

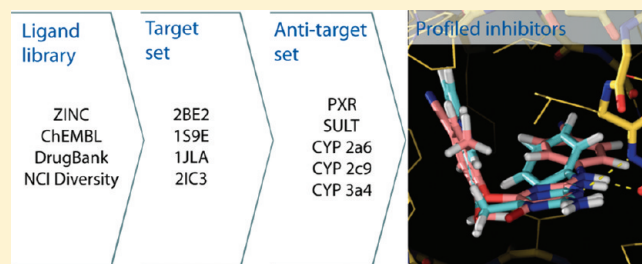
Combined Approach Using Ligand Efficiency, Cross-Docking, and Antitarget Hits for Wild-Type and Drug-Resistant Y181C HIV-1 Reverse Transcriptase

Alfonso T. García-Sosa,* Sulev Sild, Kalev Takkis, and Uko Maran

Institute of Chemistry, University of Tartu, Ravila 14a, Tartu 50411, Estonia

S Supporting Information

ABSTRACT: New hits against HIV-1 wild-type and Y181C drug-resistant reverse transcriptases were predicted taking into account the possibility of some of the known metabolism interactions. In silico hits against a set of antitargets (i.e., proteins or nucleic acids that are off-targets from the desired pharmaceutical target objective) are used to predict a simple, visual measure of possible interactions for the ligands, which helps to introduce early safety considerations into the design of compounds before lead optimization. This combined approach consists of consensus docking and scoring: cross-docking to a group of wild-type and drug-resistant mutant proteins, ligand efficiency (also called binding efficiency) indices as new ranking measures, pre- and postdocking filters, a set of antitargets and estimation, and minimization of atomic clashes. Diverse, small-molecule compounds with new chemistry (such as a triazine core with aromatic side chains) as well as known drugs for different applications (oxazepam, chlorthalidone) were highly ranked to the targets having binding interactions and functional group spatial arrangements similar to those of known inhibitors, while being moderate to low binders to the antitargets. The results are discussed on the basis of their relevance to medicinal and computational chemistry. Optimization of ligands to targets and off-targets or antitargets is foreseen to be critical for compounds directed at several simultaneous sites.



INTRODUCTION

At the time this paper was written, acquired immune deficiency syndrome (AIDS) caused by the human immunodeficiency virus (HIV) was identified as the largest cause of mortality for women at reproductive age worldwide.¹ HIV is a characteristic retrovirus that can quickly incorporate errors in its genetic replication process.² These errors provide mutations that allow HIV to change its genetic material at a high rate when coupled to a fast life cycle and natural selection.³ These features also complicate the development of an effective vaccine.³ There have been clinical successes in the treatment of HIV-1 by, chronologically, the discovery of HIV-1 protease inhibitors, HIV-1 reverse transcriptase (HIV-1 RT) inhibitors, HIV-1 entry inhibitors, and HIV-1 integrase inhibitors.^{4,5} However, these viral component enzymes are liable to mutate, have adapted, and continue to adapt to inhibitors. A viable strategy may compose a treatment combining inhibitors against each of these enzymes.

HIV-1 RT is a well-known therapeutic target for treating HIV-1 infection and AIDS since there are no human equivalent enzymes and it is essential in HIV-1 infection and disease progression. HIV-1 RT converts the viral RNA into viral DNA and has a low replication fidelity (producing a high number of errors) as it does not proofread the synthesized DNA.² Nucleoside and non-nucleoside reverse transcriptase inhibitors (NRTIs and NNRTIs, respectively) are widely used in the clinic, in addition

to, or in combination with, HIV-1 protease inhibitors.⁶ However, there are mutations that occur naturally during treatment with some NNRTIs that render the protein resistant to them. One of the most important is the Y181C mutation, where many ligands that interacted with the former aromatic side chain are incapable of binding in the same manner to the smaller cysteine residue.⁷ Computational methods can be developed to screen molecules and design drug candidates against rapidly mutating enzymes for experimental testing. Therefore, an approach using both wild-type and Y181C mutant drug-resistant reverse transcriptases was employed for drug design in the present work.

Many drug candidate compounds fail due to problems of efficacy and safety (toxicity and/or side effects).⁸ Off-target or antitarget interactions may be the source of many of these side effects or toxicity.^{9–12} Target molecules are those proteins or nucleic acids to which binding is desired to cause a pharmacological effect. Off-targets are biomolecules that are not the principal target for the desired effect and as such may produce a collateral effect, or side effect, which may be beneficial (for example, therapeutic side effects that may be further developed or used in the clinic) or not. Side effects may be harmful enough to be toxic to cells, tissues, and organs. Antitargets are either off-targets that

Received: May 6, 2011

Published: August 30, 2011

cause a strong enough reaction leading to toxicity or off-targets which are desired to be avoided to tailor the biological behavior of the compound (see cytochromes below). Compounds that have the possibility to bind to targets, off-targets, and/or anti-targets can be identified as likely to have more than one target interaction and flagged as early as possible in the discovery stages to modify those interactions.

In some cases, metabolic enzymes can be considered as antitargets. Cytochromes P450 (CYP P450s) are a good choice as antitargets since they are involved in the phase I metabolism of foreign substances, generally through oxidations that are one of the first chemical modifications of a molecule inside the bloodstream and produce metabolites that are more polar and easier to excrete.^{13,14} If a CYP enzyme is inhibited by a compound, then other drugs that are administered at the same time may have a different effect since the particular CYP enzyme being inhibited by the first drug is no longer available for metabolism of the other drug(s).¹⁵ A metabolic enzyme may convert a relatively safe compound into a much more toxic metabolite, and this would be another case where it could be treated as an antitarget. Caution must be used if a compound is not transformed by any metabolizing enzyme, because it might accumulate dangerously in the organism over time and therefore cause severe side effects or it might have very fast clearance times and difficulty reaching a therapeutically important dose and window.

In some cases, a metabolic enzyme could be another therapeutic target (desired off-target) for the compound in addition to its designed target. A prodrug may be metabolized and lead to a relatively safe and active drug compound that is required for a particular treatment, such as with dacarbazine, an anticancer drug.⁶ A possible way to increase the free concentration in plasma of compounds may be to design inhibitors against the desired target(s) that, at the same time, do not interact with CYP enzymes, which may, in turn, mean that lower doses can reach a therapeutic effect. Inhibition of an antitarget may be desired, for example, to free another drug that is bound to the antitarget and boost the free levels of the latter drug in the bloodstream. In addition to thermodynamic properties, the kinetics of binding will also dictate the speed of inhibitor release and thus the free concentration in blood of a compound. A well-known case is the HIV-1 protease inhibitor ritonavir, which also inhibits cytochromes 2d6 and 3a and enables other HIV antivirals to remain not bound to these enzymes and thus remain unmetabolized for a longer period of time.¹⁶ Therefore, a fine-tuned metabolic enzyme and target receptor inhibition is required, one that is flexible and adaptable to different design purposes, yet sensitive and reliable.

Good choices for antitargets are also the enzymes involved in phase II metabolism where the metabolite is conjugated to hydrophilic compounds, such as sugar or glucuronic acid. In the case of the sulfotransferases (SULTs), a sulfonyl group is attached to hydroxyl and amino groups on the compound to form sulfates and sulfamate conjugates, again making them easier to excrete, as well as less able to interfere with host proteins, receptors, and DNA. Other good choices for antitargets are drug transporters of phase III metabolism that can efflux drugs and proteins that create resistance to medications.¹⁵ Foreign substances are recognized by the pregnane X receptor (PXR), which also up-regulates enzymes (including CYP P450s, SULTs, and drug transporters) that will transform the substance or excrete it.¹⁵ Studies on the structure of PXR show a mostly hydrophobic binding site, with the ability to bind to a wide variety of different

Table 1. Target Protein Crystal Structures for HIV-1 Reverse Transcriptase

| protein type | PDB ID | resolution (Å) |
|-----------------------|--------|----------------|
| wild | 2BE2 | 2.43 |
| | 1S9E | 2.6 |
| drug-resistant mutant | 1JLA | 2.5 |
| | 2IC3 | 3.0 |

ligands and without large conformational changes between *apo* and *holo* forms of the receptor.¹⁷ Virtual screening and docking can make good use of this character, since the interactions between predicted binding ligands and proteins in the binding site may be controlled with one protein crystal structure. It is important to note that frequent sites of toxicity problems, the liver and small intestines, are the organs where most of these enzymes, receptors, and transporters are mainly located. The side effect and toxicity profiling of compounds by predicting the antitarget interactions of ligands may help discover similar side effects and therapeutic targets between drugs¹⁸ and drug candidates, as well as find new targets for compounds.¹⁹ Network pharmacology, or “polypharmacology”, advances the theory of designing ligands that can inhibit more than one protein and carry out the desirable (sometimes even required) ligand optimization against several targets.^{20,21} Drug repurposing, or finding new diseases that a drug or a drug candidate can aid with, is also increasing in importance.^{8,22} In addition, there can be differences between animal models of disease and toxicity and those of the human organism, and sometimes there are no human orthologue proteins in other mammals, such as for sulfotransferase 1A3.²³ A major challenge to be solved for these aims is to achieve selectivity against several targets, and for the approach to be successful, it is critical that the designed ligands have the exact level of promiscuity desired, binding to several protein targets but at the same time avoiding unwanted targets to not create over-promiscuous ligands.

In the present work, a novel drug design strategy has been developed through *in silico* screening on several wild-type and Y181C drug-resistant HIV-1 RT protein structures, using ligand efficiency (also called binding efficiency) indices and hydrogen bonding to improve the results of the scoring and ranking of molecules predicted to bind to more than one protein crystal structure, sequence, and hydration state to account for multiple mutations and conformations. In addition, a set of antitargets is collected (battery), and the results of the predicted binding of three CYP structures, one SULT structure, and one PXR structure are used to roughly and visually estimate metabolic interactions for virtual drug design. This combination is a new approach to treat drug design and metabolism together in the early stages of drug discovery and may aid in designing compounds with a desired level of interaction with targets and antitargets and in the first steps to establish an *in silico* metabolic profile. The choice of off-targets can define the off-battery set purpose. Biological interactions still require experimental confirmation, and the procedure does not distinguish between side effects and toxicity.

METHODS

Proteins. The set (or “battery”) of protein targets included the structures of wild-type HIV-1 reverse transcriptase, 2BE2 and 1S9E from the Protein Data Bank (PDB)²⁴ (see Table 1).

Table 2. Antitarget Protein Crystal Structures

| name (abbreviation) | PDB ID | resolution (Å) | ligand or inhibitor in complex |
|-------------------------------|--------|----------------|--------------------------------|
| sulfotransferase 1A3 (SULT) | 2A3R | 2.60 | L-dopamine |
| pregnane X receptor (PXR) | 1M13 | 2.15 | hyperforin |
| cytochrome P450 2a6 (CYP 2a6) | 1Z10 | 1.90 | coumarin |
| cytochrome P450 2c9 (CYP 2c9) | 1OG5 | 2.55 | (S)-warfarin |
| cytochrome P450 3a4 (CYP 3a4) | 1TQN | 2.05 | |

They have tyrosine 181 in the “up” conformation, which is the most common in the cocomplexes between this protein and known inhibitors.⁷ Structures of the drug-resistant Y181C mutant reverse transcriptases, 1JLA and 2IC3 from the PDB, were also employed. Together they represent at least two of the different clusters of HIV-1 RT structures recently observed.²⁵ The use of a set of protein structures can provide the benefit of exploring a larger region of conformational space, as well as exploring chemical space to identify compounds that would bind well to the protein in a variety of plausible conformations and/or mutations, as well as hydration states.^{26–28}

The antitarget set contained the human sulfotransferase 1A3, pregnane X receptor, and three cytochrome P450 enzymes. Table 2 shows these antitargets, their structures, and their cocrystallized ligands.

Maestro²⁹ was used to place hydrogen atoms on protein structures. The Protein Preparation Wizard³⁰ was used to assign ionic states of residues in the protein, allocate polar hydrogens, and partially minimize the protein structures while restraining the protein backbone to converge to a root means square deviation (rmsd) of 0.3 Å using the OPLS2005 force field (Impref minimization with default options). Water molecules were removed from most of the structures since they did not involve extensive interactions in the binding site, except water molecule HOH A1013 in structure 2BE2. For the latter, docking runs were performed both with and without water molecule HOH A1013 in the binding site, and the lowest energy ligand binding pose was kept.

