Review

The HIV-1 Antisense Gene ASP: The New Kid on the Block

Zahra Gholizadeh, Mohd. Shameel Iqbal, Rui Li and Fabio Romerio *

Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; zgholiz1@jhu.edu (Z.G.); miqbal9@jhmi.edu (M.S.I.); rli74@jhmi.edu (R.L.)
* Correspondence: fromeri2@jhmi.edu

Abstract: Viruses have developed incredibly creative ways of making a virtue out of necessity, including taking full advantage of their small genomes. Indeed, viruses often encode multiple proteins within the same genomic region by using two or more reading frames in both orientations through a process called overprinting. Complex retroviruses provide compelling examples of that. The human immunodeficiency virus type 1 (HIV-1) genome expresses sixteen proteins from nine genes that are encoded in the three positive-sense reading frames. In addition, the genome of some HIV-1 strains contains a tenth gene in one of the negative-sense reading frames. The so-called Antisense Protein (ASP) gene overlaps the HIV-1 Rev Response Element (RRE) and the envelope glycoprotein gene, and encodes a highly hydrophobic protein of ~190 amino acids. Despite being identified over thirty years ago, relatively few studies have investigated the role that ASP may play in the virus lifecycle, and its expression in vivo is still questioned. Here we review the current knowledge about ASP, and we discuss some of the many unanswered questions.

Keywords: HIV-1; natural antisense transcription; antisense protein (ASP); overprinting; replication; pathogenesis; spread

1. Introduction: De Novo Creation of Genes

In the majority of cases, new genes are created by transfer of existing genetic material [1]. This can occur in various forms: exon shuffling, gene duplication, retroposition, lateral gene transfer, and gene fusion or fission [1]. However, in rare cases, genes can also be created de novo. This mechanism was thought to be rare [2], but recent studies show that it occurs frequently. De novo gene creation can occur in intergenic regions and introns [3,4], but also in genomic regions that already contain a protein-coding gene—a process called 'overprinting' (Figure 1). In this case, a genomic region with an existing coding sequence in one of the six reading frames (Figure 1A) undergoes point mutations in one of the other five reading frames that generate a new *start* codon (Figure 1B) and/or remove *stop* codons (Figure 1C), giving rise to a second coding sequence. The resulting genomic region contains two overlapping open reading frames (ORFs): the ancestral or 'overprinted' gene, and the novel or 'overprinting' gene [5]. For most gene pairs, the identification of the ancestral and the de novo gene can be determined with high accuracy because the former presents a much wider phylogenetic distribution than the latter [6]. Indeed, proteins that are created de novo by overprinting do not have homologs in other organisms and are at times referred to as taxonomically restricted or 'ORFans' [7,8]. However, it is also possible that de novo proteins may be members of a family of proteins that have diverged beyond recognition or that were lost [9].
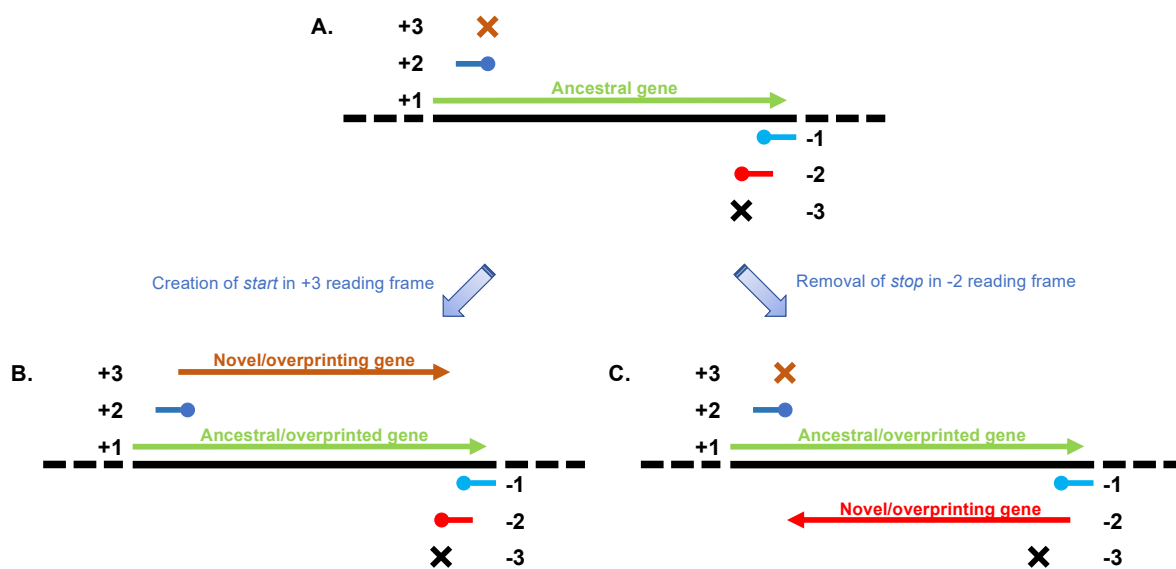
Figure 1. De novo creation of genes by overprinting. Typically, genomic regions contain a single gene encoding a protein product in one of the 6 reading frames (for example, reading frame +1 in panel (**A**). The same genomic region does not encode any protein in the other 5 reading frames, either because of absence of *start* codons (as in reading frame +3 and −3, marked with ×) or because of the presence of *stop*s shortly after a *start* codon (as in reading frames +2, −1 and −2, marked with ●). The occurrence of single point mutations in one of these 5 reading frames can give rise to a new protein-coding gene, either by creating a new *start* codon (as in reading frame +3 in panel (**B**) or by eliminating a *stop* codon (as in reading frame −2 in panel (**C**). When this occurs, the original gene is called 'ancestral' or 'overprinted', while the new one is called 'novel' or 'overprinting'. In rare cases, both events can occur over time (as in the case of HTLV-1 and HIV-1), thus giving rise to genomic regions that contain three overlapping protein-coding genes.

2. Overprinting in Viral Genomes

Overprinting has been documented in both prokaryotic and eukaryotic genomes [10–17]. However, the incidence of overlapping genes in these organisms is relatively low [18–20]. On the other hand, overprinting is quite frequent in viral genomes [21, 22]. Indeed, the first evidence of overlapping genes came from the bacteriophage ΦX174 [23], then followed by many examples in several eukaryotic viruses. Two theories have been proposed to explain the high abundance of overlapping genes in viral genomes [24]. The gene-compression theory states that error-prone viral polymerases and biophysical constraints imposed by the viral capsid drive the creation of overlapping genes that allow maximization of the amount of information that can be encoded in small genomes [25]. The gene-novelty theory asserts that de novo creation of genes is driven by selective pressure, and gives rise to gene products that provide a selective advantage to the virus, and thus become fixed in the population [5].

The high abundance of overlapping genes in viral genomes has allowed the use of statistical and computational methods to investigate the composition bias, structural features, evolution, and potential function of overlapping gene pairs and de novo proteins [5,24,26,27]. Sequence composition varies greatly between overlapping and non-overlapping genes, as well as between ancestral and novel genes in overlapping pairs [27]. Pavesi et al. showed that, in the case of overlapping pairs, the codon usage of ancestral genes is more similar to the rest of the genome than that of novel genes [26]. While this difference decreases with the age of de novo genes [6], constraints imposed by the ancestral gene might prevent the novel gene from acquiring a codon usage completely similar to that of the rest of the genome [26]. Moreover, young novel genes initially evolve rapidly under positive, or weakly purifying selection, while older novel genes evolve more slowly under increasingly stronger negative selection [6]. Interestingly, in some cases overlapping genes show 'asymmetric evolution' whereby the two members of a pair evolve at different

rates [28]. Analysis of composition bias of proteins encoded by overlapping genes found that they are enriched in high-degeneracy amino acids (arginine, leucine, serine), which may alleviate evolutionary constraints acting on the pair of overlapping genes at the DNA level [27,29]. In addition, proteins encoded by overlapping genes are enriched in amino acids with high propensity toward structural disorder (arginine, proline, serine) [5,27,30,31]. This is an essential state of many proteins that, in the absence of a binding partner, lack a stable secondary and tertiary structure, and adopt a number of rapidly interconverting structural forms rather than a particular three-dimensional structure. Structural disorder affords proteins encoded by overlapping genes greater freedom of sequence evolution without loss of function [5]. Indeed, proteins with high intrinsic structural disorder are more likely to tolerate amino acid substitutions that maintain disorder.

