RNA STRUCTURAL STUDIES: IMPLICATIONS IN HIV-1 REPLICATION AND THE ORIGINS OF LIFE

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

By

Thomas James Gremminger

Dr. Xiao Heng, Dissertation Supervisor

May 2021

The undersigned, appointed by the dean of the Graduate School,

Have examined the dissertation entitled

RNA STRUCTURAL STUDIES: IMPLICATIONS IN HIV-1 REPLICATION AND THE ORIGINS OF LIFE

Presented by Thomas James Gremminger

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

Professor Xiao Heng

Professor Marc C. Johnson

Professor Steven R. Van Doren

Professor Shi-Jie Chen

To my wife Victoria for being my best friend and the rock of our family

Acknowledgments

I would like to thank Dr. Xiao Heng for being an excellent mentor. She has molded me into the scientist I am and has taught me so much. Thank you for pushing me when I needed it and always being there for anything I needed help with. I could not have completed my PhD without you.

I would also like to thank all the members of my lab. Dr. Zhenwei Song, you were always there to help me with any experiment or question. You were instrumental to helping me development and I learned so much from you. Thank you for always being there to talk to me. I would also like to thank Juan Ji. Without her expertise my research would not be as impactful and significant as it could be. Thank you for being my friend and always looking out for me and my family. I would also like to thank Dr. Samantha Brady. Through our time in lab together you have become one of my closest friends and always made being in lab fun.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

(-)ssDNA	Strong stop DNA
^{1}H	Proton
1D	One dimension
^{1}H-^{1}H	Proton- proton
2D	Two dimensional
^{2}H	Deuterium
3'UTR	3' untranslated region
5'UTR	5' untranslated region
7MeG	7-methylguanosine
A	Adenosine
AIDS	Acquired Immunodeficiency Syndrome
C	Cytosine
CA	Capsid
CCR5	C-C chemokine receptor type 5
Cryo-EM	Cryogenic electron microscopy
CXCR-4	C-X-C chemokine receptor type 5
DIS	Dimer initiation signal
DNA	Deoxyribonucleic Acid
dsDNA	Double stranded deoxyribonucleic acid
dsRBD	Double stranded RNA binding domain
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay

ENV	Envelope
ESCRT	Endosomal sorting complexes required for transport
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FTSC	5’-fluorescein thiosemicarbazide
G	Guanosine
Gag	Group specific antigen
gRNA	Genomic RNA
HIV	Human immunodeficiency virus
hLysRS	Human lysyl-tRNA synthetase
IN	Integrase
ITC	Isothermal titration calorimetry
LTR	Long terminal repeat
MA	Matrix
NAD	Nicotinamide adenine dinucleotide
NC	Nucleocapsid
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NPC	Nuclear Pore Complex
Nt	Nucleotide
PBS	Primer Binding Site
PolyA	Polyadenylation signal
PPM	Parts per million
PPT	Polypurine tract

PR	Protease
Psi	Packaging signal
R	Repeat region
RHA	RNA Helicase A
RNA	Ribonucleic acid
RRE	Rev response element
RT	Reverse transcriptase
SAXS	Small Angle X-ray Scattering
SELEX	Systematic evolution of ligands by exponential enrichment
SD	Splice Donor
ssDNA	Single stranded deoxyribonucleic acid
TAR	*Trans*-activation response element
TFIID	Transcription factor II D
TLE	tRNA like element
U3	Unique 3 region
U5	Unique 5 region
Vif	Viral infectivity factor
Vpu	Viral Protein U
Vpr	Viral Protein R

RNA STRUCTURAL STUDIES: IMPLICATIONS IN HIV-1 REPLICATION AND THE ORIGINS OF LIFE

Thomas Gremminger

Dr. Xiao Heng, Dissertation Supervisor

ABSTRACT

Structure and function relationships are important to understand many biological processes. Biological components are not one dimensional and their structures may drive key aspects of their function. This dissertation focuses on understanding key structural elements involved in HIV-1 replication and elucidating the structure of RNA aptamers that could enhance our understanding of early life. In this dissertation we examined a key interaction in HIV-1 reverse transcription initiation, the interaction between the A-rich loop of the HIV-1 5' untranslated region (5'UTR) and the anticodon of $tRNA^{Lys3}$. We showed that this interaction is important for viral infectivity. Using NMR structural techniques, we provided indirect evidence of this interaction and showed that it important for HIV-1 gRNA integrity. Using a combination of in vitro and cell-based assays we showed that this interaction serves to protect the HIV-1 gRNA from degradation by HIV-1 RT before reverse transcription can occur. We also examined another important host factor involved in HIV-1 replication, RNA Helicase A (RHA). RHA has been shown to be important for HIV-1 infectivity and acts as a processivity enhancement factor during reverse transcription. We showed that RHA is recruited to the virion through interactions with the primer binding site (PBS) segment of the HIV-1 5'UTR. We showed that this highly conserved region adopts a three-way junction structure that is recognized and is necessary for its recruitment. We also examined novel RNA aptamers that shift the reduction potential of bound flavin. This is a common strategy for highly reactive flavoenzymes and understanding the molecular mechanisms of flavin recognition may provide insights into early life. We were able to

solve a preliminary solution structure of one of these aptamers and showed that it recognizes flavin using a scaffold of base triples. In summary, this work shows how structural features can influence biological function.

Chapter 1. Background of Human Immunodeficiency Virus and the Origins of Life

1.1 Overview

This dissertation highlights multiple projects in diverse fields. One focus of this dissertation highlights HIV-1 biology with an emphasis on virus: host interactions. Viruses are molecular parasites and require host machinery in order to replicate. Understanding these interactions is important to understanding their lifecycle and may reveal new targets for novel therapies. Chapters 2 and 3 of this dissertation focus on important virus: host interactions in HIV-1 reverse transcription. Additionally, appendix section 1 focuses on characterization of an RNA aptamer that interferes with HIV-1 reverse transcriptase maturation, which could be developed into a novel antiviral treatment for HIV-1. The remaining portion of this dissertation focuses on the RNA world hypothesis and aptamers that could provide novel insights for early life. A key focus of this dissertation is the relationship between structure and function. Understanding RNA structure leads to new insights into molecular mechanisms and thus a deeper understanding of basic biology in HIV-1 and early life.

1.2 HIV-1 Intro

Human immunodeficiency virus 1 (HIV-1) is an enveloped, single stranded, positive sense, enveloped RNA virus of the genus lentivirus and from the family Retroviridae. Left untreated, HIV-1 can develop into acquired immunodeficiency syndrome (AIDS)[1,2], a chronic disease that weakens the immune system over time, and leads to increased susceptibility to other infections and cancers. According to UNAIDS global statistics in 2019, 38.0 million people were living with HIV, 1.7 million people were newly infected, and 690,000 people died from an AIDS related illness. Treatment for HIV-1 is effective and patients receiving antiretroviral therapy,

progression to AIDS can be prevented [3]. However, emerging drug-resistant HIV variants raises the need of mechanistic study of HIV-1 replication and identification of new antiviral targets.

1.3 HIV-1 Lifecycle

HIV-1 virus targets a subset of T-lymphocytes and macrophages that display the CD4 receptor[4]. The HIV-1 viral lifecycle can be divided into two main stages of replication: early replication and late replication. The steps of viral replication can be thought of occurring in an ordered fashion as an organized series of steps [5]. Figure 1-1 shows an illustration of the HIV-1 lifecycle.

The first step of the early stage of HIV-1 replication occurs upon recognition of a CD4 receptor by the surface envelope glycoprotein gp120 [6,7]. In addition to interaction with the CD4 receptor, interaction with an additional coreceptor is also required for viral entry. Two main coreceptors have been discovered, CXCR4 and CCR5. CXCR4 and CCR5 are chemokine receptors of the G protein-coupled receptor superfamily and are expressed mainly in T-lymphocytes and macrophages, respectively [5,8]. Gp120 is composed of five variable regions (V1-V5) and five conserved regions (C1-C5) [7]. Variable region 3, is the major determinant of coreceptor specificity [9]. The ability of HIV-1 to use different coreceptors allows it to infect multiple cell types that express CD4 and leads to variable tropism.

After binding to the receptor and coreceptor, membrane fusion of the viral envelope and plasma membrane of the host cell begins, which the trimeric gp120/gp41 envelope glycoprotein complex facilitates. Binding of the receptor and coreceptor lead to conformational changes in the gp120/gp41 complex. This conformational change exposes the hydrophobic N terminal “fusion peptide” of the gp41 ectodomain that inserts into the host plasma membrane [5,10]. Insertion of the “fusion peptide” leads to fusion of the viral envelope and plasma membrane.

After fusion of the viral envelope and plasma membrane, the viral core is released into the host's cytoplasm. The viral core is composed of capsid (CA) in a conical shape, and contains the viral genome (gRNA) and several viral proteins such as: nucleocapsid (NC), reverse transcriptase (RT), integrase (IN), the accessory proteins, and host factors required for reverse transcription including $tRNA^{Lys3}$ and RNA Helicase A [11,12]. Once the core is released into the cytoplasm uncoating of the core begins but its exact timing is debated. Multiple models of capsid uncoating have been proposed: immediate uncoating, biphasic uncoating, and nuclear pore complex (NPC) uncoating. Immediate uncoating is the process of disassembly of the capsid core soon after viral entry. In the biphasic uncoating model, there is a partial disassembly of the capsid with some capsid remaining associated with the viral reverse transcription complex. The NPC uncoating model proposes that uncoating of the capsid core does not occur until it associates with the nuclear pore complex [11]. Additionally a recent paper found evidence that uncoating does not occur until after import through the nuclear pore complex suggesting a fourth model of capsid uncoating [13].

Uncoating of the capsid core is linked with reverse transcription [14,15], the hallmark step of retroviral replication. Reverse transcription is catalyzed by the viral protein RT. RT is composed of a p51/p66 heterodimer and the p66 subunit contains an active polymerase and RNase H domain [16]. Reverse transcription is a complex multistep process that begins with RT recognition of the primer for reverse transcription, $tRNA^{Lys3}$ annealed onto the gRNA [17]. Illustrated in Figure 1-2, reverse transcription begins at the primer binding site (PBS) of the 5' untranslated region (5'UTR) of the HIV-1 viral genome and proceeds from the 3' end of the primer towards the beginning of the 5' UTR [17]. This produces a DNA/RNA hybrid where the RNA portion of this hybrid is degraded by the RNase H activity of RT as it proceeds towards the 5' end[17]. Digestion of the RNA produces minus strand strong stop DNA [17]. The HIV-1 5'UTR

contains a region termed the R or repeat region that has a homologous sequence in the 3' untranslated region (3'UTR) [17]. After production of the minus strand strong stop DNA the first strand transfer event occurs. In the first strand transfer the minus strand strong stop DNA is transferred to the R region in the 3'UTR [17]. Using the minus strand strong stop DNA as a primer the minus strand is synthesized towards the 5' end of the HIV-1 genome [17]. As synthesis occurs the RNase H domain degrades the viral RNA but a purine rich sequence, termed the polypurine tract (PPT), adjacent to the unique 3 (U3) region of the 3' UTR is resistant to RNase H degradation [17]. The PPT serves as the primer for plus strand synthesis. Plus strand synthesis occurs until the 18 nucleotides of $tRNA^{Lys3}$ complimentary to the PBS are copied [17]. This allows for the second strand transfer of the plus strand to the PBS on the minus strand. From here both minus and plus strands can be completed and the full linear viral DNA is synthesized [17].

As mentioned before, RT is a heterodimer made of a p66 subunit and a p51 subunit. The p66 subunit contains the polymerase active site and the RNaseH active site [17]. The polymerase active site is similar in organization to a human hand and contains fingers, thumb, and a palm [17]. This hand is dynamic and exists in multiple conformations depending of the stage of polymerization [17]. Opening of the hand creates a nucleic acid binding cleft that allows for RT to bind to a substrate [17]. The RNaseH active site is located 17 or 18 base pairs apart from the polymerase active site depending on which nucleic acid substrate is bound [18]. As reverse transcription generates double stranded DNA, removal of the gRNA is required as reverse transcription occurs. As mentioned previously the PPT of the gRNA is resistant to RNaseH degradation. This is thought to occur for multiple reasons. The PPT is a purine rich sequence that contains a narrow minor grove, mispaired and unpaired bases, and in the structure that contains the PPT the active site does not make close contact with the RNA strand [17].

It has been thought that after reverse transcription, nuclear import of the pre-integration complex (PIC) occurs [5]. The PIC is a large nucleoprotein complex composed of the viral DNA, integrase (IN), and host factors including human HMG I(Y) protein [19,20]. Nuclear import proceeds through the nuclear pore complex (NPC). The NPC is a large multiprotein complex that facilitates transport across the nuclear membrane. NPCs are approximately 50 mDa in size and are composed of approximately 30 different proteins [11]. A number of viral proteins, viral elements, and host factors has been implicated in nuclear import including: MA, Vpr, IN, CA, and a three-stranded DNA flap present in the newly-synthesized DNA [21–23].

After nuclear import, integration of the viral DNA begins. Integration is carried out by the viral protein IN. IN is encoded by the pol gene and is produced from the Gag-pol precursor after cleavage by the viral protease. Integrase contains three domains: an N-terminal domain similar to a zinc finger, a central catalytic domain, and a C-terminal domain that binds DNA non-specifically [24]. The first step for integration is processing of the 3’ ends of viral DNA. Integrase removes two nucleotides from each 3’ end of the viral DNA. The 3’ ends then attack a pair of phosphodiester bonds of the host DNA. This pair is separated by five nucleotides and the 3’ ends are joined to the 5’ ends. The 5’ ends of the viral DNA are not joined to the host DNA and the two excess 5’ nucleotides require removal. Additionally, the single stranded region require polymerization and ligation. These additional processing steps occur by host factors [25]. The integrated form of the viral DNA is termed the provirus.

After integration, the late phase of viral replication begins. The provirus serves as the template for transcription of further viral RNA. The long terminal repeat (LTR) of HIV-1 serves as the initiation site of transcription [5,26]. The LTR is made up of three regions: the unique 3 (U3) region, the repeat region (R), and the unique 5 (U5) region [5]. Transcription begins at the transcription start site on the junction of the U3 and R regions. The U3 region contains a TATA

element that is recognized by the host transcription factor IID (TFIID) and is one of that factors that helps recruit DNA polymerase II and initiate transcription [5]. Initial transcription of RNA is abortive, and it is enhanced once the viral Tat protein binds to the transactivation response element (TAR) stem loop of the nascent RNA. Tat has also been shown to recruit other host factors that promote HIV-1 transcription and form a super elongation complex [27].

The HIV-1 genome encodes 10 genes and transcription of the provirus produces a variety of viral RNAs to express these genes [28]. Unspliced RNA encodes the Gag and Gag-pol polyproteins. In order to produce the other HIV-1 proteins more than 50 mRNA variants are produced by the host splicing machinery [29]. Splicing produces three classes of RNA sizes: 9.2kb full length unspliced RNA that encodes Gag and pol genes, 4 kb partially spliced RNAs which encode code the Env, Vif, Vpu, and Vpr proteins, and 1.8 kb completely spliced RNAs which encode Rev, Tat, and Nef[5,30]. In order to export unspliced and partially spliced RNAs the viral protein Rev and a Rev Response Element (RRE) are required. The RRE is a 250 nucleotide highly structured element in the env gene that is recognized by REV [5]. REV facilitates export of unspliced or partially spliced RNAs and binds to the RRE [5]. Additional REV proteins bind to the RRE and REV coats the RRE allow the RNA to be exported using the CRM1-mediated nuclear export pathway into the cytoplasm [30]. REV contains a nuclear localization signal that allows it to be trafficked back into the nucleus [5].

Once the RNAs have been exported from the nucleus they are translated into viral proteins or the full-length RNA is incorporated into the nascent viral particle. To form the viral particle Gag is targeted to the host cell membrane through interactions between the phospholipid phosphatidylinositol-(4,5)-bisphosphate and its matrix (MA) domain [31]. The envelop glycoproteins (ENV), which are made up of gp120 and gp41 domains, are trafficked to the host cell membrane through the secretory pathway [31]. First ENV travels through the rough

endoplasmic reticulum, then the golgi apparatus, and finally in vesicles to the membrane. Once at the membrane ENV is incorporated into the viral particle [31]. Through interactions with the NC domain of Gag the full length viral RNA is incorporated into the viral particle [31]. In order to release the viral particle from the cell membrane, the p6 domain of Gag interacts with the endosomal sorting complex required for transport (ESCRT), a host factor that is required for budding of the virus from the membrane [31]. This forms the immature virion and the viral protease cleaves the Gag precursor into the mature viral proteins [31]. Viral protease is formed from the Gag-Pol polyprotein precursor that is formed from a frameshift mutation during Gag translation [31,32]. After maturation of the virion, the virion can infect a new cell and begin its lifecycle anew.

1.4 HIV-1 5' UTR and Initiation of Reverse Transcription

The 5' UTR of the HIV genome is highly structured and highly conserved region that is involved in many steps of the viral lifecycle, including initiation of reverse transcription, genome dimerization, translation initiation, genome packaging, and RNA splicing [33]. The 5' UTR contains multiple structural elements important for viral function including the Tat Responsive Element (TAR), polyA, PBS, dimer initiation signal (DIS), splice donor (SD), and Psi packaging element [34]. The TAR and polyA form stem loops and are part of the R (repeat) region that is repeated in the 3' UTR [34]. This repeat allows for the first strand transfer step during reverse transcription [17]. The U5 region encompasses a portion of polyA and continues until the beginning of the PBS. The PBS region contains a stem loop structure called the tRNA like element (TLE) upstream of the PBS. This region is important for placement of $tRNA^{Lys3}$ onto the PBS in order to prime reverse transcription [35,36]. The PBS is followed by four stem loops, which contain the DIS, SD, Psi, and AUG start codon respectively [34].

The 5’ UTR has been shown to adopt two major conformations [37,38]. Dimeric genomic RNA has been shown to be packaged into virions [39]. Under in vitro conditions the number of guanosine nucleotides at the 5’ end of the 5’ UTR has been shown to affect the folding of downstream nucleotides and modulate the monomer: dimer equilibrium [40,41]. Transcripts that begin with single guanosine capped by a 5’-5’ triphosphate-linked 7-methylguanosine (7MeG) have been found to be monomeric and transcripts that begin with two or three guanosines capped by a 7MeG primarily exist as monomers under in vitro conditions [40,41]. In the dimeric form of the 5’UTR sites that promote dimerization and packaging, including NC binding sites and dimer promoting residues are accessible, whereas, in the monomeric form, these sites are sequestered and packaging is reduced [41]. Additionally, in the dimeric form of the 5’UTR the DIS can base pair with a second DIS and form a “kissing” intermolecular interaction [42]. This kissing interaction then forms into an extended duplex between the two RNAs [42]. In the monomeric form of the 5’ UTR the DIS is sequestered and base pairs with U5 whereas, in the dimeric form the AUG stem loop base pairs with U5 making the DIS accessible [43]. The secondary structure of monomeric and dimeric HIV-1 5’UTR is shown in Figure 1-3.

The PBS of the 5’ UTR is the site were the primer that initiates reverse transcription anneals. The PBS is composed of 18 nucleotides that are complementary to the 18 3’-terminal nucleotides of $tRNA^{Lys3}$ [44,45]. $tRNA^{Lys3}$ is packaged into virions in addition to the other tRNA isoacceptor, $tRNA^{Lys1,2}$ [16]. However, as the PBS is only complementary to $tRNA^{Lys3}$, $tRNA^{Lys1,2}$ does not prime reverse transcription [44]. Packaging of $tRNA^{Lys3}$ is accomplished through interactions between human lysyl-tRNA synthetase (hLysRS) and the viral protein Gag [35]. Studies using the HIV-1 NL4-3 sequence have shown that hLysRS recognizes the TLE near the PBS, binds to the TLE, thus releasing $tRNA^{Lys3}$, and allowing it to be annealed to the PBS [35].

Studies using the HIV-1 MAL sequence have shown that hLysRS doesn't recognize a specific structure in the 5' UTR but, recognizes the global shape of the PBS region [36].

As HIV-1 has spread worldwide its genetic variability has increased [46]. HIV-1 has been divided into groups M, O, and N [46]. Group M is divided into nine subtypes or clades: A-D, F-H, J, and K [46]. Subtype C accounts for 50% of infections but subtypes A and B are the most studied in the literature [34]. The PBS region of the 5' UTR is also subject to genetic variability. In subtype A sequences and subtype A recombinants immediately downstream of the PBS a 23-nucleotide insertion is present that is a partial duplication of the PBS[34]. This insertion is not present in subtype B sequences [34]. As a result of this insertion subtype A sequences are predicted to adopt an altered structure in the PBS region of the 5' UTR as compared to subtype B sequences [34]. In subtype A sequences the PBS is involved in a stem-loop structure in the whereas it is largely unpaired in subtype B sequences [34]. This is shown in Figure 1-4. In the unpublished results section of Chapter 3 the conservation of the PBS region is analyzed. We examined all sequences in the Los Alamos National Labs HIV-1 sequence database (https://www.hiv.lanl.gov/content/index) and divided the sequences into two groups for analysis: sequences that contained an insertion and those that did not. As these sequences were highly conserved, we hypothesized that the PBS region may fold into one of two primary secondary structures based on the presence or absence of the insertion.

Annealing of $tRNA^{Lys3}$ to the PBS does not occur spontaneously and requires the viral protein NC, which acts a chaperone and facilitates the rearrangement and annealing of $tRNA^{Lys3}$ to the PBS [34]. NC is a 55-residue protein and is very basic (pI = 9.9) [34]. It contains two zinc-knuckle domains with the consensus sequence $CX_2CX_4HX_4C$ [47]. NC has a role in almost all steps of the viral lifecycle including RNA packaging, virion assembly, reverse transcription, and integration [34]. NC contains 15 lysine or arginine residues that are positively charged at

physiological pH and may interact through nonspecific electrostatic interactions with viral RNA [34]. In the case of reverse transcription, annealing occurs by first a small number of base pairs (3-4bp) followed a "zipping up" of the rest of the 18 base pairs through interactions with NC [34,48]. The zinc knuckle domains of NC can then recognize the D loop of $tRNA^{Lys3}$ and stabilize the annealed complex [34,49].

After annealing, in order for reverse transcription to begin, RT must recognize the annealed complex to create the reverse transcription initiation complex (RTIC). The structure of this complex is complex and no high resolution structures were available until recently [50]. This Cryo-EM structure is composed of $tRNA^{Lys3}$ annealed to an RNA construct of the PBS region of the 5' UTR that is bound by RT. In this structure, RT is crosslinked to the $tRNA^{Lys3}$: viral RNA complex and the $tRNA^{Lys3}$ primer is extended by one dideoxynucleotide. This complex shows an extended $tRNA^{Lys3}$ helix and two helixes from the viral RNA extending from the nucleic acid binding cleft of RT. No additional interactions between the $tRNA^{Lys3}$ and the viral RNA were detected, besides the 18 3' nucleotides of the $tRNA^{Lys3}$ and the PBS of the viral RNA. The cryo-EM structure shows a transient stage of reverse transcription initiation where additional interactions may be disrupted.

Additional interactions between $tRNA^{Lys3}$ and the 5' UTR, besides the 18 3' nucleotides of $tRNA^{Lys3}$ and PBS, have been shown to be necessary for efficient reverse transcription and $tRNA^{Lys3}$ utilization [51–55]. These interactions are believed to stabilize the RTIC structure and enable efficient initiation of reverse transcription. Some groups have attempted to force HIV-1 to utilize other tRNAs and these experiments have been challenging. Mutations that allow utilization of other tRNAs eventually revert back to wild type sequence and utilize $tRNA^{Lys3}$ [51,52,56–58]. This suggested that other interactions between $tRNA^{Lys3}$ and the vRNA are important for $tRNA^{Lys3}$ utilization. Mutations near the A-rich loop, a highly conserved stretch of

adenosines near the PBS, were necessary to prevent reversion to utilizing $tRNA^{Lys3}$. These mutations were designed to be complementary to the anticodon residues of alternative tRNAs.

Detecting and confirming extended interactions between $tRNA^{Lys3}$ has been challenging as various groups have proposed conflicting models about the extended interactions between $tRNA^{Lys3}$ and the gRNA. In vitro and in situ probing experiments utilizing subtype A HIV-1 sequences have shown that residues in the A-rich loop formed base pairs with residues in the anticodon and anticodon stem of $tRNA^{Lys3}$ but NMR studies were unable to detect these interactions [59–63]. Additionally, cell-free reverse transcription assays showed that these interactions were important for efficient transition from initiation to elongation during reverse transcription [54,55]. tRNA selection assays utilizing subtype B sequences reported that mutations in the A-rich loop were required to sustain usage of tRNAs other than $tRNA^{Lys3}$ [51,56,64–66]. However, in vitro studies showed that an alternative region upstream of the PBS, the primer activation signal (PAS), form base pairs with the TΨC residues in $tRNA^{Lys3}$. Reverse transcription assays utilizing subtype B sequences reported that this interaction is necessary for efficient initiation of reverse transcription [67–69]. Structural studies, including chemical probing and cryo-electron microscopy, were unable to detect this interaction [50,70]. Even though these interactions may be highly important for reverse transcription, the molecular details are not fully understood, and the results are controversial.

As shown in Chapter 2, our study explores the importance of the gRNA A-rich loop: $tRNA^{Lys3}$ anticodon interaction and how this interaction is necessary for the gRNA's stability. Our study uses a combined approach of structural, biophysical, and cell-based experiments to show that the extended intermolecular interactions are important for the stability of the vRNA. We show that this mechanism is conserved in two subtypes of HIV-1 and these extended

interactions may not have been able to be observed in previous studies due to experimental annealing conditions of the $tRNA^{Lys3}$: gRNA complex.

DHX9/RNA helicase A (RHA) is a protein host factor that has been shown to be important for HIV-1 infectivity [71–73]. RHA unwinds the RNA duplex in the 3' to 5' direction in a (d)NTP dependent manner and is part of the DExH box superfamily of proteins [74]. RHA is a 1270 amino acid, 140 kDa protein composed of two double-stranded RNA binding domains (dsRBD1 and dsRBD2) at the N-terminus, followed by two RecA-like domains in the core helicase region, and an RG-rich domain on the C-terminus [75]. RHA is packaged into the HIV-1 virion and through interactions with the PBS region of the 5' UTR. Mutations that disrupted the structure of the PBS region of the 5' UTR reduced binding affinity of RHA to the 5' UTR and lowered infectivity [71]. RHA has been shown to enhance the processivity of reverse transcription by unwinding secondary and tertiary structures in the RNA [76]. However, the structural basis for how RHA is recruited to the RTIC is not clear.

Chapter 3 explores the determination of high-resolution structural data for the PBS region of HIV-1 and how RHA interacts with the PBS region. This study elucidates new structural information about the PBS region and shows that it is not largely unstructured as previously thought. It also reveals molecular details of the recruitment of RHA, a key host factor involved in the hallmark step of reverse transcription.

1.5 Treatment of HIV-1

Currently there is no cure for HIV. However, there are effective treatments that can prevent progression to AIDS and lower viral loads to undetectable levels. HIV treatment uses a variety of drugs that target different steps of the viral lifecycle. Because of the high mutation rate of HIV multiple types of drugs are used in combination therapy called HAART or highly

active antiretroviral therapy [77]. HAART is customized to the patient's viral load, symptoms, infected strain and other factors.

The first type of drugs discovered against HIV were NRTIs or nucleotide/nucleoside reverse transcription inhibitors [78]. NRTIs work by causing chain termination of the nascent DNA chain. NRTIs are chemically similar to nucleotides but cannot be incorporated as they lack a 3' hydroxyl that is necessary for formation of a 3'- 5' phosphodiester bond. Termination of the nascent DNA chain can occur during both the RNA dependent and DNA dependent DNA synthesis. Currently there are eight approved NRTIs.

Non-nucleoside reverse transcription inhibitors (NNRTIs) are another class of therapy for HIV-1. NNRTIs inhibit reverse transcription by binding to a hydrophobic pocket in RT near the active site. This induces a conformational change in RT that reduces activity. Residues from the p66 and p51 subunit form this pocket and are only formed in the presence of NNRTIs [77].

Integrase inhibitors are another class of treatment for HIV-1. Integrase inhibitors inhibit integration of viral DNA into the host genome by interacting with the integrase: viral DNA complex. Integrase inhibitors bind to the integrase: viral DNA complex and sequester the active site magnesium cofactors. This inhibits the strand transfer reaction of integrase [77].

Another target for HIV-1 therapies is the viral protease. Protease catalyzes the hydrolysis of the Gag and Gag-Pol polyproteins into the structural proteins and viral enzymes during maturation of the virion. Protease inhibitors bind to the active site of protease and act as transition state analogs and mimic the transition state during hydrolysis of the peptide backbone. Many of the protease inhibitors contain a central hydroxyl group that is present of the transition state. Another mechanism of protease inhibition is through prevention of protease dimerization. To be catalytically active, protease monomers must dimerize. Compounds such as darunavir and tipranavir prevent this dimerization [79].

HIV-1 therapies also target viral entry at multiple stages of the viral entry process. One stage that therapies target is after attachment of gp120 and CD4. Post-attachment inhibitors, such as the monoclonal antibody ibalizumab, prevent the gp120-CD4 complex from associating with the coreceptor CCR5 or CXCR4 [80]. Another class of entry inhibitors target the coreceptor CCR5. These small molecule based inhibitors bind allosterically to the transmembrane domain of CCR5 and stabilize a conformation of CCR5 that prevents binding by gp120 and native agonists of CCR5 [77]. Currently there is no approved CXCR4 antagonist [81]. One fusion inhibitor preventing fusion of the viral and cell membrane has also been approved. Fusion inhibitors prevent formation of the six helical bundle of gp41 that is necessary for fusion. Fusion inhibitors are peptide derived and bind to the domains of gp41 that form the six helical bundle [82].

HIV-1 treatments target many steps in the viral lifecycle. Development of novel treatments is necessary as HIV-1 has a high mutation rate and drug resistance is a common problem in patients. HIV-1 reverse transcription is error-prone and it is estimated that there are 1-10 mutations per 1000-10,000 nucleotides polymerize [77]. Due to the fact that he HIV-1 genome is approximately 10,000 nucleotides long 1-10 mutations will be produced during reverse transcription [77]. This leads to great genetic diversity and thus the need for multiple treatments and development of novel therapies. Additionally, understanding the basic biology of HIV-1 could lead to identification of new targets for the development of treatments.

1.6 Aptamers

Aptamers are short oligonucleotides that bind to a target with high affinity and specificity through their tertiary structure [83]. Aptamers have potential in a wide range of applications including therapeutics, diagnostics, and biosensors and have been compared to antibodies in their ability to bind targets [84]. Aptamers are generated through techniques such as systematic evolution of ligands by exponential enrichment (SELEX). In a SELEX selection,

shown in Figure 1-5, a target is incubated with a random library of single stranded oligonucleotides. These oligonucleotides are generally 40 to 100 nucleotides, which contain a middle random region and fixed sequences on both ends, which allow for sequential analysis. After incubation, unbound aptamers are removed, and bound aptamers are eluted and amplified. Aptamers that have bound the target are then used for further rounds of selection. After a number of rounds of selection, aptamers with high affinity and specificity will be enriched and can be sequenced. The process for RNA aptamer selection is similar but requires transcription of RNA oligonucleotides and reverse transcription before PCR amplification of bound aptamers. SELEX has many variations that have been developed to increase the efficiency and reliability of aptamer selection, including immunoprecipitation-coupled SELEX (IP-SELEX), cell-SELEX, capture-SELEX, capillary electrophoresis-SELEX (CE-SELEX), atomic force microscopy-SELEX (AFM-SELEX), and artificially expanded genetic information system-SELEX (AEGIS-SELEX)[83].

Aptamers have also been developed as potential therapeutics that target proteins involved in HIV-1 replication [85]. Aptamers have been developed against gp120, the envelope glycoprotein that binds CD4, in order to prevent viral attachment and entry [85]. Aptamers have also been developed against the viral protein Gag, which were shown to lower extracellular capsid levels and Gag mRNA. Additionally, other aptamers selected against Gag interfered with genome packaging by binding the nucleocapsid region of Gag. Aptamers against other viral proteins, such as nucleocapsid, Rev, and RT have also been developed.

Appendix 1 examines broad spectrum aptamers that bind to RT. This study was a collaboration between the Heng Lab and the Burke Lab and characterized the interface between the aptamer and RT and also showed that the aptamer modulated the efficiency of proteolytic cleavage of precursor RT by viral protease.

Through our collaboration with the Burke lab in characterizing aptamers that bind to HIV-1 RT we were able to join another study involving RNA aptamers. This study was a divergence from my previous expertise and focused on the origins of life and the RNA world hypothesis. This new collaboration is a joint project with the Heng Lab, Burke Lab, and Baum Lab and examines aptamers that bind flavin moieties.

1.7 RNA World Hypothesis

The RNA world hypothesis proposes that not only can RNA store heritable genetic information but also could catalyze important biological chemical reactions and that RNA or a similar polymer was the main macromolecule utilized for catalysis and storing genetic information before other macromolecules evolved [86,87]. Experiments that explore the RNA world hypothesis have attempted to investigate the formation and synthesis of nucleotides and nucleotides strands using prebiotic conditions and investigated conserved mechanisms of catalysis and replication in biology to understand the last universal common ancestor before the current domains of life were established. RNA can catalyze numerous classes of chemical reactions including oxidation/reduction reactions, carbon-carbon bond formation, acyl and phosphoryl transfers, and hydrolysis reactions [88–93]. However, while it has been shown that RNA can catalyze these reactions it unclear if RNA is able to catalyze the reactions required to support the network of pathways and metabolites present in cells.

1.8 Flavin Binding Aptamers

Flavoenzymes catalyze a diverse range of biochemical reactions including redox transfer of one or two electrons, light emission, the activation of oxygen for oxidation and hydroxylation reactions, and the dehydrogenation of metabolites, which makes the flavin moiety an important biological cofactor [94]. One mechanism that drives redox reactions involving flavin cofactors is utilizing binding energy to increase the reduction potential of the flavin cofactor and thus

increase reactivity. In an RNA world, catalytic RNAs that could increase the reduction potential of flavin would have been important for early life. Aptamers that can bind FAD and NAD+ have been selected [95–97]. However, their ability to shift reduction potential of bound flavin was not evaluated and these studies focused on the molecular details of the flavin-binding mode.

Chapter 4 examines novel aptamers that bind flavin and shift the reduction potential of the bound flavin. We have been able to characterize the binding affinity of theses aptamers to flavin moieties. We have also structurally characterized the binding of these aptamers to flavin moieties using nuclear magnetic resonance (NMR). From these experiments we have a built a preliminary structure of one of these aptamers.

1.9 NMR of RNA

NMR is a tool frequently used in this dissertation to study the structure of RNA. NMR is an excellent tool to study the structure of RNA in solution and their interactions with ligands such as proteins and small molecules [98].

NMR experiments observe signals on NMR active nuclei with half-spin, such as ^{1}H, ^{13}C, or ^{15}N [98]. As ^{1}H is naturally abundant to prepare an RNA sample for a ^{1}H NMR experiment a standard *in vitro* transcription reaction can be used to prepare an amount necessary for NMR experiments. However, the natural abundance of ^{13}C and ^{15}N is low and require using isotopically labeled NTPs in an *in vitro* transcription reaction to prepare an isotope labeled RNA for NMR experiments [98].

A simple NMR experiment that we frequently use to study an RNA is a one-dimensional (1D) ^{1}H NMR experiment. It shows chemical shift dispersion of exchangeable or non-exchangeable ^{1}H (protons), depending on the solvent conditions. Exchangeable protons will exchange with the bulk solvent and are only observable when the NMR sample is dissolved in H_2O. These protons would not be protected from exchange if the sample were dissolved in a solvent such as D_2O. One

such experiment that detects exchangeable protons is a 1D ^{1}H NMR experiment of the imino-proton region. In this experiment, the RNA sample is dissolved in at least 80% H_2O to protect the imino-protons from exchange and a spectrum is collected. For this experiment resonances from 10-15 parts per million (ppm) yield information about base pairing in the RNA helix [98]. In this region the imino-protons from guanosine and uracil give signals when they are base paired. Specifically, Watson-Crick base pairs tend to have chemical shifts of 12-15 ppm, while chemical shifts of non-canonical base pairs are often upfield shifted [98]. Another 1D ^{1}H experiment used in this dissertation examines the non-exchangeable protons of the aromatic region (approximately 6.5-8.0 ppm). One way we used this experiment in this dissertation is to probe structural changes upon annealing of $tRNA^{Lys3}$ to the HIV-1 gRNA construct PBS Segment in Chapter 2. We examined the PBS segment before and after annealing of $tRNA^{Lys3}$. Although 1D NMR spectra do not have the resolution for full assignment of the RNA, we were able to use characteristic upfield shifted H2 protons to detect structural changes upon $tRNA^{Lys3}$ annealing. This is one example of a way to use a simple 1D ^{1}H experiment to probe structural changes. Similar experiments can be done with RNA in the presence of small molecules or proteins.

NMR experiments of RNA often utilize a second dimension, which not only increases resolution, but expands information that can be inferred about the structure of the RNA. Both the imino-proton and aromatic regions can be resolved using a 2D experiment. One common experiment is a 2D ^{1}H-^{1}H NOESY experiment. In a NOESY experiment if two protons are within approximately 5 angstroms of one other this will give rise a peak in the spectra [98]. This is due to the Nuclear Overhauser Effect which correlates protons that are close together in space [98]. In a standard A-form helix NOE connectivities exist between sequential residues in RNA strand [99]. These NOE connectivities yield peaks in the spectra that can be assigned via a “NOESY walk.” Both imino-protons and aromatic protons yield NOE connectivities that can be assigned using a NOESY

walk. Figure 1-6 shows NOE connectivities of aromatic and ribose protons. These connectivities usually suggest that an A-form helix and therefore give information on which residues are involved in a A-form helix [99]. Residues that are not in an A-form helix may not have these connectivities such as unpaired residues or residues involved in loops and bulges. Assignments of RNA residues are helpful for determination of secondary structure and as restraints in three-dimensional structural calculation.

One common problem of RNA NMR experiments is their limited resolution. As the atoms in an RNA helix are in a similar chemical environment, larger (> 50 nucleotides) RNAs often yield spectra that contains overlap and is limited in resolution and information [98]. Additionally, large RNAs tumble slowly in solution which leads to broadening and loss of intensity of signals [100]. One method to overcome this is using a "divide and conquer" approach [101]. In this approach a smaller construct (or constructs) containing a portion of a larger RNA is analyzed using 2D NMR spectroscopy. Spectra is collected for the smaller construct and the large RNA. Whereas the large RNA may have a broad and highly overlapped spectra, the spectra of the small RNA construct will be of higher resolution and assignable. Using this approach care must be taken to ensure the small RNA construct folds into the same structure as the large RNA. The smaller RNA construct should have the same NOE connectivities as the larger RNA construct. This approach is used in both Chapter 2 and 3 help assign large RNAs.

Another approach to simplify spectra and enable assignment is to utilize nucleotide specific ^{2}H-labeled RNA samples [101]. In this approach RNA is prepared using in vitro transcription with nucleotides that contain site-specific deuteration (H6 or H8, H1', D2', D3', D4', D5'/D5" and D5) [101]. For example, in chapter 3 we utilized an $A^{2R}GU^{R}$-PBS-segment, which means the H2 and ribose of adenosine, ribose of uracil and guanosine are protonated, and cytidine is fully deuterated. This approach is also utilized in chapter 4. This approach simplifies the

spectra and allows unambiguous assignment as no signals are present from the atoms that are deuterated.

1D ^{1}H and 2D ^{1}H-^{1}H NMR experiments are not the only experiments useful for obtaining RNA structural information. One experiment utilized in chapter 4 is an HNN-COSY experiment. This a heteronuclear NMR experiments that utilizes ^{15}N labeled RNA and correlates ^{15}N atoms involved in base pairs and imino-protons [98]. Because of different chemical shifts of the ^{15}N atoms, canonical and non-canonical base pairs can be distinguished [98]. This is useful for determining base paring and secondary structure. Additional heteronuclear experiments can be utilized to obtain base pairing information include ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC [98].

In summary, we have used standard and modern NMR strategies to study RNAs sizing ranging from 38 nucleotides to 196 nucleotides, to investigate $tRNA^{Lys3}$ annealing to HIV-1 gRNA (Chapter 2), to investigate ligand binding of an RNA aptamer to an important biological cofactor, FMN (Chapter 4), and to investigate protein interactions between HIV-1 gRNA and an important host factor, RHA. It is one of the major biophysical tools I used in this dissertation to study RNA structure.

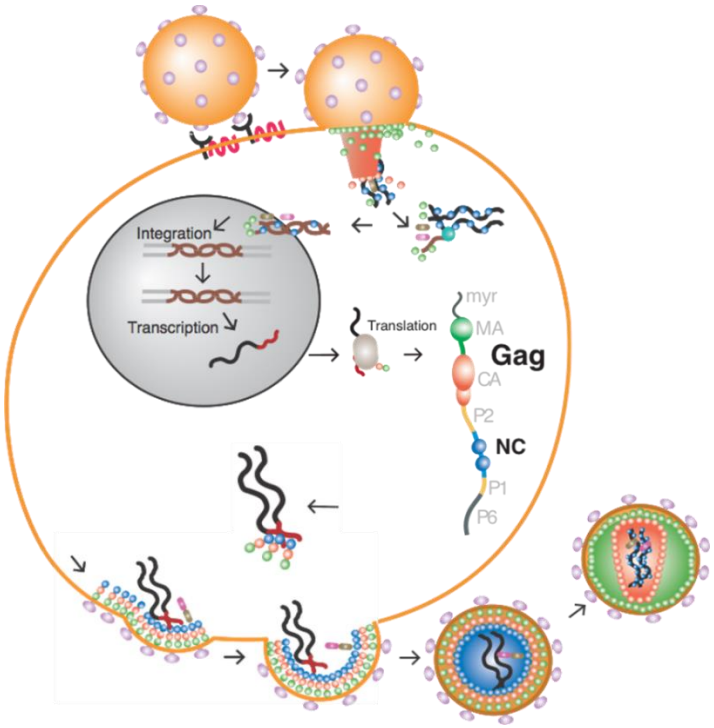


Figure 1-1. Cartoon of the HIV-1 lifecycle. The HIV-1 virion (orange ball) is coated in surface envelope glycoprotein (purple balls). Their interactions with host cell receptor (black claw) and coreceptor (red squiggle) mediate virus entry. Fusion of the viral and cell membrane begins and the viral capsid core (red cone) is released into the cytoplasm. The capsid core releases the viral genome (black lines) and reverse transcriptase (turquoise ball) transcribes the RNA genome into proviral DNA (brown helix). Note: The timing of capsid uncoating is debated and only the cytoplasmic uncoating model is shown in this graphic for simplicity. The proviral DNA is trafficked to the nucleus and integrated into the host genome (fusion of brown proviral DNA and grey lines). The integrated viral genome is transcribed into mRNA using the host machinery and is either unspliced (black line), partially spliced (not shown), or multiply spliced (not shown). The full length mRNA is exported from the nucleus and can be either translated into viral proteins or targeted to the membrane and packed into new virions. Once the virion assembles it buds from the cell and forms an immature virion. Viral protease cleaves the Gag and Gag-pol into the mature viral proteins and forms the mature infectious virion.

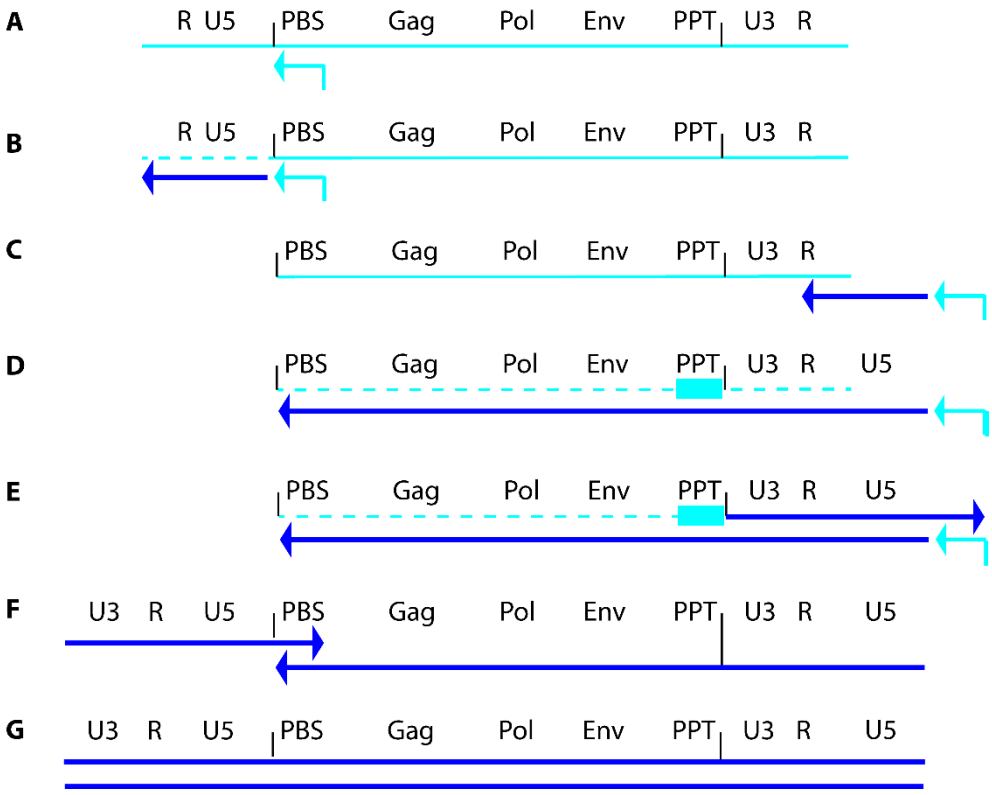


Figure 1-2. Schematic of HIV-1 reverse transcription. **A.** $tRNA^{Lys3}$ is annealed to the gRNA by NC. **B.** RT initiates reverse transcription from the 3' end of $tRNA^{Lys3}$ and proceeds towards the 5' end of the gRNA producing minus strand strong stop DNA. Dashed lines indicate degraded RNA by the RNaseH domain of RT. **C.** The first strand transfer event occurs. The nascent DNA chain is able to base pair with the R region of the 3' LTR. **D.** RT proceeds from the 3' end of the nascent DNA chain towards the 5' end of the gRNA. The PPT is resistant to RNaseH degradation. **E.** The PPT serves as the primer for plus strand synthesis and proceeds until the 18 3' nucleotides of $tRNA^{Lys3}$ are copied. **F.** A second strand transfer of the plus strand occurs. The plus strand base pairs with the PBS. **G.** The plus and minus strands are then extended to complete the proviral DNA.

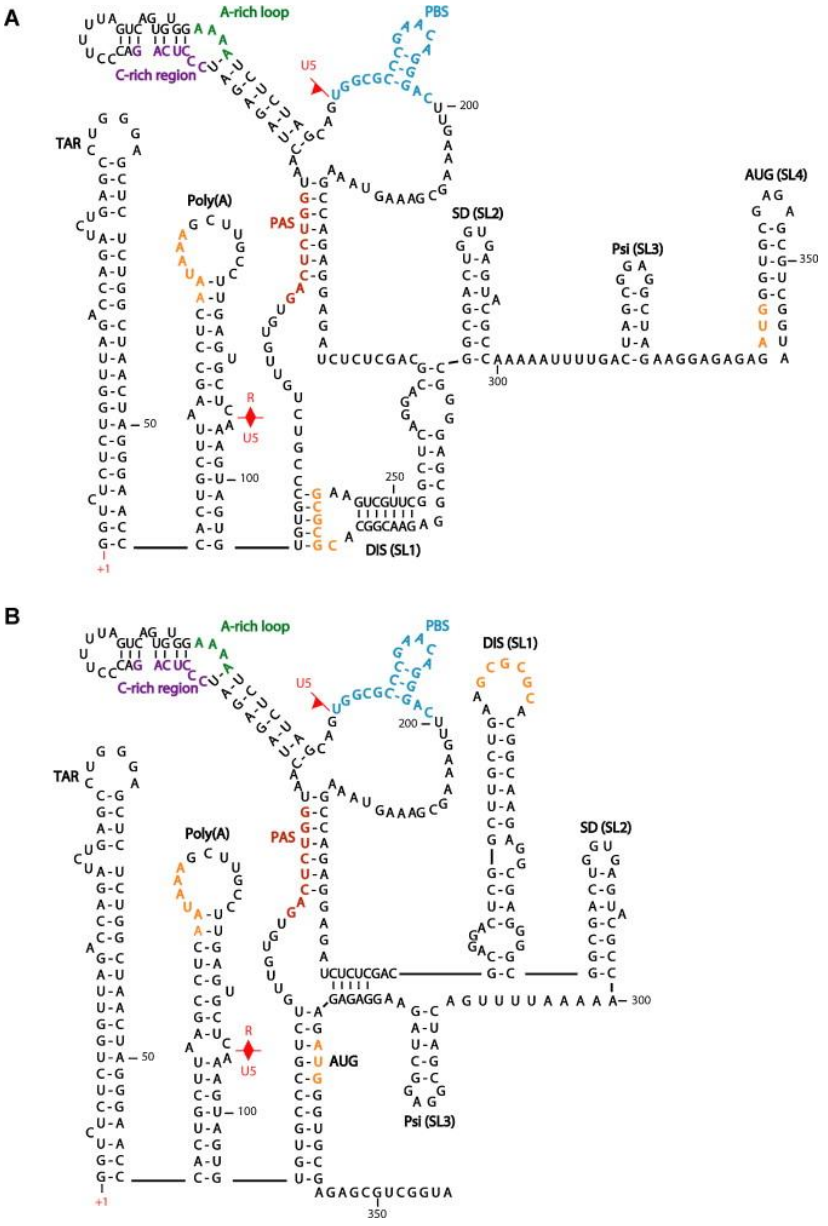


Figure 1-3. Secondary structure of the HIV-1 NL4-3 5' UTR. **A.** Secondary structure of the form of the 5' UTR in monomeric gRNA. **B.** Secondary structure of the form of the 5' UTR that is involved in gRNA dimerization. The PBS is highlighted in blue, the A-rich loop in green, and the PAS in red. The boundary of the R and U5 region is indicated in red arrows. Adapted from Sleiman, D. *et al.* Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. Adapted from *Virus Res.* **169**, 324–339 (2012).

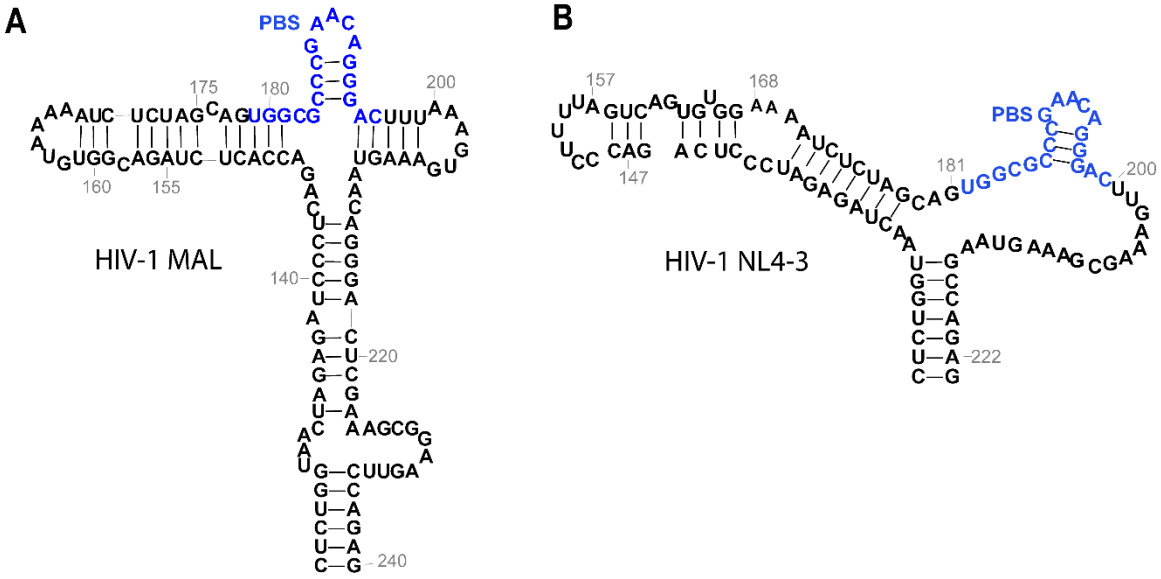


Figure 1-4. Secondary structure of PBS region of the 5' UTR of HIV-1 gRNA differs in subtype A and subtype B sequences. **A.** Secondary structure of subtype A HIV-1 MAL sequence. Secondary structure based on NMR assignment [41]. PBS is highlighted in blue. **B.** Secondary structure of subtype B HIV-1 NL4-3 sequence. PBS is highlighted in blue.

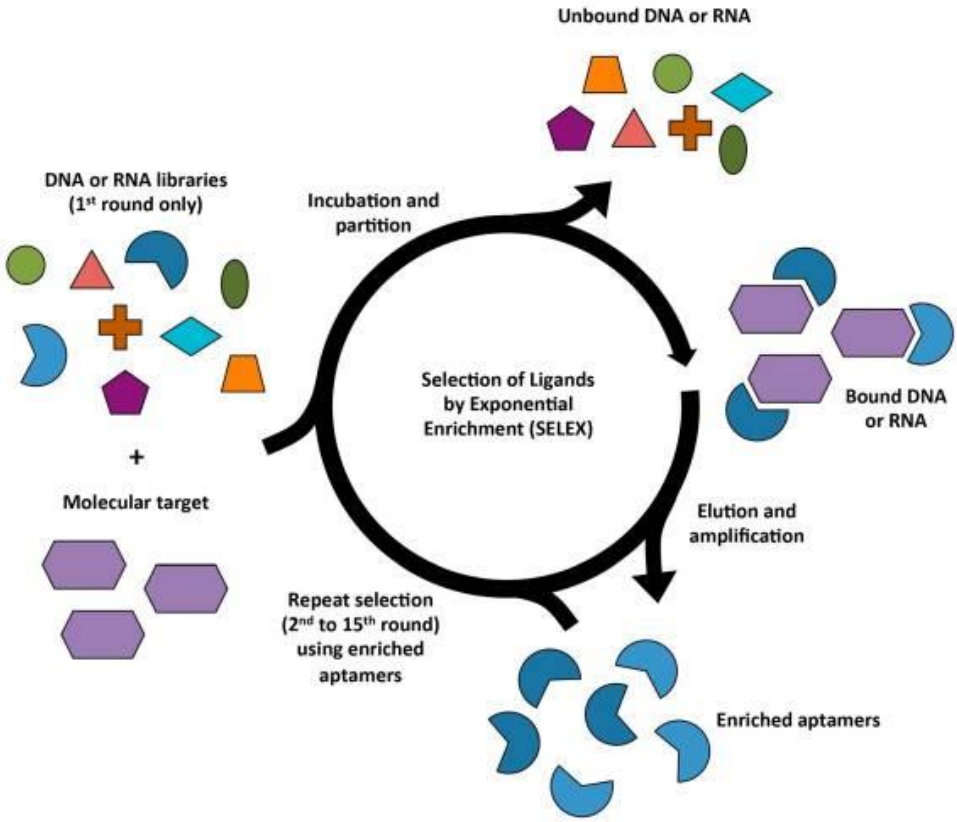


Figure 1-5. Schematic diagram of SELEX (Systematic Evolution of Ligands by EXponential Enrichment) for aptamer selection. An initial random DNA or RNA library is used and incubated with a molecular target. The unbound DNA or RNA are partitioned from the bound DNA or RNA and the bound DNA or RNA is eluted and amplified by PCR. This enriched pool is then used for further rounds of aptamer selection. Adapted from Zhang, Y *et al*. Recent advances in aptamer discovery and applications. *Molecules* **24**, (2019)

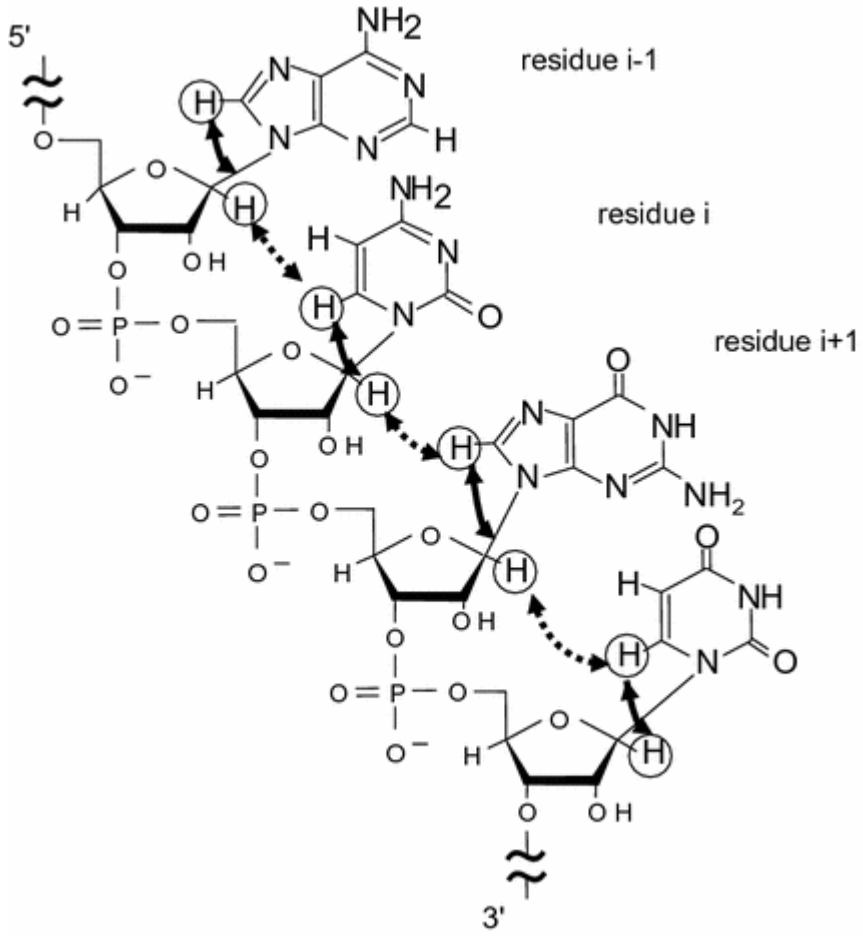


Figure 1-6. Schematic of NOE connectivities of aromatic and ribose protons. H1' of residue i has NOE connectivities to the H6 of residue i and H8 of residue i+1. Arrows indicate intra-residue NOE connectivities and dotted arrows represent inter-residue NOE connectivities. Adapted from Sakamoto, T *et al.* NMR studies on RNA. In *Experimental Approaches of NMR Spectroscopy: Methodology and Application to Life Science and Materials Science*; Springer Singapore, 2017; pp. 439–459 ISBN 9789811059667.

Chapter 2. Extended Interactions between HIV-1 Viral RNA and $tRNA^{Lys3}$ Are Important to Maintain Viral RNA Integrity

Note: This chapter is based on a research paper that was published in International Journal of Molecular Sciences in December 2020.

Abstract

The reverse transcription of the human immunodeficiency virus 1 (HIV-1) initiates upon annealing of the 3'-18-nt of $tRNA^{Lys3}$ onto the primer binding site (PBS) in viral RNA (vRNA). Additional intermolecular interactions between $tRNA^{Lys3}$ and vRNA have been reported, but their functions remain unclear. Here, we show that abolishing one potential interaction, the A-rich loop: $tRNA^{Lys3}$ anticodon interaction in the HIV-1 MAL strain, led to a decrease in viral infectivity and reduced the synthesis of reverse transcription products in newly infected cells. In vitro biophysical and functional experiments revealed that disruption of the extended interaction resulted in an increased affinity for reverse transcriptase (RT) and enhanced primer extension efficiency. In the absence of deoxyribose nucleoside triphosphates (dNTPs), vRNA was degraded by the RNaseH activity of RT, and the degradation rate was slower in the complex with the extended interaction. Consistently, the loss of vRNA integrity was detected in virions containing A-rich loop mutations. Similar results were observed in the HIV-1 NL4.3 strain, and we show that the nucleocapsid (NC) protein is necessary to promote the extended vRNA: $tRNA^{Lys3}$ interactions in vitro. In summary, our data revealed that the additional intermolecular interaction between $tRNA^{Lys3}$ and vRNA is likely a conserved mechanism among various HIV-1 strains and protects the vRNA from RNaseH degradation in mature virions.

Introduction

One of the most distinctive features of retroviruses is that their RNA genome is reverse-transcribed into double-stranded DNA [102], which is integrated into the host genome during infection [44]. Catalyzed by viral reverse transcriptase (RT) [45], the human immunodeficiency

virus 1 (HIV-1) reverse transcription initiates on a double-stranded region where the 3'-end of $tRNA^{Lys3}$ forms 18 base pairs with complementary residues in the primer binding site (PBS) (Figure 1a–c) [44,45]. It has been reported that extended interactions between $tRNA^{Lys3}$ and viral RNA (vRNA) are necessary for efficient reverse transcription and $tRNA^{Lys3}$ selection [51–55]. The extended interactions are believed to maintain the overall tertiary structure of the RNA complex to efficiently load the RT and facilitate the transition from reverse transcription initiation to elongation [54,55]. Attempts to force HIV-1 to use alternative tRNAs have been challenging, as HIV-1 RNA carrying a mutated PBS complementary to alternative tRNAs co-packaged in virions exhibited a delay in viral replication, and the PBS residues eventually reverted back to wild-type (WT) sequences. In order to prevent reversion, additional mutations in the A-rich loop region of the vRNA were introduced to be complementary to the anticodon residues of alternative tRNAs, supporting the role of the A-rich loop in the selection of $tRNA^{Lys3}$ [51,52,56–58]. The A-rich loop residues are highly conserved in HIV-1 [59,103]. A similar bulge is also present in other retroviruses, including HIV-2 and simian immunodeficiency virus (SIV) and exists at a similar distance to the PBS [59,104].

Numerous efforts have been spent to identify the extended interactions between vRNA and $tRNA^{Lys3}$ in the annealed complex. However, based on different types of experimental systems and different viral strains, there are conflicting models about the exact nature of these extended interactions. In vitro and in situ probing of the annealed RNA complex of HIV-1 subtype A revealed that the A-rich loop residues of the vRNA formed base pairs with residues in the anticodon and anticodon stem of $tRNA^{Lys3}$ [59,60], and cell-free reverse transcription assays reported that such interactions facilitated the efficient transition from initiation to elongation during reverse transcription [54,55]. However, NMR studies of the annealed complex did not detect the A-rich loop: anticodon interaction [61,63,105]. In HIV-1 subtype B, it was reported in

the tRNA selection assays that additional mutations in the A-rich loop region were necessary to maintain the usage of non-self tRNAs during virus replication, supporting the role of the A-rich loop in tRNA selection [51,56,64–66]. On the other hand, in vitro reverse transcription assays and biophysical studies of the annealed complex suggest an alternative model, in which the primer activation signal (PAS) residues upstream of the PBS form intermolecular base pairs with TΨC residues in $tRNA^{Lys3}$ [67,106,107]. The reverse transcription assays suggested that the PAS helped to promote the efficient initiation of reverse transcription [67,69,107]. However, such an interaction was not detected in chemical probing and was not supported due to the lack of an adaptive PAS mutation in long-term virus replication [70]. Additionally, a structural analysis by cryo-electron microscopy did not report any extended interactions in HIV-1 subtype B RNA [50], suggesting that these extended interactions are likely to be disrupted when cDNA is actively synthesized. Despite their prominent roles in reverse transcription, molecular details of the extended interactions in the vRNA: $tRNA^{Lys3}$ complex remain unclear.

Here, we present NMR evidence supporting the A-rich loop: $tRNA^{Lys3}$ anticodon interaction. Consistent with previous studies [55], abolishing such an interaction led to a decrease in infectivity and reduced the synthesis of reverse transcription products. Our in vitro primer extension experiments revealed that disruption of the A-rich loop: $tRNA^{Lys3}$ anticodon interaction in HIV-1 subtype A RNA resulted in an increased affinity for RT and enhanced the reverse transcription efficiency. In the absence of deoxyribose nucleoside triphosphates (dNTPs), the RNaseH activity of RT degraded the annealed vRNA: $tRNA^{Lys3}$ complex, and the complex with a disrupted A-rich loop: $tRNA^{Lys3}$ anticodon interaction underwent an accelerated degradation. Consistently, the loss of vRNA integrity was observed in virions carrying A-rich loop mutations. We also reported similar results in HIV-1 subtype B RNA, that the A-rich loop: anticodon interaction affected the RT- binding affinity, primer extension efficiency, and RNA

integrity. Additionally, we show that the nucleocapsid (NC) protein was required to promote the extended vRNA: $tRNA^{Lys3}$ interaction under in vitro conditions. Together, our results suggest a conserved mechanism among various HIV-1 strains: that the extended vRNA: $tRNA^{Lys3}$ interaction protects the vRNA in the mature virions prior to entering a new host cell.

Results

A-Rich Loop Mutations in the PBS Segment Disrupted Extended Interaction Between vRNA and $tRNA^{Lys3}$ and Led to Reduced Infectivity

In our study, a minimal HIV-1 MAL construct PBS segment (nt 123−240, with two non-native G-C pairs) was synthesized by in vitro T7 transcription to investigate the possible extended interactions between vRNA and $tRNA^{Lys3}$ (Figure 2-1a). The NMR assignment of residues in the A-rich loop hairpin and the PAS stem were made by referencing the spectra with the NMR data collected for the control RNA fragments (Figure S2-1). To test the hypothesis that the A-rich loop is involved in extended vRNA: $tRNA^{Lys3}$ interactions, G162-A167 in the PBS segment were mutated to a UNCG tetraloop enclosed by a CG base pair (PBS-MutA, Figure 2-1b). The wild-type (WT) PBS segment and PBS-MutA were annealed with $tRNA^{Lys3}$ (Figure 2-1c) by incubating the mixture at 94 ◦C for 5 min and slowly cooling down to room temperature (Experimental Procedures), and the predicted models are shown in Figure 2-1d–f. The annealed complexes were examined by an electrophoretic mobility shift assay (EMSA) (Figure 2-1g), and the results show that both the WT PBS segment and PBS-MutA formed a complex with $tRNA^{Lys3}$. The A-rich loop mutation did not disrupt the formation of the PBS-MutA: $tRNA^{Lys3}$ complex, suggesting that the A-rich loop: anticodon interaction is not the major inter-molecular interaction contributing to the formation of the vRNA: $tRNA^{Lys3}$ complex. Next, NMR data were collected for the PBS segment: $tRNA^{Lys3}$ and PBS-MutA: $tRNA^{Lys3}$ complexes. The annealing of

$tRNA^{Lys3}$ to the WT PBS segment led to a drastic spectral change (Figure S2-2). Spectra broadening was observed in the PBS segment: $tRNA^{Lys3}$ complex because of the large molecular size (63 kDa). Several characteristic peaks belonging to the stem of the A-rich loop hairpin disappeared or shifted upon tRNA annealing. In particular, A155.H2 and A174.H2 gave rise to characteristic up-field shifted chemical shifts in the PBS segment RNA, and they disappeared upon $tRNA^{Lys3}$ annealing (Figure S2-2 and Figure 2-1h). The NMR spectral changes suggest that the base pairings in the stem were altered after $tRNA^{Lys3}$ annealing (Figure 2-1b). The mutations in the PBS-MutA did not affect the A-rich loop structure in the RNA, as both the A174 and A155 peaks appeared at the same place. The annealing of $tRNA^{Lys3}$ to the PBS-MutA segment did not lead to changes of chemical shifts of residues in the A-rich loop hairpin, and the A155.H2 and A174.H2 remained unchanged (Figure 2-1h). NMR studies of the PBS segment RNA and $tRNA^{Lys3}$-annealed complexes suggest that the mutations in PBS-MutA successfully disrupted the predicted intermolecular interactions between the vRNA and $tRNA^{Lys3}$.

