Parallel Screening: A Novel Concept in Pharmacophore Modeling and Virtual Screening†

Theodora M. Steindl,‡,§ Daniela Schuster,‡ Christian Laggner,‡ and Thierry Langer*,‡,§

Institute of Pharmacy, Computer Aided Molecular Design Group, University of Innsbruck, Innrain 52c, Austria, Center for Molecular Biosciences Innsbruck (CMBI), Peter-Mair-Str.1, A-6020 Innsbruck, Austria, and Inte:Ligand GmbH, Clemens Maria Hofbauer-Gasse 6, A-2344 Maria Enzersdorf, Austria

Received May 18, 2006

Parallel screening comprises a novel in silico method to predict the potential biological activities of a compound by screening it with a multitude of pharmacophore models. Our aim is to provide a fast, large-scale system that allows for virtual activity profiling. In this proof of principle study, carried out with the software tools LigandScout and Catalyst, we present a model work for the application of parallel pharmacophore-based virtual screening on a set of 50 structure-based pharmacophore models built for various viral targets and 100 antiviral compounds. The latter were screened against all pharmacophore models in order to determine if their biological targets could be correctly predicted via an enrichment of corresponding pharmacophores matching these ligands. The results demonstrate that the desired enrichment, that is, successful virtual activity profiling, was achieved for approximately 90% of all input molecules. We discuss descriptors for output validation, as well as various aspects influencing the analysis of the obtained activity profiles, and the effect of the utilized search modus for screening.

INTRODUCTION

The generation of pharmacophore models and consequent compound/database (DB) screening to identify molecules with the desired biological effect has become a broadly applied technique in drug discovery. The classical pharmacophore approach comprises the following steps: After definition of the target of interest, relevant data are collected from a thorough literature search, for example, structural information on the target/binding site, known active ligands, interaction patterns, and so forth. This is followed by the determination of the pharmacophoric features, that is, the functional and steric requirements for ligand binding, and consequent pharmacophore model generation. A pharmacophore can form in the scientist's head, on a sheet of paper, or on the computer. It not only combines and visualizes critical interactions but also serves the purpose of checking other compounds for their ability to map the required features. With the appropriate software tool, even large DBs can be screened. Of course, these models should be validated beforehand as efficiently as possible.[1−3]

Parallel screening (PS), however, focuses on the ligand rather than on the target/pharmacophore model. Assuming there is a multitude of pharmacophore models representing a variety of different pharmacological targets, would it not be desirable to have a system to screen a molecule simultaneously against all of these models? One would obtain a hit list of matching pharmacophores and could link them to the biological targets, thereby enabling an in silico identification of macromolecular systems that will possibly be influenced by this ligand. That is exactly the aim that PS seeks to achieve: a compound screened against a set of high quality pharmacophore models will provide a hit list of mapping models, the so-called pharmacophoric profile. According to the targets encoded by these models, a pharmacological profile for the compound will emerge (Scheme 1). On one hand, this enables the virtual characterization of the biological properties of new compounds. On the other hand, the sphere of action for substances with already established activities could be enlarged, an often highly successful concept in drug development.[4,5] Of course, if the PS system includes the appropriate models, prognoses will also cover toxicity, side effects, antitargets, and metabolic pathways.[6−11] Another interesting aspect in PS is that the behavior of a compound in the system can point toward promiscuity and therefore be a warning signal. Further applications of this approach might be the fine-tuning of early results in high-throughput screening or ideas for the identification of targets hit by natural products that are therapeutically used because of long-time experience but not mechanistically characterized.[12−14]

A system for PS requires (i) a large set of pharmacophore hypotheses including the availability of a fast and reliable tool for their automatic generation and (ii) a high-speed screening platform to test one compound against a variety of models, which should also allow for analysis, visualization, and facile interpretability of the output data.

Our aim is to provide such an automatic system for fast virtual activity profiling of compounds. The number of pharmacophore hypotheses in the system is constantly growing, targeted at extensive coverage of available biologi-

* To whom correspondence should be addressed. Tel.: +43 512 507 5252. Fax +43 512 507 5269. E-mail: thierry.langer@uibk.ac.at.

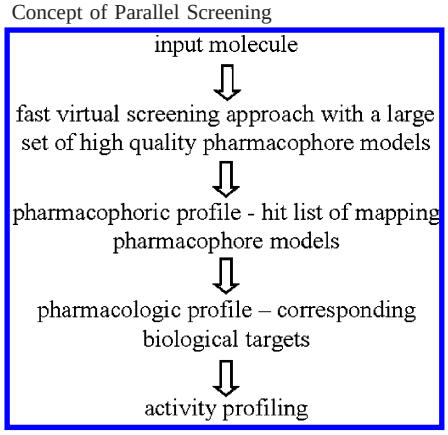
† Parts of this study have been presented at the 9th E.U. Catalyst User Group Meeting 2006, March 23, 2006, Frankfurt, Germany, and at the 231st ACS National Meeting, March 29, 2006, Atlanta, GA, as an oral presentation.

‡ University of Innsbruck and CMBI.

§ Inte:Ligand GmbH.

10.1021/ci6002043 CCC: $33.50
Published on Web 08/23/2006

Scheme 1. Concept of Parallel Screening



cal targets. This study is focused on the validation of a fraction of this system and presents a first test case. Thereby, the PS approach was tested on a large set of antiviral compounds screened manually with a set of structure-based pharmacophore hypotheses. The main question we tried to address was of course: is fast and reliable activity profiling possible via PS; that is, are the test compounds attributed correct biological activities? A similar concept using 22 diverse targets and ligand-based virtual screening with models based on molecular similarity has been published recently.[15] Aside from profiling quality, other considerations came to attention, such as, which aspects and parameters are critical for the outcome; how can the results be analyzed, visualized, and quantified; how should the enrichment be judged; and what are the benefits and the limits. The study design comprised 100 antiviral substances which were screened against 50 pharmacophore hypotheses for several viral targets utilizing two search algorithms. For easier visualization, the obtained data are shown in a hit map. We propose four descriptors for quantifying the pharmacophoric profiles of the compounds and quote several examples to explain their meaning and interpretation. Although, profile prediction is multifactorial, and up to a certain point, depending on the user's preference and mindset, several guidelines could be derived. For most ligands, the results clearly show an enrichment of correct, that is, the desired target-representing, models in the pharmacophore hit lists. Only in approximately 10% of the cases does the data point at a false target and therefore give misleading profiles.

STUDY DESIGN

1. Targets. Our study involved the selection of several viral targets which were represented in the PS system by sets of pharmacophore models. A target had to fulfill certain criteria: Because pharmacophore hypotheses were derived in a structure-based approach, the existence of a sufficient number of complexes from the Brookhaven Protein Data Bank (PDB)[16] was a requirement. Further, we focused on proteins whose inhibitions offer therapy strategies in the combat of highly relevant viral diseases. These include HIV infection, influenza, common cold, and hepatitis C.[17] For increased universal validity of the study, we aimed to provide diversity concerning the nature of the proteins as well as inhibitory mechanisms.

HIV infection—resulting in the condition of AIDS—ranks among the most well-known and intensively studied viral infections. The inhibition of two viral key enzymes, *HIV protease* and *HIV reverse transcriptase* (RT) proved highly successful and is part of every standard therapy for this infection.[18] HIV protease is responsible for the cleavage of gag and gag-pol precursor polyproteins into mature, structural and functional viral proteins. Pharmacophore model generation was based on this catalytic center.[19] RT is characteristic of retroviruses and carries out the transcription of viral RNA to double-stranded DNA, which is then integrated into the host chromosomal DNA and transcribed by the host to viral genomic and messenger RNA. For RT, we focused on allosteric inhibition.[20] Aside from their high relevance in HIV treatment, these two targets are advantageous for our study because of the large number of known inhibitors and available PDB complexes. The same applies to *influenza neuraminidase* (NA). Influenza and the possibility of a new pandemic pose a constant threat. The inhibition of influenza virus NA was a breakthrough and the first causative therapy strategy for this disease. NA, a viral envelope glycoprotein, cleaves sialic acid residues and plays a role in viral release, preventable by blockage of the substrate binding site.[21,22] The common cold—mainly ascribable to a member of the picornavirus family, the human rhinovirus (HRV)—affects large parts of the population every year and causes enormous economic damage.[23] One strategy to combat this disease derives from inhibitory occupancy of a hydrophobic pocket within the *HRV coat protein* and consequent stabilization of the latter. This causes interference with host cell attachment, virus entry, and uncoating. The final protein included in the study is *hepatitis C virus (HCV) RNA polymerase*, which is crucial for viral genome replication through the transcription of RNA. This enzyme has been intensively investigated in recent years, because of the fact that hepatitis C presents a severe condition still lacking sufficient treatment possibilities and because of the advantageous distinctness of the viral polymerase to mammalian enzymes. Interestingly, several allosteric sites have been identified, which can be targeted by inhibitors to halt the spread of the virus.[24−27] In our PS system, three of these different interaction sites are encoded as pharmacophore hypotheses, so that the protein comprises three sets of pharmacophores and molecules instead of one like the others. These five viral target proteins are summarized in Table 1.

