Parallel Screening and Activity Profiling with HIV Protease Inhibitor Pharmacophore Models

Theodora M. Steindl,*[,†,‡] Daniela Schuster,[†] Christian Laggner,[†] Karen Chuang,[§] Rémy D. Hoffmann,[#] and Thierry Langer*[,†,‡]

Institute of Pharmacy, Computer Aided Molecular Design Group, University of Innsbruck, Innrain 52c and Center for Molecular Biosciences Innsbruck (CMBI), Peter-Mair-Strasse 1, A-6020 Innsbruck, Austria, Inte:Ligand GmbH, Clemens Maria Hofbauer-Gasse 6, A-2344 Maria Enzersdorf, Austria, Accelrys Inc., 10188 Telesis Court, Suite 100, San Diego, California 92121, and Accelrys SARL, Parc Club Orsay Université, 20, rue Jean Rostand, 91898 Orsay, Cedex, France

Received July 28, 2006

Parallel Screening has been introduced as an in silico method to predict the potential biological activities of compounds by screening them with a multitude of pharmacophore models. This study presents an early application example employing a Pipeline Pilot-based screening platform for automatic large-scale virtual activity profiling. An extensive set of HIV protease inhibitor pharmacophore models was used to screen a selection of active and inactive compounds. Furthermore, we aimed to address the usually critically eyed point, whether it is possible in a parallel screening system to differentiate between similar molecules/molecules acting on closely related proteins, and therefore we incorporated a collection of other protease inhibitors including aspartic protease inhibitors. The results of the screening experiments show a clear trend toward most extensive retrieval of known active ligands, followed by the general protease inhibitors and lowest recovery of inactive compounds.

INTRODUCTION

The application of pharmacophore models is a common technique in drug discovery. The classical pharmacophore approach has recently been extended by the systematic usage of a collection of pharmacophores to determine the overall retrieval of a compound by different models. We refer to this strategy as parallel screening (PS) which means the simultaneous screening of a compound or a set of compounds against a pharmacophore series often representing different target proteins. The aim is the fast in silico determination of the biological activity profile of a molecule in order to speed up the time and cost-intensive drug discovery development process and increase its efficiency. This activity profiling concept includes several steps: From screening a compound with a set of pharmacophore models, the pharmacophoric profile—a hit list of mapping hypotheses—is received. In the next step the pharmacological targets represented by these models are identified giving the pharmacological profile. This allows then predictions on the potential biological properties of the compound—the biological profile. A first manually accomplished broad proof of principle study for the PS concept has been carried out and published only recently.[1] Thereby, 100 antiviral compounds were screened against 50 pharmacophore models representing five diverse viral targets. We could show that correct activity profiles were predicted for the majority of the ligands. The promising outcome affirmed our aim to further pursue and develop the PS concept. Thereby the primary criteria for judging profile prediction accuracy (correct/incorrect) was the ratio between the retrieved hit pharmacophores for the correct target (i.e., the target at which the compound is a known inhibitor) and the hit pharmacophores representing the most frequently hit false target (based on the assumption that a compound is inactive on all but the correct target). Aside from this first parameter for result analysis also the impact of model selectivity, model quantity for a particular target, the screening algorithm, the conformer generation mode, and the nature of the retrieved targets were investigated.

A limitation of this early investigations was the lack of automation, which made manual data generation, visualization, and analysis a highly time-consuming task. So an important point within the present study was to investigate: (i) Is PS also able to provide the expected results when possible influence factors, e.g., caused by user adjustments, are ruled out via automation of several steps? (ii) And can sensible visualization strategies be found? Our PS approach was therefore realized within the Pipeline Pilot environment employing an automated screening platform.[2,3] The latter uses the Inte:Ligand pharmacophore database, as will be explained in more detail later on.

A second major factor for investigation was the performance of PS for a set of structurally or functionally related molecules. Will the system be able to provide a notably higher enrichment for known active compounds for a particular target or will similar molecules or compounds

* Corresponding author addresses: Institute of Pharmacy, Department of Pharmaceutical Chemistry, Innrain 52c, University of Innsbruck, A-6020 Innsbruck, Austria (T.L.). Corresponding author phone: +43 512 507 5252; fax: +43 512 507 5269; e-mail: thierry.langer@uibk.ac.at. (T.L.) and e-mail: theodora.steindl@uibk.ac.at (T.M.S.).

† University of Innsbruck and Center for Molecular Biosciences Innsbruck (CMBI).

‡ Inte:Ligand GmbH.

§ Accelrys Inc.

Accelrys SARL, Parc Club Orsay Université.

10.1021/ci600321m CCC: $37.00
Published on Web 01/17/2007

acting on related target proteins obtain the same activity profiles? For example can inhibitors of a particular protease be separated from other protease inhibitors? Since for related molecules selectivity profiles do not always enable a clear distinction between active and completely inactive—we often face a continuum for a data set—one has to act with caution when analyzing the results. In other words, for structurally or functionally related molecules we want similar but not as intensive or dominant activity profiles as for known active structures, but we expect clearly higher retrieval than for inactive structures.

The selection of HIV protease as target for this study was based on several criteria: The currently available Inte:Ligand pharmacophore database comprises a large number of models for this enzyme. Furthermore, extensive information on ligands and activities are available as well as knowledge of the binding characteristics of inhibitors. Proteases seemed to provide a good example to study the ability of the system to distinguish between different classes of protease inhibitors and to enrich compounds with known activity for the desired enzyme subtype. Choosing only a single target for this test run we hoped to avoid too large data volumes that are difficult to manage and interpret.

As input molecules we created four approximately equally sized data sets to undergo PS in our system. In order to avoid biasing of the system, the literature was searched for appropriate input data sets. A set of known HIV protease inhibitors, a group of other protease inhibitors, and two sets of inactive compounds—one derived from a literature reference, the other one randomly extracted from a large virtual druglike library—were selected. After screening with the HIV protease inhibitor pharmacophore collection the obtained profiles of the data sets were analyzed for the expected trend: 1. higher enrichment for known active ligands, 2. followed by the related protease inhibitors, and 3. lowest retrieval for the inactive structures. Analysis and visualization options within the screening platform as well as with the aid of MS Excel tables and diagrams are presented.

