7366–7378 *Nucleic Acids Research, 2018, Vol. 46, No. 14* *Published online 26 May 2018*
doi: 10.1093/nar/gky369

Identification of conserved, primary sequence motifs that direct retrovirus RNA fate

Gatikrushna Singh[1], Brittany D. Rife[2], Bradley Seufzer[1], Marco Salemi[2], Aaron Rendahl[1] and Kathleen Boris-Lawrie[1,*]

[1]Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA and [2]Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, FL 32610, USA

Received April 04, 2017; Revised April 20, 2018; Editorial Decision April 23, 2018; Accepted May 16, 2018

ABSTRACT

Precise stoichiometry of genome-length transcripts and alternatively spliced mRNAs is a hallmark of retroviruses. We discovered short, guanosine and adenosine sequence motifs in the 5′untranslated region of several retroviruses and ascertained the reasons for their conservation using a representative lentivirus and genetically simpler retrovirus. We conducted site-directed mutagenesis of the GA-motifs in HIV molecular clones and observed steep replication delays in T-cells. Quantitative RNA analyses demonstrate the GA-motifs are necessary to retain unspliced viral transcripts from alternative splicing. Mutagenesis of the GA-motifs in a C-type retrovirus validate the similar downregulation of unspliced transcripts and virion structural protein. The evidence from cell-based co-precipitation studies shows the GA-motifs in the 5′untranslated region confer binding by SFPQ/PSF, a protein co-regulated with T-cell activation. Diminished SFPQ/PSF or mutation of either GA-motif attenuates the replication of HIV. The interaction of SFPQ/PSF with both GA-motifs is crucial for maintaining the stoichiometry of the viral transcripts and does not affect packaging of HIV RNA. Our results demonstrate the conserved GA-motifs direct the fate of retrovirus RNA. These findings have exposed an RNA-based molecular target to attenuate retrovirus replication.

INTRODUCTION

An RNA polymer of 9000 bases contains within it the genetic instructions to produce retroviruses that cause a range of consequences from cancer and AIDS to cellular transduction without harm. Executing these instructions requires expression of alternatively spliced transcripts and stoichiometric amounts of the unspliced 9 kb RNA. This means different populations of the 9 kb RNA experience, or subvert, the processes of intron removal and nuclear retention (1,2). While activation of unspliced RNA is a fundamental property of *Retroviridae*, whether conserved viral sequences and cognate RNA-binding protein are responsible remains an open issue.

During transcription, nascent transcripts become decorated with splicing commitment factors that direct the assembly of spliceosomes and heteronuclear (hn) ribonucleoproteins (RNPs), some of which contain nuclear retention signals and are released as a result of splicing or during nuclear export (3). Otherwise, pre-mRNAs are retained in the nucleus (3). Retroviruses exhibit a fragile balance between unspliced transcripts and their spliced isoforms (4). Even slight perturbations in this balance can attenuate viral replication by several orders of magnitude (56). The introduction of mutations that favor spliced isoforms can severely attenuate retroviral replication (7). Furthermore, positive selection maintains nonconsensus sequences of 5′ and 3′ splice sites (8). Experiments by Katz and Skalka established that activating-mutations in the splice sites of avian sarcoma-leukosis virus (ASLV) severely attenuate replication (6,7). Replicating virus reemerged after long-term passage of infected cell cultures, suggesting that second-site mutations can restore the natural ratio of spliced to unspliced RNA isoforms (6,7).

Clustered in the 5′untranslated region (UTR) of retrovirus unspliced RNA, overlapping cis-acting RNA elements govern early and late events in viral replication (1). Beginning in the 1980s, scientists discovered the ability to genetically engineer retroviral 5′UTR elements and create retroviral vectors (9–11). By deleting significant portions of the leader RNA, virion structural proteins could be expressed in trans (12,13). In contrast, fusing these leader sequences to transgenes of interest generated a retroviral vector RNA susceptible to replication by helper virus proteins (14). Henceforth, empirical studies defined the sequences necessary for the dimerization and packaging of retroviral

*To whom correspondence should be addressed. Tel: +1 612 624 2700; Fax: +1 612 624 9713; Email: kbl@umn.edu

vector RNA, which typically overlap the 5′ splice site (5′ss) (15–17).

These pioneering studies occurred before scientists discovered that retroviruses require a precise balance of unspliced and alternatively spliced transcripts to generate infectious virions (4). This discovery began with the observation that avian reticuloendotheliosis virus strains vary in transforming titer due to the amount of pre-mRNA that was processed to sub-genomic mRNA (5,18). Then, 5'intron sequences were identified that suppressed the amount of sub-genomic transcript and significantly bolstered infectious titer (5–7,15). Only decades later, molecular studies demonstrated that the activity of the spliceosome dispatches nuclear retention signals in pre-mRNA that otherwise guard against nuclear export and translation of immature mRNA (3). To fully understand retroviral replication and the molecular mechanisms these viruses use to subvert cellular post-transcriptional regulation, we must first precisely define the conserved viral sequences and cognate trans-acting factors necessary to retain pre-mRNAs in balance with spliced isoforms.

During the course of generating hybrid retroviral vectors, we noticed remarkably similar sequence patterns flanking the 5′ss of both simple and more complex retroviruses (19–21). We postulated that these sequence patterns have a vital role in viral gene regulation because their position is exclusive to the unspliced viral transcript. Here, we demonstrate that similar primary sequence motifs on both sides of the 5′ss are conserved in simple and more complex retroviruses and evaluate their activity in two representative retroviruses. We show that GA-motifs in the 5′UTR of HIV and spleen necrosis virus (SNV) are necessary to regulate retention of unspliced viral transcript and for interaction with SFPQ/PSF, a splicing regulatory protein that is conserved amongst the host species for these retroviruses.

MATERIALS AND METHODS

Evolutionary analysis of HIV group M major subtypes

HIV group M 5′UTR RNA sequences (+1–338) were obtained from the Los Alamos HIV Database (www.lanl.gov) and filtered based on the following inclusion criteria: (i) subtypes with must have greater than 10 published sequences; (ii) only one sequence was used per patient; (iii) sequences for which both the splice site and gag start codon were present (no insertions/deletions); (iv) no hypermutated sequences. These criteria resulted in 393 included sequences belonging to subtypes A1, B, C, D, circulating recombinant form (CRF)01-AE and CRF02-AG (alignment available from the authors upon request). HIV^{NL4-3} reference sequence was included in the analysis of subtype B sequences. Please see the Supplementary Methods for details of tree reconstruction and analysis.

Plasmid subcloning and mutagenesis

Derivatives of the spleen necrosis virus (SNV) proviral clone pPB101 were generated by PCR-based site-directed mutagenesis on sub-genomic template DNAs, followed by ligation into reporter or provirus plasmids. Using *Ase*I and *Sma*I restriction sites in pEGFP-N1 (Clonetech), SNV provirus sequence 1 to 2475 (5′LTR and *gag*) was ligated in-frame with enhanced green fluorescent protein (eGFP). SNV *pol* and *env* from pPB101 (nt 4788–6900) containing the env splice acceptor were ligated at the unique *Hpa*I site between *gfp* and the polyadenylation signal. 5′UTR mutations were introduced on *Ase*I/*Sal*I digested PCR products of SNV U3 (nt 1–398) from pYW205 (22). To create pSNV Δ300, pSNV WT was digested with KpnI and EclXI and ligated with a *Kpn*I–*Cla*I–*Ecl*XI linker. RU5 generated by PCR was ligated into the *Kpn*I site to generate pSNV Δ300+Rescue. All site-directed mutagenesis used the QuikChange protocol (Agilent Technologies, La Jolla, CA, USA) with mismatched oligonucleotides that are provided in the Supplementary Methods. All constructions were confirmed by restriction mapping and DNA sequence analysis. To harvest virions, culture supernatant fractions were collected through 0.2 μm filters and placed in 13 ml ultracentrifugation tubes containing 2 ml 20% sucrose and centrifuged at 35 000 rpm for 2 h in SW41 rotor (Beckman Coulter). The supernates were removed from the tubes, and then virion-associated RNA was prepared by TRIzol extraction, followed by RT-qPCR. HIV cDNA-specific primers were used as described by Wong *et al.* (23). All oligonucleotide sequences are provided in the Supplementary Methods.

Virus propagation, infection, and virion isolation

The plasmids were transfected to HEK293 cells for 4–6 h, medium was exchanged, and cells were cultured for 24–36 h. Jurkat cells or MT-4 lymphocytes (5×10^5 cells in 0.5 ml RPMI) freshly seeded in 12-well plates and were incubated with cell-free medium containing equivalent numbers of virions determined using a Gag enzyme-linked immunosorbent assay (ELISA). Culture plates were centrifuged at 2300 rpm for 2 h at room temperature. The cells were washed twice with RPMI medium and maintained at 37°C. Culture supernatants fractions were collected every 48 h and reverse transcriptase (RT) activity was determined using PicoGreen reagent according to manufacturer's instructions (Life Technologies) and using standard curve generated on recombinant RT enzyme (Worthington Biochemical Corporation).

Transfections, protein analyses, recombinant protein purification and RNA electrophoretic mobility shift assays

Please see the Supplementary Methods.

