Archives of Biochemistry and Biophysics 495 (2010) 28–34

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Original Paper

Role of the 207–218 peptide region of Moloney murine leukemia virus integrase in enzyme catalysis

Mónica L. Acevedo [a], José Jaime Arbildúa [c], Octavio Monasterio [c], Héctor Toledo [b], Oscar León [a,*]

[a] Programa de Virología, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago, Chile
[b] Programa de Biología Celular y Molecular, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Independencia 1027, Santiago, Chile
[c] Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile

ARTICLE INFO

Article history:
Received 30 September 2009
and in revised form 11 December 2009
Available online 21 December 2009

Keywords:
Integrase
Retrovirus
Mutagenesis
Flexible loop
Molecular modeling
B-factor

ABSTRACT

X-ray diffraction data on a few retroviral integrases show a flexible loop near the active site. By sequence alignment, the peptide region 207–218 of Mo-MLV IN appears to correspond to this flexible loop. In this study, residues H208, Y211, R212, Q214, S215 and S216 of Mo-MLV IN were mutated to determine their role on enzyme activity. We found that Y211A, R212A, R212K and Q214A decreased integration activity, while disintegration and 3′-processing were not significantly affected. By contrast H208A was completely inactive in all the assays. The core domain of Mo-MLV integrase was modeled and the flexibility of the region 207–216 was analyzed. Substitutions with low integration activity showed a lower flexibility than wild type integrase. We propose that the peptide region 207–216 is a flexible loop and that H208, Y211, R212 and Q214 of this loop are involved in the correct assembly of the DNA-integrase complex during integration.

Introduction

Integration of viral DNA in the genome of the infected cell is an essential property of the retroviruses. This process is catalyzed by the viral integrase in two steps. In the first reaction, the double stranded proviral DNA is processed at both 3′OH ends by integrase, releasing the terminal dinucleotides. A preintegration complex that includes viral and cellular proteins is directed to the nucleus where both ends of the processed viral DNA are inserted in a concerted reaction in the genome of the host cell. This reaction is also catalyzed by integrase (for review see Ref. [1]).

Retroviral integrases are organized in three structural domains: N-terminal domain, central domain and C-terminal domain. The first two domains are conserved among different retroviruses [2–4]. The central domain contains the triad DDE, responsible for the catalytic activity. Its structure is composed of several α-helices and β-sheets that have compact shape like other polynucleotidyl-phosphotransferases [5]. In the catalytic domain, there is a flexible loop adjacent to the active site. This loop corresponds to amino acids 140–149 in HIV-1 integrase [6] and to the boundary amino acids 207–218 in Mo-MLV integrase.

In the X-ray structures of the core domain of HIV-1 IN, the flexible loop is missing or presents different conformations with high temperature factors, and it has been suggested that the loop conformation is dependent upon the presence of substrate and inhibitors [7].

Crosslinking studies in HIV-1 integrase showed that during the disintegration reaction the substrate binds near the flexible loop [8], involving specifically residues Q148 and Y143 in an interaction with the last base of the single strand tail of the non-processed strand of the LTR [9,10]. Furthermore, disulfide crosslinking demonstrated an interaction between the terminal 5′ cytosine of the single strand tail of a disintegration substrate with Q148 in the flexible loop of HIV-1 integrase, and the authors suggested that possibly other amino acids of this flexible loop may be involved in DNA binding [11].

Modeling studies in HIV-1 IN suggested that the flexible loop is part of a hydrophobic cavity formed by amino acids: Q62, L63, H114, T115, D116, N117, D139, Y143, S147, Q148, V150, and V151 or I151. This pocket is a common binding site of IN inhibitors and it is shared with the 3′-GT dinucleotide of the unprocessed viral DNA. This hydrophobic pocket would form after DNA processing [7].

A previous model of the structure of a small region of the catalytic domain of Mo-MLV IN suggested that the region between H208 and S215 could form a loop similar to the flexible loop of HIV-1 IN [12]. Crosslinking and chemical modification of Mo-MLV integrase indicated that the 5′ region of the single strand tail of the non-processed strand of viral DNA binds near C209 in the flexible loop [12]. The aim this study was to determine the effect

* Corresponding author. Fax: +56 2 978 6124.
E-mail addresses: macevedo@med.uchile.cl (M.L. Acevedo), jarbildu@cimm.cl (J.J. Arbildúa), monaster@uchile.cl (O. Monasterio), htoledo@med.uchile.cl (H. Toledo), oleon@med.uchile.cl (O. León).

0003-9861/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2009.12.018

of amino acid substitutions in the peptide region 207–218 of Mo-MLV integrase on the catalytic activity. A model of the catalytic domain of Mo-MLV IN was constructed based on 3D-structures of retroviral integrases to analyze the flexibility of the region 207–218 in wild type and mutant INs.

Materials and methods

Materials

Crude [γ-^{32}P]ATP (7000 Ci/mmol) was purchased from Perkin Elmer. T4 polynucleotide kinase, T4 DNA ligase, and restriction enzymes were obtained from New England Biolabs. *Pfu* turbo DNA polymerase was from Stratagene. Ni^{2+}–nitrilotriacetic acid agarose (Ni^{2+}–NTA) was purchased from Novagen.

Oligonucleotide synthesis

DNA oligonucleotides (IdT)[1] were purified by electrophoresis on 20% acrylamide, 7 M urea gels [13]. The following oligonucleotides were used: T7 promoter-fwd 5′-TAATACGACTCACTATAGGG-3′, T7 terminator-rev 5′-GCTAGTTATTGCTCAGCGG-3′.

