Structural basis for spumavirus GAG tethering to chromatin

Paul Lesbats[a,1], Erik Serrao[b,c], Daniel P. Maskell[a,2], Valerie E. Pye[a], Nicola O'Reilly[d], Dirk Lindemann[e], Alan N. Engelman[b,c,3], and Peter Cherepanov[a,f,3]

[a]Chromatin Structure and Mobile DNA, The Francis Crick Institute, London NW1 1AT, United Kingdom; [b]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215; [c]Department of Medicine, Harvard Medical School, Boston, MA 02115; [d]Peptide Chemistry, The Francis Crick Institute, London NW1 1AT, United Kingdom; [e]Institute of Virology, Technische Universität Dresden, Dresden, 01307, Germany; and [f]Division of Medicine, Imperial College London W2 1PG, United Kingdom

Edited by Paul Bieniasz, Rockefeller University, New York, NY, and accepted by Editorial Board Member Stephen P. Goff April 19, 2017 (received for review December 24, 2016)

The interactions between a retrovirus and host cell chromatin that underlie integration and provirus expression are poorly understood. The prototype foamy virus (PFV) structural protein GAG associates with chromosomes via a chromatin-binding sequence (CBS) located within its C-terminal region. Here, we show that the PFV CBS is essential and sufficient for a direct interaction with nucleosomes and present a crystal structure of the CBS bound to a mononucleosome. The CBS interacts with the histone octamer, engaging the H2A–H2B acidic patch in a manner similar to other acidic patch-binding proteins such as herpesvirus latency-associated nuclear antigen (LANA). Substitutions of the invariant arginine anchor residue in GAG result in global redistribution of PFV and macaque simian foamy virus (SFV_{mac}) integration sites toward centromeres, dampening the resulting proviral expression without affecting the overall efficiency of integration. Our findings underscore the importance of retroviral structural proteins for integration site selection and the avoidance of genomic junkyards.

retrovirus | integration | nucleosome | chromatin-binding

The integration of a reverse-transcribed viral genome into a host cell chromosome is an obligatory step in the retroviral replication cycle (reviewed in ref. 1). The fate of the resulting provirus depends on the hospitality of the local chromatin environment, which can facilitate, moderate, or prevent viral expression (2, 3). It is not surprising, therefore, that retroviruses evolved genus-specific and contrasting traits with respect to the types of chromosomal loci they preferentially target. HIV-1, a highly pathogenic lentivirus, channels integration toward gene-dense chromosomal domains, integrating predominantly within active transcription units (4, 5). Recent studies uncovered a hierarchical mechanism that depends on HIV-1 capsid, the major viral structural protein (a product of the viral *gag* gene), and integrase, which interact with cellular proteins CPSF6 and LEDGF/p75, respectively (6–11). Ablation of the mRNA maturation factor CPSF6 results in the bulk of virus abandoning gene-rich domains, whereas ablation of the transcriptional coactivator LEDGF leads to a reduction of integration into transcription units while largely preserving the virus's ability to locate gene-rich domains (11). Thus, the capsid–CPSF6 and integrase–LEDGF axes appear to direct HIV-1 integration on the global and local scale, respectively. Although other members of the lentivirus genus are less studied, the similarities in their integration site distributions and conservation of the integrase–LEDGF interaction (12–14) suggest a shared strategy. Murine leukemia virus (MLV) and other studied gammaretroviruses similarly target gene-dense domains but tend to integrate in the immediate vicinity of promoters, CpG islands, and DNase-hypersensitive sites (15–18). These properties depend at least in part on the interaction between integrase and bromodomain proteins BRD2–4 (19–21). In contrast, prototype foamy virus (PFV), a member of the generally benign spumavirus retroviral genus, appears to be averse to gene-rich regions and disfavors integration into genes (22). Encouraged by earlier observations that PFV GAG can bind chromatin (23, 24), we determined a crystal structure of its chromatin-binding sequence (CBS) bound to a mononucleosome. We show that the CBS binds at the acidic patch in the surface of H2A–H2B heterodimer via an invariant conserved Arg anchor motif. Mutations disrupting these contacts alter the ability of primate spumavirus GAGs to bind nucleosomes and instigate global redistributions of integration sites into centromeric regions of chromosomes.

Results

Crystal Structure of the PFV GAG CBS in Complex with a Mononucleosome. Chromatin tethering of ectopically expressed PFV GAG depends on the CBS, which is located in the C-terminal region of the protein within Gly/Arg box II (GRII) (Fig. 1*A*) (23, 24). To test if the GAG–chromatin interaction is direct, we purified full-length hexahistidine (His_6)-tagged PFV GAG protein and used recombinant mononucleosomes assembled from bacterially expressed human histones and 147-bp DNA with a strong positioning sequence (25). WT His_6-GAG protein readily pulled down nucleosomes on Ni-nitrilotriacetic

Significance

Spumaviruses are being developed as vectors for gene-therapy applications, but how these retroviruses select genomic locations for integration remains unknown. Here we use X-ray crystallography to visualize the interaction between the spumaviral GAG protein and a nucleosome. We show that this interaction is essential for the observed distribution of spumavirus integration sites in various human cell types. Thus, despite stark differences in the mechanistic details of spumavirus and orthoretrovirus replication strategies, both retroviral subfamilies depend on their structural proteins to locate optimal integration sites.

