

Specificity of LTR DNA recognition by a peptide mimicking the HIV-1 integrase $\alpha 4$ helix

Zeina Hobaika¹, Loussine Zargarian¹, Yves Boulard², Richard G. Maroun³, Olivier Mauffret¹ and Serge Femandjian^{1,*}

¹Laboratoire de Biotechnologies et Pharmacologie génétique Appliquée (LBPA), UMR 8113 CNRS, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, ²CEA SACLAY, (DSV/iBiTec-S/SBIGeM/LBI), Bat.144, P.C. 22, 91191 Gif-sur-Yvette Cedex, France and ³Département des Sciences de la Vie et de la Terre, Faculté des Sciences, Université Saint-Joseph, CST-Mar Roukos, Beirut, Lebanon

Received March 31, 2009; Revised September 16, 2009; Accepted September 18, 2009

ABSTRACT

HIV-1 integrase integrates retroviral DNA through 3'-processing and strand transfer reactions in the presence of a divalent cation (Mg^{2+} or Mn^{2+}). The $\alpha 4$ helix exposed at the catalytic core surface is essential to the specific recognition of viral DNA. To define group determinants of recognition, we used a model composed of a peptide analogue of the $\alpha 4$ helix, oligonucleotides mimicking processed and unprocessed U5 LTR end and 5 mM Mg^{2+} . Circular dichroism, fluorescence and NMR experiments confirmed the implication of the $\alpha 4$ helix polar/charged face in specific and non-specific bindings to LTR ends. The specific binding requires unprocessed LTR ends—i.e. an unaltered 3'-processing site $CA\downarrow GT3'$ —and is reinforced by Mg^{2+} (K_d decreases from 2 to 0.8 nM). The latter likely interacts with the ApG and GpT3' steps of the 3'-processing site. With deletion of GT3', only persists non-specific binding (K_d of 100 μM). Proton chemical shift deviations showed that specific binding need conserved amino acids in the $\alpha 4$ helix and conserved nucleotide bases and backbone groups at LTR ends. We suggest a conserved recognition mechanism based on both direct and indirect readout and which is subject to evolutionary pressure.

INTRODUCTION

HIV-1 cDNA integration into the host cell chromosome is catalyzed by the virus enzyme integrase (IN) (1,2). The reaction involves two separate steps: 3' processing of the newly synthesized cDNA in the cytoplasm, and

strand transfer in the nucleus (3). Processed cDNA and IN are imported into the nucleus via a preintegration complex (PIC) including viral and host proteins (4).

The 3'-processing and the strand transfer reactions have been modeled *in vitro* using purified recombinant IN and a double stranded DNA fragment (21 base pairs) mimicking either the U5 or the U3 LTR end. IN alone, in the presence of divalent cations, performs both the 3' processing and the DNA strand transfer reactions (2). The DNA fragment plays the role of both the donor (virus DNA) and acceptor (cell DNA).

The IN monomer is a 32-kDa protein comprising three structural domains (5,6). The N-terminal domain (Nt: residues 1–49) has several helices and adopts a compact structure fixed by zinc atom coordination (7). The central domain of the catalytic core (CC: residues 50–212) bearing the active site acidic residues, Asp64, Asp116 and Glu152 (the so-called DDE catalytic triad) (8–12), belongs to a sup family of DNA/RNA strand transferases/nucleases (13–15). The C-terminal domain (Ct: residues 213–288) incorporates a Src-like domain and is involved in DNA host recognition (16).

The high resolution 3D structure of the entire enzyme bound or unbound to DNA has not yet been resolved. The main handicap to obtaining crystals for X-ray studies or in performing an NMR analysis of IN is the weak solubility of the protein. However, a low resolution structure of IN at the DNA contact level has been derived by electronic microscopy (17). This reveals an asymmetric tetrameric IN assembly, contrasting the symmetric structure provided by theoretical or semi-empirical models (18–20).

IN uses a divalent cation (either Mn^{2+} or Mg^{2+}) as co-factor (21), similar to several enzymes that perform nucleic acid phosphoryl-transfer reactions. Mg^{2+} may be the relevant factor for IN function *in vivo* as its intracellular concentration is much higher than that of Mn^{2+} (1 mM vs 10⁻⁴ mM). Moreover, Mn^{2+} augments

*To whom correspondence should be addressed. Tel: +33 1 47 40 77 42; Fax: +33 1 47 40 76 71; Email: serge.femandjian@lbpa.ens-cachan.fr

both the non-specific nuclease activity of IN (22) and the acceptance of sequence variations at the LTR extremities (23), and several mutations affecting Mn^{2+} are ineffective, which is not the case with Mg^{2+} (23,24). This difference in the selection of Mn^{2+} and Mg^{2+} also affects the efficiency of IN inhibitors and has been taken into account in the design of raltegravir and elvitegravir drugs (25). How does the divalent cation in IN function? Many uncertainties still exist. The cation may introduce conformational changes to the catalytic site, thus conferring an active structure, but it could also serve as an intermediate permitting the binding of IN to the DNA substrate (26,27).

Previously, to study binding of IN to DNA we have used a model approach involving an analogue (K156) of the amphiphile $\alpha 4$ helix lying at the surface of the IN CC (Figure 1A), and an oligonucleotide corresponding to the U5 LTR end (28). Results have highlighted the roles of Lys156 and Lys159 of the $\alpha 4$ helix and the need for an unprocessed LTR DNA end to achieve specific interaction (28). In a following paper we have shown that the $\alpha 4$ helix is the DNA recognition helix of the HTH (helix turn helix) motif (29). Here we aim to provide greater details on the interaction of the $\alpha 4$ helix with LTR ends in the presence of Mg^{2+} . Our approach involved circular dichroism (CD), fluorescence (quenching and anisotropy) and 1H -NMR spectroscopy. Results were consistent with IN recognizing viral DNA via both direct and indirect readout, in which the binding is optimal only when LTR ends are unprocessed and divalent cations are present.

MATERIALS AND METHODS

The peptides and oligonucleotides used in this study are shown in Figure 1. Some of their characteristics are also presented.

Peptides

Two versions of the peptide K156 (K156-Y and K156-W) (Figure 1B) were synthesized as previously reported (28). K156 is a helix-stabilized version of the helical $\alpha 4$ peptide (residues 147 to residues 169 in CC) (28). Briefly, several residues were replaced in parts of the helix deemed not important for DNA recognition by residues promoting helix formation. The K156 peptide backbone conformation had greater resemblance to the $\alpha 4$ helix in the protein context than the $\alpha 4$ peptide taken in isolation. It was also less aggregation prone and more adapted to the study of specific interactions, that are highly conformation dependent. The Tyr (Y) or Trp (W) aromatic residues added to the C-terminus enabled peptide concentration estimation from absorbances in UV spectra, using molar absorption coefficients of 1280 and 5600 $M^{-1}cm^{-1}$ at 280 nm for the tyrosine-containing peptide and the tryptophan-containing peptide, respectively.

The wheel presentation of the $\alpha 4$ helix and K156 peptide provided an illustration of the hydrophobic and hydrophilic domains and showed the positions of the substitutions made in the helix (Figure 1C).

Oligonucleotides

The two oligonucleotides were purchased from Eurogentec (Belgium) (Figure 1D). The choice of monomolecular hairpin-forming oligonucleotides rather than bimolecular duplex-forming oligonucleotides was motivated by the need for double helix stability under the low concentrations used in fluorescence and CD experiments (10^{-9} to $10^{-5}M$). The oligonucleotide sequences reproduce an U5 LTR end that is unprocessed and processed (deletion of GT3' on the upper strand). The central thymine of the three thymine-loop bears a fluorescein reporter allowing fluorescence measurements. The fluorescein is thought not to interfere with the binding of IN to LTR ends. Nonetheless, a version without fluorescein was prepared for CD and NMR studies.

