Supplementary Information

Human T-cell leukemia virus type 1 uses a specific tRNA^{Pro} isodecoder to prime reverse transcription

Yu-Ci Syu[1], Joshua Hatterschide[1], Christina R. Budding[1], Yingke Tang[1], and Karin Musier-Forsyth[1*].

[1]Department of Chemistry and Biochemistry, Center for RNA Biology, and Center for Retrovirus Research, The Ohio State University, Columbus, OH

*Karin Musier-Forsyth (Corresponding author): musier-forsyth.1@osu.edu; Tel : 614-292-2021

Yu-Ci Syu: ys3824@cumc.columbia.edu

Joshua Hatterschide: joshua.hatterschide@duke.edu

Christina R. Budding: budding.5@osu.edu

Yingke Tang: tang.1044@buckeyemail.osu.edu

Running Title: HTLV-1 RT primer is full-length tRNA^{Pro}

Key Words: HTLV-1, reverse transcription, tRNA primer, isodecoder, m1A58

Materials and Methods

Primer-tagging assay to determine the length of the RT primer

Total RNA from HIV-1 and HTLV-1 virions was extracted by TRIzol (Invitrogen) following the manufacturer's protocol. The viral RT primer-template complex was formed by heating the extracted total viral RNAs in 50 mM HEPES, pH 7.5 at 80°C for 2 min and then at 60°C for 2 min. A final concentration of 1 mM of $MgCl_2$ was added, followed by incubation at 37°C for 30 min, and cooling on ice for at least 30 min. To make size standards, synthetic or *in vitro* transcribed RNAs were annealed in the same manner with 1 μM tRF^{Pro}, $tRNA^{Pro}_{UGG}$, or $tRNA^{Lys}_{UUU}$ and either 0.5 μM HIV-1 or HTLV-1 PBS sequence (see Table S1). For RT extension reactions, the folded RT primer-template complexes were incubated with d/ddNTP mix, 10 mM dithiothreitol (DTT), 1x first-strand buffer (50 mM Tri-HCl, pH 8.3, 75 mM KCl, and 3 mM $MgCl_2$), and SuperScript III reverse transcriptase (Invitrogen) at 50°C for 1 h, followed by incubation at 70°C for 15 min to inactivate the reverse transcriptase. For the HIV-1 primer-template system, two nucleotides (nt), cytidine and thymine, were extended from the 3' end of the primer by addition of a d/ddNTP mix composed of 165 nM of [α-^{32}P] dCTP and 500 nM of ddTTP. For the HTLV-1 primer-template system, three nt, adenosine, cytidine, and thymine, were extended from the 3' end of the primer using a d/ddNTP mix composed of 500 nM dATP, 165 nM [α-^{32}P] dCTP, and 500 nM ddTTP. To generate a single-nt ladder, cycle sequencing reactions were carried out using a BstYI-linearized pUC19 HTLV-1 5' UTR plasmid template and a 5' ^{32}P-labeled primer (sequence 5'-CAG CCC ATC CTA TAG C-3', complementary to nt 360-375 in the template) in 1x reaction buffer (15 mM Tris-HCl and 3.5 mM $MgCl_2$, pH 9.5). Thermo sequenase DNA polymerase and ddNTP termination reaction mix (Affymetrix) were added followed by a PCR reaction according to the following parameters: 95°C for 1 min followed by 60 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min. All samples were mixed with RNA loading dye containing 50% formamide, denatured by heating at 95°C for 5 min, and loaded onto 8% polyacrylamide/8 M urea sequencing gels. Gels were exposed to phosphor screens overnight and visualized by a Typhoon FLA 9500 phosphoimager (Cytiva).

Table S1 List of RNA sequences used in this study.

RNA	Length (nt)	Sequence
HTLV-1 PBS[a]	98	5'-*GG*G CCC AUC CUA UAG CAC UCU CCA GGA GAG AAA UUU AGU ACA CAG UUG GGG GCU CGU CCG GGA UAC GAG CGC CCC UUU AUU CCC UAG GCA AUG GGC C*C*-3'
tRF^{Pro}	18	5'-AUC CCG GAC GAG CCC CCA-3'
$tRNA^{Pro}_{UGG}$	75	5'-GGC UCG UUG GUC UAG GGG UAU GAU UCU CGC UUU GGG UGC GAG AGG UCC CGG GUU CAA AUC CCG GAC GAG CCC CCA-3'
$tRNA^{Lys}_{UUU}$	76	5'-GCC CGG AUA GCU CAG UCG GUA GAG CAU CAG ACU UUU AAU CUG AGG GUC CGG GGU UCA AGU CCC UGU UCG GGC GCC A-3'

[a] The sequence underlined is the HTLV-1 18-nt PBS. The italicized nt are not encoded by HTLV-1; the two 5'-G nt were added for efficient initiation of *in vitro* transcription by T7 RNA polymerase.

