# Exploring Polypharmacology Using a ROCS-Based Target Fishing Approach

Mohamed Diwan M. AbdulHameed,* Sidhartha Chaudhury, Narender Singh, Hongmao Sun, Anders Wallqvist, and Gregory J. Tawa*

Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland 21702, United States

Ⓢ *Supporting Information*

ABSTRACT: Polypharmacology has emerged as a new theme in drug discovery. In this paper, we studied polypharmacology using a ligand-based target fishing (LBTF) protocol. To implement the protocol, we first generated a chemogenomic database that links individual protein targets with a specified set of drugs or target representatives. Target profiles were then generated for a given query molecule by computing maximal shape/chemistry overlap between the query molecule and the drug sets assigned to each protein target. The overlap was computed using the program ROCS (Rapid Overlay of

Chemical Structures). We validated this approach using the Directory of Useful Decoys (DUD). DUD contains 2950 active compounds, each with 36 property-matched decoys, against 40 protein targets. We chose a set of known drugs to represent each DUD target, and we carried out ligand-based virtual screens using data sets of DUD actives seeded into DUD decoys for each target. We computed Receiver Operator Characteristic (ROC) curves and associated area under the curve (AUC) values. For the majority of targets studied, the AUC values were significantly better than for the case of a random selection of compounds. In a second test, the method successfully identified off-targets for drugs such as rimantadine, propranolol, and domperidone that were consistent with those identified by recent experiments. The results from our ROCS-based target fishing approach are promising and have potential application in drug repurposing for single and multiple targets, identifying targets for orphan compounds, and adverse effect prediction.
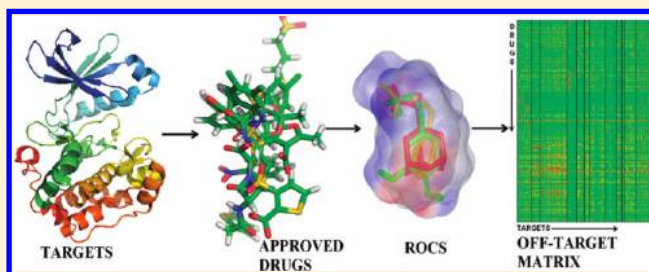
1. INTRODUCTION

Polypharmacology has emerged as a new theme in drug discovery.[1-4] In contrast to the traditional view of one drug against one target, polypharmacology focuses on the fact that one drug can hit multiple targets.[1] Polypharmacology is desirable in the case of complex diseases that involve functional modulation of multiple proteins such as cancer.[5] Identification of compounds that interact with multiple proteins in a particular disease network may be a good starting point for drug discovery. However, protein targets outside of these networks may interact with putative drugs. This may either cause unwanted side effects or it may help in the modulation of different diseases. Therefore, identification of these off-target proteins may facilitate drug repurposing and the determination of toxic liabilities. Identifying new indications for old drugs was reported to be the best and most economical way to bring a drug to market.[6]

Computational approaches have traditionally focused on studying ligand interactions with a single target and have been successfully used in lead identification and optimization studies.[7,8] These methods complement much more expensive experimental approaches to drug design and have been integrated into virtually all modern drug-discovery programs. Similarly, computational off-target profiling methods or "target fishing" are complementary to

the experimental screening approaches. It is not possible to test each compound against every possible target. The application of computational approaches in off-target prediction has been reviewed.[9,10] Many structure-based target fishing (SBTF) approaches, such as INVDOCK[11] and Target Fishing Dock (TarFisDock),[12] are reported in the literature.[8] The basic idea behind SBTF is the inverse of docking. In the usual docking experiments, a set of ligands is docked into a particular target, and the results are ranked by docking score. However, in SBTF, a single ligand is docked into many targets, and the potential targets are ranked by docking[8,12,13] or Z-score.[14] SBTF approaches are of limited utility for major drug targets like G-protein coupled receptors (GPCRs) and ion channels, because their crystal structures are not available. Nearly 50% of all recently launched drugs were reported to target GPCRs.[15] Furthermore, issues such as protein flexibility and the treatment of water-mediated interactions in the active site are other limiting factors of this approach.

Ligand-based target fishing approaches do not have these limitations. For many targets that do not have an experimentally determined structure, there is still a known set of active ligands.
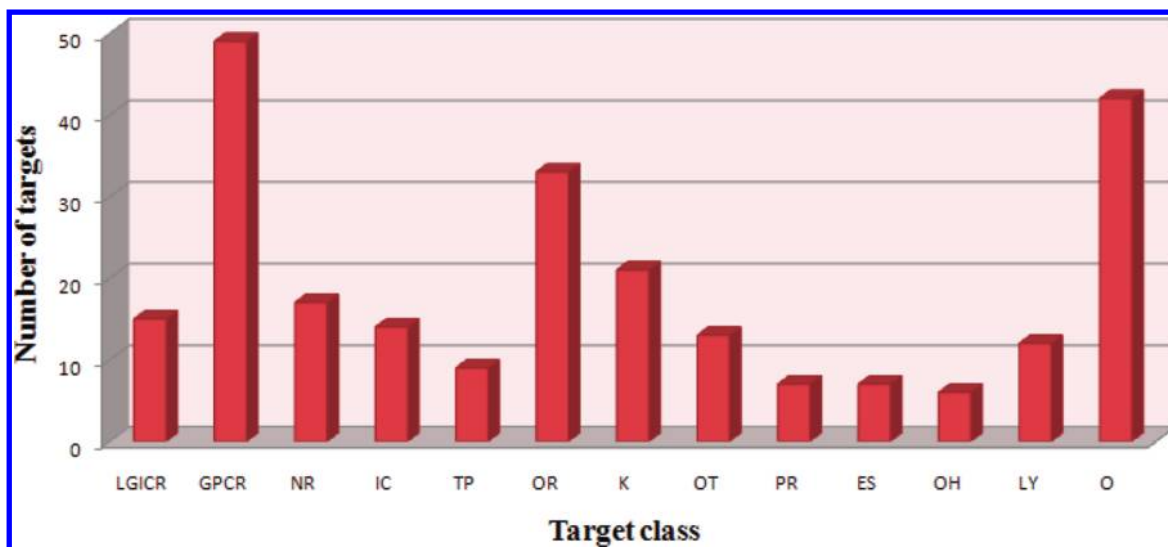
Figure 1. Number of targets in different target classes. Drug targets were grouped into 13 major classes. Abbreviations: LGICR, ligand-gated ion channel receptor; GPCR, G-protein coupled receptor; NR, nuclear receptor; IC, ion channels; TP, transporters; OR, oxido-reductase enzymes; K, kinases; OT, other transferses; PR, proteases; ES, esterases; OH, other hydrolases; LY, lyases, ligases, and isomerases; O, others.

This allows the application of ligand-based approaches in the study of a wide variety of targets. The fundamental idea underlying ligand-based approaches is that two similar ligands are likely to have similar target-binding profiles. Ligand-based target fishing approaches utilize either similarity-based screening or machine learning models. Similarity-based target fishing is conducted by determining the protein targets for screening, identifying ligands to represent those targets, and choosing the similarity method for comparing ligands. Keiser et al. have used 2D-similarity searching along with a BLAST-like statistical model to successfully predict the off-targets of a set of known drugs.[16,17] Scitegic ECFP4 and Daylight topological fingerprints were used as the descriptors for the similarity search.[17] Nettles et al. have used feature point pharmacophores (FEPOPS) and highlighted the ability of the 3D similarity search approach to identify novel scaffolds.[18] Multiple-category Bayesian modeling, Shannon Entropy Descriptors (SHED), and morphological similarity have also been used to carry out target fishing.[19−21] Among 3D-similarity search approaches, the ROCS program[22] is considered to be a de facto standard. There are many reports on the successful application of ROCS in lead identification and optimization.[23−26]

In this paper, we have explored the application of ROCS in target fishing. We used public data sources including Drug Bank[27] and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[28] to create a chemogenomic database linking drug molecules to protein targets. This allowed us to develop a ligand-based target fishing (LBTF) protocol using the ROCS program. We have extended the group fusion and inverse docking approaches to develop our ROCS-based target fishing (RBTF) approach. Group fusion refers to the use of multiple reference structures in a similarity search. On the basis of our database annotation, multiple reference structures were used to represent the targets. Typically, one or more query molecules are screened against multiple target sets. This is the inverse of traditional ligand-based screening approaches. We first validated this approach using the Directory of Useful Decoys (DUD) data set.[29] We found that, for the majority of targets, the enrichment of known actives was significantly higher using RBTF than that for which a random selection of compounds was used as the screening method. We used the RBTF method to generate a drug−target matrix. For a subset of drugs in our matrix, we identified off-targets that were recently reported in the literature.

To the best of our knowledge, this study is the first to use the 3D-shape/chemical similarity analysis program ROCS to generate off-target profiles of drugs. The results demonstrate that a shape and chemical similarity-based target fishing approach using a robust drug−target matrix can successfully identify off-targets. The methodology has potential application in the prediction of toxicity, identification of targets of orphan compounds, and drug repurposing.

2. METHODS

2.1. Creation of a Chemogenomic Database. In order for a chemogenomic database to be amenable to automated data mining, it must contain a clear annotation of targets and chemical structures.[10] The annotation will be necessary to distinguish drug target and compound classes such as bacterial targets from human targets or antibiotics from cardiovascular drugs. We used Drug Bank to obtain the initial drug-target information, and a detailed literature survey identified 245 of these targets as primary drug targets.[30,31] The drug targets were grouped according to biochemical classification into 13 major classes (Figure 1 shows the major target classes). Overall, our coverage of primary therapeutic targets agrees with the previous report of Imming et al.[30] There are 20 targets that have at least one known approved drug molecule, 17 targets that have two known drug molecules, and 208 targets that have three or more drug molecules which are known to interact with them (Figure 2). We have also included the species information for the drug targets, identifying particular drug targets as bacterial, viral, or human. The Drug Bank "target ID" was used as the standard nomenclature for the targets.

