
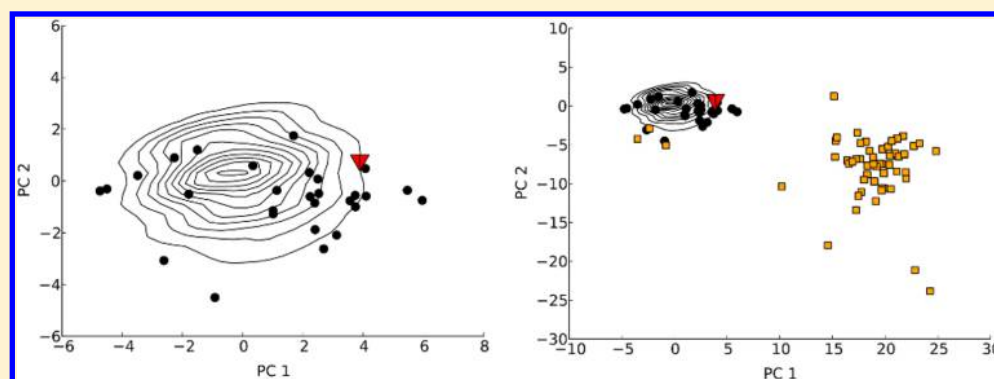


Toward a Benchmarking Data Set Able to Evaluate Ligand- and Structure-based Virtual Screening Using Public HTS Data

Martin Lindh, Fredrik Svensson, Wesley Schaal, Jin Zhang,[†] Christian Sköld, Peter Brandt, and Anders Karlén*

Organic Pharmaceutical Chemistry, Department of Medicinal Chemistry, Uppsala University, Biomedical Centre, Box 574, SE- 751 23 Uppsala, Sweden

 Supporting Information



ABSTRACT: Virtual screening has the potential to accelerate and reduce costs of probe development and drug discovery. To develop and benchmark virtual screening methods, validation data sets are commonly used. Over the years, such data sets have been constructed to overcome the problems of analogue bias and artificial enrichment. With the rapid growth of public domain databases containing high-throughput screening data, such as the PubChem BioAssay database, there is an increased possibility to use such data for validation. In this study, we identify PubChem data sets suitable for validation of both structure- and ligand-based virtual screening methods. To achieve this, high-throughput screening data for which a crystal structure of the bioassay target was available in the PDB were identified. Thereafter, the data sets were inspected to identify structures and data suitable for use in validation studies. In this work, we present seven data sets (MMP13, DUSP3, PTPN22, EPHX2, CTDSP1, MAPK10, and CDK5) compiled using this method. In the seven data sets, the number of active compounds varies between 19 and 369 and the number of inactive compounds between 59 405 and 337 634. This gives a higher ratio of the number of inactive to active compounds than what is found in most benchmark data sets. We have also evaluated the screening performance using docking and 3D shape similarity with default settings. To characterize the data sets, we used physicochemical similarity and 2D fingerprint searches. We envision that these data sets can be a useful complement to current data sets used for method evaluation.

■ INTRODUCTION

Virtual screening (VS) is used to identify bioactive compounds and new chemical starting points through computational means and is an important and integral part of modern drug discovery. A plethora of different methods have been reported, varying both in information accounted for and in the algorithms implemented.¹ Most VS techniques can be categorized either as ligand-based (LB) or structure-based (SB). LB methods rely on information from known active compounds, while the SB methods rely on information about the structure of the macromolecular target. The use of combined methods, either in data fusion protocols^{2–4} or in multistep VS where techniques with different performance and computational cost are sequentially used is common practice.⁵

To develop, evaluate, and compare VS techniques, comparison with experimental data is required. Validation of the performance of VS techniques using known active

compounds and untested decoys has been widely used. As the field has matured, more developed test sets have been created to avoid the risk of artificial enrichment⁶ and analogue bias.⁷ Many earlier benchmark data sets targeted toward SB VS have had the shortcoming that they include close analogues as actives, something which will favor 2D methods.⁸

Some of the more widely used benchmarking data sets include the Directory of Useful Decoys (DUD),⁹ the improved DUD-E,¹⁰ Demanding Evaluation Kits for Objective *In silico* Screening (DEKOIS),¹¹ and the Maximum Unbiased Validation (MUV) data set,¹² with DUD being the most cited.

Many of the publicly available benchmarking data sets, through their designs, target either SB or LB techniques, exemplified by MUV aimed at LB and DUD that is designed for

Received: September 9, 2014

Published: January 7, 2015

SB methods. Thus, comparisons may be unreliable when applying both structure- and ligand-based methods to the same data sets. For example, the design used to compile DUD/DUDE excludes decoys that are challenging for 2D fingerprint similarity methods,¹³ and the results from such methods are, thus, inherently biased. However, this problem can to some extent be relieved by including only a subset of DUD, which is selected to exclude data sets with a low number of chemotypes.^{14,15}

Many of the problems associated with VS validation were considered by Baumann et al. in the creation of the MUV data set, as both analogue bias and artificial enrichment were minimized in their study. The MUV data set was primarily aimed at validation of LB methods, but three-dimensional structures were available from the PDB for seven of the 17 targets.¹² However, as the SB approach was not the focus in the design, not all MUV data sets were well suited for docking evaluation, as exemplified by HIV RT-RNase, where there are multiple sites for inhibitor binding.¹⁶

Although designed data sets may be useful for many aspects of VS method development and intramethod comparisons, benchmarking of methods against these data sets will most likely not give an appropriate description of the performance in real case scenarios. For example, the fraction of inactive compounds will be much larger in a HTS compared to the case of VS using most designed data sets. Thus, a diverse collection of HTS data sets would give a more realistic measure of the performances of different VS methods when applied in actual drug discovery efforts. In addition, selecting active compounds from the literature often results in highly similar and optimized structure series, typically with high affinity for their targets. This will be quite different from a HTS scenario where the “hit” compounds often express a fairly low activity.

It has also been shown that scoring functions favor large molecules over small molecules.⁶ If the decoys are similar to the actives in size and in physicochemical properties, this and other potential biasing factors will not be addressed. Keeping the large structural diversity of HTS libraries can therefore be important in method development in order to avoid overfitting methods against the small number of decoys present in many of the training data sets.

Data deposited in PubChem have previously been used to construct benchmarking data sets for LB methods.^{12,17,18} In this study, we identify PubChem data sets suitable for validation of both structure and ligand-based virtual screening methods and present the methods that we have applied to identify suitable targets with accompanying bioassay data and the subsequent filtering to generate the final validation data sets. Using this strategy, we present seven data sets suitable for validation of VS methods.

METHODS

A general description of the methods is given here. Scripts developed for the process are available in the Supporting Information (Scripts S1–S3).

Data Set Compilation. HTS data were collected from The PubChem BioAssay database,¹⁹ and protein 3D structures were collected from the Brookhaven Protein Data Bank²⁰ (PDB). PubChem lists each assay (AID) with a gene number (gi) that defines the target. A gi links the bioassay to the corresponding protein target gene. We searched for HTS bioassay data sets with a primary screen with more than 1000 screened compounds. The data set should have a linked protein target,

of the same species that was used in the HTS, available in the PDB.