STUDY DESIGN

1. Pipeline Pilot - Automated PS Platform. We used a newly developed protocol for fast automatic PS of compounds against a set of pharmacophore models aimed at the determination of activity profiles for the input molecules. The so-called screening platform is a Pipeline Pilot WebPort[2] application of Catalyst software components and accepts compounds either from an SD file or an internal database from the program Catalyst.[4] The included pharmacophore models with which compounds can be screened are managed in the so-called pharmacophore database. Pharmacophore hypotheses that should be screened are selectable. The current pharmacophore database consists of over 1500 structure-based pharmacophore models provided by Inte:Ligand, built with the LigandScout software and stored in Catalyst pharmacophore data format. It covers approximately 170 targets and over 20 therapeutic classifications.[5] For detailed information on the structure-based LigandScout-Catalyst pharmacophore model generation process see refs 1 and 6−8. This data volume in the pharmacophore database is constantly growing with the aim of being able to offer a broad coverage of pharmacologically relevant targets and indications. The open platform design also supports insertion of the user's own models into the system. Input molecules may be screened against the whole collection of hypotheses or against a reduced selection which can for example represent important targets for a particular therapeutic classification.

Visualization of the screening results is realized in the screening platform through (i) a heatmap, (ii) a more detailed ligand profile display, and (iii) an Excel export table for postprocessing. In a so-called heatmap the retrieved hits are displayed against the corresponding hypotheses and highlighted in a color-coded manner according to their fitting on the model. This latter information, the score, is based on the calculation of normalized fit values, i.e., the calculated fit (how well are the model features mapped by the ligand) divided by the maximum fit. The maximum fit value equals the number of pharmacophore features, since this automatic screening approach utilizes the Catalyst CatSearch module, which requires that all features of a hypothesis have to be mapped by a hit molecule. Consequently, the maximum score (normalized fit) for a compound is one. The second depiction mode shows a more detailed protocol. Hit molecules and their fitting to the models can be viewed together with the model score. Furthermore, the outcome of a screening run can be written into an Excel table, showing molecule name, molecular weight (MW), molecular formula, mapping hypothesis, fit value, and normalized fit value. These raw data can then be further processed for example to obtain a different display style. A detailed description of the screening platform methodology will be published elsewhere. Examples showing the preliminary graphical user interface and result visualization can be viewed in Figure 1 as well as in the Results and Discussion section.

2. Pharmacophore Models. A total of 81 HIV protease inhibitor models that are currently available within the pharmacophore collection were used for this study. The hypotheses were originally generated in a structure-based approach from 47 crystal structure complexes from the Brookhaven Protein Databank (PDB).[9] Looking at the models in more detail shows that 45% of them contain so-called excluded volume spheres and/or shapes, which limit the spatial extension of a compound and allow for increased steric restriction. The fact that some PDB entries serve as source for more than one hypothesis results from the generation of shape models in addition to the original models for even higher refinement. Furthermore, in some cases a ligand is involved in too many interactions to be included in a single model: Catalyst does not support multiple features on one functional group or atom (e.g., an amine representing a positive ionizable feature as well as a hydrogen bond donor) which in some cases required the generation of more than one model for a complex. Highly selective models focused on giving only a few hits, thereby avoiding false positives, are present as well as more general hypotheses providing a good universal display of features required for anti-HIV protease activity. Selectivity and restriction were judged originally upon model generation from the outcome of screening experiments of the Derwent World Drug Index (WDI),[10] which contains over 63 000 drugs and other bioactive molecules. Thereby, the hit lists were inspected for the occurrence of known HIV protease inhibitors—the entire WDI includes approximately 170 HIV protease inhibitors according to 1D data information—and for the total

