Novel Virtual Screening Protocol Based on the Combined Use of Molecular Modeling and Electron-Ion Interaction Potential Techniques To Design HIV-1 Integrase Inhibitors

Cristina Tintori,[§] Fabrizio Manetti,[§] Nevena Veljkovic,[#] Vladimir Perovic,[#] Jo Vercammen,[‡] Sean Hayes,[||] Silvio Massa,[§] Myriam Witvrouw,[||] Zeger Debyser,[||] Veljko Veljkovic,[#] and Maurizio Botta*[,§]

Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Alcide de Gasperi 2, I-53100 Siena, Italy, Center for Multidisciplinary Research and Engineering, Institute of Nuclear Sciences VINCA, P.O. Box 522, 11001 Belgrade, Serbia, Molecular Medicine, Katholieke Universiteit Leuven and IRC KULAK, Kapucijnenvoer 33, B-3000 Leuven, Belgium, and Laboratory of Biomolecular Dynamics, Katholieke Universiteit Leuven, Celestijnenlaan 200G, B-3001 Leuven, Belgium

Received February 26, 2007

HIV-1 integrase (IN) is an essential enzyme for viral replication and represents an intriguing target for the development of new drugs. Although a large number of compounds have been reported to inhibit IN in biochemical assays, no drug active against this enzyme has been approved by the FDA so far. In this study, we report, for the first time, the use of the electron-ion interaction potential (EIIP) technique in combination with molecular modeling approaches for the identification of new IN inhibitors. An innovative virtual screening approach, based on the determination of both short- and long-range interactions between interacting molecules, was employed with the aim of identifying molecules able to inhibit the binding of IN to viral DNA. Moreover, results from a database screening on the commercial Asinex Gold Collection led to the selection of several compounds. One of them showed a significant inhibitory potency toward IN in the overall integration assay. Biological investigations also showed, in agreement with modeling studies, that these compounds prevent recognition of DNA by IN in a fluorescence fluctuation assay, probably by interacting with the DNA binding domain of IN.

INTRODUCTION

HIV-1 integrase (IN) is an essential enzyme for viral replication and represents an intriguing target for the development of new drugs.[1−3] IN catalyzes the integration of the viral DNA, previously transcribed by reverse transcriptase (RT), into the chromosomes of the host cell. The integration occurs in two temporally and spatially separated reactions, known as 3′-processing and strand transfer. The 3′-processing occurs in the cytoplasm, where IN binds the viral DNA and then removes a dinucleotide from each strand at the 3′-end adjacent to a conserved CA sequence. The protein−DNA complex is then transported into the nucleus where the strand transfer reaction takes place, and the 3′-ends of the viral DNA are covalently linked to the 5′-ends of the host DNA.[4]

Although many compounds have been reported to inhibit HIV-1 IN in biochemical assays,[5−8] no drug active against this enzyme has been approved by the FDA so far. 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, and compounds that preferentially inhibit the strand transfer reaction. The well-known diketo acids (DKAs)[9−11] and their recent derivatives[12−13] selectively inhibit the strand transfer reaction. Therefore, DKAs are referred to as IN strand transfer inhibitors (INSTIs). They have good ex vivo activities against HIV-1 replication, and their mechanism of action, thought not yet entirely explained, is likely to be a consequence of the interaction between the acid moiety and metal ion(s) within the IN active site, resulting in a functional sequestration of the critical metal cofactor(s).[14] Moreover, it was reported that they bind IN after the protein has formed a complex with viral DNA.[15] Unlike DKAs, styrylquinolines (SQs)[16,17] and pyranodipyrimidines (PDP)[18,19] are two classes of compounds characterized by the ability to inhibit the DNA binding of IN at low concentrations, and they are referred to as IN binding inhibitors (INBIs).

Within the process of drug discovery, computational methodologies are a well-established and essential tool. In fact, new technologies, such as combinatorial chemistry and high throughput screening (HTS), which allow the synthesis of millions or possibly billions of compounds, require chemists to confront an unimaginably large and diverse "chemical landscapes". Attempts must therefore be made to introduce a variety of computational techniques that allow chemists to reduce the huge molecular libraries to a more manageable size. Virtual screening (VS, also referred to as in silico screening) makes use of computational models to evaluate a specific biological activity of compounds in order to filter either existing databases or virtual libraries leading to the identification of molecules with activity against the target of interest. There are two basic approaches to VS, which depend on the available information about the 3D

* Corresponding author fax: (+39) 0577 234333; e-mail: botta@unisi.it.
[§] Università degli Studi di Siena.
[#] Institute of Nuclear Sciences VINCA.
[||] Molecular Medicine, Katholieke Universiteit Leuven and IRC KULAK.
[‡] Laboratory of Biomolecular Dynamics, Katholieke Universiteit Leuven.

10.1021/ci700078n CCC: $37.00
Published on Web 07/03/2007

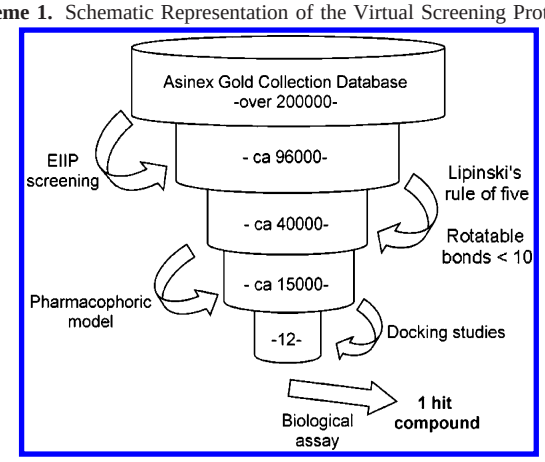
structure of the target under study. If this is known, from experimental (i.e., X-ray crystallographic studies, NMR studies) or theoretical (receptor structure built by homology modeling techniques) sources, structure-based drug design procedures can be applied (such as docking simulations). Otherwise, ligand-based drug design methods can be used (e.g., QSAR or 3D QSAR models, pharmacophoric models), based on the analysis of a number of ligands known to act with a common mechanism of action. The two VS approaches could be used separately or, alternatively, combined in the same protocol, as also described in this paper.

Both of these VS approaches are based on structural complementarity between the ligand and the binding site, and they taken into account exclusively the short-range interactions (SRI) between drug and target. However, it is well-established that not only do SRI play a key role in the recognition of a small molecule by a target but also the long-range interactions (LRI) are, of course, equally important. In this regard, it has been proposed that the electron-ion interaction potential (EIIP) represents the fundamental physical parameter of biological molecules determining their long-range properties.[20] Each of these two concepts (SRI and LRI) is necessary but not sufficient for a complete description of the target-drug interactions.

