Published online 12 March 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 9* ***5917–5928***
doi: 10.1093/nar/gku175

Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction

Sriram Aiyer[1], G.V.T. Swapna[2], Nirav Malani[3], James M. Aramini[2], William M. Schneider[4,*], Matthew R. Plumb[5], Mustafa Ghanem[6], Ross C. Larue[5], Amit Sharma[5], Barbara Studamire[6], Mamuka Kvaratskhelia[5], Frederic D. Bushman[3], Gaetano T. Montelione[2,7,†] and Monica J. Roth[1,7,†]

[1]Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers University, 675 Hoes Lane, Piscataway, NJ 08854, USA, [2]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, 679 Hoes Lane West Piscataway, NJ 08854, USA, [3]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, 3610 Hamilton Walk, Philadelphia, PA 19104, USA, [4]Department of Biochemistry, Robert Wood Johnson Medical School, UMDNJ, 675 Hoes Lane, Piscataway, NJ 08854, USA, [5]Center for Retrovirus Research and College of Pharmacy, The Ohio State University, 484 W. 12th Ave., 508 Riffe Building, Columbus, OH 43210, USA, [6]Department of Biology, Brooklyn College, 417 Ingersoll Extension and the Graduate Center of the City University of New York, Brooklyn, NY 11210, USA and [7]Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854, USA

Received November 12, 2013; Revised February 12, 2014; Accepted February 12, 2014

ABSTRACT

We report alterations to the murine leukemia virus (MLV) integrase (IN) protein that successfully result in decreasing its integration frequency at transcription start sites and CpG islands, thereby reducing the potential for insertional activation. The host bromo and extraterminal (BET) proteins Brd2, 3 and 4 interact with the MLV IN protein primarily through the BET protein ET domain. Using solution NMR, protein interaction studies, and next generation sequencing, we show that the C-terminal tail peptide region of MLV IN is important for the interaction with BET proteins and that disruption of this interaction through truncation mutations affects the global targeting profile of MLV vectors. The use of the unstructured tails of gammaretroviral INs to direct association with complexes at active promoters parallels that used by histones and RNA polymerase II. Viruses bearing MLV IN C-terminal truncations can provide new avenues to improve the safety profile of gammaretroviral vectors for human gene therapy.

INTRODUCTION

Retroviruses have been used as an important tool in developing gene therapy vectors. Their ability to stably integrate genetic information into the host genome has enabled the exploitation of these viruses for many gene delivery applications. Gammaretroviral vectors have been used successfully to rectify defects of SCID-X1 and other diseases (1). However, despite the efficiency in gene delivery, insertional mutagenesis can result in clonal expansion of cells bearing specific integrants (2), associated with the preferential integration of murine leukemia virus (MLV) vectors upstream of transcription start sites (TSS) and CpG islands near promoter regions (3). This complicates their use in gene therapy.

In the retroviral replication cycle, the viral reverse transcriptase enzyme converts the single-stranded RNA viral genome into double-stranded DNA, which is associated within a preintegration complex (PIC). MLV requires cells to undergo mitosis. The viral p12 protein, which is part of the PIC, is responsible for tethering the viral genome to the host mitotic chromatin (4–5). However, the p12 protein does not mediate targeting of the viral PIC toward genomic hotspots for retroviral integration such as TSS and CpG islands (5). The viral integrase (IN), upon entry into the host

*Present address:
William Schneider, Center for the Study of Hepatitis C, The Rockefeller University, New York, 1230 York Avenue, NY 10065, USA.
†To whom correspondence should be addressed. Tel: +1 732 235 5048; Fax: +1 732 235 4073; Email: roth@rwjms.rutgers.edu
Correspondence may also be addressed to Tel: +1 732 235 5375; Fax: 1 732 235 5779; Email: guy@cabm.rutgers.edu

nucleus, mediates the integration of the viral DNA into the host genome (6). The viral IN protein is the primary viral determinant for target-site selection (7).

It has recently been shown that the host bromo and extraterminal (BET) domain proteins Brd 2, 3 and 4 bind the viral IN protein through their conserved ET domain (8–10). The down regulation of BET proteins with siRNAs (8–9) as well as treatment with a small molecule inhibitor JQ1, which selectively impairs BET protein association with chromatin, showed decrease in preferential integration targeting at TSS and CpG islands (8–10). In the presence of LEDGF-BET protein chimeras (10), integration can be shifted toward LEDGF binding sites. *In vitro* interaction studies and coimmunoprecipitation of overexpressed MLV IN in mammalian cells have mapped the BET binding sites to different domains of MLV IN including the catalytic core domain (CCD) (9), the C-terminal domain (CTD) (8,10) and the IN C-terminus (10).

In this report, we demonstrate that the C-terminal polypeptide segment of the viral IN protein, which we refer to as the tail peptide (TP), is a key determinant in mediating the interaction of the viral IN protein with the ET domain of the BET proteins. This interaction provides a structural basis for global *in vivo* integration-site preferences. MLV virus bearing IN lacking this C-terminal 28-residue TP are viable in tissue culture (11–12) and *in vitro* (13–14). Hence, deletion of the TP does not disrupt the catalytic properties of IN. MLV IN lacking the TP lose their interaction with BET proteins, thus presenting a direct mechanism to alter target-site utilization. Virus bearing IN lacking the TP, or with it replaced with other peptides, exhibits markedly diminished viral integrations in mammalian cells near TSS, CpG islands, and known BET binding sites.

MATERIAL AND METHODS

Plasmid constructs

$IN_{1–385\ XN}$ (previously named IN *in*6215a (11)) is an infectious M-MLV clone in which a NotI restriction site was inserted at an XbaI site and results in premature truncation of the IN protein at IN position 385. Insertions into the M-MLV infectious clone expressing IN *in*6215a ((11); $IN_{1–385\ XN}$) were performed at the NotI site using oligonucleotides, as described in the Supplementary Data. The MLV IN CTD was expressed in the bacterial pET15_NESG vector (15) at the NdeI and BamHI sites. Details of the glutathione S-transferase (GST)-IN and Brd3 ET constructs along with the cloning protocols are provided in the Supplementary Data.

Protein purification for NMR studies

Protein expression and purification from the pET-based construct for MLV IN CTD was performed as previously described (15–18) with the following modifications: protein expression was induced with 1 mM IPTG at 17°C for 25 h. Induction was carried out in MJ9 media (17–18) in the presence of either ^{15}N-labeled ammonium chloride or ^{15}N ammonium chloride plus uniformly ^{13}C-enriched glucose. Following Ni-NTA resin purification (Qiagen) as per manufacturer's instructions, fractions eluted in 400 mM imidazole were pooled and concentrated to a volume of less than 250 μL using an Amicon Ultracel-3K centrifugal filter unit (Millipore). The concentrated protein fraction was then injected into an AKTA FPLC and resolved on a Superdex 75 gel filtration column (GE Healthcare) in 20 mM sodium phosphate pH 8.0, 300 mM NaCl, 50 mM potassium glutamate ($KC_5H_8NO_4$) and 5 mM 2-mercaptoethanol. The eluted fractions were then pooled and concentrated using an Amicon Ultracel-3K centrifugal filter unit (Millipore). All isotopes were purchased from Cambridge Isotopes Laboratories.

Next generation sequencing of MLV IN C-terminal truncations

Sequencing was performed exactly as described before (5). Analysis of integration sites near Brd2, 3 and 4 binding sites were correlated using ChIP-seq data (19) and analyzed as described previously (8). The statistical test rely on the variance-covariance matrix of the relative ranks of the integration sites to construct Wald-type test statistics and referred to the Chi Square distribution to obtain *P* values (20). Datasets used in the analysis: the inset box defines the data sets used in the analysis; FV fibroblast (21), HIV-1 (22), MLV ((5) and this work), WT MLV $IN_{1–408}$, MLV $IN_{1–385\ 8N}$, MLV $IN_{1–385\ 16H}$, MLV $IN_{1–385\ XN}$ (*in*6215a (11), (Supplementary Table S4).

