3342–3353 *Nucleic Acids Research, 2009, Vol. 37, No. 10* *Published online 26 March 2009*
doi:10.1093/nar/gkp187

The structure of the human tRNA^{Lys3} anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs

Yann Bilbille[1], Franck A. P. Vendeix[1], Richard Guenther[2], Andrzej Malkiewicz[3], Xavier Ariza[4], Jaume Vilarrasa[4] and Paul F. Agris[1,*]

[1]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695-7622, [2]Department Plant Pathology, North Carolina State University, Raleigh, NC 27695-7616, USA, [3]Institute of Organic Chemistry, Technical University, Łodz 90-924, Poland and [4]Department of Organic Chemistry, Faculty of Chemistry, University of Barcelona, Barcelona 08028, Spain

Received December 10, 2008; Revised March 5, 2009; Accepted March 6, 2009

ABSTRACT

Replication of human immunodeficiency virus (HIV) requires base pairing of the reverse transcriptase primer, human tRNA^{Lys3}, to the viral RNA. Although the major complementary base pairing occurs between the HIV primer binding sequence (PBS) and the tRNA's 3′-terminus, an important discriminatory, secondary contact occurs between the viral A-rich Loop I, 5′-adjacent to the PBS, and the modified, U-rich anticodon domain of tRNA^{Lys3}. The importance of individual and combined anticodon modifications to the tRNA/HIV-1 Loop I RNA's interaction was determined. The thermal stabilities of variously modified tRNA anticodon region sequences bound to the Loop I of viral sub (sero)types G and B were analyzed and the structure of one duplex containing two modified nucleosides was determined using NMR spectroscopy and restrained molecular dynamics. The modifications 2-thiouridine, s^{2}U$_{34}$, and pseudouridine, Ψ$_{39}$, appreciably stabilized the interaction of the anticodon region with the viral subtype G and B RNAs. The structure of the duplex results in two coaxially stacked A-form RNA stems separated by two mismatched base pairs, U$_{162}$•Ψ$_{39}$ and G$_{163}$•A$_{38}$, that maintained a reasonable A-form helix diameter. The tRNA's s^{2}U$_{34}$ stabilized the interaction between the A-rich HIV Loop I sequence and the U-rich anticodon, whereas the tRNA's Ψ$_{39}$ stabilized the adjacent mismatched pairs.

INTRODUCTION

Retrovirions are packaged with an enrichment of a host cell tRNA for the purpose of priming reverse transcription. Human immunodeficiency virus type 1, HIV-1 and all other lentiviruses, evolved to select and recruit the cytoplasmic tRNA^{Lys3} and the other tRNA^{Lys} isoacceptors, tRNA^{Lys1} and tRNA^{Lys2} (1). While the function of tRNA^{Lys1,2} in HIV-1 is not known, the selective packaging of HIV-1 reverse transcriptase primer tRNA^{Lys3} is required for optimizing both the annealing of tRNA^{Lys3} to viral RNA and the infectivity of the HIV-1 population (2). A considerable number of interactions are believed to occur between the human tRNA^{Lys3} and HIV-1, as determined by chemical and enzymatic probes (3). Eighteen nucleosides of the tRNA's 3′-terminal sequence (Figure 1A), part of the amino-acid accepting stem (nucleosides 66–76) and the 3′-side of the TΨC domain (tRNA nucleosides 59–65) (4), are bound to a complementary sequence in the viral RNA, the Primer Binding Sequence (PBS) (viral nucleosides 183–200), while the TΨC domain interacts with the Primer Activation Sequence (PAS) (5) (Figure 1A). The uridine-rich anticodon and adjacent nucleosides are bound to the adenosine-rich Loop I of the viral RNA (Figure 1A and C) (4). Deletion of the A-rich Loop I of HIV-1 that is complementary to the uridine-rich anticodon domain resulted in significantly reduced levels of infectivity and reduced synthesis of viral RNA (6–8). Constitution of a virus that was able to use a host cell tRNA other than tRNA^{Lys3} required not only the conversion of the PBS to the complement of the investigator-selected tRNA, but also the mutation of the AAAA sequence of HIV-1

*To whom correspondence should be addressed. Tel: +1 919 515 6188; Fax: +1 919 515 2047; Email: paul_agris@ncsu.edu

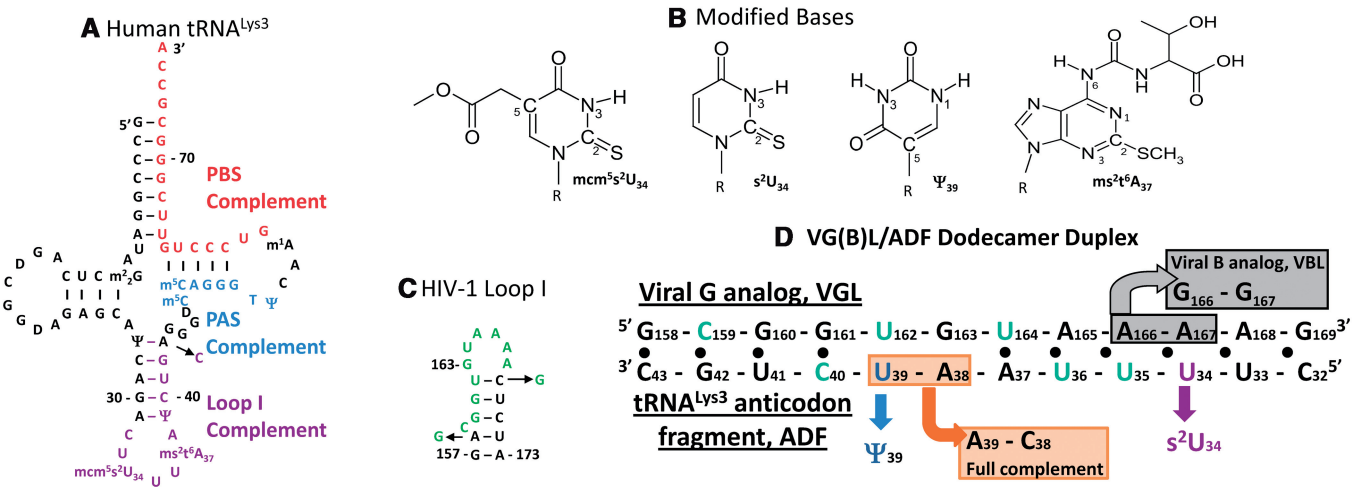

Figure 1. Sequence and secondary structure of the $tRNA^{Lys3}$ interaction with Loop I of the HIV genome. (**A**) Sequence and secondary structure of human $tRNA^{Lys3}$. The 3′-terminal, 18 nucleoside sequence of $tRNA^{Lys3}$ (in red) is complementary to the HIV-1 Primer Binding Site, PBS. Residues in blue are complementary to the HIV-1 Primer Activation Signal, PAS. In purple are the residues of the anticodon domain fragment, ADF, studied in the article. In order not to have frayed and weak termini to the dodecamer duplex, a C_{43} was substituted for A_{43} to constitute a terminal C_{43}•G_{158} base pair. The native post-transcriptional modifications are indicated according to standard nomenclature (80). (**B**) Chemical structures of the modified nucleosides in the ADF of human $tRNA^{Lys3}$. The natural modification occurring at position 34 is 5-methoxycarbonylmethyl-2-thiouridine ($mcm^5s^2U_{34}$) and that at position 37 is 2-methylthio-N6-threonylcarbamoyladenosine ($ms^2t^6A_{37}$). The uridine at position 39 is modified to pseudouridine, Ψ_{39}. The modifications studied here were 2-thiouridine (s^2U_{34}) and pseudouridine, Ψ_{39}. (**C**) The sequence and secondary structure of the A-rich Loop I of HIV-1 (residues 157–173). The residues in green (158–169) are those residues studied in this article. Residues A158 and C169 were changed to G158 and G159 in order that the dodecamer duplex have terminal G•C base pairs. (**D**) Duplexes composed of a fragment of the $tRNA^{Lys3}$ anticodon domain, the ADF, and the A-rich Loop I of HIV-1, subtype G (VGL). Constructs included the individual and doubly modified ADFs with 2-thiouridine, s^2U_{34} (purple) and pseudouridine, Ψ_{39} (blue). An additional construct of the ADF was substituted at positions 38 and 39 to create a full complement to the VGL (orange). The labeled pyrimidine nucleosides, [^{15}N3]-uridine and 4-amino-[$^{15}NH_2$]-cytidine, are designated (green). The HIV-1 subtype B construct had a sequence with two Gs substituting for A_{166} and A_{167} (gray).

Loop I to complement the anticodon sequence of the new tRNA (9–11). Changing one or the other of these sequences was not sufficient for maintaining the new interaction during prolonged culturing. Binding of the tRNA's anticodon region to the A-rich Loop I region has been associated with transition of viral replication from initiation to the highly processive elongation with a possible requirement for the tRNA's post-transcriptional modifications (12).

