Molecular Dynamics Studies on the HIV-1 Integrase Catalytic Domain

Roberto D. Lins,* James M. Briggs,* T. P. Straatsma,# Heather A. Carlson, Jason Greenwald,§ Senyon Choe,§ and J. Andrew McCammon*

*Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0365; #Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352; and §The Salk Institute, La Jolla, California 92186-5800 USA

ABSTRACT The HIV-1 integrase, which is essential for viral replication, catalyzes the insertion of viral DNA into the host chromosome, thereby recruiting host cell machinery into making viral proteins. It represents the third main HIV enzyme target for inhibitor design, the first two being the reverse transcriptase and the protease. Two 1-ns molecular dynamics simulations have been carried out on completely hydrated models of the HIV-1 integrase catalytic domain, one with no metal ions and another with one magnesium ion in the catalytic site. The simulations predict that the region of the active site that is missing in the published crystal structures has (at the time of this work) more secondary structure than previously thought. The flexibility of this region has been discussed with respect to the mechanistic function of the enzyme. The results of these simulations will be used as part of inhibitor design projects directed against the catalytic domain of the enzyme.

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) encodes three enzymes as part of the POL gene, namely the reverse transcriptase (RT), protease (PR), and integrase (IN). Drugs have been available for a number of years that target the RT and several have recently been approved that operate against the PR. In fact, cocktails of these compounds have proven to be quite effective in the clinical applications. However, in part based on problems of patient noncompliance, resistant strains are known which render each of these drugs ineffective. It has therefore become imperative that drugs be designed to target other aspects of the viral life cycle and that they be designed in such a way as to make them less prone to resistance (escape) mutations. One recent target is the HIV-1 integrase, for which a mostly complete crystal structure of the catalytic domain was published in 1994 (1ITG) (Dyda et al., 1994). A second crystal structure was published in 1996 (2ITG) which contains all residues missing in the active site from the first structure (Bujacz et al., 1996a); however, they form a rather extended conformation and have moderately high thermal factors.

The HIV-1 integrase is composed of a single polypeptide chain that folds into three functional domains (Fig. 1) (Andrake and Skalka, 1996). Three-dimensional structures have been determined for all three domains separately and all form dimers. The N-terminal domain is comprised of residues ~1–50 and has a zinc binding motif that is different from the typical zinc finger fold, based on NMR studies (Cai et al., 1997; Eijkelenboom et al., 1997), and is known to be important for protein-protein multimerization (Zheng et al., 1996; Lee et al., 1997; Heuer and Brown, 1998). The C-terminal domain, residues ~212–288, takes on an SH3-type fold (Lodi et al., 1995; Lutzke and Plasterk, 1998; Eijkelenboom et al., 1995) and is known to bind DNA strongly but nonspecifically (Lutzke and Plasterk, 1998; Khan et al., 1991; Vink et al., 1993; Woerner and Marcus-Secura, 1993; Lutzke et al., 1994; Engelman et al., 1994). All three domains are required for full catalytic activity (Vink et al., 1993; Drelich et al., 1992; Schauer and Billich, 1992), although the purified catalytic domain can carry out a so-called disintegration reaction (Chow et al., 1992; Bushman et al., 1993).

The catalytic domain, which is composed of residues 50–212, has an RNaseH-type fold and belongs to the superfamily of polynucleotidyl transferases. The active site is comprised of two Asp residues and one Glu, in the typical D,D(35)E motif, each of which is required for catalysis (Engelman and Cragie, 1992; Kulkosky et al., 1992; van Gent et al., 1993). It has fairly low sequence homology with the avian sarcoma virus (ASV) integrase (~24% identity); however, they share very high structural homology, especially in the catalytic region (Bujacz et al., 1995). The enzyme carries out at least two reactions, namely 3′ processing and strand transfer.

In the 3′-processing reaction, the integrase cleaves a dinucleotide from both 3′ ends of viral DNA adjacent to the highly conserved CA dinucleotide, in the long terminal repeat (LTR) region (Rice et al., 1996; Pommier et al., 1997). The newly exposed 3′ hydroxyl groups then undergo a concerted transesterification strand transfer reaction across a five-basepair stretch in the host DNA, such that the viral DNA is now inserted into the host chromosome. (Rice et al., 1996; Pommier et al., 1997) This results in a dangling

Received for publication 26 October 1998 and in final form 29 January 1999.

Address reprint requests to Dr. James Briggs, Department of Biology and Biochemistry, University of Houston, Houston, TX 77204–5513. Tel.: 713-743-8366; Fax: 713-743-8351; E-mail: jbriggs@uh.edu.

A preliminary report of this work was given on June 12, 1998 during the "Structural Biology of AIDS Related Proteins" meeting at the National Institutes of Health.

Roberto D. Lins is on leave from Departamento de Quimica Fundamental, Universidade Federal de Pernambuco, Recife, PE 50670-901, Brazil.

0006-3495/99/06/2999/13 $2.00

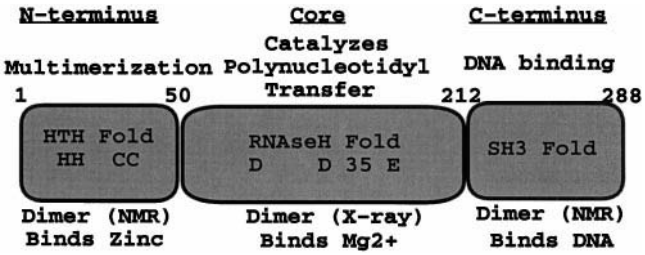

FIGURE 1 Schematic of the structure and function of the various domains of the HIV-1 integrase.

5′ dinucleotide at each end of the integration along with a five-basepair repeat at each end. These regions are presumably repaired by host DNA repair enzymes, although it is possible that the integrase may be involved in some aspects of the repair. The catalytic domain of the integrase, as mentioned above, can also carry out another reaction, known as disintegration (Chow et al., 1992; Bushman et al., 1993). This reaction is commonly used to assay for efficacy of inhibitors directed against the catalytic domain and is essentially the reverse of the strand transfer reaction.

Crystal structures are available for the catalytic core domain of the HIV-1 integrase, but each structure has missing or poorly defined regions near the active site. (Dyda et al., 1994; Bujacz et al., 1996a) Although divalent ions are known to bind within the active site, none were detected in the crystal structures (Dyda et al., 1994; Bujacz et al., 1996a). In an effort to describe the active site and address the binding of metal ions in the catalytic domain, we have performed crystallographic structure determinations and computer simulations. Throughout our exploratory molecular dynamics (MD) simulations, we observed important conformational interactions in the active region, particularly residues 141–150, which are part of the unsolved region in the crystal structures. The presence of a divalent metal in the active site, the conformations of Q148 and Y143, and secondary structure transitions constitute the salient points of our studies providing, for the first time, information about the dynamic behavior of the HIV-1 integrase catalytic domain. Furthermore, descriptions of the hydrated active site are provided which should prove useful in drug design as it provides a model for the uncomplexed active site in vivo.

METHODOLOGY

Crystallization and structure determination

The catalytic domain (residues 50–212) of the HIV-1 integrase containing a single mutation (F185K) was expressed and purified according to the published method (Dyda et al., 1994). The protein crystals were obtained by vapor diffusion at 277 K using 15 mM ammonium sulfate and 5 mM dithiothreitol, as a precipitant. This condition is similar to what has been published by Dyda et al. (1994). The integrase crystallized in the space group P3121 with unit cell dimensions $a = b = 7.26$, $c = 6.55$ nm. In order to collect the x-ray diffraction data, crystals were equilibrated in cryoprotectant prior to freezing by a nitrogen stream at 110 K.

Diffraction data for the present study were collected to 0.18 nm resolution at the 1–5 beamline at the Stanford Synchrotron Radiation Laboratory. To obtain the set of phases, rigid-body refinement was performed using the 1ITG structure coordinates as the starting model and with use of the X-Plor computer program (Brunger, 1996). Subsequent rounds of positional and temperature factor refinement were performed alternating with manual inspection of 2Fo-Fc electron density maps.