Table S2 List of RT and PCR primers used in sequencing the tRNA RT primers.

Retrovirus	Primer	Sequence
HTLV-1	RT and PCR forward primer[a]	5'-GTA AAA CGA CGG CCA G*CG TCT GTA ACT CAC GGC TTG* CAG GAG AGA AAT TTA GTA CAC AGT-3'
	PCR reverse primer[b]	5'-GGC TCG TTG GTC TAG-3'
HIV-1	RT and PCR forward primer[a]	5'-GTA AAA CGA CGG CCA G*CG TCT GTA ACT CAC GGC TTG* AGT GTG GAA AAT CTC TAG CAG-3'
	PCR reverse primer[b]	5'-GCC CGG ATA GCT CAG-3'

[a] Underlined sequences are M13 forward primer designed for Sanger sequencing. Italicized sequences are 20-nt spacers. Shaded sequences are complementary to the 5' 24-nt (for HTLV-1) and the 5' 21-nt (for HIV-1) of the (-) ssDNA.

[b] The PCR reverse primers are complementary to the 5' end of $tRNA^{Pro}$ (for HTLV-1) and $tRNA^{Lys}_{UUU}$ (for HIV-1).

Supplementary Figure S1

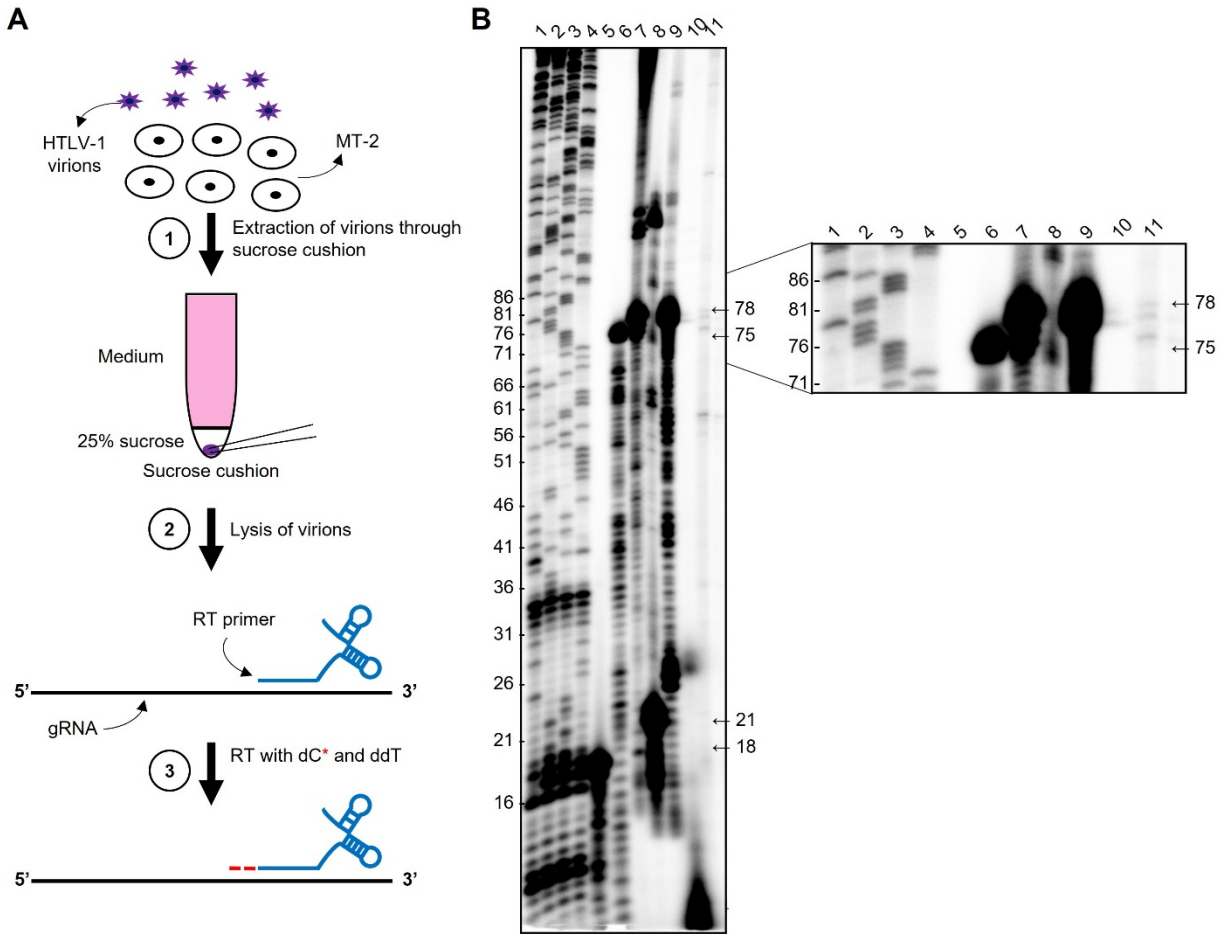
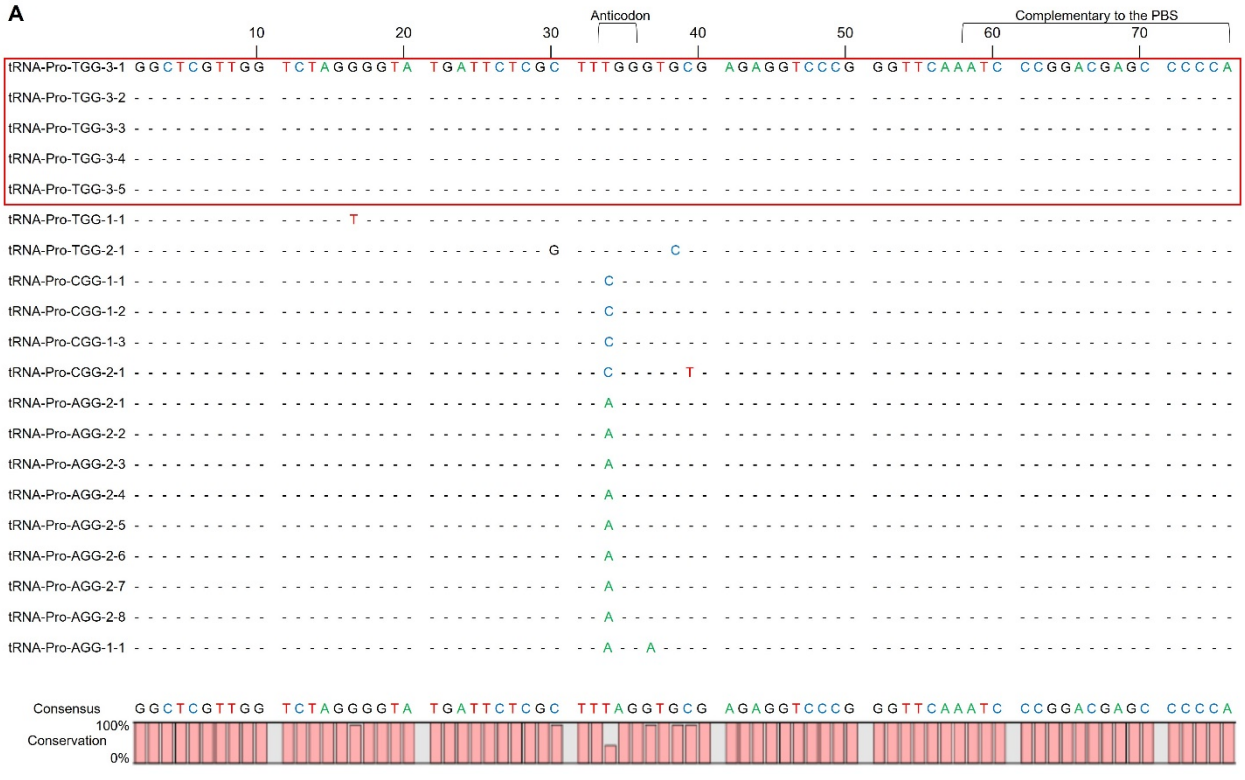


