Large-Scale Conformational Dynamics of the HIV-1 Integrase Core Domain and Its Catalytic Loop Mutants

Matthew C. Lee,* Jinxia Deng,† James M. Briggs,†‡ and Yong Duan§¶

*Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716; †Department of Biology and Biochemistry and ‡Department of Chemical Engineering, University of Houston, Houston, Texas 77204-5001; and §University of California, Davis, Genome Center and ¶Department of Applied Science, University of California, Davis, California 95616

ABSTRACT HIV-1 integrase is one of the three essential enzymes required for viral replication and has great potential as a novel target for anti-HIV drugs. Although tremendous efforts have been devoted to understanding this protein, the conformation of the catalytic core domain around the active site, particularly the catalytic loop overhanging the active site, is still not well characterized by experimental methods due to its high degree of flexibility. Recent studies have suggested that this conformational dynamics is directly correlated with enzymatic activity, but the details of this dynamics is not known. In this study, we conducted a series of extended-time molecular dynamics simulations and locally enhanced sampling simulations of the wild-type and three loop hinge mutants to investigate the conformational dynamics of the core domain. A combined total of >480 ns of simulation data was collected which allowed us to study the conformational changes that were not possible to observe in the previously reported short-time molecular dynamics simulations. Among the main findings are a major conformational change (>20 Å) in the catalytic loop, which revealed a gatinglike dynamics, and a transient intraloop structure, which provided a rationale for the mutational effects of several residues on the loop including Q^{148}, P^{145}, and Y^{143}. Further, clustering analyses have identified seven major conformational states of the wild-type catalytic loop. Their implications for catalytic function and ligand interaction are discussed. The findings reported here provide a detailed view of the active site conformational dynamics and should be useful for structure-based inhibitor design for integrase.

INTRODUCTION

The human immunodeficiency virus type-1 integrase (IN) is a key enzyme in the replication cycle of the HIV-1 virus (Asante-Appiah and Skalka, 1997). It is responsible for inserting the retroviral cDNA into the host genome, a required step for efficient viral replication. Because of its pivotal role in the replication cycle, IN has gained much attention as a novel target for drug development in recent years (Bushman, 1995; Chen et al., 2002a,b; Pani et al., 2002; Pluymers et al., 2001; Sayasith et al., 2001). However, drug discovery programs targeting IN have not been very successful. High-throughput screening programs have so far only identified one series of active compounds (Dayam and Neamati, 2003) and structure-based methodologies have been hampered by the lack of complete and detailed structural data. The difficulty in obtaining structural information of IN stems from the low solubility of the full-length IN, which presents a major challenge for x-ray and NMR analysis.

Despite the lack of detailed structural information, the basic outline of the IN's structural organization has been known for some time. Biochemical studies (Engelman and Craigie, 1992; Johnson et al., 1986) have revealed that IN is organized into three domains: the N-terminal domain, the core domain, and the C-terminal domain. Mutagenesis studies have located the catalytic site to the core domain and identified a conserved three-residue motif (D,D^{35}-E) common in many other retroviral INs and transposases (Engelman and Craigie, 1992; Kulkosky et al., 1992). Thus, the core domain has been the focal point of the structure-function investigations.

Based on functional analogy, a possible catalytic mechanism patterned after that of the *Escherichia coli* DNA polymerase I has been proposed (Beese and Steitz, 1991). However, crystallographic analysis of the core-domain has shown that the three-dimensional structure around the active site has a high degree of flexibility (Bujacz et al., 1996; Dyda et al., 1994; Greenwald et al., 1999), and the true active conformation of IN is still not completely determined. Without more detailed structural information about the structural and dynamical properties of the active site, the structural mechanism of the IN remains unsolved. One of our long-term objectives is to understand the structural mechanism through theoretical modeling. To help better define the scope of this study, we briefly summarize the relevant experimental and theoretical findings of the IN to date.

As a DNA manipulating enzyme, IN catalyzes two major reactions during the integration process: 3′-processing and strand transfer. In the first step, known as 3′-processing, just after the viral RNA genome is reverse transcribed into

Submitted December 21, 2004, and accepted for publication February 4, 2005.

Address reprint requests to Yong Duan, Tel.: 530-754-7632; Fax 530-754-9658; E-mail: duan@ucdavis.edu.

Matthew C. Lee's present address is U.S. Patent and Trademark Office, Alexandria, VA 22313.

Jinxia Deng's present address is Dept. of Pharmaceutical Sciences, University of Southern California, School of Pharmacy, Los Angeles, CA 90089.

0006-3495/05/05/3133/14 $2.00

doi: 10.1529/biophysj.104.058446

double-stranded linear cDNA (Brown et al., 1987; Lobel et al., 1989) by reverse transcriptase, IN recognizes two conserved nucleotides (5′-CA-3′) and removes two terminal nucleotides (5′-GT-3′) 3′, exposing the 3′-hydroxyl group of A for use in nucleophilic attack in the next step of catalysis (Gallay et al., 1995). After 3′-processing, IN stays bound to the viral DNA and forms a preintegration complex. In the second reaction, known as strand transfer or joining, IN catalyzes a concerted insertion of the viral DNA into the host genome using the previously exposed 3′-hydroxyl groups of the viral DNA. A 5-bp gap between the insertion points results in a 5-bp duplicate sequence flanking each side of the inserted provirus.

Because these two steps utilize the same active site and yet involve different substrates, it has been postulated that the active site conformation of IN may be different for each step. This view is supported by a recent study in which selective inhibition of the strand transfer reaction but not the 3′-processing reaction was observed (Espeseth et al., 2000). Further evidence of the conformational flexibility arose from the crystallographic analysis of the core-domain, which showed that the three-dimensional structure around the active site has a high degree of flexibility (Bujacz et al., 1996; Dyda et al., 1994; Greenwald et al., 1999).

In the available crystal structures of IN, the core domain consists of an α/β-fold with five β-strands at the center, sandwiched by six α-helices (Fig. 1). On each side of the β-core is a surface loop. The first loop (139-152) is located near the conserved catalytic triad containing several conserved residues. We refer to it here as the catalytic loop. Three conserved acidic residues (D^{64}, D^{116}, E^{152}) form a catalytic triad motif commonly found in other retroviral INs and transposases (Engelman and Craigie, 1992; Kulkosky et al., 1992) right below the catalytic loop (see Table 2 for schematics). Because of its proximity to the active site, we are most interested in the dynamics of this loop.

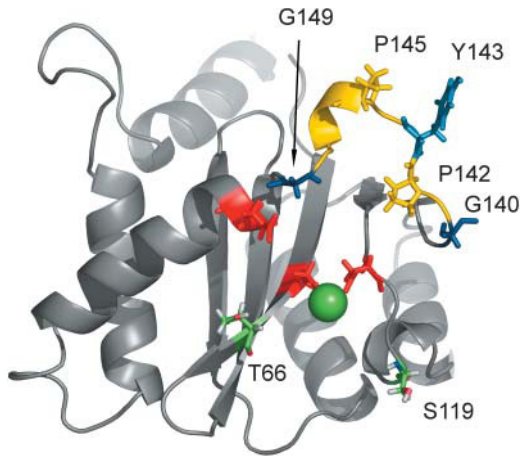

FIGURE 1 Schematics of the IN core domain. Backbone of the loop is colored in yellow (residues 141–148), the two hinge residues are colored in blue; the green sphere is the Mg^{2+}; the three conserved catalytic residues are highlighted in red; the two prolines in the loop are also highlighted in stick representation. The figure was prepared using the PyMol software (DeLano, 2002).

In a recent mutagenesis study on the core domain by Greenwald et al. (1999), the flexibility of the catalytic loop overhanging the active site was found to correlate with enzymatic activity. In that experiment, the glycine residues at each end of the catalytic loop (G^{140} and G^{149}) were mutated into alanines. The prevalent effect of the mutations is to decrease the loop flexibility as shown by crystallographic analysis that demonstrated the reduced temperature factors of the mutants in the loop region. Because of the reduced flexibility, G149A and G140A mutants exhibited ninefold and 18-fold reduction in activity, respectively, and the G140A/G149A double mutant was virtually inactive. Thus, it is now clear that the flexibility of the loop is essential for IN's enzymatic activity.

Yet, despite accumulation of several crystallographic structures of the core domain, this important catalytic loop has often been found to be either completely disordered or poorly ordered in crystals, and its structure and dynamics could have been influenced by the neighboring units through crystal contacts. Much less is known about the flexibility of the loop. What are the major modes of its native motion and how do the loop hinge mutations affect the native modes of dynamics? What are the conformations accessible to the loops in solution? How does the energy landscape look in the vicinity of the preferred (native) conformation? Answers to these questions are central to our understanding of IN's function. However, these questions cannot be easily derived from the crystal structures. It is also interesting to understand the crystal packing effect on the dynamics of exposed loops.