To test whether the mutations of the A-rich loop affect the virus infectivity, a chimeric molecular clone containing the HIV-1 MAL 5' untranslated region (UTR) and NL4.3 back- bone (Δ5'UTR and Δenv, with a reporter an enhanced green fluorescent protein (EGFP) under a cytomegalovirus (CMV) promoter [108]) was made, and viruses pseudo-typed with glycoprotein G from the vesicular stomatitis virus (VSV-G) were prepared (Figure 2-2a,b). PBS-MutA mutations were introduced to the corresponding position of the 5'UTR region in the molecular clone to assess their impact on viral infectivity. An equivalent amount of WT and MutA virions were used in single-cycle infectivity assays, in which the infectivity was quantified by measuring the EGFP signals in infected cells. The results show that the mutation of the A-rich loop led to an approximated 60% reduction of infectivity compared to the WT (Figure 2-2d).

Since the disruption of the extended vRNA: $tRNA^{Lys3}$ interaction is likely to affect

RT loading and reverse transcription initiation [55,109], the reverse transcription products of the WT and MutA infection were quantified. Specific PCR primers and a probe were used to measure the amount of reverse transcription cDNA products two hours post- infection [110]. Consistent with the infectivity results, a mutation of the A-rich loop led to a two-fold decrease in reverse transcription products compared to WT (Figure 2-2e). These data demonstrate that mutations in MutA negatively impacted the early stage of viral replication.

Mutation of the A-Rich Loop Increased the Affinity of RT Binding to the vRNA: tRNA^{Lys3} Complex and Enhanced the Primer Extension Efficiency

To test whether the mutation of the A-rich loop altered the affinity of RT binding to the vRNA: tRNA^{Lys3} complex, we employed microscale thermophoresis (MST) to measure the RT-binding affinity. Since reverse transcription initiates on the genomic RNA after packaging, a 5'UTR construct (nt 1–358) favoring the dimer conformation with a packaging signal exposed was prepared to represent the vRNA [41]. The MutA mutations were introduced to generate 5'UTR-MutA. Both the 5'UTR and 5'UTR-MutA were labeled with a 5'-fluorescein thiosemicarbazide (FTSC) and annealed to tRNA^{Lys3} at 50 nM and a 1:1 ratio. RT was titrated to reach a final concentration ranging from 0.00252 µM to 20.6 µM in the WT complex samples and 0.000763 µM to 6.2 µM in the mutant complex samples. The fluorescence signal changes upon RT titrations versus the RT concentrations that were plotted (Figure 2-3a,b). An analysis of the MST traces showed that the affinity of RT to the 5'UTR: tRNA^{Lys3} complex is 1356 ± 408 nM (Figure 2-3a), and the affinity to the 5'UTR- MutA: tRNA^{Lys3} complex is 78.3 ± 27.6 nM (Figure 2-3b). Disruption of the A-rich loop: anticodon interaction resulted in a nearly 20-fold affinity enhancement of RT binding to the tRNA^{Lys3}-annealed complex for reverse transcription initiation.

Since the annealed complex differed in structure between the PBS segment and PBS-

MutA, and the MutA virus was deficient in producing (-)cDNA in infected cells (Figure 2-2), we hypothesized that the reverse transcription efficiency is affected. To test whether mutations of the A-rich loop led to an alteration of primer extension efficiency, an in vitro primer extension assay was employed. 5'-Cy3-labeled $tRNA^{Lys3}$ was annealed to WT 5'UTR and 5'UTR-MutA and incubated with dNTPs and RT at 37 °C for various amounts of time. The polymerase interaction generated a cDNA product covalently linked to the Cy3-labeled $tRNA^{Lys3}$ primer, and the fully extended product was expected to be 255 nt long (76 nt ($tRNA^{Lys3}$) +179 nt (cDNA product)). Primer extension products were resolved by polyacrylamide gel electrophoresis and imaged by fluorescent scanning (Figure 2-3c). It appeared that the mutation of the A-rich loop led to an increased primer extension, as more full-length products were synthesized on the mutant template at each tested time point (Figure 2-3d). These data agree with the MST affinity results that the 5'UTR-MutA: $tRNA^{Lys3}$ had a stronger affinity to RT, and thus, the reverse transcription efficiency was greater than that of the WT 5'UTR: $tRNA^{Lys3}$ template.

The A-Rich Loop: Anticodon Interaction Protected vRNA from Degradation by RT in the Absence ofdNTP

Our cell-based experiments revealed that the MutA-containing virus had a deficiency in replication step(s) before or during reverse transcription. However, the in vitro binding and primer extension results show that the vRNA: $tRNA^{Lys3}$ complex with A-rich loop mutations was favored by RT and exhibited enhanced efficiency in producing cDNA products. Early reverse transcription may take place in mature virions, but the rate is limited by the dNTP concentrations in the extracellular environment [17,111–113]. Thus, there might be a window of time that RT can bind to the RNA complex but is unable to synthesize cDNA before the virion enters a host cell. While the RNaseH activity of RT mainly cleaves the RNA on a DNA: RNA duplex, the cleavage of RNA on an RNA: RNA template has been previously reported [114].

Therefore, we hypothesized that, before reverse transcription begins, RT could bind to the RNA complex and degrade the complex with its RNaseH activity. To address this possibility, we designed an in vitro degradation assay to monitor the degradation of the 5'UTR over time in the presence of RT without dNTPs (Figure 2-4a). The 5'UTR and 5'UTR-MutA were labeled with 3'-FTSC and annealed to $tRNA^{Lys3}$ to visualize the degraded RNA fragments from the 3'-end. The appearance of a degradation band slightly higher than 156 nt was observed in both the WT and mutant but was absent when no RT was added (Figure 2-4b), indicating the degradation was caused by the addition of RT. The degradation product was stronger in 5'UTR-MutA than in the WT at 40 min and 60 min. We attributed this band to be from the RNaseH activity of RT, as it did not appear when incubating the RNA complex with RTE478Q, a RT mutant reported to abolish the RNaseH activity [115–117] (Figure 4b).

Since the mutation of the A-rich loop increased the degradation of the 5'UTR in vitro, we then checked the integrity of genomic RNA in virions. According to the in vitro degradation results, RT cleaved at the 3'-end of the PBS and, thus, could generate large fragments that could be detected by PCR primers targeting the RU5 region. Thus, the genomic RNAs extracted from virions were subjected to oligo dT-coated magnetic beads to purify RNA with polyA tails and remove the degraded RNA fragments upstream of the PBS (Figure 2-4c). qRT-PCR was performed with specific primers/probes targeting RNA regions upstream and downstream of the PBS (RU5 and Gag, respectively). The genomic RNA integrity was calculated as the ratio of RU5: Gag. In each parallel experiment, the RU5: Gag ratio of the MutA virus was always lower than that of the WT, suggesting the genomic RNA in the MutA virus lost its RU5 region in mature virions. Our data show that A-rich loop mutations led to an approximately three-fold reduction of genomic RNA integrity in virions (Figure 2-4d). Based upon our in vitro results, we attributed this to be from RNaseH- mediated degradation in virions.

The A-Rich Loop: Anticodon Interaction Is Conserved in Other HIV-1 Subtypes to Protect vRNA Integrity

To understand if the A-rich loop: anticodon interaction to maintain viral genomic RNA integrity is conserved in other subtypes, we introduced A-rich loop mutations into the subtype B NL4.3 PBS region (PBS-NL4.3) to generate PBSm-NL4.3, in which the unpaired adenosines were substituted by UA pairs (Figure 2-5a,b). The infectivity of the mutant was measured in single-cycle infectivity assays using the HIV-1 NL4.3-derived reporter virus (pNL4.3-CMV-EGFP Figure 2a) [118]. Pseudo-typed viruses were propagated in 293FT cells, and virus-containing supernatants were harvested 48 h post-transfection. An equivalent amount of WTand mutant viruses were used to infect cells, and the infectivity was analyzed by monitoring the EGFP signals using flow cytometry [118]. As expected, the A-rich loop mutation in the PBSm led to modest infectivity attenuation in the single-round infectivity assay (Figure 2-5b). Reverse transcription products were quantified two hours post-infection, and the mutations in PBSm led to a two-fold reduction of cDNA synthesis in infected cells (Figure 2-5c). These data indicate that the A-rich loop mutation in HIV-1 NL4.3 also caused a replication deficiency. To address the hypothesis that the mutation alters the structure of the vRNA: $tRNA^{Lys3}$ complex and, thus, cannot prevent RT from degrading the genomic RNA, the integrity of the genomic RNA extracted from the virions was quantified using the same procedure as illustrated in Figure 2-4c. As expected, the integrity of the genomic RNA in PBSm viruses were significantly reduced than that of the WT viruses (Figure 2-5d). Collectively, these cell-based results of the NL4.3 viruses are consistent with the data collected for the MAL strain, suggesting the A-rich loop: anticodon interaction is likely conserved among various HIV-1 strains to protect the genomic RNA integrity in virions prior to entering a new host cell.

Formation of the A-Rich Loop: Anticodon Interaction Requires Nucleocapsid (NC) Under In Vitro Conditions

In order to confirm that the A-rich loop: anticodon interaction was disrupted in PBSm-NL4.3, we annealed PBS-NL4.3 and PBSm-NL4.3 with $tRNA^{Lys3}$ in vitro and investigated the complex structure by EMSA and NMR. One of the unexpected findings was that the structure of the annealed complex was different, depending on the annealing conditions. NC-promoted annealing was performed by mixing the PBS-NL4.3 RNA and $tRNA^{Lys3}$ at a 1:1 ratio with various amounts of NC at 37 °C. NC proteins were removed from the RNA complex by high salt washes (10 mM Tris-HCl, pH 7.5, 1.5 M NaCl, and 1 mM $MgCl_2$), and the annealed complexes were analyzed by EMSA. Under all conditions, a stable vRNA: $tRNA^{Lys3}$ complex was formed in the absence of Mg^{2+} (Figure 2-6c, top gel). Interestingly, when the complexes were electrophoresed in the presence of 2 mM $MgCl_2$, these complexes resolved in different bands (Figure 2-6c, bottom gel). The complex annealed with one equivalent of NC (NC: PBS-NL4.3: $tRNA^{Lys3}$ = 1:1:1) overnight showed a weak slow-migrating band (slow-complex 1) (Figure 2-6c, bottom gel, lane 3). The complexes annealed with more NC showed a new slow-migrating band (slow-complex 2). The intensity of these bands increased with the amount of NC they incubated with, even though NC was removed from the RNA samples prior to gel electrophoresis (Figure 2-6c, bottom gel, lanes 4–7). However, the heat-annealed complex (Experimental Procedures) resolved in a similar pattern as the 1-NC-annealed complex (Figure 2-6c, bottom gel, lane 8). These data suggest an alternative RNA complex structure when more NC was included in the annealing step. We suspected that the 12-NC-annealed complex exposed certain residues that mediate dimerization of the annealed complex (slow-complex 2), but these residues were in a different structure and not available to mediate dimerization when the complex was annealed by 1-NC or by heat. These fast- and slow-migration complex bands were also reported previously in NC-

promoted vRNA: $tRNA^{Lys3}$ complexes [119]. Furthermore, the position of the slow complex formed in the 1-NC-annealed complex (slow-complex 1) was lower (migrated faster) than the slow complex formed in the 12-NC- annealed complex (slow-complex 2, Figure 2-6c), suggesting structural differences in these slow complexes. Further, the NMR analysis of these heat- and NC-annealed complexes were performed to assess their structures. If an A-rich loop: anticodon interaction occurs, then the chemical shifts of residues within or close to the A-rich loop of the PBS-NL4.3 are expected to undergo large shifts. Although we were not able to complete the NMR assignment of the entire complex, we found several characteristic peaks were particularly helpful in distinguishing different RNA complex structures (Figure 2-6a, b). For example, the A147.H8 signal was detected in the spectra of both heat- and 1-NC-annealed complexes, which were of the same chemical shift and intensity as the A147.H8 signal in the PBS segment spectrum (Figure 2-6d). These data indicate that the A-rich loop is likely not affected upon $tRNA^{Lys3}$ annealing and support the complex model shown in Figure 6b. However, such a signal was significantly reduced in the 12-NC-annealed complex spectrum, demonstrating that the chemical environment of A147 in the complex annealed by 12 NC was different (Figure S2-4). We compared the intensity of the 147.H8 signal to the 157.H2, which remained unchanged in all the annealed complexes, and the result showed that approximately 50% of the A147.H8 signals disappeared/shifted upon 12-NC-promoted tRNA annealing (Table S2-1). The disappeared or shifted A147 peak supported the extended A-rich loop: anticodon interaction in the model shown in Figure 2-6a.

We then examined if the mutations in PBSm-NL4.3 disrupted the proposed A-rich loop: anticodon interaction. When annealed by a sufficient amount of NC (NC: RNA = 12: 1), PBSm-NL4.3: $tRNA^{Lys3}$ resolved into two bands that were clearly distinct from the WT complex in a native polyacrylamide gel (Figure 2-6e). Both the heat-annealed complex and NC-annealed

PBSm-NL4.3: $tRNA^{Lys3}$ complex contained a slow-complex 1 shown on the gel, similar to the heat-annealed complex formed by PBS-NL4.3 and $tRNA^{Lys3}$ (Figure 2-6e). These results suggest that the slow-complex 2 observed in the PBS-NL4.3: $tRNA^{Lys3}$ complex annealed by a sufficient amount of NC is related with the formation of the A-rich loop: anticodon interaction. This is further supported by the NMR data of the PBSm-NL4.3: $tRNA^{Lys3}$ complex annealed by 12 NC. The A147.H8 of PBSm-NL4.3 remained unchanged in the 12-NC-annealed complex (Figure 2-6f), demonstrating no structural change in or near the A-rich loop region upon $tRNA^{Lys3}$ annealing (Figure 2-6b).

While the annealing conditions affected the PBS-NL4.3: $tRNA^{Lys3}$ complex structure, we noticed that both the in vitro annealing conditions led to the same complex structure in the MAL PBS segment and 5'UTR (Figure S2-3). No slow-complex bands were observed when running the heat-annealed or NC-annealed complex on a gel containing Mg^{2+}. Additionally, heat annealing was sufficient to disrupt the formation of the A-rich loop hairpin in the PBS segment RNA: $tRNA^{Lys3}$ complex (Figure 2-1 and Figure S2-2). These results agree with the previously reported findings that heat and NC annealing resulted in the same complex structure in MAL RNAs [120].

The A-Rich Loop: Anticodon Interaction Protects the Viral RNA from the RNaseH Degradation of RT

Since NC annealing promoted the A-rich loop: anticodon interaction in PBS-NL4.3: $tRNA^{Lys3}$, we next tested if the complex is resistant to RT RNaseH degradation in the absence of dNTPs, as we observed in the MAL RNAs. Both 3'-FTSC-labeled PBS-NL4.3 and PBSm-NL4.3 were annealed to $tRNA^{Lys3}$ in the presence of 12 equivalent NC and then mixed with RT for three hours. While the WT RNA was partially degraded over time, the PBSm-NL4.3 RNA was almost completely degraded after two hours. The major degradation band was slightly higher than the

24-nt RNA ladder, indicating the major cleavage site is at or near C200, the 3'-end of the PBS (Figure 2-5a). When PBS-NL4.3 was heat-annealed to $tRNA^{Lys3}$, the complex was less resistant to RT degradation, and the degradation rate was similar to that of the PBSm-NL4.3: $tRNA^{Lys3}$ complex (Figure 2-7b). Collectively, these data demonstrate that the A-rich loop: anticodon interaction protected the viral RNA from RT degradation, and the formation of this extended interaction, in vitro, requires NC. Consistently, RTE478Q did not degrade the template RNA (Figure 2-7b), suggesting the degradation was mediated by the RNaseH activity of RT.

Primer extension experiments and MST affinity measurements were carried out to examine if the protection of the RNA integrity comes from the reduced affinity of RT binding to the PBS-NL4.3: $tRNA^{Lys3}$ complex. All of the RNA complexes tested in these assays were annealed in the presence of a sufficient amount of NC to promote the formation of the A-rich loop: anticodon base pairs. As shown in Figure 2-7c, the primer extension efficiency of the WT PBS-NL4.3 was much slower than that of the PBSm-NL4.3, agreeing with the results observed in the MAL RNAs (Figure 2-3c). An MST analysis of RT binding to the PBSm-NL4.3: $tRNA^{Lys3}$ complex suggests that the A-rich loop mutations led to a stronger affinity (dissociation constants (Kd) = 46.5 ± 29 nM, Figure 2-7d, right panel), similar to the RT affinity for PBS-MutA: $tRNA^{Lys3}$ complex (Figure 2-3b). The MST trace of RT binding to the WT complex annealed by NC exhibited a biphasic binding, that one binding event occurred at low RT concentrations and one occurred at high RT concentrations. Data fitting to a biphasic model show two binding events with Kd1 = 49.7 ± 9 nM and Kd2 = 5500 ± 1200 nM (Figure 2-7d, left panel). According to the NMR studies of the 12-NC- annealed complex, ~50% of the complex did not form the A-rich loop: anticodon interaction (Figure S2-4 and Table S2-1). Thus, we think the weak binding is likely RT binding to a vRNA: $tRNA^{Lys3}$ complex with the A-rich loop: anticodon interaction formed, as the affinity is in the same μM range as the affinity of RT for the WT MAL vRNA: $tRNA^{Lys3}$ complex (Kd = 1356 ±

408 nM, Figure 2-3a). The tight binding may come from RT binding to the annealed complex that did not successfully form the A-rich loop: anticodon interaction. Indeed, the affinity of the strong binding event was very close to the affinity measured for both the MAL and NL4.3 RNA complexes containing A-rich loop mutations.

Thus, the in vitro data of NL4.3 RNA confirmed that the A-rich loop: anticodon interaction protected the RNA from RT RNaseH degradation in the absence of dNTP. Our data demonstrate that such extended interactions between vRNA and $tRNA^{Lys3}$ exist in two HIV-1 experimental strains and could be a conserved mechanism that HIV-1 employs to protect its genomic RNA in mature virions prior to entering a newly infected cell.

Discussion

Our findings demonstrate the importance of the A-rich loop residues upstream of the 18-nt tRNA annealing site on HIV-1 replication. Supported by NMR data, the A-rich loop residues are likely involved in the proposed A-rich loop: anticodon interaction with $tRNA^{Lys3}$ (Figures 2-1h and 2-6d). The mutation of the A-rich loop residues in both the MAL and NL4.3 RNAs led to the disruption of the extended interactions. In agreement with previous studies [55], the importance of the A-rich loop and the extended vRNA: $tRNA^{Lys3}$ interactions were demonstrated by a reduced viral infectivity and deficiency in producing cDNA in infected cells (Figures 2-2 and 2-5b,c). However, we found RT preferred binding to the A-rich loop mutated vRNA: $tRNA^{Lys3}$ complex and exhibited a higher extension efficiency on the mutant primer: template complex (Figures 2-3 and 2-7c,d) than the WT primer: template complex. We therefore hypothesized that the A-rich loop: anticodon interaction could serve to protect the vRNA from degradation by the RNaseH activity of RT in the absence of dNTPs. Using an in vitro degradation assay, we showed that the RT degraded the genome at approximately the 3'-end of the 18-nt

$tRNA^{Lys3}$ annealing site. Loss of the segment upstream of the PBS in the viral genomic RNAs carrying A-rich loop mutations were also observed. In summary, our data suggest that the A-rich loop: anticodon interaction serves to protect the viral genome before efficient reverse transcription takes place.

Biophysical Evidence Supports the Extended vRNA: $tRNA^{Lys3}$ Interactions

Decades of effort has been spent to identify the extended vRNA: $tRNA^{Lys3}$ interactions in addition to the 18 base pairings between the PBS and the 3'- of $tRNA^{Lys3}$ and investigate their functions in HIV-1 viral replication. Extended intermolecular interactions between vRNA and $tRNA^{Lys3}$ are important in tRNA selection and reverse transcrip- tion initiation [51–55]. Besides the 18-nt annealing base pairs formed between the PBS and $tRNA^{Lys3}$, two major interactions have been proposed, but are yet to be confirmed by further structural evidence. One is between the A-rich loop residues and the anticodon loop of $tRNA^{Lys3}$. Attempts to force HIV to use alternative tRNAs for reverse transcription by mutating the PBS to be complementary to other tRNAs were not successful, as the PBS residues reverted after short-term cultures. However, with additional mutations in the A-rich loop region, the PBS could remain complementary to a non-self tRNA and did not revert back, supporting the role of the A-rich loop in the selection of the use of non-self tRNAs [51,52,121,122]. Liang et. al. showed that deletion of the A-rich loop in HIV-1 HXB2D (Subtype B) led to the reduced infectivity and synthesis of viral DNA. A long-term cell culture led to a partial restoration of the A-rich loop in the A-rich loop-deleted virus [55]. The other interaction is between the PAS residues and the TΨC arm of $tRNA^{Lys3}$, which was supported by cell-free reverse transcription assays and an in vitro biophysical analysis but not by tRNA adaptive selection upon a long-term cell culture [67,69,70,106,123–125]. In vitro and in situ probing of the annealed RNA complex of HIV-1 subtype A support the A-rich loop: anticodon model, and cell-free reverse transcription assays

demonstrated that such an interaction facilitates the efficient transition from initiation to elongation in reverse transcription [54]. However, NMR studies of the annealed complex were not able to detect the A-rich loop: anticodon interaction [61]. Although the results from different groups and experimental settings are controversial, a consensus has been reached that extended vRNA: $tRNA^{Lys3}$ interactions contribute to tRNA selection and modulate the reverse transcription efficiency and viral infectivity. Here, we present indirect NMR evidence supporting the A-rich loop: anticodon interaction in both MAL and NL4.3 RNAs. The disappearance of the A147.H2 and A155.H2 signals in the MAL PBS segment (Figure 2-1h) and the reduced signal of A148-H8 in the NL4.3 RNA upon $tRNA^{Lys3}$ annealing (Figure 2-6d) indicate the residues near the A-rich loop experienced drastic chemical environment changes. The mutating A-rich loop residues in both MAL and NL4.3 RNAs disrupted such extended interactions, as no chemical shift changes of MAL A147.H2, A155.H2, or NL4.3 A147.H8 were observed (Figures 2-1h and 2-6f). The mutations in this study were not designed to investigate the possible PAS: TΨC interaction, and the biological function of the PAS: TΨC interaction, if it exists, remains to be investigated.

The A-Rich Loop: Anticodon Interactions Protect Viral Genomic RNA as a Template for Reverse Transcription

The Marquet group introduced similar A-rich loop mutations into a MAL 5?UTR (nt 1–311) substituting GUAAAA (nt 162–167) by CUAUG and used the RNA as a template for $tRNA^{Lys3}$ annealing and primer extension. They reported the mutation had a slower rate of converting from the initiation phase into the elongation phase [54], and the mutant template had a slightly weaker affinity of RT binding [125]. Within the timeframe of our primer extension assays, we did observe the accumulation of reverse transcription products using the 5'UTR-MutA template, while such early pause bands were barely detected when using the WT 5'UTR template (Figure 2-3c). The discrepancy is likely caused by the low gel resolution in our studies

that were not sufficient to distinguish +3 and +5 products from the un-extended primers. Nevertheless, we observed more full-length product accumulation on the mutant template, indicative of efficient primer extension. Deletion of the A-rich loop in the viral RNA of the HXB2 strain was reported to impair reverse transcription initiation, demonstrated by reduced levels of (-)cDNA products in both cell-based and cell- free assays [55]. Our data agreed that the A-rich loop is crucial for viral replication, as MutA viruses were deficient in producing cDNA in infected cells (Figure 2-2). The discrepancies in the RT-binding affinity and apparent primer extension rate of the A-rich loop mutants are likely caused by the different experimental systems and conditions employed. The structure of the MAL 5'UTR RNA used in our in vitro assays was recently investigated by NMR and was demonstrated to adopt a dimer conformation with NC-binding sites exposed to direct viral genome packaging [41]. Thus, the PBS segment in the 5'UTR is believed to represent, or close to, its conformation in nascent virions when the genomic RNA is packaged. Our NMR data support the formation of an A-rich loop: anticodon interaction in the WT annealed complex and show that such an interaction was disrupted by A-rich loop mutations in MutA (Figure 2-1h). Using such a system, a stronger RT-binding affinity, enhanced primer extension efficiency, and accelerated degradation were observed in MutA. These results agree with our cell-based assays, where the MutA viruses were deficient in synthesizing cDNA products, because their RNA genome was partially degraded in mature virions. Furthermore, similar phenotypes were observed in NL4.3 when its A-rich loop was deleted (Figure 2-5), demonstrating the genome RNA protection strategy by the extended A-rich loop: anticodon interaction is conserved among various HIV-1 strains.

NC Annealing and Heat Annealing Lead to Differently Folded RNA Complexes In Vitro

Annealing of $tRNA^{Lys3}$ onto the PBS segment does not occur spontaneously at a

physiological temperature, as both the PBS segment and $tRNA^{Lys3}$ are highly structured. Complete annealing requires mature NC, which is produced after Gag is processed by viral protease, as in virio SHAPE activities of the viral RNA from protease-deficient virions did not show a similar pattern of complete $tRNA^{Lys3}$ annealing as that of the WT virions [126]. A recent NMR study uncovered that Murine Leukemia Virus (MLV) NC mediates primer annealing by targeting several regions of both the PBS segment and tRNApro primer to expose their complementary sequences for annealing [127], emphasizing the importance of NC in promoting tRNA annealing. Our data show that, for the MAL PBS segment: $tRNA^{Lys3}$ complex, the extended A-rich loop: tRNA anticodon interaction is formed in both heat- and NC-annealing conditions. This is consistent with the previously reported results that examined the structure of the MAL PBS segment: vRNA complexes annealed by heat and by NC using enzymatic probing [120]. However, in the case of subtype B RNA, NC is necessary to promote the extended A-rich loop: tRNA anticodon interaction in vitro. Heat can anneal $tRNA^{Lys3}$ onto the PBS-NL4.3, but both the EMSA and NMR data indicate that the structure of the complex was not the same as the NC-annealed complex (Figure 2-6). The heat-annealed complex was not resistant to RT degradation, which was similar to the complex formed between PBSm-NL4.3 and $tRNA^{Lys3}$ (Figure 2-7b). These findings are in-line with the primer extension results reported by Liang et. al, that the extension efficiency of the del-A mutant was similar to the WT, because the $tRNA^{Lys3}$ was placed on the PBS by heat [55].

Under in vitro conditions, NC did not fully promote the formation of the vRNA: $tRNA^{Lys3}$ complex with the extended A-rich-loop: anticodon interactions. In Figure 2-6c, annealing by 12 NC led to the formation of a major band of slow-complex 2 with sufficient time (three hours, lane 7). Annealing with a shorter time (0.3 h, lane 6) led to a mixture of slow-complex 2 and fast-complex. The appearance of multiple bands in the fast complex in lane 6

may be an intermediate that will form slow-complex 2 with additional time. These results demonstrate the slow kinetics of the NC-promoted annealing process. Our NMR results show that, even at a high NC: RNA ratio (12: 1), NC did not fully promote the formation of the vRNA: $tRNA^{Lys3}$ complex with the extended A-rich loop: anticodon interactions in test tubes (Figure S4 and Table S1). This could be caused by the extensive wash step to remove NC that partially affected the RNA structures. However, a recent single-molecule Förster resonance energy transfer (smFRET) also detected two distinct conformers [105] and suggested a dynamic interplay between the RT and the vRNA: $tRNA^{Lys3}$ complex [128]. While the A-rich loop: anticodon interaction was not detected in the CryoEM structure of the reverse transcription initiation complex [50], such a structure could be a transient structure that temporally protects the genomic RNA integrity.

Rediscovery of the RNaseH Activity of RT

It has been reported that, in addition to the cleavage of RNA in a RNA/DNA hybrid, the RNaseH domain of RT has the ability to cleave RNA in a RNA/RNA hybrid [115,129]. One early report contributed this to a contamination of Escherichia coli RNaseIII [130] from a recombinant protein expression. However, it was shown that the mutation of one of the residues required for binding of the divalent metal ions in the RNaseH active site of RT, E478Q, eliminated the cleavage of both RNA in an RNA/RNA hybrid and RNA in a DNA/RNA hybrid [115], suggesting that this activity is inherent to RNaseH. In our in vitro degradation assay, we employed this RT mutation strategy to show that cleavage in the vRNA: $tRNA^{Lys3}$ complex is RNaseH-specific. In both WT and MutA complexes, the degradation bands were eliminated when mixed with RTE478Q, suggesting that the degradation was RNaseH-specific. It has also been shown that, in arrested complexes, RNaseH can degrade RNA in an RNA/RNA hybrid 18-bp upstream from the 3'-end of the primer [114], due to the proximity between the RNaseH

domain of RT and the 3'-end of the PBS sequence demonstrated in the smFRET studies [128,131]. This cleavage is consistent with the size of our cleavage products in our in vitro degradation assay in both the MAL and NL4.3 RNAs (Figures 2-4b and 2-7a).

Affinity of RT to DNA/RNA, DNA/DNA, and RNA/RNA: A-Rich Loop Interaction Creates Hindrance for RT to Bind

The affinity of RT to RNA/RNA is much weaker than to DNA/RNA and DNA/DNA, with disassociation constants (Kd) of approximately 90 nM for RNA/RNA and 5 nM for DNA/RNA and DNA/DNA [132,133]. Using MST, we observed a similar Kd in the mutant vRNA: $tRNA^{Lys3}$ complexes carrying A-rich loop mutations (Figures 2-3b and 2-7d). However, the Kd observed in MAL WT vRNA: $tRNA^{Lys3}$ was approximately 20-fold weaker. These observations suggest that the A-rich loop: anticodon interaction possibly creates a steric hindrance for RT to bind. The NC-annealed NL4.3 vRNA: $tRNA^{Lys3}$ complex partially formed the A-rich loop: anticodon interaction, and thus, both the tight binding and weak binding were detected (Figure 2-7d). The RT affinity difference between the WT and MutA complexes explains the differences in observed reverse transcription efficiency, in which the greater complex binding affinity to the mutant led to efficient primer extension. As a result, the weaker complex binding affinity of RT to the WT RNA limited the chance of vRNA degradation by the RNaseH activity of RT. The biophysical data, as well as the in vitro functional results, support the hypothesis that the extended A-rich loop: anticodon interaction serves to protect the vRNA in the absence of reverse transcription and prevents it from being degraded by the RNaseH activity of RT.

Author Contributions

Conceptualization, X.H., Z.S., and T.G.; methodology, Z.S., T.G., and J.J.; data collection and analysis, Z.S., T.G., J.J., A.F., and K.W.; and writing X.H., Z.S., T.G, and J.J. All authors have read and agreed to the published version of the manuscript.

For this study I conducted all the in vitro experiments using the HIV-1 MAL RNA and the microscale thermophoresis experiments of the HIV-1 NL4.3 RNA with the help of my undergrad Avery Foster. Zhenwei Song and Kexin Weng conducted all the other in vitro HIV-1 NL4.3 experiments. Juan Ji conducted all of the cell based experiments.

Experimental Procedures

Plasmids

The MAL [134] 5' UTR plasmid was a kind gift from the Michael F. Summers Laboratory (University of Maryland, Baltimore, MD, USA). Multiple DNA fragment assembly using the primers listed in Table S2-2 was performed to prepare plasmids and molecular clones. The chimeric molecular clone pMAL5UTR-NL4.3-Chimera-EGFP was made by replacing the 5'UTR (1–366) in the pNL4.3-CMV-EGFP [108] with the MAL 5'UTR (1–366). MutA mutation was introduced to the MAL-5'UTR plasmid and the pMAL5UTR-MutA- NL4.3-Chimera-EGFP using the primers listed in Table S2-2. PBSM mutation was introduced to the NL4.3-5'UTR plasmid and the pNL4.3-CMV-EGFP in the same way. E478Q mutation was introduced to the RT expression plasmid pHXB2-RT, a kind gift from Dr. Donald Burke (University of Missouri, Columbia, MO, USA) [135,136], to generate pHXB2-RT-E478Q. All the plasmid sequences were confirmed by Sanger sequencing (DNA Core, University of Missouri, Columbia, MO, USA).

In Vitro RNA Transcription

RNA used for in vitro experiments were synthesized by T7 in vitro transcription. The DNA templates for RNA synthesis were produced by amplifying the corresponding sequences with a T7 promoter sequence. To determine optimal conditions for large-scale transcriptions, small-scale trial transcriptions were performed using varying concentrations of $MgCl_2$. Large-scale transcriptions were carried out at volumes of 7.5–10 mL in transcription buffer (40 mM Tris-HCl, pH 8.0, 5 mM dithiothreitol (DTT), 10 mM spermidine, and 0.01% (v/v)

Triton X-100) using dNTPs (12 mM each), Ribolock RNase Inhibitor (80 units, Thermo Fisher Scientific, Waltham, MA, USA), and T7 RNA polymerase (1 μM). Transcriptions reactions were held at 37 °C for 4 h and quenched with 1-M urea and 25 mM ethylenediaminetetraacetic acid (EDTA). RNAs were purified by sequencing gels, visualized by UV shadowing, electroeluted from the gel, and washed in ultra-centrifugal filters to remove acrylamide and salts.

HIV-1 NC, RT, and RTE478Q Purification

Recombinant HXB2-RT and HXB2-RT-E478Q were expressed in E. coli BL21 (DE3)-pLysS cells (Invitrogen, Carlsbad, CA, USA) with 1-mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 37◦C, respectively. The harvested cell pellets were resuspended in lysis buffer (25 mM Tris, 500 mM NaCl, 0.15 mg/mL lysozyme, and 0.4-mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0), sonicated, and centrifuged. The supernatant was applied to a cobalt column (HisPur Cobalt Resin, Thermo Fisher Scientific, Waltham, MA, USA) and purified according to the manufacturer's protocols with an extra-high-salt buffer (25 mM Tris and 1 M NaCl, pH 8.0) washing step. The protein was then subjected to size exclusion chromatography for further purification and buffer exchange and stored in buffer containing 20-mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5), 80 mM NaCl, and 2 mM tris(2-carboxyethyl)phosphine (TCEP) at −80 °C. The HIV-1 NC protein expression plasmid was a kind gift from Dr. Michael Summers (University of Maryland, Baltimore, MD, USA). The recombinant HIV-1 NC protein was expressed and purified exactly as previously described [137].

RNA Fluorescence Labeling

Cy3-labeled tRNA was prepared as described previously [76]. 3'-FTSC-labeled RNAs were prepared by incubating 20 μM of RNA in 200 μM of $NaIO_4$ and 62.5 mM of NaOAc at room temperature in the dark for 90 min, followed by incubating with an additional 0.42 mM Na_2SO_3 at room temperature for 15 min and then mixing with 0.58 mM FTSC dye (dissolved in

dimethylformamide) at 37 ◦C in the dark for 3 h. To quench the reaction, LiCl was added to a final concentration of 0.74 M, and the sample was mixed with 2.5 volumes of 100% ethanol and incubated at −80 °C for 1 h. The sample was centrifuged at 13,000× g for 20 min at 4 °C, and the supernatant was discarded. The precipitated RNA was washed twice with 75% ethanol and lyophilized for 10 min. The lyophilized RNA was dissolved in double-distilled H2O and stored for further use.

NMR Experiments

RNA used for 1D ^{1}H NMR experiments was prepared in 50–200 µM and dissolved in a $^{2}H_2O$ buffer containing 10 mM deuterated Tris-HCl (pH 8.0) and 2 mM $MgCl_2$. RNA used for 2D ^{1}H-^{1}H Nuclear Overhauser Effect Spectroscopy (NOESY) experiments were prepared at 200–300 µM in the same buffer. The NMR data were collected at 308 K on a Bruker Avance III 800 MHz spectrometer equipped with a TCI cryoprobe. Data was processed with NMRPipe and NMRDraw using the NMRBox cloud-based virtual machine [138]. Data was analyzed using NMRViewJ (NMRFX).

RNA Annealing

Heat annealing was carried out by mixing HIV-1 vRNA with $tRNA^{Lys3}$ at a 1:1 ratio in the annealing buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 1 mM $MgCl_2$, and 140 mM KCl) and incubated at 94 °C for 5 min, 85 °C for 15 min, 75 °C for 15 min, and 65 °C for 60 min. NC annealing was carried out by mixing HIV-1 vRNA with $tRNA^{Lys3}$ at a 1:1 ratio and mixing with various amounts of NC, as indicated in the Results section. NC was then removed from the RNA complex by high-salt washes with 10 mM Tris-HCl, pH 7.5, 1.5 M NaCl, and 1 mM $MgCl_2$) for the in vitro analysis.

Microscale Thermophoresis

3’-FTSC-labeled WT and mutant MAL 5’UTR RNA samples were pre-annealed to

$tRNA^{Lys3}$, and 50 nM of the complex was titrated with RT at concentrations varying from 0.00252 µM to 20.6 µM for the MAL 5'UTR: $tRNA^{Lys3}$ experiments and 0.000763 µM to 6.2 µM for the MAL 5'UTR-MutA: $tRNA^{Lys3}$ experiments. Experiments were performed using a Nanotemper Monolith and were analyzed using MO.Affinity Analysis (Nanotemper Technologies, Munich, Germany). Each experiment was repeated three times.

3'-FTSC-labeled WT and mutant NL4.3 RNA were annealed to $tRNA^{Lys3}$ in annealing buffer using NC (vRNA: tRNA: NC = 1:1:12). NC was removed by high salt washes, as described in Section 4.6. The complexes were held at 50 nM in the microscale thermophoresis experiments, and RT was varied from 0.00198 µM to 65 µM for the PBS NL4.3: $tRNA^{Lys3}$ experiments and 0.000763 µM to 25 µM for the PBSm-NL4.3: $tRNA^{Lys3}$ experiments. Experiments were performed using a Nanotemper Monolith and were analyzed using MO.Affinity Analysis (Nanotemper Technologies, Munich, Germany). Three independent replicates of the measurements were performed for the K_d quantification. For the biphasic binding curve, K_d was determined by excluding data points for the low-affinity binding mode during the analysis of the high-affinity binding mode and vice versa [139].

Primer Extension Assay

The MAL and NL4.3 vRNA were annealed to 3'-Cy3 $tRNA^{Lys3}$ in an annealing buffer by heat and by NC, respectively. The annealed complexes were mixed to 90 nM with 3 mM $MgCl_2$, 1 µM NC, Ribolock RNase Inhibitor (Thermo Fisher Scientific, Waltham, MA, USA) 0.6 units, 2.7 µM RT, 25 mM HEPES, and 80 mM NaCl. To initiate the reaction, dNTPs were added to 90 µM and held at 37 °C. Each time point was quenched with 20 mM EDTA and placed on ice. SDS was added to 12.5 mg/mL, and samples were boiled for 5 min. Proteins were removed by phenol–chloroform extraction, and the RNA samples were boiled for 3 min and loaded onto an

8% denaturing 8 M Urea polyacrylamide gel. The gels were visualized by fluorescent scanning on a Typhoon FLA 9000 (GE Healthcare, Marlborough, MA, USA).

In Vitro Degradation Assay

The 3'-FTSC-labeled HIV-1 vRNA was pre-annealed to $tRNA^{Lys3}$, as described in Section 4.6. In a 20 µL reaction, 166 nM of the RNA complex was mixed with 1.2 µM NC and 3 µM RT or RTE478Q in a buffer containing 25 mM HEPES, 80 mM NaCl, and 0.6 units of Ribolock RNase Inhibitor (Thermo-Fisher Scientific, Waltham, MA, USA). The reactions were incubated at 37 °C for the specified times and quenched by the addition of EDTA. SDS was added to 12.5 mg/mL, and samples were boiled for 5 min. Proteins were removed by phenol–chloroform extraction, and the RNA samples were boiled for 3 min and loaded onto an 8% denaturing 8 M Urea polyacrylamide gel. The gels were visualized by fluorescent scanning on a Typhoon FLA 9000 (GE Healthcare, Marlborough, MA, USA).

Cells and Viruses

The 293FT cells were propagated in Dulbecco's modified Eagle's medium (DMEM, Sigma, St.Louis, MO, USA) supplemented with 10% fetal bovine serum (FBS). WT and mutant MAL chimeric viruses were produced by transfection [108]. Briefly, 293FT cells were co-transfected with 500 ng of WT and mutant MAL molecular clones (pMAL5UTR-NL4.3-Chimera-EGFP and pMAL5UTR-MutA-NL4.3-Chimera-EGFP) and 100 ng VSV-G (AIDS Reagent Program). The cells were rinsed and cultured in 10% FBS DMEM 6 h later, and the pseudo-typed viruses were harvested 48 h post-transfection. The cells were fixed with 4% paraformaldehyde, and the transfection efficiency was examined by using flow cytometry (Accuri Flow Cytometer, BD Biosciences, San Jose, CA, USA). The WT and mutant NL4.3 pseudo-typed viruses were produced in the same way using pNL4.3-CMV- EGFP and pNL4.3-PBSm-CMV-EGFP plasmids.

Infectivity Assay

The amount of pseudo-typed viruses was quantified by p24 ELISA, using precoated MicroFleur well plates (Optofluidic Bioassay, Ann Arbor, MI, USA). Pseudo-typed viruses containing 250 ng of Gag p24 was used to infect 2×10^5 of TZM-bl cells (AIDS Reagent Program) in 12-well plates. The infected cells were harvested 20 h post-infection and fixed with 4% paraformaldehyde. The percentage of EGFP cells were measured by Accuri Flow cytometer.

Quantification of cDNA Products in Infected Cells

Equivalent amounts (250 ng of Gag p24) of WT and mutant pseudo-typed viruses were pretreated with DNase I and DpnI and then infected 2×10^5 of 293 cells in 12- well plates. The cells were harvested 2 h post-infection and subjected to DNA extraction with QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Cellular DNA (2 µL) was evaluated by qPCR using primer sets specifically targeting the reverse transcription products [110]. The WT and mutant qPCR data were normalized by the respective cellular DNA concentrations.

RNA Integrity Assay

A total of 250 µl of pseudo-typed viruses were used for RNA extraction with TRIzol Reagent (Invitrogen Carlsbad, CA, USA) following the manufacturer's user guide. The ex- tracted nucleic acids were subjected to DNase I and DpnI digestion (NEB, Ipswitch, MA, USA). The viral genomic RNA was extracted using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, Ipswitch, MA, USA), and the integrity were evaluated by qRT-PCR using the primer and probe sets listed in Table S2-1. The RT-qPCR was performed with the CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA).

Statistics Analysis

An unpaired t-test was used to compare the significance of differences between different groups with Prism (GraphPad Software, San Diego, CA, USA). The p-values less than 0.05 were considered significant. Statistical differences are indicated as follows: * $p < 0.05$, ** p

< 0.01, *** p < 0.001, and **** p < 0.0001. More than three biological replicates were applied in all the experiments.

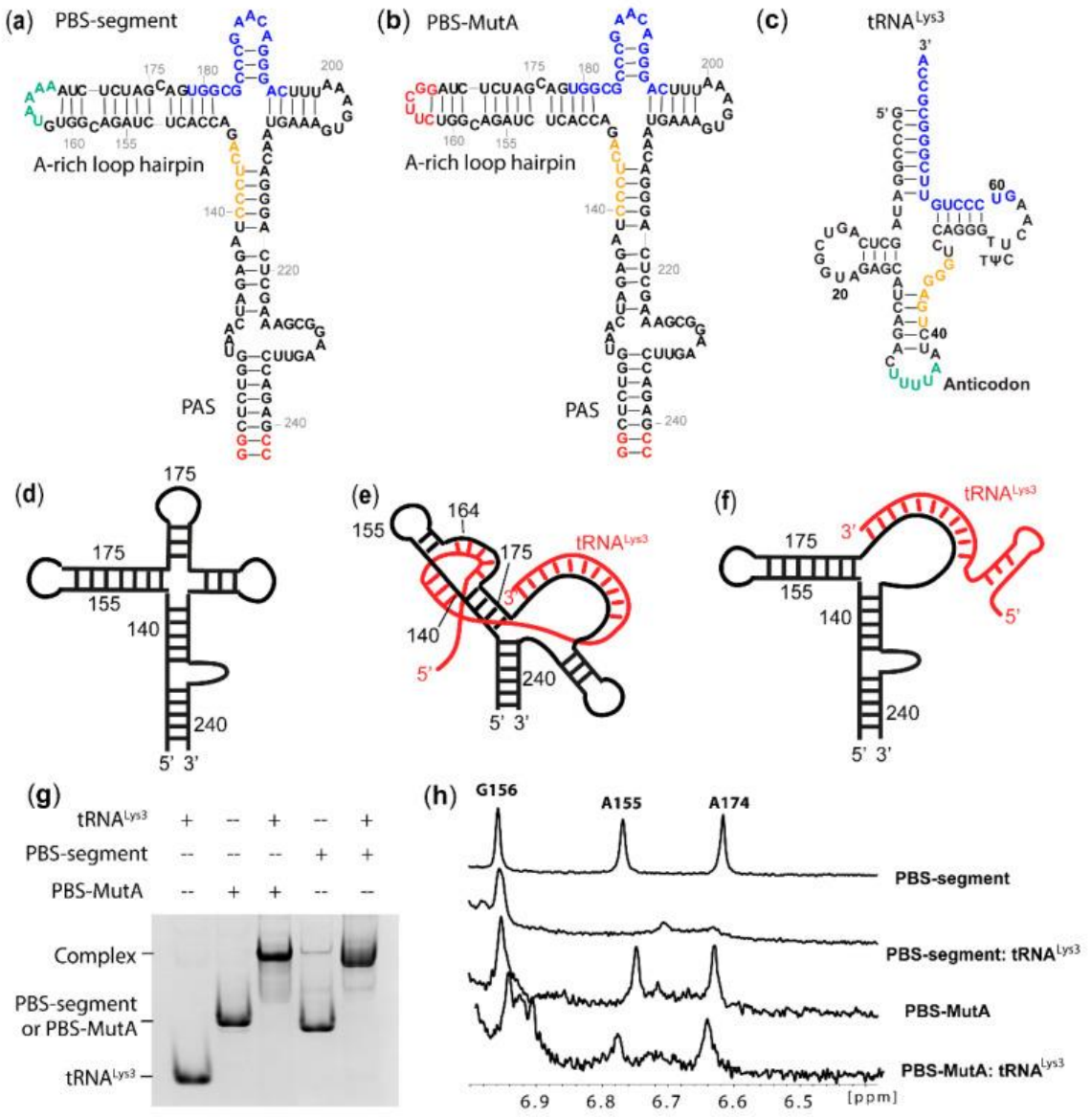


Figure 2-1. Mutation of the A-rich loop did not abolish $tRNA^{Lys3}$ annealing but disrupted the possible A-rich loop: anticodon interactions. (a) Predicted secondary structure of protein binding site (PBS) segment RNA. The A-rich loop residues complementary to the anticodon of $tRNA^{Lys3}$ are labeled in green, with additional complementary residues labeled in yellow. The 18-nt PBS is labeled in blue. Two non- native G-C pairs were added (red) to enhance the T7 transcription. (b) Predicted secondary structure of PBS-MutA, of which G162-A167 were substituted by CUUCGG (red). (c) Secondary structure of $tRNA^{Lys3}$. The 3'-18-nt residues complementary to PBS are

labeled in blue. Anticodon loop residues that are complementary to the PBS segment A-rich loop are labeled in green, with additional complementary residues labeled in yellow. (d–f) Cartoon models of the PBS segment (d), PBS segment: $tRNA^{Lys3}$ complex with extended A-rich loop: anticodon intermolecular interactions (e), and without the extended interactions (f). The MAL PBS segment is labeled in black, and $tRNA^{Lys3}$ is labeled in red. (g) Mutation of the A-rich loop did not prevent $tRNA^{Lys3}$ annealing. Complex formation was examined using an 8% native PAGE. (h) Mutation of the A-rich loop disrupted the possible A-rich loop: anticodon interactions. Portion of 1D 1HNMR of the PBS segment and PBS segment: $tRNA^{Lys3}$ complexes (top) and PBS-MutA and $tRNA^{Lys3}$ complexes (bottom). The signal-to-noise of the PBS-MutA samples were lower, because the RNA sample concentrations used for NMR data collection were lower than the wild-type (WT) RNAs.

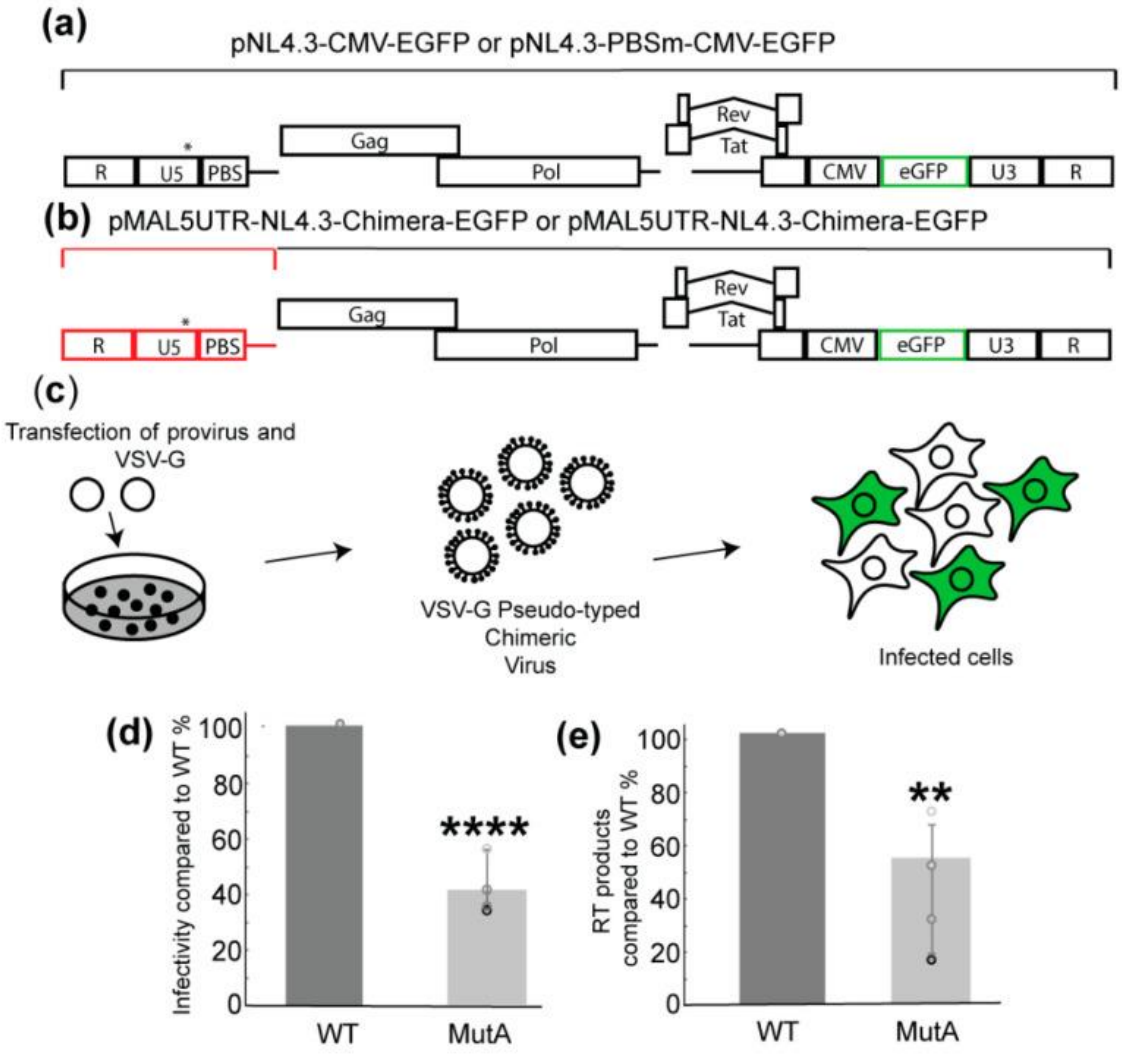


Figure 2-2. Cell-based infectivity and reverse transcription product assays to compare the replication efficiency of the WT and Mut-A viruses. (a) Map of the pro-viral plasmid used for the pseudo-typed virus production and cell-based assays. An asterisk indicates the approximate location of A-rich loop mutations. (b) A chimeric molecular clone was created using the MAL 5'UTR and NL4.3 backbone. The MAL 5'UTR is highlighted in red. An asterisk indicates the approximate location of A-rich loop mutations. (c) Schematic of a single-cycle infectivity assay. Pro-viral and vesicular stomatitis virus (VSV-G) plasmids were transfected in 293FT cells. VSV-G pseudo-typed chimeric viruses were harvested and used to infect TZM-bl cells. Flow cytometry was used to measure the percentage of EGFP-expressing cells. (d) Mutation of the A-rich loop

lowered the infectivity of the pseudo-typed virus. (e) Mutation of the A-rich loop led to reduced reverse transcription products in cells infected by the pseudo-typed virus. The cDNA products were quantified by qPCR. Graphs are representative of four experiments, average ± standard deviation; ** $p < 0.01$ and **** $p < 0.0001$.

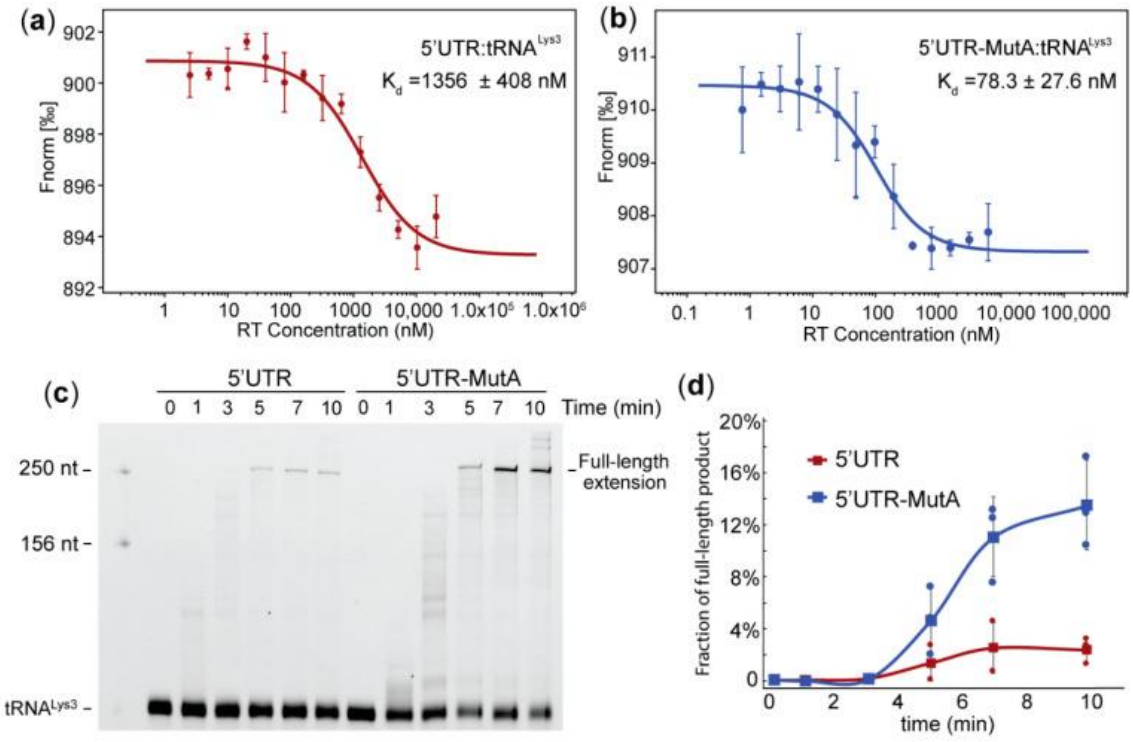


Figure 2-3. Mutation of the A-Rich loop led to increased affinity for reverse transcriptase (RT) binding to the vRNA: $tRNA^{Lys3}$ complex and increased primer extension efficiency in vitro. (a) microscale thermophoresis (MST)-binding curve of RT binding to the 5'UTR: $tRNA^{Lys3}$ complex. (b) MST-binding curve of RT binding to the 5'UTR-MutA: $tRNA^{Lys3}$ complex. Representative curves of three independent experiments are shown in (a,b), average ± standard deviation. (c) Mutation of the A-rich loop in the template RNA led to an increased primer extension. The primer extension products of Cy3-labeled $tRNA^{Lys3}$ on the 5'UTR WT (left) and 5'UTR-MutA (right) were visualized by polyacrylamide gel electrophoresis and imaged by fluorescent scanning. (d) Quantification of the full-length primer extension products averaged from three independent experiments shows enhanced primer extension efficiency upon mutation of the A-rich loop.

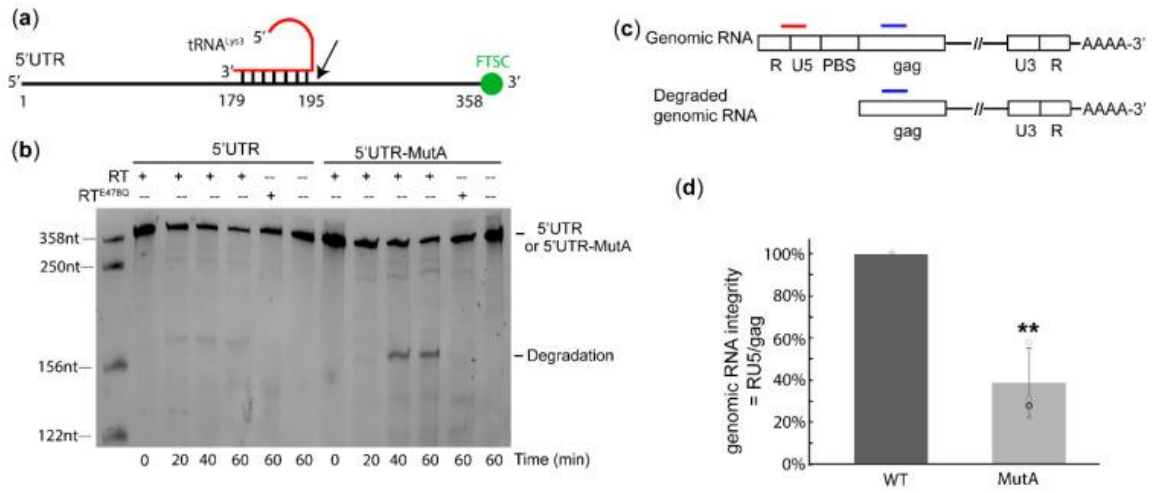


Figure 2-4. A-rich loop mutation in viral RNA (vRNA) resulted in accelerated RNA degradation in both in vitro and in cell-based assays. (a) Schematic diagram for the in vitro degradation assay. $tRNA^{Lys3}$ was annealed to 3'-fluorescein thiosemicarbazide (FTSC)-labeled 5'UTR or 5'UTR-MutA. Black arrow marks the approximate RNaseH degradation site. (b) In vitro degradation assay of the 5'UTR: $tRNA^{Lys3}$ and 5'UTR-MutA: $tRNA^{Lys3}$ complexes shows enhanced degradation of the mutant complex. (c) Schematic diagram of cell-based genomic integrity assay. Primers targeting the RNA upstream and downstream of PBS are sketched in red and blue, respectively. (d) Genomic RNA integrity assay shows decreased genomic RNA integrity upon mutation of the A-rich loop. Graph is representative of three experiments, average ± standard deviation; ** $p < 0.01$.

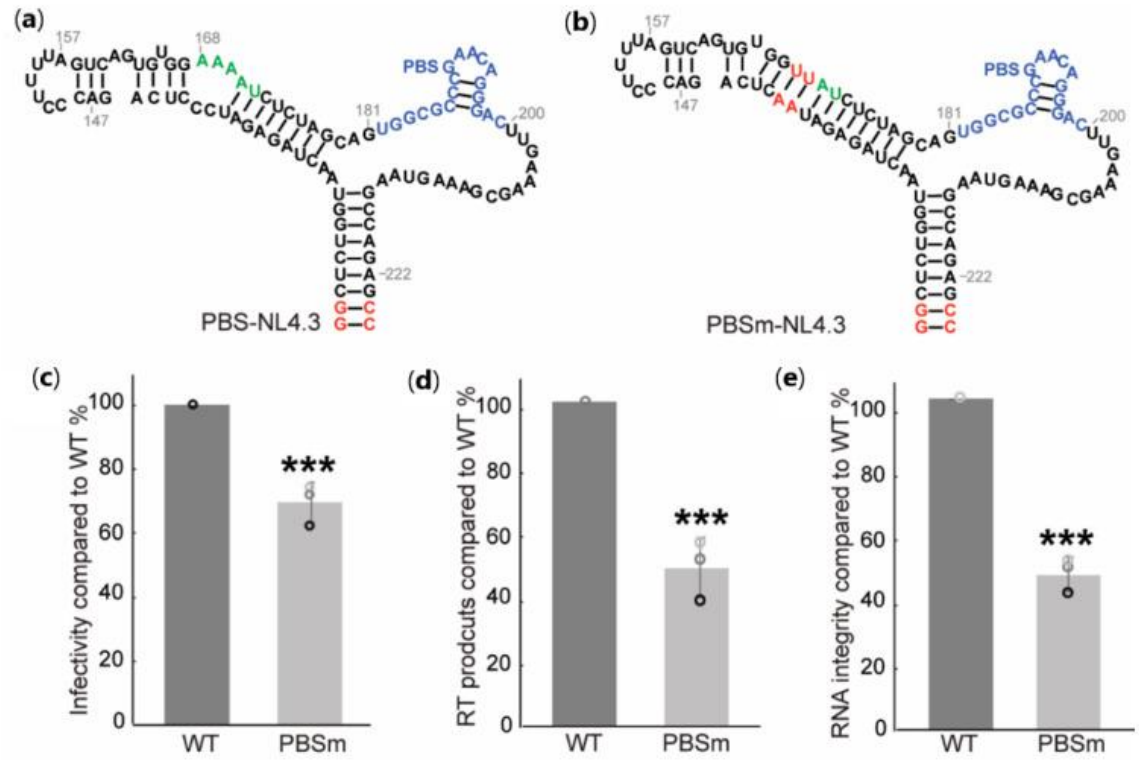


Figure 2-5. Cell-based assays show the A-rich loop residues are important for viral replication in human immunodeficiency virus 1 (HIV-1) subtype B. (a) Predicted secondary structure of PBS-NL4.3 RNA. The A-rich loop is labeled in green, the primer binding site is labeled in blue, and the non-native nucleotides are labeled in red. (b) Predicted secondary structure of PBSm-NL4.3 RNA. A PBS that is complementary to the 3'-18-nt of $tRNA^{Lys3}$ is labeled in blue, and the non- native nucleotides are highlighted in red. (c) Mutation of A-rich loop lowered the infectivity of the pseudo-typed virus in the single-round infectivity assays. The infectivity was quantified by measuring the EGFP signals and normalizing the p24 levels of the virions, and the infectivity of the mutants was normalized to the WT in each parallel experiment. (d) Mutation of the A-rich loop reduced the reverse transcription products 2 h post-infection. (e) Mutation of the A-rich loop resulted in reduced vRNA integrity in the genomic integrity assay. Graphs are representative of three independent experiments, average ± standard deviation; *** $p < 0.001$.

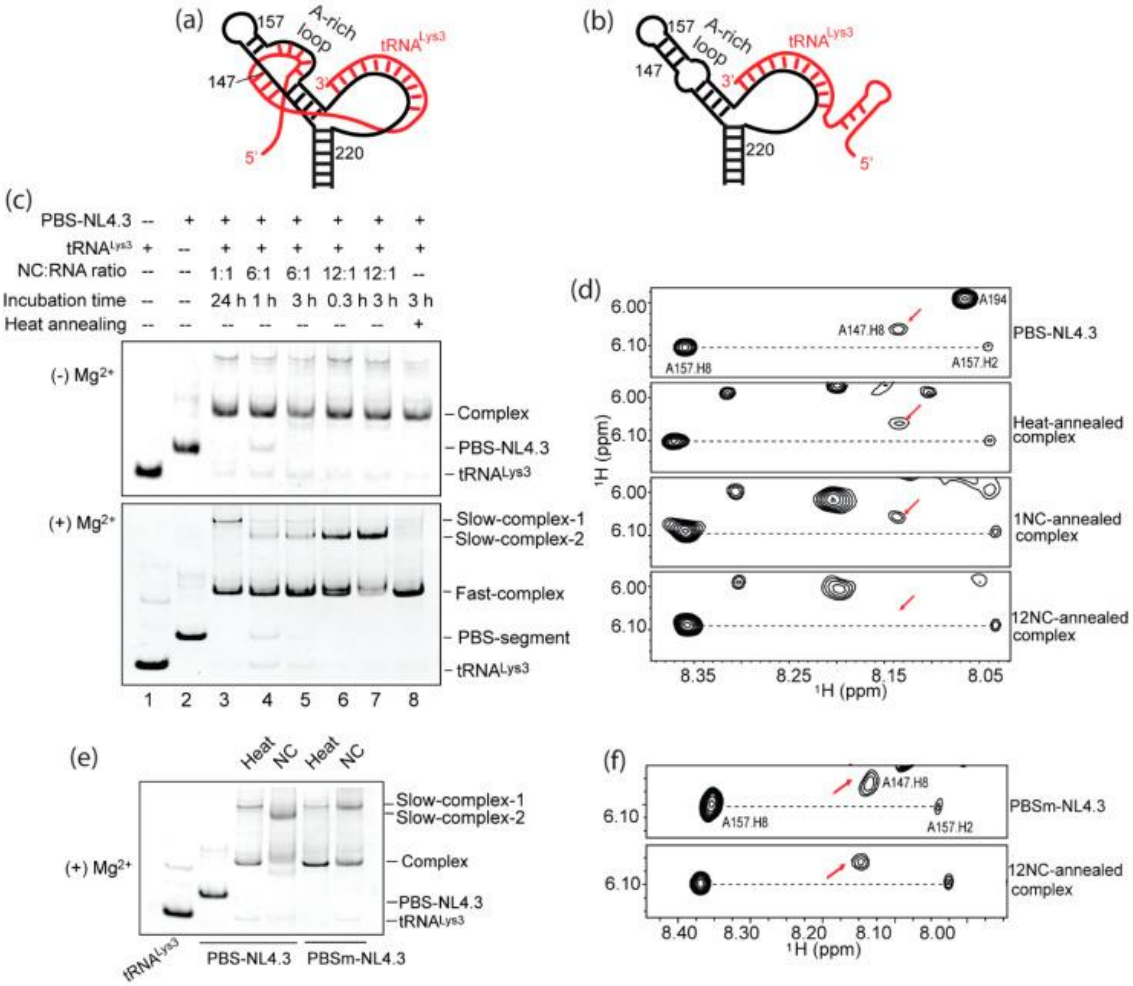


Figure 2-6. Native gel and NMR analysis of annealing conditions on PBS-NL4.3 and PBSm-NL4.3 show that the A-rich loop: anticodon interaction was promoted by nucleocapsid (NC) annealing but not by heat annealing. (a) Sketch of the A-rich loop: anticodon interaction model. PBS-NL4.3 is shown in black, and $tRNA^{Lys3}$ is shown in red. (b) Sketch of the complex model without extended interactions. (c) Complexes annealed under various conditions adopt different structures. The PBS-NL4.3: $tRNA^{Lys3}$ complexes were analyzed by polyacrylamide gel electrophoresis without Mg^{2+} (top panel) and with Mg^{2+} (bottom panel). (d) Portions of 2D 1H-1H NMR of the PBS-NL4.3: $tRNA^{Lys3}$ complexes show that only the 12:1 NC: RNA complex led to a possible A-rich loop anticodon interaction. The red arrow indicates an A147.H8 chemical shift. (e) Native gel analysis of the complexes of PBS-NL4.3:$tRNA^{Lys3}$ and PBSm-NL4.3: $tRNA^{Lys3}$

annealed by heat and by 12 NC. (f) Portions of 2D 1H-1H NMR of PBSm-NL4.3 and the PBSm-NL4.3: $tRNA^{Lys3}$ complex show that the mutation of the A-rich loop disrupted the formation of a possible A-rich loop anticodon interaction. The red arrow indicates an A147.H8 chemical shift.

