

Detection of antisense protein (ASP) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1)

Antonio Mancarella,¹ Francesco A. Procopio,¹ Tilmann Achsel,² Elisa De Crignis,^{3†} Brian T. Foley,⁴ Giampietro Corradin,⁵ Claudia Bagni,² Giuseppe Pantaleo¹ and Cecilia Graziosi^{1,*}

Abstract

The detection of antisense RNA is hampered by reverse transcription (RT) non-specific priming, due to the ability of RNA secondary structures to prime RT in the absence of specific primers. The detection of antisense RNA by conventional RT-PCR does not allow assessment of the polarity of the initial RNA template, causing the amplification of non-specific cDNAs. In this study we have developed a modified protocol for the detection of human immunodeficiency virus type 1 (HIV-1) antisense protein (ASP) RNA. Using this approach, we have identified ASP transcripts in CD4⁺ T cells isolated from five HIV-infected individuals, either untreated or under suppressive therapy. We show that ASP RNA can be detected in stimulated CD4⁺ T cells from both groups of patients, but not in unstimulated cells. We also show that in untreated patients, the patterns of expression of ASP and *env* are very similar, with the levels of ASP RNA being markedly lower than those of *env*. Treatment of cells from one viraemic patient with α -amanitin greatly reduces the rate of ASP RNA synthesis, suggesting that it is associated with RNA polymerase II, the central enzyme in the transcription of protein-coding genes. Our data represent the first nucleotide sequences obtained in patients for ASP, demonstrating that its transcription indeed occurs in those HIV-1 lineages in which the ASP open reading frame is present.

INTRODUCTION

The ASP gene is an antisense open reading frame (ORF) potentially encoding for the putative HIV-1 antisense protein (ASP) [1]. In retroviruses, antisense transcripts coding for proteins have been characterized for each member of the human T-cell leukaemia virus (HTLV) family (reviewed in [2]). In HTLV-1, antisense transcription results in the production of HTLV-1 basic zipper protein (hbz), which is involved in the regulation of adult T-cell leukaemia oncogenesis and whose highly conserved ORF is located between the *env* and *tax/rex* genes [3, 4]. An antisense protein called Aph-2 has also been found in HTLV-2 [5], and is seemingly implicated in the regulation of viral transcription and translation. Antisense proteins Aph-3 and Aph-4 have been also described in the regulation of HTLV-3 and HTLV-4 transcription [6–8].

In HIV-1, the ASP ORF is located in the envelope gene at the junction between gp120 and gp41, in the antisense reading frame –2, in a similar position to the hbz of HTLV-1 [1]. In HXB2, the reference sequence for subtype B, the ASP ORF is 570 bp long, spanning coordinates 7942–7373 [9]. A study on almost 23 000 HIV genomes showed the prevalence of the ASP ORF in the large majority of group M. In subtype B it has been recognized in 85 % of the genomes [10]. The putative ASP protein is 189 amino acids long and appears to have two potential transmembrane helices with a cytoplasmic N-terminus, indicating that it could be associated with cell membranes [1, 11, 12].

Although the ASP ORF was first discovered three decades ago, the existence of its gene product in HIV-1 infection is still controversial. Some authors support the hypothesis of ASP being part of a long non-coding (lnc) RNA with

Received 4 October 2018; Accepted 19 February 2019; Published 21 March 2019

Author affiliations: ¹Division of Immunology and Allergy, Lausanne University Hospital, Switzerland; ²Department of Fundamental Neuroscience, University of Lausanne, Switzerland; ³Department of Biochemistry, Erasmus Medical Center, Rotterdam, The Netherlands; ⁴Theoretical Biology and Biophysics Group, Los Alamos National Laboratories, Los Alamos, New Mexico, USA; ⁵Department of Biochemistry, University of Lausanne, Switzerland.

***Correspondence:** Cecilia Graziosi, cecilia.graziosi@chuv.ch

Keywords: antisense protein; ASP; HIV-1; antisense transcription, envelope.

Abbreviations: ART, antiretroviral therapy; ASP, antisense protein; ORF, open reading frame; PBMC, peripheral blood mononuclear cell; RT, reverse transcriptase.

†Present address: Clinical Trial Office, CRO Aviano National Cancer Institute, Aviano, Italy.

MH756691-MH756746.

Six supplementary figures and two supplementary tables are available with the online version of this article.

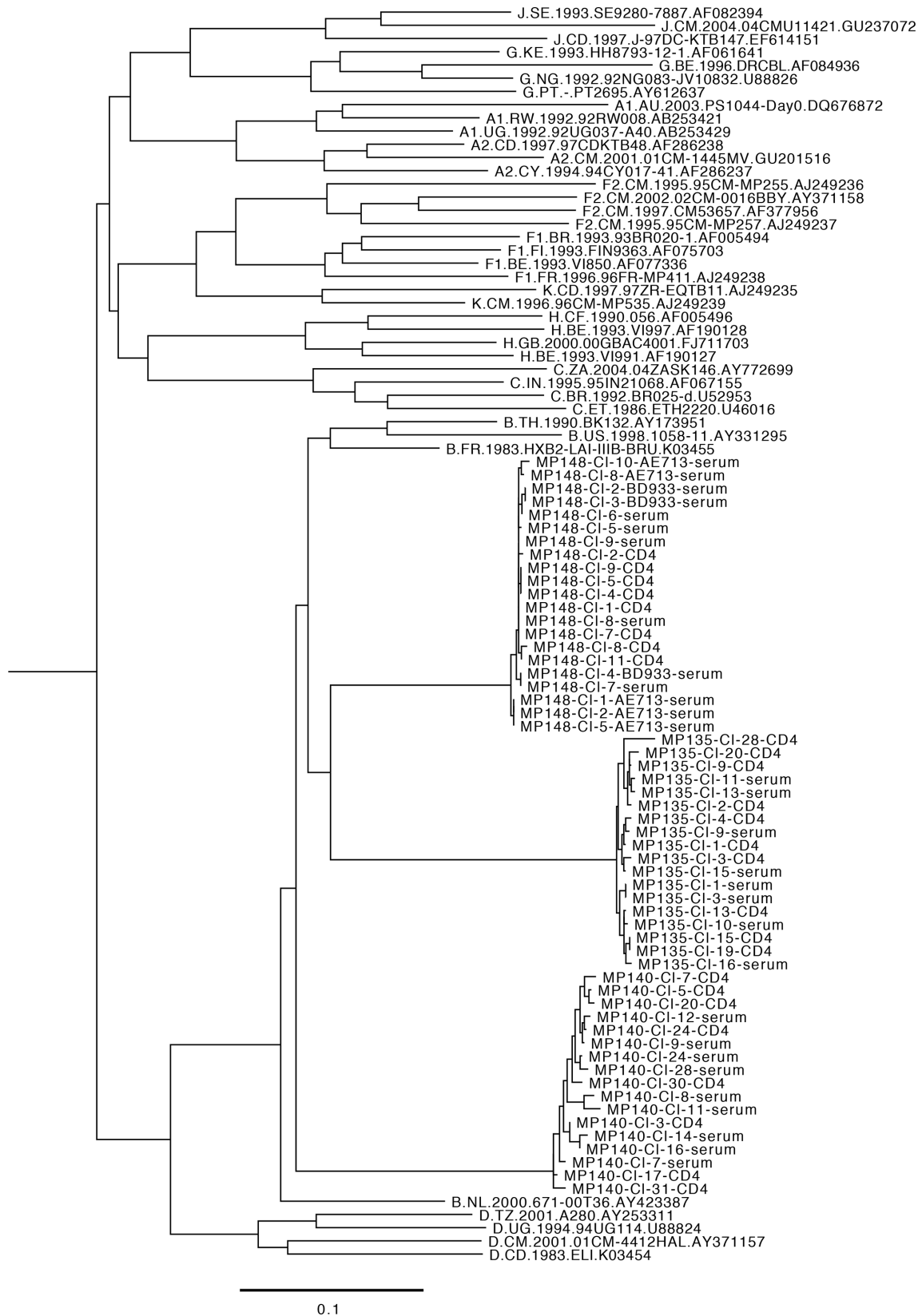


Fig. 1. Phylogenetic tree showing the alignment of the ASP and *env* sequences obtained in our patients with the reference sequences of the HIV-1 M group subtypes. The tree was inferred using the maximum-likelihood method with 1000 bootstrap replicates using MEGA 7.0 software. ASP sequences obtained in cells were converted to the *env* reverse and complementary orientation and reading frame,

prior to alignment. As shown in the tree, the sequences from patients cluster with HXB2-LAI-IIIB-BRU, the reference sequence for clade B, confirming that they are all of the B subtype.

regulatory functions [13, 14], whereas others argue that it might be a protein involved in ASP-induced autophagy to improve virion production [15]. Regardless of its real nature, bioinformatic data demonstrate that ASP, either as regulatory RNA or protein (or both?), is very likely to play a function in HIV pathogenesis, since the degree of conservation of its ORF in subtype B is higher than would be expected if it had no function at all [10]. In contrast, the lack of a good ASP ORF in many HIV-1 lineages points to this function not being crucial for survival, or at least not yet [10]. Although several studies indicate that the ASP gene does indeed give rise to several transcripts of different lengths in HIV *in vitro* [9, 16], no information is currently available about the existence of a translation product in HIV-infected individuals. The nucleotide sequence of ASP transcripts in infected individuals is still missing and in fact the bioinformatic data supporting the distribution of the ASP ORF in the various subtypes do not come from actual ASP RNA sequences, but have been extrapolated from envelope web alignments, codon-aligned in the ASP antisense reading frame (https://www.hiv.lanl.gov/content/sequence/HIV/SI_alignments/ASP.html) [10].