In this context, with the aim of identifying novel IN inhibitors, we employed an innovative VS approach based on the calculation of both short- and long-range interactions between molecules. This original VS protocol consists of the application of different sequential filters combining the EIIP technique, druglike property calculation, pharmacophoric model generation, and docking studies. Moreover, since our main goal 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, we took into account structural information coming from both known inhibitors of DNA binding and experimental data of the corresponding macromolecular target (IN), such as mutational studies and photocross-linking experiments.[21−27] The former were used to generate a pharmacophoric model containing the common features for DNA binding inhibition, while the latter were used to set up a docking protocol in the DNA-binding region. As a result of the VS of the commercial Asinex Gold Collection database, one of the selected compounds showed a significant inhibitory potency against IN in the overall integration assay. To validate this scaffold and generate SAR information, seven additional compounds were also chosen on the basis of a 2D substructure search. In vitro biological tests showed that six of these compounds are effective IN inhibitors, with IC_{50} values in the micromolar range, confirming that members of this family of compounds could be interesting hits for targeting IN. Further biological investigations (fluorescence fluctuation assay) showed, in agreement with modeling studies, that such compounds prevent recognition of DNA by IN, probably by interacting at the DNA binding domain of IN.

It is important to underline that the innovative VS protocol reported herein is able to describe the whole set of interactions between ligand and target. In fact, the ability for a small molecule to be recognized by the target was first evaluated by the EIIP technique, and then the shape complementarity with the binding site was evaluated by the common molecular modeling techniques. In addition, the preselection of molecules by means of the EIIP technique facilitates docking calculations to be carried out on a restricted number of compounds, and, since the docking filter is a very time-consuming step, the reported VS protocol results in a screening protocol faster than the structure-based virtual screening usually adopted.

Scheme 1. Schematic Representation of the Virtual Screening Protocol



EXPERIMENTAL AND COMPUTATIONAL DETAILS

Virtual Screening Protocol. The VS protocol reported in this study (Scheme 1) was based on the application of sequential filters to select a restricted number of compounds to be submitted to biological evaluation. In detail, (i) the EIIP screening was applied to the Asinex Gold Collection,[28] followed by (ii) a pruning based on the Lipinski's rule-of-five. Moreover, only compounds with a number of rotatable bonds lower than 10 were selected. Next, (iii) a three-dimensional pharmacophoric hypothesis was used as a search query on the selected compounds, retaining the molecules that adhere to all the features of the model. Finally, (iv) the binding mode of all retrieved compounds was evaluated by molecular docking using the 3D structure of HIV-1 IN. Docking calculations were performed into the IN core domain, in the region where the enzyme interacts with the DNA, widely characterized by mutation studies and photocross-linking experiments.[21−27] On the basis of docking scores as well as different structural scaffolds 12 compounds (**31**−**42**) were chosen for an in vitro assay against IN, leading to the identification of one active hit compound (**31**).

(i) EIIP Screening. The EIIP for organic molecules can be determined by the following simple equation derived from the "general model pseudopotential"[29−32]

$$W = 0.25Z^*\sin(1.04\pi Z^*)/2\pi \quad (1)$$

where Z^* is the average quasi-valence number (AQVN) determined by

$$Z^* = \Sigma^m n_i Z_i / N \quad (2)$$

where Z_i is the valence number of the ith atomic component, n_i is the number of atoms of the ith component, m is the number of atomic components in the molecule, and N is the total number of atoms. EIIP values calculated according to eqs 1 and 2 are expressed in Rydberg units (Ry).

EIIP/AQVN screening of both the learning set (1956 IN inhibitors collected in the HIV/OI Therapeutics Database,

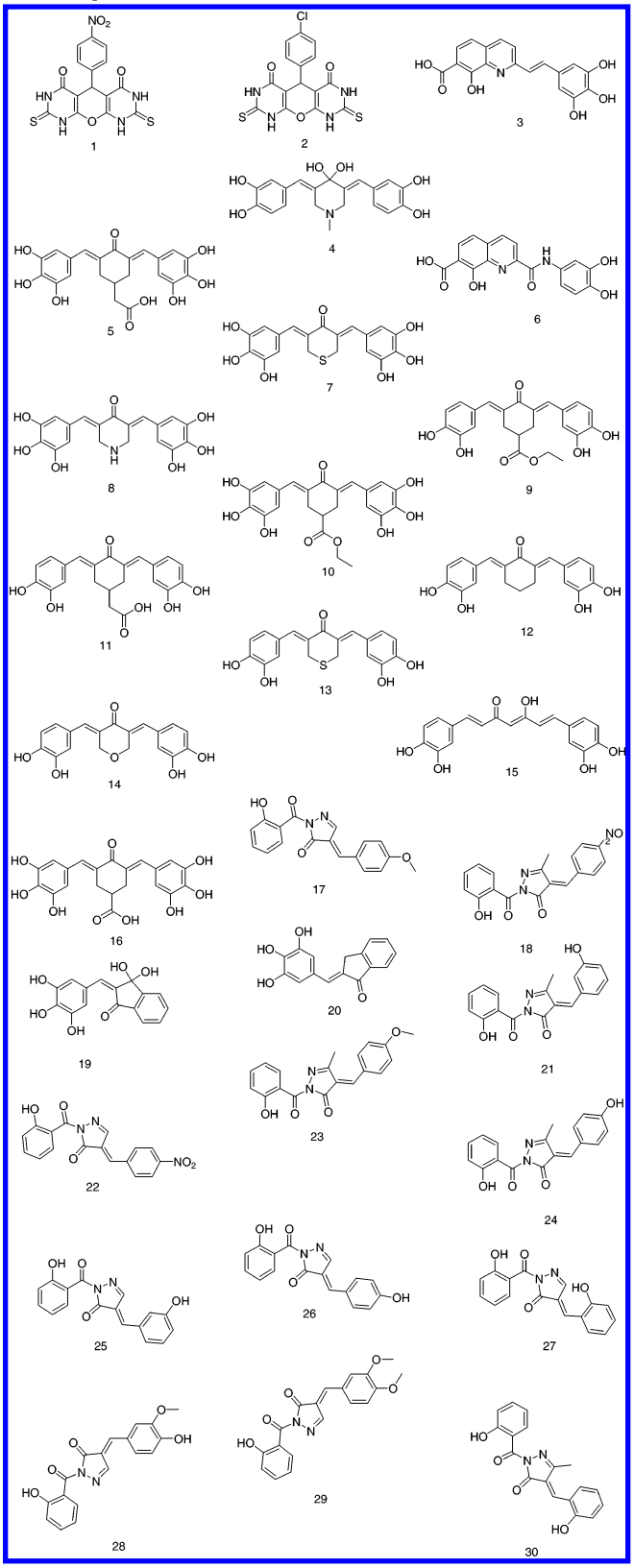
Division of AIDS, NIAID, http://chemdb2.niaid.nih.gov), the Asinex Gold Collection database, and the data set of IN inhibitors under clinical or preclinical trials was performed by the FormCal software, an in-house tool designed to execute simple and fast prescreening of large molecular libraries. Input data set utilized by FormCal is a bruto formula list of chemical compounds. The output file comprises corresponding AQVN and EIIP values as well as mean values of AQVN and EIIP for the data set analyzed. The software provides information about the distribution of the compounds according to their EIIP and AQVN values (i.e., the number of compounds with particular AQVN and EIIP values in a database analyzed and the standard deviation from the EIIP and AQVN mean values). First, FormCal was applied to the learning set of IN inhibitors to obtain both EIIP and AQVN mean values and standard deviations. Based on these data, the range for filtering IN inhibitor candidates was defined. Afterward, the Asinex Gold Collection database was screened, and compounds matching the predefined criterion were retrieved.