Modeling

This structure contained residues 57–140, 149–189, 193–210 in addition to 110 crystal waters. The missing regions (141–148, 190–192) were completed in order to perform MD simulations. A model of the first missing region (141–148) was generated from a backbone alignment against an ASV integrase structure (1ASV) (Bujacz et al., 1995), while the second missing region (190–192) was modeled based on an alignment with a previous HIV-1 integrase structure determined as described above. The primary sequence in this missing region was corrected using the residue replacement feature in the InsightII software (InsightII, 1997), because the HIV and ASV integrases do not have the same amino acid sequence (IPYNPQSQ for HIV and IPGNSQGQ for ASV).

The ionization state for each ionizable residue in the catalytic domain was predicted using a procedure available in the UHBD program (Madura et al., 1995), which is used to perform a detailed electrostatics analysis of the protein employing a continuum dielectric treatment of the solvent (Antosiewicz et al., 1994, 1996a, b). Effects of ionic strength can also be treated with this method, although an ionic strength of 0 was used in the present study. No unusual amino acid ionization states were predicted. Hydrogens were then added to the protein and crystal water molecules using the HBUILD module in CHARMM v.25 (HBUILD, 1992) based on the predicted ionization state for each residue. In the case of the simulation involving one Mg^{2+}, the metal was placed between the first two catalytic residues (D64 and D116) in the same relative position where it is observed in an ASV structure (1ASH) (Bujacz et al., 1996b). One or more counterions were added such that they were not initially within ~1.0 nm of any protein atom. The system (protein, crystal waters, counterions, and Mg^{2+}), which has a maximum dimension of ~4.0 nm, was embedded in a $6.4 \times 6.4 \times 6.4$ nm box of water. Water molecules that were found to be within 0.28 nm of any atom in the solute were removed, resulting in 2375 solute atoms and 23,187 solvent atoms for the simulation without a metal ion. The simulation with one Mg^{2+} also required the addition of two Cl^- ions and had a system size of 2378 solute atoms and 23,064 solvent atoms. The SPC/E water model (Berendsen et al., 1987) was used to describe the solvent. A 1.0 nm short-range cutoff was used for all nonbonded interactions, and long-range electrostatic interactions were treated by the PME method (Essmann et al., 1995), with a grid size of $64 \times 64 \times 64$ (i.e., there is no neglect of any electrostatic interactions). No correction was applied for the neglected long-range van der Waals interactions because they are expected to be small.

Molecular dynamics

The simulation system was equilibrated, as described below, in a stepwise fashion in order to eliminate bad atomic contacts in a gradual way. The solvent was energy-minimized, keeping the protein and counterions/Mg^{2+} fixed, using 200 steps of steepest descent, plus an additional 200 steps for the whole system. Solvent equilibration (protein + counter ions/Mg^{2+} fixed) was carried out by performing molecular dynamics for 20 ps at 298 K using a 1-fs time step. Equilibration of the solute (solvent fixed) was performed for 5 ps each at temperature intervals of 50, 100, 150, 200, 250, and 298 K with velocity reassignment every 0.5 ps and a time step of 2 fs. Finally, the whole system (solute and solvent) was equilibrated for 10 ps at 298 K using a 2-fs time step. Periodic boundary conditions, SHAKE (Ryckaert et al., 1977), and the NVT ensemble were used throughout the minimization/equilibration procedure.

Data acquisition was carried out for 1 ns at 298 K with separate relaxation times for the solvent and solute of 0.4 and 0.1 ps, respectively, using periodic boundary conditions in the NPT ensemble (pressure = 1.025×10^5 Pa). A SHAKE (Ryckaert et al., 1977) tolerance of 10^{-4} nm was applied to all bonds involving a hydrogen atom. A 2-fs time step was used and snapshots of the trajectory were taken every 0.1 ps (50 time steps).

All energy/MD calculations were performed using the AMBER95 force field (Cornell et al., 1995) and the leapfrog time step algorithm as implemented in the parallel NWChem v3.2 program (Anchell et al., 1998). Calculations were performed on IBM SP2 and Cray T3E parallel computers at SDSC and the University of Texas, respectively, using a spacial decomposition of the molecular system across processors. The physical volume was divided into rectangular boxes, each of which was sent to a processor that handled one or more units of these grouped boxes. The efficiency of the NWChem program (Anchell et al., 1998) was improved with the use of methods that minimize the number of solute-bonded interactions crossing node boundaries, by avoiding solvent bonded interactions between nodes, and by breaking the molecular system into separately treated solvent and solute parts (Anchell et al., 1998). Also, algorithms allowing dynamic balancing of the workload and which manage the communication calls between the processors were used to this aim (Anchell et al., 1998). The AMBER potential energy function (Cornell et al., 1995) used is given by,

$$V(r_i) = \sum_{n=1}^{N_b} \frac{1}{2} C_{b,n}(b_n - b_{0,n})^2 + \sum_{n=1}^{N_\theta} \frac{1}{2} C_{\theta,n}(\theta_n - \theta_{0,n})^2 + \sum_{n=1}^{N_\xi} \frac{1}{2} C_{\xi,n}(\xi_n - \xi_{0,n})^2 + \sum_{i<j}^{N} \left[\frac{C_{12,ij}}{r_{ij}^{12}} - \frac{C_{6,ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \quad (1)$$

The first three terms describe the bonded interactions, while the fourth term accounts for the nonbonded ones. The bonded terms are described by harmonic potentials for bond stretching (first term), angle bending (second term), and dihedral bending (third term). The nonbonded interactions are represented by the fourth term where the Lennard-Jones potential function describes the van der Waals interactions and the electrostatic term is given by a Coulombic function based on the separation of atom-centered point charges.

RESULTS AND DISCUSSION

Crystal structure

The starting structure coordinates for the molecular dynamics simulations are the result of crystallographic refinement and manual rebuilding using data at a significantly higher resolution; several notable changes were made to the starting model (Dyda et al., 1994). The side chains with the most significant differences from the starting model were those of His-67, Val-77, Phe-83, Ile-84, Glu-85, Arg-107, Phe-121, Lys-136, Phe-139, Lys-186, and Tyr-194. In addition to these changes, the most notable is in the backbone conformation within the region (residues 140–153) that has been previously ambiguous (Dyda et al., 1994). This region of the polypeptide chain contains the catalytic residue E152. In the present structure, residues from 148 to 153 are clearly defined in helical conformations, which is in contrast to what has been proposed and modeled in the structure of HIV-1 IN (F185H; 2ITG) (Bujacz et al., 1996a), but is close to the conformation of ASV IN (Bujacz et al., 1995) (Fig. 2).

Two systems involving the HIV-1 integrase catalytic domain were treated here, one with no metal in the active site region and a second containing one Mg^{2+} ion placed between the first two catalytic residues (D64, D116), based on alignments with crystal structures of the ASV integrase.

Molecular properties and dynamical stability

No metal in the active site

Analysis of molecular properties can provide some insight into how stable various parts of the molecular system are

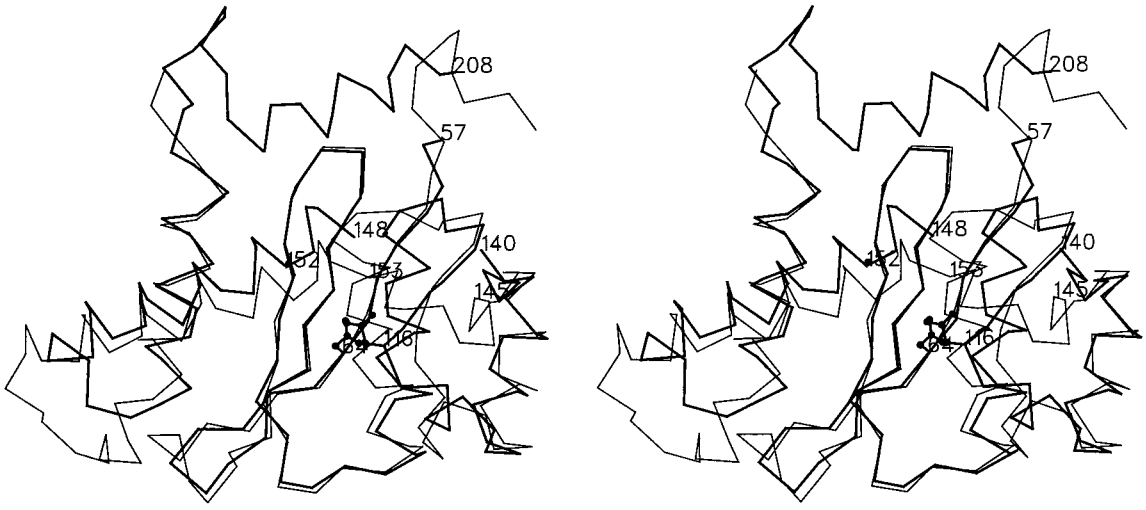

FIGURE 2 Superposition of the backbone conformations of ASV (*thin line*) and the HIV-1 (*thick line*) IN crystal structure reported herein.

with time. After a total of only 60 ps of equilibration, as described earlier, the temperature of the solute and solvent, total potential energy, and nonbonded intrasolute and solute-solvent interactions (Fig. 3), were observed to be quite constant across the 1 ns of data collection, meaning that the system was well-equilibrated. The mean temperature for the solute was 298.29 ± 1.05 K and for the solvent 297.96 ± 0.70 K.

