FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set

Timothy J. Cheeseright,* Mark D. Mackey, James L. Melville, and Jeremy G. Vinter

Cresset BioMolecular Discovery Ltd., BioPark Hertfordshire, Broadwater Road, Welwyn Garden City, Hertfordshire AL7 3AX, United Kingdom

Received April 14, 2008

FieldScreen, a ligand-based Virtual Screening (VS) method, is described. Its use of 3D molecular fields makes it particularly suitable for scaffold hopping, and we have rigorously validated it for this purpose using a clustered version of the Directory of Useful Decoys (DUD). Using thirteen pharmaceutically relevant targets, we demonstrate that FieldScreen produces superior early chemotype enrichments, compared to DOCK. Additionally, hits retrieved by FieldScreen are consistently lower in molecular weight than those retrieved by docking. Where no X-ray protein structures are available, FieldScreen searches are more robust than docking into homology models or apo structures.

INTRODUCTION

Virtual Screening (VS) has become an established method for hit finding throughout the pharmaceutical industry.[1] Where X-ray protein data are available, a structure-based approach (e.g., docking) is usually followed, where the structure of both the protein and ligand is used to prioritize molecules for experimental screening. However, current docking scoring methods have known deficiencies in predicting binding affinity;[2] indeed, there is some pessimism about whether they can even correctly rank candidates.[3] Accounting for protein flexibility adds additional complexity.[4] Despite these issues, docking is usually seen as the preferred VS method if protein structures are available.

Where a protein structure is not available, ligand-based methods must be used. These generally require one or more known active molecules, against which the molecules to be screened are compared. Various methods for calculating molecular similarity are available,[5] but most represent a molecule as a vector with the entries indicating the presence or absence of features of the molecule. These methods are often fast, but where the features are based on simple molecular topology, they can be criticized for finding close structural analogues[6,7] rather than novel structures. Such "scaffold-hopping" behavior is often required to avoid intellectual property issues. For this purpose, alternative representations of molecules are used, focusing on the types of interactions made between a ligand and a protein. One example is the use of pharmacophore perception,[8] where molecules are scored based on their ability to match generic features such as hydrophobes and hydrogen bond donors/acceptors in 3D space. Other 3D methods include those which seek to maximize volume or shape overlap, such as ROCS.[9] Related methods, popular in deriving Quantitative Structure−Activity Relationships (QSAR), use field-based approaches, in particular the molecular electrostatic potential (MEP). The most famous of these is CoMFA.[10] However, despite the use of field-based methods in 3D QSAR, surprisingly little work has been carried out on the use of molecular fields in VS. Some examples include topomer field similarity,[11,12] MOLPRINT3D,[13] and SHOP.[14,15]

One possible reason why field-based approaches have not been widely adopted for VS is that, unlike atom-based or pharmacophore-based methods, balancing efficiency and accuracy with field-based scoring is challenging. However, recently[16] we described the creation and use of the local extrema of molecular interaction fields[17] (hereafter referred to simply as "field points") to describe biologically active molecules. We have shown that molecules that bind to the same biological target could be successfully aligned and scored using these field points[18] and demonstrated their application in scaffold hopping.[19] In this paper we present FieldScreen, which uses this field-based alignment and scoring method in a VS protocol.

For a VS method to be truly useful for scaffold hopping it is not enough for it to retrieve actives; those actives should be structurally novel and leadlike. To demonstrate this in a retrospective validation, two major considerations are the data set used and the measure of success. Care must be taken when assembling a data set that the actives and the decoys (presumed inactives from which the actives must be discriminated) have similar physicochemical properties, or separation of actives and decoys may be due to trivial differences in e.g. log*P* or molecular weight (MW), a phenomenon which has been observed for both docking[20,21] and ligand-based similarity searching.[22] Recently, Huang et al. generated the Directory of Useful Decoys (DUD) to avoid these problems.[23] This set represents an excellent starting place for assessing the efficiency of structure-based virtual screening protocols. Further refinements to this data set were proposed by Good and Oprea[24] to make the data set more leadlike by filtering based on log*P* and MW cutoffs. Additionally, they clustered the actives, so that the task becomes one of retrieving chemotypes, rather than individual structures. Both modifications result in a more realistic scenario for validating ligand-based VS for scaffold hopping,[7] and we have adopted them for the experiments reported herein.

* Corresponding author phone: +44 1707 356120; e-mail: tim@cresset-bmd.com.