2. Pharmacophore Models. For each of these target proteins, 10 pharmacophores were generated on the basis of receptor−inhibitor complexes from the PDB. The three allosteric sites of HCV polymerase were represented by three, five, and two models. Structure-based pharmacophore model generation was performed with the software LigandScout.[28] The LigandScout workflow comprises the extraction and interpretation of ligands from crystal structure complexes, as well as the identification and representation of ligand−receptor interactions in pharmacophore models via appropriate features. Automatic pharmacophore creation for the PDB complexes was thereby made easily feasible. Nevertheless, thorough checking and processing of the LigandScout output is indispensable: we inspected the ligands and the proposed interaction patterns and features and compared them with information from the original PDB literature. LigandScout definition allows an atom or functional group to serve as the root for multiple features, thus resembling the situation in the binding pocket. For instance, an amine might function

Table 1. Viral Targets for Parallel Screening Study

target	disease	function	mechanism of inhibition
HIV protease	HIV infection, AIDS	cleavage of gag and gag-pol precursor polyproteins into mature, structural, and functional viral proteins	inhibition at active site
HIV reverse transcriptase	HIV infection, AIDS	synthesis of a double-stranded DNA from virus RNA for integration into host chromosomal DNA and transcription to viral genomic and messenger RNA	inhibition at allosteric site
influenza virus neuraminidase	influenza	viral envelope glycoprotein involved in viral release, cleavage of sialic acid residues of new virus particles, and host membranes	inhibition at active site
human rhinovirus coat protein	common cold	attachment to host cell receptor, viral entry, and uncoating	binding in hydrophobic pocket (capsid stabilization)
hepatitis C virus RNA polymerase	hepatitis C	viral replication, transcription of genomic RNA	inhibition at three different allosteric sites

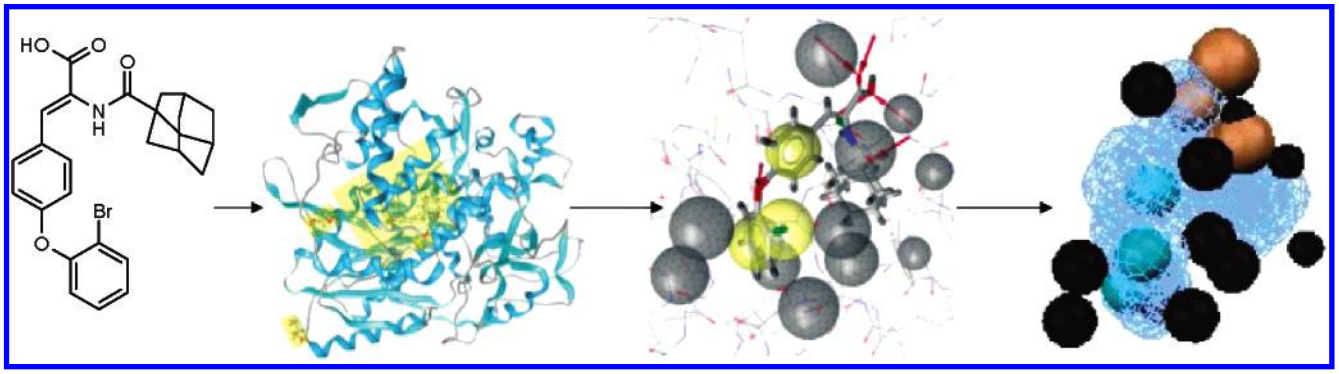


Figure 1. Process flow pharmacophore model generation: HCV polymerase inhibitor from PDB complex 1z4u (left), ligand identification and extraction (middle left) and pharmacophore identification (middle right) in LigandScout, and resulting Catalyst hypothesis (right).

as hydrogen-bond donor and as a positive ionizable feature; a hydroxyl group can face the appropriate interactions partners to accept and donate hydrogen bonds. Because such multiple features are not supported in the software Catalyst,[29] which was employed for DB screening, the LigandScout pharmacophores had to be reduced to make them compatible before they were imported into Catalyst. The process flow from the PDB complex to LigandScout pharmacophore identification and the corresponding Catalyst hypothesis is depicted in Figure 1. The models consist of the default Catalyst features except for a hydrogen-bond acceptor whose definition was enlarged to also include fluorine atoms. Excluded volume spheres and shapes provide steric restrictions, where necessary, for sufficient selectivity of the models. A LigandScout-derived Catalyst pharmacophore hypothesis should always be tested for its ability to map the minimized and conformationally characterized ligand from the original crystal structure. If the ligand could not be mapped even after model modification, the model was refused. For further validation, a fast flexible screening run of the Derwent World Drug Index (WDI)[30] was performed. It revealed that 80% of the hypotheses (40 models) find less than 500 compounds from this DB containing more than 60 000 entries. Only four of the models (8%) retrieved over 1000 hits. Aside from the total number of compounds, the occurrence of known active molecules in the hit lists was also checked. According to these investigations, the major part of the models can be attributed high selectivity. Only few models showed poorer performance. However, we decided to keep them in the pharmacophore set in order to study their impact on the results of the PS approach and to determine how activity profile accuracy is related to pharmacophore quality. Examples of hypotheses for the five different targets used in this study can be seen in Figure 2.

3. Antiviral Compounds. A total of 20 inhibitors for each of the five viral proteins was collected from PDB complexes and various literature sources giving evidence of common binding modes. Example structures can be viewed in Chart 1 and in Figure 2. Inhibitors for HIV protease target the proteolytic site. Earlier compounds display large peptidic or smaller peptidomimetic characters, such as ritonavir (**1**) or amprenavir (**2**), while more recently developed substances comprise nonpeptidic scaffolds, like the cyclic urea inhibitors, for example, DMP 323 (**3**).[31−33] For RT, we herein address solely inhibitors acting on the non-nucleoside allosteric site of the enzyme including early structures such as nevirapine (**4**) or efavirenz (**5**) and more recently established compounds such as UC 781 (**6**) combining higher potency and resilience to drug resistance.[31,32,34,35] NA inhibitors in our study cover initial transition state analogues, zanamivir (**7**)-like compounds which still carry the polar glycerol side chain of the sialic acid substrate, substances such as oseltamivir acid (**8**), where this polar part is exchanged for lipophilic residues opening a hydrophobic pocket, as well as compounds with cyclohexyl or aromatic central parts, for example, BANA 113 (**9**).[36,37] For occupancy of the hydrophobic pocket within the HRV coat protein, a long stretched topology and a predominantly hydrophobic character of the ligand is required. The so-called WIN compounds present the most prominent class of inhibitors with pleconaril (**10**) as the figure head. WIN compounds generally contain three (aromatic) rings: the central ring—a phenoxy moiety—is linked to an isoxazole by an aliphatic chain of various lengths and to the third ring via a single bond. But we also aimed to include substances with other structural compositions, such as, R 61837 (**11**) or SDZ 880-061 (**12**).[38−40] For the three HCV polymerase allosteric sites, seven, five, and eight inhibitors were selected.

