Robust Ligand-Based Modeling of the Biological Targets of Known Drugs

Ann E. Cleves and Ajay N. Jain*

UCSF Cancer Research Institute and Department of Biopharmaceutical Sciences, University of California, San Francisco, California 94143

Received November 11, 2005

Systematic annotation of the primary targets of roughly 1000 known therapeutics reveals that over 700 of these modulate approximately 85 biological targets. We report the results of three analyses. In the first analysis, drug/drug similarities and target/target similarities were computed on the basis of three-dimensional ligand structures. Drug pairs sharing a target had significantly higher similarity than drug pairs sharing no target. Also, target pairs with no overlap in annotated drug specificity shared lower similarity than target pairs with increasing overlap. Two-way agglomerative clusterings of drugs and targets were consistent with known pharmacology and suggestive that side effects and drug−drug interactions might be revealed by modeling many targets. In the second analysis, we constructed and tested ligand-based models of 22 diverse targets in virtual screens using a background of screening molecules. Greater than 100-fold enrichment of cognate versus random molecules was observed in 20/22 cases. In the third analysis, *selectivity* of the models was tested using a background of drug molecules, with selectivity of greater than 80-fold observed in 17/22 cases. Predicted activities derived from crossing drugs against modeled targets identified a number of known side effects, drug specificities, and drug−drug interactions that have a rational basis in molecular structure.

Introduction

Discovery of novel lead compounds through computational exploitation of experimentally determined protein structures, either derived from screening of databases or through focused design exercises, is well-established,[1] and methodological development within the docking field remains a very active area of investigation for a large number of research groups.[2−12] Methods for predictive computational modeling of ligand activity in the absence of protein structure have a long history and have also met with important theoretical and practical successes.[13−29] It is certainly desirable to have a high-resolution structure for a protein that is the subject of therapeutic intervention, but frequently one is not available. Further, when considering a different but related question involving the potential *secondary* effects of small molecules, the problems involving absent protein structures become worse.

Proteins whose structure and function depend on localization within cell membranes are the source of a large number of pharmacological effects, and it is unlikely that general methods for solution of structures of these protein classes will be developed in the short term. In terms of their importance as therapeutic targets, membrane-spanning G-protein coupled receptors (GPCRs) and ion channels were the primary biological targets for nine of the top 20 selling prescription drugs worldwide in the year 2000.[30] Each of the proteins that is interesting as a primary target may also be important as a source of side effects. For example, the muscarinic receptors are targeted therapeutically for urinary incontinence, but they also are thought to be primarily responsible for the frequent side effects of dry mouth, urinary retention, and sedation seen with many drugs.[31,32] Other proteins in these classes are *not* the desired targets of drugs, but they have been suggested as the explanation for serious drug side effects. For example, the hERG potassium channel is the likely effector of the lethal side effects of the antihistamine Seldane (terfenadine, which is now withdrawn from human use).[33] Membrane-bound transporter proteins (e.g. P-glycoprotein) and the metabolic enzymes (e.g. cytochrome P450 isoforms) form an increasingly important and well-characterized class of proteins that explain aspects of genetic variation in drug efficacy and many aspects of drug−drug interactions.[34−37] Modern drug discovery, being so precariously dependent on expensive human trials (or postmarket surveillance), would benefit greatly from improvements in our ability to predict drug activity on a scale that would illuminate organism-scale effects.

In this paper, we apply methods that are ligand-focused toward modeling a significant fraction of the space of known drugs, with the goal of demonstrating a first step toward computational prediction of drug side effects and drug−drug interactions. We employed four computational methods to model drug/drug similarities, target/target similarities, and to construct models of the binding requirements of the targets. The methods are described briefly in this paper (each having been validated in other reports):

1. Morphological Similarity. Given query and object molecules, this method rapidly optimizes the pose of the query to maximize 3D similarity to the object molecule. We have shown previously that the computation has the property that pairs of molecules judged to be similar tend to bind the same proteins.[19]

2. Molecular Imprinting. For computing very large numbers of pairwise similarities, it is computationally efficient to rerepresent molecular structure as a vector of similarities to a fixed set of basis molecules. Distances between these vectors are used as a surrogate for the more expensive similarity computation.[38,39]

3. Optimal Multiligand Superpositioning. Given a small number of competitive ligands for a protein binding site, this method produces an optimal superposition, maximizing pairwise similarities while minimizing total volume.[18]

4. Ligand-Based Virtual Screening. Given a superpositioning of multiple molecules that form a hypothesis as to their preferred binding mode to a target (a *model*), this method functions as a docking program to rank a set of input molecules according to their degree of fit to the model.[18]

* Correspondence author. Phone: (415) 502-7242. Fax: (650) 240-1781. E-mail: ajain@jainlab.org.

10.1021/jm051139t CCC: $33.50
Published on Web 04/22/2006

These methods and the antecedent Compass method have been the subject of a number of validation studies and have been used successfully in lead discovery and optimization, in the context of proprietary discovery projects as well as published work.[13−15,17]

The present paper addresses two questions that the prior work has not. First, will the methods yield robust results when applied to a significant fraction of the space of known drugs? Second, can systematic modeling of therapeutically relevant biological targets based solely on their known ligands make it possible to rationalize the off-target effects of drugs or possibly to *predict* them? These two questions required a curation effort to identify the specific biological targets of the space of small molecule human therapeutics. The task was challenging, since much of the pharmacology literature has focused on narrow chemical structural classes and is descriptive (as opposed to mechanistic) in describing pharmacological effects. We identified linkages between drugs and their primary and secondary biological targets, covering 979 small molecule drugs in all (of roughly 1000 approved therapeutics). Given this information, we made two types of computations. The first involved drug/drug and target/target similarities, computed on the basis of ligand structures. The second involved the induction of ligand-based models for 22 diverse targets whose performance would be quantified with respect to virtual screening and selectivity. We report results in three areas.

1. Drug and Target Similarities. Comprehensive comparison of drug/drug pairs and target/target pairs formed the basis for clustering and for direct analysis of distributions of pairwise similarities. Drug pairs sharing a target had significantly higher similarity than drug pairs sharing no target, and target pairs with no overlap in annotated drug specificity shared lower similarity than target pairs with increasing overlap.

2. Ligand-Based Virtual Screening. Ligand-based models of 22 diverse drug targets were constructed in a fully automated computation. Virtual screening experiments testing the ability of the models to identify cognate drugs against a background of screening molecules showed excellent enrichment in the great majority of cases.

3. Selectivity of Ligand-Based Models. Selectivity of the 22 models was tested by measuring the ability of the models to identify cognate drugs from a background of other drug molecules, many of which constituted easily confusable classes. Enrichment results quantitatively paralleled those in the virtual screening experiments. Analysis of the high-ranking putative false positives yielded a number of cases where there is a specific biological explanation for the predicted cross-talk among the ligands used to construct a model and the nominal nonligands found to fit the model well.

Taken together, these results represent a substantial validation of our ligand-based modeling and molecular similarity methods. They also mark a first step toward systematic computational modeling of a large enough fraction of pharmacologically relevant targets to support practical hypothesis generation of side effects and drug interactions in preclinical drug discovery.

The software that implements the algorithms described here is available free of charge to academic researchers for noncommercial use (see http://www.jainlab.org for details on obtaining the software). Molecular data sets presented herein are also available.

Methods

The following describes the methodology used in this paper, molecular data sets, detailed computational procedures, and quantification of performance.

Computational Methods. The computational methods used here have been reported in previous methodological papers focusing on molecular similarity,[19] fingerprint-based chemical indexing,[38,39] and ligand-based modeling[18] and will be described only briefly here.

Morphological Similarity. Given query and object molecules, this method rapidly optimizes the pose of the query to maximize 3D similarity to the object molecule. Figure 1A illustrates the computation. Morphological similarity is defined as a Gaussian function of the differences in molecular surface distances of two molecules at weighted observation points on a uniform grid, yielding a value from 0 to 1. The distances that drive the computation are depicted as the lines at the right of Figure 1A after the query molecule has been aligned to the object molecule. The surface distances computed include both distances to the nearest atomic surface (black lines) and distances to donor and acceptor surfaces (blue and red, respectively). In the case shown, the molecules are competitive nicotinic agonists, and their optimal similarity is 0.94, reflecting highly similar surfaces despite differences in scaffolding (an oxazole versus a pyridine) that can confound 2D methods.

The function is dependent on the relative alignment of two molecules, and the algorithm for optimizing the similarity of one molecule to the fixed conformation of another makes use of the observation points (illustrated at the left of the figure). The alignment problem can be addressed with an efficient algorithm, because the molecular observations that underlie the similarity function are local and are not dependent on the absolute coordinate frame. So, two unaligned molecules or molecular fragments that have some degree of similarity will have some corresponding set of observers that are "seeing" the same things. Optimization of the similarity of two unaligned molecules is performed by finding sets of observers of each molecule that form triangles of the same size, where each pair of corresponding points in the triangles are observing similar features. The transformation that yields a superposition of the triangles will tend to yield high-scoring superpositions of the molecules. The problem of flexibly aligning one molecule onto another is addressed with a divide and conquer algorithm, making use of molecular fragmentation and incremental construction to ameliorate the exponential dependence of conformational space on the number of rotatable bonds. The overall computation is roughly linear in the number of rotatable bonds within the query molecule, taking a few seconds per bond on standard desktop hardware. Additional details of the method can be found in two previous reports.[18,19]

Molecular Imprinting. In making large-scale computations of molecular similarity, even with fast methods, the problem posed in all-by-all computations is computationally challenging. For the work here, some computations involved on the order of 1 million such comparisons, and a surrogate computation was employed making use of molecular *imprints*. The idea is to make use of a small *basis set* of molecules to which to compute molecular similarity for a large number of molecules. The process is illustrated in Figure 1B, making use of 20 basis molecules. Each input molecule is flexibly aligned to a fixed conformation of each of the basis molecules. For each input molecule, the result is a 20-dimensional vector, where each value within each vector represents a single similarity computation (yielding a value between 0 and 1).

Distances between these vectors can be used as a computationally cheap surrogate for the direct molecular similarity computation. In Figure 1B, three molecules are shown that all target the serotonin reuptake transporter (among other things).

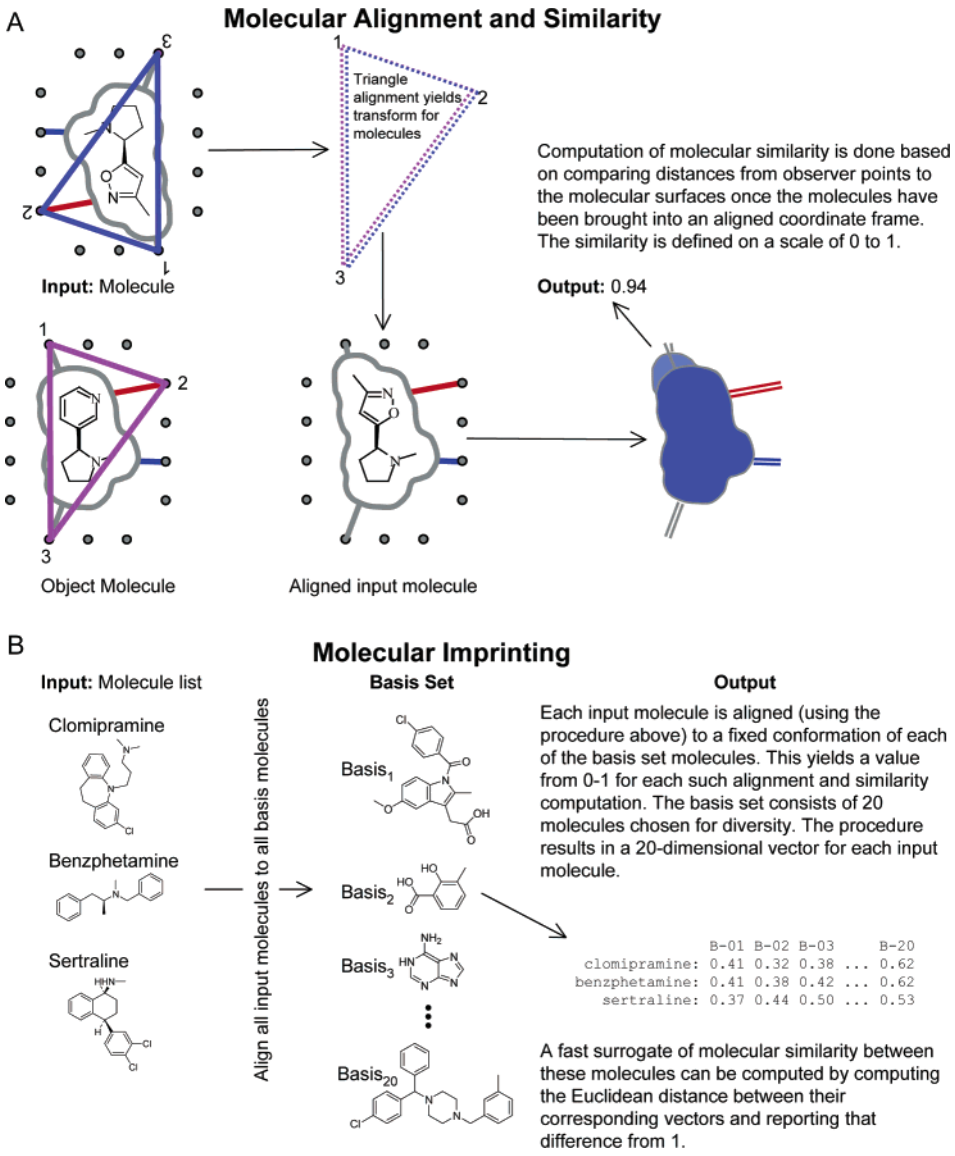

Figure 1. Molecular similarity and molecular imprinting methods (see the text for details).

