Dynamic Pharmacophore Model Optimization: Identification of Novel HIV-1 Integrase Inhibitors

Jinxia Deng,[†,‡] Tino Sanchez,[‡] Nouri Neamati,*[,‡] and James M. Briggs*[,†]

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

Received October 20, 2005

We extended the previously described dynamic pharmacophore model studies of HIV-1 integrase (IN) by considering more key residues in the active site, including Mg^{2+}. First, we applied a Monte Carlo sampling method to map the complementary features of the IN binding surface. Two types of dynamic pharmacophore models were generated. One considers Mg^{2+} as part of the IN and therefore as an excluded volume, and the other treats Mg^{2+} as a positively charged feature, representing a new type of pharmacophore model aimed to identify compounds potentially preventing Mg^{2+} binding. Second, we validated the models with 385 known active ($IC_{50} < 20\ \mu M$) and 235 ($IC_{50} > 100\ \mu M$) inactive IN inhibitors. Third, we used the derived models to screen our small molecule database. Twenty-two structurally novel compounds were tested in an in vitro assay specific for IN, and two of them showed $IC_{50} \leq 10\ \mu M$ for strand transfer reaction.

Introduction

Human immunodeficiency virus (HIV) is the causative agent of acquired immunodeficiency syndrome (AIDS), which is one of the most serious health problems in the world. HIV encodes three enzymes: protease, reverse transcriptase, and integrase (IN). Currently, treatment for HIV infection includes the combination of inhibitors of the reverse transcriptase and protease (HAART). Problems of drug toxicity and drug resistance may be reduced via the inhibition of a new HIV target. IN is 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. In addition, IN is a testable target because rapid and sensitive assays exist for measuring enzymatic activity, and crystal and NMR structures are available for use in rational structure-based drug design.[1] However, IN is a difficult system for structure-based drug discovery for a number of reasons. It contains a shallow substrate binding site positioned on the surface of the protein; it also plays a role in the formation of a multimeric complex in preintegration complexes, and finally, the lack of a full-length structure for the enzyme in the absence or presence of a DNA substrate, all complicate the rational drug design process.

Many IN inhibitors have been identified, and two have been tested in clinical trials.[2−9] 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] introduced for the first time the "dynamic" concept in IN drug design. The dynamic pharmacophore method considers multiple receptor conformations that are used to generate a receptor-based pharmacophore model. The initial model was then successfully used to identify two active inhibitors against IN ($IC_{50} \sim 25\ \mu M$) through the searching of available compound databases.[13] In that model, only hydrogen bond donor sites were identified by the methanol probes that only associated with D64 and D116. That model included neither metal ions nor residue E152 due to the limited crystallographic information for such at that time.

Our recently reported dynamic receptor-based model was derived using the LigBuilder program. The ligand orientation, as revealed in the crystal complex (1QS4), was used to define the binding site, and Mg^{2+} was treated as a charged protein atom and as an excluded volume in the pharmacophore model. Its application in database screening was successfully able to recognize some structurally novel IN inhibitors.[14] However, the limitation of this approach is the requirement of a ligand-bound complex.

Therefore, in this work, we performed pure receptor-based dynamic pharmacophore model studies to conceptually extend the previous work by (i) applying different types of functional fragments as probes to map the complementary features of the IN binding surface via a Monte Carlo sampling algorithm, (ii) building dynamic pharmacophore models by considering Mg^{2+}-as an excluded volume or a charged feature in the model, respectively, and (iii) screening databases by the derived models and selecting structurally novel molecules for an IN specific enzyme assay.

Computational Methods

The Monte Carlo software package for biomolecular systems, BOSS 4.0, was used in our study to perform a Metropolis configurational sampling of new configurations by probe molecules in association with each of the protein conformations. The initial molecular system consists of a rigid protein conformation and a sphere of hundreds of probe atoms/molecules (e.g., methyl, benzyl, methanol, ammonium, etc.) centered on the active site of the enzyme. A Metropolis Monte Carlo algorithm is then employed to randomly select and move probe groups, generating new system configurations. If the new intermolecular configuration is lower in energy than its predecessor, the new one is always accepted leading to large populations of minimum energy configurations. If the new configuration, however, is higher in energy than its predecessor, a Boltzmann factor (BF), defined in eq 1, is compared to a

* To whom correspondence should be addressed. (N.N.) Tel: 323-442-2341. Fax: 323-442-1390. E-mail: neamati@usc.edu. (J.M.B.) Tel: 713-743-8366. Fax: 713-743-8351. E-mail: jbriggs@uh.edu.
[†] University of Houston.
[‡] University of Southern California.

10.1021/jm0510629 CCC: $33.50
Published on Web 02/15/2006

random number between 0 and 1 to decide whether the new configuration will be accepted

$$\mathrm{BF} = \sum e^{-\Delta U/k_B T} \quad (1)$$

where k_B is Boltzmann's constant, U is the potential energy, and T is the temperature. If the BF is greater than the random number, the new configuration is accepted; otherwise, it is rejected. The acceptance condition can be written in the following concise fashion (eq 2) for the NPT (constant mass, pressure, and temperature) ensemble

$$\mathrm{rand}(0,1) \leq e^{-\Delta H/k_B T} \quad (2)$$

where rand(0,1) is a machine-generated random number between 0 and 1.

The size of the displacement at each iteration is governed by the maximum displacement. This is an adjustable parameter whose value is usually chosen so that approximately 50% of the trial moves are accepted.

In this work, all of the Monte Carlo simulations were performed under an NPT ensemble with the OPLS (Optimized Potentials for Liquid Simulations) all atom force field.[15,16] The BOSS program generates hundreds of nonoverlapping probes within the specified active site area. The probes only see the target protein (i.e., probe−probe interactions were set to zero). The Monte Carlo simulation is effectively a simulated annealing of the probe molecules against the protein. The probes cluster themselves within local minima in the active site region; the clustered probes define the favored binding regions in the active site for that specific functionality.

Probe. Four types of functional groups were used to sample the active surface of IN: (i) chloride (Cl^-) representing a negatively charged group, (ii) methanol (CH_3OH) representing H-bonding features (i.e., functioning as either an H-bond donor or an H-bond acceptor depending upon the neighboring amino acid residues), (iii) ammonium ion (NH_4^+) representing a positively charged group, and (iv) methane (CH_4) representing a hydrophobic feature. As mentioned, the functional group probes were effectively modeled as noninteracting solvent molecules (i.e., they do not interact with one another but do interact with the protein) in the employed Monte Carlo procedure.

