Enhancing Specificity and Sensitivity of Pharmacophore-Based Virtual Screening by Incorporating Chemical and Shape Features−A Case Study of HIV Protease Inhibitors

Deepangi Pandit,[†] Sung-Sau So,[‡] and Hongmao Sun*,[‡]

Department of Chemistry and Environmental Science, New Jersey Institute of Technology, Newark, New Jersey, and Department of Discovery Chemistry, Hoffmann−La Roche, Nutley, New Jersey

Received November 23, 2005

Virtual screening (VS), if applied appropriately, could significantly shorten the hit identification and hit-to-lead processes in drug discovery. Recently, the version of VS that is based upon similarity to a pharmacophore has received increased attention. This is due to two major factors: first, the public availability of the ZINC[1] conformational database has provided a large selection pool with high-quality and purchasable small molecules; second, new technology has enabled a more accurate and flexible definition of pharmacophore models coupled with an efficient search speed. The major goal of this study was to achieve improved specificity and sensitivity of pharmacophore-based VS by optimizing the variables used to generate conformations of small molecules and those used to construct pharmacophore models from known inhibitors or from inhibitor−protein complex structures. By using human immunodeficiency virus protease and its inhibitors (PIs) as a case study, the impact of the key variables, including the selection of chemical features, involvement of excluded volumes (EV), the tolerance radius of excluded volumes, energy windows, and the maximum number of conformers in conformation generation, was explored. Protein flexibility was simulated by adjusting the sizes of EV. Our best pharmacophore model, combining both chemical features and excluded volumes, was able to correctly identify 60 out of 75 structurally diverse known PIs, while misclassifying only 5 out of 75 similar compounds that are not inhibitors. To evaluate the specificity of the model, 1193 oral drugs on the market were screened, and 25 original hits were identified, including 5 out of 6 known PI drugs.

INTRODUCTION

Efficient identification of high-quality leads is a critical step in a modern drug discovery campaign. Although high-throughput screening (HTS) is still a dominant technique in the hit identification process, HTS suffers from two limitations: first, the compound library is mostly limited to historical corporate compounds, a fraction of which are "failed compounds" for various reasons, including poor solubility, poor pharmacokinetic behavior, or high toxicity; second, HTS can only be applied to compounds that physically exist, in sufficient quantity and purity. Computer-based virtual screening (VS)[2,3] is free of these limitations. As its name implies, VS is able to consider any compound, which could include virtual compounds from rational design or real compounds not currently possessed but readily purchasable from vendors.

VS was once a synonym for high-throughput docking.[4−6] Considering the difficulties in accounting for the flexibility of target proteins, solvation effects, and entropy terms in scoring functions, docking-based VS is still under development toward achieving consistent performance against different targets.[7,8] An alternative approach to VS is based on pharmacophore models. Pharmacophores are highly abstract representations of the requirements for biological activity. They are represented as a set of chemical features with a specified special relationship. Chemical features are characterized into several basic pharmacophore types, such as hydrogen-bond donor (HBD), hydrogen bond acceptor (HBA), cation, anion, aromatic system, and hydrophobe. Pharmacophore mapping is more forgiving than the distance-sensitive Lennard-Jones potential used in docking. Moreover, the assignment of atoms in candidate molecules to specific chemical features and the linkage of these atoms are not defined beforehand, allowing the discovery of compounds of different chemotypes that can satisfy the pharmacophore model. This also makes pharmacophore modeling a suitable tool for scaffold hopping.[9,10]

The mapping of small molecules to a pharmacophore model is extremely fast because of the use of a precomputed ensemble of conformers for each compound. This is an expensive step, but calculating the conformers for the database is done only once. Consequently, construction of a high-quality small compound conformational library is one of the preliminary requirements for successful pharmacophore-based VS. For each compound, the ionization state, stereochemistry, and tautomeric state are among the long list of properties that must be properly set. Also, reactive and insoluble compounds should be excluded. Shoichet and Irwin at U.C.S.F. have not only standardized the procedures for preparing and trimming a compound library, but they have supplied a conformational library of purchasable small compounds from major vendors. This library, called ZINC,

* To whom correspondence should be addressed. Phone: (973) 562-3870. Fax: (973) 235-6084. E-mail: Hongmao.sun@roche.com.
[†] New Jersey Institute of Technology.
[‡] Hoffmann−La Roche.

