-1	2-(4-methoxyphenylmethyl)-5-nitrobenzimidazole	4.26	3.55
115583-15-8	2-(4-chlorphenylmethyl)-5-chlorobenzimidazole	4.86	5.54
001571-92-2	2-(2-methylphenyl)-5-nitrobenzimidazole	4.87	4.13
163192-67-4	2-(2,4-dimethylphenyl)-5-nitrobenzimidazole	5.29	4.16
033138-16-8	2-(4-bromophenylmethyl)-5-chlorobenzimidazole	5.42	5.43
000120-03-6	2-(4-methylphenyl)-benzimidazole	6.90	6.80
111526-85-3	2-(4-methylphenylmethyl)-5-chlorobenzimidazole	7.09	6.21
006528-85-4	2-(2-methoxyphenyl)-benzimidazole	7.17	7.28
119691-79-1	2-(4-aminophenylmethyl)-5-chlorobenzimidazole	7.47	6.50

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

Table 15. Quinolines

Observed and calculated pK_a values for the quinoline dataset using eq. (53).

CAS no.	Chemical name	Exp pK_a	Calc pK_a
000091-22-5	quinoline	4.90	5.20
000578-67-6	5-quinolinol	5.02	4.57
004964-76-5	7-methoxyquinoline	5.03	5.07
005263-87-6	6-methoxyquinoline	5.03	5.12
000611-32-5	8-methylquinoline	5.05	5.30
000612-58-8	3-methylquinoline	5.17	5.41
000091-62-3	6-methylquinoline	5.34	5.40
000612-60-2	7-methylquinoline	5.34	5.41
000580-16-5	6-hydroxyquinoline	5.15*	4.77
000580-20-1	7-quinolinol	5.46	4.76
000491-35-0	4-methylquinoline	5.67	5.23
000091-63-4	2-methylquinoline	5.71	5.50

005332-24-1	3-bromoquinoline	2.69	3.54
000086-98-6	4,7-dichloro-quinoline	2.80	2.49
000611-33-6	8-chloroquinoline	3.12	3.89
000394-68-3	8-fluoroquinoline	3.34	3.84
000612-57-7	6-chloroquinoline	3.85	3.85
004965-36-0	7-bromoquinoline	3.87	3.63
005332-25-2	6-bromoquinoline	3.87	3.59
000877-43-0	2,6-dimethylquinoline	6.10	5.67
001198-37-4	2,4-dimethylquinoline	5.12	5.53
000580-18-7	3-quinolinol	4.28	4.80
000148-24-3	8-quinolinol	4.90	4.84
000130-16-5	8-quinolinol,5-chloro-	3.56	3.54
000938-33-0	8-methoxyquinoline	5.01	5.69
000826-81-3	2-methyl 8-quinolonol	5.55	5.16
003846-73-9	4-methyl-8-quinolinol	5.56	4.87
099607-70-2	cloquintocet-mexyl	3.75	3.56

* Value taken from Albert, A., Serjeant, E.P. *The Determination of Ionization Constants: A Laboratory Manual*; 3rd ed.; Chapman and Hall: London, 1984.

1.4 BIBLIOGRAPHY

(1)   Carlson, N. R. and Buskist, W. The Nature and Causes of Mental Disorders. *Psychology: The Science of Behaviour 5th ed*; Allyn and Bacon: Needham Heights, MA, 1997; pp 585.

(2)   Baldessarini, R. J. Drugs and the Treatment of Psychiatric Disorders. *Goodman & Gilmans The Pharmacological Basis of Therapeutics*; McGraw-Hill, 1996; pp 399-430.

(3)   Barondes, S. H. Molecules and mental illness; Scientific American Library: New York, 1993; pp 150.

(4)   *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders.*; 4th edition ed.; American Psychiatric Association: Washington DC, 1994; 273.

(5)   Meltzer, H. Y. Treatment of suicidality in schizophrenia. *Annals of the New York Academy of Sciences* **2001**, *932*, 44-60.

(6)   Mortensen, P. B.; Pedersen, C. B.; Westergaard, T.; Wohlfahrt, J.; Ewald, H.; Mors, O.; Andersen, P. K. and Melbye, M. Effects of Family History and Place and Season of Birth on the Risk of Schizophrenia. *The New England Journal of Medicine* **1999**, *340*, 603-608.

(7)   Schmidt, C. J.; Sorensen, S. M.; Kehne, J. H.; Carr, A. A. and Palfreyman, M. G. The role of 5-HT2A receptors in antipsychotic activity. *Life Sciences* **1995**, *56*, 2209-2222.

(8)   Stahl, S. Psychosis and Schizophrenia. *Essential Psychopharmacology: Neuroscientific Basis and Practical Applications*; Cambridge University Press: Cambridge, 1996; pp 254.

(9)   Jaber, M.; Robinson, S. W.; Missale, C. and Caron, M. G. Dopamine receptors and brain function. *Neuropharmacol.* **1996**, *35*, 1503-1519.

(10)   Khan, Z. U.; Mrzljak, L.; Gutierrez, A.; de la Calle, A. and Goldman-Rakic, P. S. Prominence of the dopamine D2 short isoform in dopaminergic pathways. *PNAS* **1998**, *95*, 7731-7736.

(11)   Carlsson, A. and Lindqvist, M. Effect of chlorpromazine or haloperidol on formation of 3-methoxytyramine and normetanephrine in mouse brain. *Acta. Pharmacol. Toxicol.* **1963**, *20*, 140-144.

(12)   Snyder, S. H. Amphetamine psychosis: a "model" schizophrenia mediated by catecholamines. *American Journal of Psychiatry* **1973**, *130*, 61-67.

(13)   Seeman, P.; Lee, T.; Chau-Wong, M. and Wong, K. Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature* **1976**, *261*, 717-719.

(14)   Pycock, C. J.; Kerwin, R. W. and Carter, C. J. Effect of lesion of cortical dopamine terminals on subcortical dopamine receptors in rats. *Nature* **1980**, *286*, 74-76.

(15)   Van Tol, H. H.; Bunzow, J. R.; Guan, H. C.; Sunahara, R. K.; Seeman, P.; Niznik, H. B. and Civelli, O. Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. *Nature* **1991**, *350*, 610-614.

(16)   Seeman, P.; Guan, H. C. and Van Tol, H. H. Dopamine D4 receptors elevated in schizophrenia. *Nature* **1993**, *365*, 441-445.

(17)   Reynolds, G. P. and Mason, S. L. Absence of detectable striatal dopamine D4 receptors in drug-treated schizophrenia. *European Journal of Pharmacology* **1995**, *281*, R5-6.

(18) Stefanis, N. C.; Bresnick, J. N.; Kerwin, R. W.; Schofield, W. N. and McAllister, G. Elevation of D4 dopamine receptor mRNA in postmortem schizophrenic brain. *Brain Res. Mol. Brain Res.* **1998,** *53,* 112-119.

(19) Ohmori, J.; Maeno, K.; Hidaka, K.; Nakato, K.; Matsumoto, M.; Tada, S.; Hattori, H.; Sakamoto, S.; Tsukamoto, S.; Usuda, S. and Mase, T. Dopamine D_3 and D_4 receptor antagonists: synthesis and structure--activity relationships of (S)-(+)-N-(1-Benzyl-3-pyrrolidinyl)-5-chloro-4- [(cyclopropylcarbonyl) amino]-2-methoxybenzamide (YM-43611) and related compounds. *J. Med. Chem.* **1996,** *39,* 2764-2772.

(20) Moore, K. W.; Bonner, K.; Jones, E. A.; Emms, F.; Leeson, P. D.; Marwood, R.; Patel, S.; Rowley, M.; Thomas, S. and Carling, R. W. 4-N-linked-heterocyclic piperidine derivatives with high affinity and selectivity for human dopamine D_4 receptors. *Bioorg. Med. Chem. Lett.* **1999,** *9,* 1285-1290.

(21) Rowley, M.; Collins, I.; Broughton, H. B.; Davey, W. B.; Baker, R.; Emms, F.; Marwood, R.; Patel, S.; Ragan, C. I.; Freedman, S. B.; Ball, R. and Leeson, P. D. 4-Heterocyclylpiperidines as selective high-affinity ligands at the human dopamine D_4 receptor. *J. Med. Chem.* **1997,** *40,* 2374-2385.

(22) Rowley, M.; Broughton, H. B.; Collins, I.; Baker, R.; Emms, F.; Marwood, R.; Patel, S.; Ragan, C. I.; Freedman, S. B. and Leeson, P. D. 5-(4-Chlorophenyl)-4-methyl-3-(1-(2-phenylethyl)piperidin-4-yl)isoxazole: a potent, selective antagonist at human cloned dopamine D_4 receptors. *J. Med. Chem.* **1996,** *39,* 1943-1945.

(23) Carling, R. W.; Moore, K. W.; Moyes, C. R.; Jones, E. A.; Bonner, K.; Emms, F.; Marwood, R.; Patel, S.; Fletcher, A. E.; Beer, M.; Sohal, B.; Pike, A. and Leeson, P. D. 1-(3-Cyanobenzylpiperidin-4-yl)-5-methyl-4-phenyl-1, 3-dihydroimidazol-2-one: a selective high-affinity antagonist for the human dopamine D_4 receptor with excellent selectivity over ion channels. *J. Med. Chem.* **1999,** *42,* 2706-2715.

(24) Belliotti, T. R.; Wustrow, D. J.; Brink, W. A.; Zoski, K. T.; Shih, Y. H.; Whetzel, S. Z.; Georgic, L. M.; Corbin, A. E.; Akunne, H. C.; Heffner, T. G.; Pugsley, T. A. and Wise, L. D. A series of 6-and 7-piperazinyl- and -piperidinylmethylbenzoxazinones with dopamine D_4 antagonist activity: Discovery of a potential atypical antipsychotic agent. *J. Med. Chem.* **1999,** *42,* 5181-5187.

(25) Boyfield, I.; Brown, T. H.; Coldwell, M. C.; Cooper, D. G.; Hadley, M. S.; Hagan, J. J.; Healy, M. A.; Johns, A.; King, R. J.; Middlemiss, D. N.; Nash, D. J.; Riley, G. J.; Scott, E. E.; Smith, S. A. and Stemp, G. Design and Synthesis of 2-Naphthoate Esters as Selective Dopamine D_4 Antagonists. *J. Med. Chem.* **1996,** *39,* 1946-1948.

(26) Bristow, L. J.; Collinson, N.; Cook, G. P.; Curtis, N.; Freedman, S. B.; Kulagowski, J. J.; Leeson, P. D.; Patel, S.; Ragan, C. I.; Ridgill, M.; Saywell, K. L. and Tricklebank, M. D. L-745,870, a subtype selective dopamine D_4 receptor antagonist, does not exhibit a neuroleptic-like profile in rodent behavioral tests. *J. Pharmacol. Exp. Ther.* **1997,** *283,* 1256-1263.

(27) Kesten, S. R.; Heffner, T. G.; Johnson, S. J.; Pugsley, T. A.; Wright, J. L. and Wise, L. D. Design, synthesis, and evaluation of chromen-2-ones as potent and selective human dopamine D_4 antagonists. *J. Med. Chem.* **1999,** *42,* 3718-3725.

(28) Kulagowski, J. J.; Broughton, H. B.; Curtis, N. R.; Mawer, I. M.; Ridgill, M. P.; Baker, R.; Emms, F.; Freedman, S. B.; Marwood, R.; Patel, S.; Ragan, C. I. and Leeson, P. D. 3-((4-(4-Chlorophenyl)piperazin-1-yl)-methyl)-1H-pyrrolo-2,3-b-

pyridine: an antagonist with high affinity and selectivity for the human
dopamine D_4 receptor. *J. Med. Chem.* **1996**, *39*, 1941-1942.

(29)   Tenbrink, R. E.; Bergh, C. L.; Duncan, J. N.; Harris, D. W.; Huff, R. M.; Lahti,
        R. A.; Lawson, C. F.; Lutzke, B. S.; Martin, I. J.; Rees, S. A.; Schlachter, S. K.;
        Sih, J. C. and Smith, M. W. (S)-(-)-4-[4-[2-(Isochroman-1-Yl)Ethyl]-Piperazin-
        1-Yl]Benzenesulfonamide , a Selective Dopamine D_4 Antagonist. *J. Med. Chem.*
        1996, *39*, 2435-2437.