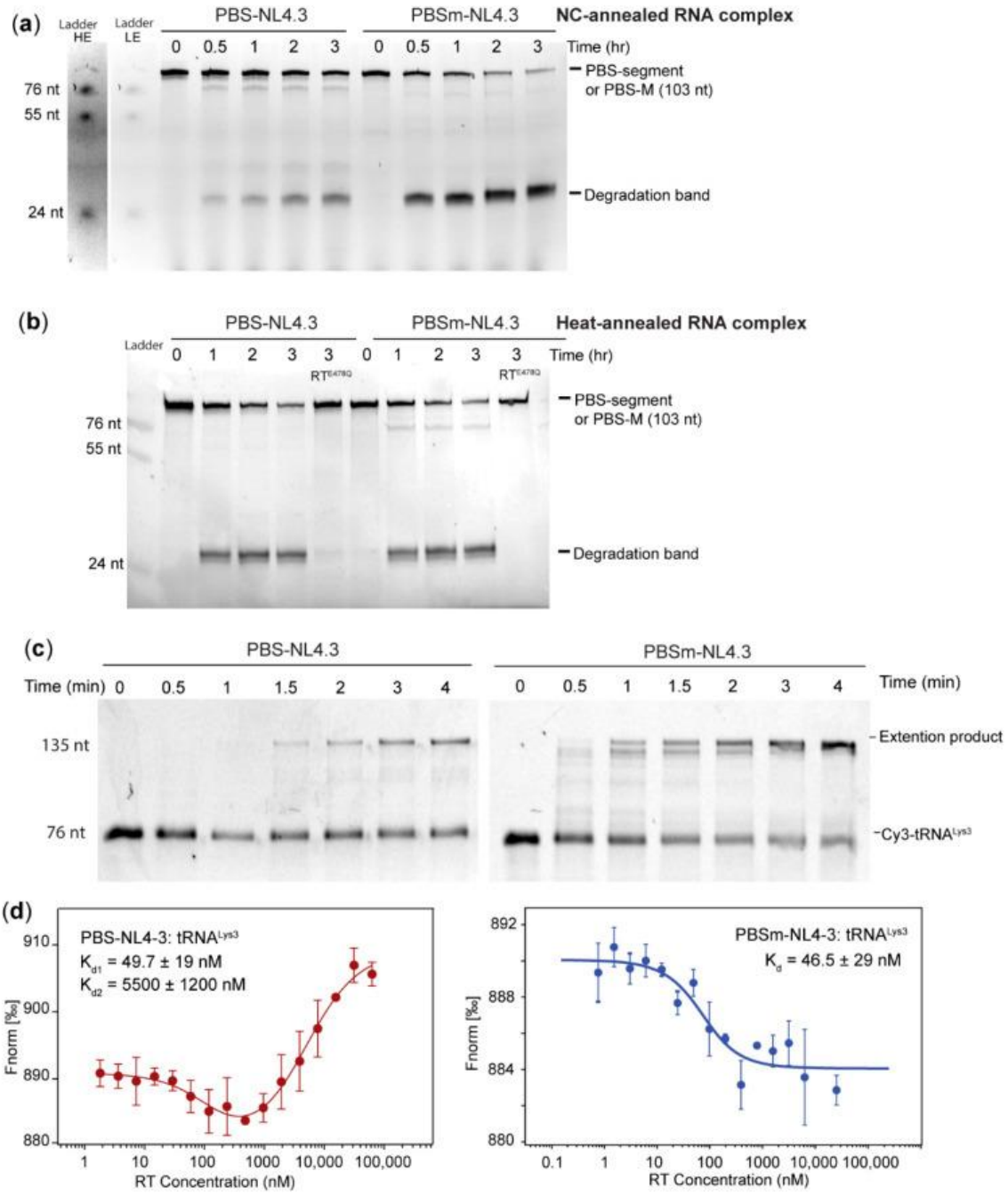


Figure 2-7. The A-rich loop: anticodon interaction is conserved in HIV-1 NL4.3 to protect the vRNA integrity. (a) The PBSm-NL4.3 RNA in the NC-annealed complex was degraded faster the WT RNA in the NC-annealed complex. Ladders are shown in contrasted adjusted levels (left panel, HE = high exposure and LE = low exposure). (b) The PBSm-NL4.3 RNA in the heat-annealed complex was degraded at a similar rate as the WT RNA in the heat-annealed complex.

(c) Primer extension on the mutant PBSm-NL4.3: $tRNA^{Lys3}$ complex was more efficient than on the WT complex. (d) MST binding curves of PBS-NL4.3: $tRNA^{Lys3}$ (left panel) and PBSm-NL4.3: $tRNA^{Lys3}$ (right panel) titrated with RT. Representative curve of three experiments is shown; average ± standard deviation.

Supplementary Materials

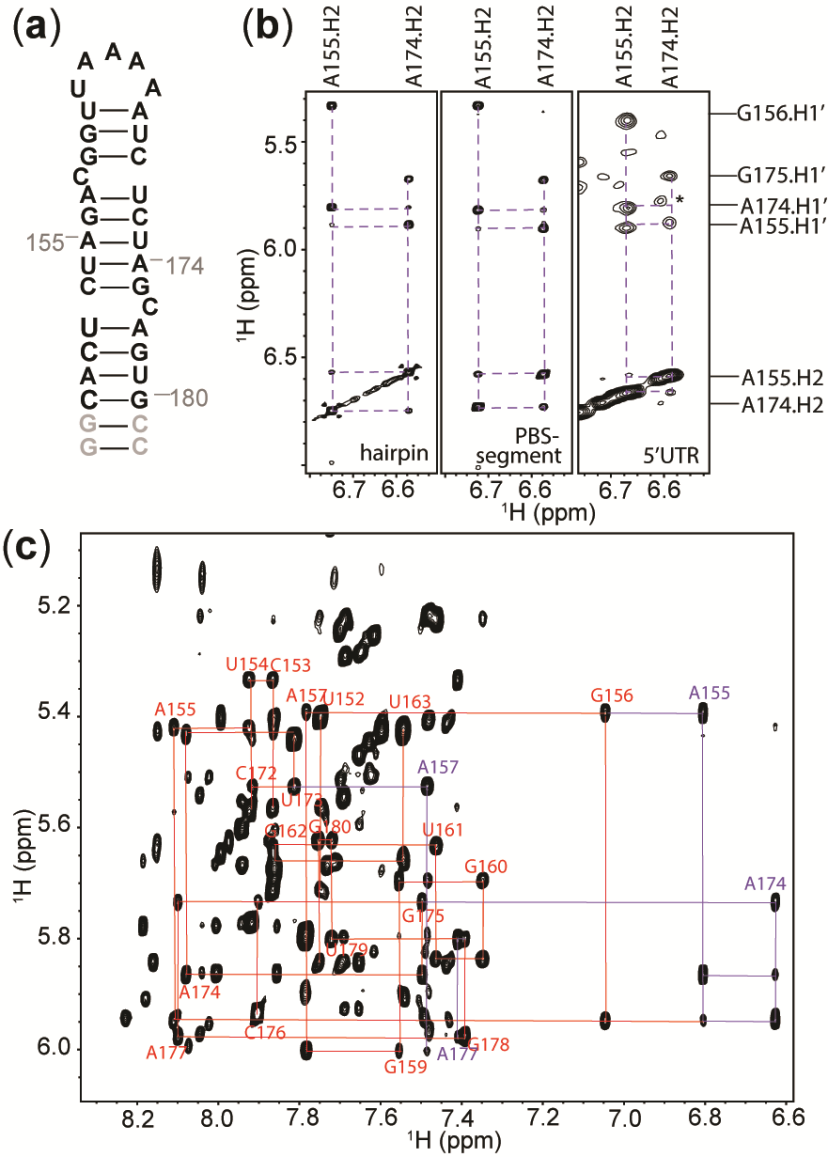


Figure S2-1. A-rich loop hairpin exhibits similar NOE patterns in all RNA constructs used in this study. (a) Secondary structure of A-rich loop hairpin control with non-native bases highlighted in gray. (b) Similar NOE patterns were observed in the hairpin control (left), PBS-segment (middle) and the 5'UTR (right). NOESY walk is labeled in purple dashed lines. (c) Sequential NOESY walk used for assignment of A-rich loop hairpin control. Red dashed lines indicate walk of H6/H8/H1' from U152 to A177. Purple dashed lines indicate adenosine H2 resonances .

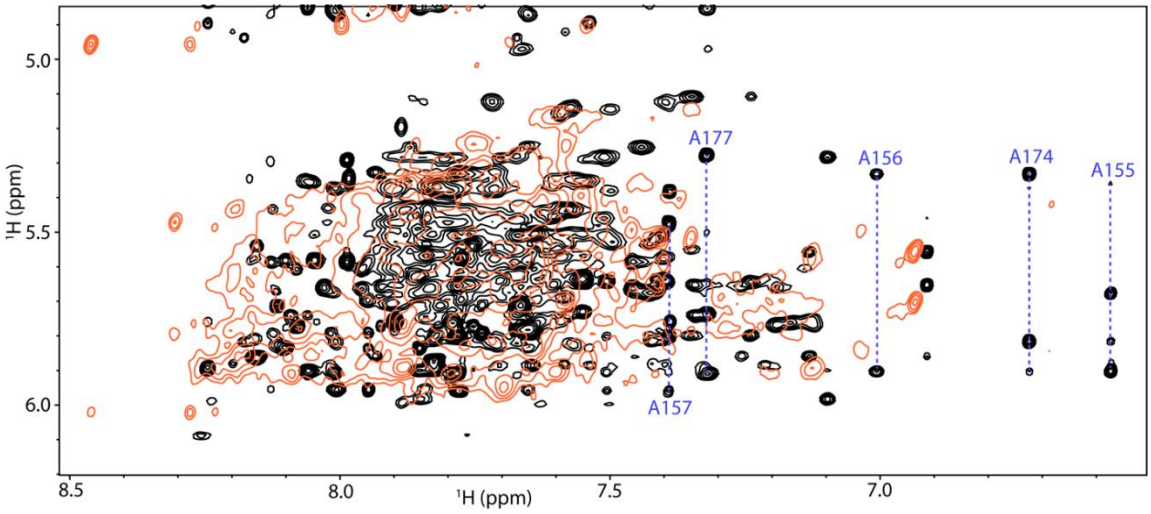


Figure S2-2. Annealing of $tRNA^{Lys3}$ to PBS-segment disrupted folding of A-rich loop hairpin. Portion of the aromatic region of 2D ^{1}H-^{1}H NOESY spectra with PBS-segment shown in black and PBS-segment: $tRNA^{Lys3}$ complex shown in orange. Resonances of A-rich loop hairpin nucleotides are shown with blue dashed lines.

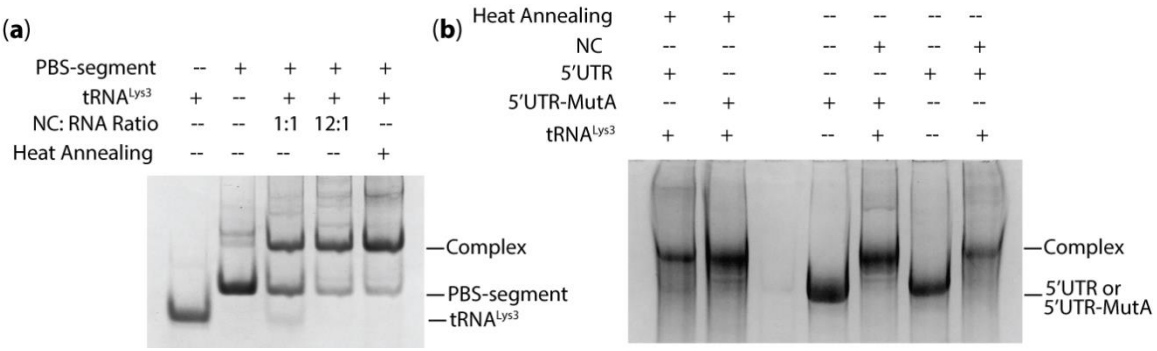


Figure S2-3. The structure of the MAL vRNA : $tRNA^{Lys3}$ complex is not sensitive to annealing conditions. (a) PBS-segment was annealed to $tRNA^{Lys3}$ using NC or heat annealing. Complexes were analyzed using a native polyacrylamide gel. (b) The 5'UTR and 5'UTR-MutA were annealed to $tRNA^{Lys3}$ using NC or heat annealing and the complexes were resolved in a native polyacrylamide gel.

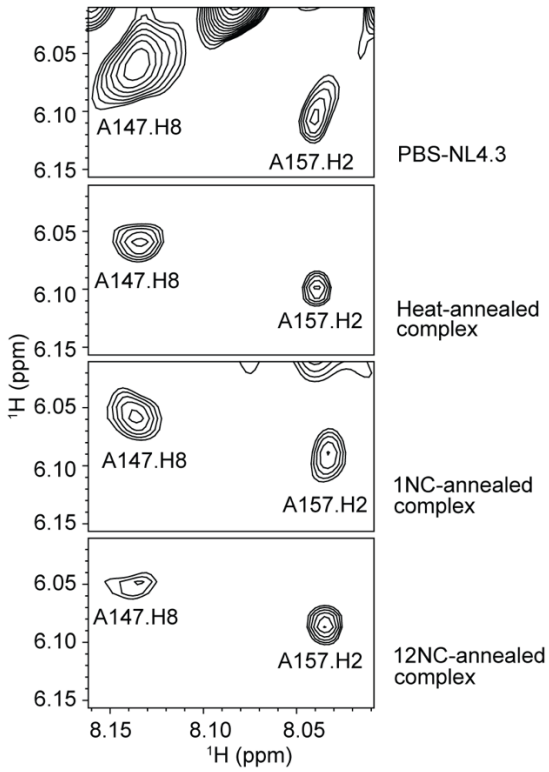


Figure S2-4. The intensity of A147.H8 decreased upon NC annealing. Portions of 2D ^{1}H-^{1}H NOESY NMR spectra of PBS-NL4.3, heat annealed complex, 1 NC annealed complex (NC: RNA=1:1), and 12 NC annealed complex (NC: RNA=12:1) are shown. The NC proteins were removed by high salt washes prior to NMR data collection. Peak intensities are quantified in table S2.

Table S2-1. Table of primers used in this study. Notes list which experiment primers were used for.

Primer	Sequence (5’-3’)	Notes
MAL5UTRNL4-3ChimeraUTRF	CTTTTTGCCTGTACTGGTCTCTCTTGTTAGACCAGGTCGAG	Construction of chimeric virus plasmid
MAL5UTRNL4-3ChimeraBackboneR	AGTACAGGCAAAAAGCAGCTGCTTATATGTAGCATCTGAGGG	Construction of chimeric virus plasmid
MAL5UTRNL4-3ChimeraBackboneF	GGTGCGAGAGCGTCGGTATTAAGCGGG	Construction of chimeric virus plasmid
MAL5UTRNL4-3ChimeraUTRR	CGACGCTCTCGCACCCATCTCTCTCCTT	Construction of chimeric virus plasmid
MAL-TAR-F	GCATCTAATACGACTCACTATAGGTCTCTCTTGTTAGAC	Production of template for transcription
MAL-358-R	CGCACCCATCTCTCTCCTTCTAGC	Production of template for transcription
MAL-MutA-F	GTCTTCGGATCTCTAGCAGTGGCGCCCGAACAG	Mutagenesis of MAL plasmid
MAL-MutA-R	TAGAGATCCGAAGACCGTCTAGAGTGGTCTGAGGGATCTCTAGTTACCAG	Mutagenesis of MAL plasmid
3'-ori-R	CTCAAGTCAGAGGTGGCGAAACCCGACAG	Mutagenesis of MAL plasmid
5'-ori-F	CACCTCTGACTTGAGCGTCGATTTTTGTGATGC	Mutagenesis of MAL plasmid
MAL-PBS-F	GCTAGTAATACGACTCACTATAGGCTCTGGTAACTAGAGA	Production of template for transcription
MAL-PBS-R	GGCTCTGGAACTTCCGCTTTCGAGTC	Production of template for transcription
BstZ17I-F	CCTTCACCTGAAATGTGTGTATACAAAATCTAGGCCAGTC	Introducing PBS-M mutant into pNJ4-3 plasmid
BstZ17I-R	CTAGGTATGGTAAATGCAGTATACTTCCTGAAGTCTTTATC	Introducing PBS-M mutant into pNJ4-3 plasmid
NL4-3-PBSM-F	AGACCCTTTTAGTCAGTGTGGTTATCTCTAGCAGTGGCGCCCGAACAG	Introducing PBS-M mutant into pNJ4-3 plasmid
NL4-3-PBSM-R	CACACTGACTAAAAGGGTCTGAGTTATCTCTAGTTACCAGAGTCACAC	Introducing PBS-M mutant into pNJ4-3 plasmid
RT-E478Q-F	CTCAGTTACAAGCAATTCATCTAGCTTTGCAGGATTCG	Introducing E478Q mutant into the pRT-Dual plasmid
RT-E478Q-R	ATGAATTGCTTGTAACTGAGTCTTCTGATTTGTTGTGTC	Introducing E478Q mutant into the pRT-Dual plasmid
pRT-Dual-F	GGCAACGCCAATCAGCAACGACTGTTTGC	Introducing E478Q mutant into the pRT-Dual plasmid
pRT-Dual-R	GCTGATTGGCGTTGCCACCTCCAGTCTG	Introducing E478Q mutant into the pRT-Dual plasmid
hRU5-F2bb	GCCTCAATAAAGCTTGCCTTGA	Quantification of cDNA/ RNA integrity assay

MAL-hRU5-R3	TAGAGTGGTCTGAGGGATCT	Quantification of cDNA/ RNA integrity assay
MAL-hRU5-probe	AGAGTCACACAACAGATGGGCACACACT	Quantification of cDNA/ RNA integrity assay
Gag-F1b	CTAGAACGATTCGCAGTTAATCCT	RNA integrity assay
Gag-R1b	CTATCCTTTGATGCACACAATAGAG	RNA integrity assay
P-HUS-103 (probe)	CATCAGAAGGCTGTAGACAAATACTGGGA	RNA integrity assay

Table S2-2: Quantification of peak intensities of 147.H8 and 157.H2 in 2D NOESY spectra of complexes annealed under different conditions.

	147.H8	157.H2	147.H8/157.H2
Heat-annealed complex	1.87	1.78	105%
1NC-annealed complex	1.90	1.77	107%
12NC-annealed complex	1.10	1.88	59%

Chapter 3. The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity.

Note: This chapter is based on a research paper that was accepted to Nucleic Acids Research in April 2021.

Abstract

HIV-1 reverse transcription initiates at the primer binding site (PBS) in the viral genomic RNA (gRNA). Although the structure of the PBS-segment undergoes substantial rearrangement upon $tRNA^{Lys3}$ annealing, the proper folding of the PBS-segment during gRNA packaging is important as it ensures loading of beneficial host factors. DHX9/RNA helicase A (RHA) is recruited to gRNA to enhance the processivity of reverse transcriptase. Because the molecular details of the interactions have yet to be defined, we solved the solution structure of the PBS-segment preferentially bound by RHA. Evidence is provided that PBS-segment adopts a previously undefined adenosine-rich three-way junction structure encompassing the primer activation stem (PAS), tRNA-like element (TLE) and tRNA annealing arm. Disruption of the PBS-segment three-way junction structure diminished reverse transcription products and led to reduced viral infectivity. Because of the existence of the tRNA annealing arm, the TLE and PAS form a bent helical structure that undergoes shape-dependent recognition by RHA double-stranded RNA binding domain 1 (dsRBD1). Mutagenesis and phylogenetic analyses provide evidence for conservation of the PBS-segment three-way junction structure that is preferentially bound by RHA in support of efficient reverse transcription, the hallmark step of HIV-1 replication.

Introduction

The HIV-1 5' untranslated region (5'UTR) consists of a series of structural elements that coordinate various viral replication events. The hallmark event of reverse transcription initiates on the 5'UTR formed by annealing of $tRNA^{Lys3}$ with the complementary primer binding site (PBS) [140]. Annealing does not occur spontaneously at physiological temperature, as both the PBS-containing segment (PBS-segment) of viral RNA and $tRNA^{Lys3}$ are highly structured. Complete annealing requires mature nucleocapsid (NC) protein, which is produced after maturation of the virion proteins [126,141–143].

Several discrete regions in the PBS-segment are known to interact with viral and host factors to promote reverse transcription [35,106,107,123,144–148]. The primer activation signal (PAS) has been reported to promote tRNA priming and necessary to enhance reverse transcription efficiency [106,107,123,144] but it is under debate whether PAS directly interacts with $tRNA^{Lys3}$ during reverse transcription initiation [70]. Lysyl-tRNA synthetase (LysRS) has been reported to facilitate the annealing step by placing $tRNA^{Lys3}$ onto the PBS-segment. The process is mediated by the interactions between the anticodon binding domain of LysRS and the tRNA-like element (TLE) within the PBS-segment [35], which mimics the $tRNA^{Lys3}$ anticodon for LysRS binding and increases the efficiency of $tRNA^{Lys3}$ annealing [145–148].

DHX9/RNA helicase A (RHA) co-assembles with the HIV-1 genomic RNA (gRNA) and bolsters virion infectivity [12,149]. RHA belongs to the DExH-box superfamily and unwinds RNA duplex in the 3'- to 5'- direction. It has two double-stranded RNA binding domains (dsRBDs) at the N-terminus, followed by two RecA-like domains in the core helicase region, and an RG-rich domain on the C-terminus. Both dsRBDs are required for RHA to interact with the 5'UTR [12,149,150]. A point mutation that disrupted the structure of the PBS-segment was deficient in interacting with the N-terminus domain of RHA (dsRBD1+dsRBD2) and resulted in reduced

infectivity [151]. RHA does not re-arrange the annealed RNA complex for reverse transcription initiation, but enhances processivity of reverse transcriptase (RT) during the elongation phase by unwinding the RNA secondary and/or tertiary structures [76].

Hence, prior to $tRNA^{Lys3}$ annealing, the PBS-segment RNA serves as a scaffold to interact with viral and host factors, including aforementioned NC, $tRNA^{Lys3}$, LysRS, RHA and possibly other yet to be characterized factors [35,106,107,123,144–148]. These interactions may be mechanistically linked to achieve appropriate primer annealing for the temporal control of reverse transcription initiation. Chemical probing, enzymatic probing, computational modeling, as well as recent small angle X-ray scattering (SAXS) data supported formation of the TLE stem loop and PAS stem in the PBS-segment, connected by a long flexible loop that encompasses the $tRNA^{Lys3}$ annealing site [152–156]. However, these structural features leave gaps in knowledge of structural basis for the recruitment of some factors, including RHA. Lack of high-resolution structural information of the PBS-segment has hampered the structure-function investigation of the RNA: protein interactions necessary for efficient reverse transcription initiation.

Here we report the solution structure of the HIV-1 PBS-segment RNA and its crucial role in RNA helicase interaction necessary for efficient reverse transcription. The results characterize a previously unidentified three-way junction structure that is composed of the TLE, tRNA annealing arm and PAS. Phylogenetic analysis of HIV-1 clinical sequences documents that the secondary structure was highly conserved among various HIV-1 strains, and computational mutagenesis scanning analysis shows that the base pairings in the bottom of the TLE stem are necessary to maintain the three-way junction structure. SAXS data of PBS-segment mutant RNAs demonstrate that disruption of base pairings at the bottom of TLE stem altered RNA folding, and reduced infectivity of progeny viruses attributable to deficient recruitment of RHA to the PBS-segment, as would be expected based on previous results [151]. Combining electrophoretic

mobility shift assay (EMSA), isothermal titration calorimetry (ITC), NMR and advanced molecular docking, we show that the three-way junction structure of the PBS-segment is specifically recognized by RHA. RHA's dsRBD1 and dsRBD2-Core domains are likely to manifest the selective shape recognition mechanism. In summary, our data demonstrated that the three-way junction structure of the PBS-segment structure is conserved to support infectivity of progeny viruses.

Results

The PBS-segment RNA adopts a three-way junction structure.

To investigate the PBS-segment structure, an RNA of nucleotides (nts) 125-223 was synthesized with two terminal G-C pairs for efficient transcription (Fig. 3-1A). NMR assignments were facilitated by referencing the PBS-segment RNA spectra with that of control RNA fragments, including TLE stem loop (nt 135-177) and PAS stem (nt 125-223 with a GAGA tetraloop connecting residue U131 and G217) (Supplementary Fig. S3-1D). Attempts to make a control RNA fragment for the tRNA annealing arm were not successful. We tried various 5'- and 3'- end residues but none of the RNA fragments gave rise to similar NMR cross peak patterns in the ^{1}H-^{1}H NOESY spectrum as that of the PBS-segment RNA. Imino proton resonances of the PBS-segment were very broad (Supplementary Fig. S3-1B-C). The base pairs in the PAS stem and TLE stem control RNAs were confirmed by referencing the imino proton spectrum with that of the control RNA fragment spectra (Supplementary Fig. S3-1E-F). Additional imino proton resonances were observed and assigned to G190, G195, U200 and G206, suggesting base pairings in the tRNA annealing arm (Supplementary Fig. S3-1C). Assignments of the non-exchangeable protons of residues in these control RNA samples were used as references for the assignment of the entire PBS-segment. To overcome the spectral overlap and enhance spectral resolution, nucleotide-specific ^{2}H-labeled samples (Experimental Procedures) were prepared for NMR data collection and assignments of non-exchangeable protons. This strategy enabled us to

assign residues in and near the tRNA annealing arm that are not covered by the control RNA fragments. In addition to imino proton resonances, we also relied on the sequential and long-range NOEs that involve adenosine H2 to obtain secondary structure information within the PBS-segment [157] (Fig. 3-1A). Comparison of the non-exchangeable proton assignments with database predictions also indicate the canonical structures in PAS and TLE region, and non-canonical structures in the junction and tRNA annealing arm (Supplementary Fig. S3-3) [158].

In agreement with previous structural models, our NMR data confirm the formation of PAS stem and the TLE stem in the PBS-segment [152–156]. The assignment of residues in the tRNA annealing arm indicate that these residues are not completely flexible and unstructured, but instead forming some non-canonical structures. While no imino proton resonances were observed for residues U182-C185 and G208-G212, sequential stacking of these residues are indicated by the NOESY walk (Supplementary Fig. S3-2). Long-range NOEs were detected and assigned to the residues in the tRNA annealing arm (nt 179-214), including A209.H2-C185.H1' (Supplementary Fig. S3-2), and A198.H2-187.H1', demonstrating base pairing in the tRNA annealing arm. However, some cross-peaks, such as G183.H8-U182.H1' and A209.H2-C185.H1', are very weak (Supplementary Fig. S3-2), suggesting structural dynamics of these residues in the tRNA annealing arm. They are likely in rapid exchange between a stacked/paired position and unstructured positions. The junction that converges PAS stem, TLE hairpin and tRNA annealing arm is adenosine-rich (A132, A133 and A216), and NOE cross peaks between A216-H2 and the H1's of G178, G217 and C134 were detected in spectra with different ^{2}H labeling strategies (Fig. 3-1B). In summary, our NMR data reveal that the PBS-segment RNA adopts a three-way structure with an adenosine-rich junction (Fig. 3-1C and D).

SAXS data were collected to obtain restraints of the overall shape of the PBS-segment RNA. Gel-filtration analysis of the PBS-segment RNA showed a minor peak before the main

elution peak, indicating that a small portion of the sample may be aggregated (Fig. 3-2A). The aggregated RNAs are not likely to give rise to measurable NMR signals due to their large molecular sizes and low population (<10%), but could introduce large errors when measuring the RNA dimensions by SAXS. To obtain better homogeneity of the RNA sample, in-line SAXS data were collected by eluting the PBS-segment RNA on a Superdex-200-Increase SEC column (GE), and SAXS data collected for samples eluted in the main peak were used for structural analysis. The averaged scattering data, Kratky Plot, Guinier plot and the distribution of the interatomic distance P(r) are shown in Fig. 3-2B and 3-2E, respectively. *Ab initio* models generated from the SAXS data recapitulated a three-way junction structure highly consistent with our NMR model (Fig. 3-2F). Efforts were made to truncate one of the arms in the RNA sample to help unambiguously identify the TLE and the tRNA annealing arm in the SAXS *ab initio* model. However, NMR analysis of the truncated PBS-segment revealed that the secondary structure was altered due to the truncation. SAXS profiles of the NMR-derived structures were back calculated and compared with the experimental SAXS data. The structures with lowest energy and χ^2 values were selected for further energy optimization by MD simulation (Fig. 3-2G). The statistics of calculated structures are summarized in Tables S3-1 and S3-4.

The base pairs within the three-way junction structure are phylogenetically conserved.

To determine if the three-way junction structure of the PBS-segment is phylogenetic conserved, analysis of HIV-1 sequences deposited in the HIV database (http://www.hiv.lanl.gov/) was performed. A total of 1,908 PBS-segment containing sequences (C125-G223) were aligned using ClustalX2 [159]. There are 421 sequences in subtypes A, G and some circulating recombinant forms containing a 23-nt insertion downstream of the PBS (56,57). For the purposes of analysis, this group of sequences were excluded as they may result in an alternative folding of the PBS-segment [41,59,105,160]. Analysis of the remaining 1,487 HIV-1

sequences from various subtypes revealed that majority of the nucleotides in the PBS-segment are highly conserved, underlining the importance of PBS-segment in virus replication. The nearly 100% conservation of the 18-nt tRNA annealing residues (U182-C199) are expected, given necessity to maintain complementary to the 3'- end of human $tRNA^{Lys3}$ primer for reverse transcription initiation. The PAS stem and the lower stem of TLE are structurally conserved, with base-pair substitution of U128-A220 by a U-G pair in the PAS stem, and conversion of G137-C175 to a U-A pair (Fig. 3-3A). Maintaining the base pairing in PAS is essential for the dimeric conformation of 5'UTR and efficient reverse transcription [37,67,69], whereas the structural conservation of the lower stem of TLE remained unclear. We therefore hypothesized that the structural integrity of TLE lower stem is necessary to maintain the overall tripartite structure of the PBS-segment.

To address this hypothesis, computational mutation scanning was performed to scan the nucleotides in the PBS-segment and predict their impact on the RNA secondary structure. We applied the Vfold2D RNA structure folding model [70,148] to compute the structures for a given RNA sequence. For the WT PBS-segment RNA (HIV-1 NL4-3), the Vfold2D-predicted structure agrees with the experimental result. We performed a total of 3×103 (n=309) single nucleotide mutations, to exhaustively predict structures for each mutant, and learned that different mutations can cause different structural changes. Critical nucleotides whose mutation cause disruption of the folding were identified. Three major types of structural changes were observed: 1) in the central three-way junction region, 2) the TLE stem loop and 3) the tRNA-annealing stem. Some of mutant structures maintain the WT three-way junction global fold while some form an alternative elongated stem-loop structure. We found that the disruption of the central three-way junction can result in folding changes in both the TLE stem loop and the tRNA annealing arm. The mutations that lead to the disruption of the central three-way junction

are distributed mostly in the lower stem of TLE (U135-A140 and U172-G178). The result shows the TLE stem is stable as long as the lower stem is maintained. Mutations in the upper stem of TLE only cause local structural variations, but do not change the overall three-junction structure. Similarly, single point mutations in PAS and the tRNA annealing arm lead to only local changes without disrupting the three-way structure of the PBS-segment. In summary, the predicted low-tolerance nucleotides reside in the lower stem of TLE, while majority of the nucleotides outside of this region can be mutated (Fig. 3-3B). These results indicate that the lower stem of TLE is important to maintain the three-way junction of the PBS-segment structure. Other regions are sensitive to mutations, but single nucleotide mutations only cause local structural changes.

***The three-way junction structure is important for virion infectivity*.**

Based on the computational mutation analysis, a series of mutations were designed to 1) disrupt the TLE loop, A157G and ΔUC-loop (C151CUUUUA157 was substituted by CGAGAG), 2) perturb tRNA annealing arm, G212C/U213C and A214U, and 3) disrupt the three-way junction structure, G137C, A140C and C173G. To avoid possible disruption of tRNA annealing and reverse transcription initiation, none of the mutation sites selected are predicted to be involved in inter-molecular interactions with $tRNA^{Lys3}$ based on previously reported $tRNA^{Lys3}$: PBS complex models (7-9,57 62, 64-66). Vfold secondary structure prediction [161] suggests that the mutant RNAs in groups 1 and 2 adopt the three-way junction structure, whereas group 3 mutants, G137C, A140C and C173G, mainly adopt a long stem loop structure without a three-way junction feature (Supplementary Fig. S3-4). Consistently, SAXS data reveal that these three-way junction disrupting mutants have a longer paired distance (D_{max}), and their shapes are different compared with the WT PBS-segment (Supplementary Fig. S3-5).

We next examined the virion infectivity of these mutants in single-round infectivity assays using an NL4-3-derived reporter vector virus. The *env* gene of the reporter virus was

replaced by an EGFP open reading frame and thus EGFP was produced when the cells become transduced [162]. TZM-bl cells were incubated with vector virus samples containing 100 ng of Gag p24 and harvested 24 hrs later. The percentage of cells producing EGFP were quantified by flow cytometry to compare infectivity. Mutations altering the TLE loop (A157G and ΔUC-loop) resulted in modest reduction of the single-round infectivity (Fig. 3-3C), consistent with the previously reported role of TLE mimicking $tRNA^{Lys3}$ to facilitate $tRNA^{Lys3}$ loading for reverse transcription initiation [35]. Mutations in the tRNA annealing arm showed none (G212C/U213C), or slightly positive impact (A214U) on viral infectivity (Fig. 3-3C). These mutations are predicted to maintain the three-way junction structure (Supplementary Fig. S3-4). On the other hand, an approximate 60% infectivity reduction was observed for the three-way junction disrupted mutants (G137C, A140C and C173G) (Fig. 3-3C).

Since the PBS-segment is the reverse transcription initiation site for (-)cDNA synthesis, we then examined the reverse transcription activity of these mutants in lymphocytes. MT4 lymphocytes (5×10^5 cells in 0.5 ml of RPMI medium) were transduced with the pseudotyped virions in cell-free medium containing 200 ng of Gag p24 by spinoculation. Six hrs post-infection, cellular DNA from the transduced MT4 cells was isolated and analyzed by qPCR using primer pairs to amplify the early and late products of reverse transcription [110]. The results demonstrated that the amount of early RT product, (-)ssDNA, in MT4 cells was significantly diminished for the three-way junction disrupted mutants (G137C, A140C and C173G) ($p< 0.001$) (Fig. 3-3D). The synthesis of late RT products was similar, indicating the significant reduction in early RT activity was carried forward. These results indicate that the three-way junction disruption decreased virion infectivity by reducing the early reverse transcription activity.

To test whether the point mutations in PBS-segment affect $tRNA^{Lys3}$ annealing or RT loading, we examined $tRNA^{Lys3}$ annealing by EMSA, compared the *in vitro* primer extension

efficiencies, and quantified the tRNA placement on viral RNA in virions. The WT and mutant RNAs were annealed with $tRNA^{Lys3}$ in the presence of NC. None of the mutations disrupted formation of an RNA: RNA duplex, and the migration rates of the mutant complexes on a polyacrylamide gel were similar to the WT (Supplementary Fig. S3-6A). *In vitro* primer extension assays were carried out by incubating the annealed RNA template with RT/dNTP, and monitoring the synthesis of (-)ssDNA. No difference in the primer extension efficiency between the WT and mutants was observed (Fig. S3-6 B-E). We then investigated $tRNA^{Lys3}$ placement on WT and mutant RNAs using RNA extracted from viral particles produced from transfected 293FT cells. The RNA was mixed with an RNase H activity reduced RT (SuperScript II, Invitrogen) and dNTPs. If the mutations in the PBS-segment do not affect $tRNA^{Lys3}$ placement on gRNA, the co-purified $tRNA^{Lys3}$ will serve as a primer to generate (-)cDNA when supplemented with RT [163]. To normalize the (-)cDNA generated in WT and mutant viral RNAs, the input gRNAs were measured by RT-qPCR with primers/probe targeting the gag region (Supplementary Fig. S3-6F). All of the mutant viral RNA successfully generated (-)cDNA at a similar level as the WT (Supplementary Fig. S3-6G). The amount of proviral DNA in each sample was quantified by qPCR to ensure minimal DNA contamination (<1% RNA copy numbers). Collectively, our data show that the three-way junction disrupting mutations did not affect $tRNA^{Lys3}$ annealing and reverse transcription under *in vitro* conditions. Their negative impact on viral replication is likely to happen prior to tRNA annealing.

RHA preferentially binds to the PBS-segment of the HIV-1 5'UTR.

We have previously reported that HIV-1 recruits host RHA during virus assembly to serve as a processivity enhancement factor for RT [76]. The recruitment is mediated by protein: RNA interactions that involve the PBS-segment of HIV-1 5'UTR [151]. Therefore, the reduced infectivity and decreased reverse transcription products of the mutant viruses (Fig. 3-3 C and D)

are likely due to the failure of RHA loading onto the PBS-segment. Our previous biophysical assays show that the N-terminal domain of RHA (dsRBD1 + dsRBD2) preferentially bind to the PBS-segment RNA [151]. To confirm that PBS-segment is the binding target of RHA, recombinant full-length RHA protein (Fig. 3-4A) expressed in a baculovirus system was purified (Fig. 3-4C). The recombinant RHA was titrated to a series of HIV-1 RNA constructs, including the 5'UTR RNA (nt 1-344), 5'UTR with TAR and PolyA truncated (5'UTR$^{\Delta TAR\text{-}PolyA}$) and with PBS-segment deleted (5'UTR$^{\Delta PBS}$). These 5'UTR constructs have the 3'-half of the AUG hairpin truncated to favor RNA dimerization (Fig. 4B) [37,43,157]. As expected, the EMSA results show that the relative affinity of RHA for 5'UTR$^{\Delta TAR\text{-}PolyA}$ is similar to 5'-UTR, but much weaker for 5'UTR$^{\Delta PBS}$ (Fig. 3-4D), as at RHA: RNA ratios of 7:1 and 8:1 (lanes 8 and 9) almost no free RNA bands were visible in the 5'UTR and 5'UTR$^{\Delta TAR\text{-}PolyA}$ samples, but significant amount of free RNA bands were observed in the 5'UTR$^{\Delta PBS}$ samples. When comparing the RHA binding affinity for smaller RNA fragments, TAR-PolyA (nt 1-104, 33 kDa) and PBS-segment (33 kDa), RHA preference towards the PBS-segment was observed (Fig. 3-4E). Together, these data demonstrate that RHA preferentially binds to PBS-segment within the 5'-UTR.

The binding interface between dsRBD1 and PBS-segment was mapped by NMR.

To map the RNA-binding residues on dsRBD1, ^{15}N-labeled dsRBD1 was titrated with PBS-segment in a series of ^{1}H-^{15}N-TROSY NMR experiments. Due to the severe precipitation issues when mixing PBS-segment RNA with dsRBD1 at NMR concentrations (200 μM), the titration was carried out at a high salt condition (500 mM KCl and 1 mM $MgCl_2$) to reduce non-specific electrostatic interactions between protein and RNA. As a result, the affinity between PBS-segment and dsRBD1 was weakened and thus fast exchange regime was observed (Fig. 3-5A). Chemical shift perturbations (CSPs) were observed for a group of residues in α1 helix (region 1), the loop connecting β1 – β2 (region 2) and α2 helix (region 3), suggesting these residues are

within or close to the binding interface (Fig. 3-5D and Supplementary Fig. S3-7). These residues are consistent with the RNA binding interface reported in the crystal structure of RHA dsRBD1 in complex with a non-specific target $(GC)_{10}$ RNA (PDB: 3vyy) (69). To identify the RNA residues within or close to the dsRBD1 binding interface, 2D ^{1}H-^{1}H NOESY spectra of the PBS-segment in complex with dsRBD1 was collected. In addition to fully protonated RNA, site-specific deuterated samples, including AG-PBS-segment and $A^{2R}C^{R}U^{R}$-PBS-segment RNAs, were used to simplify the spectra for unambiguous assignment. G129-H1' was shifted upon dsRBD1 titration (Fig. 3-5B), as we previously reported [151]. Chemical shifts of residues in the lower TLE stem region also exhibited moderate shifts upon dsRBD1 binding, including A140, U141, C142, and U172 (Fig. 3-5C). The protein and RNA residues identified within and close to the binding interface by NMR titrations are labeled in Fig. 3-5D. Both the TLE and the PAS stems are involved in dsRBD1 binding, and the distance/dimension matches the RNA binding interface in dsRBD1 (Fig. 3-5D).

With the guidance of the experimental data, a data-driven molecular docking study was exploited to model the structure of the PBS-segment: dsRBD1 complex. The crystal structure of RHA dsRBD1 (PDB: 3vyy) was used to dock the PBS-segment structure using an in-house program MDockPP [164]. Protein residues that are in the RNA binding interface of the reported crystal structure (PBD: 3vyy) and showed large CSP ($\Delta\delta > 0.05$ ppm), including F7, G13, K14, K16, M17, T18, Y21, K29, K54, K55 and Q58, were selected to generate distance restraints. The RNA residues (G129, A140, U141, C142 and U172) were also defined to be close to dsRBD1. The distance between protein and RNA was set to 7 Å because this is the shortest distance restraint that resulted in docking models without steric clashes. The lowest energy model (Fig. 3-6B) shows that region 1 (α1 helix) of dsRBD1 interacts with the minor groove of the bottom stem of TLE, region 2 is in close proximity to the PAS stem, and K54 and K55 in region 3 are near RNA

residues in or adjacent to the three-way junction (A132, A133, and U176). The TLE: regions 1 and PAS: region 2 interactions are supported by both protein and RNA NMR data (Fig. 3-5). The NMR signals for the junction residues A132, A133, and U176 were broadened and became undetectable upon dsRBD1 titration in 2D NOESY spectra, which could be caused by protein: RNA interactions. Overall, the docking model is consistent with the NMR data, as majority of the residues near or within the protein: RNA interface exhibited CSP in the NMR titrations (Supplementary Fig. S3-7).

The docking model suggests a shape recognition of the PBS-segment by dsRBD1. The tRNA-annealing arm introduces a ~108° bend of the PAS away from co-axial stacking on TLE stem, resulting in ~15 Å departure of the G129 from the straight A-form helical RNA structure model. In the crystal structure of dsRBD1: $(GC)_{10}$ complex, in order to have all the three regions to interact with RNA, region 1 is slightly distant from the RNA so that region 2 could be close. In the case of the PBS-segment, due to the angle between the TLE lower stem and PAS, dsRBD1 is closer to RNA with regions 1-3 all in contact with RNA. This leads to formation of additional hydrogen bonds between dsRBD1 and PBS-segment as compared to a straight $(GC)_{10}$ RNA shown in the crystal structure (Fig. 3-6A-B).

To validate the shape-dependent recognition predicted by the docking model, ITC experiments were performed to compare the titration isotherms of dsRBD1 into PBS-segment and a truncated tRNA annealing arm hairpin-control RNA. The hairpin-control RNA contains base pairings in the PAS stem (nt 125-131 and 217-223) and TLE lower stem (nt 134-141 and 171-178), and is enclosed by a CG pair and a UUCG tetraloop (Fig. 3-6C). The isothermal data of titrating dsRBD1 into the PBS-segment RNA were fit using two-sites binding modes, and data fitting show that the PBS-segment contains one high affinity dsRBD1 binding site (K_{d1} = 1.17 ± 0.23 μM) and a second set of weak binding sites (K_{d2} = 45 ± 8.1) (Fig. 3-6E). However, only weak

dsRBD1 binding sites was observed in the 44-nt hairpin-control RNA (Fig. 3-6D). In order to characterize these weak interactions, both protein and RNA concentrations in ITC titrations were increased to obtain reliable quantification of binding thermodynamic parameters (Experimental Procedures). One-site non-linear least square regression of the integrated heats reveals that the hairpin-control RNA can interact with N= 5.2 ± 0.2, dsRBD1 at K_d = 18.5 ± 1.1 μM, indicative of non-specific electrostatic interactions. The K_d is comparable to the second set of dsRBD1: PBS-segment binding event, suggesting that after the high affinity binding site on PBS-segment is occupied, dsRBD1 starts to non-specifically interact with the RNA. It also explains the ^{1}H-^{15}N TROSY NMR titration experiments. They were performed at high salt conditions to eliminate non-specific dsRBD1: RNA interactions (Fig. 3-5A). Assuming no cooperativity among these dsRBD1 nonspecific binding sites, the averaged thermodynamic contribution of each dsRBD1 binding site on the hairpin-control were calculated (Supplementary Table S3-5). Both PBS-segment: dsRBD1 and hairpin-control: dsRBD1 interactions are enthalpy ($\triangle H$)-driven, but the enthalpy change associated with PBS-segment binding is significantly greater than hairpin-control binding (-7845 ± 123 cal/mol per site in the PBS-segment vs. -1762 ± 37 cal/mol per site in the hairpin-control) (Supplementary Table S3-5). For a low affinity titration system, even though the binding event is near but not completely saturated at the end of titrations, K_d can be accurately determined but $\triangle H$ can be underestimated [165]. The second set of binding sites in PBS-segment is not saturated because the heat in the final injections are not near zero, and thus the $\triangle H$ for these binding sites are less negative than the dsRBD1: hairpin-control interaction (Supplementary Table S3-5). In summary, the ITC data are in agreement with the docking model that more hydrogen bonds are formed when dsRBD1 binds to a specific target RNA (PBS-segment) than a non-specific target ($(GC)_{10}$ or hairpin-control). Recognition of PBS-segment by dsRBD1 is RNA shape dependent.

Molecular docking of dsRBD2-Core onto PBS-segment suggests possible specific RHA: RNA interactions.

Previous studies suggested that dsRBD2 is indispensable in directing RHA binding to target RNAs [150,166,167]. Crystal structure of dsRBD2 in complex with a $(GC)_{10}$ RNA reveals that it binds to RNA similarly as dsRBD1, with equivalent α1 helix (region 1) and loop connecting β1- β2 (region 2) binding to two successive minor grooves and α2 helix (region 3) binding to the major groove in between (Supplementary Fig. S3-8A) [166]. Crystal structure of a homologous protein MLE reveals that its dsRBD2 is tightly associated with the helicase core domain. While its region 1 and region 3 are surface exposed and could interact with an RNA target, region 2 of dsRBD2 is surrounded by residues in RecA2 domain and is not surface exposed (Supplementary Fig. S3-8B) [168]. RHA and MLE share 51% sequence identity and 69% sequence similarity, so it is very likely that dsRBD2 of RHA is also in close contact with its RecA2 domain. Thus, *in vitro* study of dsRBD2 alone in RNA interactions may not represent real interactions with RNA in the context of full-length RHA, as the RecA2 domain of RHA may create spatial hindrance to prevent dsRBD2 region 2 residues from binding to a target RNA. To avoid this problem, we modeled the structure of RHA dsRBD2-Core (Fig. 3-4A, residue 169-1150) using homology modeling based on the MLE structure (PDB: 5aor) (Supplementary Fig. S3-8C), and docked it onto the PBS-segment RNA. The docked models were filtered using criteria that dsRBD2 regions 1 and 3 participate in RNA binding, and the dsRBD2-Core binding site does not overlap with the previously determined dsRBD1 binding site. The lowest energy model shows that the α1 helix (region 1) and K236 (region 3) of dsRBD2 interact with the three-way junction of the PBS-segment, and RecA2 residues in the Core domain make contacts with the tRNA annealing arm (Supplementary Fig. S3-8D-E). Region 2 of dsRBD2 is not involved in PBS-segment RNA binding, as it is surrounded by residues in RecA2. It is consistent with a previous study which reported substitution of a region

2 residue in dsRBD2 did not affect RHA-mediated RISC assembly [166]. In summary, the docking model suggests that the dsRBD2-Core could also contribute to the recognition of the three-way junction structure of the PBS-segment.

Discussion

The three-way junction structure of the PBS-segment is important for infectivity.

Solving the structure of the PBS-segment has been challenging as the tRNA annealing arm is highly metastable thus making crystallization difficult and traditional characterization by NMR impractical. Metastable residues in the tRNA annealing arm were proposed to be unstructured by chemical probing, enzymatic probing, SAXS and computational modeling [152–156]. In our study, we established the PBS-segment sequence boundaries encompassing the three-way junction in the HIV-1 RNA encapsidation signal delineated by prior NMR studies, in which residues C125-G130 form base pairs with C218-G223 [37,157]. The structure reported for the PBS-segment here should resemble its parent structure in the dimeric 5'UTR prior to reverse transcription initiation that directs viral RNA genome packaging [37,169]. By combining site-specific deuteration NMR strategies with SAXS methodology, EMSA with full length RHA and biological validation experiments, we were able to determine one conformation of the PBS-segment important for viral infectivity.

The identified three-way junction structure is reinforced by phylogenetic analysis of HIV sequences in patient samples, which document conserved pairings at the bottom of TLE (Fig. 3-3A). In line with the phylogenetic findings, we show that single point mutations in this region could completely disrupt the three-way junction (Fig. 3-3B). The mutants adopt relatively extended structures (Supplementary Fig. S3-4 and S3-5) that failed to efficiently produce reverse transcription products in infected cells (Fig. 3-3C and 3-3D). Our structural study of the PBS-segment implements the current knowledge of the structure of the 5'UTR of the gRNA and

provide structural basis for the recruitment of beneficial host factors during genome packaging to bolster virion infectivity.

The tRNA annealing arm is important for RHA: PBS-segment interaction

Annealing of $tRNA^{Lys3}$ onto PBS requires a chaperone, which can be the NC domain of Gag polyprotein or the processed mature NC protein [142,143,170,171]. *In vitro* primer extension on RNA extracted from protease-deficient virions exhibited lower nucleotide incorporation rate than WT, but could be rescued by incubating the RNA with mature NC protein, suggesting Gag facilitates a partial annealing of tRNA onto PBS and NC promotes a complete annealing [143]. The two-step annealing model is supported by the *in virio* SHAPE analysis of the HIV-1 viral RNA in WT and protease-deficient virions [126]. The matrix (MA) domain was reported to modulate the chaperone activity of Gag, which is negatively regulated by MA: RNA interactions and can be stimulated by MA interacting with inositol phosphate on the plasma membrane [172]. Hence tRNA annealing promoted by Gag is likely to happen at the plasma membrane where virus assembly takes place. RHA is predominately localized in nucleus and shuttles between nucleus and cytoplasm [173]. The interactions between RHA and PBS-segment may occur as early as in nucleus. It is possible that RHA is brought by the 5'UTR of gRNA to the virus assembly site, where Gag interacts with inositol phosphate and promotes tRNA annealing. The spatial and temporal control of these molecular interactions need to be investigated in future studies.

The PBS-segment structure is composed of an adenosine-rich three-way junction at the confluence of three subdomains: PAS, TLE and the tRNA annealing arm (Fig. 3-1). It is worth mentioning that a similar SAXS envelope was reported for an almost identical PBS-segment RNA from the NL4-3 isolate (nt 125-223 with three non-native G-C pairs at the 5'/3' termini) [155], but the residues in the tRNA annealing arm were single stranded and unstructured in the

reported SAXS-derived model. In our studies, base pairings and base stackings in the tRNA annealing arm were apparent in the NMR data. Sequential NOESY walk of residues in the tRNA annealing arm was observed in the NMR spectra, but the inter-residue NOE peak intensities were weak (Supplementary Fig. S3-2), suggesting that residues U182-C185 and G208-G212 are exchanging between flexible/unpaired and base stacked structures. The structural flexibility is believed to be favored for reverse transcription as the secondary structure must be unwound for tRNA annealing. The tRNA annealing arm in the *ab initio* model generated from SAXS data appeals wider than a typical A-form dsRNA (Fig. 3-2G). Both the NMR and SAXS data are in agreement with an intrinsically dynamic domain in which several residues in tRNA annealing arm remain accessible for chaperone-mediated tRNA annealing.

The importance of the TLE stem loop and the PAS stem facilitating tRNA loading and activating reverse transcription have been previously addressed [35,67,106,123]. Unlike PAS and TLE, residues in the tRNA annealing arm are less phylogenetically conserved (Fig. 3-3A). Many mutations in the tRNA annealing arm are not predicted to disrupt the three-way junction structure of the PBS-segment (Fig. 3-3B). Our study suggests that the tRNA annealing arm contributes to the overall shape of the PBS-segment RNA recognized by RHA. The bent TLE-PAS stem is preferred by dsRBD1 over the hairpin-control, a straight dsRNA helical structure which can be considered as TLE stacking on PAS (Fig. 3-6C-E). TAR-PolyA is another example of straight helical structure as PolyA co-axially stacks on TAR [41,155]. EMSA data show the affinity of dsRBD1 for TAR-PolyA is much weaker than PBS-segment (Fig. 3-4E). Thus, even though the tRNA annealing arm does not directly interact with dsRBD1, it participates in the shape-dependent dsRBD1 recognition by preventing TLE from co-axially stacking on PAS. The affinity differences indicate two RNA-binding modes of RHA: the non-specific interactions with dsRNA

when RHA functions as a helicase, and the specific interactions with PBS-segment to be recruited by HIV-1 as a beneficial factor.

RHA dsRBD1 shape-dependent recognition contributes to loading RHA to the PBS-segment

DsRBDs exist in many proteins and influence various steps of RNA metabolism. Some of the proteins have more than one dsRBD, and the contribution of each dsRBD to target RNA selection is usually not equal. MLE helicase facilitates the incorporation of the roX lncRNA into the dosage compensation complex via its dsRBDs interacting with the roX lncRNA. Studies revealed that dsRBD2 has 10-fold greater affinity than dsRBD1, and dsRBD2 plays the major role in recognition of the R2H1 and SL7 stem loops of roX lncRNA [174,175]. Both the R2H1 and SL7 stem loops contain Watson-Crick and G: U wobble base pairs that form straight A-form helices. The crystal structure of dsRBD1+dsRBD2 in complex with R2H1 shows that MLE dsRBD1 binds to RNA (PDB: 5ztm) in a similar way as RHA dsRBD1 binds to a $(GC)_{10}$ dsRNA (PDB: 3vyy). These dsRBD1: dsRNA crystal structures indicate that dsRBD1 binds to straight A-form dsRNA helices without sequence specificity.

Efforts have been made to investigate the sequence and shape recognition of dsRBDs in directing the proteins to their target RNAs. In general, dsRBDs use conserved residues in region 1, 2 and 3 to interact with two consecutive minor grooves in dsRNAs [174,176–181]. Sequence specificity is achieved by some species-specific amino acids interacting with non-canonical structures in RNA. For example, dsRBD of Rnt1p, a member of RNase III family of dsRNA endonucleases, binds to the sn47 precursor RNA to direct the endonuclease to its RNA substrates by recognition of the AGNN tetraloop of the target RNAs. NMR studies show that in addition to using regions 1-3 interacting with two consecutive minor grooves in sn47 RNA, S376 and R372 in α1 form hydrogen bonds with the 2'OH of the two non-conserved 3'-nucleotides of the tetraloop, and M368 is stacked on the ribose of the adenosine in the tetraloop [182]. The

interactions between the dsRBM (double-stranded RNA binding domains, an alternative name of dsRBD) of adenosine deaminases that act on RNA (ADARs) and its target GluR-2 RNA provide another example of specific recognition. ADARs are a group of enzymes that selectively deaminate adenosine. Solution studies of dsRBMs of ADAR2 in complex with GluR-2 RNA show that the dsRBM1 of ADAR2 binds to the top stem of the GluR-2 RNA with specific contacts between α1 of dsRBM1 and RNA tetraloop. The interactions include M84 making a sequence-specific contact with the A-U pair that is adjacent to the UCCG tetraloop, and E88 forming a hydrogen bond with the amino group of the first C in the tetraloop [183]. Therefore, in many cases the α1 residues in dsRBDs confer sequence specific and/or shape-dependent recognition of their target RNAs.

Our study shows that dsRBD1 recognizes the shape of the three-way junction structure of the PBS-segment. ITC data demonstrate that dsRBD1: PBS-segment interaction is of higher affinity as compared to binding to a straight A-form helical RNA structure (hairpin-control, Fig 3-6C-E). K5, N6, Y9, K54, K55, K29 and N30 are involved in H-bonds with RNA in both the crystal structure of dsRBD1: $(GC)_{10}$ complex and our dsRBD1: PBS-segment model. In addition, K16, T18, Y21, K29, N23 and E41 participate in H-bonds with PBS-segment, which explains the K_d and enthalpy differences between dsRBD1 binding to a straight helical RNA and PBS-segment. Some of these residues are unique in the dsRBD family [184], and therefore they are likely to contribute to the shape-dependent recognition by dsRBD1.

In summary, we show that the structure of the PBS-segment prior to tRNA annealing folds into a three-way junction structure that is important for viral infectivity. The biophysical studies of the interactions between RHA dsRBD1 and the PBS-segment provides a plausible explanation for the recruitment of a beneficial host factor in the late stage of virus replication. Although we cannot exclude the possibility of other yet-to-be discovered molecular interactions

involving PBS-segment, our solution structure of the PBS-segment provides the structural basis for further investigation of the biomolecular interactions that occur on the scaffold of the PBS-segment.

Author Contributions

For this paper I analyzed the conservation of the HIV-1 PBS segment and characterized RHA binding to the MAL 5'UTR using EMSA. Some of these results were included in the published manuscript and results not included in the manuscript are summarized in the unpublished results section of this chapter. Zhenwei Song collected all NMR and SAXS data and did structure calculation. The SAXS data collection and analysis were performed at beamline 12-ID-B of Advanced Photon Source at Argonne National Laboratory with Dr. Xiaobin Zuo. The final structure optimization was achieved through collaboration with Dr. Shi-Jie Chen lab (Yi Cheng and Jun Li). The infectivity data were collected by Juan Ji and Margaret Lange, and RT intermediates were measured through collaboration with Dr. Kathleen Boris-Lawrie lab (Gatikrushna Singh). The NMR and ITC titration data to characterize dsRBD and PBS-segment interaction were collected by Zhenwei Song, and the docking models were generated through collaboration with Dr. Xiaoqin Zou lab (Liming Qiu).

Experimental Procedures

Plasmids

The PBS-segment, 5'UTR, 5'UTR$^{\Delta TAR\text{-}PolyA}$ and 5'UTR$^{\Delta PBS}$ plasmids used for RNA *in vitro* transcription were constructed as previously described [151]. The PBS-segment mutant template plasmids for RNA transcription were generated by site-directed mutagenesis. The vector pNL4-3-CMV-EGFP used for infectivity measurement had deletion of *vif, vpr, vpu, nef and env* open reading frames (ORF), and contained a GFP ORF driven by a CMV promoter [162]. The PBS mutations in pNL4-3-CMV-EGFP for infectivity assays were generated by multiple DNA

fragment assembly using primers listed in Supplementary Tables S3-1 and S3-2. (GenBuilder Cloning Kit, GenScript). All of the sequences were confirmed by Sanger sequencing (DNA Core, University of Missouri).

RNA in vitro transcription

All of the RNA constructs, including 5'UTR, 5'UTR$^{\Delta TAR\text{-}PolyA}$, 5'UTR$^{\Delta PBS}$, TAR-PolyA, PBS-segment and mutants, TLE, PAS stem, hairpin-control and tRNA^{Lys3}, were made by *in vitro* T7 transcription reactions. The DNA templates of 5'UTR, 5'UTR$^{\Delta TAR\text{-}PolyA}$, 5'UTR$^{\Delta PBS}$, TAR-PolyA, PBS-segment and mutants were made by PCR amplifying corresponding region of the plasmids with a 5'- Top17 sequence. The template DNAs of TLE, PAS stem, hairpin-control and tRNA^{Lys3} were custom synthesized (IDT). The nucleotide specific ^{2}H labeled samples, including A^{2R}GU^{R}-, A^{2R}C^{R}U-, A^{2R}C^{R}U^{R}-, A^{2R}G^{R}-, AC- and AG-PBS-segment (nomenclature here: the letters denote the nucleosides containing ^{1}H which is visible by NMR, R=ribose, A^{2R} means H2 and ribose hydrogens are protonated; A^{2R}GU^{R}-PBS-segment means the H2 and ribose of adenosine, ribose of uracil and guanosine are protonated, and cytidine is fully deuterated) were prepared by incorporating corresponding deuterated and protonated rNTPs in T7 transcriptions as previously described [151,185]. Fully deuterated rNTPs and H5, H6-deuterated CTP and UTP were purchased from Silantes (Silantes GmbH, Munich) and Cambridge Isotope Laboratories (CIL, Andover, MA). H8-deuterated ATP and GTP were prepared in the lab [151]. ^{13}C/^{15}N labeled RNAs were synthesized by incorporating ^{13}C/^{15}N labeled rNTPs (CIL, Andover, MA) in T7 transcription.

RNA refolding

The PBS-segment RNA used for biophysical studies was refolded by preparing the RNA in 10 mM Tris-HCl, pH 7.5, incubating at 95 °C for 3 min, snap cooling on ice, and then mixing with salts to reach final salt concentrations of binding buffer A (10 mM Tris-HCl, pH 7.5, 140 mM KCl,

10 mM KCl, and 1 mM $MgCl_2$). The RNA samples were then incubated at 37 °C for 30 min prior to EMSA and ITC experiments.

Protein purification

The plasmid for recombinant RHA expression was a kind gift from Dr. William Clay Brown (University of Michigan). The recombinant RHA contained N-terminal His_6- and Mocr-tags and was expressed in insect cells. Purification of RHA was performed as previously described (20) and the protein was stored in storage buffer (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM β-mercaptoethanol and 10% glycerol). The recombinant N-terminally His_6-tagged dsRBD1 was expressed in E.coli BL21(DE3)-pLysS cells (Invitrogen) treated with 0.3 mM IPTG for 4 hrs at 37 °C. To express ^{15}N labeled dsRBD1, the bacteria were grown in the LeMaster and Richards minimal medium with $^{15}NH_4Cl$ (0.5 g/L) as the sole nitrogen source. The harvested cells were resuspended in lysis buffer (40 mM NaH_2PO_4, 300 mM NaCl, pH 7.5) with lysozyme, sonicated and centrifuged. The supernatant was applied to a Cobalt column (HisPur Cobalt Resin, Thermo Scientific), and the protein was eluted with lysis buffer containing additional 150 mM imidazole. To remove the His_6-tag, the protein was digested overnight at 4 °C with TEV protease (1 mg protease per 20 mg of His_6-dsRBD1) and dialyzed in low-salt buffer (30 mM Tris, 100 mM NaCl, 2 mM EDTA, 5 mM β-mercaptoethanol). The protein was then subjected to ion exchange and size exclusion chromatography, and stored in storage buffer at -80 °C. The plasmid for recombinant NC (HIV-1 NL4-3 strain) expression was a kind gift from Dr. Michael F. Summers (University of Maryland, Baltimore County). The NC protein was expressed and purified as described elsewhere [137]. The pRT-Dual plasmid [135] for recombinant RT expression (HIV-1 HXB2 strain) was a kind gift from Dr. Donald Burke (University of Missouri, Columbia). The RT protein was expressed and purified as described previously [160].

Cells and Viruses

293FT cells (Invitrogen) were propagated in Dulbecco's modified Eagle's medium (DMEM, Sigma) supplemented with 10% fetal bovine serum (FBS). Wildtype and mutant viruses were produced by transfection as previously described [162]. Briefly, 293FT cells were plated in 10 % FBS DMEM at density of 5×10^5 cells per well in 6-well plates overnight and then co-transfected with 500 ng of pNL4-3-CMV-EGFP or mutants together with 100ng pMD-G (AIDS Reagent Program), which encodes the vesicular stomatitis virus glycoprotein, VSV-G, for pseudotyping. The mixtures were added dropwise to the cells and incubated for 6 hrs in 5 % CO_2 at 37 ^oC, rinsed, and cultured in 2 ml of 10% FBS DMEM. The supernatant was harvested 48 hrs post-transfection by centrifugation at 1000 rpm for 10 min at 4^oC, and the vector virus preparations passed though 0.45 μm filters. The transfected cells were collected and fixed with 4% of paraformaldehyde, and transfection efficiency was monitored by detection of EGFP on an Accuri C6 Flow Cytometer (BD Biosciences, San Jose, CA). The Gag p24 was quantified by ELISA using pre-coated MicroFluere plates (Optofluidic Bioassay, MI).

Infectivity assay

Equivalent amounts of VSV-G-pseudotyped vectors viruses (15 ng of Gag p24) with or without mutations in the PBS-segment were used to transduce 2×10^5 of TZM-bl cells in 12-well plates. The infected cells were collected and fixed with 4% paraformaldehyde 24 hrs post-transduction, and the percentage of cells expressing EGPF was measured using the BD Accuri C6 Flow Cytometer (BD Biosciences, San Jose, CA). The infectivity was calculated as the percentage of EGFP-expressing cells in mutant virus infected cells compared to that of the wild type (WT).

Quantification of reverse transcription products in infected cells

MT4 lymphocytes (5×10^5 cells in 0.5 ml of RPMI medium) were seeded in 12-well plates and incubated with equivalent cell-free medium containing vector viruses normalized to 200 ng of Gag p24 and spinoculated for 2 hrs, washed twice in RPMI medium and cultured for 4 hrs. The

inactivated virus control sample was heated to 100 °C for 5 min, cooled to room temperature, centrifuged at 12,000 rpm for 1 min, and treated with DNase I for 1 hr. The infected cells were collected and washed twice with 1× phosphate buffered saline (PBS), and DNA was extracted in the QIAamp DNA blood mini kit (Qiagen). Cellular DNA (600 ng) was evaluated by qPCR with specific primer sets that detect early and late products of HIV-1 reverse transcription (Supplementary Table. S3-3) [110]. The abundance of copies was determined relative to standard curves generated on gene-specific PCR amplicons compared to mock and heat-inactivated virus infection.

In vitro tRNA annealing, primer extension and tRNA placement assays

The annealing reaction was initiated by mixing 20 pmol PBS-segment, 20 pmol $tRNA^{Lys3}$, and 600 pmol NC (~ 6 nucleotides per NC molecule) in 20 µl of buffer B (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl and 5 mM $MgCl_2$). The mixture was incubated at 37°C for 30 min, and the reaction was terminated by incubating with 0.5% SDS and 0.2 mg/ml Proteinase K at 37°C for 30 min. The RNA was purified by phenol-chloroform extraction and ran on 10% polyacrylamide gels.

Primer extension assays were performed with 5'-Cy3-labeled $tRNA^{Lys3}$ annealed on PBS-segment and mutants in the presence of NC. To label $tRNA^{Lys3}$ on the 5'-terminus, GMP-primed $tRNA^{Lys3}$ was synthesized by mixing GMP with rNTPs in T7 transcription, and the RNA was fluorophore labeled with a Cy3-NHS ester (Lumiprobe) [76,186]. The primer extension reaction was initiated by mixing 0.33 mM dNTP with a solution containing 0.1 µM template and 0.4 µM RT, incubated in 37 °C for 0, 0.5, 1.0, 1.5, 2.0, 3.0 and 4.0 min, and quenched by 25 mM EDTA. The proteins were removed from the reactions by phenol-chloroform extraction, and the primer extension products were resolved on 10% denaturing polyacrylamide gels.