It is important to underline that the EIIP value does not represent a parameter related to the 2D/3D structure of the molecules, but it is simply derived from the bruto formula. As a consequence, molecules with comparable structure could have different values of EIIP and molecules with diverse structure could have similar values. EIIP values for numerous structures could be found in the literature.[20,55]

(ii) Druglike Property Calculations. Water solubility and cell permeability have long been recognized as key molecular properties in pharmaceutical science. Drug distribution, delivery, and transport depend on such properties. In an effort to reduce time and expense of the drug discovery and development process, it becomes apparent that an early analysis of the druglikeness is essential. Lipinski's rule-of-five (molecular weight < 500; logP < 5; number of hydrogen bond donors < 5; number of hydrogen bond acceptors < 10) is a simple method to foresee the leadlike properties of compounds under investigation. In this regard, we used the *Drug Discovery* module of Cerius2[33] to check if compounds derived from the previous step (the EIIP screening) were also able to fit these molecular properties. Compounds that violated more than one of those properties were removed. Moreover, based on the fact that the majority of known IN inhibitors have a rigid structure, we decided to keep only compounds with a number of rotatable bonds lower than 10. The flexibility of the compounds was evaluated by means of Cerius2 calculating the structural descriptor termed as rotatable bonds. The resulting set of compounds was transformed into a Catalyst[33] multiconformer (three-dimensional) database.

(iii) Pharmacophoric Models. Several HIV-1 IN pharmacophores have been already reported elsewhere,[34−42] built on DKA-like strand-transfer selective inhibitors or classes of inhibitors different from those used by us. Moreover, pharmacophoric models using compounds characterized by the ability to inhibit 3′-processing at low concentrations have not been published so far. On this basis, a set of 30 compounds (able to inhibit the 3′-processing with IC_{50} values $<1\ \mu M$) was collected from the literature (Chart 1) and used to generate a common feature pharmacophoric model by means of the HipHop approach (Catalyst software).[33] It compares diverse and highly active inhibitors to derive 3D hypotheses based on common chemical features, without considering biological activities. Such an approach seems to be the most appropriate method in this case, being the activity range of the compounds under investigation too narrow to generate a statistically significant quantitative model.

Chart 1. Structures of the Training Set Compounds Used for the Pharmacophore Model Generation



Strong inhibitors of the 3′-processing reaction (with comparable IC_{50} values and expected to bind in the same way in the active site of IN) were selected to generate pharmacophoric hypotheses. Moreover, compounds were chosen in such a way to maximize the type and the number of relative positions of the chemical features shared by the molecules. Following such rules, we took into account compounds belonging to four different chemical classes: styrylquinolines,[16,17] pyranodipyimidine,[18,19] cinnamoyl-based derivatives,[43,44] and salicylhydrazines.[45] A local minimum energy conformation was built for each compound using the 2D-3D sketcher implemented in Catalyst. The conformational space of all compounds was then explored by a conformational search by means of the "best" procedure, keeping all conformations within a 20 kcal/mol window from the global minimum and specifying 250 as the maximum number of conformers. Since the studied compounds showed comparable activity values, the highest weighting value was given to compounds **1** and **2**, being the most rigid molecules among the training set compounds. Accordingly, a "principal" value of 2 was set for these two inhibitors, while a value of 1 was applied for the remaining compounds. Additionally, to force HipHop to render only those pharmacophores that codify all the chemical features of the training set compounds, the "maximum omitted features" value was set to 0 for all inhibitors as well as the misses, feature misses, and complete misses parameters. The minimum interfeature spacing was set to 200 pm, while default settings were used for the remaining parameters (a detailed description of these input parameters can be found at the Accelrys Web page: http://www.accelrys.com/doc/life/catalyst410/tutorials/cat410_tutsTOC.html). The chemical features considered in the pharmacophoric model generation were the aromatic ring (RA), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), and hydrophobic groups (HY), chosen on the basis of the chemical features present in the studied compounds.

Ten hypotheses were generated, and the highest ranked pharmacophoric model (hereafter referred to as Hypo1) was chosen to screen the multiconformational database (built in the previous step) by application of the fast flexible fitting procedure, with the aim of retrieving only compounds that fully map Hypo1.

(iv) Docking Calculations. The software Autodock 3.0[46] was chosen as the tool to perform the automatic docking of previous selected compounds in the DNA binding domain of IN and to obtain an evaluation of the energy of interaction between such compounds and the enzyme. The X-ray crystallographic structure (2.2 Å resolution, PDB accession code 1BL3)[47] of the core domain of IN (chain C, residues 50-209) was chosen as the starting three-dimensional structure for docking studies, based on the fact that all residues were well resolved. The magnesium ion was included in the calculations, while the water molecules were removed. The X-ray structure was energy minimized by means of Macromodel 8.5[48] to avoid any steric bump.

In detail, the computational protocol consisted in the application of 10 000 steps of the stepest descent algorithm or until the derivative convergence was 0.01 kcal/Åmol. OPLS-AA force-field[49] with the continuum GB/SA solvation model (solvent water) was used.

The output structure of the minimization protocol was further manipulated to prepare the input structure for Autodock calculations by removing the nonpolar hydrogens, while Kollman united-atom partial charges and solvent parameters were added. Moreover, the file containing the final sets of atomic coordinates, together with the partial charges and the solvation parameters for each atom, was edited to insert the correct parameters for magnesium ion (charge: 2.000, solvation: −19.00, and fragmental volume: 8.17).[50]

Considering the data derived from experimental studies,[21−27] we decided to define a binding site constituted by the residues K156, N155, K159, Q148, and E92, all of them involved in both the 3′-processing and in the recognition of viral DNA by IN. The grid box dimensions (40 × 40 × 36 Å with a grid spacing of 0.375 Å) were set in such a way to accommodate all amino acids cited above. During the next step, several atom probes (characterized by the same stereoelectronic properties as the atoms constituting the compounds) were moved on the grid nodes, while the interaction energy between the probe and the protein was calculated at each node. In such a way, grid maps can be generated for each atom probe, describing its interactions with the inhibitor. Autogrid 3.0, as implemented in Autodock, was used to generate grid maps. Similarly to the protein, the structures of all compounds identified from the previous pharmacophore-based selection were also prepared by adding atomic charges (Gasteiger) using Open Babel 1.100.2 software.[51] Finally, the rigid root and rotatable bonds for each compounds were defined using the autotors module of Autodock. The Lamarckian genetic algorithm (LGA) was employed to generate orientations/conformations of the compounds within the binding site. For each of the 10 independent runs, a maximum number of 100 000 operations was performed with a population size of 100 individuals. The highest ranked solutions of the ligands were taken for further analysis.

