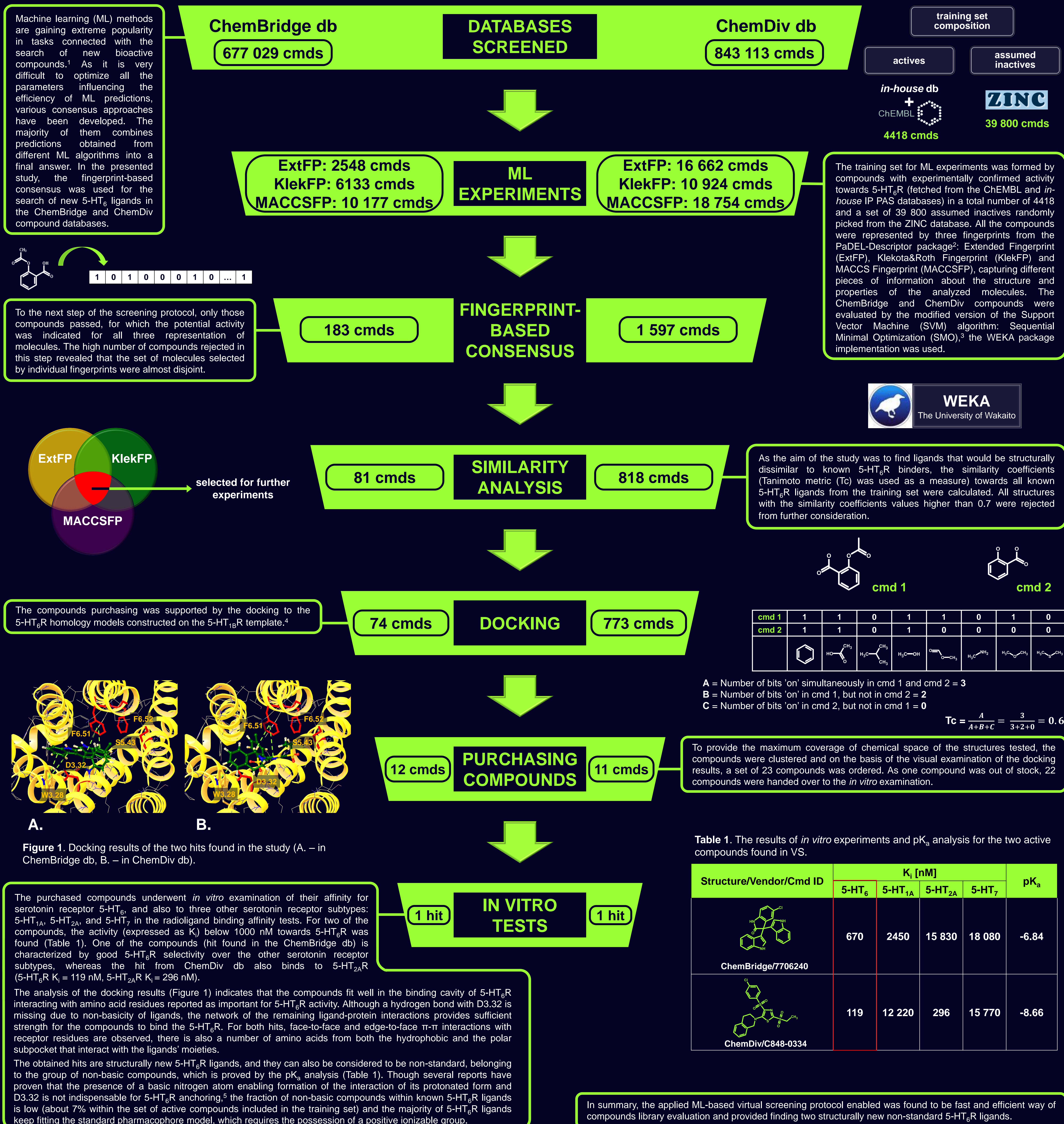# New non-basic ligands of serotonin receptor 5-HT$_6$ as a result of virtual screening based on machine learning methods