Approved drug molecules obtained from the Drug Bank database were filtered using the Filter module of the OpenEye Scientific Software[32] to remove protein-based therapeutics such as insulin and oxytocin. Filtering was carried out with the following parameters: molecular weight (150 to 800), ring systems (0 to 10), number of carbons (5 to 40), rotatable bonds (0 to 15), and allowed elements (H, C, N, O, F, S, Cl,
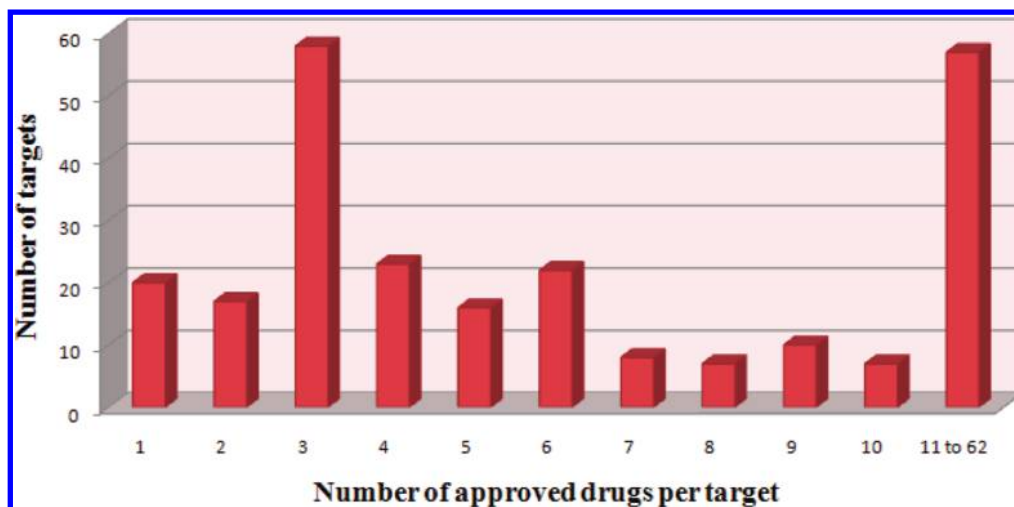
Figure 2. Number of targets based on the number of approved drugs per target.

Br, and P). After filtering, a final database of 1150 approved drug molecules was obtained. The approved drug molecules in our database were grouped into 14 major classes according to the anatomical therapeutic chemical (ATC) classification system of the World Health Organization (WHO).[33] Drug Bank and KEGG were used to obtain the ATC codes of the drugs. In the ATC system, drugs are categorized into groups at five different levels, namely the following: (1) anatomical main group, (2) therapeutic subgroup, (3) pharmacological subgroup, (4) chemical subgroup, and (5) chemical substance. The first level, which indicates the anatomical main group, consists of a one letter code; e.g., "J" refers to anti-infective agents. The Drug Bank number (DB number) was used as the standard nomenclature for approved drugs.

The 245 targets along with their approved drugs were organized into a chemogenomics matrix with rows, i, defined by drugs (1150 in number) and columns, j, defined by targets (245 in number). A 16 × 8 subsection of the matrix is shown in Figure 3A. We represent the matrix elements by the symbol, O_{ij}^{dt}, where the superscript designation, dt, represents drug-target. These matrix elements are set to either 1 or 0 depending on whether or not the drug, i, has a Drug Bank documented interaction with the protein target, j.

The importance of the matrix is that we can easily find drug sets to represent a given target by selecting an appropriate column of the matrix, scanning downward through the rows and noting where the 1's and 0's are located. In addition, we can find the targets associated with a given drug by selecting an appropriate row of the matrix, scanning horizontally across the columns, and noting where the 1's and 0's are located. The chemogenomics database is composed of the chemogenomics matrix, as shown in Figure 3A, and a structural data file containing the drug structures and associated data.

2.2. Validation Study Using DUD. DUD is one of the most commonly used data sets for the analysis and validation of structure-based and ligand-based virtual screening methods. It contains approximately 3000 active ligands[29] distributed across 40 protein targets. For every active ligand, 36 inactive "decoy" molecules were selected that are physically and chemically similar but topologically distinct from the active ligands. This approach to selecting decoys avoids the bias in screening efficiency that arises due to dissimilarity in physical properties between active and inactive compounds present in the same database. Since we

wanted to use approved drugs as target representatives, we chose 30 targets in the DUD data set that have at least one approved drug molecule. The approved drug molecules for each target (target representatives) were obtained from our chemogenomics database. A screening database was created by seeding the DUD actives into the decoys for each of the 30 targets, and the ability of target representatives to discern active from decoy compounds was analyzed. In a second study, the DUD actives were mixed with the entire decoy set (cross decoys), and the ability of the target representatives to discern active from decoy compounds was analyzed for each respective target.

Multiple conformations of the target representatives were generated by using OMEGA (Open Eye Scientific Software)[34] with the following parameters: number of allowed conformations (nconfs) = 400, root-mean-square distance (RMS) = 0.5 Å, and Ewindow = 10 kcal/mol. Ewindow is the value used to discard high-energy conformations. The Merck Molecular Force Field (MMFF) was used. The maximum allowed conformations per compound was set to 400 to ensure complete conformational coverage. The same OMEGA parameters were used to generate a single (nconfs = 1) low-energy conformation of DUD active molecules and decoys.

The ROCS program (OpenEye Scientific Software)[35] was used to carry out the virtual screens between the DUD screening databases and the target representatives. The ROCS run was carried out with the following parameters: rankby = combo and besthits = 1. In this screen, ROCS compares database compounds and target representatives by aligning the compounds such that their volumes and chemical features are as closely matched as possible. This match is represented by a combo score which ranges from 0 to 2. If the combo score is close to 2, then the molecules have an excellent shape and chemical-feature match. On the other hand, values close to 0 imply a poor shape and chemical-feature match. The screening score for a particular database compound was set to the maximum combo score between the database compound and any of the target representatives. The use of the maximum combo score is consistent with group fusion ideas[36−38] that utilize the MAX fusion rule. MAX fusion is an extreme case, where all of the data are thrown out and only the maximum value is retained.

An overview of steps used in this validation study is shown in Figure 4. For example, in the case of COX-2, there are 408
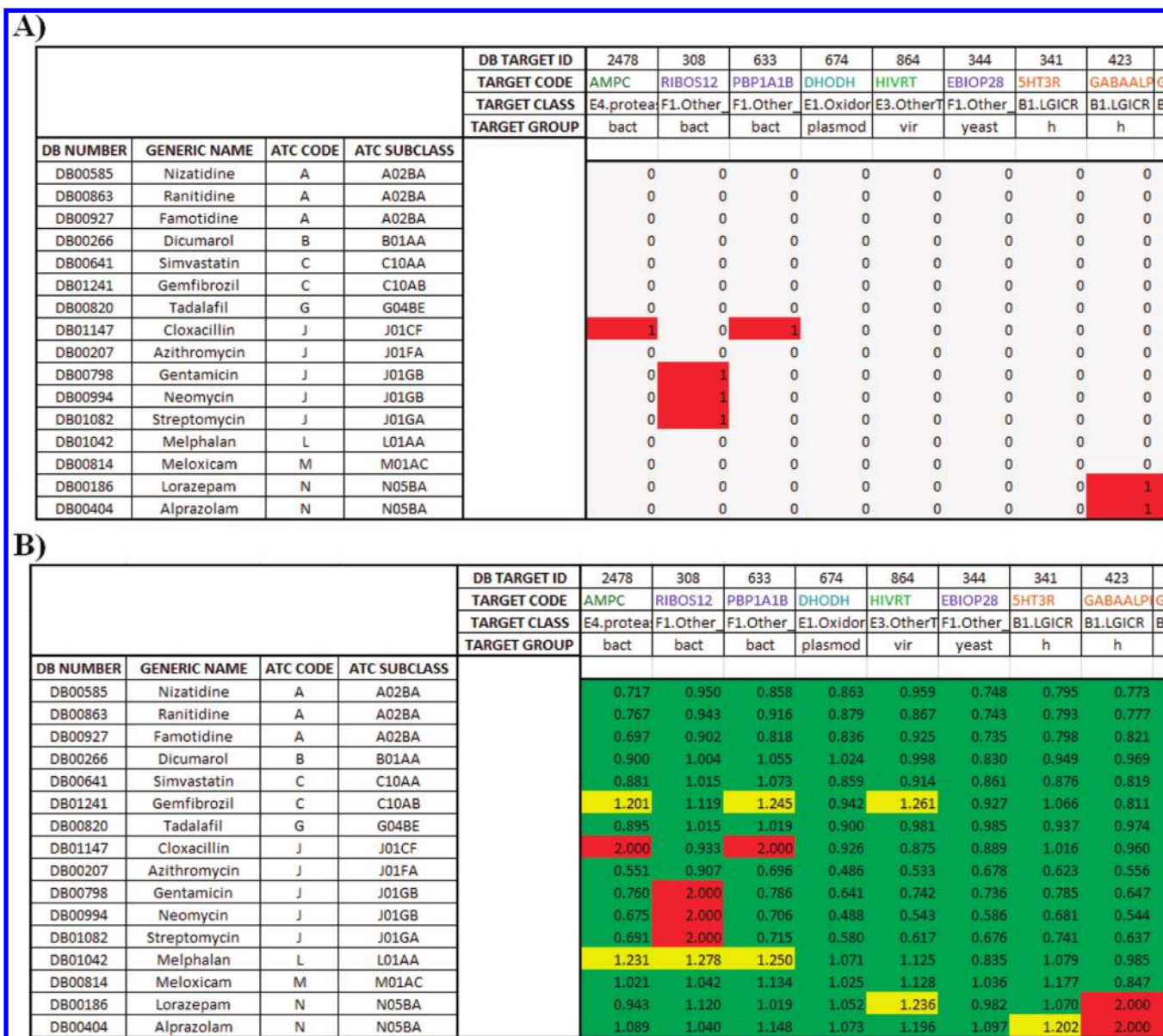
Figure 3. Chemogenomics matrix. Each drug in the matrix is annotated with its Drug Bank number, generic name, ATC code, and ATC subclass. Each target in the matrix is annotated with Drug Bank target ID, target name, and target class based on biochemical classification and species information. (A) Matrix entry values of 1 and 0 denote documented and unknown interactions, respectively, between the drug and protein. (B) Drug-target matrix. The matrix elements, O'^{dt}_{ij}, are maximum combo score values (see discussion in section 2.3). Matrix element values, O'^{dt}_{ij}, close to 2 indicate a high likelihood of interaction between the drug, i, and target, j, whereas O'^{dt}_{ij} values close to 0 indicate a small likelihood that the drug will interact with the target. Abbreviations: DB number, Drug Bank number; ATC, anatomical therapeutic chemical classification; AMPC, AmpC β-lactamase; RIBOS12, 30s ribosomal protein S12; PBP1A1B, penicillin-binding protein 1A/1B; DHODH, dihydroorotate dehydrogenase; HIVRT, HIV reverse transcriptase; EBIOP28, ergosterol biosynthetic protein 28; 5HT3R, 5HT3 receptor; GABAALP, GABA receptor subunit alpha; LGICR, ligand-gated ion channel receptor; bact, bacteria; vir, virus; h, humans.