(30)   Thurkauf, A.; Yuan, J.; Chen, X.; He, X. S.; Wasley, J. W. F.; Hutchison, A.;
        Woodruff, K. H.; Meade, R.; Hoffman, D. C.; Donovan, H. and Joneshertzog, D.
        K. 2-Phenyl-4(5)-[[4-(Pyrimidin-2-Yl)Piperazin-1-Yl]Methyl]Imidazole - a
        Highly Selective Antagonist at Cloned Human D_4 Receptors. *J. Med. Chem.*
        1997, *40*, 1-3.

(31)   Unangst, P. C.; Capiris, T.; Connor, D. T.; Heffner, T. G.; MacKenzie, R. G.;
        Miller, S. R.; Pugsley, T. A. and Wise, L. D. Chromeno[3,4-c]pyridin-5-ones:
        selective human dopamine D_4 receptor antagonists as potential antipsychotic
        agents. *J. Med. Chem.* **1997**, *40*, 2688-2693.

(32)   Sanner, M. A.; Chappie, T. A.; Dunaiskis, A. R.; Fliri, A. F.; Desai, K. A.; Zorn,
        S. H.; Jackson, E. R.; Johnson, C. G.; Morrone, J. M.; Seymour, P. A.;
        Majchrzak, M. J.; Faraci, W. S.; Collins, J. L.; Duignan, D. B.; Prete Di, C. C.;
        Lee, J. S. and Trozzi, A. Synthesis, SAR and pharmacology of CP-293,019: a
        potent, selective dopamine D_4 receptor antagonist. *Bioorg. Med. Chem. Lett.*
        1998, *8*, 725-730.

(33)   Gazi, L.; Bobirnac, I.; Danzeisen, M.; Schupbach, E.; Langenegger, D.;
        Sommer, B.; Hoyer, D.; Tricklebank, M. and Schoeffter, P. Receptor density as
        a factor governing the efficacy of the dopamine D4 receptor ligands, L-745,870
        and U-101958 at human recombinant D4.4 receptors expressed in CHO cells.
        British Journal of Pharmacology **1999**, *128*, 613-620.

(34)   Truffinet, P.; Tamminga, C. A.; Fabre, L. F.; Meltzer, H. Y.; Riviere, M. E. and
        Papillon-Downey, C. Placebo-controlled study of the D4/5-HT2A antagonist
        fananserin in the treatment of schizophrenia. *American Journal of Psychiatry*
        1999, *156*, 419-425.

(35)   Malleron, J. L.; Comte, M. T.; Gueremy, C.; Peyronel, J. F.; Truchon, A.;
        Blanchard, J. C.; Doble, A.; Piot, O.; Zundel, J. L.; Huon, C.; Martin, B.;
        Mouton, P.; Viroulaud, A.; Allam, D. and Betschart, J. Naphthosultam
        derivatives: a new class of potent and selective 5-HT2 antagonists. *Journal of
        Medicinal Chemistry* **1991**, *34*, 2477-2483.

(36)   Bailey, P. E.; Luthringer, R.; Greger, G. and Macher, J. P. A Double-Blind,
        Placebo-Controlled Study of the Tolerability (Especially Cardiovascular) and
        Efficacy of Belaperidone, a Dopamine D4 Receptor Antagonist, in Twelve
        Hospitalized Schizophrenic Patients. *40th Annual NCDEU Meeting*: Boca
        Raton, Florida, 2000.

(37)   Capuano, B.; Crosby, I. T. and Lloyd, E. J. Schizophrenia: genesis,
        receptorology and current therapeutics. *Current Medicinal Chemistry* **2002**, *9*,
        521-548.

(38)   Lopez-Rodriguez, M. L.; Benhamu, B.; Morcillo, M. J.; Murcia, M.; Viso, A.;
        Campillo, M. and Pardo, L. 5-HT4 receptor antagonists: structure-affinity
        relationships and ligand-receptor interactions. *Current Topics in Medicinal
        Chemistry (Hilversum, Netherlands)* **2002**, *2*, 625-641.

(39)   Riemer, C.; Borroni, E.; Levet-Trafit, B.; Martin, J. R.; Poli, S.; Porter, R. H. P.
        and Boes, M. Influence of the 5-HT6 Receptor on Acetylcholine Release in the

Cortex: Pharmacological Characterization of 4-(2-Bromo-6-pyrrolidin-1-ylpyridine-4-sulfonyl)phenylamine, a Potent and Selective 5-HT6 Receptor Antagonist. *Journal of Medicinal Chemistry* **2003**, *46*, 1273-1276.

(40) Pouzet, B. SB-258741: a 5-HT7 receptor antagonist of potential clinical interest. *CNS Drug Reviews* **2002**, *8*, 90-100.

(41) Abi-Dargham, A.; Laruelle, M.; Aghajanian, G. K.; Charney, D. and Krystal, J. The role of serotonin in the pathophysiology and treatment of schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences* **1997**, *9*, 1-17.

(42) Duinkerke, S. J.; Botter, P. A.; Jansen, A. A.; van Dongen, P. A.; van Haaften, A. J.; Boom, A. J.; van Laarhoven, J. H. and Busard, H. L. Ritanserin, a selective 5-HT2/1C antagonist, and negative symptoms in schizophrenia. A placebo-controlled double-blind trial. *British Journal of Psychiatry* **1993**, *163*, 451-455.

(43) Williams, J.; Spurlock, G.; McGuffin, P.; Mallet, J.; Nothen, M. M.; Gill, M.; Aschauer, H.; Nylander, P. O.; Macciardi, F. and Owen, M. J. Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. European Multicentre Association Study of Schizophrenia (EMASS) Group. *Lancet* **1996**, *347*, 1294-1296.

(44) Meltzer, H. Y.; Matsubara, S. and Lee, J. C. Classification of typical and atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin2 pKi values. *Journal of Pharmacology & Experimental Therapeutics* **1989**, *251*, 238-246.

(45) Carlsson, M. and Svensson, A. Interfering with glutamatergic neurotransmission by means of NMDA antagonist administration discloses the locomotor stimulatory potential of other transmitter systems. *Pharmacology, Biochemistry and Behavior* **1990**, *36*, 45-50.

(46) Toru, M.; Kurumaji, A. and Ishimaru, M. Excitatory amino acids: implications for psychiatric disorders research. *Life Sciences* **1994**, *55*, 1683-1699.

(47) Singh, M. M. and Kay, S. R. Therapeutic antagonism between anticholinergic antiparkinsonism agents and neuroleptics in schizophrenia. Implications for a neuropharmacological model. *Neuropsychobiology* **1979**, *5*, 74-86.

(48) Tandon, R. and Greden, J. F. Cholinergic hyperactivity and negative schizophrenic symptoms. A model of cholinergic/dopaminergic interactions in schizophrenia. *Archives of General Psychiatry* **1989**, *46*, 745-753.

(49) Gomes, U. C.; Shanley, B. C.; Potgieter, L. and Roux, J. T. Noradrenergic overactivity in chronic schizophrenia: evidence based on cerebrospinal fluid noradrenaline and cyclic nucleotide concentrations. *British Journal of Psychiatry* **1980**, *137*, 346-351.

(50) Fischel, T.; Hermesh, H.; Aizenberg, D.; Zemishlany, Z.; Munitz, H.; Benjamini, Y. and Weizman, A. Cyproheptadine versus propranolol for the treatment of acute neuroleptic-induced akathisia: a comparative double-blind study. *Journal of Clinical Psychopharmacology* **2001**, *21*, 612-615.

(51) Maas, J. W.; Miller, A. L.; Tekell, J. L.; Funderburg, L.; Silva, J. A.; True, J.; Velligan, D.; Berman, N. and Bowden, C. L. Clonidine plus haloperidol in the treatment of schizophrenia/psychosis. *Journal of Clinical Psychopharmacology* **1995**, *15*, 361-364.

(52) Lewis, D. A.; Pierri, J. N.; Volk, D. W.; Melchitzky, D. S. and TU., W. Altered GABA neurotransmission and prefrontal cortical dysfunction in schizophrenia. *Biological Psychiatry* **1999**, *46*, 616-626.

(53)   Waszczak, B. L. and Walters, J. R. Effects of GABAergic drugs on single unit activity of A9 and A10 dopamine neurons. *Brain Research Bulletin* **1980**, *5*, 465-470.

(54)   Fielding, S. and Lal, H. Behavioural aspects of neuroleptics. *Handbook of Psychopharmacology*; Plenum, 1978; pp 91-128.

(55)   Niemegeers, C. J. E.; Verbruggen, F. J. and Janssen, P. A. J. Influence of various neuroleptic drugs on shock avoidance responding in rats. I. Nondiscriminated Sidman avoidance procedure. *Psychopharmacologia* **1969**, *16*, 161-174.

(56)   Niemegeers, C. J. E.; Verbruggen, F. J. and Janssen, P. A. J. Influence of various neuroleptic drugs on shock avoidance responding in rats. II. Nondiscriminated Sidman avoidance procedure with alternate reinforcement and extinction periods and analysis of the interresponse times (IRT's). *Psychopharmacologia* **1969**, *16*, 175-182.

(57)   Niemegeers, C. J. E.; Verbruggen, F. J. and Janssen, P. A. J. Influence of various neuroleptic drugs on shock avoidance response in rats. III. Amphetamine antagonism in the discriminated Sidman avoidance procedure. *Psychopharmacologia* **1970**, *17*, 151-159.

(58)   Waddington, J. L. and O'Callaghan, E. O. What Makes an Antipsychotic 'Atypical'? Conserving the Definition. *CNS Drugs* **1997**, *7*, 341-346.

(59)   Meltzer, H. Y. Serotonin Receptors and Antipsychotic Drug Action. *Psychopharmacology Series: Stratergies in Psychotropic Drug Development*; Springer, 1993; pp 70-81.

(60)   Minchin, S. A. and Csernansky, J. G. Classification schemes for Antipsychotic Drugs. *Antipsychotics*; Springer, 1996; pp 1-21.

(61)   Nadal, R. Pharmacology of the atypical antipsychotic remoxipride, a dopamine D2 receptor antagonist. *CNS Drug Reviews* **2001**, *7*, 265-282.

(62)   Ashby, C. R., Jr. and Wang, R. Y. Pharmacological Action of the Atypical Antipsychotic Drug Clozapine: A Review. *Synapse* **1996**, *24*, 349-394.

(63)   Leucht, S.; Pitschel-Walz, G.; Abraham, D. and Kissling, W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophrenia Research* **1999**, *35*, 51-68.

(64)   Oshiro, Y.; Sato, S.; Kurahashi, N.; Tanaka, T.; Kikuchi, T.; Tottori, K.; Uwahodo, Y. and Nishi, T. Novel antipsychotic agents with dopamine autoreceptor agonist properties: synthesis and pharmacology of 7-[4-(4-phenyl-1-piperazinyl)butoxy]-3,4-dihydro-2(1H)-quinolinone derivatives. *Journal of Medicinal Chemistry* **1998**, *41*, 658-667.

(65)   Altar, A. A.; Martin, A. R. and Thurkauf, A. Antipsychotic Agents. *Burger's Medicinal Chemistry and Drug Discovery, Volume 6: Nervous System Agents*; John Wiley & Sons, 2003; pp 1-48.

(66)   Janssen, P. A.; Niemegeers, C. J. and Schellekens, K. H. Is it possible to predict the clinical effects of neuroleptic drugs (major tranquillizers) from animal data? *Arzneimittel-Forschung* **1966**, *16*, 339-346.

(67)   Janssen, P. A.; Niemegeers, C. J.; Schellekens, K. H. and Lenaerts, F. M. Is it possible to predict the clinical effects of neuroleptic drugs (major tranquillizers) from animal data? IV. An improved experimental design for measuring the inhibitory effects of neuroleptic drugs on amphetamine-or apomorphine-induced "Cheroing" and "agitation" in rats. *Arzneimittel-Forschung* **1967**, *17*, 841-854.

(68)   Janssen, P. A. and Awouters, F. H. Is it possible to predict the clinical effects of
        neuroleptics from animal data? Part V: From haloperidol and pipamperone to
        risperidone. *Arzneimittel-Forschung* **1994**, *44*, 269-277.

(69)   Gschwend, H. W. Chemical approaches to the development of neuroleptics.
        Industrial Pharmacology **1974**, *1*, 1-51.

(70)   Burki, H. R.; Sayers, A. C.; Ruch, W. and Asper, H. Effects of clozapine and
        other dibenzo-epines on central dopaminergic and cholinergic systems.
        Structure-activity relationships. *Arzneimittel-Forschung* **1977**, *27*, 1561-1565.