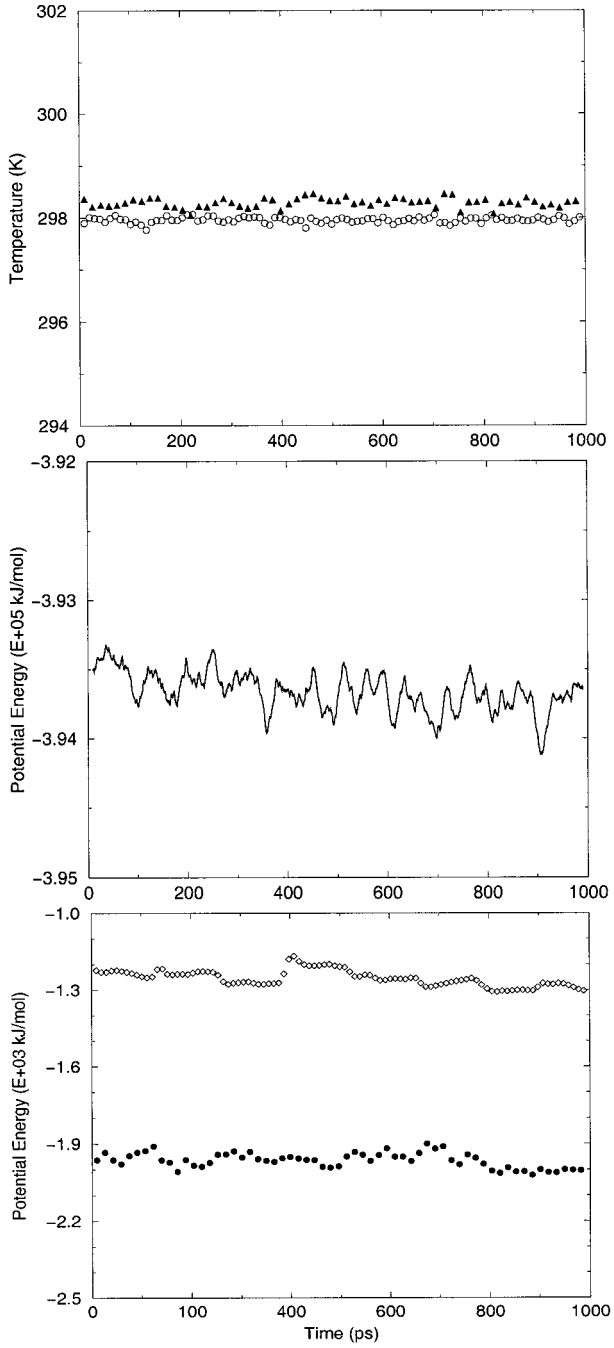

FIGURE 3 Time dependence of temperature (*top graph*) of the solvent (▲) and solute (○); total potential energy (*middle graph*); and solute-solute nonbonded interactions (bottom graph, ●) and nonbonded solute-solvent interactions (◇) for the no-metal simulation. All plots are shown as 10 ps averages. The uptick in the bottom panel at 400 ps was caused by a rearrangement of the solvent with respect to the protein due to a secondary structure transition (coil to helix) in the region 145–149. Note that after 500 ps this energy is again stabilized. This transition is well illustrated in Fig. 7, *top*.

The root-mean-squared deviations (RMSD) of all protein atoms were calculated using the What If program (Vriend, 1990) over the 1-ns trajectory taking as a reference frame the average structure (Fig. 4). The time evolution of the RMSD with respect to the average structure provides a measurement of the convergence of the dynamical properties of the protein. Each frame of the trajectory was superimposed on the reference frame using a least-squares fit method; periodicity effects, rotational, and translational motion were removed, leaving just the internal motions of the system. The RMSD from the average structure is stable after 400 ps, showing that our MD was successfully converged to an RMSD of 0.06–0.1 nm, even though two parts of the structure were model-built (a total of 11 residues). These regions and their neighborhood exhibit high flexibility, possibly explaining why their 3D structures could not be well determined by x-ray techniques. However, this result also indicates that those regions converged reasonably well, with respect to the starting structure, after 400 ps. In Fig. 5, a comparison between the C_α x-ray B-factors and the atomic fluctuations over 1 ns of MD simulation are shown. Note that the areas with the highest flexibility in our simulation are in excellent agreement with the regions that have the largest B-factors in the x-ray structure.

One metal in active site

The same analyses were performed for the HIV-1 integrase containing one metal in the active site, which was placed between the D64 and D116 residues based on alignments with the ASV integrase, as described above. All of the physical and chemical properties mentioned above were quite well-equilibrated, as in the system containing no

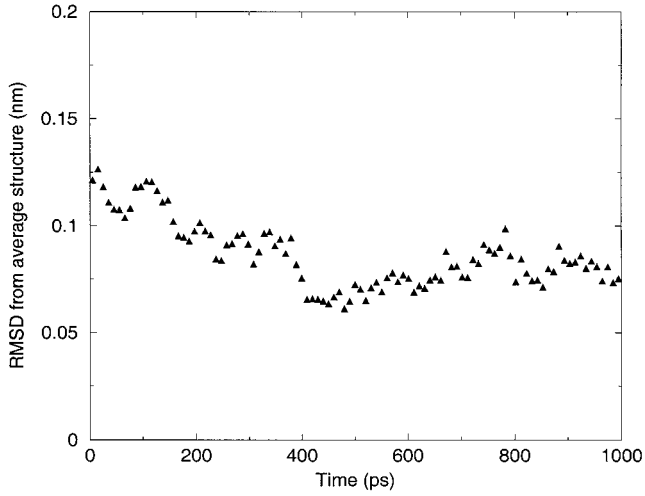

FIGURE 4 RMS deviation for all atoms of HIV-1 IN derived from the average structures during 1 ns of simulation shown as 10-ps averages.

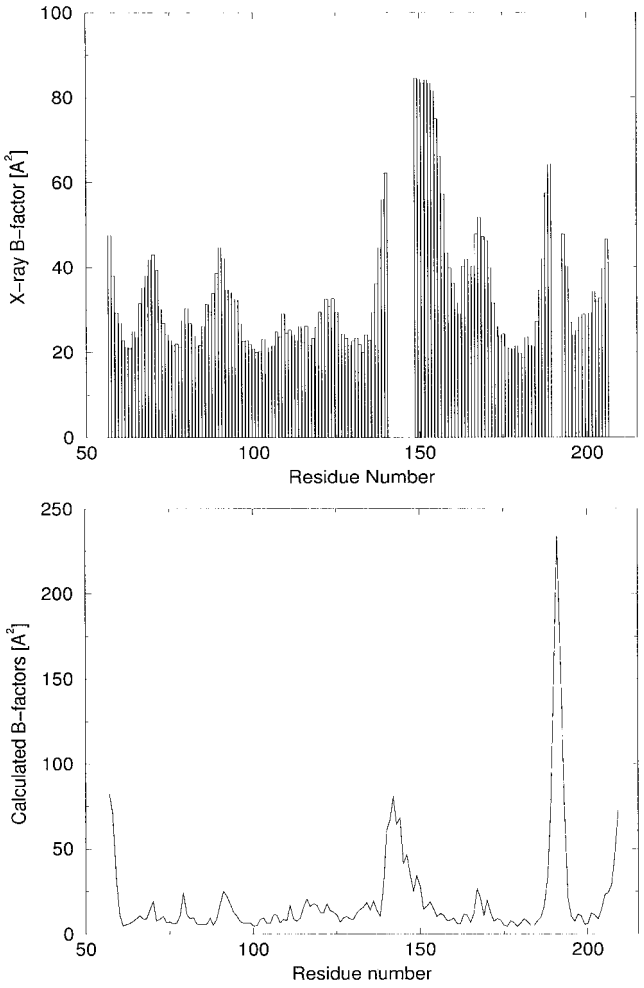

FIGURE 5 Comparison between experimental B-factors (*top graph*) and RMS fluctuations for C_α atoms (*bottom graph*) in HIV-1 IN. Note the two empty regions in the top graph, which represent the two missing regions in the crystal structure (141–148 and 190–192).

metal. The RMSD with respect to the average structure after 400 ps was maintained between 0.07 and 0.12 nm. The analyses of atomic fluctuations indicate that the same regions (141–148 and 190–192) showed the most flexibility in the two simulations.