Their similarities to the first three basis molecules are relatively low, but their similarity to the last is higher. The pattern of similarities represented within the vectors gives rise to a correlation between similarity inferred on the basis of the Euclidean distance of vector pairs and the corresponding direct computation of pairwise molecular similarity. Additional details on this method and its application to molecular diversity and bioavailability computations can be found in two previous reports.[38,39] The basic concept has been exploited by other groups as well.[40,41]

In addition to offering a method for comparing two molecules directly, we employed this method to compare two targets, based on the structures of their cognate ligands. For each pair of targets A and B (where A and B may be the same target), we compute all pairs of imprint distances of the cognate ligands of A to the cognate ligands of B. Comparisons of a ligand to itself are omitted (arising from either overlap in drug specificity between different targets or from self/self target comparison). We defined the target similarity as the 80th percentile of the ligand similarities between the targets. This was done to avoid strong dependence on outliers (which tend to skew the mean similarity in an unpredictable manner) and to focus the similarity of targets on *shared* similarity in ligands. If a significant proportion of the ligands of A are similar to the ligands of B, the target similarity is high, even if a fraction of the ligands are very different.

Optimal Multiligand Superpositioning. Given N molecules that are mutually competitive at a protein's ligand binding site as input, the object of the superpositioning method is to produce a joint superposition that is predictive of the relative bioactive poses of the input molecules. The method combines the morphological similarity function (described above) with a term to minimize overall joint molecular volume. The space of joint molecular poses is searched to maximize an objective function. The objective function is the product of (1) the sum of all pairwise similarities and (2) the total empty volume in a sphere of fixed size centered on the superimposed ligands. This biases the solutions of joint superposition to the smallest possible volume, given equivalent joint similarities. In the remainder of the paper, such superpositions will be referred to as *models* when they are used as the targets for virtual screening.

Figure 2A illustrates the methodology using three competitive μ-opioid agonists as input. The output of the procedure is a list of high-scoring overlays of the input molecules (by default 100 are provided). The highest scoring hypothesized superposition is depicted in the figure. Identifying the proper relative alignment

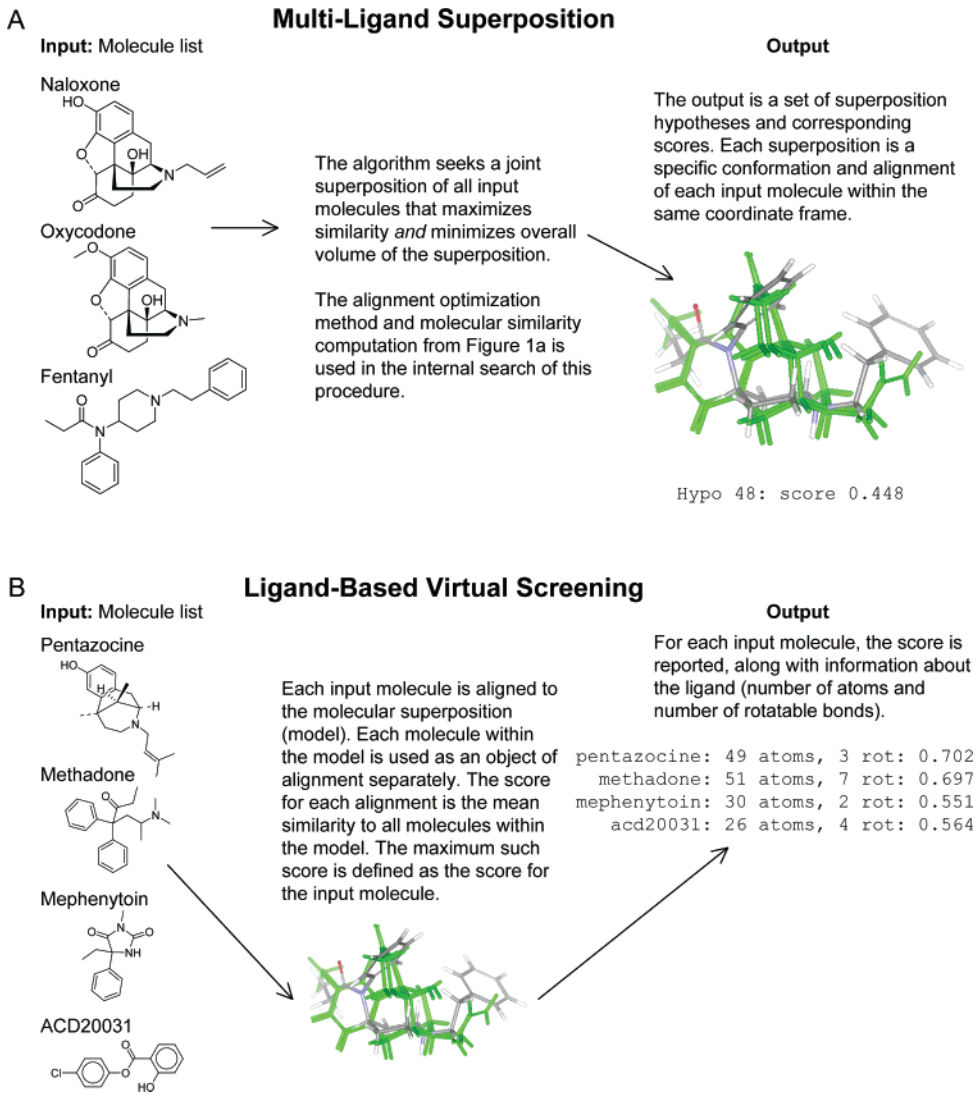

Figure 2. Generation of ligand-based model and use as a model for virtual screening (see the text for details).

of the morphinans (shown in green at right) is not challenging, but fentanyl (shown in atom color), which has a very different chemotype, is more difficult. This particular model will be discussed in some detail later. Our previous report provides additional details on the method and presents the results of screening enrichment experiments on four targets of therapeutic interest.[18]

Ligand-Based Virtual Screening. A model such as that shown in Figure 2A can be used for visualization purposes in medicinal chemistry design exercises. Such models may also serve as the target for virtual screening, in much the same was as a protein structure is used with docking algorithms. The input is the model and a list of molecules to be screened, with the object being to rank molecules on the basis of their ability to mimic the surface displayed by the ligands within the model. Given a model consisting of *M* ligands and a query ligand, the procedure yields a score between 0 and 1 along with the specific pose of the query ligand that gives rise to the reported score. The query ligand is flexibly aligned to maximize similarity to each of the *M* ligands in the model separately, resulting in a pool of poses. For each of the query ligand poses within the pool, the mean of the similarity score to the *M* model ligands is computed. The maximum such score is defined as the score of the query ligand and is returned along with the corresponding pose. So, the score of a new ligand is intended to reflect its ability to mimic, using a single pose, the model that is represented by the joint superposition of all *M* molecules.

Figure 2B depicts this process, using the model from Figure 2A and four query molecules. The first two are competitive μ-opioid agonists, the next is a drug that does not target the opioid receptor family, and the last is a typical screening molecule. The output of the procedure is shown at right of the figure, which lists some information about the query molecules and their final scores. Both pentazocine and methadone score significantly higher than the noncognate ligands. Quantification of the degree of separation the models achieve between cognate drugs and large numbers of screening compounds and noncognate drugs will be presented later. Additional details about the use of models as the targets for virtual screening, including a comparison to 2D methods, can be found in a previous report.[18]

Molecular Data Sets. We identified 1125 small molecule agents approved for human use through the National Drug Code Directory, which serves as a repository for universal product identifiers for human drugs in the United States (http://www.fda.gov/cder/ndc/). This list is dominated by agents that are commonly considered therapeutics, but there are examples of insecticides (e.g. permethrin), bulk nutrients (e.g. glucose), vitamins, and other small molecules that are not the focus of the present study. We focused on those agents whose primary use is therapeutic and whose desired biological target is either

a human protein or a protein within a viral, bacterial, or fungal human pathogen. Where possible, the biological effectors of secondary effects of the drugs were also identified.

In keeping with the postgenomic molecular characterization of biochemical networks, we sought to annotate the biological effectors of pharmacological effects down to specific binding sites on assemblies of gene products, making use of public resources such as Entrez Gene for definitive naming of specific protein subunits. In the easiest cases, a single human gene product was identified. For example, the primary target of over 20 small molecule drugs is the opioid receptor μ, which is officially named OPRM1 within Entrez Gene (GeneID 4988).[42] In other cases, a common target name such as the "$GABA_A$ receptor" corresponds to a pentameric assembly of multiple gene products, commonly the following: GABRA1, GABRB2, and GABRG2. The binding site for benzodiazepines is thought to be a cleft between the γ_2 and α_1 subunits, while the endogenous ligand GABA binds between the α_1 and β_2 subunits.[43]

These distinctions become critical in computational experiments, as the implicit assumptions frequently include competitive binding among a set of ligands. So, the fact that barbiturates and benzodiazepines both modulate the activity of the $GABA_A$ receptor *and* the fact that their binding sites are separate are part of our curated information. Of the 1125 drugs, we have annotated the primary (desired) targets of 979 and, when possible, have indicated secondary targets as well (which are generally responsible for side effects). Overall, we have identified 271 targets, many of which are the pharmacological effectors of multiple drugs. Roughly 25 primary targets cover 400 drugs, 60 cover 600, and 85 cover over 700. In the Results, we make distinctions between primary desired effects by referring to "primary" targets and side effects by referring to "secondary" ones.

In our computational experiments, we focused on drug targets for which we have identified the largest number of competitive small molecule drugs. Molecular preparation protocols (detailed below) had some impact as well on the ligands considered. For the results presented, we used the ligands of a set of 48 targets in an all-by-all molecular similarity computation. Figure 3 shows examples of ligands for 22 of these targets, which formed the basis for our ligand-based modeling computations. Both the targets and the chemical scaffolds of their ligands are diverse. Targets A−E are all proteins within bacterial, viral, and fungal pathogens, with cognate drugs including azole antifungals, β-lactam antibiotics, sulfa drugs, quinolones, and nucleoside analogues. Targets F, H, and I are diverse but are all involved in cardiac indications. Targets L−O are steroid receptor targets, including both those involved in inflammation and those whose natural ligands are the sex hormones. Targets J, K, and P are all involved (though quite differently) in analgesia, with Q and R involved in sedation. GPCRs are represented by I, J, S, and T. Within this set of targets, there is diversity in function and in ligand characteristics, but there are subsets of targets where structural overlap exists among the proteins themselves as well as their cognate ligands (e.g. the steroid and GPCR cases).

Diversity is an important feature in the computations that follow, with the goal being that the methods work well across all classes of targets and small molecules. Subtleties between related targets with related ligands are also important, with the goal being that the modeling approach will yield sufficiently specific results that, for example, androgens would not be confused with estrogens.

Computational Procedures. We used the same procedures as in our report on the Surflex-Sim ligand-based modeling method.[18] Briefly, all molecules were subject to the same preparation procedures, which involved automatic protonation, ring search, protonated nitrogen inversion, and minimization using a Dreiding-type force field. Up to 10 conformations were retained for each molecule, to account for alternative ring conformations and protonation geometries. It is important to note that the molecular superposition methods described above sample the conformational space of the ligands much further than the initial sampling used to identify energetically reasonable ring geometries, but the on-line search is currently limited to acyclic bonds, necessitating this two-step approach. The ACD screening set[7] originally contained 990 molecules, and of these, 850 were correctly processed and used as a negative control in screening enrichment experiments. The computations involving known drugs included 979 molecules, which were processed in exactly the same way as the screening compounds in order to avoid any systematic difference between the drugs and nondrugs. Of these, 230 represented the known cognate ligands for the 22 targets shown in Figure 3.

Following preparation of the molecular data, Surflex-Sim (version 1.31) was used for molecular imprinting and ligand-based hypothesis generation and testing. Generation of the molecular imprints followed the default practice ("Surflex-Sim vector LigandList BasisList ImprintFile"). The imprints computed for the 979 molecules along with the basis set of molecules are part of the data archive associated with this paper.

We generated the molecular superpositions for the 22 test cases using standard ligand-based hypothesis generation procedures and default parameters ("Surflex-Sim hypo InputMoleculeList log"). For each case, this resulted in up to 100 scored superpositions. For each target, the top scoring superposition was selected as the model for testing against two different chemical libraries: (1) the cognate drugs plus 850 screening ligands and (2) the cognate drugs of the target in question along with drugs of the other 21 targets. In the latter case, this provided a more rigorous test of the methodology with respect to selectivity than if we used all 979 drugs, which would have decreased the proportion of potentially confusable chemical structures. This also focused the background on a well-characterized set of drugs.

To evaluate the utility of these models, the two screening libraries were tested, again using standard procedures and parameters ("Surflex-Sim align_list TestLibrary HypoList logtest" where HypoList contained the pathnames to the mol2 files comprising the highest scoring superposition). The score of a ligand against a model was the maximum mean similarity of a single pose of the ligand to the individual molecules comprising the model. So, the scores reflected the extent to which a ligand could best mimic the joint superposition of molecules within a model.