The creation of a novel ORF by overprinting and its subsequent fixation in the viral population strongly suggest that the new protein provides a selective advantage [24]. Although most de novo genes encode for accessory proteins, they appear to play a role in the pathogenicity or spread of the virus [5]. Studies have shown that de novo proteins act by neutralizing the interferon and RNA interference responses of the host [32–35], by inducing apoptosis [36,37], or by promoting systemic spread of the virus [38]. Therefore, there is evidence that creation of novel genes by overprinting allows viruses to make virtue out of necessity.

3. A Special Kind of Overprinting: Antisense Genes

Most cases of overlapping genes studied to date are encoded in two different reading frames of the same DNA strand. There is also evidence of overlaps that involve three different ORFs. This is observed, for instance, in two human retroviruses: HIV-1 and HTLV-1. For HIV-1, the overlap involves the *env*, *tat*, and *rev* genes [39], and for HTLV-1 it involves the *p13/p30*, *tax*, and *rex* genes [40]. Significantly fewer are the examples in which the pair of overlapping ORFs are encoded on different DNA strands, and therefore in opposite orientations.

Without a doubt the viral antisense gene most intensely studied is the HTLV-1 basic leucine zipper (bZIP) factor (*hbz*) gene [41] (Figure 2). This gene maps in the pX region of the HTLV-1 genome, which has been called a 'gene nursery' due to the presence of high number of new genes and overprinting events [26]. The *hbz* gene is expressed via both spliced and unspliced antisense transcripts that originate from the proviral 3′LTR, which has been shown to contain a bidirectional promoter [42,43]. The negative sense promoter in the HTLV-1 3′LTR lacks a TATA box, relies on Sp1 binding sites, and is not under the control of the HTLV-1 transactivator TAX protein [43,44]. The *hbz* gene encodes for a 206-aa protein (HBZ) that plays a key role in the establishment of HTLV-1 chronic infection. Indeed, HBZ appears to promote proviral latency via interaction with several host factors that bind the viral cyclic AMP Response Elements (vCRE) in the HTLV-1 5′LTR, such as CREB, CREM and ATF-1 [41]. Interaction of HBZ with these factors precludes recruitment of TAX to the 5′LTR, and ultimately downregulates sense transcription [45]. Additionally, HBZ has been shown to prevent the binding of TAX to the host factor CBP/p300, and its recruitment of the 5′LTR [41]. There is evidence that HBZ may play a role in the leukemic process following HTLV-1 infection. Indeed, in about half of the ATL cases, the *tax* gene is not expressed due to deletion of the 5′LTR [46], epigenetic silencing of the 5′LTR [47], or mutations within the *tax* gene [48]. On the contrary, the 3′LTR and the pX region of HTLV-1 remain intact, and *hbz* is transcribed in all ATL cases [42]. In addition, knockdown of HBZ inhibits proliferation of ATL cells [49]. Finally, expression of HBZ in transgenic mice was shown to cause T cell lymphomas and inflammatory diseases that do not appear in TAX transgenic mice [50,51]. Altogether, there is strong evidence for the implication of the HTLV-1 antisense protein HBZ in leukemogenesis.
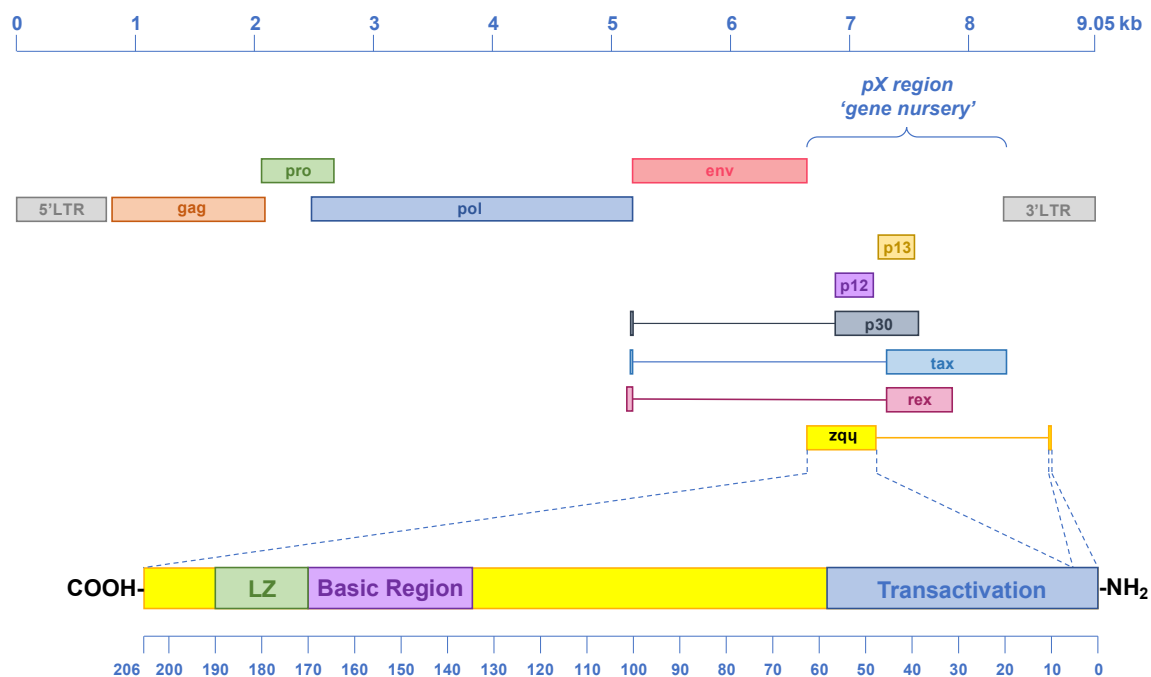
Figure 2. Genomic organization of the HTLV-1 provirus and main structural features of HBZ. The *hbz* gene is located within the *pX* region of the proviral genome gene, and it overlaps the p12 and p30 ORFs. HBZ is a basic leucine zipper (bZIP) transcription factor of 206 aa, and it contains three main domains: the transactivation domain, the basic region, and the leucine zipper (LZ). HBZ is involved in the establishment of HTLV-1 latency by preventing the recruitment of the HTLV-1 transactivator Tax to the proviral 5′LTR region.

Antisense genes are also found in other human and animal retroviruses. HTLV-2, 3 and 4 express the antisense proteins, APH-2, 3 and 4, respectively [52]. These proteins are similar to HBZ, suggesting that these ORFs were created in the common ancestor of HTLV-1/4 after it diverged from bovine leukemia virus (BLV) [26]. The simian T-cell leukemia virus (STLV) expresses the simian bZIP (SBZ) protein with function similar to HBZ [53]. An antisense ORF is present in the feline immunodeficiency virus (FIV) genome. However, while FIV is capable of producing antisense transcripts, expression of an antisense protein has not been detected [54]. Similarly, murine leukemia virus (MLV), bovine immunodeficiency virus (BIV) and BLV are capable of antisense transcription, but expression of antisense proteins by these viruses has not been demonstrated [52].

4. The HIV-1 Antisense Protein, ASP

The existence of an antisense gene in the HIV-1 proviral genome was first proposed more than 30 years ago with the identification of a highly conserved ORF in the minus strand of twelve viral isolates [55]. The *asp* ORF, located in the same genomic region as the *env* gene straddling the gp120/gp41 boundary (Figure 3), was also present in an additional 12/13 *env* sequences available (i.e., incomplete genome sequences) [55]. This newly identified HIV-1 gene was predicted to encode for an antisense protein (ASP) of ~190 residues, rich in hydrophobic amino acids and possibly associated with cellular membranes [55]. The hypothesis that the newly identified ORF might encode for a real protein was based on three lines of evidence. First, the ASP ORF is >300 nucleotides (i.e., >100 codons), which is uncommon in DNA strands complementary to known genes [56]. Second, the signal sequences necessary for production of an antisense mRNA transcript (promoter, poly-A addition signal, poly-A addition site, downstream G and T domains) are present and conserved in all sequences analyzed. Third, sequences necessary for the translation of a protein (canonical *start* and *stop* codons, a codon periodicity of 'G-nonG-N') are also present and conserved. Altogether, these analyses provided convincing

