HIV-1 Modulates the tRNA Pool to Improve Translation Efficiency

Anna van Weringh,[1,2] Manon Ragonnet-Cronin,[1,2] Erinija Pranckeviciene,[1,2] Mariana Pavon-Eternod,[3] Lawrence Kleiman,[4] and Xuhua Xia*[,1,2,5]

[1]Department of Biology, University of Ottawa, Ottawa, Ontario, Canada
[2]Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Ontario, Canada
[3]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago
[4]Lady Davis Institute for Medical Research and McGill AIDS Centre, Jewish General Hospital, Montreal, Quebec, Canada
[5]Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, Canada
*Corresponding author: E-mail: xxia@uottawa.ca.
Associate editor: Helen Piontkivska

Abstract

Despite its poorly adapted codon usage, HIV-1 replicates and is expressed extremely well in human host cells. HIV-1 has recently been shown to package non-lysyl transfer RNAs (tRNAs) in addition to the tRNA^{Lys} needed for priming reverse transcription and integration of the HIV-1 genome. By comparing the codon usage of HIV-1 genes with that of its human host, we found that tRNAs decoding codons that are highly used by HIV-1 but avoided by its host are overrepresented in HIV-1 virions. In particular, tRNAs decoding A-ending codons, required for the expression of HIV's A-rich genome, are highly enriched. Because the affinity of Gag-Pol for all tRNAs is nonspecific, HIV packaging is most likely passive and reflects the tRNA pool at the time of viral particle formation. Codon usage of HIV-1 early genes is similar to that of highly expressed host genes, but codon usage of HIV-1 late genes was better adapted to the selectively enriched tRNA pool, suggesting that alterations in the tRNA pool are induced late in viral infection. If HIV-1 genes are adapting to an altered tRNA pool, codon adaptation of HIV-1 may be better than previously thought.

Key words: HIV-1, tRNA, codon usage, translation efficiency, codon–anticodon adaptation.

Introduction

Viruses use the host translational machinery to translate their own mRNA. As a consequence, their codon usage is under selective pressure to adapt to the host transfer RNA (tRNA) pool (Sharp and Li 1987). In RNA viruses in general and HIV-1 in particular, adaptation to the host is poor despite this selection (Bahir et al. 2009). For example, according to a recent compilation of tRNAs in human genome (Chan and Lowe 2009), the AUC codon can be translated by 17 tRNA^{Ile} species, that is, 14 tRNA$^{Ile/IAU}$, and 3 tRNA$^{Ile/GAU}$, AUU can be translated by 14 tRNA$^{Ile/IAU}$ species, whereas AUA can be translated by 5 tRNA$^{Ile/UAU}$ species only. In agreement with this, human genes code Ile mostly by AUC and least by AUA. In contrast, HIV-1 genes code Ile mostly by AUA and least by AUC (Haas et al. 1996; Nakamura et al. 2000). The poor codon adaptation of HIV-1 reduces the translation efficiency of HIV-1 genes. Modifying HIV-1 codon usage according to host codon usage has been shown to increase the production of viral proteins (Haas et al. 1996; Ngumbela et al. 2008).

Conventionally, the poor concordance between HIV-1 and host codon usage is explained by the mutation hypothesis invoking high A-biased mutation at the third codon position of HIV-1 genes (Jenkins and Holmes 2003). The A-bias is mediated by the error-prone reverse transcriptase (Martinez et al. 1994; Vartanian et al. 2002) and the human APOBEC3 protein (Yu et al. 2004). The frequency of A can reach up to 40% in some HIV-1 genomes (Vartanian et al. 2002), resulting in a preponderance of A-ending codons, which are typically rarely used in the host genes (Sharp 1986; Kypr and Mrazek 1987). Although there have been claims that the A-richness in a parasitic or symbiotic genome may confer some selective advantage (Xia 1996; Keating et al. 2009), further empirical substantiation is required. In short, although avoiding A-ending codons may improve translation elongation efficiency, A-biased mutations lead to an overrepresentation of A-ending codons in HIV-1 genes.

Recent empirical studies on tRNA packaging in HIV-1, however, suggest an additional hypothesis to explain the codon usage of HIV-1. HIV-1 is known to package host tRNAs, specifically tRNA^{Lys} and tRNA^{Ile} (Jiang et al. 1993; Pavon-Eternod et al. 2010). tRNA^{Lys} is used by HIV-1 as a primer to initiate reverse transcription (Jiang et al. 1993; Marquet et al. 1995; Kleiman and Cen 2004; Kleiman et al. 2004) and is selectively packaged by HIV-1, through the interaction of lysyl-tRNA synthetase (LysRS) with Gag-Pol, Gag, and viral genomic RNA (Javanbakht et al. 2003; Kleiman et al. 2010). More specifically, Gag alone is capable

of packaging LysRS, but the presence of Gag-Pol both enhances the packaging of LysRS and is needed to incorporate tRNA^{Lys} (Mak et al. 1994; Khorchid et al. 2000; Javanbakht et al. 2003; Kleiman et al. 2010). However, both the function of tRNA^{Ile} and the mechanisms that lead to its incorporation into HIV-1 virions are unknown.