In this study, we conducted an extensive set of molecular dynamics (MD) simulations to explore the conformational space around the crystallographic native state of IN's core domain. To circumvent the limited conformational sampling ability of molecular dynamics simulations at room temperature, we chose a three-stage simulation strategy. In the first stage, we used multiple-trajectory short-time simulations. By combining the sampling ability of the multiple trajectories, we expect to sample more conformational space than single trajectory of the same length. In the second stage, we selected the most interesting trajectory in which the loop reached positions closest to the active site from the ensemble of the multiple trajectories and extended it to a much longer timescale to probe more fully the time-dependent aspect of the dynamics.

By using this two-tiered approach, we successfully followed the loop from its open state to a relatively stable closed state. The closed conformation was observed after ~8 ns of simulation in one trajectory and took another ~2 ns to become completely closed. Once closed, the loop remained in that state for nearly 30 ns. To further enhance the sampling around the closed conformation, we applied the locally

enhanced sampling (LES; Elber and Karplus, 1990) method to the closed state conformation. The enhanced sampling allowed us to observe the reopening event of the loop within ~4 ns of LES simulation time.

These simulations were further complemented by the simulations of the three loop hinge mutants (G140A, G149A, and G140A/G149A). Our results indicated that this large-scale loop motion was not observed in the mutant structures within the simulation timescale. Comparison of the major conformational states sampled by the three mutants to that sampled by the wild-type showed that the differences are mainly concentrated in the catalytic loop region. Because the gating motion was significantly hampered (completely eliminated in the case of the double mutant), we believe that this conformational dynamics is functionally relevant and is likely to play a role in catalysis.

In the following sections, we present the details and results of our simulation and discuss the implications of our findings in the context of IN structure-dynamics-function relationship and IN specific inhibitor design.

THEORETICAL METHODS

Model building

Four molecular systems were constructed, the wild-type IN, the G140A mutant, the G149A mutant, and the G149A/G140A double mutant. The wild-type model was constructed from chain B of the crystal structure 1QS4. This structure contains one Mg^{2+} ion coordinated by D^{64} and D^{116}, but has two unresolved residues (I^{141} and P^{142}) in the catalytic loop. Although, two metal ions are thought to be present for the catalysis (the second one is thought to be between E^{152} and D^{64}), we chose to include single Mg^{2+} in the simulations to probe the dynamics of the core domain in its apo-state when

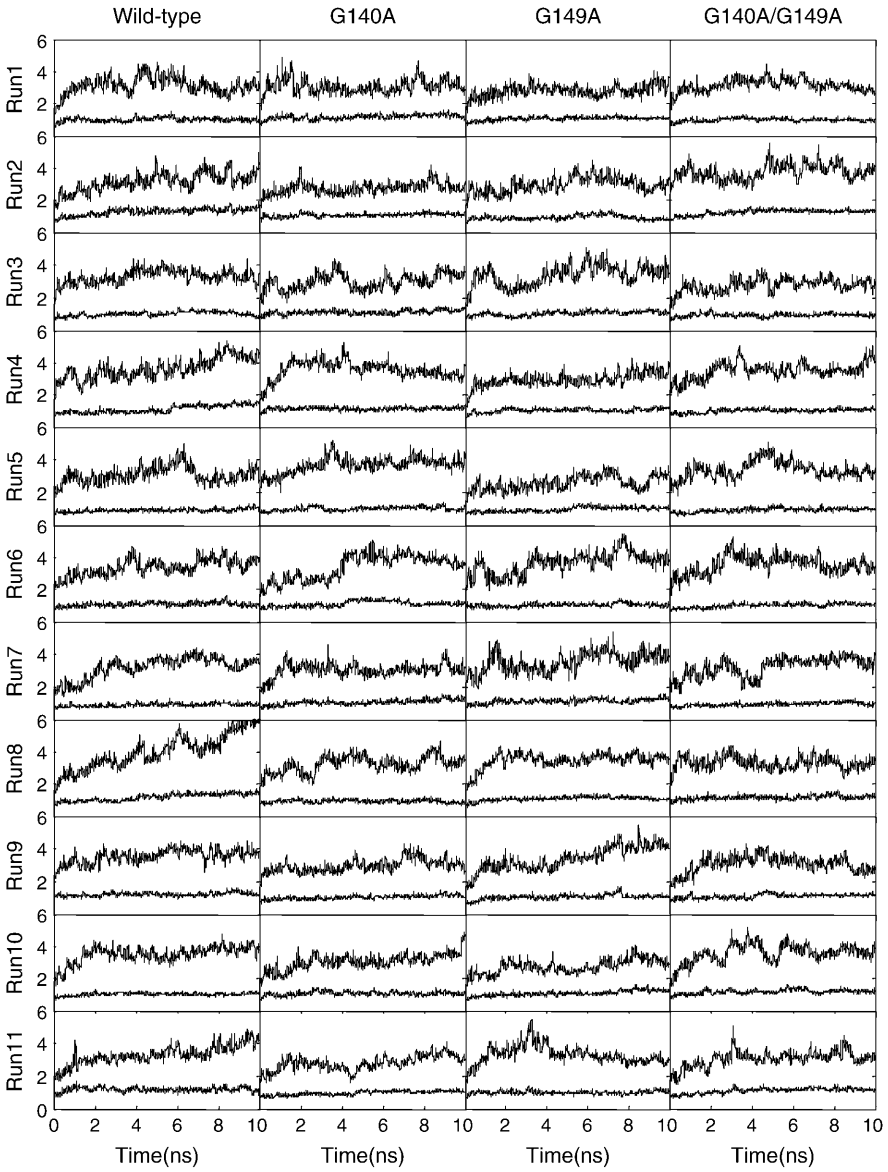

FIGURE 2 RMSD versus time. The horizontal axes represent RMSD values in Å, and the vertical axes represent time in nanoseconds. The RMSDs were calculated using C_α atoms only and fitted to the initial structure of each simulation. The black lines represent the RMSD of the whole protein, and the shaded lines represent the RMSD without the loop region (residue 140–149, 186–196).

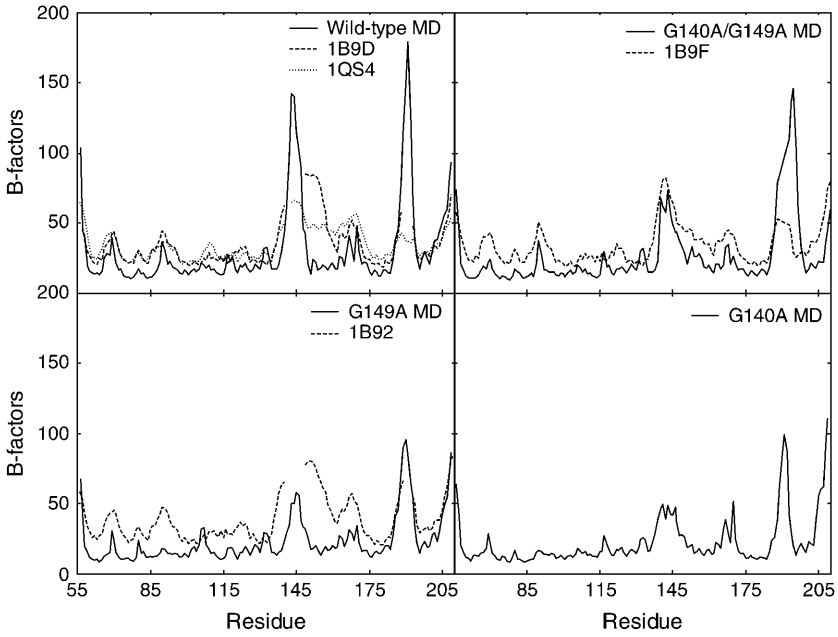

FIGURE 3 Calculated B-factors from the MD simulations and the B-factors obtained from the indicated Protein Data Bank file. Horizontal axes are residue indices, and vertical axes are B-factors in Å^2.

only one metal ion is bound. The missing coordinates were homology modeled from chain B of 1BIS, which contains these two missing residues, using the SwissPDB software (Guex and Peitsch, 1997). The completed models were then subjected to 200 cycles of steepest descent energy minimization using the GROMOS force field available in the SwissPDB software, allowing only residues 140–143 to move while holding the rest of the protein fixed. The *tleap* module in AMBER7 (Case et al., 2002) was used to prepare the input files for simulations. The Duan et al. (2003) force field (also known as ff03) was selected to represent the molecular mechanical potentials. Our experience has shown that this force field has a reasonable balance between the α-helix and β-strand conformations and therefore is more appropriate for extended-time simulations of proteins whose structure contains both α- and β-structural elements. Hydrogens were added with the *tleap* program as well, and protonation states of the ionizable side chains were assigned according to values previously predicted (Lins et al., 1999) by the UHBD (Madura et al., 1995) program. (The neutral H^{114} was protonated at the ε-position, whereas the rest of the ionizable side chains were kept at their standard protonation states.) The completed model was then solvated with TIP3P water (Jorgensen et al., 1983) in a box measuring $73 \times 77 \times 60$ Å^3. The system was neutralized with two Cl^- ions. Both mutations introduced in the 1QS4 structure to aid in crystallization (F^{185K} and W^{131E}) were changed back to their wild-type identities since we are interested in the dynamics of the domain in its native form. The parameters of Cl^- and Mg^{2+} ions were taken from the standard AMBER database.

