Targets Looking for Drugs: A Multistep Computational Protocol for the Development of Structure-Based Pharmacophores and Their Applications for Hit Discovery

Cristina Tintori,[†] Valentina Corradi,[†] Matteo Magnani,[†,‡] Fabrizio Manetti, and Maurizio Botta*

Dipartimento Farmaco Chimico Tecnologico, Università di Siena, Via Aldo Moro, I-53100 Siena, Italy

Received March 27, 2008

Pharmacophores—three-dimensional (3D) arrangements of essential features enabling a molecule to exert a particular biological effect—constitute a very useful tool in drug design both in hit discovery and hit-to-lead optimization process. Two basic approaches for pharmacophoric model generation can be used by chemists, depending on the availability or not of the target 3D structure. In view of the rapidly growing number of protein structures that are now available, receptor-based pharmacophore generation methods are becoming more and more used. Since most of them require the knowledge of the 3D structure of the ligand-target complex, they cannot be applied when no compounds targeting the binding site of interest are known. Here, a GRID-based procedure for the generation of receptor-based pharmacophores starting from the knowledge of the sole protein structure is described and successfully applied to address three different tasks in the field of medicinal chemistry.

INTRODUCTION

Pharmacophores—three-dimensional (3D) arrangements of essential features enabling a molecule to exert a particular biological effect—constitute a very useful tool in drug design.[1] In other words, molecules have affinity toward a particular receptor if they have a number of chemical features which interact favorably with the receptor and which possess a geometry complementary to it.

There are two basic approaches for pharmacophoric model generation, which depend on the available information about the 3D structure of the target under study. If this is known, either from experimental (i.e., X-ray crystallography, NMR studies) or theoretical (receptor structure built by homology modeling techniques) sources, structure-based pharmacophores can be developed. Otherwise, ligand-based pharmacophoric models can be derived from a set of ligands known to act with a common mechanism of action.[2] Once generated, a pharmacophoric model represents a versatile tool for drug design. In fact, it can be used as a query to screen databases of commercially available compounds or, alternatively, to guide chemists in the synthesis of new compounds during the hit-to-lead optimization process. Moreover, it can also be used to align molecules based on the 3D arrangement of chemical features or to develop predictive 3D QSAR models.[3] Recently, a new application of pharmacophore modeling, namely parallel screening, has been developed and successfully applied. This innovative concept consists of screening a compound by several pharmacophore models to predict its biological profile.[4−7]

Several methods have been developed over the years for pharmacophore generation, most of them (e.g., DISCO, Catalyst and GASP)[8−12] dealing with ligand-based approaches. However, some methods for the generation of receptor-based pharmacophores have been also described in recent literature. Almost all of these procedures[13−15] require the knowledge of the 3D structure of the ligand-protein complex, thus they cannot be applied when no compound targeting the binding site of interest is known (when one or more active ligands are known, the absence of the crystal structure of the complex could be overcome by molecular docking). On the contrary, the generation of receptor-based pharmacophoric models starting from the knowledge of the sole target has received little attention.[16−22]

In this study, we have applied a computational procedure able to generate receptor-based pharmacophoric models starting from the analysis of the protein target by means of the GRID software. Briefly, the protocol involves the computation of points of minimum energy for GRID Molecular Interaction Fields (MIFs)[23] in the binding site of interest, followed by the conversion of such points into pharmacophoric features. It should be emphasized that such an approach constitutes an application of the GRID MIFs and represents a further confirmation of their broad versatility.[24−28] In this regard, the use of GRID minima to describe pharmacophoric points in the context of the protein binding site has been already reported. In particular, Mason and Cheney have applied this method to study the selectivity between different serine proteases.[29,30] Furthermore, Fox and Haaksma have presented, in 2000, a comparable procedure to map the binding site of thrombin and to perform a flexible 3D database search for benzamidine-based thrombin inhibitors.[18] We have applied a similar protocol in the context of a more complex computational procedure which combines the pharmacophoric model generation with other techniques such as docking studies, molecular dynamics simulation, and conformational analysis. This approach allows the generation of structure-based pharmacophoric models also for receptor sites which have never been targeted before and for which

* Corresponding author phone: 0039 0577 234306; fax: 0039 0577 234333; e-mail: botta@unisi.it.

[†] These author contributed equally to this work.

[‡] Current address: Siena Biotech SpA, Strada del Petriccio e Belriguardo, 35 - 53100, Siena, Italy.

10.1021/ci800105p CCC: $40.75
Published on Web 10/23/2008

therefore no ligands able to interact with them are known. In this aspect, such an approach is different from the GRID-based procedure for the generation of pharmacophoric models derived from ligand-protein complexes recently reported by Ortuso et al.[15,31] The idea of the present study is that, in those cases characterized by a lack of information about a ligand-target complex, the GRID minima within the protein binding site could be used to generate pharmacophoric models to be synergistically integrated to other computational filters in an elaborate virtual screening procedure in such a way as to improve the drug discovery process.

Structure-based virtual screening usually involves docking algorithms to put compounds into an active site of the target protein, followed by application of scoring functions which in turn produce a ranking list of compounds to identify potential hits. However, several recent studies have shown that pharmacophore-based filtration in combination with docking-based virtual screening can significantly improve the probability of identifying putative candidates.[32,33] In fact, the use of pharmacophore-based searches before docking studies can significantly reduce the number of chemicals to be analyzed, thus allowing more exhaustive docking on the retained compounds and a more accurate investigation of their binding modes and interactions with the target. Moreover, when inserted into a virtual screening procedure, pharmacophoric models could also represent a useful tool to assess the reliability of docking results and to guide the final selection of compounds. In fact, although pharmacophore generation and docking calculations are two completely different approaches from a conceptual point of view, they are both aimed at hypothesizing the bioactive conformation of ligands. This implies that the orientation of compounds within the binding site and their interaction pattern, proposed by docking simulations, should be in good agreement with the ligand conformations mapping the pharmacophoric functions at best, independently from the protocol (ligand- or structure-based approach) by which the pharmacophore hypothesis has been generated. As a consequence, compounds for which a consensus between docking results and pharmacophore mapping occurs should be preferred to compounds for which the binding conformation predicted on the basis of docking studies differs from that provided by the pharmacophore approach.

On the other hand, such a protocol can be combined with molecular dynamics (MD) simulations or conformational searches to take into account the flexibility of the receptor active site.[34] In fact, receptor flexibility is still a critical issue in structure-based virtual screening methods. Although docking on multiple conformations of the receptor is an efficient way to account for receptor flexibility, it is still too CPU-demanding for large molecular libraries. In this context, the combination of receptor-based pharmacophoric models with MD or conformational searches within a virtual screening protocol allows to face the flexibility problem in a reasonable time.

The effectiveness of such a protocol has been tested in three different tasks of the medicinal chemistry, and satisfactory results were obtained in each field validating its versatility and applicability. The first case study reports a simple application of the receptor-based pharmacophoric model protocol in combination with docking studies for the research of small molecules able to inhibit the activity of thioredoxin reductase (TrxR) from *Mycobacterium tuberculosis* (MTB), one of the new targets for antitubercular therapy. The second example shows a more elaborate application of the same protocol for the search of novel HIV-1 integrase (IN) inhibitors able to affect the IN-viral DNA complex formation. Since the DNA binding site of IN is characterized by a very flexible region (namely, the loop 140−149), the flexibility of this loop was taken into account by means of a conformational search study, and the most representative conformations were used to generate different receptor-based pharmacophoric models. Finally, in the third example, we have applied the computational protocol to a protein−protein system to identify anti-HIV-1 agents affecting reverse transcriptase (RT) dimerization. In this case, the flexibility of the receptor active site has been explored through MD simulations.

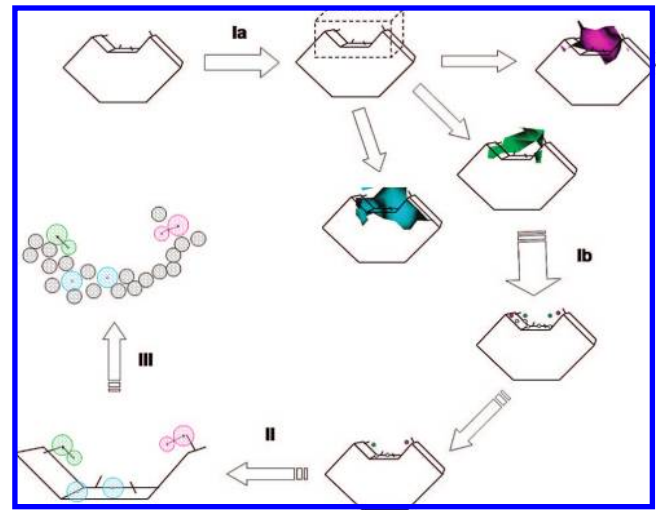


Figure 1. Schematic representation of receptor-based pharmacophoric model generation workflow.

METHODS

Virtual Screening Protocol. The virtual screening protocol used in all the case studies described here was fundamentally based on the application of two sequential pruning steps on a database of commercially available compounds in order to select a restricted number of compounds to be submitted to biological evaluation. In the first step, the structure-based pharmacophoric models, generated as described below, were used as search queries on the whole database. Subsequently, docking calculations were performed on the selected molecules within the binding site under analysis.

Receptor-Based Pharmacophore Model Generation. The procedure applied in this paper for the receptor-based pharmacophore model generation can be conveniently discussed in the context of three sequential steps, summarized in Figure 1. In detail, (Ia) MIFs are calculated for the binding sites under study, followed by (Ib) the determination of points of minimum of MIFs (i.e., the points with the lowest interaction energy). Next, (II) the selected GRID points of minimum are converted into pharmacophoric features. Finally, (III) the pharmacophoric model is refined by adding excluded volume spheres.

Ia) MIFs Calculation. The first step of the computational protocol consists of the calculation of the MIFs for the

putative binding site of the target proteins by means of the GRID program.[35] This software initially builds a grid over the molecular region of interest and then computes at each grid point the interaction energies between the protein and several molecular probes. GRID probes incorporate different physicochemical properties and are able to mimic most of the atom types and small moieties commonly found in ligands. The set of energies calculated for a given probe constitutes the MIFs of the probe-target system. In order to well characterize the main interactions between interacting molecules, at least three different probes should be used: a hydrophobic probe (DRY, termed the hydrophobic probe; C3, corresponding to a methyl group; C1=, corresponding to an sp^2 carbon), a hydrogen bond acceptor (O, that is a sp^2 carbonyl oxygen), and a hydrogen bond donor (N1, a neutral flat NH group).

