Bioinformatics and Biology Insights

ORIGINAL RESEARCH

Human Retrovirus Codon Usage from tRNA Point of View: Therapeutic Insights

Diego Frias[1,*], Joana P. Monteiro-Cunha[2,3,*], Aline C. Mota-Miranda[2,3], Vagner S. Fonseca[1], Tulio de Oliveira[4], Bernardo Galvao-Castro[3,5] and Luiz C. J. Alcantara[5]

[1]Bahia State University, Salvador, Bahia, Brazil. [2]Federal University of Bahia, Salvador, Bahia, Brazil. [3]Bahia School of Medicine and Public Health/Bahia Foundation for Science Development, Salvador, Bahia, Brazil. [4]Africa Centre for Health and Population Studies, Doris Duke Medical Research Institute, Nelson Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa. [5]Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil. *These authors contributed equally to this work.
Corresponding author email: lalcan@bahia.fiocruz.br

Abstract: The purpose of this study was to investigate the balance between transfer ribonucleic acid (tRNA) supply and demand in retrovirus-infected cells, seeking the best targets for antiretroviral therapy based on the hypothetical tRNA Inhibition Therapy (TRIT). Codon usage and tRNA gene data were retrieved from public databases. Based on logistic principles, a therapeutic score (T-score) was calculated for all sense codons, in each retrovirus-host system. Codons that are critical for viral protein translation, but not as critical for the host, have the highest T-score values. Theoretically, inactivating the cognate tRNA species should imply a severe reduction of the elongation rate during viral mRNA translation. We developed a method to predict tRNA species critical for retroviral protein synthesis. Four of the best TRIT targets in HIV-1 and HIV-2 encode Large Hydrophobic Residues (LHR), which have a central role in protein folding. One of them, codon CUA, is also a TRIT target in both HTLV-1 and HTLV-2. Therefore, a drug designed for inactivating or reducing the cytoplasmatic concentration of tRNA species with anticodon TAG could attenuate significantly both HIV and HTLV protein synthesis rates. Inversely, replacing codons ending in UA by synonymous codons should increase the expression, which is relevant for DNA vaccine design.

Keywords: codon usage, tRNA, HIV, HTLV, therapy

Bioinformatics and Biology Insights 2013:7 335–345

doi: 10.4137/BBI.S12093

This article is available from http://www.la-press.com.

Introduction

The human immunodeficiency virus (HIV) and human T-cell lymphotropic virus (HTLV) are two of the most pathogenic human retroviruses; currently, they affect around 33.5 and 20 million people worldwide, respectively.[1] Both viruses have in common many structural characteristics, the same transmission routes and the fact that they predominately, but not exclusively, infect CD4+ T cells. However, these viruses present distinct disease manifestations. HIV infection is associated with a progressive damage to the immune system which leads to the development of acquired immunodeficiency syndrome (AIDS) within a variable time frame in the great majority of infected individuals. In contrast, HTLV is associated with adult T-cell leukemia/lymphoma (ATL) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP); however, 95% of HTLV infected individuals never develop symptoms.[2] Unfortunately, there is currently no therapy capable of curing or vaccinating against either HIV or HTLV infection.

The translation of messenger ribonucleic acid (mRNA) into protein is an unidirectional process mediated by ribosomes, aminoacylated transfer RNAs (aa-tRNAs), and several factors regulating translation initiation, peptide chain elongation, and translation termination.[3–5]

The primary information gathered from non-redundant coding sequences in a genome, relevant for translational studies, is the counts of the 61 sense codons. For many species this data can be retrieved from the Codon Usage Database. (Nakamura et al, 2000).[6] The relative amount of each codon species determines the codon usage of the species. As each codon needs to be decoded by a cognate tRNA species during translation, codon usage and tRNA content have coevolved towards translation optimization.[7–12] However, tRNA statistics are only available for some model organisms, limiting the studies on the balance between codon usage and tRNA abundance, as in this work.

Most studies of codon usage evolution within species and between closely related species, as well as correlating tRNA abundance and codon usage,[7–21] were based on the unequal use of codons encoding the same amino acid, known as codon usage bias (CUB). However, recent results[22] suggest that CUB is not relevant for translation. Analyzing the correlation between ~152,000 protein chains and the associated mRNAs retrieved from public databases (PDB, Uniprot, TREMBL, EMBL), Saunders and Deane found that CUB is less informative than tRNA concentration for assigning translation speed. Previously, Tuller et al[17] obtained similar results for yeast. This finding, apparently contradictory, is because CUB does not take into account the use of amino acids that is relevant to the logistics of translation. Therefore, our model, rather than rely on CUB, is based on codon frequencies. For this reason we need to develop proper bioinformatics tools.

Pathogens and hosts are immersed in an antagonistic coevolution that results in a never-ending arms race.[23] Some of the arms developed by HIV and the human host cell have already been described. In particular, van Weringh et al[24] found changes in the tRNA pool elicited by the presence of HIV, and Kofman et al[25,26] found evidence of the existence of codon usage-mediated mechanisms of viral gene expression inhibition in mammalian cells, assumed to be independent of translation. More recently, a novel innate antiviral mechanism in the host cell was identified, in which a human protein, Schlafen 11 (SLFN11), selectively inhibits viral protein synthesis in HIV-infected cells.[12] SLFN11 is induced directly by pathogens via the interferon regulatory factor 3 (IRF3) pathway and was shown to bind specific tRNAs modulating the translation of viral proteins without affecting the translation of cell proteins.[12] However, downregulating the abundance of specific tRNA species in the cell in order to selectively inhibit viral protein synthesis can be done by interfering any of the processes occurring in the tRNA life-cycle, and not only by binding to the aminoacylated tRNA molecule.[4] For example, pseudomonic acid (mupirocin) and furanomycin, a non-proteinogenic amino acid, inhibits the isoleucyl-tRNA synthetase, interfering the amino acylation process of the tRNA species decoding codon AUA.[27] Based on these premises, and taking into account that viral mRNA is translated by the host cell translational system, we investigated the logistical support for a hypothetical antiretroviral therapy consisting of inhibiting specific transfer RNA species. Such tRNA Inhibition Therapy (TRIT) should severely attenuate the rate of viral protein translation with moderated collateral effects.

