Discovery of HIV-1 Integrase Inhibitors through a Novel Combination of Ligand and Structure-based Drug Design

A. Brigo[†,1], G. Iurcu Mustata[§,2], J.M. Briggs[2] and S. Moro*[,1]

[1]*Dipartimento di Scienze Farmaceutiche, Università degli Studi di Padova, 35131 Padova, Italy*
[2]*Department of Biology and Biochemistry, University of Houston, Houston, Texas 77024-5001, USA*

Abstract: Over the past 10 years, classical computer-aided molecular design methods have not been frequently applied for the discovery of novel HIV-1 integrase (IN) inhibitors, due to the intrinsic challenges that this enzyme presents. Therefore, a novel approach that combines the chemical information of known integrase inhibitors with the enzyme's detailed 3D structure in a stepwise fashion is proposed: (I) use of a pharmacophore model (PM), which takes into account in a weighted fashion the chemical features of known ligands, in analogous manner to the to search the Maybridge and the NCI 3D databases; (II) drug-likeness optimization; (III) virtual high-throughput screening of the hits matching the PM query against 1QS4 wild-type IN structure using different Docking/Scoring combinations; (IV) visual inspection and selection of the hits in function of: binding free energies; binding mode type within the active site; retrieval among the best 20% hits in more than 6 Docking/Scoring protocols at the same time. This approach aims at a rational selection of new potential HIV-1 integrase inhibitors.

Key Words: High Throughput Docking; HIV-1 Integrase; Diketo Acid HIV-1 Integrase inhibitors.

INTRODUCTION

Human immunodeficiency virus (HIV) type 1 is the etiological agent of acquired immunodeficiency syndrome (AIDS). HIV encodes three enzymes: reverse transcriptase, protease and integrase. Only the first two have been successfully exploited as targets for antiviral drug therapy. The emergence of strains resistant to currently available reverse transcriptase and protease inhibitors has led to the necessity to study new targets for antiretroviral therapy. An essential step in HIV replication is the integration of the transcribed double-stranded DNA into the host chromosomes, carried out by the integrase protein[1,2]. Integration occurs in two subsequent reactions [3]: in the first step, termed 3'-processing, an activated water molecule attacks at the 3'-ends of the viral DNA removing a terminal di-nucleotide from each 3'-end; in the second step, called "strand transfer", each exposed viral DNA 3'-OH ribose is activated for nucleophilic attack on the opposite strands of the host DNA becoming covalently attached to it. Divalent metal ions such as Mg^{2+} and Mn^{2+} are required for both of the reactions and for the formation of the IN complex with viral DNA, competent to carry out either process [4,5]. Given the essential role of integrase in the production of viral proteins and RNA progeny, it has become an attractive target for the design of new drugs, that would be a valuable complement to reverse transcriptase and protease inhibitors in the combined therapy of AIDS.

*Address correspondence to this author at the Dipartimento di Scienze Farmaceutiche, Università degli Studi di Padova, 35131 Padova, Italy; Tel: (++39)-049-8275704; Fax: (++39)-049-8275704;

E-mail: stefano.moro@unipd.it

[§]Present address: Emisphere Technologies, Inc., Tarrytown, NY 10591, USA

†Present address: The Genetics Company, Inc., 8952 Schlieren, Zürich, Switzerland

Over the past few years, many groups have worked on developing new HIV-1 IN inhibitors. A large number of classes of compounds, such as dioxepinones [6], quinones [6], benzoic hydrazides [7], thiazolo-thiazepines [8], salicyl-pyrazolinones [9],coumarins [10], etc. [11,12], have been reported to inhibit integrase enzymatic activity at low micromolar concentrations acting either against the 3'-processing or the strand transfer reactions.

Unfortunately, none of them are currently in the market due to either their limited potency or high toxicity. Recently, a series of compounds, typified by an aryl β-diketo motif, have been identified by Merck. This class of molecules prevents the HIV-1 replication in cells by selectively inhibiting the strand transfer process [13]. These molecules, alongwith newly published naphthyridine-[14] and azido-derivatives [15], represent one of the most promising classes of integrase inhibitors in terms of potency and selectivity. Currently, there are three HIV-1 integrase inhibitors in Phase-I clinical trials designated as L-870,810 [16], V-165, and S-1360 (compounds **11**, **12** and **13** in Fig. (**1**). The second one is a new type of inhibitor from the pyrano-di-pyrimidine class that inhibits the binding of DNA to integrase. It is structurally different from the other two integrase inhibitors, which are a diketo acid from Shionogi (S-1360) and a naphthyridine derivative from Merck (L-870,810).

The identification of novel lead structures is more than ever required after the discovery that single or multiple mutations on IN at residues T66, S153 and M154 [13] lead to high degrees of resistance against the diketo acids: screening of chemical databases may be one of the optimal methods to achieve this goal. Computer-aided molecular discovery approaches have been used in the past years targeting HIV-1 integrase [17-21] by the development of

1573-4064/05 $50.00+.00

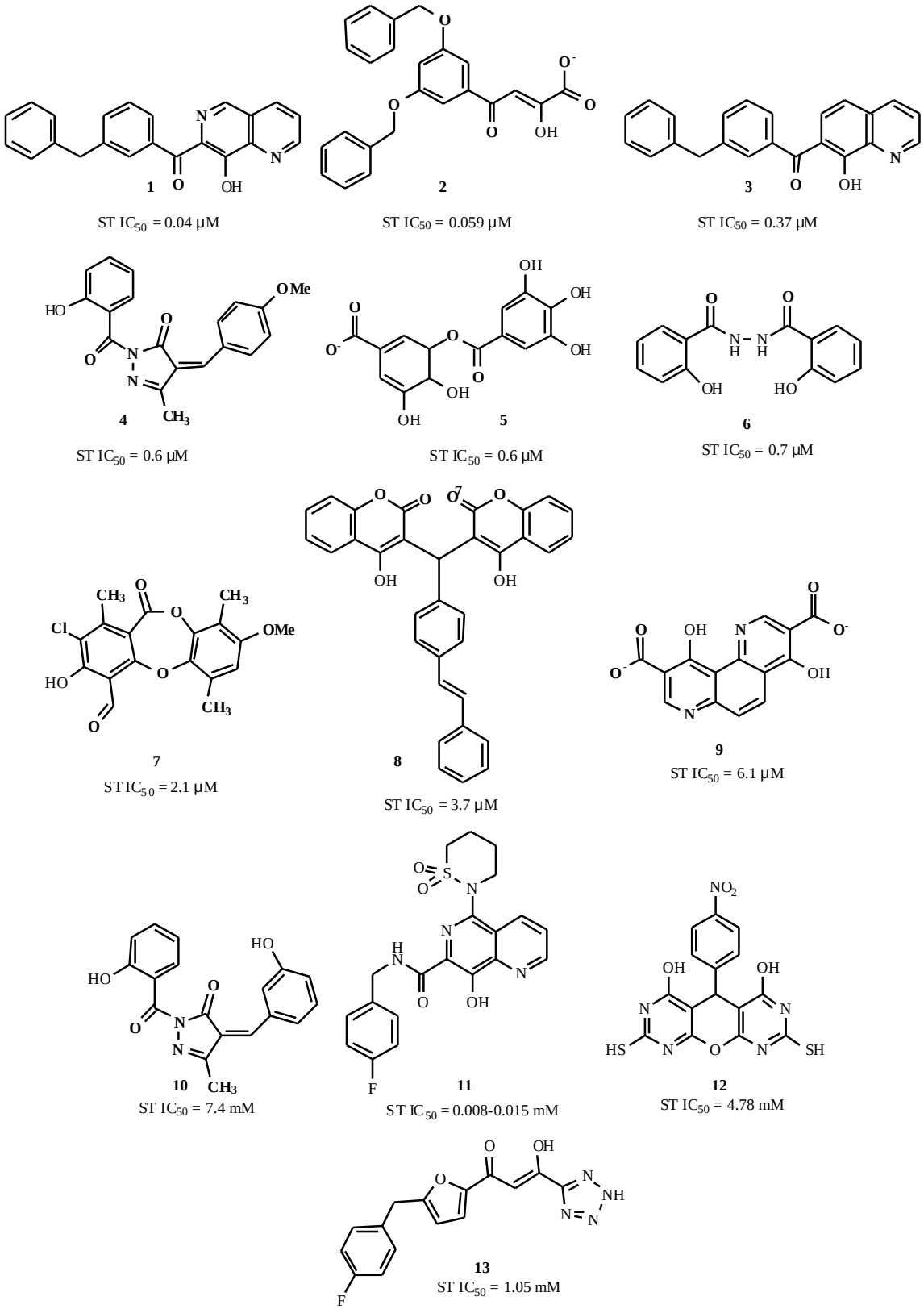


Fig. (1). Chemical structures of 13 representative HIV-1 integrase inhibitors along with their IC_{50} values against IN-catalyzed strand transfer (ST) reaction.

possible pharmacophores to be used as a query to search chemical 3D databases, in an effort to find new potential inhibitors.

