JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

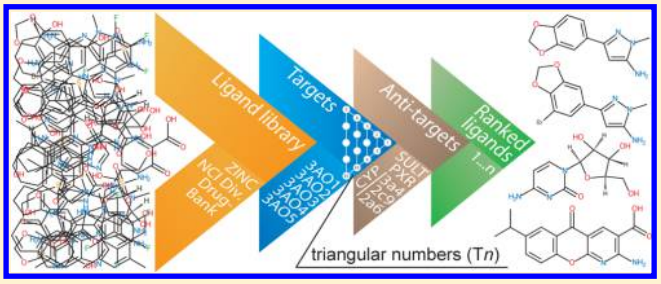
Improving the Use of Ranking in Virtual Screening against HIV-1 Integrase with Triangular Numbers and Including Ligand Profiling with Antitargets

Alfonso T. García-Sosa*,† and Uko Maran†

†Institute of Chemistry, University of Tartu, Ravila14a, Tartu 50411, Estonia

Supporting Information

ABSTRACT: A delicate balance exists between a drug molecule's toxicity and its activity. Indeed, efficacy, toxicity, and side effect problems are a common cause for the termination of drug candidate compounds and development projects. To address this, an antitarget interaction profile is built and combined with virtual screening and cross docking for new inhibitors of HIV-1 integrase, in order to consider possible off-target interactions as early as possible in a drug or hit discovery program. New ranking techniques using triangular numbers improve ranking information on the compounds and recovery of known inhibitors into the top compounds using different docking programs. This improved ranking arises from using consensus of ranks between docking programs and ligand efficiencies to derive a new rank, instead of using absolute score values, or average of ranks. The triangular number rerank also allowed the objective combination of results from several protein targets or screen conditions and several programs. Triangular number reranking conserves more information than other reranking methods such as average of scores or averages of ranks. In addition, the use of triangular numbers for reranking makes possible the use of thresholds with a justified leeway based on the number of available known inhibitors, so that the majority of the compounds above the threshold in ranks compare to the compounds that have known experimentally determined biological activity. The battery of anti- or off-targets can be tailored to specific molecular or drug design challenges. *In silico* filters can thus be deployed in successive stages, for prefiltering, activity profiling, and for further analysis and triaging of libraries of compounds.



INTRODUCTION

Acquired immune deficiency syndrome (AIDS) is a pandemic caused by the human immunodeficiency virus (HIV) that affects around 33 million people worldwide.[1] It infects 2.7 million new people, results in 3 million deaths each year, and causes severe strains on the lives, productivity, and health systems of many countries.[1] HIV-1 integrase is an interesting target against the disease since it is an essential enzyme in the mechanism of infection of HIV-1 and does not have an equivalent enzyme in the human organism, thus avoiding selectivity issues. HIV-1 integrase and reverse transcriptase are both contained in the viral RNA capsid. Once inside the host cell, viral RNA is converted into proviral DNA, which is then incorporated into the host cell's own DNA in a process aided by HIV-1 integrase.[2] The only other self-contained viral enzyme is HIV-1 protease.

Raltegravir (Isentress)[3] is the first clinically approved HIV-1 integrase inhibitor (F.D.A. Oct. 2007) and is used in patients with resistance to HIV-1 protease inhibitors. There have also been computational studies[4] on the crystal structure derived from the complex of 5CITEP with HIV-1 integrase.[5] New high resolution X-ray crystal structure determinations of HIV-1 integrase with a series of inhibitors have been disclosed.[6] These structures show the inhibitors binding tightly to the protein in an allosteric binding site, different to the conventionally used site that contains metal ions. HIV-1 integrase inhibitors show promising potential,[3] but given that virus proteins can mutate in response to a drug, there is a continuous need for new inhibitors. In particular, a combination therapy including a cocktail of inhibitors against all three virally contained enzymes (HIV-1 protease, reverse transcriptase, and integrase) in addition to an entry inhibitor, may provide a better treatment profile and outcome as well as less resistance issues.

Despite increased research and development in pharmaceutical companies, a bottleneck in the development of new drugs has appeared where drug candidate compounds fail during the clinical testing stages due to efficacy, toxicity, or side effect problems. Some of these toxic and side effects may be produced by antitarget or off-target interactions, i.e. the binding of the candidate molecule to a protein or nucleic acid which is not the target intended to cause the pharmacological effect. Safer drug design or ligand toxicity and/or side effect profiling through the discovery and development of compounds that do not bind to antitargets can identify and flag compounds early in the drug discovery and development stages. This may aid in discovering

Received: March 18, 2014
Published: October 10, 2014

compounds with less toxic or side effects, or honing selectivity to new targets, by virtue of exploring their antitarget or off-target interactions. On the other hand, antitargets may be useful for designing compounds with less metabolizing interactions, either to avoid a toxic metabolite or to increase their half-life in the bloodstream. Equally possible would be to design compounds that are better metabolized or excreted.

Some of the most important antitargets are the enzymes responsible for metabolizing foreign substances, such as cytochromes P450 (CYP P450s), which are part of phase I metabolism through oxidations, one of the first modifications of a substance in the blood. They tend to produce more polar and easier to excrete metabolites than the original compound.[7] In addition, various enzymes are involved in phase II metabolism where hydrophilic moieties are conjugated to the metabolite, such as glucuronic acid, or sulfonyl groups onto amino and hydroxyl groups to form sulfamates and sulfates (sulfotransferases - SULTs). Another important group of proteins are involved in generating resistance to certain medications by effluxing drugs, the drug transporters of phase III metabolism. The pregnane X receptor (PXR) is involved in recognizing foreign substances and up-regulating enzymes (including drug transporters, SULTs, and CYP P450s) to transform or excrete the compound.[8] Structural analysis of PXR presents a largely hydrophobic binding site that binds to a wide variety of ligands of diverse nature and that does not present large conformational changes between *apo* and *holo* forms of the receptor.[9] This relative rigidity is well-suited to virtual screening and docking since a single protein structure can capture the interactions between proteins and ligands in the binding site. Many of these enzymes, receptors, and transporters are mainly located in the liver and small intestine, common locations for side effects. Endogenous as well as exogenous substances can trigger the activation of any of the above-mentioned proteins including PXR. The study of the antitarget interactions of ligands (that is, their side effect and toxicity profiling) may help discover similar side effects between drugs[10] and drug candidates. The use of multiple off-targets may also find new targets for known drugs[11] and for other compounds in development if those interactions are detected.

Efforts have been carried out to predict binders to CYP P450s[8] (see review in ref 7), including using 2D and 3D-similarity searches.[12] Docking against a battery of antitargets can provide a more detailed representation of interactions and modify screening strategies. In the present work, *in silico* screening experiments were conducted on several HIV-1 integrase X-ray crystal structures, improving the results of the scoring and ranking of molecules by using triangular numbers and ligand efficiency calculations in search of novel inhibitors. In addition, a ligand toxicity and side effect profiling technique was employed, where the results of the predicted inhibition of one SULT, one PXR, and three CYP p450s antitargets are used to identify ligands with the best safety profile. The design of new ligands to these important targets that was carried out includes consideration toward the safety of the compounds designed from this early stage. This gives an advantage since many compounds may be good inhibitors but then fail at safety tests further along the drug design process if safety objectives are not included in the early stages.

METHODS

Targets. For HIV-1 integrase, five high-resolution X-ray crystal structures of the protein with inhibitors in a newly described allosteric site were obtained from the Protein DataBank:[13] 3AO1, 3AO2, 3AO3, 3AO4, and 3AO5. Hydrogen atoms were placed on protein structures using Maestro.[14] Protein structures were modified by using the Protein preparation wizard[15] which assigns polar hydrogens, ionic states of residues in the protein, as well as conducting partial minimizations while holding the protein backbone restrained. Water molecules were retained in the HIV-1 integrase structure, given that they can be an integral part of the binding site and provide as strong binding as nonhydrated binding sites.[16]

Small Molecule Library. A collection of ligands was built by joining the 2012 version of the ZINC database (a compilation of commercially available compounds)[17] with the National Cancer Institute's (NCI) diverse set of ligands,[18] as well as with the structures of known and in-use small-molecule drugs obtained from the DrugBank database.[19] Chemical compound libraries were pretreated by prescreening virtually the collection of ligands and removing from the ZINC database molecules with predicted unwanted, toxic, and reactive groups, such as epoxides, hydrazines, etc., as well as those molecules with undesirable predicted solubility, high number of aromatic rings, and high number of rotatable bonds. This resulted in a collection of around 70,000 compounds, fully prepared with hydrogens assigned and finally, in an ionic state corresponding to a pH of 7,[20] which were then employed for docking against all the protein structures.

