JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

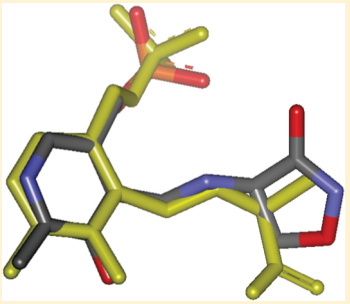
ReverseScreen3D: A Structure-Based Ligand Matching Method To Identify Protein Targets

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

Supporting Information

ABSTRACT: Ligand promiscuity, which is now recognized as an extremely common phenomenon, is a major underlying cause of drug toxicity. We have developed a new reverse virtual screening (VS) method called ReverseScreen3D, which can be used to predict the potential protein targets of a query compound of interest. The method uses a 2D fingerprint-based method to select a ligand template from each unique binding site of each protein within a target database. The target database contains only the structurally determined bioactive conformations of known ligands. The 2D comparison is followed by a 3D structural comparison to the selected query ligand using a geometric matching method, in order to prioritize each target binding site in the database. We have evaluated the performance of the ReverseScreen2D and 3D methods using a diverse set of small molecule protein inhibitors known to have multiple targets, and have shown that they are able to provide a highly significant enrichment of true targets in the database. Furthermore, we have shown that the 3D structural comparison improves early enrichment when compared with the 2D method alone, and that the 3D method performs well even in the absence of 2D similarity to the template ligands. By carrying out further experimental screening on the prioritized list of targets, it may be possible to determine the potential targets of a new compound or determine the off-targets of an existing drug. The ReverseScreen3D method has been incorporated into a Web server, which is freely available at http://www.modelling.leeds.ac.uk/ReverseScreen3D.



INTRODUCTION

Virtual screening (VS) commonly involves the molecular docking of large chemical libraries into the three-dimensional structure of a protein target. It has been established as a promising tool for identifying new lead compounds and is an integral part of the modern drug discovery process. However, the opposite task of identifying the most likely protein target(s) of a known ligand from among thousands of candidates is also a challenging task.[1,2] Proteomics-based approaches, which involve expressing all of these proteins in the presence and absence of the compound of interest, are time-consuming and expensive.[1,3] An alternative approach that has shown promise in recent years is the use of reverse (or "inverse") VS to identify the putative targets of a given compound.[1] Here the term "reverse" is used because the method is used to identify proteins that will fit with a specific compound, rather than identifying compounds that will fit with a specific protein, as is the case in standard VS.[4] By screening a compound against a protein target database *in silico*, it is possible to identify protein "hits" as potential candidates for subsequent experimental validation.[1]

In vitro screening assays for small molecules typically only permit testing against a single protein target. Therefore, off-targets, which can potentially cause side effects, or which may serve as secondary targets, are commonly overlooked.[5] Off-target identification is extremely important because there is increasing evidence to suggest that the action of drugs is much more complicated than that described by the concept of the "magic bullet", with one ligand acting on just one protein target. Indeed, a number of recent studies have revealed that protein and ligand promiscuity is a much more common phenomenon than previously thought. For instance, a recent study by Yildirim et al.[6] not only revealed that a single target can often bind multiple drugs, but also that it is far more common than expected for a single drug to modulate several different molecular targets, which may be involved in multiple diseases. In fact, drugs that bound only a single target were found to be the exception. Similarly, Paolini et al.[7] discovered that around 35% of their database of 276,122 active compounds had observed activity for more than one target. While the majority of these promiscuous compounds were active against targets within the same gene family, a significant proportion (around one-quarter) had recorded activity across different gene families. In fact, the rapid completion of numerous genome sequencing projects has revealed that proteins involved in entirely different biochemical pathways, and even residing in different tissues and organs, may possess functional binding pockets with similar shapes and physicochemical properties.[8]

Received: August 18, 2010
Published: February 28, 2011

Given the above, it is not surprising that the biggest hurdle in drug development is toxicity, which results in numerous lead compounds, especially those that are more promiscuous in nature, failing late stage clinical trials. By using reverse VS to identify potential off-targets at an early stage in the drug discovery process, it may be possible to circumvent this problem. Indeed, by identifying and subsequently fine-tuning the binding profile of a particular compound, it is possible to select those compounds for which profiles maximally modulate a disease network, while causing minimal side effects, for clinical development.[9] Similarly, reverse VS can play a significant role in elucidating the molecular mechanisms behind the unwanted side effects observed for existing drugs.[10,11] By subsequently modifying a drug to decrease its affinity to the undesirable off-target(s), while maintaining its affinity to the intended target, it may be possible to eliminate its observed side effects.

Reverse VS also has the potential to reveal new uses for existing drugs, thereby reducing both the time and costs associated with drug development.[12] Indeed, the successful launch of a single new drug is estimated to cost around U.S. $800 million and take a staggering 15 years.[13,14] By repositioning drugs that have already been approved for human use, it is possible to bypass toxicological and pharmacokinetic assessments, which together contribute almost 40% of the overall cost of bringing a new drug to market. Newly identified drug indications can be evaluated relatively quickly in phase II clinical trials, which typically only take two years and cost around $17 million.[13,14] However, the successful repositioning of drugs is commonly serendipitous, with physicians and pharmaceutical companies mainly relying on chance observations and educated guesses.[13] In order to achieve its full potential, a more systematic approach to drug repurposing is required. Through the use of reverse VS, it may be possible to identify many potential off-targets of a particular drug on a proteome-wide scale. It is likely that some of these off-targets will represent viable drug targets, therefore providing opportunities to repurpose that particular drug to target different pathways and to treat different diseases.

To date, only a handful of reverse VS methods have been described,[1,2,4,15−17] most of which involve the use of a docking algorithm for target identification. Although promising, these methods present a number of disadvantages: (i) Docking algorithms are notoriously weak predictors of binding affinity, and their performance has been shown to be highly target-dependent. (ii) The target databases are not automatically updated to reflect the ever-increasing amount of structural information available. (iii) The significant CPU time required for searching a compound against all entries in the target database discourages their use. Recently, Keiser et al.[18] used a similarity ensemble approach to compare protein targets by the 2D similarity of the ligands that they are known to bind. The authors screened 3665 U.S. Food and Drug Administraton (FDA)-approved and investigational drugs against a database of 65,241 ligands organized into 246 protein targets taken from the MDL Drug Data Report (MDDR) database.[19] Their study revealed unanticipated associations between thousands of drugs and ligand sets. Of the 30 most promising drug-target associations that were tested experimentally, 23 were confirmed, 5 of which were shown to be potent (<100 nM). Thus, their study demonstrated the power of using simple ligand-based similarity searches in reverse VS.

Here, we present ReverseScreen3D, a Web server for identifying potential target proteins that are likely to bind a given compound. It uses an automated 3D structure-based ligand matching method, based on that employed by LigMatch,[20] to search against a biologically relevant and automatically updated subset of ligands extracted from the RCSB Protein Data Bank (PDB).[21] Due to its ability to discriminate actives from decoys for various different protein targets, the LigMatch methodology was expected to be equally applicable to reverse VS. Therefore, ReverseScreen3D may serve as a valuable tool for the identification of targets of novel synthetic compounds or newly isolated natural products, as well as existing drugs with either known or unknown mechanisms of action.