Virtual screening of chemical databases is now a well-established method for finding new lead structures when the 3D structure of the target is available [22,23]. This strategy has been successfully used in the discovery of novel inhibitors against several enzymes [24-29]. Notwithstanding the availability of several crystal structures of the HIV-1 integrase catalytic core domain [30-33], this enzyme has never been the target of a virtual high-throughput screening study.

In the present contribution, a new stepwise approach to target-based lead discovery for HIV-1 integrase, combining ligand-based and structure-based methodologies is proposed. The first step consisted of the development of a pharmacophore model that could take advantage of the chemical and structural information coming from a dataset of 26 known inhibitors [34]. The molecules of Maybridge and NCI databases that matched the query pharmacophore model have been screened against 1QS4 wild-type integrase 3D structure, using a combination of three docking tools (Dock [35], Glide [36] and FlexX [37]) and four scoring functions (Dock, GlideScore, XScore [38] and FlexXScore). The choice of combined virtual high-throughput screening protocols has been dictated by the awareness that scoring is typically more important than docking in database screening, and that scoring functions performances often depend on the target active site features. However, since docking poses may significantly affect the scoring three docking tools, implementing three different algorithms have been employed [39].

The protocol described herein, led to the selection of 50 compounds whose activity is currently being tested using the HIV-1 integrase assay.

COMPUTATIONAL METHODS

Pharmacophore Model and Database Screening

A pharmacophore model derived from diverse classes of known inhibitors of HIV-1 IN has recently been reported [34]. The structures of these molecules have been used by the HypoGen module implemented in CATALYST [40], to generate a pharmacophore hypothesis that has been subsequently used as a query to perform 3D database searches. The NCI 3D database of 250,251 compounds and the Maybridge 2001 3D database of 56,000 compounds were screened against the HIV-1 IN pharmacophore model. For each hit returned, conformers were generated with an energy constraint of 20 Kcal/mol or less, above the local minimized structures, and upto a maximum number of 250 conformers. All the conformers that matched the four-feature pharmacophore model [34] were saved for further screening.

Virtual Screening

Docking programs consist of two parts: (i) an algorithm that searches the conformational, rotational and translational space available to a compound within the protein binding site [39]; (ii) a function, usually referred to as scoring function [41-44], which calculates the binding affinity or receptor-ligand complementarity. These scoring functions can be divided into three categories. The most important class consists of *empirical* scoring functions (GlideScore [36], XScore [38], and FlexXScore [37]), which approximate the binding free energy as a weighted sum of terms, describing different types of interactions between the receptor and the ligand. The second class is based on molecular mechanics force fields (such as Dock [35] energy function). Finally, knowledge-based scoring functions [44] are derived from statistical analyses of experimentally determined protein-ligand X-Ray structures.

Virtual screening studies were carried out using the following programs: Dock version 4.01 [35]. Glide [36], and FlexX [37] as implemented in the 6.9 release of the SYBYL package (TRIPOS Inc., St. Louis, MO). The purpose of using three different approaches was to compare the efficiency of their three related scoring functions, as well as their capability of finding reliable poses within IN active site. The binding free energies of the complexes target-ligand obtained by the three programs were further evaluated using the XScore [38] scoring function. With such combinations, 8 different rankings for each screened database were obtained. Dock/Dock protocol gave the best-ranked hits for both databases (Maybridge 2001 and NCI 2000) according to Dock energy function; Dock/XScore gave the ranking of the Dock poses according to XScore scoring function; Dock/Glide gave the best Dock poses ranked according to GlideScore scoring function; Glide/Glide protocol returned the best Glide poses ranked by GlideScore energy function; Glide/XScore returned the ranking of the glide poses according to XScore scoring function; Glide/Dock returned the Glide poses in function of Dock energy function; FlexX/FlexXScore protocol returned the best FlexX poses ranked by the FlexX scoring function; finally, FlexX/XScore gave the best FlexX poses ranked according to XScore scoring function. The FlexX/Dock protocol was not evaluated, since Dock scoring function was not capable to give favorable binding free energy values to most of the poses identified by FlexX (data not shown).

Preparation of the 3D Sub-databases

The hits obtained by screening the NCI 2000 and Maybridge 2001 databases were organized into two sub-groups. These sub-databases were then filtered in order to leave, for each ligand in the databases, only the conformer that gave the best fit into the pharmacophore. The process led to a 7,418-compound NCI sub-database and to a 6,174-compound Maybridge sub-database. The protonation state of ionizable groups (primary acidic and basic centers) was taken into account, assuming that these groups are ionized at a physiological pH of 7.4. Gasteiger-Marsili atomic charges [45] were added next, and the final coordinates stored in two multi mol2 files, one for each database. The compounds were then filtered in accordance with a Lipinski-like rule-of-five [46]. Only compounds with a molecular weight inferior to 25 Da or superior to 550 Da (a molecular weight of 500 Da would have been too restrictive for the type of molecules contained in our two databases) were discarded from the databases, as well as those having logP values greater than

5.0. Compounds with hydrogen bond donor and acceptor groups greater than 5 and 10 respectively, were also discarded from the databases, alongwith all the molecules that contained unsuitable atoms or functional groups (e.g. elements other than C, N, O, P, S; more than seven fluorines; more than six bromines, chlorines or iodines; peroxide; saturated terminal alkyl chains of *n*-heptane or greater). These filters left 5,153 and 6,780 compounds in Maybridge and NCI sub-databases, respectively, which had been next submitted to the high throughput docking study. In a last step, ten known IN inhibitors (compounds **1**-**10** in Fig. (**1**)) were added to both databases as control, in order to evaluate the capability of docking protocols to retrieve active compounds of diverse chemical classes among the top scorers.

Active Site Preparation

The coordinates of the protein used in the docking study were taken from the 2-ns MD simulation carried out on HIV-1 1QS4 (wild type) IN catalytic domain in complex with the Merck inhibitor L-731,988 [47]. In our previous investigation, in an effort to understand the molecular mechanism of the resistance of T66I/M154I double mutant IN to the inhibitor L-731,988 and its specific binding modes, docking studies, explicit solvent molecular dynamics simulations and free energy calculations were carried out. After the identification of the most favorable binding modes of the inhibitor L-731,988 by docking calculations, MD simulations were carried out on the following systems: T66I/M154I double mutant IN; double mutant IN in complex with L-731,988 in two different orientations; 1QS4 (wild type) IN in complex with L-731,988. The results of these simulations showed a similar dynamical behavior between T66I/M154I IN alone and in complex with L-731,988, while significant differences were observed in the mobility of the IN catalytic loop, (residues 138-149) [47]. Water molecules bridging the inhibitor to the active site residues have been identified in the simulation of T66I/M154I double mutant IN where the loop flexibility, which is considered to be essentially responsible for the enzymatic activity [48], was not diminished throughout the 2-ns production run. On the other hand, residues Q62 and Q148 were found in the simulation of 1QS4 (wild type) IN to play an important role in the interactions between the inhibitor and the protein, hampering the catalytic loop mobility [47]. The equilibrated 1QS4 IN conformation (0ns of production time) was chosen for carrying out virtual high-throughput screening because of the complete stability of the catalytic loop (residues 138-149) flexibility, and the preservation of the ligand-protein interactions throughout the 2-ns MD simulation.

The selected structure was quickly minimized with DISCOVER [40], using CHARMM force field through 200 steps of Steepest Descent. Finally, the hydrogen atoms were further minimized through 500 steps of Conjugate Gradient, keeping the non-hydrogen protein atoms fixed.

Dock 4.01 Docking

There are two major steps in the docking process: orienting the ligand into the target site and scoring the resulting complex. In the orienting step, the program Dock systematically attempts to fit each compound or conformer from a database into the binding site of the target biomolecule, so that three or more of the atoms in the database molecule overlap with a set of predefined site points in the target binding site.

Each acceptable orientation of a ligand in the binding site is then scored on a grid throughout the macromolecular target, using pre-calculated values for the protein component of the interaction energy.