A recent tRNA microarray study (Pavon-Eternod et al. 2010) demonstrated that HIV-1 packages a variety of tRNAs in addition to tRNA^{Lys} and tRNA^{Ile}. Inexplicably, specific tRNA species were packaged within each codon family. For example, tRNA$^{Ile/UAU}$ was highly packaged by HIV-1 as compared with tRNA$^{Ile/IAU}$. As aforementioned, the AUA codon read by tRNA$^{Ile/UAU}$ is rarely used in human genes but is the most common codon for Ile in HIV-1. The authors therefore speculated that the incorporation of the rare tRNA$^{Ile/UAU}$ could be related to the decoding of the unfavorable AUA codon in HIV-1 genes (Pavon-Eternod et al. 2010).

In this paper, we further tested whether tRNA packaging is related to viral translation through the comparison of packaged tRNAs to the codon usage of HIV-1. tRNAs decoding poorly adapted codons, defined as codons highly used by HIV-1 but strongly avoided by its host, were overrepresented in HIV-1 virions. This suggests that the tRNA pool for translating HIV-1 genes may be selectively altered to favor the translation of HIV-1 genes and that this altered tRNA pool is reflected in the tRNA packaged by HIV-1 virions. HIV-1 early genes were found to have similar codon usage as highly expressed host genes, but HIV-1 late genes do not. This is consistent with our interpretation that alterations to the tRNA pool occur late in infection and that HIV-1 early and late genes evolved different codon usage in response to their respective tRNA pools.

Methods

The microarray data used in this study were collected in T. Pan's laboratory, following an experimental protocol previously described in detail (Pavon-Eternod et al. 2010). In brief, tRNAs packaged in HIV-1 virions and in Gag viral-like particles (GagVLP) were quantified by microarray. GagVLP particles are HIV-1 mutant particles deficient in the frameshift required to translate Gag-Pol. The tRNA content of HEK293T cells was also measured, as cellular tRNAs are the source of tRNAs packaged into HIV-1. Ratios of viral tRNAs compared with cellular tRNA content were profiled to identify tRNAs that are concentrated in HIV-1 viral particles relative to normal host cells.

The subtype B HXB2 reference genome for HIV-1 (NC_001802) was downloaded from the National Centre for Biotechnology Information (NCBI). Coding sequences were extracted, and the relative synonymous codon usage (RSCU) was computed using DAMBE (Xia 2001; Xia and Xie 2001). RSCU is a normalized index of codon usage bias (Sharp and Li 1987). It has a value of zero for unused synonymous codons, a value of one for equally used synonymous codons and a maximum of n, where n is the number of synonymous codons in the codon family. Codon families

read by only one codon (AUG for Met and UGG for Trp) always have a value of 1 and were therefore excluded from analyses. Stop codons were also excluded from RSCU analyses. Codon adaptation index or codon adaptive index (CAI) (Sharp and Li 1987) was computed by the improved implementation in DAMBE (Xia 2007), using the codon usage table of highly expressed human genes compiled in the file Ehuman.cut in the EMBOSS package (Rice et al. 2000).

Both Gag and Gag-Pol polyproteins are translated from the full-length HIV-1 RNA genome, with Gag-Pol translated with a −1 ribosomal slippage at the slippery sequence UUUUUUA starting at position 1631 in the HXB2 reference genome. The two polyproteins therefore share the N-terminal portion but differ in the C-terminal portion. "Pol" in this manuscript refers to the C-terminal portion of Gag-Pol that differs from Gag to avoid data dependence, that is, it starts from position 1637 in the HXB2 reference genome.

Based on the tRNA data collected and RSCU values calculated, we defined two indices for viral codon usage and tRNA packaging:

$$l_{\text{codon}.i} = \log_2\left(\frac{\text{RSCU}_{i.\text{HIV}-1}}{\text{RSCU}_{i.\text{Human}}}\right), \qquad (1)$$

$$l_{\text{tRNA}.i} = \log_2\left(\frac{\text{tRNA}_{i.\text{HIV}-1}}{\text{tRNA}_{i.\text{GagVLP}}}\right), \qquad (2)$$

where i indicates sense codons (excluding AUG and UGG codon) in equation (1) and their corresponding tRNA in equation (2). The ratio tRNA$_{i.\text{HIV-1}}$/tRNA$_{i.\text{GagVLP}}$ for the calculation of $l_{\text{tRNA}.i}$ is obtained as the ratio of the two relative tRNA abundances, tRNA$_{i.\text{HIV-1}}$/tRNA$_{i.\text{Cell}}$ over tRNA$_{i.\text{GagVLP}}$/tRNA$_{i.\text{Cell}}$, quantified in the microarray experiment (Pavon-Eternod et al. 2010). $l_{\text{codon}.i}$ is a measure of the requirement of specific tRNAs by HIV-1 to translate codon i. For example, a large $l_{\text{codon.AUA}}$ suggests a strong need for HIV-1 to have AUA-decoding tRNAs. l_{tRNA} measures the degree of enrichment of each tRNA in HIV-1 relative to that in GagVLP whose packaged tRNA is similar to that of the uninfected host cell (Pavon-Eternod et al. 2010).

Because HIV-1 genes are expressed in a temporal manner, the codon usage in early (*tat*, *rev*, and *nef*) and late genes (*gag*, *pol*, *env*, *vif*, *vpu*, and *vpr*) in HIV-1 (Cullen 1991) was examined for differential codon adaptation. Because the early genes are short and RSCU is not meaningful with few codons per codon family, we set a threshold to include codon families with at least 14 codons (so as not to exclude the Ile codon family). Excluding rarely used codon families should lead to little loss of information because selection for optimal codon usage would be weak for such codon families in any case (Duret 2000).

