The Mobility of an HIV-1 Integrase Active Site Loop Is Correlated with Catalytic Activity[†,‡]

Jason Greenwald,[§,||] Vincent Le,[§,||] Scott L. Butler,[⊥] Frederic D. Bushman,[⊥] and Senyon Choe*[,§]

Structural Biology and Infectious Diseases Laboratories, The Salk Institute, La Jolla, California 92037, and Department of Chemistry and Biochemistry, University of California, San Diego, California 92037

Received March 29, 1999; Revised Manuscript Received May 14, 1999

ABSTRACT: Replication of HIV-1 requires the covalent integration of the viral cDNA into the host chromosomal DNA directed by the virus-encoded integrase protein. Here we explore the importance of a protein surface loop near the integrase active site using protein engineering and X-ray crystallography. We have redetermined the structure of the integrase catalytic domain (residues 50−212) using an independent phase set at 1.7 Å resolution. The structure extends helix α4 on its N-terminal side (residues 149−154), thus defining the position of the three conserved active site residues. Evident in this and in previous structures is a conformationally flexible loop composed of residues 141−148. To probe the role of flexibility in this loop, we replaced Gly 140 and Gly 149, residues that appear to act as conformational hinges, with Ala residues. X-ray structures of the catalytic domain mutants G149A and G140A/G149A show further rigidity of α4 and the adjoining loop. Activity assays in vitro revealed that these mutants are impaired in catalysis. The DNA binding affinity, however, is minimally affected by these mutants as assayed by UV cross-linking. We propose that the conformational flexibility of this active site loop is important for a postbinding catalytic step.

Integration of the viral cDNA into a host chromosome is required for viral replication. Prior to integration, viral cDNA is cleaved on each strand near the 3′ end by integrase (terminal cleavage), probably to remove nontemplated extra bases occasionally added by reverse transcriptase (*1*, *2*). Integrase then catalyzes the attachment of the recessed 3′ ends to the target DNA (strand transfer) (*3*−*5*). In vitro, purified integrase protein can carry out the terminal cleavage (*6*, *7*) and strand transfer reactions (*8*−*11*). Purified integrase can also carry out an apparent reversal of strand transfer, termed "disintegration" (*12*).

The structure of full-length integrase has not yet been identified. However, the structure of each of its three domains has been determined in isolation. The amino-terminal domain (amino acids 1−50) is composed of three α-helices with an embedded zinc binding site (*13*, *14*). The central domain [amino acids 50−212 (50-212[1])] is composed of mixed α-helix and β-sheet (*15*), forming a compact fold seen previously in several polynucleotidyl phosphotransferases (*16*−*19*). This domain by itself can carry out covalent chemistry on permissive disintegration substrates, an indication of its role in catalysis (*20*, *21*). A conserved amino acid sequence motif, D,D-35-E, is found in the central domain of integrases and some related prokaryotic transposases. The carboxyl-terminal domain (amino acids 213−288) is composed of a five-stranded β-barrel resembling an SH3 domain, and contributes to DNA binding (*22*, *23*). All three domains form dimers independently, and the full-length integrase is likely to act as a higher-order multimer (reviewed in ref *24*).

We have used X-ray crystallography and protein engineering to investigate the active site of HIV-1 integrase. We determined the structure of 50−212 using an independent phase set at 1.7 Å resolution. After this work was completed, two reports of 50−212 structures in new crystal forms were published, both of which contained bound metal cofactor near the active site (*25*, *26*). In our structure, we observed that the inferred active site was more visible than in the structure of Dyda et al. (*15*); however, we found that a surface loop (residues 141−148) was still disordered. In many well-studied enzymes, flexible surface loops are important for substrate binding and catalysis (*27*−*29*), leading us to focus on the role of the region of residues 140−149. We substituted Gly residues at 140 and 149, potential hinges for the integrase flexible loop, with more constrained Ala residues and examined the mutant structures and enzymatic activities. Our data indicate that the degree of conformational flexibility seen in the structures of these mutants is correlated with catalytic activity.

EXPERIMENTAL PROCEDURES

Site-Directed Mutagenesis and Protein Expression. DNA encoding the integrase 50−212 (*20*) was mutagenized to

[†] This work was supported by NIH Grants GM56553 (S.C. and F.D.B.) and AI34786 (F.D.B.). F.D.B. is a scholar of the Leukemia Society of America. J.G. is supported by a Markey fellowship and a Howard Hughes Medical Institute predoctoral fellowship.

[‡] The coordinates have been deposited at the Protein Data Bank under file names 1b9d, 1b9f, and 1b92.

* Corresponding author.

[§] Structural Biology Laboratory, The Salk Institute.

[||] University of California.

[⊥] Infectious Diseases Laboratory, The Salk Institute.

[1] Abbreviations: HIV-1, human immunodeficiency virus-1; 50−212, residues 50−212 of HIV-1 integrase; βME, β-mercaptoethanol; DTT, dithiothreitol; HT, hexahistidine tag; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; PEG, poly(ethylene glycol); MAD, multiwavelength anomalous diffraction; ASV, avian sarcoma virus.

