Molecular Dynamics Studies of the Wild-Type and Double Mutant HIV-1 Integrase Complexed with the 5CITEP Inhibitor: Mechanism for Inhibition and Drug Resistance

Maria L. Barreca,* Keun Woo Lee,† Alba Chimirri,* and James M. Briggs†
*Dipartimento Farmaco-Chimico, Università di Messina, 98168 Messina, Italy; and †Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001 USA

ABSTRACT The human immunodeficiency virus type 1 (HIV-1) integrase (IN) is an essential enzyme in the life cycle of the virus and is an attractive target for the development of new drugs useful in acquired immunodeficiency syndrome multidrug therapy. Starting from the crystal structure of the 5CITEP inhibitor bound to the active site in the catalytic domain of the HIV-1 IN, two different molecular dynamics simulations in water have been carried out. In the first simulation the wild-type IN was used, whereas in the second one the double mutation T66I/M154I, described to lead to drug resistance, was introduced in the protein. Compelling differences have been observed in these two structures during analyses of the molecular dynamics trajectories, particularly in the inhibitor binding modes and in the conformational flexibility of the loop (residues 138–149) located near the three catalytic residues in the active site (Asp^{64}, Asp^{116}, Glu^{152}). Because the conformational flexibility of this region is important for efficient biological activity and its behavior is quite different in the two models, we suggest a hypothetical mechanism for the inhibition and drug resistance of HIV-1 IN. These results can be useful for the rational design of more potent and selective integrase inhibitors and may allow for the design of inhibitors that will be more robust against known resistance mutations.

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) is the etiological agent of the acquired immunodeficiency syndrome (AIDS). Enzymes essential for the replication cycle of this virus, such as reverse transcriptase (RT), protease (PR), and integrase (IN) are important targets for the development of anti-AIDS drugs. Although effective inhibitors have been developed against RT and PR with several of them now being marketed as anti-AIDS agents, no drug active against HIV-1 integrase has been approved by the U.S. Food and Drug Administration. Because viral genome integration is essential to the replication of the retroviruses and, unlike PR and RT, IN has no known functional analog in human cells, the development of drugs able to inhibit one of the various steps involved in integration should be very useful in AIDS multidrug therapy (De Clercq, 2000).

The HIV-1 IN enzyme, which belongs to the superfamily of polynucleotidyltransferases, inserts a double-stranded DNA copy of the viral RNA genome into the chromosomes of an infected cell through two separate reactions (Engelman et al., 1991). In the first hydrolytic step, termed ''3′-processing,'' IN removes two nucleotides from each viral cDNA end adjacent to a conserved 3′-CA sequence leading to the formation of a new recessed 3′-CA-OH end. In the second reaction, called ''strand transfer'' or ''transesterification,'' the two newly processed 3′-viral DNA ends are inserted into opposite strands across a five basepair stretch of host target DNA. Two separate active sites (i.e., two distinct IN proteins) are involved in the simultaneous double strand transfer. The product of this step is a gapped intermediate product in which the 5′-phosphate ends of the viral DNA are no longer linked to the 3′-OH ends of the host DNA. Both reactions show a direct nucleophilic attack by a hydroxyl group. IN uses an activated water molecule as the nucleophile in the 3′-end processing, while in the strand transfer the newly exposed 3′-CA-OH group from the viral DNA is the nucleophile that attacks the host DNA backbone. The integration process is completed by cleavage of the unpaired dinucleotides from the 5′-ends of the viral DNA and repair of the gaps between the viral and target DNA. Although IN may be involved in these repair reactions, it is not necessary because the host cell already has the machinery to carry out such processes. In vitro, integrase can also carry out an apparent reversal of the strand transfer reaction, referred to as ''disintegration'' (Chow et al., 1992). For the integration reaction, no source of energy (e.g., no ATP) is needed and only divalent cations such as Mn^{2+} or Mg^{2+} are required for the catalytic activity (Asante-Appiah and Skalka, 1999; Wlodawer, 1999).

Retroviral IN is a 32-kDa enzyme (288 residues) encoded by the *pol* gene and is composed of one polypeptide chain that folds into three distinct functional domains: the N-terminal domain (residues 1–50), the catalytic core domain (residues 50–212), and the C-terminal domain (residues 212–288). The amino-terminal domain contains a conserved ''HH-CC'' motif that binds a Zn^{2+} ion and promotes enzyme multimerization (Zheng et al., 1996; Lee et al., 1997). The catalytic domain is composed of a mixed α-helix and β-sheet motif and contains an absolutely conserved D,D-35-E motif characterized by three acidic residues, Asp^{64}, Asp^{116}, and

Submitted March 1, 2002, and accepted for publication August 28, 2002.
Address reprint requests to Dr. James Briggs, Dept. of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001. Tel.: 713-743-8366; Fax: 713-743-8351; E-mail: jbriggs@uh.edu.

0006-3495/03/03/1450/14 $2.00

Glu^{152}; the last two residues are separated by 35 amino acids (Engelman and Craigie, 1992; Kulkosky et al., 1992; Polard and Chandler, 1995). The C-terminal domain has been shown to have a nonspecific but strong DNA binding activity similar to that of the full-length IN (Engelman et al., 1994; Vink et al., 1993). All three domains bind DNA, and each isolated domain forms a homodimer in solution. Even though all three domains are required for full catalytic activity, site-directed mutagenesis experiments have shown that the central core domain is sufficient to promote a reverse integration reaction in vitro, known as disintegration, indicating that this region contains the enzymatic catalytic center (Chow et al., 1992; Bushman et al., 1993). The structures of the three separate domains have been solved by x-ray crystallography or NMR spectroscopy (Dyda et al., 1994; Bujacz et al., 1996; Maignan et al., 1998; Goldgur et al., 1998; Greenwald et al., 1999; Eijkelenboom et al., 1995; Eijkelenboom et al., 1999; Lodi et al., 1995; Cai et al., 1997; 1998). The first crystal structure of the catalytic core domain did not reveal any bound metal ion in the active site; and a 13-residue loop and helix bounded by residues 140 and 154, which includes the third conserved amino acid Glu^{152}, was not resolved (Dyda et al., 1994).

Later, complete structures of this IN domain have been reported, but the previously unresolved loop/helix close to the active center is still not well defined because of high temperature factors, suggesting that this region is either particularly flexible or disordered in the crystal. Cross-linking studies of IN with DNA revealed that residues in this region (139–152) could be cross-linked to DNA close to the reactive phosphate (i.e., the scissile bond), placing the loop near the substrate during the reaction (Heuer and Brown, 1997; Esposito and Craigie, 1998). The loop may hold the DNA substrates in proper position while the active site carries out the integration reaction. In addition, mutation studies have shown that Tyr^{143} is involved in the integration reactions by hydrogen-bonding thereby stabilizing and directing the nucleophiles for attack (van Gent et al., 1993; Esposito and Craigie, 1998; Beese and Steitz, 1991; Chen et al., 2000). Unfortunately, the active conformation adopted by the surface loop during the integration reaction is unknown; however, it is evident that its mobility is correlated with the catalytic activity. Single and double loop hinge mutations (G140A, G149A, G140A/G149A) have reduced or abolished activity (Greenwald et al., 1999). X-ray structures are available for some of these loop hinge mutants (Greenwald et al., 1999).

Beese and Steitz proposed that, by analogy to the 3′-5′ exonuclease of *Escherichia coli* DNA polymerase I, the three conserved acidic residues in the retroviral integrases coordinate two metal ion cofactors, forming a template for DNA binding and catalysis (Beese and Steitz, 1991). However, the HIV-1 IN structures available to date contain only a single divalent metal ion bound to the catalytic core domain, while the location of the binding site for the second metal ion is not yet defined, leading to the thought that the second metal cofactor may bind together with a DNA substrate. On the other hand, structures of the avian sarcoma virus (ASV) IN catalytic domain are available with one or two noncatalytic divalent metal ions bound (Bujacz et al., 1997).

Recently, the first HIV-1 IN two-domain structure (PDB code 1EX4, residues 52–288) has been solved by x-ray crystallography (Chen et al., 2000). Two different crystal structures of a ligand bound to the HIV-1 IN have been reported. The first HIV-1 IN inhibitor complex (PDB code 1QS4) was solved by Goldgur et al. (Goldgur et al., 1999) and shows a novel Shionogi integrase inhibitor 5CITEP in complex with the HIV-1 IN catalytic domain. Furthermore, in 2001 a small-molecule binding site was identified at the dimer interface of the HIV-1 IN enzyme (PDB code 1HYV and 1HYZ) (Molteni et al., 2001).

However, the fully bioactive structure of the HIV-1 IN in vivo is still unknown. Numerous studies suggest that the smallest multimer with full catalytic activity is a tetramer or an octamer (Cai et al., 1997; Heuer and Brown, 1998). In fact, the full nuclear preintegration complex contains 50–100 copies of integrase along with structural proteins and double-stranded DNA (Bukrinsky et al., 1992; Bowerman et al., 1989). Site-directed mutagenesis and photo-crosslinking studies have identified several residues near the active site that are critical for binding viral DNA (Heuer and Brown, 1997; Esposito and Craigie, 1998; Jenkins et al., 1997).