NMR analysis of MLV CTD structure and ET interactions

Single- (^{15}N) and double-(^{13}C, ^{15}N) enriched MLV IN CTD protein samples for studies of complex formation were concentrated to approximately 200 μM–1 mM concentration and where indicated, mixed with 2 mM unlabeled Brd3 ET in a buffer containing 5% ^2H_2O, 50 mM DSS, 300 mM NaCl, 50 mM potassium glutamate, 25 mM sodium phosphate at pH 7.0 or pH 8.0 and 5 mM 2-mercaptoethanol. Samples for NMR studies at pH 6.5 were prepared as described for pH 8.0, except at 100 mM NaCl and in the absence of 2-mercaptoethanol. Sequence-specific backbone ^{1}H, ^{13}C and ^{15}N resonance assignments for free and ET-bound $IN_{329–408}$ were determined at pH 7.0 and 8.0 using standard triple NMR resonance experiments (23). Resonance assignments and the solution NMR structure determination of $IN_{329–408}$ at pH 6.5 are reported elsewhere (PDB ID: 2M9U, BMRB ID: 19299). All spectra were recorded using a Bruker Avance 800 MHz spectrometer at 25°C. NMR data were processed using NMRPipe (24) and SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The oligomeric states of both free and complexed proteins (1:1 molar ratio of IN CTD:Brd3 ET at 100 μM or 200 μM) were assessed by measurements of rotational correlation times computed from ^{15}N T_1 and T_2 nuclear relaxation measurements, as described previously (25–26). Detailed protocols and calibration data for molecular correlation time measurements based on ^{15}N nuclear relaxation rate data are provided online at (http://www.nmr2.buffalo.edu/nesg.wiki/Main_Page). Peptides used for the TP competition assay are 'WT TP'—SRLTWRVQRSQNPLKIRLTREAP; and 'mutant TP'—SRLTARVQRSQNAAAIALTREAP (Pep-

tide 2.0 Inc.). The peptides were solubilized in deionized water to a final concentration of about 3 mM.

In vitro pull down assays

Pull-down assays were performed as described (8).

Protein interaction trap assay

The protein interaction trap assay (27) was adapted and performed as described in Supplementary Data.

Data deposition

The sequences reported in this paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive (project accession number SRP021184).

RESULTS

The structure of the MLV IN CTD changes in the presence of the Brd3 ET domain

The MLV IN interacts through its CTD with the BET family members through the ET domain (8). We have determined the three-dimensional structure of the MLV IN CTD (28) (PDB ID 2M9U) using conventional triple-resonance solution-state NMR methods. It consists of an SH3 fold followed by a long unstructured tail (Figure 1A, B and C). The solution structure and NMR resonance assignments (BMRB ID 19299) of the IN CTD provide unique tools to characterize IN-BET protein interactions.

Changes in the structural environment of amino-acid residues within the IN CTD upon complex formation with the ET domain of the BET protein can be monitored using the chemical shifts of backbone and side chain NMR resonances. Changes include both chemical shift changes, and changes in amide ^{1}H resonance intensities due to altered exchange rates with solvent water protons. Both of these effects upon complex formation are referred to in this study as chemical shift perturbations (CSPs). Using standard 2D [^{15}N-^{1}H]-heteronuclear single quantum coherence (HSQC)-type NMR experiments, backbone amide, side chain amide, arginine guanido and tryptophan indole ^{15}N-^{1}H NMR resonances can be monitored. Changes in ^{15}N and/or ^{1}H resonance frequencies and/or intensities can arise from many different aspects of the complex formation, including interfacial interactions, disorder to order transitions and/or occlusion of amide protons from solvent exchange due to formation of protein-protein interfaces or ordered structure within the TP region. In the case of complex formation between the IN CTD and BET ET domains, the TP region of the CTD becomes ordered in the complex and CSPs arise from both this change in the structure of the TP region and the specific interactions of the intermolecular interface.

Using standard [^{15}N-^{1}H]-HSQC experiments at 600 MHz, IN CTD-Brd3 ET interactions could be detected at pH $\geq$ 7.0, but not at pH 6.5 that was used to solve the solution NMR structure of the IN CTD (28). Accordingly, backbone ^{15}N, ^{1}H and ^{13}C resonance assignments for IN CTD, with and without complex formation with the unlabeled Brd3 ET domain, were redetermined at pH 7.0 and pH 8.0 using standard triple-resonance NMR experiments (23). Tryptophan indole NϵH resonances appearing in the [^{15}N-^{1}H]-HSQC spectra were also assigned. These NH resonance assignments are tabulated in Supplementary Table S1.

In transitioning from pH 6.5 to pH 8.0, the HSQC spectra of the free CTD domain exhibit attenuation of surface amide proton intensities due to base-catalyzed amide proton exchange. This attenuation of amide proton intensities is illustrated in Supplementary Figure S1A and B. However, in the spectra of the complex formed between ^{15}N-enriched IN CTD and unlabeled Brd3 ET domains, many of these same amide sites do not exhibit attenuation due to solvent-exchange broadening. This is because they are occluded due to structure formation in the TP region and/or due to the interface formed with the ET domain.

Figure 1 documents the significant NH CSPs, including changes in both frequencies and intensities of resonances, of IN CTD due to complex formation with Brd3 ET. Overlaying the [^{15}N-^{1}H]-HSQC spectra at pH 8.0 for the IN CTD in the presence (Figure 1D, blue) and absence (Figure 1D, red) of Brd3 ET demonstrates that binding of the Brd3 ET domain protects specific amide resonances from solvent-exchange broadening typical of surface amide NH resonances at pH 8.0. Significant resonance frequency shifts as large as 0.5 ppm are observed for the NH resonances of R391, V392, Q393, R394, L399, K400, I401, R402, L403 and T404 (Figure 1D and Supplementary Figure S1A and B). In addition, tryptophan indole NϵH resonances for residues W347, W369 and W390 appear or shift in the presence of Brd3 ET. Other amide resonances altered in intensity and/or frequency upon complex formation with Brd3 ET include residues Q339 and L343 (Figure 1A, in ß1–ß2 loop), T359 (Figure 1A, in ß2–ß3 loop), G365 (Figure 1A, in ß3–ß4 loop) and G381–383, R387, L388, T389, W390, S395, N397, R405, E406 and A407 (Figure 1A, C-terminal region). Several of these TP-associated NH resonances are observed in the complex but not in the free CTD at pH 8 (Figure 1D and Supplementary Figure S1B). Significantly, at pH 6.5, these resonances have amide ^{1}H chemical shifts typical of disordered polypeptide segments (i.e., in the range of 7.5–8.5 ppm; Supplementary Figure S1A).

Figure 1B and C shows a representation of the sites in the IN CTD domain that are perturbed in the presence of Brd3 ET mapped onto the 3D structure characterized at pH 6.5. The core β-barrel of the SH3 fold of the IN CTD is not affected by complex formation with the Brd3 ET. Rather, 22 of the 25 non-proline residues whose amide resonances exhibit significant CSPs (i.e., either slowed in solvent exchange and/or altered in chemical shift) are localized in the TP region, C-terminal to the SH3 fold. This suggests that the heterodimeric interaction is facilitated primarily through residues localized in the disordered TP region, many of which are conserved in other INs of the gammaretroviral genera (Supplementary Figure S1C).