RESULTS

Invariant purine residues on both sides of the 5′ splice site are necessary for viral replication

When we compared the 5′UTR sequences from seven reference strains of *Retroviridae* available in GenBank, we noticed two clusters of GA-rich nucleotides (Supplementary Figure S1) were apparent in the context of otherwise different primary sequences. Phylogenetic comparisons of HIV^{NL4-3}, SNV, simian immunodeficiency virus (SIV), human T-cell leukemia virus type 1 (HTLV-1), bovine

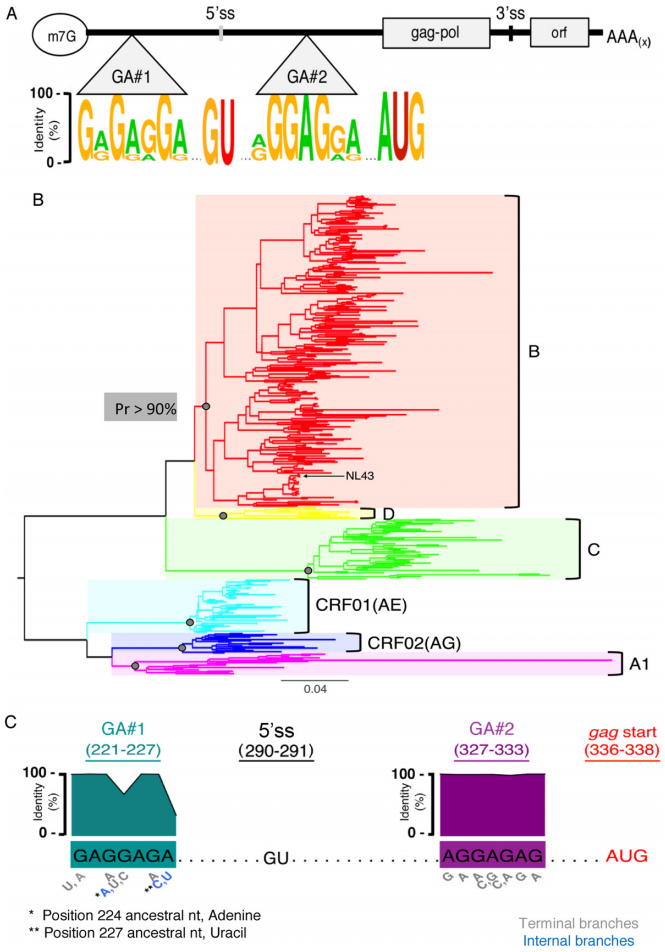

Figure 1. Conserved guanosine-adenosine (GA)-motifs flank the 5′ splice site (5'ss) of unique 5′ (U5) sequences of major HIV group M subtypes and other retroviruses. (**A**) Generic retrovirus RNA structure diagram that shows the 5′ 7methyl-guanosine cap (m7G), gag-pol open reading frame (ORF) and relative position of 3′ splice site (3'ss) that flanks downstream ORFs. Phylogenetic analysis of these GA-motifs in HIV, SNV, HTLV-1, BLV, MMLV, and RSV reference sequences revealed a conserved guanosine (G) or adenosine (A) or G and A at each position, as shown by the colorful sequence logo. Percent identity is depicted by relative height of the G or A wherein the site-specific information content (bits) is represented by the total height of all nucleotides occupying a single position within the sequence logo (44). (**B**) Bayesian phylogenetic tree reconstruction for 393 U5 sequences obtained from the Los Alamos Database. Branches are scaled in nucleotide substitutions per site (scale along the bottom) and colored/labeled according to subtype. *Gray circles indicate posterior probability >90% for subtype-specific clades. (**C**) Genomic map of the GA-motifs described in this study and corresponding site-specific substitution distributions across branches of the HIV 5'UTR evolutionary tree. Sequence conservation (percent identity) and specific nucleotide substitutions, as well as their location along internal (light blue) or terminal (gray) branches, are indicated. Asterisks indicate deviation of ancestral nucleotide (common ancestor for all analyzed sequences) from HIV^{NL4-3} sequence. The sequence logo demonstrates significant conservation within the phylogeny, and the relative height of the boxed line represents nucleotide frequency (44).

leukemia virus (BLV), Moloney murine leukemia virus (MMLV), and Rous sarcoma virus (RSV) predicted the several positions in these primary sequences have identical nucleotide composition, and the sequence 'GGAG' was present in every 5′UTR (Figure 1A). Notably, the two GA-motifs vary in orientation and spacing relative to the major 5′ss (Supplementary Figure S1).

We performed phylogenetic analysis of 393 unique sequences of HIV group M patient isolates downloaded from the Los Alamos database. We uncovered two clusters of nearly invariant guanosine and adenosines that flank the major 5′ss (average of >95% identity). The resolved quartet tree reconstruction documented insignificant levels of substitution saturation, and a positive correlation between sequence divergence and sample collection time (Figure 1B). Albeit rare, variations occurred primarily between adenosine and guanosine (G-to-A or A-to-G substitutions) and almost exclusively along external branches. Such mutations are typically characterized as transient polymorphisms and are rapidly cleared from the population due to purifying selection.

To evaluate activity associated with the similar GA-rich sequences conserved between retroviruses, we used the results from the phylogenetic analyses above as a guide to direct site-directed mutagenesis of HIV molecular clones and clones of another representative retrovirus. As summarized in Figure 1C, HIV GA#1 was conserved in five of seven residues of the motif. The fourth (G224) and seventh (A227) positions allowed some substitution flexibility, as cytosine and uridine substitutions were inferred along internal branches. Therefore, we excluded these sites as options for site-directed mutagenesis. Alternatively, the second (A222) and fifth (A225) positions were completely conserved, with substitutions absent from both external and internal branches. The sixth position (G226) only experienced a substitution to adenosine along external branches. Thus, the phylogenetic analysis revealed that the second, fifth and sixth positions were the best candidates for genetic analysis.

GA#2 was also under significant constraint. Whereas more of the sites experienced guanosine to adenosine or adenosine to guanosine substitutions, internal branch substitutions were not detected for sites within this motif, implicating critical importance to activity. Again, we chose positions within this motif for mutagenesis based on level of conservation: positions two (G328) and three (G329), which only experienced substitutions to adenosine, and position four (A330), which was slightly more flexible, experiencing substitutions to both cytosine and guanosine. We chose uridine to replace A330 because we did not observe uridine at this position in any of the 393 patient isolates, in contrast to cytosine and guanosine substitutions, which we observed along external branches. To evaluate a posited role in virus growth, we introduced the substitutions to the 5′UTR of HIV^{NL4-3} in molecular clones designated mGA#1, mGA#2 and mGA#1,2 (Figure 2A).

To propagate virus, we transfected these molecular clones into HEK293 cells. We collected cell-free medium containing viruses for use in infectivity assays on Jurkat T-cells. Jurkat cells were infected by spinoculation for 2 h, washed twice, and cultured for 21 days. At 2-day intervals, medium was collected from each sample, replenished with fresh RPMI medium, and virus in cell-free supernates were measured by RT assays. We detected infection of WT by Day 6, which peaked at Day 10, in three replicate experiments (Figure 2B). In the same time-frame, we could not observe infection by mGA#1,2 or the individual motif mu-

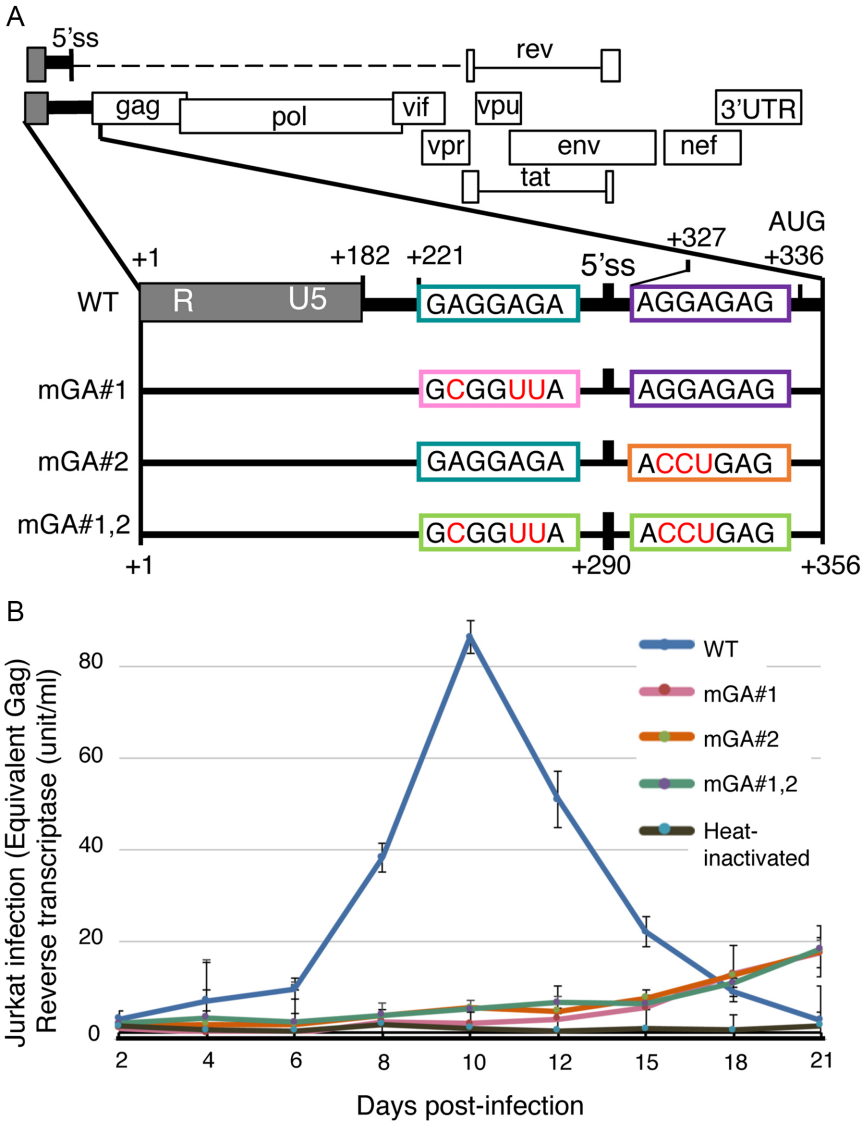

Figure 2. Guanosine-adenosine (GA)-motifs flanking both sides of the 5′ splice sites are necessary for efficient viral replication. (**A**) Diagram illustrating the HIV US RNA. The splicing pattern necessary for viral protein expression is enlarged of the 5'untranslated region (5′UTR). The gray box depicts the RU5 (+1–182) region of the long terminal repeat. The thin black line represents the total length of the 5′UTR (+1–356). Labels indicate nucleotide position of GA-motifs represented in boxes, 5′ splice site (5'ss), and start codon, AUG. The GA#1 (GAGGAGA) motif is in the teal box and the GA#2 (AGGAGAG) motif is in the purple box, shown for WT. The nucleotide substitutions within GA#1 or #2 are shown in red font (mGA#1: Pink, mGA#2: Orange, mGA#1,2: Green). Numbers represent the nucleotide position within the 5′UTR. Labeled white rectangles represent the exons of HIV genes and the gaps represent the introns removed from all spliced RNAs. (**B**) Jurkat cells were infected with cell-free supernatant fractions containing equivalent Gag (5 ng) by spinoculation. The cells were washed and maintained in fresh RPMI medium. The culture supernates were collected at 48 h intervals for 21 days and measure using a reverse transcriptase (RT) activity assay with PicoGreen. RT units per ml were determined based on a standard curve of recombinant RT enzyme. Error bars represent the mean ± standard error of three independent experiments.