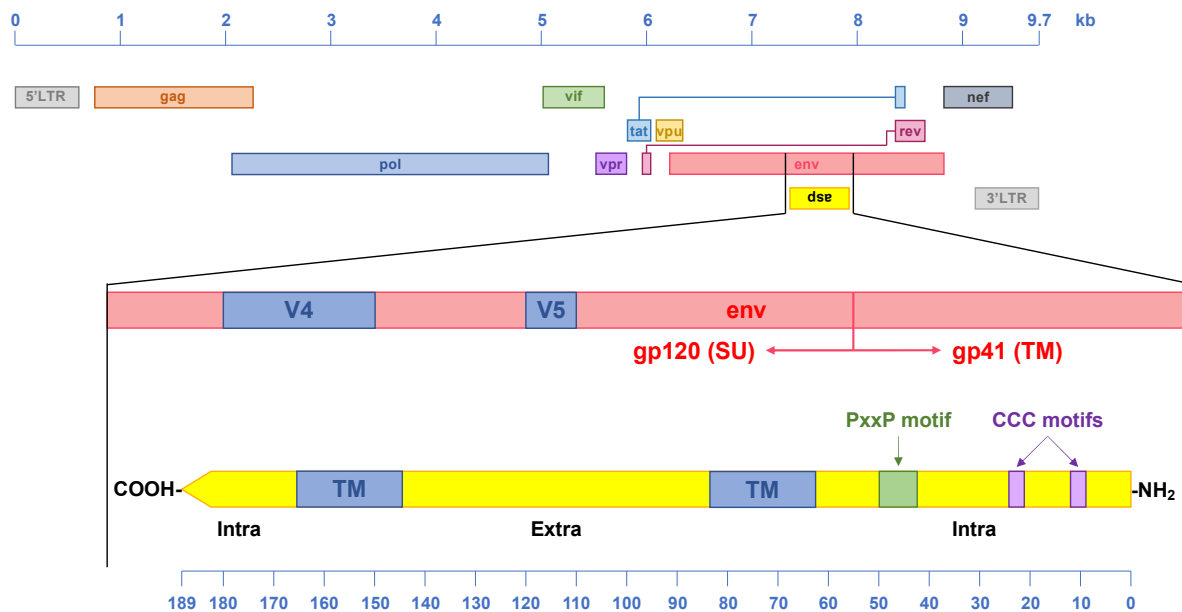theoretical evidence for the presence of a protein-coding antisense gene in the HIV-1 proviral genome [55].

Figure 3. Genomic organization of the HIV-1 provirus (group M) and main structural features of ASP. The *asp* gene is located within the *env* gene, and it straddles the gp120 (SU) and gp41 (TM) boundary. In particular, the *asp* gene overlaps the *env* gene in the region encoding the V4 and V5 loops of gp120. ASP encodes a protein of ~189 aa (the actual length varies across HIV-1 clades). It is predicted to contain two transmembrane domains (TM, between residues 63–84 and 146–167). The N-terminus and C-terminus are intracellular, while the portion between the two TM domains is extracellular. The N-terminus also includes highly conserved motifs: two cysteine-triplets, and a potential SH3 domain-binding PxxP motif.

At a glance, ASP presents several conserved features that may have functional significance (Figure 3). Although the crystal structure of ASP has not been resolved, sequence analysis and computer modeling suggest that ASP possesses two transmembrane (TM) domains comprised approximately between residues 65–85 and 145–165. The portion of the molecule between the two TM domains constitutes an extracellular loop, while the N-terminal and C-terminal ends of ASP are intracellular [57,58]. Additional features of interest are present in the N-terminal domain. They include, two closely spaced cysteine triplets located in the first 25 residues of the protein. Their function has not been investigated, but they might form disulfide bonds that stabilize the protein or coordinate heavy metals such as Zn^{2+} and Ni^{2+}. In addition, the N terminus of ASP includes a highly conserved PxxPxxP motif located between residues 40–50 of the protein. This motif is reminiscent of Src homology 3 (SH3) domain-binding motifs found in many proteins, including HIV-1 Nef [59]. Biochemical and functional studies are needed to establish whether the PxxP motif of ASP is a bona fide SH3 domain-binding motif. Finally, the intracellular portion of the C-terminal domain of ASP does not seem to contain conserved domains or motifs. This is in part due to the fact that it overlaps the V4 loop of gp120 on the opposite strand. Indeed, many HIV-1 strains and subtypes (including some of the ones with high prevalence, see below) express an ASP that is truncated shortly after the second TM domain, and that lacks the intracellular C-terminal domain altogether, suggesting that it may be dispensable for full ASP function. A more thorough and systematic experimental analysis is needed to fully evaluate the functional role of these and possibly other yet unidentified ASP domains.

5. ASP Expression in HIV-1 Infected Individuals

At least two recent reports have documented expression of HIV-1 antisense transcripts in cells collected from donors on ART using RT-qPCR [60,61] (Table 1). Our lab used a

strand-specific RT-qPCR assay to detect levels of antisense transcripts ranging from 10 to 30 copies per million resting CD4+ T cells [60]. Using a different approach, Mancarella et al. observed similar levels of antisense transcripts but only following ex vivo anti-CD3/CD28 stimulation of donor CD4+ T cells [61]. While direct evidence that ASP protein is expressed in HIV-1 infected individuals is still missing, multiple studies over two decades have documented the presence of humoral and cellular immune responses against ASP in HIV-1 infected individuals (Table 1). An early report by Vanhée-Brossollet et al. showed that sera obtained from 15 different HIV-1 infected patients, but not sera from HIV-1 negative individuals, was able to immunoprecipitate in vitro-translated ASP [62]. More recently, Savoret and colleagues demonstrated that antibodies to ASP are detectable in serum of HIV-1 patients who are off antiretroviral therapy [63].

Table 1. Evidence of ASP expression in HIV-1 infected individuals.

Target	Samples	Assay	Findings	Refs.
Anti-ASP antibodies	Sera from HIV-1 patients	IP	Sera from 15 HIV-1 patients were able to immunoprecipitate in vitro translated ASP; patients were at stage I, III and IV of infection during pre-HAART era	[62]
		LIPS	Antibodies to ASP are detectable in serum of untreated patients, but not in serum of treated patients or in serum of HIV-1 controllers	[63]
CTL responses against ASP	PBMC; acute and chronic; on and off HAART	ELISpot	PBMC from chronically infected patients both on and off HAART reacted to peptide pools spanning the ASP open reading frame	[64]
	PBMC; on and off HAART	ELISpot ICS	Detection of CD8+ T cell responses against ASP in patients off HAART; CD8+ T cells responses included production of cytokines and chemokines	[65]
	PBMC; chronic patients	ELISpot	Detection of CTL responses against several peptides matching the ASP consensus sequence	[66]
Antisense RNA	CD4+ T cells from treated patients	Strand-specific RT-qPCR	10–30 copies of HIV-1 antisense RNA per 10^6 resting CD4+ T cells without in vitro activation	[67]
	CD4+ T cells from treated and untreated patients	RT-qPCR with biotinylated RT primer	After anti-CD3/CD28 activation for 5 days: 5–10 copies of antisense RNA per 10^6 cells from treated patients; 20–2×10^6 copies of antisense RNA per 10^6 cells from untreated patients	[68]

A few studies also demonstrated the presence of CD8+ T cell responses against ASP-derived peptides in HIV-1 infected patients. Champiat et al. identified multiple alternate reading frames (ARFs) in both orientations of the HIV-1 genome, but most of them were present in the *env* region and in the same reading frame as ASP [62]. They then designed 9-mer peptide pools for each ARF, including 3 distinct pools based on the ASP ORF, and tested whether CD8+ T cells from HIV-1 infected individuals showed reactivity against these peptide pools. While no reactivity toward ASP-specific peptide pools was detected with cells from acutely-infected patients, CD8+ T cells from chronic patients (both on and off antiretroviral therapy) did react against ASP-derived peptides [62]. A subsequent study

focusing exclusively on ASP detected CD8+ T cell responses against ASP in HIV-1 positive subjects off antiretroviral therapy [63]. In addition, this study found that CD8+ T cells from these patients produced multiple cytokines and chemokines when exposed to ASP-derived peptides. Similar results were published in an independent report that investigated CD8+ T cell responses against five different antisense ORFs in the HIV-1 genome, including ASP [64].

Taken together, these studies provide indirect evidence that ASP in expressed during the course of HIV-1 infection. However, this does not indicate that ASP is a protein that plays a role in the virus lifecycle.