(71)   Phillips, S. T.; de Paulis, T.; Baron, B. M.; Siegel, B. W.; Seeman, P.; Van Tol,
        H. H.; Guan, H. C. and Smith, H. E. Binding of 5H-dibenzo[b,e][1,4]diazepine
        and chiral 5H-dibenzo[a,d]cycloheptene analogues of clozapine to dopamine and
        serotonin receptors. *Journal of Medicinal Chemistry* **1994**, *37*, 2686-2696.

(72)   Phillips, S. T.; de Paulis, T.; Neergaard, J. R.; Baron, B. M.; Siegel, B. W.;
        Seeman, P.; Van Tol, H. H.; Guan, H. C. and Smith, H. E. Binding of 5H-
        dibenzo[a,d]cycloheptene and dibenz[b,f]oxepin analogues of clozapine to
        dopamine and serotonin receptors. *Journal of Medicinal Chemistry* **1995**, *38*,
        708-714.

(73)   Schmutz, J. Neuroleptic piperazinyl-dibenzo-azepines. Chemistry and structure-
        activity relationships. *Arzneimittel-Forschung* **1975**, *25*, 712-720.

(74)   Rognan, D.; Sokoloff, P.; Mann, A.; Martres, M. P.; Schwartz, J. C.; Costentin,
        J. and Wermuth, C. G. Optically active benzamides as predictive tools for
        mapping the dopamine D2 receptor. *European Journal of Pharmacology,
        Molecular Pharmacology Section* **1990**, *189*, 59-70.

(75)   Boström, J.; Gundertofte, K. and Liljefors, T. A pharmacophore model for
        dopamine D-4 receptor antagonists. *Journal of Computer-Aided Molecular
        Design* **2000**, *14*, 769-786.

(76)   Marshall, G. R. Computer-aided drug design: the active analog approach. *Med.
        Chem., Proc. Int. Symp., 6th*, 1979 (Meeting Date 1978).

(77)   Liljefors, T. and Pettersson, I. Computer Aided Development of Three-
        Dimensional Pharmacophore Models. *A Textbook of Drug Design and
        Development*; Harwood Academic Publishers, 1997; pp 61-93.

(78)   Clark, M.; Cramer III, R. D.; Jones, D. M.; Patterson, D. E. and Simeroth, P. E.
        Comparative Molecular Field Analysis (CoMFA). Part 2. Towards its use with
        3D-structural databases. *Tetrahedron Computational Methodology* **1990**, *3*, 47-
        59.

(79)   Tripos *Sybyl*; version 6.8 ed.: 1699 South Hanley Road, St. Louis, MO 63144,
        USA.

(80)   Hilgeroth, A.; Fleischer, R.; Wiese, M. and Heinemann, F. W. Comparison of
        azacyclic urea A-98881 as HIV-1 protease inhibitor with cage dimeric N-benzyl
        4-(4-methoxyphenyl)-1,4-dihydropyridine as representative of a novel class of
        HIV-1 protease inhibitors: A molecular modeling study. *Journal of Computer-
        Aided Molecular Design* **1999**, *13*, 233-242.

(81)   Kulkarni, S. S.; Gediya, L. K. and Kulkarni, V. M. Three-dimensional
        quantitative structure activity relationships (3-D-QSAR) of antihyperglycemic
        agents. *Bioorganic & Medicinal Chemistry* **1999**, *7*, 1475-1485.

(82)   Kulkarni, S. S. and Kulkarni, V. M. Three-dimensional quantitative structure-
        activity relationship of interleukin 1-beta converting enzyme inhibitors: A
        comparative molecular field analysis study. *Journal of Medicinal Chemistry*
        1999, *42*, 373-380.

(83)   Tehan, B. G.; Lloyd, E. J. and Wong, M. G. Molecular Field Analysis of
         Clozapine Analogues in the Development of a Pharmacophore Model of
         Antipsychotic Drug Action. *Journal of Molecular Graphics and Modelling*
         2001, *19*, 417-426.

(84)   Lemmen, C. and Lengauer, T. Computational methods for the structural
         alignment of molecules. *Journal of Computer-Aided Molecular Design* **2000**,
         14, 215-232.

(85)   Jones, G.; Willett, P. and Glen, R. C. A Genetic Algorithm for Flexible
         Molecular Overlay and Pharmacophore Elucidation. *Journal of Computer-Aided
         Molecular Design* **1995**, *9*, 532-549.

(86)   Jones, G.; Willett, P. and Glen, R. C. GASP: Genetic Algorithm
         Superimposition Program. *Pharmacophore Perception, Development, and Use
         in Drug Design* **2000**, 85-107.

(87)   Tehan, B. G.; Lloyd, E. J. and Wong, M. G. Atypical Antipsychotics:
         Modelling, QSAR and Database Searching. *Rational Approaches to Drug
         Design: 13th European Symposium on QSAR*; Prous Science: Barcelona, 2001;
         pp 345-348.

(88)   Accelrys, 2001.

(89)   Kearsley, S. S. and Smith, G. *Tetrahedron Computational Methods* **1992**, *3*, 615.

(90)   Masek, B. B.; Merchant, A. and Matthew, J. B. Molecular shape comparison of
         angiotensin II receptor antagonists. *Journal of Medicinal Chemistry* **1993**, *36*,
         1230-1238.

(91)   Jewell, N. E.; Turner, D. B.; Willett, P. and Sexton, G. J. Automatic generation
         of alignments for 3D QSAR analyses. *Journal of Molecular Graphics and
         Modelling* **2001**, *20*, 111-121.

(92)   Thorner, D. A.; Wild, D. J.; Willett, P. and Wright, P. M. Similarity Searching in
         Files of Three-Dimensional Chemical Structures - Flexible Field-Based
         Searching of Molecular Electrostatic Potentials. *Journal of Chemical
         Information & Computer Sciences* **1996**, *36*, 900-908.

(93)   Pitman, M. C.; Huber, W. K.; Horn, H.; Kramer, A.; Rice, J. E. and Swope, W.
         C. FLASHFLOOD: A 3D field-based similarity search and alignment method
         for flexible molecules. *Journal of Computer-Aided Molecular Design* **2001**, *15*,
         587-612.

(94)   Labute, P.; Williams, C.; Feher, M.; Sourial, E. and Schmidt, J. M. Flexible
         alignment of small molecules. *Journal of Medicinal Chemistry* **2001**, *44*, 1483-
         1490.

(95)   Sippl, W. Receptor-based 3D QSAR analysis of estrogen receptor ligands -
         merging the accuracy of receptor-based alignments with the computational
         efficiency of ligand-based methods. *Journal of Computer-Aided Molecular
         Design* **2000**, *14*, 559-572.

(96)   Waller, C. L.; Oprea, T. I.; Giolitti, A. and Marshall, G. R. Three-Dimensional
         Qsar of Human Immunodeficiency Virus (I) Protease Inhibitors .1. A Comfa
         Study Employing Experimentally-Determined Alignment Rules. *Journal of
         Medicinal Chemistry* **1993**, *36*, 4152-4160.

(97)   Wildman, S. A. and Crippen, G. M. Evaluation of ligand overlap by atomic
         parameters. *Journal of Chemical Information and Computer Sciences* **2000**, *41*,
         446-450.

(98)   Hansch, C.; Unger, S. H. and Forsythe, A. B. Strategy in drug design. Cluster
         analysis as an aid in the selection of substituents. *Journal of Medicinal
         Chemistry* **1973**, *16*, 1217-1222.

(99)   Hansch, C.; Leo, A.; Unger, S. H.; Kim, K. H.; Nikaitani, D. and Lien, E. J. "Aromatic" substituent constants for structure-activity correlations. *Journal of Medicinal Chemistry* **1973**, *16*, 1207-1216.

(100)   Gerlach, R. W.; Kowalski, B. R. and Wold, H. O. A. Partial least-squares path modelling with latent variables. *Anal. Chim. Acta* **1979**, *112*, 417-421.

(101)   Wold, H. Soft Modeling: The Basic Design and Some Extensions. *Systems Under Indirect Observation*: Amsterdam, 1982.

(102)   Cramer, R. D. d.; Patterson, D. E. and Bunce, J. D. Comparative Molecular Field Analysis (CoMFA) 1. Effect of shape on Binding of Steriods to Carrier Protiens. *Journal of American Chemical Society* **1988**, *110*, 5959-5967.

(103)   Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A. and Rognan, D. Recommendations for CoMFA Studies and 3D QSAR Publications. *Quantitative Structure-Activity Relationships* **1994**, *13*, 1-3.

(104)   Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F. and Wade, R. C. Reliability of Comparative Molecular Field Analysis Models - Effects of Data Scaling and Variable Selection Using a Set of Human Synovial Fluid Phospholipase a(2) Inhibitors. *Journal of Medicinal Chemistry* **1997**, *40*, 1136-1148.

(105)   Pajeva, I. K. and Wiese, M. Interpretation of CoMFA results - A probe set study using hydrophobic fields. *Quantitative Structure-Activity Relationships* **1999**, *18*, 369-379.

(106)   Kellogg, G. E.; Semus, S. F. and Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *Journal of Computer-Aided Molecular Design* **1991**, *5*, 545-552.

(107)   Klebe, G.; Abraham, U. and Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (Comsia) of Drug Molecules to Correlate and Predict Their Biological Activity. *Journal of Medicinal Chemistry* **1994**, *37*, 4130-4146.

(108)   Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry* **1985**, *28*, 849-857.

(109)   Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R. and Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quantitative Structure-Activity Relationships* **1993**, *12*, 9-29.

(110)   Nilsson, J.; Wikstrom, H.; Smilde, A.; Glase, S.; Pugsley, T.; Cruciani, G.; Pastor, M. and Clementi, S. GRID/GOLPE 3D quantitative structure-activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D3 receptor subtype. *Journal of Medicinal Chemistry* **1997**, *40*, 833-840.

(111)   Pastor, M.; Cruciani, G. and Clementi, S. Smart Region Definition - a New Way to Improve the Predictive Ability and Interpretability of Three-Dimensional Quantitative Structure-Activity Relationships. *Journal of Medicinal Chemistry* **1997**, *40*, 1455-1464.

(112)   Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S. and Clementi, S. GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *Journal of Medicinal Chemistry* **2000**, *43*, 3233-3243.

(113)   Hurst, T. and Heritage, T. W. HQSAR - a highly predictive QSAR technique based on molecular holograms. *213th ACS National Meeting*; American Chemical Society: San Francisco, 1997.

(114) Mason, J. S. and Beno, B. R. Library design using BCUT chemistry-space descriptors and multiple four-point pharmacophore fingerprints: simultaneous optimization and structure-based diversity. *Journal of Molecular Graphics and Modelling* **2000**, *18*, 438-451.

(115) McGregor, M. J. and Muskal, S. M. Pharmacophore fingerprinting in QSAR and primary library design. In *PCT Int. Appl.*; GLAXO GROUP LTD (GB): America, 2000; pp 102.

(116) Mason, J. S.; Beno, B. R. and Martin, E. J. 3-D pharmacophores in drug discovery. *Current Pharmaceutical Design* **2001**, *7*, 567-597.

(117) Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* **1982**, *79*, 2554-2558.

(118) Aoyama, T.; Suzuki, Y. and Ichikawa, H. Neural networks applied to quantitative structure-activity relationship analysis. *Journal of Medicinal Chemistry* **1990**, *33*, 2583-2590.

(119) Livingstone, D. J. and Manallack, D. T. Statistics using neural networks: chance effects. *Journal of Medicinal Chemistry* **1993**, *36*, 1295-1297.

(120) So, S. S. and Karplus, M. Evolutionary Optimization in Quantitative Structure-Activity Relationship - an Application of Genetic Neural Networks. *Journal of Medicinal Chemistry* **1996**, *39*, 1521-1530.

(121) So, S. S. and Karplus, M. Genetic Neural Networks for Quantitative Structure-Activity Relationships - Improvements and Application of Benzodiazepine Affinity for Benzodiazepine/Gaba(a) Receptors. *Journal of Medicinal Chemistry* **1996**, *39*, 5246-5256.

(122) Tetko, I. V.; Livingstone, D. J. and Luik, A. I. Neural Network Studies .1. Comparison of Overfitting and Overtraining. *Journal of Chemical Information & Computer Sciences* **1995**, *35*, 826-833.

(123) Manallack, D. T. and Livingstone, D. J. Neural networks in drug discovery: have they lived up to their promise? *European Journal of Medicinal Chemistry* **1999**, *34*, 195-208.

(124) Burden, F. R.; Ford, M. G.; Whitley, D. C. and Winkler, D. A. Use of automatic relevance determination in QSAR studies using Bayesian neural networks. *Journal of Chemical Information & Computer Sciences* **2000**, *40*, 1423-1430.