METHODS

Compilation of the ReverseScreen3D Target Database. *BLAST Clusters in the RCSB PDB.* All protein chains in the RCSB PDB, except for those that contain fewer than 20 amino acids, are automatically clustered by sequence identity on a weekly basis for the "remove similar sequences" feature of the Web site (www.pdb.org). BLASTCLUST, which is a BLAST[22] score-based single-linkage clustering algorithm, is used to cluster protein chains at 30%, 40%, 50%, 70%, 90%, 95%, and 100% sequence identity with an e-value of 0.01. The resulting BLAST clusters are available at ftp://resources.rcsb.org/sequence/clusters. The BLAST clusters with a sequence identity of 90% were selected for this study because clusters at this level are most likely to provide discrimination between different proteins. Indeed, proteins were grouped into families of 90% sequence identity in the protein−ligand database BindingMOAD.[23] At a sequence identity threshold of 90%, all qualifying protein chains in the RCSB PDB were divided among a total of 25,439 BLAST clusters (24/03/2010). For the purposes of this study, each BLAST cluster was seen to represent a unique protein.

Identification of Biologically Relevant Ligands. For each BLAST cluster, all ligands were extracted from all protein chains within the cluster, and only those that met the following criteria were retained:

- Must have 10 or more heavy atoms (on the basis that biologically irrelevant molecules in PDB files, e.g., ions, are generally very small[24]).
- Must contain carbon and either nitrogen or oxygen atoms.
- Must not contain metal atoms. Also, must not be a heme group, regardless of whether or not it contains Fe.
- Must not be a known additive of crystallographic experiments (e.g., buffer and solvent molecules) (as defined by Additional File 1 from ref 25).
- Must not be a polyethylene glycol derivative.
- Must not be covalently linked to the protein. Ligands were defined as covalently linked if any ligand atoms were within a covalent bond length, plus 20% of the bond length, of any protein atoms. Bond lengths were taken from ref 26.
- Must not be greater than 5 Å away from the protein, i.e., there must be at least one protein atom within 5 Å of at least one ligand atom.

To eliminate redundancy, only the best resolution ligand (as determined by the atomic resolution of the protein crystal structure) was retained for *each* PDB ligand code in every BLAST cluster. Ligands derived from NMR were only retained if there were no equivalent ligands derived from X-ray crystallography. In the cases where there were multiple ligand entries in an NMR structure, only the conformation of the first ligand entry was added to the database. A total of 6041 of the 25,439 BLAST clusters (23.7%) were found to contain one or more unique, biologically relevant ligands.

Binding Site Discrimination. In order to differentiate between different binding sites within the same protein, all protein chains corresponding to the identified ligands within a BLAST cluster were superimposed using a geometric matching method.[27] While the coordinates of the ligands themselves were rotated and translated to reflect the superimposition of the protein chains, the ligands themselves did not influence the superimposition process. A greedy clustering algorithm was then used to assign all ligands within a BLAST cluster to a particular binding site. The RMSDs between all ligand centroids (i.e., mean *X*, *Y*, and *Z* coordinates) for a particular BLAST cluster were calculated. The ligand found to be within 5 Å of the greatest number of other ligands ("partner ligand") was assigned to the first binding site, along with its partner ligands. Once a ligand or partner ligand had been assigned to a binding site, it was not considered again. Then, from the remaining set of ligands, the ligand with the greatest number of partner ligands was assigned to a binding site, along with its partner ligands. This process was continued until all ligands in a BLAST cluster had been assigned to a binding site. A threshold of 5 Å was chosen because it was shown to be able to discriminate between cofactor and substrate binding sites, for instance in lactate dehydrogenase and enoyl-acyl carrier protein reductase. A total of 8355 ligand binding sites were identified from the 6041 BLAST clusters containing biologically relevant ligands.

The ReverseScreen3D Method. *Conformer Generation.* Since the geometric matching alignment and scoring method uses rigid conformations, multiple conformers of each query molecule must be generated. Vconf (v2.0 from VeraChem), which uses a stochastic method for conformer searching, was chosen because it has been shown to be highly effective at reproducing bioactive conformations.[28] Vconf is used to generate up to around 100 conformers of the query molecule, using an energy threshold of 25 kcal/mol, because this was shown to be essential for best conformer generation performance in a study by Kirchmair et al.[29] 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 bioactive conformations. The energy tolerance is set to zero, so that all molecules are retained regardless of the similarity between their energies, and a distance tolerance of 1 Å RMSD is used to discard highly similar conformations. No time restriction is imposed on the conformational search.

2D and 3D Ligand-Based Matching. Firstly, Open Babel 2.1.1[30,31] is used to calculate the Tanimoto coefficients between the 2D fingerprints (Daylight representation) of the query molecule and every ligand in the database. The ligand with highest 2D Tanimoto coefficient to the query molecule is selected from each binding site of each BLAST cluster. A geometric matching method described previously[27] is then used to compare all conformers of the query molecule with the 3D structure of the single ligand that was chosen from each binding site in each BLAST cluster. Each match is 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 conformer with the best atom-atom score is selected, and if there are multiple conformers with the same best score, the lowest energy conformer is selected. In order to take into account the size of both the query molecule and the ligand being compared, the 3D Tanimoto coefficient of this atom-atom score is calculated as follows

$$\mathrm{Tc} = \text{atom-atom score}/(\text{no.of ligand atoms} + \text{no.of query atoms} - \text{atom-atom score})$$

Prioritization of the Target Binding Sites. Once both the 2D and the 3D Tanimoto coefficients have been calculated between the query molecule and each ligand selected from each binding site of each BLAST cluster (i.e., a total of 8355 ligands), the 8355 binding sites are ranked by their Tanimoto coefficients with the query molecule. The results presented for ReverseScreen 2D are ranked by 2D score, and those for ReverseScreen 3D are ranked by 3D score. In the event that the score is identical for two or more compounds, the compounds are further ranked by 3D score for ReverseScreen 2D or 2D score for ReverseScreen 3D.

Validation Procedure. *Compilation of the Validation Data Set.* The validation process requires the identification of compounds with known activity against a sufficient number of protein targets, which are represented by the ligand database described previously. DrugBank[32] is a publicly available database that specifically focuses on drug-target interactions. It contains nearly 4800 annotated drug entries, including more than 1350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals, and more than 3243 experimental drugs. Detailed information is given about each drug, including chemical, pharmacological, and pharmaceutical data. Furthermore, these drugs are linked to more than 2500 nonredundant protein targets, for which there is sequence, structure, and pathway information.

The target protein sequences of all compounds in DrugBank that (a) were known to act as protein inhibitors (i.e., labeled as an approved drug or described as an inhibitor in DrugBank or in the PDB) and (b) had five or more listed targets in DrugBank were downloaded. Compounds with a molecular weight of greater than 500 Da or less than 15 heavy atoms were excluded. A BLAST search[22] of these sequences was carried out against the sequences of all protein chains in the RCSB PDB (available at ftp://ftp.wwpdb.org/pub/pdb/derived_data, accessed 29/03/2010). As with BLASTCLUST, an e-value threshold of 0.01 was used, and only hits with greater than 90% sequence identity to the target protein sequence were retained. An additional filter was used to ensure that only hits with greater than 50% coverage of the target sequence were retained. The PDB code and chain of each qualifying hit was recorded, and the number of the BLAST cluster to which it belonged was retrieved. For each compound, the total number of targets corresponding to occupied BLAST clusters (i.e., those containing one or more ligands) was calculated. Only those compounds that had five or more targets represented by occupied BLAST clusters were retained.