10.1021/ci050511a CCC: $33.50
Published on Web 02/11/2006

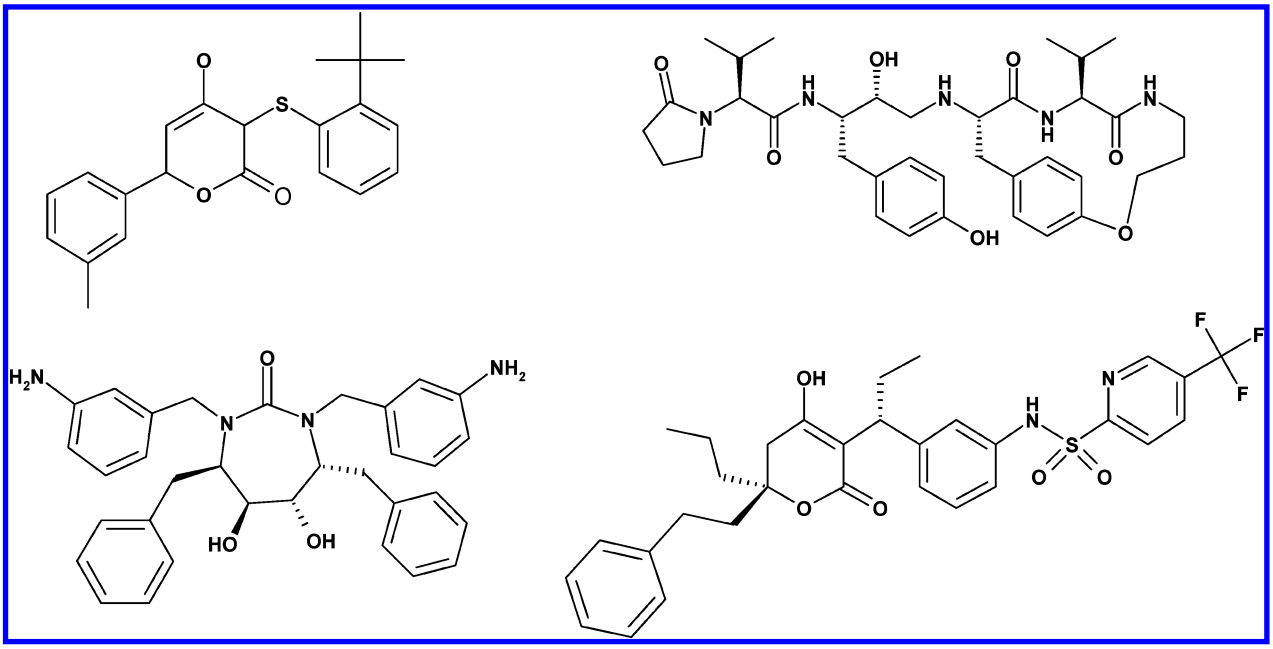

Figure 1. Selected HIV protease inhibitors showing chemical diversity.

is available free of charge[1] to the scientific community. ZINC has the potential to expand the application of VS from merely improving hit rates to identifying novel purchasable hits, thereby offering a jump start for new drug-discovery projects.

The rapid growth of the X-ray crystal structure database is another factor increasing the attractiveness of VS. Crystal structures of many important drug targets and protein−ligand complex structures have been solved in recent years, and new structures are added to the public protein database (PDB) at a pace of about 1000 novel structures per year.[11,12] The ever-increasing availability of target structures, and access to a high-quality conformational database of commercially available compounds, provides a strong platform for pharmacophore-based VS.

Pharmacophore models are tools for integrating the chemical space defined by the small compound library which utilize the existing structural information of the target and its ligands. The program Catalyst from Accelrys[13] dominated the field of conformation enumeration for small molecules and pharmacophore mapping for many years, before other software, such as OMEGA from OpenEye Science Software[14] and MOE from Chemical Computing Group,[15] became available recently. OMEGA has proven capable of retrieving the correct binding conformations of ligands against different proteins,[16] but few studies have been published on the use of MOE to build pharmacophore models and perform pharmacophore mapping. In this study, by building pharmacophore models with MOE from a recently determined crystal structure of human immunodeficiency virus protease (HIVP) in complex with a potent protease inhibitor (PI), we explored the capability of pharmacophore-based VS to select true PIs from a collection of PI-like compounds. The method reported herein was a novel approach incorporating features from both the protein and ligand, a strategy that has not received much attention from the community. We hope this example arouses interest in the community toward developing similar models for other drug targets.

COMPUTATION METHODS

Protein Systems. HIVP and its inhibitors (PIs) are often regarded as an ideal case for many reasons. First, abundant structural information accumulated over the years for HIVP has made it one of the best understood targets. There are more than 200 crystal structures of HIVP in the PDB.[17,18] Second, HIVP inhibitors are many in number and diverse in structure (Figure 1), and several inhibitors have been developed into marketed drugs. Third, HIVP is a difficult problem,[19] because of the flexibility of the flaps in HIVP[20−22] but also because the inhibitors tend to have a large number of rotatable bonds. An additional goal of the current study was to find out if pharmacophore-based VS could tackle the protein flexibility problem by adjusting excluded volumes.

The pharmacophore models were built from a newly solved X-ray crystal structure of the triple mutant $HIVP_{3X}$ (L63P, V82T, and I84V) in complex with the small molecule inhibitor TMC114 (PDB code 1T7J).[23,24] The triple mutant protein is widely used as a prototype for studying drug-resistant versions of the HIV virus.[25] To investigate the effect of mutations of the protein and selection of the initial structure on the performance of VS, structures of wild-type HIVP with TMC114 (PDB code 1T3R)[23] and Amprenavir (APV; PDB code 1HPV) bound and the multidrug-resistant (MDR) triple mutant protein with Amprenavir (PDB code 1T7I)[23] bound were also considered in the pharmacophore model construction. APV and TMC114 (Figure 2a) both fit primarily within the substrate pocket. Although TMC114 and APV are close analogues, the perturbations to the active site of HIVP upon ligand binding were different (Figure 2b).

Hydrogen atoms were added to the proteins and ligands of the crystal structures, and the resulting structures were subjected to energy minimization using MOE.[15] Heavy atoms were first fixed and then tethered in primary minimization steps. The system was minimized to a root-mean-square (RMS) gradient of less than 0.05 Å, after gradually removing the tethering force. The MMFF94x force field was applied to carry out energy minimizations.

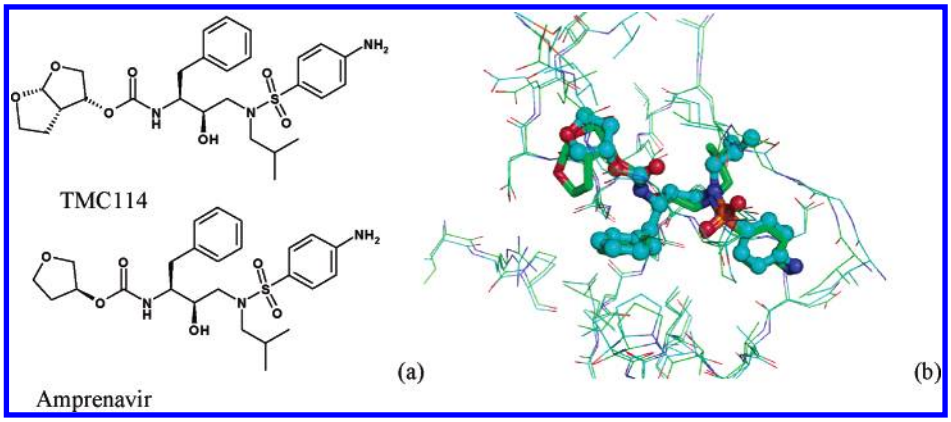

Figure 2. (a) Chemical structure of TMC114 and APV. (b) Superposition of ligand binding sites of wild-type HIVP with TMC114 (colored in green) and APV (colored in cyan) bound.

