BET proteins promote efficient murine leukemia virus integration at transcription start sites

Amit Sharma, Ross C. Larue, Matthew R. Plumb, Nirav Malani, Frances Male, Alison Slaughter, Jacques J. Kessler, Nikoloz Shkribaia, Elizabeth Coward, Sriram S. Aiyer, Patrick L. Green, Monica J. Roth, Frederic D. Bushman, and Mamuka Kvaratskhelia

The selection of chromosomal targets for retroviral integration varies markedly, tracking with the genus of the retrovirus, suggestive of targeting by binding to cellular factors. γ-Retroviral murine leukemia virus (MLV) DNA integration into the host genome is favored at transcription start sites, but the underlying mechanism for this preference is unknown. Here, we have identified bromodomain and extr-terminal domain (BET) proteins (Brd2, -3, -4) as cellular-binding partners of MLV integrase. We show that purified recombinant Brd4(1-720) binds with high affinity to MLV integrase and stimulates correct concerted integration in vitro. JO-1, a small molecule that selectively inhibits interactions of BET proteins with modified histone sites impaired MLV but not HIV-1 integration in infected cells. Comparison of the distribution of BET protein-binding sites analyzed using ChIP-Seq data and MLV-integration sites revealed significant positive correlations. Antagonism of BET proteins, via JO-1 treatment or RNA interference, reduced MLV-integration frequencies at transcription start sites. These findings elucidate the importance of BET proteins for MLV integration efficiency and targeting and provide a route to developing safer MLV-based vectors for human gene therapy.

Results

BET Proteins Specifically Interact with MLV IN. To identify cellular-binding partners of MLV IN, we used affinity capture coupled with mass spectrometry (MS). In parallel experiments, the interacting partners of MLV and HIV-1 INs from nuclear extracts of NIH 3T3 and Sup-T1 cells were compared. The semiquantitative analyses of MS peaks identified the BET proteins (Brd2, -3, and -4) as the main binding partners of MLV IN (Table 1, Table S1, and Fig. S1). Of these, Brd4 and Brd3 were the top hits in NIH 3T3 and Sup-T1 cells, respectively. Differential pull-down levels of these proteins (Table 1) could be attributable to the varying expression levels of BET proteins seen in different cell types (Fig. S2). In control experiments, no peptides from the BET proteins were detected in HIV-1 IN pull-downs. Instead, as expected (8), LEDGF/p75 was identified as the main interacting partner of HIV-1 but not MLV IN. Furthermore, the immunoblot analyses of the pull-down fractions (Fig. 1 A and B) have validated our MS results by showing that all three BET proteins selectively bind MLV IN but not HIV-1 IN. The interaction between MLV IN and Brd2 has also been detected by yeast two-hybrid experiments (19).

Next we attempted to map the interacting domains between MLV IN and full-length Brd3(1-726). MLV IN is composed of the following three distinct domains (Fig. S3A): the N-terminal domain (NTD), which like prototype foamy virus IN (20) is comprised of the NTD-extension domain and the HH-CC-type Zn finger; the catalytic core domain (CCD) containing the DDE triad that coordinates catalytic Mg2+; and the C-terminal domain (CTD), which is thought to bind DNA, but it could also have additional functions. Similar levels of ectopically expressed Brd3(1-726) were pulled down by full-length MLV IN and its CTD (Fig. 1C and Fig. S3B). In contrast, no interactions were detected between Brd3 (1-726) and the MLV IN NTD or the two domain (NTD+CCD) construct. These findings indicate that MLV IN CTD is primarily responsible and sufficient for interaction with Brd3(1-726).

To delineate the Brd3(1-726) segments (Fig. S4) interacting with MLV IN, two truncated constructs were studied: Brd3 (1-419), the N-terminal fragment which contained the two bromodomains known to interact with chromatin (21–25); and Brd3 (420-726), the C-terminal region implicated in a number of protein-protein interactions (reviewed in refs. 26–28). Full-length...
Brd3(1-726) (Fig. 1D) and its C-terminal Brd3(420-726) (Fig. 1E) bound MLV IN, whereas the N-terminal Brd3(1-419) failed to bind MLV IN (Fig. 1F). Similar results were seen with recombinant MLV IN binding to recombinant Brd4(1-720) and Brd4(462-720) but not Brd4(1-461) (Fig. S5).

BET Proteins Directly Bind and Stimulate MLV Integration in Vitro. To examine whether the observed interactions were direct or bridged by additional cellular components, we next examined interactions with purified recombinant proteins. Among the three BET proteins, only His-Brd4(1-720) was amenable to bacterial expression and purification. Importantly, Brd4(1-720) contains all of the functional domains and motifs conserved between Brd2, -3, and -4 (Fig. S4) and, thus, serves as a representative of all three BET proteins. The data in Fig. 2A show that Brd4(1-720) binds MLV but not HIV-1 IN. In control experiments, LEDGF/p75 interacted with HIV-1 but not MLV IN (Fig. 2B). Using pull-down experiments, the apparent $K_d$ value was $\sim$33 nM for MLV IN-Brd4(1-720) (Fig. 2C and D). A $K_d$ of $\sim$200 nM was measured for the HIV-1 IN-LEDGF/p75 interaction (29).