The main difficulty in detecting ASP RNA transcripts lies in their complementarity to *env* RNA sequences that are also present in infected cells. The most sensitive and specific approach for the detection of RNA transcripts is RT-PCR, with the use of the gene-specific antisense primer to prime the reaction of reverse transcription (RT) in order to synthesize the cDNA of the right polarity. In recent years a phenomenon called RT self-priming (or endogenous priming) has been observed, whereby non-specific priming of the reaction of RT occurs due to the ability of some RNA secondary structures, such as RNA hairpins or loops, to prime RT even in the absence of primers [17]. Since these structures are able to prime RT in both directions, it is not possible to assess the polarity of the initial RNA template from which the products of subsequent PCR amplifications are generated. Thus, common RT-PCR protocols, which are routinely used for sense RNA detection, are unable to

provide reliable information when applied to the study of antisense transcription.

In this study we have developed a novel approach to detecting ASP RNA based on a modification of the protocol described by Haist and coworkers [17], whereby the ASP-specific antisense primer is biotinylated and the resulting cDNA affinity-purified in order to eliminate non-specific products. Our goals were (i) to demonstrate that the ASP ORF is expressed during HIV infection and (ii) to isolate and sequence ASP RNA from patients. Our results show that ASP transcripts are easily detectable in stimulated CD4⁺ T cells isolated from untreated patients and, to a lesser extent, from patients undergoing antiretroviral therapy (ART). Our data also represent the first nucleotide sequences obtained in patients for ASP transcription products, demonstrating that ASP is indeed expressed in those HIV-1 lineages in which the ASP ORF is present.

METHODS

Study groups

Samples from serum and peripheral blood mononuclear cells (PBMCs) from three healthy donors and six HIV-infected individuals were used in this study. The subjects' features and clinical information are summarized in Table 1. Patient MP140 has been described elsewhere [18, 19]. Three patients had detectable viraemia and were not receiving ART at the time of sampling. Patients MP148 and MP140 were asymptomatic and naïve to therapy, whereas patient MP135 was an AIDS patient who had gone repeatedly on and off treatment and was untreated at the time of leukapheresis. The other three patients were undergoing ART and were aviraemic. These studies were approved by the Institutional Review Board of the Centre Hospitalier Universitaire Vaudois (CHUV) and subjects gave written informed consent.

Primary cells and sera

PBMCs from healthy donors were obtained by flebotomy. In HIV-positive individuals, PBMCs were obtained by

Table 1. Patients' features, clinical information and antiretroviral therapy status

Patient ID	Age	Sex	Stage of HIV infection	Clade	Viral load (copies ml ⁻¹)	CD4 Count (cells µ ⁻¹)	ART status
MP135	44	M	C3	B	1.6 × 10 ⁵	176	Untreated
MP140	23	M	A2	B	3.6 × 10 ⁴	427	Naïve
MP148	37	M	A1	B	2.0 × 10 ⁴	717	Naïve
MP069	42	M	A1	B	<20	1309	Treated
MP071	47	M	C3	B	<20	167	Treated
MP146	59	M	C3	B	<20	385	Treated

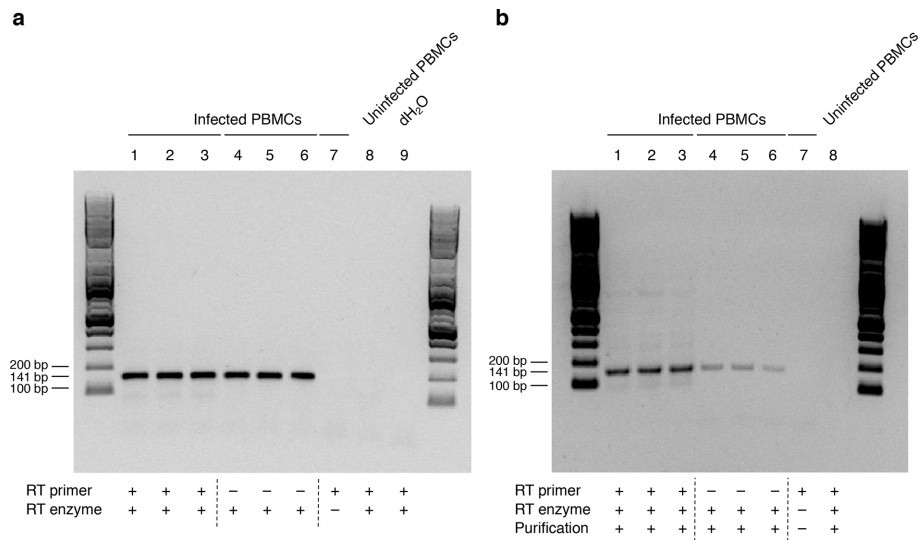


Fig. 2. Detection of ASP RNA by RT-PCR in PBMCs of one healthy donor infected *in vitro* with HXB2. (a) Detection of ASP RNA using standard RT-PCR. The ASP-specific band (141 bp) is clearly visible in the presence of the specific RT primer (lanes 1–3), although a band of the same intensity is visible in samples reverse-transcribed in the absence of primer (lanes 4–6). (b) Detection of ASP RNA using the modified Heist protocol. Reverse transcription of RNA from infected PBMCs in the presence of the biotinylated antisense primer followed by purification of the cDNA leads to a clear band of the right molecular weight (lanes 1–3). In contrast, a weaker band is still present in unpurified samples (lanes 4–6). No bands are visible in infected PBMCs reverse-transcribed in the absence of RT, uninfected PBMCs or dH₂O.

leukapheresis. Samples were stored in liquid nitrogen until use. Sera were obtained by flebotomy using S-Monovette (Sarstedt AG&Co., Nümbrecht, Germany) a few days (<1 week) before the leukapheresis. Serum was separated from blood by centrifugation for 10 min at 2000 r.p.m., aliquoted at 1 ml/tube and stored at -80°C without further manipulations. CD4⁺ T cells were isolated from PBMCs by negative selection using the EasySep Human CD4⁺ T Cell Enrichment kit (Stemcell Technologies), following the manufacturer's instructions. Unfractionated PBMCs and purified CD4⁺ T cells were cultured in complete RPMI/10% FCS/10 ng ml⁻¹ IL-2 at a concentration of 1×10^6 cells ml⁻¹ and plated in anti-CD3/CD28-coated (Becton Dickinson, San Diego, CA, USA) 48-well plates at 1 ml/well. Stimulation was carried out for 4 to 5 days, depending on the amount of cells available for each patient. Stimulated cells were pelleted (1×10^6 /tube) and the pellets were stored at -80°C until use. For the α -amanitin sensitivity assay, CD4⁺ T cells were isolated and stimulated with anti-CD3/CD28 as described above. At day 3 post-stimulation, the cells were treated with increasing concentrations of α -amanitin (5, 10, 20 and 100 $\mu\text{g ml}^{-1}$) (Sigma-Aldrich, St Louis, MI, USA) with time points post-treatment collected at 4 h, 8 h and 24 h (day 4 post-stimulation). The IC₅₀ for α -amanitin was calculated using GraphPad Prism version 8.