The $tRNA^{Lys3}$ anticodon domain has post-transcriptional modifications (5-methoxycarbonyl-methyl-2-thiouridine, $mcm^5s^2U_{34}$; 2-methylthio-N6-threonylcarbamoyladenosine, $ms^2t^6A_{37}$; and pseudouridine, Ψ_{39}) that are critical for the decoding of its cognate (AAA) and wobble codons (AAG) on the ribosome (13) and affect its interaction with the HIV Loop I (14). Fully modified, native $tRNA^{Lys3}_{SUU}$ when bound to the viral RNA protected the HIV-1 Loop-I AAAA sequence (viral nucleosides 169–172) from chemical reactivity and enzymatic cleavage (14). In contrast, the unmodified *in vitro* transcript of the tRNA was unable to block chemical reactivity of the same sequence (14). In complex with viral RNA, the tRNA anticodon was also protected from chemical and enzymatic probes (3) and the modified nucleoside $mcm^5s^2U_{34}$ (Figure 1B) can be crosslinked to the viral RNA (15). When the complementarity between the anticodon of tRNA and the A-rich Loop 1 was disrupted, all four helices of the extended tRNA/HIV interactions were destabilized (16). The importance of the thionyl of $mcm^5s^2U_{34}$ to the interaction of the tRNA's anticodon with the viral Loop I was demonstrated when its oxidation was associated with a decrease in the tRNA's protection of Loop I (14). Thus, tRNA's 2-thio modification of U_{34} may be as important in initiating viral replication (17), as it is in recognizing cognate codon AAA in translation (18).

The chronology of events that lead to the interaction of the U-rich, modified anticodon domain with the A-rich viral Loop I and formation of a duplex could involve an intermediate 'kissing-loop' interaction of the two stem and loop hairpins (19,20). However, the 5′-side of the tRNA's anticodon stem, when in complex with the PBS and Loop I, is susceptible to chemical and enzymatic probes (3). In addition, the HIV nucleocapsid protein NCp7 stably denatures $tRNA^{Lys3}$ in the presence of the complementary HIV sequences (21–23). Whatever the mechanism that leads to formation of the replication initiation complex, a duplex apparently results from the annealing of the anticodon domain to the HIV Loop I. Yet, the contributions of the modified nucleosides to the stability and structure of the resulting duplex are not known.

To investigate the effect of modified nucleosides on the stability and structure of the duplex that is formed from the anticodon domain and the viral Loop I, an experimental system was designed to mimic the linear, denatured form of the RNAs. Here, we report how two of the more important individual tRNA-modified nucleosides contribute to the stabilization of the tRNA anticodon domain interaction with both the HIV sub

(sero)type G, exhibiting the highest complementariness of the Loop I sequences with the anticodon stem and loop domain of $tRNA^{Lys3}$ (ASL), and sub(sero)type B, the most prevalent subtype in the Americas, Europe and Oceania. The modified nucleosides s^2U_{34} and Ψ_{39} were most effective at stabilizing the interaction between the tRNA's anticodon region and the viral A-loop sequence. The NMR-derived structure of one duplex containing the two modified nucleosides, s^2U_{34} and Ψ_{39} resulted in two coaxially stacked A-form RNA stems separated by two mismatched base pairs, U_{162}•Ψ_{39} and G_{163}•A_{38}, that maintained a C1′–C1′ distance between 10 and 12.5 Å that is comparable to that of the A-form RNA helix diameter.

MATERIALS AND METHODS

Oligonucleotide synthesis

The $tRNA^{Lys3}$ anticodon domain fragments (ADF) and HIV Loop I sequences corresponding to the subtypes G and B (VGL, VBL) were synthesized (Nucleic Acids Facility at North Carolina State University, Raleigh, NC) as dodecamers (Figure 1) with modified nucleosides and the ^{15}N-labeled nucleosides [$^{15}N_3$]-uridine and the 4-amino-[$^{15}NH_2$]-cytidine. The protected, s^2U phosphoramidite was synthesized as published (24). The [$^{15}NH_2$]-cytidine, and $^{15}N_3$-uridine were chemically synthesized (25) and then, derivatized to the protected 3′-phosphoramidite (ChemGenes, Ashland, MA). Oligomers were purified by HPLC (26), desalted (Waters Corporation Sep-pak columns, Milford, MA) and determination of the nucleoside composition of the RNA was performed by enzymatic hydrolysis followed reverse phase HPLC and UV spectroscopy on-the-fly (27). NMR samples were equilibrated to buffer (10 mM sodium phosphate, pH 6.5, 100 mM NaCl, 90% H_2O and 10% D_2O) using Amicon Centricon 3 concentrators (Millipore, Bedford, MA). For experiments in D_2O, the sample solutions were lyophilized and exchanged with D_2O (Cambridge Isotopes, Andover, MA). The VGL and VBL dodecamers were titrated with the various ADF oligomers and monitored by 1D NMR. The VGL/ADF duplexes had a final duplex concentration ranging from 0.5 mM to 1.2 mM.

UV and circular dichroism spectroscopy

Thermal denaturations and renaturations were monitored by UV absorbance at 260 nm with a Cary 3 spectrophotometer running WinUV version 3.00. Buffer conditions were identical to those for NMR. Three successive denaturations and renaturations were conducted over a temperature range of 5–55°C and a concentration range of RNA from 1 μM to 1 mM using cuvettes of 1.0 and 0.2 cm pathlengths. The temperature ramp rate was 1.0°C/min or 0.5°C/min with a data sampling interval of 0.2 min or 0.5 min. Circular dichroism spectra were collected (Jasco J600) at 4°C with 1 cm path length using a jacketed, cylindrical sample cell and RNA sample concentrations adjusted to ~0.2 absorbance units at 260 nm (28,29).

NMR spectroscopy

NMR spectra were collected on a Bruker DMX 500 MHz instrument equipped with a triple-resonance ^1H, ^{13}C, and ^{15}N probe and three-axis pulsed field gradient capabilities, on a Varian 600 MHz instrument equipped with a cryoprobe at NCSU, and on a Varian 800 MHz spectrometer at Duke University (Durham, NC). All NMR data were processed using NMRPIPE (30). Spectra were displayed and analyzed using NMRVIEW 5.0 (31) or SPARKY software (32). NOESY and TOCSY spectra using the MLEV-17 mixing sequence were collected at varying temperatures, and mixing times to aid in the assignment process. T_1 relaxation experiments were conducted at various temperatures on the VGL/ADF-s^2U_{34};Ψ_{39} duplex to identify H2 resonance. A DQF-COSY experiment was conducted at 20°C in D_2O with the VGL/ADF-s^2U_{34} and the VGL/ADF-s^2U_{34};Ψ_{39} duplexes. Two-dimensional heteronuclear spectra aided in the assignment process. Natural abundance ^1H-^{13}C HSQC experiments were performed on VGL/ADF-s^2U_{34} in H_2O at 4°C and VGL/ADF-s^2U_{34};Ψ_{39} in D_2O at 20°C. ^1H-^{31}P HETCOR and a ^1H-^{31}P HeteroTOCSY were collected on VGL/ADF-s^2U_{34};Ψ_{39} in D_2O at 20°C. For samples with site-specific ^{15}N labels, ^1H–^{15}N HSQC spectra were obtained in H_2O at seven temperatures from 2°C to 35°C and at the concentrations used for structural analyses.

NMR restraints generation and structure calculation

Measurement of individual peak intensities, when possible, was performed using the box integration method in the SPARKY software. The cross-peaks were qualitatively classified as strong (0.0–3.5 Å), medium (0.0–4.5 Å), or weak (0.0–5.5 Å). For peaks that overlapped, standard A-form RNA distances were used when an A-form RNA was deduced from the NMR spectra. No restraints were used for the H5′ and H5″ protons. Intervals with wide boundaries, 0.0–5.0 or 6.0 Å, were used for restraints involving the imino protons. A total of 292 restraints were used to generate the structure (Table 1). Consistent with NMR data, the nine G•C and U•A base pairs were subjected to six and four hydrogen-bond restraints, respectively (1.8–2.5 Å between the hydrogen and acceptor, and 2.7–3.5 Å between the heavy atoms). The G•U base pair was restrained using four hydrogen-bond distance restraints between the GO6 and the UN3H and N3 atoms, and between the UO2 and the GN1 and GN1H atoms. No hydrogen-bonds restraints were used for Ψ_{39}•U_{162} and G_{163}•A_{38} base pairs, but an NMR-derived, imino distance restraint between the U_{162}N3H and Ψ_{39}N3H was introduced with an interval of 0.0–6.0 Å. A total of 47 hydrogen bond distance restraints were finally used. Planarity restraints (5–50 kcal/mol $Å^2$) were applied to Watson–Crick and G•U base pairing. Because of severe overlapping in DQF-COSY and HETCOR spectra, torsion angle restraints were set with wide ranges of ±30° and not determined from the direct measurement of coupling constants. β (P-O5′-C5′-C4′), γ (O5′-C5′-C4′-C3′) and ε (C4′-C3′-O3′-P) dihedral angles were set respectively to 178 ± 30°, 54 ± 30° and −153 ± 30°. When the presence of the regular A-form

Table 1. Restraints and structural statistics

NOEs distances restraints	292
Internucleotide NOEs	139
Intranucleotide NOEs	115
Involving iminos	38
Hydrogen bonds restraints	47
Torsion angles restraints	156
Planarity restraints	11
NOEs per residue	12.2
Total restraints per residue	18.7
Distance violations (Å)	None > 0.5
Dihedral violations (°)	None > 5.0
Backbone RMSD for all atoms	
Compared to the average structure (Å)	2.55 ± 0.89
Stems RMSD:	
(G_{158}–G_{161}/C_{40}–C_{43}) (Å)	0.80 ± 0.29
(A_{166}–G_{169}/C_{32}–U_{35}) (Å)	0.86 ± 0.39
Loop and adjacent residues (G_{161}–U_{164}/A_{37}–C_{40}) (Å)	1.01 ± 0.34

RNA structure could be determined, using both the ^{31}P chemical shift and the expected connectivity patterns between the aromatic and sugar resonances, the α (O3′-P-O5′-C5′: −68 ± 30°) and ζ (C3′-O3′-P-O5′: −71 ± 30°) torsion angles were constrained to exclude the *trans* conformation. For all residues, except G_{163}, the glycosidic torsion angles, χ [O4′-C1′-(N9/N1)-(C4/C2)], were restrained to the *anti* conformation (−158 ± 30°); the intensities of all the H1′-H6 to H8 connectivities exhibited smaller volumes than the H5 to H6 connectivity. For all residues except G_{163}, the δ angles were restrained to be in C3′-endo conformation. A total number of 156 torsion angles restraints were used for the structure calculation (Table 1).