6. Expression and Functional Role of ASP in In Vitro Models

Relatively few studies have investigated the expression of ASP in infected cells, and its role in HIV-1 infection. Exploring these research areas has been hindered by several challenges. First, ASP is expressed at very low levels. The negative sense promoter (NSP) located in the HIV-1 3′LTR lacks a TATA box, is Tat-independent, and relies primarily on ubiquitous, housekeeping transcription factors [67–69]. Therefore, NSP drives the expression of antisense RNA at low levels. In addition, HIV-1 antisense transcripts are not efficiently exported into the cytoplasm, possibly due to inefficient polyadenylation. All these factors contribute to ASP being expressed in low amounts. Second, reagents needed to study ASP expression are lacking. Antibodies directed against ASP are still not commercially available. In addition, ASP is poorly immunogenic, which has made it difficult to raise potent and specific antibodies. Recently, two anti-ASP monoclonal antibodies have been reported: one is directed against a 16-aa epitope located immediately before the first predicted transmembrane domain [70], the other directed against a 13-aa epitope located between the two predicted transmembrane domains in the extracellular loop of ASP (Figure 4) [57]. However, perhaps what makes ASP especially difficult to study is, at the same time, what makes it so special and peculiar: the fact that it is a gene created de novo, and that it is entirely encased within the *env* gene. Indeed, ASP does not have any known orthologs, paralogs or xenologs [71]. Therefore, we do not have any information from other viruses that might enlighten us about a possible role for ASP in HIV-1 replication, and guide us in developing hypotheses. In addition, functional studies involving mutagenesis of ASP (e.g., deletion or substitution of large sequences or domains) would also affect the *env* sequence on the opposite DNA strand of the genome. Thus, while 'work-around' solutions are possible, the genomic location of the ASP gene makes it quite challenging to study its function.
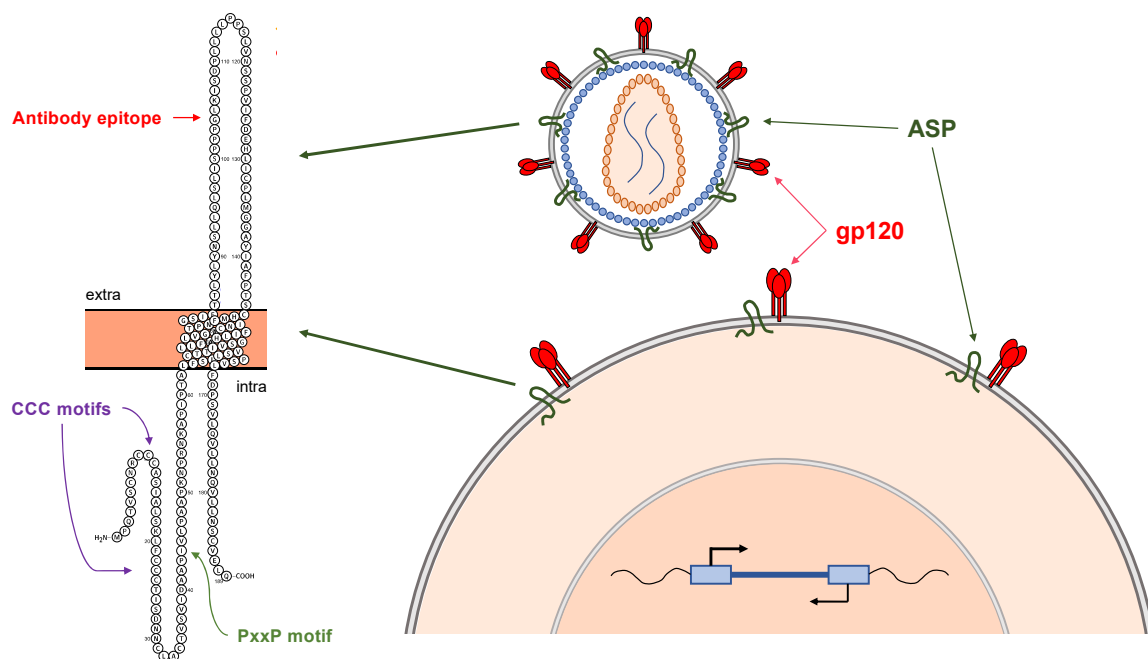
Figure 4. Association of ASP with the plasma membrane and the viral envelope. Studies from our and other groups have shown that ASP is expressed on the surface of productively infected cells. In addition, our studies demonstrate that ASP co-localizes with gp120 on the surface of primary HIV-infected CD4+ T cells and macrophages. A putative model of membrane-associated ASP is shown on the left side of the figure. According to this model, the N-terminus and C-terminus of ASP are intracellular (or intraviral), while the segment between the two transmembrane domains is exposed in the extracellular (or extra-viral) milieu. This model is supported by evidence that our monoclonal antibody directed against an epitope in the predicted extracellular loop of the protein (residues highlighted in red) can detect ASP on the cell surface in flow cytometry and confocal microscopy studies without the need from membrane permeabilization [57]. The same monoclonal antibody could detect ASP on the surface of viral particles (fluorescence correlation spectroscopy), and could capture HIV-1 virions (virion capture assay) [57].

Despite these limitations, a few publications have reported the expression of ASP in various cellular systems, and have made progress in elucidating its (possible) role in HIV-1 infection. In an early study, Briquet and Vaquero studied the sub-cellular localization of ASP in multiple in vitro cell systems, namely A3.01 cells over-expressing a fusion VSV-ASP protein, chronically infected ACH-2 cells, and SupT1 transfected with a molecular clone expressing HIV-1$_{HXB2}$ (mimicking acute infection) [72]. For detection of ASP, the authors used electron microscopy after immunostaining with anti-VSV antibodies (for VSV-ASP fusion protein), or rabbit antiserum directed against two N-terminal and one C-terminal peptides of ASP (for chronically infected ACH-2 and acutely infected SupT1 cells). In A3.01 cells over-expressing VSV-ASP, the authors found ASP associated with multiple membrane systems (plasma, mitochondrial and nuclear membranes). In chronically infected ACH-2 cells, ASP was expressed at very low levels in unstimulated cells, but after activation ASP became readily detectable both in the nucleus and in the cytoplasm. A similar expression pattern was visible also in acutely infected SupT1 cells [72]. The sub-cellular localization of ASP was also the focus of a later study from the Mesnard and Barbeau groups [58]. In their 2011 report, the authors used anti-FLAG antibodies in confocal microscopy to investigate ASP localization in Jurkat cells transfected with a construct expressing Flag-tagged ASP. These studies showed that ASP localized to the plasma membrane (with both polarized and unpolarized expression patterns) as well as with cell surface protrusion [58]. A subsequent report by the same groups confirmed these results in primary monocyte-derived macrophages infected with VSV-pseudotyped HIV-1 expressing Flag-tagged ASP [73]. More recently, our group reported a more in-depth analysis of ASP expression in several chronically infected lymphoid and myeloid cell lines and also in acutely infected primary

human CD4+ T cells and monocyte-derived macrophages (MDM) [57]. For these studies, we developed a mouse monoclonal antibody directed against a peptide that maps in the putative extracellular domain of ASP (Figure 4). Flow cytometry and confocal microscopy studies showed that ASP localizes within the nucleus of all unstimulated, non-productively infected lymphoid and myeloid cell lines. Interestingly, ASP displayed a polarized nuclear distribution, and an accumulation in regions of the nucleus that contain actively transcribed chromatin [57]. Following cell stimulation and reactivation of productive infection, we found that ASP translocates to the cytoplasm, it associates with the plasma membrane (Figure 4), and thus becomes detectable with our monoclonal antibody without cell perme-abilization, proving that this antibody recognizes an epitope exposed in the extracellular milieu. Further, our studies demonstrate that once present on the plasma membrane, a large fraction of ASP localizes in close proximity of the HIV-1 envelope glycoprotein gp120 (Figure 4). These results were confirmed by super resolution microscopy and also in acutely infected primary CD4+ T cells and MDM [57]. Expression of ASP in close proximity of gp120 on the plasma membrane suggested that ASP may be also present on the enve-lope of viral particles upon budding and release from infected cells, which we were able to demonstrate by both fluorescence correlation spectroscopy and virion capture assay (Figure 4) [57].