Spiked Compounds. A set of known inhibitors was used to spike the data set, in order to judge which methods, ranks, and structures were best suited to recover them among the top ranked compounds after screening. These six compounds were as follows: 1,3-benzodioxol-5-ol (BZX),[6] 3-(7-bromo-1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazol-5-amine (AVX),[6] 3-(1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazole-5-carboxylic acid (BMC),[6] 3-(1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazol-5-amine (833), 5-(7-bromo-1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazol-3-amine (BBY),[6] and 1-methyl-3-(thiophen-2-yl)-1H-pyrazol-5-amine (MPV_Wielens3).[6]

Consensus Screening. A procedure was developed where the dock scores between the ligand molecules and the target protein are calculated with three independent docking scoring programs (each with their own independent scoring functions and docking engines, Glide v5.0,[21] Autodock v4.0,[22] and Vina v1.1.2[23]), and only those compounds which were highly scored by consensus between the programs and scored higher or comparable to known inhibitors were selected and their dock score was combined into the mean of their values (ΔG). In addition, we calculated the dock score of the ligands with a battery of antitargets (five antitarget proteins) also using Glide, Autodock, and Vina, but now detailing the interaction between each ligand and antitarget determined by each program and also providing a new additive measure, the grand total of all the antitarget interactions. The consensus required between docking programs, scoring functions, and congruence with known inhibitor binding profiles ensures more confidence in the screening results but, at the same time, also details any possible interaction between ligands and antitargets. The use of the logical "AND" between docking programs for target ligand binding would in principle reduce the number of false positives produced, a common grievance with single methods. Also, the application of the logical "OR" among the different docking programs in binding to antitargets can be more sensitive and in principle reduce the number of false negatives, which would be

more dangerous than mispredicting a false positive for side- and off-target interactions.

Molecular Docking. Using three different docking programs and their docking approaches, results can be hedged in order to eliminate possible extremes in one docking method. The following paragraphs detail the conditions that were used for Glide, Autodock, and Vina.

For the docking calculations with Glide, a hierarchical scheme was applied using the HTVS scoring function to make a first selection based on flexible docking retaining 10% of the top compounds (1 pose per compound). This was followed by the SP scoring function to make the next selection based on flexible docking retaining 10% of the top compounds (1 pose per compound), and finally the XP scoring function was used to dock flexibly and rank all top compounds, retaining 1 pose per compound. For all stages, amide bond conformations were allowed to be varied, penalizing nonplanar solutions. A scaling factor of 0.8 was used on the protein atom size (van der Waals radii) for those atoms that have partial atomic charge smaller than 0.15, in order to account for a small degree of conformational variation or adaptation of the protein to the ligand when binding. The OPLS2001 force field was used for postdocking minimizations. The scoring procedure allows penalizing ligand binding poses where polar groups do not have a binding partner in the protein binding site, since these groups would reduce the ligand's binding energy through a high desolvation energy.

The settings used for the genetic algorithm in Autodock were as follows: 250 individuals in a population; 20,000,000 maximum energy evaluations; 27,000 maximum generations; one individual surviving into next generation; 100 genetic algorithm docking runs; and a ranked cluster analysis was performed on each docking calculation (100 runs of each ligand against each protein). The Autodock Binding Energy was used for scoring. These settings, though fairly intensive and more stringent than the default values, have provided good results in reproducing strong scores for known actives in previous studies.[24]

The settings used for the iterated local search global optimizer based on mutation and local optimization steps, accepted or rejected with a Metropolis criterion in Vina were nine modes, one central processing unit, and energy range of 1 kcal/mol.

Ligand chemical structures were visually inspected for the top-ranked compounds and judged based on their consistency between programs, their structural and chemical soundness, as well as their interactions with key protein binding partners: 1) checking that the binding mode is plausible, 2) that the ligand is in a correct site, 3) that the ligand is not protruding significantly from the binding site, 4) that the predicted binding pose has the identified privileged interactions and interaction partners, 5) proper consideration of tautomers,[25] and 6) possible ionization and hydration states[16] of ligand and protein, among other checks, such as hydrogen bonds with protein backbone atoms which are less prone to conformational and mutational variation. These hydrogen bonds can help identify those compounds which would be less sensitive to mutations which could render the protein resistant to a drug. From the docking results, structures with undesired qualities were removed.[26] The top compound structures were filtered in order to obtain compounds that were predicted to inhibit strongly HIV-1 integrase and by consensus between all docking programs.

Antitarget Battery and Rules. The battery of antitargets contained the pregnane X receptor, the sulfotransferase 1A3, and three cytochrome P450s (see Table 1). For the pregnane X receptor, structure 1M13 was used which contains the ligand hyperforin that is known to bind and activate this receptor. For sulfotransferase 1A3, cytochrome P450 2a6, and cytochrome P450 2c9, their structures contained the known active ligands L-dopamine, coumarin, and *S*-warfarin, respectively.

Table 1. Antitarget Protein Crystal Structures

name and abbreviation	PDB ID	resolution (Å)	ligand or inhibitor in complex
pregnane X receptor (PXR)	1M13	2.15	hyperforin
sulfotransferase 1A3 (SULT)	2A3R	2.60	L-dopamine
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	

For all of these antitargets except CYP P450 3a4, the screened ligand molecule would receive a value of 0 if it had a dock score lower than 0.5 kcal/mol from the dock score recorded by the cocomplexed ligand, a value of 1 if its dock score was more than 0.5 kcal/mol, and a value of 0.5 if it was within 0.5 kcal/mol. This system was introduced in order to provide a very rough approximate of equilibrium concentration values existent between bound and unbound states of protein−ligand complexes defined by their predicted interaction energies and also by the need to account for uncertainty and include a less sharp threshold than a 0/1 case based on an approximate value of binding energy. It is thus represented as

$$\text{Score} = 0.0 \text{ if } \Delta G - \Delta G\text{ref} > 0.5$$

$$\text{Score} = 0.5 \text{ if } |\Delta G - \Delta G\text{ref}| \leq 0.5$$

$$\text{Score} = 1.0 \text{ if } \Delta G - \Delta G\text{ref} < -0.5$$

where ΔGref is the dock score of the protein with cocrystallized active reference ligand, and ΔG is the dock score for a ligand bound to that protein binding site. For CYP P450 3a4, since there was no cocrystallized ligand present in the crystal structure, a value of −7.5 kcal/mol was heuristically adopted as a reference as described previously.[27] In addition, an extra series of compounds that are also known actives against the antitargets were docked to see if they would be recovered among the top-ranked in the docking procedures. These are A-792611,[28] ChapmanE1,[29] methoxsalen, sulfaphenazole, and nephazodone, for inhibiting PXR, SULT, CYP 2a6 CYP 2c9, and CYP 3a4, respectively.

Physicochemical Descriptors. The molecular weight (MW), molecular formula, hydrogen bond acceptor atom count (acceptor count), hydrogen bond donor atom count (donor count), molecular surface area (MSA), polar surface area (PSA), apolar surface area (apolarSA), number of carbons (NoC), number of heavy atoms (NHA), molecular polarizability, Balaban index, Randic index, Szeged index, Platt index, Harary index, aromatic atoms, aromatic rings, aliphatic atoms, aliphatic rings, number of atoms (atom count), number of hydrogens, sp3 carbons, bond count, rotatable bond count, Wiener index (W), hyper Wiener index (hyperWiener), and Wiener polarity were calculated with Marvin Beans.[30] The

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

logarithm of the partition coefficient between octanol and water (log*P*) of compounds was calculated with XlogP v. 3.0.[31] Some of these properties provided values for later use in the calculation of ligand efficiency indices. Tanimoto similarity coefficients were calculated using OpenBabel.[32]

Ligand Efficiency. Ligand efficiency indices, also called binding efficiency indices, have been recently introduced to normalize the free energy of binding of a ligand per unit of measure.[33−37] They have been successfully applied to improve the correlation between calculated and experimental values of drug−protein binding efficiencies,[37] as well as separating drugs from nondrug compounds.[38,39] They have also found value in being able to determine small molecule compounds that are able to disrupt (and therefore inhibit) large surface protein−protein interactions.[36] The ligand efficiency indices of ΔG/PSA (Table S1 in the Supporting Information), ΔG/MW (Table S1), ΔG/NHA (Table S1), ΔG/NoC[37] (Table 5), ΔG/W (Table S1), $\Delta G/P$[37] (Table S1), ΔG/MSA (Table 5), and ΔG/PSA[27] (Table S1) were measured for all top-scoring ligands, where ΔG is the dock score for a ligand-protein pair.

RESULTS AND DISCUSSION

Screening. The six known inhibitors were recovered among the top 2% of the screened ligand database, which can be seen as a measure of reliability and quality in the screen. The superposition of the five target protein crystal structures is shown in Figure 1, where the all-atom root-mean-square deviations compared to 3AO1 are 0.432 Å (3AO2), 0.452 Å (3AO3), 0.492 Å (3AO4), and 0.455 Å (3AO5). The general similarity between the allosteric binding sites and the same region occupied by their cocrystallized ligands can be seen. However, small differences in the binding site structure can confer different ligand screening results.