Mutagenic primers: forward and reverse primers for each substitution are listed: 5′AACTCTTCTGCGAGACCCCAAAGCTCAGGC CA-3′ and 5′ AACTCTTCTCGCTGCACAATGTAATTTCCAATCAATCCC-3′ (Y211A), 5′ AACTCTTCTTGGAGACCCCAAAGCTCAGGCCA-3′ and 5′ AACTCTTCTCCATGCACAATGTAATTTCCAATCAATCCC-3′ (Y211W), 5′ AACTCTTCTCAGCCCCAAAGCTCAGGCCAGGTA-3′ and 5′ AACTCTT CTCTGGTATGCACAATGTAATTTCCAATCAATCC-3′ (R212Q), 5′ AACTC TTCTGCGCCCCAAAGCTCAGGCCAG-3′ and 5′ AACTCTTCTCGCGTATG CACAATGTAATTTCCAATCAATCC-3′ (R212A), 5′ AACTCTTCTAAACCC CAAAGCTCAGGCCAGGTA-3′ and 5′ AACTCTTCTTTTGTATGCACAATG TAATTTCCAATCAATCC-3′ (R212K), 5′ AACTCTTCTGCGAGCTCAGGCC AGGTAGAAAGAATG-3′ and 5′ AACTCTTCTCGCGGGTCTGTATGCACA ATGTAATTTCC-3′ (Q214A), 5′ AACTCTTCTAACAGCTCAGGCCAGGTA GAAAGAATG-3′ and 5′ AACTCTTCTGTTGGGTCTGTATGCACAATGTA ATTTCCAAT-3′ (Q214N), 5′AACTCTTCTGCGTCAGGCCAGGTAGAAAG AATGAATA-3′ and 5′ AACTCTTCTCGCTTGGGGTCTGTATGCACAATGT AATTT-3′ (S215A), 5′ AACTCTTCTGCGGGCCAGGTAGAAAGAATGAAT AGAA-3′ and 5′ AACTCTTCTCGCGCTTTGGGGTCTGTATGCACAAT-3′ (S216A).

Site specific mutagenesis

Integrase mutants were generated as described in Ref. [14]. Briefly, two separated DNA fragments of the integrase gene cloned in the plasmid pETIN1 [13] were amplified by PCR using the pairs T7 promoter-rev mutagenic primer and T7 terminator-fwd mutagenic primer. The mutagenic primers had a site for the restriction enzyme EarI (5′-CTCTTCN*-3′), to generate sticky ends. The reaction contained 2.5 U of *Pfu*Turbo DNA polymerase (Stratagene), 200 μM of dNTPs, 1 μM of primers and 100 ng of pETIN1 in 50 μl in 1 X *Pfu* DNA polymerase reaction buffer. The amplification program was 94 °C for 4 min, 30 cycles of 94 °C for 30 s, 57 °C for 1 min and 72 °C for 1 min, and extension at 72 °C for 7 min. The amplified DNA fragments were purified using Wizard SV gel columns and a PCR clean-up system (Promega). Two microgram of plasmid DNA were digested with 10 U of EarI (New England Biolabs) for 2 h at 37 °C in 10 mM Bis–Tris–propane–HCl, pH 7.0; 10 mM $MgCl_2$; 1 mM DTT. Both digested DNA fragments were purified using the Wizard SV clean-up system and ligated with 400 U of T4 DNA ligase (New England Biolabs) for 2 h at room temperature in 50 mM Tris–HCl pH 7.5; 10 mM $MgCl_2$; 1 mM ATP; 10 mM DTT. The 1389 bp fragment containing the complete sequence of integrase was digested with 10 U of SacII and BamHI, purified by electrophoresis on 1% agarose gels and ligated to pETIN1 digested with SacII and BamHI. This DNA was used to transform competent *Escherichia* coli DH5α cells. Mutations were confirmed by DNA sequencing.

Purification of the integrase mutants

Recombinant wild type and mutant integrases were purified by chromatography on Ni–NTA agarose as described in Ref. [15]. The enzyme was dialyzed against 20 mM Hepes pH: 7.4, 1.5 mM DTT, 400 mM KCl, 0.1% NP-40 and 20% glycerol.

Activity assays

Oligonucleotides were labeled with [γ-^{32}P] ATP by a kinase reaction as previously described [13]. The 5′-labeled oligodeoxynucleotide 5′-GATCCGACTACCCGTCAGCGGGGGTCTTTCATT-3′ annealed to 5′-AATGAAAGACCCCCGCTGACGGGTAGTCGGATC 3′ was used for 3′-processing. The reaction contained 1 pmol labeled substrate, 20 mM MES (pH 6.2), 100 mM KCl, 10 mM $MnCl_2$, 10 mM DTT and 10% (v/v) glycerol (SST buffer) in a final volume of 15 μl. The reaction was started by adding 4.5 pmol of integrase in a final volume of 15 μl. The reaction mixture was incubated at 37 °C for 1 h. The 3′-processing reaction was stopped with 10 μg of proteinase K and incubated at 55 °C for 1 h, then the DNA products were precipitated with 0.1 volume of 3 M ammonium acetate plus 2.5 volumes of 100% ethanol at −20 °C for 16 h, centrifuged for 10 min at 4 °C and the pellet was washed with 80% ethanol. The dried products were dissolved in water and 95% formamide, 20 mM EDTA, 0.05% bromophenol blue (loading buffer) (1:1) heated at 90 °C for 5 min. The reaction products were separated on 20% acrylamide, 7 M urea in Tris/borate/EDTA (TBE) buffer (pH 7.6), dried, and exposed on a phosphor-Imager (Bio-Rad). The images were visualized and quantified using the program Quantity One 4.2.1. (Bio-Rad). Disintegration assays were performed using a 5′-end radiolabeled dumbbell substrate, as described previously [16].

