JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Virtual Affinity Fingerprints for Target Fishing: A New Application of Drug Profile Matching

Ágnes Peragovics,[†,§] Zoltán Simon,[§,∥] László Tombor,[⊥] Balázs Jelinek,[‡,§] Péter Hári,[§,∥] Pál Czobor,[⊥] and András Málnási-Csizmadia*[,†,‡,§]

[†]Department of Biochemistry, Institute of Biology, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

[‡]Molecular Biophysics Research Group, Hungarian Academy of Sciences - ELTE, Budapest, Hungary

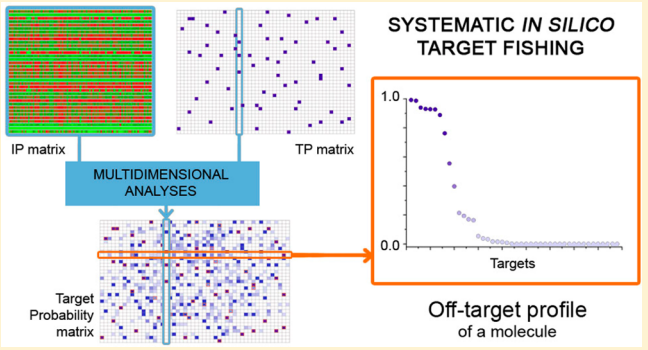
[§]Drugmotif Ltd., Szent Erzsébet krt. 14, H-2112 Veresegyház, Hungary

[∥]Printnet Ltd., Petneházy utca 52, H-1139 Budapest, Hungary

[⊥]Department of Psychiatry and Psychotherapy, Semmelweis University, Balassa utca 6, H-1083 Budapest, Hungary

Supporting Information

ABSTRACT: We recently introduced drug profile matching (DPM), a novel virtual affinity fingerprinting bioactivity prediction method. DPM is based on the docking profiles of ca. 1200 FDA-approved small-molecule drugs against a set of nontarget proteins and creates bioactivity predictions based on this pattern. The effectiveness of this approach was previously demonstrated for therapeutic effect prediction of drug molecules. In the current work, we investigated the applicability of DPM for target fishing, i.e. for the prediction of biological targets for compounds. Predictions were made for 77 targets, and their accuracy was measured by receiver operating characteristic (ROC) analysis. Robustness was tested by a rigorous 10-fold cross-validation procedure. This procedure identified targets (N = 45) with high reliability based on DPM performance. These 45 categories were used in a subsequent study which aimed at predicting the off-target profiles of currently approved FDA drugs. In this data set, 79% of the known drug-target interactions were correctly predicted by DPM, and additionally 1074 new drug−target interactions were suggested. We focused our further investigation on the suggested interactions of antipsychotic molecules and confirmed several interactions by a review of the literature.



INTRODUCTION

Finding compounds for a given target is a common computational task in a conventional medicinal chemistry program. However, by means of increasingly available bioactivity data, this approach can be reversed to finding targets for compounds. In silico target fishing[1] is an emerging field that aims at predicting biological targets of molecules based on their chemical structure. The rise of this area is in connection with that of polypharmacology,[2,3] which posits that drugs act on multiple targets in contrast with the traditional one drug−one target paradigm. As a consequence, it is likely to discover new targets even for well-known drugs.

Many in silico target prediction tools have been developed, and they were summarized by a recent review.[4] As it is common for drug development methods, target prediction tools can also be divided into two main groups: ligand-based and structure-based approaches.

Similarity search is often used among the ligand-based methods. The most common question that arises in case of similarity based virtual screening is the description of molecular structure. No universal solution seems to exist for this problem,[5] as the best representation used to characterize the molecules depends on the studied activity classes. Therefore, it is important to combine several methods for a given task, e.g. by applying data fusion techniques.[6] An approach that generates off-target profiles of drugs based on their 3D similarity has just been reported, and some of its predictions were proved by a literature survey.[7]

Several ligand-based methods apply data mining methods in order to identify unknown drug−target interactions. One of the first initiatives in this field was PASS developed by Poroikov et al.[8] It can predict the biological activity profile of a compound based on the analysis of structure−activity relationships for more than 250 000 biologically active substances. Nigsch et al. implemented the Winnow and Naive Bayesian algorithms for ligand−target prediction and compared their performance on a data set comprising 20 activity classes with 13 000 compounds.[9] They generally produced similar performance, however, significant differences were observed for the individual activity classes. The similarity ensemble approach (SEA) uses a minimal

Received: September 19, 2012

   dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

spanning tree considering ligand chemical similarity in order to clusterize 246 enzymes and receptors.[10,11] On the basis of the model, target prediction was performed for more than 3000 FDA approved drugs, and 23 suggested interactions were confirmed experimentally.

Pharmacophore based methods also proved to be successful to predict protein targets. PharmMapper employs pharmacophore models derived from structures complexed with small molecules to identify target candidates of query molecules.[12] Tamoxifen was selected as a validation example, and it was concluded that the method was successful in predicting its targets.

Molecular docking is far the most often used tool among the structure-based methods. While conventionally it is applied to identify potential ligands for a given protein, for target prediction the so-called inverse docking procedure needs to be applied (docking one ligand against multiple targets). INVDOCK[13] and TarFishDock[14] are examples of recently presented methods for predicting protein targets for small molecules based on docking against a set of proteins supposedly interacting with the ligand.

This concept has some relation to in silico affinity fingerprints,[15−17] which are a series of docking scores against a reference panel of proteins that do not include the target protein (one ligand, multiple proteins). However, this approach is not designed to find possible targets among the reference proteins. Instead, these reference proteins are used as a discriminator surface which can differentiate a wide range of compounds. In contrast to the computationally more demanding inverse docking, individual interactions are not considered here, the resulting pattern is characteristic for the studied molecules.

Affinity fingerprints were originally based on in vitro measurements;[18−20] however, the measured values were later replaced by docking scores (virtual binding free energies). In silico affinity fingerprints were successfully applied in virtual screening protocols[16,21] and focused library design.[22]

We recently introduced drug profile matching (DPM), a novel virtual affinity fingerprinting prediction method. DPM is based on the docking profiles of ca.1200 FDA-approved small-molecule drugs against a panel of nontarget proteins. Individual interactions are not investigated in the method; instead, a docking profile serves as a pattern that is characteristic for a given molecule. Our working hypothesis was that similar patterns indicate similar bioactivity of the respective molecules and this feature can be exploited for bioactivity prediction. Relevant information of the docking profiles was extracted by multidimensional statistical techniques that produced probabilities showing the likelihood of having the investigated property for each molecule. The effectiveness of this approach was already demonstrated for pharmacological effect prediction.[23] Moreover, we also showed that DPM adds additional predictive power to drug effect prediction as compared to traditional molecular similarity based approaches.[24] Candidate molecules were tested in vitro for three selected categories, and high hit rates were obtained which further proved the validity of DPM predictions (unpublished results). The system was formerly trained on pharmacological effects (medical indications) based on the categories listed in the DrugBank database. Groups based on common targets were also included among the studied categories and resulted in high classification accuracy. Therefore, as a continuation of our work, we decided to pursue a study on drug-target interaction data. Our current approach is similar to the original application of affinity fingerprints presented by Kauvar et al. In their pioneer work, the binding potencies of several compounds were measured against a reference panel of proteins and the resulting affinity fingerprints of the compounds were applied to predict their binding properties to other proteins not included in the reference panel.[18,19] In our approach, we also aim to predict interactions between the studied molecules and possible drug targets that are not represented in the reference protein set used to generate the interaction patterns of the compounds by molecular docking.