tants (mGA#1 and mGA#2) (Figure 2B). Comparable to the heat-inactivated WT virus, the mGA clones exhibit undetectable growth for up to 18 days (Figure 2B). We speculated that this resurgence is attributable to second-site mutations, based on previous studies that observed positive selection for molecular clones mutated around the 5′ss (7,17,24).

In summary, the substitution of three guanosine or adenosine residues in either GA#1 or GA#2 is sufficient to significantly attenuate HIV replication. These mutations may reduce the quantity and quality of virions by: (a) altering cellular post-transcriptional processing of the HIV primary RNA early on, which is necessary and critical to synthesize the components of virions; (b) interfering with the structural conformation of the HIV 5′UTR crucial to late events in replication or (c) corrupting early and late events in replication. Hence, we designed experiments to investigate early and late replication events in the retroviral lifecycle. Given the sequence conservation observed in the retroviruses, we posited the GA-motifs may underlay a common molecular basis for activation of the retrovirus genome-length RNA.

The guanosine and adenosine motifs flanking the 5′ splice site activate intron retention

Because the conserved guanosine and adenosine residues flank the 5′ss, 5' intron removal eliminates the second motif (mGA#2) (Figure 2A). Hence, the GA-motifs are present solely in unspliced (US) HIV RNA that encodes the virion structural proteins and genomic RNA, and are not present in the singly or multiply spliced transcripts (SS, MS, respectively). We hypothesized the GA-motifs contribute to activating the US RNA and evaluated HIV post-transcriptional expression. We isolated steady-state RNA from transfected cells and performed RT and qPCR assays. HIV gene-specific primers distinguished US RNA (gag) from the SS (vpr-env) and MS (tat) transcripts and copies of US, SS, and MS HIV transcripts in the same cDNA samples were determined using standard curves generated on HIV cDNA.

With the WT virus, the population of US transcripts was significantly greater than the populations of SS or MS transcripts (Figure 3A). In contrast, mGA#1,2, mGA#1 and mGA#2 viruses had US transcript levels that were lower than the spliced transcripts (Figure 3A). These data were statistically significant as determined from five independent experiments (Supplementary Table S1). On the other hand, the total number of HIV transcripts was not significantly different between the molecular clones (Supplementary Table S1). We concluded the GA-motif mutations downregulate the population of HIV pre-mRNA retaining introns.

Next, we performed Gag ELISAs on cell-free and cell-associated lysates from the transfected cells. As expected, Gag production from WT virus was significantly different from that of mGA#1,2, mGA#1 or mGA#2 (Figure 3B). We observed similar results in the cell-free and cell-associated samples, indicating virion release was not impaired (Figure 3B). We observed similarity between the levels of Gag protein and gag (US) RNA, suggesting that the mutations affected early events in cellular post-transcriptional processing of the HIV primary RNA that are critical to synthesize components of virions. Because this deficiency was insufficient to account for the severe viral replication defect on T-cells (Figure 2B), we posited interference in late events in replication may be attributable to deficient US RNA packaging into virions.

We proceeded to measure the number of US RNA copies in equivalent virion preparations using the RT-qPCR. HEK293 cells were transfected with HIV^{NL4-3} or GA mutants. After 24 h, the cell-free supernatant fractions were collected and filtered, and Gag was quantified by ELISA (Figure 3B). Samples containing equivalent amounts of Gag were centrifuged on 20% sucrose to collect virions and extract RNA for RT-qPCR. The virion-associated US RNA copies were determined and normalized to the cell-associated copies. We found conclusively that US RNA copies are packaged into equivalent WT virions, whereas the number of US RNA copies packaged into mGA#1,2, mGA#1 or mGA#2 virions was reduced by at least 5-fold (Figure 3C). These results demonstrate the critical importance of the guanosine-adenosine sequences to effective viral RNA packaging. Our results here recapitulate previously published mutagenesis of guanosine residues overlapping those in GA#2. These studies showed deficient dimerization and genomic RNA packaging (17). Because

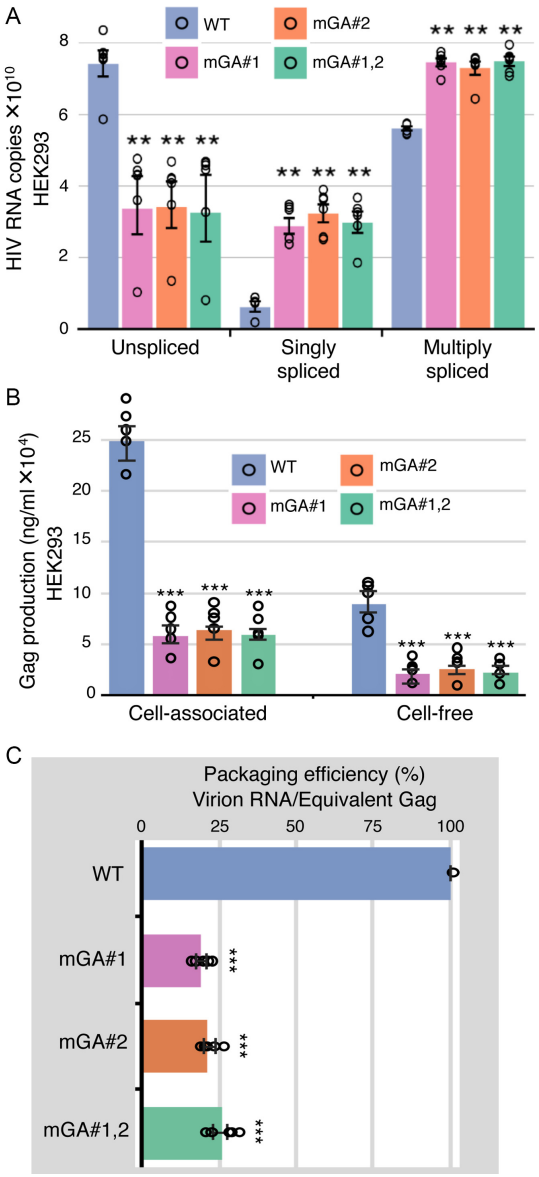

Figure 3. The GA-motifs regulate retention of unspliced HIV RNA, thereby activating virion production and infectivity. HEK293 cells were transfected with mGA#1, mGA#2, mGA#1,2 or WT and at 24 h, cell lysates and cell-free medium were harvested and HIV RNA and Gag protein content were measured. The bar graphs present the mean ± standard error of five independent experiments and the small circles indicate the range of precise values from the individual experiments. Bar colors indicate molecular clones: blue, WT; magenta, mGA#1; orange, mGA#2; green, mGA1,2. $^{**}P \leq 0.005$, $^{***}P \leq 0.001$. (**A**) RNAs were extracted in TRIzol and equivalent amounts were treated ± RT. Validated cDNA samples were determined to be free of plasmid contamination. Each cDNA was subjected to qPCR with primers specific to unspliced (US), singly spliced (SS), or multiply spliced (MS) HIV transcripts, as indicated on the x-axis. Copy numbers were calculated relative to standard curves generated using HIV cDNA. (**B**) Gag production was measured using ELISA on cell lysates and cell-free medium ($n = 5$). (**C**) Equivalent Gag containing cell-free medium was layered over 2 ml of 20% sucrose and virions were collected by centrifugation. Virion RNA was extracted using TRIzol and subjected to RT-qPCR using primers specific to US HIV RNA and the virion-associated US RNA copies were determined and normalized to the cell-associated copies. Bar graph presents packaging efficiency, defined as: US RNA copies in equivalent extracellular Gag. WT is set as 100% and error bars and *P* values document differences from the treatment groups are statistically significant.

mGA#2 is positioned in proximity to nucleotide pairings in the core encapsidation signal (Supplementary Figure S2), the mutations likely diminish interaction with Gag and possibly other components of genomic RNP.

In sum, the GA-motif mutations interfere with early and late events in HIV replication that account for the attenuated replication of mGA viruses (Figure 2B). We posit that the fundamental basis for this mechanism is upregulated alternative splicing, which is regulated by cellular RNA-binding protein(s) recruited during transcription. Previously, RNA affinity-coupled proteomics of HIV, SNV, and RSV 5′UTRs identified some of the same nuclear RNA-binding proteins, consistent with a conserved molecular basis for post-transcriptional expression (25). We sought to identify the cellular proteins that recapitulate the phenotype observed in the GA#1 and GA#2 loss-of-function mutants.