Our approach uses production logistics principles used for ensuring that each workstation is being supplied with the right product, in the right quantity and quality, at the right time.[28] The eukaryotic cell is undoubtedly the most complex chemical industry, especially its translational apparatus. The ribosome can be viewed as a molecular machine which receives protein-production orders from the cell in the form of mRNAs. Because viral mRNAs compete for resources in infected cells, mainly for ribosomes and for tRNAs, with the host mRNA pool it seems reasonable to investigate whether the translation of retroviral mRNA could be attenuated by limiting the supply of certain species of transfer RNA, but without a severe impact on the host cell. Therefore, it was assumed a TRIT drug is capable of selectively inhibiting a specific tRNA species. Under this assumption, the challenge was to identify the best targets for TRIT. With this purpose, we designed an index called T-score, proven to be maximized by tRNA species with the best TRIT potential. From this point, we will refer indistinctly to codons and their cognate tRNAs, except when strictly needed. We will mostly refer to a generic ordinal *i* which varies from 1 to 64, according to an ordering scheme described in the material and methods section.

Material and Methods

Firstly, we characterize data sources and introduce a codon ordering scheme, leading to a new arrangement of the genetic code table. Secondly, we describe how the integral differences of codon usage of the considered species were measured. Then, upon the analysis of tRNA data and identification of pyrimidine-ending synonymous tRNA-poor and tRNA-rich codon pairs in humans, we describe a translational model with the smallest logistic complexity—that is, one that minimizes resource sharing—by assuming that tRNA species are shared by the codons in such pairs. Finally, the rationale behind T-score is described.

Data sources

The overall number of codons in *Homo sapiens*, HIV-1, HIV-2, HTLV-1, and HTLV-2 genomes were retrieved from the Codon Usage Database (http://www.kazusa.or.jp/codon/ accessed on May 2013). The copy number of the genes encoding each tRNA species in human genome was obtained from the Genomic tRNA Database (http://gtrnadb.ucsc.edu/ accessed on May 2013).

Codon ordering and genetic code representation

According to the intracodon purine gradient observed in most coding sequences[29] we introduced a new codon ordering scheme. Codons were ordered according to the nucleotide type (purine or pyrimidine) in the different codon positions. Purines (A before G) come first, followed by pyrimidines (C before U). Thus, denoting by $ord(x)$ the ordinal of a given base x, we have $ord(A) = 1$, $ord(G) = 2$, $ord(C) = 3$, and $ord(U) = 4$. This implies that AAA is the first, $ord(AAA) = 1$, while UUU is the last, $ord(UUU) = 64$. For a generic codon $b_1b_2b_3$, such that $b_i = \{A,G,C,U\}$ for $i = 1,2,3$, we have $ord(b_1b_2b_3) = ord(b_3) + 4(ord(b_2) - 1) + 16(ord(b_1) - 1)$; this way, the first 32 codons begin with purine and the last 32 codons with pyrimidine. Besides, this codon ordering lead to a new arrangement of the translation table (Table 1) that resembles the classification scheme of the genetic code proposed by Wilhelm and Nikolajewa.[37] The adopted codon ordering scheme provides better insights into the symmetry of the underlying box-structure reflecting the redundancies in the genetic code that will be published elsewhere. However, in this case it was useful in identifying regularities on tRNA abundance in the human genome, as well as an atypical codon composition of HTLV.

Codon usage comparison

The codon usage pattern of a given species *s* is, in first instance, given by the list of the genomic frequencies of the codons, $0 \leq C_{i,s} \leq 1$, $i = 1,2, \ldots ,64$ such that $\sum_i C_{i,s} = 1$. The genomic frequency of codon *i* is calculated as the ratio of the counts of the codon, N_i, in the genome and the total number of codons $N = \sum_j N_j$, that is $C_i = N_i/N$. The stop-codon frequencies (at positions 49, 50, and 53) in this context were considered null. To measure the degree of similarity between codon usage patterns of two species "a" and "b", we introduced the dot product coefficient of similarity *S,* given by

$$S_{a,b} = \frac{\sum_i C_{i,a}C_{i,b}}{\sqrt{(\sum_i C_{i,a}^2)(\sum_i C_{i,b}^2)}}.$$

Table 1. Copy number of cognate tRNA genes in *Homo sapiens* by codon species.

n	Codon	aa	g	g*	n	Codon	aa	g	g*	n	Codon	aa	g	g*	n	Codon	aa	g	g*
1	AAA	Lys	16	16	17	GAA	Glu	13	13	33	CAA	Gln	11	11	49	UAA	STOP	–	–
2	AAG		17	17	18	GAG		13	13	34	CAG		20	20	50	UAG		–	–
3	AAC	Asn	32	18	19	GAC	Asp	19	10	35	CAC	His	11	6	51	UAC	Tyr	14	8,8
4	AAU		2	16	20	GAU		0	8,8	36	CAU		0	5	52	UAU		1	6,2
5	AGA	Arg(2)	6	6	21	GGA	Gly	9	9	37	CGA	Arg(4)	6	6	53	UGA	STOP	–	–
6	AGG		5	5	22	GGG		7	7	38	CGG		4	4	54	UGG	Trp	9	9
7	AGC	Ser(2)	8	4,9	23	GGC		15	10	39	CGC		0	5	55	UGC	Cys	30	16
8	AGU		0	3,1	24	GGU		0	4,9	40	CGU		7	2	56	UGU		0	14
9	ACA	Thr	6	6	25	GCA	Ala	9	9	41	CCA	Pro	7	7	57	UCA	Ser(4)	5	5
10	ACG		6	6	26	GCG		5	5	42	CCG		4	4	58	UCG		4	4
11	ACC		0	5,9	27	GCC		0	17	43	CCC		0	5	59	UCC		0	5,9
12	ACU		10	4,1	28	GCU		29	12	44	CCU		10	5	60	UCU		11	5,1
13	AUA	Ile	5	5	29	GUA	Val	5	5	45	CUA	Leu(4)	3	3	61	UUA	Leu(2)	7	7
14	AUG	Met	20	20	30	GUG		16	16	46	CUG		10	10	62	UUG		7	7
15	AUC	Ile	3	9,6	31	GUC		0	6,2	47	CUC		0	7	63	UUC	Phe	12	6,4
16	AUU		14	7,4	32	GUU		11	4,8	48	CUU		12	5	64	UUU		0	5,6