The three mutant models were constructed by replacing the appropriate residues in the above wild-type model with the mutant residues using the SwissPDB software. These mutated structures were then energy minimized, solvated, and neutralized as outlined above.

Explicit solvent molecular dynamics simulations

The initial solvated structures were first subjected to 200 steps steepest descent energy minimization, whereas the solute atoms, including both the protein and the Mg^{2+} ion, were restrained by a harmonic potential with a force constant of 100.0 kcal/mol/Å^2. After the initial solvent minimization, the entire system was minimized using 200 steps of steepest descent minimization without harmonic restraints.

The minimized structures were then subjected to an equilibration protocol in which the temperature of the systems was gradually raised from 100 K to 300 K over a 10-ps period while holding both the volume and temperature constant, followed by another 10-ps of solvent density adjustment at 300 K by holding the temperature and pressure constant while allowing the volume to change. The initial velocities were assigned randomly from a Maxwellian distribution at 100 K. At the end of the equilibration, the average temperature of the final 5 ps was ~300 K, and the average density was ~1.0 g/ml. Long range electrostatic interactions were treated with the particle mesh Ewald (Darden et al., 1993) method. Periodic boundary conditions were applied via both nearest image and the discrete Fourier transform implemented as part of the particle mesh Ewald method. All bonds involving hydrogen atoms were restrained using the SHAKE (Ryckaert et al., 1977) algorithm, allowing larger time steps (2 fs) to be taken. Global translation and rotation of the system (solvent and solute) was removed every 100 integration steps during the simulation.

The initial 20-ps stage was designed to equilibrate those particles that were added during the initial model-building process, including water molecules and hydrogen atoms, and to allow the systems to be solvated adequately. It was not intended to bring the system into ''thermodynamic equilibrium''. Therefore, the initial 20-ps trajectories were discarded and were followed by the production stage in which both pressure (1.0 ATM) and temperature (300 K) were held constant by Berendsen's coupling scheme. This allowed us to monitor the conformational transitions from the early stages. On the other hand, the first 5 ns of trajectories were excluded from more quantitative analyses, such as B-factor calculation, population, and distribution, to allow adequate equilibration of the system. In the latter cases, the effective equilibration phase was more than 5 ns in each trajectory.

A set of 44 independent simulations (11 for each model system) was conducted using the same simulation protocol. These multiple trajectories allow us to study events that are one order of magnitude slower than the simulation time of the individual trajectories. In this case, the 11-trajectory sets allowed us to study the events of ~100-ns timescale. The differences among the trajectories for each system were the initial velocities, which were assigned by choosing different random number seeds. These trajectories sampled different regions in the phase space and conformational space, as clearly shown in root mean-square deviation (RMSD) plots (Fig. 2). These multiple trajectories were intended to examine the consistency of the

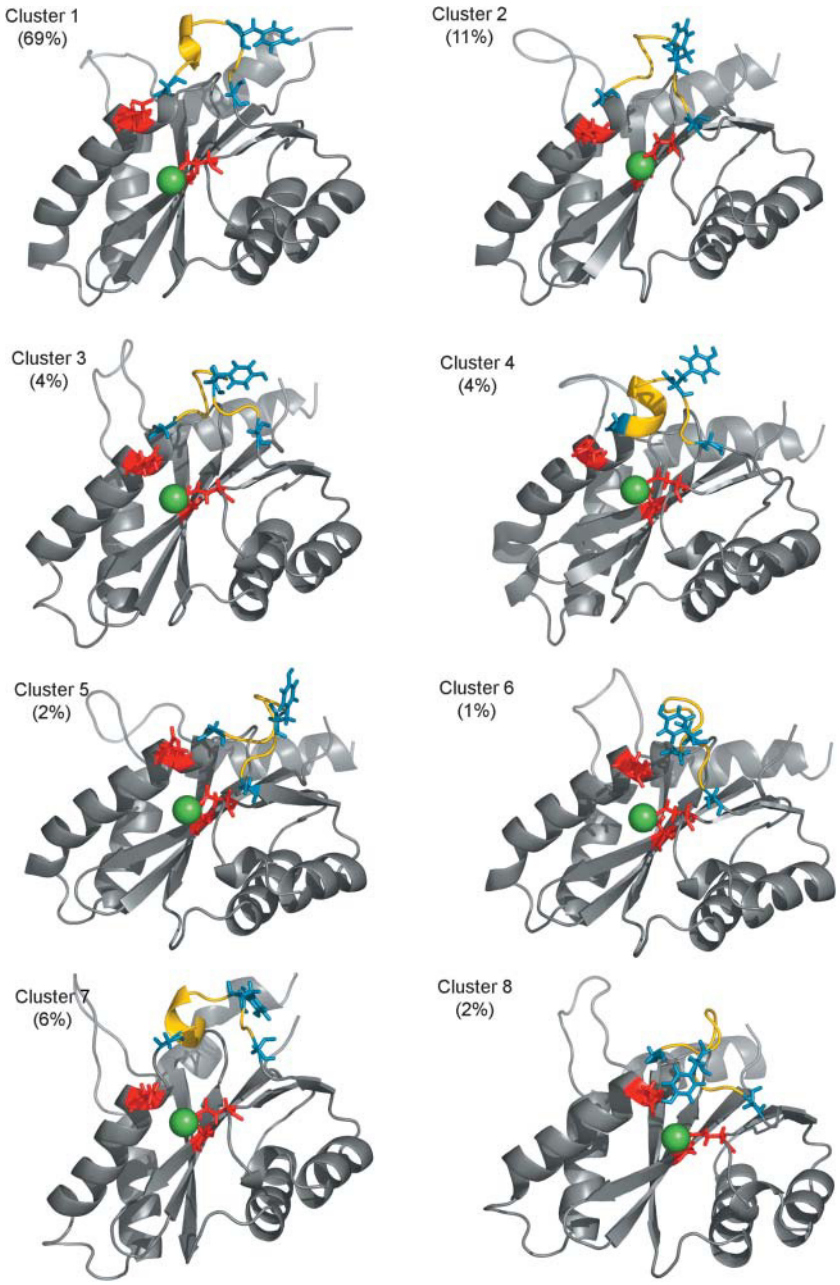

FIGURE 4 Representative structures from the 11 multiple-trajectory MD simulations of the wild-type. The percentage values indicate the percentage of snapshots belonging to each cluster.

observations and to maximize the sampling ability with limited simulation timescales.

Explicit solvent molecular dynamics simulations with LES

In this set of simulations, we took one structure from each model as our starting structures for LES simulations. The wild-type structure was taken from the final snapshot of the 40-ns extended MD simulation. Because none of the mutant loops were closed, we arbitrarily took the final structures of the last run (run 11) as starting points. We then divided the loop into three regions (140-143, 144-146, and 147-149) and replaced each segment with five duplicate copies by using the *addles* module in AMBER7. All copies belonged to the same region and had the same initial conformation as that of the template structure but were given different initial velocities to permit divergence. Four nanoseconds of data were collected for each model.

Conformation clustering

A heuristic clustering approach was used to characterize the snapshots of the MD simulations based on the C_α-RMSD of the loop region. In this method, a snapshot may become a member of its closest cluster if the C_α-RMSD is smaller than a given cutoff (1.5 Å), otherwise a new cluster is created. The C_α-RMSD was calculated after rigid body alignment of the loop and the two bracing secondary elements (the β-strand N-terminal to the loop and the α-helix C-terminal to the loop) with respect to the average structure of the cluster. This method is semilinear and is rather efficient for clustering large data sets. However, it is a heuristic method and is inherently approximate.

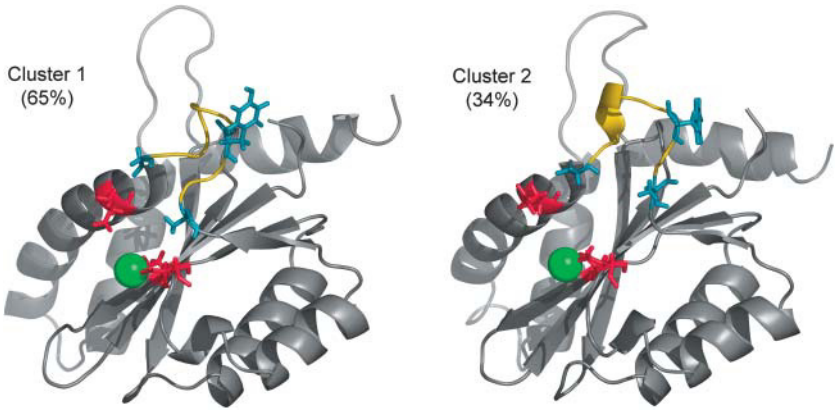

FIGURE 5 Representative structures from the 11 multiple-trajectory MD simulations of the G140A mutant. The percentage values indicate the percentage of snapshots belonging to each cluster.

