*Published online 6 December 2009* *Nucleic Acids Research, 2010, Vol. 38, No. 4* ***1367–1381***
*doi:10.1093/nar/gkp1109*

# A new type of IRES within *gag* coding region recruits three initiation complexes on HIV-2 genomic RNA

**Laure Weill[1], Laurie James[1], Nathalie Ulryck[1], Nathalie Chamond[1], Cecile H. Herbreteau[2], Theophile Ohlmann[2] and Bruno Sargueil[1,*]**

[1]CNRS UMR 8015, Laboratoire de cristallographie et RMN Biologique, Université Paris Descartes, 4 avenue de l'Observatoire, 75270 Paris Cedex 06 and [2]INSERM U758, Ecole Normale Supérieure de Lyon, Unité de Virologie Humaine, IFR 128, Lyon, F-69364 France

Received August 24, 2009; Revised November 10, 2009; Accepted November 11, 2009

## ABSTRACT

**Genomic RNA of primate lentiviruses serves both as an mRNA that encodes Gag and Gag-Pol polyproteins and as a propagated genome. Translation of this RNA is initiated by standard cap dependant mechanism or by internal entry of the ribosome. Two regions of the genomic RNA are able to attract initiation complexes, the 5′ untranslated region and the *gag* coding region itself. Relying on probing data and a phylogenetic study, we have modelled the secondary structure of HIV-1, HIV-2 and $SIV_{Mac}$ coding region. This approach brings to light conserved secondary-structure elements that were shown by mutations to be required for internal entry of the ribosome. No structural homologies with other described viral or cellular IRES can be identified and lentiviral IRESes show many peculiar properties. Most notably, the IRES present in HIV-2 *gag* coding region is endowed with the unique ability to recruit up to three initiation complexes on a single RNA molecule. The structural and functional properties of *gag* coding sequence define a new type of IRES. Although its precise role is unknown, the conservation of the IRES among fast evolving lentiviruses suggests an important physiological role.**

## INTRODUCTION

Primate lentivirus genomic RNA (gRNA) plays a dual role as an mRNA template to produce the Gag and the Gag-Pol polyproteins, and as a genome encapsidated into virions (1). Whether those two functions are mutually exclusive or if instead one requires the other depends on the virus considered. HIV-1 gRNA can fulfil both functions interchangeably and independently, while HIV-2 gRNA needs to be translated to be encapsidated (2,3). The Gag polyprotein thus produced is matured by the viral protease in four main proteins known as the matrix, the capsid, the nucleocapsid and p6 (4). Genomic RNAs, which are RNA polymerase II transcripts, are capped and could theoretically be translated following a cap dependent mechanism. However, the 5′ untranslated region (leader or 5′UTR) of lentivirus gRNA contains numerous conserved *cis*-acting elements required at various stages of the virus cycle. Those elements are conserved sequences and stable stem-loops involved in many phenomena, including transcription regulation, reverse transcription, encapsidation, non-covalent dimerisation, or splicing (5–9). Most of them are, at one stage or another, the site of RNA or protein interactions. Such characteristics are clearly detrimental for translation by a cap dependant mechanism since stable structures impede ribosome scanning through the 5′UTR to the initiation codon (10,11). Those long 5′UTRs do not contain any AUG triplet, but numerous alternative initiation codons such as CUG, UUG or GUG. For example,

*To whom correspondence should be addressed. Tel: +33 1 53 73 1568; Fax: +33 1 53 73 9925; Email: bruno.sargueil@parisdescartes.fr
Present address:
Laure Weill, Translational Control of Gene Expression group, Centre for Genomic Regulation (CRG), PRBB, C/Dr Aiguader, 88, 08003 Barcelona, Spain.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

HIV-1 and HIV-2 5′UTR conceal, respectively, 17 and 36 such triplets, among which three for HIV-1 and 11 for HIV-2 lie in a favourable 'Kozak' context (10,12,13). Albeit with a lower efficiency than the canonical AUG, such triplets can be recognised by the scanning complex and inhibit translation at the downstream *bona fide* initiation codon (14,15).

Despite the unfavourable characteristics presented by the 5′UTR, cap dependent translation of HIV-1 gRNA has been shown *in vivo* and *in vitro* (8,16–18), but seems less likely in the case of HIV-2 and $SIV_{MAC}$ (17,19).

Detailed investigations showed that initiation of Gag polyprotein translation could also be mediated by an Internal Ribosome Entry Site concealed in the 5′UTR of HIV-1 and $SIV_{MAC}$ (20–22). Interestingly enough, HIV-1 IRES is especially active during the G2/M phase of the cell cycle, when canonical cap dependent translation is inhibited (20). No IRES activity has been detected to date in HIV-2 gRNA 5′UTR. However, we previously reported that the HIV-2 *gag* coding region is endowed with an IRES activity that can recruit ribosomes on three in frame AUG, yielding three isoforms of the Gag polyprotein (23). The presence of an IRES within *gag* coding region, also demonstrated for HIV-1 and $SIV_{Mac}$ (21,24), confers to HIV-2 gRNA the potential to be efficiently translated as a leaderless RNA (23). Here we show that this property is conserved among primate lentiviruses. Using a combination of chemical/enzymatic probing and a phylogenetic analysis, we show that the *gag* open reading frame of the three viruses adopts a stable secondary structure. Mutations that destabilise conserved structural elements severely inhibit the IRES mediated translation, while the activity is retrieved with compensatory mutations that restore the pairing. Our results strongly suggest that the structure model is involved in recruiting the initiation complexes. Conservation of the IRES and its structure in such fast evolving viruses betrays the physiological importance of the observed phenomenon. Finally, we show that this HIV-2 IRES has the unique ability to attract up to three initiation complexes on a single RNA molecule. This work defines a new type of IRES, and paves the way to the molecular characterisation of IRESes present within the *gag* coding region of primate lentivirus.

## MATERIALS AND METHODS

### Constructions

RNAs were directly transcribed using T7 RNA polymerase from PCR products containing the T7 RNA polymerase promoter sequence, and purified as previously described (50,51). For the translation of full length Gag, the DNA sequence corresponding to the coding region of *gag* HIV-2 (pROD10), *gag* HIV-1 (pNL4.3) or *gag* SIVmac (mac251, 13 001) were amplified by PCR using a 3′ oligonucleotide starting at the stop codon of HIV-2 *gag*, HIV-1 *gag* or the SIV*gag* gene, and a 5′ oligonucleotide starting with the T7 promoter sequence and the 24 first nucleotides of the 5′UTR of each gene, or the 24 first nucleotides of *gag* open reading frame, to generate 5′UTR or $AUG_1$ constructs, respectively. To generate the truncated *gag* open reading frame, HIV-2 AUG1* and HIV-2 5′UTR*, the EcoR1–BamHI fragment of pcDNA3 HIV-2 Gag (23) including the whole 5′UTR and the 804 first nucleotides of *gag* open reading frame, were cloned into Puc19 digested with EcoRI and BamHI, yielding Puc19-gagHIV-2. In this context, a stop codon terminates the *gag* open reading frame three nucleotides downstream to the BamHI site. RNAs were transcribed from PCR products on this plasmid. RNA used for structure probing experiments were transcribed form PCR products on HIV-1 NL4.3 (nucleotides from +341 to +790 from the transcription site) or $SIV_{Mac}$ Mm251 (nucleotides from +537 to +952 from the transcription site) sequences, that yield RNAs starting with one extra guanosine upstream the initiation codon.

HIV-2-Gag mutated for $AUG_2$ and $AUG_3$ were transcribed form PCR products using as template the pcDNAHIV–2Gag–$AUG_1$–$CUC_2$–$CUC_3$ previously described (23). The stable stem-loop was introduced during PCR amplification using a primer starting with a T7 promoter top strand immediately followed by sequence of a stable stem-loop and then the sequences necessary to prime 14–16 nucleotides upstream from the initiation codon SL-HIV-2-AUG1: 5′-TAATACGACTCACTATAGGACCAGATCTACGCGTACGTACGCGTAGATCTG*TAGAAGATTGTGGGAGA**ATG**GGCGC-3′*, SL-HIV-2-5′UTR: 5′-TAATACGACTCACTATAGGACCAGATCTACGCGTACGTACGCGTAGATCTG*GTCGCTCTGCGGAGAGGCT-3′*. The nucleotides base paired in the stem loop are underlined, HIV-2 nucleotides are italicised, and the initiation codon is in bold.

The control RNA is a chimera containing the 5′UTR of the human β globin mRNA fused to the ßgal coding region (generous gift of E.Ricci). The RNA described in the text were transcribed from PCR products using the following oligonucleotides 5′ primers: T75′UTRglobine 5′-TAATACGACTCACTATAG*ACATTTGCTTCTGACACAACTGT-3′*, T7SL5′UTRglobine 5′-TAATACGACTCACTATAGGACAGATCTACGCGTACGTACGCGTAGATCTG*GACATTTGCTTCTGACAACTGT-3′*, T7AUG-control: 5′-TAATACGACTCACTATAGATGGTGCTAGCGGATCCCGTCG-3′, T7SL-AUG-control: 5′-TAATACGACTCACTATAG*GA*CAGATCTACGCGTACGTACGCGTAGATCTG*CTCAAACAGACACCA**ATG**GTGCTAG-3′*. the nucleotides base paired in the stem loop are underlined, 5′UTR globin or βGal nucleotides are italicised, and the initiation codon is in bold. For the four constructs we used the following oligonucleotide as a 3′ primer: BetagalW2402 5′-AGATTTGATTCAGCGATACAGCG-3′.

All mutants were obtained using the 'quick change site-directed mutagenesis' (stratagene).

### Chemical and enzymatic probing

The secondary structure of RNA HIV was probed using (DMS), N-cyclohexyl-N′-[N-methylmorpholino)-ethyl]-carbodiimid-4-toluolsulfonate (CMCT) and V1 RNAse (Ambion) as described previously (52,53). Ten picomoles of RNA were resuspended in 32 µl of 50 mM Hepes pH

7.5 (or 50 mM borate potassium pH 8 for CMCT), denatured for 2 min at 80°C. Then, 4 µl of 1.5 M KCl and 100 mM $MgCl_2$ [and 1 µl of tRNA (10 µg/µl) for RNase V1] was added and the mixture was slowly cooled to room temperature. DMS (25 mM final), CMCT (25 mM final) or RNAse V1 (0.1 or 0.01 U) was added and the mixture was incubated for 5 min (10 min for RNAse V1). The modification reaction was stopped in ice by addition of 10 µg of tRNA (DMS and CMCT), or 1 µl of SDS 1% (RNAse V1). The time and concentration of modification agent were established in order to have at the most one modification per molecule (at least 80% of RT full length product as compared to the unmodified RNA). The reaction was then immediately ethanol precipitated on dry ice in presence of 0.5 M ammonium acetate. RNA was then resuspended in 0.5 M ammonium acetate, ethanol precipitated, washed with 70% ethanol and resuspended in 5 µl of $H_2O$. Modification was revealed by reverse transcriptase using $^{32}$P-labelled primer and MMLV RNAse $H^-$ reverse transcriptase, according to the manufacturer's instructions (promega). To cover the whole sequence, we used the following primers: for SIV SIV1 5′-TTCCGCCGGGTCGTAGCC TAA-3′, SIV2 5′-CTGTTGGCACTAATGGAGCT-3′, SIV3: 5′-CACTTTCTCTTCTGCGTGAAT-3′, HIV6 5′-TGTTGCACTGGGTAATTTCCT-3′, for HIV-1: HIV-1-3′: 5′-AAACATGGGTATTACTTCTGG-3′ HIV-1-4 5′-CTGCTATTGTATTATATAATG-3′, HIV-1–5: 5′-T GGCTGTTGTTTCCTGTGTC-3′. Those products were separated on an 8 M urea 6% polyacrylamide electrophoresis gel. Radioactive products were revealed using a fluorescent screen and a storm scanner (GE healthcare). The intensity of each reverse transcriptase stops was determined using the ImageQuant 5.2 software (GE healthcare).

