Dynamic Receptor-Based Pharmacophore Model Development and Its Application in Designing Novel HIV-1 Integrase Inhibitors

Jinxia Deng,[†,§] Keun Woo Lee,[‡,||] Tino Sanchez,[§] Meng Cui,[‡] Nouri Neamati,[§,*] and James M. Briggs[†,‡,*]

Department of Chemical Engineering, University of Houston, Houston, Texas 77204-4004, Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, and Department of Pharmaceutical Sciences, School of Pharmacy, University of Southern California, 1985 Zonal Avenue, PSC304A, Los Angeles, California 90089

Received July 24, 2004

We present here a dynamic receptor-based pharmacophore model representing the complementary features of the active site region of HIV-1 integrase (IN), which was developed from a series of representative conformations of IN. Conformations of IN were sampled through a molecular dynamics study of the catalytic domain of an IN monomer, and an ensemble of representative IN structures were collected via a probability-based representative conformer sampling method that considers both the potential energy and the structural similarity of the protein conformations. The dynamic pharmacophore model was validated by a set of 128 known inhibitors, and the results showed that over 72% of the active inhibitors (IC_{50} lower than 20 μM) could be successfully identified by the dynamic model. Therefore, we screened our in-house database of commercially available compounds against this model and successfully identified a set of structurally novel IN inhibitors. Compounds **7** and **18** with IC_{50}s of 8 μM and 15 μM, respectively, against the strand transfer reaction were the most potent. Moreover, **7**, **8** and **20** showed a 5-fold selectivity for the strand transfer reaction over 3′-processing.

Introduction

HIV encodes three enzymes: reverse transcriptase, integrase, and protease. Drugs targeting the reverse transcriptase and protease have been available over 10 years. The currently employed treatment for HIV infection and AIDS is a combination of inhibitors of the reverse transcriptase and protease (HAART).[1] However, problems of drug toxicity and drug resistance require a new target for treatment of AIDS. HIV-1 integrase, a 32-kDa enzyme containing a conserved D,D(35)E catalytic motif, has been an attractive and a validated target for anti-AIDS drug design because of its crucial role in the viral life cycle and the fact that there is no cellular homologue in humans. Integrase is a testable target because rapid and sensitive assays exist for ascertaining enzymatic activity, and crystal and NMR structures are available for use in rational structure-based drug design.[2−5] However, integrase does not have a well-defined, buried binding pocket, but rather has a shallow solvent-exposed one, which has led to difficulties for structure-based drug discovery. In addition, it exists in a multimeric state in preintegration complexes. Furthermore, there is currently no structure available for the full-length enzyme. Many IN inhibitors have been identified, designed, or synthesized showing that effective inhibition of IN is possible.[3,5−8] In addition, a number of natural products inhibiting integrase have been reported recently.[9] Several pharmacophore models have been developed based on sets of the active inhibitors, and a series of novel categories of inhibitors have been discovered.[10−12] The dynamic pharmacophore method, developed by Carlson,[13] considers multiple receptor conformations that are used to generate a receptor-based pharmacophore model which is then used to search available compound databases. The protein conformations were collected at 100 ps intervals (i.e. effectively at random) from a 500-ps molecular dynamics trajectory.[13] Due to limitations in the model development which only focused on part of the active site, i.e. the functional features associated only with D64 and D116 (no metal ions or E152), only hydrogen-bond donor features were favored in the active site. Despite tremendous endeavors focused on drug discovery for IN, only two diketo compounds have thus far entered clinical trials.[14,15]

A number of integrase core domain crystal structures have become available in the past few years; however, only one complex has been solved with an inhibitor (1QS4; 5-CITEP) in the active site of the catalytic domain.[16] The inhibitor was found to associate with an essential residue, E152, by donating a hydrogen-bond. In addition, one Mg^{2+} ion was resolved in the complex chelated with D64 and D116. Therefore, we chose the complex (1QS4) containing the inhibitor, 5-CITEP, in its core domain as our initial template for a 1 ns molecular dynamics (MD) simulation (Lee, M. C., et al., unpublished results) and the subsequent pharmacophore model development which included the Mg^{2+}, D64, D116, and E152 in the active site, as well as the nearby residues K156 and K159.

In this work, we first present a collection of the representative conformations from a 1 ns molecular dynamics trajectory that were selected since they statistically represent the dynamic behavior of IN by taking into account both the similarity of the conformations

* Corresponding authors. Neamati: Tel: 323-442-2341, Fax: 323-442-1390, e-mail: neamati@usc.edu; Briggs: Tel: 713-743-8366, Fax: 713-743-8351, e-mail: jbriggs@uh.edu.
† Department of Chemical Engineering, University of Houston.
‡ Department of Biology and Biochemistry, University of Houston.
§ University of Southern California.
|| Current address: Department of Biochemistry, Gyeongsang National University, 900 Gazwa-dong, Jinju, GN 660−701 Korea.

10.1021/jm049410e CCC: $30.25
Published on Web 02/15/2005

and their potential energies. The conformational snapshots were selected via a recently developed probability-based representative conformer sampling method (PRCSM) (Lee, K. W.; Briggs, J. M., unpublished results). Second, we report on both the development of a static model based on a single conformation of the protein and on a dynamic pharmacophore model. Both models included the entire active site, rather than the subset used in the previous work.[13] Third, we discovered a set of structurally novel IN inhibitors from database screening with the dynamic pharmacophore model.

Computational Methods

Creating an Ensemble of Representative Snapshots from MD Trajectories. A 1 ns MD trajectory from a prior study (Lee M. C., et al., unpublished results) was used in this study. In that study, the crystal structure of HIV-1 integrase bound to the inhibitor 5-CITEP (PDB ID 1QS4) chain A was chosen, as it is the only one that binds the inhibitor. To complete the structure prior to the MD simulation, the full catalytic domain structure was modeled by extracting two residues (Y143 and N144) from chain B, and the remaining two (I141 and P142) from integrase structure 1BIS[17] via sequence and backbone alignments. The protein was soaked in a 72 Å × 72 Å × 72 Å box of water molecules. All electrostatic interactions were included through use of the particle mesh Ewald method. The protonation states of titratable residues were predicted via pK_a calculations as previously described.[18] A time step of 2 fs was used, and the SHAKE algorithm was employed to constrain all bonds to hydrogen atoms to their equilibrium values. A total of 1000 protein conformations saved during the MD study were mined with the PRCSM method, a probability-based representative conformational sampling method (Lee, K. W.; Briggs, J. M., unpublished results).