Flexibility, helix formation, and coil-sheet transition

No metal in active site

Essential dynamics (ED) analysis techniques (Amadei et al., 1993; van Alten et al., 1997) were applied, using the What If program (Vriend, 1990), to our 1-ns trajectories in order to identify the regions of the protein that exhibit the largest flexibility during the simulations. This technique can be used to filter out local vibrational motions across an MD trajectory from the correlated ones. From our MD trajectories, a covariance matrix was built for the entire catalytic domain, and diagonalized to identify the 10 most significant motions of the solute along their eigenvectors (Fig. 6). Upon examination of the probability distribution of the displacements, for these projections, it was noted that the first two modes did not fit a Gaussian distribution (Fig. 6). This indicates a high anharmonicity and flexibility along these modes. The largest contributions to eigenvector 1 are located in the region 190–192, while the contributions to eigenvector 2 come primarily from residues 141–148 (Fig. 6). These results were expected because these regions are either missing or have high B-factors in the previously determined structures (Dyda et al., 1994; Bujacz et al., 1996a). Therefore, it should be noted in Fig. 6 that the large contribution of region 190–192, and its neighborhood, to the eigenvectors with eigenvalues indicate that it is very flexible.

A detailed analysis of the time-dependent secondary structural changes using the DSSP program (Kabsh and Sander, 1983) revealed that a portion of our modeled loop adopted a helical conformation, thus extending the α-4 helix (involving residues Q146–G149). This extension represents one additional turn of the helix. It was also observed that G149 assumed a helical conformation during the equilibration. The formation of this extra turn in the α-helix occurs during the first 300 ps of the simulation and then becomes stable after ~500 ps. This finding is not in agreement with the experimental HIV-1 integrase crystal structure 2ITG (Bujacz et al., 1996a), where an extended loop is found from G141 up to S153.

The second missing region (190–192) and its neighborhood have high flexibility, as mentioned above; however, our secondary structure analysis shows a coil transition into a small antiparallel β-sheet. This behavior was observed in both simulations, with and without metal, but this region predominantly adopts a random coil conformation.

One metal in active site

The ED (Amadei et al., 1993; van Alten et al., 1997) analysis for the system with one metal in the active site yielded the same results concerning protein flexibility as for the case with no metals; however, the secondary structure analysis (DSSP) (Kabsh and Sander, 1983) of the modeled region hints at some instability of the helical behavior for the same region that became an extension (extra turn) of the helix α-4 in the system with no metal (Fig. 7). In this case, the helix stretches during the dynamics simulation such that an analysis with DSSP (Kabsh and Sander, 1983) in this region reveals that these residues are less often in an ideal helical range, being stable only from Q148.

The greater stabilization of this region in the system with no metal ion is associated with a hydrogen-bonding interaction between the side chains of Q62 and Q148. This interaction was observed to be due to a χ_1 torsional movement of Q148, which is located at the start of the extra turn of the helix in the no metal simulation. This result provides some evidence that maintenance of a certain degree of flexibility in this region (141–148) may be required for

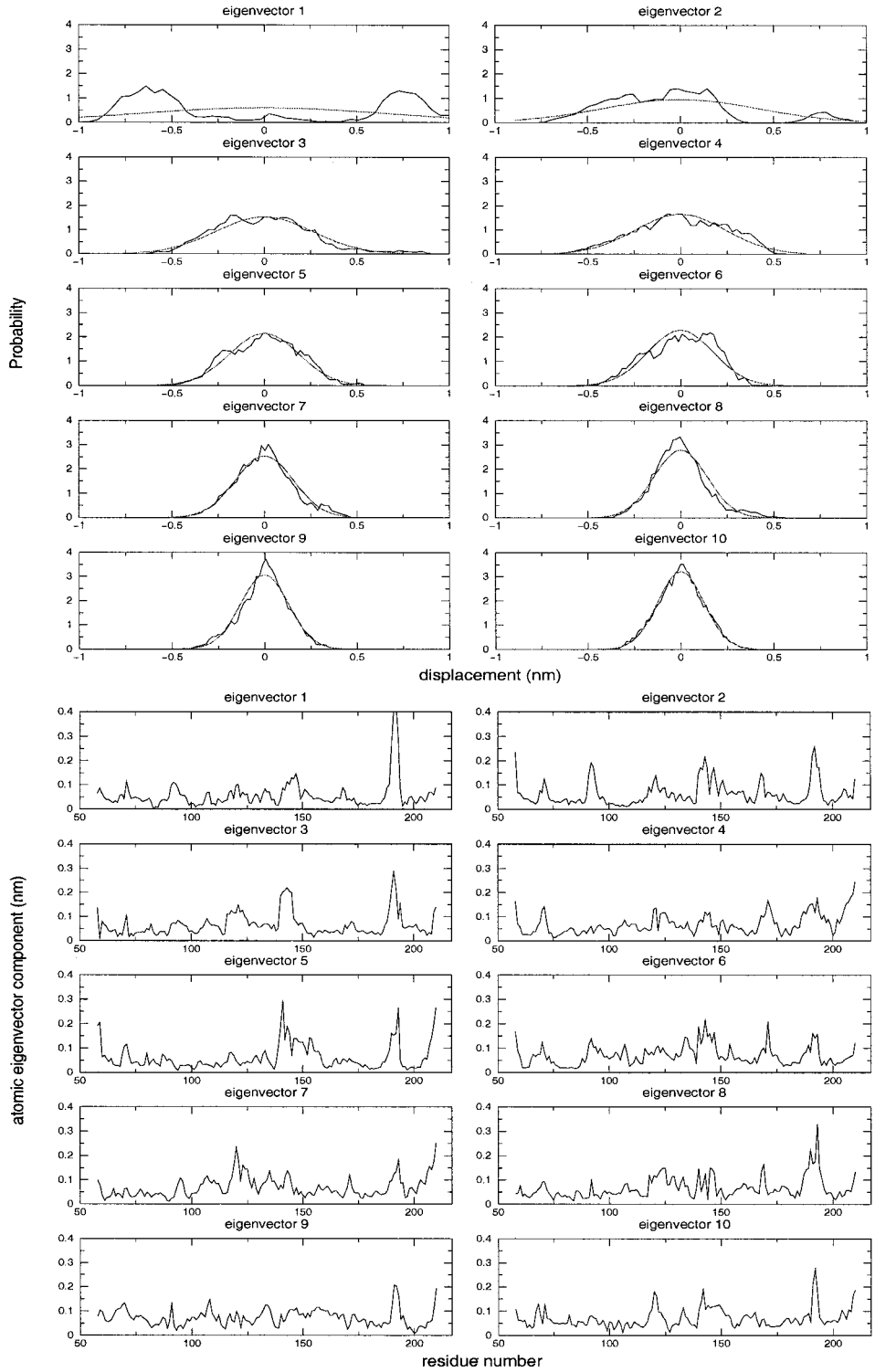

FIGURE 6 Displacements for the 10 highest eigenvectors (*top set*) and their corresponding atomic fluctuations throughout the 1 ns of MD for the no-metal simulation (*bottom set*).

enzymatic activity. However, the presence of one metal ion reduces the amplitude of some of the protein's vibrational motions. The minima and maxima of the three eigenvalues from the ED analyses for both systems are shown in Fig. 8, showing the amplitudes of those flexibilities. Arrows highlight the region 141–148, which is clearly more stable for the one-metal simulation (*arrow b*).

