Selective Pressure Against AUG Triplets in the 5′ Untranslated Region of Human Immunodeficiency Virus Type 1 Supports Cap-Dependent Translation Initiation Mechanism

Sam Khalouei[1] and Xuhua Xia[2,3]

[1]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada. [2]Department of Biology and Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Ontario, Canada. [3]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Ontario, Canada.

Abstract: The human immunodeficiency virus type 1 (HIV-1) is dependent on the host transcription and translation machinery to complete its life cycle. Different hypotheses have been postulated regarding the mechanism of translation initiation in HIV-1. The cap-dependent ribosomal scanning hypothesis predicts selection against AUG usage in optimal context in the HIV-1 5′UTR to avoid initiation at a wrong AUG by the scanning ribosome which would hamper the detection of the true downstream translation initiation codon. Recent reports have shown evidence of cap-independent translation initiation mechanisms in HIV-1, such as the direct binding of ribosome at internal ribosome entry sites (IRES), distant from the 5′ cap. It has been proposed that the IRES-dependent mechanism of translation initiation, allows the ribosome to bypass the stable secondary structures in the HIV-1 5′UTR. This hypothesis predicts no selective pressure against AUG usage in optimal context in the HIV-1 5′UTR since any such AUGs would be embedded in the stable secondary structures. We evaluated these two hypotheses based on their prediction regarding selective pressure against AUG usage in optimal context in the HIV-1 5′UTR. Our results show that there is indeed a selective pressure against AUG usage in optimal context in the HIV-1 5′UTR which supports the cap-dependent translation initiation hypothesis. That is, the HIV-1 mRNAs are translated either by the cap-dependent mechanism alone or in addition to IRES-dependent mechanism in the presence of 5′UTR stable secondary structures. They are clearly not translated by the latter mechanism alone.

Keywords: HIV-1, 5′UTR, translation initiation, selective pressure

Background

HIV-1 virus, which is responsible for the AIDS epidemic, mainly infects human $CD4^+$ T-lymphocytes (Ho et al. 1986). After entry into the cell, reverse transcription of the genomic RNA results in a double-stranded DNA, which is transferred to the nucleus and integrated into the human chromosome with the aid of the virus-coded integrase protein (Cimarelli and Darlix, 2002). Even though some of the enzymes that are critical to the HIV-1 replication cycle are coded by the viral genome, the virus is dependent on the host transcription and translation machinery. More than 30 different mRNAs are produced from alternative splicing of the full-length HIV-1 transcript, which are divided into three different classes; unspliced 9kb mRNAs, intermediate 4kb mRNAs, and short 2kb mRNAs (Schwartz et al. 1990; Purcell and Martin, 1993). Splicing of HIV-1 mRNAs is a complex mechanism due to the presence of constitutive and alternatively used 5′ splice donor (SD) and 3′ splice acceptor (SA) sites. There are several weak competing splice sites in the middle of the genome, and their alternative selection determines which proteins are coded by the mature mRNA (Arrigo et al. 1990; Guatelli et al. 1990; Furtado et al. 1991). These various forms of HIV-1 mRNAs are transferred to the cytoplasm where they undergo translation by the host ribosomes.

Translation of eukaryotic mRNAs follows a cap-dependent ribosomal scanning mechanism (CDRSM) whereby a pre-initiation complex comprised of the small 40S ribosomal subunit, $tRNA_i^{Met}$, and multiple translation initiation factors binds at the 5′ end of the mRNA and scans linearly in the 5′ to 3′ direction until it reaches the first AUG in an optimal context (Kozak, 1989). The 60S ribosomal subunit then binds to the complex and the translation of a polypeptide chain initiates. The optimal context for translation initiation in mammalian species, known as the Kozak consensus,

Correspondence: Xuhua Xia, University of Ottawa, 30 Marie Curie, Ottawa, Ontario, Canada K1N 6N5. Tel: 1-613-562-5800; Ext: 6886; Fax: 1-613-562-5486; Email: xxia@uottawa.ca

is RCCaugG, with the –3R (R = purine) and +4G being particularly important (Kozak, 1986; Kozak, 1987; Kozak, 1997).

