Multiway calibration in 3D QSAR
Nilsson, Jonas

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1998

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
1.1 Schizophrenia

Schizophrenia is one of the most common psychiatric disorders and approximately 1 % of the world's population suffer from severe symptoms occupying more than half of the beds in psychiatric clinics. Schizophrenia is distributed over the whole population independent of sex, location, social class or color of the skin. A schizophrenic patient is frequently described as a person with a “Dr Jekyll and Mr. Hyde” personality, but the diagnosis of schizophrenia is more complex than that. Generally, the symptoms are divided into two classes: positive (reality distortion) and negative symptoms (psycho-motor poverty syndrome). Each patient is different and could suffer from more or less of positive or negative symptoms. The positive symptoms, e.g., delusions, hallucinations, grandiosity, excitement, hostility and disorganization are more easily identified, as compared to the negative symptoms. Examples of negative symptoms are apathy, attentional impairment, affective blunting, asociality, poverty of speech and anhedonias that may be difficult to distinguish from either depression or side effects caused by medication with typical antipsychotic drugs.2,3

Approximately half of the schizophrenia patients will experience periods with severe depression during the course of their illness. Consequently, the detection of depressive symptoms4 is very important, since, about 7 %–10 % of all schizophrenia patients commit suicide.5

: Treatment of Schizophrenia

The diagnosis of schizophrenia is just as complex as the medication to suppress the symptoms. There is no real cure against schizophrenia and most patients are bound to medication for the rest of their lives. The drug of choice is more often a trade-off between clinical efficacy and EPS (Extra Pyramidal Syndrome) or other side-effects. EPS are the side-effects, e.g., major movement disorders elicited by typical antipsychotic drugs. In general, low-potency drugs, e.g., chlorpromazine (1) or thioridazine (2), are more sedative and hypotensive than high-potency drugs, e.g., fluphenazine (3) and haloperidol (4) which, in turn, produce more EPS than low-potency agents.
Thus, patients that are highly agitated and excited may be better off with a drug as chlorpromazine. On the contrary, if there is no need for sedation and no history of unusual sensitivity to EPS, high-potency drugs as haloperidol (4) or fluphenazine (3), are most likely prescribed.

Recently, risperidone (5) at fixed doses of 2, 6 and 16 mg/day, has been reported to have higher efficacy and elicit fewer EPS than haloperidol (4).\textsuperscript{6,7} It was found that risperidone,\textsuperscript{7} at daily doses of 6 mg, was more effective than haloperidol and placebo against both negative and positive subscales of PANSS (Positive and Negative Syndrome Scale).\textsuperscript{8} At higher doses, no advantage of risperidone over haloperidol was demonstrated. In the same investigation,\textsuperscript{7} it was shown that risperidone can suppress TD (Tardive Dyskinesia) but whether it was superior to other typical neuroleptics was not clear.

![Chemical structures](1-chlorpromazine.png) ![Chemical structures](2-thioridazine.png) ![Chemical structures](3-fluphenazine.png) ![Chemical structures](4-haloperidol.png) ![Chemical structures](5-risperidone.png) ![Chemical structures](6-clozapine.png) ![Chemical structures](7-olanzapine.png)
Just as risperidone (5), clozapine (6) belongs to the new generation of antipsychotics often classified as atypical. An atypical antipsychotic drug produces, by definition, fewer EPS than typical antipsychotics and clozapine is the compound used as reference for new antipsychotics. The advantages of clozapine (6) over classical antipsychotics are manifold: a) it is effective in the treatment of both positive and negative symptoms; b) it is more effective in treatment-refractory patients and c) it produces fewer EPS. However, clozapine is also an example of what often is referred to as a “dirty drug” since it has affinity to a large number of different receptors (see Table 1.1). As a consequence, side-effects like hypersalivation (may be due to peripheral drug actions) and weight gain must be considered. In addition, agranulocytosis, a potentially fatal blood disorder has been observed in patients medicated with clozapine.

Figure 1.1 The nigrostriatal and the mesolimbic systems. The presynaptic part of the nigrostriatal system originates in A9 and A8 and its axons terminates, mainly, in the forebrain. The mesolimbic system, runs parallel to the nigrostriatal system, and originates in the VTA (A10) with projections to a number of areas, e.g., accumbens, septum and cerebral cortex (frontal, cingulate and entorinal).

Olanzapine (7, LY170053) is a novel atypical antipsychotic with similar binding profile as clozapine (Table 1.1). In studies of schizophrenic patients, olanzapine was demonstrated to be effective in the treatment of both negative and positive symptoms with few EPS. Additionally, the potency of olanzapine in reversing the effects of d-amphetamine was greater, as compared to clozapine. (The reversal of the inhibitory effects of d-amphetamine on A10 cells [Figure 1.1] has been hypothesized to be predictive of clinical antipsychotic efficacy). These findings are consistent with the fact that olanzapine is a DA D₂ antagonist in vivo, and is more potent at DA D₂ receptors in vitro, as compared to clozapine (Table 1.1).
Table 1.1 Receptor affinity for a selection of drugs used in the treatment of schizophrenia. All values are $K_i$ (nM).

<table>
<thead>
<tr>
<th>receptor</th>
<th>chlorpromazine</th>
<th>thioridazine</th>
<th>trifluoperazine</th>
<th>haloperidol</th>
<th>risperidone</th>
<th>clozapine</th>
<th>olanzapine</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>32</td>
<td>32</td>
<td>5</td>
<td>10</td>
<td>75</td>
<td>85</td>
<td>31</td>
</tr>
<tr>
<td>$D_{2L}$</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>0.5</td>
<td>1.5</td>
<td>60-150</td>
<td>11</td>
</tr>
<tr>
<td>$D_{2S}$</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>1.5</td>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td>$D_3$</td>
<td>~4</td>
<td>2</td>
<td></td>
<td>2</td>
<td>7</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>$D_4$</td>
<td>34</td>
<td>31</td>
<td></td>
<td>2</td>
<td>7</td>
<td>9-54</td>
<td>27</td>
</tr>
<tr>
<td>$D_5$</td>
<td></td>
<td></td>
<td></td>
<td>27</td>
<td></td>
<td>35-400</td>
<td></td>
</tr>
<tr>
<td>5HT$_{1A}$</td>
<td>3635</td>
<td>643</td>
<td></td>
<td>1714</td>
<td>16</td>
<td>875</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>5HT$_{2A}$</td>
<td>7</td>
<td>48</td>
<td>52</td>
<td>74</td>
<td>0.6</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>5HT$_{2C}$</td>
<td>12</td>
<td>60</td>
<td>295</td>
<td>5755</td>
<td>16</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>5HT$_{4}$</td>
<td>4</td>
<td>7</td>
<td>17</td>
<td>&gt;5000</td>
<td>425</td>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>5HT$_7$</td>
<td>21</td>
<td>70</td>
<td>8</td>
<td>263</td>
<td>1.4</td>
<td>6</td>
<td>104</td>
</tr>
<tr>
<td>M$_1$</td>
<td>25</td>
<td>3</td>
<td></td>
<td>1700</td>
<td>&gt;3000</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>M$_2$</td>
<td>150</td>
<td>14</td>
<td></td>
<td>2500</td>
<td>&gt;3000</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>M$_3$</td>
<td>67</td>
<td>15</td>
<td></td>
<td>&gt;3000</td>
<td>&gt;3000</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>M$_4$</td>
<td>40</td>
<td>9</td>
<td></td>
<td>2700</td>
<td>&gt;3000</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>M$_5$</td>
<td>42</td>
<td>13</td>
<td></td>
<td>1800</td>
<td>&gt;3000</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>H$_1$</td>
<td></td>
<td></td>
<td></td>
<td>3630</td>
<td>&gt;10000</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>~5</td>
<td>~7</td>
<td></td>
<td>46</td>
<td>&gt;10000</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>360</td>
<td></td>
<td></td>
<td>2904</td>
<td>8</td>
<td>228</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td></td>
<td></td>
<td></td>
<td>&gt;10000</td>
<td>&gt;10000</td>
<td>&gt;10000</td>
<td>&gt;10000</td>
</tr>
</tbody>
</table>

