



# HHS Public Access

Author manuscript

*Nat Struct Mol Biol.* Author manuscript; available in PMC 2010 July 01.

Published in final edited form as:

*Nat Struct Mol Biol.* 2010 January ; 17(1): 83–89. doi:10.1038/nsmb.1687.

## A sequence similar to tRNA<sub>3</sub><sup>Lys</sup> gene is embedded in HIV-1 U3/R and promotes minus strand transfer

**Dorota Piekna-Przybylska, Laura DiChiacchio, David H. Mathews, and Robert A. Bambara**  
Department of Biochemistry and Biophysics and Center for RNA Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA.

### Abstract

We identified a sequence embedded in the U3/R region of HIV-1 RNA that is highly complementary to human tRNA<sub>3</sub><sup>Lys</sup>. The free energy of annealing to tRNA<sub>3</sub><sup>Lys</sup> is significantly lower for this sequence and the primer-binding site than for other similar length viral sequences. The only interruption in complementarity is a 29-nucleotide segment inserted where a tRNA intron would be expected. The insert contains the TATA box for viral RNA transcription. The embedded sequence includes a nine-nucleotide segment previously reported to aid minus strand transfer by binding the primer tRNA<sub>3</sub><sup>Lys</sup>. Reconstituting transfer *in vitro*, we show that including segments from the embedded sequence in the acceptor template, beyond the nine nucleotides, further increases transfer efficiency. We propose that a tRNA<sub>3</sub><sup>Lys</sup> gene was incorporated during HIV-1 evolution and retained largely intact because of its roles in transcription and strand transfer.

---

The human immunodeficiency virus 1 (HIV-1), like other retroviruses, initiates its replication in the cell by converting its single stranded RNA genome into double-stranded DNA, a process called reverse transcription. Synthesis of this DNA is a multi-step process catalyzed by the viral enzyme reverse transcriptase (RT). In all retroviruses, the RT utilizes a cellular tRNA as a primer to initiate cDNA synthesis. The HIV-1 virus primes with human tRNA<sub>3</sub><sup>Lys</sup>, which anneals the 18 nucleotides at its 3'-terminus to a complementary primer binding site (PBS) near the 5' end of the viral RNA genome. In early steps of first (minus) strand synthesis a cDNA segment is made that spans from the PBS to the 5' end of the RNA genome. The cDNA synthesis copying the rest of the viral genome is continued after the cDNA fragment copied at the 5' end of the RNA genome is transferred to its 3' end. This process is called minus strand transfer. A similar strand transfer event also takes place later, during the second (plus) strand DNA synthesis<sup>1</sup>.

Efficient minus strand transfer occurs by a sophisticated mechanism that relies on structural features of the viral RNA, and protein factors and activities. The presence of two terminal repeat (R) elements, one at each end of the genomic RNA, facilitates the transfer. The

---

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence should be addressed to R.A.B. (Robert\_Bambara@URMC.Rochester.edu).

**AUTHOR CONTRIBUTIONS** R.A.B., D.P.P. and D.H.M. designed the project, D.P.P. identified the intron, performed sequence alignment, designed and performed biochemical experiments, wrote the manuscript, L.D. and D.H.M. performed computational analysis of virus genomes for hybridization with tRNA<sub>3</sub><sup>Lys</sup>, R.A.B. edited the manuscript.

cDNA generated at the 5' end of the genome, called the minus-strand strong-stop DNA ((-) ssDNA), can interact with the 3' end of the RNA genome through complementarity with the R elements. This interaction can be initiated before the extension of (-) ssDNA reaches the 5' end of the RNA genome<sup>2</sup>, but transfers are most likely completed after full-length synthesis of the (-) ssDNA<sup>3</sup>. These two transfer steps are called invasion and terminus transfer, respectively, and for both, the template RNA used to make the (-) ssDNA must be at least partly removed by the intrinsic RNase H activity of the RT<sup>4</sup>.

Another important factor in HIV-1 minus strand transfer is nucleocapsid protein (NC). NC has a chaperone activity facilitating rearrangement of folding conformations in nucleic acids molecules<sup>1</sup>. NC is proposed to stabilize the interactions between complementary sequences such as the interaction of the tRNA primer and the PBS RNA<sup>5</sup> and the (-) ssDNA with the R element to which it transfers at 3' end of HIV genome<sup>6</sup>. However, it is also known that NC destabilizes the secondary structures such as hairpins in the RNA genome, which would otherwise stall the progression of the RT during reverse transcription<sup>7</sup>.

Studies *in vitro* demonstrated that tRNA<sub>3</sub><sup>Lys</sup> participates in a mechanism that stimulates minus strand transfer<sup>8,9</sup>. The enhancement is proposed to occur through direct interaction between nucleotides at positions 38-46 in the tRNA<sub>3</sub><sup>Lys</sup> and a nonanucleotide segment located within the U3 region adjacent to the R element at the 3' end of the HIV genome. In this context the tRNA<sub>3</sub><sup>Lys</sup> would function not only as a primer for initiating reverse transcription, but also as a bridging factor, that brings distant viral ends together, stimulating strand transfer through a proximity effect<sup>8,9</sup>.

The tRNA<sub>3</sub><sup>Lys</sup> has also been proposed to interact with sequences near the PBS. These interactions were analyzed extensively for several retroviruses, such as ASLV (avian sarcoma and leukosis viruses)<sup>10</sup>, HIV-111-14 and HIV-215 and also for yeast retrotransposon Ty116. Additional contacts between the tRNA and the genomic RNA were proposed to occur in the U5 region immediately 5' of the template PBS. These intermolecular interactions apparently occur transiently during the assembly of the initiation complex essential for tRNA-primed reverse transcription. In the case of HIV-1, proposed tRNA-U5 interactions are extensive, involving contacts with nucleotides in the anticodon loop, stem of the anticodon loop, the variable loop and stem of the T $\psi$ C loop of tRNA<sub>3</sub><sup>Lys</sup><sup>11,13,14,17</sup>. It was proposed that progress of cDNA synthesis displaces these tRNA-U5 duplexes during initiation, freeing the stem of the anticodon loop of the tRNA so it can interact with the complementary nonanucleotide in U3 for stimulation of minus strand transfer<sup>8</sup>. The nine nucleotides of the tRNA-U3 contact are 5'CCCUCAGAU3', a highly conserved sequence among HIV species.

Considering that tRNA<sub>3</sub><sup>Lys</sup> has multiple complementary interactions within U5, and that interaction with a nine nucleotide sequence in U3 promotes strand transfer, we considered the possibility that tRNA<sub>3</sub><sup>Lys</sup> has additional biologically relevant interactions within the U3 region or other parts of the HIV-1 genome. Prediction of low free energy regions of hybridization between human tRNA<sub>3</sub><sup>Lys</sup> and the entire HIV-1 RNA genome revealed the presence of a gene-length tRNA-like sequence in U3/R extensively complementary to the primer tRNA, and including the 9nt motif. Since an entire full-length tRNA sequence is

unlikely to have evolved within HIV-1 to perform viral functions, we hypothesize that this sequence is descended from an embedded ancient tRNA gene. This observation encouraged us to investigate whether interactions beyond the 9nt motif involving this tRNA sequence promote reverse transcription.

## RESULTS

### A tRNA gene-like sequence is present in the LTR of HIV-1

We used a computational approach to scan the entire HIV-1 genome for hybridization potential with tRNA<sub>3</sub><sup>Lys</sup> 18,19. Two regions in HIV-1 stood out in this analysis, having substantial hybridization stability with tRNA<sub>3</sub><sup>Lys</sup>. These are the PBS and a region in U3/R within the long terminal repeat (LTR) that encompasses the originally observed nine-nucleotide sequence, but that extends the length of the tRNA (Fig. 1a). The only interruption of the complementarity is a 29-nucleotide segment inserted in approximately the position expected for a tRNA intron (Fig. 1b). To test the statistical significance of this discovery, we generated ten thousand random sequences with nucleotide content identical to tRNA<sub>3</sub><sup>Lys</sup> by shuffling the tRNA<sub>3</sub><sup>Lys</sup> sequence. The PBS and U3/R sequences hybridize to tRNA<sub>3</sub><sup>Lys</sup> with stabilities that are unlikely to occur by random chance. Figure 2 shows a histogram plot for the lowest folding free energy of hybridization of the shuffled sequences to the entire HIV genome and indicates the stabilities of the tRNA<sub>3</sub><sup>Lys</sup> hybridization to the PBS ( $-80.5$  kcal mol<sup>-1</sup>) and to the U3/R region ( $-74.4$  kcal mol<sup>-1</sup>). It is also worth noting that the third most stable site to which tRNA<sub>3</sub><sup>Lys</sup> can hybridize has a predicted stability of  $-62.7$  kcal mol<sup>-1</sup>, which is close to that expected by random chance and suggests that there are only two true hybridization sites. The full length near-complementarity between tRNA<sub>3</sub><sup>Lys</sup> and the U3/R region of HIV-1 suggests that an entire tRNA gene was incorporated during evolution of the HIV-1 genome.