Development of the Pharmacophore Models. The minimized complex structure of $HIVP_{3X}$ and TMC114 was used to create pharmacophore models for VS, by using the Pharmacophore Query Editor of MOE.[15] The pharmacophore scheme of PCH (polarity−charge−hydrophobicity) was applied throughout the study. Locations of chemical features were set to those suggested by MOE, while the tolerance radius of each feature was adjusted to achieve better balance between sensitivity and specificity. Selected pharmacophoric features were assigned as "essential", and other features were presented for a partial match. The active-site atoms of HIVP, those within a 4.5 Å distance from any atom of the ligand, were used to generate excluded volumes. No other volume constraints were defined.

Databases. Two ligand databases were prepared to evaluate pharmacophore models. The drug database consisted of 75 unique ligands selected from the data set provided by Meagher and Carlson.[26] Our selection was based on biological activity. The original database contains PIs with IC_{50} values ranging from 1.3 pM to 52 μM. Since the biological data are from different labs, we applied a generous cutoff of 400 nM of K_i or IC_{50} for the selection of PIs. Wherever the cocrystallized structures of PIs were available, ligands were extracted in the bound conformation from the cocrystal structures in order to retrieve the correct stereoisomer. Ligands with no bound crystal structures were built with MOE.[15] The database of 75 druglike non-PIs was generated from a data set of 85 compounds also provided by Meagher and Carlson,[26] by removing 10 close analogues. The original data set was created from the Comprehensive Medicinal Chemistry Index. To select compounds that would be physically similar to PIs, only those ligands with a molecular weight between 300 and 1200, with at least one aromatic group and at least one HBD but without reported HIVP activity, were selected.[26] All ligands were initially minimized using the MMFF94x force field before submitting them to OMEGA.[14] A list of both the known PIs and druglike non-PI compounds used in this study are provided in the Supporting Information.

Using OMEGA, conformations were enumerated for each compound employing an energy cutoff of 5 kcal/mol, since typically a bound conformer is one of the low-energy conformers, although not necessarily the lowest-energy conformer.[16,27] To allow the generation of conformers for all of the compounds in the databases, the maximum number of rotatable bonds and maximum number of flexible ring systems were set to 40 and 10, respectively. A heavy-atom root-mean-square deviation (RMSD) of 0.6 Å was used; that is, conformers with heavy-atom RMSDs less than 0.6 Å would be considered duplicates. Final optimization was performed with the MMFF on conformers which were generated by varying torsions and evaluated by the Dreiding force field. A Sheffield solvation correction term was used to simulate an aqueous environment.[14]

A conformational database containing 1193 orally administrated marketed drugs was prepared,[28] by following the standard procedures mentioned above. The virtual conformations were generated from 2D structures. The chiralities of six PI drugs were examined and corrected before enumeration.

To examine the effects of the number of conformers and the energy window on the quality of the conformational ensemble, and on the success rate of the resulting models, a maximum of 400 and 1000 lowest-energy conformers were retained using the energy cutoffs of 5 and 20 kcal/mol, respectively.

Criteria for Model Evaluation. Generated pharmacophore models were used to screen the databases containing the 75 PIs and the 75 PI-like compounds. MOE was used to carry out VS. A ligand was counted as a hit if at least one of the enumerated conformers could be aligned to the pharmacophore model. The aim of this study was to develop a model based on as little structural information as possible, which would be capable of identifying the maximum number of active compounds and a minimal number of false positives.

RESULT AND DISCUSSION

Pharmacophore Query Construction. The recently published cocrystal structure of $HIVP_{3X}$ and TMC114 was used for model optimization, because TMC114 is highly active against both wild-type HIVP and a triple mutant widely used to study MDR.

Chemical features play a pivotal role in pharmacophore-based VS. Successful selection of chemical features stems from a thorough understanding of the target system. Like many other aspartic proteases, HIVP is characterized by two proximate aspartic acids, one from each monomer, to form the catalytic site.[21] A transition state analogue such as an aliphatic hydroxyl group in a ligand which can mimic a tetrahedral intermediate for amide bond hydrolysis[29,30] is