Different hypotheses have been postulated regarding the mechanism of translation initiation in HIV-1 (for a review see Yilmaz et al. 2006). Despite the supporting evidence for the CDRSM hypothesis (Schwartz et al. 1992; Miele et al. 1996; Anderson et al. 2007), there are some obstacles associated with this mechanism of translation initiation in the HIV-1 virus. All HIV-1 transcripts share a 289-nucleotide sequence as part of their 5′UTR, which forms into stable stem-loop structures (Baudin et al. 1993; Berkhout and van Wamel, 2000). The 5′UTR is the most conserved region of the HIV-1 genome (Berkhout, 1996) and contains several functional domains which are important for virus replication. These include transactivation response region (TAR) (Selby et al. 1989), polyadenylation signal (polyA) (Klasens et al. 1999), primer binding site (PBS) (Beerens and Berkhout, 2002), RNA dimmer initiation signal (DIS) (Laughrea and Jette, 1994), the major splice donor site (SD1) (Purcell and Martin, 1993), and the packaging signal (Ψ), which is only present in unspliced mRNAs (Lever et al. 1989). The stable stem-loop structures formed by these domains have been shown to inhibit translation initiation by the scanning ribosome (Parkin et al. 1988; SenGupta et al. 1990; Geballe and Gray, 1992; Svitkin et al. 1994; Miele et al. 1996).

The possibility of alternative cap-independent translation initiation mechanisms in HIV-1 and other viruses was introduced by the observation of internal ribosome entry site (IRES)-mediated translation initiation in picornaviruses (Jang et al. 1988; Pelletier and Sonenberg, 1988). Through this mechanism, the ribosome directly enters at IRES to initiate the translation without having to bind at the 5′cap and scan the viral mRNA. Accumulating evidence point to the presence of IRES-dependent translation initiation in *retroviridae* family as well (Berlioz and Darlix, 1995; Lopez-Lastra et al. 1997; Deffaud and Darlix, 2000; Ohlmann et al. 2000), which suggests that this form of translation initiation might be an important part of the retroviral life cycle, including that of HIV-1. This mechanism would allow bypassing the stable 5′ secondary structures of the HIV-1 5′UTR, alleviating the barriers imposed by CDRSM on translation initiation. We refer to this IRES-dependent mechanism in the presence of 5′UTR Stable Secondary Structures (SSS) as SSS/IRES hypothesis.

The CDRSM and SSS/IRES-dependent translation initiation hypotheses have different predictions in regard to selective pressure against AUG usage in the HIV-1 5′UTR. The CDRSM hypothesis predicts that there should be a selective pressure against AUG usage in optimal context in the 5′UTR region. This is to avoid initiation at a wrong AUG by the scanning ribosome which would hamper the detection of the true downstream translation initiation codon. The SSS/IRES-dependent translation initiation hypothesis, on the other hand predicts no selective pressure against these AUGs because they would be embedded in the secondary structures and not exposed to the scanning ribosome (Fig. 1). The predictions from the different translation initiation scenarios are summarized in Table 1.

In this study we have assessed the predictions of the CDRSM and SSS/IRES-dependent translation initiation hypotheses in HIV-1, with regard to selective pressure against AUG usage in optimal context in the HIV-1 5′UTR. Our results demonstrate this selective pressure and further support the CDRSM mechanism in HIV-1.

Materials and Methods

Calculating the observed and expected number of AUGs in optimal context in the 5′UTR

The NCBI HIV-1 type genome (NCBI Genome: NC_001802), hereafter referred to as HIV-1_{NC1802}, was used to analyze the selective pressure against AUG usage in optimal context in the 5′UTR region. This sequence, which belongs to HIV-1 subtype B, has been curated by the NCBI staff and is the same as HXB2, which is the reference sequence of the Los Alamos National Library (LANL) HIV sequence database (Kuiken et al. 2003). The only difference is that HIV-1_{NC1802} lacks the first 454 bases of HXB2 and has the Vpu start codon (defective ACG in HXB2) corrected to ATG.

