

Solution Conformation and Dynamics of the HIV-1 Integrase Core Domain^{*[S]}

Received for publication, February 11, 2010, and in revised form, March 30, 2010 Published, JBC Papers in Press, April 1, 2010, DOI 10.1074/jbc.M110.113407

Nicholas C. Fitzkee[‡], James E. Masse[‡], Yang Shen[‡], David R. Davies[§], and Ad Bax^{‡1}

From the [‡]Laboratory of Chemical Physics and the [§]Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

The human immunodeficiency virus type 1 (HIV-1) integrase (IN) is a critical enzyme involved in infection. It catalyzes two reactions to integrate the viral cDNA into the host genome, 3' processing and strand transfer, but the dynamic behavior of the active site during catalysis of these two processes remains poorly characterized. NMR spectroscopy can reveal important structural details about enzyme mechanisms, but to date the IN catalytic core domain has proven resistant to such an analysis. Here, we present the first NMR studies of a soluble variant of the catalytic core domain. The NMR chemical shifts are found to corroborate structures observed in crystals, and confirm prior studies suggesting that the $\alpha 4$ helix extends toward the active site. We also observe a dramatic improvement in NMR spectra with increasing $MgCl_2$ concentration. This improvement suggests a structural transition not only near the active site residues but also throughout the entire molecule as IN binds Mg^{2+} . In particular, the stability of the core domain is linked to the conformation of its C-terminal helix, which has implications for relative domain orientation in the full-length enzyme. ¹⁵N relaxation experiments further show that, although conformationally flexible, the catalytic loop of IN is not fully disordered in the absence of DNA. Indeed, automated chemical shift-based modeling of the active site loop reveals several stable clusters that show striking similarity to a recent crystal structure of prototype foamy virus IN bound to DNA.

An essential step in the life cycle of retroviruses is the integration of their reverse-transcribed genome into that of the infected host. In human immunodeficiency virus type 1 (HIV-1), the viral integrase (IN)² is both necessary and sufficient to complete the integration reaction *in vitro* (1–3), although other proteins play a role *in vivo* (4). The integration mechanism occurs via a two-step process. First, during 3' processing, two nucleotide bases are removed from the 3' end of the U3 and U5 ends of the linear viral cDNA (1, 5–8). A stable nucleoprotein complex is then formed until the DNA is integrated into the host genome (9, 10). Then, during the second step of integra-

tion, strand transfer, the target DNA is cleaved and viral cDNA is integrated by joining the 5' end of the target sequence to the recessed 3' end of the viral DNA (1, 7, 8, 11, 12). The final removal of the unpaired 5' viral DNA ends and the ligation of the nicked DNA are thought to be carried out by cellular enzymes (13–15). Additionally, IN is observed to catalyze the reverse reaction, whereby the spliced cDNA is removed from the host genome (16). The biological relevance of this disintegration reaction is presently unclear, but it provides a useful means to characterize functionality of IN variants. Because of its functional requirement for viral replication, IN is an attractive target for inhibition, and drug development targeting of IN has been commercially successful (17). Further work inhibiting the mechanism of IN is ongoing and shows promise (18, 19).

HIV-1 integrase is organized into three domains, an N-terminal zinc-binding domain (residues 1–50), a catalytic core domain (residues 50–212), and a C-terminal Src homology 3 domain (3, 20–22). Full-length IN exists in a dimer-tetramer equilibrium (23), but it is thought that at least a tetramer is required for biological activity (24). The function of each domain during integration remains unclear, but the C-terminal domain has nonspecific DNA affinity, and constructs without this domain fail to form tetramers (23). Importantly, the catalytic core domain alone can catalyze the disintegration reaction, and mutations to catalytic residues Asp⁶⁴, Asp¹¹⁶, or Glu¹⁵² (the Asp-Asp-35-Glu motif) abolish all activity of the enzyme (3, 25). Thus, whereas all three domains work in concert to perform integration, the core domain alone is directly responsible for chemistry preprocessing and strand transfer.

The core domain has been studied extensively, and crystallography has yielded insight into its function, both when isolated (24, 26–31) and when paired with either the N-terminal (32) or C-terminal (33) domains. The structures reveal a dimeric core domain with an RNase H-type fold. The dimer is thought to be biologically relevant (24), but several regions of the structure, including a loop containing catalytic residue Glu¹⁵² (residues 140–153), are often disordered or involved in crystal packing interactions. Furthermore, mutagenesis studies, where inherently flexible Gly residues in this catalytic loop are replaced with Ala, suggest that the dynamics of the active site are important for catalysis (29). This is not surprising, as the core domain presumably must catalyze both the 3' processing as well as the strand transfer reactions. The active site is known to complex divalent cations such as Mn^{2+} and Mg^{2+} (27), and binding of these cations stabilizes IN as evidenced by protease protection assays (34–36).

* This work was supported, in whole or in part, by the National Institutes of Health Intramural Research Program of the NIDDK and the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH.

[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. S1–S8 and Tables S1 and S2.

¹ To whom correspondence should be addressed: 5 Memorial Dr., Bethesda, MD 20892. E-mail: bax@nih.gov.

² The abbreviations used are: IN, integrase; PRE, paramagnetic relaxation enhancement; PDB, Protein Data Bank; PFV, prototype foamy virus; DMSO, dimethyl sulfoxide.

Recently, the structure of an IN homolog from prototype foamy virus (PFV) was crystallized in complex with DNA (37). The catalytic loop in this structure extends helix $\alpha 4$ in the N-terminal direction, and residues 146–148 immediately precede $\alpha 4$ as a short stretch of 3_{10} helix. It is this 3_{10} helix that fits into the pocket formed by the recessed 3' viral DNA end. This study represents a significant step forward to understanding integrase biochemistry; nevertheless, it is not currently known whether the $\alpha 4$ helix kinks in the absence of DNA, nor is the interplay between dynamics and catalysis well understood. Computational studies on the core domain particle have yielded conflicting results: short time scale simulations suggest that the active site loop samples α -helical conformations (38), whereas longer time scale simulations suggest that this loop acts as a gate over the other catalytic residues in IN (39). Experimental validation of these simulations to date has been limited.

Solution NMR provides a powerful means to characterize biological macromolecules, elucidating both structure and dynamics, provided that adequate spectra can be obtained. It has the additional advantage of bypassing the requirement for crystallization, although it has its own difficulties for sample preparation. Solution NMR has been particularly useful for the study of HIV proteins, including studies of components of the Gag polyprotein (40–45). Although the structures of the N- and C-terminal domains of IN have also been determined by solution NMR (46–51), poor solubility along with the need for high salt concentrations have hindered NMR characterization of the catalytic domain (52). Here, we present the solution characterization of a soluble dimeric variant of the IN core domain (IN^{50–212}), Q53E,C56S,W131E,F185K,Q209E, which yields good NMR spectra in the presence of 40 mM MgCl₂. Having assigned backbone resonances of 97% of the residues, we show that the chemical shifts are consistent with crystal structures of the core domain. We report the dynamics of IN^{50–212} and find that the catalytic loop is only moderately dynamic on the picosecond to nanosecond time scale. Additionally, we investigate a structural transition induced by Mg²⁺ binding that dramatically improves the NMR spectral quality. This conformational shift, while requiring relatively high concentrations of MgCl₂, may reveal important clues about the structural stability of IN, including effects on the N- and C-terminal domain orientations. Finally, we study the interaction of the IN core domain with raltegravir, a commercially available drug for IN inhibition (17).

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The F185K,W131E IN^{50–212} construct was used in earlier crystallographic studies (30). Initial screening revealed that this construct had limited solubility at NaCl concentrations below 250 mM and pH \leq 7.0, so mutagenesis was performed to lower the pI of the protein. Mutagenesis sites were located at the protein termini to avoid undesired effects on catalysis. The final variant, Q53E,C56S,W131E,F185K,Q209E, was created using the QuikChange II mutagenesis kit (Agilent) and confirmed using DNA sequencing. Further testing revealed this variant to be more soluble and better suited for NMR conditions (up to 700 μ M at pH 7.0, 150 mM NaCl).

Deuterated, ¹⁵N- and ¹³C-labeled IN^{50–212} was expressed in BL21 Star(DE3) *Escherichia coli* cells (Invitrogen). A modified M9 growth medium was used (53), supplemented with 1 g/liter of ¹³C/²H/¹⁵N-IsoGro (Sigma). [¹H/¹³C]Glucose was used in the medium to reduce ¹H T₁ values by retaining partial protonation of methyl groups. After an overnight growth at 37 °C in 90% D₂O, 1 ml of culture was transferred to 1 liter of medium in 99% D₂O. The resulting protein is nearly fully deuterated at all H ^{α} positions, and characterization by mass spectrometry indicated that the overall deuteration fraction of the protein was 85–90%. The culture was grown at 37 °C to an A₆₀₀ of 0.6 and induced with 1 mM isopropyl thiogalactoside for 6 h. Cells were then harvested by centrifugation and stored overnight at –80 °C.

Residue specifically labeled IN^{50–212} was expressed using methods as outlined previously (54, 55). Briefly, unlabeled amino acids were added to unlabeled M9 growth medium at a concentration of 60 mg/liter. This mixture contained all amino acids except those that would ultimately be labeled, e.g. Val would be excluded if ¹⁵N-Val labeling was desired. BL21 Star(DE3) cells were grown in this medium and induced with 1 mM isopropyl thiogalactoside when they reached A₆₀₀ of 0.6. Fifteen minutes before induction, labeled amino acids were added to the medium (60 mg/liter), as well as the transaminase inhibitors disodium succinate, oxaloacetate, and sodium maleate (250 mg/liter each) (56). Cells were grown for 1.5 h and harvested by centrifugation. Residue specifically labeled constructs were made for ¹⁵N-Ile, ¹⁵N-Leu, ¹⁵N-Lys, ¹⁵N-Met, ¹⁵N-Phe/Tyr, and ¹⁵N-Val. Scrambling with other amino acids was very low, typically less than 5%.

