Supporting Information

for

From Virtual to Real Screening for D3 Dopamine Receptor Ligands

Evgeny Byvatov, Britta C. Sasse, Holger Stark, and Gisbert Schneider*

Contents

1. Brief description of experimental procedures (Synopsis)
2. Experimental results of the first screening round
3. Docking of ligands into a homology model
4. SVM training and active learning
5. Regression SVM model for predicting the $\log \frac{K_{i}^{D3}}{K_{i}^{D2}}$ ratio
6. Compound ranking based on the similarity to the reference compounds
7. Homology model of the D$_3$ receptor
8. Binding studies
9. References
1. Brief description of experimental procedures

Data sets and descriptors. 395 compounds with reported affinity at D2 and D3 receptors served as the SVM training sets. These compounds were synthesized analogues of BP897 and related structures reported in literature.[1] Each compound was represented by a fingerprint of 3PP pharmacophores using MOE version 2004.05 (Chemical Computing Group, Montreal). The individual 3PP pharmacophore was represented as a triangle. We considered all possible triangles with their vertexes located at atom centers. Presence or absence of a certain triangle defines the one or zero state of the corresponding bit of the fingerprint. Triangles were distinguished by the type of atom at vertexes and by the path length of their edges. The vertex was either donor (D) and planar (Dpl), acceptor (A) and planar (Apl), polar (P), or hydrophobic (H) and planar (Hpl) as defined by the atom-types implemented in MOE.[4] Lengths of the edges were calculated along the molecular graph, no estimation of the 3D structure of molecule was performed at this stage.

Binding studies. Human dopamine D2s and dopamine D3 receptors were expressed in stably transfected Chinese hamster ovary (CHO) cells.[2,3] In brief, radioligand binding screening was performed on cell membrane preparations in triplicate by using 0.2 nM [³H]spiperone (Amersham Biosciences, Freiburg, Germany). Nonspecific binding was determined in the presence of 10 µM BP897. For rapid screening the compounds have been tested at four concentrations (10 µM, 1 µM, 0.1 µM, and 0.01 µM) in two independent experiments. Competition binding data were analyzed using the software GraphPad Prism (San Diego, CA, USA), using a non-linear least squares fit. Ki values were calculated from the IC50 values according to the Cheng-Prusoff equation.[4]

2. Results of binding studies of the first screening round

Table S1. Dopamine receptor affinities of compounds from the first virtual screening round (from IBS catalogue)

<table>
<thead>
<tr>
<th>No.</th>
<th>Chemical Structure</th>
<th>$K_i(D_2)$ [µM]$^a$</th>
<th>$K_i(D_3)$ [µM]$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td><img src="image" alt="Chemical Structure S1" /></td>
<td>&lt;2</td>
<td>&lt;2</td>
</tr>
<tr>
<td>S2</td>
<td><img src="image" alt="Chemical Structure S2" /></td>
<td>&gt; 2</td>
<td>= 2</td>
</tr>
<tr>
<td>S3</td>
<td><img src="image" alt="Chemical Structure S3" /></td>
<td>2·6</td>
<td>2·6</td>
</tr>
<tr>
<td>S4</td>
<td><img src="image" alt="Chemical Structure S4" /></td>
<td>&gt;6</td>
<td>2·6</td>
</tr>
<tr>
<td>S5</td>
<td><img src="image" alt="Chemical Structure S5" /></td>
<td>&gt;6</td>
<td>2·6</td>
</tr>
<tr>
<td>S6</td>
<td><img src="image" alt="Chemical Structure S6" /></td>
<td>&gt;6</td>
<td>2·6</td>
</tr>
<tr>
<td>S7</td>
<td><img src="image" alt="Chemical Structure S7" /></td>
<td>&gt;6</td>
<td>&gt;6</td>
</tr>
<tr>
<td>S8</td>
<td><img src="image" alt="Chemical Structure S8" /></td>
<td>&gt;6</td>
<td>&gt;6</td>
</tr>
<tr>
<td>S9</td>
<td><img src="image" alt="Chemical Structure S9" /></td>
<td>&gt;6</td>
<td>&gt;6</td>
</tr>
<tr>
<td>S10</td>
<td><img src="image" alt="Chemical Structure S10" /></td>
<td>&gt;6</td>
<td>&gt;6</td>
</tr>
<tr>
<td>S11</td>
<td><img src="image" alt="Chemical Structure S11" /></td>
<td>&gt;6</td>
<td>&gt;6</td>
</tr>
</tbody>
</table>