### A tRNA gene-like sequence is present in SIV, but not in FIV

Since the first discovery of HIV, numerous relatives of this virus have been identified. HIV diverged into several lineages upon multiple cross-species infections involving SIV (simian immunodeficiency virus) from the nonhuman primates<sup>20,21</sup>. The HIV-2 and HIV-1 groups M, N, and O likely represent zoonotic transmission of different SIV groups<sup>22</sup>. FIV (feline immunodeficiency virus), infecting cats, is thought to be the virus from which SIV had originated<sup>23</sup>, and therefore an ancestor of HIV.

The high genomic diversity of HIV results from the error-prone nature of reverse transcriptase and from recombination events<sup>1</sup>. The LTR sequences do not encode proteins, which would tend to restrict evolution, thus evolution of this region could be fast. In order to determine whether the tRNA gene-like sequence is preserved within the HIV groups, we aligned their LTR sequences. We also analyzed the homology of long terminal repeats in SIV and FIV, to establish whether the tRNA gene-like sequence is also present in these two related immunodeficiency viruses.

The LTR alignments revealed that the tRNA gene-like sequence is well preserved among different subtypes of the HIV-1 group and also in SIV infecting chimpanzees (Fig. 3a),

which is considered the most closely related immunodeficiency virus from nonhuman primates<sup>20</sup>. The LTR regions of the HIV-2 group are substantially longer than in HIV-1, but evidence for a tRNA gene-like sequence in U3/R region was apparent. Similarly, a tRNA gene-like sequence is present in the SIV that infects the monkey sooty mangabey (SIV<sub>smm</sub>), a virus from which HIV-2 is thought to originate<sup>21</sup> (Fig. 3b). However, LTR sequences of FIV are much shorter than those of SIV and HIV, and show no sequence evidence of the tRNA-like sequence. It is possible that the tRNA gene-like sequence was never present in the LTR of FIV, or, because FIV is thought to have infected cats for millions of years, it could have been eliminated during evolution.

Since the SIV viruses encompass appreciable diversity, we also performed the analysis of SIV genomes representing different species, including the recently discovered ancient lemur lentivirus pSIVgml identified at the base of the phylogenetic relationship tree of the primate lentiviruses<sup>24</sup>. For comparison we also analyzed sequences of HIV-1 and HIV-2. The results from scanning of their genomes against tRNA<sub>3</sub><sup>Lys</sup> (*P*-values of homology) are summarized in Table 1, and indicate that significant homology to tRNA<sub>3</sub><sup>Lys</sup> can be found in seven of twenty analyzed SIV species including SIV<sub>cpz tan</sub>, SIV<sub>mnd-2</sub>, SIV<sub>drl</sub> and SIV<sub>hoest</sub> with the lowest *P*-values of homology. These SIV species do not group together in the SIV family tree. Overall, acquisition of the tRNA gene-like sequence is evident in viruses from the SIV family, but it is not uniformly preserved, suggesting that biological roles of the sequence vary in different viruses. Moreover, while the tRNA-like sequence is evident in both HIV-1 and HIV-2 from alignment, its preservation is not uniform.

### Minus strand transfer with U3 segments of tRNA-like gene

Our prediction of hybridization showed three major segments of interaction with the tRNA gene-like sequence, any or all of which could form biologically relevant interactions with the tRNA<sub>3</sub><sup>Lys</sup> primer. Two segments are located in the U3 region; one is complementary to the 3' region of tRNA<sub>3</sub><sup>Lys</sup>, and is referred to here as segment 1 (S-1), and the other is the nine-nucleotide segment (9nt) previously analyzed<sup>8,9</sup> (Fig. 4a). The third segment (S-3) is within the transacting responsive (TAR) hairpin of the R element, present at both ends of the HIV RNA genome, and already a fundamental part of the invasion mechanism of minus strand transfer<sup>25</sup>. Our initial experiments evaluated whether segment 1 can promote minus strand transfer. To analyze the role of tRNA/U3 interactions in driving minus strand transfer we previously developed a reconstituted system *in vitro*<sup>9</sup>. It consists of a primer (DNA oligonucleotide or human tRNA<sub>3</sub><sup>Lys</sup>) and two RNA templates with sequences of the 5' end (donor) and 3' end (acceptor) of the HIV-1 RNA genome. The donor RNA template spans 199 nucleotides from the genome 5' end including the R element, U5 region and PBS. The newly synthesized cDNA extended from the primer by purified viral RT can transfer to the acceptor RNA template that also has an R sequence, but with a 5' extension with the sequence of the natural U3 region. Because the donor extension product is shorter than the product of transfer and subsequent extension on the acceptor, these two products are readily distinguishable by gel mobility.

Our previous analysis showed that an acceptor A97(-54) representing sequence between 9022-9172 of HIV-1, which includes the R element and a fragment of U3 including the

nine-nucleotide motif (9033-9041), exhibited highly efficient minus strand transfer, when tRNA<sub>3</sub><sup>Lys</sup> was used as a primer<sup>9</sup>. Transfer efficiency reached 80%. This high efficiency was proposed to occur through cooperation of an invasion-driven mechanism, occurring within the R elements, and a proximity effect achieved by interactions of tRNA<sub>3</sub><sup>Lys</sup> with the nine complementary nucleotides in the U3 region. To determine whether segment 1 in U3 can also influence transfer, we extended the acceptor at its 5' end to 76 nucleotides beyond the R element to include this sequence. In addition, we performed the transfer reactions for this study using donor-acceptor template combinations sufficiently short at the 3' ends that they could not utilize the invasion mechanism. The designed acceptor, A19(-76), contained only the part of the R element not involved in the invasion-driven mechanism, but with 19nt of homology to the 5' end of the donor template (Fig. 4b). Direct terminus transfer was the only allowed mode<sup>1</sup>. Enhancement of transfer could then only occur by a proximity effect mechanism, presumably mediated by sequences complementary to the tRNA<sub>3</sub><sup>Lys</sup> in the U3 region. The advantage of designing the system in this manner is that the basal level of transfer was low. This increases the sensitivity with which enhancement by proximity effects could be measured.

In initial experiments we used the acceptor A19(-54) as a control. This acceptor encompasses the nine-nucleotide transfer-enhancing motif in U3 and allows for terminal transfer with efficiency of about 20%, similar to that previously reported<sup>9</sup> (Figs. 4b and 5). Our mutation or deletion of the complementary nine nucleotides in this acceptor reduced the transfer efficiency to below 10% (Fig. 5). A similar low level of transfer was also observed for acceptor A19(-20)<sup>9</sup>, which does not have any complementarity to tRNA<sub>3</sub><sup>Lys</sup> (data not shown). However, the presence of additional sequences 5' of the nine-nucleotide motif in acceptor A19(-76) caused a significant increase in terminal transfer efficiency, namely up to 70%. Such high efficiency of transfer for systems *in vitro* was previously reported only in reactions with substrates capable of utilizing the invasion mechanism<sup>9</sup>. Mutation or deletion of the nine-nucleotide segment in acceptor A19(-76) caused a drop of transfer efficiency, but only to about 45% (A19(-76)m1; Fig. 5) and 56% (A19(-76)Δ; Fig. 5), respectively, indicating that segment 1 alone can significantly contribute to efficient minus strand DNA transfer.

