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, Michael M Thomson.

AIDS Res Hum Retroviruses. 2013 Jul;29(7):1075-8.

TITLE

Identification of new and unusual *rev* and *nef* transcripts expressed by a HIV-1 primary isolate

RUNNING TITLE: Unusual splicing in a HIV-1 primary isolate.

AUTHORS

Yolanda Vega[1], Elena Delgado[1], Cristina Carrera[1], Paloma Nebreda[1], Aurora Fernández-García[1], María Teresa Cuevas[1], Lucía Pérez-Álvarez[1], Michael M Thomson[1]*.

[1]Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain.

CORRESPONDING AUTHOR

Michael M. Thomson, MD, PhD

Instituto de Salud Carlos III

Centro Nacional de Microbiología

Ctra. Majadahonda-Pozuelo, Km. 2

28220 Majadahonda (Madrid)

Spain

Tel: 34-91-8223900

Fax: 34-91-5097014

e-mail: mthomson@isciii.es

ABSTRACT

We analyzed RNA splice site usage in three HIV-1 subtype B primary isolates through RT-PCR amplification of spliced RNAs using a fluorescently labeled primer, with computerized size determination and quantitation 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 5 nucleotides upstream of A4a, previously only identified 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 a 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 ANT70C[7]; 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. AZT at 10 μM 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 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 GeneMapper software program (Applied Biosystems, Carlsbad, CA), which can accurately determine sizes of PCR products by running size standards labeled with a fluorophore emiting 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 reported for subtype B viruses (results not shown). However, in one isolate, P2149-3, several peaks with unexpected sizes were detected (Fig. 2). Four of the peaks derived from DS RNAs corresponded to PCR products with calculated sizes 303, 352, 377, and 426 nucleotides (nt), which are 4 or 5 nt longer than expected for previously described *rev* RNAs using 3'ss A4a 1.4a.7, 1.2.4a.7, 1.3.4a.7, and 1.2.3.4a.7, respectively, which were also detected, but with weaker signals than those of the unexpected products. Additional peaks, with sizes 281, 336, 386, and 410 nt, which do not correspond to known HIV-1 transcripts, were also detected (Fig. 2). To identify the transcripts corresponding to the unexpected peaks, PCR products of DS RNAs from day 2 postinfection were cloned and sequenced. Of the 40 sequenced clones (Table 1), 28 were predicted to code for Nef, nine for Rev, two for Tat, and one for Vpr. Among the nine *rev* RNA-derived clones, eight contained unusual or unreported splice junctions (Fig. 3, Table 1). Five of these used a 3'ss located 5 nt upstream of A4a. This 3'ss has been reported to be used in a ~0.3 kb cloned fragment of the T cell-line adapted subtype B isolate SF2 and in the group O virus ANT70, and was

designated A4d[7]. The exon composition of the P2149-3 clones using A4d was 1.4d.7 in four and 1.3.4d.7 in one. Three other *rev* clones of P2149-3 derive from a transcript using a previously unreported 3'ss located 20 nt upstream of A4c, which was designated A4h, with exon composition 1.4h.7 (Fig. 3). Among the 28 *nef* clones, two used unusual 3'ss. One spliced from 5' splice site D1 to 3'ss A5a, located 4 nt downstream of A5 (the usual 3'ss of *nef* RNAs). S5a has been reported to be used occasionally by HTLV-IIIB[10] and p89.6[4] isolates *in vitro* and by another virus *in vivo*[5]. Another *nef* RNA used 3'ss A7a, located 28 nt upstream of A7, previously identified only as a minor 3'ss in HXB2 isolate[1]. A third *nef* transcript had the exon composition 1.3.7, which involves splicing from D1 to 3'ss A7, previously only detected in a very small minority of transcripts in p89.6 isolate through next generation sequencing[4].

The detection of A4d and A4h splice site usage in P2149-3 by sequencing allowed to assign seven GeneMapper peaks with unexpected sizes to *rev* RNAs: 1.4d.7 (303 nt), 1.2.4d.7 (352 nt), 1.3.4d.7 (377 nt), 1.2.3.4d.7 (426 nt), 1.4h.7 (336 nt), 1.2.4h.7 (386 nt), and 1.3.4h.7 (410 nt). The remaining peak, of 281 nt, corresponds to *nef* 1.3.7 transcript. The analysis of PCR products derived from SS RNAs of P2149-3 also revealed peaks with sizes predicted for *env* RNAs using A4d and A4h (results not shown).

The relative usage of the different 3'ss by *rev* RNAs in P2149-3 was quantified by measuring the areas under the fluorescent peaks (Table 2). A4d was used preferentially at all time points (mean 62.3% in the four days of the assay), followed by A4a (18.6%), A4h (13%), and A4b (6.1%).