The concerted two-end integration assay is a modification of that previously described [17]. One pmole of the 5′-labeled substrate used for strand transfer was incubated in SST buffer with 4.5 pmol of wt or mutant IN for 30 min in ice. Then 100 pmol of pUC18 DNA were added and incubated at 37 °C for 1 h in a final volume of 15 μl. The reaction was stopped with 3 μl of 0.1 M EDTA, 5% SDS plus 1 μl with 10 μg of proteinase K and incubated at 55 °C for 1 h. The reaction products were separated by electrophoresis on 1% agarose gels. The gel was dried and exposed on a Phosphor-Imager (Bio-Rad) and the products were quantified using the program Quantity One 4.2.1.

Structural modeling

Mo-MLV integrase is composed of three domains; however, a reliable model was obtained only for the central catalytic domain that comprises amino acids 116–273. The 3D model was built with the program MODELLER v6.2 [18] using three templates: 1vsd.pdb (residues 55–199) of the ASV integrase core domain, crystallized with Mg^{2+}, 1c6v.pdb (residues 57A–205A) of SIV integrase (catalytic domain + DNA binding domain comprise residues 50–293) and 1ex4.pdb (residues 56A–205A) of the HIV-1 integrase catalytic core and C-terminal domain. In the first step of the modeling, sequences of known INs structures were aligned with the malign3d command of MODELLER, and the profile Hidden Markov Models (HMM profile) [19] of the PFAM family PF00665 (Integrase core domain, seed alignment) [20] was used with HMMer v2.3.2 to align

[1] *Abbreviations used:* ASV, avian sarcoma virus; HIV, human immunodeficiency virus; Mo-MLV, Moloney murine leukemia virus; LTR, long terminal repeat; RMSD, root mean square deviation, wt, wild type.

the Mo-MLV sequence against the structural alignment. The second step involved choosing complementary templates to perform the actual modeling. The reliability of the 3D structure obtained was confirmed with the statistical potentials software PROSA (Protein Structure Analysis) [21] and DOPE (Discrete Optimized Protein Energy) [22].

The secondary structure assignment program (DSSP) [23], was used to determine the conformation of the region 207–218 in the 3D model of the catalytic domain of Mo-MLV integrase.

A loop modeling routine (MODELLER v6.2) was used for the wild type and point mutation Mo-MLV 3D-structures, comprising amino acids 207–218 with a total of 1000 models for each structure [24]. The calculations were performed in a 20 PIII 1 GHz CPU cluster.

Alpha-carbon displacement by B-factor

The flexibility of region 207–216 in the structures of Mo-MLV wild type and the mutants were represented as Cα B-factors. In summary, from the 1000 loop configurations obtained for Mo-MLV wild type or one of the mutants, the coordinates of those with an energy value of 100 or less were selected, and for each of the Cα amino acids the mean-square displacement (μ^2) was calculated. With the assumption of an isotropic movement the B-factor is:

$$\text{B-factor} = 8 \times \pi^2 \times \mu x^2$$

Statistical analysis

Statistical significance of experimental data sets of flexibility by atomic displacement parameter was assessed using Student's t-test, performed with the Microsoft Excel program. A value of >99% was considered as a statistically significant difference.

Results

Mo-MLV IN mutagenesis

In order to examine the role of the peptide region 207–218 of Mo-MLV IN on enzyme activity, several residues from within this region were substituted by site directed mutagenesis. Some of these residues were selected either on the basis of their conservation among several integrases or in consideration of specific functional groups that could potentially interact with DNA. Fig. 1 shows a comparison of the amino acid sequences of six INs from different species. The residues H208, Y211, R212, Q214, S215 and S216 were selected for mutagenesis.

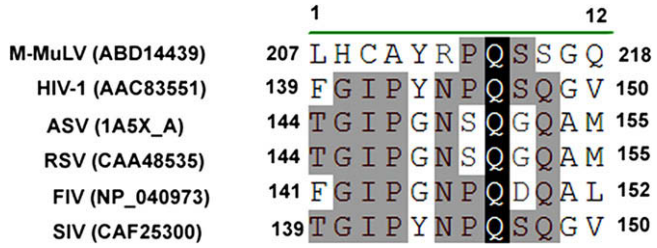

Fig. 1. Amino acid sequence alignment of the region corresponding to flexible loop of several retroviral integrases. Amino acid sequences were aligned with the program AlignX of Vector NTI suite 9 (Invitrogen). Identical amino acids are indicated in black shading; conservative amino acids are in dark grey.

Integrase mutants, encoding a His-tag at the carboxyl terminus, were expressed in *E. coli* BL21 (DE3) and purified as described in Materials and methods. Approximately 2–3 mg of enzyme was obtained for each of the mutants from 400 ml of cell culture. In the functional studies described below, at least three different enzyme preparations of the mutant INs were tested.

Catalytic activity

To determine the effect of each mutation on Mo-MLV IN we measured the two-end concerted integration activity of the IN mutants using a 30-mer processed LTR substrate and a plasmid as an acceptor (Fig. 2A). The products of this reaction have been well characterized previously [15,17,25,26]. As expected for wt, IN two products were observed (Fig. 3A, lane 1). These products correspond to a single end non-concerted integration (NCI) and a two-end concerted integration (CI). In these assays the labeled substrate migrated off the gel in order to get a good separation of the integration products.