(125) Burden, F. R. and Winkler, D. A. Robust QSAR Models Using Bayesian Regularized Neural Networks. *Journal of Medicinal Chemistry* **1999**, *42*, 3183-3187.

(126) Mackay, D. J. C. Bayesian methods for back-propagation networks. *Models of Neural Networks III*; Springer: New York, 1994; pp 211-254.

(127) Neal, R. M. *Bayesian Learning for Neural Networks*; Springer: New York, 1996.

(128) Tetko, I. V.; Kovalishyn, V. V. and Livingstone, D. J. Volume learning algorithm artificial neural networks for 3D QSAR studies. *Journal of Medicinal Chemistry* **2001**, *44*, 2411-2420.

(129) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N. and Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235-242.

(130) Blundell, T. L.; Bedarkar, S.; Rinderknecht, E. and Humbel, R. E. Insulin-like growth factor: a model for tertiary structure accounting for immunoreactivity and receptor binding. *Proceedings of the National Academy of Sciences of the United States of America* **1978**, *75*, 180-184.

(131) Lewis, R. A. and Leach, A. R. Current methods for site-directed structure generation. *Journal of Computer-Aided Molecular Design* **1994**, *8*, 467-475.

(132) Pearlman, D. A. and Murcko, M. A. CONCERTS: dynamic connection of fragments as an approach to de novo ligand design. *Journal of Medicinal Chemistry* **1996**, *39*, 1651-1663.

(133) Bohm, H. J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design* **1992**, *6*, 61-78.

(134) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R. and Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **1982**, *161*, 269-288.

(135) Rarey, M.; Kramer, B.; Lengauer, T. and Klebe, G. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* **1996**, *261*, 470-489.

(136) Goodsell, D. S.; Morris, G. M. and Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *Journal of Molecular Recognition* **1996**, *9*, 1-5.

(137) Bissantz, C.; Folkers, G. and Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry* **2000**, *43*, 4759-4767.

(138) Stahl, M. and Rarey, M. Detailed analysis of scoring functions for virtual screening. *Journal of Medicinal Chemistry* **2001**, *44*, 1035-1042.

(139) Bissantz, C.; Bernard, P.; Hibert, M. and Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **2003**, *50*, 5-25.

(140) Attwood, T. K. and Findlay, J. B. C. Fingerprinting G-Protein-Coupled Receptors. *Protein Engineering* **1994**, *7*, 195-203.

(141) Kurogi, Y. and Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Current Medicinal Chemistry* **2001**, *8*, 1035-1055.

(142) Gustavsson, A.; Kihlen, M. and Uppenberg, J. Structure Based Focusing of Compound Libraries. *Rational Approaches to Drug Design: 13th European Symposium on QSAR*; Prous Science: Barcelona, 2001; pp 447-450.

(143) Lipinski, C. A.; Lambando, F.; Dominy, B. W. and Feenoy, P. J. Experimental and Computational approaches to estimate solubility and permeability in drug discovery and development. *Adv. Drug Dev. Res.* **1997**, *23*, 3-25.

(144) DiMasi, J. A. Risks in New Drug Development: Approval Success Rates for Investigational Drugs. *Clinical Pharmacology & Therapeutics* **2001**, *69*, 297-307.

(145) Schüürmann, G. Modelling pKa of Carboxylic Acids and Chlorinated Phenols. *Quantitative Structure-Activity Relationships* **1996**, *15*, 121-132.

(146) Gancia, E.; Montana, J. G. and Manallack, D. T. Theoretical Hydrogen Bonding Parameters for Drug Design. *Journal of Molecular Graphics and Modelling* **2001**, *19*, 349-362.

(147) Tehan, B. G.; Lloyd, E. J.; Wong, M. G.; Pitt, W. R.; Montana, J. G.; Manallack, D. T. and Gancia, E. Estimation of pK(a) using semiempirical molecular orbital methods. Part 1: Application to phenols and carboxylic acids. *Quantitative Structure-Activity Relationships* **2002**, *21*, 457-472.

(148) Tehan, B. G.; Lloyd, E. J.; Wong, M. G.; Pitt, W. R.; Gancia, E. and Manallack, D. T. Estimation of pK(a) using semiempirical molecular orbital methods. Part

2: Application to amines, anilines and various nitrogen containing heterocyclic compounds. *Quantitative Structure-Activity Relationships* **2002**, *21*, 473-485.

(149) Schrödinger *Jaguar*; http://www.schrodinger.com/Products/jaguar.html: 1500 SW First Ave, Suite 1180, Portland, Oregan, 97201, U.S.A.

(150) Beck, B.; Breindl, A. and Clark, T. QM/NN QSPR models with error estimation: Vapor pressure and LogP. *Journal of Chemical Information & Computer Sciences* **2000**, *40*, 1046-1051.

(151) Gross, K. C. and Seybold, P. G. Substituent effects on the physical properties and pK(a) of aniline. *International Journal of Quantum Chemistry* **2000**, *80*, 1107-1115.

(152) Gross, K. C.; Seybold, P. G.; Peralta-Inga, Z.; Murray, J. S. and Politzer, P. Comparison of quantum chemical parameters and Hammett constants in correlating pK(a) values of substituted anilines. *Journal of Organic Chemistry* **2001**, *66*, 6919-6925.

(153) ACD/Labs *ACD Labs*; version 3 ed.: 133 Richmond St. West Suite 605, Toronto, ON, M5H 2L3, Canada.

(154) Manallack, D. T.; Tehan, B. G.; Gancia, E.; Hudson, B. D.; Ford, M. G.; Livingstone, D. J.; Whitley, D. C. and Pitt, W. R. A Consensus Neural Network-Based Technique for Discriminating Soluble and Poorly Soluble Compounds. *Journal of Chemical Information and Computer Sciences* **2003**, *43*, 674-679.

(155) Artursson, P. and Karlsson, J. Correlation between oral drug absorption in humans and apparent permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880-885.

(156) MDL *MDL Metabolite Database*; http://www.mdli.com/products/products.html.

(157) MDL *MDL Toxicity Database*; http://www.mdli.com/products/products.html.

(158) Butina, D.; Segall, M. D. and Frankcombe, K. Predicting ADME properties in silico: methods and models. *Drug Discovery Today* **2002**, *7*, S83-S88.

(159) Sali, A.; Sanchez, R. and Badretdinov, A. *Modeller 4*; 4 ed.: Rockefeller University, New York.

(160) Sali, A. and Overington, J. Derivation of rules for comparative protein modelling from a database of protein structure alignments. *Protein Science* **1994**, *3*, 1582-1596.

(161) Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P. and Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 A resolution. *Journal of Molecular Biology* **1999**, *291*, 899-911.

(162) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M. and Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, *289*, 739-745.

(163) Teeter, M. M.; Froimowitz, M.; Stec, B. and Durand, C. J. Homology Modeling of the Dopamine D-2 Receptor and Its Testing by Docking of Agonists and Tricyclic Antagonists. *Journal of Medicinal Chemistry* **1994**, *37*, 2874-2888.

(164) Trumpp-Kallmeyer, S.; Hoflack, J.; Bruinvels, A. and Hibert, M. Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. *Journal of Medicinal Chemistry* **1992**, *35*, 3448-3462.

(165) Javitch, J. A.; Li, X. C.; Kaback, J. and Karlin, A. A Cysteine Residue in the Third Membrane-Spanning Segment of the Human D_2 Dopamine Receptor Is Exposed in the Binding-Site Crevice. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 10355-10359.

(166) Javitch, J. A.; Fu, D. Y. and Chen, J. Y. Residues in the Fifth Membrane-Spanning Segment of the Dopamine D_2 Receptor Exposed in the Binding-Site Crevice. *Biochemistry* **1995**, *34*, 16433-16439.

(167) Javitch, J. A.; Fu, D. Y.; Chen, J. Y. and Karlin, A. Mapping the Binding-Site Crevice of the Dopamine D_2 Receptor by the Substituted-Cysteine Accessibility Method. *Neuron* **1995**, *14*, 825-831.

(168) Fu, D. Y.; Ballesteros, J. A.; Weinstein, H.; Chen, J. Y. and Javitch, J. A. Residues in the Seventh Membrane-Spanning Segment of the Dopamine D_2 Receptor Accessible in the Binding-Site Crevice. *Biochemistry* **1996**, *35*, 11278-11285.

(169) Javitch, J. A.; Ballesteros, J. A.; Weinstein, H. and Chen, J. Y. A Cluster of Aromatic Residues in the Sixth Membrane-Spanning Segment of the Dopamine D_2 Receptor Is Accessible in the Binding-Site Crevice. *Biochemistry* **1998**, *37*, 998-1006.

(170) Javitch, J. A.; Ballesteros, J. A.; Chen, J.; Chiappa, V. and Simpson, M. M. Electrostatic and aromatic microdomains within the binding-site crevice of the D_2 receptor: contributions of the second membrane-spanning segment. *Biochemistry* **1999**, *38*, 7961-7968.

(171) Simpson, M. M.; Ballesteros, J. A.; Chiappa, V.; Chen, J.; Suehiro, M.; Hartman, D. S.; Godel, T.; Snyder, L. A.; Sakmar, T. P. and Javitch, J. A. Dopamine D_4/D_2 receptor selectivity is determined by A divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. *Molecular Pharmacology* **1999**, *56*, 1116-1126.

(172) Javitch, J. A.; Shi, L.; Simpson, M. M.; Chen, J. Y.; Chiappa, V.; Visiers, I.; Weinstein, H. and Ballesteros, J. A. The fourth transmembrane segment of the dopamine D_2 receptor: Accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry* **2000**, *39*, 12190-12199.

(173) Shi, L.; Simpson, M. M.; Ballesteros, J. A. and Javitch, J. A. The first transmembrane segment of the dopamine D2 receptor: Accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry* **2001**, *40*, 12339-12348.

(174) Shi, L. and Javitch, J. A. The binding site of aminergic G protein-coupled receptors: The transmembrane segments and second extracellular loop [Review]. *Annual Review of Pharmacology & Toxicology* **2002**, *42*, 437-467.

(175) Javitch, J. A. The substituted-cysteine accessibility method. *Structure-Function Analysis of G Protein-Coupled Receptors*; Wiley-Liss: New York, 1999; pp 21-41.

(176) Mansour, A.; Meng, F.; Meador-Woodruff, J. H.; Taylor, L. P.; Civelli, O. and Akil, H. Site-directed mutagenesis of the human dopamine D2 receptor. *European Journal of Pharmacology* **1992**, *227*, 205-214.

(177) Bikker, J. A.; Trumpp-Kallmeyer, S. and Humblet, C. G-Protein coupled receptors: models, mutagenesis, and drug design. *Journal of Medicinal Chemistry* **1998**, *41*, 2911-2927.

(178) Howard, P. and Meylan, W. Physical/Chemical Property Database (PHYSPROP); 1999 ed.; Syracuse Research Corporation, Environmental Science Center: North Syracuse NY, 1999.

(179) Fukui, K.; Yonezawa, T. and Nagata, C. Theory of substitution in conjugated molecules. *Bulletin of Chemical Society of Japan* **1954**, *27*, 423-427.

(180) Grüber, C. and Buss, V. Quantum-Mechanically Calculated Properties for the Development of Quantitative Structure Activity Relationships (QSARs). pKa

Values of Phenols and Aromatic and Aliphatic Carboxylic Acids. *Chemosphere* **1989**, *19*, 1595-1609.

(181) Citra, M. J. Estimating the pKa of Phenols, Carboxylic Acids and Alcohols from Semi-Empirical Quantum Chemical Methods. *Chemosphere* **1999**, *38*, 191-206.

(182) Pallas *pKalc*; 5 ed.; CompuDrug International, Inc.: P.O.B. 160, Budapest, 1255, Hungary.

