Identification of New and Unusual *rev* and *nef* Transcripts Expressed by an HIV Type 1 Primary Isolate

Yolanda Vega, Elena Delgado, Cristina Carrera, Paloma Nebreda, Aurora Fernández-García, María Teresa Cuevas, Lucía Pérez-Álvarez, and Michael M. Thomson

Abstract

We analyzed RNA splice site usage in three HIV-1 subtype B primary isolates through reverse transcriptase polymerase chain reaction (RT-PCR) amplification of spliced RNAs using a fluorescently labeled primer, with computerized size determination and quantification of PCR products, which were also identified by clone sequencing. In one isolate, P2149-3, unusual and unreported spliced transcripts were detected. This isolate preferentially used for *rev* RNA generation a 3' splice site (3'ss) located five nucleotides upstream of A4a, previously identified only in a T cell line-adapted virus and in a group O isolate, and designated A4d. P2149-3 also used an unreported 3'ss for *rev* RNA generation, designated A4h, located 20 nucleotides upstream of 3'ss A4c. Additionally, unusual *nef* RNAs using 3'ss A5a and A7a and with exon composition 1.3.7 were identified. The identification of several unusual and unreported spliced transcripts in an HIV-1 primary isolate suggests a greater diversity of splice site usage in HIV-1 than previously appreciated.

HIV-1 RNAs ARE TRANSCRIBED from a single promoter at the 5' long terminal repeat and their relative expression is regulated through the alternative usage of splice sites. According to splicing patterns, HIV-1 transcripts can be assigned to three major categories^{1,2} (Fig. 1): (1) unspliced RNA, coding for Gag-Pol and Pol polyproteins; (2) doubly spliced (DS) transcripts, generated by excision of introns overlapping Gag-Pol and Vpu-Env open reading frames, coding for Tat, Rev, Nef, and Vpr proteins; and (3) singly spliced (SS) transcripts, generated by excision of the Gag-Pol intron, coding for Env, Vpu, Vif, Vpr, and a truncated Tat protein. A fourth class of short spliced RNAs, using 3' splice sites (3'ss) near the HIV-1 genome's 3' end, has been identified in two isolates *in vitro*^{3,4} and in a minority of viruses *in vivo*⁵; most of these transcripts are predicted to code for a 34 amino acid peptide in the C-terminus of Nef,⁵ but their function still remains to be defined. The complexity of HIV-1 splicing is further increased by two additional factors: (1) the usage of redundant 3'ss for generation of *rev* RNAs, of which seven have been reported, A4a, A4b, and A4c in subtype B viruses,^{1,2} A4f, A4g, and A4c in subtype C viruses,⁶ A4d in the subtype B isolate SF2 and in the group O virus ANT70C, and A4e in ANT70C⁷; and (2) the optional incorporation of small noncoding exons in the leader sequence: exons 2 or 3 or both in *tat*, *rev*, *nef*, and *env* RNAs and exon 2 in *vpr* RNAs.

Most studies on HIV-1 splicing have been done using T cell line-adapted viruses or cloned subgenomic fragments derived from them. Here we analyze splice site usage in three primary HIV-1 subtype B isolates, P2149-3, X2102, and X2210,^{8,9} in an *in vitro* acute infection assay of peripheral blood mononuclear cells (PBMCs).⁶ PBMCs prestimulated with phytohemagglutinin and interleukin-2 were exposed to virus at a multiplicity of infection of 0.1 infectious particles per cell for 2 h. Zidovudine (AZT) at 10 μ M was added at 8 h postinfection to avoid subsequent rounds of infection. Cells were collected on days 1 through 4 and total RNA was extracted. HIV-1 splicing patterns were analyzed through reverse transcriptase polymerase chain reaction (RT-PCR) followed by nested PCR, using primers recognizing sequences in the outermost exons common to either all DS or all SS HIV-1 RNAs. Reagents and PCR conditions were as described,⁶ with the forward primer in the nested PCR 5'-labeled with VIC fluorophore, which allowed analysis through electrophoresis in an automated sequencer using the GeneMapper software program (Applied Biosystems, Carlsbad, CA), which can accurately determine sizes of PCR products by running size standards labeled with a fluorophore emitting light at a different wavelength in the same capillary and quantify them by measuring peak areas.