We next examined the effects of Brd4(1-720) in vitro-integration assays catalyzed by MLV IN (Fig. 2E and F). Addition of Brd4(1-720) to the reactions significantly enhanced the biologically relevant concerted two-end integration products. Comparative analysis of strand-transfer activities of MLV and HIV-1 INs revealed that the levels of stimulation for the integration activities of MLV IN by Brd4(1-720) and HIV-1 IN by LEDGF/p75 were comparable [Fig. 2F and G; $\sim$564% for MLV IN+Brd4(1-720) and $\sim$689% for HIV-1 IN+LEDGF/p75]. In control experiments, no stimulation of strand transfer activities of MLV IN by LEDGF/p75 or HIV-1 IN by Brd4(1-720) were

Table 1. List of top protein hits from NIH 3T3 and Sup-T1 cells specifically binding GST-MLV IN or GST-HIV-1 IN

<table>
<thead>
<tr>
<th>Protein</th>
<th>NIH 3T3 MLV IN</th>
<th>HIV-1 IN</th>
<th>Sup-T1 MLV IN</th>
<th>HIV-1 IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brd4</td>
<td>164</td>
<td>18</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Brd2</td>
<td>90</td>
<td>7</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Brd3</td>
<td>88</td>
<td>25</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>LEDGF/p75</td>
<td>ND</td>
<td>35</td>
<td>ND</td>
<td>15</td>
</tr>
</tbody>
</table>

“Unweighted spectrum count” values for each protein are shown from representative runs of four independent experiments. “ND” indicates that the listed protein was not detected.

Brd3(1-726) (Fig. 1D) and its C-terminal Brd3(420-726) (Fig. 1E) bound MLV IN, whereas the N-terminal Brd3(1-419) failed to bind MLV IN (Fig. 1F). Similar results were seen with recombinant MLV IN binding to recombinant Brd4(1-720) and Brd4(462-720) but not Brd4(1-461) (Fig. S5).

BET Proteins Directly Bind and Stimulate MLV Integration in Vitro. To examine whether the observed interactions were direct or bridged by additional cellular components, we next examined interactions with purified recombinant proteins. Among the three BET proteins, only His-Brd4(1-720) was amenable to bacterial expression and purification. Importantly, Brd4(1-720) contains all of the functional domains and motifs conserved between Brd2, -3, and -4 (Fig. S4) and, thus, serves as a representative of all three BET proteins. The data in Fig. 2A show that Brd4(1-720) binds MLV but not HIV-1 IN. In control experiments, LEDGF/p75 interacted with HIV-1 but not MLV IN (Fig. 2B). Using pull-down experiments, the apparent $K_d$ value was $\sim$33 nM for MLV IN-Brd4(1-720) (Fig. 2C and D). A $K_d$ of $\sim$200 nM was measured for the HIV-1 IN-LEDGF/p75 interaction (29).

We next examined the effects of Brd4(1-720) in vitro-integration assays catalyzed by MLV IN (Fig. 2E and F). Addition of Brd4(1-720) to the reactions significantly enhanced the biologically relevant concerted two-end integration products. Comparative analysis of strand-transfer activities of MLV and HIV-1 INs revealed that the levels of stimulation for the integration activities of MLV IN by Brd4(1-720) and HIV-1 IN by LEDGF/p75 were comparable [Fig. 2F and G; $\sim$564% for MLV IN+Brd4(1-720) and $\sim$689% for HIV-1 IN+LEDGF/p75]. In control experiments, no stimulation of strand transfer activities of MLV IN by LEDGF/p75 or HIV-1 IN by Brd4(1-720) were

Fig. 1. BET proteins specifically interact with MLV integrase. (A) GST pull-down of NIH 3T3 cell lysate with either GST-MLV IN or GST-HIV-1 IN and immunoblotting with Brd2, -3, and -4 antibodies. (B) Coomassie-stained SDS/PAGE gel of GST pull-down products from A showing that similar levels of GST-MLV IN or GST-HIV-1 IN bound to glutathione Sepharose beads. (C) GST pull-down of HEK293T cell lysate expressing FLAG-Brd3(1-726) with GST-MLV IN, GST-MLV IN CTD, GST-MLV IN NTD, and GST-MLV IN NTD-CCD and immunoblotting with FLAG antibody. (D-F) Affinity pull-down with GST-MLV IN and HEK293T cell lysate expressing FLAG-Brd3(1-726) (D), GFP or GFP-Brd3(420-726) (E), and FLAG-Brd3(1-419) (F), respectively. Immunoblotting with FLAG or GFP antibody. "Input lysate (10%)" indicates 10% of indicated whole-cell lysate used for pull-down. "Beads + lysate" indicates control pull-down without GST-tagged protein.
observed (Fig. 2 F and G). Thus, BET proteins and LEDGF/p75 specifically stimulate MLV and HIV-1 integration, respectively.

Antagonism of BET Proteins Reduces MLV Integration in Infected Cells. To elucidate the role of the BET proteins in MLV replication, we down-regulated these cellular proteins using either siRNA or shRNA. The individual down-regulation of BET proteins using a siRNA approach impaired MLV but not HIV-1 expression by ~25% for Brd2, ~32% for Brd3, and ~27% for Brd4 (Fig. S6 A–C). Concurrent down-regulation of all three BET proteins by siRNA impaired MLV but not HIV-1 expression by ~37% (Fig. S6 D–F). In contrast, as reported previously (30, 31), HIV-1 expression was enhanced upon Brd4 down-regulation (Fig. S6 C and F). More effective down-regulation of individual BET proteins was obtained with shRNA, which impaired MLV but not HIV-1 expression by ~39% for Brd2, ~37% for Brd3, and ~49% for Brd4 (Fig. S7 A and B). Our results in Fig. 1 suggest that Brd2, -3, and -4 are redundant in interaction with MLV IN. Attempts to knock down all three proteins concurrently using shRNA were lethal to the cells.

Therefore, we used the cell-permeable small molecule JQ-1, which selectively inhibits interactions of all three BET proteins with cognate-modified histone sites (32). JQ-1 inhibited MLV expression in a dose-dependent manner (Fig. 3A). In contrast, as expected (30, 31), HIV-1 expression was enhanced in JQ-1–treated cells. Thus, our results indicate that the specific inhibition of BET proteins selectively impairs MLV expression.