To check the tRNA placement on the WT and mutant viral RNAs, 200 µl of WT and mutant pseudotyped viruses were harvest from transfected 293FT cells and treated with Dnase I and DpnI at 37°C for 90 min. The viral RNA was extracted with TRIzol (Invitrogen). To check if the viral RNA was pre-annealed with $tRNA^{Lys3}$ to prime reverse transcription, 2 ng of extracted RNA was mixed with dNTP (10 mM each), 5X First-Strand Buffer (Invitrogen), 0.1 M DTT and RNaseOUT™ (40 units/µL), and the mixture was incubated at 42°C for 2 min. Primer extension was initiated by adding 1 µl (200 units) of SuperScript™ II RT (Invitrogen), and the reaction was incubated at 42°C for 50 min. To quantify the (-)cDNA products, 2 µl of cDNA generated from the previous step was mixed with the primer/probe set targeting the RU5 region (Table S3-3) for qPCR (Luna® Universal Probe qPCR Master Mix, NEB). To quantify the total gRNA, the primer/probe set (Table S3-3) targeting gag region was used in RT-qPCR (Luna® Universal Probe One-Step RT-qPCR, NEB). To detect if plasmid DNA were carried over through RNA extraction, the same amount of RNA (0.2 ng) and primer/probe set targeting the gag region were used in qPCR. Both qPCR and RT-qPCR were performed on CFX96 Real-Time PCR Detection System (Bio-Rad).

NMR spectroscopy

The PBS-segment and its control fragment samples were prepared (200 – 300 µM) and pre-incubated in 10 mM Tris-d_{11}, pD=7.5 and 1 mM $MgCl_2$ at 37°C for 30 min. Two-dimensional ^1H-^1H NOESY and ^1H-^1H TOCSY data for control RNA fragments, and ^1H-^1H NOESY data for fully protonated and site-specific deuterated PBS-segment RNA and in complex with dsRBD1 were collected in D_2O (99.96%, CIL) at 308K. ^1H-^{13}C HMQC data were collected for $^{13}C/^{15}N$ labeled control RNA fragments. One- and two-dimensional imino proton spectra were collected for the PBS-segment RNA and control RNAs (200-300 µM) prepared in buffer containing 10% D_2O + 90% H_2O. HNNCOSY data were collected for $^{13}C/^{15}N$ labeled TLE stem loop RNA. ^1H-^{15}N TROSY data

were collected for 200 µM dsRBD1 titrated with 0, 54 µM, 108 µM, 136 µM and 160 µM of the PBS-segment RNA in binding buffer B (10 mM Tris-d_{11}, pD=7.5, 500 mM KCl and 1 mM $MgCl_2$) in 10% D_2O and 90% H_2O at 318 K. All the NMR data were collected on a Bruker Avance III 800 MHz spectrometer equipped with TCI cryoprobe (NMR Core, University of Missouri). The NMR data were processed by NMRPipe [187] and analyzed by NMRViewJ [188].

SAXS data collection and processing

SAXS data of the PBS-segment RNA were collected at Sector 18 of the Advanced Photon Source, Argonne National Laboratory. In-line data were collected by eluting 9 mg/ml PBS-segment on a Superdex-200-Increase SEC column (GE) at 0.75 ml/min flowing rate. SAXS data of the mutant RNA samples were collected at beamline 12-ID-B of Advanced Photon Source at Argonne National Laboratory, using an in-line AKTA micro FPLC setup with a Superdex 75 Increase 5/150 GL size exclusion column. The wavelength, λ, of X-ray radiation was set to 0.9322 Å. Scattered X-ray intensities were measured using a Pilatus 2M detector. The sample-to-detector distance was set such that the detecting range of momentum transfer q [$=4\pi \sin\theta/\lambda$, where 2θ is the scattering angle] was 0.005-0.85 $Å^{-1}$. The sample passed through the FPLC column and was loaded to a flow cell for SAXS measurements. The flow cell is a cylindrical quartz capillary 1.5 mm in diameter and 10 µm wall thickness. The exposure time was set to 1 second to reduce radiation damage and data were collected at every other second. The 2-D scattering images were converted to 1-D SAXS (I(q) vs q) curves through azimuthally averaging after solid angle correction and then normalizing with the intensity of the transmitted X-ray beam flux, using the beamline software. The SAXS data were then processed using ATSAS suite 2.8 and 3.0 [189]. The radius of gyration (R_G) was determined by the Guinier plot in its linearity region and pairwise distribution was plotted using GNOM [190]. The *ab initio* models were generated by averaging 10 independent DAMMIF runs on DAMAVER [191].

Structure calculation.

The PBS-segment structure was initially calculated using NMR-derived restraints by CYANA [192]. Standard torsion angle restraints were used for regions of A-helical geometry in PAS stem and TLE stem loop. Standard hydrogen bonding restraints for Watson-Crick pairs in PAS stem (nt 123-131 and 217-225) and TLE stem loop (nt 135-150, 158-167 and 171-177), and tRNA annealing arm (nt 187-189 and 195-197), and standard phosphate distances in A-form RNA helical structures in these regions were employed to maintain the major groove width in an A-form helical structure (Table 3-1) [192,193]. The theoretical solution scattering curves of the lowest structures were calculated and compared with the experimental SAXS data by CRYSOL v2.8 [194]. A total of 10 structures of lowest energy and lowest χ^2 values were selected for further molecular dynamics (MD) refinement. For each of the 10 structures, 1 ns MD simulation was performed in 1 mol/L NaCl solution under 300 K and 1 bar with NMR restraints using Amber software package [195], and the structure of lowest χ^2 value ranging from 1.8 to 2.5 in the trajectory was extracted for further refinement. In the third step, Xplor-NIH software [196] was used to further refine the structures to reduce χ^2 values. For each of the 10 structures, 100 simulated annealing MD simulations were performed with NMR and SAXS restraints. After Xplor-NIH refinement, the χ^2 values of the 1000 (10×100) structures were reduced to 1.35 – 1.5. The 10 structures of the lowest energies were deposited to the RCSB protein databank (PDB: 7lva).

Computational mutagenesis scanning

Computational mutation scanning was performed using the Vfold2D RNA structure folding model [161,197–199] to predict the RNA structure with single nucleotide variation at each nucleotide position. Structure prediction of a total of 309 mutants containing single point

mutation in the PBS-segment was carried out. Models of lowest energy were analyzed and grouped based on the structural changes.

EMSA

EMSA was carried out by preparing mixture solutions of 0.5 μM RNA (5'UTR, 5'UTR$^{\Delta TAR\text{-}PolyA}$ or 5'UTR$^{\Delta PBS}$) and recombinant RHA at concentrations of 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4 μM in binding buffer A. To compare RHA binding to PBS-segment and TAR-PolyA, 0.5 μM RNA was incubated with 0, 1, 2, 3 and 4 μM of RHA. The mixtures were incubated at 37°C for 10 min and resolved on 1.5% agarose gels. The gels were imaged using Gel Doc XR+ (Bio-Rad). EMSA experiments were repeated three times and the representative gel is shown.

ITC

The ITC experiments were carried out by titrating 610 μM of dsRBD1 into 17 μM PBS-segment and titrating 1.15 mM of dsRBD1 into 18 μM hairpin-control RNA at 30°C on a VP-ITC (MicroCal, GE Healthcare). Heat of dilution titrations were performed by titrating the dsRBD1 protein into a matching buffer in the same experimental settings. Both protein and RNA samples were prepared in binding buffer B. The baseline was corrected by subtracting the heat of dilution, and the data were fitted using "one-site" non-linear least square regression for hairpin-control and "two-site" non-linear least square regression for PBS-segment. The statistics of thermodynamic parameters (N, K_d, ΔH, ΔS and ΔG) were summarized from three experiments (Supplementary Table S3-5).

Molecular docking

Molecular docking of dsRBD1 onto PBS-segment was guided by NMR experimental data. Using the crystal structure of human RHA dsRBD1 (PDB: 3vyy) and 1040 PBS-segment RNA structures that had been relaxed in MD simulations, a free docking by our in-house MDockPP program [164] was carried out to generate a total of 56,160,000 potential complex structures.

The modeled complex structures were subsequently sorted by the protein-RNA interaction-specific scoring function ITScorePR in a descending order based on their energy scores [200]. Then, the NMR-derived constraints related to residues of the PBS segment and the dsRBD1 were incorporated to filter out compliant candidates as described below. On the RNA side, restraints impose a maximum distance of 7 Å between both H1' and C1 of G129 in PBS-segment and dsRBD1 because of the significantly shifted H1' of G129 upon dsRBD1 titration, and a minimum distance of 5 Å between atom and C2 of A220 of PBS and dsRBD1 because A220-H2 remains unshifted in the NMR titration spectra. Additionally, restraints requesting U141 and U172 of PBS-segment be within 7 Å of dsRBD1 were also applied. One the protein side, restraints were generated from the ^{1}H-^{15}N TROSY titration data, which demand the following residues, F7, G13, K14, K16, M17, T18, Y21, K29, K54, K55 and Q58, be within 7 Å of PBS-segment. These residues exhibited significant chemical shift perturbation upon PBS-segment titration ($\Delta\delta > 0.05$ ppm). Distance restraints were also set for residue Y43 that should not be within 5 Å distance of RNA because no chemical shift perturbation was observed in the TROSY titration. Any qualifying candidate must satisfy all the specified restraints, resulting in a total of 69 modeled complex structures for further evaluation with experimental data. The best model with lowest energy was further optimized with UCSF Chimera [201].

Using Modeller [202], the structure of the dsRBD2-Core (RHA without dsRBD1 and RGG domains, 149-1150) was constructed by homology modeling based on the template structure of Drosophila MLE helicase (PDB: 5aor) [168], an orthologous protein sharing 51% sequence identity with the human RHA. Molecular docking was performed using our in-house docking program, MDockPP [164]. A 6° angular interval was used to sample the relative orientations between the dsRBD2-Core and PBS-segment RNA. The resulting 54,000 potential binding poses were then filtered with three constraints: (1) The α1 helix of the dsRBD2 interacts with the RNA,

(2) Region 3 residues (K235, K236) interact with the RNA, and (3) The dsRBD2-Core does not have any steric clash with the dsRBD1 domain when both bind to the RNA. For constraints (1) and (2), a 7Å distance cutoff was used to define the proximity to the RNA. Namely, the constraint (1) was met if at least one atom in a residue was within the cutoff distance to the RNA for every residue in the α1 helix. The constraint (2) regarding K235 or K236 was likewise determined. The constraint (3) was based on the fact that both dsRBD1 and dsRBD2-Core bind to the RNA simultaneously in physiological condition. The assessment of this constraint was based on the buried solvent accessible surface area (SASA) between the dsRBD1 and the dsRBD2-Core. The SASA was calculated using NACCESS [203]. The buried SASA equals $(SASA_{dsRBD1} + SASA_{dsRBD2-Core} - SASA_{dsRBD1+dsRBD2-Core})/2$, where $SASA_{dsRBD1}$ and $SASA_{dsRBD2-Core}$ stand for the SASA of dsRBD1 alone and the SASA of the dsRBD2-Core alone, respectively. $SASA_{dsRBD1+dsRBD2-Core}$ represents the SASA of the dsRBD1 and the dsRBD2-Core when they were viewed as a single structure after the dsRBD1: RNA structure and the dsRBD2-Core: RNA structure were superimposed on the PBS-segment RNA. Due to the inaccuracy in numerical calculations and the specific algorithm of the NACCESS, it is possible to obtain a non-zero, albeit small, SASA value for two molecules that are close but not touching each other. To counter this problem, a threshold value of 30 $Å^2$ was used to determine if two molecules have atomic clashes. The dsRBD1 and the dsRBD2-Core were deemed to have no atomic clash if the calculated buried SASA was below this threshold value.

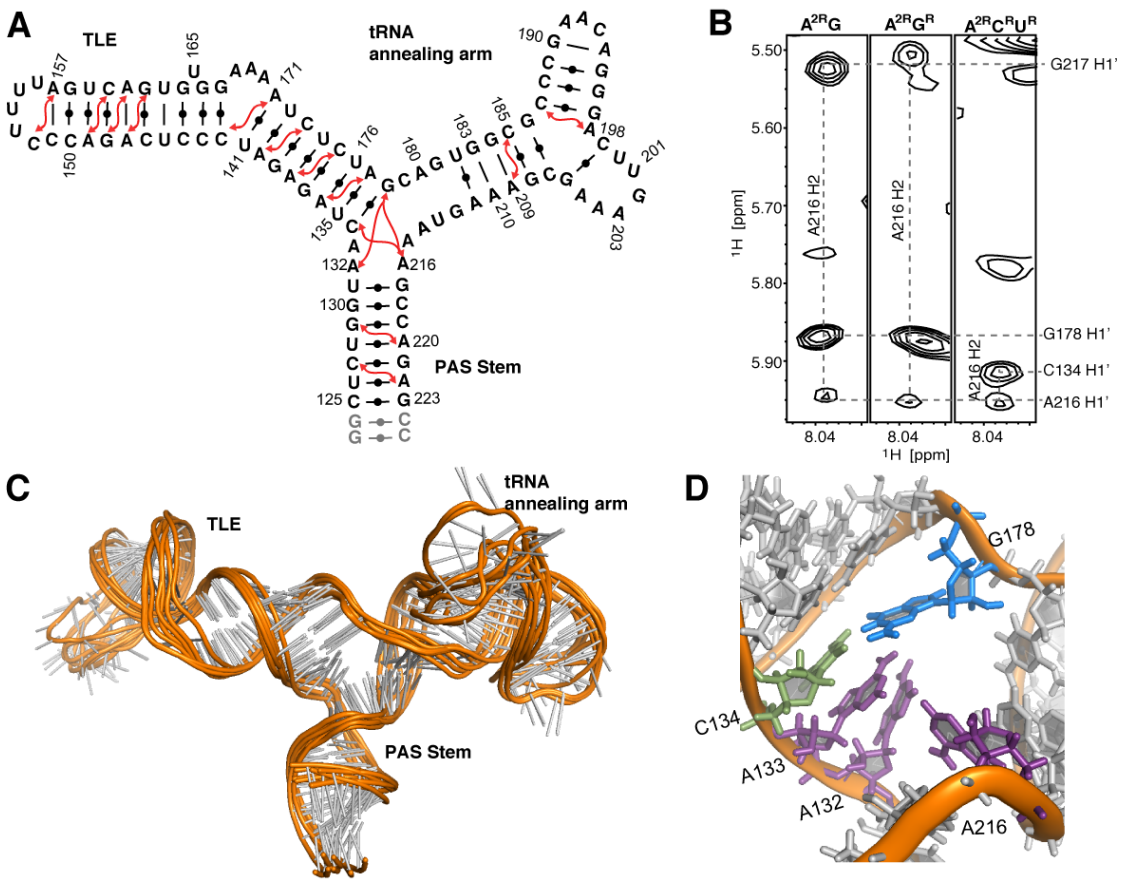


Figure 3-1. PBS-segment adopts a three-way junction structure. A. Secondary structure of the PBS-segment supported by NMR assignment is shown. Red arrows denote the inter-helical NOEs involving adenosine H2 protons. **B**. Portions of the 2D NOESY spectra with NOE connectivities supporting the three-way junction are shown. From left to right, spectra of $A^{2R}G$-, $A^{2R}G^{R}$- and $A^{2R}C^{R}U^{R}$- labeled PBS-segment. (See the nomenclature in the Experimental Procedures). **C**. NMR-derived secondary structure model of the PBS-segment. **D**. A zoomed-in view of the adenosine-rich three-way junction in the PBS-segment.

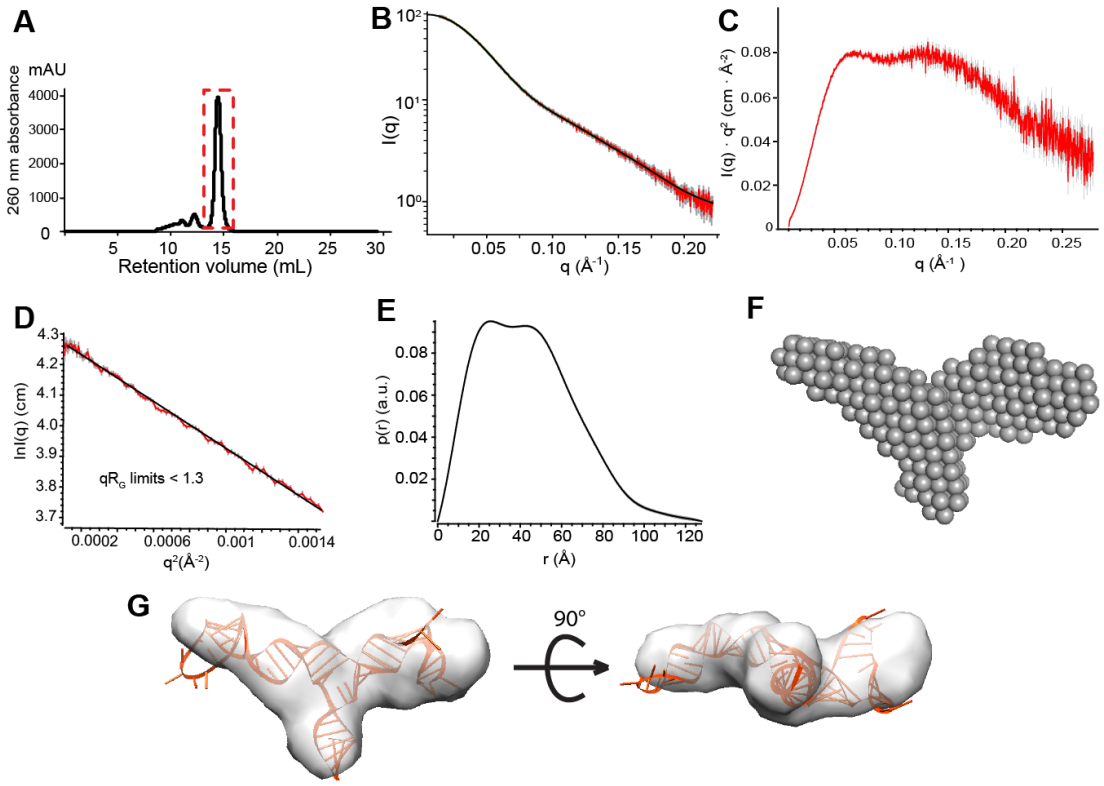


Figure 3-2. SAXS data of the PBS-segment support a three-way junction structure are shown.

A. SEC profile of the PBS-segment RNA is shown. The SAXS diffraction data of red dash boxed region were used for structural analysis of the PBS-segment. **B-E**. Averaging scattering profile (B), Kratky plot (C), Guinier plot (D) and the pair distance distribution function (E) for the scattering curve are shown. **F.** An *ab initio* model of the PBS-segment generated with DAMMIF is shown. **G.** Different views of the *ab initio* model superimposed with a representative structure of PBS-segment calculated by NMR-derived restrains and refined with the SAXS data.

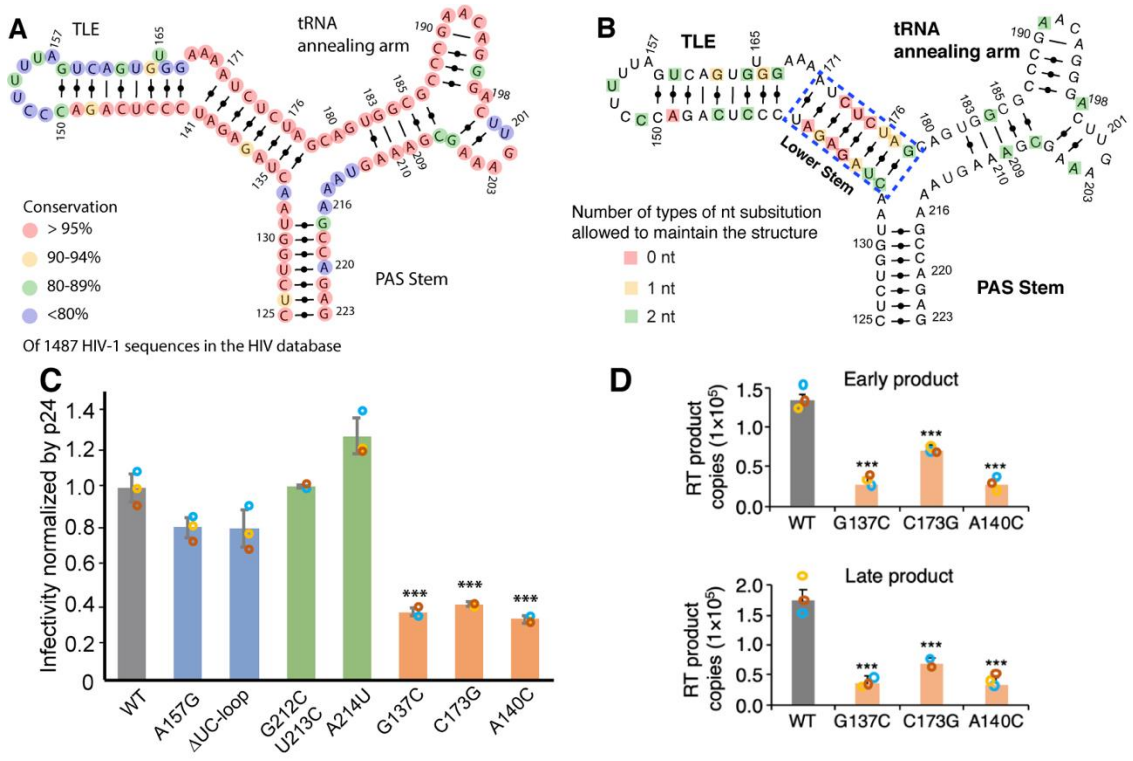


Figure 3-3. The three-way junction structure of the PBS-segment is conserved for infectivity. **A**. Phylogenetic analysis of 1487 clinical sequences spanning the PBS-segment is shown. The conservation at each nucleotide position is color labeled. **B.** Mutational scanning was performed to substitute each one nucleotide and predict the secondary structure of each mutant. The red, yellow, and green squares highlight the residues can be substituted by 0, 1 and 2 nucleotides in order to maintain the three-way junction structure. Unlabeled residues do not contribute to the three-way junction structure in a sequence specific manner. The low stem of TLE is shown in blue dashed box. **C**. Mutations that disrupt the three-way structure of the PBS-segment reduced the virus infectivity. Infectivity was normalized by the p24 protein level. Gray bar, WT; blue bars, mutations in the TLE loop; green bars, mutations in the tRNA annealing arm; orange bar, mutations in the TLE stem. **D**. Mutations that disrupt the three-way junction structure of the PBS-segment produced diminished HIV early reverse-transcription intermediates in target cells.

Cyan, red, and yellow circles represent data from three independent experiments. Statistically significant differences were measured by Student's T test: *** $p \leq 0.001$.

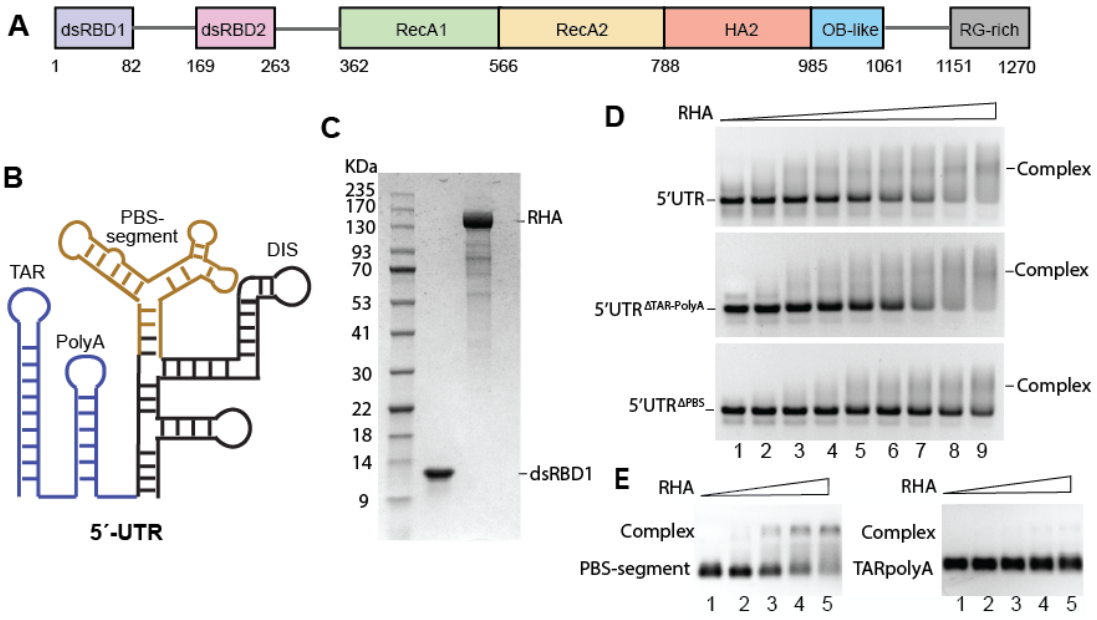


Figure 3-4. RHA preferentially binds to the PBS-segment region of the 5'UTR**. A**. The domain organization of RHA is shown. **B**. The secondary structural model of the 5'UTR is shown. TAR-PolyA, blue; PBS-segment, brown. **C**. A representative SDS-PAGE shows the molecular mass markers (left), recombinant dsRBD1 purified from *E.Coli* (middle), and recombinant RHA purified from insect cells (right). **D**. EMSA of recombinant RHA binding to 5'UTR, 5'UTR$^{\Delta TAR\text{-}PolyA}$ and 5'UTR$^{\Delta PBS}$ (from top to bottom). The RNA samples were pre-dimerized, and the final mixture contained 0.5 µM of RNA and various concentrations of RHA (from lane 1 to 9, RHA concentrations were 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 µM). The unbound RNA and complex were separated in a 1.5% native agarose gel. The representative gel of three independent experiments is shown. **E**. EMSA of recombinant RHA binding to PBS-segment and TAR-PolyA. The final mixtures contained 0.5 µM of RNA and 0, 1, 2, 3 and 4 µM of RHA (from lane 1 to 5). The representative gel of three independent experiments is shown.

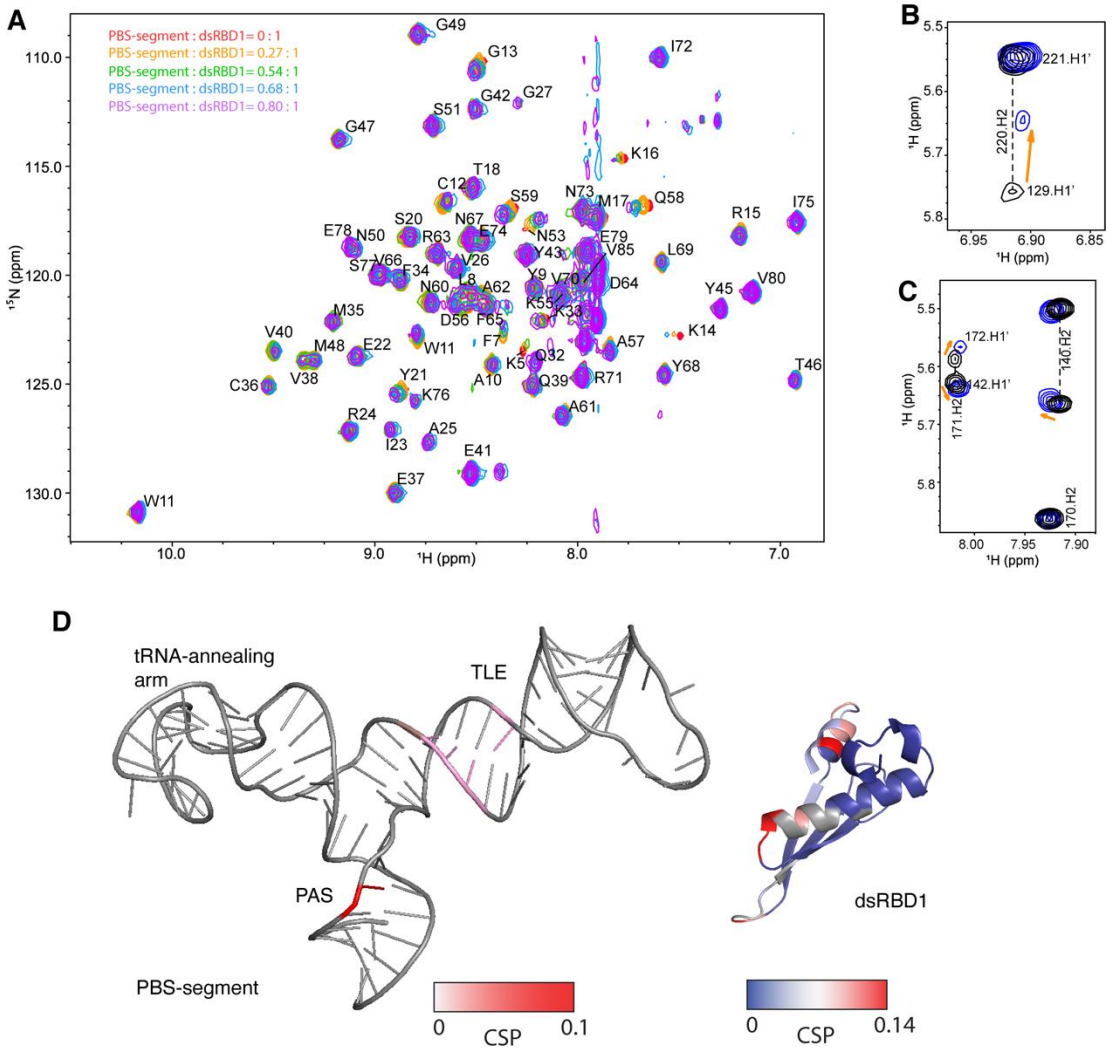


Figure 3-5. The dsRBD1:PBS-segment binding interface was mapped by NMR. **A.** ^{1}H-^{15}N TROSY spectra were collected for dsRBD1 titrated with various molar equivalents of PBS-segment RNA. The RNA: protein ratios are listed on the top right corner. **B-C**. Portions of the 2D NOESY spectra of the PBS-segment RNA (black) and in complex with one molar equivalent of dsRBD1 protein (blue) are shown. **D**. Chemical shift perturbations are mapped on the structure of PBS-segment and dsRBD1.

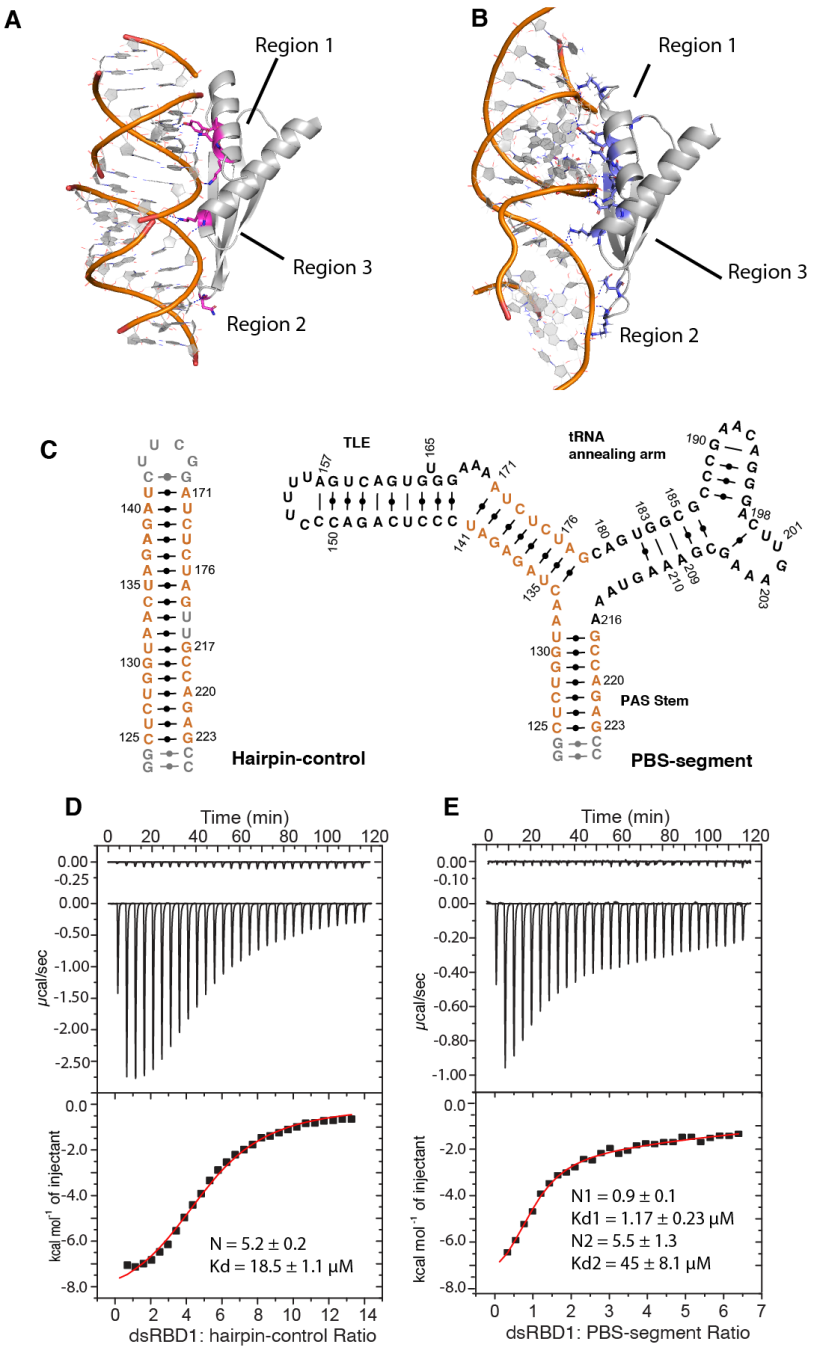


Figure 3-6. Binding of dsRBD1 to PBS-segment is of higher affinity than binding to a straight A-form dsRNA helix. **A.** Crystal structure of dsRBD1 in complex with $(GC)_{10}$ (PDB: 3vyy) shows that regions 1-3 interact with two consecutive minor grooves of dsRNA. The dsRBD1 residues forming hydrogen bonds with RNA are colored in magenta. **B**. NMR data derived model of dsRBD1: PBS-segment predicts more H-bonds between protein and RNA. The dsRBD1 residues

predicted to form hydrogen bonds with RNA are colored in blue. The hydrogen bonds are shown in blue dash lines. **C.** The secondary structure of Hairpin-control (left) and PBS-segment (right) are shown. The sequence identical residues located in both RNAs are colored in orange. **D**. ITC profile of titrating dsRBD1 into hairpin-control suggests the RNA contains several dsRBD1 binding sites. The heat of dilution profile is presented on the top panel. **E.** ITC profile of titrating dsRBD1 into PBS-segment shows PBS-segment contains one high-affinity dsRBD1 binding site and multiple weak bind sites. The heat of dilution profile is presented on the top panel.

Table 3-1. NMR Restraints and statistics of calculated structures

Cyana [1]	
NOE-derived restraints	504
Intra-residue	184
Sequential	306
Long range |i-j|>1	14
H-bond restraints	170
NOE restraints/residue	4.89
Target function (Å^2)	0.2 ± 0.008
Upper distance viol. (Å^2)	0.0114 ± 0.0002
Lower distance viol. (Å^2)	0.0026 ± 0.0003
Sum VDW viol. (Å^2)	1.25 ± 0.85
RMSD (Å)	2.9 ±1.6
SAXS	
χ^2	1.4 ± 0.02
MolProbity analysis [2]	
Clashscore	0, 100th percentile (0)
Probably wrong sugar pucker (%)	1 (0.97)
Bad backbone conformation	3 (12.6)
Bad bonds (%)	0 (0)
Bad angles (%)	2 (0.05)

[1] Statistics of 10 lowest energy models

[2] The 10 structures refined by MD simulations were evaluated using the MolProbity Server (89,90).

Supplementary Materials

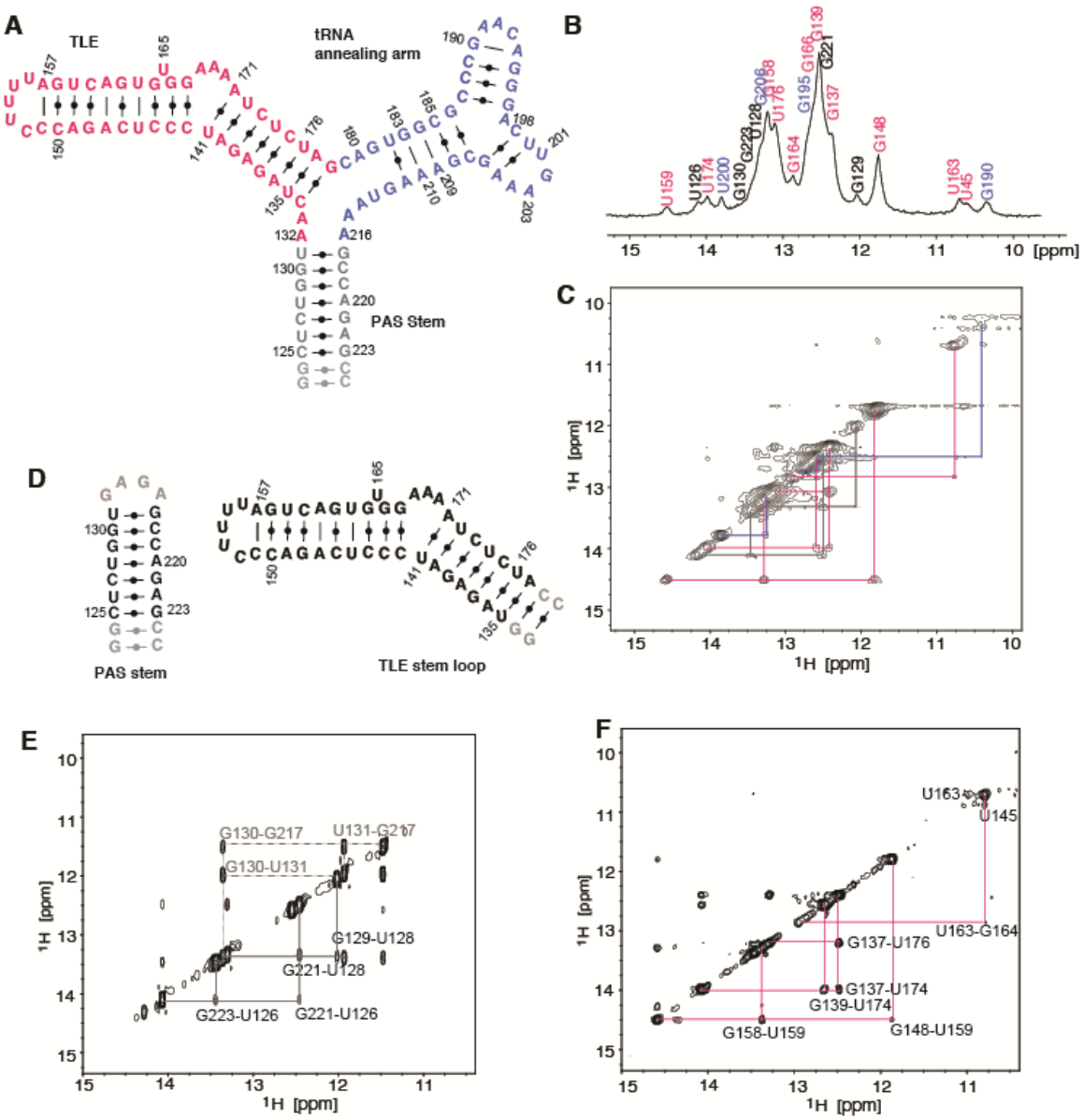


Figure S3-1. Imino proton spectrum supports the formation of the TLE and PAS stem in the PBS-segment. **A**. Secondary structure of the PBS-segment. **B**. 1D imino proton spectrum of the PBS-segment. Assignment of chemical shifts from PAS stem, TLE and tRNA annealing arm are shown in gray, red and blue, respectively. **C**. 1H-1H NOESY spectrum for the imino-proton correlation of the PBS-segment RNA. The imino-proton correlations in PAS, TLE and tRNA annealing arm are traced in gray, red and blue respectively. **D**. Secondary structure of PAS stem and TLE stem loop

RNAs to serve as controls for NMR assignment. E-F. 1H-1H NOESY spectrum for the imino-proton correlation of PAS stem (E) and TLE stem loop (F). The dashed lines denote NOE connectivities for the non-native residues in the PAS stem that do not exist in PBS-segment.

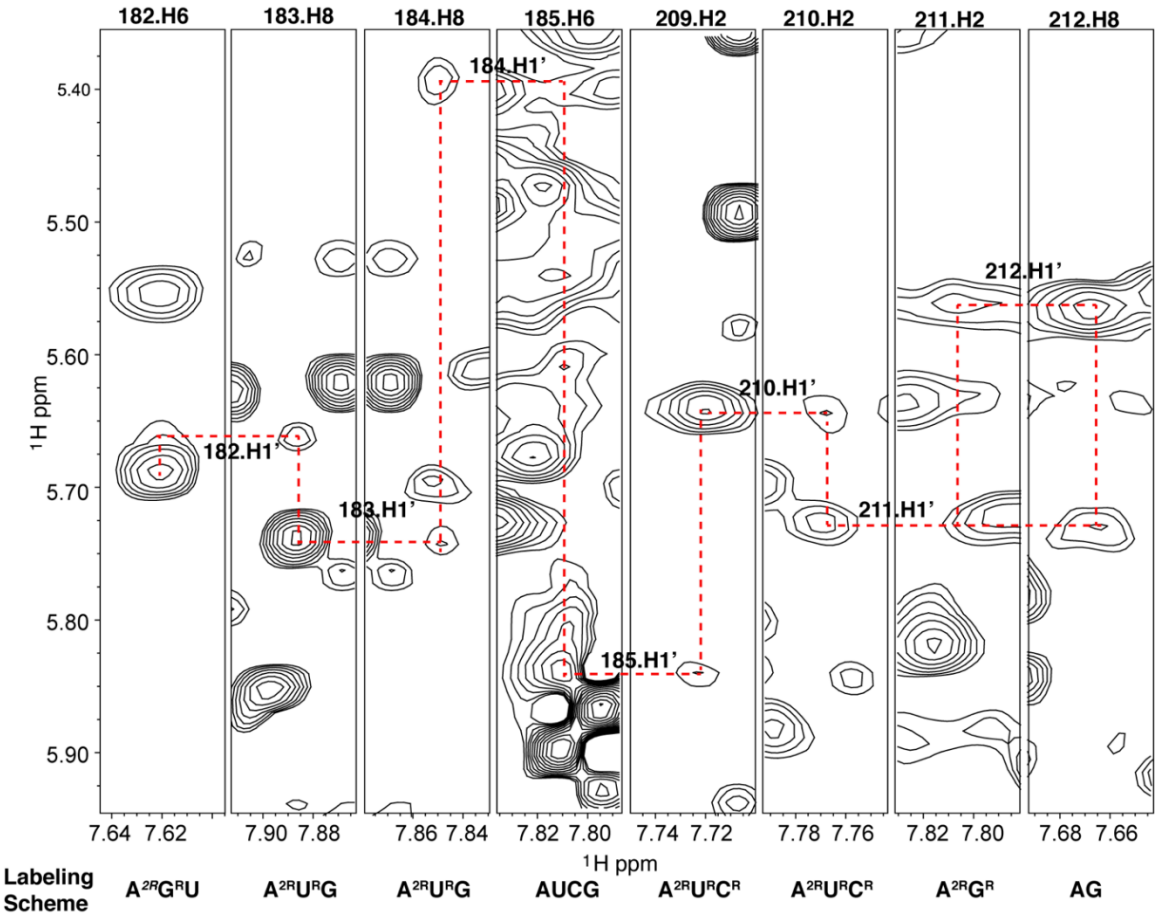


Figure S3-2. Plots of 1H-1H NOESY spectra from PBS-segment collected with different labeling strategies. NOE connectivities of residues in the tRNA annealing arm of the PBS-segment are shown in red dashed line. Base stacking from U182 to C185 (left four strips) and from A209 to A212 (right three strips) were observed. The cross-helical NOE between A209.H2 and C185.H1' (the 5th strip from the left) indicate the distance proximity between the two RNA chains. The RNA labeling scheme is shown on the bottom.

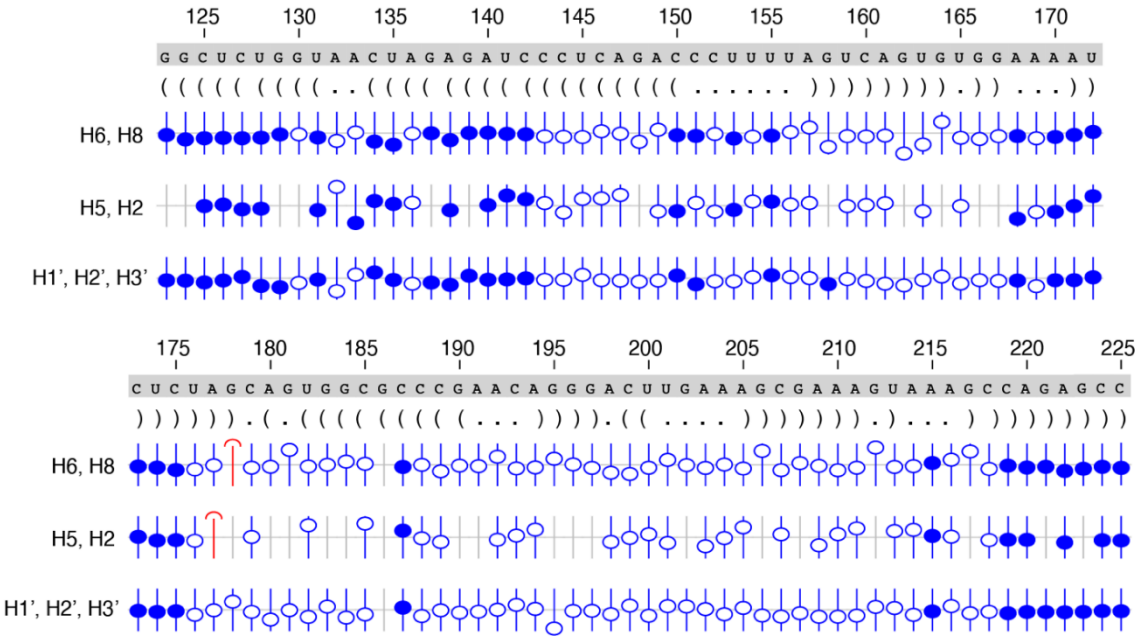


Figure S3-3. The sequence, secondary structure, and NMR assignment validation for PBS-segment are shown. The secondary structure is shown in Vienna dot-bracket format. The assignments of H6/H8, H5/H2, and H1'/H2'/H3' were validated by NMRView chemical shift prediction software. Filled circles denote the assigned proton chemical shift deviation from reported chemical shifts in the BMRB database, and the open circles denote the chemical shift deviation from chemical shift predictions. Assignments that fall outside of the tolerance range (± 0.5 ppm) are indicated with a red arrow. A177 and G178 are near the three-junction and thus their chemical shifts are away from predicted positions

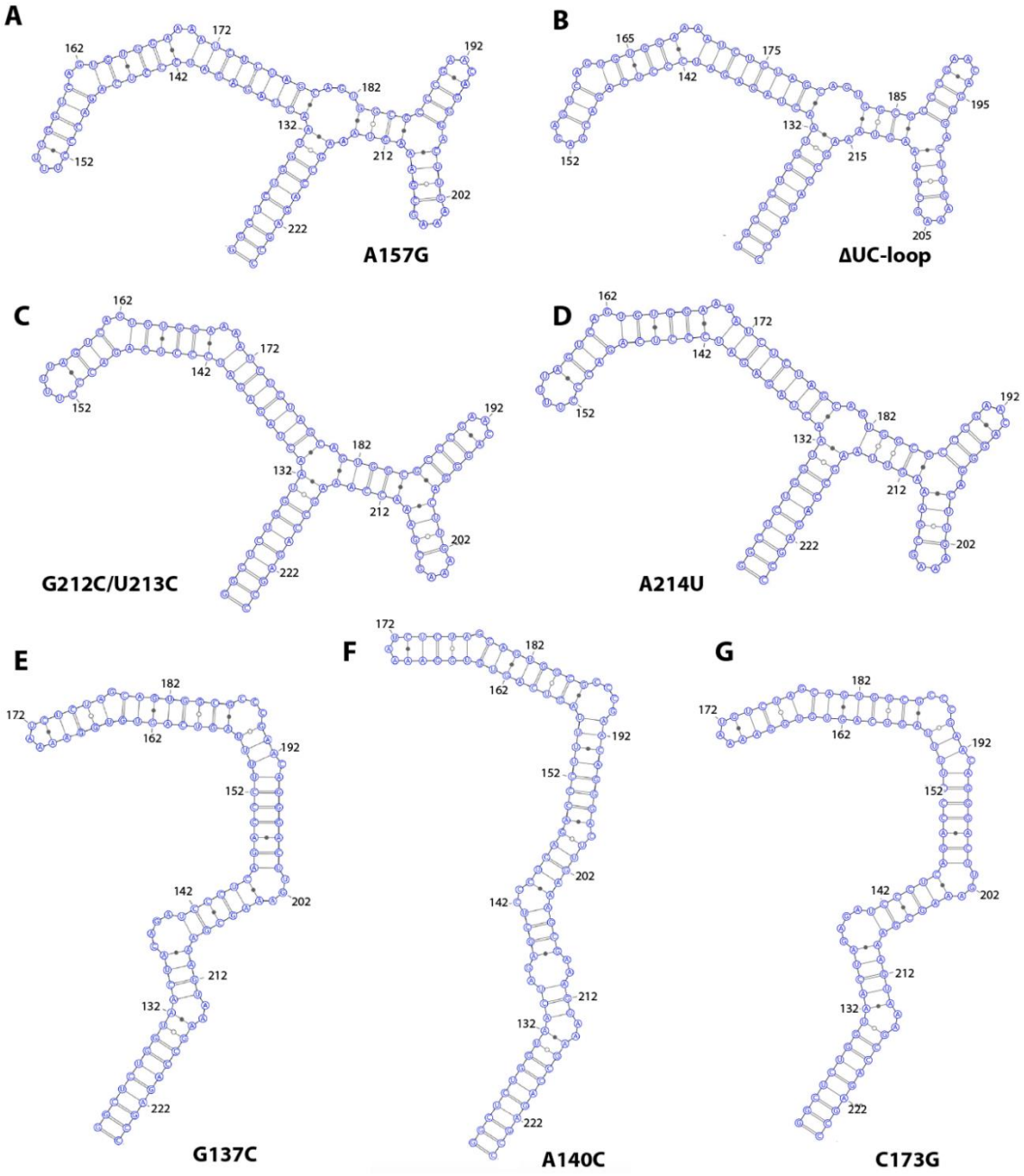


Figure S3-4: Predicted secondary structures of the PBS-segment mutants that were tested in the viral infectivity assays. Mutations in the TLE loop, including **A.** A157G, **B.** ΔUCloop, and mutations in the tRNA annealing arm, including **C.** G212C/U213C and **D.** A214U, do not disrupt the three-way junction structure. Mutations in the lower TLE stem, including **E.** G137C, **F.** A140C

and **G.** C173G resulted in an extended structure. All predicted secondary structures were performed using Vfold2D RNA structure folding model.

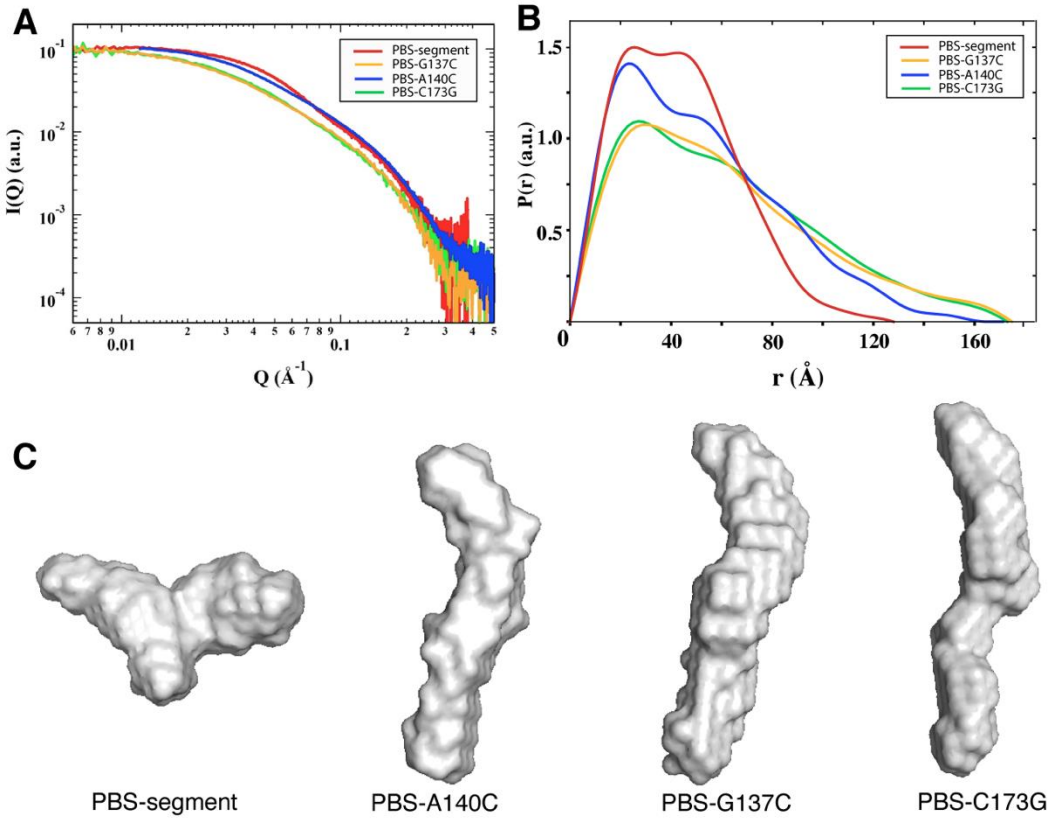


Figure S3-5. SAXS data of the PBS-segment RNA and mutants show that point mutations in the lower stem of TLE altered the RNA structure. A. Overlay of averaging scattering profiles of PBS-segment, PBS-G137C, PBS-A140C and PBS-C173G is shown. B. Overlay of the pair distance distribution function of PBS-segment and the mutants is shown. C. Ab initio models were of PBS-segment and the mutants were generated by DAMMIF.

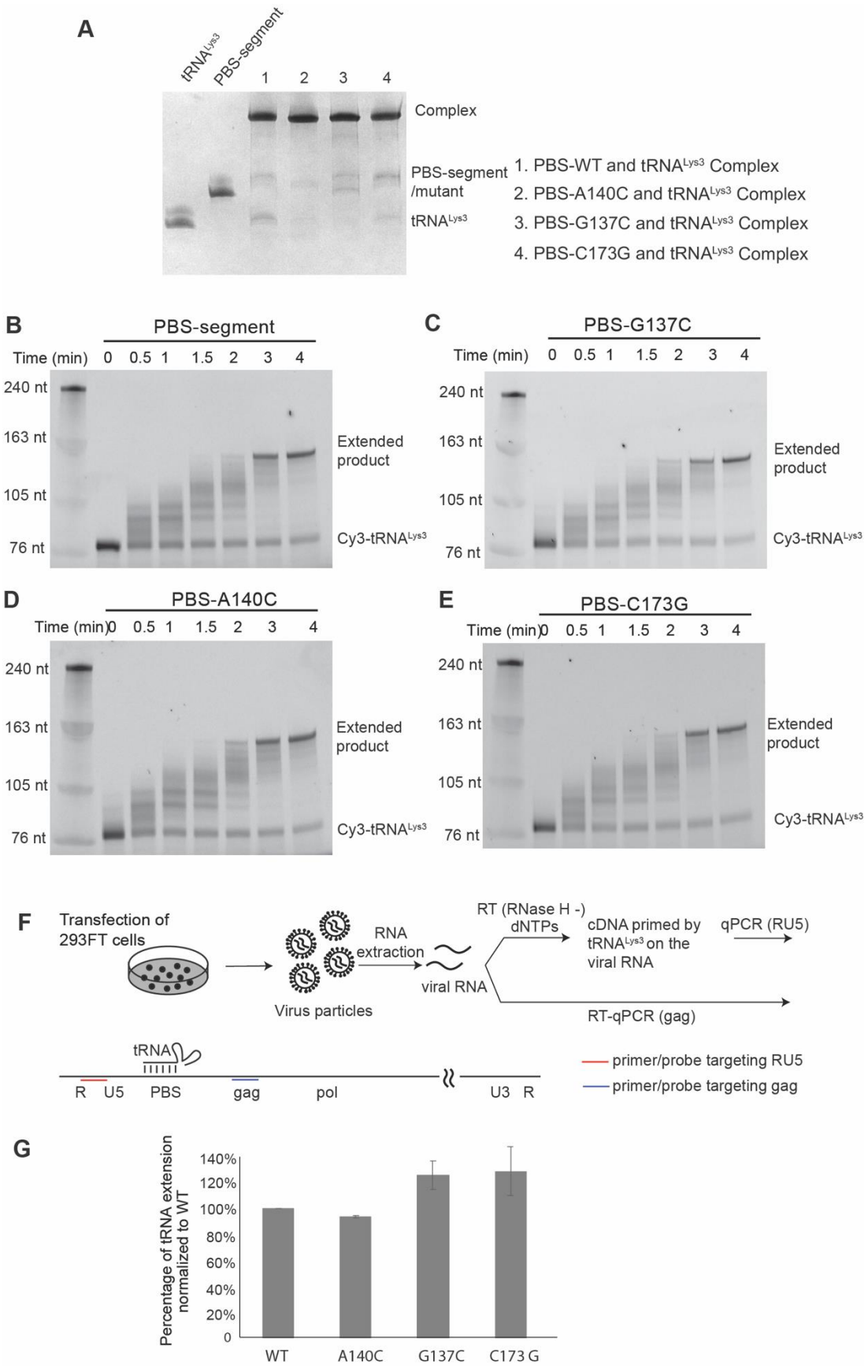

A
tRNALys3
PBS-segment
1
2
3
4
Complex
PBS-segment /mutant
tRNALys3
1. PBS-WT and tRNALys3 Complex
2. PBS-A140C and tRNALys3 Complex
3. PBS-G137C and tRNALys3 Complex
4. PBS-C173G and tRNALys3 Complex
B
PBS-segment
Time (min) 0 0.5 1 1.5 2 3 4
240 nt
163 nt
105 nt
76 nt
Extended product
Cy3-tRNALys3
C
PBS-G137C
Time (min) 0 0.5 1 1.5 2 3 4
240 nt
163 nt
105 nt
76 nt
Extended product
Cy3-tRNALys3
D
PBS-A140C
Time (min) 0 0.5 1 1.5 2 3 4
240 nt
163 nt
105 nt
76 nt
Extended product
Cy3-tRNALys3
E
PBS-C173G
Time (min) 0 0.5 1 1.5 2 3 4
240 nt
163 nt
105 nt
76 nt
Extended product
Cy3-tRNALys3
F
Transfection of 293FT cells
Virus particles
RNA extraction
viral RNA
RT (RNase H -) dNTPs
cDNA primed by tRNALys3 on the viral RNA
qPCR (RU5)
RT-qPCR (gag)
tRNA
R
U5
PBS
gag
pol
U3
R
primer/probe targeting RU5
primer/probe targeting gag
G
Percentage of tRNA extension normalized to WT
140%
120%
100%
80%
60%
40%
20%
0
WT
A140C
G137C
C173 G

Figure S3-6. The mutations in the PBS-segment TLE lower stem did not affect $tRNA^{Lys3}$ annealing, in vitro primer extension, and $tRNA^{Lys3}$ placement on viral RNA. A. The NCpromoted RNA duplex formed between PBS-segment mutants and $tRNA^{Lys3}$ shifted to the same position as the WT PBS-segment: $tRNA^{Lys3}$. B-E. The in vitro RT primer extension products of Cy3-labled $tRNA^{Lys3}$ on PBS-segment (**B**), PBS-G137C (**C**), PBS-A140C (**D**), and PBS-C173C (**E**) were resolved on 10% denaturing polyacrylamide gels. **F**. To investigate $tRNA^{Lys3}$ placement on gRNA, RNA was extracted from vector viruses produced within 48 h post transfection. Virion RNA complexes were isolated and incubated with RNaseH deficient RT (SuperScript II). cDNA intermediates of tRNA annealing and extension by RT were detected by qPCR with primers complementary to RU5. The input gRNA amount was quantified by RT-qPCR of gag region. Negative control qPCR to amplify the gag region was also performed on viral RNA to establish minimal contamination from proviral DNA (gag DNA copy numbers were less than 1% of the RNA copy numbers measured by RT-qPCR). **G**. The amount of cDNA synthesized by extending the $tRNA^{Lys3}$ annealed on the mutant viral RNA was quantified and compared to that of the WT.

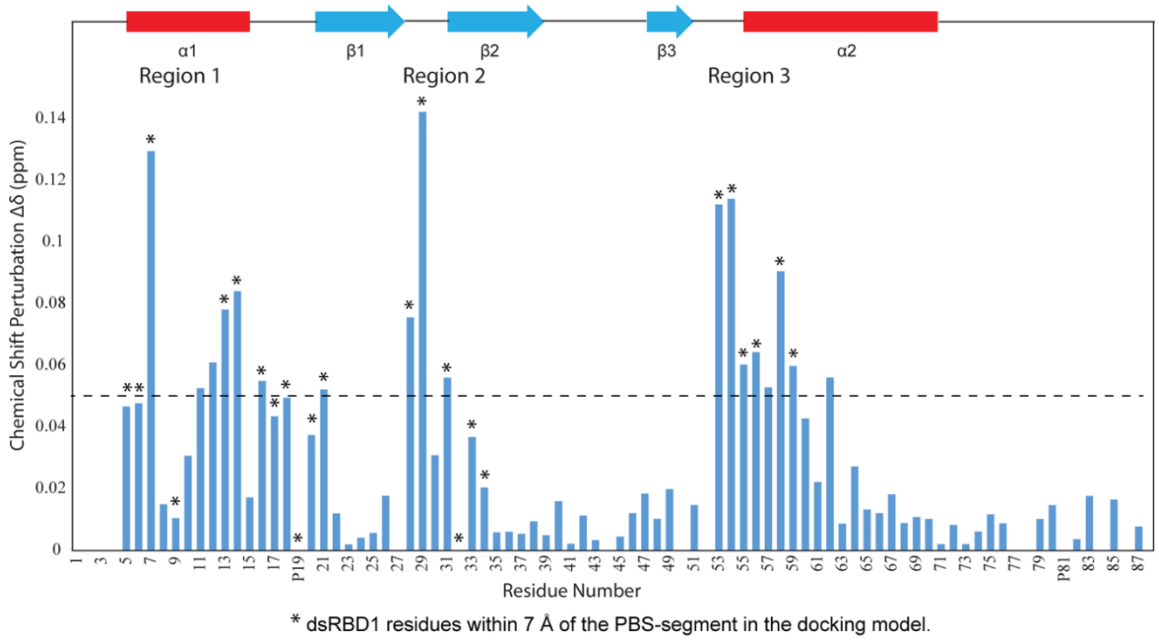


Figure S3-7. CSPs of dsRBD1 1H-15N backbone resonances upon the addition of PBS-segment RNA (the molar ratio of protein/RNA was 1:0.8) are plotted. Perturbations were calculated using the formula: Δδ = [δHN 2 + (δN/6.51)2] 1/2 and plotted against the primary structure. Secondary structure elements based on the crystal structure of human RHA dsRBD1 (PDB: 3vyy) is shown on the top. No signals were detected for P19 or P81 as prolines do not have an amide proton. Amide protons of some residues, including G27, Q32, N50, S77 and E78, were not detected under the experimental conditions and thus their CSPs were not plotted. Asterisks (*) denote residues that are within 7 Å of the PBS-segment RNA in the docking model.

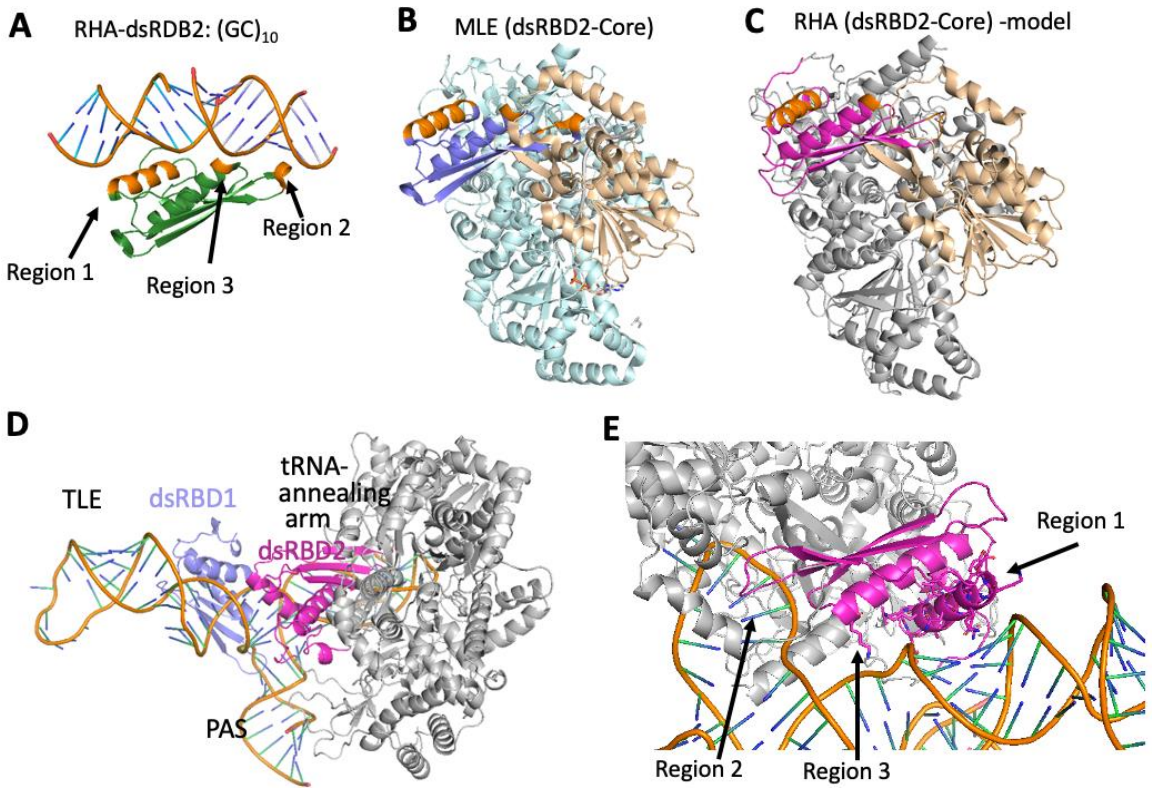


Figure S3-8. A model of dsRBD2-Core docking onto PBS-segment presents possible interactions between RHA and PBS-segment. **A**. Cartoon view of RHA dsRBD2 (green) in complex with (GC)10, with the RNA binding residues (regions 1-3) shown in orange (PDB: 3vyy). **B**. Cartoon view of MLE dsRBD2-Core. The dsRBD2 domain is shown in purple with the equivalent regions 1-3 shown in orange (PDB: 5aor). The RecA2 domain is shown in wheat. **C**. Cartoon view of RHA dsRBD2-Core model generated by homology modeling. The dsRBD2 domain is shown in magenta, and regions 1-3 are shown in orange. The RecA2 domain is shown in wheat. **D**. A model of RHA docked onto PBS-segment with the lowest energy score and does not interfere the dsRBD1 binding site. Purple: dsRBD1; pink, dsRBD2; grey, core domain of RHA. **E**. A zoom-in view of the dsRBD2:PBS interface. α1 (region1) binds to a minor groove in TLE lower stem and K236 (region 3) binds to the adjacent major groove. Residues in region 2 are not in the binding interface because it is surrounded by RecA2 domain.