In the present study, DPM predictions were made based on 77 targets extracted from the DrugBank database that contain at least 10 registered molecules in order to provide sufficient amount of information about the active molecules. It should be noted that there is no overlap between the reference protein set used for creating the interaction patterns and the investigated 77 targets. The reference protein set consists of only nontarget proteins. Similar to our previous work, the accuracy of DPM predictions was assessed by receiver operating characteristic (ROC) analysis, while robustness was measured via 10-fold cross-validation. On the basis of the calculated prediction properties, 45 targets possessing sufficient prediction power were selected for further analysis. Predicted off-target profiles with this reduced target set were examined in order to reveal new drug−target interactions. For many drug molecules, significantly more targets were predicted with high probability than it was originally registered in the database. Predicted off-target profiles were examined for selected molecules, and the validity of several suggested interactions was demonstrated by a review of the literature.

METHODS

Drug Profile Matching Method. DPM was described in detail in our recent publications.[23,24] The key steps of the method and the analyses used to describe its accuracy and robustness are presented briefly in the following.

Creation of the Interaction Pattern Matrix. Here 1177 FDA approved small-molecule drugs were extracted from DrugBank database and were docked to 135 nontarget proteins from RCSB Protein Data Bank (PDB) (Table S1, Supporting Information). Docking was performed using DOVIS 2.0 software (DOcking-based VIrtual Screening),[25] AutoDock4 docking engine,[26] Lamarckian genetic algorithm, and X-SCORE scoring function.[27] The geometrical center of the original ligand was used as a center of the docking box, box size and grid spacing were set to 22.5 and 0.375 Å, respectively. Twenty-five docking runs were performed for each job, and the best docking scores for each drug−protein complex were used to form the interaction pattern (IP) matrix. In this matrix, drugs are organized into rows while proteins are in the columns therefore each row represents the IP of a given drug against the reference protein set.

For a more detailed description of IP generation see the Supporting Information.

Creation of the Target Profile Matrix. Target information on 1177 FDA approved small-molecule drugs was extracted from DrugBank database. For 20 molecules no target information could be obtained; these molecules were excluded from further analysis. The resulting 1157 drugs were assigned to 1163 targets that were reviewed manually. The number of categories was reduced to 995 by merging cohesive target groups (for example, DNA topoisomerase 4 subunit A and subunit B were combined to produce the final DNA

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

topoisomerase 4 category). Figure 1 shows the distribution of the registered drugs for these 995 targets. Remarkably, to 628

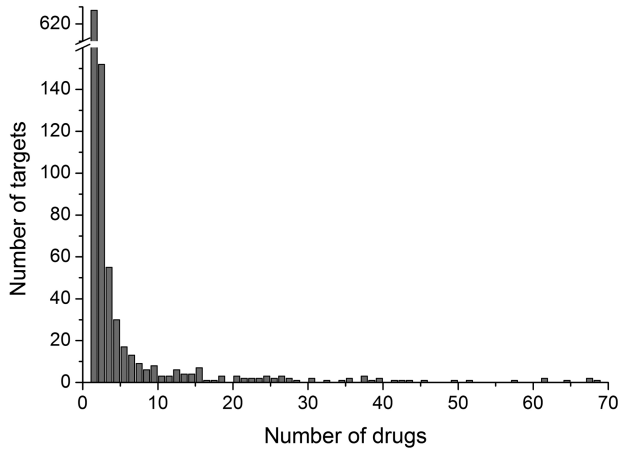


Figure 1. Distribution of the registered drugs for the original 995 targets. Number of targets with a given number of approved drugs is displayed. Note that for more than 60% of the targets only one molecule is assigned.

targets only one drug is assigned in the database, raising difficulties to exploit this information for prediction with our method. There are only 6 targets having more than 60 assigned drugs: histamine H1 (68 drugs), muscarinic acetylcholine receptor M1 (67 drugs), alpha 1A adrenergic receptor (67 drugs), DNA (64 drugs) dopamine D2 receptor (61 drugs), and GABA receptor (61 drugs). The mean of the registered drugs to the 995 targets is 3.6, supporting the general view of polypharmacology that drugs act on multiple targets. According to our previous experience gained in therapeutical effect predictions, DPM requires 10 active molecules for sufficient classification. Thus, from the original 995 target groups, only 77 could be kept for the analyses having at least ten registered molecules. This investigated target set is independent of the reference protein set used to generate the interaction patterns. A binary matrix called Target Profile (TP) matrix was created based on these 77 groups that displays whether a drug interacts with a given target according to DrugBank. ("1" marks the presence of the interaction while "0" indicates that a given drug-target interaction was not documented). Targets are organized into columns whereas drugs are in the rows of this matrix; therefore, one row represents the DrugBank documented target profile of a given drug. Since many targets were excluded due to the fact that they have less than 10 active molecules, there are several drugs whose target profile is empty. This issue does not raise problems for DPM since the statistical analyses are performed separately for each target (i.e., column by column), as it is described in the following section.

Creation of the Target Probability Matrix. Canonical correlation analysis (CCA) was performed between the IP matrix and each target to generate a factor pair having as high correlation as possible via linear combination of the original variables. This factor pair was used as an input for linear discriminant analysis (LDA) that yielded classification functions which were applied to calculate the probability for each drug-target pair. These probability values were used to create the target probability matrix. Any row of this matrix represents the predicted off-target profile of a given drug. In contrast to the binary target profile matrix, the values in this matrix are continuous, and therefore assignment of a given target to a particular drug also depends on the used probability threshold.

An example on a small data set that illustrates the different steps of the DPM method resulting in the final probability values is presented in the Supporting Information.

Receiver Operating Characteristic Analysis. Receiver operating characteristic (ROC) analysis was used for assessing the accuracy of the classification functions. To create a ROC curve for each target group, the true positive rate (TPR) was plotted as a function of the false positive rate (FPR) using a sliding cutoff parameter from 0 to 1 for the probabilities. Molecules are reclassified at each cutoff value, labeling compounds as "positive" if they have a greater probability for a given target than the applied cutoff point and "negative" in the opposite case. TPR (also called sensitivity) is the portion of positives classified correctly, while FPR (1-sensitivity) is the rate of negatives which were wrongly classified as positive. To produce a quantitative summary measure of the ROC curve, the area under the curve (AUC) was calculated. Perfect classification results in an AUC of 1, because in that case there exists a cutoff value above that all positive molecules but no negative molecules are classified as positive and thus the curve runs through the (0,1) point. Therefore, the closer the calculated AUC value is to 1, the better the classification. A random classification would result in a diagonal ROC curve (AUC of 0.5), representing a method with no ability to distinguish active and inactive molecules. New measures have been introduced recently such as BEDROC that also take into account the shape of the ROC curves,[28] resulting in higher values for those curves that rise steeply along the x axis, meaning that known actives are indentified at the top of the list. Calculation of the BEDROC metric was performed in our earlier work,[23] but it did not result in different conclusions than the use of the AUCs. Therefore, we decided to use AUC values in the current work.

10-fold Cross-validation. In order to evaluate the robustness of the results and control for possible overfitting, 10-fold cross-validation was performed. The data was divided into 10 complementary subsets. Each subset was used as a test set for validation while the residual subsets were combined to produce the training set. In each round of the validation, CCA and LDA was performed on the training set and probabilities were predicted for the test set that show the likelihood of interacting with a given target for each test molecule. Accordingly, the classification function was created without considering the test set, ensuring that the test set was completely independent of the training set. Variable selection was not performed in the cross-validation loop as the same set of the predefined 135 nontarget proteins was used in each round of the validation. This process was repeated for each of the 10 subsets, and the probability values for each of the originally registered drugs to a given target were averaged to produce a single measure (mean probability value, MPV) that indicates the robustness of the studied target. This process was repeated 100 times for each target group to eliminate the impact of the distribution of molecules on the results. The outcomes of the 100 runs were combined to create the investigated mean MPVs that describe the robustness of a given target, i.e. to what extent the classification can be generalized on external data. The closer the MPV to 1, the better is the performance of the method on the external data for the studied target group.

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

A validation based on ChEMBL data for a subset of the investigated interactions is presented in the Supporting Information.

Target Selectivity Analysis. In order to assess the target selectivity of the studied drugs, the number of predicted targets above a certain probability limit (>0.8) was counted. To ensure nonbiased analysis, from the 77 original targets only the 45 highly reliable targets with the best robustness values (mean MPV > 0.5) were used. The predicted interactions of antipsychotics was investigated in more detail by a literature survey.