First, the desired number of clusters was selected based on a structural similarity measurement by backbone root mean-square deviation (RMSD) comparisons of all snapshots. For example, the 1st snapshot was compared to the remaining 999 snapshots saved from the MD simulation saved at 1 ps intervals. Those with a RMSD less than a selected tolerance value, e.g. 1.5 Å, were assigned to the family identified with this 1st snapshot. Similarly, the procedure was performed for the remaining 999 snapshots until finally 1000 families were formed. Second, for the largest families, the relative probability of each conformation was estimated via a Boltzmann factor ($e^{-\Delta U/RT}$), where the energy difference was compared with the minimum within the trajectory. Third, the cluster that has the largest total sum of the Boltzmann factors of the members would be selected as the 1st cluster. Hence, the 1st cluster will represent the most probable structure during the simulation time. Finally, the structure most close to the average structure in the cluster was chosen as the representative of the cluster members.

Next, to generate the 2nd cluster and representative conformer, the same procedure as above was carried out on the remaining snapshot structures. However, the members of the 1st cluster are first removed from the original set of structures. This routine is repeated until the desired number of conformers is obtained. We chose a tolerance of 1.5 Å as the RMSD cutoff which resulted in the top 10 clusters containing about 50% of the 1000 conformers.

Pharmacophore Model Development. We used the Pocket module of LigBuilder, a multipurpose structure-based drug design program,[19] to generate a receptor-based pharmacophore model according to the interactive features identified by the program, such as hydrogen-bond donor, hydrogen-bond acceptor, hydrophobic features, etc. The input for Pocket is the 3D structure of the protein (the complete chain A of 1QS4, containing one Mg^{2+} ion) in PDB format and a ligand (5-CITEP) in Mol2 format in complex with the protein, which is used by LigBuilder to specify the active site. The interaction mapping was performed for each of the 10 representative protein snapshots in addition to the crystal complex. The ligand (5-CITEP) position in each representative snapshot was determined by superimposing protein backbone with the crystal structure of the complex. The modeled snapshot complex was minimized via 100 cycles of steepest descent followed by 1000 steps of conjugate gradient using NWChem (v4.0)[20] software package with a 5 Å layer of water molecules. LigBuilder was first used to generate a shell covering the ligand and all of the surrounding protein residues (i.e. those within 5 Å of each atom in the ligand) and then subdivided it into regularly spaced (0.5 Å) grid points. Water molecules were removed during the feature generation procedure by LigBuilder. The static pharmacophore was derived directly from each selected snapshot and the crystal complex; accordingly eleven static pharmacophore models were derived. The dynamic model was generated by superimposing all of the static models, and the features of this model were defined by the well clustered elements from the superimposition. For example, a hydrogen-bond donor feature in the dynamic model was determined by a cluster containing hydrogen-bond donor elements from at least three static models. Therefore, the location was the average coordinates of all the elements in the cluster and the size is determined by the RMSD.

Database Screening. Catalyst 4.6[21] was used to generate search queries according to the pharmacophore model derived from LigBuilder program and to search our in-house database of available chemical compounds via the "Best Compare/Fit" search routine.

Experimental Methods

Materials, Chemicals, and Enzymes. All compounds were dissolved in DMSO, and the stock solutions were stored at −20 °C. The γ [^{32}P]-ATP was purchased from either Amersham Biosciences or ICN. The expression systems for the wild-type IN and soluble mutant IN^{F185KC280S} were generous gifts of Dr. Robert Craigie, Laboratory of Molecular Biology, NIDDK, NIH, Bethesda, MD.

Preparation of Oligonucleotide Substrates. The oligonucleotides 21top, 5′-GTGTGGAAAATCTCTAGCAGT-3′, and 21bot, 5′-ACTGCTAGAGATTTTCCACAC-3′, were purchased from Norris Cancer Center Microsequencing Core Facility (University of Southern California) and purified by UV shadowing on polyacrylamide gel. To analyze the extent of 3′-processing and strand transfer using 5′-end-labeled substrates, 21top was 5′-end labeled using T_4 polynucleotide kinase (Epicenter, Madison, WI) and γ [^{32}P]-ATP (Amersham Biosciences, Piscataway, NJ or ICN). The kinase was heat-inactivated and 21bot was added in 1.5-molar excess. The mixture was heated at 95 °C, allowed to cool slowly to room

temperature, and run through a spin 25 minicolumn (USA Scientific, Ocala, FL) to separate annealed double-stranded oligonucleotide from unincorporated material.

IN Assays. To determine the extent of 3′-processing and strand transfer, wild-type IN was preincubated at a final concentration of 200 nM with the inhibitor in reaction buffer (50 mM NaCl, 1 mM HEPES, pH 7.5, 50 μM EDTA, 50 μM dithiothreitol, 10% glycerol (w/v), 7.5 mM $MnCl_2$, 0.1 mg/mL bovine serum albumin, 10 mM 2-mercaptoethanol, 10% dimethyl sulfoxide, and 25 mM MOPS, pH 7.2) at 30 °C for 30 min. Then, 20 nM of the 5′-end ^{32}P-labeled linear oligonucleotide substrate was added, and incubation was continued for an additional 1 h. Reactions were quenched by the addition of an equal volume (16 μL) of loading dye (98% deionized formamide, 10 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue). An aliquot (5 μL) was electrophoresed on a denaturing 20% polyacrylamide gel (0.09 M Tris-borate pH 8.3, 2 mM EDTA, 20% acrylamide, 8 M urea).

Gels were dried, exposed in a PhosphorImager cassette, and analyzed using a Typhoon 8610 Variable Mode Imager (Amersham Biosciences) and quantitated using ImageQuant 5.2. Percent inhibition (%I) was calculated using the following equation:

$$\%I = 100 \times [1 - (D - C)/(N - C)]$$

where C, N, and D are the fractions of 21-mer substrate converted to 19-mer (3′- processing product) or strand transfer products for DNA alone, DNA plus IN, and IN plus drug, respectively. The IC_{50} values were determined by plotting the logarithm of drug concentration versus percent inhibition to obtain the concentration that produced 50% inhibition.

Results and Discussion

Representative Snapshots Ensemble. Selecting representative structures from MD snapshots was addressed over 10 years ago by considering the frames with low potential energies and clustering the snapshots RMSD comparisons.[22] The PRCSM sampling method which was first used in this study employs a rational and statistical method for identifying representative conformations from a large number of configurations generated from MD or Monte Carlo (MC) simulations (Lee, K. W.; Briggs J. M., unpublished results). Because it considers both the similarity of all available snapshots by RMSD and the energetic probability evaluated by a Boltzmann factor, it identifies diverse conformational representatives which reflect the most probable conformations from the ensemble of available snapshots − the most populated cluster will not necessarily be the highest ranked cluster. For example, if cluster 1 has one member less than cluster 2, but the sum of Boltzmann factors of cluster 1 is larger than that for cluster 2, the former is the top ranked cluster. Ten snapshots representing the top 10 clusters (Table 1) cover 554 of the 1000 available frames, suggesting that the resulting dynamic pharmacophore model should be more statistically meaningful than that from randomly selected frames.

