LigMatch: A Multiple Structure-Based Ligand Matching Method for 3D Virtual Screening

Sarah L. Kinnings and Richard M. Jackson*

Institute of Molecular and Cellular Biology and Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

We have developed a new virtual screening (VS) method called LigMatch and evaluated its performance on 13 protein targets using a filtered and clustered version of the directory of useful decoys (DUD). The method uses 3D structural comparison to a crystallographically determined ligand in a bioactive 'template' conformation, using a geometric hashing method, in order to prioritize each database compound. We show that LigMatch outperforms several other widely used VS methods on the 13 DUD targets. We go on to demonstrate that improved VS performance can be gained from using multiple, structurally diverse templates rather than a single template ligand for a particular protein target. In this case, a 2D fingerprint-based method is used to select a ligand template from a set of known bioactive conformations. Furthermore, we show that LigMatch performs well even in the absence of 2D similarity to the template ligands, thereby demonstrating its robustness with respect to purely 2D methods and its potential for scaffold hopping.

INTRODUCTION

Virtual screening (VS) is used to rank databases of chemical compounds according to decreasing probability of activity against a particular target. This prioritization means that only those molecules with significant probabilities of activity are screened experimentally, therefore decreasing the costs associated with the drug discovery process.[1] With a number of success stories already reported,[2] VS methods are increasingly playing an important role in drug discovery.[3] When a crystal structure of the target protein is available, a structure-based approach will commonly be used.[4] Structure-based methods often involve docking small molecules into the structure of the target protein and measuring the resulting interactions.[3] However, it is well-known that current docking scoring methods are generally poor predictors of binding affinity,[5] and even their ability to correctly rank candidates has been questioned.[6] When the structure of the target protein is unavailable, a ligand-based approach must be followed. Ligand-based approaches require one or more known active molecules against which the database compounds can be compared.[4] There are numerous examples in the literature of where ligand-based methods have been shown to outperform docking algorithms.[7−10] Alternatively, when the structure of the target protein in complex with a ligand is available, the 3D bioactive conformation of the ligand will commonly be used as the 'template', i.e., reference molecule against which the database compounds are compared. Various different structure-based, ligand-centric methods often use the bioactive conformations of template ligands, including shape-matching algorithms,[11] molecular field descriptors,[12−14] and pharmacophore fingerprints.[15−17] Alternatively, low-energy conformations,[8] or even 2D representations,[18−20] of the known actives may be used as template ligands when no structural information is available.

One of the main challenges when using any ligand-centric approach is to find the appropriate balance between search specificity, which is required for target selectivity, and search flexibility, which is necessary for removing any dependency on the template ligand.[3] Indeed, ligand-based methods, particularly those based on simple molecular topology, are often criticized for finding close structural analogues instead of discovering novel structures, a concept known as 'scaffold hopping'. Scaffold hopping is a highly desirable outcome of VS because it not only provides ways of improving the pharmacological properties of known lead compounds, but it also allows the exploration of unpatented regions of chemical space. VS methods are only truly useful for scaffold hopping if they are able to retrieve actives that are both lead-like and structurally novel, and a major consideration when demonstrating this in a retrospective validation is the data set of actives and decoys used. It is important that the decoys have similar physicochemical properties to the actives, otherwise the actives may simply be distinguished from the decoys based on trivial differences, such as molecular weight or logP.[4] Huang et al.,[21] recently constructed the directory of useful decoys (DUD) in which the decoys were specifically selected to be physicochemically similar yet topologically dissimilar to the actives, thereby providing a more realistic scenario for the validation of ligand-based VS. In this study, we use a version of the DUD that has been modified by Good and Oprea[22] to ensure its suitability for benchmarking ligand-centric VS methods.

The vast majority of similarity methods reported to date have made use of only a single bioactive template ligand. However, by focusing on only a single bioactive molecule, it is likely that actives with dissimilar structures will be overlooked, leading to a high false-negative rate.[23] With the increasing availability of crystal structures of multiple,

* Corresponding author. Telephone: +44 (0)113 343 2592. Fax: +44 (0)113 343 3167. E-mail: r.m.jackson@leeds.ac.uk.

STRUCTURE-BASED LIGAND MATCHING METHOD

J. Chem. Inf. Model., Vol. 49, No. 9, 2009 **2057**

structurally diverse, bioactive ligands for a given protein target, there has been recent interest in the use of multiple template ligands in VS.[1] Indeed, fingerprint searching is often more effective when multiple template ligands are used due to an increase in the amount of information available for the calculations.[24,25] Accordingly, several different approaches have been applied to multiple template-based fingerprint searching, including consensus fingerprints,[26] centroid fingerprints,[24,25,27] bit-scaling techniques[28] and nearest-neighbor methods.[24,25,28−31] A number of studies comparing these different search strategies have shown that the nearest-neighbor methods and the centroid methods often give the best performance.[27,29] Consensus fingerprints, centroid fingerprints, and bit-scaling techniques emphasize bit positions that are shared by all template ligands and are, therefore, likely to account for their observed bioactivity.[25] Although they are powerful methods, they may be of less use when applied to the structurally diverse molecules that are typical of high-throughput screening experiments.[29] Indeed, they are more likely to rediscover the common features of known bioactive molecules rather than discover new chemotypes, which is a desirable feature of scaffold-hopping approaches. Conversely, the nearest-neighbor method individually calculates the similarity of each compound in the database to all of the template ligands, and then either the largest observed similarity score is selected to represent each database compound (MAX fusion rule) or the mean of the similarity scores is taken for the *k* nearest neighbors (SUM fusion rule).[25] Both Hert et al.[29] and Whittle et al.[31] showed that the best recovery of active compounds was achieved when ranking the database molecules according to the MAX fusion rule. In this study we have extended this 2D approach by using the Tanimoto coefficient and the MAX fusion rule to select a template ligand against which each database compound can be compared in terms of its 3D structure, using a geometric hashing method.[32]

Here we report a new VS method called LigMatch, which uses a geometric hashing method[32] in order to compare database compounds with the 3D bioactive conformation of a template ligand. We evaluate the performance of LigMatch on protein targets specifically selected from the DUD to allow a direct comparison with a recent study carried out by Cheeseright et al.[4] For the 13 DUD single template targets, we show that the performance of LigMatch is better for six of them than for any of the four other popular VS methods, i.e., FieldScreen,[4] DOCK,[4,21] ROCS,[33] and a 2D fingerprint-based method.[34,35] Subsequently, we show that performance is highly dependent on the choice of ligand used as the template, and we demonstrate that the use of a 2D method to select the template ligand from a set of multiple template ligands can improve the performance of LigMatch. Finally, we systematically remove similarity between the template ligands and the database molecules and show that LigMatch is able to perform well even in the absence of 2D similarity to the template ligands, thereby demonstrating its potential for scaffold hopping.