Tanimoto Diversity Calculation. Two-dimensional hashed chemical fingerprints, that encode topological properties of the chemical graph up to six bonds, were generated using ChemAxon's JChem based software for each drug molecule. The process resulted in a 4096-bit-long binary fingerprint for each drug. Then, ChemAxon Similarity plugin was used to calculate the Tanimoto similarity for each possible drug pair on the basis of these fingerprints:

$$\mathrm{SIM(A,\ B)} = \frac{c}{a + b + c}$$

where a is the number of bits on in molecule A, b is the number of bits on in molecule B, and c is the number of bits in common in both structures.

Comparing identical molecules results in a similarity value of 1, while the calculated similarity is 0 when two molecules have no bits in common. The average Tanimoto similarity (referred to as Tanimoto diversity) was calculated for each of the studied targets to quantify the structural distribution of the registered molecules. Considerable structural similarity exists among the ligands of a given target if the Tanimoto diversity exceeds 0.6. If this value is less than 0.4, a target group is considered structurally heterogeneous.

The Statistical Analysis System for Windows (version 9.2; SAS Institute, Cary, NC) was used for the implementation of all analyses.

RESULTS AND DISCUSSION

Figure 2 displays a graphical summary of the drug profile matching method applied for target fishing. Virtual binding affinity values obtained by docking 1157 FDA-approved drugs to 135 nontarget proteins were entered into a matrix, where each row displays the interaction pattern (IP) of a given drug against this protein set. On the basis of the target information extracted from DrugBank for the studied molecules, a binary matrix called target profile (TP) matrix was created which shows whether a given drug-target interaction is documented in the DrugBank database. A two-step multidimensional method (CCA and LDA) was applied on these matrices to yield probabilities for each drug that indicates the likelihood of interacting with a given target. These probabilities were entered into the target probability matrix where each row shows the predicted off-target profile of a given drug. It should be noted that these values do not yield any information about the strength of the suggested drug-target interaction that requires in vitro measurements in order to be determined.

Receiver Operating Characteristic Analysis. Overall classification accuracy of DPM was measured by receiver operating characteristic (ROC) analysis which is based on the list of drugs sorted by descending probability for a selected target (a column in the target probability matrix). Table 1 lists the obtained AUC values while Figure 3 shows their distribution for the 77 studied target groups. All AUC values were

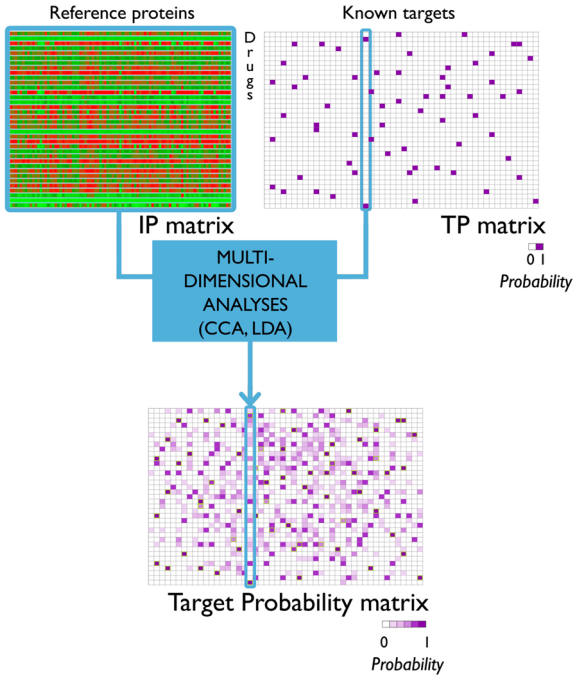


Figure 2. Graphical summary of the drug profile matching method applied for target fishing. The interaction pattern (IP) matrix contains the calculated binding free energies for the studied 1157 drugs on the reference panel of 135 nontarget proteins. The target profile (TP) matrix shows the known drug-target interactions in a binary coded form (purple cells mark the presence of the interaction while white cells indicate that a given drug−target interaction was not documented in DrugBank). These matrices were subjected to a two-step multidimensional analysis (canonical correlation analysis, CCA, and linear discriminant analysis, LDA) that resulted in the target probability matrix that consists of the predicted probabilities for each drug−target pair.

above 0.92, meaning that excellent classification was obtained by DPM for target prediction. Perfect classification (i.e., AUC of 1) occurred for three categories, registered ligands of both target groups share high degree of structural similarity (fluoroquinolone antibiotics targeting DNA topoisomerase 4, sulphanilamides targeting dihydropteroate synthase, and steroid molecules targeting progesterone receptor; Tanimoto diversities of 0.766, 0.505, and 0.545, respectively; see Table 1 for the complete list of the Tanimoto diversities). Structural similarity of registered ligands can be observed for several other target groups among the best categories (glucocorticoid receptor, peptidoglycan synthetase ftsl, penicillin binding protein 2A). However, target groups comprising of structurally diverse compounds also obtained high AUC values (0.998 for cholinesterase and 0.997 for monoamine oxidase A; Tanimoto diversities of their registered ligands are 0.241 and 0.380, respectively). This is in an agreement with our previous finding that DPM can effectively handle classes comprising of structurally diverse molecules.[24] These are the cases where DPM has additional prediction power compared to traditional similarity based approaches. The worst but still excellent AUC of 0.922 was obtained for neuronal acetylcholine receptor, target of mainly barbiturate molecules (Tanimoto diversity of 0.358).

Cross-validation. To check the validity of the obtained classifications, an independent 10-fold cross-validation was performed. The MPVs of the 100 runs were averaged to

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Prediction and Validation Properties of the Studied 77 Target Groups[a]