active molecules and 13 289 decoys in DUD. The total of 13 697 molecules was used as the query set for the first screening run, i.e., DUD target-focused decoy screens. Thirty-six approved drugs, which are known to interact with COX-2, were extracted from our chemogenomics database and were used as target representatives. The query set (i.e., 13 697 molecules) was used to screen the target representatives. The similarity of each molecule in the query set to every molecule in the target representative set was calculated, and the maximum combo score was selected. Each DUD query molecule will now have an associated maximum combo score that gives the similarity between the DUD query molecule and the target representative set.

The resulting file was sorted according to combo score. A receiver-operating characteristic (ROC) plot was generated, and the AUC was computed. Similar computations were carried out for all 30 targets using DUD target-focused decoys and cross decoys. Ideally, if the target representatives are capable of identifying the actives, then higher AUC values (close to one) are expected.

2.3. Generation of the Drug−Target Matrix. Figure 5 shows the workflow for the creation of a drug−target matrix. In the first step, we constructed an 1150 × 245 chemogenomics database as discussed in section 2.1. In the second step, we constructed a 1150 × 1150 drug−drug similarity matrix. This was done by using ROCS to align and generate combo scores
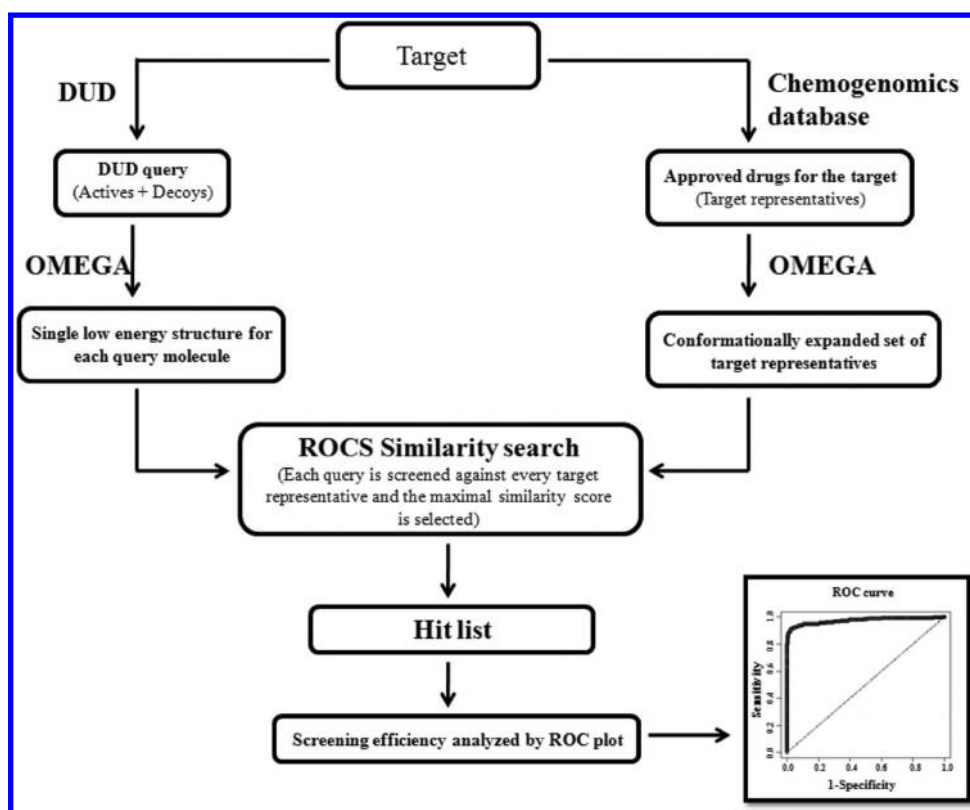
Figure 4. Validation study using Directory of useful Decoys (DUD).

for each and every pair of drugs. Each target representative drug molecule was conformationally expanded using OMEGA, and the query drug molecule was prepared with the same parameters as those used for the DUD data set. When two drug molecules, drug 1 and drug 2, were compared using ROCS, all pairwise alignments and combo score values between the conformational sets of the two molecules were evaluated. The final combo score between drug 1 and drug 2 was then set to the maximum combo score generated from the set of pairwise alignments. The drug−drug matrix is shown in step 2 of Figure 5, where O_{ij}^{dd} represents the maximum combo score between drugs i and j, and the superscript designation, dd, represents drug−drug.

In step 3, we combine the information from the chemogenomics database (step 1 in Figure 5) with the drug−drug matrix (step 2 in Figure 5) to create a drug−target matrix (step 3 in Figure 5). We collected all of the known drugs, $\{i\}_j$, for each target, j, from the chemogenomics matrix by selecting an appropriate column of the matrix, j, scanning downward through the rows, i, and noting the set of row locations, $\{i\}_j$, with matrix elements, O_{ij}^{dt}, equal to 1 (described in section 2.1). For each drug−target pair, i and j, we evaluated the maximum ROCS combo score between the drug, i, and the set of target representatives, $\{i\}_j$. We used this maximum combo score to populate the drug−target matrix element values, $O_{ij}^{\prime dt}$. Here, the prime designation is added to differentiate the drug−target matrix from the chemogenomics matrix. The off-target profile of a drug, i, is simply the vector of matrix element values, $O_{ij}^{\prime dt}$, $j = 1−N$, where N is the number of protein targets (245). Matrix element values, $O_{ij}^{\prime dt}$, close to 2 indicate a high likelihood of interaction between the drug, i, and target, j, whereas $O_{ij}^{\prime dt}$
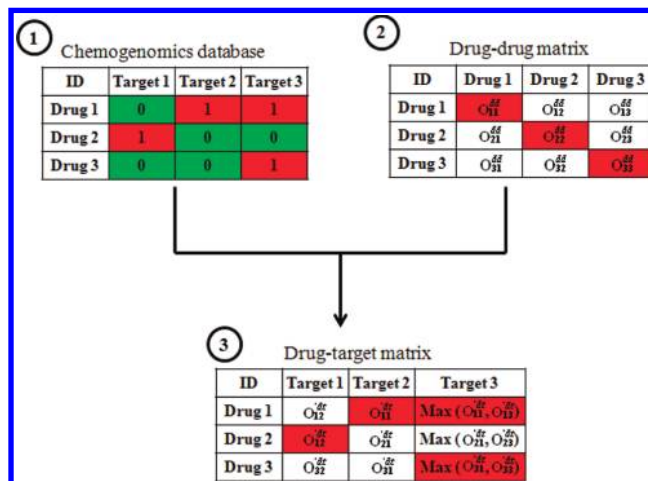

Figure 5. Schematic representation of drug−target matrix development. (1) The chemogenomics database lists all known approved drugs for the targets. The matrix elements O_{ij}^{dt} have a value of 1 (which is marked in red color) if there is a known interaction between drug, i, and target, j, and 0 (marked in green color) if there is no known interaction. (2) Drug−drug matrix elements, O_{ij}^{dd}, are generated from pairwise combo scores of each drug i with all other drugs j from the chemogenomics matrix. (3) Drug−target matrix elements, $O_{ij}^{\prime dt}$, are generated by combining the information from matrices in steps 1 and 2. For example, there is no known link between drug 1 and target 1 (see step 1). Drug 2 is the only known inhibitor of target 1. So, $O_{1,2}^{dd}$ was used as a link between drug 1 and target 1. When more than one known drug exists for a target, then the maximum combo score is taken. For example, in the case of drug 2 and target 3, which has two known drugs (drugs 1 and 3), the matrix element is given by $O_{2,3}^{\prime dt} = \text{Max}(O_{2,1}^{dd}, O_{2,3}^{dd})$.

values close to 0 indicate a small likelihood that the drug will interact with the target. A snapshot of the final version of the drug–target matrix is shown in Figure 3B. Any row of the matrix will give us the off-target profile of a drug. For the comparison of the ROCS result to the 2D similarity approach, we used the scitegic ECFP4 fingerprint in pipeline pilot.[39,40] In order to compare with the Similarity Ensemble Approach (SEA), which is a well-known target fishing application, we used the SEA search tool along with the ChEMBL database and ECFP4 descriptors as options.[41]

2.4. External Test Set. In order to check the ability of the approach to identify the off-targets of new molecules, we used an external test set. Fourteen drug molecules were identified from the literature, for each of which a new off-target has been reported recently.[16,17] We used this as an external test of molecules to further validate the use of our ROCS Based Target Fishing (RBTF) model. In order to facilitate the comparison between ligands for a particular target, we converted the combo scores to Z scores. As mentioned earlier, in our matrix, each query drug molecule is represented as a row. The Z score is calculated using the formula

$$Z_{ij} = \frac{(X_{ij} - \mu_i)}{\sigma_i}$$

where X_{ij} is the combo score for a drug i to target j, μ_i is the mean of all combo scores for that query drug across 245 targets in the row, and σ_i is the standard deviation of all combo scores for that query drug across the row.