Notes: g, copy number of the cognate tRNA gene; g*, copy number of the cognate tRNA gene after tRNAs sharing within synonymous codons pairs comprised by pyrimidine ending tRNA-poor and tRNA-rich codons (shadowed cells).

which differs from most common indexes used for codon usage studies because it is not based on CUB and varies between 0 and 1. S ranges from 0 to 1 because it is the cosine of the angle spanned by the pair of codon frequency vectors being compared ($[C_{1,a}, C_{2,a}, \ldots, C_{64,a}]$ and $[C_{1,b}, C_{2,b}, \ldots, C_{64,b}]$) on a 64-dimensional Euclidean hyperspace, and the two vectors have no negative components.

Translation model

We used the relative codon frequencies in each synonymous pyrimidine-ending codon pairs to establish a translational balance between tRNA gene number and pair-wise synonymous codon frequency. Denote by g_i and g_j the number of tRNA genes decoding codon i (tRNA-poor) and j (tRNA-rich), respectively, in a tRNA-sharing pair (column g in Table 1). Let $G = g_i + g_j$ denote the overall tRNA gene number in the codon pair; denote also by N_i and N_j the genomic codon counts for codons i and j, respectively, retrieved from the human genome data. The codon-pair wise relative frequency of the tRNA-poor codon is given by $f_i = N_i/(N_i + N_j)$. Thus, we calculate what we call functional tRNA gene numbers as $g_i^* = f_i G$ and $g_j^* = G - g_i^*$, shown at column g^* in Table 1. Note that for codons not forming tRNA-sharing pairs, that is, ending with purine (A or G), we repeat on column g^* the genomic tRNA gene number g. Functional tRNA gene numbers were used to calculate the relative tRNA abundance $t_{i,h} = g_i^*/\sum_j g_j^*$ needed for calculating *T-score*.

Therapeutic score calculation

In productive systems, the ratio of demand (d_i) and supply (s_i) for a generic resource i is a measure of the imbalance ($u_i = 1 - d_i/s_i$) of the supply chain for this resource. The imbalance is positive ($u_i > 0$) when the supply exceeds the demand, and negative ($u_i < 0$) when the supply does not match the demand. Taking into account that virus and host mRNA pools compete for the same finite translational resources in the host ($s_{i,h}$), the performance of the supply chain will differ between virus ($u_{i,v} = 1 - d_{i,v}/s_{i,h}$) and host ($u_{i,h} = 1 - d_{i,h}/s_{i,h}$) perspectives as their demands differ. There are two main translational resources, ribosomes and tRNA species, for which virus and host compete. Here if we focus tRNA species decoding each sense codon i, then the demands are given by the codon frequencies, that is $d_{i,v} = C_{i,v}$ and $d_{i,h} = C_{i,h}$, for virus and host, respectively, while the supply is given by the relative abundance of the cognate tRNA species, that is, $s_{i,h} = t_{i,h}$. In fact, we hypothesized that for some tRNA species there may exist a negative imbalance for the virus ($u_{i,v} < 0$) but a positive imbalance for the host ($u_{i,h} > 0$). If this case ever exists, it must hold that $\Delta u_i = u_{i,h} - u_{i,v} >> 0$; thus by calculating the difference of imbalance Δu_i for each codon i in a given virus-host system, and looking at the maximum

$max(\Delta u_i)$, we can identify the species of tRNA that could be therapeutically inhibited (decreasing $s_{i,h}$) to worsen the imbalance for the virus, but without affecting (turning negative) imbalance for the host. In order to prioritize codons less abundant in the human genome and to minimize the impact on the host cell, Δu_i was divided by the frequency of the codon i in the human genome $C_{i,h}$, thus resulting in the T-score index formula

$$T\text{-}score_i = \frac{\Delta u_i}{C_{i,h}}$$

given in explicit form as

$$T\text{-}score_i = \frac{C_{i,v} - C_{i,h}}{t_{i,h} C_{i,h}}$$

In accordance with T-score definition, tRNA targets for TRIT must have large positive T-score values with respect to the average over all tRNA species.

Results and Discussion

Codon usage comparison

The codon frequencies of HIV and HTLV types 1 and 2 were plotted using the same scale (Fig. 1A). There are slight differences between types but a large difference between retroviruses. Strikingly, the data show that the bias toward codons beginning with purine observed in most organisms,[29] assumed to be a reminiscence of an ancestral R:N:Y (puRine:aNy:pYrimidine) codon,[30] is not present in HTLV. Such dissociation from a universal pattern opens questions about HTLV's origin and evolution, which deserves further attention but is beyond the scope of this paper. To quantify the departure of HTLV from the universal compositional pattern consisting of a purine gradient from the third to the first codon position,[29] we introduced a new compositional feature given by the ratio of codons with purine and pyrimidine in the first codon position (R1/Y1). The R1/Y1 values for HTLV are less than 1 (0.59 for HTLV-1, 0.82 for HTLV-2, whilst HIV-1, HIV-2 and human have R1/Y1 values of 1.81, 1.57, and 1.39, respectively (Fig. 1B).

The similarity index values (S) of the codon frequency patterns are shown in Table 2. HIV-1 and HIV-2 are ~98% similar, while HTLV-1 and HTLV-2 are only ~88% similar. Compared with human, the codon usages of retroviruses have similarities varying from ~76% (human versus HTLV-1) to ~87% (human versus HIV-2).