target	*n*	AUC	10-fold cross-validation mean	10-fold cross-validation std	Tanimoto diversity
acetylcholinesterase	18	0.991	0.322	0.047	0.271
alpha-1A adrenergic receptor	67	0.951	0.622	0.016	0.382
alpha-1B adrenergic receptor	39	0.946	0.467	0.021	0.399
alpha-1D adrenergic receptor	23	0.965	0.321	0.038	0.398
alpha-2A adrenergic receptor	51	0.945	0.492	0.020	0.378
alpha-2B adrenergic receptor	30	0.976	0.436	0.024	0.394
alpha-2C adrenergic receptor	26	0.982	0.386	0.027	0.395
androgen receptor	14	0.993	0.656	0.048	0.429
angiotensin-converting enzyme	13	0.997	0.503	0.037	0.618
arachidonate 5-lipoxygenase	13	0.988	0.112	0.038	0.336
ATP-binding cassette transporter subfamily C member 8	13	0.997	0.494	0.068	0.495
Beta-1 adrenergic receptor	37	0.982	0.722	0.020	0.508
Beta-2 adrenergic receptor	41	0.987	0.734	0.017	0.515
calmodulin	15	0.982	0.227	0.036	0.409
cAMP-specific 3′,5′-cyclic phosphodiesterase 4	12	0.989	0.353	0.039	0.456
carbonic anhydrase 1	20	0.993	0.459	0.019	0.417
carbonic anhydrase 2	20	0.997	0.512	0.021	0.429
carbonic anhydrase 4	16	0.996	0.578	0.027	0.448
cholinesterase	12	0.998	0.227	0.044	0.241
cytochrome P450 51	12	0.999	0.317	0.053	0.569
D(1) dopamine receptor	43	0.958	0.668	0.013	0.423
D(2) dopamine receptor	61	0.965	0.656	0.011	0.410
D(3) dopamine receptor	25	0.992	0.422	0.030	0.420
D(4) dopamine receptor	21	0.988	0.357	0.034	0.409
delta-type opioid receptor	22	0.974	0.587	0.017	0.564
dihydropteroate synthase	10	1.000	0.800	0.042	0.505
DNA	64	0.965	0.575	0.016	0.330
DNA gyrase	15	0.996	0.737	0.024	0.705
DNA topoisomerase 2	21	0.994	0.691	0.034	0.512
DNA topoisomerase 4	13	1.000	0.820	0.028	0.766
estrogen receptor	27	0.972	0.694	0.031	0.418
gamma-aminobutyric acid receptor	61	0.985	0.674	0.013	0.337
glucocorticoid receptor	37	0.999	0.901	0.010	0.636
glutamate receptor NOS	34	0.957	0.563	0.013	0.333
histamine H1 receptor	68	0.957	0.695	0.010	0.383
kappa-type opioid receptor	23	0.977	0.488	0.017	0.502
monoamine oxidase A	10	0.997	0.488	0.061	0.380
Mu-type opioid receptor	28	0.982	0.553	0.013	0.552
muscarinic acetylcholine receptor M1	67	0.953	0.656	0.009	0.396
muscarinic acetylcholine receptor M2	49	0.944	0.536	0.015	0.373
muscarinic acetylcholine receptor M3	45	0.954	0.566	0.016	0.403
muscarinic acetylcholine receptor M4	30	0.961	0.557	0.020	0.396
muscarinic acetylcholine receptor M5	26	0.965	0.535	0.018	0.391
neuronal acetylcholine receptor	35	0.922	0.519	0.014	0.358
penicillin-binding protein 1A	24	0.991	0.589	0.034	0.698
penicillin-binding protein 1b	22	0.990	0.597	0.041	0.699
penicillin-binding protein 2	12	0.997	0.406	0.065	0.751
penicillin-binding protein 2B	15	0.995	0.452	0.053	0.720
penicillin-binding protein 2a	15	0.998	0.610	0.045	0.708
penicillin-binding protein 3	24	0.994	0.670	0.037	0.681
penicillin-binding proteins 1A/1B	18	0.997	0.724	0.055	0.690
peptidoglycan synthetase ftsI	12	0.999	0.528	0.060	0.793
peroxisome proliferator-activated receptor	18	0.987	0.413	0.034	0.399
potassium voltage-gated channel subfamily H member 2	20	0.976	0.381	0.038	0.431
progesterone receptor	14	1.000	0.717	0.039	0.545
prostaglandin G/H synthase 1	38	0.970	0.612	0.020	0.345
prostaglandin G/H synthase 2	42	0.976	0.649	0.020	0.356
reverse transcriptase	10	0.996	0.392	0.086	0.530
sodium channel protein type 10	14	0.991	0.463	0.041	0.505
sodium channel protein type 5	27	0.959	0.345	0.025	0.396
sodium-dependent dopamine transporter	26	0.967	0.432	0.028	0.422
sodium-dependent noradrenaline transporter	39	0.971	0.650	0.016	0.402
sodium-dependent serotonin transporter	35	0.962	0.594	0.026	0.410
translocator protein	12	0.992	0.705	0.031	0.558
tubulin	11	0.998	0.434	0.059	0.554
voltage-dependent L-type calcium channel	11	0.992	0.664	0.045	0.581
voltage-dependent T-type calcium channel	14	0.987	0.344	0.064	0.356
voltage-dependent calcium channel	15	0.940	0.493	0.027	0.493
16S rRNA	15	0.999	0.632	0.042	0.589
5-hydroxytryptamine 1A receptor	37	0.969	0.560	0.026	0.406
5-hydroxytryptamine 1B receptor	25	0.992	0.483	0.031	0.467
5-hydroxytryptamine 1D receptor	24	0.994	0.545	0.035	0.495
5-hydroxytryptamine 2A receptor	57	0.949	0.640	0.013	0.411
5-hydroxytryptamine 2B receptor	15	0.984	0.248	0.038	0.441
5-hydroxytryptamine 2C receptor	32	0.981	0.512	0.023	0.413
5-hydroxytryptamine 3 receptor	17	0.983	0.293	0.044	0.379
5-hydroxytryptamine 7 receptor	11	0.997	0.216	0.055	0.478

[a]For each studied target, the number of active molecules (*n*), the AUC values, and the results of the 10-fold cross-validation (mean and standard deviation of MPV) are listed. To quantify the chemical diversity of the molecules registered to a given target, the average Tanimoto similarity (Tanimoto diversity, see Methods) was calculated for each target group.

produce the further investigated mean MPVs for each target. Table 1 and Figure 4 display the mean MPVs with standard deviation for the 77 targets. This value is used to counter overfitting, which is a known phenomena of multidimensional statistical techniques, and shows whether the classification functions could capture relevant features for the studied targets.

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

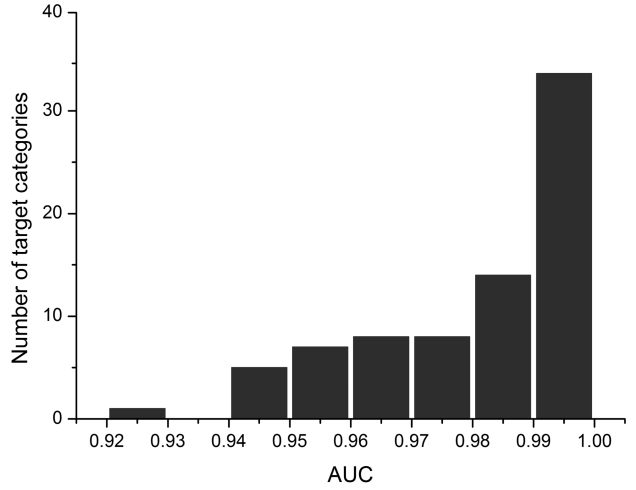


Figure 3. Distribution of AUC values for the studied 77 target groups. ROC analyses were performed to describe classification accuracy. All of the calculated AUC values were greater than 0.92, indicating that a near perfect classification was obtained for the studied targets.

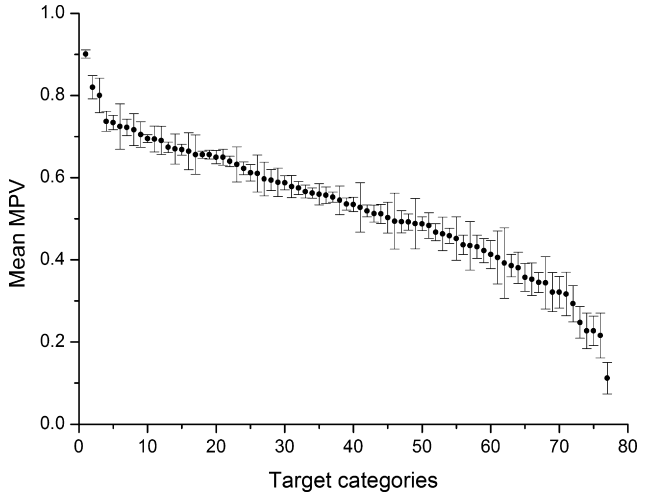


Figure 4. Means of mean probability values (mean MPVs) with standard deviations obtained from 10-fold cross-validation. Mean MPVs calculated from 10-fold cross-validation were used to assess to robustness of the predictive models. The higher an obtained mean MPV, the greater the resistance of the system to the information removal.

According to our former analyses,[24] we consider groups having mean MPV > 0.5 reliably predictable, since it indicates that DPM can classify the majority of the registered molecules into the respective target class. In this work, 45 of the studied 77 categories met this criterion, including important pharmaceutical targets such as angiotensin-converting enzyme and carbonic anhydrases, whose inhibitors are widely used antihypertensive agents and diuretics. D(2) dopamine receptor and histamine H1 receptor also exceeded the threshold, their antagonists are known as antipsychotics and antiallergic agents. High mean MPV is obtained for prostaglandin G/H synthase 1 and 2 (often referred to as cyclooxygenase 1 and 2, mean MPVs of 0.612 and 0.649), key enzymes in the mechanism of action of nonsteroidal anti-inflammatory agents. Thirty-one target groups possess medium mean MPV (0.2 < mean MPV < 0.5) such as acetylcholinesterase or reverse transcriptase (mean MPV of 0.322 and 0.392, respectively). These categories are not entirely cohesive based on their IPs, and the redistribution of molecules might improve the reliability of predictions. Only one target, arachidonate 5-lipoxygenase, produced low mean MPV (mean MPV < 0.2), indicating that DPM fails to recognize the originally registered molecules of this target group in external data. Remarkably, this worst obtained mean MPV of 0.112 is considerably higher than the lowest value for effect prediction (0.0028 for antioxidant).[23] This is in agreement with our expectations that the use of targets improves the prediction power of DPM compared to the more diverse medical effect categories.