Fluorescence measurements

The intrinsic fluorescence quantum yield and fluorescence anisotropy studies were carried out with a Jobin-Yvon Fluoromax II instrument (HORIBA Jobin Yvon, France) equipped with an Ozone-free 150 W xenon lamp. Samples (800 μl) were placed at 5°C in thermally jacketed 1 cm \times 0.5 cm quartz cells. At least ten measurements for each titration point were recorded with an integration time of 1 s. Fluorophores were either tryptophan or fluorescein purposely fixed to the peptide or the oligonucleotide, respectively.

In fluorescence anisotropy, ($A = (I_{\parallel} - GI_{\perp}) / (I_{\parallel} + 2GI_{\perp})$), parallel (I_{\parallel}) and perpendicular (I_{\perp}) emission components were measured in L-format with 4 nm excitation and emission slit widths. With fluorescein as the fluorophore, the excitation was performed at 488 nm and the emission was recorded at 516 nm in the case of LTR34 and at 515 nm in the case of LTR32. Fluorescein-labeled oligonucleotides were diluted to the desired concentration (10 nM) in 800 μl of assay buffer (Na/Na₂ phosphate, pH 6, $I = 0.1$ in the presence or absence of 5 mM $MgCl_2$). Peptides were stepwise diluted. For each anisotropy measurement, the parallel (I_{\parallel}) and the perpendicular (I_{\perp}) intensities of the background solution (i.e. buffer and protein contributions) were automatically subtracted from the sample value, calculating the G-value correction each time.

In quenching experiments, intrinsic fluorescence of K156-W was measured at a concentration of 400 nM in 800 μl of reaction buffer (with or without 5 mM $MgCl_2$). Excitation at 290 nm provided an emission between 300 and 480 nm, using 2 and 5 nm excitation and emission slit widths, respectively. Maximal emission was measured at 355 nm. Titration isotherms illustrating the binding of Mg^{2+} to K156 were expressed as $1 - F/F^0$, where F^0 is the fluorescence in the absence of Mg^{2+} .

CD spectroscopy

CD spectra were recorded on a CD6 dichrograph (HORIBA Jobin Yvon, France). Measurements were calibrated with (+)-10-camphorsulfonic acid. Oligonucleotide and peptide concentrations varied from 6 to 12 μM in phosphate buffer pH 6, $I = 0.1$, with and

without Mg^{2+} . Samples were placed in jacketed cells with a 1 mm path length, minimizing thermal drift. To allow the solutions to reach their equilibrium state, these were incubated for 10 min at the chosen temperature. Spectra, recorded in 1 nm steps, were averaged over ten scans and corrected for the base line. They were presented as differential molar absorptivity per residue, $\Delta\epsilon$ ($M^{-1}cm^{-1}$), as a function of wavelength, between 260 and 185 nm for peptides and between 200 and 330 nm for DNA, peptide–DNA complexes or peptide alone. In the latter case, aliquots of peptide were added to LTR34 or LTR32 (12 μM) and the control spectrum of the LTR without ligand was subtracted from that of the complex. Effects of β -hydroxycarbonyl compounds were estimated by adding aliquots of the drug in the presence and absence of Mg^{2+} . The α -helical content of peptides was obtained using the relation: $P_{\alpha} = -[\Delta\epsilon_{222} \times 10]$ (P_{α} : percentage of α -helix; $\Delta\epsilon_{222}$: CD per residue at 222 nm) (30).

NMR spectroscopy

Proton NMR spectra were recorded on Bruker Avance 500 and 700 MHz (equipped with a TCI cryo-probe) spectrometers and were processed with the SPARKY program. Samples (500 μl) were prepared with 1 mM K156 and LTR34 diluted in 90% phosphate buffer (10 mM Na/Na₂ pH 6.5)–10% ²H₂O. TSP was used as an internal chemical shift reference. Standard two-dimensional NMR experiments were recorded at 10°C (31). Spectral widths were enlarged to 12 (for the peptide) and 20 p.p.m. (for the DNA and the complexes). The NOESY mixing times were fixed at 150, 200, 300 and 450 ms. The Clean-TOCSY (32) sequences were collected with MLEV-17 spin-locking fixed at 40, 60, 80, 100 and 120 ms. Solvent suppression was achieved by applying a WATERGATE pulse sequence. NOE relative intensities were measured from 300 ms mixing time NOESY spectra with solvent presaturation. Intensities (volume and height of connectivity) were determined by the Sparky program (Goddard and Kneller, University of California, San Francisco) and were confirmed by the TopSpin program (Bruker). The ³J _{α HNH} coupling constants were measured on NH peaks in 1D NMR spectra. Chemical shifts were determined from both TOCSY and NOESY spectra.

RESULTS AND DISCUSSION

Design of the peptide K156

The conformation of the $\alpha 4$ peptide in solution is mainly unordered (28) and is therefore very different of the $\alpha 4$ helix structure observed in the context of the protein (Figure 1A). Thus, the $\alpha 4$ peptide cannot be used to give an account of the role of the $\alpha 4$ helix in the recognition of DNA the enzyme. A more suitable model is peptide K156, a structural analogue of the $\alpha 4$ peptide displaying a higher helical content and thereby a pre-organized structure for interaction (Figure 1B). Most amino acid substitutions were made in ensuring that hydrophobic and hydrophilic surfaces were not significantly altered (Figure 1C). We also took into account the previously reported data of mutagenesis (33) to ensure that no amino acid

important to 3'-processing was replaced. Substitutions were as follows: (i) Gly149–Ala, Gly163–Ala, Val151–Leu, Ile161–Leu, Ile162–Leu (alanine and leucine being strongly involved in helix formation, especially compared with side chain deprived glycine and β -branched valine/isoleucine, respectively); and (ii) Val150 Lys and Ser153 Glu, allowing the formation of two $i-i+3$ pairs with stabilizing electrostatic interactions between positively and negatively charged side chains (Lys150 Glu153 and Glu153 Lys156). Examination of results of the variation in HIV-1 group M IN indicates that among all the substitutions performed in K156, only one, that of Ser153 (Ser153 Glu), confers resistance to raltegravir and elvitegravir (34).

Binding of Mg^{2+} to LTR34 and LTR32

The *in vivo* activity of Mg^{2+} is similar to that of a solution with 0.5–1.0 mM Mg^{2+} and 0.15 M monovalent salt (35). Both monovalent and divalent cations interact with the DNA surface and neutralize the phosphate negative charges, decreasing the repulsive forces between negative or positive charges. Mg^{2+} is a compact ion with a small atomic radius, facilitating its coordination with oxygen atoms in DNA (36,37). It is also a net hydrogen bond donor through its rigid octahedral primary solution shell (38); hydrogen bonds occur with bases in particular sequences (39,40). In fact, Mg^{2+} is involved in sequence-specific binding to the major and minor grooves of DNA, as well as non-specific binding to backbone phosphates (41). Often, the structures, the dynamics and the ligand interactions are modified by the cation, but each effect depends on the groove where the binding occurs.