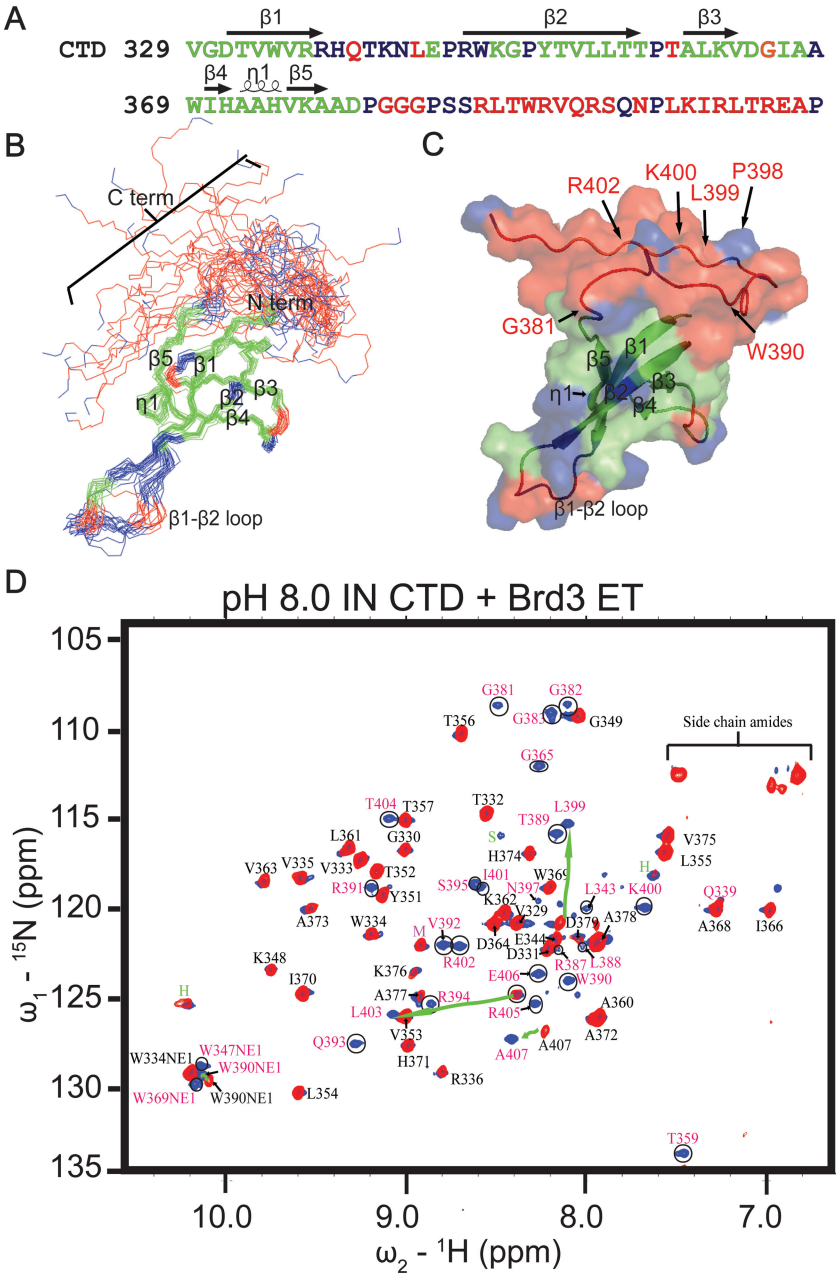

Figure 1. NMR analysis of MLV IN: Brd3 ET interaction. **A.** MLV IN CTD sequence is displayed with the following color codes: green indicates backbone amide resonance chemical shifts that were the same for IN CTD in the presence or absence of Brd3 ET at pH 8.0; red indicates backbone amide resonances that are observed in the presence of Brd3 ET, but solvent exchange broadened in the absence of complex formation, and/or amide resonances that exhibit frequency shifts upon complex formation; blue indicates backbone amide resonance assignments that could not be determined at pH 8.0 either in the presence or absence of the Brd3 ET domain. Residues for which HN amide assignments could not be determined in either free or ET-bound CTD at pH 8.0 include R337, H338, T340, K341, N342, R346, W347, A367, S385, S386, Q396, as well as proline residues P345, P350, P358, P380, P384, P398 and P408 which lack amide protons. **B.** Cα backbone trace, along with key structural features, of an ensemble of 20 conformers of MLV IN CTD from amino acids 329–408 (PDB ID 2M9U) is shown in this panel with the same color codes as described in panel A. **C.** Ribbon representation of a single MLV IN CTD conformer is shown within a transparent view of a surface space fill model. Color code is the same as in panel A with key structural features and specific amino-acid residues that show significant CSPs and/or reduced amide proton exchange broadening marked in red. All images were generated using PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.) **D.** NMR spectrum of the IN CTD–Brd3 ET complex. Overlay of [^{15}N^{-1}H]-HSQC spectra of ^{15}N-enriched IN CTD construct $IN_{329–408}$ at pH 8.0, 300 mM NaCl either with (blue) or without (red) unlabeled Brd3ET. The stoichiometric ratio of $IN_{329–408}$ (1 mM) and Brd3 ET (2 mM) was 1:2 at the concentrations indicated. Backbone amide resonances that are not affected by complex formation are labeled with sequence-specific assignments in black; assigned amide resonances that are not observable due to solvent-exchange broadening in the absence of ET, but become observable upon complex formation, as well as resonances exhibiting significant CSPs upon complex formation are labeled in magenta. All amide peak resonances not observable due to solvent-exchange broadening in the absence of ET, but become observable upon complex formation are marked with black circles; some of these could not be unambiguously assigned at pH 8.0. The curved green arrows indicate the CSPs due to complex formation of the amide resonances assigned to residues L399, L403, A407 and the side-chain indole NHε resonance of W390. Tryptophan W347 and W369 NHε side chain indole resonances with significant proton exchange rate reduction due to complex formation are also indicated. Peak resonances labeled in green are assigned to the non-cleavable affinity tag.

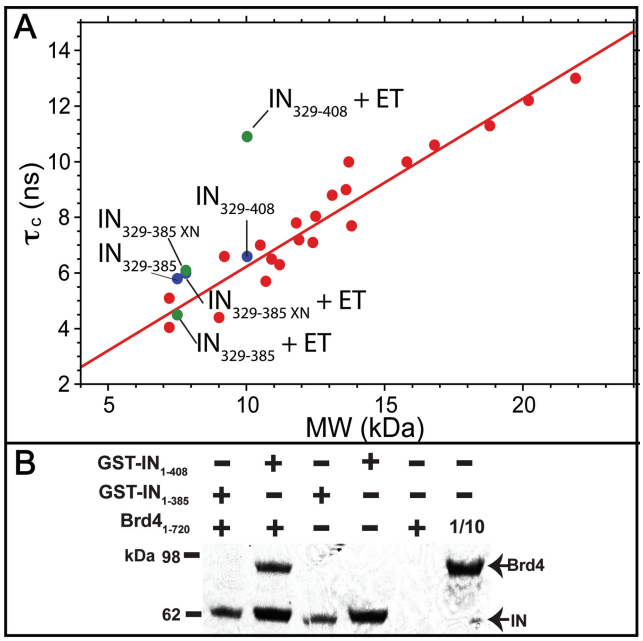

Figure 2. MLV IN interacts with the BET family through the IN TP. **A.** Rotational correlation time measurements. Plot of rotational correlation time (τ_c) computed from ^{15}N T_1/T_2 relaxation rate measurements versus molecular weight. Known monomeric protein standards are indicated in red. Data for the three IN CTD constructs ($IN_{329–408}$, $IN_{329–385\ XN}$ and $IN_{329–385}$) are indicated in blue (individually) and green (in the presence of the Brd3 ET domain). The molar ratio of the IN CTD constructs and Brd3 ET proteins was 1:1 at 100 μM. Plots of the ^{15}N T_1 and T_2 nuclear relaxation data for each sample are presented in Supplementary Figure S2. **B.** Interaction of MLV IN TP with Brd4. GST pull-down experiments performed with WT GST-MLV $IN_{1–408}$ and IN ΔC construct GST-MLV $IN_{1–385}$ with $Brd4_{1–720}$. Coomassie stain of SDS/PAG of GST pull-down products. Components of individual reactions are indicated as well as 10% of the purified $Brd4_{1–720}$ input sample. The predicted molecular of the GST-MLV $IN_{1–408}$ fusion protein is 69 kDa. Positions of molecular weight standards are indicated on the left.

The absence of the IN CTD TP disrupts the heterodimeric interaction

Having established that the heterodimeric IN CTD-Brd3 ET interaction is facilitated primarily by residues in the TP that are largely disordered in the IN free CTD (28), we next explored the impact of truncating these disordered residues on complex formation. Molecular rotational correlation time (τ_c) measurements, based on simple 1D ^{15}N nuclear relaxation rate measurements, can be used to estimate molecular mass changes upon complex formation under precisely defined conditions (25). These data are summarized in Figure 2A with supporting data in Supplementary Figure S2. A calibrated plot of τ_c versus molecular weight (Figure 2A) indicates that at 100 μM concentration, 300 mM NaCl, pH 8.0 and 25°C, MLV CTD construct $IN_{329–408}$ behaves as a monomer (τ_c = 6.6 ns, MW ~ 10.5 kDa). Addition of stoichiometric amounts of unlabeled Brd3 ET results in a heterodimer (τ_c = 10.9 ns, MW ~ 18 kDa). Two MLV IN CTD constructs truncated at the C-terminus (ΔC), $IN_{329–385}$ and $IN_{329–385\ XN}$ were also found to behave as monomers in either the presence or absence of unlabeled Brd3 ET (i.e., τ_c = 4.5 - 6.1 ns, MW ~ 8 ± 1 kDa), as illustrated in Figure 2A. Under these conditions, removal of the IN TP prevents heterodimer formation.