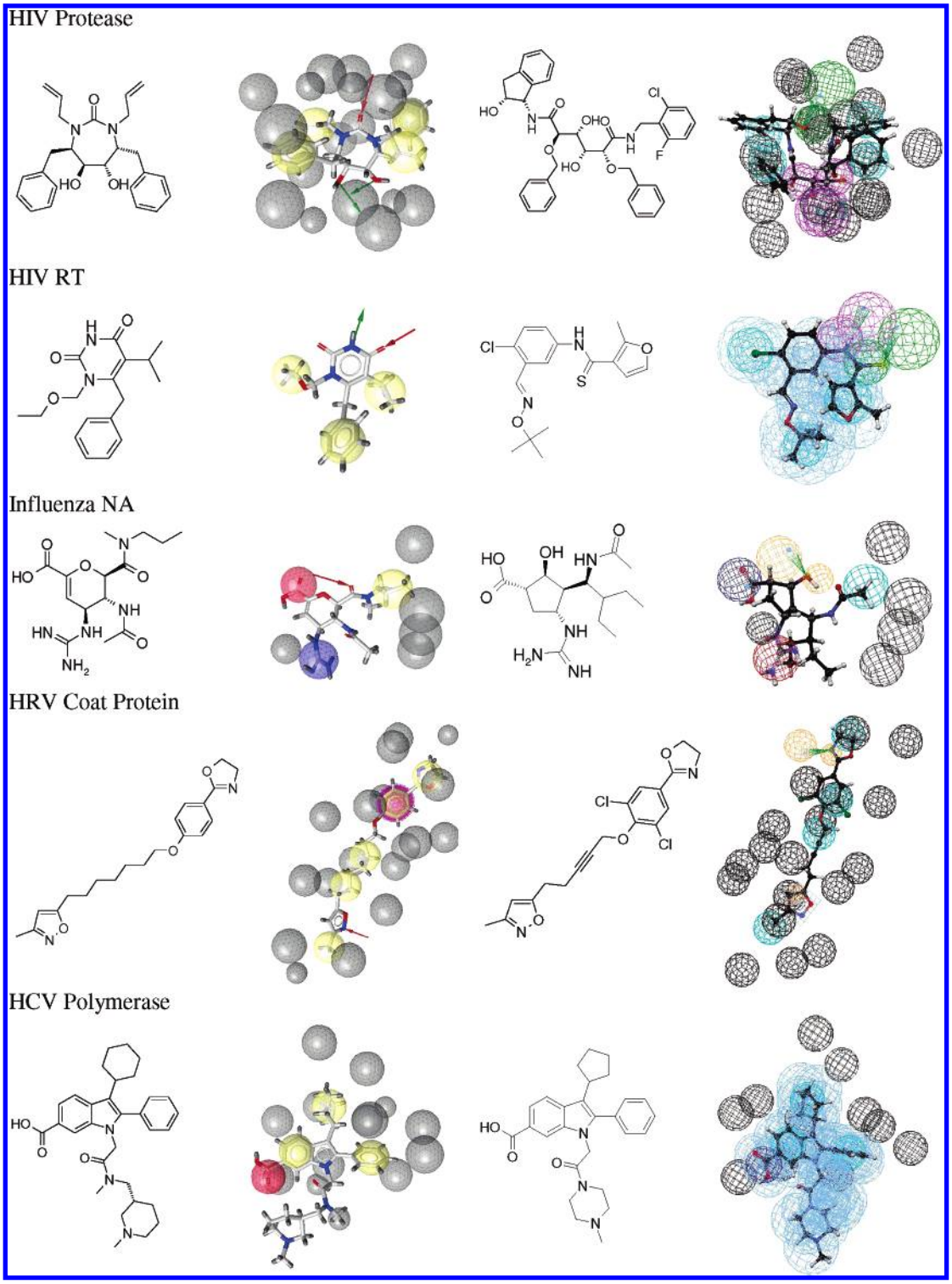
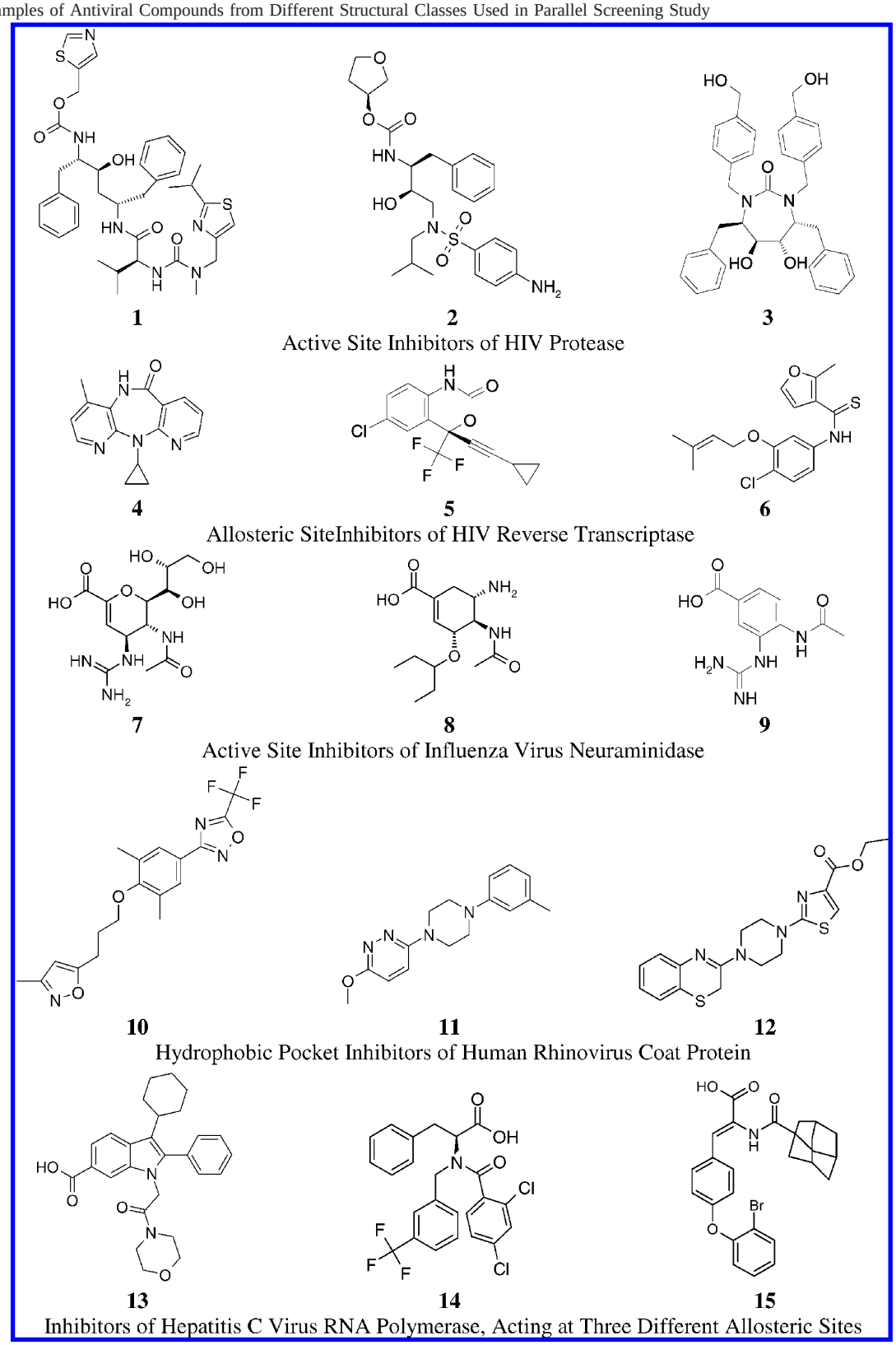


Figure 2. Examples of LigandScout hypotheses (left column of models) and corresponding Catalyst hypotheses (right column of models) for five study targets derived from PDB Entries 1hwr, 1rt1, 2qwf, 2r04, and 2brl mapped onto known inhibitors used in this study. Feature color coding for LigandScout: green (hydrogen-bond donor), red (hydrogen-bond acceptor), yellow (hydrophobic), red (negative ionizable), blue (positive ionizable), black (excluded volume sphere), and magenta (ring aromatic). Feature color coding for Catalyst: magenta (hydrogen-bond donor), green (hydrogen-bond acceptor), orange (hydrogen-bond acceptor including fluorine atoms), cyan (hydrophobic), blue (negative ionizable), red (positive ionizable), black (excluded volume sphere), light blue (shape), and orange vectorized feature with plane (ring aromatic).