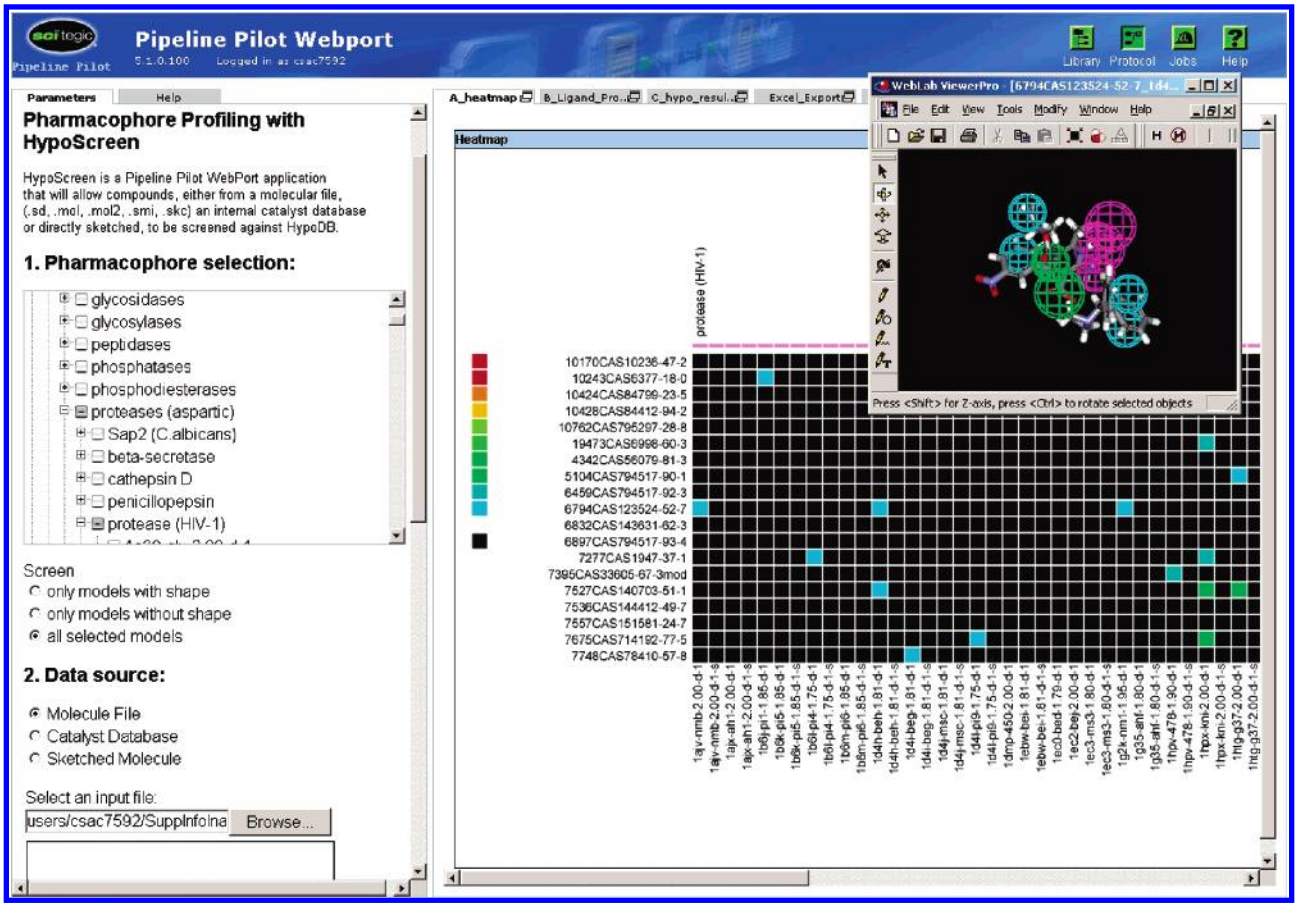


Figure 1. Left: graphical user interface showing pharmacological targets, selectable pharmacophores, and in- and output determinants. Right: screening results for one of the inactive data sets in a heatmap with hit ligands (vertically) and mapping models (horizontally), color coding from red boxes for high score compounds to light blue boxes for low score compounds, black boxes where compounds are not found by the models. Top right: image of a ligand mapped to a hypothesis.

amount of hits, respectively. Thereby, 94% (76 hypotheses) find less than 500 and 2.5% (2 hypotheses) over 2000 WDI hits. Examples of these hypotheses are shown in Figure 2. Despite differences in restrictivity all models are designed to show typical anti-HIV protease features. It is assumed that such a pool of hypotheses applied in a parallel way is well suited to identify activity profiles. Using only one model or restricting the screening only to very selective models would mean the loss of a lot of information and would not give such a good overall view of the data set properties.

3. Input Molecules. Four input data sets of approximately equal size were prepared for the PS approach. From a recently published study by Meagher et al. 89 known HIV protease inhibitors and 85 druglike inactive compounds were gained.[11−14] For further affirmation of the results, another set of presumably inactive molecules was taken from a virtual library of over 12 000 druglike diverse compounds generated with the software tool ilib diverse as described in detail elsewhere.[7,15−18] Focusing on desired conformity with the active ligands with regard to MW and number of rotatable bonds, 80 structures of this library were randomly chosen (average MW of 490 and average number of rotatable bonds of 16 versus 480 and 20, respectively, in the actives set). The fourth set of molecules comprises 79 inhibitors of other proteases referenced in a review by Leung et al.[19] These consist of ligands for all four major classes of protease enzymes: serine (24 molecules), cysteine (20 molecules), metallo (16 molecules), and most importantly the aspartic proteases (19 molecules), of which HIV protease is a member. This latter group includes renin, plasmempsin, cathepsin D, and secreted aspartic protease ligands. Moderate anti-HIV protease activity cannot be ruled out or is even known for some of these other protease inhibitors since not all of them show highly distinct selectivity profiles—a fact that must be taken into consideration when inspecting the results. General inhibitors like pepstatin A[20−22] **4** (Chart 1) were deliberately included. Whereas peptidic or peptidomimetic scaffolds provoke an overall structural similarity among the majority of all protease inhibitors in this study, especially other aspartic protease inhibitors can be considered most closely related to the group of active compounds.

In order to estimate the structural similarity of the HIV and other protease inhibitors used for this study and to eliminate the possibility that simple 2D similarity analysis could also discriminate between these two data sets, similarity calculations between the 168 compounds were performed. For this task, we used Pipeline Pilot which contains a component to cluster molecules based on Tanimoto similarities between user-selected descriptor sets.[2,23] As 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 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