In order to validate that the truncation of the TP region did not affect the 3D structure of the rest of the IN CTD, we also recorded [^{15}N-^{1}H]-HMQC NMR spectra of the IN ΔC constructs (Supplementary Figure S3). These data demonstrate that truncation of the C-terminal disordered TP region does not affect the 3D structure of the rest of the CTD domain, including the core SH3 fold. The activity of IN constructs truncated at residue 385 has been extensively characterized *in vitro* and *in vivo* (11–14,29–30. Additional characterization with respect to viral titer, reverse transcriptase activity, minus-strand strong stop and plus-strand extension, Alu-PCR and two-end integration assays are described in the Supplementary Data.

These NMR studies of the IN CTD in the presence of the Brd3 ET domain provide biophysical evidence of the two proteins interacting, with changes in the MLV IN CTD NMR spectra clustering within the disordered IN C-terminal TP. The interaction with the Brd3 ET domain requires the C-terminal polypeptide region, and is suppressed by deletion of residues 386–408.

Additional protein interaction studies corroborate the role of the TP region in the structural basis of Brd3 ET binding. Figure 2B and Supplementary Figure S4 document direct binding between GST-MLV $IN_{1–408}$ and full-length Brd2, Brd3 and Brd4 constructs. This interaction was lost with constructs lacking the CTD TP (e.g., GST-MLV $IN_{1–385}$ in Figure 2B and Supplementary Figure S4). Yeast two-hybrid studies were also used to analyze the interaction between the mouse Brd2 ET domain and Moloney MLV (M-MLV) IN ($mIN_{1–408}$) (Table 1). The interaction between the LexA-mIN bait plasmid and the GAL4-ET domain prey plasmid in the colony lift assay was readily detected, as indicated by the blue color observed in the β-galactosidase assays. This interaction was stronger than the dimerization between M-MLV IN monomers (pSH2-mIN plus pGADNOT-mIN or pACT2-mIN). Deletion of the MLV IN C-terminal TP (pSH2-$mIN_{1–385\ XN}$; IN ΔC), markedly decreased the interaction with the Brd2 ET domain (pACT2-Brd2 ET; Table 1). Negative controls remained white, other than a very weak background interaction observed between the MLV IN expression construct (pGADNOT-5'mIN) and the empty vector (pSH2–1). Proteins expressing the HIV-1 p66 and p51 reverse transcriptase subunits (pSH2-p66 and pGADNOT-p51) served as positive controls.

Taken together, the protein interaction data summarized in Figure 2, Table 1 and Supplementary Figures S3 and S4 unequivocally demonstrate that the TP region of MLV IN, including residues 386–408, is critical for interactions with the ET domains of Brd2, Brd3 and Brd4 BET family members.

TP competition assay shows sequence-dependent disruption of the heterodimeric interaction

The importance of the TP region was further validated using a 24 amino-acid peptide (WT TP; Figure 3A) in a competition assay to disrupt the MLV IN CTD and Brd3 ET complex. The TP region contains a $_{390}WX_7PLK^{I}/_{L}R_{402}$

Table 1. Interaction between M-MLV IN and Brd2 ET LexA DNA binding domain fusions

GAL4 AD Fusions	pSH2-1	pSH2-p66	pSH2-mIN$_{1—408}$	pSH2-mIN$_{1–385\ XN}$
pACT2-empty	–	–	–	–
pACT2-mIN	–	–	+/–	+
pACT2-Brd2-ET	–	nd	+++	+
pGADNOT-empty	–	–	–	–
pGADNOT-p51	–	+++	–	–
pGADNOT-mIN	–	–	+/–	+
pGADNOT-5'mIN	+/–	–	+	++
pGADNOT- Ddxp18	–	nd	++	++

Qualitative ß-galactosidase yeast colony lift assays. Results represent the average colorimetric values from six independent transformation reactions and their corresponding ß-galactosidase assays. p51 and p66 represent the components of the HIV-1 reverse transcriptase heterodimer. The LexA DBD bait vectors are pSH2 and its derivatives; the GAL4-AD vectors pACT2 and pGADNOT (or their derivatives) are the prey vectors. Ddxp18 was previously isolated as an IN interactor in a yeast-two hybrid screen (27). Interaction key: (–) white; (+/–) pale blue; (+) light blue; (++) intermediate blue; (+++) dark blue; (nd) not determined (27).

motif that is conserved across gammaretroviral IN proteins (Supplementary Figure S1C). Accordingly, in addition to the wild-type (WT) peptide, a second 24-residue mutant TP in which all five of the conserved amino acids in the consensus motif are replaced with alanine (Figure 3A) was also studied.

Rotational time (τ_c) measurements, summarized in Supplementary Table S2 and supported by data in Supplementary Figure S5, demonstrate that the 24-residue WT TP disrupts the heterodimeric IN CTD–Brd3 ET complex. [^{15}N-^{1}H]-HSQC spectra presented in Figure 3B further demonstrate that the WT TP disrupts the complex formed between ^{15}N-enriched IN CTD (construct IN$_{329–408}$) and Brd3 ET. Adding the WT TP peptide to the complex of ^{15}N-enriched IN CTD bound to unenriched Brd3 ET dramatically changes the spectrum, resulting in an [^{15}N-^{1}H]-HSQC similar to that of the free full-length IN CTD. The spectrum of ^{15}N-enriched IN CTD in the presence of unenriched ET and WT TP is shown superimposed on the spectrum of *free* IN CTD in Figure 3B. The mutant TP, replacing five key residues with alanine, does not disrupt the complex between IN CTD (construct IN$_{329–408}$) and Brd3 ET (Figure 3C), that is, the spectrum of the complex is not altered by adding the mutant TP peptide. These data demonstrate the importance of some or all of these five residues in the energetics of complex formation.

Truncation of the CTD C-terminal TP results in decreased integration at TSS and CpG islands.

A hallmark of MLV integration is the preferential integration within 2 kb of TSS and CpG islands (Figure 4A and B). The down regulation and inhibition of BET proteins have been shown to influence this integration bias. The BET protein inhibitor JQ1 blocks binding of BET proteins to modified histones but maintains the interaction of BET proteins with IN. Truncation of the MLV IN CTD rather resulted in the direct loss of interaction with BET family members (Figure 2, Supplementary Figure S4 and Table 1). Mapping the integration site profile for MLV lacking the IN C-terminal TP thus tests whether the binding of BET family proteins drives MLV integration to TSS and CpG islands and presents a direct mechanism to alter the MLV integration preference.

Viruses lacking the C-terminal TP (ΔC) are viable in tissue culture and multiple tags have been inserted into the C-terminal segment of MLV IN (11,12,30–31). The integration target-site distribution of three constructs lacking the C-terminal 23 amino acids of IN were examined (IN$_{1–385\ XN}$ (in6215a) (11)), IN$_{1–385\ 8N}$ and IN$_{1–385\ 16H}$) each differing in the non-viral amino-acid sequence tags at their terminus (Supplementary Data). Viral titers were within 2–4-fold of WT for all three IN ΔC virus (Supplementary Table S3) and cells transfected with all three IN ΔC constructs were positive for reverse transcriptase by the end of the second passage of cells (day 8 post transfection; Supplementary Figure S6A). Additionally, accumulation of reverse transcription intermediates and the copy number of viral integrants of the IN ΔC constructs were within 2-fold of WT MLV (Supplementary Figure S6B and C). All three MLV IN ΔC constructs showed markedly diminished preference for integration at TSS (Figure 4A, red arrow) and CpG islands (Figure 4B, red arrow) compared to the experimental WT MLV IN control and a compilation of published MLV integration sites generated by infection of 293T cells (Supplementary Table S4). For example, integration of IN ΔC constructs plus or minus 1 kb of TSS averaged ∼2.5% of the total integrants, whereas ∼12% of the WT MLV mapped within this interval. Levels of the MLV IN ΔC constructs remained above the published levels for the Foamy virus and HIV-1 integrants in 293 and 293T (Figure 4 and Supplementary Table S4). Loss of localization to TSS and CpG islands was not specific for oncogenes; analysis of housekeeping genes showed a similar decrease in targeting their promoter regions (Supplementary Figure S7). Analysis of the local sequence bias at the site of the target DNA duplication revealed that the IN ΔC constructs maintained the characteristic palindromic consensus sequence of MLV (Supplementary Figure S8) at the scissile bonds (32). These results indicate that the loss of the IN C-terminal TP, which is the key IN:BET interaction domain, results in redistribution of viral integration sites away from promoter regions without grossly altering virus viability (11).