In order to confirm that segment 1 indeed stimulates the minus strand DNA transfer reaction, we designed an acceptor with a mutation that eliminated potential tRNA interaction with segment 1, but allowed interaction with the nine-nucleotide segment (A19(-76)m3; Fig. 6). We also designed the acceptor with mutations in both segments that did not allow for any interactions with tRNA<sub>3</sub><sup>Lys</sup> (A19(-76)m2). The results are displayed in Figure 6, and confirm that segment 1 contributes significantly to minus strand transfer efficiency. Mutations only in this segment reduced the stimulation to about 25%, a level similar to that measured with the shorter acceptor, A19(-54), having only the nine-nucleotide motif. Few transfer products were observed when the acceptor was long but had mutations in both segments. The transfer efficiency in that case was about 10%, comparable to that of acceptor A19(-54) with a mutated or deleted nine-nucleotide sequence.

As an additional control we performed reactions with the same acceptors, but primed with a DNA oligonucleotide complementary to primer binding site. This system does not allow

interaction of the U3 sequences in the acceptor RNA with the donor-primer. For all acceptors the transfer products were barely detected, instead intense bands reported fold-back products. These products were present at a level similar to those in the control reaction performed without acceptor RNA, which, of course, precludes transfer (Fig. 6). The fold-back products represent the donor extension products folded back at their 3' ends, allowable because their sequence is the complement of the TAR hairpin, and then further extended by copying the donor cDNA template to its 5' end. Fold-back products, prevalent when transfer is inefficient, are a dead-end alternative to transfer, that the natural transfer system has evolved to suppress<sup>1,26</sup>.

Taken together, these results demonstrate that segment 1 can promote efficient minus strand transfer. Interestingly, segment 1 corresponds to the sequence in tRNA<sub>3</sub><sup>Lys</sup> known to be involved in interaction with the PBS region. This suggests that the 3' end of tRNA<sub>3</sub><sup>Lys</sup> used as a primer to start reverse transcription is already released from interactions with the PBS by the time it interacts with segment 1 to facilitate transfer.

## DISCUSSION

Nine nucleotides in the U3 of HIV-1, which are complementary to nucleotides (38-46) in the primer tRNA<sub>3</sub><sup>Lys</sup>, were found to stimulate minus strand DNA transfer *in vitro* by holding together the 5' and 3' regions of the viral RNA genome<sup>8,9</sup>. The nine-nucleotide motif in U3 might have been created by mutations generated during reverse transcription. It would then have been maintained though evolutionary pressure because it can promote viral replication. However, our computational alignment of tRNA<sub>3</sub><sup>Lys</sup> against the HIV-1 genome revealed that the nine-nucleotide segment is, in fact, a part of the larger sequence with additional complementarities to the tRNA<sub>3</sub><sup>Lys</sup> molecule. The complementary sequence spans segments of both the U3 and R elements. Since it is a full-length tRNA sequence, it is more likely to be an acquired cellular gene, than a sequence that evolved within the virus.

Our mutational analysis using a reconstituted system for minus strand transfer *in vitro* showed that extended complementarities in the U3 region, beyond the nine-nucleotide segment, promote greater transfer efficiency. The actual mechanism by which complementary sequences in U3 that are 5' of the nine-nucleotide motif promote transfer could be complex. They would certainly act by augmenting the proximity mechanism promoted by the nine-nucleotide segment. These other sequences also resemble the PBS sequence, and so could help to displace the primer tRNA from the PBS. The release of tRNA<sub>3</sub><sup>Lys</sup> from the PBS is necessary to allow copying of the template region of the PBS into DNA, and copying of the 3' end of the tRNA into DNA, both in preparation for second strand DNA transfer<sup>1</sup>.

A 29-nucleotide long segment near the middle of the tRNA-like sequence has no complementarity to tRNA<sub>3</sub><sup>Lys</sup>, but has the appropriate length and position expected for a tRNA intron. This suggests that the whole region originated from a tRNA gene incorporated into the LTR in an early stage of HIV-1 evolution. The preservation of the sequence in the LTR might partly derive from its role as a section of the HIV-1 promoter. The tRNA gene-like sequence includes the TATAA box and sites recognized by several transcription



factors<sup>27</sup>, probably created by mutations generated during reverse transcription and maintained by evolutionary pressure. It also spans the ascending stem of the TAR hairpin recognized by TAT protein, a viral activator promoting efficient transcriptional elongation<sup>28</sup>. Mutations within these sites substantially reduce viral expression<sup>29</sup> and virus propagation<sup>30</sup>.

Because phylogenetic analyses suggest that HIV and immunodeficiency viruses in other species adapted to and developed in new hosts after cross-species infections, we suspect that this tRNA gene-like sequence did not originate from human DNA. The virus infecting chimpanzees (SIVcpz) is indicated as the most recent ancestor of HIV-1<sup>20</sup>. A cross species origin of HIV-1 is suggested by the presence of an apparent intron in the tRNA-like sequence, since human tRNA<sub>3</sub><sup>Lys</sup> lacks an intron<sup>31</sup>. Introns are also absent in chimpanzee and monkey, but abundant with diverse sequences in cat genomes<sup>31</sup>. However, tRNAs share some homology within the D- and T- arms, thus it is possible that a different tRNA gene was incorporated and during retrovirus evolution some of it evolved to be complementary to tRNA<sub>3</sub><sup>Lys</sup>. HIV-related viruses were identified in sheep, goats, horse, cattle and cats, but only the cat virus FIV appears to be a close relative of HIV and SIV<sup>24,32</sup>. Our sequence alignment within the LTR regions revealed a high homology of HIV-1 with SIVcpz within the tRNA gene-like sequence, but not with FIV. Because FIV is thought to be a very old virus, the tRNA-like sequence in U3/R might have once been present in FIV but lost during divergent evolution.

Determination of the origin of the tRNA-like sequence should provide valuable clues about the ancestry of HIV. We know that vertebrate retroviruses can acquire new genes from their hosts during development<sup>33</sup>. However, it is also possible that a tRNA gene was incorporated much earlier; conceivably when the viral ancestor was a mobile segment of the host genome that evolved into an infectious virus. Evolutionary studies suggest that all retroviruses evolved from retrotransposons, which are ubiquitous mobile pieces of DNA replicating through RNA intermediates<sup>34</sup>. Phylogenetic analyses of the RNase H domains of RTs revealed that the group of LTR retrotransposons could have originated from the non-LTR group<sup>35</sup>, and later become the ancestors of the retroviruses<sup>36,37</sup>.

There are striking similarities between vertebrate retroviruses and LTR retrotransposons in both replication and integration mechanisms, and their “genome” organization. For some LTR retrotransposons, e.g. Ty3 in yeast, an intracellular virus-like particle had been observed<sup>38</sup>. Interestingly, *gypsy*-like elements in *Drosophila* previously classified as LTR retrotransposons, are now considered to be endogenous retroviruses<sup>39</sup>. The Ty3/*gypsy*-like LTR retrotransposons are indicated as a group from which vertebrate retroviruses evolved, or a group that evolved from the same ancestor as *Retroviridae*<sup>37,40</sup>. Interestingly, Ty3 LTR retrotransposons in yeast reveals unique integration preferences, namely in the vicinity of genes transcribed by RNA polymerase III, notably tRNA genes<sup>41</sup>. Similar integration specificity was found among non-LTR retrotransposons TRE in social amoeba<sup>42</sup>. Conceivably, the tRNA gene-like sequences in HIV and SIV are relics of a tRNA gene incorporated by an ancient retrotransposon. Notably, tRNA-like sequences are commonly found in non-LTR retrotransposons, like SINEs and LINEs<sup>43-45</sup>. In addition, Bak and Jorgensen showed that a number of retroviruses and LTR retrotransposons have conserved

sequences of the D- and T- arms of tRNA within the LTR<sup>46</sup>. Moreover, a recent computer analysis revealed the presence of tRNA-like sequences in various transposable elements and viruses in different taxonomic divisions<sup>47</sup>. Incorporation of tRNA genes into LTR of retrotransposons and retroviruses might have been common throughout history.