Available crystal structures linked to bioassay data were reviewed, and only structures with a reported resolution below 3 Å were taken further. These structures were evaluated using a semiautomatic method to locate crystal structures with a bound drug-like ligand. We defined non-drug-like ligands as DNA fragments, common cofactors (e.g., ATP, CTP, NADPH, hemoglobin), modified cofactors, modified amino acid residues, small ions, or compounds that were used to facilitate crystallization (for a complete list of ligands that were defined as nondrug-like, see Script S2). Only crystal structures binding a drug-like ligand were kept. Crystal structures with a DNA fragment or cofactor bound were regarded as having multiple binding sites and were not progressed. Finally, the resulting list of potential targets was reviewed manually. Targets for which we identified structures with covalently bound ligands were discarded. Bioassays where the data originated from cell-based or multiplex assays were discarded due to the uncertainty in the actual target evaluated. The scope of each bioassay was also considered in order to ensure that the assay was designed to locate competitive inhibitors.

For any pair of primary and secondary screens, an active compound was marked as active if it was reported active in both these screens (the active and inactive status of a compound were defined based on the PubChem annotation). Any compound with a different result with respect to being active between the screens was not used in the data sets. The remaining inactive compounds were used as decoys. The final protein structure was taken from the PDB entry with the highest resolution and the fewest missing atoms. At the time of the study, this resulted in the structure with the highest resolution for all targets. Electron densities were downloaded from the Uppsala Electron Density Server²¹ and inspected to make sure the ligands and ligand binding sites were well-defined.

To further characterize the hits as true positives, the Hill slope for each active was extracted from the PubChem data. Active compounds with a Hill slope <0.5 or >2.0 were flagged as potential false positives. In addition, compounds matching a PAINS filter²² (implemented in InstantJChem²³) were also flagged as potential false positives.^{24,25}

Data Set Characterization. Physicochemical Characterization and Artificial Enrichment. Canvas²⁶ was used to generate physicochemical descriptors for all tested compounds after desalting and neutralization with the Schrödinger utilities “neutralizer” and “desalter.” Properties calculated were the atomic logP (AlogP), number of hydrogen bond donors (HBD) and hydrogen bond acceptors (HBA), molar refractivity (MR), molecular weight (MW), the molecular polarizability (Polar),²⁷ polar surface area (PSA), and number of rotatable bonds (RB). Properties were scaled to unit variance. Thereafter, a principal component analysis (PCA) was performed for each target separately to determine the first three principal components. Resulting scores were calculated for all test compounds and for the crystal structure ligand by projection into this three-dimensional space. Similarity to the crystal structure ligand was calculated as the Euclidean distance in the score space for all active and inactive compounds. Sorting by similarity was performed to calculate enrichment factors (eq 1, *vide infra*) and retrieval rates. The kernel density²⁸ plots were produced from the PCA described above using the first two

principal components. The python scipy.stats module was used to prepare the plots.²⁹

Chemotype Diversity among Active Compounds and Analogue Bias. MOLPRINT2D³⁰ 32-bit hashed fingerprints, using Mol2 atom types with all bonds treated as equivalent, were calculated in Canvas²⁶ for all tested compounds. The Tanimoto similarities to the ligand found in the crystal structure were then used for characterizing the risk of analogue bias in the data set compared to the ligand found in the X-ray structure. This was performed by ranking all tested compounds on the basis of the calculated Tanimoto distance, followed by calculations of enrichment factors (eq 1, *vide infra*) and retrieval rates.

Scaffold Diversity among Actives. To get a rough estimate of the number of scaffolds present among the actives, the number of unique Bemis–Murcko scaffolds³¹ were calculated using InstantJChem.²³ The ratio of this quantity and the number of active compounds will give an indication on any potential analogue bias in the data set using an arbitrary 2D fingerprint query.

Virtual Screening. Data Set Preparation. The compounds were prepared using LigPrep³² with Epik,^{33–35} generating tautomers, enantiomers, protonation states, and 3D structures. All compounds containing stereogenic centers were assumed to have been evaluated as stereoisomeric mixtures, and therefore, all combinations of stereoisomers were generated and only the best scoring isomer considered. If the inhibitor in the target protein was bound to a metal, metal binding states were generated in LigPrep.

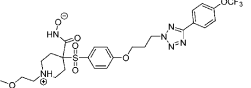
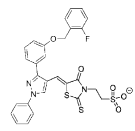
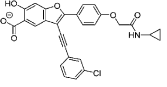
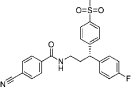
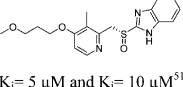
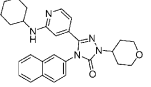
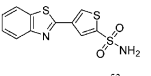
Crystal Structure Preparation. Crystal structures were downloaded from the PDB and prepared using Protein Preparation Wizard.^{4,36} For the PDB entries containing more than one subunit with an active site, chain A was used. Residues for which atoms were missing were left as is, and when prompted for residue alternate positions, the first possibility was always chosen. Hydrogen atoms were added and ionization and tautomeric states were generated using PROPKA³⁷ for protein and Epik for ligands. If a metal was present in the active site of the protein, metal binding states were included (PDB files: 3O2X, 3PGL). After optimization of the hydrogen bonding networks, water molecules with less than three hydrogen bonds to nonwaters were removed, and minimizations of the hydrogen atoms were performed. To achieve the final structures, restrained minimizations were performed with heavy atoms converged to a maximum of 0.3 Å RMSD.

Docking Studies. Grids for docking were prepared with Glide^{38–41} by selecting atoms in a 20 Å box surrounding the crystallized ligand. Dockings were performed using Glide in SP mode with default settings. Docking score was used to rank compounds.

3D Shape Similarity Screening. The 3D-shape screening was performed using ROCS.^{42,43} The crystal ligand was used as the query, set in the protonation state generated by the Protein Preparation Wizard (see Table 1). The structures were not minimized. The data sets prepared using LigPrep above were also used as input to OMEGA^{44,45} for pose generation prior to ROCS screening. Default settings were used for the runs and the compounds ranked according to their TanimotoCombo score (ROCS default).

Enrichment Calculations. To evaluate the success of each VS technique, we calculated enrichment factors (EF) according to eq 1.