In vitro infections

HIV-negative PBMCs were resuspended at 1×10^6 cells ml⁻¹ in complete RPMI/10% FBS, plated in six-well plates (3 ml/well) and treated with 50 U ml⁻¹ of IL-2 and 3 $\mu\text{g ml}^{-1}$ of

PHA. After 3 days, the cells were washed, resuspended in complete RPMI and incubated with HIV-1_{HXB2} (37.6 ng/10⁶ cells) in the presence of IL-2 and polybrene. The infection was carried out for 4 days and monitored daily by flow cytometric detection of intracellular p24.

Development of patient-specific RT-PCR primers

Given the high degree of variability of the envelope region in which ASP is comprised, patient-specific primers had to be developed in two steps, a HXB2-specific step and a patient-specific step. In the HXB2-specific step, a primer pair called PanASP was identified on the sequence of HXB2, amplifying a fragment of 1094 bp, substantially exceeding the ASP ORF both upstream and downstream (PanASPF – ACCAAGCCTCCTACTATCATTATG; PanASPR – GCA-CATTGTAACATTAGTAGAGCA). PanASP primers were used to amplify proviral DNA from each patient. In the patient-specific step, the proviral DNA PCR product was sequenced (direct sequencing of amplification products) and the sequence obtained was used to design primers internal to the PanASP fragment, which were specific for each individual patient. Several oligoes, both sense and antisense, were identified within the PanASP fragment, at various distances upstream and downstream from the ASP ORF. The oligoes were tested in the patient's proviral DNA in various forward/reverse combinations and under various stringency conditions. The pairs resulting in the best amplifications (the longest fragment with the cleanest and most intense signal) were selected for ASP detection.

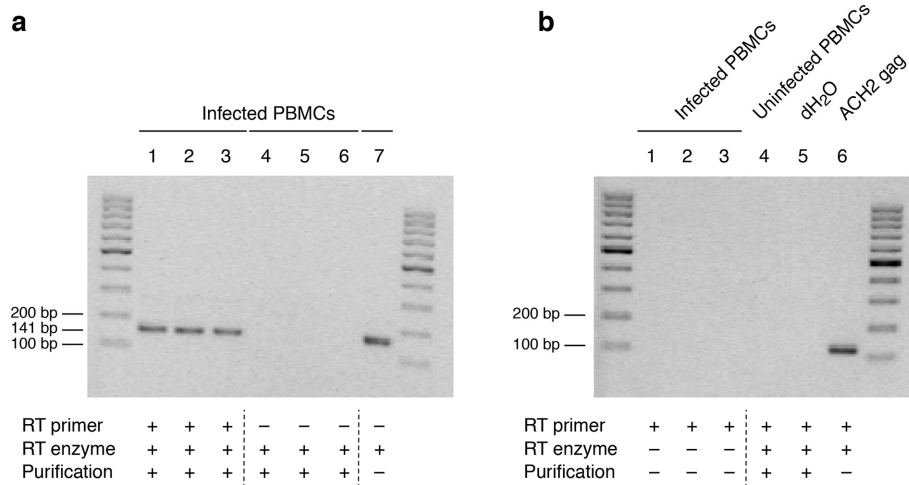


Fig. 3. Detection of ASP RNA in PMBCs of one healthy donor infected *in vitro* with HXB2 by RT-PCR following RNA linearization. (a) The ASP band is clearly visible in linearized RNA reverse-transcribed in the presence of the biotinylated specific primer followed by purification of the cDNA (lanes 1–3). In contrast, no bands can be seen in purified cDNA from primer minus RT-PCR reactions (lanes 4–6), although a band is clearly visible in the primer minus reaction in the absence of purification. (b) No bands can be detected in RNA from infected PMBCs in RT minus controls, uninfected PMBCs or water. The positive control is ACH-2 gag RT-PCR.

Additional ASP oligoes amplifying shorter fragments (<200 bp) were also designed based on the HXB2 sequence, to be used for screening purposes (ASP141F: TGCACCAC TCTTCTCTTTGC; ASP141R: TAACAACAATGGG TCCGAGA; ASP171F: CCCTCATATCTCCTCCTCCA; ASP171R: TAAAACAAATTATAAACATGTGGC). All primers were obtained from Integrated DNA Technologies (San Jose, CA, USA).

Reverse transcription and PCR reactions

Total RNA was extracted using the RNeasy Mini kit (Qiagen, Hilden, Germany) and treated with the TURBO DNA-free kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). RT reactions were carried out in a total volume of 50 µl using SuperScript III Reverse Transcriptase (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Just prior to RT, the RNA was fully denatured for 5 min at 94 °C and quickly cooled down in iced water. For each reaction, 1–3 µg of RNA template, depending on the sample, was used along with 2 pmol of a biotinylated version of the ASP-specific antisense primer and 10 mM of dNTPs. Reactions were carried out for 1 h at 55 °C prior to purification of the ASP biotinylated cDNA by streptavidin-coated magnetic beads (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA).

Standard PCR reactions were performed in a total volume of 50 µl, using the Platinum *Taq* DNA Polymerase High Fidelity kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). PCR reactions were carried out over 35 cycles of amplification. For analysis of genomic *env*, HIV genomic RNA was extracted from 1 ml serum using the QiAmp Viral RNA Mini kit (Qiagen, Hilden, Germany). The reverse

transcriptase conditions, PCR and cloning manipulations were the same as those used for ASP, with the exception of the RT primer, which was antisense to *env* and not biotinylated.

Cloning and sequencing

Amplified products were cloned into pCR 2.1 vector using the TA cloning kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Clones bearing inserts were identified by colony PCR and sequenced using the BigDye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) on an Applied Biosystems Automated Sequencer.

Nested real-time RT-PCR

Patient-specific qPCR primers and FAM and TAMRA probes were synthesized by IDT-Integrated DNA Technologies (Coralville, IA, USA) and designed using the IDT PrimerQuest Tool available on the company's website (<https://eu.idtdna.com/Primerquest/Home/Index>). Total RNA was extracted from purified CD4⁺ T cells, reverse-transcribed with biotinylated primers and affinity-purified as described above. Pre-amplification reactions of samples' cDNA and standard curve plasmid DNA were carried out in a total volume of 25 µl containing 1 µl of template, 0.1 µl Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), 0.2 µM primers, 0.2 mM dNTPs and 1.5 µM to 2.5 µM Mg²⁺ (depending on the primers used). The reactions were carried out as follows: initial denaturation for 2 min at 95 °C, amplification through 18 cycles (30 s at 95 °C, 30 s at 50 °C, 40 s at 68 °C) and a 7 min extension at 68 °C. qPCR reactions were carried out in a total volume of 20 µl containing 1 µl of pre-

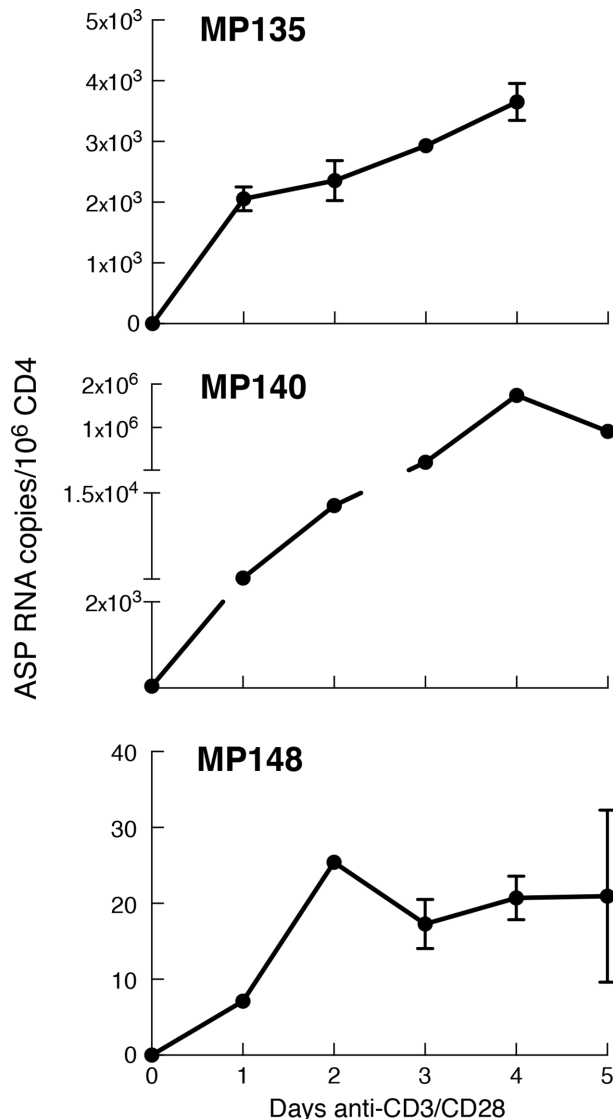


Fig. 4. Quantitative analysis of ASP RNA expression in CD4⁺ T cells isolated from three untreated, viraemic patients after stimulation with anti-CD3/28. In patients MP135 and MP140, ASP RNA expression peaks at day 4 post-stimulation, while in patient MP148, a peak in the expression of ASP RNA can already be observed at day 2. The results are expressed as ASP RNA copies/million CD4⁺ T cells. Each point in the time course represents the mean value of triplicate PCR reactions.