A solvent-accessible surface of the receptor active site was generated using a 1.4-Å probe radius. Dock provides the sphgen [35] program as a method to generate a set of overlapping spheres that geometrically represent the binding site. A negative image of the surface invaginations is thus created, and spheres are constructed using the molecular surface previously calculated. Each sphere touches the molecular surface at two points and has its radius along the surface normal of one of the points. For the receptor, each sphere center is "outside" the surface, and lies in the direction of a surface normal vector. Spheres are calculated over the entire surface, producing approximately one sphere per surface point. This very dense representation is then filtered to keep only the largest sphere associated with each receptor surface atom. The filtered set is then clustered on the basis of radial overlap between the spheres using a single linkage algorithm. The area was restricted to the known IN active site, including the flexible loop (comprising residues 138-149) and the residues such as Asp64, Asp116, Glu152, Lys156, Lys159, Gln62 and His67 that, by site-mutation experiments and by computational studies [47], are known to make crucial interactions with IN substrates. Finally, 34 spheres were retained, sufficient for exhaustively describing IN binding site.

To compute interaction energies, a 3D grid with a 0.30 Å spacing was centered on the active site. Energy scoring grids were obtained by using an all atom model and a distance-dependent dielectric function ($\varepsilon = 4r$) with a 10 Å cutoff. Amber95 atomic charges [49] were assigned to all protein atoms. Databases molecules were then docked into the protein active site by matching the sphere centers with ligand atoms. A flexible docking of all molecules with subsequent energy minimization was performed as follows: multiple anchors search; maximum of 1,000 orientations, manual matching with a minimum of 3 nodes (node: number of site, points/ligand atoms used in distance matching) and a maximum of 10, a distance tolerance of 0.5 Å, simultaneous relaxation of the base fragments as well as of all peripheral segments, and final relaxation of the entire molecule. Typically, very flexible ligands, i.e.those with more than 12 rotatable bonds, are hard to treat properly using the conformational search algorithms during the docking process. This leads to poses that are not completely reliable, consequently affecting the scoring process. Moreover, compounds either with a too high or a too low number of heavy atoms do not reflect the typical features of drug-like or lead-like molecules, representing chemical entities that will be inadequate under the medicinal chemistry points of view. Therefore, compounds with more than 12 flexible bonds were discarded, as well as compounds with less than 6 or more than 60 heavy atoms. All docking poses were energy

minimized using 100 iterations and 10 minimization cycles with a convergence value of 0.1 kcal/mol. The poses were evaluated and the ligands ranked using the energy scoring function implemented in Dock, which is an interaction force field score that includes Van Der Waals and electrostatic terms. The energy score was used because it has been demonstrated to be the most robust among the three possible scores implemented in Dock (energy score, contact score, chemical score). The top 20% of energy scoring molecules of each screened databases was saved for the analysis. The protocol described above, after a calibration carried out using 10 known IN inhibitors, compounds **1-10** in Fig. (**1**), proved to give the best ranking consistency and the shortest energy gap between the most and the least experimentally active inhibitor.

Glide Docking

The same relaxed conformation of HIV-1 1QS4 IN catalytic domain used for the studies with Dock was set up for Glide. A cubic bounding box (14.0 Å x 14.0 Å x 14.0 Å) was set in order to include all essential residues and protein portion for the enzyme-substrate interactions. Glide generates conformations internally, and passes them through a series of filters. The first one places the ligand center at various grid positions of a 1 Å grid and rotates it around the 3 Euler angles; crude score values and geometric filters discard the unlikely binding modes. The next stage involves a grid-based force-field evaluation and refinement of docking solutions, including torsional (with OPLS-AA force field) and rigid-body movements of the ligand. A small number of surviving docking solutions can be evaluated by a Monte Carlo procedure to try to minimize the energy score. Glide protocol is tunable mainly by changing the Van Der Waals radii values. In the case of our system, treating the full Van Der Waals radii for the target atoms and for the ligand atoms led to the best results on the ten-compound validation set. The final energy evaluation is done with GlideScore function, which is a modified version of ChemScore [50] with different weighting factors for each term and an additional steric repulsive term. A penalty term that adds 3 Kcal mol^{-1} to the score for each polar group buried into a lipophilic environment was included in GlideScore. The top 20% scorers ranked according to GlideScore function were saved for both screened database.

FlexX Docking

The docking algorithm in FlexX is based on an incremental construction strategy, which consists of three phases [37]:

1. **Base Selection.** The ligand is fragmented into components by severing at all acyclic single bonds. Then FlexX automatically forms a set of alternative base fragments by selecting single components or combination of them.

2. **Base Placement.** The base fragments are the first parts of the ligand that are placed into the active site. Two algorithms are in use: the first one superposes triples of compatible interaction points in the active site; the second one, called line matching, matches pairs of interaction centers with pairs of interaction points if the number of the previous placement is too low. Clustering and clash tests follow in order to reduce the large number of solutions [51].

3. **Incremental Construction**. Starting with the different base placements, the complete ligand is constructed by adding the remaining ligands components, in compliance with the torsional database, step by step.

The ligands databases were docked into the HIV-1 IN active site using the default FlexX parameter settings contained in the FlexX version implemented in SYBYL 6.9 package. FlexX differentiates between three kinds of stereocenters: pseudo-R/S, R/S, and Z/E, which can be handled as additional degrees of freedom during the docking calculation. FlexX then automatically generates the stereoisomer with the best fit to the binding site. Final scores were calculated for all FlexX solutions (up to 5 per ligand) with FlexXScore scoring function [37].

XScore Scoring Function

The ranked docking solutions obtained with Dock, Glide and FlexX were then re-evaluated by empirical scoring function XScore [38]. This program allows the prediction of binding affinity of ligands in complex (with a conformation/ orientation predicted by the aforementioned docking algorithms) with the target. Unlike force field based scoring functions, XScore includes terms accounting for Van Der Waals interaction, hydrogen bonding, deformation penalty and hydrophobic effects. A novel feature is that three different algorithms have been implemented to calculate the hydrophobic effect term [38].

The compounds ranked in function of the binding free energies calculated by the scoring functions implemented in the three docking programs alongwith those ranked by XScore, were then compared and visually inspected.

Consensus Scoring

Each of the four scoring functions have their respective strengths and weaknesses, which can lead to different scoring functions performing better, depending on the class of molecules in question. Our approach consisted of merging the results of different scoring functions and of taking into account only those molecules that received a high rank by more than one scoring function (consensus scoring) [52].

The docked poses obtained with the three docking algorithms were rescored using the aforementioned four scoring functions, without performing any relaxation of the bound ligand and keeping the atomic coordinates of the target fixed. The eight docking/scoring protocols, whose results were evaluated, are the following: Dock/Dock; Dock/ XScore; Dock/GlideScore; Glide/GlideScore; Glide/XScore; Glide/Dock; FlexX/FlexXScore; FlexX/XScore. The common hits that were retrieved within the top 20% scorers by 8, by 7 and by 6 protocols were visually inspected in order to select a suitable number of molecules to be tested in the integrase assay according to their ranking, their binding mode within the target enzyme, and their molecular structure.

RESULTS AND DISCUSSION

Target-based Drug Design Protocol

As already mentioned, HIV-1 IN has been a challenging target for the design of novel inhibitors. A novel approach that takes into account all the available information about HIV-1 IN inhibition is presented herein. In Fig. (**2**), a flowchart of our protocol is depicted. The molecules contained in two publicly available databases (Maybridge and NCI) were filtered in a stepwise fashion, Fig. (**2**). The first step involved the development of a pharmacophore model [34], which, used as a query, allowed filtering out 11% of Maybridge and 2.9% of NCI. The resulting molecules were refined by discarding those compounds that did not fulfill the drug-likeness rules [46], Fig. (**2**). This further step led to retain a total of 11,933 compounds that were next submitted to the high-throughput docking studies, using 8 different docking/scoring protocols. The common hits retrieved by 8 out of 8, by 7 out of 8 and by 6 out of 8 docking/scoring protocols were visually inspected in order to select the most significant molecules to be tested in the IN assay. In order to select suitable ligands for the IN assay, the following features were taken into account: (i) formation of direct interactions between the metal co-factor and ligands functional groups; (ii) presence of hydrogen bonds formed by ligands with IN residues Gln148 and Asn155, or Lys159 or Glu152: these specific interactions were hypothesized to be significant for IN inhibition [47]; (iii) suitable (low energy) binding conformation within the active site; (iv) ligand chemical structures.