To estimate the structural diversity of compounds used for this study and to eliminate the possibility that a simple 2D similarity analysis could also discriminate between inhibitors of the five different target proteins, similarity calculations between the 100 analyzed compounds were performed. For this task, we used Pipeline Pilot, which contains a component to cluster molecules on the basis of Tanimoto similarities between user-selected descriptor sets.[41,42] As the descriptor set, we chose Scitegic's proprietary method for calculating structural fingerprints, the extended connectivity fingerprints (ECFP), a method that is used for the characterization of molecules, indexing the environments of

Chart 1. Examples of Antiviral Compounds from Different Structural Classes Used in Parallel Screening Study



every atom in a molecule by using up to 4 billion different structural features. A detailed description of this method is given in the Pipeline Pilot user guide. More specifically, the ECFP_6 setting was used, which takes into account all those neighbor atoms within a diameter of six bonds when calculating the features for each atom. A maximum distance setting was used to define the maximum allowed dissimilarity between a compound and that compound that forms the center of a cluster. If the distance is above this value, a new cluster is formed. According to the different values employed for the maximum distance—0.7, 0.5, and 0.3—we obtained 30, 53, and 81 clusters, respectively. These results confirm the validity of our pharmacophore approach and give evidence of sufficient ligand diversity. Even with the most tolerant setting 0.7, the 100 analyzed molecules, which represent 20-membered inhibitor sets for only five targets,

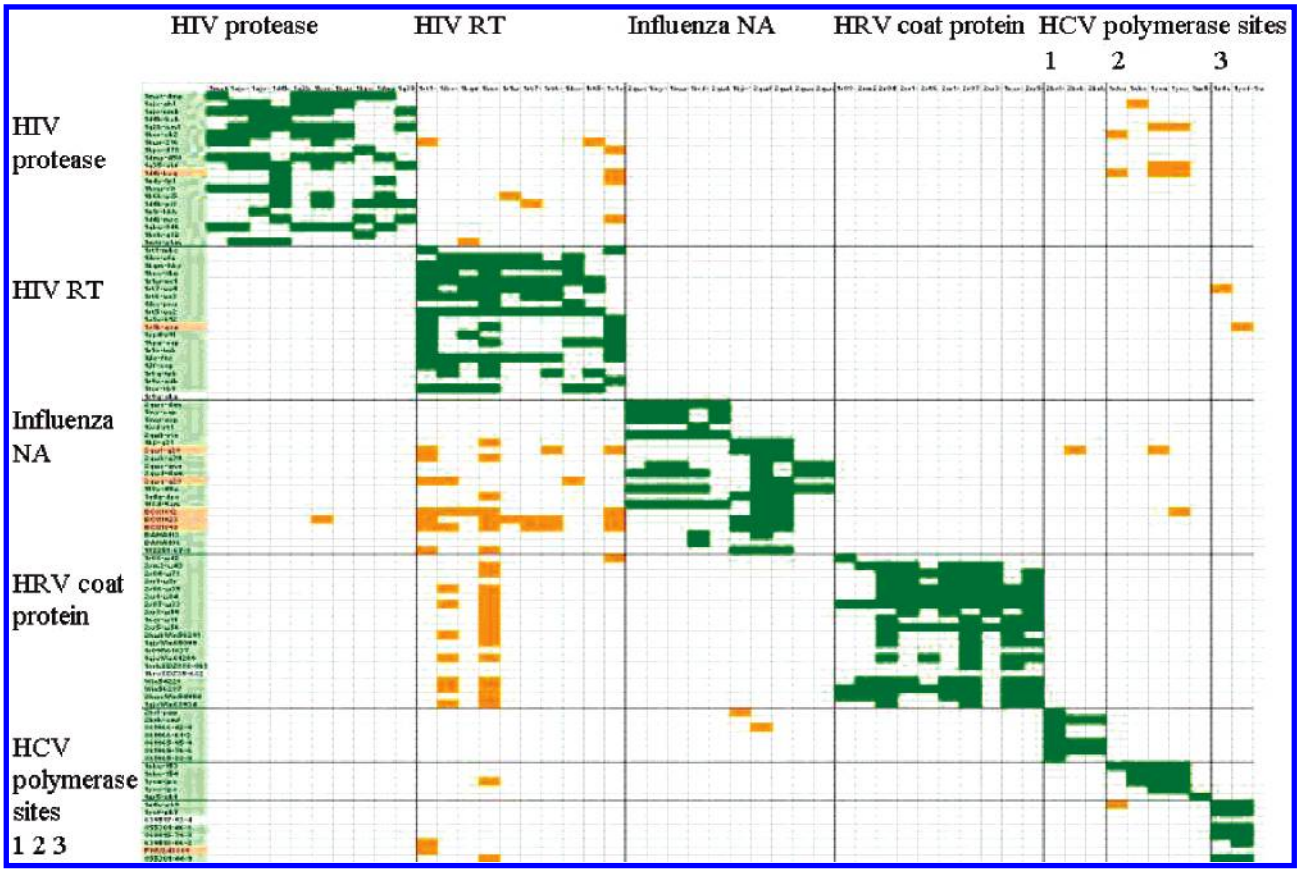


Figure 3. Hit map representing the results from parallel screening approaches with fast flexible search. Green signal, compound is found with model from correct target; orange signal, compound is found with model from another target; light green highlighting, correct activity profiling for a compound; light orange highlighting, incorrect activity profiling for a compound; and white, lack of profile predictability.

were put into as many as 30 classes. Whereas the largest cluster contains 14 ligands, 11 groups are formed only by single compounds.

For pharmacophore fitting calculations within Catalyst, 3D structure minimization of the 100 inhibitors was performed using the default Catalyst CharmM force field parameters followed by conformational model generation using the following settings: a maximum of 250 conformers, the best generation algorithm, and an energy threshold of 20 kcal above the lowest-energy conformation.

4. PS and Analysis. All 100 antiviral compounds were screened manually against the 50 structure-based pharmacophore hypotheses using the fast and the best flexible search algorithm in Catalyst. The results were analyzed; that is, a particular inhibitor was found by particular models and could therefore be linked to particular targets. For easier data interpretability, the outcome was visualized in tables, so-called hit maps (Figures 3 and 4 and in the Supporting Information). To quantify the data, we decided on four validation descriptors that were calculated to express the quality of activity profile prediction: for each of the compounds, a pharmacophore hit list was retrieved, that is, a set of hypotheses by which this particular ligand is found. Color coding shows correct and incorrect results. A correct model represents the target for which the ligand is a known inhibitor. Incorrect means that the model was built for another target. The first step was the calculation of the percentage of correct and incorrect models in the pharmacophore hit list. Furthermore, we were interested to what extent the available correct models were retrieved and which false target was most extensively identified. The relation between the last two parameters was found to be critical for activity profile prediction accuracy.

Additionally, the pharmacophore models were validated in a similar manner inspecting the hit map vertically: the percentage of known active versus inactive compounds in the obtained hit list plays a vital role for pharmacophore quality. Furthermore, it is of interest to determine the most frequently found false inhibitor class and the reasons why its members are found.

Finally, the switch from the fast to the best flexible search as well as several other critical parameters were analyzed for the impact on the selectivity of the PS system and the accuracy of activity profiling.

RESULTS AND DISCUSSION

The results of the screening experiments were visualized in a hit map, see Figures 3 and 4. Pharmacophore models are shown in the first line and must be inspected vertically, while ligands depicted in the first row can be viewed horizontally, with green boxes indicating the retrieval of a compound by a model for the correct target and orange boxes meaning retrieval with a model from another target, where the molecule is presumably inactive. A critical issue in this context is the following: in our system, we assume that one molecule displays activity only at one particular target and is inactive at all of the others—an assumption which has not been tested and verified in biological assays. However, this simplification is necessary because, without the same, no