amplification PCR products diluted 1:5 in dH₂O, 0.9 μ M of each qPCR primer and 0.25 μ M of qPCR probe, using the TaqMan Gene Expression Master Mix (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA). The reaction conditions were 10 min at 95 °C followed by 40 cycle of amplification (15 s at 95 °C and 1 min at 60 °C). For the standard curve, pCR2.1 clones carrying patient-specific ASP amplicons were used (see 'Cloning and sequencing' above). Each plasmid was utilized at 3×10^0 – 3×10^6 copies to generate the PCR standard curve and the number of copies per

1 μ l of cDNA was normalized to 1×10^6 total CD4 cells. Data were acquired using the Applied Biosystems StepOne Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) and analysed with the provided software.

Nested duplex RT-PCR for quantification of α -amanitin inhibition

Total RNA was extracted as described above, and ASP RNA levels were normalized to U6 snRNA (small nuclear RNA). Duplex RT reactions were performed using 1 μ g RNA template and biotinylated reverse primers specific to ASP (Pan-ASP R, see above) and U6 snRNA [20]. Following RT, cDNAs were affinity-purified and resuspended in 10 μ l of dH₂O. Pre-amplification of both ASP and U6 targets was carried out in separate wells, adding 1 μ l of duplex RT reaction to a PCR mix containing 0.1 μ l Platinum Taq DNA Polymerase High Fidelity (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA), 0.2 μ M of each gene-specific primer, 0.2 mM each dNTPs, 2.5 μ l 10 \times Taq buffer and dH₂O up to 25 μ l for each sample. Pre-amplification was performed using the following thermocycling conditions: 2 min at 95 °C, 30 s at 95 °C, 30 s at 55 °C and 30 s at 68 °C for a total of 14 cycles, followed by 7 min final extension at 68 °C. At the end of pre-amplification, 100 μ l of dH₂O was added to each well (1:5 dilution). The qPCR was performed using TaqMan Gene Expression Master Mix (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) and primers/probes specific to ASP and U6. Probes were labelled with FAM and TAMRA dyes at 5' and 3', respectively. All primers and probes were obtained from IDT Integrated DNA Technologies (Coralville, IA, USA). Each qPCR reaction was carried out in triplicate in a total volume of 20 μ l containing 1 μ l of the diluted pre-amplification product, 0.9 μ M of each qPCR primer and 0.25 μ M probe. Reactions were carried out as follows: 10 min at 95 °C, followed by 15 s at 95 °C and 1 min at 60 °C for a total of 40 cycles. Data were acquired using the StepOnePlus Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) and analysed with the provided software. The $\Delta\Delta C_T$ method was then used to process these data to calculate relative gene expression for the ASP RNA.

Nucleotide and aminoacid sequence analyses

Nucleotide sequence editing and alignments were performed with Sequencer, BioEdit and CLUSTAL X software.

Phylogenetic analyses

Phylogenetic trees were inferred by the maximum-likelihood method with 1000 bootstrap replicates using MEGA 7.0 software.

RESULTS

In this study we have developed a new approach to the detection of antisense RNA, based on a modification of the protocol proposed by Haist *et al.* [17]. Using this approach we were able to identify ASP RNA in CD4⁺ T lymphocytes

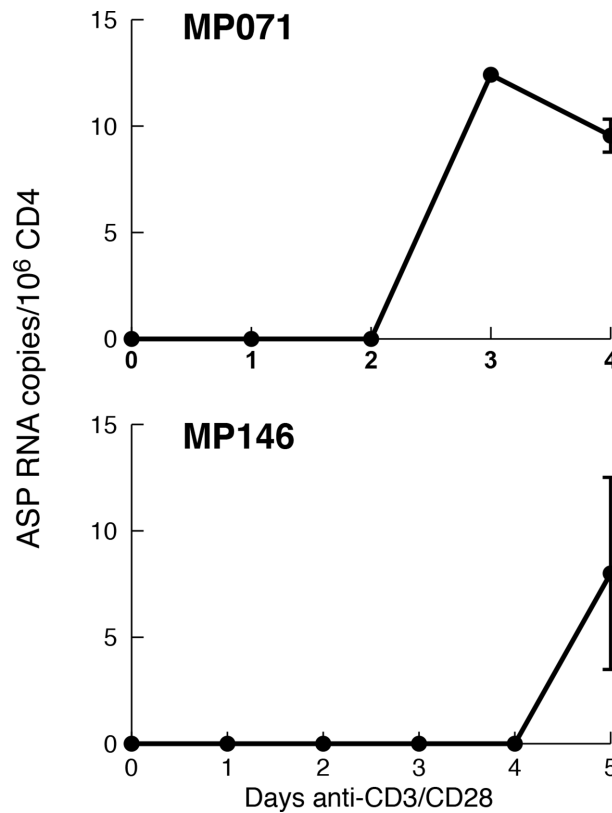


Fig. 5. Quantitative analysis of ASP RNA expression in CD4⁺ T cells isolated from two ART-treated, aviraemic patients after stimulation with anti-CD3/28. In both patients, no ASP can be detected in unstimulated cells (day 0), with low levels of ASP RNA only becoming detectable at day 2 (MP071) and day 4 (MP146) of the time course. The results are expressed as ASP RNA copies/million CD4⁺ T cells. Each point in the time course represents the mean value of triplicate PCR reactions.

isolated from five HIV-infected individuals. Three of our subjects were not receiving therapy at the time of sampling and had detectable viraemia, whereas the other two were undergoing ART and were aviraemic (Table 1). ASP RNA isolated from the three viraemic, untreated patients was cloned and sequenced. Given the high genetic diversity that characterizes *env* V4 and V5 in patients during the asymptomatic stage at both inter- and intra-patient level [21–23], the question arises of whether the products of ASP transcription are also characterized by hypervariability, at least to some degree. To address this point, and to assess whether differences existed in terms of length variants between the RNA transcript and the genomic sequence pools, the genomic *env* region corresponding to ASP in the serum of each patient was also cloned and sequenced. The phylogenetic relations among the HIV isolates in the three viraemic patients and their subtype assignment are shown Fig. 1. The tree was inferred using the maximum-likelihood method with 1000 bootstrap replicates using MEGA 7.0 software. Sequences from both cells and serum were aligned with the reference sequences of the HIV-1 M group subtypes, which were obtained from the HIV sequence database (<https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>).

All sequences, both ASP and *env*, were in the *env* reading frame. As shown in Fig. 1, the sequences from both ASP and *env* all cluster with the HXB2-LAI-IIIB-BRU, confirming that they are all of the B subtype. In total, 56 nucleotide sequences were obtained, 26 from CD4⁺ T lymphocytes (ASP RNA transcripts) and 30 from serum (*env* genomic RNA). The sequences are available in GenBank under accession numbers MH756691–756716 (RNA transcripts) and MH756717–756746 (genomic RNA).