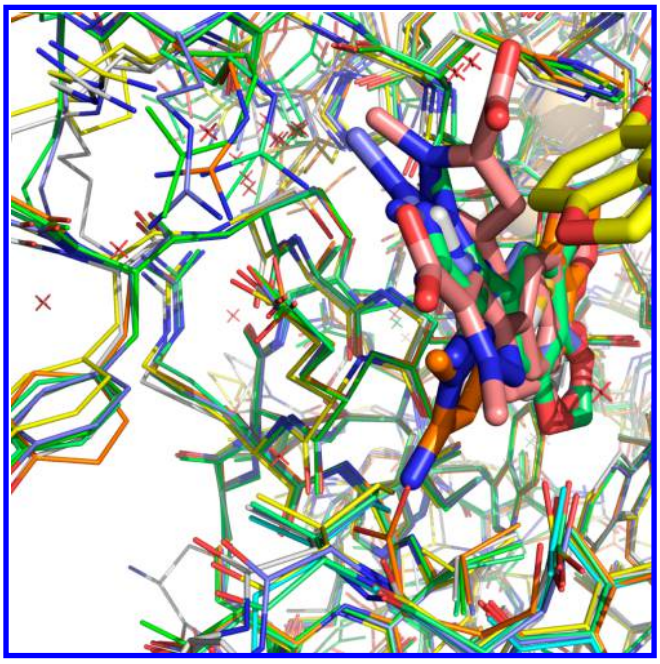


Figure 1. Superposition of the allosteric binding sites of 3AO1 (backbone in yellow), 3AO2 (backbone in slate), 3AO3 (backbone in salmon), 3AO4 (backbone in gold/orange), and 3AO5 (backbone in green).

To discern the best protein crystal structure (among five), scoring and docking procedure (among three), physicochemical property, ligand efficiency indices, as well as ranking system, the ranks of the known inhibitors were compared among all run cases. The best procedure would recall the highest number of known inhibitors among the top ranks, and the number of these in the top 15 ranked compounds was examined. The results are shown in Table 2.

The number of known inhibitors recalled in the top 15 ranked compounds (recall) show that the best protein crystal structures for inhibitor recall were 3AO4 for Glide XP and Autodock 4 and 3AO2 for Vina. Also observable, even simple measures such as bond and atom counts could recall the known inhibitors; however, as expected, such simple measures provided many ties in the top ranked compounds and are not very discriminate among compounds. Dock score by itself was not among the best recalling procedures, but ligand efficiencies provided better recall power. Among these, the best recall was provided by ΔG/MSA and ΔG/NoC.

Ranking with Triangular Numbers. To investigate a possible reranking measure or number, the best protein cases for each docking algorithm and scoring procedure were used, that is, 3AO4 for Glide and Autodock and 3AO2 for Vina. Given that the absolute score values provided by the docking programs are variable, a different way of obtaining a consensus between the ranks of the compounds according to each docking program, ligand efficiency, and physicochemical parameter was developed. Thus, the sum of ranks (SOR) of all the known inhibitors was computed for a number of physicochemical properties and ligand efficiencies (see Table 3). The ideal or perfect value for SOR would be 1 + 2 + 3 + 4 + 5 + 6, i.e. 21, for the case when the known inhibitors are the six top ranked compounds. This series of numbers is known as triangular or triangle numbers (Tn) and represent a divergent number series with applications in a variety of fields, conceptually analogous to factorial numbers, but the result is the sequential summation of n with $n - 1$, instead of multiplication as in factorial, and the final result being always $n(n+1)/2$. Figure 2 shows a graphical representation of the triangular number sequence progression.

From Figure 2, it can be seen how as the rank number of a compound rises, the triangular number also increases, not in a linear fashion, but according to the triangular number sequence, so that for a compound ranked in place three, the sum of all the ranks of compounds until that point will be six. This results in a numerical measure that allows the use of the number of known inhibitors used as part of the controls in a virtual screen to give a threshold with which to objectively judge the derived scored ligands. For example, if among the top ranked ligands the majority of the highest scoring compounds are known inhibitors, such as if the six known inhibitors are included within the top 11 compounds, then a ligand ranked as high as T(11) = 66 or higher, would have a reasonable predicted chance to have binding to the target in question, since the known inhibitors show that docking profile. It also allows to define a buffer, i.e., a number of compounds of experimentally unknown activity but strong docking rank can be considered if they are very close to the known inhibitors' ranks. This triangular number reranking also allows the combination of several ranks from the several screening runs. This is needed in order not to use other measures of combining docking results, such as the average of the scores or of the ranks, which may be misleading if one or two results strongly skew the average in

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

Table 2. Number of Six Known Inhibitors Recalled in the Top Ranked 15 Compounds According to Five Crystal Structures, Three Docking Programs, Their Consensus, Different Ligand Efficiency Indices, and Ligand Properties[a]

property	3AO1	**3AO2**	3AO3	**3AO4**	3AO5	consensus of all structures and all methods
ΔG	0; 0; 0	2; 0; 0	0; 0; 0	3; 2; 0	0; 0; 0	0.5
ligand efficiency index:						
ΔG/MW	2; 2; 2	2; 1; 3	3; 2; 3	4; 4; 2	2; 3; 2	2.5
ΔG/Wiener	2; 3; 2	4; 2; 4	4; 3; 4	5; 4; 3	2; 4; 4	3.3
ΔG/MSA	3; 4; 3	4; 2; 5	4; 4; 5	0; 4; 5	3; 5; 5	3.7
ΔG/PSA	0; 0; 0	0; 0; 0	0; 0; 0	4; 0; 0	0; 1; 0	0.3
ΔG/NHA	2; 3; 2	4; 1; 5	3; 3; 5	5; 4; 4	2; 4; 4	3.4
ΔG/NoC	4; 4; 4	4; 4; 5	4; 4; 5	3; 4; 5	3; 5; 5	4.2
physicochemical property:						
MW	2	2	2	2	2	2
Wiener	2	2	2	2	2	2
MSA	5	5	5	5	5	5
PSA	0	0	0	0	0	0
apolarSA	4	4	4	4	4	4
donor count	1	1	1	1	1	1
acceptor count	0	0	0	0	0	0
rotatable bonds	4	4	4	4	4	4
atom count	5	5	5	5	5	5
number of hydrogens	5	5	5	5	5	5
NHA	4	4	4	4	4	4
NoC	5	5	5	5	5	5
sp3 carbons	5	5	5	5	5	5
fraction sp3 carbons	5	5	5	5	5	5
molecular polarizability	3	3	3	3	3	3
aliphatic rings	1	1	1	1	1	1
aromatic rings	5	5	5	5	5	5
aromatic atoms	5	5	5	5	5	5
Balaban	0	0	0	0	0	0
Harary	2	2	2	2	2	2
Bond count	5	5	5	5	5	5
hyperWiener	2	2	2	2	2	2
Platt	5	5	5	5	5	5
Randic	5	5	5	5	5	5
rings	2	2	2	2	2	2
Szeged	4	4	4	4	4	4
Wiener polarity	4	4	4	4	4	4

[a]In order: Glide XP; Autodock 4; Vina.

any direction. However, the rerank still makes use of consensus between good or bad results from different runs or screens.

The logarithm was also calculated in order to have numbers in the same order of magnitude: log(SOR), where the ideal (perfect) value of six out of six known inhibitors in the first six ranks would amount to log(Tn), i.e. 1.32, while a good value could be given by the logarithm of T(11), i.e. 1.82. A further measure is to subtract the value for the ideal recall of the known six inhibitors from the log(SOR), to give log(SOR) − log(Tn), where n = 6, and the better ranking would be provided by values as close to zero as possible. A rerank using factorial numbers: log(SOR) − log(n!) was also tried (results not shown), but triangular numbers gave better results. Table 3 shows these reranks. The median of ranks of the known inhibitors is also presented, where the best ranking would also show the lowest number. Finally, the number of known inhibitors in the top ranked 15 compounds is also shown for comparison.

Table 3 presents the performance of different physicochemical properties and ligand efficiencies for the best protein structure case for each docking program. In cases where the performed measure is better than the good thresholds, then the value is marked in bold. Some values show better than ideal values, which are the effect of ties among high ranks, but are considered as good as the ideal cases. Tables 2 and 3 show that some ligand efficiencies perform well, such as the commonly used ΔG/NHA. However, the best in this case seem to be ΔG/NoC, ΔG/MSA, and ΔG/Wiener. These are interesting new ligand efficiencies, that are best suited for the protein system and docking procedures in this case. The results show that ΔG/MW and ΔG/NHA do not have to be always the best suited for every protein−ligand system and that other ligand efficiencies can find utility for docking and virtual screening for HIV-1 integrase.