While the expression pattern and sub-cellular localization of ASP have been addressed in several studies that yielded largely consistent results, the role of ASP in viral replication has been explored to a much lesser degree. This is in part due to the lack of ASP homologs and the fact that *asp* and *env* are overprinting genes. To assess a potential role of ASP in viral replication, Clerc et al. substituted a single nucleotide within the twelfth codon of the ASP gene in the infectious HIV-1 molecular clone NL4-3 strain, which replaced a cysteine codon with a *stop* codon (TGC > TGA), thus resulting in the early termination of ASP [58]. At the same time, this substitution resulted in a synonymous mutation in *env* on the opposite strand. Infection of Jurkat cells with equal amounts of the wild type and ASP-deficient HIV-1$_{NL4-3}$ did not show any difference in the replication dynamics of the two viruses over 14 days. However, this negative result could be due to various reasons. For example, as discussed above, our studies showed that ASP is present in the HIV-1 viral envelope and is exposed to the 'extra-viral' milieu [57]. This might suggest a possible accessory role of ASP during the early steps of HIV-1 infection (e.g., attachment, fusion, or entry). In the study by Clerc and colleagues that sought to evaluate differences in the replication rates of wildtype vs. ASP-deficient HIV-1, the use of a cell line model expressing high cell surface levels of CD4 and CXCR4 might have allowed optimal infection efficiency even in the absence of ASP. Moreover, T cell lines offer sustained transcription levels, which might contribute to minimize the effects caused by the loss of an 'accessory' gene. Therefore, it is possible that the experimental system may have contributed to hide small but appreciable differences in the replication efficiency between wildtype and ASP-deficient HIV-1.

Two studies from the Barbeau group provided evidence that ASP expression might be responsible for the induction of autophagy in infected cells. In the first report, the authors found that codon optimization improved ASP expression levels and facilitated its detection after transient transfection [74]. Further, this study showed that ASP forms high molecular weight aggregates that are stable even in the presence of detergents and reducing agents. More interestingly, this report demonstrated that overexpression of ASP induces autophagy, as shown by a punctate distribution of ASP reminiscent of autophagosomes, and by induction of autophagy markers LC3b-II and Beclin 1. Moreover, inhibition of autophagy with 3-methyladenine resulted in increased levels of ASP, suggesting that it may be degraded during the last step of the autophagy process [74]. In a more recent study, Liu et al. reported that ASP is ubiquitinated, and that it promotes autophagy by interacting with the host factor, p62 [70]. In addition, the authors demonstrated that ASP from all HIV-1 clades induces autophagy [70]. Both DNA and RNA viruses utilize autophagy to their advantage during their replication cycle [75]. In the case of HIV-1, autophagy is required

for processing of GAG during infection of macrophages, and it significantly increases viral production [76].

These findings are very intriguing and point to two potential roles of ASP in HIV-1 replication. When present on the viral surface, ASP might be involved in virus entry, whereas when expressed within infected cells, it might promote autophagy. These roles are not mutually exclusive, because they occur at different stages of the virus life cycle. An early study did not show a reduction in viral replication when the ASP ORF was disrupted [58], but this may be due to limitations of the system that was used. Indeed, the role of ASP in HIV-1 replication may be small when observed in the limited setting of certain in vitro models that monitor a few cycles of viral replication, and yet significant when observed in the context of the infection in vivo and/or at the population level. Therefore, future studies will require a more careful evaluation of the model system used to determine whether ASP plays a role HIV-1 replication.

7. Origin, Conservation and Evolution of ASP

An important and still unanswered question that may help shed light onto the role, if any, that ASP plays in HIV-1 replication, spread and pathogenesis is '*where does ASP come from?*'.

This question was the focus of a 2016 study published by Cassan and colleagues [71]. Here, the authors analyzed ~23,000 HIV-1 (groups M, N, O and P), SIVcpz, and SIVgor *env* sequences obtained from ~3900 infected humans and nonhuman primates. These sequences were aligned to the reference HIV-1$_{HXB2}$ sequence, which contains an ASP ORF of 189 codons between nucleotide positions 1717 and 1151 of the *env* gene in the −2 frame. To identify a similar ORF in the ~23,000 sequences being analyzed, the authors searched for the presence of an open reading frame of >150 codons (identified by canonical *start* and *stop* codons) located between the same two reference positions in *env* indicated above. Using these criteria, they found that 77% of the sequences from HIV-1 group M—which is responsible for the world pandemic—contain an ORF with a median length of 182 codons. When looking at high prevalence subtypes A, B, C, G and CRF01_AE (combined prevalence 81%) and low prevalence subtypes D, F, J, H and K (combined prevalence 3%), the ASP ORF was present in 84% of the former vs. 45% of the latter. Moreover, the ASP ORF was found in less than 1.5% of HIV-1 sequences that belong to non-pandemic HIV-1 groups N, O and P (combined prevalence ~0.1%). Finally, an ASP ORF as defined based on the criteria outlined above was absent in the sequences of SIVcpz and SIVgor that were analyzed. However, it is interesting to note that the median length of the ASP ORF in SIV becomes longer as the corresponding viral strain approaches phylogenetically to HIV-1 group M (66 codons for SIVcpz_Pts and 125 codons for SIVcpz_Ptt). Therefore, these studies showed that an ASP ORF of >150 aa is present almost exclusively in the pandemic HIV-1 group M, and much less frequently in non-pandemic HIV-1 groups O, N and P. In addition, the study found a statistically significant positive correlation between the frequency with which the ASP ORF was present in each HIV-1 subtype and the prevalence of the subtype, suggesting that ASP might be involved in virus spread. The authors also noted that, while 16% of HIV-1 subtypes A, B, C, G and CRF01_AE lack the ASP ORF, they also lack the *nef* gene at a similar frequency [77].

In the same report, Cassan and colleagues performed computer simulations to estimate the likelihood that the presence of ASP in 77% of HIV-1 group M strains was due to chance [71]. They found that the likelihood of finding a 180-codon ORF in the -2 frame of a gene the same size as *env* or of a genome the same size as HXB2 were 3% and 19%, respectively—much lower than the actual frequency of group M sequences with the ASP ORF. They concluded that the presence of ASP at such high frequency was unlikely due to mere chance [71].

The evolution and selective pressure of *asp* were investigated in three different studies. The first report is the same study by Cassan et al. discussed above [71]. These authors found that the canonical *start* codon of the ASP ORF is conserved in 97% of the HIV-1

group M sequences. However, conservation of the ASP *start* codon at such high frequency was not imposed by the sequence of *env* on the opposite strand, because point mutations are possible that would cause loss of the *start* codon for *asp* while producing a synonymous codon in *env* [71]. The authors made a similar observation in regard to *stop* codons: there are 11 sites along the ASP ORF that could potentially mutate and create a *stop* without affecting the amino acid sequence of ENV. While seven of these sites are actual *stop* codons in sequences of HIV-1 groups N, O and P, they are *stop*s in only 0.5% of the group M sequences. Moreover, all sequences belonging to HIV-1 group M subtype A contain an early *stop* codon located 12 codons downstream of the *start* codon in *asp*. However, 90% of these sequences contain an alternative *start* codon located 17 codons downstream of the early *stop* codon [71]. Therefore, these results show a selective pressure to conserve an intact ASP ORF by maintaining a *start* codon (or by creating it a new one when lost, as in the case of subtype A sequences), and by avoiding early *stop* codons [71].

Two subsequent reports analyzed co-evolution of the ASP and ENV open reading frames. Dimonte studied the correlation between mutations in ASP and ENV, and CCR5 vs. CXCR4 coreceptor usage [78]. The author first used the G2P algorithm to predict the tropism of ~24,000 HIV-1 subtype B *env* sequences deposited in the Los Alamos HIV Sequence Database, and then analyzed the association of mutations in ASP with predicted coreceptor usage. He found that mutations in 36 amino acids of ASP are associated with preferential CCR5 or CXCR4 usage. A number of these mutations localize in the loop between the two transmembrane domains of ASP, which we have shown to be exposed in the extracellular or 'extra-viral' environment, thus suggesting a possible involvement of ASP in cell entry [57]. Next, Dimonte investigated the association between mutations in ASP and in the V3 loop of gp120, and coreceptor usage. The gp120 V3 loop becomes exposed upon binding of gp120 with CD4, contacts the coreceptor (CCR5 or CXCR4), and mediated virus entry into the target cell [79]. The author found eight sites on ASP where specific mutations are associated in a statistically significant manner with mutations in the V3 loop of gp120 and with CCR5/CXCR4 tropism; six of these eight mutations map in the extracellular/viral loop of ASP [78]. It should be underscored that the ASP ORF overlaps the *env* gene in the region of the V4 and V5 loops, but not V3. Therefore, association of specific residues in ASP and in the V3 loops of ENV is not genetically linked, but possibly functionally linked.