We analyzed sequence features in the genome of P2149-3 that could explaing its unusual usage of splice sites for *rev* RNA generation. The usual elements of the metazoan 3'ss include an AG at the 3' end of the intron, a branch point site (BPS),

usually 18-40 nt upstream of the AG, with the loosely conserved mammalian consensus sequence YNYURAY (the underline 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 9 pyrymidines interspersed with 2 purines (Fig. 4), which coincides with the PPT used by 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 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 pyrymidine-rich segment with 7 pyrymidines and 3 interspersed purines. Just upstream of this segment, there is a sequence (UACAAAU) whith 6 nt coincident with the consensus mammalian BPS sequence and 5 potential base-pairings with U2 snRNP (underlined), whose complementarity to the BPS correlates positively to 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 consecutives pyrymidines. Although the strongest PPTs contain long stretches of consecutive pyrymidines, stretches of alternating purines and pyrymidines 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 previosuly used for studies on HIV-1 splicing

(HXB2-LAI-IIIB-BRU, NL4-3, SF2, Ba_L, BH10) have an AG one nt downstream of the AG of A4h, and the sequences potentially used as PPT for A4h are identical to 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 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 one 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 4 potential base-pairings with U2 snRNP (underlined), which would make it a weaker branch point than the sequence in P2149-3, which has 5 potential base-pairings with U2 snRNP.

In summary, we report for the first time in a 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 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 accessions JF808039-JF808078.

REFERENCES

1. 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: 2519-29.

2. 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: 6365-78.

3. 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: 1825-8.

4. 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; 40:10345-55.

5. 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.

6. 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.

7. 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: 9764-72.

8. 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: 1019-25.

9. 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.

10. 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: 258-70.

11. Swanson AK and Stoltzfus CM. Overlapping cis sites used for splicing of HIV-1 env/nef and rev mRNAs. J Biol Chem 1998; 273: 34551-7.

12. 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: 4939-52.

13. Wu J and Manley JL. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev 1989; 3: 1553-61.

14. Coolidge CJ, Seely RJ, and Patton JG. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic Acids Res 1997; 25: 888-896.

15. Caputi M, Freund M, Kammler 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: 6517-26.

16. Korber B, Gaschen B, Yusim K, *et al*. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull 2001; 58: 19-42.

17. 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: 316-27.

FIGURE LEGENDS

Fig. 1. Schematic depiction of RNA splicing in HIV-1 subtype B. Open reading frames are shown as open boxes and exons as black bars. Only exons generated by splicing at commonly used splice sites are shown. Exons are named as previously[1,2]. All spliced transcripts incorporate exon 1. Optionally, exons 2 or 3 or both can be incorporated into *tat*, *rev*, *nef*, or *env* transcripts, and exon 2 into *vpr* transcripts. Proteins encoded in spliced RNAs are indicated on the right of the 3' exon.

Fig. 2. GeneMapper analysis of DS RNAs expressed by P2149-3. The analysis corresponds to day 2 after acute infection of PBMCs. Green peaks represent PCR products and orange peaks represent size standards. Size of PCR product, encoded gene, and exon composition (named as in previous studies[1,2,5]) predicted according to PCR product size are shown on top or on the side of each peak. Peaks with unexpected sizes, according to usual splice site usage by subtype B viruses, are marked with interrogation signs.

Fig. 3. Sequences surrounding new and unusual splice junctions identified in P2149-3 DS RNAs. The arrow pointing downwards 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 main text), are signaled, 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 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 main text) are indicated with arrows and those potentially base-pairing with U2 snRNP at these sites are underlined.

Fig. 1

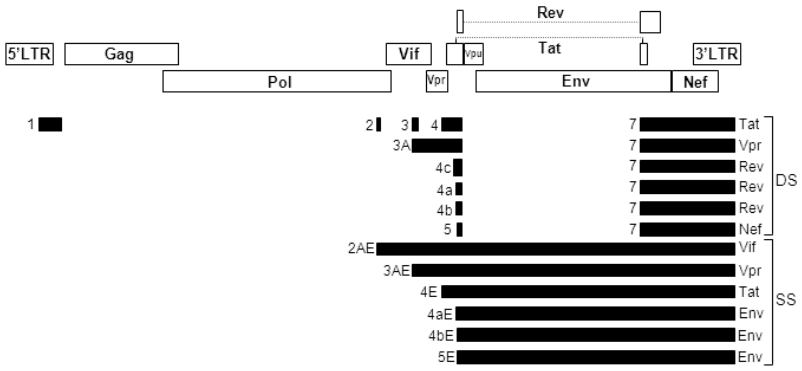

Rev
5'LTR
Gag
Vif
Vpu
Tat
3'LTR
Pol
Vpr
Env
Nef
1
2
3
4
3A
4c
4a
4b
5
7
Tat
Vpr
Rev
Rev
Rev
Nef
DS
2AE
3AE
4E
4aE
4bE
5E
Vif
Vpr
Tat
Env
Env
Env
SS