`,` **2.9 BIBLIOGRAPHY**

(1)   Minchin, S. A. and Csernansky, J. G. Classification schemes for Antipsychotic
        Drugs. *Antipsychotics*; Springer, 1996; pp 1-21.
(2)   Conley, R. R.; Tamminga, C. A.; Kelly, D. L. and Richardson, C. M. Treatment-
        resistant schizophrenic patients respond to clozapine after olanzapine non-
        response. *Biolgical Psychiatry* **1999**, *46*, 73-77.
(3)   Ashby, C. R., Jr. and Wang, R. Y. Pharmacological Action of the Atypical
        Antipsychotic Drug Clozapine: A Review. *Synapse* **1996**, *24*, 349-394.
(4)   Leysen, J. E.; Janssen, P. M.; Schotte, A.; Luyten, W. H. and Megens, A. A.
        Interaction of antipsychotic drugs with neurotransmitter receptor sites in vitro
        and in vivo in relation to pharmacological and clinical effects: role of 5HT2
        receptors. *Psychopharmacology* **1993**, *112*, S40-54.
(5)   Buckley, P. F. and Meltzer, H. Y. Treatment of Schizophrenia. *Textbook of
        Psychopharmacology*; American Psychiatric Press, 1995; pp 615-639.
(6)   Meltzer, H. Y.; Matsubara, S. and Lee, J. C. Classification of typical and
        atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin2
        pKi values. *Journal of Pharmacology & Experimental Therapeutics* **1989**, *251*,
        238-246.
(7)   Wilson, J. M.; Sanyal, S. and Van Tol, H. H. Dopamine D_2 and D_4 receptor
        ligands: relation to antipsychotic action. *European Journal of Pharmacology*
        1998, *351*, 273-286.
(8)   Bøgesø, K. P.; Liljefors, T.; Arnt, J.; Hyttel, J. and Pedersen, H. Octoclothepin
        enantiomers. A reinvestigation of their biochemical and pharmacological
        activity in relation to a new receptor-interaction model for dopamine D_2 receptor
        antagonists. *Journal of Medicinal Chemistry* **1991**, *34*, 2023-2030.
(9)   Seeman, P.; Westman, K.; Protiva, M.; Jilek, J.; Jain, P. C.; Saxena, A. K.;
        Anand, N.; Humber, L. and Philipp, A. Neuroleptic receptors: stereoselectivity
        for neuroleptic enantiomers. *European Journal of Pharmacology* **1979**, *56*, 247-
        251.
(10)   Jaunin, A.; Petcher, T. J. and Weber, H. P. (+)-(S)-2-Chloro-10,11-dihydro-11-
        (4-methylpiperazin-1-yl)-dibenzo(b,f)thiepin. *Journal of Chemical Society,
        Perkin Transactions 2* **1977**, 186.
(11)   CCDC *Cambridge Crystallographic Data Centre*; 5.18 ed.: 12 Union Road.
        Cambridge. CB2 1EZ. UK.
(12)   Accelrys *CAChe*; version 3.1 ed.: The Medawar Centre, Oxford Science Park,
        Oxford, OX4 4GA, UK.
(13)   Tripos *Sybyl*; version 6.8 ed.: 1699 South Hanley Road, St. Louis, MO 63144,
        USA.
(14)   Clark, M.; Cramer III, R. D. and Van Opdenbosch, N. Validation of the general
        purpose Tripos 5.2 force field. *Journal of Computational Chemistry* **1989**, 982-
        1012.
(15)   Gasteiger, J. and Marsili, M. Iterative partial equalization of orbital
        electronegativity - rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219-
        3228.
(16)   *MacroModel*; 5 ed.: Department of Chemistry, Columbia University New York,
        NY 10027, USA.
(17)   *Amsol*; 6.6 ed.: Department of Chemistry, University of Minnesota Minneapolis,
        Minnesota 55455-0431, USA.

(18) Petcher, T. J. Topology of Dopamine Receptors. *The Role of Brain Dopamine* **1985**, 47-56.

(19) Liljefors, T. and Pettersson, I. Computer Aided Development of Three-Dimensional Pharmacophore Models. *A Textbook of Drug Design and Development*; Harwood Academic Publishers, 1997; pp 61-93.

(20) ACD/Labs *ACD Labs*; version 3 ed.: 133 Richmond St. West Suite 605, Toronto, ON, M5H 2L3, Canada.

(21) Liljefors, T. and Bogeso, K. P. Conformational analysis and structural comparisons of (1R,3S)-(+)- and (1S,3R)-(-)-tefludazine, (S)-(+)- and (R)-(-)-octoclothepin, and (+)-dexclamol in relation to dopamine receptor antagonism and amine-uptake inhibition. *Journal of Medicinal Chemistry* **1988**, *31*, 306-312.

(22) Phillips, S. T.; de Paulis, T.; Baron, B. M.; Siegel, B. W.; Seeman, P.; Van Tol, H. H.; Guan, H. C. and Smith, H. E. Binding of 5H-dibenzo[b,e][1,4]diazepine and chiral 5H-dibenzo[a,d]cycloheptene analogues of clozapine to dopamine and serotonin receptors. *Journal of Medicinal Chemistry* **1994**, *37*, 2686-2696.

(23) Phillips, S. T.; de Paulis, T.; Neergaard, J. R.; Baron, B. M.; Siegel, B. W.; Seeman, P.; Van Tol, H. H.; Guan, H. C. and Smith, H. E. Binding of 5H-dibenzo[a,d]cycloheptene and dibenz[b,f]oxepin analogues of clozapine to dopamine and serotonin receptors. *Journal of Medicinal Chemistry* **1995**, *38*, 708-714.

(24) Liegeois, J. F.; Rogister, F. A.; Bruhwyler, J.; Damas, J.; Nguyen, T. P.; Inarejos, M. O.; Chleide, E. M.; Mercier, M. G. and Delarge, J. E. Pyridobenzoxazepine and pyridobenzothiazepine derivatives as potential central nervous system agents: synthesis and neurochemical study. *Journal of Medicinal Chemistry* **1994**, *37*, 519-525.

(25) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A. and Rognan, D. Recommendations for CoMFA Studies and 3D QSAR Publications. *Quantitative Structure-Activity Relationships* **1994**, *13*, 1-3.

(26) Cramer, R. D. d.; Patterson, D. E. and Bunce, J. D. Comparative Molecular Field Analysis (CoMFA) 1. Effect of shape on Binding of Steriods to Carrier Protiens. *Journal of American Chemical Society* **1988**, *110*, 5959-5967.

(27) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *Journal of Computational Chemistry* **1999**, *20*, 730-748.

(28) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *Journal of Computational Chemistry* **1999**, *20*, 720-729.

(29) Tintelnot, M. and Andrews, P. Geometries of functional group interactions in enzyme-ligand complexes: guides for receptor modelling. *Journal of Computer-Aided Molecular Design* **1989**, *3*, 67-84.

(30) Bruno, I. J.; Cole, J. C.; Lommerse, J. P. M.; Rowland, R. S.; Taylor, R. and Verdonk, M. L. Isostar - a Library of Information About Nonbonded Interactions. *Journal of Computer-Aided Molecular Design* **1997**, *11*, 525-537.

(31) Hendrickson, J. B. machine computation of the common rings. *J. Am. Chem. Soc.* **1961**, *83*, 4537-4547.

(32) Irurre, J.; Narbon, E.; Serra, B. and Teixido, J. Conformational Barrier Calculation of Dibenzo[b,e]Azepine Derivatives by Semi-Empirical Method AM1. *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications.*; Prous Science, 1995; pp 329-331.

(33) Petcher, T. J. and Weber, H. P. Conformations of some semi-rigid neuroleptic drugs. Part 1. Crystal structures of loxapine, clozapine, and HUF-2406 monohydrate {2-chloro-11-(4-methylpiperazin-1-yl)dibenzo[b,e][1,4]diazepine monohydrate}. *Journal of the Chemical Society, Perkin Transactions 2: Physical Organic Chemistry (1972-1999)* **1976**, 1415-1420.

(34) Fillers, J. P. and Hawkinson, S. W. The structure of 8-chloro-11-(4-methyl-1-piperazinyl)-5H-dibenzo[b,e][1,4]diazepine dihydrobromide, clozapine dihydrobromide. *Acta Crystallographica, Section B: Structural Crystallography and Crystal Chemistry* **1982**, *B38*, 1750-1753.

(35) Boström, J.; Gundertofte, K. and Liljefors, T. A pharmacophore model for dopamine D-4 receptor antagonists. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 769-786.

(36) Campiani, G.; Butini, S.; Gemma, S.; Nacci, V.; Fattorusso, C.; Catalanotti, B.; Giorgi, G.; Cagnotto, A.; Goegan, M.; Mennini, T.; Minetti, P.; Di Cesare, M. A.; Mastroianni, D.; Scafetta, N.; Galletti, B.; Stasi, M. A.; Castorina, M.; Pacifici, L.; Ghirardi, O.; Tinti, O. and Carminati, P. Pyrrolo[1,3]benzothiazepine-based atypical antipsychotic agents. Synthesis, structure-activity relationship, molecular modeling, and biological studies. *Journal of Medicinal Chemistry* **2002**, *45*, 344-359.

(37) Karle, I. L.; Britts, K. and Brenner, S. 1-Cyclohexenyl-1-cyclobutenedione. *Acta. Crystallogr.* **1964**, *17*, 1506.

(38) Karle, I. L. and Dragonette, K. S. 1,2,3,4-Tetraphenyl-cis,cis-butadiene. *Acta. Crystallogr.* **1965**, *19*, 500.

(39) Takahashi, H. J.; Nozawa, K. and Kawai, K. I. Epurpurin C dimethyl ether. *Chem. Pharm. Bull.* **1996**, *44*, 2227.

(40) Ferguson, G.; Marsh, W. C.; Restivo, R. J. and Lloyd, D. 2,3-Dibenzoyl-5-nitrocyclopentadiene. *J. Chem. Soc. Perkin Trans.2* **1975**, 998.

(41) Fornies-Marquina, J. M.; Courseille, C.; Busetta, B. and Hospital, M. Dienestrol Cycladiene oestrogenic activity. *Acta. Crystallogr., Sect. B* **1972**, *28*, 655.

(42) Lloyd, E. J. and Andrews, P. R. A Common Structural Model for Central Nervous System Drugs and their Receptors. *Journal of Medicinal Chemistry* **1986**, *29*, 453-462.

(43) Leysen, J. E.; Niemegeers, C. J. E.; Van Nueten, J. M. and Laduron, P. M. [3H]Ketanserin (R 41 468), a selective 3H-ligand for serotonin2 receptor binding sites. Binding properties, brain distribution, and functional role. *Mol. Pharmacol.* **1982**, *21*, 301-314.

(44) De Paulis, T.; Betts, C. R.; Smith, H. E.; Mobley, P. L.; Manier, D. H. and Sulser, F. Synthesis of clozapine analogs and their affinity for clozapine and spiroperidol binding sites in rat brain. *Journal of Medicinal Chemistry* **1981**, *24*, 1021-1026.

(45) Pearlman, R. S. 3D Molecular structures: generation and use in 3D searching. *3D QSAR in Drug Design, Theory, Methods and Applications.*; ESCOM Science Publishers: Leiden, 1993; pp 41-79.

(46) Besler, B. H.; Merz, J. K. M. and Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *Journal of Computational Chemistry* **1990**, *11*, 431-439.

(47) Pajeva, I. K. and Wiese, M. Interpretation of CoMFA results - A probe set study using hydrophobic fields. *Quantitative Structure-Activity Relationships* **1999**, *18*, 369-379.

(48)   Clark, M. and Cramer, R. D. The Probability of Chance Correlation Using
       Partial Least Squares (Pls). *Quantitative Structure-Activity Relationships* **1993**,
       12, 137-145.

(49)   Liegeois, J. F.; Dupont, L. and Delarge, J. Application de l'analyse CoMFA
       (Comparative Molecular Field Analysis) à une série de dérivés tricycliques
       apparentés à la clozapine. *J. Pharm. Belg.* **1992**, *47*, 100-108.

(50)   Martin, A. R. Antipsychotic Agents. *Burgers Medicinal Chemistry and Drug
       Discovery Volume 5: Therapeutic Agents*; Wiley-Interscience Publication, 1997;
       pp 195-253.

(51)   Baldessarini, R. J. Drugs and the Treatment of Psychiatric Disorders. *Goodman
       & Gilmans The Pharmacological Basis of Therapeutics*; McGraw-Hill, 1996; pp
       399-430.

(52)   Ohmori, J.; Maeno, K.; Hidaka, K.; Nakato, K.; Matsumoto, M.; Tada, S.;
       Hattori, H.; Sakamoto, S.; Tsukamoto, S.; Usuda, S. and Mase, T. Dopamine D_3
       and D_4 receptor antagonists: synthesis and structure--activity relationships of
       (S)-(+)-N-(1-Benzyl-3-pyrrolidinyl)-5-chloro-4- [(cyclopropylcarbonyl) amino]-
       2-methoxybenzamide (YM-43611) and related compounds. *J. Med. Chem.* **1996**,
       39, 2764-2772.