Secondary structure was modelled with Mfold, an energy minimisation software developed by Zuker, and using chemical probing data to provide constraints (34).

### Toeprinting assays

For toeprinting assays, 4.5 pmol of RNA were heated for 2 min at 80°C, slowly cooled to room temperature in 20 mM Tris pH 8, 0.5 mM $MgCl_2$ and 75 mM KCl for 10 min. RNA is incubated for 5 min at 30°C in 15 µl RRL (promega) in the presence of 0.5 mM magnesium acetate, 75 mM potassium acetate, 20 µM amino acid and 8 U RNasin (promega) in absence or in presence of 4.8 mM cycloheximide. The reaction was diluted to 40 µl final with 20 mM Tris pH 7.6, 2.5 mM magnesium acetate, 2 mM DTT and 0.25 mM spermidine before add 1 µl $^{32}$P-labelled primer. We used two primers for HIV-2, HIV-2-16: 5′-CCTGATTCTTTCTAATTCAT-3′ ($AUG_1$) and HIV-2-6: 5′-TGTTGCACTGGGTAATTTCCT-3′ ($AUG_2$ and $AUG_3$) and one for HIV-1: HIV-1-3: 5′-AG TTTATATTGTTTCTTTCCC-3′. The reaction was then incubated 5 min at 30°C. Primer extension was performed for 45 min at 30°C using 200 U of MMLV reverse transcriptase in the presence of 400 µM each of the four dNTP. The reaction was subsequently extracted with phenol, then with chloroform and was finally ethanol precipitated in presence of 0.5 M ammonium acetate and resuspended in 12 µl of loading buffer (95% formamide, 20 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue) including RNase A (10 µg/ml). The reaction was incubated for 15 min at 50°C, then 10 min at 95°C. Toeprint reactions were analysed on 6% denaturing polyacrylamide gels and signal quantification was performed using a STORM phosphorimager.

### Assembly and analysis of ribosomal complexes

Ribosomal complexes were assembled on $^{32}$P-labelled HIV-2 RNA. First, 20 pmol of RNA were heated for 2 min at 80°C, slowly cooled to room temperature in 20 mM Tris pH 8, 0.5 mM $MgCl_2$ and 75 mM KCl for 10 min. Rabbit reticulocyte lysate (FlexiR-RRL-Promega) was preatreated with either 4.8 mM cycloheximide or 1 mM GMPpNp for 10 min at 30°C. RNA was incubated at 30°C during 20 min in 170 µl or pretreated RRL in presence of 0.5 mM magnesium acetate, 75 mM potassium acetate, 20 µM amino acid and 8 U RNasin (promega). Reactions were stopped on ice, then were layered over 10–50% sucrose gradients (25 mM Tris pH 8, 6 mM $MgCl_2$, 75 mM KCl) and sedimented by ultracentrifugation at 39 000 r.p.m. in an SW41.1 rotor for 4 h at 4°C. Fractions were then collected and the amount of RNA in each fraction determined by scintillation counting.

### Translation assay

RNA (0.5 pmol) was incubated for 30 min at 30°C in a volume final of 10 µl containing 7 µl RRL (promega) in the presence of 0.5 mM magnesium acetate, 75 mM potassium acetate, 20 mM of each amino acid (except methionine) and 0.2 mCi/ml[$^{35}$S]methionine and 8 U Rnasin (Promega). The reaction was stopped with 90 µl of protein loading buffer. Ten microlitre of the reaction was loaded on a 12% SDS–PAGE gel. The electrophoresis were analysed and quantified using a storm phosphorimager (GE Healthcare). Each experiment was repeated independently three times using different RNA preparation and always in parallel with a corresponding WT construct.

## RESULTS

### Gag HIV-2, HIV-1 and $SIV_{Mac}$ are efficiently translated as leaderless RNA

We, and others, have recently shown that HIV-2, HIV-1 and $SIV_{Mac}$ *gag* coding regions contain one or several IRES able to trigger translation from the first initiation codon, and from in frame AUG triplets (21,23,24). We furthermore demonstrated that the HIV-2 *gag* coding region lacking the 5′UTR is efficiently translated, with no obvious leaky scanning from the first to the second AUG triplet (23). To extend this property to other primate lentiviruses, we compared the translation efficiency of leaderless (HIV-1–$AUG_1$ and $SIV_{MAC}$–$AUG_1$) or 5′UTR containing HIV-1 and $SIV_{Mac}$ *gag* gene (HIV-1-5′UTR and SIV-5′UTR) in rabbit reticulocyte

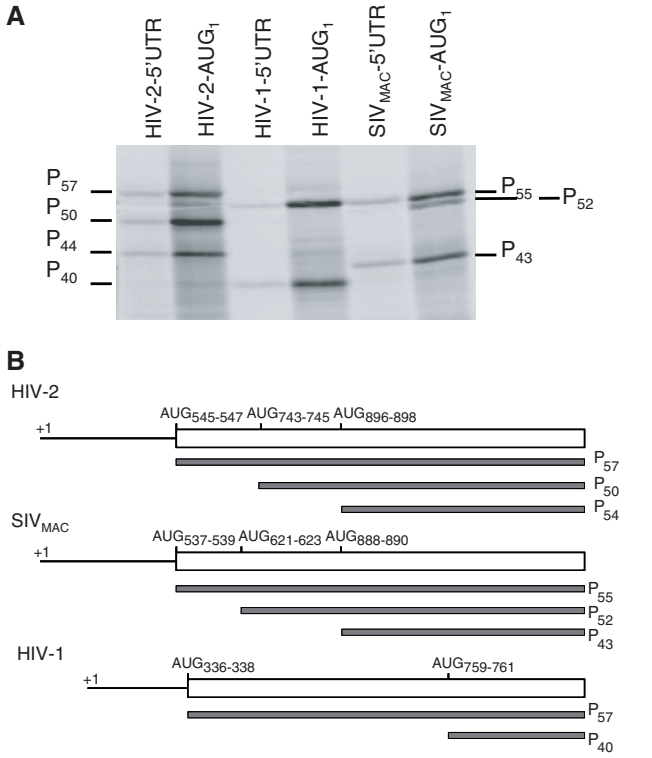


**Figure 1.** Lentiviral genomic RNA are more efficiently translated in absence of their 5′UTR. (**A**) Translation of HIV-2 rod, HIV-1 pNL4.3 and $SIV_{Mac}$ Mm 251 *gag* gene in rabbit reticulocytes lysate in presence of $^{35}$S-methionine. For each lentivirus, the RNA template translated comprises the *gag* entire open reading frame. Constructs with 5′UTR (HIV-1-5′UTR, HIV-2-5′UTR, $SIV_{MAC}$-5′UTR) start at the first nucleotide of the genomic RNA, constructs labelled devoid of 5′UTR (HIV-1–$AUG_1$, HIV-2–$AUG_1$, $SIV_{MAC}$–$AUG_1$) start with an extra non encoded guanosine before the first AUG codon. None of the transcripts were capped. Translation reactions contain 0.05 μM RNA and were carried out for 60 min as described in 'Materials and Methods' section. Proteins translated are referred to as $P_x$, $X$ being its apparent molecular weight. (**B**) Schematic representation of HIV-2 rod, HIV-1 pNL4.3 and $SIV_{MAC}$ Mm 251 *gag* gene, the initiation codons and the corresponding protein produced.

lysate (RRL) (Figure 1). As previously observed, the translation of both types of constructs yield several isoforms for each virus (21,23,24). The translation of HIV-2 rod genomic RNA yields three isoforms of 57, 50 and 44 kDa which are initiated from three in frame AUGs present within the matrix portion of the *gag* open reading frame (23). $SIV_{MAC}$ genomic RNA translation mainly yields two isoforms: the full length 55 kDa polyprotein and 43 kDa protein initiated from the in frame $AUG_{888–890}$ homologous to HIV-2 third initiation codon (24). We have detected an additional 52 kDa minor isoform, the translation of which is probably initiated from $AUG_{621–623}$ (Figure 1A and B). Translation of HIV-1 *gag* mRNA yields the previously described 55 and 40 kDa proteins (21). Of note, the shortest HIV-1 isoform is not homologous to any of HIV-2 or $SIV_{Mac}$ isoforms, because it is initiated on $AUG_{759–761}$, within the capsid part of the polyprotein. In all three cases, translation of the leaderless open reading frame is clearly more efficient than when it is preceded by its cognate 5′UTR (Figure 1A). HIV-2 *gag* in different context, was compared to a chimeric control gene constituted of the human β globin 5′UTR followed by a portion of the β gal coding sequence. This control was engineered to contain the same number of methionine than the full length Gag, in order to facilitate translation efficiency comparison (Figure 2). Under the conditions we used, the translation of the leaderless HIV-2 gag mRNA is 25-fold more efficient than when preceeded by a 5′UTR (Figure 2A and B). This is in clear contrast with the translation of an uncapped control gene which translation is 7-fold reduced upon 5′UTR deletion (Figure 2A and B). In order to simplify the constructs nomenclature, the three initiation codons: $AUG_{545–547}$, $AUG_{743–745}$ and $AUG_{898–890}$ (numbering from + 1, the transcription start) are termed $AUG_1$, $AUG_2$ and $AUG_3$, respectively.

It has been unambiguously shown both *in vitro* and *ex vivo* that, in a bicistronic context, the coding region itself attracts the ribosomes on $AUG_1$ (23,25). However, it is not clear if, in a leaderless context, the 5′ terminus of the RNA threads through the 40S subunit canal, or if the ribosomes are internally recruited. To preclude 5′ entry of the ribosomes, a stable hairpin ($\Delta G = -21.3$ kCal) was added 15 nucleotides upstream of the initiation codon of HIV-2–$AUG_1$ (Figure 2C). The presence of a 5′ stem-loop strongly reduced the translation of a control gene containing the globin 5′UTR, and almost abolishes the translation of the cognate leaderless RNA [Figure 2A and B and (11,26)]. Here again, HIV-2 *gag* mRNA appears atypical since its translation remains quantitatively and qualitatively unaffected when preceded by a stable stem-loop [please note that for all the experiments presented in Figure 2, RNAs were shown to be homogenous and stable in the RRL (see Supplementary Data S1)]. This indicates that initiation complexes, including those translating from $AUG_1$, were recruited by an internal initiation event. However, we can not tell if the ribosomes translating from $AUG_2$ and $AUG_3$ were positioned by a direct internal initiation event or through a 'land and scan' process.