Recently, Hazuda et al. have presented a new class of integrase inhibitors with a diketo acid functionality (Hazuda et al., 2000). Among them, L-731,988 and L-708,906 are two of the most active molecules and exhibit the unique ability to inhibit the strand transfer reaction. Individual and combined mutations in HIV-1 IN at Thr^{66}, Met^{153}, and Met^{154} have been reported to confer some degree of resistance to one or more inhibitors.

Using the x-ray structure of the 5CITEP inhibitor bound to the IN catalytic core domain as the starting point, we have carried out two different molecular dynamics (MD) simulations. In the first simulation, the wild-type IN was used for our studies, while in the second one the double mutation T66I/M154I was introduced in the protein. The 5CITEP inhibitor was present in the structures of both simulations.

The aims of the present work were to 1), understand the relationship between ligand-IN complex formation and functional response; 2), obtain evidence of the role of the loop close to the active site in the catalytic activity; 3), hypothesize the molecular mechanisms for HIV-1 IN drug resistance; and finally 4), obtain information for receptor-based drug design.

We observed substantial differences between these two structures during the MD simulations, particularly in the inhibitor binding mode and in the structure and dynamics of the surface loop located near the three catalytic residues in

the active site (Asp^{64}, Asp^{116}, Glu^{152}). On this basis, our ongoing work is the rational design and testing of new HIV-1 IN inhibitors.

COMPUTATIONAL METHODS

Structures

Protein

The x-ray crystallographic structure of the HIV-1 IN catalytic core domain (subunit A, residues 56–140, 145–209) complexed with the Shionogi inhibitor 5CITEP (PDB code 1QS4) (Goldgur et al., 1999) was used as the initial structure in both simulations. This crystal structure contains one Mg^{2+} coordinated by Asp^{64} and Asp^{116} and has some unresolved residues in the flexible loop Gly^{140}–Gly^{149} adjacent to the active site, connecting helices α4 and α5. The unresolved region (residues 141–144) was manually completed based on the conformation of the homologous loop in the crystal structure of the uncomplexed core domain of HIV-1 integrase 1BIS (Goldgur et al., 1998). Residue Tyr^{143} in the loop, which has been proposed to be involved in the catalytic mechanism, was modeled based on its conformation in monomer B of 1QS4, in which it is pointed toward the active site. All hydrogens were added to the proteins according to previous electrostatically based pK_a predictions (Lins et al., 1999) using the UHBD program (Madura et al., 1995), where no unusual amino acid ionization states were predicted (i.e., Asp^{-1}, Glu^{-1}, Arg^{+1}, His^{0}, Lys^{+1}, C-$Term^{-1}$, N-$Term^{+1}$).

The region including residues 139–146 was minimized using 100 steps of steepest-descent minimization employing a distance-dependent dielectric and the Discover/CVFF force field in the InsightII package (InsightII, 2000). Residues Thr^{66} and Met^{154} were each computationally replaced with Ile to obtain the double mutant system using the InsightII/Biopolymer module. The starting orientations of the new Ile side chains (Fig. 1) were chosen after consideration of the physical environment of the two residues, trying to initially avoid steric clashes.

Ligand

The starting structure of 5CITEP was that present in the complex with the HIV-1 IN (PDB code 1QS4). The semiempirical PM3 method was used to calculate the partial atomic charges of the inhibitor (Table 1).

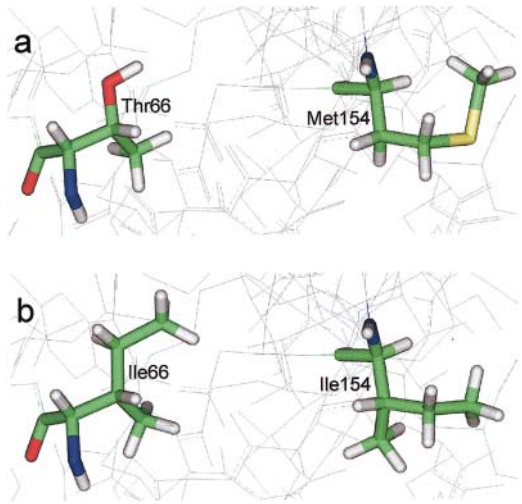

FIGURE 1 Orientation of residues 66 and 154 (*a*) in the crystal structure and (*b*) in the initial model of T66I/M154I HIV-1 IN complex.

TABLE 1 Atom types and charges for 5CITEP

Atom name	Atom type	Charge
C1	CA	−0.087
C2	CA	−0.058
C3	CN	−0.246
C4	CB	−0.030
C5	CA	−0.022
C6	CA	−0.120
N7	NA	0.422
C8	CW	−0.188
C9	C*	−0.229
C10	C	0.407
C11	CM	−0.245
C12	C	0.306
C13	CC	−0.040
N14	NB	−0.141
N15	NB	−0.046
N16	NB	−0.015
N17	NB	−0.200
O18	O	−0.347
O19	OH	−0.176
CL20	XC	0.075
H1	HA	0.117
H2	HA	0.109
H3	HA	0.128
H4	H	0.069
H5	H4	0.151
H6	HC	0.147
H7	HO	0.259

All of the force field parameters necessary to describe the bonded and nonbonded interactions involving the ligand atoms were kindly provided by Ni et al., 2001. The augmented (i.e., with hydrogens or mutations) wild-type and T66I/M154I IN proteins both complexed with 5CITEP, were used as initial structures for our MD simulations.

Molecular dynamics simulations

Two independent 1.75 ns MD simulations with an equilibration phase of 80 ps each were carried out using the parallel NWChem v.3.3.1 program (Anchell et al., 1999) with the 1995 version of the all-atom AMBER force field (Cornell et al, 1995). The molecular trajectory for the two systems generated by the molecular dynamics simulations were analyzed using the Analysis module in NWChem v.4.0 (Harrison et al., 2000) and the DeCipher module in the InsightII package.

The same MD simulation protocol was applied for both simulations. For our studies we used the NPT ensemble (pressure = 1.025×10^5 Pa) with explicit solvent and periodic boundary conditions. The integration of the equations of motion was carried out using the leap-frog algorithm (Hockney and Eastwood, 1981); all bonds involving hydrogen atoms were constrained to their equilibrium values by means of the SHAKE algorithm (Ryckaert et al., 1977). During the equilibration phase, the velocities were reassigned every 0.5 ps according to a Maxwell-Boltzmann distribution. A short-range spherical cutoff of 10 Å was used for all nonbonded interactions while the smooth particle-mesh Ewald (PME) method (Essmann et al., 1995) was applied to have an adequate treatment of the long-range electrostatic energies and forces.

The complexes were immersed in a cubic box ($72 \times 72 \times 72$ Å) of extended simple point charge (SPC/E) water molecules (Berendsen et al., 1987). Water molecules that were within 2.6 Å of any nonhydrogen atom of the protein were deleted. The solvated systems were neutralized by addition of two chloride ions (whose positions were randomly chosen and were initially far away from the molecular surface of the protein) and then relaxed

by performing 100 steps of steepest descent followed by 1000 steps of conjugate gradient energy minimization to remove the strain in the initial structure before the MD trajectory.

For the simulation with the wild-type IN/5CITEP complex, the entire system consisted of 2403 solute atoms (protein, inhibitor, one Mg^{2+}, and two Cl^- counter ions) and 33,999 solvent atoms, whereas for the double mutant IN/5CITEP system the number of solute and solvent atoms were 2410 and 33,998, respectively.

MD simulation studies consist of equilibration and production phases. In the first stage of equilibration, the solute (protein, counter ions, Mg^{2+}, and ligand) was fixed and the solvent (water molecules) was equilibrated for 30 ps of molecular dynamics at 298 K using an integral time step of 1 fs. Then, the solvent was constrained while the protein was heated from 50 K to 298 K in increments of 50 K, using intervals of 5 ps each with a 1 fs time step. The full (i.e., solute and solvent) systems were subjected to MD at 298 K using a 1 fs time step over 80 ps, during which time the systems achieved stability.

After the equilibration period, the production phase was run for 1.75 ns for each system, using an integral time step of 2 fs with a constant temperature of 298 K kept constant by weak coupling to a heat bath, and separate relaxation times of 0.4 and 0.1 ps were used for the solvent and solute, respectively. The coordinates of the system were saved every 0.1 ps during the simulations for later analyses.

Essential dynamics analyses

Essential dynamics (ED) analyses were performed for both 1.75 ns MD trajectories. Each frame of the trajectory was superimposed onto the equilibrated geometry to remove rotational and translational movements and to isolate only the internal motions of the system. ED analyses were focused on the movement of the 154 C_α atoms of the protein. The Cartesian coordinates for the atoms result in 462 dimensional displacement vectors. The 462×462 covariance matrix was constructed based on the coordinates from the MD trajectories and then diagonalized to obtain its eigenvalues and eigenvectors. Movements in the essential subspace were projected along the most important eigenvectors from the analyses. These overall ED analyses were performed using the WHATIF program (Vriend, 1990).

RESULTS AND DISCUSSION

We have performed two MD simulations, one of the wild-type and one of the double mutant T66I/M154I IN, both complexed with the 5CITEP inhibitor. At the beginning of this work, 5CITEP was the only crystal structure of an inhibitor bound to the active site (PDB code 1QS4), so we decided to use this ligand for our studies to have a good starting model. The integrase inhibitor 5CITEP, 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2*H*-tetrazol-5-yl)-propenone (Fig. 2), is one of a group of Shionogi integrase inhibitors with 50% inhibitory concentrations (IC_{50}) of 2.3 ± 0.1 μM and 2.1 ± 0.1 μM by multiple plate integration assay (MIA) and multiple plate preincubation assay (MPA), respectively (Goldgur et al., 1999).