Ib) Determination of Points of Minimum Energy. In the next step, the points of minimum of MIFs are calculated, thus identifying the regions of most favorable interaction between each probe and the protein. The number of points of minimum calculated in the putative binding sites for the selected probes is generally too high to allow the construction of useful pharmacophoric models. This makes necessary the selection of a reduced subset of the most suitable points to be codified into pharmacophoric features. The criteria that should be taken into account for the selection of GRID points of minimum are as follows: (1) The interaction energy: for each probe, the points which most favorably interact with the target protein should be preferred. (2) The position within the binding site: points located within cavities or clefts of the binding region are more suited than points exposed on the protein surface. (3) The distance from other minima: the distance between the selected points should be carefully evaluated, in order to avoid the presence (in the resulting pharmacophoric models) of features that, if too close, could not be simultaneously mapped by the ligand moieties or, if too distant, could force the selection of only large ligands. In this regard, 1 and 15 Å could represent suitable threshold values for minimum and maximum distance between points, respectively. (4) The presence (in the proximity of the point of minimum of N1 and O probes) of a protein functional group able to act as hydrogen bond acceptor or hydrogen bond donor, respectively. Concerning the last criterion, it is important to point out that unlike hydrophobic features, which simply require the presence of a hydrophobic moiety in a given region of space, hydrogen bond donor and acceptor features also contain a directional component. This implies that the construction of such features requires the definition not only of the region of space in which a hydrogen bond donor (or acceptor) group should be located but also of the orientation of such a group, which should be directed toward its hydrogen bond acceptor (or donor) counterpart. On the other hand, the energy calculated by GRID for a probe at each point of the grid depicts the overall interaction between the probe and all the atoms of the protein, and it is therefore the result of several concomitant factors and complex patterns of interactions. Moreover, in many cases, a number of residues, and not only a single amino acid, significantly concurs to the interaction with a given probe. As a consequence, the residues and atoms which play the most important role in determining the value of the interaction energies are not always easy to identify. For this reason, only those points of minimum for N1 and O probes (used to explore hydrogen bond interactions) should be selected for which the protein atoms mainly responsible for the favorable interaction energy are clearly detectable and are located at hydrogen bond distance. In fact, the knowledge of such atoms and their suitable position is strictly necessary to properly define the direction of the hydrogen bond donor and acceptor features corresponding to the points of minimum of N1 and O probes, respectively. When available, information derived from biological or biochemical experiments, such as mutation data or photocross-linking experiment could be of great help for the selection of the most appropriate points of minimum.

II) Generation of Pharmacophoric Models. In the third step, the 3D coordinates of the selected GRID minima were used to arrange the pharmacophoric features. Accordingly, the minimum points of the O probe are converted into hydrogen bond acceptor features, the minimum points of N1 probe are converted into hydrogen bond donor features, and the minimum points of the hydrophobic probes are converted into hydrophobic features. Hydrogen bond donor and acceptor features are directed toward the atoms of the protein which act as hydrogen bond acceptors or donors, respectively. Noteworthy, the radius of the spheres, representing the tolerance of the features, should be set in such a way as to allow a certain degree of tolerance during the search of compounds. In fact, GRID points of minimum should not be intended as rigid spatial constraints but rather as indicators of a region of the binding site that is expected to favorably interact with a given chemical function of ligands. Consistent with this, we have to keep in mind that GRID minima are not isolated spots of favorable interaction in an environment of negligible or unfavorable interactions but are simply the lowest energy representatives of a variable number of neighbor points all bringing common information, namely the presence of a region of favorable interaction between the probe and the target protein.

III) Refinement of Pharmacophoric Models. Finally, to mimic the boundary of the active site, excluded volume spheres are added to the generated 3D pharmacophoric hypothesis. These spheres represent regions of space that cannot be occupied by any ligand portion.

The use of excluded volumes allows for codifying into the structure-based pharmacophoric models information about not only the points of favorable interaction with the binding site, as derived by GRID analysis, but also about the shape and the steric hindrance of the binding site itself (thus exploiting as much as possible the knowledge of the 3D structure of the protein target). In this regard, it has been previously shown that the addition of excluded volume spheres to a pharmacophoric hypothesis could reduce the number of hits and also the number of false positives in a hit list by a factor of 2−5.[36]

RESULTS AND DISCUSSION

Case Studies. To validate the use of the computational protocol described above and to show its usefulness to the discovery of hits toward targets of current therapeutic interest, we have applied it to address three different tasks of medicinal chemistry: 1) the research of small-molecule inhibitors of the thioredoxin reductase (TrxR) enzyme, 2) the discovery of a novel class of compounds able to inhibit

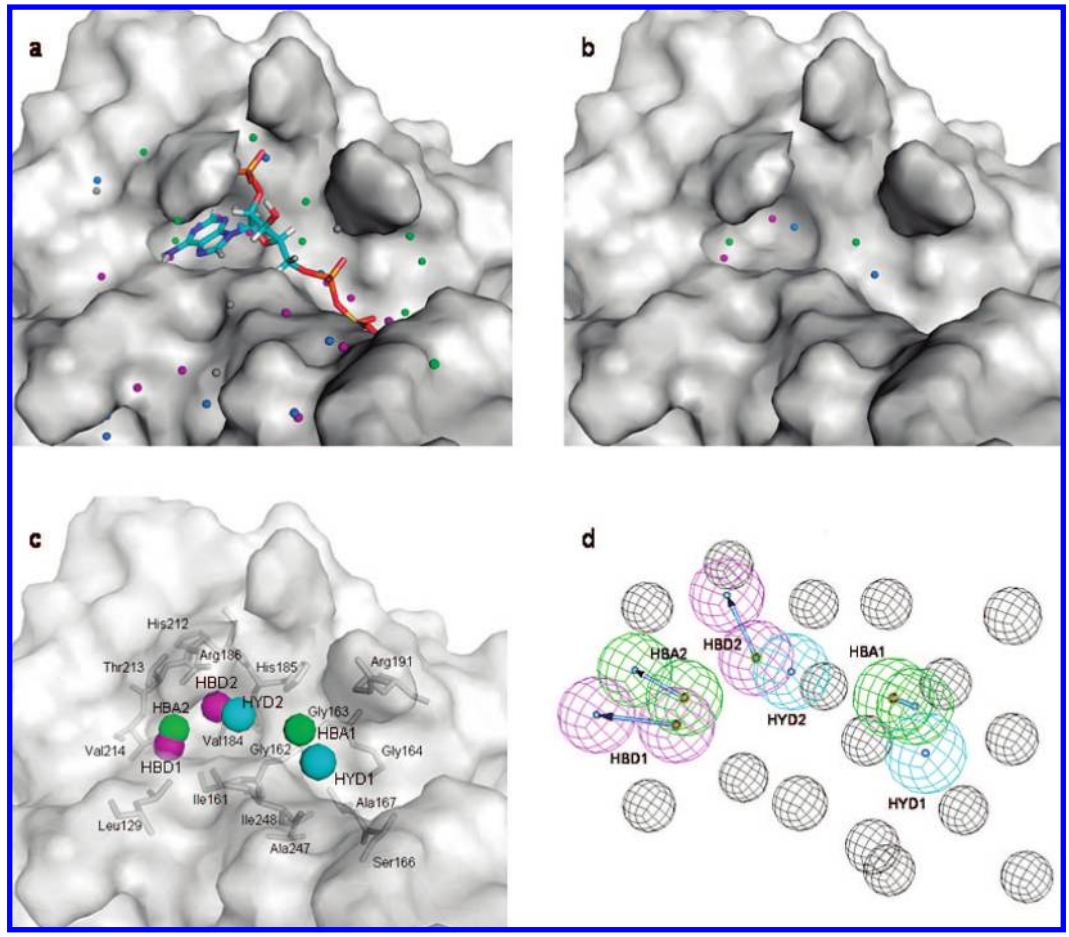


Figure 2. a) GRID points of minimum in the NADPH binding site of TrxR for probes C3 (blue), DRY (gray), N1 (violet), and O (green). NADPH is displayed in stick representation. b) GRID minima selected for pharmacophore generation. c) Functions constituting the six-feature pharmacophoric model. Pharmacophore functions are color-coded (cyan, HYD; violet: HBD; green: HBA); the protein surface is shown in transparency, and residues are used to arrange the excluded volume spheres are represented as sticks. d) Final structure-based pharmacophore, consisting of six features and nineteen excluded volume spheres (colored in black; for sake of clarity, three spheres are omitted).

the formation of the complex between HIV-1 IN and viral DNA, and 3) the research of small molecules affecting the HIV-1 RT dimerization.

In all cases, pharmacophores were used as filters of a more complex virtual screening protocol to select, from the Asinex Gold Collection Database, a restricted number of compounds to be submitted to biological evaluation.

Thioredoxin Reductase. TrxR catalyzes the transfer of reducing equivalents from pyridine nucleotides to disulfide/dithiol compounds and promotes catalysis via the flavin adenine dinucleotide (FAD) and a redox active disulfide. In particular, TrxR catalyzes the reduction of a protein substrate, namely the thioredoxin, by NADPH:

$$\text{thioredoxin-}S_2 + \text{NADPH} + H^+ \rightleftharpoons \text{thioredoxin-}(SH)_2 + NADP^+$$

NADPH, TrxR, and thioredoxin constitute the thioredoxin system, which is ubiquitous from archea to human and is responsible for providing reducing equivalents inside the cells. Another thiol redox system found in cells is the glutathione system, consisting of NADPH, glutathione reductase, glutathione, and glutaredoxin. Similar to the thioredoxin system, the glutathione system utilizes NADPH as its source of reducing equivalents. The glutathione system is absent in MTB. As a consequence, the thioredoxin system is only responsible for maintaining reducing conditions inside the mycobacterial cell, and it is thought to play a fundamental role in the pathogen resistance to the oxidative stress exerted by the innate immune response. This is the reason why the MTB TrxR constitutes an attractive target for antitubercular therapy, even more interesting since no inhibitors of this enzyme have been reported so far.