Pharmacophore Model Development

The purpose of the pharmacophore hypothesis generation is not only to accurately predict the activity of the training set, but also to verify whether the pharmacophore models are able to predict the activities of compounds that are not included in the training set and classify them correctly as active or inactive. For HIV-1 IN it is not possible to derive a completely reliable pharmacophore model from a protein-ligand complex, since the only available crystal structure seems to be unsuitable for this purpose, due to crystal packing effects [53]. The high B-factor values found for the catalytic loop region comprised between residues 138-149 [31], which is essential for the catalytic activity [48], demonstrate that this loop is extremely flexible and its involvement in IN catalytic activity cannot be structurally demonstrated especially when a substrate or an inhibitor binds to the active site. Site-directed mutagenesis studies [54,55] revealed that residues Asp64, Asp116 and Glu152 are crucial for the enzymatic activity. Based on that, the most rational way to proceed was to develop first a pharmacophore model, taking into account the chemical features of known HIV-1 IN inhibitors. A detailed description of the model is given elsewhere [34], therefore, only a brief description will be provided herein. The

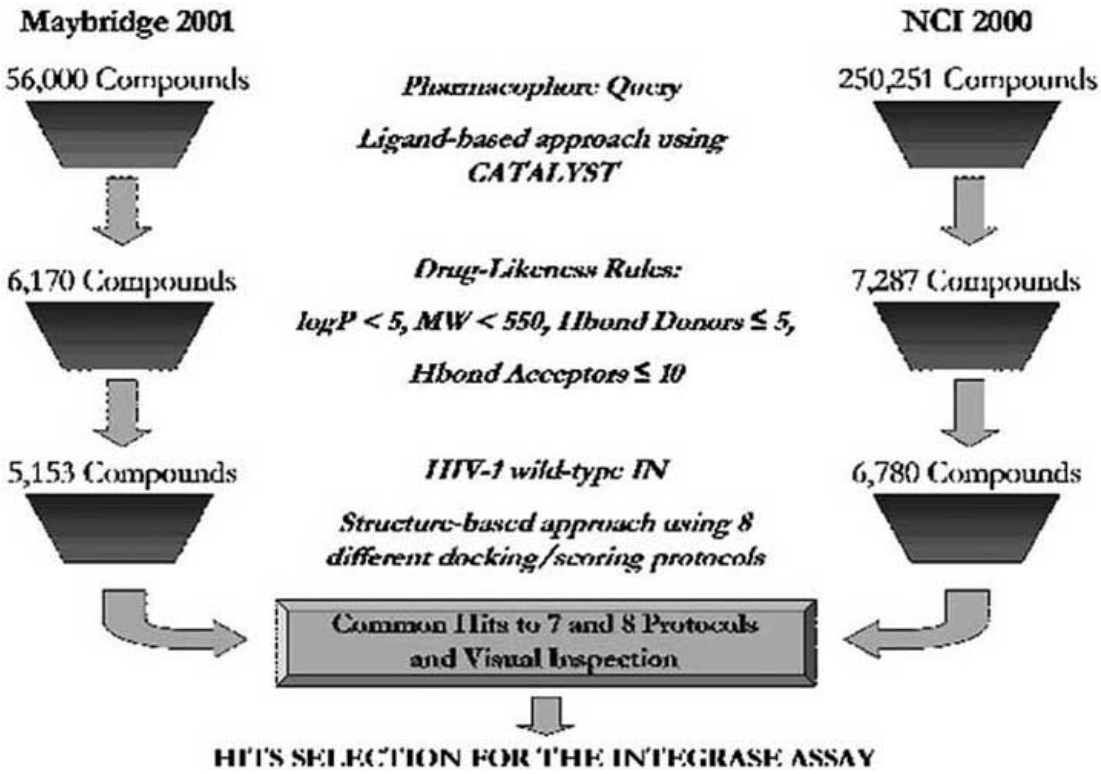


Fig. (2). Flowchart representation of the subsequent filters applied to Maybridge and NCI databases. The first selection was undertaken with a pharmacophore model query, obtained from a ligand-based approach [34]. Secondly, drug-likeness of the resulting compounds was evaluated by applying Lipinski-like rules [46]. Thirdly, HIV-1 wild-type IN 3D structure was used to screen the remaining molecules. Finally, visual inspection of the resulting top ranked poses led to the ultimate selection of the molecules to be tested in the integrase assay.

HypoGen module in the program CATALYST [40], was used in order to evaluate in a weighted fashion, the functional groups of known inhibitors in function of their biological activity.

The best hypothesis, after a validation process with a test set of 14 IN known inhibitors, was used as a query for 3D database search [34].

The identification of four essential chemical features [34] allowed the identification of 13,457 compounds (out of 306,251) that contained the pharmacophore pattern in one or more of their previously generated conformers. These matching molecules represented the 4.4% of the starting 3D databases and required further refinement. All the compounds that did not fulfill, the drug likeness filters (see the Computational Methods Section) were discarded, leaving 5,153 and 6,780 compounds in Maybridge and NCI sub-databases, respectively, which were next submitted to the high throughput docking study.

Virtual Screening

Virtual high-throughput docking was carried out with a refinement purpose, rather than with a common screening intent, since the databases that were filtered with the docking protocols resulted from a pharmacophore-based database search, Fig. (**2**). This approach represents a novel combination of a ligand-based drug design method (i.e. pharmacophore search), with a classical structure-based drug design.

HIV-1 integrase represents a challenging case study from a virtual screening point of view, because of several potential disadvantages: (i) its active site is shallow and solvent-exposed; (ii) there is one metal ion coordinated to two important active site residues, that is also supposed to be involved in ligands molecular recognition [56-59]; (iii) most of the known ligands bind with low binding constants (micromolar range); (iv) only one IN crystal structure in complex with an inhibitor [33] is currently available, in which the inhibitor position seems to be affected by crystal packing [53].

Initial docking attempts with several settings were undertaken for each docking program, using a test set of 10 known inhibitors, compounds **1-10** in Fig. (**1**), one for each of the most active chemical classes. Our purpose was to find a virtual ranking of these compounds as similar as possible to the experimental ranking, having, at the same time, the shortest energy gap between the top ranked compound and the least experimentally active compound. A protocol employing a multiple anchor conformation search alongwith a non-automated matching resulted in the method of choice. Dock 4.01 implements three possible scoring functions (energy score, chemical score, and contact score). In our case, the results were computed with the Dock energy score, which revealed to be the most robust among the three. Within the top 6 Dock/Dock scorers, 5 out of the top 6 experimentally scored compounds were retrieved (even though the order was not consistent), with the only exception of compound **4** (Fig. (**1**)), which was ranked as 8^{th}, Fig. (**4**). Compound **2** was ranked as first in Dock/Dock, while it was ranked as 2^{nd} in the experimental values; moreover, compound **10** in Fig. (**1**), was correctly ranked as 10^{th}. The poses identified by this docking protocol were re-evaluated using the XScore scoring function, and the resulting ranking resulted worse than the previously described one, even though the energy gap was 1.09 Kcal/mol. Dock/Glide protocol ranked compounds **1** and **2,** Fig. (**1**) as 10^{th} and 7^{th}, respectively, instead of 1^{st} and 2^{nd}, retrieving only 3 out of the top 6 most experimentally active compounds. With the Glide/Glide protocol, 5 out of 6 most active compounds were retrieved with compounds **1** and **2,** Fig. (**1**), ranked as 5^{th} and 6^{th} respectively, instead of 1^{st} and 2^{nd}. The Glide/XScore protocol was better-performing: only 4 compounds were retrieved among the 6 most active, but the ranking was more consistent with the experimental data.. Glide/Dock protocol retrieved 3 out of 6 most active compounds with compounds **1** and **2** (Fig. (**1**)), ranked as 2^{nd} and 1^{st} instead of 1^{st} and 2^{nd}. FlexX/FlexXScore protocol could retrieve 4 out of 6 experimentally active compounds. FlexX/XScore protocol slightly improved the performance by retrieving 5 out of 6 most active compounds having compound **1** in Fig. (**1**), ranked as 2^{nd} instead of 1^{st} (see Fig. (**3**) for more details).

Finding a virtual screening protocol, capable of discriminating small differences in HIV-1 IN inhibitors activity, is not the purpose of the present paper. However, having active compounds ranked as similarly as possible to the experimental studies, represents an excellent starting point for any virtual screening.