Running ReverseScreen3D on the Validation Data Set. ReverseScreen3D was run on a total of 20 multitarget small molecule inhibitors that passed the above criteria. In order to avoid self−self matches and therefore falsely high enrichment rates in the validation study, all database ligands with 2D Tanimoto coefficients to the query molecule of greater than or equal to 0.85 were excluded. Furthermore, because the database ligands were taken from the RCSB PDB, it was possible that self−self matches could occur below this threshold, and so PDB ligand codes (HETIDs) were also used to exclude matching ligands. It is worth noting that once ligands that were highly similar to the query molecule had been removed, some of the BLAST clusters were empty. As a result, the number of

Table 1. Validation Results[a]

small molecule inhibitor	no. of heavy atoms	no. of rotatable bonds	EF @ 1%		EF @ 2%		EF @ 5%		EF @ 10%	
			2D	3D	2D	3D	2D	3D	2D	3D
liothyronine	23	3	66.7	**100.0**	50.0	50.0	20.0	20.0	10.0	10.0
2-deoxy-2,3-dehydro-N-acetyl-neuraminic acid	20	4	55.6	**88.9**	33.3	44.4	20.0	17.8	10.0	10.0
levothyroxine	24	3	60.0	**80.0**	40.0	40.0	16.0	16.0	8.0	8.0
acetazolamide	13	2	66.7	**66.7**	33.3	33.3	13.3	13.3	6.7	6.7
etretinate	26	10	20.0	**60.0**	30.0	30.0	20.0	12.0	10.0	10.0
quercetin	22	1	0.0	**60.0**	0.0	30.0	0.0	12.0	4.0	6.0
2-[(formyl-hydroxy-amino)-methyl]-heptanoic acid [1-(2-hydroxymethyl-pyrrolidine-1-carbonyl)-2-methyl-propyl]-amide	27	11	50.0	**50.0**	25.0	25.0	10.0	10.0	5.0	5.0
adapalene	31	3	50.0	**50.0**	25.0	25.0	10.0	20.0	5.0	10.0
alitretinoin	22	8	33.3	**50.0**	16.7	25.0	6.7	13.3	5.0	8.3
3-beta-hydroxy-5-androsten-17-one	21	0	14.3	**42.9**	21.4	28.6	8.6	11.4	4.3	5.7
sulindac	25	5	33.3	**33.3**	16.7	16.7	10.0	6.7	6.7	5.0
staurosporine	35	2	22.2	**33.3**	11.1	16.7	4.4	8.9	2.2	5.6
flavopiridol	29	2	28.6	**28.6**	14.3	21.4	8.6	11.4	5.7	5.7
acitretin	24	9	60.0	**20.0**	40.0	20.0	16.0	12.0	10.0	8.0
dasatinib	34	4	60.0	**20.0**	30.0	10.0	16.0	8.0	8.0	8.0
N-acetyl-D-allosamine	15	2	16.7	**16.7**	8.3	8.3	3.3	3.3	1.7	3.3
simvastatin	30	6	0.0	**11.1**	0.0	5.6	2.2	4.4	1.1	4.4
azathioprine	19	2	20.0	**0.0**	10.0	0.0	4.0	0.0	8.0	2.0
resveratrol	17	3	20.0	**0.0**	20.0	20.0	16.0	12.0	8.0	6.0
atorvastatin	41	11	16.7	**0.0**	16.7	0.0	13.3	0.0	6.7	3.3
mean			**34.7**	**40.6**	**22.1**	**22.5**	**10.9**	**10.6**	**6.3**	**6.6**
sd			**21.4**	**28.8**	**13.0**	**13.4**	**6.1**	**5.6**	**2.7**	**2.4**

[a] Enrichment factors are given at 1, 2, 5, and 10% for the 2D and the 3D methods for the 20 selected small molecule inhibitors. The results are ordered by descending EFs at 1% for the 3D method (highlighted in bold).

identifiable targets for a particular compound was sometimes less than five. This factor, in addition to the fact that there would be fewer than 8355 binding sites to search against, was taken into account when calculating enrichment factors (EFs). EFs were calculated at the 1, 2, 5, and 10% levels for both the 2D and the 3D methods for all 20 compounds. EFs were calculated on the basis of the first occurrence of a target BLAST cluster, regardless of the binding site.

RESULTS AND DISCUSSION

The Protein Target Database. The current target database contains a total of 8355 ligand binding sites in a total of 6041 unique proteins, as identified by BLAST clusters at the 90% sequence identity level. The majority (73.4%) of these proteins have one identified ligand binding site, while a significant proportion (19.5%) have two identified binding sites, and 4.7, 1.4, and 0.6% have three, four, and five identified binding sites, respectively. An increasingly smaller percentage of proteins have more than five identified binding sites, with the largest number of binding sites identified for a single protein being 14. Each of the 8355 identified binding sites contains one or more unique ligands. Because the first identified binding site in a protein always has the greatest number of ligands, it is often the case that a protein with multiple binding sites will possess a single binding site containing the majority of the ligands, and a number of other binding sites containing small numbers of, or even single, ligands. The proteins containing binding sites with the greatest number of unique ligands are HIV-1 protease, cell division protein kinase 2 (CDK2), thrombin heavy chain, carbonic anhydrase II, and beta-secretase 1, with 158, 145, 134, 114, and 107 unique ligands, respectively. Note that all of these proteins, with the exception of CDK2, have additional binding sites containing a very small number of unique ligands. The large number of different ligands cocrystallized with these proteins reflects their importance as widely studied drug targets.

Presentation of Data. In the following sections, the ReverseScreen 2D and 3D methods are firstly validated using a set of 20 multitarget inhibitors. For this initial validation study, 20 different inhibitors were chosen that had at least five known targets with representation in the ligand database (see the Methods section). This requirement for five or more targets was to allow the calculation of meaningful enrichment factors. Secondly, the performance of the ReverseScreen 2D and 3D methods are assessed on the multitarget drug 4OH-tamoxifen, and the results compared with that of three other reverse VS methods described in the literature. Finally, a number of case studies are presented in which the strengths and weaknesses of the ReverseScreen 2D and 3D methods are discussed. For the final set of individual case studies, inhibitors were selected that had either three or four known targets with representation in the ligand database. In this way, a distinct set of results could be provided for the purpose of validation. Those inhibitors for which the performance of the ReverseScreen 2D or 3D methods was remarkable in some way were selected as the individual case studies.

dx.doi.org/10.1021/ci1003174 |J. Chem. Inf. Model. 2011, 51, 624–634

Table 2. Retrieval of 11 Targets of 4OH-Tamoxifen by the ReverseScreen 2D and 3D Methods, TarFisDock, PharmMapper, and INVDOCK[a]

target name	evidence (ref)	ReverseScreen2D score	ReverseScreen2D rank (%)	ReverseScreen3D score	ReverseScreen3D rank (%)	TarFisDock (%)	PharmMapper (%)	INVDOCK
estrogen receptor	33	0.66	**0.01**	0.69	**0.01**	10.00+	**0.01**	y
17β-hydroxysteroid dehydrogenase	34	0.06	97.71	0.39	**1.37**	3.87	**0.25**	y
dihydrofolate reductase	35	0.30	3.27	0.47	**0.17**	**0.57**	**0.40**	y
glutathione transferase	36	0.38	**1.01**	0.41	**0.64**	**1.72**	**0.67**	y
prostaglandin synthase	37	0.40	**0.91**	0.38	**1.95**	10.00+	**1.70**	y
collagenase	38	0.30	3.09	0.43	**0.37**	3.01	**1.86**	y
3α-hydroxysteroid dehydrogenase	39	0.11	42.74	0.39	**1.30**	10.00+	2.30	y
protein kinase C	40	0.12	31.84	0.41	**0.55**	10.00+	3.04	y
calmodulin	40	0.34	**1.87**	0.36	2.91	10.00+	4.07	n
alcohol dehydrogenase	41	0.20	10.99	0.26	23.90	10.00+	11.19	y
immunoglobulin	42	0.28	4.24	0.42	**0.46**	**1.43**	13.70+	y

[a] In each case, the % of the ranked database observed before retrieving the target 4OH-tamoxifen is given, except in the case of INVDOCK, where a target was either retrieved ("y") or not retrieved ("n"). For TarFisDock and PharmMapper, only the top 10% and 13.7% of the ranked databases were provided in the Supporting Information, respectively, and so any targets not retrieved within these thresholds are indicated with a "+" sign, i.e., 10.00+ or 13.70+. For the ReverseScreen methods, the 2D and 3D Tanimoto coefficients between 4OH-tamoxifen and the selected target ligand are also given. All solutions retrieved in the top 2% are highlighted in bold, and the best solution for a given target is underlined.