(53)   Moore, K. W.; Bonner, K.; Jones, E. A.; Emms, F.; Leeson, P. D.; Marwood, R.;
       Patel, S.; Rowley, M.; Thomas, S. and Carling, R. W. 4-N-linked-heterocyclic
       piperidine derivatives with high affinity and selectivity for human dopamine D_4
       receptors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1285-1290.

(54)   Carling, R. W.; Moore, K. W.; Moyes, C. R.; Jones, E. A.; Bonner, K.; Emms,
       F.; Marwood, R.; Patel, S.; Fletcher, A. E.; Beer, M.; Sohal, B.; Pike, A. and
       Leeson, P. D. 1-(3-Cyanobenzylpiperidin-4-yl)-5-methyl-4-phenyl-1, 3-
       dihydroimidazol-2-one: a selective high-affinity antagonist for the human
       dopamine D_4 receptor with excellent selectivity over ion channels. *J. Med.
       Chem.* **1999**, *42*, 2706-2715.

(55)   Rowley, M.; Collins, I.; Broughton, H. B.; Davey, W. B.; Baker, R.; Emms, F.;
       Marwood, R.; Patel, S.; Ragan, C. I.; Freedman, S. B.; Ball, R. and Leeson, P.
       D. 4-Heterocyclylpiperidines as selective high-affinity ligands at the human
       dopamine D_4 receptor. *J. Med. Chem.* **1997**, *40*, 2374-2385.

(56)   Unangst, P. C.; Capiris, T.; Connor, D. T.; Heffner, T. G.; MacKenzie, R. G.;
       Miller, S. R.; Pugsley, T. A. and Wise, L. D. Chromeno[3,4-c]pyridin-5-ones:
       selective human dopamine D_4 receptor antagonists as potential antipsychotic
       agents. *J. Med. Chem.* **1997**, *40*, 2688-2693.

(57)   Bristow, L. J.; Collinson, N.; Cook, G. P.; Curtis, N.; Freedman, S. B.;
       Kulagowski, J. J.; Leeson, P. D.; Patel, S.; Ragan, C. I.; Ridgill, M.; Saywell, K.
       L. and Tricklebank, M. D. L-745,870, a subtype selective dopamine D_4 receptor
       antagonist, does not exhibit a neuroleptic-like profile in rodent behavioral tests.
       J. Pharmacol. Exp. Ther. **1997**, *283*, 1256-1263.

(58)   Kulagowski, J. J.; Broughton, H. B.; Curtis, N. R.; Mawer, I. M.; Ridgill, M. P.;
       Baker, R.; Emms, F.; Freedman, S. B.; Marwood, R.; Patel, S.; Ragan, C. I. and
       Leeson, P. D. 3-((4-(4-Chlorophenyl)piperazin-1-yl)-methyl)-1H-pyrrolo-2,3-b-
       pyridine: an antagonist with high affinity and selectivity for the human
       dopamine D_4 receptor. *J. Med. Chem.* **1996**, *39*, 1941-1942.

(59)   Rowley, M.; Broughton, H. B.; Collins, I.; Baker, R.; Emms, F.; Marwood, R.;
       Patel, S.; Ragan, C. I.; Freedman, S. B. and Leeson, P. D. 5-(4-Chlorophenyl)-4-
       methyl-3-(1-(2-phenylethyl)piperidin-4-yl)isoxazole: a potent, selective

antagonist at human cloned dopamine D_4 receptors. *J. Med. Chem.* **1996**, *39*, 1943-1945.

(60)   Sanner, M. A.; Chappie, T. A.; Dunaiskis, A. R.; Fliri, A. F.; Desai, K. A.; Zorn, S. H.; Jackson, E. R.; Johnson, C. G.; Morrone, J. M.; Seymour, P. A.; Majchrzak, M. J.; Faraci, W. S.; Collins, J. L.; Duignan, D. B.; Prete Di, C. C.; Lee, J. S. and Trozzi, A. Synthesis, SAR and pharmacology of CP-293,019: a potent, selective dopamine D_4 receptor antagonist. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 725-730.

(61)   Tang, A. H.; Franklin, S. R.; Himes, C. S.; Smith, M. W. and Tenbrink, R. E. PNU-96415E, a potential antipsychotic agent with clozapine-like pharmacological properties. *J. Pharmacol. Exp. Ther.* **1997**, *281*, 440-447.

(62)   Tenbrink, R. E.; Bergh, C. L.; Duncan, J. N.; Harris, D. W.; Huff, R. M.; Lahti, R. A.; Lawson, C. F.; Lutzke, B. S.; Martin, I. J.; Rees, S. A.; Schlachter, S. K.; Sih, J. C. and Smith, M. W. (S)-(-)-4-[4-[2-(Isochroman-1-Yl)Ethyl]-Piperazin-1-Yl]Benzenesulfonamide , a Selective Dopamine D_4 Antagonist. *J. Med. Chem.* **1996**, *39*, 2435-2437.

(63)   Kesten, S. R.; Heffner, T. G.; Johnson, S. J.; Pugsley, T. A.; Wright, J. L. and Wise, L. D. Design, synthesis, and evaluation of chromen-2-ones as potent and selective human dopamine D_4 antagonists. *J. Med. Chem.* **1999**, *42*, 3718-3725.

(64)   Perrone, R.; Berardi, F.; Colabufo, N. A.; Leopoldo, M. and Tortorella, V. N-[2-[4-(4-Chlorophenyl)Piperazin-1-Yl]Ethyl]-3-Methoxybenzamide - a Potent and Selective Dopamine D_4 Ligand. *J. Med. Chem.* **1998**, *41*, 4903-4909.

(65)   Boyfield, I.; Brown, T. H.; Coldwell, M. C.; Cooper, D. G.; Hadley, M. S.; Hagan, J. J.; Healy, M. A.; Johns, A.; King, R. J.; Middlemiss, D. N.; Nash, D. J.; Riley, G. J.; Scott, E. E.; Smith, S. A. and Stemp, G. Design and Synthesis of 2-Naphthoate Esters as Selective Dopamine D_4 Antagonists. *J. Med. Chem.* **1996**, *39*, 1946-1948.

(66)   Belliotti, T. R.; Wustrow, D. J.; Brink, W. A.; Zoski, K. T.; Shih, Y. H.; Whetzel, S. Z.; Georgic, L. M.; Corbin, A. E.; Akunne, H. C.; Heffner, T. G.; Pugsley, T. A. and Wise, L. D. A series of 6-and 7-piperazinyl- and -piperidinylmethylbenzoxazinones with dopamine D_4 antagonist activity: Discovery of a potential atypical antipsychotic agent. *J. Med. Chem.* **1999**, *42*, 5181-5187.

(67)   Thurkauf, A.; Yuan, J.; Chen, X.; He, X. S.; Wasley, J. W. F.; Hutchison, A.; Woodruff, K. H.; Meade, R.; Hoffman, D. C.; Donovan, H. and Joneshertzog, D. K. 2-Phenyl-4(5)-[[4-(Pyrimidin-2-Yl)Piperazin-1-Yl]Methyl]Imidazole - a Highly Selective Antagonist at Cloned Human D_4 Receptors. *J. Med. Chem.* **1997**, *40*, 1-3.

(68)   Clark, M.; Cramer III, R. D. and Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comp. Chem.* **1989**, 982-1012.

(69)   Stewart, J. J. P. MOPAC: a semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design* **1990**, *4*, 1-105.

(70)   Stewart, J. J. P. *MOPAC 93: a general-purpose semiempirical quantum mechanics package*; '93 ed.; Fujitsu: Frank J. Seiler Research Laboratory, United States Air Force Academy, CO 80840.

(71)   Taylor, R.; Kennard, O. and Versichel, W. Geometry of the N-H.....O=C Hydrogen Bond. 1. Lone-Pair Directionality. *J. Am. Chem. Soc.* **1983**, *105*, 5761-5766.

(72)   Taylor, R.; Kennard, O. and Versichel, W. Geometry of the N-H.....O=C
       Hydrogen Bond. 2. Three-Center ("Bifurcated") and Four-Center ('Trifucated")
       Bonds. *J. Am. Chem. Soc.* **1984**, *106*, 244-248.

(73)   Jones, G.; Willett, P. and Glen, R. C. A Genetic Algorithm for Flexible
       Molecular Overlay and Pharmacophore Elucidation. *Journal of Computer-Aided
       Molecular Design* **1995**, *9*, 532-549.

(74)   Jones, G.; Willett, P. and Glen, R. C. GASP: Genetic Algorithm
       Superimposition Program. *Pharmacophore Perception, Development, and Use
       in Drug Design* **2000**, 85-107.

(75)   Conradi, R. A.; Burton, P. S. and Borchardt, R. T. Physico-chemical and
       Biological Factors that Influence a Drug's Cellular Permeability by Passive
       Diffusion. *Lipophilicity in Drug Action and Toxicology.*; VCH Publishers, 1996;
       pp 233-252.

(76)   Lipinski, C. A.; Lambando, F.; Dominy, B. W. and Feenoy, P. J. Experimental
       and Computational approaches to estimate solubility and permeability in drug
       discovery and development. *Adv. Drug Dev. Res.* **1997**, *23*, 3-25.

(77)   Houttemane, C.; Boivin, J. C.; Nowogrocki, G.; Thomas, D. J.; Bonte, J. P. and
       Debaert, M. N-(2-Diethylamino-ethyl)-2-methoxy-5-methylsulfonyl-benzamide,
       Tiapride. *Acta Crystallogr.,Sect.C* **1983**, *39*, 585.

(78)   Adachi, T.; Mizoguchi, J. I.; Hayashi, Y.; Yamashoji, Y.; Kanehisa, N.; Kai, Y.
       and Inoue, Y. 4-Amino-5-chloro-2-methoxy-N-((2S,4S)-1-ethyl-2-
       hydroxymethyl-4-pyrrolidinyl)-2-ethoxybenzamide, Agent TKS159. *Acta
       Crystallogr.,Sect.C* **1998**, *54*, 1527.

(79)   Blaton, N. M.; Peeters, O. M.; De Ranter, C. J.; Denisoff, O. and Molle, L. 4-
       Amino-5-chloro-N-((2-diethylamino)-ethyl)-2-methoxybenzamide
       hydrochloride monohydrate, Primperan. *Cryst.Struct.Commun.* **1980**, *9*, 857.

(80)   Blaton, N. M.; Peeters, O. M.; De Ranter, C. J.; Denisoff, O. and Molle, L. 5-
       (Aminosulfonyl)-N-((1-ethyl-2-pyrrolidinyl)-methyl)-2-methoxybenzamide
       hydrochloride, Sulpiride hydrochloride. *Cryst.Struct.Commun.* **1981**, *10*, 833.

(81)   Blaton, N. M.; Peeters, O. M.; De Ranter, C. J.; Denisoff, O. and Molle, L. 5-
       Methylsulfonyl-N-((2-diethylamino)-ethyl)-2-methoxybenzamide
       hydrochloride,Tiapride hydrochloride derivative of psychotropic and
       antidyskinetic agent. *Cryst.Struct.Commun.* **1982**, *11*, 1357.

(82)   Bradley, G.; Ward, T. J.; White, J. C.; Coleman, J.; Taylor, A. and Rhodes, K. F.
       endo-N-(((8-methyl-8-azabicyclo(3.2.1)octan-3-yl)amino)carbonyl)-2-
       (cyclopropylmethoxy)benzamide, WAY-100289. *J.Med.Chem.* **1992**, *35*, 1515.

(83)   Collin, S.; Vercauteren, D. P.; Evrard, G.; Durant, F.; Tollenaere, J. P. and
       Moereels, H. Cisapride monohydratea non-dopamine blocking gastrokinetic
       agent. *J.Mol.Struct.* **1989**, *214*, 159.

(84)   Dapporto, P. and Sega, A. Diethylmethyl-(2-(p-(o-propyloxybenzamido)-
       benzoyloxy)-ethyl)-ammonium bromide. *Acta Crystallogr.,Sect.C* **1986**, *42*, 474.

(85)   De Winter, H. L.; Verlinde, C. L.; Blaton, N. M.; Peeters, O. M. and De Ranter,
       C. J. 4-Amino-N-1-((1-ethyl-2-pyrrolidinyl)methyl)-5-(ethylsulfonyl)-2-
       methoxybenzamide, Amisulpride. *Acta Crystallogr.,Sect.C* **1990**, *46*, 313.

(86)   Durant, F.; Renard, P.; De Beys, V. and Evrard, G. exo-4-Amino-5-bromo-2-
       methoxy-N-(8-(2-chlorophenyl-methyl)-8-azabicyclo(3.2.1)oct-3-yl)-
       benzamide. *Bull.Soc.Chim.Belg.* **1984**, *93*, 923.