Author contributions: P.L. and P.C. designed research; P.L., E.S., D.P.M., and N.O. performed research; N.O. and D.L. contributed new reagents/analytic tools; P.L., E.S., D.P.M., V.E.P., D.L., A.N.E., and P.C. analyzed data; and P.L., A.N.E., and P.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.B. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The structure factors and the refined model reported in this paper have been deposited in the Protein Data Bank (PDB) database (accession ID code 5MLU). Integration site sequencing data reported in this paper have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) database (accession ID code GSE97973).

[1]Present address: Microbiologie Fondamentale et Pathogénicité, UMR5234, CNRS-Université de Bordeaux, Structure Fédérative de Recherche Transbiomed, 33076 Bordeaux, France.

[2]Present address: School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom.

[3]To whom correspondence may be addressed. Email:peter.cherepanov@crick.ac.uk or alan_engelman@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1621159114/-/DCSupplemental.

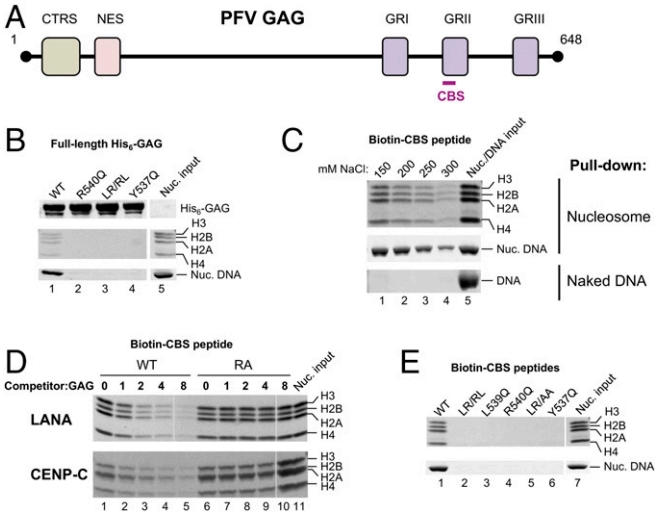

Fig. 1. PFV GAG CBS–nucleosome interaction in vitro. (*A*) Schematic representation of PFV GAG. CBS, chromatin-binding sequence; CTRS, cytoplasmic targeting and retention signal; GRI-III, glycine-arginine box I-III; NES, nuclear export signal. (*B*) His_6-tag pull-down experiments using full-length WT (lane 1) and indicated mutant His_6-GAG constructs (lanes 2–4; LR/RL reversed the natural Leu539–Arg540 sequence to Arg539–Leu540). Protein and DNA recovered on Ni-NTA beads were separated by SDS/PAGE and detected by staining with Coomassie and ethidium bromide, respectively. Migration positions of histones or DNA are indicated to the right of the gel. Lane 5 shows input quantity of nucleosomes used in each pull down. (*C*) Streptavidin pull-down of nucleosomes or naked DNA with biotinylated GAG CBS peptide in the presence of 150–300 mM NaCl (lanes 1–4). (*D*) Lanes 1–5: nucleosomes incubated with biotinylated PFV GAG CBS in the absence or presence of indicated excess WT LANA (*Upper*) or CENP-C (*Lower*) peptide. Lanes 6–10: mutant LANA R9A and CENP-C R717A peptides deficient for nucleosome-binding were used as specificity controls (RA). (*E*) Streptavidin pull-down of nucleosomes with WT (lane 1) or mutant (lanes 2–6; LR/AA: L539A/R540A) biotinylated PFV GAG CBS peptides.

acid (NTA) agarose beads (Fig. 1*B*, lane 1). Furthermore, biotinylated CBS peptide efficiently pulled down the nucleosomes but not free DNA on streptavidin agarose beads (Fig. 1*C*), indicating that the CBS is minimally sufficient for the interaction. To visualize the CBS–chromatin interface, we soaked nucleosome crystals in the presence of a synthetic peptide comprising PFV GAG residues 535–550 and collected X-ray diffraction data to 2.8-Å resolution (*SI Appendix*, Table S1). The structure was solved by molecular replacement, and the initial difference *Fo–Fc* Fourier map revealed positive density for the bound peptide (*SI Appendix*, Fig. S1*A*). Sixteen residues of PFV GAG (Gly535–Gly550) spanning the CBS region could be built into the map, and the model was refined to an R_{free} of 25.1% (*SI Appendix*, Fig. S1*B* and Table S1). The peptide adopts an extended conformation, tracking across the protein side of the nucleosome core particle (Fig. 2*A*). A molecular surface area of 2,070 $Å^2$ is buried upon the formation of the CBS–nucleosome complex, involving one H2A–H2B heterodimer and both H3 chains. The interactions with the H2A–H2B heterodimer account for >75% of the GAG–nucleosome contact area. Here, the side chain of PFV GAG Arg540 projects into the H2A–H2B acidic patch to interact with H2A carboxylates Glu61, Asp90, and Glu92 (Fig. 2*B*). GAG Tyr537 and Leu539 make hydrophobic contacts with H2A residues Tyr57 and Ala60 as well as with H2B Val41 and Val45. The interactions with the H2A–H2B heterodimer are further supported by hydrogen bonds between the hydroxyl group of GAG Tyr537 and the side chains of H2B Gln44 and H2A Glu56 and, additionally, between the main chain carbonyl of GAG Tyr544 and the amide of H2A Glu91. CBS–histone H3 contacts primarily involve GAG Tyr549, which makes hydrophobic interactions with Leu109, Leu126, and Arg129 and also hydrogen bonds with His113 and the main chain carboxylate of the C-terminal Ala135.