HIV-1_{NC1802} is a viral genome (i.e. packaged in the virus particle) and consists of 5′-R-U5-3′ in the 5′ end and 5′-U3-R-3′ in the 3′ end. The splice sites position numbering by Purcell and Martin (Purcell and Martin, 1993) corresponds to the HIV-1 vector pNL4-3 (NCBI Genome: AF324493),

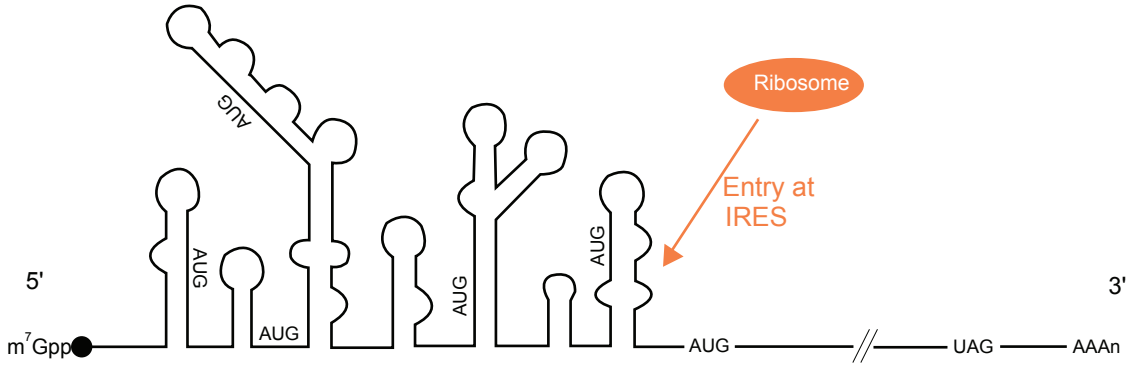

Figure 1. SSS/IRES-dependent translation initiation.
A prediction of secondary structure of HIV-1 Vif 5′UTR, adapted and modified from Yilmaz et al. 2006. The location of AUGs and IRES are hypothetical.

which is a provirus sequence (i.e. integrated into host genome), with identical repeats, 5′-U3-R-U5-3′, at both ends. Their numbering starts from +1 at the beginning of the U3 region. In order to annotate the splice site locations in HIV-1_{NC1802}, which lacks the initial U3 and terminal U5 regions, we aligned the two sequences using Blast2 program at NCBI (Altschul et al. 1990), and identified the corresponding splice site positions. The position of start and stop codons for the eight HIV-1 ORFs; gag-pol, vif, vpr, tat, rev, vpu, env, and nef, along with their sequence lengths are shown in Table 2.

There are more than 30 alternatively spliced HIV-1 transcripts (Purcell and Martin, 1993), each with a different combination of exons. Table 2 shows the most abundant transcript of each gene, according to figure 4 of the Purcell and Martin (1993), along with the comprising exons. The 5′UTR of these transcripts were extracted for analysis and their nucleotide frequencies were calculated using the DAMBE program (Xia, 2001; Xia and Xie, 2001). The expected number of AUGs for each sequence, NE_{AUG}, is obtained by the following formula (Xia, 2007, pp. 4–6):

$$NE_{AUG} = pA \times pU \times pG \times (L - 3 + 1)$$

where L stands for the sequence length, 3 is the length of the query sequence (i.e. AUG), and pA, pU, and pG are the frequencies of A, U, and G nucleotides, respectively. The +4G of the Kozak consensus has been shown to be important in translation initiation especially in the absence of the −3R (Kozak, 1986; Kozak, 1987; Kozak, 1997), therefore the optimal contexts of AUG in this study are considered as the presence of either or both −3R and +4G; Rnnaug (AUG_{-3R}), RnnaugG (AUG_{-3R+4G}), and YnnaugG (AUG_{-3Y+4G}), where "aug" is the translation initiation codon and "n" represents any nucleotides. The expected number of AUG_{-3R}, AUG_{-3R+4G}, and AUG_{-3Y+4G} were calculated as $pR \times pA \times pU \times pG \times (L - 6 + 1)$ (where $pR = pA + pG$), $pR \times pA \times pU \times pG \times (L - 7 + 1) \times pG$, and $pY \times pA \times pU \times pG \times (L - 7 + 1) \times pG$ (where $pY = pC + pU$), respectively. The observed numbers of AUGs in the three different optimal contexts were also recorded for each 5′UTR.