The error margin is related to the cutoff used in the clustering; clusters generated using a small cutoff can be highly accurate. To further enhance the accuracy, the clusters were filtered by removing structures that were far away from the average of the cluster as measured by the RMSD. The removed snapshots were then compared to the existing clusters.

RESULTS AND DISCUSSION

Multiple trajectory (10-ns) results

Fig. 2 shows the main-chain RMSD of the protein core, excluding the loop regions, and of the whole protein for all 44 trajectories. The core RMSD averaged over individual trajectories ranged from 0.89 Å to 1.27 Å with an overall 1.05 ± 0.09 Å when averaged over all 44 trajectories. The RMSD indicated that the overall structures were well maintained; the protein core was quite stable throughout the simulations. This high stability is clearly demonstrated by the B-factors calculated from the simulations (discussed later). On the other hand, the loops showed notably higher degrees of fluctuations, as exemplified by the overall RMSD of the whole protein. When the entire protein was considered, the average main-chain RMSD ranged from 5.00 Å to 2.56 Å for an overall average of 3.27 ± 0.41 Å. The large RMSD indicates that loops underwent conformational changes.

Crystal structures of the wild-type (1B2D), G149A mutant (1B92), and G149A/G140A double mutant (1B9F) were solved by Greenwald et al. (1999). The structure of the single mutant G140A has yet to be solved due to poor crystallization. In Fig. 3 we compared the crystallographic B-factors to those calculated from our MD trajectories by $B = (8\pi^2)/(3)\langle \Delta r^2 \rangle$ where $\langle \Delta r^2 \rangle$ is the mean-square atomic fluctuation averaged over the last 5 ns of the 11 simulations on each of the wild-type and mutants. The correlation between the experimental values and the MD calculated values, excluding the loop regions, were in the range of 0.7 to 0.8.

One notable difference is found in the interface loop region where the crystallographic B-factors in all of the crystal structures are generally lower than the MD B-factors. This is mainly due to crystal packing effects. In the case of 1QS4, the interface loop is packed against another subunit. In the case of 1B9D and 1B9f, the loops are packed against structures from neighboring unit cells due to crystallographic symmetry. By comparison, the MD simulations mimicked the protein in solution phase; there were no protein neighbors to form stabilizing interactions with the interface loop, hence the interface loop was allowed a greater degree of freedom.

Although the B-factor offers a convenient yardstick by which one can compare experimental and simulation results, one should note that the crystallographic B-factor measures both the thermal fluctuation (including conformational heterogeneity of the proteins) and global translation and rotation. In our calculated B-factors, however, the global translation/rotation has been removed by rigid-body alignment. Therefore, one might expect that the calculated B-factors should be smaller than the experimental ones. However, one should also consider other factors, such as crystal packing, which can play a role and can reduce the flexibility of the parts involved in the crystal contacts. Notwithstanding these differences, comparison with crystallographic B-factors offers a qualitative assessment of the simulations.

To summarize the nearly 40,000 snapshots, we performed a clustering analysis using the heuristic algorithm outlined in the Methods section. Note that in this analysis, only the conformational states of the catalytic loop were clustered to highlight the conformational transitions of the loop observed in the simulations. The clustering was based on the C_α-RMSD computed by aligning the loop and the two bracing secondary elements, β-5 (residues 135–149) and α-4 (residues 150–160), from each end of the loop.

Figs. 4–7 show the representative structures of the clusters identified. As shown in the figures, the majority of the loop conformations are distributed in various types of open conformations. Among the clusters generated from the wild-type simulations (Fig. 4), clusters 6 and 8 are in the closed form. The main difference between these two clusters is at the P^{142}/P^{145}. The loop conformational change in cluster 8 is mainly localized on residue P^{142}, whereas in cluster 6, both P^{142} and P^{145} have changed position.

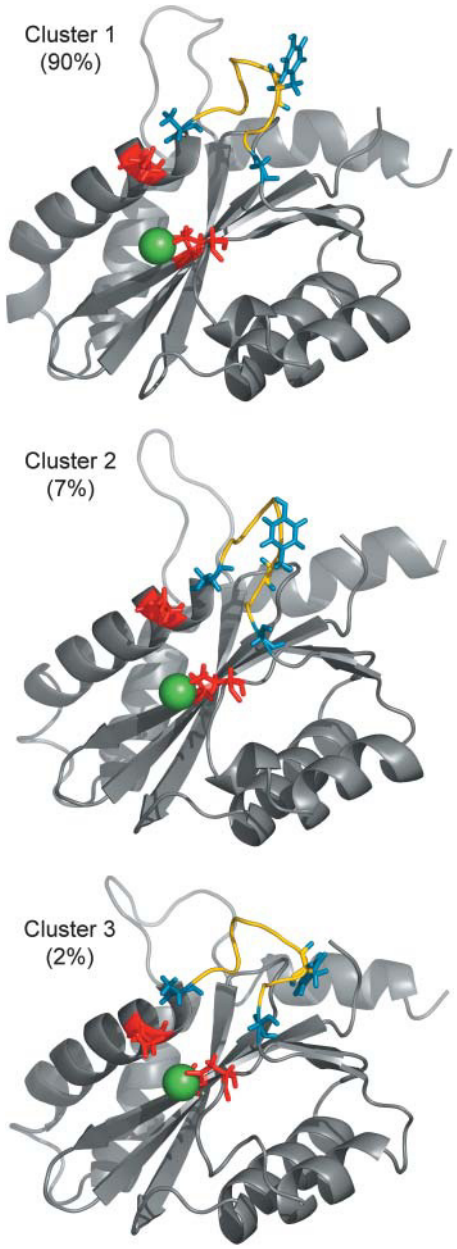

FIGURE 6 Representative structures from the 11 multiple-trajectory MD simulations of the G149A mutant. The percentage values indicate the percentage of snapshots belonging to each cluster.

The closed conformation makes up ~3% of all snapshots (Fig. 4, *clusters 6* and *8*). Given a total of 110 ns simulations, we would estimate a time constant of ~3.6 μs, assuming a two-state kinetics. Thus, the event is a relatively slow process. On the other hand, the small fraction of closed conformation observed in the LES simulation (discussed later) suggests that the loop favors the open conformation. This is in agreement with the currently available crystallographic structures whose resolved portion of the loop is mostly found in the open conformation.

Another interesting conformation is cluster 7 shown in Fig. 4 in which the loop is bent backward with Y^{143} pointing away from the active site, making contact with the hydrophobic patch consisting of residues I^{60}, V^{79}, and A^{80}. This provides a possible explanation for the reduced activity of Y143L mutation (Table 1). A possible scenario is that the L^{143} of the mutant makes stronger contact with the hydrophobic patch and stabilizes the loop into the open conformation.

Extended-time MD results

To explore the loop motion in greater detail, we extended the wild-type trajectory that showed the closed conformation to ~40 ns (including the initial 10 ns) and closely monitored the conformational state of the loop. Fig. 8 *a* shows the time course of C_α-RMSD of this extended wild-type simulation. To facilitate tracking the position of the loop, we defined two planes (one for the loop, one for the active site) and monitored the angle and distance between the two planes as a measure of the gating motion. The loop plane is made up of the two hinge-C_α atoms and the Y^{143}-C_α atom; the active site plane is made up of the two hinge-C_α atoms and the Mg^{2+} atom. Fig. 8 *b* shows the gate behavior as a function of time.

Starting from the initial open conformation, the loop began to dip forward toward the catalytic triad at ~8 ns. By 10 ns, the loop was completely bent over and assumed a closed conformation. It remained in the closed conformation for another 10 ns and then opened slightly for a brief moment at ~20 ns before closing again for another 20 ns. We stopped the simulation after seeing that the loop attempted to open up (see Fig. 4, *cluster 5*) for a second time at ~40 ns when the loop appeared to move upward but did not return to the completely open position found in the initial structure.

To characterize the snapshots of this long simulation, we also performed a clustering analysis using all snapshots from this trajectory. Four major clusters were found in this trajectory, two in the closed state and two in the open state. Fig. 9 shows the backbones of these four states superposed on one another and the position of Y^{143} in each structure.

In extending the simulation, our goal was to gain a more complete view of the loop dynamics and to gain insights into the perturbations due to the mutations. To facilitate our analysis of the motion, we applied essential dynamics (ED) analysis. ED analysis (Amadei et al., 1993) is a motion analysis technique that was developed specifically to reduce the complexity of the data to aid in the extraction of meaningful insights. In this analysis, the principal components of the protein's motion are extracted from the variance-covariance matrices $\Delta A_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle$ where x_i and x_j are, respectively, the ith and jth dimension of the coordinates from which the fast modes of local motions and thermal fluctuations are filtered out to reveal the slower, correlated modes of motions that are more likely to be relevant to biological function.