Target Selectivity Analysis. On the basis of the validation results, 45 targets were selected for further analysis on the previously defined drug set comprising of 1157 FDA-approved drugs. The predicted off-target profiles of the investigated drugs for these targets were sorted by descending probability and were plotted to produce a so-called target selectivity plot for each drug. Figure 5 displays typical selectivity plots for four drug molecules. In the case of the antiasthmatic agent cinalukast, no targets were predicted among the applied target set therefore its selectivity plot consists of only low probability values. For the antihypertensive agent benazepril, its well-known target the angiotensin-converting enzyme was assigned with a probability of 1.00. The second highest target probability value is only 0.29 therefore benazepril is a good example of a selective drug concerning our studied targets. Olanzapine, a second generation antipsychotic shows a nonselective predicted target profile with high probability for several targets, mainly dopamine, serotonin, and muscarinic receptor subtypes in agreement with the literature. It is a well-known issue that in case of CNS drugs, the selectively nonselective (sic) drugs offer higher efficacy than the single-target acting drugs.[29] Thus, their polypharmacology, i.e. affecting multiple targets rather than acting on one single target is essential for a therapeutic effect. For the antiparkinson drug apomorphine, several targets were predicted with high and medium probability, among them unknown interactions can also be found that need further investigation in order to be proved.

Predicted Drug−Target Interactions. Investigating the subset of the selected 45 targets in the binary target profile matrix revealed that 1435 drug−target interactions were originally registered in DrugBank for this data (value of 1). Comparison with the target probability matrix resulted in 79% precision as DPM could correctly predict (>0.8 probability value) 1138 drug−target interactions of them. Applying this probability threshold for the unregistered compounds (value of 0 in the target profile matrix), 1074 new drug−target interactions were predicted. These predictions can originate from classification errors; however, considering the known incompleteness of bioactivity databases, part of the predictions may be correct.

Predicted Interactions of Antipsychotics. We examined some of the top drug−target predictions among the 1074 suggested interactions, and this review revealed that several predictions can be proved by the literature. We focused our investigation on the predicted interactions of antipsychotic molecules. According to our medical effect database presented in our former study,[23] the studied drug set contains 45 known antipsychotics for which all of the predicted targets above 0.8 probability threshold were collected. The resulting list comprises of 21 antipsychotics for which 84 drug−target interactions were suggested that were not documented in the DrugBank database. An extensive literature survey revealed that

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

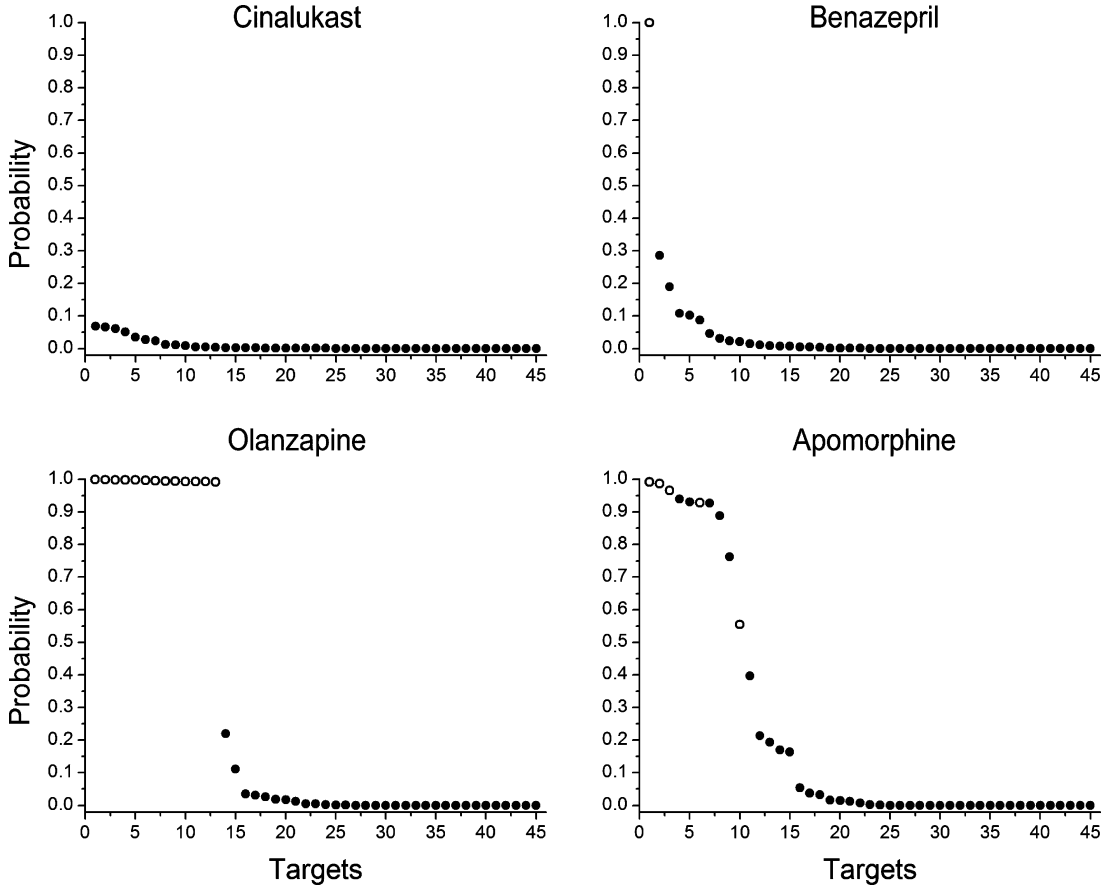


Figure 5. Examples of the selectivity plots. In a selectivity plot, predicted probability values are plotted as a function of the particular targets which are ordered by descending probability values. Hollow circles mark those targets that were already assigned to the studied molecule.

38 of the suggested interactions are already reported. Results of the survey are summarized in Table 2 for each antipsychotic. For six molecules, no drug-target interactions could be confirmed; however for the remaining drugs, potentially valuable drug−target interactions were predicted.

Fluphenazine is a first generation antipsychotic used for the treatment of schizophrenia and other psychotic disorders. In our prediction it gained high prediction value for alpha 1 adrenergic and 5-HT1A and 5-HT2A serotonergic receptors. While investigating the well-known weight gain inducing side effect of first and second generation antipsychotics, the research by Kroeze and co-workers also measured the binding affinity of fluphenazine to alpha adrenergic and serotonergic receptors.[30] Other publications also confirm an existing receptor−ligand binding both with the use of human[31,32] and rodent[33,34] 5-HT2A receptors. Also high prediction values were measured for alpha 1 adrenergic, 5-HT1A, 5-HT2A serotonergic, and M1 muscarinic acethylcholine receptors in the case of the phenothiazine derivative perphenazine which is structurally very similar to the above-mentioned fluphenazine. Literature search also confirmed an existing receptor binding for the serotonergic and adrenergic receptors.[30,32]

Sertindole is a second generation antipsychotic with well-known dopaminergic and serotonergic effects. Our results show a high prediction value for DRD 1, 5-HT1D, M1, and M2 muscarinic receptors. All four predictions were confirmed by the literature search including human and animal samples as well.[35−39]

A high prediction value was gained for different kinds of muscarinic acethylcholine receptors (in some cases all five subtypes, in other cases only a few of the existing subtypes) in the case of several compounds. A literature survey confirmed a positive receptor−ligand interaction in the case of chlorpromazine,[40−43] mesoridazine,[40] loxapine,[40,44] and sertindole[39] but failed to prove direct receptorial interaction for example in the case of prochlorpromazine or triflupromazine although these compounds have well-known adverse effects in clinical practice associated with the cholinergic autonomous nervous system (e.g., dry mouth, constipation, urinary retention, blurred vision, etc.).