The inhibitor forms a dimer with another molecule of 5CITEP at a crystallographic dimer interface, so the binding site of this inhibitor in the crystal is rather unlikely to resemble a physiologically relevant binding configuration, as also recently reported in computational studies of 5CITEP optimized binding modes in the absence of the crystallographic environment (Sotriffer et al., 2000).

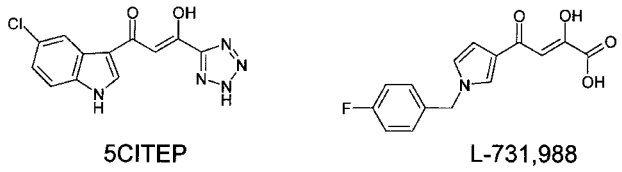

FIGURE 2 5CITEP and L-731,988 HIV-1 IN inhibitors.

Therefore, the 5CITEP/IN complex has been subjected to 1.75 ns of MD in explicit solvent with the aim of gaining information about the bioactive conformation of the ligand within the catalytic site and the mechanism of inhibition.

On the other hand, to understand the effect of drug resistant mutations on the inhibitor binding mode, we performed another MD simulation with the Ile modeled at residues 66 and 154. Even though this double mutation was reported to confer resistance to the diketo acid inhibitors (Hazuda et al., 2000), it is important to note that 5CITEP is also a diketo acid inhibitor and is an analog of L-731,988 (Fig. 2), where the carboxylic group is replaced with one of its typical bioisosteres, e.g., the tetrazole ring. Considering this observation, it is reasonable to think that the same mutations conferring resistance to L-731,988 should have a similar effect on 5CITEP.

Analysis of the dynamics trajectories

Thermodynamic properties such as temperature and energies were monitored during the MD simulations and all converged to stable values. The results presented here refer to the starting structures (i.e., after the equilibration of 80 ps) and average structures for the two MD simulations of wild-type and T66I/M154I inhibitor complexes.

Protein folding of the wild-type and T66I/M154I IN

Using the DSSP module (Kabsch and Sander, 1983) implemented in the ICMlite v.2.7 software (Abagyan et al., 1994), secondary structural analyses were performed for the two average structures over the whole simulation period, and their folds are shown in Fig. 3. Additionally, the HBPLUS software (McDonald and Thornton, 1994) has been used to calculate and compare the hydrogen-bonds in the two proteins. To qualify an interaction as a hydrogen-bond, we have used the default geometric criterion as defined in the program: the donor-acceptor heavy atom distance had to be shorter than 3.9 Å, the maximum distance between a hydrogen atom donor and an acceptor of 2.5 Å, and the donor-hydrogen-acceptor angle had to be larger than 90°.

We compared the fold of our models with the x-ray structure of the catalytic core domain used as the starting point, which consists of a central five-stranded β-sheet with six surrounding α-helices, the same as the crystal forms

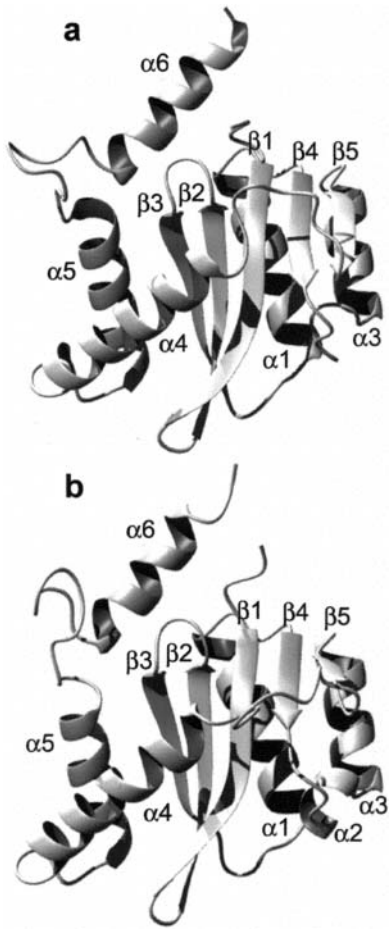

FIGURE 3 Ribbon diagram of the (*a*) wild-type and (*b*) T66I/M154I HIV-1 IN complexes from average structures over the entire simulations. The figures were prepared using the program MOLMOL (Koradi et al., 1996).

obtained to date of the HIV-1 IN catalytic domain. Some notable structural changes were found in our complexes and detailed analyses were performed. In Table 2 we summarize the characteristics of the secondary structure in both the modeled and x-ray crystal structures.

In the wild-type complex (Fig. 3 *a*), the major difference from previously reported unliganded core domain structures is with the lack of helix $\alpha 2$, originally composed of residues Ser^{119} through Thr^{122}. In fact, after the first 200 ps this short helix is no longer present because hydrogen-bonds between Thr^{97}-Asn^{120} and Gly^{118}-Phe^{121} were broken in the early stages of the dynamics simulation. Moreover, in this model the side chains of Asn^{120} and Phe^{121} are not involved in σ-π interactions, because the amine group from Asn^{120} is now located at ~4.0 Å from the plane of the phenyl group, as compared with a distance of ~3.3 Å in the x-ray and T66I/M154I structures (Maignan et al., 1998). Interestingly, the double mutant model (Fig. 3 *b*) has instead the same α-β fold of the initial structure.

The secondary structural analyses reveal a significant difference between the two complexes in the area between residues 136 and 166, which includes the flexible loop close to the active center. In the x-ray structure as well as in the wild-type model, $\beta 5$ consists of four residues (136–139) whereas in the double mutant it is shorter; after the first 300 ps of the production phase for the double mutant Glu^{138} and Phe^{139} are no longer in a $\beta-$strand conformation.

On the other hand, in the wild-type system the long antiparallel helix $\alpha 4$, that forms a part of the active site and contains the catalytic carboxylate Glu^{152}, is defined by residues 149–165, which is different from T66I/M154I that shows $\alpha 4$ from residues 150 to 166, as in the latest published apoenzyme structures. As a result of these different arrangements, the loop in the two models has a different length, consisting of 9 (140–148) and 12 residues (138–149) in the wild-type and double mutant complexes, respectively. Furthermore, the native enzyme contains two additional hydrogen-bonds involving residues in the loop (Gln^{62}-Ser^{147} and Asn^{144}-Gln^{146}), which are not present in the double mutant complex.

This initial information already suggested that in the native protein complex with the drug, the loop near the catalytic residues might be more constrained and less flexible than in the double mutant enzyme complex with the inhibitor. A slight difference was moreover observed in the length of $\beta 4$ that is one residue longer (Thr^{115}) in the wild-type complex with respect to the starting structure and the T66I/M154I complex.

Overall perturbative changes in the two models

The root-mean-square deviations (RMSD) of the C_α atoms with respect to the starting and average coordinates over the two MD simulations are plotted in Fig. 4. It can be seen clearly that after the first 500 ps the wild-type IN is fairly stable during the entire remaining period of MD simulation, with an RMSD value with respect to the starting structure of ~1.3 Å. The T66I/M154I system shows a larger flexibility, and the RMSD increases from 0.6 to 2.3 Å over the first 900 ps of the whole trajectory. After this period, the system stabilizes to around 1.9 Å. Using the average structures as the

TABLE 2 Characteristics of the secondary structure in all three IN complexes

Complex	Sheet 1	Sheet 2	Sheet 3	Helix 1	Sheet 4	Helix 2	Helix 3	Sheet 5	Helix 4	Helix 5	Helix 6
Wild-type IN	60–68	71–78	83–89	94–107	112–115	—	124–132	136–139	149–165	172–185	196–208
T66I/M154I IN	60–68	71–78	83–88	94–107	112–114	119–122	124–132	136–137	150–166	172–183	196–206
X-ray structure	60–68	71–78	83–89	94–105	112–114	119–122	124–132	136–138	150–166	172–185	196–206

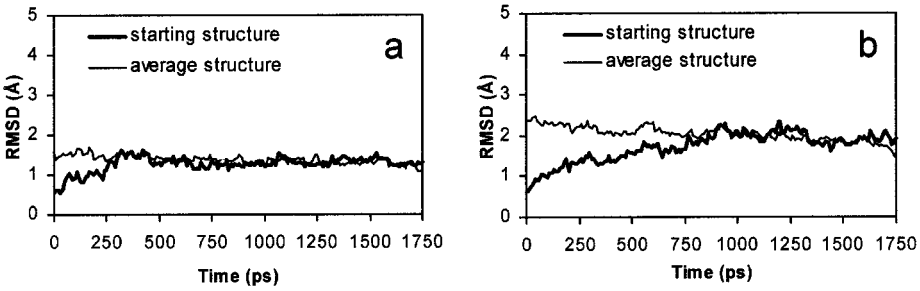

FIGURE 4 Root-mean-square deviation (RMSD) of the C_α atoms for the (*a*) wild-type and (*b*) T66I/M154I IN complexes as a function of simulation time.

reference, the mean RMSD value is 1.3 Å for the wild-type and 2.0 Å for the double mutant complexes, respectively.