Protein purification was performed as described by Goldgur *et al.* (27). The final dialysis buffer for IN^{50–212} was 150 mM NaCl, 20 mM HEPES, pH 6.8, 40 mM MgCl₂, 6% D₂O, and 0.02% NaN₃. Additionally, EDTA-free Complete Protease Inhibitor (Roche Applied Science) was added to reduce proteolytic degradation during storage. The dialyzed protein was then concentrated by ultrafiltration to 500–700 μ M and used for NMR measurements. The core domain was confirmed to be dimeric in both the presence and absence of MgCl₂ using a Wyatt DAWN EOS multiangle laser light scattering detector, yielding an estimated molecular mass of 35.5 kDa.

NMR Measurements—NMR measurements were carried out at 298 K on Bruker Avance 600, DRX800, and Avance 900 spectrometers equipped with cryogenic triple-resonance probeheads and *z* axis pulsed field gradients. Backbone assignment was performed using TROSY-based versions of the HNCO, HNCA, HNCB, HN(CO)CA, HN(CA)CO, and HN(COCA)CB triple resonance experiments (57) with modifications to the TROSY block for gradient selection and improved water suppression (58). Triple resonance pulse programs were also modified to allow for long nitrogen evolution times using the mixed constant time strategy (59). Triple resonance experiments used acquisition times of 45 ms (90 complex points) in the ¹⁵N dimension. For experiments measuring C ^{α} and C ^{β} resonances, 30 complex points were measured in the ¹³C dimension, with total acquisition times of 12 (C ^{α}) and 2 ms (C ^{β}). Acquisition times in the ¹³C' dimension were 36 ms (60 complex points). ¹H acquisition times varied from 60 to 70 ms using 512 complex

Conformation and Dynamics of HIV Integrase Core Domain

points. A three-dimensional ^{15}N - ^1H MT-PARE HMQC-NOESY spectrum was acquired on IN $^{50-212}$ at the 800 MHz field with a 150-ms NOE mixing time (59). This experiment used EBURP and reverse EBURP-selective $^1\text{H}^{\text{N}}$ pulses (60) for the first two 90° pulses as well as for the final readout block, to improve water suppression and increase the repetition rate of the experiment (61). Acquisition time in the $^{15}\text{N}\{F_2\}$ dimension was 43 ms using 110 complex points. For the ^1H dimension $\{F_1\}$, 128 complex points were measured with an acquisition time of 40 ms.

TROSY HNCO-based experiments were employed to measure R_1 and $R_{1\rho}$ relaxation rates at 600 and 900 MHz (62). An additional hyperbolic secant (600 MHz) or WURST (900 MHz) pulse was applied to ^{13}C nuclei at the midpoint of the R_1 and $R_{1\rho}$ relaxation periods to eliminate nitrogen-carbon cross-correlated relaxation effects. At 600 MHz (900 MHz), the $^{13}\text{C}'$ acquisition time was 22 ms (9 ms) over 20 (12) complex points, and ^{15}N evolution was measured for 70 ms (55 ms) using 140 (165) complex points. The spectra were folded extensively in the $^{13}\text{C}'$ dimension to reduce overall measurement time. At 600 MHz, R_1 and $R_{1\rho}$ rates were measured using two points assuming an exponential decay, measured at 40 and 880 ms for R_1 and 2 and 35 ms for $R_{1\rho}$. At 900 MHz these times are identical for R_1 , but 2 and 33 ms for $R_{1\rho}$. To improve sensitivity, the latter time point was measured for twice as many scans as the earlier time point (63). Steady state $^{15}\text{N}\{-^1\text{H}\}$ NOE values in IN $^{50-212}$ were measured at 600 MHz. Because of the long recycle delays required to avoid proton saturation during the no-NOE reference experiment, a two-dimensional TROSY readout was used instead of the three-dimensional TROSY-HNCO. In this experiment, ^{15}N was acquired for 150 ms (300 complex data points), and ^1H acquisition was 60 ms (512 complex points). $R_{1\rho}$ was measured using a 2-kHz spin lock field, and R_2 was calculated for this field using R_1 and $R_{1\rho}$ (64).

Data Processing and Chemical Shift Assignment—All NMR data were processed using the NMRPipe/NMRDraw software package (65). Peak picking and initial assignment was performed with the aid of the AutoLink package (66) within CARA (67). Site-specifically labeled variants were used to confirm and complete the assignments manually. For relaxation experiments, peaks were picked using NMRDraw and fit to exponential decays using in-house scripts.

Structure Ensemble Generation—Chemical shift-derived structures were created using the CS-ROSETTA protocol (68). In the first set of calculations, all backbone and $^{13}\text{C}^\beta$ chemical shifts were used to construct a backbone model for the core domain monomer using the standard CS-ROSETTA algorithm. Note that use of CS-ROSETTA to generate a monomer structure for this homodimeric protein requires prior knowledge from the x-ray structures that the protein is not a domain-swapped dimer. The latter would require an alternate computational approach, for which IN falls beyond the current size limit of CS-ROSETTA (69). In the monomer structures generated by CS-ROSETTA, the conformation of the active site is not sampled optimally because the ROSETTA energy is calculated over the entire molecule. Hence, models that make favorable interactions at the active site, but have higher energies elsewhere would not be selected in the final scoring process as the

total energy would remain high. To better sample low energy conformations of the catalytic loop that are in agreement with NMR chemical shifts, a second set of CS-ROSETTA calculations was performed. For this round, fragment selection (70) was once again carried out by using the backbone and $^{13}\text{C}^\beta$ chemical shifts; however, during refinement, all residues were fixed to the x-ray coordinates (PDB entry 1QS4, chain C) except for the catalytic loop (residues 139–153) and N-terminal residues 50–57. Specifically, the loop-relax protocol in ROSETTA 3.0 was used to generate 3,000 all-atom models, where conformational sampling was confined to the catalytic loop. These 3,000 models were clustered into 20 groups according to the conformations sampled by the active site loop. For each model in each cluster, the ROSETTA full-atom energy was calculated for the residues in the active site loop only. To this energy a chemical shift component was added to favor agreement between experimental and SPARTA-predicted chemical shifts, following the standard protocol (68). Then, for each of the lowest energy clusters, a representative low energy model is selected. Six such models, representing the six lowest energy clusters, are finally selected as representative of the ensemble. Repeating the analysis on PDB entry 1BL3 produced nearly identical clusters.

Relaxation Analysis— ^{15}N R_1 , R_2 , and $^{15}\text{N}\{-^1\text{H}\}$ NOE values at 600 MHz and ^{15}N R_1 and R_2 values at 900 MHz were used to determine an isotropic diffusion tensor for IN $^{50-212}$. Residues used for tensor optimization were filtered by the following criteria: only residues existing in regular secondary structure were selected for analysis. Additionally, residues were removed with a chemical shift order parameter <0.7 (71, 72), or a $^{15}\text{N}\{-^1\text{H}\}$ NOE <0.7 . The resulting list of residues was: 60, 66, 71, 72, 82, 83, 89, 92, 93, 98, 99, 100, 105, 106, 112, 113, 115, 117, 120, 126, 127, 129, 130, 132, 133, 134, 135, 136, 137, 138, 139, 155, 157, 159, 161, 163, 164, 168, 175, 176, 180, 181, 182, 183, 184, and 185. The diffusion tensor was fit using ModelFree 4.15 (73) and FAST-ModelFree (74). Lipari-Szabo order parameters (75–77) were obtained for all residues from the 600 MHz data. Model selection was performed by FAST-ModelFree (74) using the previously determined, fixed isotropic diffusion tensor.

Magnesium and Manganese Titrations—Magnesium titrations were performed by direct addition into 340 μM IN $^{50-212}$, using a 2 M stock solution of MgCl_2 . ^{15}N - ^1H TROSY-HSQC spectra were recorded at 2, 10, 20, and 40 mM MgCl_2 on a Bruker Avance 750 MHz spectrometer. Spectral quality improved substantially at MgCl_2 concentrations above 10 mM, but reasonable TROSY HNCA/HN(CO)CA spectra could also be obtained at 5 mM MgCl_2 , and these were used to follow changes in C^α chemical shifts. To test for specific binding of $\text{Mg}^{2+}/\text{Mn}^{2+}$, we added 10 and 50 μM MnCl_2 to the 40 mM MgCl_2 standard NMR buffer. pH stability before and after the titrations was confirmed by direct measurement. Chemical shift changes as a function of Mg^{2+} concentration were quantified using a weighted mean chemical shift change,

$$\Delta\delta_{\text{avg}} = \sqrt{\frac{1}{2}(\Delta\delta_{\text{HN}}^2 + (\Delta\delta_{\text{N}}/5)^2)} \quad (\text{Eq. 1})$$

as described by Grzesiek *et al.* (78). Peaks present in at least three of the four spectra were used to determine a two-state K_d

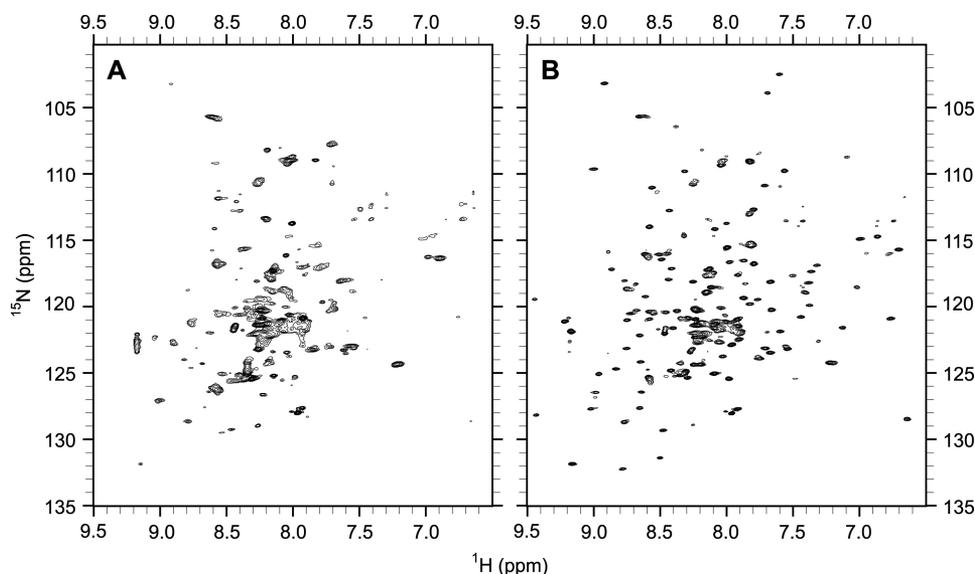


FIGURE 1. ^1H - ^{15}N TROSY-HSQC NMR spectrum of IN^{50-212} at 800 MHz, 298 K, and pH 7.0 in 5 mM MgCl_2 (A) and 40 mM MgCl_2 (B). All other buffer conditions are identical, and the spectra are plotted at similar contour levels.

value. Fit parameters were bootstrapped using the R statistical computing package (79).