Binding data are obtained from the following references:17,20-28

The Dopamine Hypothesis of Schizophrenia

Extensive research during the last decades, have given rise to many different hypotheses over the pathophysiology of schizophrenia. In 1963 Carlsson and Lindquist$^{29}$ reported that neuroleptics, e.g., chlorpromazine (1), increase dopamine (8) and noradrenaline (9) turnover in rat brain, and postulated that this effect is caused by blockade of catecholamine receptors. These findings form the basis of the DA hypothesis of schizophrenia. Accordingly, it was found that DA agonists can induce psychosis similar to acute paranoid schizophrenia$^{30,31}$ and that neuroleptics inhibit dopaminergic activity.$^{32-34}$ Based on the fact that dopamine-mimetic drugs elicit hallucinations, and that other neuroleptics cause rigidity, Van Rossum$^{35}$ suggested that schizophrenia may be caused by overactivity in certain dopaminergic pathways.$^{36}$ As further support of the DA hypothesis, the clinical doses of neuroleptics and antipsychotics were found to correlate very well with their ability to block DA D$_2$ receptors.$^{37,38}$ Furthermore, the correlations between the clinical efficacy of neuroleptic drugs
and the \textit{in vitro} binding affinity of the muscarinic cholinergic, histaminergic (H\textsubscript{1}), serotonergic (5-HT\textsubscript{2}) and \(\alpha\textsubscript{1}\) receptors, were poor.\textsuperscript{39}

There are two major dopaminergic neuronal systems that project in the forebrain: the nigrostriatal (A9) and the mesolimbic (A10) systems (Figure 1.1). Parkinsonism is a consequence of degeneration of neuronal pathways in the A9 system, and EPS induced by treatment with typical antipsychotics is caused by blockade of dopamine receptors in the same system. Consequently, it was postulated that the symptoms of schizophrenia originated from hyperactivity in the \textit{mesolimbic} dopaminergic systems (A10).\textsuperscript{34,40}

In agreement with the DA hypothesis is the clinical observation that patients with Parkinson’s disease do not develop schizophrenia.\textsuperscript{38}

\textbf{Serotonin Hypothesis of Schizophrenia}

The first indications that serotonin (10, 5-HT) might be involved in the pathophysiology of schizophrenia came with the discovery that certain ergots (\textit{e.g.}, lysergic acid diethylamide (11)), with structural resemblance to 5-HT, were hallucinogenic and induced many of the symptoms of schizophrenia.\textsuperscript{41}

\begin{center}
\begin{tabular}{ccc}
8 dopamine; DA & 9 noradrenaline; NA & 10 serotonin; 5-HT & 11 (+)-LSD
\end{tabular}
\end{center}

Today, several atypical antipsychotic drugs (\textit{e.g.}, clozapine (6), olanzapine (7) and risperidone (5)) with affinity to one or several serotonin receptor subtypes (Table 1.1) are known. Clozapine for example, the prototypical atypical antipsychotic drug, has been shown to have high affinity towards, at least, four different serotonin receptors including 5-HT\textsubscript{2A}, 5-HT\textsubscript{2C}, 5-HT\textsubscript{6} and 5-HT\textsubscript{7}.\textsuperscript{20,24,26,42} Meltzer \textit{et al.},\textsuperscript{26} however, showed that most putative atypical antipsychotic drugs could be classified by their 5-HT\textsubscript{2A}/D\textsubscript{2} affinity ratios.

It has been found that full or partial 5-HT\textsubscript{1A} agonists reverse catalepsy in rat.\textsuperscript{43,44} Catalepsy in the rat is predictive for extrapyramidal side-effects in man.\textsuperscript{45}

\textbf{Muscarinic Hyperactivity in Schizophrenia}

Tandon \textit{et al.}\textsuperscript{46} suggested that hyperactivity of muscarinic cholinergic receptors had a role in the pathogenesis of negative symptoms of schizophrenia. Their observations showed that unmedicated schizophrenics displayed symptoms, (\textit{e.g.}, reduced pain perception, hyper-salivation and increased
water intake), which resemble a muscarinic receptor hyperactive state. Occasionally, anti-cholinergic drugs have been reported effective in treating negative symptoms of schizophrenia.\textsuperscript{47,48} It was not clear, however, whether the negative symptoms of schizophrenia or the neuroleptic induced EPS, were reduced.

Since the atypical antipsychotic drug clozapine (6), has high affinity towards all five muscarinic receptors (Table 1.1), one may predict that the degree of atypicality is related to its cholinergic activity. However, Bolden et al.\textsuperscript{27} could not find any clear pattern in their investigation. Taken together, according to the investigations of Boldens et al.\textsuperscript{27} and others,\textsuperscript{11,49,50} it is not clear whether anticholinergic activity is essential for an atypical antipsychotic drug or not.

: The Noradrenaline Receptor

The relationship between noradrenaline (9, NA) and schizophrenia was first studied by Stein et al.\textsuperscript{51} in 1971. However, significant and reproducible research established a relationship between NA levels in the limbic forebrain and the intensity of the schizophrenic symptoms first in 1990.\textsuperscript{52} Recently Breier et al.\textsuperscript{53} demonstrated a direct correlation between the ability of clozapine (6) to elevate plasma NA levels with its ability to improve positive symptoms of schizophrenia. As yet, no selective drugs towards the $\alpha_1$, $\alpha_2$ or $\beta$ receptors with high efficacy in man have been reported.\textsuperscript{54}

1.2 Molecular Biology of Dopamine Receptors

Until now, five central dopamine receptor subtypes have been discovered, distributed with the highest concentrations in the putamen, caudate nucleus and the nucleus accumbens. As can be seen in Figure 1.2, the different dopamine receptor subtypes are not homogeneously distributed in the brain. Instead each subtype is concentrated in specific small areas. Generally, three major dopamine pathways\textsuperscript{38,55} are discussed: the nigrostriatal, the mesolimbic and the tuberoinfundibular pathways. The first two pathways (Figure 1.1) control voluntary movement and regulate emotional behavior, respectively. The tuberoinfundibular pathway regulates the secretion of prolactin from the pituitary,\textsuperscript{56} and is thus, not mentioned in the context of schizophrenia. The nigrostriatal pathway (A9) has cellbodies in the substantia nigra, with long axons projecting in the corpus striatum (Figure 1.1). The abundance of DA D\textsubscript{1} and D\textsubscript{2} receptors\textsuperscript{57-60} in the nigrostriatal system is high while the presence of DA D\textsubscript{3} receptors is very low.\textsuperscript{55,61} Instead, high concentration of mRNA for the DA D\textsubscript{3} receptors are found in the limbic areas\textsuperscript{55,61,62} (Figure 1.2), suggesting DA D\textsubscript{3} receptors to be involved in emotional and cognitive disorders (e.g., schizophrenia). The mesolimbic neuronal pathway (A10) has cellbodies in the ventral tegmental area (VTA) of the brainstem, with cells projecting in the limbic system (Figure 1.1).
Introduction to the Medicinal Chemistry of Schizophrenia

Figure 1.2 The distribution of dopamine receptors in the human brain as determined by the concentrations of mRNA for the respective receptor subtypes in the different brain areas.

: G Protein-Coupled Receptors

The dopamine receptors belong to a class of proteins normally referred to as the G protein-coupled receptor (GPCR) superfamily. To date, no X-ray crystallographic structure of a GPCR is resolved, but along with molecular cloning and receptor binding studies the amino acid sequence of all five human DA receptors have been elucidated (Table 1.2). In 1993, Schertler et al. provided evidence that the bovine rhodopsins, G protein-coupled receptors active as the photoreceptors in rod cells, were arranged in seven $\alpha$-helices. In 1990, Henderson et al. presented a high quality 3D model of bacteriorhodopsin, also a G protein-coupled receptor, based on cryo-microscopy experiments. Further refinements of this model have recently been published by Grigorieff et al., Unger et al. and Kimura et al. The latter group collected structural data from bacteriorhodopsin crystals at 3.0 Å resolution with 90% completeness using electron cryo-microscopy. Although the function of bacteriorhodopsin is different from rhodopsin both proteins bind retinal in a similar way, and have similar topology with seven transmembrane helices.

The receptor protein may be folded through the cellular membrane forming seven hydrophobic trans-membrane $\alpha$-helices connected, alternately, via intra- and extra-cellular loops. The amino terminal (N-terminal) and the carboxylic terminal (C-terminal) of the receptor protein reside at the extra- and the intracellular sides of the cell-membrane, respectively.

The intrinsic activity of a DA agonist is mediated by a signal transduction across the cellular membrane. That is, the drug-receptor interaction most probably induces a conformational change in the receptor protein which in turn, activates a G protein coupled to the third intracellular loop. Accordingly, the activated G protein stimulates (or inhibits) adenylyl cyclase (see Table 1.2) to
produce cAMP (a so-called second messenger) from AMP (Figure 1.3), which influences various processes in the cytosol.