The two models exhibit different dynamical behavior; the wild-type displays conformational rigidity higher than that of the native enzyme. Fig. 4 is helpful in conveying the dynamics of the overall structures; however, we were also interested in seeing which parts of the protein were more flexible and affected by the double resistance mutation, so we monitored the root-mean-square fluctuations of the C_α atoms (i.e., somewhat analogous to crystallographic B-factors) for each residue during MD simulations. Fig. 5 compares the calculated B-factors with the experimental ones from the x-ray structure.

Our data indicate that the larger fluctuations are concentrated in two regions corresponding to residues 138–149 and 187–195, which are flexible loops on the surface of the protein as reported in uncomplexed core domain structures. The second loop, 187–195, exhibits almost the same behavior in both simulations except for residue Gly^{189}, which is hydrogen-bonded to Tyr^{194} for the entire length of the simulation of the double mutant complex, whereas this interaction is never present in the wild-type dynamics study. In any event, this region is far away from the active site, so we focused our attention on the loop containing residues 138–149, because this region has been found to adopt several different conformations in previous structures, and its mobility appears to be correlated with IN catalytic activity. Moreover, our secondary structural analyses showed that the length of the loop is not constant in the two complexes,

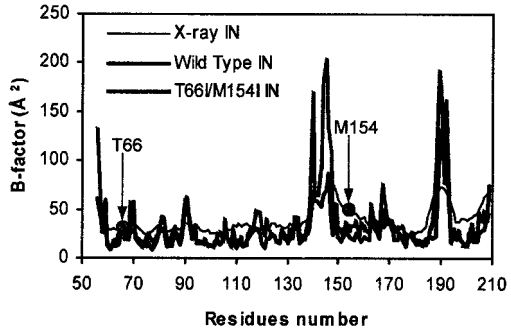

FIGURE 5 Plots of the experimental and calculated B-factors of the C_α atoms as a function of residue number along the chain. The positions of residues 66 and 154 are marked in the original crystal structure.

being shorter in the wild-type system (residues 140–148) in comparison to the double mutant T66I/M154I (residues 138–149).

It is obvious from the plot in Fig. 5 that the loop did not undergo the same conformational changes in the two simulations. The T66I/M154I complex shows larger fluctuations for the residues 138–149 than in the wild-type model indicating that the double mutation affected the dynamics of this region.

This result seemed extremely important to us. Previous molecular, Brownian, and essential dynamics studies of the uncomplexed wild-type IN (Lins et al., 1999; Weber et al., 1998; Lins et al., 2000) showed that the same regions (138–150 and 185–195) were the most flexible parts of the HIV-1 protein. The calculated B-factors for the uncomplexed enzyme simulation (Lins et al., 2000) are higher compared with those of the wild-type IN-5CITEP complex, but similar to the values obtained from the T66I/M154I complex simulation.

This result may explain a portion of the mechanism of inhibition by 5CITEP, i.e., by decreasing the mobility of the active site loop. Of course, 5CITEP bound to the wild-type IN physically blocks substrate access to the active site. It also suggests that the mechanism by which the double mutation leads to inhibitor resistance is by restoring the loop to wild-type flexibility, apparently necessary for catalytic activity. Furthermore, the double mutant complex shows a slightly greater structural fluctuation of residues Gly^{82}, Ser^{119}, Thr^{122}, Gly^{163}, and Asp^{167} compared to the wild-type complex. In addition, the inhibitor in the double mutant complex moves deeper into the binding pocket, apparently no longer blocking substrate binding nor tying up the catalytic residues.

Dynamic behavior of the active site

We have monitored the distance between the fixed center of the catalytic site and the carboxyl groups of Asp^{64}, Asp^{116}, and Glu^{152} as a function of simulation time to gain insights into the conformational changes of the three conserved residues. The center of the catalytic site was defined using a pseudoatom based on the center of mass of all nonhydrogen atoms in the three catalytic residues Asp^{64}, Asp^{116}, and Glu^{152}. The average distances between Asp^{64}, Asp^{116},

and Glu^{152} to the geometric center of the active site are 2.64 Å, 5.48 Å, and 7.76 Å for the wild-type and 3.12 Å, 5.17 Å, and 5.18 Å for the double mutant, respectively. The fluctuations of the Asp^{64} and Asp^{116} are in the same range in both complexes, so we can conclude that these residues are not significantly influenced by the mutations. On the other hand, T66I/M154I has an effect on residue Glu^{152}. Examination of the average structures revealed that the side chain of Glu^{152} faces the general direction of the two catalytic aspartates in the double mutant complex, whereas in the wild-type complex this catalytic residue points away from the two other active carboxylates.

In the wild-type complex the backbone of Glu^{152} is engaged in hydrogen-bonds with the O of Gln^{148} and the N of Lys^{156} over the entire simulation. In the double mutant system, the interaction with Lys^{156} is preserved, while that with Gln^{148} is replaced by Gly^{149}. In both systems the side chain atoms of Glu^{152} are not involved in any intramolecular contacts, but are involved in hydrogen-bonds with solvent molecules. However, in the wild-type complex, atom OE2 of Glu^{152} maintains a water-mediated interaction to the keto oxygen of 5CITEP.

As we already mentioned, in all HIV-1 IN structures reported to date, only a single bound metal ion is found at the active site, which may suggest that the second ion, most likely between the first and third catalytic residues (Asp^{64} and Glu^{152}) as seen in ASV IN structures, binds only in the presence of a DNA substrate.

In both models the magnesium ion is complexed in an octahedral fashion by Asp^{64}, Asp^{116}, and four water molecules, a configuration already observed in native crystal forms. In the wild-type complex these coordinated water molecules are involved in hydrogen-bonds with Asp^{64}, Asp^{116}, Cys^{65}, and Asn^{117}, whereas the last interaction is not present in the double mutant.

We also found that in both average complex structures, the conserved Glu^{152} did not create a proper metal-binding site for the second cofactor, most likely because of the presence of the inhibitor. In fact, in our models the distance between the carboxylate carbons of Asp^{64} and Glu^{152} is 9.98 Å in the wild-type and 8.40 Å in the double mutant, compared with a distance of 8.00 Å in the model coming from an MD simulation of the catalytic domain of the HIV-1 IN containing two Mg^{2+} ions (Lins et al., 2000) and 7.20 Å in the x-ray structure of the catalytic domain of ASV IN with two divalent metal ions in the active site (PDB code 1VSH) (Bujacz et al., 1997).

Conformational changes of the loop close to the active site

It was evident from the results of the secondary structural analyses and the RMS fluctuations for C_α atoms in the wild-type and T66I/M154I simulations that the loop residues 138–149 showed conformational differences between the two complexes. This region therefore needed to be carefully explored in an effort to understand the influence of its local motion on the catalytic mechanism of the HIV-1 IN enzyme.

The RMSD of the loop (C_α atoms) clearly shows the behavior of this region over the entire length of the simulations (Fig. 6). The figure indicates that the conformation of the loop is stable during the course of the MD trajectory of the wild-type complex, showing an average RMSD below 1.5 Å. On the contrary, during the first 900 ps of the double mutant T66I/M154I IN complex simulation, the RMSD increased from 0.6 to 8.7 Å. This large conformational change started at ~300 ps when Glu^{138} and Phe^{139} were no longer in a β-strand conformation and the loop began to widen. From 900 ps to 1350 ps the loop did undergo some structural adjustments, and in the last 400 ps the RMSD fluctuated between 4.6 and 6.8 Å.

These loop transitions were accompanied by changes in hydrogen-bonding. In particular, the principal determinants of the different backbone dynamics in the two systems might be residues Gln^{62} and Gln^{148}. We think that the hydrogen-bonds involving these two residues in the wild-type complex are the main cause of the rigidity of the catalytic loop, anchoring it to other residues in the protein. Residue Gln^{148} is at or just beyond the end of helix $\alpha 4$ that contains Glu^{152}, whereas Gln^{62} is on the opposite side of the active site, two residues away from the catalytic Asp^{64}. In fact, the hydrogen-bonds between Gln^{62}–Gln^{148}, Gln^{148}–Glu^{152}, and Gln^{62}–Ser^{147} are almost always present during the dynamics simulation of the wild-type complex. Furthermore, Gln^{62} has a strategic position in holding the backbone loop in nearly fixed conformation, because it seems to occupy the center of the catalytic site and works as an anchor controlling the mobility of the region 138–149.

In the double mutant enzyme, all of these interactions were disrupted after the first 300 ps of the simulation. Consequently, Gln^{62} participates in a hydrogen-bond with Ile^{151}, whereas Gln^{148} exhibits only water-bridging interactions. However, further analyses were performed with the aim of better comprehending the significance of the loop motions in both systems. The distance between the C_α atoms of Gly^{140} and Gln^{148}, which can be taken as a measure of the

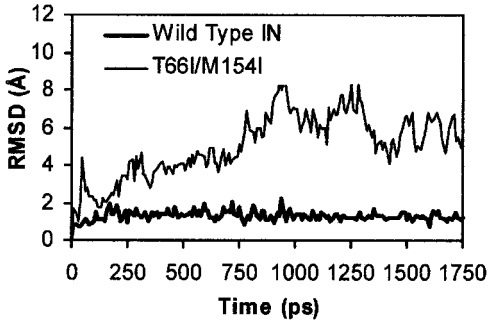

FIGURE 6 RMSD over the trajectory of the C_α atoms of the loop 138–149 taking as the reference frame the starting structure.

width of the active loop, is 11.35 Å and 12.76 Å in the average structures of the wild-type and double mutant complexes, respectively.