Table S3-1. Primers for pNL4-3-CMV-EGFP mutagenesis

Mutants	Primers for Amplicon1	Primers for Amplicon2
A157G	BstZ17I-F; A157G-R	A157G-F; BstZ17I-R
ΔUC-Loop	BstZ17I-F; ΔUC-Loop-R	ΔUC-Loop-F; BstZ17I-R
G212C/U213C	BstZ17I-F; 212213-R	212213-F; BstZ17I-R
A214U	BstZ17I-F; A214U-R	A214U-F; BstZ17I-R
G137C	BstZ17I-F; G137C-R	G137C-F; BstZ17I-R
C173G	BstZ17I-F; C173G-R	C173G-F; BstZ17I-R
A140C	BstZ17I-F; A140C-R	A140C-F; BstZ17I-R

Table S3-2. Primer list for pNL4-3-CMV-EGFP mutagenesis

Primer	5’-3’ sequence
BstZ17I-F	CCTTCACCTGAAATGTGTGTATACAAAATCTAGGCCAGTC
BstZ17I-R	CTAGGTATGGTAAATGCAGTATACTTCCTGAAGTCTTTATC
A157G-F	TTTGGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAAC
A157G-R	TCCACACTGACCAAAAGGGTCTGAGGGATCTCTAGTTACCAG
ΔUC-Loop-F	CAGAGGAGACTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAAC
ΔUC-Loop-R	ACTGAGTCTCCTCTGAGGGATCTCTAGTTACCAGAGTCACACAACAG
212213-F	GCGAAACCAAAGCCAGAGGAGATCTCTCGACGCAG
212213-R	TGGCTTTGGTTTCGCTTTCAAGTCCCTGTTCG
A214U-F	GCGAAAGTTAAGCCAGAGGAGATCTCTCGACGCAG
A214U-R	TGGCTTAACTTTCGCTTTCAAGTCCCTGTTCG
G137C-F	GTAACTACAGATCCCTCAGACCCTTTTAGTCAGTGTGG
G137C-R	AGGGATCTGTAGTTACCAGAGTCACACAACAGACG
C173G-F	AATGTCTAGCAGTGGCGCCCGAACAG
C173G-R	CCACTGCTAGACATTTTCCACACTGACTAAAAGG
A140C-F	TGGTAACTAGAGCTCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTC
A140C-R	GAGCTCTAGTTACCAGAGTCACACAACAGACGGGCACACAC

Table S3-3. Primer list for quantitative PCR to measure the RT products

Primer		Sequence
KB2284	Early product sense	5’ GCC TCA ATA AAG CTT GCC TTG A 3’
KB2285	Early product antisense	5’ TGA CTA AAA GGG TCT GAG GGA TCT 3’
KB2363	Late product sense	5’ TGT GTG CCC GTC TGT TGT GT 3’
KB2364	Late product antisense	5’ GAG TCC TGC GTC GAG AGA TC 3’
KB3267	CCR5 sense	5’ CCA GAA GAG CTG AGA CAT CCG 3’
KB2368	CCR5 antisense	5’ GCC AAG CAG CTG AGA GGT TAC T 3’
RU5-F2	RU5 sense	5’ GCC TCA ATA AAG CTT GCC TTG A -3’
RU5-R3	RU5 antisense	5’ TAG AGT GGT CTG AGG GAT CT -3’
RU5-p	RU5 probe	5’ FAM-AGA GTC ACA CAA CAG ATG GGC ACA CAC T-TAMRA-3’
gag-F	gag sense	5’ CTA GAA CGA TT CGC AGT TAA TCC T 3’
gag-R	gag antisense	5’ CTA TCC TTT GAT GCA CAC AAT AGA G 3’
gag-P	gag probe	5’ FAM-CAT CAG AAG GCT GTA GAC AAA TAC TGG GA-TAMRA 3’

Table S3-4. SAXS statistics

Sample	PBS-segment
a. Sample Details	
Organism	HIV-1: pNL4-3
Source	NIH AIDS Reagent Program
Description	HIV-1 RNA pNL4-3 (125-223) synthesized with two terminal G-C pairs
Molecular masses from chemical composition (kDa)	33.3
SEC-SAXS column	GE Superdex 200 Increase 10/300
Loading concentration (mg/ml)	9
Injection volumn (μl)	200
Flow rate (ml/min)	0.75
Average concentration in combined data frames (mg/ml)	~ 2.2 (1.4 - 2.66)
Solvent composition	10 mM Tris, 140 mM KCl, 10 mM NaCl, 1 mM MgCl2, pH 7.5
b. SAXS data collection parameters	
Source and instrument	BioCAT 18ID beam line with Pilatus3 X 1M detector at the Advanced Photon Source (APS), Argonne National Laboratory storage ring (Lemont, IL, USA)
Wavelength (Å)	1.03
Sample-detector distance (m)	3.5

q-measurement range ($Å^{-1}$)	0.0054 – 0.3807
Exposure time (s) & number	876 successive 1 s frames
Sample temperature (°C)	25
c.Software employed	
SAXS data reduction	Beamline software
SAXS data Basic Analyses	PRIMUS/qt from ATSAS 2.8.0/3.0.0
Shape/bead modelling	DAMMIF from ATSAS 2.8.0
Atomistic modelling	CRYSOL from ATSAS 3.0.0
d. Structural parameters	
Guinier analysis	
I(0) (cm^{-1})	73.96 ± 0.11
R_g(Å)	33.59 ± 0.09
q-range ($Å^{-1}$)	0.0113 – 0.0386
qR_gmax	1.3
Coefficient of correl. R2	0.997
P(r) Analysis	
I(0) (cm^{-1})	74.4
R_g (Å)	34.50
d_{max} (Å)	128
q-range ($Å^{-1}$)	0.0123 -0.238
GNOM total est.	0.81
M from I(0) (kDa)	40.0
e. Shape model-fitting results	

DAMMIF (default parameters, 10 calculations)	
q-range for fitting ($Å^{-1}$)	0.0123 -0.238
Symmetry, anisotropy assumptions	P1, none
NSD (standard deviation)	0.731 (0.041)
χ^2 range	1.30 – 1.55
f. Atomistic modelling	
CRYSOL (with default parameters, No constant subtraction)	
χ^2,	1.40 ± 0.03
Predicted R_g (Å)	34.80 ± 1.69
Vol (Å),	32117 ± 334
Ra (Å),	1.45 ± 0.11
Dro ($eÅ^{-3}$)	0.026 ± 0.003
e. Data and model deposition ID	
SASDB	SASDJU7

Unpublished Results

As mentioned previously, to determine if the three-way junction structure of the PBS-segment is phylogenetic conserved, 1940 sequences that contained the 5'UTR were downloaded from the HIV-1 database (http://www.hiv.lanl.gov/) from all groups, subtypes, and recombinants. 1908 sequences were used for alignment due to incomplete sequence or nonstandard characters. Sequences were aligned with ClustalX2 using nucleotides from the PBS region, which included the primer activation signal, A-rich loop, primer binding site, and the 23 nucleotide insertion. The 23 nucleotide insertion occurs downstream of the PBS and contains a partial duplication of the PBS. The sequences were separated into two groups: those that contained a 23 nucleotide insertion and those that did not. Of the 1908 sequences 1487 sequences (78%) did not contain a 23 nucleotide insertion after the PBS and 421 (22%) sequences contained a 23 nucleotide insertion that was identical or similar to the HIV-1 MAL 23 nucleotide insertion.

The sequences that contained a 23 nucleotide insertion that was identical or similar to the HIV-1 MAL 23 nucleotide insertion were then realigned. From this alignment the sequences aligned into two distinct groups that differed in the position of the 23 nucleotide insertion. The linker sequence between the PBS and 23 nucleotide insertion differs in length between the two groups (Figure 3-7). Subgroup 1 contained the MAL sequence and its conservation was mapped to the secondary structure of MAL PBS segment (Figure 3-8). Subgroup 1 has a high level of conservation. The consensus sequence of subgroup 2 was used to create a secondary structure using mfold. The secondary structure has a similar secondary structure to the MAL PBS segment but does not form a stem after the PBS (Fig 3-9). We hypothesize that sequences that contain an insertion will have a secondary structure similar to the MAL PBS segment with a flexible hairpin after the PBS that may or may not be formed.

To test if recruitment of RHA via binding to the PBS segment is conserved in other HIV-1 subtypes, we utilized an EMSA using an HIV-1 subtype A sequence, MAL. To confirm that the PBS-segment is the binding target of RHA, recombinant full-length RHA protein expressed in a baculovirus system was purified (Fig. 4C). The recombinant RHA was titrated to a series of HIV-1 RNA constructs, including the MAL 5'UTR RNA (nt 1-358), MAL 5'UTR with PBS-segment deleted (MAL 5'UTR$^{\Delta PBS}$), and MAL 5'UTR annealed to tRNA^{Lys3}. These 5'UTR constructs have the 3'-half of the AUG hairpin truncated to favor RNA dimerization. Unexpectedly, the EMSA results show that RHA binds to all the constructs with similar affinity (Figure 3-10). We propose two possible explanations, Our MAL 5'UTR construct may have the PBS segment in a conformation that does not allow RHA to bind with high affinity and RHA may bind non-specifically to other regions of the MAL 5'UTR, or RHA is recruited in HIV-1 MAL using another region of the MAL 5'UTR and deleting or alter the PBS segment will not affect RHA binding. We were not able to confirm that RHA is recruited via the PBS segment in HIV-1 MAL, however the MAL PBS segment has been shown to adopt a similar global shape as the HIV-1 NL4-3 PBS segment [36] and further characterization of the binding between RHA and the MAL 5'UTR is necessary.

TGGCGCCCGAACAGGGACTTGAAAGCGAAAGTTAATAGGGACTCGAAAGCGAAAGTTCCAGAG

Subgroup 1

TGGCGCCCGAACAGGGACTTGAAGTTAATAGGGACTCGAAAGCGAAAGTTCCAGAG

Subgroup 2

Figure 3-7. Selection of consensus sequences of subgroups from sequences that contain a 23 nucleotide insertion. The PBS is highlighted in purple and the 23 nucleotide insertion is highlighted in green.

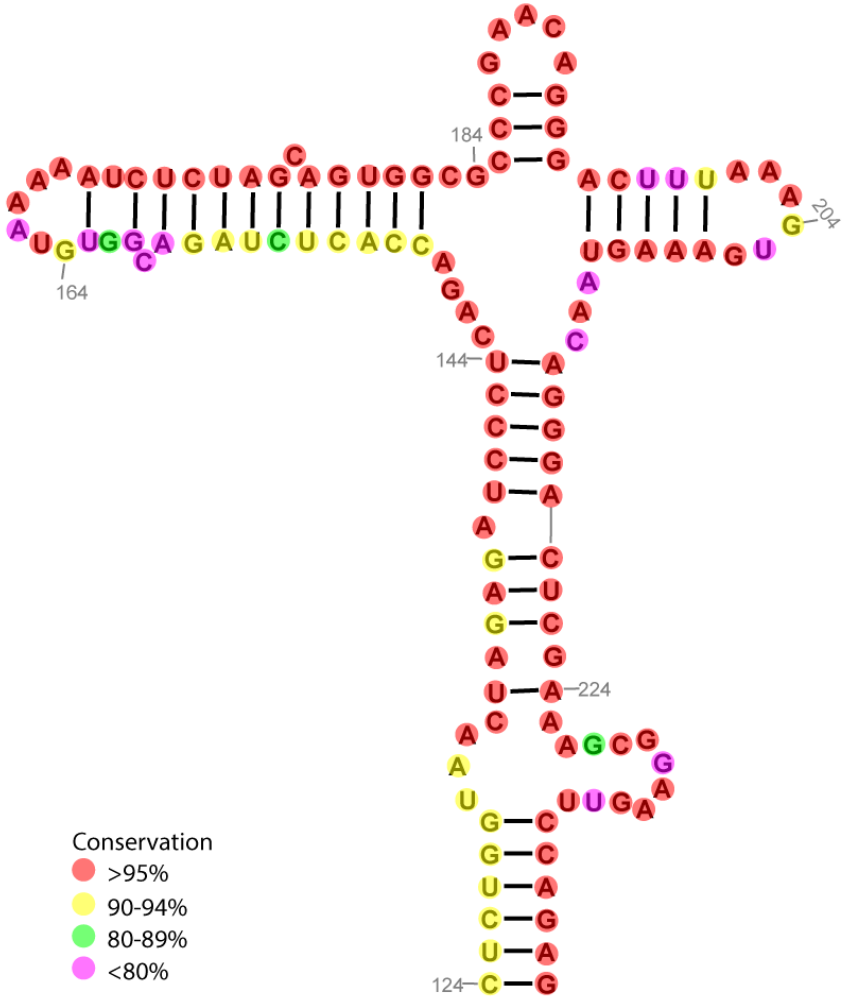


Figure 3-8. Conservation of PBS segment in sequences from subgroup 1 that contain a 23 nucleotide insertion.

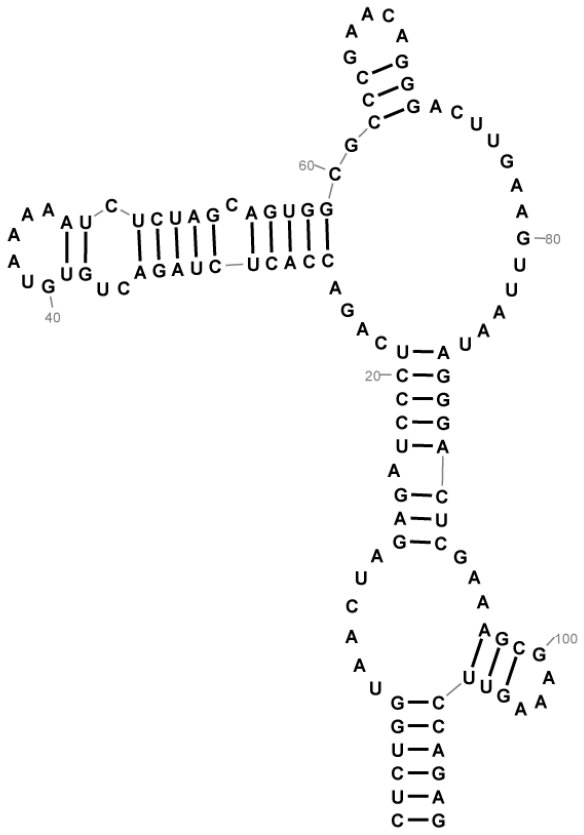


Figure 3-9. Predicted secondary structure (mfold) of consensus sequence from insertion containing sequences subgroup 2.

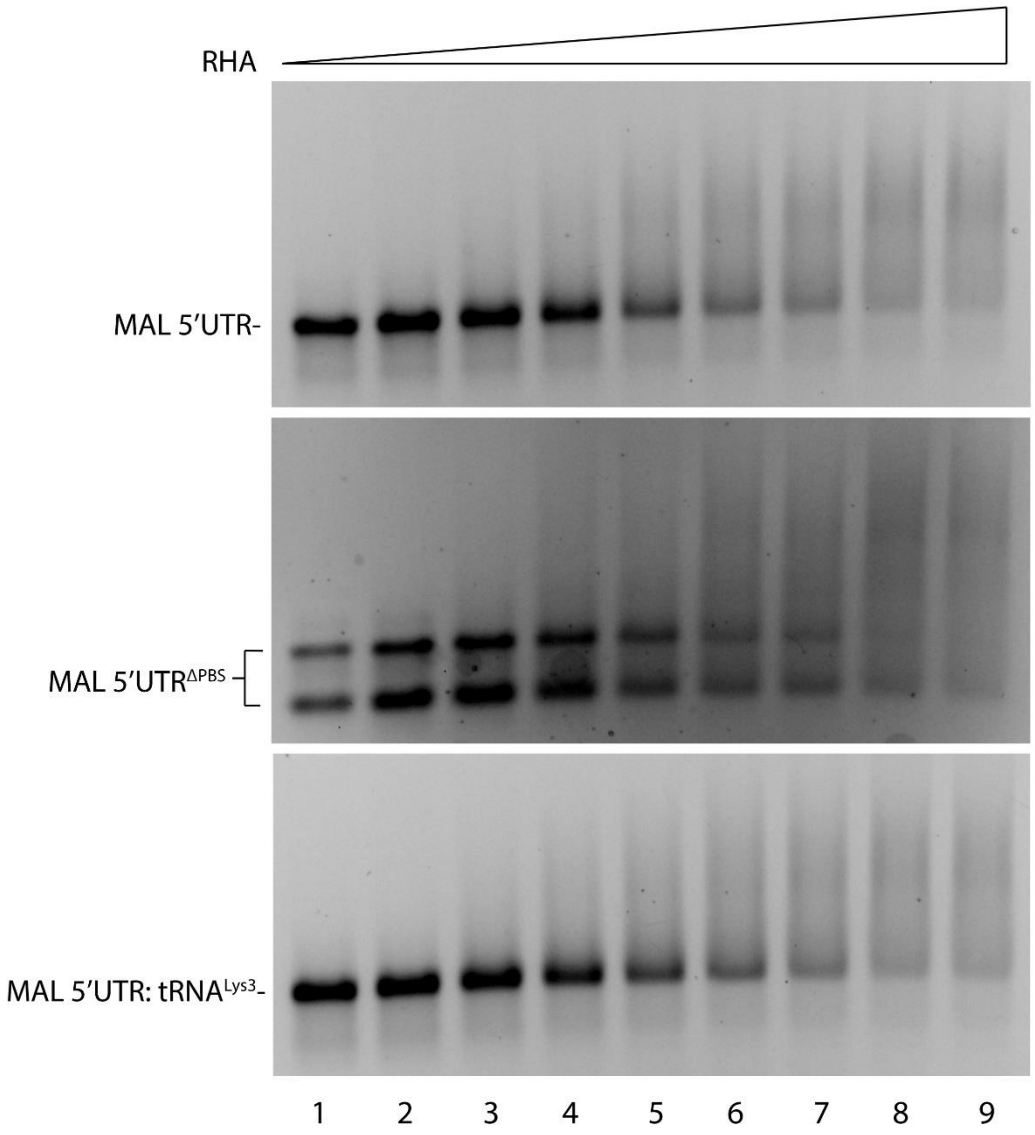


Figure 3-10. EMSA of recombinant RHA binding to MAL 5'UTR, MAL 5'UTRΔPBS , and MAL 5'UTR: tRNALys3 (from top to bottom). The RNA samples were pre-dimerized, and the final mixture contained 0.5 µM of RNA and various concentrations of RHA (from lane 1 to 9, RHA concentrations were 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 µM). The unbound RNA and complex were separated in a 1.5% native agarose gel.

Chapter 4. Structural Studies of RNA Aptamers that Can Shift the Redox Potential of Bound Flavins

Abstract

The RNA world hypothesis postulates that RNA was once a carrier of genetic information and catalyst. RNA enzymes would have needed to carry out redox reactions. Here we report a novel aptamer, X2B2, that binds flavin and shifts its reduction potential, opening the door for redox ribozymes. Using isothermal titration calorimetry, we show that this aptamer and a mutant are able to bind flavin moieties with high affinity. This mutant, X2B2-C14U, was able to shift the reduction potential even further than X2B2. Preliminary structural studies have shown that X2B2-C14U contains a scaffold of base triples that support flavin binding and reduction potential shift may be driven by base stacking interactions with an un-base paired FMN.

Introduction

Many modern protein enzymes utilize small molecule cofactors to catalyze chemical reactions, such as flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) which carry out one-electron and two-electron transfer reactions. The RNA world hypothesis argues that in early life RNA molecules could have been utilized as carriers of genetic information and participated in catalysis of chemical reactions [204–206].

Modern enzymes that utilize flavin moieties participate in many biochemical reactions such as, redox transfer of one or two electrons, light emission, the activation of oxygen for oxidation and hydroxylation reactions, and the dehydrogenation of metabolites[94]. Redox reactions can be driven by utilizing binding energy upon binding flavin moieties in order to increase the reduction potential of the flavin moiety and increase reactivity. In early life, catalytic RNAs that bind flavin and increase its reduction potential would have been important. Aptamers that can bind FAD and NAD+ have been selected[95–97]. However, their ability to

shift reduction potential of bound flavin was not evaluated and these studies focused on the elucidating the molecular details of the flavin-binding mode. In this study, we report an RNA aptamer that preferentially binds to the oxidized form of flavin. By comparing the spectroscopic signatures of our FAD-specific aptamer to other reported flavin aptamers, we were able to identify additional aptamers with similar effects on the bound flavin. We have been able to characterize the binding affinity of theses aptamers to flavin moieties. We have also structurally characterized the binding of these aptamers to flavin moieties using nuclear magnetic resonance (NMR). From these experiments we have a built a preliminary structure of one of these aptamers.

Results

Aptamer isolation proceeded using a standard SELEX approach. FAD was immobilized on a solid support and the pool RNA sequences contained a 42-nucleotide random region flanked by primer binding sites. Negative selection as carried out using an ADP-modified column to remove sequences binding to adenosine. After 6 selection rounds, the enriched pool was mutagenized and selection pressure for low k_{off} rates was implemented for an additional 6 rounds prior to cloning. Four candidate aptamers were identified and showed a preference for FAD over $FADH_2$. The strongest binder, designated 12.29, was truncated a 38-mer designated X2B2 based on in-line probing and 1-D NMR analysis.

To determine if aptamer X2B2 was capable of shifting the reduction potential of the bound flavin, we used the xanthine oxidase assay used for flavoproteins [207]. In this assay, xanthine oxidase reduces a common oxidizing agent, methylene viologen. The methylene viologen then reduces the bound flavin, as well as a reference dye chosen to have a reduction potential near that of the bound flavin. The reduction of the flavin and the reference dye is monitored spectroscopically and can be used to determine the reduction potential of the bound

flavin. As shown in Figure 4-1 free FAD has a reduction potential of -209 mV. When bound to an aptamer not specific for FAD redox state, a very small shift in potential is observed. However, for aptamer X2B2, the reduction potential for the flavin shifts -11 mV. Additionally, for X2B2, upon binding FMN and riboflavin, the reduction potential is shifted -22 an -14 mV respectively. We also tested if a previous aptamer that had been selected for FMN binding could also shift the reduction potential [95]. We found that this aptamer, 35FMN-2, was also able to shift the reduction potential -48 mV.

To understand how the aptamer is accomplishing this effect, we introduced mutations to pinpoint nucleotides in the proposed aptamer structure that might be important for flavin binding. Many of the mutants lost the ability to bind flavin, but one mutation produced interesting effects. Changing C14 to U (designated X2B2-C14U) resulted a shift in reduction potential of -50 mV (Fig. 4-1) upon binding FMN. X2B2-C14U shifts the reduction potential of bound FAD and riboflavin by -37 and -25 mV, respectively.

In order to understand if there was a correlation between reduction potential shift and binding energy we characterized the interactions of FMN, FAD, and riboflavin with aptamer X2B2 and X2B2-C14U using isothermal titration calorimetry (ITC). We determined aptamer X2B2 had a K_d of 522 ± 60.42 nM for FMN, while aptamer X2B2-C14U binds more tightly, with a K_d of 191 ± 66.07 nM (Figure 4-2). FAD and riboflavin bound with decreased affinity and these results are summarized in Table 4-1. We did not detect a linear correlation between reduction potential shift and binding energy but the aptamer that caused the greatest shift in reduction potential, X2B2-C14U, bound FMN with the highest affinity.

Two dimensional ^{1}H-^{1}H NOESY NMR experiments were carried out to identify the location of FMN binding in these aptamers and to understand the role certain nucleobases, specifically the ones that led to substantial changes upon mutation, play in the binding of FMN

and the observed shift in reduction potential. To enhance spectral resolution and aid in assignment, nucleotide-specific ^{2}H-labeled samples (Experimental Procedures) were prepared for NMR data collection of non-exchangeable protons. We collected spectra for X2B2-C14U: FMN using AC, AG, GU, and $A^{2R}U^{R}CG$ labeling strategies. The labeled samples enabled us to identify the key residues of the FMN binding site and aided in assignment of other residues. We found that FMN has NOEs with U14, G15, and U18. These interactions occurred between the methyl groups and aromatic protons of the isoalloxazine ring of FMN and key interactions can be seen in Figure 4-3.

In addition to 2D NOESY data from non-exchangeable protons we collected a ^{1}H-^{1}H NOESY of X2B2-C14U: FMN that included the imino proton region of the spectra. To aid in assignment of this spectrum and to gather more information on the base pairing of X2B2-C14U: FMN we collected an HNNCOSY spectrum using ^{13}C, ^{15}N labeled X2B2-C14U: FMN (unlabeled) and an HNN-COSY spectrum using ^{13}C, ^{15}N labeled X2B2-C14U: riboflavin-(dioxopyrimidine-$^{13}C_4$, $^{15}N_2$). This riboflavin contains ^{13}C and ^{15}N on the uracil like ring of the isoalloxazine ring of riboflavin. These experiments showed that an A: U Hoogsteen edge base pair was present in X2B2-C14U: FMN and that FMN had NOEs with this A: U base pair and a guanosine nucleotide (Figure 4-4c). Through comparison of these spectra and our non-exchangeable 2D NOESY spectra we were able to assign this base pair to adenosine 26 and uracil 18. Additionally, a new peak appeared in the ^{13}C, ^{15}N labeled X2B2-C14U: riboflavin-(dioxopyrimidine-$^{13}C_4$, $^{15}N_2$) HNN-COSY spectrum (Figure 4-4b). This peak is from the uracil like ring of the isoalloxazine ring of riboflavin. No additional peaks were detected that indicated riboflavin-(dioxopyrimidine-$^{13}C_4$, $^{15}N_2$) was base paired. Thus, we concluded that FMN does not base pairto any residue in X2B2-C14U.

Through our assignments of these spectra we were able to build a secondary structure model of X2B2-C14U: FMN (Figure 4-5). X2B2-C14U: FMN forms 8 standard Watson-Crick base pairs followed by five base triples. These base triples form a scaffold for FMN to bind and FMN is sandwiched by a base triple below and 15G above. The imino proton data also showed that the U like edge is near 19G. Our analysis is still ongoing, and we are still searching for NMR evidence that gives more information about the contacts above FMN.

To understand the molecular details of FMN's interaction with X2B2-C14U we have calculated a preliminary solution structure using NMR-derived restraints by CYANA [192]. Standard torsion angle restraints, standard hydrogen bonding restraints for Watson-Crick base pairs, and standard phosphate distances in A-form RNA helical structures were employed to maintain A-form helical geometry in the first 8 base pairs (nt 1-8, and 31-38) [192,193]. For the five base triples standard hydrogen bonding restraints were used based on the secondary structure derived from our NMR analysis. NOE restraints of this region, the five base triples and of residues involved in FMN recognition were also utilized. Figure 4-6 shows an ensemble of 10 structures derived using these restraints. These structures agree well in regions below FMN binding but are flexible above the FMN binding site. More restraints and further refinement of this structure is necessary. Figure 4-7a-c shows the five potential base triples that are used as a scaffold for FMN binding. These base triples do not base stack perfectly in our current model. Figure 4-7d shows the FMN binding site. FMN is recognized by the fifth base triple below and the region above is flexible. This model shows FMN stacking with 15G above and G19 nearby.

Discussion

This work represents the first time that an aptamer has been shown to shift the reduction potential of bound flavin. We were able to characterize and aptamer X2B2 and a mutant X2B2-C14U that were able to shift the reduction potential of bound FMN, FAD, and

riboflavin. Upon examining a previous aptamer that was selected for FMN, 35FMN-2, we showed that is also shifted the reduction potential upon binding FMN.

We were able to characterize the binding affinity using ITC and found that the highest reduction potential shift correlates with the highest binding affinity of X2B2 and X2B2-C14U binding FMN. We didn't see a linear trend between the reduction potential shift upon FAD and riboflavin binding, but this could be due to the high standard deviation of some of the ITC data. One other possibility is that binding affinity is not directly correlated with reduction potential shift and further analysis of the molecular contacts between the aptamers and FAD and riboflavin is necessary.

Our preliminary structure shows many non-canonical base pairs and base triples. Based on our NMR assignment were able to conclude that there are many non-canonical interactions. Our current secondary structure model is only model we have been able to derive that satisfies all our NMR assignments. In one previous structural study of an aptamer that binds FMN, FMN uses its U-like edge to base pair with an adenosine residue via its Hoogsteen edge[97]. We had considered that FMN may base pair with an adenosine but we unable to detect evidence of base pairing (Figure 4-4). Thus, in our model FMN is not base paired and is recognized via base stacking of its isoalloxazine ring. Another surprising element of this model is the large number of base triples that stack on top of one another. This may not be typical but stacking of multiple base triples has been observed. In the *Tetrahymena* ribozyme the active site contains stacking of 4 base triples [208]. Additionally, the previous structural study of a FMN binding aptamer showed that FMN stacks onto a base triple below it [97]. As FMN's isoalloxazine ring contains 3 rings, stacking onto a base triple may be necessary for efficient binding.

One aspect that our study is not currently able to address is what are the molecular mechanisms that drive the difference in the reduction potential between X2B2 and X2B2-C14U.

Why does X2B2-C14U shift the reduction potential further than X2B2? Our study shows that a key residue involved in FMN recognition is U14. U14 is the one mutation that differentiates X2B2-C14U from X2B2. Our current model suggests that U14 may be involved in a U14-U18-A26 base triple. In this base triple U14 can potentially have two hydrogen bonds with U18. In X2B2 this residue is cytosine. If this cytosine participates in a C14-U18-A26 base triple it can only have one potential hydrogen bond with U18. This may make C14 more flexible and prevent efficient base stacking that drives a larger shift in reduction potential. Further structural analysis of the X2B2 aptamer is necessary.

Author Contributions

This work is an ongoing collaboration between the Heng Lab, Burke Lab (University of Missour), and Baum Lab (St. Louis University). For this study, the Burke Lab selected the initial aptamer and Jack Samuelian of the Baum Lab characterized the reduction potential shift of the aptamers. I collected ITC data, NMR data, and did preliminary structural calculations.

Experimental Procedures

In vitro RNA transcription

RNA used for in vitro experiments were synthesized by T7 in vitro transcription. The DNA templates for RNA synthesis were purchased as primers and annealed using heat (Integrated DNA Technologies, Coralville, Iowa, USA). To determine optimal conditions for large-scale transcriptions, small-scale trial transcriptions were performed using varying concentrations of $MgCl_2$. Large-scale transcriptions were carried out at volumes of 7.5–10 mL in transcription buffer (40 mM Tris-HCl, pH 8.0, 5 mM dithiothreitol (DTT), 10 mM spermidine, and 0.01% (v/v) Triton X-100) using dNTPs (12 mM each), Ribolock RNase Inhibitor (80 units, Thermo Fisher Scientific, Waltham, MA, USA), and T7 RNA polymerase (1 μM). Transcription reactions were held at 37 °C for 4 h and quenched with 1-M urea and 25 mM ethylenediaminetetraacetic acid

(EDTA). RNAs were purified by sequencing gels, visualized by UV shadowing, electroeluted from the gel, and washed in ultra-centrifugal filters to remove acrylamide and salts. The nucleotide specific ^{2}H labeled samples, including AC, AG, GU, and $A^{2R}U^{R}CG$ (Nomenclature: the letters denote the nucleosides containing ^{1}H which is visible by NMR, R=ribose, A^{2R} means H2 and ribose hydrogens are protonated; $A^{2R}U^{R}CG$ means the H2 and ribose of adenosine are protonated, cytosine and guanosine are fully protonated) were prepared by incorporating corresponding deuterated and protonated rNTPs in T7 transcriptions as previously described [151,185]. Fully deuterated rNTPs and H5, H6-deuterated CTP and UTP were purchased from Silantes (Silantes GmbH, Munich) and Cambridge Isotope Laboratories (CIL, Andover, MA). H8-deuterated ATP and GTP were prepared in the lab [151]. $^{13}C/^{15}N$ labeled RNAs were synthesized by incorporating $^{13}C/^{15}N$ labeled rNTPs (CIL, Andover, MA) in T7 transcription.

Reduction potential shift determination

Mixed in a sealable quartz cuvette is 100 mM HEPES, pH 7.5, 200 mM potassium chloride, 15 mM magnesium chloride, 700 µM xanthine, 30 µM methyl viologen, 20 µM reference dye (anthraquinone-2-sufonic acid or phenosafranin), 20 µM flavin (FAD, FMN, or riboflavin), 20 µM RNA aptamer, 5 mM glucose, 75 µg/mL glucose oxidase, 20 µg/mL catalase, and 10-30 µg/mL xanthine oxidase. Xanthine oxidase is added last to initiate the reaction. The cuvette is sealed with a septum, placed into the UV-Vis spectrophotometer, and argon is continuously flushed over the headspace in the cuvette at a slow rate to ensure no oxygen leaks in during the experiment. Spectra is collected using an Agilent Cary 100 UV-Vis spectrophotometer every 60 seconds for ~3 hours, or until no more reduction of the flavin and reference dye is seen, and data between 280-700 nm is collected in 1 nm increments. Reduction potentials can be calculated by using the Nernst equation (Eq. 1), where R is the gas constant, F is Faraday's constant, n is the number of electrons transferred, and T is the

temperature in Kelvin. At room temperature, and for a 2 e^- redox reaction, the RT/nF term is equal to 12.5.

$$E = E_m + \frac{RT}{\mathrm{n}F}\ln\left(\frac{[\mathrm{Ox}]}{[\mathrm{Red}]}\right)$$

Electrochemical potentials of the flavin and the reference dye in solution are at equilibrium ($E_D = E_{Fl}$), so it is possible to express the Nernst equations of the flavin and the reference dye as equaling each other. (Eq. 2)

$$E_{m,D} + \frac{RT}{\mathrm{n}F}\ln\left(\frac{[D_{ox}]}{[D_{red}]}\right) = E_{m,Fl} + \frac{RT}{\mathrm{n}F}\ln\left(\frac{[Fl_{ox}]}{[Fl_{red}]}\right)$$

From the change in midpoint potential from that of free flavin, the ratio of $K_{d(red)}$ to $K_{d(ox)}$ can be determined using the following two equations:

$$\Delta\Delta G = -zF\Delta E$$

$$\frac{K_{d2}}{K_{d1}} = e^{-\Delta\Delta G/RT}$$

Isothermal titration calorimetry experiments

The ITC experiments were carried out by titrating 100 µM of FMN, FAD, or riboflavin into 5 µM X2B2 or X2B2-L14CU at 20°C (35°C for X2B2-L14CU: FMN titration) on a VP-ITC (MicroCal, GE Healthcare). Heat of dilution titrations were performed by titrating FMN, FAD, or riboflavin into a matching buffer in the same experimental settings. RNA and flavin samples were prepared in 10 mM Tris HCl pH 7.5 and 20 mM $MgCl_2$. The baseline was corrected by subtracting the heat of dilution, and the data were fitted using “one-site” non-linear least square regression. The statistics of thermodynamic parameters (N, K_d, ΔH, ΔS and ΔG) were summarized from three experiments (Table 4-1).

NMR Spectroscopy

X2B2-L14CU: FMN were prepared (200 – 600 μM) and pre-incubated in 10 mM Tris-d_{11}, pD=7.5 and 20 mM $MgCl_2$ at 37°C for 30 min. Two-dimensional ^{1}H-^{1}H NOESY data for fully protonated and site-specific deuterated samples were collected in D_2O (99.96%, CIL) at 308K and 293K. One- and two-dimensional imino proton spectra were collected for X2B2-C14U: FMN in buffer containing 10% D_2O + 90% H_2O. HNNCOSY data were collected for $^{13}C/^{15}N$ labeled X2B2-C14U: FMN and $^{13}C/^{15}N$ labeled X2B2-C14U: riboflavin-(dioxopyrimidine-$^{13}C_4$, $^{15}N_2$). All the NMR data were collected on a Bruker Avance III 800 MHz spectrometer equipped with TCI cryoprobe (NMR Core, University of Missouri). The NMR data were processed by NMRPipe [187] and analyzed by NMRViewJ [188].

Structure Calculation

The X2B2-C14U structure was initially calculated using NMR-derived restraints by CYANA [192]. Standard torsion angle restraints, standard hydrogen bonding restraints for Watson-Crick base pairs, and standard phosphate distances in A-form RNA helical structures were employed to maintain A-form helical geometry in the first 8 base pairs (nt 1-8, and 31-38) [192,193]. For the five base triples standard hydrogen bonding restraints were used based on the secondary structure derived from our NMR analysis. NOE restraints of this region, the five base triples and of residues involved in FMN recognition were also utilized.

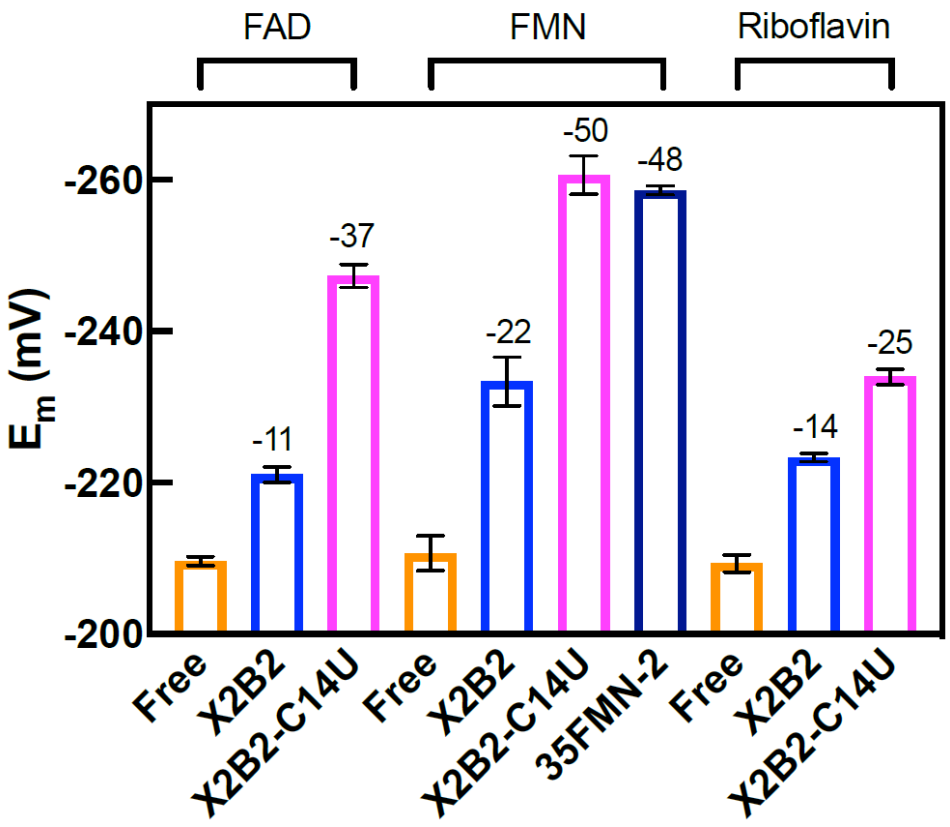


Figure 4-1. Reduction potential of flavin moieties are shifted upon binding by aptamers. Bar graph summarizing reduction potential shift of flavin binding aptamers X2B2, X2B2-C14U, and 35FMN-2. Flavin moieties are listed about the bar graph. Free FAD, FMN, and riboflavin are shown in orange. X2B2 is shown in blue, X2B2-C14U is shown in purple, and 35FMN-2 is shown in dark blue.

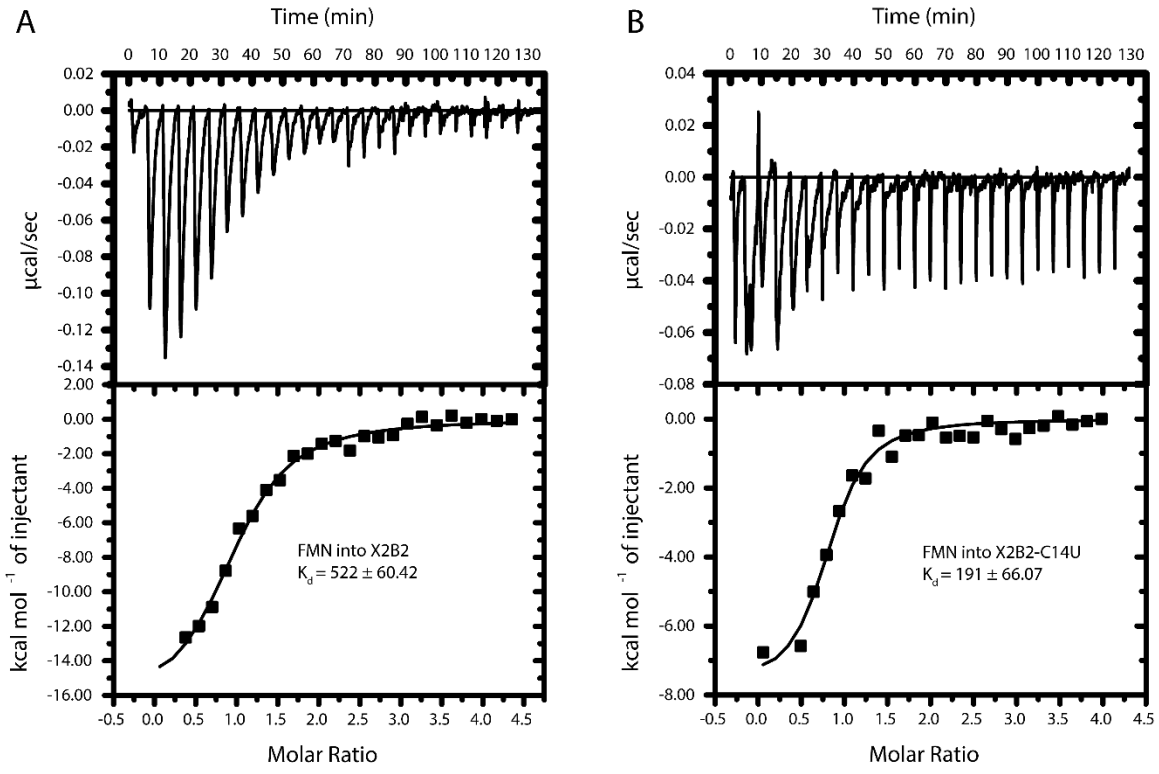


Figure 4-2. Representative titrations of FMN into X2B2 and X2B2-C14U. Left panel shows representative ITC profile of titrating FMN into X2B2. Right panel shows ITC profile of titrating FMN into X2B2-C14U. K_d is representative of three individual titrations and standard deviation is listed in each panel.

FMN into X2B2		
	Average	Standard Deviation
N	1.01	0.17
Kd (nM)	522	60.42
H (cal/mol)	-10850	1270
S (cal/mol/deg)	-23	8.15

FMN into X2B2-C14U		
	Average	Standard Deviation
N	1.00	0.18
Kd (nM)	191	66.07
H (cal/mol)	-6685	1089
S (cal/mol/deg)	8.04	4.50

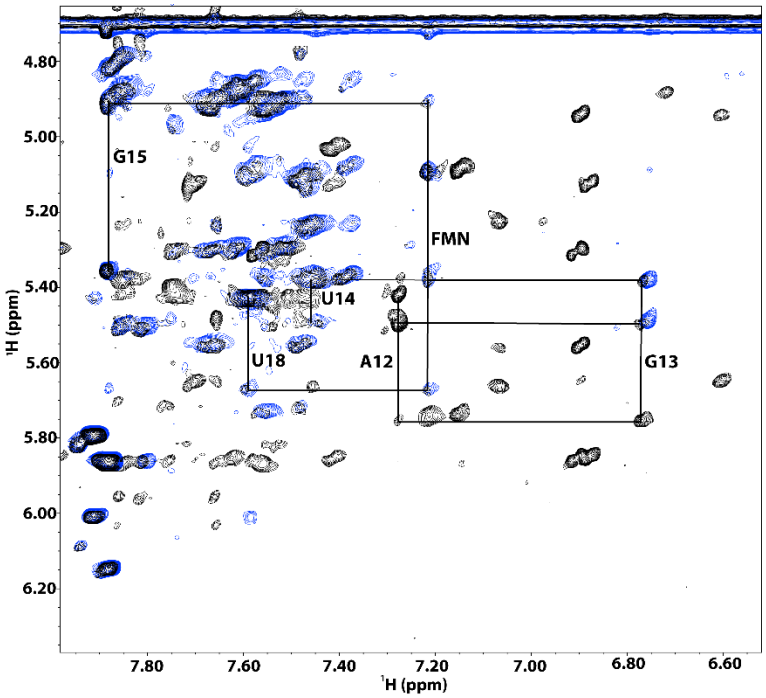
FAD into X2B2		
	Average	Standard Deviation
N	1.64	0.21
Kd (nM)	732	340
H (cal/mol)	-4620	967
S (cal/mol/deg)	12.45	4.15

FAD into X2B2-C14U		
	Average	Standard Deviation
N	1.13	0.05
Kd (nM)	1186	195
H (cal/mol)	-17653	2491
S (cal/mol/deg)	-30.17	8.32

Riboflavin into X2B2		
	Average	Standard Deviation
N	1.26	0.12
Kd (nM)	1240	393
H (cal/mol)	-11897	2974
S (cal/mol/deg)	-13.21	10.46

Riboflavin into X2B2-C14U		
	Average	Standard Deviation
N	0.97	0.21
Kd (nM)	548	202
H (cal/mol)	-9080	1553
S (cal/mol/deg)	-2.20	6.05

Table 4-1. Binding parameters of titration of FMN, FAD, and riboflavin into X2B2 and X2B2-C14U. Parameters listed are representative of 3 individual experiments.


G15
FMN
U14
U18
A12
G13
4.80
5.00
5.20
5.40
5.60
5.80
6.00
6.20
^{1}H (ppm)
7.80
7.60
7.40
7.20
7.00
6.80
6.60
^{1}H (ppm)

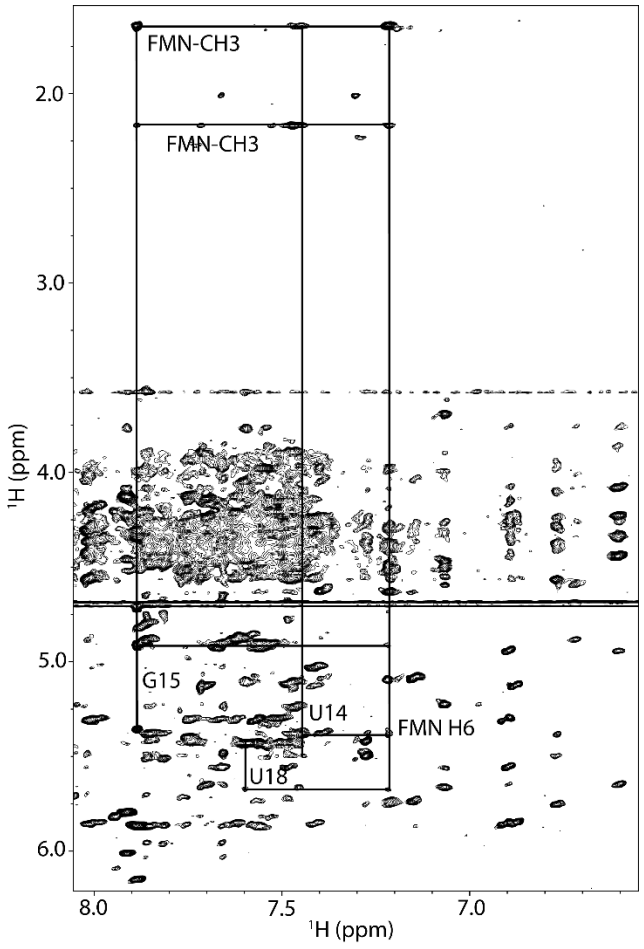

FMN-CH3
FMN-CH3
2.0
3.0
4.0
5.0
6.0
^{1}H (ppm)
G15
U14
FMN H6
U18
8.0
7.5
7.0
^{1}H (ppm)

Figure 4-3. 2D NOESY experiments reveal key residues involved in FMN binding to X2B2-C14U. Top panel shows a selection of the fully protonated 2D ^{1}H-^{1}H NOESY spectrum of X2B2-C14U: FMN in black overlaid with a selection of GU protonated (AC deuterated) 2D ^{1}H-^{1}H NOESY X2B2-C14U: FMN in blue. A NOESY walk of residues involved in FMN binding are shown with black lines and labeled. Bottom panel shows a selection of the fully protonated 2D ^{1}H-^{1}H NOESY with a NOESY walk of residues involved in FMN binding. FMN-CH3 represent protons from the methyl groups of FMN.

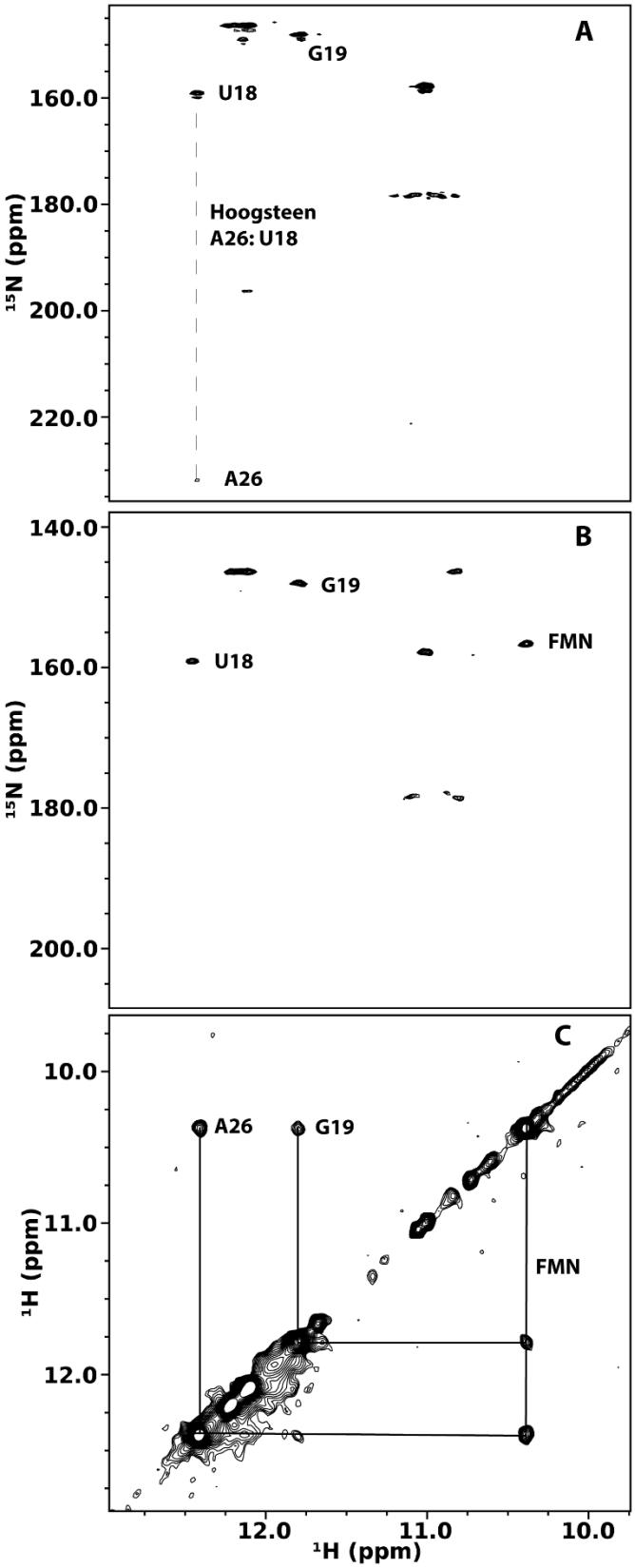

A
G19
U18
Hoogsteen
A26: U18
A26
^{15}N (ppm)
160.0
180.0
200.0
220.0
B
G19
FMN
U18
^{15}N (ppm)
140.0
160.0
180.0
200.0
C
A26
G19
FMN
^{1}H (ppm)
10.0
11.0
12.0
12.0
11.0
10.0
^{1}H (ppm)

Figure 4-4. HNN-COSY and 1H-1H imino proton NOESY experiments reveal base pairings and residues involved in FMN recognition in X2B2-C14U. A) HNN-COSY of ^{13}C, ^{15}N labeled X2B2-C14U: FMN (unlabeled) shows evidence of A26: U18 Hoogsteen edge base pair based on ^{15}N chemical shift. A peak corresponding to an unpaired G is also labeled. B) HNN-COSY of ^{13}C, ^{15}N labeled X2B2-C14U: riboflavin-(dioxopyrimidine-$^{13}C_4$, $^{15}N_2$) reveals a new peak corresponding to the U-like ring of riboflavin/FMN. No additional peaks that correlate with FMN were observed. Note that the A26 peak present in panel A) was visible but only at high contour levels. C) ^{1}H-^{1}H Iminio proton NOESY of X2B2-C14U: FMN shows FMN is near unpaired G and A26: U18 base pair. Cross-peaks shown using black lines.

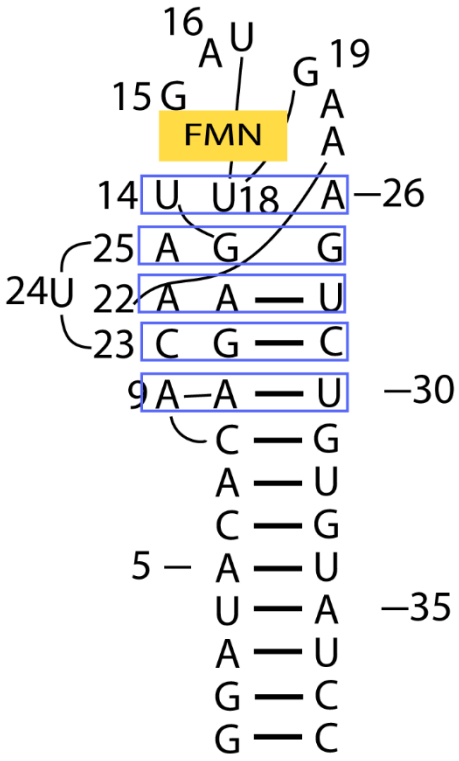


Figure 4-5. Secondary Structure of X2B2-C14U: FMN. Potential base triples are indicated using blue boxes. FMN is shown highlighted in a yellow box.

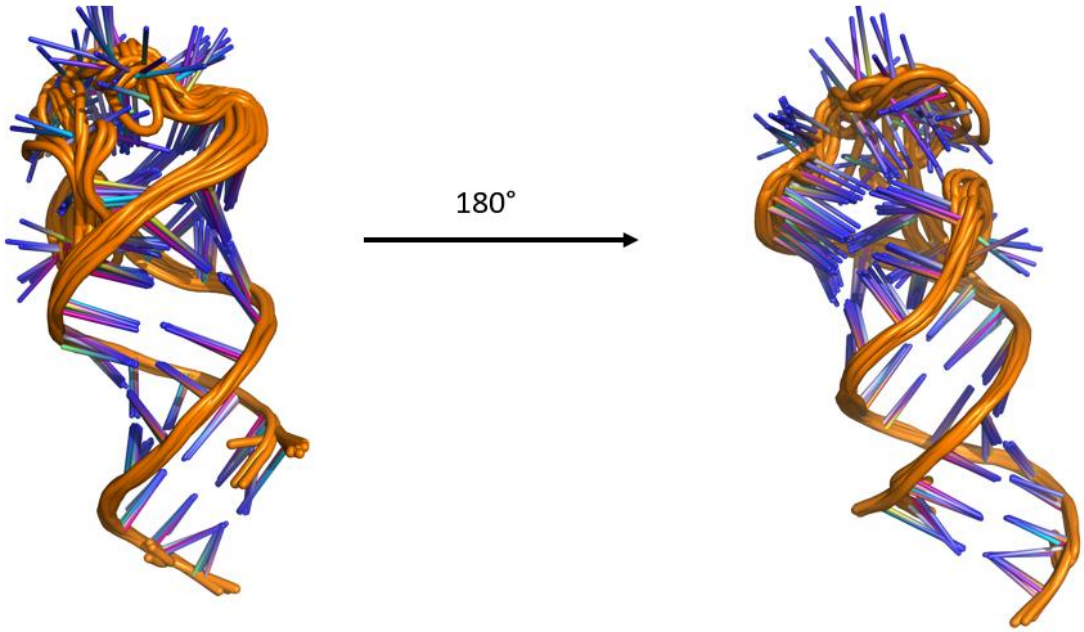


Figure 4-6. Ensemble of 10 structures of X2B2-C14U: FMN calculated using NMR derived restraints by CYANA. Shown in cartoon view. Ensemble rotated 180° in right panel.

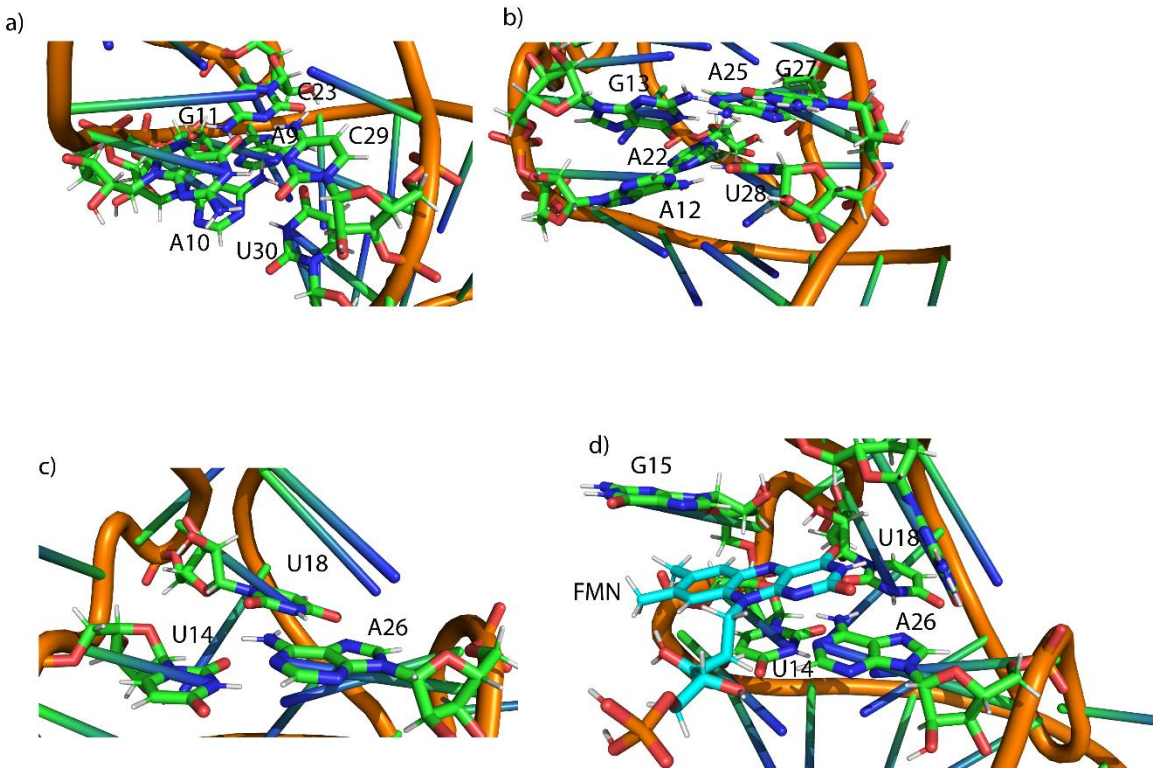


Figure 4-7. X2B2-C14U: FMN adopts non-canonical base pairing in order to bind FMN. A) First two of five base triples in X2B2-C14U: FMN. Nucleotides participating are labeled. B) Third and fourth of five base triples in X2B2-C14U: FMN. Nucleotides participating are labeled. C) Fifth base triple in X2B2-C14U: FMN. Nucleotides participating are labeled. D) FMN binding site in X2B2-C14U: FMN. Nucleotides that interact with FMN are labeled.

Chapter 5. Summary

The work presented in this dissertation combines functional biochemical approaches, biophysical approaches, cell-based approaches, and structural studies to understand how structure influences function in HIV-1 and the origins of life. In Chapter 2 of this dissertation we explored a key virus: host interaction in HIV-1. We showed that HIV-1 gRNA is protected through interactions with the host $tRNA^{Lys3}$. This interaction takes place during the initiation of the hallmark step of HIV-1, reverse transcription. We showed that this mechanism is conserved in two HIV-1 subtypes and were able to confirm its biological significance using numerous cell-based assays. In Chapter 3 of this dissertation, we elucidated the structure of PBS-segment of HIV-1, a conserved region of the HIV-1 5'UTR that is the site of initiation of reverse transcription. We showed that its tripartite structure is necessary for the binding of another important host factor, RHA. RHA specifically recognizes this three-way junction in order to be recruited into the virion to serve as a processivity enhancer of HIV-1 RT and thus to bolster virion infectivity. Chapter 4 examines an ongoing study that shows novel aptamers that shift the reduction potential of bound flavins and could serve as a potential redox ribozymes in support of RNA's proposed role in the origin of life. We were able to characterize the interactions between flavin and these aptamers and build a preliminary structural model that show the key contacts between flavin and the aptamers. In Appendix 1, we were involved in a collaborative project that looked at the molecular mechanisms of an aptamer that is a broad-spectrum HIV-1 RT inhibitor. This work showed that the aptamer bound to RT and modulated its maturation by HIV-1 PR. My part in that work showed that mutations in the RNA aptamer led to a different shape as demonstrated by SAXS envelope and was the cause of mutant aptamers inability to bind RT.

$tRNA^{Lys3}$ is the primer that HIV-1 uses to initiate reverse transcription [44]. $tRNA^{Lys3}$ anneals to the PBS in the 5'UTR of HIV-1. Through long term cell culture experiments attempting

to utilize alternative tRNAs it was shown that the 18-nt PBS sequence wasn't the only determinant of $tRNA^{Lys3}$ utilization in HIV-1 [34]. This early observation led to the identification of other sites where $tRNA^{Lys3}$ could possibly interact with the 5'UTR [34]. Over the years there has been controversy if these interactions occur and their importance. Our work focused on one potential interaction, the interaction between the A-rich loop of the 5'UTR and the anticodon of $tRNA^{Lys3}$. Through mutation of the A-rich loop we were able to show that this interaction is important for viral infectivity. We also examined this interaction using NMR and found indirect evidence of this interaction occurring. Through a combination of in vitro functional assays and cell-based data we showed that the A-rich loop: anticodon interaction is a mechanism that HIV-1 uses to protect its gRNA and is conserved in multiple HIV-1 subtypes.

The nature of the interactions between $tRNA^{Lys3}$ and the HIV-1 5'UTR has been controversial. Varied results between studies may have been due to experimental conditions and the diversity of HIV-1 strains examined. In our work we showed that a key experimental condition, the experimental parameters of $tRNA^{Lys3}$ annealing to the HIV-1 5'UTR led to different complexes. We showed that using NC to anneal $tRNA^{Lys3}$ to the 5'UTR may be necessary for the A-rich loop: anticodon interaction to occur. This could explain the controversial results past groups have obtained and the inability to observe the A-rich loop: anticodon interaction in certain HIV-1 strains.

In our previous work we have shown that RHA is important for viral infectivity and enhances the processivity of reverse transcription in HIV-1 [71,76]. Through a combination of structural, biophysical, computational, and cell-based data we showed that RHA specifically recognizes the shape of the PBS-segment in HIV-1. We showed that this region is highly conserved and mutations that disrupt its three-way junction structure led to lowered infectivity. Using ITC, we were able to show that the PBS contains a high affinity binding site for RHA

dsRBD1 and its interaction is specific as compared to a hairpin-control RNA. As RHA is a key host factor involved in reverse transcription our work suggests that the three-way junction structure serves as a scaffold and drives RHA recruitment into the HIV-1 virion.

The flavin moiety is an important biological cofactor that is involved in many diverse types of chemical reactions [94]. As such it is important biological cofactor and its role in early life could have also been important. In this work we examined novel aptamers that provide new insight into the RNA world hypothesis and show that these aptamers shift the reduction potential of bound flavins. Using NMR based structural studies we were able to obtain a preliminary structure of one of these aptamers. Our analysis provides key molecular details between the interaction between flavin and the aptamer and could help enable future development of novel ribozymes that would give furthers insights into early life.

In summary, this work combines a diverse array of approaches, including functional biochemical studies, structural studies, cell-based assays, and computational approaches to gain further insight into key virus: host factor interactions in HIV-1 and provides insight into the origins of life. We have shown key structural elements in HIV-1 and flavin binding aptamers that are important for their function. This research shows the importance of understanding RNA structure in order to understand biological function.

Appendix 1. Binding interface and impact on protease cleavage for an RNA aptamer to HIV-1 reverse transcriptase

Note: This chapter is based on a research paper that was published in Nucleic Acids Research in March 2020.

Abstract

RNA aptamers that bind HIV-1 reverse transcriptase (RT) inhibit RT in enzymatic and viral replication assays. Some aptamers inhibit RT from only a few viral clades, while others show broad-spectrum inhibition. Biophysical determinants of recognition specificity are poorly understood. We investigated the interface between HIV-1 RT and a broad–spectrum UCAA-family aptamer. SAR and hydroxyl radical probing identified aptamer structural elements critical for inhibition and established the role of signature UCAA bulge motif in RT-aptamer interaction. HDX footprinting on RT ± aptamer shows strong contacts with both subunits, especially near the C-terminus of p51. Alanine scanning revealed decreased inhibition by the aptamer for mutants P420A, L422A and K424A. 2D proton nuclear magnetic resonance and SAXS data provided constraints on the solution structure of the aptamer and enable computational modeling of the docked complex with RT. Surprisingly, the aptamer enhanced proteolytic cleavage of precursor p66/p66 by HIV-1 protease, suggesting that it stabilizes the productive conformation to allow maturation. These results illuminate features at the RT-aptamer interface that govern recognition specificity by a broad-spectrum antiviral aptamer, and they open new possibilities for accelerating RT maturation and interfering with viral replication.

Introduction

Aptamers are nucleic acids that can be selected via Systematic Evolution of Ligands by EXponential Enrichment against specific target(s). RNA aptamers selected to bind HIV-1 reverse

transcriptase (RT) inhibit the protein's enzymatic activity in biochemical assays and they block HIV replication in cell culture[118,209–214]. The inhibitory function of aptamers against RT enzymatic activity comes from their ability to compete with viral primer/template (p/t) for RT binding [209,215–219], although many of the molecular details of the interactions between RNA aptamers and various RTs are still poorly understood. In addition to binding nucleic acids, RT contacts a network of viral and host proteins during HIV-1 replication. Alteration of these interactions could potentially provide additional mechanisms for aptamer-mediated interference with viral replication, a possibility that we explore here with respect to protease-mediated maturation. Aptamers can bind the mature heterodimer RT, and recent evidence suggests that they also bind to the precursor homodimer in the cytoplasm during viral assembly [214]; therefore, we reasoned that they could modulate protein-protein interaction involving RT.

For DNA aptamers, efforts have been made to elucidate the RT-binding interface using different approaches, including crystallization and mass spectrometry footprinting [216,220]. In contrast, information on the interface between RT and RNA aptamers is still limited, with most structural studies only focused on RT complexes with pseudoknot aptamers such as T1.1 [216,221–223], which are known to be sensitive to RT amino acid sequence variations [214,224]. Several structural families of anti-HIV RNA aptamers have been described based on conserved signature motifs, including family 1 pseudoknots (F1Pk), family 2 pseudoknots (F2Pk), 6/5 asymmetric loop motif ((6/5)AL) and UCAA-bulge motif (UCAA) [135,209,213,214,224,225]. F1Pk aptamers are highly specific for RTs that encode arginine at position 277, as K277 RTs are not susceptible to inhibition by F1Pk pseudoknots such as aptamer T1.1 [214,224]. In contrast, UCAA and (6/5)AL aptamers can inhibit RTs from diverse lentiviruses and thus have been considered as broad-spectrum inhibitors [135,213,214].

Aptamers from each structural family likely make distinct molecular contacts, and the broad-spectrum aptamers may recognize conserved regions among phylogenically diverse RTs. Information on RT-aptamer binding interfaces from different aptamer structural families will provide insight for understanding the mechanism of broad-spectrum inhibition and for engineering nucleic acid tools for differential recognition of HIV-1. Here, we have defined critical RNA–protein molecular interactions for a broad-spectrum RNA aptamer from both the aptamer and RT perspectives, focusing on 148.1, a UCAA-family aptamer that emerged from a 'PolyTarget' selection against a panel of RTs from different HIV strains, including HIV-1 Group M subtypes A, B, and A/E, HIV-1 Group O, and HIV-2 [135]. The UCAA motif definition includes two conserved base pairs (AC/GU) on the 5' side of the unpaired UCAA within a relatively simple stem-loop structure. The broad-spectrum aptamer 148.1t1 [135] is the smallest UCAA variant (44 nt) tested to date that satisfied the essential requirements of the motif, making it a promising subject for structural studies. Using biochemical and chemical approaches, we identified the 38 nt-binding core of aptamer 148.1t1 (named 148.1-38m) and elucidated the interface of the complex between RT and aptamer 148.1-38m. Alanine scanning mutagenesis of this region revealed decreases in susceptibility for specific mutant RTs toward inhibition by 148.1-38m. 2D NMR and SAXS established structural features of the apoRNA in solution, and computational modeling revealed a 3D structure of the bound complex. Furthermore, we investigated the ability of the aptamer to interfere with RT-PR interactions. Information on RT-aptamer binding interfaces from different aptamer structural families will provide insight for understanding the mechanism of broad-spectrum inhibition.