Static Pharmacophore Models. Each static pharmacophore model was developed directly from the application of LigBuilder to the individual representative snapshot, or the crystal structure. Accordingly, a total of 11 separate static pharmacophore models were generated. Although all water molecules were removed during sampling the active site, the important water−IN interactions were characterized by hydrogen-bond donor or hydrogen-bond acceptor features. Figure 1 illustrates the feature definition by the program and shows the static model based on the crystal structure complex. Since the program needs the ligand as input to define the active site, the derived complementary features map the ligand's functional groups, as depicted in Figure 1a. In the pharmacophore model, each feature

Table 1. Ensemble of the Representative Conformations Selected by the Probability-Based Representative Conformer Sampling Method (PRCSM)[a]

cluster[b]	cluster population	sum of Boltzmann factors ($\sum e^{-\Delta U/kT}$)[c]	MD frame #[d]
1	78	39.4	RC119
2	79	36.7	RC44
3	59	29.6	RC737
4	55	28.9	RC439
5	51	26.2	RC685
6	52	25.8	RC844
7	42	22.2	RC811
8	43	22.0	RC316
9	42	20.4	RC548
10	41	20.3	RC226

[a] The total number of frames represented in these 10 clusters: 79+78+59+.....+42+41 = 542 out of 1000 frames. [b] The rank is according to the sum of the Boltzmann factors. [c] $\Delta U = E_i - E_{min}$, E_i is the potential of the ith frame in the cluster, E_{min} is the minimum potential in all 1000 snapshots. [d] The frame number represents the snapshot taken at certain time, which has the closest RMSD with the average structure within the cluster, the original corresponding snapshot in the collection of 1ns trajectory; for example, RC119 refers to the representative conformation (RC) snapshot taken at 119 ps.

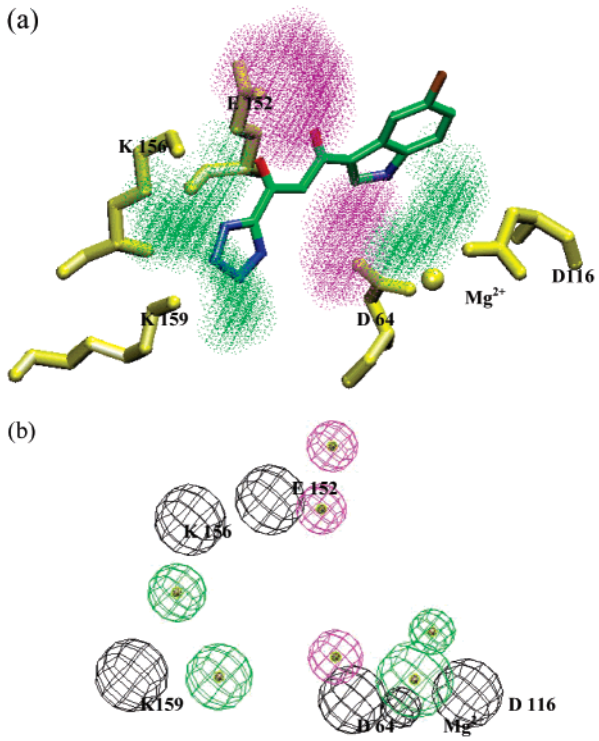


Figure 1. (a) Interactions between the ligand and IN derived from the crystal structure complex by LigBuilder. Green/blue/red-colored sticks represent 5-CITEP (hydrogens not displayed), and yellow sticks are key residues of the protein (D64, D116, E152, K156, and K159). The metal ion, Mg^{2+} is displayed as a yellow sphere. Condensed dots are colored by magenta or green representing the locations of hydrogen-bond donors and hydrogen-bond acceptors, respectively. (b) The static pharmacophore model from the crystal complex. The black mesh spheres represent the excluded volumes defined by the key residues listed above, green for hydrogen-bond acceptors, and magenta for hydrogen-bond donors.