The simian immunodeficiency virus genome (SIV, M58410) was downloaded from NCBI as well as reference HIV-1 sequences for subtypes A1 (AB253421), A2 (AF286237), C (U52953), and J (EF614151) from the Los Alamos National Database. The human T-lymphotropic virus-1 (HTLV-1, NC_001436), another retrovirus, was also obtained from NCBI for comparative purposes. For HTLV-1, early genes are *rex* and *tax*, and late genes are *gag*, *pol*, and
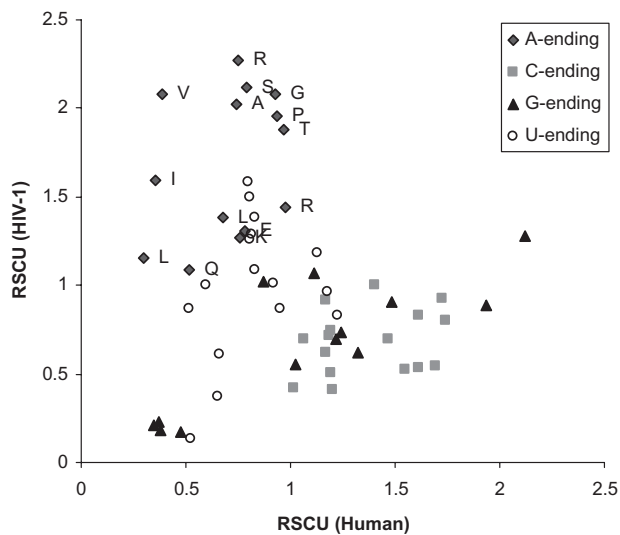
FIG. 1. RSCU of HIV-1 compared with RSCU of highly expressed human genes. Data points for codons ending with A, C, G, or U are annotated with different combinations of colors and symbols. A-ending codons exhibit strong discordance in their usage between HIV-1 and human and are annotated with their coded amino acids.

pro (Yoshida 2005). For SIV, early genes are *tat*, *rev*, and *nef*, and late genes are *gag*, *pol*, *vif*, *vpx*, and *env*.

When normality or linearity assumptions were violated, we provide output from both parametric and nonparametric methods. All statistical tests are two tailed and were carried out in SAS (SAS Institute Inc. 1989).

Results

Poorly Adapted HIV-1 Codons are Mostly A-ending

Viral codon adaptation is considered poor when codons highly avoided by the host genes are highly used by viral genes, as host codon usage is considered a reliable indicator

Table 1. Relationship between codon usage measured by RSCU for human and HIV-1 ($RSCU_{Human}$ and $RSCU_{HIV}$) and packaging of host tRNA by HIV-1. tRNA abundance data were kindly provided by Dr T.P. Rank(I_{codon}) and Rank(I_{tRNA}) are significantly and positively correlated ($r = 0.5780$, $P = 0.0304$).

AA (Codon)	$RSCU_{Human}$	$RSCU_{HIV}$	Rank (I_{codon})	$tRNA_{HIV}$[a]	$tRNA_{GagVLP}$[b]	Rank (I_{tRNA})
Arg(AGA)	0.97	1.44	8	0.0494	0.0277	4
Arg(AGG)	1.03	0.56	4	0.0660	0.0544	2
Ile(AUA)	0.24	1.59	14	1.3397	0.0614	14
Ile(AUY)	2.64	1.41	3	0.2672	0.1371	5
Leu(UUA)	0.68	1.38	11	0.0900	0.0374	8
Leu(UUG)	1.32	0.62	1	0.0450	0.0496	1
Lys(AAA)	0.76	1.27	9	0.6405	0.0340	13
Lys(AAG)	1.24	0.73	5	1.0081	0.0601	12
Gly(GGA)	0.93	2.08	12	0.0708	0.0290	9
Gly(GGB)	3.07	1.92	6	0.2016	0.0486	10
Val(GUA)	0.39	2.08	13	0.0662	0.0282	7
Val(GUB)	3.61	1.92	2	0.0739	0.0485	3
Thr(ACA)	0.97	1.94	10	0.0481	0.0215	6
Thr(ACB)	3.03	2.06	7	0.2522	0.0347	11

[a] $tRNA_{HIV}$: the relative tRNA abundance of HIV-1 virion versus human HEK293T cells.
[b] $tRNA_{GagVLP}$: the relative tRNA abundance of GagVLP versus human HEK293T cells.

of the availability of cognate tRNAs (Bulmer 1987). To identify HIV-1 codons poorly adapted to its host, the RSCU values for HIV-1 were plotted against human RSCU values (fig. 1). RSCU measures the degree of bias in codon usage for each codon family. The codon usage of HIV-1 genes correlated poorly with that of host genes (Pearson $r = -0.1470$, $P = 0.2665$; Spearman $r = 0.1829$, $P = 0.1657$, fig. 1). A-ending codons are particularly highly used by HIV-1 but strongly avoided by the host.

