

Prediction of HIV-1 Integrase/Viral DNA Interactions in the Catalytic Domain by Fast Molecular Docking

Adeyemi A. Adesokan,^{†,‡} Victoria A. Roberts,^{†,‡} Keun Woo Lee,^{†,‡} Roberto D. Lins,[§] and James M. Briggs^{*,†}

Department of Biology and Biochemistry, Houston Science Center, Room 402D, University of Houston, Houston, Texas 77204-5001, Department of Molecular Biology, MB4, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Laboratory of Physical Chemistry, ETH-Hnggerberg, CH-8093 Zurich, Switzerland

Received April 21, 2003

This study details the separate analyses of binding specificity of HIV-1 integrase (IN) and viral B-DNA forms through ligand–receptor docking studies by means of a fast molecular docking method. The application of solvated electrostatics with the University of Houston Brownian Dynamics Program (UHBD) and configurational sampling by the Daughter of Turnip (DOT) docking program resulted in the computation of energies of more than 113 billion configurations for each ligand–receptor docking study, a procedure considered computationally intractable a few years ago. A specific binding pattern of viral DNA to the IN catalytic domain region has been predicted as a result of these calculations. In a representative docked configuration, we observe the 3'-hydroxyl of the conserved deoxyadenosine to be close to one of the two divalent metal ions that are necessary for catalysis. A superimposition of our energy-minimized docked complex on representative structures from a molecular dynamics (MD) simulation of a crystallographically resolved IN/inhibitor complex revealed an overlap of viral DNA with the inhibitor, indicating that the bound inhibitor might operate by blocking substrate binding. The DOT docking calculation also identified a second, adjacent DNA-binding site, which we believe is the nonspecific host DNA binding site. The binding pattern predicted by DOT complements previous electrostatics, MD simulation, photo-cross-linking, and mutagenesis studies and also provides a further refinement of the IN/viral DNA binding interaction as a basis for new structure-based design efforts.

Introduction

The human immunodeficiency virus type 1 (HIV-1) encodes three enzymes as part of the POL gene, namely, the reverse transcriptase, protease, and integrase. HIV integrase is the retroviral protein responsible for the insertion of a double-stranded DNA copy into host chromosome, thereby recruiting host cell machinery for the synthesis of viral proteins. It represents the third main enzyme target for inhibitor design. Inhibitors of HIV reverse transcriptase and protease have been developed, but the emergence of mutant HIV strains has caused these drugs to become ineffective. Therefore, there is an increasing need for the development of new antiviral agents directed against the DNA integration step in the HIV replication cycle.

HIV-1 integrase (IN) is composed of a single polypeptide chain that folds into three functional domains¹ and belongs to the polynucleotidyl transferase superfamily. The active site is located in the catalytic domain, which binds two Mg²⁺ ions that are required for catalysis.^{1,2} The active site consists of two Asp residues and one Glu in the conserved D,D(35)E motif, each of which is required for catalysis.^{2–4} Attempts to elucidate the complete three-dimensional structural characteristics of the IN active site by X-ray crystallography have resulted in structures with relatively high B factors in the active

site region, which contains residues 141–148. Thus, an average structure resulting from a 500 ps molecular dynamics (MD) simulation of the catalytic domain containing two Mg²⁺ ions and a dianionic phosphate group (HPO₄²⁻) has been used in this study.^{5,6} This structure was used because of the probability that it represents a better model for docking than a structure modeled in the absence of a HPO₄²⁻ substrate.

IN carries out at least two reactions: 3'-processing and strand transfer.^{7–9} In the 3'-processing reaction, IN hydrolyzes a dinucleotide from both 3'-ends of viral DNA adjacent to the highly conserved CA dinucleotide in the long terminal repeat region.^{10,11} The newly exposed 3'-hydroxyl groups then undergo a concerted transesterification strand transfer reaction across a 5-base pair stretch in the host DNA such that the viral DNA is now inserted into the host chromosome flanked on each side by the same 5-base pairs.^{10,11} The catalytic domain of the IN can also carry out another reaction known as disintegration.¹² In this reaction, IN catalyzes a strand cleavage on a Y-shaped substrate. This reaction generates two DNA fragments and represents a reverse strand transfer.

The lack of detailed structural information about IN/substrate interactions, apart from photo-cross-linking results, has so far hindered the search for strong and selective IN inhibitors. Instead, drug design efforts have become dependent on pharmacophore hypotheses based on structures of known inhibitors.^{13–15} The first receptor-based pharmacophore model was presented recently and was developed with a view to taking into account

* To whom correspondence should be addressed. E-mail: jbriggs@uh.edu. Phone: (713) 743-8366. Fax: (713) 743-8351.

[†] University of Houston.

[‡] The Scripps Research Institute.

[§] ETH-Hnggerberg.

[‡] These authors have contributed equally to this work.