As the sequences beyond $AUG_3$ do not influence translation efficiency (23,25), we constructed a truncated form of the *gag* gene which ends by an artificial stop codon at position 1351 (807 nucleotides downstream to $AUG_1$), this construct is termed HIV-2–$AUG_1^*$. The apparent molecular weight of the three isoforms translated from this construct are 30, 23 and 17 kDa instead of 57, 50 and 44 kDa from the full length construct (Figure 3A). The improved resolution allowed us to detect a minor fourth isoform running at 25 kDa. This protein had been previously observed, with the full length construct, and termed p52 (25). It has been attributed to an alternative initiation event on a cryptic CUG and/or UUG codon at position 692–694 and 695–697, respectively. We confirmed the identity of this protein, by mutating both alternative initiation codons into non-initiation codons $UUC_{692–694}$ and $UCG_{695–697}$ (HIV-2–$AUG_1^*$-alt). Translation of this mutant in RRL does no longer yield the 25 kDa protein while the three other isoforms remain efficiently produced (Figure 3B). Interestingly, initiation at $AUG_2$ is stimulated in HIV-2–$AUG_1^*$-alt, this could reveal that

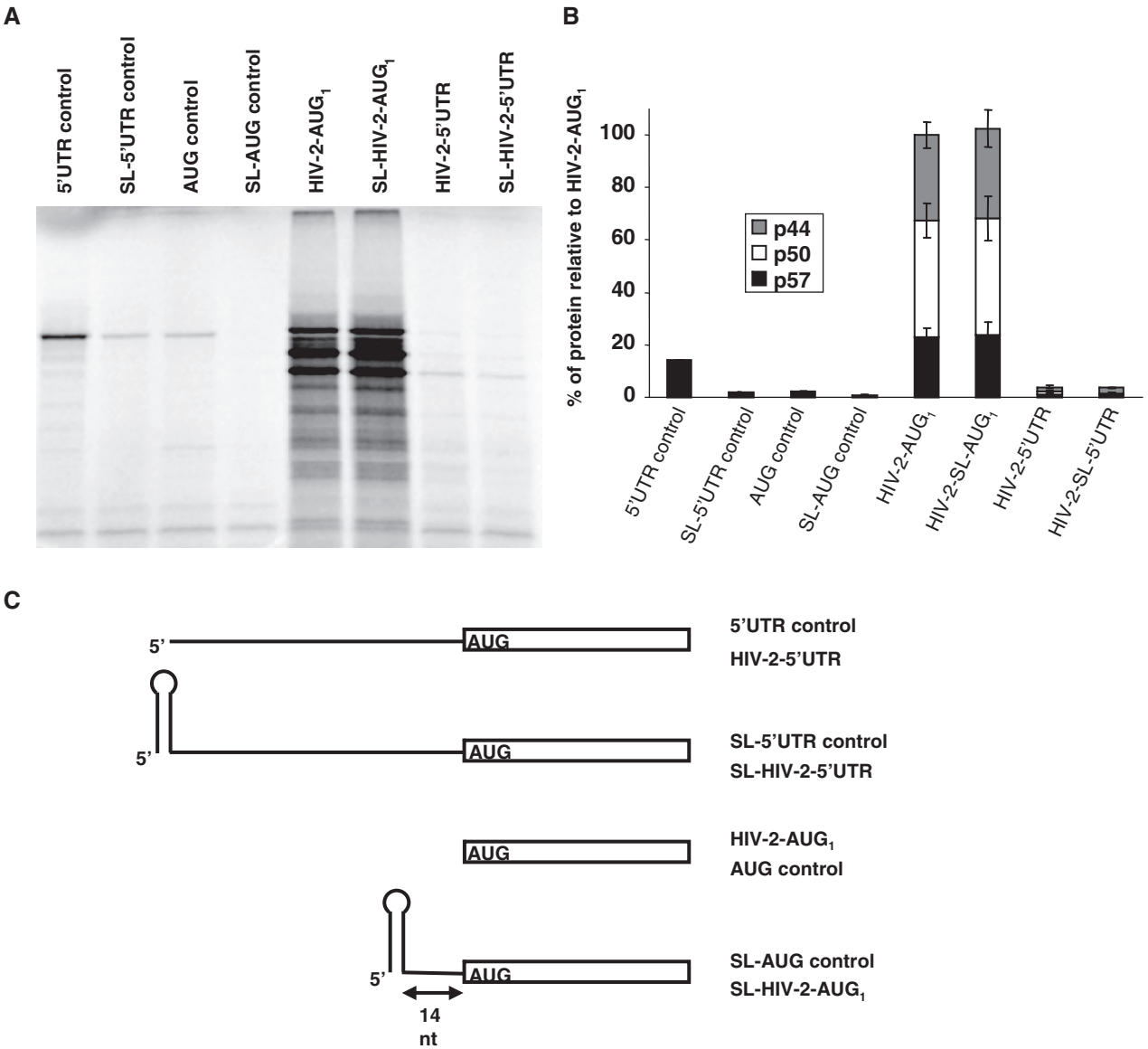


**Figure 2.** Leaderless *gag* mRNA is efficiently transcribed by internal entry of the ribosome. *In vitro* translation of HIV-2 gag compared to translation of a control gene in presence of $^{35}$S methionine. The control gene (5′UTR control) is a chimeric construct comprising the 5′UTR of the human β-globin upstream of the β-galactosidase open reading frame. AUG-control corresponds to the β-galactosidase open reading frame preceded by a single G, SL-5′UTRcontrol and SL-AUG-control, respectively, correspond to 5′UTR control and AUG-control preceded by a stable 13 base pair stem-loop (see 'Material and Methods' section). Those controls have been designed so that the protein translated contains the same number of methionines (13) as the full length HIV-2 gag. Therefore the yield in the different protein translated directly reflects their relative efficiencies of translation. HIV-2- 5′UTR is the whole *gag* gene preceded by its cognate 5′ UTR, HIV-2- AUG1 is the *gag* open reading frame preceded by a single G, HIV-2-SL-5′UTR and HIV-2- SL–AUG1 are the corresponding constructs preceded by a stable stem-loop. Translation reactions were as described in the 'Material and Methods' in presence of 0.05 μM of RNA for 30 min, experiments were repeated three times independently. (**A**) SDS–PAGE analysis of the different proteins translated in RRL; (**B**) Quantitative analysis of the different proteins translated in RRL: the results of three independent experiments were quantified, normalised to the total amount of proteins (three isoforms) produced by translation of HIV-2–$AUG_1$, and averaged. The *Y*-error bars correspond to ± the standard deviation with the 5′UTR-control used as internal standard. (**C**) Schematic representation of the different constructs used.

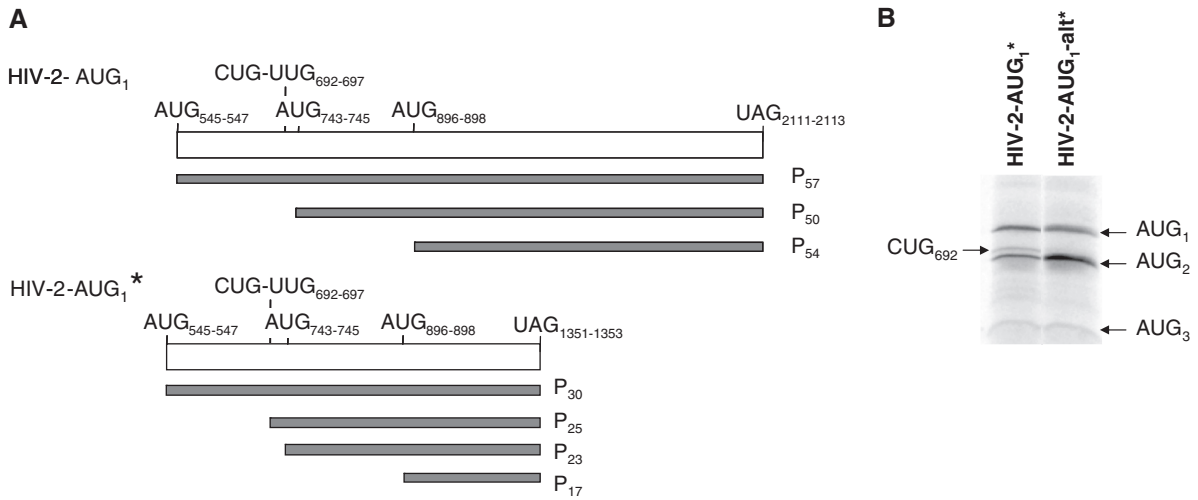


**Figure 3.** A fourth isoform is translated from alternative initiation codons. (**A**) Schematic representation of the truncated HIV-2 transcript (HIV-2–$AUG_1^*$), and the resulting proteins. (**B**) *In vitro* translation of the truncated form (+1 to 1641) of HIV-2–$AUG_1^*$ and HIV-2–$AUG_1^*$alt (HIV-2 mutated on the two alternative codons: $CUG_{692–694}$ and $UUG_{695–697}$ to $UUC_{692–694}$ and $UCG_{695–697}$). Translation of HIV-2–$AUG_1^*$ yields four isoforms of 30, 25, 23 and 17 kDa. The mutated HIV-2-$AUG_1^*$ alt no longer supports the translation of the p25 isoform. Both constructs were translated in standard conditions with 0.05 μM of RNA for 30 min, as described in 'Material and Methods' section.

initiation on $CUG_{692}/UUG_{695}$ may originate from initiation complexes scanning to $AUG_2$.

### Three initiation complexes are recruited on a single RNA molecule

We next analysed the initiation complexes recruited on HIV-2–$AUG_1$ by sedimentation on sucrose gradients. Body-labelled HIV-2–$AUG_1$ was incubated in RRL pre-treated with translation inhibitor, and formed initiation complexes were separated on sucrose gradient. Ribosomal complexes (48S and 80S) were unambiguously identified by (i) comparison with the profile obtained with a control gene, (ii) modification of the profiles depending on the inhibitor used and (iii) analysis of the UV profile of the same gradients which detects 40S and 60S subunits.