Finally, a recent report used a new computer algorithm called OLGenie to study co-evolution of the *env* and *asp* genes [80]. The ratio of nonsynonymous and synonymous mutations (i.e., nucleotide substitutions that alter or do not alter, respectively, the amino acid sequence of the protein product) within a gene is represented as d_N/d_S, and is an indicator of selective pressure. In protein-coding genes, synonymous mutations typically exceed nonsynonymous mutations ($d_N/d_S < 1$), which are often deleterious (especially in functionally-constrained regions of the gene), and therefore are removed by negative or 'purifying' selection. However, when nonsynonymous mutations provide a selective advantage, they will become fixed in the population by positive selection. In the special case of overlapping genes—such as in the case of *env* and *asp*—synonymous mutations in one of the two ORFs could be nonsynonymous in the other, and this may impact the d_N/d_S value. The OLGenie algorithm was developed to study evolution of overlapping genes [80]. When applied to study *asp*, OLGenie found that this gene is under intense purifying selection ($d_N/d_S = 0.29$). This study strongly suggests that mutations occurring within the ASP ORF are disproportionately synonymous, indicating an evolutionary effort to maintain the amino acid sequence of ASP [80].

Altogether, the studies described above demonstrate that the ASP ORF is present almost exclusively in pandemic HIV-1 subtypes, and that the percentage of viral strains with a full-length ASP within each HIV-1 subtype correlates with the worldwide prevalence of the subtype. It also appears that a significant evolutionary effort has been invested toward the maintenance of an intact ORF (preservation of *start* codon and avoidance of *stop* codons) and the conservation of the amino acid sequence (low d_N/d_S ratio). Finally, there is

evidence of a correlation between certain ASP variants and viral tropism. When considering the small size and the complexity of the HIV-1 genome, and the intense immune pressure acting on the virus, it would seem highly unlikely that the *asp* gene appeared by chance, and that it would be conserved if it did not encode for a protein product, or that such a protein did not provide any selective advantage to the virus.

8. Concluding Remarks

Since its identification more than three decades ago, there have been very few studies on the antisense protein of HIV-1. That is surprising when considering the strong association between the presence of a full-length ASP ORF and HIV-1 subtype prevalence, and also how conserved the ASP sequence is among pandemic HIV-1 strains. Since the antisense protein HBZ has been proven to play a crucial role in the pathogenesis of HTLV-1, it is equally reasonable to hypothesize that its HIV-1 counterpart, ASP, plays an important role in the HIV-1 pandemic.

The role of ASP in HIV-1 replication, spread and pathogenesis remains largely unknown. However, the fact that the ASP ORF is present only in pandemic HIV-1 strains, and that its conservation and length correlate with subtype prevalence [71], along with the evidence that the *asp* gene presents synonymous mutations at a disproportionately high rate [80] strongly suggest that ASP is involved in the virus life cycle. Studies from our and other groups have begun to make some progress in that direction, and these lines of investigation are likely to continue and to produce valuable information. Achieving a complete understanding of the role that ASP has played in inter- and intra-species spread, and that it continues to play in intra-host spread will require new efforts, new research tools, and—most importantly—new and original hypotheses.

Author Contributions: Conceptualization: F.R.; writing—original draft preparation: F.R., Z.G., M.S.I., R.L.; writing—review and editing: F.R., Z.G., M.S.I., R.L.; visualization: F.R., Z.G., M.S.I., R.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Long, M.; Betran, E.; Thornton, K.; Wang, W. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **2003**, *4*, 865–875. [CrossRef] [PubMed]
2. Jacob, F. Evolution and tinkering. *Science* **1977**, *196*, 1161–1166. [CrossRef] [PubMed]
3. Sorek, R. The birth of new exons: Mechanisms and evolutionary consequences. *Rna* **2007**, *13*, 1603–1608. [CrossRef] [PubMed]
4. Li, C.Y.; Zhang, Y.; Wang, Z.; Zhang, Y.; Cao, C.; Zhang, P.W.; Lu, S.J.; Li, X.M.; Yu, Q.; Zheng, X.; et al. A human-specific de novo protein-coding gene associated with human brain functions. *PLoS Comput. Biol.* **2010**, *6*, e1000734. [CrossRef]
5. Rancurel, C.; Khosravi, M.; Dunker, A.K.; Romero, P.R.; Karlin, D. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J. Virol.* **2009**, *83*, 10719–10736. [CrossRef]
6. Sabath, N.; Wagner, A.; Karlin, D. Evolution of viral proteins originated de novo by overprinting. *Mol. Biol. Evol.* **2012**, *29*, 3767–3780. [CrossRef]
7. Fischer, D.; Eisenberg, D. Finding families for genomic ORFans. *Bioinformatics* **1999**, *15*, 759–762. [CrossRef]
8. Wilson, G.A.; Bertrand, N.; Patel, Y.; Hughes, J.B.; Feil, E.J.; Field, D. Orphans as taxonomically restricted and ecologically important genes. *Microbiology* **2005**, *151*, 2499–2501. [CrossRef]
9. Yooseph, S.; Sutton, G.; Rusch, D.B.; Halpern, A.L.; Williamson, S.J.; Remington, K.; Eisen, J.A.; Heidelberg, K.B.; Manning, G.; Li, W.; et al. The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol.* **2007**, *5*, e16. [CrossRef] [PubMed]
10. Delaye, L.; Deluna, A.; Lazcano, A.; Becerra, A. The origin of a novel gene through overprinting in Escherichia coli. *BMC Evol. Biol.* **2008**, *8*, 31. [CrossRef]