Table 1. Cocrystallized Ligands Used for 1D, 2D, and 3D Screening and Reported Activity^a

target	co-crystallized ligand ^b	#X-ray structures ^c	PDB code	resolution (Å)
MMP13	 $K_i = 0.00013 \mu\text{M}^{47}$	72	3O2X ⁴⁷	1.90
DUSP3 ^d	 $\text{IC}_{50} = 0.074 \mu\text{M}^{48}$	2	3F81 ⁴⁸	1.90
PTPN22	 $\text{IC}_{50} = 0.26 \mu\text{M}^{49}$	2	4J51 ⁴⁹	2.30
EPHX2	 $\text{IC}_{50} = 0.006 \mu\text{M}^{50}$	15	3I28 ⁵⁰	1.95
CTDSP1 ^e	 $K_i = 5 \mu\text{M}$ and $K_i = 10 \mu\text{M}^{51}$	2	3PGL ⁵¹	2.35
MAPK10	 $\text{IC}_{50} = 0.016 \mu\text{M}^{52}$	31	3OY1 ⁵²	1.70
CDK5	 $\text{IC}_{50} = 0.551 \mu\text{M}^{53}$	5	4AU8 ⁵³	1.90

^aThe number of X-ray structures in the PDB containing the ligand, PDB identifier, and the resolution of the crystal structure used for docking. ^bThe ligand is drawn in the protonation state given by the protein preparation wizard. ^cNumber of crystal structures found with scripts using gi-numbers. ^dThe full structure is not visible in the crystal structure. The gray portion does not appear in crystal structure. ^eAmbiguous electron density for part of the ligand.

$$\text{EF} = \frac{\text{tp}/(\text{tp} + \text{fp})}{A/T} \quad (1)$$

In eq 1, tp is the number of true positives, fp the number of false positives, A the total number of active compounds in the data set, and T the total number of compounds in the data set. The Boltzmann-enhanced discrimination of the receiver operating characteristic (BEDROC) was also calculated, and we used $\alpha = 20$ for the calculations.⁴⁶ In the discussion of the results, the more common EFs were used in this study.

RESULTS

Data Set Compilation. The workflow for the data set compilation is shown in Figure 1. Searching the PubChem BioAssay database, 1256 bioassays were identified as being associated with a protein target and containing more than 1000 screened compounds (most of these correspond to primary

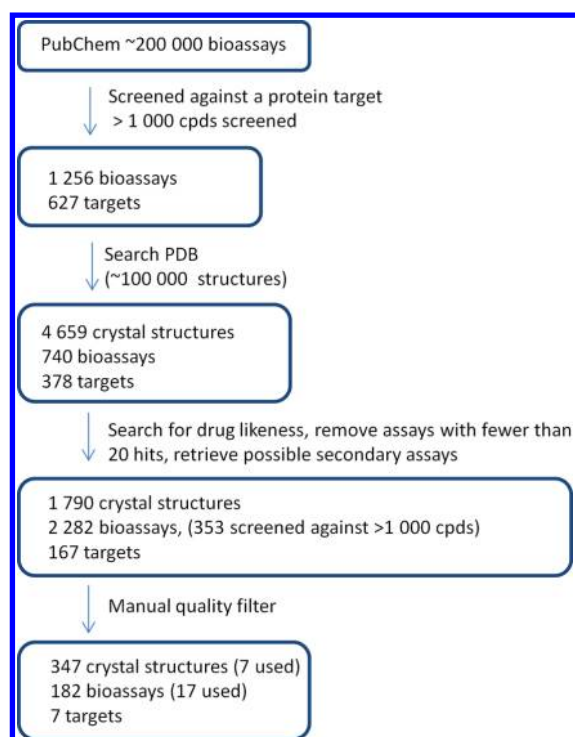


Figure 1. A graphical overview of the strategy used to mine data from PubChem and the PDB to derive the final data set.

assays with measurements at only one concentration). These 1256 bioassays had been performed against 627 unique protein targets. A total of 378 of these were linked to at least one crystal

structure in the PDB. For many targets, several crystal structures had been solved, giving a total of 4659 PDB structures. The 378 targets had 740 bioassay records associated with them.

In the next step, the 4659 crystal structures were filtered to identify PDB entries containing a drug-like ligand. Assays which identified less than 20 active compounds were removed. A total of 167 targets with 1790 associated PDB structures remained after this step. Data for all assays associated with these 167 targets were retrieved, 2282 bioassays in total.

After the manual quality filter step, only five targets with associated active and inactive compounds were identified passing our criteria: matrix metalloproteinase 13 (MMP13), dual specificity phosphatase 3 (DUSP3), protein tyrosine phosphatase nonreceptor type 22 (PTPN22), epoxide hydrolase 2—cytoplasmic (EPHX2), and carboxyl-terminal domain RNA polymerase II polypeptide A small phosphatase 1 isoform 1 (CTDSP1).

The exclusion of ATP binding proteins in general led to the disqualification of the entire kinome. These correspond to a well characterized class of drug targets, usually explored by ATP competitive inhibitors. Therefore, we went back and reanalyzed our hit list. As a result, two additional targets could be included in the benchmark data set: mitogen-activated protein kinase 10 (MAPK10/JNK3) and cyclin-dependent protein kinase 5 (CDK5).

Finally, the electron densities were examined for each selected crystal structure, focusing on the ligand and the ligand binding site. For all targets in the final data set, the electron densities were found to be satisfactory except for DUSP3 (PDB entry 3F81) and CTDSP1 (PDB entry 3PGL). In 3F81, a part

Table 2. Selected Targets with Their PubChem Assay Identifiers and Assay Readout, and the Number of Active and Inactive Compounds in the Final Data Sets

target	primary AID	confirmatory AID	PubChem classification		Hill coeff. filter ^c	PAINS filter ^d	# actives after filtration
			# actives (# of scaffolds) ^a	# inactives ^b			
MMP13	570 ^e	735 ^{e,f} 734 ^{e,g} 769 ^{e,h}	19 (17)	64 814	2	0	17
DUSP3 (VHR)	1654 ⁱ	1878 ⁱ	52 (41)	289 236	37	9	14
PTPN22 (LYP)	1779 ^j	435 024 ^j	28 (25)	292 058	12	4	13
EPHX2 (sEH)	707 ^j	1026 ^j	44 (36)	90 262	3	0	40
CTDSP1 (SCP1)	493 091 ⁱ	540 281 ^{i,k} 540 297 ^{i,l} 540 329 ^{m,n}	369 (218)	337 634	200	44	152
MAPK10 (JNK3)	746 ^o	1284 ^o	57 (38)	59 405	1	0	56
CDK5	488 839 ^p	504 545 ^{p,q} 504 546 ^{p,r}	23 (15)	334 339	0	0	23

^aThe number in parentheses is the number of unique Bemis–Murcko scaffolds found among the active compounds. ^bThe number of unique inactive compounds after Ligprep. ^cThe number of active compounds with a Hill coefficient < 0.5 or > 2.0. ^dThe number of active compounds flagged with PAINS filter. ^eAssay type FRET. ^fDose–response confirmatory screen. ^gSingle point assay searching for true positives among possible assay artifacts. ^hDose–response confirmatory screen among possible assay artifacts. ⁱAssay type absorbance. ^jAssay type fluorescence. ^kSingle point confirmatory screen. ^lDose–response confirmatory screen. ^mAssay type differential scanning fluorimetry. ⁿSingle point orthogonal confirmatory screen. ^oAssay type TF-FRET. ^pAssay type luminescence. ^qDose–response confirmatory screen. ^rCounter-screen used to remove possible false positive compounds.