In a virtual screening study, two issues have to be typically addressed: (i) hit rates, i.e. percentage of known true inhibitors, among the top 5%, 10% and 20% scorers; (ii) docking accuracy. Any virtual screening protocol should be able to retrieve at least 80% of the known inhibitors among the top 20% scorers, in order to provide an acceptable number of false positives within the top ranked hits. The analysis was carried out separately on both our databases, because of the rather different chemical features that typify the compounds belonging to Maybridge and to NCI databases. In Fig. (**4**), the ranking of each HIV-1 integrase inhibitor proposed by the 4 different scoring functions from the three independent docking poses is represented. XScore was found to be the best-performing function in retrieving the known inhibitors among the top 20% scorers for both databases. It significantly improved the hit rates either when coupled with Dock or Glide poses, going from 50% in Dock/Dock upto 100% and 90% in Dock/XScore for Maybridge and NCI, respectively. Even though GlideScore performed well when used to rank Glide poses, retrieving 80% of known inhibitors among the top 20% scorers, XScore was able to improve Glide/GlideScore hit rates upto 100%, Fig. (**4**).

If the differences in the enrichment factors among the top 5%, 10% and 20% scorers for the eight different protocols are compared, Fig. (**5**), it can be observed that XScore is capable of significantly increasing the percentage of known inhibitors retrieved among the top 5% and 10% scorers. Glide/XScore achieved 50% and 60% of hit rates among the top 10% scorers for Maybridge and NCI from the 30% retrieved by Glide/Glide scheme, Fig. (**5**). The enrichment of known inhibitors among the top 5% scorers was significant

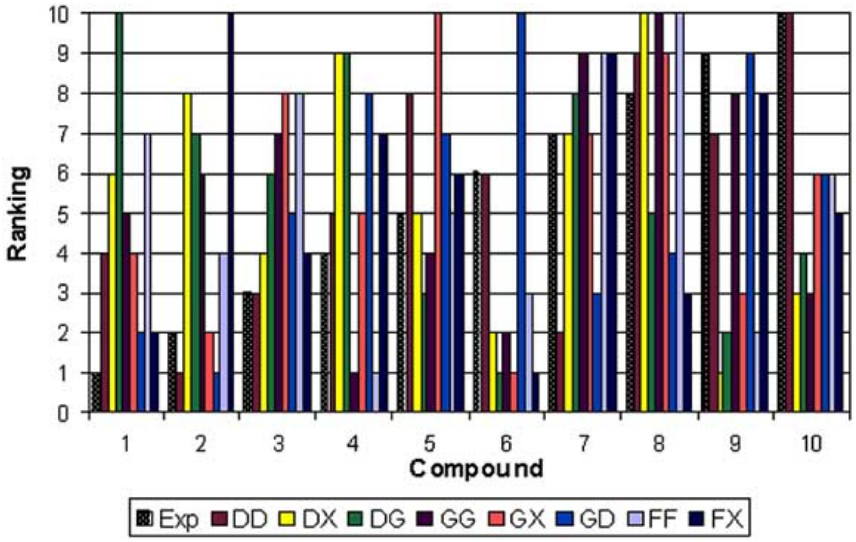


Fig. (3). Graphical representation of the experimental IC_{50} and the binding free energy of the 10 active molecules (compounds **1-10** in Figure **1**) used as a control in the sub-databases virtual high throughput docking, according to each docking/scoring protocol. Abbreviations are: **Exp**, integrase assay; **DD**, Dock/Dock; **DX**, Dock/XScore; **DG**, Dock/GlideScore; **GG**, Glide/GlideScore; **GX**, Glide/XScore; **GD**, Glide/Dock; **FF**, FlexX/FlexXScore.

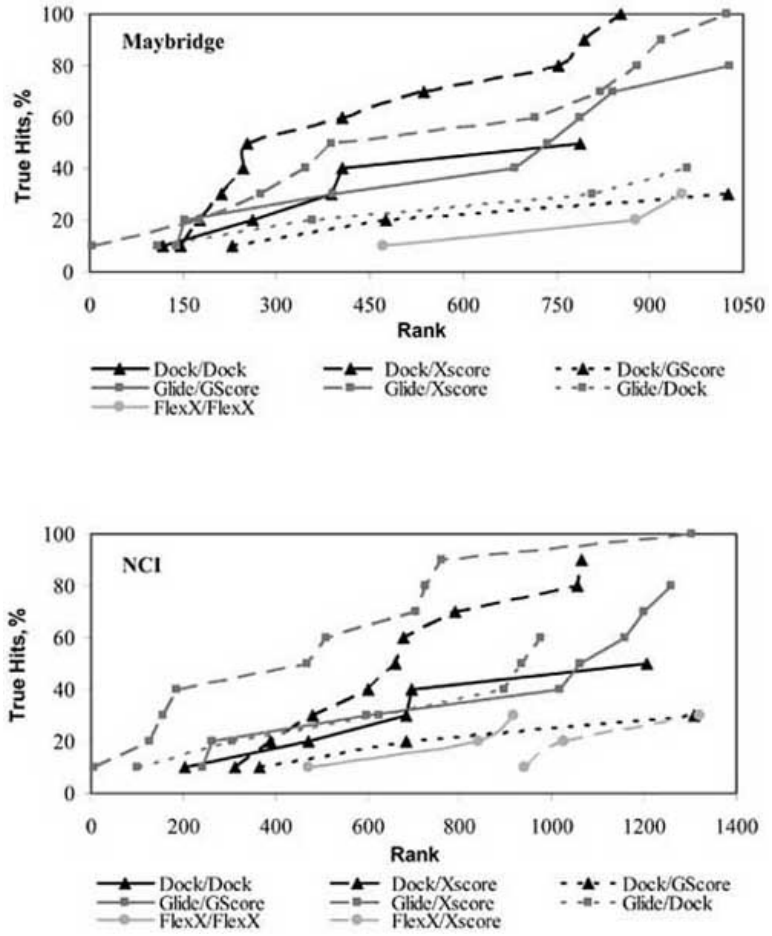


Fig. (4). Ranking of HIV-1 IN inhibitors among the top 20% scorers of Maybridge and NCI databases using a combination of three docking programs (Dock, Glide and FlexX) and four scoring functions (Dock, GlideScore, FlexXScore and XScore).

for Dock/XScore protocol in Maybridge screening and for Glide/XScore protocol in NCI screening, Fig. **(5)**.

The scheme FlexX/XScore did not retrieve any of the known inhibitors among the top 20% scorers for Maybridge, while for NCI it was only able to confirm the 30% of known inhibitors retrieved by FlexX/FlexX scheme. As already pointed out, HIV-1 integrase has a shallow and hydrophilic active site that presents many polar groups suitable for forming hydrogen bonds with the base ligands fragment. The incremental construction process allows the ligands to grow into the solvent-exposed region, due to the lack of conformational restrictions within the outer region of IN binding site. As a consequence, the hit rates of FlexX poses do not receive a good rank by Xscore, because only one portion of the ligands forms geometrically correct hydrogen bonds.

It is not trivial to explain the performance of each scoring function with respect to their particular features. The force field based Dock energy score seemed to be the worst performing in our system, regardless of the docking poses. This might be explained with the insufficient strength of Van Der Waals and electrostatics terms of the equation for describing the binding energies of ligands, docked far away from the inner region of IN binding pocket. For such types of binding sites, hydrogen bonds are expected to be the most important protein-ligand interactions. GlideScore and FlexXScore have strong directional H-bonding terms and should perform very well in our case. This is true for the GlideScore coupled with Glide docking tool, while FlexXScore poor performances are mainly attributable to the FlexX poses, that place portions of the ligands in the solvent-exposed region. The best-performing scoring function was XScore, which was capable of improving active compounds retrieval independently from the docking poses. Nevertheless, FlexX poses did not receive a good ranking by XScore, suggesting that ligands interacting with IN by means of hydrogen bonds involving only a small fragment of