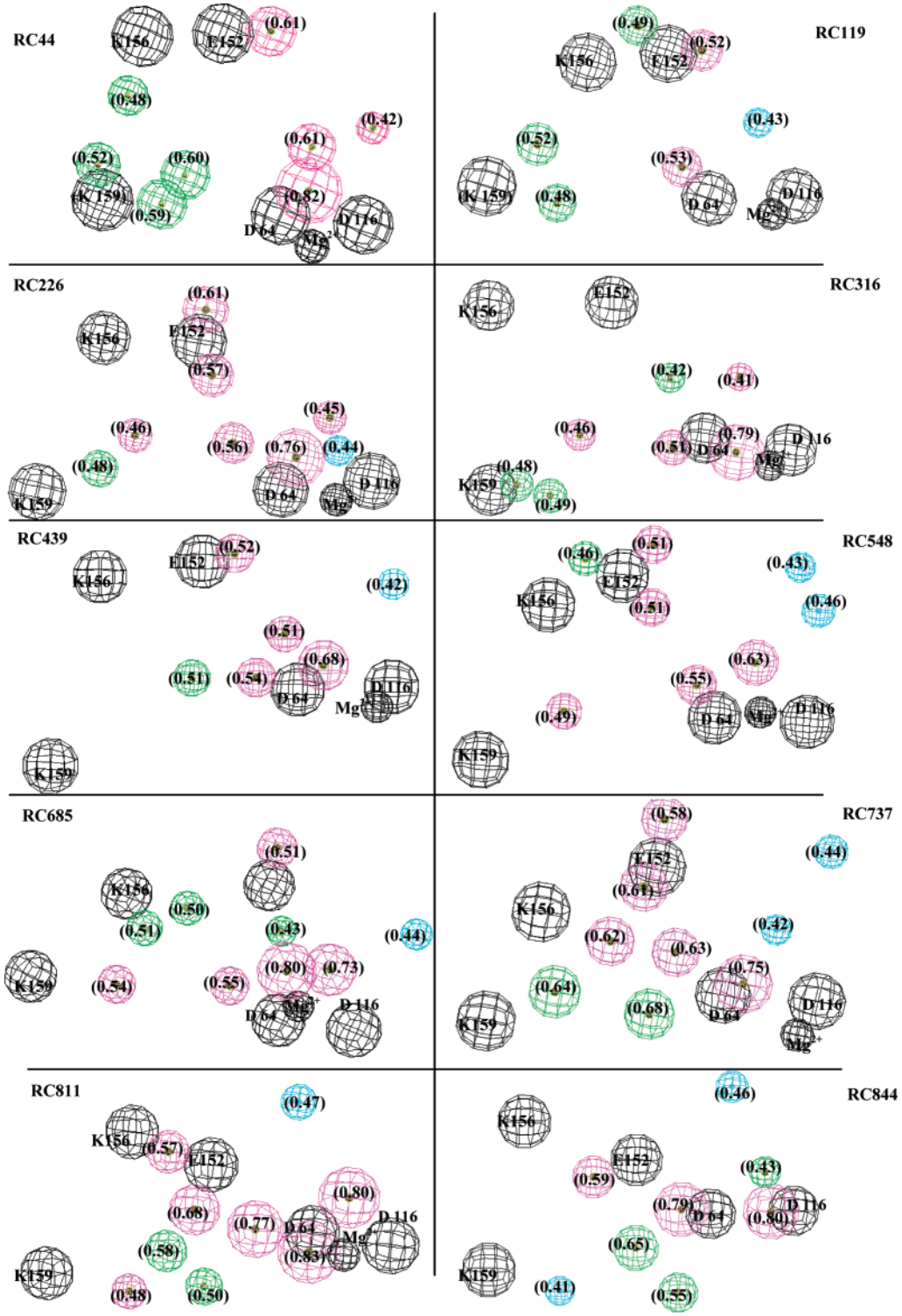


Figure 2. Characteristics of the static pharmacophore model from each single snapshot. The black mesh spheres represent the excluded volumes defined by the key residues listed above, green for HBA, magenta for HBD, and cyan for hydrophobic feature. The weight of each element is labeled accordingly.

site was identified by the interactions between the inhibitor and the protein, containing features such as hydrogen-bond donor, hydrogen-bond acceptor, and/or hydrophobic sites, plus excluded volumes (Figures 1b and 2). We define the key residues to be D64 (C_γ), D116 (C_γ), E152 (C_δ), K156 (C_ϵ), K159 (C_ϵ), and one Mg^{2+} as excluded volumes to avoid identifying small molecules that may overlap with these residues of the protein. A radius of 1.5 Å was chosen for the excluded volumes,[13] except for Mg^{2+}, for which we used a radius of 0.86 Å.[23] Coordinates of functional features are determined by the position of probe atoms identified by the Pocket program; the size of each pharmacophore feature is directly related to the area occupied by the functional probes. The original values of the feature site were in the range of 0.41−0.83 Å. These values were too small to be directly used as radii of the pharmacophore elements. Therefore, the radius of each element was increased by a factor of 1.5, 2, or 3 times that of the RMSDs. Since feature overlapping occurs in some of the models with

Table 2. Characteristics of the Refined Static Queries and Database Screening

query ID	frame	#[a] HBA	#[a] HBD	#[a] HPhob	# hits	sites removed from original static models (as shown in Figure 2)
#1	RC44	3	3	0	1	HBD site with weight 0.42; HBA site with weight 0.48
#2	RC119	3	2	0	5	hydrophobic site with weight 0.43
#3	RC226	1	4	0	1	HBD sites with weight 0.45, 0.46; the only hydrophobic site with weight 0.44
#4	RC316	2	3	0	52	HBA site with weight 0.42, HBD site with weight 0.41
#5	RC439	1	4	0	13	hydrophobic site with weight 0.42
#6	RC737	2	4	0	4	both hydrophobic sites; HBD site with weight 0.58
#7	RC811	1	5	0	1	HBD site with weight 0.48, HBA site with weight 0.50, hydrophobic site with weight 0.47

[a] HBA: hydrogen-bond acceptor, HBD: hydrogen-bond donor, HPhob: hydrophobic site.

radii three times the RMSD, and while no hits were identified from models with radii of 1.5 times the site RMSD, we finally selected models with radii that were twice the size of the RMSD. Figure 2 displays the static models from the selected MD frames, respectively.

Each model still contained too many features, making them way too specific, so they needed to be further refined. We have made two assumptions herein: (1) models with a greater number of features should be more accurate in representing structural and chemical features of the target active site, and (2) if a compound could map the refined query, it is regarded to map the original model. Taking these two together, the principle of the refinement was to retain as many functional features sites as possible while keeping all of the excluded volume areas. The procedure involved keeping at least five functional feature sites plus the excluded volume sites. The five point models were generated by removing feature elements one at a time (and then in multiples) from the static model. The queries were then used to search the database in order to find the ones that performed the best. The elements with the lowest weights (as shown in Figure 2), as indicated by the value of their radii, were removed first when generating the five point models. The refined models were used as searching queries to screen the available compound database. Some queries were not able to identify any molecules in the database. In Table 2 are summarized the features and total number of hits which were returned by the queries corresponding to the refined static models. Obviously, none of the refined static queries contain a hydrophobic feature, as shown in the table, because of its relatively low weight. We applied Lipinski's rule of five[24] (i.e. molecular weights less than 500 g/mol, an absolute partition coefficient between octanol/water less than 5, and the number of hydrogen-bond donors and acceptors less than 5 and 10, respectively) as a filter to the hits in order to remove those compounds in conflict with any two of the rules. It is clear that, even though some refined models could return hits, the number of known inhibitors that are matched is rather poor, indicating that considering only a single structure of the IN in model development is interior to using more than one.

Dynamic Pharmacophore Model. Our dynamic pharmacophore model is a conceptually improved hypothesis compared to the previously reported model[13] since we considered the entire active site of IN in addition to superimposing 11 static pharmacophore models which were developed from selected representative protein conformations identified using our PRCSM method. The structural diversity was measured in terms of the RMSD with respect to the reference frame of the most probable snapshot taken at 119 ps (RC119 in Table 1), and the superimposition performed according to five key residues in the active site, i.e., D64, D116, E152, K156, and K159. The RMSD values of all of the 10 selected snapshots ranged from 0 to 1.149 Å, indicating the well-converged behavior of IN during the 1 ns simulation. Interestingly, recently reported docking simulations,[25] regarding predicting the ligand binding mode in several IN snapshots collected from 2 ns MD studies, identified a previously uncharacterized open trench, suggesting the necessity to explore the conformational space of the IN through longer time-scale MD simulations. Therefore, the dynamic pharmacophore model accounts for the inherent flexibility of the active site by considering the average structural and chemical features of the ensemble of conformations. The elements of the dynamic model were identified as the conserved/consensus feature sites, containing at least three elements[13] within the conserved area of the superimposed static pharmacophore models. Accordingly, the position of each pharmacophore element was determined by the average coordinates of all feature sites within the conserved volume, and the radius was set to the average RMSD of the individual features. The Mg^{2+} ion was treated as part of the protein and therefore was represented as one of the excluded volumes in the pharmacophore model development, and the location was the average position of Mg^{2+} in all of the selected snapshots. The characteristics of the dynamic pharmacophore model are listed in Table 3. In Figure 3a is shown the process used to obtain the dynamic model by overlaying all of the static pharmacophore models with the encircled areas reflecting the positions of the dynamic elements. In Figure 3b, the resulting dynamic pharmacophore model is displayed, indicating all possible interaction features within the entire active site of IN. Three of the hydrogen-bond donor features, i.e., HBD3, HBD4, and HBD5, are near the key residues D64 and D116 and are also near to the features identified in a previous study.[13] Additional