Quantification of Performance. Evaluation of the results of the computations emphasizes the enrichment of known ligands over other ligands based on a ranking generated from virtually screening libraries consisting of cognate ligands mixed within a background of other ligands (as seen in a number of recent reports of both docking and molecular similarity[3−7,9,18]). Quantification of the degree of separation between true positive ligands and false positives was done by using receiver operating characteristic (ROC) curves along with the corresponding areas underneath the curves. Given a set of scores for positives and negatives, the ROC curve plots the true positive proportion (Y axis) with the corresponding false positive proportion (X axis) at all possible choices of some threshold that would mark a binary distinction between a prediction of positive or negative

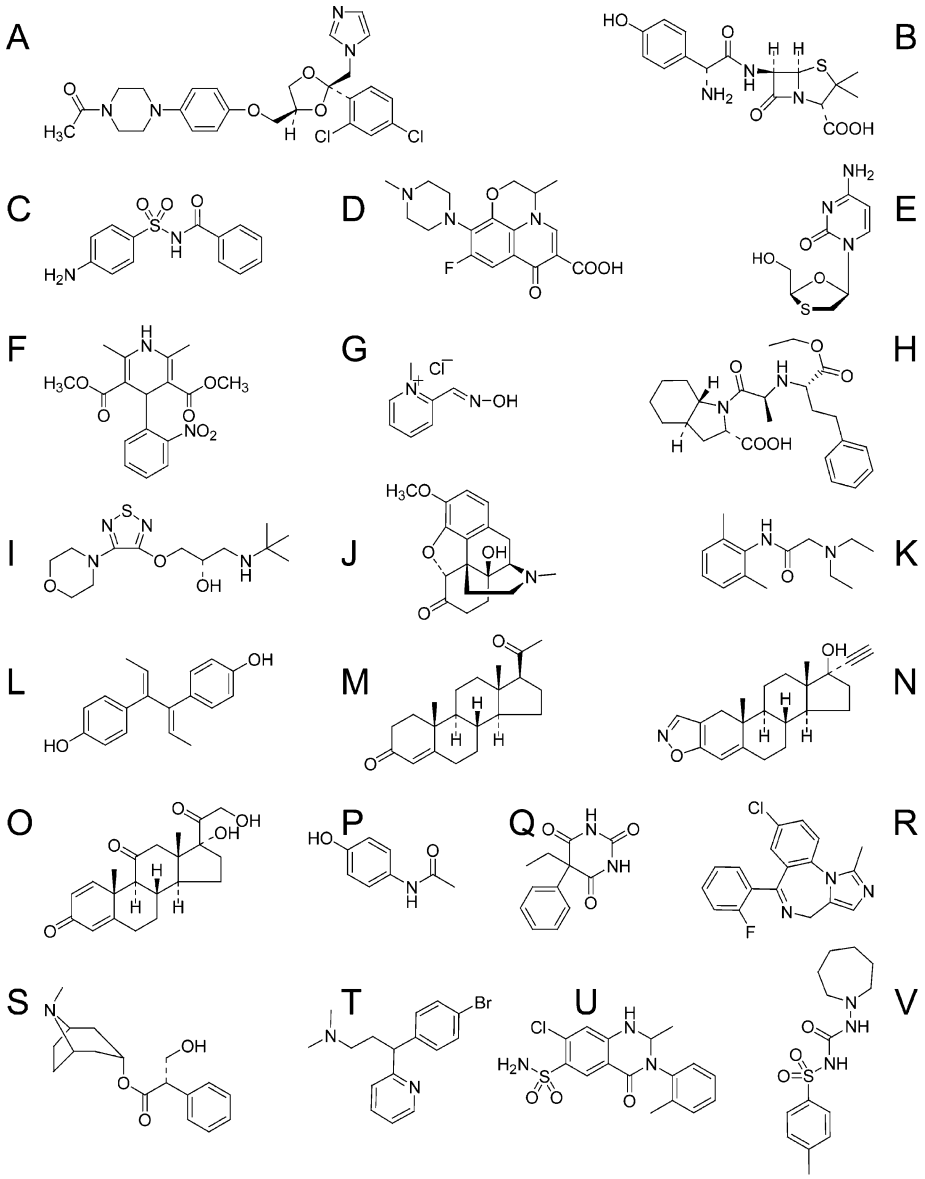

Figure 3. Examples of compounds used for model construction for each of 22 different biological targets. For each compound, the target name and compound name are given: (A) lanosterol demethylase, ketoconazole; (B) D-Ala-D-Ala carboxypeptidase, amoxicillin; (C) dihydropteroate synthase, sulfabenzamide; (D) DNA gyrase, levofloxacin; (E) HIV reverse transcriptase, lamivudine; (F) L-type calcium channel, nifedipine; (G) acetylcholinesterase, pralidoxime; (H) angiotensin I converting enzyme, trandolapril; (I) β-1,2,3-adrenergic receptor, timolol; (J) opioid receptor μ, oxycodone; (K) voltage-gated Na^+ channel, lidocaine; (L) estrogen receptor, dienestrol; (M) progesterone receptor, progesterone; (N) androgen receptor, danazol; (O) gluco/corticosteroid receptor, prednisone; (P) COX-I COX-II, acetaminophen; (Q) $GABA_A$ receptor barbiturate site, phenobarbital; (R) $GABA_A$ receptor benzodiazepine site, midazolam; (S) muscarinic acetylcholine receptor, hyoscyamine; (T) histamine receptor, brompheniramine; (U) NaCl cotransporter renal, metolazone; (V) sulfonylurea receptor, tolazamide.

class membership. The perfect ROC curve goes from [0,0] to [0,1] to [1,0] and results in an area of 1.0. Complete intermixing of positive and negative scores gives an area of 0.5, with areas less than 0.5 reflecting the case where true positives are ranked lower than false positives.

We also report screening enrichment values, which have a more intuitive interpretation. The result of a virtual screening exercise, in practice, is to take a small percentage of the top-ranked compounds and test them experimentally for activity against the target of interest. Theoretical enrichment rates (the fold excess of observed hits to expected hits given a selected subset of a library) are computable from the data that underlie ROC analyses. Enrichment rates are dependent on the proportion of the library chosen for screening, which is based on the score threshold applied to define the subset. With large libraries, enrichment rates simplify to the ratio between true and false positive rates at different proportions of the top ranked molecules.[18] Maximal enrichment values are typically seen with the very highest ranked molecules within the library.

Results

In what follows, we present three primary results, based on application of the four methods described above (see Methods for details about the data sets, computational methods, and specific procedures).

Drug and Target Similarities. Due to the size of our data sets, pairwise computation of molecular similarities required on the order of a million individual ligand/ligand similarities. Rather than employ the morphological similarity method directly, we employed the surrogate molecular imprinting approach (which is much faster) to infer similarities in these experiments. Our previous work focused on the use of this technique for

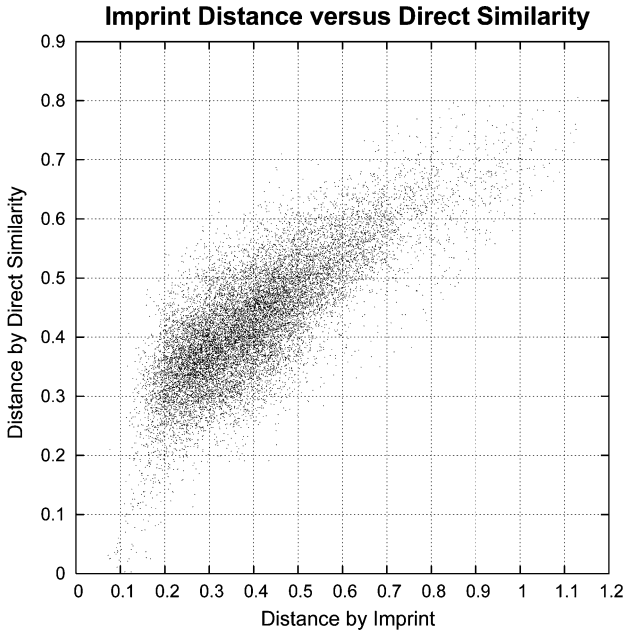

Figure 4. Plot of the relationship between pairwise molecular distance computed by imprint and by direct molecular similarity computations. The Pearson correlation coefficient is 0.79. Molecular pairs with close distances measured by imprint identify molecular pairs that have close distances measured by molecular similarity, and vice versa. In particular, for the closest 10% of pairs by similarity, over 80% of the imprint distances are within the lowest quartile of imprint distances overall. Also, for the closest 10% of pairs by imprint, over 80% of the similarity distances fall within their lowest quartile overall. For large distances, the correspondence in pairs that are identified by each method is even tighter. Thus, imprint distances may be used as a fast surrogate computation in place of direct molecular similarity.

computations involving screening compounds.[38] Consequently, for this work, we wanted to verify that the method yielded the expected results within the space of small molecule drugs. Figure 4 shows a plot comparing the distance between pairs of molecules computed by direct molecular similarity and by the surrogate of imprint distance for each pair. We computed over 15 000 pairwise distances among the 979 drugs in the present study. The overall correlation between the two methods was 0.79 by Pearson correlation. Importantly, molecular pairs with close distances measured by imprint identify molecular pairs that have close distances measured by direct molecular similarity, and vice versa. Thus, imprint distances may be used as a fast surrogate computation in place of direct molecular similarity.

Clustering. These fast distance computations are particularly useful in clustering applications. Figure 5 shows a two-way hierarchical clustering of 48 drug targets and their cognate drugs (single-linkage hierarchical agglomerative clustering with an optimization of the rendering order of the dendrogram).[44] The intertarget distances were computed from the Euclidean distances between the imprints of the ligands of the respective targets. The interdrug distances were computed by Euclidean distance between the imprints of molecule pairs. The Methods section contains additional details.

At the top of Figure 5, the full target and drug clustering are shown, with the target dendrogram at left and the drug dendrogram at the top. Below that, two subsets are enlarged, with the orientation rotated clockwise. For the target clustering, note that the tree structure induced is in the same spirit as classical pharmacology, with the characterization of the effects of drugs being driven by drug structure. In our clustering, targets that group together within common subtrees have ligands that are similar under the imprint-based distance metric, with the analogous observation for drugs that group together. If it were the case that the computed distances between ligands were unrelated to the biological effectors of their pharmacology, we would not observe the formation of blocks of black in the two-way clustering.

These blocks, in the target dimension, indicate a series of drugs that all bind the same target where the imprint distances between the drugs was sufficient to lead to the grouping. We observe a number of sensible target groupings. For example, the steroid receptors that are targets of the sex hormones (androgen, estrogen, and progesterone) segregate tightly, with the glucocorticoid and mineralocorticoid receptors also clustering together. We observe a number of the amine-typeGPCRs segregating, with the muscarinic and histamine receptors grouping closely together, as expected from the frequent overlap among the ligands of the targets.

In the enlargement of the two small subsets of the drugs from the full dendrogram, not all targets are populated with cognate drugs, since the drugs hit a subset of the 48 targets overall. We observe a striking enrichment of drug groupings with overlapping annotated targets. All of the drugs within the top enlarged subtree share at least one target: the serotonin (5HT) reuptake transporter. These drugs include some first-generation tricyclic antidepressants (e.g., clomipramine) that have broad effects against many targets. The drugs also include sertraline (Zoloft), sibutramine (Meridia), and benzphetamine (Didrex), whose effects are substantially more specific against the transporters. In clinical practice, sibutramine and benzphetamine are used for weight loss, with sertraline used as an antidepressant. Note that the structures of the drugs within this group exhibited wide structural diversity, but the methods used for structure comparison were not dominated by 2D structural differences among the drugs.

Within the lower half of the bottom enlarged block, we see a separate group of psychopharmaceuticals typified by promazine (Sparine), with the chief difference being a lack of activity against the reuptake transporters of the top block. Several of these are used as antipsychotic agents, but the agents have a wide variety of effects against a number of biological targets, and the specific selectivity profiles define their clinical uses. The top half of the bottom block includes primarily first-generation antihistamines, such as diphenhydramine (Benadryl), which have almost universal muscarinic side effects. With one exception, all of the drugs within the bottom block were annotated as including effects against the histamine H1 receptor, with nearly all sharing the muscarinic receptor as a documented target. The single apparent outlier (based on annotation) within the bottom block is methadone, which is a μ-opioid receptor agonist used clinically in treatment of opiate dependence. However, methadone's side effects include dry mouth, urinary retention, sweating, and reduced bowel motility,[45] which are all associated with muscarinic activity.[46] In the context of this molecular similarity driven group assignment, the μ-opioid ligands were among the least well cosegregated, but we believe that much of the dispersion can be explained by widely varying side effects of the drugs owing to disparate off-target specificities. Note also that the μ-opioid receptor itself segregated away from the histamine and muscarinic receptors as well, despite methadone's placement.

Distributional Analysis of Pairwise Similarities. Clustering diagrams can be very useful as visualization tools and may give rise to suggestive observations, but they do not directly support

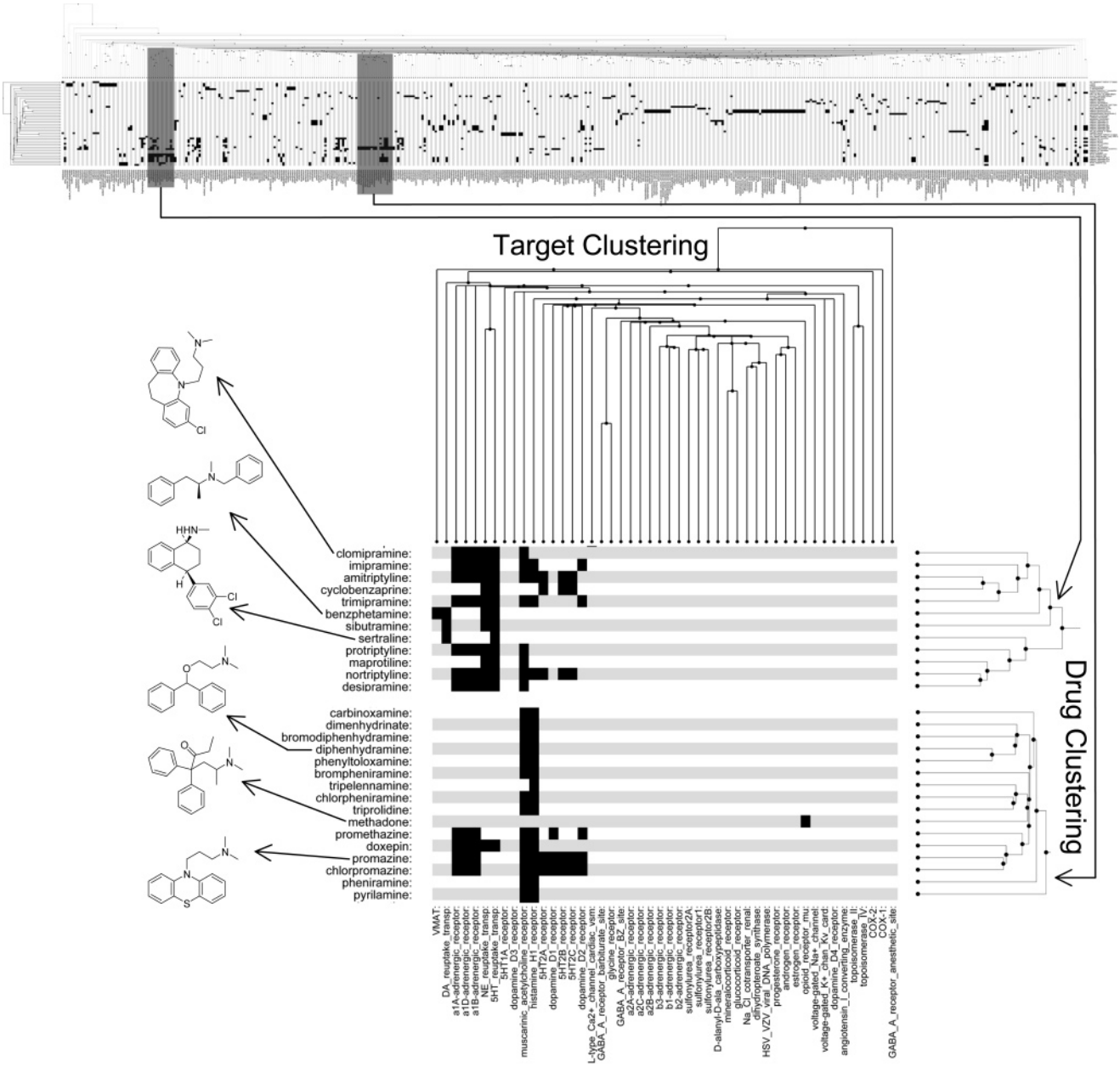

Figure 5. Two-way hierarchical clustering of drug targets and drugs. A full clustering of 48 targets and their cognate drugs is shown across the top. The two shaded areas are depicted below (rotated clockwise), with the target dendrogram across the top and two portions of the drug dendrogram at the right. The target dendrogram results from considering the imprint-based distances of the ligands of each of the 48 drug targets, with the ligands of each target considered as a group. The drug dendrogram results from considering each drug individually. Black blocks indicate that a particular drug has a known effect against a particular target. The appearance of blocks of black, both in the vertical and horizontal directions, indicates that the targets and the drugs segregate sensibly on the basis of considerations of molecular structure alone.