To pinpoint the replication steps affected by JQ-1 treatment, we quantified viral DNA forms longitudinally, including the minus-strand strong-stop extension products (MSSEs), plus-strand extension products (PSEs), 2-LTR circles, and integrated proviruses (33). JQ-1 treatment did not alter MSSEs or PSEs (Fig. 3 B and C), indicating that MLV reverse transcription was not affected. In contrast, a significant increase in 2-LTR circles was observed upon JQ-1 treatment (Fig. 3D). These dead-end products serve as surrogate markers and are known to increase in quantity in the presence of defective integration (33). Quantitation of integrated proviruses by Alu-based quantitative (q) PCR revealed a significant reduction in integrated MLV proviruses upon treatment with JQ-1 (Fig. 3E). In contrast, JQ-1 treatment had no significant effects on HIV-1 late reverse-transcription products (Fig. 3F), 2-LTR circles (Fig. 3G), or integrated proviruses (Fig. 3H).

Additionally, we evaluated the effects of JQ-1 on the integration of a MLV-based retroviral vector. JQ-1 treatment significantly inhibited integration in a dose-dependent manner, which correlated closely with the levels of reduction of MLV vector expression (Fig. 3 I and J). Taken together, these results indicate that the inhibition of the BET proteins by JQ-1 selectively impaired MLV integration and, hence, subsequent gene expression.

BET Proteins Target MLV Integration to Transcription Start Sites. We next evaluated whether there is a correlation between MLV-integration sites and the chromatin-binding sites of BET proteins mapped using ChIP-Seq data (21). For this, MLV-integration sites in HEK293T cells were compared within a 1-kb window to Brd2, -3, and -4 binding sites (21). Fig. 4A shows that compared with HIV-1 or avian sarcoma leukemia virus (ASLV), MLV significantly favored integration near-binding sites of the BET proteins.
In contrast, MLV did not favor integration near-binding sites for heterochromatin protein 1 (HP1α and HP1β), which were mapped in the same study, are known to be enriched in heterochromatin and serve here as controls. We next examined MLV-integration sites near the promoters (within 1-kb window) bound by the BET proteins. The heatmap in Fig. 4B shows that in comparison with HIV-1 or ASLV, MLV significantly favors integration near promoters associated with the BET proteins. In contrast, MLV did not favor integration near promoters bound by HP1α and HP1β.

To dissect the role of BET proteins in MLV integration-site selection, we analyzed the distribution of 11,968 unique integration sites in cells treated with JQ-1 or a pool of Brd2, -3, and -4 siRNAs [Brd(2+3+4)i] by 454 pyrosequencing. As expected (6, 34) in control experiments with either no inhibitor or scrambled siRNA (Sci) MLV integration was favored (39% of integration events) within 2-kb distance from RefSeq transcription start sites (Fig. 4 C–F). The JQ-1 treatment significantly reduced the frequency of MLV integration at transcription start sites in a dose-dependent manner (Fig. 4 C and D). Moreover, concurrent down-regulation of all three BET proteins by siRNA also significantly reduced the frequency of MLV integration at transcription start sites (Fig. 4 E and F). The residual MLV integration at transcription start sites observed with JQ-1 or siRNA treatments is likely attributable to suboptimal inhibition or down-regulation of BET proteins, respectively. Additionally, other cellular or viral factors may also contribute to residual targeting. Taken together, these results indicate that BET proteins target MLV integration to transcription start sites.

**Discussion**

The distributions of integrated proviruses in host genomes are not random and appear to be genus-specific. For example, γ-retroviruses such as MLV exhibit strong bias for integrating in the vicinity of transcription start sites and CpG islands (6), whereas lentiviruses, including HIV-1, prefer to integrate into active genes (12, 35). Previous studies (7) with chimeric viruses have shown that IN is the principal viral determinant for integration-site selectivity. The distribution of integration sites for a chimeric HIV-1 virus with the IN coding region replaced with its MLV counterpart was markedly different from wild-type HIV-1 but closely resembled MLV (7).

The mechanism of integration-site targeting is best understood for HIV-1. The site selectivity of HIV-1 integration is controlled by cellular protein LEDGF/p75 (9, 11, 12), which functions as a bimodal tether. The LEDGF/p75 C-terminal region directly interacts with MLV IN (9, 11, 12), which functions as a bimodal tether. The LEDGF/p75 C-terminal region directly engages lentiviral INs, whereas its N-terminal region containing the PWWP domain recognizes trimethylated H3 tails in chromatin (36–38) and, accordingly, directs HIV-1 integration into actively transcribed genes (9–12).

Our present studies reveal that the BET protein-mediated interaction links MLV IN to transcriptional start sites in chromatin. In particular, our data suggest that the BET proteins act as bimodal tethers (Fig. S8), with the C-terminal fragment directly interacting with MLV IN (Fig. 1) and the N-terminal bromodomains binding to acetylated H3 and H4 tails (32), which are found at the transcription start sites and strongly correlate with MLV-integration sites (34) in genomic DNA. These findings support the notion that INs from different retroviral genera...
have adopted different chromatin-binding tethers to effectively integrate viral DNAs into specific features of chromatin.