Examination of the MnCl_2 peak attenuation and comparison with the crystal structure 1BL3 (28) revealed the possibility of a second Mn^{2+} binding site not seen in the crystal structure. In the fast exchange limit, the weighted average relaxation rate for a single proton site experiencing paramagnetic relaxation enhancement (PRE) resulting from a bound Mn^{2+} is given by,

$$R_{\text{eff}} = \frac{K_d^{-1}x}{1 + K_d^{-1}x} R_0 + \frac{1}{1 + K_d^{-1}x} (R_0 + A\tau^{-6}) \quad (\text{Eq. 2})$$

where R_0 is the intrinsic proton relaxation rate, K_d is the Mn^{2+} dissociation constant, x is the concentration of Mn^{2+} , and A is a scaling constant for the PRE term, determined primarily by the protein tumbling time τ_c (80–82). If NMR line shapes are Lorentzian, and ignoring magnetization decay during the fixed duration transfer delays in the NMR pulse sequence, the peak height attenuation is simply the ratio of rates without and with Mn^{2+} . Using the crystallographic Mg^{2+} ion, and assuming equivalent binding and relaxation rates across residues 60–180, we fit R_0 , K_d , and A to the data, and then used those values to optimize a second Mg^{2+} binding site location near the C-terminal helix. The χ -squared statistic measuring the difference between the observed and predicted attenuation of residues 181–212 was optimized as a function of Mn^{2+} position using the R statistics package (79). The resulting fit (red curve in Fig. 4B) was robust to small changes in R_0 , K_d , and A and consistently produced the same Mn^{2+} site (Fig. 2, B and C). Repeating the analysis for PDB entries 1BIU (27) and 1QS4 (30) yielded very similar locations for the second Mn^{2+} .

Raltegravir Purification and Titration—Raltegravir (Isentress, Merck) tablets were obtained commercially. After removal of the tablet coating, tablets were crushed and suspended in deuterated dimethyl sulfoxide (DMSO). After centrifugation for 10 min at $16,000 \times g$, the supernatant was collected. A one-dimen-

sional ^1H NMR spectrum was recorded and compared with synthetic, pure raltegravir obtained from Dr. Yves Pommier. The comparison revealed the tablet extract to be $\sim 90\%$ pure, and this material was used in titrations without further purification. Raltegravir was quantified by NMR using a trimethylsilyl propionic acid standard, and stock solution concentrations were 45–50 mM in DMSO. The raltegravir extinction coefficient in DMSO at 341 nm was determined to be $9960 \pm 120 \text{ M}^{-1} \text{ cm}^{-1}$. The uncertainty in this number was estimated from the variation of peak volumes in the one-dimensional ^1H NMR spectrum.

Titrations were performed on a Bruker Avance 750 MHz spectrometer using a 49 mm stock solution of raltegravir and a NMR sample of $245 \mu\text{M}$ IN^{50-212} . Raltegravir was added in steps of 200–1000 μM up to a drug:protein ratio of 14:1. The final DMSO concentration was 6.5% (v/v). To assess the effects of DMSO on IN^{50-212} , the titration was repeated with DMSO alone, and peak positions of a ^{15}N - ^1H TROSY-HSQC were compared at corresponding titration points. Although chemical shift changes were observed during the raltegravir titration, nearly identical changes were observed with DMSO alone, indicating that the effects of raltegravir on the isolated core domain are very small.

RESULTS

Chemical Shift Assignments—The F185K,W131E IN^{50-212} variant has been used in crystallization studies and is catalytically competent for 3' processing (27). However, when studied by NMR, this construct yielded very poor spectra and required high salt concentrations, which had a strong adverse impact on the sensitivity of our experiments. Site-directed mutagenesis was employed to develop a soluble variant of IN^{50-212} amenable to NMR characterization. First, we introduced C56S to eliminate potential disulfide bonding between monomers. It is known that this mutation does not hinder disintegration (27). Then, working from the termini to avoid interfering with the active site, we added two Asp residues to lower the protein pI. These residues are not visible in most crystal structures and therefore should not affect the conformation of the active site. The final variant, Q53E,C56S,W131E,F185K,Q209E, exhibited good solubility at 150 mM NaCl, pH 6.8, a substantial improvement over the original variant, which required NaCl concentrations larger than 250 mM. In this paper, IN^{50-212} designates this particular variant of the integrase core domain.

Although IN^{50-212} was soluble, spectra were still unfit for NMR assignment purposes (Fig. 1A). Initial experiments included 5 mM MgCl_2 (27), but increasing the MgCl_2 concentration was found to dramatically increase the spectral quality

Conformation and Dynamics of HIV Integrase Core Domain

to the point where triple resonance assignment techniques could be employed (Fig. 1B). Not only did the line width of existing peaks become more uniform, but more peaks were observed as well. At 40 mM MgCl₂, the marginal improvement in spectral quality diminished, and no new peaks were observed beyond this concentration. Characterizing the IN^{50–212} variant by gel filtration chromatography revealed that, whereas still eluting as a dimer, the particle had diminished in size somewhat, possibly indicating a structural collapse of a partially disordered region (data not shown).

The standard suite of TROSY triple resonance experiments were employed to assign backbone H^N, N, C', C^α, and C^β resonances on ¹⁵N-, ¹³C-, ²H-labeled IN^{50–212} (57). However, because of spectral overlap, several assignments remained ambiguous even after all other connectivities were established. To resolve this problem, ¹⁵N residue-specific labeling was employed to identify particular classes of amino acids and to confirm existing assignments. Additionally, a three-dimensional ¹⁵N-¹H MT-PARE HMQC-NOESY spectrum was recorded to determine H^N-H^N connectivities through space where traditional J-correlated methods failed to yield unambiguous results. The collected spectra allowed for the assignment of 97% of all non-prolyl backbone and ¹³C^β resonances from residues 51 to 212 as well as two sets of assignments for residues 205 to 212. Assignments have been submitted to the BMRB data base (accession number 16695; supplemental Table S1).

Structural Information from Chemical Shifts—Because of the size of dimeric IN^{50–212} (36 kDa) and its relatively unfavorable NMR properties, high quality NMR spectra of uniformly protonated samples were not obtainable, precluding extensive side chain ¹H assignment and traditional ¹H NOE-based methods for structure determination that are successful on smaller proteins (83, 84). Although work is ongoing to apply small angle x-ray scattering and residual dipolar coupling constraints to develop refined models of the integrase core domain, recent developments have allowed a wealth of structural information to be inferred from chemical shifts alone.

It is well known that C^α and C^β secondary chemical shifts (*i.e.* their deviations from random coil values) are strongly correlated with backbone torsion angles and thereby with secondary structures in folded proteins (85, 86). We characterized secondary shifts in IN^{50–212} and compared them with the secondary structure reported for chain C in the crystal structure of PDB entry 1BL3 (28) (supplemental Fig. S1). This is the highest resolution crystal structure of the core domain obtained to date in the presence of Mg²⁺. The agreement between the NMR observations and the x-ray data is very good, and the TALOS+ secondary structure predicted from chemical shifts (72) predicts every unit of secondary structure observed in the crystal structure (supplemental Fig. S1). Some minor differences are seen at the ends of helices and strands, but these may be attributed to fraying in solution, especially considering the dynamic nature of the molecule (see below).

To further characterize the structure of IN^{50–212} in solution, we created a CS-ROSETTA model of a core domain monomer (68). This algorithm relies on backbone atom chemical shifts to select possible peptide fragments, which are then assembled and scored according to the ROSETTA scoring functions. The

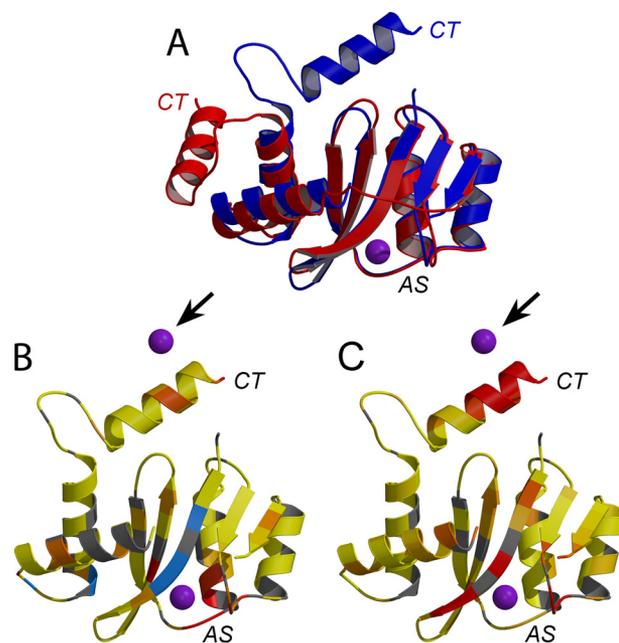


FIGURE 2. Structural modeling from IN^{50–212} NMR data. A, the lowest energy CS-ROSETTA monomeric model (red) based on H^N, N, C^α, C', and C^β chemical shifts, overlaid on the crystal structure at 2.0-Å resolution (PDB entry 1BL3) (28). The backbone root mean square deviations is 1.62 Å for residues 60–186. Because the CT helix (residues 196–209) interacts across the dimer interface, its position is ill-defined in the NMR model. B, a summary of chemical shift changes observed at moderate (10 mM) and high (40 mM) MgCl₂ concentrations. Residues are shaded yellow or red, corresponding, respectively, to smaller or larger chemical shift differences. Resonances that are broadened beyond detection at 10 mM MgCl₂ are shaded blue. Data from B are plotted as described in the legend to Fig. 4A. C, a summary of peak intensity attenuations observed when 50 μM MnCl₂ is added to IN^{50–212} in 40 mM MgCl₂. Large changes are shaded red, intermediate changes are orange, and smaller attenuations are yellow. Data from C are plotted as described in the legend to Fig. 4B. Overlapped or undetected resonances in B and C are shaded gray. Mg²⁺ ions are displayed as purple spheres, and in B and C the Mn²⁺/Mg²⁺ binding site determined by PRE is marked with an arrowhead. The C-terminal helices are labeled CT, and the active site is labeled AS. Molecular graphics were created with MOLSCRIPT (106) and RASTER3D (107).