We used CD and fluorescence spectroscopy to examine the effect of Mg^{2+} on both unprocessed LTR34 DNA and processed LTR32 DNA. We also studied ¹H NMR data in the case of LTR34 DNA. In CD experiments addition of MgCl₂ (to a final concentration of 5 mM) to a solution of LTR34 or LTR32 in phosphate buffer pH 6, at 5°C, did not produce noticeable change in the intensity or in the shape of spectra, which continued to present B DNA characteristics with a positive signal at ~ 280 nm and a negative one at ~ 250 nm (data not shown). The NMR analysis of LTR34 DNA in the absence of Mg^{2+} has been previously reported (42). Despite the large number and poor dispersion of proton resonances inherent to nucleic acids and the rather long size of our oligonucleotide (a stem of 17 base pairs), almost all the ¹H-resonances were assigned. The addition of Mg^{2+} to LTR34 did not significantly modify the chemical shifts (Supplementary Table S1). LTR34 maintains the B-DNA form with or without Mg^{2+} , which is consistent with the CD results.

Although Mg^{2+} does not visibly modify the DNA conformation, the titration isotherms obtained by fluorescence anisotropy show that the divalent cation binds to DNA. A deep transition in the curve occurs in the millimolar range for both LTR34 and LTR32 (data not shown). Curve analysis considering a simple two-state reversible equilibrium between Mg^{2+} ions and

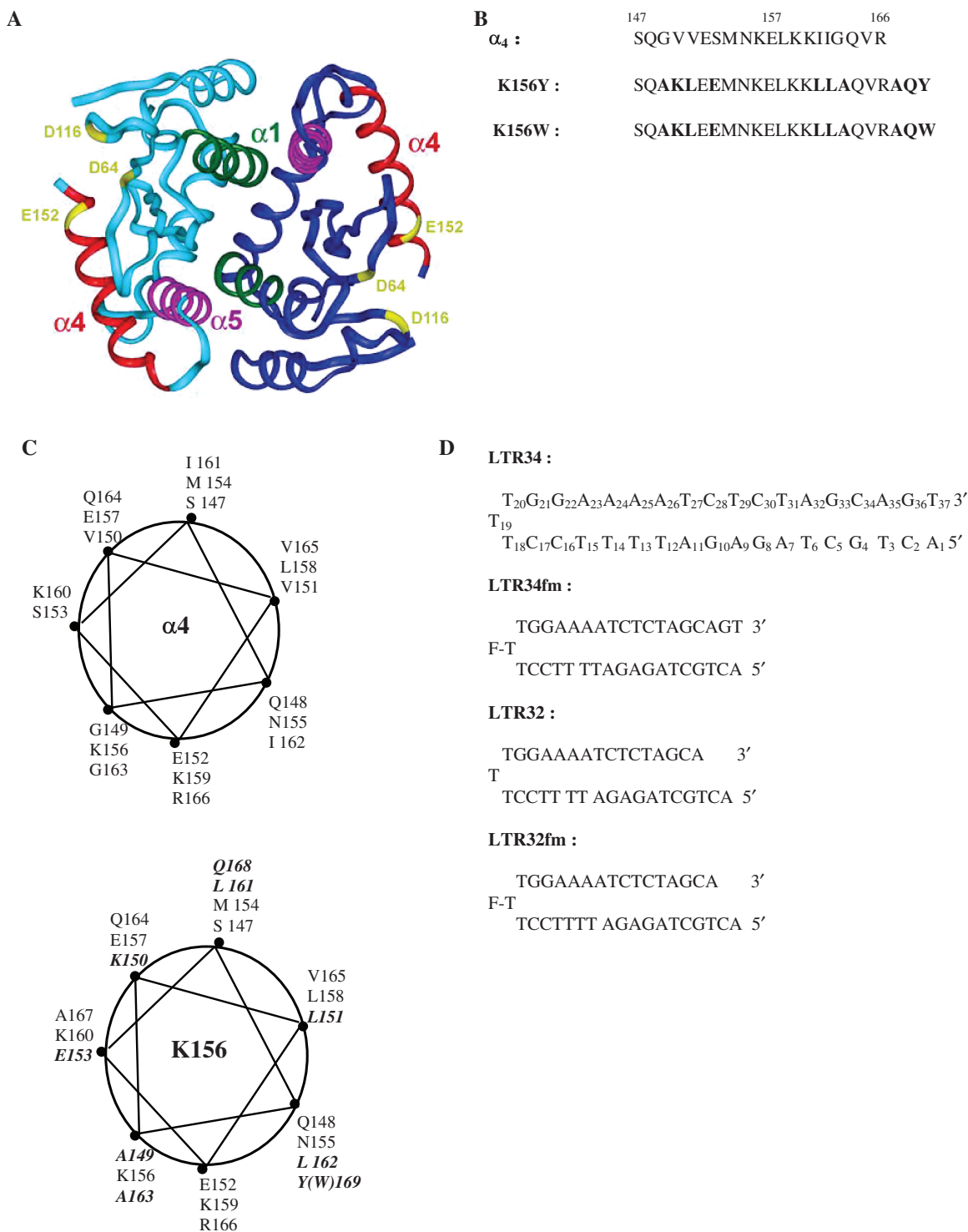


Figure 1. (A) Crystal structure of the dimeric catalytic core domain showing the α_4 helix at the protein surface [after Dyda *et al.* (10)]. (B) Peptides used in this study: sequences of α_4 and analogues K156Y and K156W with the substituted and added residues in bold. (C) Wheel representations of α_4 (top) and K156 (bottom) with the substituted and added residues in bold. (D) Oligonucleotides used in this study: unprocessed LTR34 under the hairpin form (three thymine loop and 17 base pair stem) numbered from 5' to 3', the processed oligonucleotide LTR32, and their fluoresceinated versions LTR34fm and LTR32fm.

DNA, which is an oversimplified approach (43) using 'GraphPad Prism' (Non-linear 'Least Squares'), provides K_d values of ~ 60 mM for LTR34 and ≥ 80 mM for LTR32.

Binding of Mg^{2+} to K156

The interaction of Mg^{2+} —which is assumed to act as a cofactor for the catalytic reaction and as a stabilizer of the IN–DNA complex—with the K156 peptide, was assessed

by CD, NMR and fluorescence spectroscopy. The CD spectra of K156 in the presence and absence of Mg^{2+} are given in Supplementary Figure S1. In the absence of Mg^{2+} , the CD spectra of K156 at 12 μM displayed two negative bands, at ~ 225 and ~ 208 nm, and a positive band, at ~ 190 nm, typical of the α helix (28). The α helix content based on CD intensity at ~ 225 nm (30) is of $\sim 25\%$. This partly depends on the composition and sequence of the peptide. There are several negatively charged Glu and positively charged Lys residues displaying $i + 3$ or $i + 4$ spacing within the K156 sequence. Their distribution allows either ion pair formation (Lys150–Glu153, Glu152–Lys156, Glu153–Lys156 and Glu157–Lys160) or ion pair repulsion (Glu153–Glu157, Lys156–Lys159 and Lys156–Lys160). The sum of the effects results either in a stabilization or destabilization of the helix, but salts modulate the intensity of effects.

We found that addition of $MgCl_2$ to K156 slightly increased band intensities, and thus helix stability. Our previous NMR analysis, which mainly investigated the K156 backbone, has shown that K156 adopts a rather stable helix structure in buffer and aqueous media at pH 6 (28). Here, we extended the analysis to amino acid side chains groups, as side chains are directly implicated in the DNA–peptide interactions. We measured K156 chemical shifts in the presence and absence of Mg^{2+} (Supplementary Table S2). The addition of Mg^{2+} to K156 produced some weak chemical shift variation and a selective broadening of correlations, visible in the TOCSY spectra. The β and γ correlations of the catalytic acid residue Glu152 resulted in a weak increase in intensity. In contrast, the αH , NH and γH protons of Gln168, the βH protons of Lys159, the γH proton of Lys160, and the βH and γH protons of Leu161 displayed a weak decrease in intensity. One of the δ proton signals of Lys159 was no longer visible. The only noticeable chemical shift variation involved the γH proton of Glu153, which led to a shift of -0.1 p.p.m.