BET proteins have been implicated in numerous aspects of medicine such as cancer, inflammation, obesity, and HIV latency (26–28, 30, 31, 39). Most relevant to this study, Brd4 acts as an attachment site for the tethering of papilloma viral genomes to the mitotic chromosomes. The E2 protein of bovine papilloma virus (BPV) tethers the viral genome to mitotic chromosomes via bimodal interaction of its C-terminal DNA-binding domain with viral DNA and the N-terminal transactivation domain with the

Fig. 4. Antagonism of BET proteins reduces MLV-integration frequencies at the transcription start sites. (A and B) Analysis of integration frequencies of ASLV, HIV-1, and MLV relative to BET proteins (Brd2, -3, and -4) or HP1αβ chromatin sites in HEK293T cells. The chromatin sites and promoters bound by BET proteins or HP1αβ were quantified using ChIP-Seq data (21). (A) Heatmap depicting association of integration sites with BET proteins or HP1αβ chromatin sites. (B) Heatmap depicting association of integration sites with promoters bound by BET proteins or HP1αβ. The frequency of integration sites relative to the matched random controls was quantified using the receiver operating characteristic area method (44). The color key depicts enrichment or depletion of chromatin sites or promoters bound by indicated protein near integration sites. All comparisons of MLV to HIV or ASLV achieved P < 0.001 (Wald statistic). (C–F) Percentage of MLV-integration sites found within each interval surrounding RefSeq transcription start sites (TSSs) in HEK293T cells. The integration sites near TSSs were compiled onto a single start site, and the frequencies were mapped. The x axis depicts the distance (in kb) relative to the TSSs (set at 0). The y axis depicts the percentage of integration sites in the indicated window. For comparison, integration sites of HIV-1 (9) and MLV (45, 46) in HEK293T cells are shown. (C) Dose-dependent effect of JQ-1 on MLV-integration frequencies at the TSSs. Percentage of MLV-integration sites within each interval surrounding TSSs in HEK293T cells treated with indicated concentrations of JQ-1 inhibitor or DMSO (indicated as “0 nM JQ-1”). (D) Percentage of MLV integration within 2-kb distance from TSSs. All samples achieved statistical significant (**P < 0.001; Fisher’s exact test) compared with 0 nM JQ-1 treatment. (E) Effect of concurrent down-regulation of BET proteins on MLV-integration frequencies at the TSSs. Percentage of MLV-integration sites within each interval surrounding TSSs in HEK293T cells transfected with scrambled siRNA (indicated as “Sci”) or a pool of Brd2, -3, and -4 siRNAs (indicated as “Brd(2+3+4)”). (F) Percentage of MLV integration within 2-kb distance from TSSs. Brd(2+3+4) achieved statistical significant (**P = 0.009; Fisher’s exact test) compared with Sci.
C-terminal region of Brd4 (40, 41). This E2–Brd4 interaction plays a critical role for BPV genome segregation by ensuring that the BPV epimodal DNA is retained in the nucleus after cell division.

In this report, we have uncovered a role for BET proteins in promoting and targeting MLV integration to transcription start sites. These findings will facilitate the development of safer MLV-based vectors for human gene therapy. For example, the inclusion of the experimental drug JO-1, in combination with MLV-based vectors, could minimize the risks of insertional activation of protooncogenes.

Materials and Methods

MS-Based Proteomic Analysis. MS experiments and data analysis were performed as described previously (42). See SI Materials and Methods for more details.

Plasmids and Cloning. C-terminal FLAG-tagged human full-length Brd3(1-726) (amino acids 1–726; NM_007371) and Brd4(1-720) (amino acids 1–720; NM_014299) were purchased from Origene. Truncated Brd3(1-419) (amino acids 1–419) and Brd4(1-461) (amino acids 1–461) (N-terminal fragments containing two bromodomains) and Brd3(420-726) (amino acids 420–726) and Brd4(462-720) (amino acids 462–720) (C-terminal fragments containing the protein-protein interaction domains) were subcloned into pCMV6-AC-GFP or pEX-N-His (Origene). GST-MLV IN NTD/catalytic core domain (CCD) (amino acids 1–408), GST-MLV IN N-terminal domain (NTD) (amino acids 1–105), and GST-MLV IN C-terminal domain (CTD) (amino acids 286–408) were made by subcloning into pGEX-2T. GST-MLV IN NTD/catalytic core domain (CCD) (amino acids 1–290) was generated by the addition of a stop codon at amino acid 291. Other clones were as described: His-lens epithelium-derived growth factor (LEDGF/p75) (1) and His-MLV IN (2).

Expression and Purification of Recombinant Proteins. All proteins were expressed in BL-21(DE3) cells at 37 °C and induced for 3.5 h with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). All cell pellets were lysed by sonication and clarified by centrifugation (20,000 × g for 45 min). His-Brd4(1-720), His-Brd4(1-461), and His-Brd4(462-720) were purified using a 20–500 mM imidazole gradient on a HisTrap HP column. GST-MLV IN, GST-MLV IN NTD, GST-MLV IN CTD, and GST-MLV IN NTD-CCD were gradually adjusted to 0.8 M (NH₄)₂SO₄, loaded onto a phenyl Sepharose column, and eluted with decreasing gradient of (NH₄)₂SO₄. GST–HIV-1 IN was loaded onto glutathione Sepharose beads and eluted with 50 mM glutathione. Pooled fractions of all but GST-MLV IN CTD were diluted to 125 mM NaCl, loaded onto a Heparin column, and eluted with a gradient of 0–1 M NaCl. GST-MLV IN CTD was diluted to 50 mM NaCl, loaded onto a CM Sepharose column, and eluted with a gradient of 0–1 M NaCl. His-Brd4(1-720), His-Brd4(1-461), and His-Brd4(462-720) were further purified with size-exclusion chromatography using a HiLoad 16/60 Superdex 200 column. Other proteins were purified as described; His-LEDGF/p75 and His-MLV IN using a protocol developed for His-HIV-1 IN (1). All chromatography columns were purchased from GE Healthcare.

Immunoblotting. Standard immunoblotting procedures were used with the following antibodies: Brd2 (catalog no. 5848; Cell Signaling Technology), Brd3 (catalog no. sc-81202; Santa Cruz Biotechnology), Brd4 (catalog no. 75898; Abcam), GAPDH (catalog no. AHP 1628T; Serotec), FLAG (catalog no. TA100023; Origene), and TurboGFP (catalog no. AB513; Evrogen).