Translation model

Until now, a total of 506 tRNA genes in the human genome have been annotated. However, those genes encode only 48 tRNA species (anticodons). Therefore, there are 13 tRNA-less sense codons in the human genome (see column g in Table 1). Those tRNA-less codons must be decoded by a tRNA species cognate to their synonymous codon(s) via wobble base pairing.[31,32] We noticed that in the 16 pairs of synonymous codons ending with pyrimidine (eight with U and eight with C) on the standard genetic code, there is a significant bias in the number of cognate tRNA genes, being one codon tRNA-poor, with no or very few tRNA genes and the other tRNA-rich, and with a number of tRNA genes above the average genome (see column g in Table 1). Interestingly, this regularity is amino acid-independent and is determined by the chemical structure or size of the nucleotide at the third codon position, purine or pyrimidine. Purines consist of two carbon-nitrogen chemical rings, whereas pyrimidines have only one ring. We found that only pyrimidine-ending codons have scarce tRNA genes (see shadowed cells in Table 1). We also noticed that all 13 tRNA-less codons form part of such tRNA-poor and tRNA-rich codon pairs. Since regularities in nature are always related to functional constraints, this finding suggests that the 16 tRNA-poor codons are translated mainly by the tRNA species of their synonymous tRNA-rich codons, giving rise to a model of tRNA sharing at pyrimidine-ending codon-pair level.

It also suggests that the most frequent wobble codon-anticodon pairings in the human cell are U-G and C-A, depending on the third base of the tRNA-poor codon.

Under this assumption we can calculate the fraction of tRNA cognate to the tRNA-rich codon in a pyrimidine-ending synonymous codon pair that is more likely used to decode the tRNA-poor codon; this can be done in order to obtain a balance of translation (see material and methods). In columns g and g^* in Table 1, we show the tRNA gene numbers before (genomic) and after redistribution (tRNA-sharing model based), respectively.

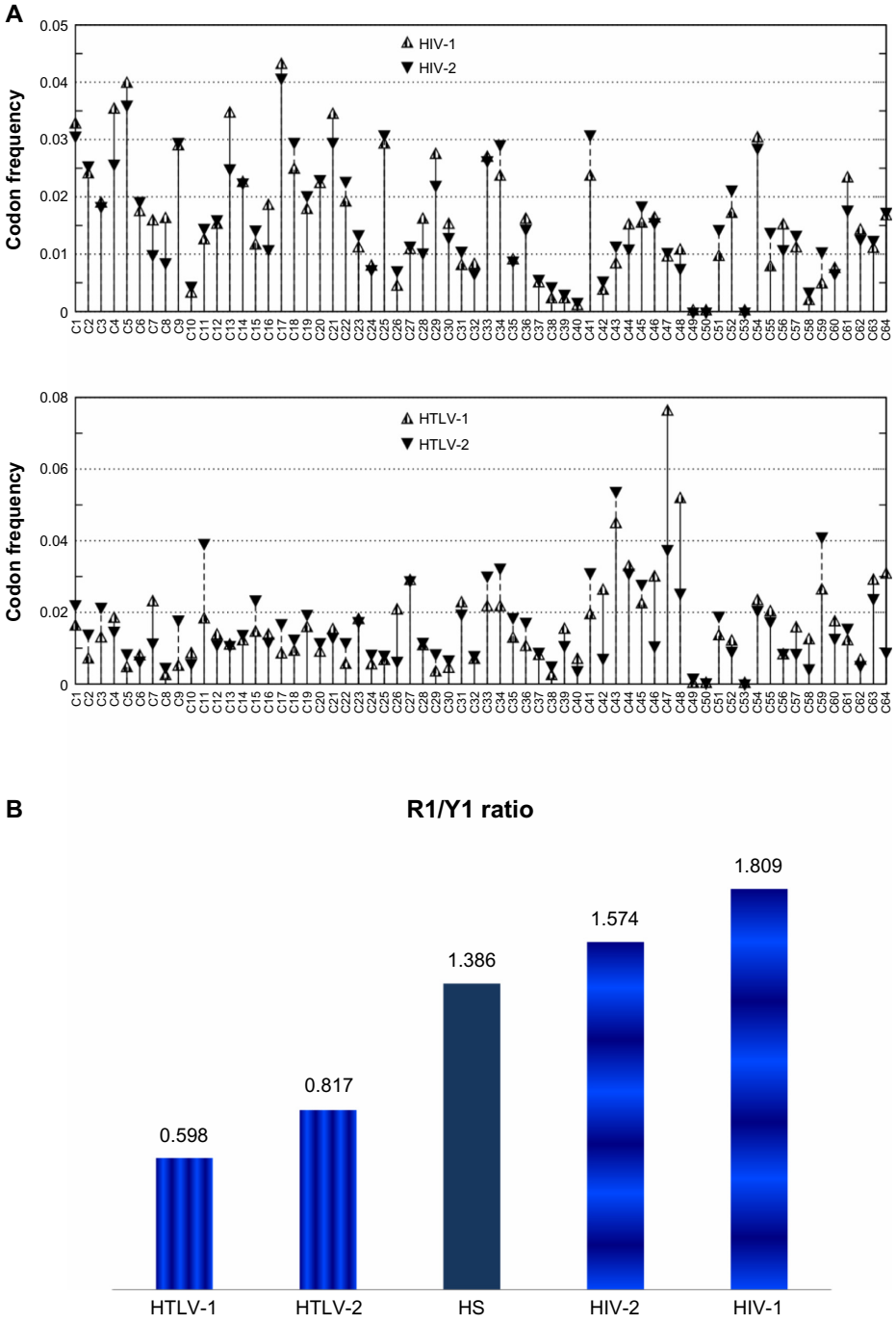

Figure 1. (**A**) Codon frequencies in human retroviruses (HIV and HTLV). *Upper frame*: HIV-1 and HIV-2. *Bottom frame*: HTLV-1 and HTLV-2. The codon number C_i, $i = 1,2, \ldots ,64$ is given by the formula $i = ord(b_3) + 4 * (ord(b_2) - 1) + 16 * (ord(b_1) - 1)$, where b_k is the number of the nucleotide at codon position $k = 1,2,3$. Nucleotides are numbered as: $A = 1$, $G = 2$, $C = 3$ and $U = 4$. Codons 49, 50 and 53 are non-coding codons and their frequencies were set as zero in all cases. The frequency of the first 32 codons, beginning with purine, in the genome of HTLV is lower than that observed in the HIV genome. (**B**) Ratio of purine and pyrimidine at first codon position in retroviruses and human genome. Note that the HIV genome is richer in codons starting with purines than the human genome, whereas the human genome is more rich in purine beginning codons than the HTLV genome.