Table 3. Virtual Screening Results for Each Target Showing the Number of Active Compounds Found at 100, 1000, and 10 000 Screened Compounds and the Enrichment Factor (EF) at 1% and 10% of the Database^a

target	method	first 100 ^b	first 1000 ^b	first 10 000 ^b	EF 1%	EF 10%	BEDROC alpha = 20
MMP13	docking	0	1	9 (8)	5.3	3.7	0.14
	3D similarity	0	2 (1)	4 (3)	11	1.0	0.10
	2D similarity	0	3 (2)	5 (4)	11	2.1	0.16
	1D similarity	0	0	2 (2)	0.0	1.1	0.02
DUSP3	docking	0	3 (3)	18 (16)	14	4.8	0.31
	3D similarity ^c	—	—	—	—	—	—
	2D similarity	0	1	5 (5)	5.8	1.9	0.12
	1D similarity	1	2 (2)	14 (12)	9.6	3.7	0.24
PTPN22	docking	0	4 (3)	7 (5)	21	4.3	0.30
	3D similarity	0	0	3 (3)	0.0	1.4	0.09
	2D similarity	0	1	3 (3)	3.6	3.2	0.14
	1D similarity	0	1	3 (2)	3.6	3.6	0.16
EPHX2	docking	0	2 (2)	25 (20)	4.5	5.0	0.26
	3D similarity	1	1	5 (5)	2.3	1.1	0.07
	2D similarity	1	2 (2)	12 (11)	2.3	2.7	0.16
	1D similarity	0	1	11 (10)	2.3	2.3	0.09
CTDSP1	docking	3 (3)	10 (10)	87 (58)	9.8	3.6	0.24
	3D similarity	0	2 (2)	6 (6)	0.8	0.5	0.03
	2D similarity	0	0	5 (4)	0.3	0.7	0.03
	1D similarity	0	1	11 (10)	1.1	1.0	0.05
MAPK10	docking	0	6 (5)	15 (12)	3.5	2.1	0.12
	3D similarity	0	1	6 (6)	1.8	0.7	0.03
	2D similarity	0	1	18 (13)	1.8	2.3	0.13
	1D similarity	0	4 (1)	5 (2)	7.0	0.9	0.08
CDK5	docking	0	1	3 (3)	8.7	4.8	0.22
	3D similarity	0	0	0	0.0	0.4	0.01
	2D similarity	0	0	0	0.0	0.0	0.004
	1D similarity	0	0	3 (2)	4.3	2.6	0.14

^aBoldface numbers indicate highest EF and BEDROC values. ^bThe number in parenthesis is the number of Bemis–Murcko scaffolds found among the hits. ^cLigand density of X-ray structure not completely defined, and part of the ligand structure is missing. Target omitted from 3D similarity screening.

of the ligand structure is missing. Since no other structure with a drug-like molecule is available for this target in the PDB, we decided to keep 3F81. This means that this target was omitted from 3D-shape screening since no accurate bioactive conformation of the ligand was available. In 3PGL, the poor density makes the position and orientation of the ligand ambiguous. However, since the ligand is present in the structure, we still used it in the data set characterization. The final data sets together with the VS results are presented in Tables 1 and 2.

The data sets show a high ratio of the number of inactive to active compounds. On average, over the seven data sets there are around 2500 inactive compounds to every active with a range from 1042 for the MAPK10 data set to 14 539 for the CDK5 data set. This is a major difference compared to many designed decoy sets where there are typically a lot fewer inactive compounds to every active. In DUD-E, there are 50 inactive compounds for each active compound, and in MUV there are 500 inactive compounds for each active compound. For each target, the bound crystal structure ligand is shown in Table 1.

Characterization of Data Sets and VS Performance.

We have used four different VS techniques to characterize the seven data sets: physicochemical similarity (1D similarity), 2D fingerprint similarity, 3D shape similarity, and docking. We have chosen commonly used software tools from each field. Since our goal was to investigate the properties and suitability of the selected data sets and not to optimize the enrichment or compare VS methods, default settings were used with the primary aim to provide a characterization of the data sets and the selected X-ray structures. The VS results are shown as retrieval rates, enrichment factors (EF), and BEDROC values in Table 3 and as Receiver Operating Characteristic (ROC) curves in Figure 2. Generally, docking was able to enrich active compounds, whereas 3D-shape similarity, physicochemical similarity, and 2D-fingerprint searches are more variable in the results.

Assessment of chemical diversity among the active compounds is shown in Table 2 as the number of Bemis–Murcko scaffolds found in each data set. As indicated by the ratio between the number of scaffolds and the number of active compounds, the diversity is high.

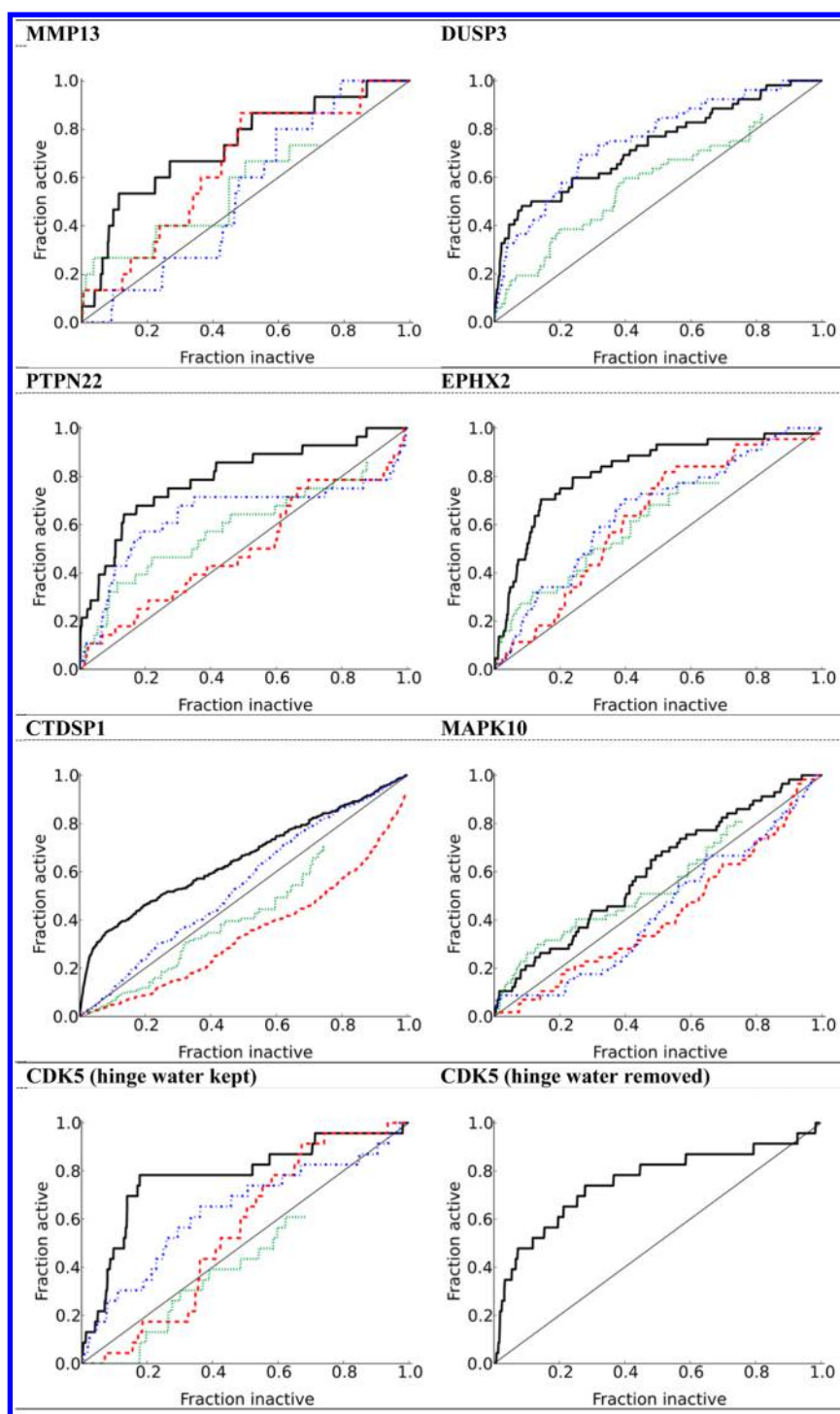


Figure 2. Receiver operating characteristic (ROC) curves showing the results for docking (black), 3D similarity (red), 2D similarity (green), 1D similarity (blue), and a thin black line indicating random performance. DUSP3 ligand density of the X-ray structure was not completely defined; therefore there are no ROCS results available.