Table 1. Three different scenarios involving CDRSM and SSS/IRES and their respective predictions. Scenarios 1 and 3 have the same prediction.

Scenarios	CDRSM	SSS/IRES	Predictions
1	Yes	No	Strong selection against AUG in 5′UTR
2	No	Yes	No selection against AUG in 5′UTR
3	Yes	Yes	Strong selection against AUG in 5′UTR

Table 2. The HIV-1 coding sequences and corresponding 5´UTRs.
The position of start and stop codons are indicated according to the NCBI HIV-1 type sequence (NCBI Genome: NC_001802). Start position corresponds to the position of A in AUG start codon and stop position corresponds to the third base position of the stop codon. The most abundant transcripts and annotation of exons are obtained from Purcell and Martin (1993). Vpu always appears with Env on bicistronic mRNAs, therefore Env1 is considered as the most abundant transcript for Vpu as well.

Gene name	Start codon	Stop codon	Length	Most abundant transcript: comprising exons	5´UTR Length
gag-pol	336	4642	4307	Full length primary transcript	335
vif	4587	5165	579	Vif2: 1 + 2E	417
vpr	5105	5396	292	Vpr1 (2kb mRNA): 1 + 3a + 7 Vpr3 (4kb mRNA): 1 + 3E Note: both transcripts have the same 5´UTR	458
tat	5377	7970	261 (excluding the intron)	Tat1 (2kb mRNA): 1 + 4 + 7 Tat5 (4kb mRNA): 1 + 4E Note: both transcripts have the same 5´UTR	342
rev	5516	8199	351 (excluding the intron)	Rev2: 1 + 4a + 7	304
vpu	5608	5856	249	Env1: 1 + 5E	374
env	5771	8341	2571	Env1: 1 + 5E	537
nef	8343	8963	621	Nef2: 1 + 5 + 7	776

Calculating the observed and expected number of AUGs in optimal context in the HIV-1 concatenated 5´UTR

Except for parts of the terminal LTR regions, the rest of the HIV-1 viral genome is made up almost entirely of overlapping coding regions. The only exception is a short 16-nucleotide sequence between the major splice donor site 4 (SD4) and the *vpu* start codon (see figure 1 in Purcell and Martin, 1993). It should be noted that due to the complex and overlapping nature of the HIV-1 genome, all non-coding segments, except exon1, overlap with a coding sequence (e.g. non-coding exons 2 and 3 are in the *gag* and *vif* ORFs, respectively). They are, however, part of a 5´UTR in the major HIV-1 transcripts.

The concatenated 5´UTR region of HIV-1_{NC1802} was assembled by connecting the following seven segments; exon1 (1:289), beginning of exon2E up to *vif* start codon (4459:4586), beginning of exon3a up to *vpr* start codon (4936:5104), beginning of exon4 up to *tat* start codon (5324:5376), beginning of exon4a up to *rev* start codon (5501:5515), beginning of exon5E up to *env* start codon (5523:5770), and beginning of exon7 up to *nef* start codon (7925:8342). The resulting sequence is 1320 nucleotides long. In addition to aforementioned statistical analyses, the concatenated 5´UTR was divided into nucleotide triplets. In order to test the significance of deviation of the observed number of AUG_{-3R} from the expected, two-way tables were constructed which contained the observed and expected numbers of AUG_{-3R} and Non-AUG_{-3R} triplets in the concatenated 5´UTR, and the following chi-square test was conducted;

$$X^2 = \frac{(NO_{AUG_{-3R}} - NE_{AUG_{-3R}})^2}{NE_{AUG_{-3R}}} + \frac{(NO_{NonAUG_{-3R}} - NE_{NonAUG_{-3R}})^2}{NE_{NonAUG_{-3R}}}$$

where "NO" refers to the observed number and "NE" refers to the expected number of AUG_{-3R} and Non-AUG_{-3R} triplets in the 5´UTR.

