Published in final edited form as:

Proteins. 2017 April ; 85(4): 647–656. doi:10.1002/prot.25245.

X-ray Crystal Structure of the N-Terminal Region of Moloney Murine Leukemia Virus Integrase and its Implications for Viral DNA Recognition

Rongjin Guan^{1,2,3}, Sriram Aiyer⁴, Marie L. Cote⁵, Rong Xiao^{1,2,3}, Mei Jiang^{1,2,3}, Thomas B. Acton^{1,2,3}, Monica J. Roth^{4,*}, and Gaetano T. Montelione^{1,2,3,*}

¹Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway NJ 08854

²Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway NJ 08854

³Department of Biochemistry and Molecular Biology, Rutgers, The State University of New Jersey, Piscataway NJ 08854

⁴Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway NJ 08854

⁵Department of Biochemistry, Robert Wood Johnson Medical School, UMDNJ, Piscataway NJ 08854

Abstract

The retroviral integrase (IN) carries out the integration of a dsDNA copy of the viral genome into the host DNA, an essential step for viral replication. All IN proteins have three general domains, the N-terminal domain (NTD), the catalytic core domain (CCD), and the C-terminal domain (CTD). The NTD includes an HHCC zinc finger-like motif, which is conserved in all retroviral IN proteins. Two crystal structures of Moloney murine leukemia virus (M-MuLV) IN N-terminal region (NTR) constructs that both include an N-terminal extension domain (NED, residues 1-44) and an HHCC zinc-finger NTD (residues 45-105), in two crystal forms are reported. The structures of IN NTR constructs encoding residues 1-105 (NTR₁₋₁₀₅) and 8-105 (NTR₈₋₁₀₅) were determined at 2.7 and 2.15 Å resolution, respectively and belong to different space groups. While both crystal forms have similar protomer structures, NTR₁₋₁₀₅ packs as a dimer and NTR₈₋₁₀₅ packs as a tetramer in the asymmetric unit. The structure of the NED consists of three anti-parallel β -strands and an α -helix, similar to the NED of prototype foamy virus (PFV) IN. These three β -strands form an extended β -sheet with another β -strand in the HHCC Zn²⁺ binding domain, which is a unique structural feature for the M-MuLV IN. The HHCC Zn²⁺ binding domain structure is similar to that in HIV and PFV INs, with variations within the loop regions. Differences between the PFV and MLV IN NEDs localize at regions identified to interact with the PFV LTR and are compared with established biochemical and virological data for M-MuLV.

*To whom correspondence should be addressed: Monica J. Roth, Department of Pharmacology, Robert Wood Johnson Medical School-Rutgers University, 675 Hoes Lane West, Piscataway NJ 08854, roth@rwjms.rutgers.edu. Gaetano T. Montelione, Center for Advanced Biotechnology and Medicine, Rutgers – The State University of New Jersey, 679 Hoes Lane West, Piscataway NJ 08854, gtm@rutgers.edu.

Keywords

Retroviral integrase; N-terminal extension domain; Zn²⁺ finger protein; Homology modeling; N-terminal domain; DNA integration

Introduction

The integration of double-stranded retroviral DNA into the host chromatin is catalyzed by the viral encoded integrase (IN) protein. The IN protein is comprised of an N-terminal zinc binding domain (NTD), a catalytic core domain (CCD), and a C-terminal domain (CTD) (for a review see¹). Integration involves the coordinated assembly of the IN proteins, the two viral DNA termini, and the host target DNA into an intasome complex. Insights into the assembled intasome have been obtained from structural studies of the PFV IN^{2–5}, RSV IN⁶, MMTV IN⁷ and molecular modeling based upon these structures. Target DNA binding sites localize within the CCD and CTD^{4,8–10}, whereas viral long-terminal repeat (LTR) DNA recognition localizes to all three domains, including the N-terminal region (residues 1–105)^{2,4,6,7}.

The IN NTD contains two conserved histidine and two conserved cysteine residues (HHCC), which are required for Zn²⁺ binding¹¹. In HIV and RSV IN proteins, the NTD containing this invariable HHCC motif is ~ 50 residues in length^{12–14}. However, M-MuLV IN has an additional 44 residues preceding the HHCC Zn-finger subdomain^{15,16}, with 19% sequence similarity (ClustalW) with the NED found in the prototype foamy virus (PFV) IN². We refer to this segment of NED followed by NTD, of M-MuLV, PFV, and related INs, as the N-terminal Region (NTR, residues 1–105) of IN. In PFV IN, the NTR consists of the N-terminal extension domain (NED, residues 1–48) and the N-terminal domain (NTD, residues 51–105)².

The IN protein is expressed within the Gag-Pol precursor protein and is excised through proteolytic cleavage by the viral protease within the viral particle^{1,17}. Upon infection, the IN protein is associated with the reverse transcription complex (RTC), which transitions to the preintegration complex (PIC) upon formation of the newly generated double-stranded viral DNA^{1,18}. For M-MuLV, proteins associated with the PIC include the capsid (CA), p12, IN viral proteins and cellular proteins including barrier-to-autointegration factor 1 (BAF)^{18–22}. Ultimately, integration occurs in the nucleus through interaction of the M-MuLV IN protein with the cellular bromodomain and extra-terminal (BET) proteins^{23–26}. The conformational states of the IN, from the precursor state through the intasome complex, must therefore be highly dynamic.