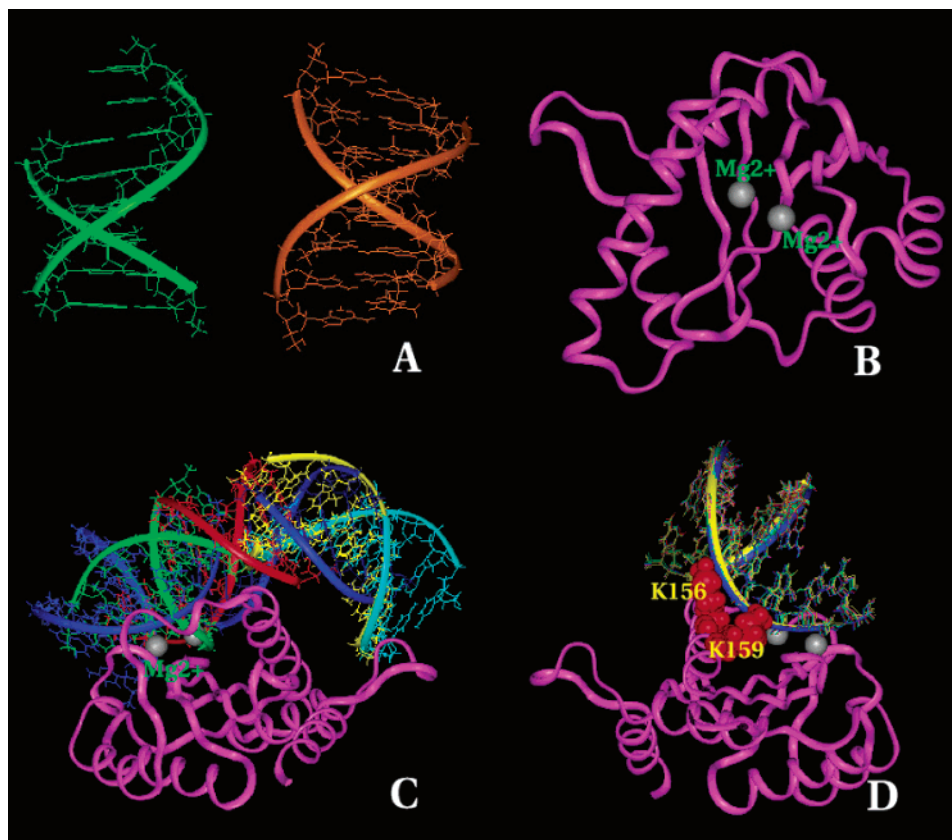


Figure 1. (A) Intact (right) and 3'-processed viral DNA (left) substrates used in the docking simulations. (B) IN structure obtained from a 500 ps MD simulation of 1BIS featuring two Mg^{2+} ions and a HPO_4^{2-} ligand. (C) Results from rough docking simulation ($12^\circ \times 12^\circ$ rotational screen) featuring 3'-processed viral DNA clustering over the IN active site. (D) Refined docking simulation ($6^\circ \times 6^\circ$ rotational screen) showing 3'-processed viral DNA in close interaction with K156 and K159.

the effects of conformational flexibility of the active site.^{16,17} These models have so far been successful in the discovery of new IN lead inhibitors through database searches, but uncertainties still remain concerning details of enzyme/ligand interactions and effects of the molecular framework that are not taken into account by the pharmacophore models. Hence, the main purpose of this study was to obtain docked complexes of IN and viral B-DNA forms that were consistent with previous computational and experimental data.^{6,18–20} The elucidation of the structure of such a complex would be essential to the structure-based study of IN and specifically to the search for stronger and more selective inhibitors.

In the absence of detailed experimental data, computational docking methods have become a vital tool for probing the interactions of enzymes with inhibitors and substrates mainly because of recent improvements in search algorithms and energy functions.^{21–25} The docking program DOT²¹ provides a complete search of all orientations between two rigid molecules by systematic rotation and translation of one molecule about another. In this study, DOT was used to perform a search of the ligand (DNA) translated in 54 000 different orientations about the receptor (IN), which was enclosed in an implicitly solvated $128 \times 128 \times 128$ point grid. The application of implicitly solvated electrostatics computed by UHBD²⁶ and sampled by DOT resulted in the computation of energies of more than 113 billion configurations for each ligand–receptor docking study. Since DOT performs a complete search, long-range as

well as short-range interactions can be examined.²³ Herein, we describe the application of DOT to the prediction of a specific binding pattern of viral DNA to the IN catalytic domain region. Our calculated complex exhibits the coordination of the newly exposed 3'-hydroxyl to one of the magnesium ions in the IN catalytic domain. Analyses of solvated minimizations carried out on the representative resultant docked configuration identified viral DNA interactions with D64, D116, E152, K156, and K159, all of which are residues experimentally determined to be in close contact with DNA and critical for IN function.^{18–20}

Computational Methodology

Coordinate Preparation. It is known that sequences at the viral DNA ends are intimately involved in integration because they serve as biochemical substrates for the integrase enzyme and as attachment (att) sites for the integrated provirus.²⁷ Hence, in this study we used two DNA forms (Figure 1A) that were based on a specific viral DNA nucleotide sequence known to be critical for IN/DNA binding.²⁸ The first was the unprocessed DNA form, representing the preprocessed substrate with no cleavage at the viral ends. The unprocessed viral DNA molecule was eight base pairs with the nucleotide sequence 5'-CTAGCAGT-3'. The second DNA type was the 3'-processed DNA molecule, a representation of the product of the IN processing reaction, featuring a 5'-CA overhang resulting from the cleavage of the 3'-GT dinucleotide adjacent to the conserved 3'-CA.²⁸ This 3'-processed DNA form consisted of 14 nucleotides, with the cleaved strand being 5'-CTAGCA-3' and the second strand being 5'-ACTGCTAG-3'. These B-form viral DNA molecules were modeled using the biopolymer module of the InsightII²⁹ molecular modeling software pack-

age. The total charges on the 3'-processed and unprocessed viral DNA were -12 and -14 , respectively. Hydrogen atoms were added by means of the PROTONATE utility using AMBER³⁰ force field parameters.

The IN structure used in this study was obtained from a 500 ps MD simulation that included two Mg^{2+} ions and one HPO_4^{2-} ^{5,6} and started from the crystallographic structure 1BIS³¹ (Figure 1B). Polar hydrogen atoms for IN were also added by PROTONATE using the AMBER force field. The total charge on the IN coordinates was $+4$, with each magnesium atom assigned a charge of $+2$. The HPO_4^{2-} was removed before the docking simulation.

Molecular Descriptions Used in the DOT Calculation.

For the DOT calculation, viral DNA with a diameter of 20 \AA and a length of $\sim 30 \text{ \AA}$ was assigned as the moving molecule while IN with a diameter of 54 \AA was assigned as the stationary molecule.³² Each molecule was held internally rigid throughout the docking studies. These assignments provided an advantage because they allowed the use of finer spacing for the grid because the minimum dimension of the grid must be $2M + S$, where M and S represent the diameters of the moving and stationary molecules, respectively.²³

Electrostatic and shape potentials for the stationary molecule (IN) and atomic charges for the moving molecule (DNA) are required for the DOT energy calculations. IN was centered on a cubic grid of 128 \AA per side with a spacing between grid points of 1 \AA . The electrostatic potential of IN was calculated on this grid with the UHBD program at a temperature of 300 K , an ionic strength of 150 mM , a solvent dielectric of 80 , and a dielectric of 3 for the protein interior. The shape of the stationary molecule was an approximate van der Waals potential, defined by an excluded volume consisting of all grid points inside the van der Waals volume and a surrounding 3 \AA favorable layer.²¹ Each moving molecule atom lying within the favorable layer contributed -0.1 kcal/mol to the total energy. When the van der Waals term was not included in the energy calculation, the shape of the stationary molecule was defined solely by its excluded volume. The shape of the moving molecule (DNA) was defined by its atomic positions. The DNA charge distribution was represented by partial charges at its atomic positions, with assigned charges taken from the AMBER library.