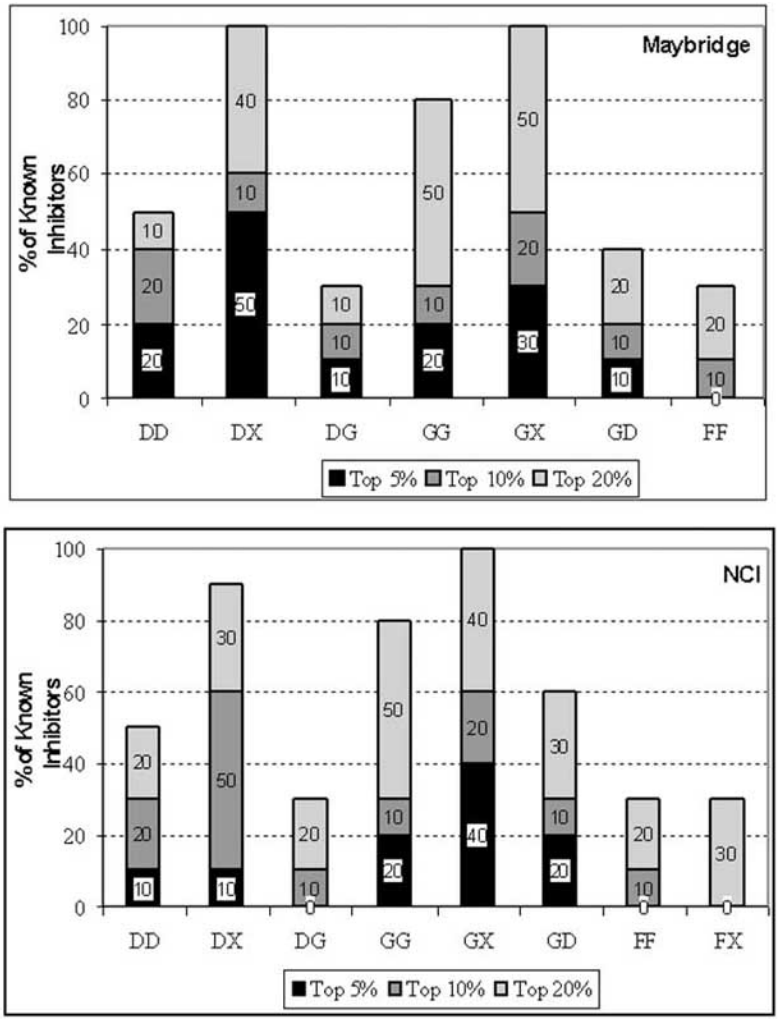


Fig. (5). Hit rates (% of known IN inhibitors) among the top 5% (black bars), the top 10% (dark gray bars) and the top 20% (light gray bars) scorers. Abbreviations for docking/scoring protocols are: **DD**, Dock/Dock; **DX**, Dock/XScore; **DG**, Dock/GlideScore; **GG**, Glide/GlideScore; **GX**, Glide/XScore; **GD**, Glide/Dock; **FF**, FlexX/FlexXScore.

the compound, having the other fragments completely solvent-exposed, are low-ranked by XScore. In order to see whether the failure of rankings may be due to docking inaccuracies and to confirm the statements above, visual inspection was carried out on the top 200 scorers of each docking/scoring protocol. Given the absence of X-Ray poses, the following features were considered as very significant: (i) hydrogen bonds formed by the ligands with residues Gln148, Glu152, Asn155, Lys159: a hydrogen bond network involving two or more of these residues at the same time is likely to lead to IN inhibition [47]; (ii) direct interactions between the Mg^{2+} co-factor with the ligand; (iii) IN binding site exploration; (iv) correlation between ranking and ligand binding modes. In Fig. (**6**), the poses of the same ligand obtained by the three docking programs are compared. This ligand was ranked among the top 100 scorers according to each docking/scoring protocol, but its poses are significantly different, demonstrating that scoring is typically more important than docking accuracy. FlexX pose retrieved only one hydrogen bond with His67 side chain, involving the base fragment ring placed close to the Mg ion. The remaining portion of the ligand is rather far away from the center of IN active site, forming two hydrogen bonds with Lys159 side chain from the solvent-exposed region, Fig. (**6**). On the other hand, Glide results showed all of the compounds bound within the active site region, retrieving the aforementioned hydrogen bond network in more than 70% of the inspected hits. In Fig. (**6**) the Glide pose of the considered compound is in the middle of the active site and forms three hydrogen bonds with Gln148, Glu152 and Lys159. These interactions are part of the previously identified hydrogen bond network [47] and are supposed to lead to IN catalytic loop (residues 138-149) constrain. The catalytic loop loss of flexibility induced by the formation of these interactions has been hypothesized to be responsible for HIV-1 IN inhibition [47,60]. The overall quality of Dock poses was somehow in the middle between Glide and FlexX, since 35% of the inspected hits interacted by hydrogen bonds with the aforementioned IN residues, while the remaining compounds were placed far away from those residues by Dock, even though they were still within the active site region, Fig. (**6**).

The comparison of the different docking/scoring schemes clearly suggests, at least for the present system, that Glide/XScore is the most reliable protocol in terms of both docking accuracy and correlation between poses and ranking. However, Dock/XScore should not be neglected, since it was capable of retrieving 90% and 100% of controls within the top 20% scorers of Maybridge and NCI respectively.

In order to select the compounds to be tested in the integrase assay, the top 20% scorers ranked according to each Dock/XScore and Glide/XScore protocols were first compared and the common hits saved. Subsequently, in order to render the hits choice more conclusive, all of the top 20% scorers of the 8 different protocols employed for both Maybridge and NCI databases (Table **1**) were compared. The common hits retrieved by 6 or more protocols were visually inspected, and those compounds that fulfilled the features described above (in terms of binding modes and also of chemical structure) were selected for the experimental assay (Table **2**). It is remarkable that the protocols using XScore as scoring function had the greatest number of common hits

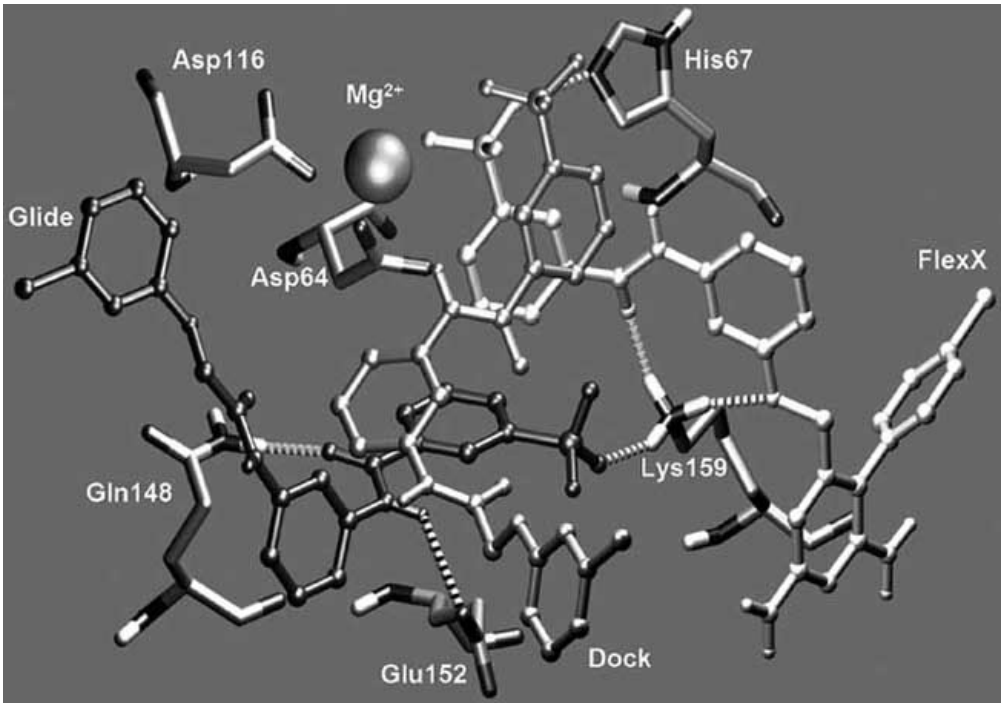


Fig. (6). Overlay of the docking poses obtained by the three different docking programs (Dock, Glide and FlexX) for the same ligand [62]. In the Glide pose the ligand forms hydrogen bonds (displayed in white thick dashed lines) with Gln148, Glu152 and Lys159. Dock and FlexX poses are rather far away from the active site: the former pose does not form any hydrogen bond, while the latter pose forms only two hydrogen bonds with Lys159.