## Selectively Packaged tRNAs Decode Poorly Adapted Codons

Given that A-biased mutation dramatically reduces codon adaptation in HIV-1 genes (Berkhout et al. 2002) and that most discordance between HIV-1 codon usage and host codon usage is due to A-ending codons (fig. 1), we investigated whether tRNAs packaged by HIV-1 were those that decode A-ending codons, by using the indices defined in the previous section, I_{codon} and I_{tRNA}. The indices are such that codons with large I_{codon} values correspond to codons poorly adapted to human tRNA pool, and large values of I_{tRNA} correspond to tRNAs that are more concentrated in HIV-1 virion than in the normal host cell. If the tRNA pool during the translation of HIV-1 late genes is altered in favor of translating HIV-1 genes and if relative abundance of packaged tRNA reflect the relative abundance in the tRNA pool, then we would expect a positive correlation between I_{codon} and I_{tRNA}.

Based on the tRNA microarray data (Pavon-Eternod et al. 2010), seven pairs of I_{codon} and I_{tRNA} values for codon families (Arg, Ile, Leu, Lys, Gly, Val, and Thr) were computed (table 1). These seven families were the only ones that used probes on the tRNA microarray that distinguished A-ending from non–A-ending codons (Pavon-Eternod et al. 2010). Only families for which information on all tRNAs was available were studied in order to characterize the relationship between I_{codon} and I_{tRNA}.

I_{tRNA} and I_{codon} were ranked and found to be significantly positively correlated, with $r = 0.5780$ and $P = 0.0304$ (Pearson and Spearman correlations are identical for ranked data). $tRNA^{Lys}$ is known to be selectively packaged for the purpose of reverse transcription and its packaging should not be related to HIV-1 translation. The removal of Lys from the analysis resulted in an increased correlation between the ranked I_{codon} and I_{tRNA} to 0.6853 ($P = 0.0139$).

The data in table 1 may represent two groups: that of A-ending codons, whose I_{codon} is increased by the A-biased mutation, and that of non–A-ending codons, whose codon–anticodon adaptation is not disrupted by the A-biased mutation. In order to take into account the influence of A-biased mutation, ranked I_{codon} and I_{tRNA} values were, therefore, plotted separately for A-ending and non–A-ending codons (fig. 2). A strong positive correlation between I_{codon} and I_{tRNA} was maintained in both A-ending and non–A-ending groups (fig. 2) suggesting that tRNA packaging is related to codon usage of all codons, not only
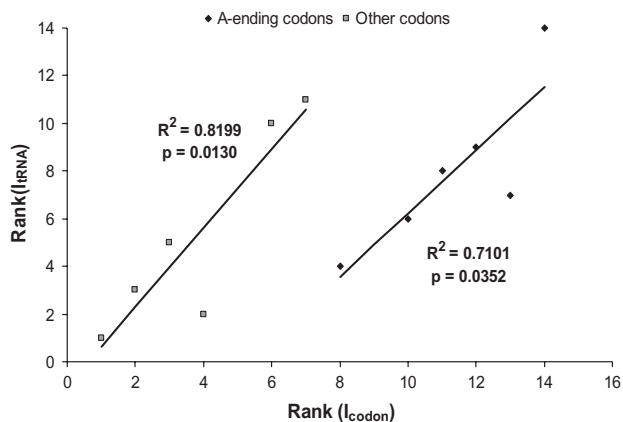
FIG. 2. Relationship between ranked I_{codon} and I_{tRNA} for A-ending and non–A-ending codons. Correlation lines are plotted with R^2 values to indicate the goodness of fit. Lys data were not included because tRNA^{Lys} enrichment is for the purpose of reverse transcription and may have little to do with codon–anticodon adaptation.

A-ending codons. We may view the regression line for the A-ending codons (fig. 2) as being pushed to the right by the A-biased mutation.

Differential Codon Adaptation between Early and Late HIV-1 Genes

Packaging of tRNAs into HIV-1 virions occurs late in the HIV-1 life cycle, when translation of the late genes has occurred, as these are the structural proteins that make up HIV-1 particles. We therefore investigated whether there were differences in the codon usage between HIV-1 early and late genes. If the tRNA pool is altered during the translation of HIV-1 late genes, then HIV-1 early and late genes would adapt to different tRNA pools and may exhibit different codon adaptation. In particular, the early genes are expected to have codon adaptation similar to that of highly expressed human genes because they are expected to share the same tRNA pool for translation.

HIV-1 early genes have RSCU significantly and positively correlated with that of human genes (fig. 3a, Pearson $r = 0.32$, $P = 0.0413$; Spearman $r = 0.2776$, $P = 0.0789$). In contrast, HIV-1 late genes have RSCU negatively correlated with that of human genes, although the correlation is not statistically significant (fig. 3a, Pearson $r = -0.1435$, $P = 0.3704$; Spearman $r = -0.1847$, $P = 0.2477$). These results are consistent for different HIV-1 subtypes A1, A2, C, and J, as well as for SIV, and are consistent with the interpretation that HIV-1 early and late genes may be translated in different tRNA pools and have evolved different codon usage in response to the different tRNA pools.

In corroboration, the CAI values of early genes are significantly higher than those of late genes when highly expressed human genes were used as a reference for computing CAI (two-sample t-test assuming unequal variances, $t = 2.8099$, degree of freedom [df] = 4, $P = 0.04832$, table 2). This further suggests that HIV-1 early genes have adapted to the tRNA pool, where host genes are normally translated and evolved codon usage similar to those of

highly expressed host genes. HIV-1 late genes, however, have evolved a different codon usage pattern.