Validation. A common measure of the performance of VS methods is the enrichment factor (EF), which measures the enrichment of actives within a given percentile of a ranked list of screened compounds. The EF can be similarly applied to measure the performance of reverse VS methods by identifying known protein targets from a target database. As mentioned previously, 20 different multitarget compounds were selected as test cases for the validation process. Further information about these compounds can be found in Table S1 of the Supporting Information. Table 1 provides EF values at 1, 2, 5, and 10% for both the 2D and the 3D methods for the 20 selected compounds. EF values of greater than 1.0 signify enrichment with respect to random. When averaged across all 20 compounds, the mean EF for the ReverseScreen 3D method at the 1% level is 40.6, indicating highly significant enrichment with respect to random. While the mean EFs are similar (within 0.5 of one another) for both the 2D and the 3D methods at the 2, 5, and 10% levels, the 3D method can be seen to outperform the 2D method at the 1% level. Indeed, the EFs at 1% are 40.6 and 34.7 for the 3D and 2D methods, respectively. These findings suggest that while both methods are good at retrieving actives from the database, the 3D method is able to retrieve them slightly sooner than the 2D method. Since only a small proportion of the ranked database would be subject to further experimental screening in reality, early enrichment (i.e., at the 1% level) is of great importance.

However, the 2D method is shown to be complementary to the 3D method as it is able to retrieve targets that are missed by the 3D method. For instance, at the 1% level, it is able to retrieve targets for atorvastatin, azathioprine, and resveratrol, none of which are retrieved by the 3D method. On the other hand, the 3D method is able to retrieve targets for quercetin and simvastatin at the 1% level, while the 2D method is not. This suggests that both methods each have their own strengths and weaknesses, and in order to carry out the most rigorous target search for a compound of interest, they should both be investigated.

4OH-Tamoxifen Case Study. Tamoxifen, which is used as an adjuvant therapy in the treatment of breast cancer, has been shown to be a multitarget drug. Indeed, tamoxifen and its active metabolite, 4-hydroxytamoxifen (4OH-tamoxifen), are known to interact with at least 11 different targets.[4] To further test the reliability of ReverseScreen3D, we used it to identify potential target proteins for 4OH-tamoxifen from our database. We have chosen to investigate 4OH-tamoxifen because it is believed to be responsible for the antiestrogenic effects of tamoxifen inside the human body.[33] Furthermore, three different reverse VS methods have used 4OH-tamoxifen as a test case for validation. Table 2 shows the 11 known targets of 4OH-tamoxifen, and their retrieval by ReverseScreen (both the 2D and 3D methods), TarFisDock,[1] PharmMapper,[16] and INVDOCK.[4] Note that the performance of ReverseScreen is not directly comparable with that of the other methods because they employed dissimilar strategies to search against entirely different databases. More specifically, INVDOCK was used to dock 4OH-tamoxifen into 1040 structures, representing 38 types of proteins known to potentially cause side effects, while TarFisDock was used to dock the query molecule into 698 structures covering 371 drug targets in the "potential drug target database" (PDTD).[3] PharmMapper, on the other hand, was used to screen 4OH-tamoxifen against a database consisting of 7302 receptor-based pharmacophore models. Although different in nature, we feel that a general comparison between the methods is useful for the purposes of this study.

For ReverseScreen, the 3D method can be seen to outperform the 2D method when retrieving all of the targets, except for prostaglandin synthase, calmodulin, and alcohol dehydrogenase. On average, the targets are retrieved after 3.06 and 17.97% of the ranked database using the 3D and 2D methods, respectively. The poor mean percentage retrieval by the 2D method can be attributed to its failure to retrieve three targets in particular; 17β-hydroxysteroid dehydrogenase, 3α-hydroxysteroid dehydrogenase, and protein kinase C, which are retrieved after only 97.71, 42.74, and 31.84% of the ranked database, respectively. These failures are due to the lack of relevant database ligands with 2D similarity to 4OH-tamoxifen. In spite of this, the 3D

Table 3. Examples of Compounds Where the ReverseScreen 3D Method Outperforms the 2D Method When Retrieving Known Protein Targets[a]

inhibitor	target	ligand	2D score	2D rank (%)	ligand	3D score	3D rank (%)
purvalanol	CDK2	OLO	0.62	0.11	OLO	0.40	0.19
	CRK2	INR	0.23	**46.96**	INR	0.27	**6.49**
	MAPK1	OLO	0.53	0.14	OLO	0.43	0.16
	MAPK3	5ID	0.32	**30.74**	5ID	0.26	**7.33**
pyridoxyl-N,O-cycloserylamide-5-monophosphate	alr	PP3	0.17	**12.75**	PP3	0.65	0.39
	arnB	PMP	0.12	**59.11**	PMP	0.73	0.04
	pabC	DCS	0.23	0.92	DCS	0.69	0.17
nojirimycine tetrazole	GLU1	GLC	0.18	**41.45**	GLC	0.53	**2.24**
	MA1	NAG	0.25	**2.21**	GOX	0.69	0.04
	PYGM	HTP	0.63	0.01	HTP	0.65	0.17

[a] All solutions not retrieved in the top 2% are highlighted in bold.

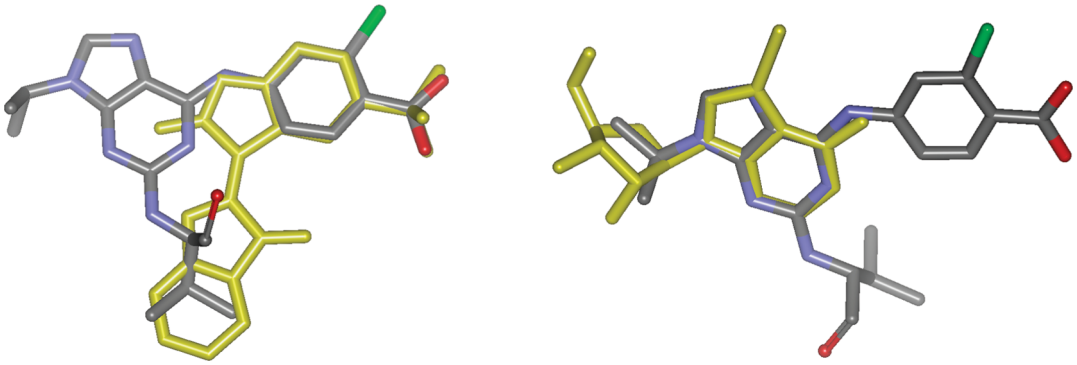


Figure 1. Superimpositions of the best scoring conformers of purvalanol (colored by element) onto each of the two template ligands [left, INR (from 1v0o) and right, 5ID (from 2zoq), both shown in yellow] by the geometric matching method.

similarities of these ligands to 4OH-tamoxifen are detected by the geometric matching method, therefore, enabling all three of these targets to be retrieved within 1.5% of the ranked database when using the 3D method. Alcohol dehydrogenase is the only target for which both the 3D and the 2D methods perform poorly. It is interesting to note that both TarFisDock and PharmMapper also perform poorly for this particular target.