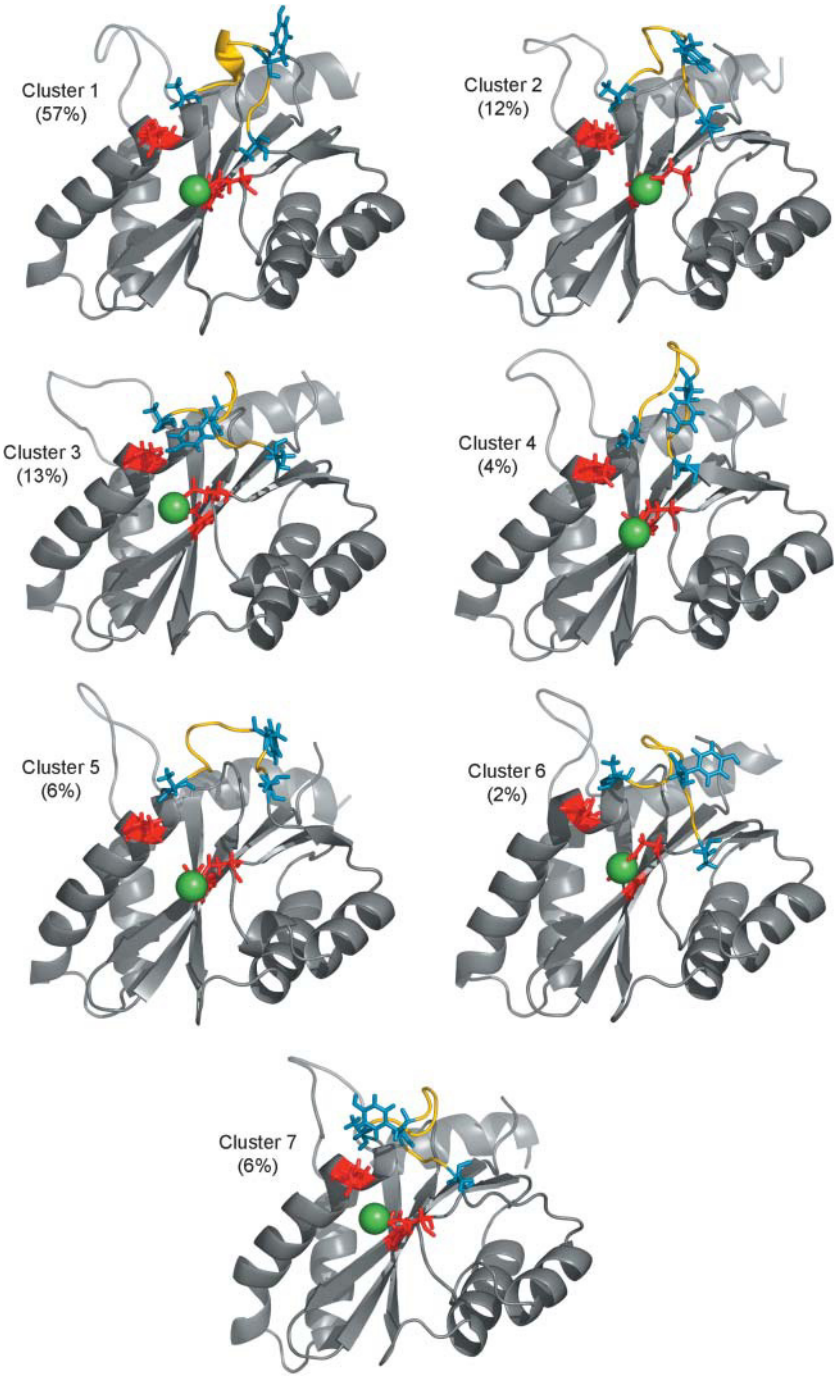

FIGURE 7 Representative structures from the 11 multiple-trajectory MD simulations of the G140A/G149A double mutant. The percentage values indicate the percentage of snapshots belonging to each cluster.

In our analysis, we selected the C_α atoms to describe the backbone conformations, and the variance-covariance matrices were calculated by averaging over the simulations. The results of the ED analysis are the collection of eigenvectors that represent the correlated modes of motion and their associated eigenvalues that specify the amplitudes of the motion. Earlier studies have shown that, when reduced to the essential space, the first few eigenvectors are sufficient to describe most of a protein's correlated motions (de Groot et al., 1996). Herein we present only the first essential modes for the extended-time simulations. For comparison, we also show the representative ED modes for the three mutants.

In Fig. 10, we show the maximum (*red*) and minimum (*yellow*) projected structures of the first essential mode for each of the four systems. They were superimposed to highlight the extrema of each essential motion. In the wild-type simulation, the primary mode of motion is the opening and closing of the catalytic loop. The red and yellow spheres in the middle of the loop represent Y^{143} at the minimum and maximum positions, respectively. This residue occupies the

TABLE 1 Catalytic activity of various IN core domain single-point mutants found in the literature.

Res	Mut	Cl	IN	DIS	Res	Mut	Cl	IN	DIS
52	G-52V*	100	100	100	120	N-120S§	165	148	120
53	Q-53K*	100	100	100		N-120Q§	103	78	102
	Q-53C†	100	100	50 ~ 75		N-120L‡	>50	10 ~ 50	>50
	Q-53L‡	>50	>50	>50	121	F-121I*	100	10 ~ 50	100
58	L-58V*	100	100	100	123	S-123A¶	71	68	136
59	G-59S*	100	100	100	125	T-125A††	ND	100–20	ND
61	W-61R*	0	0	0	127	K-127T‡	>50	>50	>50
62	Q-62A§	4.5	7.6	7.6	131	W-131N*	100	>100	100
	Q-62N§	19	11	6.3	134	G-134D*	100	100	100
	Q-62E**	20 ~ 50	20 ~ 50	50 ~ 100	136	K-136R**	50 ~ 100	50 ~ 100	50 ~ 100
	Q-62A**	5 ~ 20	5 ~ 20	50 ~ 100		K-136E**	50 ~ 100	50 ~ 100	50 ~ 100
64	D-64V‡	10 ~ 49	0	0		K-136A**	1 ~ 5	1 ~ 5	50 ~ 100
	D-64N¶	<0.3	0.5	0.3	140	G-140A‡‡	ND	ND	6
	D-64E¶	<0.3	0.6	0.3	142	P-142V‡	>50	>50	>50
	D-64V§§	0	0	0	143	Y-143G¶¶	100	100	100
66	T-66A§	22	53	49		Y-143L‡	>50	>50	>50
	T-66A††	ND	60–80	ND		Y-143D*	100	100	100
	T-66A¶	22	42	91	144	N-144V‡	>50	>50	>50
	T-66A‡	>50	>50	>50	145	P-145I‖	0	0	0
67	H-67S§	140	133	52	146	Q-146R |	100	100	100
75	V-75P††	ND	ND	ND	147	S-147A‡	>50	>50	>50
77	V-77L*	10 ~ 50	10 ~ 50	10 ~ 50	148	Q-148L§	0	0	17
78	H-78R*	0	0	0		Q-148L‡	<10	<10	>50
80	A-80S*	100	100	100	149	G-149A‡‡	ND	ND	11
81	S-81R*	ND	ND	ND	150	V-150E‖	80 ~ 100	25 ~ 80	80 ~ 100
	S-81R†	active	Active	10 ~ 49	151	V-151A¶¶	ND	ND	ND
	S-81A‡	>50	>50	>50	152	E-152V‖	0	0	0
85	E-85W*	100	100	100		E-152G†	trace	trace	trace
87	E-87Q*	100	100	100		E-152D¶	<0.3	1.8	7.2
90	P-90D‖	20 ~ 80	<5	0		E-152Q¶	<0.3	<0.1	<0.1
92	E-92A§	25	24	49		E-152L‡	0%	0	0
	E-92N§	24	26	61	153	S-153R¶	24	22	48
	E-92A**	50 ~ 100	50 ~ 100	50 ~ 100	155	N-155E§	3.8	7.9	5.6
	E-92Q**	50 ~ 100	50 ~ 100	50 ~ 100		N-155K§	2.2	8.1	2.8
	E-92K**	5 ~ 20	5 ~ 20	50 ~ 100		N-155L§	18	16	12
93	T-93A††	ND	25–55	ND	156	K-156I‖	0	0	<5
	S-93P*	100	10 ~ 50	100		K-156E§	11	9	7.9
	S-93A‡	>50	>50	>50	158	L-158F‖	25 ~ 80	80 ~ 100	25 ~ 80
103	K-103Q‡	>50	>50	>50	159	K-159N§	21	23	17
104	L-104P*	0	0	0		K-159S§	26	36	22
	R-107A*	100	100	100		K-159V‡	>50	10 ~ 50	>50
107	R-107L‡	>50	>50	>50	166	R-166L‡	>50	>50	>50
109	P-109N*	100	100	100	172	L-172M‖	80 ~ 100	25 ~ 80	25 ~ 80
110	I-110R*	100	100	>100	177	L-177Q*	100	>100	100
111	T-111I*	100	100	100	179	A-179P¶¶	ND	ND	ND
114	H-114S‡	>50	>50	>50		F-185K |;**	100	100	100
115	T-115A¶	80	95	140	185	F-185K§	100	>100	100
	T-115V‡	>50	>50	>50	186	K-186Q‡	>50	>50	>50
116	D-116I†	0	0	0	195	S-195A¶¶	ND	60–80	ND
	D-116N¶	<0.3	<0.1	<0.1		R-199C†	100	100	100
	D-116E¶	1.7	4.8	38	199	R-199S‡	>50	>50	>50
	D-116I‡	0	<10	0	206	T-206A‡	>50	>50	>50
117	N-117S§	30	35	39	211	K-211N¶¶	ND	32	ND
	N-117Q§	40	59	73	219	K-219N‡	>50	>50	>50
	N-117Q¶	26	57	99	228	K-228I‡	10 ~ 50	10 ~ 50	>50
	N-117I‡	>50	>50	>50					