10.1021/ci800110p CCC: $40.75
Published on Web 11/08/2008

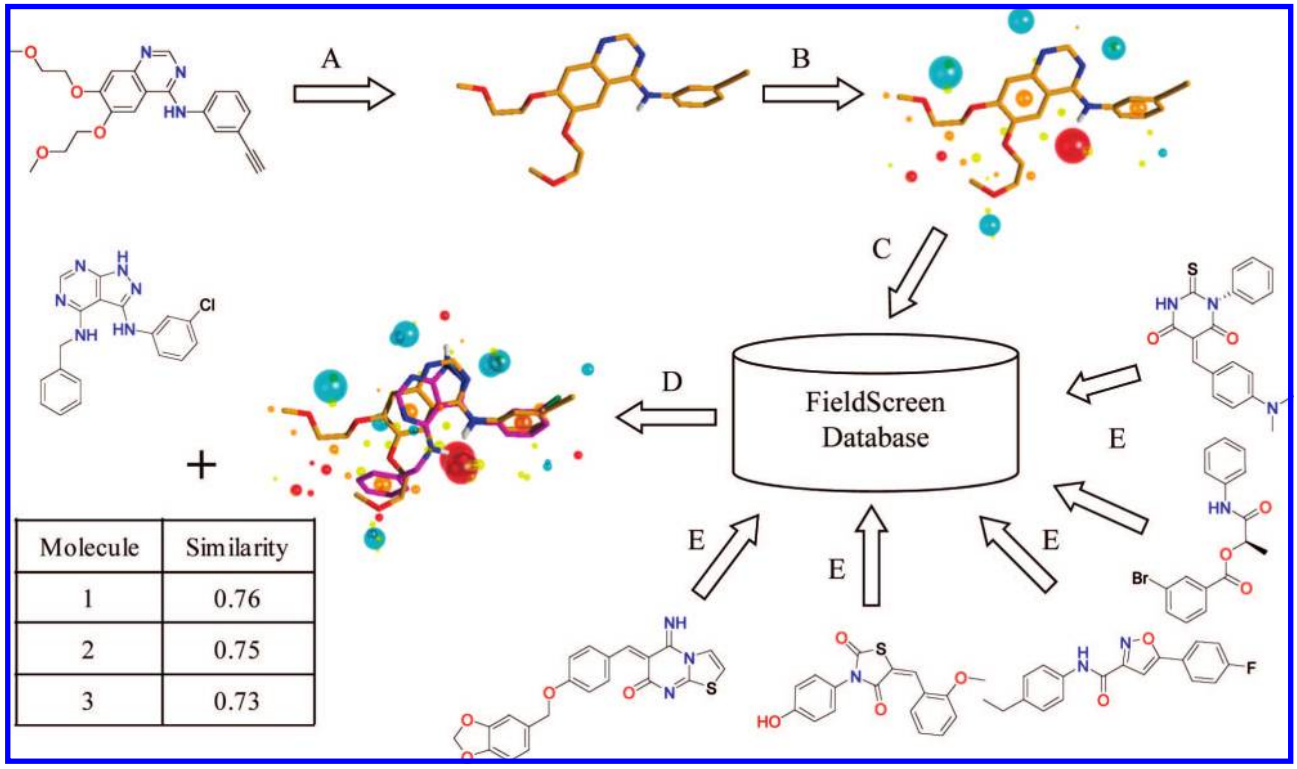

Figure 1. Schematic representation of the steps involved in searching the FieldScreen database: (A) select an active molecule and convert it to a relevant conformation; (B) add field points to the search ligand in the specified conformation to produce the FieldScreen search query, which consists of a ligand and its field points in a specified conformation; (C) search the FieldScreen database by alignment of every structure using field points; (D) retrieve the top scoring compounds (score expressed as a molecular similarity) as 3D alignments to the search query or as 2D structures. The FieldScreen database (E) is populated by exploration of conformations of all molecules with field point patterns added to and stored with each conformation.

The methods used to quantitatively assess the success of VS have also recently come under scrutiny.[25,26] The traditionally used enrichment factors[27] suffer from instability due to arbitrary cutoffs, and approaches using the Receiver Operating Characteristic (ROC) curve, which provides a measure of accuracy across the entire database, have become popular.[28] However, enrichment factors have the advantage of placing an emphasis on the retrieval of active over decoy compounds in the first few percent of the database, which is of greatest practical importance. Several modifications to the enrichment factor and ROC measures have been advocated to account for their perceived shortcomings, including RIE,[29] BEDROC,[30] and pROC.[31] However, consideration of the chemotype enrichment adds a further layer of complexity. Given the design of FieldScreen as a tool for finding diverse actives in as small a portion of a database as possible, we place emphasis on early enrichment, with a modification for chemotype retrieval.

METHODS

The process involved in performing a FieldScreen search is outlined in Figure 1. The first stage (A) requires the identification of an active ligand and conversion of this to a single 3D conformation. Since the main hypothesis of 3D ligand-based VS is that the binding properties of actives to a particular target are similar to that of a known active in the bioactive conformation, it is preferable to select this bioactive conformation as a search query. The FieldScreen search query is obtained from the addition of a field point pattern to the ligand in the chosen conformation (B). This query is compared to every conformation in the FieldScreen database (C), the results from an individual molecule being the best score obtained from comparison of all the molecule's conformations. The final results list (D) contains a similarity score for each molecule (allowing a flexible number of results to be taken forward to wet screening); a 3D alignment of the best scoring conformation (and, where relevant, protonation state) of each molecule and a 2D representation of each molecule. Populating a FieldScreen database (E) is achieved by conformation exploration (using our conformation hunter, XedeX[32]) and addition of field points to each conformation. In the sections that follow, we provide further details on these steps.

In terms of computational time, most VS methods use a filter approach, with more computationally expensive methods applied to fewer molecules later in the screening process. FieldScreen also offers this capability, and on a moderate-sized Linux cluster (20 CPUs), 1000000 structures can be screened within a day.

Conformation Search. Our alignment and scoring technique uses rigid conformations. Hence, each molecule must be conformationally explored prior to screening. Each conformation in a FieldScreen database is generated using XedeX, a conformation search algorithm based on the XED force field;[33] XedeX employs a Monte Carlo process to randomize all rotatable torsions. The resulting proto-conformations are then minimized, and the ensemble is filtered to remove closely related conformations and high energy structures. To avoid problems with structures adopting folded conformations due to the lack of explicit or implicit

solvation, attractive nonbonded forces are turned off during minimization. No more than 50 conformations are retained for each molecule.

Field Point Generation. Each molecule within FieldScreen is described by a set of four molecular fields.[16] The first three fields are defined as the energy of interaction of the molecule with a positive, negative, or neutral probe atom (with the steric properties of sp3-hybridized oxygen), with positive (unfavorable) interaction energies clipped to zero. A positive probe results in a field complementary to regions of negative charge on the ligand, and hence this is termed the "negative" field; a negative probe generates a "positive" field; a neutrally charged probe only accounts for steric interactions and generates a "surface" field. A fourth "hydrophobic" field is defined by an empirical density function with the following form

$$V_h = -K_h \sum_{j=1}^{n} \frac{E_{pj}}{r_{pj}^6} \quad (1)$$

where V_h is the hydrophobic score of the neutral probe p with the molecule containing n atoms, K_h is a constant, r_{pj} is the distance between the probe and the jth atom, and E_{pj} is a force field constant based on atom type, such that nonpolar atoms have large values, hydrogen atoms have a lower weighting, and electronegative groups have zero weighting. The effect of this potential is localize large hydrophobic field values within nonpolar regions of the molecule (e.g., at the center of aromatic rings).

Rather than use the full field around a molecule (by sampling on a grid, for example), we use the locally maximal field values for each conformation as our primary descriptor. These are located by simplex optimization from multiple starting points around the surface of the molecule. Each such local maximum is termed a "field point", and its position and size (the field intensity) are stored. Full details of the field equations and the field point generation process have been published previously.[16]

Field Alignment and Scoring. The detailed methods for the alignment of a conformation population to a search query were described in an earlier publication.[16] Briefly, a distance matrix of the field points for a conformation is generated and compared with that of the search query. Colored cliques are generated by finding sets of field points with similar interpoint distances and the same types which are present in both molecules. The clique is scored using the size of the clique and the size of the field points in the clique. Top scoring cliques are used to generate alignments of the molecules by fitting to the field points in the clique. The resulting alignment is scored by evaluating the field of one molecule at the positions of the field points of the other (all field points, not just those used to define the clique) and vice versa. This score is then normalized to a similarity value using the Dice similarity metric. The overlay is then further improved through a simplex optimization.

The clique process is extremely fast, taking a few milliseconds for each pair of conformations and 1−3 s for a typical search query against a conformation population. As a modification to the scoring, FieldScreen allows the use of "excluded volumes" to constrain the ligand based search. These volumes penalize atoms in any alignment that enter the excluded volume, reducing the overall similarity score. The most common use for this is to use a protein structure to define an excluded volume around its ligand, such that any alignment that has a steric clash to the protein is penalized. The excluded volumes used are typically both loose and soft: only alignments with significant overlaps with the protein structure will be penalized. We refer to this use of protein structure as an excluded volume in FieldScreen searches as 'FieldScreen+Protein' in the results below.