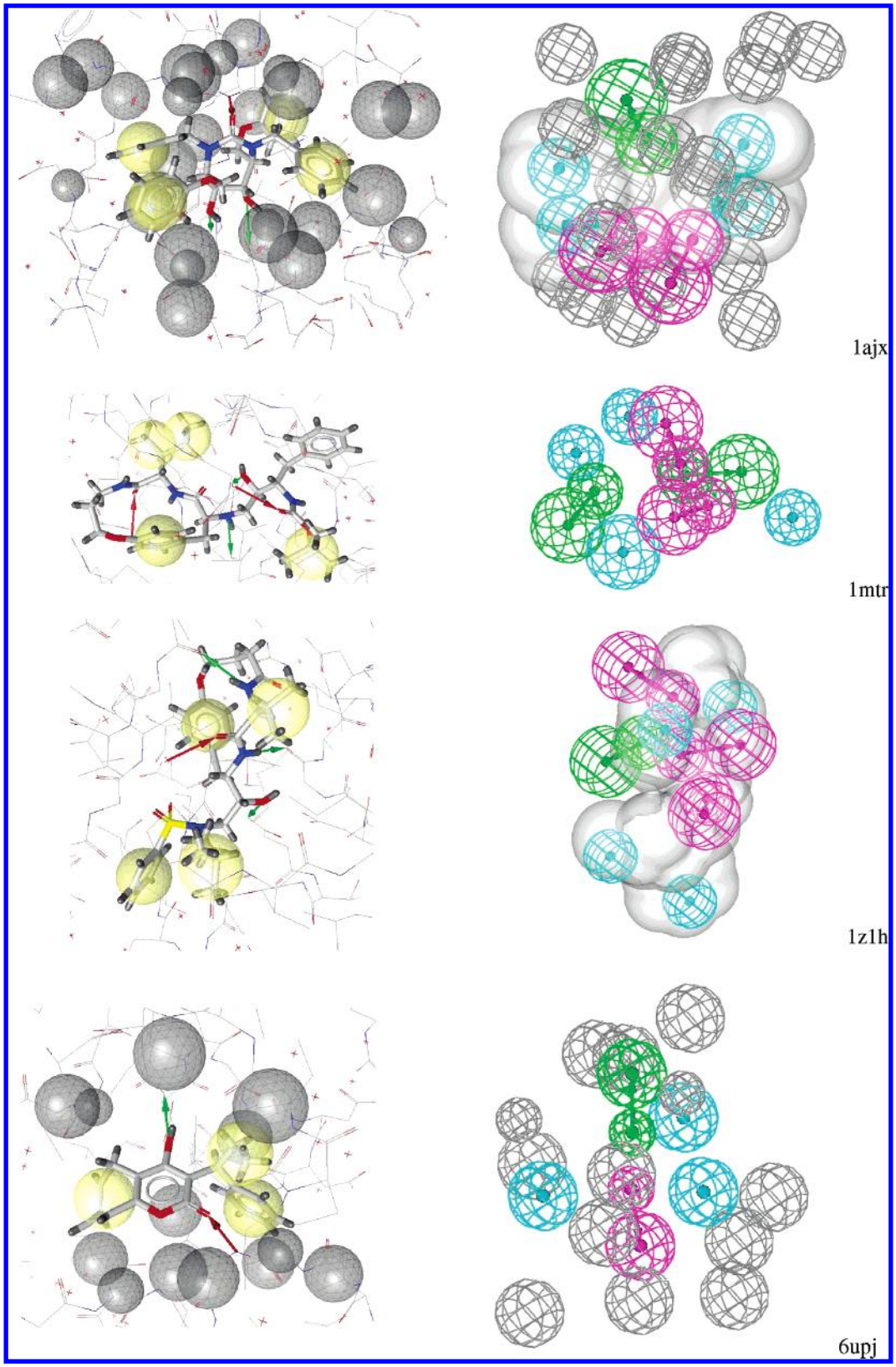
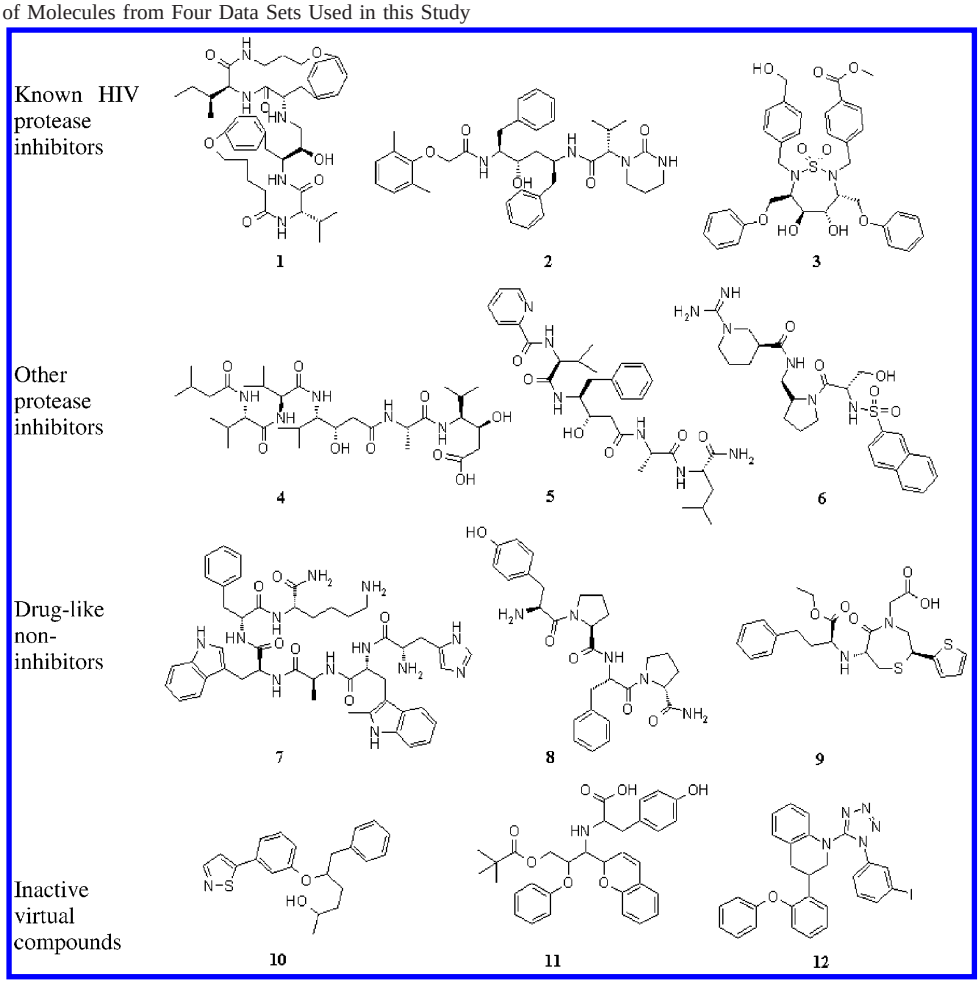


Figure 2. Examples of structure-based HIV protease inhibitor pharmacophore models derived from various PDB complexes present in the pharmacophore database. Left column: LigandScout pharmacophores with cocrystallized ligand. Right column: Resulting Catalyst pharmacophores. Four-letter PDB code on the side of the models. Feature color coding LigandScout: green: hydrogen bond donor; red: hydrogen bond acceptor; yellow: hydrophobic; black: excluded volume sphere. Feature color coding Catalyst: magenta: hydrogen bond donor; green: hydrogen bond acceptor; cyan: hydrophobic; black: excluded volume sphere; gray: shape.