quantitative conclusions. We have shown previously that the morphological similarity metric is well-correlated with competitive ligand binding,[19] and we presented data above that the imprint-based distance computation is a good surrogate, but the *direct* question of whether the imprint-based surrogate similarity metric will yield higher similarities for drug pairs that share a target than for drug pairs that do not has not been formally addressed. Figure 6A shows the cumulative histogram of the two relevant distributions of imprint-based similarities. The distribution of pairwise similarities for drugs sharing at least one target is shifted significantly to the right ($p \ll 0.01$ by *t*-test). This is the quantitative reason behind the appearance of black blocks in the drug dimension of the clusterings shown in Figure 5.

We carried out a similar computation for comparing target pairs by constructing five groups of target pairs. The first such group consisted of target pairs that shared no annotated drug overlap (936 total pairs), the second consisted of pairs that shared low overlap (defined as 1−19%, 82 total pairs), the third consisted of pairs with medium overlap (20−79%, 92 pairs), the fourth consisted of pairs with high overlap (80−99%, 13 pairs), and the fifth consisted of self/self pairs (100% overlap; 48 "pairs"). An example of a target pair with no annotated drug overlap was the glucocorticoid receptor and topoisomerase II; low overlap was exemplified by the μ-opioid receptor and the muscarinic acetylcholine receptor; medium overlap was exemplified by the histamine and muscarinic receptors; high overlap was exemplified by COX-I and COX-II.

Recall from the Methods that ligand identities arising from drug overlap between targets or from self/self target comparisons are not included in the target similarity computation. Figure 6B shows the cumulative histograms for all five target pair groups,

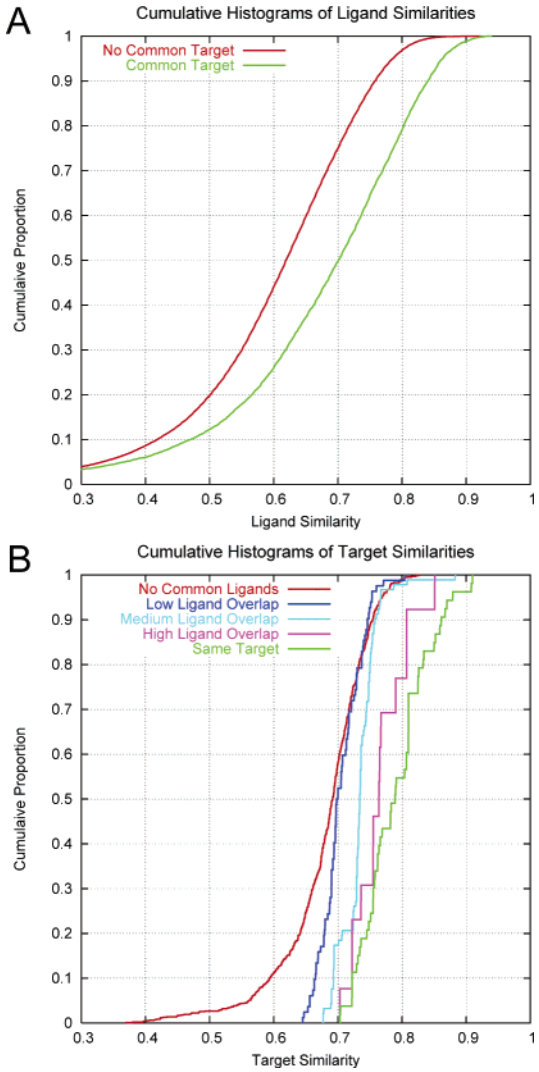

Figure 6. Cumulative histograms showing the degree of separation between drug/drug similarities and target/target similarities under different conditions. Panel A illustrates the separation between drug pair similarities for drugs that share a target (green) and drugs that share no target (red). The similarities for drugs that share a target are significantly higher as a population ($p \ll 0.01$ by t-test). Panel B illustrates a similar feature for target similarities inferred from their cognate drugs' similarities. The red curve depicts the intertarget similarities for targets that are annotated as sharing no cognate drugs (936 target pairs). The green curve depicts the self/self target similarity (48 targets total). Blue, cyan, and purple depict, respectively, the intertarget similarities for low, medium, and high overlap pairs of targets (82, 92, and 13 pairs total).

with similarity increasing monotonically from no target overlap through each of the cases with increasing target overlap. The differences between the no overlap pair set with all other sets are highly statistically significant ($p \ll 0.01$ by t-test). The no-, low-, and medium-overlap pair sets compared with the same target pair set were similarly significant. The high overlap case compared with the case of the same target distribution was not significant by t-test at $p = 0.05$. Note that there are a number of examples where no annotated overlap exists, for example, androgen and estrogen receptor pair, but where both in the clustering of Figure 5 and in the computation of target distance here (the target similarity was 0.81) the computational method suggests significant overlap. Such cases will be considered further in the Discussion.

Ligand-Based Virtual Screening. We built ligand-based models of each of the 22 targets, prototypical ligands for which are shown in Figure 3. These targets had the largest number of known drugs from our curation effort, so it was possible to induce models based on superpositions of two or three drugs for each target while having enough remaining cognate drugs to test each model. In each case the drugs used to construct the models were chosen randomly. Choice of two versus three drugs was based on the total number of identified drugs. The Methods section has additional details regarding model construction.

The issue of chemotype diversity in retrieval of cognate ligands based on very limited information in model induction deserves attention, and the μ-opioid receptor is a suitable example. Figure 7 shows the results of model induction for the μ-opioid receptor. Three molecules were used; two were morphinans (naloxone and oxycodone) and one was not (fentanyl). Recall from the discussion above that pure molecular similarity computations did not result in aggregation of all μ-opioid ligands into a single subtree of our clustering. Considering the structural diversity present within the drugs, this should not be surprising. Whereas the very rigid morphinan derivatives cannot display much variation in molecular surface, molecules such as fentanyl can. Notwithstanding this diversity, the superposition of fentanyl onto the moprhinans in the model construction is convincing. The amine functionality is perfectly superimposed, with the carbonyl oxygen corresponding to an important hydrogen-bond acceptor, based on the structure–activity relationships evident from the other known ligands. The hydrophobic portions of fentanyl are also well-matched to the morphinan volumes.

While the superposition itself is convincing, the proof of utility lies in the ability of such a model to yield a ranking of molecules in a virtual screen where true ligands are ranked above nonligands. We conducted a screen of a library against the model, where the library included the cognate μ-opioid ligands mixed with a set of screening compounds (see Methods for details). Figure 7 shows the structures of six different cognate μ-opioid ligands, ranked by their position in the screen. The number above each ligand is the percentile within the ranking. The ranking illustrates that the model is sufficiently accurate to identify even nonmorphinans at very low false positive levels. Computation of a full ROC curve based on the scores of the cognate ligands and nonligands yielded an area of 0.982 with a maximal enrichment of 283 of cognate ligands over nonligands at the top of the ranking.

Figures 8 shows the corresponding ROC plots for each of the 22 targets for which models were constructed. In all but three cases, retrieval of over 70% of the true positives was achieved with false positive rates of less than 5%. Table 1 reports the ROC areas and enrichment rates for the 22 screening runs. The maximal enrichment rates exceeded 100-fold in 20/22 cases. These results compare favorably with the best reported performance of docking methods.[4,9]

Selectivity of Ligand-Based Models. A common, and not unreasonable, criticism of virtual screening experiments constructed as just described is that the background molecules may not be druglike and therefore represent an easy case for measuring enrichment. Further, such experiments do not address a key issue in drug design, that of selectivity. The question is whether a computational model will appropriately distinguish between confusable ligands (e.g., androgen versus estrogen receptor ligands). To address this issue, we ran an additional 22 virtual screens as above, but we employed *noncognate drugs* as the background for each of our models. These were the annotated drugs of the 21 other targets. We had expected this to be a challenging task, given the presence of confusable ligands

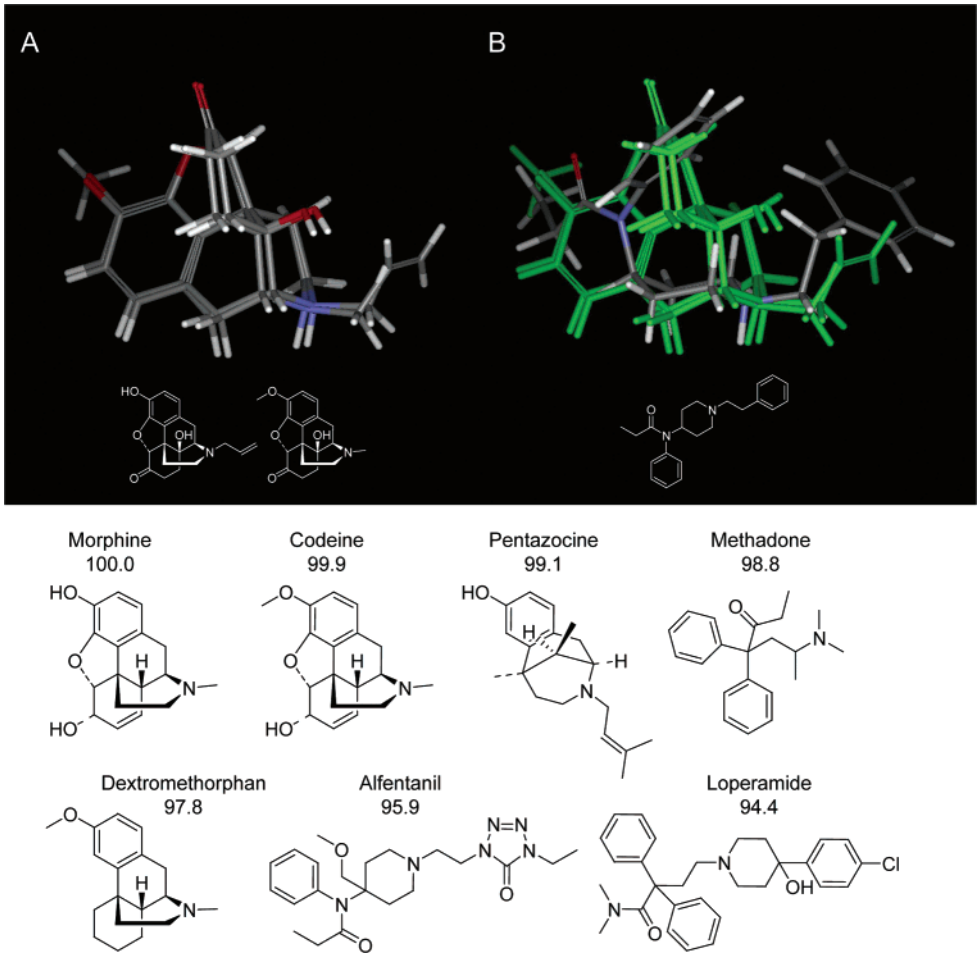

Figure 7. Model based on opioid receptor μ ligands. Panel A shows the superposition of naloxone and oxycodone, both used in the model, and classic opioid ligand structure (morphinan derivatives). Panel B superimposes fentanyl, which is a non-morphinan. It is a competitive agonist and was also used in the model. The molecules shown below the graphic are all μ-opioid ligands and were tested against the three-ligand superposition shown in B using two different background chemical libraries. Above each molecule is the name along with the percentile ranking among the background of screening compounds.

Table 1. Enrichment of Cognate Drugs Using a Screening Compound Background for 22 Different Biological Targets

target name	max. enrichment	ROC area
L-type calcium channel	850	1.000
histamine receptor	850	1.000
$GABA_A$ barbiturate site	744	0.999
NaCl cotransporter	708	0.998
HIV reverse transcriptase	425	0.998
$GABA_A$ benzodiazepine site	765	0.991
progesterone receptor	142	0.991
dihydropteroate synthase	773	0.988
DNA gyrase	283	0.986
D-Ala-D-Ala carboxypeptidase	447	0.985
muscarinic acetylcholine receptor	304	0.983
β-adrenergic receptor	283	0.983
acetylcholinesterase	283	0.982
opioid receptor μ	283	0.982
lanosterol demethylase	283	0.977
sulfonyl urea receptor	213	0.971
angiotensin I converting enzyme	567	0.962
estrogen receptor	121	0.955
gluco/corticosteroid receptor	63	0.941
voltage-gated sodium channel	155	0.925
androgen receptor	121	0.908
COX-I/COX-II	9	0.831

within the background. For example, in the case of the μ-opioid model (shown in Figure 7), the presence of many ligands of amine-type GPCRs within the screening library presented a potential challenge. However, the ROC area and maximal enrichment were 0.980 and 121-fold, both comparable to the results above using screening compounds as the background. On the basis of the observations from the clustering exercise where μ-opioid ligands were spread out among the drugs instead of being tightly segregated, the specificity of the μ-opioid model is somewhat surprising. It appears that by inducing a model that requires *simultaneous* similarity to a specific conformation of each of multiple superimposed ligands, we are better able to segregate cognate ligands than in the case where we are asking a less constrained question about molecular similarity. Put more concretely, methadone fits very well into the model of μ-opioid activity, but methadone can also look like the ligands of other targets as well, whereas other opioids cannot (notably the relatively rigid morphinans).