Data Sets. Release 2 of The Directory of Useful Decoys (DUD) was downloaded from http://dud.docking.org (accessed August 29, 2007) and contains known actives for 40 targets as well as decoys chosen to be physicochemically similar but topologically dissimilar to the actives. The DUD decoy data set comprises 95172 compounds enumerated as 102209 structures (with some molecules being represented with multiple tautomer and protonation states). Two molecules (ZINC03997305 and ZINC03997306) required minor editing to correct the drawn structures (protonation of an aromatic nitro group). Three molecules (ZINC03831927, ZINC04629420, and ZINC04633281) from the Factor Xa active set required editing due to incorrectly drawn amidino groups. After conformer enumeration using XedeX, each DUD decoy was represented by an average of 31 conformations. Two structures failed to produce any valid conformations: ZINC01292381 (which is drawn with an unlikely geometry) and ZINC04273472 (which contains a pentavalent nitrogen atom).

Because DUD was curated to test docking algorithms, it should not be used in validating ligand-based similarity methods without further processing.[34] Recently, Good carried out an analysis of the DUD actives and filtered them for lead-likeness (Alog$P$ < 4.5, MW < 450), followed by clustering based on a reduced graph representation.[24,35] We have used this filtered and clustered set of actives as a starting point for generating a subset of DUD suitable for our purposes. This helps avoid two problems with many retrospective screening analyses: retrieval of large molecules with unsuitable physicochemical properties and retrieval of trivial analogues. In order to restrict our analysis to data with a sufficient number of chemotypes, we treated each cluster as a separate chemotype and selected a subset of 13 targets from DUD, for each of which there were at least 15 clusters. For assessing the chemotype retrieval ability of each VS method, we used a subset of the "own" decoy sets described by Huang et al.[23] These contain only the decoys designed to be physicochemically similar and topologically dissimilar to the actives for a particular target. Each set was then filtered using the same AlogP and MW cutoff used by Good to filter the actives. For the purposes of assessing the MW of the actives retrieved by each VS method, we combined all the decoy sets available and did not apply any property filters. We call this the "all" decoy set. Details of the targets and number of actives, clusters, and decoys are provided in Table 1.

Search Queries. Unlike docking, ligand-based VS methods require a search query, based on known bioactive molecules. For 3D methods, ideally the bioactive conformation (or as close as can be deduced) should also be known. In FieldScreen, the combination of a conformation with its field points is the search query. Changes to the nature of the ligand, the conformation of the ligand, or the field point pattern derived from the conformation can all have an effect on the search process. For direct comparison with the DOCK

Table 1. Selected DUD Data Set Used in This Study As Modified by Good[24] Showing the Number of Decoys and the Number of Clusters of Actives for Each Target[a]

target	number of clusters	number of actives[b]	number of decoys	number of decoys per active	PDB code of search query
ace	18	46	1721	37.4	1o86
ache	18	100	3659	36.6	1eve
cdk2	32	47	1754	37.3	1ckp
cox2	44	212	11240	53.0	1cx2
egfr	40	365	14220	39.0	1m17
fxa	19	64	1817	28.4	1f0r
hivrt	17	34	1388	40.8	1rt1
inha	23	57	2411	42.3	1p44
p38	20	137	5554	40.5	1kv2
pde5	22	26	1506	57.9	1xp0
pdgfrb	22	124	5041	40.7	1t46[c]
src	21	98	4912	50.1	2src
vegfr2	31	48	2396	49.9	1fgi[d]

[a] Abbreviations: ace, angiotensin-converting enzyme; ache, acetylcholinesterase; cdk2, cyclindependent kinase 2; egfr, epidermal growth factor receptor; fxa, factor Xa; hivrt, HIV reverse transcriptase; inha, enoyl ACP reductase; p38, P38 mitogen activated protein; pde5, phosphodiesterase 5; pdgfrb, platelet derived growth factor receptor kinase; src, tyrosine kinase SRC; vegfr2, vascular endothelial growth factor receptor. [b] Where an active is represented by multiple tautomeric or protomeric forms, it is only counted once. [c] Protein structure was a homology model; the search query is gleevec, the ligand for the template, c-Kit kinase. The ligand coordinates are taken from DUD, not the crystal structure. [d] Protein is apo structure (PDB code 1vr2); search query is SU5402, ligand for the homologous FGFr1 kinase. The ligand coordinates are taken from DUD, not the crystal structure.

results, we used the ligand Huang et al.[23] used for defining the active site for DOCK (the "DUD ligand"). In most cases, the conformation of the search molecule was determined by the X-ray coordinates of the ligand-protein complex. However, in the case of pdgfrb, the ligand was taken from a homology model based on c-Kit kinase-gleevec complex. For vegfr2, an apo structure, the ligand was based on the FGFr1 kinase-SU4502 complex. While neither query has an experimentally determined bioactive conformation determined for the target, it should be noted that both are known active inhibitors for the targets. Search queries were prepared using Accelrys DSVisualizer 1.6 and 1.7 by downloading the appropriate ligand-protein crystal structure from the PDB, removing all water molecules, manually correcting the bonding pattern in the ligand, and adding hydrogen atoms in positions suggested by the H-bonding network (adding formal charges as appropriate). The ligand 3D coordinates were exported as an SDF file, the empty protein as a mol2 file.

Analysis of Virtual Screening Performance. For a qualitative assessment of virtual screening, enrichment curves provide a useful visual guide. For quantitative purposes, we use two measures. To assess early enrichment, we use the ROC Enrichment (ROCE),[25] defined as the ratio of the true positive rate to the false positive rate, for a given proportion of the known decoys having been observed. This is similar to the Enrichment Factor metric but does not show a dependence on the ratio of actives to decoys. We report ROCE values at 0.5%, 1%, 2%, and 5% of the true decoys having been screened, as recommended by Nicholls and Jain.[36] Additionally, we report the Area Under the Curve of the ROC (AUC), which provides a measure of the performance of the VS methods across the entire ranked database. Because we are interested in chemotype retrieval, we make use of the arithmetic weighting modification, as derived for the AUC by Clark and Webster-Clark.[31] This weights the contribution of each active to the AUC by an amount inversely proportional to the size of the population of the chemotype of which it is a member. Given the close relationship between AUC and ROCE, this modification can also be trivially applied to the ROCE, and we report arithmetic weighted AUC (awAUC) and ROCE (awROCE). Additionally, we provide the uncorrected AUC and ROCE values in the Supporting Information.