Molecular modeling of the VGL/ADF-s^2U_{34};Ψ_{39} structure was achieved using CNS 1.1 (33). One hundred structures were calculated using standard NMR restraint annealing protocol (34). Structures with low total energy, without or with a low number of dihedral or distance restraint violations were chosen for further analysis. The helical parameters of the structure were analyzed using X3DNA software (35). To validate our structures, back calculations of the H5, H6 and H8 protons resonance chemicals shifts were conducted using MOLMOL (Johnson-Bovey model), or NUCHEMICS software (36). The output structures were visualized with MOLMOL (37) and PYMOL (38).

RESULTS

Thermodynamic stability and affinities of HIV-1 A-loop interactions with variously modified fragments of the $tRNA^{Lys3}$ anticodon domain

The thermal stabilities and the circular dichroism spectra of the duplexes formed between the U-rich anticodon domain dodecamer (ADF) of $tRNA^{Lys3}$ and the A-rich HIV Loop I indicated that the s^2U_{34} and Ψ_{39} facilitated the interaction of the two RNAs. The dodecamer sequences, based on previous studies of the complex formed between $tRNA^{Lys3}$ and the HIV-1 RNA (3),

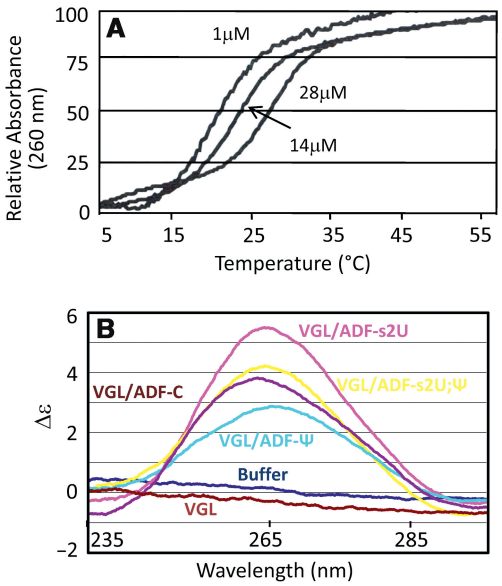

Figure 2. Thermal stability and base stacking of the VGL/ADF duplex. (**A**) UV-monitored, thermal stability assay of the VGL/ADF-s^2U_{34} duplex. The duplex was repeatedly denatured and renaturated and monitored by UV-absorbance. The T_m of a duplex is concentration dependent. Results from monitoring three concentrations of the VGL/ADF-s^2U_{34} duplex are shown: 1, 14 and 28 μM. (**B**) Circular dichroism spectra of the VGL/ADF duplexes. The circular dichroism spectrum of each duplex was collected with the RNA concentration of the duplex or the VGL alone adjusted to give an absorbance of 0.2 units (A_{260}) at 25°C. Spectra were collected at 4°C and were not corrected for the background (buffer).

were constituted with substitution of C_{43} for A_{43} to reduce fraying of the duplex by forming a terminal C_{43}•G_{158} base pair. This construct allowed us to focus our studies on the contributions made by the two internal base modification chemistries, s^2U_{34} and Ψ_{39}, and the possible mismatch base pairs. As it would be expected for the duplexes, the melting temperatures, T_m, of these bimolecular interactions were concentration dependent (Figure 2A) allowing the binding affinities to be extracted from the results of the UV-monitored thermal experiments (Table 2). The contribution of modified nucleosides to the stability of the VGL/ADF complex was dependent on the $tRNA^{Lys3}$ modification introduced. Relative to the unmodified ADF (T_m = 30.0°C at 100 μM; K_d = 320 μM), the individual incorporations of Ψ_{39} and s^2U_{34} raised the melting temperature of the VGL/ADF duplexes by +1.2°C and +3.9°C, and enhanced the affinity of the ADF for the VGL by some 8- and 72-fold, respectively. Incorporation of both the Ψ_{39} and s^2U_{34} modifications significantly increased the T_m (+7.2°C) and the affinity of the ADF for the viral sequence (Table 2). The affinity of the doubly modified ADF-s^2U_{34};Ψ_{39} (K_d = 14.2 μM) for the VGL was significantly stronger than that of the unmodified ADF (K_d = 324 μM) (Table 2). Interestingly, the affinity of the fully modified anticodon stem and loop domain for its cognate codon AAA in the A-site of the *Escherichia coli* ribosome was comparable (K_d = 3.1 μM) to that of the ADF-s^2U_{34};Ψ_{39} for the VGL, while ribosomal codon binding of the unmodified stem and loop

Table 2. Thermodynamic parameters of VGL and VBL duplex formation with various ADF constructs

VGL or VBL	ADF of $tRNA^{Lys3}$	K_d (μM)	ΔG°_{298}	$\Delta\Delta G^\circ_{298}$	T_m (100 μM)
VGL	ADF	324.0	−4.8	–	30.0
	ADF-Ψ_{39}	43.0	−5.9	1.1	31.2
	ADF-s^2U_{34}	4.5	−7.3	2.5	33.9
	ADF-s^2U_{34};Ψ_{39}	14.2	−6.6	1.8	37.2
VBL	ADF	270.0	−4.9	0.1	30.6
	ADF-s^2U_{34}	5.7	−7.2	2.4	33.0
VGL	ADF of $tRNA^{Lys2}$	>1900	–	–	

Gibbs standard free energy in kcal/mol was calculated at 25°C from: $\Delta G^\circ = -RT\ln K$.

domain was too low to be determined (E. Gustilo and P.F. Agris, personal communication). The standard free energy values, ΔG°, from duplex formation of the VGL with the ADF-Ψ_{39}, ADF-s^2U_{34}, and ADF-s^2U_{34};Ψ_{39} were lower than that formed from the VGL and the unmodified ADF (Table 2). The ΔG° value for the doubly modified duplex was between that of the two singularly modified duplexes. We expect that the free energy of the VGL/ADF-s^2U_{34};Ψ_{39} duplex is a composite of local enthalpic and entropic contributions, influenced by the effect of modified nucleosides on local structure and stability. Thus, it is not surprising that the ΔG° of the doubly modified duplex was not simply the sum of the two duplexes containing only one modification each (s^2U_{34} or Ψ_{39}).

A fourth duplex (VGL/ADF-C), formed from a fully complementary ADF sequence in which $A_{39}C_{38}$ substituted for $U_{39}A_{38}$, exhibited a T_m (42.2°C) that was 8.3°C higher than the VGL/ADF-s^2U_{34};Ψ_{39}, and 12°C higher than the duplex formed with the unmodified ADF and two mismatched pairs. In addition to increasing thermal stability, the s^2U_{34} and Ψ_{39} enhanced the degree of base stacking within the VGL/ADF duplexes (Figure 2B), as monitored by circular dichroism spectroscopy (39). The spectral ellipticity of the VGL/ADF was increased with the addition of the single and combined modifications. The duplex with ADF-s^2U_{34} exhibited the highest degree of ellipticity and presumably the greatest degree of base stacking. Surprisingly, the ellipticity of the VGL/ADF-C was comparable to that of the VGL/ADF-Ψ_{39}, and lower than that of the VGL/ADF-s^2U_{34} (Figure 2B). Here, we see again that the ellipticity of the doubly modified complex, VGL/ADF-s^2U_{34};Ψ_{39}, is not simply the sum of that of the two singularly modified complexes VGL/ADF-s^2U_{34} and VGL/ADF-Ψ_{39}.

The ADFs had four contiguous uridines (U_{33}–U_{36}) complementary to the four contiguous adenosines of the dodecamer of the HIV-1 Loop I subtype G (VGL), but were still capable of binding to the Loop I substitutions that are found in other subtypes, such as the two guanosines of the B-subtype (VBL) (Figure 1D). The affinities of the VBL for the unmodified ADF and the ADF-s^2U_{34} were comparable to the VGL (Table 2). There are three human $tRNA^{Lys}$ species. Notably, an unmodified ADF with the sequence of the human $tRNA^{Lys1,2}_{CUU}$ having a C_{34} was hardly bound by the VGL (Table 2).