The CBS–Nucleosome Interface Is Essential for the Interaction with Nucleosomes and Chromatin. A number of chromatin-binding factors interact with the histone face of the nucleosome (26), including Kaposi's sarcoma herpesvirus (KSHV) latency-associated nuclear antigen (LANA) (27) and centromere protein C (CENP-C) (28). A recurrent feature of these proteins is the use of an Arg anchor residue to make contacts with carboxylates of the H2A–H2B acidic patch, although the rest of the interactions are not conserved (26). The nucleosome-binding pose of PFV GAG CBS partially overlaps with those of acidic patch binders, and Arg540, a residue invariant among spumaviruses, plays the role of the Arg anchor (Fig. 2 *C* and *D*). In agreement with the structures, peptides spanning the chromatin-binding motifs of KSHV LANA and CENP-C efficiently competed with PFV GAG CBS for nucleosome binding (Fig. 1*D*, lanes 1–5). In contrast, the mutant forms of the peptides, unable to interact with nucleosomes because of substitutions of the respective Arg anchor residues, in large part lost the ability to compete with PFV GAG (Fig. 1*D*, lanes 6–10). Furthermore, substitutions of PFV GAG Tyr537, Leu539, and Arg540 in the context of the CBS peptide or full-length protein abolished the interaction with nucleosomes (Fig. 1 *B* and *E*).

To evaluate the importance of the CBS–nucleosome interaction for GAG chromatin tethering, we produced HT1080 cells stably expressing FLAG-tagged PFV and macaque simian foamy virus (SFV_{mac}) GAG proteins. In agreement with previous observations (23, 24), the WT GAG proteins were nuclear and colocalized with DNA throughout interphase and during mitosis (Fig. 3*A*). The Arg anchor mutations R540Q (PFV) and R541Q (SFV_{mac}) abolished nuclear accumulation and chromosome binding of both proteins. Phenocopying the CBS-deletion mutant of PFV GAG (23, 24), both point mutants accumulated in the cytoplasm and partly colocalized with pericentrin, a marker for the microtubule-organizing center, during interphase (Fig. 3*A* and *SI Appendix*, Fig. S2). In addition, SFV_{mac} GAG R541Q appeared to decorate the entire spindle apparatus during mitosis.

Loss of CBS Function Leads to Catastrophic Redistribution of Spumavirus Integration Sites to Centromeric Regions. To assess the importance of GAG tethering to chromatin during infection, we produced single-cycle virus particles using codon-optimized constructs encoding PFV GAG (WT or R540Q), POL (WT or catalytically inert integrase mutant D185N/E221Q), and ENV along with a GFP reporter transfer vector (29). The vector particles were purified by ultracentrifugation through 20% sucrose cushions and quantified by immunoblotting using polyclonal anti-GAG antisera. Analysis of the viral lysates with anti-integrase antibodies confirmed that the mutations did not perturb the relative packaging of the structural and enzyme components (Fig. 4). Human fibrosarcoma HT1080 cells and MRC5 fibroblasts were infected with equal amounts of the virus preparations and expanded to dilute nonintegrated viral DNA. Five days postinfection, the cells were subjected to flow cytometry to quantify the percentage of GFP^+ cells, and integrated proviral DNA content was measured by real-time PCR using primers specific to the PFV LTR region. As expected, the D185N/E221Q active site mutation in integrase abolished both infectivity and integration. The substitution of the GAG Arg anchor did not affect the levels of integrated proviral DNA but led to a twofold defect in the ability to render target cells GFP^+ (Fig. 4), suggesting a defect in viral gene expression. When HT1080 cells were infected with concentrated vector, a specular pattern of viral GAG antigen was readily detected at the cell periphery when the cells were fixed immediately following spinoculation (0 h postinfection) (Fig. 3*B*). In agreement with published observations (23, 30), 4 h postinfection GAG colocalized with centrosomes in every interphase cell observed and with condensed chromosomes in mitotic cells (Fig. 3*B* and *SI Appendix*,