Strand-Transfer Assays. The in vitro-concerted integration assays were carried out as described previously for MLV IN (2) and HIV-1 IN (3), with the minor changes. The assays contained 5′-Cy5-labeled viral donor DNA (1 μM), target DNA (pBR322, 300 ng), purified recombinant His-MLV IN (0.3 μM), and purified recombinant His-Brd4(1-720) (indicated concentrations). Assays were performed by the addition of purified recombinant His-Brd4(1-720) to the His-MLV IN, donor, and target DNA substrates. The reaction products were separated by agarose gel (1.25%) electrophoresis and visualized by Typhoon 9410 Imager (GE Healthcare) set at 670-nm band-pass filter to visualize the Cy5 signal.

For accurate, in-solution, quantitative analysis of the strand-transfer reaction products, homogeneous time-resolved fluorescence (HTRF)-based assays were carried out, using previously reported reaction conditions (1, 2). The assays contained 5′-Cy5-labeled viral donor DNA (50 nM), biotin-labeled target DNA (10 nM), purified recombinant His-MLV IN, or His-HIV-1 IN (200 nM) and purified recombinant His-Brd4(1-720) or His-LEDGF/p75 (2 μM). Assays were performed by the addition of purified recombinant His-Brd4(1-720) or His-LEDGF/p75 to the respective reactions containing IN, donor, and target DNA substrates. The strand-transfer products were detected after addition of europium chelate–streptavidin–Lance reagent (2 nM; PerkinElmer). The HTRF signal was recorded using a PerkinElmer Multimode Enspire plate reader using 314 nm for excitation wavelength and 668 and 620 nm for the wavelength of the acceptor and donor emission, respectively.

Cells, Viruses, Viral Vectors, Transfections, and Drug Treatment. HEK293T and NIH 3T3 cells were cultured in Dulbecco’s modified eagle medium (Invitrogen), 10% (vol/vol) FBS (Invitrogen), and 1% (vol/vol) antibiotic (Gibco) at 37 °C and 5% CO₂. All viral vectors and viruses were generated by transfecting HEK293T cells using either TransIT-2020 (Mirus) or X-tremeGENE HP (Roche) transfection reagents following the manufacturer’s protocol. Vector- and virus-containing supernatant was collected after 48 h, filtered, and, if needed, concentrated by ultracentrifugation. To generate MLV LTR–luciferase (Luc) vector, the following plasmids were used: pFB-Luc (MLV vector plasmid; Agilent Technologies), pHIT60 (packaging construct), and pMD.G (vesicular stomatitis virus-glycoprotein G (VSV-G) envelope plasmid). MLV LTR-GFP vector was generated as described previously using pMLV LTR-GFP (MLV vector plasmid), pCGP (packaging construct), and pMD.G (VSV-G envelope plasmid) (4). MLV pseudotyped with VSV-G was generated using pNCA-C EnvNeo⁺ (5) and pMD.G. Luciferase reporter HIV-1 (HIV-1–Luc) pseudotyped with VSV-G was generated using pNL4-3.LucEnv⁺ and pMD.G. All transductions and infections were carried out in the presence of 8 μg/mL Polybrene (Sigma).

JQ-1 was purchased from Cayman Chemical. Cells were preincubated with JQ-1 or diluent (DMSO) for 2 h, and subsequent viral infections or transductions were performed in the presence of JQ-1. Cells were then cultured in the presence of JQ-1 for indicated time before analysis. Luciferase activity was detected using a commercially available kit (Promega), and cell extracts were prepared using 1× reporter lysis buffer (Promega).
Quantitative PCR Analysis. Levels of MLV minus-strand strongest products, plus-strand extension products (PSEs), 2-LTR circles, and integrated copies of provirus (Alu-PCR reactions) were quantified as described previously (5). Levels of HIV-1 late reverse-transcription products, 2-LTR circles, and integrated copies of provirus (Alu-PCR reactions) were quantified as described previously (6). All viral and vector stocks were treated with DNaseI (60 U/mL; Ambion) before infections to avoid plasmid DNA contamination. DNA from infected cells was isolated at indicated times using a DNeasy Blood & Tissue kit (Qiagen). TaqMan-based quantitative (q) PCR reactions were prepared using iTaq Supermix (Bio-Rad). All reactions were normalized to GAPDH using SYBR Green-based (iQ SYBR Green Supermix; Bio-Rad) qPCR primers: GAPDH forward (5′-TGG ATA TTG CCA TCA ATG ACC-3′) and GAPDH reverse (5′-GAT GGC ATG GAC TGT GGT CAT G-3′). Relative quantification analysis was performed using the 2−ΔΔCT method (7).

siRNA and shRNA. siRNA pools targeting Brd2 (ON-TARGETplus SMARTpool L-004935-00-0010), Brd3 (ON-TARGETplus SMARTpool L-004936-00-0010), Brd4 (ON-TARGETplus SMARTpool L-004937-00-0010), and nontargeting siRNA pool (ON-TARGETplus Nontargeting pool-0001810-10-20) were purchased from Dharmacon. HEK293T cells were seeded in a six-well plate and transfected twice with 100 nM siRNA pool at 24 and 48 h postseeding. The cells were then transduced or infected with the indicated vector or virus at 72 h postseeding. Knock-down of Brd2, -3, and -4, 100 nM each siRNA pool (total of 300 nM siRNA) were used. All siRNA transfections were performed using Oligofectamine (Invitrogen) and OptiMEM (Gibco) following the manufacturer’s protocol.

