

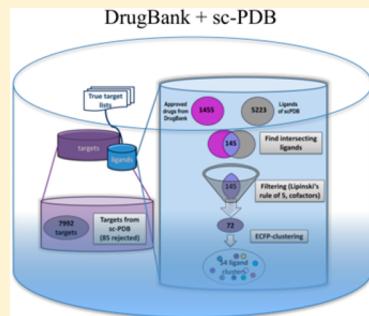
Benchmark Data Sets for Structure-Based Computational Target Prediction

Karen T. Schomburg and Matthias Rarey*

Center for Bioinformatics, University of Hamburg, Bundesstrasse 43, 20146 Hamburg, Germany

Supporting Information

ABSTRACT: Structure-based computational target prediction methods identify potential targets for a bioactive compound. Methods based on protein–ligand docking so far face many challenges, where the greatest probably is the ranking of true targets in a large data set of protein structures. Currently, no standard data sets for evaluation exist, rendering comparison and demonstration of improvements of methods cumbersome. Therefore, we propose two data sets and evaluation strategies for a meaningful evaluation of new target prediction methods, i.e., a small data set consisting of three target classes for detailed proof-of-concept and selectivity studies and a large data set consisting of 7992 protein structures and 72 drug-like ligands allowing statistical evaluation with performance metrics on a drug-like chemical space. Both data sets are built from openly available resources, and any information needed to perform the described experiments is reported. We describe the composition of the data sets, the setup of screening experiments, and the evaluation strategy. Performance metrics capable to measure the early recognition of enrichments like AUC, BEDROC, and NSLR are proposed. We apply a sequence-based target prediction method to the large data set to analyze its content of nontrivial evaluation cases. The proposed data sets are used for method evaluation of our new inverse screening method *iRAISE*. The small data set reveals the method’s capability and limitations to selectively distinguish between rather similar protein structures. The large data set simulates real target identification scenarios. *iRAISE* achieves in 55% excellent or good enrichment a median AUC of 0.67 and RMSDs below 2.0 Å for 74% and was able to predict the first true target in 59 out of 72 cases in the top 2% of the protein data set of about 8000 structures.



INTRODUCTION

Phenotypic effects of drugs are the result of interactions of the drug with protein targets, i.e., primary targets for which they were designed for as well as off-targets. In order to design selective drugs rationally, the knowledge about all possible targets of a lead is essential. Computational methods can support this task by predicting targets for a small molecule.^{1–5} Structure-based computational methods predict not only targets but also binding modes that support the rational design of drugs without critical adverse effects. While the use of virtual screening methods in drug design is well established today,⁶ computational target prediction methods still have to convince this research field.⁷ Evaluation of new target prediction methods on real world use cases is necessary for the demonstration of their usability in the drug design workflow.

Because computational methods make predictions based on models with several approximations, evaluation experiments need to be able to reveal limitations of the overall approach. Evaluations are either prospective, testing predicted effects with experiments, or retrospective, measuring to which extent a method can reproduce already available data. While prospective evaluation is considered as more trustworthy, retrospective evaluation has several advantages. First, prospective experimental studies are expensive, time consuming, and not available to most method developers. Second, if retrospective studies are established in a field of research, newly developed methods can

easily be compared to the state of the art. Whatever evaluation method is chosen, it should be designed in a way to represent real use cases⁸ to be of use for application scientists.

Lately, it was discussed how useful and systematic current evaluation experiments and performance metrics are (special issue of the Journal of Computer-Aided Molecular Design⁹). Simple metrics reported often, like EFs (enrichment factors), are not comparable on different data sets because they are dependent on the ratio of actives in a data set as well as on the cutoff factor. The AUC (area under the ROC curve) is independent of a cutoff factor, but it is not suited for assessing early enrichment. In docking/virtual screening as well as in inverse screening, usually the methods do not classify molecules or targets clearly as true positives or negatives; they rank all data points in a list. Therefore, the evaluation experiments and performance metrics have to show how well the method is capable of so-called “early recognition”, i.e., ranking true targets to the very front of the list. For assessing this capability in a metric, the BEDROC (Boltzman-enhanced discrimination of ROC) metric¹⁰ and the NSLR¹¹ (normalized sum of logarithmic ranks) were recently proposed.

Inverse virtual screening is a decade younger than regular virtual screening and only few approaches exist.^{12–14} For classic

Received: February 28, 2014

Published: August 1, 2014

Table 1. Overview of Number of Protein Complexes, Number of Query Ligands, and Composition of Two Data Sets Recommended for Evaluation of Structure-Based Inverse Screening Methods

data set	number of protein complexes	number of query ligands	composition/purpose
selectivity	three different serine protease proteins with three structures each, two phosphodiesterase isoenzymes with three structures each, and three histone deacetylase isoenzymes represented by eight structures	17	small data set with defined true negative annotation for proof-of-concept studies
Drugs/sc-PDB	7992 structures of 3127 proteins (mapping by UniProtID)	72	large data set for statistical performance evaluation on drug-like ligands

protein–ligand docking, several retrospective experiments are used for evaluation per default, e.g., redocking studies with the Astex diverse set¹⁵ and enrichment experiments on the DUD data set.¹⁶ While for ligand-based structure prediction methods a standard data set has already been presented,¹⁷ but for structure-based inverse screening approaches, so far no standard evaluation strategy has been established.

The inverse docking method Invdock developed by Chen et al.^{12,13} is evaluated by docking eight compounds against 1040 cavities of 38 proteins and comparing the results to activities reported in literature. TarFisDock developed by Li et al.¹⁴ is evaluated by screening two compounds against the TTD (therapeutic target database) and evaluation of how many true targets are ranked in the top two and top five entries. The sc-PDB was screened with four diverse compounds by Paul et al.¹⁸ who report the enrichment factor at 5% for evaluation. Following, Kellenberger et al.¹⁹ used the same four compounds for evaluation of four different ranking strategies and report the ranks of true targets along with the AUC. Meslamani et al.²⁰ compare the capability of ranking true targets of ligand similarity-based methods and protein–ligand structure-based methods on 157 diverse ligands and the sc-PDB²¹ protein structures. For evaluation, the rank of the first found true target is reported and evaluated. However, this data set lacks a reliable annotation of negatives, which prevents true enrichment experiments on this data set.

In summary, the evaluation of the so far published inverse virtual screening approaches was limited to few individual test cases by using a limited number of control samples rather than using quantitative analysis, and even if a larger data set was used, no thorough statistical evaluation was possible. Therefore, for the evaluation of structure-based inverse screening, a data set is needed that allows quantitative and systematic analysis of the computational model representing a use case similar to real application scenarios.

Consequently, we propose two data sets for the evaluation of structure-based inverse virtual screening methods: a small one consisting of three protein classes and inhibitors with defined selectivity for detailed proof of concept studies and a large one consisting of 72 ligands and 7992 protein structures for large scale statistical evaluation. Further, we propose metrics for the evaluation of the large data set. In the following, we will introduce the data sets, discuss the construction, propose evaluation metrics, and discuss the results of the recently published inverse screening tool iRAISE²² on these data sets. Furthermore, we compare iRAISE's performance to a sequence-based target prediction method on the large data set.

■ DATA PREPARATION

The requirements of an evaluation data set for structure-based target prediction concern the ligands, the targets, and the annotation of true positives and true negatives. The ligands

should be diverse and ideally cover the chemical space of potential application studies, e.g., being drug-like. The number of ligands and 3D structures of targets should be high enough to allow statistical evaluation and to demonstrate the methods capability to handle large amounts of targets. The correct annotation of negatives needs to be assured, which is not trivial because often only true targets are reported for ligands in the literature. The positive/negative ratio of targets for the ligands shall enable enrichment experiments, i.e., a few actives need to be complemented with many inactives.

Because it is hardly possible to meet all requirements in one data set, two sets will be presented here: a small data set based on well-known protein–ligand interactions allowing the analysis of prediction selectivity and a large one allowing statistical enrichment studies.

In Table 1, the composition of the two data sets is summarized. The small selectivity data set consists of three target classes with two to three proteins and ligands with well-annotated selectivity in the respective target class. The well-studied selectivity of the ligands allows a clear separation of true negatives and true positives, and the similarity of the proteins of one target class poses a nontrivial prediction problem. The Drugs/sc-PDB data set allows statistical evaluation and enrichment studies consisting of 72 ligands that are chosen from approved drugs and nearly 8000 protein structures.

Selectivity Data Set. *Selectivity Data Set Composition.* For proof-of-concept studies, a small data set is composed consisting of three target classes of which the proteins are highly similar and for which inhibitors with well-known selectivity exist.

The first target class are serine proteases from the subclass of trypsin-like serine proteases: trypsin, thrombin, and factor Xa (TTFXa). Serine proteases were chosen for three reasons. First, the well-studied inhibitors for the serine proteases trypsin, thrombin, and factor Xa allow a clear definition of true positives and true negative protein targets. Second, the overall sequence and structure of serine proteases is highly conserved, as well as the catalytic triad of aspartate, histidine, and serine in the active site.^{23–25} Therefore, sequence-based target prediction methods usually fail on this target class.²⁶ Third, the selective inhibition of thrombin and factor Xa proteins, relevant in the blood coagulation cascade, is of strong pharmacological interest.