The significance of tRNA-like sequences found in some LTR retrotransposons and viruses is not clear. In HIV-1, the U3/R sequence is complementary to the tRNA<sub>3</sub><sup>Lys</sup> used by the virus as a primer to start reverse transcription. The tRNA<sub>3</sub><sup>Lys</sup> might take advantage of these complementarities and form interactions that bring both ends of the RNA viral genome into proximity to promote minus strand DNA transfer<sup>8,9</sup>. Other roles for the tRNA-like sequence in HIV-1 could derive from the functions of the D- and T- arms of the tRNA gene representing the internal RNA polymerase III control elements called the A- and B- boxes, respectively<sup>48</sup>. The presence of at least one of the tRNA boxes in proximity to RNA polymerase II promoters stimulates genes expression<sup>49</sup>. The boxes are binding sites for transcription factor TFIIC, which is also required for targeting of tRNA genes during integration of Ty3 and TRE retrotransposons<sup>50,51</sup>. Interestingly, it was demonstrated that HIV-1 integration might occur in proximity to Alu and L1Hs elements<sup>52,53</sup>, where conserved tRNA boxes were also identified<sup>43,45</sup>. Moreover, the tRNA boxes and TFIIC might help in preventing silencing of gene expression, since they are known to restrain heterochromatin spreading in genomic DNA<sup>54</sup>. These considerations suggest that a tRNA-like sequence or A- or B- boxes in the LTR is important for replication and/or mobility of retrotransposons and possibly some viruses today. The lack of a tRNA gene-like sequence in the LTRs of HIV2, FIV and many SIV species indicates that the sequence is not absolutely required for minus strand DNA transfer to occur, but the presence of complementary sequences to primer tRNA at the 3' ends of other retroviruses might significantly stimulate this reaction.

Overall, we suggest that the tRNA-like sequence in U3/R of HIV-1 is a remnant of an event in which a mobile ancestor of HIV acquired a cellular tRNA sequence. We believe that this sequence has been retained in recognizable form because it supports efficient minus strand transfer and transcription of the viral genome. Further investigation of this sequence offers opportunities to probe the ancestry of HIV and the roles of RNA structure in strand transfer.

## ONLINE METHODS

### Materials

We purchased purified human tRNA<sub>3</sub><sup>Lys</sup> from Bio S&T (Lachine, QC) and DNA oligonucleotides from Integrated DNA Technologies, Inc. (Coralville, IA). We obtained pNL4-3 molecular clone (from Dr. Malcolm Martin) through the AIDS Research and Reference Reagent Program (Division of AIDS, NIAID, National Institutes of Health). Dr. Robert J. Gorelick generously provided HIV-1 NC (71 amino acids). We purified HIV-1 reverse transcriptase (p66/p51 heterodimer) (RT) as described previously<sup>55</sup>. We purchased the 32P labeled substrates from Perkin-Elmer Life Sciences.



## Scan of HIV-1 RNA genome for hybridization to tRNA<sub>3</sub><sup>Lys</sup>

We performed bimolecular RNA secondary structure prediction using the Bimol program from RNAstructure19. This program predicts base pairs between interacting RNA sequences without allowing unimolecular base pairs and predicts the lowest free energy hybridization site and suboptimal sites of hybridization18. We calculated *P*-values to determine significance of homology using a previously described method56.

## Sequence alignment

We took the consensus sequences for each HIV and SIV group from the HIV Database (<http://www.hiv.lanl.gov/>). In order to perform alignments, we used a multiple sequence alignment program T-coffee57 available at the EMBL-EBI website.

## Preparation of donor and acceptor RNA substrates

In order to prepare the donor and acceptor RNA templates, we first amplified the genomic sequences of HIV-1 (pNL4-3) by PCR using *Vent* DNA polymerase (BioLabs). We generated the DNA templates for transcription *in vitro* with the following sets of oligonucleotides: oligomers 1 and 2 (see below) for the donor template corresponding to the first 199nt at 5' end of the HIV-1 RNA genome; oligomers 3 and 4 for acceptor template A19(-54) corresponding to region 9022-9094 in the HIV-1 RNA genome; oligomers 5 and 4 for A19(-54)m with mutations in the 9nt motif; oligomers 6 and 4 for A19(-54)Δ with deletions of the 9nt motif; oligomers 7 and 4 for A19(-76) corresponding to region 9000-9094 in the HIV-1 RNA genome. We used the set of two overlapping oligonucleotides to generate DNA templates for the following acceptors: oligomers 8 and 4 for A19(-20) corresponding to region 9055-9097 in the HIV-1 RNA genome; oligomers 9 and 10 for A19(-76)m1 with mutation in the 9nt segment; oligomers 11 and 10 for A19(-76)Δ with deletion of the 9nt segment; oligomers 12 and 13 for A19(-76)m2 with mutation of segment 1; oligomers 14 and 13 for A19(-76)m3 with mutations in segment 1 and the 9nt motif. The forward oligonucleotides (1, 3, 5, 6, 7, 8, 9, 11, 12, 14) for all DNA templates contain the sequence of T7 RNA promoter. Using the Ambion T7-MEGAshortscript kit (Applied Biosystems) we transcribed *in vitro* the RNA templates, then purified by polyacrylamide/urea gel electrophoresis and resuspended in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA buffer. RNAs were quantitated by UV absorption using a GeneQuant II from Amersham Biosciences.

## DNA oligonucleotides

**1.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGT CTC TCT GGT TAG ACC AGA TC-3'; **2.** 5'-GTC CCT GTT CGG GCG CCA CTG CTA GAG A-3'; **3.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGA GTG GCG AGC CCT CAG ATG-3'; **4.** 5'-GGT CTA ACC AGA GAG ACC CAG TAC AGG CAA AAA GCA G-3'; **5.** 5'-TAA TAC GAC TCA CTA TAG GGA GTG GCG AGT ATA TAT ATG CTG CAT ATA AGC AGC TGC TTT TTG CC-3'; **6.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGA GTG GCG AGG CTG CAT ATA AGC AGC TGC TTT TTG-3'; **7.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGC GTG GCC TGG GCG GGA CTG GGG AG-3'; **8.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGC TGC TTT TTG CCT GTA CTG-3; **9.** 5'-GGA TCC TAA TAC

GAC TCA CTA TAG GGC GTG GCC TGG GCG GGA CTG GGG AGT GGC GAG CTC  
 CTC GAT GCT GCA TAT AAG CAG CTG CTT TTT GC-3'; **10.** 5'-GGT CTA ACC AGA  
 GAG ACC CAG TAC AGG CAA AAA GCA GCT GCT TAT ATG CAG C-3'; **11.** 5'-  
 GGA TCC TAA TAC GAC TCA CTA TAG GGC GTG GCC TGG GCG GGA CTG GGG  
 AGT GGC GAG GCT GCA TAT AAG CAG CTG CTT TTT GC-3'; **12.** 5'-GGA TCC  
 TAA TAC GAC TCA CTA TAG GGC GTG GCC AAC CAG GCC AAG GGG AGT GGC  
 GAG CCC TCA GAT GCT GCA TAT AAG CAG CTG CTT TTT-3'; **13.** 5'-GGT CTA  
 ACC AGA GAG ACC CAG TAC AGG CAA AAA GCA GCT GCT TAT ATG CAG CAT  
 C-3'; **14.** 5'-GGA TCC TAA TAC GAC TCA CTA TAG GGC GTG GCC AAC CAG GCC  
 AAG GGG AGT GGC GAG CTC CTC GAT GCT GCA TAT AAG CAG CTG CTT  
 TTT-3' ; **15.** 5'-TGG CGC CCG AAC AGG GAC TTG AAC CCT GGA CCC-3'; **16.** 5'-  
 GTC TGA TGC TCT ACC GAC TGA GCT ATC CGG G-3'