COMPUTATIONAL METHODS

Active and Decoy Molecules. The DUD (http://dud.docking.org/r2/, accessed 11/21/2008) provides a well-defined, unbiased set of active and decoy compounds that is publicly available for the validation of docking algorithms. It contains sets of known actives for 40 target proteins, and sets of decoys that have been specifically selected to be physicochemically similar yet topologically dissimilar to these actives. Five molecular descriptors (molecular weight, logP, number of hydrogen-bond acceptors, number of hydrogen-bond donors, and number of rotatable bonds) were used to select decoys with similar physicochemical properties to the actives, while topological dissimilarities were calculated based on fingerprints. The strict criteria used in the decoy selection process mean that the DUD provides a stringent test with which to evaluate VS performance. As one of the best and most comprehensive collections of actives and decoys, the DUD is currently considered to be the industry standard for the validation of docking algorithms.[23] However, since the DUD was designed as a benchmarking set for docking, it requires further processing to make it suitable for use in the validation of ligand-based methods.[4] Indeed, in its raw state, there are many targets for which nearly all actives are trivial analogues of a central structure, causing it to be an inappropriate benchmark for 3D molecular similarity methods.[36] In a recent study, Good and Oprea[22] filtered the DUD actives for lead-likeness (molecular weight (Mw) < 450 and AlogP < 4.5) before clustering them according to chemotype using a reduced graph representation. In this way, they not only removed large molecules with unsuitable physicochemical properties, but they also addressed the problem of analogue bias. This filtered and clustered set of actives was used in the validation of FieldScreen, a field-based method reported recently by Cheeseright et al.[4] In order to restrict their analyses to targets with a sufficient number of chemotypes, the authors selected a subset of 13 targets from the DUD, which each possessed at least 15 different clusters of actives. They then filtered the sets of decoys that corresponded to each of these targets using the same Mw and AlogP thresholds that were used by Good and Oprea when filtering the actives.

The sets of active and decoy molecules corresponding to each of the 13 selected target proteins (ace, ache, cdk2, cox2, egfr, fxa, hivrt, inha, p38, pde5, pdgfrb, src, and vegfr2) were downloaded directly from the DUD (http://dud.docking.org/r2/; DUD release 2, 10/22/2006) in mol2 file format. In order to allow a direct comparison with the results presented in the paper by Cheeseright et al., ranked lists of actives and decoys were obtained from the authors, and actives and decoys that were absent from their results were removed from our data set. It is worth noting that where multiple tautomeric forms of a molecule were present in the DUD, only a single tautomer was retained for our data set.

Conformer Generation. Since our geometric hashing alignment and scoring method uses rigid conformations, multiple conformers of each active and decoy molecule were generated using Openeye's OMEGA.[37] OMEGA, which uses a deterministic rule-based method for conformer searching, was chosen because it has been shown to be very reliable and effective at reproducing bioactive conformations.[38] An energy window of 25 kcal mol^{-1} was used since it was shown to be essential for best conformer generation performance in a study by Kirchmair et al.[39] Indeed, a lower energy threshold could result in valuable conformations being discarded, whereas a higher energy threshold may produce high-energy conformations that are not likely to represent

Table 1. Summary of the Data Sets Used in This Study

target	PDB code (and ligand code) of single template used[a]	no. of possible templates[b]	no. of actives (and active conformers)[c]	no. of decoys (and decoy conformers)[d]	no. of clusters of actives[e] (and no. of clusters represented by the possible templates)[f]
angiotensin-converting enzyme (ace)	1o86 (**LPR**)	6	46 (2 310)	1,705 (119 737)	19 (3)
acetylcholinesterase (ache)	1eve (**E20**)	26	100 (5 162)	3,652 (166 556)	19 (4)
cyclin-dependent kinase 2 (cdk2)	1ckp (**PVB**)	131	47 (1 525)	1,545 (54 966)	32 (14)
cyclooxygenase-2 (cox2)	1cx2 (**558**)	4	207 (1 675)	9,335 (336 371)	43 (4)
epidermal growth factor receptor (egfr)	1m17 (**AQ4**)	10	365 (10 980)	12,626 (526 281)	40 (1)
factor Xa (fxa)	1f0r (**815**)	71	63 (5 298)	1,809 (143 852)	19 (6)
HIV reverse transcriptase (hivrt)	1rt1 (**MKC**)	63	32 (605)	1,353 (37 853)	15 (9)
enoyl ACP reductase (inha)	1p44 (**GEQ**)	11	57 (1 696)	2,401 (84 357)	23 (1)
P38 mitogen activated protein (p38)	1kv2 (**B96**)	53	137 (2 545)	5,220 (141 564)	20 (4)
phosphodiesterase (pde5)	1xp0 (**VDN**)	8	26 (616)	1,504 (97 937)	22 (1)
platelet derived growth factor receptor kinase (pdgfrb)	1t46 (**STI**)[g]	n/a	120 (2 126)	4,529 (251 313)	22 (n/a)
tyrosine kinase SRC (src)	2src (**ANP**)	13	98 (1 395)	4,546 (312 785)	21 (1)
vascular endothelial growth factor receptor (vegfr2)	1fgi (**SU1**)[h]	16	48 (1 631)	2,376 (132 311)	31 (0)

[a] The crystallographic ligand used as the single template in our study. It is identical to those used in the studies by Huang et al. and Cheeseright et al. [b] The total number of different ligands retrieved from all crystal structures of the target protein present in the RCSB PDB. These ligands were used for the multiple template method, and include the ligand used in the single template method (except in the case of vegfr2). [c] The number of active molecules taken from the DUD, and the number of conformers of these actives generated using OMEGA. [d] The number of decoy molecules taken from the DUD, and the number of conformers of these decoys generated using OMEGA. [e] Good and Oprea assigned each active to a cluster based on chemotype, therefore allowing an arithmetic weighting procedure to be used to remove analogue bias from the results. [f] The most similar active to each template ligand (greater than or equal to a Tanimoto coefficient threshold of 0.85) was selected, and the total number of clusters to which the selected actives belonged was then calculated. [g] Since no crystal structure of pdgfrb was available, the ligand coordinates for the template were taken from a homology model based on a c-Kit kinase-gleevec complex. [h] Ligand coordinates of SU5402 (ligand code; SU1) from the closely related fgfr1 protein kinase (PDB code; 1fgi) were used as the template.

bioactive conformations. A diversity threshold (root-mean-square deviation, rmsd) of 1 Å was also used to discard similar conformations, and a maximum of 100 conformers were generated per molecule. These parameters were used in the work by Sperandio et al.[3] because they represented an optimal balance between speed and performance and agreed well with the study carried out by Kirchmair et al. For each target, a set of actives comprising of the original coordinates of the active molecules from the DUD, plus all generated conformers of these actives, was compiled. A set of decoys was compiled for each target in the same way. The numbers of actives and decoys for each target are given in Table 1.