When expressed as isolated domains, the NTD of HIV and NTR of M-MuLV IN form dimers in solution^{27–29}. The HIV IN protein, lacking an NED subdomain, dimerizes through its HHCC Zn-finger²⁷. However, the HHCC Zn-finger (residues 51–105) of M-MuLV IN alone is not sufficient for stable dimer formation²⁹. The NTR segment (i.e. NED plus NTD) of M-MuLV IN is capable of complementing IN bearing NTR deletions, as measured by *in vitro* IN activity assays^{28,29}. However, the HHCC Zn-finger alone is not sufficient for complementation²⁹. Therefore, the NED subdomain is required for M-MuLV IN activity²⁹.

Within the PFV intasome, the IN NTR is observed as a monomer; and dimers of PFV NTR are not observed within any available crystal structures^{2,4}.

To gain insight into structure-function relationships of the M-MuLV IN NTR, we determined its 3-dimensional (3D) structure. Here we report X-ray crystal structures of M-MuLV IN NTR crystallized in two crystal forms. Form I, a construct containing residues 1-105 (with a C-terminal His₆ tag), in space group R32, was determined to 2.7 Å resolution. Form II, a shorter construct consisting of residues 8-105, IN₈₋₁₀₅, in space group P2₁, was solved to 2.15 Å resolution. These structures reveal that key residues, shown to be important for strand transfer activity²⁹ contribute to an extensive basic surface. By analogy with the X-ray crystal structure of PFV IN bound to the viral LTR², this basic surface of M-MuLV appears to function in binding the LTR DNA, a key step in the mechanism of IN integration. Overall, the structural and subdomain organization of the M-MuLV IN NTR is similar to that of PFV IN NTR, with some key structural differences that include regions of the PFV NTR that are important in binding the LTR in the intasome complex². The key LTR DNA binding residues of PFV IN NTR are biophysically quite different than the corresponding residues in the M-MuLV IN NTR structure. This suggests that the details of the recognition of the M-MuLV IN for its cognate LTR DNA sequences are distinct from that of the PFV IN:LTR DNA interaction.

Materials and Methods

Protein expression and purification

A DNA fragment encoding M-MuLV IN NTR₁₋₁₀₅ (residues 1-105) was generated by PCR from the full-length gene and cloned into the bacterial expression vector pET21_NESG³⁰, with a short C-terminal purification tag “LEHHHHHH”. The plasmid was then transformed into *E. coli* BL21(DE3) cells (Stratagene). These cells were grown in either minimal media containing selenium-methionine or LB media (1L) at 37 °C to O.D.₆₀₀ of ~0.8 units, and induced with 1 mM isopropyl-β-D-thiogalactoside overnight at 17 °C. The bacteria were pelleted by centrifugation, and resuspended in 1x PBS buffer by mild sonication to release the soluble target protein. After high-speed centrifugation, the supernatant was applied to a 5 ml His-tag affinity column (GE Healthcare), and eluted with a linear (50–500 mM) imidazole gradient. Further purification was accomplished by size exclusion chromatography using a HighLoad 26/60 Superdex S75 column (GE Healthcare). The purified protein was over 95% pure based on SDS PAGE, and was also validated by MALDI-TOF mass spectrometry (13.6 kDa).

In an effort to optimize crystallization, a second IN NTR construct (IN NTR₈₋₁₀₅, residues 8-105) was generated by PCR and expressed in pET15-TEV-NESG^{30,31}, encoding a TEV protease cleavage site after the N-terminal His-tag. The procedures used for the expression and purification of IN NTR₈₋₁₀₅ are similar to the procedures described above, with the additional step of treating with TEV protease followed by passing the cleaved product through a His-tag affinity column (GE Healthcare), and a final gel filtration purification using a Superdex S75 column, as described elsewhere^{30,31}.

Crystallization and data collection

The purified M-MuLV IN NTR₁₋₁₀₅ was concentrated to 10 mg/ml in 0.05 M MES at pH 6.5 and stored at -80°C prior to crystallization. The initial crystallization screening was performed at the high-throughput screening (HTS) facility at Hauptman-Woodward Research Institute (HWI) located in Buffalo, NY, where 1536 crystallization conditions were screened using the microbatch method³². Initial crystallization hits were further optimized manually to obtain diffraction quality crystals. The addition of detergents was important in improving the diffraction quality of the crystals. Optimal conditions for crystallization were obtained at room temperature in 2.0 M sodium malonate, 0.1 M sodium acetate at pH 5.0, in the presence of 0.05% Anapoe X-305.

Diffraction of M-MuLV IN NTR₁₋₁₀₅ crystals was first tested using a Rigaku R-Axis IV⁺⁺ detector. The crystals were harvested directly from the drops and flash-frozen in liquid nitrogen. A complete three-wavelength multiple anomalous dispersion (MAD) data set to 2.7 Å was then collected at beamline X4A of the National Synchrotron Light Source (NSLS), the Brookhaven National Laboratory (BNL). The HKL2000 package³³ was used to index, integrate and scale the data (Table 1).