The third intracellular loop exhibit the largest sequence dissimilarities among the different DA receptors. The DA D$_1$ and DA D$_5$ receptors have relative short third intracellular loops, are coupled to G$_s$ proteins and have long C-terminal tails. These two receptors, the D$_1$-like receptors, stimulate the activity of adenylyl cyclase and the pharmacological functions from known ligands are more or less, identical.$^{54,61}$ The D$_2$-like receptors, i.e., D$_2$, D$_3$ and D$_4$ receptors on the other hand, all have long third intracellular loops with short C-terminal tails, might couple to G$_i$ proteins (or G$_0$ proteins) and inhibit adenylyl cyclase. Interesting, two forms of the DA D$_2$ receptor have been found, differing in 29 amino acids in the third intracellular loop,$^{70-72}$ and seem to have identical pharmacology but their presence in various cerebral tissues differ.$^{70,73}$ Hence, a difference in functionality is likely still to be found. The long and the short forms of the D$_2$ receptor (i.e., D$_{2L}$ and D$_{2S}$) consist of 443 and 414 amino acid residues (see Table 1.2), respectively. The genes of the dopamine receptor superfamily can be divided into two different categories: 1) intronless genes that codes for the D$_1$-like receptors and 2) genes with their coding sequences in discontinuous DNA segments (exons) separated by sequences (introns) that do not form a part of the mature mRNA. The latter category of genes are found in the D$_2$-like family of dopamine receptors, which explains the occurrence of a long and a short form of the DA D$_2$ receptor. In the biosynthesis of mRNA a mechanism called alternative
splicing, in which a given exon in the pre-mRNA is either present or absent in the final mRNA, result in two different proteins coded by the same gene.\(^6\)

![Chemical structures]

Table 1.2: Specifics of the human DA receptors.

<table>
<thead>
<tr>
<th></th>
<th>D(_1)</th>
<th>D(_2)</th>
<th>D(<em>{2L}/D(</em>{2S})</th>
<th>D(_3)</th>
<th>D(_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td># a.a.</td>
<td>446(^a)</td>
<td>477(^b)</td>
<td>443(^c)/414</td>
<td>400(^d)</td>
<td>467(^e)</td>
</tr>
<tr>
<td>mRNA location</td>
<td>neostriatum</td>
<td>hypothalamus, hippocampus</td>
<td>neostriatum</td>
<td>isl. of Calleja, n. accumbens</td>
<td>frontal cortex, hippocampus</td>
</tr>
<tr>
<td>Adenyl cyclase</td>
<td>stimulates</td>
<td>stimulates</td>
<td>inhibits</td>
<td>?</td>
<td>inhibits</td>
</tr>
<tr>
<td>Agonists</td>
<td>SKF38393 (36)</td>
<td>SKF38393 (36)</td>
<td>bromocriptine (12)</td>
<td>PD128907(13)(^f)</td>
<td>PD1(16)(^g)</td>
</tr>
<tr>
<td></td>
<td>SKF82958 (37)</td>
<td>bromocriptine (12)</td>
<td>PD128907(13)(^f)</td>
<td>7-OH-DPAT (25)</td>
<td>PD1(16)(^g)</td>
</tr>
<tr>
<td>Antagonists</td>
<td>SCH23390 (35)</td>
<td>SCH23390 (35)</td>
<td>haloperidol (4)</td>
<td>AJ76(14)(^h)</td>
<td>clozapine (6)</td>
</tr>
<tr>
<td></td>
<td>UH232(15)(^h)</td>
<td>UH232(15)(^h)</td>
<td>UH232(15)(^h)</td>
<td>PD2(17)(^i)</td>
<td>PD2(17)(^i)</td>
</tr>
</tbody>
</table>

\(^a\) Ref.57,59,60,74; \(^b\) Ref.75-77; \(^c\) Ref.72,78-86; \(^d\) Ref.87-89; \(^e\) Ref.90-93; \(^f\) Ref. 94; \(^g\) Ref.95; \(^h\) Ref. 96; \(^i\) Ref.97
1.3 Computer-Assisted Molecular Design

The last decade new tools have become available for drug design including computational chemistry, high-throughput screening and combinatorial chemistry. Still, no matter how advanced our technology have developed or how fast new compounds can be synthesizes and tested, a medicinal chemist simply has two major questions to answer: Do I understand the structure activity-relationship for this series of compounds and which compound should I synthesize next? In order to provide answers to these questions two general work procedures are followed. First, one may try to build a 3D model, e.g., homology modeling of the target protein and, accordingly, dock ligands into the active site of the protein. Simply, a potent ligand fits into the receptor while an inactive ligand does not. The second approach is more basic where ligands, initially, are superimposed on mutual tentative interaction points (e.g., lone pair of electrons and midpoints of aromatic rings) with the receptor. Consequently, attempts to explain the potency of the ligands by comparison of their structures and their relative 3D orientations can be done. This approach is generally called an “active analogue approach” or structure-activity relationship (SAR).

Homology Modeling

Until recently, we had no knowledge about the amino acid sequence and less knowledge about the secondary structure of G protein-coupled receptors (GPCRs). Therefore, computational chemists have utilized the structure elucidated for bacteriorhodopsin (see above), although the sequence homology is poor, to create 3D models of GPCRs. Recently, the dopamine D2 receptor was constructed based on the coordinates from bacteriorhodopsin itself.

The first problem in GPCR modeling is to determine which amino acid residues that reside in the transmembrane domains, or stated differently, the alignment of the seven helices. Computer programs that can perform this kind of alignment, are available. The dopamine receptors presented in Table 1.3 were downloaded from EMBL in Heidelberg and aligned using the ‘whatif’ program. The alignments were by no means perfect and refinements were applied manually as presented in Table 1.3.

Dopamine agonists are believed to interact with the third and the fifth transmembrane domain, while antagonists additionally interact with the seventh domain as reviewed by Savarese et al. and Teeter et al. The endogenous neurotransmitter, dopamine (8), most likely interacts with three amino acid residues located at helices three and five. The protonated nitrogen forms a salt-
bridge with Asp_{114} on helix three while, simultaneously, hydrogen bonds are formed between the \(m\)- and \(p\)-OH and Ser_{194} and Ser_{197} on helix five, respectively.

### Table 1.3 Transmembrane domain amino acid sequences from the five dopamine receptor proteins. Alignment was originally taken from EMDL but was refined, manually, to obtain maximum sequence overlap between the receptors.

<table>
<thead>
<tr>
<th>TM1</th>
<th>Asp_{80}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>A T L L T L L I A V I V F G N V L V C M A V S</td>
</tr>
<tr>
<td>D3</td>
<td>A L Y S C A L L I L A I V F G N G L V C M A V I</td>
</tr>
<tr>
<td>D4</td>
<td>L V G G V L I G A V L A G N S L V C V S V A</td>
</tr>
<tr>
<td>D1</td>
<td>A C F L S L L I L S T L L G N T L V C A A</td>
</tr>
<tr>
<td>D5</td>
<td>A C L L T L L I I W T L L G N V L V C A A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TM2</th>
<th>Asp_{114}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>L I V S L A V A D L L V A T L V M P W V V Y L E</td>
</tr>
<tr>
<td>D3</td>
<td>L V V S L A V A D L L V A T L V M P W V V Y L E</td>
</tr>
<tr>
<td>D4</td>
<td>S V I V S L A A A D L L L A L L V L P L F V Y</td>
</tr>
<tr>
<td>D1</td>
<td>F F V I S L A V D L L L V A V L V M P W K A V A E</td>
</tr>
<tr>
<td>D5</td>
<td>V F I V S L A V D L F V A L L V M P W K A V A E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TM3</th>
<th>Ser_{194}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>I F V T L D V M M C T A S I L N L C A I S I</td>
</tr>
<tr>
<td>D3</td>
<td>V F V T L D V M M C T A S I L N L C A I S I</td>
</tr>
<tr>
<td>D4</td>
<td>A L M A M D V M C T A S I F N L C A I S V</td>
</tr>
<tr>
<td>D1</td>
<td>N I W V A F D I M C T A S I L N L C V I S V</td>
</tr>
<tr>
<td>D6</td>
<td>V W V A F D I M C T A S I L N L C V I S V</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TM4</th>
<th>Ser_{194}Phe_{198}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>P A F V V Y S S I V S F Y V P F I V T L L V Y I</td>
</tr>
<tr>
<td>D3</td>
<td>P D F V I Y S S V S F Y L P F G V T V L V Y A</td>
</tr>
<tr>
<td>D4</td>
<td>Q L L I G A T W L S A A A P V L C G L N</td>
</tr>
<tr>
<td>D1</td>
<td>Y V V Y S S V C S F F L P C P L M L L L Y W</td>
</tr>
<tr>
<td>D6</td>
<td>V M V G L A W T L S I I S F I P V Q L S W</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TM5</th>
<th>Phe_{190}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>M L V A I I V G V F I I C W L P F F I T H I L N</td>
</tr>
<tr>
<td>D3</td>
<td>M V A I I V G V A F I V C W L P F F L T H V L</td>
</tr>
<tr>
<td>D4</td>
<td>V L P V V V G A F L L C W T P F F V V H I</td>
</tr>
<tr>
<td>D1</td>
<td>T L S V I M G V F V C C W L P F F I L N C I L</td>
</tr>
<tr>
<td>D6</td>
<td>T L S V I M G V F V C C W L P F F I L N C M V</td>
</tr>
</tbody>
</table>