If a common technique is used, such as the consensus or average of the dock scores of the six known inhibitors, then a good retrieval of the known inhibitors cannot be achieved. The average dock scores for all six known inhibitors would occupy low places in the ranking of the compounds. For Glide XP it would correspond to a rank of 148, for Autodock 4 to a rank of 100, for Vina to a rank of 243, and for the consensus of all docking programs, the average of their dock scores would rank

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

Table 3. Reranking of Docked Ligands and Improvement of Early Retrieval of Known Positives

Glide 3AO4; AutoDock4 3AO4; Vina 3AO2	sum of ranks (SOR) good value < 67 ideal = T*n* = 21	Log(SOR) good is < 1.82 ideal = 1.32	Log(SOR) − log(T*n*) ideal ≤ 0 good value < 0 .5	median of ranks of known inhibitors good < 35	recall of known inhibitors in top 15 compds ideal = 6 good ≥ 4
Δ*G*	700; 533; 1182	2.85; 2.73; 3.07	1.52; 1.40; 1.75	66.5; 76; 195.5	0; 2; 0
Δ*G*/MW	103; 162; 199	2.01; 2.21; 2.3	0.69; 0.89; 0.98	**12.5; 9.5; 10.5**	3; 3; 3
Δ*G*/Wiener	**63**; 67; **58**	**1.80**; 1.83; **1.76**	**0.48**; 0.5; **0.44**	**12; 9; 11.4**	**4; 4; 6**
Δ*G*/MSA	**30**; 70; **39**	1.48; 1.85; 1.59	**0.15**; 0.52; **0.27**	**5; 3.5;** 7	**5; 4; 5**
Δ*G*/PSA	677; 577; 830	2.83; 2.76; 2.92	1.51; 1.44; 1.6	93.5; 83; 144.5	0; 0; 0
Δ*G*/NHA	**60**; 153; **50**	**1.78**; 2.18; **1.7**	**0.46**; 0.86; **0.38**	**9; 3.5; 9.5**	**4; 4; 5**
Δ*G*/NoC	**35**; 98; **23**	**1.54**; 1.99; **1.36**	**0.22**; 0.67; **0.04**	**5; 3.5; 3.5**	**5; 4; 5**
MW	160	2.20	0.88	**13**	2
Wiener	71	1.85	0.53	**15**	2
MSA	**47**	**1.67**	**0.35**	**8**	**5**
PSA	688	2.84	1.52	**132**	0
apolarSA	**42**	**1.62**	**0.30**	**4.5**	**4**
donor count	192	2.28	0.96	**20**	1
acceptor count	495	2.69	1.37	90	0
rotatable bonds	**48**	**1.68**	**0.36**	**7**	**4**
atom count	**24**	**1.38**	**0.06**	**4**	**5**
number of hydrogens	**20**	**1.30**	**-0.02**	**3.5**	**5**
NHA	**60**	**1.78**	**0.46**	**11**	**4**
NoC	**28**	**1.45**	**0.12**	**4**	**5**
sp3 carbons	**14**	**1.15**	**-0.18**	**3**	**5**
fraction sp3 carbons	**18**	**1.26**	**-0.07**	**3.5**	**5**
molecular polarizability	**67**	1.83	0.50	**11.5**	3
aliphatic rings	207	2.32	0.99	**17**	1
aromatic rings	**28**	**1.45**	**0.12**	**1**	**5**
aromatic atoms	**33**	**1.52**	**0.20**	**3.5**	**5**
Balaban	731	2.86	1.54	110.5	0
Harary	72	1.86	0.54	**14**	2
bond count	**32**	**1.51**	**0.18**	**6**	**5**
hyperWiener	74	1.87	0.55	**15.5**	2
Platt	30	**1.48**	**0.15**	**5**	**5**
Randic	**29**	**1.46**	**0.14**	**5.5**	**5**
rings	562	2.75	1.43	133	2
Szeged	**65**	**1.81**	**0.49**	**12**	**4**
Wiener polarity	**58**	**1.76**	**0.44**	**10**	**4**

in place 118. Clearly, better methods of reranking are needed. Given the data known and published on these inhibitors, new reranks such as those based on triangular numbers can improve the reranking.

The consensus (average) of ranks of all docking programs and all proteins gives 3.7 and 4.2 as the best retrievers of known inhibitors for Δ*G*/MSA and Δ*G*/NoC, as shown in Table 2, which are also the best retrievers according to the triangular numbers, but the triangular numbers can give an indication with which protein and with which docking procedure and with which ranking score the best results were achieved. The average values mask this information.

The average of ranks for the six known inhibitors, for five proteins gives rank binding = 118; rank Δ*G*/MW = 12.8; rank Δ*G*/Wiener = 9.5; rank Δ*G*/MSA = 4.5; rank Δ*G*/PSA = 100.8; rank Δ*G*/NHA = 7.5; rank Δ*G*/NoC = 4.7. Again, Δ*G*/MSA and Δ*G*/NoC are the best reranking ligand efficiencies, but the average values hide all knowledge of performance. That is, there is information loss when using only an average value based on average of ranks or average of scores. Reranking with triangular numbers allows selecting the best protein case, as well as the best method, procedure, and ligand efficiency for reranking compounds and finding the case with highest amount of known inhibitors in the top ranked compounds.

In addition, the use of triangular numbers for reranking makes possible the use of thresholds with a justified leeway based on the number of available known inhibitors, so that the majority of the compounds above the threshold in ranks compare to the compounds that have known experimentally determined biological activity. In this particular case, with six known inhibitors, taking 11 compounds means the majority of these compounds would be active, and hence the triangular number for 11 is 66. Compounds ranked higher than this threshold would have a good possibility of being a strong binder to the target, since they are reranked as high as the known active compounds.

Receiver-operator curves were computed for each of the docking program/scoring function-derived efficiency indices Δ*G*/MW, Δ*G*/NHA, Δ*G*/NoC, and Δ*G*/MSA (see Figures S1−S3 in the Supporting Information). The worst performance was observed in each case to be the simple Δ*G* provided by the simple score. Δ*G*/MW and the widely used Δ*G*/NHA provided improvement in some cases, but the best performance for early recovery of known inhibitors among the top ranked compounds was provided by the relatively simple Δ*G*/NoC

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

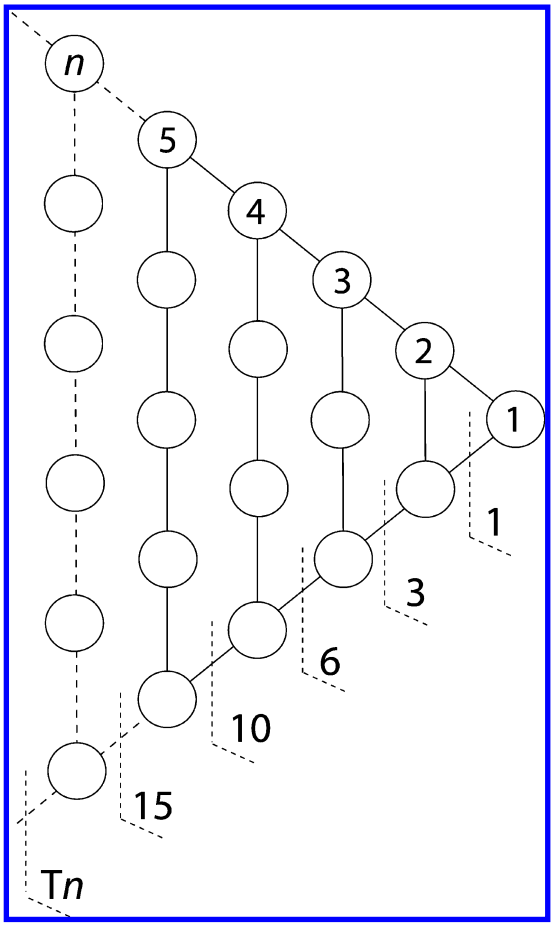


Figure 2. Triangular number (T*n*) sequence and increase in size according to *n*.

and ΔG/MSA. The areas under the curves (AUC) of the receiver-operator curves (ROC)[40−43] (shown in Table 4) for all these cases show numerically these improvements.

Table 4. Areas under the Curve (AUCs) for Receiver-Operator Curves (ROCs) for Ranking and Docking Procedures

	ΔG	ΔG/MW	ΔG/NHA	ΔG/NoC	ΔG/MSA
Glide XP on 3AO4	0.437	0.947	0.975	0.991	0.994
Autodock4 on 3AO4	0.538	0.876	0.884	0.932	0.957
Vina on 3AO2	0.268	0.891	0.982	0.989	0.998

As was the case with the triangular number reranking, the AUC values demonstrate that ΔG/MSA and ΔG/NoC are best suited for recalling known inhibitors in the present screening case and that the T*n* reranks are valid as shown by the widely used measure of AUC to evaluate screening and docking procedures by recalling known inhibitors early.