Our strategy to inhibit the activity of MTB TrxR was aimed at identifying compounds competing with NADPH for its binding site. We could take advantage of the experimentally determined structure of the MTB TrxR (PDB entry: 2A87, 3 Å resolution).[37] Being aware that the lack of selectivity could represent a serious limitation for agents competing with a physiological substrate that, like NADPH, binds to a variety of enzymes and is involved in a multitude of metabolic regulatory processes, we based our search on distinctive elements that were expected to differentiate the NADPH binding site of MTB TrxR from the NADPH binding sites of other enzymes, so as to achieve a selective inhibition.

GRID-Based Binding Site Analysis. The structure of MTB TrxR (chain A) was energy minimized after removal of NADPH, and the NADPH binding region was subjected to GRID analysis. The points of minimum extracted from the

MIFs of C3, DRY, N1, and O are displayed in Figure 2a. Several minima were identified in the zone of interest. However, we took into account only the few points located in the region directly involved in interaction with NADPH and therefore deemed as the most suitable for inhibition of NADPH binding. In particular, we focused on the pocket embedding the adenine moiety of NADPH (hereafter referred to as the adenine binding pocket), since this pocket was better defined than that occupied by the nicotinamide ring of the ligand (referred to as the nicotinamide binding pocket).

Development of Structure-Based Pharmacophores. Despite the restricted number of points of minimum detected within the NADPH binding site, a selection of the most suitable minima was still necessary to avoid the construction of too restrictive pharmacophoric queries. The points of minimum to be translated into pharmacophoric features were selected according to the criteria described above. Noteworthy, an additional criterion was applied in this case, consisting of the bias of the selection toward interaction points that were specific for the binding site and were not exploited by the natural ligand. In this way, we intended to reduce the risk of lack of selectivity of the candidate compounds. The final set of six GRID minima selected for pharmacophore development is shown in Figure 2b. Four of the selected points were concentrated within the adenine binding pocket, while the remaining points were positioned slightly farther away, along the groove that connects the adenine and the nicotinamide binding pockets and that interacts with the ribose and phosphate moieties of NADPH.

The conversion of the selected points into pharmacophoric functions gave rise to a six-feature pharmacophoric model consisting of two hydrophobic (HYD1 and HYD2), two hydrogen bond donor (HBD1 and HBD2), and two hydrogen bond acceptor (HBA1 and HBA2) features (Figure 2c). The projection points of the HBD1 and HBD2 features were placed on the oxygen atom of the backbone carbonyl group of Val214 and His212, respectively, while the projection points for the HBA features were arranged according to the 3D coordinates of the N atom of the backbone NH of Gly163 (HBA1) and Val214 (HBA2). The radius of all the spheres, representing the tolerance of the features, was set to 1.5 Å (the default value). The pharmacophore was completed by the addition of nineteen excluded volume spheres, representing the protein counterpart and characterizing the steric limitations of the binding site. They were placed on the centroids of all the residues surrounding the pharmacophoric functions shown in Figure 2c. In particular, two distinct centroids for backbone and side chain atoms were determined for residues whose side chains were exposed on the surface of the binding pocket and/or markedly concurred to define the shape of the binding site. Given the high number of excluded volume spheres, their radius was lowered from 1.5 (the default value) to 1 Å, with the aim of reducing the restrictiveness that could arise from a too marked excluded volume component in the pharmacophoric models. The final pharmacophoric model is shown in Figure 2d. This was one of the four pharmacophoric queries used in the following database search, the other three being simplified five-feature hypotheses (Table 1) derived from the original model by alternately removing one of the three hydrogen bond functions of the adenine binding pocket. Such features (namely, HBD1, HBD2, and HBA2) were alternately removed, as their concurrent presence could constitute an excessively restrictive requirement, due to their closeness in space. Conversely, the HYD2 feature was kept in all the simplified models to ensure the presence of a hydrophobic moiety within a cavity that naturally accommodates an aromatic nucleus. It is worth noting that the GRID-based protocol used here for structure-based pharmacophore model generation allows for taking into account all the points of favorable interaction between an hypothetical ligand and the protein binding site. In this aspect, it is different from methods that deal with an occupied binding site for deriving the pharmacophore model, since these are strictly dependent on the binding mode of the bound ligand (one pharmacophore model accounts for only one binding mode). Figure 3 shows a comparison between the pharmacophore Hypo 1 described above (Figure 3a) and a pharmacophoric model based on the 2A87 crystallographic structure of the TrxR-NADPH structure, generated by means of LigandScout software (Figure 3b). As shown in Figure 3c, though referring to the same binding site, the two pharmacophoric models are markedly different in terms of both nature and location of their features. The LigandScout pharmacophore contained 6 features: one aromatic ring, two hydrogen bond acceptors, two positive ionizable, one hydrogen bond donor, and nine excluded volumes. These features describe the crucial interactions involving NADPH and the residues Arg187, Arg191, His185, Arg127, Gly163, Ser166, and Arg292. Differently, Hypo 1 was built to mimic favorable interactions with the binding site that are not exploited by NADPH, in order to achieve a selective inhibition toward the TrxR enzyme. In detail, Hypo 1 features are involved in hydrogen-bond interactions with the amino acids Val214, His212, and Gly163 as well as in hydrophobic interactions with His185 and Ile248.

Table 1. Feature Composition of Structure-Based Pharmacophores for NADPH Inhibitors and Results Obtained from the Search of the Asinex Database[a]

search query	pharmacophoric features	no. of hits	% of the Asinex database
Hypo 1	HYD1, HYD2, HBD1, HBD2, HBA1, HBA2	871	0.4
Hypo 2	HYD1, HYD2, HBD1, HBA1, HBA2	9537	4.8
Hypo 3	HYD1, HYD2, HBD2, HBA1, HBA2	17512	8.8
Hypo 4	HYD1, HYD2, HBD1, HBD2, HBA1	1744	0.9

[a] The sum of hits is more than the total of compounds derived from pharmacophore search (Figure 4), as some compounds were selected by more than one hypothesis.

Virtual Screening of the Asinex Database. The structure-based pharmacophoric models generated in the previous step (Table 1) were combined with docking techniques for the virtual screening of the Asinex Gold Collection, a database consisting of over 200000 commercially available compounds.[38] A flowchart depicting the various steps of the virtual screening, including the database filtration and subsequent docking studies, is shown in Figure 4. The pharmacophoric hypotheses were applied as the first filter to screen the whole Asinex Database (Figure 4 and Table 1) and led to the selection of a subset of 22290 compounds (11% of the total entries).

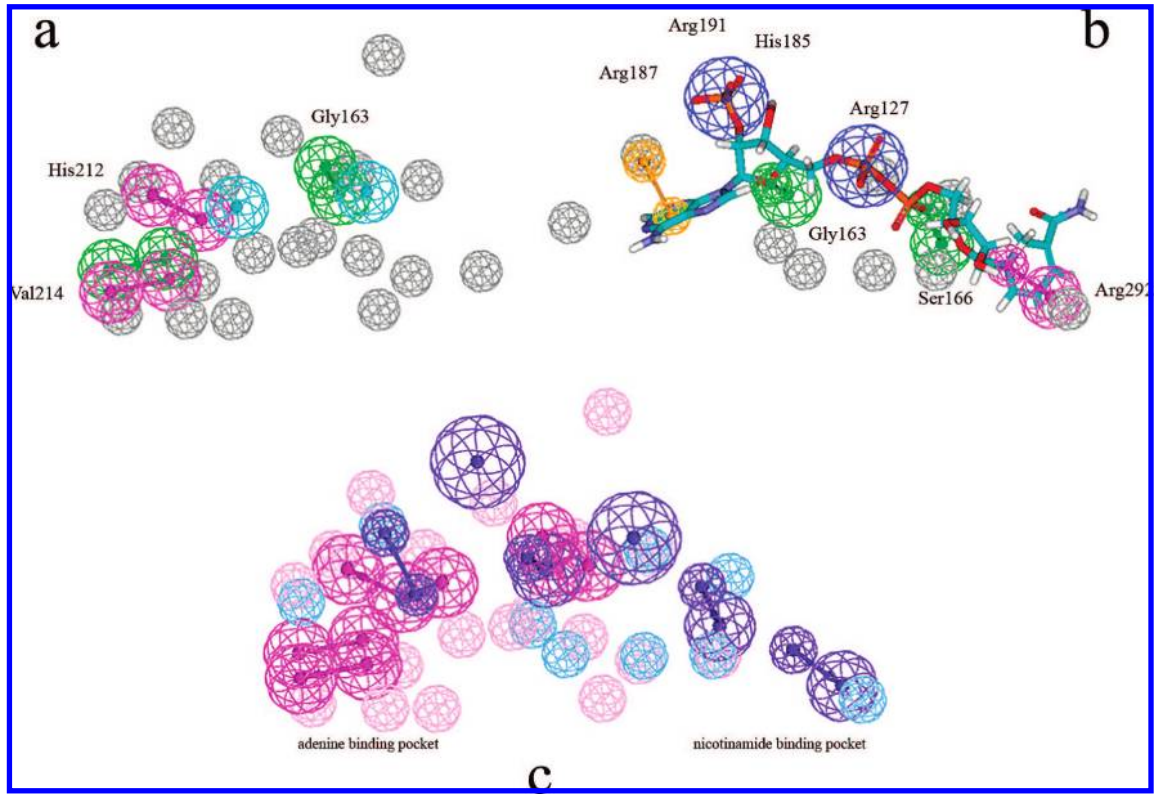


Figure 3. a) Graphical representation of the pharmacophoric hypothesis Hypo 1 consisting of two hydrogen bond acceptors (green), two hydrogen bond donors (magenta), and two hydrophobic features (cyan). b) Graphical representation of LigandScout pharmacophore model derived from the crystallographic complex between NADPH and MTB TrxR. NADPH is displayed in stick representation. The model consists of one aromatic ring (orange), two hydrogen bond acceptors (green), two positive ionizable (blue), and one hydrogen bond donor (magenta). c) Superimposition between Hypo 1 (magenta) and the LigandScout pharmacophore (blue). Both hypotheses have features located in the adenine binding pocket as well as in the groove that connects the adenine and the nicotinamide binding pockets, while only the LigandScout pharmacophore has features positioned in the nicotinamide binding pocket. As a whole, the two pharmacophores express a different interaction pattern with the residues of the binding site.