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 rather tolerant value of 0.7 employed for the maximum distance, we obtained 38 clusters. This high

Chart 1. Examples of Molecules from Four Data Sets Used in this Study



number of clusters gives evidence of sufficient ligand diversity. Looking at the different clusters in more detail confirmed that most of them contain functionally diverse members: HIV protease inhibitors, other aspartic, serine, cysteine, and/or metallo protease inhibitors mixed in the same cluster. Exceptional in this context was one cluster made up solely of six HIV protease inhibitors with cyclic urea scaffolds. Whereas the two largest clusters contain 19 ligands, 12 groups are formed only by single compounds. Applying the same similarity calculation procedure could then show sufficient diversity and ligand mixture in clusters for all four data sets as well as for an analysis of the HIV protease inhibitor data set and the HIV protease inhibitors from the PDB complexes used for pharmacophore model design. These results confirm the validity of our pharmacophore approach.

These four compound sets—known actives, druglike non-inhibitors, inactive virtual compounds from ilib diverse, and other protease inhibitors—provided the input data for our PS approach with 81 HIV protease inhibitor pharmacophore models. Examples of studied molecules are shown in Chart 1.

RESULTS AND DISCUSSION

The four input data sets for this study were built in Catalyst, energy minimized, saved as single conformer SD files, and submitted to the screening platform. From the pharmacophore collection all currently included HIV protease inhibitor pharmacophore models were selected. The molecules underwent conformational model generation and screening according to the default settings of the protocol (fast conformer generation with a maximum of 255 conformers, fast flexible search algorithm for screening). Approximate calculation time was 0.2 s per molecule and hypothesis. For a fast and easy visual analysis the results of these four screening runs were inspected within the heatmap depiction as well as in the more detailed protocol (see examples in Figures 3 and 4). Already from the size of the heatmaps it becomes clear that the active compound set is found most extensively by the pharmacophores followed by the other protease inhibitors and the two inactive sets, for which the number of retrieved compounds as well as the number of hit pharmacophores decreases. Inspecting the heatmaps vertically reveals several frequently hitting pharmacophores that, according to the colored boxes, map significantly more of the hit structures than other models from the collection, indicating more general models.

For numerical analysis of the outcome we inspected the data written into an Excel report table (Figure 5). Although the calculated and normalized fit values can provide very helpful information, they were not considered here. We only focused on the point whether a molecule is found or not by a particular pharmacophore hypothesis and therefore obtains an activity profile for potential HIV protease inhibition. This simple yes or no decision philosophy allows a fast determination of the overall performance of the screening experiments with regard to the general activity profiles returned

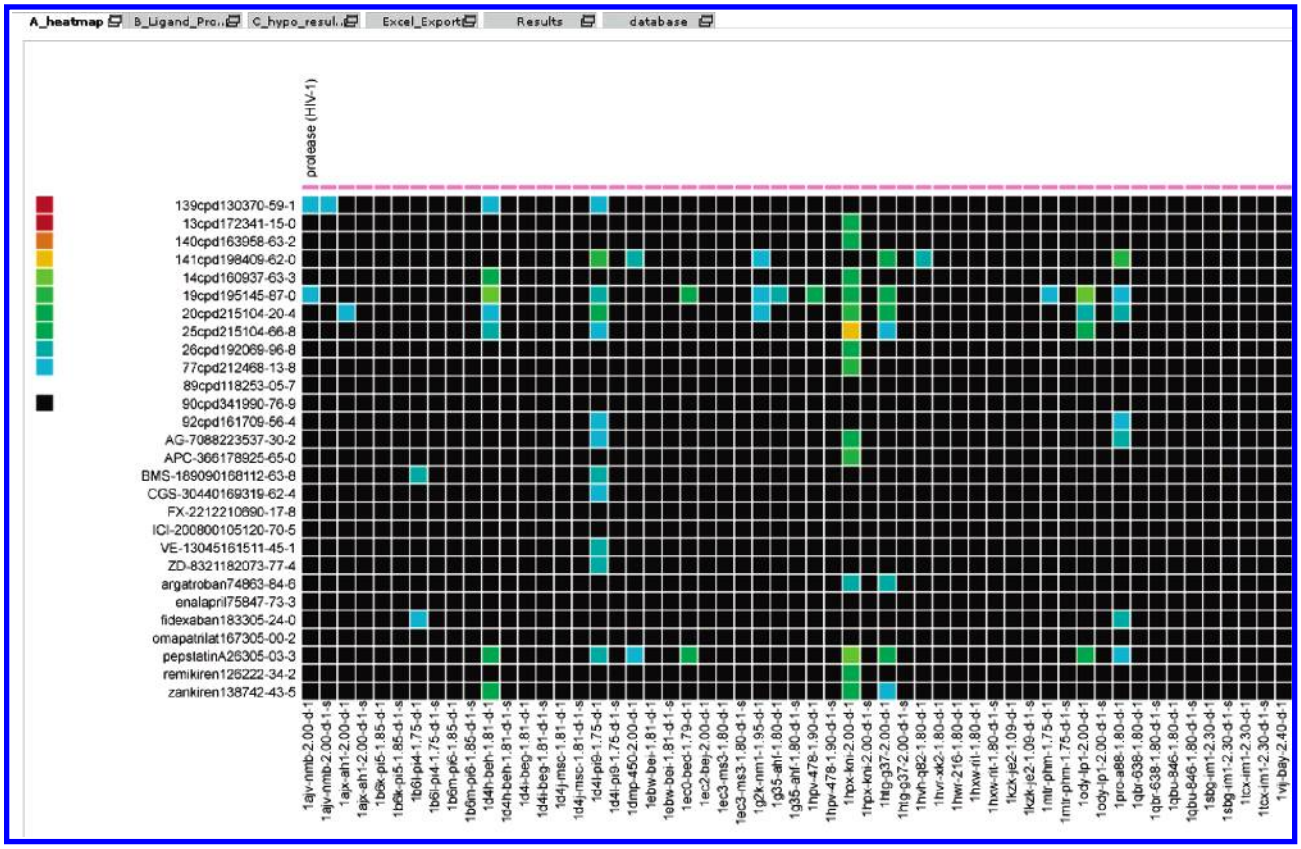


Figure 3. Overall display of the screening results for the other protease inhibitor set in the heatmap mode (detailed description in Figure 1).

for the four molecule collections. Also outstanding profiles of single ligands or the performance of a particular model can thus be investigated. Even easier interpretability than from the Excel table can be achieved by converting the information into diagrams (Figure 6). The trend that known anti-HIV compounds get more intense profiles by the HIV protease pharmacophores than the group of other protease inhibitors or inactive compounds is clearly observable from three curve parameters: (i) the decrease of curve height from one group to the next, (ii) the decreasing number of models producing these profiles, and (iii) the decreasing number of retrieved molecules (or rather the percentage of retrieved molecules because of the slightly differing test set component numbers).