Protein. A total of 10 snapshots of IN were collected at equal time intervals from prior molecular dynamics (MD) simulations,[17] which were performed on the wild-type IN (PDB code 1QS4[18] in which the ligand 5-CITEP was removed). Choosing snapshots derived from ligated complexes is an important choice since it has been demonstrated that the use of ligand-complexed protein structures greatly enhances the probability of correctly identifying complex structures by docking.[19,20] Each snapshot with no water molecules was used as a target for the minimization of probes via the described Monte Carlo procedure. The metal ion (Mg^{2+}) between D64 and D116 in the catalytic site was retained in the protein conformations. The protein structure was treated as a solute molecule and was kept entirely rigid and immobile during the Monte Carlo procedure.

Parameter Assignment. The side chain oxygen (O_δ) of residue Q62 was selected to be the center of the active site[13] and was used to define a 19 Å sphere containing hundreds of probe molecules. A boundary force constant of 5 kcal/mol was used to keep the small molecules from wandering beyond the sphere, while no probes were initially within 1.5 Å of the protein surface, as previously described.[13] In addition to the protein being kept rigid during the energy minimizations, the probes were also held internally rigid. During each Monte Carlo step, a probe was randomly chosen and translated up to 0.15 Å and rotated by 15° around a randomly chosen Cartesian axis. The solute−solvent cutoff was set to 150 Å so that each probe interacted with the entire protein (the protein has dimensions of ca. 62 Å × 62 Å × 62 Å). The Monte Carlo studies were carried out under 1 atmospheric pressure and a dielectric constant of 1.0. The number of configurations generated during the annealing procedure was 1 000 000 at each of the following temperatures, 300, 200, 100, and 0 °C, and 2 000 000 at −100 °C (i.e., a total of 6 000 000 configurations were sampled per probe).

Probe Clustering. At the end of a Monte Carlo run, each probe was assigned to a given cluster, which was sorted according to interaction energies with the protein. For example, the probe with the most favorable interaction energy was the parent of the first cluster (e.g., cluster 1). All other probes within a user-defined RMSD (2.0 Å in this study) were assigned to cluster 1. The parent of the second cluster was the probe molecule with the most favorable interaction energy that was not already a member of cluster 1. All probes within the same RMSD (2.0 Å) to parent 2, but that were not part of cluster 1, were assigned to cluster 2, and so on, until all probes were assigned to local minima (i.e., clusters) on the surface of the receptor. InsightII,[21] the Swiss PDB Viewer,[22] and VMD[23] were used to analyze the clusters.

Developing the Receptor-Based Pharmacophore Models. All 10 of the protein snapshots together with their associated probe clusters (represented by the cluster parents) were overlaid via five residues (D64, D116, E152, K156, and K159) in order to identify the consensus probe binding sites. The identified consensus sites were used to locate the elements of the dynamic pharmacophore model. We defined a consensus site as containing at least four clusters covering a wide range of the MD conformations; that is, a site was not considered as conserved if it was only observed for structures from the beginning or the end of the MD simulation. The centers and the radii of the feature sites were calculated from the Cartesian coordinates of the central atoms (C of CH_4, O of CH_3OH, Cl^-, and N of NH_4^+, respectively) of all parent probes in an individual consensus site. Therefore, the center of each functional feature site in the pharmacophore model was equal to the average position of the central atoms of the parents in each site, and the radii were set to the RMSD of all of the central parent atoms in the site. Catalyst[24] was then used to represent the spatial arrangement of the pharmacophore elements.

Model Validation and Database Screening. The model was validated by searching it against known IN inhibitors collected from publications in the last 10 years. Two small databases containing 385 active inhibitors ($IC_{50} < 20\ \mu M$ for either the 3′-processing or the strand transfer reaction) and 235 inactive compounds ($IC_{50} > 100\ \mu M$ for both 3′-processing and strand transfer), respectively, were build by the CatDB module in Catalyst. The validated model was used as a searching query to screen a small-molecule compounds database with the Catalyst software.

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 or ICN). The kinase was heat-inactivated, and 21bot was added in 1.5 M 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, a 20 nM concentration 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, and 8 M urea).

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

$$\% I = 100 \times [1 - (D - C)/(N - C)] \quad (3)$$

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 vs percent inhibition to obtain the concentration that produced 50% inhibition.

Results and Discussion

Monte Carlo Procedure and Probe Clusters. Hundreds of copies of the probe molecules were used to sample the substrate binding surface of IN via a Monte Carlo approach, functioning as a simulated annealing docking procedure, to identify favorable binding sites in the vicinity of the D,D(35)-E motif. In addition, K156 and K159 have been identified experimentally as being critical for DNA binding via photocrosslinking studies.[25,26] Electrostatics calculations[27] support the proposed function of these two key residues. The sampling sphere was 19 Å in radius and covers the above key residues. Figure 1a shows the initial sphere with 450 methanol probe molecules automatically generated by BOSS. Similarly, the initial evenly distributed probe-filled spheres were generated for the other probes (i.e., Cl^-, NH_4^+, and CH_4). Figure 1b shows the clustered probes at the end of the Monte Carlo-simulated annealing calculations. Clearly, well-formed clusters were identified for methanol probes close to the key residues of D64, D116, E152, K156, and K159. Because we concentrate only on the active site in this work, clusters that were formed near the top of the loop (140−149) or far from the active site were removed from consideration.

In the simulated annealing docking studies of Cl^-, although few probes gathered close to Mg^{2+} and between K156 and K159 as well, condensed clusters were not observed in the active site.