shRNA pools targeting Brd2 [The RNAi Consortium (TRC) lentiviral shRNA RHS4533-EG4064], Brd4 [TRC lentiviral shRNA RHS4533-EG8019], Brd4 [TRC lentiviral shRNA RHS4533-EG23476], and nontargeting shRNA (TRC lentiviral Lentiviral pKO.1 Empty Vector RHS4080) were purchased from Dharmacon. The Brd2 shRNA pool contained six unique shRNAs (TRCN0000006308, TRCN0000006309, TRCN0000006310, TRCN0000006311, TRCN0000006312, and TRCN0000023963). The Brd3 shRNA pool contained five unique shRNAs (TRCN0000021374, TRCN0000021375, TRCN0000021376, TRCN0000021377, and TRCN0000021378). The Brd4 shRNA pool contained six unique shRNAs (TRCN0000021424, TRCN0000021425, TRCN0000021426, TRCN0000021427, TRCN0000088478, and TRCN00000088479). The pLKO.1 Empty Vector contains the scrambled stuffer sequence (5′-ACCCGGACACTGAGCAGCTTTTTTGAAATTC-3′).

Isolation of Integration Sites and Analysis of Integration-Site Distribution. Isolation of MLV-integration sites was performed as described previously (4), with some changes. Briefly, HEK293T cells with indicated treatment were transduced with MLV LTR-GFP vector. Cells were harvested 15 d posttransduction, and the genomic DNA was purified using DNeasy Blood & Tissue kit (Qiagen). MLV-integration sites were isolated using ligation-mediated PCR. Genomic DNA was fragmented with dsDNA Fragmentase (NEB), followed by linker ligation. Host-vector DNA junctions were amplified by nested PCR using bar-coded primers. Purified PCR products were sequenced on the 454 GS-Junior (Roche). Three independent transductions were performed for each sample, and sample triplicates were separately bar coded with the second pair of PCR primers.

The 454 sequencing reads were first decoded by requiring a perfect match to the sample barcode and were subsequently trimmed. The resulting collections of reads were quality-filtered by requiring a 95% match to the LTR primer and a 100% match to the flanking LTR region. Only sequences that began within 3 bp of the LTR end and showed unique best alignments to the human genome by BLAST-like alignment tool (BLAT) (hg18, Version 36.1; >98% match score) were considered as authentic integration sites.

Bioinformatics and Statistical Analyses. ChiP-Seq data on bromo-domain and extraternal domain (BET) proteins and HP1 proteins was from Leroy et al. (8). The data were curated by removing any peaks which overlapped with the control data. Avian sarcoma leukosis virus (ASLV)-, HIV-1-, and MLV-integration sites (9–12) and matched random controls were tested for proximity within 1 kb of ChiP-Seq annotation and summarized using the receiver operating characteristic (ROC) area method as described previously (4, 13). The ROC value of each comparison is depicted as a tile in a heatmap. Statistical significance in heat maps was assessed to determine whether the ROC areas calculated were significantly different from one another or from 0.5 (matched random controls). All of the tests rely on the variance–covariance matrix of the relative ranks of the integration sites to construct Wald-type test statistics. For comparisons between integration-site sets for specific genomic features, the Wald statistics were calculated and referred to the χ2 distribution to obtain P values as described (12).

12. Brady T, et al. (2009) Integration target site selection by a resurrected human genome by BLAST-like alignment tool (BLAT) (hg18, Version 36.1; >98% match score) were considered as authentic integration sites.
Fig. S1. Identification of BET proteins (Brd2, -3, and -4) as GST-MLV IN binding partners in NIH 3T3 cells. (A) Representative MS/MS data for one of the Brd4 peptides is shown. Identified peptide sequences in Brd4 (B), Brd2 (C), and Brd3 (D) are indicated in bold and highlighted in yellow. The amino acid sequence coverage was 23% (317/1400 aa) for Brd4, 39% (315/798 aa) for Brd2, and 26% (190/731 aa) for Brd3.
Fig. S2. Relative amounts of BET proteins (Brd2, -3, and -4) in NIH 3T3 and Sup-T1 cells. Indicated amounts of whole-cell lysates were subjected to SDS/PAGE and immunoblotting with GAPDH and Brd2, -3, and -4 antibodies. GAPDH immunoblot shows equivalent protein loading between the indicated cell lines.