Development of a modified RT-PCR protocol for antisense RNA detection

The experimental approach for the detection of ASP RNA by RT-PCR was developed in PBMCs from one healthy donor acutely infected *in vitro* with HXB2. In our first attempt at detecting ASP RNA, we amplified a short fragment (primer pair ASP141) using a standard antisense primer (ASP141R) to perform RT. As shown in Fig. 2a, the amplifications were successful, as we could clearly see a band of the expected size (141 bp) in the HIV-1-infected PBMCs (lanes 1–3). However, a band of the same molecular weight and of the same intensity was also visible in RT reactions performed without the RT primer (lanes 4–6). In

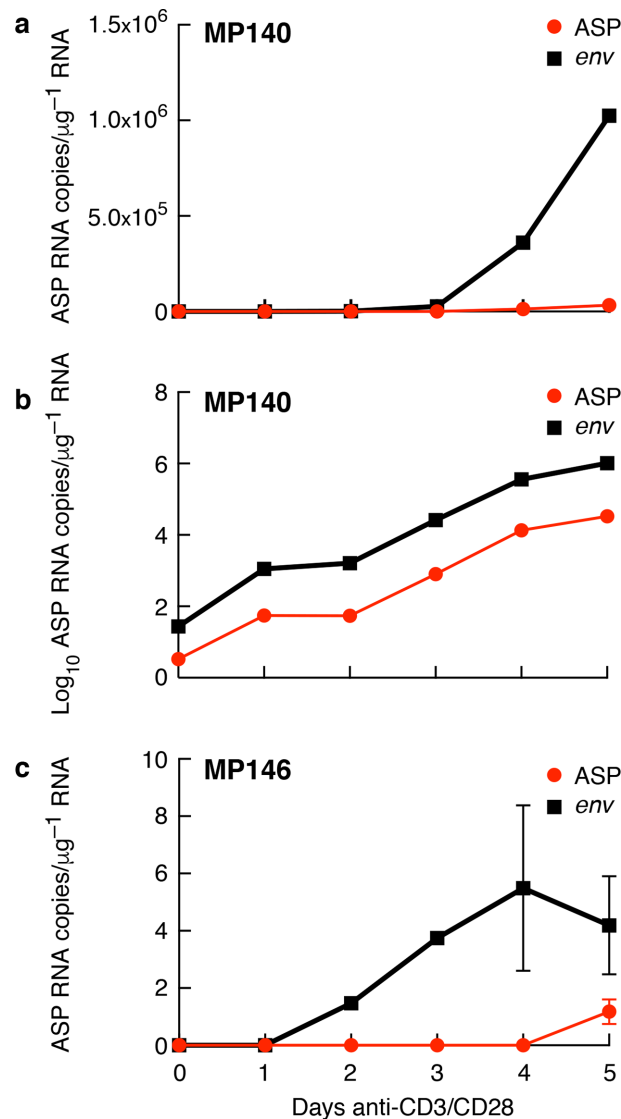


Fig. 6. Quantitative analysis of ASP and *env* RNA in CD4⁺ T cells stimulated with anti-CD3/CD28 in one treated (MP146) and one untreated (MP140) patient in the same qRT-PCR. (a) Linear graphic representation of ASP and *env* RNA in one patient in the absence of therapy (MP140), showing that *env* expression occurs at levels that are much higher than those of ASP. (b) The same data as above are represented on a logarithmic scale, showing that ASP and *env* share similar patterns of expression. (c) Linear graphic representation of ASP and *env* RNA in one aviraemic patient undergoing ART (MP146). ASP and *env* are both characterized by low levels of expression, with *env* levels being higher than ASP levels and already becoming detectable at day 3 post-stimulation.

contrast, no bands were detected in either RT minus controls or in uninfected PBMCs (lanes 7–8), indicating that the bands in the RT primer minus controls were not due to either leftover genomic DNA in the RNA or to cross-contamination during PCR.

Our next step was to perform the RT using a biotinylated version of the specific antisense primer, followed by purification of the biotinylated cDNA, as described by Haist *et al.* [17]. The results of this approach are shown in Fig. 2(b). A clear band is present in RT reactions performed with the biotinylated primer (lanes 1–3), whereas a band

characterized by markedly lower intensity is present in primer minus controls (lanes 4–6). Although very encouraging, these results were not yet satisfactory. In their paper, Haist and coworkers also obtained some residual non-specific cDNA, which they eliminated by increasing the stringency of the bead washes [17]. On our part, we wanted to prevent the synthesis of non-specific products and only amplify ASP RNA. We reasoned that, if indeed RNA secondary structures could prime the RT reaction, then the answer was total linearization of the RNA prior to RT, combined with biotinylation of the specific antisense primer and

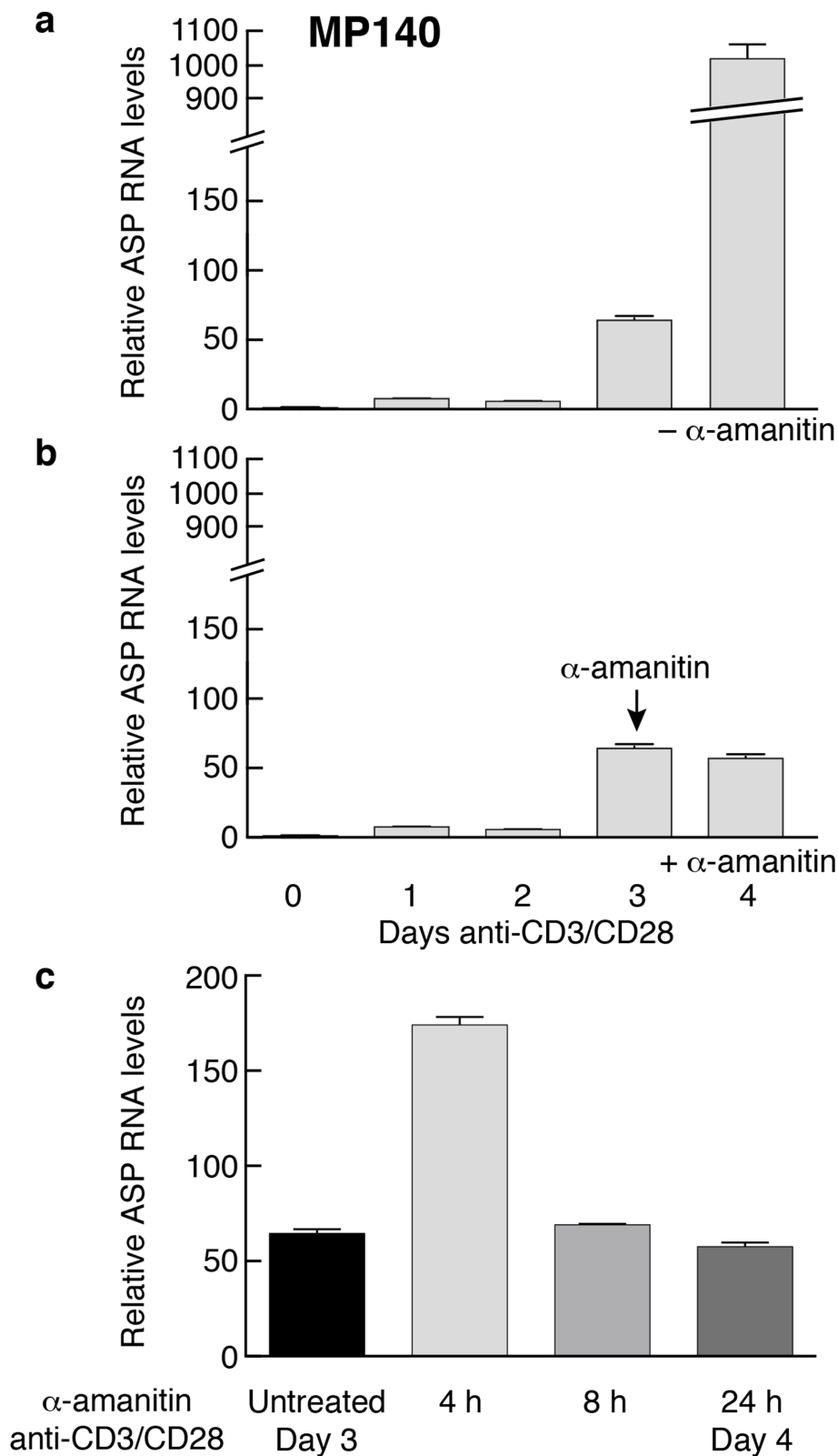


Fig. 7. Inhibition of ASP expression in anti-CD3/CD28-stimulated CD4⁺ T cells from one untreated, viraemic patient (MP140) following treatment with 20 $\mu\text{g ml}^{-1}$ α -amanitin. (a) Time-course of ASP expression in absence of α -amanitin, showing a sharp increase of the relative ASP RNA levels between day 3 and day 4 post-stimulation. (b) If α -amanitin is added to the cultures at a concentration of 20 $\mu\text{g ml}^{-1}$, the increase in relative ASP RNA levels is inhibited. (c) Inhibition of ASP expression in anti-CD3/CD28-stimulated CD4⁺ T cells from one untreated, viraemic patient (MP140) following treatment with 20 $\mu\text{g ml}^{-1}$ α -amanitin.

ml⁻¹ at day 3 post-stimulation, the peak in ASP RNA expression at day 4 is inhibited. (c) Time course of ASP RNA inhibition showing that ASP RNA levels decrease rapidly between 4 and 8 h after initiation of treatment.

cDNA purification prior to PCR. The manufacturer's protocol for SuperScript III Reverse Transcriptase, the enzyme we were using, recommended denaturing RNA at 65 °C. In order to achieve total linearization of the RNA, we decided to increase the denaturation temperature up to 94 °C, followed by immediate cooling of the tube in iced water. As shown in Fig. 3(a), the linearization step still produces a band of the right molecular weight (lanes 1–3), with total elimination of non-specific amplification products (lanes 4–6). In contrast, amplification of unpurified cDNA obtained by RT of linearized RNA in the absence of the RT primer still produces a band (lane 7). No signals were detected in RT minus controls (Fig. 3b, lanes 1–3), uninfected PBMCs or water (Fig. 3b, lanes 4 and 5). Sequencing of this band directly from the PCR reaction confirmed that the amplified product was indeed ASP (data not shown).