Since the moving molecule is represented by its atomic centers only, some atoms of the moving molecule can approach within 1.4 \AA (molecular surface radius) of the stationary molecule centers.^{22,23} This allows for some induced fit but can create problems with calculation of the electrostatic term. Moving molecule atoms that lie between the molecular and solvent-accessible surfaces of the stationary molecule can result in unrealistically large contributions to the electrostatic energy. To alleviate this problem, we imposed limits on the magnitude of the electrostatic potential of the stationary molecule. The values of the electrostatic potential grid for IN were limited to the extreme positive and negative potentials found on the IN solvent-accessible surface. This surface represents the closest approach of the center of a water molecule ($\sim 2.9 \text{ \AA}$ between atom centers). Limits were set to an upper value of $+6 \text{ (kcal/mol)/e}$ and a lower value of $-6.0 \text{ (kcal/mol)/e}$. Initially limits of $\pm 20 \text{ (kcal/mol)/e}$ were used, but this resulted in a much larger number of favorable energy false positives.

DOT: Systematic Search and Calculation of Intermolecular Energies. The DOT program centers each rotational orientation of the moving molecule at each position on the grid that surrounds the stationary molecule. For each configuration of the two molecules, the intermolecular energy was calculated as either a sum of electrostatic and van der Waals terms or the electrostatic energy alone. The calculation is independent of the number of atoms in each molecule but rather depends on the number of grid points and the number of rotational orientations of the moving molecule. The mathematics of the DOT calculation requires periodic boundary conditions, so the grid is repeated in all directions. Therefore, it is imperative that the grid be large enough that the stationary molecule

potentials are close to zero at the grid boundary.²³ A cubic grid of 128 points on a side with a 1 \AA spacing between points was sufficiently large and of reasonable fineness for the translational search. For the translational search, the moving molecule was centered at each grid point (128^3 or about 2.1 million positions). A set of $54\,000$ rotational orientations provided a resolution of 6° for the rotational search. Together, the rotational and translational search resulted in over 113 billion configurations ($128 \times 128 \times 128 \times 54\,000 = 113$ billion) for the two molecules. The calculations were performed on three nodes of a Beowulf cluster or on a farm of Sun workstations.

Energy Computation by Correlation Functions. The DOT program computes energies by the placement of the moving molecule in a potential field generated by the fixed molecule. The total energy of each configuration is described as a correlation function.²¹ Given a potential field $S(\mathbf{r})$ and a probe function $M(\mathbf{r})$ as descriptions of the stationary and moving molecules, respectively, the energy is given by²¹

$$E = \int M(\mathbf{r}) S(\mathbf{r}) \, d\mathbf{r}$$

In the case where the moving molecule is rotated by an angle ϑ and displaced from the origin by a vector \mathbf{r}_0 , the energy of the system is given by²¹

$$E_\vartheta(\mathbf{r}_0) = \int M_\vartheta(\mathbf{r} - \mathbf{r}_0) S(\mathbf{r}) \, d\mathbf{r}$$

For the electrostatic energy term, $S(\mathbf{r})$ is the electrostatic potential field of IN and $M(\mathbf{r})$ is the set of partial charges for viral DNA. For the van der Waals term, $S(\mathbf{r})$ is a mask function that represents the volume of space occupied by the fixed molecule.²¹

Correlations can be calculated by means of the convolution theorem. The electrostatic and nonbonded energy functions are described on a grid of N points (128^3). Evaluating the correlation directly requires $N \times N$ multiplications, in this case 128^6 . Evaluating the correlation using fast Fourier transforms costs three fast Fourier transforms, each proportional to $N \log N$, and N multiplications, so the computational cost is proportional to $N(3 \log N + 1)$.²¹ Multiple correlation functions are required, but only the moving molecule function changes. Therefore, the fast Fourier transforms for the stationary molecule need to be calculated only once, resulting in a cost proportional to $N(2 \log N + 1)$.²²

Solvated Energy Minimizations. The DOT program uses rigid molecules for its complete translational and rotational search, providing initial complexes for methods that allow flexibility. Here, we used energy minimizations (EM) to provide some flexibility to the docked complex. The system, which comprises the 3'-processed viral DNA/IN complex, was immersed in a $60 \text{ \AA} \times 60 \text{ \AA} \times 56 \text{ \AA}$ water box. The system had $19\,721$ total atoms, including 5635 water molecules. A 12 \AA cutoff and CVFF force field parameters were used in the EM calculations. Initially 500 steps of EM using the steepest descent algorithm was applied followed by another 5000 steps using the conjugate gradient algorithm. All EM calculations were performed on an SGI Octane workstation by means of the DISCOVER³³ program.

Results

Our goal in these docking studies was to identify the binding site for viral DNA on the IN catalytic domain. This domain, however, binds both target host DNA and viral DNA. Even though we used a DNA fragment with a viral DNA sequence known to be critical for IN binding, this fragment was likely to have favorable interactions with the nonspecific target DNA binding site as well as with the viral DNA binding site. Furthermore, the binding sites may not be complete. For example, the viral DNA has substantial interactions

with the IN C-terminal domain,³⁴ which was not included in these docking studies.

To take these issues into account, we ranked configurations found by the DOT docking program in two different ways. The interaction energy calculated by DOT is a sum of electrostatic and van der Waals terms. The electrostatic term must be included to represent the essential interactions between the negatively charged DNA and the complementary, positively charged regions of the protein. The van der Waals term used by DOT is simply a count of DNA atoms within 3 Å of the IN surface in each configuration. This term will emphasize the DNA-binding site with the greater contact area on the IN catalytic domain. Two calculations were carried out: the first in which configurations were ranked solely by the electrostatic interaction energy; the second in which configurations were ranked by the sum of both energy terms. The first method should show well-ranked solutions for both DNA-binding sites, whereas the second focuses on the site with the greatest surface area.