Fig. S3. MLV IN domains used in the in vitro GST pull-down assay. (A) Schematic representation of MLV IN domains. MLV IN is composed of three distinct structural and functional domains. The NTD contains the NTD-extension domain (NED), and the Zn finger (HH-CC type) and has been implicated in multimerization and DNA substrate binding; the CCD contains the three amino acid (DDE) triad that coordinates catalytic Mg$^{2+}$; and the CTD, which could have multiple functions, including binding to DNA substrates and interacting with other proteins. The amino acid residue numbers are indicated for each domain and key residues. Also shown are the schematic representations of the three MLV IN domains fused with GST, which were used in the in vitro GST pull-down assays: “GST-MLV IN NTD” (containing only the NTD of MLV IN), “GST-MLV IN NTD-CCD” (a two-domain construct containing the NTD and the CCD), and “GST-MLV IN CTD” (containing only the CTD of MLV IN). (B) Similar levels of GST-MLV IN or indicated domains bound to glutathione Sepharose beads were used in the in vitro GST pull-down assay. Purified recombinant GST-MLV IN (lane 1), GST-MLV IN NTD (lane 2), GST-MLV IN NTD-CCD (lane 3), and GST-MLV IN CTD (lane 4) were bound to glutathione Sepharose beads and analyzed by SDS/PAGE. Coomassie-stained image of a representative gel is shown. In parallel, the bound proteins were also used for the in vitro GST pull-down of HEK293T cell lysate expressing FLAG-Brd3(1-726), and the pull-down products were subjected to SDS/PAGE and immunoblotted with FLAG antibody as shown in Fig. 1D.
Fig. S4. Features and domain organization of BET proteins. (A) Phylogenetic analysis of indicated bromodomain-containing proteins (BRD). Phylogenetic tree was generated using protein sequences of indicated Brd proteins and Phylogeny.fr One Click software. As shown, the BET proteins (Brd2, -3, and -4) cluster together. These three proteins belong to the BET (bromodomains and extraterminal domain) protein family (Brd2, -3, -4, and -T). The defining characteristics of this protein family are the presence of dual-tandem bromodomains and an extraterminal domain (ET). The extended BET family consists of Brd1, -2, -3, -4, -7, -8, -9, and -T. The majority of work on bromodomain-containing proteins has been directed to Brd2, -3, -4, and -T. Brd2, -3, and -4 are ubiquitously expressed and have been implicated in cell-cycle control, DNA replication, and transcriptional control, whereas BrdT has expression limited to testis only. (B) Domain organization of BET proteins. The two N-terminal bromodomains I and II (BD I and BD II) are known to bind acetylated histones H3 and H4. Motifs “A” and “B” consists of conserved regions among the BET proteins. The C-terminal region contains the ET and the Ser/Glu/Asp-rich region (SEED) domains known to interact with transcription factors, histone modification enzymes, and chromatin-modifying factors, as well as proteins from various viral families. The SEED is a conserved domain with glutamic and aspartic acid residues interspersed between polyserine residues. Two isoforms of human Brd4 are known to be expressed, a short isoform (amino acids 1–720) and a long isoform (amino acids 1–1362; shown above), which contains a C-terminal motif (CTM) motif of ∼38 aa. This CTM motif is largely unstructured and could play a role in protein–protein interactions. The numbers indicate the amino acid coordinates of different domains and motifs in BET proteins. (C) Protein sequence alignment of the C-terminal region of human Brd2, -3, and -4 showing the conserved domains and motifs. Alignments were performed with Clustal Omega using default settings and the following fragments: Brd2(458-801), Brd3(420-726), and Brd4(462-720) (UniProt accession nos. P25440, Q15059, and O60885, respectively). Motif B, ET, and SEED domains are highlighted in red, green, and yellow, respectively.
Fig. S5. Brd4(462-720) is sufficient to interact with MLV IN. To map the Brd4 region directly interacting with purified recombinant GST-MLV IN, purified recombinant His-Brd4(1-720), His-Brd4(1-461), and His-Brd4(462-720) were used in the in vitro GST pull-down assay. The His-Brd4(1-461) contains the N-terminal bromodomains known to associate with chromatin, whereas the His-Brd4(462-720) C-terminal fragment containing Motif B, ET, and SEED domains has been implicated in interactions with different proteins. Purified recombinant GST-MLV IN was bound to glutathione Sepharose beads and then incubated with 5 μM His-Brd4(1-720), His-Brd4(1-461), or His-Brd4(462-720). Control pull-down reactions without GST-MLV IN but with indicated recombinant His-Brd4 were also examined. The pull-down products were analyzed by SDS/PAGE. Coo massie-stained gel of a representative GST pull-downs is shown. *Input* indicates 10% of indicated recombinant Brd4 protein used for the pull-downs. **“GST-MLV IN”** input indicates one-third load of GST-MLV IN used for the pull-downs.

Fig. S6. siRNA-mediated down-regulation of BET proteins specifically reduces MLV expression. (A) siRNA-mediated down-regulation of individual BET proteins. HEK293T cells were transfected consecutively with 100 nM nonsilencing scrambled (Sci) or Brd2, -3, or -4 siRNAs (Brd2i, Brd3i, Brd4i) for 48 h. Equivalent whole-cell lysates were subjected to SDS/PAGE and immunoblotted with GAPDH, Brd2, -3, and -4 antibodies to verify down-regulation of BET proteins. GAPDH immunoblot shows equal protein loading. Numbers below the images show relative amounts of protein as measured by ImageJ software. (B and C) siRNA-mediated down-regulation of individual BET proteins modestly but specifically reduces MLV expression. The siRNA-transfected cells from A were transduced with VSV-G pseudotyped MLV LTR-Luc vector (B) or HIV-1–Luc (C). Luciferase assay was performed at 24 h posttransduction. The luciferase signal obtained for the nonsilencing scrambled control (Sci) was set to 100%. Bars represent means ± SD (n = 3; ***P < 0.0001 measured by one-way ANOVA; multiple comparisons of the individual siRNA treatment to the Sci control used Dunnett simultaneous test). (D) Concurrent down-regulation of BET proteins. HEK293T cells were transfected consecutively with 300 nM Sci or pool of Brd2, -3, and -4 siRNAs (indicated as “Brd(2+3+4)i”) for 48 h. Equivalent whole-cell lysates were immunoblotted as done in A. Numbers below the images show relative amounts of protein as measured by ImageJ software. (E and F) Concurrent down-regulation of BET proteins specifically reduces MLV expression. In parallel, siRNA-transfected cells from D were transduced with VSV-G–pseudotyped MLV LTR-Luc vector (E) or HIV-1–Luc (F). Luciferase assay was performed as done in B and C. Bars represent means ± SD (n = 3; *P < 0.05; **P < 0.01; ***P < 0.001 by Student t test).
Fig. S7. shRNA-mediated down-regulation of BET proteins specifically reduces MLV expression. HEK293T cells were transduced with lentiviral vectors encoding shRNAs directed against nonsilencing scrambled (Scram) or Brd2, -3, or -4 for 48 h. Cells were then treated with puromycin for an additional 72 h to select for transduced cells. (A) Equivalent whole-cell lysates were subjected to SDS/PAGE and immunoblotted with GAPDH or Brd2, -3, and -4 antibodies to verify down-regulation of BET proteins. GAPDH immunoblot shows equal protein loading. (B and C) Puromycin-selected cells were transduced with VSV-G-pseudotyped MLV LTR-Luc vector (B) or HIV-1-Luc (C). Luciferase assay was performed at 24 h posttransduction. The luciferase signal obtained for the nonsilencing scrambled control (Scram) was set to 100%. All bars represent means ± SD (n = 3; ***P < 0.0001 measured by one-way ANOVA; multiple comparisons of the individual shRNA treatment to the Scram control used Dunnett simultaneous test).
Fig. S8. The proposed role of the BET proteins in MLV integration. The proposed model shows that the BET proteins could act as bimodal tethers with the C-terminal fragment directly interacting with MLV intasome (MLV IN complex with viral DNA) and the N-terminal bromodomains binding to acetylated H3 and H4 tails, which are found at the transcription start sites and strongly correlate with MLV-integration sites in the genomic DNA.