Error Estimation. In order to estimate the error associated with ROCE and AUC values we used a bootstrapping procedure. For each result set, 20% of the data was randomly removed, and the metrics were recalculated. This was repeated 10000 times: the standard deviations of these bootstrapped values obtained are taken as the error in the value calculated for the entire data set.

Comparison against Docking. To give a meaningful comparison of FieldScreen to established VS methods we have reanalyzed the DOCK energy scores available at the DUD Web site, http://dud.docking.org/r2/energies (accessed August 25, 2007), using the same measures we applied to analyzing FieldScreen. We removed from the results list any molecule which was not successfully processed by both DOCK and FieldScreen, so that exactly the same actives and decoys were assessed for all methods.

RESULTS

The performance of FieldScreen on the selected DUD data sets is presented in the enrichment graphs of Figure 2, along with the equivalent analysis of the DOCK results. Note that these graphs present the results obtained with the arithmetic weighting analysis method, i.e. each active is given a weight inversely proportional to the size of the cluster that it belongs to.

Arithmetic weighted ROCE values at 0.5%, 1%, 2%, and 5% are shown in the first two columns of Tables 2−5 and show similar trends for all four cutoffs. For illustrative purposes, values for ROCE@1% are shown in Figure 3. For each of the four cutoffs, the Wilcoxon matched-pairs signed-rank test indicates a significant difference between the performance of DOCK and FieldScreen ($p < 0.05$). In Figure 4 and Table 6 we show the awAUC values, evaluating chemotype enrichment across the entire ranked database. Values below 0.5, indicating subrandom enrichment, are not shown. Across all targets, there is no statistically significant difference between DOCK and FieldScreen. However, FieldScreen demonstrates performance well above random for the majority of targets, and for some targets the difference between DOCK and FieldScreen is substantial (e.g., cox2, egfr, inha, pde5).

The effect of using protein structure as excluded volume is also considered in the third column of Tables 2−6 and in Figures 3 and 4. Performance is clearly target dependent, and, overall, no statistically significant difference between the use of protein excluded volume and its absence in FieldScreen was found. For some targets, we observe a meaningful increase in retrieval, and, notably, the only case where the inclusion of protein information shows a marked

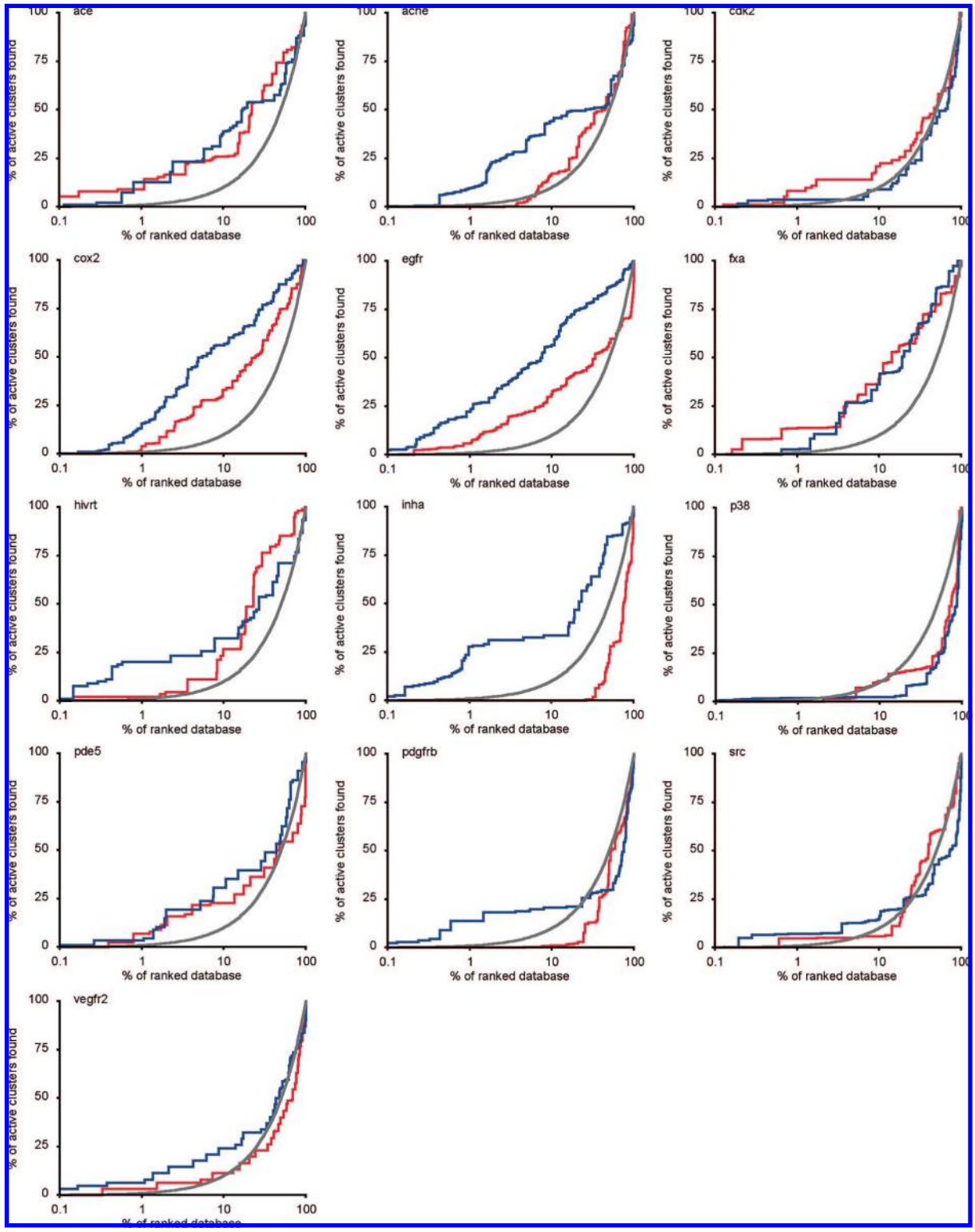

Figure 2. Chemotype enrichment plots for DOCK (red line), FieldScreen (blue line), and random performance (gray line) for 13 targets from the DUD data set. For FieldScreen the search ligands are specified in Table 1.

decrease in early retrieval (as measured by the ROCE@1%) is for pdgfrb, where the protein structure used is a homology model.