<u>Sabina Podlewska</u>,[a,b] Rafał Kurczab[a], Grzegorz Satała[a], Andrzej J. Bojarski[a]

[a]Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków
[b]Faculty of Chemistry, Jagiellonian University, 3 Ingardena Street, 30-060 Kraków
e-mail: smusz@if-pan.krakow.pl

Machine learning (ML) methods are gaining extreme popularity in tasks connected with the search of new bioactive compounds.[1] As it is very difficult to optimize all the parameters influencing the efficiency of ML predictions, various consensus approaches have been developed. The majority of them combines predictions obtained from different ML algorithms into a final answer. In the presented study, the fingerprint-based consensus was used for the search of new 5-HT$_6$ ligands in the ChemBridge and ChemDiv compound databases.

DATABASES SCREENED

ChemBridge db — 677 029 cmds

ChemDiv db — 843 113 cmds

training set composition

actives

in-house db + ChEMBL — 4418 cmds

assumed inactives

ZINC — 39 800 cmds

ML EXPERIMENTS

ExtFP: 2548 cmds
KlekFP: 6133 cmds
MACCSFP: 10 177 cmds

ExtFP: 16 662 cmds
KlekFP: 10 924 cmds
MACCSFP: 18 754 cmds

The training set for ML experiments was formed by compounds with experimentally confirmed activity towards 5-HT$_6$R (fetched from the ChEMBL and *in-house* IP PAS databases) in a total number of 4418 and a set of 39 800 assumed inactives randomly picked from the ZINC database. All the compounds were represented by three fingerprints from the PaDEL-Descriptor package[2]: Extended Fingerprint (ExtFP), Klekota&Roth Fingerprint (KlekFP) and MACCS Fingerprint (MACCSFP), capturing different pieces of information about the structure and properties of the analyzed molecules. The ChemBridge and ChemDiv compounds were evaluated by the modified version of the Support Vector Machine (SVM) algorithm: Sequential Minimal Optimization (SMO),[3] the WEKA package implementation was used.

`1 0 1 0 0 0 0 1 0 ... 1`

FINGERPRINT-BASED CONSENSUS

183 cmds

1 597 cmds

To the next step of the screening protocol, only those compounds passed, for which the potential activity was indicated for all three representation of molecules. The high number of compounds rejected in this step revealed that the set of molecules selected by individual fingerprints were almost disjoint.

WEKA The University of Wakaito

SIMILARITY ANALYSIS

ExtFP, KlekFP, MACCSFP → selected for further experiments

81 cmds

818 cmds

As the aim of the study was to find ligands that would be structurally dissimilar to known 5-HT$_6$R binders, the similarity coefficients (Tanimoto metric (Tc) was used as a measure) towards all known 5-HT$_6$R ligands from the training set were calculated. All structures with the similarity coefficients values higher than 0.7 were rejected from further consideration.

DOCKING

74 cmds

773 cmds

The compounds purchasing was supported by the docking to the 5-HT$_6$R homology models constructed on the 5-HT$_{1B}$R template.[4]

cmd 1

cmd 2

cmd 1	1	1	0	1	1	0	1	0
cmd 2	1	1	0	1	0	0	0	0

A = Number of bits 'on' simultaneously in cmd 1 and cmd 2 = **3**
B = Number of bits 'on' in cmd 1, but not in cmd 2 = **2**
C = Number of bits 'on' in cmd 2, but not in cmd 1 = **0**

$$Tc = \frac{A}{A+B+C} = \frac{3}{3+2+0} = 0.6$$

PURCHASING COMPOUNDS

12 cmds

11 cmds

To provide the maximum coverage of chemical space of the structures tested, the compounds were clustered and on the basis of the visual examination of the docking results, a set of 23 compounds was ordered. As one compound was out of stock, 22 compounds were handed over to the *in vitro* examination.