Selection of Single-Template Ligands. For each target protein, we used the coordinates of the ligand used in the studies by Huang et al.[21] and subsequently by Cheeseright et al., which are derived from those in the RCSB Protein Data Bank (PDB)[40] but with added hydrogen atoms. Since no crystal structure of pdgfrb was available, the ligand coordinates for the template were taken from a homology model based on a c-Kit kinase-gleevec complex (PDB code, 1t46; ligand code, STI). For vegfr2, the ligand coordinates of SU5402 (ligand code, SU1) from the closely related fgfr1 protein kinase (PDB code, 1fgi) were used as the template. Details of each of the 13 templates used (referred to as the 'DUD ligands') are given in Table 1.

Scoring Using the Geometric Hashing Method. A geometric hashing method described previously[32] was used to compare all actives and decoys for a particular protein target with the template ligand. The geometric hashing method proceeds as follows: A series of possible triplets consisting of three atoms forming a triangle is generated for both the database molecule and the template ligand. All possible

triplets between the two molecules are then compared, and those with the same atom types at triangle vertices (all atoms of the same element are treated as equivalent) and similar interatomic distances are defined as a match. A least-squares fitting routine is then used to perform a transformation for each match, resulting in a translational and rotational matrix that maps one 3D triplet onto another. Each match is then given a score corresponding to the number of coincident atoms in the mapping, and only the highest score, known as the 'atom–atom score', is retained.

The atom–atom score was determined for every conformer of each active or decoy molecule, with both the best and the mean atom–atom score of all conformers being assigned to that particular molecule. The best and the mean atom–atom scores were then normalized to take into account the maximum possible number of matching atoms, by dividing by the number of atoms present in the active or decoy molecule. The resulting list of actives and decoys was subsequently ranked by both the descending normalized best score and descending normalized mean score. Since the normalized mean score gave slightly better performance overall (see the Supporting Information, Table SI 1), it was used in the studies reported herein. In terms of computational time, the geometric hashing method is able to process around 130 conformers per minute on a single processor, which equates to almost four million conformers in one day on a Linux cluster of 20 CPUs.

Scoring Using Other Methods. In order to provide a comparison with LigMatch, four other methods were compared, using exactly the same sets of actives and decoys for each target. Both the FieldScreen and the DOCK rankings were obtained from Cheeseright et al.[4,14] In addition, we ran ROCS and a 2D fingerprint-based method. ROCS is a shape-

STRUCTURE-BASED LIGAND MATCHING METHOD

J. Chem. Inf. Model., Vol. 49, No. 9, 2009 **2059**

based method that assesses the volume overlap of two molecules, using Gaussians parametrized according to the hard-sphere volume of heavy atoms.[11] We used the ROCS combo score (default) to rank the lists of actives and decoys. The combo score is the sum of the shape Tanimoto and the scaled color score, where the former is a measure of the shape similarity and the latter is a measure of the chemical match between two molecules. Finally, in order to compare the performance of LigMatch with a 2D fingerprint-based method, Tanimoto coefficients were used to prioritize the actives and the decoys for each target. Open Babel[34,35] was used to calculate the Tanimoto coefficients, using Daylight fingerprints,[41] between the sets of actives and decoys and their corresponding templates. The actives and decoys for each target were then ranked by descending Tanimoto coefficient.

Comparison of Methods on Single Templates. Since receiver operator characteristic (ROC) curve analysis is considered to be the best approach for characterizing the performance of VS methods,[23] ROC curves were used to compare the overall performance of the five different methods. An arithmetic weighting procedure[4,42] was applied in order to remove any analogue bias from the results. As mentioned previously, all active molecules for each target had been assigned by Good and Oprea to 15 or more different clusters, based on their chemotype (http://dud.docking.org/clusters/, accessed 01/14/2009). Each active molecule was given a weighting that was equal to the inverse of the size of the cluster to which it belonged. For instance, if five active molecules with similar chemotypes had been assigned to the same cluster, each active within that cluster would be given a weighting of one-fifth. Where there were actives that were absent from the Cheeseright et al. results (due to a failure to be processed by both FieldScreen and DOCK), the weightings of the remaining actives were modified accordingly. The number of clusters of actives for each target is given in Table 1.

LigMatch Multiple-Template Method. A protein sequence corresponding to the relevant chain of each of 12 different target proteins was downloaded from the RCSB PDB. Since the X-ray crystal structure of pdgfrb had not been solved at the time of writing, no further analysis was conducted on this target. A protein—protein BLAST[43] search with an e-value cutoff of e^{-40} and a sequence identity cutoff of 70% was carried out against the sequences of all proteins present in the RCSB PDB (http://www.rcsb.org/, accessed 09/16/2008 (kinases) and 02/10/2009 (nonkinases)). Protein structures with bound ligands were downloaded, and where there were multiple structures with the same ligand, only one representative structure was retained. Ligands were inspected to ensure that they were biologically relevant (resulting in the removal of small molecules, such as lipids and ions) and bound in the correct site, i.e., the same site as the DUD ligand. Ligands meeting these criteria were selected, so that for each target there was a set of different ligands with known crystallographic coordinates that would serve as templates against which actives and decoys could be scored. Note that the crystal structure of the DUD ligand was also included in each set of templates. The total number of templates for each target is given in Table 1 (for further details refer to the Supporting Information). Simply using

LigMatch to compare the active and decoy molecules to every individual template molecule for each target would be a time-consuming process, especially for those targets with a large number of templates, e.g., cdk2, which has a total of 131 templates. Therefore, for each active or decoy molecule, a single template was selected from all of the templates available for a particular target by using Open Babel to calculate Tanimoto coefficients between that active or decoy molecule and each of the available template ligands. In accordance with the MAX fusion rule, the template with the highest Tanimoto coefficient was chosen as the template against which that active or decoy molecule and all of its conformers would be scored using the geometric hashing method.

In order to quantify any overlap between the template ligands and the actives in the data set, the active with the highest Tanimoto coefficient, greater than or equal to a threshold of 0.85, to each template was selected. This threshold was chosen because it is generally accepted that molecules with Tanimoto coefficients of greater than 0.85 will exhibit similar biological activity.[44−46] For each target, the total number of different clusters to which the selected actives belonged was then calculated. From Table 1 it can clearly be seen that the overlap between the templates and the actives is relatively low, i.e., the templates represent only a small number of clusters of actives in each case.

Comparison of Single and Multiple Template Methods Using Ligmatch. A common measure of the performance of VS methods is the enrichment factor (EF), which measures the number of actives found within a compound database at a particular stage of VS. ROC enrichment (ROCE) differs slightly from the EF because it expresses the percentage of actives observed as a proportion of the percentage of the *decoys* observed rather than as a proportion of the percentage of the *entire data set* observed. Thus, ROCE does not show a dependence on the ratio of actives to decoys in the data set.[47] Since early enrichment is an important measure of performance in VS, arithmetic weighted ROCE (awROCE) was calculated at the 1% level for all 12 targets in order to compare the LigMatch single template method with the LigMatch multiple template method. awROCE was calculated at the 1% level because it is likely that, in the drug discovery process, a similarly small proportion of the ranked database would be subject to further experimental screening. As with the EF, ROCE values of greater than 1.0 signify enrichment with respect to random. To provide a measure of VS performance across the entire data set, the arithmetic weighted area under the ROC curve (awAUC) was also calculated for all targets, using both the LigMatch single and multiple template methods. In contrast to the EF, the AUC shows no dependence on the ratio of actives to decoys in a data set. Ideal distributions of actives and decoys result in an AUC value approaching 1.0, whereas random distributions result in a value of 0.5.