Active site conformation

There are two factors accounting for the stabilization of the helix in the absence of a metal ion: a hydrogen bond between the side chains of Q62 and Q148 and a solvent bridge between the hydroxyl of Y143 and a carboxylate oxygen of E152. The interaction of Q62 with Q148, which

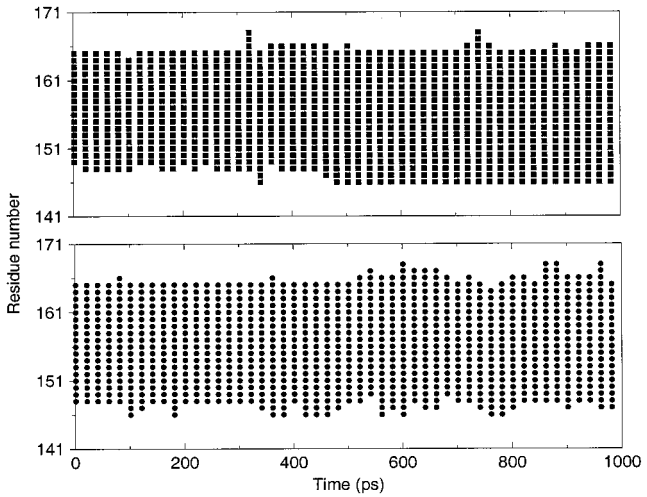

FIGURE 7 Time-dependent secondary structure analyses for the helix α-4 for the no-metal (*top graph*) and one-metal (*bottom graph*) simulations. The analyses were carried out for the entire 1-ns simulations using 10-ps snapshots. Each symbol represents a residue considered to be helical.

is located at the start of the extra turn of the helix, anchors the helix to other residues in the active site. Over the second half of the simulation, when the helix is most stable, Y143 and E152 are almost continually bridged by one or two water molecules, as can be seen in Fig. 9. This interaction reduces the strain from the flexible loop adjacent to the helix. Perhaps the most interesting role of the water in the active site is the interaction between Q148 and the three essential residues (D64, D116, and E152). Q148 usually participates in three single-solvent-bridged interactions to those key residues (Fig. 9). D64 and D116 are often bridged by two water molecules, and E152 and D116 are occasionally linked in the same fashion.

There is a marked difference in the water structure observed in the active site during the one-metal simulation (Fig. 10). First, it should be noted that the magnesium ion is always complexed in an octahedral fashion by D64, D116, and four water molecules. The average distance between the magnesium ion and the oxygens of the coordinating water molecules is 0.208 nm. Adding 0.14 nm for the radius of a water molecule, the radius for the solvated magnesium ion is just under 0.35 nm. It is interesting that most of the steric conflicts that arise from the accommodation of the ion are borne by the water and the side chains of Q148 and Y143. Q148 can no longer occupy the center of the active site with a hydrogen bond to Q62 and water-mediated interactions with D64, D116, and E152. Instead, it is pushed back toward the helix and interacts with E152, D116, and the metal ion through three two-water solvent bridges. Y143 can no longer interact with E152, as Q148 now occupies its previous position. The addition of the ion appears to make the active site more hydrophilic, as there are fewer residue-

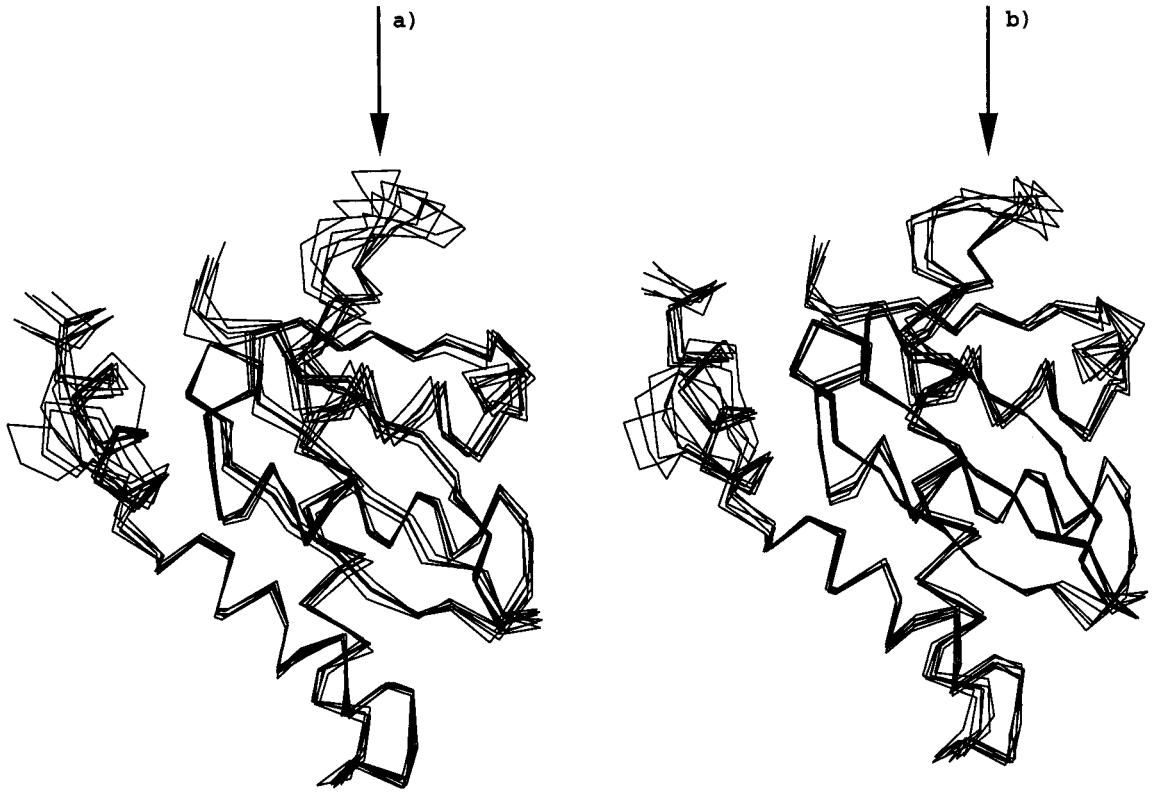

FIGURE 8 Superposition of HIV-1 IN structures containing the three highest eigenvectors over the three lowest ones for the no-metal (*left*) and one-metal (*right*) simulations. Arrows *a*) and *b*) highlight the region 141–148, which is missing in the crystal structure.

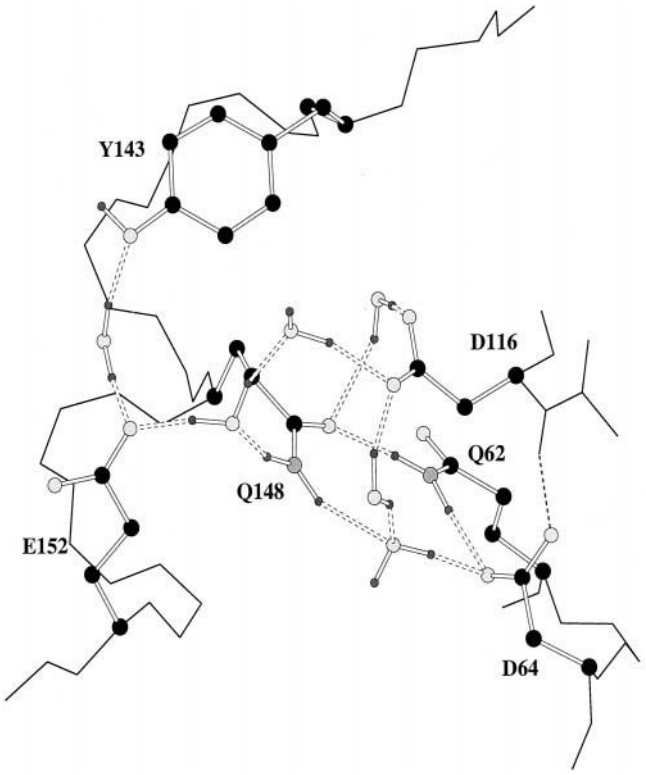

FIGURE 9 Representation of protein-protein and protein-solvent interactions seen in the active site. The only water molecules displayed are involved in bridging contacts between side chains. The hydrogen bond between D64 and the protein backbone is also shown on the right. The backbone of the helix and flexible loop runs from the top to the left corner.