For the other two chosen target classes, also structurally highly similar and well-studied examples were chosen. The second target class consists of phosphodiesterases (PDE). For this target class, isoenzyme-specific inhibitors are known. There are PDE-selective inhibitors that are either only active at PDE4 or at PDE5.^{27,28} As a third target class, histone deacetylases (HDAC) were chosen because here also subclasses-selective inhibitors exist. Selective HDAC inhibitors are either active on subclass I or on subclass II. Of these subclasses, HDAC2 and HDAC8 were selected for subclass I and HDAC4 as representative for subclass

II because most assays as well as crystal structures are available for these representatives.²⁹

Protein Structures of the Selectivity Data Set. For each of the proteins of the selectivity data set, three structures were collected from the Protein Data Bank (PDB)³⁰ to account for protein conformation changes applying the following criteria. Each structure had to be co-crystallized with a different ligand to account for different ligand-induced protein conformations. This ligand was not allowed to be one of the inhibitors with which the structures shall be screened in the experiment (see below for the inhibitors). Then the structures with the best resolution were chosen (see Table 2 for the list of all structures). For the HDAC

Table 2. List of PDB Structures Contained in the Selectivity Data Set

PDB codes of protein structures (UniProt ID)	
phosphodiesterases	
PDE4	2QYN (Q04899), 2QYK (Q04899), 3TVX (P27815)
PDES	3HC8 (O76074), 3SHY (O76074), 3TGE (O76074)
histone deacetylases	
HDAC2 (class I)	3MAX (Q92769), 4LY1 (Q92769)
HDAC8 (class I)	1T67 (Q9BY41), 1W22 (Q9BY41), 3SFF (Q9BY41)
HDAC4 (class II)	2VGJ (P56524), 2VGM (P56524), 4CBY (P56524)
serine proteases	
trypsin	2G5N (P00760), 2G8T (P00760), 3GY2 (P00760)
thrombin	2BVR (P00734), 3RM2 (P00734), 3SI4 (P00734)
factor Xa	2JKH (P00742), 2YSF (P00742), 3KL6 (P00742)

target class, the number of available structures was limited; thus, for HDAC2, only two structures were found. The numbers of structures is on purpose limited to three per protein in order to allow inverse screening methods to demonstrate their ability of handling protein flexibility. Methods that do not depend on flexibility sampled by protein conformation ensembles can show that they are able to place each inhibitor into each of the three protein structures independent of the starting protein conformations. Methods dependent on protein conformation ensembles have to show how they cope with the limited number of conformations, a situation often occurring in real application cases.

Ligands of the Selectivity Data Set. As query ligands for the TTFXa target class, five inhibitors extracted from DrugBank³¹ and BRENDA³² were chosen (DrugBankID or PubChemID in brackets): benzamidine (DB03127) and pefabloc (DB07347) as general serine protease inhibitors, apixaban (DB06605) and rivaroxaban (DB06228) as factor Xa inhibitors, and melagatran (ID of prodrug ximelagatran: DB04898) as thrombin inhibitor. For the PDE target class, the general inhibitors caffeine (DB00201), paraxanthine (CID 4687), and theophylline (DB00277), the PDE4 inhibitors piclamilast (DB01791), drotaverine (DB06751), and roflumilast (DB01656), and the PDES inhibitors avanafil (DB06237), sildenafil (DB00203), and tadalafil (DB00820) were extracted from DrugBank³¹ and PubChem.³³ As HDAC class I and class II inhibitors, vorinostat (DB02546) and trichostatin A (CID 444732) were extracted from PubChem, and as selective HDAC class I inhibitor, mocetinostat (CID 9865515) was extracted.

The 3D structures of all inhibitors were collected from PubChem.³³

Validation Experiments with the Selectivity Data Set. An inverse screening method has to show that it predicts the correct targets for the selective inhibitors of three target classes. In particular, it shall be used to demonstrate that a method is not only able to find true positives, but that it also correctly identifies true negatives. This evaluation is possible for the selectivity data set because the activities of the inhibitors are clearly defined, which is not the case in large data sets where for many protein–ligand combinations it is not experimentally validated if they interact or not.

Large Data Set: Drugs/sc-PDB. Drugs/sc-PDB Data Set

Composition. The large scale evaluation data set represents real use cases and allows statistical analysis. It was designed under the following criteria. First, a large number of high quality diverse protein structures should be contained. Further, the known related ligands should cover drug-like chemical space, and an automatic definition of true positives and true negatives should be possible. To achieve the latter, approved drugs are used as ligands, exploiting the fact that drugs are well studied with respect to selectivity. Thus, it is more likely that interacting proteins are in fact known. For protein target structures, we chose sc-PDB, a subset of protein–ligand complexes from the Protein Data Bank relevant for drug design fulfilling certain quality criteria.²¹ This data set with more than 8000 structures is large enough for statistical evaluation. Thus, the large data set is a combination of approved drugs and sc-PDB and will be called in the following Drugs/sc-PDB data set. In Figure 1, the scheme of the data set creation process is shown.

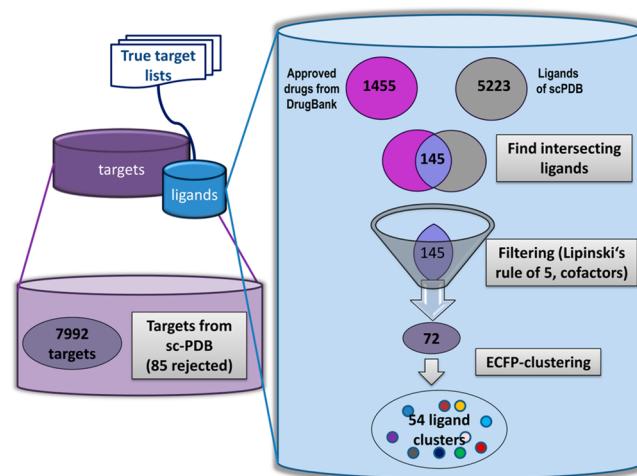


Figure 1. Scheme of the composition and creation of the Drugs/sc-PDB data set.

Protein Structures of the Drugs/sc-PDB Data Set. The 2012 version of sc-PDB³⁴ contains 8077 protein–ligand complexes. We downloaded all PDB files directly from the Protein Data Bank. Of these, 85 are discarded due to several errors during initialization of the reference ligand or the protein (51 structures), due to an erroneous reference ligand (25 structures), or due to obsolete PDB codes (9). Please refer to the Supporting Information for a list of the PDB codes of the discarded as well as the remaining structures.

Ligands of the Drugs/sc-PDB Data Set. The ligands for the data set were selected applying the following criteria:

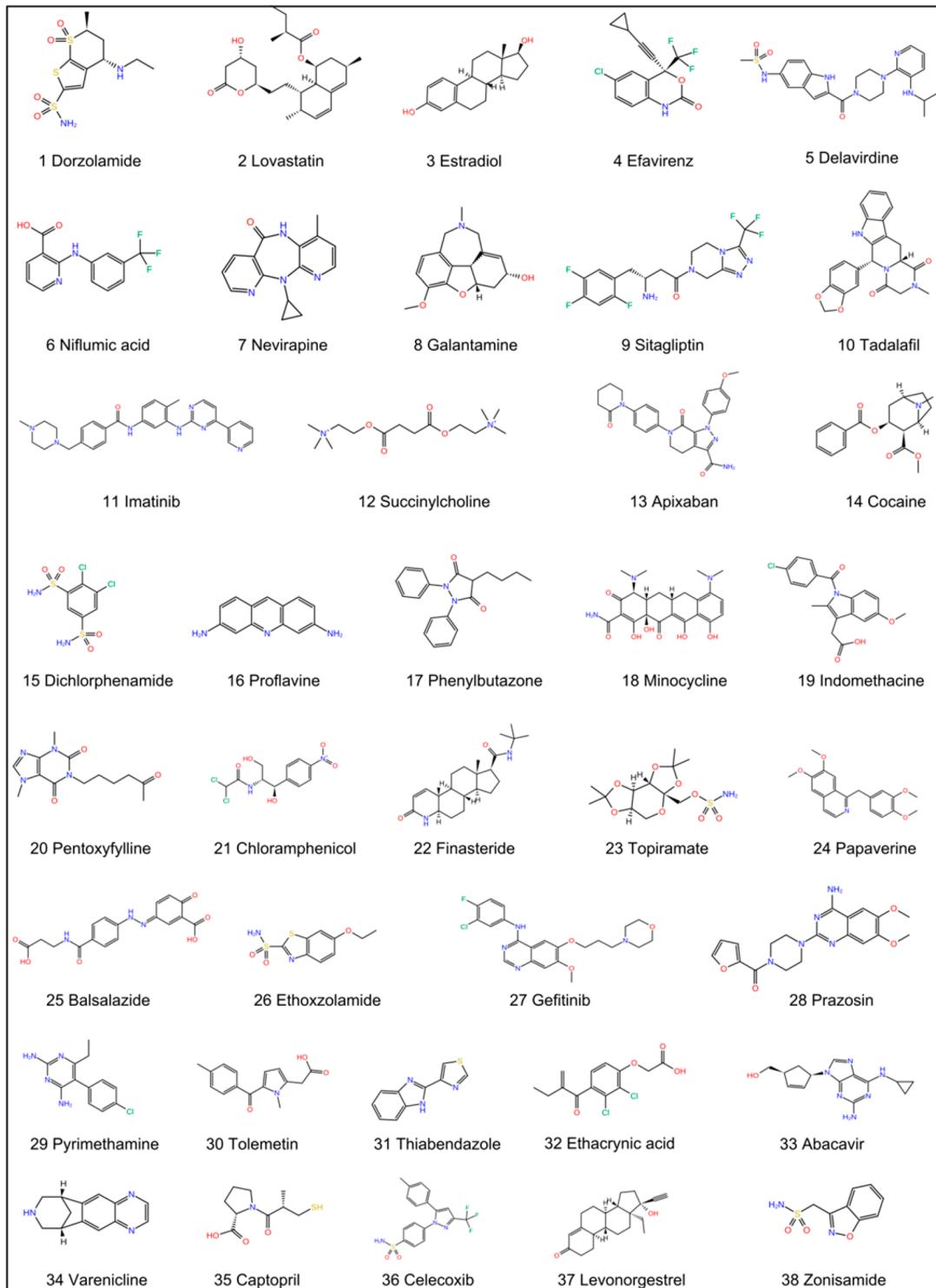


Figure 2. Structure diagrams of drugs 1–38 of the data set Drugs/sc-PDB that are singletons after the ECFP clustering.