Reranked Ligands. A selection of the most promising compounds for inhibiting HIV-1 integrase is shown in Table 5.

Table 5 shows multiple information, such as the MW and log*P* of the compound, as well as the binding profile of each compound against the protein target as measured by the ligand efficiencies ΔG/MSA and ΔG/NoC with each calculation being verified by consensus (average) of the three docking programs. The ligand efficiency measure ΔG/MSA combines the pharmacokinetic measure of permeability or distribution estimated by the molecular surface area with the pharmacodynamic measure of binding affinity. Remarkably, the top-ranked compounds have very strong ΔG/MSA and ΔG/NoC values, due to their favorable binding and small size. Values of ΔG/NoC around −0.5 kcal·NoC/mol were identified among known drugs,[37] so compounds in Table 5 would have a high binding capability concentrated in a small size given strongly negative values of ΔG/NoC. Also, this ligand efficiency index is straightforward to calculate and easily understood, given that the number of carbons in a compound is directly related to the compound's size and hydrophobicity.

The top ranked compounds have some similarities to the known inhibitors. They are small compounds, with low molecular weight, and, in addition, they all contain a small aromatic group in close vicinity to a small polar group. However, the top-ranked ligands have high chemical diversity and are not structurally related to the known inhibitors. As such, they are valuable new chemical motifs for exploring their activities and those of series of their derivatives. All top ranked ligands had Tanimoto similarity coefficients (Tc) lower than 0.21 with the known inhibitors, and almost all of the top-ranked ligands had a Tc lower than 0.34 among themselves, with the sole exception being cytarabine (**9**) and compound **3** with a Tc = 0.57. The similarity between these latter two molecules arises from both containing a hydroxymethyl tetrahydrofuran group, with cytarabine possessing in addition a pyrimidine group, while compound **3** has a benzimidazole group. For both compounds, the tetrahydrofuran group participates in polar interactions, while the pyrimidine or benzimidazole groups interact with aromatic and hydrophobic groups in the protein binding site.

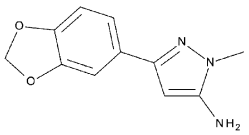
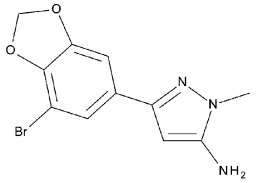
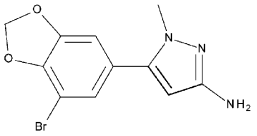
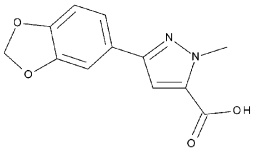
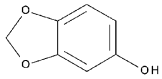
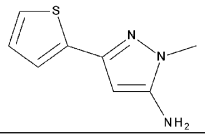
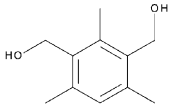
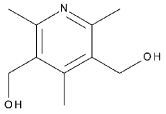
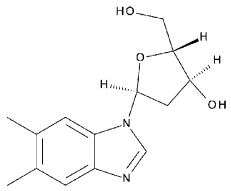
Two well-known drugs were among the top-ranked compounds: amlexanox (**10**) is an antiallergic, topical antiulcer, and anti-inflammatory agent, while cytarabine (**9**) is a nucleoside analog antineoplastic agent used in the treatment of leukemia that also has antiviral properties. They also possess a hydrophobic/aromatic group and a polar group in a moderate to small sized overall molecular structure. Cytarabine has been successfully used to treat Burkitt and meningeal lymphoma in HIV positive patients,[44−46] as well as decreasing HIV-1 receptor expression and reducing susceptibility to HIV-1 of human T-lymphoid cells.[47]

All of the compounds in Table 5 bind in an allosteric binding site that is separate from the binding site containing a mobile loop and metal ions. The binding mode of both the known inhibitor **833** and of the highly ranked amlexanox (**10**) is shown in Figure 3.

The allosteric binding site is comprised of residues including the following: Val 77 → Ile 84, Trp 108 A, Trp 108 B, Pro 109 B, Val 150 → Ile 161, Val 180, His 185, Val 201, Ile 204, and Asp 207. The known inhibitor **833** makes hydrogen bonds with Ile 204 backbone carbonyl and Asp 207 side chain carboxylate, while amlexanox donates a hydrogen bond to Asp 207 carboxylate and receives a hydrogen bond from the Arg 187 backbone NH. Both ligands make hydrophobic contacts to Val 201 and Ile 204, as well as extensive $\pi-\pi$ interactions with Tyr 83, Trp 108 chain A, Trp 108 chain B, and His 185.

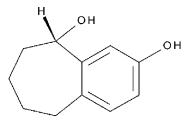
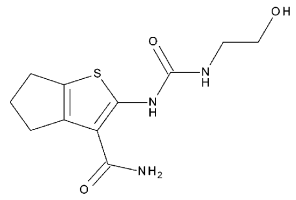
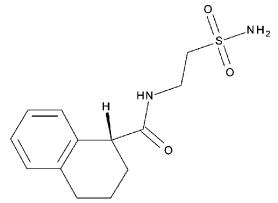
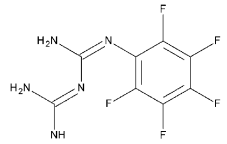
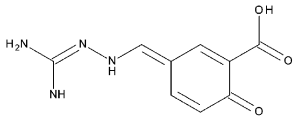
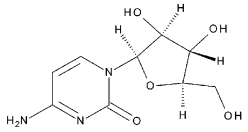
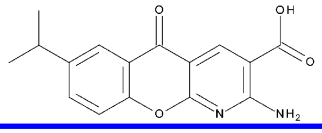
The compounds in Table 5 were used to calculate their values according to DrugLogit equations,[38] to provide a measure of the similarity of a compound to approved small-molecule drugs. According to DrugLogit's general eq 6, where probability ($P_{eq\ 6}$) of a compound to be classified as a drug versus a nondrug compound is a function of aliphatic ring

Table 5. Top Ranked Ligands and Known Inhibitors for HIV-1 Integrase in an Allosteric Binding Site[a]

833		MW = 216 logP = 3.14	ΔG/MSA = -0.029 ΔG/NoC = -0.729	Peq6 = 0.92 Peq19 = 0.97
AVX		MW = 296 logP = 3.94	ΔG/MSA = -0.029 ΔG/NoC = -0.788	Peq6 = 0.89 Peq19 = 0.98
BBY		MW = 296 logP = 3.14	ΔG/MSA = -0.028 ΔG/NoC = -0.764	Peq6 = 0.92 Peq19 = 0.99
BMC		MW = 246 logP = 2.45	ΔG/MSA = -0.025 ΔG/NoC = -0.598	Peq6 = 0.86 Peq19 = 0.95
BZX		MW = 138 logP = 1.37	ΔG/MSA = -0.032 ΔG/NoC = -0.769	Peq6 = 0.92 Peq19 = 0.96
MPV_Wielens3		MW = 180 logP = 2.41	ΔG/MSA = -0.028 ΔG/NoC = -0.808	Peq6 = 0.84 Peq19 = 0.65
1		MW = 180 logP = 1.71	ΔG/MSA = -0.022 ΔG/NoC = -0.555	Peq6 = 0.79 Peq19 = 0.87
2		MW = 181 logP = 0.21	ΔG/MSA = -0.025 ΔG/NoC = -0.686	Peq6 = 0.80 Peq19 = 0.77
3		MW = 262 logP = 2.77	ΔG/MSA = -0.022 ΔG/NoC = -0.545	Peq6 = 0.93 Peq19 = 0.97

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

Table 5. continued

4 	MW = 178 logP = 2.25	ΔG/MSA = -0.029 ΔG/NoC = -0.645	Peq6 = 0.92 Peq19 = 0.97
5 	MW = 269 logP = -1.33	ΔG/MSA = -0.021 ΔG/NoC = -0.623	Peq6 = 0.43 Peq19 = 0.54
6 	MW = 282 logP = 0.77	ΔG/MSA = -0.023 ΔG/NoC = -0.638	Peq6 = 0.65 Peq19 = 0.80
7 	MW = 267 logP = 0.82	ΔG/MSA = -0.029 ΔG/NoC = -0.950	Peq6 = 0.69 Peq19 = 0.92
8 	MW = 222 logP = 0.41	ΔG/MSA = -0.026 ΔG/NoC = -0.743	Peq6 = 0.82 Peq19 = 0.11
9: Cytarabine 	MW = 243 logP = -1.1	ΔG/MSA = -0.028 ΔG/NoC = -0.803	Peq6 = 0.77 Peq19 = 0.72
10: Amlexanox 	MW = 283 logP = 3.08	ΔG/MSA = -0.029 ΔG/NoC = -0.928	Peq6 = 0.96 Peq19 = 0.88

[a]MW = molecular weight (g/mol); ΔG/MSA = consensus predicted free energy of binding divided by the molecular surface area (kcal/mol·Å^2); ΔG/NoC = consensus predicted free energy of binding divided by the number of carbons atoms (kcal/mol·NoC).

count + log*P* − MW, all of the known inhibitors had $P_{eq\ 6}$ > 0.84, and most compounds **1**−**10** had $P_{eq\ 6}$ values higher than 0.8, with compound **5** having the lowest value of 0.43. This would indicate that all compounds except **5** have a combination of properties similar to drug compounds. Compound **5** appears to be very hydrophilic, with the lowest log*P* in Table 5 of −1.33 that is not very common among drugs, as is shown by the terms in DrugLogit eq 6.