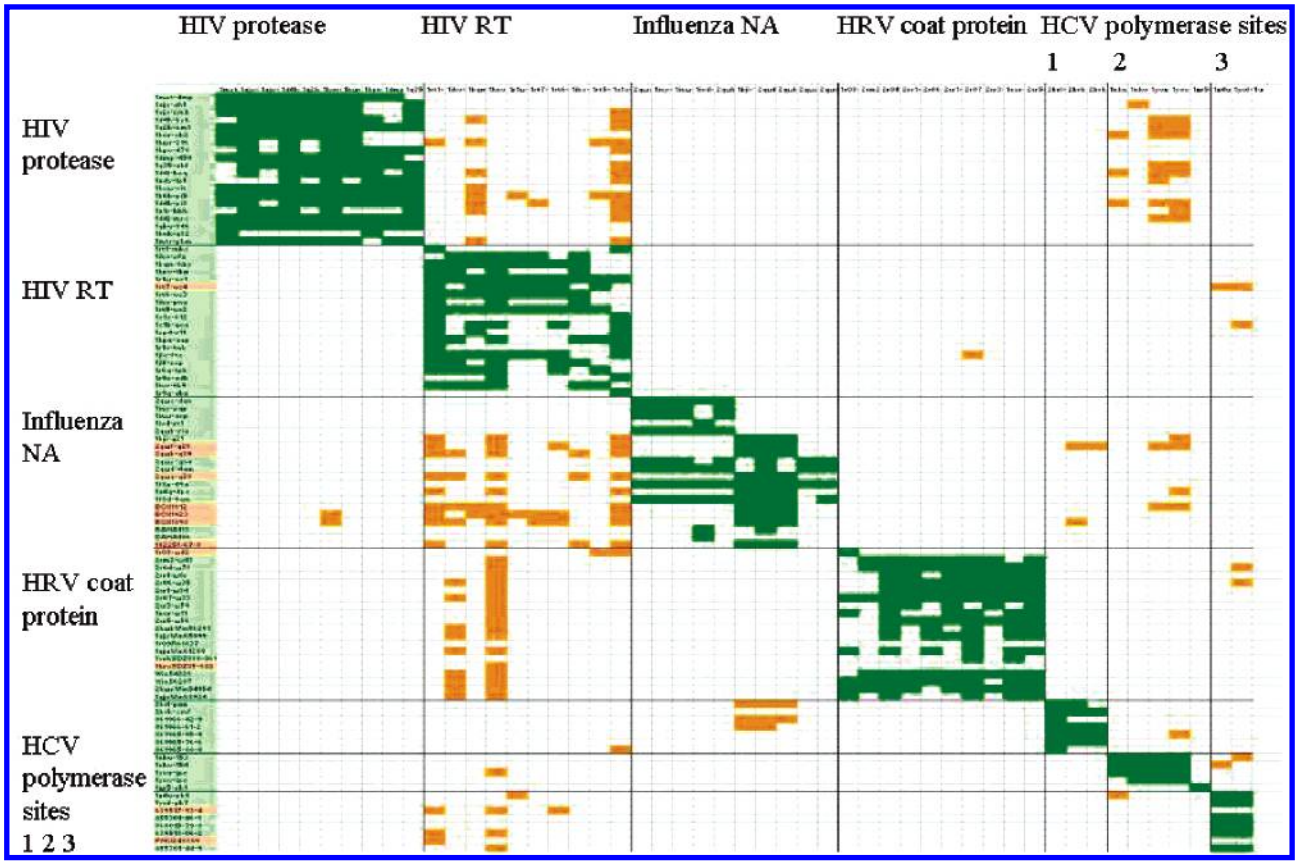


Figure 4. Hit map representing the results from parallel screening approaches with best flexible search. For description of color coding, see Figure 3.

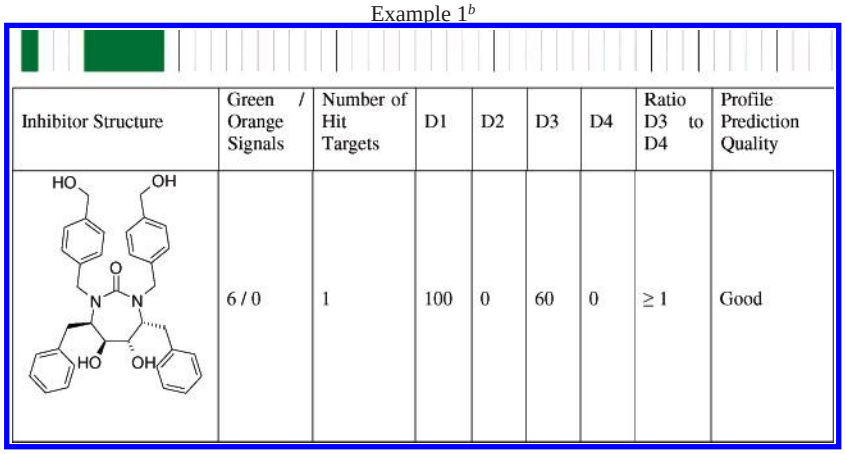
analysis and validation of the activity profiling outcome would be possible. Implicitly, it should be kept in mind that what will be called an incorrect or false model or a misleading prediction later on might in fact be correct and only indicative of a new, so far unknown activity spectrum of an inhibitor.

Already from the green and orange signals in the hit maps, enrichment is observable: inspecting the maps horizontally, for many ligands, preferentially green signals appear, that is, models representing the correct target. However, to quantify the pharmacophoric profile obtained in the screening experiments for a compound, that is, a list of pharmacophores mapping this compound, four validation descriptors were chosen: primarily, the pharmacophore hit list was inspected for the fraction of hypotheses representing the correct target and those built for another target; descriptor 1 (D1) is the percentage of correct models in the profile list, descriptor 2 (D2) is the percentage of incorrect models in the profile list, descriptor 3 (D3) is the percentage of models for the correct target, and descriptor 4 (D4) is the percentage of models for one specific false target. Already, this early step provides information for the assessment of profile prediction quality. The higher that D1 is in comparison to D2, the better the prognosis for profiling. However, D1 > D2 does not necessarily mean that a correct prediction has been achieved. This would only be true for a system with an equal number of models for each target. Because of the fact that for some target proteins just very few PDB entries have been published, while others have been characterized and investigated in numerous complexes, such an equalized situation will not exist in the system, requiring further description variables. When performing PS for a ligand, it is critical to know the following: how many models exist for the target, at which the ligand is a known inhibitor, and how many of all of these models find the particular ligand? Furthermore, models representing other targets have to be investigated. Of all wrongly identified targets, that with the best retrieval of the ligand—that is, the highest percentage of mapping models—is the most interesting one, because this is the target to which the activity profile might misleadingly point. Examples of the PS results for several inhibitors also showing the calculation of these descriptors and their impact for activity profile prediction can be viewed in Table 2.

The accuracy of activity profiling mainly depends on the last two descriptors D3 and D4. As described above, it is not that important how many correct versus incorrect models can be found in the pharmacophore hit list. The critical point is how the false models are distributed: if they are distributed among a variety of targets not showing considerable enrichment anywhere, PS will not be guided to a false target. However, caution is required for ligands found frequently and by a multitude of diverse models, because this could point to a promiscuous and therefore unwanted compound. On the other hand, if there is a clear enrichment of the returned hypotheses for one particular false target, one might be misled to believe the compound would be active there. It then depends on whether the correct or one specific false target is better represented in the pharmacophore hit list, that is, with a greater percentage of hypotheses. This question can be addressed by determining the ratio between D3 and D4. A ratio higher or at least equal to 1 points to the correct target and gives beneficial profiling information. On the

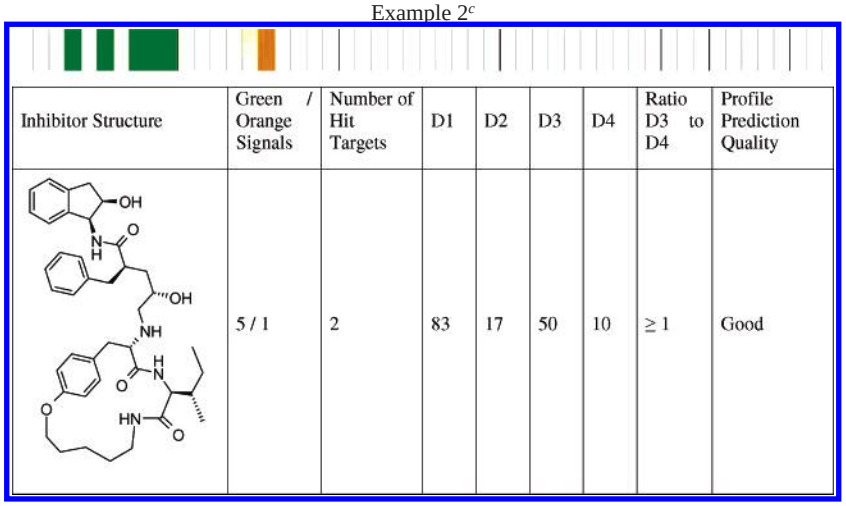
Table 2. Examples of Parallel Screening Results for Several Inhibitors, Descriptor Calculation, and Activity Profile Prediction from the Fast Flexible Search Hit Map[a]

Example 1[b]