### Labeling of tRNA<sub>3</sub><sup>Lys</sup> and the DNA primer

We prepared 5'-radiolabeled tRNA<sub>3</sub><sup>Lys</sup> as described previously with slight modifications<sup>9</sup>. We heat annealed a 1.5-fold excess of two DNA oligonucleotides 15 and 16 to tRNA<sub>3</sub><sup>Lys</sup> at 65°C for 5 min in buffer 50 mM Tris-HCl (pH 8.0), 75 mM KCl, and gradually cooled to 37°C. The purpose of this annealing process was to prevent the tRNA from folding into a configuration that would inhibit labeling. Next, we treated the reaction mixture with shrimp alkaline phosphatase (SAP, Fermentas) at 37°C for 60 min., and then incubated at 65°C for 25 min to inactivate the enzyme. Following slow cooling to 37°C, we treated the reaction mixture with [ $\gamma$ -<sup>32</sup>P]ATP (6000 (222TBq) Ci mmol<sup>-1</sup>). We separated the radiolabeled tRNA<sub>3</sub><sup>Lys</sup> and DNA oligomers by electrophoresis in a 6% polyacrylamide/urea gel, and then eluted the tRNA<sub>3</sub><sup>Lys</sup> overnight from the gel, precipitated with ethanol, and resuspended in 10 mM Tris-HCl (pH 8.0), 1 mM EDTA buffer. We labeled the DNA primer with sequence complementary to the PBS region (5'-GTC CCT GTT CGG GCG CCA-3') at the 5'-end using [ $\gamma$ -<sup>32</sup>P]ATP (6000 (222 TBq) Ci mmol<sup>-1</sup>) and T4 polynucleotide kinase (Roche) and separated it from unincorporated radionucleotides using a Micro Bio-Spin column (Bio-Rad).

### Strand Transfer Assay

We heat-annealed the tRNA<sub>3</sub><sup>Lys</sup> primer or DNA primer to donor RNA by incubation at 95°C for 5 min and slow cooling to 37°C. Then, we added acceptor templates and incubated with 200% NC (100% NC is 7 nt per NC molecule) for 3 min. Next we added RT to the substrate and after 4 min of incubation at 37°C, we initiated reactions with MgCl<sub>2</sub> and dNTPs. We mixed primer, donor, and acceptor at a ratio of 2:1:2. Final reaction mixtures contained 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 1 mM DTT, 1 mM EDTA, 32 nM HIV-1 RT, 6 mM MgCl<sub>2</sub>, and 50  $\mu$ M dNTPs. We incubated reactions at 37°C, and terminated them at appropriate times with 1 volume of termination dye (10 mM EDTA, pH 8.0, 90% formamide (v/v), and 0.1% (w/v) each xylene cyanole and bromophenol blue). We then resolved products by polyacrylamide-urea gels and analyzed using a PhosphorImager (GE Healthcare) and ImageQuant (version 2.1). We estimated sizes of DNA products by using a 5'-radiolabeled 10 bp DNA ladder (Invitrogen).

## ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health research grants GM049573 (to R.A.B.) and GM076485 (to D.H.M.). We received additional support from the Rochester Developmental Center for AIDS Research (NIH P30-AI78498). We are grateful to Dr. Robert J. Gorelick, of the National Cancer Institute, for NC used in these studies. We thank Dr. Thomas H. Eickbush, Dr. Harold Smith, Dr. Lata Balakrishnan, and students of the Bambara laboratory group for helpful comments.

## ABBREVIATIONS

<b>RT</b>	reverse transcriptase
<b>NC</b>	nucleocapsid protein
<b>LTR</b>	long terminal repeat
<b>PBS</b>	primer binding site
<b>(-)ssDNA</b>	minus-strand strong-stop DNA
<b>9nt</b>	nine nucleotide motif
<b>S-1</b>	segment 1
<b>S-3</b>	segment 3
<b>TAR hairpin</b>	transacting responsive hairpin
<b>SIV</b>	simian immunodeficiency virus
<b>FIV</b>	feline immunodeficiency virus
<b>TRE</b>	tRNA gene-targeted retrotransposable elements
<b>SINE</b>	short interspersed nuclear elements
<b>LINE</b>	long interspersed nuclear elements

## REFERENCES

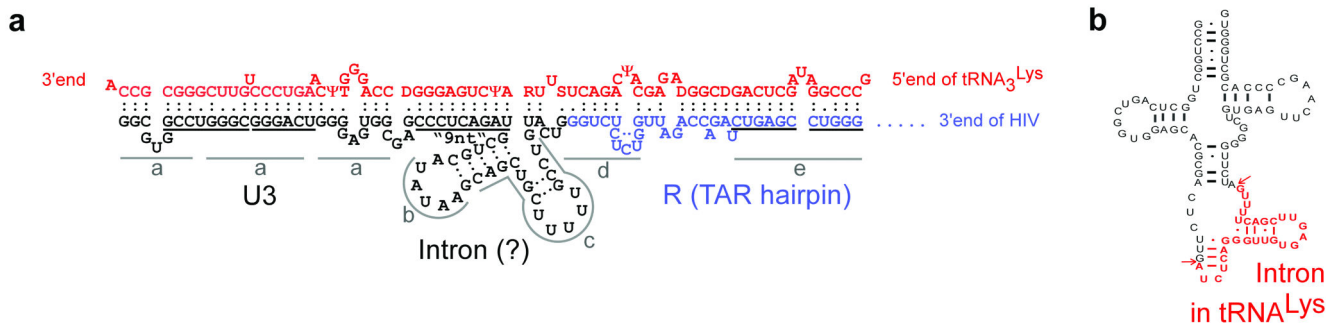
1. Basu VP, et al. Strand transfer events during HIV-1 reverse transcription. *Virus Res.* 2008; 134:19–38. [PubMed: 18279992]
2. Chen Y, Balakrishnan M, Roques BP, Bambara RA. Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. *J. Biol. Chem.* 2003; 278:38368–75. [PubMed: 12878597]
3. Peliska JA, Benkovic SJ. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science.* 1992; 258:1112–8. [PubMed: 1279806]
4. Negroni M, Buc H. Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. *Proc. Natl. Acad. Sci. U S A.* 2000; 97:6385–90. [PubMed: 10829081]
5. Feng YX, et al. The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.* 1999; 73:4251–6. [PubMed: 10196321]
6. You JC, McHenry CS. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* 1994; 269:31491–5. [PubMed: 7989315]
7. Rodriguez-Rodriguez L, Tsuchihashi Z, Fuentes GM, Bambara RA, Fay PJ. Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J. Biol. Chem.* 1995; 270:15005–11. [PubMed: 7541033]

8. Brule F, et al. In vitro evidence for the interaction of tRNA(3)(Lys) with U3 during the first strand transfer of HIV-1 reverse transcription. *Nucleic Acids Res.* 2000; 28:634–40. [PubMed: 10606665]
9. Song M, Balakrishnan M, Gorelick RJ, Bambara RA. A succession of mechanisms stimulate efficient reconstituted HIV-1 minus strand strong stop DNA transfer. *Biochemistry.* 2009; 48:1810–1819. [PubMed: 19192967]
10. Aiyar A, Cobrinik D, Ge Z, Kung HJ, Leis J. Interaction between retroviral U5 RNA and the T psi C loop of the tRNA(Trp) primer is required for efficient initiation of reverse transcription. *J. Virol.* 1992; 66:2464–72. [PubMed: 1548772]
11. Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B. Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* 1993; 268:25269–72. [PubMed: 7503978]
12. Zhang Z, Kang SM, Li Y, Morrow CD. Genetic analysis of the U5-PBS of a novel HIV-1 reveals multiple interactions between the tRNA and RNA genome required for initiation of reverse transcription. *RNA.* 1998; 4:394–406. [PubMed: 9630246]
13. Iwatani Y, Rosen AE, Guo J, Musier-Forsyth K, Levin JG. Efficient initiation of HIV-1 reverse transcription in vitro. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. *J. Biol. Chem.* 2003; 278:14185–95. [PubMed: 12560327]
14. Beerens N, Groot F, Berkhout B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.* 2001; 276:31247–56. [PubMed: 11384976]
15. Berkhout B, Schoneveld I. Secondary structure of the HIV-2 leader RNA comprising the tRNA-primer binding site. *Nucleic Acids Res.* 1993; 21:1171–8. [PubMed: 8464701]
16. Friant S, Heyman T, Wilhelm ML, Wilhelm FX. Extended interactions between the primer tRNAi(Met) and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic Acids Res.* 1996; 24:441–9. [PubMed: 8602356]
17. Kleiman L. tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life.* 2002; 53:107–14. [PubMed: 12049192]
18. Mathews DH, Burkard ME, Freier SM, Wyatt JR, Turner DH. Predicting oligonucleotide affinity to nucleic acid targets. *RNA.* 1999; 5:1458–69. [PubMed: 10580474]
19. Mathews DH, et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U S A.* 2004; 101:7287–92. [PubMed: 15123812]
20. Peeters M, et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS.* 1989; 3:625–30. [PubMed: 2512955]
21. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature.* 1989; 339:389–92. [PubMed: 2786147]
22. Tebit DM, Nankya I, Arts EJ, Gao Y. HIV diversity, recombination and disease progression: how does fitness “fit” into the puzzle? *AIDS Rev.* 2007; 9:75–87. [PubMed: 17694675]
23. Goudsmit, J.; *Beyond, SIV.; HIV. Viral Sex: The Nature of AIDS.* Oxford University Press US; 1997. The Cat Connection; p. 127-142.
24. Gifford RJ, et al. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc. Natl. Acad. Sci. U S A.* 2008; 105:20362–7. [PubMed: 19075221]
25. Chen Y, Balakrishnan M, Roques BP, Fay PJ, Bambara RA. Mechanism of minus strand strong stop transfer in HIV-1 reverse transcription. *J. Biol. Chem.* 2003; 278:8006–17. [PubMed: 12499370]
26. Guo J, Henderson LE, Bess J, Kane B, Levin JG. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* 1997; 71:5178–88. [PubMed: 9188585]
27. Pereira LA, Bentley K, Peeters A, Churchill MJ, Deacon NJ. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. *Nucleic Acids Res.* 2000; 28:663–8. [PubMed: 10637316]