Ten $^3J_{\alpha HNH}$ were accessible in the 1D spectra of K156 recorded in the presence and the absence of Mg^{2+} . The residues analyzed were Ala149, Lys150, Glu152, Met154, Asn155, Asp157, Lys159, Lys160, Ala163 and Ala167, which are evenly distributed along the entire chain length and, thereby may be used as good indicators for the whole backbone structure. The average of $^3J_{\alpha HNH}$ values in the absence of Mg^{2+} was 4.4 ± 0.3 Hz, a value consistent with that found for other helical structures (31). The addition of Mg^{2+} decreased the average value to 4.2 ± 0.3 Hz, suggesting a weak helix stabilization. The observed decrease in $^3J_{\alpha HNH}$ values was accompanied by a variation in several typical $NH_{(i)}-NH_{(i+1)}$ and $\alpha H_{(i)}-NH_{(i+1)}$ NOEs. For instance, $NH_{(i)}-NH_{(i+1)}$ correlations of the linked residues Ala149–Lys150, Glu153–Met154 and Asn155–Lys156, which show well-delineated cross-peaks, resulted in a noticeable increase of intensity (140, 70 and 40, respectively), consistent with helix stabilization.

We used fluorescence spectroscopy to determine the thermodynamic variables for Mg^{2+} and K156 binding. The signal of tryptophan (excitation at 290 nm, emission at 300–480 nm) purposely incorporated at the C-end of the peptide was used as the fluorophore. The change in

fluorescence induced by Mg^{2+} was weak, indicating that the conformational change was also weak, and was likely limited to only some Mg^{2+} -binding positions. Treatment of the titration curve yielded an apparent K_d value of 2.5 μM (data not shown).

CD analysis of the peptide binding to DNA

CD is a convenient method for analyzing both peptide and oligonucleotide conformations, and conformational changes accompanying complex formation. The CD spectra of LTR34 and LTR32 were typical of B DNA, and these spectra remained unchanged upon Mg^{2+} addition (data not shown). In contrast, there were slight changes in the K156 spectrum on Mg^{2+} addition, which is consistent with a more stable helix (Supplementary Figure S1). We investigated the binding of K156 to processed and unprocessed LTR ends in the presence of Mg^{2+} . Previous experiments performed in the absence of Mg^{2+} have shown that the GT3' dinucleotide deleted upon 3'-processing is essential for the specific binding between IN and virus DNA (28). Mixing LTR34 and K156 in the presence of Mg^{2+} resulted in a spectrum in the 190–260 nm region that clearly differed from the sum of individual K156 and LTR34 spectra. Since no changes were detected between 260 and 300 nm, an UV region rather specific to DNA, we deduced that the changes observed in the 190–260 nm UV region were due to conformational variations affecting the sole peptide. The difference spectra (i.e. spectra of DNA–K156 complexes at the 1:1 ratio minus the spectrum of unbound K156) showed that LTR34 stabilizes the K156 helix, but LTR32 does not, thereby confirming the contribution of the LTR GT3' dinucleotide to complex formation (Figure 2A and B).

Analysis of the peptide binding to DNA by fluorescence

Direct implication of the $\alpha 4$ helix in interactions with LTR ends has been suggested in several *in vitro* and *in vivo* experiments (23,28,44–57). Here, the DNA– $\alpha 4$ peptide binding analysis was performed by monitoring the anisotropy signal of fluorescein linked to the hairpin oligonucleotides (Figure 1D). The binding isotherms were related to the total average amount of K156 binding (Figure 3A and B). The LTR34 isotherm had two phases: a small increase of anisotropy accompanied by an inflection in the nanomolar range, followed by a steeper increase with an inflection in the micromolar range. Curve analysis provided a K_{d1} value of ~ 0.8 nM (first binding) and a K_{d2} value ≥ 100 μM (second binding). In contrast, the binding of K156 to LTR32 provided a monophasic isotherm with a K_d value ≥ 100 μM , which was similar to the LTR34 K_{d2} value. Comparison of these values with our previous K_d values obtained in the absence of Mg^{2+} (i.e. $K_{d1} \sim 2.1$ nM and $K_{d2} \sim 55$ μM for LTR34 and $K_d \sim 65$ μM for LTR32) (28) reveals that Mg^{2+} strengthens the high affinity binding between K156 and LTR34; however, it impairs low affinity binding in both LTR34 and LTR32.

Thus, high affinity binding may be dependent on directional hydrogen bonds, which are known for their resistance to changes in salt concentration, and possible

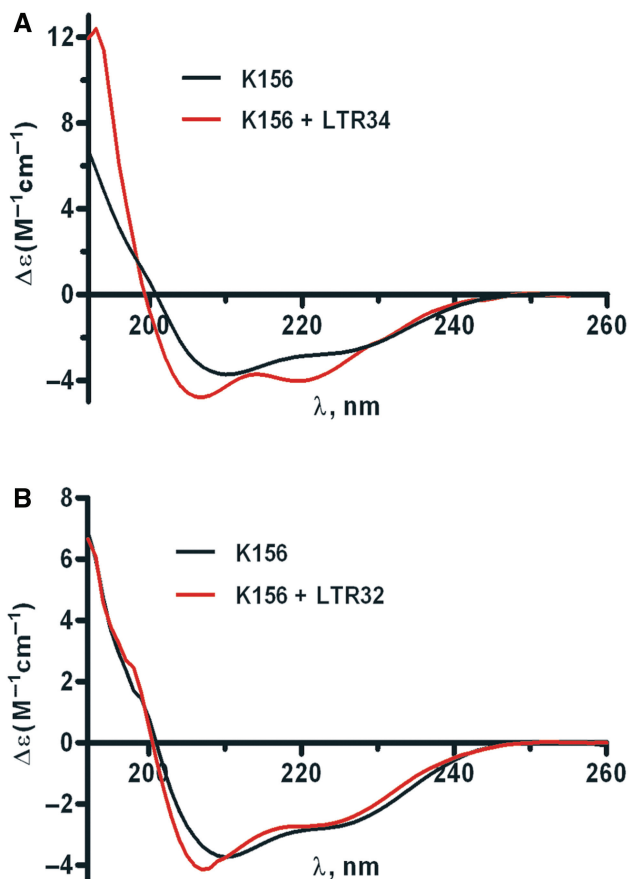


Figure 2. CD of unbound and bound K156 in the presence of Mg²⁺. (A) Spectrum of K156 at 12 μ M (Black line); difference spectrum [K156 + LTR34]—LTR34 at 12 μ M each (Red line). (B) Spectrum of K156 at 12 μ M (Black line); difference spectrum [K156 + LTR32]—LTR32 at 12 μ M each (Red line). Phosphate buffer pH 6, $I = 0.1$, at 5°C, MgCl₂ 5mM final concentration.

non-polar forces that are generally strengthened by salt addition (58). In contrast, low affinity binding that is impaired by salt addition is mediated by ionic interactions between negatively charged phosphate groups of the DNA backbone and positively charged side chain groups of the peptide.