Figure 9 shows the ROC plots using the background of noncognate drugs in the screens, which quantify model selectivity. The results are very similar to those shown in Figure 8, both in absolute terms and with respect to the rank order of performance of the models. Table 2 reports the ROC areas and enrichment rates of these specificity screens, with 17/22 exceeding 80-fold enrichment. Antihistamines were perfectly separated from both the screening compounds and from the other drugs, notably including antimuscarinic compounds. Steroid receptor models were generally very successful in avoiding confusion among the different steroid activity classes. The

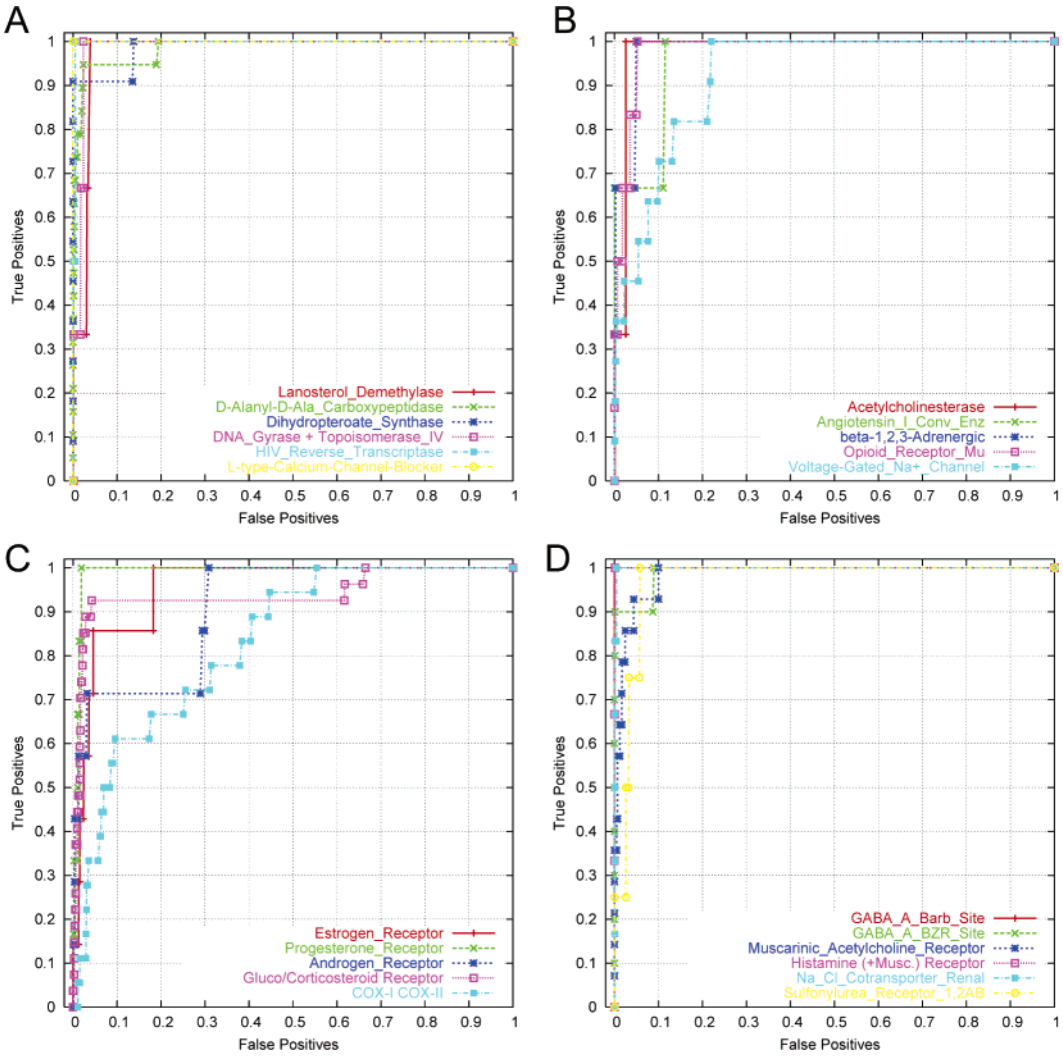

Figure 8. ROC plots reflecting the enrichment of cognate ligands against a background of screening compounds.

Table 2. Selectivity for Cognate Drugs over Noncognate Drugs for 22 Different Biological Targets

target name	max. enrichment	ROC area
L-type calcium channel	246	1.000
histamine receptor	245	1.000
$GABA_A$ barbiturate site	209	0.995
NaCl cotransporter	241	0.999
HIV reverse transcriptase	123	0.996
$GABA_A$ benzodiazepine site	212	0.982
progesterone receptor	80	0.981
dihydropteroate synthase	237	1.000
DNA gyrase	81	0.971
D-Ala-D-Ala carboxypeptidase	192	0.984
muscarinic acetylcholine receptor	85	0.962
β-adrenergic receptor	82	0.963
acetylcholinesterase	82	0.980
opioid receptor μ	121	0.980
lanosterol demethylase	81	0.986
sulfonyl urea receptor	61	0.902
angiotensin I converting enzyme	163	0.898
estrogen receptor	51	0.946
gluco/corticosteroid receptor	187	0.931
voltage-gated sodium channel	21	0.857
androgen receptor	34	0.855
COX-I/COX-II	5	0.732

weakest retrieval is seen with the COX-I/II model. This is not terribly surprising, since the NSAIDS (typified by aspirin, acetaminophen, and naproxen) not only display divergent specificity for the COX-I/II enzymes but they display a host of different side effects. For example, both aspirin and acetaminophen are nonspecific with respect to COX-I/II, but the former has significant gastrointenstinal bleeding complications and cardioprotective effects that the latter lacks.

Discussion

Our results represent an expansion and generalization of the validation of the four computational methods used. The morphological similarity and molecular imprinting approaches exhibited intuitive behavior when applied both to segregation of drugs and drug targets, both in a qualitative sense in the clustering and quantitatively when considering the underlying distributions. The computations involving the Surflex-Sim ligand-based modeling and virtual screening methods are a substantial test of such an approach, with explicit models built that cover roughly one-quarter of approved small molecule therapeutics.[18] Our focus in previous work was methodological, and we showed that the Surflex-Sim methodology quantitatively outperformed 2D methods, but the validation was limited to four targets.[18] In the present work, the 22 biological targets that were the subject of modeling represent a broad diversity of biology and pharmacology. Further, the structural diversity of drugs in most of the cases was qualitatively as high as in our previous report. The performance we observed paralleled that reported earlier. To achieve 60−70% recovery of known cognate ligands, typically between 1 and 5% of the random screening ligands would be found as false hits. This level of performance

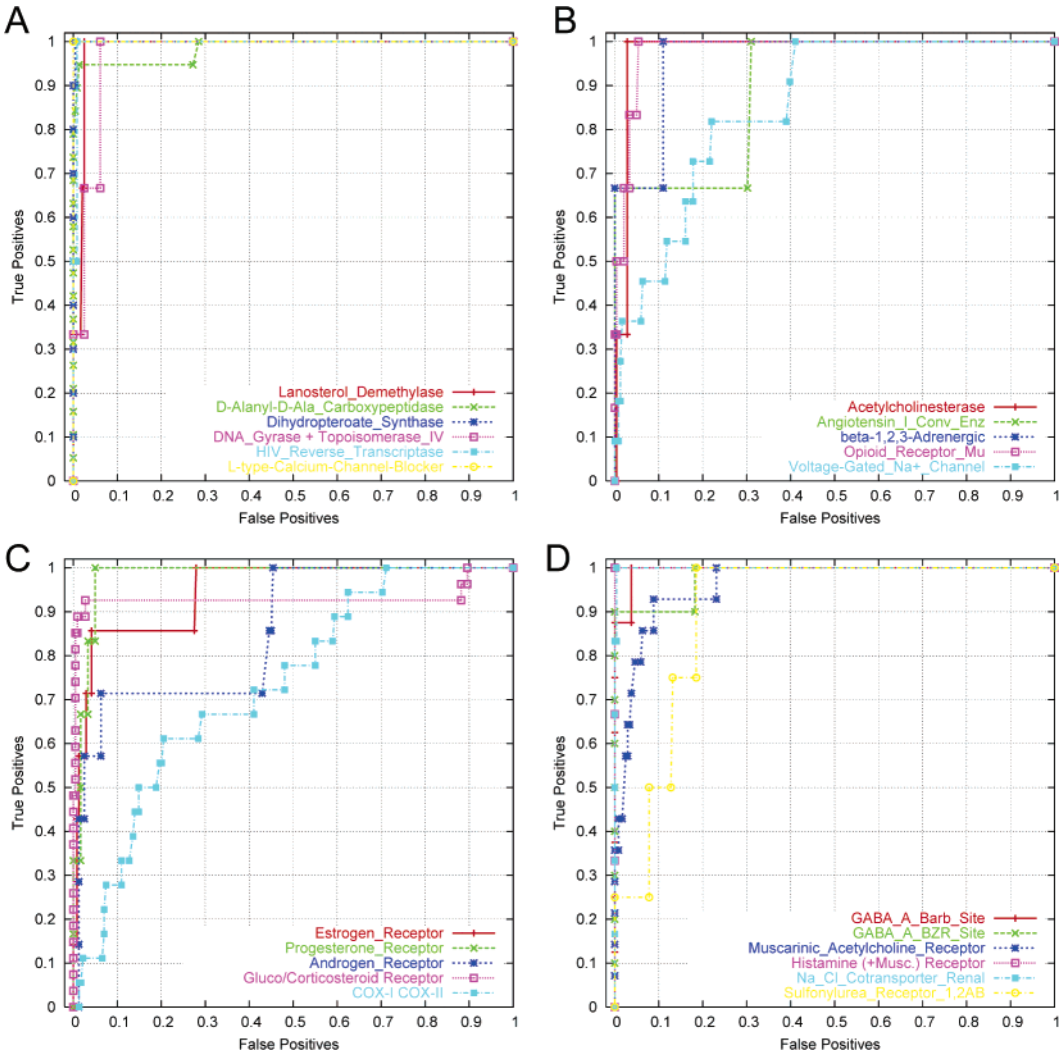

Figure 9. ROC plots reflecting the enrichment of cognate ligands against a background of other drugs.

is competitive with that of the best available docking methods when exploiting well-determined protein structures.[2−5,7,9] As before, cognate ligands with widely differing chemotypes were identified at very low false positive rates. Robust performance in this large-scale test has a significant practical impact, offering a highly automated method for predictive modeling for lead identification and optimization in cases where protein structures are unavailable.

Robust modeling of many targets offers an additional potential benefit. One of the most challenging aspects of modern drug discovery is the extent to which non-target-related side effects must be discovered through clinical trials or, worse, through clinical practice. Side effects and drug−drug interactions are observed with a great number of therapeutic drugs on the market today. These undesirable activities may stem from specific binding to an unintended target and may occur at any point in the absorption, distribution, metabolism, and elimination of drugs. For many drugs (and almost certainly for drugs in the modern discovery process), the primary (cognate) biological target has been identified, but secondary (noncognate) targets, transport routes, and metabolic pathways often remain unspecified. Consider what could be gained by systematic modeling of as many targets of pharmacological effects as possible, given the available data regarding the biological effectors of such effects (both identities and structures where available) and the respective ligands. With sufficiently accurate models of a large enough proportion of pharmacologically relevant proteins, computational experiments might reveal hypotheses about undesirable effects of drugs that are testable using in vitro methods.

We are not claiming that our methodology is fully up to this challenge, but there is some reason for optimism. The optimism derives partially from our observations about the breadth of applicability and quantitative sensitivity and specificity of the models and similarity methods. However, observations relating to the nominal *mistakes* of the methods contribute as well. For example, the target clustering shown in Figure 5 shows a number of groupings directly supported by annotated overlaps in their cognate drugs, but we also observe some groupings without such support. For example, we see the sex hormone nuclear receptors grouped together despite no annotation of cross-talk among the receptors and noncognate ligands. It turns out that there are a number of documented examples of androgen ligands binding the estrogen receptor and vice versa.[47] A more incongruous grouping places a cardiac potassium channel near the several amine-type GPCRs. But recall from the Introduction the established effects of terfenadine (a histamine antagonist) against hERG.[33] This is a voltage-gated potassium channel, as is the potassium channel seen in the target clustering among the amine-type GPCRs. These examples suggest that putative overlaps that are revealed by considering ligand similarities might reveal biologically relevant pharmacology.