Results

Determination of the binding core of broad-spectrum aptamer 148.1

To investigate the RNA structural features required for binding of aptamer 148.1-38m to RT, mutated and truncated variants were evaluated for their abilities to inhibit purified RT (Figure A-1; Supplementary Table AS1-1 and Figure AS1-1). Based on the predicted secondary structure of aptamer 148.1t1, an additional 9 nt segment on the 5' end of this aptamer was replaced with GGG fragment (to aid transcription efficiency) to create a 38 nt version denoted 148.1-38m. Aptamer 148.1-38m retained the ability to bind and inhibit RT, with a normalized fraction extended product (NFEP) value <0.2 (Figure A1-1 and Supplementary Figure AS1-1A). RT inhibition was not significantly sensitive to base pairing or deletion of the 5'GGG overhang (Supplementary Figure AS1-1) or to replacing the AAUCU loop with GGA (Figure A1-1A and A1-1B). In contrast, inhibition was lost upon deletion of the two unpaired Us (U6 and U14) or pairing them with As inserted into the opposite strand (NFEP values 0.78 and 0.88, respectively) (Figure A1-1B). A representative gel for primer extension assay is shown in Supplementary Figure AS-S2. Deleting U6 or replacing it with C abolished inhibition (Figure A1-1C), and the variant with U6 deletion (148.1-38m delU6) served as a non-binding control in later binding studies. Although deleting U14 did not affect aptamer inhibition, changing U14 to C abolished the inhibition (Figure A1-1D) potentially by changing the overall predicted structure. When U14 was replaced by C, the signature UCAA bulge is predicted to become an AC/UUCAA asymmetric internal loop in the U14C mutant but is predicted to be maintained in the U14 deletion mutant (Supplementary Figure AS1-3). The signature UCAA bulge was also sensitive to modifications, as single variations at any of the four position (12 variants) and a scrambled variant of the bulge, AACU, also abolished inhibition of RT (NFEP values > 0.6) (Figure A1-1C and Supplementary Figure AS1-4). The results from modification studies indicate that the unpaired U6 and the UCAA bulge are critical to 148.1-38m's inhibition of RT, while the stem loop was not highly sensitive to the alteration tested here. Although findings of 148.1-38m's UCAA

bulge and stem loop are consistent with the previously identified consensus for UCAA aptamers 80.103 and 80.111 [213], aptamer 148.1-38m differs in the role of the first unpaired U from the aptamers' 5'-end, as deletion of the unpaired nucleotides of aptamer 80.103 did not affect inhibition [213].

UCAA bulge of 148.1-38m does not directly interact with RT

We next looked at the RNA–protein interface by identifying nucleotides with changes in solvent accessibility of their sugar moieties upon binding to RT. Aptamers 148.1-38m and 148.1-38m delU6 were subjected to hydroxyl radical cleavage in the presence or absence of RT (Figure A1-2 and Supplementary Figure AS1-5), and cleavage intensities at each position were compared for free and bound aptamer under identical treatment [216,226–228]. The non-binding control aptamer, 148.1-38m delU6, failed to bind (Supplementary Figure AS1-5A) or inhibit RT (Figure A1-1C), and its hydroxyl radical cleavage pattern was unchanged by the addition of RT. A quantitative representation of cleavage patterns clearly identified differences in cleavage intensities between bound and unbound aptamer 148.1-38m (Figure A1-2 and Supplementary Figure AS1-5B). Addition of protein induced a hypersensitive cleavage site at the sugar of nucleotide A30 within the UCAA bulge. Sugars of the neighboring nucleotides A31 and C29 and nucleotides A19 and U22 of the stem loop were also sensitive to cleavage, albeit to a much lesser degree (Figure A1-2A). These observations indicated that the sugar moieties of the nucleotides in the UCAA bulge become more exposed to solvent when the aptamer is in complex with RT. In contrast, RT protected nucleotides C9 to C13 opposite the UCAA bulge and G25 to U27 adjacent to the UCAA bulge from cleavage. These data suggest that the UCAA bulge might not be buried in the complex but rather remains exposed and shapes the aptamer's 3D structure, helping the two stems of the aptamer adopt a conformation that is appropriate for docking into the RT binding pocket.

RT makes distinct interactions with specific and broad-spectrum inhibitors

Aptamer binding surfaces on RT were identified by hydrogen deuterium exchange coupled with mass spectrometry (HDX/MS), using complexes formed between RT and 148.1-38m. As a reference, HDX/MS was also performed using the minimum binding core of F1Pk aptamer 70.05 (70.05core2) [209,224] (Supplementary Figure AS1-6). We predicted that the two complexes would leave different footprints on RT due to the differences in aptamer structure. For the HDX/MS experiment, the p51 subunit of RT was isotopically labeled with ^{15}N so that digested peptides from p51 could be differentiated from those of p66 [229,230].

Differences in HDX/MS for the RT heterodimer in the presence and absence of aptamers 148.1-38m and 70.05core2 were mapped onto the atomic coordinates of the RT crystal structure without substrate (PDB ID: 1DLO) [231] (Figure A1-3A). Aptamer 148.1-38m protected p66 in the fingers, palm, and C-terminal region of the thumb and the C-terminal of RNase H domain, in addition to protecting p51 in the fingers, palm and thumb subdomains and near the C-terminal domain (Figure A1-3A and Supplementary Figure AS1-7). Aptamer 70.05core2 protected p66 peptides located in the fingers, N-terminal region of the palm and the thumb subdomain and the N-terminal region of the RNase H domain, in addition to protecting p51 in the finger and connection subdomains (Figure A1-3A and Supplementary Figure AS1-8). Although there was some similarity in the footprints of the two aptamers, each aptamer formed a distinct set of contacts with the target RT protein, especially within the p51 subunit. These results are consistent with previous observations that both aptamers compete with p/t for binding to RT [209,213,216,219]. Protection at peptides covering position 277 was observed in both complexes but were notably were higher in RT-70.05core2 than in RT–148.1 complex (Supplementary Figure AS1-9), consistent with the finding that R277K mutation abolished the inhibitory function of F1Pk aptamers but not UCAA aptamers [213,214,224]. In the absence of

nucleic acid, RT exists predominantly in a fingers-and-thumb-together ('closed') conformation [231–234]. Previous HDX studies of interactions between RT and efavirenz, a non-nucleoside RT inhibitor, observed exposed regions in the p66 fingers and thumb domains that were associated with formation of an 'open' conformation between the two domains [230]. We found no evidence of similarly exposed peptides in either RT–aptamer complex, suggesting that our observation is not linked to conformational changes in the finger and thumb domains.

P420A, L422A and K424A mutants are less susceptible to 148.1-38m inhibition

Two regions of RT were strongly stabilized against HDX upon binding to aptamer 148.1-38m. Region 1 (positions 418-NTPPLKL-425) in the p51 subunit was protected to a remarkably high degree by aptamer 148.1-38m, and also modestly by aptamer 70.05core2 (Figure A1-3B and Supplementary Figure AS1-7B). Within this region, positions 419, 420, 422 and 425 are highly conserved among phylogenically diverse RTs. The remaining positions (418, 421, 423 and 424) were conserved within group M (HXB2, NL4-3, 93TH and 94CY) and SIVcpz (SIV-US) (Supplementary Figure AS1-10A). Region 2 (positions 534-AWVPAHGIGGNEQVDKLVSAG-555) in the p66 subunit was only protected by aptamer 148.1-38m but not by aptamer 70.05core2, although the magnitude of the protection was less than in Region 1. Almost all amino acid in this region are conserved among RTs from different HIV-strains [235].

To explore the significance of Region 1 in RT–aptamer interactions, each amino acid was individually mutated to alanine in the p51 subunit (Supplementary Figure AS1-10B). The singly substituted p51 subunits were co-expressed with wild-type p66 and purified to yield eight enzymatically active mutant RTs (N418A, T419A, P420A, P421A, L422A, V423A, K424A and L425A) (Supplementary Figure AS1-11). Aptamers 148.1-38m and 70.05core2 were both able to inhibit all eight mutants, at least to some degree; however, mutants P420A, L422A and K424A were noticeably less susceptible than the other RTs to inhibition by aptamer 148.1-38m (NFEP

> 0.7, Figure A1-4A), while all eight mutants remained susceptible to 70.05core2 inhibition (NFEP 0.15 to 0.45). The difference in inhibition by aptamers 148.1-38m and 70.05core2 is especially evident for RT mutants L422A and K424A (P = 0.0016 and 0.0004, respectively) and for N418A (P = 0.02), even though both aptamers inhibit N418A with NFEP < 0.6. These observations underscore the functional significance of differences in molecular contacts between aptamer structural families. Nevertheless, none of these alanine substitutions yielded the degree of resistance obtained with an R277K mutant RT, which extended essentially all of the input p/t to full-length product in the presence of 70.05core2, even while remaining susceptible to inhibition by 148.1-38m (Figure A1-4A). Polymerase activities of the three mutants (P420A, L422A and K424A) were similar to, or slightly less than, that of WT RT (Figure A1-4B and Supplementary Figure AS1-12), ruling out the possibility that their reduced sensitivities to inhibition by 148.1-38m could be due to enhanced abilities to bind or extend the DNA substrate. Mutant T419A was also included as a internal control and its polymerase activity was similar to that of WT RT. Overall, the alanine-scanning mutations of Region 1 caused more severe effects on the enzymatic inhibition by 148.1-38m than by 70.05core2, consistent with the significantly higher protection of this fragment in the 148.1–38m complex than in the RT–70.05core2 complex. Amino acids 420, 422 and 424 of p51 appear to be involved in establishing the RT–148.1–38m interface.

When both aptamers were titrated against four RTs (WT HXB2, T419A, P420A and R277K), aptamer 148.1-38mer inhibited P420A weakly compared to its inhibition of HXB2, T419A and R277K (Figure A1-5A). Inhibition of P420A was partially rescued at high concentration of aptamer. This weak resistance is consistent with the slight decrease in binding affinity of 148-38m for P420A compared to WT HXB2 (Supplementary Figure AS1-13). In contrast, RT mutant R277K was not susceptible to inhibition by aptamer 70.05core2 at any tested concentration (up

to 180 nM), while all other RTs were equivalently inhibited in a concentration-dependent manner by this aptamer (Figure A1-5B), again highlighting differences in amino acid sequence sensitivities for aptamers from different structural families. Since full-length aptamer 70.05 had been previously demonstrated to be unable to bind R277K [214], the strong resistance phenotype of R277K mutant against 70.05core2 was expected.

Structural constraints for aptamer 148.1-38m in solution

2D ^{1}H-^{1}H NOESY NMR spectra were collected for aptamer 148.1-38m and peaks were assigned (Supplemental Figure AS1-14) to establish initial structural constraints for the free (apo) RNA. Base stacking interactions were observed for residues in the lower and upper stems. Residues U6, A19-U22 and C29 gave rise to strong cross-peaks, indicating that these residues are flexible in the structure. Interestingly, cross-peaks of U14 were broad, suggesting that it may be engaged in ordered structure rather than acting as a flexible bulge. An NOE between A30-H2 and U14-H1' was observed, and the H8 of G25 was shifted unusually upfield (Figure A1-6A). These data suggest that the UCAA bulge likely makes tertiary contact with U14 and residues nearby. To examine the impact of U6 and U14 on the structure of the aptamer, we collected 2D ^{1}H-^{1}H NOESY data for delU6+delU14. The chemical shifts of residues neighboring U6 and U14 in aptamer 148.1-38m exhibited large changes due to the deletion (Figure A1-6B), indicating a role for these residues in re-organizing local structure. U16-A24 remained unaffected, demonstrating that these U deletions had minimal impact on the structure of the apical loop region. These observations are consistent with RT inhibitory assays showing that the GGA stem loop mutant did not reduce the RT inhibitory function of the aptamer (Figure A1-1B). The chemical shifts of UCAA and their neighboring residues were sensitive to the U6 and U14 deletions, even though they were distant from the deleted Us in primary sequence and in the folded secondary structure, consistent with the tertiary contact suggested above.

To determine whether deletion of U6 and U14 altered the structure of the 148.1-38m, in-line small angle X-ray scattering (SAXS) data were collected. SAXS provides global information about RNA size and shape in solution. Deletion of both U6 and U14 altered the SAXS scattering profile (Figure A1-7A). Comparison of the pairwise distribution plot shows that 148.1-38m has a slightly longer max dimension (*x*-axis intercept in Figure A1-7B). To further understand the structural differences between 148.1-38m and 148.1-38m dU we used DAMMIF to reconstruct *ab initio* models of each RNA. Comparison of these models shows an alteration of the RNA structure upon deletion of the two Us. Specifically, 148.1-38m has a more compact structure than 148.1-38m dU. Deletion of U6 and U14 disrupts the tertiary contacts between the top stem with the UCAA bulge, leading to a more extended RNA fold not favored by RT binding (Figure A1-7C).

Docking model of aptamer 148.1-38m with HIV-1 RT

An iterative strategy of experimental constraints and computational modeling was employed to obtain a 3D model of aptamer 148.1-38m in complex with HIV-1 RT. The 3D structure of RNA aptamer 148.1-38m was computationally modeled with the assistance of the observed NMR experimental data. Secondary structures of 148.1-38m derived from NMR data (Supplementary Figure AS1-15A) and predicted from sequence by using the free energy-based approach Vfold2D model [161,236] (Supplementary Figure AS1-15B) essentially agreed with the representation in Figure A1-1A. Therefore, the secondary structure inferred from NMR data and the initial 3D structures from the templated-based Vfold3D model [161,237,238] were used as input for the IsRNA coarse-grained molecular dynamics simulation method [239] to generate a set of putative 3D structures (see Supplementary Figure AS1-16 for representative 3D apo RNA conformations). The simulated coarse-grained structures were then used as input to generate

refined all-atom 3D structure models using NAMD simulations with energy minimization [240] based on the CHARMM force field [241] for subsequent docking modeling.

The structure of the complex between aptamer 148.1-38m and HIV-1 RT was then predicted through molecular docking. Specifically, 70 modeled all-atom 3D structures of aptamer 148.1-38m obtained from the above computational study were individually docked onto the crystal structure of the HIV-1 RT protein (PDB ID: 1DLO [231]) using our in-house docking program, MDockPR [164]. Using a 6-degree orientation sampling interval, 54 000 docking poses of protein–RNA complex structures were generated for each of the 70 modeled aptamer structures, for a total of 3 780 000 potential poses. These docking poses were further filtered with five experimental constraints: First, Region 1 (N418 - L425 in the p51 subunit of HIV-1 RT) should have at least one close contact (≤5Å) with the aptamer, based on the observed HDX protection in this region. Second, U6 of the aptamer should have at least one close contact, based on the observations above that U6 deletion or substitution resulted in loss of the inhibitory function. The filtering resulted in 47 425 candidate poses, which can be categorized into two distinct types of binding modes. One is characterized by a single contact between the RNA aptamer and the protein, where only a cluster of nucleotides that are close to each other in space forms the single binding interface with the protein. The other type of binding mode is characterized by multiple contacts with the protein by more than one nucleotide cluster. Only poses of the multiple-contact mode were retained, because it is more typical of physiological interactions, as evidenced by the supermajority of protein/RNA complexes deposited in the PDB. The remaining 12 400 complex structures were cascaded to two extra filters, which imposed the constraints that the RNA should have at least one close contact (≤5Å) with Region 2 (A534 – G555 of p66) in Figure A1-3 (based on strong HDX protection in this region), and none of the sugars of the UCAA bulge is in contact with the protein (based on hydroxyl radical probing).

At last, the remaining 2807 candidate poses were restricted to having the termini of the aptamer pointing away from the protein to allow for longer stems on either end of the aptamer, based on the observed tolerance for such extensions. The 1255 surviving candidates were ranked by ITScorePR [164,200] and then further consolidated into 301 clusters, defined as groups of structures with an L-RMSD less than 12Å from each other, are represented by a representative. We examined the top 20 best-scored clusters and selected the final complex structure. The resulting final model was locally minimized with the ITScorePR score, and the optimized structure is reported as the predicted complex structure of aptamer 148.1-38m and HIV-1 RT (Figure A1-8).

As shown in Figure A1-8, the aptamer contacts both Region 1 and Region 2 of the protein. The overall shape complementarity in the complex structure is moderate, with a buried solvent accessible surface area of ~1900Å^2 upon binding, which is ~27% less than the 2600Å^2 that is buried for a pseudoknot aptamer in complex with RT [222]. A number of amino acids are predicted to lie within 5 to 8 Å of the aptamer (Supplementary Table AS1-2) and are thus in position to contact the aptamer directly. These contacts include potential specific interactions between the U6 and the His361 of subunit p66 (<5Å) and between C29 and Arg284 of p66 (<8Å). Fitting the RNA model selected for RT docking against the experimental SAXS data above calculated a χ^2 value of 0.2, suggesting that 148.1-38m RNA does not experience a major shape change upon RT binding.

148.1-38m enhanced protease cleavage of p66 homodimer

During the HIV replication cycle, the immature HIV RT is produced as a homodimer with two p66 subunits. Then the p51 subunit is formed by the proteolytic cleavage of one of the p66 subunits [242,243]. Although p51 consists of the same amino acids as p66 except for the cleaved region, the domains of two subunits are oriented differently in space. The p66 subunit is

responsible for the enzymatic function of RT whereas p51 plays a structural role and aids in binding to nucleic acids. The cleavage of p66/p66 to form p66/p51 is important for the maturation of HIV and is catalyzed by HIV-1 protease (PR) [243,244]. Previous studies showed that interruption of this proteolytic cleavage led to virion instability [245]. Because the PR cleavage site that gives rise to p51 is buried in the mature heterodimer, it has been suggested that the p66/p66 precursor unfolds or adopts different conformations along its maturation pathway [243,244,246–248].

The strongly protected 418–425 peptide of p51 is located 15 amino acids away from the PR cleavage site of p66. Therefore, *in vitro* PR cleavage assays were performed to test the impact of aptamer 148.1-38m on RT maturation. When p66/p66 was incubated with PR in the absence of aptamers, cleavage of p66 homodimer was inefficient (Figure A1-9 and Supplementary Figure AS1-17, lane 2), similar to previous observations [249]. A modified version of full-length aptamer 148.1 was used in which the pyrimidines (Y) were replaced with 2'-F-substituted pyrimidines (148.1-2'FY) to reduce aptamer degradation during the assay. The 2'FY modification did not interfere with aptamer-p66/p66 or aptamer-p66/p51 interactions (Supplementary Figure AS1-18). Surprisingly, the presence of 148.1-2'FY significantly enhanced PR cleavage of p66 (Figure A1-9A, lane 3). Another broad-spectrum aptamer, single-chain DNA RT1t49(-5) [250], may also enhance cleavage slightly, although the stimulation is not statistically significant. Our results suggested that interaction with 148.1 aptamer might drive p66/p66 toward the productive conformation, allowing proteolytic processing of p66/p66 homodimer to p66/p51.

Discussion

In vitro inhibition of RT enzymatic activity by nucleic acid aptamers derives from competition with p/t for binding to RT. Thus, details of their interactions with RT are crucial for

understanding the molecular basis of inhibition. The UCAA aptamer family was previously identified as broad-spectrum, capable of inhibiting phylogenetically diverse RTs [135,213,214]. This work extends those observations, revealing details of the interaction between HIV-1 RT and 148.1-38m, representing the UCAA aptamer family. The complementary structural information obtained in our study from both the RNA and protein perspectives provides critical guidance for understanding the RT–aptamer interface and the structural basis of RT–nucleic acid specificity. In the absence of nucleic acids, RT is in a 'closed' conformation with the fingers and thumb domains together [233,234], but it adopts an 'open' conformation with the fingers and thumb domains apart in the presence of dsDNA, DNA/RNA duplex [232,251,252] or non-nucleoside reverse transcriptase inhibitors (NNRTIs) [222,253–257]. The rearrangement from closed to open conformation when RT is in complex with an NNRTI was evident in prior HDX/MS analyses [230,258]. When RT is in complex with RNA pseudoknot aptamer T1.1, the 'closed' conformation is maintained [216,222], with the tips of the finger and thumb domains in contact [222]. The fact that no region of increased exposure was evident in the present study when RT binds to either 148.1-38m or 70.05core2 suggests that both of these aptamers bind to the fingers and thumb domains together 'closed' conformation of the apo-protein. Aptamer 148.1-38m also appears to remain close to its solution structure upon binding RT. To allow for the possibility of induced-fit binding (RNA changes structure upon RT binding), all of the predicted 3D models for the apo-RNA were included for protein docking. However, the RNA structure in the final complex model is consistent with the SAXS data, suggesting a nearly-rigid fit binding with only minor local rearrangements to expose the sugar of A30 to solvent and enhance its sensitivity to hydroxyl radicals.

Our HDX/MS data show significant differences in the level of protection between footprinting of RT-148.1-38m and that of RT-70.05core2 in peptides comprising Region 1 (418-

425 of p51) and Region 2 (534–555 of p66). The fact that footprinting of F1Pk aptamer 70.05core2 was found in p51 is consistent with the prior observation of an extensive interaction between another F1Pk aptamer and p51 in an aptamer-RT co-crystal [222]. Amino acids 420, 422 and 424 are especially important in establishing the RT–148.1–38m interaction, since alanine mutation at these positions diminished aptamer inhibition *invitro*. The partial resistance phenotype was specific to 148.1–38m, and did not impair inhibition by F1Pk aptamer 70.05core2. The amino acid sequence of the 418–425 peptide is highly conserved among naïve and NRTI-treated patients [235]. Alignment of this region among phylogenetically diverse RTs also shows a high level of conservation with variations found only at three positions (N418S, V423I and K424R), and even then only for MVP5180 and EHO, the two RTs that are most distant from group M subtype B (HXB2) in the panel (Supplementary Figure AS1-10). The charge and hydrophobicity of the peptide were maintained for each of these natural variations. Interaction with RT at this conserved region might be important for RT recognition of 148.1–38m, allowing the aptamer to inhibit a panel of different RTs and potentially slowing the emergence of viable, aptamer-resistant virus.

Aptamers have been used as tools to study protein–protein interactions [259,260], but never as a tool for exploring interactions between viral proteins. During the viral life cycle, HIV-1 RT interacts with many different viral and host proteins, including HIV-1 integrase (IN), nucleocapsid (NC) and PR. RT-IN interaction during reverse transcription and integration can enhance RT activity by stimulating initiation and elongation of viral DNA synthesis [261–263], while RT interactions with IN can inhibit DNA 3'-end processing by IN [264]. RT-NC interaction increases the proportion of long cDNA transcripts by RT [265], stimulates DNA strand transfer reactions and modulates RNase H activity [266,267]. Interaction between RT and PR is critical for RT maturation [242,244], proteolytic processing of the RT-IN cleavage site [268], RT activity

[269] and PR activity [270]. Cleavage of the p66/p66 homodimer by purified PR to yield p66/p51 heterodimer *in vitro* has been previously reported to be slow and inefficient [249], and this processing may require partial unfolding of the beta strand β1 of the RNaseH domain [243,244,247–249,271]. Our data indicate that RNA aptamer 148.1–38m and potentially DNA aptamer RT1t49(-5) both enhance PR cleavage of p66/p66. The ability of anti-RT aptamers to increase PR cleavage of p66/p66 *in vitro* has not been previously reported. However, tRNA was shown to enhance the proteolytic processing of p66/p66 homodimer to p66/p51 [248]. Previous studies reported that non-nucleoside reverse transcriptase inhibitors could induce premature activation of PR via stimulation of Gag-Pol dimerization [272,273]. It may be possible to exploit the effect of aptamers on RT–PR interaction for studying and manipulating HIV maturation by PR, a critical process that is still not fully understood. Aptamers have been exploited as inhibitors of reverse transcription. However, their potential to interrupt the network of interactions among viral proteins during integration, maturation, packaging and other steps in viral life cycle could open a new avenue for blocking viruses and for studying biological processes during infection.

Author Contributions

For this work I collected and analyzed SAXS data on the wild type and mutant aptamer.

Experimental Procedures

Purification of RT and the RT mutants

Plasmids expressing mutant RT were generated using Phusion site-direct mutagenesis on a modified pRT-Dual plasmid (originally containing HXB2 RT) [135]. Mutations were confirmed by Sanger sequencing (University of Missouri DNA Core Facility). RT expressing plasmids were heat shock transformed into *Escherichia coli* BL21(DE3)pLysS competent cells and cells were incubated overnight at 37°C on LB-Agar plates (35 μg/ml streptomycin). Clonal

isolates from each plate were recovered and incubated for ~16 h at 37°C with 225 rpm shaking in 10 ml of 2xYT broth supplemented with 35 µg/ml streptomycin. Cultures were then diluted into 1 L of 2xYT broth with streptomycin. Bacteria were allowed to grow until A_{600nm} reached 0.8. Protein overexpression was induced with 1 mM IPTG. After a 4-h incubation, samples were centrifuged for 15 min at 4000 xg. Supernatants were decanted and cell pellets were frozen at −80°C until purification. For purification, cell pellets were resuspended in lysis buffer (25 mM Tris–HCl pH 8.0, 500 mM NaCl, 1 mM PMSF, 0.15 mg/ml lysozyme) and subjected to four rounds of ultrasonication, with 30 s of sonication on ice, followed by 2 min of rest on ice in each round. Cell lysates were then centrifuged for 15 min at 4°C and 12 000 × *g* to remove cell debris. Cell extracts were passed through a 0.45 µm filter to remove insoluble material and applied to a Ni-NTA agarose affinity column (Qiagen) for His tag purification. Purification was performed according to the manufacturer's protocol with an added 40 ml of high salt wash (2 M NaCl) to remove any endogenous nucleic acids bound to RTs. Eluted protein samples were pooled and quantified using UV absorbance at 280 nm on a NanoDrop 1000 spectrophotometer (Thermo Fisher) using an extinction coefficient at 280 nm of $2.6012 \times 10^5\ M^{-1}\ cm^{-1}$. Purified proteins were concentrated and exchanged into 2× storage buffer (100 mM HEPES pH 7.5, 100 mM NaCl) using Amicon Ultra Centrifugal Filters (Millipore Sigma). Proteins were stored at −80°C after addition of glycerol to 50% (v/v).

Primer extension assay

DNA-dependent DNA polymerase activities of RTs were performed as previously described [135]. In brief, 20 nM of RT (in-house purified HXB2 and its mutant variants) was incubated with 100 nM of aptamer in extension buffer (50 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$) on ice for 10 min, then a mixture of primer/template/dNTP was added (final concentration: 10 nM primer, 20 nM template and 100 µM of each dNTP) [135]. DNA-

dependent DNA polymerase activities of RTs were assayed using a 31-nt DNA template (5'-CCATAGATAGCATTGGTGCTCGAACAGTGAC-3') and a complementary, 5'-Cy3-labeled 18-nt DNA primer (5'-Cy3-GTCACTGTTCGAGCACCA-3'). The extension reaction was incubated at 37°C for 10 min. Reaction was stopped by adding 2× polyacrylamide gel electrophoresis (PAGE) loading dye (95% formamide, 50 mM ethylenediaminetetraacetic acid (EDTA), xylene and bromophenol blue). Samples were separated on 10% denaturing (8M urea) PAGE. Gels were scanned on a Typhoon FLA9000 imager (GE Healthcare) and quantified using Multi Gauge software (Fujifilm). *P*-values were calculated using unpaired *t*-test computed by GraphPad Prism.

RT polymerase activity assays

Polymerase activity of RT was monitored via time course primer extension assays essentially as above, except that reactions were stopped at various time points (0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 min) prior to analysis by denaturing gel electrophoresis as above. The fraction of primer converted to full-length product as a function of time was plotted using GraphPad Prism.

K_d determinations

Aptamers of interest were 5'-radiolabeled using [^{32}P]ATP and polynucleotide kinase. Labeled aptamer (~40 000 cpm) was added to binding reactions containing various concentrations of subtype-B RT. Binding buffer contained 140 mM KCl, 1 mM $MgCl_2$ and 50 mM Tris–HCl pH 7.5 (final concentration). Binding reactions were mixed on ice, then moved to room temperature and incubated for 15 min before loading onto pre-wet nitrocellulose membranes under vacuum and washing with 1× binding buffer. Aptamer–RT complexes were retained on the membrane, and radioactivity was determined via scintillation counting. The 'no RT' reaction was used as a control to determine protein-independent contribution to signal (typically <3%) and the amount of radioactivity present in the unfiltered sample was set to 100%. The protein-dependent signal was determined by subtracting the value obtained for the no RT control from

the value obtained for the sample. Values at each RT concentration were fit to a one-site specific binding graph Y = B_{max}*X/(K_d + X) using Prism software version 6.2. In the equation, B_{max} is the maximum fraction bound, K_d is the dissociation constant, X is RT concentration and Y is fraction bound.

Hydroxyl radical footprinting

Hydroxyl radical footprinting reactions were performed using methods similar to those previously described for mapping nucleic acid-RT binding interactions [216,219,226,274]. 5'-radiolabeled aptamers (~40 000 cpm) were incubated in 30 mM KCl, 2 mM $MgCl_2$ 30 mM Tris–HCl (pH 7.5) in the presence or absence of 500 nM HIV-1 RT and equilibrated for 15 min on ice in a total volume of 20 µl at the bottom of a 1.7 ml centrifuge tube. The hydroxyl radical footprinting solution was generated by placing 1 µl of a freshly prepared Fe(II)-EDTA solution from powders (1 mM $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$ and 2 mM EDTA final concentration unless otherwise noted), a 1 µl of 2% hydrogen peroxide (freshly diluted from a 30% stock) and 1 µl of 100 mM sodium ascorbate) as three separate drops on the side of the tube. The reaction was initiated by simultaneously combining the three 1 µl drops with the aptamer solution by briefly spinning the tube in a centrifuge. The reactions were quenched by quickly adding 1 µl of 100 mM thiourea, 2.5 µl of 3 M sodium acetate and 1 µl of 20 µg/µl glycogen after 2 min incubation, unless otherwise noted.

Hydrogen-deuterium exchange

Solution-phase amide HDX experiments were carried out with a fully automated system described previously [229] with slight modifications. Briefly, 4 µl of 10 µM full-length RT alone or in the presence of ligand at a molar ratio of 1.1 and 1.5:1 ligand:RT, respectively, was diluted to 20 µl with D_2O and incubated at 4°C for 10, 30, 60, 900 or 3600 s. Following exchange, unwanted forward or back exchange was minimized and the protein was denatured with a quench solution

(5 M urea, 50 mM TCEP and 1% v/v TFA) at 1:1 ratio to protein. Samples were then passed through an in-house prepared immobilized pepsin column at 50 µl min^{-1} (0.1% v/v TFA, 15°C) and the resulting peptides were trapped on a C18 trap column (Hypersil Gold, Thermo Fisher). The bound peptides were then gradient-eluted (5-50% CH_3CN w/v and 0.3% w/v formic acid) across a 1 × 50 mm C18 HPLC column (Hypersil Gold, Thermo Fisher) for 5 min at 4°C. The eluted peptides were then analyzed directly using a high resolution Orbitrap mass spectrometer (LTQ Orbitrap XL with ETD, Thermo Fisher). Each HDX experiment was performed in triplicate. To identify peptides, MS/MS experiments were performed with a LTQ Orbitrap mass spectrometer over a 70 min gradient. Product ion spectra were acquired in a data-dependent mode and the five most abundant ions were selected for the product ion analysis. The MS/MS *.raw data files were converted to *.mgf files and then submitted to Mascot (Matrix Science, London, UK; http://www.matrixscience.com) for peptide identification. Peptides with a Mascot score of 20 or greater were included in the peptide set used for HDX detection. The MS/MS Mascot search was also performed against a decoy (reverse) sequence and false positives were ruled out. The MS/MS spectra of all the peptide ions from the Mascot search were further manually inspected and only the unique charged ions with the highest Mascot scores were used in estimating the sequence coverage. The intensity weighted average m/z value (centroid) of each peptide isotopic envelope was calculated with the latest version of our in-house developed software, HDX Workbench [275].

2D ^{1}H-^{1}H NMR

Purified RNA samples were lyophilized overnight and dissolved in D_2O with 10 mM dTris–HCl, pH 7.0, 140 mM KCl, 10 mM NaCl and 1 mM $MgCl_2$ to a final concentration of ~700 µM. NMR data were collected at 308K on a Bruker Avance III 800 MHz spectrometer equipped with TCI cryoprobe (University of Missouri, Columbia). Two dimensional ^{1}H-^{1}H NOESY

data were processed with NMRPipe and NMR Draw [187], and analyzed with NMRViewJ (One Moon Scientific, NJ, USA).

SAXS data collection and analysis

148.1-38m and 148.1-38m delU were prepared by *invitro* transcription using DNA templates containing a T7 promoter. DNA templates were prepared by phosphoramidite method oligonucleotide synthesis (Integrated DNA Technologies). Transcription products were subjected to denaturing gel electrophoresis overnight. RNA was visualized by UV-shadowing, extracted from the gel through electro-elution (Elutrap, Whatman) and washed in ultra-centrifugal filters (Amicon, Millipore Sigma). Prior to travel to the beamline the RNA samples were lyophilized. On site the samples were dissolved in 10 mM Tris–HCl (pH 7.5) to a concentration of 100 μM. In order to refold the RNA the samples were boiled for 5 min and cooled quickly in ice. Salts were then added to achieve a final concentration 40 mM KCl, 10 mM NaCl and 1 mM $MgCl_2$ and the RNA was incubated for 5 min.

The samples were then applied to a size exclusion column connected inline to the beamline. SAXS data were collected as the samples eluted from the column (Argonne beamline 12-ID-B). ATSAS [189] was used to average the sample data and the buffer scattering was subtracted. DAMMIF was used to generate *ab initio* models. Pairwise distribution plots were created using GNOM by goodness of fit to the raw SAXS data and the total quality estimate. *Ab initio* models were visualized using PyMOL.

3D structural prediction of the apoRNA

Using the template-generated 3D structures by Vfold3D as initial states, we ran coarse-grained Molecular Dynamics simulations in IsRNA model with the secondary structures as constraints, including the tertiary contact suggested by the NMR data (A30-H2 to U14-H1' distance restraint (distance < 5 Å)), to sample a 3D structure ensemble. Coarse-grained

Molecular Dynamics simulations were implemented with Langevin dynamics at temperature T = 300 K with the integration time step set to Δt = 1fs. The simulation length was 5 ns and the structure snapshots were recorded every 5ps. The top 10% snapshot structures with the lowest energies in IsRNA model were collected and then clustered according to pairwise root-mean-square deviation (RMSD) with a threshold of 5 Å. The centroid structures of the clusters were selected as candidate 3D structures, resulting a total of 70 modeled structures for aptamer 148.1-38m.

Computational docking of the aptamer–RT complex

The RNA aptamer structures were docked to HIV-1 RT using our own protein–RNA docking program, MDockRP [164]. MDockRP is based on the Fast Fourier Transformation algorithm [276], and is implemented with the pairwise shape complementarity function and electrostatic energy function of the protein–protein docking program, ZDOCK [277]. The GPU-based MDockRP created a total of 54 000 docking modes for each model aptamer structure on HIV-1 RT.

Protease cleavage assay

The 2.4 pmol of p66/p66 RT was incubated with 45.4 pmole aptamer on ice in binding buffer (50 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$). Note that a modified version of aptamer 148.1 was used in which the pyrimidines (Y) were replaced with 2'F-substituted pyrimidines (148.1-2'FY) to reduce aptamer degradation during the assay. After 10 min, 45.4 pmole HIV PR (Abcam) was added and the reaction was incubated at 37°C for 6 h. RT without aptamer or PR was used as control for background and was set up in the same reaction conditions. The cleavage reaction was stopped by adding 4× Laemmli protein dye (Biorad). Samples were loaded onto 10% sodium dodecyl sulphate (SDS)-PAGE. Cleavage of p66/p66 by PR was detected via immunoblot. Proteins in the SDS-PAGE were transferred to PVDF membrane (transferring buffer

contained 50 mM Tris pH 7.5, 380 mM glycine and 20% methanol) using semi-dry electroblotting method. Transferring current was set at 100 mA for 2 h. Membrane was washed three times with phosphate-buffered saline (PBS)-Tween (80 mM Na_2HPO4, 1.5 M NaCl, 20 mM KH_2PO_4, 30 mM KCl and 0.5% Tween 20, pH 7.5) before blocking with milk (5% milk in PBS-Tween). After 45 min, membrane was incubated in primary rabbit anti-RT antibody (This reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: Cat. #6195 and lot #140226) (1:4000 dilution in 3% PBS-Tween milk) overnight at 4°C. After washing three times with PBS-Tween, membrane was incubated in secondary horseradish peroxidase anti-rabbit (1:10 000 dilution in 4% PBS-Tween milk) for 1 h. Membrane was washed three times with PBS-Tween. Development of the membrane started by loading the Immobilon Classico Western HRP substrate (Millipore Sigma) onto the washed membrane. Luminescence was detected using the UVP Biospectrum Imaging System (Analytikjena). Image was captured using Vision Works software. Intensities of p66 and p51 bands corresponding to each reaction sample were quantified using ImageJ, then the p51:p66 ratio was calculated. The p51:p66 ratio corresponding to p66 homodimer alone (no PR and no aptamer) was set as background and was subtracted from the p51:p66 ratio of each sample. The adjusted p51:p66 ratios were then plotted via GraphPad Prism 6.2

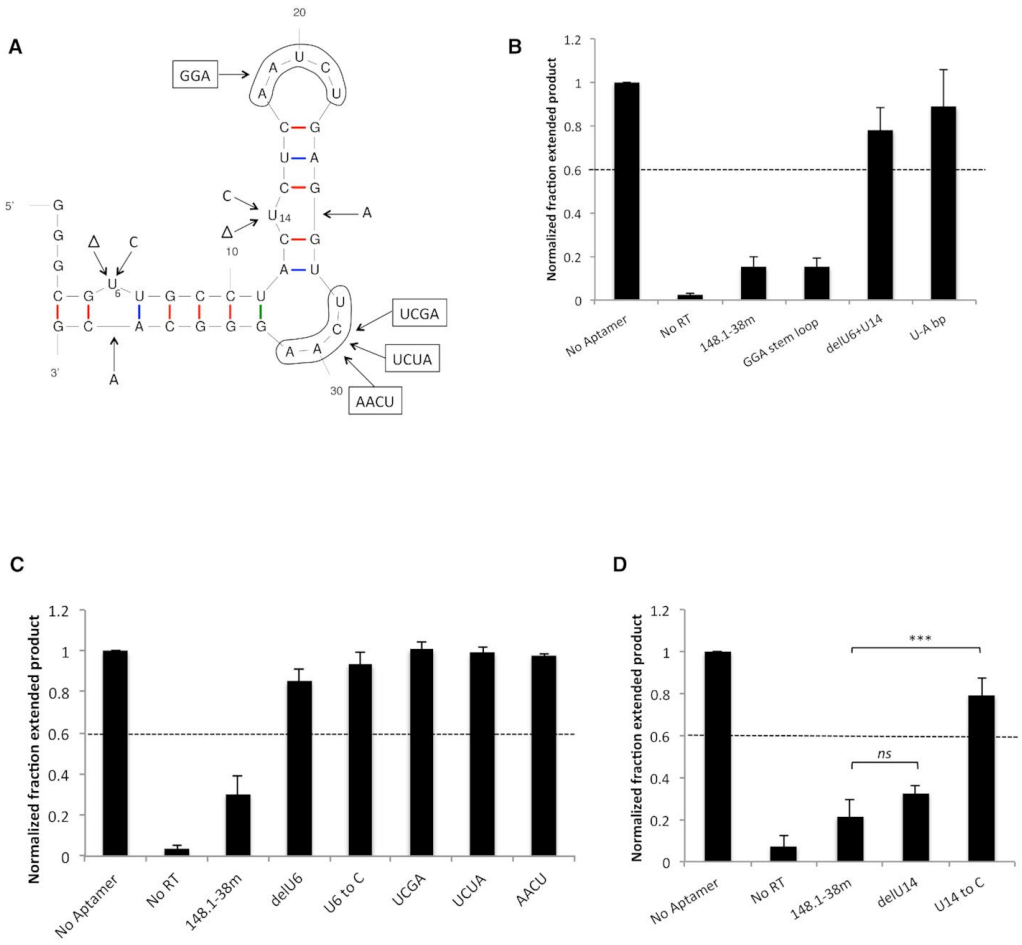


Figure A1-1. Binding core and critical structural features of 148.1-38m. **A** Modifications on 148.1-38m aptamer. The 2D predicted structure of the aptamer was generated via Mfold (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form). Red, blue and green sticks depict C-G, A-U and non-Watson–Crick base pairings, respectively. (**B**), (**C**) and (**D**) Inhibition of RT by 148.1-38m variants measured via primer extension. For each RT-aptamer pair, the fraction extended product was measured and normalized to that of the 'No Aptamer' reaction. The fraction of primers extended to full-length product was the highest (NFEP value = 1) in the absence of aptamer (designated 'No Aptamer'). In the absence of RT, p/t was not able to extend. Variants that gave NFEP values higher than 0.6 were considered 'not inhibitory.' Values presented on the graphs were the results from three independent experiments. The error bars represent the standard deviation of the NFEP values. (**B**) Base pairing or deletion of both

unpaired Us (U6 and U14) abolished aptamer inhibition whereas replacing AAUCU stem loop with minimum GGA sequence did not affect inhibition of the aptamers. (**C**) The unpaired U6 and the signature UCAA bulge were also critical to 148.1-38m inhibition. The UCGA, UCUA and AACU variants failed to inhibit RT. (**D**) Variant with deleted U14 was still inhibitory. However replacing U14 by C14 abolished aptamer inhibition. *ns* ($P > 0.05$), *** ($P < 0.001$).

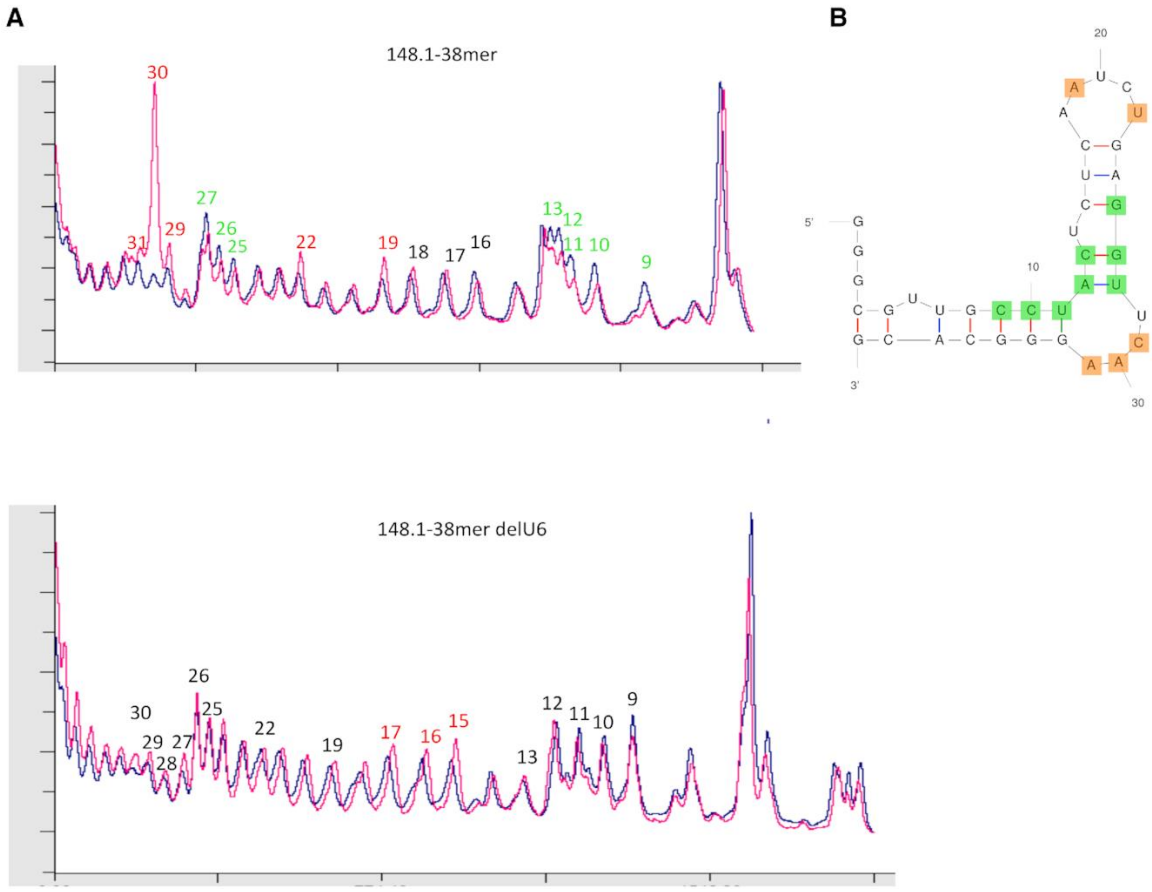


Figure A1-2. Hydroxyl radical footprinting profiles of 148.1-38m and its delU6 variant in the presence (magenta) or absence (blue) of RT. **A.** Cleavage patterns from hydroxyl radical digestions are shown for 148.1-38m (45) and the delU6 variant (18). When in complex with RT, 148.1-38m was subjected to hyper-sensitive cleavage ('exposed') at position A30 of the UCAA bulge. Positions A19, U22, A31 and C29, were also sensitive to cleavage but not to the same level as A30. In contrast, C9 to C13 and G25-U27 experienced decrease in level of cleavage ('protected') when RT was present. **B**. Protected (green) and exposed (orange) positions are mapped onto predicted structure of 148.1-38m.

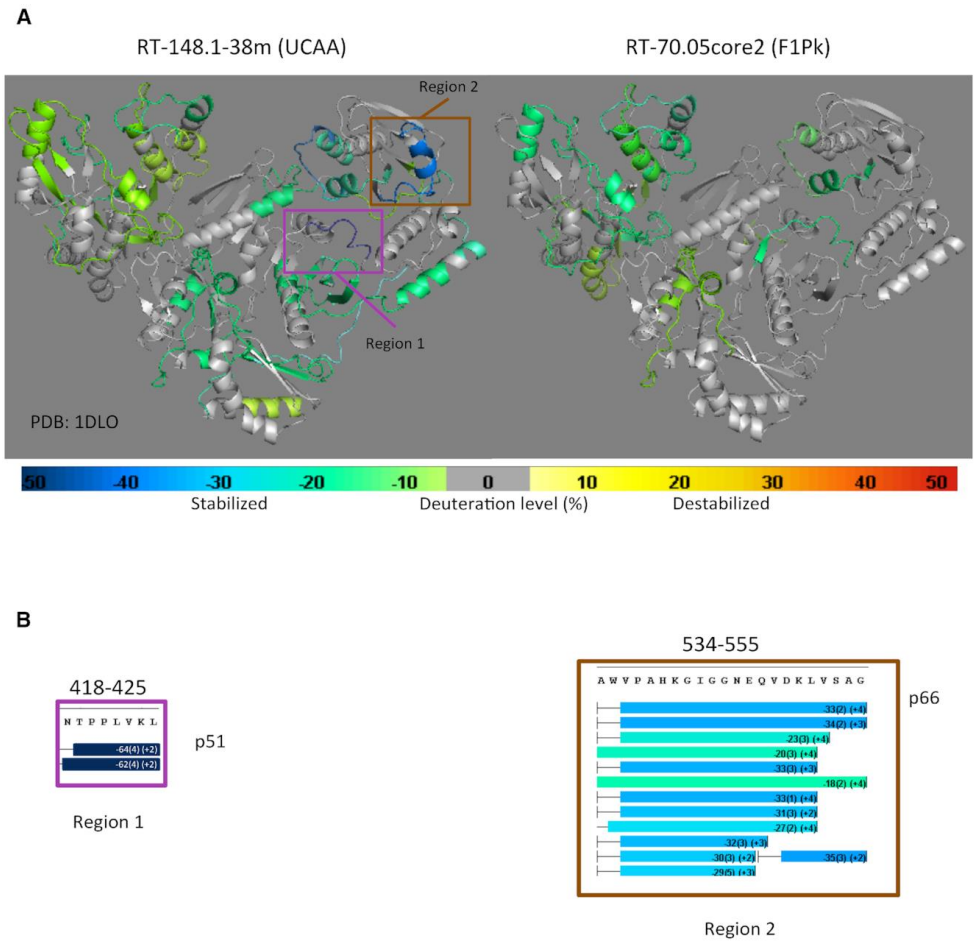


Figure A1-3. HDX/MS analysis of RT-148.1-38m (left) versus RT–70.05core2 (right) complexes. **A**. The difference in HDX of RT in the presence and absence of the 148.1-38m and 70.05core2 aptamers are mapped onto the crystal structure of ligand-free RT (PDB ID: 1DLO). Change in HDX is color-scaled. Lime green to dark blue indicates decreasing percent deuterium uptake (protection) compared to apo RT. Yellow to bright red represents increasing percent deuterium uptake (exposure). Peptides in gray are those having no net change. **B**. Highly protected peptides on RT in the presence of 148.1-38m. When in complex with RT, peptide 418-425 of the p51 subunit (magenta box- Region 1) and peptide 524–555 of the p66 subunit (brown box- Region 2) are strongly protected, especially the 418–425 peptide (dark blue), relative to the remaining regions of RT.

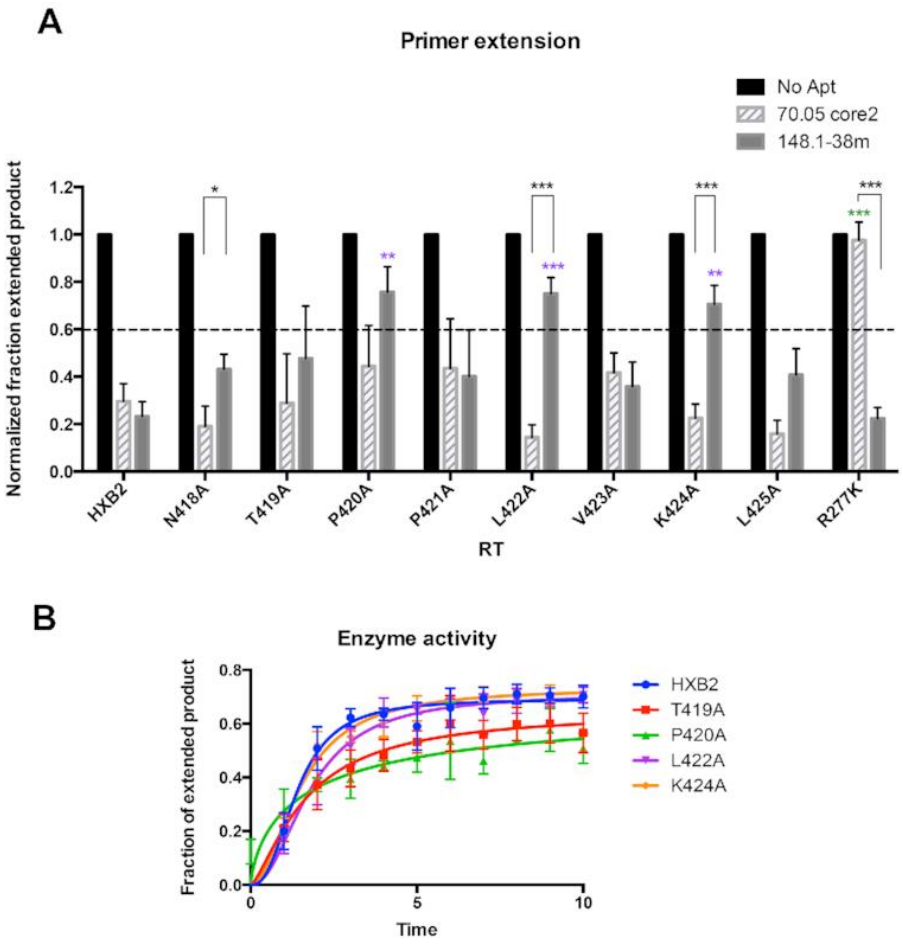


Figure A1-4. Alanine-scanning mutagenesis of the 418-NTPPLVKL-425 peptide of p51. **A.** Values on the y-axis represent the normalized fraction extended product (NFEP). For each RT-aptamer pair, the fraction of primer converted to full-length product was determined by quantifying band intensities using Multigauge software then normalized to that of RT-No Aptamer. In the absence of aptamer (No Aptamer), the fraction of primers extended to full-length product was the highest. This NFEP value is defined as '1.' In the absence of RT, p/t was not able to extend. Reactions in which the NFEP values were higher than 0.6 were considered to be 'not inhibitory.' Values presented on the graphs were the results from three independent experiments. The error bars represent the standard deviation of the NFEP values. Inhibition of mutant RTs by 70.05core2 and 148.1-38m via primer extension assay. R277K is fully resistant to 70.05core2 and was used as control for the assay. Among tested mutants, P420A, L422A and K424A had reduced

susceptibility to 148.1-38m inhibition. In contrast, 70.05core2 inhibits all mutants except for R277K. Significant difference between inhibition of 148.1-38m against HXB2 and mutant RT is indicated on top of the specific mutant (purple). Significant difference between inhibition of 70.05core2 against HXB2 and mutant RT is indicated on top of the specific mutant (green). In black is the significant difference between inhibition of 148.1-38m and 70.05core2 against the same mutant RT. * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$). **B**. Enzymatic activity of mutant RTs compared to WT RT (HXB2). Activity was measure by primer extension assay and represented as the fraction of extended product over the time of reaction. Values presented on the graphs were the results from three independent experiments. The error bars represent the standard deviation of the fraction of extended product values.

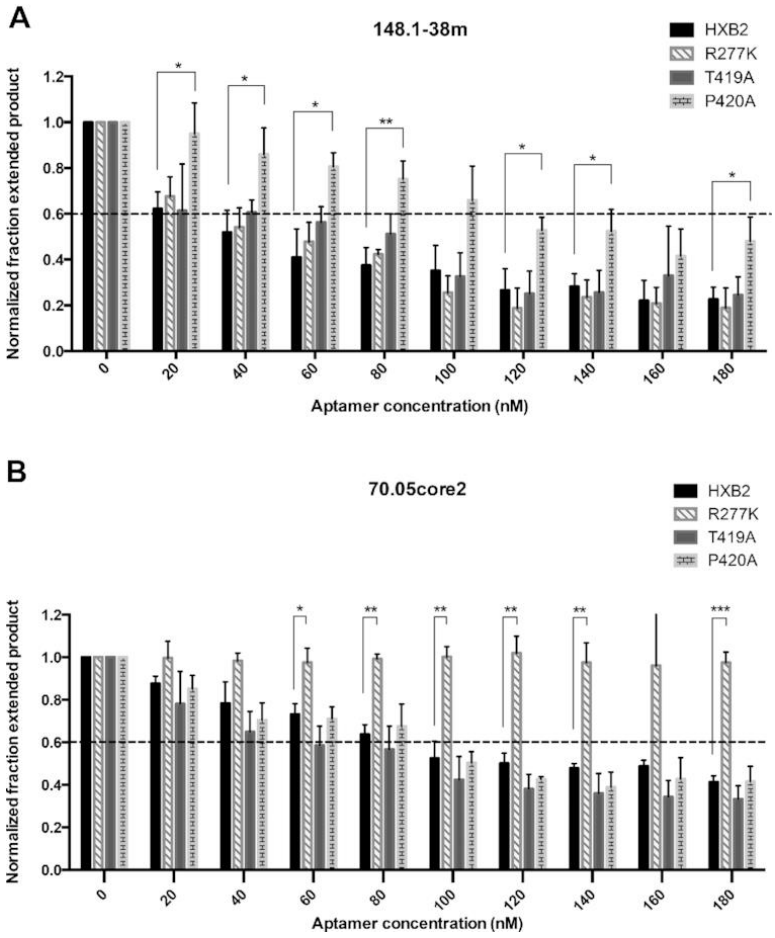


Figure A1-5. Inhibition profiles of WT and mutant RTs in response to increasing concentration of 148.1-38m (**A**) and 70.05core2 (**B**). Values shown represent triplicate measurements (±SD) of three independent experiments. **A.** Inhibitory phenotype of 148.1-38m against P420A can be partially rescued with high aptamer concentration (160 and 180 nM). Inhibition profiles of R277K and T419A are similar to WT HXB2. Significant difference between inhibition of 148.1-38m against HXB2 and P420A mutant is represented in the graph. * ($P < 0.05$), ** ($P < 0.01$). (B) R277K was used as a control and stayed fully resistant to 70.05core2 under any tested concentration. T419A and P420A behaved similarly to WT HXB2 in the presence of 70.05core2 indicating that change in susceptibility observed with P420A is specific only to 148.1-38m. Significant difference between inhibition of 148.1-38m against HXB2 and P420A mutant is represented in the graph. * ($P < 0.05$), ** ($P < 0.01$), *** ($P < 0.001$).

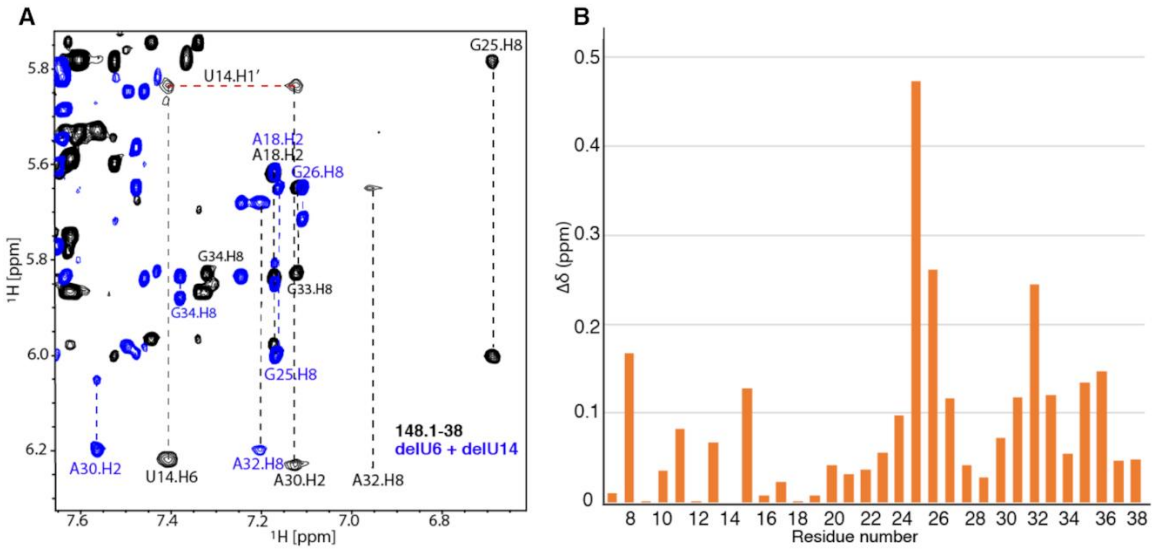


Figure A1-6. NMR study of 148.1-38m and its mutant RNA. **A**. Overlay of a representative region of the 2D NOESY NMR spectra collected for 148.1-38m (black) and delU6+delU14 (blue). H2, H6 and H8 connectivities are denoted in vertical dashed lines. The connectivity between A30-H2 and U14-H1' is shown (horizontal red dashed line). **B**. Absolute values of the chemical shift differences (△δ) of each residue in the 148.1-38m and delU6+delU14 are plotted. The chemical shifts of H8 of purines and H6 of pyrimidines are used to represent the chemical shift change of each residue.

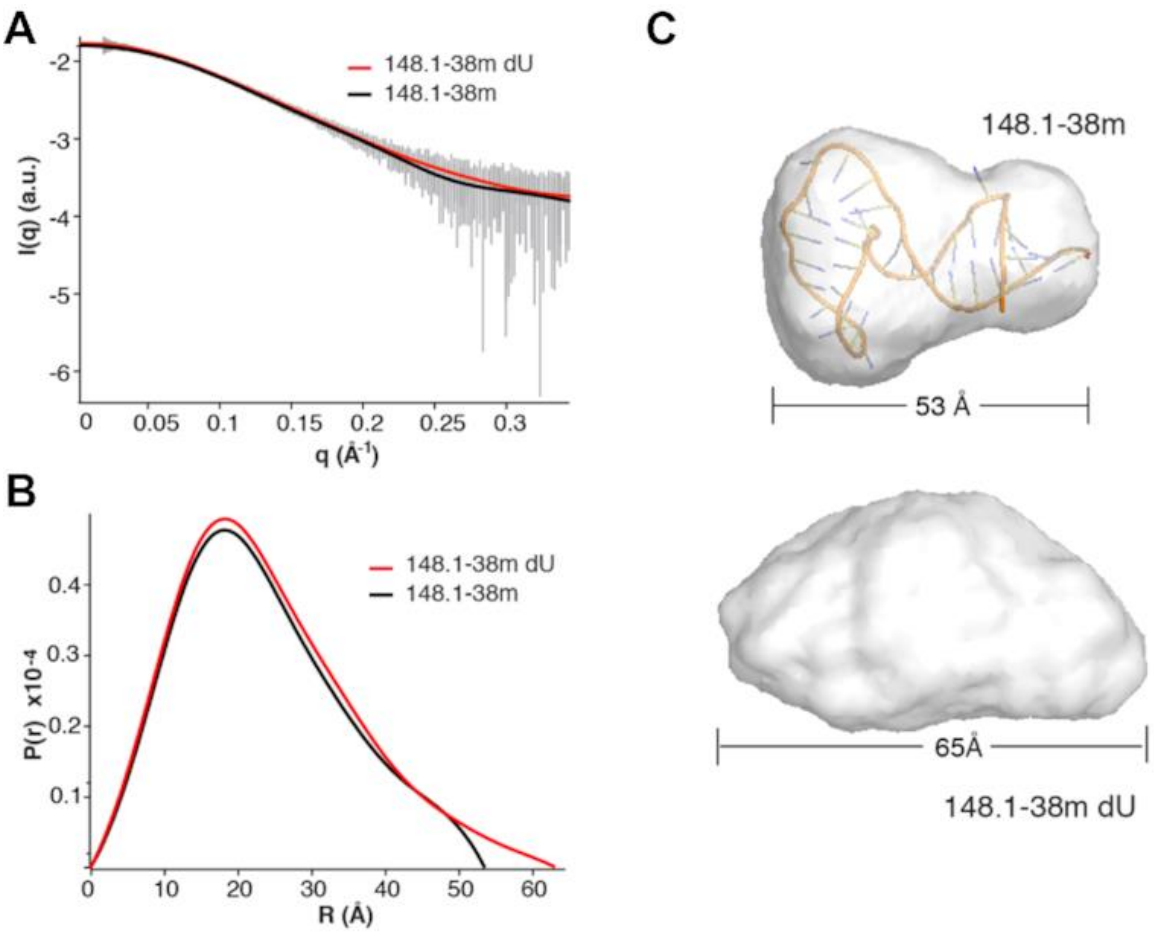


Figure A1-7. Deletion of uracil alters SAXS scattering profile. **A**. Line of best fit of scattering data. 148.1-38m denoted by black line. 148.1-38m dU denoted by red line. Error bars of averaged scattering data in gray. **B**. Pair wise distribution plot of 148.1-38m (black line) and 148.1-38m (red line). (C) *Ab initio* models of 148.1-38m (top, overlaid with docked conformation of RNA from Figure 8) and 148.1-38m dU (bottom).

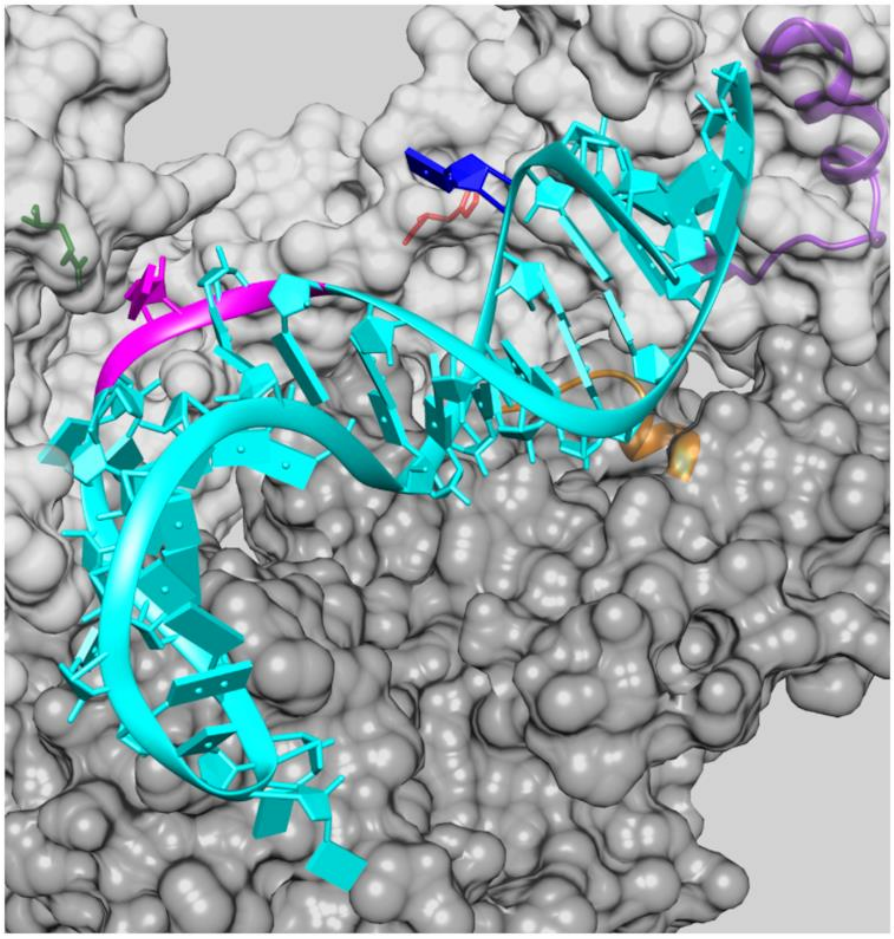


Figure A1-8. Modeled structure of the aptamer 148.1-38m in complex with the HIV-1 RT protein. The two subunits of protein, i.e. p51 and p66, are presented as dark gray and light gray surface, respectively. Region 1 of p51 is highlighted in golden ribbon representation and Region 2 of p66 in purple ribbon. The side chains of two residues of p66, HIS361 and ARG284, are displayed in stick representation and colored red and green, respectively. The RNA aptamer is depicted in cyan ribbon representation, except for the UCAA segment, which is emphasized as a stretch of a magenta ribbon. The bases of the aptamer are shown as cyan slabs, except for U6 and C29, which are colored blue and magenta, respectively.