From Table 1.3 it is also clear that these amino acid residues are preserved in all five receptor subtypes, supporting this hypothesis. (In order not to confuse, the numbering of the amino acid residues starts with number one from the extra-cellular end, and is valid for the DA \(D_{2L}\) receptor if not stated otherwise.) Trump-Kallmeyer et al.\(^{107}\) also suggested that the OH-group from Tyr_{416} on
helix seven helps to stabilize the transmitter-receptor complex by interacting with the charged dopamine nitrogen. Another interesting feature is the narrow aromatic cleft, defined by Phe$_{390}$ on helix six and Phe$_{198}$ on helix five, that may interact with the flat aromatic part of catecholamine related ligands. Finally, site-directed mutagenesis study have shown$^{73,114,115}$ the importance of Asp$_{80}$ in the regulation of D$_2$ affinity for drugs, coupling to adenylate cyclase and sensitivity to Na$^+$ and pH. Sodium decreases the binding of D$_2$ agonists (e.g., dopamine (8))$^{114,115}$ and increases binding for some substituted benzamides (e.g., epidepride (18) and sulpiride (40))$^{114,115}$ but does not affect binding of other D$_2$ antagonists (e.g., spiperone (19)).$^{68,115}$ This amino acid residue, Asp$_{80}$, is preserved in all five dopamine receptors (see Table 1.3) and Teeter $et$ $al.$$^{68}$ called it the ‘sodium site’.

| Structure-Activity Relationships (SAR) |

Although homology modeling is a direct consequence of recent analytical refinements (e.g., cloning, crystallization and X-ray techniques) and increased computer efficiency the technique is not, as yet, able to explain the structure-activity relationship (SAR) for most ligands. Computer models of receptors cannot, for instance, account for the obvious flexibility of the receptor protein, nor predict the conformational change of the receptor caused when an agonist binds to the active site. There are, simply, to many uncertain parameters that we cannot simulate. Therefore, traditional methods where the activity and inactivity of ligands are explained by superimposition of mutual and possible interaction points with the receptor, are as valid today as for ten years ago. This approach is, generally, referred to as “the active analogue approach”,$^{103}$ upon which many recent computational methods rely (e.g., CoMFA$^{117}$). During the last two decades several attempts to construct models that can explain the SAR of dopamine receptor ligands have been presented. In the following, a review of the most successful agonist and antagonist models will be given, and the SAR of dopaminergic compounds will be discussed.

| The McDermed Dopamine Receptor Concept |

One of the first dopamine receptor models was presented by McDermed $et$ $al.$$^{118}$ in 1979, and was two dimensional (Figure 1.4). The model was based on two tentative interaction points with the receptor: one for the basic nitrogen atom and one for the hydroxyl group $meta$ to the ethylamine chain. Additionally, a steric boundary defined the receptor excluded volume and explained the low affinity for ligands possessing steric bulk (i.e., a substituent or a part of the molecule) protruding into this region (see below). McDermed and coworkers rationalized their model on the fact that dopamine receptor agonists like 20 ((6aR)-apomorphine) and 22 ((2R)-5,6-di-OH-dipropylaminotetralin) have the dopamine moiety in its $\alpha$-rotameric conformation, while dopamine receptor agonists like 21 ((6aS)-isoapomorphine) and 23 ((2R)-6,7-di-OH-dipropylaminotetralin) have the dopamine moiety in the $\beta$-rotameric conformation, and have to be flipped and rotated in order to fit properly into the presumed active site, as illustrated in Figure 1.4. The same model could
rationalize the inactivity of 21 since one of the aromatic rings protrudes into the steric boundary and render binding to the tentative active site impossible (Figure 1.4).

**Figure 1.4** The McDermed receptor concept defined by two tentative interaction points with the receptor and a steric boundary preventing binding with “bulky” ligands.

The mono-hydroxy DPATs*, 24 and 25, exhibit the same stereo-isomer as 22 and 23, with the S-enantiomer of the 5-OH-DPAT (the α-rotameric conformation) and the R-enantiomer of the 7-OH-DPAT (the β-rotameric conformation) being the more potent isomers.\(^{119}\) Both conformers display affinity to the DA D\(_{2L}\) and the D\(_3\) receptors.\(^{120}\) However, the α-conformer (24) display less preference for the DA D\(_3\) receptor as compared to the β-conformer (25), with a D\(_{2L}/D_3\) ratio of 26 and 60, respectively. No preference for the α- or β-conformer was observed for the DA D\(_{4.2}\) receptor.\(^{120}\) The triflate analogues, 26 and 27, displayed similar dopamine agonist profiles as their hydroxy counterparts, 24 and 25, both in *in-vitro* binding assays and in *in-vivo* biochemical and behavioral assays in rat.\(^{121}\) The *in vitro* affinity towards the DA receptors were lower for the triflated compounds, as compared to the hydroxy compounds. Interestingly, 26 was found to have mixed

---

* DPAT is short for di-(n-propyl)-aminotetralin.
DA/5-HT\textsubscript{1A} properties after oral administration not observed after subcutaneous administration. This suggest that active metabolite(s) may be formed. (Similar findings were found for 8-OTf-DPAT, a potent 5-HT\textsubscript{1A} receptor agonist where the dominating metabolite, the mono-propyl analogue, turned out to be more potent \textit{in vivo} than 8-OTf-DPAT itself.\cite{121})

![Figure 1.5](image)

\textbf{Figure 1.5} The Extended McDermed model as presented by Liljefors \textit{et al.} (a) Molecules 28 and 29 superimposed with the receptor excluded volume marked. (b) The same superimposition as in (a) but now flipped 90 degrees forward. For clarity, only hydrogens on the propyl groups are shown.

\textbf{The Extended McDermed Receptor Model}

Liljefors\textsuperscript{122,123} and Wikström\textsuperscript{124} extended the McDermed dopamine receptor concept to a receptor model able to explain the activity and inactivity of a larger number of structurally related compounds. Initially, two modes of receptor interaction and two \(N\)-alkyl directions were defined, by superimposition of the nitrogen atoms and the midpoints of the aromatic rings of compounds 28 and 29 in their calculated lowest energy conformations\textsuperscript{122} (Figure 1.5). The aromatic rings were constrained to be coplanar.

![Chemical structures](image)

\textbf{28 (4aS,10bS)} \hspace{2cm} \textbf{29 (4aR,10bR)} \hspace{2cm} \textbf{30 (S)-3-PPP} \hspace{2cm} \textbf{30 (R)-3-PPP}

The sizes and orientations of the \(N\)-alkyl substituents are, according to the authors,\textsuperscript{122,123} of crucial importance in order to understand the dopaminergic activity and presynaptic selectivity. The model in Figure 1.5(a) defines one “upward” and one “downward” direction for the \(N\)-alkyl substituents. The “downward“ direction is a narrow cleft complementary to maximum a n-propyl group, while the “upward” direction is sterically less restricted.\textsuperscript{122} The model was rationalized since compounds 20 and 29, both having a \(N\)-n-propyl substituent, are potent pre- and postsynaptic agonists and display high enantioselectivity. The \(N\)-n-butyl analogues of the same compounds were
found inactive at both pre- and post-synaptic receptors,\textsuperscript{122,124} most likely due to the fact their \(N\)-n-butyl substituents are too large to fit into the “downward” propyl-cleft. The potency of the \(N\)-n-butyl analogue\textsuperscript{125} of 28, could be explained since its \(N\)-n-butyl protrudes “upwards” in the less sterically restricted direction when aligned properly into the “active site” (Figure 1.5(a)). In fact, compounds with as large substituents as phenylethyl or thiopenethyl groups directed “upwards” may still be active.\textsuperscript{122,126,127}