- The set of unique ligands of all approved drugs downloaded from DrugBank³¹ and the set of unique ligands from the sc-PDB were created using MONA,³⁵ resulting in 1455 and 5223 ligands, respectively.
- Of the two unique ligand sets, the intersecting ligand set was built with MONA. Ligands were also considered equal if differing in tautomeric or protonation state because these states are not definitely assignable from the PDB files.³⁶ The intersecting set contains 145 ligands.

- The 145 ligands were filtered with Lipinski's rule of 5, and co-factors were excluded. Because co-factors bind to separate pockets that are not used in the sc-PDB data set, the annotation of true targets is hardly possible. This filtering leaves a set of 78 ligands. Furthermore, six ligands were removed because the list of targets in the DrugBank contained only DNA or RNA and no protein. Thus, the final set consists of 72 ligands.

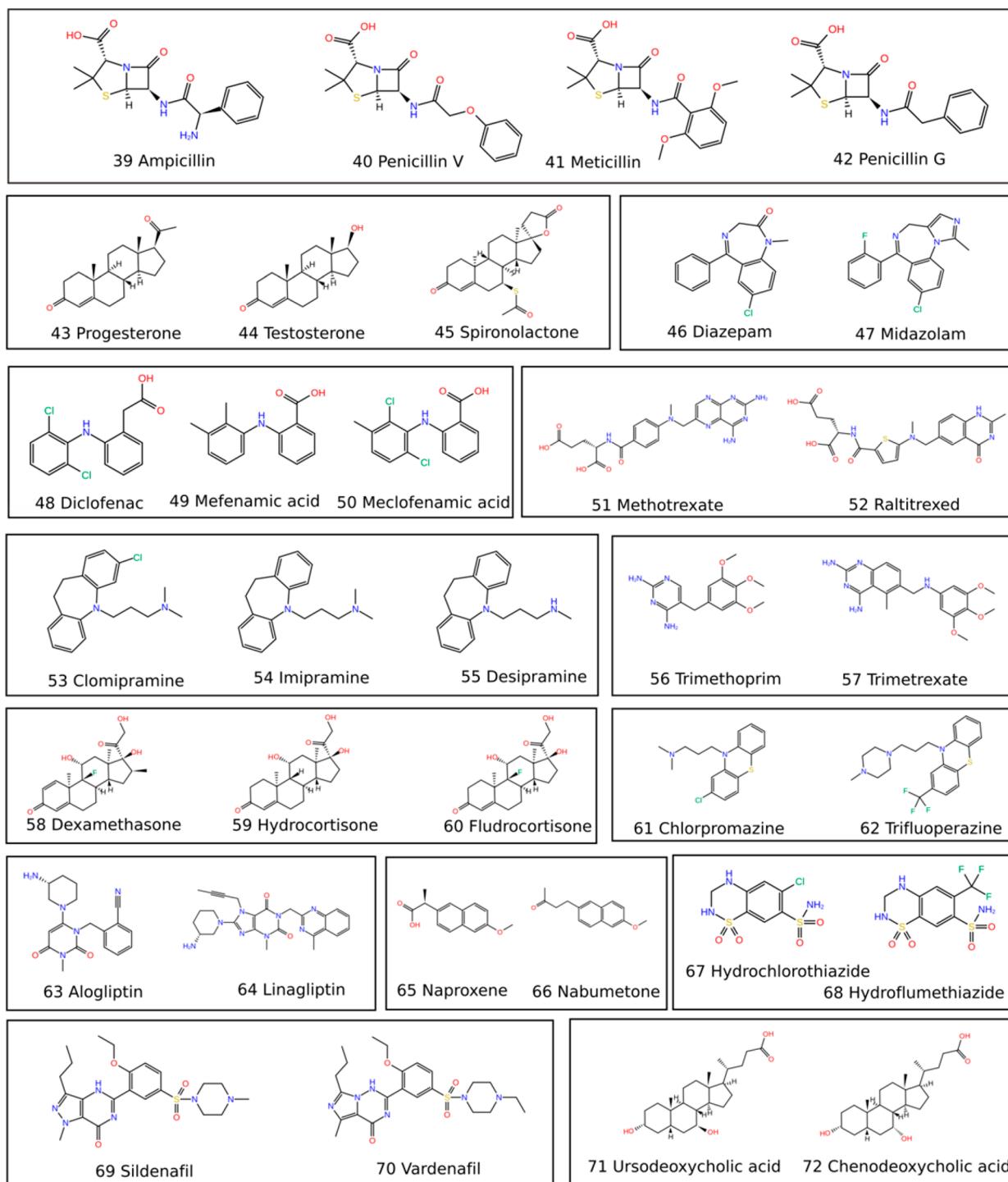


Figure 3. Structure diagrams of drugs 39–72 of the Drugs/sc-PDB data set. Each cluster of compounds is surrounded by a frame.

Finally, as some of the 72 ligands are structurally highly similar, varying only in substituents, the ligands were clustered with ECFP circular fingerprints (with three iterations)³⁷ using the Tanimoto coefficient with a similarity threshold of 0.2. In Figure 2, the 38 ligands are shown that are only present in a single cluster. In Figure 3, the clusters of ligands with two or more members are shown.

True Positives of the Drugs/sc-PDB Data set. For each drug, a list of true positive structures of the sc-PDB is assigned automatically.

A structure is added to the list of true positives if

- ...the ligand is co-crystallized with it. Because the data set was composed by the intersection of sc-PDB ligands and drugs, each ligand has at least one structure with which it is co-crystallized in the data set.
- ...the structure has the same EC number as a target listed in DrugBank for the drug.
- ...the name of the structure as contained in the “MOLECULE” field in the PDB file is the same as a name of a target listed in the DrugBank for the drug.

Table 3. List of Number of Targets Listed in the DrugBank and Number of True Positive Protein Structures Assigned to Each Ligand of the Drugs/sc-PDB Data Set^a

Nr	Drug name	DrugBank ID	No of targets listed in DrugBank	No of positive structures in the sc-PDB	Nr	Drug name	DrugBank ID	No of targets listed in DrugBank	No of positive structures in the sc-PDB
1	Dorzolamide	DB00869	3	152	37	Levonorgestrel	DB00367	4	107
2	Lovastatin	DB00227	3	29	38	Zonisamide	DB00909	5	161
3	Estradiol	DB00783	4	80	39	Ampicillin	DB00415	2	17
4	Efavirenz	DB00625	1	114	40	Penicillin V	DB00417	3	36
5	Delavirdine	DB00705	1	114	41	Meticillin	DB01603	1	3
6	Niflumic acid	DB04552	4	53	42	Penicillin G	DB01053	1	3
7	Nevirapine	DB00238	1	114	43	Progesterone	DB00396	3	90
8	Galantamine	DB00674	3	37	44	Testosterone	DB00624	3	54
9	Sitagliptin	DB01261	1	48	45	Spironolactone	DB00421	2	31
10	Tadalafil	DB00820	2	91	46	Diazepam	DB00829	3	13
11	Imatinib	DB00619	10	188	47	Midazolam	DB00683	2	2
12	Succinylcholine	DB00202	2	3	48	Diclofenac	DB00586	6	71
13	Apixaban	DB06605	1	87	49	Mefenamic acid	DB00784	2	23
14	Cocaine	DB00907	4	2	50	Meclofenamic acid	DB00939	2	22
15	Dichlorphenamide	DB01144	4	152	51	Methotrexate	DB00563	8	165
16	Proflavine	DB01123	3	141	52	Raltitrexed	DB00293	2	51
17	Phenylbutazone	DB00812	3	22	53	Clomipramine	DB01242	7	23
18	Minocycline	DB01017	7	7	54	Imipramine	DB00458	7	31
19	Indomethacine	DB00328	8	153	55	Desipramine	DB01151	8	31
20	Pentoxifylline	DB00806	3	98	56	Trimethoprim	DB00440	2	160
21	Chlormaphenicol	DB00446	3	51	57	Trimetrexate	DB01157	1	125
22	Finasteride	DB01216	2	9	58	Dexamethasome	DB01234	5	13
23	Topiramate	DB00273	3	152	59	Hydrocortisone	DB00741	2	8
24	Papaverine	DB01113	1	91	60	Fludrocortisone	DB00687	3	38
25	Balsalazide	DB01014	4	97	61	Chlorpromazine	DB00477	6	6
26	Ethoxzolamide	DB00311	1	152	62	Trifluoperazine	DB00831	8	42
27	Gefitinib	DB00317	1	140	63	Alogliptin	DB06203	1	48
28	Prazosin	DB00457	1	5	64	Linagliptin	DB08882	1	48
29	Pyrimethamine	DB00205	1	125	65	Naproxene	DB00788	2	25
30	Tolmetin	DB00500	3	25	66	Nabumetone	DB00461	2	24
31	Thiabendazole	DB00730	1	1	67	Hydrochlorothiazide	DB00999	1	153
32	Ethacrynic acid	DB00903	2	15	68	Hydroflumethiazide	DB00774	1	153
33	Abacavir	DB01048	3	137	69	Sildenafil	DB00203	1	38
34	Varenicline	DB01273	1	1	70	Vardenafil	DB00862	1	39
35	Captopril	DB01197	3	31	71	Ursodeoxycholic acid	DB01586	1	22
36	Celecoxib	DB00482	3	65	72	Chenodeoxycholic acid	DB06777	1	19

^aLigands grouped by purple frames are assigned to the same cluster by the ECFP clustering.

- ...the structure has the same UniProtKB ID as a target of the ligand listed in the DrugBank. UniProtKB IDs were assigned to each PDB code by using SIFTS.³⁸

The list of true positive PDB codes assigned by this list can be found in the Supporting Information, as well as the list of target names extracted for each drug from DrugBank. In Table 3, the number of different protein targets listed in DrugBank and the number of positive protein structures in the target data set are listed for each ligand. The number of positive structures shows that for some proteins many structures are contained in the sc-PDB, like for dorzolamide, where 152 structures of only three different proteins are found, in contrast to, e.g., cocaine, which has four protein targets of which, however, in the structure data set in total, only two true positive structures are contained.