Performance of LigMatch in the Absence of Template Similarity. The method used to select the template ligand in the LigMatch multiple template method was modified in order to investigate the effects of reduced 2D similarity between the templates and the active and decoy molecules. Accordingly, the ligand with the highest Tanimoto coefficient to an active or decoy molecule, *below or equal to* a certain threshold, was chosen as the template for that particular molecule. Thresholds of 1.0 (i.e., no threshold),

0.8, 0.6, 0.4, and 0.2 were used, and where there were no templates with Tanimoto coefficients below or equal to a certain threshold, the relevant active or decoy molecules were given a normalized mean score and a weighting of zero. The mean enrichment across all 12 target proteins was calculated for each threshold, before being plotted as a ROC curve to give a generalized view of the overall performance. The above procedure was also performed using the 2D similarity method where the Tanimoto coefficient was used as the basis for prioritizing the actives and the decoys. In this way, it was possible to compare the performance of LigMatch with the 2D similarity method in an unbiased way at low template similarity.

RESULTS AND DISCUSSION

Comparison of Methods on Single Templates. The performance of LigMatch on the benchmark of 13 selected target protein templates used in the study by Cheesewright et al. is shown by the awROC curves in Figure 1. The performance of FieldScreen, DOCK, ROCS, and the 2D fingerprint-based method on the same data set is also given for comparison. LigMatch gives the best performance out of all of the methods for six of the targets, in particular for ache, inha, p38, src, and vegfr2. For the targets where other methods outperform LigMatch, such as ace and egfr, it is never the worst performing method. For these two targets, ROCS and the 2D fingerprint-based method give superior performance. However, they appear unable to maintain this level of performance across all targets, particularly for fxa, p38, and vegfr2, where their performance is close to random. The performance of the methods in terms of enrichment (awROCE at 1, 2, and 5%) is given in Table 2. Although ROCS outperforms LigMatch at the 1% level for six of the 13 targets, it gives very poor performance for many of the other targets, in particular for src and vegfr2, where it fails to retrieve any actives at the 1% level. On average, across all 13 targets, LigMatch outperforms each of the other methods at all three enrichment levels. In addition, the standard deviations across the targets are lower than those of the 2D fingerprint method and ROCS (which are the two most competitive methods in terms of their performance).

One target family that would appear to be particularly troublesome for many of the methods is the protein kinase family, which makes up six of the 13 targets, cdk2, egfr, p38, pdgfrb, src, and vegfr2. Interestingly, Huang et al. commented on the high failure rates observed for the protein kinase targets when using DOCK. Furthermore, in a recent study where they studied the performance of ROCS against all 40 DUD targets, Kirchmair et al.[23] also showed that, when compared to other target families, the lowest performance, on average, was achieved for the protein kinases. A possible explanation of this failure could lie within the characteristics of the highly conserved protein kinase ATP binding site. Indeed, the ATP binding site is highly flexible and can alter its shape and size in response to ligand binding. Therefore, the use of a single ligand as a template ligand, or indeed a single protein structure as a template for docking, could be insufficient to represent the wide diversity of binding modes exhibited by the different ligands that bind a single protein kinase target (for a recent paper examining the structural diversity of protein kinase binding sites see Kinnings and

Jackson (2009)[48]). With the exception of egfr, LigMatch gives the best performance overall for the protein kinase targets. For p38, src, and vegfr2 in particular, LigMatch performed well, while several of the other methods failed to perform better than random.

Comparison of Single and Multiple Template Methods Using LigMatch. We have compared the LigMatch single template method (described above) with using multiple, structurally diverse templates for a target protein cocrystallized with different bioactive ligands. In this case, the 2D fingerprint was used to select a ligand template based on the Tanimoto coefficient. The effect of using multiple templates, rather than single templates, on the performance of LigMatch is shown in Figures 2 and 3. All awROCE values at 1% (Figure 2) are well above 1.0 (ranging from ~10−70), and all awAUC values (Figure 3) are above 0.5, indicating that both single and multiple template methods perform much better than random across all targets. The multiple template method generally outperforms the single template method in terms of both early enrichment (awROCE at 1%) and enrichment across the entire data set (awAUC). As shown in Figure 2, the multiple template method improves the early enrichment for seven out of 12 targets. For one target, ace, there is no change, and for the remaining four targets, early enrichment is better using the single template method.

In general, the greatest improvements in early enrichment from using the multiple template method can be observed where early enrichment is relatively low using the single template method. The awAUC values in Figure 3 show a very similar trend, suggesting that the results achieved for early enrichment are consistent with enrichment across the entire data set. For instance, pde5, p38, and cdk2, which have low early enrichment with the single template method, show the greatest percentage increases with the multiple template method (300, 185, and 79% improvement, respectively). Achieving good enrichment does not simply depend upon analogs of the template ligands being present in the data set of actives. Indeed, for pde5, the target showing the greatest improvement in enrichment, only one out of a total of 22 clusters of actives were represented by the eight different ligand templates (see Table 1). The targets that show a decrease in performance with the multiple template method already have the greatest early enrichment with the single template method, i.e. src, followed by inha and then vegfr2. In all cases except src, the early enrichment is never significantly reduced by moving to a multiple template model, and in many cases, it is significantly improved, especially for pde5, p38, fxa, hivrt, and cdk2. It would appear that the multiple template method removes some of the randomness from choosing an appropriate template, leading to a more robust method overall.