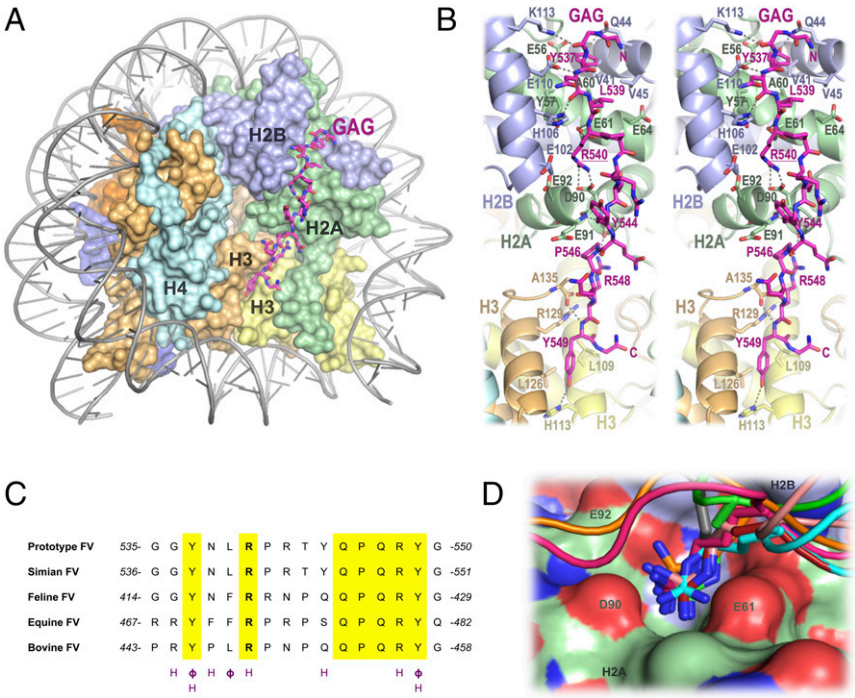

Fig. 2. Crystal structure of the PFV GAG-CBS–nucleosome complex. (*A*) Overview of the structure. Core histones are shown as a surface representation, with H2A in green, H2B in pale blue, H3 in orange and yellow, and H4 in cyan. DNA is shown as a cartoon colored in gray, and the GAG CBS peptide is shown as sticks with carbon atoms in magenta. (*B*) Wall-eye stereoview of the CBS–histone interface. The GAG CBS peptide is shown as a stick, and histones are shown as cartoons with select residues as sticks; hydrogen bonds are shown as dashed lines. (*C*) Amino acid sequence alignment of GAG CBSs from PFV, SFV_{mac}, and feline, equine, and bovine foamy viruses. Invariant residues are highlighted in yellow, and the Arg anchor is shown in bold; residues involved in hydrogen bonds and hydrophobic interactions with the core histones are indicated with the magenta symbols H and ϕ, respectively. (*D*) Superposition in the nucleosome acid patch of the conserved arginine anchor motif from KHSV LANA (PDB ID code 1ZLA, cyan) (27), regulator of chromosome condensation 1 (RCC1, PDB ID code 3MVD, green) (50), CENP-C (PDB ID code 4X23, orange) (28), Sir3 (PDB ID code 3TU4, salmon) (51), protein regulator of cytokinesis 1 (PRC1, PDB ID code 4R8P, gray) (52), human cytomegalovirus immediate-early protein 1 (IE1) (PDB ID code 5E5A, red) (53), and PFV GAG (pink). Histones are shown in surface representation with oxygen and nitrogen atoms in red and blue, respectively, and carbon atoms in green (H2A) or light blue (H2B).

Fig. S3). Strikingly, although GAG R540Q PFV retained the ability to concentrate near or on centrosomes, it failed to relocalize to chromosomes (Fig. 3*B* and *SI Appendix*, Fig. S3), phenocopying ectopically expressed R540Q PFV GAG-FLAG (Fig. 3*A*).

To test if the GAG–chromatin interactions play a role in spumaviral integration site selection, we mapped integration sites of WT and R540Q PFV in a range of cell lines and compared them with a reference set of integration sites obtained using recombinant PFV intasomes and deproteinized human DNA in vitro. In agreement with previous observations (22, 31) and in sharp contrast to HIV-1 and MLV, PFV disfavored integration into genes (Fig. 5*A* and *SI Appendix*, Table S2). Thus, only 31.7% of vector integration events mapped to transcription units in HT1080 cells, 15% lower than the levels observed for the in vitro reference sample (46.4%, $P < 10^{-320}$); results of comprehensive statistical tests are given in Dataset S1. Furthermore, the virus showed marked preference for integration into constitutive lamina-associated regions (cLADs) and dark Giemsa-positive cytobands ($P < 10^{-320}$). Surprisingly, the propensity of PFV to integrate into these deep heterochromatic regions varied widely, depending on the nature of target cell line. Thus, although the PFV vector showed twofold preference for integration into cLADs in HT1080 cells, it slightly disfavored integration into these regions in HepG2 cells ($P < 10^{-128}$). Furthermore, although PFV showed only a minor preference for integration near transcription start sites (TSSs) ($P < 10^{-5}$) and insignificant preference for CpG islands in HT1080 cells ($P > 0.2$), it integrated into these regions three times more frequently than expected in HepG2 cells ($P < 10^{-320}$). Despite these differences, under all conditions, PFV integrated into cLADs significantly more frequently than either HIV-1 or MLV and integrated near TSSs and CpG islands much less frequently than MLV ($P < 10^{-82}$) (Fig. 5*A*) (16–18, 22).

The R540Q mutation profoundly affected the distribution of integration sites across the panel of target cell lines (Fig. 5*A*). The mutant virus integrated significantly less frequently in the vicinities of TSSs ($P < 10^{-37}$) and CpG islands ($P < 10^{-24}$) and preferred regions with lower local gene densities ($P < 10^{-15}$), indicating a general retreat from loci associated with gene expression. However, although the mutation significantly stimulated integration events into cLADs ($P < 10^{-320}$) and Giemsa-positive cytobands ($P < 10^{-85}$) in HepG2 cells, it dampened them in HT1080 cells ($P < 10^{-175}$) (Fig. 5*A*). This unexpected discordance prompted us to broaden the focus of the integration site analysis to a more global scale. Inspection of integration site densities at the chromosomal level revealed that the R540Q mutation led to a massive redistribution of integration toward centromeres in all studied cell lines (Fig. 5*A*). Using α-satellite–specific quantitative PCR as an independent method, we estimated that the R540Q mutant integrated 7.8 ± 1.1 times more frequently in the vicinity of α-satellites than the WT control in HT1080 cells. Finally, to confirm that the observed phenotypes are independent of the vector construct or the original viral isolate, we studied integration site distributions of non–codon-optimized vectors based on PFV and SFV_{mac} (32, 33). Arg anchor substitution variants of either vector displayed twofold defects in the ability to transduce HT1080 cells while integrating with nearly WT efficiency (*SI Appendix*, Fig. S4). Furthermore, the R540Q/R541Q mutations greatly enhanced their propensity to integrate into centromeres (Fig. 5 *A* and *B*).