Interestingly, in HTLV-1 that infects the same type of host cell as HIV-1, both early and late genes have their RSCU positively correlated with host RSCU, with similar slopes (fig. 3b). The positive correlation in RSCU between HTLV-1 and human genes is highly significant (Pearson $r = 0.4982$, $P < 0.0001$, Spearman $r = 0.4688$, $P = 0.0002$). Both HIV-1 and HTLV-1 are retroviruses with RNA genomes, but HTLV-1 is exceptional in that it does not have a strong A-biased mutation (van Hemert and Berkhout 1995; Van Dooren et al. 2004). HTLV-1 relies for the most part on the host polymerase to replicate through clonal expansion of infected cells rather than undergoing iterative replication cycles like HIV-1 (Strebel 2005). The substitution rate of HTLV-1 is consequently lower, about 5.2×10^{-6} substitutions/site/year (Hanada et al. 2004; Van Dooren et al. 2004), whereas that of HIV-1 is around 2.5×10^{-3} substitutions/site/year (Hanada et al. 2004). Thus, although HTLV-1 infects the same cells as HIV-1, that is, human CD4+ T cells (Rimsky et al. 1988) and both viruses are therefore subject to the same selective pressures on codon usage by the host tRNA pool, mutations are less likely to disrupt codon–anticodon adaptation in HTLV-1 than in HIV-1 as they occur at a lower rate in the former. The contrast between HIV-1 and HTLV-1 codon usage (fig. 3) suggests that the difference in codon adaptation between HIV-1 early and late genes (fig. 3a) may be a derived feature evolved after the divergence of the HIV-SIV lineage from HTLV-1 in response to the high A-biased mutation disrupting codon adaptation in HIV-1 late genes that need to be mass translated to produce viral particles.

Discussion

tRNA species reading poorly adapted codons (high I_{codon}) were found to be highly packaged (high I_{tRNA}) by HIV-1 virions, particularly tRNAs that read A-ending codons. As tRNAs that read A-ending codons are rare in human cells, their higher levels of packaging is surprising. It would be tempting to hypothesize that rare tRNAs are actively packaged and delivered into a newly infected cells to alter the tRNA pool. However, this hypothesis cannot be true for two reasons. First, reverse transcription and translation in HIV-1 are two spatially and temporally separate processes. Reverse transcription is completed early in the viral life cycle, soon after viral entry and uncoating, whereas viral translation occurs much later, particularly if the virus enters latency (Saksela et al. 1993). Second, the few tRNAs carried in a single HIV-virion (Huang et al. 1994; Halwani et al. 2004) would minimally affect the cellular tRNA pool and could therefore contribute little to the improvement of viral translation. We therefore suggest an alternative hypothesis. That is, the tRNA pool when HIV-1 late genes are translated is enriched with tRNAs decoding A-ending codons common in HIV-1 late genes. The packaging of tRNA into the HIV-1 virion is passive and reflects the relative tRNA abundance in the altered tRNA pool.
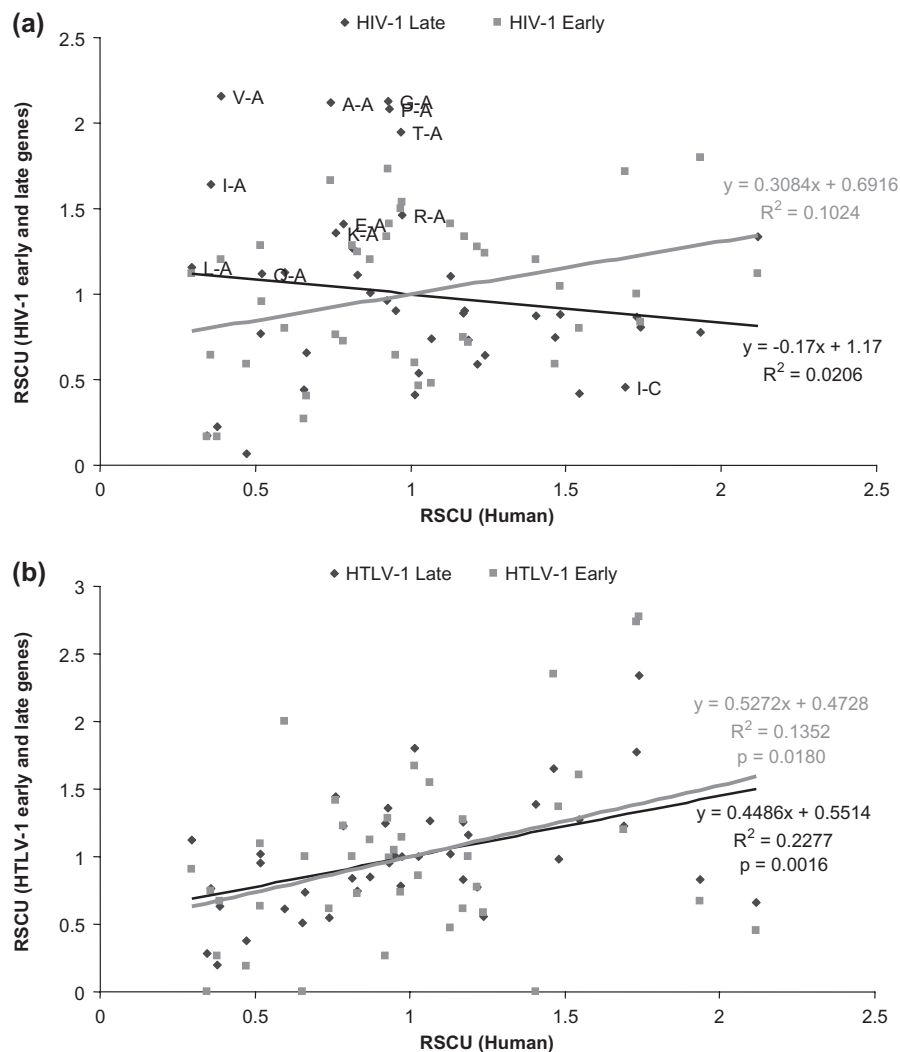
Fig. 3. RSCU of HIV-1 *a*) and HTLV-1 *b*) compared with RSCU of highly expressed human genes, with early (pink) and late (dark blue) genes shown with different colors. Linear regression lines are plotted with the fitted equations, R^2 and P values to indicate statistical significance.