Stabilization of high affinity binding in unprocessed LTR (LTR34) by Mg²⁺ suggests the binding of a divalent cation to the processing site CA↓GT3'. Mg²⁺ is known for its strong preference for the guanine base, although the adjacent base pair in the sequence usually cooperates in this interaction (38). Note that, in the crystal structure of the [d(CCAGTACTGG)]₂ duplex, the divalent ion binds the 5'CAGT3' sequence in the major groove, at either the ApG or GpT step (41,59). The solvated Mg²⁺ formed hydrogen bonds with guanine and adenine bases through its water molecules; the major groove was compressed at the ion binding site. This resulted in an opening of the minor groove, exactly where IN is supposed to perform its nucleophilic attack at LTR ends. Clearly, in a mechanism of two-ion catalysis, as the one generally assumed for IN, an Mg²⁺

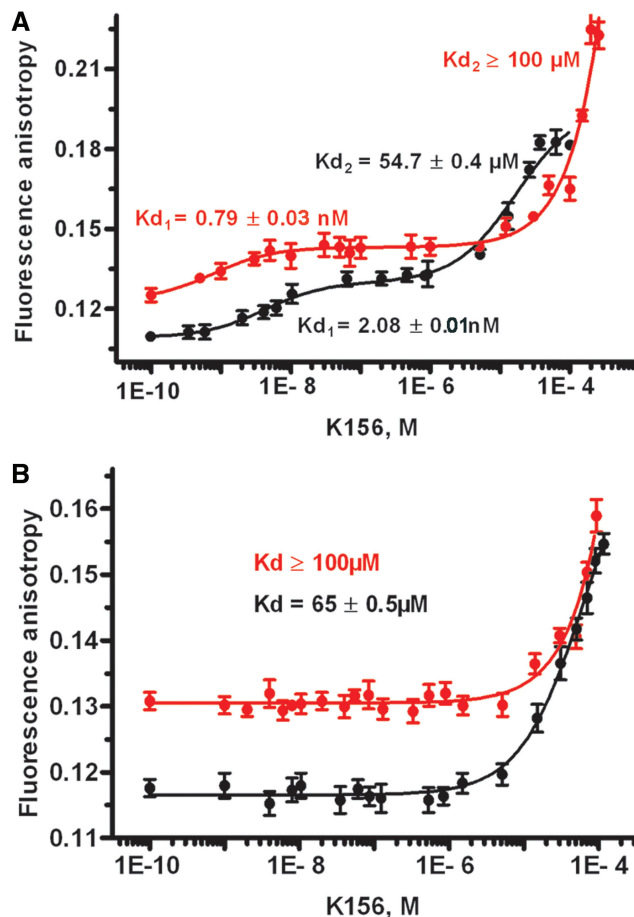


Figure 3. Fluorescence anisotropy titration of (A) LTR34 by peptide K156 in presence (Red line) and absence (Black line) of Mg²⁺; (B) LTR32 by peptide K156 in presence (Red line) and absence (Black line) of Mg²⁺. Phosphate buffer pH 6, $I = 0.1$, at 5°C, MgCl₂ 5mM final concentration.

ion positioned in the major groove may play the role of the second ion. Once detected and liganded by the carboxylate groups of catalytic residues this Mg²⁺ may assist the enzyme to cleave the appropriate phosphodiester bond. Note that in crystallographic studies the less compact Ca²⁺ also binds with high efficiency to the ApG and GpT steps, but its binding mode differs from that of Mg²⁺ (41). The same difference in binding may exist between Mg²⁺ and Mn²⁺, explaining the biological differences observed between these two ions in *in vitro* experiments.

Analysis of peptide binding to DNA by ¹H NMR

Proton chemical shift deviations (CSDs) may aid the identification of interacting regions in structures of complexed proteins and nucleic acids (60). In general, the CSDs of only one partner are used to certify the interaction (61). Here we investigated CSDs of both K156 and LTR34 to map the interface of the peptide–DNA complex in the presence of Mg²⁺. Fluorescence experiments indicated a significant difference between the high affinity binding K_d (~1 nM) and the low affinity binding K_d ($\geq 100 \mu$ M); these

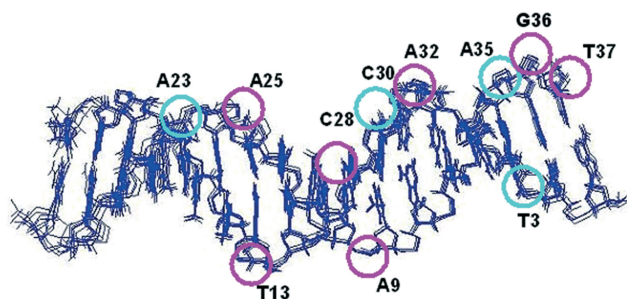


Figure 4. The LTR34 double helix affected by K156 interactions. The LTR structures presented are taken from Renisio *et al.*, 2005. Purple and cyan circles indicate the bases whose chemical shifts are most affected (CSDs: purple ≥ 0.1 ppm; 0.05 ppm \leq cyan < 0.1 ppm) by K156 (DNA:peptide ratio of 1:5).

conditions allowed an NMR analysis of binding in which the specific complex occurs more frequently than the non-specific complex. Yet, at the millimolar concentrations required by NMR, an equimolecular 1:1 complex was not attainable without material aggregation. To avoid a too much signal broadening and achieve a convenient mapping of the K156 and LTR interface, spectra were recorded at DNA-peptide ratios of 1:5 and 1:2.

α H-NH regions of the TOCSY spectra of K156 alone and K156 with DNA (DNA:peptide ratio of 1:5) are given in the Supplementary Figure S2. The specific K156 binding site on LTR34 was fully saturated at this ratio, although some non-specific binding was likely to occur. The CSDs for the base and ribose protons of LTR34 bound to K156 at a DNA: peptide ratio of 1:5, relative to unbound DNA (in the presence of Mg^{2+}), are given in Supplementary Table S3. Assuming that only CSDs values ≥ 0.05 p.p.m. indicate interaction we can infer that eleven nucleotide bases interact with the K156 amino acid residues (Figure 4). These are asymmetrically distributed between the two strands, with eight of them on the upper strand (the one that undergoes processing) and only three on the lower strand. Four of the affected bases belong to the same base pairs, A35.T3 and A25.T13, whereas two other bases, C28 and A9, are in close spatial proximity in the double helix. The A35 nucleotide, constitutive of the highly conserved CpA step, essential to the viral DNA integration, is among these nucleotides. Once again, the role of GT3' next to the conserved CpA step was highlighted: both the guanine and thymine base protons displayed significant CSDs. The large CSD of the guanine imino proton (0.23 p.p.m.) suggested a significant change in base arrangement; however, the corresponding signal in the NOESY spectrum showed no change, suggesting that stability of the G36.C2 base pair was not impaired by the interaction of LTR34 with K156.

The high number of desoxyribose sugars affected by the binding of K156 was also remarkable (Supplementary Table S3). Binding contacts extend inward, about 15 or 16 base pairs from the right end of the LTR (Figure 4). Remarkably, the binding pattern was similar to the 3'-processing pattern observed by Esposito and Craigie with 21 base pair LTR DNAs and the whole enzyme

(23). These authors found that in addition to the six commonly outermost nucleotides AGCACT3', the adenine (AAAA) tract in the distal LTR was also involved in viral DNA recognition by IN; our previous work (28) also suggested the implication of the adenine tract in binding between K156 and LTR. Furthermore, present findings resemble those on the stepwise increase in LTR length versus IN binding, mutational analysis, chemical modifications and photo cross-linking experiments, as well as those of DNA protection against DNase in the IN-viral LTR DNA complex (23,47,62–64). That the full enzyme recognizes the whole oligonucleotide (23) is conceivable, but how the relatively small $\alpha 4$ peptide recognizes both the terminal ACGT3' and the internal A-tract remains unclear.