3. RESULTS AND DISCUSSION

We generated a chemogenomics matrix of known drug–target interactions. This matrix is sparse because approved drug molecules have documented interactions with only a few of the 245 primary targets. Of the 128 potential drug–protein interactions shown in Figure 3A, there are only 7 documented activities. There are two reasons for this as follows: (1) the drugs were designed with a particular target in mind, thereby minimizing the potential for off-target activity, and (2) the drugs were never tested against the off-targets. In this work, we have used approved drug molecules as the target representatives and ROCS as the similarity method to fill in the blanks of this sparse matrix.

3.1. Validation Using the DUD Set. We first tested our idea of using approved drug molecules as target representatives using the DUD data set. The ability of chosen target representatives from Drug Bank to retrieve the DUD actives seeded into DUD decoys was studied for each target. Cross-decoy screens were also performed where the screening database was the set of DUD actives for a particular target seeded into the entire DUD ligand set, which includes the decoys for all of the other targets. The enrichment was analyzed using the AUC values from ROC plots. The results from the target-focused screen and cross-decoy screen are shown in Table 1, and the ROC plots for all 30 targets are shown in Figure 6. Our results show that the use of approved drug molecules can retrieve active molecules from decoys in most of the test case studies. If we consider AUC values greater than 0.8 as excellent, between 0.7 and 0.8 as good, between 0.6 and 0.7 as fair, between 0.5 and 0.6 as poor, and less than 0.5 as failed, then our target-focused screening strategy produced good or better enrichment for 20 of the 30 targets tested (67% success rate). Cross-decoy screening gave a 77% success rate. This is reasonable

Table 1. Screening of Target-Focused and Cross Decoys of DUD Using Approved Drugs As Target Representatives

target[a]	no. of approved drugs	no. of DUD active ligands	LBTF target-focused screen AUC	LBTF cross-screen AUC
DHFR	9	407	0.99	0.99
NA	1	49	0.97	0.99
COX-2	36	408	0.97	0.98
HMGR	8	31	0.97	0.93
EGFR	4	458	0.96	0.99
thrombin	6	68	0.95	0.83
ACHE	14	101	0.93	0.96
PNP	2	30	0.93	0.98
PR	12	26	0.92	0.95
ACE	12	48	0.91	0.94
MR	4	15	0.88	0.92
TK	6	22	0.88	0.99
COX-1	29	18	0.84	0.94
HIV-PR	6	61	0.83	0.70
ADA	4	37	0.82	0.93
AR	10	73	0.81	0.92
GR	14	77	0.80	0.89
RXRa	3	20	0.76	0.82
PDE5	7	76	0.73	0.71
PDGFR	4	169	0.71	0.83
FXa	1	146	0.66	0.20
GPB	1	52	0.65	0.88
SRC	1	159	0.63	0.68
COMT	3	10	0.62	0.73
PPARg	4	82	0.62	0.21
HIV-RT	11	40	0.56	0.72
InhA	1	86	0.50	0.49
VEGFR2	2	78	0.49	0.56
ALR2	2	26	0.49	0.60
AmpC	2	21	0.40	0.56

[a]Abbreviations: DHFR, dihydrofolate reductase; NA, neuraminidase; COX-2, cyclooxygenase-2; HMGR, hydroxymethylglutaryl-CoA reductase; EGFR, epidermal growth factor receptor; ACHE, acetylcholinesterase; PNP, purine nucleoside phosphorylase; ACE, angiotensin-converting enzyme; PR, progesterone receptor; MR, mineralocorticoid receptor; TK, thymidine kinase; COX-1, cyclooxygenase-1; ADA, adenosine deaminase; AR, androgen receptor; HIV-PR, HIV protease; GR, glucocorticoid receptor; PDE5, phosphodiesterase 5; RXRa, retinoic X receptor; PPARg, peroxisome proliferator activated receptor γ; PDGFR, platelet derived growth factor receptor kinase; SRC, tyrosine kinase SRC; COMT, catechol O-methyltransferase; HIV-RT, HIV reverse transcriptase; GPB, glycogen phosphorylase β; FXa, Factor Xa; InhA, enoyl ACP reductase; VEGFR2, vascular endothelial growth factor receptor 2; ALR2, aldose reductase 2; AmpC, AmpC β-lactamase; AUC, area under the curve.

as target-focused decoys are more challenging cases for the retrieval of actives from decoys.

Enrichment obtained from one target to the next varies considerably and is highly dependent on the selection of target representatives.[24] For progesterone receptor (PR), we obtained a target-focused AUC value of 0.92, whereas for enoyl ACP reductase (InhA), we obtained a lower AUC value of 0.5. Figure 7 shows PR target representatives (top) and a representative set of DUD actives (bottom) along with the combo score. All 12 target representatives of PR contain a cyclopenta-phenanthrene ring system. Yet, this set was able to identify diverse DUD actives with different scaffolds such as dihydro-quinoline (ZINC03832321) and chromeno[3,4-f]quinoline (ZINC03831939). The combo
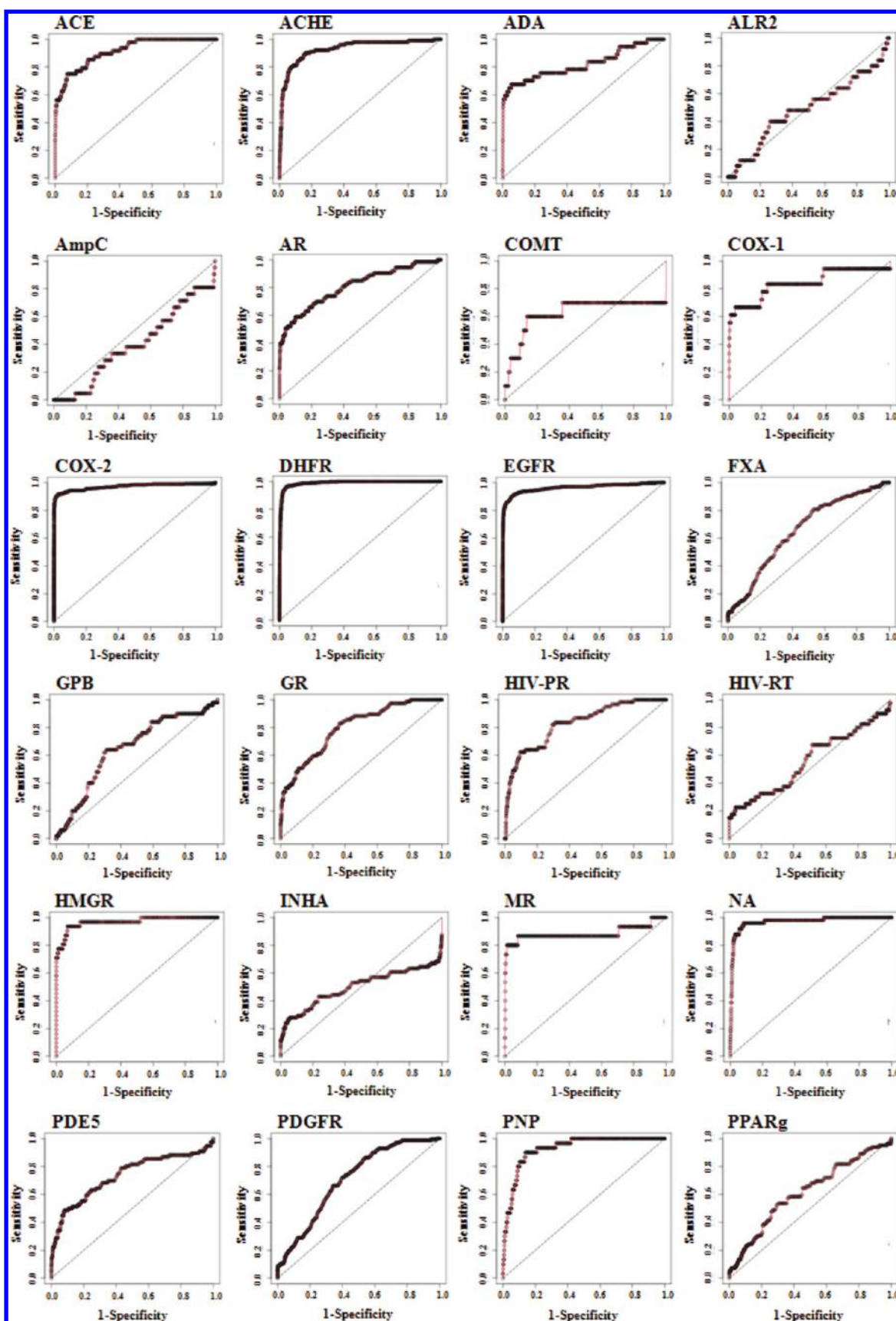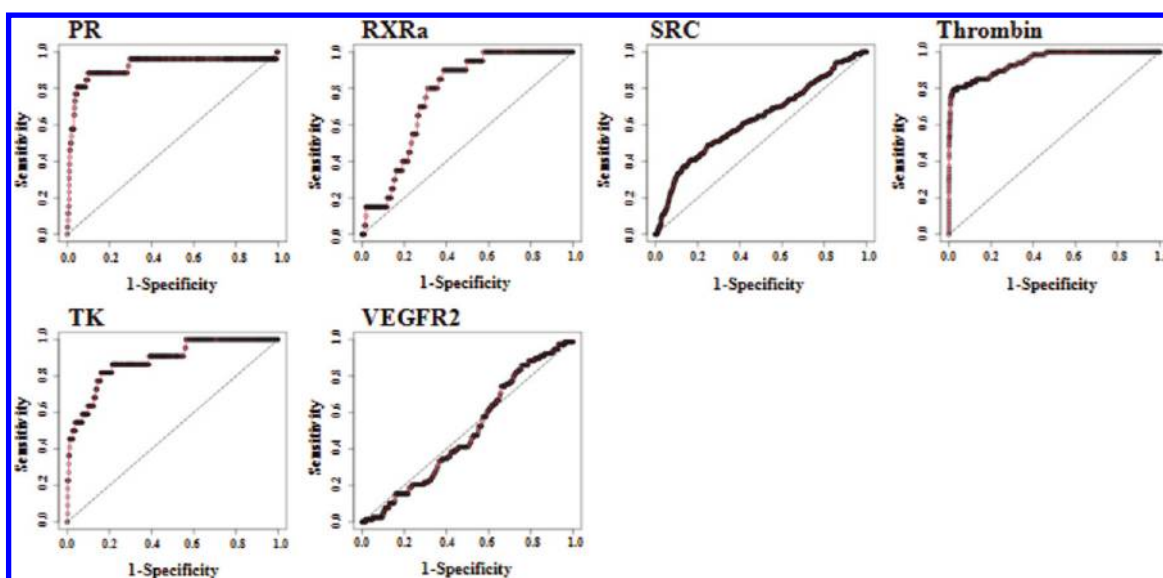
Figure 6. continued