10.1021/bi9907173 CCC: $18.00
Published on Web 06/18/1999

Table 1: X-ray Data and Phasing and Refinement Statistics

space group		$P3_121$							
data set		50−212 (λ = 1.0162 Å)			G149A (λ = 1.54 Å)			G140A/G149A (λ = 1.08 Å)	
resolution (Å)		20−1.7			20−2.0			20−1.7	
completeness (%)/outer 5%		99.2/96.9			98.4/91.9			95.3/98.5	
R_{sym} (%)/outer 5%		4.2/31.9			3.9/25.9			5.4/21.2	
no. of protein/water atoms		1094/110			1094/100			1147/69	
R-factor/R_{free} (%)[a]		22.4/27.8			21.9/28.9			24.3/28.0	
rmsd bond distance		0.020			0.019			0.015	
rmsd angle distance[b]		0.033			0.037			0.032	
				MAD Phasing Statistics					
resolution (Å)	9.49	5.98	4.27	3.35	2.76	2.34	2.04	1.80	overall
no. of reflections	52	317	713	1319	2076	2966	4041	5281	16765
phasing power									
λ_1[c] anomalous	0.67	0.83	1.17	1.11	1.28	1.18	0.84	0.52	0.92
λ_1 isomorphous	−	−	−	−	−	−	−	−	−
λ_2 anomalous	0.82	0.96	1.33	1.34	1.46	1.35	0.98	0.57	1.03
λ_2 isomorphous	1.68	3.15	3.28	3.20	3.76	3.75	2.71	1.40	2.69
λ_3 anomalous	0.42	0.72	0.61	0.57	0.62	0.59	0.40	0.17	0.43
λ_3 isomorphous	1.81	3.10	3.60	3.50	4.03	4.24	3.26	2.11	3.26
λ_4 anomalous	0.23	0.25	0.20	0.17	0.22	0.17	0.10	0.07	0.12
λ_4 isomorphous	1.45	2.17	2.47	2.18	2.47	2.76	2.23	1.53	2.19
mean FOM	0.60	0.78	0.86	0.85	0.87	0.86	0.79	0.60	0.76

[a] R_{free} was calculated with 5% of the data. No cutoff was used for statistics or refinement. [b] REFMAC defines angles as 1−3 bond distances. [c] λ_1 = 1.0162 Å. λ_2 = 1.0458 Å. λ_3 = 1.0464 Å. λ_4 = 1.0596 Å.

contain the F185K change known to improve solubility (*30*) using the Excite Mutagenesis Kit (Stratagene). All catalytic domain fragments used in this study contained this mutation. The G149A and G140A/G149A mutations were constructed in a similar fashion, and expressed in *Escherichia coli* from a pET28 vector (Novagen) as described previously (*30*). Briefly, *E. coli* cells containing the pET28 vector with 50−212 were grown for 4 h at 37 °C and then lysed by sonication in nickel loading buffer [20 mM HEPES (pH 7.5), 500 mM NaCl, and 5 mM βME]. The lysate was cleared by centrifugation at 100000*g*, loaded onto a nickel column (Qiagen), washed with 10 mM imidazole, and then eluted with a gradient of 20 to 100 mM imidazole. The eluted protein was dialyzed into thrombin cleavage buffer [40 mM Tris (pH 8.0), 300 mM NaCl, 2 mM $CaCl_2$, 5 mM βME, and 5% glycerol], digested with thrombin, and then run over a benzamidine−Sepharose column and then a nickel column to remove thrombin and uncleaved protein. The cleaved protein was dialyzed into storage buffer [20 mM HEPES (pH 7.5), 500 mM NaCl, and 1 mM DTT]. The vector expressing the G140A change alone was constructed by mutagenizing a synthetic integrase gene containing F185K (GenBank accession number AF029884 kindly provided by P. Brown) by replacing wild-type DNA with DNA encoding the G140A change.

Disintegration and UV Cross-Linking. The HIV-1 integrase catalytic domain (residues 50−212) containing an amino-terminal hexahistidine tag (HT) was overexpressed and purified essentially as described previously (*15*, *20*). Disintegration assays were carried out as described in refs *12* and *20*. Briefly, disintegration substrates were end-labeled with ^{32}P and then incubated with 50−212 or the constrained-loop mutants at 37 °C (0.8 pmol of DNA per 10 μL of reaction mixture). Aliquots were taken at various time points, separated by denaturing gel electrophoresis, and quantified with a phosphorimager. Initial rates were determined at early times when less than 20% of the substrate had been consumed. UV cross-linking studies were carried out as previously described (*31*). End-labeled disintegration substrate was mixed with 50−212 or each of the constrained-loop mutants, each at 5 μM, in the presence of integration buffer containing 5 mM Mn^{2+} and 0.4 pmol of DNA in a 20 μL final volume, incubated on ice for 30 min, and then irradiated at 254 nm (3 mV) for 3 min. Samples were then boiled in SDS loading buffer and separated by SDS−PAGE. Cross-linked products were visualized as labeled complexes, which exhibited the molecular masses expected for that of the integrase domain and the labeled DNA. Results are shown for protein containing the HT. Assays without the HT yielded reduced amounts of product but in the same relative ratios (data not shown). Tests of the terminal cleavage and strand transfer activities of integrase under the conditions described in ref *8* showed no effect of the cacodylate buffer on activity in the millimolar range.