The abbreviations for each of the columns are Res, Residue number; Mut, nature of the mutation; CI, 3′-processing activity; IN, strand transfer activity; DIS, disintegration activity. Numbers are in percent activity.

*van den Ent et al. (1998).
†Leavitt et al. (1993).
‡van Gent et al. (1992).
§(Gerton et al. (1998).
¶Engelman and Craigie (1992).
‖van den Ent et al. (1998).
**Engelman et al. (1997).
††Cannon et al. (1994).
‡‡Greenwald et al. (1999).
§§Drelich et al. (1992).
¶¶Tsurutani et al. (2000).
‖‖Sayasith et al. (2000).

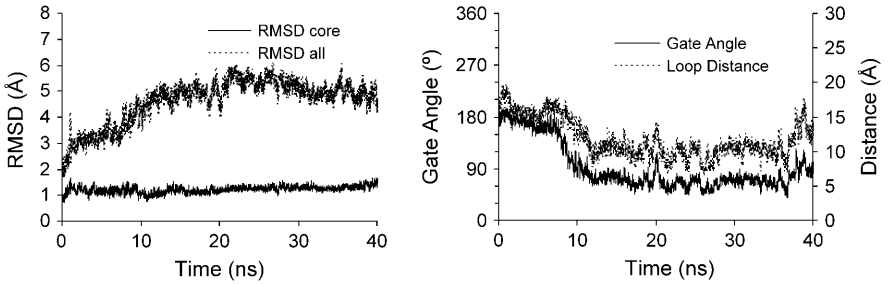

FIGURE 8 (*a*) RMSD versus time plot of the 40-ns wild-type simulation. The black line represents the C_α-RMSD including the loops, and the shaded line represents the C_α-RMSD excluding the loops. (*b*) Loop position versus time. The shaded line represents the angle of opening of the gate, and the black line represents the distance between the tip of the loop to the center of the catalytic site.

tip of the loop and has been implicated in catalysis (Beese and Steitz, 1991; Chen et al., 2000; Esposito and Craigie, 1998; van Gent et al., 1993). In the ''loop closing'' mode, the Y^{143} was brought closer to the other three catalytic residues (also depicted in spheres of *red* and *yellow*) to form a four-point active site configuration. The distance between the C_α atoms of Y^{143} in the two projected structures is 13.8 Å. This large-scale motion has never been observed in previous MD simulations largely due to the smaller timescales sampled in those simulations, and hence represents a new possible active site conformation. The blue and magenta spheres represent the positions of the two loop hinges, and the bottom two green spheres represent the positions of T^{66} and S^{119}. We show these two residues as anchoring points around the active site to facilitate the visual comparison.

In comparison to the wild-type, the catalytic loop in the three mutant systems showed significant reduction in mobility. The G149A single mutant still retained some residual wild-type loop motion with a 7.07 Å maximum separation between the open and closed conformations (C_α of P^{145}), but the catalytic loop in the G140A mutant was virtually stationary and the double mutant showed a slight backward bending motion that formed a more open active site configuration.

The geometry of the IN catalytic loop is asymmetric and contains one proline residue near each end of the loop (P^{142} and P^{145}). Because proline side chains are rigid in nature, these two residues are responsible for much of the ordered internal structure of the loop. In the open conformation of the loop, the C_γ atom of P^{142} points toward the active site, whereas the C_γ atom of P^{145} points away from the active site. In the completely closed conformation, the orientations of these two residues become reversed. As in most β-α types of loops (Oliva et al., 1997), the region near the C-terminal end, around P^{145}, has a more helical character and the region near the N-terminal end, around P^{142}, has more of an extended and β-character. During the simulations, residues P^{145}, Q^{146}, and S^{147} form a transient 3_{10}-helix ~20% of the time in the wild-type, 50% in G149A, 37% in G40A, and 12% in the G140A/G149A double mutant. Available mutagenesis data show that P^{145} and Q^{148} are particularly sensitive to mutations, suggesting that the transient 3_{10}-helical region plays a role in IN's function. Because of the extensive hydrogen-bonding interactions involved, we hypothesize that the hydrogen-bonding network in this region might contribute to the loop motion.

The role of Y^{143} in catalysis has received some attention recently. Based on the proposed structural arrangement of the active site of *E. coli* polymerase I (Beese and Steitz, 1991), it has been suggested that the role of Y^{143} may be to similarly stabilize the activated water molecule. Mutagenesis studies have shown that mutations at this position shift the preference of nucleophile during the 3′-processing reaction from water to alcohol (van Gent et al., 1992, 1993; Vink et al., 1991), which appears to support this hypothesis.

In our simulations, we observed that the wild-type Y^{143} side chain has a high degree of mobility as shown in Fig. 11. Interestingly, in the three mutant simulations, the orientations of Y^{143} are predominately pointing toward the active site. This change of orientation was previously interpreted as

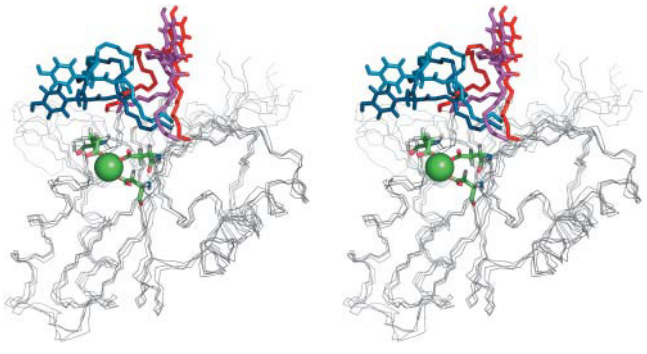

FIGURE 9 Wall-eyed stereogram of the four representative conformations identified from the 40-ns wild-type simulation. The catalytic loop is color coded to correspond to the color coding scheme of Table 2 (*maroon*, cluster 1; *blue*, cluster 2; *red*, cluster 3; and *pink*, cluster 4). The three conserved catalytic residues are colored by atom (C, *green*; O, *red*; and H, *white*). The Mg^{2+} ion is shown in sphere representation.

TABLE 2 Number of snapshots and relative populations identified from the 40-ns wild-type trajectory

	Wild-type (long)	
Cluster ID	# Snapshot	% of total
1	1271	31%
2	2059	50%
3	566	14%
4	237	6%