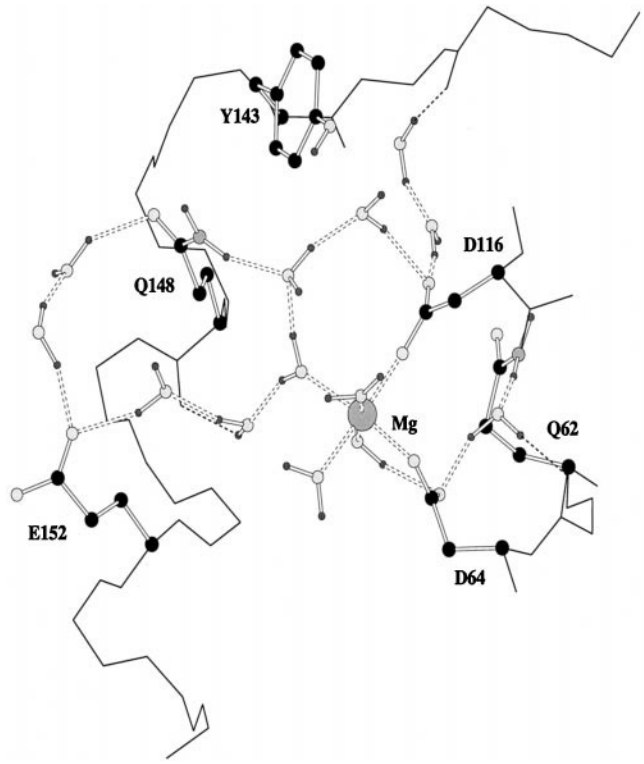

FIGURE 10 The active site in the presence of a divalent cation. Note the octahedral coordination of the ion and the increased number of water molecules as compared with the no-metal simulation in Fig. 9. Although D64 is no longer able to form a hydrogen bond to the backbone, D116 has gained an occasional solvent-mediated interaction to the flexible loop.

residue contacts and longer solvent-bridged contacts. The displacement of Q148 by the hydrated magnesium ion forces a change in orientation of Y143 and a destabilization of the helix through the loss of favorable interactions for both Q148 and Y143.

Distance measurements between the C_γ of D64 and D116 reveal that the separation between the side chains of these two residues is fairly stable across the MD simulations, ~0.7–0.8 nm most of the time; however, distances involving those residues and the C_γ of E152 show large fluctuations in the no-metal case. These fluctuations are due to the existence of two conformational preferences for the χ_1 angle of E152 at −60 and −180/180°. Distance fluctuations are smaller when the magnesium ion is present in the active site. Complexation of the Mg^{2+} by D64 and D116 causes χ_1 of D116 to have only one conformational preference after the system properties equilibration (see Fig. 11).

The third catalytic residue in the HIV-1 integrase active site, E152, is missing in the 1ITG crystal structure but is present in 2ITG with very high B-factors. Even though our initial crystal structure has a part of this residue (up to C_β) and lacks Mg^{2+}, it still has high B-factors. The equivalent catalytic residue in ASV IN (E157) also has substantial flexibility and the distances to the other two catalytic residues (D64 and D121) are quite variable and seem to be dependent on the presence or absence of divalent metals in the active site. Distance measurements between the carboxyl groups of the catalytic residues of the three ASV IN structures with no metals, as well as for our average structures of HIV-1 IN with no metals in the active site, are shown in Table 1. Table 2 shows the distance measurements for the HIV-1 IN average structure with one metal, in addition to the three published ASV IN structures containing one divalent ion, all in the same metal-binding site. According to our studies, the HIV-1 IN active site is somewhat wider than in the ASV IN with respect to the distances among the catalytic residues, largely due to the flexibility of E152, and the binding of Mg^{2+} does not produce significant conformational changes in the active site. These dimensions are probably sensitive to the presence of a metal in the second site which has been proposed to coordinate to the third catalytic residue, E152, as in the ASV IN structures containing two metals.

Another parameter studied was the angle formed among the carboxyl groups of the catalytic residues. For both systems studied, this angle converged after 500 ps to similar values, but the fluctuations are visibly smaller for the system with one metal ion in the active site (Fig. 12). This is rather surprising considering the differences seen in Figs. 9 and 10, and it may imply a functional role.

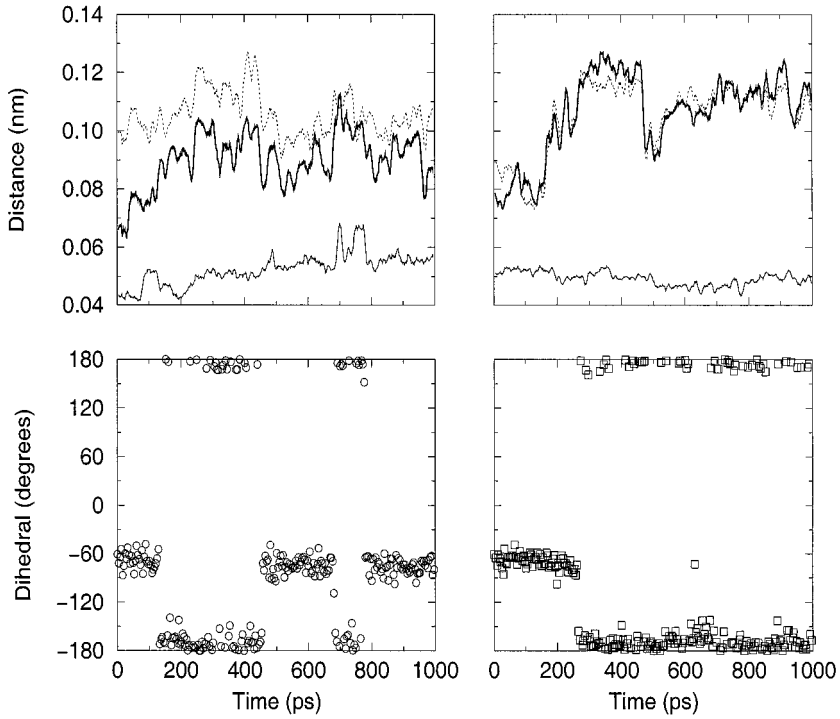

FIGURE 11 Distances between carboxyl groups of the catalytic triad (*top graphs*) throughout the simulations are shown along with the χ_1 dihedral angle of D116 (*bottom graphs*) for the no-metal (*left*) and one-metal (*right*) simulations. The thin solid line represents the D64–D116 distance, the dotted line D116–E152, and the thick line the distance between D64 and E152. All plots are shown as 10-ps averages.

Other important residues

We stress that Y143 has been pointed out as an important residue in the catalytic process. Experimental observations reveal a conserved tyrosine residue in the catalytic domain close to the active site among a variety of retroviral integrases (Engelman and Cragie, 1992). However, statistical analyses show that tyrosines are rarely located in surface loops, ranking as only 16th most often found residue in this kind of secondary structure element (Kwasigroch et al., 1996). Mutagenesis studies have shown that in the HIV-2 integrase catalytic core domain, replacement of this tyrosine by a leucine does not reduce the enzymatic activity; however, it shifts the favored nucleophile for terminal cleavage when Mn^{2+} is present (i.e., it affects the balance between hydrolysis/3′-processing, and alcoholysis/strand transfer) (van Gent et al., 1993; van den Ent et al., 1998). By using the photo-cross-linking technique, the region between residues 139–152 was identified as the one interacting with DNA (Heuer and Brown, 1997). These facts, taken together, constitute convincing evidence that Y143 seems to play the secondary role of stabilizing and directing the nucleophile for most efficient and balanced catalysis, similar to that proposed for DNA polymerase I (Beese and Steitz, 1991). The sequence comparison between the first missing region (141–148) in HIV-1 integrase and its corresponding region in ASV integrase is shown below:

(140)GIPYNPQSQGVVES(153): HIV-1 Integrase

(145)GIPGNSQGQAMVER(158): ASV Integrase

In ASV, a glycine residue takes the place of the tyrosine (Y143) in the HIV-1 integrase sequence. However, when the heavy atoms of the catalytic triad (D64, D116, and E152) in the average structure from the no-metal simulation were superimposed on the equivalent residues in the ASV integrase (D64, D121, and E157), it revealed that the hydroxyl group of Y143 in HIV-1 integrase was found structurally close to the amide oxygen of residue Q151 in ASV IN. Note also that the values for the angles involving the carboxyl groups of the last two catalytic residues and the hydroxyl group of Y143 in HIV-1 IN and the amide one in ASV IN are similar, 74.2° and 78.1°, respectively. This angle was found to be relatively constant during our simulations in both cases despite the different orientation of