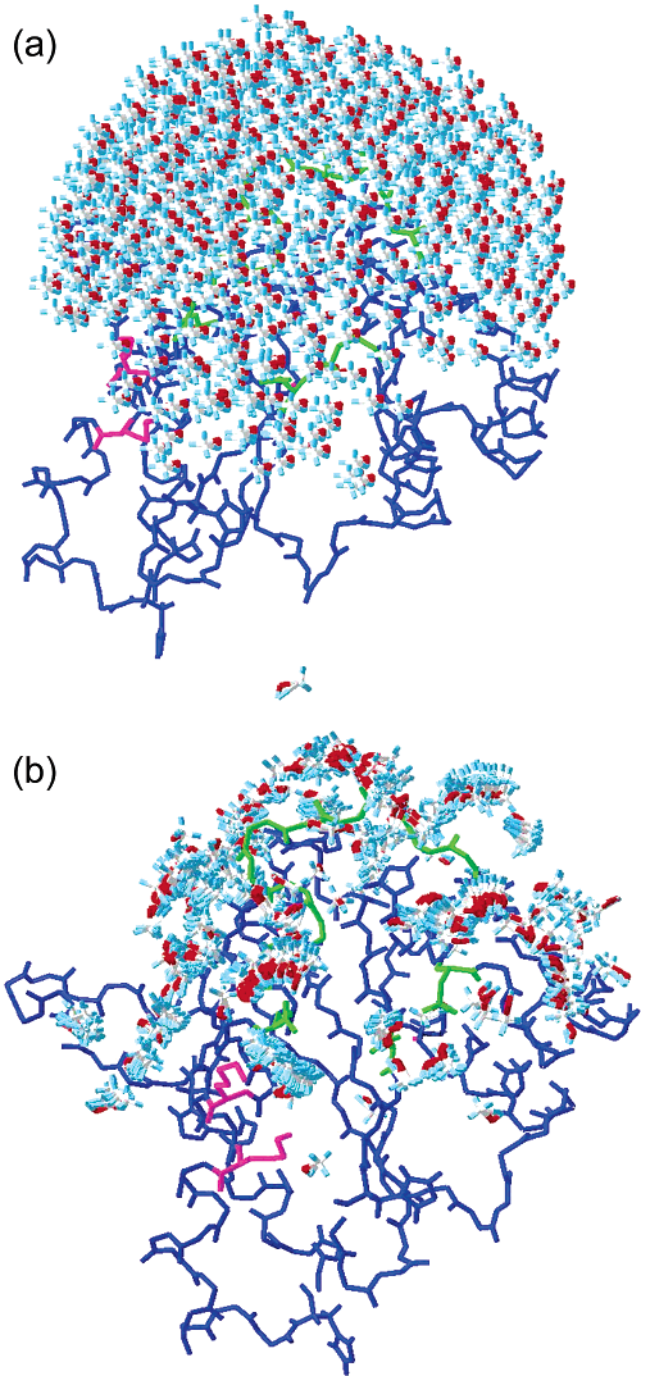


Figure 1. (a) Starting methanol probe sphere generated by BOSS on a well-equilibrated protein structure. (b) Clusters identified by methanol at the end of the Monte Carlo procedure. D64, D116, E152, and loop residues 140−149 are in green; K156 and K159 residues are in magenta. The sphere center is defined by atom, O_ϵ, of Q62 of IN, and the backbone of IN is in blue.

Actually, the places where few Cl^- probes gathered near Mg^{2+}, or K156, and K159 were able to be mapped by methanol clusters as well, indicating that complementary negatively charged features were not favorably detected in the active site of the selected snapshots. Apparently, this phenomenon is consistent with the electrostatic surface potential in the active site of the IN, as shown in Figure 2 (Mg^{2+} was removed), where a strong negative potential (in red) dominates the active site. For the same reason, however, the distribution of electrostatic potential accurately reflects the location of the condensed clusters formed by positively charged fragments, as represented by NH_4^+ in this

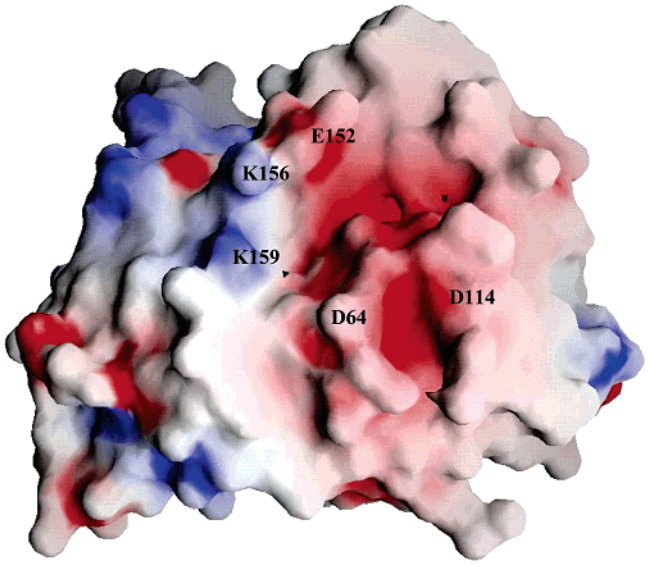


Figure 2. HIV-1 IN surface potential estimated by Grasp. Blue represents positive potential, and red represents negative surface potential. Mg^{2+} was not considered during the surface potential generation.

work. Interestingly, CH_4 probes, representing the hydrophobic feature, were not observed favorably binding in the active site of the selected snapshots. Therefore, only H-bonding and positively charged features were identified complementary to the ensemble of our selected snapshots.

The pharmacophore model based on a single snapshot, so it is referred to as a static pharmacophore model, has been shown to perform worse than ones considering multiple protein conformations.[13,14] Figure 3a illustrates the H-bond interactions between the methanol probe and a selected snapshot of IN, a well-equilibrated structure. Obviously, parent probes (sites 1 and 2) interact with D64 and D116 in such a way that probe hydrogens are pointing toward the protein residues and, thus, were regarded as H-bond donor sites. Likewise, methanol probes (sites 3 and 4) interact with K156 and K159 by accepting hydrogens from those two residues and thus were treated as H-bond acceptor sites. However, methanol probes (site 5) near E152 could form H-bonds with both E152 and K156 by acting as both an H-bond donor and an H-bond acceptor at the same time; therefore, this site was treated as both an H-bond donor and an H-bond acceptor site in the model. Site 6, in Figure 3a, was located by an ammonium cluster, which is complementary to the negative potential in the active site, as shown in Figure 2. Accordingly, Figure 3b shows the pharmacophore model corresponding to the interactions shown in Figure 3a; the excluded volumes were defined by the three catalytic active site residues (D64, D116, and E152), the Mg^{2+} ion, and two additional key residues (e.g., K156 and K159). The radii of the excluded volumes were set to 0.86 Å[28] for Mg^{2+} and 1.5 Å for the rest (representing key amino acid side chains).[13]

Dynamic Pharmacophore Model. The clustered probe configurations for selected snapshots are summarized in Figure 4, as exemplified by the case of methanol docking. Similar to the probe clusters observed in the well-equilibrated frame (Figure 1b), highly orientated clusters formed near the key residues (D64, D116, E152, K156, and K159) and some others were on the top of the active site loop (140−149) or far from the active surface. The protein conformations together with the associated clusters of CH_3OH and NH_4^+ were superimposed via the five key residues, D64, D116, E152, K156, and K159,