DrugLogit also provides specific equations for organs or diseases, to distinguish drugs and molecules for particular pharmacological categories. According to DrugLogit eq 19, where probability ($P_{eq\ 19}$) of a compound being classified in the specific pharmacological drug category of anti-infectives for systemic use is a function of ring count − NoC − PSA + Wiener polarity, all of the known inhibitors had $P_{eq\ 19}$ values > 0.95, except MPV_Wielens_cmc_3, with a $P_{eq\ 19}$ of 0.65. The top-ranked ligands all had high $P_{eq\ 19}$ values, except compounds **5** and **8**, with 0.54 and 0.11, respectively. This means that most compounds have a combination of physicochemical properties similar to known anti-infective drugs, except **5** and **8**, that may have too large PSA values; compound **5**, in addition, being not very similar to known drugs in general. An analysis of this sort can help to profile compounds as to their broad mechanism of action and localization in the body. According to Bayesian predictions[48] that use hydrogen bond acceptor count, hydrogen bond donor count, log*P*, MW, NHA, PSA, Δ*G*, and some ligand efficiencies, all known inhibitors and top-ranked compounds had properties similar to known drugs, with their probability of having properties similar to drug compounds

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

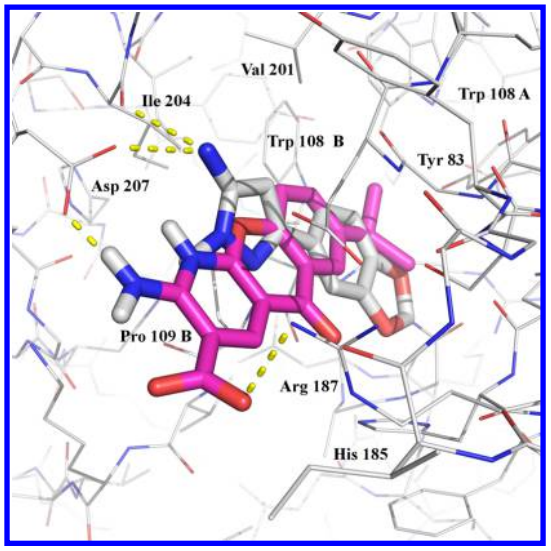


Figure 3. Docked binding mode for small molecule drug amlexanox (**10**, in magenta) with HIV-1 integrase structure 3AO4, superposed with native inhibitor **833** (3-(1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazol-5-amine, in white). Ligand polar hydrogens are shown in white, oxygen atoms in red, and nitrogen atoms in blue. Intermolecular hydrogen bonds are shown as yellow dashed lines.

being higher than their probability of having properties similar to nondrug compounds.

All of the compounds in Table 5 pass Lipinski's Rule of Five, a measure of the likely oral bioavailability of these chemical compounds.[49] In addition, they have PSA values lower than 140 Å^2, atom count within 20 to 70 atoms, and less than ten rotatable bonds, which are in line with their general drug-likeness,[50] and have MW values lower than 300 g/mol, as is the case for leads.[51,52] The compounds profiled and shown in Table 5 also have good ADMET scores according to Gleeson et al.[53] at below or close to 1.0, comparable to most oral drugs (see Table S2 in the Supporting Information), except **2** (2.16) and **5** (2.04). However, the known drug cytarabine also has an ADMET score of 1.91.

Antitargets. The thresholds for recording a hit against an antitarget were based on the GlideScore, Autodock binding energy, and Vina score of native ligands where available, as well as the score for extra compounds that are also known actives, and are shown in Table 6. There is a general concordance in the scores for each ligand according to each docking program.

A small degree of protein flexibility is taken into account by the scaling of the van der Waals radii of the protein atoms while docking, and proteins such as PXR do not change substantially their conformation between *apo* and *holo* states, but large flexibility in the protein targets may not be taken into account. Also, compounds may or may not be metabolized in the body, and it can be either the compound or its metabolite(s) that may be responsible for any unwanted side effects.

The five antitarget proteins compose a battery for generating color-coded antitarget battery profiles in Figure 4, that summarize the binding interactions for each top-ranked ligand from Table 5 determined separately by Glide, Autodock, and Vina. For the antitarget interaction profile, if the sum of any string for a compound is under five, then it may be regarded as moderately safe. If the sum of the strings is over five, or if any of the strings has a total value of five, then the compound may be flagged as needing close attention. These limits are empirical and can be changed to a required level.

Table 6. Thresholds and Scores (in kcal/mol) for Antitarget Hit Determination

antitarget	Glide XPScore	Autodock 4	Vina
PXR, cocrystallized ligand Hyperforin	−7.7	−12.5	−10.3
A-792611	−16.0	−15.3	−11.8
SULT, cocrystallized ligand L-dopamine	−6.3	−7.5	−6.0
ChapmanE1	−7.1	−8.6	−9.8
CYP 2a6, cocrystallized ligand coumarin	−7.6	−6.8	−8.5
methoxsalen	−8.0	−8.4	−7.1
CYP 2c9, cocrystallized ligand S-warfarin	−8.7	−9.4	−10.0
sulfaphenazole	−7.0	−9.1	−7.6
CYP 3a4	−7.5	−7.5	−7.5
nephazodone	−4.1	−11.7	−8.2

The high-ranked compounds in Table 5 and Figure 4 had a moderate inhibition profile against the antitarget proteins, perhaps due to their relatively small size that provide lower interaction energies with the antitargets. Compound **7** was predicted to have low interactions with CYP 3a4, a prediction that is corroborated by experimental results of inactivity against this enzyme.[54] Also important to note is that the known drugs cytarabine (**9**) and amlexanox (**10**), which have a long safe use record, each had a low interaction profile with the antitargets. This indicates that the antitarget battery might provide a useful tool to profile compounds for interactions off-target to their pharmaceutical target.

Most of the proposed inhibitors in Table 5 are small compounds with relatively simple chemical structures, relatively rigid compounds due to at least one aromatic cycle. Given their small size, their ligand efficiencies are good, which also means that they can be modified and optimized more easily than larger compounds.[55] They also have good (strongly negative) ligand efficiency values as compared for example to threshold values of −0.24 kcal/mol·NHA based on known protein−protein interaction inhibitors[36] and in line with other studies[37] (Table S1 in the Supporting Information). Improvements in ΔG/MSA and ΔG/NoC could also result in more favorable desolvation energies for the ligands and thus improved pharmacokinetics. Compounds designed to have deep efficiency index values benefit from being easier to develop,[56] as well as possibly having less side effects.[57]

There is a balance between a drug's benefits and the side effects it may produce, which is dependent on the severity of the disease, the availability of therapeutical options, among other things. No drug is completely safe, in the sense that any drug can produce side effects. The severity of these can depend intrinsically on the nature of the drug, as well as the patient, dose, interactions, health of the patient, impurities, *etc.* Designing drugs that have lower toxicity and side effects is indeed required. None of the compounds had strong interactions with CYP 2a6. Interaction with a CYP does not automatically mean that a compound is toxic, but it shows that it can be metabolized into a different substance that may have a different activity and side effects which must be monitored. If a compound inhibits a CYP enzyme, then other drugs that are administered at the same time may have a different effect since the particular CYP enzyme that has been inhibited by the first