We performed a similar analysis in two other HIV-1 subtypes, namely subtypes A and C. A sequence of HIV-1 subtype A was selected by

specifying "A" as the *subtype*, "HIV-1" as the *organism*, and "complete genome" as the *genomic region* in the LANL HIV sequence database (Kuiken et al. 2003) search interface at (http://www.hiv.lanl.gov/components/sequence/HIV/combined_search_s_tree/search.html). We then selected the first sequence in the returned list of sequences (i.e. NCBI Nucleotide: AB098330) for our analysis. The same approach was used to select a HIV-1 subtype C sequence (i.e. NCBI Nucleotide: AB023804), by specifying "C" as the *subtype* in the search interface. The concatenated 5´UTR in these two sequences were assembled by aligning them against the HIV-1_{NC1802} sequence and joining the corresponding aforementioned seven segments. The resulting concatenated 5´UTR regions were comprised of 1351 and 1319 nucleotides for HIV-$1_{AB098330}$ and HIV-$1_{AB023804}$, respectively.

Calculating the observed number of AUGs in optimal contexts in 5´UTR of publicly available HIV-1 sequences

We used the aforementioned LANL HIV sequence database search interface to extract all available HIV-1 sequences that contain exon 1, which is shared among all HIV-1 transcripts. This region was selected by specifying 456 and 743 as the start and end of the genomic region, respectively, and selecting HIV-1 as the organism at the search interface. As of July 2008, there were 1142 such sequences, which were downloaded in FASTA format. Using a Python script, we calculated the frequency of AUGs in optimal contexts in these sequences.

Results and Discussion

Selective pressure against AUG usage in optimal context in HIV-1 5´UTRs

Table 3 shows the nucleotide frequencies of the 5´UTRs in the major HIV-1 transcripts (see methods). With G being the most frequent nucleotide in most cases, these frequencies do not conform to the HIV-1_{NC1802} A-rich genome with 36% A, 18% C, 24% G, and 22% U, obtained using the DAMBE program. The observed and expected numbers of AUGs in optimal context in the 5´UTR of HIV-1 major transcripts are shown in table 4. In all cases, the observed number is less than the expected value. In fact, other than the AUG_{-3R} located in the 5´UTR of env, which corresponds to the upstream vpu start codon (GuaAUGc), there are no other AUGs in optimal context in the 5´UTR regions.

Selective pressure against AUG usage in optimal context in HIV-1 concatenated 5´UTRs

As previously mentioned, the first 289 nucleotides are shared among all HIV-1 5´UTR sequences. We therefore decided to assess the selective pressure against AUG usage in optimal context in the concatenated non-overlapping 5´UTR sequence comprised of 1320 nucleotides. This sequence contains 32.6% A, 19.9% C, 26.4% G, and 21.1% U, which match the A-richness and C-poorness of the HIV-1_{NC1802} genome. There are no AUG_{-3R+4G} and AUG_{-3Y+4G} observed in the concatenated 5´UTR despite the expected numbers of 3.7 and 2.6,

Table 3. Nucleotide frequencies of the 5´UTR of major HIV-1 mRNAs.

Name	A	C	G	U	Sum	pA	pC	pG	pU
gag/pol	82	79	104	70	335	0.2448	0.2358	0.3104	0.209
vif	119	92	122	84	417	0.2854	0.2206	0.2926	0.2014
vpr	129	107	122	100	458	0.2817	0.2336	0.2664	0.2183
tat	85	81	105	71	342	0.2485	0.2368	0.307	0.2076
rev	68	79	90	67	304	0.2237	0.2599	0.2961	0.2204
vpu	98	91	108	77	374	0.262	0.2433	0.2888	0.2059
env	172	109	141	115	537	0.3203	0.203	0.2626	0.2142
nef	214	177	224	161	776	0.2758	0.2281	0.2887	0.2075

Table 4. The expected and observed number of AUGs in HIV-1 5′UTRs of major transcripts in different optimal contexts.

Name	NExp. AUG_{-3R}	NObs. AUG_{-3R}	NExp. AUG_{-3R+4G}	NObs. AUG_{-3R+4G}	NExp. AUG_{-3Y+4G}	NObs. AUG_{-3Y+4G}
gag/pol	2.91	0	0.90	0	0.72	0
vif	4.01	0	1.17	0	0.85	0
vpr	4.07	0	1.08	0	0.89	0
tat	2.96	0	0.91	0	0.73	0
rev	2.27	0	0.67	0	0.62	0
vpu	3.17	0	0.91	0	0.74	0
env	5.59	1	1.46	0	1.05	0
nef	7.19	0	2.07	0	1.60	0

respectively, which implies a selective pressure against these AUGs.