Crystallization and Data Collection. Crystals of 50−212 were obtained by vapor diffusion at 4 °C using 15% PEG8000, 100 mM cacodylic acid (pH 6.5), 200 mM ammonium sulfate, 5 mM DTT, 5 mM $MnCl_2$, and 5 mM $MgCl_2$, as a precipitant (modified from that described in ref *15*). Crystals of the mutants G149A and G140A/G149A were also obtained under the same condition. The amino-terminal hexahistidine tag was removed by treatment with thrombin for the crystallization studies. Each protein crystallizes isomorphously in space group $P3_121$ with the following unit cell dimensions: $a = b = 72.6$ Å and $c = 65.5$ Å with one monomer per asymmetric unit. To collect X-ray diffraction data, crystals were equilibrated in cryoprotectant containing 30% ethylene glycol prior to freezing in a nitrogen stream at 110 K. Multiwavelength anomalous diffraction (MAD) data for 50−212 were collected at the ALS 5.0.2 beamline on a Quantum 4 CCD detector (ADSC). G140A/G149A data were collected at the SSRL 7-1 beamline on a 30 cm Mar imaging plate (Mar Research) and G149A data on a MacScience imaging plate detector (DIP2020k) with a rotating anode equipped with platinum-coated focusing mirrors. All data were integrated and scaled with DENZO and SCALEPACK (*32*) except the MAD data (50−212) which were processed with MOSFLM (*33*) and SCALA (*33*).

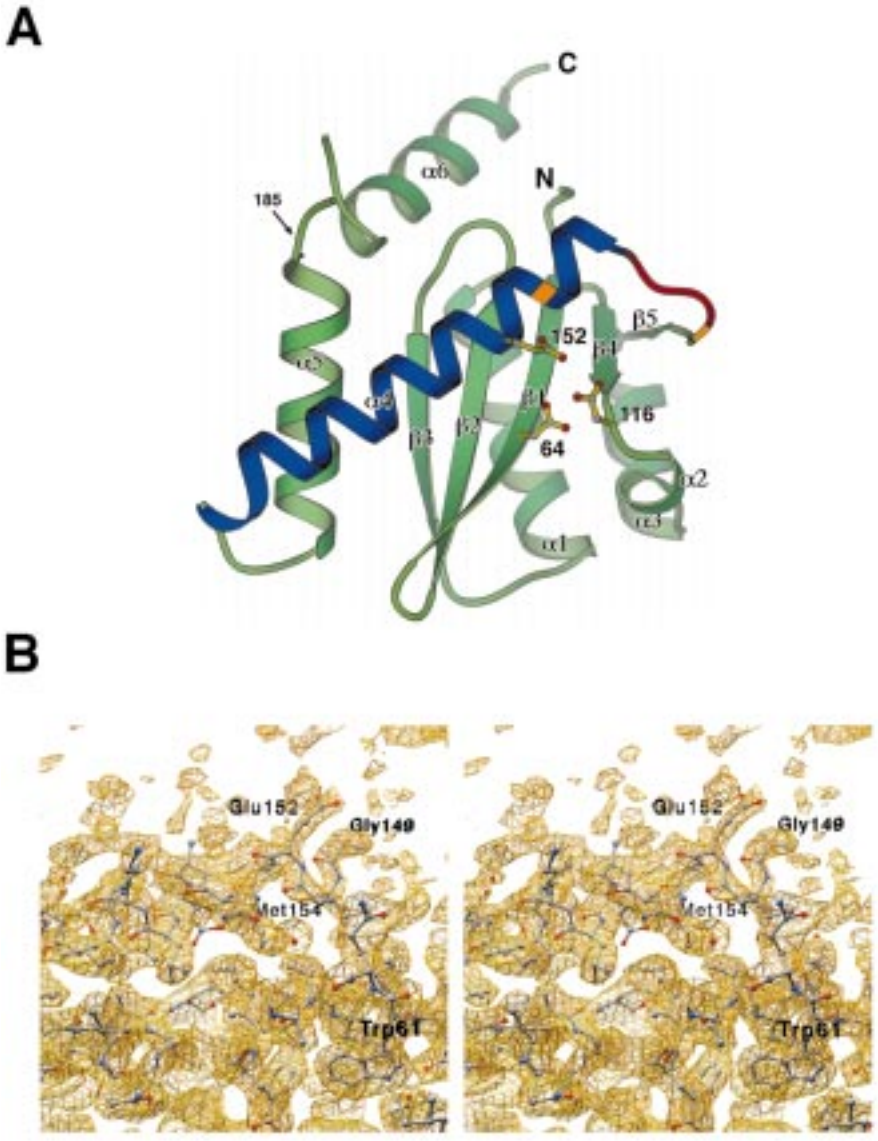

FIGURE 1: (A) Ribbon diagram of 50−212 (G140A/G149A) with α4 in front of the β-sheet core and colored blue. Three catalytic residues are shown with carboxylate oxygen atoms in red. The two positions that were mutated to alanine are in orange, and the poorly defined part of the catalytic loop is in red (residues 141−144). This figure was created with Setor (*48*). (B) Stereoview of the experimental electron density in the region of the disordered loop overlaid with the final model. The map was calculated using $F_{\lambda 1}$ and the MAD phases to 2.2 Å and is contoured at 0.75σ.