The hypothesis of an altered tRNA pool for translating HIV-1 late genes explains the difference in codon adaptation between HIV-1 early and late genes (fig. 3*a*). The early genes are translated in the tRNA pool that normally translate cellular proteins and therefore have codon usage similar to that of highly expressed human genes (table 2, fig. 3*a*). The late genes are translated in the altered tRNA pool

Table 2. Frequency of A residues, length, and CAI for the three HIV-1 early (*tat*, *rev*, and *nef*) and five late (*gag-pol*, *vif*, *vpu*, *vpr*, and *env*) coding sequences (CDS).

Gene	CDS (bp)	CAI
tat	261	0.66875
rev	351	0.66211
nef	621	0.67523
gag	1503	0.62784
pol	3012	0.58139
vif	579	0.61941
vpr	291	0.64272
vpu	249	0.49068
env	2571	0.61924

with enriched tRNAs decoding A-ending codons and therefore would experience weakened selection against A-ending codons. This weakened selection, coupled with the A-biased mutation, explains the preponderance of the A-ending codons in HIV-1 late genes.

Our findings highlight a problem in empirical characterization of translation efficiency of HIV-1 genes. Several empirical studies have shown that modifying HIV-1 codon usage according to the codon usage of highly expressed host genes can substantially increase the production of viral proteins of HIV-1 late genes (Haas et al. 1996; Ngumbela et al. 2008). However, these studies were conducted with the normal tRNA pool in the host cell and did not take into consideration of the possible alteration of the host tRNA pool during the translation of HIV-1 late genes. Our results suggest that HIV-1 late genes may be translated more efficiently in the altered tRNA pool with enriched tRNA species decoding A-ending codons frequently used in HIV-1 late genes. Thus, the empirically characterized translation efficiency of HIV-1 late genes may be an underestimate.

So far we have interpreted the difference in codon adaptation between the early and the late HIV-1 genes as being caused by two different tRNA pools. That is, the HIV-1 early genes are translated in a tRNA pool when normal translation of host genes is still taking place, whereas the HIV-1 late genes are translated when the tRNA pool has been selectively enriched with tRNA species decoding A-ending codons in HIV-1 late genes. However, as one reviewer has pointed out, the difference in codon adaptation between early and late HIV-1 genes could also be due to a lower mutation rate in the early genes than in the late genes. This hypothesis is reasonable given that 1) A-biased mutation in HIV-1 disrupts codon–anticodon adaptation and 2) HTLV-1 that infects the same type of host cells but has a lower mutation rate than HIV-1 has its protein-coding genes exhibiting better codon–anticodon adaptation than those in HIV-1 (fig. 3b). Thus, differential mutation rate can plausibly result in different codon adaptation. If HIV-1 early genes have lower (A-biased) mutation rate than HIV-1 late genes, then we would expect the early genes to exhibit better codon adaptation than the late genes. The two hypotheses are not mutually exclusive and both have the same prediction (i.e., the A-ending codons are less frequent in HIV-1 early genes than in HIV-1 late genes), but which one might play a more significant role?

As there is no direct experimental data on whether HIV-1 early genes and late genes differ in mutation rate, we checked the mutation hypothesis by examining the frequency of A at codon positions 1 and 2. Many studies have shown that mutation bias is not only reflected in the third codon position but also in the first and second codon positions (Sueoka 1961; Lobry 2004). If A-biased mutation is stronger in the late genes than in the early genes, then A_{12} should be higher in the late genes than in the early genes. However, there is no significant differences in A_{12} between the early and the late genes, and A_{12} for the early *tat* gene (=0.3276) is in fact greater than A_{12} for two late genes vpr (=0.3144) and vpu (=0.3133). Although this result cannot exclude the possibility that the late genes have a higher A-biased mutation rate than the early genes, we are inclined to conclude that the mutation hypothesis is empirically not strong.