dx.doi.org/10.1021/ci500300u | J. Chem. Inf. Model. 2014, 54, 3172−3185

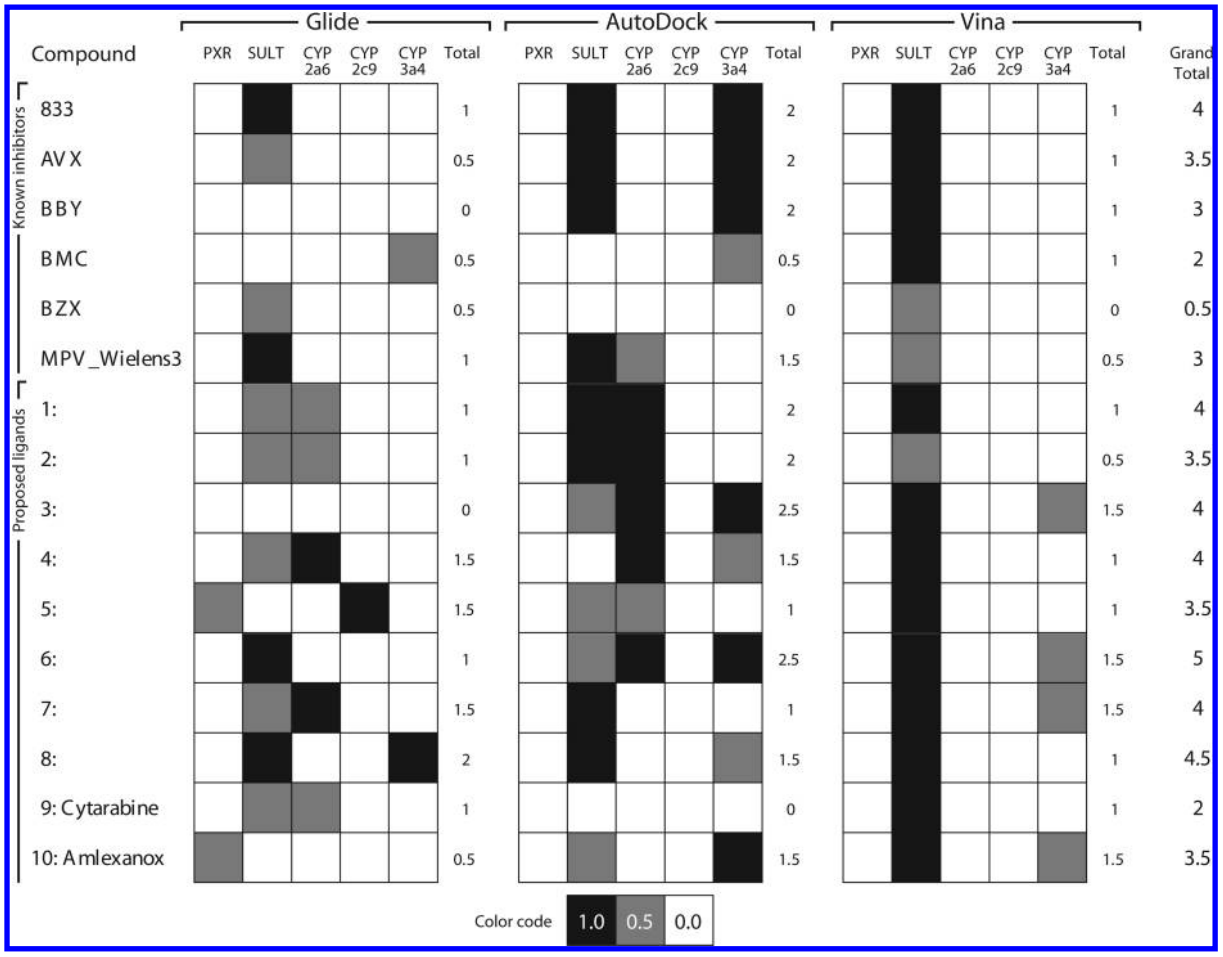


Figure 4. Interaction matrix between known and proposed ligands and antitargets.

drug is no longer available for metabolism of the other drug(s).[9] Developing HIV-1 integrase inhibitors that do not interact with CYP enzymes may increase their free concentration in plasma, which may mean that lower doses are needed to reach a therapeutic effect. CYP enzymes have a wide spectrum of substrates. The CYP 2a subfamily has a preference for coumarin and 3-ketosteroids substrates, such as testosterone.[58] CYP 3a4 also metabolizes a wide range of compounds, including small and large therapeutic compounds from different classes of drugs.[8] Among the irreversible substrates of CYPs are furans and thiophenes, dichloro- and trichloroethylenes, benzodioxoles, hydrazines, thioamides, acetylenes, epoxides, secondary amines, conjugated systems, isothiocyanates, and dithiocarbamate.[59] SULT prefers alcohol and amine substrates, and PXR also has a wide variety of substrates.[8]

Sometimes, inhibition of an antitarget may be desired, for example to free a different drug that is bound to the antitarget and boost the free levels of the latter drug in the bloodstream (a type of drug−drug interaction). This is the case of ritonavir, which inhibits cytochromes 3a and 2d6 and enables other HIV antivirals to remain unbound from this enzyme and thus unmetabolized for a longer period of time.[60] Also, if a candidate drug compound does not interact with any metabolizing enzyme, it may accumulate dangerously in the organism and therefore cause severe side effects during prolonged use. Other times, the metabolism of a relatively safe compound may convert it into a much more toxic metabolite. At yet other times, the metabolism of a prodrug will lead to the relatively safe and active drug compound which is desired and required for a particular treatment. An example of the latter is the cancer drug dacarbazine which is an alkylating agent that cross-links DNA and promotes cancerous cell destruction. Demethylation by cytochrome P450 enzymes transform this prodrug into a compound which degrades spontaneously into its active form.[2]

However, our filtering and docking procedure allowed retrieving only those compounds with good inhibition profiles against HIV-1 integrase. In preliminary results and work in progress, we observed that only a few compounds had any predicted activity against both HIV-1 integrase and HIV-1 reverse transcriptase targets, although there have been some compounds reported,[60] and none of them were predicted to strongly bind to both HIV-1 integrase and reverse transcriptase enzymes simultaneously. It may be necessary to combine inhibitors for each enzyme into a large molecule to achieve dual inhibition.[61]

The types of antitargets can be extended and used in the manner described in the present work and be tailored to suit individual design projects, ligand side effect and off-target profiling, or antitargets of interest, such as in polypharmacology.[62,63] In addition, the coming of age and widening use of chemogenomics and pharmacogenomics implies that metabolizing enzymes with a high degree of polymorphisms which confer different metabolic rates (such as is seen in CYP P450 2c9)[64] need different inhibition profiles taken into account to better tailor treatments to individuals and population groups.

CONCLUSIONS

As shown in this work, compound structure libraries can be filtered by screening out those molecules that might interact with specific antitargets. This may lead to paradigms of safer drug design, profiling metabolism and side effects of ligands (their antitarget or off-target activity), as well as their activities against their targets, as early on as possible in the drug design process. Alternatively, those molecules which are active but that have been flagged as possible binders to specific antitargets can be followed more closely with a goal to optimize against or in favor of binding to the antitargets in question. The battery of antitargets can be built to suit each different optimization strategy and calibrated according to specific antitargets. At the same time, the battery of targets can also include other proteins, identifying novel target interactions for the studied compounds.

Several interesting, small-sized compounds with new chemistry were predicted to inhibit HIV-1 integrase through a recently described allosteric site. Many of these compounds, including two available small molecule drugs, had low inhibition profiles against known metabolism proteins, as well as a good bioavailability profile, and good ligand efficiency values in line with published values. The results also indicate a way of improving inhibitors by achieving better ΔG/MSA and ΔG/NoC ligand efficiency values, either through an increase in affinity with the target or a reduction in molecular surface area and/or reducing the number of carbon atoms.

The proposed procedure also shows how approved drugs that are already in use and are relatively safe have low interactions with the antitarget battery that we have studied, and thus this approach can provide a method to include in drug discovery and design programs for improving the metabolic safety profile of candidate compounds. Pharmacodynamics, pharmacokinetics, and off-target interactions can thus be optimized in parallel. Computational chemistry makes possible several successive stages for screens of ADME properties, toxicity, and undesirable side effects: both prefiltering compound libraries to remove known functional groups and compounds which might have chemical reactivity, low solubility, low efficiency issues, etc., as well as docking against several targets and antitargets to give further assessment of the interactions of compounds with new targets.

Due to variability in the absolute scores provided by docking programs, a different way of obtaining a rerank by consensus between the ranks of the compounds was developed according to each docking program, ligand efficiency, and physicochemical parameter. This reranking based on sum of ranks and triangular numbers allows to calculate a numerical measure that can logically and objectively compare different docking procedures, protein crystal structures, ligand efficiencies, and physicochemical properties, in order to select the best screening method and limit on top-ranked compounds and, thus, the best way to recover known inhibitors in the top-ranked ligands and ligands that may have activity against a protein target if they are comparable to such known inhibitors. The triangular number reranking uses consensus between ranks, hedging results, and avoiding extreme values, as well as avoiding other ways of merging several run results, such as averages of ranks, absolute scores, average of scores, among others.

ASSOCIATED CONTENT

Supporting Information

Table S1, consensus ligand efficiency index values for the top-ranked compounds; Figures S1−S3, receiver-operator curves for Glide XP, Autodock 4, and Vina; Table S2, ADMET score of Gleeson et al. for the top-ranked 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.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Estonian Ministry for Education and Research (grant SF0140031Bs09).

REFERENCES

(1) *2008 Report on the Global AIDS Epidemic*; UNAIDS: Geneva, Switzerland, 2008.

(2) Patrick, G. L. *An Introduction to Medicinal Chemistry*, 3rd ed.; Oxford University Press: NY, USA, 2005.