Enzyme Assays. *Overall Integration Assay*. To determine the susceptibility of HIV-1 IN to different compounds, an enzyme-linked immunosorbent assay (ELISA) was performed. This assay uses an oligonucleotide substrate of which one oligo (5′-ACTGCTAGAGATTTTCCACACTGACTAAAAGGGTC-3′) is labeled with biotin on the 3′-end and the other oligo is labeled with digoxigenin at the 5′-end. For the overall integration assay, the second 5′-digoxigenin labeled oligo is 5′-GACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA**GT**-3′. The integrase was diluted in 750 mM NaCl, 10 mM Tris pH 7.6, 10% glycerol, and 1 mM β-mercapto ethanol. To perform the reaction, 4 μL of diluted IN (corresponding to a 1.6 μM concentration of wt IN) and 4 μL of annealed oligos (7 nM) were added in a final reaction volume of 40 μL containing 10 mM $MgCl_2$, 5 mM DTT, 20 mM HEPES pH 7.5, 5% PEG, and 15% DMSO. The reaction was carried out for 1 h at 37 °C. These reactions were followed by an immunosorbent assay on avidin coated plates.[52]

Fluorescence Fluctuation Assay. The oligonucleotides INT1-TMR5 and INT2 (Eurogentec, Seraing, Belgium) were used for the fluorescent assay with a sequence corresponding to the U5 LTR ends of the HIV-1 genome. INT1-TMR5 (5′-tetramethylrhodamine-TGTGGAAAATCTCTAGCAGT-3′) has a tetramethylrhodamine attached at the 5′-end, and INT2 (5′-ACTGCTAGAGATTTTCCACA-3′) is the complementary oligonucleotide of INT1-TMR5. They were synthesized and purified by Eurogentec. The final DNA

Table 1. Compounds Reported as Potent IN Inhibitors

no.	HIV-1 IN inhibitor	references
1	L-870810	Hazuda, D. et al. J. *Science* **2000**, *287*, 646−650.
2	S-1360	Billich, A. *Curr. Opin. Invest. Drugs* **2003**, *4*, 206−209.
3	FZ41	Bonnenfant, S. et al. *J. Virol.* **2004**, *78*, 5728−5736.
4	dicafeoylquinic acid	Robinson, W. E. et al. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6326−6331.
5	JTK-303/GS-9137	Matsuzaki, Y. et al. 13th Conference of Retroviruses and Opportunistic Infections. Denver, U.S.A., 2006.
6	L-731988	Wai, J. S. et al. *J. Med. Chem.* **2000**, *43*, 4923−4926.
7	ITI-367	Haffar, O. et al. *J. Virol.* **2005**, *79*, 13028−13036.

concentration of the fluorescent dsDNA was determined using the ConfoCor I Zeiss (EVOTEC, Jena, Germany). All chemicals used were of analytical grade.

A commercial FCS set up (the ConfoCor I) was used. In this configuration, a C-Apochromat 40×/1.2W objective lens was used, the pinhole diameter was 45 μm in all experiments, and a typical size of the excitation volume was 1.4 fl. The 543 nm line of the HeNe ion laser was applied in all experiments, attenuated with a 0.3 OD neutral density filter. The laser beam was focused at about 180 μm above the bottom of the Nunc cuvettes (Nalge Nunc International, Naperville, IL) in a typical volume of 10 μL. The data electronics and software (Borland Delphi) were used as described earlier.[53] The assay was carried out using the same reaction conditions as the ELISA assay and the method as described earlier.[54]

RESULTS AND DISCUSSION

EIIP-Based Compound Filtering. It has been proposed that the number of valence electrons and EIIP, representing the main energy term of valence electrons, are essential physical parameters determining the long-range properties of biological molecules.[20] Moreover, it has been demonstrated that there is a strong correlation between EIIP of organic molecules and their biological activity (mutagenicity, carcinogenicity, toxicity, antibiotic, and cytostatic activity, etc.).[31,32] Previously, an EIIP-based VS technique has been successfully applied together with other complementary structural approaches for selection of candidate HIV-1 entry inhibitors.[55]

Here, we describe an EIIP/AQVN-based filter, suitable for the rapid in silico prescreening of large chemical libraries, to be applied in order to eliminate compounds with inadequate LRI properties. In detail, such a filter is established by analyses of compounds reported to exhibit anti-IN activity, irrespective of their mode of action and collected from the HIV/OI Therapeutics Database. The EIIP and AQVN values were calculated for 1956 compounds, leading to the anti-IN mean values of 0.0798 Ry for EIIP and 3.1 for AQVN. These values were considered as characteristic properties of HIV-1 IN inhibitors and subsequently used to define the interval for screening the Asinex Gold Collection database. In fact, the range of acceptable values either for EIIP or AQVN parameters was set to the corresponding mean ± 25% of standard deviation, leading to EIIP values ranging from 0.044 to 0.116 Ry and AQVN values ranging from 3.0 to 3.2. This simple rule-based filter was applied to the Asinex database (Figure 1) and led to the preselection of about 96 000 compounds to be used for further virtual screening.

Additionally, as a proof of the concept, we calculate the EIIP and AQVN values for the HIV-1 IN inhibitors which

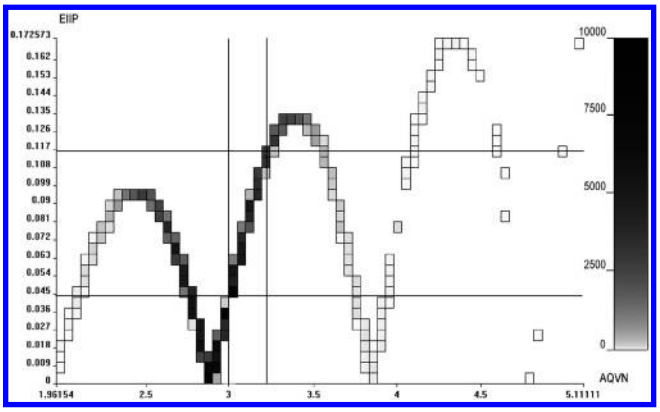


Figure 1. The distribution of compounds contained in the Asinex Gold Collection database according to their EIIP and AQVN characteristics. Data about the number of compounds with particular AQVN and EIIP values are coded according to the right greyscale bar (black indicates 10 000 compounds, white indicates 0 compounds, and gray indicates an intermediate number of compounds). Rectangular area encompasses approximately 96 000 compounds that have EIIP and AQVN values within the range accepted for HIV-1 integrase inhibitor candidates, i.e., EIIP (0.044−0.116 Ry) and AQVN (3.0−3.2).