$^a$ $K_i$ values were measured in CHO cells stably expressing hD$_{2a}$ and hD$_{3}$ receptors by using $[^3H]$spiperone (two experiments). All compounds were approximately aligned according to the basic nitrogen.
3. Docking of ligands into a homology model. To get an idea of a potential binding pose of the found actives, we constructed a homology model of the dopamine D₃ receptor. Docking of the compounds into a homology model of human D₃ receptor highlights two potential hydrogen bonds (Figure S1): one to Ser192, and a second one between the basic amine and Asp110. It was previously shown that mutation of Ser192 to Ala (S192A) leads to approximately ten-fold reduced ligand binding to the mutated D₃ receptor.[¹] The importance of optimal hydrogen bonding interaction between the hydroxyl group of the well-studied ligand $R^\ast(+)$-7-OH-DPAT and Ser192 is also supported by SAR data, which shows that replacement of the hydroxyl group by a methoxy group in $R^\ast(+)$-7-OH-DPAT reduces its binding affinity by 100-fold.[²]

A potential hydrogen bonding pattern between the protonated basic amine of the ligands and Asp110 is also visible in the homology model. It was previously demonstrated that the basic amine function is important for receptor-ligand interaction for many different GPCR ligands.[³] For the D₂ and D₃ receptors the interaction with Asp110 (Asp114 for D₂) is a generally accepted hypothesis.[⁴]

![Figure S1. A) Docking of compound 4 into a homology model of the human dopamine D₃ receptor. Parts of the predicted binding pocket are shown. Potential hydrogen bonds between the ligand and Asp110 and Ser192 are indicated by dotted lines. The predicted lipophilic pocket for part C of the molecule (cf. Figure 1) is represented by a Connolly surface. B) Superposition of docked structures of compounds 3-6. The overlapping of the basic amine, as well as part C of the molecules is observed. Potential hydrogen bond acceptors that potentially interact with Asp110 are in close vicinity. Compound 7 is not present in the alignment as it probably adopts a different mode of binding. (Note: compound numbering according to Table 1 and Figure 1 of the main manuscript).](image-url)
Residues corresponding Phe345 and Phe346 were shown to be important for ligand binding in many different GPCRs.\textsuperscript{[5]} Phe345 of the D\textsubscript{3} receptor corresponds to the Phe389 of D\textsubscript{2}; its mutation to Ala was shown to abolish the binding of several ligands.\textsuperscript{[6]} Although we observed these residues to be in contact with the docked ligand in the homology model, we cannot unambiguously identify face-to-face or face-to-edge aromatic interactions between rings of these two residues and the aromatic moiety of our compounds. This is easily explained by the inaccuracy of the constructed homology models due to low sequence identity (28\%) between the D\textsubscript{3} receptor and the rhodopsin template. We wish to stress that although the docking experiments were able to propose a common binding mode for several ligands (Figure S1b), and the model of the binding site is in accordance with receptor mutation studies, the homology model must be treated with great caution. Unarguably, homology models have their value in molecular modeling,\textsuperscript{[7]} but we wish to stress that one should consider our D\textsubscript{3} receptor model only as an “idea generator” potentially guiding the following steps of hit exploration and generation of structure-activity relationships.

Since the compounds selected were taken from public compound collections it is clear that some of the compounds were already used in other investigations e.g., \textbf{4} for adrenergic and serotonergic receptors (cf. Ref. 9 in main manuscript).