MAD Phasing and Model Building. From the earlier studies (*15*), it was known that two free sulfhydryl groups of Cys 65 and Cys 130 in the 50−212 structure were covalently modified by the cacodylate buffer in the crystallization solution to form Cys-S-AsO$(CH_3)_2$. We took advantage of the presence of two As atoms to derive an unbiased experimental phase set for an independent structure determination (Table 1). To obtain the MAD phases, four data sets from four wavelengths near the absorption edge of As ($\lambda = 1.046$ Å) were collected from a single frozen crystal using the inverse beam technique. Initial phases were calculated with MLPHARE (*33*) to 1.8 Å resolution. The resulting electron density maps were of high enough quality to build a complete model, including most of the side chains, with the program O (*34*). Refinement with REFMAC (*33*) and addition of waters using ARP (*33*) led to the current *R*-factor of 22.4% against all data in the resolution range of 20.0−1.7 Å. The temperature factors were refined using the default restraints of 1.5σ, 1.0σ, and 1.5σ for bonded atoms, 1−3 nonbonded atoms, and 1−4 nonbonded atoms, respectively. The refined model was then used to calculate phases for the $2F_o - F_c$, $F_o - F_c$, and $F_o - F_o$ maps for the two mutant data sets (G149A and G140A/G149A). In the G140A/G149A data, residues 145−148 became clearly defined as an extension of α4 by four residues toward its N-terminus. Water molecules were independently added in three different models. In all three models, the main chain conformations of all residues lie in favored and allowed regions in the Ramachandran plots.

RESULTS

Fold of the 50−212 Monomer. We carried out an independent structure determination of 50−212 at 1.7 Å resolution, derived from a MAD phase set. For this experiment, we used the previously determined positions of two