For those targets that performed poorly using the single template method, but then showed a significant improvement upon using the multiple template method, the unsuitability of the DUD ligand may have been a contributing factor. For instance, p38 is a mitogen-activated protein kinase (PDB code, 1kv2), with the template ligand 1-(5-tert-butyl-2-*p*-tolyl-2h-pyrazol-3-yl)- 3-[4-(2-morpholin-4-yl-ethoxy)-naphthalen-1-yl]-urea (ligand code, B96). The 185% increase in early enrichment when using the multiple template method on this target is possibly due to the unsuitability of B96 as
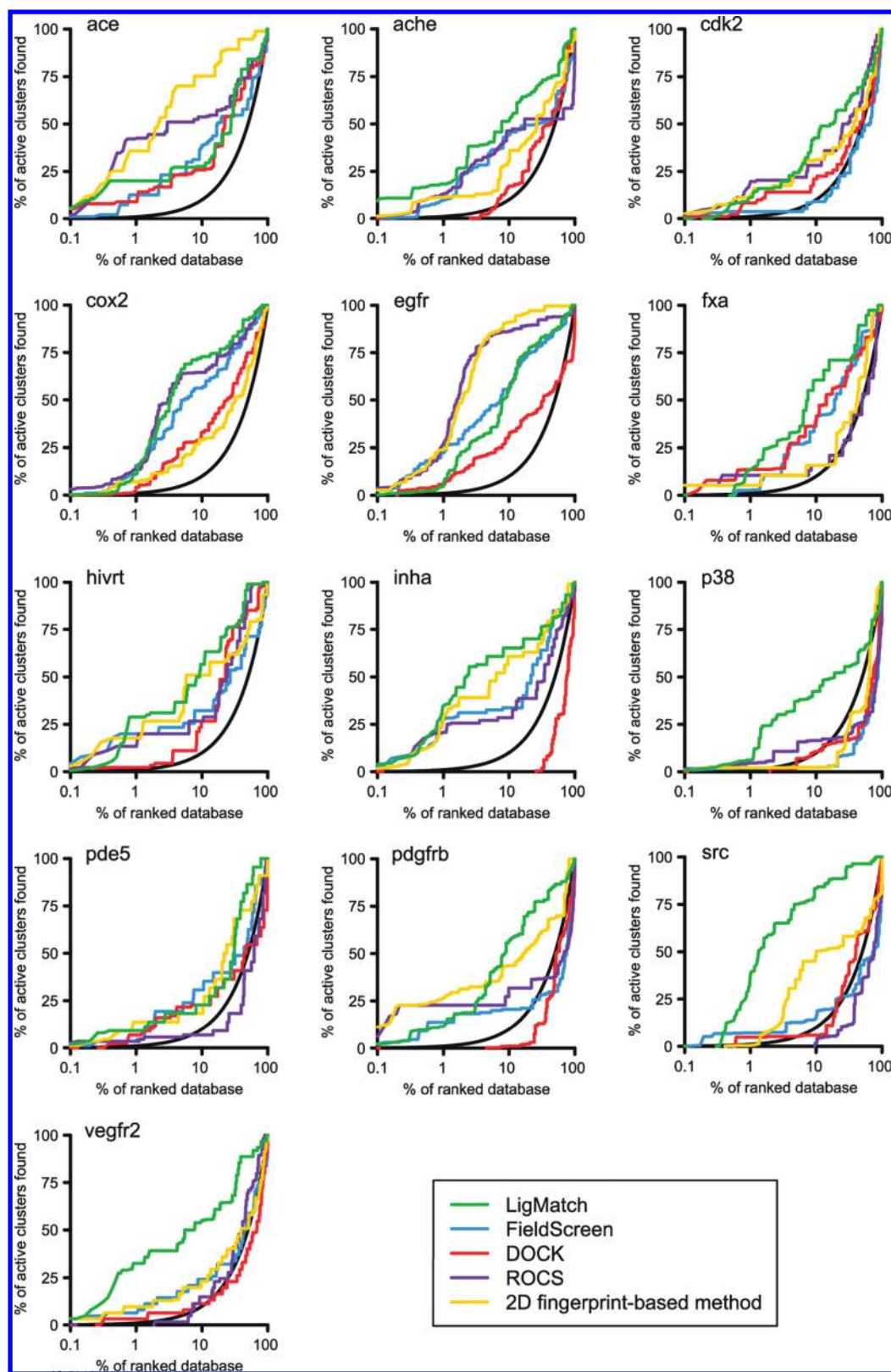
Figure 1. Arithmetic weighted ROC (awROC) curves to compare the performance of five different VS methods across the 13 selected DUD targets. The lines are colored as follows: LigMatch (green), FieldScreen (blue), DOCK (red), ROCS (purple), 2D fingerprint-based method (yellow), and random performance (black).

a template ligand. Indeed, B96 is a large ligand (MW > 500) that is bound in the rarely seen 'DFG-out' conformation of the protein.[49] There are many known actives which receive low scores when using B96 in the single template method. For instance, the lowest ranked active (ZINC03815765), which was only retrieved after observing >99% of the data

set, is shown in Figure 4. The predicted binding pose is wrong with the molecule positioned in the back of the binding cleft and a major steric clash with a leucine residue (Leu167). However, when using the multiple template method, 4-(4-fluorophenyl)-1-(4-piperidinyl)-5-(2-amino-4-pyrimidinyl)-imidazole (SB4) is selected as the template, and

Table 2. Arithmetic Weighted ROC Enrichment (awROCE)[a] Values at 1, 2, and 5% for Five Different VS Methods Across the 13 Selected DUD Targets

target	LigMatch			FieldScreen			DOCK			ROCS			2D fingerprint-based method		
	1%	2%	5%	1%	2%	5%	1%	2%	5%	1%	2%	5%	1%	2%	5%
ace	20.0	10.0	5.4	12.6	8.9	4.7	14.2	8.4	4.6	44.5	22.9	10.2	48.2	29.6	14.0
ache	20.8	19.1	8.1	20.4	13.5	7.3	0.0	0.0	0.8	23.7	15.4	7.2	11.9	6.0	2.9
cdk2	15.9	8.3	5.5	3.8	1.9	0.8	9.9	7.0	2.8	20.3	10.2	4.4	15.9	8.8	5.2
cox2	41.5	28.1	13.8	29.5	17.8	10.4	5.3	7.1	5.5	48.1	29.3	12.8	11.0	6.1	4.3
egfr	23.1	15.1	7.5	29.5	18.1	9.5	6.8	7.0	4.5	77.6	39.6	17.1	71.7	41.0	17.5
fxa	26.0	16.4	11.0	2.8	5.4	5.4	13.7	7.0	6.2	10.5	5.3	2.1	5.3	5.3	2.1
hivrt	31.1	15.6	8.7	20.0	11.7	5.1	2.2	2.2	2.2	20.0	10.0	4.0	26.7	13.3	10.2
inha	51.1	28.3	12.2	31.2	15.6	6.5	0.0	0.0	0.0	25.6	12.8	5.4	39.1	19.6	9.6
p38	24.2	15.1	7.6	1.8	0.9	0.5	0.0	0.0	0.4	4.9	5.4	3.2	2.1	1.1	0.4
pde5	11.4	6.8	3.6	4.5	9.7	4.8	6.8	8.0	4.3	4.5	2.8	1.4	13.6	6.8	3.6
pdgfrb	18.2	9.5	8.4	13.6	9.1	3.8	0.0	0.0	0.0	22.7	11.4	4.5	29.5	16.2	7.1
src	60.0	32.6	15.2	7.0	3.7	2.5	4.8	2.6	1.0	0.0	0.0	0.0	0.4	4.9	8.0
vegfr2	39.1	19.6	10.0	8.1	7.3	3.5	3.2	3.2	1.3	0.0	0.8	0.3	9.7	6.5	4.0
mean	29.4	17.3	9.0	14.2	9.5	5.0	5.1	4.0	2.6	23.3	12.8	5.6	21.9	12.7	6.8
sd	14.5	8.2	3.4	10.8	5.7	3.0	5.0	3.4	2.2	22.3	11.7	5.1	20.7	11.5	4.9