HIV-2-$AUG_1$ RNA was incubated in GMP-PNP pretreated RRL. GMP-PNP, a non-hydrolysable analogue of GTP, inhibits hydrolysis of GTP by eIF2 upon AUG recognition by the 48S particle. This prevents 60S subunit joining and causes 48S complexes accumulation on mRNA. After separation of the initiation complexes, most of the RNA sediments as hnRNP (fraction 5–9), and as a peak corresponding to the 48S (fraction 14–16). The 48S peak is shouldered (fraction 18) and smears until fraction 22, suggesting the presence of heavier complexes (Figure 4A). This experiment was repeated on a lysate pre-treated with cycloheximide, a translocation inhibitor which stalls ribosomes and induces the accumulation of 80S complexes on initiation codons. Under those conditions, the 48S peak in fraction 15 almost disappears, mostly to the benefit of a heavier complex in fraction 20 corresponding to the 80S complex. A significant amount of the mRNA sediments as a second peak centred on fraction 24, which is shouldered by a third peak reaching a maximum at fraction 26 (Figure 4A). We then incubated a capped and poly-adenylated globin mRNA in untreated RRL, freeze translation by adding cycloheximide, and analysed the sedimentation profile of polysomes paused on the mRNA. We found that the first 80S complex sediments in fraction 20, while the two following complexes corresponding to two and three 80S ribosomes translating on the same RNA sediment in fraction 24 and 26, respectively (Figure 4B). To confirm that the observed complexes were specific for HIV-2 coding region, we next analysed the sedimentation profile of 172-nucleotide long RNA transcribed from a human genomic sequence that do not contain any initiation triplet (Sequence available in Supplementary Data S2). As expected, such an RNA yields only one peak (fraction 5) corresponding to the RNA complexed to non specific protein (Figure 4B). In order to determine if the formation of those initiation complexes relies on the initiations codons, we mutated both $AUG_2$ and $AUG_3$ to non-initiation codons (HIV-2–$AUG_1$–$CUC_2$–$CUC_3$). Incubated in a cycloheximide pre-treated RRL this leaderless RNA sediments as a single peak corresponding to one 80S complex (Figure 4C). Furthermore, the exact same pattern of sedimentation with no decrease in intensity of the 80S peak is obtained when using a leaderless RNA preceded by a stable stem-loop (SL–HIV-2–$AUG_1$–$CUC_2$–$CUC_3$, Figure 4D) confirming the internal recruitment on $AUG_1$. In summary, we show here that three 80S complexes are stalled on the WT *gag* sequence in cycloheximide pretreated RRL, while only one is observed upon mutation of the two internal AUG codons. Although there is no natural control for this, it is tempting to interpret the three lighter complexes observed in GMP–PNP pre-treated lysate as multiple 48S complexes.

To determine the exact location of those initiation complexes we used the 'toeprinting' approach. To this end, RNAs were incubated in a GMP–PNP or

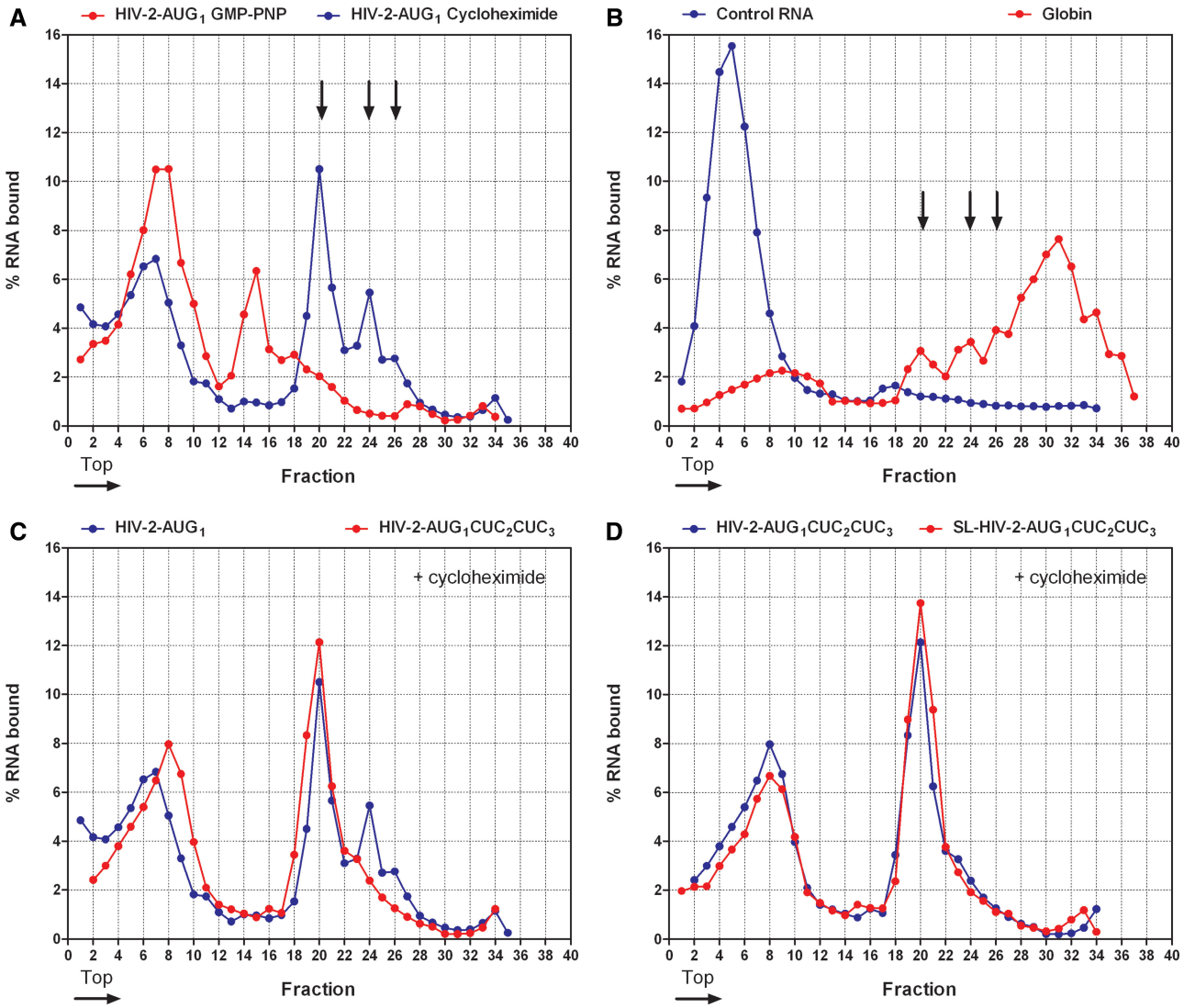


**Figure 4.** Three initiation complexes are present on a single RNA molecule. Sucrose gradient analysis of the initiation complexes were performed as described in 'Materials and Methods' section. Initiation complexes formed were then analysed on a 10–50% sucrose gradient. Fractions were collected and counted. (**A**) HIV-2–$AUG_1$ was incubated in RRL pre-treated with GMP–PNP (red line) or cycloheximide (blue line) pre-treated RRL. Body labelled mRNA were incubated in RRL pretreated with GMP–PNP (to analyse 48S complexes) or cycloheximide (to analyse 80S complexes). Arrows indicate respectively, the fractions in which sediments one 80S complex (fraction 20), two 80S complexes (fraction 24) and three 80S complexes (fraction 26), as determined by the analysis of the polysome profile (Figure 4B). (**B**) Capped polyadenylated globin mRNA (red line) was incubated for 10 min in untreated RRL, then 4.8 mM cycloheximide was added, and the mix was analysed on a 10–50% sucrose gradient. As control, a 172-nucleotide-long mRNA containing no AUG was incubated in untreated RRL (control RNA, blue line), and separated on a 10–50% sucrose gradient. Arrows indicate respectively, the fractions in which sediments one 80S complex (fraction 20), two 80S complexes (fraction 24) and three 80S complexes (fraction 26). (**C**) HIV-2–$AUG_1$ (blue line) and HIV-2–$AUG_1$–$CUC_2$–$CUC_3$ (red line) were incubated in cycloheximide pre-treated RRL. (**D**) HIV-2–$AUG_1$–$CUC_2$–$CUC_3$ (blue line) and SL HIV-2–$AUG_1$–$CUC_2$–$CUC_3$ (red line) were incubated in cycloheximide pre-treated RRL.

cycloheximide treated RRL to pause initiation complexes. The presence and the position of the initiation complex were then revealed by reverse transcription using radiolabeled primers. Depending on the experimental system several premature reverse transcriptase stops 13–21 nucleotides downstream to the A of the initiation codon are expected (27–31). We detected clear cycloheximide dependent reverse transcriptase stops downstream to each of the AUG triplets on the leaderless RNA. For the two internal initiation codons, we observed one stop at $G_{761}$, 18 nucleotides downstream to $AUG_2$, and four stops at $G_{912}A_{913}C_{914}C_{915}$ located 17–20 nucleotides 3′ of $AUG_3$ (Figure 5B and C). Initiation on $AUG_1$ is clearly more enigmatic since it has previously been shown that the initiation complex contacts several nucleotides upstream from the initiation codon, and covers ~15 nucleotides up- and downstream the initiation triplet (13). We therefore compared the position of the initiation codon paused on $AUG_1$ on WT and leaderless RNA. As shown on Figure 5A, the results obtained were similar on both RNA. The presence of the initiation complex induces five reverse transcriptase stops at $C_{562}G_{563}U_{564}C_{565}U_{566}$, 17–21 nucleotides downstream to $AUG_1$. The specificity of those signals is demonstrated by