The SFPQ/PSF binding to unspliced RNA is abrogated by mutation of the GA-motifs

Of fifty-eight proteins identified in the prior 5′UTR proteomic analysis, three are splicing regulatory factors already known to modulate HIV Gag production: SRSF1, SRSF10 and SFPQ/PSF (25). SRSF1 binds GA-sequence motifs in the HIV gag ORF and exon splicing enhancer 3, which modulates splicing (26–28). Further, SRSF1 competes with Tat to bind the HIV transactivation responsive region (TAR) in the first 58 nucleotides of the HIV 5′UTR (28). SRSF10 regulates the alternative splicing of HIV (28,29). SFPQ/PSF is involved in Rev/Rev-responsive element-dependent expression of HIV RNA through an interaction with sequences in the 5′intron (30,31). Given these established roles for RNA-binding proteins, we investigated activity in context of the 5′UTR. To begin, we carried out HIV RNA co-immunoprecipitation assays (RIP) to establish whether endogenous SFPQ/PSF, SRSF1 or SRSF10 co-precipitated HIV WT RNA and any differential binding attributable to mGA#1 and 2.

HEK293 cells were transfected with the HIV molecular clones for 24 h, washed with PBS and treated with RIPA buffer containing protease inhibitors. IgG or specific antiserum was conjugated to magnetic beads and incubated at 4°C with lysates for 2 h. The conjugated immune complexes were washed four times with cold RIPA buffer. RNA was extracted from the isolated immune complexes, and the input lysates and subjected to RT-qPCR with HIV-specific primer pairs. We measured the difference in US RNA copies in the immune complexes over the matched IgG controls.

Western blotting of the input samples verified the presence of each RNA-binding protein in the cell lysates. The IP samples demonstrated enrichment of SFPQ/PSF, SRSF1 and SRSF10 immune complexes by the specific antibodies and not the IgG negative control (Figure 4, representative of three experiments). Previous HIV studies had documented the co-localization of SFPQ/PSF with p54nrb (NONO) and implicated a significant role for these proteins in nuclear retention (32). Consistent with those results, western blot with anti-p54nrb demonstrated IP of SFPQ/PSF and p54nrb heterodimer (Figure 4). IPs of SRSF1 and SRSF10 did not exhibit association with p54nrb. Table 1 summarizes the results of three experiments measuring the copies of WT or mGA#1,2 US RNA enriched in each immune complex relative to IgG controls. The SFPQ/PSF IP significantly enriched WT RNA compared with mGA#1,2 RNA ($P \leq 0.01$). For the SRSF1 and SRSF10 precipitates, no significant enrichment of WT RNA over the mGA#1,2 RNA was observed ($P \leq 0.4$). Taken together, these results demonstrate SFPQ/PSF binds HIV US RNA. This binding is sensitive to mGA#1,2. By contrast, SRSF1 and SRSF10 bind HIV US RNA and are not sensitive to m#GA1,2. Hence, we prioritized SFPQ/PSF for further study.

The *in vitro* binding of SFPQ/PSF to the HIV 5′UTR is disrupted by GA-motif mutations

Previously HIV US RNA was observed to bind SFPQ/PSF through sequences in the gag ORF, while the 5′ss and flanking sequences were deleted (30,33,34). We sought to verify direct interaction between the 5′UTR and SFPQ/PSF, positing that mGA#1,2 mutation could disrupt this interaction. First, we synthesized WT and mGA-5′UTRs (+1–356) using T7 RNA polymerase. We purified recombinant (r) human SFPQ/PSF, SRSF1 and SRSF10; the latter two served as negative controls. Protein purity was demonstrated by SDS-PAGE (Figure 5A). Gel-filtration chromatography recapitulated previous *in vitro* results that rSFPQ/PSF forms homodimers (Supplementary Figure S3) (35). Each HIV 5′UTR was incubated at a 1:3 molar ratio with recombinant protein in EMSA binding buffer at room temperature for 1 h. The RNA-protein complexes were resolved by native gel electrophoresis in 1% agarose. We observed robust electrophoretic mobility shifts of WT RNA by rSFPQ/PSF, rSRSF1 and rSRSF10 (Figure 5B, Supplementary Figure S4A and B). The shifts were abrogated by heat inactivation of each protein (designated by Δ). The incubation of SRSF1 or SRSF10 with the mGA#1, mGA#2 or mGA#1,2 UTR produced robust mobility shifts (Supplementary Figure S4A and B). On the other hand, SFPQ/PSF did not shift mGA#1, mGA#2, or mGA#1,2 RNA (Figure 5B). These results indicate that GA-motifs, alone or in combination, are necessary for direct binding of SFPQ/PSF homodimer to the HIV 5′UTR (Figure 5B). In agreement with the RNA IP of HIV RNA in cells, SFPQ/PSF binds directly to the HIV 5′UTR through an interaction that is sensitive to these GA mutations, whereas SRSF1 and SRSF10 recapitulate *in vitro* binding that is insensitive to GA mutations. We proceeded to conduct a cell-based analysis of SFPQ/PSF activity on the post-transcriptional expression of HIV unspliced and alternatively spliced RNAs.

Because previous studies had reported that overexpression of SFPQ/PSF increases MS RNAs and reduces US RNA (30), we sought to validate these observations using our RT-qPCR assay. We also evaluated overexpression of SRSF1 or SRSF10 because previous studies observed this overexpression to increase US transcripts and reduce SS and MS RNAs (26,27). HEK293 cells were transfected for 24 h with molecular clones and plasmids encoding epitope-tagged SRSF1, SRSF10, or SFPQ/PSF. Cells were collected, lysed in RIPA buffer, and subjected to immunoblotting to evaluate levels of the FLAG-tagged proteins and Gag (Figure 6A). Gag ELISAs on the cell lysates and cell-free supernatant fractions were used to quantify Gag expression (Figure 6B). The RT-qPCR with HIV-specific primers (US, SS, and MS) was performed to quan-

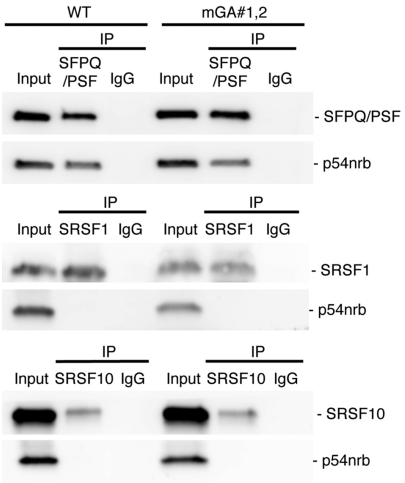

Figure 4. RNA-protein co-precipitation enriched SFPQ/PSF-p54nrb heterodimer. RNA co-immunoprecipitation (IP) was conducted, and RT-qPCR measured the unspliced WT and mGA#1,2 HIV RNA association with SFPQ/PSF, SRSF1, SRSF10 and whether or not SFPQ/PSF co-precipitated p54nrb. HEK293 cells transfected with HIV WT or GA#1,2 molecular clones. Cells were washed, cell contents were harvested in RIPA, and nuclei were removed by centrifugation. Lysates were incubated with magnetic beads conjugated with IgG or antiserum to either SFPQ/PSF, SRSF1 or SRSF10 for 2 h at 4°C. The beads were washed four times in RIPA and RNA was extracted, or bound proteins were eluted and immunoblotted.

Table 1. HIV unspliced RNA co-immunoprecipitation by SFPQ/PSF is sensitive to substitutions in the GA-motifs

RNA co-precipitate	Replicate Experiment			*P*-value*
HIV molecular clone	Exp. 1	Exp. 2	Exp. 3	
SFPQ/PSF IP				
WT	2.7×10^6	1.0×10^7	1.3×10^7	
mGA#1,2	1.0×10^6	1.5×10^6	1.8×10^6	
Ratio	*2.7*	*6.9*	*7.3*	*Baseline*
SRSF1 IP				
WT	1.1×10^6	1.8×10^6	1.8×10^6	
mGA#1,2	1.0×10^6	1.2×10^6	1.6×10^6	
Ratio	*1.1*	*1.5*	*1.1*	*0.04*
SRSF10 IP				
WT	7.4×10^5	1.3×10^6	1.4×10^6	
mGA#1,2	8.2×10^5	1.1×10^6	1.3×10^6	
Ratio	*0.9*	*1.2*	*1.1*	*0.03*

HEK293 cells were transfected with the indicated HIV molecular clones for 24 h and cell lysates were prepared in RIPA buffer and aliquots used for RNA extraction in TRIzol (Input) or incubated with the indicated antiserum or IgG control (Figure 4). The IP reactions were extracted in TRIzol and co-precipitated RNAs were subjected to RT-qPCR. The US RNA copies were determined relative to standard curves and normalized to Input RNA copies. The reported values are the difference between RNA copies in the IP and IgG negative control. The ratios measure enrichment in co-precipitated WT RNA compared to mGA#1,2 RNA. *$P \leq 0.05$, Significant difference from ratio of SFPQ/PSF.