tRNA and codon frequency balance

The expression profile of human genes was assumed to be uniform, not only due to the lack of expression data in human T-cells, but with the aim of giving the same importance to all genes, including those with low expression in host cells. This is important when an antiviral therapeutic approach that attenuates selectively tRNA abundances is being investigated. In the absence of better data, we assumed a uniform profile of tRNA genes transcription, that is, that tRNA

Table 2. Similarity of codon usage patterns between HIV, HTLV and human genomes.

Compared species	Similarity index—*S*
HIV-1/HIV-2	0.981
HTLV-1/HTLV-2	0.882
HIV-1/HTLV-1	0.615
HIV-2/HTLV-1	0.645
HIV-1/HTLV-2	0.726
HIV-2/HTLV-2	0.775
Human/HIV-1	0.841
Human/HIV-2	0.874
Human/HTLV-1	0.762
Human/HTLV-2	0.832

abundance is determined by the number of genes of each tRNA species. In the upper frame of Figure 2, we show a bar plot of relative codon frequency c_i and relative tRNA abundance t_i for human. The correlation coefficient was R = 0.6325. In the bottom frame of Figure 2, the relative frequency of tRNA $[(t_i - c_i)/c_i]$ was plotted. A negative value indicates deficit of tRNA and positive value tRNA a surplus. Data suggest that there are eleven codons (AAC, AAU, ACG, AUG, AUU, CAA, UGC, UGU, UCG, and UUA) in human genome favored with abundant tRNA species $[(t_i - c_i)/c_i \geq 0.5]$ and five codons (AGC, AGU, CCC, CCU, CUG) neglected $[(t_i - c_i)/c_i \leq -0.5]$. We concluded that the tRNA species decoding such neglected codons should never be selected as targets for TRIT.

T-score analysis in retroviruses

The *T-scores* for the tRNA species decoding the 61 sense codons in HIV-1, HIV-2, HTLV-1, and HTLV-2 are plotted in Figure 3A. The best tRNA targets for TRIT are those having highest positive scores (see material and methods). The results for HIV and HTLV show that approximately 90% of the codons have T-scores in ranging from −85 to 100 (Fig. 3B). Therefore, we assumed that tRNA species with T-score >> 400 are the best TRIT targets, whilst negative outliers are the worst choices (Fig. 3B).

HIV-1 and HIV-2 have the same five best tRNA targets for TRIT decoding codons AGA, AUA, GUA, CUA, and UUA. The worst tRNA target (negative outlier) for HIV types 1 and 2 is the tRNA species decoding the codon CGU. It seems that HIV types 1

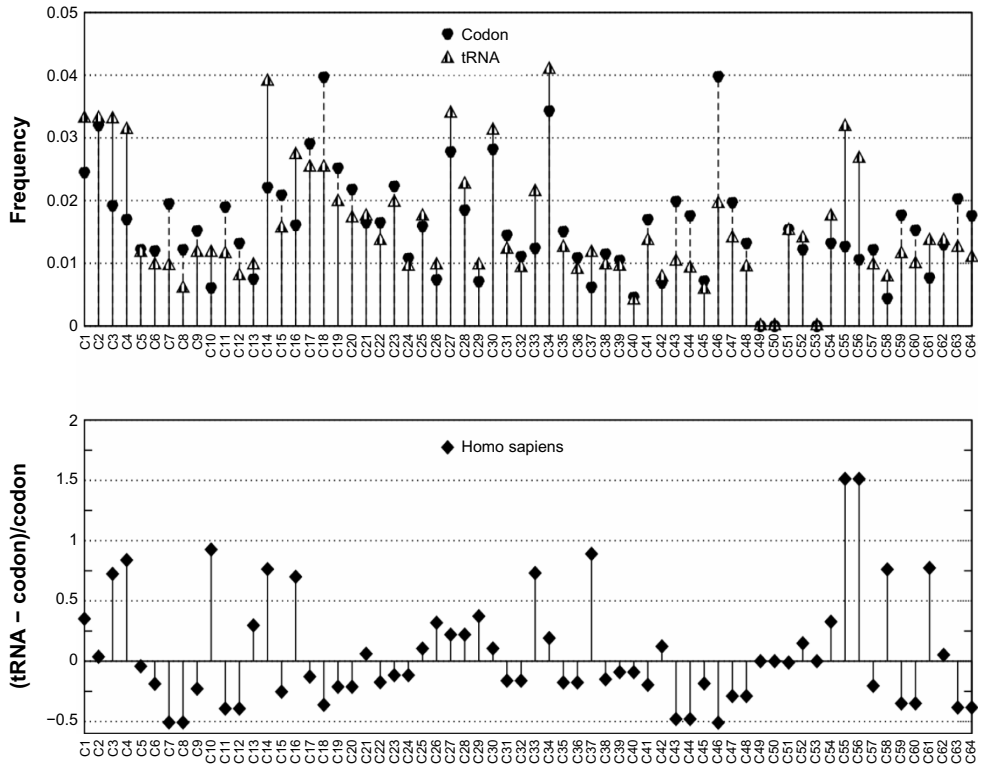

Figure 2. *Upper frame*: Codon frequencies (C_i) and modified cognate tRNA species (t_i^*) in human genome. *Bottom frame*: Relative balance of codon frequency and cognate tRNA species relative abundance $(t_i^* - C_i)/C_i$ for each codon. The codons with the greatest relative tRNA deficit (~−50%) are CUG, AGC and AGU, followed by CCC and CCU. There are also eleven codons with tRNA superavit greater than 50%.