Liljefors \textit{et al.}\textsuperscript{122,123} also investigated the biological active conformations of the enantiomers of 3-PPP, \textit{i.e.}, compounds (\textit{S})-30 and (\textit{R})-30. Pharmacologically, the (\textit{R})-30 enantiomer displays classical pre- and postsynaptic receptor agonist properties, while the (\textit{S})-30 enantiomer is a presynaptic agonist with postsynaptic antagonistic properties (Table 1.4). In an attempt to explain the opposed profiles of the enantiomers (\textit{R})-30 was fitted in the model with the \(N\)-n-propyl directed in the “downward” propyl-cleft and (\textit{S})-30 with the \(N\)-n-propyl directed “upwards”. The superimposition of (\textit{R})-30 on (4a\textit{R},10b\textit{R})-29 exerts an excellent fit explaining the pre- and postsynaptic properties of (\textit{R})-29. The (\textit{R})-30 does not fit in its “global energy minimum conformation” but the \(N\)-n-propyl directed in the lipophilic “propyl-cleft” helps to stabilize the compound in the agonist conformation. Liljefors and coworkers\textsuperscript{122} found and defined two different conformations for compound (\textit{S})-30 to take: one agonist and one antagonist conformation. The requirement to activate the postsynaptic receptors seems to be a demand for lipophilicity around the nitrogen. The (\textit{S})-30 has no n-propyl directed into the “propyl-cleft”. Obviously, the methylene group alfa to the nitrogen in the piperidine ring that is directed into the “propyl-cleft” is not lipophilic enough to maintain the (\textit{S})-30 in a postsynaptic activating conformation. Thus, (\textit{S})-30 does not activate postsynaptic receptors. In compound (4a\textit{R},10b\textit{S})-28, however, the methylene group is maintained in the “propyl-cleft”, since the OHBJ/Q skeleton is rigid, and the postsynaptic receptors can be activated. The (\textit{S})-30 can, as well, assume a low energy conformation that fits into the model in Figure 1.5 and explain its presynaptic agonist properties.

The model in Figure 1.5(a) cannot explain the inactivity of compounds 31 and 32 since the van der Waal volume of 30 fits perfectly while the volume from 32 is too large. The orientations of the propyl groups in Figure 1.5 which can be oriented either in an anti or a gauche conformation with respect to the nitrogen lone pair of electrons provides an explanation. Liljefors \textit{et al.}\textsuperscript{123} concluded that if one propyl assumes an anti conformation the other one must be gauche and \textit{vice versa}. If the “downward” oriented propyl group is oriented in an anti conformation (Figure 1.5(b)) the steric boundary in front of the nitrogen atom becomes more narrow as compared to in Figure 1.5(a) and the propyl cleft is located above the plane (Figure 1.5(b)). In this improved model, Figure 1.5(b), neither of the inactive compounds 31 and 32 fit while the active compounds in Table 1.4 do.
Table 1.4 Intrinsic activity of the ligands discussed

<table>
<thead>
<tr>
<th>compound</th>
<th>presynaptic agonism</th>
<th>postsynaptic agonism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ED$_{50}$, nmol/kg</td>
<td>motor activity$^b$</td>
</tr>
<tr>
<td>limbic</td>
<td>striatum</td>
<td>dose µmol/kg sc</td>
</tr>
<tr>
<td>(R)-20$^{c,d}$</td>
<td>190</td>
<td>2.3</td>
</tr>
<tr>
<td>(S)-24$^d$</td>
<td>3.7</td>
<td>0.31</td>
</tr>
<tr>
<td>(R)-25$^d$</td>
<td>9.5</td>
<td>0.31</td>
</tr>
<tr>
<td>(S)-26$^e$</td>
<td>830</td>
<td>12.5</td>
</tr>
<tr>
<td>(R)-27$^c$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(S)-28$^d$</td>
<td>14</td>
<td>1.30</td>
</tr>
<tr>
<td>(4aS,10bS)-29$^d$</td>
<td>4</td>
<td>1.06</td>
</tr>
<tr>
<td>(4aR,10bR)-29$^e$</td>
<td>800</td>
<td>213</td>
</tr>
<tr>
<td>(S)-30$^d$</td>
<td>1000</td>
<td>1300</td>
</tr>
<tr>
<td>(R)-30$^c$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>33$^f$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>34$^f$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$^a$ Measured indirectly as inhibition of DA synthesis rate (see ref 124); $^b$ Motor activity measured in motility meters on raserpinized rats (see ref 124); $^c$ Data taken from ref 122; $^d$ Data taken from ref 124; $^e$ Data taken from ref 128; $^f$ Data taken from ref 129; $^g$ Values expressed as percentage of saline controls; mean ± SEM.

It was demonstrated for the triflated aminotetralines (e.g., 8-OTf-DPAT and 24) that the triflate group induced biochemical changes as compared to their hydroxy analogues (see above). This was confirmed also for the OHB[f]Qs in experiments performed by Sonesson et al.$^{121,129}$ They found that compound (±)-33 was inactive as an agonist, even at high doses (50 µmol/kg), although the hydroxyl analogue ((±)-28) is a potent agonist. Instead, presynaptic DA receptor antagonistic properties was demonstrated for (±)-33, by the increase of DOPA levels in nonpretreated habituated rats.$^{129}$ Additionally, (±)-33 also decreased significantly the locomotor activity to 56 ± 4 %. Obviously, (±)-33 is a compound with postsynaptic agonistic and presynaptic antagonistic properties. The more flexible analogue, compound (±)-34, did not portray the same affinity for postsynaptic receptors as (±)-33.

The triflate group has, obviously, great impact on the phenyl ring due to its electron withdrawal ability$^{130}$ distorting the conjugated aromatic system. However, information concerning the physicochemical properties of the (aryl-)triflate group is very sparse, and to date, no X-ray crystallographic structure of a (aryl-)triflate group has been resolved. In Chapter 3, the triflate group will be discussed further, also in comparison with other sulfonaryl esters.
Introduction to the Medicinal Chemistry of Schizophrenia

Dopamine D<sub>1</sub> Agonist and Antagonist Models

The McDermed receptor concept is an example of an early working model of the dopamine receptor proven useful in the design of new potent ligands. Today, however, with five different dopamine receptor subtypes known and sophisticated molecular modeling tools available we aim for models that enable us to explain and understand the structure-activity relationships within as well as in-between different receptor subtypes.

The SAR of the DA D<sub>1</sub> receptor subtype has extensively been scrutinized in the literature during the last decade.<sup>131-133</sup> A challenge has been, and still is, to fully explain the ligand-receptor interactions of the potent benzazepines. To date, no theory exist that explains why compound <sup>35</sup> (SCH23390) is a selective D<sub>1</sub> antagonist while the structurally similar compound, <sup>36</sup> (SKF38393), is a potent agonist at DA D<sub>1</sub> receptors. Compound <sup>35</sup> has a selectivity for the DA D<sub>1</sub> over DA D<sub>2</sub> receptors with a factor 2093, and fully inhibits dopamine stimulated adenylyl cyclase<sup>131</sup> (K<sub>i</sub> = 0.47 ± 0.06 nM). In contrast, compounds <sup>36</sup>,<sup>37</sup> and <sup>38</sup> are reported as potent and selective agonist for the D<sub>1</sub> receptors, all able to activate adenylyl cyclase to the same extent as dopamine (8).<sup>133</sup>

![Chemical Structures](https://example.com/structures.png)

Pettersson et al.<sup>134,135</sup> have through extensive conformational analysis and electrostatic potential calculations proposed biologically active conformations for <sup>35</sup> and <sup>36</sup>. Both compounds were proposed to have their seven rings in chair conformations with the N-methyl (or N-H) and the 1-phenyl rings in (pseudo-)equatorial positions. Additionally, the plane of the 1-phenyl rings do not deviate more than 30 degrees from being orthogonal to the plane of the catechol aromatic ring in the main skeleton, since the energy penalty for such rotations would be too large.<sup>134</sup> In the case of compound <sup>38</sup> and other rigid compounds<sup>134</sup> deviations larger than 30 degrees are possible (i.e., an accessory phenyl ring closer to coplanar with the catechol phenyl ring), since no energy penalties are involved. As further support of this hypothesis, probing of the electrostatic surroundings of <sup>35</sup> by means of the GRID program<sup>136</sup> with two different probes (e.g., the cationic NH<sub>3</sub> probe and the anionic carboxy oxygen probe), indicates that the hydroxyl group and the accessory phenyl ring may interact with the same receptor site via electrostatic interactions.<sup>135</sup>

Mottola et al. suggested the following agonist pharmacophore derived from the analogues of benzazepine and <sup>38</sup>: 1) the two catechol hydroxyl groups; 2) the nitrogen atom (ca. 7 Å from the m-hydroxyl) and 3) the accessory phenyl ring (ca. 5 Å from the catechol ring). Despite the calculations performed by Pettersson et al.<sup>134</sup> Mottola et al.<sup>133</sup> defined the accessory phenyl ring, in their
pharmacophore, to be close to coplanar with the catechol ring. Finally, Mottola et al. conclude that alkylation on the nitrogen diminishes the affinity for DA D$_1$ receptors, defining a steric receptor boundary in that direction.