It is not known whether changes to the altered tRNA pool are actively induced by viral proteins or if this is part of a host response to viral infection. Host cells react to infection via various responses if uninhibited by the virus. Though HIV-1 has been shown to block the innate anti-viral response (Goujon and Malim 2010), other stress responses such as the unfolded protein response (UPR) can induce shutdown of transcription of ribosomal RNAs as well as repression of translation via phosphorylation of eukaryotic translation initiation factor eIF-2α (DuRose et al. 2009). Translation of HIV-1 proteins is not inhibited by the repression of host proteins because HIV-1 messages can be translated by a cap-independent internal ribosome entry site-mediated mechanism (Yilmaz et al. 2006). Shutdown or drastic alteration of host protein and RNA expression would af-

fect the tRNA pool, as demand for tRNAs for host proteins would be diminished. The UPR response is induced in response to the accumulation of highly expressed viral proteins, particularly the structural glycoproteins (Kaufman 2002) and has been demonstrated to be activated by numerous viruses, including hepatitis C (Chan and Egan 2009), severe acute respiratory syndrome (Minakshi et al. 2009), Japanese Encephalitis virus (Su et al. 2002), and Coxsackie B2 virus (Zhang et al. 2010). UPR responses are activated late in viral infections during translation of structural proteins, in agreement with the observed differences between early and late HIV-1 genes, suggesting that this host response might play a role in the alteration of the tRNA pool during HIV-1 infection.

In short, our finding is consistent with the following scenario. There are two tRNA pools in the host cell at two different times. The first is when the translation machinery is translating host proteins (and the early HIV-1 proteins), and the second is when the host translation machinery has been usurped by HIV-1 to translate mainly HIV-1 late proteins. This second tRNA pool has been selectively enriched with tRNAs translating A-ending codons. The packaged tRNA in the virion is simply a reflection of the relative abundance of this second tRNA pool. The main difference between the two tRNA pools is that the second tRNA pool features relatively more tRNAs translating A-ending codons. Because the early genes are translated in the first tRNA pool with relatively few tRNA translating A-ending codons, they should avoid the use the A-ending codons (similar to human genes). In contrast, the HIV-1 late genes are translated in the second tRNA pool containing relatively more tRNAs decoding A-ending codons, the selection against using A-ending codons is weaker and these late genes have significantly more A-ending codons.

Understanding the translation of HIV-1 genes has potential biomedical implications. Although a plethora of drugs are available for the treatment of HIV-1, none fully suppresses its replication and HIV-1 eventually develops resistance. New drug targets are always needed, and the translation of mRNA into proteins is a point of intervention in the HIV-1 life cycle not currently targeted (de Clercq 2007). Understanding the mechanisms used by HIV-1 to properly express its genes could suggest such novel drug targets. Reduced translational efficiency, particularly of structural genes that are needed for the formation of new particles, could decrease viral success.

Acknowledgments

References

Bahir I, Fromer M, Prat Y, Linial M. 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol.* 5:311.

Berkhout B, Grigoriev A, Bakker M, Lukashov VV. 2002. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. *AIDS Res Hum Retroviruses.* 18:133–141.

Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. *Nature* 325:728–730.

Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37:D93–D97.

Chan S-W, Egan P. 2009. Effects of hepatitis C virus envelope glycoprotein unfolded protein response activation on translation and transcription. *Arch Virol.* 154:1631–1640.

Cullen BR. 1991. Regulation of HIV-1 gene expression. *Faseb J.* 5:2361–2368.

de Clercq E. 2007. The design of drugs for HIV and HCV. *Nat Rev Drug Discov.* 6:1001–1018.

Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 16:287–289.

DuRose JB, Scheuner D, Kaufman RJ, Rothblum LI, Niwa M. 2009. Phosphorylation of eukaryotic translation initiation factor 2alpha coordinates rRNA transcription and translation inhibition during endoplasmic reticulum stress. *Mol Cell Biol.* 29:4295–4307.

Goujon C, Malim MH. 2010. Characterization of the alpha interferon-induced postentry block to HIV-1 infection in primary human macrophages and T cells. *J Virol.* 84:9254–9266.

Haas J, Park E-C, Seed B. 1996. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol.* 6:315–324.

Halwani R, Cen S, Javanbakht H, Saadatmand J, Kim S, Shiba K, Kleiman L. 2004. Cellular distribution of Lysyl-tRNA synthetase and its interaction with Gag during human immunodeficiency virus type 1 assembly. *J Virol.* 78:7553–7564.

Hanada K, Suzuki Y, Gojobori T. 2004. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Mol Biol Evol.* 21:1074–1080.

Huang Y, Mak J, Cao Q, Li Z, Wainberg MA, Kleiman L. 1994. Incorporation of excess wild-type and mutant tRNA3(Lys) into human immunodeficiency virus type 1. *J Virol.* 68:7676–7683.

Javanbakht H, Halwani R, Cen S, Saadatmand J, Musier-Forsyth K, Gottlinger H, Kleiman L. 2003. The interaction between HIV-1 gag and human Lysyl-tRNA synthetase during viral assembly. *J Biol Chem.* 278:27644–27651.

Jenkins GM, Holmes EC. 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92:1–7.

Jiang M, Mak J, Ladha A, Cohen E, Klein M, Rovinski B, Kleiman L. 1993. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. *J Virol.* 67:3246–3253.

Kaufman RJ. 2002. Orchestrating the unfolded protein response in health and disease. *J Clin Invest.* 110:1389–1398.