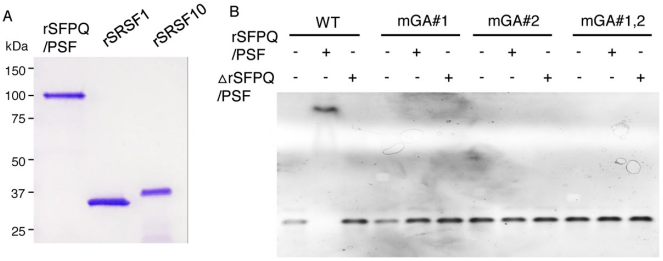

Figure 5. Three-nucleotide substitutions in either GA-motif abrogate in-solution binding of SFPQ/PSF to the HIV 5′UTR. SFPQ/PSF binding to 5′UTR was evaluated *in vitro* by synthetic RNA electrophoretic mobility shift assays (EMSAs) and *in vivo* by RNA immunoprecipitation (RIP). Binding of recombinant proteins to the 5′UTR was affected by three-nucleotide mutations in the GA-motifs. (**A**) Recombinant human SFPQ/PSF (runs at 100 kDa) (30,45), SRSF1 (28 kDa) and SRSF10 (37 kDa) were expressed in *Escherichia coli*, purified using Ni-NTA followed by gel-filtration chromatography (Supplementary Figure S3), resolved by electrophoresis on SDS-polyacrylamide 4–15% gradient gels (Bio-Rad), and stained by Coomassie Brilliant blue. EMSAs were performed on WT, mGA#1, mGA#2, mGA#1,2 *in vitro*-transcribed synthetic RNA using native or heat-treated recombinant polypeptide (Supplementary Figure S3). (**B**) Using EMSA, rSFPQ/PSF protein bound to WT, but not to mGA#1, mGA#2, or mGA#1,2. EMSA was performed using synthetic WT, mGA#1, mGA#2, or mGA#1,2 RNAs (+1–356) and 1:3 molar equivalents of recombinant SFPQ/PSF that was native or heat-inactivated, incubated with the DNA probe for 1 h at room temperature, resolved in 1% agarose, and visualized with ethidium bromide and ultraviolet light.

tify the steady state of HIV RNAs. From these experiments, we observed that exogenous expression of FLAG-SFPQ/PSF reproducibly reduced the numbers of US and SS HIV transcripts, which are Rev/RRE-dependent, and increased abundance of MS RNAs, which are Rev/RRE-independent (Figure 6C) (Supplementary Table S2). In contrast, SRSF1 and SRSF10 increased the US isoforms and reduced both SS and MS transcripts. Our qPCR results recapitulated that SFPQ/PSF overexpression acts on the pool of HIV transcripts susceptible to Rev/RRE transactivation, influencing metabolism at a step downstream of pre-mRNA processing (30,26).

Next, we infected Jurkat T-cells with equivalent aliquots of Gag p24 using spinoculation. Twenty-four hours later, supernates were collected and screened by Gag ELISA. The virions from mock-transfected or FLAG-SFPQ/PSF-transfected cells exhibited equivalent infectivity. By comparison, FLAG-tagged SRSF1 or SRSF10 diminished virion infectivity (Figure 6D). These experiments showed that exogenous SFPQ/PSF affected early, but not late, events in viral replication. By contrast, SRSF1 and SRSF10 altered both RNA expression and virion infectivity in the Jurkat target cells. These results reinforce the qualitative differences between SFPQ/PSF and the SRSF1 and SRSF10 RNA-binding proteins on HIV RNA.

Balanced expression of HIV RNA isoforms is abrogated by SFPQ/PSF downregulation

RNA interference by shRNA expression plasmids was used to downregulate SFPQ/PSF in HEK293 cells. Twenty-four hours post-transfection, cell lysates were harvested for western blotting and extraction of cellular RNA for RT-qPCR with *sfpq/psf* and *gapdh* gene-specific primers. Western blotting documented SFPQ/PSF protein was downregulated (Figure 7A) and RT-qPCR verified sfpq/psf mRNA was

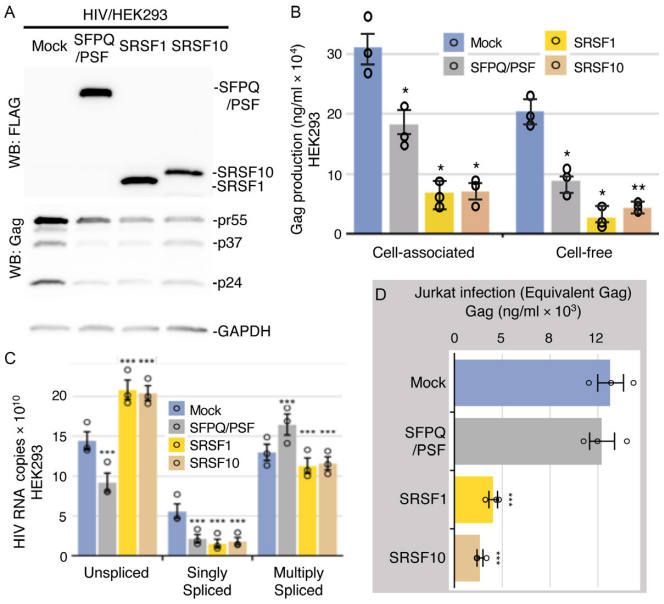

Figure 6. The exogenous expression of SRSF1 or SFSF10, but not SFPQ/PSF, diminishes virion infectivity. HEK293 cells were co-transfected with pNL4-3 and expression plasmids encoding FLAG-tagged SFPQ/PSF, SRSF1 or SRSF10 for 24 h. Cell lysates and cell-free medium were collected and assayed. Bar graphs present mean and standard error of three independent experiments, and the precise values from the individual experiments are denoted by small circles. Statistical significance: $*P \leq 0.05$, $**P \leq 0.005$ and $***P \leq 0.001$. Bar colors indicate exogenous expression of: blue, mock; gray, FLAG-SFPQ/PSF; yellow, FLAG-SRSF1; brown, FLAG-SRSF10. (**A**) Equal amounts of cell lysate were resolved by SDS-PAGE and immunoblotted with primary antiserum to FLAG, Gag, or GAPDH, followed by HRP-conjugated secondary rabbit or mouse antibodies as needed. Results of a typical experiment are presented. (**B**) ELISA results of Gag content in equivalent volumes of cell-associated or cell-free samples, as labeled on the x-axis. (**C**) RNA preparations were subjected to cDNA preparation, followed by qPCR with HIV-specific primers. RNA abundance (copy number) was determined based on standard curves generated with cDNA from HIV-infected cells. (**D**) Jurkat cells (1×10^6) were infected in a 12-well tissue culture plate with equivalent cell-free Gag (5 ng) supernates derived from Figure 6B using spinoculation for 2 h, washed twice in RPMI medium, and cultured for 24 h. The cell-free media were collected from the cultured cells and subjected to Gag ELISA.

downregulated relative to the gapdh control (Supplementary Figure S5A). Results of five experiments documented statistically significant downregulation of SFPQ/PSF RNA (Supplementary Figure S5A). These cells were transfected with HIV molecular clones with or without GFP-PSF expression plasmid. After 24 h, cell lysates were harvested, and samples prepared for immunoblotting or RT-qPCR (Figure 7A, Supplementary Figure S5A).

Immunoblotting with SFPQ/PSF (Figure 7A) ascertained endogenous SFPQ/PSF was downregulated and rescued through exogenous expression of GFP-PSF (*PSF rescue*). Fluorescence microscopy documented the transfection were efficient (Supplementary Figure S5B). Gag production from WT was significantly reduced by SFPQ/PSF downregulation to a level similar to the mGA#1, #2 or #1,2 molecular clones absent the SFPQ/PSF depletion (Figure 7B).

WT Gag levels were completely rescued by exogenous GFP-PSF, whereas mGA#1, #2 or #1,2 were nonresponsive (Figure 7B). The RNA analysis documented similar trends between the molecular clones; SFPQ/PSF downregulation significantly reduced WT US and increased SS and MS transcript to levels similar to mGA#1, #2 or #1,2 (Figure 7C). Physiologic expression of GFP-PSF completely restored balanced expression in the WT molecular clone, whereas mGA#1,2 or #1 or #2 were nonresponsive (Figure 7C). Statistical analysis of replicate experiments results documented strong significance to the conclusion SFPQ/PSF downregulation and rescue recapitulated the activity of mGA molecular clones (Supplementary Table S3). The results of the WT molecular clone recapitulated those of Kula *et al.* (31), indicating SFPQ/PSF depletion acts on HIV pre-mRNA processing.

To analyze HIV RNA packaging efficiency, the cell-free medium from these virus-producing cells was collected. Gag ELISA was performed to determine sample volumes having equivalent Gag p24 (Figure 7B). Samples containing equivalent amount of Gag were centrifuged on 20% sucrose to collect virions to extract RNA for RT-qPCR and the virion-associated US RNA copies were determined and normalized to the cell-associated copies (from Figure 7C). SFPQ/PSF downregulation did not change the packaging efficiency of WT RNA. While the exogenous expression of GFP-PSF was sufficient to rescue the copies of WT RNA in cells, packaging efficiency remained similar to control (Figure 8A). The mGA#1, #2 and #1,2 HIV RNA packaging efficiency was reduced to ∼20% and remained the same during SFPQ/PSF depletion or rescue (Figure 8A) and recapitulated the results of Figure 3C.

Equivalent aliquots of the virus analyzed in Figure 7B (5 ng Gag p24) were used to infect MT-4 cells by spinoculation for 2 h. The cells were washed twice and then cultured in fresh medium. Virus growth was measured after 48 h by Gag ELISA of cell-free medium. As shown in Figure 8B, the growth of WT virus that was generated in cell experiencing SFPQ/PSF downregulation was as infectious as WT control, indicating SFPQ/PSF activity in producer cells does not inhibit late steps in HIV replication. In agreement with the viral growth curves in Figure 2B, the mGA#1, mGA#2, and mGA#1,2 virions had poor infectivity (Figure 8B), which is attributable to deficient packaging of virion RNA.