The GeneMapper analyses revealed that in two isolates peak sizes were consistent with common splice site usage

TABLE 1. DOUBLY SPLICED TRANSCRIPTS OF P2149-3 IDENTIFIED BY CLONE SEQUENCING

Gene	Exons ^a	n	% DS ^b	% Gene ^c
<i>nef</i>	1.5.7	13	32.5	46.4
	1.2.5.7	1	2.5	3.6
	1.3.5.7	11	27.5	39.3
	1.5a.7	1	2.5	3.6
	1.3.5.7a	1	2.5	3.6
Total <i>nef</i>	1.3.7	1	2.5	3.6
		28	70	
	1.4d.7	4	10	44.4
	1.3.4d.7	1	2.5	11.1
Total <i>rev</i>	1.4h.7	3	7.5	33.3
	1.4a.7	1	2.5	11.1
Total <i>rev</i>		9	22.5	
<i>tat</i>	1.4.7	2	5	100
<i>vpr</i>	1.3A.7	1	2.5	100
Total		40		

^aExons are named as in Schwartz *et al.*,¹ Purcell *et al.*,² Bilodeau *et al.*,⁷ Carrera *et al.*,⁵ and in the current study (see the main text).

^bPercent of all doubly spliced (DS) transcript clones.

^cPercent of DS transcript clones coding for the same gene.

denotes the branch point), and a polypyrimidine tract (PPT) downstream of the BPS. P2149-3 has the AG dinucleotide at the intron end adjacent to A4d and an upstream PPT with nine pyrimidines interspersed with two purines (Fig. 4), which coincides with the PPT used by the NL4-3 isolate for splicing at A4a and A4b. Two BPS have been identified in NL4-3 for splicing at these two sites,¹¹ one of which is also used by the SF2 isolate.⁷ The sequences at these two BPS in P2149-3 are identical to those of NL4-3, and therefore they could also potentially be used for splicing at A4d, which is located 22 and 16 nt, respectively, downstream of these BPS.

The relatively infrequent usage of A4a and A4b in P2149-3 may be explained by the linear scanning mechanism model for 3'ss recognition,¹² by which the nt after the first AG downstream of the BPS (which in P2149-3 is adjacent to A4d) is preferentially selected as 3'ss. With regard to A4h, P2149-3 has an AG at the intron end adjacent to this splice site, and upstream of it there is a pyrimidine-rich segment with seven

TABLE 2. RELATIVE USAGE OF 3'ss FOR *rev* RNA GENERATION IN P2149-3

3'ss	% <i>rev</i> RNAs ^a				
	Day 1	Day 2	Day 3	Day 4	Mean
A4b	8.3	5.1	6.3	4.8	6.1
A4a	20.7	20.1	16.4	17.4	18.6
A4d	57.0	58.5	65.0	68.6	62.3
A4h	14.0	16.3	12.3	9.2	13.0
Total	100	100	100	100	100

^aNumbers represent the percent of the sum of the peak areas of PCR products derived from *rev* RNAs using each 3'ss, relative to the total sum of the peak areas from all *rev* RNAs.

pyrimidines and three interspersed purines. Just upstream of this segment, there is a sequence (UACAAAU) with 6 nt coincident with the consensus mammalian BPS sequence and five potential base pairings with U2 snRNP (underlined), whose complementarity to the BPS correlates positively with splicing efficiency.¹³ The relatively infrequent usage of A4h in P2149-3 may derive from the suboptimal PPT sequence, which is interrupted by three interspersed purines and contains not more than two consecutive pyrimidines. Although the strongest PPTs contain long stretches of consecutive pyrimidines, stretches of alternating purines and pyrimidines can function as PPT to promote usage of a downstream 3'ss if the tract is close to the splice site.¹⁴ In the case of A4h, its usage may also be promoted by a downstream exon splice enhancer (GAR ESE).¹⁵