Table 1. Number of Common Hits Retrieved Among the Top 20% Scorers Ranked According to Each Protocol. In Parenthesis, the Corresponding Percentage of the Considered Top Scorers is Indicated

MAY[1]	DD*	DX*	GG*	GX*	GD*	DG*	FF*	FX*	NCI[2]
DD		502 (38.6%)	279 (21.5%)	292 (22.5%)	348 (26.8%)	362 (27.8%)	251 (19.3%)	290 (22.3%)	DD
DX	301 (29.2%)		220 (16.9%)	535 (41.1%)	304 (23.4%)	520 (40.0%)	260 (20.0%)	477 (36.7%)	DX
GG	231 (22.4%)	172 (16.7%)		358 (27.5%)	237 (18.2%)	259 (19.9%)	299 (23.0%)	302 (23.2%)	GG
GX	154 (14.9%)	390 (37.9%)	358 (34.8%)		422 (32.5%)	299 (23.0%)	348 (26.8%)	520 (40.0%)	GX
GD	253 (24.6%)	192 (18.6%)	237 (23.0%)	230 (22.3%)		271 (20.8%)	338 (26.0%)	328 (25.2%)	GD
DG	230 (22.3%)	371 (36.0%)	193 (18.7%)	205 (19.9%)	198 (19.2%)		211 (16.2%)	292 (22.5%)	DG
FF	230 (22.3%)	226 (21.9%)	239 (23.2%)	295 (28.6%)	279 (27.1%)	186 (18.1%)		551 (42.4%)	FF
FX	170 (16.5%)	339 (32.9%)	212 (20.6%)	415 (40.3%)	271 (26.3%)	204 (19.8%)	416 (40.4%)		FX

[1]MAY indicates the Maybridge 2001 3D database. Results related to it are reported behind the gray diagonal on the left side of the table.

[2]NCI indicates the NCI 2000 3D database. Results related to it are reported above the gray diagonal on the right side of the table.

*DD indicates Dock/Dock protocol; DX indicates Dock/XScore; GG indicates Glide/GlideScore; GX indicates Glide/XScore; GD indicates Glide/Dock; DG indicates Dock/GlideScore; FF indicates FlexX/FlexXScore; FX indicates FlexX/XScore.

among the top 20% scorers when compared to each other in a pair wise fashion (Table **1**). Even though the poses identified for each ligand by the three different docking algorithms were very dissimilar, the scoring by XScore was not influenced, since about 40% of the top 20% scorers were retrieved.

In Table **2**, the number of common hits retrieved by more than 5 docking/scoring protocols, alongwith the percentage of compounds displaying the interactions described above, is schematically represented. Most of those (from 68% up to 100%) gave interactions with Gln148, Asn155 and Lys159, reproducing the hydrogen bond network that was hypothesized to be essential for the catalytic loop constrain [47].

The compounds that were eventually selected for the activity assay (the experimental results will be reported elsewhere) included β-diketo acids, sulfonamides, benzoic-hydrazides, urea derivatives, coumarin derivatives, oxalamide derivatives, N-carbonyl-amides and salicyl-hydrazide derivatives. Interestingly, our protocol retrieved two compounds that are strongly related to previous studies [17,61]. Molecule **a** in Table.**3** is a derivative of three compounds that showed very high activity against both 3'-processing and strand transfer with values for the latter enzymatic reaction spanning from 0.29 up to 14.3 μM [17]. Moreover, molecule **b** in Table **3** had already been selected and tested by Neamati and co-workers, showing an activity of 6.2 μM against the strand transfer reaction [61].

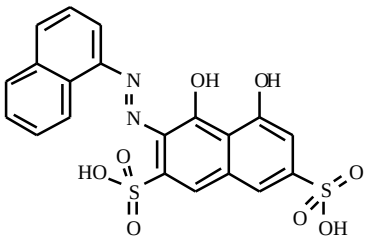
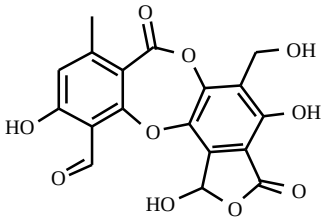
CONCLUSIONS

An HIV-1 integrase-targeted drug design approach has been presented herein. The application of subsequent filters allowed the reduction of the number of compounds to be inspected significantly, speeding up the lead candidate selection process. The development of a pharmacophore model taking advantage of the chemical information contained in the structure of known integrase inhibitors [34], led to the selection of compounds typified by specific chemical features out of more than 300,000 molecules contained in Maybridge and NCI databases. The resulting sub-databases were subsequently screened by means of a high throughput docking methodology, employing different combinations of docking algorithms and scoring functions. In the case of our target, Glide poses showed clearly to be the most suitable for our virtual screening study in terms of exploration of the active site space. Dock and FlexX poses did not seem to be fully reliable for integrase, even though no relationship between docking accuracy and hits ranking could be found in the present contribution. The best performing scoring function, in terms of true hits retrieval was Xscore, when combined with Dock and Glide docking

Table 2. Number of Common Hits Retrieved by More Than 5 Docking/Scoring Protocols

Databases	*Common Hits Retrieved in 6/8 Protocols*	*% of Hits Displaying Significant Interactions*	*Common Hits Retrieved in 7/8 Protocols*	*% of Hits Displaying Significant Interactions*	*Common Hits Retrieved in 8/8 Protocols*	*% of Hits Displaying Significant Interactions*
Maybridge	26	68%	3	100%	2	100%
NCI	64	71%	16	75%	2	100%

Table 3. Retrieved Compounds that are Correlated to Previous Studies [6,16]

	Structure	NCI Database Code	Protocols*	IC_{50} (μM)
a		45208	7/8	-
b		87509	6/8	6.2

*Number of Docking/Scoring protocols that ranked each compound among the top 20% scorers.

algorithms. Dock and FlexX scoring functions did not perform well, in some cases, not being able to score experimentally active compounds. The final application of a consensus-like approach allowed the selection of 116 compounds retrieved by more than 5 docking/scoring protocols among the top 20% scorers. A representative compound for each of the chemical classes identified by our combined methodology is currently being tested for the activity against integrase.

The results presented herein showed that our drug-discovery approach is valuable for a rational selection of new HIV-1 integrase potential inhibitors. However, given the specific characteristics of this enzyme, it cannot be considered as a general method for any type of target.

ACKNOWLEDGEMENTS

This work was supported the NIH Program on Structural Biology of AIDS Related Proteins (GM56553) and by Associazione Italiana per la Ricerca sul Cancro (AIRC), Milan, and the Italian Ministry for University and Research (MIUR), Rome, Italy. S. Moro is strongly grateful to Dr. Paolo Braiuca (Department of Pharmaceutical Sciences, University of Trieste, Italy) for carrying out Sybyl docking studies. The authors wish to acknowledge also Dr. Irwin Kuntz for making the software Dock (version 4.01) available.

ABBREVIATIONS

HIV-1 = Human immunodeficiency virus type 1

1QS4 = HIV-1 integrase catalytic domain with double solubility mutation F185K/W131E

IN DM = 1QS4 with double resistance mutation (T66I/M154I)

REFERENCES

[1] Asante-Appiah, E.; Skalka, A.M. *Antivir. Res.* **1997**, *36*, 139.
[2] Hindmarsh, P.; Leis, J. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 836.
[3] Brown, P.O.; Coffin, J.M., Hughes, S.H.; Varmus, H.E. Ed.; Cold Spring Harbor Lab. Press, Plainview: NY, 1997.
[4] Ellison, V.; Brown, P.O. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7316.
[5] Wolfe, A.L.; Felock, P.J.; Hastings, J.C.; Uncapher Blau, C.; Hazuda, D.J. *J. Virol.* **1996**, *70*, 1424.
[6] Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M.C.; Milne, G.W.A.; Proksa, B.; Pommier, Y. *J. Med. Chem.* **1997**, *40*, 942.
[7] Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S.; Milne, G.W.A.; Pommier, Y.; Burke, T.R.Jr. *J. Med. Chem.* **1997**, *40*, 937.
[8] Neamati, N.; Turpin, J.A.; Winslow, H.E.; Christensen, J.L.; Williamson, K.; Orr, A.; Rice, W.G.; Pommier, Y.; Garofalo, A.; Brizzi, A.; Campiani, G.; Fiorini, I.; Nacci, V. *J. Med. Chem.* **1999**, *42*, 3334.
[9] Neamati, N.; Hong, H.; Owen, J.M.; Sunder, S.; Winslow, H.E.; Christensen, J.L.; Zhao, H.; Burke, T.R.Jr.; Milne, W.A.; Pommier, Y. *J. Med. Chem.* **1998**, *41*, 3202.
[10] Zhao, H.; Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Milne, G.W.A.; Pommier, Y.; Burke, T.R.Jr. *J. Med. Chem.* **1997**, *40*, 242.
[11] Fesen, M.R.; Kohn, K.W.; Leteurtre, F.; Pommier, Y. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2399.
[12] Robinson, W.E.; Reinecke, M.G.; Abdel-Malek, S.; Jia, Q.; Chow, S.A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6926.
[13] Hazuda, D.J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J.A.; Espeseth, A.; Gabryelski, L.; Schleif, W.; Blau, C.; Miller, M.D. *Science* **2000**, *287*, 646.
[14] Zhuang, L.; Wai, J.S.; Embrey, M.W.; Fisher, T.E.; Egbertson, M.S.; Payne, L.S.; Guare, J.P. Jr.; Vacca, J.P.; Hazuda, D.J.; Felock, P.J.; Wolfe, A.L.; Stillmock, K.A.; Witmer, M.V.; Moyer, G.; Schleif, W.A.; Gabryelski, L.J.; Leonard, Y.M.; Lynch, J.J.Jr.; Michelson, S.R.; Young, S.D. *J. Med. Chem.* **2003**, *46*, 453.
[15] Zhang, X.; Pais, G.C.G.; Svarovskaia, E.S.; Marchand, C.; Johnson, A.A.; Karki, R.G.; Nicklaus, M.C.; Pathak, V.K.; Pommier, Y.; Burke, Jr., T.R. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1215.
[16] Hazuda, D.J.; Neville, J.A.; Gomez, R.P.; Samson, M.J.; Wai, J.S.; Zhuang, L.; Fisher, T.E.; Embrey, M.; Guare J.P. Jr.; Egbertson, M.S.; Vacca, J.P.; Huff, J.R.; Felock, P.J.; Witmer, M.V.;