An important consideration in the assessment of any VS method is the nature of the molecules that are retrieved. Therefore we calculated the average MW of the top scoring 500 compounds that were retrieved from the "all" decoys data set by FieldScreen and performed the same analysis for the DOCK results. Results are shown in Figure 5. For each query we plot the average MW (with standard deviation) against the MW of the search query. Figure 5 shows that FieldScreen consistently returns molecules with a lower MW than DOCK ($p < 0.003$, Wilcoxon matched pairs signed-rank test), generally in the range 300−370 Da. Additionally, there is only a very slight increase in the MW of the molecules as the MW of the search query increases. For most ligand-based methods (e.g., structural fingerprints) it would be expected that the size of the search query would have a

Table 2. awROC Enrichment at 0.5% Values for the Three Virtual Screening Methods, DOCK, Across the Thirteen DUD Targets

target	DOCK	FieldScreen	FieldScreen+Protein[a]
ace	17.9	14.7	12.6
ache	0.0	16.7	17.3
cdk2	4.2	7.5	7.5
cox2	1.9	48.8	62.8
egfr	7.6	52.4	52.2
fxa	27.0	0.0	10.5
hivrt	4.4	40.0	40.0
inha	0.0	56.7	59.8
p38	0.0	3.7	3.7
pde5	4.5	6.8	6.8
pdgfrb	0.0	27.3	4.5
src	0.0	13.7	13.7
vegfr2	6.5	12.9	12.9

[a] FieldScreen results using a protein structure as an excluded volume.

Table 3. awROC Enrichment at 1% Values for the Three Virtual Screening Methods Across the Thirteen DUD Targets

target	DOCK	FieldScreen	FieldScreen+Protein[a]
ace	14.2	12.6	11.6
ache	0.0	20.4	21.9
cdk2	9.9	3.8	3.8
cox2	5.3	29.5	40.3
egfr	7.6	29.5	32.7
fxa	13.7	2.8	18.4
hivrt	2.2	20.0	20.0
inha	0.0	31.2	30.5
p38	0.0	1.8	1.8
pde5	6.8	4.5	5.7
pdgfrb	0.0	13.6	3.0
src	4.8	7.0	7.0
vegfr2	3.2	8.1	14.5

[a] FieldScreen results using a protein structure as an excluded volume.

Table 4. awROC Enrichment at 2% Values for the Three Virtual Screening Methods Across the Thirteen DUD Targets

target	DOCK	FieldScreen	FieldScreen+Protein[a]
ace	8.4	8.9	6.0
ache	0.0	13.5	13.3
cdk2	7.0	1.9	1.9
cox2	7.1	17.8	25.4
egfr	7.0	18.1	20.5
fxa	7.0	5.4	10.5
hivrt	2.2	11.7	11.7
inha	0.0	15.6	16.0
p38	0.0	0.9	1.2
pde5	8.0	9.7	5.1
pdgfrb	0.0	9.1	1.5
src	2.6	3.7	6.2
vegfr2	3.2	7.3	7.3

[a] FieldScreen results using a protein structure as an excluded volume.

Table 5. awROC Enrichment at 5% Values for the Three Virtual Screening Methods Across the Thirteen DUD Targets

target	DOCK	FieldScreen	FieldScreen+Protein[a]
ace	4.6	4.7	4.1
ache	0.8	7.3	7.3
cdk2	2.8	0.8	1.5
cox2	5.5	10.4	11.3
egfr	4.5	9.5	10.5
fxa	6.2	5.4	7.1
hivrt	2.2	5.1	6.4
inha	0.0	6.5	6.6
p38	0.4	0.5	0.5
pde5	4.3	4.8	4.3
pdgfrb	0.0	3.8	0.9
src	1.0	2.5	2.6
vegfr2	1.3	3.5	3.9

[a] FieldScreen results using a protein structure as an excluded volume.

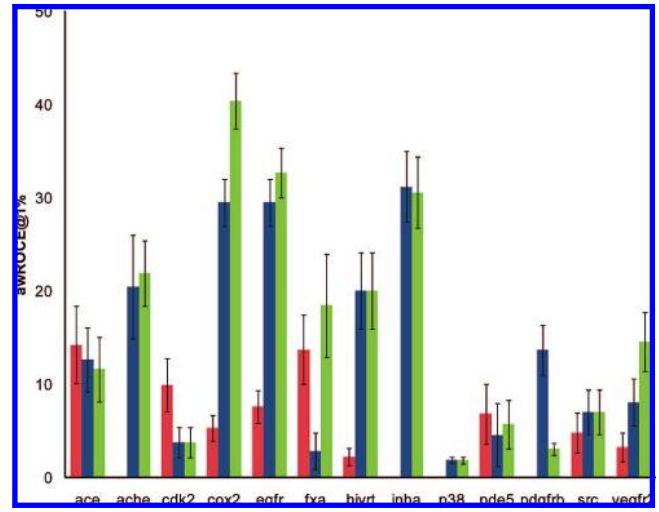

Figure 3. ROC enrichments at 1% with an arithmetic-weighted chemotype correction, for DOCK (red), FieldScreen (blue), and FieldScreen including excluded volume data (green) for each target.

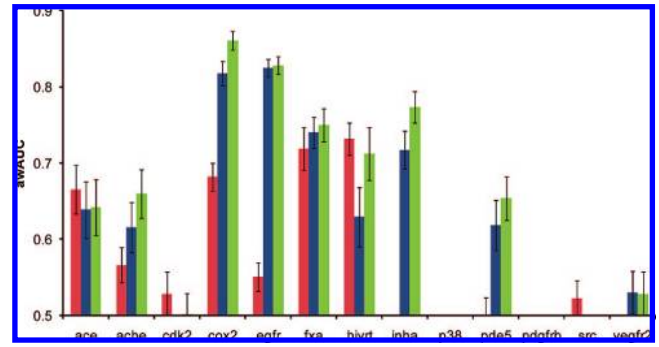

Figure 4. Area Under the ROC Curve values, with an arithmetic-weighted chemotype correction, for DOCK (red), FieldScreen (blue), and FieldScreen including excluded volume data (green) for each target. Worse-than-random performance (awAUC < 0.5) is not shown.

major influence on the size of the returned compounds. The horizontal line in Figure 5 indicates the average MW, 372 Da, of the "all" decoys. For all but the two heaviest queries, the average MW of the molecules retrieved by FieldScreen in the top 500 are below this value, suggesting that FieldScreen does not simply favor molecules of average MW for a given data set. The relative independence of FieldScreen results on the size of the search ligand can be rationalized by considering the complexity of the molecules that are present in a database and the probability of a large molecule having a high similarity to a search query. Larger molecules tend to have more field points, which leads to more opportunities for the fields of the database molecule and the search query to be mismatched, leading to lower similarities. Hence, the probability that two large (e.g., 500 Da) molecules

Table 6. awAUC Values for the Three Virtual Screening Methods Across the Thirteen DUD Targets

target	DOCK	FieldScreen	FieldScreen+Protein[a]
ace	0.67	0.64	0.64
ache	0.57	0.62	0.66
cdk2	0.53	0.44	0.50
cox2	0.68	0.82	0.86
egfr	0.55	0.82	0.83
fxa	0.72	0.74	0.75
hivrt	0.73	0.63	0.71
inha	0.26	0.72	0.77
p38	0.35	0.27	0.26
pde5	0.48	0.62	0.65
pdgfrb	0.39	0.40	0.40
src	0.52	0.39	0.46
vegfr2	0.42	0.53	0.53

[a] FieldScreen results using a protein structure as an excluded volume.