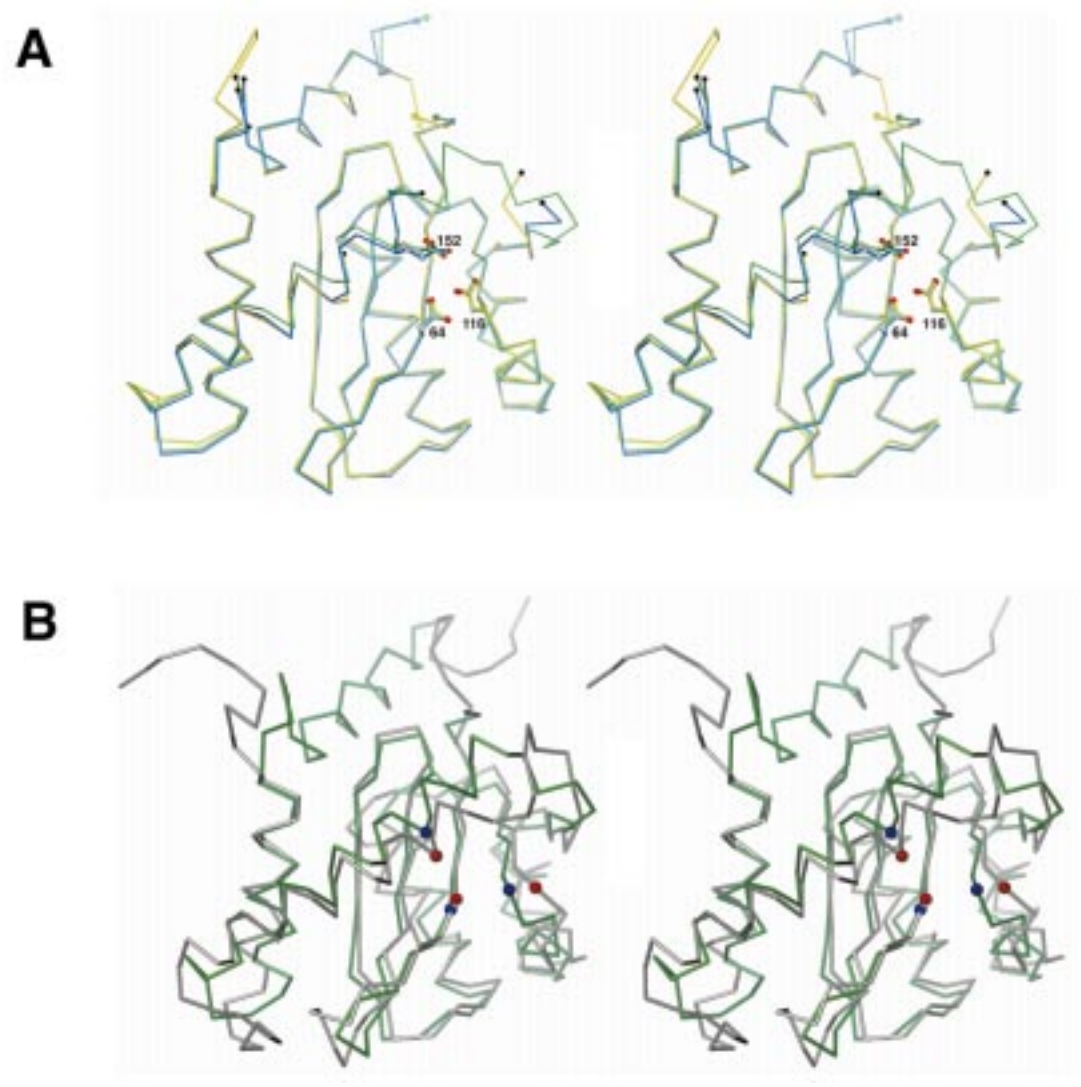

FIGURE 2: (A) Superposition of the structure described by Dyda et al. (*15*) in yellow (residues 56−140 and 154−208), the MAD-phased structure of the wild-type 50−212 in blue (residues 57−140, 149−189, and 193−210), and 50−212 (G140A/G149A) in green (residues 56−188 and 193−210). The view is in the same orientation as Figure 1A. Chain breaks are denoted by black dots and termini by colored balls. The three conserved acidic residues are displayed with their side chains to highlight the similarity of their conformations. (B) Superposition of 50−212 (G140A/G149A) in green with the ASV integrase catalytic domain in gray in the same orientation as Figure 1A. Cα positions of three catalytic residues are shown in blue for HIV and in red for ASV structures. This figure was created with Setor (*48*).

As atoms that become covalently bound to the integrase molecule in the presence of cacodylate buffer and a reducing agent (DTT) (*15*). As seen previously (*15*, *25*, *35*), the structure of the catalytic core consists of a central antiparallel β-sheet ($\beta1-\beta5$) flanked by α-helices on both sides (α1, α3, and α5 on one side and α4 on the other). Two subunits form a dimer interface through α1, α3, α5, and α6, whereas α4 (in blue in Figure 1A), on the opposite side of the β-sheet core, forms a part of the active site pocket.

The 1.7 Å MAD-phased structure differs from that reported in ref *15* in that it extends α4 on its N-terminal side to Gly 149 (Figure 1B), providing a complete description of the three conserved acidic residues (Asp 64, Asp 116, and Glu 152; see Figure 2A) inferred to form part of the active site. The region from residue 141 to 148 remains disordered. Furthermore, the C-terminal helix (α6) is now continuous to Thr 210. The other differences are in the conformations of some side chains, including several residues which clearly exist in two conformations and are modeled as such (Leu 63, Thr 66, Glu 138, and Lys 186). The model also includes a sulfate whose binding in a basic patch (Lys 71, Arg 166, and His 171) was confirmed by its anomalous signal at 1.54 Å.

Comparison of the currently available structures indicates that covalent modification by cacodylate causes structural changes which may obstruct metal binding. The structures of 50−212 in this study did not reveal any bound metal ion in the active site, even when the crystals were soaked at a concentration as high as 0.5 M $MnCl_2$. The three conserved acidic residues are thought to utilize metal in several polynucleotide phosphotransfer enzymes related to integrase, so it is presumed that metal atoms should bind. The binding of cacodylate causes a series of structural movements in the region around $\beta5$ which result in the side chains of Asp 64 and Asp 116 moving away from the metal-binding conformation found in refs *25*, *26*, *36*, and *37*. The binding of Mg^{2+} or Mn^{2+}, in contrast, does not alter the structure, as revealed by comparison of the structures of the HIV and ASV

integrase catalytic domains in the absence of cacodylate but in the presence and absence of metal (*25*, *36*, *38*). Nevertheless, 50−212 retains its disintegration activity in cacodylate buffers (unpublished data). Possibly, the presence of substrate stabilizes the metal binding and permits catalysis, at least disintegration.

A surface loop (residues 190−192) was disordered in the 1.7 Å MAD structure. This loop has been found to adopt several different conformations in previously reported structures (*15*, *25*, *35*), indicating its mobility. Possibly contributing to the mobility in this loop is the F185K substitution, which was included to improve solubility. Given that this mutant obstructs replication of HIV (*39*), it seems possible that the F185K change disrupts the native conformation in this region.

The Active Site Conformation of the Constrained-Loop Mutants. To explore the function of the active site loop, we replaced two glycines that flank the loop and potentially act as hinges with alanines. 50−212 mutants containing G140A, G149A, and G140/G149A were generated and the structures of the latter two determined (the G140A mutant did not yield diffraction-quality crystals). Both structures exhibit increased rigidity of the active site loop. In G140A/G149A, the major difference from nonmutant structures is in the extension of α4 to Pro 145. The backbone trace of G140A/G149A for residues 141−144 of the active site loop is poorly defined (red in Figure 1A) with temperature factors reaching 82.1. In G149A, the increased rigidity of α4 is less profound, and is evident in the lower relative temperature factors of the helix as well as in the presence of weak density near the borders of the disordered region.