In the wild-type complex, a hydrogen-bond interaction was formed in the early stage of the dynamics simulation between Gly^{140} and Thr^{115} and remains throughout the dynamics run, preventing Gly^{140} from moving. At the beginning of the T66I/M154I complex production run, the backbone oxygen of Gly^{140} maintained a hydrogen-bond to NE2 of Gln^{148}; however, at around 300 ps this interaction was lost and Gly^{140} was no longer involved in any hydrogen-bonds over the remainder of the simulation time. Gly^{140} was therefore not ''constrained'' so its backbone was permitted to move, resulting in more loop motion and a wider active site.

The time evolution of the angle between the planes defined for the three catalytic residues and the loop 138–149 over the entire simulation period helped us to reveal the real movements of this region and to quantify how much it moves away or toward the active center.

The wild-type IN maintains an angle value of ~75° indicating that once more the loop in this system has reduced mobility. On the contrary, in the T66I/M154I complex this angle is initially 75° and during the first 250 ps of the simulation it dramatically decreases to 20°, showing that the loop moves toward the catalytic site (after 1500 ps, the average value is 27°).

These results completely confirmed our initial idea that the double drug-resistant mutation restored high (i.e., uncomplexed wild-type) mobility in the catalytic loop whereas the drug constrains the loop in the wild-type.

Role of Tyr^{143} in the catalytic activity

Tyr^{143} in the active site loop has been suggested to play a secondary but important role in the catalytic activity of the integrase (Esposito and Craigie, 1998; Chen et al., 2000; Lins et al., 1999; Tsurutani et al., 2000), so we focused some attention on this residue. Tyr^{143} is located in the disordered loop near the catalytic site and, based on the proposed mechanism for the integration process as being like the Tyr in the 3′-5′ exonuclease activity of *E. coli* polymerase I (Beese and Steitz, 1991), this residue might be involved in the enzymatic activity by stabilizing and properly orienting the nucleophile through a hydrogen-bond interaction, consequently assisting the catalysis of the hydrolytic and phosphoryl transfer reactions.

An important difference in the conformation and dynamics of Tyr^{143} has been observed in the trajectory analyses of the wild-type drug complex and double mutant drug complex simulations. In the wild-type model, the distance between atoms OE1 of Glu^{152} and the OH of Tyr^{143} increases during the first 800 ps from 13 Å to ~21 Å. So, even though the backbone of the loop containing residues 138–149 is largely stable in the wild-type drug complex simulation, the side chain of Tyr^{143} moves away from the active site. Conversely, in the T66I/M154I complex, the loop exhibits high mobility although Tyr^{143} does not move very far from the starting point, maintaining an average distance over the entire simulation of ~10.1 Å. This compares well with the average distance of 10.06 Å calculated for a 1 ns MD simulation of the same model of the catalytic domain of the wild-type IN without inhibitor (unpublished results). The differences in the conformation of the loop and especially with Tyr^{143} between the two models can easily be seen in Fig. 7, where the average structures were superimposed by their backbones.

Our observation that Tyr^{143} has been substantially deformed in the wild-type complex, where a lack of enzymatic activity is seen, supports the hypothesis that its conformation and mobility are important for the integration process and that this residue might be a central player in the mechanism of the integration process. In addition, mutation of Tyr^{143} was shown to affect the ratio between 3′-processing and strand transfer activities (van Gent et al., 1993).

Inhibitor binding mode

The interaction between HIV-1 IN and 5CITEP in the two complexes as well as the influence of the inhibitor on the structural and dynamical properties of the active site region have been clarified by analyzing the trajectory data obtained from the MD simulations.

It is evident from the RMSD of all nonhydrogen atoms of 5CITEP (Fig. 8) that the inhibitor behaves quite differently in the wild-type versus how it behaves in the resistant double mutant, and that surprisingly its binding mode is more stable in the double mutant model. In fact, in the T66I/M154I simulation the inhibitor remains in the same conformation during almost the entire simulation (average RMSD from the initial structure is 1.1 Å), whereas in the wild-type its position did undergo substantial changes, up to a maximum RMSD of 9.0 Å at 980 ps, stabilizing only in the latter part of the trajectory (average RMSD is ~4.5 Å). Other analyses of the trajectory show that the inhibitor does not move in a simple way. The distance between the center of the catalytic

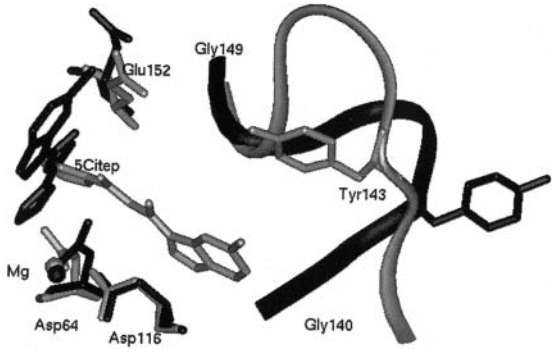

FIGURE 7 Superimposed average structures from the MD simulations of the wild-type (*black*) and double mutant (*gray*) IN.

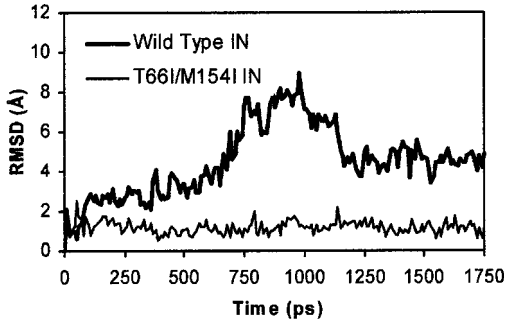

FIGURE 8 RMSD from the initial structure for all nonhydrogen 5CITEP atoms as a function of simulation time.

triad (D64, D116, and E152) and the center of the indole and tetrazole rings of the ligand for the two models as a function of the simulation time (Fig. 9) gave us some insights into the differential behavior of the two portions of the ligand, even though in both simulations the intramolecular hydrogen-bond of the keto-enol group was preserved throughout the dynamics runs. The binding mode of 5CITEP in the two average structures is shown in Figs. 7 and 11.

Wild-type complex

The structure used as the starting point for the production run showed that 5CITEP interacted through hydrogen-bonds with Asp^{64}, Cys^{65}, Thr^{66}, and Asn^{155}.

During the first 200 ps, the indole ring moves a little away from the center of the active site, while the tetrazole portion moves rapidly toward the catalytic triad; in particular, it moves closer to Asp^{64} and Asn^{155}. From 300 ps to ~500 ps of simulation time, the hydrogen-bond involving 5CITEP and Thr^{66} is gradually lost and the chloroindole ring, possibly due to steric interactions with Ile^{141}, Tyr^{143}, and Gln^{148}, starts to move away from the original binding site and toward the solvent. Simultaneously, the keto-enol group moves toward Gln^{148} while the tetrazole portion approaches Asn^{155}. Starting at 560 ps the tetrazole ring shifts the side chain of Asn^{155} from its position (atom ND2 of Asn^{155} moves ~3.5 Å) and slips into a pocket consisting of Asp^{64}, Cys^{65}, Thr^{66}, Val^{75}, Ile^{151}, Glu^{152}, Asn^{155}, and the magnesium ion with its four coordinated water molecules. The inhibitor is at this point almost perpendicular to the plane defined by the three catalytic residues, with the tetrazole portion buried in the pocket, the keto-enol group toward Glu^{152}, and the indole ring solvent-exposed.

Between 500 ps and 1 ns of the dynamics run, occasional hydrogen-bonds occur between 5CITEP and Asp^{64}, Cys^{65}, Asn^{155}, Gln^{148}, and Ile^{151}. It seems to us that because of these forming and breaking bonds, the inhibitor oscillates up and down within its binding site. At ~1100 ps, the indole ring moves closer to Gln^{148} and Glu^{152}, while Asn^{155} is within hydrogen-bonding distance of the enol hydroxyl. Except for the constant presence of the hydrogen-bond with Cys^{65}, the ligand does not show direct interactions with other residues in the protein in the latter phase of the simulation.

According to x-ray analysis, the 5CITEP inhibitor is hydrogen-bonded to several residues in the active site, such as Thr^{66}, Gln^{148}, Asn^{155}, Glu^{152}, Lys^{156}, and Lys^{159}. As we described earlier, this optimal hydrogen-bonding geometry was lost thereafter due to competition with solvent molecules. In fact, using the average structure it was observed that the wild-type IN-5CITEP binding mode contains indirect interactions via bound waters: the keto oxygen of 5CITEP exhibits a solvent-mediated interaction to one of the carboxylate oxygens of Glu^{152}, while a bridge of two water molecules is present between the nitrogen of the indole ring of the ligand and the Mg^{2+} ion.