Evaluation by Electrostatic Energy: IN/Viral DNA Complexes. A preliminary test run using a coarse rotational sampling for the DNA (a resolution of about 12°) was performed on IN with DOT. The top 30 configurations from the 3'-processed DNA molecule had intermolecular energies calculated by DOT ranging from -17.4 to -15.2 kcal/mol. Of these 30 solutions, the top 10 lay over the HIV-1 IN active site, with 8 of 10 having the CA 5' overhang generally oriented toward the Mg²⁺ ions in the active site (Figure 1C).

These results prompted a finer rotational search with a set of 54 000 orientations, which provided a 6° resolution. Results from this docking simulation revealed the 30 best 3'-processed viral DNA configurations clustered over the active site of IN. The top 30 configurations ranged from -19.0 to -17.6 kcal/mol, highlighting the advantage of the finer rotational sampling. Within the top 10 configurations there is a single tight cluster featuring eight solutions in contact with K159 and K156 (Figure 1D). The results from the IN/intact DNA docking simulation had the top 30 solutions within an energy range of -20.6 to -17.8 kcal/mol. Clustering over the active site residues was also observed in this case.

The top 30 solutions from the docking study using the 3'-processed viral DNA showed two general types of binding over the IN active site. In one, the DNA lies along the protein surface with contacts to K156 and K159. This is represented by a tight cluster of 8 of the top 10 solutions (Figure 1D), with other solutions in the top 30 showing similar alignments. There is also a second type of binding in which one end of the DNA lies against the active site. The best ranked configuration in this category was ranked 11th by electrostatic energy.

This configuration exhibits good agreement with experimental data^{6,18-20} for bound viral DNA and shows extensive potential contacts of the conserved nucleotides with active site residues (parts A and C of Figure 2). In this configuration, the 3'-hydroxyl of the conserved deoxyadenosine is within hydrogen-bonding distance to D64 and within bonding distance to the Mg²⁺ coordinated by D64. This configuration also has hydrogen bonds between the 3'-oxygen of the conserved cytosine and NZ of K159, between the O3' of the guanine and

T66, and between O5' of the conserved cytosine and H67. All of these residues have been hypothesized or shown to be crucial for IN/viral DNA and IN/inhibitor binding (Figure 2).^{18-20,35,37} Hydrogen bonds were also observed between 3'-processed viral DNA and IN residues E69, E92, and G70. The 3'-processed viral DNA also contacts IN residues C65, D64, D116, and N155. Similar interactions were observed within the unprocessed viral DNA/IN complex.

A solvated minimization was carried out on this complex of the 3'-processed viral DNA with IN. Minimization caused only slight movement of viral DNA and IN side chains. Superimposition of the docked and minimized complexes resulted in an RMSD of 1.0 Å for IN and 0.9 Å for viral DNA. Analyses of the optimized structure identified hydrogen bonds between the O3' of the conserved deoxyadenosine and E152 and between the 5'-deoxyadenosine overhang and D116 (parts B and D of Figure 2). Hydrogen bonds were also present between the O1P of the conserved cytosine and E69 and between O2P of the conserved deoxyadenosine and T66.

Superimposition of Docked Complex with Relevant MD Structures. To further ascertain the validity of our docking results, we superimposed an energetically favored 3'-processed viral DNA/IN docked complex on a 2 ns MD simulation of IN and the inhibitor 1-(5-chloroindol-3-yl)-3-(tetrazolyl)-1,3-propadione enol (5CITEP).³⁶ Superimposition was carried out on snapshots of the IN/5CITEP system at 500, 1000, 1500, and 2000 ps of the MD simulation.³⁶ Results revealed a viral DNA overlap with 5CITEP (parts E and F of Figure 2), indicating that the inhibitor may act by physically blocking the binding of the substrate DNA molecule(s). The 5CITEP tetrazole ring was seen to overlap with the conserved deoxyadenosine.

Superimposition of the 3'-processed viral DNA was also carried out on an average structure resulting from a 500 ps MD simulation of IN featuring two Mg²⁺ ions and a dianionic phosphate group (HPO₄²⁻).¹³ This was done to serve as a means of comparing the docked position of 3'-processed viral DNA and HPO₄²⁻, which had been removed for the docking simulation. The HPO₄²⁻, which was modeled in such a way as to mimic a portion of the DNA backbone of a substrate,¹³ was aligned in such a way with the docked viral DNA backbone to reflect the position occupied by the phosphate of 3'-guanine of viral DNA prior to the 3'-processing reaction.

Evaluation by Total Energy: IN/Target DNA Complex. The DOT search was repeated with configurations being ranked by total energy. The 30 best 3'-processed DNA solutions were centered over residue K156. Most were tightly clustered, as represented by 9 of the top 10 solutions (Figure 3). In this configuration, the DNA major groove lies over the helix with K156, and the phosphate backbone contacted K186 and K188, which are on a loop more distant from the binding site. Although the phosphate backbones of most of these top-ranked solutions were closely aligned, they varied by translation along the DNA axis. These solutions are similar to the cluster results from the runs ranked by electrostatic energy alone (Figure 1D) but are translated more toward the loop containing K186 and K188. This positioning allows greater contact with the IN surface,