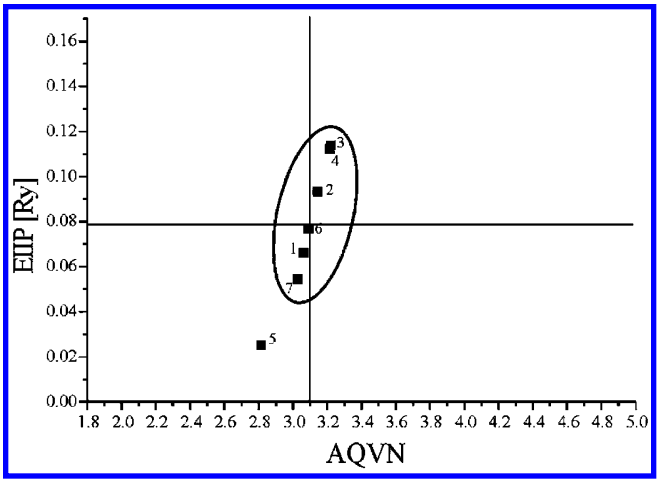


Figure 2. The AQVN and EIIP values for HIV-1 integrase inhibitors listed in Table 1. The *X*-axis represents the AQVN values, while the *Y*-axis represents corresponding EIIP values (Ry).

are currently or were previously involved in preclinical and clinical trials (Table 1). This analysis reveals that the LRI characteristics of these compounds are clustered around the standard for anti-IN AQVN and EIIP values, with the exception of the JTK-303/GS-9137 (Figure 2).

Pharmacophoric Models and Database Search. Pharmacophoric modeling requires molecules showing a large variety in chemical structure and which interact through the same binding mechanism with the target of interest. Therefore, among compounds known in the literature as IN inhibitors, we chose to focus on compounds acting at the

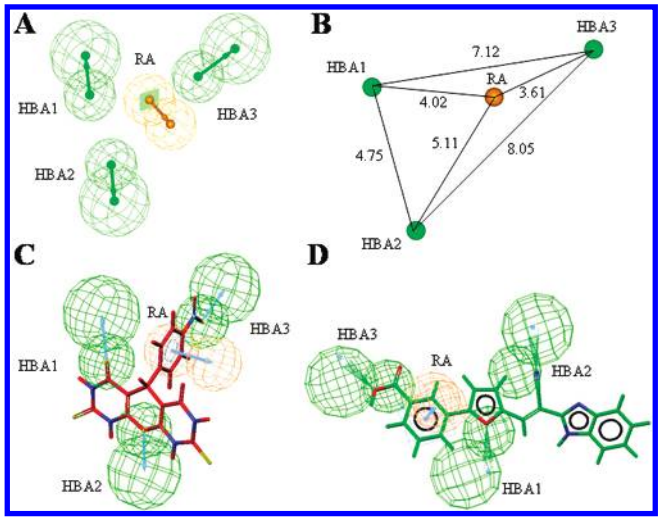
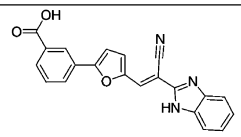
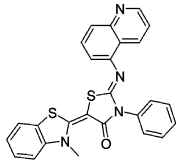
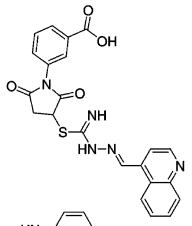
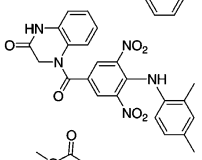
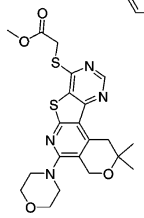
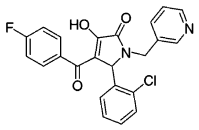
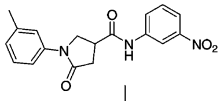
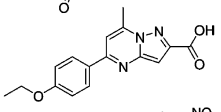
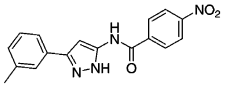
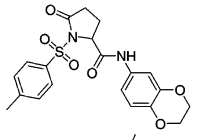
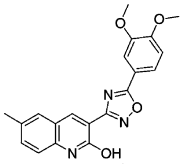
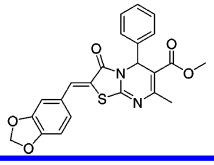


Figure 3. A. A graphical representation of the pharmacophoric hypothesis Hypo1 consisting of three hydrogen bond acceptors (HBA, green) and an aromatic ring (RA, orange); B. a schematic representation of the distances (expressed in Å, with a tolerance of ±0.8 Å) between the features; C. the alignment of compound **1** (red) on Hypo 1; and D. the alignment of compound **31** (green) on Hypo 1.

DNA binding step. Starting from a set of 30 highly active 3′-processing inhibitors (**1**−**30**, Chart 1), a common feature hypothesis generation (HipHop routine, Catalyst software) was carried out, leading to 10 pharmacophoric hypotheses with scores ranging from 229 to 267. Among them, the highest ranked model (Hypo1) was used for subsequent calculations. Hypo1 (Figure 3) consisted of three hydrogen bond acceptors (HBA1, HBA2, HBA3) almost equidistant from an aromatic ring (RA). Superposition of the training set compounds onto Hypo1 resulted in fast fit scores ranging from 0.95 to 4 (since Hypo1 has four equiweighted features and each of them has a weight of 1, 4 is the maximum value for the fit of a ligand into the model). Hypo 1 was then used as a 3D query to screen the 40 000 compounds (converted in a multiconformational database) by means of the "Fast Flexible Search Databases/Spreadsheets" option. Only structures able to map all the features of the pharmacophoric template were retrieved. As a result, out of the about 40 000 entries constituting the database, Hypo 1 selected about 20 000 compounds, which were in turn reduced to about 15 000 entries by application of a 0.95 cutoff in terms of the fit value (0.95 was the lowest fit score found for training set compounds aligned to Hypo1). As an example, Figure 3C,D shows the superimposition of compounds **1** and **31**, respectively, on Hypo 1.