(87)   Foresti, E.; Riva di Sanseverino, L. and Sabatino, P. 4-Methoxy-N,N'-bis(3-
       pyridylmethyl)-isophthalamide monohydrate, Picotamide monohydrate. *Acta
       Crystallogr.,Sect.C* **1986**, *42*, 220.

(88)   Furuya, T.; Iwanami, S.; Takenaka, A. and Sasada, Y. N-((2RS,3SR)-1-Benzyl-
        2-methyl-3-pyrrolidinyl)-5-chloro-2-methoxy-4-methylaminobenzamide
        hydrochloride neuroleptic drug with antipsychotic activity. *Acta
        Crystallogr.,Sect.C* **1986**, *42*, 117.

(89)   Furuya, T.; Iwanami, S.; Takenaka, A. and Sasada, Y. N-((2RS,3SR)-1-Benzyl-
        2-methyl-3-pyrrolidinyl)-5-chloro-2-methoxy-4-methylaminobenzamide
        hydrochloride monohydrate. *Acta Crystallogr.,Sect.C* **1986**, *42*, 117.

(90)   Hogberg, T.; Ramsby, S.; de Paulis, T.; Stensland, B.; Csoregh, I. and Wagner,
        A. (-)-(S)-3-Bromo-N-((1-ethyl-2-pyrrolidinyl)methyl)-6-methoxysalicylamide,
        dopamine receptor blocking activity. *Mol.Pharmacol.* **1986**, *30*, 345.

(91)   Wong, M. Personal Communication, unpublished data, 2000.

(92)   Koch, M. H. J. and Dideberg, O. 1-Benzyl-4-(2,6-dioxo-3-phenyl-3-
        piperidyl)piperidine Benzitimide. *Acta Crystallogr.,Sect.B*, **1973**, *29*, 369.

(93)   Brine, G. A.; Stark, P. A.; Carroll, F. I. and Singh, P. trans-(+-)-1-Benzyl-C-4-
        cyano-R-3-methyl-N-phenyl-4-piperidinamine. *J.Heterocycl.Chem.* **1994**, *31*,
        513.

(94)   Furuya, T.; Iwanami, S.; Takenaka, A. and Sasada, Y. N-((3RS,5SR)-1-Benzyl-
        5-methyl-3-pyrrolidinyl)-5-chloro-2-methoxy-4-methylaminobenzamide
        hydrochloride. *Acta Crystallogr.,Sect.C* **1986**, *42*, 1345.

(95)   Spek, A. L. (+)-1-Benzyl-4-(2,6-dioxo-3-phenyl-3-piperidyl)piperidine
        hydrobromide hemihydrate, Dexetimide hydrobromide hemihydrate. *Acta
        Crystallogr.,Sect.B*, **1976**, *32*, 1870.

(96)   Humblet, C.; Durant, F.; Evrard, G. and Koch, M. H. J. 8-((4-
        Methylphenyl)methyl)-1-phenyl-1,3,8-triazaspiro(4,5)decan-4-one. *Acta
        Crystallogr.,Sect.B* **1976**, *32*, 2878.

(97)   Kharchenko, V. G.; Chałaya, S. N.; Struchkov, Y. T.; Espenbetov, A. A.;
        Litvinov, O. V. and Komyag, N. T. 2-Benzoyl-5-phenyl-pyrrole.
        Khim.Get.Soedin.,SSSR **1984**, 1355.

(98)   L'abbe, G.; Van Stappen, P.; Toppet, S.; Germain, G. and Scheefer, G. 4,5-
        bis(Methoxycarbonyl)-3-methyl-2-phenyl-pyrrole. *Bull.Soc.Chim.Belg.* **1983**,
        92, 193.

(99)   Lian-He Yu; Yang-Fu Ming; Mei-Gong Fan and Zhu, Y. 3-Acetyl-1,2-dimethyl-
        4-isopropyl-5-phenylpyrrole. *Youji Huaxue(J.Org.Chem.)* **1996**, *16*, 53.

(100)   Lian-He Yu; Yang-Fu Ming; Mei-Gong Fan and Zhu, Y. 3-Acetyl-1,2,4-
        trimethyl-5-(p-methoxyphenyl)pyrrole. *Youji Huaxue(J.Org.Chem.)* **1996**, *16*,
        53.

(101)   Lianhe Yu; Yangfu Ming and Fan, M. (E)-(1,2-Dimethyl-4-isopropyl-5-phenyl)-
        3-pyrryl-ethylidene-(isopropylidene)succinnic anhydride pyrryl fulgide
        photochromic. *Mol.Cryst.Liq.Cryst.Sci.Technol.,Sect.A* **1997**, *297*, 115.

(102)   Singh, P.; Scott, L. and Myers, J. A. 1,2,5-Trimethyl-3-phenyl-2H-isoindole-4,7-
        dione. *Cryst.Struct.Commun.* **1982**, *11*, 1091.

(103)   Yoneda, F.; Motokura, M.; Kamishimoto, M.; Nagamatsu, T.; Otagiri, M.;
        Uekama, K. and Takamoto, M. 8-(p-Chlorophenyl)-9-methoxy-7-methyl-9-
        deazatheophylline. *Chem.Pharm.Bull.* **1982**, *30*, 3187.

(104)   Boström, J.; Gundertofte, K. and Liljefors, T. A 3-D Pharmacophore Model for
        Dopamine D₄ Receptor Antagonists. *Molecular Modelling and Prediction of
        Bioactivity*; Kluwer Academic/Plenum Publishers: Copenhagen, 1999; pp 382-
        383.

(105)   Liao, Y.; DeBoer, P.; Meier, E. and Wikstrom, H. Synthesis and
        pharmacological evaluation of triflate-substituted analogues of clozapine:

identification of a novel atypical neuroleptic. *Journal of Medicinal Chemistry* **1997**, *40*, 4146-4153.

(106) Liao, Y.; Venhuis, B. J.; Rodenhuis, N.; Timmerman, W.; Wikstrom, H.; Meier, E.; Bartoszyk, G. D.; Bottcher, H.; Seyfried, C. A. and Sundell, S. New (sulfonyloxy)piperazinyldibenzazepines as potential atypical antipsychotics: chemistry and pharmacological evaluation. *Journal of Medicinal Chemistry* **1999**, *42*, 2235-2244.

(107) Bandoli, G. and Nicolini, M. Solid state conformational parameters in dibenzo[b,f]heteroepin drugs: crystal structures of 3-methoxy-10-methyl-11-phenyldibenzo[b,f]thiepine and 3-allyloxy-10-ethyl-11-phenyldibenzo[b,f]oxepine. *Journal of Crystallographic and Spectroscopic Research* **1982**, *12*, 425-447.

(108) Liegeois, J.-F.; Eyrolles, L.; Ellenbroek, B. A.; Lejeune, C.; Carato, P.; Bruhwyler, J.; Geczy, J.; Damas, J. and Delarge, J. New Pyridobenzodiazepine Derivatives: Modifications of the Basic Side Chain Differentially Modulate Binding to Dopamine (D4.2, D2L) and Serotonin (5-HT2A) Receptors. *Journal of Medicinal Chemistry* **2002**, *45*, 5136-5149.

3.2 BIBLIOGRAPHY

(1)   GlaxoWellcome Intelligent Drug Design. *Nature Suppl.* **1996**, *384*, 1-5.

(2)   Gudermann, T.; Numberg, B. and Schultz, G. Receptors and G Proteins as Primary Components of Transmembrane Signal Transduction .1. G-Protein-Coupled Receptors - Structure and Function [Review]. *Journal of Molecular Medicine-Jmm* **1995**, *73*, 51-63.

(3)   Dahl, S. G.; Edvardsen, O. and Sylte, I. Molecular dynamics of dopamine at the D2 receptor. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 8111-8115.

(4)   Findlay, J. and Eliopoulos, E. Three-dimensional modelling of G protein-linked receptors. [erratum appears in Trends Pharmacol Sci 1991 Mar;12(3):81]. *Trends in Pharmacological Sciences* **1990**, *11*, 492-499.

(5)   Hibert, M. F.; Trumpp-Kallmeyer, S.; Bruinvels, A. and Hoflack, J. Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Molecular Pharmacology* **1991**, *40*, 8-15.

(6)   Trumpp-Kallmeyer, S.; Hoflack, J.; Bruinvels, A. and Hibert, M. Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. *Journal of Medicinal Chemistry* **1992**, *35*, 3448-3462.

(7)   Teeter, M. M.; Froimowitz, M.; Stec, B. and Durand, C. J. Homology Modeling of the Dopamine D-2 Receptor and Its Testing by Docking of Agonists and Tricyclic Antagonists. *Journal of Medicinal Chemistry* **1994**, *37*, 2874-2888.

(8)   Boström, J. Dopamine D_4 receptor antagonists pharmacophore, 3D-QSAR and 7TM modeling. In *Medicinal Chemistry*; Royal Danish School of Pharmacy: Copenhagen, 2000; pp 146.

(9)   Van Rhee, A. M.; Fischer, B.; Van Galen, P. J. and Jacobson, K. A. Modelling the P2Y purinoceptor using rhodopsin as template. *Drug Design & Discovery* **1995**, *13*, 133-154.

(10)   Sali, A.; Sanchez, R. and Badretdinov, A. *Modeller 4*; 4 ed.: Rockefeller University, New York.

(11)   Sali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **1993**, *234*, 779-815.

(12)   Schoonman, M. J.; Knegtel, R. M. and Grootenhuis, P. D. Practical evaluation of comparative modelling and three ling methods. *Computers & Chemistry* **1998**, *22*, 369-375.

(13)   Tehan, B. G.; Wong, M. G.; Lloyd, E. J. and Chalmers, D. K. Analysis of Agonism by Dopamine at the Dopaminergic D2 G-Protein Coupled Receptor based on Comparative Modelling of Rhodopsin. *Molecular Simulations* **2001**, *in press*.

(14)   Sali, A. and Overington, J. Derivation of rules for comparative protein modelling from a database of protein structure alignments. *Protein Science* **1994**, *3*, 1582-1596.

(15)   Shacham, S.; Topf, M.; Avisar, N.; Glaser, F.; Marantz, Y.; Bar-Haim, S.; Noiman, S.; Naor, Z. and Becker, O. M. Modeling the 3D structure of GPCRs from sequence. *Medicinal Research Reviews* **2001**, *21*, 472-483.

(16)   Yeagle, P. L.; Choi, G. and Albert, A. D. Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in unactivated and activated forms. *Biochemistry* **2001**, *40*, 11932-11937.

(17) Bikker, J. A.; Trumpp-Kallmeyer, S. and Humblet, C. G-Protein coupled receptors: models, mutagenesis, and drug design. *Journal of Medicinal Chemistry* **1998**, *41*, 2911-2927.

(18) Cho, W.; Taylor, L. P.; Mansour, A. and Akil, H. Hydrophobic Residues of the D-2 Dopamine Receptor Are Important for Binding and Signal Transduction. *Journal of Neurochemistry* **1995**, *65*, 2105-2115.

(19) Javitch, J. A.; Li, X. C.; Kaback, J. and Karlin, A. A Cysteine Residue in the Third Membrane-Spanning Segment of the Human D_2 Dopamine Receptor Is Exposed in the Binding-Site Crevice. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 10355-10359.

(20) Javitch, J. A.; Fu, D. Y. and Chen, J. Y. Residues in the Fifth Membrane-Spanning Segment of the Dopamine D_2 Receptor Exposed in the Binding-Site Crevice. *Biochemistry* **1995**, *34*, 16433-16439.

(21) Javitch, J. A.; Fu, D. Y.; Chen, J. Y. and Karlin, A. Mapping the Binding-Site Crevice of the Dopamine D_2 Receptor by the Substituted-Cysteine Accessibility Method. *Neuron* **1995**, *14*, 825-831.

(22) Javitch, J. A.; Ballesteros, J. A.; Chen, J.; Chiappa, V. and Simpson, M. M. Electrostatic and aromatic microdomains within the binding-site crevice of the D_2 receptor: contributions of the second membrane-spanning segment. *Biochemistry* **1999**, *38*, 7961-7968.

(23) Mansour, A.; Meng, F.; Meador-Woodruff, J. H.; Taylor, L. P.; Civelli, O. and Akil, H. Site-directed mutagenesis of the human dopamine D2 receptor. *European Journal of Pharmacology* **1992**, *227*, 205-214.

(24) Muller, G. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. *Current Medicinal Chemistry* **2000**, *7*, 861-888.