Ligand Libraries. To compute receiver-operator characteristics curves (ROCs; as false-positive rate vs true-positive rate), as well as their area under the curve (AUC), 60 known active and potent ligands downloaded from the ChEMBL database of biochemical assays³¹ were docked (positive controls), in addition to a test subset of 1140 randomly picked compounds (no molecular property filtering) from the ZINC database³² to act as decoys (negative controls), as well as a separate test with the Schrödinger collection of druglike ligand decoys (average molar mass (MW) of 400 g/mol, negative controls),³³ since they have the closest molecular properties to those of the known actives and are negative controls with which to measure the ability of the proposed procedure to recover known inhibitors. Decoys are presumed inactive molecules that have molecular properties close to those of active compounds to provide “difficult” test cases for docking procedures to recover the true actives from a background of similar compounds. The ROC and AUC values were calculated using the Python module CROC.³⁴

For HIV-1 RT, the known and in-clinical-use drugs efavirenz, etravirine, and nevirapine, as well as the known inhibitor rilpivirine, were employed as reference compounds. The National Cancer Institute’s diverse set of ligands,³⁵ a second set of ZINC ligands

comprising the full ZINC database collection of commercially available compounds (ca. 7 million compounds),³² and the structures of known and in-use drugs obtained from the DrugBank database³⁶ completed the sources for the library. The large ZINC library was then treated through virtual prescreening, originally prepared for RAC1 screening (currently unpublished), but also used in avian influenza drug design.³⁷ This was done by removing molecules with undesirable predicted solubility, predicted toxic and reactive groups (including α,β -unsaturated ketones⁷), and more than six rings, but keeping those with rotatable bonds between 5 and 12, polar surface area between 25 and 180 Å², molar mass between 300 and 650 g/mol, log *P* between –3 and +5, and up to seven atoms in any given ring, as well as less than six fused rings, using Instant JChem.³⁸ In the resulting collection of ca. 70 000 compounds (containing ca. 65 000 ZINC compounds), hydrogens were assigned in an ionic state corresponding to a pH close to 7.0 with LigPrep, and then employed for docking against the protein structures to find interesting hits.

Target and Antitarget Threshold Rules. A slightly different procedure was followed for the selection of target and antitarget hits. This was done by design to have extra confidence in that the target hits were predicted by consensus between docking programs/scoring functions (prospective part), while not requiring consensus between programs for antitarget hits, since in the latter case we would be interested if either of the programs/scoring functions detected an interaction (safety part).

Targets. A procedure was developed where compounds were docked and ranked to the protein targets with both Glide (version 55110)^{40–43} and Autodock (version 4).⁴⁴ For the targets, only those compounds that had deep interaction scores (negative) according to both programs, and lower (deeper) than those obtained with these programs for known inhibitors and drugs (see nevirapine), were considered for the top ranks. This can be considered an extra confidence procedure to minimize false positives for the targets, requiring correspondence between different programs and scoring functions. Consensus between scoring procedures has been documented to improve the performance over experimental high-throughput screening, increase the hit rates of active compounds, and decrease the false-positive rate as compared to the use of a single scoring function.^{45,46} Cross-docking was carried out between the native (cognate) ligands of each of the wild-type protein structures into the other, as well as between the native ligands of the mutant protein structures into the other.

Antitargets. At the same time, the compounds were docked and ranked against each of the antitarget proteins, and the interaction of each compound with each protein was recorded. The procedure for antitargets is designed for extra safety, in that any strong interaction (low score, deeply negative) between a ligand and protein by either docking program/scoring function is recorded to reduce false negatives for the antitargets. For each antitarget, the docked ligand molecule would receive a new numerical score of (i) 0 if it had a score higher (less negative) than 0.5 kcal/mol from the interaction score recorded by the cocomplexed native ligand in the protein crystal structure to indicate a lack of strong binding, (ii) 1/2 if it was within 0.5 kcal/mol of the score recorded by the cocomplexed ligand (i.e., no more than 0.5 kcal/mol above or below the score recorded by the cocomplexed native ligand in the protein crystal structure) to indicate a border case of binding, or (iii) 1 if its score was deeper (more negative) than 0.5 kcal/mol from the score recorded by the cocomplexed ligand to indicate a strong binding interaction.

These new scores provide a rough estimate of the likely hits against the several antitargets, and they are coded by color according to their value (0, 0.5, or 1). These new color codes are used to generate a visual representation of interactions as shown in Figures 6 and 7. These new scores also enable an interaction array to be built. A hit string can be constructed by adding the values of the individual interactions against the set of antitargets for a ligand. A combined hit string, also called a combined array sum, is the sum of all interactions with all antitarget proteins for both docking programs for the same ligand. An individual hit string, or individual array sum, is the sum of all interactions with all proteins for one ligand using one docking program/scoring function. That is, in the present method, with the present set of antitargets, the maximum combined hit string or combined array sum would be 10, and the maximum individual hit string or individual array sum would be 5.

Antitarget Validation. Extra sets of known active compounds (positive controls, obtained through literature searches and the ChEMBL database), as well as known nonbinding compounds (inactives obtained through literature searches), and the Schrödinger ligand decoy set (presumed inactives that have molecular properties similar to those of active compounds, e.g., average MW of 400 g/mol) were spiked into the ligand data set and docked into the set of antitargets to test if the procedure would recover them among the top ranks. These compounds for SULT, PXR, CYP 2a6, CYP 2c9, and CYP 3a4, respectively, were, for the actives, (SULT) ChapmanE-1,⁴⁷ (PXR) A-792611,⁴⁸ (CYP 2a6) 8-methoxyorsoralen (methoxsalen),⁴⁹ (CYP 2c9) sulfaphenazole,⁵⁰ and (CYP 3a4) nephazodone⁵¹ and, for the inactives, (SULT) 2,2-dimethyl-3-vinylcyclohexane,⁵² (PXR) VelaparathiU-10,⁵³ (CYP 2a6) ϵ -caprolactone,⁵⁴ (CYP 2c9) RaoS-31,⁵⁵ and (CYP 3a4) felbamate.⁵⁶ In addition, active and potent compounds from the ChEMBL bioassay database included ChEMBL IDs 167055 and 169033 for SULT, 456237, 457977, 59030, and 606702 for PXR, 178938, 179621, 179669, 179704, 214859, 214990, 36099, 361153, 368883, 369285, and 386124 for CYP 2a6, 1109, 455975, 455976, 456181, 456432, 457087, 458566, 458567, 463577, 463976, 464595, and 514730 for CYP 2c9, and 1089957, 270271, 271580, 507731, 573665, 583954, 75, and 98745 for CYP 3a4.³¹

Docking Schemes. A hierarchical scheme was employed in Glide using the high-throughput virtual screening (HTVS) scoring function to make a first selection based on flexible ligand, rigid protein docking, retaining 10% of the top compounds (one pose per compound). Then the standard precision (SP) scoring function made the next selection based on flexible ligand, rigid protein docking, retaining 10% of the top compounds (one pose per compound), and finally the extra precision (XP) scoring function docked ligands flexibly and ranked all top compounds, retaining one pose per compound. Conformations of amide bonds were allowed to be varied for all stages penalizing non-planar solutions. The protein atom size (van der Waals radii) was scaled by a factor of 0.8 for those atoms that have partial atomic charge smaller than 0.15. Scaling the van der Waals radii of protein atoms during docking is a standard procedure in the Virtual Screening Workflow,³⁰ which aims at reducing hard clashes between ligand and protein atoms. Many proteins such as PXR remain in a similar conformation in both *apo* and *holo* states, but large flexibility in the protein targets may not be taken into account. Since PXR is reported to bind to several ligands without major conformational rearrangement,¹⁷ we used the *holo* structure of PXR as described in structure 1M13. Postdocking

minimizations were carried out using the default options of the OPLS2001 force field for a maximum of 100 conjugate gradient steps with a distance-dependent dielectric model with a setting of 2.0 (where the effective dielectric constant is 2.0 times the distance between the interacting pair of atoms). The scoring procedure penalizes ligand binding poses where polar groups do not have a binding partner in the protein binding site as their high desolvation energy would disfavor the ligand's binding energy.

The settings used for the genetic algorithm in Autodock were as follows: number of individuals in population, 250; maximum number of energy evaluations, 2 million; maximum number of generations, 27 000; number of top individuals to survive to next generation, 1; number of genetic algorithm runs, 100. In addition, a ranked cluster analysis was performed on each docking calculation (100 runs of each ligand against each protein) to determine the cluster with the best energy and population of result poses.

To improve the reliability of the docking results, they were verified, including checking that (a) the ligand is in a correct site, (b) the ligand is not protruding significantly from the binding site, (c) the binding mode is plausible, and (d) the predicted binding pose has the identified privileged interactions and interaction partners, and proper consideration of (e) tautomers,⁵⁷ and (f) possible ionization and hydration states^{26–28} of the ligand and protein, among other checks. If carefully conducted and analyzed using control compounds, consensus docking results can be valuable to identify binding geometries and interaction partners. The results were inspected visually for the top-ranked compounds, and structures were judged on the basis of their structural and chemical soundness, as well as their consistency between programs. Compounds that were predicted to score deeply to both the wild-type and drug-resistant reverse transcriptases, and by consensus between both docking programs, were filtered to obtain the top compounds. The consensus scores were calculated as

$$\begin{aligned}\Delta G_{\text{consensus}}^{\text{wild type}} &= (\Delta G_{\text{XP GlideScore}}^{2\text{BE2}} + \Delta G_{\text{XP GlideScore}}^{1\text{S9E}} \\ &\quad + \Delta G_{\text{Autodock}}^{2\text{BE2}} + \Delta G_{\text{Autodock}}^{1\text{S9E}}) / 4 \\ \Delta G_{\text{consensus}}^{\text{mutant}} &= (\Delta G_{\text{XP GlideScore}}^{1\text{JLA}} + \Delta G_{\text{XP GlideScore}}^{2\text{IC3}} \\ &\quad + \Delta G_{\text{Autodock}}^{1\text{JLA}} + \Delta G_{\text{Autodock}}^{2\text{IC3}}) / 4 \\ \Delta G_{\text{consensus}}^{\text{total}} &= (\Delta G_{\text{consensus}}^{\text{wild type}} + \Delta G_{\text{consensus}}^{\text{mutant}}) / 2\end{aligned}$$

Compounds with scores deeper than the threshold were then reranked on the basis of their interactions with key protein binding partners, such as hydrogen bonds with protein backbone atoms, that can be less prone to mutational variation and conformational variation. The latter may help identify those compounds less sensitive to mutations that could make the protein resistant to a drug. The electronegative heteroatoms with at least one free pair of electrons in thiophenes, furans, NH_2 -phenyl, and X-phenyl ($X = \text{F}, \text{Cl}, \text{Br}$) were not counted as hydrogen bond acceptors in the hydrogen bond-based ranking since they are weak acceptors. Additional characterization was also provided by several ligand efficiency indices.

Chemical structure searches were performed postdocking for the highest ranked hits using SciFinder.⁵⁸ The compounds with an exact match were analyzed.

Molecular Physicochemical Properties and Efficiency Indices. Marvin Beans⁵⁹ were used to calculate the molecular surface area (MSA), polar surface area (PSA), MW, rotatable bonds,

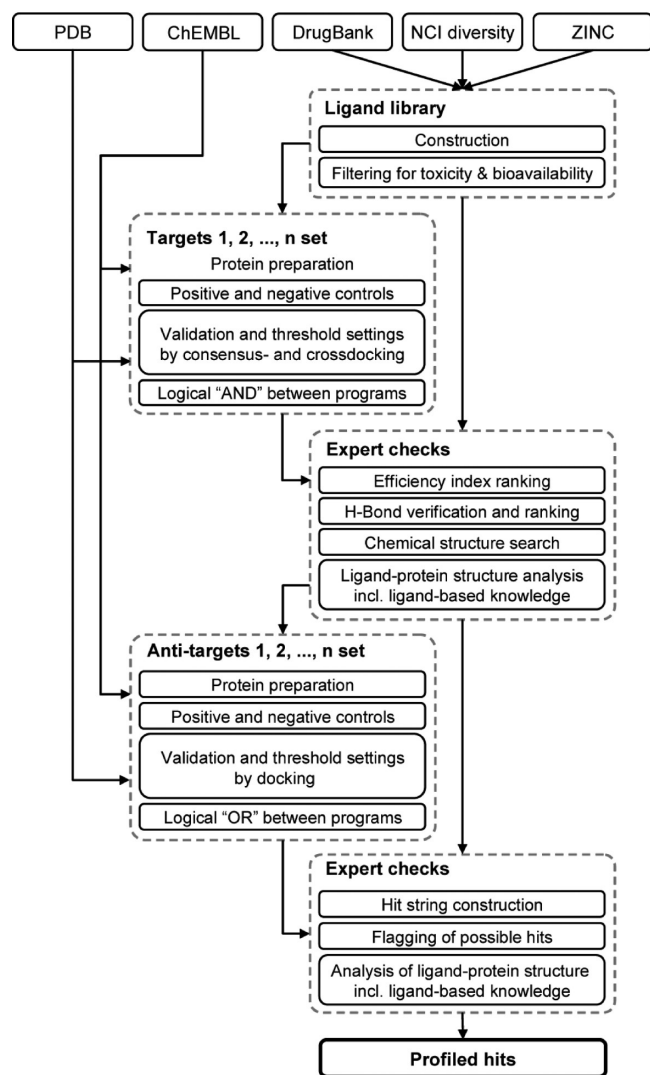


Figure 1. Method workflow.