Keating CP, Hill MK, Hawkes DJ, et al. (11 co-authors). 2009. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. *Nucleic Acids Res.* 37:945–956.

Khorchid A, Javanbakht H, Wise S, Halwani R, Parniak MA, Wainberg MA, Kleiman L. 2000. Sequences within Pr160gag-pol affecting the selective packaging of primer tRNA(Lys3) into HIV-1. *J Mol Biol.* 299:17–26.

Kleiman L, Cen S. 2004. The tRNALys packaging complex in HIV-1. *Int J Biochem Cell Biol.* 36:1776–1786.

Kleiman L, Halwani R, Javanbakht H. 2004. The selective packaging and annealing of primer tRNALys3 in HIV-1. *Curr HIV Res.* 2:163–175.

Kleiman L, Jones CP, Musier-Forsyth K. 2010. Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett.* 584:359–365.

Kypr J, Mrazek JAN. 1987. Unusual codon usage of HIV. *Nature* 327:20.

Lobry JR. 2004. Life history traits and genome structure: aerobiosis and G+C content in bacteria. *Lect Notes Comput Sci.* 3039:679–686.

Mak J, Jiang M, Wainberg MA, Hammarskjold ML, Rekosh D, Kleiman L. 1994. Role of Pr160gag-pol in mediating the selective incorporation of tRNA(Lys) into human immunodeficiency virus type 1 particles. *J Virol.* 68:2065–2072.

Marquet R, Isel C, Ehresmann C, Ehresmann B. 1995. tRNAs as primer of reverse transcriptases. *Biochimie.* 77:113–124.

Martinez MA, Vartanian J-P, Simon W-H. 1994. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc Natl Acad Sci U S A.* 91:11787–11791.

Minakshi R, Padhan K, Rani M, Khan N, Ahmad F, Jameel S. 2009. The SARS coronavirus 3a protein causes endoplasmic reticulum stress and induces ligand-independent downregulation of the type 1 interferon receptor. *PLoS ONE.* 4:e8342.

Nakamura Y, Gojobori T, Ikemura T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28:292.

Ngumbela KC, Ryan KP, Sivamurthy R, Brockman MA, Gandhi RT, Bhardwaj N, Kavanagh DG. 2008. Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 *gag* in intact T cells. *PLoS ONE.* 3:e2356.

Pavon-Eternod M, Wei M, Pan T, Kleiman L. 2010. Profiling non-lysyl tRNAs in HIV-1. *RNA.* 16:267–273.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16:276–277.

Rimsky L, Hauber J, Dukovich M, Malim MH, Langlois A, Cullen BR, Greene WC. 1988. Functional replacement of the HIV-1 rev protein by the HTLV-1 rex protein. *Nature* 335:738–740.

Saksela K, Muchmore E, Girard M, Fultz P, Baltimore D. 1993. High viral load in lymph nodes and latent human immunodeficiency virus (HIV) in peripheral blood cells of HIV-1-infected chimpanzees. *J Virol.* 67:7423–7427.

SAS Institute Inc. 1989. SAS/STAT user's guide. Version 6. Cary (NC): SAS Institute Inc.

Sharp PM. 1986. What can AIDS virus codon usage tell us? *Nature* 324:114.

Sharp PM, Li WH. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:1281–1295.

Strebel K. 2005. APOBEC3G & HTLV-1: inhibition without deamination. *Retrovirology.* 2:37.

Su HL, Liao CL, Lin YL. 2002. Japanese encephalitis virus infection initiates endoplasmic reticulum stress and an unfolded protein response. *J Virol.* 76:4162–4171.

Sueoka N. 1961. Correlation between base composition of deoxyribonucleic acid and amino acid composition of proteins. *Proc Natl Acad Sci U S A.* 47:1141–1149.

Van Dooren S, Pybus OG, Salemi M, et al. (11 co-authors). 2004. The low evolutionary rate of human T-cell lymphotropic virus type-1 confirmed by analysis of vertical transmission chains. *Mol Biol Evol.* 21:603–611.

van Hemert FJ, Berkhout B. 1995. The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral

genome organization and cellular tRNA availability. *J Mol Evol.* 41:132–140.

Vartanian J-P, Henry M, Wain-Hobson S. 2002. Sustained G->A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. *J Gen Virol.* 83:801–805.

Xia X. 2007. An improved implementation of codon adaptation index. *Evol Bioinform.* 3:53–58.

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

Xia X. 1996. Maximizing transcription efficiency causes codon usage bias. *Genetics* 144:1309–1320.

Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered.* 92:371–373.

Yilmaz A, Bolinger C, Boris-Lawrie K. 2006. Retrovirus translation initiation: issues and hypotheses derived from study of HIV-1. *Curr HIV Res.* 4:131–139.

Yoshida M. 2005. Discovery of HTLV-1, the first human retrovirus, its unique regulatory mechanisms, and insights into pathogenesis. *Oncogene* 24:5931–5937.

Yu Q, Chen D, König R, Mariani R, Unutmaz D, Landau NR. 2004. APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. *J Biol Chem.* 279:53379–53386.

Zhang HM, Ye X, Su Y, Yuan J, Liu Z, Stein DA, Yang D. 2010. Coxsackievirus B3 infection activates the unfolded protein response and induces apoptosis through downregulation of p58IPK and activation of CHOP and SREBP1. *J Virol.* 84:8446–8459.