Docking Studies. The software Autodock was chosen to perform automatic docking simulations of selected compounds in the DNA binding domain of IN and to evaluate the interaction energy between such compounds and the enzyme. To this aim, the docking protocol was first checked on all 30 compounds constituting the training set. In detail, the three-dimensional structure of IN (entry 1BL3 of the Brookhaven Protein Data Bank)[47] was relaxed through an energy minimization routine (all-atom Amber* force field, Polack-Ribiere conjugate gradient method, continuum solvation with water, extended cutoffs, convergence at 0.01 kJ/mol·Å) to avoid any steric bump that may occur in the X-ray structure. Next, a docking protocol was applied to find all the profitable interactions between known 3′-processing

Table 2. Structure and Antiviral Activity of Compounds Selected by VS Protocol[a]

Compound	Compound code	Structure	IC_{50} μM (standard deviation)
31	BAS-0314191		69 (13)
32	BAS-0112410		> 214
33	BAS-0227358		>223
34	BAS-0308666		>217
35	BAS-0338312		>217
36	BAS-1815208		>236
37	BAS-0728912		>295
38	BAS-1516086		>336
39	BAS-2231176		>310
40	BAS-8979893		>240
41	BAS-8978276		>275
42	BAS-0663648		>230

[a] IC_{50} values are referred to as the results of the overall integration assay. Data represent the mean values of at least two separate experiments.

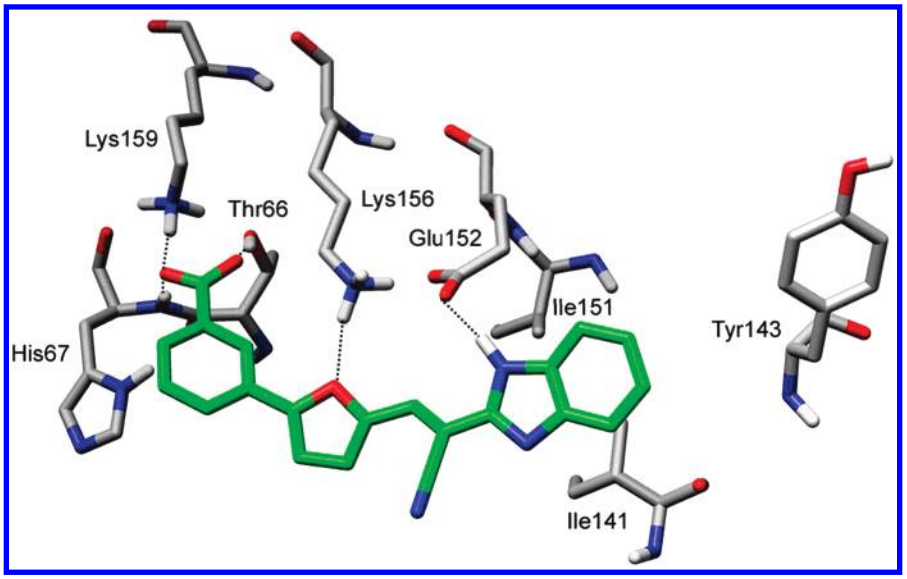


Figure 4. The binding mode of compound **31** (green) within the HIV-1 IN core domain. For the sake of clarity, nonpolar hydrogens were omitted, hydrogen bond interactions are represented by black dotted lines, and only a few residues of the binding site were displayed.

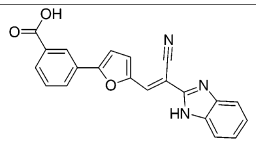
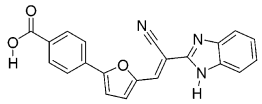
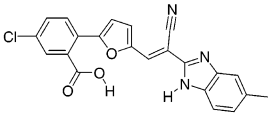
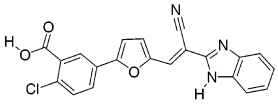
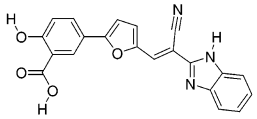
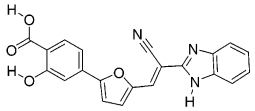
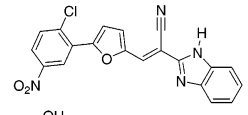
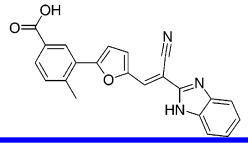
inhibitors (training set compounds) and IN. Inspection of results showed that all the inhibitors showed an energy of binding in the range between −6.2 and −8.5 kcal/mol.

As a consequence, as a part of the compound selection process, the docking procedure was applied to compounds selected during the pharmacophore-based search. Results were collected according to the binding energy values, retaining about 300 compounds with energy of binding greater than or equal to −6.2 kcal/mol (corresponding to the lowest energy of binding found among the training set compounds). Obviously, being the docking protocol only based on active compounds, the filter will be able, in principle, to select compounds that could be potentially active, but it will not be able to discard potentially inactive compounds. In the last step, on the basis of three different parameters (namely, energy of binding, structural diversity, and commercially availability of the appropriate amount for biological tests), 12 compounds (**31**−**42**, Table 2) were purchased and submitted to an in vitro assay.

Biological Investigations. Biological data, reported in Tables 2 and 3, showed that compound BAS-0314191 (**31**) inhibited IN activity, in the overall integration assay, with an IC_{50} value of 69 μM. In addition, we performed further biological investigations to assay the effects of the new compounds on the formation of IN-viral DNA complexes independently from the catalytic process. In fact, the capability of such a compound to inhibit the IN-viral DNA complex formation was assessed through a fluorescence fluctuation assay, showing its ability to inhibit the binding of IN to viral DNA with a K_i value of 19 μM. Moreover, among the 7 analogues (**43**−**49**) of the hit compound (**31**) coming from a substructure search (see below), six compounds inhibited IN in the overall integration assay with IC_{50} values from 10 to 90 μM.

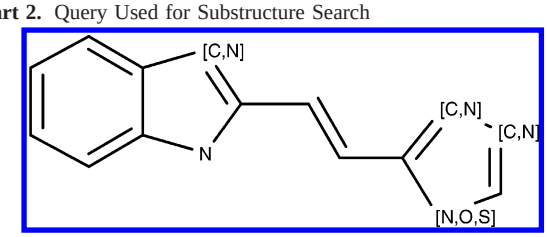
Substructure Search. On the basis of biological results on compound **31**, the ranking list obtained with the above-described flowchart was further analyzed to find the next congeneric compounds. In detail, to make the research faster, a substructure codifying the chemical features of compound **31** (Chart 2) was used. A small set of additional 7 compounds were selected (**43**−**49**). Their structures and biological data are reported in Table 3.

Table 3. Structure and Inhibitory Activity of Compounds Selected by Substructure Search[a]

Compound name	Compound code	Structure	IC_{50} μM (standard deviation)
31	BAS-0314191		69 (13)
43	BAS-01948546		> 281
44	BAS-02514584		25 (0.0)
45	BAS-0314123		90 (4.3)
46	BAS-0330647		35 (12.6)
47	BAS-0442786		71 (48.0)
48	BAS-0717929		10 (6.1)
49	BAS-07562890		53 (5.8)

[a] IC_{50} values relate to the inhibition in the overall integration assay. Data represent the mean values of at least two separate experiments.