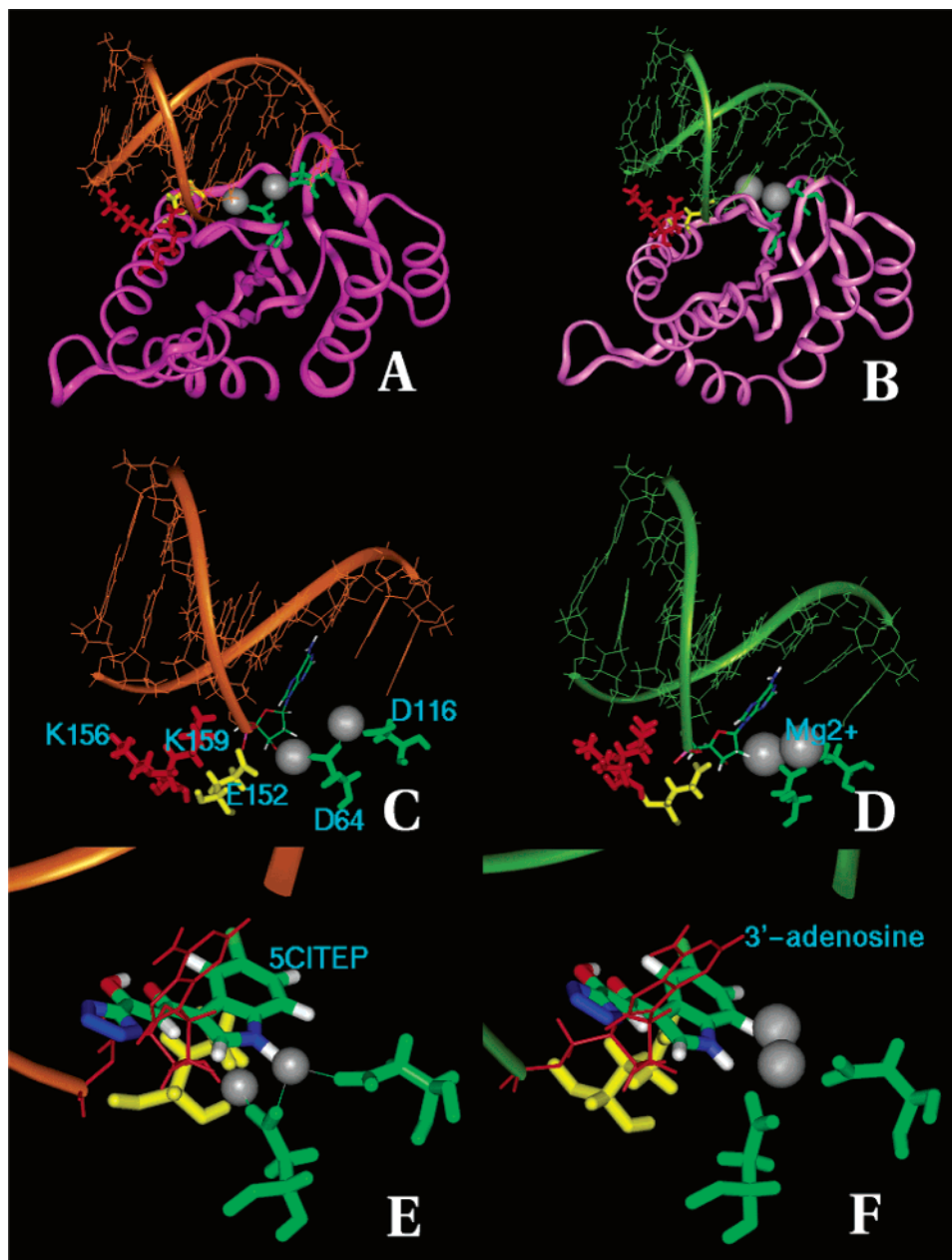


Figure 2. (A) Docked IN/viral DNA complex featuring the IN active site residues K156 and K159. The recessed 3'-hydroxyl of the 3'-processed viral DNA is positioned near the Mg^{2+} ion coordinated by D64. (B) Energy-minimized model of docked IN/viral DNA complex. (C) Detailed representation of docked viral DNA in relation to the IN active site residues. Viral DNA is seen to be positioned such that the 3'-hydroxyl of the conserved deoxyadenosine is located near the active site residues. (D) Energy-minimized docked viral DNA in relation to the active site residues of IN. (E) Detailed view of the docked viral DNA superimposed on a 2 ns MD simulation of the IN/5CITEP complex. The inhibitor is shown to overlap with the conserved deoxyadenosine. (F) View of 5CITEP overlapping with the energy-minimized, docked processed viral DNA.

which was encouraged by addition of the van der Waals term. The van der Waals term contributed about -13 kcal/mol to the total energy for 30 top-ranked configurations. Lower ranked solutions with less favorable van der Waals energy were translated more over the active site, like those in Figure 1D.

Discussion

Available structures of the IN active site region consist of a flexible loop comprising important residues such as E152 and Y143, the last of which has been shown to play a secondary role in catalysis.^{4,37,38} This loop region has been difficult to characterize, however, because of relatively high experimental X-ray B factors

for most residues in this region.¹³ IN has been observed to be receptive to very diverse ligands,³⁸ which might be due to the absence of any clefts on the IN surface. Instead, the active site is a comparatively large and shallow substrate binding region on the surface of the enzyme.³⁵ The docking studies presented herein aim to help shed more light on IN/substrate interactions. To achieve this, we need to assess the structures and the computational methods upon which our studies are based.

Molecular Structure. The IN structure we used in this study was obtained from a prior molecular dynamics simulation.^{5,6} It represents a realistic platform because it includes both Mg^{2+} ions required for catalysis,

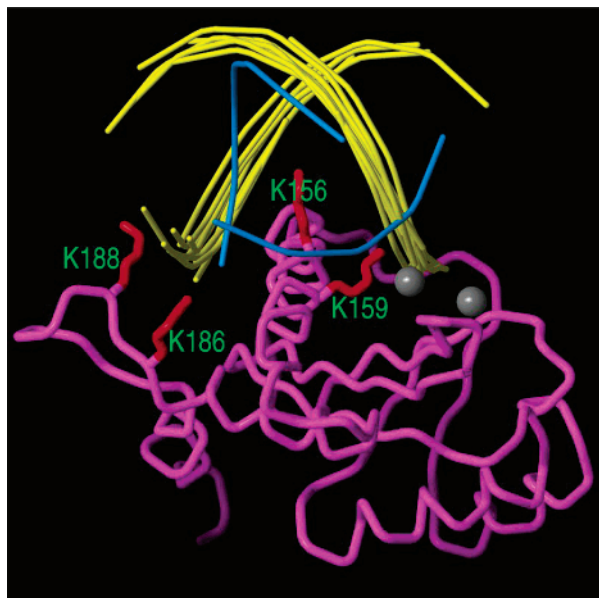


Figure 3. Model for bound target DNA. The most favorable configurations found by DOT as ranked by the sum of electrostatic and van der Waals energies are centered over K156 rather than over the catalytic site metal ions. Nine of the top 10 DNA solutions have closely aligned phosphate backbones (shown as yellow tubes). The DNA major groove lies over the helix containing K156, and the phosphate backbone is between K186 and K188 (left). One of the top 10 (blue, rank 7) is also centered over K156 but in a different orientation. View is as in Figure 1D.

whereas the crystallographic structure upon which it was based has only one active site metal ion. Furthermore, the molecular dynamics structure provides a model for the disordered loop and includes a substrate fragment (HPO_4^{2-}) that mimics the viral DNA backbone. This substrate was not included as part of the IN structure in these docking studies but is useful for comparison with our predicted position of bound viral DNA.