Figure 6. ROC curves for 30 DUD targets using approved drug molecules of the respective targets as target representatives. Target abbreviations are given in Table 1. Sensitivity is the fraction of truly active compounds selected from the virtual screening workflow, and 1-specificity is the fraction of inactive compounds selected from the virtual screening workflow.
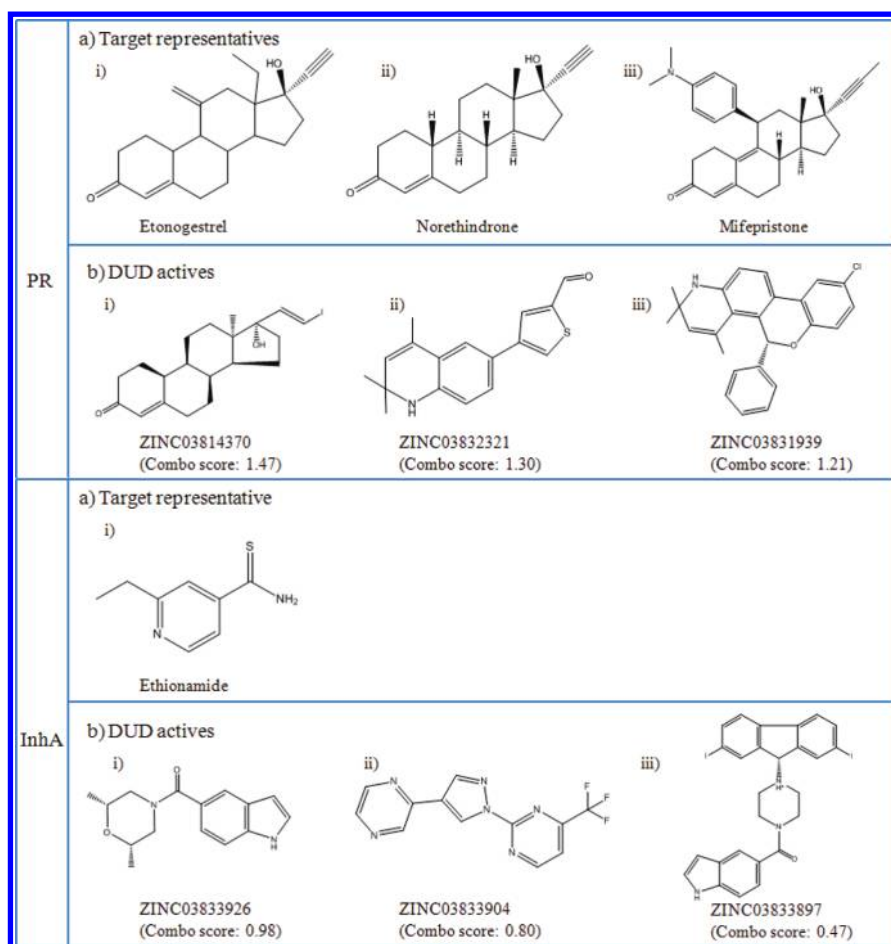

Figure 7. Structures of target representatives and representative DUD actives for PR and InhA. Abbreviations: PR, progesterone receptor; InhA, enoyl ACP reductase.

score between the target representative norethindrone and the dihydro-quinoline derivative (ZINC03832321) is 1.30, whereas the 2D similarity between these two molecules calculated using ECFP4 fingerprint gives a Tanimoto value of 0.08. This high-lights the fact that the 3D-overlap facilitates enrichment even for compounds which are not found to be similar in 2D. In the case

of InhA, the target representative, ethionamide, is a compact rigid structure. On the other hand, the DUD actives for InhA are large molecules with multiple rotatable bonds. These differences are consistent with the low AUC value of 0.5 that we obtained for InhA.

Most of the targets that produced lower enrichment in this study have a lower number of available target representatives. For example, the targets AmpC, ALR2, and VEGFR2 each have only two target representatives (Table 1) and achieved AUC values of 0.40, 0.49, and 0.49, respectively. Our conclusion from these studies is that the use of approved drugs as target representatives is reasonable with a 67% success rate of retrieving DUD actives. However, the specific examples outlined also underscore the limitations of our approach. In order for the LBTF method to be successful, there must be some similarity between the drugs used to represent the target and the active compounds that are sought. Although ROCS has been shown to be successful in scaffold hopping,[22] it is not expected to identify completely different scaffolds as exemplified in the case of InhA. Enrichment depends solely on how well the target-ligand set overlaps with the actives to be found in the database. If the target is only represented by one or two ligands, then the probability of nonoverlap with active compounds in the database may increase. Overall, this experiment validates our target fishing approach, demonstrating that it is possible to predict the activity of an unknown compound against a protein target by evaluating its similarity to drugs that have a documented protein target activity.

3.2. Generation of the Drug−Target Matrix. We have identified 245 primary drug targets which can be arranged into 13 classes. For each target, 1150 drugs were collected and classified using ATC codes. The drug target, histamine H1 receptor, was annotated with the highest number of approved drug molecules (64) in our list, followed by muscarinic M1 and dopaminergic D1 receptors, which were found to interact with 49 approved drug molecules. There are 208 targets in our list which have three or more drug molecules that are known to interact with them. The workflow for generating the final drug−target matrix is shown in Figure 5. A snapshot of a small subsection of the matrix is shown in Figure 3B, and the full drug−target matrix between 1150 drugs and 245 targets is shown in Figure 8. The red and yellow regions are the signals or alerts for potential off-target interactions in this matrix.

The value of the matrix elements of the drug−target matrix, ranging from 0 to 2, represents the likelihood of interaction between the drug and the target. The success of our DUD validation study supports this observation. In addition, the drug−target matrix is dense; i.e., every drug has a computed interaction value with every protein target. By contrast, the chemogenomics database (Figure 3A and Figure 5, part 1) derived from Drug Bank is sparse because the matrix is limited to reported interactions between drugs and proteins. As such, the drug−target matrix extends our ability to study drug−protein relationships beyond those documented in the literature or in public sources such as Drug Bank.