MLV IN ΔC integration sites lose association with known BET protein binding sites in 293 cells

In 293 cells, binding sites of BET proteins (Brd2, Brd3 and Brd4) have been identified by chromatin immunoprecipita-

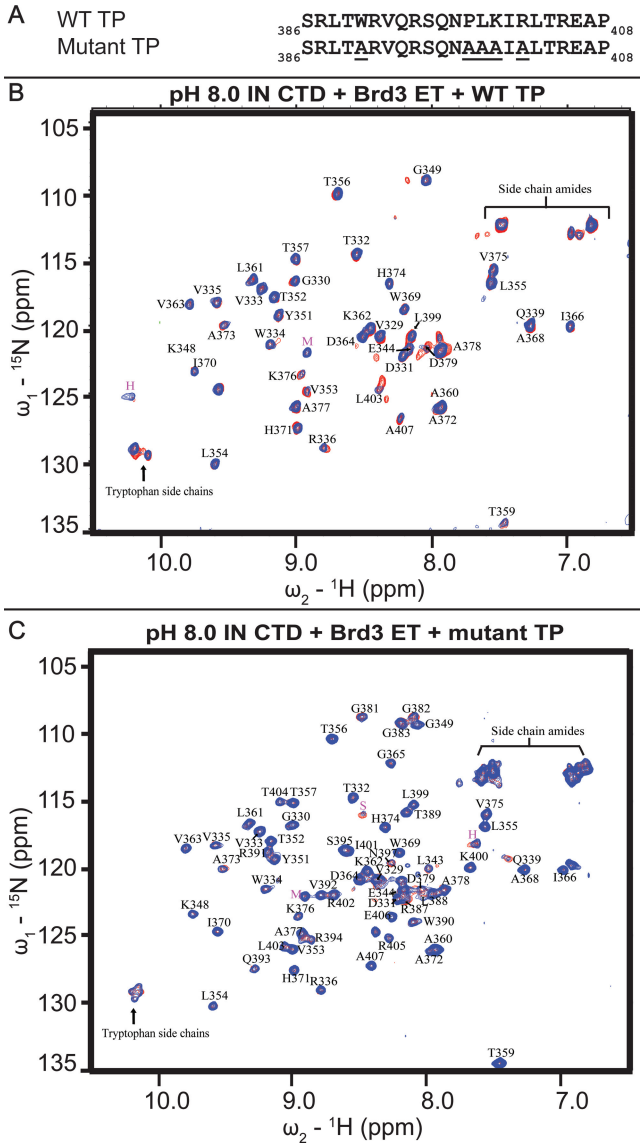

Figure 3. Mutating the consensus TP sequence inhibits interaction of MLV IN CTD with Brd3 ET. **A.** The 24-residue WT TP sequence is displayed on top and the mutant TP sequence is displayed on bottom. Underlined residues indicate amino-acid residues mutated to alanine. **B.** Comparison of [^{15}N-^{1}H]-HSQC spectra of the complex formed between ^{15}N-enriched IN CTD (construct $IN_{329–408}$) and unlabeled Brd3 ET in the presence of the WT TP (blue) and the free ^{15}N-enriched IN CTD (construct $IN_{329–408}$) spectra (red; same as Supplementary Figure S1B). The stoichiometric ratio of $IN_{329–408}$ (200 μM) and Brd3 ET (200 μM) was 1:1 and the peptide was added at 3-fold molar excess (600 μM). Under these conditions, the WT TP disrupts the complex by binding to the ET domain, resulting in a spectrum for CTD that is different from that of the complex, but essentially identical to that of free IN CTD. **C.** Comparison of [^{15}N-^{1}H]-HSQC spectra of ^{15}N-enriched IN CTD (construct $IN_{329–408}$) and unlabeled Brd3 ET in the presence (red) or absence (blue) of the mutant TP. The stoichiometric ratio of $IN_{329–408}$ (200 μM) and Brd3 ET (200 μM) was 1:1 and the mutant peptide was added at 3-fold excess (600 μM). The mutant TP does not disrupt the complex, and the spectrum is not changed when the peptide is added. In both panels, backbone amide resonances are labeled in black and peak resonances labeled in magenta are assigned to the non-cleavable affinity tag. Buffer conditions are as follows—buffer 1: 25 mM sodium phosphate pH 8.0, 300 mM NaCl, 50 mM potassium glutamate, 5 mM 2-mercaptoethanol. Buffer 2: 25 mM sodium phosphate pH 8.0, 360 mM NaCl, 60 mM potassium glutamate, 6 mM 2-mercaptoethanol.

tion and mapped onto the human genome (19). The correlation of these BET protein binding sites and integration sites of MLV WT IN and IN ΔC viruses was examined (Figure 4C). Analysis was performed using the total BET protein binding sites (Figure 4C: promoters + within genes + intergenic regions) and as well as those limited to specific promoter regions (Figure 4D). WT IN integrations correlated with identified Brd2, Brd3 and Brd4 binding sites compared to matched random controls throughout the host chromosomes. Integrants obtained from all three IN ΔC isolates showed a marked decrease (~2-fold, $P < 0.001$) compared to the WT IN integrants. Interestingly, this effect was localized to within 100 bp of a known BET protein-binding site. Approximately 9–13% of BET protein binding sites map within 2 kb of genes (19). Analysis of the MLV integrants from IN ΔC isolates indicated a more pronounced decrease in association with BET protein binding sites (4-fold) located within promoters upon loss of the IN C-terminal TP (Figure 4D), mapping to within 100 bp of the identified Brd2, Brd3 or Brd4 sites. These results indicate that the loss of the IN C-terminal tail results in the loss of targeting to identified binding sites of BET proteins.

DISCUSSION

The results of this study establish the molecular mechanism of MLV retroviral integration into TSS and CpG islands by tethering MLV IN with BET family proteins Brd2, Brd3 and Brd4 (Figure 5). Using solution NMR and biochemical studies, the interaction was localized predominantly to the C-terminal polypeptide tail segment of the MLV IN protein, including residues 386–408. In the absence of the BET proteins, the MLV IN C-terminal peptide (TP) is unstructured. It becomes structured upon complex formation with the Brd3 ET domain. This disorder-to-order transition may contribute a significant entropic component to the energetics of complex formation. The TP is non-essential for virus viability; however, viruses lacking the TP show marked diminution of integrations at the TSS and CpG islands, which are favored targets of MLV DNA integration in chromosomes. Loss of the MLV IN TP also correlated with the loss of association with BET protein binding sites within 293 cells.

Backbone CSPs upon complex formation indicate a change in the environment of the corresponding backbone atoms. Such changes implicate the corresponding region of the polypeptide or protein in the molecular recognition process, but do not necessarily demonstrate a direct contact of the corresponding residue in the intermolecular interface; that is, CSPs do not necessarily indicate direct intermolecular interactions. Site directed mutations could be used to validate the role of specific atoms in interfacial interactions, but the interpretation of such data requires structural data demonstrating that such sequence changes do not disrupt the structure of the CTD or the CTD–ET complex distal from the site of mutation. In any case, the extensive CSP data for free and ET-bound IN CTD, summarized in Figure 1 and Supplementary Table S1, provide unequivocal evidence for structural changes in C-terminal TP region of IN CTD upon complex formation, providing a structural basis for the interaction between these two domains.