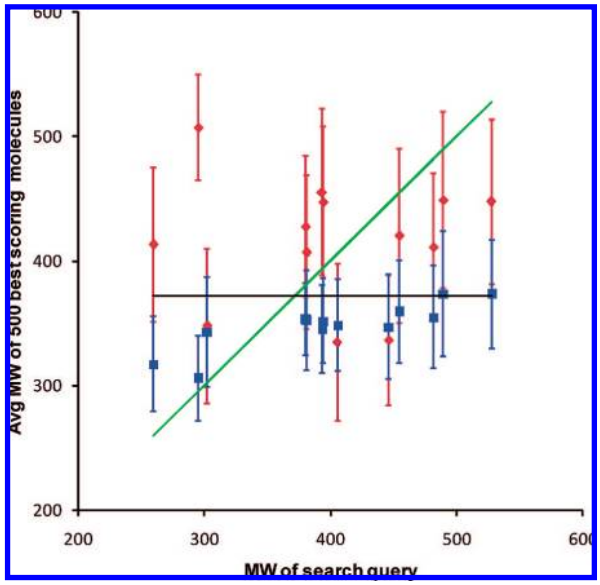

Figure 5. Average molecular weight (MW) of the top scoring 500 compounds in the "all" decoys data set plotted against the MW of the search query for DOCK results (red) and FieldScreen searches (blue). The green trend line shows the expected average MW of retrieved queries, if the virtual screening method retrieved molecules with a MW equal to that of the search query. The black horizontal line represents the average MW of the entire data set.

have mutually matching features is smaller than the probability of a smaller compound (e.g., 370 Da) matching two-thirds of the features of a large query. The fact that FieldScreen preferentially returns smaller hits is of great utility in virtual screening, as these are more likely to be amenable to hit-to-lead development.

DISCUSSION

To further elucidate the different aspects of FieldScreen's VS performance, we have chosen six targets to discuss in greater depth.

ache. The acetylcholinesterase results for FieldScreen are fairly typical and show a significant enrichment. The ROC enrichment values are also substantially above the scores for DOCK, which fails to perform above random at any of the assessed cutoffs. The average MW of the first 500 compounds retrieved using the 1eve ligand (MW 350) is 353 Da showing a significant advantage over DOCK (428 Da). The standard deviation in the average MW of the FieldScreen results is also smaller than for DOCK (29 Da versus 57 Da). These MW results are typical of most searches with FieldScreen consistently retrieving smaller molecules than DOCK with less variation in the size.

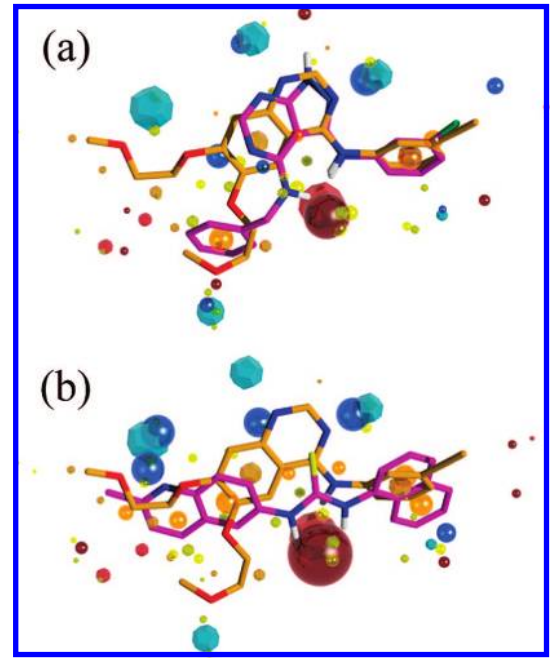

Figure 6. Alignments obtained from searching for egfr actives showing the search query (orange, dodecahedra used to show field points) aligned to the retrieved molecule (magenta, spheres used to show field points). (a) Active retrieved at position 96 in the search of the "all" decoy set. (b) Decoy false positive found at position 42 of the search of "all" decoy set.

egfr. Epidermal growth factor receptor inhibitors were retrieved by FieldScreen with excellent efficiency. This is surprising given that the search query contains two methoxyethyl ether groups that appear to have little interaction with the protein and possess high temperature factors. Interestingly, removal of these flexible chains resulted in a slightly lower enrichment (data not shown) reflecting the fact that many of the actives seem to have solvating groups that align to this region in FieldScreen. The average MW of the first 500 compounds retrieved was 345 Da, over 100 Da less than those retrieved by DOCK (455 Da). This reflects the solvent exposed binding site of egfr, where the binding site definition for DOCK can tolerate large molecules. Conversely, a FieldScreen search uses the shape of the ligand to effectively constrain the size of the retrieved molecules. Visual inspection of the top scoring alignments not only indicates some excellent overlays, even of relatively diverse active molecules (Figure 6a), but also gives an indication of how the search could be improved further. Figure 6b shows an alignment of a high scoring false positive containing a thiourea group. There is a good match across the bottom (as viewed) part of the inhibitor. However, the N1 of the 4-aminoquinazoline of the search query is known to make a strong interaction with an H-bond donor in the protein, and a matching interaction is not present in the false positive. Adding a constraint to the search query to force matches to this field region would be expected to significantly improve the results but is beyond the scope of the automated procedure we employed for this study.

fxa. The retrieval of Factor Xa actives by FieldScreen yields only moderate enrichments. However, the addition of the protein structure as an excluded volume results in large improvements. The majority of fxa inhibitors are large flexible molecules that bind in elongated conformations. As FieldScreen is a rigid fitting method the performance may well be improved by more in-depth conformational sampling. It is worth noting that the fxa decoy set is not well matched to the actives in terms of charge. 793 of 5689 fxa decoys are positively charged, while 62 of 67 actives are positively charged. A simple rank on positive charge count thus performs very well (62 actives in the top 855 compounds). Field-based methods are potentially susceptible to domination by charged molecules, as the electrostatic field around a formally charged molecule in vacuo is much larger than that around an uncharged molecule. Although the electrostatic fields used in FieldScreen contain a correction for this,[16] there is the question as to whether the enrichments we obtain for fxa are due to simple physicochemical reasons. However, we note that the FieldScreen search retrieves some of the uncharged actives relatively early: e.g. ZINC03815573 is the sixth active retrieved at position 56 in the database. Therefore, the enrichments we achieve for this target, although moderate, are not artifactual. FieldScreen is capable of finding uncharged mimics for charged molecules, which is often an important goal in VS.