NMR data describing the binding of K156 were obtained at a DNA:peptide ratio of 1:2 (at 700 MHz). Under these conditions, $\sim 50\%$ of K156 was bound to LTR34. Thus, the so measured CSDs should be weaker than CSDs obtained for an equimolecular complex. Both the side chain and backbone protons of several K156 residues underwent shifts in the complex. The backbone NH and α H protons may be influenced by local functional group contacts, and reflect to a good degree the changes to residues nearest the binding site (61,65). The chemical shifts for bound and unbound K156 are given in Supplementary Table S4. Overall, α H CSDs were larger than NH CSDs (an average of 0.09 ppm versus an average of 0.03 p.p.m.). The α Hs and NHs of residues Asn155, Lys156, Lys159, Glu(Ser)153, Glu152, Leu(Ile)161 and Leu(Ile)162 were most affected during binding (Figure 5). Unfortunately, Gln148 could not be identified in our spectra. All of these residues, except Leu161(Ile) and Leu162(Ile), formed the polar/charged face of the $\alpha 4$ helix (Figure 6). The binding of residues Lys156, Lys159 and Gln148 to viral LTR ends has already been shown through cross-linking experiments (23,44–46,48,49,51), whereas the binding of the Asn155 residue has been predicted by molecular modeling (57). Among these, the two positively charged residues Lys156 and Lys159 are required for high affinity binding between K156 and the U5 LTR (28). Lys156, Lys159 and Asn155 have been further shown to interact with the L731–988 derivative (66) and 5CITEP (67), two IN inhibitors supposed to interact at the DNA-protein interface. On another hand the CSD exhibited by Leu161 could be attributed to changes caused by interactions occurring with residues in close proximity or by the dissociation of K156 multimers stabilized by hydrophobic residues (68).

Possible DNA-peptide contacts

The LTR34 oligonucleotide is composed of 34 nucleotides, and 11 of the first 15 nucleotides interact with K156 via their bases (Figure 4). Most of these are important for 3'-processing (23). There is a good statistical correlation between the frequency of nucleobase conservation in DNAs and the interaction of these bases with protein amino acid side chains (69). At the same time, most of the $\alpha 4$ helix amino acids interacting with LTR DNA are themselves conserved residues, recognized for

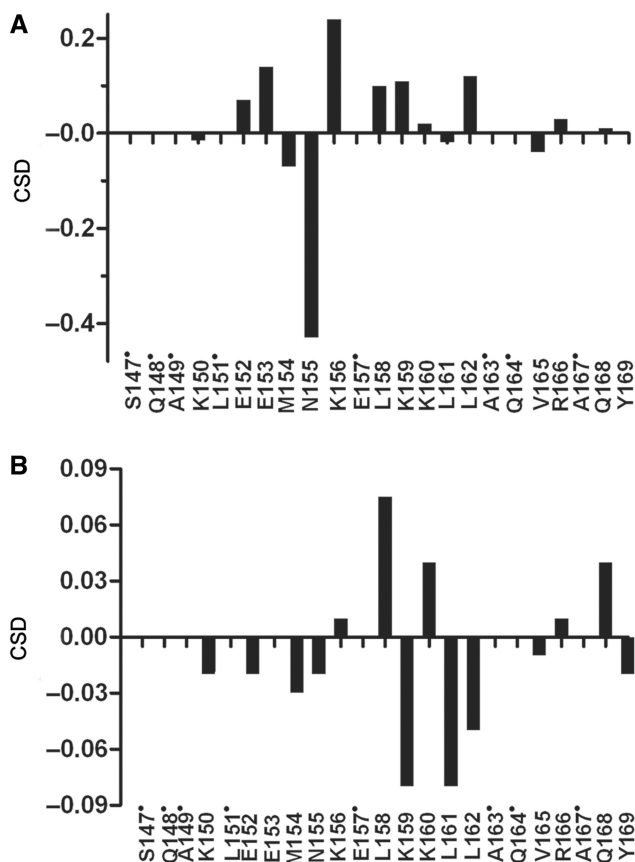


Figure 5. CSDs for the α Hs (A) and NHs (B) of K156 bound to LTR34 at a DNA:peptide ratio of 1:2. Non-assigned residues are indicated by a dot.

their role in the enzyme catalytic activity and virus infectivity (33,45). This suggests that DNA–protein recognition proceeds through the interaction of conserved complementary domains. Tight binding requires an optimal number of contacts between the amino acid chains and the bases, and also the backbone phosphates and sugars. These contacts include ionic interactions, van der Waals forces, and hydrogen bonds. Lys and Gln belong to a category of residues having side chains forming bidentate interactions with DNA bases, which in general provide higher specificity than those using single hydrogen bonds (70). The positively charged side chain of Lys is also frequently found in ionic interactions with DNA phosphate groups. For instance, the conserved Lys159 has been shown to interact with the adenine base and phosphate group of invariant CpA (47,48). Moreover, Lys159, together with Lys156, belongs to the Lys-rich sequence 156 Lys Glu Leu Lys Lys160, which has been implicated in the specific binding of IN to DNA (33). The base of adenine in CpA (and that of its complementary thymine) is also in contact with the catalytic Glu152 residue of the α 4 helix (54). The Gln148 residue (not identified in our spectra) has been shown to interact with the adenine (23) and the cytosine (46) of the 5'AC overhang in processed LTR.

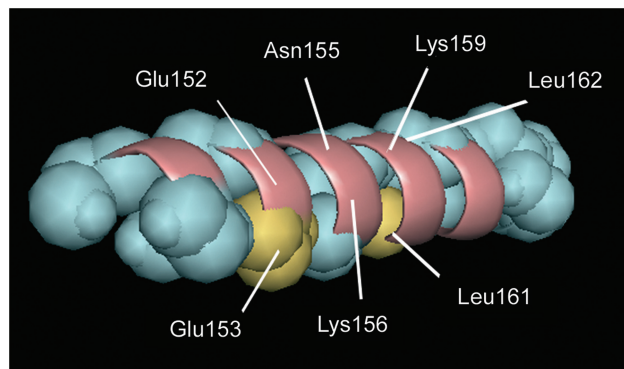


Figure 6. Backbone of the amphipathic K156 helix with residues of its polar/charged face (pink) that are most affected by LTR34, at a DNA:peptide ratio of 1:2. Yellow residues (Glu156, Leu161 and Leu162) are substituted residues showing CSD variations upon complex formation.

However, most importantly, Mg^{2+} is implicated in the specific recognition of U5 LTR by the α 4 helix of IN. How Mg^{2+} impacts the recognition? Mg^{2+} displays a high preference for guanine in crystal structures of DNA duplexes that are in either the A or B form (41,71). In several oligonucleotide crystals, Mg^{2+} has been identified within the major and the minor groove of ApG and GpT steps (41), which in LTRs are constitutive to the 3'-processing site. The base binding of Mg^{2+} in the major groove may cause helix bending by base-roll compression toward the major groove (41), which may then favor DNA cleavage from the minor groove. In fact, the CSDs for both aromatic and sugar protons strongly suggest that both the base and the scissile phosphodiester group of adenine are involved in interactions. Thus, binding of the Glu 152 carboxylic group may be mediated by the preferential binding of Mg^{2+} to this group.