11. Ribrioux, S.; Brungger, A.; Baumgarten, B.; Seuwen, K.; John, M.R. Bioinformatics prediction of overlapping frameshifted translation products in mammalian transcripts. *BMC Genom.* **2008**, *9*, 122. [CrossRef]
12. Chung, W.Y.; Wadhawan, S.; Szklarczyk, R.; Pond, S.K.; Nekrutenko, A. A first look at ARFome: Dual-coding genes in mammalian genomes. *PLoS Comput. Biol.* **2007**, *3*, e91. [CrossRef]
13. McVeigh, A.; Fasano, A.; Scott, D.A.; Jelacic, S.; Moseley, S.L.; Robertson, D.C.; Savarino, S.J. IS1414, an Escherichia coli insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene. *Infect. Immun.* **2000**, *68*, 5710–5715. [CrossRef] [PubMed]
14. Michel, A.M.; Choudhury, K.R.; Firth, A.E.; Ingolia, N.T.; Atkins, J.F.; Baranov, P.V. Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* **2012**, *22*, 2219–2229. [CrossRef] [PubMed]
15. Bergeron, D.; Lapointe, C.; Bissonnette, C.; Tremblay, G.; Motard, J.; Roucou, X. An out-of-frame overlapping reading frame in the ataxin-1 coding sequence encodes a novel ataxin-1 interacting protein. *J. Biol. Chem.* **2013**, *288*, 21824–21835. [CrossRef] [PubMed]
16. Vanderperre, B.; Lucier, J.F.; Bissonnette, C.; Motard, J.; Tremblay, G.; Vanderperre, S.; Wisztorski, M.; Salzet, M.; Boisvert, F.M.; Roucou, X. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS ONE* **2013**, *8*, e70698. [CrossRef] [PubMed]
17. Fellner, L.; Simon, S.; Scherling, C.; Witting, M.; Schober, S.; Polte, C.; Schmitt-Kopplin, P.; Keim, D.A.; Scherer, S.; Neuhaus, K. Evidence for the recent origin of a bacterial protein-coding, overlapping orphan gene by evolutionary overprinting. *BMC Evol. Biol.* **2015**, *15*, 283. [CrossRef]
18. Zhou, Q.; Zhang, G.; Zhang, Y.; Xu, S.; Zhao, R.; Zhan, Z.; Li, X.; Ding, Y.; Yang, S.; Wang, W. On the origin of new genes in Drosophila. *Genome Res.* **2008**, *18*, 1446–1455. [CrossRef]
19. Toll-Riera, M.; Bosch, N.; Bellora, N.; Castelo, R.; Armengol, L.; Estivill, X.; Alba, M.M. Origin of primate orphan genes: A comparative genomics approach. *Mol. Biol. Evol.* **2009**, *26*, 603–612. [CrossRef]
20. Ekman, D.; Elofsson, A. Identifying and quantifying orphan protein sequences in fungi. *J. Mol. Biol.* **2010**, *396*, 396–405. [CrossRef]
21. Belshaw, R.; Pybus, O.G.; Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* **2007**, *17*, 1496–1504. [CrossRef]
22. Chirico, N.; Vianelli, A.; Belshaw, R. Why genes overlap in viruses. *Proc. Biol. Sci.* **2010**, *277*, 3809–3817. [CrossRef] [PubMed]
23. Barrell, B.G.; Air, G.M.; Hutchison, C.A., 3rd. Overlapping genes in bacteriophage phiX174. *Nature* **1976**, *264*, 34–41. [CrossRef] [PubMed]
24. Pavesi, A. New insights into the evolutionary features of viral overlapping genes by discriminant analysis. *Virology* **2020**, *546*, 51–66. [CrossRef] [PubMed]
25. Pavesi, A.; De Iaco, B.; Granero, M.I.; Porati, A. On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. *J. Mol. Evol.* **1997**, *44*, 625–631. [CrossRef]
26. Pavesi, A.; Magiorkinis, G.; Karlin, D.G. Viral proteins originated de novo by overprinting can be identified by codon usage: Application to the "gene nursery" of Deltaretroviruses. *PLoS Comput. Biol.* **2013**, *9*, e1003162. [CrossRef]
27. Pavesi, A.; Vianelli, A.; Chirico, N.; Bao, Y.; Blinkova, O.; Belshaw, R.; Firth, A.; Karlin, D. Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS ONE* **2018**, *13*, e0202513. [CrossRef]
28. Pavesi, A. Asymmetric evolution in viral overlapping genes is a source of selective protein adaptation. *Virology* **2019**, *532*, 39–47. [CrossRef]
29. Carter, C.W., Jr. Simultaneous codon usage, the origin of the proteome, and the emergence of de-novo proteins. *Curr. Opin. Struct. Biol.* **2021**, *68*, 142–148. [CrossRef]
30. Karlin, D.; Ferron, F.; Canard, B.; Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* **2003**, *84*, 3239–3252. [CrossRef]
31. Willis, S.; Masel, J. Gene Birth Contributes to Structural Disorder Encoded by Overlapping Genes. *Genetics* **2018**, *210*, 303–313. [CrossRef]
32. Wensman, J.J.; Munir, M.; Thaduri, S.; Hornaeus, K.; Rizwan, M.; Blomstrom, A.L.; Briese, T.; Lipkin, W.I.; Berg, M. The X proteins of bornaviruses interfere with type I interferon signalling. *J. Gen. Virol.* **2013**, *94*, 263–269. [CrossRef]
33. Van Knippenberg, I.; Carlton-Smith, C.; Elliott, R.M. The N-terminus of Bunyamwera orthobunyavirus NSs protein is essential for interferon antagonism. *J. Gen. Virol.* **2010**, *91*, 2002–2006. [CrossRef]
34. Li, F.; Ding, S.W. Virus counterdefense: Diverse strategies for evading the RNA-silencing immunity. *Annu. Rev. Microbiol.* **2006**, *60*, 503–531. [CrossRef]
35. Chellappan, P.; Vanitharani, R.; Fauquet, C.M. MicroRNA-binding viral protein interferes with Arabidopsis development. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10381–10386. [CrossRef]
36. Chen, W.; Calvo, P.A.; Malide, D.; Gibbs, J.; Schubert, U.; Bacik, I.; Basta, S.; O'Neill, R.; Schickli, J.; Palese, P.; et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med.* **2001**, *7*, 1306–1312. [CrossRef]
37. Boehme, K.W.; Hammer, K.; Tollefson, W.C.; Konopka-Anstadt, J.L.; Kobayashi, T.; Dermody, T.S. Nonstructural protein sigma1s mediates reovirus-induced cell cycle arrest and apoptosis. *J. Virol.* **2013**, *87*, 12967–12979. [CrossRef]
38. Taliansky, M.; Roberts, I.M.; Kalinina, N.; Ryabov, E.V.; Raj, S.K.; Robinson, D.J.; Oparka, K.J. An umbraviral protein, involved in long-distance RNA movement, binds viral RNA and forms unique, protective ribonucleoprotein complexes. *J. Virol.* **2003**, *77*, 3031–3040. [CrossRef]
39. Singh, M. No vaccine against HIV yet–are we not perfectly equipped? *Virol. J.* **2006**, *3*, 60. [CrossRef]