Experiment Setup and Evaluation Metrics of the Drugs/sc-PDB Data Set. For measuring the capability of a screening method to rank true targets first, i.e., to obtain a good early enrichment, various performance metrics exist. We propose to

apply the following metrics, including well-established as well as new metrics assessing the very early enrichment performance:

- The enrichment factor at 1%, 2%, and 5%. The enrichment factor has well-known drawbacks.^{10,11,39} However, if calculated on the same data set, the values are comparable, and the simplicity of the metric makes it attractive because it is intuitively understandable.
- The AUC (area under the ROC curve). The AUC is also well established and is a measure of the complete screening performance in one number. It is also intuitively interpretable.
- The BEDROC (Boltzmann-enhanced discrimination of ROC) metric¹⁰ with factor $\alpha = 2$. The BEDROC metric applies decreasing exponential weight to measure the early enrichment. It has the advantage of being bounded by 0 (equal to random enrichment) and 1. A BEDROC with $\alpha = 2$ corresponds approximately to EF(0.5).⁴⁰
- The NSLR (normalized sum of logarithmic ranks) metric.³⁹ The NSLR metric weights the ranks of the true

positives logarithmically to measure the early enrichment and is also bounded by 0 and 1 (perfect ranking). A further advantage of this metric is that it is parameter-free.

METHODS

Inverse Screening. We used the proposed data sets for the evaluation of *iRAISE*, our newly developed inverse screening method.²² In brief, *iRAISE* (*inverse RApid Index-based Screening Engine*) is based on a triangle descriptor representation of the active site and the query ligand adapted from the TrixX approach.⁴¹ The RAISE technology is based on a bitmap index and a protein structure database that store precalculated triangle descriptors and initialized protein structures. For scoring, a five-step scoring cascade is applied that takes into account a reference cutoff score and the coverage of the ligand and the active site. The scoring cascade accounts for pocket diversity and thus enables inter-score comparison among proteins. Finally, a Gaussian-based weighting score estimates if a score is relevant for a protein pocket. This Gauss-weighting score is defined as the average score of all 84 ligands of the Astex diverse set¹⁵ and thus is protein pocket specific. In order to get a trustworthy average score, at least 20 of the 84 ligands must be scored for each protein pocket. Pockets with less than 20 successful dockings are not considered in evaluation, reducing, for example, the 7992 protein structures of the sc-PDB to 7915 structures considered in the evaluation. The Gauss-weighting score is used to weight the score from the scoring cascade, resulting in the final *iRAISE* score. Thus, the final score is easily interpretable, i.e., scores greater than the average score for the respective pocket are statistically relevant. This scoring enables a dynamic interpretation of the results. Instead of always using a given percentage of the top scored hits, those hits that scored better than average can be assessed experimentally.

The measures applied by *iRAISE* help to overcome the difficulties of inter-target ranking of protein–ligand scoring functions¹⁴ for diverse protein pockets. For details, see Schomburg.²²

For all experiments described in this paper, for the query ligand, up to 200 conformations were sampled with CONFECT.⁴²

pBLAST Sequence-Based Target Prediction. As a sequence-based target prediction method, protein-BLAST⁴³ from the NCBI/BLAST server was used with default parameters for the protein-BLAST algorithm with the exception of setting the number of results returned to maximum. As sequence database, we chose the complete PDB and later on filtered the results for protein entries that are also contained in the sc-PDB. As sequence query, we used the protein from the sc-PDB with which the ligand is co-crystallized because each of the 72 ligands of the Drugs/sc-PDB data set has at least one co-crystallized protein in the data set. If there are several structures available, the one with the alphabetically first PDB code was chosen. We downloaded the fasta-sequences for the proteins from the Protein Data Bank. The score returned from the protein-BLAST server was then used to rank the proteins from the sc-PDB set.

RESULTS

Selectivity Data Set. In Table 4, the screening results on the selectivity data set are summarized. The *iRAISE* score for each inhibitor/protein combination of the three target classes is listed. In *iRAISE*, if a ligand is scored with equal to or greater than 1.0, a protein is considered a target for this ligand. The scores listed are

Table 4. *iRAISE* Scores on the Selectivity Data Set^a

Inhibitor	iRAISE-score		
	PDE4	PDE5	
Drotaverine	1.0	0.9	
Piclamilast	1.4	1.1	
Roflumilast	1.2	0.9	
Avanafil	0.8	1.1	
Sildenafil	1.6	0.8	
Tadalafil	1.2	1.2	
Caffeine	1.0	0.9	
Paraxanthine	1.2	1.1	
Theophylline	1.0	1.2	
Histone Deacetylases			
	HDAC2	HDAC8	HDAC4
Mocetinostat	2.1	1.0	1.1
Vorinostat	1.6	1.4	1.1
Trichostatin A	1.3	1.7	1.2
Serine Proteases			
	Thrombin	Trypsin	FactorXa
Benzamidine	0.7	0.8	1.0
Pefabloc	1.3	1.2	1.0
Apixaban	--	0.9	--
Rivaroxaban	0.7	1.3	--
Melagatran	--	0.9	--

^aFor each inhibitor, its best score of the three structures of the respective protein is shown. If an inhibitor is active at the respective protein, the cell is highlighted in green; if it is inactive, the cell is highlighted in red. A hyphen indicates that no structure of the respective protein was hit by *iRAISE* for this inhibitor. Scores highlighted in bold red numbers are incorrectly predicted results of the respective method.

the “best” (highest) scores of the protein structures representing the proteins.

For the target class of phosphodiesterases, three PDE4 selective inhibitors, three PDE5 selective inhibitors, and three general PDE inhibitors are contained in the data set. For all PDE4 inhibitors, *iRAISE* correctly identifies the true targets, and only one negative target is falsely hit. For the PDE5 inhibitors, the classification is correct for avanafil but false for sildenafil. For tadalafil, the PDE5 structure is classified correctly, but the PDE4 structures are classified incorrectly.

In the case of the histone deacetylases, *iRAISE* finds all true positives but falsely classifies HDAC4 as a target for mocetinostat. Benzamidine and pefabloc are general serine protease inhibitors. For pefabloc, *iRAISE*’s scores are correct, but for benzamidine, only factor Xa is classified as a true target. For apixaban, correctly, no poses are generated in thrombin or factor Xa, but the low score in the trypsin protein scores is false. For rivaroxaban, the scores correctly map true negatives and true positives, while for melagatran, the negatives trypsin and factor Xa are correctly identified. However, it is the only case where *iRAISE* fails to produce any pose in a true target. Thus, this case was further examined.

In the PDB, a structure of thrombin co-crystallized with melagatran exists (PDB code 4BAH). Using this structure, *iRAISE* produces a pose of melagatran with a score of 1.22, which is better than the score for the factor Xa structure. Thus, *iRAISE* does not fail in general for this inhibitor, but the available structures have conformational changes that prevent *iRAISE* from creating a binding mode for melagatran. In Figure 4, the alignment of the two thrombin structures, 4BAH and 3SI4, is shown. In 3SI4, Ile174 is rotated into the active site, which would produce a clash with the binding mode of melagatran as in 4BAH.

In summary, the selectivity data set contains 29 true positive data points and 13 true negative data points. Of these, *iRAISE*

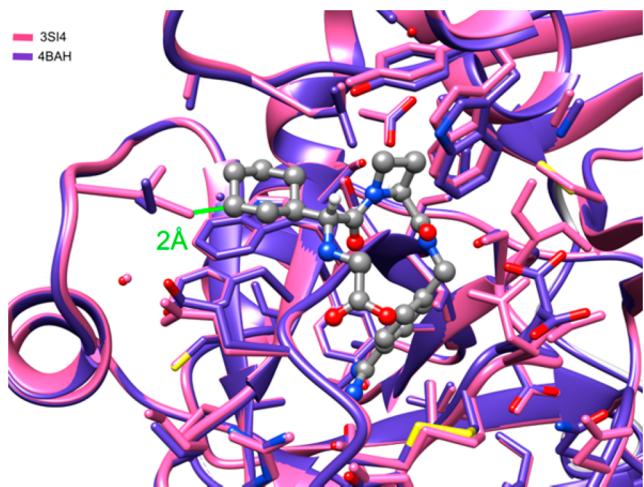


Figure 4. Alignment of the two thrombin structures 4BAH in purple and 3SI4 in pink with melagatran co-crystallized in 4BAH. Melagatran is shown in gray. In 3SI4, melagatran cannot be placed in the same binding mode as in 4BAH due to a different conformation of Ile174 that would create a clash with the inhibitor.

correctly identifies 24 true positives and 9 true negatives and thus reaches a 70% success rate for true negative prediction and 83% success rate for true positive predictions. This experiment shows that *iRAISE* is able to selectively classify proteins as targets or nontargets, but it is also dependent on the available protein conformations in the structure data set to account for target flexibility.

Drugs/sc-PDB Data Set. Binding Pose Prediction. *iRAISE*'s capability of predicting the correct binding mode was assessed by calculating the RMSD between the poses predicted by *iRAISE* (30 best ranked poses for each target) and the co-crystallized ligands (each ligand has at least one co-crystallized target in the protein data set). The RMSD of the ligands redocked into their co-crystallized targets ranges between 0.45 and 7.55 Å (for 7 of the 72 ligands, no binding pose in the co-crystallized protein structures was created by *iRAISE* because in the scoring cascade poses scored too low compared to the reference ligand are discarded). The median RMSD (of the best RMSD among the top 30 poses) is 1.13 Å, and the average RMSD is 1.66 Å. The RMSDs of 74% of the 72 ligands are below 2.0 Å (see Supporting Information for a list of all RMSDs).