Wiener index (W , a topological intramolecular connectivity index describing the summation of the edges in the shortest paths between all the heavy atoms), hydrogen bond acceptor atom count, hydrogen bond donor atom count, and logarithm of the partition coefficient between octanol and water ($\log P$) of compounds. Some of these properties were used in addition to the number of carbons (NoC) for calculating ligand efficiency indices as $\Delta G_{\text{consensus}}^{\text{total}} / \text{NF}$, where $\Delta G_{\text{consensus}}^{\text{total}}$ is the consensus, combined binding score between ligand and all target proteins and for both docking programs and NF is the normalization factor, such as MW, W , PSA, etc. OpenBabel⁶⁰ was used for calculating Tanimoto similarity coefficients.

The described methodological steps were used sequentially as indicated in the schematic representation of the workflow in Figure 1. The workflow shows the starting points in terms of data sources, construction, validation, and analysis of protein batteries, as well as the stops for expert checks between those steps.

Many of the procedures in Figure 1 are run manually, and any fully automated workflow would need to ensure that all the checks in the method workflow, such as threshold setting, binding mode evaluation, and ligand structure comparison, are adequately included. Many of the programs run in Figure 1 are not

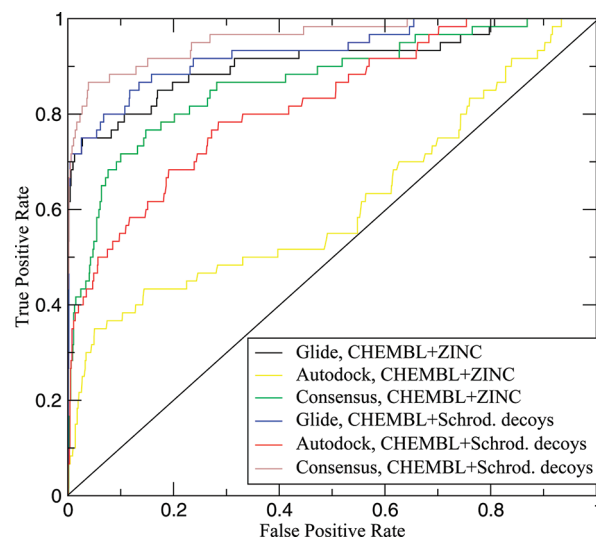


Figure 2. ROCs for screening of known actives and decoys against HIV-RT compared to a random control (diagonal line).

free to redistribute, but the programs to control them written in-house (such as launching the docking programs and postprocessing the ligand files) are available by request from the authors, and input files are shown in Supporting Information files S1 and S2.

RESULTS AND DISCUSSION

ROCs and AUCs. Figure 2 shows the ROC curves for the consensus screening, as well as for the individual programs/scoring functions, and a random control (diagonal line). The area under the curves was also calculated for each procedure, resulting in 0.91 for Glide, 0.62 for Autodock, and 0.87 for the consensus. They also had high enrichment factors at 5% of the database screened of 75, 35, and 50, respectively. These high values for the consensus procedure indicate early retrieval of known and potent inhibitors and usefulness in screening and classifying actives and decoys in the top ranks.

From Figure 2, it can be seen that Glide performed better than Autodock when using ChEMBL actives plus the small subset of random ZINC compounds as decoys. This affects the consensus procedure, since it degrades the quality of retrieval of actives. Consensus scoring may decrease the performance of an individual method if one of the methods is markedly worse than the other. However, consensus scoring is a hedging method that allows diminishing the risk of false positives and false negatives that could arise when only one docking program/scoring function is considered. When using the ChEMBL actives plus the Schrödinger ligand decoy set, Glide again performs better than Autodock, but here the consensus procedure performs better than either of the methods on their own.

Cross-docking between native ligands and related protein structures showed reasonable agreement: The cognate ligand for wild-type structure 2BE2 (HET-ID R22) docked into the wild-type structure 1S9E gave a heavy atom rmsd of 1.43 Å (XP GlideScore, -12.24 kcal/mol; Autodock4, -11.07 kcal/mol). The cognate ligand for wild-type structure 1S9E (HET-ID ADB) docked into the wild-type structure 2BE2 without a tightly bound water molecule gave an rmsd of 1.9 Å (XP GlideScore, -11.11 kcal/mol; Autodock4, -9.64 kcal/mol), while that docked into the

wild-type structure 2BE2 including the tightly bound water molecule gave an rmsd of 1.34 Å (XP GlideScore, -10.66 kcal/mol; Autodock4, -9.36 kcal/mol), which indicates that the water molecule produces binding poses with closer resemblance to the original cognate structure. The cognate ligand for mutant structure 1JLA (HET-ID TNK) docked into the structure of mutant protein 2IC3 gave an rmsd of 0.61 Å (XP GlideScore, -13.09 kcal/mol; Autodock4, -8.73 kcal/mol). The cognate ligand for mutant structure 2IC3 (HET-ID HBY) docked into the structure of mutant protein 1JLA gave an rmsd of 1.56 Å (XP GlideScore, -12.63 kcal/mol; Autodock4, -7.99 kcal/mol). For the mutant proteins, XP GlideScore appeared to score the cross-docked ligands deeper than Autodock4, showing a better agreement between scoring procedures for closely related proteins. All cross-docked ligands showed good docking poses with a smaller than 2 Å rmsd from the original X-ray pose.

Profiled Ligands. Table 3 shows the top-ranked ligands, as well as known drugs used against HIV-1 RT, in addition to known drugs used for different diseases. The new compounds are ranked according to the number of hydrogen bonds formed between the ligand and the protein main chain atoms (only those compounds that made two hydrogen bonds to the main chain of the protein—i.e., two asterisks—in most of the protein structures were selected). Compounds are further ranked according to the ligand efficiency index of predicted binding score/polar surface area ($\Delta G_{\text{consensus}}^{\text{total}}/\text{PSA}$). $\Delta G_{\text{consensus}}^{\text{total}}/\text{NHA}$ and $\Delta G_{\text{consensus}}^{\text{total}}/\text{MW}$ are also presented in Table 3. The other four calculated ligand efficiencies are shown in Table S1 in the Supporting Information. The total change in apolar solvent-accessible surface area upon binding ($\Delta\text{SASA}_{\text{apolar}}$) is also presented in Table 3 as an approximation to the total change in apolar molecular surface area, which has been shown to be correlated with the binding free energy for a database of thermodynamic measurements,⁶¹ in addition to the percentage of the total apolar surface area ($\%\Delta\text{SASA}_{\text{apolar}}$) as a measure of the fit between the ligand and protein binding site.

The consensus scoring results for known drugs suggest a broad agreement with experimental phenomena; etravirine and efavirenz were predicted to have a probability to bind to both wild-type and mutant proteins. Most of the compounds included in Table 3 have at least two hydrogen bonds to the protein main chain in all predicted protein–ligand structures. The predicted binding score of each compound was computed against the wild-type and drug-resistant protein targets (four proteins, five structures), as well as against the antitargets (five proteins). This gives the predicted interactions between each ligand with 10 protein structures in total, each calculation verified by both docking programs, to design drug candidates which may then be tested experimentally.

The number of hydrogen bonds to the protein main chain atoms (both to wild-type and drug-resistant mutant proteins) is used to rank compounds in Table 3 for five reasons: (1) they have been suggested in the context of kinase inhibitors to give an indication of how reliable the docking calculation may be to avoid false positives;⁶² (2) protein main chain atoms are frequently easier to assign to electron density maps from X-ray crystal structures, as well as having lower *B* factors, than side chain atoms, and therefore, their positioning may be more reliable than that of side chain atoms; (3) protein side chain atoms are more mobile and have increased conformational variability compared to main chain atoms, so counting hydrogen bonds to the main chain may detect interactions that persist for longer periods of time; (4) interactions between the ligand and the protein main chain

atoms may leave residual mobility in the protein that could contribute favorably to the entropic part of the binding energy;⁶³ (5) protein main chain atoms are much less frequently changed due to residue mutations than side chain atoms, so accounting for them to have a larger weight than side chain hydrogen bond interactions can help discover and design compounds with less exposure to the frequent viral protein mutations.

The free energy of binding of a ligand can be normalized per unit of measure thanks to the recently introduced ligand efficiency (or binding efficiency) indices.^{64–70} This normalization effect is important as it can remove size effects and therefore optimize compounds on the basis of their effective binding and pharmacokinetic related properties. Our previous results using ligand efficiencies include the improvement of the correlation between calculated and experimental values of drug–protein binding efficiencies,⁷⁰ as well as separating drugs from nondrug compounds.⁷¹ Small-molecule compounds able to disrupt (and therefore inhibit) large surface protein–protein interactions were also found to be characterized by their ligand efficiency, specifically a value of $\Delta G/\text{NHA}$ lower than -0.24 kcal/(mol·NHA).⁶⁸ This is reasonable if we consider that a typical submicromolar potent drug compound that has 50 or fewer heavy atoms would have a $\Delta G/\text{NHA}$ of around -0.2 kcal/(mol·NHA) or lower. Assuming a submicromolar potency compound with a molar mass lower than 500 g/mol (as in one of the Lipinski rule thresholds⁷²) would imply an efficiency index $\Delta G/\text{MW}$ lower than -0.02 (kcal·g)/mol². Alternatively, considering values of $\text{NHA} = 35$ and $\text{MW} = 350$ g/mol for a lead-like compound, values of $\Delta G/\text{NHA}$ and $\Delta G/\text{MW}$ of around -0.29 kcal/(mol·NHA) and -0.029 (kcal·g)/mol², respectively, would be reasonably expected. A good prospective candidate compound would be required to have efficiency indices similar to or lower than these described values. A normalized comparative measure (therefore, a whole molecule property) such as a ligand efficiency index can help to better characterize a ligand than only a ΔG score based on atom type or functional group contributions, or only on its molecular properties alone, since it has been suggested that functional group contributions to protein–ligand binding are not additive.⁷⁴ The ligand efficiency measure of $\Delta G/\text{PSA}$ (PSA is the polar surface area of the compound, Å²) can combine the pharmacodynamic measure of the binding affinity with the pharmacokinetic measure of permeability or distribution estimated by the polar surface area. Using the data of different drugs with several protein targets from our previous study,⁷⁰ correlations between experimental $\Delta G/\text{PSA}$ and GlideXP-calculated $\Delta G/\text{PSA}$ with an $R^2 = 0.794$ ($n = 26$) and between experimental $\Delta G/\text{PSA}$ and Autodock-calculated $\Delta G/\text{PSA}$ with an $R^2 = 0.811$ ($n = 26$) are achieved. Considering a submicromolar inhibitor with a PSA of ca. 140 Å² (suggested as a top limit for cell permeability),⁷⁴ we would expect a $\Delta G/\text{PSA}$ threshold of ca. -0.07 kcal/(mol·Å²), and designed compounds would need to have a $\Delta G/\text{PSA}$ value similar to or lower than this threshold. All of our proposed compounds in Table 3 meet this requirement.