As mentioned previously, TarFisDock is a reverse protein−ligand docking Web server which can be used to screen molecules of interest against a target database containing 698 protein structures. TarFisDock retrieves three of the known targets of 4OH-tamoxifen within the top 2% of the ranked database and a further two targets within the top 5%. The remaining six targets are retrieved after more than 10% of the ranked database. As a comparison, the ReverseScreen 3D method retrieves nine targets within 2% and a further one target within 5%, while the 2D method retrieves four targets within 2% and a further three targets within 5%. While TarFisDock can be seen to outperform the ReverseScreen 2D method for four of the targets, there is not a single target where TarFisDock outperforms the ReverseScreen 3D method. PharmMapper is another reverse VS method that uses a pharmacophore mapping approach to identify potential targets of a query compound from a database comprising 7302 pharmacophore models. It is able to retrieve six known targets of 4OH-tamoxifen within 2% of the ranked database and a further three targets within 5%. Therefore, its performance is more comparable to the ReverseScreen 3D method than that of TarFisDock. However, the ReverseScreen 3D method is able retrieve seven of the targets sooner than PharmMapper. In particular, while PharmMapper was unable to retrieve immunoglobulin within the top 1000 of the ranked database (i.e., within 13.70% of the ranked database), the ReverseScreen 3D method retrieves this target within 0.46% of the ranked database, therefore indicating its reliability. The automated inverse docking procedure, INVDOCK, was able to retrieve all targets except for calmodulin, which was retrieved within 1.87 and 2.91% of the ranked database using the ReverseScreen 2D and 3D methods, respectively. However, direct comparison between INVDOCK and the other methods is difficult because of the binary nature of its predictions.

Examples Where the 3D Method Outperforms the 2D Method

Purvalanol (DB02733). Four of this protein kinase inhibitor's five targets that are listed in DrugBank are represented by the ligand database; cell division protein kinase 2 (CDK2), cell division control protein 2 homologue (CRK2), mitogen-activated protein kinase 1 (MAPK1), and mitogen-activated protein kinase 3 (MAPK3) (Table 3). While CDK2 and MAPK1 were retrieved within 0.2% of the ranked database by both the 2D and 3D methods, the 2D method was only able to retrieve CRK2 and MAPK3 after 46.96 and 30.74% of the ranked database, respectively. However, the 3D method was able to provide some improvement, retrieving CRK2 and MAPK3 after 6.49 and 7.33%, respectively. This suggests that although the selected template ligands (indirubin-5-sulfonate, INR, and

dx.doi.org/10.1021/ci1003174 |J. Chem. Inf. Model. 2011, 51, 624–634

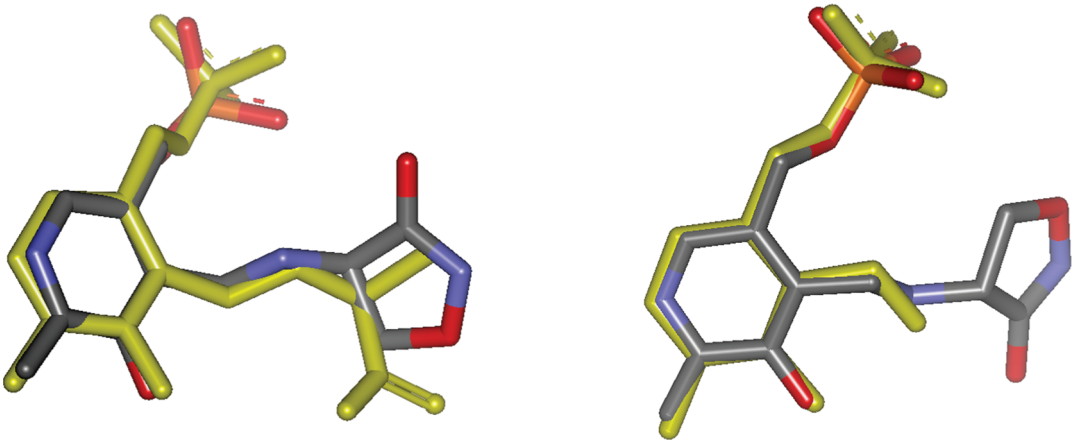


Figure 2. Superimpositions of the best scoring conformers of pyridoxyl-N,O-cycloserylamide-5-monophosphate (colored by element) onto each of the two template ligands [left, PP3 (from 1l6f) and right, PMP (from 1mdo), both shown in yellow] by the geometric matching method.

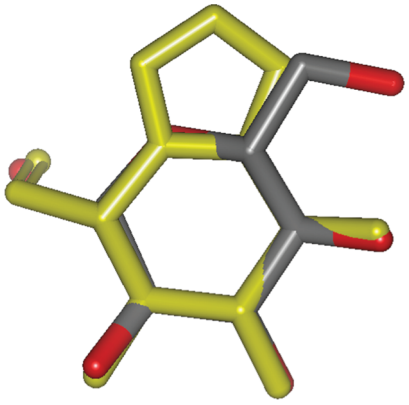


Figure 3. Superimposition of the best scoring conformer of nojirimycine tetrazole (colored by element) onto the selected template ligand [GLC (from 1e55), shown in yellow] by the geometric matching method.

5-iodotubercidin, 5ID) did not show significant 2D similarity to purvalanol, the geometric matching method was able to identify 3D similarity between a purvalanol conformer and the bioactive conformations of the template ligands, as shown in Figure 1.

Pyridoxyl-N,O-cycloserylamide-5-monophosphate (DB03-579). All three targets of this antimicrobial agent that are listed in DrugBank are represented by the ligand database; alanine racemase (alr), UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase (arnB), and aminodeoxychorismate lyase (pabC) (Table 3). While both the 2D and the 3D method were able to retrieve pabC within 1% of the ranked database, the 2D method failed to retrieve alr and arnB. Indeed, while the 3D method was able to retrieve alr and arnB within 0.39% and 0.04% of the ranked database, respectively, the 2D method could only retrieve them within 12.75% and 59.11%, respectively. It is interesting to note that the 2D Tanimoto scores between pyridoxyl-N,O-cycloserylamide-5-monophosphate and the selected ligands, pyridoxyl-alanine-5-phosphate (PP3, for alr) and pyridoxamine-5′-phosphate (PMP, for arnB) were only 0.17 and 0.12, respectively, whereas their 3D Tanimoto scores were much higher at 0.65 and 0.73, respectively. Again, their 3D similarity is evident from the alignment between pyridoxyl-N,O-cycloserylamide-5-monophosphate and each of the template ligands by the geometric matching method (Figure 2).

Nojirimycine Tetrazole. This transition state analogue inhibitor of glycogen phosphorylase has three targets listed in DrugBank, all of which are represented in the ligand database; glycogen phosphorylase, muscle form (PYGM), beta-glucosidase, chloroplast precursor (GLU1), and myrosinase (MA1) (Table 3). PYGM and MA1 were retrieved within 2.5% of the ranked database by both the 2D and the 3D methods. However, while GLU1 was retrieved within 2.5% by the 3D method, it was only retrieved within 41.45% by the 2D method. This is likely to be the result of the inhibitor's relatively low 2D score of 0.18 and relatively high 3D score of 0.53 to the selected template ligand, alpha-D-glucose (GLC). This high 3D score is reflected in the good 3D alignment between nojirimycine tetrazole and alpha-D-glucose by the geometric matching method, as shown in Figure 3.