Most subtype B isolates previously used for studies on HIV-1 splicing (HXB2-LAI-III-BRU, NL4-3, SF2, Ba_L, BH10) have an AG 1 nt downstream of the AG of A4h, and the sequences potentially used as PPT for A4h are identical to those of P2149-3 (Fig. 4). The obvious question is why the AG in the mentioned viruses is not used as 3'ss but A4h is used in P2149-3. One possible reason may be that in P2149-3 the AG at the A4c position is mutated to AC, which would lead to activation of A4h or a compensatory increase in its usage, in accordance with cryptic 3'ss activation or compensatory increases in 3'ss usage occurring when a nearby competing 3'ss is inactivated through mutation.¹¹ However, there must be additional factors, since the SF2 isolate lacks the intronic AG adjacent to A4c, which is mutated to CG, and does not use upstream AGs for *rev* RNA splicing.⁷ One may be that the sequence at A4h of P2149-3 is AAG/G, which is closer to the mammalian 3'ss consensus (YAG/G) than the sequences 1 nt downstream in SF2 and other viruses used in previous studies (AAG/U), and therefore it may represent a better candidate for selection as 3'ss by the splicing machinery. In addition, in SF2 the sequence at the putative branch point of A4h (UAACAAU) differs from that in P2149-3, having only four potential base pairings with U2 snRNP (underlined), which would make it a weaker branch point than the sequence in P2149-3, which has five potential base pairings with U2 snRNP.

In summary, we report for the first time in an HIV-1 primary isolate (P2149-3) the preferential usage for *rev* RNA generation of 3'ss A4d, previously detected only in a cloned fragment of a T cell line-adapted isolate and in a group O virus.⁷ P2149-3 also used a previously unreported 3'ss, A4h, for *rev* RNA generation, and three unusual *nef* transcripts were also identified in this isolate. These results point to a

FIG. 3. Sequences surrounding new and unusual splice junctions identified in P2149-3 DS RNAs. The arrow pointing downward signals the splice junction. 5' and 3' splice sites involved in splicing, named as in previous studies^{1,2,5,7} and in this study (see the main text), are indicated, with nucleotide positions in the HXB2 genome in parentheses. Nearby splice sites are also indicated. Splice junctions shown are, from top to bottom, D1/A4d, D3/A4d, D1/A4h, D1/A5a, and D4/A7a.

FIG. 4. Intronic and exonic sequences surrounding *rev* RNA splice sites in P2149-3. Sequences are aligned with those of NL4-3 and SF2 isolates. AG dinucleotides in the intron ends adjacent to 3'ss are in bold type. Polypyrimidine tracts potentially used for splicing at A4d and A4h in P2149-3 are boxed. In the NL4-3 and SF2 sequences, branch sites previously identified for *rev* RNA splicing^{7,11} are within ellipses. Nucleotides in P2149-3 potentially used as branch points for splicing at A4d and A4h (see the main text) are indicated with arrows and those potentially base pairing with U2 snRNP at these sites are underlined.

greater diversity in splice site usage by HIV-1 RNAs than previously appreciated. The high multiplicity of 3'ss used for *rev* RNA generation, currently numbering eight with A4h, may derive from the fact that Tat, in whose coding sequence *rev* 3'ss are located, is one of the most variable HIV-1 proteins,¹⁶ and from the absolute dependence of HIV-1 replication on Rev expression, in whose absence viral replication cannot be rescued by proteins secreted by nearby HIV-1-infected cells, as occurs with Tat.¹⁷

Acknowledgments

We thank the personnel at the Genomic Unit of Centro Nacional de Microbiología, Instituto de Salud Carlos III, for technical assistance in sequencing and GeneMapper analyses. This work was funded by Ministerio de Economía y Competitividad (Spain), Plan Nacional de I+D+I, through grants SAF2007-61688 and SAF2010-2096.