NMR studies of the VGL/ADF interaction: exchangeable proton resonances identification and assignment

A doubly modified (s^2U_{34};Ψ_{39}) ADF bound to the VGL, VGL/ADF-s^2U_{34};Ψ_{39}, was studied by NMR. Three other duplexes were investigated by NMR: a fully base-paired complement (VGL/ADF-C); a duplex containing only the pseudouridine modification, ADF-Ψ_{39} (VGL/ADF-Ψ_{39}); and a duplex containing only the 2-thiouridine modification, ADF-s^2U_{34} (VGL/ADF-s^2U_{34}) (Figure 1D). These additional constructs aided in the assigning of NMR signals and in elucidating the effect of the different modified nucleosides of $tRNA^{Lys3}$ on the interaction of its anticodon domain with the HIV-1 A-rich Loop I (Figure 1D). All four duplexes were studied using classical procedures (40,41). One dimensional spectra of the various duplexes exhibited exchangeable protons resonances between 12.5 and 15 ppm that were attributed to seven hydrogen-bonded imino protons involved in standard Watson–Crick base pairs for the VGL/ADF-Ψ_{39}, VGL/ADF-s^2U_{34} and VGL/ADF-s^2U_{34};Ψ_{39} duplexes (Figure 3 and Supplementary Data). The imino proton resonance of s^2U_{34} was shifted downfield (14.23 ppm) compared to the corresponding resonance of the similar unmodified U_{34} (13.85 ppm). The downfield shift of the base paired imino proton of s^2U_{34} has been previously observed (42) and could indicate that the sulfur is contributing to a stronger hydrogen bond through a change in ring current or stacking interactions.

Imino proton resonances between 10 and 12 ppm corresponded to those protons involved in non-canonical base pairs. The characteristic chemical shifts (43) and observation of an intense NOE connectivity between the N1H of G_{160} and N3H of U_{41}, together with the imino–imino NOE connectivity between N3H of U_{41} and N1H of G_{161} allowed the unambiguous assignment of the imino protons of G_{160} and U_{41} in the VGL/ADF-C NOESY spectrum. The formation of this base pair was observed in all four duplexes including the VGL/ADF-s^2U_{34};Ψ_{39} (Supplementary Data). The imino protons resonances of Ψ_{39}-N1H, Ψ_{39}-N3H and U_{162}-N3H were identified and assigned on the basis of the VGL/ADF-Ψ_{39} and VGL/ADF-s^2U_{34};Ψ_{39} NOESY spectra at low temperatures (2°C and 4°C, respectively) and of the ^{15}N-HSQC spectra. The chemical shifts of the Ψ_{39} and U_{162} imino protons resonances agreed with previously published studies (44). Moreover, a weak imino–imino NOE connectivity between the Ψ_{39}-N3H and the U_{162}-N3H was observed on the 200 ms NOESY spectra of the VGL/ADF-Ψ_{39} duplex (Supplementary Data). In addition, NOE connectivities were observed between Ψ_{39}-N1H and A_{38}H2′ and A_{38}H3′ on the NOESY spectra (data not shown). A broad imino resonance at 11.50 ppm was assigned to the imino proton of G_{163}. Despite this proton being protected from exchange with solvent, no NOE connectivity between this proton and other exchangeable or non-exchangeable protons could be observed.

The thermal denaturations of the VGL/ADF-C, VGL/ADF-Ψ_{39} and VGL/ADF-s^2U_{34};Ψ_{39} duplexes were followed by NMR (Supplementary Data) and were

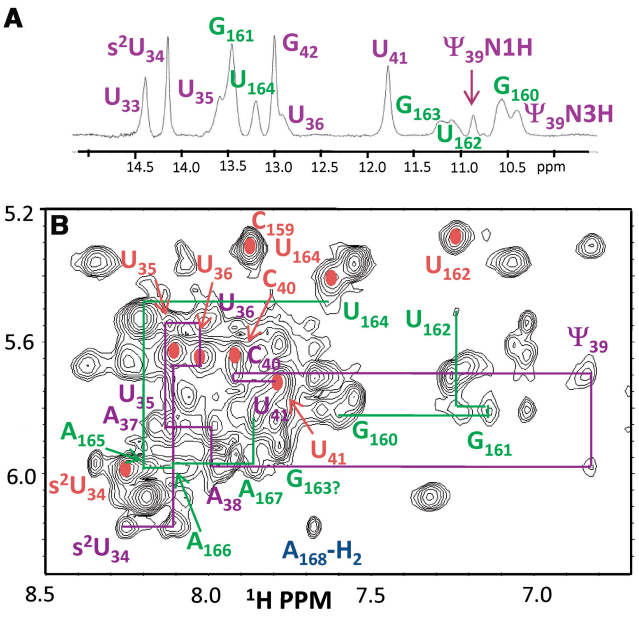

Figure 3. Imino proton resonances and NOESY sequential assignment NMR spectra of the VGL/ADF. (**A**) A 1D NMR spectrum of VGL/ADF-$s^2U_{34};\Psi_{39}$ (90% H_2O/10% D_2O and at 10°C). The region of the spectrum in which imino protons resonate (9.5–15.0 ppm) is shown. (**B**) Sequential assignment of the aromatics to H1′ resonances of the VGL/ADF-$s^2U_{34};\Psi_{39}$ duplex. The NOESY spectrum was recorded at 10°C (90% H_2O/10% D_2O) with a mixing time of 300 ms. For clarity, only a portion of the 'NOE walk' is displayed. The residues belonging to HIV-1 Loop I and to the tRNA are in green and purple, respectively. The H5-H6 connectivities are shown in rust.

consistent with the results of the UV-monitored melting studies. The VGL/ADF-C exhibited a T_m between 40°C and 45°C, whereas the T_m values of the VGL/ADF-Ψ_{39} and VGL/ADF-$s^2U_{34};\Psi_{39}$ were between 25°C and 30°C.

Non-exchangeable protons resonances identification and assignment

The non-exchangeable resonances of the VGL/ADF-$s^2U_{34};\Psi_{39}$ duplex were identified and assigned with a variety of homo- and heteronuclear NMR experiments using standard procedures (41) and the help of the three other duplexes. Using the different duplexes, all of the aromatics and H1′ resonances were assigned, excepted that of G_{163} in VGL/ADF-s^2U_{34}, VGL/ADF-Ψ_{39} and the VGL/ADF-$s^2U_{34};\Psi_{39}$ (Figure 3B and Supplementary Data). Although assignment of the G_{163} H8 and H1′ resonances remains tentative because of severe overlapping of signals, an NOE connectivity was observed between the U_{164}H6 and the G_{163}H1′ in the VGL/ADF-Ψ_{39} duplex, as well as a strong NOE between the U_{164}H6 and G_{163}H2′ in the NOESY spectra of the VGL/ADF-Ψ_{39} and VGL/ADF-$s^2U_{34};\Psi_{39}$ duplexes at low mixing times. The chemical shifts of the H5 and H6 protons are known to be most affected by the ring current from the 5′ adjacent base (36). The H5 and H6 chemical shifts of U_{164} were consistent with protons experiencing a ring current effect from its 5′-adjacent G_{163} (36), indicating that the G_{163} and U_{164} were stacked, and consequently that G_{163} is inside the helix. The U_{162}H6 and H5 chemical shifts were also consistent with a uridine experiencing the ring current of G_{161} in spite of a more important displacement toward the high-field for the U_{162}H6 chemical shift as compared to U_{164}H6. Moreover, the volume of the U_{162}H5 to H6 NOE connectivity was less than all of the other H5-H6 NOE connectivities indicating that the U_{162} was more dynamic than the other pyrimidines. An inter-strand NOE connectivity between the A_{165}H2 and the A_{37}H2 was observed for the VGL/ADF-$s^2U_{34};\Psi_{39}$.

The Ψ_{39}H6 proton resonated at 6.8 ppm, characteristic of a pseudouridine (45). An aromatic to aromatic NOE connectivity was observed between Ψ_{39}H6 and A_{38}H8, indicating that Ψ_{39} and A_{38} were stacked. Because of overlapping peaks, we were not able to unambiguously identify the connectivities of A_{38} to the A_{37} H2. The A_{168}H2 proton exhibited an unusual NOE connectivity with the s^2U_{34}H1′, indicating an enhanced stacking that we attributed to the 2-thio modification. The chemical shifts of A_{38}H8 and H1′ were consistent with an adenosine involved in a typical A-form RNA. The A_{38}H2 proton, identified and assigned by T1 relaxation experiments and confirmed through the natural abundance ^{1}H-^{13}C-HSQC spectra, relaxed slowly, as did other H2 resonances for the VGL/ADF-$s^2U_{34};\Psi_{39}$ duplex. Taken together along with strong sequential NOEs between the aromatic to H2′ protons, this information showed that the entire duplex adopted a conformation similar to a canonical A-form RNA. Moreover, the data indicated that all the bases were stacked and inserted into the helix.

Sugar conformations and backbone geometry

The DQF COSY spectra showed that all the residues of the VGL/ADF-$s^2U_{34};\Psi_{39}$ duplex adopted the C3′-*endo* sugar pucker ($J_{H1'-H2'} < 3$ Hz) characteristic of an A-form RNA helix. A narrow spread of the ^{31}P chemical shifts (2.2 ± 0.9 ppm) around that attributed to A-RNA helix-like values indicated that on average the phosphate backbone conformed to a regular A-RNA conformation without significant deformation (Supplementary Data) (41). Each of the aromatic to H1′ connectivities of the VGL/ADF-$s^2U_{34};\Psi_{39}$ had an intensity significantly lower than that for the H5–H6 indicating that all the bases, including Ψ_{39}, adopted the *anti* form of the glycosidic torsional angle (46). Nevertheless, the resonance overlaps and tentative assignment for G_{163} did not allow us to determine its glycosidic torsional angle.