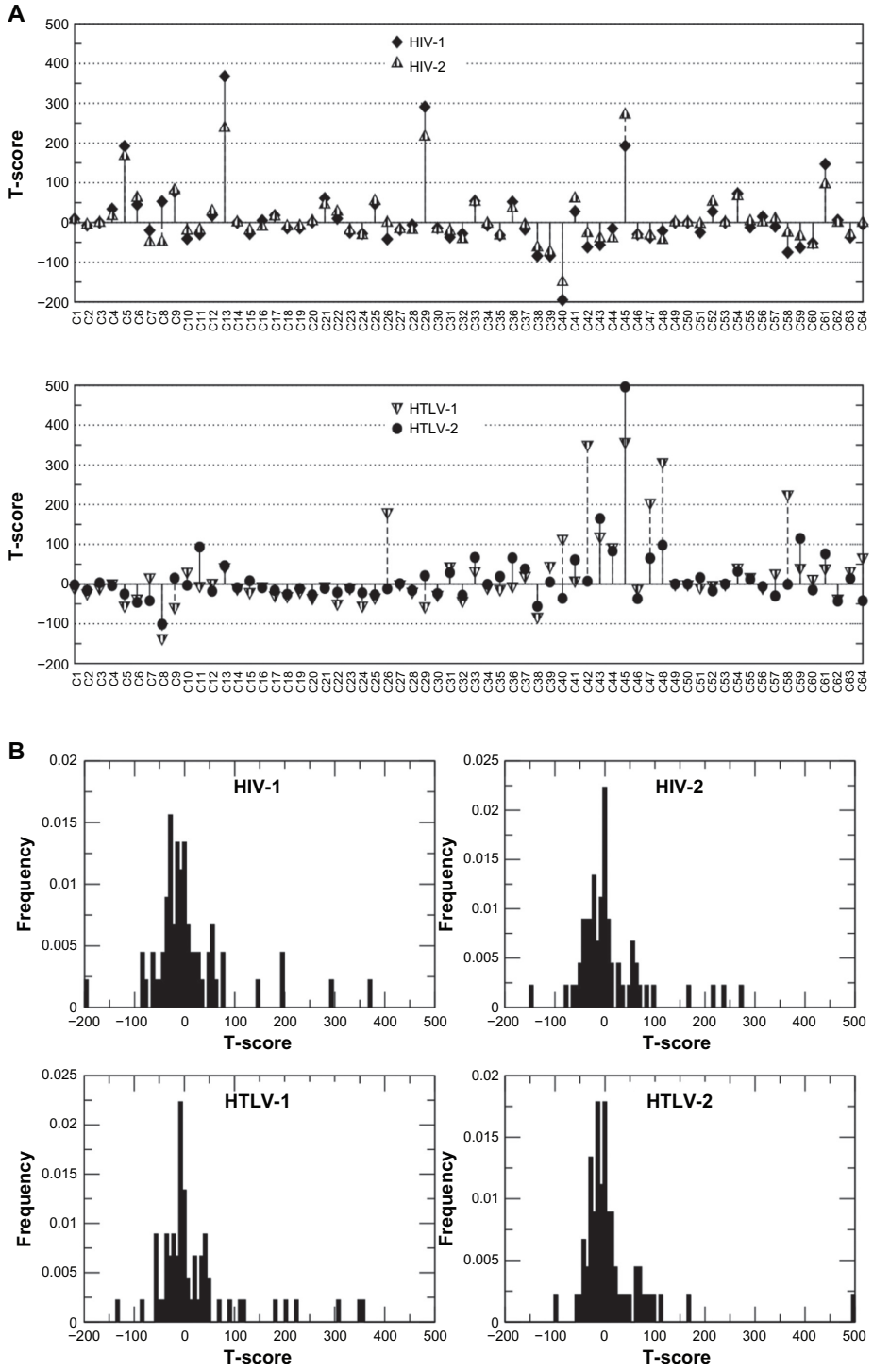

Figure 3. (**A**) Codon T-score plot for HIV-1 and HIV-2 (top) and HTLV-1 and HTLV-2 (bottom). The best tRNA targets takes maximum positive values. (**B**) Distribution of T-score in retroviruses. The T-score of most (~85% in HTLV and ~90% in HIV) tRNA targets belong to the interval [–100, 100]. The best TRIT targets are positive outliers (T-score ≥100).

and 2 have the same tRNA targets due to the high intraspecies similarity of their codon usages, yielding a similarity index of $S_{HIV\text{-}1,\ HIV\text{-}2} = 0.98$ (Table 2). Interestingly, four of the five TRIT targets in HIV types 1 and 2 end in UA, and carry isoleucine, valine and leucine, which are large hydrophobic residues (LHR) known to have a central role in protein folding.[33,34] More precisely, the HIV cone-shaped capsid depends on the correct folding of the side chains of the N-terminus domain (NTD) ring that forms a

hydrophobic region.[36] They found that the correct folding of the side chains depends on four valine (24, 26, 27, and 59), three leucine (20, 52, and 56), and three alanine (22, 31, and 65). This finding suggests that inhibiting the found target tRNA species in HIV-infected patients should affect the assembly of capsids limiting the production of virions.

Regarding HTLV, we found only three tRNA targets for HTLV-2 decoding codons CCC, CUA, and UCC, and one negative outlier decoding codon AGU. However, we found eight tRNA targets for HTLV-1 decoding codons GCG, CGU, CCG, CCC, CUA, CUC, CUU, and UCG, and the same negative outlier. Analyzing the T-score results for HTLV, we noticed that the tRNA-poor CCC codon is one of the five neglected codons in human cells, and that the tRNA species decoding its tRNA-rich synonymous codon CCU was excluded from the target list. The difference in tRNA targets between types 1 and 2 of HTLV can be explained by the large dissimilarity of their codon usages ($1 - S_{HTLV\text{-}1,\ HTLV\text{-}2} = 0.22$) (Table 2).