ASP transcripts have been detected in CD4⁺ T cells isolated from viraemic, untreated HIV-infected subjects following stimulation with anti-CD3/CD28. Having developed a reliable tool for antisense RNA detection, we were curious to see whether we could also assess the presence of ASP RNA in HIV-infected individuals. Our first attempts to detect ASP transcripts were carried out in total PBMCs, either resting or stimulated with anti-CD3/CD28, from the three viraemic subjects, MP135, MP140 and MP148. This approach, however, did not produce any results (data not shown). Attempts at detecting ASP RNA in purified, resting CD4⁺ T cells using a panel of oligoes amplifying target sequences of various sizes (ASP141, ASP171, PanASP, patient-specific primers) were equally unsuccessful (data not shown). In contrast, ASP detection in purified CD4⁺ T cells stimulated with anti-CD3/CD28 produced a clear, specific signal in each of the three patients tested (Figs S1–S3, available in the online version of this article). The kinetics of expression measured by qPCR show that in MP135 and MP140, the ASP RNA increased over the first 4 days of stimulation, peaking at day 4 (MP140), whereas in MP148 the peak had already been reached at day 2 post-stimulation (Fig. 4).

Low levels of ASP RNA can be detected in treated aviraemic patients after stimulation with anti-CD3/CD28

Our next question was whether ASP could be detected in treated subjects with undetectable viraemia. To this end, we analysed ASP expression by qPCR in anti-CD3/CD28-stimulated CD4⁺ T cells isolated from three additional patients, MP069, MP071 and MP146, who were undergoing suppressive therapy at the time of sampling. In patient MP069 we could not visualize any ASP bands, regardless of the starting template (proviral DNA, total RNA) and the primers (Pan-ASP, ASP171, ASP141) used (data not shown). In the absence of the amplification of proviral DNA, we could not develop sequence-specific primers for this patient. As a

consequence, we cannot say whether the failure to amplify ASP RNA was due to low similarity between the primer and template sequences preventing amplification from occurring, or to a real lack of ASP expression in this subject. In the other two treated patients, MP071 and MP146, low levels of ASP RNA could be detected at days 3 and 5 post-stimulation, respectively (Fig. 5). The levels of expression in these two patients were markedly lower than those observed in patients who had detectable viraemia, corresponding to 10–15 copies/million CD4⁺ T cells.

Similar patterns of expression of ASP and *env* RNA could be observed in one untreated, viraemic patient (patient MP140). Next we wanted to know whether there were differences in the kinetics of expression of ASP and *env* in the same HIV-infected individual. To answer this question, we analysed ASP and *env* RNA expression at different time points following stimulation of cells with anti-CD3/CD28 in one treated (MP146) and one untreated (MP140) patient in the same qRT-PCR. The results of this experiment are shown in Fig. 6 and Table S1. In patient MP140, both ASP and *env* were detected, although with a profound difference in transcription levels, with ASP quantified at 3.3×10^4 copies μg^{-1} of total RNA and *env* at 1×10^6 copies μg^{-1} (Fig. 6a and Table S1). Interestingly, the curves of expression of the two genes follow identical trends, as can be seen clearly if the data are plotted on a logarithmic scale (Fig. 6b). In MP146, which was treated and aviraemic, the expression of both genes was very low, with *env* peaking at day 4 post-stimulation with 5.5 copies μg^{-1} of total RNA and ASP peaking at day 5 with 1.17 copies μg^{-1} (Fig. 6c and Table S1). Given the low levels of ASP RNA and its delayed expression, it is not possible to assess the similarities between the two curves in this patient.

α -amanitin greatly reduces ASP RNA transcription in one patient in the absence of therapy (patient MP140)

Next we tested the sensitivity of ASP RNA to α -amanitin. To this end, we stimulated CD4⁺ T cells isolated from patient MP140 with anti-CD3/CD28, and added α -amanitin at increasing concentrations of 5, 10, 20 and 100 $\mu\text{g ml}^{-1}$ at day 3 post-stimulation, just prior to the peak in ASP expression, which in this patient occurred at day 4 (Fig. 4). Cells were harvested at 4, 8 and 24 h post-treatment and ASP RNA was quantified and normalized using U6 snRNA, a housekeeping gene transcribed by RNAP III [24] and thus not affected by our working concentrations of α -amanitin. As shown in Fig. 7(a) and Table S2, the addition of 20 $\mu\text{g ml}^{-1}$ α -amanitin at day 3 post anti-CD3/CD28 stimulation strongly inhibited the expression of ASP. This inhibitory effect could be fully appreciated at day 4, the peak day for ASP expression in this patient. Fig. 7(b) shows the kinetics of ASP inhibition at 4, 8 and

24 h post-treatment, with most of the inhibition occurring between 4 and 8 h following the addition of α -amanitin at $20 \mu\text{g ml}^{-1}$. When lower concentrations of α -amanitin were used, corresponding to 5 or $10 \mu\text{g ml}^{-1}$, there were no significant inhibitory effects between time points at 5 and $10 \mu\text{g ml}^{-1}$, although a trend towards decreased levels of ASP RNA could already be observed between the 4 h time points at 5 and $10 \mu\text{g ml}^{-1}$ (Fig. S4). The IC_{50} for α -amanitin, determined by nonlinear regression using GraphPad Prism version 8, was equal to $8.435 \mu\text{g ml}^{-1}$ (Fig. 5).

Similar sequence variants can be detected in cells and serum in untreated, viraemic patients during early HIV infection. ASP RNA is synthesized from the complementary strand of *env* spanning V4 and V5, two hypervariable regions that have been shown to harbour marked length polymorphism at intra-patient level [21–23]. As a consequence, it was possible that differences may exist between the pool of length variants in ASP RNA from cells and in the genomic *env* RNA pool. To answer this question, genomic *env* from serum of the viraemic patients was also cloned and sequenced. As shown in Table 2, the ASP length variants that are dominant in cells are the same as those in genomic *env* from serum. No ASP RNA was detected in patients' serum (Fig. S6).

ASP ORFs can be of variable length due to different types of stop codons

In HXB2, the ASP ORF is characterized by a TAG at positions 7375–7373, overlapping on the plus strand with the codons for phenylalanine (F383) and tyrosine (Y384) of the gp120 motif CGGEFFY, a highly conserved sequence located at the junction C3/V4. Since this is the only stop codon found in HXB2 ASP, and since it is located in a very conserved region, we refer to it as 'canonical', meaning a TAG stop codon with the same sequence and genomic location as in HXB2. ASP ORFs of different length were identified in patients for the various clones. Table 2 summarizes the length variants (i.e. the distance between the canonical start and stop codons) of the ASP ORF in cells and serum. In patient MP135, 18 sequences were obtained, 10 from CD4 cells (ASP) and 8 from serum (*env*). In sequences from cells, one clone had the full canonical ASP ORF, resulting in a theoretical peptide of 192 amino acids (aa). Of the remaining sequences, three clones had a TGA stop codon internal

to the canonical reading frame, resulting in a shorter ORF of 115 aa, whereas the other six had a TGA stop codon downstream from the canonical TAG, resulting in an ORF of 212 aa, i.e. a little longer than the ORF found in HXB2. No canonical ASP ORF was found in genomic RNA from serum. In patient MP140, 17 clones were obtained, 8 from cells (ASP) and 9 from serum (*env*). In both groups, about half of the clones carried the canonical full coding region of 187 aa, whereas the other half carried a shorter variant of 104 aa. In patient MP148, a total of 21 sequences were obtained, 8 from CD4 and 13 from serum. In all clones, only a truncated ORF of 172 aa was found in both cells and serum.