The SNV GA-motifs recapitulate activation of unspliced RNA and direct binding to SFPQ/PSF

Given the conservation of similar GA-motifs between HIV and other retroviruses, we sought to characterize their activity in another representative retrovirus. For this purpose, we chose SNV because previous studies described 5′UTR deletions that produced log reductions in virus titer on D17 cells that were only partially explained by deficient SNV RNA packaging (15). We introduced deletions to the SNV 5′UTR and the molecular clones were propagated in D17 cells. We evaluated Gag protein using immunoblot. As shown in Figure 9A, the deletions overlapped GA-motif 1 (Δ165) or GA-motif 2 (Δ300) and derivative Δ300+Rescue was constructed to determine whether or not Δ300 is rescued by GA-motif 1.

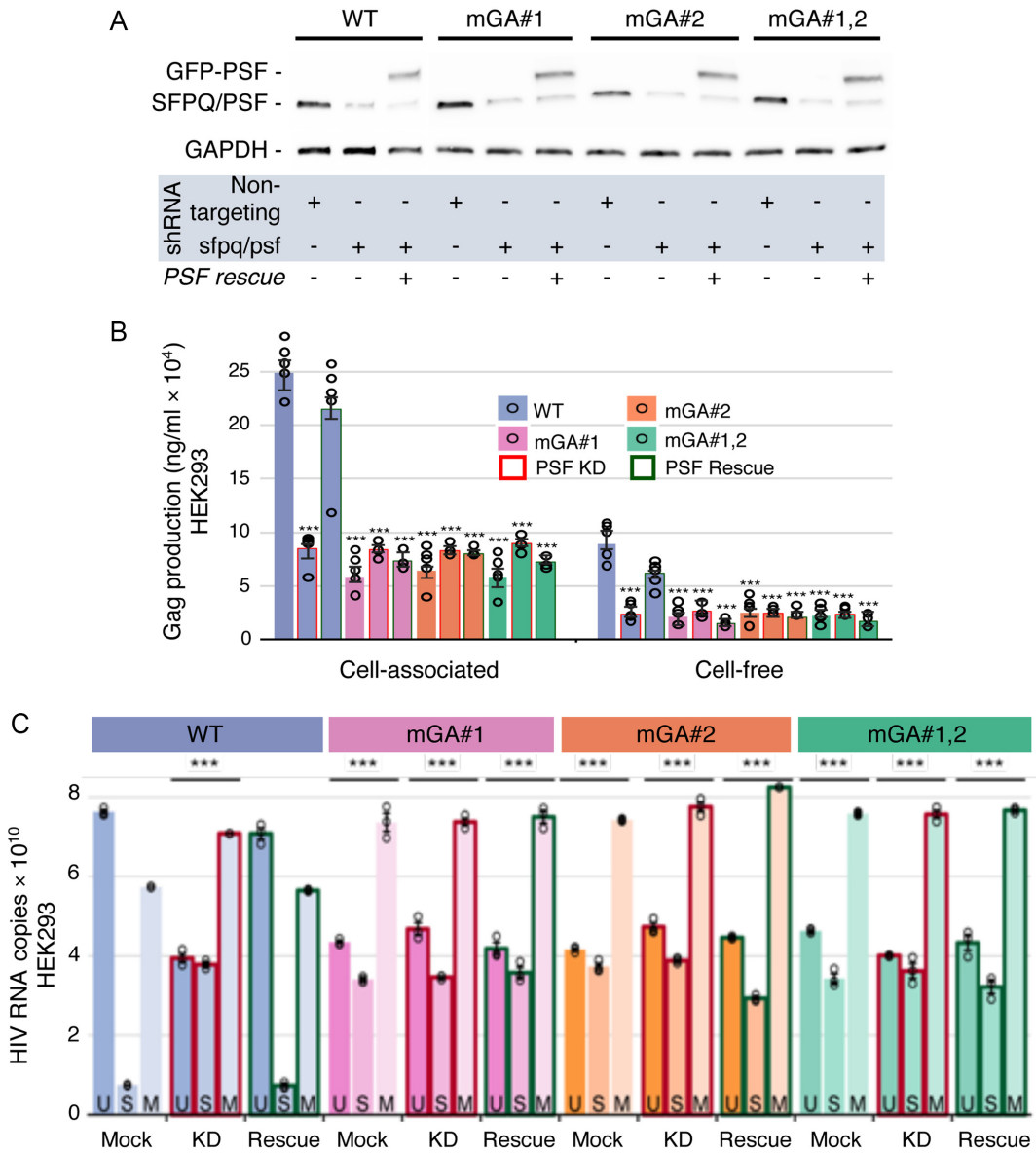

Figure 7. Downregulation of SFPQ/PSF attenuates retention of HIV US RNA, which is rescued by the exogenous expression of SFPQ/PSF. HEK293 cells were transfected with a plasmid encoding SFPQ/PSF shRNAs for 24 h, and then they were co-transfected with a GFP-PSF expression plasmid (*PSF rescue*) (30) and pNL4-3 WT or mGA#1, #2 or #1,2 molecular clones. Cells and cell-free medium were harvested for lysate preparation and RNA isolation. Statistical significance: ***$P \leq 0.001$. Dots are values from the individual experiments. (**A**) Representative immunoblot using antiserum to SFPQ/PSF or GAPDH on equal amounts of cell lysate subjected to SDS-PAGE and transfer. (**B**) Gag ELISAs on cell lysate and cell-free medium samples were executed and the values derived from the standard curve is presented. The percentage of cell-free Gag relative to total Gag was similar among the samples, indicating no difference in viral particle release. Data indicate the mean of three independent experiments with standard error. (**C**) RNA preparations from the transfected cells were subjected to reverse transcription with random hexamers and qPCR with primers specific to unspliced (U), singly spliced (S), or multiply spliced (M) HIV cDNA. RNA standard curves were generated using cDNA from HIV-infected cells using each primer set. The shRNA downregulated samples are designated by red borders on the bars, and downregulation and transfected rescue plasmid have green borders. Bars indicate the mean of five (WT, PSF KD, PSF Rescue and mGA#1, mGA#2 and mGA#1,2) or three (mGA-PSF KD and mGA-PSF Rescue) independent experiments.

Given uneven transient transfection efficiencies in these cells, we selected permanently transfected cells for further experimentation. SNV copies were normalized using PCR with gag-specific primers. SDS-PAGE gels were loaded with samples containing equivalent gag copy numbers in 3-fold dilution series, and anti-eGFP antiserum was used to verify expression of the eGFP-tagged Gag structural protein (Figure 9B). The SNV WT cell lysates validated robust Gag production. Compared to WT, Δ165 and Δ300 exhibited weak Gag production; the dilution series indicated a reduction by a factor of ≥9. In the case of Δ300+Rescue, robust Gag production was restored to levels similar to WT. Based on results of replicate experiments, we concluded that the SNV GA#1 motif is functionally redundant with the SNV GA#2 motif. We posited the SNV GA-motifs recapitulate the binding to SFPQ/PSF and activate retention of SNV transcripts in the US form.

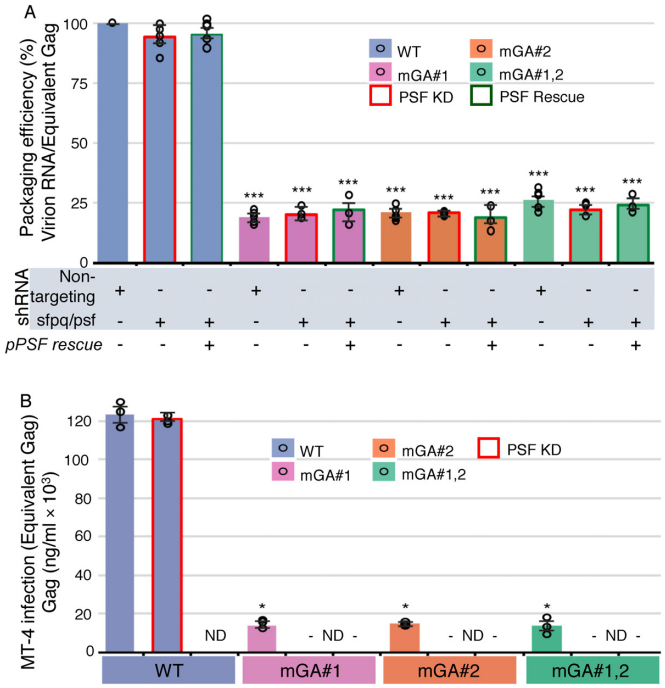

Figure 8. Downregulation of SFPQ/PSF does not hinder genomic RNA packaging or infectivity of HIV GA#1, #2, or #1,2 viruses. HEK293 cells were transfected with sfpq/psf shRNA plasmid or mock control for 24 h, and then they were co-transfected with GFP-PSF or mock expression plasmid and pNL4-3 WT or mGA molecular clones for 24 h. Cell-free supernatant fractions were filtered and normalized to Gag equivalents. Virions were collected by centrifugation on 20% sucrose. Virion RNA was extracted using TRIzol and subjected to RT-qPCR using primers specific to unspliced (US) HIV RNA and the virion-associated US RNA copies were determined and normalized to the cell-associated copies. Data from samples receiving shRNA are designated by red borders on the bars; samples treated with shRNA and rescue plasmid have green borders. Results present the mean ± standard error of three independent experiments. Statistical significance: $*P \leq 0.05$, $***P \leq 0.001$. ND: Experiments not done. (**A**) RNA was isolated from virions and subjected to RT with random hexamers, followed by qPCR with HIV-specific primers. US HIV RNA copies were determined using standard curves, and packaging efficiency was determined by normalization of virion RNA to the cell-associated copies ($n = 5$: WT, PSF KD, PSF Rescue and mGA#1, mGA#2 and mGA#1,2; $n = 3$: PSF KD and PSF Rescue for GA mutants). (**B**) The cell-free medium from producer cells containing equivalent Gag (5 ng) was incubated with MT-4 lymphocytes by spinoculation for 2 h, and then the lymphocytes were washed twice. After 48 h, the infections were monitored by Gag ELISA of MT-4 culture media.