In this vein, in addition to considering the quantitative separation of cognate from noncognate drugs for each of our

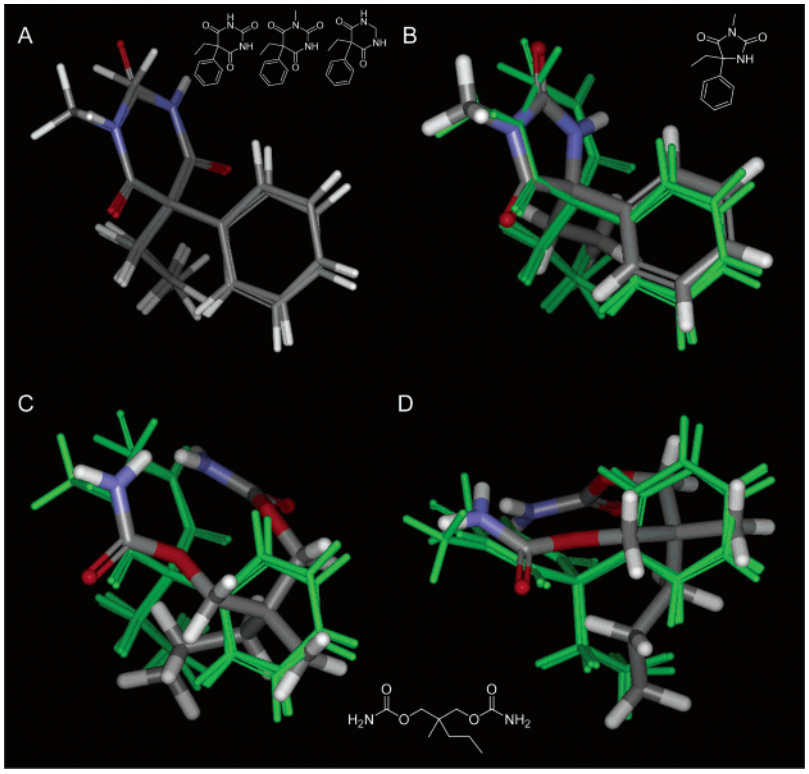

Figure 10. Model based on $GABA_AR$ barbiturate site ligands. Panel A shows the superposition of phenobarbital, mephobarbital, and primidone, which were the three ligands used to induce the model. Panel B shows the superposition of mephenytoin onto the model (model ligands in green). Panels C and D show the superposition of meprobamate onto the model from two slightly different orientations (the view in D is tilted back). The lack of structural variation in the barbiturates leads to excellent statistical performance of the model against both chemical library backgrounds. However, the model is able to identify drugs of different chemotypes that have been shown to have overlapping effects (see the text for details).

22 models, we also analyzed the composition of the top-ranking *noncognate* drugs. In the case of any particular model (target A) where multiple ligands of another target turned up as high-scoring (cognate ligands of target B), there were four interesting situations: (1) primary target overlap, where the ligands of targets A and/or B actually bind to both targets; (2) tertiary target overlap, where the ligands of A and B each are known to bind target C, causing unintentional pharmacological effects; (3) drug transporter overlap, where the ligands of A and B share active transporter proteins; and (4) drug metabolism overlap, where the ligands of A and B share enzymatic metabolic machinery. The following discussion illustrates several examples of these overlaps, each of which relate to a side effect or drug−drug interaction of clinical significance.

Case 1: Primary Target Overlap. Primary target overlap is the instance in which two drugs both bind and affect the same biological targets (as in the sex hormone nuclear receptor case alluded to above). Our first example of target overlap concerns barbiturates and sodium channel (SCN) antagonists, two classes of drugs that target ligand-gated ion channels. The $GABA_A$ receptor ($GABA_AR$) is a pentameric GABA-gated chloride channel with binding sites for barbiturates, benzodiazepines, and the natural ligand GABA. SCNs are heterotrimers consisting of a channel-forming, 24-transmembrane domain α subunit with two regulatory β subunits. SCN antagonists such as antiarrhythmics, local anesthetics, and anticonvulsants bind a common receptor site on the SCN α subunit, albeit in a nonidentical manner.[48]

Three barbiturates were used to generate the $GABA_AR$ model, and their structures and resulting superposition are shown in Figure 10A. The annotated ligands of $GABA_AR$ exhibited relatively little structural diversity, and six cognate barbiturates were the highest scoring molecules in the screen. However, among the 14 highest scoring molecules in the screen were four SCN antagonists, specifically the anticonvulsants phenytoin, mephenytoin, ethotoin, and felbamate. The converse effect was also observed. Three SCN drugs were used to generate the SCN model: two anticonvulsants (mephenytoin, topiramate) and a Class Ib antiarrhythmic (lidocaine). The 27 highest scoring ligands included all nine barbiturate drugs in the set. Figure 10B shows the superposition of mephenytoin onto the $GABA_AR$ model.

It has been established that the $GABA_AR$ and SCN drugs overlap in their effects on the two proteins. The three barbiturates used to generate the $GABA_AR$ model have anticonvulsant actions as well as the traditional sedative effects of other barbiturates. Phenobarbital and pentobarbital have been shown to inhibit SCN function like the anticonvulsant phenytoin.[49,50] Further, pentobarbital has been shown to be an antagonist of SCNs in human brain and heart.[51−53] The barbiturates and the prototypical SCN antagonists are nearly 100 years old,[54] and the early pharmacology was, of course, phenomenological. The primary targets for these drugs were well-established by the early 1990s.[51,52,55] The effects of the barbiturates on SCNs were established roughly concomitantly.[49−53] The multidecade gap between pharmacological and biological characterization was mostly due to the slow evolution in biological investigation methods, but even within the context of modern biology, we believe that comprehensive computational modeling will help to suggest direct experiments that will *accelerate* the linkage between pharmacological observation and specific biology.

The anxiolytic meprobamate was included in our noncognate drug screen, having been initially annotated as targeting the benzodiazepine site of $GABA_AR$.[46] In fact, meprobamate was the next highest scoring molecule after the barbiturates within the $GABA_AR$ barbiturate site model. Meprobamate was syn-

thesized and became pharmacologically characterized in the 1950s, with its clinical effects being similar to those of benzodiazepines.[46,54] Figure 10, panels C and D, show meprobamate superimposed onto the $GABA_AR$ barbiturate site model. Meprobamate is a nonbarbiturate chemically, but it has been shown (i) to activate directly the $GABA_AR$ in a barbiturate-like manner,[56] and (ii) to enhance allosterically benzodiazepine binding to the $GABA_AR$ in a manner similar to that of barbiturates.[57] It may be experimentally challenging to show direct competitive binding of meprobamate with barbiturates to $GABA_AR$, but the idea is well motivated by the computational results illustrated in Figure 10.

Our second example of primary target overlap was observed using our models for three GPCRs: muscarinic acetylcholine receptor (mAChR), histamine H1 receptor (H1R), and μ-opioid receptors (muR). As seen from the ROC analysis above, these models were highly accurate in separating true positives from false positives with both background sets. However, these three models scored each other's ligands as the majority of the top 25 molecules. There is ample evidence that this computational overlap reflects real biological effects. For example, many first-generation antihistamines such as brompheniramine, chlorpheniramine, and diphenhydramine are H1 receptor antagonists but also have central and peripheral antimuscarinic side effects, such as sedation and dry mouth.[58,59] The antimuscarinic effects of some H1 antagonists have been quantified in binding assays; mequitazine, cyproheptazine, clemastine, diphenylpyraline, promethazine, homochlorcyclizine, and alimemazine have high affinity for muscarinic receptors with dissociation constants ranging from 5 to 38 nM.[60] Even third-generation antihistamines such as desloratadine have been shown to be potent muscarinic antagonists in competitive assays and to have antimuscarinic effects in vivo, albeit at doses greater than recommended for antihistamine therapy.[32,61]

The μ-opioids show variation in the degree of muscarinic side effects such as dry mouth and urinary retention, and this was reflected in our mAChR screen. Variation in the side effects of the class of compounds was also reflected in our clustering analysis, with the compounds segregating in several separate subgroups instead of within a single one. The μ-opioids fentanyl and morphine were both highly ranked in our mAChR screen. Fentanyl has been shown to be a muscarinic antagonist in a competitive binding assay,[62] and activation of spinal muscarinic receptors is thought to contribute to the analgesic effect of morphine.[63] The opioid loperamide was scored low using our mAChR model, and interestingly, dry mouth is a less common side effect of this drug (except in cases of overdose).[58]

Since assays for direct binding and for functional activity are increasingly available, we believe that the use of large-scale predictive computational modeling of target activity may offer a practical method for producing testable hypotheses in preclinical drug development. While it may be prohibitive to experimentally screen for hundreds of different biological activities, the cost of screening for a much smaller number based on predictions of the sort we have described may be feasible.

Case 2: Tertiary Target Overlap. Another case we observed in our noncognate drug screen was similar to primary target overlap but involves neither the primary target of the drugs used to construct a model nor the primary target of the high-scoring noncognate drugs. What we term a tertiary target is a common *secondary* target to ligands of both primary targets. Our first example from above involving the barbiturate $GABA_AR$ agonists and SCN antagonists shows a tertiary target overlap in addition to their primary target overlap. The barbiturates phenobarbital and pentobarbital and the SCN antagonists felbamate, phenytoin, and topiramate have been shown to inhibit glutamatergic NMDA receptors.[50,64−68]

There is a subtlety in uncovering these tertiary target effects that is worth discussing. In the case of primary target overlap, we are uncovering exactly what the models are supposed to uncover: competitive ligands at the site being modeled (e.g. the antihistamine clemastine binding the muscarinic receptor). In the case of tertiary target overlap, an additional benefit is the uncovering of effects that have their roots in the lack of perfect specificity inherent in defining a functional model of a protein binding site based on the structures of its ligands. When the ligands for a protein are not perfectly specific and share common secondary targets, the model we induce will likely identify ligands of both targets. This is quite different from the situation we would see if we had a well-determined structure of our primary target and were making use of an effective docking strategy. In this case, we would identify primary target overlap (e.g., mephenytoin binding the barbiturate $GABA_AR$ site). But in cases where there is *no* primary target overlap between two ligands, but there is overlap with a tertiary target, a protein structure-based method would reveal no linkage between the two ligands.

The second example of a potential common tertiary target falls into this category and was observed with cyclooxygenase (COX) inhibitors and nucleoside antivirals. COX-I and -II both catalyze an endoperoxide synthase and a peroxidase reaction in the conversion of arachidonate to the prostaglandin precursor PGH2. COX-I and -II are inhibited by a set of chemically diverse molecules (collectively known as NSAIDS), the majority of which are organic acids that act as reversible competitive inhibitors.[46] Despite vast structural diversity in the COX screen, several cognate NSAIDs were among the 25 highest scoring molecules. However, there was a surprising high-scoring chemical class. Six antiviral nucleoside analogues were among the 30 highest scoring molecules. These included three nucleoside reverse transcriptase inhibitors (nRTIs; didanosine, lamivudine, stavudine) and three HSV DNA polymerase inhibitors (acyclovir, ganciclovir, and trifluridine). There is an unusual side effect in common among these nucleoside drugs, which do not appear to have *any* direct COX-I/II effects; they are all associated with mitochondrial damage and apoptosis. The gastric mucosal cell death effects of NSAIDs have been proposed to involve the mitochondrial apoptotic pathway.[69] The same has been suggested for ganciclovir-associated cytotoxicity.[70]

While the conjecture is somewhat speculative, we suggest that the observed overlap among the NSAIDs and the nucleoside analogues may involve shared targets in the human apoptotic pathway, possibly involving modulation of Bcl-2 or Bcl-XL. The NSAIDs sulindac, indomethacin, mesalazine and COX-II-selective inhibitors have been shown to induce apoptosis in colon carcinoma cells in vitro.[71−74] The chemotherapeutic nucleoside analogue gemcitabine is a potent inducer of apoptosis. While it was not included in our noncognate drug database, it scored as well as acyclovir within the COX model. Resistance to gemcitabine-induced apoptosis is conferred by high expression of Bcl-2 or Bcl-XL,[75,76] and it has been suggested that direct inhibition of Bcl-2 or Bcl-XL function should serve as a novel strategy for pancreatic cancer therapy.[77] Further, induction of apoptosis has been suggested as a mechanism of mitochondrial toxicity resulting from HIV therapy involving nRTIs.[78,79]

It is possible that these seemingly unrelated classes of drugs (NSAIDs and nucleoside antivirals) all induce the apoptotic pathway by binding to Bcl-2 or Bcl-XL proteins. Due to the

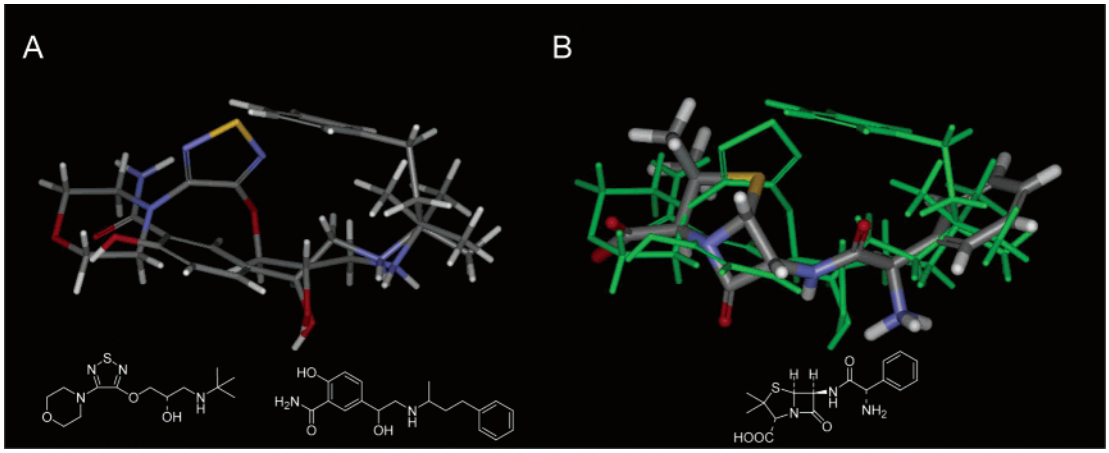

Figure 11. Model based on β-adrenergic site ligands. Panel A shows the superposition of timolol and labetalol, which were the two ligands used to induce the model. Panel B shows the superposition of ampicillin onto the model (model ligands in green). While it does not appear to be the case that ampicillin directly affects the β-adrenergic receptors, it appears that both the adrenergic ligands and ampicillin are substrates for some of the same transporters (see the text for details).

central role of Bcl-2 and Bcl-XL in apoptosis, binding assays have been developed, and small molecules that bind Bcl-2 and Bcl-XL are being investigated as potential new cancer therapeutics.[80] Our data suggesting a tertiary target overlap, coupled with evidence from the literature, point to a potentially specific shared role for NSAIDs and nucleoside antivirals in the modulation of apoptosis. We propose that direct interaction with Bcl-2 and Bcl-XL would be an interesting avenue to investigate.