To characterize the HTS data sets, a principal component analysis (PCA) of simple physicochemical descriptors calculated for each compound was performed for each of the seven data sets. In Figure 3, the distributions of inactive compounds in principal component space are shown as contour plots of kernel density estimates. This illustrates the distribution of actives (black dots) in relation to the much larger pool of inactive compounds in the 1D descriptor space. In general, the first principal component (PC1) corresponds to size and PC2

to polarity (the loadings for each plot are found in Figures S5–S11 in the SI). The left panels show a zoom-in on the active compounds, in all cases overlapping with the majority of the inactive compounds. The right panes also include the 100 top scored compounds from docking, highlighting that they in some targets are very dissimilar to the actives; in particular, they tend to be larger. The 1D enrichments shown in Table 3 and Figure 2 obviously depend on both the distribution of actives and inactive compounds in the 1D descriptor space and the

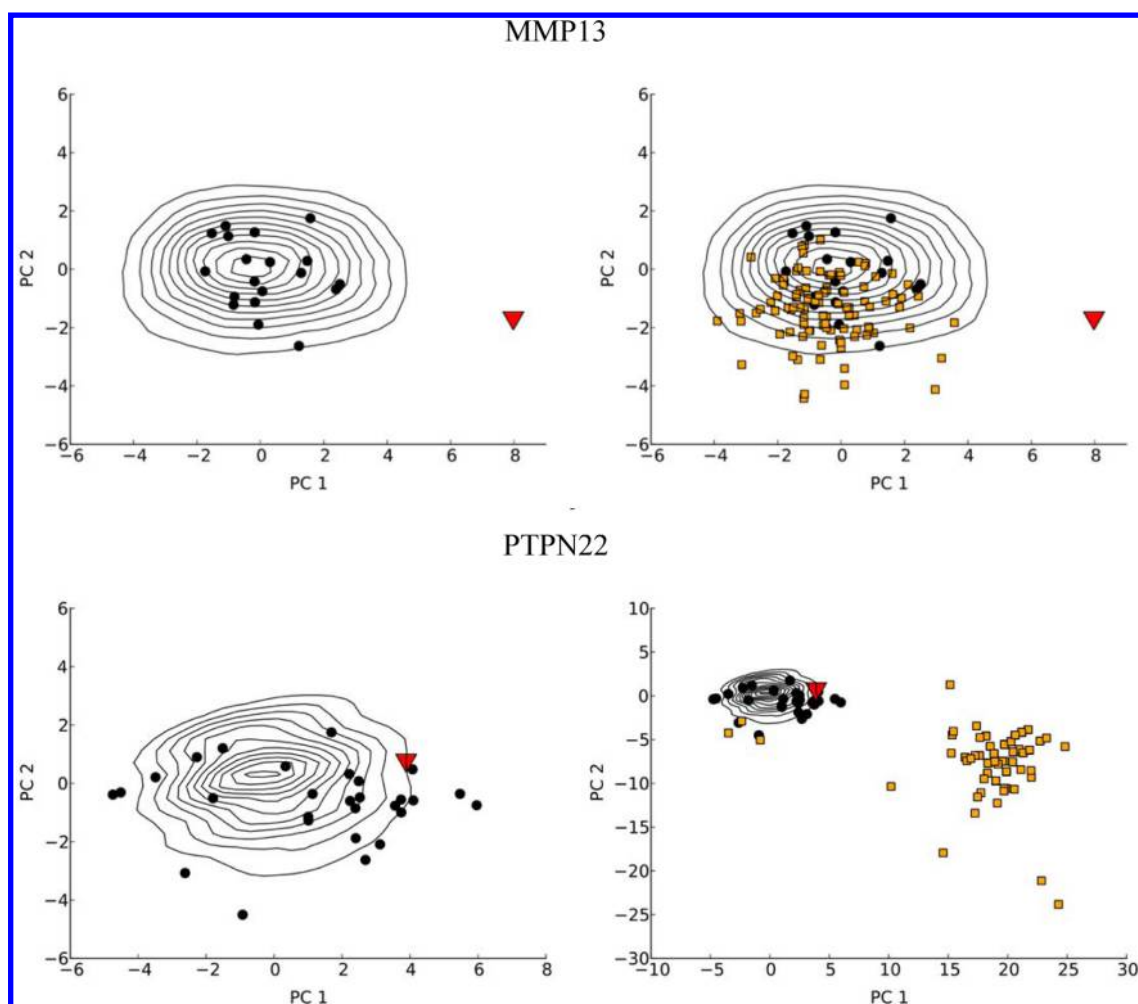


Figure 3. Example of contour plots of kernel density estimates for the data sets showing the scores from the first two principal components of the PCA of physicochemical descriptors. The active compounds are shown as black dots, crystal ligands as red triangles, and the 100 compounds with the best docking scores as orange squares. For clarity, each PCA plot is shown in two different scales with the left pane being a zoom-in on the actives.

distance to the crystal structure ligand (Figure 3, red triangles). As seen in the plots, the active compounds distribute over large areas, in most cases overlapping with the maximum density of inactive compounds. Therefore, any enrichment based on 1D physicochemical properties alone will likely be low. Box plots showing the distribution of MW, LogP, HBA, and HBD in the data sets can be found in Figures S1–S4 in the SI.

DISCUSSION

Current Benchmarking Data Sets. Several benchmarking data sets are available today and are frequently used for evaluation of VS methods. Attractive features of recent benchmarking sets include the large diversity of target classes and measures to prevent artificial enrichment and analogue bias. However, there are still some problems associated with many of these data sets. For example, the data on activity is used regardless of species; the active molecules are drawn from already optimized series, making the binding affinity unrealistically high for a screening scenario. The compounds have been evaluated in whole cell assays making the actual target uncertain. Active compounds may bind allosterically, and the decoy molecules have not been confirmed to be inactive. Specific designs of data sets may also render the data inappropriate for some virtual screening methods. For example,

the removal of decoys topologically similar to active compounds can make the data less challenging for ligand-based VS methods. In this work, we have tried to minimize the influence of these potential problems, and below we present how these data sets were compiled. We also describe their characterization as well as how they perform in VS.