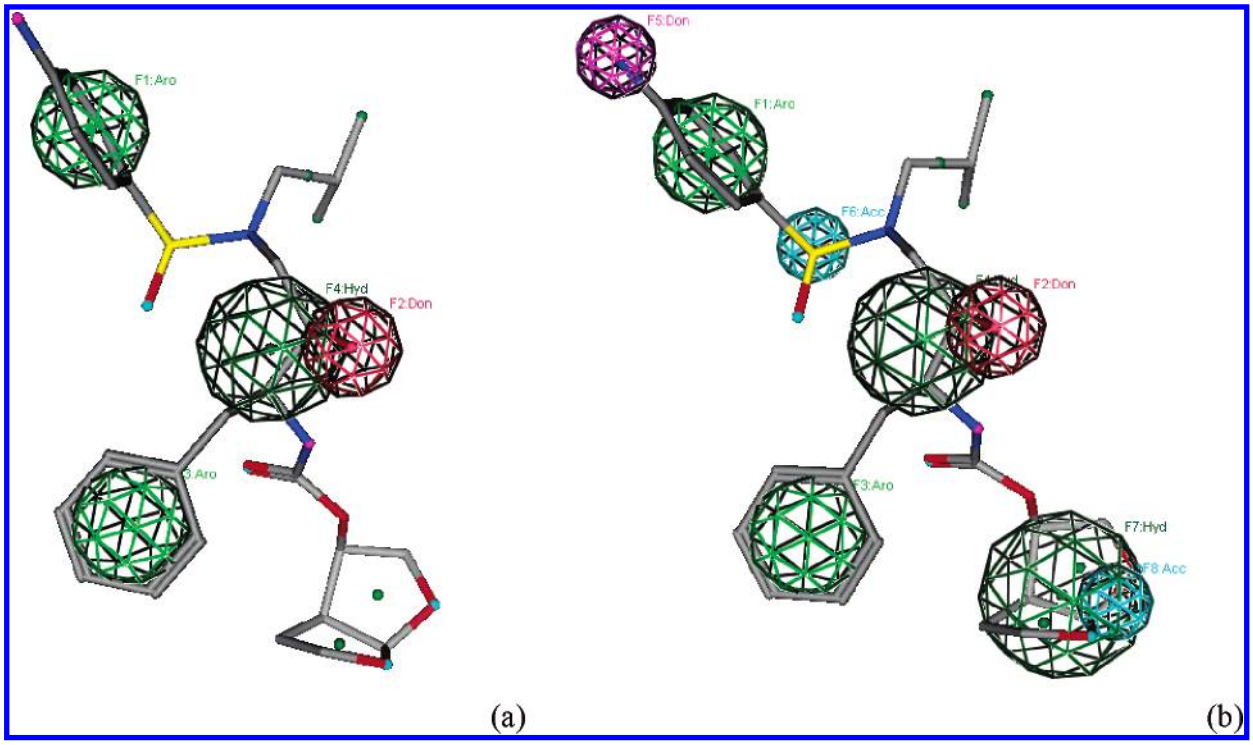

Figure 3. Core pharmacophoric features (a) and complete pharmacophoric features (b) of the pharmacophore model.

often present in potent HIVP inhibitors. Therefore, the first chemical feature selected for the pharmacophore model was a HBD, represented by the hydroxyl group of TMC114. Although a hydroxyl group can serve as either a HBD or a HBA, the pharmacophoric feature was specified as HBD to reflect its function and, thereby, to enhance the specificity of the model. To enhance the basicity of the proton on a hydroxyl group, a hydrophobic environment is generally required. Requiring the combination of a HBD and a hydrophobe in proximity further improved the specificity of the model. Two other aromatic sites, which were formed by the side chains of ALA28, VAL32, ILE47, ILE50, and LEU76, and LEU23, PRO81, and VAL84, respectively, were also selected. This formed a four-point pharmacophore model (Figure 3a). The tolerance radii of HBD, hydrophobe, and aromatic sites were adjusted from the default values of 0.8, 1.8, and 1.2 Å to 1.0, 1.5, and 1.2 Å, respectively, to achieve an optimal balance of sensitivity and specificity. The model based on these four "essential" features was used to screen the PI database and produced a success rate of 87% (65/75). Screening the non-PI database yielded a false positive rate that was somewhat high—25% (19/75).

It was clear from these results that the simple four-point pharmacophore model was able to correctly identify active PIs but lacked sufficient specificity to exclude PI-like noninhibitors. In an attempt to further optimize this model so that the success rate could be sustained while reducing the false positive rate, a set of four additional features consisting of two HBAs, one HBD, and one hydrophobic region were identified (Figure 3b) from the structure of TMC114. It was found that the presence of any one of these additional four features in conjunction with the four original essential features improved specificity. This model, which allowed a partial match for five features—four original features and one out of the four additional features—was able to screen the drug database with a success rate of 87% (65/75) and the nondrug database with a false positive rate of 15% (11/75).

Effect of Excluded Volume. The essence of molecular recognition is the complementarity of both electrostatics and shape at the interface of two molecules, which is also widely appreciated as the general guideline for de novo drug design. Chemical features define the essential requirements for a ligand to form favorable electrostatic interactions with its target protein, but it is obvious that not all of the compounds that meet these requirements can bind to the target, because the size and shape of the compounds may not be appropriate. Including shape features into a pharmacophore query will significantly improve the specificity of the model, since each protein has a uniquely shaped ligand binding site.[31] Excluded volumes (EV) derived from the coordinates of active-site atoms of the target protein represent a simple but accurate way to model the receptor shape (Figure 4). In an attempt to further improve the specificity of the pharmacophore model, we decided to unite EV to the existing chemical features. As mentioned previously, protein flexibility can confound molecular docking, so the key issue was how to account for it in the EV method. The solution was based on the fact that the size of EV is adjustable—reducing the size of EV is equivalent to allowing higher protein flexibility. Obviously, reduction of EV gives more room for ligands to fit in and reduces the specificity of the resulting pharmacophore model. Therefore, the size of EV determines a compromise between specificity and sensitivity and, thereby, has an intimate effect on the hit rate. We studied the effect of the size of the EV using the model which allowed a partial match for five features (four essential features and one out of four additional features). As illustrated in Figure 5, an EV of 0.5 and 0.6 Å lowers the false positive rate significantly from 15% with no EV applied to 7% and 5% with reasonable success rates of 80% and 72%, respectively. When the radius of EV increased to 1.0 Å, the success rate