There is a wide molecular diversity in the predicted binders in Table 3 that is reflected in their low Tanimoto similarity coefficients that have a mean value of 0.24 and a median value of 0.21. Furthermore, all of the small-molecule compounds in Table 3 pass Lipinski's rule-of-five,⁷² which may give an indication of them having good bioavailability, in addition to the compounds having less than 10 rotatable bonds and polar surface areas smaller than 140 Å² (with the exception of hit 6, $\text{PSA} = 144.86$ Å²). Many of the known inhibitors have moderate

Table 3. HIV-1 Reverse Transcriptase Profiled Ligands against Wild-Type and Drug-Resistant Mutant Strains^a

| Name or ID | Structure | $\Delta G/PSA$ | $\Delta G/NHA$ | $\Delta G/MW$ | $\Delta SASA_{apolar}$, % $\Delta SASA_{apolar}$ |
|--|-----------|----------------|----------------|---------------|--|
| efavirenz **, ** (**, *) -12.36, -10.75 | | -0.30 | -0.55 | -0.037 | 393.7, 88% |
| etravirine **, ** (*, **) -9.44, -9.90 | | -0.08 | -0.35 | -0.022 | 357, 87% |
| rilpivirine **, *(none, *) -9.83, -8.91 | | -0.10 | -0.33 | -0.026 | 320.6, 85% |
| nevirapine none, none(*, none) -7.14, -6.92 | | -0.12 | -0.36 | -0.027 | 206.5, 79% |
| oxazepam *, ** (**, *) -11.13, -11.6 | | -0.18 | -0.57 | -0.040 | 313.1, 86% |
| chlorthalidone *, ** (**, *) -10.95, -9.89 | | -0.10 | -0.47 | -0.031 | 283.9, 82% |
| 1: ZINC01645740 **, **** (**, ****) -12.12, -10.12 | | -0.13 | -0.51 | -0.035 | 252.1, 79% |
| 2: ZINC00880460 **, ** (**, **) -11.21, -10.98 | | -0.12 | -0.41 | -0.030 | 342.2, 84% |
| 3: ZINC05451764 **, **** (**, ****) -10.91, -10.99 | | -0.10 | -0.39 | -0.029 | 364.4, 85% |

Table 3. Continued

| Name or ID Hydrogen bond count (stars) ΔG w.t., ΔG mut. | Structure | $\Delta G/PSA$ | $\Delta G/NHA$ | $\Delta G/MW$ | $\Delta SASA_{apolar}$, $\% \Delta SASA_{apolar}$ |
|---|-----------|----------------|----------------|---------------|---|
| 4: ZINC08192573 *** (**,***) -11.72, -12.0 | | -0.10 | -0.41 | -0.031 | 383.2, 86% |
| 5: ZINC05606139 ** (**,***) -12.23, -11.53 | | -0.09 | -0.40 | -0.029 | 128, 60% |
| 6: ZINC02154371 ** (**,***) -10.84, -9.79 | | -0.07 | -0.40 | -0.029 | 258.9, 77% |
| 7: ZINC08391692 ** (**,***) -12.92, -11.71 | | -0.15 | -0.51 | -0.036 | 258.9, 77% |
| 8: ZINC03339025 ** (**,*) -11.81, -8.93 | | -0.12 | -0.40 | -0.029 | 328.7, 83% |
| 9: ZINC06589885 ** (**,*) -13.14, -11.49 | | -0.13 | -0.46 | -0.033 | 298.6, 80% |
| 10: ZINC08251923 ** (**,*) -12.91, -11.91 | | -0.10 | -0.44 | -0.033 | 333.8, 83% |
| 11: 211340NSC.1 ** (**,*) -10.12, -10.02 | | -0.08 | -0.48 | -0.033 | 215.7, 77% |

^a ΔG w.t. = consensus binding score to wild-type protein (kcal/mol), ΔG mut. = consensus binding score to mutant drug-resistant protein (kcal/mol), $\Delta G/PSA$ = combined predicted binding score divided by the polar surface area (kcal/(mol·Å²)), $\Delta G/NHA$ = combined predicted binding score divided by the number of heavy atoms (kcal/(mol·NHA)), $\Delta G/MW$ = combined predicted binding score divided by the molar mass ((kcal·g)/mol²), $\Delta SASA_{apolar}$ = change in the total buried apolar solvent-accessible surface area upon binding (Å²), and $\% \Delta SASA_{apolar}$ = change in the total buried apolar solvent-accessible surface area upon binding divided by the total apolar solvent-accessible surface area and expressed as a percentage. The name or ID of each compound is followed by asterisks indicating the number of hydrogen bonds between the ligand and protein main chain atoms. The order of asterisks is for wild-type structures 1S9E and 2BE2 (and in parentheses for drug-resistant 2IC3 and 1JLA).

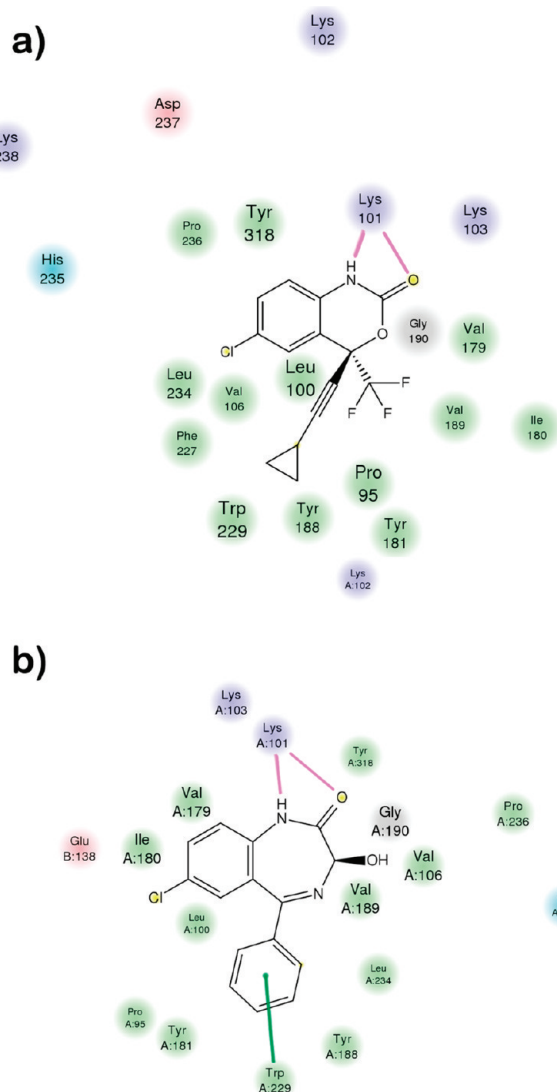


Figure 3. Binding site interactions between HIV-1 RT and (a) docked efavirenz and (b) docked oxazepam. Magenta lines are hydrogen bonds to main chain atoms, green lines are aromatic–aromatic interactions, green residues are hydrophobic contacts, cyan residues are polar contacts, blue residues are positively charged contacts, and red residues are negatively charged contacts. The residue font size indicates the distance (larger letters are closer).

$\Delta G/PSA$ values within limits of -0.04 to -0.62 kcal/(mol $\cdot\text{\AA}^2$) determined for a previously studied set of different drugs and protein targets with known activity and crystal structure.⁷⁰ By improving $\Delta G/PSA$ values (either by increasing the binding affinity and/or by decreasing the polar surface area, i.e., decreasing the overall size and/or decreasing the number of polar functional groups in the molecule), there may be scope for improvement of HIV-1 RT inhibitors. Additional ligand efficiency indices are shown in Table S1 in the Supporting Information. Compounds with low efficiency index values may benefit from being easier to develop,⁷⁵ as well as having possibly less side effects.⁷⁶ Fragment-based drug design and ligand efficiency make likely the adoption of smaller and less complex compounds. This may indeed be desirable for further optimization and chemical modification of compounds, though conceivably, very small compounds can be more promiscuous and bind to a variety of targets,⁷⁷ as

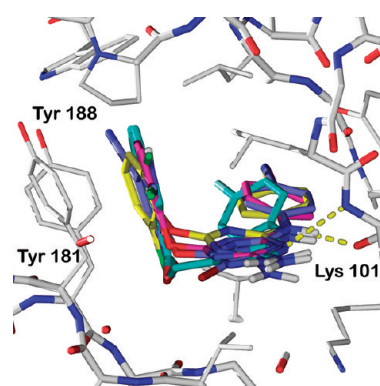


Figure 4. Original X-ray native (cognate) ligand (in yellow), docked binding poses for the known drug etravirine (in blue) and ligand 10 (in cyan), and docked native (cognate ligand, in magenta) pose (rmsd = 0.83 Å) in the binding site of HIV-1 reverse transcriptase (1S9E, in white). Hydrogen bonds to the main chain atoms are indicated by yellow dashes. Hydrogens are not shown for the protein. Polar hydrogens are shown for the ligands.

well as be more likely to cross the blood–brain barrier and have central nervous system effects that may or may not be a desired feature of a compound. All of the proposed compounds have predicted $\Delta G/MW$, $\Delta G/NHA$, $\Delta G/W$, $\Delta G/P$, and $\Delta G/NoC$ values that are within the limits from previous studies.^{37,67–70} In particular, their $\Delta G/NHA$ values are lower than the identified -0.24 kcal/(mol $\cdot NHA$) threshold that can define small-molecule inhibitors for protein–protein interactions.⁶⁸ Their predicted $\Delta G/NHA$, $\Delta G/MW$, and $\Delta G/PSA$ values are also deeper than our defined thresholds for bioavailable, cell-permeable, small-molecule, lead-like submicromolar compounds of -0.29 kcal/(mol $\cdot NHA$), -0.029 (kcal $\cdot g$)/mol², and -0.07 kcal/(mol $\cdot \text{\AA}^2$), respectively. Their predicted $\Delta G/W$ values are also between the limits of -0.001 and -0.07 kcal/mol determined in our previous studies on drug–protein systems.^{37,67,69,70} Their $\Delta SASA_{apolar}$ values (using a rough estimation of $\Delta CSA \approx 0.75(\Delta SASA)$)⁶¹ are also within the limits determined by Olsson et al. for a number of thermodynamically measured systems,⁶¹ and their % $\Delta SASA_{apolar}$ values show a high degree of burying of the apolar surfaces of both the protein and ligand, indicating a good fit.

The binding site of HIV-1 RT is nearly completely buried and enclosed within the protein structure. Efavirenz makes extensive hydrophobic interactions, as well as aromatic interactions, and donates and receives a hydrogen bond, respectively, to the backbone C=O and NH groups of Lys 101. It is interesting to note the chemical similarity of the known drugs oxazepam and chlorthalidone to efavirenz. They also have the same binding mode and similar interactions with HIV-1 RT compared to efavirenz. These include two fused rings, one aromatic with a chlorine substitution in the same position, an amide or urethane group on the top part of the second ring (similar to a quinolin-2-one), and a hydrophobic group on the lower part of the rings in a similar spatial arrangement. The binding interactions of efavirenz and oxazepam are shown in Figure 3. The binding interactions of chlorthalidone are similar, the ligands having the same orientation seen in Table 3. This may indicate that these widely used drugs may also bind to HIV-1 RT similarly to the approved drug efavirenz; that is, they may compose a pharmacophore with the same groups and with a similar distribution in space for important chemical factors responsible for interaction.