Docking Method. The DOT program performs a complete translational and rotational search between two molecules and is highly efficient on parallel computers. The computational time is not dependent on the size of the molecules, allowing interactions between macromolecules of any size to be examined. The potentials used to describe the molecules are intentionally approximate to account for some induced fit upon complex formation, which is not allowed in the DOT studies because each molecule is held internally rigid. The DOT program was designed²¹ and successfully used^{22,23} to examine macromolecular interactions for which electrostatic forces are important. This makes it particularly appropriate for studies of DNA/protein interactions.

Computational docking of DNA fragments to the IN catalytic domain has also been performed with the Autodock program.⁴³ Autodock allows for some internal flexibility; however, the computational time is dependent on the size and number of flexible torsion angles in the moving molecule. This is particularly problematic with DNA, which has a large number of flexible torsional angles for each base pair. This effectively limits the size of DNA docking with Autodock to a dinucleotide.

Given the absence of structural data, the importance of HIV IN as a drug target, and the complexity of the

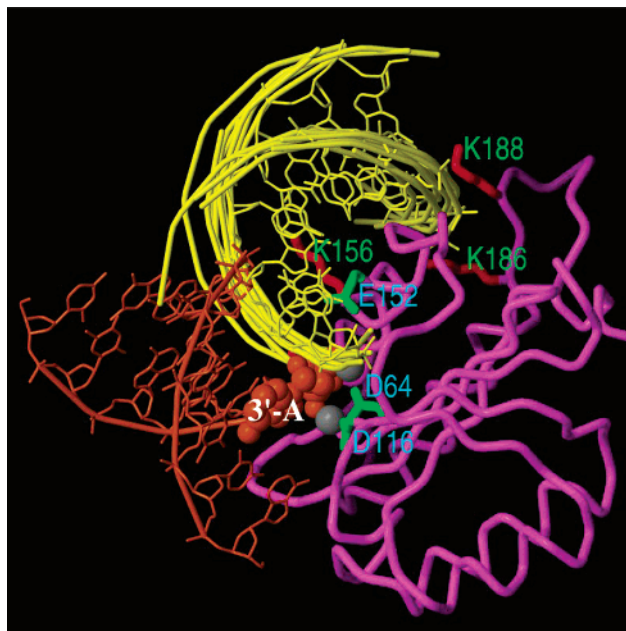


Figure 4. Model for the complex of the IN catalytic domain with bound viral and target DNA. The viral DNA structure shown in Figure 2 (orange, with conserved 3'-adenosine shown by spheres) and the target DNA (yellow, tight cluster shown in Figure 3 with complete structure shown for the top-ranked solution) intersect over the catalytic site of IN.

IN structure and reaction, computational approaches are currently the best way to provide structural models for IN/DNA complexes. Although each type of docking algorithm has its own strengths and weaknesses, each can contribute to understanding IN/DNA interactions. Furthermore, different techniques can complement each other, providing a more complete picture.

Model for Viral and Target DNA Binding to the IN Catalytic Domain. Information about IN/DNA binding has so far been limited to mainly photo-cross-linking and mutagenesis studies that have identified residues in contact with viral DNA.^{18–20,34} Mutagenesis studies show that K156 and K159 are critical for IN/DNA binding.¹⁹ Furthermore, photo-cross-linking studies also suggest the close interaction of the conserved adenosine with K159.¹⁹ Mutation of either residue resulted in the loss of both 3'-processing and strand transfer activities *in vitro*.¹⁹ Hence, our analysis of the DOT searches for viral DNA focused on DNA/IN complexes with possible close interactions with K156 and K159.

Two distinct modes of DNA binding near the catalytic site were found, providing a model for simultaneously bound viral and target DNA to the IN catalytic domain (Figure 4). This model is consistent with a model proposed from structural and cross-linking studies and with a model of the complete IN complex based on fluorescence anisotropy.⁴¹ Our target DNA position also agrees well with computational docking studies of dinucleotide binding.⁴³ The DOT result shows how the many dinucleotide phosphate docking sites over K156, K186, and K188 can be accommodated by intact double-stranded DNA. The site for viral DNA binding suggested by a recent dinucleotide docking study⁴³ appears less consistent with our model and other proposed models.^{41,34} It may be that the small dinucleotide fragment

finds some positively charged regions on the IN surface that cannot accommodate a continuous strand of DNA. There was no cluster of DNA configurations in the 2000 top-ranked solutions found by DOT over residues K111 and K136, which are proposed as part of the viral DNA site by the dinucleotide studies. On the other hand, their predicted site could represent a portion of a DNA-binding site that also uses other domains from the IN multimeric complex, and therefore, this region was ranked more poorly by DOT than the DNA-binding sites near the catalytic site.

Comparison of Predicted Viral DNA Interactions and Experimental Data. Since the viral DNA sequence is highly conserved (unlike that for the target DNA sequence), inhibitor design for this site should be more relevant than that for the target DNA binding site. Therefore, we focused on refining IN/viral interactions by energy minimization. The optimized IN/viral DNA complex featured the O1P of the 3'-adenosine in the A(5) position in a hydrogen-bonded interaction with residue K156. The minimized IN/viral DNA complex also identified possible interactions between the NZ atom of K159 and the H4' atom of the sugar ring of CYT in the other DNA strand.