An Example Where the 2D Method Outperforms the 3D Method

10-Propargyl-5,8-dideazafolic Acid. All four targets of this antineoplastic agent that are listed in DrugBank are represented in the ligand database; bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS) and thymidylate synthase (TS) from *Escherichia coli*, *Lactobacillus casei*, and *Pneumocystis carnii* (Table 4). While DHFR-TS and *P. carnii* TS are retrieved within 1% of the ranked database by both the 2D and the 3D methods, *E. coli* TS and *L. casei* TS are retrieved within 1% of the ranked database by the 2D method only. Indeed, the 3D method only retrieves *E. coli* TS and *L. casei* TS after 2.93 and 5.85% of the ranked database, respectively. While this is not such a big difference as seen in the previous examples, it is of significance because early enrichment is highly desirable. Interestingly, the selected template ligands 10-parpargyl-5,8-dideazafolate-4-glutamic acid (PFG, for *E. coli* TS) and 5-hydroxymethylene-6-hydrofolic acid (THF, for *L. casei* TS) show relatively high 2D scores to 10-propargyl-5,8-dideazafolic acid of 0.75 and 0.47, respectively, yet relatively low 3D scores of 0.28 and 0.26, respectively. The relatively high 2D similarity between 10-parpargyl-5,8-dideazafolate-4-glutamic acid and the two template ligands can be seen form the molecular structures of the compounds shown in Figure 4. All three molecules are quite complex with multiple rotatable bonds and so are likely to be conformationally very flexible. Therefore, the chance of one

Table 4. Example of a Compound Where the ReverseScreen2D Method Outperforms the ReverseScreen3D Method When Retrieving Known Protein Targets[a]

inhibitor	target	ligand	2D score	2D rank (%)	ligand	3D score	3D rank (%)
10-propargyl-5,8-dideazafolic acid	DHFR-TS	F89	0.74	0.08	DHF	0.42	0.16
	TS (*E.coli*)	PFG	0.75	0.05	PFG	0.28	**2.93**
	TS (*L.casei*)	THF	0.47	0.55	THF	0.26	**5.85**
	TS (*P.carnii*)	F89	0.74	0.07	F89	0.31	0.90

[a] All solutions not retrieved in the top 2% are highlighted in bold.

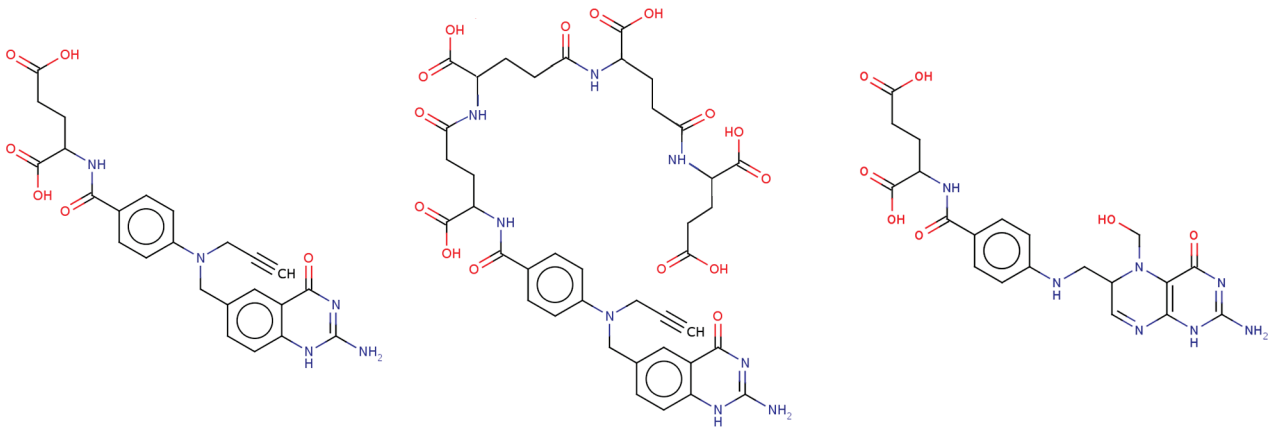


Figure 4. Chemical structure of 10-propargyl-5,8-dideazafolic acid (left) and its selected templates, 10-parpargyl-5,8-dideazafolate-4-glutamic acid (PFG, middle) and 5-hydroxymethylene-6-hydrofolic acid (THF, right).

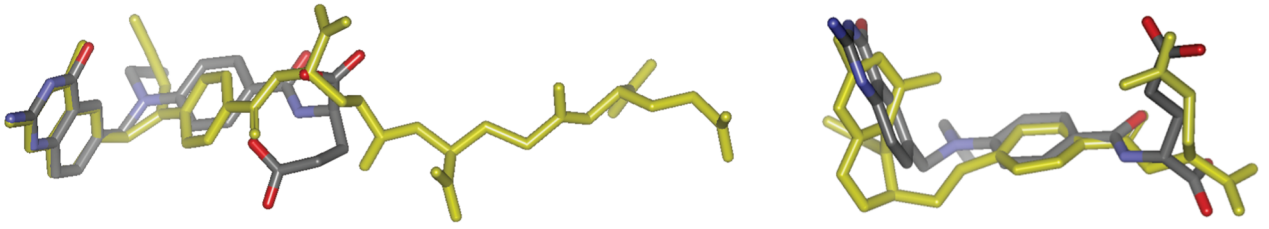


Figure 5. Superimpositions of the best scoring conformers of 10-propargyl-5,8-dideazafolic acid (colored by element) onto each of the two template ligands [left, PFG (from 2bbq) and right, THF (from 2tdd), both shown in yellow] by the geometric matching method.

Table 5. Example of Compounds Where Both the ReverseScreen2D and 3D Methods Fail To Retrieve Known Protein Targets[a].

inhibitor	target	ligand	2D score	2D rank (%)	ligand	3D score	3D rank (%)
nitrofurazone	gor	FAD	0.13	**12.14**	NAD	0.14	**84.46**
	mdh	NAD	0.13	**15.21**	NAD	0.16	**76.18**
	poxB	FAD	0.13	**14.41**	TDP	0.18	**69.04**
tretinoin	ALDH-E1	NAD	0.06	**64.74**	NAD	0.18	**89.67**
	ALDH1A2	NAD	0.05	**91.23**	NAD	0.26	**57.64**
	RARG	156	0.17	**7.74**	156	0.44	**2.66**

[a] All solutions not retrieved in the top 2% are highlighted in bold.

of the conformers of 10-propargyl-5,8-dideazafolic acid exhibiting a good 3D match with either of the bioactive conformations of the template ligands is significantly reduced. Indeed, Figure 5 shows that the 3D superimposition of 10-propargyl-5,8-dideazafolic acid onto each the template ligands is rather poor. Furthermore, the higher molecular weight of PFG, compared with 10-propargyl-5,8-dideazafolic acid means that 10-propargyl-5,8-dideazafolic acid is only able to cover around half of PFG.