TABLE 1 Distances between carboxyl groups (nm) of the catalytic triad D,D,35E motif in three ASV integrases and in the HIV-1 integrase average structure for the no-metal simulation

Residues	HIV-1 IN (no metal)	1ASV	1ASW	1VSE
D64–D116	0.513	0.553	0.567	0.544
D116–E152	0.894	1.183	0.854	0.899
D64–E152	1.051	0.900	0.612	0.698

TABLE 2 Distances between carboxyl groups (nm) of the catalytic triad D,D,35E motif in three ASV integrases and in the HIV-1 integrase average structure for the one-metal simulation

Residues	HIV-1 IN (one metal)	1VSI	1VSF	1VSD
D64–D116	0.485	0.544	0.512	0.487
D116–E152	1.055	0.899	0.907	0.840
D64–E152	1.055	0.698	0.740	0.671

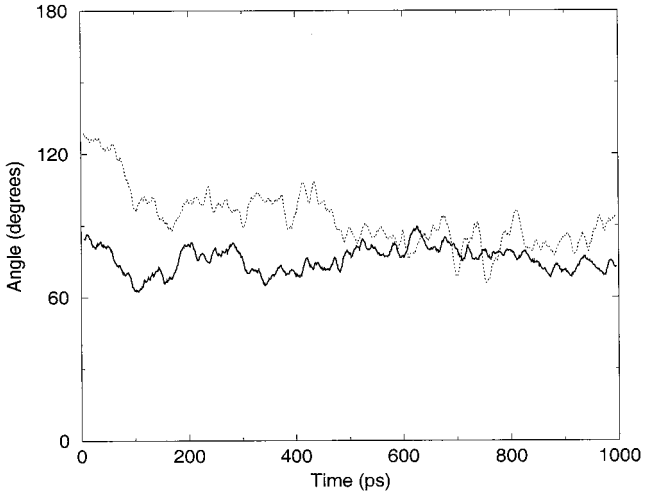

FIGURE 12 Time-dependent variation of the angle formed between the carboxyl group of the catalytic residues shown as 10-ps averages for the no-metal (*dotted line*) and one-metal (*solid line*) simulations.

Y143 (with and without metal in the active site; see Fig. 13). Again, the variation in the no-metal case is larger than when one metal is present. This behavior shows the existence of a spatial relationship between Y143 and the active site. This part of the molecule (residues 141–148) constitutes an important region for the enzymatic mechanism and its behavior could point to the need for flexibility for efficient catalytic activity.

In addition to Y143, residues Q148 and N117, among others, were also highlighted as critical residues for the catalytic activity of the HIV-2 integrase in the work of Vink and coworkers (Vink et al., 1993). Some of the conformational changes of Q148 in the absence of metals observed in our simulation have already been discussed. The distance between the C_δ atoms of Q62 and Q148 plotted against the distance between the C_γ atoms of D64 and D116 shows a significant correlation among these four residues for the system with no metal (Fig. 14). This indicates that the interaction between Q62 and Q148 is possible only if D116 assumes a different conformation.

Another important residue for normal enzymatic activity is N117, which is located very close to the active site, but its side chain is not pointed in that direction. The one-metal simulation showed this residue to be rather stable as compared with the crystal structure; however, in the no-metal MD simulation, the χ_1 of N117 distorts in the direction of the active site (Fig. 15).

CONCLUSIONS

MD simulations of the HIV-1 integrase containing no metal or one metal in the active site were successfully carried out, resulting in well-equilibrated molecular systems. Analyses of time-dependent properties showed that the motions in regions 141–148 and 190–192 have large amplitudes, which agrees very well with the experimental difficulty in determining the structures of these regions via x-ray crystallography. Except for those regions, the whole protein was found to be very stable.

Conformational flexibility of the secondary structure in the region 146–149, between the system containing no metal or one metal in the active site, implies some importance for flexibility in this region, which could be a requirement for efficient biological activity. This notion is further supported by a recent result, in which two mutations (G140A and G149A) introduced near this region rigidly extended the helical conformation toward the N-terminal side of α-4 up to P145, which resulted in a significant decrease in its catalytic activity (Greenwald et al., 1999, personal communication). The active site of the HIV-1

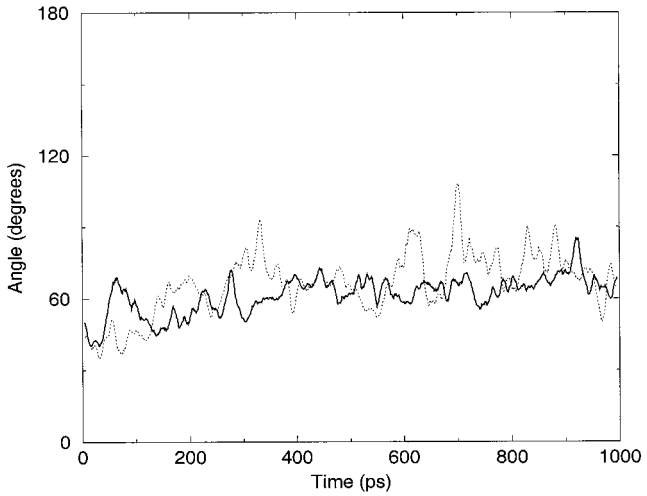

FIGURE 13 Angle formed by the hydroxyl oxygen of Y143 and the carboxyl groups of the last two catalytic residues (D116 and E152) shown as 10-ps averages for the no-metal (*dotted line*) and one-metal (*solid line*) simulations.

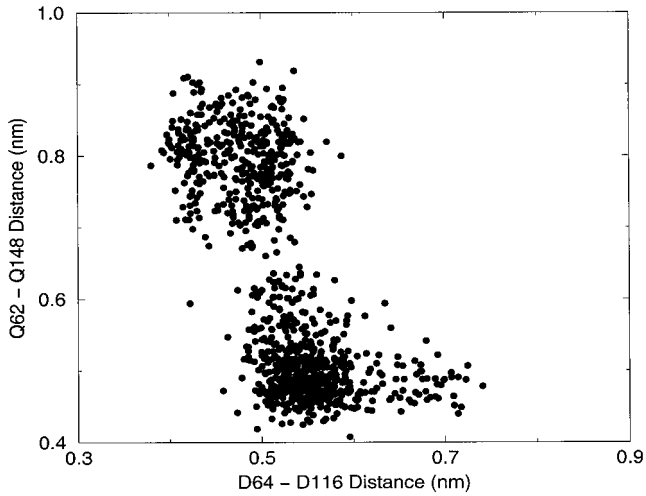

FIGURE 14 Correlation of the distances between the C_δ of Q62–Q148 and C_γ between D64 and D116 shown as 10-ps averages from the no-metal simulation.

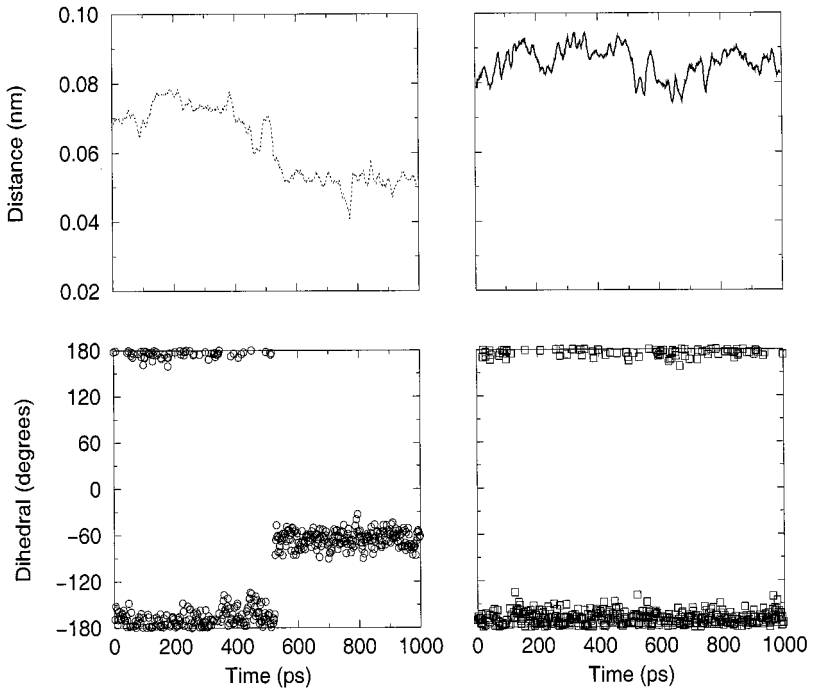

FIGURE 15 Distances between the center of mass of the amide group of N117 and the center of mass of the first two catalytic residues (*top graphs*) shown as 10-ps averages for the no-metal (*dotted line*) and one metal (*solid line*) simulations. The χ_1 dihedral angle is shown in both cases; no-metal (*bottom left graph*) and one-metal (*bottom right graph*).

integrase has defined angular dimensions; however, pairwise distances were shown to be metal-dependent.