Fig. 2

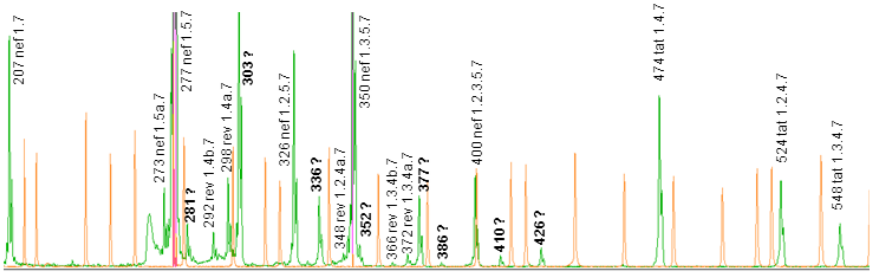

207 nef 1.7
273 nef 1.5a.7
277 nef 1.5.7
281 ?
292 rev 1.4b.7
298 rev 1.4a.7
303 ?
326 nef 1.2.5.7
336 ?
348 rev 1.2.4a.7
350 nef 1.3.5.7
352 ?
366 rev 1.3.4b.7
372 rev 1.3.4a.7
377 ?
386 ?
400 nef 1.2.3.5.7
410 ?
426 ?
474 tat 1.4.7
524 tat 1.2.4.7
548 tat 1.3.4.7

Fig. 3

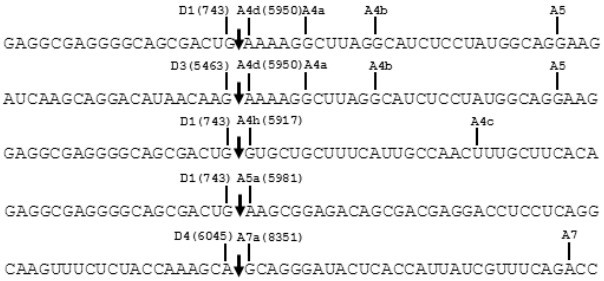
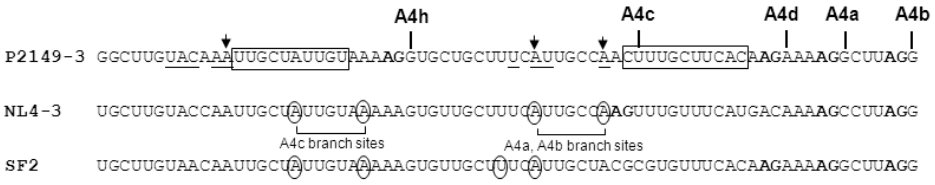

D1(743) A4d(5950) A4a A4b A5
GAGGCGAGGGGCAGCGACUG▼AAAAGGCUUAGGCAUCUCCUAUGGCAGGAAG
D3(5463) A4d(5950) A4a A4b A5
AUCAAGCAGGACAUAACAAG▼AAAAGGCUUAGGCAUCUCCUAUGGCAGGAAG
D1(743) A4h(5917) A4c
GAGGCGAGGGGCAGCGACUG▼GUGCUGCUUUCAUUGCCAACUUUGCUUCACA
D1(743) A5a(5981)
GAGGCGAGGGGCAGCGACUG▼AAGCGGAGACAGCGACGAGGACCUCCUCAGG
D4(6045) A7a(8351) A7
CAAGUUUCUCUACCAAAGCA▼GCAGGGAUACUCACCAUUAUCGUUUCAGACC

Fig. 4


A4h
A4c
A4d
A4a
A4b
P2149-3 GGCUUGUACAAAUUGCUAUUGUAAAAGGUGCUGCUUUCAUUGCCAACUUUGCUUCACAAGAAAAGGCUUAGG
NL4-3 UGCUUGUACCAAUUGCUAUUGUAAAAAGUGUUGCUUUCAUUGCCAAGUUUGUUUCAUGACAAAAGCCUUAGG
A4c branch sites
A4a, A4b branch sites
SF2 UGCUUGUAACAAUUGCUAUUGUAAAAGUGUUGCUUUCAUUGCUACGCGUGUUUCACAAGAAAAGGCUUAGG