(25) Simpson, M. M.; Ballesteros, J. A.; Chiappa, V.; Chen, J.; Suehiro, M.; Hartman, D. S.; Godel, T.; Snyder, L. A.; Sakmar, T. P. and Javitch, J. A. Dopamine D_4/D_2 receptor selectivity is determined by A divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. *Molecular Pharmacology* **1999**, *56*, 1116-1126.

(26) Shi, L.; Simpson, M. M.; Ballesteros, J. A. and Javitch, J. A. The first transmembrane segment of the dopamine D2 receptor: Accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry* **2001**, *40*, 12339-12348.

(27) Fu, D. Y.; Ballesteros, J. A.; Weinstein, H.; Chen, J. Y. and Javitch, J. A. Residues in the Seventh Membrane-Spanning Segment of the Dopamine D_2 Receptor Accessible in the Binding-Site Crevice. *Biochemistry* **1996**, *35*, 11278-11285.

(28) Javitch, J. A.; Ballesteros, J. A.; Weinstein, H. and Chen, J. Y. A Cluster of Aromatic Residues in the Sixth Membrane-Spanning Segment of the Dopamine D_2 Receptor Is Accessible in the Binding-Site Crevice. *Biochemistry* **1998**, *37*, 998-1006.

(29) Javitch, J. A.; Shi, L.; Simpson, M. M.; Chen, J. Y.; Chiappa, V.; Visiers, I.; Weinstein, H. and Ballesteros, J. A. The fourth transmembrane segment of the dopamine D_2 receptor: Accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry* **2000**, *39*, 12190-12199.

(30) Javitch, J. A. The substituted-cysteine accessibility method. *Structure-Function Analysis of G Protein-Coupled Receptors*; Wiley-Liss: New York, 1999; pp 21-41.

(31)   Pearson, W. R. Flexible sequence similarity searching with the FASTA3 program package. *Methods in Molecular Biology* **2000**, *132*, 185-219.

(32)   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W. and Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403-410.

(33)   Jones, D. T. *Threader 2*; 2.5 ed.: Department of Biological Sciences, University of Warwick.

(34)   Leach, A. R. and Editor *Molecular Modeling: Principles and Applications*, 1996; 616 pp.

(35)   Srinivasan, N. and Blundell, T. L. An evaluation of the performance of an automated procedure for comparative modeling of protein tertiary structure. *Protein Engineering* **1993**, *6*, 501-512.

(36)   Bates, P. A.; Kelley, L. A.; MacCallum, R. M. and Sternberg, M. J. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *PROTEINS* **2001**, *Suppl 5*, 39-46.

(37)   Laskowski, R. A.; MacArthur, M. W.; Moss, D. S. and Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystalography* **1993**, *26*, 283-291.

(38)   Hooft, R. W. W.; Vriend, G.; Sander, C. and Abola, E. E. Errors in protein structures. *Nature* **1996**, *381*, 272-272.

(39)   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. and Karplus, M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4*, 187-217.

(40)   Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K. and Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics* **1999**, *151*, 283-312.

(41)   Braun, W. and Gō, N. Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm. *Journal of Molecular Biology* **1985**, *186*, 611-626.

(42)   Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M. and Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**, *289*, 739-745.

(43)   Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E. and Downing, K. H. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology* **1990**, *213*, 899-929.

(44)   Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P. and Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 A resolution. *Journal of Molecular Biology* **1999**, *291*, 899-911.

(45)   Sander, C. and Schneider, R. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **1991**, *9*, 56-68.

(46)   Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G. and Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **1992**, *12*, 345-364.

(47)   Hubbard, S. J. and Thornton, J. M. *NACCESS*; 2.1.1 ed.: Department of Biochemistry and Molecular Biology, University College London.

(48)   Tripos *Sybyl*; version 6.8 ed.: 1699 South Hanley Road, St. Louis, MO 63144, USA.

(49)   Weiner, S. J.; Kollman, P. A.; Nguyen, D. T. and Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *Journal of Computational Chemistry* **1986**, *7*, 230-252.

(50)   Bairoch, A. and Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **2000**, *28*, 45-48.

(51)   Galtier, N.; Gouy, M. and Gautier, C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in the Biosciences* **1996**, *12*, 543-548.

(52)   Stefanini, E.; Marchisio, A. M.; Devoto, P.; Vernaleone, F.; Collu, R. and Spano, P. F. Sodium-dependent interaction of benzamides with dopamine receptors. *Brain Res* **1980**, *198*, 229-233.

(53)   Neve, K. A.; Cox, B. A.; Henningsen, R. A.; Spanoyannis, A. and Neve, R. L. Pivotal role for aspartate-80 in the regulation of dopamine D2 receptor affinity for drugs and inhibition of adenylyl cyclase. *Molecular Pharmacology* **1991**, *39*, 570-578.

(54)   Riek, R. P.; Rigoutsos, I.; Novotny, J. and Graham, R. M. Non-alpha-helical elements modulate polytopic membrane protein architecture. *Journal of Molecular Biology* **2001**, *306*, 349-362.

(55)   Neve, K. A.; Cumbay, M. G.; Thompson, K. R.; Yang, R.; Buck, D. C.; Watts, V. J.; DuRand, C. J. and Teeter, M. M. Modeling and mutational analysis of a putative sodium-binding pocket on the dopamine D2 receptor. *Molecular Pharmacology* **2001**, *60*, 373-381.

(56)   Rowley, M.; Bristow, L. J. and Hutson, P. H. Current and novel approaches to the drug treatment of schizophrenia. *Journal of Medicinal Chemistry* **2001**, *44*, 477-501.

(57)   Gazi, L.; Bobirnac, I.; Danzeisen, M.; Schupbach, E.; Langenegger, D.; Sommer, B.; Hoyer, D.; Tricklebank, M. and Schoeffter, P. Receptor density as a factor governing the efficacy of the dopamine D4 receptor ligands, L-745,870 and U-101958 at human recombinant D4.4 receptors expressed in CHO cells. *British Journal of Pharmacology* **1999**, *128*, 613-620.

4.6 BIBLIOGRAPHY

(1)   Bowman, W. C. and Rand, M. J. *Textbook of Pharmacology*; 2nd ed.; Blackwell
       Scientific Publications: Oxford, 1980.

(2)   Albert, A. and Serjeant, E. P. *The Determination of Ionization Constants*; 2nd
       ed.; Chapman and Hall: London, 1971.

(3)   ACD/Labs *ACD Labs*; version 3 ed.: 133 Richmond St. West Suite 605,
       Toronto, ON, M5H 2L3, Canada.

(4)   Pallas *pKalc*; 5 ed.; CompuDrug International, Inc.: P.O.B. 160, Budapest, 1255,
       Hungary.

(5)   Schrödinger *Jaguar*; http://www.schrodinger.com/Products/jaguar.html: 1500
       SW First Ave, Suite 1180, Portland, Oregan, 97201, U.S.A.

(6)   Grüber, C. and Buss, V. Quantum-Mechanically Calculated Properties for the
       Development of Quantitative Structure Activity Relationships (QSARs). pKa
       Values of Phenols and Aromatic and Aliphatic Carboxylic Acids. *Chemosphere*
       1989, *19*, 1595-1609.

(7)   Citra, M. J. Estimating the pKa of Phenols, Carboxylic Acids and Alcohols from
       Semi-Empirical Quantum Chemical Methods. *Chemosphere* **1999**, *38*, 191-206.

(8)   Fukui, K.; Yonezawa, T. and Nagata, C. Theory of substitution in conjugated
       molecules. *Bulletin of Chemical Society of Japan* **1954**, *27*, 423-427.

(9)   Gancia, E.; Montana, J. G. and Manallack, D. T. Theoretical Hydrogen Bonding
       Parameters for Drug Design. *Journal of Molecular Graphics and Modelling*
       2001, *19*, 349-362.

(10)   Howard, P. and Meylan, W. Physical/Chemical Property Database
       (PHYSPROP); 1999 ed.; Syracuse Research Corporation, Environmental
       Science Center: North Syracuse NY, 1999.

(11)   Lipinski, C. A.; Lambando, F.; Dominy, B. W. and Feenoy, P. J. Experimental
       and Computational approaches to estimate solubility and permeability in drug
       discovery and development. *Adv. Drug Dev. Res.* **1997**, *23*, 3-25.

(12)   Perrin, D. D. *Dissociation Constants of Organic Bases in Aqueous Solution*;
       Butterworths Scientific Publications: 88 Kingsway, W.C.2, London., 1965.

(13)   Tripos Inc.: 1699 South Hanley Road, St. Louis, MO 63144, USA.

(14)   Abraham, M. H.; Duce, P. P.; Prior, D. V.; Barratt, D. G.; Morris, J. J. and
       Taylor, P. J. Hydrogen bonding. Part 9. Solute proton donor and proton acceptor
       scales for use in drug design. *Journal of the Chemical Society Perkins
       Transactions II* **1989**, 1355-1375.

(15)   Bondi, A. *Journal of Physical Chemistry* **1964**, *68*, 441.

(16)   Sadowski, J. and Gasteiger, J. From atoms and bonds to three-dimensional
       atomic coordinates: automatic model builders. *Chemical Reviews* **1993**, *93*,
       2567-2581.

(17)   Stewart, J. J. P. *Mopac program package, quantum chemistry program exchange
       no. 455.*

(18)   Accelrys *Tsar*; version 3.21 ed.: The Medawar Centre, Oxford Science Park,
       Oxford, OX4 4GA, UK.

(19)   Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F. and Stewart, J. J. P. Development
       and use of quantum mechanical molecular models. AM1: a new general purpose
       quantum mechanical molecular model. *Journal of the American Chemical
       Society* **1985**, *107*, 3902-3909.

(20) Serjeant, E. P. and Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution.*; Pergamon Press: Headington Hill Hall, Oxford, OX3 OBW, England, 1979.

(21) Topliss, J. G. and Edwards, R. P. Chance Factors in Studies of Quantitative Structure-Activity Relationships. *Journal of Medicinal Chemistry* **1979**, *22*, 1238-1244.

(22) Schüürmann, G. Modelling pKa of Carboxylic Acids and Chlorinated Phenols. *Quantitative Structure-Activity Relationships* **1996**, *15*, 121-132.

(23) Coulson, C. A. *VALENCE*; 2nd ed.; Oxford University Press: Oxford University Press, Amen House, London E.C.4, 1961; 417.

(24) Gross, K. C.; Seybold, P. G.; Peralta-Inga, Z.; Murray, J. S. and Politzer, P. Comparison of quantum chemical parameters and Hammett constants in correlating pK(a) values of substituted anilines. *Journal of Organic Chemistry* **2001**, *66*, 6919-6925.

(25) Gross, K. C. and Seybold, P. G. Substituent effects on the physical properties and pK(a) of aniline. *International Journal of Quantum Chemistry* **2000**, *80*, 1107-1115.

5.5 BIBLIOGRAPHY

(1)   Selassie, C. D. History of Quantitative Structure-Activity Relationships. *Burgers Medicinal Chemistry and Drug Discovery Volume 1: Drug Discovery*; Wiley-Interscience Publication, 2003; pp 1-48.

(2)   Sheridan, R. P. The most common chemical replacements in drug-like compounds. *Journal of Chemical Information and Computer Sciences* **2002**, *42*, 103-108.

(3)   Egan, W. J.; Walters, W. P. and Murcko, M. A. Guiding molecules towards drug-likeness. *Current Opinion in Drug Discovery & Development* **2002**, *5*, 540-549.

(4)   Tripos Inc.: 1699 South Hanley Road, St. Louis, MO 63144, USA.

(5)   Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *Journal of Computational Chemistry* **1999**, *20*, 720-729.

(6)   Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *Journal of Computational Chemistry* **1999**, *20*, 730-748.

(7)   Fukui, K.; Yonezawa, T. and Nagata, C. Theory of substitution in conjugated molecules. *Bulletin of Chemical Society of Japan* **1954**, *27*, 423-427.

(8)   Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F. and Stewart, J. J. P. Development and use of quantum mechanical molecular models. AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society* **1985**, *107*, 3902-3909.

(9)   Simpson, M. M.; Ballesteros, J. A.; Chiappa, V.; Chen, J.; Suehiro, M.; Hartman, D. S.; Godel, T.; Snyder, L. A.; Sakmar, T. P. and Javitch, J. A. Dopamine D_4/D_2 receptor selectivity is determined by A divergent aromatic microdomain contained within the second, third, and seventh membrane-spanning segments. *Molecular Pharmacology* **1999**, *56*, 1116-1126.