p38. Neither DOCK nor FieldScreen perform well in retrieving actives for p38 and fail to perform above random in ranking the entire data set, although FieldScreen results show above random chemotype ROC enrichment at 0.5% and 1%. This was a surprising result for us, because we have carried out a successful prospective search for p38 inhibitors using FieldScreen.[37] The DUD ligand for p38 (from pdb code 1kv2) is large (MW >500) and bound to the "DFG out" conformation of the protein.[38] A ligand bound to the "closed" form of the protein would be more suitable, as most published p38 inhibitors bind to the closed form. To test this, we repeated the FieldScreen search, using the ligand from pdb code 1ouk. This pyridinylimidazole ligand binds across the hinge region of p38 and extends down through the ribose-binding region toward the phosphate-binding pocket. With this ligand, we obtained much improved results (awAUC = 0.74, awROCE@1% = 29.4). The use of a more relevant crystal structure would therefore probably also improve the DOCK results.

pdgfrb. For pdgfrb, a homology model was used as a target for the original DOCK results and as the excluded volume in our FieldScreen search. Consequently, the DUD ligand we used for the search query in FieldScreen does not have crystallographically determined coordinates. The early enrichments obtained by FieldScreen were satisfactory (awROCE ranged from 4 to 27, depending on the cutoff), whereas DOCK failed to find any actives at the 5% cutoff. Interestingly, the pdgfrb search was the only one that suffered a significant drop in enrichment on addition of the protein excluded volume. Visual examination of the pdgfrb homology model-ligand complex shows a highly enclosed binding site in which the ligand makes numerous favorable contacts with the protein. It is entirely reasonable to suppose that relatively small errors in the position of one or more protein amino acid side chains in the homology model could cause clashes with ligands, a factor that is more likely considering the enclosed nature of the binding site. Thus adding the excluded volume to the FieldScreen search query penalizes actives at the expense of smaller decoys. Similar arguments apply to the poor performance of DOCK for this target.

vegfr2. The protein structure used for vegfr2 was taken from X-ray data of the apo form. In contrast to pdgfrb, the addition of excluded volume information to the vegfr2 search query has no deleterious effect on enrichment, and DOCK performance is also somewhat improved. However, FieldScreen returns molecules that are significantly smaller than does DOCK (306 Da versus 507 Da on average) and close in size to the search query, 295 Da. Visual inspection of the apo structure reveals a possible reason for the large difference between the methods. A large open cleft between two lobes of the protein is defined as the active site with the search ligand residing at one end of the cleft. It is reasonable to believe that the semiautomated docking employed by Huang et al.[23] produced an elongated active site definition that did not exclude longer molecules that fit across a larger portion of the cleft. Favorable docking scores are thus obtained for large molecules. In contrast, the ligand-based method has a much smaller definition of the "active site" (i.e., the size of the search ligand), and hence larger molecules are correctly down-weighted. Recent protein−ligand crystal structures of vegfr2[39] show that this extended cleft may not be present in the complexed form of the protein, and therefore using one of these structures may significantly improve the DOCK results.

cdk2. DOCK outperforms FieldScreen on cdk2, and, unlike the case of fxa, the addition of protein excluded volume does not improve FieldScreen's performance. An investigation of the effect of protein structure on the success of cdk2 inhibitor docking by Thomas and co-workers[40] indicates that using a cdk2 protein structure with a large binding site volume maximizes the probability of docking success. In 1ckp, residues 153−163 of the T-loop are disordered, and hence their coordinates are not present. Additionally, examination of the relative positions of Lys33, Leu38, and Asp145 in 1ckp indicates that the binding site has a large volume relative to other cdk2 crystal structures. Both of these features indicate that the binding site of 1ckp is conducive to docking success. Conversely, the MW of the DUD ligand is 260 Da, whereas the average MW of the cdk2 actives is 331 Da. This suggests that the search molecule used in FieldScreen is smaller than the actives we are attempting to retrieve, and therefore the query is underdefined. Examination of the known protein−ligand crystal structures of cdk2 with Relibase[41] shows that ligands can exhibit a variety of binding modes in the hinge region of cdk2 and are not limited to the relatively small area defined by the ligand in 1ckp. Thus the use of a larger search query could reasonably be expected to improve the FieldScreen results. Nonetheless, cdk2 remains a challenging target for ligand-based approaches due to the large number of different binding modes known.

CONCLUSIONS

FieldScreen is a practical, high performance ligand-based virtual screening system. To validate FieldScreen, we have used a clustered and filtered subset of DUD, which ameliorates several problems with publically available data sets,

including issues of lead-likeness and structural homogeneity. In conjunction with bootstrapped chemotype-corrected metrics as a measure of enrichment, this provides a rigorous challenge for FieldScreen as a scaffold-hopping VS method. Taken as a whole, the results are extremely encouraging, with FieldScreen demonstrating a statistically significant superiority in early chemotype retrieval, compared to DOCK. Additionally, the top scoring molecules from FieldScreen have a smaller MW than those obtained from docking, suggesting that FieldScreen returns more leadlike hits.

The results we report here demonstrate that 3D ligand-based virtual screening can deliver state-of-the-art performance under conditions where docking has hitherto been first choice as well as when no crystal structural data are available, as others have also noted.[42] Exemplifying this point is worse than random early enrichments of DOCK with pdgfrb, a protein structure based on homology modeling. Conversely, FieldScreen achieved genuine enrichments. Similarly, while DOCK achieves above-random early enrichments in the case of vegfr2, where the protein structure used for docking was an apo form, FieldScreen was able to improve on these enrichments. These results indicate that field-based VS could be integrated with chemogenomics, using small molecule ligand queries from targets with known homology to the desired target.[43] Other VS scenarios, where FieldScreen is augmented by structure-based data, are possible. For example, where some docking/scoring regimes are too expensive to be carried out in a truly high throughput manner (e.g., Glide XP[44]) a possible solution is to use FieldScreen as an initial filter to reduce the search space to a few thousand molecules and then perform low-throughput scoring on this smaller data set.

Several optimizations of the VS protocol outlined here can be envisaged. As noted in the discussion of the results for the egfr target, the ability to alter the weighting given to different field types or applying constraints to ensure certain field points are matched would ease integration of any existing SAR in the definition of the search query. The use of multiple simultaneous search queries at least partially removes the limitation that any given query molecule is only likely to retrieve actives with a similar binding mode. These enhancements are already available in FieldScreen, and their effect will be the subject of future publications.