Enrichment Evaluation. The performance of *iRAISE* in enrichment experiments and its capability of early recognition of true positives is evaluated as proposed by the metrics EF1%, EF2%, EF5%, AUC, BEDROC($\alpha = 2$), and NSLR. The results for each of the 38 singleton ligands are shown in Figure 5 and for the clustered ligands in Figure 6. In the top row, the median of each metric of all 72 ligands is given for comparison. Of all 72 ligands, the medians are 0.67 (AUC), 0.54 (BEDROC), and 0.28 (NSLR).

The metrics used can be exploited for classifying the enrichment potency of the inverse screening. For categorizing the achieved enrichments as “excellent, good, medium, or bad”, the following criteria were applied:

- excellent: AUC > 0.7, BEDROC > 0.6, and EF1% > 3
- good: AUC > 0.6, BEDROC > 0.6, and EF1% > 3 (two of the three criteria need to be fulfilled)
- medium: BEDROC > 0.4
- bad: all others

The classification is given for each ligand in the last column of Figures 5 and 6. In total, 20 ligands are classified with excellent enrichment, 20 with good enrichment, 21 with medium enrichment, and only 11 were classified with bad enrichment. Thus, for 55% of the test cases, good or excellent enrichment could be achieved with *iRAISE* screening, while the method failed for about 15% of the test cases.

In Figure 7, for each classification, the ROC plot of one example is shown. Naproxen (ROC plot, Figure 7A; for the 2D structure, see Figure 3, number 6S) is classified as “excellent” enrichment. Its targets are prostaglandin G/H synthases and glucuronosyltransferases. *iRAISE* is able to identify all structures as true targets and rank them to the beginning of the target list. The enrichment of prazosin (Figure 7B; for the 2D structure, see Figure 2, number 28) is classified only as “good” because all five true positive structures are correctly identified, but no true positive is found in the top 2.5% of the ranked list. Minocycline (Figure 7C; for the 2D structure, see Figure 2, number 18) is classified with “medium” enrichment. Only two out of a total of seven true positives are found. Only one of the five matrix-metalloprotease-9 structures is hit and also only one of the two collagenase structures. For ursodeoxycholic acid (Figure 7D; for the 2D structure, see Figure 3, number 71), the enrichment is classified as “bad”. The ranking of the 22 true positive structures of the protein aldo-ketoreductase tends to random distribution.

Like in the two examples shown for medium and bad enrichments, the reason for weak performance is mostly the fact that of the true positive structures in the data set not all are correctly identified by *iRAISE* screening. Because in *iRAISE*, protein flexibility is not handled internally, it is highly dependent on the correct protein conformation for binding the respective ligand, which is certainly a limitation of the method.

One example is methotrexate, which is classified with medium enrichment. The very first true positive structure of dihydrofolate reductase is found at rank 2 of the 7915 structures. However, of a total of 165 true positive structures of the true target proteins dihydrofolate reductase and thymidylate synthase contained in the data set, only 41 are identified as true positives. The fact that dihydrofolate reductase is a difficult target for docking due to its flexibility has already been observed by others.¹⁸ The loop regions around the active site of dihydrofolate reductase are highly flexible⁴⁴ and especially the movement of the Met20-loop can be observed in the crystal structures of this data set (e.g., structure of PDB code 1DLS with bound methotrexate and structure of PDB code 1JOM with bound leucovorin). In order to be able to correctly identify all dihydrofolate reductase structures, the method would need to be able to include protein backbone flexibility.

Prediction of Unknown Targets: 1. Analyzing Clustering Effects. Clustering the drugs shows an interesting effect. Although the molecule structures in one cluster are nearly identical, the target annotation in the DrugBank differs, as does the target prediction performance of *iRAISE*. The reason for the annotation difference may be high activity changes caused by small structure changes (activity cliffs) or missing target annotations in DrugBank. We compared the screening performance of *iRAISE* for the drugs with annotated targets from the DrugBank to its screening performance if the true positive target structures of one cluster were combined. Exemplarily, the ROC curves, annotated true positive proteins, and number of true positive structures for both annotation modes are shown in Figure 8 for the cluster containing fludrocortisone, dexametha-

	EF1%	EF2%	EF5%	AUC	BEDROC	NSLR	Classification
MEDIAN	3.34	3.64	2.8	0.67	0.54	0.28	G
1 Dorzolamide	17.14	11.54	7.5	0.67	0.58	0.47	G
2 Lovastatin	3.45	5.18	6.89	0.67	0.55	0.29	G
3 Estradiol	7.51	9.39	7.75	0.83	0.73	0.46	E
4 Efavirenz	29	16.26	8.24	0.69	0.59	0.51	G
5 Delavirdine	0	0.44	1.23	0.52	0.36	0.20	B
6 Niflumic acid	3.78	2.84	2.64	0.70	0.54	0.28	G
7 Nevirapine	2.64	4.39	2.63	0.57	0.42	0.26	M
8 Galantamine	5.42	2.71	2.16	0.73	0.57	0.29	G
9 Sitagliptin	0	1.04	0.42	0.47	0.30	0.14	B
10 Tadalafil	5.5	3.85	5.05	0.73	0.61	0.38	E
11 Imatinib	3.2	2.13	2.23	0.56	0.40	0.26	M
12 Succinylcholine	33.4	16.7	6.66	0.66	0.54	0.26	G
13 Apixaban	0	1.15	0.92	0.56	0.41	0.22	M
14 Cocaine	0	0	0	0.86	0.73	0.24	G
15 Dichlorphenamide	16.48	12.52	7.63	0.73	0.64	0.51	E
16 Proflavine	0	0	0.57	0.60	0.42	0.22	M
17 Phenylbutazone	0	0	0	0.44	0.26	0.10	B
18 Minocycline	0	7.16	2.86	0.59	0.45	0.19	M
19 Indomethacine	1.96	2.62	2.09	0.58	0.42	0.26	M
20 Pentoxyfylline	3.07	3.58	4.28	0.77	0.63	0.36	E
21 Chlormaphenicol	0	2.95	1.57	0.60	0.43	0.21	M
22 Finasteride	22.26	16.7	6.66	0.69	0.58	0.32	G
23 Topiramate	8.57	6.26	4.6	0.63	0.52	0.37	G
24 Papaverine	15.41	12.11	7.69	0.81	0.72	0.49	E
25 Balsalazide	0	0.52	0.62	0.48	0.32	0.16	B
26 Ethoxzolamide	3.95	4.61	4.34	0.79	0.66	0.44	E
27 Gefitinib	0.72	2.15	1.14	0.57	0.41	0.24	M
28 Prazosin	0	0	11.99	0.95	0.89	0.39	G
29 Pyrimethamine	0.8	1.6	1.92	0.55	0.40	0.23	M
30 Tolmetin	4.01	4.01	3.2	0.81	0.66	0.31	E
31 Thiabendazole	0	0	0	0.95	0.89	0.33	G
32 Ethacrynic acid	0	0	0	0.70	0.50	0.19	M
33 Abacavir	0.73	0.73	1.46	0.66	0.51	0.29	M
34 Varenicline	100.19	50.09	19.99	0.99	0.98	0.55	E
35 Captopril	0	0	0.64	0.53	0.36	0.15	B
36 Celecoxib	9.25	5.39	2.77	0.63	0.48	0.28	G
37 Levonorgestrel	10.3	9.83	5.6	0.67	0.55	0.39	G
38 Zonisamide	9.33	7.78	6.08	0.86	0.76	0.53	E

Figure 5. Performance metrics for the 38 singleton ligands of the Drugs/sc-PDB data set. The classification abbreviations are E for excellent enrichment, G for good enrichment, M for medium enrichment, and B for bad enrichment.

sone, and hydrocortisone (for the 2D structures, see Figure 3, numbers 60, 58, and 59, respectively).

For fludrocortisone, the enrichment in fact improves with all 42 structures as targets. Thus, iRAISE suggests that the additional structures are also targets for this ligand. The four structures added are cholesterol side-chain cleavage enzyme structures, of which dexamethasone is known as an inhibitor. However, for the other two ligands, the enrichment gets worse if all the targets are combined, suggesting that these proteins are no targets for these drugs, despite the ligand similarity. For dexamethasone, the added structures of the mineralocorticoid receptor, however, are ranked at the very beginning at ranks 1, 8, and 23 and at ranks 113 and 2749. This effect has already been observed in experiments, as literature reporting the inhibition of the mineralocorticoid receptor by dexamethasone shows.^{45–49}

Further evaluation has to show if the suggestions of iRAISE that were not confirmed by literature so far are correct.

Prediction of Unknown Targets: 2. Analyzing Predictions by iRAISE. A practical application of inverse screening methods is

the prediction of unknown targets for a given ligand. Successful drug repurposing projects show that the target annotation for drugs is far from being complete. Therefore, we studied the target predictions of iRAISE for the cluster of the antidepressants imipramine, desipramine, and clomipramine in detail in order to see which targets next to the annotated targets from the DrugBank for this cluster were predicted by iRAISE.

