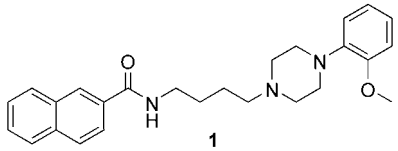
From Virtual to Real Screening for D_3 Dopamine Receptor Ligands

Evgeny Byvatov,[a] Britta C. Sasse,[b] Holger Stark,[b] and Gisbert Schneider*[a]

Imbalance of the dopaminergic system is involved in various neurological and neuropsychiatric disorders, for example, Parkinson's disease, schizophrenia, and drug abuse.[1] Selective attraction of one dopamine receptor subtype could represent an improved therapeutic approach or at least a good way to evaluate the (patho)physiological functions of this subtype in the disorder. Here we focused on the dopamine D_3 receptor, since this subtype plays an important neuroregulatory role in several diseases and possesses a distinct localization in the central nervous system.[2] As D_3 receptors display high sequence identity to D_2 receptors, cross-reactivity is a problem for most compounds used. Although this field of research has been worked on for decades, many lead structures have unsatisfying selectivity. Since numerous described compounds with diverse structural elements showed some D_3 receptor preference, we focused on these elements—first by virtual and then by real screening of the most promising compounds—to find new lead candidates for further optimization.

Virtually screened synthetic compounds from collections of Specs (229685 compounds from release June 2003, Specs, Delft, The Netherlands) and Interbioscreen (IBS; 25601 compounds from release February 2004, Interbioscreen, Moscow, Russia) were investigated as potentially selective ligands at dopamine D_3 receptors. We performed this screening by using analogues of BP897 (**1**), a D_3 receptor-preferring partial agonist



1

in clinical development, and related structures as a starting point. Virtual screening was performed in two stages. In the first stage, we trained a support vector machine (SVM) on the reference set and constructed a filter for D_3 receptor-selective ligands. Based on the prediction of this virtual filter, eleven compounds from the IBS collection and the reference BP897 were tested for binding affinity at D_2 and D_3 receptors. In the second stage, we performed a similarity search with the most promising candidate molecule from the first round against the Specs collection. The parameters for this similarity search were extracted from the SVM model of the BP897 analogues. Four out of five compounds exhibited nanomolar affinity at the D_3 receptor, including a novel scaffold structure. The K_i value for the best molecule was 40 ± 6 nM.

[a] *E. Byvatov, Prof. Dr. G. Schneider*
Beilstein Endowed Chair for Cheminformatics
Johann Wolfgang Goethe-Universität
Fachbereich Chemische und Pharmazeutische Wissenschaften
Institut für Organische Chemie und Chemische Biologie
Marie-Curie-Straße 11, 60439 Frankfurt am Main (Germany)
Fax: (+49) 69-798-29826
E-mail: g.schneider@chemie.uni-frankfurt.de

[b] *B. C. Sasse, Prof. Dr. H. Stark*
Johann Wolfgang Goethe-Universität
Institut für Pharmazeutische Chemie
Marie-Curie-Straße 9, 60439 Frankfurt am Main (Germany)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author: construction of an homology model for the D_3 receptor, docking of compounds into the constructed homology model, and analysis of predicted binding modes. The Supporting Information also includes full details of SVM training and the binding studies.

Ligand-based virtual screening

We used analogues of BP897 and related structures as a reference active set.[3] The compounds from this set possess the following features: i) a lipophilic amine moiety, that is, phenylpiperazine in BP897, ii) a spacer, usually a linear tetramethylene chain, and iii) a hydrophobic residue connected by an amide bond, which has proven to be favorable for high receptor affinity.[3] In order to fulfill structural requirements for high-affinity binding, the basic nitrogen connected to the aryl group through an aliphatic linker was preserved. For all compounds in this series, K_i values of D_2 and D_3 receptor affinities were screened in radioligand binding assays as described.[3]

Compounds were encoded by three-point pharmacophore (3PP) fingerprints available from the MOE software suite.[4] For the first virtual screening round, an SVM was trained on the prediction of potential D_3 receptor ligands. We defined molecules that have measured K_i values below 1 μM for the D_2 or D_3 receptor as "active" compounds (331 out of 395 reference compounds).[5] For cross-validation, this active set was split into four nonoverlapping subsets. During validation we "mimicked" a real screening experiment by addition of compounds known to bind to the D_2 or D_3 receptor to the screening database and estimated the efficiency with which these compounds were retrieved from the screened database. For this, we ranked all screening compounds based on the SVM predictions and optimized SVM parameters, so that the compounds that we mixed with the screening data were at the top of the ranked list.[6–8] The observed enrichment gave an estimation of the expected percentage of active compounds from the IBS dataset that are among the top 1% of the ranked compounds. In the cross-validation study, 50.6 ± 1.3% of the known active compounds were retrieved within 1% of the IBS collection—a result that is significantly above random screening. The training procedure with parameter optimization lasted less than 30 minutes on a Linux cluster with 16 CPUs.