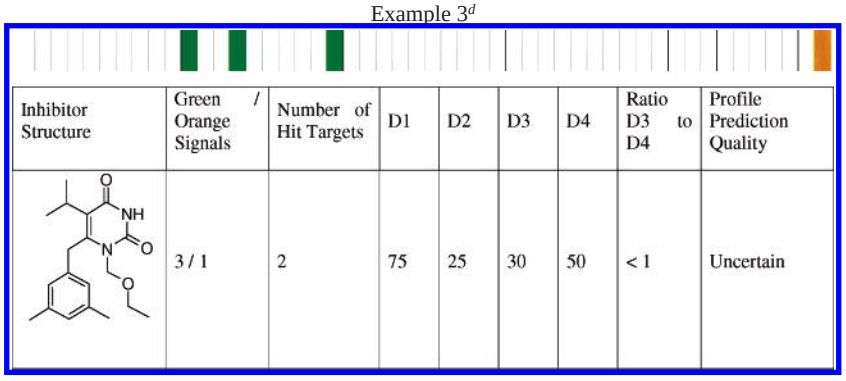
Inhibitor Structure	Green / Orange Signals	Number of Hit Targets	D1	D2	D3	D4	Ratio D3 to D4	Profile Prediction Quality
	6 / 0	1	100	0	60	0	≥ 1	Good

Example 2[c]



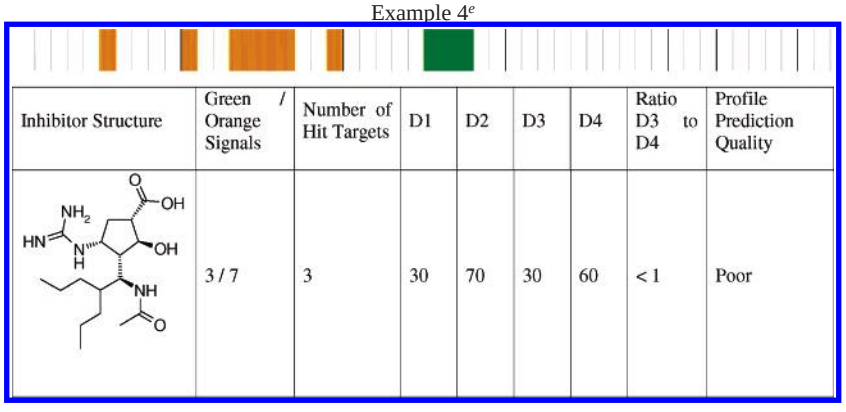
Inhibitor Structure	Green / Orange Signals	Number of Hit Targets	D1	D2	D3	D4	Ratio D3 to D4	Profile Prediction Quality
	5 / 1	2	83	17	50	10	≥ 1	Good

Example 3[d]



Inhibitor Structure	Green / Orange Signals	Number of Hit Targets	D1	D2	D3	D4	Ratio D3 to D4	Profile Prediction Quality
	3 / 1	2	75	25	30	50	< 1	Uncertain

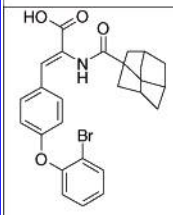
Example 4[e]



Inhibitor Structure	Green / Orange Signals	Number of Hit Targets	D1	D2	D3	D4	Ratio D3 to D4	Profile Prediction Quality
	3 / 7	3	30	70	30	60	< 1	Poor

Table 2 (Continued)

Example 5[f]

Inhibitor Structure	Green / Orange Signals	Number of Hit Targets	D1	D2	D3	D4	Ratio D3 to D4	Profile Prediction Quality
	2 / 1	1 (Two Allosteric Sites)	67	33	100	20	≥ 1	Good

[a] D1: Percentage of correct models in pharmacophoric profile list. D2: Percentage of incorrect models in pharmacophoric profile list. D3: Percentage of models for correct target. D4: Percentage of models for one specific false target. [b] HIV protease inhibitor identified by six pharmacophore models. Exclusively green signals in the hit map indicate 100% correct hit pharmacophores. Out of the 10 HIV protease pharmacophores in the system, 60% are found. Activity profile prediction is correct, straight forward, and easily interpretable. [c] HIV protease inhibitor identified by six pharmacophore models. Five green signals in the hit map indicate 83% correct and 17% incorrect hit pharmacophores. Out of the 10 HIV protease pharmacophores in the system, 50% are found. The orange signal represents one of the 10 models for RT resulting in a percentage of 10% for models found for one specific false target. Activity profile prediction is correct and easily interpretable because of the equalized situation between the two identified targets. [d] HIV RT inhibitor identified by four pharmacophore models. Three green signals in the hit map indicate 75% correct and 25% incorrect hit pharmacophores. Out of the 10 RT pharmacophores in the system, 30% are found. The orange signal represents one of the two models for an HCV polymerase allosteric site, resulting in a percentage of 50% for models found for one specific false target. Activity profile prediction is incorrect according to the D3/D4 ratio; however, it is hard to interpret because of the unequalized number of models for the two identified targets. [e] Influenza NA inhibitor identified by 10 pharmacophore models. Three green signals in the hit map indicate 30% correct and 70% incorrect hit pharmacophores. Out of the 10 NA pharmacophores in the system, 30% are found. The orange signals represent two false targets: one comes from the HIV protease and six from the RT. In this equalized situation, that makes RT the most extensively found specific false target with a percentage of 60%. Activity profile prediction is incorrect according to the D3/D4 ratio, the outcome clearly misleading to RT. [f] HCV polymerase identified by three pharmacophore models. Two green signals in the hit map indicate 67% correct and 33% incorrect hit pharmacophores. Both of the only two pharmacophores for this allosteric site in the system are found (100%). The orange signal represents one of five models for another polymerase allosteric site, resulting in a percentage of 20% for models found for one specific false site. Activity profile prediction is correct and easily interpretable because the two identified interaction sites both represent the same target.

contrary, if D3/D4 is lower than 1, activity profile prediction will lead to a false target. The lower the ratio is, the lower the chances are that the correct locus of biological action can be identified.

This ratio was selected to be the dominating parameter to determine the accuracy of prediction performance achieved in our screening approach and was therefore calculated for each compound in the hit map. First, the results obtained with the fast flexible search algorithm in Catalyst were inspected, and they gave convincing results: while a majority of inhibitors (89%) was predicted to have correct activity profiles, in only 8% of the cases, the ratio was lower than 1, and therefore, the prediction accuracy was doubtful. For three of the compounds, no mapping hypothesis was found, and therefore, no clues concerning their biological activities were gained. For better visualization in the hit map, inhibitors in the first row are color-coded representing the outcome of this ratio calculation with light green highlighting characterizing D3/D4 $\geq$ 1, light orange a ratio lower than 1, and white indicating the lack of profile predictability.

Aside from this fundamental analysis parameter—the ratio between D3 and D4—the hit map was studied further. Investigation of the profiles obtained for the screened ligands as well as an assessment of the performance of the pharmacophore models revealed other decisive aspects influencing profile prediction and should therefore be taken into consideration. Consequently, several guidelines emerged:

Guideline 1. To investigate the influence of the used search algorithm, the outcome of the best versus the fast flexible search was compared. Experiences from DB screening show that the best flexible search often retrieves a largely increased number of hits, provoking the question, do the results of the study thereby become more unselective? In fact, the outcome was hardly altered: Still 88% of all inhibitors obtained a profile ratio higher or equal to 1 and were therefore predicted to be correct activity profiles. For 12%, misleading predictions were returned. Although this seems to be a slight deterioration, the enrichment within the correct targets could be seen much more clearly: more green signals were observed, resulting in a higher density, which simplifies activity profiling vitally, compare Figures 3 and 4. In the course of quantifying this observation, an increase of D3 for the best versus D3 for the fast flexible search simultaneously with an unchanged value for D4 best compared to D4 fast was defined an improvement in profile prediction for a compound switching from the fast to the best screening modus. Conversely, D3 best/D3 fast = 1 and D4 best/D4 fast > 1 indicates deterioration. Looking at our total ligand set, 32 improvements face not even half as many (15) deteriorations. When inspecting only the 85 molecules with correct profiles (from the fast and best searches), 28 of them obtained better and eight of them poorer (nevertheless, still correct) profile predictions with the best screening mode, whereas the remaining structures were not influenced.

Incorrect or lacking profiles with the fast or best flexible search have been observed for 15 compounds. When looking at the impact of switching from the fast to the best for these critical ligands, four profile prediction improvements and seven deteriorations can be seen.

Furthermore, it can be seen that the best flexible search returns activity profiles for all compounds. In this case study, it can be seen as a way of fine-tuning the approach without really compromising selectivity. We suggest the best flexible search for screening approaches of DB subsets and for cases where no or only very restricted activity profiles can be found for a compound, or where no distinct profile is achieved. The best search can always be helpful to affirm the outcome of the fast search; however, it might sometimes not be feasible because of extensive, hard manageable data output.