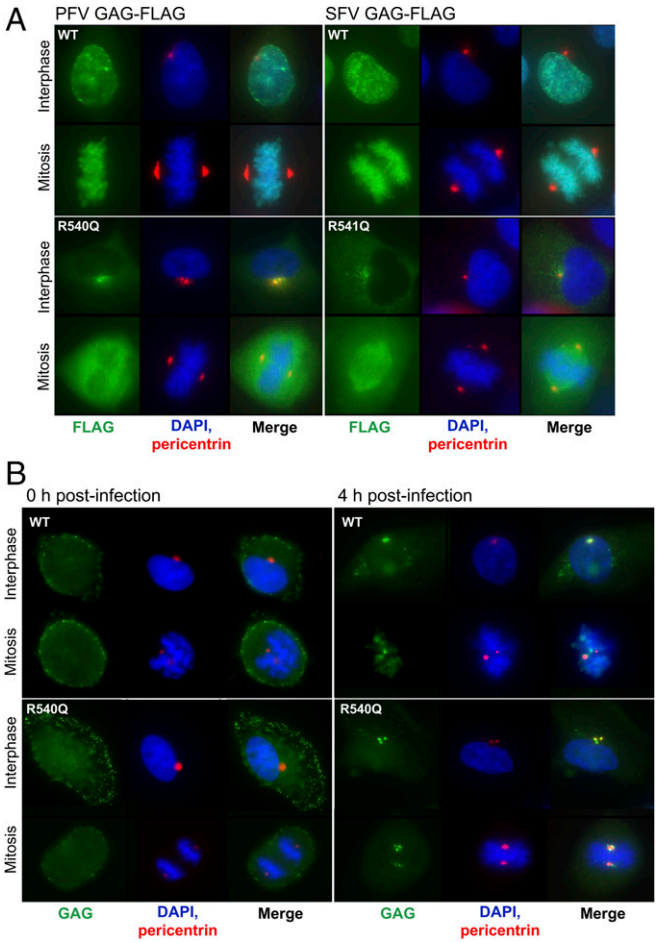

Fig. 3. The Arg anchor residue is essential for chromosomal tethering of PFV and SFV_{mac} GAG. (*A*) HT1080 cells stably expressing FLAG-tagged PFV (WT or R540Q) (*Left*) or SFV_{mac} GAG (WT or R541Q) (*Right*) were stained with anti-FLAG (green) or anti-pericentrin antibody (red); DNA was visualized with DAPI (blue). (*B*) GAG localization during PFV infection. HT1080 cells infected with PFV vector by spinoculation at 4 °C were fixed immediately or after 4-h incubation at 37 °C. GAG was detected using polyclonal anti-PFV Gag antiserum (green); pericentrin (red) and chromosomal DNA (blue) were visualized as in *A*.

Discussion

Here we show that spumaviral GAG interacts directly with chromatin, engaging the H2A–H2B acidic patch via the Arg anchor motif located within its CBS. The GAG residues involved in this interface are critical for binding nucleosomes in vitro, and the anchor residue Arg540 is essential for stable GAG chromatin tethering in cells. The structure of the PFV CBS–nucleosome complex underscores similarities as well as striking differences among H2A–H2B acidic patch binders. Coalescing at the essential Arg anchor residue (Fig. 2*D* and *SI Appendix*, Fig. S5), these proteins compete for exclusive binding to a nucleosome face (Fig. 1*D*). At the same time, the histone surface presents ample freedom for the chromatin interactors to adopt a wide range of binding configurations (*SI Appendix*, Fig. S5). The PFV GAG CBS peptide tracks across the entire protein face of the nucleosome, making contacts with highly conserved residues within core regions of histones H2A, H2B, and H3. We conclude that the CBS is a pan-nucleosome binder, and, although it is essential to maintain the observed primate spumavirus integration site distribution profiles (Fig. 5*A* and *B*), it is unlikely to determine them. Similarly, the structure of the PFV intasome bound to the nucleosome core particle highlighted conserved nucleosome features that are critical for integration into chromatin but do not explain the observed integration site profiles (22). The propensity of R540Q GAG to persist near centrosomes (Fig. 3 *A* and *B*) provides a clue as to why the mutant viral integration sites cluster around centromeres: As chromosomes approach centrosomes at the end of mitosis, the centromeric regions could come in closer contact with the viral nucleoprotein complexes. Binding of GAG to centrosomes is integral to intracellular trafficking and biogenesis of the spumavirus (34–36), and it would be of interest to identify the centrosomal receptor for spumaviral GAG.