\textbf{4. SVM training and active learning.} For constructing SVM models we used the SVM-light package.\textsuperscript{[8]} Details of the SVM training protocol can be found elsewhere.\textsuperscript{[8,10]} The prediction of a trained SVM is given by Equation S1.

\[ f(x) = \sum_{i} \alpha_i K(x, x_i^{sv}) + b \text{, where } K(x, y) = ((x \cdot y)s + 1)^5 \]  

(S1)

The greater \( f \) the higher is the probability for a compound to be active. \( x \) and \( y \) are molecular fingerprint vectors, \( x_i^{sv} \) are support vectors, i.e. molecular fingerprints that define the exact shape of the separating SVM hyperplane. The kernel function \( K \) defines the complexity of the surface that will be constructed. We used a fifth order polynomial kernel for all SVM models. Kernel parameter \( s \) was optimized to achieve improved ranking of compounds.\textsuperscript{[10]} For active learning, we considered all compounds with \( K_i < 1 \ \mu M \) for D\textsubscript{2} or D\textsubscript{3} receptors (331 compounds) as active compounds; they were labeled as “Class” (C). 50,000 “Non-Class” (NC) substances were selected from the IBS collection (~240,000). The NC compounds were selected for minimal
distance to the SVM hyperplane allowing a more fine-grained re-sampling of the “near-active-compound” space. The resulting filter consisted of two consecutive SVM models: the first SVM model ranked all available IBS compounds with respect to the distance to the active reference set, the second SVM model re-orders the most promising candidates.

SVM training aimed at maximizing the enrichment factor, expressed by the percentage of active compounds retrieved in the top 1% of a ranked screened database. It was done by standard four-fold cross-validation.11

5. Regression SVM model for predicting the $\log \frac{K_{D3}}{K_{D2}}$ ratio. SVM-light was used to construct the model. The $<q^2>$ value was used as the criterion for optimization [Eq. (S2)]:

$$<q^2> = 1 - \frac{\sum_i (f^{i}_{\text{measured}} - f^{i}_{\text{predicted}})^2}{\sum_i (f^{i}_{\text{measured}} - \langle f^{i}_{\text{measured}} \rangle)^2}$$  

(2)

Here $f^{i}_{\text{measured}}$ and $f^{i}_{\text{predicted}}$ are measured and predicted $\log \frac{K_{D3}}{K_{D2}}$, where $\langle f^{i}_{\text{measured}} \rangle$ is the averaged ratio of measured binding constants. For model optimization four-fold cross-validation was applied, in which the model was trained on the compounds excluding the validation subset, and then $<q^2>$ was calculated for the compounds of the validation subset. Parameters of the SVM were optimized to achieve maximum performance for the four validation subsets. The final model was trained using all molecules yielding optimized SVM parameters $s$ (Eq.1) and $w$ epsilon width for the regression tube.8

6. Compound ranking based on the similarity to the reference compounds. The selection of the final set of compounds was performed by using a similarity measure, where the distance between molecules was calculated by Equation S3.

$$L(M1, M2) = \sum_i w_i (x^{M1}_i - x^{M2}_i)$$  

(S3)
Here, \( L \) is a distance between molecules \( M1 \) and \( M2 \); \( f_{i}^{M1} \) and \( f_{i}^{M2} \) are 3PP fingerprint bits, and \( w_{i} \) are the weights of the features estimated by the SVM model. The weights \( w \) were extracted from the SVM model by estimating their influence on the predicted activity [Eq. (S4)].

\[
\sum_{i=1}^{n} \left( f_{\text{pred}}(x_{i=1}^{k}) - f_{\text{pred}}(x_{i=0}^{k}) \right) \frac{w_{i}}{n}
\]

The summation is over all active compounds (\( n = 395 \)) of the reference set. The \( f_{\text{pred}}(x_{i=1,0}^{k}) \) is the prediction of the \( \log \frac{K_{D3}}{K_{D2}} \) for compound \( x^{k} \), where fingerprint \( i \) is set to one or zero.

7. Homology model of the D\(_3\) receptor. The transmembrane region of the dopamine D\(_3\) receptor was homology-modeled based on a 2.8 Å resolution rhodopsin crystal structure (PDB-code 1F88).\(^{[12]}\) The sequence alignment for the D\(_3\) receptor and rhodopsin was obtained from the multiple sequence and Hidden Markov model of rhodopsin-like GPCRs from the PFAM database.\(^{[13]}\) For homology modeling, energy minimization and structural analysis of protein we used MOE software package.\(^{[14]}\) The active site of protein was predicted by analyzing the positions of the following residues, that are known to be important for ligand binding to D\(_3\) receptor: Asp110, Ser192, Phe345, Phe346.\(^{[5,15]}\) Compounds were docked into the proposed active site using the MOE built-in docking routine. For scoring partial charges from MMFF94s were used.\(^{[16]}\) We generated 100 different docked conformations for each compound. A final conformation of each ligand was manually selected, taking into consideration mutation data for the D\(_3\) receptor.\(^{[1,2,5]}\) We selected conformations that possess potential hydrogen bonds to Ser192, and between the positively charged amine of the ligands and Asp110.