Imipramine, desipramine (active metabolite of imipramine), and clomipramine are highly similar in structure (for the 2D structures, see Figure 3, numbers 54, 55, and 53, respectively). Common annotated targets for these ligands in the DrugBank are the serotonin transporter, noradrenalin transporter, and 5-hydroxytryptamine receptor. Clomipramine further has annotated the targets serum albumin and glutathione S-transferase. Imipramine has further annotated the targets adrenergic receptor, histamine H1 receptor, and muscarinic acetylcholine receptor. Looking at iRAISE's hit lists of clomipramine, imipramine, and desipramine, the high rank of structures of acetylcholinesterase (ACHE) for all three ligands stands out. For

	EF1%	EF2%	EF5%	AUC	BEDROC	NSLR	Classification
MEDIAN	3.34	3.64	2.8	0.67	0.54	0.28	G
39 Ampicillin	0	2.95	2.35	0.67	0.49	0.20	M
40 Penicillin V	2.78	2.78	2.78	0.61	0.47	0.24	M
41 Meticillin	0	0	0	0.60	0.45	0.45	M
42 Penicillin G	0	0	6.66	0.78	0.66	0.26	G
43 Progesterone	20.04	14.47	8.44	0.71	0.61	0.46	E
44 Testosterone	16.7	12.99	6.66	0.74	0.64	0.43	E
45 Spironolactone	12.93	11.31	4.51	0.64	0.52	0.32	G
46 Diazepam	0	3.85	3.07	0.79	0.65	0.28	G
47 Midazolam	0	0	0	0.66	0.49	0.16	M
48 Diclofenac	4.23	4.23	2.82	0.71	0.55	0.30	G
49 Mefenamic acid	4.36	8.71	4.35	0.78	0.65	0.32	E
50 Meclofenamic acid	13.66	9.11	4.35	0.79	0.65	0.35	E
51 Methotrexate	3.64	3.64	2.66	0.58	0.43	0.29	M
52 Raltitrexed	0	1.96	1.57	0.59	0.44	0.22	M
53 Clomipramine	13.07	10.89	6.95	0.72	0.61	0.34	E
54 Imipramine	6.46	4.85	3.22	0.57	0.43	0.23	M
55 Desipramine	3.23	4.85	1.93	0.52	0.38	0.19	B
56 Trimethoprim	2.5	2.5	2.25	0.60	0.45	0.28	M
57 Trimetrexate	9.62	8.82	6.24	0.70	0.60	0.44	E
58 Dexamethasome	23.12	11.56	6.15	0.85	0.76	0.43	E
59 Hydrocortisone	0	12.52	5	0.74	0.60	0.26	G
60 Fludrocortisone	13.18	10.55	5.26	0.64	0.52	0.31	G
61 Chlorpromazine	0	0	0	0.67	0.53	0.19	M
62 Trifluoperazine	4.77	2.39	1.43	0.54	0.38	0.19	B
63 Alogliptin	2.09	2.09	0.83	0.56	0.39	0.18	B
64 Linagliptin	0	0	0	0.49	0.33	0.14	B
65 Naproxene	12.02	10.02	8.79	0.87	0.77	0.42	E
66 Nabumetone	4.17	8.35	5.83	0.82	0.71	0.36	E
67 Hydrochlorothiazide	18.99	16.37	11.1	0.85	0.79	0.63	E
68 Hydroflumethiazide	18.34	15.39	10.06	0.78	0.71	0.58	E
69 Sildenafil	2.64	1.32	1.58	0.55	0.40	0.19	M
70 Vardenafil	2.57	2.57	1.02	0.53	0.38	0.18	B
71 Ursodeoxycholic acid	4.55	2.28	2.73	0.51	0.37	0.18	B
72 Chenodeoxycholic acid	0	0	2.1	0.78	0.65	0.28	G

Figure 6. Performance metrics for clustered ligands of the Drugs/sc-PDB data set. The classification abbreviations are E for excellent enrichment, G for good enrichment, M for medium enrichment, and B for bad enrichment. The frames around entries indicate clusters of drugs.

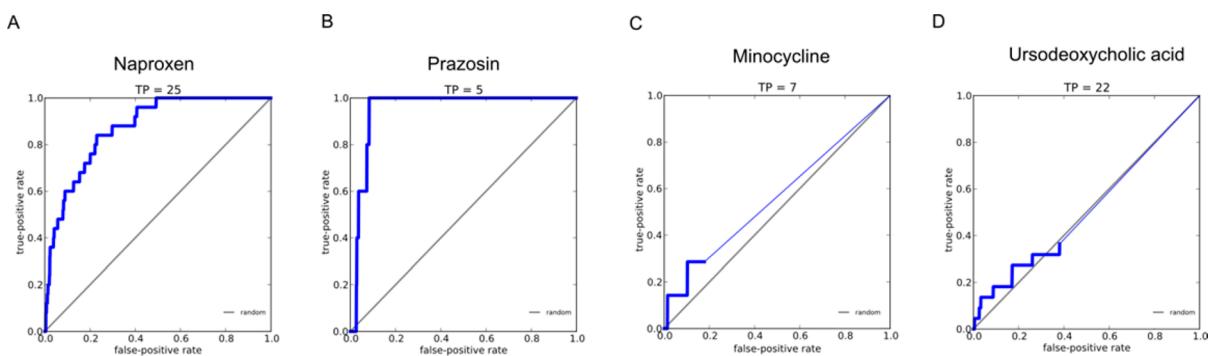


Figure 7. ROC curves of four ligands of the Drugs/sc-PDB data set. Examples for “excellent” enrichment (A), “good” enrichment (B), “medium” enrichment (C), and “bad” enrichment (D). TP = number of true positive structures.

clomipramine, a structure of ACHE was ranked at position 3, for imipramine at rank 50, and for desipramine at rank 12 (see the Supporting Information for a list of the first 50 ranked targets for these three ligands). A literature study verified the competitive inhibition of ACHE by the antidepressants paroxetine, imip-

amine, clomipramine, and sertraline.⁵⁰ In Figure 9 A–C, iRAISE’s poses of the three ligands in the ACHE structure (PDB code 1GPN) are shown. The co-crystallized ligand is shown in Figure 9D. For imipramine and desipramine, which differ in their structure only in a methyl group, almost the same

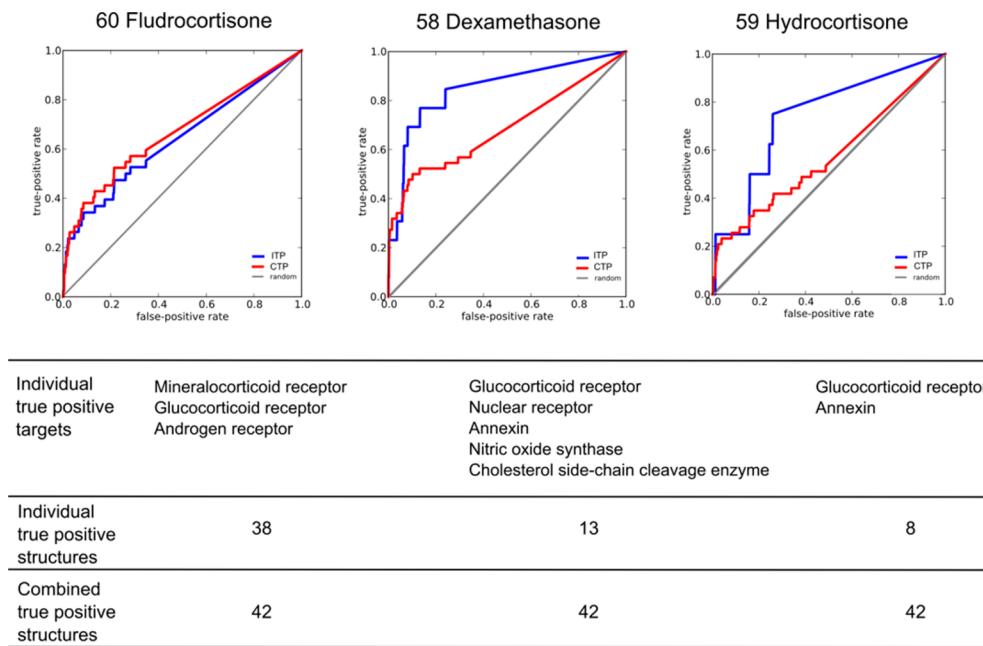


Figure 8. ROC curves of three drugs grouped by the ECFP-clustering in the same cluster. Comparison of ROC curves for true positives based on the target annotations by DrugBank (blue lines) and by combination of the targets annotations for all three ligands (red line). ITP = individual true positives; CTP = combined true positives.

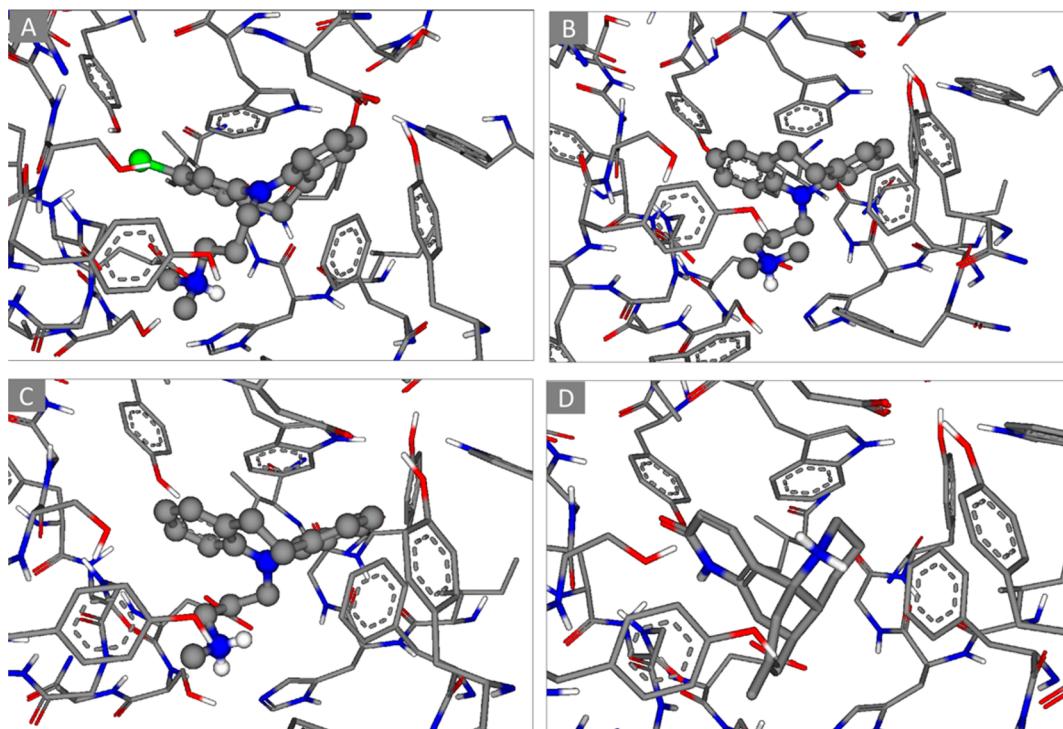


Figure 9. Poses of antidepressant drugs in acetylcholinesterase structures (PDB: 1GPN): (A) clomipramine, (B) imipramine, (C) desipramine, and (D) co-crystallized ligand huperzine B (PDB HET code HUB).

pose is created. In the generated pose for clomipramine, the ring scaffold is placed in the reversed orientation due to the space-demanding chloride that is not present in the other structures.