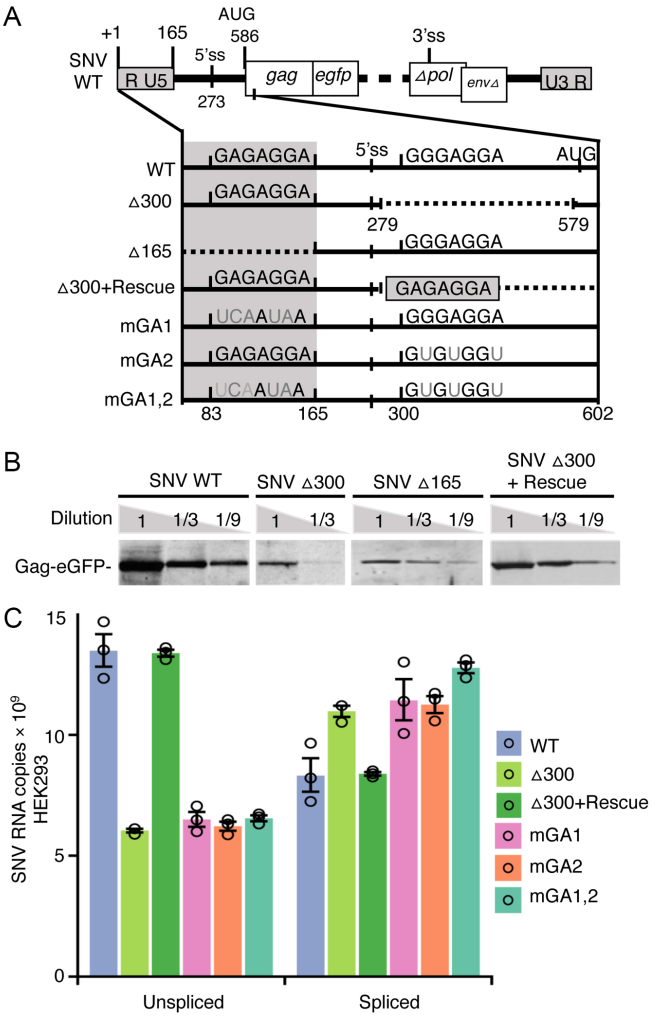

Figure 9. Two SNV GA-motifs contribute to retention of SNV unspliced mRNA translated to Gag. (**A**) Diagram of SNV WT and derivatives. 5′terminal rectangles, R, U5 regions of the 5′ SNV LTR; mutations in RU5 and distal 5′leader on line diagram; SNV RU5; 5′ and 3′ splice site (ss); *gag, pol,* and *env* genes are labeled, *green fluorescent protein (gfp)* gene; 3′R, U3 represented in gray box. The black solid line represents the RNA used in the experiment and dotted line represents deleted sequence. The numbers represent the nucleotide positions. The GA-motif substitution mutations are labeled in gray text. (**B**) D17 cells were permanently transfected with SNV molecular clones, and copy number was determined by real-time PCR. Protein aliquots standardized to SNV copy number were serially diluted 3-fold, loaded to immunoblots that were incubated with GFP antiserum. (**C**) HEK293 cells were transfected with SNV WT or mutant clones. Twenty-four hours post-transfection, RNA was isolated and subjected to RT-qPCR with SNV-specific primers to detect unspliced and spliced transcripts. The RNA copies were calculated using a standard curve generated from cellular cDNA. Bars and error bars represent the mean ± standard error, calculated on the log scale and converted back. Small circles indicate the values from individual experiments.

The EMSA examined direct binding between rSFPQ/PSF and the SNV 5′UTR. SNV WT (+1–602) exhibited robust mobility shift upon binding to rSFPQ/PSF, which was eliminated by heat inactivation of the recombinant protein (Supplementary Figure S6A, lanes 1–3). SNV Δ165 or SNV Δ300 transcripts containing either one of the GA-motifs did not shift upon addition of rSFPQ/PSF (Supplementary Figure S6A, lanes 4–9), whereas Δ300+Rescue exhibited the robust shift of WT sequence (Supplementary Figure S6A, lanes 10–12). These data suggest that SFPQ/PSF binds the SNV 5′UTR, and that the GA-motifs are critical to this interaction.

To precisely evaluate the SNV GA-motifs, we generated *in vitro* transcripts with targeted substitutions: GA-motif 1 (mGA1), GA-motif 2 (mGA2), GA-motif 1 and 2 (mGA1,2) (Figure 9A) and evaluated these transcripts in RNA electrophoretic mobility shift assays (EMSAs). As shown in Supplementary Figure S6B, the mutations eliminate direct binding of rSFPQ/PSF to SNV WT RNA. We introduced the GA-motif substitution mutations into the SNV vectors and transiently transfected HEK293 cells for 24 h. The cellular RNA was harvested, converted to cDNA, and subjected to qPCR with SNV *gag*- and *env*- specific

primers to measure the US RNA and spliced transcripts, respectively. Three independent experiments were performed, the average and standard error were calculated; robust statistically significant differences from WT were identified (Supplementary Table S4).

The US RNA was detected at higher levels for WT than for Δ300 (Figure 9C). The spliced transcripts were more abundant in Δ300 than in WT (Figure 9C). Δ300+Rescue restored the US RNA and spliced RNAs to the levels observed for WT. These results account for the diminished Gag production observed in Figure 9B. Each of the GA-motif mutants exhibited diminished US RNA and more abundant spliced RNA at levels similar to those of Δ300 (Figure 9C). From these data, we conclude that the SNV GA-motifs are required to retain US RNA at WT levels. Deletions or substitution of the GA-motifs significantly reduces Gag, the major virion structural protein. Our results add to the previous characterization of SNV vector RNAs experiencing severe replication defects (15). Deletions that include the GA-motifs downregulate SNV US RNA, the production of virions, and gRNP packaging, which taken together, account for the log-reduction observed in vector virus titer (15).

In sum, SNV and HIV GA-motifs are necessary to activate retention of US transcripts, and mutation of these motifs is sufficient to derepress alternatively splicing. From these data, we conclude that the phylogenetic analysis and functional analysis of two representative retroviruses has identified a common molecular basis activating the US genome-length RNA.

DISCUSSION

Retroviruses balance the relative amounts of alternatively spliced primary transcripts with US transcripts, but a conserved mechanism or primary sequence regulating this fundamental property remain unknown. We have identified nearly invariant GA-motifs on both sides of the 5′ss in several retroviruses and used phylogenetic reconstruction of ∼400 unique HIV isolates from patients to guide site-directed mutagenesis of two representative retroviruses. The evidence provided here indicates the GA-motifs regulate the balance of primary transcripts that become spliced or are retained in the US form.

SFPQ/PSF protein binds directly to the 5′UTR sequences of HIV and SNV in a manner dependent on the GA-motifs. Downregulation of SFPQ/PSF recapitulates the phenotype of the GA-motifs mutants. The quantitative comparisons of US, SS and MS mRNA demonstrated SFPQ/PSF acts on the HIV primary transcript upstream of RRE/Rev activity; depletion reduced US and increased SS and MS RNAs. On the other hand, overexpression of SFPQ/PSF perturbs a step downstream of pre-mRNA processing. Because of the high sensitivity afforded by the quantitative RNA analysis, the results have resolved fundamentally different outcomes of depletion or exogenous expression of SFPQ/PSF previously observed (30,31).

SFPQ/PSF is conserved in the natural hosts of the seven retroviruses investigated here (Supplementary Figure S7). The alignment of SFPQ/PSF from human, chimpanzee, macaque, bovine, murine, feline, and avian species revealed that the RNA-recognition motifs (RRM) (RRM1: 297–369 aa; RRM2: 371–452 aa) and C-terminal domain are highly conserved (Supplementary Figure S7). Extensive cell-based and biophysical studies have determined that SFPQ/PSF can exist as a homodimer or as a heterodimer with p54nrb (35,36). Well-characterized, the SFPQ/PSF homo- and hetero-dimers form with high affinity between the RRM2 of one monomer and the NOPS domain of the other monomer (35). While the results of gel-filtration chromatography show rSFPQ/PSF exists as a homodimer, the SFPQ/PSF immunoprecipitated from cells is heterodimeric with p54nrb/NONO (Figure 4). Because impaired polypeptide dimerization is caused by NOPS domain truncation (35), future experiments to disrupt this interface are expected to downregulate HIV gene expression at the post-transcriptional level.

SFPQ/PSF activity implicates a role for HIV RNA in paraspeckles

SFPQ/PSF regulation of HIV RNA was first documented by Felber and colleagues (30). In that study, SFPQ/PSF was demonstrated to bind the gag ORF. Mutation of adenosine and uridine nucleotides diminished this binding, leading to the designation of instability sequences, which contribute to Rev/Rev-responsive element (RRE)-dependence of the US and SS HIV RNAs. However, the test RNAs lacked the 5′ss and flanking residues, inclusive of GA-motif 2 (30,33,34), and a role for SFPQ/PSF in activating the retention of HIV US transcripts was not investigated. Because other studies have demonstrated a role for SFPQ/PSF in activating nuclear retention of adenosine-rich RNAs (31,32,37,38), the current findings warrant additional experiments to reconcile the role of SFPQ/PSF binding at the 5′UTR or within the gag ORF.