H208A was inactive for concerted integration (Fig. 3A, lane 12). Y211A was 2.5% active (Fig. 3A, lane 2 and Table 1); however, Y211W (Fig. 3A lane 3), which maintains the hydrophobicity of the side chain at position 211, was 54% active (Table 1). R212A, R212Q and R212K (Fig. 3A, lanes 4–6) decreased concerted integration activity to less than 10% (Table 1). Q214A decreased the activ-

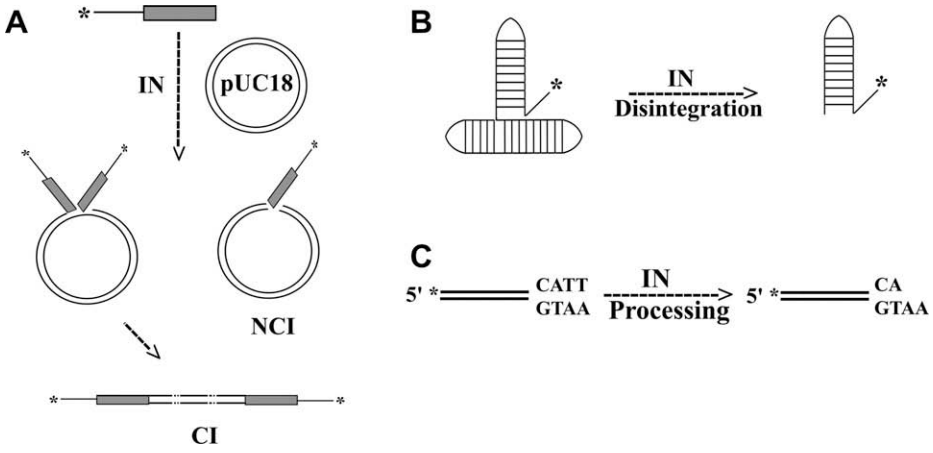

Fig. 2. Chemical reactions catalyzed by IN. In all these assays the substrates were labeled with ^{32}P (*) at the 5′-end. (A) For concerted integration assay a double stranded DNA mimicking the sequence of the processed U5 LTR end was used as the donor substrate and pUC18 was used as the acceptor DNA. Concerted integration (CI) of two donor substrates into the acceptor DNA yields a linear product, while non-concerted integration (NCI) yields circular products. (B) For disintegration assay a dumbbell oligonucleotide mimicking the integration intermediate was used. IN catalyzes the release of viral and target DNAs. (C) For 3′-processing an oligonucleotide containing the sequence of the LTR end of the Mo-MLV DNA was used. IN catalyzes the removal of two nucleotides of the labeled DNA after CA, releasing a −2 product.