resulting structure is not a true NMR solution structure and is less than ideal for two reasons. First, no experimental data were used to determine side chain conformations, although nine detected long range ¹H^N-¹H^N NOEs were included and helped to ensure correct pairing of the β-strands. Second, dimer calculations performed on IN^{50–212} with CS-ROSETTA did not converge because of the size of the system. As a result, we were limited to generating models for the monomeric unit (Fig. 2A, red conformation). In the top 10 low-energy CS-ROSETTA models, the position of the C-terminal helix is ill-defined, retaining its helical fold yet sampling multiple orientations. The R₁/R₂ values for this helix are similar to those for residues 60–186 (Fig. 3, discussed below), indicating that the C-terminal helix has the same rotational diffusion properties as the rest of the core domain and is not actually sampling multiple conformations. The CS-ROSETTA simulation likely lacks sufficient restraints to orient the C-terminal helix correctly, because it has no dimeric partner to pack against. Thus, we have no reason to doubt the crystallographic orientation, which pairs the two helices along the dimer interface. Excluding the C-terminal helix, the CS-ROSETTA model for the remainder of the core domain agrees well with crystal structures, and the backbone

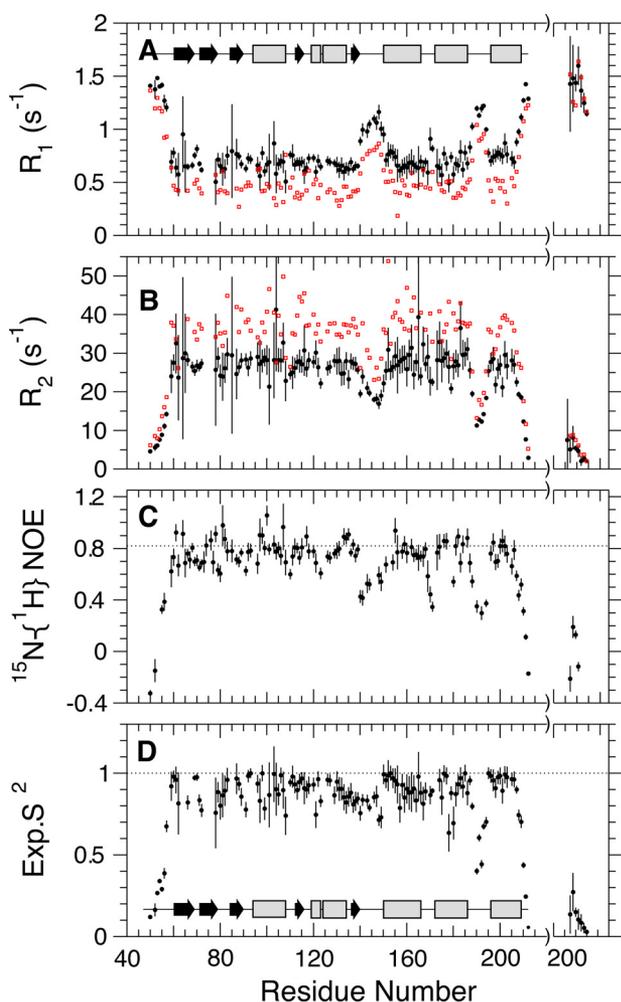


FIGURE 3. ^{15}N relaxation values measured for IN^{50-212} at 298 K. *A*, R_1 values; *B*, R_2 values; and *C*, $^{15}\text{N}\{-^1\text{H}\}$ NOE values at 600 MHz (black circles) and 900 MHz (red squares). Uncertainties are estimated from spectral noise and are comparable between 600 and 900 MHz. In *C*, the theoretical maximum NOE at 600 MHz (0.82) is indicated as a dashed line (108). *D*, Lipari-Szabo order parameters determined using measurements from panels *A*–*C*, as described in the text. *IN* exhibits some dynamics on the picosecond-nanosecond time scale in the active site region (residues 140–153), but several of the residues in the 188–196 loop are markedly more dynamic on this time scale. The alternate assignments for C-terminal residues 204–212 (broken axis at the right) indicate that these assignments correspond to a dynamically disordered state. The secondary structure from PDB entry 1BL3, shown in panels *A* and *D*, is plotted as helices (outlined rectangles) and strands (solid black arrows).

root mean square deviations for residues 60–186 differ by only 1.6 Å relative to PDB entry 1BL3, chain C (Fig. 2*A*, blue conformation).

Partial Unfolding of the C-terminal Helix—During the assignment process, more resonances were observed in the HNCO spectrum than were expected given the number of backbone amides in the core domain. Additionally, two strong resonances were observed near 7.9 ^1H ppm and 128 ^{15}N ppm, where the C-terminal residue typically appears. Linking both C termini to their preceding residues resulted in two sets of connectivities: one that became less and less intense and impossible to trace after seven residues, and one that could be assigned with the rest of the core domain. Because both fragments show C^β resonances consistent with the C-terminal residues, initially we suspected that the alternate assignment was simply a proteolytic fragment of the C terminus. However, closer inspection

of the NOESY data reveals cross-peaks between both sets of assignments, suggestive of conformations in slow exchange (supplemental Fig. S3). The secondary shifts for the shorter assignment fragment are near zero for both C^α and C^β resonances indicating that this conformation is highly disordered. At 40 mM MgCl_2 , it is also the less populated conformation based on peak intensity. Assuming a two-state model and using peak volumes determined by NMRPipe (65), we estimate an apparent K_{eq} of 1.7 ± 0.6 favoring the folded conformation of the C-terminal helix at 40 mM MgCl_2 . Given the ratio of cross-peaks to diagonal peaks as well as the NOE mixing time (150 ms) it is possible to calculate k_{ex} , the sum of forward and backward exchange rates (87). Although kinetics are harder to determine because of the overlap observed for cross-peaks in the NOESY spectrum, the ratios observed for residues 206–207 and 209–212 are consistent and averaged to $0.8 \pm 0.2 \text{ s}^{-1}$.

To probe whether the original W131E,F185K construct also showed C-terminal helix unfolding, we recorded tr-HSQC spectra at 40 mM MgCl_2 . Two C-terminal peaks were observed, although the peak corresponding to the unfolded form was much weaker, less than 1% of the peak corresponds to the primary conformation (data not shown). Therefore, it is likely that the Q209E mutation used in our studies affects the stability of the final helix. Interestingly, even at 40 mM MgCl_2 , ~30% of the backbone amide resonances remain missing in the F185K,W131E spectrum. If Q209E destabilizes the C-terminal helix, one would expect the original construct to yield a better spectrum, but this is not the case. It is possible that this helix modulates the conformation of the rest of IN^{50-212} depending on how it folds, and this may be important for full-length integrase function.

Fast Time Scale Dynamics—To probe dynamics of the core domain, we measured ^{15}N R_1 , R_2 , and heteronuclear $^{15}\text{N}\{-^1\text{H}\}$ NOE values at both 600 and 900 MHz on IN^{50-212} (Fig. 3 and supplemental Table S2). These values characterize the flexibility of the core domain on the picosecond to nanosecond time scale and are useful for determining which parts of the chain are disordered. Assessed by heteronuclear NOE values <0.6 , the most dynamic regions of the protein (after the termini) are residues 140–153, near the active site loop, and residues 185–195, before the final C-terminal helix. ^{15}N relaxation data indicate that N-terminal residues 50–56, lacking electron density in the X-ray structure, are also dynamically disordered in solution, suggesting that the Q53E,C56S mutations have little impact on the ordered structure in the core domain. The active site is known to be conformationally promiscuous (26), and corresponding residues in RNase H (88) and bacteriophage Mu transposase (89) are also flexible. Many crystal structures lack density for the loop containing the catalytic residue Glu¹⁵² (24, 27). Similarly, multiple conformations have been observed for residues 180–195, and in the extreme cases both short stretches of α -helix (27) as well as β turns (32) have been observed. The conformational heterogeneity of these regions seen in crystals is therefore consistent with the backbone dynamics observed in solution.

To further investigate IN^{50-212} dynamics, we used the ModelFree (73) and Fast ModelFree (74) programs to determine the isotropic tumbling time τ_c and Lipari-Szabo order parameters

Conformation and Dynamics of HIV Integrase Core Domain

(75–77). The τ_c for the core domain particle is 20.5 ± 0.2 ns, which is somewhat larger than expected for a 36-kDa dimeric system. Multiangle light scattering confirmed that IN^{50–212} is fully dimeric at concentrations of 55 μ M, but partial unfolding of the C-terminal helix or transient aggregation of the core domain particles at the higher concentrations used in NMR may explain the elevated τ_c value. Of note is that the loop containing Glu¹⁵² is somewhat dynamic on the picosecond-nanosecond time scale, whereas the backbone of catalytic residues Asp⁶⁴ and Asp¹¹⁶ are found to be well ordered. In the Model-Free analysis, residues 185–195 remain highly dynamic compared with residues 140–153, suggesting that the C-terminal loop is the most disordered in the molecule, even more so than the active site loop. Comparing the dynamics observed at the active site and the 185–195 loop, it is interesting to note that the latter shows behavior typical for a disordered loop, with the order parameter lowest near its midpoint and gradually increasing when approaching the ends of the loop. In contrast, residues 140–153 show a more limited degree of internal dynamics, which remains relatively flat across the entire active site loop, suggestive of concerted motions between distinct states, rather than simple increasing disorder toward its midpoint.

In agreement with the secondary chemical shift data, the ¹⁵N relaxation data and NOE values also indicate that residues 205–212 in the alternate conformation are disordered (Fig. 3, right). Although this conformation appears to be promoted by the mutations used to solubilize IN^{50–212}, it is nevertheless, surprising that these residues are in slow exchange, as helix fraying typically occurs on the fast exchange, submicrosecond time scale. Because of the disorder seen for the alternate conformation, it is likely that the entire C-terminal helix unfolds, with residues prior to residue 205 not observable because of base-catalyzed hydrogen exchange, which increases when further removed from the carboxyl terminus. This would reconcile the observed dynamics with the slow exchange implied by multiple resonance peaks.