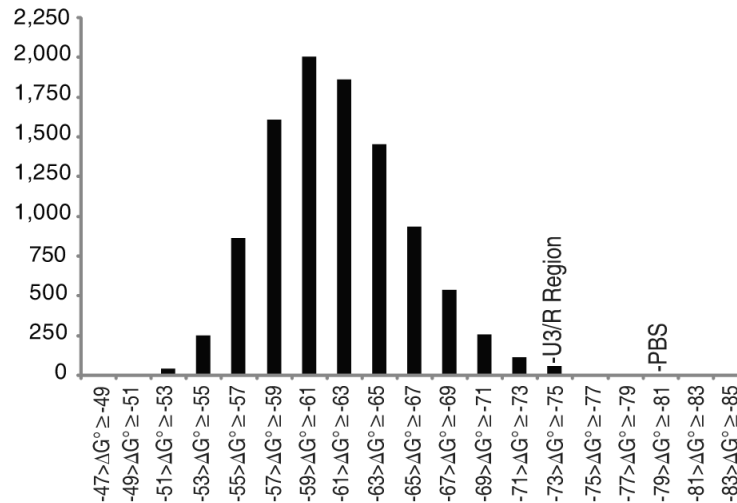
28. Rana TM, Jeang KT. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* 1999; 365:175–85. [PubMed: 10328810]
29. Naghavi MH, Schwartz S, Sonnerborg A, Vahlne A. Long terminal repeat promoter/enhancer activity of different subtypes of HIV type 1. *AIDS Res. Hum. Retroviruses.* 1999; 15:1293–303. [PubMed: 10505678]
30. de Arellano, E. Ramirez; Martin, C.; Soriano, V.; Alcami, J.; Holguin, A. Genetic analysis of the long terminal repeat (LTR) promoter region in HIV-1-infected individuals with different rates of disease progression. *Virus Genes.* 2007; 34:111–6. [PubMed: 17160552]
31. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997; 25:955–64. [PubMed: 9023104]
32. Carpenter MA, O'Brien SJ. Coadaptation and immunodeficiency virus: lessons from the Felidae. *Curr. Opin. Genet. Dev.* 1995; 5:739–45. [PubMed: 8745072]
33. Halpern CC, Hayward WS, Hanafusa H. Characterization of some isolates of newly recovered avian sarcoma virus. *J. Virol.* 1979; 29:91–101. [PubMed: 219237]
34. Temin HM. Origin of retroviruses from cellular moveable genetic elements. *Cell.* 1980; 21:599–600. [PubMed: 6254661]
35. Malik HS, Eickbush TH. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* 2001; 11:1187–97. [PubMed: 11435400]
36. Eickbush, TH. Origin and evolutionary relationships of retroelements. In: Morse, SS., editor. *The Evolutionary Biology of Viruses.* Raven Press; New York: 1994. p. 121-157.
37. Eickbush, TH.; Malik, HS. Origins and evolution of retrotransposons. In: Craig, NL.; Craigie, R.; Gellert, M.; Lambowitz, AM., editors. *Mobile DNA.* Vol. II. American Society of Microbiology Press; Washington, DC: 2002. p. 1111-1144.
38. Hansen LJ, Chalker DL, Orlinsky KJ, Sandmeyer SB. Ty3 GAG3 and POL3 genes encode the components of intracellular particles. *J. Virol.* 1992; 66:1414–24. [PubMed: 1371165]
39. Kim A, et al. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U S A.* 1994; 91:1285–9. [PubMed: 8108403]
40. Llorens C, Fares MA, Moya A. Relationships of gag-pol diversity between Ty3/Gypsy and Retroviridae LTR retroelements and the three kings hypothesis. *BMC Evol. Biol.* 2008; 8:276. [PubMed: 18842133]
41. Boeke JD, Devine SE. Yeast retrotransposons: finding a nice quiet neighborhood. *Cell.* 1998; 93:1087–9. [PubMed: 9657139]
42. Szafranski K, et al. Non-LTR retrotransposons with unique integration preferences downstream of *Dictyostelium discoideum* tRNA genes. *Mol. Gen. Genet.* 1999; 262:772–80. [PubMed: 10628860]
43. Paoletta G, Lucero MA, Murphy MH, Baralle FE. The Alu family repeat promoter has a tRNA-like bipartite structure. *EMBO J.* 1983; 2:691–696. [PubMed: 16453450]
44. Lawrence CB, McDonnell DP, Ramsey WJ. Analysis of repetitive sequence elements containing tRNA-like sequences. *Nucleic Acids Res.* 1985; 13:4239–52. [PubMed: 3839306]
45. Bertling WM. Full length L1 retroposons contain tRNA-like sequences near the 5' terminus--hypothesis on the replication mechanism of retroposons. *J. Theor. Biol.* 1989; 138:185–94. [PubMed: 2558257]
46. Bak AL, Jorgensen AL. RNA polymerase III control regions in retrovirus LTR, Alu-type repetitive DNA, and papovavirus. *J. Theor. Biol.* 1984; 108:339–48. [PubMed: 6748697]
47. Frenkel FE, Chaley MB, Korotkov EV, Skryabin KG. Evolution of tRNA-like sequences and genome variability. *Gene.* 2004; 335:57–71. [PubMed: 15194190]
48. Geiduschek EP, Tocchini-Valentini GP. Transcription by RNA polymerase III. *Annu. Rev. Biochem.* 1988; 57:873–914. [PubMed: 3052292]
49. Oliviero S, Monaci P. RNA polymerase III promoter elements enhance transcription of RNA polymerase II genes. *Nucleic Acids Res.* 1988; 16:1285–93. [PubMed: 2831496]

50. Kirchner J, Connolly CM, Sandmeyer SB. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science*. 1995; 267:1488–91. [PubMed: 7878467]
51. Siol O, et al. Role of RNA polymerase III transcription factors in the selection of integration sites by the dictyostelium non-long terminal repeat retrotransposon TRE5-A. *Mol. Cell. Biol.* 2006; 26:8242–51. [PubMed: 16982688]
52. Stevens SW, Griffith JD. Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1s repetitive elements. *Proc. Natl. Acad. Sci. U S A.* 1994; 91:5557–61. [PubMed: 8202527]
53. Stevens SW, Griffith JD. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* 1996; 70:6459–62. [PubMed: 8709282]
54. Noma K, Cam HP, Maraia RJ, Grewal SI. A role for TFIIC transcription factor complex in genome organization. *Cell.* 2006; 125:859–72. [PubMed: 16751097]
55. Roda RH, et al. Role of the Reverse Transcriptase, Nucleocapsid Protein, and Template Structure in the Two-step Transfer Mechanism in Retroviral Recombination. *J. Biol. Chem.* 2003; 278:31536–46. [PubMed: 12801926]
56. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA.* 2004; 10:1507–17. [PubMed: 15383676]
57. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 2000; 302:205–17. [PubMed: 10964570]