SFPQ/PSF RNPs accumulate in paraspeckles, which are nuclear bodies that become disaggregated during cell cycling, treatment with actinomycin D, or downregulation of long noncoding RNA NEAT1 (39,40). Given that cell division is a requirement for productive replication of genetically simple retroviruses (41), and NEAT1 downregulation (42) or actinomycin D (43) abrogate Rev/RRE activity, an ancillary role for paraspeckles in retention of retroviral US transcripts is a valid aim for future experiments. Further, experiments addressing whether crosstalk is required between SFPQ/PSF bound at 5′UTR and the gag ORF may also hold merit. We speculate the retroviral GA-motifs and the cognate RNA-binding protein described in this study provide the foundation for complete understanding how retroviruses subvert post-transcriptional control by cellular RNPs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Barbara Felber at the National Cancer Institute for providing the expression plasmids encoding SFPQ/PSF and shRNAs. The authors acknowledge

Deepali Singh, Nicole Placek for preliminary experiments on SNV and Carla Mavian for a draft of Supplementary Figure S7.

FUNDING

National Institutes of Health [R01 CA108882 and P50GM103297]. Funding for open access charge: University of Minnesota funds (to K.B.L.).
Conflict of interest statement. None declared.

REFERENCES

1. Leblanc,J., Weil,J. and Beemon,K. (2013) Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdiscip. Rev. RNA.*, **4**, 567–580.
2. Butsch,M. and Boris-Lawrie,K. (2002) Destiny of unspliced retroviral RNA: ribosome and/or virion? *J. Virol.*, **76**, 3089–3094.
3. Lykke-Andersen,J. (2001) mRNA quality control: Marking the message for life or death. *Curr. Biol.*, **11**, R88–R91.
4. Coffin,J.M., Hughes,S.H. and Varmus,H. (1997) *Retroviruses Plainview*, Cold Spring Harbor Laboratory Press, Plainview.
5. Miller,C.K. and Temin,H.M. (1986) Insertion of several different DNAs in reticuloendotheliosis virus strain T suppresses transformation by reducing the amount of subgenomic mRNA. *J. Virol.*, **58**, 75–80.
6. Katz,R.A. and Skalka,A.M. (1990) Control of retroviral RNA splicing through maintenance of suboptimal processing signals. *Mol. Cell. Biol.*, **10**, 696–704.
7. Katz,R.A., Kotler,M. and Skalka,A.M. (1988) cis-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implication for mechanisms of control. *J. Virol.*, **62**, 2686–2695.
8. Tazi,J., Bakkour,N., Marchand,V., Ayadi,L., Aboufirassi,A. and Branlant,C. (2010) Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. *FEBS J.*, **277**, 867–876.
9. Shimotohno,K. and Temin,H.M. (1981) Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. *Cell*, **26**, 67–77.
10. Wei,C.M., Gibson,M., Spear,P.G. and Scolnick,E.M. (1981) Construction and isolation of a transmissible retrovirus containing the src gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1. *J. Virol.*, **39**, 935–944.
11. Tabin,C.J., Hoffmann,J.W., Goff,S.P. and Weinberg,R.A. (1982) Adaptation of a retrovirus as a eucaryotic vector transmitting the herpes simplex virus thymidine kinase gene. *Mol. Cell. Biol.*, **2**, 426–436.
12. Mann,R., Mulligan,R.C. and Baltimore,D. (1983) Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell*, **33**, 153–159.
13. Watanabe,S. and Temin,H.M. (1983) Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. *Mol. Cell. Biol.*, **3**, 2241–2249.
14. Miller,A.D. and Rosman,G.J. (1989) Improved retroviral vectors for gene transfer and expression. *Biotechniques*, **7**, 980–982.
15. Yang,S. and Temin,H.M. (1994) A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *EMBO J.*, **13**, 713–726.
16. Katz,R.A., Terry,R.W. and Skalka,A.M. (1986) A conserved cis-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. *J. Virol.*, **59**, 163–167.
17. Russell,R.S., Hu,J., Beriault,V., Mouland,A.J., Laughrea,M., Kleiman,L., Wainberg,M.A. and Liang,C. (2003) Sequences downstream of the 5′ splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J. Virol.*, **77**, 84–96.
18. Miller,C.K., Embretson,J.E. and Temin,H.M. (1988) Transforming viruses spontaneously arise from nontransforming reticuloendotheliosis virus strain T-derived viruses as a result of increased accumulation of spliced viral RNA. *J. Virol.*, **62**, 1219–1226.
19. Boris-Lawrie,K. and Temin,H.M. (1995) Genetically simpler bovine leukemia virus derivatives can replicate independently of Tax and Rex. *J. Virol.*, **69**, 1920–1924.
20. Kucerova,L., Altanerova,V., Altaner,C. and Boris-Lawrie,K. (1999) Bovine leukemia virus structural gene vectors are immunogenic and lack pathogenicity in a rabbit model. *J. Virol.*, **73**, 8160–8166.
21. Hull,S. and Boris-Lawrie,K. (2002) RU5 of Mason-Pfizer monkey virus 5' long terminal repeat enhances cytoplasmic expression of human immunodeficiency virus type 1 gag-pol and nonviral reporter RNA. *J. Virol.*, **76**, 10211–10218.
22. Butsch,M., Hull,S., Wang,Y., Roberts,T.M. and Boris-Lawrie,K. (1999) The 5′ RNA terminus of spleen necrosis virus contains a novel posttranscriptional control element that facilitates human immunodeficiency virus Rev/RRE-independent Gag production. *J. Virol.*, **73**, 4847–4855.
23. Wong,R.W., Balachandran,A., Ostrowski,M.A. and Cochrane,A. (2013) Digoxin suppresses HIV-1 replication by altering viral RNA processing. *PLoS Pathog.*, **9**, e1003241.
24. Abbink,T.E.M. and Berkhout,B. (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J. Virol.*, **82**, 3090–3098.
25. Stake,M., Singh,D., Singh,G., Marcela,H.J., Kaddis,M.R., Parent,L.J. and Boris-Lawrie,K. (2015) HIV-1 and two avian retroviral 5′ untranslated regions bind orthologous human and chicken RNA binding proteins. *Virology*, **486**, 307–320.
26. Jablonski,J.A. and Caputi,M. (2009) Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *J. Virol.*, **83**, 981–992.
27. Paz,S., Lu,M.L., Takata,H., Trautmann,L. and Caputi,M. (2015) SRSF1 RNA recognition motifs are strong inhibitors of HIV-1 replication. *J. Virol.*, **89**, 6275–6286.
28. Caputi,M., Freund,M., Kammler,S. and Asang,C. (2004) A bidirectional SF2 / ASF- and SRp40-Dependent splicing enhancer regulates human immunodeficiency virus type 1 rev, env, vpu, and nef gene expression. *Society*, **78**, 6517–6526.
29. Shkreta,L., Blanchette,M., Toutant,J., Wilhelm,E., Bell,B., Story,B.A., Balachandran,A., Cochrane,A., Cheung,P.K., Harrigan,P.R. *et al.* (2017) Modulation of the splicing regulatory function of SRSF10 by a novel compound that impairs HIV-1 replication. *Nucleic Acids Res.*, **45**, 4051–4067.
30. Zolotukhin,A.S., Michalowski,D., Bear,J., Smulevitch,S.V, Traish,A.M., Peng,R., Patton,J., Shatsky,I.N. and Felber,B.K. (2003) PSF acts through the human immunodeficiency virus type 1 mRNA instability elements to regulate virus expression. *Mol. Cell. Biol.*, **23**, 6618–6630.
31. Kula,A., Gharu,L. and Marcello,A. (2013) HIV-1 pre-mRNA commitment to Rev mediated export through PSF and Matrin 3. *Virology*, **435**, 329–340.
32. Yarosh,C.A., Iacona,J.R., Lutz,C.S. and Lynch,K.W. (2015) PSF: nuclear busy-body or nuclear facilitator? *Wiley Interdiscip. Rev.*, **6**, 351–367.
33. Felber,B.K., Hadzopoulou-Cladaras,M., Cladaras,C., Copeland,T. and Pavlakis,G.N. (1989) rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 1495–1499.
34. Schwartz,S., Campbell,M., Nasioulas,G., Harrison,J., Felber,B.K. and Pavlakis,G.N. (1992) Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J Virol.*, **66**, 7176–7182.
35. Lee,M., Sadowska,A., Bekere,I., Ho,D., Gully,B.S., Lu,Y., Iyer,K.S., Trewhella,J., Fox,A.H. and Bond,C.S. (2015) The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res.*, **43**, 3826–3840.
36. Fox,A.H., Bond,C.S. and Lamond,A.I. (2005) P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell*, **16**, 5304–5315.
37. Cho,S., Moon,H., Loh,T.J., Oh,H.K., Williams,D.R., Liao,D.J., Zhou,J., Green,M.R., Zheng,X. and Shen,H. (2014) PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion. *Biochim. Biophys. Acta*, **1839**, 517–525.
38. Ray,P., Kar,A., Fushimi,K., Havlioglu,N., Chen,X. and Wu,J.Y. (2011) PSF suppresses tau exon 10 inclusion by interacting with a stem-loop structure downstream of exon 10. *J. Mol. Neurosci.*, **45**, 453–466.

39. Zhang,Q. and Jeang,K.-T. (2013) Long non-coding RNAs (lncRNAs) and viral infections. *Biomedicine*, **3**, 34–42.
40. Zhang,Z. and Carmichael,G.G. (2001) The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell*, **106**, 465–475.
41. Fritsch,E. and Temin,H.M. (1977) Formation and structure of infectious DNA of spleen necrosis virus. *J. Virol.*, **21**, 119–130.
42. Zhang,Q., Chen,C.Y., Yedavalli,V.S.R.K. and Jeang,K.T. (2013) NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio*, **4**, e00596–e00612.
43. Iacampo,S. and Cochrane,A. (1996) Human immunodeficiency virus type 1 Rev function requires continued synthesis of its target mRNA. *J. Virol.*, **70**, 8332–8339.
44. Schneider,T.D., Stormo,G.D., Gold,L. and Ehrenfeucht,A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, **188**, 415–431.
45. Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.*, **7**, 393–406.