Structure of VGL/ADF-$s^2U_{34};\Psi_{39}$

The spectral data indicated that the structure of the VGL/ADF-$s^2U_{34};\Psi_{39}$ duplex was similar to a standard double helical A-form RNA, that all the bases were inserted into the helix and that the 2×2 internal loop 5′-$U_{162}G_{163}$-3′/3′-$\Psi_{39}A_{38}$-5′ was structurally organized. One hundred structures were calculated using a simulated annealing protocol (34). Among the 100 calculated structures, 41 converged to satisfy all of the NOE distance and dihedral angle restraints within 0.5 Å and 5°,

respectively. The U_{162}•Ψ_{39} and G_{163}•A_{38} mismatch base pairs were observed with great attention. Among the 41 accepted structures, 23 presented an imino G•A base pair and the lowest total energy. The other 18 structures presented different kinds of G•A base pairs. Of the 23 structures exhibiting an imino G•A base pair, the 10 structures with the lowest total energies were selected for further analysis. In order to validate the selected structures, a chemical shift back-calculation was conducted for the H5 and H6 chemical shifts of C_{159} (standard), U_{162} and U_{164}, and for the Ψ_{39}H6 and G_{163}H8. Indeed, the chemical shifts of the H5 and H6 were most influenced by the 5′ bases and thus, could be used as probes to validate our structure. The back-calculated chemical shifts for the U_{164}H5 and the H6 were in excellent agreement with the observed chemical shifts for U_{164} (Supplementary Data). This confirmed the stacking of the U_{164} and G_{163} and moreover, tended to prove that the imino G•A base pair observed in the 10 selected structures was consistent with the NMR data. The predicted Ψ_{39}H6 chemical shift was consistent with the observed chemical shift. The back-calculated and experimental chemical shifts for U_{162}H5 were almost identical (Supplementary Data). This contrasted with U_{162}H6 for which these chemical shift values displayed a difference of ~0.7 ppm. In this case, the position of U_{162} in relation to G_{161} seemed to be different from that observed in our structures.

The overall structures are well defined with a mean backbone RMSD for all residues of 2.55 ± 0.89 Å. The structures have two canonical, A-form RNA stem regions of four and six Watson–Crick base pairs including the wobble G•U, respectively, separated by a structured 2×2 internal loop. Relative to the mean structure, the backbone RMSDs were 0.80 ± 0.29 Å and 0.86 ± 0.39 Å, respectively, for the G_{158}-G_{161}/C_{40}-C_{43} (Figure 4A) and the A_{166}-G_{169}/C_{32}-U_{35} (Figure 4B) A-form RNA stems. However, the backbone RMSD for the loop and the

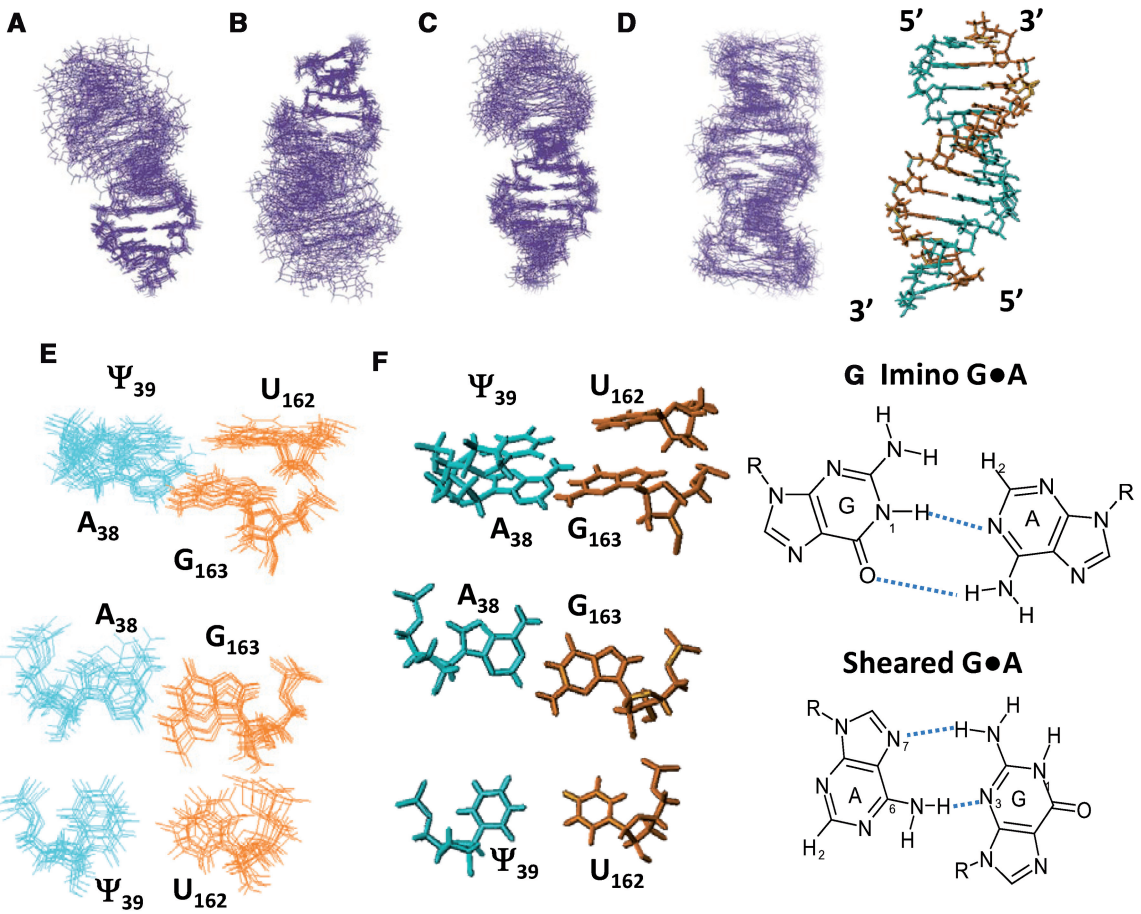

Figure 4. Structure of the VGL/ADF-s^2U_{34};Ψ_{39} duplex. The structure of the VGL/ADF-s^2U_{34}; Ψ_{39} duplex was determined from NMR-restrained, molecular dynamics. Members of a family of 10 structures without distance or torsion angle violations were superimposed. Both the U_{162}•Ψ_{39} and imino G_{163}•A_{38} base pairings are within the structures. (**A**) Best fit for the stem residues G_{158}-G_{161}/C_{40}-C_{43}. (**B**) Best fit for the stem residues A_{166}-G_{169}/U_{35}-C_{32}. (**C**) Best fit for internal residues G_{161}-U_{164}/A_{37}-C_{40}. (**D**) Superposition of the ten best structures (left) and the mean structure (right). (**E**) The G_{163}•A_{38} and U_{162}•Ψ_{39} base stacking (top) and the geometry of the two base pairings (bottom). (**F**) The mean G_{163}•A_{38} and U_{162}•Ψ_{39} base stacking (top) and mean geometry of the two base pairings (bottom). (**G**) The imino and sheared G•A base pairs. *Top:* The imino G•A base pair employs the GN1H as a donor for the AN1 acceptor, whereas the carbonyl oxygen of G is the acceptor for an amino proton of A. *Bottom:* The sheared G•A pair employs an amino proton of G as a donor with the AN7 as the acceptor, and an amino proton of A as a donor for the N3 of G.

immediately adjacent residues G_{161}-U_{164}/A_{37}-C_{40} was greater than that of the two A-form RNA stems with a value of 1.01 ± 0.34 Å (Figure 4C).

Both G_{163}, being loosely restrained, and U_{162} exhibited conformational dynamics (Figure 4E and F). The average C1′–C1′ distance between U_{162} and Ψ_{39} was 11.4 ± 0.5 Å, more than the conventional U•U base pair that averages a C1′–C1′ distance of ~9 Å (47–49). The average hydrogen bond distance observed between the U_{162}N3H and the Ψ_{39}N3H was 3.8 ± 0.6 Å. This distance is consistent with the moderately weak NOE observed between the two imino protons. In comparison, distances between the U_{162}N3 and Ψ_{39}O4 and between the U_{162}O4 and Ψ_{39}N3 were on average 5.4 ± 0.6 Å and 4.4 ± 0.9 Å, respectively. The former distance excluded the formation of a direct hydrogen bond, and possibly the latter as well, but did not exclude the formation of a water-mediated hydrogen bond for either. The aforementioned distance between the U_{162} and Ψ_{39} agreed with the formation of an imino base pair between G_{163} and A_{37}. Indeed, the C1′–C1′ distance between G_{163} and A_{37} was nearly 12.4 ± 0.2 Å in the 10 selected structures. The average distance between Ψ_{39}N1 and the Ψ_{39}OP1 was 4.3 ± 0.6 Å. Once again, this distance did not exclude the possible formation of a hydrogen bond or a water-mediated hydrogen bond between the N1H imino proton of Ψ_{39} and a phosphate atom, such as that observed in crystallographic structures (50).