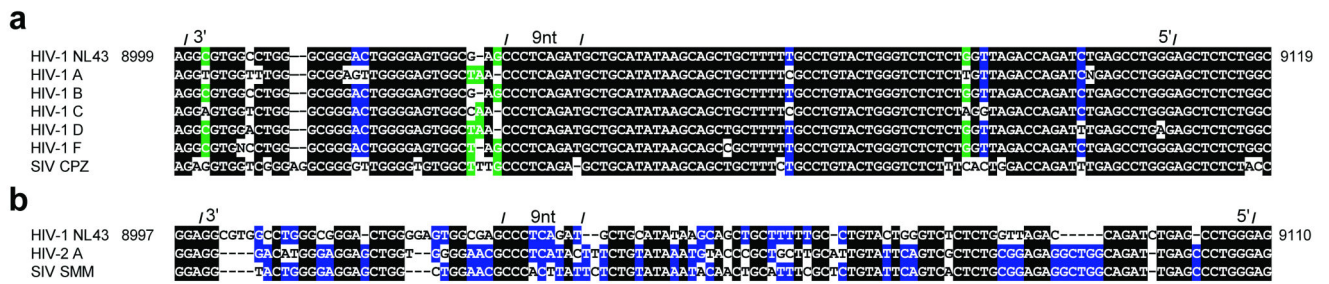




**Figure 1.** A sequence in HIV-1 U3/R complementary to tRNA<sub>3</sub><sup>Lys</sup>. **(a)** The nine-nucleotide segment in U3 (“9nt”) proposed to interact with tRNA<sub>3</sub><sup>Lys</sup> is a part of a longer sequence with additional complementarities to tRNA<sub>3</sub><sup>Lys</sup>, and could derive from an ancient tRNA gene incorporated at an early stage of HIV-1 evolution. The sequence was identified through application of computational methods, as a region of HIV-1 with a low free energy for hybridization to tRNA<sub>3</sub><sup>Lys</sup>. A 29-nucleotide segment of the tRNA gene-like sequence does not have complementarity to tRNA<sub>3</sub><sup>Lys</sup> and is present in a position expected for an intron. The TATA box (b) and sites recognized by several transcription factors (a - Sp1, Sp3; c - RBF-2; d — NFκB) as well as TAT protein (e) are marked in gray. For the list of all transcription factor sites, see review<sup>27</sup>. The U3 and R elements and the transacting responsive (TAR) sequence are indicated. The black underlines indicate likely regions of interaction with tRNA<sub>3</sub><sup>Lys</sup>. **(b)** An example of a tRNA with an intron present in its gene. The sequence shown is tRNA<sup>Lys</sup> (gene scaffold\_208396) of domestic cat (*Felis catus*).

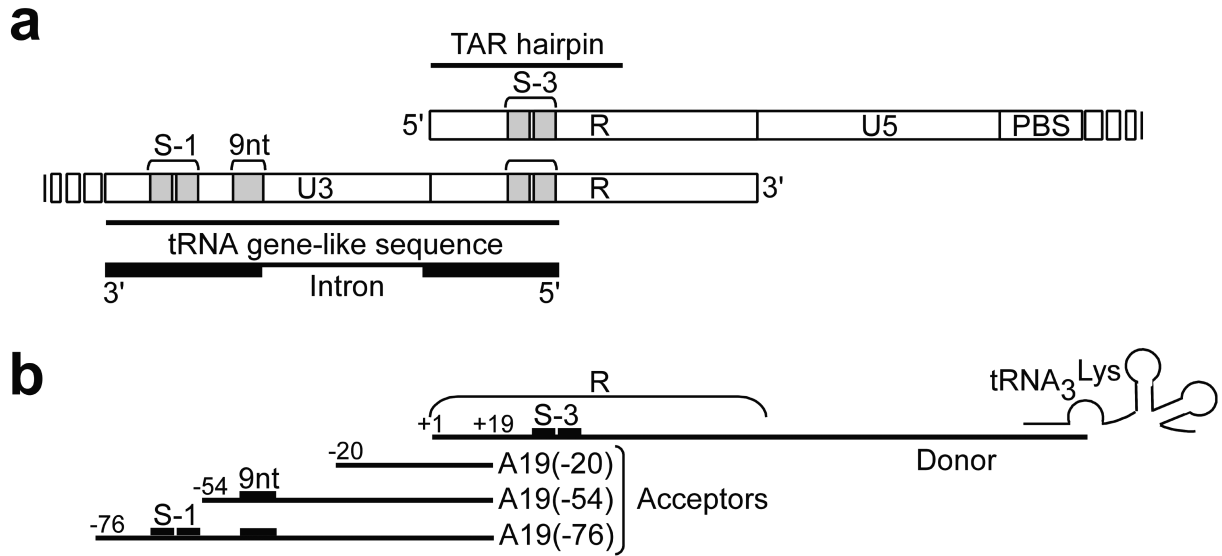


**Figure 2.** Distribution of binding stabilities in kcal mol<sup>-1</sup> of shuffled tRNA<sub>3</sub><sup>Lys</sup> to HIV-1. Ten thousand random sequences were generated with nucleotide content identical to tRNA<sub>3</sub><sup>Lys</sup> and used to scan the HIV-1 genome for hybridization. The figure presents the histogram plot for the free energy of hybridization of the shuffled sequences to the entire HIV genome. The hybridization free energies for tRNA<sub>3</sub><sup>Lys</sup> hybridization to the PBS and to the U3/R region are indicated for comparison.