Figure S1. Reverse transcription (RT) primer tagging assays showed that full-length tRNA^{Pro} serves as the HTLV-1 RT primer. (A) Flowchart of method used to concentrate HTLV-1 virions produced from MT-2 cells and label the RT primer with [α-^{32}P] dCTP (dC*). (B) Primer tagging assays to determine the length of the extended RT primer. The first four lanes show a dideoxynucleotide sequencing ladder. Lane 5 is 5' ^{32}P-labeled tRF^{Pro} (18 nt); lane 6 is 5' ^{32}P-labeled $tRNA^{Pro}_{UGG}$ (75 nt); lane 7 was prepared by annealing *in vitro* transcribed $tRNA^{Lys}_{UUU}$ to the HIV-1 PBS and extending by two nt (78 nt); lanes 8 and 9, 21-nt and 78-nt standards were prepared by annealing tRF^{Pro} or $tRNA^{Pro}_{UGG}$ to the HTLV-1 PBS, respectively, and extending them by three nt; lanes 10 and 11 show the results obtained upon primer extension of RNA isolated from HIV-1 virions (lane 10) and HTLV-1 virions (lane 11).

Supplementary Figure S2



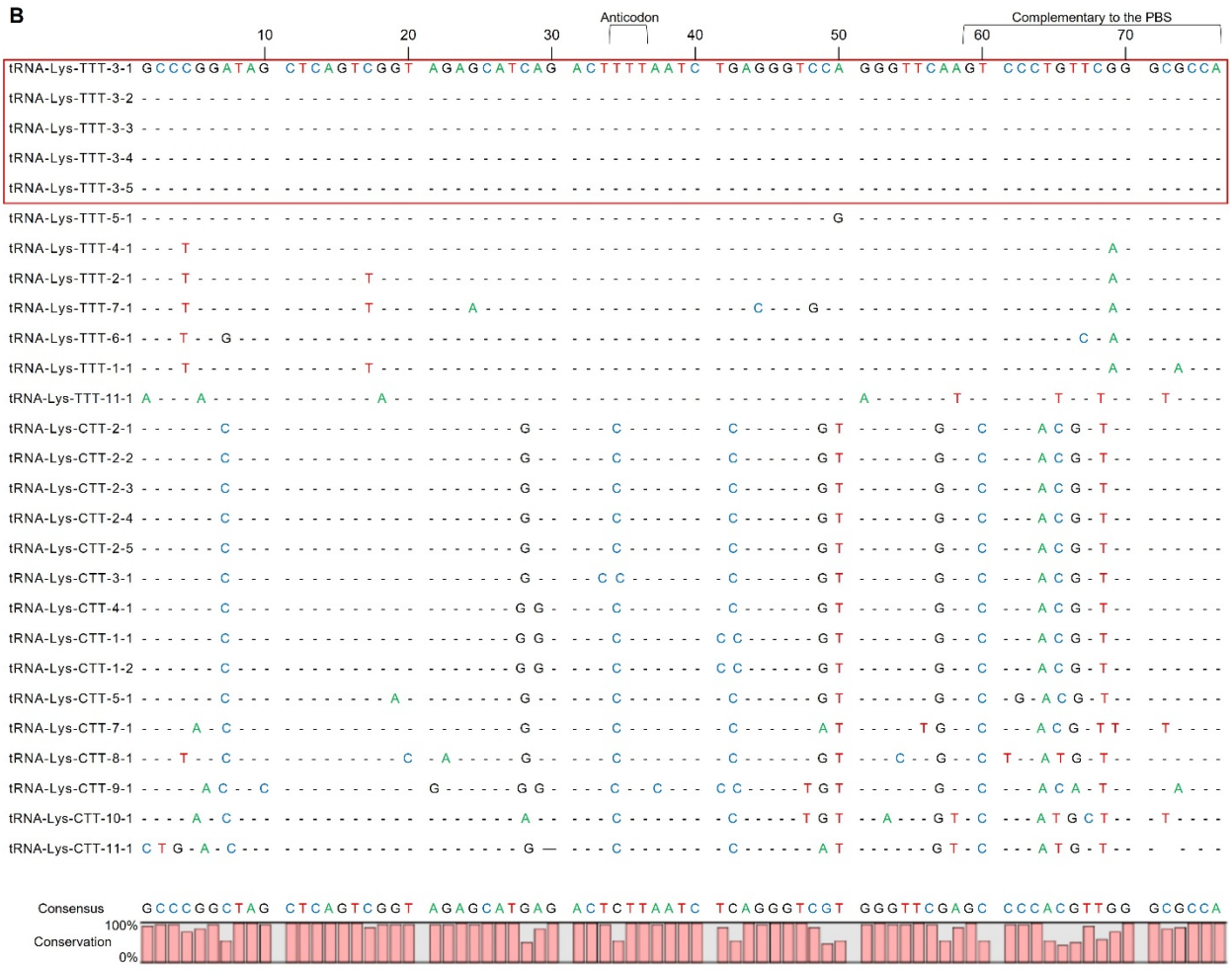


Figure S2. Sequence alignment of human tRNA^{Pro} and tRNA^{Lys} genes. Sequence alignment of 20 human $tRNA^{Pro}_{NGG}$ genes (A) and 27 human $tRNA^{Lys}_{YGG}$ genes (B). tRNA sequences are from the