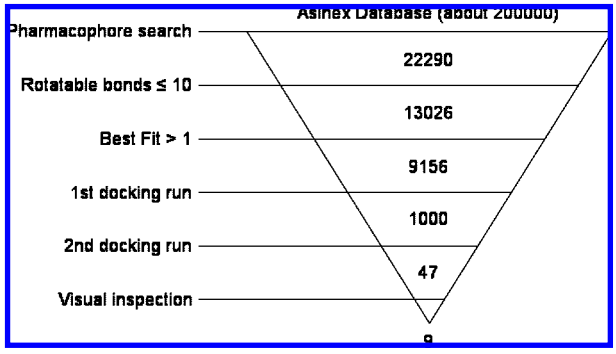
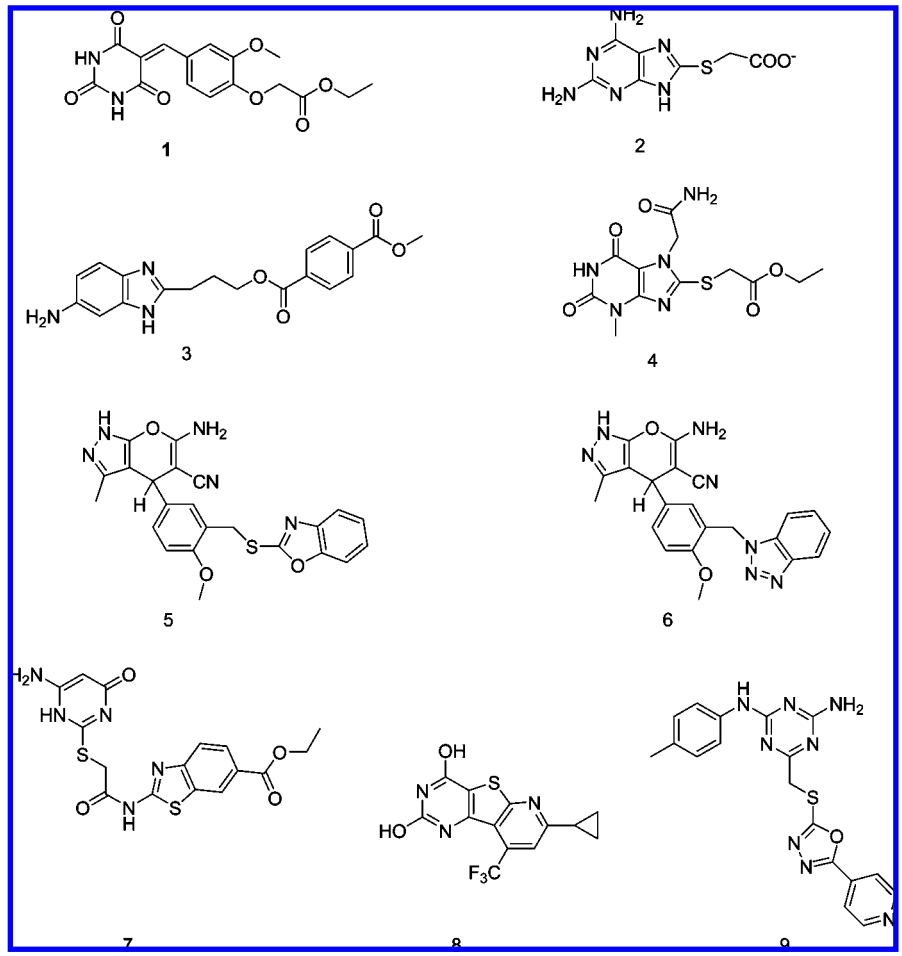


Figure 4. Computational protocol employed for the virtual screening of the Asinex database aimed at identifying potential inhibitors of the NADPH binding site.

The number of rotatable bonds and the Best Fit value (a measure of how well a ligand matches the pharmacophoric features) were used as sequential filters to remove chemicals either too flexible or poorly fitting the pharmacophore models, respectively. It is worth noting that a Best Fit threshold of 1 was considered an appropriate value to keep or discard compounds when the maximum fit value ranged from 4 to 6 (see the Experimental Section).

The resulting database, comprising 9156 compounds, was then subjected to docking analysis, performed with the software GOLD.[39] Docking experiments were carried out in two consecutive runs (Figure 4) to optimize the balance between the quality of docking and the time required for calculations. Accordingly, two different automatic (i.e., ligand dependent) GA parameter settings of Gold were employed. In the first phase of docking studies, the compounds derived from database filtration were analyzed using the Search efficiency parameter set to 50%, thus allowing for speeding up the process and handling a large number of ligands in a reasonable computing time, though losing something in accuracy. The top 1000 (a reasonable number for accurate docking calculations in the next step) scored hits were then subjected to the second docking run, for which the Search efficiency parameter was set to 100%. In this way, the optimal settings for each ligand were applied, slowing down the calculations but making the analysis more exhaustive. The resulting complexes were visually inspected, and only the 47 compounds that were found to bind to the NADPH binding site in good agreement with the pharmacophoric features (and thus well fitting the GRID points of minimum) were retained. Among them, a small subset of nine putative TrxR inhibitors (Chart 1) was finally selected for purchasing, taking into account the following criteria: predicted binding energy, complementarity between the bound ligand and the protein surface in terms of size and shape, and investigation of different structural classes. The selected compounds were purchased from Asinex and submitted to biological testing. Among them, one candidate (compound **1** in Chart 1) endowed with interesting inhibitory activity (IC_{50}=16 μM) toward MTB thioredoxin reductase emerged. Figure 5 shows the superimposition of compound

Chart 1. Structures of Compounds Selected by Virtual Screening Protocol (Figure 4) for Identifying Inhibitors of the NADPH Binding Site of TrxR[a]



[a] Compound **1** emerged as active (the % of inhibition was 90, 85, and 53 at 100, 50, and 25 μM, respectively). Compounds **2**−**9** were inactive (the % of inhibition at 100 μM was lower than 20).

1 on Hypo2 (the simplified pharmacophore through which the ligand has been identified) and its docking pose within the binding site. We can observe that there is a good agreement between pharmacophore mapping and docking results, the main difference being the location of the methoxy substituent on the phenyl ring which is not predicted to be involved in any interaction with the binding pocket.

HIV-1 Integrase. HIV-1 IN is an essential enzyme for viral replication and represents an intriguing target for the development of new drugs.[40−42] Inhibitors active toward IN can be divided into two groups: compounds able to inhibit the 3′-processing reaction by interfering with the interaction of IN with viral DNA (IN binding inhibitors or INBIs) and compounds that preferentially inhibit the strand transfer reaction (IN strand transfer inhibitors or INSTIs). In this context, the aim of our research was the discovery of a novel class of IN inhibitors belonging to the INBIs category and thus acting as inhibitors of the IN-viral DNA complex formation. The DNA binding site of IN has been widely characterized by mutational studies and photocross-linking experiments[43−49] and consists of the residues Lys156, Lys159, Glu152, and Asn155 and a loop formed by the amino acids from 140 to 149. HIV-1 IN is a very difficult system for drug design since it has an active site that is shallow and solvent exposed. In addition, it is also characterized by the very flexible loop 140−149, difficult to be solved.[50] The first crystallographic structures of IN core domains deposited in the Protein Data Bank did not show the loop region. To date, two complete crystal structures of the core domain are available.[51,52] These structures disagree in the previously unsolved regions. In fact, the flexible loop adjacent to the active site is oriented away from the catalytic residues in the crystal structure solved by Goldgur et al.,[52] while the structure solved by Maignan et al.[51] places the flexible loop over the top of the active site and positions Glu152 closer to the other catalytic residues. In 2000, Carlson et al.[53] reported the first receptor-based pharmacophoric model for the active site of IN which takes into account the flexibility of the enzyme by using different conformations of the catalytic core domain derived from MD simulation. However, since the study was completed prior to the availability of the X-ray structures with solved loop regions, the latter structures were not included in that research. In this paper, we report a detailed study of the loop region based on an approach different than those described by other authors. For these calculations, we have used the Macro-Model LOOP tool (a specific conformational search method for protein loops).[54] Starting from the primary sequence of the region to model, it generates a variety of structures that are both diverse and geometrically appropriate for the protein loop and minimizes the energy of these structures. In particular, 50000 conformations were generated, and, among

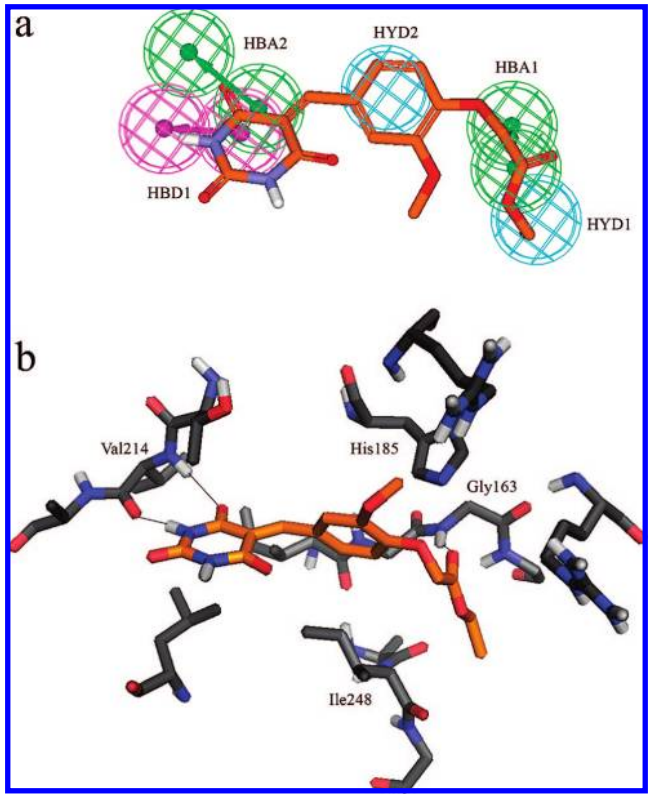


Figure 5. a) Alignment of compound **1** (orange) on Hypo 2. b) Binding mode of compound **1** within the NADPH binding site of MTB TrxR. For the sake of clarity, nonpolar hydrogens are omitted, hydrogen bond interactions are represented by black lines, and only a few residues of the binding site are displayed.

them, only those with an energy that fell within a 10 kcal/mol window above the global minimum were kept (44 conformations). Selected conformations, together with the two crystallographic structures (46 conformations in total), were submitted to a cluster analysis using the g_cluster as implemented in the Gromacs package (version 3.3.1).[55] As a result, 10 clusters were obtained using an rmsd of 0.8 Å (referred to as backbone atoms). The most populated cluster, formed by 25 conformations, contains the structure solved by Maignan et al. (PDB code: 1BL3).[51] The second cluster consists of 6 conformations, while the third one, with 5 conformations, contains the X-ray structure solved by Goldgur et al. (PDB code: 1BIS).[52] Three additional two-membered clusters and four clusters constituted by only one conformation were also found. Therefore, the collection of structures from the LOOP conformational search method appears to provide a reasonable ensemble of configurational states for this study. Three structures, representative of the most populated clusters, were used for next calculations (the two crystallographic conformations and the minimum energy conformer of the second cluster, Figure 6a).