Residues Y143, Q148, and N117 and D116 showed significant conformational changes with respect to the initial structure only for the simulation with no metal, which means that the presence of metals close to the active site stabilizes this region as well as the loop formed between 141 and 145. Therefore, the absence of metal produces a stabilization of the region 146–149 into an α-helical conformation.

Just before submission of this manuscript, two experimental studies describing new HIV-1 IN crystal structures without and with one magnesium ion in the active site was published (Goldgur et al., 1998; Maignan et al., 1998). Our findings such as the absence of large changes in the active site upon binding the first magnesium ion, the position of E152 being closer to D64 and D116 in the active site, an extended α-4 helix and its metal-dependent stability and coil-sheet transitions in the neighborhood of the second missing region, are all in very good agreement with this new experimental paper (Goldgur et al., 1998; Maignan et al., 1998). This shows the predictive capability of MD simulations. It should also be noted that a preliminary report of molecular dynamics simulations that qualitatively demonstrated some of these structural features was presented prior to publication of the experimental results (Nicklaus et al., 1995).

The present study provides insights into the dynamics of the HIV-1 IN and the importance of the first metal ion in the active site, which can be useful for future studies involving rational drug design. Simulations involving two magnesium ions (ASV IN alignment-based) are underway. The results of that simulation can help to more completely characterize the HIV-1 IN active site conformation and address questions about the stability of the magnesium in the second metal site.

Drs. Volkhard Helms, Wolfgang Weber, and Rick Bushman are acknowledged for many useful conversations and advice. HAC is grateful to the American Cancer Society for a postdoctoral fellowship (PF-4427) and also thanks the La Jolla Interfaces in Science Training Program. The authors thank the San Diego Supercomputer Center for grants of computer time (to J.M.B. and J.A.M.). Gratitude is also expressed to Molecular Simulations, Inc., San Diego, CA for generously providing us with the InsightII and Quanta software. This project is supported by the NIH Program on Structural Biology of AIDS Related Proteins (GM56553). The NWChem computational chemistry package for parallel computers used in this study was developed by the High Performance Computational Chemistry Group, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, and funded by the Offices of Biological and Environmental Research, Computational and Technology Research, and Basic Energy Sciences in the U.S. Department of Energy. Pacific Northwestern Laboratory is operated for the U.S. Department of Energy by Battele Memorial Institute under contract ACO6_76RLO 1830.

REFERENCES

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

Anchell, J., E. Apra, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, M. Dupuis, K. Dyall, G. Fann, H. Fruchtl, M. Gutowski, R. Harrison, A. Hess, J. Jaffe, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nichols, J. Nieplocha, A. Rendall, M. Stave, T. P. Straatsma, H. Taylor, G. Thomas, K. Wolinski, and A. Wong. 1998. NWChem, A Computational Chemistry Package for Parallel Computers, Version 3.2. High Performance Computational Chemistry Group, Pacific Northwestern National Laboratory, Richland, WA 999352.

Andrake, M. D., and A. M. Skalka. 1996. Retroviral integrase, putting the pieces together. *J. Biol. Chem.* 271:19633–19636.

Antosiewicz, J., J. M. Briggs, A. E. Elcock, M. K. Gilson, and J. A. McCammon. 1996a. Computing the ionization states of proteins with a detailed charge model. *J. Comput. Chem.* 17:1633–1644.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996b. The determinants of pK_as in proteins. *Biochemistry.* 35:7819–7833.

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.

Brunger, A. T. 1996. X-Plor Version 3.8: A System for X-ray Crystallography and NMR. Yale University Press, New Haven, CT.

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

Bujacz, G., M. Jaskolski, J. Alexandratos, A. Wlodawer, G. Merkel, R. A. Katz, and A. M. Skalka. 1995. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* 253:333–346.

Bujacz, G., M. Jaskolski, J. Alexandratos, A. Wlodawer, G. Merkel, R. A. Katz, and A. M. Skalka. 1996b. The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure.* 4:89–96.

Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield, and R. Cragie. 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., 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.

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. *JACS.* 117:5179–5197.

Drelich, M., R. Wilhelm, and J. Mous. 1992. Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 in protein in vitro. *Virology.* 188:459–468.

Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Cragie, 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., F. M. van den Ent, A. Vos, J. F. Doreleijers, K. Hard, T. D. Tullius, R. H. Plasterk, R. Kaptein, and R. Boelens. 1997. The solution structure of the amino-terminal hhcc domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr. Biol.* 7:739–746.

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

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

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., F. Dyda, A. B. Hickman, T. M. Jenkis, 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.

HBUILD. 1992. Polar Hydrogen Parameter Set for CHARMm Version 22 (Polar Hydrogens Only). Molecular Simulations, Inc., Waltham, MA.

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.

InsightII. 1997. Molecular Simulations, Inc., San Diego, CA.

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

Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka. 1991. Retroviral integrase domains: DNA binding and the recognition of ltr sequences. *Nucleic Acids Res.* 19:851–860.

Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. 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 transposons. *Mol. Cell Biol.* 12:2331–2338.

Kwasigroch, J.-M., J. Chomilier, and J.-P. Mornon. 1996. A global taxonomy of loops in globular proteins. *J. Mol. Biol.* 259:855–872.

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

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.

Lutzke, R. A. P., and R. H. A. Plasterk. 1998. Structure-based mutational analysis of the c-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.* 72:4841–4848.

Lutzke, R. A. P., C. Vink, and R. H. A. Plasterk. 1994. Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.* 22:4125–4131.

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 structure 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.

Nicklaus, M. C., Y. Pommier, A. Mazumder, and G. W. A. Milne. 1995. HIV-1 integrase inhibitors. 3D searching and active site docking. *210th ACS National Meeting.* Chicago, IL, Aug. 20–24.

Pommier, Y., A. A. Pilon, K. Bajaj, A. Mazumder, and N. Neamati. 1997. HIV-1 integrase as a target for antiviral drugs. *Antiviral Chem. Chemother.* 8:463–483.

Rice, P., R. Cragie, and D. R. Davies. 1996. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* 6:76–83.

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.

Schauer, M., and A. Billich. 1992. The N-terminal region of HIV-1 integrase required for integration activity, but not for DNA binding. *Biochem. Biophys. Res. Commun.* 185:874–880.

van Alten, D. M. F., B. DeGroot, J. B. C. Findlay, H. J. C. Berendsen, and A. Amadei. 1997. A comparison of techniques for calculating protein essential dynamics. *J. Comp. Chem.* 18:169–181.

van den Ent, F. M. I., A. Vos, and R. H. A. Plasterk. 1998. Mutational scan of the human immunodeficiency virus type 2 integrase protein. *J. Virol.* 72:3916–3924.

van Gent, D. C., A. A. M. O. 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. A. M. O. Groeneger, and R. H. A. Plasterk. 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. *Nucleic Acids Res.* 21: 1419–1425.

Vriend, G. 1990. What If—a molecular modeling and drug design program. *J. Mol. Graphics.* 8:52–56.

Woerner, A. M., and C. J. Marcus-Secura. 1993. Characterization of a DNA binding domain in the c-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Res.* 21:3507–3511.

Zheng, R., T. M. Jenkins, and R. Cragie. 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.