It should be noted that in our MD simulation of the wild-type, the inhibitor adopted a completely different orientation as compared to the x-ray structure and furthermore showed a different binding site with respect to the ones presented in previous computational docking studies (Sotriffer et al., 2000). In fact, the authors reported two main binding positions of the 5CITEP to the IN core domain. In the most frequently occurring and most favorable result, the ligand was found with the main molecular plane aligned ''horizontally'' within the active site, whereas in our work the ligand is perpendicular to the plane defined by the three catalytic residues. However, in both cases, the keto-enol oxygen pointed toward Glu^{152} and Asn^{155}, and the chloroindole is

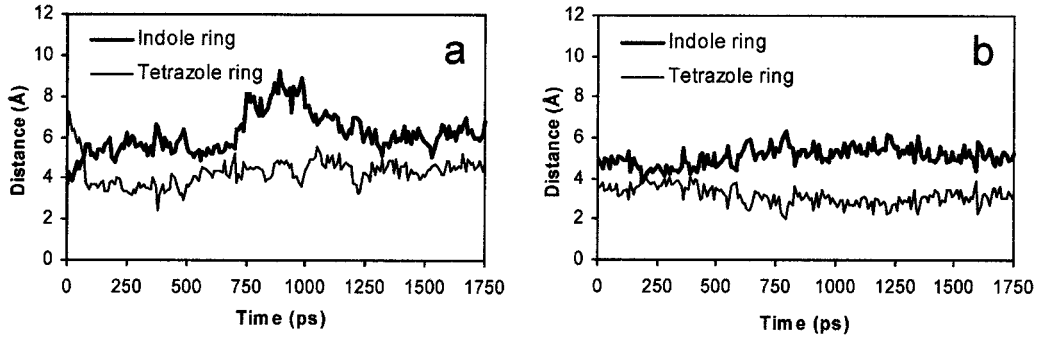

FIGURE 9 Time evolution of the distance between the fixed center of the catalytic triad (D64, D116, and E152) and the center of the indole and tetrazole rings of the ligand for the (*a*) wild-type and the (*b*) double mutant complexes.

located in proximity to Gln^{148}. The second result of the docking was found less frequently, and the inhibitor showed an ''opposite'' orientation with respect to the one we have presented: the chlorine became completely buried, the keto-enol oxygens are oriented toward the Asn^{155} amide, and the indole NH remained solvent-exposed.

Double mutant complex

In T66I/M154I, both the indole and tetrazole portions are nearly stable during the course of the MD trajectory, with an average distance from the catalytic center of ~5.6 and 3.0 Å, respectively. The inhibitor binding mode is therefore preserved along the entire dynamics simulation, and 5CITEP is involved in hydrogen-bonding interactions with the protein (a nitrogen atom of the tetrazole ring with a side chain oxygen atom of Asp^{64} and the indole NH with the backbone oxygen of His^{114}). Furthermore, the keto-enol group exhibits hydrogen-bonds only with solvent molecules. In this model, 5CITEP occupies a cleft that is formed by Gln^{62}, Leu^{63}, Asp^{64}, Ile^{66}, His^{114}, Thr^{115}, Asp^{116}, Glu^{138}, Phe^{139}, Ile^{141}, Tyr^{143}, Ser^{147}, Gln^{148}, Ile^{151}, Glu^{152}, Asn^{155}, and the metal ion with one coordinated water molecule.

In particular, the chloroindole portion is oriented toward Asn^{155}, while the tetrazole ring lies between Asp^{116} and Gln^{148}. Furthermore, it is important to note that Tyr^{143} is above the ligand, which does not interfere with the orientation of this residue and also allows it to point toward Glu^{152}. However, in this double mutant system the ligand is more deeply embedded in the protein than in the x-ray structure of the wild-type.

So, in both of our molecular dynamics simulations, most of the initial IN-5CITEP hydrogen-bonding interactions are gradually lost and other interactions, such as van der Waals interactions, appear to contribute largely to the stability of the inhibitor in the two complexes.

Essential dynamics studies

To support our MD results, we have performed principal component analyses (i.e., essential dynamics). In protein MD simulations, many atomic fluctuations occur, although few represent correlated motions. In general, the correlated motions are less than 1% of all possible motions. Essential dynamics (ED) is a method that can filter out all of the locally confined fluctuations and vibrational motions in a macromolecular system to focus on the few large and global structural motions (Amadei et al., 1993).

By diagonalizing the covariance matrix, the anharmonic and large-scale motions of the protein are isolated from the much larger number of remaining, mostly harmonic and small-scale motions. It has been established that the large-scale anharmonic motions in the essential subspace are often correlated to the vital functions of the protein.

To visualize the largest correlated motions of the proteins, two projected structures representing the minimal and maximal amplitudes along the first eigenvector were superimposed. It is worth noting that this ED analysis method shows the entire range of the motion rather than the dynamic flexibility of the region, which can be shown by atomic fluctuation calculations.

The characteristic movements of the catalytic core domains of the WT and the double mutant (T66I/M154I) were compared by superimposing the structures, as displayed in Fig. 10, for wild-type (Fig. 10 *a*) and the double mutant (Fig. 10 *b*). The significant differences between the two structures are in both loop regions, i.e., residues 138–149 (*Loop A*) and residues 185–194 (*Loop B*). For quantitative comparison of the differences, the C_α-C_α distances of each maximal and minimal structure were measured for all 154 residues (Fig. 10 *c*). The plot clearly shows that the amplitudes of the motions of both loops were significantly increased by the double mutation.

It has been previously reported that the flexibility of the active site loop is important for the enzymatic activity (Greenwald et al., 1999). These ED results suggest that the range of the loop motion may be correlated to the enzymatic activity. The double mutant has inhibitor resistance. Although it is not known whether the inhibitor can still bind to the doubly mutated enzyme, our simulations suggest that even if it does bind, it should not interfere with substrate binding (because it does not block the active site) and wild-type loop dynamics exist in the double mutant even with drug bound. At this point we hypothesize that the double mutant regains the dynamical characteristics of the active site loop of the wild-type even with the inhibitor bound. We therefore propose that dynamical flexibility of the loop is critical for enzymatic activity and the double inhibitor resistant mutant recovers activity by exhibiting wild-type loop flexibility, even in the presence of bound inhibitor.

Effect of the double mutation

The differential behavior of the loop and the ligand in the two systems needs to be explained in the context of the mutation of residues Thr^{66} and Met^{154}. At the beginning of the wild-type simulation, Thr^{66} made hydrogen-bonds with Ile^{73}, Asn^{155}, His^{67}, and the inhibitor. After the first 500 ps of dynamics simulation, the ligand-Thr^{66} interaction is broken and replaced by a hydrogen-bond with Lys^{159}. At around 800 ps, the Thr^{66}-Asn^{155} hydrogen-bond is also disrupted and only the interactions with Ile^{73} and Lys^{159} remain for the remainder of the trajectory.

Conversely, for the entire simulation of the double mutant model, Ile^{66} was involved in the only conserved hydrogen-bonding interaction with the backbone of Ile^{73}. Because Thr^{66} was mutated to Ile^{66} in the double mutant system, it was evident that during the simulation, the inhibitor could

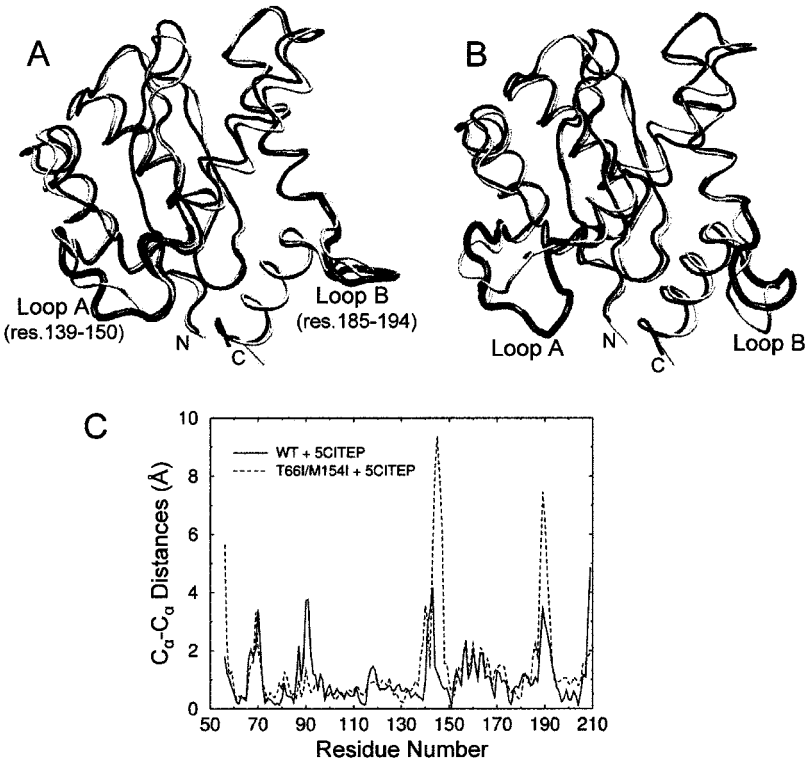

FIGURE 10 The largest correlated motions for the HIV IN catalytic core domains of (*a*) wild-type and (*b*) double mutant with a bound inhibitor, 5CITEP. For each figure, two protein trace structures from the maximal and minimal projections along the largest eigenvector were superimposed to show the largest correlated motions (maximum in *black* and minimum in *gray*). The two major moving loops, designated as Loop A and Loop B, are highlighted in thick black. Each structure was slightly rotated to show each loop structure clearly. For quantitative comparisons of the two structures, the 154 C_α-C_α distances of each maximal and minimal structure were measured (*c*).

not move into the same binding pocket available in the wild-type enzyme (Fig. 11 *a*) because the side chain of Ile^{66} occupied the center of that pocket (Fig. 11 *b*).