Magnesium and Manganese Binding—The addition of MgCl₂ to NMR samples of IN^{50–212} dramatically improved the spectral quality. To investigate the causes of this improvement, we performed titrations on the core domain, from 2 to 40 mM MgCl₂. During this titration, peaks were observed to both shift and sharpen as MgCl₂ was added, indicating intermediate exchange (supplemental Figs. S3–S5). The weighted mean chemical shift differences between 10 and 40 mM MgCl₂ reveals that most residues in IN^{50–212} are affected by Mg²⁺ concentration to some extent, but catalytic residues Asp⁶⁴ and Asp¹¹⁶ are very strongly affected as indicated by the fact that they do not become visible in NMR spectra until at least 20 mM MgCl₂ is present (supplemental Fig. S4). The catalytic loop (residues 140–150) and the start of $\alpha 4$ are visible at lower MgCl₂ concentrations, and significant chemical shift changes are observed for these residues (Fig. 4A; mapped onto the structure in Fig. 2B). Additionally, another region that is particularly sensitive to MgCl₂ concentration comprises residues 90–96. This region is not known to be involved in 3' preprocessing or strand transfer. Nevertheless, Glu⁹² coordinates the Mg²⁺ ion observed crystallographically (27), and mutations to this residue confer

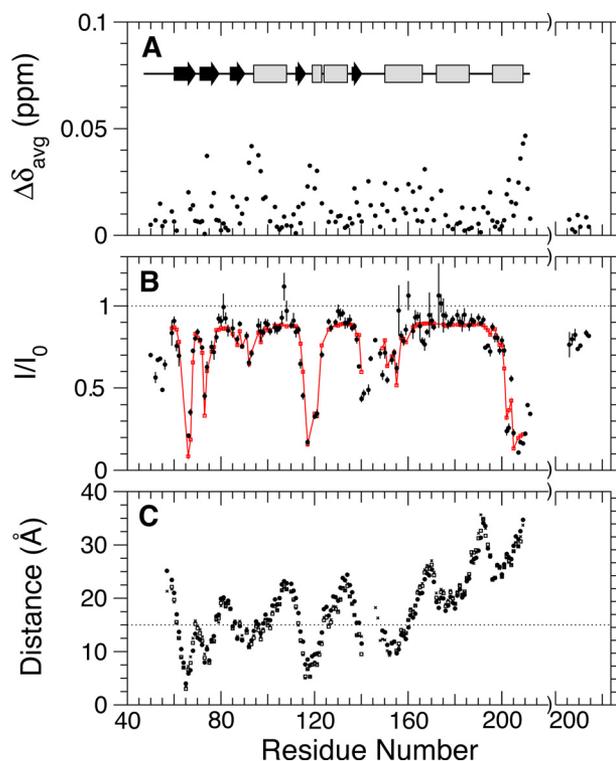


FIGURE 4. Summary of Mg²⁺/Mn²⁺ binding effects. A, normalized chemical shift differences for IN^{50–212} in a TROSY-HSQC at 10 mM MgCl₂ relative to 40 mM MgCl₂. The weighted average $\Delta\delta_{\text{avg}}$ is calculated as in Equation 1 (78). The chemical shift changes are most pronounced near the active site residues, with peaks near Asp⁶⁴ and Asp¹¹⁶, and elevated values for residues 140–150. A structurally proximal region near Glu⁹² is also affected. Changes are also observed at the C-terminal helix. B, peak intensity ratios in a TROSY-HSQC resulting from paramagnetic relaxation as 50 μ M MnCl₂ is added to IN^{50–212} in 40 mM MgCl₂. Black circles show the experimental data points, and the red curve shows the predicted attenuation using the Mg²⁺ position in PDB entry 1BL3 as well as the additional C-terminal binding site described in the text and shown in Fig. 2. B and C. C, distances between H^N atoms and the crystallographic Mg²⁺ observed in PDB entries 1BIU (●), 1BL3 (□), and 1QS4 (×). Residues that are close to the Mg²⁺ in C are more attenuated in B, except for residues near the C terminus. Thus, crystallographic Mg²⁺ alone cannot solely account for the observed attenuation in B. The absence of PRE effects in the C-terminal alternate assignments (broken axis at right in A and B) indicates that this conformation does not bind Mn²⁺. The secondary structure from 1BL3 is overlaid in panel A.

resistance to the IN inhibitors raltegravir (90) and elvitegravir (91, 92), suggesting that this loop may be important for catalysis. Although many peaks disappear entirely below 5 mM MgCl₂ (including all of the catalytic residues), we are able to determine apparent K_d values from our experiments. Residues Gly¹¹⁸, Asn¹²⁰, Thr¹¹⁵, and Asn¹⁵⁵, proximal to the active site, are most sensitive to MgCl₂, with a K_d of 1.75 ± 0.18 mM. Residues Ala⁹¹, Thr⁹³, Glu⁹⁶, and Tyr⁹⁹ have a K_d of 1.84 ± 0.28 mM, and probably track the same binding event. Other residues titrate with weaker affinities but correspondingly larger uncertainties, e.g. Gly¹⁶³ with a K_d of 3.7 ± 1.0 mM (supplemental Fig. S3). The K_d values for all residues generally differ by no more than 2 S.D. of experimental uncertainty.

Integrase is functional when Mg²⁺ or Mn²⁺ is present as a divalent cation (93, 94), and it is therefore reasonable to expect that the core domain will bind Mn²⁺ just as it binds Mg²⁺. Manganese has the additional property of being paramagnetic and therefore increases relaxation rates of nearby nuclei when bound. To identify potential Mn²⁺ binding sites, we performed

titrations of MnCl_2 into IN^{50-212} at 40 mM MgCl_2 . This high background concentration of MgCl_2 masks any nonspecific Mn^{2+} binding effects that would otherwise be observed. As assessed by peak intensity changes, the attenuation profile at 50 μM MnCl_2 reveals several clusters of residues in IN^{50-212} where attenuation caused by Mn^{2+} binding is strong (see Fig. 4B; mapped onto the structure in Fig. 2C).

We compared the observed attenuation profile (Fig. 4B) with crystal structures to determine whether the crystallographic Mg^{2+} binding site could account for the observed PRE. For most of the residues in IN^{50-212} , the minimum distance between H^N atoms (added using REDUCE (95)) and the crystallographically observed Mg^{2+} ion correlate very well (Fig. 4, C, compare with B). Specifically, the correlation is very good for residues 60–180. After residue 180, however, the correlation breaks down, indicating that the crystallographic Mg^{2+} site is insufficient to describe Mn^{2+} binding for these residues. This finding suggests that another Mn^{2+} binding site exists near the C-terminal helix. To test this, we used the crystallographic Mg^{2+} binding site to calibrate a $1/r^6$ potential for residues 60–180 (see “Experimental Procedures”). Then, this potential was used to position an additional Mn^{2+} binding site that could reproduce the broadening seen under identical conditions for residues 181–212. The resulting curve, using the x-ray coordinates as a reference frame, shows good agreement between experimental and predicted PRE rates (Fig. 4B). The additional $\text{Mn}^{2+}/\text{Mg}^{2+}$ binding site is able to reproduce the data for the final helix, and when combined with the original binding site, the overall attenuation is reproduced very well.

Because of the symmetry requirements, the second Mn^{2+} binding site is forced to lie on the C_2 axis of the symmetrical dimer (top of the molecule in Fig. 2, B and C). However, attempting to fit multiple Mn^{2+} atoms without this constraint identified essentially the same position. The Mn^{2+} binding site lies in a cleft between the two Glu²⁰⁹ side chains, and this suggests that the binding site was introduced artificially via the Q209E mutation made to stabilize IN^{50-212} . To confirm this observation, a similar set of experiments was carried out on F185K,W131E IN^{50-212} , without the additional Q209E mutation (data not shown). Spectral quality was poorer without the Q209E mutation, but MgCl_2 binding was observed near the catalytic loop (residues 64 and 140–157), and spectral quality improved for these residues as MgCl_2 was added, similar to what was observed with Q209E (supplemental Fig. S4). Constructs with and without Q209E exhibit virtually indistinguishable active site chemical shifts, and bind Mg^{2+} at the active site in a similar way, indicating that the Q209E mutation does not significantly affect the catalytic loop. On the other hand, Mn^{2+} or Mg^{2+} binding effects at the C terminus were absent for the F185K,W131E variant. Moreover, we also do not observe binding for the IN^{50-212} conformer that contains the disordered C-terminal helix (Fig. 4, A and B, right). Given these results and the position of the additional binding site, we hypothesize that the C-terminal helix in wild type IN^{50-212} is unstable, and the additional engineered Mg^{2+} binding site in our variant helps rigidify this helix, thereby stabilizing the entire core domain particle, leading to better NMR spectra. However, the Q209E mutation destabilizes this helix when Mg^{2+} or Mn^{2+} is not

present (supplemental Fig. S4). At this point it is unclear to what extent the stability of the C-terminal helix impacts the function of the full-length enzyme. However, the structure of this helix controls the position of the DNA binding domain relative to the catalytic core domain, which differs between models of the full-length enzyme and crystal structures of the hexameric catalytic core plus DNA binding domain construct (33).

Raltegravir Binding—Raltegravir is a member of a class of diketo acid IN inhibitors currently approved for clinical use (17). It functions primarily by blocking integrase at the strand transfer step, but at higher concentrations it also inhibits 3' processing (96, 97). Previous studies have shown that raltegravir fails to bind integrase tightly in the absence of DNA, nor does it appear to interact with the core domain alone (96). However, given that NMR is well suited to detect even very weak levels of binding up to K_d values greater than 1 mM, we attempted to titrate raltegravir into solutions containing IN^{50-212} . Raltegravir (obtained commercially) was solubilized to high concentrations in DMSO and added to a final drug:protein ratio of 14:1 and a final DMSO concentration of 6.5% (v/v). Residues 143–152 were observed to shift, but shift changes were very small after correction for the contribution from DMSO alone, with the maximum chemical shift change, $\Delta\delta_{\text{avg}}$, being 0.03 ppm. Binding curves had not approached a plateau by this point in the titration. Assuming a typical lower limit of 0.1–0.2 ppm chemical shift change upon ligand binding for at least some of the protein backbone amide resonances, this indicates less than about 20% drug binding at this point in the titration, *i.e.* $K_d \geq \sim 10$ mM. Thus, in agreement with previous studies, we find no evidence of significant binding to the core domain alone in the absence of target DNA, and work is continuing to develop a minimal system that binds both DNA and inhibitors.