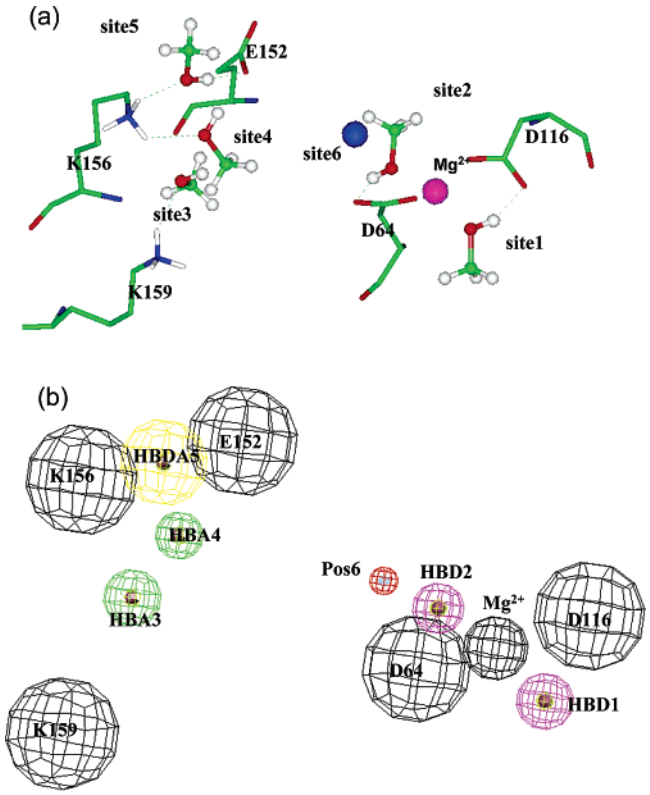


Figure 3. (a) H-bond interactions of probes with a well-equilibrated IN snapshot. (b) Static pharmacophore model derived from methanol- and ammonium-docked clusters associated with single snapshots. Black represents excluded volumes located by the key residues. Green represents H-bond acceptor (HBA), and pink represents H-bond donor (HBD). Yellow represents the feature that could function as both donor and acceptor.

in the catalytic domain (Figure 5a). Each sphere is defined by the central atom (O for CH_3OH and N for NH_4^+) of the parent molecule in the cluster. The grouped spheres in the aligned, overlaid structures represent consensus features. Seven groups were identified by the CH_3OH and labeled as P1−P7 and two by NH_4^+ probes, i.e., P8−P9. Similar to the methanol probe interactions with both K156 and E152 in Figure 3, each probe in group P6 and P5 in Figure 5a is observed sharing the same interaction mode with their target snapshots. Therefore, the properties defined by P5 and P6 could be either/both H-bond donor and/or H-bond acceptor, as labeled as HBDA in Figure 5b,c. Figure 5b,c shows the dynamic pharmacophore model corresponding to Figure 5a. However, in Figure 5b, the Mg^{2+} was treated as an excluded volume, as in previous work,[13,14] and is referred to as model I hereafter. It should be noted that while the Mg^{2+} was treated as an excluded volume in the pharmacophore model, it was treated as an atom with a charge of +2 during the docking of feature probes. In addition, Mg^{2+} was also considered as a positive-charged feature in a pharmacophore model with a radius defined by the RMSD of all grouped Mg^{2+} atoms from the 10 snapshots. This model, represented in Figure 5c, is referred to as model II hereafter. The characteristics of the dynamic pharmacophore models are summarized in Table 1. In model I, the radius of Mg^{2+} is 0.86 Å,[28] but in model II, it was set to 1.48 Å as the RMSD of all of the particles in that group.

As compared with a previously published dynamic pharmacophore model for IN, which was developed by Carlson et al.,[13] the current model conceptually extends the ligand sampling space to cover a larger portion of the substrate binding surface.