Guideline 2. The quality/selectivity of the pharmacophore models in the hit list is a critical point to consider, when determining the activity profile of a compound and choosing the targets for biological testing. Retrieval of a compound with a highly selective model should obviously be attributed greater significance than retrieval with a less selective model. This means that, on one hand, information on model quality must be available and, on the other hand, that preferentially selective models should serve as input for the system. Also, an automatic integration of this parameter in profiling might be possible, adding weight to targets represented by highly selective models in the pharmacophore hit list and reducing the significance of those identified by low-selectivity models.

The impact of model quality on the profile of a compound, however, varies depending on model quantity for a target. An unselective model solely representing a target will produce far more deteriorating results in PS than a single low-performing model among a set of selective models for a target. Therefore, rather, the overall performance of a hypothesis pool is vital as could be seen in a test where the least selective model of the system, generated from an HIV RT−inhibitor complex, was left out in the hit map. For the fast flexible search, no alterations of the predicted profiles could be seen and the situation D3/D4 being higher or lower than 1 remained the same for all molecules. The best flexible search revealed a profile prediction improvement for one NA inhibitor, deterioration for one RT inhibitor, and alteration from incorrect to not accomplishable profiling for one HRV coat protein targeting compound.

Guideline 3. An important aspect is the entire number of hypotheses by which a target is represented in the system, because it influences the significance of the calculated descriptor values. A high percentage of models from a specific target finding a particular ligand indicates activity at that target. If a large set of models exists for a target, this adds even more weight to that effect and offers additional security for activity profiling. For instance, 100% of 20 models finding a ligand weights heavier than 100% of 2 models. Because this effect is elusive and its actual impact hard to assess, it has not been included in the analysis process. In general, extensive retrieval of a compound for a target well-represented in pharmacophore space can be interpreted as a firm indicator for biological activity there.

Guideline 4. When inspecting the pharmacophoric and, later, the pharmacological profile obtained for a compound, detailed investigation of the returned targets can prove very helpful. (i) It is critical to know whether the targets are structurally or functionally related. Do they all belong to the same class of enzymes; for instance, are they all proteases? A ligand widely recognized by models for protease targets can very reliably be forecasted as a protease inhibitor. Thus, such a situation also raises the question of ligand selectivity, a task often difficult to handle with pharmacophore models. (ii) In very advantageous cases, the models lead to macromolecules involved in the treatment of the same diseases. The targets may, for example, belong to the same physiological pathway or cascade (like the renin−angiotensin−aldosterone system), or very rarely, the models define diverse interaction sites at the same protein. The last example in Table 2 shows such a case: An HCV polymerase inhibitor is predicted active at two different allosteric binding sites of this enzyme. This means that, although the exact interaction site for this ligand cannot be foreseen so easily, the necessity for this is not that critical here, because testing against HCV and especially against the polymerase clearly seems advisable. (iii) There is further the question of the meaningfulness of a ligand profile with regard to drug development. Compounds for application in the anti-infective sector, for example, should be checked for the desired lack of interactions with mammalian enzymes. Does the profile suggest a promiscuous compound? Does the pharmacophore hit list contain many models belonging to metabolizing proteins or antitargets indicating an unfavorable, high metabolic rate for a compound or severe side effect and probably failure in later phases of drug development?

Guideline 5. To discard the possibility that the conformer generation mode and conformer number might significantly influence the study outcome, exemplarily, five inhibitors underwent another conformational model generation. Thereby, conformer generation type specification in Catalyst was set to fast, creating a maximum of 250 and 100 conformers. Because these parameters can be expected to have an effect primarily for very flexible structures, the inhibitor with the highest number of rotatable bonds for each of the five proteins was selected. For compounds with equal flexibility, higher molecular weight was consulted for prioritization. The maximum number of rotatable bonds seen among HIV protease inhibitors was 33, for RT 11, for NA 17, for HRV coat protein 15, and for HCV polymerase 14. These 10 conformational models of the five compounds were searched with all 50 pharmacophore models using the fast as well as the best flexible search, and the hit pharmacophores were compared with the prior results (best conformer generation with a maximum of 250 conformers). Although the hereby obtained pharmacophoric profiles differed in some cases from previous signals, the pharmacological profile prediction quality—the ratio between D3 and D4—was generally not altered, indicating mostly independence of the conformer number and generation type. Similar results are also confirmed in a study by Kirchmair et al.[43] The only exception was a highly flexible HIV protease inhibitor whose former correct prediction changed to an incorrect profile for fast conformer generation of a maximum of 100 conformers when applying the fast flexible search. Except for cases where large DBs are inserted into the PS system or where calculation time is a limiting factor, preferentially highly sophisticated conformational models should be used for the input mol-

ecules. Thereby, potentially important pharmacophoric profiles are not missed because of an insufficient ligand representation.

Hit map color coding not only allows easy recognition of enrichment but also renders fast assessment of the critical areas possible. These contain ligands or hypotheses that perform considerably worse in comparison with other components of the system. Vertical inspection of the hit map shows a multitude of high-quality pharmacophores deducible by the domination of green signals in the rows. Especially, HIV protease, influenza NA, and HRV coat protein models provide hit lists comprising mainly known active inhibitors. More problematic is the situation for RT hypotheses, whose hit lists often comprise a high percentage of false positives. The latter mainly include influenza NA inhibitors, especially those from the oseltamivir type with a hydrophobic side chain, and furthermore, HRV coat protein and HIV protease inhibitors. This can at least partly be explained by the commonness of binding requirements for these sites. For instance, RT and HRV coat protein bindings are dominated by hydrophobic interactions. What puts the RT in such an unfavorable position is the fact that in many cases only few features make up the interaction pattern, such as only some hydrophobic areas and maybe one or two hydrogen bonds (usually to Lys 101).[44] Not even improvement with a shape applied for most of these models can balance their low selectivity completely. In particular, for many NA inhibitors, false recognition by RT models is the principal reason for misleading activity profiles. The situation is reversed when you inspect the hit map horizontally and look at the RT inhibitors. Their retrieval by correct models is nearly unexceptional because of the high selectivity of the other models in the system. Another area of overlap is caused by HCV polymerase allosteric site 2 models, which in some cases show considerable recognition of HIV protease inhibitors. One of them is even attributed a false activity profile because of this effect. Again, a different situation occurs the other way around because polymerase inhibitors do not seem to be found by protease inhibitor models.

Despite a variety of influencing parameters, a rapid method for activity profile estimation and validation could be developed for the antiviral compounds in this trial set. The promising outcome of this case study augurs well for the upscaling and automation of PS in a larger system and its application for fast and reliable in silico activity profiling for compounds. More validation studies are in progress, expanding targets, models, and ligands as further proofs of concept. Certainly, these studies will also address the possibilities and limitations of PS when applied for challenging cases like closely related proteins or similar ligands acting on different targets.

EXPERIMENTAL SECTION

Molecular modeling studies were carried out with prerelease versions as well as version 1.0 of the LigandScout software using an Athlon 1800 PC running Windows 2000 and with Catalyst version 4.11 on a PC with a Pentium IV processor/2.8 GHz running Fedora Core. If not mentioned otherwise in the text, default parameter settings of the programs were used.

CONCLUSIONS

The study shows the first example for a pharmacophore-based extensive PS approach aimed at the in silico activity profile prediction for antiviral compounds searched with a large set of chemical-feature-based models. Successful activity profiling was achieved in the majority of cases independent from the search algorithm. Descriptors for the analysis of output data are discussed as well as the influence of factors such as pharmacophore quality or total number of models for a target for the interpretation of an obtained activity profile. Upscaling and automation of this approach will provide a system for fast virtual PS of compounds against a variety of targets and prediction of potential biological activities, which will offer new possibilities in drug development.

ACKNOWLEDGMENT

We thank Dr. Rémy Hoffmann (Accelrys SARL, Paris) for performing the screening operations of the WDI.