In contrast to spumaviruses, orthoretroviral GAG proteins are processed during viral maturation into separate polypeptides, including capsid, as well as into a range of smaller accessory proteins. P12 is an essential product of the gammaretroviral *gag* gene and is a component of the MLV preintegration complex (37, 38). In striking similarity to PFV GAG, MLV P12 associates with mitotic chromosomes, and this property depends on an Arg-rich motif within its C-terminal region (39). Although deletions of or substitutions within this putative nucleosome-binding region abolish MLV infectivity, its replacement with heterologous chromatin-binding peptides, such as the PFV GAG CBS, are tolerated (39, 40). Thus the loss of chromatin-tethering function via a *gag* product seems to be even more catastrophic for MLV than for the spumaviruses, which retain the ability to complete integration upon CBS abrogation (Fig. 4 and *SI Appendix*, Fig. S4) (23). In this regard, the phenotype of the spumaviral GAG CBS mutants is more similar to those of HIV-1, which display dramatic redistributions of integration sites away from gene-rich chromosomal domains without severe loss of virus titer when capsid binding to the mRNA maturating factor CPSF6 is disrupted (10, 11,

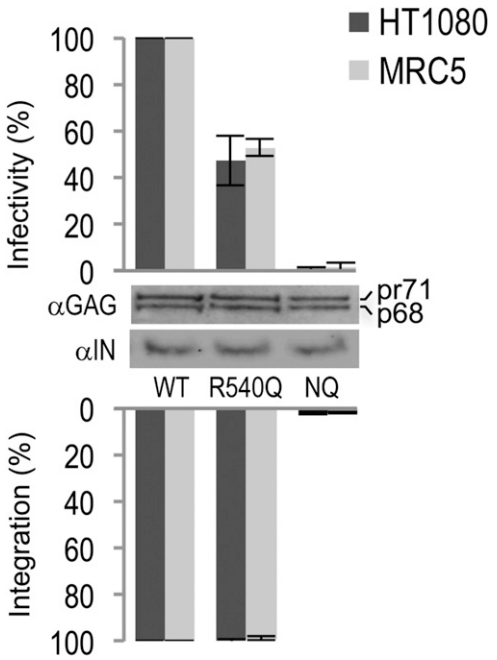

Fig. 4. Infectivity analyses of the PFV GAG R540Q mutant. Equivalent amounts of PFV vector particles carrying WT or R540Q GAG or catalytically inert D185N/E221Q integrase (NQ) were used to infect HT1080 or MRC5 cells. Five days postinfection, the total proviral content was measured by quantitative real-time PCR, and GFP^+ cells were counted by flow cytometry as relative measures of successful integration and infectivity, respectively. The bar charts represent mean relative values from a series of infections with varied viral inputs normalized by GAG protein content and resulting in ~5–50% GFP conversion in the WT sample. Error bars are SDs determined from at least three independent infections; the WT values in each experiment were set to 100%. Immunoblots detecting GAG isoforms (pr71 and p68) and integrase are shown in the middle.

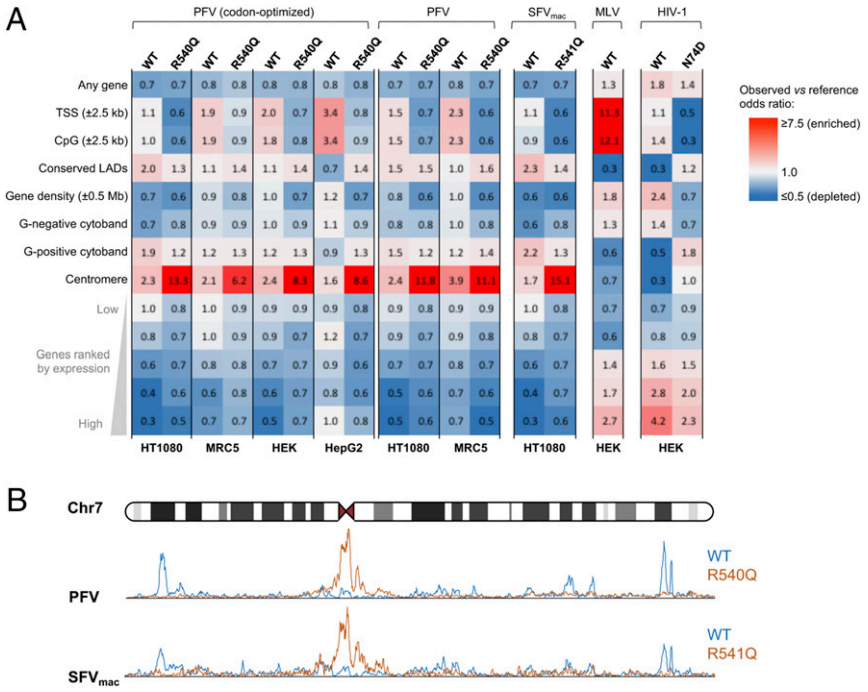

Fig. 5. Integration site distribution of arginine anchor motif-mutant virus particles. (*A*) Integration frequency of the indicated viruses near annotated genomic features expressed as odds ratios and shown in a heatmap. Enriched features (odds ratio >1) compared with reference are highlighted in red, and depleted features (odds ratio <1) are shown in blue, as indicated by the legend to the right. Raw data are given in *SI Appendix*, Table S2. HIV-1 and MLV integration sites in HEK293T cells alongside the matched random control simulated data were from published work (11). HT1080, HEK293T, MRC5, or HepG2 cell line-specific gene-expression activity was reported previously (GEO accession codes GSE58968, GSE11892, GSE63577, and GSE87958) (54–57). All observable differences between WT and the corresponding mutant data were statistically significant ($P < 0.05$) (Dataset S1). (*B*) PFV (WT or R540Q) and SFV_{mac} (WT or R541Q) integration site density along human chromosome 7 in a sliding window of 500 Mbp. Integration site densities of WT and mutant vectors are shown as blue and orange traces respectively.