The application of "active learning" further increased the enrichment to 91.8 ± 1.2% of validation actives in the top 1% of the ranked IBS collection (for details of the SVM training procedure and the active-learning concept, see Supporting Information). This was a consequence of the more fine-grained compound sampling from the neighborhood of the known actives in pharmacophore space.

 DOI: 10.1002/cbic.200400400

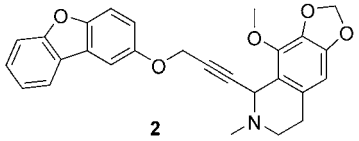
Selection of D_3 receptor-specific ligands

We trained a regression SVM to predict the logarithm of the ratio between K_i values for D_2 and D_3 receptors. The $\langle q^2 \rangle$ of the fourfold cross validation was 0.40 ± 0.15. The relatively low $\langle q^2 \rangle$ is explained by the marked similarity between D_2 and D_3 receptor-binding behavior.[2] The final prediction system was a combination of the two virtual filters described above: binary SVM optimized with active learning, and regression SVM. First, we selected compounds that were similar to the reference set, then we ranked them according to the predicted $\log(K_iD_3/K_iD_2)$ to pick up potential D_3 receptor-selective compounds. The list of the selected molecules obtained was further processed manually so as to exclude compounds with potentially reactive groups or poor solubility. Compounds that are too similar to the reference set were also excluded in order to identify compounds with novel scaffolds. K_i measurement followed a similar protocol as for the BP897 analogues.[3]

Results

Individual compounds exhibited preferential binding at the D_3 receptor, although K_i values for most of the molecules are in the micromolar range, if any could be determined at all (cf. Supporting Information). This observation can be explained by the bias introduced during manual post-selection of molecules. We avoided a pronounced similarity to BP897-like compounds; this obviously resulted in lowering the D_2 and D_3 binding activity.

In order to further increase D_3 receptor affinity, we optimized compound **2** using a similarity searching approach. Molecule **2**



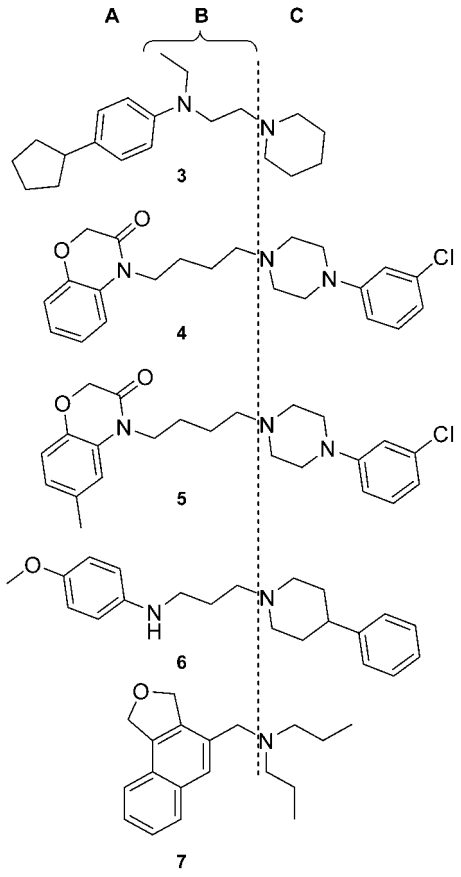
2

was the only ligand found in the first virtual screening round with an experimental $K_i < 2$ μM at the D_3 receptor and K_i value of 2–6 μM at the D_2 receptor. For similarity calculations, we employed a modified distance metric for 3PP fingerprints space, in which fingerprints were weighted based on their importance in our SVM regression model (cf. Supporting Information). This procedure allowed for the selection of compounds that are similar to **2**, focusing on features that were considered to be important for interaction with the receptor. Very similar compounds and compounds with reactive groups were again excluded manually. The testing results for the selected molecules are given in Table 1. The chemical structures of the tested molecules are shown in Scheme 1, aligned at their basic nitrogen, which is assumed to be essential for this type of G protein-receptor binding. As can be seen from Table 1, all active compounds possess a common pattern of the aromatic residue coupled to a potential hydrogen-bond donor (assuming the protonated form) and separated by an aryl moiety from the positively charged amine with an adjacent ring system.

Table 1. Dopamine receptor affinities of compounds from the second virtual screening round (from Specs catalogue)

Molecule	K_i (D_2) ±SD [nM][a]	N[b]	K_i (D_3) ±SD [nM][a]	N[b]
3	1414±516	2	1408±1068	2
4	554±97	4	40±6	4
5	417±60	8	139±17	5
6	201±48	8	96±21	7
7	4395±497	6	914±307	6

[a] K_i values (mean value with standard deviation (SD)) were measured in CHO cells stably expressing hD_{2s} and hD_3 receptors by using [^{3}H]spiperone. [b] Number of experiments.