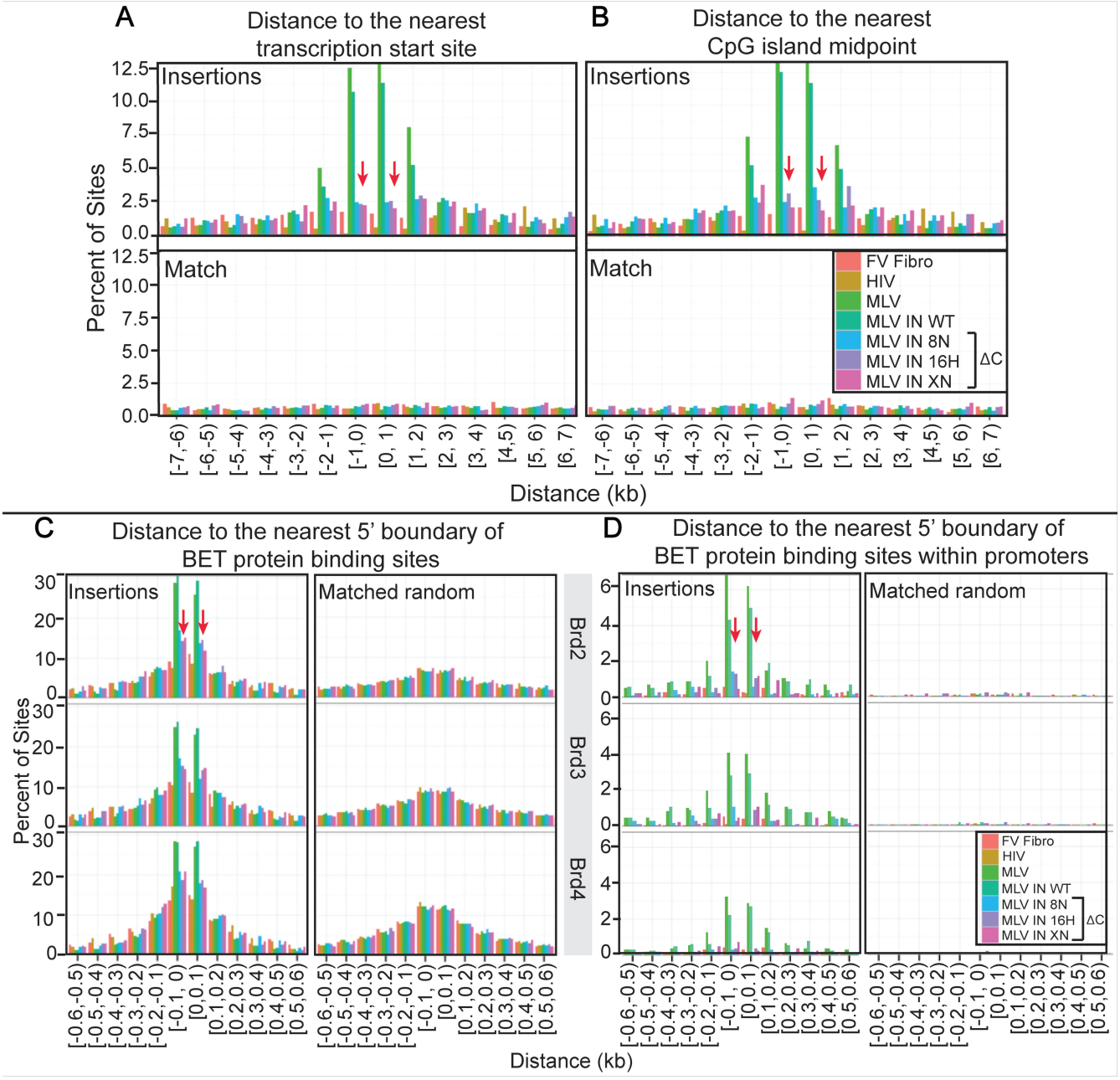

Figure 4. MLV IN ΔC integrations lose association with TSS, CpG islands and known BET protein binding sites in 293 cells. **A**. Percentages of integration sites in 293 cells are plotted with respect to the distance from the annotated TSS compared to matched random controls (MRCs) (50). **B.** Percentages of integration sites are plotted with respect to the distance from the 5' end of the nearest CpG islands compared to MRCs. **C.** Total BET binding sites in 293 cells. Integration sites were measured for its proximity to the 5' boundary of known BET protein binding sites compared to MRCs. MRCs are selected randomly from the host genome respective to a restriction site and should have no relation to chromatin sites. **D.** BET protein chromatin binding sites that overlap with promoters as defined in LeRoy et al. ((19); additional file 8). Loss of association of the IN ΔC versus the WT IN within 100 bp of chromatin sites bound by BET proteins was statistically different with $P < 0.001$ for both the total BET sites and those localized to specific promoters, respectively. Square bracket denotes inclusion of the limit, while parenthesis denotes exclusion.

The C-terminus of MLV IN is non-essential for IN enzymatic activity *in vitro* (13) and for virus viability in tissue culture (12,31), however second-site revertants have been isolated from modified IN ΔC virus where the IN C-terminal tail was restored (31). The MLV IN TP overlaps with the Env signal peptide in an alternative reading frame. The sequence conservation within the IN TP/4070A amphotropic Env overlap region displays a bias toward maintenance of the IN reading frame (33). This suggests that the C-terminal segment of IN may have a functional role *in vivo*. Alignment of retroviral IN proteins with known targeting to TSS and CpG islands show conservation of the sequence $_{390}$WRVQRSQN**PLKIR**LTR$_{405}$ located at the MLV IN C-terminus (Supplementary Figure S1C). This homology ex-

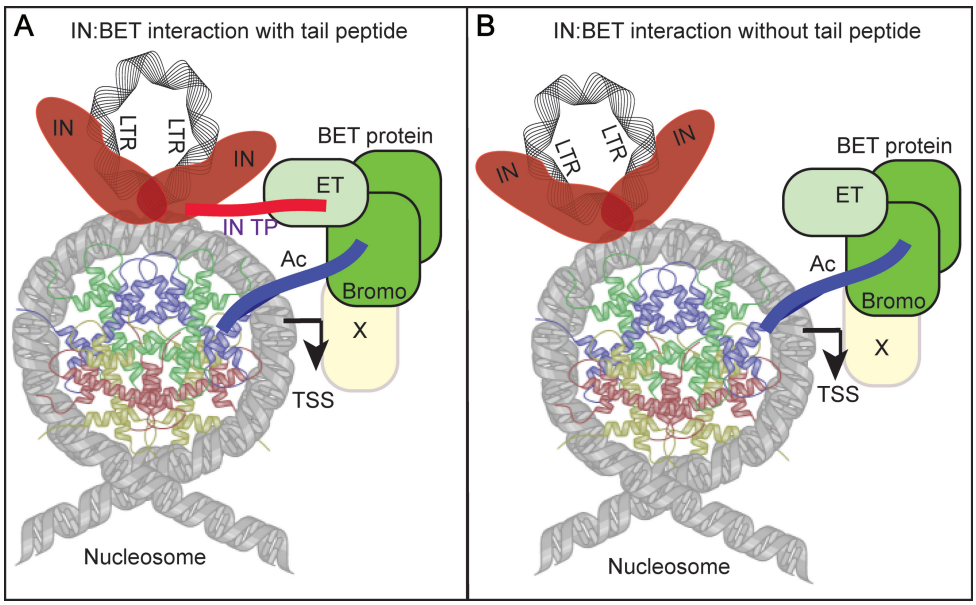

Figure 5. Model for MLV integration. **A.** Assembly of BET proteins on acetylated histone tail (blue) in the presence of additional host factors and RNA Pol II (together marked as X) to TSS. MLV IN interacts with the ET domain of BET proteins through the IN TP (red) resulting in a preponderance of integrations near TSS. **B.** Loss of MLV IN TP results in loss of association with BET proteins and results in decreased targeting to BET binding sites and TSS.

tended to the core consensus of $_{390}WX_7PLKJR_{402}$ (Figure S1C; J = I/L) in RaLV, GaLV and KoRV, but the integration site preferences for these viruses have not been analyzed. It is predicted that these would be critical for the interaction with BET family members. Indeed, the peptide competition assay (Figure 3 and Supplementary Table S2) shows that one or more of residues W390, P398, L399, K400 and R402 contribute to the energetics of complex formation. Additionally, IN W390A has been shown to have reduced binding affinity for Brd4 ET (10).

The function of the IN C-terminal tail appears analogous to that of the histone (34–35) and RNA polymerase II tails (36), as an otherwise unstructured docking site for additional proteins or regulatory factors (Figure 5). Interestingly, the NS1 protein of influenza A H3N2 subtype is proposed to use a related histone mimic as a means to associate with the human PAF1 transcription elongation complex and thus inhibit the antiviral response (37).

Targeting of IN integration to active promoter regions is beneficial for expression of the viral genome for subsequent rounds of infection. However, targeting to promoter regions only accounts for approximately 25% of all MLV retroviral integrations. Loss of targeting through the BET proteins does not affect the majority of MLV integration events. The local sequence selection at the site of integration (LOGOs analysis) remain unchanged in the absence of the IN TP. The quantitation of integration copy numbers of the three replicating IN ΔC isolates was statistically similar to WT IN and consistent with the 2-fold variation observed for a single round of infection (12). Additionally, reverse transcription of the viral genome into dsDNA is not grossly affected by the absence of the IN TP (Supplementary Figure S6B and C). Variations in viral titers (between 2- and 4-fold) and the time course of viral spread were observed between the three IN ΔC isolates. These assays depend on the level of expression of the viral or transgene mRNA. Loss of targeting to highly expressed promoter regions can result in a decrease in the total level of viral mRNA produced and thus be reflected in minor variations in viral titer and subsequent proviral integrations. Alternatively, the differences in the IN terminal amino acids may affect the association with other unidentified host factors.