E152 constitutes part of the D,D-35-E motif common to all retroviruses, retrotransposons, and some bacterial transposases.²⁻⁴ Experiments have implicated E152 in binding to the conserved deoxyadenosine.⁴⁰ Minimized results of the docked IN/processed viral DNA complex shows hydrogen bonding between E152 and the O3' of the conserved deoxyadenosine, complementing previous experimental studies. This processed viral DNA had the CA 3' end in a favorable configuration, oriented toward the active site with hydrogen bonds and van der Waals interactions observed with significant residues. This is consistent with studies that have shown that viral DNA ends are intimately involved in integration.⁷ We also observed hydrogen-bonding interactions between the optimized viral DNA and D116, another key residue of the active site. D116 has also been identified by recent docking simulations to form hydrogen bonds with the inhibitors 5CITEP and 1,2,5,8-tetrahydroxyanthraquinone (quinalizarin, QLZ).⁴²

The optimized viral DNA/IN docked complex also identified hydrogen-bonding interactions with other residues that have been observed to be in contact with inhibitors such as E92, H67, and T66. A hydrogen-bonding interaction was also observed between processed viral DNA and E69. This is worth noting because this residue is yet unexplored and may provide a further step in the development of new strategies to target HIV integration.

Photo-cross-linking results from Esposito et al.²⁸ suggest that with the conserved adenosine binding in the vicinity of K159 and E152, the 5'-cytosine overhang interacting with Q62, and the 5'-adenosine overhang should be in contact with Q148. However, in our optimized docked structure, the 5'-adenosine is about 8 Å away from Q148. Because IN is active only in the multimer form, we suggest that the multimer may play a role in an *in trans* stabilization of viral DNA using residues Q148 and Q62.

Studies on the IN strand transfer reaction have shown that the presence of the divalent cations is

required for covalent joining involving the recessed 3'-hydroxyl ends of viral DNA to 5'-phosphates of host DNA. The calculated complex from our docking simulations complement the above hypothesis, since we identified the 3'-hydroxyl of the conserved deoxyadenosine to be within 2.8 Å of D64 and 0.9 Å of the Mg²⁺ coordinated by D64.

Superimposition of the 3'-processed viral DNA/IN complex on the IN/HPO₄²⁻ structure shows that the modeled phosphate is in the general vicinity of the position occupied by the phosphate of the 3'-guanine prior to the 3'-processing of the GT dinucleotide. The positioning of the phosphate with respect to the predicted DNA configuration suggests the possibility of a 3',5' exonuclease reaction that has been suggested to be the governing mechanism for IN action.⁴⁴ The mechanism, and our model, places the 3'-OH of the viral DNA in proximity to a Mg²⁺ ion in the active site of the integrase. This metal ion has the role of stabilizing the 3'-OH terminus, making it easier to deprotonate and participate in the nucleophilic attack on a P atom of the host DNA.

A superimposition of the unprocessed and 3'-processed viral DNA forms on the IN/5CITEP inhibited complex revealed inhibitor overlap with the conserved deoxyadenosine, suggesting a steric blocking mode of action of the inhibitor that prevents substrate binding. This complements results from docking and superimposition studies carried out by Sottriffer et al., which reveal a considerable overlap between the adenine of docked dAMP and chloroindole of docked 5CITEP.⁴² This result would make sense because some inhibitors such as 5CITEP are nucleotide mimics.⁴²

Conclusion

The results from the docking of viral DNA to the IN catalytic domain by means of the DOT docking program provide the first detailed molecular binding model of an IN/DNA complex consistent with experimental data. A coherent picture of DNA interactions of the viral DNA terminal end with IN active site residues is presented, which should prove to be helpful in future inhibitor design efforts. The minimized docked complex reveals hydrogen-bonding interactions between the 3'-processed DNA terminal end and catalytic residues such as D116 and E152. In addition, the docked configuration identified an overlap between an inhibitor and viral DNA nucleotides, providing an initial insight into the inhibitory mechanism of 5CITEP.

Acknowledgment. The authors thank the San Diego Supercomputer Center and Prof. Lynn Ten Eyck for the DOT program and Drs. C. Sottriffer and H. Ni for providing an IN/5CITEP trajectory. We also thank Dr. Maria Letizia Barreca for useful advice and careful reading of the manuscript. This project was supported by the NIH Program on Structural Biology of AIDS Related Proteins (Grant GM56553). V.A.R. received support from the NSF (Grant DBI 99-04559). Gratitude is also expressed to the National Resource Allocation Committee for a grant of supercomputer time to J.M.B. Additional computing, data storage, and visualization resources were provided by the Institute for Molecular Design (IMD) and the Texas Learning and Computation Center at the University of Houston (UH). Gratitude is

also expressed to Accelrys, Inc. for software licenses provided to the IMD at UH.