DISCUSSION

This work represents the first study of the IN catalytic core domain by NMR. The data presented here confirm many prior findings about integrase to apply to the solution state, but they also reveal a few surprises. In particular, the dramatic effect of the high magnesium concentration on stability of the core domain was unexpected and may provide clues about the importance of dynamics during catalysis. Additionally, the ModelFree ¹⁵N relaxation analysis presented here confirms that the active site catalytic loop from residues 140 to 153 is dynamic on the picoseconds-nanosecond time scale, but it does not show the characteristics often seen for flexible linker regions in proteins. Instead, it shows dynamic behavior of moderate amplitude that is relatively homogeneous across the active site loop, suggestive of concerted motions between distinctly different conformations. Finally, our work with drug binding highlights some of the difficulties that will arise for IN as structure-activity relationships are studied using NMR.

Previous studies have characterized a Mg^{2+} - or Mn^{2+} -induced conformational change in integrase. Asante-Appiah and Skalka (34) were the first to characterize this effect on the full-length IN construct. They identified the surprising result that increased metal concentrations (up to 10 mM) were able to confer resistance to proteolytic digestion of IN. Furthermore, they

Conformation and Dynamics of HIV Integrase Core Domain

found that this behavior was unique to HIV-1 IN, and they proposed that the relevant conformational change occurred between the core and C-terminal domains. Later work confirmed that the metal-driven conformational changes were influenced by at least two binding sites: one at the active site and one at an unknown binding site elsewhere in the molecule (35). It was hypothesized that this second binding site changes the relative orientation of the IN domains, which resulted in the increased proteolytic protection observed at higher metal concentrations.

Our work confirms that magnesium induces significant conformational changes in the catalytic domain of IN. In particular, the active site residues Asp⁶⁴, Asp¹¹⁶, and Glu¹⁵² show peak broadening at low MgCl₂ concentrations, indicating exchange with at least one other (possibly disordered) conformation (supplemental Fig. S4). Thus, we can confirm that metal binding at the active site is essential for conformational stability in IN. An unexpected result is the degree to which other, non-catalytic residues are affected by the addition of MgCl₂ to the sample. Even in the original F185K,W131E construct, spectral quality improves as magnesium concentration is increased, indicating that this effect is not simply a result of the IN^{50–212} mutations used in our study. Examining the chemical shift changes as a function of MgCl₂ (Fig. 2B and supplemental Figs. S1 and S5), it is clear that magnesium may interact with multiple regions of the core domain to act as a general stabilizing agent.

Highest quality NMR spectra were obtained with the Q209E mutation, which introduces an additional, specific Mg²⁺/Mn²⁺ binding site between the two C-terminal helices in the catalytic core dimer. Introducing this binding site was unexpected, because Gln²⁰⁹ is rarely seen as a helix in IN crystal structures, but we hypothesize that this binding site may be directly responsible for the improved NMR spectra over F185K,W131E integrase. Given that the final loop between helices $\alpha 5$ and $\alpha 6$ is highly dynamic, the orientation of the C-terminal helix may be ill-defined in solution. Introducing the binding site may help stabilize the helical packing, thereby stabilizing the entire core domain-fold. Although the Q209E binding site is artificial, it may nonetheless hint at another, biologically relevant metal binding site. Many Asp and Glu residues line the helix connecting the core and C-terminal domains, and a slight reorientation of the crystallographically observed helix may be sufficient to form such a site (33). This site may correspond to the additional site hypothesized by Asante-Appiah *et al.* (35).

Uncertainty exists as to whether the IN active site binds one or two Mg²⁺/Mn²⁺ ions in the absence of DNA substrate. The proposed mechanism requires two metal cations (98), and two Mn²⁺ ions are observed to bind the active site in DNA-bound PFV IN (37). Non-physiological metals such as Zn²⁺, Cd²⁺, and Ca²⁺ have been complexed to avian sarcoma virus IN as well as HIV IN in the absence of DNA (33, 99), but to date no crystal structure exists with two Mg²⁺ or Mn²⁺ ions in the active site. Because it has poorer geometry, it is likely that the second binding site has weaker affinity (100). Unfortunately, our data cannot resolve this issue. We do not observe positive evidence for two binding sites from chemical shift changes during titrations,

and the chemical shift titrations for residues near the active site can each be fit with a single apparent K_d value. This suggests that if multiple metal binding sites are present, K_d values for them cannot be very different. The fact that apparent K_d values for different active site residues do not differ much beyond the experimental error from one another (supplemental Fig. S3) confirms this conclusion. So, even though our data are compatible with two Mg²⁺ ions binding with similar affinities, it is equally compatible with the case where only one Mg²⁺ binds.

We also note that our observations on IN stability *versus* magnesium concentration may be helpful for producing more crystal structures of the core domain as well as other multidomain constructs. To date, crystal structures have been determined at relatively low concentrations of MgCl₂, typically 5 mM or less (27, 29, 31). At this concentration, the conformational change we observe is far from complete, and our spectra continue to improve up to concentrations of 40 mM MgCl₂. Although the lower protein concentrations used in crystallization translate to a somewhat lower required concentration of MgCl₂, we expect that fully saturating IN metal binding will aid in the production of new high-quality crystal structures.

A topic of key interest in IN biochemistry is the structure and function of the active site. Given that the core domain must catalyze both 3' processing and strand transfer, it is not surprising that the domain would be dynamic and populate multiple conformations in its unbound state. Despite the wealth of structural information available on the core domain, little is known about the conformation of IN during catalysis. Recently, a co-crystal of PFV integrase and DNA has been determined, and this promises to guide our understanding of HIV integrase structure and function (37). In their structure, Hare *et al.* (37) observed an extended $\alpha 4$ helix with a kink for residues homologous to Pro¹⁴⁵ and Gln¹⁴⁶. These two residues are in the 3_{10} helix and interact with the viral DNA end to stabilize the nucleoprotein particle. In the apo state, it is known from other x-ray studies that the catalytic loop can sample multiple conformations and is relatively unstable compared with the rest of the molecule (24, 26, 27). It is not known, however, whether the active site samples the kinked conformation seen in the PFV IN structure. One insightful study by Greenwald and co-workers (29) made Gly to Ala mutations to rigidify the catalytic loop. They found that G140A and G149A variants of IN reduced catalytic rates to 6–12% of their normal levels, and a double mutant construct virtually abolished activity. Their crystal structures indicated that these variants stabilize a helical conformation, extending the N-terminal end of the $\alpha 4$ helix to Pro¹⁴⁵ in the G140A,G149A variant. Given that the PFV IN crystal structure is kinked at Gly²¹⁸ (corresponding to Gly¹⁴⁹ in HIV-1 IN), it is easy to see how mutations to this residue could affect catalysis. It is less clear why mutations at Gly¹⁴⁰ would influence catalysis. In PFV, the analogous residue is Ser²⁰⁹, which is even more sterically constrained than Ala in the G140A variant. It is therefore likely that dynamics play an important role in catalysis as well.

Because of the uncertainty in the mechanism of catalysis, computational studies have examined the conformation and dynamics of the active site loop in the core domain apo state. An initial molecular dynamics study by Lins *et al.* (38) observed a

Conformation and Dynamics of HIV Integrase Core Domain

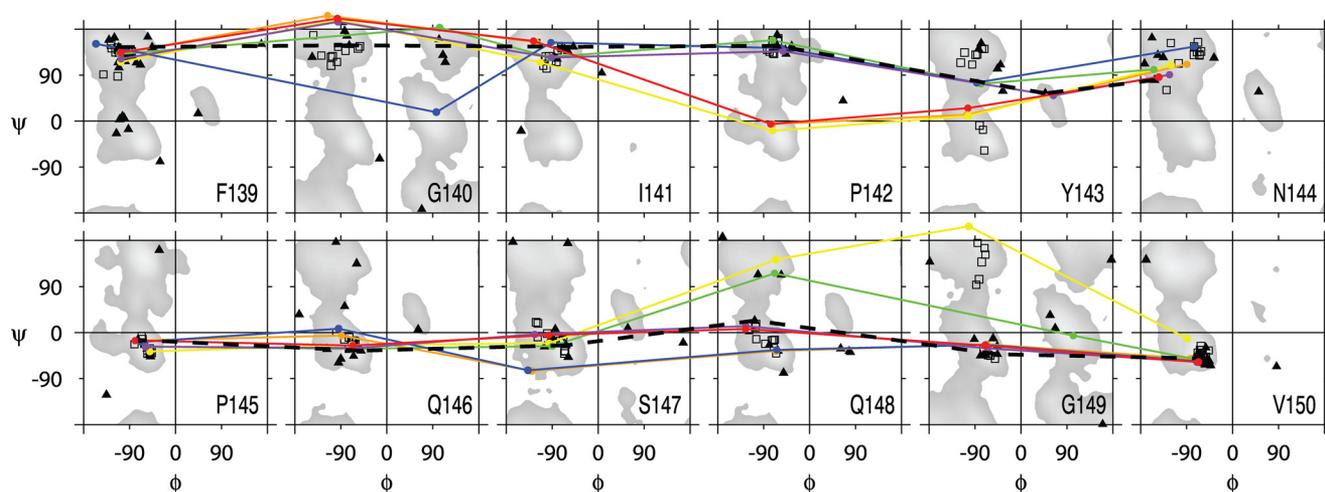


FIGURE 5. Comparison of the CS-ROSETTA (68) model of the active site loop with the available IN⁵⁰⁻²¹² structures. Each Ramachandran map corresponds to a residue in integrase, from Phe¹³⁹ to Val¹⁵⁰. These active site residues are the least well defined in crystal structures, with φ , ψ values for x-ray structures that include (subsets of) these residues shown as *filled black triangles*. TALOS+ (72) was independently used to select 3-residue peptide fragments with φ , ψ values based on chemical shifts. These φ , ψ values are shown as *open black squares*. The φ , ψ values of six low-energy CS-ROSETTA clusters are shown as *connected filled circles*, with colors corresponding to the conformations shown in Fig. 6. The conformations suggest that the N terminus of helix α 4 frays in solution, commensurate with concerted motions of the loop containing Tyr¹⁴³. The conformation of corresponding residues in the PFV IN-DNA complex are shown as a *black dashed line*.

tendency for the helix extension at the N-terminal end of α 4, and dynamics on the picosecond time scale were commensurate with elevated crystallographic temperature factors. A later study extended the time scale of simulations to 40 ns (39). During this time, the catalytic loop was observed to close across the surface of the other catalytic residues, leading the authors to propose a gating mechanism for these residues. Additionally, both papers support an important role for Tyr¹⁴³, where the side chain increases enzymatic efficiency by helping to position an activated water nucleophile during catalysis.