Comparison to Sequence-Based Target Prediction. For evaluation of the performance of pBLAST on the Drugs/sc-PDB data set, we calculated the same metrics as used for assessing the performance of iRAISE for 71 of the 72 ligands of the Drugs/sc-PDB data set using the ranking by pBLAST score. One ligand

was excluded from the evaluation because it only has the one true positive in the data set with which it was co-crystallized.

In Table S, the median AUC, median NSLR, and median BEDROC for the target predictions of the 71 ligands based on pBLAST are reported in comparison to those of iRAISE. On the basis of these medians, iRAISE's target prediction performance is superior in this data set. However, because the medians of these performance measures on the data set only show one snapshot of

Table 5. Median Performance Metrics for Enrichment of the Drugs/sc-PDB Data Set for iRAISE and pBLAST

	median AUC	median NSLR	median BEDROC
iRAISE	0.67	0.28	0.54
pBLAST	0.56	0.14	0

the total performance, the performance is now compared in detail for each ligand.

In Figure 10, for each of the 71 ligands, the AUC (Figure 10A), BEDROC (Figure 10B), and NSLR (Figure 10C) are shown in histograms. The ligands are in all diagrams sorted on the abscissa by the AUC of the pBLAST predictions. These histograms show for some ligands very good enrichment results for the sequence-based predictions with AUC values up to 1.0 as well as BEDROCs and NSLRs of up to 1.0. However, for more than half of the ligands, the sequence-based target predictions fail, which is the reason for the unsatisfactory median performance metrics.

In order to evaluate further on which examples the sequence-based predictions outperform those of iRAISE, the number of

different proteins represented in the true target structures is plotted in Figure 10D. The ligands on which the performance metrics are highest are mostly ligands for which only structures of one or two different proteins are contained in the true positive structure set. However, the sequence-based method also performs well on one ligand (dichlorphenamide) with four different protein targets and also on the ligand (imatinib) with the highest number of true positive structures from different protein targets (10 different protein targets). Nevertheless, the four different proteins for dichlorphenamide are different carbonic anhydrases, and for imatinib, the 10 different targets are all from the protein family of protein kinases, which both exhibit high sequence similarity.

Two conclusions can be drawn from this experiment. First, the Drugs/sc-PDB data set is suitable for evaluation of structure-based methods because it contains examples that cannot be assessed with sequence-based methods. Second, iRAISE is able to identify true off-targets with low sequence identity to known targets, and thus, it is able to perform well on cases not predictable with sequence-based target prediction. These

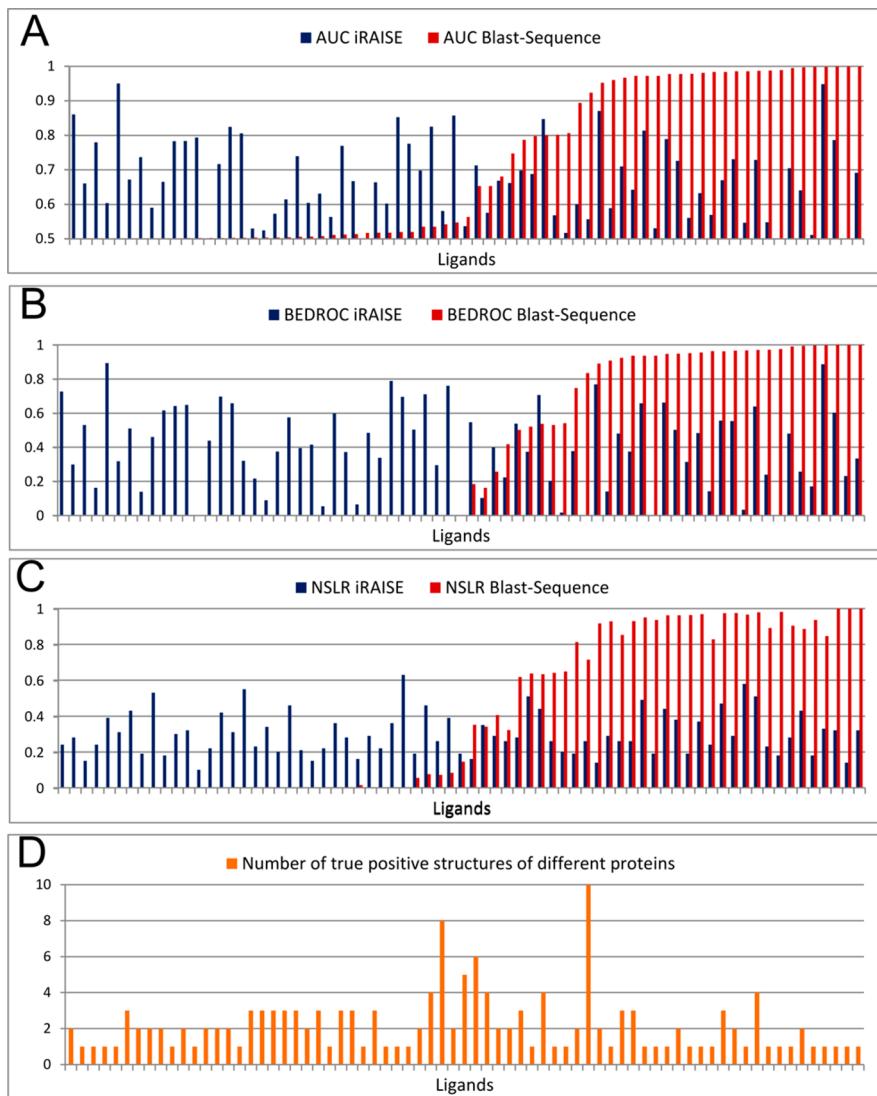


Figure 10. AUC, BEDROC, and NSLR for comparison of iRAISE and pBLAST target prediction on 71 ligands of the Drugs/sc-PDB data set. All ligands in all four plots are sorted on the abscissa by ascending AUC of the pBLAST predictions. The fourth plot shows the number of different proteins present in the true target structures. On the ordinate, the value of the metric labeled in the heading of each diagram is plotted.

findings support the strategy of using a combination of sequence-based and structure-based methods, i.e., identification of “obvious” targets with sequence-based methods while using structure-based methods to identify true off-targets.

CONCLUSION

In contrast to standard virtual screening, for inverse screening, no standard evaluation representing real world use cases has been established. We propose standard evaluation methods based on two data sets and performance metrics able to measure “early recognition”. The two newly composed data sets contain only openly available data. A small one consisting of three protein classes allows proof-of-concept studies and evaluation of specificity versus sensitivity. A large one based on drugs and the sc-PDB allows annotation of true positive proteins due to the well-studied selectivity of drugs. True positives were annotated automatically using DrugBank, the EC number, and the UniProtID of the proteins. Although also with this data set a save true negative annotation is not possible, the number of false negative assignments should be reduced. The metrics that we proposed are already established metrics like AUC and EF as well as relatively new metrics suited better for “early enrichment” recognition like BEDROC and NSLR. As these metrics measure the screening performance of a method in one value, we preferred to list several measures that focus on different aspects instead of choosing a single one. On the basis of these metrics, a classification of enrichments into excellent, good, medium, and bad is proposed, allowing a fast quality assessment. Applying a sequence-based method on the data set demonstrates that it contains cases that cannot be predicted by sequence similarity.

As a first use of the proposed evaluation strategy, we evaluated our inverse screening method *iRAISE* on the proposed data sets showing its capability of early recognition and its ability to predict unknown targets for small molecules. The selectivity enhancement of *iRAISE*’s prediction leads to fewer false positives and to selection of proteins in the correct conformation. Success rates of 70% for true negative identification and 83% for true positive identification are achieved. Evaluation on the large data set of 72 ligands and 7915 structures resulted in a total of 41 excellent or good enrichments. Further analysis has to show why bad enrichment was observed for 11 ligands. Comparison with a sequence-based method showed superior performance for prediction of off-targets. This underlines our understanding of how to use these methods. Structure-based methods shall be used in complement to sequence-based methods because both have strengths and weaknesses and perform well on different examples. Furthermore, sequence-based methods are only applicable if at least one known protein target for a ligand exists, and they do not predict binding modes.

Even more importantly for the emerging field of structure-based target identification, the two data sets can act as a benchmark set for further developments. All data that has been used in this study is publicly available, and all necessary details for using the data sets are reported.

ASSOCIATED CONTENT

Supporting Information

List of used and discarded structures from the sc-PDB protein structures. List of targets for the Drugs/sc-PDB data set extracted from the DrugBank. True positive PDB codes for each ligand of the Drugs/sc-PDB data set. RMSDs of the Drugs/sc-PDB data set of the binding modes predicted by *iRAISE*. Lists of the first 50 predicted targets for clomipramine, imipramine, and desipr-

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

AUTHOR INFORMATION

Corresponding Author

*E-mail: rarey@zbh.uni-hamburg.de. Tel. 004940428387350.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the Structural Chemogenomics Group of the Universite de Strasbourg for creating and providing the sc-PDB freely. This project is part of the Excellence Cluster in Excellence Initiative by the State of Hamburg “Fundamentals of Synthetic Biological Systems (SynBio)” (www.tu-harburg.de/synbio).