Figure 1. Docking results of the two hits found in the study (A. – in ChemBridge db, B. – in ChemDiv db).

IN VITRO TESTS

1 hit

1 hit

The purchased compounds underwent *in vitro* examination of their affinity for serotonin receptor 5-HT$_6$, and also to three other serotonin receptor subtypes: 5-HT$_{1A}$, 5-HT$_{2A}$, and 5-HT$_7$ in the radioligand binding affinity tests. For two of the compounds, the activity (expressed as K$_i$) below 1000 nM towards 5-HT$_6$R was found (Table 1). One of the compounds (hit found in the ChemBridge db) is characterized by good 5-HT$_6$R selectivity over the other serotonin receptor subtypes, whereas the hit from ChemDiv db also binds to 5-HT$_{2A}$R (5-HT$_6$R K$_i$ = 119 nM, 5-HT$_{2A}$R K$_i$ = 296 nM).

The analysis of the docking results (Figure 1) indicates that the compounds fit well in the binding cavity of 5-HT$_6$R interacting with amino acid residues reported as important for 5-HT$_6$R activity. Although a hydrogen bond with D3.32 is missing due to non-basicity of ligands, the network of the remaining ligand-protein interactions provides sufficient strength for the compounds to bind the 5-HT$_6$R. For both hits, face-to-face and edge-to-face π-π interactions with receptor residues are observed, there is also a number of amino acids from both the hydrophobic and the polar subpocket that interact with the ligands' moieties.

The obtained hits are structurally new 5-HT$_6$R ligands, and they can also be considered to be non-standard, belonging to the group of non-basic compounds, which is proved by the pK$_a$ analysis (Table 1). Though several reports have proven that the presence of a basic nitrogen atom enabling formation of the interaction of its protonated form and D3.32 is not indispensable for 5-HT$_6$R anchoring,[5] the fraction of non-basic compounds within known 5-HT$_6$R ligands is low (about 7% within the set of active compounds included in the training set) and the majority of 5-HT$_6$R ligands keep fitting the standard pharmacophore model, which requires the possession of a positive ionizable group.

Table 1. The results of *in vitro* experiments and pK$_a$ analysis for the two active compounds found in VS.

Structure/Vendor/Cmd ID	K$_i$ [nM]				pK$_a$
	5-HT$_6$	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_7$	
ChemBridge/7706240	670	2450	15 830	18 080	-6.84
ChemDiv/C848-0334	119	12 220	296	15 770	-8.66

In summary, the applied ML-based virtual screening protocol enabled was found to be fast and efficient way of compounds library evaluation and provided finding two structurally new non-standard 5-HT$_6$R ligands.

References

1. Melville, J. L.; Burke, E. K.; Hirst, J. D. *Comb. Chem. High Throughput Screen.* **2009**, *12*, 332.
2. Yap, C. W. E. I. *J. Comput. Chem.* **2010**, *32*, 1466.
3. Platt, J.C. Tech. Report (Miscrosoft research) **1998**, MSR-TR-98-14,1.
4. Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katrich, V.; Han, G. W.; Liu, W.; Huang, X. P.; Vardy, E.; McCorvy, J. D.; Gao, X.; Zhou, X. E.; Melcher, K.; Zhang, C.; Bai, F.; Yang, H.; Yang, L.; Jiang, H.; Roth, B. L.; Cherezov, V.; Stevens, R. C.; Xu, H. E. *Science* **2013**, *340*, 610.
5. Ivachtchenko, A. V.; Dmitriev, D. E.; Golovina, E. S.; Kadieva, M. G.; Koryakova, A. G.; Kysil, V. M.; Mitkin, O. D.; Okun, I. M.; Tkachenko, S. E.; Vorobiev, A. A. *J. Med. Chem.* **2010**, *53*, 5186.