Examples Where Both the 2D and 3D Methods Fail

Nitrofurazone. This approved anti-infective agent has four targets listed in DrugBank, three of which are representated in the ligand database; glutathione reductase (gor), malate dehydrogenase (mdh), and pyruvate dehydrogenase, cytochrome (poxB) (Table 5). However, none of these three targets are identified within the top 10% of the ranked database by either the 2D or the 3D method. The 2D method gives the best performance, retrieving gor, mdh, and poxB after 12.14, 15.21,

dx.doi.org/10.1021/ci1003174 |*J. Chem. Inf. Model.* 2011, 51, 624–634

and 14.41%, respectively, whereas the 3D method only retrieves them after 84.46, 76.18, and 69.04% of the ranked database, respectively. The low 2D and 3D scores are hardly surprising given that the selected template ligands FAD, NAD, and TDP are significantly larger and much more complex molecules than nitrofurazone. In fact, the selected template ligands are all cofactors which bind to the targets at different sites from nitrofurazone. The cofactors were selected as template ligands because there were no other appropriate ligands sufficiently similar to nitrofurazone in the ligand database. As a result, the method was unable to identify nitrofurazone's known targets.

Tretinoin. This approved antineoplastic agent has a total of nine different targets listed in DrugBank, three of which have representation in the ligand database; retinal dehydrogenase 1 (ALDH-E1), retinal dehydrogenase 2 (ALDH1A2), and retinoic acid receptor gamma-1 (RARG) (Table 5). While both the 2D and 3D methods are able to retrieve RARG within 10% of the ranked database, they fail to retrieve the other two targets. Indeed, the 2D method is only able to retrieve ALDH-E1 and ALDH1A2 after 64.74 and 91.23%, respectively, while the 3D method is only able to retrieve them after 89.67 and 57.64%, respectively. Again, the poor performance of both methods for these particular targets is due to the selection of cofactors as template ligands, as a result of there being a lack of appropriate ligands sufficiently similar to tretinoin in the ligand database. It is worth mentioning that the case where the method does perform better (i.e., for RARG) is when tretinoin is compared to 4-[3-oxo-3-(5,5,8,8-tetramethyl-5,6,7,8-tetrahydro-napthalen-2-yl)-propenyl]-benzoic acid (PDB ligand code, 156), a suitable ligand in the appropriate binding site.

Current Limitations. The major limitation of the methodology applied in this study is the poor structural coverage of the human genome, as well as other disease-related genomes. For instance, although there are around 6500 unique human protein structures currently present in the RCSB PDB, this only represents less than 30% of protein-coding genes in the human genome. Furthermore, the ReverseScreen3D methodology relies on the crystallographic determination of bioactive ligand conformations, and many of these solved proteins may not have been cocrystallized with a suitable ligand. As mentioned previously, only 6041 out of the 25,439 protein clusters (23.7%) in the RCSB PDB were found to contain one or more biologically relevant ligands. Since not all potential targets of a compound of interest will have been determined experimentally, and even if they have, they may not contain a suitable ligand, it is difficult to ascertain how many targets may have been missed as a result. However, even using the limited structural information that is currently available, we have been able to retrieve the known targets of a number of existing multitarget drugs, therefore demonstrating the value of the methodology at the present time. As the structural coverage of the human genome and other disease related genomes increases as a result of continuing efforts in structure determination, so will the ability of ReverseScreen3D to predict the targets and off-targets of existing and potential drugs on a proteome-wide scale. Indeed, around 150 new structures are deposited in the RCSB PDB every week, and this continuous update is transferred to the ligand database on a quarterly basis.

CONCLUSIONS

We have developed a new reverse VS method called ReverseScreen3D, which can be used to predict the potential targets of a compound of interest from a target database containing only the structurally determined bioactive conformations of known ligands. The target database consists of 8355 ligand binding sites identified in a total of 6041 unique proteins from the RCSB PDB, and each binding site contains one or more biologically relevant, unique ligands. ReverseScreen3D uses a 2D fingerprint-based method to select a ligand template from each of the 8355 binding sites, before carrying out a 3D structural comparison to the selected ligand using a geometric matching method in order to prioritize each binding site in the database. Alternatively, the ReverseScreen method can prioritize each binding site in the database based on the 2D Tanimoto score only, thereby excluding the final 3D comparison step. By carrying out further experimental screening on the prioritized list of targets, it may be possible to determine the potential targets of a new compound or determine the off-targets of an existing drug, thereby revealing the molecular mechanisms of unwanted side-effects or even discovering new therapeutic indications for a drug.

We have evaluated the performance of the ReverseScreen method, both including and excluding the final 3D comparison step, on 20 small molecule protein inhibitors with five or more annotated targets in DrugBank, and shown that both methods are able to successfully retrieve their targets in the majority of cases. When compared with random screening, the 2D and 3D methods both provided a highly significant enrichment of true targets within the most highly ranked targets in the database. While the mean EFs were very similar for both the 2D and the 3D methods at the 2, 5, and 10% levels, the 3D method outperformed the 2D method at the more important 1% level. However, although the 3D method was able to retrieve some targets that the 2D method missed at the 1% level, the 2D method was also able to retrieve a number of targets that the 3D method missed at the 1% level. Furthermore, we have shown that the 3D method was able to retrieve nine out of the eleven known targets of 4OH-tamoxifen within 2% and a further one target within 5% of the ranked database, while the 2D method was able to retrieve four targets within 2% and a further three targets within 5% of the ranked database. Our results for 4OH-tamoxifen compared very favorably with those of three existing reverse VS methods; INVDOCK, TarFisDock, and PharmMapper.

Further investigation of a number of multitarget compounds for which the performance of the ReverseScreen 2D and 3D methods differed, revealed that the 3D method was able to perform well even in the absence of 2D similarity to the template ligands, thereby demonstrating its robustness with respect to purely 2D methods and revealing its potential for scaffold hopping. The opposite situation was also found to be true, although perhaps to a lesser extent. In a very small number of cases, both methods failed to retrieve some of the targets, although this was due to lack of structural representation in the ligand database, rather than a failure of the methodology. Because the ReverseScreen2D and 3D methods each appeared to have their own strengths and weaknesses, we recommended that in order to carry out the most rigorous target search for a compound of interest, both methods should be investigated by the user.

Availability. Both the ReverseScreen2D and 3D methods have been incorporated into a Web server, which is freely available online at http://www.modelling.leeds.ac.uk/ReverseScreen3D. The interface provides both "step-by-step" and "quick submit" functionalities, depending on the needs of the user. The "step-by-step" option guides the user through each step of the

process, from conformer generation, to the 2D similarity search, and then the final optional 3D similarity search. The "quick submit" option allows the user to submit their query compound, select whether or not they wish to include a 3D similarity search, and leave it to run through each of these steps automatically. In both cases, the results of the similarity search(es) can be automatically e-mailed to the user as soon as the process is complete. Note that for the Web server, only database ligands with 2D similarity scores to the query compound that are within the top 10% of the prioritized database are included in the 3D search, thereby significantly reducing the amount of CPU time required at minimal cost to excluding genuine 3D matches. Furthermore, up to 50 conformers of the query molecule are generated instead of up to around 100 conformers, as in the validation process. The current ligand database can be fully screened against a typical query compound using both the 2D and 3D similarity searches within around eight CPU hours. The Web server is automatically updated on a quarterly basis in order to reflect the ever-increasing amount of structural information available in the RCSB PDB.

■ ASSOCIATED CONTENT

Supporting Information. Information about the 20 different multitarget compounds that were selected for the validation procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: r.m.jackson@leeds.ac.uk; phone: +44 (0)113 343 2592; fax: +44 (0)113 343 3167.

■ ACKNOWLEDGMENT

This work was funded by the BBSRC in the form of a studentship for S.L.K.

■ REFERENCES

(1) Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H. TarFisDock: A Web server for identifying drug targets with docking approach. *Nucleic Acids Res.* **2006**, *34* (Web Server issue), W219–24.