Atomic coordinates of the conserved catalytic residues in three related 50−212 structures [Asp 64 and Asp 116 in Dyda et al. (*15*) and Asp 64, Asp 116, and Glu 152 in two structures reported here] exhibit few differences regardless of the degree of order in α4 (Figure 2A). On the other hand, comparison between G140A/G149A and the ASV integrase catalytic core domain structures reveals that α4 and the loop preceding it have significant positional differences. Figure 2B shows that α4 is shifted in its position in relation to the rest of the molecule between HIV and ASV enzymes. However, the catalytic residues from the two enzymes retain a common spatial orientation and distance relationship between their Cα positions (Glu 152, Asp 64, and Asp 116 from HIV and Glu 157, Asp 64, and Asp 121 from ASV; Figure 2B).

Activities of 50−212 and the Constrained-Loop Mutants. The catalytic activities of constrained-loop mutants were tested in vitro. 50−212 alone cannot carry out normal terminal cleavage or strand transfer, but does catalyze disintegration (*20*), a reaction resembling the reversal of integration that is apparently more permissive (*12*) (diagrammed beside Figure 3A). After 20 min, 25 pmol of 50−212 converted 18% of the substrate (0.8 pmol of DNA per reaction) to product, for a conversion rate of 7.2×10^{-15} mol/min (Figure 3A, lanes 2−5). The rates were slower with the single mutants, 4×10^{-16} and 8×10^{-16} mol/min for G140A (lanes 6−9) and G149A (lanes 10−13), respectively. Product formation by G140A/G149A was almost undetectable with a rate of 8×10^{-17} mol/min (lanes 14−17). Thus, the single mutants impair disintegration activity 9−18-fold,

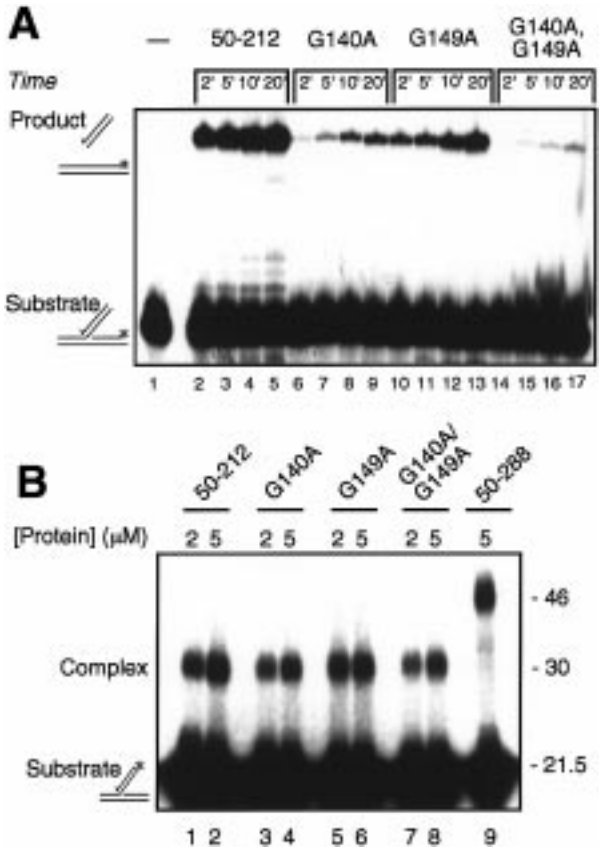

FIGURE 3: Disintegration and DNA binding activities of 50−212, G140A, G149A, and G140A/G149A. (A) Time course analysis of disintegration. Reaction products were denatured, separated by electrophoresis on a DNA-sequencing type gel, and visualized by autoradiography. Disintegration substrates were labeled as marked by the asterisk. Lane 1 is a control without protein. (B) Assay of DNA binding by UV cross-linking. DNA substrates were labeled on the indicated DNA strand (asterisk). Proteins and concentrations tested are indicated above each lane. Numbers to the right of the gel indicate the mobilities of protein molecular mass standards.

while the double mutant nearly eliminates activity (about a 90-fold decrease).

DNA Binding by 50−212 and the Constrained-Loop Mutants. The reduced activity seen with the constrained-loop mutants could be a result of either impaired DNA binding or impaired catalysis. To address this issue, DNA binding was assayed by UV cross-linking. Assays with wild-type 50−212 and each of the constrained-loop mutants yielded cross-linked products in similar abundance (Figure 3B, lanes 1−8). No products were seen when no protein was added (data not shown). As a control, a longer form of integrase 50−288 (F185K/C260S) was also tested by cross-linking. This protein yielded a cross-linked complex, but as expected with slower mobility than that seen with the 50−212 derivatives (Figure 3B, lane 9).

Quantitation revealed that DNA binding is not greatly perturbed in the constrained-loop mutants. The amount of complex formed with each single mutant was within 2-fold of the wild-type amount, while that of the double mutant was within 4-fold. The defects in activity, in contrast, were 9−18-fold for the single mutants and >90-fold for the double mutant. Since the changes in the loop do not strongly influence DNA binding, these data support a model in which the correct loop sequence is important in a reaction step after the DNA binding step.