HIV Protease Inhibitors Set. Looking at the overall profile of the group of known active compounds shows that 59 out of the 89 ligands (i.e., 66%) were retrieved by one or several models. Three-quarters of the total pharmacophores used (61 hypotheses) retrieved hits from the active compound group. When adding up how often a match occurs between the 59 retrieved structures and the 61 mapping hypotheses, you come up with a total of 263 mappings for the PS run of this first data set. The second criterion for analysis of profile attribution—the curve height—is represented by the ligand-(s) found most often by different models. In our study ligand **1** (Chart 1) was hit by 21 hypotheses (26% of maximal possible hit model number of 81). Eight other active molecules were also hit by 10 or more models. The average hit pharmacophore number for a mapped ligand is 4.5. Out of the 61 fitting models each of them occurs on average 4.3 times in the hit structure profiles.

Other Protease Inhibitors Set. As could be expected, the second most intensive anti-HIV profile was displayed for the group of other protease inhibitors. Of these 79 structures 28 (35%) were found 72 times by 20 different hypotheses (25%) with a maximum retrieval of two compounds which were found seven times each (9%). Deducible from these data, a hit pharmacophore will occur approximately 3.6 times, and a hit ligand will be found on average by 2.6 models.

Inactive Compounds Sets. The first group of inactive molecules, the druglike noninhibitors, performed as follows: 31 screening hits divide into 14 hitting models (17%) and 15 mapped molecules (18%). Two structures were hit by a maximum of five pharmacophores (6%). From the virtual ilib diverse set no more than 8% were displayed as hits by only four hypotheses (5%) and a highest occurrence of a molecule of two times (2%). An explanation for the higher retrieval of the compounds of the first group can be given from their overall larger structures (compared to the virtual molecules). Since they mainly present natural compounds or thereof derived ligands, many functional groups are available to fulfill the pharmacophore requirements. Besides, several large peptides or peptidomimetic structures are present in the set. The assumption that many of these hits are only found because of their size is confirmed by the fact that none of the hit pharmacophores contains a shape as steric restriction.

These results are summarized in Table 1: Explicitly, known active compounds performed best in the PS approach with the highest compounds (column 2) and pharmacophores (column 3) retrieval rates, highest average hit model number

Figure 4. Display of the screening results for the active compound set in the more detailed pharmacophore profile mode: 2D hit compound structure (left), 3D hit compound structure and pharmacophore mapping (middle, selectable), and bar chart showing hit models, their score, and model information (right).

	A	B	C	D	E	F
1	Name	Fit_Value	FIT_VALUE_NORMALIZED	hypo	MolWt	MolFormula
2	Molecule 101	3.19063	0.638126	6upj-niu-2.34-d-1	311.38	C19H21NO3
3	Molecule 101	3.19063	0.531771667	6upj-niu-2.34-d-1-s	311.38	C19H21NO3
4	Molecule 124	0.824713	0.1649426	6upj-niu-2.34-d-1	564.68	C35H36N2O5
5	Molecule 142	2.117	0.352833333	2AID-THK-1.9	504.711	C32H44N2O3
6	Molecule 177	0.468788	0.0937576	6upj-niu-2.34-d-1	339.451	C20H21NO2S
7	Molecule 177	0.468788	0.078131333	6upj-niu-2.34-d-1-s	339.451	C20H21NO2S
8	Molecule 172	1.8424	0.307066667	1hpx-kni-2.00-d-1	518.608	C30H34N2O6
9	Molecule 138	0.0429533	0.00859066	6upj-niu-2.34-d-1	548.68	C35H36N2O4

Figure 5. MS Excel report of the screening results for the inactive ilib diverse compound set.

per found inhibitor (column 5), and highest average (column 6) as well as total (column 4) attribution of different pharmacophores to a hit structure. The trend in column 4 is also detectable from curve peaks in Figure 6. Moving on from the actives to the other three data sets, a decrease of all these calculated description values can be observed, whereby for the related group of other protease inhibitors still reasonable retrieval could be achieved.

Several molecules as well as pharmacophores give outstanding results and deserve more detailed examination: Most noticeable is certainly the top retrieval of HIV protease inhibitor **1** by a total of 21 models. This good performance can only partly be explained by its large size and its many functional groups that allow mapping on many different pharmacophores, since the set also contains many other large peptidic/peptidomimetic and even closely related cyclic structures. As can be expected for such a sterically demanding structure, three-quarters of the mapped pharmacophores do not contain a shape feature for steric restriction. Further frequent active hitters (found with more than 10 models) contain peptidomimetic structures like lopinavir **2** or the cyclic urea inhibitor **3**. An inhibitor of the *Plasmodium falciparum* aspartic protease plasmempsin **5** derived from the general aspartic protease inhibitor pepstatin A **4** as well as a metalloprotease inhibitor were the most frequently (seven times each) found compounds from the related data set. From the druglike noninhibitors a large macrocyclic structure (MW above 900) and an equally large oxytocin analogue **7** were each found by a maximum of five models. Again none of the shape models is among the hit pharmacophores for these compounds. The situation is reversed for the inactive ilib diverse structures, where hits were considerably smaller. Compound **10** for example was found by a model from a PDB complex of HIV protease with a small, nonpeptidic inhibitor (PDB entry 6upj, Figure 2).