41, 42). Our findings highlight that the involvement of GAG in integration site selection is conserved in both retroviral subfamilies, despite the great evolutionary distance separating them.

Primate spumaviruses have broad host-tissue tropism and are capable of replicating in a wide range of cell types in vitro (43, 44), although the most active viral replication appears to occur in superficial epithelium (45). Comparison of PFV integration site profiles in our panel of human cell lines revealed unexpected variability, with HT1080 fibrosarcoma and HepG2 hepatocytes on opposing sides of the spectrum. Thus, although PFV integration is strongly biased toward deep heterochromatic regions, such as cLADs, in HT1080, it largely avoids them in HepG2. The converse is true for TSSs and CpG islands, which become strongly preferred in HEK293T, MRC5 fibroblasts, and even more so in HepG2 cells. Moreover, in the latter cell line, the virus appears to lose its ability to discriminate against highly active transcription units, a trait that is highly prominent in HT1080 cells (Fig. 5*A*) (22). Nevertheless, in all cell lines studied here (Fig. 5*A* and *SI Appendix*, Table S2), as well as in primary fibroblasts and hematopoietic cells (31), PFV disfavors integration into transcription units and integrates into regions of considerably lower gene density than the lentivirus HIV-1 or the gammaretrovirus MLV (Fig. 5*A*). Spumaviral infection is not associated with severe pathology, and perhaps this retroviral subfamily evolved to self-moderate by integrating into less active regions of host cell genomes. It is tempting to speculate that, akin to lentiviruses and gammaretroviruses, spumaviruses use chromatin-associated host factor(s), possibly recognized by integrase, to facilitate integration site selection. Variable expression of a targeting factor could explain the observed differences in integration profiles among the studied cell lines.

We have not been able to detect interaction between the PFV intasome and GAG, and their respective interfaces with the nucleosome do not overlap (Fig. 2*A* and ref. 22). It is possible that GAG could associate with the preintegration complex as a consequence of its ability to bind nucleic acids (24). Alternatively, the GAG–chromatin interface may occur before viral capsid uncoating and preintegration complex assembly. Our data show that GAG–chromosomal tethering is critical to the virus's ability to avoid genomic junkyards, such as centromeric regions. On the face of the massive redistribution of integration sites toward centromeres, it is not surprising that the R540Q vectors displayed a significant reduction in transgene expression (Fig. 4 and *SI Appendix*, Fig. S4).

Materials and Methods

Recombinant PFV and SFV_{mac} GAG proteins and human histones were produced in bacteria and purified as detailed in *SI Appendix, SI Materials and Methods*. Mononucleosomes were assembled using a DNA construct with the Widom 601 positioning sequence (25) and crystallized according to established procedures (46). The vector and in vitro integration sites were sequenced using linker-mediated PCR in conjunction with the Illumina technology as described (4, 47, 48). Additional experimental details are given in *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Ophelie Cosnefroy for helpful discussions; Bas van Steensel for sharing definitions of human lamina-associated regions (49); the ENCODE consortium for providing HepG2 gene-expression data (GEO accession code GSE87958); the staff at I03, Diamond Light Source for assistance with X-ray data collection; and Phil Walker and Andrew Purkiss for X-ray crystallography software support. This work was supported by NIH Grants GM082251 (to A.N.E. and P.C.), AI039394 (to A.N.E.), and AI060354 (to the Harvard University Center for AIDS Research); Deutsche Forschungsgemeinschaft LI 621/10-1 and LI 621/11-1 SPP1923 Grants (to D.L.); and the Francis Crick Institute (P.C.), which receives its core funding from Cancer Research UK (FC001061), the UK Medical Research Council (FC001061), and the Wellcome Trust (FC001061).