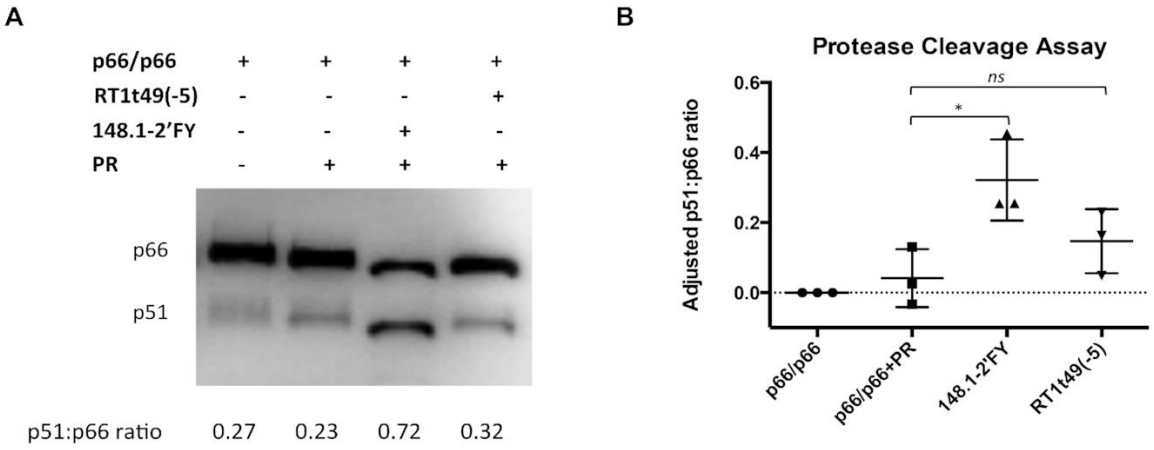


Figure A1-9. Effect of aptamers on protease cleavage of p66. Aptamer used in this assay were the RNA aptamer 148.1-38m and DNA aptamer RT1t49(-5). **A**. Representative protein immunoblot showing PR-mediated cleavage of p66/p66. The presence of p66 and p51 in each sample was detected by Western blot. Free p66/p66 RT was used as a control for background signal of p51. Additional data can be found in Supplementary Figure 9. **B**. Quantitative representation of three individual experiments. Intensities of p66 and p51 bands corresponding to each reaction sample were quantified using ImageJ then the p51:p66 ratio was calculated. The p51:p66 ratio correlated with free p66/p66 was set as background and was subtracted from the p51:p66 ratio of each remaining sample. The adjusted p51:p66 ratio then was plotted via Prism software version 6.2. *ns* ($P > 0.05$), * ($P < 0.05$). Values shown represent triplicate measurements (±SD) of three independent experiments.

Supplementary Materials

148.1 full length	GGGCAUAAGGUAUUUAAUUCCAUACAGCUACACCGACGCACGUUGC CUACUCUCAAUCUGAGGU**UCAA**GGGCACGAAAACGAGGUCUCAUGC UGACAUGCGAGAUUGAUUCGGAUGCUCCGGUAGCUCAACUCG
148.1-38m	GGGCGUUGCCUACUCUCAAUCUGAGGU**UCAA**GGGCACG
148.1-38mer del GGG	GGUUGCCUACUCUCAAUCUGAGGU**UCAA**GGGCACC
148.1-38m GGA	GGUUGCCUACUCUC**GGA**GAGGU**UCAA**GGGCACC
148.1-38mer GGG-CCC	**GGG**CGUUGCCUACUCUCAAUCUGAGGU**UCAA**GGGCACG**CCC**
148.1-38m A-U bp	GGGCGUUGCCUACUCUCAAUCUGAG**A**GU**UCAA**GGGCA**A**CG
148.1-38mer delU6	GGGCGUGCCUACCUCAAUCUGAGGU**UCAA**GGGCACG
148.1-38m U6 to C	GGGCGCUGC**C**UACUCUCAAUCUGAGGU**UCAA**GGGCACG
148.1-38mer delU14	GGGCGUUGCCUACCUCAAUCUGAGGU**UCAA**GGGCACG
148.1-38m U14 to C	GGGCGUUGCCUACCCU**C**AAUCUGAGGU**UCAA**GGGCACG
148.1-38m-UCGA	GGGCGUUGCCUACUCUCAAUCUGAGGU**UCGA**GGGCACG
148.1-38m-UCUA	GGGCGUUGCCUACUCUCAAUCUGAGGU**UCUA**GGGCACG
148.1-38mer-AACU	GGGCGUUGCCUACUCUCAAUCUGAGGU**AACU**GGGCACG
70.05core2	CAAGAUCCGAGGCAGAACGGGAAAAUCUGCGAAGU

Table AS1-1. List of aptamers and their sequences. The signature bulge UCAA is underlined. Blue color indicates mutated/added nucleotides.

	5 Å cutoff	8 Å cutoff
Amino Acid Residue	Nucleotide	Nucleotide
p66: LYS11		G26
p66: LYS12		G25, G26, U27
p66: LYS13	G25, G26, U27, U28	A24, G25, G26, U27, U28
p66: PRO14	A18, A24, G25, G26	U16, C17, A18, A24, G25, G26, U27
p66: GLY15	G25	U16, C17, A18, G23, A24, G25, G26
p66: MET16	G25, G26, U27	C15, A18, G25, G26, U27, U28
p66: ASP17		G25
p66: ARG78		U27
p66: GLU79		U27
p66: LEU80		U28
p66: ASN81	U28	G26, U27, U28
p66: LYS82	G26, U27, U28	G25, G26, U27, U28
p66: ARG83		G25, G26, U27, U28
p66: THR84	G26	G25, G26, U27, U28
p66: GLN85		G25, G26, U27, U28
p66: ASP86	G26, U27, U28	G25, G26, U27, U28
p66: LYS154	U28	G25, G26, U27, U28
p66: ARG277		C29
p66: SER280		C29
p66: LYS281	C29	C29, A30
p66: ARG284	C29	U28, C29, A30
p66: GLY359	U6	U6, G8
p66: ALA360	U6	U6, U7, G8
p66: HIS361	U6	G5, U6, U7, G8
p66: THR362		U6

	5 Å cutoff	8 Å cutoff
Amino Acid Residue	Nucleotide	Nucleotide
p51: PRO9		U22
p51: VAL10		U22, A24
p51: LYS11	A19, U20, U22, G23, A24	A18, A19, U20, C21, U22, G23, A24
p51: LEU12	G23, A24	A19, U22, G23, A24, G25
p51: LYS13	U22, G23, A24	A19, U22, G23, A24, G25
p51: PRO14	U22, G23, A24	A19, U20, C21, U22, G23, A24, G25
p51: GLY15	G23, A24, G25	C13, A19, U22, G23, A24, G25, G26
p51: MET16	G23, A24, G25	A12, G23, A24, G25
p51: ASP17	A24, G25	G23, A24, G25, G26, U27
p51: GLY18	G25	G23, A24, G25
p51: LYS20		G25, G26, U27, U28
p51: PRO55		G25
p51: TYR56	G25	A24, G25
p51: GLU79		U11
p51: LYS82		C10, U11, A12
p51: ARG83		A12, G23, A24
p51: THR84		U22, G23, A24
p51: GLN85	U22	U22, G23, A24
p51: ASP86		U22, G23
p51: PHE87		U22
p51: TRP88		C21, U22
p51: PHE124		A24
p51: LYS126		A24, G25
p51: TYR127	A24	G23, A24, G25
p51: GLY262		G38

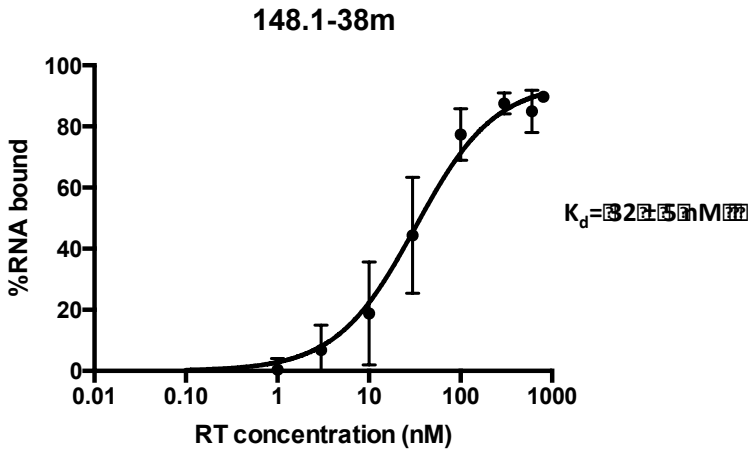
p66: GLN407		G8, C9
p66: ASP443	C40	C39, C40
p66: GLY444	C39, C40	C39, C40, C41
p66: ALA445	C40, C41	G3, C4, C39, C40, C41
p66: ALA446	C40, C41	G2, G3, C4, C39, C40, C41
p66: ASN447	G3, C40, C41	G2, G3, C4, C40, C41
p66: ARG448	G2, G3, C4	G2, G3, C4, C40, C41
p66: GLU449	G3, C41	G1, G2, G3, C4, C41
p66: THR450		G3, C40, C41
p66: LYS451		G3, C4
p66: LEU452		C40, C41
p66: GLY453		C40, C41
p66: LYS454		C40, C41
p66: THR472		G3, C4, G5
p66: THR473	C4, G5	G3, C4, G5, U6, C39, C40
p66: ASN474	G3, C4, G5, C39, C40	G2, G3, C4, G5, U6, G38, C39, C40, C41
p66: GLN475	C4, G5, U6, G38	G3, C4, G5, U6, U7, G38, C39, C40
p66: LYS476	C4, G5	C5, G5, U6
p66: THR477		C4, G5, C39, C40
p66: GLU478		C4, G5, G38, C39, C40
p66: LEU479	U6	G5, U6
p66: ASP498	G38, C39, C40	G38, C39, C40
p66: SER499	G38	G5, U6, G38, C39
p66: GLN500	C37, G38	G5, U6, U7, A36, C37, G38, C39
p66: TYR501	G5, U6, U7	C4, G5, U6, U7, G8, C37, G38, C39
p66: ALA502		G5, U6, U7
p66: GLY504		U7
p66: ILE505	G5, U6	G5, U6, U7
p66: LYS512		U6
p66: SER513	U6	U6

p51: ASN265	G38	C37, G38, C39
p51: TRP266		C37, G38
p51: ALA267		G38
p51: GLN269	A36, C37, G38	A36, C37, G38
p51: ILE270		C37
p51: ASP324		C10, U11
p51: ILE326		C9, C10
p51: GLU328		G8, C9, A36
p51: TYR342		G8, C9, C35, A36
p51: GLN343		C35
p51: GLU344		C35, A36
p51: PRO345	C35, A36	C35, A36, C37
p51: PHE346	A36, C37	A36, C37, G38
p51: LYS388	C9, C10	C9, C10, U11, G34
p51: PHE389	C10	C9, C10
p51: LYS390	C9	G8, C9, C10, G34, C35, A36
p51: LEU391		G8, C9, C10
p51: ILE393		G8, C9
p51: GLN394		C9
p51: LYS395		C9
p51: GLU413		C10, U11
p51: TRP414	C10	C9, C10, U11
p51: GLU415	C9, C10	G8, C9, C10, U11, G34, C35
p51: PHE416	G8, C9	G8, C9, C10
p51: VAL417	G8, C9	G8, C9, C10
p51: ASN418	G8, C9	U7, G8, C9
p51: THR419		G8

p66: GLU514	U6	U6
p66: SER515	U6	U6
p66: GLU516		U6
p66: VAL518		G5, U6
p66: TRP535	G38	C37, G38, C39
p66: VAL536		G38, C39
p66: PRO537	C39	G38, C39
p66: ALA538	G38, C39	G38, C39, C40
p66: HIS539	C39, C40	G2, G3, G38, C39, C40, C41
p66: LYS540		C39, C40
p66: ILE542		C39
p66: ASN545		C39, C40
p66: VAL548		C40
p66: ASP549	C40	C39, C40, C41
p66: LYS550		C40, C41
p66: LEU551		C41
p66: VAL552	C40, C41	C39, C40, C41
p66: SER553	C40, C41	C39, C40, C41
p66: ALA554	C40, C41	C40, C41
p66: GLY555	C41	C40, C41
p66: ILE556		C40, C41

Table AS1-2. Amino acids in HIV-1 RT p66 (left set of columns) and p51 (right set of columns) that are predicted to lie within 5Å or within 8Å of aptamer 148.1-38m in the bound complex.

a



b

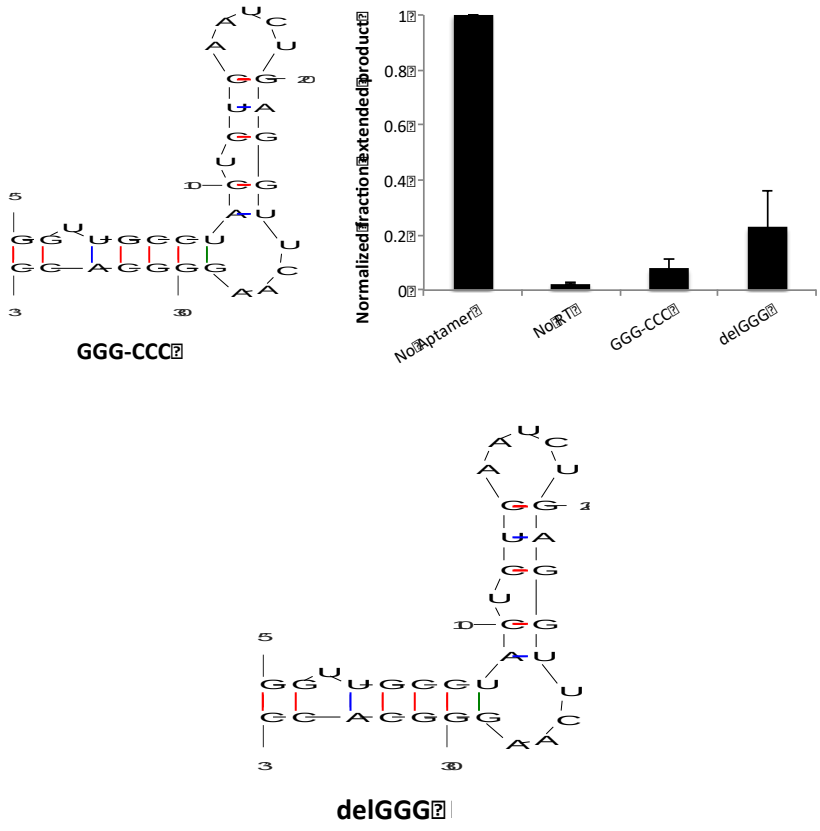


Figure AS1-1. Binding affinity of the minimum core 148.1-38m for RT and the impact of the 5'GGG overhang on 148.1-38m inhibition. **A.** 148.1-38m bind to RT with a calculated K_d=32 ± 5 nM. **B.** For efficient transcription of the aptamer, the 5'GGG overhang was included in 148.1-

38m sequence. Primer extension assays revealed base pairing (GGG-CCC) or that deletion (delGGG) of this 5'GGG did not affect inhibition of the aptamer against RT. Values shown represent triplicate measurements (±SD) of three independent experiments.

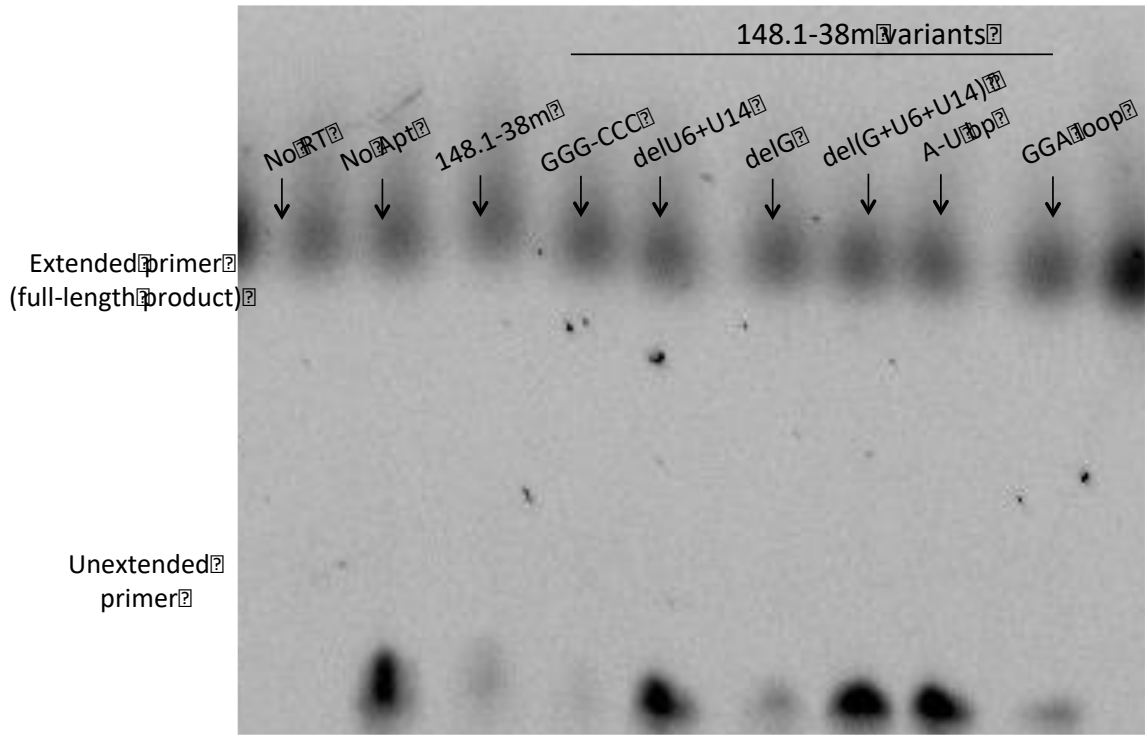


Figure AS1-2. Representative gel for primer extension assay. The assay allow evaluation of aptamer's ability to inhibit polymerase activity of RT. In the absence of RT (No RT), the primer cannot be extented. In the absence of aptamer (No Apt), RT able to extend the primer to full length product. Variant del(G+U6+U14) is an inactivate mutation and is not investigated further in the study.

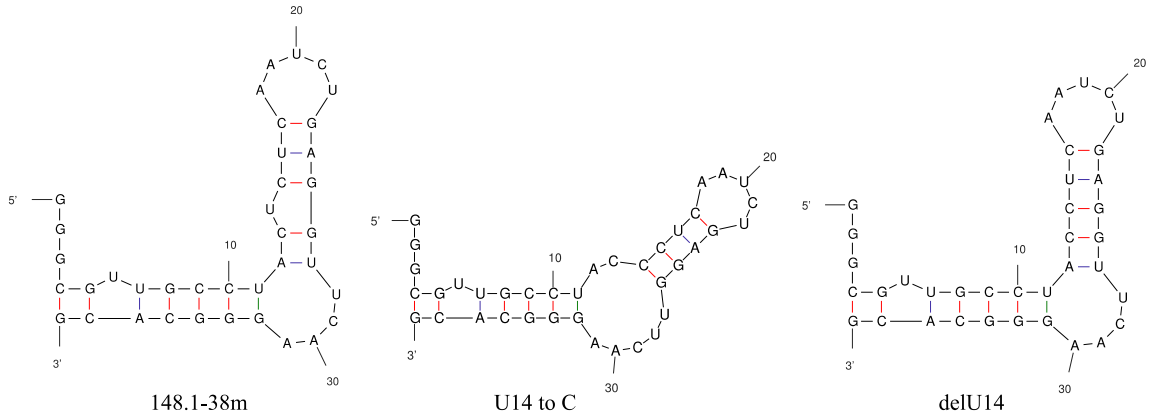


Figure AS1-3. Predicted 2D structure of 148.1-38m and its variants. When U14 was replaced by C, the signature UCAA bulge is predicted to became an AC/AACUU asymmetric internal loop. In contrast, deletion of U14 did not alter the aptamer's predicted structure (the UCAA bulge was maintained.) All predicted structures were generated using Mfold.

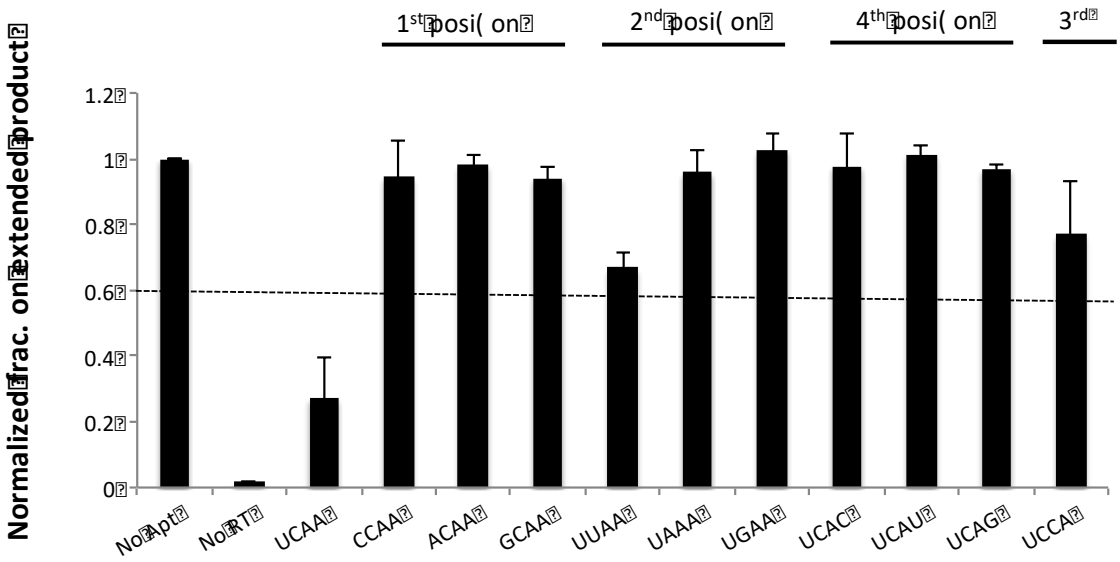


Figure AS1-4. Inhibition of additional single-variants at different positions of UCAA bulge assayed via primer extension. Calculation of NFEP values and the criteria for "not inhibitory" were stated in Fig. 1. The data show that all tested variants represented in the panel failed to inhibited RT (NFEP> 0.6). Values presented on the graphs were the results from three independent experiments. The error bars represent the standard deviation of the NFEP values.

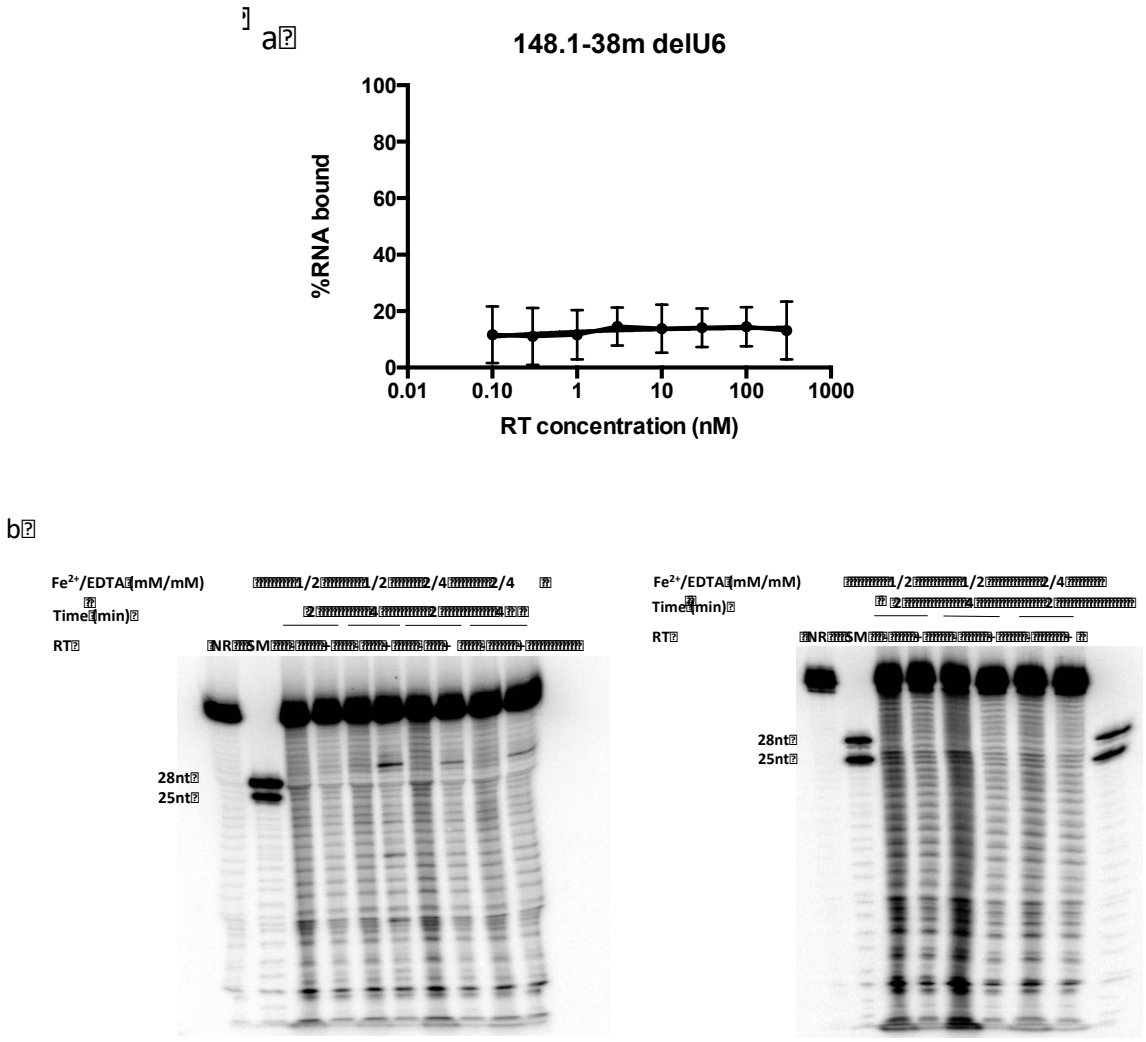


Figure AS1-5. A. Aptamer 148.1-38m delU6 failed to bind RT. Values shown represent triplicate measurements (±SD) of three independent experiments. **B.** Hydroxyl radical footprinting (HRF) profiles of 148.1-38m (left) and its delU6 variant (right) in the presence or absence of RT. NR: No reaction (aptamer only). SM: size marker. Details on buffer conditions and reaction processes can be found in Materials and Methods.

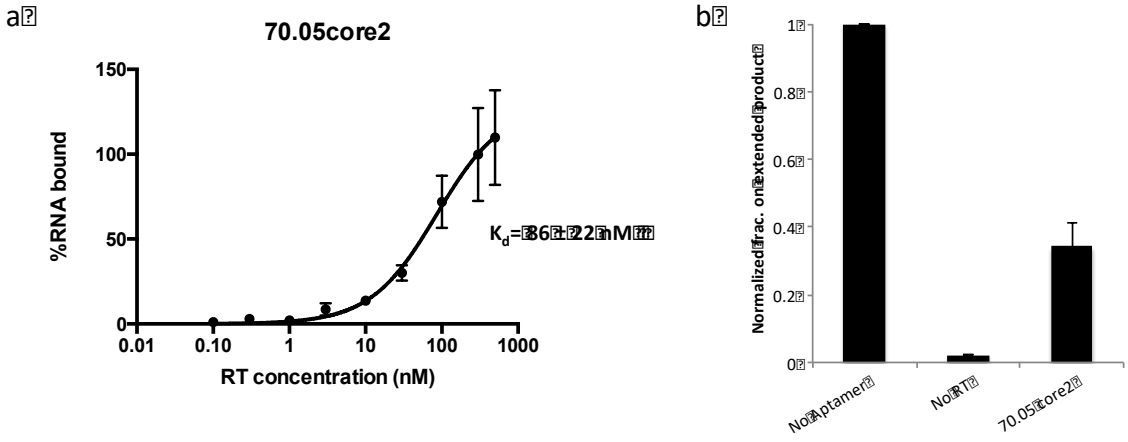
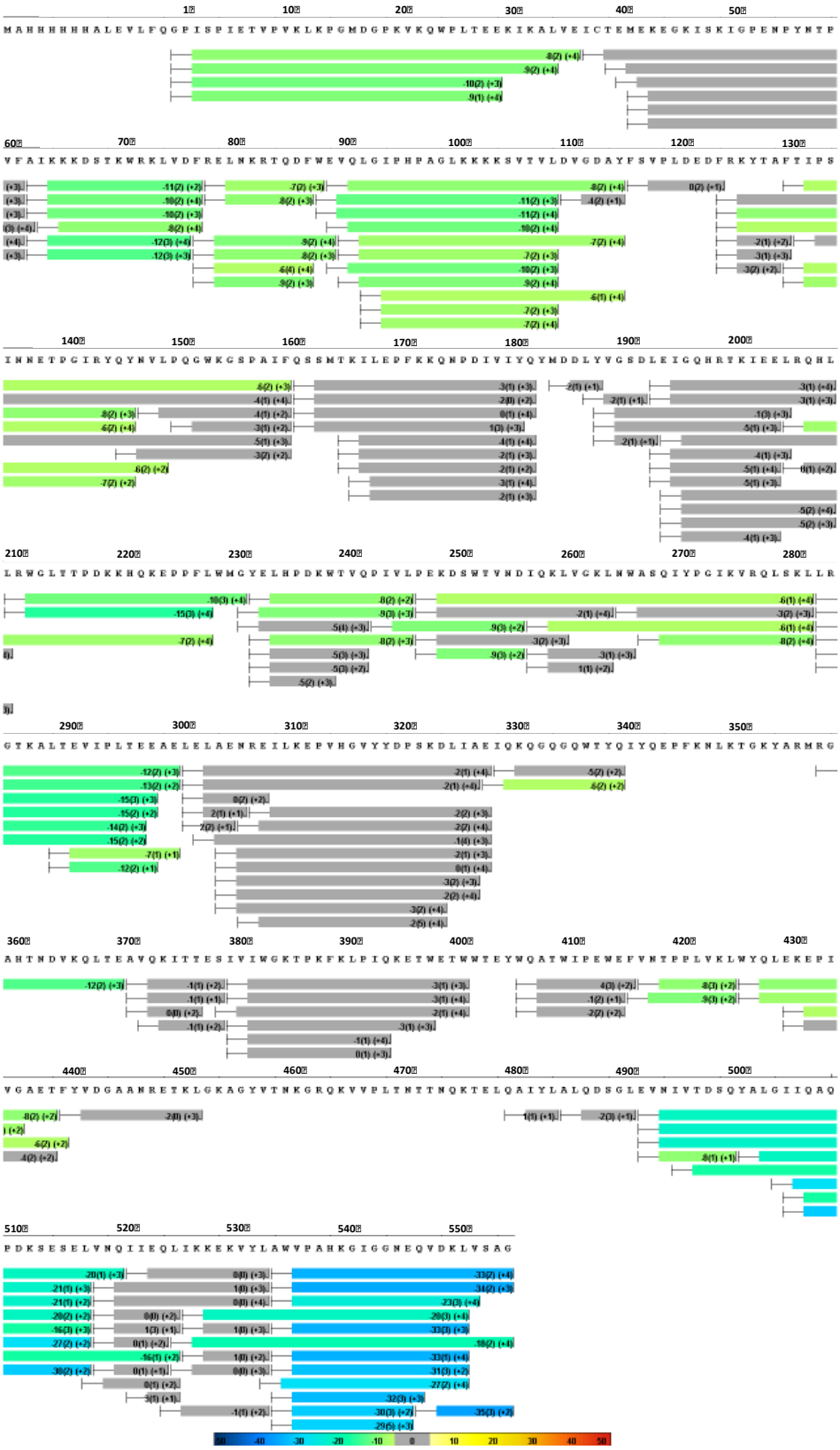


Figure AS1-6. A. Aptamer 70.05core2 binds RT with Kd= 86 ± 22 nM. **B.** 70.05core2 inhibits RT in primer extension assays. Values shown represent triplicate measurements (±SD) of three independent experiments.

a



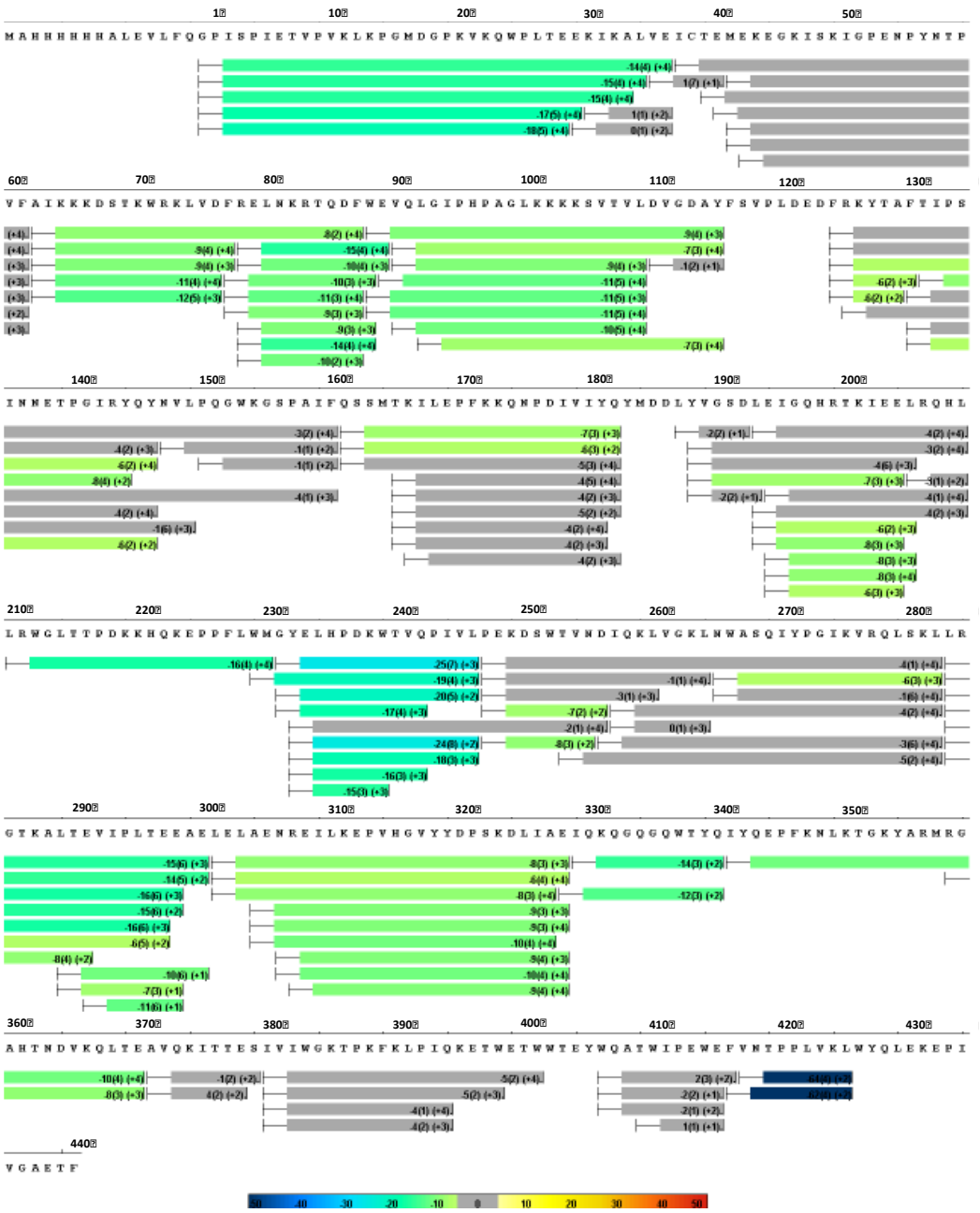
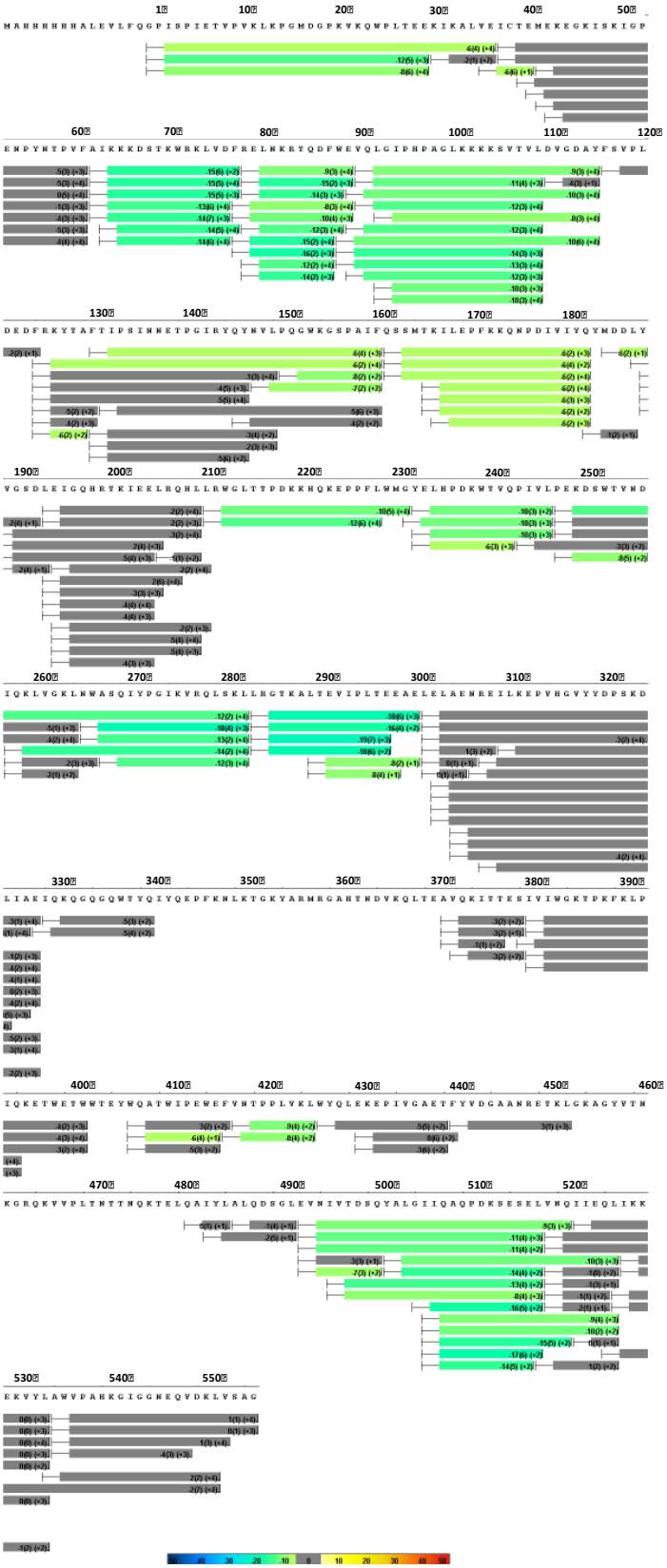


Figure AS1-7. Heat maps of the difference in HDX signal in the presence and absence of 148.1-38m for (**A**) unlabeled p66 and (**B**) 15N labeled p51 in p66/p51 RT. For each subunit, numbering of the first amino acid of RT (Pro) started at 1. The bars below the sequence represent peptides resolved by HDX-MS. Colors represent the average percentage of deuterium uptake based on the color scale shown at the bottom of the figure. Within each bar the three numbers are the

average percent deuterium uptake over five time points (10, 30, 60, 900, and 3600 sec), SD value and peptide charge state, respectively. Grey: no net change. Lime green to dark blue: decreasing percent of deuterium exchanging. Yellow to bright red: increase in percent deuterium uptake. Higher resolution available at https://academic.oup.com/nar/article/48/5/2709/5707188



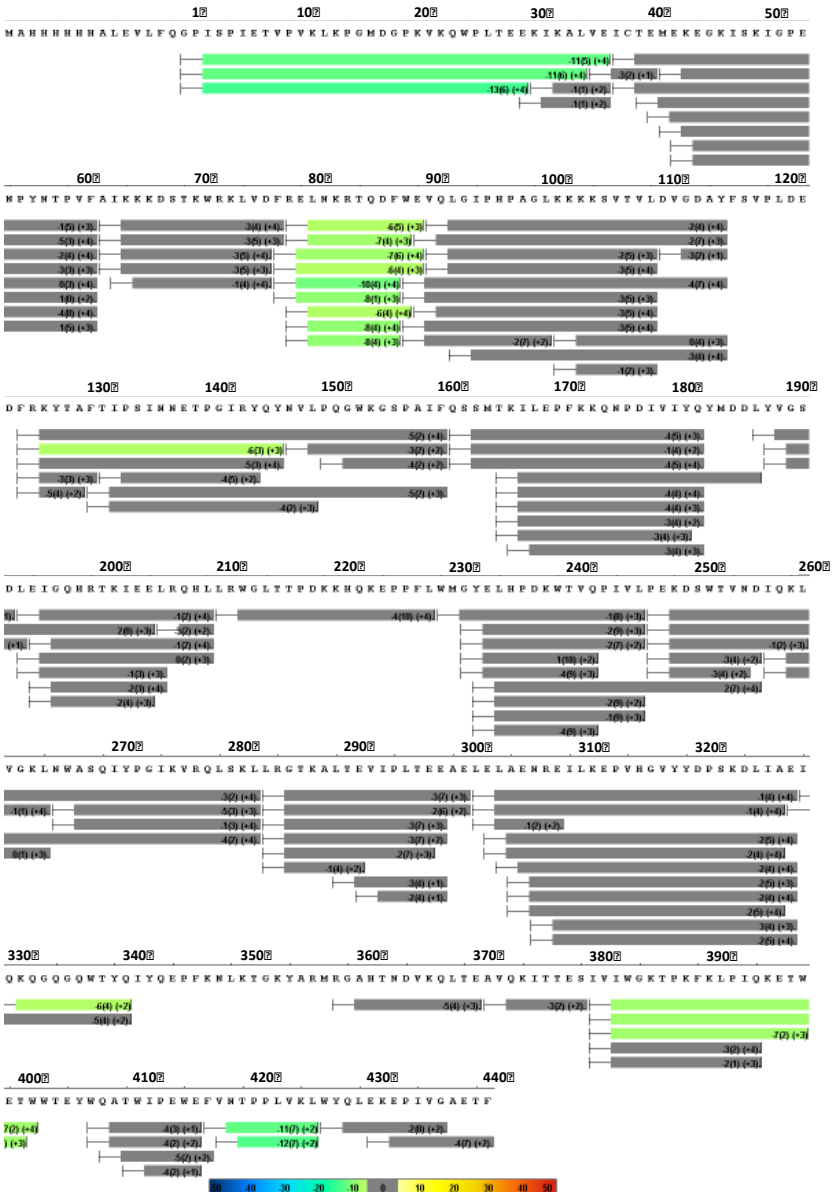


Figure AS1-8. Heat maps of difference in HDX signal in the presence and absence of 70.05core2 for (**A**) unlabeled p66 and (**B**) ^{15}N labeled p51 in p66/p51 RT. Numbering of amino acids, color scheme and interpretation of the heat map are presented as in Figure AS1-7. Higher resolution available at https://academic.oup.com/nar/article/48/5/2709/5707188

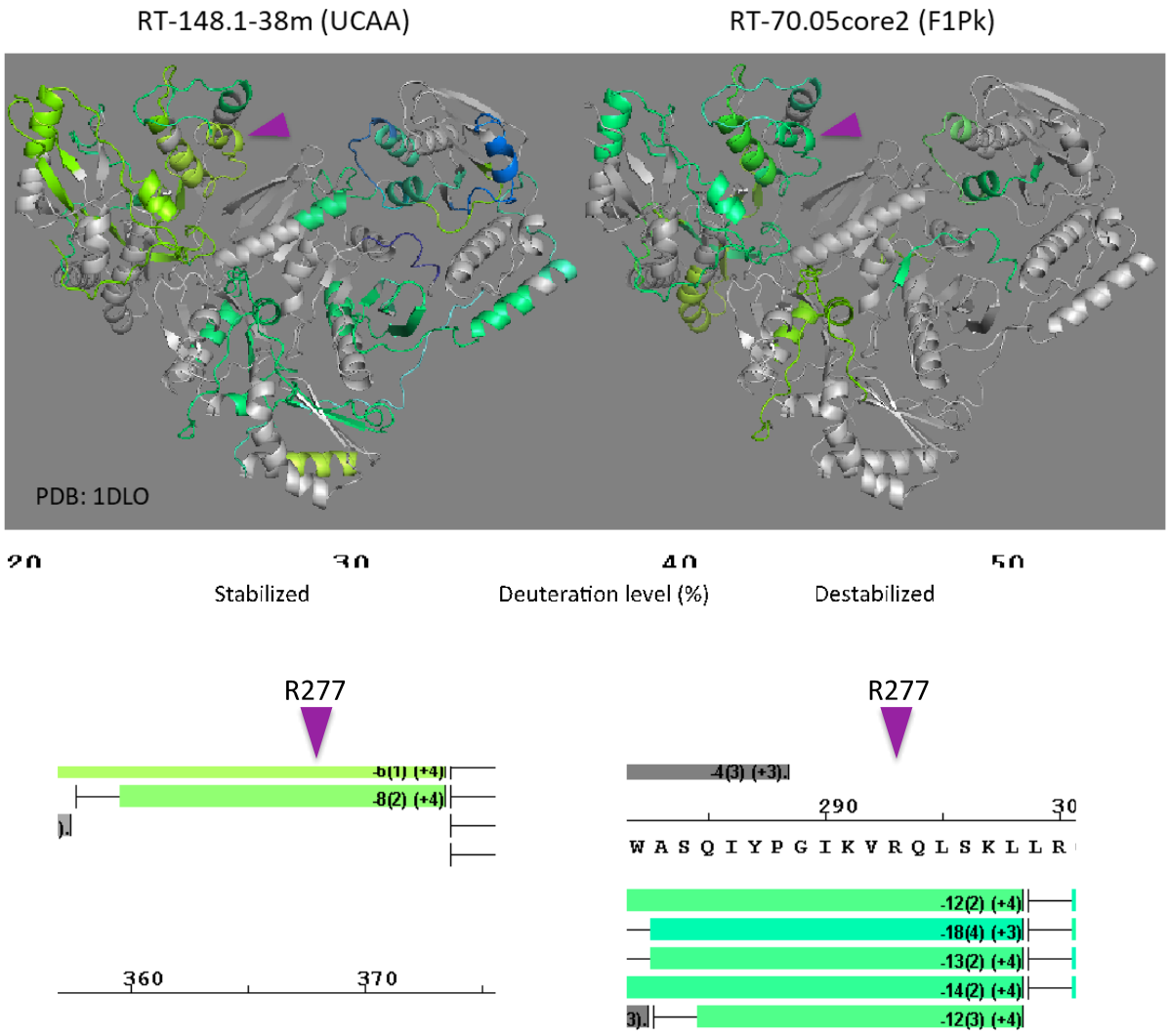


Figure AS1-9. Difference in the percent deuterium uptake at peptide containing amino acid R277 of RT-148.1-38m (left) compared to that of RT-70.05core2 (right). The difference in HDX of RT in the presence and absence of the 148.1-38m and 70.05core2 aptamers are mapped onto the crystal structure of ligand-free RT (PDB: 1DLO). The color scheme is presented as in Fig. 3. At position R277 (magenta arrow), the decrease in percent deuterium uptake corresponding to RT-70.05core2 is twice of that of RT-148.1-38m.

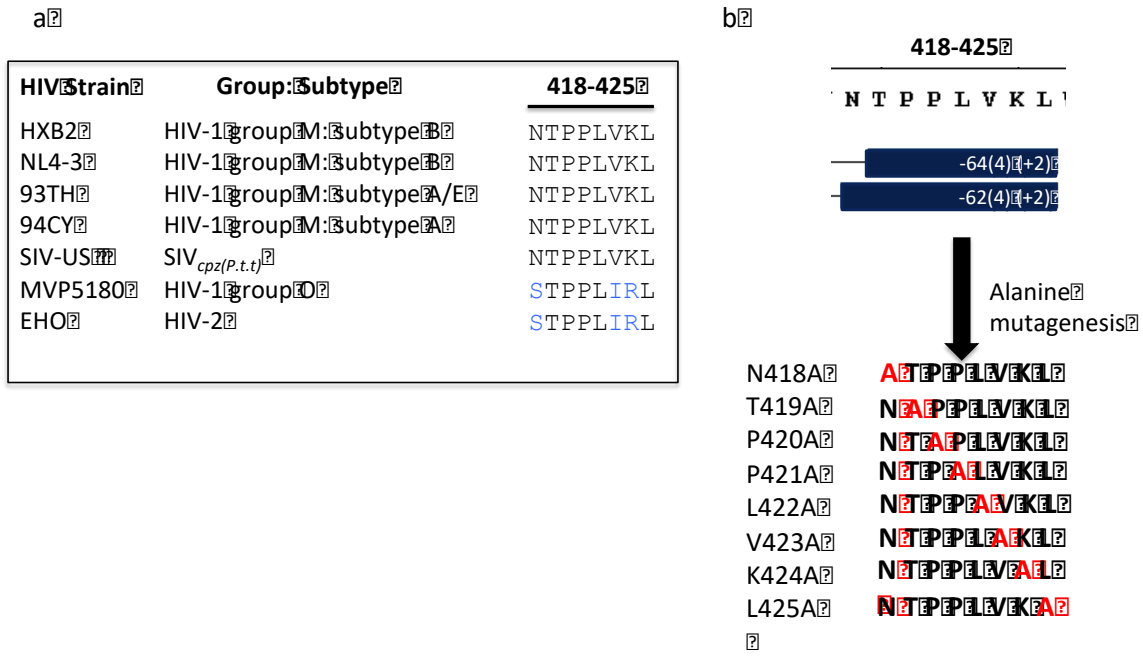


Figure AS1-10. Alanine mutagenesis study on the 418-NTPPLVKL-425 peptide of p51 subunit. **A**. Alignment of the 418-425 peptide from phylogenetically diverse RTs shows the level of conservation at each position. **B**. Schematic representation of the mutagenesis. Alanine mutations were introduced to each position of the peptide 418-425, one at a time. Dark-blue bars below the peptide indicate a strong decrease in percent deuterium uptake at the 418-NTPPLVKL-425 peptide when RT is in complex with 148.1-38m.

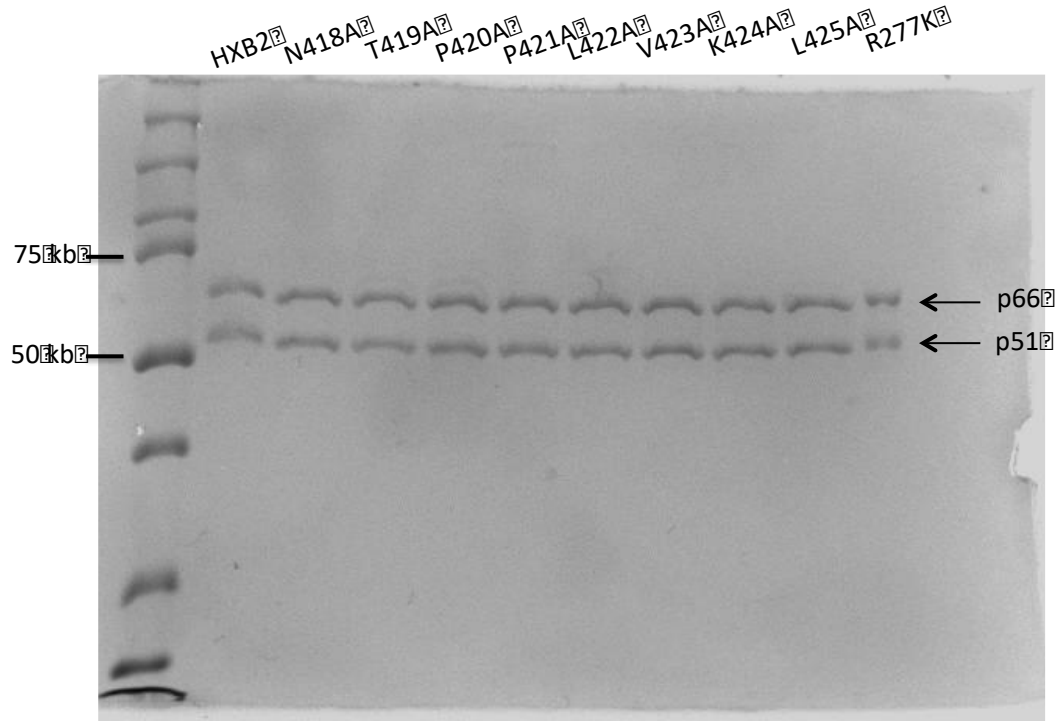


Figure AS1-11. Gel demonstrating the purity of WT and mutant RTs used in the manuscript. An aliquot of each RT are load on 10% SDS-PAGE and the gel is visualized by Coomassie staining. The gel shows no visible sign of additional bands other than those indicating the 2 subunits (p66 and p51) of RT. The upper and lower bands represent the p66 subunit and the p51 subunit of RT, respectively.

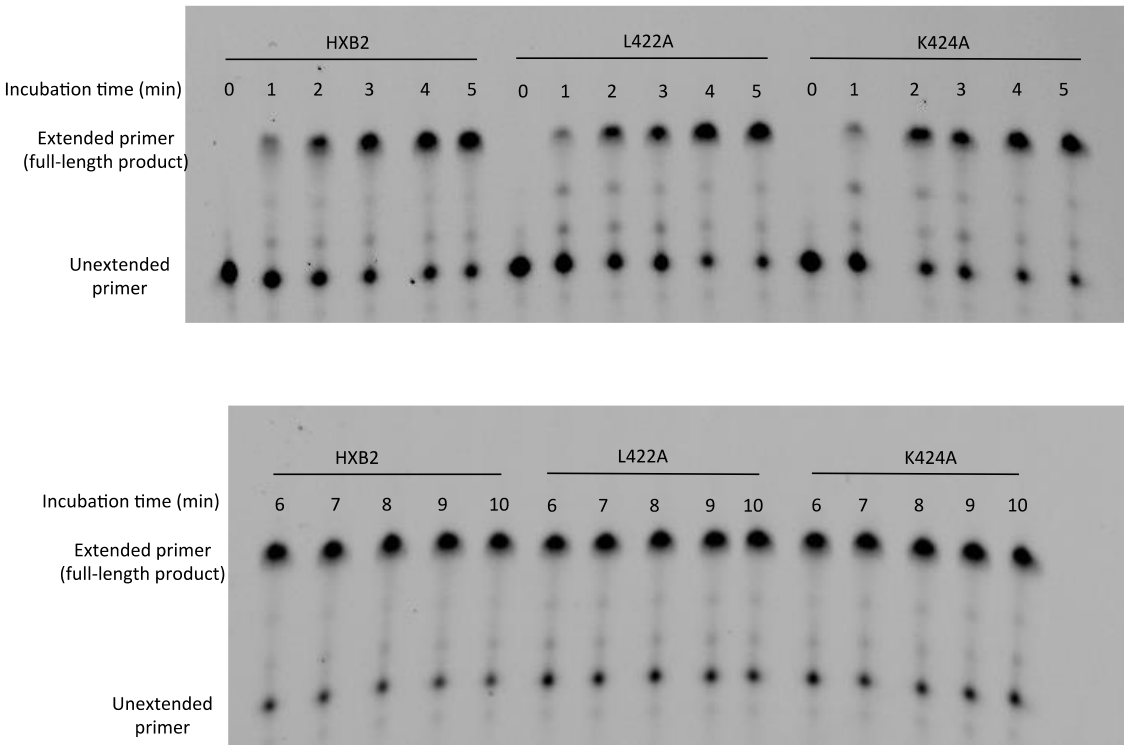


Figure AS1-12. A representative gel demonstrating polymerase activity of WT and mutant RTs (see Fig. 4b). Polymerase activity was evaluated using time course primer extension assay (more detailed information is in Material and Methods).

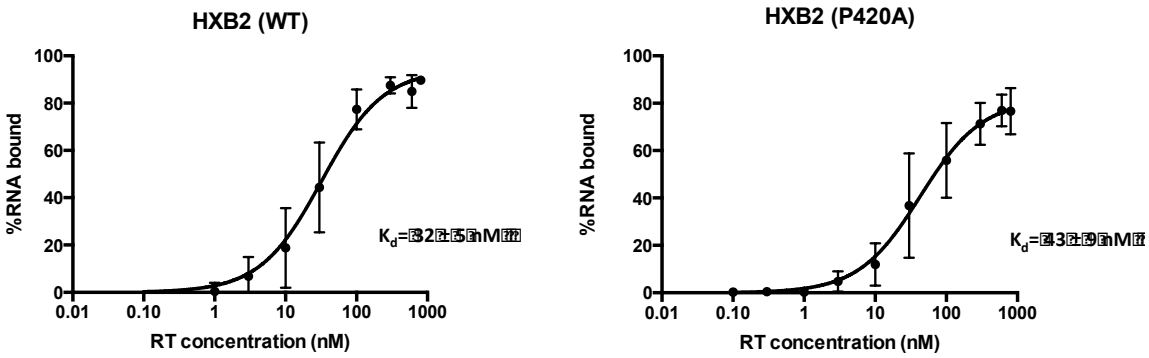


Figure AS1-13. Binding of 148.1-38m to wild-type HXB2 (left) and HXB2 (P420A) (right). Values shown represent triplicate measurements (±SD) of three independent experiments.

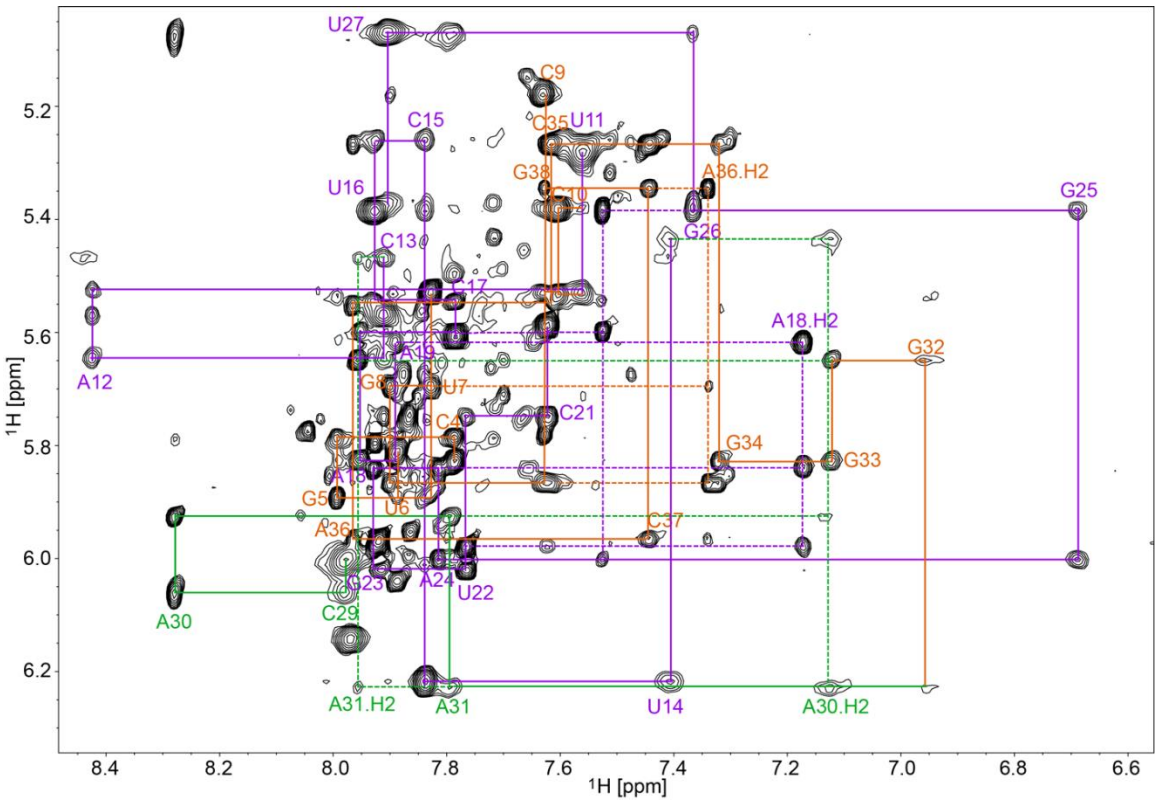


Figure AS1-14. Portion of the 2D 1H-1H NOESY spectrum collected for the 148.1-38 RNA in D2O. Inter- and intra- residue ribose-to-aromatic proton NOE connectivities are labeled with solid lines. Adenosine C2-H proton NOE connectivities are marked in dashed lines. The bottom stem residues C4-U11 and G32-G38 are labeled in orange. The top stem loop residues A12-U27 are labeled in purple. UCAA residues are labeled in green.

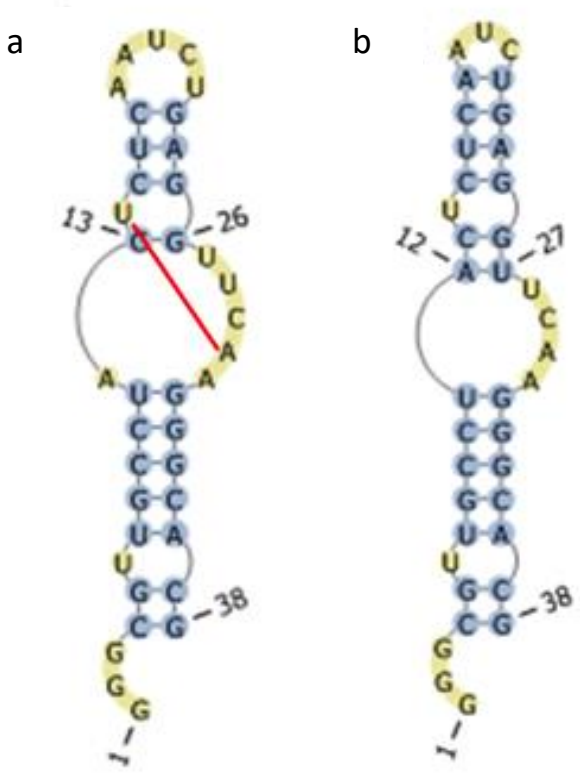


Figure AS1-15. 2D structure of RNA aptamer 148.1-38m. **A**. The 2D structure suggested by NMR data (with a tertiary contact between U14-A30 labeled by red line). **B**. The prediction from the free energy-based Vfold2D model.

Figure AS1-16. Representative predicted 3D structures for RNA aptamer 148.1-38m in the free state. The figure shows ten typical conformations predicted from the Vfold3D and IsRNA models. The red structure was selected for the final docking calculation. Three cytosines (CCC-3') were extended at the 3'-end to form base pairs with nucleotides G1-G3 at the 5'-end (5'-GGG).

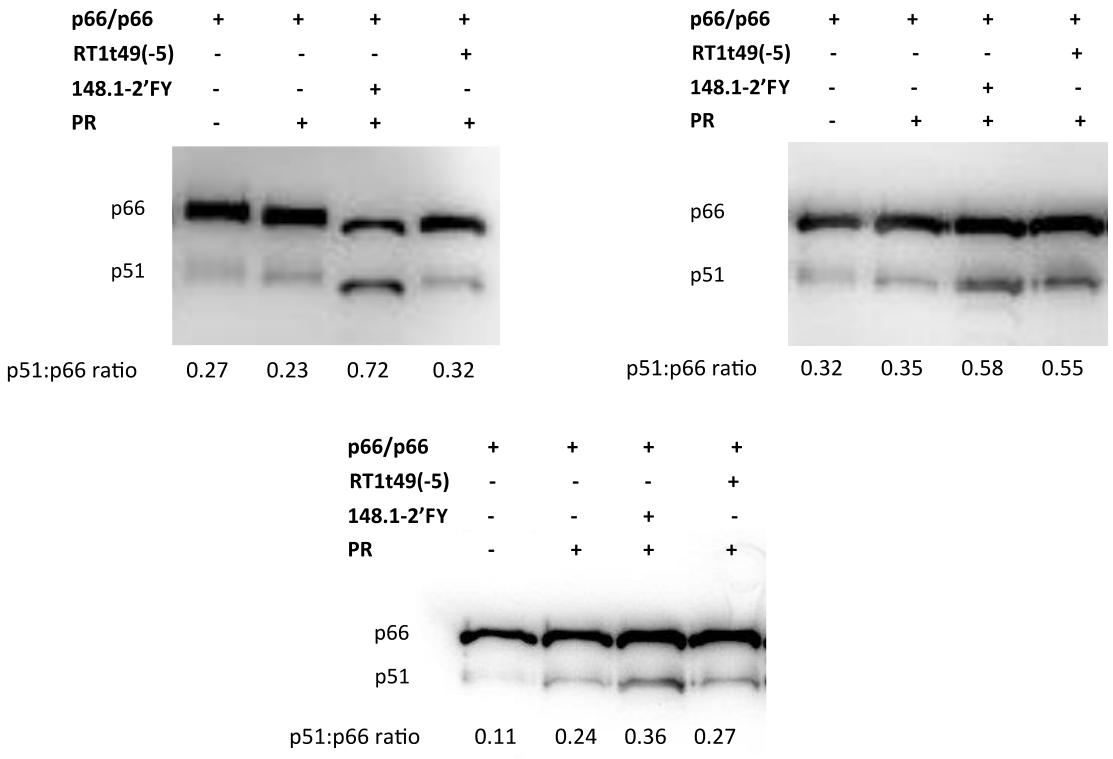


Figure AS1-17. PR cleavage of p66/p66 homodimer, analyzed by immunoblot analysis. Images show blots generated from three experiments, as plotted in Fig. 6. In the presence of 148.1-38m, PR cleavage of p66 was enhanced, indicated by an increase in intensity of p51 band compared to no aptamer samples in all three experiments.

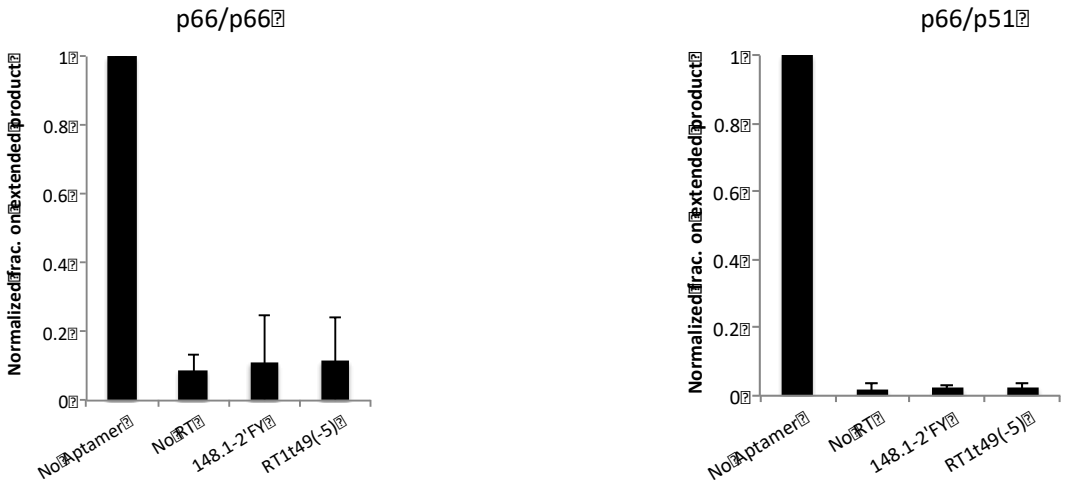


Figure AS1-18. Inhibition of (**A**) p66/p66 and (**B**) p66/p51 RTs by 2'F pyrimidine (2'FY) modified 148.1-38m and by ssDNA aptamer RT1t49(-5). 2'F modification of 148.1-38m did not interfere with RT-aptamer interaction, as both the 2'-OH and 2'FY versions of the aptamer efficiently inhibit homodimer and heterodimer forms of RT.