Recently, it was reported that three residues in the MLV IN CCD are important in the interaction with BET proteins (9). It is unclear whether these mutations compromise MLV IN catalytic activities or indirectly affect interactions with BET proteins. Despite an intact TP region, interaction with BET proteins was lost in the context of these CCD mutations. It is possible that the presence of a C-terminal FLAG tag used in that study might compromise the binding potential of the TP region. Loss of the IN TP though did not compromise IN catalytic activities within the CCD or those requiring multimerization including two-end concerted integration (Supplementary Figure S9).

Interestingly, BET family members have been implicated in other viral systems and used to regulate viral transcription as well as tethering the viral DNA complexes to mitotic chromosomes. Both the KSHV LANA protein and the Merkel cell polyomavirus T antigen bind Brd4 through protein–protein interactions with the ET domain (38–40). In addition, various animal and human papilloma viruses interact with the host protein Brd4 (41). For MLV, integration at BET protein binding sites positions the virus within transcriptionally active regions of the host chromosomes and should facilitate viral gene expression.

Multiple retroviral and retrotransposon systems utilize host proteins to influence their position of viral integration. Integration of HIV-1 is directed within genes through asso-

ciation with the host factor LEDGF/p75 (22,42–44). Yeast Ty1 and Ty3 target integration into tRNA genes. Ty3 interacts with Brf1, TFIIB and TBP, which are involved in Pol III transcription (45). Recent studies of Ty1 targeting to tRNA genes implicate a possible histone modification near Pol III transcription sites as a driving factor for recognition (46). For Tf1, it is known that a chromodomain at the C-terminus interacts with the host Atf1 protein to drive integration into genes associated with environmental stresses (47).

The function of the ET domain of BET family members is not well defined. Recent studies have identified NSD3, JMJD6, CHD4, GLTSCR1 and ATAD5 proteins as binding partners to the BET family members ET domain (48) and thus the ET domain functions to recruit specific effector proteins to regulate transcriptional activity. It is not known if binding of the MLV IN TP interferes with the function of the ET domain and/or the association of cellular proteins to this domain. The solution structure of the Brd4 ET domain consists of three α-helices plus a loop structure, with no close structural homologs (49). Binding to IN is localized to the acidic patch in the α2–α3 loop region (9–10). The solution structure of the ET domain:IN CTD complex will be interesting to elucidate.

Mapping of WT integrated MLV proviruses shows favored targeting within 2 kb of the TSS and CpG islands (3). For the WT IN, correlation with the chromatin binding sites of BET proteins showed a tighter association, favoring integration within 100 bp of the defined BET protein binding sites for Brd2, 3 and 4. The two analyses measure distinct features, as the positions of the BET sites with respect to the TSS are not defined. Loss of the IN TP did not affect a panel of additional features, including integration in gene dense regions (50). MLV integrants are reported to be associated with the supermarkers (51) containing H3K4me1, H3K4me3, H3K9ac chromatin marks and STAT1 binding sites. Indeed, comparison of the integration sites of the WT IN versus the IN ΔC constructs from this study with the epigenetic maps of HeLa and resting T cells correlated the loss of the IN TP with the loss of integrants at H3K4me1, H3K4me3, H3K9ac chromatin marks, STAT1 binding sites and H2AZ sites (data not shown). Understanding the additional proteins that assemble with the preintegration:BET complexes will shed light on the molecular mechanisms associated with recognizing these supermarkers as well as those that stabilize gene or promoter-specific targets. This knowledge has direct implications for understanding the potential of murine-based vectors for gene activation resulting from insertional viral mutagenesis.

Eliminating the interactions with BET family members has implications for the use of gammaretroviral vectors for gene therapy and gene delivery. Foamy viruses, which have limited bias to integrate at TSS and CpG islands have been shown to be apathogenic in humans (52). Vectors lacking the IN TP would decrease targeting to promoter regions of all classes of genes, not specifically oncogenes. This nonetheless has the potential to decrease the probability of promoter/enhancer insertional mutagenesis due to viral integration at promoter regions. The modulation of overall expression of the transgene due to chromosomal positional effects might become more evident in the absence of BET protein targeting. Current studies are aimed at defining the potential of oncogene activation resulting from integration of MLV bearing the IN ΔC proteins.

ACCESSION NUMBER

The sequences reported in this paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive (project accession number SRP021184).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online, including [53–67].

ACKNOWLEDGMENT

We would like to thank Roland Felkner, Jared Sharp and Christopher Warren for technical assistance.

FUNDING

National Institutes of Health [RO1 GM070837, GM088808 to M.J.R., AI052845 to F.D.B., 5SC2GM088095, PSCOOC-40-199 to B.S., AI062520 to M.K.]; National Institute of General Medical Sciences Protein Structure Initiative [U54-GM094597 to G.T.M.]. Funding for open access charge: NIH [RO1 GM070837, GM088808].

REFERENCES

1. Fischer,A., Hacein-Bey-Abina,S. and Cavazzana-Calvo,M. (2013) Gene therapy of primary T cell immunodeficiencies. *Gene*, **525**, 170–173.
2. Hacein-Bey-Abina,S., Von Kalle,C., Schmidt,M., McCormack,M.P., Wulffraat,N., Leboulch,P., Lim,A., Osborne,C.S., Pawliuk,R., Morillon,E. *et al.* (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*, **302**, 415–419.
3. Wu,X., Li,Y., Crise,B. and Burgess,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.
4. Elis,E., Ehrlich,M., Prizan-Ravid,A., Laham-Karam,N. and Bacharach,E. (2012) p12 tethers the murine leukemia virus pre-integration complex to mitotic chromosomes. *PLoS Pathog.*, **8**, e1003103.
5. Schneider,W.M., Brzezinski,J.D., Aiyer,S., Malani,N., Gyuricza,M., Bushman,F.D. and Roth,M.J. (2013) Viral DNA tethering domains complement replication-defective mutations in the p12 protein of MuLV Gag. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 9487–9492.
6. Engelman,A., Mizuuchi,K. and Craigie,R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211–1221.
7. Lewinski,M.K., Yamashita,M., Emerman,M., Ciuffi,A., Marshall,H., Crawford,G., Collins,F., Shinn,P., Leipzig,J., Hannenhalli,S. *et al.* (2006) Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.*, **2**, e60.
8. Sharma,A., Larue,R.C., Plumb,M.R., Malani,N., Male,F., Slaughter,A., Kessl,J.J., Shkriabai,N., Coward,E., Aiyer,S.S. *et al.* (2013) BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 12036–12041.
9. Gupta,S.S., Maetzig,T., Maertens,G.N., Sharif,A., Rothe,M., Weidner-Glunde,M., Galla,M., Schambach,A., Cherepanov,P. and Schulz,T.F. (2013) Bromo and ET domain (BET) chromatin regulators serve as co-factors for murine leukemia virus integration. *J. Virol.*, **87**, 12721–12736.