Charifson and coworkers$^{131,132}$ proposed a pharmacophore for DA D$_1$ antagonists, very similar to the agonist pharmacophore of Mottola et al. but with the second hydroxyl group replaced by a chlorine atom, derived from analogues of 35 and the tetrahydroisoquinoline 39. Also in the tetrahydroisoquinoline series elongation of the nitrogen alkyl (i.e., longer than methyl) group was not favorable for DA D$_1$ receptor binding, although the interatomic distance between the chlorine and the nitrogen is decreased as compared to the benzazepines.$^{132}$ Additionally, the stereochemistry for the tetrahydroisoquinolines is reversed as compared to the benzazepines; both (R)-SCH23390 (35) and (S)-39 are potent and selective DA D$_1$ antagonists.$^{132}$

Dopamine D$_2$ Antagonist Pharmacophores

Among dopamine antagonist, benzamides are generally characterized by a high selectivity for the DA D$_2$ receptor subtype with low affinity for the DA D$_1$ receptor and other non-dopamine receptors. Additionally, the pharmacological profile of benzamides in general is unique, with low propensity to induce neurological side effects (e.g., extrapyramidal syndromes and tardive dyskinesia)$^{137}$ and effective in the treatment of negative symptoms in schizophrenia. Thus, it seems a lot to gain by learning about the SAR of benzamides. In 1981 Olson et al.$^{138}$ proposed a pharmacophore based on
potent, but flexible, dopaminergic compound such as sulpiride (40), haloperidol (4), pimozide (43) and molindone (41). The pharmacophore comprised four different elements as pictured in Figure 1.6 (adopted from Olson et al.): 1) the benzamide phenyl ring (π-π interaction); 2) the amide carbonyl oxygen atom (π-π interaction) or, alternatively, an aromatic moiety (π-π interaction); 3) the basic nitrogen atom (‘NH-‘COO’ interaction) and 4) a lipophilic tail corresponding to the heterocyclic moiety of pimozide (Figure 1.6, 43). Interestingly, the authors introduce the carbonyl oxygen as an isosteromer to a phenyl ring in the ligand-receptor interaction, supported by a crystal structure where a similar interaction has been observed.

In order to validate their model Olsen et al. attempted to include the pharmacophoric elements in one single structure, and the resulting (−)-(4aR,8aR)-piquindone (42) turned out to have similar properties as haloperidol (4) but with significant decreased cataleptogenic liability. For a computational chemist a rigid and potent compound (Table 1.5), like 42, is ideal to utilize as a starting point for molecular modeling and SAR studies. Accordingly, Rognan et al. used (−)-(4aR,8aR)-piquindone (42) for elucidation of the SAR from a number of optically active DA D2 antagonists.

From the crystallographic structure of (−)-42 three initial pharmacophoric elements were defined (Figure 1.7): an aromatic ring Ar1 (pyrrole); a dipole coplanar to Ar1 (amide carbonyl); a basic nitrogen atom with a lone pair of electrons directed orthogonal to the Ar1 plane towards a dummy atom Du representing the receptor site (e.g., an aspartic acid). The aromatic pharmacophoric element Ar2 was defined by superimposition of low energy conformations of domperidone (44) with the crystallographic structure of (−)-42 on the chlorophenyl ring (from domperidone), the pyrrole ring (from piquindone) and the Du in the direction of the nitrogen lone pair of electrons, respectively. In one high quality fit the second phenyl ring of 44 is protruding into and defines the Ar2 pharmacophore element (Figure 1.7).
Table 1.5 Affinities of the discussed antagonists for the dopamine D$_2$ receptor subtype

<table>
<thead>
<tr>
<th>Compound</th>
<th>$K_i$ (nM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 chlorpromazine</td>
<td>4.1</td>
</tr>
<tr>
<td>4 haloperidol</td>
<td>1.8</td>
</tr>
<tr>
<td>19 spiperone</td>
<td>0.04</td>
</tr>
<tr>
<td>6 clozapine</td>
<td>220</td>
</tr>
<tr>
<td>40 (S)-sulpiride</td>
<td>7.6</td>
</tr>
<tr>
<td>42 (4aR,8aR)-piquindone</td>
<td>4.1</td>
</tr>
<tr>
<td>43 pimozide</td>
<td>2.5</td>
</tr>
<tr>
<td>44 domperidone</td>
<td>0.7</td>
</tr>
<tr>
<td>45 (R)-DO749</td>
<td>2.4</td>
</tr>
</tbody>
</table>

$^a$ Data from reference 137 obtained by inhibition of $[^{125}]$iodosulpiride to rat striatal membranes.

Another interesting feature of this series of compounds is the stereochemistry. In benzamides with short substituents (e.g., ethyl, propyl and allyl) attached to the pyrrolidine nitrogen atom the activity resides solely, in the (S)-enantiomer$^{137,139}$ (see Table 1.5, (S)-sulpiride). If, however, the substituent is replaced by a benzyl group (45) the stereochemistry is switched; the activity now resides in the (R)-enantiomer$^{137,139}$ (see Table 1.5, (R)-DO749). In order to account for the stereochemistry in the pharmacophoric model (Figure 1.7) a third aromatic element Ar$_3$ was defined for large pyrrolidine substituents and a small hydrophobic pocket HP for small pyrrolidine substituents for benzamides with (R) and (S) configuration, respectively.

Later, in Chapters 4, 5 and 6, two different series of benzamides will be discussed as they form the bases for a couple of publications$^{140,141}$ involving 3D QSAR and multivariate statistical analysis. Furthermore, structure-activity relationships of benzamides will be dealt with also in these chapters.

1.4 Quantitative Structure-Activity Relationships (QSAR)

So far, two different approaches to design new drugs have been discussed: homology modeling and the active analogue approach.$^{103}$ Despite sophisticated computational tools,$^{142,143}$ with all known dopamine receptors cloned and site mutagenesis techniques available, the active analogue approach is still the most widely used approach to design new drugs. A speculative reason for this may be the fact that in receptor modeling too many parameters need to be estimated: conformational flexibility in the receptor protein, solvent dependency (dielectric constant), pH at the active site (protonated ligands or not), water improved receptor binding$^{144}$, induced fit$^{145,146}$ and many more. Hence, most drugs are developed with ideas based on structure-activity relationships and simple molecular modeling studies. In 1964 Hansch and Fujita$^{147}$ introduced a way to predict biological activity from theoretically derived molecular descriptors, often referred to as Hansch-analysis. The ideas behind the Hansch analysis are as valid today as in 1964 since theoretically (not necessarily) generated descriptors are assumed to be independent of the ligands conformations. More important, if the
predictive ability of the model is high enough, valuable time will not be spent synthesizing inactive compounds.

**Physicochemical Molecular Descriptors**

The applicability of physicochemical descriptors is manifold. First, prediction of the biological activity (e.g., receptor affinity) as introduced by Hansch et al.\(^\text{147}\) (see below). Second, they enable us to search large molecular databases (e.g., Chemical Abstracts structural database) by simply define a physicochemical profile for the target molecule.\(^\text{148-150}\) Third, recently developed techniques like combinatorial chemistry\(^\text{100-102}\) and high-throughput screening,\(^\text{98,99}\) call for methods in order to design the synthesis of the most diverse compounds and for the definition of new specific targets, respectively.

Examples of commonly used physicochemical descriptors are listed in Table 1.6. which may be divided into steric, electrostatic and hydrophobic (lipophilic) types of descriptors, although many of them are a combination of all three types. Some of these descriptors are listed in the literature, while some may be obtained through analytical experiments\(^\text{151,152}\) or through computational calculations.\(^\text{142}\) Steric descriptors like FW, L, B and VdWV (Table 1.6) are simply different attempts to quantify the molecular structure; electrostatic descriptors like \(\sigma_m\), HOMO or LUMO portray the electronic features of a compound (e.g., the influence of a specific substituent on an aromatic system) and the hydrophobic descriptors \(\pi\), logP and logD accounts for a molecules ability to, for instance penetrate the blood-brain barrier.\(^\text{153,154}\) Taken together, the array of descriptors collected for a specific molecule compares to the fingerprint from a human being, hence, it should be unique.