Table 3. Characteristic of the Dynamic Pharmacophore Model Derived from the Overlay of 11 Static Pharmacophore Models

feature site[a]	*X* (Å)	*Y* (Å)	*Z* (Å)	*R*
HBA1	−8.99	−0.68	−5.55	1.87
HBA2	−12.85	0.42	−4.10	1.85
HBD1	−9.83	4.11	−5.84	1.91
HBD2	−8.21	8.14	−5.72	0.97
HBD3	−2.16	3.82	−8.95	1.04
HBD4	−3.08	1.42	−6.85	0.61
HBD5	−5.89	1.64	−8.03	1.07
HPhob	−1.75	3.27	−13.78	0.62
D64	−3.96	0.05	−9.84	1.50
D116	0.69	0.81	−9.10	1.50
E152	−9.33	6.53	−8.54	1.50
K156	−14.62	5.24	−7.90	1.50
K159	−15.99	−2.76	−4.76	1.50
Mg^{2+}	−0.91	0.21	−11.11	0.86

[a] HBA, hydrogen-bond acceptor, HBD, hydrogen-bond donor, HPhob, hydrophobic element.

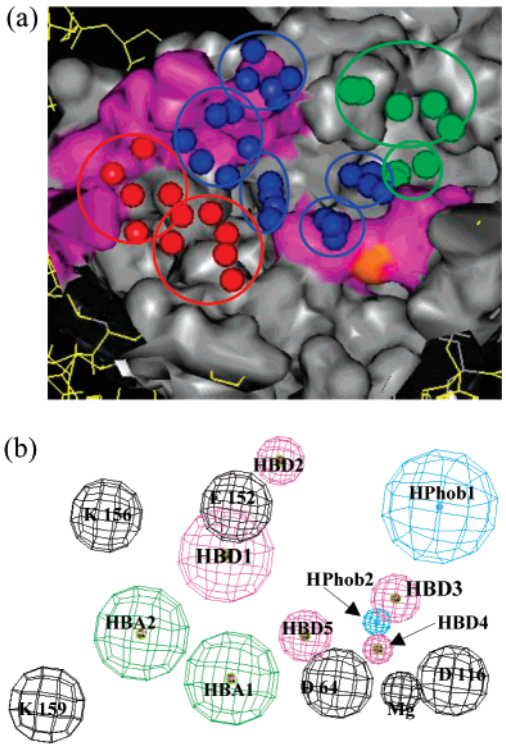


Figure 3. Schematic illustration of the method for generating the dynamic pharmacophore model. (a) Superimposing 11 static pharmacophore models to identify the conserved or consensus features. Molecular surface displays of frame RC119 as a reference. The magenta areas represent residues D64, D116, E152, K156, and K159, and an orange surface for Mg^{2+}. The green spheres are for hydrophobic sites, blue sphere for hydrogen-bond donor, and red sphere for hydrogen-bond acceptor. (b) The dynamic pharmacophore model according to the consensus features. The black mesh spheres are for excluded volumes defined by the key residues listed above, green for HBA, magenta for HBD, and cyan represents the corresponding hydrophobic feature.

sites were found in this study, namely, two hydrogen-bond donors, labeled as HBD1 and HBD2 complementary to E152, and two hydrogen-bond acceptors, i.e. HBA1 and HBA2 complementary to K156 and K159. Additionally, two hydrophobic features were located in this comprehensive dynamic model.

As with the static models, the complete dynamic pharmacophore with nine functional feature sites (five hydrogen-bond donor, two hydrogen-bond acceptor, and two hydrophobic sites) plus six excluded volume sites is way too specific to be a searching query. Hence, the model was refined in a similar way as that for the static models. However, each refined query must contain at least one hydrogen-bond donor near E152, one donor near D64 or D116, one acceptor, and at least one hydrophobic feature. Multiple independent queries containing at least five feature sites plus six excluded volumes were produced.

Consistent with a prior study,[13] we also find that the dynamic model generated from multiple protein conformations performed much better than any of the static models. Hence, we collected 128 inhibitors with diverse skeletons and published recently[10,11,26−45] in order to evaluate the performance of the dynamic pharmacophore models. The details of known inhibitors mapping the dynamic model are available in the Supporting Information. In Table 4 are reported the statistical analyses of the performance of the model. Most of the compounds contain data for both the 3′-processing and strand transfer reactions. The molecules were sorted according to their values of IC_{50} against the 3′-processing reaction only. Among the 128 compounds, eight molecules with various templates (compounds **121**−**128**, Supporting Information) were reported inactive because of high IC_{50} values for both 3′-processing and strand transfer and therefore were collected to test the model for false positives. Only one (compound **121**, Supporting Information) of the eight was falsely recognized by the model probably due to the hydroxyl and ketone groups positioned properly to map the model. None of the seven remaining molecules have hydrogen-bond donors, and one of them also does not contain any hydrogen-bond acceptors. A small number of the active compounds failed to be recognized by the model, even though some of the compounds contain one or two ketone groups that could act as hydrogen-bond acceptors; their relatively rigid conformations, however, prevent the molecules from mapping with the model. Some molecules lack hydrogen-bond donors but have one or two acceptors; for example, thiazolothiazepine inhibitors[35] should not fit the model and actually were not recognized by the model. Seventy-seven active inhibitors among the 128 compounds (compounds **1**−**83** except three molecules for which strand transfer IC_{50} values are not available and another three with IC_{50}'s against strand transfer activity greater than 20 μM, as listed in Supporting Information) have both IC_{50}s for 3′-processing and strand transfer below 20 μM, and over 72% could be predicted from at least one of the refined dynamic queries. Moreover, if we consider only inhibition of the 3′-processing reaction, among 83 active compounds (**1**−**83** in the list) with IC_{50} values below 20 μM, ca. 75% could be easily identified by the model. In terms of inhibition of the strand transfer reaction, among 108 active inhibitors collected in the subset with IC_{50} values lower than 20 μM, nearly 65% could be identified by the dynamic model. Furthermore, we investigated a larger subset containing molecules with IC_{50} values lower than 100 μM for 3′-processing and/or strand transfer reactions. One set of 3′-processing inhibitors contained 108 molecules, 70% of which were identified by the model, while 66% of another set containing 116 strand transfer inhibitors matched the dynamic queries.