A quick visual analysis of the drug−target matrix provides many insights (Figure 8). For example, the anti-infective agents (marked by ATC code J) show the least off-target effects because these drugs were mainly designed to target bacterial proteins essential for survival in human hosts. Column 1

(Figure 8) is composed of pathogen targets. Most notably, the population of red matrix elements for the anti-infective agents
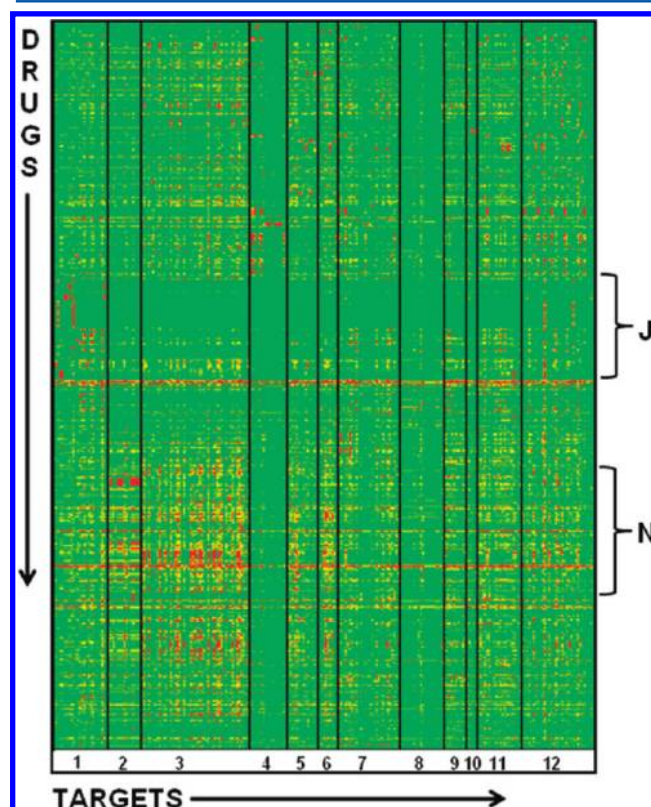

Figure 8. Drug−target matrix (1150 drugs × 245 targets) generated using the RBTF approach. Color coding: Red reflects regions with combo scores of 1.4 to 2.0 and represents potential off-target interactions. Yellow shows borderline cases of off-target interaction with combo scores of 1.2 to 1.4. Green reflects regions with combo scores below 1.2. We do not expect drug−protein interactions in green regions. J and N are ATC codes which represent anti-infectives for systemic use and drugs acting on the nervous system, respectively. Columns are labeled on the basis of different classes of targets. column 1, pathogen targets; column 2, ligand-gated ion channel receptors; column 3, G-protein coupled receptors; column 4, nuclear receptors; column 5, ion channels; column 6, transporters; column 7, oxidoreductases; column 8, kinases; column 9, other transferases; column 10, proteases; column 11, other enzymes; column 12, other targets.

in column 1 is much higher than for any other column (target class) of the matrix. In contrast to anti-infectives, the drugs acting on central nervous system targets (grouped by ATC code N) show many off-target alerts. This category of drugs includes many GPCR ligands. Our drug−target matrix agrees with a previous study demonstrating that GPCR ligands produce the most promiscuous polypharmacology-based profiles.[37]

A closer analysis of specific compounds highlights the potential of this matrix. For example, rimantadine is an antiviral compound, but it is also predicted to have interaction with N-methyl-ᴅ-aspartate (NMDA) 3A receptor. Interestingly, our preliminary analysis of the literature shows that rimantadine is an NMDA antagonist and has been reported to be of benefit to patients with Parkinson's disease.[42] We further analyzed whether this can be identified by simple 2D-similarity analysis. The chemogenomics database allows us to quickly retrieve the target representative molecule. The off-target flag was

generated on the basis of the 3D-similarity between rimantadine and memantine with a combo score of 1.54. We used ECFP4 fingerprints[39,40] for all 2D-similarity analysis. The 2D-similaritiy between these two compounds gives a Tanimoto value of 0.23. SEA is considered to be a standard 2D-based off-target prediction program.[17] In addition to 2D-similarity calculation, it gives an expectation value based on a statistical scoring scheme.[17] When analyzed with the SEA search tool, it did not give any predicted off-targets for rimantadine. This shows an example where a potential off-target of a compound could be missed if we look at the 2D-similarity alone. In addition to NMDA receptors, our RBTF predicts that rimantadine has potential off-target interactions with targets like adrenergic receptors, muscarinic receptors, serotonin transporter, and acetycholinesterase. Further study on this drug against these new off-targets will help us to understand its neuropharmacological properties.

Chlorphenesin is a centrally acting muscle relaxant with antibacterial properties. On the basis of the 3D-similaritiy with dyphylline (combo score = 1.64), RBTF predicts phospho-diesterase-4A (PDE4A) as a potential off-target for this molecule. These two compounds share a lower 2D similarity with a Tanimoto value of 0.19. The predicted off-target effect is in agreement with a previous report.[43] Celecoxib is a well-known cyclooxygenase-2 (COX-2) inhibitor. RBTF predicts a potential interaction with carbonic anhydrase (CA) based on the 3D-similarity with brinzolamide, a known CA inhibitor. The combo score between these two molecules is 1.26. Literature evidence shows the CA inhibitory activity of celecoxib.[44] The 2D-similarity between these two molecules has a Tanimoto value of 0.12.

Finally, desloratadine is an antihistaminergic compound which is predicted to interact with muscarinic (M1) receptor. Desloratadine and cyclizine share a higher 3D-similarity with a combo score of 1.49, whereas the 2D Tanimoto between these two compounds is 0.08. Desloratadine was reported to have nanomolar affinity to the M1 receptor *in vitro*.[45] Thus, most of the compounds highlighted above have lower 2D-similarity, but RBTF is able to correctly predict the off-targets based on the 3D-similarity. These examples show that new insight can be obtained from a 3D approach, and it also highlights the potential of the 3D approach to complement the 2D approaches. The structure of these molecules along with similarity score is given in the Supporting Information.

3.3. Validation Using External Test Set and Potential Applications. *3.3.1. External Test set.* Fourteen test molecules were collected from the literature for which a new off-target has been reported recently.[16,17] We used this as an external test of molecules to further validate our RBTF model. The final form of the drug–target matrix generated for an example test molecule (query) is shown in Figure 9. The combo scores were

converted into Z scores. We read across a row of the drug–target matrix, for each of the 14 test molecules, to extract the matrix values (Z scores) for each of the 245 primary targets. The targets were then sorted into decreasing order by maximum Z score value. We then collected the identities of the newly published off-targets for each test molecule and determined their positions in the sorted target lists. Off-targets that have the same score received the same ranking number, and the next target received the next immediate ranking number. The results of this calculation are shown in Table 2. If a newly published target appeared within the top 5% (top 12) of the sorted target list for each test molecule, then we deemed the RBTF protocol a success. Analysis of Table 2 shows that the RBTF protocol was able to correctly predict at least one of the newly identified off-targets for 10 of 14, or 71%, of the test molecules (molecules 1−8, 11 and 14). For molecules 9, 12, and 13, the top off-targets were ranked 64, 28, and 48, respectively. The reported off-target (5HT5A) was not present in our target list for molecule 10. These results are significant because some of the test molecules are not present in our chemogenomics database (italicized drugs in Table 2), and for those test molecules present, we were able to predict interaction with protein targets that were not previously documented (italicized proteins in column 3 of Table 2). It should be noted that other targets which occur within the top 5% of our list (not reported in the table) could be potential off-targets and candidates for future testing. Some examples are discussed below to highlight this point. The test set molecules were not known, until recent experiments,[16,17] to interact with these off-targets, and it is very gratifying to note that our RBTF approach was able to identify most of these unknown off-targets.

For example, dimetholizine (first molecule in Table 2) was recently reported to have antihistaminergic and antihypertensive action.[17] However, this molecule was not present in Drug Bank from which our chemogenomic database was derived. Our RBTF approach (outlined in Figure 9) has correctly predicted the recently identified off-targets α1A, α1B, α1D adrenergic receptors, D2 dopamine receptor, and 5HT1A serotonergic receptor (Table 2, column 5 and Figure 10A). Moreover, the histamine H1 receptor (column 3 of Table 2) was also identified as a potential target (with rank 6), which agrees with its well-known antihistaminergic activity.

Fluanisone (Sedalande), another molecule in our test set, was reported to be a neuroleptic.[17] This molecule is also not present in our chemogenomic database, and there are no known targets assigned to it. Our RBTF protocol (Figure 9) correctly predicted the recently identified off-targets α1A, α1B, α1D adrenergic receptors, and the 5HT1D serotonergic receptor (column 5 of Table 2). Dopamine D2 and 5HT2A receptors (column 3 of Table 2) are well-known targets of sedalande.[46,47] D2 is ranked first in the off-target hit list (column 4 of Table 2 and Figure 10B), and the 5-HT2A receptor is ranked third.

Fluoxetine (Prozac) is another well-known drug in our list which inhibits the serotonin transporter. This drug is present in our chemogenomic database, but its association with the recently identified off-target β1 adrenergic receptor was not. In fact, the β1 adrenergic receptor is ranked fifth in our list. A literature analysis shows that fluoxetine was known to have a weak binding affinity for the norepinephrine transporter (K_i = 1560 nM) and dopamine transporter (K_i = 6670 nM).[48] Fluoxetine was also known to have histamine H1 receptor

③	Drug–target matrix		
ID	**Target 1**	**Target 2**	**Target 3**
Drug 1	O_{12}^{dt}	O_{11}^{dt}	Max (O_{11}^{dt}, O_{21}^{dt})
Drug 2	O_{11}^{dt}	O_{21}^{dt}	Max (O_{11}^{dt}, O_{21}^{dt})
Drug 3	O_{32}^{dt}	O_{31}^{dt}	Max (O_{31}^{dt}, O_{31}^{dt})
Query 4 (Unknown)	O_{42}^{dt}	O_{41}^{dt}	Max (O_{41}^{dt}, O_{41}^{dt})

Figure 9. The drug–target matrix generated for a test molecule (shown as query 4). The workflow is similar to that shown in Figure 5.