[a] Equivalent awROC values are provided in the Supporting Information.
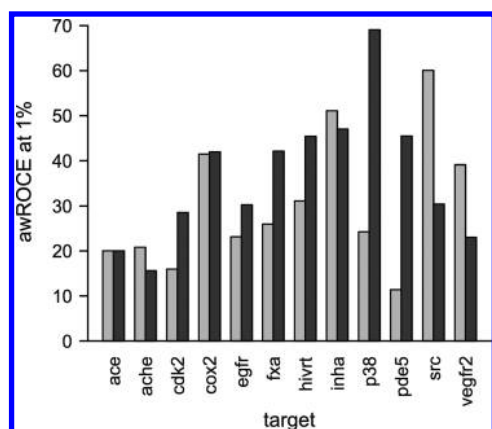
Figure 2. Arithmetic weighted ROC enrichment (awROCE) at 1% for both the single (light gray) and the multiple template methods (dark gray) using LigMatch for each of the 12 DUD targets.
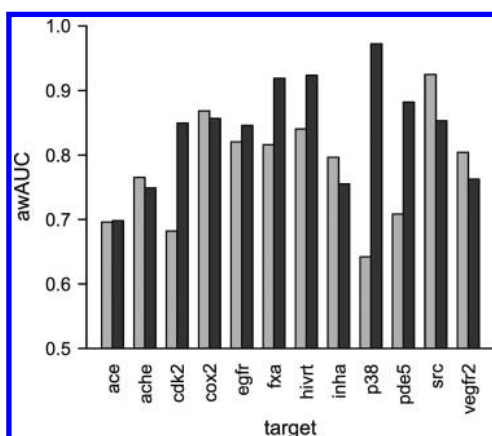
Figure 3. Arithmetic weighted area under the curve (awAUC) for both the single (light gray) and the multiple template methods (dark gray) using LigMatch for each of the twelve DUD targets.

Figure 4. The best scoring conformation of a p38 active molecule (ZINC03815765, colored by element) superimposed onto the single template ligand B96 (shown in yellow). The protein structure (PDB code, 1kv2) is shown as a ribbon and colored by secondary structure (Leu167 is shown in green stick representation). N.B. the protein structure was not considered in this study.

the resulting binding pose is much more realistic (Figure 5), causing the active to be given a much higher score and to be retrieved after observing <10% of the data set. There are many other actives for which this is also the case, and success with p38, in particular, is greatly increased by adopting the multiple template approach, since it is possible to retrieve

actives that bind to *both* the DFG-in and the DFG-out conformations of the protein.

Another example of a target with a weakly performing template ligand is the cyclin-dependent kinase cdk2. Indeed, purvalanol B has a relatively small molecular weight of 260 Da when compared with the mean molecular weight of all cdk2 actives, which is 331 Da.[4] As a result, highly scoring database compounds are limited to those that occupy the relatively small region defined by purvalanol B. Cdk2 is the target for which there are by far the greatest number of possible templates: 131 different ligands have been cocrystallized with cdk2 structures in the RCSB PDB (as of 09/16/2008). These ligands exhibit a wide variety of different binding modes in the hinge region of cdk2, and they bind in conformations that extend well beyond the area defined by purvalanol B. While a single template is simply unable to
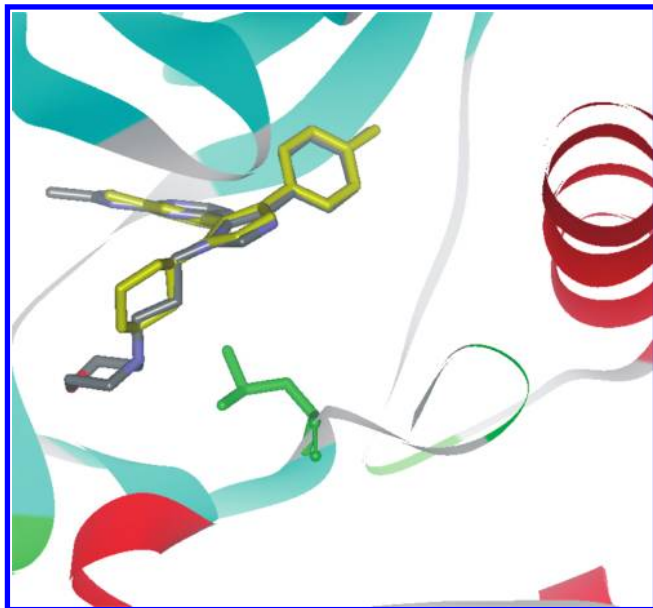
Figure 5. The best scoring conformation of a p38 active molecule (ZINC03815765, colored by element) superimposed onto the multiple template ligand SB4 (shown in yellow) using the geometric hashing method. The protein structure (PDB code, 1bl7) is shown as a ribbon and colored by secondary structure (Leu167 is shown in green stick representation).

Figure 6. Arithmetic weighted ROC (awROC) curves for the target pde5 using eight different ligands as template ligands. The original DUD ligand (ligand code, VDN) is represented by the thick green line, and the other lines are colored as follows: IBM (red), 5GP (yellow), CIA (blue), 3P4 (dark green), VIA (orange), 7CA (pale pink), and WAN (purple).

Figure 7. Arithmetic weighted ROC (awROC) curves for the target src using 13 different ligands as template ligands. The original DUD ligand (ligand code, ANP) is represented by the thick green line, and the other lines are colored as follows: NBS (dark gray), MPZ (purple), S03 (yellow), P01 (dark pink), 24A (dark green), HET (brown), H8H (blue), RBS (turquoise), DJK (orange), STI (red), SR2 (light gray), and STU (pale pink).

represent the large number of different binding modes exhibited by cdk2 ligands, a multiple template method has the potential to retrieve a set of actives with diverse binding modes. Indeed, using the multiple template method is shown to increase both early and overall enrichment by almost 100% relative to the single template method, which produces relatively low enrichment when compared with the other targets (see Figures 2 and 3).