A high probability of possible interaction with alpha 1 adrenergic and type 1 histaminergic receptors was also predicted several times (as mentioned above and also for prochlorperazine, mesoridazine, thiotixen and triflupromazine, pimozide, and prochlorperazine, respectively). These receptors are also associated with adverse effects typical for the antipsychotic drug class, such as orthostatic hypotension, rhinitis in the case of alpha 1 adrenergic and sedation, and weight gain for H1 receptor. And again, as with muscarinic receptors, the literature search confirmed a direct receptor−compound interaction only in some part of the cases (Table 2).

A possible interpretation of the large number of false positive targets can be the incompleteness of the target database. To investigate this issue, a validation study for a small fraction of the false positive interactions was performed by using the ChEMBL database. We could confirm that 10% of the predicted false positives are in fact true positives according to the ChEMBL database (see the Supporting Information). Thus, ChEMBL provided additional information on drug−target interactions compared to DrugBank but could not validate the

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Results of the Literature Survey Performed for Antipsychotics[a]

name	target	predicted probability	result	ref
acepromazine	5-hydroxytryptamine 2C receptor	0.986	no data	
	muscarinic acetylcholine receptor M1	0.949	yes	*b*
	muscarinic acetylcholine receptor M2	0.976	yes	*b*
	muscarinic acetylcholine receptor M3	0.839	no data	
	muscarinic acetylcholine receptor M4	0.971	no data	
	muscarinic acetylcholine receptor M5	0.979	no data	
acepromazine	5-hydroxytryptamine 2A receptor	0.996	no data	
	5-hydroxytryptamine 2C receptor	0.926	no data	
	alpha-1A adrenergic receptor	0.963	no data	
	D(1) dopamine receptor	0.999	no data	
	D(2) dopamine receptor	0.997	no data	
	muscarinic acetylcholine receptor M1	0.930	no data	
	muscarinic acetylcholine receptor M2	0.958	no data	
	muscarinic acetylcholine receptor M4	0.915	no data	
	muscarinic acetylcholine receptor M5	0.975	no data	
carphenazine	5-hydroxytryptamine 2A receptor	0.949	no data	
	alpha-1A adrenergic receptor	0.912	no data	
chlorpromazine	muscarinic acetylcholine receptor M1	0.923	yes	40,41,43
	muscarinic acetylcholine receptor M2	0.955	yes	40−43
	muscarinic acetylcholine receptor M3	0.911	yes	30,40,41,43
	muscarinic acetylcholine receptor M5	0.936	yes	40,41,43
chlorprothixene	5-hydroxytryptamine 1A receptor	0.962	no data	
	alpha-1A adrenergic receptor	0.977	no data	
	sodium-dependent noradrenaline transporter	0.984	yes	45
	sodium-dependent serotonin transporter	0.993	yes	45
droperidol	5-hydroxytryptamine 1A receptor	0.873	yes	46
	5-hydroxytryptamine 1D receptor	0.938	no data	
fencamfamine	sodium-dependent noradrenaline transporter	0.914	no data	
flupenthixol	5-hydroxytryptamine 1A receptor	0.814	yes	47
	muscarinic acetylcholine receptor M2	0.926	no data	
	muscarinic acetylcholine receptor M3	0.855	no data	
	muscarinic acetylcholine receptor M4	0.974	no data	
	muscarinic acetylcholine receptor M5	0.980	no data	
fluphenazine	5-hydroxytryptamine 1A receptor	0.869	yes	30,47
	5-hydroxytryptamine 2A receptor	0.991	yes	30,31,33,47,48
	Alpha-1A adrenergic receptor	0.936	yes	30,32,49
loxapine	DNA	0.856	no data	
	histamine H1 receptor	0.925	yes	30,32,44,49
	muscarinic acetylcholine receptor M1	0.942	yes	40,44
	muscarinic acetylcholine receptor M4	0.865	yes	40
mesoridazine	alpha-1A adrenergic receptor	0.860	yes	32,49,50
	D(1) dopamine receptor	0.996	yes	51,52
	muscarinic acetylcholine receptor M1	0.900	yes	40
	muscarinic acetylcholine receptor M2	0.948	yes	40
	muscarinic acetylcholine receptor M3	0.949	yes	40
	muscarinic acetylcholine receptor M4	0.943	yes	40
	muscarinic acetylcholine receptor M5	0.906	yes	40
methotrimeprazine	5-hydroxytryptamine 1A receptor	0.862	no data	
perphenazine	5-hydroxytryptamine 1A receptor	0.950	yes	30
	5-hydroxytryptamine 2A receptor	0.982	yes	30
	alpha-1A adrenergic receptor	0.906	yes	30,32,49
	muscarinic acetylcholine receptor M1	0.917	no data	
pimozide	histamine H1 receptor	0.932	yes	30,37
prochlorperazine	5-hydroxytryptamine 1A receptor	0.930	no data	
	5-hydroxytryptamine 2A receptor	0.965	no data	
	5-hydroxytryptamine 2C receptor	0.900	yes	33,53
	alpha-1A adrenergic receptor	0.986	yes	32,49
	D(1) dopamine receptor	0.976	no data	
	histamine H1 receptor	0.905	yes	32,49
	muscarinic acetylcholine receptor M1	0.944	no data	
	muscarinic acetylcholine receptor M2	0.899	no data	

dx.doi.org/10.1021/ci3004489 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. continued

name	target	predicted probability	result	ref
	muscarinic acetylcholine receptor M3	0.947	no data	
	muscarinic acetylcholine receptor M5	0.860	no data	
propericiazine	5-hydroxytryptamine 1A receptor	0.907	no data	
	5-hydroxytryptamine 2A receptor	0.985	no data	
	D(2) dopamine receptor	0.984	no data	
	muscarinic acetylcholine receptor M1	0.906	no data	
sertindole	5-hydroxytryptamine 1D receptor	0.878	yes	30,37,38
	D(1) dopamine receptor	0.875	yes	35
	muscarinic acetylcholine receptor M1	0.945	yes	39
	muscarinic acetylcholine receptor M2	0.826	yes	39
thiothixene	alpha-1A adrenergic receptor	0.851	yes	30
trifluperazine	5-hydroxytryptamine 1A receptor	0.908	yes	30,54
	5-hydroxytryptamine 2A receptor	0.995	yes	30,31,48,54−56
	D(1) dopamine receptor	1.000	yes	57,58
triflupromazine	5-hydroxytryptamine 1A receptor	0.872	no data	
	5-hydroxytryptamine 2A receptor	0.987	no data	
	alpha-1A adrenergic receptor	0.962	no data	
	histamine H1 receptor	0.976	yes	59
	muscarinic acetylcholine receptor M4	0.831	no data	
	muscarinic acetylcholine receptor M5	0.935	no data	
zuclopenthixol	5-hydroxytryptamine 1A receptor	0.894	no data	
	5-hydroxytryptamine 2C receptor	0.914	no data	
	muscarinic acetylcholine receptor M1	0.808	no data	

[a]For each studied antipsychotic, the predicted drug−target interactions (probability > 0.8) are displayed. An extensive literature survey revealed those interactions for that evidence already exists and the corresponding reference is provided. [b]Not listed in DrugBank table "Targets" but mentioned in the "Pharmacology" section.

false positive interactions so widely. The reason might be that targets which are important in the clinical effect are in the focus of the majority of the databases and receptors which mediate adverse effects are not so well documented. Another interpretation can be that these, usually general and not easily quantifiable side effects such as dry mouth and constipation for example, are traditionally considered as anticholinergic, but in some cases, these might be at least partially mediated by other transmitter systems as well in line with the model of polypharmacology.