Table 2. Prediction of Off-Targets for Test Molecules Using the RBTF Approach

no.	drug[a]	known action/target[b,c]	RBTF rank	off-targets (recently identified)	RBTF rank
1	*dimetholizine*	*antihistamine (H1 histamine receptor)*	4	D2 (K_i = 180 nM)	1
				α1A (K_i = 1.2 nM)	2
				α1B (K_i = 14 nM)	2
				α1D (K_i = 7 nM)	2
				5HT1A (K_i = 140 nM)	4
2	*fluanisone*	neuroleptic (D2)	1	α1A (K_i = 1.2 nM)	2
				α1B (K_i = 14 nM)	2
		5HT2A	3	α1D (K_i = 7 nM)	2
				5HT1D (K_i = 140 nM)	6
3	*indoramin*	*adrenergic receptor (α1A receptor)*	3	D4 (K_i = 18 nM)	3
4	paroxetine	SERT	1	β1 antagonist (K_i = 1000 nM)	3
5	methadone	μ opiod receptor	1	M3 antagonist (K_i = 1000 nM)	4
6	fluoxetine	SERT	1	β1 antagonist (K_i = 4400 nM)	5
		noradrenaline transporter	1		
		dopamine transporter	1		
		H1 histamine receptor	3		
7	domperidone	D2	1		
		hERG	1	α1B (K_i = 530 nM)	2
8	*DM tryptamine*	*serotonergic (5HT receptor)*	2	5HT1B (K_i = 130 nM)	2
				5HT7 (K_i = 210 nM)	6
				5HT2A (K_i = 130 nM)	15
9	*denopamine*	*cardiotonic (β1 receptor)*	2	β3 agonist (K_i = 2100 nM)	64
10	*mebhydrolin*	*Antihistamine (H1 Histamine receptor)*	1	5HT5A (K_i = 130 nM)	not listed
11	*ifenprodil*	*NMDAR*	15	μ opioid (K_i = 1400 nM)	12
12	*tetrabenazin*	*VMAT2*	1	α2A (K_i = 960 nM)	28
				α2C (K_i = 1300 nM)	29
13	diphemanil	M3	1	Δ opioid (K_i = 1400 nM)	48
14	*RO-25-6981*	*NMDA*	4	D4 (K_i = 120 nM)	6
				SERT (K_i = 1400 nM)	18
				noradrenaline transporter (K_i = 1300 nM)	18

[a]Test molecules which are not present in our database are italicized. [b]Known targets with no previously identified interaction with the test molecule are italicized. [c]Abbreviations: SERT, serotonin transporter; D2, dopamine receptor-2; α 1A receptor, α 1A adrenergic receptor; D4, dopamine receptor-4; M3, muscarinic receptor M3; VMAT2, vesicular monoamine transporter-2; NMDA, NMDA receptor.

antagonist activity with an IC$_{50}$ value of 1.9 μM.[49] The norepinephrine transporter and dopamine transporter were ranked first, and the histamine H1 receptor was ranked third in our off-target hit list for fluoxetine (Figure 10C). Furthermore, hERG ranks seventh in our off-target hit list (not shown in Table 2, but shown in Figure 10C), which is consistent with previous work demonstrating that fluoxetine inhibits hERG with an IC$_{50}$ of 0.7 μM.[50] Nine of the 14 drugs in the test set were not in our database, and therefore they have no target assignments. The nine compounds are as follows: fluanisone, dimetholizine, indoramin, mebhydrolin, denopamine, DMtryptamine, tetrabenazine, ifenprodil, and RO-25-6981 (Table 2). Our RBTF approach was able to assign the correct targets in eight of the nine cases. This also highlights the potential application of RBTF in assigning targets to orphan compounds. Orphan compounds are compounds with known pharmacological activity but with unknown macromolecular target.[10] Fluanisone is only known as a neuroleptic, but by using RBTF, we were able to assign potential targets or off-targets to it. Indoramin's known pharmacological action is as an adrenergic blocker and antihypertensive. Adrenergic α1A and α1B receptors were identified as potential targets with ranks three and four, respectively. In the case of mebhydrolin, the recently identified off-target 5HT5A receptor is not present in our list of targets. However, RBTF was able to identify the histamine

H1 receptor as its target (rank 1; combo score = 1.62). The off-target hit lists of other test set molecules are given in the Supporting Information.

3.3.2. hERG Toxicity. One of the important applications of developing the off-target profiles of drug molecules is to understand potential toxicity due to interactions with unwanted targets. The hERG potassium channel is a well-known target which is implicated in cardiac toxicity.[51] We explored the potential of our RBTF protocol to predict the interaction of drugs with hERG. There are 14 approved drugs in our chemogenomics database which are known to interact with hERG and which serve as target representatives. The 14 approved drugs served as target representatives, and the overlap with the query molecule is given by combo score with values between 0 and 2. We converted it into a Z score as explained in the Methods section. RBTF predicts that propranolol will interact with hERG, and a quick search of the literature shows that propranolol inhibits hERG.[52] In our chemogenomics database (Figure 3), propranolol was not associated with hERG activity, but we demonstrated via our RBTF protocol that propranolol has hERG activity. Through literature analysis, we were able to confirm the hERG interaction for at least five drugs (shown in Table 3), which produced an alert in our RBTF screen.[50,53]
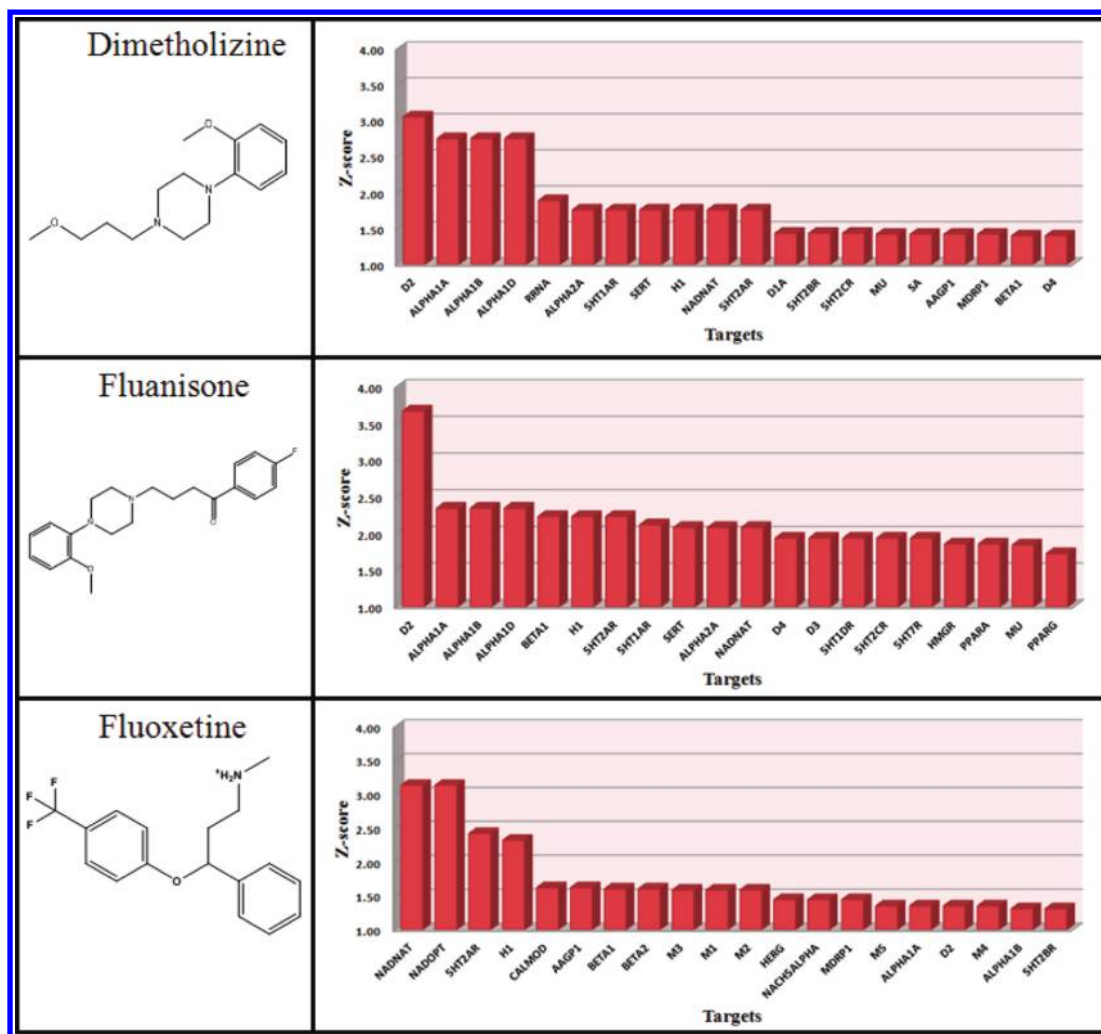
Figure 10. Off-target hits using RBTF for (A) dimetholizine, (B) Fluanisone, and (C) Fluoxetine. Abbreviations: ALPHA, α adrenergic receptor; D, dopamine receptor; BETA, β adrenergic receptor; H1, histamine H1 receptor, 5HT, serotonin receptor; SERT, serotonin transporter; NADNAT, sodium-dependent noradrenaline transporter; MU, μ opioid receptor; SA, serum albumin; AAGP1, alpha-1-acid glycoprotein 1; MDRP1, multidrug resistance protein 1; HMGR, HMG-CoA reductase; PPARA, peroxisome proliferator-activated receptor α; PPARG, peroxisome proliferator-activated receptor γ; NADOPT, sodium-dependent dopamine transporter; CALMOD, calmodulin; M, muscarinic receptor; A1, adenosine A1 receptor; HERG, hERG potassium ion channel; NACH5ALPHA, sodium channel protein type 5 subunit α.

Table 3. Analysis of the Off-Target Profile of Approved Drugs against hERG

	combo score[a]	Z score
domperidone	1.42	3.15
droperidol	1.46	2.75
fluoxetine	1.39	1.45
atomoxetine	1.38	1.11
haloperidol	1.33	1.79
propronalol	1.49	2.20

[a]Combo score ranges from 0 to 2 and represents the volume and chemical features overlap.