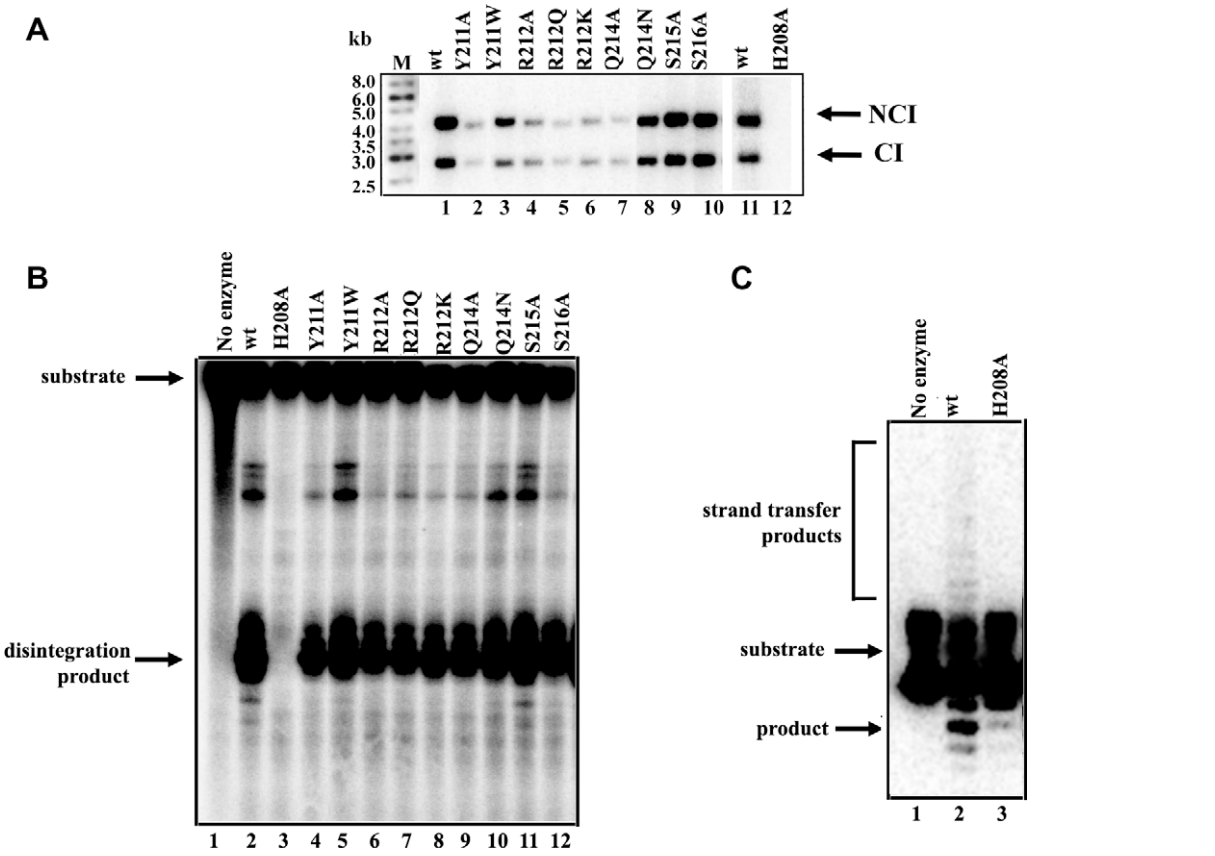

Fig. 3. Enzyme activity assays of wild type and IN mutants. (A) Integration assays were carried out with 4.5 pmol of wild type and mutant INs and the products were electrophoresed on agarose gels. The upper product corresponds to non-concerted integration (NCI) and the lower product to concerted integration (CI). The substrate migrated off the gel. The position of the mass standard marker is indicated (M) on the left. (B) Disintegration activity was tested with 4.5 pmol of the enzyme and 1 pmol of the [^{32}P]-labeled oligonucleotide. The position of the substrates and products of the reaction are indicated by the arrows on the left. (C) 3′-processing activity was determined as described in Methods; 4.5 pmol of enzyme were used in each assay. Reactions were carried out at 37 °C and the products were analyzed by electrophoresis on polyacrylamide denaturing gels. Strand transfer products larger than the input substrate are indicated by a bracket. The position of the substrate and −2 product are indicated by the arrows on the left. Recombinant INs are indicated on top of each lane. Lane numbers are indicated below the gel.

ity to about 6% (Fig. 3A, lane 7, Table 1), in contrast to Q214N, which was 84% active (Fig. 3A, lane 8, Table 1). S215A and S216A showed an activity similar to wild type IN (Fig. 3A, lanes 9 and 10 and Table 1).

Retroviral integrases catalyze the *in vitro* disintegration of viral DNA. This reaction has provided a way to examine the role of the ss-tail in a single site integration intermediate. We determined the disintegration activity of the IN mutants on a dumbbell substrate (Fig. 2B). The results of these assays are shown in Fig. 3B and Table 1. All the mutants except H208A (Fig. 3B, lane 3) were active for disintegration. However, the activity of Y211A, R212A, R212Q, R212K and Q214A (Fig. 3B, lanes 4, 6–9) decreased between 26% and 55% compared to wt IN (Fig. 3B lane 2 and Table 1).

Table 1
Percentage of relative activity[a] of the mutants of the region 207–218 of IN on each activity assay.

Mutant IN	Integration		Disintegration
	Concerted	Non-concerted	
Wild type	100	100	100
H208A	0.7 ± 1	0.6 ± 1	0.5 ± 1
Y211A	2.5 ± 2	5.4 ± 3	45 ± 7
Y211W	54 ± 6	54 ± 9	90 ± 5
R212A	9 ± 6	11 ± 5	68 ± 3
R212Q	7 ± 2	8 ± 2	71 ± 3
R212K	7 ± 2	7 ± 2	74 ± 5
Q214A	6 ± 1	6 ± 1	73 ± 6
Q214N	84 ± 10	82 ± 7	81 ± 7
S215A	105 ± 11	114 ± 5	91 ± 7
S216A	110 ± 13	108 ± 14	79 ± 4

[a] Percent of mean* activity of the mutants IN relative to the wild type IN ± standard deviation.
* Three different enzyme preparations of the mutant INs were tested.

In 3′-processing (Fig. 2C) only H208A was significantly affected. As shown in Fig. 3C the 3′-processing activity was abolished. The very small amount of product seen in H208A was due to a contaminant nonspecific nuclease present in the preparation as shown by using a mutated LTR substrate (not shown).

By photo-crosslinking we found that H208A was able to bind a DNA dumbbell substrate (not shown) indicating that the loss of activity in H208A was not due to a loss of DNA binding.

Model of the region 207–218 of Mo-MLV integrase

Mo-MLV IN has not been crystallized and in order to explain the mutagenesis results, a model of 158 residues of the core domain of Mo-MLV IN (residues 116–273) was obtained as described in Material and methods and shown in Fig. 4. We found that region 207–218 corresponds to a loop structure. Using the program DSSP, we concluded that the loop of Mo-MLV IN contains ten amino acids that include L207 to S216. K206 and G217 were the anchor residues of the loop ($loop_{207-216}$) (Fig. 4).

Using the loop modeling routine of MODELLER v6.2 we obtained 1000 models for the $loop_{207-216}$. These models were analyzed in plots of Cα RMSD against their energy score obtained by MODEL-