genomic tRNA database (http://gtrnadb.ucsc.edu/), and the alignments were generated with CLC sequence viewer 7 software. Specific tRNA isodecoders identified in this work to be the RT primers used by HTLV-1 or HIV-1 are in red boxes; only nt that differ from these sequences are explicitly shown in the other sequences. The tRNA genes are annotated by two numbers. The first number is transcript ID, indicating a specific tRNA isodecoder. Genes with different transcript IDs have different sequences. The lower the number, the higher the confidence that this gene encodes a tRNA. The second number is a gene locus ID. Genes with the same transcript ID but different gene locus ID have the same sequence but are transcribed from different gene loci (Chan et al. 2021).

Supplementary Figure S3

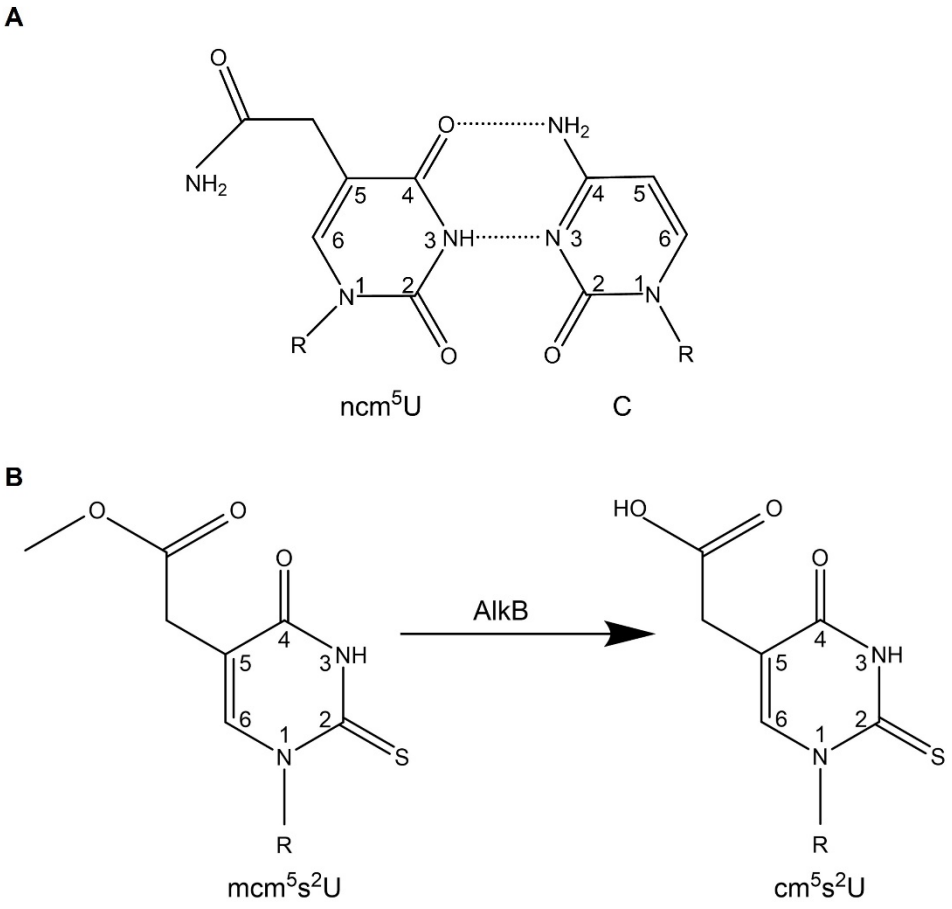


Figure S3. Structure of the ncm^{5}U:C base pair and Alk B demethylation reaction. (A) Two hydrogen bonds form between the N3 hydrogen in ncm^{5}U (left) and the N3 nitrogen in C (right) and between the 4-carbonyl oxygen in ncm^{5}U and the N4 hydrogen in C. (B) The terminal methyl group at the C4 position of mcm^{5}s^{2}U is removed by AlkB, resulting in cm^{5}s^{2}U.