The refined queries were used to screen our in-house available compounds database, and nearly 400 unique compounds meet the requirement of Lipinski's rule of

Table 4. Summary of the Dynamic Model Performance against Known IN Inhibitors

	IC_{50} < 20 μM[a]			IC_{50} < 100 μM[a]		IC_{50} > 100 μM[a]
	both 3′ and ST	3′	ST[1]	3′	ST	both 3′ and ST
subset sizes[b]	77	83	108	108	116	8
hits[c]	56	62	70	76	76	1
%[d]	72.7	74.7	64.8	70.4	65.5	12.5

[a] 3′:3′-processing, ST: strand transfer reaction. [b] Total number of known inhibitors collected in the subset. [c] Total number of inhibitors that could be recognized by the model. [d] Percentage of molecules mapping the dynamic queries. For example, in the subset of 83 compounds with IC_{50}'s lower than 20 μM for the 3′-processing reaction, 62 among them could be identified by the model; therefore, the % = (62/83) × 100 = 74.7%.

Table 5. Inhibition of HIV-1 Integrase Wild-type and Soluble Mutant by Selected Hits

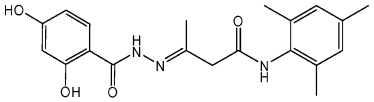
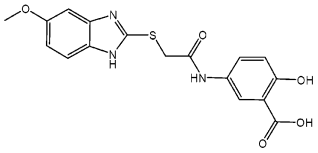
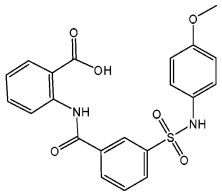
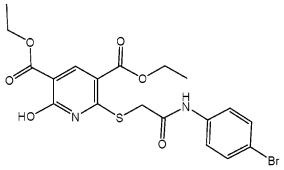
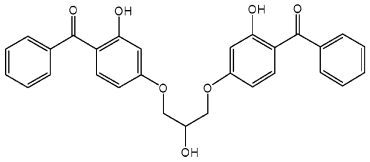
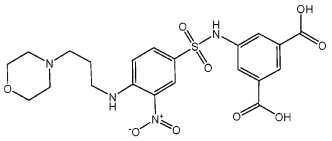
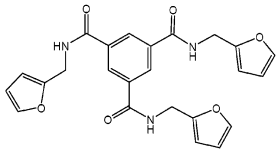
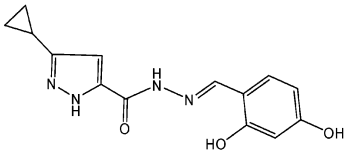
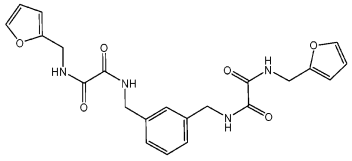
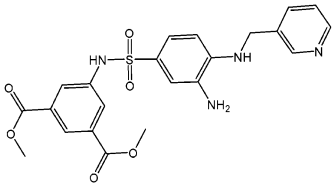
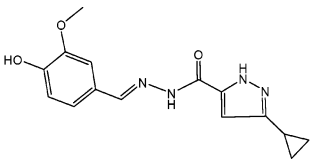
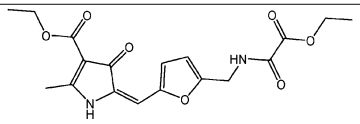
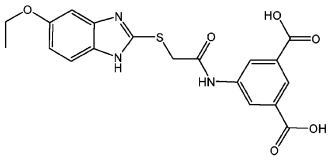
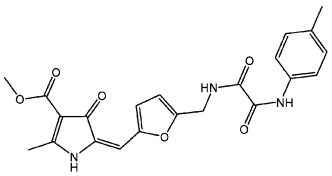
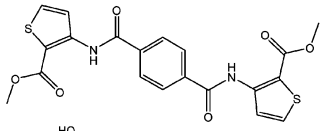
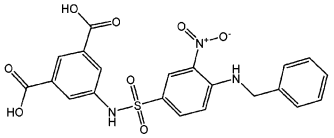
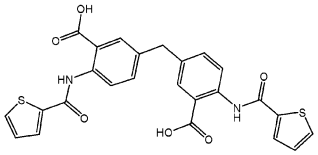
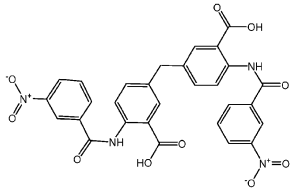
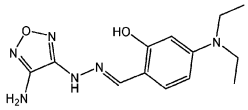
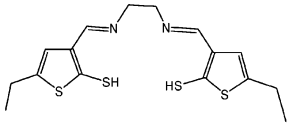
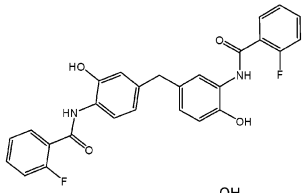
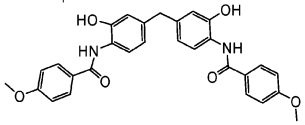
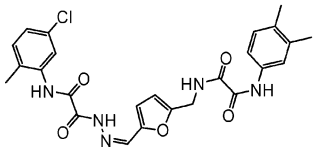
Cmpd	2-D structure	M.W.	Experimental data (μM) 3'-proc.	Strand Transfer
1		369	> 1000	150
2		373	> 100	> 100
3		426	355	300
4		483	> 100	> 100
5		484	> 100	> 100
6		508	140	100
7		447	37 ± 3	8 ± 1
8		286	177 ± 12	43 ± 7
9		438	> 100 (1000)[a]	> 100 (40)[a]
10		470	>100	>100
11		300	>100	>100

Table 5 (Continued)

Cmpd	2-D structure	M.W.	Experimental data (μM) 3'-proc.	Strand Transfer
12		376	>>100	>>100
13		415	>>100	>>100
14		423	>>100	81 ± 17
15		444	>>100	97 ± 5
16		471	>>100	66 ± 29
17		506	63 ± 11	38 ± 5
18		584	15 ± 4	13 ± 4
19		290	>> 100	>> 100
20		368	>100 (> 100)[a]	21 ± 3 (6)[a]
21		474	>100 (205)[a]	> 100 (33)[a]
22		498	> 100	> 100
23		509	>100 (510)[a]	> 100 (70)[a]

[a] Soluble mutant (F185K, C280S).