1. Lesbats P, Engelman AN, Cherepanov P (2016) Retroviral DNA Integration. *Chem Rev* 116:12730–12757.
2. Van Lint C, Bouchat S, Marcello A (2013) HIV-1 transcription and latency: An update. *Retrovirology* 10:67.
3. Gérard A, et al. (2015) The integrase cofactor LEDGF/p75 associates with Iws1 and Spt6 for postintegration silencing of HIV-1 gene expression in latently infected cells. *Cell Host Microbe* 17:107–117.
4. Schröder AR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521–529.
5. Singh PK, et al. (2015) LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev* 29:2287–2297.
6. Cherepanov P, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 278:372–381.
7. Shun MC, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 21:1767–1778.
8. Ciuffi A, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11:1287–1289.
9. Lee K, et al. (2010) Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* 7:221–233.
10. Schaller T, et al. (2011) HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog* 7:e1002439.
11. Sowd GA, et al. (2016) A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc Natl Acad Sci USA* 113:E1054–E1063.
12. Llano M, et al. (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.
13. Cherepanov P (2007) LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. *Nucleic Acids Res* 35:113–124.
14. Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A (2014) Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* 42:10209–10225.
15. Mitchell RS, et al. (2004) Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2:E234.
16. De Ravin SS, et al. (2014) Enhancers are major targets for murine leukemia virus vector integration. *J Virol* 88:4504–4513.
17. LaFave MC, et al. (2014) MLV integration site selection is driven by strong enhancers and active promoters. *Nucleic Acids Res* 42:4257–4269.
18. Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300:1749–1751.
19. Sharma A, et al. (2013) BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc Natl Acad Sci USA* 110:12036–12041.
20. Gupta SS, et al. (2013) Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J Virol* 87:12721–12736.
21. De Rijck J, et al. (2013) The BET family of proteins targets Moloney murine leukemia virus integration near transcription start sites. *Cell Reports* 5:886–894.
22. Maskell DP, et al. (2015) Structural basis for retroviral integration into nucleosomes. *Nature* 523:366–369.
23. Müllers E, Stirnnagel K, Kaulfuss S, Lindemann D (2011) Prototype foamy virus gag nuclear localization: A novel pathway among retroviruses. *J Virol* 85:9276–9285.
24. Tobaly-Tapiero J, et al. (2008) Chromatin tethering of incoming foamy virus by the structural Gag protein. *Traffic* 9:1717–1727.
25. Lowary PT, Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 276: 19–42.
26. McGinty RK, Tan S (2016) Recognition of the nucleosome by chromatin factors and enzymes. *Curr Opin Struct Biol* 37:54–61.
27. Barbera AJ, et al. (2006) The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* 311:856–861.
28. Kato H, et al. (2013) A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* 340:1110–1113.
29. Hamann MV, et al. (2014) The cooperative function of arginine residues in the prototype foamy virus Gag C-terminus mediates viral and cellular RNA encapsidation. *Retrovirology* 11:87.
30. Stirnnagel K, et al. (2012) Differential pH-dependent cellular uptake pathways among foamy viruses elucidated using dual-colored fluorescent particles. *Retrovirology* 9:71.
31. Trobridge GD, et al. (2006) Foamy virus vector integration sites in normal human cells. *Proc Natl Acad Sci USA* 103:1498–1503.
32. Duda A, et al. (2004) Prototype foamy virus envelope glycoprotein leader peptide processing is mediated by a furin-like cellular protease, but cleavage is not essential for viral infectivity. *J Virol* 78:13865–13870.
33. Yap MW, et al. (2008) Restriction of foamy viruses by primate Trim5alpha. *J Virol* 82: 5429–5439.
34. Lehmann-Che J, et al. (2007) Centrosomal latency of incoming foamy viruses in resting cells. *PLoS Pathog* 3:e74.
35. Petit C, et al. (2003) Targeting of incoming retroviral Gag to the centrosome involves a direct interaction with the dynein light chain 8. *J Cell Sci* 116:3433–3442.
36. Yu SF, Eastman SW, Linial ML (2006) Foamy virus capsid assembly occurs at a pericentriolar region through a cytoplasmic targeting/retention signal in Gag. *Traffic* 7: 966–977.
37. Yuan B, Li X, Goff SP (1999) Mutations altering the Moloney murine leukemia virus p12 Gag protein affect virion production and early events of the virus life cycle. *EMBO J* 18:4700–4710.
38. Prizan-Ravid A, et al. (2010) The Gag cleavage product, p12, is a functional constituent of the murine leukemia virus pre-integration complex. *PLoS Pathog* 6:e1001183.
39. Schneider WM, et al. (2013) Viral DNA tethering domains complement replication-defective mutations in the p12 protein of MuLV Gag. *Proc Natl Acad Sci USA* 110: 9487–9492.
40. Wight DJ, et al. (2012) The gammaretroviral p12 protein has multiple domains that function during the early stages of replication. *Retrovirology* 9:83.
41. Koh Y, et al. (2013) Differential effects of human immunodeficiency virus type 1 capsid and cellular factors nucleoporin 153 and LEDGF/p75 on the efficiency and specificity of viral DNA integration. *J Virol* 87:648–658.
42. Saito A, et al. (2016) Capsid-CPSF6 interaction is dispensable for HIV-1 replication in primary cells but is selected during virus passage in vivo. *J Virol* 90:6918–6935.
43. Mergia A, Leung NJ, Blackwell J (1996) Cell tropism of the simian foamy virus type 1 (SFV-1). *J Med Primatol* 25:2–7.
44. Soliven K, et al. (2013) Simian foamy virus infection of rhesus macaques in Bangladesh: Relationship of latent proviruses and transcriptionally active viruses. *J Virol* 87: 13628–13639.
45. Murray SM, et al. (2008) Replication in a superficial epithelial cell niche explains the lack of pathogenicity of primate foamy virus infections. *J Virol* 82:5981–5985.
46. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* 319:1097–1113.
47. Chatterjee AG, et al. (2014) Serial number tagging reveals a prominent sequence preference of retrotransposon integration. *Nucleic Acids Res* 42:8449–8460.
48. Serrao E, Cherepanov P, Engelman AN (2016) Amplification, next-generation sequencing, and genomic DNA mapping of retroviral integration sites. *J Vis Exp* (109): e53840.
49. Kind J, et al. (2015) Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 163:134–147.
50. Makde RD, England JR, Yennawar HP, Tan S (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 467:562–566.
51. Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE (2011) Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* 334:977–982.
52. McGinty RK, Henrici RC, Tan S (2014) Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* 514:591–596.
53. Fang Q, et al. (2016) Human cytomegalovirus IE1 protein alters the higher-order chromatin structure by targeting the acidic patch of the nucleosome. *eLife* 5:5.
54. Deyle DR, et al. (2014) A genome-wide map of adeno-associated virus-mediated human gene targeting. *Nat Struct Mol Biol* 21:969–975.
55. Richard H, et al. (2010) Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res* 38:e112.
56. Marthandan S, et al. (2015) Similarities in gene expression profiles during in vitro aging of primary human embryonic lung and foreskin fibroblasts. *BioMed Res Int* 2015:731938.
57. Consortium EP; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.