References

- (1) Andrade, M. D.; Skalka, A. M. Retroviral integrase, putting the pieces together. *J. Biol. Chem.* **1996**, *271*, 19633–19636.
- (2) Kulosky, J.; Jones, K. S.; Katz, R. A.; Mack, J. P. G.; Skalka, A. M. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposons, integrases and bacterial insertion sequence transposons. *Mol. Cell. Biol.* **1992**, *12*, 2331–2338.
- (3) Engelman, A.; Cragie, R. Identification of amino acid residues critical for human immunodeficiency virus type 1 protein *in vitro*. *J. Virol.* **1992**, *66*, 6361–6369.
- (4) van Gent, D. C.; Groeneger, A. A. M. O.; Plasterk, R. H. A. Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Res.* **1993**, *21*, 3373–3377.
- (5) Lins, R. D.; Briggs, J. M.; Straatsma, T. P.; Carlson, H. A.; Greenwald, J.; Choe, S.; McCammon, J. A.; Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys. J.* **1999**, *76*, 2999–3011.
- (6) 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 electrostatic calculations. *Pharmacol. Ther.* **2000**, *85*, 123–131.
- (7) Katz, R. A.; Skalka, A. M. The retroviral enzymes. *Annu. Rev. Biochem.* **1994**, *63*, 133–173.
- (8) Farnet, C. M.; Bushman, F. D. HIV cDNA integration: molecular biology and inhibitor development. *AIDS* **1996**, *10* (Suppl. A), S3–S11.
- (9) Asante-Appiah, E.; Skalka, A. M. Molecular mechanisms in retrovirus DNA integration. *Antiviral Res.* **1997**, *36*, 139–156.
- (10) Rice, P.; Cragie, R.; Davies, D. R. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **1996**, *6*, 76–83.
- (11) Pommier, Y.; Pilon, A. A.; Bajaj, K.; Mazunder, A.; Neamati, N. HIV integrase as a target for antiviral. *Antiviral Chem. Chemother.* **1997**, *8*, 463–483.
- (12) Chow, S. A.; Vincent, K. A.; Ellison, V.; Brown, P. O. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* **1992**, *255*, 723–726.
- (13) Nicklaus, M. C.; Neamati, N.; Hong, H.; Mazunder, A.; Sunder, S.; Chen, J.; Milne, G. W. A.; Pommier, Y. HIV-1 integrase pharmacophore discovery of inhibitors through three-dimensional database searching. *J. Med. Chem.* **1997**, *40*, 920–929.
- (14) Hong, H.; Neamati, N.; Wang, S.; Nicklaus, M. C.; Mazunder, A.; Zhao, H.; Burke, T. R.; Pommier, Y.; Milne, G. W. A. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. *J. Med. Chem.* **1997**, *40*, 930–936.
- (15) Neamati, N.; Hong, H.; Mazunder, A.; Wang, S.; Sunder, S.; Nicklaus, M. C.; Milne, G. W. A.; Proska, B.; Pommier, Y. Depsides and depsidones as inhibitors of HIV-1 integrase: discovery of novel inhibitors through 3D database searching. *J. Med. Chem.* **1997**, *40*, 942–995.
- (16) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Development of a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.
- (17) Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Methods for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* **1999**, *103*, 10213–10219.
- (18) Huer, T. S.; Brown, P. O. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active site enzyme–DNA complex by photo-crosslinking. *Biochemistry* **1998**, *36*, 10655–10665.
- (19) Jenkins, T. M.; Esposito, D.; Engelman, A.; Cragie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.* **1997**, *16*, 6849–6859.
- (20) Heuer, T. S.; Brown, P. O. Photo-crosslinking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase–DNA complex. *Biochemistry* **1998**, *37*, 6667–6678.
- (21) Ten Eyck, L. F.; Mandell, J.; Roberts, V. A.; Pique, M. E. *Proceedings of the 1995 ACM/IEEE Supercomputing Conference*, San Diego, December 3–8, 1995; Hayes, A., Simmons, M., Eds.; IEEE Computer Society Press: Los Alamitos, CA, 1995. http://www.supercomp.org/sc95/proceedings/636_LTEN/SC95.HTM
- (22) Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovoy, V.; Mitchell, J.; Nelson, E.; Tsilgeny, I.; Ten Eyck, L. F. Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* **2001**, *14*, 105–113.
- (23) Roberts, V. A.; Pique, M. E. Definition of the interaction domain for cytochrome *c* on cytochrome *c* oxidase: Prediction of the docked complex by a complete systematic search. *J. Biol. Chem.* **1999**, *274*, 38051–38060.
- (24) Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* **1996**, *6*, 402–406.
- (25) Bamborough, P.; Cohen, F. E. Modeling protein–ligand complexes. *Curr. Opin. Struct. Biol.* **1996**, *6*, 236–241.
- (26) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; McCammon, J. A. Electrostatics and diffusion of molecules in solution—simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* **1995**, *62*, 187–197.
- (27) Katzman, M.; Katz, R. Substrate recognition by retroviral integrases. *Adv. Virus Res.* **1999**, *52*, 371–395.
- (28) Esposito, D.; Cragie, R. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein–DNA interaction. *EMBO J.* **1998**, *17*, 5832–5843.
- (29) *InsightII*; Accelrys, Inc.: San Diego, CA (<http://www.accelrys.com>).
- (30) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5*; University of California: San Francisco, 1997.
- (31) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Cragie, R.; Davies, D. R. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9150–9154.
- (32) *DOT User's Guide*, version 1.0 alpha; The Computational Center for Macromolecular Structures, San Diego Supercomputer Center: San Diego, CA, 1998.
- (33) *DISCOVER*; Accelrys, Inc.: San Diego, CA (<http://www.accelrys.com>).
- (34) Gao, K.; Butler, S. L.; Bushman, F. Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J.* **2001**, *20*, 3565–3576.
- (35) Goldgur, Y.; Cragie, R.; Cohen, G. H.; Fujiwara, T.; Yoshinaga, 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.
- (36) Ni, H.; Sotriffer, C.; McCammon, J. A. Ordered water and ligand mobility in the HIV-1 integrase 5-CITEP complex: A molecular dynamics study. *J. Med. Chem.* **2001**, *44*, 3043–3047.
- (37) Lutzke, R. A. P.; Plasterk, R. H. A. Structure-based mutation analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and protein binding. *J. Virol.* **1998**, *72*, 4841–4848.
- (38) Lutzke, R. A. P.; Vink, C.; Plasterk, R. H. A. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.* **1994**, *22*, 4125–4131.
- (39) Pommier, Y.; Neamati, N. Inhibitors of human immunodeficiency virus integrase. *Adv. Virus Res.* **1999**, *52*, 427–458.
- (40) Gerton, J. L.; Brown, P. O. The core domain of HIV-1 integrase and viral DNA substrates. *J. Biol. Chem.* **1997**, *272*, 25809–25815.
- (41) Podtelezchnikov, A. A.; Gao, K.; Bushman, F. D.; McCammon, J. A. Modeling HIV-1 integrase complexes based on their hydrodynamic properties. *Biopolymers* **2002**, *68*, 110–120.
- (42) Sotriffer, C. A.; Ni, H.; McCammon, J. A. Active site binding mode of HIV-1 integrase inhibitors. *J. Med. Chem.* **2000**, *43*, 4109–4117.
- (43) Perryman, A. L.; McCammon, J. A. Autodocking dinucleotides to the HIV-1 integrase core domain: Exploring possible binding sites for viral and genomic DNA. *J. Med. Chem.* **2002**, *45*, 5624–5627.
- (44) Beese, L. S.; Steitz, T. A. Structural basis for the 3′-5′-exonuclease activity of *Escherichia coli* DNA polymerase I. *EMBO J.* **1991**, *10*, 25–23.

JM0301890