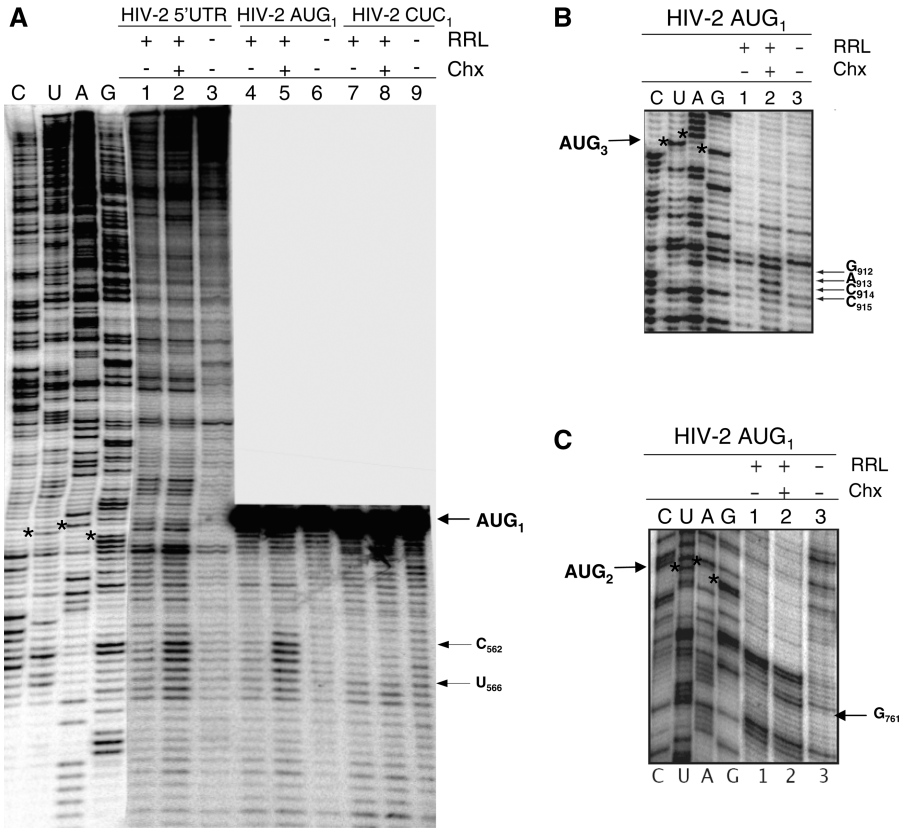


**Figure 5.** Toeprinting analysis of the initiation complexes paused on HIV-2 $AUG_1$–Gag. (**A**) Toeprinting analysis of the initiation complex present on $AUG_1$ initiation codon of HIV-2–$AUG_1$ RNA. Initiation complexes were mapped on three different RNA: the 5′ UTR containing construct, HIV-2–5′UTR, the leaderless RNA, HIV-2–$AUG_1$ and a version of the leaderless mutated for the 5′ proximal AUG: HIV-2–$CUC_1$. RNA were reverse transcribed with the labelled primer HIV-2-16 ('Materials and Methods' section) in three different conditions: incubated in RRL in standard translation conditions (RRL +, Cycloheximide −, lanes 1,4,7), in RRL pretreated with 4.8 μM cycloheximide (Chx) (RRL +, cycloheximide +, lanes 2,5,8), incubated in water (RRL −, cycloheximide −, lanes 3,6,9). Elongation products were run along with a sequence realised on HIV-2-5′UTR with the HIV-2-16 primer. The bases of the initiation codon are indicated by red stars. Reverse transcriptase stops specific for the AUG containing constructs in presence of cycloheximide are identified on the right side of the autoradiogram. (**B**) Toeprinting analysis of the initiation complex present on the $AUG_2$ initiation codon of HIV-2–$AUG_1$ RNA. Initiation complexes were mapped on HIV-2–$AUG_1$ RNA. RNA was reverse transcribed with the labelled primer HIV-2-6 ('Materials and Methods' section) in three different conditions: incubated in RRL in standard translation conditions (RRL +, cycloheximide −, lane 1), in RRL pretreated with 4.8 μM cycloheximide (RRL +, cycloheximide +, lanes 2), incubated in water (RRL −, cycloheximide −, lanes 3). Elongation products were run along with a sequence realised on HIV-2–$AUG_1$ RNA using the labelled HIV-2-6 primer ('Materials and Methods' section). The bases of the initiation codon are indicated by red stars. Reverse transcriptase stops specific for the AUG containing constructs in presence of cycloheximide are identified on the right side of the autoradiogram. (**C**) Toeprinting analysis of the initiation complex present on the $AUG_3$ initiation codon of HIV-2–$AUG_1$ RNA. Legend is identical to panel 5B.

their dependence on an intact initiation triplet (Figure 5A, lanes 7–9). Comparable results were obtained with HIV-1 *gag* leaderless RNA (Figure 6) on which presence of the 80S complex arrest primer elongation on $G_{353}G_{354}U_{355}A_{356}U_{357}$ 17–21 nucleotides before reaching $AUG_{336–338}$ (homologous to HIV-2 $AUG_1$). Interestingly, we were unable to map any initiation complex on HIV-1 or HIV-2 5′ proximal AUG when the RRL was treated with GMP–PNP, although those complexes are merely present on the RNA according to the sedimentation profile (see above). This is likely to reflect the poor stability of the 48S on the 5′ proximal AUG lacking upstream sequences.

## Secondary structure model for HIV-1, HIV-2 and $SIV_{Mac}$ *gag* coding region

The properties of some viral IRES rely on the integrity of their tertiary folding (32). Based on chemical and enzymatic probing, we previously proposed a secondary structure model for the HIV-2 *gag* coding region (23). In this study, we modelled HIV-1 (laboratory strain NL4.3) and $SIV_{MAC}$ (isolate Mm251) *gag* coding region, according to experimental and phylogenetic data. Leaderless HIV-1–$AUG_1$ and $SIV_{MAC}$–$AUG_1$ RNA were subjected to the action of CMCT, DMS or to the cleavage by the Cobra Venom RNAseV1 (see 'Materials and Methods' section). DMS and CMCT specifically modify unpaired adenosine and cytosine or uracil and guanosine, respectively. RNAse V1 provides a complementary information since it specifically cleaves stacked nucleotides, revealing mostly double-stranded region (33). The nucleotides targeted by those reagents are revealed by reverse transcription initiated with a $^{32}$P-labelled primer (see Supplementary Data S3 for examples of results). DMS and CMCT treatments were performed in presence or absence of $Mg^{2+}$ in the

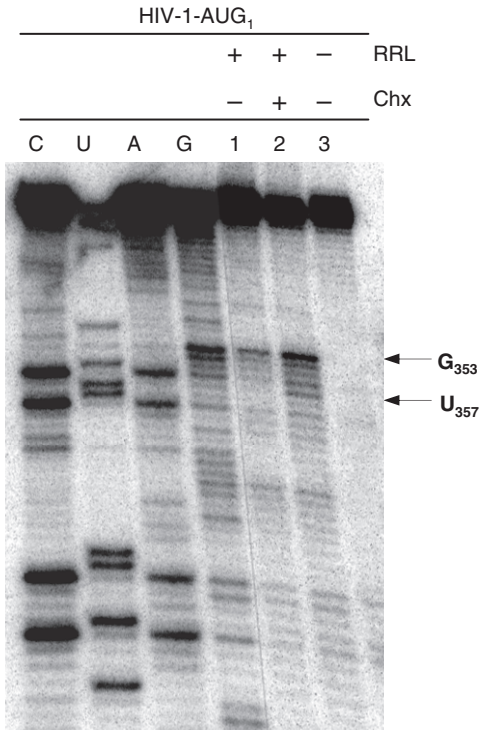


**Figure 6.** Toeprinting analysis of the initiation complex present on HIV-1 $AUG_1$ initiation codon. Same as 4B, but initiation complexes were mapped on HIV-1–$AUG_1$ RNA. RNA was reverse transcribed with the labelled primer HIV-1-3. Elongation products were run along with a sequence realised on HIV-1–$AUG_1$ RNA using the labelled HIV-1-3 primer.

reaction buffer, in order to gain insight into the influence of divalent ions on the folding. It is generally considered that $Mg^{2+}$ ions stabilise weak pairings and are required for the formation of long distance base pairings and pseudoknot. According to the intensity of the RT stop induced we ranked the hits in two classes: the weak hits (2–3-fold increased compared to the untreated control RNA) and strong hits (over 3-fold). Conducting those processes in parallel, we obtained an extensive map of nucleotide accessibility that was used as constraints for Mfold. In a first round of modelling, we used only strong single-strand hits persisting in presence of $Mg^{2+}$ ions, the models were further refined using weak hits, RNAse V1 cleavage sites and a phylogenetic study (see Supplementary Data S4, S5 and S6). For that matter, it should be noted that a multitude of sequences are available for HIV-1, we chose a subset of sequences representative of each groups and subtypes, plus one sequence from its homologue infecting the chimpanzee (SIVcpz). For SIV, we modelled the sequence of $SIV_{Mac}$ (macaque isolate Mm251) and compare it with 15 SIV sequences infecting different subspecies of macaque ($SIV_{Mac}$, $SIV_{MNE}$ and $SIV_{STM}$) and sooty mongabey monkeys ($SIV_{SMN}$). A secondary structure model previously established for HIV-2 rod was assessed using 18 HIV-2 sequences available from the Los Alamos HIV database (http://www.hiv.lanl.gov/content/index) and compared with the HIV-1 and SIV models. Please note that HIV-2 model was only slightly rearranged (see 'Discussion' section), although it may deceptively appear very different from our previously published model due to changes in nomenclature and design. *Gag* sequences are clearly constrained by their coding function, leaving few opportunities for compensatory changes. Nevertheless we observed frequent conservative changes (transition on one side of the helix that does not disturb the pairing, as U–G to C–G or A–U to G–U), and scarce compensatory changes. The folding was considered consolidated by the phylogenetic analysis when equivalent secondary structures can be formed in all isolates (the results of the detailed phylogenetic study are presented in the Supplementary Data S4, S5 and S6).

In the resulting models, most of the DMS and CMCT hits are in single-stranded regions, although weak hits were tolerated within unstable stems or at the extremity of helices. The majority of nuclease V1 sites are within helical regions, but some of them left in single-strand region could reflect a yet unidentified pairing, or nucleotide stacking (nucleotide at a base of an helix, or within a structured loop for example). Finally, nucleotides protected from modification upon $Mg^{2+}$ addition were attributed to long range base pairings, unstable stem or left single-stranded, possibly reflecting an unidentified long-range interaction.

The three models share common elements, although $SIV_{Mac}$ and HIV-1 sequences are only 60 and 80% homologous to HIV-2 sequences respectively (Figure 7). The structural similarity is mainly observed in the 5′ part of the sequence. The parings have been named $P_x$ in their order of appearance from the 5′ terminus.

The initiation codon is immediately followed by P1, a short G–C rich stem-loop. The sequence of the stem is conserved within a considered species, although the sequence in the loop varies (see Supplementary Data S4, S5 and S6). In HIV-2, it is followed by stem-loop termed P1-2, conserved in all HIV-2 isolates, but which has no homologues in HIV-1 and $SIV_{Mac}$. The most striking feature of this model is P2; a central long-range base-pairing that can be modelled in the three species although with different sequences. In all three cases some nucleotides within this stem are susceptible to DMS or CMCT modification in absence of $Mg^{2+}$ ions. Numerous conservative changes are observed within this stem in the three species considered. P2 is fragmented in two and three segments in HIV-2 and $SIV_{Mac}$, respectively. It closes a central wheel from which emerge two to three stem-loops. Notably, the first, P3 is a G–C rich stem which exposes a large A / purine-rich single-stranded region. In the case of HIV-1 and HIV-2, P3 is capped by a large unstructured loop in which we have modelled few base pairs which could account for the few V1 nuclease cleavages observed in this region. For $SIV_{Mac}$, P3 is constituted of two segments interrupted by an A rich internal loop. In HIV-2 and $SIV_{Mac}$ secondary structures, two stem-loops are found directly 3′ of P3. The first, P3-2, is well supported by both the structure probing data and the phylogenetic analysis (see Supplementary Data S4 and S5). Nevertheless, in $SIV_{Mac}$ this structure, essentially composed of A–U base pairs is formed only upon