Stillmock, K.A.; Danovich, R.; Grobler, J.; Miller, M.D.; Espeseth, A.S.; Jin, L.; Chen, I.; Lin, J.H.; Kassahun, K.; Ellis, J.D.; Wong, B.K.; Xu, W.; Pearson, P.G.; Schleif, W.A.; Cortese, R.; Emini, E.; Summa, V.; Holloway, M.K.; Young, S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11233.

[17] Nicklaus, M.C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G.W.A.; Pommier, Y. *J. Med. Chem.* **1997**, *40*, 920.

[18] Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M.C.; Mazumder, A.; Zhao, H.; Burke, T.R.Jr.; Pommier, Y.; Milne, G.W.A. *J. Med. Chem.* **1997**, *40*, 930.

[19] Neamati, N.; Hong, H.; Sunder, S.; Milne, G.W.A.; Pommier, Y. *Mol. Pharm.* **1997**, *52*, 1041.

[20] Hong, H.; Neamati, N.; Winslow, H.E.; Christensen, J.L.; Orr, A.; Pommier, Y.; Milne, G.W. *Antivir. Chem. Chemother.* **1998**, *9*, 461.

[21] Barreca, M.L.; Rao, A.; De Luca, L.; Zappala, M.; Gurnari, C.; Monforte, P.; De Clercq, E.; Van Maele, B.; Debyser, Z.; Witvrouw, M.; Briggs, J.M.; Chimirri, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1450.

[22] Kontoyianni, M.; McClellan, L.M.; Sokol, G.S. *J. Med. Chem.* **2004**, *47*, 558.

[23] Kontoyianni, M.; Sokol, G.S.; McClellan, L.M. *J. Chem. Inf. Comput. Sci.* **2005**, *1*, 11.

[24] Kraemer, O.; Hazemann, I.; Podjarny, A.D.; Klebe, G. *Prot.: Struc. Func. Bioinformat.* **2004**, *55*, 814.

[25] Schapira, M.; Raaka, B.M.; Samuels, H.H.; Abagyan, R. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1008.

[26] Bohem, H.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, T.; Meunier-Keller, N.; Mueller, F. *J. Med. Chem.* **2000**, *43*, 2664.

[27] Soelaiman, S.; Wei, B.Q.; Bergson, P.; Lee, Y.; Shen, Y.; Mrksich, M.; Shoichet, B.K.; Tang, W. *J. Biol. Chem.* **2003**, *278*, 25990.

[28] Schoichet, B.K.; McGovern, S.L.; Wei, B.; Irwin, J.I. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439.

[29] Evers, A.; Klebe, G. *J. Med. Chem.* **2004**, *47*, 5381.

[30] Dyda F,H.A.; Jenkins T.M.; Engelman, A.; Craigie, R.; Davies, D.R. *Science* **1994**, *266*, 1981.

[31] Goldgur Y.D.F.; Hickman, A.B.; Jenkins, T.M.; Craigie, R., Davies, D.R. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9150.

[32] Maignan S.G.J.; ZhoupLiu, Q.; Clement-Mella, C.; Mikol, V. *J. Mol. Biol.*, *282*, 359.

[33] Goldgur Y.C.R.; Cohen, G.H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D.R. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13040.

[34] Iurcu Mustata, G.; Brigo, A.; Briggs, J.M. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1447.

[35] Kuntz, I. D., Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. *J. Mol. Biol.* **1982**, *161*.

[36] Schrodinger, I.; Schrodinger, Inc.: Portland, OR, 2001.

[37] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.

Received: 06 October, 2004 Accepted: 04 January, 2005

[38] Wang, R.; Lai, L.; Wang, S. *J. Comp. Aid. Mol. Des.* **2002**, *16*, 11.

[39] Brooijmans, N.; Kuntz, I.D. *Ann. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335.

[40] Accelrys, Inc.: San Diego, CA.

[41] Ajay Murcko, M.A. *J. Med. Chem.* **1995**, *38*, 4953.

[42] Tame, J.R.H. *J. Comp. Aid. Mol. Des.* **1999**, *13*, 99.

[43] Bohem, H.J.; Stahl, M. *Med. Chem. Res.* **1999**, *9*, 445.

[44] Gohlke, H.; Klebe, G. *Curr. Opin. Struct. Biol.* **2001**, *11*, 231.

[45] Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3228.

[46] Lipinski, C. A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Advanced Drug Delivery Reviews* **1997**, *23*, 3.

[47] Brigo, A.; Lee, K.W.; Fogolari, F.; Iurcu Mustata, G.; Briggs, J.M., *Prot.: Struc. Func. Bioinformat., In press*.

[48] Greenwald, J.; Le, V.; Butler, S.L.; Bushman, F.D.; Choe, S. *Biochemistry* **1999**, *38*, 8892.

[49] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, J.K.M.; Ferguson, D.M.; Spellmeyer, D.M.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[50] Eldridge, M.; Murray, C.W.; Auton, T.A.; Paolini, G.V.; Lee, R.P. *J. Comp. Aid. Mol. Des.* **1997**, *11*, 425.

[51] Rarey, M.; Wefing, S.; Langauer, T. *J. Comp. Aid. Mol. Des.* **1996**, *10*, 41.

[52] Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, W.P. *J. Med. Chem.* **1999**, *42*, 5100.

[53] Sotriffer, C.A.; Ni, H.; McCammon, J.A. *J. Am. Chem. Soc.* **2000**, *122*, 6136.

[54] Kulkosky, J.; Jones, K.S.; Katz, R.A.; Mack, J.P.; Skalka, A.M. *Mol. Cell. Biol.* **1992**, *12*, 2331.

[55] Engelman, A., Craigie, R. *J. Virol.* **1992**, *66*, 6361.

[56] Marchand, C.; Johnson, A.A.; Karki, R.G.; Pais, G.C.G.; Zhang, X.; Cowansage, K.; Patel, T.A.; Nicklaus, M.C.; Burke, T.R. Jr.; Pommier, Y. *Mol. Pharm.* **2003**, *64*, 600.

[57] Pais, G.C.G.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K.; Svarovskaia, E.S.; Pathak, V.K.; Tang, Y.; Nicklaus, M.; Pommier, Y.; Burke, Jr. T.R. *J. Med. Chem.* **2002**, *45*, 3184.

[58] Grobler, J.A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A.S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J.S.; Young, S.; Vacca, J. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6661.

[59] Neamati, N.; Lin, Z.; Karki, R.G.; Orr, A.; Cowansage, K.; Strumberg, D.; Pais, G.C.G.; Voigt, J.H.; Nicklaus, M.C.; Winslow, H.E.; Zhao, H.; Turpin, J.A.; Yi, J.; Skalka, A.M.; Burke, T.R.; Pommier, Y. *J. Med. Chem.* **2002**, *45*, 5661.

[60] Barreca, M. L.; Lee, K.W.; Chimirri, A.; Briggs, J.M. *Biophys. J.* **2003**, *84*, 1450.

[61] Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M.C.; Milne, G.W.A.; Proksa, B.; Pommier, Y. *J. Med. Chem.* **1997**, *40*, 942.

[62] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.