**Hansch Analysis**

In the early 1960s, Hansch and co-workers\(^\text{147}\) investigated the possibility of expressing a relationship between structural and physicochemical properties and biological activity, quantitatively. Typically, properties as logP, \(\sigma\), or \(E_s\) representing 1-octanol/water partition coefficient, the well-known Hammett constant and Tafts steric descriptors (Table 1.6), respectively, were used as descriptors. In the Hansch analysis, the biological activity \(\log(1/C)\) were correlated with the descriptors using Multiple Linear Regression (MLR, see Chapter 2), also called Ordinary Least Squares (OLS).\(^\text{155,156}\) The so-called Hansch equation (Equation 1.1) comprise the relationship established by Hansch \textit{et al.}, where \(a, b, c, d\) and \(e\) are constants obtained through regression analysis.

\[
\log\left(\frac{1}{C}\right) = -a(\log P)^2 + b\log P + c \sigma + dE_s + e \tag{1.1}
\]

As was emphasized by Van de Waterbeemd,\(^\text{157}\) Hansch analysis is a method aiming at describing the relationship between only a few variables and the biological activity and should not be considered too much as a predictive model. By employing MLR for the regression analysis a couple of crucial items need to be considered: 1) keep the ratio of compounds to variables greater than approximately
five and 2) multicollinearity may cause spurious solutions. Today, methods to circumvent these problems are available, where the original variables are replaced by underlying orthogonal latent variables (i.e., PCR and PLS in Chapter 2). The predictive ability of a model may then be validated with crossvalidation\textsuperscript{155} or by predictions of an external test set.

: Molecular Diversity and Experimental Design

![Figure 1.8](image)

\textbf{Figure 1.8} The basic skeleton of trans-OHB[f]Qs (a) and 6-methoxybenzamides (b) used by Nilsson \textit{et al.} and Norinder \textit{et al.}, respectively.

In Chapter 3, physicochemical descriptors are employed to guide the selection of which compounds to synthesize.\textsuperscript{158} From a library of a 88 tentative OHB[f]Qs (Figure 1.8(a)), only the 15 most diverse compounds were selected to be synthesized by means of a factorial design in the Principal Properties (PPs). The PPs are score-vectors obtained from a Principal Component Analysis (PCA, see Chapter 2) of the physicochemical descriptors, and comprise the variation that significantly discriminate between the compounds. Norinder \textit{et al.}\textsuperscript{159} used physicochemical descriptors in order to increase the understanding of the structure-activity relationships of a series of benzamides (Figure 1.8(b)). They used a fractional factorial design in the first three PPs, for the selection of 16 representative compounds, out of 70, for the training set. In the following regression analysis the original descriptors (not the PPs, thus) from the training set were correlated with the biological activity (pIC\textsubscript{50} for the DA D\textsubscript{2} receptor), using PLS. The remaining compounds were utilized as a test set in order to estimate the predictability of the PLS model. The profound difference in the approaches used in Chapter 3 and by Norinder \textit{et al.}, should be obvious. Nilsson \textit{et al.} generated descriptors for whole molecules, \textit{e.g.}, logP was used rather than the contribution from single substituents (\(\pi\)). The objective with that investigation was to select the most diverse compounds to synthesize, not necessarily to create a predictive model. In contrast, Norinder \textit{et al.} already had a large data set with compounds tested for biological activity and their aim was to elucidate also the influence of single substituent positions on the biological activity. Therefore, the training set was selected such that the diversity in each position R2, R3 and R5, in Figure 1.8(b), was maximized. The conclusions drawn from this investigation will be discussed further in Chapter 6, where this data set was analyzed with 3D QSAR (see next section) using multilinear PLS\textsuperscript{160} as regression method.

: Molecular Fields
The obvious extension of the Hansch analysis is modeling where molecular flexibility is considered. The concept of Comparative Molecular Field Analysis (CoMFA)\textsuperscript{161} as presented by Cramer \textit{et al.}\textsuperscript{117} in 1988 is such a method, commercially available as a module in the molecular modeling package SYBYL.\textsuperscript{142}

\textbf{Figure 1.9} The three dimensional grid used in CoMFA to generate molecular field descriptors. For clarity, grid points within the grid are omitted.
Table 1.6 The most commonly used molecular descriptors available in the literature or through computational calculations.

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Abbr.</th>
<th>type(^a)</th>
<th>ref.</th>
<th>Descriptor</th>
<th>short</th>
<th>type(^a)</th>
<th>ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula Weight</td>
<td>FW</td>
<td>ste</td>
<td></td>
<td>Ionisation constant</td>
<td>pK(_a)</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>Hammett constant</td>
<td>(\sigma_m)</td>
<td>ele</td>
<td>163</td>
<td>Swain-Lupton field</td>
<td>F</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>Hammett constant</td>
<td>(\sigma_p)</td>
<td>ele</td>
<td>163</td>
<td>Swain-Lupton res.</td>
<td>R</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>Tafts polar constant</td>
<td>(\sigma^*)</td>
<td>ele</td>
<td>157</td>
<td>VdWals Volume</td>
<td>VdWVV ste</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Tafts steric parameter</td>
<td>(E_s)</td>
<td>ste</td>
<td>157</td>
<td>VdWals Area</td>
<td>VdWA ste</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Hansch aromatic fragment</td>
<td>(\pi)</td>
<td>lip</td>
<td>162,163</td>
<td>Connolly Surface Vol.</td>
<td>CoVo ste</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Lipophilicity</td>
<td>logP</td>
<td>lip</td>
<td>157,162</td>
<td>Connolly Surface Area</td>
<td>CoAr ste</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Lipophilicity (pH=7.4)</td>
<td>logD</td>
<td>lip</td>
<td>157,162</td>
<td>Electronic Energy</td>
<td>ELEC ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Connectivity index (Randic)</td>
<td>(1\chi)</td>
<td>ste</td>
<td>157</td>
<td>Core-Core interaction</td>
<td>CoCo ste</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Connectivity index</td>
<td>(2\chi)</td>
<td>ste</td>
<td>157</td>
<td>Heat of Formation</td>
<td>HoFo ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Molar Refractivity</td>
<td>MR</td>
<td>ele</td>
<td>164</td>
<td>Ionization potential</td>
<td>HOMO ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Verloop Sterimol</td>
<td>L</td>
<td>ste</td>
<td>157</td>
<td>Electron affinity</td>
<td>LUMO ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Verloop Sterimol</td>
<td>B</td>
<td>ste</td>
<td>157</td>
<td>Dipole moment</td>
<td>Dipo ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Eudismic Index</td>
<td>El</td>
<td></td>
<td>165</td>
<td>Point charges</td>
<td>chaX ele</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Ionization constant</td>
<td>pK(_a)</td>
<td></td>
<td>162</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ele electronic; ste steric; lip lipophilicity

Molecular fields basically are three dimensional representations of the steric, electrostatic and hydrophobic surroundings of a molecule. A molecular field is generated by enclosing the molecule in a three dimensional grid (Figure 1.9) and assigning nonbonded interactions between a probe atom and the molecule in each grid point. Obviously, the difference between two different types of fields is the algorithm with which the nonbonded interactions are calculated. In SYBYL/CoMFA the steric field interaction energies (\(E_{ste}\)) are Lennard-Jones potentials, also referred to as steric 6-12 potentials,\(^{166}\) which are sensitive to changes in the distance between the probe and the atoms (\(r_i\)) as can be seen in Equation 1.2. \(N\) is the number of atoms in the molecule; \(A\) and \(B\) are constants characteristic for the probe atom type and the type of the \(i\)th atom in the molecule, respectively.

\[
E_{ste} = \sum_{i=1}^{N} \left[ \frac{A}{r_i^{12}} - \frac{B}{r_i^{6}} \right]
\]  

(1.2)

The electrostatic field interaction energies are less influenced by the distances between the probe and the atoms but instead the charge of the probe and the point charges of the atoms are important. In addition, \(E_{ele}\) is very sensitive to spatial dielectric behavior of the environment\(^{167}\) and a distance-dependent dielectric term has been proposed.\(^{168}\) The magnitude of the electrostatic potential (\(E_{ele}\)) between two ions with charges \(Q\) and \(q\) separated by a distance \(r\) is given by Coulomb’s law\(^{166}\):

\[
E_{ele} = \sum_{i=1}^{N} \frac{Qq_i}{K\zeta} \left[ \frac{1}{r} + \frac{1}{\sqrt{r^2 + 4s_q\zeta^2}} \right]
\]  

(1.3)

where \(N\) is the number of atoms in the molecule; \(Q\) is the charge of the probe atom; \(q_i\) is the point charge on the \(i\)th atom and \(K\) is a constant term. In this formula a homogenous protein phase and a homogenous solution phase with dielectrics \(\zeta\) and \(\epsilon\), respectively, are assumed to be present.\(^{167}\) The depth of each protein atom (\(s_p\)) in the protein phase is assessed by counting the number of
neighboring protein atoms whose nuclei lie within 4 Å. For the probe atom the depth ($s_Q$) is calculated similarly. Consequently, Equation 1.3 leads to an effective dielectric of $\zeta$ when the pairwise groups of atoms are so deep in the protein and so close together that the solvent effects can be neglected. However, when one or both of the atoms approach the surface of the protein the effective dielectric becomes ($\zeta + \varepsilon$)/2 since the term $4s_Qs_q$ is set to zero.