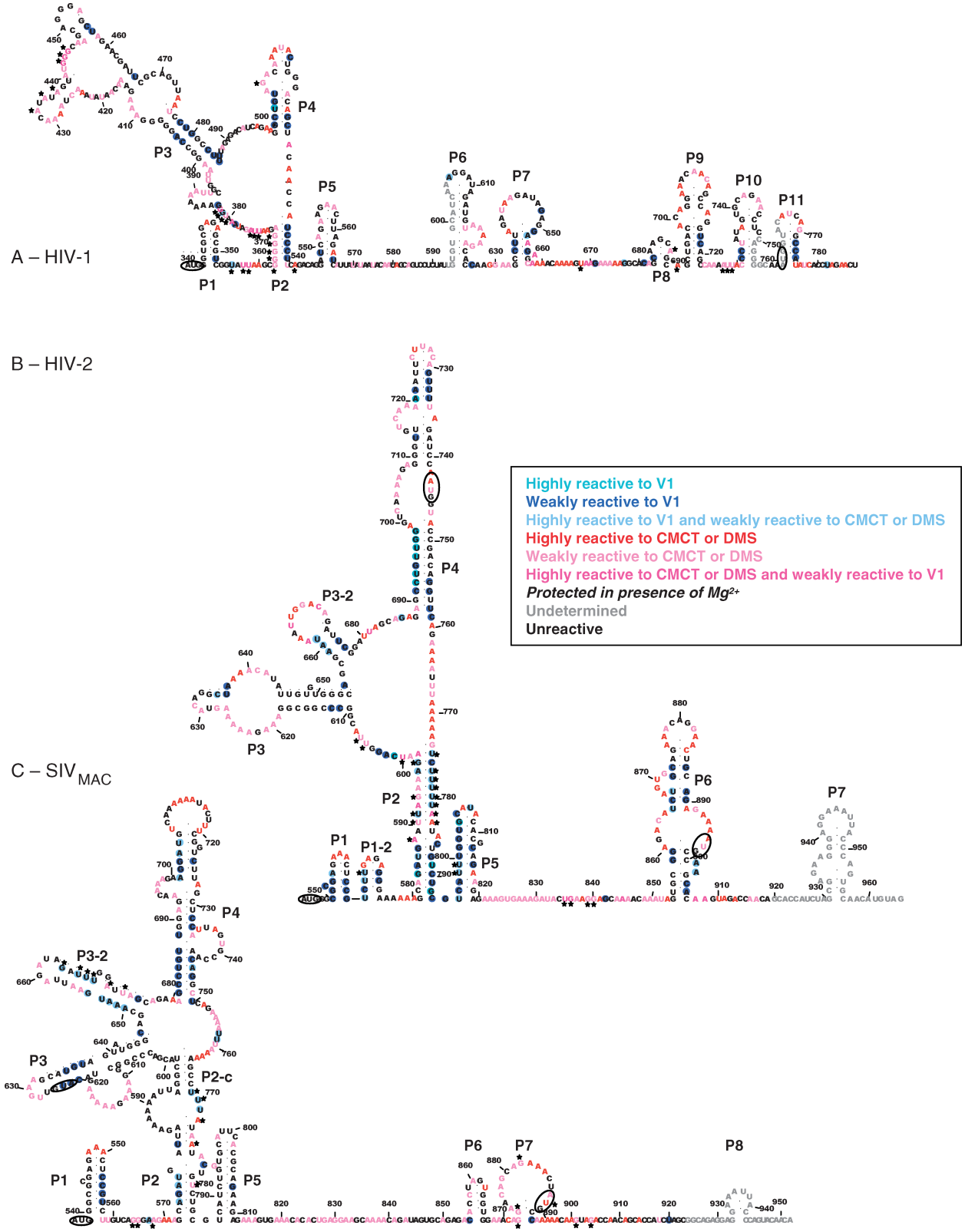


**Figure 7.** Secondary structure model of HIV-2, $SIV_{Mac}$ and HIV-1 coding region. Colours reflect the accessibility of the nucleotides to the different probes as indicated in the boxed area. 'Weakly reactive' refers to reverse transcriptase stops enhanced 2–3-fold compared to the untreated RNA, 'strongly reactive' position are enhanced 4 fold and over. The asterisk denotes position protected by at least 2-fold in presence of magnesium. Initiation codons are circled. Pairings are designated by $P_n$ in order of appearance from 5′ to 3′. Segments of a same pairings are numbered $P_{n-x}$, and pairings that appeared to be non-conserved between the three virus are designated by $P_{x_1-x_2}$ with $x_1$ referring to the preceding conserved stem-loop. (**A**) Secondary model for HIV-1 *gag* coding sequences, from $AUG_{336-338}$ to $G_{788}$. (**B**) Secondary model for HIV-2 *gag* coding sequences, from $AUG_{545-547}$ to $G_{965}$. (**C**) Secondary model for $SIV_{Mac}$ *gag* coding sequences, from $AUG_{537-539}$ to $U_{952}$.

$Mg^{2+}$ addition. The second, P4 is a long helical domain constituted of three segments. The second initiation triplet of *gag* HIV-2, lies within the 3′ portion of the first internal loop of the P4 domain. Of note, this initiation codon is not conserved and is only present in the well-studied rod isolate (see Supplementary Data S4). For HIV-1, only one relatively short stem-loop has been modelled between P3 and the 3′ portion of P2, it was named P4 although no sequence, structural or functional homology links it with either HIV-2 or $SIV_{Mac}$ P4. The last common feature of the three viruses structure is the P5 stem-loop that immediately follows the 3′ sequences of P2. The secondary structure models for the sequences located in 3′ of P5 are fairly divergent from a virus to another. For $SIV_{Mac}$, this region appears mostly unstructured. Most of the 30 nucleotides downstream of HIV-2 P5 are hit by CMCT or DMS except for few bases including four nucleotides that are protected from modification upon $Mg^{2+}$ addition. This region was left single-strand, and is followed by a helical region P6 in which is embedded $AUG_{896–898}$. HIV-2 P6 is fully substantiated by the experimental probing profile and the phylogenetic analysis. For HIV-1 sequences, we have modelled four helical regions, numbered P6–P9, the latest comprising $AUG_{759–761}$ from which is initiated HIV-1 shorter isoform p40. No obvious functional or structural homology can be extrapolated between HIV-1 and HIV-2 P6. For HIV-1 sequences lying from $U_{580}$ to $A_{660}$, we have modelled two alternative structures whether organised as two stem-loops P6 and P7, or as a two segment stem-loop that embeds P6 (see Supplementary Data S7). Both models are supported by the phylogenetic study, and the co-existence of both structures would better explain the modification profile obtained than a single conformation.

### Defining the IRES determinants by directed mutagenesis

To assay the importance of the secondary structure modelled for the translation, we have mutated two of the conserved elements and evaluated the mutation effects on the IRES activity. We first introduced mutations on the 5′ or the 3′ side of P2: mutP2-5′: $_{581}$GCA GAU GAA$_{589}$ to $_{581}$*AAU* GAU *C*AA$_{589}$ and mutP2-3′: $_{785}$ACU GUC UGC$_{793}$ to $_{785}$AC*G* GUC *AUU*$_{793}$, both mutants drastically destabilise the first segment of P2 [$\Delta G_{mutP2–5'} - \Delta G_{WT} = 6.9$ kcal and $\Delta G_{mutP2–3'} - \Delta G_{WT} = 5.6$ kcal (34)]. Both HIV-2 $AUG_1$mutP2-5′ and HIV-2 $AUG_1$mutP2-3′ are severely impaired for *in vitro* translation. The proportion of the three isoforms remains unchanged as compared to the wild type, but the yield of each protein is 5-fold reduced in both mutants (Figure 8). The two sets of mutations, were combined on the same RNA, thus restoring P2 pairing with a new sequence ($\Delta G_{mutP2–5'/3'} - \Delta G_{WT} = 1.3$ kcal). The resulting leaderless RNA, HIV-2 $AUG_1$mutP2-5′/3′, is almost as efficiently translated as the WT constructs (Figure 8). This strongly suggests that P2 is an important structural determinant of the IRES activity, while its sequence is not. As a first step to characterise another element of the coding region structure, we deleted the whole P4 helical segment taking care of preserving the open reading frame (mutΔP4: 5′$_{682}$A G$_{683}$ (Δ) $_{759}$CA GAA$_{763}$ 3′). The coding region of this mutant is 75-nucleotides shorter than the wild type, and $AUG_{756–758}$ has been deleted, thus its translation should yield only two isoforms. The molecular weight of the longer should be 2750 Da less than its wild-type homologue while the shorter should remain unchanged. Indeed we observed the two expected proteins; the yield of the shortest isoform is not affected, while the translation efficiency of the 'full length' protein is 3-fold reduced, taking into account that the mutΔP4 misses one methionine as compared to the WT (Figure 9). This shows that the presence of the helical domain P4 is essential for the optimal IRES activity, although we can not tell at this stage whether it contains an important structure or sequence.

## DISCUSSION

### A new type of IRES is present in the coding region of lentivirus

We and other had previously shown that primate lentivirus *gag* coding region conceals one or several IRESes that directs ribosomes on several initiation codons (21,23,24). This confers to HIV-2 *gag* coding region the unique property to be more efficiently translated as a leaderless RNA, than when it is preceded by its cognate 5′UTR. Here we show that this phenomenon is conserved among primate lentiviruses. The present study also demonstrates that the translation efficiency of the leaderless RNA relies on the structural integrity of the coding region. It is tempting to speculate that *in vitro* the presence of the 5′ UTR directly or indirectly interferes with the active folding of the IRES. Two interactions between the 5′UTR and the coding region have been described and could interfere with the productive folding of the IRES (see below). Most interestingly, one of them occludes the initiation codon. It has recently been shown that in HIV-2 a minor splicing event disrupts this interaction, thus stimulating *gag* translation (35). However, it should be noted that this interaction does not influence HIV-1 translation and no such splicing event seems possible (35,36).

HIV Gag IRES shows several other peculiarities that distinguishes it from other described IRES. First, the analysis of the initiation complexes paused on HIV-2 *gag* coding region yielded interesting results clearly distinguishing the lentiviral IRES from other viral IRESes described. We report here that up to three 48S or 80S complexes can be formed on the same mRNA. At this stage, we can not tell if the recruitment of complexes is ordered, and if RNA structural rearrangements may happen after recruiting each complex. This is in clear contrast with HCV or CrPV IRESes which can recruit only one initiation complex. Those IRESes adopt a three-dimensional structure that specifically binds the 40S ribosomal subunit. At this step, most of the surface of the folded RNA is hindered by one 40S subunit (32, 37–39). In those cases, it would clearly not be possible to accommodate more than one 40S subunit on a single RNA in contrast to what we observed with HIV-2-*gag*.