It should be remarked that despite the large dissimilarities between codon usages of HIV and HTLV species, ranging from 0.23 for HIV-2 and HRLV-2 to 0.39 for HIV-1 and HTLV-1, we noticed that the codon CUA is decoded by one of the best tRNA targets for TRIT for the four studied retrovirus species (HIV-1, HIV-2, HTLV-1, and HTLV-2). Thus, a TRIT drug designed for inhibiting the tRNA with anticodon UAG could probably control both HIV and HTLV types 1 and 2 replication. Finally, it must be noted that, unlike in HIV cases, we found three tRNA targets in HTLV-1 or HTLV-2 decoding codons ending in pyrimidine (CGU, CUC and CUU). For this reason, the calculated T-score values for such tRNA targets depend on the chosen translation model

Conclusion

In this article, we addressed for the first time the balance between tRNA and codon usage in virus-host systems, seeking novel therapeutic alternatives. More precisely, the current study aimed to investigate the viability of a hypothetical TRIT expected to attenuate the translation of retroviral proteins without inducing collateral damage to the host cells. A T-score that quantifies the potential of each tRNA species to be targeted by a TRIT was derived from a steady state balance at ribosome of the relative tRNA species' abundance and cognate codon relative frequencies, in both host and viral mRNA pools. In this case, viability is, on first instance, ensured by differences between the virus and the host codon usage profiles. Available data show significant differences between codon-frequency profiles in retroviruses and humans, which varies from 13% (HIV-2) to 24% (HTLV-1). It is important to remark that a significant singularity in the HTLV codon usage profile was found, since it does not follow the universal pattern observed in most known organisms consisting of a purine gradient towards first codon position. This finding suggests further studies addressing HTLV origin and evolution are needed.

In the next step, we looked for tRNA targets using T-score. To achieve that, a putative distribution of the relative abundance of tRNA species effectively translating each of the 61 sense codons was needed. Available human tRNA gene data show that there are only 48 anticodon species encoded in the genome. Therefore, using wobble pairing theory and logistic principles, we designed a low-complexity translation model that was used to calculate a complete tRNA frequency profile in the human cell, assuming uniform tRNA gene expression. Subsequently, calculating the T-score profile for each retrovirus-host system, we showed that, in all virus types, the T-score of most codons are distributed in the range of −85 to approximately 100 (Fig. 3B). Codons with the greatest T-scores were outside the main distribution (positive outliers), leading to two conclusions: (1) T-score is a robust linear discriminator which performs in a similar way with different genomes like HIV and HTLV, and (2) codons (their cognate tRNAs) with T-score >100 are all TRIT targets in retrovirus-human cases. Analyzing such TRIT targets in HIV-1, we observed that four of the five codons ending in UA, which encode LHR, were identified between the five TRIT targets. Moreover, codon CUA is also the best of the three TRIT targets in HLTV-2, as well as the best of eight TRIT targets in HTLV-1. This coincidence allowed us to suggest the design of a single TRIT drug, theoretically effective against both types of HIV and HTLV. Our data also suggest that substituting adenine with guanine at the third codon position of tRNA-poor codons ending in UA in the HIV genome should increase the availability of tRNA for the mutant codon, and that it may then enhance

HIV protein synthesis. The later can be relevant for increasing antigenicity of vaccine candidates. Previous results support this hypothesis; in particular, Lemey et al[38] found a correlation between synonymous codon substitutions in the virus genome and the replication rate of HIV which determines the disease progression. As synonymous substitutions do not change the amino acid encoded, such increases in the replication rate of the virus are unlikely to be associated with escape from the humoral and cell-mediated immune responses; they are more likely due to an improvement in translation rate as a result of choosing better codons, which is, having more abundant tRNA. This assumption was also supported by the results of Zhou et al[39] which found that papillomavirus capsid protein expression depends on the match between codon usage and tRNA availability. In our work we found that codons ending in UA have the highest T-score, indicating a deficit of tRNA decoding such non-rare codons during translation of HIV proteins in the human cell. However, further studies are needed for testing our in silico-born hypothesis.

Currently, there are five classes of antiretroviral agents combined in Highly Active Antiretroviral Therapy (HAART). They act at different stages in the lifecycle of HIV, inhibiting entry, reverse transcription, integration, and protein cleavage. If an HIV infection becomes resistant to standard HAART, there are limited options. Some patients may benefit from clinical trials of new drugs, but this opportunity is very limited in the developing world. In this scenario, TRIT represents a new and promising alternative for controlling the replication of retroviruses, which can give rise to a new class of antiretroviral drug for HAART. One or two nonsynonymous mutations in the viral genome can provide resistance to certain HAART drugs. Therefore, it is highly relevant to design personalized antiretroviral therapies by screening for resistance mutations in the strains collected for each HIV patient. (Altman et al, 2009).[40] However, to escape from TRIT the virus needs to select several synonymous mutations of the codons decoded by the target tRNA, which does not happen easily by chance.

Acknowledgements

Bioinformatics analysis was performed at the LASP/CPqGM/FIOCRUZ Bioinformatics Unit, supported by Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) and the Brazilian Ministry of Health (PN-DST/AIDS-MS). The authors also acknowledge the helpful contributions of Dr. Tulio de Oliveira.

Author Contributions

Conceived and designed the experiments: DF, JPMC, LCJA, BGC. Analyzed the data: DF, JPMC, VSF. Wrote the first draft of the manuscript: DF, JPMC. Contributed to the writing of the manuscript: DF, JPMC, ACMM, LCJA. Agree with manuscript results and conclusions: TO, BGC. Jointly developed the structure and arguments for the paper: DF, JPMC, LCJA. Made critical revisions and approved final version: TO, LCJA. All authors reviewed and approved of the final manuscript.