Chart 2. Query Used for Substructure Search



Binding Mode Analysis. Analysis of the minimized complexes between the inhibitors **31**, **44**−**49**, and IN was performed with the aim of hypothesizing the way by which they prevent the recognition of DNA and, thus, getting insights on the peculiar chemical features that influenced their activity. Inspection of results showed that all the inhibitors exhibited a common orientation within the IN core domain. The predicted binding mode of the minimum energy conformation of the hit compound **31** (with an estimated free energy of binding of −8.3 kcal/mol) is shown in Figure 4. In detail, the benzoic ring was located in the region usually occupied by the phosphate groups of viral DNA and characterized by hydrogen bond contacts with polar amino acids. Electrostatic interactions between the acidic group of the inhibitor and the side chains of Thr66 and Lys159 as well as with the backbone NH group of His67 contributed significantly to the stabilization of the complex. The benzimidazolyl moiety was directed toward the flexible loop constituted by residues 140−149 and its NH group interacted with the carboxylate of Glu152. The furan oxygen was involved in an additional hydrogen bond with the terminal moiety of Lys156. Moreover, hydrophobic interactions between the aromatic side chain of Tyr143 and the benzimidazole moiety of the ligand were found as well as positive van der Waals contacts involving the side chain of Ile151, Ile141, Gly149, and Glu152 (the aliphatic portion of the side chain). It is interesting to note that compound **43**, differing from the hit compound **31** only for the substitution pattern of the benzoic phenyl group, was found to be inactive. Results from docking simulations accounted for its inactivity, being that compound **43** was unable to interact with Lys159, differently from the active inhibitors **31** and **44**−**49**. This is in full agreement with the fact that Lys159 is known to be a critical residue for the binding with IN.

CONCLUSIONS

In this work, we present an innovative and fast VS protocol that takes into account both short- and long-range interactions between interacting molecules. In fact, EIIP calculations (accounting for the long-range properties of organic molecules), together with molecular modeling tools (able to evaluate the short-range interactions between interacting molecules), were employed with the aim of identifying new IN inhibitors. Application of this powerful VS protocol to screen a database of commercially available compounds facilitated the identification of a novel structural family of IN inhibitors, which necessitated the biological evaluation of only 0.006% of the compound library. Further biological investigations (fluorescence fluctuation assay), in agreement with modeling studies, showed that these compounds prevent recognition of DNA by IN, probably by interacting at the DNA binding domain of IN itself. In addition, this protocol adopts a holistic approach, incorporating both long- and short-range molecular interactions, druglikeness, pharmacophore recognition, and docking studies. By extension of this logic, one could envisage that the further evolution of this protocol will include efficient filter elements which incorporate an intestinal absorption, such as polar surface area[56] and toxicological predictors.[57−60]

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 the work. We acknowledge Linda Desender and Martine Michiels for excellent technical assistance.

REFERENCES AND NOTES

(1) Nair, V. Novel inhibitors of HIV integrase: the discovery of potential anti-HIV therapeutic agent. *Curr. Pharm. Des.* **2003**, *9*, 2553−2565.
(2) 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.
(3) Gupta, S. P.; Nagappa, A. N. Design and development of integrase inhibitors as anti-HIV agents. *Curr. Med. Chem.* **2003**, *10*, 1779−1794.
(4) Mazumder, A.; Gazit, A.; Levitzki, A.; Nicklaus, M.; Yung, J.; Kohlhagen, G.; Pommier, Y. Effects of tyrphostins, protein kinase inhibitors, on human immunodeficiency virus type 1 integrase. *Biochemistry* **1995**, *34*, 15111−15122.
(5) Dayam, R.; Deng, J.; Neamati, N. HIV-1 integrase inhibitors:2003−2004 update. *Med. Res. Rev.* **2006**, *26*, 271−309.
(6) Young, S. D. Inhibition of HIV-1 integrase by small molecules: the potential for a new class of AIDS chemotherapeutics. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 402−410.
(7) Craigie, R. HIV integrase, a brief overview from chemistry to therapeutics. *J. Biol. Chem.* **2001**, *276*, 23213−23216.
(8) Cotelle, P. Patented HIV-1 integrase inhibitors (1998−2005). *Recent Pat. Anti-Infect. Drug Discovery* **2006**, *1*, 1−15.
(9) Yoshinaga, T. S.; Fujishita, T.; Fujiwara, T. In vitro activity of a new HIV-1 integrase inhibitor in clinical development. 9th Conference on Retroviruses and Opportunistic Infections, Seattle, WA, 2002.
(10) Young, S. L-870,810: a potent antiviral HIV integrase inhibitor with potential clinical utility. XIV International AIDS Conference, West Point, PA, 2002.
(11) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K.; Grobler, J. A.; Espeseth, A.; Gabrielski, L.; Schleif, W.; Blau, C.; Miller, M. D. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **2000**, *287*, 646−650.
(12) Zhuang, L.; Wai, J. S.; Embrey, M. W.; Fisher, T. E.; Egbertson, M. S.; Payne, L. S.; Guare, J. P.; 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. Design and synthesis of 8-hydroxy-[1,6]-naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. *J. Med. Chem.* **2003**, *46*, 453−456.
(13) Sato, M.; Motomura, T.; Aramaki, H.; Matsuda, T.; Yamashita, M.; Ito, Y.; Kawakami, H.; Matsuzaki, Y.; Watanabe, W.; Yamataka, K.; Ikeda, S.; Kodama, E.; Matsuoka, M.; Shinkai, H. Novel HIV-1 integrase inhibitors derived from quinolone antibiotics. *J. Med. Chem.* **2006**, *49*, 1506−1508.
(14) Grobler, J. A.; Stilmock, 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. and Hazuda D. J. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6661−6666.
(15) Espeseth, A. S.; Felock, P.; Wolfe, A.; Witmer, M.; Grobler, J.; Anthony, N.; Egbertson, M.; Melamed, J. Y.; Young, S.; Hamill, T.; Cole, J. L.; Hazuda, D. J. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11244−11249.
(16) Deprez, E.; Barbe, S.; Kolaski, M.; Leh, H.; Zouhiri, F.; Auclair, C.; Brochon, J. C.; Le Bret, M.; Mouscadet, J. F. Mechanism of HIV-1 integrase inhibition by styrylquinoline derivatives in vitro. *Mol. Pharmacol.* **2004**, *65*, 85−98.
(17) Zouhiri, F.; Mouscadet, J. F.; Mekouar, K.; Desmaële, D.; Savouré, D.; Leh, H.; Subra, F.; Le Bret, M.; Auclair, C.; D'Angelo, J. Structure-activity relationship and binding mode of styrylquinolines as potent