In the case of AUG_{-3R}, while the expected number is 14.1, there is only one observed AUG_{-3R} in the concatenated 5′UTR. The significance of the difference between the observed and expected number in this case was measured by analyzing the sequence as consisting of 440 (≈ 1320/3) AUG_{-3R} and Non-AUG_{-3R} triplets (Table 5). A χ^2 test revealed that the difference is significant ($X^2 = 12.6$, $p < 0.0005$), indicating a selective pressure against AUG_{-3R} in the 5′UTR.

Since HIV-1_{NC1802} sequence belongs to subtype B, a similar analysis was performed on HIV-1 subtype A (HIV-$1_{AB098330}$) and subtype C (HIV-$1_{AB023804}$). Subtypes A, B, and C account for 12%, 10%, and 50% of all infections worldwide, respectively (Hemelaar et al. 2006). Different HIV-1 subtypes may possess distinct 5′UTR secondary structures which could potentially affect the selective pressure against AUG usage in optimal context. As in the case of HIV-1_{NC1802}, there are no AUG_{-3R+4G} and AUG_{-3Y+4G} observed in the concatenated 5′UTR of HIV-$1_{AB098330}$ (expected numbers of 2.3 and 1.6, respectively) and HIV-$1_{AB023804}$ (expected numbers of 2.2 and 1.5, respectively) sequences. Furthermore, despite the expected number of 14.7 for AUG_{-3R} in HIV-$1_{AB098330}$ 5′UTR, there is only one AUG_{-3R} in this region ($X^2 = 13.2$, $p < 0.0005$). Similarly, despite an expected number of 14.8 for AUG_{-3R} in HIV-$1_{AB023804}$ 5′UTR, there is only one AUG_{-3R} in this region ($X^2 = 13.3$, $p < 0.0005$). These results point to a selective pressure against AUG_{-3R} in the 5′UTR of HIV-1 subtypes A and C.

Table 5. The expected and observed number of AUG_{-3R} and Non-AUG_{-3R} triplets in the concatenated HIV-1 5′UTR.

	Concatenated 5′UTR		
	AUG_{-3R} triplets	Non-AUG_{-3R} triplets	Sum
Observed	1	439	440
Expected	14.1	425.9	440

Selective pressure against AUG usage in optimal context in 5′UTR of publicly available HIV-1 sequences

In order to expand our analysis, we analyzed exon 1, which is part of the 5′UTR that is shared among HIV-1 transcripts, in all publicly available HIV-1 sequences containing this region. From the 1142 downloaded sequences, 21 (1.8%) contained AUG_{-3R} (one subtype A, twelve subtype B, seven subtype C, and one unknown subtype), 9 (0.8%) contained AUG_{-3Y+4G} (two subtype B and seven subtype C), and 2 (0.2%) contained AUG_{-3R+4G} (both subtype B).

Overall, the results presented in this study demonstrate a selective pressure against the AUG_{-3R} and suggest possible selection against AUG_{-3R+4G} and AUG_{-3Y+4G} in the 5′UTR region of HIV-1_{NC1802}, which corroborates the results obtained by other groups regarding the underrepresentation of AUGs in HIV-1 5′UTR (Das et al. 1998). Our similar findings in sequences of HIV-1 subtype A and C, suggest the generality of a selective pressure against AUG usage in optimal context in the