Furthermore, model selectivity in this PS system was investigated. We identified several hypotheses that were frequent hitters across the different data sets. This observation

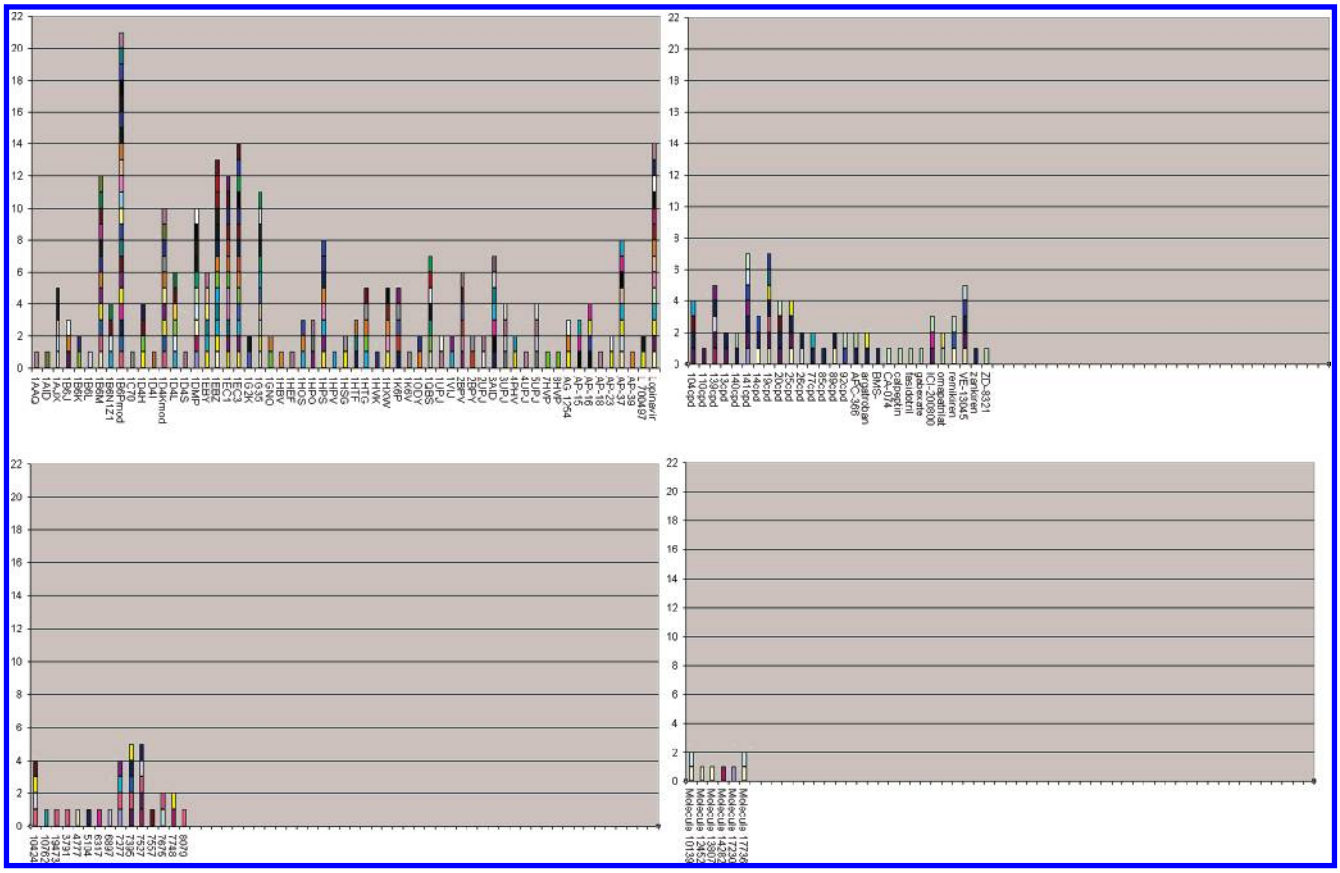


Figure 6. Diagram display of the screening results for the data set of known HIV protease inhibitors (upper left), other protease inhibitors (upper right), drug-like non-inhibitors (lower left), and inactive virtual compounds (lower right). Hit molecules are displayed on the *x*-axis, number of the mapped hypotheses on the *y*-axis.

Table 1. Numerical Analysis of the Screening Results of the Four Data Sets

	%age of hit molecules from entire data set	no./%age of hit models for all retrieved hit molecules	no./%age of hit models for best retrieved hit molecules	av hit model no. for all retrieved molecules	av occurrence of a hit model in hit molecules profiles
known HIV protease inhibitors	66	61/75	21/26	4.5	4.3
other protease inhibitors	35	20/25	7/9	2.6	3.6
druglike noninhibitors	18	14/17	5/6	2.1	2.2
inactive virtual compounds	8	4/5	2/2	1.3	2

points toward low selectivity in cases where active and inactive molecules are found. When HIV and other protease inhibitors are found, general protease inhibitor recognition is achieved by a particular model. As mentioned before, since especially some of the other aspartic protease inhibitors are known to also act on HIV protease, retrieval by a model does not necessarily prove a lack of anti-HIV selectivity. Highest hit rates from all four data sets were achieved by the obviously low selectivity models 1hpx and 6upj gained from crystal structures with a larger peptidomimetic and a small, nonpeptidic HIV protease inhibitor, respectively. Models providing good selectivity mainly for the two protease inhibitor groups are 1d4l or 1mtr. Clear domination of known active compound recognition was for example observed for models from PDB entries 1ajx, 1ody, and 1z1h. Such information is valuable for further application of the system, e.g., when choosing compounds for biological testing. Examples of striking models mentioned in this paragraph are shown in Figure 2. In all screening experiments there was no hypothesis that returned more hits from one of the inactive compounds sets than from the known HIV protease inhibitors.

We also attempted to investigate the reasons, why the pharmacophore set failed to identify 30 known active compounds out of the group of 89 HIV protease inhibitors. Visual inspection of these compounds which wrongly did not obtain anti-HIV activity profiles showed no outstanding differences from the retrieved active structures (e.g., completely different structural compound classes, etc.). Neither did property comparisons (number of rotatable bonds, MW, number of hydrogen bond donors and acceptors, number of hydrophobic features, steric requirements) offer an explanation. Therefore, most likely conformational model generation mode and the screening algorithm are major factors. In order to confirm this theory the missed inhibitors were provided with refined conformational models (best quality, maximum number of conformers 250, and an energy range of 20 kcal/mol above the energy minimum) in Catalyst and underwent screening experiments. The model obtained from PDB complex 1ody which had demonstrated reasonable HIV

protease selectivity alone returned 25 of the 30 missed ligands. A combination of the lower selectivity models 1hpx and 6upj provided a hit list of 28 structures. Obviously the ligands do not possess structural characteristics preventing them from being found by the utilized model set. The incapability of the automatic PS approach to identify them rather seems to be caused by insufficient coverage of conformational space using the default parameters.