Those predictions for which no literature evidence exists might be demonstrated experimentally since it is also possible that a given drug was not tested against the predicted off-targets. On the basis of our previous DPM analysis using medical effects instead of targets, we already obtained valuable predictions which were validated by in vitro experiments with a high hit rate of 47−84% (unpublished results).

CONCLUSIONS

In this paper, the applicability of DPM for in silico target fishing was investigated using 77 target classes, each containing at least 10 active molecules. High classification accuracies were obtained in all cases. The robustness of the prediction results was checked by 10-fold cross-validation which revealed those targets for that the performance of DPM is highly reliable. These 45 categories were used in a subsequent analysis which aimed at predicting the off-target profiles (limited to the studied categories) of currently approved FDA drugs. 79% of the known drug−target interactions in this data set were correctly predicted by DPM. Additionally 1074 new drug−target interactions were suggested. A pilot study was presented that aimed at confirming part of the suggested drug−target interactions for antipsychotic molecules by a literature survey. 45% of the 84 suggested interactions were demonstrated and references were provided.

Our study supports the theory of polypharmacology by pointing out that drugs usually act on several targets and have a characteristic off-target profile that contains valuable information for future drug development. DPM is able to find previously unknown pharmaceutical targets of the studied compounds; therefore, the method may serve as a good starting point for drug repositioning that aims at finding new medical applications of well-known drug molecules.

ASSOCIATED CONTENT

Supporting Information

Detailed description of IP generation, example calculation on a small dataset that illustrates the different steps of the DPM method, validation of the predicted drug−target interactions by ChEMBL data, and Table S1: list of the names and the Protein Data Bank entries of the 135 proteins used. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +36 1 372 2500 ext. 8780. Fax: +36 1 381 2172. E-mail: malna@elte.hu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the Hungarian Academy of Sciences, HAS-ELTE research group ID: 01055. The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. [208319]. This work was

supported by the European Union and cofinanced by the European Regional Development Fund (New Széchenyi Plan KMOP-1.1.1-09/1-2009-0045).

■ ABBREVIATIONS

AUC, area under the curve; CCA, canonical correlation analysis; DPM, drug profile matching; TP, target profile; FDA, Food and Drug Administration; FPR, false positive rate; IP, interaction pattern; LDA, linear discriminant analysis; PDB, Protein Data Bank; ROC, receiver operating characteristic; TPR, true positive rate

■ REFERENCES

(1) Nettles, J. H.; Jenkins, J. L.; Bender, A.; Deng, Z.; Davies, J. W.; Glick, M. Bridging chemical and biological space: "target fishing" using 2D and 3D molecular descriptors. *J. Med. Chem.* **2006**, *49*, 6802−6810.

(2) Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4*, 682−690.

(3) Metz, J. T.; H., P. Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks. *Curr. Opin. Chem. Biol.* **2010**, *14*, 498−504.

(4) Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V.; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; Bender, A. From in silico target prediction to multi-target drug design: current databases, methods and applications. *J. Proteom.* **2011**, *74*, 2554−2574.

(5) Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods? *Drug Discov. Today* **2002**, *7*, 903−911.

(6) Salim, N.; Holliday, J.; Willett, P. Combination of fingerprint-based similarity coefficients using data fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435−442.

(7) AbdulHameed, M. D.; Chaudhury, S.; Singh, N.; Sun, H.; Wallqvist, A.; Tawa, G. J. Exploring polypharmacology using a ROCS-based target fishing approach. *J. Chem. Inf. Model.* **2012**, *52*, 492−505.

(8) Poroikov, V.; Filimonov, D.; Lagunin, A.; Gloriozova, T.; Zakharov, A. PASS: identification of probable targets and mechanisms of toxicity. *SAR QSAR Environ. Res.* **2007**, *18*, 101−110.

(9) Nigsch, F.; Bender, A.; Jenkins, J. L.; Mitchell, J. B. Ligand-target prediction using Winnow and naive Bayesian algorithms and the implications of overall performance statistics. *J. Chem. Inf. Model.* **2008**, *48*, 2313−2325.

(10) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197−206.

(11) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175−181.

(12) Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609−614.

(13) Chen, X.; Ung, C. Y.; Chen, Y. Can an in silico drug-target search method be used to probe potential mechanisms of medicinal plant ingredients? *Nat. Prod. Rep.* **2003**, *20*, 432−444.

(14) Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **2006**, *34*, W219−224.

(15) Briem, H.; Kuntz, I. D. Molecular similarity based on DOCK-generated fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.

(16) Lessel, U. F.; Briem, H. Flexsim-X: a method for the detection of molecules with similar biological activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 246−253.

(17) Fukunishi, Y.; Hojo, S.; Nakamura, H. An efficient in silico screening method based on the protein-compound affinity matrix and its application to the design of a focused library for cytochrome P450 (CYP) ligands. *J. Chem. Inf. Model.* **2006**, *46*, 2610−2622.

(18) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **1995**, *2*, 107−118.

(19) Kauvar, L. M.; Villar, H. O.; Sportsman, J. R.; Higgins, D. L.; Schmidt, D. E., Jr. Protein affinity map of chemical space. *J. Chromatogr. B Biomed. Sci. Appl.* **1998**, *715*, 93−102.

(20) Bender, A.; Jenkins, J. L.; Glick, M.; Deng, Z.; Nettles, J. H.; Davies, J. W. "Bayes affinity fingerprints" improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? *J. Chem. Inf. Model.* **2006**, *46*, 2445−2456.

(21) Briem, H.; Lessel, U. F. In vitro and in silico affinity fingerprints: Finding similarities beyond structural classes. *Perspect. Drug Discovery Des.* **2000**, *20*, 231−244.

(22) Fukunishi, Y.; Mikami, Y.; Takedomi, K.; Yamanouchi, M.; Shima, H.; Nakamura, H. Classification of chemical compounds by protein-compound docking for use in designing a focused library. *J. Med. Chem.* **2006**, *49*, 523−533.

(23) Simon, Z.; Peragovics, A.; Vigh-Smeller, M.; Csukly, G.; Tombor, L.; Yang, Z.; Zahoranszky-Kohalmi, G.; Vegner, L.; Jelinek, B.; Hari, P.; Hetenyi, C.; Bitter, I.; Czobor, P.; Malnasi-Csizmadia, A. Drug effect prediction by polypharmacology-based interaction profiling. *J. Chem. Inf. Model.* **2012**, *52*, 134−145.

(24) Peragovics, A.; Simon, Z.; Brandhuber, I.; Jelinek, B.; Hari, P.; Hetenyi, C.; Czobor, P.; Malnasi-Csizmadia, A. Contribution of 2D and 3D Structural Features of Drug Molecules in the Prediction of Drug Profile Matching. *J. Chem. Inf. Model.* **2012**, *52*, 1733−1744.

(25) Jiang, X.; Kumar, K.; Hu, X.; Wallqvist, A.; Reifman, J. DOVIS 2.0: an efficient and easy to use parallel virtual screening tool based on AutoDock 4.0. *Chem. Cent. J.* **2008**, *2*, 18.

(26) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1152.

(27) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided. Mol. Des.* **2002**, *16*, 11−26.

(28) Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488−508.

(29) Roth, B. L.; Sheffler, D. J.; Kroeze, W. K. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* **2004**, *3*, 353−359.

(30) Kroeze, W. K.; Hufeisen, S. J.; Popadak, B. A.; Renock, S. M.; Steinberg, S.; Ernsberger, P.; Jayathilake, K.; Meltzer, H. Y.; Roth, B. L. H1-histamine receptor affinity predicts short-term weight gain for typical and atypical antipsychotic drugs. *Neuropsychopharmacology* **2003**, *28*, 519−526.

(31) Seeman, P.; Tallerico, T. Antipsychotic drugs which elicit little or no parkinsonism bind more loosely than dopamine to brain D2 receptors, yet occupy high levels of these receptors. *Mol. Psychiatry* **1998**, *3*, 123−134.