M-MuLV IN NTR₈₋₁₀₅ was concentrated to 5 mg/ml in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM DTT, 0.02% NaN₃, and screened at HWI for initial crystallization screening, as described above³², followed by manual optimization. The best crystals of IN₈₋₁₀₅ NTR were grown in 1.0 M K₂HPO₄, 0.1 M sodium acetate at pH 4.6, in the presence of 0.05% Anapoe X-305. The crystals were briefly soaked in cryo buffer containing mother liquid plus 15% glycerol and flash frozen in liquid N₂. Preliminary diffraction analysis was carried out with a home X-ray source, and provided diffraction to 2.5 Å resolution. Higher resolution data to 2.15 Å were collected at NSLS beamline X29A at BNL using their mail-in data collection service³⁴. The data processing was performed by HKL2000³³. The resulting crystallographic statistics are listed in Table 1.

Structure determination and refinement

The crystal structure of M-MuLV IN NTR₁₋₁₀₅ was determined by MAD methods using the program *Solve*³⁵. Density modification and automated model building were performed with *Resolve*³⁶. Initially, approximately 50% of the residues could be automatically built by *Resolve*. Significant effort was spent on manual tracing and model building, and the model was gradually completed. Structure refinement was carried out with program *PHENIX*³⁷, and the final refined structure was validated by *MOLPROBITY*³⁸. Manual model building and adjustment were performed using *COOT*³⁹. The final structure refinement and geometry statistics are summarized in Table 1.

The crystal structure of IN NTR₈₋₁₀₅ was determined by molecular replacement with program *Phase*⁴⁰, using the structure of IN NTR₁₋₁₀₅ as a search model. Structure refinement, model building and validation were all performed as detailed above. The final structure was optimized with *PDB_REDO*⁴¹, which improved the model quality. The statistics for the final structure refinement and model geometry are summarized in Table 1.

Interface residues and buried surface areas for the two crystal structures were computed using the EBI-PISA server⁴².

Results and Discussion

Overall structure of M-MuLV IN-NTR₁₋₁₀₅

The crystals of M-MuLV IN NTR₁₋₁₀₅, belong to space group R32, with two M-MuLV IN NTR₁₋₁₀₅ molecules in the asymmetric unit (Fig. 1A). The structure was determined by MAD to 2.7 Å resolution. The dimer interface is relatively small, burying only 378 Å² of surface, and mainly includes the following key interacting residues of T60, L62, S63, K66, A69, L70, R73, Q99, V100, N101, A102 and S103. The first 10 amino acid residues in both molecules are missing due to lack of electron densities, suggesting that these first 10 residues are highly flexible. The M-MuLV protease cleavage recognition site for Gag-Pol polyprotein processing resides in these initial 5 IN residues¹⁷. The C-terminal tag (LEHHHHHH) was also not detected at this resolution. The two molecules in the asymmetric unit are very similar, having a backbone C^α RMSD of 0.008 Å for residues 12-104. Each molecule contains both the NED and NTD HHCC Zn²⁺-finger domains (Fig. 1B). The ion Zn²⁺ is coordinated by residues H57, H61, C94 and C97 (Fig. 1C).

The NED is composed of three short anti-parallel β-strands and one α-helix (Fig. 2A). Interestingly, the β3 strand of the NED interacts with the β4 strand in the α3-α4 loop of the HHCC Zn²⁺ binding domain, forming an extended β-sheet (Fig. 1B), which stabilizes the orientations of the NED and NTD with respect to one another. Similar to other IN proteins, the M-MuLV IN NTD HHCC Zn²⁺-finger domain is composed of 3 core helices (α2, α3 and α4), and a C-terminal α-helix α5 (Fig. 2A). Additionally, there is a short β-strand within the α3-α4 loop, β4, previously not identified in other IN proteins (Fig. 2A). Sequence alignment of the M-MuLV NTD with that of PFV, HIV-1, HIV-2, and maedi-visna virus (MVV) (Fig. 2B) indicates that the α3-α4 loop in M-MuLV IN (residues 75-82) is longer than that in all other IN proteins, allowing for formation of the β-sheet that bridges the NED and NTD regions.

Overall structure of M-MuLV IN₈₋₁₀₅

Treatment of the M-MuLV IN₈₋₁₀₅ construct with TEV removes the N-terminal His₆ tag and results with the N-terminal Ser-His sequence upstream of the IN₈₋₁₀₅ sequence. The resulting IN₈₋₁₀₅ construct was crystallized, and its structure determined to 2.15 Å resolution (Fig. 3A, B). These crystals belong to space group P2₁, where there are four molecules in the asymmetric unit. Although M-MuLV IN NTR₈₋₁₀₅ is a shorter protein (lacking both the His tag and seven N-terminal IN residues), crystallized under different conditions, and belongs to a different space group compared to M-MuLV IN NTR₁₋₁₀₅, the structures of the individual protomers are nearly identical.