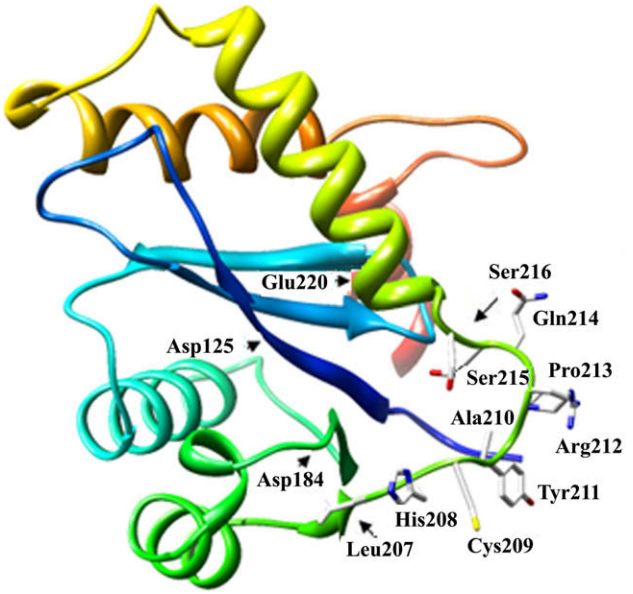

Fig. 4. Core domain model of wild type Mo-MLV integrase. Integrase is shown as ribbons. The IN model was built by MODELLER v6.2. from amino acid 116–273. This model was selected at random from among the lower energy models group. Amino acids of the loop are labeled. Anchor and catalytic residues of the loop are indicated by the arrows.

LER. Plots of wild type and mutant INs showed a similar biphasic profile with a first component with energy values of less than 100, where different loop conformations produce similar energy values, and a second component where increasing values of RMS showed increasing values of energy. The behavior of the first component represents the flexibility of the loop, and the second component the less probable configurations (higher energy) (Fig. 5).

Analysis of the loop$_{207–216}$ flexibility

The flexibility was analyzed by atomic displacement analysis of alpha-carbon expressed as B-factor. Mg^{2+} was not considered in this analysis since in the template 1vsd.pdb it is located at 6 Å from the closest amino acid residue of the loop and the flexibility of the loop is intrinsic to its amino acid sequence. Fig. 6 shows the flexibility profile for each of the loop mutants. The red symbols represent mutants with less than 40% activity with respect to wild type integrase, and blue symbols are mutants with more than 40% activity. Mutant S216A had the most flexible profile whereas mutant R212K was the least flexible. In general, mutants having lower activity showed lower flexibility; this is best observed between residues 210 and 212; the region which corresponds to the central residue 211 in the flexibility plot of Fig. 6 ($p < 0.01$). This is the same location where the functional residues Y211 and R212 were found and very close to the essential residue H208. The flexibility results suggest that not only residues H208, Y211 and R212 were important for Mo-MLV integrase function, but the flexibility of the loop$_{207–216}$ is relevant especially in the 209–212 region.

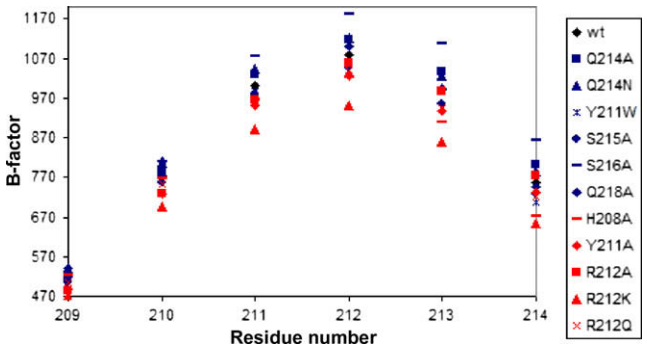

Fig. 6. Loop flexibility expressed as B-factor. The plot shows the average B-factor for a five residues window, with the position of the central residue in the *x*-axis. Blue symbols represent the loop mutants with an activity higher than 40% with respect to Mo-MLV wild type, and red symbols represent mutants with less than 40% activity of concerted integration.

Discussion

In this study, we carried out mutagenesis studies on the region 207–218 of Mo-MLV IN and analyzed the *in vitro* activity. The two-end concerted integration requires the formation of an IN tetramer stably associated with a pair of viral DNA ends [27]. This assay allowed us to determine whether mutations in the flexible loop could have an effect on protein oligomerization to form a productive com-

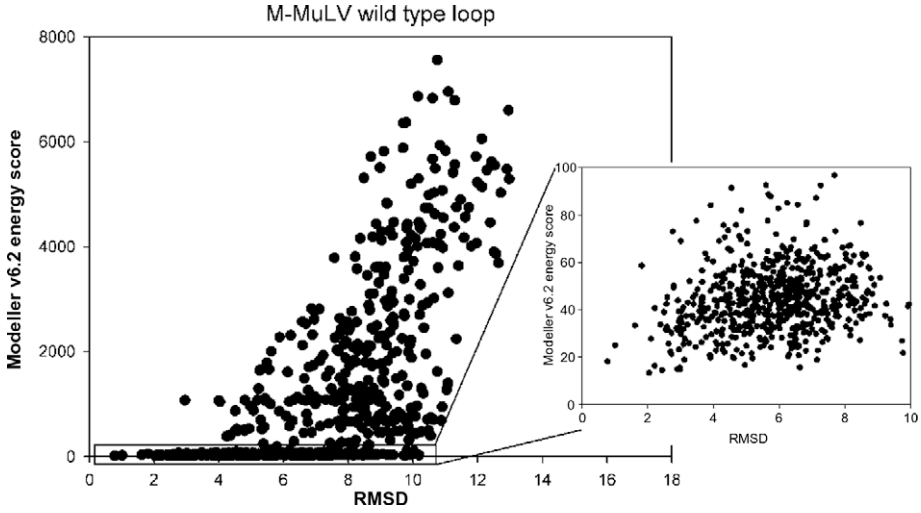

Fig. 5. Energy values of a thousand different conformations of Mo-MLV wt loop (residues 207–216). Each loop conformation has a MODELLER v6.2 energy value plotted against its structural similarity (RMSD) with respect to minimum energy loop coordinates. The loops with similar low energy values and different RMSD represent flexible regions and higher energy values mean less probable loop configurations. The zoom window shows the representation of different loop configurations (RMSD) with similar low energy scores, selected in this study as the allowed loop conformations.

plex. Mutation of H208 that corresponds to G140 in HIV-1 IN (Fig. 1) impairs 3′-processing, disintegration and concerted integration. For HIV-1 IN it was reported that G140 and G149, which are located at the ends of the loop, act as a hinge. Mutations of these amino acids reduced the flexibility of the loop and decreased catalytic activity. Our results suggest that for Mo-MLV, H208 could also act as a hinge, however, since catalytic activity was completely abolished we suggest that this residue has an essential role in catalytic activity.