Biblography

1. Weiss, R.A. How does HIV cause AIDS? *Science (80-.).* **1993**, *260*, 1273–1279, doi:10.1126/science.8493571.
2. Sepkowitz, K.A. AIDS — The First 20 Years. *N. Engl. J. Med.* **2001**, *344*, 1764–1772, doi:10.1056/nejm200106073442306.
3. Moore, R.D.; Chaisson, R.E. Natural history of HIV infection in the era of combination antiretroviral therapy. *Aids* **1999**, *13*, 1933–1942, doi:10.1097/00002030-199910010-00017.
4. Dalgleish, A.G.; Clapham, P.R.; Crawford, D.H.; Greaves, M.F.; Weiss, R.A.; Beverley, P.C.; Clapham, P.R.; Crawford, D.H.; Greaves, M.F.; Weiss, R.A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **1984**, *321*, 763–767.
5. Freed, E.O. HIV-1 Replication. *Somat. Cell Mol. Genet.* **2002**, *26*, 13–33.
6. Mcdougal, J.S.; Kennedy, M.S.; Sligh, J.M.; Cort, S.P.; Mawle, A.; Nicholson, J.K.A. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science (80-.).* **1986**, *231*, 382–385, doi:10.1126/science.3001934.
7. Kwong, P.D.; Wyatt, R.; Robinson, J.; Sweet, R.W.; Sodroski, J.; Hendrickson, W.A. Structure of an HIV gp 120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 1998, *393*, 648–659.
8. Berger, E.A.; Murphy, P.M.; Farber, J.M. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* 1999, *17*, 657–700.
9. Speck, R.F.; Wehrly, K.; Platt, E.J.; Atchison, R.E.; Charo, I.F.; Kabat, D.; Chesebro, B.; Goldsmith, M.A. *Selective Employment of Chemokine Receptors as Human Immunodeficiency Virus Type 1 Coreceptors Determined by Individual Amino Acids within the Envelope V3 Loop*; 1997; Vol. 71;.

10. Blumenthal, R.; Durell, S.; Viard, M. HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* 2012, *287*, 40841–40849.

11. Campbell, E.M.; Hope, T.J. HIV-1 capsid: The multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* 2015, *13*, 471–483.

12. Roy, B.B.; Hu, J.; Guo, X.; Russell, R.S.; Guo, F.; Kleiman, L.; Liang, C. Association of RNA helicase A with human immunodeficiency virus type 1 particles. *J. Biol. Chem.* **2006**, *281*, 12625–12635, doi:10.1074/jbc.M510596200.

13. Li, C.; Burdick, R.C.; Nagashima, K.; Hu, W.-S.; Pathak, V.K. HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, doi:10.1073/pnas.2019467118.

14. Rankovic, S.; Varadarajan, J.; Ramalho, R.; Aiken, C.; Rousso, I. Reverse Transcription Mechanically Initiates HIV-1 Capsid Disassembly. *J. Virol.* **2017**, *91*, doi:10.1128/jvi.00289-17.

15. Cosnefroy, O.; Murray, P.J.; Bishop, K.N. HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology* **2016**, *13*, 58, doi:10.1186/s12977-016-0292-7.

16. Isel, C.; Ehresmann, C.; Marquet, R. Initiation of HIV reverse transcription. *Viruses* **2010**, *2*, 213–243, doi:10.3390/v2010213.

17. Hu, W.S.; Hughes, S.H. HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, doi:10.1101/cshperspect.a006882.

18. Beilhartz, G.L.; Götte, M. HIV-1 ribonuclease H: Structure, catalytic mechanism and inhibitors. *Viruses* **2010**, *2*, 900–926, doi:10.3390/v2040900.

19. Miller, M.D.; Farnet, C.M.; Bushman, F.D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **1997**, *71*,

5382–5390, doi:10.1128/jvi.71.7.5382-5390.1997.

20. Delelis, O.; Carayon, K.; Saïb, A.; Deprez, E.; Mouscadet, J.F. Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* 2008, *5*, 114.

21. Rivière, L.; Darlix, J.-L.; Cimarelli, A. Analysis of the Viral Elements Required in the Nuclear Import of HIV-1 DNA. *J. Virol.* **2010**, *84*, 729–739, doi:10.1128/jvi.01952-09.

22. Francis, A.C.; Marin, M.; Prellberg, M.J.; Palermino-Rowland, K.; Melikyan, G.B. HIV-1 Uncoating and Nuclear Import Precede the Completion of Reverse Transcription in Cell Lines and in Primary Macrophages. *Viruses* **2020**, *12*, 1234, doi:10.3390/v12111234.

23. Selyutina, A.; Persaud, M.; Lee, K.; KewalRamani, V.; Diaz-Griffero, F. Nuclear Import of the HIV-1 Core Precedes Reverse Transcription and Uncoating. *Cell Rep.* **2020**, *32*, doi:10.1016/j.celrep.2020.108201.

24. Delelis, O.; Carayon, K.; Saïb, A.; Deprez, E.; Mouscadet, J.F. Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* 2008, *5*, 114.

25. Craigie, R.; Bushman, F.D. HIV DNA integration (2012).pdf. **2012**, 1–18, doi:10.1101/cshperspect.a006890.

26. Liu, R.D.; Wu, J.; Shao, R.; Xue, Y.H. Mechanism and factors that control HIV-1 transcription and latency activation. *J. Zhejiang Univ. Sci. B* 2014, *15*, 455–465.

27. Liu, R.D.; Wu, J.; Shao, R.; Xue, Y.H. Mechanism and factors that control HIV-1 transcription and latency activation. *J. Zhejiang Univ. Sci. B* 2014, *15*, 455–465.

28. Emery, A.; Zhou, S.; Pollom, E.; Swanstrom, R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* **2017**, *91*, 2515–2531, doi:10.1128/jvi.02515-16.

29. Emery, A.; Zhou, S.; Pollom, E.; Swanstrom, R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* **2017**, *91*, 2515–2531, doi:10.1128/jvi.02515-16.

30. Emery, A.; Swanstrom, R. HIV-1: To Splice or Not to Splice, That Is the Question. *Viruses*

2021, *13*, 181, doi:10.3390/v13020181.

31. Freed, E.O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* 2015, *13*, 484–496.

32. Sundquist, W.I.; Kra, H. HIV-1 Assembly, Budding, and Maturation. **2012**, 1–24.

33. Bieniasz, P.; Telesnitsky, A. Multiple, switchable protein: RNA interactions regulate human immunodeficiency virus type 1 assembly. *Annu. Rev. Virol.* **2018**, *5*, 165–183, doi:10.1146/annurev-virology-092917-043448.

34. Sleiman, D.; Goldschmidt, V.; Barraud, P.; Marquet, R.; Paillart, J.C.; Tisné, C. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res.* **2012**, *169*, 324–339, doi:10.1016/j.virusres.2012.06.006.

35. Jones, C.P.; Saadatmand, J.; Kleiman, L.; Musier-Forsyth, K. Molecular mimicry of human tRNALys anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. *RNA* **2013**, *19*, 219–229, doi:10.1261/rna.036681.112.

36. Comandur, R.; Olson, E.D.; Musier-Forsyth, K. Conservation of tRNA mimicry in the 5'-untranslated region of distinct HIV-1 subtypes. *Rna* **2017**, *23*, 1850–1859, doi:10.1261/rna.062182.117.

37. Heng, X.; Kharytonchyk, S.; Garcia, E.L.; Lu, K.; Divakaruni, S.S.; Lacotti, C.; Edme, K.; Telesnitsky, A.; Summers, M.F. Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J. Mol. Biol.* **2012**, *417*, 224–239, doi:10.1016/j.jmb.2012.01.033.

38. Brigham, B.S.; Kitzrow, J.P.; Reyes, J.P.C.; Musier-Forsyth, K.; Munro, J.B. Intrinsic conformational dynamics of the HIV-1 genomic RNA 5'UTR. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 10372–10381, doi:10.1073/pnas.1902271116.

39. Nikolaitchik, O.A.; Dilley, K.A.; Fu, W.; Gorelick, R.J.; Tai, S.-H.S.; Soheilian, F.; Ptak, R.G.;

Nagashima, K.; Pathak, V.K.; Hu, W.-S. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. *PLoS Pathog.* **2013**, *9*, e1003249, doi:10.1371/journal.ppat.1003249.

40. Kharytonchyk, S.; Monti, S.; Smaldino, P.J.; Van, V.; Bolden, N.C.; Brown, J.D.; Russo, E.; Swanson, C.; Shuey, A.; Telesnitsky, A.; et al. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 201616627, doi:10.1073/pnas.1616627113.

41. Brown, J.D.; Kharytonchyk, S.; Chaudry, I.; Iyer, A.S.; Carter, H.; Becker, G.; Desai, Y.; Glang, L.; Choi, S.H.; Singh, K.; et al. Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science (80-.).* **2020**, *368*, 413–417, doi:10.1126/science.aaz7959.

42. Keane, S.C.; Van, V.; Frank, H.M.; Sciandra, C.A.; McCowin, S.; Santos, J.; Heng, X.; Summers, M.F. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13033–13038, doi:10.1073/pnas.1614785113.

43. Lu, K.; Heng, X.; Summers, M.F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **2011**, *410*, 609–633, doi:10.1016/j.jmb.2011.04.029.

44. Marquet, R.; Isel, C.; Ehresmann, C.; Ehresmann, B. tRNAs as primer of reverse transcriptases. *Biochimie* **1995**, *77*, 113–124, doi:10.1016/0300-9084(96)88114-4.

45. Rhim, H.; Park, J.; Morrow, C.D. Deletions in the tRNA(Lys) primer-binding site of human immunodeficiency virus type 1 identify essential regions for reverse transcription. *J. Virol.* **1991**, *65*, 4555–4564, doi:10.1128/jvi.65.9.4555-4564.1991.

46. Geretti, A.M. HIV-1 subtypes: Epidemiology and significance for HIV management. *Curr. Opin. Infect. Dis.* **2006**, *19*, 1–7, doi:10.1097/01.qco.0000200293.45532.68.

47. Green, L.M.; Berg, J.M. Retroviral nucleocapsid protein-metal ion interactions: Folding and sequence variants (zinc/nucleic acid-binding protein/metal-binding domain). *Proc.*

Natl. Acad. Sci. USA **1990**, *87*, 6403–6407, doi:NEJMoa1001286 [pii]10.1056/NEJMoa1001286.

48. Hargittai, M.R.S.; Gorelick, R.J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004**, *337*, 951–968, doi:10.1016/j.jmb.2004.01.054.

49. Barraud, P.; Gaudin, C.; Dardel, F.; Tisné, C. New insights into the formation of HIV-1 reverse transcription initiation complex. *Biochimie* **2007**, *89*, 1204–1210, doi:10.1016/j.biochi.2007.01.016.

50. Larsen, K.P.; Mathiharan, Y.K.; Kappel, K.; Coey, A.T.; Chen, D.H.; Barrero, D.; Madigan, L.; Puglisi, J.D.; Skiniotis, G.; Puglisi, E.V. Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature* **2018**, *557*, 118–122, doi:10.1038/s41586-018-0055-9.

51. Wakefield, J.K.; Kang, S.M.; Morrow, C.D. Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His). *J Virol* **1996**, *70*, 966–975.

52. Zhang, Z.; Kang, S.M.; LeBlanc, A.; Hajduk, S.L.; Morrow, C.D. 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* **1996**, *226*, 306–317, doi:10.1006/viro.1996.0658.

53. Kang, S.-M.; Wakefield, J.K.; Morrow, C.D. 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* **1996**, *222*, 401–14, doi:10.1006/viro.1996.0437.

54. Isel, C.; Lanchy, J.M.; Le Grice, S.F.J.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. *EMBO J.*

1996, *15*, 917–24.

55. Liang, C.; Li, X.; Rong, L.; Inouye, P.; Quan, Y.; Kleiman, L.; Wainberg, M.A. The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J Virol* **1997**, *71*, 5750–5757.

56. Kang, S.M.; Zhang, Z.; Morrow, C.D. Identification of a human immunodeficiency virus type 1 that stably uses tRNA(Lys1,2) rather than tRNA(Lys,3) for initiation of reverse transcription. *Virology* **1999**, *257*, 95–105, doi:10.1006/viro.1999.9615.

57. Das, A.T.; Klaver, B.; Berkhout, B. Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *J. Virol.* **1995**, *69*, 3090–3097, doi:10.1128/jvi.69.5.3090-3097.1995.

58. Li, X.; Mak, J.; Arts, E.J.; Gu, Z.; Kleiman, L.; Wainberg, M.A.; Parniak, M.A. Effects of alterations of primer-binding site sequences on human immunodeficiency virus type 1 replication. *J. Virol.* **1994**, *68*, 6198–6206, doi:10.1128/jvi.68.10.6198-6206.1994.

59. Isel, C.; Marquet, R.; Keith, G.; Ehresmann, C.; Ehresmann, B. Modified nucleotides of tRNA3/(Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* **1993**, *268*, 25269–25272.

60. Isel, C.; Westhof, E.; Massire, C.; Le Grice, S.F.J.; Ehresmann, B.; Ehresmann, C.; Marquet, R. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* **1999**, *18*, 1038–1048, doi:10.1093/emboj/18.4.1038.

61. Puglisi, E.V.; Puglisi, J.D. Secondary structure of the HIV reverse transcription initiation complex by NMR. *J Mol Biol.* **2011**, *410*, 863–874, doi:10.1016/j.jmb.2011.04.024.Secondary.

62. Coey, A.; Larsen, K.; Puglisi, J.D.; Puglisi, E.V. Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *Rna* **2016**,

22, 1689–1698, doi:10.1261/rna.056804.116.

63. Tisne, C.; Roques, B.P.; Dardel, F. The Annealing Mechanism of HIV-1 Reverse Transcription Primer onto the Viral Genome. *J. Biol. Chem.* **2004**, *279*, 3588–3595, doi:10.1074/jbc.M310368200.

64. Dupuy, L.C.; Kelly, N.J.; Elgavish, T.E.; Harvey, S.C.; Morrow, C.D. Probing the Importance of tRNA Anticodon: Human Immunodeficiency Virus Type 1 (HIV-1) RNA Genome Complementarity with an HIV-1 That Selects tRNAGlu for Replication. *J. Virol.* **2003**, *77*, 8756–8764, doi:10.1128/jvi.77.16.8756-8764.2003.

65. Kang, S.M.; Zhang, Z.; Morrow, C.D. 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.* **1997**, *71*, 207–217, doi:10.1128/jvi.71.1.207-217.1997.

66. Zhang, Z.; Kang, S.; Li, Y.U.N.; Morrow, C.D. 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. **1998**, 394–406.

67. Ooms, M.; Cupac, D.; Abbink, T.E.M.; Huthoff, H.; Berkhout, B. The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. *Nucleic Acids Res.* **2007**, *35*, 1649–1659, doi:10.1093/nar/gkm046.

68. Beerens, N.; Groot, F.; Berkhout, B. Initiation of HIV-1 Reverse Transcription Is Regulated by a Primer Activation Signal *. **2001**, *276*, 31247–31256, doi:10.1074/jbc.M102441200.

69. Huthoff, H.; Bugala, K.; Barciszewski, J.; Berkhout, B. On the importance of the primer activation signal for initiation of tRNA lys3 -primed reverse transcription of the HIV-1 RNA genome. **2003**, *31*, 5186–5194, doi:10.1093/nar/gkg714.

70. Goldschmidt, V.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Does the HIV-1 primer

activation signal interact with tRNA3Lys during the initiation of reverse transcription. *Nucleic Acids Res.* **2003**, *31*, 850–859, doi:10.1093/nar/gkg187.

71. Boeras, I.; Song, Z.; Moran, A.; Franklin, J.; Brown, W.C.; Johnson, M.; Boris-Lawrie, K.; Heng, X. DHX9/RHA Binding to the PBS-Segment of the Genomic RNA during HIV-1 Assembly Bolsters Virion Infectivity. *J. Mol. Biol.* **2016**, doi:10.1016/j.jmb.2016.04.011.

72. Li, J.; Tang, H.; Mullen, T.M.; Westberg, C.; Reddy, T.R.; Rose, D.W.; Wong-Staal, F. A role for RNA helicase A in post-transcriptional regulation of HIV type 1. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 709–714, doi:10.1073/pnas.96.2.709.

73. Naji, S.; Ambrus, G.; Cimermančič, P.; Reyes, J.R.; Johnson, J.R.; Filbrandt, R.; Huber, M.D.; Vesely, P.; Krogan, N.J.; Yates, J.R.; et al. Host cell interactome of HIV-1 Rev includes RNA helicases involved in multiple facets of virus production. *Mol. Cell. Proteomics* **2012**, *11*, doi:10.1074/mcp.M111.015313.

74. Lee, C.G.; Hurwitz, J. A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3’ to 5’ direction. *J. Biol. Chem.* **1992**, *267*, 4398–4407, doi:10.1016/s0021-9258(18)42849-9.

75. Zhang, S.; Grosse, F. Domain structure of human nuclear DNA helicase II (RNA helicase A). *J. Biol. Chem.* **1997**, *272*, 11487–11494, doi:10.1074/jbc.272.17.11487.

76. Brady, S.; Singh, G.; Bolinger, C.; Song, Z.; Boeras, I.; Weng, K.; Trent, B.; Brown, W.C.; Singh, K.; Boris-Lawrie, K.; et al. Virion-associated, host-derived DHX9/RNA helicase A enhances the processivity of HIV-1 reverse transcriptase on genomic RNA. *J. Biol. Chem.* **2019**, *294*, 11473–11485, doi:10.1074/jbc.RA119.007679.

77. Arts, E.J.; Hazuda, D.J. HIV-1 antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, doi:10.1101/cshperspect.a007161.

78. Young, F.E. The role of the FDA in the effort against AIDS. *Public Health Rep.* **1988**, *103*,

242–245.

79. Ghosh, A.K.; Osswald, H.L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *J. Med. Chem.* 2016, *59*, 5172–5208.

80. Henrich, T.J.; Kuritzkes, D.R. HIV-1 entry inhibitors: Recent development and clinical use. *Curr. Opin. Virol.* 2013, *3*, 51–57.

81. Huang, L.S.M.; Snyder, E.Y.; Schooley, R.T. Strategies and progress in CXCR4-targeted anti-HIV therapeutic development. *Clin. Infect. Dis.* **2021**, doi:10.1093/cid/ciab160.

82. Eggink, D.; Berkhout, B.; W. Sanders, R. Inhibition of HIV-1 by Fusion Inhibitors. *Curr. Pharm. Des.* **2010**, *16*, 3716–3728, doi:10.2174/138161210794079218.

83. Zhang, Y.; Lai, B.S.; Juhas, M. Recent advances in aptamer discovery and applications. *Molecules* 2019, *24*.

84. Zhou, J.; Rossi, J. Aptamers as targeted therapeutics: Current potential and challenges. *Nat. Rev. Drug Discov.* 2017, *16*, 181–202.

85. Bala, J.; Chinnapaiyan, S.; Dutta, R.K.; Unwalla, H. Aptamers in HIV research diagnosis and therapy. *RNA Biol.* 2018, *15*, 327–337.

86. Robertson, M.P.; Joyce, G.F. The origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, 1, doi:10.1101/cshperspect.a003608.

87. Bernhardt, H.S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biol. Direct* 2012, *7*, 23.

88. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **1982**, *31*, 147–157, doi:10.1016/0092-8674(82)90414-7.

89. Illangasekare, M.; Sanchez, G.; Nickles, T.; Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science (80-.).* **1995**, *267*, 643–647, doi:10.1126/science.7530860.

90. Seelig, B.; Jäschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **1999**, *6*, 167–176, doi:10.1016/S1074-5521(99)89008-5.

91. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849–857, doi:10.1016/0092-8674(83)90117-4.

92. Dai, X.; De Mesmaeker, A.; Joyce, G.F. Cleavage of an amide bond by a ribozyme. *Science (80-.).* **1995**, *267*, 237–240, doi:10.1126/science.7809628.

93. Tsukiji, S.; Pattnaik, S.B.; Suga, H. An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.* **2003**, *10*, 713–717, doi:10.1038/nsb964.

94. Fraaije, M.W.; Mattevi, A. Flavoenzymes: Diverse catalysts with recurrent features. *Trends Biochem. Sci.* 2000, *25*, 126–132.

95. Burgstaller, P.; Famulok, M. Isolation of RNA Aptamers for Biological Cofactors by In Vitro Selection. *Angew. Chemie Int. Ed. English* **1994**, *33*, 1084–1087, doi:10.1002/anie.199410841.

96. Lauhon, C.T.; Szostak, J.W. RNA Aptamers that Bind Flavin and Nicotinamide Redox Cofactors. *J. Am. Chem. Soc.* **1995**, *117*, 1246–1257, doi:10.1021/ja00109a008.

97. Fan, P.; Suri, A.K.; Fiala, R.; Live, D.; Patel, D.J. Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **1996**, *258*, 480–500, doi:10.1006/jmbi.1996.0263.

98. Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **2003**, *4*, 936–962, doi:10.1002/cbic.200300700.

99. Sakamoto, T.; Otsu, M.; Kawai, G. NMR studies on RNA. In *Experimental Approaches of NMR Spectroscopy: Methodology and Application to Life Science and Materials Science*; Springer Singapore, 2017; pp. 439–459 ISBN 9789811059667.

100. Zhang, H.; Keane, S.C. Advances that facilitate the study of large RNA structure and

dynamics by nuclear magnetic resonance spectroscopy. *Wiley Interdiscip. Rev. RNA* 2019, *10*, e1541.

101. Barnwal, R.P.; Yang, F.; Varani, G. Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.* **2017**, *628*, 42–56, doi:10.1016/j.abb.2017.06.003.

102. Baltimore, D. Viral RNA-dependent DNA Polymerase. *Nature* **1970**, *226*, 1209–1211.

103. Goldschmidt, V.; Paillart, J.C.; Rigourd, M.; Ehresmann, B.; Aubertin, A.M.; Ehresmann, C.; Marquet, R. Structural variability of the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* **2004**, *279*, 35923–35931, doi:10.1074/jbc.M404473200.

104. Berkhout, B.; Schoneveld, I. Secondary structure of the HIV-2 leader RNA comprising the tRNA-primer binding site. *Nucleic Acids Res.* **1993**, *21*, 1171–1178, doi:10.1093/nar/21.5.1171.

105. Coey, A.; Larsen, K.; Puglisi, J.D.; Viani Puglisi, E. Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *Rna* **2016**, *22*, 1689–1698, doi:10.1261/rna.056804.116.

106. Beerens, N.; Berkhout, B. The tRNA Primer Activation Signal in the Human Immunodeficiency Virus Type 1 Genome Is Important for Initiation and Processive Elongation of Reverse Transcription. *J. Virol.* **2002**, *76*, 2329–2339, doi:10.1128/jvi.76.5.2329-2339.2002.

107. Beerens, N.; Groot, F.; Berkhout, B. Initiation of HIV-1 Reverse Transcription is Regulated by a Primer Activation Signal. *J. Biol. Chem.* **2001**, *276*, 31247–31256, doi:10.1074/jbc.M102441200.

108. Lange, M.J.; Burke, D.H. Screening inhibitory potential of anti-HIV RT RNA aptamers. *Methods Mol. Biol.* **2014**, *1103*, 11–29, doi:10.1007/978-1-62703-730-3_2.

109. Lanchy, J.M.; Ehresmann, C.; Le Grice, S.F.; Ehresmann, B.; Marquet, R. Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO J.* **1996**, *15*, 7178–87.

110. Mbisa, J.L.; Delviks-Frankenberry, K.A.; Thomas, J.A.; Gorelick, R.J.; Pathak, V.K. Real-Time PCR Analysis of HIV-1 Replication Post-entry Events. *Methods Mol Biol. 2009* **2009**, *485*, 55–72, doi:10.1007/978-1-59745-170-3.

111. Zhang, H.; Dornadula, G.; Pomerantz, R.J. Natural endogenous reverse transcription of HIV-1. *J. Reprod. Immunol.* **1998**, *41*, 255–260, doi:10.1016/S0165-0378(98)00062-X.

112. Lori, F.; di Marzo Veronese, F.; de Vico, A.L.; Lusso, P.; Reitz, M.S.; Gallo, R.C. Viral DNA carried by human immunodeficiency virus type 1 virions. *J. Virol.* **1992**, *66*, 5067–5074, doi:10.1128/jvi.66.8.5067-5074.1992.

113. Trono, D. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *J. Virol.* **1992**, *66*, 4893–4900, doi:10.1128/jvi.66.8.4893-4900.1992.

114. Gotte, M.; Fackler, S.; Hermann, T.; Perola, E.; Cellai, L.; Gross, H.J.; Grice, S.F.J. Le; Heumann, H. HIV-1 reverse transcriptase-associated RNase H cleaves RNA/RNA in arrested complexes: implications for the mechanism by which RNase H discriminates beween RNA/RNA and RNA/DNA. **1995**, *14*, 833–841.

115. Ben-artzi, H.; Zeelon, E.; Le-grice, S.F. j.; Gorecki, M.; Panet, A. Characterization of the double stranded RNA dependent RNase activity associated with recombinant reverse transcriptases. *Nucleic Acids Res.* **1992**, *20*, 5115–5118, doi:10.1093/nar/20.19.5115.

116. Mizrahi, V.; Brooksbank, R.; Nkabinde, N. Mutagenesis of the Conserved Aspartic Acid 443, Glutamic Acid 478, Asparagine 494, and Aspartic Acid 498 Residues in the Ribonuclease H Domain of p66/p51 Human Immunodeficiency Virus Type I Reverse Transcriptase. *J. Biol. Chem.* **1994**, *269*, 19245–19249.

117. Destefano, J.J.; Seehra, J.; Mccoy, J.; Laston, D.; Albone, E.; Fay, P.J.; Bambara, R.A. Characterization of an RNase H deficient mutant of human immunodeficiency virus-1 reverse transcriptase having an aspartate to asparagine change at position 498. *Biochim. Biophys. Acta* **1994**, *1219*, 380–388.

118. Lange, M.J.; Sharma, T.K.; Whatley, A.S.; Landon, L.A.; Tempesta, M.A.; Johnson, M.C.; Burke, D.H. Robust suppression of HIV replication by intracellularly expressed reverse transcriptase aptamers is independent of ribozyme processing. *Mol. Ther.* **2012**, *20*, 2304–2314, doi:10.1038/mt.2012.158.

119. Li, X.; Quan, Y.; Arts, E.J.; Li, Z.; Preston, B.D.; de Rocquigny, H.; Roques, B.P.; Darlix, J.L.; Kleiman, L.; Parniak, M.A.; et al. Human immunodeficiency virus Type 1 nucleocapsid protein (NCp7) directs specific initiation of minus-strand DNA synthesis primed by human tRNA(Lys3) in vitro: studies of viral RNA molecules mutated in regions that flank the primer binding site. *J. Virol.* **1996**, *70*, 4996–5004, doi:10.1128/jvi.70.8.4996-5004.1996.

120. Brulé, F.; Marquet, R.; Rong, L.; Wainberg, M.A.; Roques, B.P.; Le Grice, S.F.J.; Ehresmann, B.; Ehresmann, C. Structural and functional properties of the HIV-1 RNA-tRNA3Lys primer complex annealed by the nucleocapsid protein: Comparison with the heat-annealed complex. *Rna* **2002**, *8*, 8–15, doi:10.1017/S1355838202010981.

121. Wakefield, J.K.; Wolf, A.G.; Morrow, C.D. Human immunodeficiency virus type 1 can use different tRNAs as primers for reverse transcription but selectively maintains a primer binding site complementary to tRNA(3Lys). *J. Virol.* **1995**, *69*, 6021–6029, doi:10.1128/jvi.69.10.6021-6029.1995.

122. Rigourd, M.; Goldschmidt, V.; Brulé, F.; Morrow, C.D.; Ehresmann, B.; Ehresmann, C.; Marquet, R. Structure-function relationships of the initiation complex of HIV-1 reverse transcription: The case of mutant viruses using tRNAHis as primer. *Nucleic Acids Res.*

2003, *31*, 5764–5775, doi:10.1093/nar/gkg754.

123. Beerens, N.; Berkhout, B. Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *Rna* **2002**, *8*, 357–369, doi:10.1017/S1355838202028194.

124. Beerens, N.; Jepsen, M.D.E.; Nechyporuk-Zloy, V.; Krüger, A.C.; Darlix, J.L.; Kjems, J.; Birkedal, V. Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *Rna* **2013**, *19*, 517–526, doi:10.1261/rna.035733.112.

125. Goldschmidt, V.; Rigourd, M.; Ehresmann, C.; Le Grice, S.F.J.; Ehresmann, B.; Marquet, R. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. *J. Biol. Chem.* **2002**, *277*, 43233–43242, doi:10.1074/jbc.M205295200.

126. Seif, E.; Niu, M.; Kleiman, L. In virio SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology* **2015**, *12*, 40, doi:10.1186/s12977-015-0171-7.

127. Miller, S.B.; Yildiz, F.Z.; Lo, J.A.; Wang, B.; D'Souza, V.M. A structure-based mechanism for tRNA and retroviral RNA remodelling during primer annealing. *Nature* **2014**, *515*, 591–595, doi:10.1038/nature13709.

128. Coey, A.T.; Larsen, K.P.; Choi, J.; Barrero, D.J.; Puglisi, J.D.; Puglisi, E.V. Dynamic Interplay of RNA and Protein in the Human Immunodeficiency Virus-1 Reverse Transcription Initiation Complex. *J Mol Biol.* **2018**, *430*, 5137–5150, doi:10.1016/j.jmb.2018.08.029.Dynamic.

129. Ben-Artzi, H.; Zeelon, E.; Gorecki, M.; Panet, A. Double-stranded RNA-dependent RNase activity associated with human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 927–931, doi:10.1073/pnas.89.3.927.

130. Hostomsky, Z.; Hudson, G.O.; Rahmati, S.; Hostomska, Z. RNase D, a reported new activity associated with HIV-1 reverse transcriptase, displays the same cleavage specificity as Escherichia coli RNase III. *Nucleic Acids Res.* **1992**, *20*, 5819–5824, doi:10.1093/nar/20.21.5819.

131. Liu, S.; Harada, B.T.; Miller, J.T.; Le Grice, S.F.J.; Zhuang, X. Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1453–1460, doi:10.1038/nsmb.1937.

132. Vaccaro, J.A.; Singh, H.A.; Anderson, K.S. Initiation of minus-strand DNA synthesis by human immunodeficiency virus type 1 reverse transcriptase. *Biochemistry* **1999**, *38*, 15978–15985, doi:10.1021/bi990945x.

133. Kati, W.M.; Johnson, K.A.; Jerva, L.F.; Anderson, K.S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **1992**, *267*, 25988–25997.

134. Alizon, M.; Wain-Hobson, S.; Montagnier, L.; Sonigo, P. Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients. *Cell* **1986**, *46*, 63–74, doi:10.1016/0092-8674(86)90860-3.

135. Alam, K.K.; Chang, J.L.; Lange, M.J.; Nguyen, P.D.M.; Sawyer, A.W.; Burke, D.H. Poly-Target Selection Identifies Broad-Spectrum RNA Aptamers. *Mol. Ther. - Nucleic Acids* **2018**, *13*, 605–619, doi:10.1016/j.omtn.2018.10.010.

136. Nguyen, P.D.M.; Zheng, J.; Gremminger, T.J.; Qiu, L.; Zhang, D.; Tuske, S.; Lange, M.J.; Griffin, P.R.; Arnold, E.; Chen, S.J.; et al. Binding interface and impact on protease cleavage for an RNA aptamer to HIV-1 reverse transcriptase. *Nucleic Acids Res.* **2020**, *48*, 2709–2722, doi:10.1093/nar/gkz1224.

137. Lee, B.M.; De Guzman, R.N.; Turner, B.G.; Tjandra, N.; Summers, M.F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **1998**, *279*, 633–649,

doi:10.1006/jmbi.1998.1766.

138. Maciejewski, M.W.; Schuyler, A.D.; Gryk, M.R.; Moraru, I.I.; Romero, P.R.; Ulrich, E.L.; Eghbalnia, H.R.; Livny, M.; Delaglio, F.; Hoch, J.C. NMRbox: A Resource for Biomolecular NMR Computation. *Biophys. J.* **2017**, *112*, 1529–1534, doi:10.1016/j.bpj.2017.03.011.

139. Hill, K.J.; Rogers, L.C.; Njenda, D.T.; Burke, D.H.; Sarafianos, S.G.; Sönnerborg, A.; Neogi, U.; Singh, K. Strain-specific Effect on Biphasic DNA Binding by HIV-1 Integrase. *Aids* **2019**, *33*, 588–592, doi:10.1097/QAD.0000000000002078.

140. Coffin, J.M.; Hughes, S.H.; Varmus, H.E. *Retroviruses*; Coffin, J.M., Hughes, S.H., Varmus, H.E., Eds.; Cold Spring Harbor (NY), 1997; ISBN 0-87969-571-4.

141. De Rocquigny, H.; Gabus, C.; Vincent, A.; Fournie-Zaluski, M.C.; Roques, B.; Darlix, J.L. Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 6472–6476, doi:10.1073/pnas.89.14.6472.

142. Hargittai, M.R.S.; Gorelick, R.J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004**, *337*, 951–968, doi:10.1016/j.jmb.2004.01.054.

143. Cen, S.; Khorchid, A.; Gabor, J.; Rong, L.; Wainberg, M.A.; Kleiman, L. Roles of Pr55gag and NCp7 in tRNA3Lys Genomic Placement and the Initiation Step of Reverse Transcription in Human Immunodeficiency Virus Type 1. *J. Virol.* **2000**, *74*, 10796–10800, doi:10.1128/jvi.74.22.10796-10800.2000.

144. Beerens, N.; Jepsen, M.D.E.; Nechyporuk-Zloy, V.; Krüger, A.C.; Darlix, J.-L.; Kjems, J.; Birkedal, V. Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA* **2013**, *19*, 517–26, doi:10.1261/rna.035733.112.

145. Cen, S.; Khorchid, A.; Javanbakht, H.; Gabor, J.; Stello, T.; Shiba, K.; Musier-Forsyth, K.; Kleiman, L. Incorporation of Lysyl-tRNA Synthetase into Human Immunodeficiency Virus Type 1. *J. Virol.* **2001**, *75*, 5043–5048, doi:10.1128/jvi.75.11.5043-5048.2001.

146. Javanbakht, H.; Cen, S.; Musier-Forsyth, K.; Kleiman, L. Correlation between tRNALys3 aminoacylation and its incorporation into HIV-1. *J. Biol. Chem.* **2002**, *277*, 17389–17396, doi:10.1074/jbc.M112479200.

147. Guo, F.; Cen, S.; Niu, M.; Javanbakht, H.; Kleiman, L. Specific Inhibition of the Synthesis of Human Lysyl-tRNA Synthetase Results in Decreases in tRNALys Incorporation, tRNA3LysAnnealing to Viral RNA, and Viral Infectivity in Human Immunodeficiency Virus Type 1. *J. Virol.* **2003**, *77*, 9817–9822, doi:10.1128/jvi.77.18.9817-9822.2003.

148. Cen, S.; Javanbakht, H.; Niu, M.; Kleiman, L. Ability of Wild-Type and Mutant Lysyl-tRNA Synthetase To Facilitate tRNALys Incorporation into Human Immunodeficiency Virus Type 1. *J. Virol.* **2004**, *78*, 1595–1601, doi:10.1128/jvi.78.3.1595-1601.2004.

149. Bolinger, C.; Sharma, A.; Singh, D.; Yu, L.; Boris-Lawrie, K. RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res.* **2010**, *38*, 1686–1696, doi:10.1093/nar/gkp1075.

150. Xing, L.; Niu, M.; Kleiman, L. In Vitro and In Vivo Analysis of the Interaction between RNA Helicase A and HIV-1 RNA. *J. Virol.* **2012**, *86*, 13272–13280, doi:10.1128/jvi.01993-12.

151. Boeras, I.; Song, Z.; Moran, A.; Franklin, J.; Brown, W.C.; Johnson, M.; Boris-Lawrie, K.; Heng, X. DHX9/RHA Binding to the PBS-Segment of the Genomic RNA during HIV-1 Assembly Bolsters Virion Infectivity. *J. Mol. Biol.* **2016**, *428*, 2418–2429, doi:10.1016/j.jmb.2016.04.011.

152. Damgaard, C.K.; Andersen, E.S.; Knudsen, B.; Gorodkin, J.; Kjems, J. RNA Interactions in the 5′ Region of the HIV-1 Genome. *J. Mol. Biol.* **2004**, *336*, 369–379,

doi:10.1016/j.jmb.2003.12.010.

153. Abbink, T.E.M.; Berkhout, B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 rna occludes the gag start codon. *J. Biol. Chem.* **2003**, *278*, 11601–11611, doi:10.1074/jbc.M210291200.

154. Wilkinson, K.A.; Gorelick, R.J.; Vasa, S.M.; Guex, N.; Rein, A.; Mathews, D.H.; Giddings, M.C.; Weeks, K.M. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* **2008**, *6*, 883–899, doi:10.1371/journal.pbio.0060096.

155. Jones, C.P.; Cantara, W.A.; Olson, E.D.; Musier-Forsyth, K. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3395–3400, doi:10.1073/pnas.1319658111.

156. Kenyon, J.C.; Prestwood, L.J.; Le Grice, S.F.J.; Lever, A.M.L. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.* **2013**, *41*, doi:10.1093/nar/gkt690.

157. Keane, S.C.; Heng, X.; Lu, K.; Kharytonchyk, S.; Ramakrishnan, V.; Carter, G.; Barton, S.; Hosic, A.; Florwick, A.; Santos, J.; et al. Structure of the HIV-1 RNA packaging signal. *Science (80-.).* **2015**, *348*, 917–921, doi:10.1126/science.aaa9266.

158. Barton, S.; Heng, X.; Johnson, B.A.; Summers, M.F. Database proton NMR chemical shifts for RNA signal assignment and validation. *J. Biomol. NMR* **2013**, *55*, 33–46, doi:10.1007/s10858-012-9683-9.

159. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; Mcgettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948, doi:10.1093/bioinformatics/btm404.

160. Gremminger, T.; Song, Z.; Ji, J.; Foster, A.; Weng, K.; Heng, X. Extended Interactions

between HIV-1 Viral RNA and tRNALys3 Are Important to Maintain Viral RNA Integrity. *Int. J. Mol. Sci.* **2020**, *22*, 58, doi:10.3390/ijms22010058.

161. Xu, X.; Zhao, P.; Chen, S.J. Vfold: A web server for RNA structure and folding thermodynamics prediction. *PLoS One* **2014**, *9*, doi:10.1371/journal.pone.0107504.

162. Lange, M.J.; Burke, D.H. Screening inhibitory potential of anti-HIV RT RNA aptamers. *Methods Mol. Biol.* **2014**, *1103*, 11–29, doi:10.1007/978-1-62703-730-3_2.

163. Huang, Y.; Wang, J.; Shalom, A.; Li, Z.; Khorchid, A.; Wainberg, M.A.; Kleiman, L. Primer tRNA3Lys on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. *J. Virol.* **1997**, *71*, 726–728, doi:10.1128/jvi.71.1.726-728.1997.

164. Xu, X.; Qiu, L.; Yan, C.; Ma, Z.; Grinter, S.Z.; Zou, X. Performance of MDockPP in CAPRI rounds 28-29 and 31-35 including the prediction of water-mediated interactions. *Proteins Struct. Funct. Bioinforma.* **2017**, *85*, 424–434, doi:10.1002/prot.25203.

165. Turnbull, W.B.; Daranas, A.H. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J. Am. Chem. Soc.* **2003**, *125*, 14859–14866, doi:10.1021/ja036166s.

166. Fu, Q.; Yuan, Y.A. Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). *Nucleic Acids Res.* **2013**, *41*, 3457–3470, doi:10.1093/nar/gkt042.

167. Ranji, A.; Shkriabai, N.; Kvaratskhelia, M.; Musier-Forsyth, K.; Boris-Lawrie, K. Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J. Biol. Chem.* **2011**, *286*, 5328–5337, doi:10.1074/jbc.M110.176339.

168. Prabu, J.R.; Müller, M.; Thomae, A.W.; Schüssler, S.; Bonneau, F.; Becker, P.B.; Conti, E.

Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. *Mol. Cell* **2015**, *60*, 487–499, doi:10.1016/j.molcel.2015.10.011.

169. Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E.L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A.; et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science (80-.).* **2011**, *334*, 242–245, doi:10.1126/science.1210460.

170. Feng, Y.-X.; Campbell, S.; Harvin, D.; Ehresmann, B.; Ehresmann, C.; Rein, A. 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*, 4251–4256, doi:10.1128/jvi.73.5.4251-4256.1999.

171. Roldan, A.; Warren, O.U.; Russell, R.S.; Liang, C.; Wainberg, M.A. A HIV-1 minimal gag protein is superior to nucleocapsid at in vitro tRNA3Lys annealing and exhibits multimerization-induced inhibition of reverse transcription. *J. Biol. Chem.* **2005**, *280*, 17488–17496, doi:10.1074/jbc.M501310200.

172. Jones, C.P.; Datta, S.A.K.; Rein, A.; Rouzina, I.; Musier-Forsyth, K. Matrix Domain Modulates HIV-1 Gag's Nucleic Acid Chaperone Activity via Inositol Phosphate Binding. *J. Virol.* **2011**, *85*, 1594–1603, doi:10.1128/jvi.01809-10.

173. Fujita, H.; Ohshima, T.; Oishi, T.; Aratani, S.; Fujii, R.; Fukamizu, A.; Nakajima, T. Relevance of nuclear localization and functions of RNA helicase A. *Int. J. Mol. Med.* **2005**, *15*, 555–560.

174. Lv, M.; Yao, Y.; Li, F.; Xu, L.; Yang, L.; Gong, Q.; Xu, Y.Z.; Shi, Y.; Fan, Y.J.; Tang, Y. Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in

Drosophila. *Nucleic Acids Res.* **2019**, *47*, 3142–3157, doi:10.1093/nar/gky1308.

175. Ankush Jagtap, P.K.; Müller, M.; Masiewicz, P.; Von Bülow, S.; Merret Hollmann, N.; Chen, P.C.; Simon, B.; Thomae, A.W.; Becker, P.B.; Hennig, J. Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1,2 of Drosophila helicase Maleless. *Nucleic Acids Res.* **2019**, *47*, 4319–4333, doi:10.1093/nar/gkz125.

176. Masliah, G.; Barraud, P.; Allain, F.H.T. RNA recognition by double-stranded RNA binding domains: A matter of shape and sequence. *Cell. Mol. Life Sci.* 2013, *70*, 1875–1895.

177. Jayachandran, U.; Grey, H.; Cook, A.G. Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. *Nucleic Acids Res.* **2015**, *44*, 1924–1936, doi:10.1093/nar/gkv1508.

178. Masliah, G.; Maris, C.; König, S.L.; Yulikov, M.; Aeschimann, F.; Malinowska, A.L.; Mabille, J.; Weiler, J.; Holla, A.; Hunziker, J.; et al. Structural basis of si RNA recognition by TRBP double-stranded RNA binding domains . *EMBO J.* **2018**, *37*, doi:10.15252/embj.201797089.

179. Partin, A.C.; Zhang, K.; Jeong, B.C.; Herrell, E.; Li, S.; Chiu, W.; Nam, Y. Cryo-EM Structures of Human Drosha and DGCR8 in Complex with Primary MicroRNA. *Mol. Cell* **2020**, *78*, 411-422.e4, doi:10.1016/j.molcel.2020.02.016.

180. Lazzaretti, D.; Bandholz-Cajamarca, L.; Emmerich, C.; Schaaf, K.; Basquin, C.; Irion, U.; Bono, F. The crystal structure of Staufen1 in complex with a physiological RNA sheds light on substrate selectivity. *Life Sci. Alliance* **2018**, *1*, doi:10.26508/lsa.201800187.

181. Bou-Nader, C.; Barraud, P.; Pecqueur, L.; Pérez, J.; Velours, C.; Shepard, W.; Fontecave, M.; Tisné, C.; Hamdane, D. Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Res.*

2019, *47*, 3117–3126, doi:10.1093/nar/gky1302.

182. Wu, H.; Henras, A.; Chanfreau, G.; Feigon, J. Structural basis for of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8307–8312, doi:10.1073/pnas.0402627101.

183. Stefl, R.; Oberstrass, F.C.; Hood, J.L.; Jourdan, M.; Zimmermann, M.; Skrisovska, L.; Maris, C.; Peng, L.; Hofr, C.; Emeson, R.B.; et al. The Solution Structure of the ADAR2 dsRBM-RNA Complex Reveals a Sequence-Specific Readout of the Minor Groove. *Cell* **2010**, *143*, 225–237, doi:10.1016/j.cell.2010.09.026.

184. Fierro-Monti, I.; Mathews, M.B. Proteins binding to duplexed RNA: One motif, multiple functions. *Trends Biochem. Sci.* 2000, *25*, 241–246.

185. Milligan, J.F.; Uhlenbeck, O.C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **1989**, *180*, 51–62, doi:10.1016/0076-6879(89)80091-6.

186. Rinaldi, A.J.; Suddala, K.C.; Walter, N.G. Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.* **2015**, *1240*, 63–95, doi:10.1007/978-1-4939-1896-6_6.

187. Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277–293, doi:10.1007/BF00197809.

188. Johnson, B.A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **2004**, *278*, 313–352, doi:10.1385/1-59259-809-9:313.

189. Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H.D.T.; Kikhney, A.G.; Hajizadeh, N.R.; Franklin, J.M.; Jeffries, C.M.; et al. ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular

solutions. *J. Appl. Crystallogr.* **2017**, *50*, 1212–1225, doi:10.1107/S1600576717007786.

190. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503, doi:10.1107/S0021889892001663.

191. Franke, D.; Svergun, D.I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009**, *42*, 342–346, doi:10.1107/S0021889809000338.

192. Güntert, P.; Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR* **2015**, *62*, 453–471, doi:10.1007/s10858-015-9924-9.

193. Tolbert, B.S.; Miyazaki, Y.; Barton, S.; Kinde, B.; Starck, P.; Singh, R.; Bax, A.; Case, D.A.; Summers, M.F. Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J. Biomol. NMR* **2010**, *47*, 205–219, doi:10.1007/s10858-010-9424-x.

194. Svergun, D.; Barberato, C.; Koch, M.H. CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773, doi:10.1107/S0021889895007047.

195. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005, *26*, 1668–1688.

196. Schwieters, C.D.; Kuszewski, J.J.; Tjandra, N.; Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**, *160*, 65–73, doi:10.1016/S1090-7807(02)00014-9.

197. Cao, S.; Giedroc, D.P.; Chen, S.J. Predicting loop-helix tertiary structural contacts in RNA pseudoknots. *RNA* **2010**, *16*, 538–552, doi:10.1261/rna.1800210.

198. Cao, S.; Chen, S.J. Structure and stability of RNA/RNA kissing complex: With application to HIV dimerization initiation signal. *RNA* **2011**, *17*, 2130–2143, doi:10.1261/rna.026658.111.

199. Cao, S.; Chen, S.J. Predicting kissing interactions in microRNA-target complex and assessment of microRNA activity. *Nucleic Acids Res.* **2012**, *40*, 4681–4690, doi:10.1093/nar/gks052.

200. Huang, S.Y.; Zou, X. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* **2014**, *42*, e55, doi:10.1093/nar/gku077.

201. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612, doi:10.1002/jcc.20084.

202. Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* **2016**, *2016*, 5.6.1-5.6.37, doi:10.1002/cpbi.3.

203. Hubbard, S.; Thornton, J. NACCESS 1993.

204. Benner, S.A.; Ellington, A.D.; Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 7054–7058, doi:10.1073/pnas.86.18.7054.

205. White, H.B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976**, *7*, 101–104, doi:10.1007/BF01732468.

206. Cech, T.R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, 1–5, doi:10.1101/cshperspect.a006742.

207. Massey, V. A Simple Method for the Determination of Redox Potentials. In Proceedings of the Flavins and Flavoproteins 1990; 1990; pp. 59–66.

208. Guo, F.; Gooding, A.R.; Cech, T.R. Structure of the tetrahymena ribozyme: Base triple

sandwich and metal ion at the active site. *Mol. Cell* **2004**, *16*, 351–362, doi:10.1016/j.molcel.2004.10.003.

209. Burke, D.H.; Scates, L.; Andrews, K.; Gold, L. Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.* **1996**, *264*, 650–666, doi:10.1006/jmbi.1996.0667.

210. Chaloin, L.; Lehmann, M.J.; Sczakiel, G.; Restle, T. Endogenous expression of a high-affinity pseudoknot RNA aptamer suppresses replication of HIV-1. *Nucleic Acids Res.* **2002**, *30*, 4001–4008, doi:10.1093/nar/gkf522.

211. Fisher, T.S.; Joshi, P.; Prasad, V.R. Mutations That Confer Resistance to Template-Analog Inhibitors of Human Immunodeficiency Virus (HIV) Type 1 Reverse Transcriptase Lead to Severe Defects in HIV Replication. *J. Virol.* **2002**, *76*, 4068–4072, doi:10.1128/jvi.76.8.4068-4072.2002.

212. Held, D.M.; Kissel, J.D.; Saran, D.; Michalowski, D.; Burke, D.H. Differential susceptibility of HIV-1 reverse transcriptase to inhibition by RNA aptamers in enzymatic reactions monitoring specific steps during genome replication. *J. Biol. Chem.* **2006**, *281*, 25712–25722, doi:10.1074/jbc.M604460200.

213. Whatley, A.S.; Ditzler, M.A.; Lange, M.J.; Biondi, E.; Sawyer, A.W.; Chang, J.L.; Franken, J.D.; Burke, D.H. Potent inhibition of HIV-1 reverse transcriptase and replication by nonpseudoknot, “UCAA-motif” RNA aptamers. *Mol. Ther. - Nucleic Acids* **2013**, *2*, e71, doi:10.1038/mtna.2012.62.

214. Lange, M.J.; Nguyen, P.D.M.; Callaway, M.K.; Johnson, M.C.; Burke, D.H. RNA-protein interactions govern antiviral specificity and encapsidation of broad spectrum anti-HIV reverse transcriptase aptamers. *Nucleic Acids Res.* **2017**, *45*, 6087–6097, doi:10.1093/nar/gkx155.

215. Tuerk, C.; Macdougal, S.; Gold, L. RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 6988–6992, doi:10.1073/pnas.89.15.6988.

216. Ditzler, M.A.; Bose, D.; Shkriabai, N.; Marchand, B.; Sarafianos, S.G.; Kvaratskhelia, M.; Burke, D.H. Broad-spectrum aptamer inhibitors of HIV reverse transcriptase closely mimic natural substrates. *Nucleic Acids Res.* **2011**, *39*, 8237–8247, doi:10.1093/nar/gkr381.

217. Kensch, O.; Connolly, B.A.; Steinhoff, H.J.; McGregor, A.; Goody, R.S.; Restle, T. Hiv-1 reverse transcriptase-pseudoknot RNA aptamer interaction has a binding affinity in the low picomolar range coupled with high specificity. *J. Biol. Chem.* **2000**, *275*, 18271–18278, doi:10.1074/jbc.M001309200.

218. Kissel, J.D.; Held, D.M.; Hardy, R.W.; Burke, D.H. Single-stranded DNA aptamer RT1t49 inhibits RT polymerase and RNase H functions of HIV type 1, HIV type 2, and SIVCPZ RTs. *AIDS Res. Hum. Retroviruses* **2007**, *23*, 699–708, doi:10.1089/aid.2006.0262.

219. Kissel, J.D.; Held, D.M.; Hardy, R.W.; Burke, D.H. Active site binding and sequence requirements for inhibition of HIV-1 reverse transcriptase by the RT1 family of single-stranded DNA aptamers. *Nucleic Acids Res.* **2007**, *35*, 5039–5050, doi:10.1093/nar/gkm420.

220. Miller, M.T.; Tuske, S.; Das, K.; DeStefano, J.J.; Arnold, E. Structure of HIV-1 reverse transcriptase bound to a novel 38-mer hairpin template-primer DNA aptamer. *Protein Sci.* **2016**, *25*, 46–55, doi:10.1002/pro.2776.

221. Green, L.; Waugh, S.; Binkley, J.P.; Hostomska, Z.; Hostomsky, Z.; Tuerk, C. Comprehensive Chemical Modification Interference and Nucleotide Substitution Analysis of an RNA Pseudoknot Inhibitor to HIV-1 Reverse Transcrip tase. *J. Mol. Biol.* **1995**, *247*,

60–68, doi:10.1006/jmbi.1994.0122.

222. Jaeger, J.; Restle, T.; Steitz, T.A. The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.* **1998**, *17*, 4535–4542, doi:10.1093/emboj/17.15.4535.

223. Aeksiri, N.; Songtawee, N.; Gleeson, M.P.; Hannongbua, S.; Choowongkomon, K. Insight into HIV-1 reverse transcriptase - Aptamer interaction from molecular dynamics simulations. *J. Mol. Model.* **2014**, *20*, doi:10.1007/s00894-014-2380-8.

224. Held, D.M.; Kissel, J.D.; Thacker, S.J.; Michalowski, D.; Saran, D.; Ji, J.; Hardy, R.W.; Rossi, J.J.; Burke, D.H. Cross-Clade Inhibition of Recombinant Human Immunodeficiency Virus Type 1 (HIV-1), HIV-2, and Simian Immunodeficiency Virus SIVcpz Reverse Transcriptases by RNA Pseudoknot Aptamers. *J. Virol.* **2007**, *81*, 5375–5384, doi:10.1128/jvi.01923-06.

225. Ditzler, M.A.; Lange, M.J.; Bose, D.; Bottoms, C.A.; Virkler, K.F.; Sawyer, A.W.; Whatley, A.S.; Spollen, W.; Givan, S.A.; Burke, D.H. High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.* **2013**, *41*, 1873–1884, doi:10.1093/nar/gks1190.

226. Götte, M.; Maier, G.; Gross, H.J.; Heumann, H. Localization of the active site of HIV-1 reverse transcriptase- associated RNase H domain on a DNA template using site-specific generated hydroxyl radicals. *J. Biol. Chem.* **1998**, *273*, 10139–10146, doi:10.1074/jbc.273.17.10139.

227. Tullius, T.D.; Greenbaum, J.A. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* 2005, *9*, 127–134.

228. Jain, S.S.; Tullius, T.D. Footprinting protein - DNA complexes using the hydroxyl radical. *Nat. Protoc.* **2008**, *3*, 1092–1100, doi:10.1038/nprot.2008.72.

229. Chalmers, M.J.; Busby, S.A.; Pascal, B.D.; He, Y.; Hendrickson, C.L.; Marshall, A.G.; Griffin,

P.R. Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2006**, *78*, 1005–1014, doi:10.1021/ac051294f.

230. Goswami, D.; Tuske, S.; Pascal, B.D.; Bauman, J.D.; Patel, D.; Arnold, E.; Griffin, P.R. Differential Isotopic Enrichment To Facilitate Characterization of Asymmetric Multimeric Proteins Using Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **2015**, *87*, 4015–4022, doi:10.1021/acs.analchem.5b00372.

231. Hsiou, Y.; Ding, J.; Das, K.; Clark, A.D.; Hughes, S.H.; Arnold, E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: Implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **1996**, *4*, 853–860, doi:10.1016/S0969-2126(96)00091-3.

232. Kensch, O.; Restle, T.; Wöhrl, B.M.; Goody, R.S.; Steinhoff, H.J. Temperature-dependent equilibrium between the open and closed conformation of the p66 subunit of HIV-1 reverse transcriptase revealed by site-directed spin labelling. *J. Mol. Biol.* **2000**, *301*, 1029–1039, doi:10.1006/jmbi.2000.3998.

233. Rodgers, D.W.; Gamblin, S.J.; Harris, B.A.; Ray, S.; Culp, J.S.; Hellmig, B.; Woolf, D.J.; Debouck, C.; Harrison, S.C. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 1222–1226, doi:10.1073/pnas.92.4.1222.

234. Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 303–308, doi:10.1038/nsb0495-303.

235. Santos, A.F.A.; Lengruber, R.B.; Soares, E.A.; Jere, A.; Sprinz, E.; Martinez, A.M.B.; Silveira, J.; Sion, F.S.; Pathak, V.K.; Soares, M.A. Conservation patterns of HIV-1 RT connection and RNase H domains: Identification of new mutations in NRTI-treated patients. *PLoS One*

2008, *3*, doi:10.1371/journal.pone.0001781.

236. Cao, S.; Chen, S.J. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* **2005**, *11*, 1884–1897, doi:10.1261/rna.2109105.

237. Cao, S.; Chen, S.J. Physics-based de novo prediction of RNA 3D structures. *J. Phys. Chem. B* **2011**, *115*, 4216–4226, doi:10.1021/jp112059y.

238. Xu, X.; Chen, S.J. A method to predict the 3D structure of an RNA scaffold. *Methods Mol. Biol.* **2015**, *1316*, 1–11, doi:10.1007/978-1-4939-2730-2_1.

239. Zhang, D.; Chen, S.J. IsRNA: An Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *J. Chem. Theory Comput.* **2018**, *14*, 2230–2239, doi:10.1021/acs.jctc.7b01228.

240. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 2005, *26*, 1781–1802.

241. Brooks, B.R.; Brooks, C.L.; Mackerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614, doi:10.1002/jcc.21287.

242. Chattopadhyay, D.; Evans, D.B.; Deibel, M.R.; Vosters, A.F.; Eckenrode, F.M.; Einspahr, H.M.; Hui, J.O.; Tomasselli, A.G.; Zurcher-Neely, H.A.; Heinrikson, R.L.; et al. Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by in vitro processing of p66 with recombinant HIV-1 protease. *J. Biol. Chem.* **1992**, *267*, 14227–14232, doi:10.1016/s0021-9258(19)49701-9.

243. Hostomska, Z.; Matthews, D.A.; Davies, J.F.; Nodes, B.R.; Hostomsky, Z. Proteolytic release and crystallization of the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem.* **1991**, *266*, 14697–14702, doi:10.1016/s0021-

9258(18)98742-9.

244. Tomasselli, A.G.; Sarcich, J.L.; Barrett, L.J.; Reardon, I.M.; Howe, W.J.; Evans, D.B.; Sharma, S.K.; Heinrikson, R.L. Human immunodeficiency virus type-1 reverse transcriptase and ribonuclease h as substrates of the viral protease. *Protein Sci.* **1993**, *2*, 2167–2176, doi:10.1002/pro.5560021216.

245. Abram, M.E.; Parniak, M.A. Virion Instability of Human Immunodeficiency Virus Type 1 Reverse Transcriptase (RT) Mutated in the Protease Cleavage Site between RT p51 and the RT RNase H Domain. *J. Virol.* **2005**, *79*, 11952–11961, doi:10.1128/jvi.79.18.11952-11961.2005.

246. Zheng, X.; Pedersen, L.C.; Gabel, S.A.; Mueller, G.A.; Cuneo, M.J.; DeRose, E.F.; Krahn, J.M.; London, R.E. Selective unfolding of one Ribonuclease H domain of HIV reverse transcriptase is linked to homodimer formation. *Nucleic Acids Res.* **2014**, *42*, 5361–5377, doi:10.1093/nar/gku143.

247. London, R.E. Structural maturation of HIV-1 reverse transcriptase—a metamorphic solution to genomic instability. *Viruses* 2016, *8*.

248. Ilina, T. V.; Slack, R.L.; Elder, J.H.; Sarafianos, S.G.; Parniak, M.A.; Ishima, R. Effect of tRNA on the Maturation of HIV-1 Reverse Transcriptase. *J. Mol. Biol.* **2018**, *430*, 1891–1900, doi:10.1016/j.jmb.2018.02.027.

249. Sluis-Cremer, N.; Arion, D.; Abram, M.E.; Parniak, M.A. Proteolytic processing of an HIV-1 pol polyprotein precursor: Insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1836–1847, doi:10.1016/j.biocel.2004.02.020.

250. Schneider, D.J.; Feigon, J.; Hostomsky, Z.; Gold, L. High-Affinity ssDNA Inhibitors of the Reverse Transcriptase of Type 1 Human Immunodeficiency Virus. *Biochemistry* **1995**, *34*,

9599–9610, doi:10.1021/bi00029a037.

251. Jacobo-Molina, A.; Ding, J.; Nanni, R.G.; Clark, A.D.; Lu, X.; Tantillo, C.; Williams, R.L.; Kamer, G.; Ferris, A.L.; Clark, P.; et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6320–6324, doi:10.1073/pnas.90.13.6320.

252. Chung, S.; Miller, J.T.; Lapkouski, M.; Tian, L.; Yang, W.; Le Grice, S.F.J. Examining the role of the HIV-1 reverse transcriptase p51 subunit in positioning and hydrolysis of RNA/DNA hybrids. *J. Biol. Chem.* **2013**, *288*, 16177–16184, doi:10.1074/jbc.M113.465641.

253. Kohlstaedt, L.A.; Wang, J.; Friedman, J.M.; Rice, P.A.; Steitz, T.A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science (80-.).* **1992**, *256*, 1783–1790, doi:10.1126/science.1377403.

254. Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; et al. High resolution structures of HIV-1 RT from four RT—inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302, doi:10.1038/nsb0495-293.

255. Das, K.; Martinez, S.E.; Bauman, J.D.; Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat. Struct. Mol. Biol.* **2012**, *19*, 253–259, doi:10.1038/nsmb.2223.

256. Sharaf, N.G.; Ishima, R.; Gronenborn, A.M. Conformational plasticity of the NNRTI-binding pocket in HIV-1 reverse transcriptase: A fluorine nuclear magnetic resonance study. *Biochemistry* **2016**, *55*, 3864–3873, doi:10.1021/acs.biochem.6b00113.

257. Sarafianos, S.G.; Marchand, B.; Das, K.; Himmel, D.M.; Parniak, M.A.; Hughes, S.H.; Arnold, E. Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. *J. Mol. Biol.* 2009, *385*, 693–713.

258. Seckler, J.M.; Barkley, M.D.; Wintrode, P.L. Allosteric suppression of HIV-1 reverse transcriptase structural dynamics upon inhibitor binding. *Biophys. J.* **2011**, *100*, 144–153, doi:10.1016/j.bpj.2010.11.004.

259. Hartig, J.S.; Najafi-Shoushtari, S.H.; Grüne, I.; Yan, A.; Ellington, A.D.; Famulok, M. Protein-dependent ribozymes report molecular interactions in real time. *Nat. Biotechnol.* **2002**, *20*, 717–722, doi:10.1038/nbt0702-717.

260. Huang, C.C.; Cao, Z.; Chang, H.T.; Tan, W. Protein-protein interaction studies based on molecular aptamers by affinity capillary electrophoresis. *Anal. Chem.* **2005**, *76*, 6973–6981, doi:10.1021/ac049158i.

261. Wu, X.; Liu, H.; Xiao, H.; Conway, J.A.; Hehl, E.; Kalpana, G. V.; Prasad, V.; Kappes, J.C. Human Immunodeficiency Virus Type 1 Integrase Protein Promotes Reverse Transcription through Specific Interactions with the Nucleoprotein Reverse Transcription Complex. *J. Virol.* **1999**, *73*, 2126–2135, doi:10.1128/jvi.73.3.2126-2135.1999.

262. Dobard, C.W.; Briones, M.S.; Chow, S.A. Molecular Mechanisms by Which Human Immunodeficiency Virus Type 1 Integrase Stimulates the Early Steps of Reverse Transcription. *J. Virol.* **2007**, *81*, 10037–10046, doi:10.1128/jvi.00519-07.

263. Tekeste, S.S.; Wilkinson, T.A.; Weiner, E.M.; Xu, X.; Miller, J.T.; Le Grice, S.F.J.; Clubb, R.T.; Chow, S.A. Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. *J. Virol.* **2015**, *89*, 12058–12069, doi:10.1128/jvi.01471-15.

264. Hehl, E.A.; Joshi, P.; Kalpana, G. V.; Prasad, V.R. Interaction between Human Immunodeficiency Virus Type 1 Reverse Transcriptase and Integrase Proteins. *J. Virol.* **2004**, *78*, 5056–5067, doi:10.1128/jvi.78.10.5056-5067.2004.

265. Drummond, J.E.; Mounts, P.; Gorelick, R.J.; Casas-Finet, J.R.; Bosche, W.J.; Henderson,

L.E.; Waters, D.J.; Arthur, L.O. Wild-type and mutant HIV type 1 nucleocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase. *AIDS Res. Hum. Retroviruses* **1997**, *13*, 533–543, doi:10.1089/aid.1997.13.533.

266. Cameron, C.E.; Ghosh, M.; Le Grice, S.F.J.; Benkovic, S.J. Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand transfer efficiency are suppressed by HIV nucleocapsid protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6700–6705, doi:10.1073/pnas.94.13.6700.

267. Peliska, J.A.; Balasubramanian, S.; Giedroc, D.P.; Benkovic, S.J. Recombinant HIV-1 Nucleocapsid Protein Accelerates HIV-1 Reverse Transcriptase Catalyzed DNA Strand Transfer Reactions and Modulates RNase H Activity. *Biochemistry* **1994**, *33*, 13817–13823, doi:10.1021/bi00250a036.

268. Nishitsuji, H.; Yokoyama, M.; Sato, H.; Yamauchi, S.; Takaku, H. Identification of amino acid residues in HIV-1 reverse transcriptase that are critical for the proteolytic processing of Gag-Pol precursors. *FEBS Lett.* **2011**, *585*, 3372–3377, doi:10.1016/j.febslet.2011.09.034.

269. Böttcher, M.; Grosse, F. HIV-1 protease inhibits its homologous reverse transcriptase by protein-protein interaction. *Nucleic Acids Res.* **1997**, *25*, 1709–1714, doi:10.1093/nar/25.9.1709.

270. Goobar-Larsson, L.; Luukkonen, M.B.G.; Unge, T.; Schwartz, S.; Utter, G.; Strandberg, B.; öberg, B. Enhancement of HIV-1 proteinase activity by HIV-1 reverse transcriptase. *Virology* **1995**, *206*, 387–394, doi:10.1016/S0042-6822(95)80054-9.

271. Wapling, J.; Moore, K.L.; Sonza, S.; Mak, J.; Tachedjian, G. Mutations That Abrogate Human Immunodeficiency Virus Type 1 Reverse Transcriptase Dimerization Affect Maturation of the Reverse Transcriptase Heterodimer. *J. Virol.* **2005**, *79*, 10247–10257,

doi:10.1128/jvi.79.16.10247-10257.2005.

272. Figueiredo, A.; Moore, K.L.; Mak, J.; Sluis-Cremer, N.; De Bethune, M.P.; Tachedjian, G. Potent nonnucleoside reverse transcriptase inhibitors target HIV-1 Gag-Pol. *PLoS Pathog.* **2006**, *2*, 1051–1059, doi:10.1371/journal.ppat.0020119.

273. Jochmans, D.; Anders, M.; Keuleers, I.; Smeulders, L.; Kräusslich, H.G.; Kraus, G.; Müller, B. Selective killing of human immunodeficiency virus infected cells by non-nucleoside reverse transcriptase inhibitor-induced activation of HIV protease. *Retrovirology* **2010**, *7*, doi:10.1186/1742-4690-7-89.

274. Metzger, W.; Schickor, P.; Heumann, H. A cinematographic view of Escherichia coli RNA polymerase translocation. *EMBO J.* **1989**, *8*, 2745–2754, doi:10.1002/j.1460-2075.1989.tb08416.x.

275. Pascal, B.D.; Willis, S.; Lauer, J.L.; Landgraf, R.R.; West, G.M.; Marciano, D.; Novick, S.; Goswami, D.; Chalmers, M.J.; Griffin, P.R. HDXWorkbench: Software for the analysis of H/D exchange MS data. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1512–1521, doi:10.1007/s13361-012-0419-6.

276. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.A.; Aflalo, C.; Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2195–2199, doi:10.1073/pnas.89.6.2195.

277. Chen, R.; Li, L.; Weng, Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins Struct. Funct. Genet.* **2003**, *52*, 80–87, doi:10.1002/prot.10389.

Vita

Thomas James Gremminger was born to Daniel and Patricia Gremminger. He has always been interested in science. During high school Thomas became interested in chemistry. After graduating from Windsor High School, he attended Missouri University of Science and Technology. He majored in chemistry and discovered his interest in biochemistry while taken biochemistry courses. Thomas graduated from Missouri University of Science and Technology and decided to pursue a doctoral degree at the University of Missouri.

During graduate school Thomas was awarded a *University of Missouri Life Sciences Fellowship* and an *Excellence in Electron Microscopy Fellowship 2017* (awarded to collect preliminary data for electron microscopy projects). Thomas was also granted a University of Missouri NMR Core Mini Grant to collect NMR data for RNA structural studies. On April 23, 2021 he successfully defended his Ph.D. in Biochemistry under the supervision of Dr. Xiao Heng.