ABBREVIATIONS

EF, enrichment factor; BEDROC, Boltzman-enhanced discrimination of ROC; ROC, receiver operating characteristic; PDB, Protein Data Bank; TTFXa, trypsin/thrombin/factor Xa; PDE, phosphodiesterases; HDAC, histone deacetylases; NSLR, normalized sum of logarithmic ranks; ACHE, acetylcholinesterase

REFERENCES

- (1) Jenkins, J. L.; Bender, A.; Davied, J. W. In silico target fishing: Predicting biological targets from chemical structure. *Drug Discovery Today* **2006**, *3* (4), 413–421.
- (2) Ekins, S.; Mesters, J.; Testa, B. In silico pharmacology for drug discovery applications to targets and beyond. *Br. J. Pharmacol.* **2007**, *152*, 21–37.
- (3) Rognan, D. Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* **2007**, *152*, 38–52.
- (4) Hopkins, A. L. Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4* (11), 682–690.
- (5) Rognan, D. Structure-based approaches to target fishing and ligand profiling. *Mol. Inf.* **2010**, *29*, 176–187.
- (6) Tanrikulu, Y.; Krüger, B.; Proschak, E. The holistic integration of virtual screening in drug discovery. *Drug Discovery Today* **2013**, *18*, 358–364.
- (7) Rognan, D. Structure-based approaches to target fishing and ligand profiling. *Mol. Inf.* **2010**, *29*, 176–187.
- (8) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. Recognizing pitfalls in virtual screening: A critical review. *J. Chem. Inf. Model.* **2012**, *52*, 867–881.
- (9) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.
- (10) 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.
- (11) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 213–228.
- (12) Chen, Y. Z.; Ung, C. Y. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. *J. Mol. Graph. Model.* **2001**, *20* (3), 199–218.

- (13) 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), 217–226.
- (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–W224.
- (15) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein–ligand docking performance. *J. Med. Chem.* **2007**, *50* (4), 726–741.
- (16) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.
- (17) Koutsoukas, A.; Lowe, R.; KalantarMotamedi, Y.; Mussa, H. Y.; Klaffke, W.; Mitchell, J. B. O.; Glen, R. C.; Bender, A. In silico target predictions: Defining a benchmarking dataset and comparison of performance of the multiclass naïve bayes and Parzen–Rosenblatt window. *J. Chem. Inf. Model.* **2013**, *53* (8), 1957–1966.
- (18) Paul, N.; Kellenberger, E.; Bret, G.; Müller, P.; Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **2004**, *54* (4), 671–680.
- (19) Kellenberger, E.; Foata, N.; Rognan, D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J. Chem. Inf. Model.* **2008**, *48* (5), 1014–1025.
- (20) Meslamani, J.; Li, J.; Sutter, J.; Stevens, A.; Bertrand, H.-O.; Rognan, D. Protein-ligand-based pharmacophores: Generation and utility assessment in computational ligand profiling. *J. Chem. Inf. Model.* **2012**, *52* (4), 943–955.
- (21) Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: A database for identifying variations and multiplicity of ‘druggable’ binding sites in proteins. *Bioinformatics* **2011**, *27* (9), 1324–1326.
- (22) Schomburg, K. T.; Bietz, S.; Briem, H.; Henzler, A.; Urbaczek, S.; Rarey, M. Facing the challenges of structure-based target prediction by inverse virtual screening. *J. Chem. Inf. Model.* **2014**, *54*, 1676–1686.
- (23) Czodrowski, P.; Sotriffer, P. A.; Klebe, G. Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pka calculations and itc experiments. *J. Mol. Biol.* **2007**, *367* (5), 1347–1356.
- (24) Di Fenza, A.; Heine, A.; Koert, U.; Klebe, G. Understanding binding selectivity toward trypsin and factor Xa: The role of aromatic interactions. *ChemMedChem* **2007**, *2* (3), 297–308.
- (25) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J.-M.; Wienen, W.; Pripke, H. V. M.; Kaufmann, I. K.; Ries, U. J.; Hauel, N. H. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure* **2001**, *9* (1), 29–37.
- (26) Glinca, S.; Klebe, G. Cavities tell more than sequences: Exploring functional relationships of proteases via binding pockets. *J. Chem. Inf. Model.* **2013**, *53* (8), 2082–2092.
- (27) Spina, D. PDE4 inhibitors: Current status. *Br. J. Pharmacol.* **2008**, *155* (3), 308–315.
- (28) Wright, P. J. Comparison of phosphodiesterase type 5 (PDE5) inhibitors. *Int. J. Clin. Pract.* **2006**, *60* (8), 967–975.
- (29) Dokmanovic, M.; Clarke, C.; Marks, P. A. Histone deacetylase inhibitors: Overview and perspectives. *Mol. Cancer Res.* **2007**, *5* (10), 981–989.
- (30) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissenig, H.; Shindyalov, I. N.; Bourne, B. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (31) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. Drugbank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668–D672.
- (32) Schomburg, I.; Chang, A.; Placzek, S.; Söhngen, C.; Rother, M.; Lang, M.; Munaretto, C.; Ulas, S.; Stelzer, M.; Grote, A.; Scheer, M.; Schomburg, D. BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. *Nucleic Acids Res.* **2013**, *41*, D764–D772.
- (33) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. Pubchem: Integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217–241.
- (34) ftp://cheminfo.u-strasbg.fr (accessed January 2013).
- (35) Hilbig, H.; Urbaczek, S.; Groth, I.; Heuser, S.; Rarey, M. MONA – Interactive manipulation of molecule collections. *J. Cheminform.* **2013**, *5*, 38.
- (36) Urbaczek, S.; Kolodzik, A.; Groth, I.; Heuser, S.; Rarey, M. Reading PDB: Perception of molecules from 3D atomic coordinates. *J. Chem. Inf. Model.* **2012**, *53* (1), 76–87.
- (37) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754.
- (38) Velankar, S.; Dana, J. M.; Jacobsen, J.; Van Ginkel, G.; Gane, P. J.; Luo, J.; Oldfield, T. J.; O'Donovan, C.; Martin, M.-J.; Kleywegt, G.-J. Sifts: Structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.* **2013**, *41*, D483–D489.
- (39) Venkatraman, V.; Pérez-Nueno, V.; Mavridis, L.; Ritchie, D. W. Comprehensive comparison of ligand-based virtual screening tools against the DUD dataset reveals limitations of current 3D methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079–2093.
- (40) Riniker, S.; Landrum, G. A. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *J. Cheminform.* **2013**, *5*, 26.
- (41) Schlosser, J.; Rarey, M. Beyond the virtual screening paradigm: Structure-based searching for new lead compounds. *J. Chem. Inf. Model.* **2009**, *49*, 800–809.
- (42) Schärfer, C.; Schulz-Gasch, T.; Hert, J.; Heinzerling, L.; Schulz, B.; Inhester, T.; Stahl, M.; Rarey, M. CONFECT: Conformations from an expert collection of torsion patterns. *ChemMedChem* **2013**, *8* (10), 1690–1700.
- (43) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
- (44) Osborne, M. J.; Schnell, J.; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. Backbone dynamics in dihydrofolate reductase complexes: Role of loop flexibility in the catalytic mechanism. *Biochemistry* **2001**, *40* (33), 9846–9859.
- (45) Hagley, R. D.; Watlington, C. O. Down-regulation of the mineralocorticoid receptor by dexamethasone in an amphibian kidney cell line (A6). *Endocr. Res.* **1996**, *22* (3), 221–235.
- (46) Marshall, D. R.; Gus Rodriguez, G.; Thomson, D. S.; Nelson, R.; Capolina, A. α -Methyltryptamine sulfonamide derivatives as novel glucocorticoid receptor ligands. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 315–319.
- (47) Hudson, A. R.; Roach, S. L.; Higuchi, R. I.; Philipps, D. P.; Bissonnette, R. P.; Lamph, W. W.; Yen, J.; Li, Y.; Adams, M. E.; Valdez, L. J.; Vassar, A.; Cuervo, C.; Kallel, E. A.; Gharbaoui, C. J.; Mais, D. E.; Miner, J. N.; Marshke, K. B.; Rungra, D.; Negro-Vilar, A.; Zhi, L. Synthesis and characterization of nonsteroidal glucocorticoid receptor modulators for multiple myeloma. *J. Med. Chem.* **2007**, *50* (19), 4699–4709.
- (48) Jaroch, S.; Berger, M.; Huwe, C.; Krolkiewicz, K.; Rehwinkel, H.; Schäcke, H.; Schmees, N.; Skuballa, W. Discovery of quinolines as selective glucocorticoid receptor agonists. *Bioorg. Med. Chem. Lett.* **2010**, *20* (19), 5835–5838.
- (49) Regan, J.; Lee, T. W.; Zindell, R. M.; Bekkali, Y.; Bentzien, J.; Gilmore, T.; Hammach, A.; Kirrane, T. M.; Kukulka, A. J.; Kuzmich, D.; Nelson, R. M.; Proudfoot, J. R.; Ralph, M.; Pelletier, J.; Souza, D.; Zuvela-Jelaska, L.; Nabozny, G.; Thomson, D. S. Quinol-4-ones as steroid A-ring mimetics in nonsteroidal dissociated glucocorticoid agonists. *J. Med. Chem.* **2006**, *49*, 7887–7896.
- (50) Ahmed, M.; Rocha, J. B. T.; Mazzanti, C. M.; Hassan, W.; Morsch, V. M.; Loro, V. L.; Thome, G.; Schetinger, M.R.C. Comparative study of the inhibitory effect of antidepressants on cholinesterase activity in Bungarussindanus (krat) venom, human serum and rat striatum. *J. Enzyme Inhib. Med. Chem.* **2008**, *23* (6), 912–917.