40. Pluta, A.; Jaworski, J.P.; Douville, R.N. Regulation of Expression and Latency in BLV and HTLV. *Viruses* **2020**, *12*, 1079. [CrossRef]
41. Matsuoka, M.; Mesnard, J.M. HTLV-1 bZIP factor: The key viral gene for pathogenesis. *Retrovirology* **2020**, *17*, 2. [CrossRef] [PubMed]
42. Satou, Y.; Yasunaga, J.; Yoshida, M.; Matsuoka, M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 720–725. [CrossRef] [PubMed]
43. Yoshida, M.; Satou, Y.; Yasunaga, J.; Fujisawa, J.; Matsuoka, M. Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene. *J. Virol.* **2008**, *82*, 9359–9368. [CrossRef] [PubMed]
44. Laverdure, S.; Polakowski, N.; Hoang, K.; Lemasson, I. Permissive Sense and Antisense Transcription from the 5′ and 3′ Long Terminal Repeats of Human T-Cell Leukemia Virus Type 1. *J. Virol.* **2016**, *90*, 3600–3610. [CrossRef] [PubMed]
45. Clerc, I.; Polakowski, N.; Andre-Arpin, C.; Cook, P.; Barbeau, B.; Mesnard, J.M.; Lemasson, I. An interaction between the human T cell leukemia virus type 1 basic leucine zipper factor (HBZ) and the KIX domain of p300/CBP contributes to the down-regulation of tax-dependent viral transcription by HBZ. *J. Biol. Chem.* **2008**, *283*, 23903–23913. [CrossRef]
46. Miyazaki, M.; Yasunaga, J.; Taniguchi, Y.; Tamiya, S.; Nakahata, T.; Matsuoka, M. Preferential selection of human T-cell leukemia virus type 1 provirus lacking the 5′ long terminal repeat during oncogenesis. *J. Virol.* **2007**, *81*, 5714–5723. [CrossRef]
47. Taniguchi, Y.; Nosaka, K.; Yasunaga, J.; Maeda, M.; Mueller, N.; Okayama, A.; Matsuoka, M. Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. *Retrovirology* **2005**, *2*, 64. [CrossRef]
48. Fan, J.; Ma, G.; Nosaka, K.; Tanabe, J.; Satou, Y.; Koito, A.; Wain-Hobson, S.; Vartanian, J.P.; Matsuoka, M. APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo. *J. Virol.* **2010**, *84*, 7278–7287. [CrossRef]
49. Arnold, J.; Zimmerman, B.; Li, M.; Lairmore, M.D.; Green, P.L. Human T-cell leukemia virus type-1 antisense-encoded gene, Hbz, promotes T-lymphocyte proliferation. *Blood* **2008**, *112*, 3788–3797. [CrossRef]
50. Satou, Y.; Yasunaga, J.; Zhao, T.; Yoshida, M.; Miyazato, P.; Takai, K.; Shimizu, K.; Ohshima, K.; Green, P.L.; Ohkura, N.; et al. HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. *PLoS Pathog.* **2011**, *7*, e1001274. [CrossRef]
51. Zhao, T.; Satou, Y.; Matsuoka, M. Development of T cell lymphoma in HTLV-1 bZIP factor and Tax double transgenic mice. *Arch. Virol.* **2014**, *159*, 1849–1856. [CrossRef]
52. Savoret, J.; Mesnard, J.M.; Gross, A.; Chazal, N. Antisense Transcripts and Antisense Protein: A New Perspective on Human Immunodeficiency Virus Type 1. *Front. Microbiol.* **2020**, *11*, 625941. [CrossRef]
53. Miura, M.; Yasunaga, J.; Tanabe, J.; Sugata, K.; Zhao, T.; Ma, G.; Miyazato, P.; Ohshima, K.; Kaneko, A.; Watanabe, A.; et al. Characterization of simian T-cell leukemia virus type 1 in naturally infected Japanese macaques as a model of HTLV-1 infection. *Retrovirology* **2013**, *10*, 118. [CrossRef]
54. Briquet, S.; Richardson, J.; Vanhee-Brossollet, C.; Vaquero, C. Natural antisense transcripts are detected in different cell lines and tissues of cats infected with feline immunodeficiency virus. *Gene* **2001**, *267*, 157–164. [CrossRef]
55. Miller, R.H. Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. *Science* **1988**, *239*, 1420–1422. [CrossRef]
56. Casino, A.; Cipollaro, M.; Guerrini, A.M.; Mastrocinque, G.; Spena, A.; Scarlato, V. Coding capacity of complementary DNA strands. *Nucleic Acids Res.* **1981**, *9*, 1499–1518. [CrossRef]
57. Affram, Y.; Zapata, J.C.; Gholizadeh, Z.; Tolbert, W.D.; Zhou, W.; Iglesias-Ussel, M.D.; Pazgier, M.; Ray, K.; Latinovic, O.S.; Romerio, F. The HIV-1 Antisense Protein ASP Is a Transmembrane Protein of the Cell Surface and an Integral Protein of the Viral Envelope. *J. Virol.* **2019**, *93*. [CrossRef]
58. Clerc, I.; Laverdure, S.; Torresilla, C.; Landry, S.; Borel, S.; Vargas, A.; Arpin-Andre, C.; Gay, B.; Briant, L.; Gross, A.; et al. Polarized expression of the membrane ASP protein derived from HIV-1 antisense transcription in T cells. *Retrovirology* **2011**, *8*, 74. [CrossRef] [PubMed]
59. Saksela, K.; Cheng, G.; Baltimore, D. Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef+ viruses but not for down-regulation of CD4. *EMBO J.* **1995**, *14*, 484–491. [CrossRef]
60. Zapata, J.C.; Campilongo, F.; Barclay, R.A.; DeMarino, C.; Iglesias-Ussel, M.D.; Kashanchi, F.; Romerio, F. 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. [CrossRef]
61. Mancarella, A.; Procopio, F.A.; Achsel, T.; De Crignis, E.; Foley, B.T.; Corradin, G.; Bagni, C.; Pantaleo, G.; Graziosi, C. Detection of antisense protein (ASP) RNA transcripts in individuals infected with human immunodeficiency virus type 1 (HIV-1). *J. Gen. Virol.* **2019**, *100*, 863–876. [CrossRef] [PubMed]
62. Vanhee-Brossollet, C.; Thoreau, H.; Serpente, N.; D'Auriol, L.; Levy, J.P.; Vaquero, C. A natural antisense RNA derived from the HIV-1 *env* gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. *Virology* **1995**, *206*, 196–202. [CrossRef]
63. Savoret, J.; Chazal, N.; Moles, J.P.; Tuaillon, E.; Boufassa, F.; Meyer, L.; Lecuroux, C.; Lambotte, O.; Van De Perre, P.; Mesnard, J.M.; et al. A Pilot Study of the Humoral Response against the AntiSense Protein (ASP) in HIV-1-Infected Patients. *Front. Microbiol.* **2020**, *11*, 20. [CrossRef]
64. Champiat, S.; Raposo, R.A.; Maness, N.J.; Lehman, J.L.; Purtell, S.E.; Hasenkrug, A.M.; Miller, J.C.; Dean, H.; Koff, W.C.; Hong, M.A.; et al. Influence of HAART on alternative reading frame immune responses over the course of HIV-1 infection. *PLoS ONE* **2012**, *7*, e39311. [CrossRef]

65. Bet, A.; Maze, E.A.; Bansal, A.; Sterrett, S.; Gross, A.; Graff-Dubois, S.; Samri, A.; Guihot, A.; Katlama, C.; Theodorou, I.; et al. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* **2015**, *12*, 15. [CrossRef]

66. Berger, C.T.; Llano, A.; Carlson, J.M.; Brumme, Z.L.; Brockman, M.A.; Cedeno, S.; Harrigan, P.R.; Kaufmann, D.E.; Heckerman, D.; Meyerhans, A.; et al. Immune screening identifies novel T cell targets encoded by antisense reading frames of HIV-1. *J. Virol.* **2015**, *89*, 4015–4019. [CrossRef]

67. Michael, N.L.; Vahey, M.T.; d'Arcy, L.; Ehrenberg, P.K.; Mosca, J.D.; Rappaport, J.; Redfield, R.R. 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. [CrossRef]

68. Bentley, K.; Deacon, N.; Sonza, S.; Zeichner, S.; Churchill, M. Mutational analysis of the HIV-1 LTR as a promoter of negative sense transcription. *Arch. Virol.* **2004**, *149*, 2277–2294. [CrossRef]

69. Kobayashi-Ishihara, M.; Yamagishi, M.; Hara, T.; Matsuda, Y.; Takahashi, R.; Miyake, A.; Nakano, K.; Yamochi, T.; Ishida, T.; Watanabe, T. HIV-1-encoded antisense RNA suppresses viral replication for a prolonged period. *Retrovirology* **2012**, *9*, 38. [CrossRef] [PubMed]

70. Liu, Z.; Torresilla, C.; Xiao, Y.; Nguyen, P.; Caté, C.; Barbosa, K.; Rassart, E.; Cen, S.; Bourgault, S.; Barbeaub, B. HIV-1 Antisense Protein of Different Clades Induces Autophagy and Associates with the Autophagy Factor p62. *J. Virol.* **2019**, *93*, e01757-18. [CrossRef]

71. Cassan, E.; Arigon-Chifolleau, A.M.; Mesnard, J.M.; 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. [CrossRef] [PubMed]

72. Briquet, S.; Vaquero, C. Immunolocalization studies of an antisense protein in HIV-1-infected cells and viral particles. *Virology* **2002**, *292*, 177–184. [CrossRef]

73. Laverdure, S.; Gross, A.; Arpin-Andre, C.; Clerc, I.; Beaumelle, B.; Barbeau, B.; Mesnard, J.M. HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells. *J. Virol.* **2012**, *86*, 13785–13789. [CrossRef] [PubMed]

74. Torresilla, C.; Larocque, E.; Landry, S.; Halin, M.; Coulombe, Y.; Masson, J.Y.; Mesnard, J.M.; Barbeau, B. Detection of the HIV-1 minus-strand-encoded antisense protein and its association with autophagy. *J. Virol.* **2013**, *87*, 5089–5105. [CrossRef] [PubMed]

75. Dreux, M.; Chisari, F.V. Viruses and the autophagy machinery. *Cell Cycle* **2010**, *9*, 1295–1307. [CrossRef] [PubMed]

76. Kyei, G.B.; Dinkins, C.; Davis, A.S.; Roberts, E.; Singh, S.B.; Dong, C.; Wu, L.; Kominami, E.; Ueno, T.; Yamamoto, A.; et al. Autophagy pathway intersects with HIV-1 biosynthesis and regulates viral yields in macrophages. *J. Cell Biol.* **2009**, *186*, 255–268. [CrossRef] [PubMed]

77. Pushker, R.; Jacque, J.M.; Shields, D.C. Meta-analysis to test the association of HIV-1 nef amino acid differences and deletions with disease progression. *J. Virol.* **2010**, *84*, 3644–3653. [CrossRef]

78. Dimonte, S. Different HIV-1 *env* frames: gp120 and ASP (antisense protein) biosynthesis, and theirs co-variation tropic amino acid signatures in X4- and R5-viruses. *J. Med. Virol.* **2017**, *89*, 112–122. [CrossRef]

79. Bowder, D.; Hollingsead, H.; Durst, K.; Hu, D.; Wei, W.; Wiggins, J.; Medjahed, H.; Finzi, A.; Sodroski, J.; Xiang, S.H. Contribution of the gp120 V3 loop to envelope glycoprotein trimer stability in primate immunodeficiency viruses. *Virology* **2018**, *521*, 158–168. [CrossRef] [PubMed]

80. Nelson, C.W.; Ardern, Z.; Wei, X. OLGenie: Estimating Natural Selection to Predict Functional Overlapping Genes. *Mol. Biol. Evol.* **2020**, *37*, 2440–2449. [CrossRef]