Case 3: Drug Transporter Overlap. The preceding two cases shared a similar feature: a linkage between two drugs would manifest by some shared pharmacological effect due to modulation of a shared biological target. The case of drug transporter overlap manifests as drug−drug interactions involving dosage. These linkages share the feature with tertiary target overlap that docking methods are likely to be less effective in uncovering them than ligand-based approaches. Small molecule transport proteins are expressed in many tissues, such as intestine, brain, liver, and kidney, and can affect one or more stages in the adsorption, distribution, and elimination of drugs.[81] For example, the enteric transporter PEPT1 mediates intestinal absorption of small peptides, amino β-lactam antibiotics, and other peptide-like drugs.[82] Conversely, ATP-dependent transporters such as the p-glycoprotein (P-gp) and MRP2, which are located on the apical brush border membrane and have very broad substrate specificity, can decrease intestinal absorption of drugs via efflux into the intestinal lumen.[81] Drugs that interact with the same transporter can affect each other's pharmacokinetic profiles through competition for or inhibition of the transporter. Since some transporters have very broad substrate specificity, it is not particularly significant when our models identify sets of ligands that are their joint substrates. These interactions typically manifest as *increases* in the concentration of one of the drugs involved in such an interaction. We present an example where a specific set of structural features may explain a drug−drug interaction that manifests as a *decrease* in the concentration of drug over time.

We observed this in the screen of our β-adrenergic receptor (β-AR) model. The β-AR model was generated using the β-blockers labetolol and timolol. The model performed well, with cognate β-AR antagonists (nadolol and metipranolol) as the highest scoring molecules in the screen. Figure 11A shows the derived superposition. Surprisingly, six β-lactam antibiotics were also among the 25 highest scoring ligands in the β-AR test. Figure 11b shows ampicillin superimposed onto the two ligands comprising the model. Note that the amine functionality superimposes well, particularly in the context of the overlap between the phenyl of ampicillin and the *tert*-butyl of timolol. Further, the carboxylate of ampicillin is superimposed with a hydrogen-bond acceptor functionality of both modeled ligands.

There is evidence of drug interactions between these two classes of drugs, which we hypothesize to involve an overlap in absorptive transport. Ampicillin has been observed to result in decreased atenolol absorption in patients.[83] This is a general feature of orally dosed β-blockers: atenolol, nadolol, propranolol, and timolol have established drug interactions with penicillins in general and ampicillin in particular.[45] β-Lactam antibiotics, including penicillins and cephalosporins, have been shown to be substrates of multiple transporters such as PEPT1 and PEPT2, as well as members of the multidrug resistance (MDR/MRP), organic anion (OAT), and organic cation (OCT) families.[36,84] Note that while the OCT family of transporters are named for cation transport, they are known to transport zwitterions as well.[84] Studies suggest that β-AR antagonists may be substrates for the organic cation transporter OCT2.[85,86] Another mechanism by which one drug can reduce the intestinal absorption levels of another is by inducing the expression of human CYP3A4[87] or by enhancing expression of the P-gp efflux pump.[88] However, we found no evidence that β-lactam antibiotics induce the expression of the cytochromes that metabolize β-blockers (e.g. CYP2D6) or increase levels of P-gp.

The biopharmaceutics classification system (BCS), proposed by Amidon et al. and adopted by the FDA, classifies therapeutic agents based on mechanistic approaches to the drug absorption and dissolution processes for predicting in vivo pharmacokinetic performance.[89] Wu and Benet have proposed a modification to the BCS system that may be better for predicting overall drug disposition, including routes of drug elimination and the effects of efflux and absorptive transporters on oral drug absorption.[37] In their modified BDDCS system (Biopharmaceutics Drug Disposition Classification System), classifications are driven by solubility (as with BCS) but make use of metabolism instead of the permeability criteria of BCS. Within BCS, the β-lactams are class 3 drugs (high solubility, low permeability) as are atenolol and nadolol, but other β-blockers fall within other classes (e.g. labetalol, metoprolol, and propranolol are class I). With the BDDCS system, many drugs remain in the same nominal class as under BDS, but for different reasons. In particular the BCS class 3 drugs atenolol and nadolol remain BDDCS class 3 but are classed as such on the basis of high solubility and poor metabolism (rather than poor permeability).

In the BDDCS system, Wu and Benet note that absorptive transporter effects are frequently important in the intestinal absorption of class 3 drugs, which is consistent with our hypothesis.

Recall, though, that the orally dosed β-blockers, as a group irrespective of BCS or BDDCS classification, show drug interactions with the β-lactam antibiotics.[45] Both classification systems are based on measurable properties such as solubility, permeability, and extent of metabolism, but neither was designed to directly predict drug interactions that are based on transport phenomena. Our observations suggest that there are subtle structure-based effects that explain certain drug−drug interactions as those we have detailed between the β-lactam antibiotics and the β-blockers. We would propose that explicit modeling of the substrates of the various transporters along with large-scale computations of the sort we have done will further refine our ability to characterize and predict drug−drug interactions and may help to guide preclinical drug evaluation.

Case 4: Drug Metabolism Overlap. Overlap in drug metabolism can manifest in the same fashion as with drug transport overlap. A primary pathway of drug metabolism is mediated by the cytochrome P450 (CYP) enzymes. For example, grapefruit juice contains inhibitors of intestinal CYP3A4 and thus reduces presystemic metabolism of some cardiovascular drugs, which leads to overdose toxicity.[90] Apart from shared interactions with broadly acting metabolic enzymes, one example of metabolic overlap in our results is seen with barbiturates and SCN antagonists. Note that these also exhibit both primary and tertiary target overlap. The SCN antagonist felbamate inhibits CYP2C19 and this has been suggested to be the mechanism by which felbamate increases plasma concentrations of phenytoin and phenobarbital.[91]

Conclusions

We have reported the results of a series of computational experiments on the product of a curation effort that annotated the specific effectors of pharmacological effects for most known drugs. Molecular similarity and imprinting methods showed utility in inducing a taxonomy of drugs and drug targets in a way that went beyond classical pharmacological description of chemical classes and biological effects. The clusters of drugs were consistent with their primary activities but also suggested secondary effects. The performance levels of ligand-based virtual screening on numerous diverse targets were competitive with the best docking methods, and we believe that ligand-based methods offer a viable and productive means to accelerate drug discovery in the same way that protein-structure-based methods have come to be used. The performance we observed in terms of selectivity, in cases where confusable drugs and drug targets existed, further strengthens the case for broad application of methods such as these.

We observed four situations where predicted activities of drugs had not been initially annotated. Two types manifest as side effects (primary and tertiary target overlap) and two manifest as drug−drug interactions involving effects on pharmacokinetics (drug transporter or metabolism overlap). Interestingly, broad application of a successful docking approach based on protein structure would reveal a linkage between different drug classes *only* in the case of primary target overlap. However, a number of our observations did *not* involve primary target overlap, but revealed interesting relationships. In these cases, making use of models based on ligand structure was critical in uncovering the effects. This work represents a step toward systematic computational modeling of a large enough fraction of pharmacologically relevant targets to support practical hypothesis generation in preclinical drug discovery.

While our work has focused on a uniform methodology, we would advocate modeling pharmacologically interesting targets using the best available methods for each. Some groups have reported success in making use of homology modeled structures as the targets of docking, even in very hard cases such as presented by GPCRs.[92,93] Others have applied QSAR to modeling the substrate specificity of transporters such as P-gp.[94] We hope that our observation of the utility of ligand-based methods in uncovering shared nonprimary effects would inform the strategies of others. We expect methods that seek to directly model the physical structures of proteins to uncover linkages only in the case of primary target overlap.

A number of groups have developed methods that address ligand-based activity modeling.[20,21,27] Cramer, in particular, has addressed 15 different targets in a single report using Topomer CoMFA, with promising results.[21] He described four cases of increasing generality in modeling, which pose progressive challenges for the topomer method (which considers molecular fragments). Case 4 involved modeling chemical series with negligible homology, which poses a difficulty for fragment-based methods and any 2D-based methods. We believe that we have convincingly demonstrated utility in Cramer's case 4 on a large and diverse set of targets. A particularly attractive feature of the methods reported here is that the computational procedures are *not* labor-intensive. They are fully automatic and do not require careful selection of the ligands used to induce a model.

In our paper introducing the Surflex-Sim methodology,[18] we said, "It should be possible to enable rapid virtual screening against many tens of biological targets, which might prove to be of use in suggesting potential side-effect modulators of molecules undergoing development toward clinical application." We believe that the present work demonstrates the feasibility of that goal.

Acknowledgment. The authors are grateful to Les Benet and Kathy Giacomini for helpful discussions of the role of membrane-bound drug transporters. The authors gratefully acknowledge NIH for partial funding of the work (grants GM070481 and CA64602). A.N.J. has a financial interest in BioPharmics LLC, a biotechnology company whose main focus is in the development of methods for computational modeling that are relevant for drug discovery.

References

(1) Walters, P. W.; Stahl, M. T.; Murcko, M. A. Virtual screening−An overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(2) Jain, A. N. Virtual screening in lead discovery and optimization. *Curr Opin. Drug. Discovery Dev.* **2004**, *7*, 396−403.
(3) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499−511.
(4) Pham, T. A.; Jain, A. N. Parameter estimation for scoring protein−ligand interactions using negative training data. *J. Med. Chem.* **2006**, in press.
(5) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57*, 225−242.
(6) Miteva, M. A.; Lee, W. H.; Montes, M. O.; Villoutreix, B. O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48*, 6012−6022.
(7) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(8) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., 3rd. Comparative study of several algorithms for flexible ligand docking. *J Comput. Aided Mol. Des.* **2003**, *17*, 755−763.

(9) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(10) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(11) Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein−ligand interaction. *Proteins* **2002**, *49*, 457−471.

(12) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.

(13) Jain, A. N.; Dietterich, T. G.; Lathrop, R. H.; Chapman, D.; Critchlow, R. E., Jr.; Bauer, B. E.; Webster, T. A.; Lozano-Perez, T. A shape-based machine learning tool for drug design. *J. Comput. Aided. Mol. Des.* **1994**, *8*, 635−652.

(14) Jain, A. N.; Koile, K.; Chapman, D. Compass: Predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.* **1994**, *37*, 2315−2327.

(15) Jain, A. N.; Harris, N. L.; Park, J. Y. Quantitative binding site model generation: Compass applied to multiple chemotypes targeting the 5-HT1A receptor. *J. Med. Chem.* **1995**, *38*, 1295−1308.

(16) Jain, A. N. Chemical analysis by morphological similarity. US patent 6,470,305, 2002.

(17) Perkins, E.; Sun, D.; Nguyen, A.; Tulac, S.; Francesco, M.; Tavana, H.; Nguyen, H.; Tugendreich, S.; Barthmaier, P.; Couto, J.; Yeh, E.; Thode, S.; Jarnagin, K.; Jain, A. N.; Morgans, D.; Melese, T. Novel inhibitors of poly(ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. *Cancer Res.* **2001**, *61*, 4175−4183.

(18) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947−961.

(19) Jain, A. N. Morphological similarity: A 3D molecular similarity method correlated with protein−ligand recognition. *J. Comput. Aided Mol. Des.* **2000**, *14*, 199−213.

(20) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative MOLECULAR FIELD Analysis (CoMFA). Effect of shape on binding of steroid to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(21) Cramer, R. D. Topomer CoMFA: A design methodology for rapid lead optimization. *J. Med. Chem.* **2003**, *46*, 374−388.

(22) Pitman, M. C.; Huber, W. K.; Horn, H.; Kramer, A.; Rice, J. E.; Swope, W. C. FLASHFLOOD: A 3D field-based similarity search and alignment method for flexible molecules. *J. Comput. Aided Mol. Des.* **2001**, *15*, 587−612.

(23) Kramer, A.; Horn, H. W.; Rice, J. E. Fast 3D molecular superposition and similarity search in databases of flexible molecules. *J. Comput. Aided Mol. Des.* **2003**, *17*, 13−38.

(24) Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput. Aided Mol. Des.* **1999**, *13*, 1−10.

(25) Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast flexible superposition of drug-size molecules. *J. Comput. Aided Mol. Des.* **1999**, *13*, 35−49.

(26) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502−4520.

(27) Nicholls, A.; MacCuish, N. E.; MacCuish, J. D. Variable selection and model validation of 2D and 3D molecular descriptors. *J. Comput. Aided Mol. Des.* **2004**, *18*, 451−474.

(28) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. 2. The relationship between alignment solutions obtained from conformationally rigid and flexible matching. *J. Comput. Aided. Mol. Des.* **2000**, *14*, 39−51.

(29) Meurice, N.; Maggiora, G. M.; Vercauteren, D. P. Evaluating molecular similarity using reduced representations of the electron density. *J. Mol. Model. (Online)* **2005**, *11*, 237−247.

(30) Renfrey, S.; Featherstone, J. Structural proteomics. *Nat. Rev. Drug Discovery* **2002**, *1*, 175−176.

(31) Dmochowski, R.; Staskin, D. R. The q−T interval and antimuscarinic drugs. *Curr. Urol. Rep.* **2005**, *6*, 405−409.

(32) Howell, G., 3rd; West, L.; Jenkins, C.; Lineberry, B.; Yokum, D.; Rockhold, R. In vivo antimuscarinic actions of the third generation antihistaminergic agent, desloratadine. *BMC Pharmacol.* **2005**, *5*, 13.

(33) Fernandez, D.; Ghanta, A.; Kauffman, G. W.; Sanguinetti, M. C. Physicochemical features of the HERG channel drug binding site. *J. Biol. Chem.* **2004**, *279*, 10120−10127.

(34) Daly, A. K. Pharmacogenetics of the cytochromes P450. *Curr. Top. Med. Chem.* **2004**, *4*, 1733−1744.

(35) Sakaeda, T.; Nakamura, T.; Okumura, K. Pharmacogenetics of drug transporters and its impact on the pharmacotherapy. *Curr. Top. Med. Chem.* **2004**, *4*, 1385−1398.

(36) Dresser, M. J.; Leabman, M. K.; Giacomini, K. M. Transporters involved in the elimination of drugs in the kidney: Organic anion transporters and organic cation transporters. *J. Pharm. Sci.* **2001**, *90*, 397−421.

(37) Wu, C. Y.; Benet, L. Z. Predicting drug disposition via application of BCS: Transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm. Res.* **2005**, *22*, 11−23.