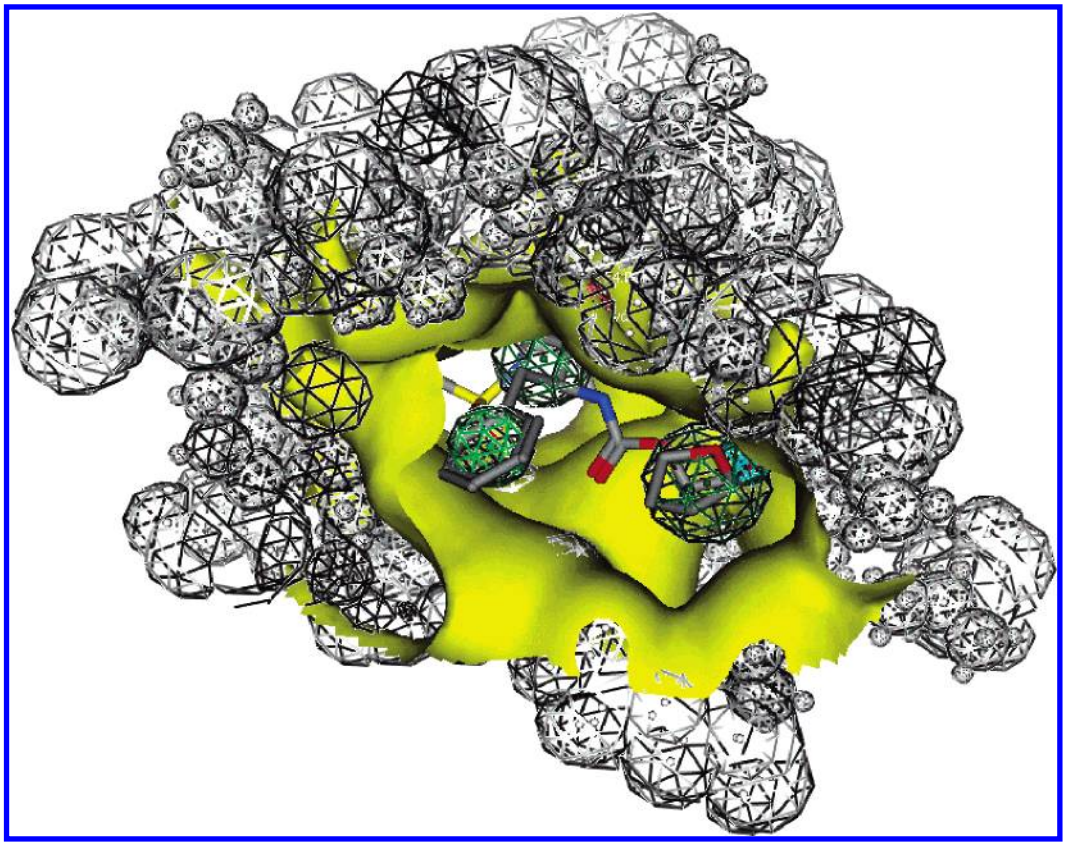

Figure 4. Excluded volumes (EV), shown as gray spheres, well defining the shape of the HIV-protease-active site. The surface of HIV protease is shown in yellow color.

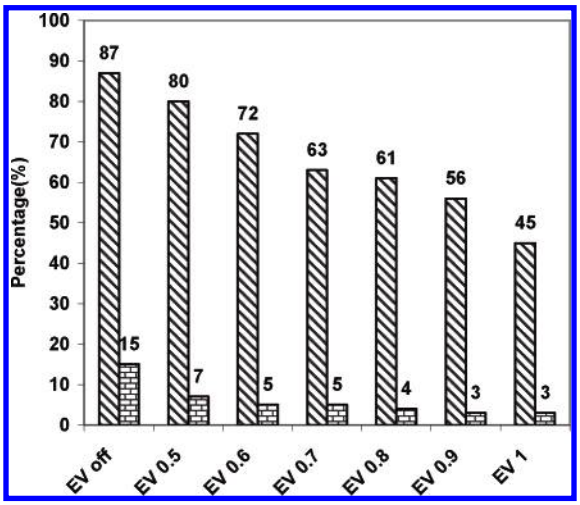

Figure 5. Effect of excluded volume on the performance of the pharmacophore model; crosshatched lines represent the success rate; bricks are the false positive rate. EV represents excluded volume in angstroms (Å).

dropped to half of that without EV. This was a good indicator of the high flexibility of HIVP, given the fact that the suggested EV radius by MOE was between 1.5 and 2.5 Å. We decided to select the model with an EV of 0.5 Å for studies of other protein systems.

A close inspection of the false positives and false negatives demonstrated that the false positives were either peptides or flexible molecules rich in functional groups (Figure 6), which had a tendency to fit into multiple pharmacophore models. If we narrowed down the query by requiring an aliphatic hydroxyl group as the essential HBD in order to mimic the intermediate of the enzyme reaction, only flavin adenine dinucleotide, or MCMC00006171, remained in the hit list. On the other hand, a large fraction of false negatives lacked one or both of the two "essential" aromatic rings defined in the pharmacophore model (Figure 7). Since the pharmacophore model was built on the basis of an asymmetric molecule, a couple of *c*2 symmetric PIs were not correctly classified. Two Warfarin derivatives,[32] which lacked a tetrahedral carbon linked directly to the hydroxyl group, were also missing from the hit list.

Table 1. Impact of the Selection of the Protein−Inhibitor Complex Structure Used to Construct a Pharmacophore Query on the Performance of VS, as Indicated by Success Rate and False Positive Rate

system		PIs (total: 75)	non-PIs (total: 75)
Amprenavir	wild type	63 (84%)	9 (12%)
	mutant 3X	60 (80%)	8 (11%)
TMC114	wild type	64 (85%)	7 (9%)
	mutant 3X	60 (80%)	5 (7%)

Effect of Protein Systems. To study the generality of this model and its applicability to other protein systems, we developed similar models based on $TMC114_{wt}$, $APV_{wt,}$ and APV_{3X} (Table 1). The new models generated did not deviate significantly from the original optimized model. The success rates were also comparable, in the range of 80−85%, with false positive rates in the range of 7−12%. This showed the strength, specificity, and robustness of the model. It appeared that, if optimized chemical features were selected, along with an EV which defined the shape of the receptor with appropriate looseness, the model would be able to effectively screen out PIs with a low false positive rate.

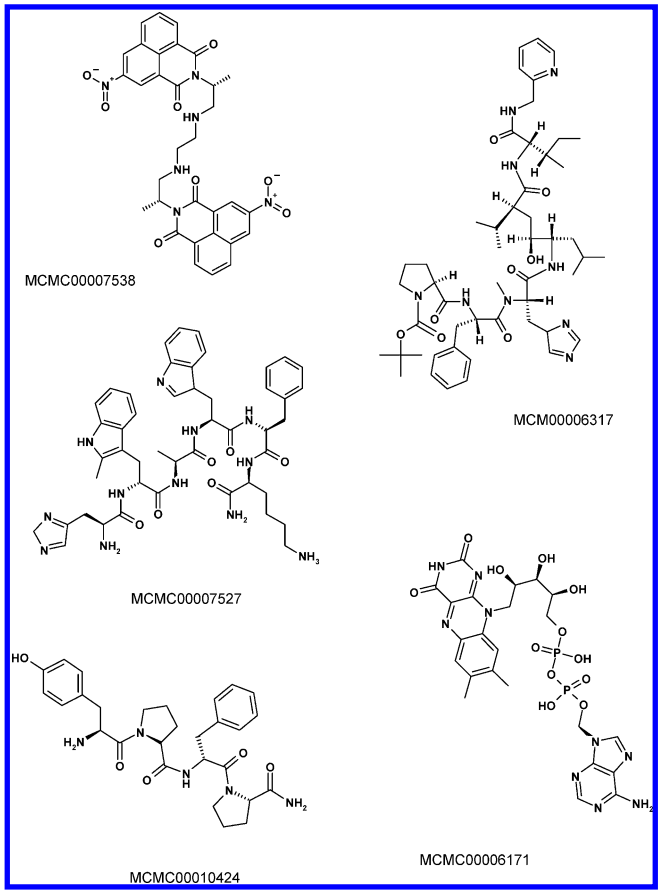

Figure 6. Structures of false positives from screening 75 PI-like noninhibitors.