Table 4. Docking to Antitargets^a

| antitarget + ligand | GlideScore (kcal/mol) | Autodock binding free energy (kcal/mol) | ligand | GlideScore (kcal/mol) | Autodock binding free energy (kcal/mol) |
|---------------------------|--------------------------|--|---------------|--------------------------|--|
| SULT | | | | | |
| L-dopamine (cocrystal) | -6.3 | -7.5 | | | |
| ChapmanE1 | -7.1 | -7.5 | Schrod_484875 | >0 | +6.7 |
| CHEMBL167055 | -7.0 | -7.3 | Schrod_510180 | >0 | -4.4 |
| CHEMBL169033 | -7.5 | -6.3 | Schrod_558540 | +3.9 | -2.0 |
| Schrod_452534 | -3.7 | -5.4 | Schrod_621157 | >0 | -2.4 |
| Schrod_480983 | >0 | -1.4 | Schrod_721049 | >0 | -3.1 |
| Schrod_481019 | >0 | -6.7 | | | |
| PXR | | | | | |
| hyperforin (cocrystal) | -7.7 | -12.5 | | | |
| A-792611 | -16.0 | -17.6 | Schrod_481019 | >0 | -7.4 |
| CHEMBL456237 | -9.9 | -9.1 | Schrod_484875 | -7.3 | -10.8 |
| CHEMBL457977 | -10.0 | -8.5 | Schrod_510180 | >0 | -5.1 |
| CHEMBL59030 | -12.9 | -12.6 | Schrod_558540 | >0 | -2.5 |
| CHEMBL606702 | -12.1 | -12.7 | Schrod_621157 | >0 | -4.9 |
| Schrod_452534 | -2.7 | -5.4 | Schrod_721049 | >0 | -10.0 |
| Schrod_480983 | >0 | -6.0 | | | |
| CYP 2a6 | | | | | |
| coumarin (cocrystal) | -7.6 | -6.8 | | | |
| methoxsalen | -8.0 | -8.4 | CHEMBL369285 | -7.1 | -6.7 |
| CHEMBL178938 | -7.2 | -7.1 | CHEMBL386124 | -9.7 | -8.9 |
| CHEMBL179621 | -9.9 | -8.7 | Schrod_452534 | >0 | +141.4 |
| CHEMBL179669 | -9.9 | -8.7 | Schrod_480983 | >0 | +280.2 |
| CHEMBL179704 | -9.9 | -8.7 | Schrod_481019 | >0 | +171.2 |
| CHEMBL214859 | -7.9 | -7.8 | Schrod_484875 | >0 | +196.5 |
| CHEMBL214990 | -9.2 | -8.3 | Schrod_510180 | >0 | +31.4 |
| CHEMBL360999 | -8.1 | -6.5 | Schrod_558540 | >0 | +4.5 |
| CHEMBL361153 | -8.8 | -7.1 | Schrod_621157 | >0 | +228.6 |
| CHEMBL368883 | -8.2 | -6.6 | Schrod_721049 | >0 | +234.5 |
| CYP 2c9 | | | | | |
| S-warfarin (cocrystal) | -8.7 | -9.4 | | | |
| sulfaphenazole | -9.2 | -9.1 | CHEMBL464595 | -10.3 | -11.5 |
| CHEMBL1109 | -9.2 | -9.2 | CHEMBL514730 | -10.4 | -11.6 |
| CHEMBL455975 | -8.9 | -11.6 | Schrod_452534 | -4.9 | -9.3 |
| CHEMBL455976 | -10.6 | -11.4 | Schrod_480983 | >0 | -7.7 |
| CHEMBL456181 | -10.7 | -10.9 | Schrod_481019 | -8.3 | -9.9 |
| CHEMBL456432 | -9.1 | -9.8 | Schrod_484875 | -2.8 | -9.6 |
| CHEMBL457087 | -8.3 | -10.1 | Schrod_510180 | >0 | -4.4 |
| CHEMBL458566 | -8.6 | -9.4 | Schrod_558540 | -0.8 | -1.4 |
| CHEMBL458567 | -8.3 | -9.3 | Schrod_621157 | >0 | -7.2 |
| CHEMBL463577 | -10.6 | -11.6 | Schrod_721049 | -2.9 | -7.6 |
| CHEMBL463976 | -9.2 | -9.5 | | | |
| CYP 3a4 | | | | | |
| nephazodone | -9.6 | -11.7 | Schrod_452534 | -1.4 | +47.2 |
| CHEMBL1089957 | -6.3 | -23.3 | Schrod_480983 | >0 | +20.8 |
| CHEMBL270271 | -7.2 | -12.5 | Schrod_481019 | -4.3 | +5.5 |
| CHEMBL271580 | -6.5 | -12.6 | Schrod_484875 | >0 | +23.6 |
| CHEMBL507731 | -6.5 | -26 | Schrod_510180 | >0 | +0.6 |
| CHEMBL573665 | -6.7 | -23.2 | Schrod_558540 | >0 | -2.6 |
| CHEMBL583954 | -6.5 | -26.6 | Schrod_621157 | >0 | +27.0 |
| CHEMBL75 | -6.3 | -17.9 | Schrod_721049 | >0 | +11.8 |
| CHEMBL98745 | -9.8 | -11.4 | | | |

^a Thresholds in bold.

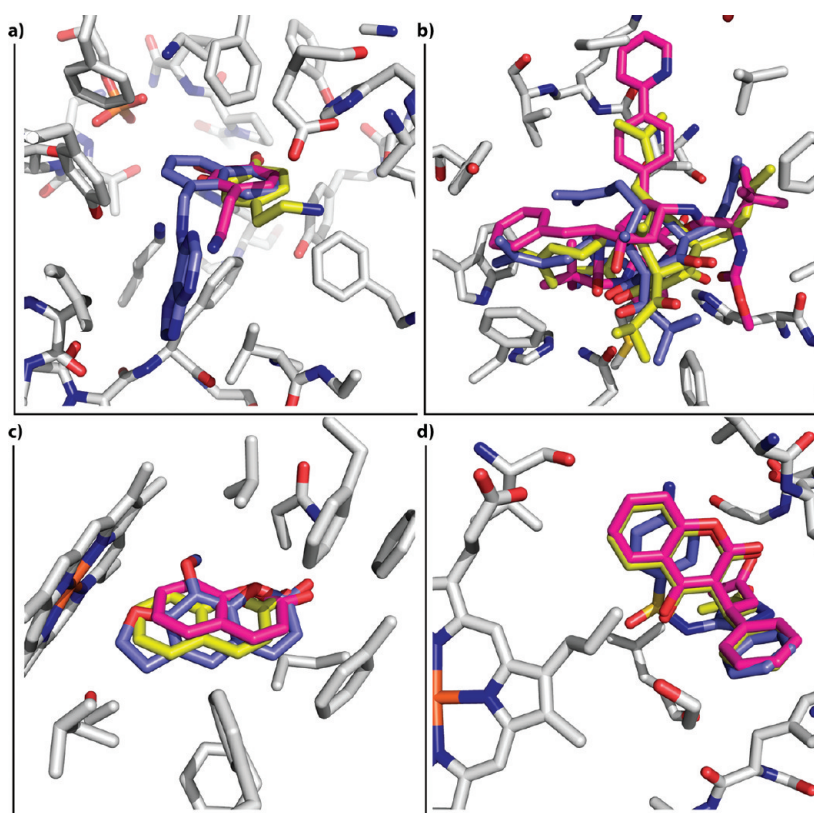


Figure 5. Superposition of binding modes between the extra set of known active binders (backbone in blue) and the original X-ray native (cognate) cocrystallized ligands (backbone in yellow) and the docked native (cognate) ligands (backbone in magenta) for the set of antitargets (backbone in white). rmsd's between the original X-ray cognate ligand and the docked cognate ligand binding pose: (a) SULT, rmsd = 1.81 Å; (b) PXR, rmsd = 1.97 Å; (c) CYP2a6, rmsd = 0.95 Å; (d) CYP2c9, rmsd = 0.54 Å.

Chlorthalidone is an antihypertensive/diuretic monosulfamyl molecule used in oral form and indicated in the management of hypertension.⁷⁸ Oxazepam is used in the treatment of anxiety, alcohol withdrawal, and insomnia. From our calculations, it may appear that these well-known and in-clinical-use drugs could find a novel indication against this target and for the treatment of HIV-1. Oxazepam has been studied and shown to be safe to coadminister together with the nucleoside HIV-1 RT inhibitor zidovudine.^{79,80} A track record for known drugs of widespread use in the clinic may provide confidence in the safety of using them for possible new therapeutic indications. Here, the use of several *in silico* techniques, both ligand- and structure-based, provides a clearer picture of possible ligands than that allowed by using only one technique.⁸¹

Most of the highest ranked compounds (those with the best interaction scores, efficiency indices, and hydrogen bonds to the protein backbone) made two correlated hydrogen bonds with the protein backbone lysine 101, accepting a hydrogen bond from Lys 101 N and donating one to Lys 101 O. This interaction was highly rewarded by the XP GlideScore scoring function since correlated hydrogen bonds (pairs of acceptor/acceptor, acceptor/donor, or donor/donor atoms that are only one rotatable bond apart)⁴³ receive a larger reward to the binding score than isolated ones,⁴³ and in addition, they were formed in the hydrophobically enclosed volume in the binding site that was inaccessible even to the water molecules that the program uses to flood the binding site. The latter feature was further rewarded by the scoring function. This interaction is also formed by the

known and in-clinical-use drugs efavirenz and etravirine. Through the desolvation term in the Autodock4 scoring function, they also scored deeply for Autodock. Many of the top compounds presented a central triazine functional group. This group has an appropriate structure and geometry that allowed these compounds to interact with the protein Lys 101 backbone atoms in a binding mode similar to that of known inhibitors. The drug etravirine and the inhibitor rilpivirine have the closely related pyrimidine group as the central core. Different aryltriazines have recently been reported to have wild-type HIV-1 RT activity,^{81,82} which shows that our method can identify compounds with chemical motifs similar to those reported to be experimentally active. Figure 4 shows the binding poses for etravirine and compound **10** docked into structure 1S9E and the close similarity between their binding interactions. Tyr 181 is shown in the front left part of the figure, with Tyr 188 in the back left. Lysine 101 is on the right-hand side of Figure 4.

Compound **10**, for example, also participates in many hydrophobic and aromatic interactions in the binding site as well as in hydrogen bonds to the main chain atoms of Lys 101. It makes hydrophobic contacts with Leu 100, Val 179, and Leu 234 and has aromatic contacts between its indole ring and the aromatic side chains of Trp 229, Tyr 181, and Tyr 188, as well as between its substituted benzene ring and Tyr 318. The 3-D view of the interactions and binding modes can be seen in Figure 4, while 2-D ("flat") representations are shown in Figure S1 in the Supporting Information. The polar part of the molecule points toward the opening of the binding site and would be more in

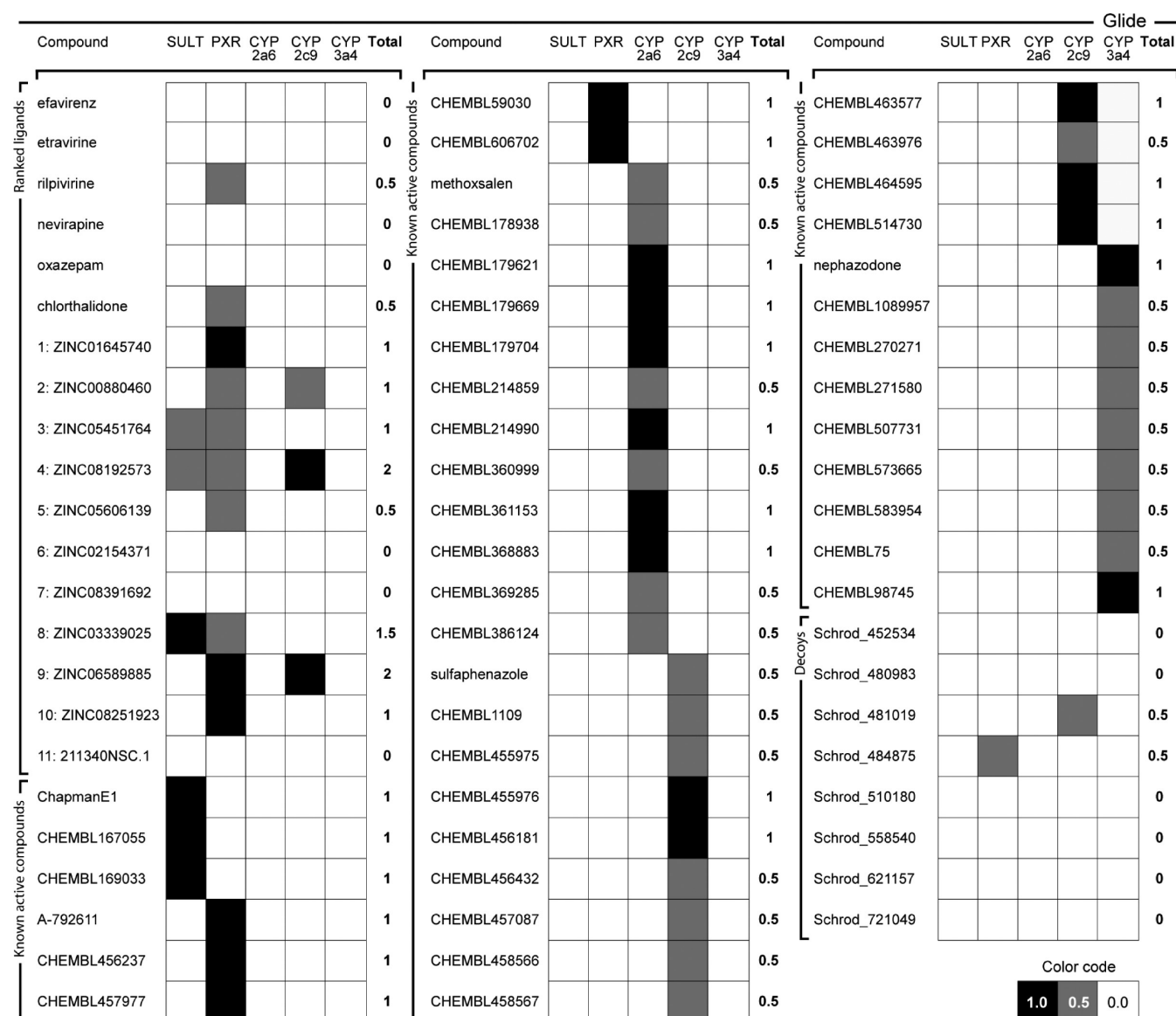


Figure 6. Interaction matrix between ligands and antitargets calculated with Glide.

contact with water molecules, therefore reducing the energetic cost of desolvation, and interacts with polar groups at the mouth of the binding site, such as Lys 101 and Glu 138 (chain B). Etravirine and efavirenz have similar binding interactions, only etravirine uses its substituted benzene rings and efavirenz uses its cyclopropylethynyl group to interact with the aromatic side chains of Tyr 181 and Tyr 188 in the binding site (3-D view in Figure 4, 2-D view in Figure S2 in the Supporting Information). Our procedure retrieved compounds that would interact well with either tyrosine or cysteine in the position sequence number 181 in chain A. In addition, the close similarity in binding poses between the original X-ray (cognate, not optimized) ligand and its docked binding pose ($\text{rmsd}(\text{heavy atoms}) = 0.83 \text{ \AA}$), and between these and the known drug etravirine and proposed ligand **10**, is clear from Figure 4.