Dependency on Choice of Single Template. The greatest improvement in early enrichment (Figure 2) when using the multiple as opposed to the single template method was observed for the target pde5, which improved by 300%. Conversely, the greatest decrease in early enrichment was observed for the target src, which showed a decrease of 49%. Furthermore, early enrichment of pde5 using the single template method was the lowest observed across all targets, while that of src was the highest; therefore, the effects of using alternative ligands as the single template were investigated for these two targets. The pde5 actives and decoys were scored against all eight possible pde5 templates, while the src actives and decoys were scored against all 13 possible src templates, and awROC curves were plotted (Figures 6 and 7). From Figure 6, it is clear that the pde5 DUD ligand, vardenafil (ligand code, VDN), resulted in the worst enrichment out of all eight possible templates. Therefore, for this target, the multiple template method could only improve upon the enrichment resulting from the single template method. The opposite is true for src. Figure 7 shows that the src DUD ligand, phosphoaminophosphonic acid-adenylate ester (ligand code: ANP), gave the best enrichment out of all 13 possible templates. As a result, the multiple template method is likely to result in a decrease in enrichment for this target. Therefore, the judicious choice of a suitable single template can, in practice, mean the difference between VS success and failure. This further strengthens the case for using the multiple template method.
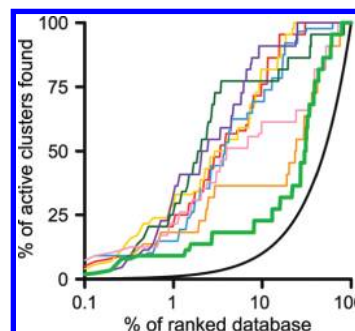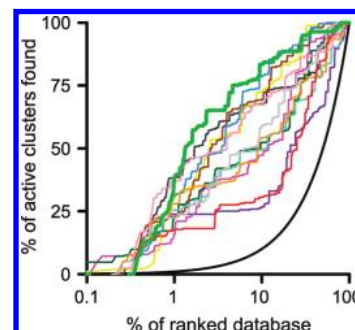
In a recent study, Sheridan et al.[50] carried out ligand-based VS where each target protein was represented by a maximum of five crystal structures. Their results showed that the relative effectiveness of different VS methods was highly dependent on the template ligand against which the compound database was screened. In many studies, the coupling between a particular protein structure and ligand would appear to be arbitrary. This is illustrated clearly in Figures 6 and 7, where the ROC curves that result from using different templates for the same target vary dramatically. The DUD ligand happened to be the ligand that gave the best performance out of all possible template ligands for src, but the one that gave the worst performance for pde5.[21] For this very reason, it is better to use a multiple template method, which guarantees good, although perhaps not optimal, performance, rather than a single template method, which has the potential to result in very poor performance, as observed for pde5.

Performance of LigMatch in the Absence of Template Similarity. One important aspect for the performance of all methods that rely on 2D and 3D ligand similarity or shape matching is that performance is dependent upon the amount of information about known actives used by the method. Clearly if this information content is high (having representative coverage of all chemotypes in the actives list), then the power of these methods (particularly 2D methods) to distinguish actives from decoys will also be high. We have attempted to create a benchmark where similarity is system-
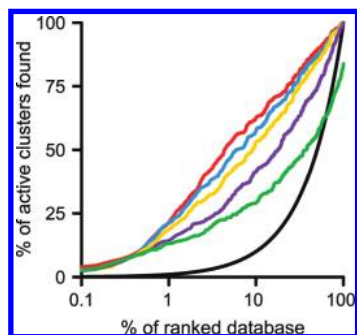
Figure 8. Mean arithmetic weighted ROC (awROC) curves across 12 selected DUD targets, using different Tanimoto thresholds in the selection criteria for the LigMatch multiple template method. The native ligand with the best Tanimoto coefficient to an active or decoy molecule that was below or equal to a certain threshold was chosen as the template for that particular molecule. The lines are colored as follows: threshold of 1.0 (i.e., no threshold) (red), 0.8 (blue), 0.6 (yellow), 0.4 (purple), 0.2 (green), and random performance (black).
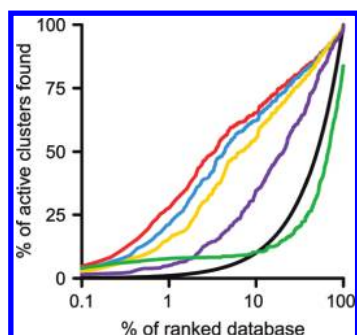

Figure 9. Mean arithmetic weighted ROC (awROC) curves across 12 selected DUD targets, using different Tanimoto thresholds in the selection criteria for the 2D fingerprint-based method. The native ligand with the best Tanimoto coefficient to an active or decoy molecule that was below or equal to a certain threshold was chosen as the template for that particular molecule. The lines are colored as follows: threshold of 1.0 (i.e., no threshold) (red), 0.8 (blue), 0.6 (yellow), 0.4 (purple), 0.2 (green), and random performance (black).

atically removed from the ligand data set in order to test the ability of LigMatch to perform at lower levels of similarity. When averaged across all 12 targets, both the LigMatch (Figure 8) and the 2D fingerprint-based (Figure 9) multiple template methods show similarly strong enrichment at Tanimoto thresholds of 1.0, 0.8, and 0.6 (i.e., removing possible templates above this level). However, LigMatch outperforms the 2D fingerprint-based method at the lower thresholds of 0.4 and 0.2. LigMatch gives superior performance particularly with respect to early enrichment at the 0.4 threshold, and it gives superior performance across the entire ranked data set at the 0.2 threshold. At the 0.4 threshold, awROC values at the 1% level are 14.3 for LigMatch and 5.1 for the 2D fingerprint-based method, while at the 0.2 threshold these values are 13.2 and 7.8 respectively.

These results clearly illustrate the added value that LigMatch provides when compared with a 2D fingerprint-based method. Indeed, as expected the 2D fingerprint-based method depends upon the existence of a sufficiently similar template ligand against which each database molecule can be compared. When this similarity is removed, this method is no longer successful at retrieving actives. By incorporating an additional 3D similarity search in the form of a geometric hashing method, we have demonstrated that it is still possible to retrieve actives at low levels of template similarity. Although they may be dissimilar to the template ligands in terms of their 2D fingerprints, such actives may still have a high affinity to the target. The LigMatch multiple template method, therefore, has the potential to screen a database and identify putative bioactive compounds that are structurally novel; such 'template-hopping' behavior is a highly desirable attribute for the drug discovery process.

CONCLUSIONS

We have developed a new VS method called LigMatch and evaluated its performance on 13 protein targets. In order to overcome issues concerning lead-likeness and analogue bias, we have validated LigMatch using a clustered and filtered version of the DUD, together with chemotype-corrected metrics as a measure of enrichment. When using a single ligand as the template ligand, LigMatch was shown to outperform four other VS methods (FieldScreen, DOCK, ROCS, and a 2D fingerprint-based method) in terms of awROCE at the 1, 2, and 5% levels, when averaged across all targets. We subsequently showed that the LigMatch method is more powerful when used in conjunction with multiple templates, where first it uses 2D similarity (Tanimoto coefficient) and the MAX fusion rule to select a template ligand for each compound in the database. A geometric hashing method is then used to score the 3D structure of each compound against the 3D structure of its chosen template ligand. The resulting list of compounds is subsequently ranked according to normalized mean 3D similarity scores.