Our CS-ROSETTA model for the core domain lacks the long range constraints typically used to generate high quality NMR macromolecular models, but the correlation between backbone torsions and chemical shifts is well established (68, 72, 85, 101, 102), and chemical shift-based modeling can provide some insight on the structural organization of the active site. Because the CS-ROSETTA energy of the entire IN⁵⁰⁻²¹² monomer is dominated by its overall molecular topology, low energy conformers do not necessarily correspond to optimal low energy active site conformations. Therefore, we also carried out CS-ROSETTA calculations for the catalytic loop (residues 139–153) on an otherwise fixed crystal structure of the core domain. The backbone conformations of the six low-energy CS-ROSETTA clusters suggest that the kinked 3_{10} helical conformation observed in PFV is also sampled in solution along with other conformations (Figs. 5 and 6). The lowest energy model (Figs. 5 and 6, *red conformation*) is in striking agreement with the PFV IN structure for residues 145–153. In addition to being kinked at residues 149, this model adopts a 3_{10} helical conformation for residues 145–147 as defined by the DSSP software package (103). In addition to the DNA-bound conformation it is clear that other conformations can be sampled by the catalytic loop as well. Chemical shifts and sequential H^N-H^N NOE intensities (supplemental Figs. S7 and S8) suggest extensive sampling of helical conformations from Pro¹⁴⁵ onwards, but we also find two clusters that trace an extended conformation for

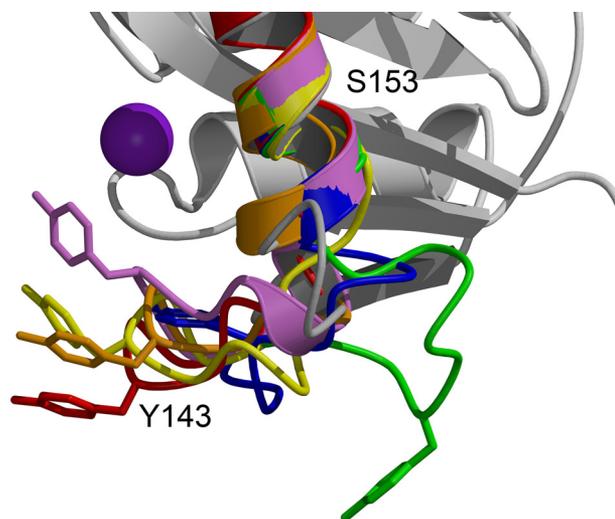


FIGURE 6. A close-up of CS-ROSETTA optimized conformations of residues 139–153. The top 6 simulated conformations (color) are overlaid onto the crystal structure (PDB entry 1BL3, chain C; *gray*). Each conformation represents a low-energy structure compatible with chemical shifts in the catalytic loop region. Colors correspond to torsion angles shown in Fig. 5. C α atoms for residues Tyr¹⁴³ and Ser¹⁵³ are labeled, and the crystallographic Mg²⁺ ion is displayed as a *purple sphere*. The side chain heavy atoms for Tyr¹⁴³ are shown for each conformation.

these residues (Figs. 5 and 6, *green* and *yellow* conformations), similar to what is observed in crystal structures complexed to Mg²⁺ (PDB entry 1BL3; Fig. 6, *gray* conformation). Although the other clusters possess near α -helical φ , ψ torsions, the hydrogen bonding patterns suggest that helix α 4 undergoes fraying as well (Fig. 6).

Two remaining residues are also of interest, Tyr¹⁴³ and Asn¹⁴⁴. The chemical shifts for Asn¹⁴⁴ are most compatible with the northwest quadrant of the Ramachandran plot, but the residues have a broad distribution roughly centered at the φ , ψ values for polyproline II helix. The backbone torsions for Tyr¹⁴³ are even less well converged, and only one of our clusters is in agreement with the left-handed helical conformation observed

Conformation and Dynamics of HIV Integrase Core Domain

in the PFV IN crystal structure (Fig. 5, *purple* conformation). The TALOS+ fragments are split approximately evenly between right-handed helix and strand conformations, and correspondingly most of our clusters sample the left-hand side of the Ramachandran plot for Tyr¹⁴³. Even though this residue (Tyr²¹²) differs in PFV, the overall topology of the catalytic loop is conserved for our models.

In light of prior simulations on the core domain, our data reconcile two conflicting models for the catalytic loop conformation. The structures of Greenwald *et al.* (29) and the simulations of Lins *et al.* (38) both suggest a helical conformation for the catalytic loop, in general agreement with the PFV integrase-DNA crystal structure. Alternatively, longer simulations by Lee *et al.* (39) suggest that Tyr¹⁴³ shifts conformations as helix $\alpha 4$ partially unfolds, effectively gating the active site. Our CS-ROSETTA calculations, which are based on experimental chemical shifts, and our H^N-H^N NOE data (supplemental Fig. S7) support both models. Most of the time, residues 139–153 sample helical basins, but large scale, occasional structural excursions are also possible. Thus, even without the Gly to Ala mutations made by Greenwald and co-workers (29), the residues N-terminal of helix $\alpha 4$ appear to sample helix in solution. We hypothesize that their mutations lock in the helical conformation and prevent the excursions into other conformations seen in the wild type protein. These excursions may be required to sample the appropriate 3_{10} helix observed in PFV integrase, or they may be needed to accommodate the structural strain experienced during DNA binding.

Although structural modeling based on chemical shifts alone is tentative, our ¹⁵N relaxation data support the above conclusions. With order parameters in the 0.7–0.8 range and heteronuclear NOE values ranging from 0.4 to 0.7, the residues in the catalytic loop are not nearly as dynamic as would be expected for a random coil (*cf.* the N- and C termini in Fig. 3). Instead, dynamics indicate that the region is fairly stable on the picosecond-nanosecond time scale. Except for the N and C termini, and in agreement with temperature factors and simulations, the most dynamic region of the molecule at this time scale is the loop from residues 185 to 195 (38, 39). As a point of reference, a Lipari-Szabo order parameter, S^2 , of 0.7 corresponds to angular diffusion within a cone subtending 28 degrees (77). Thus, the dynamics of the active site loop fall somewhere between the extremes of disordered and fully folded. Although our chemical shift-based simulations support conformational excursions of the catalytic loop (39), we do not find evidence for gating on the picosecond-nanosecond time scale. Given the CS-ROSETTA models, it is likely that these motions occur but that they happen on longer time scales. In fact, other studies have found helix formation to occur on a time scale of microseconds (104), and loop gating can even occur on time scales up to milliseconds in some cases (105). The conformational change observed in molecular dynamics simulations may therefore be slower than we could reasonably expect to detect with ¹⁵N relaxation measurements, which then report on dynamics in each of the stable clusters shown in Figs. 5 and 6 individually.

Now that all three domains of IN have been assigned by NMR, it is tempting to design experiments testing structure-activity relationships with inhibitors and DNA substrates. Such

experiments would in principle be faster and more straightforward than crystallographic studies. Unfortunately, our attempts to characterize full-length IN by NMR so far have remained unsuccessful (data not shown), and our results with raltegravir binding demonstrate that studies of IN by NMR will continue to be challenging. Diketo acid inhibitors of IN have been shown to require the presence of DNA substrate (96), and the extremely weak binding we observe for the free core domain confirms this. Moreover, we have shown that the solvents used for IN inhibitors, such as DMSO, affect the chemical shifts of IN itself at concentrations below 5% (v/v). Therefore, structure-activity studies will have to account for these effects. An additional complication is the sheer size of the protein-DNA complex, which for PFV approaches 270 kDa (37). Although *de novo* studies of such large systems tends to be extremely difficult, novel TROSY-based technology combined with extensive deuteration can make such studies feasible provided that assignments of its domains are available.

Acknowledgments—We thank Dr. Yves Pommier for providing a sample of raltegravir, Dr. John M. Louis for assistance with multiangle light scattering, and Drs. Robert Craigie, Christina Bodner, Jinfa Ying, and Lisa Parsons for helpful discussions. We also thank Dr. Peter Cherepanov for sharing coordinates of the PFV IN structure after online publication, before release from the PDB. Use of the NIH Biowulf cluster is gratefully acknowledged.