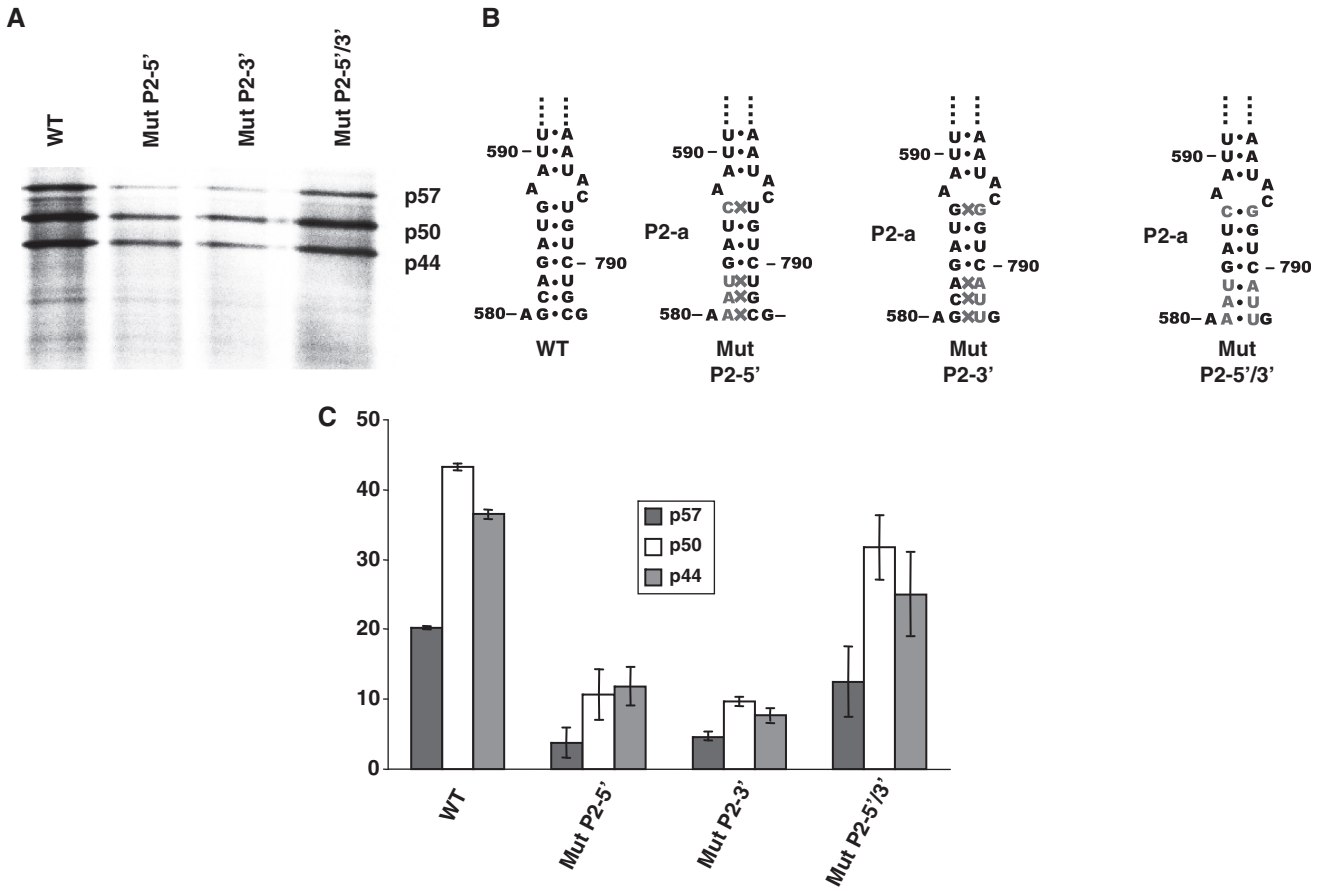


**Figure 8.** P2 is a major determinant of HIV-2 Gag translation efficiency. Mutant HIV-2–$AUG_1$ leaderless RNA were translated in RRL for 30 min in presence of 0.05 µM RNA. The control WT RNA and the mutants were translated in parallel in presence of $^{35}S$ methionine. (**A**) SDS–PAGE analysis of the translation of HIV-2–$AUG_1$ mutated in P2: MutP2-5′ harbours mutations in the 5′ strand of P2 stem, MutP2-3′ harbours mutations in the 3′of P2 stem, both destabilise P2 pairing, while mutP2-5′/3′ combines the two sets of mutations which restores the predicted P2 pairing. (**B**) Schematic representation of the mutants and their effect on P2 pairing. (**C**) Quantitative analysis of the different proteins translated in RRL: the results of three independent experiments were quantified, normalised to the total amount of proteins (three isoforms) produced by translation of HIV-2–$AUG_1$, and averaged. The *Y*-error bars correspond to ± the standard deviation with HIV-2–$AUG_1$ used as internal standard.

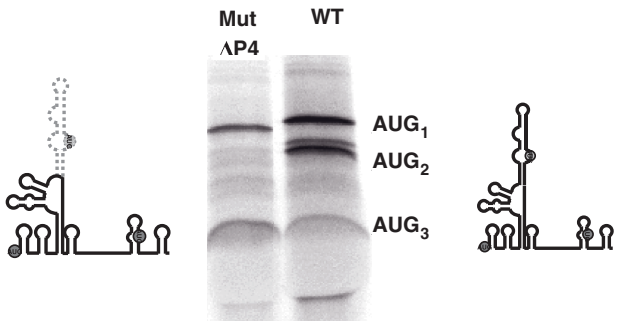


**Figure 9.** P4 helical domain contains an important determinant for HIV-2 Gag translation. Autoradiogram of the SDS–PAGE of HIV-2–$AUG_1^*$ (WT) and HIV-2–$AUG_1^*$ Δ P4 (MutΔP4) translation in RRL in presence of $^{35}S$ methionine. The secondary structure of the two constructs is represented on each side of the autoradiogram. Please note that the full length isoform produced from HIV-2–AUG1* ΔP4 misses one methionine as compared to the WT.

In this respect, HIV *gag* IRES may be compared to type II picornaviruses IRES, exemplified by the FMDV IRES which directs initiation from two in frame initiation codons (40). In this case it has not been investigated if two initiation complexes can coexist on the same RNA.

Using different primers and cycloheximide, we were able to identify initiation complexes on all three AUG triplets of HIV-2 *gag*. In each case, the toeprints observed are consistent with the initiation codon present in the P or the E site of 80S ribosome (27,28,30,41). The signals observed appear to be relatively weak with respect to the translation efficiency and the large amount of complexes observed on sucrose gradients. This could be inherent to the system we study. Indeed, reverse transcriptase is expected to disrupt the structure of the RNA 3′ to the AUG triplet where, in this peculiar case, lie the determinants necessary for recruiting the initiation complex. This could destabilise most of the 80S/mRNA complexes, although they are efficiently formed as shown by the sucrose gradient analysis. This would also explain why we were unable to detect GMP–PNP paused 48S complexes which association is

expected to be more labile than the 80S. However, the reverse transcriptase stops are specific, since they are observed only upon cycloheximide addition and disappear upon mutation of the AUG initiation codon to a CUC triplet. The toeprints observed on HIV-1 and HIV-2 $AUG_1$ are particularly interesting because sequences upstream of the initiation codon, usually lying within the mRNA channel are missing. The multiple stops observed probably also reflect the poor stability of the complex. These toeprints are reminiscent of those observed with initiation complexes reconstituted in absence of eIF1, conditions upon which the otherwise inefficient formation of initiation complexes on a 5′ proximal AUG is strongly stimulated (41). Besides, we show here that alternative initiation triplets are efficiently recognised although they are not in an optimal 'Kozak' context. Knowing that eIF1 is a major determinant for non-AUG initiation triplet discrimination (42), it is questionable whether or not the complexes recruited on HIV-2 *gag* contain eIF1.

Another peculiarity lies in the ability of HIV-2 *gag* IRES to support translation from at least four initiation sites, including an alternative initiation triplet located within P4 ($CUG_{692–694}$). This is another marked difference with the HCV and CrPV IRESes which direct initiation from a single initiation site precisely defined by the structure (32).

### An RNA structure responsible for the IRES activity

In this study we modelled the secondary structure of an isolated fragment of the *gag* open reading frame, sufficient for internal entry of the ribosome in three different viruses. Those models were established after the reactivity of individual nucleotides towards chemical and enzymatic probes, and the obtained models were confronted to the phylogenetic data available for each virus. We had previously reported a model for HIV-2 *gag* coding region (23), that was re-evaluated to better take into account the probing results and the phylogenetic data. The region spanning from $U_{600}$ to $C_{654}$ had been folded as two stem-loops, it is now modelled as a single G–C rich hairpin P3, which is fully substantiated by the probing data, and is furthermore similar to the P3 stem-loop modelled for $SIV_{MAC}$ and HIV-1 sequences. On the same bases, we also remodelled the sequences from $G_{794}$ to $A_{854}$ as one stem-loop, P5, followed by a long single-strand region. Secondary structure models for HIV-1 *gag* gene have previously been established using RNA comprising the whole 5′ UTR and a fraction or the whole coding sequence, *in vitro* or in various *in vivo* contexts (43–45). Those three models are fairly similar and comparable to ours, except for interactions between the 5′UTR and the open reading frame that are obviously absent from our model. Nucleotides surrounding $AUG_{336–338}$ can interact with the U5 region within the 5′UTR (43,45,46), here this region was part of a short stem-loop, P1. Both structures may alternatively exist, or P1 might be an artefact due to the shortened RNA construct we probed. In any case, we have previously reported that the P1 and P1-2 region of HIV-2 *gag* could be deleted without entailing any modification of the IRES properties of the open reading frame (23). Similarly, because of the absence of the 5′UTR in our construct, we could not detect the poly(A)-P3 loop pseudo-knot described for HIV-1 gRNA (43,45,47), the corresponding region within the coding region ($G_{443}$–$C_{449}$) is not, or marginally, structured in our model and is thus available for such an interaction. Interestingly, P2, the basal part of P3, and P5 are similar in all models published to date, and those features are the most obvious structural similarities in the three primate viruses. For the three viruses, we found that many nucleotides within the long-range base pairing P2 are protected from modification upon magnesium addition. In HIV-1, they were also found to be weakly hit by single-strand specific nuclease (43,44), and are protected from a single-strand-specific chemical probe by the nucleocapsid in virions (45). Here we show that this conserved pairing is important for HIV-2 *gag* IRES to promote translation from the three AUG. Mutations on each sides of P2 equally affects the three isoforms production. This could mean that recruitment on the three AUG relies on common determinants or, that P2 destabilisation leads to an overall unproductive alternative folding. The helical domain, P4 also influences the IRES activity although it is not crucial. The fact that the translation is not completely abolished in our mutants may reflect an indirect rather than a direct role for those structures. In other words, P2 and P4 may not be critical binding sites, but may be important to maintain the global architecture of the IRES and/or prevent alternative folding. Those structures could, for example, be important to expose the A-rich sequences present in the P2–P3–P4 domain and known to be potential IRES determinants (48,49). Such an hypothesis could explain how sequences spanning form $AUG_1$ to $AUG_2$ or from $AUG_2$ to $AUG_3$ retain some of the IRES activity although neither P2 nor P4 can not be formed in any of those two fragments (25).