Supporting Information Available: Hit maps depicted in Figures 3 and 4. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Langer, T.; Wolber, G. Pharmacophore definition and 3D searches. *Drug Discovery Today: Technol.* **2004**, *1*, 203−207.
(2) Guener, O.; Clement, O.; Kurogi, Y. Pharmacophore modeling and three-dimensional database searching for drug design using Catalyst: Recent advances. *Curr. Med. Chem.* **2004**, *11*, 2991−3005.
(3) Guener, O. F. The impact of pharmacophore modeling in drug design. *IDrugs* **2005**, *8*, 567−572.
(4) Kubinyi, H. In search for new leads. In *EFMC − Yearbook*; LD Organisation: Louvain-la-Neuve, Belgium, 2003.
(5) Böhm, H. J.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; Spektrum Akademischer Verlag GmbH: Heidelberg, Germany, 1996; pp 451−466.
(6) Schuster, D.; Langer, T. The identification of ligand features essential for PXR activation by pharmacophore modeling. *J. Chem. Inf. Model.* **2005**, *45*, 431−439.
(7) Sanguinetti, M. C.; Mitcheson, J. S. Predicting drug-hERG channel interactions that cause acquired long QT syndrome. *Trends Pharmacol. Sci.* **2005**, *26*, 119−124.
(8) Clement, O. O.; Guener, O. F. Use of pharmacophores in predictive ADME. In *Proceedings of the 3rd Pharmacokinetic Profiling in Drug Research: Biological, Physicochemical, and Computational Strategies (LogP2004, Lipophilicity Symposium)*, Zurich, Switzerland, Feb 29−Mar 4, 2004; Testa, B., Ed.; Verlag Helvetica Chimica Acta: Zurich, Switzerland, 2004; pp 381−393.
(9) Klabunde, T.; Evers, A. GPCR antitarget modeling: Pharmacophore models for biogenic amine binding GPCRs to avoid GPCR-mediated side effects. *ChemBioChem* **2005**, *6*, 876−889.
(10) Norinder, U. In silico modelling of ADMET − A minireview of work from 2000 to 2004. *SAR QSAR Environ. Res.* **2005**, *16*, 1−11.
(11) Oloff, S.; Zhang, S.; Sukumar, N.; Breneman, C.; Tropsha, A. Chemometric analysis of ligand receptor complementarity: Identifying complementary ligands based on receptor information (CoLiBRI). *J. Chem. Inf. Model.* **2006**, *46*, 844−851.
(12) Do, Q.-T.; Renimel, I.; Andre, P.; Lugnier, C.; Muller, C. D.; Bernard, P. Reverse pharmacognosy: Application of selnergy, a new tool for lead discovery. The example of e-viniferin. *Curr. Drug Discovery Technol.* **2005**, *2*, 161−167.
(13) Myers, A. G.; LaPorte, J. R.; Xing, C. Assay for identifying biological targets of polynucleotide-binding compounds. U.S. Patent 2004/248100, Dec. 9, 2004.
(14) Tulp, M.; Bohlin, L. Rediscovery of known natural compounds: Nuisance or goldmine? *Bioorg. Med. Chem.* **2005**, *13*, 5274−5282.
(15) Cleves, A. E.; Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **2006**, *49*, 2921−2938.
(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gililand, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(17) Zuckerman, A. J.; Banatvala, J. E.; Pattison, J. R.; Griffiths, P.; Schoub, B. *Principles and Practice of Clinical Virology*, 4th ed.; John Wiley & Sons Ltd: Chichester, U. K., 2000.
(18) Stolk, L. M. L.; Lueers, J. F. J. Increasing number of anti-HIV drugs but no definite cure. Review of anti-HIV drugs. *Pharm. World Sci.* **2004**, *26*, 133−136.
(19) Tomasselli, A. G.; Heinrikson, R. L. Targeting the HIV-protease in AIDS therapy: A current clinical perspective. *Biochim. Biophys. Acta* **2000**, *1477*, 189−214.
(20) Jonckheere, H.; Anne, J.; De Clercq, E. The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.* **2000**, *20*, 129−154.
(21) Moscona, A. Neuraminidase inhibitors for influenza. *N. Engl. J. Med.* **2005**, *353*, 1363−1373.
(22) Zhang, J.; Xu, W. Recent advances in anti-influenza agents with neuraminidase as target. *Mini-Rev. Med. Chem.* **2006**, *6*, 429−448.
(23) Ruuskanen, O.; Hyypia, T. Rhinovirus: Is it really a relevant pathogen? In *Microbe-Host Interface in Respiratory Tract Infections*; Ramilo, O., Ed.; Horizon Bioscience: Wymondham, U. K., 2005; pp 291−317.
(24) Wang, M.; Ng, K. K. S.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bedard, J.; Morin, N.; Nguyen-Ba, N.; Alaoui-Ismaili, M. H.; Bethell, R. C.; James, M. N. G. Non-nucleoside analogue inhibitors bind to an allosteric site on HCV NS5B polymerase. *J. Biol. Chem.* **2003**, *278*, 9489−9495.
(25) Pfefferkorn, J. A.; Greene, M. L.; Nugent, R. A.; Gross, R. J.; Mitchell, M. A.; Finzel, B. C.; Harris, M. S.; Wells, P. A.; Shelly, J. A.; Anstadt, R. A.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481−2486.
(26) Di Marco, S.; Volpari, C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfi, A. Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **2005**, *280*, 29765−29770.
(27) Condon, S. M.; LaPorte, M. G.; Herbertz, T. Allosteric inhibitors of hepatitis C NS5B RNA-dependent RNA polymerase. *Curr. Med. Chem.* **2005**, *4*, 99−110.
(28) Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 160−169. LigandScout available from Inte:Ligand GmbH, www.inteligand.com/ligandscout.
(29) Ekins, S.; Erickson, J. A. A pharmacophore for human pregnane X receptor ligands. *Drug Metab. Dispos.* **2002**, *30*, 96−99.
(30) Derwent World Drug Index. http://www.derwent.com/ (accessed Jul 2006).
(31) De Clercq, E. Highlights in the development of new antiviral agents. *Mini-Rev. Med. Chem.* **2002**, *2*, 163−175.
(32) De Clercq, E. Antiviral drugs in current clinical use. *J. Clin. Virol.* **2004**, *30*, 115−133.
(33) De Lucca, G. V.; Lam, P. Y. S. De novo design, discovery and development of cyclic urea HIV protease inhibitors. *Drugs Future* **1998**, *23*, 987−994.
(34) Ren, J.; Nichols, C.; Bird, L.; Chamberlain, P.; Weaver, K.; Short, S.; Stuart, D. I.; Stammers, D. K. Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation nonnucleoside inhibitors. *J. Mol. Biol.* **2001**, *312*, 795−805.
(35) Balzarini, J. Current status of the nonnucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr. Top. Med. Chem.* **2004**, *4*, 921−944.
(36) Chand, P.; Kotian, P. L.; Morris, P. E.; Bantia, S.; Walsh, D. A.; Babu, Y. S. Synthesis and inhibitory activity of benzoic acid and pyridine derivatives on influenza neuraminidase. *Bioorg. Med. Chem.* **2005**, *13*, 2665−2678.
(37) Alymova, I. V.; Taylor, G.; Portner, A. Neuraminidase inhibitors as antiviral agents. *Curr. Drug Targets: Infect. Disord.* **2005**, *5*, 401−409.
(38) Hadfield, A. T.; Diana, G. D.; Rossmann, M. G. Analysis of three structurally related antiviral compounds in complex with human rhinovirus 16. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14730−14735.
(39) Oren, D. A.; Zhang, A.; Nesvadba, H.; Rosenwirth, B.; Arnold, E. Synthesis and activity of piperazine-containing antirhinoviral agents and crystal structure of SDZ 880-061 bound to human rhinovirus 14. *J. Mol. Biol.* **1996**, *259*, 120−134.
(40) Chapman, M. S.; Minor, I.; Rossmann, M. G.; Diana, G. D.; Andries, K. Human rhinovirus 14 complexed with antiviral compound R 61837. *J. Mol. Biol.* **1991**, *217*, 455−463.
(41) *Pipeline Pilot*, version 5.0.1.100; Scitegic: San Diego, CA, 2006.
(42) Willet, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(43) Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T. Comparative analysis of protein-bound ligand conformations with respect to Catalyst's conformational space subsampling algorithms. *J. Chem. Inf. Model.* **2005**, *45*, 422−430.
(44) Barreca, M. L.; Rao, A.; De Luca, L.; Zappala, M.; Monforte, A.-M.; Maga, G.; Pannecouque, C.; Balzarini, J.; De Clercq, E.; Chimirri, A.; Monforte, P. Computational strategies in discovering novel nonnucleoside inhibitors of HIV-1 RT. *J. Med. Chem.* **2005**, *48*, 3433−3437.

CI6002043