Compilation of Data Sets. Despite the vast number of bioassay data in PubChem, the identification of data sets suitable for evaluation of both LB and SB VS methodologies is nontrivial. Without applying unreasonably rigorous data quality filters, we could only identify seven data sets that met the quality criteria set regarding both the bioassay data and the target 3D structure (Figure 1). The first major reduction in the number of suitable HTS data sets came from the requirement that the screen had to include more than 1000 compounds, a requirement that only 1256 of the 200 000 analyzed bioassays met. Thus, the vast majority of the PubChem data sets are simply too small to be considered for a HTS-based VS benchmark data set. The next applied filters removed 74% of the remaining potential HTS data sets. Either the data sets could not be connected to a PDB 3D structure with a cocrystallized drug-like ligand or the HTS had too few active compounds. The subsequent manual step provided the last quality filter. Some of the manually removed targets were

lacking a useful secondary assay. In some cases, the assay identified by the scripts was found to be a counter-screen for a campaign against another target. Multiplexed, cell based, and coupled assays where there may be multiple reasons for assay readouts have been excluded, and we chose to include only active compounds that could be verified in a confirmatory assay. When crystal structures included covalently bound or non-druglike ligands, they were excluded. Finally some assays were not used since they were part of ongoing projects.

One particularly difficult aspect to assess is the quality of the HTS data. This aspect of using PubChem Bioassay data has recently been discussed in detail by Blucher and McWeeney covering aspects of batch, plate, and positional variations as well as the general lack of metadata.⁵⁴ HTS data are often contaminated by artifacts originating from compounds prone to forming aggregates⁵⁵ or by compounds more directly interfering with the detection method. To address these problems, we have analyzed the compounds by means of a structural PAINS filter and a Hill slope filter (Table 2). These filters had very different effects on the different data sets. For CDKS, no compounds were found in either filter, whereas for CTDSP1 and DUSP3 the filtering on the Hill slope raised a warning flag for over half of the active compounds. We tested what would happen if we removed these compounds and calculated enrichment factors; see Table S1 in the SI. In the affected data sets, docking scores were generally improved, while the change in 1D, 2D, and 3D results was more mixed. As another example of unreliable HTS data, we found the cocrystallized ligand in 3PGL (CTDSP1) present in the HTS data set reported as inactive. The compound is reported to have a K_i of 5 μM ,⁵¹ an activity close to the cutoff value used in the primary screen which might explain why it was not flagged as active.

Another potential problem when classifying active and inactive compounds is that compounds able to inhibit the measured process may do so noncompetitively by binding to allosteric binding sites. This will severely compromise the results of VS, as the wrong binding site or wrong template ligand may be used for the screening. Therefore, we initially excluded targets binding a cofactor or more than one substrate. However, as the filtering process advanced, it became clear that some of the ATP binding targets in the PubChem BioAssay database would make a good contribution to the benchmark data set. Therefore, the two protein kinases (MAPK10, CDKS) that made it through the selection process were also included. Both these proteins have ATP as a cofactor and use proteins as substrates.⁵⁶ As protein–protein interactions are significantly more difficult to inhibit by small molecules compared to ATP binding, and because both these targets have been very well characterized in terms of ligand binding, they were included in the final data set. However, it should be noted that protein kinases in principle may bind inhibitors in either an enzymatically active form or in an inactive form, with large differences in the shape of the binding site and in the pharmacophore.

Data Set Characterization. Artificial enrichment and analogue bias are two potential problems in validation data sets. For example, a data set consisting of active compounds markedly different in physicochemical properties from the inactive compounds would be a different challenge to docking algorithms compared to a data set where active compounds and inactive compounds are more similar. In the case of DUD and MUV, the data sets are constructed in such a way that each

active compound is embedded in inactive compounds that have similar physicochemical properties. In this study, no measure was taken to eliminate any artificial enrichment resulting from skewed physicochemical distributions. Instead, all data sets were characterized by means of enrichments based on physicochemical similarity to the cocrystallized ligand and spread in physicochemical space. As suggested by Bender and Glen, it is beneficial to compare the results of 3D-ligand- or structure-based methods with the results of a physicochemical similarity method.⁸ Outperforming the physicochemical screening indicates that the applied methods' performance is not based on artificial enrichment. We found that the enrichment based on physicochemical similarity is lower than docking in most cases with a median EF at 1% of 3.6 (span 0 to 9.6). As seen in Figures 3 and S12, the ligand found in the X-ray structure is in many cases physicochemically different from the screened compounds and from the active compounds.

For several of the targets, the most highly scored compounds following docking are outliers in physicochemical space (Figures 3 and S12, right panel). This behavior has also been reported previously.⁵⁷ In several benchmarking data sets (DUD, DUD-E, MUV, DEKOIS), only decoys that have similar physicochemical properties as the active compounds are included. Therefore, many of the highly scored compounds would not be present in a designed data set. We believe these outliers can be important to include in a benchmarking data set in order to detect such bias.

Analogue bias results from too little diversity among the active compounds. This will lead to an unsatisfactory evaluation of methods using structure similarity as a discriminator as either none or all active compounds will be enriched. In the construction of this benchmark data set, all screening sets were characterized by 2D similarity searches using the ligand found in the crystal structure. For this task, MOL-PRINT2D^{58,59} was selected on the basis of the results from Sherman and co-workers, indicating that this is a good baseline fingerprint.³⁰ Compared to some previous studies of data sets for evaluating VS,^{30,60} the analogue bias in the current data sets is found to be comparably low with an average EF at 1% of 2.3. An exception is MMP13 that has an EF at 1% as high as 11 and possibly also DUSP3 that has an EF at 1% of 5.8. Moreover, the high ratio between the number of Bemis–Murcko scaffolds among the active compounds and the number of active compounds indicates that there is a substantial diversity among the hits (Table 2). Thus, analogue bias is not to be expected to be a major issue for the query structures used in the present study.

We envisage that the presented data sets are suitable for evaluation of both LB and SB methods. However, for each target, the enrichment obtained of a newly evaluated VS technique should be compared to both the 1D and 2D baselines of the ligands screened for that data set. In general, the docking method applied managed to outperform enrichments from 1D, 2D, and 3D ligand-based methods. This was not surprising considering that a ligand-based searching only considers the structure of one ligand, whereas docking makes use of the entire protein. However, our results clearly are not in line with the performance of LB methods using the DUD or DUD-E data set, where LB methods outperformed SB methods.^{61,62}

The contour plots in Figures 3 and S12 describe the distribution of test compounds in the physicochemical space. It is clear that the crystal structure ligand used to create the query

molecule for the LB methods in several cases is considerably different from both active and inactive compounds (e.g., MAPK10, PTPN22, MMP13, and DUSP3). Thus, it is not surprising that the performance of the LB methods is lower than for docking. In this study, the LB 3D method had very similar performance to the computationally less expensive 2D method.

Using default settings and ignoring any prior knowledge about the target might seem an unfair assessment of virtual screening tools. However, the intention with this approach was to set a baseline for shape screening and docking algorithms by minimizing the human bias going into the process. In general, the performance of the docking protocol indicates a clear gain in efficiency.