DISCUSSION

Modifications stabilize the interaction of the tRNA^{Lys3} anticodon domain with the HIV-1 Loop I

Replication of the human immunodeficiency virus (HIV) requires base pairing of the reverse transcriptase primer, human tRNA^{Lys3}, to the viral RNA. Though conclusions from our UV-monitored studies of the thermal denaturations and structures of the duplexes are limited by the fact that the duplex sequence was altered to avoid frayed ends, the circular dichroism spectra and the NMR-derived structure underline the contributions of the 2-thio group and the pseudouridine to the stability of the interaction between the Loop I and the ADF. One of the most dramatic affects on nucleotide conformation and thermodynamic stabilization is the substitution of the carbonyl oxygen at C2 of uridine by a sulfur atom (s^2U). The van der Waals' radius of sulfur being some 20% larger than oxygen contributes to the greatly increased stability of the 3′-*endo*, gauche$^+$ ring pucker and *anti* conformation of s^2U and causes the adjacent nucleotides to favor 3′-*endo* ring pucker and A-form geometry (14,18,51–54). The extremely stable nucleoside conformation of s^2U and that of the adjacent nucleosides stabilizes the canonical base pairings. In the NMR spectra, the downfield chemical shifts for the H5, H6 and H1′ protons relative to that of the unmodified uridine are indicative of an altered ring current that would contribute to the enhanced base pairing, as well as to base stacking. Though the N1 proton of Ψ has long been suspected to be responsible for contributing through direct hydrogen bonding to base pairing, little evidence has been found in investigations of RNAs. However, observation of the N1 proton resonance indicates that it is sufficiently protected from chemical exchange with the solvent water to invoke a stabilizing effect on base pairing through water-mediated hydrogen bonding, enhanced base stacking, or both. As with previous studies of Ψ (55,56), we saw no evidence for hydrogen bonding of the N1 proton of Ψ. Enhanced base stacking is observed when Ψ is in stems or in loops (57–59) and this could contribute to the greater stability of the U-rich anticodon domain interaction with the A-rich HIV Loop I. The modification $ms^2t^6A_{37}$ is missing from the studies presented here. This modification excludes the formation of a canonical U•A base pair across the anticodon loop (60). The $ms^2t^6A_{37}$ and the t^6A_{37} modifications decreased the overall T_m of the ASL (19,60). However, crystallographic studies of the ASL bound to cognate and wobble codons on the ribosome (61) indicate that the $ms^2t^6A_{37}$ modification would act as a hydrophobic platform in stabilizing the base pairing between the tRNA anticodon and the series of As in the Loop I of HIV-1.

The HIV-1 Loop I/tRNA^{Lys3} ADF structure

The dodecamer duplex formed between the VGL and the ADF-s^2U_{34};Ψ_{39} exhibited an A-form RNA conformation with a relatively well-organized internal loop composed of two mismatch base pairs. The NMR data clearly demonstrated that C_{40}, Ψ_{39}, A_{38} and A_{37} are stacked and inserted into the helix. Nevertheless, despite a relatively well-organized internal loop, the VGL/ADF-s^2U_{34};Ψ_{39} duplex exhibited a melting temperature lower than the full complement, VGL/ADF-C. The U_{162} residue was dynamic, as suggested by the weakness of the H5–H6 connectivity in comparison to other uridines or cytidines. The difference between the observed and the back-calculated chemical shifts indicated that the stacking observed in the structures between G_{161} and U_{162} does not reflect the NMR data. An intense NOE connectivity is expected between the uridine imino proton of residues involved in a U•U base pair (49,62). In those base pairs, the average distance between the two imino protons is ~2.5 Å, significantly lower than the average distance observed in our structures between U_{162}N3H and Ψ_{39}N3H (3.8 Å). Although a classical U•U base pairing cannot be formed, the observed distance between acceptor and donor does not exclude the formation of a water-mediated hydrogen bond stabilizing the Ψ•U interaction.

A G•A base pair can adopt more than nine conformations, but two most often observed are the imino and the sheared (43,63) (Figure 4G). The type of G•A base pair depends on the neighboring base sequences and pH (64). In the imino G•A conformation, an NOE connectivity between the GN1H and the AH2 is observed (65), whereas in the sheared G•A the GN1H proton is observed at very low temperatures (66). Conformational dynamics coupled with the presence of the neighboring weak Ψ_{39}•U_{162} pair could explain the broad proton resonance attributed to the G_{163}N1H, and could explain why

the expected connectivity between G_{163}N1H and A_{38}H2 could not be observed. The chemical shift of the U_{164}H5-H6 cross peak and its back-calculated chemical shift were similar and in good agreement with a uridine experiencing a 5′-guanosine (G_{163}) ring current. This is supportive of an imino G•A base pairing. Moreover, it has been suggested that a single G•A base pairing cannot adopt the sheared G•A conformation observed in a tandem G•A pair (67). Formation of an imino base pair between G_{163} and A_{38} would result in an average C1′–C1′ distance of 12.5 Å. Such a base pair could prevent the formation of a Ψ•U base pair with a C1′–C1′ distance of ∼9Å. We observed that the average C1′–C1′ distance for the Ψ_{39} and U_{162} pair was 11.4 Å. These results are consistent with the ^{31}P spectrum of the VGL/ADF-s^2U_{34};Ψ_{39}, and previous studies showing that in contrast to a sheared G•A, an imino G•A base pair resulted in little to no deformation of the backbone (68,69). A model of a 2×2 internal loop composed of 5′-UG-3′/3′-UA-5′ was proposed from NMR and thermal stability studies of a model rRNA system (70). Despite a general pre-organization of the 2X2 internal loop, no hydrogen-bond was observed between the two U's, and the G was external to the helix. Nevertheless, a G•A base pair was not excluded and it was suggested that an imino G•A base pair was possible. We found that the 5′-UG-3′/3′-ΨA-5′ was relatively well-organized with stacking of the C_{40}, Ψ_{39} and A_{38}, and that a water-mediated hydrogen-bond between Ψ_{39} and U_{162} probably existed. Furthermore, G_{163} was inside the helix and probably involved in an imino G•A base pair that would be supported by the NMR data.

The HIV Loop I/tRNA anticodon interaction: a 'kissing' loop complex?

Interactions between the A-rich Loop I and the anticodon domain of $tRNA^{Lys3}$ are essential for priming the initiation of HIV-1 reverse transcription (8,9,71). In fact, the A-rich Loop I region or the complex formed by the A-rich Loop I with the $tRNA^{Lys3}$ anticodon could represent promising targets for developing new therapies. Polyamides target the A-loop sequence (PNA_{AL}), and by binding the A-loop prevent $tRNA^{Lys3}$ from annealing to the viral RNA. Consequently, initiation of reverse transcription *in vitro* is blocked (71). Therefore, an understanding of the interaction between the HIV-1 A rich Loop I and the $tRNA^{Lys3}$ anticodon is imperative.

The annealing of the tRNA primer to the PBS requires unwinding of the acceptor and TΨC stems of the tRNA cloverleaf structure, thus releasing the tRNA sequence complementary to the PBS. Reverse transcriptase (RT) binding facilitates the annealing of the tRNA to the PBS (72). The nucleocapsid protein (NC) is known to have general unwinding and strand renaturation activities and this protein or its precursor Gag has been implicated in the placement of tRNA onto the viral genome (73,74). HIV-1 NC is a short basic protein with two zinc finger domains and functions as a nucleic-acid chaperone (75). Like RT, NC appears to destabilize base pairing in tRNA molecules, without a complete melting of the structure (76). Moreover, it has been suggested that the annealing of tRNA to PBS could open the TΨC stem of the tRNA, allowing residues 48–55 to interact with the PAS portion of HIV-1 (77). Our expanded duplex model was constructed with the purpose of determining the effect of modifications on the binding the HIV Loop I by $tRNA^{Lys3}$. The equilibrium binding constant that we observed for the s^2U and Ψ-modified 12 mer to the Loop I sequence ($K_d = 14.2\,\mu M$) was indicative of a higher affinity, but consistent with that found for the kissing loop interactions of the fully modified anticodon stem and loop ($K_d = 71\,\mu M$) (19). There is evidence that the mechanism of duplex formation may include an initial kissing loop interaction (19,20). In an annealing experiment monitored by NMR, the anticodon stem of a 'recombinant' $tRNA^{Lys3}$ produced in *E. coli* remained intact while the loop was reported to have undergone a conformational exchange without a directly observable loop–loop interaction (78). Chemical probing of infected cells and virions indicated that a kissing loop interaction of the anticodon and Loop I could precede and drive the annealing of the two RNAs (79). Therefore, the unwinding of the $tRNA^{Lys3}$ anticodon stem and that of the Loop I stem may occur subsequent to a kissing loop interaction. Whatever the chronology of the steps in forming a duplex, the interaction between the U-rich anticodon loop of the primer $tRNA^{Lys3}$ and the A-rich Loop I of HIV is stabilized by the tRNA's natural modifications.

HIV subtypes include the A/G forms predominantly from Africa, and the B form predominantly from Europe and North America, representing almost all sequenced isolates. These isolates use $tRNA^{Lys3}$ as the primer for reverse transcription, as do all lentiviruses. In addition, they have in common the A-rich Loop I sequence bound by the $tRNA^{Lys3}$ anticodon with the one difference of the B form having the $G_{166}G_{167}$ substitution for the $A_{166}A_{167}$ in the A/G subtype. The VBL/ADF duplex had a G_{166}•U_{35} wobble pair adjacent to a modified G_{167}•s^2U_{35} base pair, and exhibited a stability little different from that of the ADF-s^2U_{34} bound to the VGL. The Loop I of the viral A/G subtype was protected from chemical probing in the initiation complex for reverse transcription (79). However, the Loop I of the viral B subtype was not protected. Therefore, the interaction of the $tRNA^{Lys3}$ anticodon with the Loop I of the viral B subtype may not be as stable as indicated by equilibrium binding experiments. It is possible that our *in vitro* system might not adequately reflect the *in vivo* assembly of the initiation complex for the B subtype of the virus.