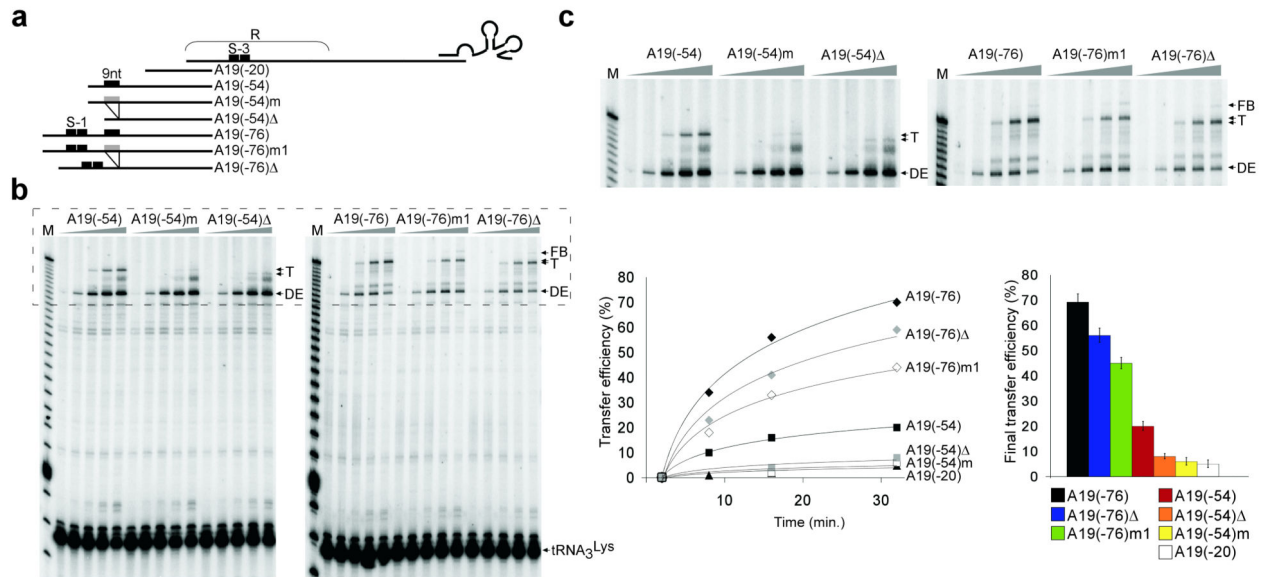


**Figure 3.**

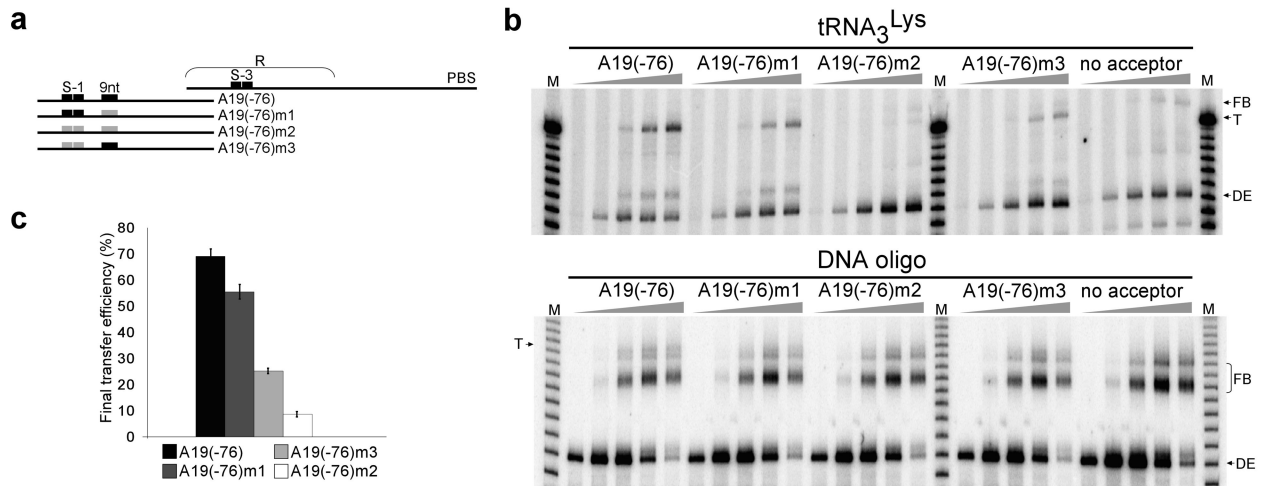
The tRNA gene-like sequence is well preserved in HIV-1 and SIVcpz. The LTR of HIV-1 with the tRNA gene-like sequence (HIV-1 NL43), was aligned with (a) LTRs of several HIV-1 subtypes and SIVcpz, as well as with (b) HIV-2, and SIVsmm. The tRNA gene-like sequence with the nine-nucleotide segment is indicated. 3', 5' and 9nt indicate respectively 3' end, 5' end and nine nucleotide motif in complementary tRNA<sub>3</sub><sup>Lys</sup>. The sequences of each HIV and SIV group represent the consensus sequences for each virus subtype.

**Figure 4.**

A system for analysis of the influence of tRNA<sub>3</sub><sup>Lys</sup>/U3 interactions during minus strand DNA transfer. (a) Schematic of the 5' and 3' end regions of the HIV-1 RNA genome, where minus strand transfer occurs. The location of the tRNA gene-like sequence is marked and three segments (S-1, 9nt, S-3) possibly supporting interactions between tRNA<sub>3</sub><sup>Lys</sup> and the HIV-1 genome are indicated. Two segments, S-1 and 9nt are present in U3. S-3 is within the TAR hairpin in the R elements at both ends of the RNA genome. (b) The RNA templates representing the 5' end (donor) and 3' end (acceptor) of HIV-1 used to measure minus strand transfer *in vitro*. The donor template spans the R, U5 and PBS. The acceptors were shortened at the 3' end to include only 19nt of the R element. This excludes the invasion-driven mechanism of transfer, but allows for terminus transfer. The 5' end of different acceptors is extended to include 20 nt of the U3 region with no complementarity to tRNA<sub>3</sub><sup>Lys</sup> (A19(-20)); 54nt of U3 with the 9nt segment complementary to 3' stem of the anticodon loop (A19(-54)); and 76nt of U3, which additionally has S-1 complementary to nucleotides at 3' end of the tRNA<sub>3</sub><sup>Lys</sup> (A19(-76)).

**Figure 5.**

Analysis of the  $tRNA_3^{Lys}/U3$  interaction during minus strand DNA transfer. **(a)** Schematic of the experimental design has the same description as in the legend of Figure 4b. Segment 9nt with mutations eliminating predicted interactions with  $tRNA_3^{Lys}$  is marked with gray color. Acceptors without marked segment 9nt (A19(-54) $\Delta$ , A19(-76) $\Delta$ ) represent RNA templates with deletion of this motif. **(b)** Gel fractionation of the transfer reaction mixtures with labeled human  $tRNA_3^{Lys}$  used as a primer and sampled at 2, 4, 8, 16, and 32 min. Transfer products (T), donor extension products (DE), and folding back products (FB) are indicated. The portion of the gel picture in the dashed frame is magnified on the right. **(c)** Quantitative data showing transfer efficiencies of each reaction in a time-dependent manner (left). Results were plotted for the transfer reaction with A19(-76) ( $\blacklozenge$ ), A19(-76) $\Delta$  ( $\blacklozenge$ ), A19(-76)m1 ( $\diamond$ ), A19(-54) ( $\blacksquare$ ), A19(-54) $\Delta$  ( $\blacksquare$ ), A19(-54)m ( $\square$ ), and A19(-20) ( $\blacktriangle$ ) (gel not shown). The final transfer efficiencies of the 32 min reactions are on the right. The transfer efficiency at each time point here and in Figure 6 was calculated as  $100\% - T/(T + DE + FB)$ . Values here and in Figure 6 were averaged from at least three independent experiments.



**Figure 6.** Effects of mutations within U3 on the minus strand transfer reaction primed from tRNA<sub>3</sub><sup>Lys</sup> and the DNA oligonucleotide. **(a)** Schematic of the experimental design has the same description as in the legend of Figure 4b. Segments S-1 and 9nt with mutations eliminating predicted interactions with tRNA<sub>3</sub><sup>Lys</sup> are marked with grey color. **(b)** Transfer reaction mixtures with different acceptors and primed with tRNA<sub>3</sub><sup>Lys</sup> (upper) and DNA oligomer (lower) were sampled at 2, 4, 8, 16, 32 min. and fractionated on a gel. **(c)** Quantitative data showing final transfer efficiencies of 32 min. for reactions primed with tRNA<sub>3</sub><sup>Lys</sup>.



**Table 1**

The summary of *P*-values representing the homology of the LTR to tRNA<sub>3</sub><sup>Lys</sup> in HIV-1, two examples of HIV-2 and different SIV species.

<b>virus</b>	<b><i>P</i>-value</b>
HIV-1	<b>2.14×10<sup>-14</sup></b>
SIVcpz tan	1.89×10 <sup>-8</sup>
SIVmnd-2	2.76×10 <sup>-6</sup>
SIVlhoest	1.99×10 <sup>-5</sup>
SIVdrl	5.47×10 <sup>-5</sup>
SIVagm tan	1.88×10 <sup>-4</sup>
SIVagm aa	8.36×10 <sup>-4</sup>
SIVagm	6.24×10 <sup>-3</sup>
SIVsyc	1.31×10 <sup>-2</sup>
SIVmac	3.03×10 <sup>-2</sup>
SIVcpz ×5	3.72×10 <sup>-2</sup>
SIVsm	7.43×10 <sup>-2</sup>
SIVden	9.00×10 <sup>-2</sup>
SIVstm	1.03×10 <sup>-1</sup>
pSIVgml	1.76×10 <sup>-1</sup>
SIVrcm	3.01×10 <sup>-1</sup>
SIVmnd-1	3.72×10 <sup>-1</sup>
SIVagm sab	3.87×10 <sup>-1</sup>
HIV-2eho	4.20×10 <sup>-1</sup>
SIVwrc	5.44×10 <sup>-1</sup>
HIV-2ben	5.80×10 <sup>-1</sup>
SIVsun	5.87×10 <sup>-1</sup>
SIVcpz	6.06×10 <sup>-1</sup>