The performance of docking/scoring is ultimately dependent on starting from a crystal structure in a conformation resembling the conformation binding the HTS hits. As these usually are suboptimal in terms of interacting with the protein, selecting a high-affinity crystal structure ligand might result in a protein conformation different from the conformation binding low affinity HTS hits.⁶³ Another parameter of importance for the performance of docking algorithms is the treatment of water molecules. The most interesting example in this study is CDK5, where a water is found to bridge between the ligand and the hinge in the structure 4AU8.⁵³ Following our protocol for protein preparation, this water molecule was left intact, dramatically altering the size and polarity of the binding site. To evaluate the impact of this unconventional water treatment, we also performed the Glide docking with the water removed (Figure 2). Not unexpectedly, the early enrichment was improved, illustrating the importance of having prior knowledge about the target class when preparing the structure for the virtual screening.

The number of hits found among the first 100, 1000, and 10 000 compounds tested is included to give an indication of the performance of virtual screening employed for purchasing focused libraries. As seen in Table 3, the result is far from impressive. This study advises against selecting a library smaller than 1000 compounds unless prior knowledge can strengthen the trust in the virtual screening protocol selected.

CONCLUSIONS

We have mined the PubChem database for HTS data in order to create data sets suitable for evaluating both LB and SB VS methods. In total, seven data sets fulfilling our criteria were identified. All compounds annotated as active have been verified in a confirmatory assay. The assay types and target macromolecules were selected so that the included compounds have a high chance of acting on the target by binding to a defined binding pocket. Each data set has been linked to a corresponding target 3D structure in the PDB that also binds a drug-like molecule and is suitable for docking analysis. We propose that it would be valuable to supplement any docking evaluation using, for example, the DUD-E data sets with the seven targets described in this paper. Such an exercise would perhaps give a more realistic picture of enrichment factors and the number of active compounds that can be expected to be found in the first 100, 1000, or 10 000 compounds screened. Additionally, evaluation and optimization of virtual screening funnels and combined methods would particularly benefit from these larger data sets with a representative number of active compounds. The presence of physicochemically diverse decoys in these data sets is valuable as VS methods tend to

overestimate the score of some physicochemically extreme molecules.

As more data become available from PubChem and PDB, the use of HTS data for evaluation of VS has the potential to grow to cover more target classes and more diverse screening libraries.

Most of the problems associated with creating benchmarking data sets stem from the lack of reliable biochemical data. The aim of an HTS is to find one or a few active compounds that can be used as chemical probes or as chemical starting points in a drug discovery project. To create an even better and more reliable high-quality VS benchmark data set, the study protocol ought to therefore be designed differently. An improved protocol would include the use of orthogonal assays as well as a means to characterize the binding mode of the actives. Such data sets would be of great importance for validating and improving virtual screening tools and methods.

ASSOCIATED CONTENT

Supporting Information

Box plots of common descriptors. Loading plots for PCA calculations. VS results after PAINS and hill slope filtering. Python scripts used for data mining. Structures of all active compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>. The data sets used in this article are available upon request from the corresponding author.

AUTHOR INFORMATION

Corresponding Author

*E-mail: anders.karlen@orgfarm.uu.se.

Present Address

[†]Department of Chemistry, Umeå University Linnaeus väg 10, SE-901 87 Umeå, Sweden

Author Contributions

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

Funding

Carl Trygger Foundation

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.K. thanks the Carl Trygger Foundation. The authors also acknowledge Open Eye and ChemAxon for generously providing their software.

REFERENCES

- (1) Lill, M. Virtual Screening in Drug Design. *Methods Mol. Biol. (Clifton, N.J.)* **2013**, 993, 1–12.
- (2) Huang, S.-Y.; Grinter, S. Z.; Zou, X. Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. *Phys. Chem. Chem. Phys.* **2010**, 12, 12899–12908.
- (3) Svensson, F.; Karlén, A.; Sköld, C. Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods. *J. Chem. Inf. Model.* **2011**, 52, 225–232.
- (4) Sastry, G. M.; Inakollu, V. S. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, 53, 1531–1542.

- (5) Drwal, M. N.; Griffith, R. Combination of Ligand- and Structure-Based Methods in Virtual Screening. *Drug Discovery Today Technol.* **2013**, *10*, e395–e401.
- (6) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual Screening Using Protein-Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.
- (7) Good, A. C.; Hermsmeider, M. A.; Hindle, S. A. Measuring CAMD Technique Performance: A Virtual Screening Case Study in the Design of Validation Experiments. *J. Comput. Aided. Mol. Des.* **2004**, *18*, 529–536.
- (8) Bender, A.; Glen, R. C. A Discussion of Measures of Enrichment in Virtual Screening: Comparing the Information Content of Descriptors with Increasing Levels of Sophistication. *J. Chem. Inf. Model.* **2005**, *45*, 1369–1375.
- (9) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
- (10) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.
- (11) Vogel, S. M.; Bauer, M. R.; Boeckler, F. M. DEKOIS: Demanding Evaluation Kits for Objective in Silico Screening—a Versatile Tool for Benchmarking Docking Programs and Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2650–2665.
- (12) Rohrer, S. G.; Baumann, K. Maximum Unbiased Validation (MUV) Data Sets for Virtual Screening Based on PubChem Bioactivity Data. *J. Chem. Inf. Model.* **2009**, *49*, 169–184.
- (13) Irwin, J. Community Benchmarks for Virtual Screening. *J. Comput. Aided. Mol. Des.* **2008**, *22*, 193–199.
- (14) Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. *J. Chem. Inf. Model.* **2008**, *48*, 2108–2117.
- (15) Good, A. C.; Oprea, T. I. Optimization of CAMD Techniques 3. Virtual Screening Enrichment Studies: A Help or Hindrance in Tool Selection? *J. Comput. Aided. Mol. Des.* **2008**, *22*, 169–178.
- (16) Bauman, J. D.; Patel, D.; Dharia, C.; Fromer, M. W.; Ahmed, S.; Frenkel, Y.; Vijayan, R. S. K.; Eck, J. T.; Ho, W. C.; Das, K.; Shatkin, A. J.; Arnold, E. Detecting Allosteric Sites of HIV-1 Reverse Transcriptase by X-Ray Crystallographic Fragment Screening. *J. Med. Chem.* **2013**, *56*, 2738–2746.
- (17) Schierz, A. C. Virtual Screening of Bioassay Data. *J. Cheminform.* **2009**, *1*, 21.
- (18) Butkiewicz, M.; Lowe, E. W.; Mueller, R.; Mendenhall, J. L.; Teixeira, P. L.; Weaver, C. D.; Meiler, J. Benchmarking Ligand-Based Virtual High-Throughput Screening with the PubChem Database. *Molecules* **2013**, *18*, 735–756.
- (19) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H. PubChem's BioAssay Database. *Nucleic Acids Res.* **2012**, *40*, D400–12.
- (20) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (21) Kleywegt, G. J.; Harris, M. R.; Zou, J. Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallogr.* **2004**, *60*, 2240–2249.
- (22) Petrova, T.; Chuprina, A.; Parkesh, R.; Pushechnikov, A. Structural Enrichment of HTS Compounds from Available Commercial Libraries. *Med. Chem. Commun.* **2012**, *3*, 571.
- (23) *Instant JChem. 6.1*; ChemAxon: Budapest, Hungary, 2013.
- (24) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAIS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740.
- (25) Mok, N. Y.; Maxe, S.; Brenk, R. Locating Sweet Spots for Screening Hits and Evaluating Pan-Assay Interference Filters from the Performance Analysis of Two Lead-like Libraries. *J. Chem. Inf. Model.* **2013**, *534*–544.
- (26) *Canvas*, version 1.5; Schrödinger, LLC: New York, 2012.
- (27) Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533–8542.
- (28) McCabe, P.; Korb, O.; Cole, J. Kernel Density Estimation Applied to Bond Length, Bond Angle, and Torsion Angle Distributions. *J. Chem. Inf. Model.* **2014**, *54*, 1284–1288.
- (29) Jones, E.; Oliphant, T.; Peterson, P. *SciPy: Open Source Scientific Tools for Python*.
- (30) Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. Analysis and Comparison of 2D Fingerprints: Insights into Database Screening Performance Using Eight Fingerprint Methods. *J. Mol. Graphics Model.* **2010**, *29*, 157–170.
- (31) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
- (32) *LigPrep*, version 2.5; Schrödinger, LCC: New York, 2012.
- (33) *Epik*, version 2.3; Schrödinger, LCC: New York, 2012.
- (34) Shelley, J.; Cholleti, A.; Frye, L.; Greenwood, J.; Timlin, M.; Uchimaya, M. Epik: A Software Program for pK_a Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput. Aided. Mol. Des.* **2007**, *21*, 681–691.
- (35) Greenwood, J.; Calkins, D.; Sullivan, A.; Shelley, J. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput. Aided. Mol. Des.* **2010**, *24*, 591–604.
- (36) *Protein Preparation Wizard*; Schrödinger, LLC: New York, 2012.
- (37) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (38) *Glide*, version 5.8; Schrödinger, LCC: New York, 2013.
- (39) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (40) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (41) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (42) *ROCS*, version 3.1; OpenEye Scientific Software: Santa Fe, NM, 2011.
- (43) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74–82.
- (44) *OMEGA*, version 2.5; OpenEye Scientific Software: Santa Fe, NM, 2012.
- (45) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (46) 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.
- (47) Barta, T. E.; Becker, D. P.; Bedell, L. J.; Easton, A. M.; Hockerman, S. L.; Kiefer, J.; Munie, G. E.; Mathis, K. J.; Li, M. H.; Rico, J. G.; Villamil, C. I.; Williams, J. M. MMP-13 Selective A-Sulfone Hydroxamates: A Survey of P1' Heterocyclic Amide Isosteres. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2820–2822.
- (48) Wu, S.; Vossius, S.; Rahmouni, S.; Miletic, A. V.; Vang, T.; Vazquez-Rodriguez, J.; Cerignoli, F.; Arimura, Y.; Williams, S.; Hayes, T.; Moutschen, M.; Vasile, S.; Pellicchia, M.; Mustelin, T.; Tautz, L. Multidentate Small-Molecule Inhibitors of Vaccinia H1-Related