Non-canonical stop codons are located in variable regions of gp120

Non-canonical stop sites mostly occurred within the canonical ORF, giving rise to truncated ORFs, in regions that corresponded to gp120 V4 and V5. In patient MP135, 3/10 clones from CD4 RNA and 2/8 from serum RNA had a TGA stop codon occurring at the amino terminus of the corresponding *env* V5, giving rise to a truncated ORF of 115 aa. In addition, a longer ORF of 212 aa was also identified in 6/10 clones from cells and 6/8 from serum. In these clones, the TAG canonical stop disappeared due to a C=>A mutation, leading to the substitution of phenylalanine F383 with leucine L383 and giving rise to the sequence CGGEFLY, which not only does not interrupt the *env* reading frame, but is also quite conserved as an alternative to CGGEFFY. Due to this mutation, the ASP ORF extended to the C-terminus of C3, a region that has been shown to harbour variability to some degree at both inter- and intra-patient level [23]. In patient MP140, four/eight clones in cells and five/nine in serum were found to be carrying a stop codon in V5. Finally, in patient MP148, only one truncated ASP variant was identified, with a premature stop codon occurring in V4 (Table 2).

The *env* fragment on the plus strand of non-canonical ASP ORFs does not appear to be defective

Regarding the plus strand sequence, a clearly defective *env* in which the reading frame was interrupted by one or more stop codons was only observed in three clones from cells

Table 2. Localization of ASP stop codons in gp120. Asterisks indicate the canonical stop codon at the junction C3/V4. Clones carrying the canonical ORF are in bold

Patient ID	No. of clones		Region of stop codon	ORF length (aa)
	CD4 (ASP)	Serum (<i>env</i>)		
MP135	1/10	0/8	C3*	192
	3/10	2/8	V5	115
	6/10	6/8	C3	212
MP140	4/8	4/9	C3*	187
	4/8	5/9	V5	104
MP148	8/8	13/13	V4	172

and one clone from serum. In patient MP135, the only CD4 clone carrying the canonical ORF also carried a defective *env* on the plus strand. In contrast, in patient MP140, the clones carrying the canonical ASP also had a seemingly functional *env* sequence on the plus strand, at least in the fragment analysed in this study. In this patient, two clones were found to be carrying the defective *env*, one in cells and the other in serum. In both cases, the ASP sequence was non-canonical with a stop site in V5. In patient MP148, a defective *env* was also identified in one clone from CD4. In all the other clones, the sequence of *env* appeared to be in-frame, although we cannot exclude the occurrence of frame-shifts and/or stop codons in regions other than ASP.

DISCUSSION

By using a modification of the protocol proposed by Haist *et al.* [17], we detected and sequenced ASP RNA transcripts in CD4⁺ T lymphocytes in three viraemic patients, who were either naïve to therapy or untreated. We also detected ASP expression in two additional patients who were undergoing ART and had undetectable viraemia, although at levels that were lower than those found in untreated subjects. In all patients, in order to detect ASP RNA, CD4⁺ T cells had to be stimulated *in vitro* with anti-CD3/CD28, since no ASP RNA could be detected in resting cells. In a previous study, Zapata *et al.* proposed that ASP is a regulatory RNA involved in HIV latency [14]. In their paper, they reported low levels (10–30 copies of RNA/10⁶ CD4⁺ T cells) of ASP expression in resting CD4⁺ T cells from three patients under suppressive ART for over 24 months. We also detected ASP expression in treated patients at similar levels to those described by Zapata and coworkers; in our case, however, CD4⁺ T lymphocytes had to be stimulated for 2 (MP071) to 4 (MP146) days in order to see ASP expression, since no signal could be detected in resting cells. This discrepancy could be explained by the fact that they used a regular RT-PCR protocol that does not involve any RNA denaturation step. We have shown that in the absence of a rigorous denaturation, non-specific synthesis of sense RNA (i.e. *env*) may occur to some extent, which could explain the low signal they detected in their samples. We also show that the expression of ASP and *env* share similar profiles over time, which is in contrast with the proposed role of ASP as a regulator of latency [14]. In fact, if that were the case, an inverse correlation between *env* (sense) and ASP (antisense) transcription should be expected [2]. The finding that in untreated, viraemic patients, the transcription levels of *env* are over 2 logs more abundant than those of ASP fully agrees with the data reported by Laverdure *et al.* [25], indicating that in activated primary CD4⁺ T cells infected *in vitro*, 3'LTR activity is very low, up to 1000-fold lower than the 5' LTR sense transcription. Our data clearly show that infected CD4⁺ T cells can express ASP regardless of the stage of the disease, at least in those patients in which the ASP ORF is present. We also show that cells need to be stimulated in order to produce ASP RNA and that even cells from ART-treated patients are able to express ASP,

although to a lesser extent than cells from untreated ones. In addition, our data represent the first nucleotide sequences of ASP transcripts in infected subjects.

The ASP ORF is known to be quite conserved among isolates of subtype B [10], and in fact we found it in all of the analysed clones. However, the length of this ORF was variable, due to a shift of the stop codon upstream or downstream from its canonical position in HXB2, leading to ASP ORFs that were shorter or longer than the HXB2 ORF. In HXB2, the ASP stop site is a TAG spanning the codons for F383 and Y384 of the motif CGGEFFY, at the junction C3/V4 on the plus strand. Cassan *et al.* [10] defined this kind of stop codon as being imposed by the coding of Env. Indeed, this motif of gp120 is highly conserved and contains several functional and structural sites involved in Env–host biological signalling, such as part of the coreceptor-binding site outside V3 (F383) and part of the recombinant human monoclonal antibody IgG1b12 epitope (Y384) [26]. The degree of conservation of this motif is so high that it can be found unchanged not only across HIV-1 subtypes and groups, but also in other species of lentiviruses, such as simian immunodeficiency virus (SIV). In cells, although the ASP ORF was present in all the clones, the canonical (i.e. as found in HXB2) TAG site could only be observed in two patients, MP135 (in 12.5 % of the clones) and MP140 (in exactly 50 % of the clones). Given the very low number of sequences analysed in this study, these values can be considered to be quite relevant. Interestingly, no canonical ASP ORF was found in MP148, who was also the earliest of our patients and had been infected for less than a year. Our inability to recover the canonical ORF in this patient does not necessarily imply its absence early during infection, since this could have been due to its frequency being lower than our detection limit, as in the case of patient MP135, in which the canonical ASP ORF was recovered from cells but not from serum.

In addition to the stop codons imposed by the coding of *env*, Cassan *et al.* [10] also observed stop codons due to mutations on the plus strand that can appear/disappear without modifying *env*, seemingly corresponding to those that we observed in V5, V4 and the C-terminus of C3 on the plus strand. The obvious question at this point is whether ORFs carrying alternative (i.e. non-canonical) stop codons may still encode for a functional ASP RNA or protein. It would be logical to think that functional gene products are more likely to be encoded by ORFs that are relatively constant, including the position of the stop codon, which must be located in a region that is conserved enough to yield transcription/translation products sharing the same structural features, as in the case of the HXB2 canonical TAG. In this perspective, the actual occurrence of the ASP canonical stop codon in such a highly conserved region appears to be the product of a complex process of selection, whereby the transcription of a functional ASP may occur regardless of the high variability of its *env* template. However, the translation of novel ORFs created by the

mutational gain or loss of start and stop codons has been proposed as one of the mechanisms underlying the evolution of gene overlaps [27]. Thus, the canonical ORF as found in HXB2 could be considered to be the actual functional ASP ORF, whereas the frequent occurrence of alternative stop sites in highly variable regions could be explained by the ASP ORF being in the evolution process [10] through the constant generation of random-length mutants.

ASP and *env* are located in the same region of the HIV-1 genome, overlapping one another in opposite orientations. As a matter of fact, overlapping genes are a common feature of viruses, in which they typically code for accessory proteins involved in viral pathogenicity or spread [28, 29]. Overlapping proteins have been shown to have a sequence composition that is globally biased towards amino acids with a high codon degeneracy, such as arginine, leucine, or serine [28, 30], in order to compensate for mutations that may occur on the 'ancestral' or 'overprinted' protein to which they overlap. In addition, it has been shown that most overlapping proteins have been created anew and that most proteins created anew are orphans, meaning that they are restricted to just one species or genus [28]. Interestingly, not only are leucine and serine among the most abundant amino acids in the sequence of ASP [1, 11, 12], but also no clear links have been found between ASP and any other known three-dimensional structures in the Protein Data-bank (PDB) (A. V. Kajava, personal communication), suggesting that ASP may indeed be a protein-coding orphan gene, restricted solely to the M group of HIV-1. Sensitivity to α -amanitin is additional evidence for the existence of an ASP gene product. RNA polymerases (RNAP) are classified based on their differential sensitivity to α -amanitin. In animal cells, RNAP II is generally very sensitive, showing 50 % inhibition at 2–20 $\mu\text{g ml}^{-1}$ of α -amanitin [31–33], whereas RNAP I and RNAP III are resistant to higher concentrations, up to 1 mg ml^{-1} for RNAP I, and 100–150 $\mu\text{g ml}^{-1}$ for RNAP III. The fact that ASP expression is already inhibited by α -amanitin at a concentration of 20 $\mu\text{g ml}^{-1}$ indicates that it is likely associated with RNAP II, the enzyme responsible for mRNA synthesis during the transcription of protein-coding genes.