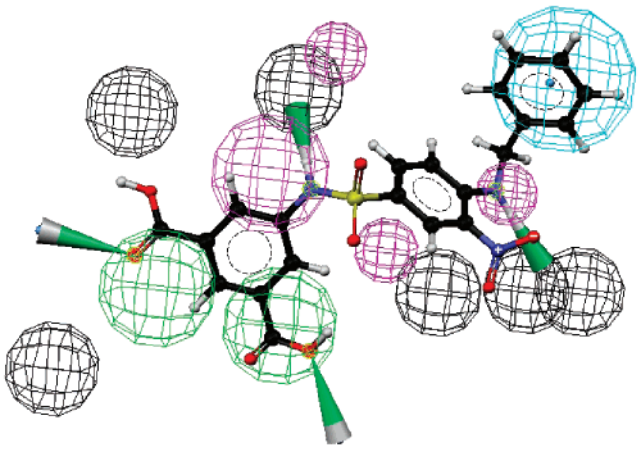


Figure 4. Mapping of compound **16** onto the one of the refined dynamic pharmacophore queries.

five.[23] Not surprisingly, many hits were recognized by multiple queries. Some of the compounds were analogues of published inhibitors, for example, sulfonamide containing molecules. Moreover, most of the compounds identified by a static model were able to be recognized by the dynamic model queries. All of this appears to suggest that the dynamic pharmacophore model derived from multiple static pharmacophore models is a reasonable hypothesis, taking into consideration the structural dynamics of the protein, and therefore is more reliable than any of the static models in identifying hits through screening databases of small molecules.

IN Catalytic Activity Assays for the Selected Novel Hits. In this work, combining our experiences with small molecule IN inhibitors, we focused only on the novel scaffolds identified by the dynamic queries. As a result, among the 394 hits, we selected 23 molecules with novel structures to be assayed (Table 5). Three of them (**7**, **17**, **18**) have IC_{50}s less than 100 μM against the 3′-processing reaction. Compounds **6**, **7**, **8**, **14**, **15**, **16**, **17**, **18**, **20** (Table 5) effectively blocked the strand transfer reactions at IC_{50}s less than 100 μM. In particular, **7** was the most potent with an IC_{50} of 8 μM for strand transfer and 37 μM for 3′-processing. Compound **18** inhibited both the strand transfer and 3′-processing reactions with IC_{50} values less than 15 μM. Interestingly, compounds **7**, **8**, and **20** showed a 4−5-fold selectivity for the strand transfer reaction. Additionally, compound **1** showed 6-fold selectivity for the strand transfer reaction. Furthermore, some compounds were also tested for inhibition of catalytic activities of the soluble mutant of IN (i.e. F185K, C280S) where compound **20** was found to be nearly 20-fold selective against strand transfer as compared with 3′-processing. Compound **16** was identified by multiple refined queries which were systematically derived from the dynamic pharmacophore model. Figure 4 shows a mapping of **16**, onto one of the 7-feature containing dynamic query. This mapping orientation indicates that a compound matching (1) the hydrogen-bond acceptors near K156, and K159, (2) hydrogen-bond donors near the conserved D,D-(35)E catalytic motif, and (3) hydrophobic feature such as an aromatic ring (Figure 4) could bind to the protein in such a way that inhibits its enzymatic activity.

Conclusions

A dynamic pharmacophore model was developed by the superimposition of static pharmacophore models generated from an ensemble of HIV-1 integrase catalytic domain conformations obtained from a molecular dynamics simulation. The receptor-based pharmacophore models contain a variety of pharmacophore features (e.g. hydrogen-bond acceptor and donor, hydrophobic, and excluded volumes). Database searches against the dynamic model yielded a series of structurally novel compounds which were assayed and found to be active inhibitors, suggesting the necessity to apply the dynamic model to search larger databases to identify potential leads. In addition, the sampling and the pharmacophore model development methods used in this study have been effectively combined and could be extended to other systems for structure-based drug discovery.

Acknowledgment. J.M.B. gratefully acknowledges the National Institutes of Health (GM56553) and the Robert A. Welch Foundation (E-1497) for support of this work. This research was also supported in part by a grant of supercomputer time to J.M.B. by the National Science Foundation cooperative agreement ACI-9619020 through computing resources provided by the National Partnership for Advanced Computational Infrastructure at the Texas Advanced Computing Center, University of Texas, Austin. J.M.B. thanks Accelrys, Inc. (http://www.accelrys.com) for making the InsightII software and associated modules available to us through the Institute for Molecular Design at the University of Houston. Work in N.N.'s laboratory was supported by funds from GlaxoSmithKline Drug Discovery and Development Award. J.D. would like to thank Prof. H. A. Carlson for helpful advice, and Prof. B. M. Pettitt for fruitful discussions.

Supporting Information Available: Structures of published inhibitors used in this study to validate our dynamic pharmacophore model are available free of charge via the Internet at http://pubs.acs.org.

References

(1) Coffin, I. M.; Hughes, S. H.; Varmus, H. E. *Retroviruses*; Cold Spring Harbor Laboratory Press: New York, 1997.
(2) Chen, I. J.; Neamati, N.; MacKerrell, A. D., Jr. Structure-based inhibitor design targeting HIV-1 integrase. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 217−234.
(3) Neamati, N.; Sunder, S.; Pommier, Y. Design and Discovery of HIV-1 Integrase Inhibitors. *DDT* **1997**, *2*, 487−198.
(4) Pommier, Y.; Marchand, C.; Neamati, N. Retroviral integrase inhibitors year 2000: update and perspectives. *Antiviral Res.* **2000**, *47*, 139−148.
(5) Neamati, N. Structure-based HIV-1 integrase inhibitor design: a future perspective. *Expert Opin. Investig. Drugs* **2001**, *10*, 281−296.
(6) Fesen, M. R.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J. et al. Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. *Biochem. Pharmacol.* **1994**, *48*, 595−608.
(7) Marchand, C.; Neamati, N.; Pommier, Y. In vitro human immunodeficiency virus type 1 integrase assays. *Methods Enzymol.* **2001**, *340*, 624−633.
(8) Dayam, R.; Neamati, N. Small-molecule HIV-1 integrase inhibitors: the 2001−2002 update. *Curr. Pharm. Des.* **2003**, *9*, 1789−1802.
(9) Singh, S. B.; Ondeyka, J. G.; Schleif, W. A.; Felock, P.; Hazuda, D. J. Chemistry and structure−activity relationship of HIV-1 integrase inhibitor integracide B and related natural products. *J. Nat. Prod.* **2003**, *66*, 1338−1344.
(10) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A. et al. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem.* **1997**, *40*, 930−936.