Quinolin-2-ones (such as **2**) and chromen-2-ones (such as **6**) were also repeatedly present, which have close resemblance to the known inhibitor efavirenz and to the drug oxazepam. A few purines (such as **5** and **11**) and related rings (such as the indole **10**)

were also found. Compounds with charged groups such as ZINC05353109 (not shown) had relatively deep scores that are the result of ionic bonds between these compounds and the proteins. However, this can also result in a too high polar surface area, which reduces the compound's $\Delta G/\text{PSA}$ efficiency index, as is the case with **11**.

Commercially available collections such as those compiled in the ZINC database typically contain a large amount of compounds that seem to have been synthesized for other design efforts, for example, kinase inhibitor design, as many contain the typical sugar unit, a purine or pyrimidine base, and a polar tail intended to mimic ATP for ATP-competitive kinase inhibitors. For some compounds docked in this study, interaction with the backbone atoms of Lys 101 of HIV-1 RT echoes the interactions formed between kinase inhibitors and the backbone atoms of the hinge Leu 83 in cyclin-dependent kinase 2 (CDK2).²⁷ The ZINC database also contains compounds that resemble tricyclic antidepressants, nucleosides, and other typical core structures from different therapeutic areas. However, our filtering and docking

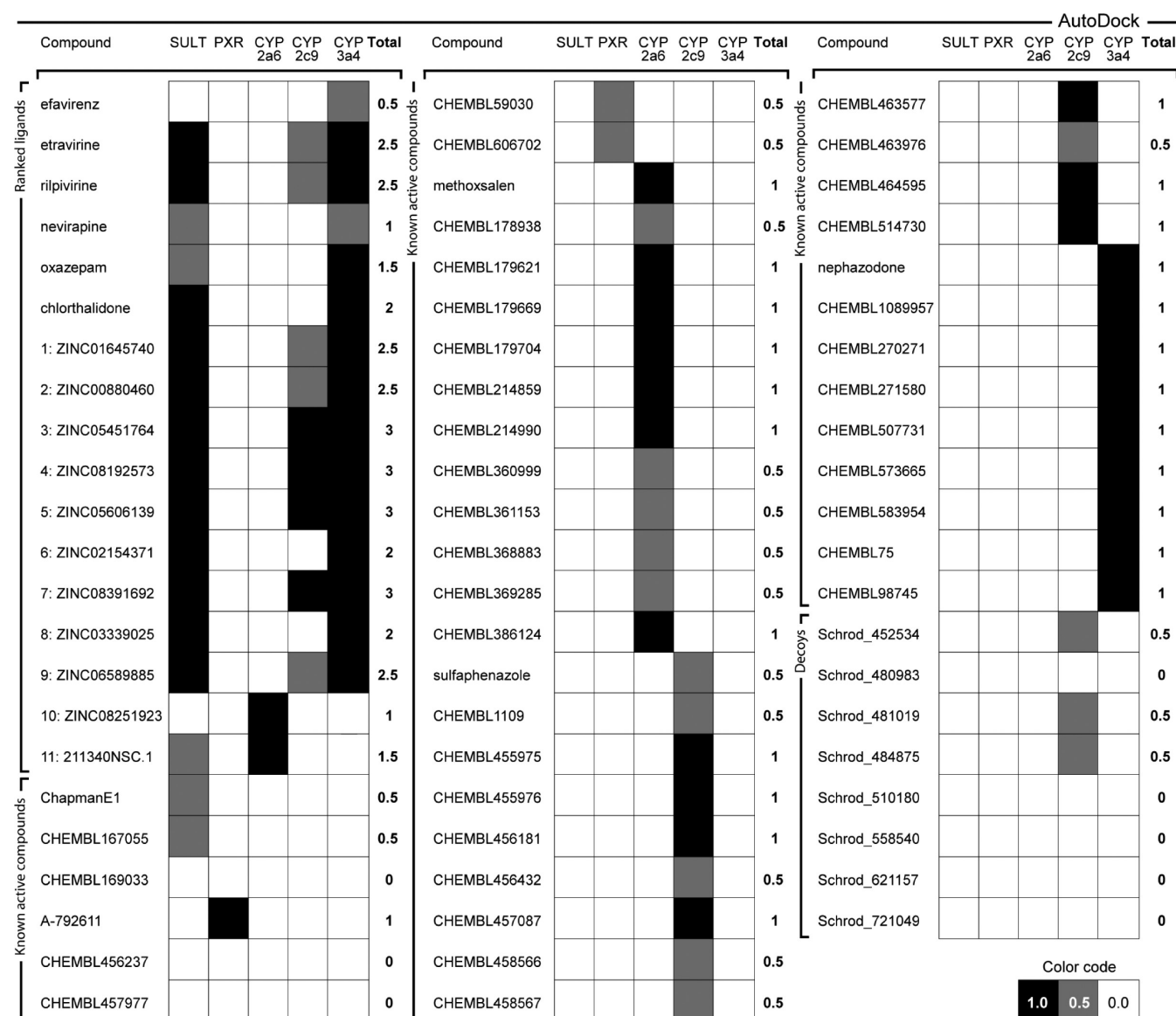


Figure 7. Interaction matrix between ligands and antitargets calculated with Autodock.

method retrieved those compounds with good predicted properties against both wild-type and drug-resistant HIV-1 RT, as well as recovering the known binders in Table 3 in the top 1% of the screened database. As confirmation that compounds may have more than one pharmacological activity, a chemical structural search on the proposed compounds showed that antitumor properties (as a p53 activator) have been determined for 11 (NSC211340)⁸⁴ and micromolar anti-influenza,⁸⁵ as well as antitumor (leukemia) properties⁸⁶ have been determined for 1 (ZINC01645740, also named CAS18020459-5 and NSC676348). Together, the analysis of the ligand compound libraries, such as the NSC and ZINC collections, shows that they must be searched and filtered carefully.

Antitargets. The antitarget binding sites varied from the small, enclosed, and solvated CYP binding sites to the very tight binding site of SULT and to the large, hydrophobic PXR binding site. A pharmacophore has been suggested for PXR that has predominantly hydrophobic as well as a few hydrogen bond donor and acceptor features.^{87–91} CYP 2a6 has been proposed to

be bound by ligands forming a pharmacophore containing a planar hydrophobic element, in addition to a single hydrogen bond donor/acceptor.⁹² CYP 2c9 binding has been proposed to have a hydrophobic group plus an anionic moiety and/or one^{93,94} or two⁹⁵ hydrogen bond acceptors 7 Å from the site of metabolism. CYP 3a4 can accommodate several substrates simultaneously in its binding site,^{96–98} which increases its binding features,⁹⁹ and a pharmacophore composed of two hydrogen bond acceptors, one hydrogen bond donor, and a hydrophobic region has been proposed,¹⁰⁰ among others.^{13–15} Many of these proteins are promiscuous in the sense that they accept a range of differently sized ligands due to the size of their binding pocket (PXR, CYP3a4) and/or their flexibility (CYPs). The thresholds for recording a hit against an antitarget are shown in Table 4 and are based on both the GlideScore and Autodock score of native ligands. A value of -7.5 kcal/mol was heuristically adopted for CYP P450 3a4 as a threshold value since there was no cocrystallized ligand present in the structure and is comparable to the calculated value of known ligands. Table 4 also shows the

recorded interaction score of the extra set of known active compounds, as well as some Schrödinger decoys.

The general concurrence in the scores of the cocrystallized known ligands in Table 4 with those of the extra, spiked set of known inhibitors for the same antitargets, as well as the result that none of the known nonbinders (inactives) nor most of the decoys scored deep enough as compared to the respective cocrystallized ligand, suggests that our procedure may be valid to compare to experimentally observed phenomena of probable binders vs nonbinders. This was also confirmed by the close similarity in binding modes between the extra set of known binders and the native (cognate) cocrystallized ligands, as well as the docked native (cognate) cocrystallized ligands, as seen in Figure 5. The rmsd of heavy atoms between the original X-ray cognate ligand (not optimized) and its docked pose was less than 2 Å for all cases.

A visual representation of the predicted compound antitarget interaction arrays from Table 4 is shown in Figures 6 and 7, where the binding interactions determined, respectively, by Glide and Autodock against the five antitargets for all the compounds in Table 4 are presented in color code. Heuristically, compounds that have a high array sum (combined hit string, i.e., sum of all protein interactions with all docking programs) or a high individual array count (individual hit string, i.e., sum of all protein interactions with one docking program) may be flagged for warning to establish a rapid method to detect problem molecules early during hit identification. There is less consensus between the docking programs as shown for the HIV-1 RT hits, but this is by design, since for antitarget hits we are interested in identifying compounds that may have any predicted interaction by *either* of the programs, whereas for HIV-1 RT, the onus was on finding those hits with high ranks by *both* programs. XP GlideScore performed better than Autodock4 at early retrieval of actives, though both performed reasonably well at scoring antitarget positive and negative controls (actives and inactive + decoys, respectively).

A perspective similar to that presented in Figures 6 and 7 for the target proteins and the ligands shows that most of the compounds have a strong interaction with all of the target proteins using both docking programs (all squares black or gray). This enabled selection with a higher degree of confidence of those candidate compounds for all of the targets in question, in consensus between different methods (logical “AND”), as well as detection of those with possible interactions against the antitarget set, as indicated by either docking program (logical “OR”). The antitargets can be further assembled and used in the manner described in the present work and be tailored to suit antitargets of interest, individual design projects, or ligand side effect and off-target profiling.

Another way of including consideration of possible toxic or metabolic effects is by ligand similarity searches¹⁰¹ and keeping track of known functional groups that can cause metabolism and toxicity problems, such as polyhalogenated compounds, aromatic nitro groups, bromoarenes, and hydrazines, for example,⁶ or by calculating GRID molecular interaction fields for the antitarget site and comparing them to those of the substrate (as in MetaSite).¹⁰² Interaction sites for molecules with CYP proteins for different fragment patterns can be scored on the basis of the interaction of typical functional groups in the different CYP binding sites.¹⁰³ The profiled compounds in Table 3 and Figure 5 were also verified according to the ADMET score of Gleeson et al.,¹⁰⁴ and the vast majority had scores under 1 (i.e., they were comparable to known oral drugs; see Table S2 in the Supporting Information).

In addition to broadly distinguishing likely binders for the HIV-1 RT wild-type and drug-resistant mutants, and their known inhibitors and nonbinders, the known HIV-1 RT inhibitors had low to moderate interaction arrays against the set of antitargets (though less consensus between programs, Figures 6 and 7, which was introduced by design), and the known antitarget inhibitors had strong interactions (deep scores) with the antitargets (by broad consensus between programs, Table 4). The majority of the highly ranked ligands in Table 3 also have low to moderate predicted interaction arrays against the antitarget set (Figures 6 and 7). Together, these results suggest the validity of our proposed model to distinguish candidate compounds against targets of interest while recognizing those that can have a probability of interaction with an antitarget. It must be stressed that this does not supplant the need for experimental determination of metabolic effects, but provides an early warning system to flag compounds for further investigation and consideration in drug design.

CONCLUSIONS

The results show that as early on as possible in the drug discovery process, during hit identification (hit discovery) and before lead optimization, a process should be established where compound structure libraries are simultaneously screened against specific antitargets and on-targets, and this may help to identify early some of the possible metabolism effects of ligands, as well as their activities. Each different optimization strategy may build its own set of antitargets, which can be calibrated according to specific constraints. Simultaneously, novel targets for interaction of the studied compounds can be identified by including other proteins in the set of targets. Off-targets may be included as extra targets if beneficial or as antitargets if nondesired. Though a small group of different conformations (the important Tyr 181 side chain conformation, in three different structures—one of them including a tightly bound water molecule) and a mutation (the important, drug-resistance-conferring Y181C, in two different structures) were included in this study, more conformations of the same protein generated through X-ray crystal structure determination, NMR solution structure determination, and/or MD simulations can be added to expand both the target and antitarget sets. This may help to include protein and protein–ligand complex flexibility and perhaps induced-fit effects. Flexibility may be more of an issue for certain proteins over others that may be more rigid (for example, PXR). Any predictions made computationally should be finally tested by experiment.