Mutations within the loop may produce local effects on protein conformation depending on the side chains and their interaction with neighboring amino acids. Substitution Y211A decreased concerted integration more than 50-fold. However, the fact that Y211W only decreased activity by two-fold suggests that hydrophobicity of tyrosine in the region 207–218 is important to maintain enzyme activity. Modeling studies in HIV-1 IN have shown that Y143, which is in the same position as Y211 of Mo-MLV IN, is located in a hydrophobic pocket and is essential for activity [7]. Three mutants of R212 caused a significant decrease of integration, suggesting that the side chain of this amino acid is important for assembly of the active DNA–protein complex. Glutamine at position 214 of Mo-MLV is highly conserved in retroviral integrases. We found that Q214A decreased concerted integration; however, Q214N was 80% active. Concerted integration was not affected when S215 or S216 were substituted by alanine. For HIV-1 IN Q148, which corresponds to S215 in Mo-MLV, a model has been proposed suggesting that Q148 could be involved in hydrogen bonding with the 5′ single strand tail of the LTR [11]. In that study Q148A was not active in strand transfer in contrast to Q148N. It is interesting to note that processing was not affected in Q148A and Q148N. Furthermore, in the mutant Q148H [28] 3′-processing and strand transfer were severely reduced, although strand transfer was not abolished as in Q148A [11]. The decrease in 3′-processing could be explained by a steric effect of histidine since this residue is oriented towards the active site, although this needs to be investigated further.

The ratio between the concerted integration and the non-concerted integration products was approximately the same in all the tested mutants, suggesting that mutations affect the catalytic activity in both one and two site strand transfer. These results suggest that oligomerization to form a two site concerted integration complex is not being affected in these mutants. This is in agreement with protein crosslinking studies which indicate that in solution the mutants described in this study form the same oligomers as wt IN (D. Henriquez, results not published).

Optimal disintegration requires binding of the ss-tail of the non-processed strand of the viral DNA end to the enzyme [16]. This binding is essential to position the integration intermediate in the active site [12]. When we analyzed the disintegration activity of the mutant INs, substitution H208A blocked disintegration. This result again highlights the importance of this residue for IN activity.

The flexibility of $loop_{207-216}$ of Mo-MLV IN mutants was determined by alpha-carbon displacement by B-factor and compared with enzyme activity. Y211A was less flexible in the first half of the $loop_{207-216}$ between residues 210–214, while Y211W had a similar flexibility to wild type IN (Fig. 6). These results suggest that the reduction of flexibility of the $loop_{207-216}$ could decrease the activity of Y211A. Molecular dynamics simulation studies on HIV-1 integrase have suggested that Y143 interacts with a nearby amino acid residue of integrase causing a conformational change of this loop, allowing viral DNA to get closer to the catalytic site [29]. The three mutants of R212 decreased the flexibility of the $loop_{207-216}$, in agreement with the reduction of the activity. Atomic displacement analysis showed that the flexibility of $loop_{207-216}$ for substitution S215A was similar to wild type IN. However, the loop substitution S216A was more flexible in our model (Fig. 6). Therefore, our results suggest that S215A and S216A are active IN probably because they increase or maintain the flexibility of this loop. Our results show that, except for Q214, substitutions that decreased the flexibility of the $loop_{207-216}$ of Mo-MLV IN also decreased both concerted and non-concerted integration. Both the flexibility of the $loop_{207-216}$ and enzyme activity of the mutants Y211W, S215A and S216A were similar to wild type IN. Furthermore, the mutant C209A, which is active for concerted integration [30], has a similar flexibility to wild type IN (data not shown).

Our results demonstrate that the region 207–218 of the integrase of Mo-MLV is involved in the catalytic activity. Substitutions of the residues H208, Y211, R212 and Q214 in this region have a significant effect on the catalytic activity. Taken together, our results indicate that, similar to HIV-1 and ASV IN, Mo-MLV IN has a flexible loop near the catalytic site ($loop_{207-216}$) and the amino acid residues H208, Y211, R212, and Q214 of the flexible loop of Mo-MLV IN play a role in the correct assembly of the protein–nucleic acid complex involved in integration.

The dynamic nature of IN and the different conformations that it presents between 3′-processing and integration has been analyzed [31,32] and different orientations between IN and substrate were suggested in both assays. The flexible loop of Mo-MLV IN may not be involved in 3′-processing, but could be important in integration, because a conformational change of the enzyme may occur after 3′-processing facilitating protein–protein interactions needed for assembly of the integration complex. The flexibility of the loop would be fundamental in this conformational change, which is supported by our results of activity and flexibility, and other reports [32]. The lack of correlation between the loop flexibility and both processing and disintegration can be explained by the fact that an active oligomer is not required for these reactions [33,34].