(11) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazumder, A.; Sunder, S. et al. HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920−929.
(12) Hong, H.; Neamati, N.; Winslow, H. E.; Christensen, J. L.; Orr, A. et al. Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. *Antiviral Chem. Chemother.* **1998**, *9*, 461−472.
(13) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L. et al. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100−2114.
(14) Young, S. L-780,810: a potent antiviral HIV integrase inhibitor with potential clinical utulity. Presented at The XIV International AIDS Conference, Barcelona, 2002.
(15) Yoshinaga, T.; Sato, A.; Fujishita, T.; Fujiwara, T. Presented at 9th Conference on Retroviruses and Opportunistic Infections: Seattle, WA, 2002.
(16) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T. et al. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13040−13043.
(17) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R. et al. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 9150−9154.
(18) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J. et al. Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys. J.* **1999**, *76*, 2999−3011.
(19) Wang, R.; Gao, Y.; Lai, L. LigBuilder: A multi-purpose program for structure-based drug design. *J. Mol. Modelling* **2000**, *6*, 498−516.
(20) NWChem *A Computational Chemistry Package for Parallel Computers, High Performance Computational Chemistry Group*; 4.0 ed.; Pacific Northwest National Laboratory: Richland, WA 99352.
(21) Catalyst; 4.6 ed.; Accelrys Inc: San Diego, CA.
(22) O'Connor, S. D.; Smith, P. E.; al-Obeidi, F.; Pettitt, B. M. Quenched molecular dynamics simulations of tuftsin and proposed cyclic analogues. *J. Med. Chem.* **1992**, *35*, 2870−2881.
(23) Shannon, R. D. Revised Effective Ionic Radii in Halides and Chalcogenides. *Acta Crystallogr.* **1976**, *A32*, 751.
(24) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(25) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H. et al. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879−1881.
(26) Lin, Z.; Neamati, N.; Zhao, H.; Kiryu, Y.; Turpin, J. A. et al. Chicoric acid analogues as HIV-1 integrase inhibitors. *J. Med. Chem.* **1999**, *42*, 1401−1414.
(27) Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S. et al. Hydrazide-containing inhibitors of HIV-1 integrase. *J. Med. Chem.* **1997**, *40*, 937−941.
(28) Neamati, N.; Mazumder, A.; Zhao, H.; Sunder, S.; Burke, T. R., Jr. et al. Diaryl sulfones, a novel class of human immunodeficiency virus type 1 integrase inhibitors. *Antimicrob. Agents Chemother.* **1997**, *41*, 385−393.
(29) Mazumder, A.; Neamati, N.; Sunder, S.; Schulz, J.; Pertz, H. et al. Curcumin analogues with altered potencies against HIV-1 integrase as probes for biochemical mechanisms of drug action. *J. Med. Chem.* **1997**, *40*, 3057−3063.
(30) Neamati, N.; Mazumder, A.; Sunder, S.; Owen, J. M.; Schultz, R. J. et al. 2-Mercaptobenzenesulphonamides as novel inhibitors of human immunodeficiency virus type 1 integrase and replication. *Antiviral Chem. Chemother.* **1997**, *8*, 485−495.
(31) Neamati, N.; Hong, H.; Mazumder, A.; Wang, S.; Sunder, S. et al. Depsides and depsidones as inhibitors of HIV-1 integrase: discovery of novel inhibitors through 3D database searching. *J. Med. Chem.* **1997**, *40*, 942−951.
(32) Neamati, N.; Hong, H.; Sunder, S.; Milne, G. W.; Pommier, Y. Potent inhibitors of human immunodeficiency virus type 1 integrase: identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors. *Mol. Pharmacol.* **1997**, *52*, 1041−1055.
(33) Neamati, N.; Hong, H.; Owen, J. M.; Sunder, S.; Winslow, H. E. et al. Salicylhydrazine-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site. *J. Med. Chem.* **1998**, *41*, 3202−3209.
(34) Zhao, H.; Neamati, N.; Hong, H.; Mazumder, A.; Wang, S. et al. Coumarin-based inhibitors of HIV integrase. *J. Med. Chem.* **1997**, *40*, 242−249.
(35) Marchand, C.; Zhang, X.; Pais, G. C.; Cowansage, K.; Neamati, N. et al. Structural determinants for HIV-1 integrase inhibition by beta-diketo acids. *J. Biol. Chem.* **2002**, *277*, 12596−12603.
(36) Neamati, N.; Turpin, J. A.; Winslow, H. E.; Christensen, J. L.; Williamson, K. et al. Thiazolothiazepine inhibitors of HIV-1 integrase. *J. Med. Chem.* **1999**, *42*, 3334−3341.
(37) Pais, G. C.; Zhang, X.; Marchand, C.; Neamati, N.; Cowansage, K. et al. Structure activity of 3-aryl-1,3-diketo-containing compounds as HIV-1 integrase inhibitors. *J. Med. Chem.* **2002**, *45*, 3184−3194.
(38) Zhang, X.; Pais, G. C.; Svarovskaia, E. S.; Marchand, C.; Johnson, A. A. et al. Azido-Containing aryl beta-Diketo acid HIV-1 integrase inhibitors. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1215−1219.
(39) Singh, S. B.; Jayasuriya, H.; Salituro, G. M.; Zink, D. L.; Shafiee, A. et al. The complestatins as HIV-1 integrase inhibitors. Efficient isolation, structure elucidation, and inhibitory activities of isocomplestatin, chloropeptin I, new complestatins, A and B, and acid-hydrolysis products of chloropeptin I. *J. Nat. Prod.* **2001**, *64*, 874−882.
(40) Zhang, X.; Neamati, N.; Lee, Y. K.; Orr, A.; Brown, R. D. et al. Arylisothiocyanate-containing esters of caffeic acid designed as affinity ligands for HIV-1 integrase. *Bioorg. Med. Chem.* **2001**, *9*, 1649−1657.
(41) Ridley, C. P.; Reddy, M. V.; Rocha, G.; Bushman, F. D.; Faulkner, D. J. Total synthesis and evaluation of lamellarin alpha 20-Sulfate analogues. *Bioorg. Med. Chem.* **2002**, *10*, 3285−3290.
(42) Rowley, D. C.; Hansen, M. S.; Rhodes, D.; Sotriffer, C. A.; Ni, H. et al. Thalassiolins A−C: new marine-derived inhibitors of HIV cDNA integrase. *Bioorg. Med. Chem.* **2002**, *10*, 3619−3625.
(43) Lee, J. Y.; Park, J. H.; Lee, S. J.; Park, H.; Lee, Y. S. Styrylquinazoline derivatives as HIV-1 integrase inhibitors. *Arch. Pharm. (Weinheim)* **2002**, *335*, 277−282.
(44) Singh, S. B.; Zink, D. L.; Heimbach, B.; Genilloud, O.; Teran, A. et al. Structure, stereochemistry, and biological activity of integramycin, a novel hexacyclic natural product produced by Actinoplanes sp. that inhibits HIV-1 integrase. *Org. Lett.* **2002**, *4*, 1123−1126.
(45) Hazuda, D. J.; Felock, P.; Witmer, M.; Wolfe, A.; Stillmock, K. et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **2000**, *287*, 646−650.

JM049410E