5′UTR of different HIV-1 subtypes. These results further support the CDRSM hypothesis of translation initiation in HIV-1, i.e. they are consistent with translation initiation scenarios 1 and 3, but contradict the prediction from scenario 2 (Table 1). It should be noted that the possibility of usage of non-AUG translation initiation codons by HIV-1, such as CUG (Yilmaz et al. 2006), is not considered in our approach. While nearly all eukaryotic genes use AUG as the initiation codon, non-AUG codons are sometimes found to initiate translation initiation. For example, a CUG codon in Moloney murine leukemia virus (Mo-MuLV) genome, located in the sequence context of ACCCUGG at positions 354–359, appears to serve as an alternative initiation codon (Prats et al. 1989). However, the initiation efficiency of CUG in Mo-MuLV have not been evaluated by mutation analysis such as that done by Kozak (Kozak, 1986; Kozak, 1997), and subsequent analyses suggest that the CUG might be used as in an internal ribosome entry mechanism instead of the conventional cap-dependent translation mechanism (Berlioz and Darlix, 1995; Vagner et al. 1995). It should also be noted that in NCBI's Mo-MuLV sequence (NCBI Genome: NC_001501), there is an AUG instead of CUG at positions 357–359, therefore the original finding of a CUG at positions 357–359 may have been incorrect. Non-AUG initiation codons in eukaryotes are typically involved in certain regulatory mechanisms and are poorly translated by the cap-dependent translation mechanisms. In short, cap-dependent translation implies that only AUG, but not non-AUG codons, should be underrepresented at the 5′UTR of mRNAs. This is apparent when one analyzes protein-coding genes of any eukaryotic genome. Thus the frequency of CUG translation initiation codons is irrelevant to our present study.

While there are some experimental results indicating the presence of IRES in HIV-1 (Buck et al. 2001; Brasey et al. 2003), HIV-2 (Herbreteau et al. 2005), and simian immunodeficiency virus (SIV) (Nicholson et al. 2006), currently used methodology for validating IRES remains problematic (Kozak, 2001). Our results suggest that even though CDRSM and SSS/IRES-dependent translation initiation mechanisms may co-exist, the latter is either a weak mechanism or a recent innovation that has not yet produced evolutionary consequences on the distribution of AUGs in optimal context in the 5′UTR. These speculations call for further experimental analysis of IRES activity in HIV-1.

Acknowledgments

This project is supported by NSERC's Discovery, Strategic and RTI Grants. We thank G. Drouin, M. Dumontier, N. Goto, X. Yao, and G. Palidwor for comments and discussion.

Disclosure

The authors report no conflicts of interest.

References

Altschul, S.F., Gish, W., Miller, W. et al. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215:403–10.

Anderson, J.L., Johnson, A.T., Howard, J.L. et al. 2007. Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs. *J. Virol.*, 81:4664–76.

Arrigo, S.J., Weitsman, S., Zack, J.A. et al. 1990. Characterization and expression of novel singly spliced RNA species of human immunodeficiency virus type 1. *J. Virol.*, 64:4585–8.

Baudin, F., Marquet, R., Isel, C. et al. 1993. Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.*, 229:382–97.

Beerens, N. and Berkhout, B. 2002. The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. *J. Virol.*, 76:2329–39.

Berkhout, B. 1996. Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucleic. Acid. Res. Mol. Biol.*, 54:1–34.

Berkhout, B. and van Wamel, J.L. 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *Rna.*, 6:282–95.

Berlioz, C. and Darlix, J.L. 1995. An internal ribosomal entry mechanism promotes translation of murine leukemia virus gag polyprotein precursors. *J. Virol.*, 69:2214–22.

Brasey, A., Lopez-Lastra, M., Ohlmann, T. et al. 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.*, 77:3939–49.

Buck, C.B., Shen, X., Egan, M.A. et al. 2001. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J. Virol.*, 75:181–91.

Cimarelli, A. and Darlix, J.L. 2002. Assembling the human immunodeficiency virus type 1. *Cell Mol. Life Sci.*, 59:1166–84.

Das, A.T., van Dam, A.P., Klaver, B. et al. 1998. Improved envelope function selected by long-term cultivation of a translation-impaired HIV-1 mutant. *Virology*, 244:552–62.

Deffaud, C. and Darlix, J.L. 2000. Rous sarcoma virus translation revisited: characterization of an internal ribosome entry segment in the 5′ leader of the genomic RNA. *J. Virol.*, 74:11581–8.

Furtado, M.R., Balachandran, R., Gupta, P. et al. 1991. Analysis of alternatively spliced human immunodeficiency virus type-1 mRNA species, one of which encodes a novel tat-env fusion protein. *Virology*, 185:258–70.

Geballe, A.P. and Gray, M.K. 1992. Variable inhibition of cell-free translation by HIV-1 transcript leader sequences. *Nucleic. Acids Res.*, 20:4291–7.

Guatelli, J.C., Gingeras, T.R. and Richman, D.D. 1990. Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. *J. Virol.*, 64:4093–8.