EXPERIMENTAL SECTION

Molecular modeling studies were carried out with the prerelease version dated June 9th 2006 of a parallel screening software, a Pipeline Pilot WebPort application, using an Athlon 1800 PC running Windows 2000 as a Pipeline Pilot WebPort client and a Pentium IV 2.8 GHz PC running Fedora Core 4 with Pipeline Pilot 5.1 and Catalyst version 4.11 as the server machine. If not mentioned otherwise in the text, default parameter settings of the programs were used. For molecules without indicated stereo information racemates were used. The overall speed of the protocol was 0.2 s per molecule and hypothesis.

CONCLUSIONS

The successful application of the PS concept is shown for a large collection of HIV protease inhibitor pharmacophore models that are used to screen known active, known inactive, and closely related presumably inactive compounds. The critical step in this study was to set up a procedure for switching from the so far manually carried out experiments to a fully automated large-scale system for PS and visualization. Another significant point emerging within the work is that we could give evidence that also for compounds closely related in structure or function correct profiles and clear trends can be achieved using our screening method. These promising results give encouragement for further studies and confirm that successful in silico activity profiling is possible in a PS approach. Thus the utilized collection of pharmacophores can be seen as a valuable tool for lead identification of potential new HIV protease inhibiting substances.

ACKNOWLEDGMENT

We thank Konstantin Poptodorov (Accelrys, Cambridge, U.K.), Jason Benedict, and Samuel Toba (Accelrys San Diego) for their support with establishing the screening protocol.

REFERENCES AND NOTES

(1) Steindl, T.; Schuster, D.; Laggner, C.; Langer, T. Parallel Screening: A Novel Concept in Pharmacophore Modelling and Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 2146−2157.

(2) SciTegic Inc. and Accelrys Inc., Pipeline Pilot (TM), San Diego. http://www.scitegic.com (accessed Jan 2007).

(3) Steindl, T. M.; Langer, T. Pharmacophore-based Parallel Screening. Presented at the *9th European Catalyst User Group Meeting*, Frankfurt, Germany, March 23, 2006. http://www.accelrys.com/products/catalyst/usergroup/2006/EU/slides/steindl.pdf (accessed Nov 20, 2006)

(4) *Accelrys Inc., Catalyst, version 4.11*; San Diego, CA, 2005. http://www.accelrys.com (accessed Jan 2007).

(5) 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. Available from Inte:Ligand GmbH, www.inteligand.com/ligandscout.

(6) Krovat, E. M.; Fruehwirth, K. H.; Langer, T. Pharmacophore Identification, in Silico Screening, and Virtual Library Design for Inhibitors of the Human Factor Xa. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 146−159.

(7) Schuster, D.; Laggner, C.; Steindl, T. M.; Palusczak, A.; Hartmann, R. W. et al. Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. *J. Chem. Inf. Model.* **2006**, *46*, 1301−1311.

(8) Schuster, D.; Laggner, C.; Steindl, T. M.; Langer, T. Development and validation of an in silico P450 profiler based on pharmacophore models. *Curr. Drug Discovery Technol.* **2006**, *3*, 1−48.

(9) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(10) Derwent World Drug Index. http://www.derwent.com/ (accessed Jan 2007).

(11) Meagher, K. L.; Lerner, M. G.; Carlson, H. A. Refining the Multiple Protein Structure Pharmacophore Method: Consistency across Three Independent HIV-1 Protease Models. *J. Med. Chem.* **2006**, *49*, 3478−3484.

(12) *Comprehensive Medicinal Chemistry Database (MDL)*; MDL Information Systems Inc.: San Leandro, CA, 2003.

(13) Meagher, K. L.; Carlson, H. A. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.

(14) *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990.

(15) Steindl, T.; Langer, T. Docking Versus pharmacophore model generation: A comparison of high-throughput virtual screening strategies for the search of human rhinovirus coat protein inhibitors. *QSAR Comb. Sci.* **2005**, *24*, 470−479.

(16) Wolber, G.; Langer, T. CombiGen: a novel software package for the rapid generation of virtual combinatorial libraries. In *Rational Approaches to Drug Design, Proceedings of the European Symposium on Quantitative Structure-Activity Relationships, 13th*, Duesseldorf, 27th Aug. - 1st Sep. 2000; Höltje, H.-D., Sippl, W., Eds.; Prous Science: Barcelona, Spain, 2000; pp 390−399.

(17) Wolber, G. *ilib diverse, version 1.02*; Inte:Ligand GmbH: Maria Enzersdorf, Austria, 2005. www.inteligand.com (accessed month year).

(18) Langer, T.; Wolber, G. Virtual combinatorial chemistry and in silico screening: Efficient tools for lead structure discovery? *Pure Appl. Chem.* **2004**, *76*, 991−996.

(19) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease inhibitors: Current status and future prospects. *J. Med. Chem.* **2000**, *43*, 305−341.

(20) Mothes, E.; Shoeman, R. L.; Traub, P. Pepstatin A: Polymerization of an oligopeptide. *Micron* **1994**, *25*, 189−217.

(21) Frishman, W. H.; Fozailoff, A.; Lin, C.; Dike, C. Renin inhibition: a new approach to cardiovascular therapy. *J. Clin. Pharmacol.* **1994**, *34*, 873−880.

(22) Tumminello, F. M.; Bernacki, R. J.; Gebbia, N.; Leto, G. Pepstatins: aspartic proteinase inhibitors having potential therapeutic applications. *Med. Res. Rev.* **1993**, *13*, 199−208.

(23) Willet, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.

CI600321M