CONCLUSION

Our data support the idea of a tight complex between viral DNA and IN formed through both specific interactions, involving amino acid side chains with DNA bases, and non-specific interactions involving amino acid side chains and DNA backbone groups. This suggests a recognition of viral DNA at the U5 LTR end that combines a direct and an indirect readout mechanism. In the indirect readout mechanism, the DNA sequence governs the geometry of the binding site, and thus the quality of interaction among amino acid side chains and backbone sugar and phosphate groups is strongly related to the sequence of the DNA. The direct readout mechanism involves direct contacts between protein functional groups and the DNA bases in appropriate geometry. Overall, the precise recognition of LTRs by IN is the result of the formation of a large network of interactions between the two partners. In this respect, the mechanism presented here resorts to many of the same basic features commonly found in the formation of specific DNA–protein complexes. Yet, it significantly differs from the

mechanism, mostly of indirect readout type, proposed on the bases of chemical modifications and nucleotide substitutions, as well as of a stepwise increase in DNA–ligand complexity (47,64,72). Such modifications in DNA may have effects on structural and energetic properties of local but also distant interactions, preventing a straightforward analysis of the binding.

Obviously, the mechanism through which IN recognizes the LTR ends requires further understanding, and the analysis must be extended to other regions of the protein. Yet, our results already shed light on the implication of the $\alpha 4$ helix and the divalent Mg^{2+} ion in the specific binding between IN and viral LTR DNA. The specific binding involves complementary conserved amino acid residues and nucleotides, suggesting that the IN $\alpha 4$ helix contains most of the specificity features that are required for a precise recognition of viral DNA, especially at the 3'-processing site. Clearly, a model that would reproduce the relevant binding properties of the enzyme could improve our understanding of the mechanisms underlying complex formation and might also facilitate the design of drugs blocking viral integration.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

SIDACTION grant (to S.F.) and by the French-Lebanese program CEDRE (05 SF21/L14 to S.F. and R.M.). Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

- Craigie,R., Mizuuchi,K., Bushman,F.D. and Engelman,A. (1991) A rapid in vitro assay for HIV DNA integration. *Nucleic Acids Res.*, **19**, 2729–2734.
- Asante-Appiah,E. and Skalka,A.M. (1997) Molecular mechanisms in retrovirus DNA integration. *Antiviral Res.*, **36**, 139–156.
- Katz,R.A. and Skalka,A.M. (1994) The retroviral enzymes. *Annu. Rev. Biochem.*, **63**, 133–173.
- Sherman,M.P. and Greene,W.C. (2002) Slipping through the door: HIV entry into the nucleus. *Microbes. Infect.*, **4**, 67–73.
- van Gent,D.C., Oude Groeneger,A.A. and Plasterk,R.H. (1993) Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini. *Nucleic Acids Res.*, **21**, 3373–3377.
- Engelman,A., Bushman,F.D. and Craigie,R. (1993) Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.*, **12**, 3269–3275.
- Cai,M., Zheng,R., Caffrey,M., Craigie,R., Clore,G.M. and Gronenborn,A.M. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.*, **4**, 567–577.
- Engelman,A. and Craigie,R. (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. *J. Virol.*, **66**, 6361–6369.
- Bujacz,G., Jaskolski,M., Alexandratos,J., Wlodawer,A., Merkel,G., Katz,R.A. and Skalka,A.M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure.*, **4**, 89–96.
- Dyda,F., Hickman,A.B., Jenkins,T.M., Engelman,A., Craigie,R. and Davies,D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.
- Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P. and Skalka,A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell Biol.*, **12**, 2331–2338.
- Leavitt,A.D., Shiue,L. and Varmus,H.E. (1993) Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J. Biol. Chem.*, **268**, 2113–2119.
- Rice,P. and Mizuuchi,K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209–220.
- Davies,J.F., Hostomska,Z., Hostomsky,Z., Jordan,S.R. and Matthews,D.A. (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.
- Katayanagi,K., Okumura,M. and Morikawa,K. (1993) Crystal structure of Escherichia coli RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single Mg^{2+} -binding site. *Proteins*, **17**, 337–346.
- Musacchio,A., Noble,M., Pauptit,R., Wierenga,R. and Saraste,M. (1992) Crystal structure of a Src-homology 3 (SH3) domain. *Nature*, **359**, 851–855.
- Ren,G., Gao,K., Bushman,F.D. and Yeager,M. (2007) Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA. *J. Mol. Biol.*, **366**, 286–294.
- Podtelezhnikov,A.A., Gao,K., Bushman,F.D. and McCammon,J.A. (2003) Modeling HIV-1 integrase complexes based on their hydrodynamic properties. *Biopolymers*, **68**, 110–120.
- Karki,R.G., Tang,Y., Burke,T.R. Jr and Nicklaus,M.C. (2004) Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *J. Comput. Aided Mol. Des.*, **18**, 739–760.
- Wielens,J., Crosby,I.T. and Chalmers,D.K. (2005) A three-dimensional model of the human immunodeficiency virus type 1 integration complex. *J. Comput. Aided Mol. Des.*, **19**, 301–317.
- Beese,L.S. and Steitz,T.A. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.
- Engelman,A. and Craigie,R. (1995) Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J. Virol.*, **69**, 5908–5911.
- Esposito,D. and Craigie,R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.*, **17**, 5832–5843.
- Katzman,M., Katz,R.A., Skalka,A.M. and Leis,J. (1989) The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J. Virol.*, **63**, 5319–5327.
- Pommier,Y., Johnson,A.A. and Marchand,C. (2005) Integrase inhibitors to treat HIV/AIDS. *Nat. Rev. Drug Discov.*, **4**, 236–248.
- Yang,W., Lee,J.Y. and Nowotny,M. (2006) Making and breaking nucleic acids: two- Mg^{2+} -ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
- Asante-Appiah,E., Seeholzer,S.H. and Skalka,A.M. (1998) Structural determinants of metal-induced conformational changes in HIV-1 integrase. *J. Biol. Chem.*, **273**, 35078–35087.
- Zargarian,L., Benleumi,M.S., Renisio,J.G., Merad,H., Maroun,R.G., Wieber,F., Mauffret,O., Porumb,H., Troalen,F. and Femandjian,S. (2003) Strategy to discriminate between high and low affinity bindings of human immunodeficiency virus, type 1 integrase to viral DNA. *J. Biol. Chem.*, **278**, 19966–19973.
- Merad,H., Porumb,H., Zargarian,L., Rene,B., Hobaika,Z., Maroun,R.G., Mauffret,O. and Femandjian,S. (2009) An unusual helix turn helix motif in the catalytic core of HIV-1 integrase binds viral DNA and LEDGF. *PLoS One*, **4**, e4081.
- Zhong,L. and Johnson,W.C. Jr. (1992) Environment affects amino acid preference for secondary structure. *Proc. Natl Acad. Sci. USA*, **89**, 4462–4465.
- Wüthrich,K. (1986) *NMR of Protein and Nucleic Acids*. Wiley, New York, NY.