ACKNOWLEDGMENT

We thank Dr. Andy Good of Bristol Myers Squibb for early communication of the filtered list of decoys.

Supporting Information Available: Table of ROC AUC, ROC enrichments, and names and ranks of the actives for the results in Figures 3 and 4, and table of average MW of retrieved compounds from Figure 5. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) *Virtual Screening in Drug Discovery*. Alvarez, J., Shoichet, B., Eds.; CRC Press: Boca Raton, 2005.
(2) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(3) Tirado-Rives, J.; Jorgensen, W. L. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein-Ligand Binding. *J. Med. Chem.* **2006**, *49*, 5880–5884.
(4) Sutherland, J. J.; Nandigam, R. K.; Erickson, J. A.; Vieth, M. Lessons in Molecular Recognition. 2. Assessing and Improving Cross-Docking Accuracy. *J. Chem. Inf. Model.* **2007**, *47*, 2293–2302.
(5) Bender, A.; Glen, R. C. Molecular Similarity: A Key Technique in Molecular Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
(6) Flower, D. R. On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379–386.
(7) Good, A. C.; Hermsmeier, 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) Kirchmair, J.; Ristic, S.; Eder, K.; Markt, P.; Wolber, G.; Laggner, C.; Langer, T. J. Fast and Efficient in Silico 3D Screening: Toward Maximum Computational Efficiency of Pharmacophore-Based and Shape-Based Approaches. *J. Chem. Inf. Model.* **2006**, *47*, 2182–2196.
(9) Rush, T. S., III; Grant, J. A.; Mosyak, L.; Nicholls, A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* **2005**, *48*, 1489–1495.
(10) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(11) Jilek, R. J.; Cramer, R. D. Topomers: A Validated Protocol for their Self-Consistent Generation. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1221–1227.
(12) Cramer, R. D.; Jilek, R. J.; Guessregen, S.; Clark, S. J.; Wendt, B.; Clark, R. D. "Lead Hopping". Validation of Topomer Similarity as a Superior Predictor of Similar Biological Activities. *J. Med. Chem.* **2004**, *47*, 6777–6791.
(13) Bender, A.; Mussa, H. Y.; Gill, G. S.; Glen, R. C. Molecular Surface Point Environments for Virtual Screening and the Elucidation of Binding Patterns (MOLPRINT 3D). *J. Med. Chem.* **2004**, *47*, 6569–6583.
(14) Ahlström, M. M.; Ridderström, M.; Luthman, K.; Zamora, I. Virtual Screening and Scaffold Hopping Based on GRID Molecular Interaction Fields. *J. Chem. Inf. Model.* **2005**, *45*, 1313–1323.
(15) Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold HOPping by GRID-Based Similarity Searches. *J. Med. Chem.* **2007**, *50*, 2708–2717.
(16) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.* **2006**, *46*, 665–676.
(17) Apaya, R. P.; Lucchese, B.; Price, S. L.; Vinter, J. G. The matching of electrostatic extrema: A useful method in drug design? A study of phosphodiesterase III inhibitors. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 33–43.
(18) Low, C. M. R.; Vinter, J. G. Rationalizing the Activities of Diverse Cholecystokinin 2 Receptor Antagonists Using Molecular Field Points. *J. Med. Chem.* **2008**, *51*, 565–573.
(19) Low, C. M. R.; Buck, I. M.; Cooke, T.; Cushnir, J. R.; Kalindjian, S. B.; Kotecha, A.; Pether, M. J.; Shankley, N. P.; Vinter, J. G.; Wright, L. Scaffold Hopping with Molecular Field Points: Identification of a Cholecystokinin-2 (CCK2) Receptor Pharmacophore and Its Use in the Design of a Prototypical Series of Pyrrole- and Imidazole-Based CCK2 Antagonists. *J. Med. Chem.* **2005**, *48*, 6790–6802.
(20) Pan, Y. P.; Huang, N.; Cho, S.; MacKerell, A. D. Consideration of Molecular Weight during Compound Selection in Virtual Target-Based Database Screening. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 267–272.
(21) 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.
(22) 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.
(23) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
(24) 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.
(25) Nicholls, A. What do we know and when do we know it. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239–255.
(26) Jain, A. N. Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 201–212.
(27) Kearsley, S. K.; Sallamack, S.; Fluder, E. M.; Andose, J. D.; Mosley, R. T.; Sheridan, R. P. Chemical similarity using physiochemical property descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 118−127.

(28) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547.

(29) Sheridan, R. P.; Singh, S. B.; Fluder, E. M.; Kearsley, S. K. Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1395–1406.

(30) 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.

(31) Clark, R. D.; Webster-Clark, D. J. Managing Bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 141–146.

(32) XedeX. http://www.cresset-bmd.co.uk/xedex.shtml (accessed Aug 29, 2007).

(33) Vinter, J. G. Extended electron distributions applied to the molecular mechanics of intermolecular interactions. II Organic Complexes. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 417–426.

(34) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193–199.

(35) Directory of Useful Decoys. http://dud.docking.org/clusters/summary.pdf (accessed Aug 29, 2007).

(36) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.

(37) Cheeseright, T. J.; Melville, J. L.; Holm, M.; Lehmann, F.; Luik, S.; Göttert, M.; Laufer, S. Novel lead structures for p38 MAP kinase via FieldScreen virtual screening. Manuscript in preparation.

(38) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268–72.

(39) Harris, P. A.; Cheung, M.; Hunter, R. N.; Brown, M. L.; Veal, J. M.; Nolte, R. T.; Wang, L.; Liu, W.; Crosby, R. M.; Johnson, J. H.; Epperly, A. H.; Kumar, R.; Luttrell, D. K.; Stafford, J. A. Discovery and evaluation of 2-anilino-5-aryloxazoles as a novel class of VEGFR2 kinase inhibitors. *J. Med. Chem.* **2005**, *48*, 1610–1619.

(40) Thomas, M. P.; McInnes, C.; Fischer, P. M. Protein Structures in Virtual Screening: A Case Study with CDK2. *J. Med. Chem.* **2006**, *49*, 92–104.

(41) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.

(42) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74–82.

(43) Martin, R. E.; Green, L. G.; Guba, W.; Kratochwil, N.; Christ, A. Discovery of the First Nonpeptidic, Small-Molecule, Highly Selective Somatostatin Receptor Subtype 5 Antagonists: A Chemogenomics Approach. *J. Med. Chem.* **2007**, *50*, 6291–6294.

(44) Zhou, Z.; Felts, A. K.; Friesner, R. A.; Levy, R. M. Comparative Performance of Several Flexible Docking Programs and Scoring Functions: Enrichment Studies for a Diverse Set of Pharmaceutically Relevant Targets. *J. Chem. Inf. Model* **2007**, *47*, 1599–1608.

CI800110P