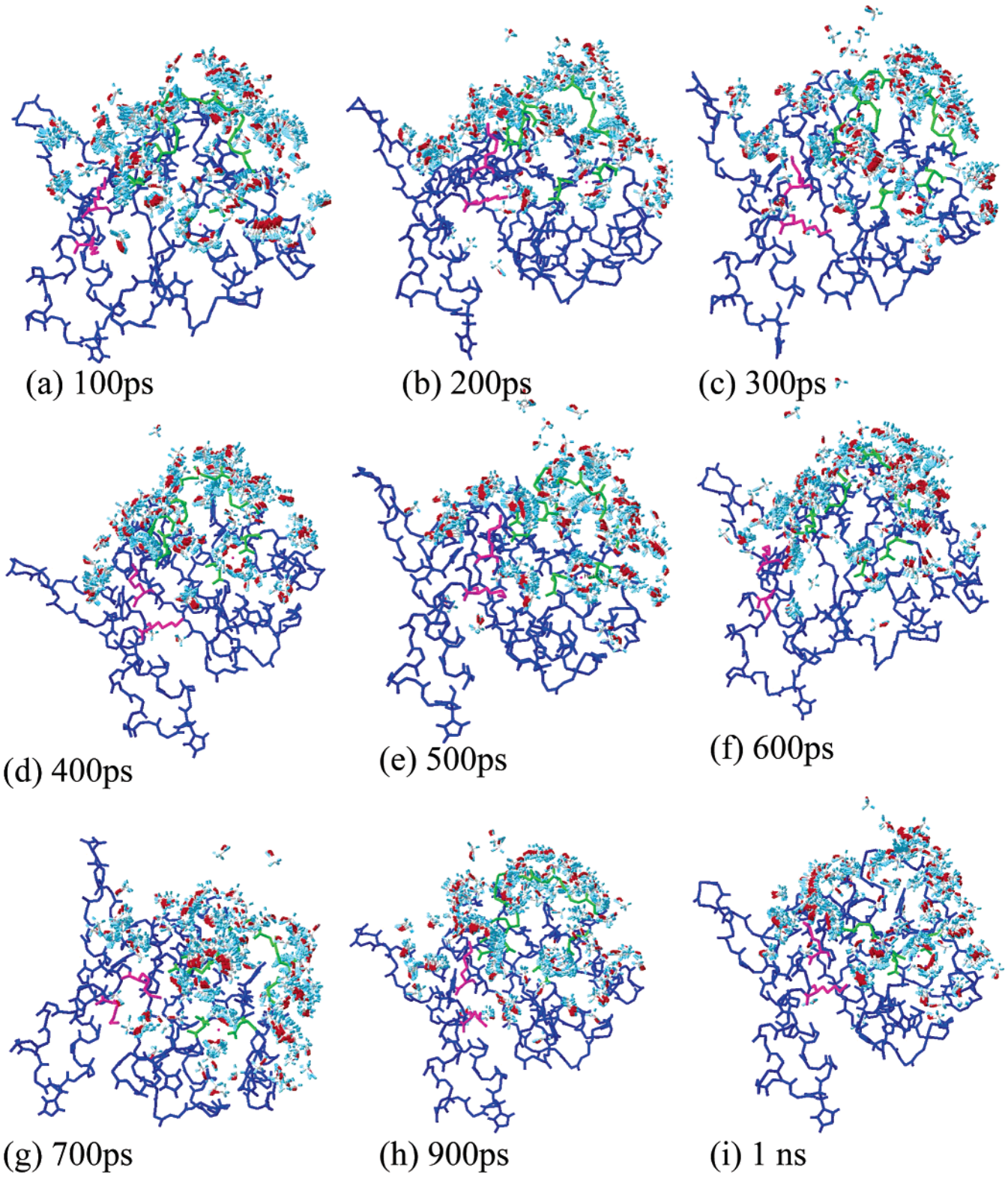


Figure 4. Methanol-clustered configurations generated from Monte Carlo simulations on each selected snapshot. Blue represents the IN backbone, and cyan/red molecular model represents the methanol probe molecule. D64, D116, E152, and loop residues 140−149 are in green. K156 and K159 residues are in magenta.

Figure 6 shows a superimposition of the two models in order to illustrate their differences (excluded volumes are not shown). The figure was generated by superimposing each model against a well-equilibrated structure, the starting snapshot for production phase of MD, which was used as a reference structure for overlaying of the models. The current model concentrates on the protein comprised of the D,D(35)E motif and residues K156 and K159, while the previous work focused mainly in the area of D64, D116, and the loop. Not surprisingly, our HDB2 feature, which reflects potential interactions between substrate with D64 and D116, align well with site D4 in the previous study.

Model Validation and Database Searching. The original models (Figure 5b,c) contain too many features and are thus too stringent to be used in database searching (i.e., no molecules

Table 1. Characteristics of the Dynamic Pharmacophore Model

feature[a]	location			radius (Å)
	X (Å)	*Y* (Å)	*Z* (Å)	
HBD 1	4.88	−0.06	4.84	1.24
HBD 2	2.60	2.37	4.71	0.97
HBA 3	−4.60	−3.20	10.16	1.24
HBA 4	−5.26	1.00	7.69	1.34
HBDA 5	−8.09	6.30	6.46	1.55
HBDA 6	−6.24	4.12	7.16	0.69
HBD 7	−2.93	6.26	6.71	1.34
Pos8	−1.03	2.24	2.48	0.79
Pos9	0.54	−0.07	4.26	1.2
K156	−9.81	3.19	6.71	1.50
K159	−7.61	−4.22	10.49	1.50
D64	1.64	0.10	2.45	1.50
D116	6.18	1.99	2.17	1.50
E152	−4.59	5.59	4.62	1.50
Mg^{2+}	4.05	1.10	4.11	0.86 (model I), 1.48 (model II)

[a] HBD, H-bond donor; HBA, H-bond acceptor; HBDA, could be both H-bond donor and acceptor; and Pos, positively ionizable feature.

can be found with nine spatially distributed pharmacophore elements); thus, they were refined in the manner as previously described.[13,14] The refining procedure is similar to that used in our recent work,[14] where we have suggested that a potent inhibitor compound should interact with D64, D116, E152, and also K156 and K159 at same time. Therefore, each refined model should contain at least one feature near E152, one feature near K156 or K159, and at least one feature near D64 or D116. Each refined model is an independent search query. The performance of the original models (model I in Figure 5b and model II in Figure 5c) was validated in terms of the performance of each refined query. Twenty independent queries were derived from both original models. Interestingly, any four-feature query containing a positively charged group (Pos8 or Pos9 in model I or Mg^{2+} in model II) did not identify any hits, probably due to the fact that most of the published active/inactive inhibitors lack this feature. However, from both biological studies and our electrostatic potential prediction (Figure 2), IN has a strong negative potential surface in the binding pocket although the substrate DNA binding mode is not known precisely. This phenomenon prompted us to design novel types of molecules, which are predicted to favorably interact with the negatively charged IN binding surface. Therefore, we further reduced the model II into a three-feature model containing the positively charged feature. Not surprisingly, very few published inhibitors could be recognized by the model. However, the three-feature query from model II was applied to screen our database. Regarding the validation of the queries derived from model I, only those that could identify at least 20 hits from the 385 active inhibitors are reported in Table 2. The selectivity index (SI) of each model was defined and computed, i.e., the ratio of the identification of actives over the incorrectly identified inactive

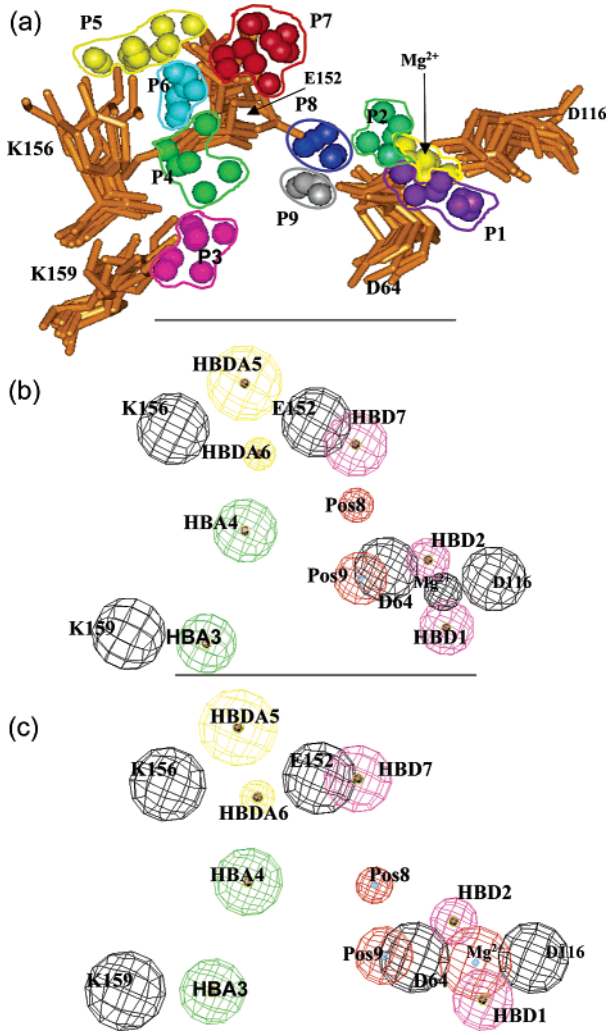


Figure 5. (a) Identification of the conserved binding sites from the overlain clusters. (b) Dynamic pharmacophore model derived from the multiple snapshots. Green represents HBA, and pink represents HBD. Yellow represents a feature that could be both a donor and an acceptor. Black represents the excluded volume, and red represents the positively charged feature.