8. Binding studies. CHO-D\(_{2S}\) cells, expressing the recombinant human D\(_2\) (short) dopamine receptor gene,\(^{[17]}\) were grown in Dulbecco’s Modified Medium/Nutrient Mixture F12 1:1 Mixture supplemented with 2 mm glutamine, 10% foetal bovine serum, and 10 μL mL\(^{-1}\) penicillin/streptomycin in an atmosphere of 5% CO\(_2\) at 37 °C (Gibco, Karlsruhe, Germany). Human D\(_3\) receptors stably expressed in CHO cells as
previously described by Sokoloff et al.\textsuperscript{[18]} were used. The cell line was cultured in Dulbecco’s Modified Eagle Medium supplemented with 2 mM glutamine, and 10\% dialysed fetal bovine serum, and were grown in an atmosphere of 5\% CO\textsubscript{2} at 37 °C (Gibco). Human D\textsubscript{2S} and D\textsubscript{3} receptors expressing cell lines were grown to confluence. The medium was removed, and the cells were washed with 10 ml PBS buffer (140 mM NaCl, 3 mM KCl, 1.5 mM KH\textsubscript{2}PO\textsubscript{4}, 8 mM Na\textsubscript{2}HPO\textsubscript{4}, pH 7.4) at 4 °C. After removing the wash buffer, the cells were scraped from the flasks into 15 mL of ice-cold media, and centrifuged at 3000 rpm for 10 min at 4 °C. After centrifugation the medium was removed and the supernatant resuspended in ice-cold Tris-HCl buffer containing 5 mM MgCl\textsubscript{2}, pH 7.4 and disrupted with a Polytron and centrifuged at 20000 rpm, for 30 min at 4 °C. The pellet was resuspended by sonication in ice-cold Tris-HCl buffer (containing 5 mM MgCl\textsubscript{2}, pH 7.4), membrane aliquots were stored at -70 °C. Determination of membrane protein was carried out by the method of Bradford.\textsuperscript{[19]} Cell membranes containing human D\textsubscript{2S} and D\textsubscript{3} receptors from CHO cells were thawed, rehomogenized with sonication at 4 °C in Tris-HCl, pH 7.4 containing 120 mM NaCl, 5 mM KCl, 2 mM CaCl\textsubscript{2} and 1 mM MgCl\textsubscript{2} (incubation buffer), and incubated with 0.2 nM \textsuperscript{[3H]}spiperone (106 Ci mmol\textsuperscript{-1}, Amersham Biosciences, Freiburg, Germany), and drug diluted in incubation buffer. Nonspecific binding was determined in the presence of 10 \mu M BP897 (prepared by same of the authors). Incubations were run at 25 °C for 120 min, and terminated by rapid filtration through PerkinElmer GF/B glass fibre filters (PerkinElmer Life Sciences, Rodgau, Germany) coated in 0.3% polyethylenimine (Sigma-Aldrich, Taufkirchen, Germany) using an Inotech cell harvester (Inotech AG, Dottikon, Switzerland). Unbound radioligand was removed with four washes of 1 mL of ice-cold 50 mM Tris-HCl buffer, pH 7.4, containing 120 mM NaCl. The filters were soaked in 8 mL Beta plate scint scintillator and counted using a PerkinElmer MicroBeta\textsuperscript{®}Trilux scintillation counter (PerkinElmer Life Sciences). Competition binding data were analysed by the software GraphPad Prism (2000, version 3.02, San Diego, CA, USA), using non-linear least squares fit. For fast screening the compounds have been tested at four concentrations (10 \mu M, 1 \mu M, 0.1 \mu M, and 0.01 \mu M) in triplicate carrying out two binding experiments for human dopamine D\textsubscript{2S} and for human dopamine D\textsubscript{3} receptors. \(K_i\) values were calculated from the IC\textsubscript{50} values according to Cheng-Prusoff equation.\textsuperscript{[20]}
9. References