Hemelaar, J., Gouws, E., Ghys, P.D. et al. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *Aids*, 20:W13–23.

Herbreteau, C.H., Weill, L., Decimo, D. et al. 2005. HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat. Struct. Mol. Biol.*, 12:1001–7.

Ho, D.D., Rota, T.R. and Hirsch, M.S. 1986. Infection of monocyte/macrophages by human T lymphotropic virus type III. *J. Clin. Invest.*, 77:1712–5.

Jang, S.K., Krausslich, H.G., Nicklin, M.J. et al. 1988. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.*, 62:2636–43.

Klasens, B.I., Thiesen, M., Virtanen, A. et al. 1999. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. *Nucleic. Acids Res.*, 27:446–54.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44:283–92.

Kozak, M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic. Acids. Res.*, 15:8125–48.

Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.*, 108:229–41.

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.*, 16:2482–92.

Kozak, M. 2001. New ways of initiating translation in eukaryotes? *Mol. Cell Biol.*, 21:1899–907.

Kuiken, C., Korber, B. and Shafer, R.W. 2003. HIV sequence databases. *AIDS. Rev.*, 5:52–61.

Laughrea, M. and Jette, L. 1994. A 19-nucleotide sequence upstream of the 5′ major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA. *Biochemistry*, 33:13464–74.

Lever, A., Gottlinger, H., Haseltine, W. et al. 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.*, 63:4085–7.

Lopez-Lastra, M., Gabus, C. and Darlix, J.L. 1997. Characterization of an internal ribosomal entry segment within the 5′ leader of avian reticuloendotheliosis virus type A RNA and development of novel MLV-REV-based retroviral vectors. *Hum. Gene. Ther.*, 8:1855–65.

Miele, G., Mouland, A., Harrison, G.P. et al. 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.*, 70:944–51.

Nicholson, M.G., Rue, S.M., Clements, J.E. et al. 2006. An internal ribosome entry site promotes translation of a novel SIV Pr55(Gag) isoform. *Virology*, 349:325–34.

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.*, 275:11899–906.

Parkin, N.T., Cohen, E.A., Darveau, A. et al. 1988. Mutational analysis of the 5′ non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *Embo. J.*, 7:2831–7.

Pelletier, J. and Sonenberg, N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, 334:320–5.

Prats, A.C., De Billy, G., Wang, P. et al. 1989. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J. Mol. Biol.*, 205:363–72.

Purcell, D.F. and Martin, M.A. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.*, 67:6365–78.

Schwartz, S., Felber, B.K., Benko, D.M. et al. 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J. Virol.*, 64:2519–29.

Schwartz, S., Felber, B.K. and Pavlakis, G.N. 1992. Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. *Mol. Cell Biol.*, 12:207–19.

Selby, M.J., Bain, E.S., Luciw, P.A. et al. 1989. Structure, sequence, and position of the stem-loop in tar determine transcriptional elongation by tat through the HIV-1 long terminal repeat. *Genes. Dev.*, 3:547–58.

SenGupta, D.N., Berkhout, B., Gatignol, A. et al. 1990. Direct evidence for translational regulation by leader RNA and Tat protein of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.*, 87:7492–6.

Svitkin, Y.V., Pause, A. and Sonenberg, N. 1994. La autoantigen alleviates translational repression by the 5′ leader sequence of the human immunodeficiency virus type 1 mRNA. *J. Virol.*, 68:7001–7.

Vagner, S., Waysbort, A., Marenda, M. et al. 1995. Alternative translation initiation of the Moloney murine leukemia virus mRNA controlled by internal ribosome entry involving the p57/PTB splicing factor. *J. Biol. Chem.*, 270:20376–83.

Xia, X. 2001. Data analysis in molecular biology and evolution. Boston: Kluwer Academic Publishers.

Xia, X. 2007. Bioinformatics and the cell: modern computational approaches in genomics, proteomics and transcriptomics. *Springer.*

Xia, X. and Xie, Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.*, 92:371–3.

Yilmaz, A., Bolinger, C. and Boris-Lawrie, K. 2006. Retrovirus translation initiation: Issues and hypotheses derived from study of HIV-1. *Curr. HIV. Res.*, 4:131–9.