The anticodon loop nucleoside sequences of $tRNA^{Lys1,2}$ are identical to that of $tRNA^{Lys3}$ except for a C_{34} substituting for $mcm^5s^2U_{34}$ and t^6A for $ms^2t^6A_{37}$. Though the C_{34} that would pair with G_{167} of the VBL would be expected to improve the stability of the duplex, C_{34} would miss pair with A_{167} of the VGL and the stability of the duplex would decrease. The $tRNA^{Lys1,2}$ are not used as the primer in the natural viral isolates. The viral primer binding site, the A-rich Loop I sequence and the recruitment of the host cell $tRNA^{Lys3}$ by the HIV

proteins all play important roles in maintaining the virus' dedication to $tRNA^{Lys3}$ as the primer of reverse transcriptase. The detailed biochemistry and structural biology of the recruitment and relaxation of the $tRNA^{Lys3}$ structure and its annealing to the HIV and other lentivirus genomes are yet to be determined.

Accession number

The coordinates of the 10 lowest energy structures, NMR restraints and VGL/ADF-s^2U_{34};Ψ_{39} chemical shifts have been deposited in the Protein Data Bank (accession number: 2K7E) and in the Biological Magnetic Resonance Bank (BMRB entry: 15915). This material is available free of charge via the Internet at http://pubs.acs.org.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Hanna Sierzputowska-Gracz (NCSU NMR Facility) for her advice in NMR spectroscopy, Winnell Newman of the NCSU Nucleic Acid Facility for syntheses, and Eric Westhof for discussions about the geometries and occurrence of G•A base pairs.

FUNDING

National Institutes of Health [R01 GM23037 to P.F.A.]; the National Science Foundation [MCB MCB-0548602 and MCB-9986011 to P.F.A.]; the TRIoH Integrated Project of the European Union [LSHB-CT-2003-503480 to J.V.]; Polish Committee for Scientific Research Grant [to M.A.]. Funding for open access charge: National Science Foundation (MCB-0548602).

Conflict of interest statement. A potential conflict arises in that Dr Andrzej Malkiewicz is a major shareholder in a start-up biotech company (TRANA Discovery, Inc.) interested in the annealing of the HIV primer tRNA to the HIV genome.

REFERENCES

1. Mak,J. and Kleiman,L. (1997) Primer tRNAs for reverse transcription. *J. Virol.*, **71**, 8087–8095.
2. Gabor,J., Cen,S., Javanbakht,H., Niu,M. and Kleiman,L. (2002) Effect of altering the tRNA(Lys)(3) concentration in human immunodeficiency virus type 1 upon its annealing to viral RNA, GagPol incorporation, and viral infectivity. *J. Virol.*, **76**, 9096–9102.
3. Isel,C., Ehresmann,C., Keith,G., Ehresmann,B. and Marquet,R. (1995) Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.*, **247**, 236–250.
4. Aiyar,A., Ge,Z. and Leis,J. (1994) A specific orientation of RNA secondary structures is required for initiation of reverse transcription. *J. Virol.*, **68**, 611–618.
5. Beerens,N., Groot,F. and Berkhout,B. (2001) Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.*, **276**, 31247–31256.
6. Kang,S.M., Wakefield,J.K. and Morrow,C.D. (1996) Mutations in both the U5 region and the primer-binding site influence the selection of the tRNA used for the initiation of HIV-1 reverse transcription. *Virology*, **222**, 401–414.
7. Liang,C., Li,X., Rong,L., Inouye,P., Quan,Y., Kleiman,L. and Wainberg,M.A. (1997) The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J. Virol.*, **71**, 5750–5757.
8. Wakefield,J.K., Kang,S.M. and Morrow,C.D. (1996) Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His). *J. Virol.*, **70**, 966–975.
9. Kang,S.M., Zhang,Z. and Morrow,C.D. (1997) Identification of a sequence within U5 required for human immunodeficiency virus type 1 to stably maintain a primer binding site complementary to tRNA(Met). *J. Virol.*, **71**, 207–217.
10. Li,Y., Zhang,Z., Wakefield,J.K., Kang,S.M. and Morrow,C.D. (1997) Nucleotide substitutions within U5 are critical for efficient reverse transcription of human immunodeficiency virus type 1 with a primer binding site complementary to tRNA(His). *J. Virol.*, **71**, 6315–6322.
11. Zhang,Z., Kang,S.M., LeBlanc,A., Hajduk,S.L. and Morrow,C.D. (1996) Nucleotide sequences within the U5 region of the viral RNA genome are the major determinants for an human immunodeficiency virus type 1 to maintain a primer binding site complementary to tRNA(His). *Virology*, **226**, 306–317.
12. Isel,C., Lanchy,J.M., Le Grice,S.F., Ehresmann,C., Ehresmann,B. and Marquet,R. (1996) Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. *EMBO J.*, **15**, 917–924.
13. Agris,P.F. (2004) Decoding the genome: a modified view. *Nucleic Acids Res.*, **32**, 223–238.
14. Isel,C., Marquet,R., Keith,G., Ehresmann,C. and Ehresmann,B. (1993) Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.*, **268**, 25269–25272.
15. Skripkin,E., Isel,C., Marquet,R., Ehresmann,B. and Ehresmann,C. (1996) Psoralen crosslinking between human immunodeficiency virus type 1 RNA and primer tRNA3(Lys). *Nucleic Acids Res.*, **24**, 509–514.
16. Isel,C., Keith,G., Ehresmann,B., Ehresmann,C. and Marquet,R. (1998) Mutational analysis of the tRNA3Lys/HIV-1 RNA (primer/template) complex. *Nucleic Acids Res.*, **26**, 1198–1204.
17. Tisne,C., Rigourd,M., Marquet,R., Ehresmann,C. and Dardel,F. (2000) NMR and biochemical characterization of recombinant human tRNA(Lys)3 expressed in Escherichia coli: identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription. *RNA*, **6**, 1403–1412.
18. Ashraf,S.S., Sochacka,E., Cain,R., Guenther,R., Malkiewicz,A. and Agris,P.F. (1999) Single atom modification (O−>S) of tRNA confers ribosome binding. *RNA*, **5**, 188–194.
19. Bajji,A.C., Sundaram,M., Myszka,D.G. and Davis,D.R. (2002) An RNA complex of the HIV-1 A-loop and tRNA(Lys,3) is stabilized by nucleoside modifications. *J. Am. Chem. Soc.*, **124**, 14302–14303.
20. Puglisi,E.V. and Puglisi,J.D. (1998) HIV-1 A-rich RNA loop mimics the tRNA anticodon structure. *Nat. Struct. Biol.*, **5**, 1033–1036.
21. Dib-Hajj,F., Khan,R. and Giedroc,D.P. (1993) Retroviral nucleocapsid proteins possess potent nucleic acid strand renaturation activity. *Protein Sci.*, **2**, 231–243.
22. Khan,R. and Giedroc,D.P. (1992) Recombinant human immunodeficiency virus type 1 nucleocapsid (NCp7) protein unwinds tRNA. *J. Biol. Chem.*, **267**, 6689–6695.
23. Tsuchihashi,Z. and Brown,P.O. (1994) DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.*, **68**, 5863–5870.
24. Agris,P.F., Malkiewicz,A., Kraszewski,A., Everett,K., Nawrot,B., Sochacka,E., Jankowska,J. and Guenther,R. (1995) Site-selected introduction of modified purine and pyrimidine ribonucleosides into RNA by automated phosphoramidite chemistry. *Biochimie*, **77**, 125–134.