It has been suggested that the flexible loop could mediate the separation of the LTR tails [6]. This function would require a high flexibility of the loop to allow conformational changes to accommodate the processed LTR. This conformational change would cause the two single strand tails to become closer to the active site, stabilizing a symmetric DNA–protein complex for correct integration.

Acknowledgments

This work was funded by grant Fondecyt 1040409.
We thank Dr. Monica J. Roth for helpful discussions.

References

[1] S.P. Goff, Nat. Rev. Microbiol. 5 (2007) 253–263.
[2] A.P. Eijkelenboom, F.M. van den Ent, A. Vos, J.F. Doreleijers, K. Hård, T.D. Tullius, R.H. Plasterk, R. Kaptein, R. Boelens, Curr. Biol. 7 (1997) 739–746.
[3] M. Cai, R. Zheng, M. Caffrey, R. Craigie, G.M. Clore, A.M. Gronenborn, Nat. Struct. Biol. 4 (1997) 567–577.
[4] P.J. Lodi, J.A. Ernst, J. Kuszewski, A.B. Hickman, A. Engelman, R. Craigie, G.M. Clore, A.M. Gronenborn, Biochemistry 34 (1995) 9826–9833.
[5] D. Esposito, R. Craigie, Adv. Virus Res. 52 (1999) 319–333.
[6] J. Wielens, I.T. Crosby, D.K. Chalmers, J. Comput. Aided Mol. Des. 19 (2005) 301–317.
[7] X. Chen, M. Tsiang, F. Yu, M. Hung, G.S. Jones, A. Zeynalzadegan, X. Qi, H. Jin, C.U. Kim, S. Swaminathan, J.M. Chen, J. Mol. Biol. 380 (2008) 504–519.
[8] T.S. Heuer, P.O. Brown, Biochemistry 36 (1997) 10655–10665.
[9] T.M. Jenkins, D. Esposito, A. Engelman, R. Craigie, EMBO J. 16 (1997) 6849–6859.
[10] D. Esposito, R. Craigie, EMBO J. 17 (1998) 5832–5843.
[11] A.A. Johnson, W. Santos, G.C.G. Pais, C. Marchand, R. Amin, T.R. Burke, G. Verdine, Y. Pommier, J. Biol. Chem. 281 (2006) 461–467.
[12] J. Vera, B. Valenzuela, M.J. Roth, O. Leon, Biol. Res. 41 (2008) 69–80.
[13] C.B. Jonsson, G.A. Donzella, M.J. Roth, J. Biol. Chem. 268 (1993) 1462–1469.
[14] J.K. Ko, J. Ma, Am. J. Physiol. Cell Physiol. 288 (2005) CI273–CI278.
[15] R.A. Villanueva, C.B. Jonsson, J. Jones, M.M. Georgiadis, M.J. Roth, Virology 316 (2003) 146–160.
[16] G.A. Donzella, O. Leon, M.J. Roth, J. Virol. 72 (1998) 1691–1698.
[17] F. Yang, M.J. Roth, L. Virol. 75 (2001) 9561–9570.
[18] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779–815.
[19] S.R. Eddy, Bioinformatics 14 (1998) 755–763.
[20] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, et al., Nucleic Acids Res. 36 (2008) D281–288.

[21] M.J. Sippl, Proteins 17 (1993) 355–362.
[22] M.Y. Shen, A. Sali, Protein Sci. 15 (2006) 2507–2524.
[23] W. Kabsch, C. Sander, Biopolymers 22 (1983) 2577–2637.
[24] A. Fiser, R.K. Do, A. Sali, Protein Sci. 9 (2000) 1753–1773.
[25] P. Lesbats, M. Métifiot, C. Calmels, S. Baranova, G. Nevinsky, M.L. Andreola, V. Parissi, Nucleic Acids Res. 36 (2008) 7043–7058.
[26] J.A. Zham, S. Bera, K.K. Pandey, A. Vora, K. Stillmock, D. Hazuda, D.P. Grandgenett, Antimicrob. Agents Chemother. 52 (2008) 3358–3368.
[27] M. Li, M. Mizuuchi, T.R. Burke Jr., R. Craigie, EMBO J. 25 (2006) 1295–1305.
[28] O. Delelis, I. Maler, L. Na, L. Tchertanov, V. Calvez, A.G. Marcelin, F. Subra, E. Deprez, J.F. Mouscadet, Nucleic Acids Res. 37 (2009) 1193–1201.
[29] L. De Luca, G. Vistoli, A. Pedretti, M.L. Barreca, A. Chimirri, Biochem. Biophys. Res. Commun. 336 (2005) 1010–1016.
[30] F. Yang, J.A. Seamon, M.J. Roth, Virology 291 (2001) 32–45.
[31] A. Alian, S.L. Griner, V. Chiang, M. Tsiang, G. Jones, G. Birkus, R. Geleziunas, A.D. Leavitt, R.M. Stroud, Proc. Natl. Acad. Sci. USA 106 (2009) 8192–8197.
[32] J. Greenwald, V. Le, S.C. Butler, F.D. Bushman, S. Choe, Biochemistry 38 (1999) 8892–8898.
[33] A. Faure, C. Calmels, C. Desjobert, M. Castroviejo, A. Caumont-Sarcos, L. Tarrago-Litvak, S. Litvak, V. Parissi, Nucleic Acids Res. 33 (2005) 977–986.
[34] A.B. Hickman, I. Palmer, A. Engelman, R. Craigie, P. Wingfield, J. Biol. Chem. 269 (1994) 29279–29287.