REFERENCES

- Bushman, F. D., Fujiwara, T., and Craigie, R. (1990) *Science* **249**, 1555–1558
- LaFemina, R. L., Schneider, C. L., Robbins, H. L., Callahan, P. L., LeGrow, K., Roth, E., Schleif, W. A., and Emini, E. A. (1992) *J. Virol.* **66**, 7414–7419
- Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3428–3432
- Turlure, F., Devroe, E., Silver, P. A., and Engelman, A. (2004) *Front. Biosci.* **9**, 3187–3208
- Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2525–2529
- Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol.* **63**, 5319–5327
- Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell* **62**, 829–837
- Sherman, P. A., and Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5119–5123
- Bowerman, B., Brown, P. O., Bishop, J. M., and Varmus, H. E. (1989) *Genes Dev.* **3**, 469–478
- Farnet, C. M., and Haseltine, W. A. (1991) *J. Virol.* **65**, 1910–1915
- Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) *Cell* **63**, 87–95
- Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1339–1343
- Yoder, K. E., and Bushman, F. D. (2000) *J. Virol.* **74**, 11191–11200
- Craigie, R. (2001) *J. Biol. Chem.* **276**, 23213–23216
- Daniel, R., Greger, J. G., Katz, R. A., Taganov, K. D., Wu, X., Kappes, J. C., and Skalka, A. M. (2004) *J. Virol.* **78**, 8573–8581
- Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) *Science* **255**, 723–726
- Grinsztejn, B., Nguyen, B. Y., Katlama, C., Gatell, J. M., Lazzarin, A., Vittecoq, D., Gonzalez, C. J., Chen, J., Harvey, C. M., and Isaacs, R. D. (2007) *Lancet* **369**, 1261–1269
- Pommier, Y., Johnson, A. A., and Marchand, C. (2005) *Nat. Rev. Drug Discov.* **4**, 236–248
- Semenova, E. A., Marchand, C., and Pommier, Y. (2008) *Adv. Pharmacol.* **56**, 199–228

Conformation and Dynamics of HIV Integrase Core Domain

20. Andrade, M. D., and Skalka, A. M. (1996) *J. Biol. Chem.* **271**, 19633–19636
21. Esposito, D., and Craigie, R. (1999) *Adv. Virus Res.* **52**, 319–333
22. Chiu, T. K., and Davies, D. R. (2004) *Curr. Top. Med. Chem.* **4**, 965–977
23. Jenkins, T. M., Engelman, A., Ghirlando, R., and Craigie, R. (1996) *J. Biol. Chem.* **271**, 7712–7718
24. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266**, 1981–1986
25. Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338
26. Bujacz, G., Alexandratos, J., Qing, Z. L., Clément-Mella, C., and Wlodawer, A. (1996) *FEBS Lett.* **398**, 175–178
27. 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
28. Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clément-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* **282**, 359–368
29. Greenwald, J., Le, V., Butler, S. L., Bushman, F. D., and Choe, S. (1999) *Biochemistry* **38**, 8892–8898
30. Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., and Davies, D. R. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13040–13043
31. Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W., Bushman, F. D., Siegel, J. S., and Choe, S. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57**, 536–544
32. Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20**, 7333–7343
33. Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8233–8238
34. Asante-Appiah, E., and Skalka, A. M. (1997) *J. Biol. Chem.* **272**, 16196–16205
35. Asante-Appiah, E., Seeholzer, S. H., and Skalka, A. M. (1998) *J. Biol. Chem.* **273**, 35078–35087
36. Yi, J., Asante-Appiah, E., and Skalka, A. M. (1999) *Biochemistry* **38**, 8458–8468
37. Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010) *Nature* **464**, 232–236
38. Lins, R. D., Briggs, J. M., Straatsma, T. P., Carlson, H. A., Greenwald, J., Choe, S., and McCammon, J. A. (1999) *Biophys. J.* **76**, 2999–3011
39. Lee, M. C., Deng, J., Briggs, J. M., and Duan, Y. (2005) *Biophys. J.* **88**, 3133–3146
40. Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. (1992) *Biochemistry* **31**, 9150–9157
41. Tang, C., Ndassa, Y., and Summers, M. F. (2002) *Nat. Struct. Biol.* **9**, 537–543
42. Gatanaga, H., Suzuki, Y., Tsang, H., Yoshimura, K., Kavlick, M. F., Nagashima, K., Gorelick, R. J., Mardy, S., Tang, C., Summers, M. F., and Mitsuya, H. (2002) *J. Biol. Chem.* **277**, 5952–5961
43. Katoh, E., Louis, J. M., Yamazaki, T., Gronenborn, A. M., Torchia, D. A., and Ishima, R. (2003) *Protein Sci.* **12**, 1376–1385
44. Newman, J. L., Butcher, E. W., Patel, D. T., Mikhaylenko, Y., and Summers, M. F. (2004) *Protein Sci.* **13**, 2101–2107
45. Saad, J. S., Miller, J., Tai, J., Kim, A., Ghanam, R. H., and Summers, M. F. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11364–11369
46. 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
47. Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hård, K. (1995) *Nat. Struct. Biol.* **2**, 807–810
48. Eijkelenboom, A. P., van den Ent, F. M., Vos, A., Doreleijers, J. F., Hård, K., Tullius, T. D., Plasterk, R. H., Kaptein, R., and Boelens, R. (1997) *Curr. Biol.* **7**, 739–746
49. Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1997) *Nat. Struct. Biol.* **4**, 567–577
50. Eijkelenboom, A. P., Sprangers, R., Hård, K., Puras Lutzke, R. A., Plasterk, R. H., Boelens, R., and Kaptein, R. (1999) *Proteins* **36**, 556–564
51. Eijkelenboom, A. P., van den Ent, F. M., Wechselberger, R., Plasterk, R. H., Kaptein, R., and Boelens, R. (2000) *J. Biomol. NMR* **18**, 119–128
52. Hickman, A. B., Palmer, I., Engelman, A., Craigie, R., and Wingfield, P. (1994) *J. Biol. Chem.* **269**, 29279–29287
53. Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., and Craigie, R. (1998) *J. Biomol. NMR* **11**, 97–102
54. Ikura, M., Krinks, M., Torchia, D. A., and Bax, A. (1990) *FEBS Lett.* **266**, 155–158
55. Torchia, D. A., Sparks, S. W., and Bax, A. (1989) *Biochemistry* **28**, 5509–5524
56. Tong, K. I., Yamamoto, M., and Tanaka, T. (2008) *J. Biomol. NMR* **42**, 59–67
57. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wüthrich, K. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13585–13590
58. Pervushin, K. V., Wider, G., and Wüthrich, K. (1998) *J. Biomol. NMR* **12**, 345–348
59. Ying, J., Chill, J. H., Louis, J. M., and Bax, A. (2007) *J. Biomol. NMR* **37**, 195–204
60. Geen, H., and Freeman, R. (1991) *J. Magn. Reson.* **93**, 93–141
61. Schanda, P., and Brutscher, B. (2005) *J. Am. Chem. Soc.* **127**, 8014–8015
62. Chill, J. H., Louis, J. M., Baber, J. L., and Bax, A. (2006) *J. Biomol. NMR* **36**, 123–136
63. Jones, J. A., Hodgkinson, P., Barker, A. L., and Hore, P. J. (1996) *J. Magn. Reson. B* **113**, 25–34
64. Davis, D. G., Perlman, M. E., and London, R. E. (1994) *J. Magn. Reson. B* **104**, 266–275
65. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277–293
66. Masse, J. E., and Keller, R. (2005) *J. Magn. Reson.* **174**, 133–151
67. Keller, R. L. J. (2004) *Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer-aided Resonance Assignment*. Ph.D. thesis, pp. 1–149, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland
68. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperki, T., Montelione, G. T., Baker, D., and Bax, A. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4685–4690
69. Das, R., André, I., Shen, Y., Wu, Y., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperki, T., and Baker, D. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18978–18983
70. Kontaxis, G., Delaglio, F., and Bax, A. (2005) *Methods Enzymol.* **394**, 42–78
71. Berjanskii, M. V., and Wishart, D. S. (2005) *J. Am. Chem. Soc.* **127**, 14970–14971
72. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) *J. Biomol. NMR* **44**, 213–223
73. Mandel, A. M., Akke, M., and Palmer, A. G., 3rd (1995) *J. Mol. Biol.* **246**, 144–163
74. Cole, R., and Loria, J. P. (2003) *J. Biomol. NMR* **26**, 203–213
75. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc.* **112**, 4989–4991
76. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4546–4559
77. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4559–4570
78. Grzesiek, S., Bax, A., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., Tjandra, N., and Wingfield, P. T. (1997) *Protein Sci.* **6**, 1248–1263
79. R Development Core Team (2009) *R: a Language and Environment for Statistical Computing*, Vienna, Austria
80. Battiste, J. L., and Wagner, G. (2000) *Biochemistry* **39**, 5355–5365
81. Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004) *J. Am. Chem. Soc.* **126**, 5879–5896
82. Clore, G. M., and Iwahara, J. (2009) *Chem. Rev.* **109**, 4108–4139
83. Clore, G. M., and Gronenborn, A. M. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 29–63
84. Wagner, G. (1990) *Prog. NMR Spect.* **22**, 101–139
85. Spera, S., and Bax, A. (1991) *J. Am. Chem. Soc.* **113**, 5490–5492
86. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) *J. Mol. Biol.* **222**, 311–333
87. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, New York

Conformation and Dynamics of HIV Integrase Core Domain

88. Davies, J. F., 2nd, Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. A. (1991) *Science* **252**, 88–95
89. Rice, P., and Mizuuchi, K. (1995) *Cell* **82**, 209–220
90. Baldanti, F., Paolucci, S., Gulminetti, R., Brandolini, M., Barbarini, G., and Maserati, R. (2010) *J. Med. Virol.* **82**, 116–122
91. Shimura, K., Kodama, E., Sakagami, Y., Matsuzaki, Y., Watanabe, W., Yamataka, K., Watanabe, Y., Ohata, Y., Doi, S., Sato, M., Kano, M., Ikeda, S., and Matsuoka, M. (2008) *J. Virol.* **82**, 764–774
92. Goethals, O., Clayton, R., Van Ginderen, M., Vereycken, I., Wagemans, E., Geluykens, P., Dockx, K., Srijbos, R., Smits, V., Vos, A., Meersseman, G., Jochmans, D., Vermeire, K., Schols, D., Hallenberger, S., and Hertogs, K. (2008) *J. Virol.* **82**, 10366–10374
93. Engelman, A., and Craigie, R. (1992) *J. Virol.* **66**, 6361–6369
94. Engelman, A., and Craigie, R. (1995) *J. Virol.* **69**, 5908–5911
95. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) *J. Mol. Biol.* **285**, 1735–1747
96. Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J. Y., Young, S., Hamill, T., Cole, J. L., and Hazuda, D. J. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11244–11249
97. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000) *Science* **287**, 646–650
98. Steitz, T. A., and Steitz, J. A. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502
99. Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrade, M., Katz, R. A., and Skalka, A. M. (1997) *J. Biol. Chem.* **272**, 18161–18168
100. Jaskolski, M., Alexandratos, J. N., Bujacz, G., and Wlodawer, A. (2009) *FEBS J.* **276**, 2926–2946
101. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol.* **239**, 363–392
102. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR.* **13**, 289–302
103. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
104. Eaton, W. A., Munoz, V., Thompson, P. A., Henry, E. R., and Hofrichter, J. (1998) *Acc. Chem. Res.* **31**, 745–753
105. Wang, Y., Berlow, R. B., and Loria, J. P. (2009) *Biochemistry* **48**, 4548–4556
106. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950
107. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524
108. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* **28**, 8972–8979