inhibitors of HIV-1 integrase and replication of HIV-1 in cell culture. *J. Med. Chem*. **2000**, *43*, 1533−1540.
(18) Pannecouque, C.; Pluymers, W.; Van Maele, B.; Tetz, V.; Cherepanov, P.; De Clercq, E.; Witvrouw, M.; Debyser, Z. New class of HIV integrase inhibitors that block viral replication in cell culture. *Curr. Biol*. **2002**, *12*, 1169−1177.
(19) Witvrouw, M.; van Maele, B.; Vercammen, J.; Hantson, A.; Engelborghs, Y.; De Clercq, E.; Pannecouque, C.; Debyser, Z. Novel inhibitors of HIV-1 integration. *Curr. Drug Metab*. **2004**, *5*, 291−304.
(20) Veljkovic, V. *A theoretical approach to preselection of carcinogens and chemical carcinogenesis*; Gordon & Breach: New York, 1980.
(21) 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.
(22) 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.
(23) 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.
(24) 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.
(25) Gerton, J. L.; Ohgi, S.; Olsen, M.; Derisi, J.; Brown, P. O. Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *J. Virol*. **1998**, *6*, 5046−5055.
(26) 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.
(27) Dirac, A. M. G.; Kjems, J. Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur. J. Biochem.* **2001**, *268*, 743−751.
(28) Further information at the Web site http://www.asinex.com/prod/gold.html (accessed June 2005).
(29) Veljkovic, V.; Slavic, I. Simple general-model pseudopotential. *Phys. Lett*. **1972**, *29*, 105−107.
(30) Veljkovic, V. The dependence of the Fermi energy on the atomic number. *Phys. Lett.* **1973**, *45A*, 41−42.
(31) Veljkovic, V.; Lalovic, D. General model pseudopotential for positive ions. *Phys. Lett*. **1973**, *45A*, 59−60.
(32) Veljkovic, V.; Lalovic, D. Theoretical prediction of mutagenicity and carcinogenicity of chemical substances. *Cancer Biochem. Biophys*. **1976**, *1*, 295−298.
(33) Cerius2 and Catalyst are distributed by Accelrys, Inc., San Diego, CA.
(34) Barreca, M. L.; Ferro, S.; Rao, A.; De Luca, L.; Zappala, M.; Monforte, A. M.; Debyser, Z.; Witvrouw, M.; Chimirri, A. Pharmacophore-based design of HIV-1 integrase strand-transfer inhibitors. *J. Med. Chem.* **2005**, *48*, 7084−7088.
(35) Barreca, M. L.; Rao, A.; De Luca, L.; Zappala, M.; Gurnari, C. et al. Efficient 3D database screening for novel HIV-1 IN inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1450−1455.
(36) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S.; Chen, J.; Milne, G. W.; Pommier, Y. HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920−929.
(37) Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S.; Nicklaus, M. C.; Milne, G. W.; Proksa, B.; Pommier, Y. Depsides and depsidones as inhibitors of HIV-1 integrase: discovery of novel inhibitors through 3D database searching. *J. Med. Chem.* **1997**, *40*, 942−951.
(38) Hong, H.; Neamati, N.; Winslow, H. E.; Christensen, J. L.; Orr, A.; Pommier, Y.; Milne, G. W. Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. *Antiviral Chem. Chemother*. **1998**, *9*, 461−472.
(39) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem*. **2000**, *43*, 2100−2114.
(40) Mustata, G. I.; Brigo, A.; Briggs, J. M. HIV-1 IN pharmacophore model derived from diverse classes of inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1447−1454.
(41) Dayam, R.; Sanchez, T.; Clement, O.; Shoemaker, R.; Sei, S.; Neamati, N. Beta-diketo acid pharmacophore hypothesis 1. Discovery of a novel class of HIV-1 integrase inhibitors. *J. Med. Chem.* **2005**, *48*, 111−120.
(42) Dayam, R.; Sanchez, T.; Neamati, N. Diketo acid pharmacophore. 2. Discovery of structurally diverse inhibitors of HIV-1 integrase. *J. Med. Chem*. **2005**, *48*, 8009−8015.
(43) Artico, M.; Di Santo, R.; Costi, R.; Novellino, E.; Greco, G.; Massa, S.; Tramontano, E.; Marongiu, M. E.; De Montis, A.; La Colla, P. Geometrically and conformationally restrained cinnamoyl compounds as inhibitors of HIV-1 integrase: synthesis, biological evaluation and molecular modeling. *J. Med. Chem.* **1998**, *41*, 3948−3960.
(44) Buolamwini, J. K.; Assefa, H. CoMFA and CoMSIA 3D QSAR and Docking Studies on Conformationally-Restrained Cinnamoyl HIV-1 Integrase Inhibitors: Exploration of a Binding Mode at the Active Site. *J. Med. Chem*. **2002**, *45*, 841−852.
(45) Neamati, N.; Hong, H.; Owen, J. M.; Sunder, S.; Winslow, H. E.; Christensen, H. Z.; Burke, T. R.; Milne, G. W. A.; Pommier, Y. Salicylhydrazine-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site. *J. Med. Chem.* **1998**, *41*, 3202−3209.
(46) 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.
(47) 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.
(48) *Macromodel, version 8.5*; Schroedinger, LLC: Portland, OR, 2003.
(49) 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.
(50) See "Solvation parameters assignment". http://www.accelrys.com/doc/pdf/affinity2K.pdf (accessed June 2007).
(51) Further information at the Web site http://babel.ifarchive.org/ (accessed June 2007).
(52) Hwang, Y.; Rhodes, D.; Bushman, F. Rapid microtiter assays for poxvirus topoisomerase, mammalian type IB topoisomerase and HIV-1 integrase: application to inhibitor isolation. *Nucleic Acids Res.* **2000**, *28*, 4884−4892.
(53) Van, Craenenbroeck, E.; Vercammen, J.; Matthys, G.; Beirlant, J.; Marot, C.; Hoebeke, J.; Strobbe, R.; Engelborghs, Y. Heuristic statistical analysis of fluorescence fluctuation data with bright spikes: application to ligand binding to the human serotonin receptor expressed in Escherichia coli cells. *Biol. Chem*. **2001**, *382*, 355−361.
(54) Vercammen, J.; Maertens, G.; Gerard, M.; De Clercq, E.; Debyser, Z.; Engelborghs, Y. DNA-induced polymerization of HIV-1 integrase analyzed with fluorescence fluctuation spectroscopy. *J. Biol. Chem.* **2002**, *277*, 38045−38052.
(55) Veljkovic, V.; Veljkovic, N.; Este, J. A.; Hüther, A.; Dietrich, U. *Curr. Med. Chem.* **2007**, *14*, 133−155.
(56) Ertl, P.; Rohde, B.; Selzer, P. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(57) Greene, N. Computer systems for the prediction of toxicity: an update. *Adv. Drug Delivery Rev.* **2002**, *54*, 417−431.
(58) Richard, A. M. Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet? *Mutat. Res.* **1998**, *400*, 493−507.
(59) Judson, P. N.; Vessey, J. D. A comprehensive approach to argumentation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1356−1363.
(60) Judson, P. N.; Marchant, C. A.; Vessey, J. D. Using argumentation for absolute reasoning about the potential toxicity of chemicals. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1364−1370.

CI700078N