(3) Summa, V.; Petrocchi, A.; Bonelli, F.; Crescenzi, B.; Donghi, M.; Ferrara, M.; Fiore, F.; Gardelli, C.; Gonzalez Paz, O.; Hazuda, D. J.; Jones, P.; Kinzel, O.; Laufer, R.; Monteagudo, E.; Muraglia, E.; Nizi, E.; Orvieto, F.; Pace, P.; Pescatore, G.; Scarpelli, R.; Stillmock, K.; Witmer, M. V.; Rowley, M. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* **2008**, *51*, 5843−5855.

(4) Alves, C. N.; Martí, S.; Castillo, R.; Andrés, J.; Moliner, V.; Tuñón, I.; Silla, E. A quantum mechanics/molecular mechanics study of the protein-ligand interaction for inhibitors of HIV-1 integrase. *Chem.—Eur. J.* **2007**, *13*, 7715−7724.

(5) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13040−13043.

(6) Wielens, J.; Headey, S. J.; Deadman, J. J.; Rhodes, D. I.; Parker, M. W.; Chalmers, D. K.; Scanlon, M. J. Fragment-based design of ligands targeting a novel site on the integrase enzyme of human immunodeficiency virus 1. *ChemMedChem.* **2011**, *6*, 258−261.

(7) Ortiz de Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **2010**, *110*, 932−948.

(8) *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.

(9) 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.

(10) Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, A. J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321*, 263−266.

(11) Keiser, M. J.; Etola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175−U48.

(12) Freitas, R. F.; Bauab, R. L.; Montanari, C. A. Novel Application of 2D and 3D-similarity searches to identify substrates among cytochrome P450 2c9, 2d6, and 3a4. *J. Chem. Inf. Model.* **2010**, *50*, 97−109.

(13) Protein Data Bank. Research Collaboratory for Structural Bioinformatics. http://www.pdb.org/pdb/home/home.do (accessed February 2013).

(14) *Maestro*, version 9.4; Schrödinger, LLC: New York, 2013.
(15) Protein Preparation Wizard, Schrödinger release 2013; Schrödinger, LLC: New York, 2013.
(16) García-Sosa, A. T. Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J. Chem. Inf. Model.* **2013**, *53*, 1388−1405.
(17) 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.
(18) NationalCancerInstitute/NationalInstitutesofHealthUSA.DTPs-DiversitySetInformation. http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html (accessed February 2013).
(19) 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.
(20) *LigPrep*, version.2.6.044; Schrödinger, LLC: New York, 2013.
(21) *Virtual Screening Workflow*; Schrödinger, LLC: New York, 2013.
(22) 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.
(23) Trott, O.; Olson, A. J. Autodock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.
(24) García-Sosa, A. T.; Sild, S.; Maran, U. Docking and virtual screening using distributed grid technology. *QSAR Comb. Sci.* **2009**, *28*, 815−821.
(25) Martin, Y. C. Let's not forget tautomers. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 693−704.
(26) Nichols, S. E.; Domaoal, 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.
(27) García-Sosa, A. T.; Sild, S.; Takkis, K.; Maran, U. Combined approach using ligand efficiency, cross-docking, and antitarget hits for wild-type and drug-resistant Y181C HIV-1 reverse transcriptase. *J. Chem. Inf. Model.* **2011**, *51*, 2595−2611.
(28) 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.
(29) Chapman, E.; Ding, S.; Shultz, P. G.; Wong, C. A potent and highly selective sulfotransferase inhibitor. *J. Am. Chem. Soc.* **2002**, *124*, 14524−14525.
(30) *Marvin* v5.6.0.1; ChemAxon: 2011. http://www.chemaxon.com (accessed February 2013).
(31) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140−2148.
(32) *OpenBabel* v2.3.1. http://openbabel.org (accessed Feb 24, 2013).
(33) 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.
(34) Hopkins, A. L.; Groom, C. R.; Alex, A. Ligand efficiency: A useful metric for lead system. *Drug Discovery Today* **2004**, *9*, 430−431.
(35) Abad-Zapatero, C.; Metz, J. T. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* **2005**, *10*, 464−469.
(36) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein−protein interfaces. *Nature* **2007**, *450*, 1001−1009.
(37) 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.
(38) García-Sosa, A. T.; Oja, M.; Hetényi, C.; Maran, U. DrugLogit: Logistic discrimination between drugs and non-drugs including disease-specificity by assigning probabilities based on molecular properties. *J. Chem. Inf. Model.* **2012**, *52*, 2165−2180.
(39) García-Sosa, A. T.; Oja, M.; Hetényi, C.; Maran, U. Disease-specific differentiation between drugs and non-drugs using principal component analysis of their molecular descriptor space. *Mol. Inf.* **2012**, *31*, 369−383.
(40) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to optimize shape-based virtual screening: choosing the right query and including chemical information. *J. Chem. Inf. Model.* **2009**, *49*, 678−692.
(41) Tawa, G. J.; Baber, J. C.; Humblet, C. J. Computation of 3D queries for ROCS based virtual screens. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 853−868.
(42) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 179−190.
(43) Nicholls, A. What do we know and when do we know it? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239−255.
(44) Wang, E.; Straus, D. J.; Teruya-Feldstein, J.; Qin, J.; Portlock, C.; Moskowitz, C.; Goy, A.; Hedrick, E.; Zelenetz, A. D.; Noy, A. Intensive chemotherapy with cyclophosphamide, doxorubicin, high-dose methotrexate/ifosfamide, etoposide, and high-dose cytarabine (CODOX-M/IVAC) for human immunodeficiency virus-associated Burkitt lymphoma. *Cancer* **2003**, *98*, 1196−1205.
(45) Montoto, S.; Wilson, J.; Shaw, K.; Heath, M.; Wilson, A.; McNamara, C.; Orkin, C.; Nelson, M.; Johnson, M.; Bower, M.; Cwynarski, K. Excellent immunological recovery following CODOX-M/IVAC, an effective intensive chemotherapy for HIV-associated Burkitt's lymphoma. *AIDS* **2010**, *24*, 851−856.
(46) Mazhar, D.; Stebbing, J.; Lewis, R.; Nelson, M.; Gazzard, B. G.; Bower, M. The management of meningeal lymphoma in patients with HIV in the era of HAART: intrathecal depot cytarabine is effective and safe. *Blood* **2006**, *107*, 3412−3414.
(47) Groschel, B.; Kaufmann, A.; Cinatl, J.; Doerr, H. W.; Cinatl, J., Jr. Cytarabine treatment of human T-lymphoid cells induces decreased HIV-1 receptor expression and reduced HIV-1 susceptibility. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 1433−1437.
(48) García-Sosa, A. T.; Maran, U. Drugs, non-drugs, and disease category specificity: organ effects by ligand pharmacology. *SAR QSAR Environ. Res.* **2013**, *24*, 585−597.
(49) 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.
(50) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.
(51) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308−1315.
(52) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discovery Today* **2003**, *8*, 876−877.
(53) Gleeson, M. P.; Hersey, A.; Montanari, D.; Overington, J. Probing the links between in vitro potency, ADMET and physico-chemical properties. *Nat. Rev. Drug Discovery* **2011**, *10*, 197−208.
(54) National Institutes of Health. PubChem BioAssay. **qHTS Assay for Inhibitors and Substrates of Cytochrome P450 3A4.** http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=8854 and **qHTS Assay for Activators of Cytochrome P450 3A4.** http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=885 (accessed December 2013).
(55) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660−672.
(56) 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.

(57) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5*, 358−364.
(58) Ioannides, C. *Cytochromes P450 role in the metabolism and toxicity of drugs and other xenobiotics*; RSC Publishing: Cambridge, UK, 2008.
(59) 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.
(60) Billamboz, M.; Bailly, F.; Barreca, M. L.; De Luca, L.; Mouscadet, J.-F.; Calmels, C.; Andreola, M.-L.; Witvrouw, M.; Christ, F.; Debyser, Z.; Cotelle, P. Design, synthesis, and biological evaluation of a series of 2-hydroxyisoquinoline-1,3(2*H*,4*H*)-diones as dual inhibitors of Human Immunodeficiency Virus type 1integrase and the reverse transcriptase RNAse H domain. *J. Med. Chem.* **2008**, *51*, 7717−7730.
(61) Wang, Z.; Bennett, E. M.; Wilson, D. J.; Salomon, C.; Vince, R. Rationally designed dual inhibitors of HIV reverse transcriptase and integrase. *J. Med. Chem.* **2007**, *50*, 3416−3419.
(62) Hopkins, A. L.; Mason, J. S.; Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **2006**, *16*, 127−136.
(63) Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **2008**, *4*, 682−690.
(64) García-Martín, E.; Martínez, C.; Ladero, J. M.; Agúndez, J. A. Interethnic and intraethnic variability of CYP2C8 and CYP2C9 polymorphisms in healthy individuals. *Mol. Diagn. Ther.* **2006**, *10*, 29−40.