32. Griesinger, C., Otting, G., Wuthrich, K. and Ernst, R.R. (1988) Clean Tocsy for H-1 spin system-identification in macromolecules. *J. Am. Chem. Soc.*, **110**, 7870–7872.
33. Semenova, E.A.M.C.a.P.Y. (2008) HIV-1 integrase inhibitors: update and perspectives. *Adv. Pharmacol.*, **56**, 199–228.
34. Rhee, S.Y., Liu, T.F., Kiuchi, M., Zioni, R., Gifford, R.J., Holmes, S.P. and Shafer, R.W. (2008) Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology*, **5**, 74.
35. Draper, D.E. (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.*, **95**, 5489–5495.
36. Sissi, C. and Palumbo, M. (2009) Effects of magnesium and related divalent metal ions in topoisomerase structure and function. *Nucleic Acids Res.*, **37**, 702–711.
37. Cowan, J.A. (2002) Structural and catalytic chemistry of magnesium-dependent enzymes. *Biometals*, **15**, 225–235.
38. Subirana, J.A. and Soler-Lopez, M. (2003) Cations as hydrogen bond donors: a view of electrostatic interactions in DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 27–45.
39. Misra, V.K. and Draper, D.E. (1998) On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113–135.
40. Manning, G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.*, **11**, 179–246.
41. Chiu, T.K. and Dickerson, R.E. (2000) 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.*, **301**, 915–945.
42. Renisio, J.G., Cosquer, S., Cherrak, I., El Antri, S., Mauffret, O. and Femandjian, S. (2005) Pre-organized structure of viral DNA at the binding-processing site of HIV-1 integrase. *Nucleic Acids Res.*, **33**, 1970–1981.
43. Misra, V.K. and Draper, D.E. (1999) The interpretation of Mg(2+) binding isotherms for nucleic acids using Poisson-Boltzmann theory. *J. Mol. Biol.*, **294**, 1135–1147.
44. Gao, K., Butler, S.L. and Bushman, F. (2001) Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J.*, **20**, 3565–3576.
45. Lu, R., Limon, A., Ghory, H.Z. and Engelman, A. (2005) Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase. *J. Virol.*, **79**, 2493–2505.
46. Johnson, A.A., Santos, W., Pais, G.C., Marchand, C., Amin, R., Burke, T.R. Jr, Verdine, G. and Pommier, Y. (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. *J. Biol. Chem.*, **281**, 461–467.
47. Agapkina, J., Smolov, M., Barbe, S., Zubin, E., Zatsepina, T., Deprez, E., Le, B.M., Mouscadet, J.F. and Gottikh, M. (2006) Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA. *J. Biol. Chem.*, **281**, 11530–11540.
48. Jenkins, T.M., Esposito, D., Engelman, A. and Craigie, R. (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.*, **16**, 6849–6859.
49. Drake, R.R., Neamati, N., Hong, H., Pilon, A.A., Sunthakar, P., Hume, S.D., Milne, G.W. and Pommier, Y. (1998) Identification of a nucleotide binding site in HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **95**, 4170–4175.
50. Lutzke, R.A., Vink, C. and Plasterk, R.H. (1994) Characterization of the minimal DNA-binding domain of the HIV integrase protein. *Nucleic Acids Res.*, **22**, 4125–4131.
51. Heuer, T.S. and Brown, P.O. (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry*, **37**, 6667–6678.
52. Dirac, A.M. and Kjems, J. (2001) Mapping DNA-binding sites of HIV-1 integrase by protein footprinting. *Eur. J. Biochem.*, **268**, 743–751.
53. Engelman, A., Liu, Y., Chen, H., Farzan, M. and Dyda, F. (1997) Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. *J. Virol.*, **71**, 3507–3514.
54. Gerton, J.L., Ohgi, S., Olsen, M., DeRisi, J. and Brown, P.O. (1998) Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions. *J. Virol.*, **72**, 5046–5055.
55. Harper, A.L., Skinner, L.M., Sudol, M. and Katzman, M. (2001) Use of patient-derived human immunodeficiency virus type 1 integrases to identify a protein residue that affects target site selection. *J. Virol.*, **75**, 7756–7762.
56. Chen, A., Weber, I.T., Harrison, R.W. and Leis, J. (2006) Identification of amino acids in HIV-1 and avian sarcoma virus integrase subsites required for specific recognition of the long terminal repeat Ends. *J. Biol. Chem.*, **281**, 4173–4182.
57. Perryman, A.L. and McCammon, J.A. (2002) AutoDocking dinucleotides to the HIV-1 integrase core domain: exploring possible binding sites for viral and genomic DNA. *J. Med. Chem.*, **45**, 5624–5627.
58. Leberman, R. and Soper, A.K. (1995) Effect of high salt concentrations on water structure. *Nature*, **378**, 364–366.
59. Kielkopf, C.L., Ding, S., Kuhn, P. and Rees, D.C. (2000) Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG)(2). *J. Mol. Biol.*, **296**, 787–801.
60. Varani, G., Chen, Y. and Leeper, T.C. (2004) NMR studies of protein-nucleic acid interactions. *Methods Mol. Biol.*, **278**, 289–312.
61. Fielding, A.J., Usselman, R.J., Watmough, N., Simkovic, M., Frerman, F.E., Eaton, G.R. and Eaton, S.S. (2008) Electron spin relaxation enhancement measurements of interspin distances in human, porcine, and Rhodobacter electron transfer flavoprotein-ubiquinone oxidoreductase (ETF-QO). *J. Magn Reson.*, **190**, 222–232.
62. Ellison, V. and Brown, P.O. (1994) A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc. Natl Acad. Sci. USA*, **91**, 7316–7320.
63. Vink, C., Banks, M., Bethell, R. and Plasterk, R.H. (1994) A high-throughput, non-radioactive microtiter plate assay for activity of the human immunodeficiency virus integrase protein. *Nucleic Acids Res.*, **22**, 2176–2177.
64. Bugreev, D.V., Baranova, S., Zakhara, O.D., Parissi, V., Desjober, C., Sottofattori, E., Balbi, A., Litvak, S., Tarrago-Litvak, L. and Nevinsky, G.A. (2003) Dynamic, thermodynamic, and kinetic basis for recognition and transformation of DNA by human immunodeficiency virus type 1 integrase. *Biochemistry*, **42**, 9235–9247.
65. Hunter, T.M., McNae, I.W., Liang, X., Bella, J., Parsons, S., Walkinshaw, M.D. and Sadler, P.J. (2005) Protein recognition of macrocycles: binding of anti-HIV metalloproteinase inhibitors to lysozyme. *Proc. Natl Acad. Sci. USA*, **102**, 2288–2292.
66. Sechi, M., Carta, F., Sannia, L., Dallochio, R., Dessi, A., Al-Safi, R.I. and Neamati, N. (2009) Design, synthesis, molecular modeling, and anti-HIV-1 integrase activity of a series of photoactivatable diketo acid-containing inhibitors as affinity probes. *Antiviral Res.*, **81**, 267–276.
67. Goldgur, Y., Craigie, R., Cohen, G.H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. and Davies, D.R. (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proc. Natl Acad. Sci. USA*, **96**, 13040–13043.
68. Maroun, R.G., Zargarian, L., Stocklin, R., Troalen, F., Jankowski, C.K. and Femandjian, S. (2005) A structural study of model peptides derived from HIV-1 integrase central domain. *Rapid Commun. Mass Spectrom.*, **19**, 2539–2548.
69. Mirny, L.A. and Gelfand, M.S. (2002) Structural analysis of conserved base pairs in protein-DNA complexes. *Nucleic Acids Res.*, **30**, 1704–1711.
70. Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, **8**, 937–946.
71. Egli, M., Tereshko, V., Mushudov, G.N., Sanishvili, R., Liu, X. and Lewis, F.D. (2003) Face-to-face and edge-to-face pi-pi interactions in a synthetic DNA hairpin with a stilbenediether linker. *J. Am. Chem. Soc.*, **125**, 10842–10849.
72. Wang, J.Y., Ling, H., Yang, W. and Craigie, R. (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.*, **20**, 7333–7343.