Sequences of PCR clones derived from P2149-3 DS transcripts have been deposited in GenBank under accession numbers JF808039–JF808078.

Author Disclosure Statement

No competing financial interests exist.

References

- Schwartz S, Felber BK, Benko DM, Fenyo EM, and Pavlakis GN: Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J Virol* 1990;64:2519–2529.
- Purcell DF and Martin M: Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* 1993;67:6365–6378.
- Smith J, Azad J, and Deacon N: Identification of two novel human immunodeficiency splice acceptor sites in infected T cell lines. *J Gen Virol* 1992;73:1825–1828.
- Ocwieja KE, Sherrill-Mix S, Mukherjee R, et al.: Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res* 2012;40:10345–10355.
- Carrera C, Pinilla M, Pérez-Álvarez L, and Thomson MM: Identification of unusual and novel HIV type 1 spliced transcripts generated in vivo. *AIDS Res Hum Retroviruses* 2010;26:815–820.
- Delgado E, Carrera C, Nebreda P, et al.: Identification of new splice sites used for generation of *rev* transcripts in human immunodeficiency virus type 1 subtype C primary isolates. *PLoS One* 2012;7:e30574.
- Bilodeau PS, Domsic JK, and Stoltzfus CM: Splicing regulatory elements within tat exon 2 of human immunodeficiency virus type 1 (HIV-1) are characteristic of group M but not group O HIV-1 strains. *J Virol* 1999;73:9764–9772.
- Cuevas MT, Fernández-García A, Pinilla M, et al.: Biological and genetic characterization of HIV type 1 subtype B and nonsubtype B transmitted viruses: Usefulness for vaccine candidate assessment. *AIDS Res Hum Retroviruses* 2009;26:1019–1025.
- Fernández-García A, Cuevas MT, Muñoz-Nieto M, et al.: Development of a panel of well-characterized human immunodeficiency virus type 1 isolates from newly diagnosed patients including acute and recent infections. *AIDS Res Hum Retroviruses* 2010;25:93–102.
- Furtado MR, Balachandran R, Gupta P, and Wolinski SM: Analysis of alternatively spliced human immunodeficiency virus type-1 mRNA species, one of which encodes a novel Tat-Env fusion protein. *Virology* 1991;185:258–270.
- Swanson AK and Stoltzfus CM: Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. *J Biol Chem* 1998;273:34551–34557.
- Smith CW, Chu TT, and Nadal-Ginard B: Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol Cell Biol* 1993;13:4939–4952.
- Wu J and Manley JL: Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev* 1989;3:1553–1561.
- Coolidge CJ, Seely RJ, and Patton JG: Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* 1997;25:888–896.
- Caputi M, Freund M, Kammiller S, Asang C, and Schaal H: A bidirectional SF2/ASF- and SRp40-dependent splicing enhancer regulates HIV-1 *rev*, *env*, *vpu*, and *nef* gene expression. *J Virol* 2004;78:6517–6526.
- Korber B, Gaschen B, Yusim K, et al.: Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* 2001;58:19–42.
- Verhoef K, Klein A, and Berkhout B: Paracrine activation of the HIV-1 LTR promoter by the viral Tat protein is mechanistically similar to trans-activation within a cell. *Virology* 1996;225:316–327.

Address correspondence to:
Michael M. Thomson
Instituto de Salud Carlos III
Centro Nacional de Microbiología
Ctra. Majadahonda-Pozuelo, Km. 2
28220 Majadahonda, Madrid
Spain

E-mail: mthomson@isciii.es