Table 2. Model Validation by a Known Inhibitors Training Set

model type	model I							model II
query	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
features contained[a]	$A_{3,6}D_{1,7}$	$A_{3,4}D_{2,7}$	$A_{4,6}D_{2,7}$	$A_{4,5}D_{2,7}$	$A_4D_{2,6,7}$	$A_3D_{2,6,7}$	$A_5D_{2,5,7}$	A_4D_7Mg
active no.	51	89	91	34	49	54	23	7
active %[b]	21.7	37.8	38.7	14.5	20.9	23.0	9.8	3.0
inactive no.	3	13	18	10	7	3	5	2
inactive %[c]	0.8	3.4	4.7	2.6	1.8	0.8	1.3	0.5
SI	27.9	11.1	8.2	5.6	11.7	30.0	7.8	5.8
database	10	363	851	62	109	2	10	162

[a] $A_{x,y}D_{m,n}$: refined query, meaning this query containing feature HBA*x*, HBA*y*, HBD*m*, and HBD*n*, as shown in Figure 5 and Table 1. For example, $A_{3,6}D_{1,7}$ is the refined four-feature model keeping HBA3, HBDA6 (here as H-bond acceptor only), HBD1, and HBD7. A_4D_7Mg is an example of a refined three-feature model II, which regards Mg^{2+} as a positively charged feature, and additionally, it contains HBA4 and HBD7 as labeled in Figure 5 and Table 1. Likewise, $D_{2,6,7}A_3$ is another four-feature refined query containing HBD2, HBD7, HBD6 (here as H-bond donor only), and HBA3. [b] Active %: the percentage of recognized inhibitors over the overall active inhibitors in the database. For example, $A_{3,6}D_{1,7}$ could map out 51 active molecules from 235 active inhibitors; therefore, the active % of this model is 51/235 = 21.7%. [c] Inactive %: the percentage of recognized inhibitors over the overall inactive inhibitors in the database. For example, $A_{3,6}D_{1,7}$ could map out three active molecules from 385 active inhibitors; therefore, the inactive % of this model is 3/385 = 0.7%.

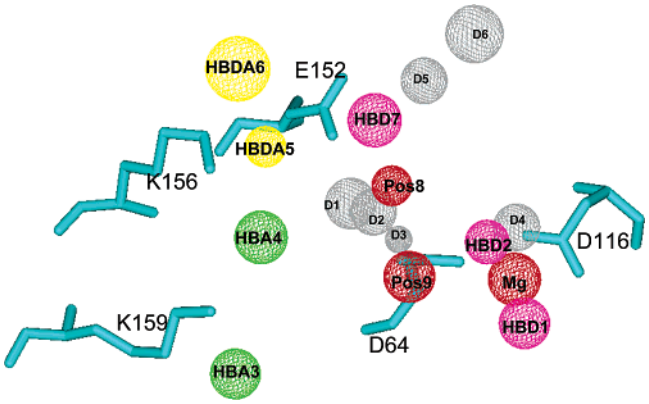


Figure 6. Comparison of the current dynamic pharmacophore model with a previously published pharmacophore model developed by Carlson et al.,[13] indicating that the model derived from the current work provides a more comprehensive description of the IN substrate binding site. Gray spheres (D1−D6) represent the H-bond donors in Carlson's model. The remaining spheres represent the features comprising the dynamic pharmacophore model derived from this work, consistent with Figure 5c. Magenta represents HBD, and green represents HBA. Red represents the positively charged feature. Yellow indicates that the feature could function as both an H-bond donor and an H-bond acceptor with neighboring atoms.

compounds, as defined by eq 4

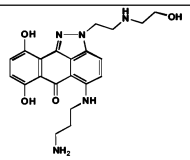
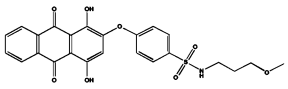
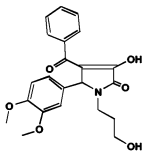
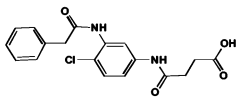
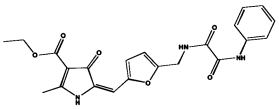
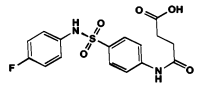
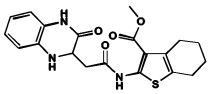
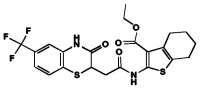
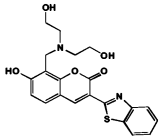
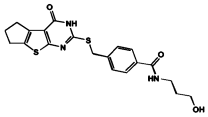
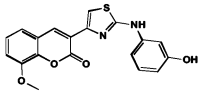
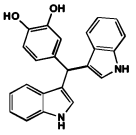
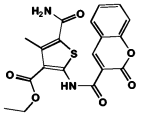
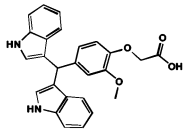
$$\mathrm{SI} = \left(\frac{R_{\mathrm{active}}}{R_{\mathrm{inactive}}}\right)/\left(\frac{D_{\mathrm{active}}}{D_{\mathrm{inactive}}}\right) \qquad (4)$$

where R_{active} is the total number of active inhibitors identified by the query; R_{inactive} is the total nonzero number of inactive compounds identified by the query (i.e., inactive compounds that fit the model and would therefore be incorrectly predicted to be active); $D_{\mathrm{active}} = 385$ and is the total number of collected known active inhibitors in our validation database; and $D_{\mathrm{inactive}} = 235$ and is the total number of collected known inactive inhibitors in the database.

This SI equation (eq 4) takes into account the differences in size between our active and inactive inhibitor databases. In an extreme ideal case, if we assume that only one inactive compound was falsely identified and all active compounds could be mapped by a query, the SI value could therefore be as high as 235. In Table 2, query Q6 showed the highest selectivity with an SI value of 30.0, among the listed queries. Some active molecules could be identified by multiple queries. For example, 34 hits in the active inhibitor database were identified by query Q4 (Table 2), but all of these molecules were able to be recognized by query Q3 as well. Queries Q2 and Q3 could map nearly 40% of the known active inhibitors. Interestingly, Q6, having the highest SI value of 30, maps 57 (54 active and three inactive compounds from Table 2) of the known inhibitors but could map only two compounds from the small molecule database. Q2 and Q3, on the other hand, identified over a hundred molecules. Not surprisingly, there are overlaps between the hit lists from database searching with the various queries. For example, 363 hits were identified by Q2, and 851 were identified by Q3. Among the two sets, 122 molecules were recognized by both queries, and nine of them were also recognized by Q7.