Effect of Energy Window and Maximum Conformers. In pharmacophore-based VS, flexibility of proteins can be accounted for by adjusting the EV size, while flexibility of ligands is represented by an ensemble of conformers. The efficacy of VS is dependent on the quality of the conformers generated for the compounds in the database, in addition to the quality of the pharmacophore model. We used OMEGA to enumerate conformers in this study. OMEGA is a rule-based method that generates conformations by first disassembling a molecule into fragments at rotatable bonds and then reassembling the fragments on the basis of the sorted order of the fragment energies.[14] To study the effect of the energy window and number of conformers generated on the performance of the model, we fixed the RMSD to 0.6 Å, as suggested by Boström et al.[16] The energy window, number of conformers, and RMSD are interrelated. Allowing a higher energy window, in general, means a requirement for a larger number of conformers to complete the ensemble, given RMSD. Increasing the RMSD value increases the rejection rate, thus making it possible to fulfill a complete ensemble with a lower number of conformers, given the energy window, but at a cost of longer computing time. Considering the limitations of disk space and computational power when generating a large database containing millions of compounds, the number of conformers will most likely be limited to a few hundred. Another important thing to notice is that, for a given number of conformers, OMEGA generates that number of conformers that meets the requirements of the energy window and RMSD value, not necessarily the best or most representative conformers. If the number of conformers is not large enough to represent the whole ensemble, the resulting set of conformers is incomplete. An incomplete ensemble cannot properly represent the conformational space of its parent compound. Of course, the above discussion is relevant only for highly flexible molecules, since for a

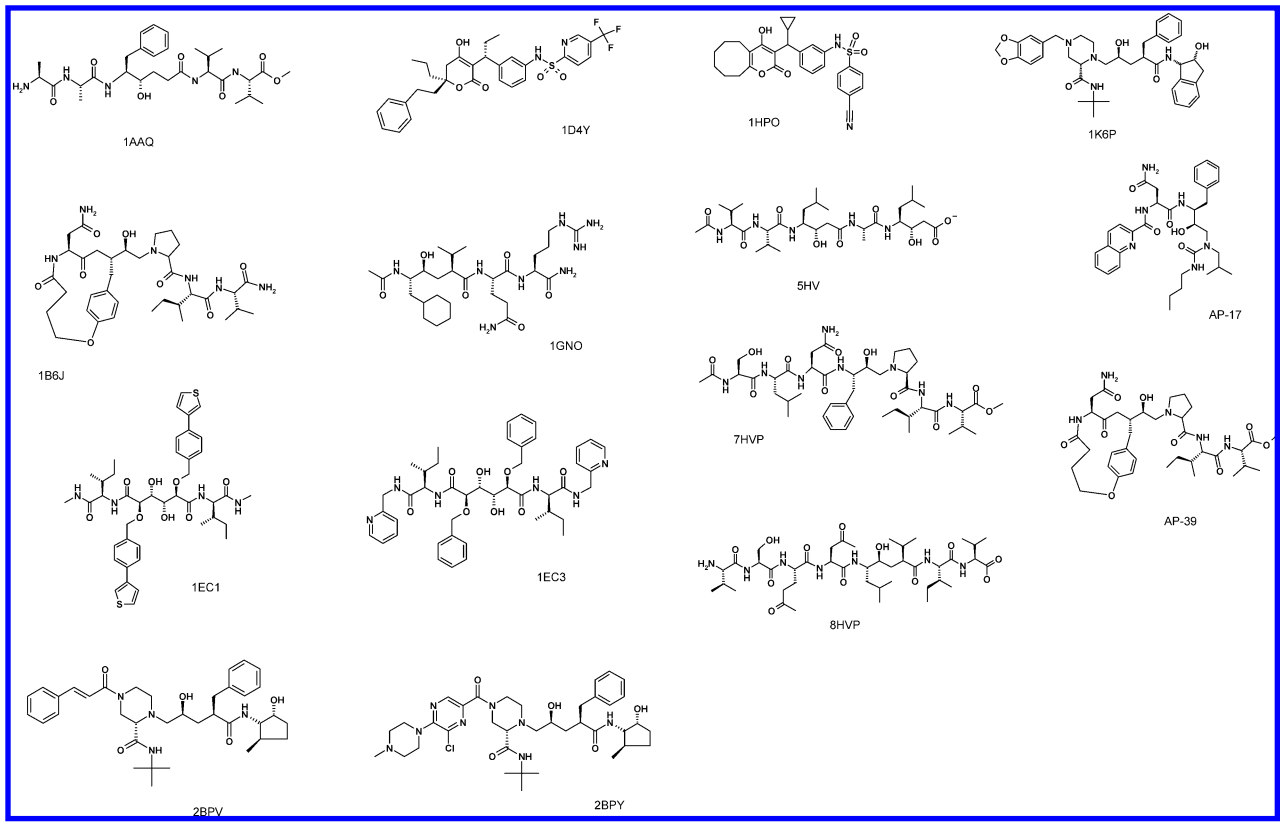

Figure 7. Structures of misclassified known PIs.