(38) Mount, J.; Ruppert, J.; Welch, W.; Jain, A. N. IcePick: A flexible surface-based system for molecular diversity. *J. Med. Chem.* **1999**, *42*, 60−66.

(39) Ghuloum, A. M.; Sage, C. R.; Jain, A. N. Molecular hashkeys: A novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J. Med. Chem.* **1999**, *42*, 1739−1748.

(40) Briem, H.; Kuntz, I. D. Molecular similarity based on DOCK-generated fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.

(41) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **1995**, *2*, 107−118.

(42) Maglott, D.; Ostell, J.; Pruitt, K. D.; Tatusova, T. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **2005**, *33*, D54−58.

(43) Sigel, E. Mapping of the benzodiazepine recognition site on GABA-(A) receptors. *Curr. Top. Med. Chem.* **2002**, *2*, 833−839.

(44) Olshen, A. B.; Jain, A. N. Deriving quantitative conclusions from microarray expression data. *Bioinformatics* **2002**, *18*, 961−970.

(45) Lacy, C. F.; Armstrong, L. L.; Goldman, M. P.; Lance, L. L. *Drug Information Handbook*, 13th ed.; Lexi-Comp: Hudson, OH, 2005.

(46) Hardman, J. G.; Limbird, L. E.; Gilman, A. G. *Goodman and Gilman's: The Pharmacological Basis of Therapeutics*, 10th ed.; McGraw-Hill: New York, 2001.

(47) Sonneveld, E.; Jansen, H. J.; Riteco, J. A.; Brouwer, A.; van der Burg, B. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. *Toxicol. Sci.* **2005**, *83*, 136−148.

(48) Ragsdale, D. S.; McPhee, J. C.; Scheuer, T.; Catterall, W. A. Common molecular determinants of local anesthetic, antiarrhythmic, and anticonvulsant block of voltage-gated Na+ channels. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9270−9275.

(49) Mullin, M. J.; Hunt, W. A. Ethanol inhibits veratridine-stimulated sodium uptake in synaptosomes. *Life Sci.* **1984**, *34*, 287−292.

(50) Czapinski, P.; Blaszczyk, B.; Czuczwar, S. J. Mechanisms of action of antiepileptic drugs. *Curr. Top. Med. Chem.* **2005**, *5*, 3−14.

(51) Frenkel, C.; Duch, D. S.; Recio-Pinto, E.; Urban, B. W. Pentobarbital suppresses human brain sodium channels. *Brain Res. Mol. Brain Res.* **1989**, *6*, 211−216.

(52) Frenkel, C.; Duch, D. S.; Urban, B. W. Molecular actions of pentobarbital isomers on sodium channels from human brain cortex. *Anesthesiology* **1990**, *72*, 640−649.

(53) Wartenberg, H. C.; Wartenberg, J. P.; Urban, B. W. Pharmacological modification of sodium channels from the human heart atrium in planar lipid bilayers: Electrophysiological characterization of responses to batrachotoxin and pentobarbital. *Eur. J. Anaesthesiol.* **2003**, *20*, 354−362.

(54) O'Neil, M. J.; Smith, A.; Patricia, E. H.; Obenchain, J. R.; Gallipeau, J. A. R.; D'Arecca, M. A. *Merck Index: An Encyclopedia of Chemicals, Drugs, & Biologicals*, 13th ed.; Merck & Co.: Rahway, NJ, 2001.

(55) DeLorey, T. M.; Olsen, R. W. Gamma-aminobutyric acidA receptor structure and function. *J. Biol. Chem.* **1992**, *267*, 16747−16750.

(56) Rho, J. M.; Donevan, S. D.; Rogawski, M. A. Barbiturate-like actions of the propanediol dicarbamates felbamate and meprobamate. *J. Pharmacol. Exp. Ther.* **1997**, *280*, 1383−1391.

(57) Koe, B. K.; Minor, K. W.; Kondratas, T.; Lebel, L. A.; Koch, S. W. Enhancement of benzodiazepine binding by methaqualone and related quinazolines. *Drug Dev. Res.* **1986**, *7*, 255−268.

(58) BeDell, L. S. *Mosby's Complete Drug Reference*; Mosby-Year Book, Inc.: St. Louis, 1997.

(59) Gonzalez, M. A.; Estes, K. S. Pharmacokinetic overview of oral second-generation H1 antihistamines. *Int. J. Clin. Pharmacol. Ther.* **1998**, *36*, 292−300.

(60) Kubo, N.; Shirakawa, O.; Kuno, T.; Tanaka, C. Antimuscarinic effects of antihistamines: Quantitative evaluation by receptor-binding assay. *Jpn. J. Pharmacol.* **1987**, *43*, 277−282.

(61) Cardelus, I.; Anton, F.; Beleta, J.; Palacios, J. M. Anticholinergic effects of desloratadine, the major metabolite of loratadine, in rabbit and guinea-pig iris smooth muscle. *Eur. J. Pharmacol.* **1999**, *374*, 249−254.

(62) Hustveit, O.; Setekleiv, J. Fentanyl and pethidine are antagonists on muscarinic receptors in guinea-pig ileum. *Acta Anaesthesiol. Scand.* **1993**, *37*, 541−544.
(63) Chen, Y. P.; Chen, S. R.; Pan, H. L. Systemic morphine inhibits dorsal horn projection neurons through spinal cholinergic system independent of descending pathways. *J. Pharmacol. Exp. Ther.* **2005**, *314*, 611−617.
(64) Sofia, R. D.; Gordon, R.; Gels, M.; Diamantis, W. Comparative effects of felbamate and other compounds on *N*-methyl-D-aspartic acid-induced convulsions and lethality in mice. *Pharmacol. Res.* **1994**, *29*, 139−144.
(65) Brown, L. M. Pentobarbital differentially inhibits *N*-methyl-D-aspartate and kainate-stimulated [^{3}H]noradrenaline overflow in rat cortical slices. *Gen. Pharmacol.* **1995**, *26*, 1603−1606.
(66) Charlesworth, P.; Jacobson, I.; Richards, C. D. Pentobarbitone modulation of NMDA receptors in neurones isolated from the rat olfactory brain. *Br. J. Pharmacol.* **1995**, *116*, 3005−3013.
(67) Harty, T. P.; Rogawski, M. A. Felbamate block of recombinant *N*-methyl-D-aspartate receptors: Selectivity for the NR2B subunit. *Epilepsy Res.* **2000**, *39*, 47−55.
(68) Schwendt, M.; Duncko, R.; Makatsori, A.; Moncek, F.; Johansson, B. B.; Jezova, D. Involvement of glutamate neurotransmission in the development of excessive wheel running in Lewis rats. *Neurochem. Res.* **2003**, *28*, 653−657.
(69) Tanaka, K.; Tomisato, W.; Hoshino, T.; Ishihara, T.; Namba, T.; Aburaya, M.; Katsu, T.; Suzuki, K.; Tsutsumi, S.; Mizushima, T. Involvement of intracellular Ca2+ levels in nonsteroidal anti-inflammatory drug-induced apoptosis. *J. Biol. Chem.* **2005**, *280*, 31059−31067.
(70) Beltinger, C.; Fulda, S.; Kammertoens, T.; Uckert, W.; Debatin, K. M. Mitochondrial amplification of death signals determines thymidine kinase/ganciclovir-triggered activation of apoptosis. *Cancer Res.* **2000**, *60*, 3212−3217.
(71) Shiff, S. J.; Koutsos, M. I.; Qiao, L.; Rigas, B. Nonsteroidal antiinflammatory drugs inhibit the proliferation of colon adenocarcinoma cells: Effects on cell cycle and apoptosis. *Exp. Cell. Res.* **1996**, *222*, 179−188.
(72) Elder, D. J.; Halton, D. E.; Hague, A.; Paraskeva, C. Induction of apoptotic cell death in human colorectal carcinoma cell lines by a cyclooxygenase-2 (COX-2)-selective nonsteroidal anti-inflammatory drug: Independence from COX-2 protein expression. *Clin. Cancer Res.* **1997**, *3*, 1679−1683.
(73) Smith, M. L.; Hawcroft, G.; Hull, M. A. The effect of non-steroidal anti-inflammatory drugs on human colorectal cancer cells: Evidence of different mechanisms of action. *Eur. J. Cancer* **2000**, *36*, 664−674.
(74) Reinacher-Schick, A.; Schoeneck, A.; Graeven, U.; Schwarte-Waldhoff, I.; Schmiegel, W. Mesalazine causes a mitotic arrest and induces caspase-dependent apoptosis in colon carcinoma cells. *Carcinogenesis* **2003**, *24*, 443−451.
(75) Bold, R. J.; Chandra, J.; McConkey, D. J. Gemcitabine-induced programmed cell death (apoptosis) of human pancreatic carcinoma is determined by Bcl-2 content. *Ann. Surg. Oncol.* **1999**, *6*, 279−285.
(76) Schniewind, B.; Christgen, M.; Kurdow, R.; Haye, S.; Kremer, B.; Kalthoff, H.; Ungefroren, H. Resistance of pancreatic cancer to gemcitabine treatment is dependent on mitochondria-mediated apoptosis. *Int. J. Cancer* **2004**, *109*, 182−188.
(77) Mohammad, R. M.; Wang, S.; Banerjee, S.; Wu, X.; Chen, J.; Sarkar, F. H. Nonpeptidic small-molecule inhibitor of Bcl-2 and Bcl-XL, (−)-Gossypol, enhances biological effect of genistein against BxPC-3 human pancreatic cancer cell line. *Pancreas* **2005**, *31*, 317−324.
(78) Tolomeo, M.; Mancuso, S.; Todaro, M.; Stassi, G.; Catalano, M.; Arista, S.; Cannizzo, G.; Barbusca, E.; Abbadessa, V. Mitochondrial disruption and apoptosis in lymphocytes of an HIV infected patient affected by lactic acidosis after treatment with highly active anti-retroviral therapy. *J. Clin. Pathol.* **2003**, *56*, 147−151.
(79) Caron, M.; Auclair, M.; Lagathu, C.; Lombes, A.; Walker, U. A.; Kornprobst, M.; Capeau, J. The HIV-1 nucleoside reverse transcriptase inhibitors stavudine and zidovudine alter adipocyte functions in vitro. *Aids* **2004**, *18*, 2127−2136.
(80) Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* **2005**, *435*, 677−681.
(81) Mizuno, N.; Niwa, T.; Yotsumoto, Y.; Sugiyama, Y. Impact of drug transporter studies on drug discovery and development. *Pharmacol. Rev.* **2003**, *55*, 425−461.
(82) Liang, R.; Fei, Y. J.; Prasad, P. D.; Ramamoorthy, S.; Han, H.; Yang-Feng, T. L.; Hediger, M. A.; Ganapathy, V.; Leibach, F. H. Human intestinal H+/peptide cotransporter. Cloning, functional expression, and chromosomal localization. *J. Biol. Chem.* **1995**, *270*, 6456−6463.
(83) Schafer-Korting, M.; Kirch, W.; Axthelm, T.; Kohler, H.; Mutschler, E. Atenolol interaction with aspirin, allopurinol, and ampicillin. *Clin. Pharmacol. Ther.* **1983**, *33*, 283−288.
(84) Van Bambeke, F.; Michot, J. M.; Tulkens, P. M. Antibiotic efflux pumps in eukaryotic cells: Occurrence and impact on antibiotic cellular pharmacokinetics, pharmacodynamics and toxicodynamics. *J. Antimicrob. Chemother.* **2003**, *51*, 1067−1077.
(85) Ott, R. J.; Giacomini, K. M. Stereoselective interactions of organic cations with the organic cation transporter in OK cells. *Pharm. Res.* **1993**, *10*, 1169−1173.
(86) Dudley, A. J.; Bleasby, K.; Brown, C. D. The organic cation transporter OCT2 mediates the uptake of beta-adrenoceptor antagonists across the apical membrane of renal LLC-PK(1) cell monolayers. *Br. J. Pharmacol.* **2000**, *131*, 71−79.
(87) Kliewer, S. A.; Moore, J. T.; Wade, L.; Staudinger, J. L.; Watson, M. A.; Jones, S. A.; McKee, D. D.; Oliver, B. B.; Willson, T. M.; Zetterstrom, R. H.; Perlmann, T.; Lehmann, J. M. An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* **1998**, *92*, 73−82.
(88) Matheny, C. J.; Lamb, M. W.; Brouwer, K. R.; Pollack, G. M. Pharmacokinetic and pharmacodynamic implications of P-glycoprotein modulation. *Pharmacotherapy* **2001**, *21*, 778−796.
(89) Amidon, G. L.; Lennernas, H.; Shah, V. P.; Crison, J. R. A theoretical basis for a biopharmaceutic drug classification: The correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm. Res.* **1995**, *12*, 413−420.
(90) Bailey, D. G.; Dresser, G. K. Interactions between grapefruit juice and cardiovascular drugs. *Am. J. Cardiovasc. Drugs* **2004**, *4*, 281−297.
(91) Glue, P.; Banfield, C. R.; Perhach, J. L.; Mather, G. G.; Racha, J. K.; Levy, R. H. Pharmacokinetic interactions with felbamate. In vitro−in vivo correlation. *Clin. Pharmacokinet.* **1997**, *33*, 214−224.
(92) Petrel, C.; Kessler, A.; Dauban, P.; Dodd, R. H.; Rognan, D.; Ruat, M. Positive and negative allosteric modulators of the Ca2+-sensing receptor interact within overlapping but not identical binding sites in the transmembrane domain. *J. Biol. Chem.* **2004**, *279*, 18990−18997.
(93) Petrel, C.; Kessler, A.; Maslah, F.; Dauban, P.; Dodd, R. H.; Rognan, D.; Ruat, M. Modeling and mutagenesis of the binding site of Calhex 231, a novel negative allosteric modulator of the extracellular Ca-(2+)-sensing receptor. *J. Biol. Chem.* **2003**, *278*, 49487−49494.
(94) Gombar, V. K.; Polli, J. W.; Humphreys, J. E.; Wring, S. A.; Serabjit-Singh, C. S. Predicting P-glycoprotein substrates by a quantitative structure−activity relationship model. *J. Pharm. Sci.* **2004**, *93*, 957−968.

JM051139T