We have demonstrated that VS performance is highly dependent on the choice of ligand used as the template ligand. If the template is unsuitable, the method will undoubtedly fail. In many cases, it was shown that one template was not suitable for representing all possible binding modes observed for actives of a particular target. It is for this reason that the multiple template method is generally superior, in terms of retrieving actives, to the single template method. Furthermore, LigMatch gives the best results when a large number of template ligands are available for a particular target. Interestingly, the greatest improvement in awROCE at 1% when using the multiple as opposed to the single template method was seen for pde5 (300%), followed by p38 (185%), cdk2 (79%), fxa (62%), and hivrt (46%), four of which were the targets with the greatest numbers of template ligands to choose from (see Table 1). Although these increases in performance may seem relatively small when compared to the increase in the number of templates used, it is important to remember that a large proportion of the total actives will often be retrieved using a single template ligand, therefore leaving only a small number of additional actives available for retrieval. As a result, it is unlikely that an increase in active retrieval that is proportionate to the number of templates used will ever be achieved.[30] Furthermore, we have shown that the templates used for the multiple template method generally represented a small proportion of the different clusters of actives in the data set (see Table 1). Therefore, it is likely that the performance of the multiple template method would improve with better representation of these clusters of actives by the template ligands.

LigMatch was also shown to perform well even in the absence of 2D similarity between the templates and actives, therefore demonstrating that it has the potential to identify structurally novel bioactive molecules. In this study LigMatch was only tested using the bioactive conformations of template ligands. However, the developers of ROCS have shown that a single low-energy conformation can be used successfully as a template ligand for VS.[8] Therefore, LigMatch could also be applied to protein targets for which no crystal structure with a bound inhibitor is available, simply by using low-energy conformations of known actives as the template ligands.

ACKNOWLEDGMENT

Supporting Information Available: Information about the ligands used in the multiple template method, awROC curves comparing the performance of the single and the multiple template methods, and awROCE/awROC values at 0.5, 1, 2, and 5%, and awAUC values for the single and multiple template methods are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* **2004**, *2* (22), 3256–66.

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

(3) Sperandio, O.; Andrieu, O.; Miteva, M. A.; Vo, M. Q.; Souaille, M.; Delfaud, F.; Villoutreix, B. O. MED-SuMoLig: a new ligand-based screening tool for efficient scaffold hopping. *J. Chem. Inf. Model.* **2007**, *47* (3), 1097–110.

(4) 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* (11), 2108–17.

(5) 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* (20), 5912–31.

(6) 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* (20), 5880–4.

(7) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* **2005**, *48* (17), 5448–65.

(8) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50* (1), 74–82.

(9) 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* (4), 1504–19.

(10) Perez-Nueno, V. I.; Ritchie, D. W.; Rabal, O.; Pascual, R.; Borrell, J. I.; Teixido, J. Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.* **2008**, *48* (3), 509–33.

(11) 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* (5), 1489–95.

(12) 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* (26), 6569–83.

(13) Bergmann, R.; Linusson, A.; Zamora, I. SHOP: scaffold HOPping by GRID-based similarity searches. *J. Med. Chem.* **2007**, *50* (11), 2708–17.

(14) 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* (2), 665–76.

(15) Doddareddy, M. R.; Choo, H.; Cho, Y. S.; Rhim, H.; Koh, H. Y.; Lee, J. H.; Jeong, S. W.; Pae, A. N. 3D pharmacophore based virtual screening of T-type calcium channel blockers. *Bioorg. Med. Chem.* **2007**, *15* (2), 1091–105.

(16) Low, C. M.; 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* (22), 6790–802.

(17) Schuster, D.; Maurer, E. M.; Laggner, C.; Nashev, L. G.; Wilckens, T.; Langer, T.; Odermatt, A. The discovery of new 11beta-hydroxysteroid dehydrogenase type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J. Med. Chem.* **2006**, *49* (12), 3454–66.

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

(19) Chen, X.; Reynolds, C. H. Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1407–14.

(20) Eckert, H.; Bajorath, J. Design and evaluation of a novel class-directed 2D fingerprint to search for structurally diverse active compounds. *J. Chem. Inf. Model.* **2006**, *46* (6), 2515–26.

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

(22) 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* (3−4), 169–78.

(23) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* **2009**, *49* (3), 678–92.

(24) Geppert, H.; Horvath, T.; Gartner, T.; Wrobel, S.; Bajorath, J. Support-vector-machine-based ranking significantly improves the effectiveness of similarity searching using 2D fingerprints and multiple reference compounds. *J. Chem. Inf. Model.* **2008**, *48* (4), 742–6.

(25) Tovar, A.; Eckert, H.; Bajorath, J. Comparison of 2D fingerprint methods for multiple template similarity searching on compound activity classes of increasing structural diversity. *ChemMedChem* **2007**, *2* (2), 208–17.

(26) Shemetulskis, N. E.; Weininger, D.; Blankley, C. J.; Yang, J. J.; Humblet, C. Stigmata: an algorithm to determine structural commonalities in diverse datasets. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (4), 862–71.

(27) Schuffenhauer, A.; Floersheim, P.; Acklin, P.; Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (2), 391–405.

(28) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Profile scaling increases the similarity search performance of molecular fingerprints containing numerical descriptors and structural keys. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (4), 1218–25.

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

(30) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. New methods for ligand-based virtual screening: use of data fusion and machine learning to enhance the effectiveness of similarity searching. *J. Chem. Inf. Model.* **2006**, *46* (2), 462–70.

(31) Whittle, M.; Gillet, V. J.; Willett, P.; Alex, A.; Loesel, J. Enhancing the effectiveness of virtual screening by fusing nearest neighbor lists: a comparison of similarity coefficients. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1840–8.

(32) Brakoulias, A.; Jackson, R. M. Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: an automated all-against-all structural comparison using geometric matching. *Proteins* **2004**, *56* (2), 250–60.

(33) ROCS, Version 2.3.1; Openeye Scientific Software, Inc.: Santa Fe, NM, 2007.

(34) The Open Babel Package, Version 2.1.1; http://openbabel.sourceforge.net/. Accessed May, 2008.

(35) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L. The Blue Obelisk-interoperability in chemical informatics. *J. Chem. Inf. Model.* **2006**, *46* (3), 991–8.

(36) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 193–9.

(37) *OMEGA*, Version 2.2.1; Openeye Scientific Software, Inc.: Santa Fe, NM, 2007.

(38) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21* (5), 449–62.

(39) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **2006**, *46* (4), 1848–61.

(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–42.

(41) James, C. A., Weininger, D. *Daylight Theory Manual.* Daylight Chemical Information Systems Inc.: Irvine, CA., 2006.

(42) Clark, R. D.; Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 141–6.

(43) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403–10.

(44) Matter, H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.* **1997**, *40* (8), 1219–29 .

(45) Thimm, M.; Goede, A.; Hougardy, S.; Preissner, R. Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1816–22.

(46) Wallach, I.; Lilien, R. The protein-small-molecule database, a non-redundant structural resource for the analysis of protein-ligand binding. *Bioinformatics* **2009**, *25* (5), 615–20.

(47) Nicholls, A. What do we know and when do we know it. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 239–55.

(48) Kinnings, S. L.; Jackson, R. M. Binding site similarity analysis for the functional classification of the protein kinase family. *J. Chem. Inf. Model.* **2009**, *49* (2), 318–29.

(49) 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* (4), 268–72.

(50) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 257–65.