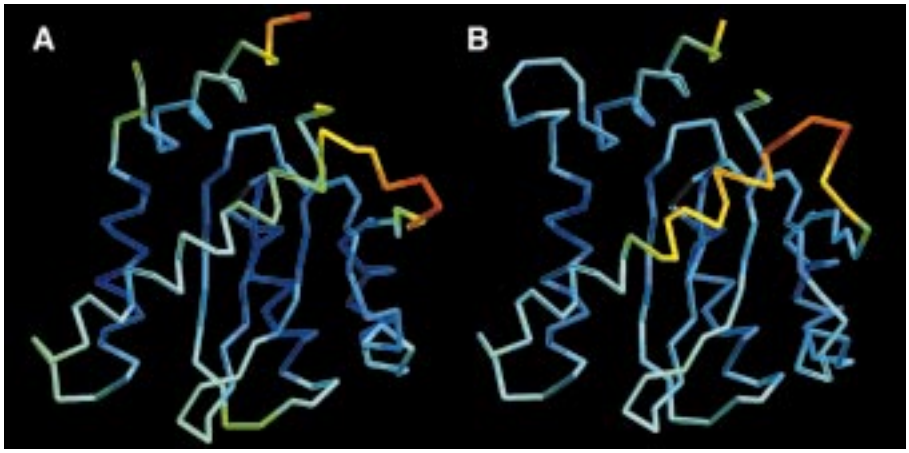

FIGURE 4: Two integrase structures color-coded by their relative temperature factors (*43*). The relative temperature factor is calculated as the number of standard deviations from the mean within one structure. The lowest is blue, and the highest is red with a continuous spectrum in between. Only the Cα atoms were used in the calculation. The low, average, and high Cα temperature factors for each structure are listed below. The program O (*34*) was used to make this figure. (A) For G140A/G149A, the low is 19.1, the average 34.0, and the high 82.1. (B) For 50−212(W130E) molecule 2 of the P1 crystal form (*25*), the low is 10.9, the average 29.7, and the high 90.0. The residue with the highest temperature factor in each model is Pro 142 (G140A/G149A) and Asn 144 (W130E).

DISCUSSION

The structures of HIV-1 integrase 50−212 presented here provide a detailed picture of the integrase active site. The conformations of the three acidic residues inferred to be involved in catalysis are all visualized in defined conformations. The Cα positions of these residues conform to that predicted by the two-metal mechanism for catalysis proposed by Beese and Steitz to describe the function of enzymes of this class (*40*). After this work was completed, two other groups also reported structures determined in the absence of bound cacodylate that confirm this point (*25*, *26*).

In our 1.7 Å MAD structure and in other structures of integrase catalytic domains, a flexible loop (amino acids 140−149) is located near the conserved acidic residues. Cross-linking studies of integrase with DNA (*41*, *42*) revealed that residues in this region (139−152) could be cross-linked to DNA near the reactive phosphate, placing the loop near the substrate during the reaction. Figure 4 illustrates the difference in flexibility of the loop between the G140A/G149A mutant and a 50−212 structure (*25*). The latter structure was chosen for comparison since it contains a fully traced active site loop. The structures in the crystal forms studied here with wild-type loop sequences do not have fully defined loop conformations, emphasizing the flexibility of the wild-type loops. The structures are color-coded by relative temperature factor (*43*), an indication of uncertainty of position and hence probable mobility. The relative temperature factor was greatly reduced by the glycine to alanine substitutions, confirming that the position of the loop was constrained. The catalytic activity of the G140A/G149A mutant was greatly reduced compared with that of the wild type, and each single mutant alone diminished activity. DNA binding was not significantly altered by comparison, indicating that the mutants are affecting primarily a postbinding step. Two models could account for the impairment of catalytic activity observed in the constrained-loop mutants. The mutants could reduce activity by altering the energetics of interconversion between different conformations required at different steps in the catalytic cycle. Alternatively, the structure adopted by the constrained loop mutants may represent a nonfunctional conformation which is more stable than the functional ones.

Previous studies of the activities of integrase mutants bolster the idea that the loop is important for catalysis and help distinguish between models for the mutant defect. Normally, integrase carries out hydrolysis of the viral cDNA 3′ end prior to integration. However, under some in vitro conditions, integrase can recruit nucleophiles other than water to participate in this reaction, such as glycerol (*44*, *45*). Mutants have been identified that alter the ratio of usage of different nucleophiles, implicating these residues in catalysis. Such mutants have been identified in residues 143 and 146−148 (*46*), implying that the flexible loop studied here interacts with the incoming nucleophile. Given that four residues in this loop show this effect, it seems simplest to propose that some of the reduction in catalytic efficiency of the mutants is a result of altering the mobility of the loop rather than mutating residues directly involved in catalysis. Also emphasizing the importance of Gly 140 and Gly 149 is the fact that substituting these residues abrogates distintegration, while changes in the other loop residues that were tested do not affect disintegration under the reported conditions (*46*). Thus, we presently favor the view that the loop must move during the catalytic cycle, though further evidence is needed on this point.

Recently, Robinson and co-workers isolated a derivative of HIV showing reduced sensitivity to the integrase inhibitor L-chicoric acid and found that the substitution of G140 with serine contributed to the phenotype (*47*). On the basis of the data presented here, that change would be expected to diminish integrase activity. It is conceivable that the inhibitor reduces the activity of the wild type more than that of G140S, making the mutation favorable under the selective pressure of L-chicoric acid.