During the entire simulation of the double mutant complex, Ile^{154} makes hydrogen-bonds involving Val^{150}, Leu^{158}, and Glu^{157}; on the other hand, in the wild-type trajectory all of these interactions except the last one are conserved.

However, even if the interactions involving Ile^{154} are almost the same in both trajectories, the bulkier side chain of Ile^{154} seems to influence the conformation of neighboring residues. One of these residues is Glu^{152} which i), is a catalytic residue and is able to orient itself toward the other two active carboxylates even when the 5CITEP inhibitor is present in the double mutant and ii), stabilizes the inhibitor binding position in the wild-type complex through a water-bridge. Furthermore, Ile^{154} belongs to the helix α4 that, as we have shown, exhibits a different average length in the two systems (i.e., wild-type enzyme plus inhibitor and double mutant enzyme plus inhibitor).

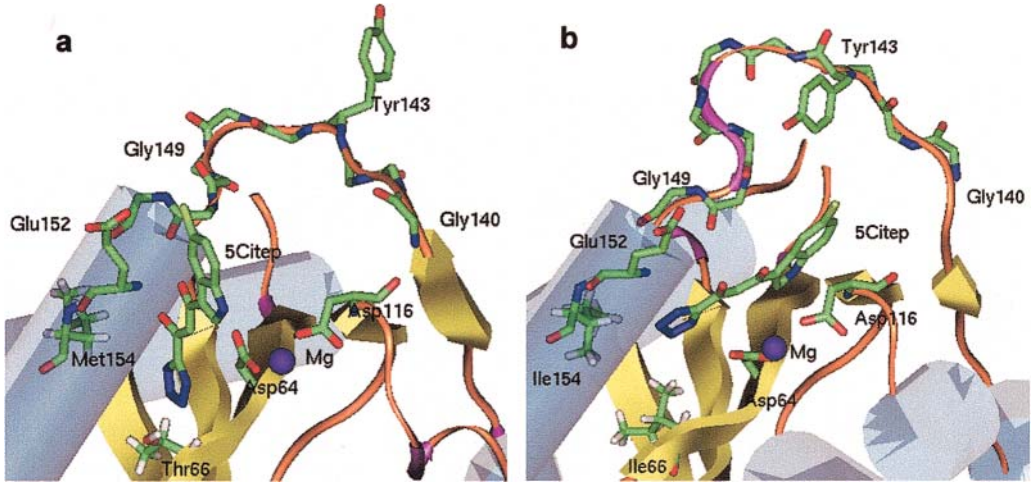

FIGURE 11 Comparison of the average structures from the (*a*) wild-type and the (*b*) double mutant T66I/M154I IN complex MD simulations.

Hypothetical mechanism for inhibition and drug resistance of IN

All of these results and observations led us to hypothesize a mechanism for the inhibition and drug resistance of IN. Recent DNA cross-linking studies with Rous sarcoma virus (RSV) IN suggested a model in which at least four monomers are required for concerted integration of two viral DNA ends into a target DNA (Yang et al., 2000). The catalytic activity is carried out only by two of the subunits, while the other ones help to position the substrate and stabilize the complex. Using as a template the structure of the two-domain (residues 49–286) RSV IN x-ray structure modeled with both viral cDNA ends and target DNA (kindly provided by F. D. Bushman), we have made hypotheses about the function of the inhibitor and double resistance mutation. In fact, the wild-type and double mutant average structures were superimposed with the RSV IN crystal form via their C_α atoms.

Fig. 12 shows, for both systems, the three conserved residues Asp^{64}, Asp^{116}, and Glu^{152}; the Mg^{2+} metal ion; the backbone of the catalytic loop with the Tyr^{143} side chain; the recessed 3′-OH group of a viral DNA end; the host cell chromosomal DNA phosphate group; and the orientation of the inhibitor.

The presence of the 5CITEP inhibitor in the wild-type enzyme could be important for stabilizing a nonfunctional conformation of the HIV-1 IN enzyme (Fig. 12 *a*). In fact, the inhibitor might affect the conformation of the surface loop near the active site that adopts a more constrained conformation with the side chain of Tyr^{143} pointing in the opposite direction of the catalytic center and of both DNA ends. Furthermore, in the wild-type complex, 5CITEP adopts a position that might inhibit the formation of the IN-DNA complex by occupying the DNA substrate-binding site and, considering the two-metal mechanism for the IN catalysis, might prevent the binding of the second divalent cation to its likely site located between residues Asp^{64} and Glu^{152}.

When the Thr^{66} and Met^{154} residues are mutated to Ile, HIV-1 IN might be able to assume a functional conformation with Tyr^{143} in the flexible loop and Glu^{152} pointing toward the active site (Fig. 12 *b*). This model seems to be an acceptable snapshot of the integration process, showing that in the double mutant the inhibitor position does not completely disturb the catalytic activity of the IN enzyme.

CONCLUSION

We have reported results herein of two 1.75 ns molecular dynamics (MD) simulations of the wild-type IN and the double mutant T66I/M154I both complexed with the 5CITEP inhibitor. Detailed analyses of structural perturbations in the enzyme-inhibitor complexes indicate that the two proteins display a distinct dynamical behavior. Major and significant structural differences were found at the catalytic residue Glu^{152}, the flexible loop 138–149 close to the active site, and in the mode in which the inhibitor binds to the active site of the enzyme.

In fact, the binding of 5CITEP to the wild-type IN caused several conformational changes especially in the loop 138–149 near the catalytic center, which achieved a rigid backbone configuration. Simultaneously, a dramatic conformational change of the side chain of Tyr^{143} brought this residue far away from the active site. The structure adopted by the constrained loop appears to represent a nonfunctional conformation of the HIV-1 IN enzyme.

Conversely, the MD simulation of the double resistance mutant IN (T66I/M154I) complex displayed a different conformational transition in the loop 138–149 and a high

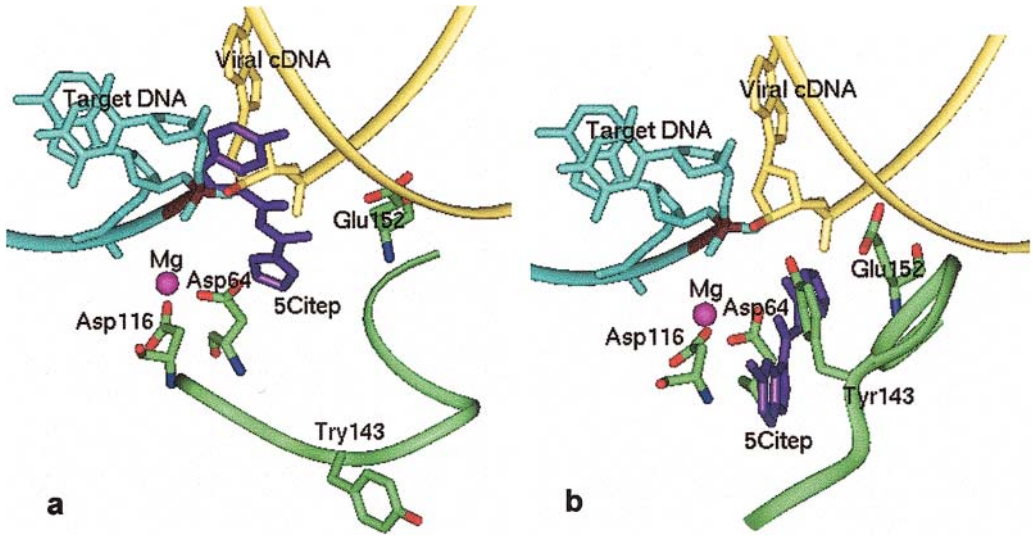

FIGURE 12 Hypothetical proposed mechanism for the HIV-1 IN inhibition (*a*) 5CITEP/wild-type IN complex, and (*b*) drug resistant 5CITEP/T66I/M154I IN complex, by MD studies.

degree of mobility in this region. The orientation of the Tyr^{143} within the catalytic loop is quite different in this model, pointing completely toward the active site.

In addition, the binding mode of the inhibitor is quite different in the two systems. In the wild-type complex the ligand has the tetrazole ring oriented approximately toward the pocket while the indole ring remains solvent-exposed; in the T66I/M154I complex, the inhibitor kept the same configuration throughout the molecular dynamics simulation and is more deeply buried in the active site than in the wild-type complex.

This work has provided insights into the detailed mechanism of action of the 5CITEP inhibitor and suggested the mechanistic cause of IN inhibition and drug resistance. Based on these results, our short-term aim is the design of suitable selective ligands able to inhibit integrase function.

We are currently using snapshots of the structures taken throughout the MD trajectories in docking studies of known HIV-1 IN inhibitors and in the development of rigid and dynamic (Carlson et al., 2000) pharmacophore models for use in 3D database searching.