According to the structure conservation and our mutagenesis data, we expect P2, P3 and P5 to be the major determinants involved in ribosome recruitment. We can only speculate on the role of this IRES, but the structural and functional conservation testifies for its physiological role. It is obvious from our experiments that it promotes a very efficient 5′UTR independent translation initiation. This property could ensure Gag production when the 5′UTR inhibition is relieved, for example, upon structural rearrangement induced by the binding of a protein, or a splicing event. The identification of a conserved core for lentiviral *gag* IRES will enable subsequent studies towards the elucidation of the molecular mechanism. If its peculiarity is confirmed, the phenomenon would be a therapeutic target of choice.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

Authors would like to thank Dr Nicolas Locker for critical reading of the manuscript, E. Ricci for generous

gift of plasmid. Research in their laboratory is funded by ANRS (Agence Nationale de recherche contre le S.I.D.A.), a grant from ANR (Agence Nationale pour la Recherche) and an ATIP from the CNRS. L.J. was a recipient of a fellowship from ANRS, L.W. was a recipient form a fellowship from SIDACTION.

## FUNDING

ATIPE from CNRS and a grant from the 'Agence Nationale pour la Recherche sur le S.I.D.A' (ANRS). Funding for open access charge: CNRS.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Butsch,M. and Boris-Lawrie,K. (2002) Destiny of unspliced retroviral RNA: ribosome and/or virion? *J. Virol.*, **76**, 3089–3094.
2. Butsch,M. and Boris-Lawrie,K. (2000) Translation is not required To generate virion precursor RNA in human immunodeficiency virus type 1-infected T cells. *J. Virol.*, **74**, 11531–11537.
3. Kaye,J.F. and Lever,A.M. (1999) Human immunodeficiency virus types 1 and 2 differ in the predominant mechanism used for selection of genomic RNA for encapsidation. *J. Virol.*, **73**, 3023–3031.
4. Ganser-Pornillos,B.K., Yeager,M. and Sundquist,W.I. (2008) The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.*, **18**, 203–217.
5. Abbink,T.E. and Berkhout,B. (2007) HIV-1 reverse transcription: close encounters between the viral genome and a cellular tRNA. *Adv. Pharmacol.*, **55**, 99–135.
6. Abbink,T.E. and Berkhout,B. (2008) RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J. Virol.*, **82**, 3090–3098.
7. Gatignol,A. (2007) Transcription of HIV: Tat and cellular chromatin. *Adv. Pharmacol.*, **55**, 137–159.
8. Lever,A.M. (2007) HIV-1 RNA packaging. *Adv. Pharmacol.*, **55**, 1–32.
9. Paillart,J.C., Shehu-Xhilaga,M., Marquet,R. and Mak,J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.*, **2**, 461–472.
10. Kozak,M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol. Cell. Biol.*, **9**, 5134–5142.
11. Kozak,M. (1980) Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell*, **19**, 79–90.
12. Kozak,M. (1997) Recognition of AUG and alternative initiator codons is augmented by G in position + 4 but is not generally affected by the nucleotides in positions + 5 and + 6. *EMBO J.*, **16**, 2482–2492.
13. Pisarev,A.V., Kolupaeva,V.G., Pisareva,V.P., Merrick,W.C., Hellen,C.U. and Pestova,T.V. (2006) Specific functional interactions of nucleotides at key -3 and + 4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.*, **20**, 624–636.
14. Peabody,D.S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.*, **264**, 5031–5035.
15. Touriol,C., Bornes,S., Bonnal,S., Audigier,S., Prats,H., Prats,A.C. and Vagner,S. (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol. Cell*, **95**, 169–178.
16. Miele,G., Mouland,A., Harrison,G.P., Cohen,E. and Lever,A.M. (1996) The human immunodeficiency virus type 1 5′ packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J. Virol.*, **70**, 944–951.
17. Ricci,E.P., Soto Rifo,R., Herbreteau,C.H., Decimo,D. and Ohlmann,T. (2008) Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem. Soc. Trans.*, **36**, 690–693.
18. Yilmaz,A., Bolinger,C. and Boris-Lawrie,K. (2006) Retrovirus translation initiation: issues and hypotheses derived from study of HIV-1. *Curr. HIV Res.*, **4**, 131–139.
19. Balvay,L., Lastra,M.L., Sargueil,B., Darlix,J.L. and Ohlmann,T. (2007) Translational control of retroviruses. *Nat. Rev. Microbiol.*, **5**, 128–140.
20. Brasey,A., Lopez-Lastra,M., Ohlmann,T., Beerens,N., Berkhout,B., Darlix,J.L. and Sonenberg,N. (2003) The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J. Virol.*, **77**, 3939–3949.
21. Buck,C.B., Shen,X., Egan,M.A., Pierson,T.C., Walker,C.M. and Siliciano,R.F. (2001) The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J. Virol.*, **75**, 181–191.
22. Ohlmann,T., Lopez-Lastra,M. and Darlix,J.L. (2000) An internal ribosome entry segment promotes translation of the simian immunodeficiency virus genomic RNA. *J. Biol. Chem.*, **275**, 11899–11906.
23. Herbreteau,C.H., Weill,L., Decimo,D., Prevot,D., Darlix,J.L., Sargueil,B. and Ohlmann,T. (2005) HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat. Struct. Mol. Biol.*, **12**, 1001–1007.
24. Nicholson,M.G., Rue,S.M., Clements,J.E. and Barber,S.A. (2006) An internal ribosome entry site promotes translation of a novel SIV Pr55(Gag) isoform. *Virology.*, **349**, 325–334.
25. Ricci,E.P., Herbreteau,C.H., Decimo,D., Schaupp,A., Datta,S.A., Rein,A., Darlix,J.L. and Ohlmann,T. (2008) In vitro expression of the HIV-2 genomic RNA is controlled by three distinct internal ribosome entry segments that are regulated by the HIV protease and the Gag polyprotein. *RNA*, **14**, 1443–1455.
26. Gallie,D.R., Ling,J., Niepel,M., Morley,S.J. and Pain,V.M. (2000) The role of 5′-leader length, secondary structure and PABP concentration on cap and poly(A) tail function during translation in Xenopus oocytes. *Nucleic Acids Res.*, **28**, 2943–2953.
27. Anthony,D.D. and Merrick,W.C. (1992) Analysis of 40 S and 80 S complexes with mRNA as measured by sucrose density gradients and primer extension inhibition. *J. Biol. Chem.*, **267**, 1554–1562.
28. Dmitriev,S.E., Pisarev,A.V., Rubtsova,M.P., Dunaevsky,Y.E. and Shatsky,I.N. (2003) Conversion of 48S translation preinitiation complexes into 80S initiation complexes as revealed by toeprinting. *FEBS Lett.*, **533**, 99–104.
29. Fraser,C.S., Hershey,J.W. and Doudna,J.A. (2009) The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat. Struct. Mol. Biol.*, **16**, 397–404.
30. Kozak,M. (1998) Primer extension analysis of eukaryotic ribosome-mRNA complexes. *Nucleic Acids Res.*, **26**, 4853–4859.
31. Wilson,J.E., Pestova,T.V., Hellen,C.U. and Sarnow,P. (2000) Initiation of protein synthesis from the A site of the ribosome. *Cell*, **102**, 511–520.
32. Kieft,J.S. (2008) Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.*, **33**, 274–283.
33. Brunel,C. and Romby,P. (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol.*, **318**, 3–21.
34. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
35. Strong,C.L., Lanchy,J.M., Dieng-Sarr,A., Kanki,P.J. and Lodmell,J.S. (2009) A 5′UTR-spliced mRNA isoform is specialized for enhanced HIV-2 gag translation. *J. Mol. Biol.*, **391**, 426–437.
36. Abbink,T.E., Ooms,M., Haasnoot,P.C. and Berkhout,B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*, **44**, 9058–9066.
37. Pfingsten,J.S., Costantino,D.A. and Kieft,J.S. (2006) Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science.*, **314**, 1450–1454.
38. Schuler,M., Connell,S.R., Lescoute,A., Giesebrecht,J., Dabrowski,M., Schroeer,B., Mielke,T., Penczek,P.A., Westhof,E. and Spahn,C.M. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat. Struct. Mol. Biol*, **13**, 1092–1096.
39. Spahn,C.M., Kieft,J.S., Grassucci,R.A., Penczek,P.A., Zhou,K., Doudna,J.A. and Frank,J. (2001) Hepatitis C virus IRES

RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science*, **291**, 1959–1962.
40. Andreev,D.E., Fernandez-Miragall,O., Ramajo,J., Dmitriev,S.E., Terenin,I.M., Martinez-Salas,E. and Shatsky,I.N. (2007) Differential factor requirement to assemble translation initiation complexes at the alternative start codons of foot-and-mouth disease virus RNA. *RNA*, **13**, 1366–1374.
41. Pestova,T.V. and Kolupaeva,V.G. (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.*, **16**, 2906–2922.
42. Lomakin,I.B., Shirokikh,N.E., Yusupov,M.M., Hellen,C.U. and Pestova,T.V. (2006) The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.*, **25**, 196–210.
43. Damgaard,C.K., Andersen,E.S., Knudsen,B., Gorodkin,J. and Kjems,J. (2004) RNA interactions in the 5′ region of the HIV-1 genome. *J. Mol. Biol.*, **336**, 369–379.
44. Henriet,S., Richer,D., Bernacchi,S., Decroly,E., Vigne,R., Ehresmann,B., Ehresmann,C., Paillart,J.C. and Marquet,R. (2005) Cooperative and specific binding of Vif to the 5′ region of HIV-1 genomic RNA. *J. Mol. Biol.*, **354**, 55–72.
45. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
46. Abbink,T.E. and Berkhout,B. (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.*, **278**, 11601–11611.
47. Paillart,J.C., Skripkin,E., Ehresmann,B., Ehresmann,C. and Marquet,R. (2002) In vitro evidence for a long range pseudoknot in the 5′-untranslated and matrix coding regions of HIV-1 genomic RNA. *J. Biol. Chem.*, **277**, 5995–6004.
48. Roberts,L.O. and Groppelli,E. (2009) An atypical IRES within the 5′ UTR of a dicistrovirus genome. *Virus Res.*, **139**, 157–165.
49. Terenin,I.M., Dmitriev,S.E., Andreev,D.E., Royall,E., Belsham,G.J., Roberts,L.O. and Shatsky,I.N. (2005) A cross-kingdom internal ribosome entry site reveals a simplified mode of internal ribosome entry. *Mol. Cell Biol.*, **25**, 7879–7888.
50. Sargueil,B., Hampel,K.J., Lambert,D. and Burke,J.M. (2003) In vitro selection of second site revertants analysis of the hairpin ribozyme active site. *J. Biol. Chem.*, **278**, 52783–52791.
51. Sargueil,B., McKenna,J. and Burke,J.M. (2000) Analysis of the functional role of a G.A sheared base pair by in vitro genetics. *J. Biol. Chem.*, **275**, 32157–32166.
52. James,L. and Sargueil,B. (2008) RNA secondary structure of the feline immunodeficiency virus 5′UTR and Gag coding region. *Nucleic Acids Res.*, **36**, 4653–4666.
53. Weill,L., Louis,D. and Sargueil,B. (2004) Selection and evolution of NTP-specific aptamers. *Nucleic Acids Res.*, **32**, 5045–5058.