Table S1. List of the top 10 unique protein hits from NIH 3T3 cells specifically binding GST-MLV IN or GST-HIV-1 IN

<table>
<thead>
<tr>
<th>Identified proteins</th>
<th>Accession no.</th>
<th>Molecular mass, kDa</th>
<th>HIV-1 IN</th>
<th>MLV IN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GST-MLV IN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Bromodomain-containing protein BRD4</td>
<td>gi</td>
<td>18308125 (+3)</td>
<td>156</td>
<td>ND</td>
</tr>
<tr>
<td>2. Bromodomain-containing protein BRD2</td>
<td>gi</td>
<td>71067345 (+2)</td>
<td>88</td>
<td>ND</td>
</tr>
<tr>
<td>3. Bromodomain containing protein BRD3</td>
<td>gi</td>
<td>148676419 (+2)</td>
<td>80</td>
<td>ND</td>
</tr>
<tr>
<td>4. Splicing factor 3B subunit 3</td>
<td>gi</td>
<td>19527174</td>
<td>136</td>
<td>ND</td>
</tr>
<tr>
<td>5. Splicing factor, arginine/serine-rich 7, 35 kDa, isoform CRA_a</td>
<td>gi</td>
<td>119620768 (+16)</td>
<td>16</td>
<td>ND</td>
</tr>
<tr>
<td>6. 60S acidic ribosomal protein P2</td>
<td>gi</td>
<td>83745120</td>
<td>12</td>
<td>ND</td>
</tr>
<tr>
<td>7. mCG50680</td>
<td>gi</td>
<td>148670393 (+5)</td>
<td>37</td>
<td>ND</td>
</tr>
<tr>
<td>8. mCG121875</td>
<td>gi</td>
<td>148676875 (+8)</td>
<td>30</td>
<td>ND</td>
</tr>
<tr>
<td>9. mCG8513</td>
<td>gi</td>
<td>148708252 (+4)</td>
<td>26</td>
<td>ND</td>
</tr>
<tr>
<td>10. 60S ribosomal protein L18</td>
<td>gi</td>
<td>12840700 (+5)</td>
<td>22</td>
<td>ND</td>
</tr>
<tr>
<td><strong>GST–HIV-1 IN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. LEDGF/p75, PC4 and SFRS1-interacting protein</td>
<td>gi</td>
<td>19527168</td>
<td>60</td>
<td>35</td>
</tr>
<tr>
<td>2. Protein virilizer homolog</td>
<td>gi</td>
<td>124487145 (+4)</td>
<td>207</td>
<td>16</td>
</tr>
<tr>
<td>3. Collagen α-1(I) chain precursor</td>
<td>gi</td>
<td>34328108 (+1)</td>
<td>138</td>
<td>11</td>
</tr>
<tr>
<td>4. Probable ATP-dependent RNA helicase DDX17 isoform 1</td>
<td>gi</td>
<td>40068493 (+1)</td>
<td>73</td>
<td>9</td>
</tr>
<tr>
<td>5. Zinc finger CCCH type containing 13</td>
<td>gi</td>
<td>223461012 (+1)</td>
<td>204</td>
<td>7</td>
</tr>
<tr>
<td>6. Myosin, light chain 12A</td>
<td>gi</td>
<td>12851268 (+4)</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>7. Caldesmon 1, isoform CRA_a</td>
<td>gi</td>
<td>148681748 (+2)</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>8. Sister chromatid cohesion protein PDS5 homolog B</td>
<td>gi</td>
<td>263346616 (+3)</td>
<td>79</td>
<td>3</td>
</tr>
<tr>
<td>9. SAFB-like, transcription modulator, isoform CRA_a</td>
<td>gi</td>
<td>148694252 (+5)</td>
<td>115</td>
<td>3</td>
</tr>
<tr>
<td>10. Unconventional myosin-lc isoform a</td>
<td>gi</td>
<td>124494242 (+3)</td>
<td>120</td>
<td>2</td>
</tr>
</tbody>
</table>

The Mascot search engine and Scaffold 3 software were used to identify, list, and rank the protein hits. The following settings of the Scaffold 3 viewer were used: minimum identification probability for proteins, 99%; minimum peptide number, 2; and minimum identification probability for peptides, 95%. A total of 185 proteins were identified, with 118 of these proteins (~64% overlap) being found in both MLV and HIV-1 IN samples. Of the remaining 67 unique protein hits, the top 10 proteins that specifically bind GST-MLV IN or GST–HIV-1 IN are shown. Contaminating proteins such as human keratin and bacterial proteins were removed from the analysis. Four separate MS-based proteomic experiments were performed: two each with NIH 3T3 and Sup-T1 cells. In all four experiments the BET proteins specifically bound GST-MLV IN and not GST–HIV-1 IN. Conversely, LEDGF/p75 bound GST–HIV-1 IN and not GST-MLV IN. The unweighted spectrum count values for identified proteins are shown in the last two columns. ND indicates that the peptides from the listed protein were not detected.