10. De Rijck,J., de Kogel,C., Demeulemeester,J., Vets,S., El Ashkar,S., Malani,N., Bushman,F.D., Landuyt,B., Husson,S.J., Busschots,K. *et al.* (2013) The BET family of proteins targets Moloney murine leukemia virus integration near transcription start sites. Cell Rep., **5**, 886–894.
11. Roth,M.J. (1991) Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. *J. Virol.*, **65**, 2141–2145.
12. Schneider,W.M., Wu,D.T., Amin,V., Aiyer,S. and Roth,M.J. (2012) MuLV IN mutants responsive to HDAC inhibitors enhance transcription from unintegrated retroviral DNA. *Virology*, **426**, 188–196.
13. Jonsson,C.B., Donzella,G.A., Gaucan,E., Smith,C.M. and Roth,M.J. (1996) Functional domains of Moloney murine leukemia virus integrase defined by mutation and complementation analysis. *J. Virol.*, **70**, 4585–4597.
14. Donzella,G.A., Jonsson,C.B. and Roth,M.J. (1996) Coordinated disintegration reactions mediated by Moloney murine leukemia virus integrase. *J. Virol.*, **70**, 3909–3921.
15. Xiao,R., Anderson,S., Aramini,J., Belote,R., Buchwald,W.A., Ciccosanti,C., Conover,K., Everett,J.K., Hamilton,K., Huang,Y.J. *et al.* (2010) The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol.*, **172**, 21–33.
16. Acton,T.B., Xiao,R., Anderson,S., Aramini,J., Buchwald,W.A., Ciccosanti,C., Conover,K., Everett,J., Hamilton,K., Huang,Y.J. *et al.* (2011) Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods Enzymol.*, **493**, 21–60.
17. Jansson,M., Li,Y.-C., Jendeberg,L., Anderson,S., Montelione,G.T. and Nilsson,B. (1996) High-level production of uniformly 15N- and 13C-enriched fusion proteins in Escherichia coli. *J. Biomol. NMR*, **7**, 131–141.
18. Schneider,W.M., Tang,Y., Vaiphei,S.T., Mao,L., Maglaqui,M., Inouye,M., Roth,M.J. and Montelione,G.T. (2010) Efficient condensed-phase production of perdeuterated soluble and membrane proteins. *J. Struct. Funct. Genomics*, **11**, 143–154.
19. LeRoy,G., Chepelev,I., DiMaggio,P.A., Blanco,M.A., Zee,B.M., Zhao,K. and Garcia,B.A. (2012) Proteogenomic characterization and mapping of nucleosomes decoded by Brd and HP1 proteins. *Genome Biol.*, **13**, R68.
20. Brady,T., Lee,Y., Ronen,K., Malani,N., Berry,C., Bieniasz,P. and Bushman,F. (2009) Integration target site selection by a resurrected human endogenous retrovirus. *Genes Dev.*, **23**, 633–642.
21. Trobridge,G.D., Miller,D.G., Jacobs,M.A., Allen,J.M., Kiem,H.P., Kaul,R. and Russell,D.W. (2006) Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 1498–1503.
22. Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J. and Bushman,F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
23. Moseley,H.N., Monleon,D. and Montelione,G.T. (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Methods Enzymol.*, **339**, 91–108.
24. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
25. Rossi,P., Swapna,G.V., Huang,Y.J., Aramini,J.M., Anklin,C., Conover,K., Hamilton,K., Xiao,R., Acton,T.B., Ertekin,A. *et al.* (2010) A microscale protein NMR sample screening pipeline. *J. Biomol. NMR*, **46**, 11–22.
26. Aramini,J.M., Ma,L.C., Zhou,L., Schauder,C.M., Hamilton,K., Amer,B.R., Mack,T.R., Lee,H.W., Ciccosanti,C.T., Zhao,L. *et al.* (2011) Dimer interface of the effector domain of non-structural protein 1 from influenza A virus: an interface with multiple functions. *J. Biol. Chem.*, **286**, 26050–26060.
27. Studamire,B. and Goff,S.P. (2008) Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology*, **5**, 48.
28. Aiyer,S., Rossi,P., Schneider,W.M., Chander,A., Roth,M.J. and Montelione,G.T. Solution NMR structure of the C-terminal domain from murine leukemia virus integrase (MLV-IN), Northeast Structural Genomics Consortium (NESG) Target OR41A. Northeast Structural Genomics Consortium. (doi: 10.2210/pdb2m9u/pdb; epub ahead of print).
29. Puglia,J., Wang,T., Smith-Snyder,C., Cote,M., Scher,M., Pelletier,J., John,S., Jonsson,C. and Roth,M. (2006) Revealing domain structure through linker-scanning analysis of the murine leukemia virus (MuLV) RNase H and MuLV and human immunodeficiency virus type 1 Integrase proteins. *J. Virol.*, **80**, 9497–9510.
30. Seamon,J.A., Adams,M., Sengupta,S. and Roth,M.J. (2000) Differential effects of C-terminal molecular tagged integrase on replication competent Moloney murine leukemia virus. *Virology*, **274**, 412–419.
31. Seamon,J.A., Miller,C., Jones,K.S. and Roth,M.J. (2002) Inserting nuclear targeting signals onto a replication-competent M-MuLV affects viral export and is not sufficient for cell cycle independent infection. *J. Virol.*, **76**, 8475–8484.
32. Wu,X., Li,Y., Crise,B., Burgess,S.M. and Munroe,D.J. (2005) Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.*, **79**, 5211–5214.
33. Ey,P., Freeman,N., Bela,B., Li,P. and McInnes,J. (1997) Nucleotide sequence of the murine leukemia virus amphotropic strain 4070A integrase (IN) coding region and comparative structural analysis of the inferred polypeptide. *Arch. Virol.*, **142**, 1757–1770.
34. Taverna,S.D., Li,H., Ruthenburg,A.J., Allis,C.D. and Patel,D.J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.*, **14**, 1025–1040.
35. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705
36. Jasnovidova,O. and Stefl,R. (2013) The CTD code of RNA polymerase II: a structural view. *Wiley Interdiscip. Rev. RNA*, **4**, 1–16.
37. Marazzi,I., Ho,J.S., Kim,J., Manicassamy,B., Dewell,S., Albrecht,R.A., Seibert,C.W., Schaefer,U., Jeffrey,K.L., Prinjha,R.K. *et al.* (2012) Suppression of the antiviral response by an influenza histone mimic. *Nature*, **483**, 428–433.
38. Platt,G.M., Simpson,G.R., Mittnacht,S. and Schulz,T.F. (1999) Latent nuclear antigen of Kaposi's sarcoma-associated herpesvirus interacts with RING3, a homolog of the Drosophila female sterile homeotic (fsh) gene. *J. Virol.*, **73**, 9789–9795.
39. You,J., Srinivasan,V., Denis,G.V., Harrington,W.J. Jr, Ballestas,M.E., Kaye,K.M. and Howley,P.M. (2006) Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen interacts with bromodomain protein Brd4 on host mitotic chromosomes. *J. Virol.*, **80**, 8909–8919.
40. Wang,X., Li,J., Schowalter,R.M., Jiao,J., Buck,C.B. and You,J. (2012) Bromodomain protein Brd4 plays a key role in Merkel cell polyomavirus DNA replication. *PLoS Pathog.*, **8**, e1003021.
41. McBride,A.A., Sakakibara,N., Stepp,W.H. and Jang,M.K. (2012) Hitchhiking on host chromatin: how papillomaviruses persist. *Biochim. Biophys. Acta*, **1819**, 820–825.
42. Llano,M., Vanegas,M., Fregoso,O., Saenz,D., Chung,S., Peretz,M. and Poeschla,E.M. (2004) LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.*, **78**, 9524–9537.
43. Shun,M.-C., Raghavendra,N.K., Vandegraaff,N., Daigle,J.E., Hughes,S., Kellam,P., Cherepanov,P. and Engelman,A. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev.*, **21**, 1767–1778.
44. Craigie,R. and Bushman,F.D. (2012) HIV DNA Integration. *Cold Spring Harb. Perspect. Med.*, **2**, a006890.
45. Qi,X. and Sandmeyer,S. (2012) In vitro targeting of strand transfer by the Ty3 retroelement integrase. *J. Biol. Chem.*, **287**, 18589–18595.
46. Baller,J., Gao,J., Stamenova,R., Curcio,M. and Voytas,D. (2012) A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. *Genome Res.*, **22**, 704–713
47. Levin,H.L. and Moran,J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.*, **12**, 615–627.
48. Rahman,S., Sowa,M.E., Ottinger,M., Smith,J.A., Shi,Y., Harper,J.W. and Howley,P.M. (2011) The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol. Cell. Biol.*, **31**, 2641–2652.

49. Lin,Y.J., Umehara,T., Inoue,M., Saito,K., Kigawa,T., Jang,M.K., Ozato,K., Yokoyama,S., Padmanabhan,B. and Guntert,P. (2008) Solution structure of the extraterminal domain of the bromodomain-containing protein BRD4. *Protein Sci.*, **17**, 2174–2179.
50. Berry,C., Hannenhalli,S., Leipzig,J. and Bushman,F.D. (2006) Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.*, **2**, e157.
51. Santoni,F.A., Hartley,O. and Luban,J. (2010) Deciphering the code for retroviral integration target site selection. *PLoS Comput. Biol.*, **6**, e1001008.
52. Lindemann,D. and Rethwilm,A. (2011) Foamy virus biology and its application for vector development. *Viruses*, **3**, 561–585.