(32) Richelson, E.; Nelson, A. Antagonism by neuroleptics of neurotransmitter receptors of normal human brain in vitro. *Eur. J. Pharmacol.* **1984**, *103*, 197−204.

(33) Roth, B. L.; Ciaranello, R. D.; Meltzer, H. Y. Binding of typical and atypical antipsychotic agents to transiently expressed 5-HT1C receptors. *J. Pharmacol. Exp. Ther.* **1992**, *260*, 1361−1365.

(34) Canton, H.; Verriele, L.; Millan, M. J. Competitive antagonism of serotonin (5-HT)2C and 5-HT2A receptor-mediated phosphoinositide (PI) turnover by clozapine in the rat: a comparison to other antipsychotics. *Neurosci. Lett.* **1994**, *181*, 65−68.

(35) Arnt, J.; Skarsfeldt, T. Do novel antipsychotics have similar pharmacological characteristics? A review of the evidence. *Neuropsychopharmacology* **1998**, *18*, 63−101.

(36) Balle, T.; Perregaard, J.; Ramirez, M. T.; Larsen, A. K.; Soby, K. K.; Liljefors, T.; Andersen, K. Synthesis and structure-affinity relationship investigations of 5-heteroaryl-substituted analogues of the antipsychotic sertindole. A new class of highly selective alpha(1) adrenoceptor antagonists. *J. Med. Chem.* **2003**, *46*, 265−283.

(37) Richelson, E.; Souder, T. Binding of antipsychotic drugs to human brain receptors focus on newer generation compounds. *Life Sci.* **2000**, *68*, 29−39.

(38) Schotte, A.; Janssen, P. F.; Gommeren, W.; Luyten, W. H.; Van Gompel, P.; Lesage, A. S.; De Loore, K.; Leysen, J. E. Risperidone compared with new and reference antipsychotic drugs: in vitro and in vivo receptor binding. *Psychopharmacology (Berl)* **1996**, *124*, 57−73.

(39) Wermuth, C. G. Selective optimization of side activities: another way for drug discovery. *J. Med. Chem.* **2004**, *47*, 1303−1314.

(40) Bolden, C.; Cusack, B.; Richelson, E. Antagonism by antimuscarinic and neuroleptic compounds at the five cloned human muscarinic cholinergic receptors expressed in Chinese hamster ovary cells. *J. Pharmacol. Exp. Ther.* **1992**, *260*, 576−580.

(41) Bymaster, F. P.; Felder, C. C.; Tzavara, E.; Nomikos, G. G.; Calligaro, D. O.; McKinzie, D. L. Muscarinic mechanisms of antipsychotic atypicality. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **2003**, *27*, 1125−1143.

(42) Kovacs, I.; Yamamura, H. I.; Waite, S. L.; Varga, E. V.; Roeske, W. R. Pharmacological comparison of the cloned human and rat M2 muscarinic receptor genes expressed in the murine fibroblast (B82) cell line. *J. Pharmacol. Exp. Ther.* **1998**, *284*, 500−507.

(43) Liegeois, J. F.; Bruhwyler, J.; Damas, J.; Nguyen, T. P.; Chleide, E. M.; Mercier, M. G.; Rogister, F. A.; Delarge, J. E. New pyridobenzodiazepine derivatives as potential antipsychotics: synthesis and neurochemical study. *J. Med. Chem.* **1993**, *36*, 2107−2114.

(44) Liao, Y.; Venhuis, B. J.; Rodenhuis, N.; Timmerman, W.; Wikstrom, H.; Meier, E.; Bartoszyk, G. D.; Bottcher, H.; Seyfried, C. A.; Sundell, S. New (sulfonyloxy)piperazinyldibenzazepines as potential atypical antipsychotics: chemistry and pharmacological evaluation. *J. Med. Chem.* **1999**, *42*, 2235−2244.

(45) Tatsumi, M.; Jansen, K.; Blakely, R. D.; Richelson, E. Pharmacological profile of neuroleptics at human monoamine transporters. *Eur. J. Pharmacol.* **1999**, *368*, 277−283.

(46) Gentil, B.; Macquin-Mavier, I.; Lienhart, A.; Harf, A. Droperidol prevents serotonin-induced bronchospasm in the guinea pig. *Anesth Analg.* **1991**, *72*, 612−615.

(47) Toll, L.; Berzetei-Gurske, I. P.; Polgar, W. E.; Brandt, S. R.; Adapa, I. D.; Rodriguez, L.; Schwartz, R. W.; Haggart, D.; O'Brien, A.; White, A.; Kennedy, J. M.; Craymer, K.; Farrington, L.; Auh, J. S. Standard binding and functional assays related to medications development division testing for potential cocaine and opiate narcotic treatment medications. *NIDA Res Monogr.* **1998**, *178*, 440−466.

(48) Roth, B. L.; Tandra, S.; Burgess, L. H.; Sibley, D. R.; Meltzer, H. Y. D4 dopamine receptor binding affinity does not distinguish between typical and atypical antipsychotic drugs. *Psychopharmacology (Berl)* **1995**, *120*, 365−368.

(49) Richelson, E. Neuroleptic binding to human brain receptors: relation to clinical effects. *Ann. N.Y. Acad. Sci.* **1988**, *537*, 435−442.

(50) Bylund, D. B. Interactions of neuroleptic metabolites with dopaminergic, alpha adrenergic and muscarinic cholinergic receptors. *J. Pharmacol. Exp. Ther.* **1981**, *217*, 81−86.

(51) Black, J. L.; Richelson, E. Antipsychotic drugs: prediction of side-effect profiles based on neuroreceptor data derived from human brain tissue. *Mayo Clin. Proc.* **1987**, *62*, 369−372.

(52) Poling, A.; Cleary, J.; Berens, K.; Thompson, T. Neuroleptics and learning: effects of haloperidol, molindone, mesoridazine and thioridazine on the behavior of pigeons under a repeated acquisition procedure. *J. Pharmacol. Exp. Ther.* **1990**, *255*, 1240−1245.

(53) Herrick-Davis, K.; Grinde, E.; Teitler, M. Inverse agonist activity of atypical antipsychotic drugs at human 5-hydroxytryptamine2C receptors. *J. Pharmacol. Exp. Ther.* **2000**, *295*, 226−232.

(54) Scott, M. K.; Martin, G. E.; DiStefano, D. L.; Fedde, C. L.; Kukla, M. J.; Barrett, D. L.; Baldy, W. J.; Elgin, R. J., Jr.; Kesslick, J. M.; Mathiasen, J. R.; et al. Pyrrole mannich bases as potential antipsychotic agents. *J. Med. Chem.* **1992**, *35*, 552−558.

(55) Seeman, P.; Corbett, R.; Van Tol, H. H. Atypical neuroleptics have low affinity for dopamine D2 receptors or are selective for D4 receptors. *Neuropsychopharmacology* **1997**, *16*, 93−110 discussion 111−135.

(56) Yevich, J. P.; New, J. S.; Smith, D. W.; Lobeck, W. G.; Catt, J. D.; Minielli, J. L.; Eison, M. S.; Taylor, D. P.; Riblet, L. A.; Temple, D. L., Jr. Synthesis and biological evaluation of 1-(1,2-benzisothiazol-3-yl)- and (1,2-benzisoxazol-3-yl)piperazine derivatives as potential antipsychotic agents. *J. Med. Chem.* **1986**, *29*, 359−369.

(57) Creese, I.; Burt, D. R.; Snyder, S. H. Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *J. Neuropsychiatry Clin. Neurosci.* **1996**, *8*, 223−226.

(58) Burt, D. R.; Creese, I.; Snyder, S. H. Properties of [3H]haloperidol and [3H]dopamine binding associated with dopamine receptors in calf brain membranes. *Mol. Pharmacol.* **1976**, *12*, 800−812.

(59) Tran, V. T.; Chang, R. S.; Snyder, S. H. Histamine H1 receptors identified in mammalian brain membranes with [3H]mepyramine. *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 6290−6294.