A number of interesting, structurally diverse, small-sized compounds were found that may interact with both wild-type and Y181C drug-resistant HIV-1 reverse transcriptases. A few known drugs with a different indication were identified as possible binders of HIV-1 reverse transcriptases. Furthermore, compounds resembling the known drugs efavirenz and etravirine were discovered, in addition to compounds with new chemistry. Compounds containing a triazine core and aromatic side chains may represent an alternative structural core and drug design route that is based on the similarity of ligand chemical functional group arrangement, binding mode, and interactions to the protein to those of the known inhibitors etravirine and rilpivirine. We also indicate a way of improving candidate HIV-1 RT binders by achieving lower $\Delta G/PSA$ ligand efficiency values. All of the proposed compounds are predicted to have good bioavailability, cell permeability, and buried total apolar surface area.

Many of these ligands had low predicted interaction arrays against some of the known metabolism proteins. The known drugs chlorthalidone and oxazepam had quite low predicted interaction against the set of antitargets, as well as molecular structures and binding patterns (poses) to HIV-1 reverse transcriptase similar to those of the known drug efavirenz. This may show how approved drugs that are already in use and are relatively safe have low interactions with the antitarget set that we have studied, and thus, our procedure may provide a way to design sets of antitargets and interaction arrays that could, in principle, help to improve the metabolic interactions of candidate compounds. Our approach included broadly following experimentally observed binders and nonbinders for both targets and antitargets, not for supplanting experimental testing, which is still required, but to provide an early warning system for multitarget optimization.

Structural searches carried out postdocking allowed finding other uses for the proposed compounds, and such searches should be carried out as standard procedure in virtual screening. In addition, short postdocking minimizations can resolve small strains and deviations from a *trans*, planar ester or amide conformation in docked ligand poses.

■ ASSOCIATED CONTENT

S Supporting Information. Input files for docking, ligand efficiency indices for the compounds in Table 3, two-dimensional views of the interactions and binding modes of etravirine, as well as those of compound **10**, with HIV-1 reverse transcriptase, and ADMET scores for all proposed compounds. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alfonsog@ut.ee; phone: +372 737 5270; fax: +372 737 5264.

■ ACKNOWLEDGMENT

We thank the Estonian Ministry for Education and Research (Grant SF0140031Bs09), Estonian Science Foundation, and University of Tartu High Performance Computing Centre for computing resources.

■ REFERENCES

- (1) World Health Organization (WHO). *Women and Health: Today's Evidence Tomorrow's Agenda* [Online]; WHO: Geneva, Switzerland, 2009. <http://www.who.int/gender/documents/9789241563857/en/index.html> (accessed March 19, 2009).
- (2) Preston, B. D.; Poiesz, B. J.; Loeb, L. A. Fidelity of HIV-1 Reverse-Transcriptase. *Science* **1988**, *242*, 1168–1171.
- (3) Ewald, P. W. Guarding against the Most Dangerous Emerging Pathogens. *Emerging Infect. Dis.* **1996**, *2*, 245–257.
- (4) Wilson, L. E.; Gallant, J. E. The Management of Treatment-Experienced HIV-Infected Patients: New Drugs and Drug Combinations. *Clin. Infect. Dis.* **2009**, *48*, 214–221.
- (5) Stanic, A.; Grana, J. C. Review of Antiretroviral Agents for the Treatment Of HIV Infection. *Formulary* **2009**, *44*, 47–54.
- (6) Patrick, G. L. *An Introduction to Medicinal Chemistry*, 3rd ed.; Oxford University Press: New York, 2005; pp 235, 450–473, and 510–511.
- (7) Nichols, S. E.; Domoal, R. A.; Thakur, V. V.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L. Discovery of Wild-Type and Y181C

Mutant Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors Using Virtual Screening with Multiple Protein Structures. *J. Chem. Inf. Model.* **2009**, *49*, 1272–1279.

(8) Petsko, G. When Failure Should Be the Option. *BMC Biol.* **2010**, *8*, 61–61.

(9) Nidhi; Glick, M.; Davies, J. W.; Jenkins, J. L. Prediction of Biological Targets for Compounds Using Multiple-Category Bayesian Models Trained on Chemogenomics Databases. *J. Chem. Inf. Model.* **2006**, *46*, 1124–1133.

(10) Azzaoui, K.; Hamon, J.; Faller, B.; Whitebread, S.; Jacoby, E.; Bender, A.; Jenkins, J. L.; Urban, L. Modeling Promiscuity Based on in Vitro Safety Pharmacology Profiling Data. *ChemMedChem* **2007**, *2*, 874–880.

(11) Gregori-Puigjané, E.; Mestres, J. Coverage and Bias in Chemical Library Design. *Curr. Opin. Chem. Biol.* **2008**, *12*, 359–365.

(12) Keiser, M. J.; Roth, B. L.; Armbruster, B. N.; Ernsberger, P.; Irwin, J. J.; Shoichet, B. K. Relating Protein Pharmacology by Ligand Chemistry. *Nat. Biotechnol.* **2007**, *25*, 197–206.

(13) Ortiz de Montellano, P. R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* **2010**, *110*, 932–948.

(14) Sun, H.; Scott, D. O. Structure-Based Drug Metabolism Predictions for Drug Design. *Chem. Biol. Drug Des.* **2010**, *75*, 3–17.

(15) *Antitargets. Prediction and Prevention of Drug Side Effects*; Vaz, R. J., Klabunde, T., Eds.; Series: Methods and Principles in Medicinal Chemistry; Wiley-VCH: Weinheim, Germany, 2008; pp xix–xxv.

(16) Kumar, G. N.; Rodrigues, A. D.; Buko, A. M.; Denissen, J. F. Cytochrome P450-Mediated Metabolism of the HIV-1 Protease Inhibitor Ritonavir (ABT-538) in Human Liver Microsomes. *J. Pharmacol. Exp. Ther.* **1996**, *277*, 423–431.

(17) Nhan, C.-H.; Beglov, D.; Rudnitskaya, A. N.; Kozakov, D.; Waxman, D. J.; Vajda, S. The Structural Basis of Pregnane X Receptor Binding Promiscuity. *Biochemistry* **2009**, *48*, 11572–11581.

(18) Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, A. J.; Bork, P. Drug Target Identification Using Side-Effect Similarity. *Science* **2008**, *321*, 263–266.

(19) Xie, L.; Evangelidis, T.; Bourne, P. E. Drug Discovery Using Systems Biology: Weak Inhibition of Multiple Kinases May Contribute to the Anti-Cancer Effect Of Nelfinavir. *PLoS Comput. Biol.* **2011**, *7*, e1002037.

(20) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can We Rationally Design Promiscuous Drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127–136.

(21) Hopkins, A. L. Network Pharmacology: The Next Paradigm in Drug Discovery. *Nat. Chem. Biol.* **2008**, *4*, 682–690.

(22) Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijter, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting New Molecular Targets for Known Drugs. *Nature* **2009**, *462*, 175–181.

(23) Eisenhofer, G.; Coughtrie, M. W. H.; Goldstein, D. S. Dopamine Sulphate: An Enigma Resolved. *Clin. Exp. Pharmacol. Physiol.* **1999**, *26*, S41–S53.

(24) Protein Data Bank. Research Collaboratory for Structural Bioinformatics. <http://www.pdb.org/pdb/home/home.do> (accessed April, 7, 2010).

(25) Paris, K. A.; Haq, O.; Felts, A. K.; Das, K.; Arnold, E.; Levy, R. M. Conformational Landscape of the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Non-Nucleoside Inhibitor Binding Pocket: Lessons for Inhibitor Design from a Cluster Analysis of Many Crystal Structures. *J. Med. Chem.* **2009**, *52*, 6413–6420.

(26) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including Tightly-Bound Water Molecules in de Novo Drug Design. Exemplification through the in Silico Generation of Poly(ADP-Ribose) Polymerase Ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624–633.

(27) García-Sosa, A. T.; Mancera, R. L. The Effect of a Tightly Bound Water Molecule on Scaffold Diversity in the Computer-Aided de Novo Design of CDK2 Inhibitors. *J. Mol. Model.* **2006**, *12*, 422–431.

- (28) García-Sosa, A. T.; Mancera, R. L. Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand-Protein Complex. *Mol. Inf.* **2010**, *29*, 589–600.
- (29) *Maestro*, version 9.1; Schrödinger, LLC: New York, 2007.
- (30) *Protein Preparation Wizard*, version 1.0; Schrödinger, LLC: New York, 2009.
- (31) European Molecular Biology Laboratory—European Bioinformatics Institute. ChEMBL Chemical Compound Database. <https://www.ebi.ac.uk/chembl/db/> (accessed April 14, 2011).
- (32) Irwin, J. J.; Shoichet, B. K. ZINC: A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
- (33) Schrödinger, Ltd. Schrödinger Glide Ligand Decoy Set. <http://www.schrodinger.com/downloadcenter> (accessed July 10, 2011).
- (34) Swamidass, S. J.; Azencott, C.-A.; Daily, K.; Baldi, P. A CROC Stronger Than ROC: Measuring, Visualizing, and Optimizing Early Retrieval. *Bioinformatics* **2010**, *26*, 1348–1356.
- (35) National Cancer Institute/National Institutes of Health USA. DTPs Diversity Set Information. http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html (accessed April 7, 2010).
- (36) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.* **2006**, *34*, D668–D672.
- (37) García-Sosa, A. T.; Sild, S.; Maran, U. Design of Multi-Binding-Site Inhibitors, Ligand Efficiency, and Consensus Screening of Avian Influenza H5N1 Wild-Type Neuraminidase and of the Oseltamivir-Resistant H274Y Variant. *J. Chem. Inf. Model.* **2008**, *48*, 2074–2080.
- (38) *Instant JChem*, version 2.1; ChemAxon Ltd.: Budapest, Hungary, 2007. <http://www.chemaxon.com>.
- (39) *LigPrep*, version 2.1; Schrödinger, LLC: New York, 2007.
- (40) *Virtual Screening Workflow*; Schrödinger, LLC: New York, 2007.
- (41) 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.
- (42) 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.
- (43) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (44) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (45) Yang, J.-M.; Chen, Y.-F.; Shen, T.-W.; Kristal, B. S.; Hsu, D. F. Consensus Scoring Criteria for Improving Enrichment in Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 1134–1146.
- (46) Friedman, R.; Caflich, A. Discovery of Plasmepsin Inhibitors by Fragment-Based Docking and Consensus Scoring. *ChemMedChem* **2009**, *4*, 1317–1326.
- (47) Chapman, E.; Ding, S.; Shultz, P. G.; Wong, C. A Potent and Highly Selective Sulfotransferase Inhibitor. *J. Am. Chem. Soc.* **2002**, *124*, 14524–14525.
- (48) Healan-Greenberg, C.; Waring, J. F.; Kempf, D. J.; Blomme, E. A.; Tirona, R. G.; Kim, R. B. A Human Immunodeficiency Virus Protease Inhibitor Is a Novel Functional Inhibitor of Human Pregnane X Receptor. *Drug Metab. Dispos.* **2008**, *36*, 500–507.
- (49) MaenPaa, J.; Juvonen, R.; Raunio, H.; Rautio, A.; Pelkonen, O. Metabolic Interactions of Methoxsalen and Coumarin in Humans and Mice. *Biochem. Pharmacol.* **1994**, *48*, 1363–1369.
- (50) Baldwin, S. J.; Bloomer, J. C.; Smith, G. J.; Ayrton, A. D.; Clark, S. E.; Chenery, R. J. Ketoconazole and Sulfaphenazole as the Respective Selective Inhibitors of P450 3A and 2C9. *Xenobiotica* **1995**, *25*, 261–270.
- (51) Spina, E.; Santoro, V.; D'Arrigo, C. Clinically Relevant Pharmacokinetic Drug Interactions with Second-Generation Antidepressants: An Update. *Clin. Ther.* **2008**, *30*, 1206–1227.
- (52) Cole, G. B.; Keum, G.; Liu, J.; Small, G. W.; Satyamurthy, N.; Kepe, V.; Barrio, J. R. Specific Estrogen Sulfotransferase Substrates and Molecular Imaging Probe Candidates. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 6222–6227.
- (53) Veluparthi, U.; Wittman, M.; Liu, P. Y.; Carboni, J. M.; Lee, F. Y.; Attar, R.; Balimane, P.; Clarke, W.; Sinz, M. W.; Hurlburt, W.; Patel, K.; Discenza, L.; Kim, S.; Gottardis, M.; Greer, A.; Li, A. X.; Saulnier, M.; Yang, Z.; Zimmermann, K.; Trainor, G.; Vyas, D. Discovery and Evaluation of 4-(2-(4-Chloro-1H-pyrazol-1-yl)ethylamino)-3-(6-(1-(3-fluoro-propyl)piperidin-4-yl)-4-methyl-1H-benzo[d]imidazol-2-yl)-pyridin-2(1H)-one (BMS-695735), an Orally Efficacious Inhibitor of Insulin-like Growth Factor-1 Receptor Kinase with Broad Spectrum Antitumor Activity. *J. Med. Chem.* **2008**, *51*, 5897–5900.
- (54) Juvonen, R. O.; Gynther, J.; Pasanen, M.; Alhava, E.; Poso, A. Pronounced Differences in Inhibition Potency of Lactone and Non-Lactone Compounds for Mouse and Coumarin 7-Hydroxylases (CYP2A5 and CYP2A6). *Xenobiotica* **2000**, *30*, 81–92.
- (55) Rao, S.; Aoyama, R.; Schrag, W.; Trager, W. D.; Rettie, A.; Jones, J. P. A Refined 3-Dimensional QSAR of Cytochrome P450 2C9: Computational Predictions of Drug Interactions. *J. Med. Chem.* **2000**, *43*, 2789–2796.
- (56) Egnell, A.-C.; Houston, J. B.; Boyer, C. S. Predictive Models of CYP3A4 Heteroactivation: In Vitro-in Vivo Scaling and Pharmacophore Mapping. *J. Pharmacol. Exp. Ther.* **2005**, *312*, 926–937.
- (57) Martin, Y. C. Let's Not Forget Tautomers. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 693–704.
- (58) Chemical Abstracts Service, American Chemical Society, Scifinder, 2010. <https://scifinder.cas.org> (accessed Dec 5, 2010).
- (59) *Marvin*, version 5.3.8; ChemAxon, Ltd.: Budapest, Hungary, 2010. <http://www.chemaxon.com>.
- (60) OpenBabel, version 2.1.0. <http://openbabel.org> (accessed Feb 1, 2010).
- (61) Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.* **2008**, *384*, 1002–1017.
- (62) Perola, E. Minimizing False Positives in Kinase Virtual Screens. *Proteins: Struct., Funct., Bioinf.* **2006**, *64*, 422–435.
- (63) Ladbury, J. E.; Klebe, G.; Freire, E. Adding Calorimetric Data to Decision Making in Lead Discovery: A Hot Tip. *Nat. Rev. Drug Discovery* **2010**, *9*, 23–27.
- (64) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The Maximal Affinity of Ligands. *Proc. Natl. Acad. U.S.A.* **1999**, *96*, 9997–10002.
- (65) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand Efficiency: A Useful Metric for Lead System. *Drug Discovery Today* **2004**, *9*, 430–431.
- (66) Abad-Zapatero, C.; Metz, J. T. Ligand Efficiency Indices as Guideposts for Drug Discovery. *Drug Discovery Today* **2005**, *10*, 464–469.
- (67) Hetényi, C.; Maran, U.; García-Sosa, A. T.; Karelson, M. Structure-Based Calculation of Drug Efficiency Indices. *Bioinformatics* **2007**, *23*, 2678–2685.
- (68) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.
- (69) García-Sosa, A. T.; Sild, S.; Maran, U. Docking and Virtual Screening Using Distributed Grid Technology. *QSAR Comb. Sci.* **2009**, *28*, 815–821.
- (70) García-Sosa, A. T.; Hetényi, C.; Maran, U. Drug Efficiency Indices for Improvement of Molecular Docking Scoring Functions. *J. Comput. Chem.* **2010**, *31*, 174–184.
- (71) García-Sosa, A. T.; Maran, U.; Hetényi, C. Calibration of Drug-Likeness beyond Binding Affinity. Submitted for publication.
- (72) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches To Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