The structures reported here may find several uses. The full visualization of the active site provides information required for the structure-based design of inhibitors. The availability of mutations that hold the loop in a defined conformation provides a means of probing the loop function genetically by the isolation of second site revertants in virus.

Moreover, the structure of the constrained-loop mutants might be used as an alternate template for structure-based design, in which inhibitors are developed that stabilize the loop in a conformation that is catalytically impaired.

ACKNOWLEDGMENT

We thank members of the Salk Institute Structural Biology Laboratory and Infectious Disease Laboratory for suggestions, materials, and comments on the manuscript. We thank J. Briggs, J. A. McCammon, J. Siegel, and members of their laboratories for helpful discussions. Data were collected at SSRL and ALS which are operated by the Department of Energy, Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the Department of Energy, Office of Biological and Environmental Research.

REFERENCES

1. Patel, P. H., and Preston, B. D. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 549−553.
2. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997) *J. Virol. 71*, 5382−5390.
3. Fujiwara, T., and Mizuuchi, K. (1988) *Cell 54*, 497−504.
4. Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987) *Cell 49*, 347−356.
5. Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989) *Cell 58*, 47−54.
6. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol. 63*, 5319−5327.
7. Sherman, P. A., and Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5119−5123.
8. Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1339−1343.
9. Bushman, F. D., Fujiwara, T., and Craigie, R. (1990) *Science 249*, 1555−1558.
10. Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell 62*, 829−837.
11. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) *Cell 63*, 87−95.
12. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) *Science 255*, 723−726.
13. Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T., Plasterk, R. H. A., Kaptein, R., and Boelens, R. (1997) *Curr. Biol. 7*, 739−746.
14. Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol. 4*, 567−577.
15. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science 266*, 1981−1986.
16. Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990) *Science 249*, 1398−1405.
17. Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994) *Cell 78*, 1063−1072.
18. Rice, P., and Mizuuchi, K. (1995) *Cell 82*, 209−220.
19. Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science 252*, 88−95.
20. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3428−3432.
21. Vink, C., Oude Groeneger, A. M., and Plasterk, R. H. A. (1993) *Nucleic Acids Res. 21*, 1419−1425.
22. Eijkelenboom, A. P. A. M., Puras Lutzke, R. A., Boelens, R., Plasterk, R. H. A., Kaptein, R., and Hard, K. (1995) *Nat. Struct. Biol. 2*, 807−810.
23. Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995) *Biochemistry 34*, 9826−9833.
24. Hansen, M. S. T., Carteau, S., Hoffmann, C., Li, L., and Bushman, F. (1998) in *Genetic Engineering. Principles and Methods* (Setlow, J. K., Ed.) pp 41−62, Plenum Press, New York and London.
25. Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 9150−9154.
26. Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., and Mikol, V. (1998) *J. Mol. Biol. 282*, 359−368.
27. Weston, S. A., Lahm, A., and Suck, D. (1992) *J. Mol. Biol. 226*, 1237−1256.
28. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1995) *Science 269*, 656−663.
29. Rice, P. A., and Steitz, T. A. (1994) *Structure 2*, 371−384.
30. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R., and Craigie, R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6057−6061.
31. Engelman, A., Hickman, A. B., and Craigie, R. (1994) *J. Virol. 68*, 5911−5917.
32. Otwinowski, Z. (1993) in *Proceedings of the CCP4 study weekend* (Sawyer, L., Isaacs, N., and Burley, S., Eds.) pp 56−62, Daresbury Laboratory, Warrington, U.K.
33. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. D 50*, 760−776.
34. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A 47*, 100−119.
35. Bujacz, G., Alexandratos, J., Qing, Z. L., Clement-Mella, C., and Wlodawer, A. (1996) *FEBS Lett. 398*, 175−178.
36. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1996) *Structure 4*, 89−96.
37. Katayanagi, K., Okumura, M., and Morikawa, K. (1993) *Proteins 17*, 337−346.
38. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol. 253*, 333−346.
39. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem. 271*, 7712−7718.
40. Beese, L. S., and Steitz, T. A. (1991) *EMBO J. 10*, 25−33.
41. Heuer, T. S., and Brown, P. O. (1997) *Biochemistry 36*, 10655−10665.
42. Esposito, D., and Craigie, R. (1998) *EMBO J. 17*, 5832−5843.
43. Parthasarathy, S., and Murthy, M. R. (1997) *Protein Sci. 6*, 2561−2567.
44. Engelman, A., Mizuuchi, K., and Craigie, R. (1991) *Cell 67*, 1211−1221.
45. Vink, C., Yeheskiely, E., van der Marel, G. A., van Boom, J. H., and Plasterk, R. H. A. (1991) *Nucleic Acids Res. 19*, 6691−6698.
46. van den Ent, F. M., Vos, A., and Plasterk, R. H. (1998) *J. Virol. 72*, 3916−3924.
47. King, P. J., and Robinson, W. E., Jr. (1998) *J. Virol. 72*, 8420−8424.
48. Evans, S. V. (1993) *J. Mol. Graphics 11*, 134−138.

BI9907173