Funding

Author(s) disclose no funding sources.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

References

1. Chan JK, Greene WC. Dynamic roles for NF-κB in HTLV-I and HIV-1 retroviral pathogenesis. *Immunol Rev*. 2012;246(1):286–310.
2. Proietti FA, Carneiro-Proietti AB, Catalan-Soares BC, Murphy EL. Global epidemiology of HTLV-I infection and associated diseases. *Oncogene*. 2225;24(39):6058–68.
3. Varenne S, Buc J, Lloubes R, Lazdunski C. Translation is a nonuniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol*. 1984;180(3):549–76.
4. Baker KE, Coller J. The many routes to regulating mRNA translation. *Genome Biology*. 2006;7(12):332.
5. Zouridis H, Hatzimanikatis V. Effects of codon distributions and tRNA competition on protein translation. *Biophys J*. 2008;95(3):1018–33.

6. Nakamura Y, Gojobori T, Ikemura T. Nucleic Acids Res. Codon usage tabulated from the international DNA sequence databases; its status 1999. Jan 1, 1999;27(1):292.
7. Ikemura T, Ozeki H. Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring Harb Symp Quant Biol*. 1983;47 Pt 2:1087–97.
8. Ikemura T. Codon usage and tRNA content in unicellular and multi-cellular organisms. *Mol Biol Evol*. 1985;2(1):13–34.
9. Bulmer M. Coevolution of codon usage and transfer RNA abundance. *Nature*. 1987;325:728–30.
10. Xia X. How optimized is the translational machinery in Escherichia coli, Salmonella typhimurium and Saccharomyces cerevisiae? *Genetics*. 1998;149(1):37–44.
11. Rocha EP. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res*. 2004;14(11):2279–86.
12. Li M, Kao E, Gao X, et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature*. 2012;491(7422):125–8.
13. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol*. 1981;151(3):389–409.
14. Urrutia AO, Hurst LD. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics*. 2001;159(3):1191–9.
15. Berkhout B, Grigoriev A, Bakker M, Lukashov VV. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. *AIDS Res Hum Retroviruses*. 2002;18(2):133–41
16. Lavner Y, Kotlar D. Codon bias as a factor in regulating expression via translation raten in the human genome. *Gene*. 2005;345(1):127–38.
17. Tuller T, Kupiec M, Ruppin E. Determinants of protein abundance and translation efficiency in S. cerevisiae. *PLoS Comput Biol*. 2007;3(12):e248.
18. Horn D. Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*. 2008;9(2):1–11.
19. Shah P, Gilchrist MA. Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genet*. 2010;6(9):e1001128.
20. Belalov IS, Lukashev AN. Causes and implications of codon usage bias in RNA viruses. *PLoS ONE*. 2013;8(2):e56642.
21. Suzuki H, Saito R, Tomita M. 2009. Measure of synonymous codon usage diversity among genes in bacteria. *BMC Bioinformatics*. 2009;10:167.
22. Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. *Nuc Acid Res*. 2010;38(19):6719–28.
23. Benton MJ. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*. 2009;323(5915):728–32.
24. van Weringh A, Ragonnet-Cronin M, Pranckeviciene E, Pavon-Eternod M, Kleiman L, Xia X. HIV-1 modulates the tRNA pool to improve translation efficiency. *Mol Biol Evol*. 2011;28(6):1827–34.
25. Kofman A, Graf M, Deml L, Wolf H, Wagner R. Codon usage-mediated inhibition of HIV-1 gag expression in mammalian cells occurs independently of translation. *Tsitologiia*. 2003;45(1):94–100.
26. Kofman A, Graf M, Bojak A, et al. HIV-1 gag expression is quantitatively dependent on the ratio of native and optimized codons. *Tsitologiia*. 2003;45(1):86–93.
27. Hughes J, Mellows G. Inhibition of isoleucyl-transfer ribonucleic acid synthetase in Escherichia coli by pseudomonic acid. *Biochem J*. 1978;176(1):305–18.
28. Ghiani G, Laporte G, Musmanno R. *Introduction to Logistics Systems Planning and Control*. New York: John Wiley & Sons; 2004.
29. Carels N, Vidal R, Frias D. Universal features for the classification of coding and non-coding DNA sequences. *Bioinform Biol Insights*. 2009;3:37–49.
30. Shepherd JC. Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. *Proc Natl Acad Sci U S A*. 1981;78(3):1596–600.
31. Crick FH. Codon-anticodon pairing: the wobble hypothesis. *J Mol Biol*. 1966;19(2):548–55.
32. Hermann T, Westhof E. Non-Watson—Crick base pairs in RNA-protein recognition. *Chem Biol*. 1999;6(12):R335–43.
33. White SH, Jacobs RE. Statistical distribution of hydrophobic residues along the length of protein chains. Implications for protein folding and evolution. *Biophys J*. 1990;57(4):911–21.
34. Jayaraj V, Suhanya R, Vijayasarathy M, Anandagopu P, Rajasekaran E. Role of large hydrophobic residues in proteins. *Bioinformation*. 2009;3(9):409–12.
35. Pornillos O, Ganser-Pornillos BK, Yeager M. Atomic level modeling of the HIV capsid. *Nature*. 2011;469(7330):424–7.
36. Lemey P, Kosakovsky Pond SL, Drummond AJ, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol*. 2007;3(2):e29.
37. Wilhelm T, Nikolajewa S. A new classification scheme of the genetic code. *J Mol Evol*. 2004;59(5):598–605.
38. Sharp PM, Li WH. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*. 1987;15(3):1281–95.
39. Zhou J, Liu WJ, Peng SW, Sun XY, Frazer I. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *J Virol*. 1999;73(6):4972–82.
40. Altmann A, Däumer M, Beerenwinkel N, et al. 2009. Predicting the response to combination antiretroviral therapy: retrospective validation of geno-2pheno-THEO on a large clinical database. *J Infect Dis*.199(7):999–1006.