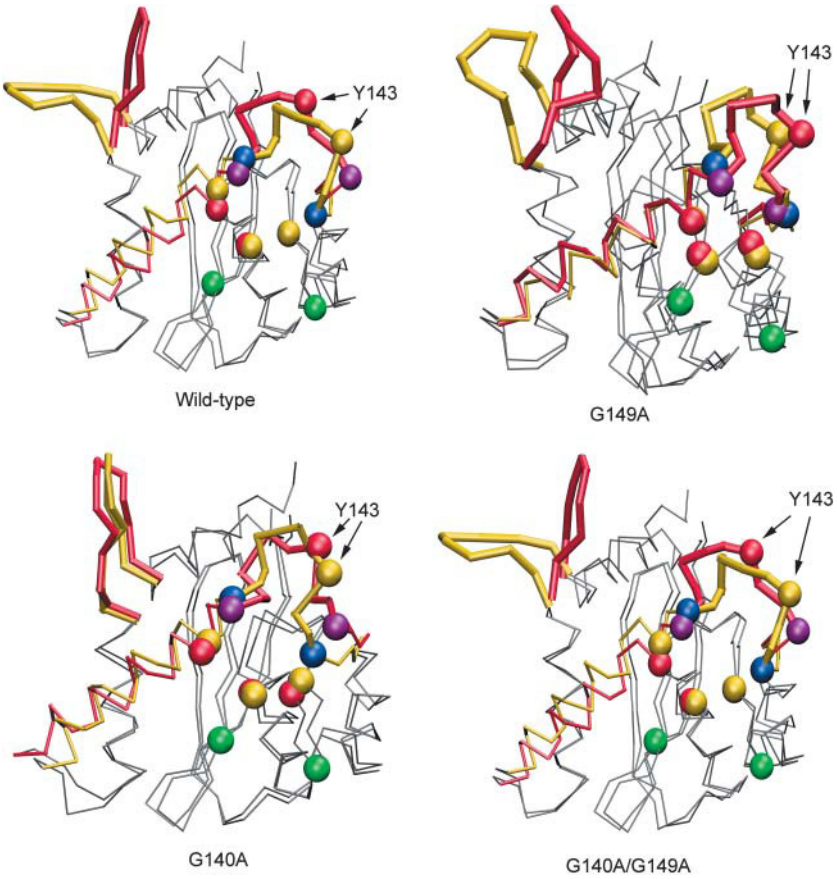

FIGURE 10 First essential modes from ED analysis. The figures were generated by projecting the entire molecular dynamics trajectory onto the first eigenvector of each respective trajectory. The minimum structure is colored in yellow, and the maximum is red. The motions of each of the shown essential modes are bounded by these two extrema. The red and yellow spheres in the middle of the loop show Y^{143} at the minimum and maximum positions, respectively. The blue and magenta spheres represent the positions of the two loop hinges. The two green spheres are T^{66} and S^{119}. The three catalytic residues are shown in red and yellow spheres.

evidence that Y^{143} plays a significant role in catalysis, and the inward-pointing conformation of Y^{143} was generally assumed to be the active conformation. Here our simulation suggests that the loop flexibility is necessary to position Y^{143} such that it is in close proximity to the substrate DNA when the loop is in the closed form. Thus, the function of Y^{143} is closely linked to the dynamics of the loop.

Closed to open transition by LES

The limited simulation time, although already considerably long in comparison to the typical simulations on systems of similar size, precluded the possibility of the observation of reversible open-close events. This could largely be due to kinetically trapped closed or open conformations. In the extended-time simulation, once the loop entered into the closed conformation, it never completely opened up again throughout the 40-ns trajectory. In other 10-ns simulations, despite the large degree of conformational flexibility in the loop regions, reversible conformational changes were rare because of the timescale. One interesting question was the following: can the loop in the completely closed conformation open up again? This question can be more effectively addressed by advanced sampling techniques such as LES (Elber and Karplus, 1990; Roitberg and Elber, 1991) without resorting to exceedingly long conventional MD simulations. The LES method is based on a mean-field theory in which enhanced sampling is achieved through making multiple copies of parts of the protein. As a result of enhanced sampling, the barriers separating local minima in the LES calculation are lower, which enhances the probability of barrier-crossing events. In these simulations, the multiple copies are allowed to diverge in the simulations by assigning different initial velocities.

The starting structures of the LES simulations were taken from the final structures from either the extended simulations or from one of the trajectories. These structures were already well equilibrated in terms of side-chain orientation and solvent environment because they were subjected to at least 10 ns of MD simulation. We observed that the divergence of the copies in all four systems leveled off after ~1 ns of simulation, and in the case of the double mutant, there was even a slight decrease (data not shown). In the wild-type LES simulations, we observed that the loop returned to its open conformation after 1.5 ns and remained open for the remainder of the simulation. In the three mutants, the LES simulation also resulted in the open form of the loop. In short, the LES simulations suggest that the open conformation is the preferred state of the catalytic loop. This is in agreement with the experimental observations that the loops are in the open state. Nevertheless, our simulations, starting

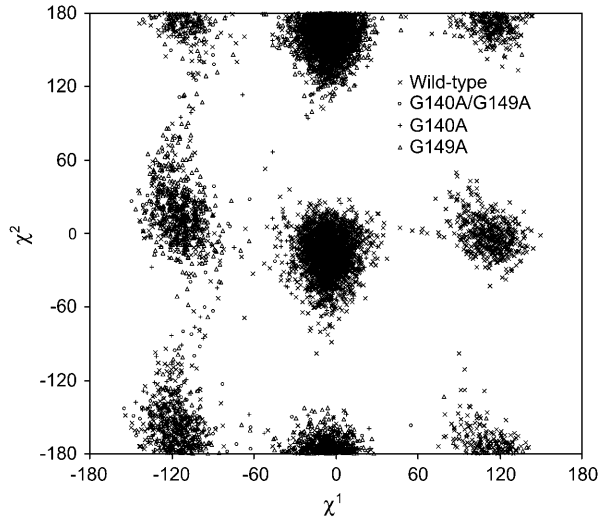

FIGURE 11 Y^{143} χ-angle distribution. Positive χ_1 corresponds to the phenol group of Y^{143} pointing away from the active site, and positive χ_2 corresponds to the hydrogen of the hydroxyl group on the phenol ring pointing away from the active site.

from the open conformation, were able to sample the closed conformation.

The mobility observed in the simulations highlights the need to take into account protein dynamics to design effective inhibitors. In the case of IN, the conformational flexibility is intimately linked to the active site conformation. The high degree of flexibility implies plasticity of the active site and the ability to accommodate changes (e.g., mutation). It also implies that the active site may undergo considerable conformational changes upon binding to ligands. If so, one may wish to consider the correlated motions between ligands and IN to design effective inhibitors. In the closed conformation, the loop actually forms a canopy overhanging the active site, forming a deeper pocket than the open form. We speculate this conformation may be used to identify ligands that may stabilize the closed conformation of IN. The benefits of such a ligand are twofold. In addition to the obvious benefit that it may be a lead compound for further inhibitor design, it may also serve as an experimental tool to investigate the dynamics of the loop.

In light of the high degree of conformational variability exhibited by the loop, one may conclude that the functionally important catalytic loop is much less ordered than the rest of the protein. One may also anticipate that binding to the substrate can reduce the mobility and make the loop more ordered. Because of close proximity, the active site significantly changed its shape when the loop moved from the open form to the closed form (Fig. 12). Clearly, for efficacy of drug design, these structural variations should be explored.

We superimposed 5CITEP onto the binding site of the wild-type open and closed conformations obtained from clustering analysis. The location of 5CITEP also overlaps in

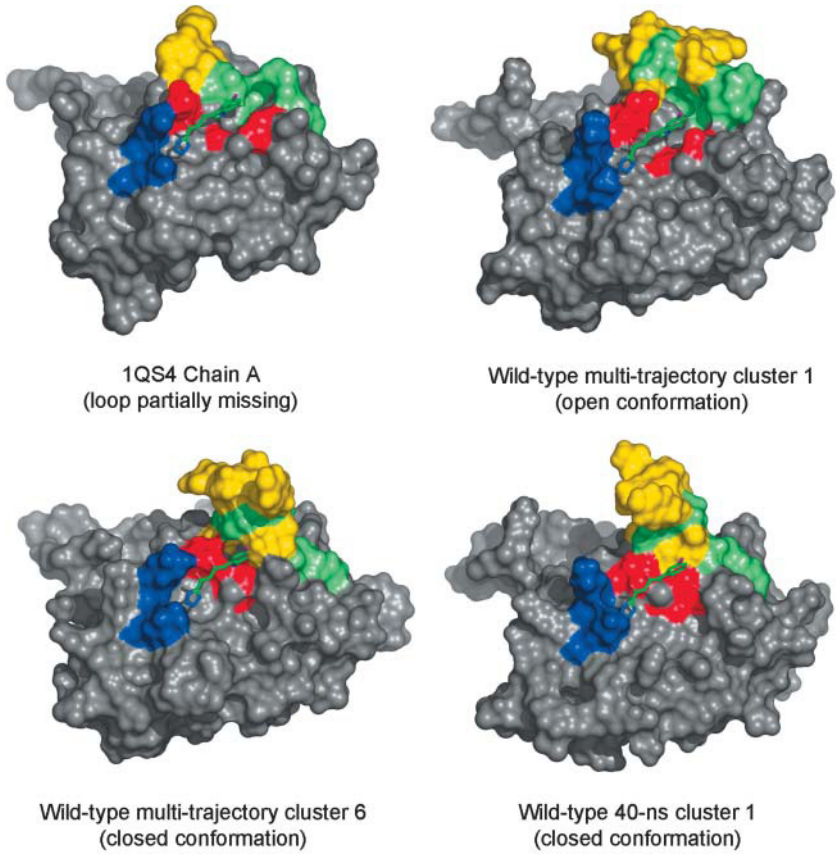

FIGURE 12 Molecular surface representation of three representative structures from the wild-type simulations and the crystal structure 1QS4, chain A, with the inhibitor 5CITEP superimposed in the active site using the orientation and position found in the crystal structure. The blue patches represent residues K^{156}, K^{159} known to interact with DNA; the red patches represent the three conserved catalytic residues; the yellow patches represent the catalytic loop; and the green patches represent the newly identified alternative binding trench for 5CITEP from docking studies by Schames et al. (2004). Some residues belong to both the loop region and the trench region (F^{139}, G^{140}, P^{142}, and Q^{148}).

between the E^{152} and D^{64} where the second metal ion is supposed to coordinate during catalysis. This suggests a possible mechanism for 5CITEP's inhibitory action. Note that the green patch actually makes up the underside of the loop and becomes buried in the closed conformation. In the open form, this ''trench'' is fully exposed and forms a very well-defined shape for binding of ligands, but in the closed form, the trench is completely buried. These conformations may offer a better platform for further structure-based inhibitor design efforts. We speculate that a bidentate ligand fitting into both binding sites may be a stronger inhibitor that can bind across the grooves formed by residues C^{65}, T^{66}, H^{67}, Q^{92}, Q^{148}, E^{152}, K^{156}, and K^{159}. An example of such a ligand is the L-chicoric acid that has two rigid wings joined by a common flexible center. This may provide an explanation of why the L-chicoric acid is so far the most potent IN inhibitor.