- (73) Baum, B.; Muley, L.; Smolinski, M.; Heine, A.; Hangauer, D.; Klebe, G. Non-Additivity of Functional Group Contributions in Protein-Ligand Binding: A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.* **2010**, *397*, 1042–1054.
- (74) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.
- (75) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-Based Lead Discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660–672.
- (76) Leeson, P. D.; Springthorpe, B. The Influence of Drug-like Concepts on Decision-Making in Medicinal Chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881–890.
- (77) Chen, Y.; Shoichet, B. K. Molecular Docking and Ligand Specificity in Fragment-Based Inhibitor Discovery. *Nat. Chem. Biol.* **2009**, *5*, 358–364.
- (78) U.S. National Library of Medicine, National Institutes of Health, Health & Human Services. Daily Med. <http://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?id=6741> (accessed Apr 7, 2010).
- (79) Mole, L.; Israelski, D.; Bupp, J.; Ohanley, P.; Merigan, T.; Blaschke, T. Pharmacokinetics of Zidovudine Alone and in Combination with Oxazepam in the HIV Infected Patient. *J. Acquired Immune Defic. Syndr.* **1993**, *6*, 56–60.
- (80) Taburet, A. M.; Singlas, E. Drug Interactions with Antiviral Drugs. *Clin. Pharmacokinet.* **1996**, *5*, 385–401.
- (81) Sheridan, R. P.; Kearsley, S. K. Why Do We Need So Many Chemical Similarity Search Methods? *Drug Discovery Today* **2002**, *7*, 903–907.
- (82) Xiong, Y.-Z.; Chen, F.-E.; Balzarini, J.; De Clercq, E. Non-Nucleoside HIV-1 Reverse Transcriptase Inhibitors. Part 11: Structural Modulations of Diaryltriazenes with Potent Anti-HIV Activity. *Eur. J. Med. Chem.* **2008**, *43*, 1230–1236.
- (83) Mahajan, D. H.; Pannecouque, C.; De Clercq, E.; Chikhaliya, K. H. Synthesis and Studies of New 2-(Coumarin-4-yloxy)-4,6-(substituted)-S-triazine Derivatives as Potential Anti-HIV Agents. *Arch. Pharm.* **2009**, *342*, 281–290.
- (84) Wang, W. G.; Kim, S. H.; El-Deiry, W. S. Small-Molecule Modulators of p53 Family Signaling and Antitumor Effects in p53-Deficient Human Colon Tumor Xenografts. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11003–11008.
- (85) Hisaki, M.; Imabori, H.; Azuma, M.; Suzutani, T.; Iwakura, F.; Ohta, Y.; Kawanishi, K.; Ichigobara, Y.; Node, M.; Nishide, K.; Yoshida, I.; Ogasawara, M. Synthesis and Anti-Influenza Virus Activity of Novel Pyrimidine Derivatives. *Antiviral Res.* **1999**, *42*, 121–137.
- (86) Kimura, H.; Katoh, T.; Kajimoto, T.; Node, M.; Hisaki, M.; Sugimoto, Y.; Majima, T.; Uehara, Y.; Yamori, T. Modification of Pyrimidine Derivatives from Antiviral Agents to Antitumor Agents. *Anticancer Res.* **2006**, *26*, 91–97.
- (87) Bachmann, K.; Patel, H.; Batayneh, Z.; Slama, J.; White, D.; Posey, J.; Ekins, S.; Gold, D.; Sambucetti, L. PXR and the Regulation of ApoA1 and HDL-Cholesterol in Rodents. *Pharmacol. Res.* **2004**, *50*, 237–246.
- (88) Ekins, S.; Chang, C.; Mani, S.; Krasowski, M. D.; Reschly, E. J.; Iyer, M.; Kholodovych, V.; Ai, N.; Welsh, W. J.; Sinz, M.; Swaan, P. W.; Patel, R.; Bachmann, K. Human Pregnane X Receptor Antagonists and Agonists Define Molecular Requirements for Different Binding Sites. *Mol. Pharmacol.* **2007**, *72*, 592–603.
- (89) Ekins, S.; Erickson, J. A. A Pharmacophore for Human Pregnane-X-Receptor Ligands. *Drug Metab. Dispos.* **2002**, *30*, 96–99.
- (90) Schuster, D.; Langer, T. The Identification of Ligand Features Essential for PXR Activation by Pharmacophore Modeling. *J. Chem. Inf. Model.* **2005**, *45*, 431–439.
- (91) Ekins, S.; Kortagere, S.; Iyer, M.; Reschly, E. J.; Lill, M. A.; Redinbo, M. R.; Krasowski, M. D. Challenges Predicting Ligand-Receptor Interactions of Promiscuous Proteins: The Nuclear Receptor PXR. *PLoS Comput. Biol.* **2009**, *5*, e1000594.
- (92) Rahnasto, M.; Wittkindt, C.; Juvonen, R. O.; Turpeinen, M.; Petsalo, A.; Pelkonen, O.; Poso, A.; Stahl, G.; Holtje, H. D.; Raunio, H. Identification of Inhibitors of the Nicotine Metabolising CYP2A6 Enzyme—An *In Silico* Approach. *Pharmacogenomics J.* **2008**, *8*, 328–338.
- (93) Jones, B. C.; Hawksworth, G.; Horne, V. A.; Newlands, A.; Morsman, J.; Tute, M. S.; Smith, D. A. Putative Active Site Template Model for Cytochrome P450C9 (Tolbutamide Hydroxylase). *Drug Metab. Dispos.* **1996**, *24*, 260–266.
- (94) Mancy, A.; Broto, P.; Dijols, S.; Dansette, P. M.; Mansuy, D. The Substrate Binding Site of Human Liver Cytochrome P450 2C9: An Approach Using Designed Tienilic Acid Derivatives and Molecular Modeling. *Biochemistry* **1995**, *34*, 10365–10375.
- (95) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three- and Four-Dimensional-Quantitative Structure Activity Relationship (3D/4D-QSAR) Analyses of CYP2C9 Inhibitors. *Drug Metab. Dispos.* **2000**, *28*, 873–878.
- (96) Shou, M.; Grogan, J.; Mancewicz, J. A.; Krausz, K. W.; Gonzalez, F. J.; Gelnoin, H. V.; Korzekwa, K. R. Activation of CYP2A4: Evidence for the Simultaneous Binding of Two Substrates in a Cytochrome P450 Active Site. *Biochemistry* **1994**, *33*, 6450–6455.
- (97) Korzekwa, K. R.; Krishnamary, N.; Shou, M.; Ogai, A.; Parise, R. A.; Rettie, A. E.; Gonzalez, F. J.; Tracy, T. S. Evaluation of Atypical Cytochrome P450 Kinetics with Two-Substrate Models: Evidence That Multiple Substrates Can Simultaneously Bind to Cytochrome P450 Active Sites. *Biochemistry* **1998**, *37*, 4137–4147.
- (98) Domanski, T. L.; He, Y. A.; Khan, K. K.; Roussel, F.; Wang, Q.; Halpert, J. R. Phenylalanine and Tryptophan Scanning Mutagenesis of CYP3A4 Substrate Recognition Site Residues and Effect on Substrate Oxidation and Cooperativity. *Biochemistry* **2001**, *40*, 10150–10160.
- (99) Vaz, R. J.; Zamora, I.; Li, Y.; Reiling, S.; Shen, J.; Cruciani, G. The Challenges of *In Silico* Contributions to Drug Metabolism in Lead Optimization. *Expert Opin. Drug Metab. Toxicol.* **2010**, *7*, 851–861.
- (100) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-Dimensional-Quantitative Structure Activity Relationship Analysis of Cytochrome P-450 3A4 Substrates. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 424–433.
- (101) Mestres, J.; Martín-Couce, L.; Gregori-Puigjané, E.; Cases, M.; Boyer, S. Ligand-Based Approach to *In Silico* Pharmacology: Nuclear Receptor Profiling. *J. Chem. Inf. Model.* **2006**, *46*, 2725–2736.
- (102) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting Drug Metabolism: A Site of Metabolism Prediction Tool Applied to the Cytochrome P450 2c9. *J. Med. Chem.* **2003**, *46*, 2313–2324.
- (103) Rydberg, P.; Gloriam, D. E.; Olsen, L. The SMARTCyp Cytochrome P450 Metabolism Prediction Server. *Bioinformatics* **2010**, *26*, 2988–2989.
- (104) Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J. Probing the Links between *In Vitro* Potency, ADMET and Physicochemical Properties. *Nat. Rev. Drug Discovery* **2011**, *10*, 197–208.