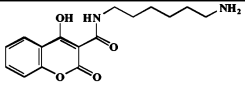
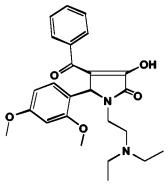
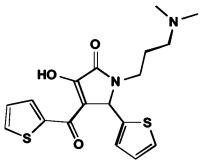
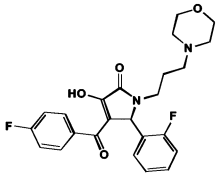
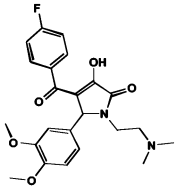
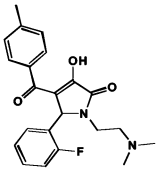
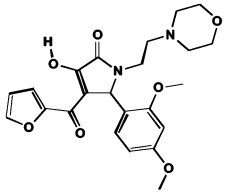
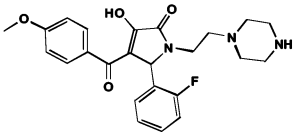
Biological Activity. On the basis of both structural novelty and the frequency of being identified by multiple queries, we selected 22 molecules for in vitro assay specific for the IN enzyme, and the results are presented in Tables 3 and 4. Five compounds exhibited an IC_{50} of less than 100 μM. Compound **1**, also known as piroxanthrone, has anticancer activity and recently was reported as a human topoisomerase II α inhibitor.[29]

Table 3. Inhibition of Catalytic Activities of HIV-1 IN by Structurally Novel Molecules Identified from the Four-Feature Queries Derived from Model I

compounds	Structure	IC_{50} (μM)[a] 3′-proc	IC_{50} (μM)[a] ST	Query mapped
1		18 ± 3	9 ± 5	**Q2, Q3, Q5, Q7, Q8**
2		100	23 ± 6	**Q2, Q3, Q4**
3		> 100	> 100	**Q2, Q3**
4		> 100	> 100	**Q3**
5		>100	40 ± 12	**Q3**
6		> 100	> 100	**Q3**
7		> 100	> 100	**Q3**
8		> 100	> 100	**Q3**
9		850	350	**Q2**
10		> 100	> 100	**Q2**
11		> 100	> 100	**Q2, Q3**
12		53 ± 24	10 ± 4	**Q3**
13		90 ± 17	56 ± 2	**Q3**
14		> 100	> 100	**Q2, Q3**

[a] 3′-proc, 3′-processing reaction; ST, strand transfer.

Table 4. Inhibition of Catalytic Activities of HIV-1 IN by Structurally Novel Molecules Identified from a Query Q8 Derived from Model II

compounds	Structure	IC_{50} (μM)[a]	
		3'-proc	ST
15		> 100	> 100
16		> 1000	> 1000
17		> 100	> 100
18		> 1000	> 1000
19		> 1000	> 1000
20		> 1000	> 1000
21		> 1000	395.0
22		> 1000	630

[a] 3′-proc, 3′-processing reaction; ST, strand transfer.

Interestingly, compound **1** also inhibits IN with IC_{50} values of 18 ± 3 and 9 ± 5 μM for the 3′-processing and strand transfer reactions, respectively (Table 3). Compound **2**, with a very different side chain than previously reported anthraquinones,[30,31] showed selectivity for the strand transfer reaction. Compound **12** has an equivalent potency for the strand transfer reaction with an IC_{50} value of 10 ± 4 μM and at least 5-fold selectivity over the 3′-processing reaction. In addition, compound **1** was able to map to at least five models (Q1, Q3, Q4, Q7 and Q8), suggesting that multiple binding modes exist and indicating a high tendency for the molecule to favorably interact with the target. Figure 7a−c shows a map of compound **1** against three of the five models by which the compound was identified. A

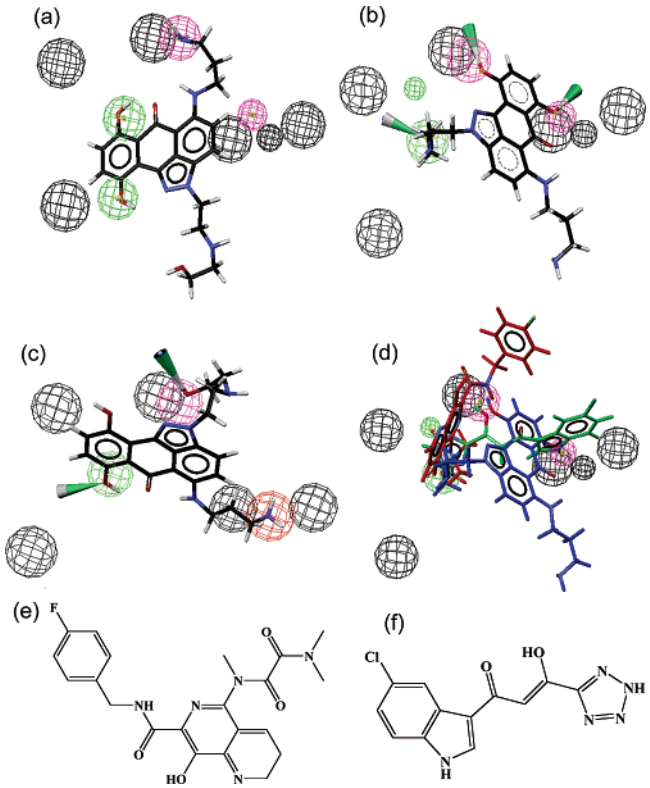


Figure 7. Representative molecules mapping against the selected queries. Black spheres represent excluded areas, green balls represent HBA, and pink balls represent HBD. The red sphere represents the positive ionizable feature. (a) Compound **1** mapping onto Q2. (b) Compound **1** mapping onto Q3. (c) Compound **1** mapping onto Q8. (d) Overlaying the mapping of three compounds onto Q3. Compound **1** is shown in blue, the same mapping as in part b. L870,812[32] is shown in red, and CITEP[18] is shown in green.

recently published inhibitor, L870,812[32] (Figure 7e), and 5-CITEP[18] (Figure 7f) were overlaid with compound **1** by mapping them to the model Q3 (Figure 7d). The mapping of 5-CITEP on Q3 is very similar to that as revealed in the X-ray structure of its complex with IN; however, L870,812 maps with the two H-bond acceptor/donors (HBA4, HBDA6) near K156 and K159 and one donor (HBD7) near E152, suggesting that multiple binding modes exist.