GRID-Based Binding Site Analysis. The three structures of IN were inserted in a Grid box large enough to accommodate all the residues involved in the binding between IN and viral DNA. Then, GRID MIFs were calculated for each structure by using C1=, O, and N1 probes. The IN structures were graphically visualized together with the GRID points of minimum, and only a few points were chosen, according to the criteria described above.

Development of Structure-Based Pharmacophores. The GRID minimum points were used as centroids to position the corresponding pharmacophoric features by means of Catalyst software.[56] Subsequently, excluded volume spheres were added to the 3D pharmacophoric models. The spheres were set on the positions occupied by the Cα of amino acids located within 7 Å from the hypothesis. The radius of the spheres was set at the default value (1.5 Å). Moreover, an excluded volume sphere with a radius of 0.8 Å was located on the Mg^{2+} ion, according to its ionic radius. As a result, three pharmacophores were generated (namely, Hypo1, Hypo2, and Hypo3) from the crystallographic structures 1BL3 and 1BIS and from the best conformation found in the second cluster of conformers above-described. A representation of the final models is shown in Figure 6. The pharmacophores are constituted by hydrogen bond acceptors (green), hydrogen bond donors (magenta), hydrophobic features (cyano), and excluded volume spheres. All the hypotheses are formed by a hydrogen bond acceptor (HBA1) and a hydrogen bond donor (HBD1) feature directed toward the amino group of Lys159 and the backbone carbonyl group of Glu152, respectively. These features fall in a region far from the loop 140−149 and thus are not influenced by its conformation. On the contrary, the models quite differ in the features close to the loop. In detail, Hypo1 consists of two hydrogen bond donors (HBD2 and HBD3) and two hydrophobic (HYD1 and HYD2) features in addition to HBA1 and HBD1, for a total of six features (Figure 6b). The projection points of the HBD1 and HBD2 features were placed on the oxygen atoms of the backbone carbonyl group of Pro142 and Gln148, respectively. Hypo2, a five-feature pharmacophoric model, is formed by a hydrogen bond donor (HBD2) and two hydrophobic (HYD1 and HYD2) features, together with HBA1 and HBD1 (Figure 6c). The projection point of the HBD1 feature was arranged according to the 3D coordinates of the oxygen atom of the backbone carbonyl group of Gly140. Hypo3 consists of a hydrogen bond acceptor (HBA2), a hydrogen bond donor (HBD2), and two hydrophobic (HYD1 and HYD2) features, in addition to HBA1 and HBD1 (Figure 6d). The direction of HBA2 and HBD2 is toward the backbone NH of Val150 and the backbone carbonyl group of Gly140, respectively.

Virtual Screening of the Asinex Database. The receptor-based pharmacophoric models were in turn used to screen the Asinex Gold Collection Database[38] followed by the evaluation of the binding energy of the selected molecules and by docking calculations. In detail, the pharmacophoric hypotheses were applied as the first filter to screen the whole database, without any prefiltering process. Since Hypo1 and Hypo3 failed to retrieve hits from the database, several simplified five-feature pharmacophores were generated from the original models by considering all the possible combinations and used within the virtual screening protocol. As a result, about 20000 compounds were selected. Subsequently, to discard too flexible compounds, the number of rotatable bonds was used as a filter on remaining compounds and only molecules with less than 10 rotatable bonds were kept while the molecules that poorly mapped the pharmacophoric models (Best Fit value <1) were removed. Next, in order to further reduce the number of compounds to be submitted to exhaustive docking calculations, the binding energy of the selected compounds was evaluated by applying the protocol described below.

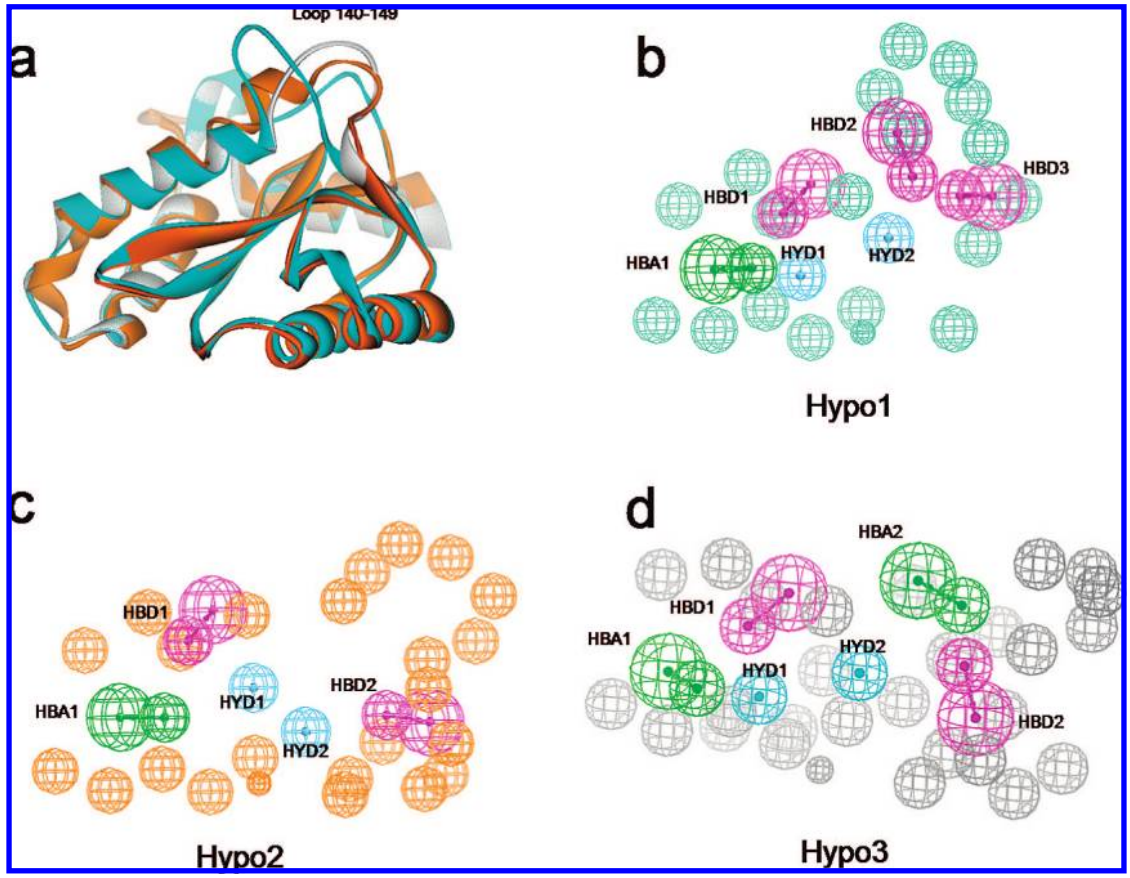


Figure 6. a) 3D structures of IN chosen for receptor-based pharmacophore model generation: 1BL3 (light green), 1BIS (orange), structure derived from conformational search (gray). b) Hypo1. c) Hypo2. d) Hypo3. Pharmacophore functions are color-coded (cyan, HYD; magenta: HBD; green: HBA). The excluded volume spheres are colored light green, orange, and gray for Hypo1, Hypo2, and Hypo3, respectively, in agreement with the color used for the corresponding IN structure. For sake of clarity, some spheres are omitted.

Each ligand identified by a particular pharmacophore was superimposed to it with Catalyst,[56] and the best conformation in terms of fit to the pharmacophore was considered as the bioactive conformation. On this basis, the superimposition pathway of the selected molecules on the pharmacophores also gives the binding mode of these compounds into the IN binding site. Accordingly, the bioactive conformation of each compound was inserted into the IN binding site, and the integrase-ligand complexes were minimized with Macromodel 8.5[54] using the OPLS-AA force-field.[57] Magnesium ion was included in the calculations, whereas water molecules were removed. Then, the interaction energy between the ligands and the target was calculated through four different scoring functions: the scoring function of Autodock 3.0[58] and those implemented in the XSCORE package.[59] Finally, docking studies were performed on the retrieved compounds by means of the GOLD software.

On the basis of three different criterions, (i) the matching between the compound interactions and the pharmacophoric features, (ii) the consensus score, and (iii) the diversity of the structural scaffolds, the first group of ten compounds was submitted to biological investigations and one hit compound was identified. This compound showed inhibitory potency toward IN in both 3′-processing and strand transfer assays (IC_{50} values of 25 μM and 3 μM, respectively), and it was found to be an effective inhibitor of HIV-1 replication in cells with EC_{50} of 30 μM in MT4/MTT experiments (Debyser Z., Witvrouw, M., unpublished data). Biological studies on additional selected compounds are ongoing.

RT Dimerization. The interest of drug design has recently focused on the development of small molecules which interfere with protein−protein interaction.[60,61] In this regard, the computational protocol described herein could be applied to identify the key interactions at the protein−protein interface and to encode them in a pharmacophoric model. This hypothesis could be then used to search a database of ligand candidates in such a way that compounds consistent with the parmacophoric model are selected.