Scheme 1. Compounds selected for testing based on their similarity to compound **2**. Structures were aligned according to the position of the basic nitrogen (dotted line). Three different parts of the molecules were distinguished: A) an aromatic moiety, B) an aliphatic linker, C) a hydrophobic part connected through a basic nitrogen.

Although the most active compound in this series, **4**, shows nanomolar affinity at the D_3 receptor accompanied by a tenfold D_3 receptor preference in comparison to its D_2 receptor affinity it must be stressed that **4**[9] and **5** are quite similar to the reference set. By using compound libraries, one can hardly expect to retrieve totally unknown lead candidates. Nevertheless, compounds **2**, **3**, **6**, and especially **7** disclose some novel structural features that result in first hits as well as promising new leads for dopamine-receptor subtype ligands in this over-

crowded area of drug development. Together with the other data obtained from virtual and real-compound screening (cf. Supporting Information) one can extract structural characteristics that have not or have only rarely been applied to dopamine D_3 receptor ligands. Compounds **6** and **7** already display slight D_3 receptor preferences; this shows the success of our approach, and give good hopes for **7** for further optimization that is distinct from well-known structure–activity relationships. For the first time, iterative virtual screening cycles with SVM have been successfully applied to entirely ligand-based searching for novel ligands. The concept offers a rapid way to identify lead-structure candidates with minimal experimental effort, even in the absence of receptor-structure information.

Acknowledgements

We thank Dr. P. Sokoloff (INSERM, Paris), and Dr. J. Shine (The Garvan Institute of Medical Research, Sydney) for providing cell lines expressing dopamine D_3 and D_2 receptors, respectively. Technical assistance from C. Bereiter is highly acknowledged. This research was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt am Main.

Keywords: bioinformatics · drug discovery · G protein-coupled receptors · pharmacophores · screening · similarity

[1] a) A. E. Hackling, H. Stark, *ChemBioChem* **2002**, *3*, 946 – 961; b) J. N. Joyce, *Pharmacol. Ther.* **2001**, *90*, 231 – 259.

[2] a) J. R. Bunzow, H. H. Van Tol, D. K. Grandy, P. Albert, J. Salon, M. Cristie, C. A. Machida, K. A. Neve, O. Civelli, *Nature* **1988**, *336*, 783 – 787; b) P. Sokoloff, B. Giros, M. P. Martres, M. L. Bouthenet, J. C. Schwartz, *Nature* **1990**, *347*, 146 – 151; c) H. H. Van Tol, J. R. Bunzow, H. C. Guan, R. K. Sunahara, P. Seeman, H. B. Niznik, O. Civelli, *Nature* **1991**, *350*, 610 – 614.

[3] a) U. R. Mach, A. E. Hackling, S. Perachon, S. Ferry, C. G. Wermuth, J. C. Schwarz, P. Sokoloff, H. Stark, *ChemBioChem* **2004**, *5*, 508 – 518; b) A. E. Hackling, R. Ghosh, S. Perachon, A. Mann, H. D. Höltje, C. G. Wermuth, J. C. Schwarz, W. Sippl, P. Sokoloff, H. Stark, *J. Med. Chem.* **2003**, *46*, 3883 – 3899.

[4] MOE (Molecular Operating Environment) version 2004.05 by Chemical Computing Group, Inc., Montréal (Canada) http://www.chemcomp.com

[5] M. Pilla, S. Perachon, F. Sautel, F. Garridol, A. Mann, C. G. Wermuth, J. C. Schwartz, B. J. Everitt, P. Sokoloff, *Nature* **1999**, *400*, 371 – 375.

[6] a) G. Schneider, P. Schneider in *Chemogenomics in Drug Discovery* (Eds.: H. Kubinyi, G. Müller), Wiley-VCH, Weinheim, **2004**, pp. 341 – 376; b) E. Byvatov, G. Schneider, *Appl. Bioinf.* **2003**, *2*, 67 – 77.

[7] T. Joachims in *Advances in Kernel Methods—Support Vector Learning* (Eds.: B. Schölkopf, C. Burges, A. Smola), MIT-Press, Cambridge, MA, **1999**, pp. 41 – 56.

[8] a) E. Byvatov, G. Schneider, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 993 – 999; b) E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882 – 1889.

[9] a) M. J. Mokrosz, P. Kowalski, T. Kowalska, *Arch. Pharm.* **1999**, *332*, 373 – 379; b) M. J. Mokroz, S. Charakchieva-Minol, P. Kowalski, *Arch. Pharm.* **2001**, *334*, 25 – 29.

Received: November 9, 2004
Published online on April 25, 2005