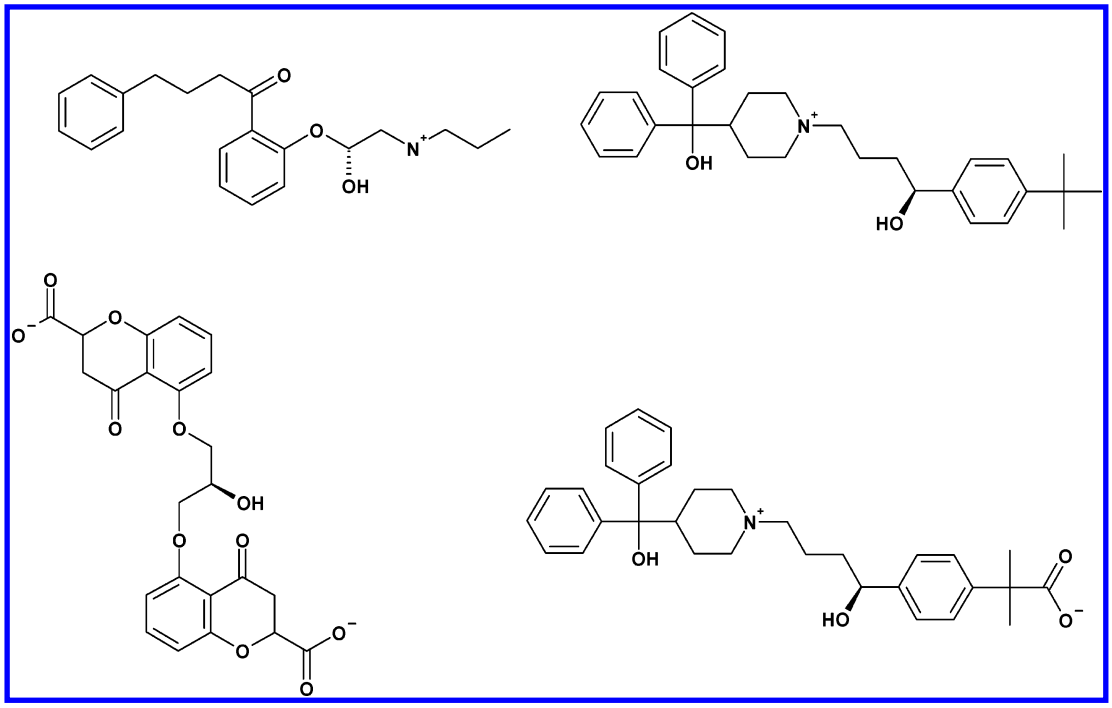

Figure 8. Representative false positives from screening the 1193-drug database.

Table 2. Impact of the Energy Window and Maximum Number of Conformers on the Performance of VS, as Indicated by Success Rate and False Positive Rate

energy window (kcal/mol)	maximum number of conformers	PIs (total: 75)	non-PIs (total: 75)
5	400	60 (80%)	5 (7%)
	1000	61 (81%)	8 (11%)
20	400	58 (77%)	7 (9%)
	1000	62 (83%)	9 (12%)

relatively rigid molecule, a complete ensemble will be reached before hitting the limitation of conformation number. We enumerated a database of conformers imposing an energy cutoff of 5 kcal/mol and 1000 as the maximum number of conformers for both PI and non-PI compounds. Compared to the database generated using a 5 kcal/mol cutoff and 400 as the maximum number of conformers, the success rate increased to 81% from 80% but the false positive rate also increased to 11% from 7% (Table 2). To study the effect of energy cutoff, we created similar databases with energy cutoffs of 20 kcal/mol and maximum numbers of conformers set to 400 and 1000. The database with an energy cutoff of 20 kcal/mol and a maximum number of conformers of 400 gave a success rate of 77% and a false positive rate of 9%. The slight reduction in the success rate might result from an incomplete ensemble. The database with an energy cutoff of 20 kcal/mol and a maximum number of conformers of 1000 gave a success rate of 83% and a false positive rate of 12% (Table 2). In summary, databases enumerated with larger energy windows and numbers of conformers identified more hits, but the differences were not significant. Therefore, to enumerate conformations for a library of medium-sized compounds with reasonable flexibility, an energy cutoff of 5 kcal/mol and a number of conformers of 400 are recommended.

Specificity of the Model. The 1193 oral drugs were screened with the pharmacophore model of both chemical and shape features. It turned out that 25 hits (2.1% hit rate)

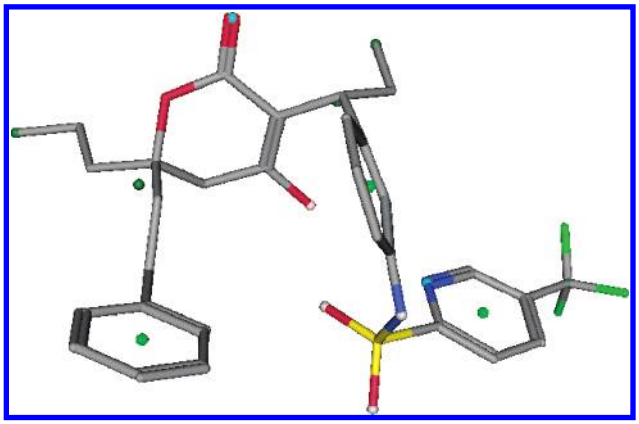

Figure 9. HIVP drug Tipranavir with PCH pharmacophoric features assigned.

were identified, among which five known HIVP drugs were Saquinavir, Ritonavir, Nelfinavir, Indinavir, and Amprenavir. Since all six known HIVP drugs contain multiple chiral centers, it is very important to make sure the right stereoisomer is selected for each drug; otherwise, the wrong stereoisomer cannot map into the pharmacophore model. The number of false positives can be further reduced from 20 to 4, if it was specified that the essential HBD must be a hydroxyl group to mimic the transition state. (Figure 8) The only false negative, Tipranavir, is highly active against wild-type HIVP with an IC_{50} of 50 nM[32] but cannot map into the pharmacophore model, due to the lack of a hydrophobic site next to the essential HBD (Figure 9). Actually, the local environment of the HBD is hydrophobic, but MOE did not recognize it because of a nearby double bond and lactone. Tipranavir is the only PI drug with the hydroxyl group bound to an sp^2, instead of an sp^3, carbon atom.