In this regard, the present case study represents an application of the structure-based pharmacophore design protocol for the identification of small molecules that interfere with the HIV-1 RT heterodimer stability. In fact, for the research of new antiviral drugs, the dimerization process is a possible target, involving the coupling between the two subunits of RT (namely, p55 and p66). The subunit association is predominantly mediated by hydrophobic interactions between the two connection subdomains, where a cluster of six tryptophan residues at codons 398, 401, 402, 406, 410, and 414 constitutes the tryptophan repeat motif. Interestingly, this tryptophan cluster is highly conserved among primate lentiviral RTs. Several studies have suggested that the hydrophobic residues in the Trp-motif are important for RT dimerization.[62,63] Therefore, the tryptophan repeat motif represents an attractive drug target for chemotherapeutic intervention, and molecules targeting this motif are likely to inhibit or impair RT maturation. In this context, since no nonpeptidic inhibitors of the RT dimerization targeting the connection subdomains are known until now,

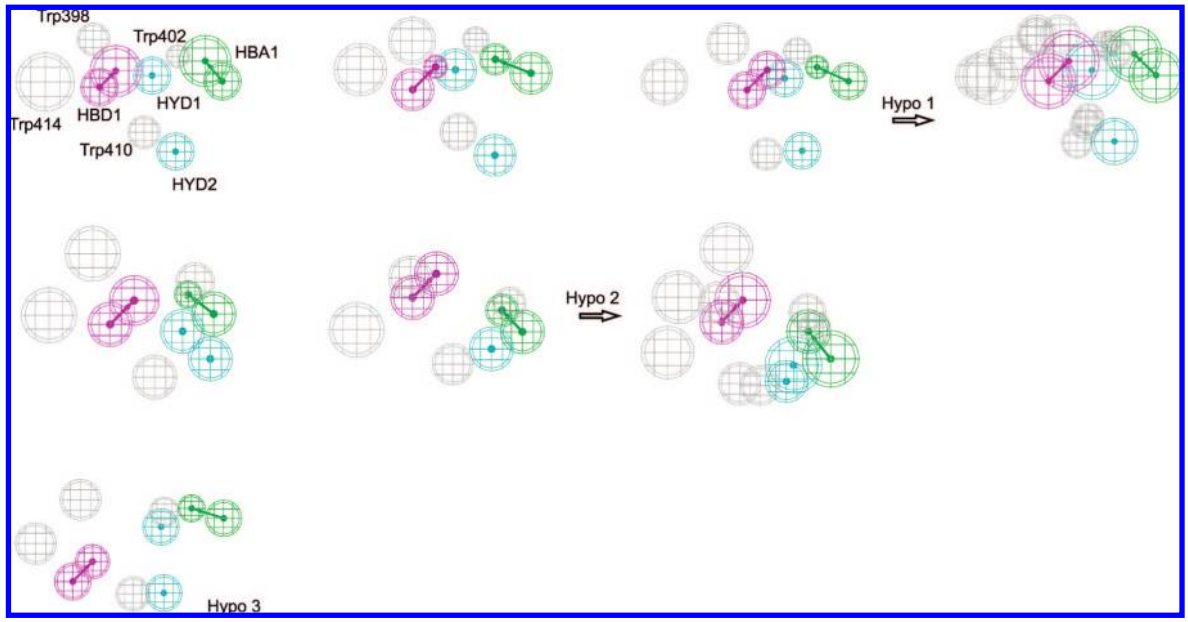


Figure 7. The three structure-based pharmacophoric hypothesis of the p66 connection subdomain were derived from six different frames chosen from MD simulation. Pharmacophoric features are color coded: hydrophobic (HYD), cyan; hydrogen bond donor (HBD), magenta; hydrogen bond acceptor (HBA), green; excluded-volumes, black.

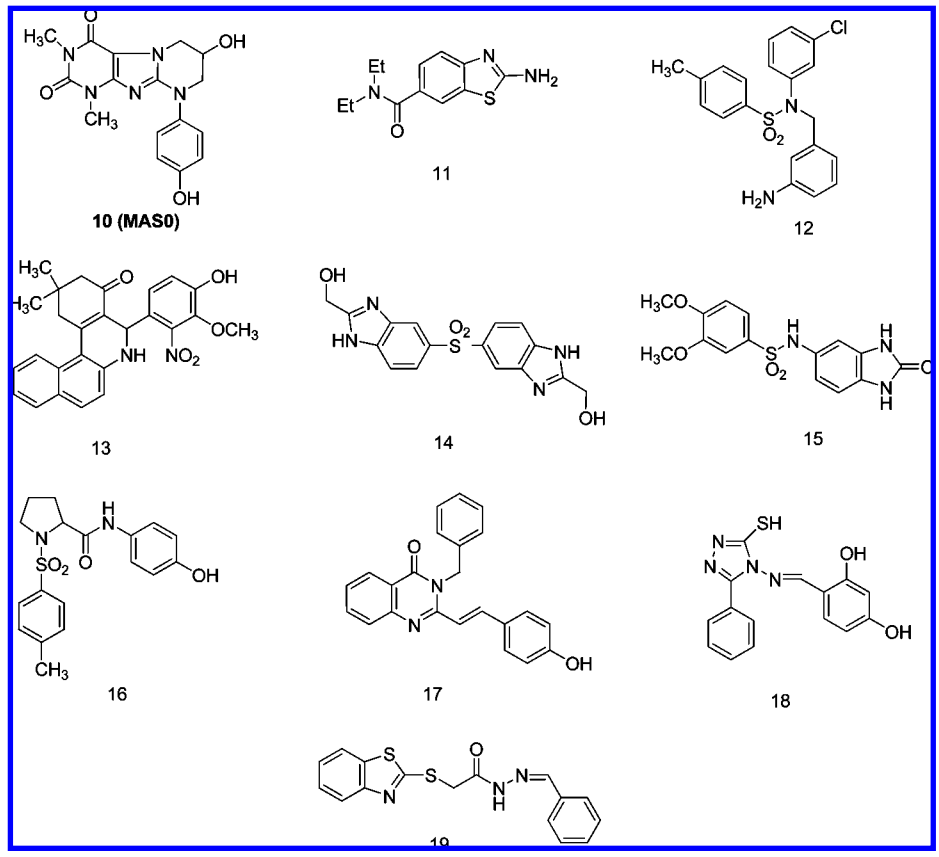
we followed a structure-based ligand design approach to identify compounds disturbing the RT subunit interaction with a mechanism of action targeting the p66 tryptophan cluster. In particular, p66 (taken from the crystallographic structure of the HIV-1 RT, entry 1RTH of the Brookhaven Protein Data Bank)[64] was first submitted to MD calculations by means of the software package NAMD (version 2.5)[65] with the CHARMM27 force field.[66] Hydrogen atoms were added by means of the psfgen package. The p66 subunit was embedded in a sphere of water molecules (60 Å radius) applying spherical boundary conditions. The starting structure was optimized with 1000 steps of conjugate gradient energy minimization to remove unfavorable contacts. MD simulations were carried out at 310 K for 1 ns, collecting snapshot structures every 1 ps. The Langevin Dynamics procedure, with a dumping factor of 5 ps^{-1}, was used to control the temperature.

GRID-Based Binding Site Analysis. From the MD trajectory, six snapshots, characterized by different conformations of relevant residues of the connection subdomain (in particular, Trp402 and Trp410), were chosen. For each snapshot, a 39 × 33 × 14 Å grid (NPLA parameter set to 4), centered on Trp402, was defined. MIFs between residues within the grid and three probes (hydrophobic, DRY; hydrogen bond donor, N1; hydrogen bond acceptor, O) were computed to describe hydrophobic interactions and hydrogen bond contacts.

Development of Structure-Based Pharmacophores. The best interaction points between probes and the putative binding site were identified for each snapshot and converted into pharmacophoric models. Minima identified by computing MIFs between the DRY probe and the tryptophan residues Trp402 and Trp410 corresponded to hydrophobic features HYD1 and HYD2 (defined by aliphatic groups and aromatic moieties). Minima for N1 and O probes were replaced by the center of hydrogen bond donor (HBD1) and acceptor (HBA1) features, respectively. Hydrogen bond acceptor (HBA1) feature is directed toward the NH group of the indolic side chain of Trp402, while the hydrogen bond donor (HBD1) feature is directed toward the backbone carbonyl group of Thr409, respectively. Excluded-volume spheres (corresponding to residues Trp398, Trp402, Trp410, and Trp414) were added to each of the six models to better define size and shape of the binding site and to avoid the identification of compounds during the next virtual screening procedure that may overlap portions of the protein, thus causing steric protein−ligand clashes. Finally, the six pharmacophoric models derived from the selected MD snapshots were merged into three pharmacophoric hypotheses using the Merge Hypothesis/Features option of the View Hypothesis Workbench module of the software Catalyst.[56] Merging was carried out on the basis of the distance tolerance value. As a result, Hypo1 was obtained by merging pharmacophoric models derived from three snapshots, Hypo2 was obtained by merging models from two snapshots, while Hypo3 corresponded to one pharmacophoric model as obtained from the corresponding MD snapshot (Figure 7).

Virtual Screening of the Asinex Database. The three pharmacophoric models were used as three-dimensional queries to perform a virtual screening of databases of commercially available compounds. With this aim, entries of the Asinex Gold Collection[38] able to fit at least one of the pharmacophoric models were retrieved. As a result, about 39000, 11000, and 600 entries were kept by Hypo1, Hypo2, and Hypo3, respectively. Selected compounds were then filtered on the basis of Lipinski's rule-of-five, to retain only druglike entries. Finally, only compounds showing the highest fit value to the pharmacophoric models were kept (fit value >1). As a whole, 91 compounds were selected for docking studies from Hypo1, 19 from Hypo2, and 4 from Hypo3 and were docked into the p66 connection subdomain by means of the Autodock software. Three different protein structures were used for docking calculations. In particular, for compounds selected with Hypo1, the MD snapshot corresponding to the first pharmacophoric model used to define Hypo1 was employed. For compounds selected with

Chart 2. Structures of Compounds Selected by Virtual Screening Protocol To Identify Molecules Affecting the HIV-1 RT Dimerization[a]



[a] Compound **10**, also namely MAS0, was identified as a hit compound. Compounds **11**−**19** were not able to reduce the association of the two RT subunits and to affect either the polymerase or the RNase H activity.[67]

Hypo2, docking calculations were performed using the MD snapshot corresponding to the first model that defined Hypo2. Finally, compounds derived from Hypo3 were docked into the MD snapshot used to build Hypo3.