4. CONCLUSION

Polypharmacology-based methods can augment modern drug-discovery efforts in a range of applications, from the repurposing of existing drugs toward new protein targets, to predicting side-effect profiles for drug compounds, to designing novel drugs with lower toxicity and higher efficacy. Generation of the polypharmacology-based profile of drugs and new lead compounds is a challenging task. In this study, we developed a novel approach to address this issue.

We generated a chemogenomic database that links known target proteins and drugs. This allowed us to use approved drug molecules as target representatives. We then used a 3D-shape and chemistry-based similarity search to develop the off-target profile of drugs. We validated this approach with the DUD data set using both target-focused decoys and cross decoys. By using our RBTF protocol, we were able to identify many off-targets of drugs which were recently reported in the literature. Overall, this is a simple and fast approach that demonstrates that a shape and chemical similarity-based target fishing approach starting with a chemogenomic database can successfully generate polypharmacology-based profiles. The methodology has potential application in the prediction of toxicity, identification of targets of orphan compounds, and drug repurposing.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Three figures that show the off-target profile via RBTF for 11 test set molecules and one figure that shows the 3D/2D similarity of 4 molecules and respective target representative compound.

This material is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: mabdulhameed@bioanalysis.org; gtawa@bioanalysis.org.

■ REFERENCES

(1) Merino, A.; Bronowska, A. K.; Jackson, D. B.; Cahill, D. J. Drug profiling: knowing where it hits. *Drug Discovery Today* 15 (17−18), 749−756.

(2) Tsaioun, K.; Bottlaender, M.; Mabondzo, A. ADDME--Avoiding Drug Development Mistakes Early: central nervous system drug discovery perspective. *BMC Neurol.* **2009**, *9* (Suppl 1), S1.

(3) Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discovery* **2004**, *3* (8), 711−5.

(4) Schuster, D.; Laggner, C.; Langer, T. Why drugs fail--a study on side effects in new chemical entities. *Curr. Pharm. Des.* **2005**, *11* (27), 3545−59.

(5) Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4* (11), 682−90.

(6) Ashburn, T. T.; Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discovery* **2004**, *3* (8), 673−83.

(7) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432* (7019), 862−5.

(8) Rognan, D. Structure-Based Approaches to Target Fishing and Ligand Profiling. *Mol. Inf.* **2010**, *29* (3), 176−187.

(9) Loging, W.; Harland, L.; Williams-Jones, B. High-throughput electronic biology: mining information for drug discovery. *Nat. Rev. Drug Discovery* **2007**, *6* (3), 220−30.

(10) Jenkins, J. L. B. A.; Davies, J. W. In silico target fishnig: predicting biological targets from chemical structure. *Drug Discovery Today: Technol.* **2006**, *3* (4), 413−421.

(11) Chen, Y. Z.; Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **2001**, *43* (2), 217−26.

(12) 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* (Web Server issue), W219−24.

(13) Li, L.; Bum-Erdene, K.; Baenziger, P. H.; Rosen, J. J.; Hemmert, J. R.; Nellis, J. A.; Pierce, M. E.; Meroueh, S. O., BioDrugScreen: a computational drug design resource for ranking molecules docked to the human proteome. *Nucleic Acids Res.* 38 (Database issue), D765−73.

(14) Yang, L.; Luo, H.; Chen, J.; Xing, Q.; He, L. SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a chemical-protein interactome. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W406−12.

(15) Rai, B. K.; Tawa, G. J.; Katz, A. H.; Humblet, C., Modeling G protein-coupled receptors for structure-based drug discovery using low-frequency normal modes for refinement of homology models: application to H3 antagonists. *Proteins* 78 (2), 457−73.

(16) 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* (2), 197−206.

(17) 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* (7270), 175−81.

(18) 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* (23), 6802−10.

(19) Nidhi.; Glick, M.; Davies, J. W.; Jenkins, J. L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **2006**, *46* (3), 1124−33.

(20) Mestres, J.; Martin-Couce, L.; Gregori-Puigjane, E.; Cases, M.; Boyer, S. Ligand-based approach to in silico pharmacology: nuclear receptor profiling. *J. Chem. Inf. Model.* **2006**, *46* (6), 2725−36.

(21) Cleves, A. E.; Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **2006**, *49* (10), 2921−38.

(22) Rush, T. S. 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48* (5), 1489−95.

(23) Kirchmair, J.; Distinto, S.; Schuster, D.; Spitzer, G.; Langer, T.; Wolber, G. Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates. *Curr. Med. Chem.* **2008**, *15* (20), 2040−53.

(24) Tawa, G. J.; Baber, J. C.; Humblet, C. Computation of 3D queries for ROCS based virtual screens. *J. Comput.-Aided Mol. Des.* **2009**, *23* (12), 853−68.

(25) Gundersen, E.; Fan, K.; Haas, K.; Huryn, D.; Steven Jacobsen, J.; Kreft, A.; Martone, R.; Mayer, S.; Sonnenberg-Reines, J.; Sun, S. C.; Zhou, H. Molecular-modeling based design, synthesis, and activity of substituted piperidines as gamma-secretase inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15* (7), 1891−4.

(26) Vijayan, R. S.; Prabu, M.; Mascarenhas, N. M.; Ghoshal, N. Hybrid Structure-Based Virtual Screening Protocol for the Identification of Novel BACE1 Inhibitors. *J. Chem. Inf. Model.* **2009** .

(27) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36* (Database issue), D901−6.

(28) Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **2010**, *38* (Database issue), D355−60.

(29) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789−801.

(30) Imming, P.; Sinning, C.; Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discovery* **2006**, *5* (10), 821−34.

(31) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5* (12), 993−6.

(32) *Filter*, 2.1.0; OpenEye Scientific Software: Santa Fe, NM, 2010.

(33) WHOCC - ATC/DDD Index. http://www.whocc.no/atc_ddd_index/ (accessed Dec. 22, 2010).

(34) *OMEGA*, 2.4.1; OpenEye Scientific Software: Santa Fe, NM, 2010.

(35) *ROCS*, 3.0.0; OpenEye Scientific Software: Santa Fe, NM, 2009.

(36) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11* (23−24), 1046−53.

(37) Gregori-Puigjane, E.; Mestres, J. A ligand-based approach to mining the chemogenomic space of drugs. *Comb. Chem. High Throughput Screen.* **2008**, *11* (8), 669−76.

(38) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1177−85.

(39) Hassan, M.; Brown, R. D.; Varma-O'brien, S.; Rogers, D. Cheminformatics analysis and learning in a data pipelining environment. *Mol. Divers.* **2006**, *10* (3), 283−99.

(40) Morgan, H. L. The generation of a unique machine description for chemical structures-A technique developed at chemical sbstracts service. *J. Chem. Doc.* **1965**, *5*, 107−112.

(41) SEArch. http://sea.bkslab.org/search/ (accessed Nov. 13, 2011).

(42) Singer, C.; Papapetropoulos, S.; Gonzalez, M. A.; Roberts, E. L.; Lieberman, A. Rimantadine in Parkinson's disease patients experiencing peripheral adverse effects from amantadine: report of a case series. *Mov. Disord.* **2005**, *20* (7), 873−7.

(43) Edelson, J.; McMullen, J. P. Interactions of chlorphenesin and divalent metal ions with phosphodiesterase. *Arch. Int. Pharmacodyn. Ther.* **1976**, *223* (1), 24−33.

(44) Weber, A.; Casini, A.; Heine, A.; Kuhn, D.; Supuran, C. T.; Scozzafava, A.; Klebe, G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J. Med. Chem.* **2004**, *47* (3), 550−7.

(45) Cardelus, I.; Anton, F.; Beleta, J.; Palacios, J. M. Anticholinergic effects of desloratadine, the major metabolite of loratadine, in rabbit and guinea-pig iris smooth muscle. *Eur. J. Pharmacol.* **1999**, *374* (2), 249−54.

(46) van Wijngaarden, I.; Kruse, C. G.; van Hes, R.; van der Heyden, J. A.; Tulp, M. T. 2-Phenylpyrroles as conformationally restricted benzamide analogues. A new class of potential antipsychotics. 1. *J. Med. Chem.* **1987**, *30* (11), 2099−104.

(47) van Luijtelaar, E. L.; Drinkenburg, W. H.; van Rijn, C. M.; Coenen, A. M. Rat models of genetic absence epilepsy: what do EEG spike-wave discharges tell us about drug effects? *Methods Find Exp. Clin. Pharmacol.* **2002**, *24* (SupplD), 65−70.

(48) Cashman, J. R.; Voelker, T.; Zhang, H. T.; O'Donnell, J. M. Dual inhibitors of phosphodiesterase-4 and serotonin reuptake. *J. Med. Chem.* **2009**, *52* (6), 1530−9.

(49) Wong, D. T.; Bymaster, F. P.; Reid, L. R.; Threlkeld, P. G. Fluoxetine and two other serotonin uptake inhibitors without affinity for neuronal receptors. *Biochem. Pharmacol.* **1983**, *32* (7), 1287−93.

(50) Staudacher, I.; Schweizer, P. A.; Katus, H. A.; Thomas, D. hERG: protein trafficking and potential for therapy and drug side effects. *Curr. Opin. Drug Discovery Dev.* 13 (1), 23−30.

(51) Sanguinetti, M. C.; Mitcheson, J. S. Predicting drug-hERG channel interactions that cause acquired long QT syndrome. *Trends Pharmacol. Sci.* **2005**, *26* (3), 119−24.

(52) Yao, X.; McIntyre, M. S.; Lang, D. G.; Song, I. H.; Becherer, J. D.; Hashim, M. A. Propranolol inhibits the human ether-a-go-go-related gene potassium channels. *Eur. J. Pharmacol.* **2005**, *519* (3), 208−11.

(53) Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers. *J. Med. Chem.* **2002**, *45* (18), 3844−53.