25. Ariza,X. and Vilarrasa,J. (2000) High-yielding preparation of [3-15N]cytidine, [4-15NH2]cytidine, and [3-15N, 4-15NH2]cytidine. *J. Org. Chem.*, **65**, 2827–2829.
26. Guenther,R.H., Gopal,D.H. and Agris,P.F. (1988) Purification of transfer RNA species by single-step ion-exchange high-performance liquid chromatography. *J. Chromatogr.*, **444**, 79–87.
27. Gehrke,C.W., Kuo,K.C., McCune,R.A., Gerhardt,K.O. and Agris,P.F. (1982) Quantitative enzymatic hydrolysis of tRNAs: reversed-phase high-performance liquid chromatography of tRNA nucleosides. *J. Chromatogr.*, **230**, 297–308.
28. Dao,V., Guenther,R.H. and Agris,P.F. (1992) The role of 5-methylcytidine in the anticodon arm of yeast tRNA(Phe): site-specific Mg2 + binding and coupled conformational transition in DNA analogs. *Biochemistry*, **31**, 11012–11019.
29. Guenther,R.H., Hardin,C.C., Sierzputowska-Gracz,H., Dao,V. and Agris,P.F. (1992) A magnesium-induced conformational transition in the loop of a DNA analog of the yeast tRNA(Phe) anticodon is dependent on RNA-like modifications of the bases of the stem. *Biochemistry*, **31**, 11004–11011.
30. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
31. Johnson,B.A. and Blevins,R.A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
32. Goddard,T.D. and Kneller,D.G. (2007) *Sparky 3*. University of California, San Francisco.
33. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.*, **54**, 905–921.
34. Stein,E.G., Rice,L.M. and Brunger,A.T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J. Magn. Reson.*, **124**, 154–164.
35. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
36. Cromsigt,J.A., Hilbers,C.W. and Wijmenga,S.S. (2001) Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J. Biomol. NMR*, **21**, 11–29.
37. Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.
38. DeLano,W.L. (2006) The PyMOL Molecular Graphics System. *DeLano Scientific LLC, San Carlos, CA, USA*.
39. Fasman,G.D. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules*. Plenum Press, New York.
40. Cromsigt,J., van Buuren,B., Schleucher,J. and Wijmenga,S. (2001) Resonance assignment and structure determination for RNA. *Methods Enzymol.*, **338**, 371–399.
41. Wijmenga,S. and van Buuren,B. (1998) The use of NMR methods for conformational studies of nucleic acids. *Progr. Nuclear Magn. Reson. Spectr.*, **32**, 287–387.
42. Kumar,R.K. and Davis,D.R. (1997) Synthesis and studies on the effect of 2-thiouridine and 4-thiouridine on sugar conformation and RNA duplex stability. *Nucleic Acids Res.*, **25**, 1272–1280.
43. Rudisser,S. and Tinoco,I. Jr (2000) Solution structure of Cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G.A mismatches. *J. Mol. Biol.*, **295**, 1211–1223.
44. Schejter,E., Roy,S., Sanchez,V. and Redfield,A.G. (1982) Nuclear Overhauser effect study of yeast tRNAVal 1: evidence for uridine-pseudouridine base pairing. *Nucleic Acids Res.*, **10**, 8297–8305.
45. Newby,M.I. and Greenbaum,N.L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA*, **7**, 833–845.
46. Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York
47. Baeyens,K.J., De Bondt,H.L. and Holbrook,S.R. (1995) Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. *Nat. Struct. Biol.*, **2**, 56–62.
48. Lietzke,S.E., Barnes,C.L., Berglund,J.A. and Kundrot,C.E. (1996) The structure of an RNA dodecamer shows how tandem U-U base pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure*, **4**, 917–930.
49. SantaLucia,J. Jr, Kierzek,R. and Turner,D.H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry*, **30**, 8242–8251.
50. Arnez,J.G. and Steitz,T.A. (1994) Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry*, **33**, 7560–7567.
51. Agris,P.F., Sierzputowska-Gracz,H., Smith,W.S., Malkiewicz,A., Sochacka,E. and Nawrot,B. (1992) Thiolation of uridine carbon-2 restricts the motional dynamics of the transfer RNA wobble position nucleoside. *J. Am. Chem. Soc.*, **114**, 2652–2656.
52. Sierzputowska-Gracz,H., Sochacka,E., Malkiewicz,A., Kuo,K.C., Gehrke,C.W. and Agris,P.F. (1987) Chemistry and structure of modified uridines in the anticodon, wobble position of transfer RNA are determined by thiolation. *J. Am. Chem. Soc.*, **109**, 7171–7177.
53. Smith,W.S., Sierzputowska-Gracz,H., Sochacka,E., Malkiewicz,A. and Agris,P.F. (1992) Chemistry and structure of modified uridine dinucleosides are determined by thiolation. *J. Am. Chem. Soc.*, **114**, 7989–7997.
54. Testa,S.M., Disney,M.D., Turner,D.H. and Kierzek,R. (1999) Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron. *Biochemistry*, **38**, 16655–16662.
55. Griffey,R.H., Davis,D., Yamaizumi,Z., Nishimura,S., Bax,A., Hawkins,B. and Poulter,C.D. (1985) 15N-labeled Escherichia coli tRNAfMet, tRNAGlu, tRNATyr, and tRNAPhe. Double resonance and two-dimensional NMR of N1-labeled pseudouridine. *J. Biol. Chem.*, **260**, 9734–9741.
56. Tworowska,I. and Nikonowicz,E.P. (2006) Base pairing within the Ψ32, Ψ39-modified anticodon arm of *Escherichia coli* $tRNA^{Phe}$. *J. Am. Chem. Soc.*, **128**, 15570–15571.
57. Davis,D.R. (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.*, **23**, 5020–5026.
58. Desaulniers,J.P., Chang,Y.C., Aduri,R., Abeysirigunawardena,S.C., SantaLucia,J. Jr. and Chow,C.S. (2008) Pseudouridines in rRNA helix 69 play a role in loop stacking interactions. *Org. Biomol. Chem.*, **6**, 3892–3895.
59. Yarian,C.S., Basti,M.M., Cain,R.J., Ansari,G., Guenther,R.H., Sochacka,E., Czerwinska,G., Malkiewicz,A. and Agris,P.F. (1999) Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic Acids Res.*, **27**, 3543–3549.
60. Stuart,J.W., Gdaniec,Z., Guenther,R., Marszalek,M., Sochacka,E., Malkiewicz,A. and Agris,P.F. (2000) Functional anticodon architecture of human tRNALys3 includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t6A. *Biochemistry*, **39**, 13396–13404.
61. Murphy,F.V., Ramakrishnan,V., Malkiewicz,A. and Agris,P.F. (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
62. Nikonowicz,E.P. and Pardi,A. (1992) Three-dimensional heteronuclear NMR studies of RNA. *Nature*, **355**, 184–186.
63. Lee,J.C. and Gutell,R.R. (2004) Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J. Mol. Biol.*, **344**, 1225–1249.
64. Gao,X. and Patel,D.J. (1988) G(syn).A(anti) mismatch formation in DNA dodecamers at acidic pH: pH-dependent conformational transition of G.A mispairs detected by proton NMR. *J. Am. Chem. Soc.*, **110**, 5178–5182.
65. Wu,M., SantaLucia,J. Jr and Turner,D.H. (1997) Solution structure of (rGGCAGGCC)2 by two-dimensional NMR and the iterative relaxation matrix approach. *Biochemistry*, **36**, 4449–4460.
66. Li,Y., Zon,G. and Wilson,W.D. (1991) NMR and molecular modeling evidence for a G.A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA*, **88**, 26–30.
67. Aboul-ela,F., Koh,D., Tinoco,I. Jr and Martin,F.H. (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,T). *Nucleic Acids Res.*, **13**, 4811–4824.

68. Greene,K.L., Jones,R.L., Li,Y., Robinson,H., Wang,A.H., Zon,G. and Wilson,W.D. (1994) Solution structure of a GA mismatch DNA sequence, d(CCATGAATGG)2, determined by 2D NMR and structural refinement methods. *Biochemistry*, **33**, 1053–1062.
69. Gautheret,D., Konings,D. and Gutell,R.R. (1994) A major family of motifs involving G.A mismatches in ribosomal RNA. *J. Mol. Biol.*, **242**, 1–8.
70. Shankar,N., Xia,T., Kennedy,S.D., Krugh,T.R., Mathews,D.H. and Turner,D.H. (2007) NMR reveals the absence of hydrogen bonding in adjacent UU and AG mismatches in an isolated internal loop from ribosomal RNA. *Biochemistry*, **46**, 12665–12678.
71. Kaushik,N., Talele,T.T., Monel,R., Palumbo,P. and Pandey,V.N. (2001) Destabilization of tRNA3(Lys) from the primer-binding site of HIV-1 genome by anti-A loop polyamide nucleotide analog. *Nucleic Acids Res.*, **29**, 5099–5106.
72. Oude Essink,B.B., Das,A.T. and Berkhout,B. (1995) Structural requirements for the binding of tRNA Lys3 to reverse transcriptase of the human immunodeficiency virus type 1. *J. Biol. Chem.*, **270**, 23867–23874.
73. Feng,Y.X., Campbell,S., Harvin,D., Ehresmann,B., Ehresmann,C. and Rein,A. (1999) 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.*, **73**, 4251–4256.
74. Fu,W., Ortiz-Conde,B.A., Gorelick,R.J., Hughes,S.H. and Rein,A. (1997) Placement of tRNA primer on the primer-binding site requires pol gene expression in avian but not murine retroviruses. *J. Virol.*, **71**, 6940–6946.
75. Levin,J.G., Guo,J., Rouzina,I. and Musier-Forsyth,K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Progr. Nucleic Acid Res. Mol. Biol.*, **80**, 217–286.
76. Tisne,C., Roques,B.P. and Dardel,F. (2001) Heteronuclear NMR studies of the interaction of tRNA(Lys)3 with HIV-1 nucleocapsid protein. *J. Mol. Biol.*, **306**, 443–454.
77. Abbink,T.E. and Berkhout,B. (2008) HIV-1 reverse transcription initiation: a potential target for novel antivirals? *Virus Res.*, **134**, 4–18.
78. Tisne,C., Roques,B.P. and Dardel,F. (2004) The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J. Biol. Chem.*, **279**, 3588–3595.
79. Goldschmidt,V., Paillart,J.C., Rigourd,M., Ehresmann,B., Aubertin,A.M., Ehresmann,C. and Marquet,R. (2004) Structural variability of the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.*, **279**, 35923–35931.
80. Rozenski,J., Crain,P.F. and McCloskey,J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196–197.