The structure-based ligand design approach revealed 10 compounds (Chart 2) which were initially screened applying a well established redimerization assay. The pharmacophores-based virtual screening procedure and additional filters applied in this case study constituted a successful procedure to identify compounds affecting RT heterodimer stability. In fact, one of these molecules (MAS0, compound **10** in Chart 2) strongly reduced the association of the two RT subunits. 50% inhibition of RT reassociation was observed at 150 μM. MAS0 shows a dose-dependent, simultaneous inhibition of both the polymerase as well as the RNase H of the RT, following preincubation with t1/2 of about 2 h, yielding IC_{50} values of 155 and 111 μM, respectively. MAS0 is supposed to interfere with the dimer equilibrium by shifting the equilibrium from an active to an inactive dimer by keeping the enzyme in the inactive conformation. This study represents the first reported successful rational screen for non peptidic RT dimerization inhibitors that target the connection subdomain of the two RT subunits and that can represent attractive hit compounds for the development of novel therapeutic agents.[67]

CONCLUSIONS

In this work, a variety of computational approaches have been combined into a multistep procedure for the generation and use of receptor-based pharmacophoric models. The resulting protocol, based on the application of GRID (a software commonly used and widely appreciated in drug design) in combination with other techniques such as docking studies, molecular dynamics, and conformational search, is expected to be a very useful tool in drug discovery when no compounds targeting the binding site of interest are known. Interestingly, this approach permits facing in an easy and fast manner a critical aspect of the molecular modeling such as the flexibility of the binding site. Some possible applications of such a procedure have been discussed in the text by the description of three different case studies. In the field of medicinal chemistry, good results have been obtained in each of the explored case studies, thus confirming the reliability of this approach. Finally, results reported here confirm GRID as a very reliable and versatile software, with an extreme flexibility and a wide range of possible applications.

EXPERIMENTAL SECTION

GRID-Based Binding Site Analysis. Computation of MIFs over the putative binding sites was carried out with software GRID, version 22.[35] Box dimensions were defined so as to accommodate all the residues constituting the binding site of interest. The determination of MIF points of minimum required the execution of GRID calculations separately for each probe and the use of the *export GRIDKONT* option. The KONT files corresponding to each MIF were processed by means of the Minim and Filmap programs (both implemented in the GRID package), which collect all points within a certain energy threshold value, allowing the interpolation of the closest ones. We have used a threshold of 0 Kcal/

mol, in a way to exclude all the points with a positive energy of interaction. The result of this procedure is a file with a .pdb extension containing the 3D coordinates of the points of minimum of the interaction energy, together with the corresponding energy values.

Construction of Structure-Based Pharmacophores. The pharmacophoric models were generated using the software Catalyst, version 4.10.[56] The GRID points of minimum selected for probes C3, C1= or DRY, N1, and O were converted into hydrophobic, hydrogen bond donor, and hydrogen bond acceptor features, respectively. The pharmacophoric functions were arranged according to the 3D coordinates of the corresponding GRID minima (which, in turn, were reported in the PDB files generated by the Minim and Filmap programs); the coordinates of the protein atoms chosen as projection points as well as the centroids of residues for excluded volume spheres were derived from the structures of the target proteins. The structure-based pharmacophores were interactively built using the View Hypothesis workbench in Catalyst. Functions were first selected from the feature dictionary, and then their 3D coordinates and tolerance were set using the Constraint Tolerance control panel (*Constraints/Define Constraint/Location* menu item). The coordinates and the radius of the excluded volume spheres were set through the Excluded Volume Tolerance control panel (*Constraints/Add Excluded Volume* menu item). The View Hypothesis workbench was also used for the removal of functions from the original pharmacophores to obtain the simplified hypotheses. The weight of each feature in pharmacophoric model was set to 1; therefore, the maximum fit value of any ligand alignment with these models is equal to the total number of features forming the hypothesis.

Database Search. The database search by means of pharmacophoric models was performed with Catalyst. The Asinex Gold Collection was downloaded in SDF format from the Web and converted into a database of 3D structures with the Catalyst *catDB* command (FAST method, maximum number of conformers = 250); in such a way, a conformational model consisting of a maximum of 250 conformers was generated for each compound, so as to reproduce the flexibility of molecules during the database search. The resulting database was stored in the Catalyst database format (BDB). The View Database workbench was used to compute the number of rotatable bonds and the Best Fit values and to discard compounds falling outside the threshold values.

Docking Studies. Docking studies were performed on the previously selected compounds within the TrxR binding site as well as into the DNA binding domain of IN using the software package Gold 3.0.1.

The ChemScore was chosen as fitness function. The GA parameter settings of Gold were employed using the Search efficiency set to 100%. In the TrxR case study, preliminary docking studies were performed with the Search efficiency set to 50% in order to reduce the number of compounds to be submitted to more exhaustive docking calculations. Finally, results differing less than 1.5 Å in ligand-all atom rmsd were clustered together. For each inhibitor, the first ranked solution as well as the lowest energy conformation of the most populated cluster were analyzed by comparing the binding mode with the pharmacophoric model mapping. On the other hand, Autodock 3.0.5 was used for docking calculations in the p66 RT subunit. Since Autodock precalculates grid maps to represent the receptor when calculating the interaction energy with a ligand, this software was chosen to check if an agreement does exist between results of the two grid-based approaches (MIFs-based calculations and docking studies), required by the complex binding site topology (see below). To prepare the input structures of the selected compounds for docking calculations, a geometry optimization was performed (using the Gaussian03 program,[68] semiempirical Hamiltonian AM1, Gaussian, Inc., Wallingford, CT, U.S.A.), according to a procedure already reported elsewhere.[69,70] Charges were computed by a Hartree−Fock calculation with a 6 31G(d) basis set, according to the Merz−Singh−Kollman procedure. Finally, the structures of the compounds, together with charge values, were imported in AutoDockTools (ADT, graphical user interface for Autodock) to automatically define rigid root and rotatable bonds. The protein structures were imported in ADT and manipulated by removing nonpolar hydrogens, while Kollman united-atom partial charges and solvent parameters were added. The Lamarckian genetic algorithm (LGA) was employed to explore the possible orientations/conformations of the inhibitors into the binding site. For each compound, the following protocol was applied: 200 independent LGA runs, a population size of 400 individuals, and a maximum number of 1000000 energy evaluations. Results differing by less than 1 Å in positional root-mean-square deviations were clustered together. Results of the docking simulations were analyzed on the basis of the cluster analyses and the values of the binding/docking energy. It is also important to point out that the final hit list was generated according to either pharmacophore mapping or docking results. In fact, compounds with a docked binding mode in disagreement with their alignment to the pharmacophore were also considered as putative hits, if their docked poses were consistent with the grid maps precalculated by means of the docking software. This means that a more permissive principle was applied in this case to allow the choice of hit compounds to be tested. This was necessary because of the binding site topology. In fact, the putative binding site for compounds able to affect the RT heterodimer stability is not a classical active site or pocket, but it is a portion of the protein surface where the two RT subunits interface and establish protein−protein interactions. Moreover, the three pharmacophoric hypotheses applied in this virtual screening protocol are relatively simple (if compared with the two previous case-studies), having only four excluded-volumes and four features, selected on the basis of available mutagenesis data. Pharmacophoric models were derived from minimum energy points of previous calculated MIFs, and only a restricted set of minima points was used to generate the features. In that way, orientation of compounds on the pharmacophore can represent only one of the possible binding modes of the selected entries against the protein surface. Consequently, in this case, the conformational space of complexes between selected compounds and the target protein can be better explored using the docking procedure. For this reason, compounds satisfying both the pharmacophores mapping and showing a reasonable docking result (in terms of docking maps, cluster analysis, and energy values) were selected for biological assays.

ACKNOWLEDGMENT

This study was supported by grants from the European TRIoH Consortium (LSHB-2003-503480). We would like to thank Asinex for partial support of this work. We are indebted to Molecular Discovery for access to the GRID code. Financial support from the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN 2005037820) is gratefully acknowledged. We would like to thank Zeger Debyser and Myriam Witvrouw (Katholieke Universiteit Leuven, Belgium), Tobias Restle and Dina Grohmann (Institut für Molekulare Medizin Universitätsklinikum Schleswig-Holstein, Campus Lübeck Ratzeburger Allee, Lübeck, Germany), and Leopold Flohé and Timo Jaeger (MOLISA GmbH, Molecular Links Sachsen-Anhalt, Germany) for biological data on HIV-1 IN, RT dimerization, and TrxR, respectively.

REFERENCES AND NOTES

(1) Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting Molecular Interactions *in Silico*: I. A Guide to Pharmacophore Identification and its Application to Drug Design. *Curr. Med. Chem.* **2004**, *11*, 71–90.

(2) Guner, O. F. The Impact of Pharmacophore Modeling in Drug Design. *IDrugs* **2005**, *7*, 567–572.

(3) Khedkar, S. A.; Malde, A. K.; Coutinho, E. C.; Srivastava, S. Pharmacophore Modeling in Drug Discovery and Development: an Overview. *Med. Chem.* **2007**, *2*, 187–197.

(4) Steindl, T. M.; Schuster, D.; Laggner, C.; Langer, T. Parallel Screening: a Novel Concept in Pharmacophore Modeling and Virtual Screening. *J. Chem. Inf. Model.* **2006**, *46*, 2146–2157.

(5) Steindl, T. M.; Schuster, D.; Wolber, G.; Laggner, C.; Langer, T. High-Throughput Structure-Based Pharmacophore Modelling as a Basis for Successful Parallel Virtual Screening. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 703–715.

(6) Steindl, T. M.; Schuster, D.; Laggner, C.; Chuang, K.; Hoffmann, R. D.; Langer, T. Parallel Screening and Activity Profiling with HIV Protease Inhibitor Pharmacophore Models. *J. Chem. Inf. Model.* **2007**, *47*, 563–571.

(7) Markt, P.; Schuster, D.; Kirchmair, J.; Laggner, C.; Langer, T. Pharmacophore Modeling and Parallel Screening for PPAR Ligands. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 575–590.

(8) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A Fast New Approach to Pharmacophore Model Mapping and Its Application to Dopaminergic and Benzodiazepine Agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102.

(9) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. Identification of Common Functional Configurations Among Molecules. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571.

(10) Kurogi, Y.; Güner, O. F. Pharmacophore Modeling and Three-Dimensional Database Searching for Drug Design Using Catalyst. *Curr. Med. Chem.* **2001**, *9*, 1035–1055.

(11) Jones, G.; Willett, P.; Glen, R. C. A Genetic Algorithm for Flexible Molecular Overlay and Pharmacophore Model Elucidation. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 532–549.

(12) Patel, Y.; Gillet, V. J.; Bravi, G.; Leach, A. R. A Comparison of the Pharmacophore Model Identification Programs: Catalyst, DISCO and GASP. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 653–681.

(13) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169.

(14) Chen, J.; Lai, L. Pocket v.2: Further Developments on Receptor-Based Pharmacophore Modeling. *J. Chem. Inf. Model.* **2006**, *46*, 2684–2691.