(VHR) Phosphatase Decrease Proliferation of Cervix Cancer Cells. *J. Med. Chem.* **2009**, *52*, 6716–6723.

(49) He, Y.; Liu, S.; Menon, A.; Stanford, S.; Oppong, E.; Gunawan, A. M.; Wu, L.; Wu, D. J.; Barrios, A. M.; Bottini, N.; Cato, A. C. B.; Zhang, Z.-Y. A Potent and Selective Small-Molecule Inhibitor for the Lymphoid-Specific Tyrosine Phosphatase (LYP), a Target Associated with Autoimmune Diseases. *J. Med. Chem.* **2013**, *56*, 4990–5008.

(50) Eldrup, A. B.; Soleymanzadeh, F.; Taylor, S. J.; Muegge, I.; Farrow, N. A.; Joseph, D.; McKellop, K.; Man, C. C.; Kukulka, A.; De Lombaert, S. Structure-Based Optimization of Arylamides as Inhibitors of Soluble Epoxide Hydrolase. *J. Med. Chem.* **2009**, *52*, 5880–5895.

(51) Zhang, M.; Cho, E.; Burstein, G. Selective Inactivation of a Human Neuronal Silencing Phosphatase by a Small Molecule Inhibitor. *ACS Chem. Biol.* **2011**, 511–519.

(52) Probst, G. D.; Bowers, S.; Sealy, J. M.; Truong, A. P.; Hom, R. K.; Galemme, R. a; Konradi, A. W.; Sham, H. L.; Quincy, D. a; Pan, H.; Yao, N.; Lin, M.; Tóth, G.; Artis, D. R.; Zmolek, W.; Wong, K.; Qin, A.; Lorentzen, C.; Nakamura, D. F.; Quinn, K. P.; Sauer, J.-M.; Powell, K.; Ruslim, L.; Wright, S.; Chereau, D.; Ren, Z.; Anderson, J. P.; Bard, F.; Yednock, T. a; Griswold-Prenner, I. Highly Selective c-Jun N-Terminal Kinase (JNK) 2 and 3 Inhibitors with in Vitro CNS-like Pharmacokinetic Properties Prevent Neurodegeneration. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 315–319.

(53) Malmström, J.; Viklund, J.; Slivo, C.; Costa, A.; Maudet, M.; Sandelin, C.; Hiller, G.; Olsson, L.-L.; Aagaard, A.; Geschwindner, S.; Xue, Y.; Vasänge, M. Synthesis and Structure-Activity Relationship of 4-(1,3-Benzothiazol-2-yl)-Thiophene-2-Sulfonamides as Cyclin-Dependent Kinase 5 (cdk5)/p25 Inhibitors. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5919–5923.

(54) Blucher, A. S.; McWeeney, S. K. Challenges in Secondary Analysis of High Throughput Screening Data. *Pac. Symp. Biocomput.* **2014**, 114–124.

(55) Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **2006**, *49*, 7274–7277.

(56) Arkin, M. R.; Wells, J. A. Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing towards the Dream. *Nat. Rev. Drug Discovery* **2004**, *3*, 301–317.

(57) García-Sosa, A. T.; Hetényi, C.; Maran, U. Drug Efficiency Indices for Improvement of Molecular Docking Scoring Functions. *J. Comput. Chem.* **2010**, *31*, 174–184.

(58) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708–1718.

(59) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* **2003**, *44*, 170–178.

(60) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of Topological, Shape, and Docking Methods in Virtual Screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519.

(61) Von Korff, M.; Freyss, J.; Sander, T. Comparison of Ligand- and Structure-Based Virtual Screening on the DUD Data Set. *J. Chem. Inf. Model.* **2009**, *49*, 209–231.

(62) Zhang, Y.; Yang, S.; Jiao, Y.; Liu, H.; Yuan, H.; Lu, S.; Ran, T.; Yao, S.; Ke, Z.; Xu, J.; Xiong, X.; Chen, Y.; Lu, T. An Integrated Virtual Screening Approach for VEGFR-2 Inhibitors. *J. Chem. Inf. Model.* **2013**, *53*, 3163–3177.

(63) Kuglstatter, A.; Wong, A.; Tsing, S.; Lee, S. W.; Lou, Y.; Villaseor, A. G.; Bradshaw, J. M.; Shaw, D.; Barnett, J. W.; Browner, M. F. Insights into the Conformational Flexibility of Bruton's Tyrosine Kinase from Multiple Ligand Complex Structures. *Protein Sci.* **2011**, *20*, 428–436.