Inasmuch as the active form of IN in vivo requires at least a dimer, the conformational change involved may not be limited to the secondary or tertiary structural levels; it may likely extend to the quaternary structural level as indicated by the dynamic nature of the interface loop (data not shown). But because the full-length structure of the IN is still unavailable and the exact functional multimeric state of IN has not been unambiguously determined, our research focus is on understanding the structure and dynamics of the core domain.

CONCLUSIONS

The IN is an interesting enzyme both because it is clinically important and because it is a member of the transposase/IN family of enzymes that are believed to be the main driving forces behind evolution. In this study, we focused on understanding the structure-function-dynamics relationships of the catalytic core domain, paying special attention to the catalytic loop. Starting from the experimental 5CITEP-bound crystal structure, our simulations demonstrated large-scale conformational changes of the catalytic loop. Analysis of protein loop conformations is a notoriously difficult task because they do not have easily identifiable regular geometric patterns. Until recently, analysis and classification of loops were done manually, and languages used to describe these structures remained mostly qualitative (Espadaler et al., 2004; Oliva et al., 1997, 1998; Turcotte et al., 2001). In our work, we have attempted a systematic characterization of the loop dynamics in an effort to further our understanding of this important enzyme. We summarize the lessons learned as follows:

The wild-type catalytic loop has a slow mode of motion that closes and opens the space around the active site, and the transient formation of a 3_{10}-helical structure by the residues around P^{145} appears to be a major influence in the dynamics of the loop. Because the hinge mutations sterically hinder the loop from closing, the associated loss of activity strongly suggests that this closing-opening conformational change is functionally important. The seven conformations identified by clustering analysis provide a starting point for further work on docking and virtual screening studies. The dynamics of Y^{143} has been further clarified in our simulations and is closely linked to the loop conformation. Judging from the high flexibility rendered by the catalytic loop, one may speculate that the role of the loop is to provide the mobility to allow Y^{143} to access the substrate easily and to allow easy release of the products.

We gratefully acknowledge the usage of VMD, Pymol, and SwissPDB Viewer. Computer time was provided by the Pittsburgh Supercomputer Center.

This work was supported by research grants from the National Institutes of Health (GM64458 and GM067168 to Y.D., GM56553 to J.M.B.) and the Robert A. Welch Foundation (J.M.B.).

REFERENCES

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

Asante-Appiah, E., and A. M. Skalka. 1997. Molecular mechanisms in retrovirus DNA integration. *Antiviral Res*. 36:139–156.

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.

Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1987. Correct integration of retroviral DNA in vitro. *Cell*. 49:347–356.

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.

Bushman, F. 1995. Targeting retroviral integration. *Science*. 267:1443–1444.

Cannon, P. M., W. Wilson, E. Byles, S. M. Kingsman, and A. J. Kingsman. 1994. Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. *J. Virol*. 68:4768–4775.

Case, D. A., D. A. Pearlm, J. W. Caldwell, T. E. Cheatham 3rd, J. Wang, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, and others. 2002. AMBER. Version 7. University of San Francisco, San Francisco, CA.

Chen, I. J., N. Neamati, and A. D. MacKerell Jr. 2002a. Structure-based inhibitor design targeting HIV-1 integrase. *Curr. Drug Targets Infect. Disord*. 2:217–234.

Chen, I. J., N. Neamati, and A. D. MacKerell Jr. 2002b. Structure-based inhibitor design targeting HIV-1 integrase. *Curr. Drug Targets Infect. Disord*. 2:217–234.

Chen, J. C., J. Krucinski, L. J. Miercke, J. S. Finer-Moore, A. H. Tang, A. D. Leavitt, and R. M. Stroud. 2000. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. USA*. 97:8233–8238.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys*. 98:10089–10092.

Dayam, R., and N. Neamati. 2003. Small-molecule HIV-1 integrase inhibitors: the 2001—2002 update. *Curr. Pharm. Des*. 9:1789–1802.

de Groot, B. L., D. M. van Aalten, A. Amadei, and H. J. Berendsen. 1996. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys. J*. 71:1707–1713.

DeLano, W. L. 2002. The PyMOL Molecular Graphics System. Version 0.95. DeLano Scientific, San Carlos, CA.

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.

Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, and others. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012.

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.

Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–9175.

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., Y. Liu, H. Chen, M. Farzan, and F. Dyda. 1997. Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. *J. Virol.* 71:3507–3514.

Espadaler, J., N. Fernandez-Fuentes, A. Hermoso, E. Querol, F. X. Aviles, M. J. Sternberg, and B. Oliva. 2004. ArchDB: automated protein loop classification as a tool for structural genomics. Nucleic Acids Res. (database issue) 32:D185–D188.

Espeseth, A. S., P. Felock, A. Wolfe, M. Witmer, J. Grobler, N. Anthony, M. Egbertson, J. Y. Melamed, S. Young, T. Hamill, and others. 2000. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proc. Natl. Acad. Sci. USA.* 97:11244–112449.

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.

Gallay, P., S. Swingler, C. Aiken, and D. Trono. 1995. HIV-1 infection of nondividing cells: C-terminal tyrosine phosphorylation of the viral matrix protein is a key regulator. *Cell.* 80:379–388.

Gerton, J. L., S. Ohgi, M. Olsen, J. DeRisi, and P. O. Brown. 1998. Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *J. Virol.* 72:5046–5055.

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.

Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PDBViewer: an environment for comparative protein modeling. *Electrophoresis.* 18:2714–2723.

Johnson, M. S., M. A. McClure, D. F. Feng, J. Gray, and R. F. Doolittle. 1986. Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc. Natl. Acad. Sci. USA.* 83:7648–7652.

Jorgensen, W. L., J. Chandrasekhar, J. Madura, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

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.

Leavitt, A. D., L. Shiue, and H. E. Varmus. 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J. Biol. Chem.* 268:2113–2119.

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.

Lobel, L. I., J. E. Murphy, and S. P. Goff. 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. *J. Virol.* 63:2629–2637.

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

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles, and M. J. Sternberg. 1997. An automated classification of the structure of protein loops. *J. Mol. Biol.* 266:814–830.

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles, and M. J. Sternberg. 1998. Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction. *J. Mol. Biol.* 279:1193–1210.

Pani, A., A. G. Loi, M. Mura, T. Marceddu, P. La Colla, and M. E. Marongiu. 2002. Targeting HIV: old and new players. *Curr. Drug Targets Infect. Disord.* 2:17–32.

Pluymers, W., E. De Clercq, and Z. Debyser. 2001. HIV-1 integration as a target for antiretroviral therapy: a review. *Curr. Drug Targets Infect. Disord.* 1:133–149.

Roitberg, A., and R. Elber. 1991. Modeling side chains in peptides and proteins: application of locally enhanced sampling and simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95:9277–9287.

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

Sayasith, K., G. Sauve, and J. Yelle. 2000. Characterization of mutant HIV-1 integrase carrying amino acid changes in the catalytic domain. *Mol. Cells.* 10:525–532.

Sayasith, K., G. Sauve, and J. Yelle. 2001. Targeting HIV-1 integrase. *Expert Opin. Ther. Targets.* 5:443–464.

Schames, J. R., R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. Ni, and J. A. McCammon. 2004. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* 47:1879–1881.

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.

Turcotte, M., S. H. Muggleton, and M. J. Sternberg. 2001. Automated discovery of structural signatures of protein fold and function. *J. Mol. Biol.* 306:591–605.

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

van Gent, D. C., A. A. Groeneger, and R. H. Plasterk. 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc. Natl. Acad. Sci. USA.* 89:9598–9602.

van Gent, D. C., A. A. Oude Groeneger, and R. H. 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., E. Yeheskiely, G. A. van der Marel, J. H. van Boom, and R. H. Plasterk. 1991. Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic Acids Res.* 19:6691–6698.