J.M.B. thanks and acknowledges funding for this work from the National Institutes of Health. M.L.B. is grateful to the Consiglio Nazionale delle Ricerche (CNR) for its Short-Term Mobility fellowship. We thank Accelrys (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. Supercomputer time for this project was granted to J.M.B. by National Resource Allocation Committee and made available on facilities at the San Diego Supercomputer Center and at the University of Texas, Austin.

REFERENCES

Abagyan, R. A., M. M. Totrov, and D. N. Kuznetsov. 1994. ICM – a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* 15:488–506.

Amadei, A., A. B. Linssen, and H. J. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.

Anchell, J., E. Apra, D. Bernholdt, P. Borowski, E. Bylaska, T. Clark, D. Clerc, H. Dachsel, W. de Jong, M. Deegan, M. Dupuis, K. Dyall, D. Elwood, G. Fann, H. Fruchtl, E. Glendenning, M. Gutowski, R. Harrison, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nichols, J. Nieplocha, A. Rendall, M. Rosing, G. Sandrone, M. Stave, T. P. Straatsma, H. Taylor, G. Thomas, J. van Lenthe, T. Windus, K. Wolinski, A. Wong, and Z. Zhang. 1999. NWChem, A Computational Chemistry Package for Parallel Computers, Version 3.3.1. High Performance Computational Chemistry Group, Pacific Northwestern National Laboratory, Richland, WA.

Asante-Appiah, E., and A. M. Skalka. 1999. HIV-1 integrase: structural organization, conformational changes, and catalysis. *Adv. Virus Res.* 52:351–369.

Beese, L. S., and T. A. Steitz. 1991. Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10:25–33.

Berendsen, H. J. C., J. R. Grigerra, and T. P. Straatsma. 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271.

Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* 3:469–478.

Bujacz, G., J. Alexandratos, Z. L. Qing, C. Clement-Mella, and A. Wlodawer. 1996. The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant. *FEBS Lett.* 398:175–178.

Bujacz, G., J. Alexandratos, A. Wlodawer, G. Merkel, M. Andrake, R. A. Katz, and A. M. Skalka. 1997. Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* 272:18161–18168.

Bukrinsky, M. I., N. Sharova, M. P. Dempsey, T. L. Stanwick, A. G. Bukrinskaya, S. Haggerty, and M. Stevenson. 1992. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. USA.* 89:6580–6584.

Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield, and R. Craigie. 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. USA.* 90:3428–3432.

Cai, M., Y. Huang, M. Caffrey, R. Zheng, R. Craigie, G. M. Clore, and A. M. Gronenborn. 1998. Solution structure of the His12 → Cys mutant of the N-terminal zinc binding domain of HIV-1 integrase complexed to cadmium. *Protein Sci.* 7:2669–2674.

Cai, M., R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn. 1997. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 4:567–577.

Carlson, H. A., K. M. Masukawa, K. Rubins, F. D. Bushman, W. L. Jorgensen, R. D. Lins, J. M. Briggs, and J. A. McCammon. 2000. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43:2100–2114.

Chen, Z., Y. Yan, S. Munshi, Y. Li, J. Zugay-Murphy, B. Xu, M. Witmer, P. Felock, A. Wolfe, V. Sardana, E. A. Emini, D. Hazuda, and L. C. Kuo. 2000. X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50–293)—an initial glance of the viral DNA binding platform. *J. Mol. Biol.* 296:521–533.

Chow, S. A., K. A. Vincent, V. Ellison, and P. O. Brown. 1992. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science.* 255:723–726.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

De Clercq, E. 2000. Novel compounds in preclinical/early clinical development for the treatment of HIV infections. *Rev. Med. Virol.* 10: 255–277.

Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davies. 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science.* 266: 1981–1986.

Eijkelenboom, A. P., R. A. Lutzke, R. Boelens, R. H. Plasterk, R. Kaptein, and K. Hard. 1995. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* 2:807–810.

Eijkelenboom, A. P., R. Sprangers, K. Hard, R. A. Puras Lutzke, R. H. Plasterk, R. Boelens, and R. Kaptein. 1999. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins.* 36:556–564.

Engelman, A., and R. Craigie. 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.* 66:6361–6369.

Engelman, A., A. B. Hickman, and R. Craigie. 1994. The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* 68:5911–5917.

Engelman, A., K. Mizuuchi, and R. Craigie. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell.* 67:1211–1221.

Esposito, D., and R. Craigie. 1998. Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* 17:5832–5843.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

Goldgur, Y., R. Craigie, G. H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, and D. R. Davies. 1999. Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc. Natl. Acad. Sci. USA*. 96:13040–13043.

Goldgur, Y., F. Dyda, A. B. Hickman, T. M. Jenkins, R. Craigie, and D. R. Davies. 1998. Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc. Natl. Acad. Sci. USA*. 95:9150–9154.

Greenwald, J., V. Le, S. L. Butler, F. D. Bushman, and S. Choe. 1999. The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*. 38:8892–8898.

Harrison, R. J., J. Nichols, T. P. Straatsma, M. Dupuis, E. J. Bylaska, G. Fann, T. Windus, E. Apra, J. Anchell, J. D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, B. de Jong, M. Deegan, K. Dyall, D. Elwood, H. Fruchtl, E. Glendenning, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nieplocha, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, K. Wolinski, A. Wong, and Z. Zhang. 2000. NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.0. High Performance Computational Chemistry Group, Pacific Northwestern National Laboratory, Richland, WA.

Hazuda, D. J., P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, and M. D. Miller. 2000. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*. 287:646–650.

Heuer, T. S., and P. O. Brown. 1997. Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry*. 36: 10655–10665.

Heuer, T. S., and P. O. Brown. 1998. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry*. 37:6667–6678.

Hockney, R. W., and J. W. Eastwood. 1981. Computer Simulations Using Particles. McGraw Hill, New York.

InsightII. 1999. Accelrys, Inc., San Diego, CA.

Jenkins, T. M., D. Esposito, A. Engelman, and R. Craigie. 1997. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J*. 16:6849–6859.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.

Koradi, R., M. Billeter, and K. Wuthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph*. 14: 51–55.

Kulkosky, J., K. S. Jones, R. A. Katz, J. P. Mack, and A. M. Skalka. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol*. 12:2331–2338.

Lee, S. P., J. Xiao, J. R. Knutson, M. S. Lewis, and M. K. Han. 1997. Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. *Biochemistry*. 36:173–180.

Lins, R. D., A. Adesokan, T. A. Soares, and J. M. Briggs. 2000. Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacol. Ther*. 85:123–131.

Lins, R. D., J. M. Briggs, T. P. Straatsma, H. A. Carlson, J. Greenwald, S. Choe, and J. A. McCammon. 1999. Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophys. J*. 76:2999–3011.

Lodi, P. J., J. A. Ernst, J. Kuszewski, A. B. Hickman, A. Engelman, R. Craigie, G. M. Clore, and A. M. Gronenborn. 1995. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry*. 34:9826–9833.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun*. 91:57–95.

Maignan, S., J. P. Guilloteau, Q. Zhou-Liu, C. Clement-Mella, and V. Mikol. 1998. Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J. Mol. Biol*. 282:359–368.

McDonald, I. K., and J. M. Thornton. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol*. 238:777–793.

Molteni, V., J. Greenwald, D. Rhodes, Y. Hwang, W. Kwiatkowski, F. D. Bushman, J. S. Siegel, and S. Choe. 2001. Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallogr. D Biol. Crystallogr*. 57:536–544.

Ni, H., C. A. Sotriffer, and J. A. McCammon. 2001. Ordered water and ligand mobility in the HIV-1 integrase-5CITEP complex: A molecular dynamics study. *J. Med. Chem*. 44:3043–3047.

Polard, P., and M. Chandler. 1995. Bacterial transposases and retroviral integrases. *Mol. Microbiol*. 15:13–23.

Ryckaert, J. P., G. Ciccitti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constrains: molecular dynamics of n-alkanes. *J. Comp. Phys*. 23:327–341.

Sotriffer, C. A., H. Ni, and J. A. McCammon. 2000. HIV-1 Integrase inhibitor interaction at the active site: prediction of binding modes unaffected by cristal packing. *J. Am. Chem. Soc*. 122:6136–6137.

Tsurutani, N., M. Kubo, Y. Maeda, T. Ohashi, N. Yamamoto, M. Kannagi, and T. Masuda. 2000. Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells. *J. Virol*. 74:4795–4806.

van Gent, D. C., A. A. M. Oude Groeneger, and R. H. A. Plasterk. 1993. Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Res*. 21:3373–3377.

Vink, C., A. M. Oude Groeneger, and R. H. Plasterk. 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res*. 21:1419–1425.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph*. 8:52–56.

Weber, W., H. Demirdjian, R. D. Lins, J. M. Briggs, R. Ferreira, and J. A. McCammon. 1998. Brownian and essential dynamics studies of the HIV-1 integrase catalytic domain. *J. Biomol. Struct. Dyn*. 16:733–745.

Wlodawer, A. 1999. Crystal structures of catalytic core domains of retroviral integrases and role of divalent cations in enzymatic activity. *Adv. Virus Res*. 52:335–350.

Yang, Z. N., T. C. Mueser, F. D. Bushman, and C. C. Hyde. 2000. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol*. 296:535–548.

Zheng, R., T. M. Jenkins, and R. Craigie. 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA*. 93:13659–13664.