Although these two structures were from different space groups, the molecular packing in the respective crystals is very similar. In the M-MuLV IN NTR₈₋₁₀₅ structure, the 4 molecules form a butterfly shaped symmetric “dimer-of-dimers”. A total of six interacting interfaces are present, burying a combined area of 1996.8 Å². The interfaces are stabilized

by 1 salt bridge and 11 hydrogen bonded interactions. These include the salt bridge between D54 (chain A) and K95 (chain D) and the following hydrogen bonded interactions – Q58 (chain C) and Q58 (chain B), Q58 (chain C) and H61 (chain B), K66 (chain B) and N101 (chain A), Q99 (chain B) and R73 (chain A), N101 (chain B) and K66 (chain A), S103 (chain B) and R73 (chain A), H75 (chain C) and Y78 (chain A), Y78 (chain C) and H75 (chain A), Y78 (chain D) and H75 (chain B) and two hydrogen bonds between A102 (chain D) and R73 (chain C). Additionally, in the structure of the M-MuLV IN NTR₁₋₁₀₅, the 4 molecules that interact to form the butterfly shaped “dimer-of-dimers”, similar to that of the M-MuLV IN NTR₈₋₁₀₅ structure, can be identified with molecules from the neighboring asymmetric unit. Fig. 3B highlights the overlap of the dimer identified for M-MuLV IN NTR₁₋₁₀₅ with the M-MuLV IN NTR₈₋₁₀₅ dimer-of-dimers. Superimposition of the corresponding “dimers” of the two structures has a C α RMSD of 0.89 Å. Preparative gel filtration demonstrates that M-MuLV IN NTR₁₋₁₀₅^{28,29} forms dimers in solution, with no identifiable tetramers. The shorter M-MuLV IN NTR₈₋₁₀₅ construct also forms dimers with little evidence for oligomerization, based on analytical gel filtration with static light scattering detection (AGF-MALS)³¹ (Fig. S1). Therefore, the butterfly shaped tetramer of M-MuLV IN NTR₈₋₁₀₅ appears to be stabilized by crystal packing interactions. In contrast, the two molecules in the asymmetric unit of the M-MuLV IN NTR₁₋₁₀₅ structure do not have extensive intermolecular interactions. It is important to recognize that dimers (or ‘dimers of dimers’) observed in crystal lattice environments are not necessarily biologically relevant. Without further evidence, we cannot conclude that the dimers we observed in these crystal structures represent the dimer(s) that form in solution, or that these dimers are relevant to structures that are formed by full-length IN in its biological contexts. Moreover, mutating the interface residues in the crystal structure, K66 or R73, did not disrupt dimer formation in gel filtration experiments²⁹.

Structural and functional comparison with PFV IN-NTD and HIV IN-NTDs

Fig. 4 shows the structural comparison of NTRs from M-MuLV and PFV, and NTDs from HIV-1, HIV-2 and MVV. HIV-1, HIV-2, and MVV INs lack the NED N-terminal extension subdomain that is present in the M-MuLV and PFV INs. In all five structures, the overall fold of the HHCC Zn-finger NTD domains, and positioning of the zinc ion are similar. HIV-1 and HIV-2 INs lack the C-terminal helix α 5 in their isolated NTDs. The NTD structures vary primarily in the sizes of the surface loops. Within the NTD, the α 2 region of PFV is shorter, providing a larger loop between α 2 and α 3. Similarly, the loops between α 3 and α 4 are quite distinct. All three viral NTDs that consist solely of the Zn-finger subdomain contain an extremely short loop consisting of 4 residues between the α 3 and α 4 helices (Fig. 2B). In contrast this region is greatly expanded in both virus INs encoding NED regions, with M-MuLV IN containing the largest loop, providing the additional β -strand that forms the β -sheet with the NED (Fig. 2B).

Although the PFV and M-MuLV NEDs are structurally similar to one another, and share an overall similar fold, overlay of the two proteins show differences in the relative orientation of the NED with respect to the NTD (Fig. 4A, B and Fig. S2, S3). Significantly, the formation of the extended β -sheet in M-MuLV IN defines the orientation of the NED with respect to the NTD.

The PFV IN has been shown to interact with viral DNA using basic residues in two loop regions of the NED (i.e., residue K26 between $\alpha 1$ and $\beta 1$, and residue R41 between $\beta 2$ and $\beta 3$)². Sequence alignment (Fig. 2A) shows that both of these loop regions, including these basic residues (indicated by green dots on the PFV IN sequence) are deleted in the M-MuLV NED. This implies either that the M-MuLV NED does not form analogous close contact interactions with the viral LTR, or that the M-MuLV viral LTR forms an altered or bent structure, possibly associated with the highly conserved GGGG sequence at positions 10–13 nucleotides upstream of the viral termini that undergoes strand transfer^{43–45}.