Prior to calculation of the electrostatic field point charges of the atoms need to be calculated. In the original CoMFA article charges were calculated by the method of Gasteiger and Marsili as was implemented in SYBYL. Today, however, several options are available although most often point charges are estimated by semi-empirical AM1 single point calculations.

There are also variations of the above described steric field. Kroemer et al. replaced the steric 6-12 potential interactions with atom-based indicator variables. That is, they assigned the values 30 or 0 kcal/mol to a grid point if an atom was present in the adjacent small cube or not. Similar, Floersheim et al. assigned values of either 1 or 0 to a grid point, depending on whether the grid point was within, or outside, the van der Waals radius of any atom in the target molecule.

In the GRID program a different and intuitively more appealing (authors comment) approach is used. Each grid point is assigned with the sum of three different non-bonded interactions, i.e., $E_{ste}$, $E_{ele}$ and $E_{hb}$, as in Equation 1.4. The two former terms are calculated as in Equations 1.2 and 1.3, while the hydrogen bonding contribution is calculated as in Equation 1.5. C and D are tabulated values for specific atoms; d is the distance between the atoms; m is usually four but the whole $E_{hb}$ term is set to zero when $\theta \leq 90^\circ$. If the probe group donates the hydrogen bond it is assumed that the probe can orient itself in order to form the most effective hydrogen bond, and the $\cos \theta$ is set to unity.

$$E_{tot} = \sum E_{ste} + \sum E_{ele} + \sum E_{hb}$$  \hspace{1cm} (1.4)

$$E_{hb} = \left[ C/d^6 - D/d^4 \right] \cos^m \theta$$  \hspace{1cm} (1.5)

Consequently, and in contrast to other methods (e.g., SYBYL/CoMFA), fields generated in the GRID program portrays the specifics of certain probes. Thus, it is the users task to select a probe that best reflects what needs to be investigated. For instance, a water molecule, a carbon atom or a Ca$^{+2}$ ion could be chosen to display hydrogen bonding, steric and electrostatic characteristics of the ligands, respectively. Goodford has demonstrated how GRID probes explicitly can reflect individual properties of specific chemical groups attached to the target molecule.

Fields generated in SYBYL or GRID are the most commonly used since they are commercially available. However, other molecular descriptors are available, e.g., Molecular Shapes, Molecular Lipophilicity Potentials, Molecular Similarity Indices (i.e., CoMSIA) and Molecular Similarity Matrices.

Independent of which program used to generate molecular fields, the following parameters need to be specified by the user: 1) the grid size; 2) the grid resolution, i.e., grid points per Å; 3) the type of probes and 4) the probe charge.

**Comparative Molecular Field Analysis (CoMFA)**

25
As stated above, the logical extension of the Hansch analysis is Comparative Molecular Field Analysis (CoMFA), where the physicochemical parameters are replaced or combined with field descriptors. In analogy with the Hansch analysis, in CoMFA the molecular fields are correlated with the biological activity. CoMFA takes the three dimensional conformations of the molecules into consideration, hence, several crucial items need to be considered.

First, in order to find low energy conformations, or rather the global minimum energy conformation of each compound under investigation, conformational analyses are conducted. According to the Boltzman distribution, a low energy conformation is more abundant than a high energy conformation and, consequently, also more likely to be involved in the ligand-receptor interaction.

Second, all molecules must be aligned in the same coordinate system and several options to perform this are possible. If a pharmacophore is available, one might choose to superimpose all molecules on mutual and likely interaction points with the receptor (see Chapter 4). The compounds could be docked into the active site of a receptor homology model (see Chapter 6) or, alternatively, the molecular fields could be superimposed in a least square manner. Independent of which alignment approach that is employed, in CoMFA the differences between the aligned molecular fields are correlated with the biological activity. Therefore, an alignment procedure where the global overlap between structurally related compounds (Chapter 4) are maximized, is likely to perform just as good as a more elaborated and rational alignment (Chapter 6). This is the case when the ligands are flexible and the rationale is pure statistical: If the alignment is not performed with maximized overlap between the molecules an increased level of insignificant variation, i.e., noise is inevitable. Noise may, or may not, affect the predictability detrimentally.

Third, molecular fields are generated first when the molecules are properly aligned.

In Figure 1.10, a typical CoMFA data set consisting of \( I \) molecules characterized with a single molecular field is shown. In order to perform the PLS analysis each grid, with the dimensions \( J, K \) and \( L \), is unfolded to form a row with \( JKL \) number of columns. Traditionally, only one response variable is considered in CoMFA, e.g., the affinity for a central dopamine receptor and, therefore, a bilinear PLS1 algorithm is used for the regression analysis. The theory covering the basics of PLS analysis is discussed in Chapter 2, but specific details important for CoMFA are pointed out in the following.
Figure 1.10 The unfolding of the molecular fields from the $N$ molecules into a matrix $X$ with $I$ rows and $JKL$ columns. In CoMFA, the biological activity is usually represented in one column ($y$). The number columns in $X$ equals the number of grid points in each grid.

In PLS, the original variables are replaced by latent variables, i.e., linear combinations of the original variables and, therefore, the number of objects should be larger than the number of latent variables included in the model (see Chapter 2). However, each additional component adds also insignificant variation to the model and should be added only if it improves the predictability. The predictability is usually estimated with crossvalidation as described in Chapter 2.

The results from a CoMFA model are often interpreted by studying contour plots of the PLS-coefficients $b_{\text{PLS}}$ (Equation 1.6, see also Chapter 2).

$$\hat{y} = b_1 x_1 + b_2 x_2 + \ldots + b_p x_p = X b_{\text{PLS}}$$

(1.6)

In SYBYL/CoMFA a steric and an electrostatic contour plot is generated, and each $b_{\text{PLS}}$ coefficient is multiplied with the standard deviation for the corresponding variable. In effect, it is an enhancement of the $b_{\text{PLS}}$ coefficients in grid points where the variation is large. This is performed in order to simplify the interpretations. The data set in Chapter 4 is analyzed using GRID/GOLPE which produces slightly different contour plots, since each field corresponds to a specific probe. For example, in the contour plot from a water probe, regions are revealed where hydrogen bonding is favorable for high affinity, considering steric, electrostatic and hydrogen bonding interactions, simultaneously. Molecular fields generated from the water (OH2), carbon (C3) and the calcium (CA+2) probes in the GRID-program, will be used for the description of the molecules included in the two data sets scrutinized in this thesis.

1.5 References

Chapter 1

38. Seeman, P. Dopamine Receptors and the Dopamine Hypothesis of Schizophrenia. *Synapse* **1987**, *1*, 133-152.
Chapter 1


79. Robakis, N.K.; Mohamadi, M.; Fu, D.J.; Sambamurit, K.; Refolo, L.M. Human Retina D_2 Receptor cDNAs have Multiple Polyadenylation Sites and Differ from a Pituitary Clone at the 5' Non-Coding Region. Nucleic Acids Res. 1990, 18, 1299.


111. EMBL in Heidelberg. http://swift.embl-heidelberg.de/7tm.


119. Copinga, S. PhD Thesis. The 2-Aminotetralin System as a Structural Base for New Dopamine- and Melatonin-Receptor Agents, Dept. of Medicinal Chemistry, University of Groningen, The Netherlands. 1994


Chapter 1


34