(2) Paul, N.; Kellenberger, E.; Bret, G.; Muller, P.; Rognan, D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. *Proteins* **2004**, *54* (4), 671–680.

(3) Gao, Z.; Li, H.; Zhang, H.; Liu, X.; Kang, L.; Luo, X.; Zhu, W.; Chen, K.; Wang, X.; Jiang, H. PDTD: A Web-accessible protein database for drug target identification. *BMC Bioinf.* **2008**, *9*, 104.

(4) Chen, Y. Z.; Ung, C. Y. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand−protein inverse docking approach. *J. Mol. Graph. Modell.* **2001**, *20* (3), 199–218.

(5) Zahler, S.; Tietze, S.; Totzke, F.; Kubbutat, M.; Meijer, L.; Vollmar, A. M.; Apostolakis, J. Inverse in silico screening for identification of kinase inhibitor targets. *Chem. Biol.* **2007**, *14* (11), 1207–1214.

(6) Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M. Drug-target network. *Nat. Biotechnol.* **2007**, *25* (10), 1119–1126.

(7) Paolini, G. V.; Shapland, R. H.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global mapping of pharmacological space. *Nat. Biotechnol.* **2006**, *24* (7), 805–815.

(8) Weber, A.; Casini, A.; Heine, A.; Kuhn, D.; Supuran, C. T.; Scozzafava, A.; Klebe, G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: New pharmacological opportunities due to related binding site recognition. *J. Med. Chem.* **2004**, *47* (3), 550–557.

(9) Hopkins, A. L. Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4* (11), 682–690.

(10) Xie, L.; Li, J.; Bourne, P. E. Drug discovery using chemical systems biology: Identification of the protein−ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* **2009**, *5* (5), e1000387.

(11) Xie, L.; Wang, J.; Bourne, P. E. In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Comput. Biol.* **2007**, *3* (11), e217.

(12) Nobeli, I.; Favia, A. D.; Thornton, J. M. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **2009**, *27* (2), 157–167.

(13) Chong, C. R.; Sullivan, D. J., Jr. New uses for old drugs. *Nature* **2007**, *448* (7154), 645–646.

(14) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: New estimates of drug development costs. *J. Health. Econ.* **2003**, *22* (2), 151–185.

(15) Cai, J.; Han, C.; Hu, T.; Zhang, J.; Wu, D.; Wang, F.; Liu, Y.; Ding, J.; Chen, K.; Yue, J.; Shen, X.; Jiang, H. Peptide deformylase is a potential target for anti-Helicobacter pylori drugs: Reverse docking, enzymatic assay, and X-ray crystallography validation. *Protein Sci.* **2006**, *15* (9), 2071–2081.

(16) Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: A Web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W609–W614.

(17) Muller, P.; Lena, G.; Boilard, E.; Bezzine, S.; Lambeau, G.; Guichard, G.; Rognan, D. In silico-guided target identification of a scaffold-focused library: 1,3,5-Triazepan-2,6-diones as novel phospholipase A2 inhibitors. *J. Med. Chem.* **2006**, *49* (23), 6768–6778.

(18) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462* (7270), 175–181.

(19) Schuffenhauer, A.; Zimmermann, J.; Stoop, R.; van der Vyver, J. J.; Lecchini, S.; Jacoby, E. An ontology for pharmaceutical ligands and its application for in silico screening and library design. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (4), 947–955.

(20) Kinnings, S. L.; Jackson, R. M. LigMatch: A multiple structure-based ligand matching method for 3D virtual screening. *J. Chem. Inf. Model* **2009**, *49* (9), 2056–2066.

(21) 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), 235–242.

(22) 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.

(23) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (mother of all databases). *Proteins* **2005**, *60* (3), 333–340.

(24) Dessailly, B. H.; Lensink, M. F.; Orengo, C. A.; Wodak, S. J. LigASite: A database of biologically relevant binding sites in proteins with known apo-structures. *Nucleic Acids Res.* **2008**, *36* (Database issue), D667–D673.

(25) Strombergsson, H.; Kleywegt, G. J. A chemogenomics view on protein−ligand spaces. *BMC Bioinformatics* **2009**, *10* (Suppl 6), S13.

(26) Kotz, J. C.; Treichel, P., Townsend, J. R. Bonding and Molecular Structure. *In Chemistry and Chemical Reactivity*, 7th ed.; Lockwood, L., McGahey, P., Eds.; Thomson Brooks/Cole: Canada, 2009; Vol. 2, p 387.

(27) 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), 250–260.

(28) Dalton, J. A. R. The Homology Modelling of Protein−Ligand Interactions. Ph.D. Thesis, University of Leeds, 2009.

(29) 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), 1848–1861.

(30) *The Open Babel Package*, version 2.1.1. http://openbabel.org (accessed 07/05/2008).

(31) 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), 991–998.

(32) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34* (Database issue), D668–D672.

(33) Favoni, R. E.; de Cupis, A. Steroidal and nonsteroidal oestrogen antagonists in breast cancer: basic and clinical appraisal. *Trends Pharmacol. Sci.* **1998**, *19* (10), 406–415.

(34) Santner, S. J.; Santen, R. J. Inhibition of estrone sulfatase and 17 beta-hydroxysteroid dehydrogenase by antiestrogens. *J. Steroid Biochem. Mol. Biol.* **1993**, *45* (5), 383–390.

(35) Levine, R. M.; Rubalcaba, E.; Lippman, M. E.; Cowan, K. H. Effects of estrogen and tamoxifen on the regulation of dihydrofolate reductase gene expression in a human breast cancer cell line. *Cancer Res.* **1985**, *45* (4), 1644–1650.

(36) Nuwaysir, E. F.; Daggett, D. A.; Jordan, V. C.; Pitot, H. C. Phase II enzyme expression in rat liver in response to the antiestrogen tamoxifen. *Cancer Res.* **1996**, *56* (16), 3704–3710.

(37) Ritchie, G. A. The direct inhibition of prostaglandin synthetase of human breast cancer tumor tissue by tamoxifen. *Recent Results Cancer Res.* **1980**, *71*, 96–101.

(38) Abbas Abidi, S. M.; Howard, E. W.; Dmytryk, J. J.; Pento, J. T. Differential influence of antiestrogens on the in vitro release of gelatinases (type IV collagenases) by invasive and non-invasive breast cancer cells. *Clin. Exp. Metastasis* **1997**, *15* (4), 432–439.

(39) Lax, E. R.; Rumstadt, F.; Plasczyk, H.; Peetz, A.; Schriefers, H. Antagonistic action of estrogens, flutamide, and human growth hormone on androgen-induced changes in the activities of some enzymes of hepatic steroid metabolism in the rat. *Endocrinology* **1983**, *113* (3), 1043–1055.

(40) Rowlands, M. G.; Budworth, J.; Jarman, M.; Hardcastle, I. R.; McCague, R.; Gescher, A. Comparison between inhibition of protein kinase C and antagonism of calmodulin by tamoxifen analogues. *Biochem. Pharmacol.* **1995**, *50* (5), 723–726.

(41) Messiha, F. S. Leu-enkephalin, tamoxifen and ethanol interactions: Effects on motility and hepatic ethanol metabolizing enzymes. *Gen. Pharmacol.* **1990**, *21* (1), 45–48.

(42) Paavonen, T.; Aronen, H.; Pyrhonen, S.; Hajba, A.; Andersson, L. C. The effect of toremifene therapy on serum immunoglobulin levels in breast cancer. *APMIS* **1991**, *99* (9), 849–853.