Additional positions within the NTR of PFV IN have been identified to be in contact with DNA. These include residues G68, R69, N84, R86, N106, and K107². Of these positions, only residue N84 is conserved in M-MuLV IN. Previous mutational analysis of M-MuLV IN NTR of residues potentially involved in DNA interactions focused on basic residues conserved in related gammaretroviruses including MCF, Friend MLV, AKV, FeLV and RaLV²⁹. From these studies, mutations within the NED including K24A, K33A/K34A, and K41A did not alter *in vitro* integration activities, whereas mutation within the NTD including K66A/K68A and R73A dramatically decreased both viral titer and *in vitro* 3' cleavage and strand transfer. The R83A mutant complements CCD/CTD constructs for strand transfer but not for 3' processing²⁹. These basic residues are highlighted in red in Fig. 5A. The surface electrostatic potential distribution on the surface of the M-MuLV NTR (Fig. 5B) further suggests that these basic residues within the NTR form a DNA-binding cluster on one face of the domain (Fig. 5C, D). A sequence alignment of critical regions of M-MuLV and PFV IN (Fig. 6A) shows key basic residues of M-MuLV IN (marked in magenta and underlined), which when mutated affect strand transfer *in vitro*²⁹, along with analogous residues of PFV IN (marked in red) that interact with the LTR in the crystal structure of the PFV IN – LTR complex². These key residues of M-MuLV and PFV INs are in similar positions in superimposed structures of the corresponding NTRs.

These conserved basic residues of the M-MuLV NTR also form contacts with DNA in a simple homology model of the M-MuLV IN-LTR DNA complex (Fig. 6B), which was constructed based on the crystal structure of the PFV IN-LTR DNA complex². Significantly, the positioning of M-MuLV IN residues K66/K68 in this simple model places them in close proximity with the viral LTR, particularly residue K68 which interacts at sequence positions 9–10 bases from the termini of the LTR that undergoes strand transfer. In this model, the sequences within the NED are held in position by the novel β -sheet formed between NED and NTD, which results in a relatively constrained conformation when compared with the PFV NTR. Disruption of this β -sheet would allow conformational flexibility in the relative orientation of the NED and NTD, potentially enabling optimal interaction with the viral DNA. Mutational analysis has indicated that the basic residues within the NED are not essential for virus viability (reference²⁹ and data not shown). Within the PFV NTR, the NED does not provide base-specific contacts with the viral LTR but rather only enables stabilization of the interaction through contacts with the phosphodiester backbone². Consequently, this model does not rule out an additional role for the NED in interacting with host proteins.

The role of the NED of the NTR of M-MuLV and PFV INs has been postulated to improve stabilization of the viral LTR:IN interaction². For HIV-1 IN, enhanced activity and solubility was observed when fused with the non-specific DNA binding domain (Sso7d) as a surrogate for NED⁴⁶. However, this enhanced activity was not affected by substituting amino acids in Sso7d that contribute to non-specific DNA binding. This suggests that the Sso7d fusion is blocking formation of a non-productive complex, rather than directly aiding in the engagement of viral LTR. Interestingly, the fusion of the PFV NED to the HIV IN NTD does not yield a highly active protein⁴⁶.

Our analysis of these X-ray crystal structures together with published mutagenesis and virology studies^{28,29} suggest that specific structural features of the M-MuLV IN NED can provide additional stabilizing contacts in IN-LTR complex; at least some key determinants for conferring LTR specificity must reside in the NED. Moreover, the NEDs appear to have evolved to specifically function only in the context of their respective retroviral INs.

Data deposition

Coordinates and structure factors for these M-MuLV IN NTR₁₋₁₀₅ and IN NTR₈₋₁₀₅ X-ray crystal structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID: 3NNQ and 4NZG respectively). Note that the residue numbering in these PDB entries is shifted by +1 residue relative to the literature numbering used in this manuscript.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the staff of the Hauptman-Woodward Research Institute high-throughput crystallization screening facility for initial crystallization screening. X-ray diffraction data for M-MuLV IN NTR₁₋₁₀₅ were measured on beamline X4A of the National Synchrotron Light Source (NSLS). We thank John Schwanof and Randy Abramowitz for their help in data collection. X-ray diffraction data for M-MuLV IN NTR₈₋₁₀₅ were measured on NSLS beamline X29A where the financial support comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy and from the National Center for Research Resources of the National Institutes of Health, and we thank Dr Howard Robinson for his help with data collection. G.T. Montelione and R. Xiao are associated with Nexomics Biosciences, Inc., a structural biology consulting company. Since completing this work, T. Acton has moved to Evotech, Inc. This work was supported as a Community Outreach activity of the NIGMS Protein Structure Initiative grant U54 GM094597 (GTM) and by grants R01GM070837 and R01GM110639 (MR).

References

1. Craigie R, Bushman FD. HIV DNA Integration. Cold Spring Harb Perspect Med. 2012; 2(7):a006890. [PubMed: 22762018]
2. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature. 2010; 464(7286):232–236. [PubMed: 20118915]
3. Hare S, Maertens GN, Cherepanov P. 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. EMBO J. 2012; 31(13):3020–3028. [PubMed: 22580823]
4. Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature. 2010; 468:326–329. [PubMed: 21068843]
5. Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, Lindemann D, Engelman AN, Costa A, Cherepanov P. Structural basis for retroviral integration into nucleosomes. Nature. 2015; 523(7560):366–369. [PubMed: 26061770]

6. Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, Aihara H. Crystal structure of the Rous sarcoma virus intasome. *Nature*. 2016; 530(7590):362–366. [PubMed: 26887497]
7. Ballandras-Colas A, Brown M, Cook NJ, Dewdney TG, Demeler B, Cherepanov P, Lyumkis D, Engelman AN. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature*. 2016; 530(7590):358–361. [PubMed: 26887496]
8. Aiyer S, Rossi P, Malani N, Schneider WM, Chandar A, Bushman FD, Montelione GT, Roth MJ. Structural and sequencing analysis of local target DNA recognition by MLV integrase. *Nucleic Acids Res*. 2015; 43(11):5647–5663. [PubMed: 25969444]
9. Nowak MG, Sudol M, Lee NE Jr, WMK, Katzman M. Identifying amino acid residues that contribute to the cellular-DNA binding site on retroviral integrase. *Virology*. 2009; 389:141–148. [PubMed: 19447461]
10. Serrao E, Ballandras-Colas A, Cherepanov P, Maertens GN, Engelman AN. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology*. 2015; 12(1):39. [PubMed: 25924943]
11. Johnson MS, McClure MA, Feng DF, Gray J, Doolittle RF. Computer analysis of retroviral pol genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc Natl Acad Sci USA*. 1986; 83:7648–7652. [PubMed: 2429313]
12. Eijkelenboom APAA, Ent FMIvd, Vos A, Doreleijers JF, Hard K, Tullius TD, Plasterk RHA, Kaptein R, Boelens R. The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Current Biology*. 1997; 7(10):739–746. [PubMed: 9368756]
13. Wang J-Y, Ling H, Yang W, Craigie R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J*. 2001; 20:7333–7343. [PubMed: 11743009]
14. Bojja RS, Andrade MD, Merkel G, Weigand S, Dunbrack RL Jr, Skalka AM. Architecture and assembly of HIV integrase multimers in the absence of DNA substrates. *J Biol Chem*. 2013; 288(10):7373–7386. [PubMed: 23322775]
15. Jonsson CB, Donzella GA, Gaucan E, Smith CM, Roth MJ. Functional domains of Moloney murine leukemia virus integrase defined by mutation and complementation analysis. *J Virol*. 1996; 70(7):4585–4597. [PubMed: 8676485]
16. Jonsson CB, Roth MJ. Role of the His-Cys finger of Moloney murine leukemia virus integrase protein in integration and disintegration. *J Virol*. 1993; 67(9):5562–5571. [PubMed: 8350412]
17. Menendez-Arias L, Gotte D, Oroszlan S. Moloney murine leukemia virus protease: bacterial expression and characterization of the purified enzyme. *Virology*. 1993; 196(2):557–563. [PubMed: 8372434]
18. Fassati A, Goff SP. Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. *J Virol*. 1999; 73(11):8919–8925. [PubMed: 10515996]
19. Bowerman B, Brown PO, Bishop JM, Varmus HE. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes & Development*. 1989; 3:469–478. [PubMed: 2721960]
20. Lee MS, Craigie R. A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci USA*. 1998; 95:1528–1533. [PubMed: 9465049]
21. Prizan-Ravid A, Elis E, Laham-Karam N, Selig S, Ehrlich M, Bacharach E. The Gag cleavage product, p12, is a functional constituent of the murine leukemia virus pre-integration complex. *PLoS Pathogens*. 2010; 6(11):e1001183. [PubMed: 21085616]
22. Suzuki Y, Craigie R. Regulatory mechanisms by which barrier-to-autointegration factor blocks autointegration and stimulates intermolecular integration of Moloney murine leukemia virus preintegration complexes. *J Virol*. 2002; 76:12376–12380. [PubMed: 12414981]
23. De Rijck J, de Kogel C, Demeulemeester J, Vets S, El Ashkar S, Malani N, Bushman FD, Landuyt B, Husson SJ, Busschots K, Gijssbers R, Debyser Z. The BET family of proteins targets Moloney murine leukemia virus integration near transcription start sites. *Cell Rep*. 2013; 5(4):886–894. [PubMed: 24183673]
24. Aiyer S, Swapna GVT, Malani N, Aramini JM, Schneider WM, Plumb MR, Ghanem M, Larue RC, Sharma A, Studamire B, Kvaratskhelia M, Bushman FD, Montelione GT, Roth MJ. Altering

- murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res.* 2014; 42(9):5917–5928. [PubMed: 24623816]
25. Gupta SS, Maetzig T, Maertens GN, Sharif A, Rothe M, Weidner-Glunde M, Galla M, Schambach A, Cherepanov P, Schulz TF. Bromo and ET domain (BET) chromatin regulators serve as co-factors for murine leukemia virus integration. *J Virol.* 2013; 87(23):12721–12736. [PubMed: 24049186]
 26. Sharma A, Larue RC, Plumb MR, Malani N, Male F, Slaughter A, Kessl JJ, Shkriabai N, Coward E, Aiyer SS, Green PL, Wu L, Roth MJ, Bushman FD, Kvaratskhelia M. BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci USA.* 2013; 110(29):12036–12041. [PubMed: 23818621]
 27. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol.* 1997; 4(7):567–577. [PubMed: 9228950]
 28. Yang F, Leon O, Greenfield NJ, Roth MJ. Functional interactions of the HHCC domain of Moloney murine leukemia virus integrase revealed by non-overlapping complementation and zinc dependent dimerization. *J Virol.* 1999; 73:1809–1817. [PubMed: 9971758]
 29. Yang F, Seamon JA, Roth MJ. Mutational analysis of the N-terminus of Moloney murine leukemia virus integrase. *Virology.* 2001; 291:32–45. [PubMed: 11878874]
 30. Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccocanti C, Conover K, Everett J, Hamilton K, Huang YJ, Janjua H, Kornhaber G, Lau J, Lee DY, Liu G, Maglaqui M, Ma L, Mao L, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GV, Tang Y, Tong S, Wang D, Wang H, Zhao L, Montelione GT. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.* 2011; 493:21–60. [PubMed: 21371586]
 31. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccocanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma LC, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GV, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J Struct Biol.* 2010; 172(1):21–33. [PubMed: 20688167]
 32. Luft JR, Collins RJ, Fehrman NA, Lauricella AM, Veatch CK, DeTitta GT. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol.* 2003; 142(1):170–179. [PubMed: 12718929]
 33. Otwinowski, Z., Minor, W. *Methods Enzymol.* Vol. 276. Academic Press; 1997. Processing of X-ray diffraction data collected in oscillation mode; p. 307–326.
 34. Robinson H, Soares AS, Becker M, Sweet R, Heroux A. Mail-in crystallography program at Brookhaven National Laboratory's National Synchrotron Light Source. *Acta Crystallogr D Biol Crystallogr.* 2006; 62(Pt 11):1336–1339. [PubMed: 17057336]
 35. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr.* 1999; 55(Pt 4):849–861. [PubMed: 10089316]
 36. Terwilliger TC. Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr.* 2000; 56(Pt 8):965–972. [PubMed: 10944333]
 37. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66(Pt 2):213–221. [PubMed: 20124702]
 38. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins.* 2003; 50(3):437–450. [PubMed: 12557186]
 39. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010; 66(Pt 4):486–501. [PubMed: 20383002]
 40. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40(Pt 4):658–674. [PubMed: 19461840]

41. Joosten RP, Joosten K, Murshudov GN, Perrakis A. PDB_REDO: constructive validation, more than just looking for errors. *Acta Crystallogr D Biol Crystallogr*. 2012; 68:484–496. [PubMed: 22505269]
42. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007; 372(3):774–797. [PubMed: 17681537]
43. Donzella GA, Jonsson CB, Roth MJ. Influence of substrate structure on disintegration activity of Moloney murine leukemia virus integrase. *J Virol*. 1993; 67(12):7077–7087. [PubMed: 8230431]
44. O'Neill RR, Buckler CE, Theodore TS, Martin MA, Repaske R. Envelope and long terminal repeat sequences of a cloned infectious NZB xenotropic murine leukemia virus. *J Virol*. 1985; 53(1):100–106. [PubMed: 2981327]
45. Golemis EA, Speck NA, Hopkins N. Alignment of U3 region sequences of mammalian type C viruses: identification of highly conserved motifs and implications for enhancer design. *J Virol*. 1990; 64(2):534–542. [PubMed: 2153223]
46. Li M, Jurado KA, Lin S, Engelman A, Craigie R. Engineered hyperactive integrase for concerted HIV-1 DNA integration. *PLoS One*. 2014; 9(8):e105078. [PubMed: 25119883]
47. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*. 1999; 15(4):305–308. [PubMed: 10320398]
48. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.2r3pre. 2010.
49. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010; 38(Web Server issue):W545–549. [PubMed: 20457744]
50. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*. 2011; 67(Pt 4):235–242. [PubMed: 21460441]

Figure 1. Overall crystal structure of M-MuLV IN NTR₁₋₁₀₅

(A) The two M-MuLV IN NTD₁₋₁₀₅ molecules in the asymmetric unit. The two Zn²⁺ ions are shown as brown spheres and labeled. N- and C-termini in both molecules are indicated with "N" and "C", respectively. (B) Each M-MuLV IN NTR₁₋₁₀₅ has two regions, the NED and NTD subdomain. There are 3 β-strands and one α-helix in the NED subdomain, and four α-helices and 1 short β-strand in the NTD subdomain. (C) The Zn²⁺ binding site in the NTD subdomain, which consists of residues H57, H61, C94 and C97.

Figure 2. Sequence and structural alignment of M-MuLV (MMLV)

(A) Structure-based sequence alignment of M-MuLV and PFV IN-NTR. The NED is defined as residues 1-48 (blue line), and the NTD as residues 51-102 (green line) of PFV². The 8 residues in PFV IN NTR marked with green dots are those contacting DNA directly in the complex formed between the PFV IN NTR and LTR DNA². TT- tight turns. (B) Sequence alignment of the NTD regions among M-MuLV, PFV, HIV-1, HIV-2, and MVV INs. Dark red boxes indicate residues that show sequence identity. Sequence numbering in panels A and B is for M-MuLV IN. Sequence alignments and identities were generated using the ESript server⁴⁷.