CONCLUSION

The practical goal of pharmacophore-based VS is not to identify all the hits but to identify a reliable set of interesting

hits to offer a jump start for a new project. It was demonstrated that a pharmacophore model derived from a single crystal structure, where locations of features and tolerance radii were not optimized against the entire training set, was able to successfully identify 80% of the structurally diverse PIs with a false positive rate of only 7% for PI-like noninhibitors. It was also illustrated that the inclusion of excluded volumes improved the specificity of the model significantly. The results indicated that the performance of VS was not very sensitive to the selection of original protein structures, or to mutations, when the EV tolerance radius was set to 0.5 Å. Protein flexibility is expected to be more precisely reflected if the sizes of EV could be set according to B factors of the atoms in the ligand binding site, or variation of the ligand binding site atoms upon binding of different ligands. In addition, it was shown that, as long as the ensemble of conformers was nearly complete, incrementation of the energy window and number of conformers in enumerating conformations had little effect on the performance of VS. The search speed was excellent—screening of the 1193-drug library by one-hit-per-molecule, including outputting a database containing the hits in docking mode, took less than 30 s on an HP workstation xw8000 with a 3.06 GHz CPU.

Of course, pharmacophore-based VS has its limitations. First, it cannot rank the hits, unless the binding energy is computed for the hits in a subsequent step; second, pharmacophore models are not accurate enough to address the selectivity issue among highly homologous proteins; third, a single pharmacophore model with reasonable specificity cannot cover all the hits. The last two limitations reflect the compromise between sensitivity and specificity of a model. Improvement of the specificity of a model is always at the cost of losing sensitivity, and vice versa. Nevertheless, the encouraging results presented in this study suggest that VS has the potential to play a major role in hit identification under situations where either a suitable bioassay is not available or screening of an entire corporate compound library is not desirable or possible. The application of pharmacophore-based VS can also be extended to the development of fast-follower drugs, where a high-quality crystal structure of the target in complex with a potent ligand is already available.

ACKNOWLEDGMENT

The authors thank Drs. David Fry and Bradford Graves for critical reading of the manuscript. The authors are thankful to Dr. Heather A. Carlson for sending us the data sets of PIs and non-PI compounds. D.P. appreciates financial support for her summer internship at Hoffmann−La Roche Inc. in Nutley.

Supporting Information Available: A list of HIV protease inhibitors selected on the basis of biological activity and structural diversity. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(2) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening − an overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(3) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894. FIELD Reference Number: 113 FIELD, Journal Code: 101124171 FIELD Call Number.
(4) Toledo-Sherman, L. M.; Chen, D. High-throughput virtual screening for drug discovery in parallel. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 414−421.
(5) Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 407−413.
(6) Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model.* [Online] **2003**, *9*, 47−57.
(7) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(8) Schulz-Gasch, T.; Stahl, M. Scoring functions for protein−ligand interactions: a critical perspective. *Drug Discovery Today Technol.* **2004**, *1*, 231−239.
(9) Green Darren, V. S. Virtual screening of virtual libraries. *Prog. Med. Chem.* **2003**, *41*, 61−97. FIELD Reference Number: 148, FIELD Journal Code: 0376452 FIELD Call Number.
(10) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.
(11) Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank. *Acta Crystallogr., Sect. D* **2002**, *58*, 899−907.
(12) Deshpande, N.; Addess, K. J.; Bluhm, W. F.; Merino-Ott, J. C.; Townsend-Merino, W.; Zhang, Q.; Knezevich, C.; Xie, L.; Chen, L.; Feng, Z.; Green, R. K.; Flippen-Anderson, J. L.; Westbrook, J.; Berman, H. M.; Bourne, P. E. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.* **2005**, *33*, D233−237.
(13) *Catalyst*; Accelrys, Inc.: San Diego, CA. http://www.accelrys.com/products/catalyst/ (accessed January 2006).
(14) *Omega*; OpenEye Science Software: Sana Fe, NM. http://www.eyesopen.com/products/applications/omega.html (accessed January 2006).
(15) *MOE*; Chemical Computing Group Inc.: Montreal, Canada. http://www.chemcomp.com/software.htm (accessed January 2006).
(16) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449−462.
(17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(18) Vondrasek, J.; van Buskirk, C. P.; Wlodawer, A. Database of three-dimensional structures of HIV proteinases. *Nat. Struct. Biol.* **1997**, *4*, 8.
(19) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962−976.
(20) Wlodawer, A.; Gustchina, A. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys. Acta* **2000**, *1477*, 16−34.
(21) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein−Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97*, 1359−1472.
(22) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **2000**, *301*, 1207−1220.
(23) Surleraux, D. L. N. G.; Tahri, A.; Verschueren, W. G.; Pille, G. M. E.; de Kock, H. A.; Jonckers, T. H. M.; Peeters, A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M.-P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. T. P. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **2005**, *48*, 1813−1822.
(24) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Wigerinck, P.; de Bethune, M. P.; Schiffer, C. A. Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. *J. Virol.* **2004**, *78*, 12012−12021.
(25) Molla, A.; Granneman, G. R.; Sun, E.; Kempf, D. J. Recent developments in HIV protease inhibitor therapy. *Antiviral Res.* **1998**, *39*, 1−23.
(26) 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.
(27) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.
(28) Vieth, M.; Siegel, M. G.; Higgs, R. E.; Watson, I. A.; Robertson, D. H.; Savin, K. A.; Durst, G. L.; Hipskind, P. A. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* **2004**, *47*, 224−232.

(29) Rich, D. H. Pepstatin-derived inhibitors of aspartic proteinases. A close look at an apparent transition-state analogue inhibitor. *J. Med. Chem.* **1985**, *28*, 263−273.
(30) Wiley, R. A.; Rich, D. H. Peptidomimetics derived from natural products. *Med. Res. Rev.* **1993**, *13*, 327−384.
(31) Greenidge, P. A.; Carlsson, B.; Bladh, L. G.; Gillner, M. Pharmacophores incorporating numerous excluded volumes defined by X-ray crystallographic structure in three-dimensional database searching: application to the thyroid hormone receptor. *J. Med. Chem.* **1998**, *41*, 2503−2512.
(32) Chrusciel, R. A.; Strohbach, J. W. Nonpeptidic HIV protease inhibitors. *Curr. Top. Med. Chem.* **2004**, *4*, 1097−1114.

CI050511A