Conclusions

Multiple functional group probes were used to sample a broad binding surface on the IN catalytic domain via Monte Carlo procedures. Favorable binding locations were used to define complementary features of the IN active site. An ensemble of conformations collected at 100 ps intervals from an MD simulation was used to build dynamic pharmacophore models reflecting native protein dynamics. We have reported two types of models containing all possible feature elements as located by the feature probes: (i) in model I, Mg^{2+} chelating D64 and D116 as resolved via X-ray was treated as a charged part of the protein and therefore was regarded as an excluded volume, and (ii) for the first time, we attempted to build a new type of model, referred to as model II, in which Mg^{2+} was treated as a positively charged probe site. Both types of models were validated by attempting to match them with known inhibitors. Model I performed much better than model II. Our own experimental testing indicates that a compound mapping multiple refined queries shows high potency against IN.

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. thanks Prof. H. A. Carlson for helpful advice and Prof. B. M. Pettitt for many fruitful discussions.

References

(1) Dayam, R.; Deng, J.; Neamati, N. HIV-1 integrase inhibitors: 2003−2004 update. *Med. Res. Rev.* **2006**, in press.
(2) Neamati, N.; Sunder, S.; Pommier, Y. Design and discovery of HIV-1 integrase inhibitors. *Drug Discovery Today* **1997**, *2*, 487−198.
(3) Zhang, X.; Neamati, N.; Lee, Y. K.; Orr, A.; Brown, R. D.; Whitaker, N.; Pommier, Y.; Burke, T. R., Jr. Arylisothiocyanate-containing esters of caffeic acid designed as affinity ligands for HIV-1 integrase. *Bioorg. Med. Chem.* **2001**, *9*, 1649−1657.
(4) Fesen, M. R.; Pommier, Y.; Leteurtre, F.; Hiroguchi, S.; Yung, J.; Kohn, K. W. Inhibition of HIV-1 integrase by flavones, caffeic acid phenethyl ester (CAPE) and related compounds. *Biochem. Pharmacol.* **1994**, *48*, 595−608.
(5) Marchand, C.; Neamati, N.; Pommier, Y. In vitro human immunodeficiency virus type 1 integrase assays. *Methods Enzymol.* **2001**, *340*, 624−633.
(6) Dayam, R.; Neamati, N. Small-molecule HIV-1 integrase inhibitors: The 2001−2002 update. *Curr. Pharm. Des.* **2003**, *9*, 1789−1802.
(7) Singh, S. B.; Zink, D. L.; Heimbach, B.; Genilloud, O.; Teran, A.; Silverman, K. C.; Lingham, R. B.; Felock, P.; Hazuda, D. J. 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.
(8) Young, S. L-780,810: A potent antiviral HIV integrase inhibitor with potential clinical utulity. The XIV International AIDS Conference, Barcelona, 2002.
(9) Yoshinaga, T.; Sato, A.; Fujishita, T.; Fujiwara, T. 9th Conference on Retroviruses and Opportunistic Infections, Seattle, WA, 2002.
(10) Zhao, H.; Neamati, N.; Sunder, S.; Hong, H.; Wang, S.; Milne, G. W.; Pommier, Y.; Burke, T. R., Jr. Hydrazide-containing inhibitors of HIV-1 integrase. *J. Med. Chem.* **1997**, *40*, 937−941.
(11) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazumder, A.; Zhao, H.; Burke, T. R., Jr.; Pommier, Y.; Milne, G. W. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem.* **1997**, *40*, 930−936.
(12) 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.
(13) 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.
(14) Deng, J.; Lee, K. W.; Sanchez, T.; Cui, M.; Neamati, N.; Briggs, J. M. Dynamic receptor-based pharmacophore model development and its application in designing novel HIV-1 integrase inhibitors. *J. Med. Chem.* **2005**, *48*, 1496−1505.
(15) Jorgensen, W. L. *BOSS*, 4.0 ed.; Yale University: New Haven, CT.
(16) Jorgensen, W. L.; Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimizations for crystal of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(17) Lee, M. C.; Deng, J.; Briggs, J. M.; Duan, Y. Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophys. J.* **2005**, *88*, 3133−3146.
(18) Goldgur, Y.; Craigie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, T.; Fujishita, T.; Sugimoto, H.; Endo, T.; Murai, H.; Davies, D. R. 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.
(19) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209−225.
(20) Barril, X.; Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432−4443.
(21) *InsightII*, 2000 ed.; Accelrys Inc.: San Diego.
(22) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(23) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph* **1996**, *14*, 33−38, 27−38.
(24) *Catalyst*, 4.7 ed.; Accelrys Inc.: San Diego, CA.
(25) Jenkins, T. M.; Esposito, D.; Engelman, A.; Craigie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photocrosslinking. *EMBO J.* **1997**, *16*, 6849−6859.
(26) Drake, R. R.; Neamati, N.; Hong, H.; Pilon, A. A.; Sunthankar, P.; Hume, S. D.; Milne, G. W.; Pommier, Y. Identification of a nucleotide binding site in HIV-1 integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4170−4175.
(27) Lins, R. D.; Adesokan, A.; Soares, T. A.; Briggs, J. M. Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacol. Ther.* **2000**, *85*, 123−131.
(28) Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in Halides and Chalcogenides. *Acta Crystallogr.* **1976**, *A32*, 751−767.
(29) Christmann-Franck, S.; Bertrand, H. O.; Goupil-Lamy, A.; der Garabedian, P. A.; Mauffret, O.; Hoffmann, R.; Fermandjian, S. Structure-based virtual screening: An application to human topoisomerase II alpha. *J. Med. Chem.* **2004**, *47*, 6840−6853.
(30) Singh, S. B.; Jayasuriya, H.; Dewey, R.; Polishook, J. D.; Dombrowski, A. W.; Zink, D. L.; Guan, Z.; Collado, J.; Platas, G.; Pelaez, F.; Felock, P. J.; Hazuda, D. J. Isolation, structure, and HIV-1-integrase inhibitory activity of structurally diverse fungal metabolites. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 721−731.
(31) Fesen, M. R.; Kohn, K. W.; Leteurtre, F.; Pommier, Y. Inhibitors of human immunodeficiency virus integrase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2399−2403.
(32) Hazuda, D. J.; Young, S. D.; Guare, J. P.; Anthony, N. J.; Gomez, R. P.; Wai, J. S.; Vacca, J. P.; Handt, L.; Motzel, S. L.; Klein, H. J.; Dornadula, G.; Danovich, R. M.; Witmer, M. V.; Wilson, K. A.; Tussey, L.; Schleif, W. A.; Gabryelski, L. S.; Jin, L.; Miller, M. D.; Casimiro, D. R.; Emini, E. A.; Shiver, J. W. Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science* **2004**, *305*, 528−532.

JM0510629