Figure 3. Crystal structure of M-MuLV IN₈₋₁₀₅

(A) The four molecules in the asymmetric unit, each in shown in different colors, form a dimer-of-dimers. The Zn^{2+} ions are colored in the same as the protein in which it is located. (B) Superimposition of the common dimer from the M-MuLV IN NTR₁₋₁₀₅ structure onto M-MuLV IN NTR₈₋₁₀₅ structure. The two monomers used in fitting in M-MuLV IN₈₋₁₀₅ are colored the same as in (A), and the other two are in gray. The two monomers from M-MuLV IN NTD₁₋₁₀₅ are colored in magenta and blue, respectively.

Figure 4. Structure comparison of M-MuLV IN-NTR with other virus IN NTRs and NTDs
 (A) Crystal structure of M-MuLV IN NTR₁₋₁₀₅ structure (PDB ID: 3NNQ). (B) Crystal structure of PFV IN NTR₁₋₁₀₅ (PDB ID: 3L2Q). (C) Crystal structure of HIV-1 IN NTD (PDB ID: 1K6Y). (D) Crystal structure of HIV-2 IN NTD (PDB ID: 3F9K). (E) Crystal structure of MVV IN NTD (PDB ID: 3HPG). In the case of (C), (D) and (E), the NED region present in M-MuLV and PFV INs is absent from these INs. All panels show the HHCC motif as stick projections, coordinating a zinc ion that is shown as a brown dot. Panels A and B NTR's consist of the NED (blue vertical bar) and NTD (green), while panels C, D and E consist of only the NTD (green vertical bars). In panels A and B the NED is comprised of the $\alpha 1$ helix and $\beta 1$, $\beta 2$ and $\beta 3$ strands. In all five panels, the NTD is comprised of only α helices, except for M-MuLV, which has a $\beta 4$ strand.

Figure 5. Surface electrostatic potential of M-MuLV IN NTR

(A) Ribbon representation of M-MuLV NTR₁₋₁₀₅ in aquamarine with basic residues and their side chains marked in brown. N- and C-termini are indicated with “N” and “C”, respectively. (B) Surface electrostatic potential of M-MuLV IN NTR₁₋₁₀₅ with red and blue colors indicating surface electrostatic potentials at ± 2 kT, at an ionic strength of 0.15 M and 298 K. (C) Transparent representation of panel B with the ribbon representation within. (D) View of the image in panel C rotated by 90°, showing basic residues (highlighted in brown) predicted to interact with DNA, largely clustering on one face. Images were generated using PyMol⁴⁸.

Figure 6. Overlay of M-MuLV and PFV IN NTRs

(A) Sequence alignment adapted from Fig. 2 indicating key residues and secondary structural features of M-MuLV and PFV IN NTRs. Sequences between the boxed regions are not shown. Key M-MuLV residues are marked in magenta and those of PFV are marked in red. Residues that are important for *in vitro* integration activity for M-MuLV are underlined. (B) Model of M-MuLV NTR – LTR complex based on the X-ray crystal structure of the homologous PFV NTR – LTR complex². PFV IN NTR (green) with DNA-binding residues represented with red sticks. Key residues predicted to be important in DNA binding for the M-MuLV IN NTR are represented with magenta sticks. Both crystal forms of M-MuLV IN NTR are depicted in the figure as blue (M-MuLV IN₈₋₁₀₅) and light blue (M-MuLV IN₁₋₁₀₅) and overlaid with PFV IN NTR. For ease of representation only the side chains of M-MuLV IN₈₋₁₀₅ are displayed in magenta. The top and bottom images of panel B differ by a 180° rotation. N- and C-termini are indicated with “N” and “C”, respectively. In both panels, the

transferred strand during integration is numbered in the 3'-5' direction; the CA dinucleotide occupies positions 2 and 1, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Data collection and refinement statistics

	NTR ₁₋₁₀₅			NTR ₈₋₁₀₅
Data collection				
Space group	<i>R</i> 32			<i>P</i> 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	112.82, 112.82, 115.53			44.66, 38.44, 135.27
α , β , γ (°)	90, 90, 120			90, 91.70, 90
	Peak	Inflection	Remote	
Wavelength (Å)	0.97917	0.97940	0.96863	1.07500
Resolution (Å)	2.8 (2.80–2.90) *	2.7 (2.70–2.80)	3.0 (3.00–3.11)	2.15 (2.15–2.23)
<i>R</i> _{merge} (%)	10.5 (62.0)	10.0 (88.6)	10.3 (51.4)	8.1 (60.5)
<i>I</i> / σ <i>I</i>	22.56 (3.52)	21.55 (2.56)	22.82 (4.59)	12.10 (2.40)
Completeness (%)	99.7 (100)	99.8 (100)	99.7 (100)	98.1 (97.0)
Redundancy	8.8 (8.8)	10.2 (9.6)	8.0 (8.1)	6.7 (6.2)
Refinement				
Resolution (Å)	2.69			2.15
No. reflections	7829			24869
<i>R</i> _{work} / <i>R</i> _{free} (%)	22.9/27.4			23.4/26.5
No. atoms				
Protein	1574			3103
Ligand/ion	10			120
Water	36			61
B-factors (Å ²)				
Protein	94.83			62.52
Ligand/ion	70.67			80.06
Water	71.49			52.04
R.m.s. deviations				
Bond lengths (Å)	0.674			1.321
Bond angles (°)	0.004			0.010

* Values in parentheses are for highest-resolution shell.