(15) Ortuso, F.; Langer, T.; Alcaro, S. GBPM:GRID-Based Pharmacophore Model: Concept and Application Studies to Protein-Protein Recognition. *Bioinformatics* **2006**, *22*, 1449–1455.

(16) Böhm, H.-J. The Computer Program LUDI: a New Method for the De Novo Design of Enzyme Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.

(17) Deng, J.; Sanchez, T.; Neamati, N.; Briggs, J. M. Dynamic Pharmacophore Model Optimization: Identification of Novel HIV-1 Integrase Inhibitors. *J. Med. Chem.* **2006**, *49*, 1684–1692.

(18) Fox, T.; Haaksma, E. E. J. Computer Based Screening of Compound Databases: 1. Preselection of Benzamidine-Based Thrombin Inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 411–425.

(19) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: a Program for Structure Generation. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 127–153.

(20) Murray, C. W.; Clark, D. E.; Auton, T. R.; Firth, M. A.; Li, J.; Sykes, R. A.; Waszkowycz, B.; Westhead, D. R.; Young, S. C. PRO_SELECT: Combining Structure-Based Drug Design and Combinatorial Chemistry for Rapid Lead Discovery. 1. Technology. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 193–207.

(21) Murray, C. M.; Cato, S. J. Design of Libraries to Explore Receptor Sites. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207–217.

(22) Eksterowicz, J. E.; Evensen, E.; Lemmen, C.; Brady, G. P.; Lanctot, J. K.; Bradley, E. K.; Saiah, E.; Robinson, L. A.; Grootenhuis, P. D.; Blaney, J. M. Coupling Structure-Based Design with Combinatorial Chemistry: Application of Active Site Derived Pharmacophores with Informative Library Design. *J. Mol. Graphics Modell.* **2002**, *20*, 469–477.

(23) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(24) Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G. New and Original pK(a) Prediction Method Using Grid Molecular Interaction Fields. *J. Chem. Inf. Model.* **2007**, *6*, 2172–2181.

(25) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *2*, 279–294.

(26) Zamora, I.; Afzelius, L.; Cruciani, G. Predicting Drug Metabolism: a Site of Metabolism Prediction Tool Applied to the Cytochrome P450 2C9. *J. Med. Chem.* **2003**, *12*, 2313–2324.

(27) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. Grid-Independent Descriptors (GRIND): a Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *17*, 3233–3243.

(28) Crivori, P.; Zamora, I.; Speed, B.; Orrenius, C.; Poggesi, I. Model Based on GRID-Derived Descriptors for Estimating CYP3A4 Enzyme Stability of Potential Drug Candidates. *J. Comput.-Aided Mol. Des.* **2004**, *3*, 155–166.

(29) Mason, J. S.; Cheney, D. L. Ligand-Receptor 3D Similarity Studies Using Multiple 4-Point Pharmacophores. *Pac. Symp. Biocomput.* **1999**, *4*, 456–467.

(30) Mason, J. S.; Cheney, D. L. Library Design and Virtual Screening Using Multiple 4-Point Pharmacophore Fingerprints. *Pac. Symp. Biocomput.* **2000**, *5*, 573–584.

(31) Ortuso, F.; Alcaro, S. Langer, T. GRID-Based Pharmacophore Models: Concept and Application Examples. In *Methods and Principles in Medicinal Chemistry: Pharmacophores and Pharmacophore Searches*, 1st ed.; Langer, T., Hoffmann R. D., Eds.; Wiley-VCH: Weinheim, Germany, 2006; Vol. 32, pp 151−170.

(32) Polgar, T.; Keseru, G. M. Virtual Screening for Beta-Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749–3755.

(33) Evers, A.; Klabunde, T. Structure-Based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the Alpha1A Adrenergic Receptor. *J. Med. Chem.* **2005**, *48*, 1088–1097.

(34) Alberts, I. L.; Todorov, N. P.; Källblad, P.; Dean, P. M. Ligand Docking and Design in A Flexible Receptor Site. *QSAR Comb. Sci.* **2005**, *24*, 503–507.

(35) GRID, version 22b; Molecular Discovery Ltd.: Pinner, Middlesex, U.K., 2004.

(36) 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.

(37) Akif, M.; Suhre, K.; Verma, C.; Mande, S. C. Conformational Flexibility of Mycobacterium Tuberculosis Thioredoxin Reductase: Crystal Structure and Normal-Mode Analysis. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 1603–1611.

(38) http://www.asinex.com/prod/gold.html (accessed June 2007).

(39) (a) Jones, G.; Willett, P.; Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53. (b) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748. (c) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609–623.

(40) Nair, V. Novel Inhibitors of HIV Integrase: the Discovery of Potential Anti-HIV Therapeutic Agent. *Curr. Pharm. Des.* **2003**, *9*, 2553–2565.

(41) Pluymers, W.; De Clercq, E.; Debyser, Z. HIV-1 Integration as a Target for Antiretroviral Therapy: a Review. *Curr. Drug Targets Infect. Disord.* **2001**, *1*, 133–149.
(42) Gupta, S. P.; Nagappa, A. N. Design and Development of Integrase Inhibitors as Anti-HIV Agents. *Curr. Med. Chem.* **2003**, *10*, 1779–1794.
(43) Drelich, M.; Wilhelm, R.; Mous, J. Identification of Amino Acid Residues Critical for Endonuclease and Integration Activities of HIV-1 Integrase Protein *in Vitro*. *Virology* **1992**, *188*, 459–468.
(44) Engelman, A.; Craige, R. Identification of Conserved Amino Acid Residues Critical for Human Immunodeficiency Virus Type 1 Integrase Function *in Vitro*. *J. Virol.* **1992**, *66*, 6361–6369.
(45) Kulkosky, J.; Jones, K. S.; Kats, R. A.; Mack, A. M. S. Residues Critical for Retroviral Integrative Recombination in a Region That Is Highly Conserved Among Retroviral/Retrotransposon Integrases and Bacterial Insertion Sequence Transposases. *Mol. Cell. Biol.* **1992**, *12*, 2331–2338.
(46) Leavitt, A. D.; Shiue, L.; Varmus, H. E. Site-Directed Mutagenesis of HIV-1 Integrase Demonstrates Differential Effects on Integrase Functions *in Vitro*. *J. Biol. Chem.* **1993**, *268*, 2113–2119.
(47) Gerton, J. L.; Ohgi, S.; Olsen, M.; Derisi, J.; Brown, P. O. Effects of Mmutations in Residues Near the Active Site of Human Immunodeficiency Virus Type 1 Integrase on Specific Enzyme-Substrate Interactions. *J. Virol.* **1998**, *6*, 5046–5055.
(48) Heuer, T. S.; Brown, P. O. Mapping Features of HIV-1 Integrase Near Selected Sites on Viral and Target DNA Molecules in an Active Enzyme-DNA Complex by Photo-Cross-Linking. *Biochemistry* **1997**, *36*, 10655–10665.
(49) Dirac, A. M. G.; Kjems, J. Mapping DNA-Binding Sites of HIV-1 Integrase by Protein Footprinting. *Eur. J. Biochem.* **2001**, *268*, 743–751.
(50) Greenwald, J.; Le, V.; Butler, S. L.; Bushman, F. D.; Cjoe, S. The Mobility of an HIV-1 Integrase Active Site Loop Is Correlated with Catalytic Activity. *Biochemistry* **1999**, *38*, 8892–8898.
(51) Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. Crystal Structures of the Catalytic Domain of HIV-1 Integrase Free and Complexed with its Metal Cofactor: High Level of Similarity of the Active Site with Other Viral Integrases. *J. Mol. Biol.* **1998**, *282*, 359–368.
(52) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. Three New Structures of the Core Domain of HIV-1 Integrase: an Active Site That Binds Magnesium. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150–9154.
(53) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, A. J. Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.
(54) *Macromodel, version 8.5*; Schrodinger, LLC: Portland, OR, 2003.
(55) (a) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56. (b) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
(56) Catalyst, version 4.10; Accelrys, Inc.: San Diego, CA, 2005.
(57) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(58) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Alghorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(59) Wang, R.; Lai, L.; Wang, S. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.
(60) Ma, B.; Nussinov, R. Trp/Met/Phe Hot Spots in Protein-Protein Interactions: Potential Targets in Drug Design. *Curr. Top. Med. Chem.* **2007**, *10*, 999–1005.
(61) Hershberger, S. J.; Lee, S. G.; Chmielewski, J. Scaffolds for Blocking Protein-Protein Interactions. *Curr. Top. Med. Chem.* **2007**, *10*, 928–942.
(62) Mulky, A.; Sarafianos, S. G.; Jia, Y.; Arnold, E.; Kappes, J. C. Identification of Amino Acid Residues in the Human Immunodeficiency Virus Type-1 Reverse Transcriptase Tryptophan-Repeat Motif That Are Required for Subunit Interaction Using infectious Virions. *J. Mol. Biol.* **2005**, *349*, 673–684.
(63) Wapling, J.; Moore, K. L.; Sonza, S.; Mak, J.; Tachedjian, G. Mutations That Abrogate Human Immunodeficiency Virus Type 1 Reverse Transcriptase Dimerization Affect Maturation of the Reverse Transcriptase Heterodimer. *J. Virol.* **2005**, *79*, 10247–10257.
(64) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High Resolution Structures of HIV-1 RT from Four RT-Inhibitor Complexes. *Nat. Struct. Biol.* **1995**, *4*, 293–302.
(65) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
(66) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(67) Grohmann, D.; Corradi, V.; Elbasyouny, M.; Baude, A.; Horenkamp, F.; Laufer, S. D.; Manetti, F.; Botta, M.; Restle, T. Small Molecule Inhibitors Targeting HIV-1 Reverse Transcriptase Dimerization. *ChemBioChem* **2008**, *9*, 916–922.
(68) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, revision C.02*; Gaussian Inc.: Wallingford, CT, 2004.
(69) Saparpakorn, P.; Kim, J. H.; Hannongbua, S. Investigation on the Binding of Polycyclic Aromatic Hydrocarbons with Soil Organic Matter: a Theoretical Approach. *Molecules* **2007**, *12*, 703–715.
(70) Hemmateenejad, B.; Tabaei, S. M. H.; Namvaran, F. Computer-Aided Design of Potential Anti-HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors by Contraction of β-Ring in TIBO Derivatives. *J. Mol. Struct. (Theochem)* **2005**, *732*, 39–45.

CI800105P