Further studies are under way to characterize the full-length transcript of ASP and assess its subcellular distribution in infected subjects. The discovery of a new HIV antigen would represent an important step in our understanding of HIV pathogenesis and perhaps open up new perspectives in the development of novel anti-HIV drugs and vaccines.

Funding information

This work received no specific grant from any funding agency.

Acknowledgements

We would like to thank Craig Fenwick for valuable feedback, Patrizia Amelio, Alessandra Noto and Matthieu Perreau for technical discussions, all the people in the Laboratory of AIDS Immunopathogenesis for technical assistance, and John and Aaron Weddle from VSB Associated, Inc. for excellent artwork. Finally, we would like to offer special

thanks to all the patients, without whom this work would not have been possible.

Author contributions

AM: conceptualization, methodology, validation, investigation, visualization. FAP: methodology, validation, investigation. TA: investigation, methodology, resources. EDC: methodology, review and editing. BTF: formal analysis, review and editing. GC: supervision, review and editing. CB: resources. GP: resources, funding. CG: methodology, validation, resources, original draft preparation, review and editing, visualization, supervision, project administration.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

These studies were approved by the Institutional Review Board of the Centre Hospitalier Universitaire Vaudois (CHUV) and subjects gave written informed consent.

References

1. Miller RH. Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. *Science* 1988;239:1420–1422.
2. Barbeau B, Mesnard JM. Making sense out of antisense transcription in human T-cell lymphotropic viruses (HTLVs). *Viruses* 2011;3:456–468.
3. Cavanagh MH, Landry S, Audet B, Arpin-André C, Hivin P *et al.* HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated. *Retrovirology* 2006;3:15.
4. Usui T, Yanagihara K, Tsukasaki K, Murata K, Hasegawa H *et al.* Characteristic expression of HTLV-1 basic zipper factor (HBZ) transcripts in HTLV-1 provirus-positive cells. *Retrovirology* 2008;5:34.
5. Halin M, Douceron E, Clerc I, Journo C, Ko NL *et al.* Human T-cell leukemia virus type 2 produces a spliced antisense transcript encoding a protein that lacks a classic bZIP domain but still inhibits Tax2-mediated transcription. *Blood* 2009;114:2427–2438.
6. Chevalier SA, Ko NL, Calattini S, Mallet A, Prévost MC *et al.* Construction and characterization of a human T-cell lymphotropic virus type 3 infectious molecular clone. *J Virol* 2008;82:6747–6752.
7. Larocque É, André-Arpin C, Borowiak M, Lemay G, Switzer WM *et al.* Human T-cell leukemia virus type 3 (HTLV-3) and HTLV-4 antisense-transcript-encoded proteins interact and transactivate Jun family-dependent transcription via their atypical bZIP motif. *J Virol* 2014;88:8956–8970.
8. Switzer WM, Salemi M, Qari SH, Jia H, Gray RR *et al.* Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. *Retrovirology* 2009;6:9.
9. Kobayashi-Ishihara M, Yamagishi M, Hara T, Matsuda Y, Takahashi R *et al.* HIV-1-encoded antisense RNA suppresses viral replication for a prolonged period. *Retrovirology* 2012;9:38.
10. Cassan E, Arigon-Chifolleau AM, Mesnard JM, Gross A, Gascuel O. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. *Proc Natl Acad Sci USA* 2016;113:11537–11542.
11. Briquet S, Vaquero C. Immunolocalization studies of an antisense protein in HIV-1-infected cells and viral particles. *Virology* 2002;292:177–184.
12. Clerc I, Laverdure S, Torresilla C, Landry S, Borel S *et al.* Polarized expression of the membrane ASP protein derived from HIV-1 antisense transcription in T cells. *Retrovirology* 2011;8:74.
13. Saayman S, Ackley A, Turner AW, Famiglietti M, Bosque A *et al.* An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription. *Mol Ther* 2014;22:1164–1175.
14. Zapata JC, Campilongo F, Barclay RA, Demarino C, Iglesias-Ussel MD *et al.* The human immunodeficiency virus 1 ASP RNA promotes viral latency by recruiting the Polycomb Repressor

- Complex 2 and promoting nucleosome assembly. *Virology* 2017; 506:34–44.
15. Torresilla C, Larocque É, Landry S, Halin M, Coulombe Y *et al.* Detection of the HIV-1 minus-strand-encoded antisense protein and its association with autophagy. *J Virol* 2013;87:5089–5105.
 16. Michael NL, Vahey MT, D'Arcy L, Ehrenberg PK, Mosca JD *et al.* Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J Virol* 1994;68:979–987.
 17. Haist K, Ziegler C, Botten J. Strand-specific quantitative reverse transcription-polymerase chain reaction assay for measurement of arenavirus genomic and antigenomic RNAs. *PLoS One* 2015;10: e0120043.
 18. Noto A, Procopio FA, Banga R, Suffiotti M, Corpataux JM *et al.* CD32⁺ and PD-1⁺ lymph node CD4 T cells support persistent HIV-1 transcription in treated aviremic individuals. *J Virol* 2018;92.
 19. Banga R, Procopio FA, Ruggiero A, Noto A, Ohmiti K *et al.* Blood CXCR3⁺ CD4 T cells are enriched in inducible replication competent HIV in aviremic antiretroviral therapy-treated individuals. *Front Immunol* 2018;9:144.
 20. Wang XS, Gong JN, Yu J, Wang F, Zhang XH *et al.* MicroRNA-29a and microRNA-142-3p are regulators of myeloid differentiation and acute myeloid leukemia. *Blood* 2012;119:4992–5004.
 21. Castro E, Bélair M, Rizzardi GP, Bart PA, Pantaleo G *et al.* Independent evolution of hypervariable regions of HIV-1 gp120: V4 as a swarm of N-Linked glycosylation variants. *AIDS Res Hum Retroviruses* 2008;24:106–113.
 22. Belair M, Dovat M, Foley B, Mayerat C, Pantaleo G *et al.* The polymorphic nature of HIV type 1 env V4 affects the patterns of potential N-glycosylation sites in proviral DNA at the intrahost level. *AIDS Res Hum Retroviruses* 2009;25:199–206.
 23. Guglietta S, Pantaleo G, Graziosi C. Long sequence duplications, repeats, and palindromes in HIV-1 gp120: length variation in V4 as the product of misalignment mechanism. *Virology* 2010;399: 167–175.
 24. Kunkel GR, Maser RL, Calvet JP, Pederson T. U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci USA* 1986;83:8575–8579.
 25. Laverdure S, Gross A, Arpin-André C, Clerc I, Beaumelle B *et al.* HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells. *J Virol* 2012;86:13785–13789.
 26. Korber B, Gnanakaran S. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. *Curr Opin HIV AIDS* 2009;4:408–417.
 27. Belshaw R, Pybus OG, Rambaut A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res* 2007; 17:1496–1504.
 28. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J Virol* 2009;83: 10719–10736.
 29. Chirico N, Vianelli A, Belshaw R. Why genes overlap in viruses. *Proc Biol Sci* 2010;277:3809–3817.
 30. Pavesi A, de Iaco B, Granero MI, Porati A. On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. *J Mol Evol* 1997;44:625–631.
 31. Keding C, Gniazdowski M, Mandel JL, Gissinger F, Chambon P. Alpha-amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. *Biochem Biophys Res Commun* 1970;38:165–171.
 32. Jacob ST, Sajdel EM, Munro HN. Specific action of alpha-amanitin on mammalian RNA polymerase protein. *Nature* 1970;225:60–62.
 33. Lindell TJ, Weinberg F, Morris PW, Roeder RG, Rutter WJ. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* 1970;170:447–449.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.