HIV-1 5'UTR RNA STRUCTURES: CONNECTIONS TO INFECTIOUS LIFECYCLE AND POTENTIAL THERAPEUTIC APPLICATIONS

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

Seth A. Staller

Dr. Xiao Heng, Dissertation Supervisor

July 2021

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

Have examined the dissertation entitled

HIV-1 5'UTR RNA STRUCTURES: CONNECTIONS TO INFECTIOUS LIFECYCLE AND POTENTIAL THERAPEUTIC APPLICATIONS

Presented by Seth A. Staller

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 Peter Tipton

Professor Shi-jie Chen

Professor Kent Gates

For my father, Steve

~~Question everything~~. Ask great questions.

In sterquiliniis invenitur.

-INSERT BLANK-

Acknowledgements

I want to first and foremost thank my wonderful wife, Caroline, for supporting me and helping me through the graduate program.

I want to thank my mentor, Dr. Xiao Heng, for taking me under her wing and agreeing to mentor me and dealing with all the struggles that came along the way. She is an excellent scientist and person of outstanding integrity. Her passion for science and '*data!'* is infectious and uplifting. I would not have gotten this far without her help and compassion.

I also want to thank the senior members of my lab, Dr. Zhenwei Song and Juan Ji. Zhenwei always has valuable insights for executing experiments of great quality, and we have several great conversations about the world. Juan has been a source of positivity and warmth, providing boosts of encouragement. I would also like to thank Dr. Samantha Brady and Dr. Thomas Gremminger, two former graduate students of the lab, for all their help as well.

I want to thank Dr. Peter Tipton, who has been something of a mentor to me. He has helped me and shown compassion to me in my darkest moments. He has supported me. He has gone to bat for me. He has talked to me when I needed someone to talk to. There is not a timeline where I had the opportunity to finish my graduate studies without him.

Thank you to my remaining committee members, Dr. Shi-jie Chen and Dr. Kent Gates, for staying with me through peculiar circumstances, and their extenuating support.

I would briefly like to thank Dr. Steve Van Doren, for his compassion, mercy, and aid.

I want to thank all the members of my bible study group, including Gene and Nancy, John and Luella, and John and Sherry, for their love.

I want to thank the University of Missouri Biochemisty Department for funding support while I finished my graduate studies. Thank you to both Dr. Gerald Hazelbauer and Dr. Michael Chapman for not tossing me away and giving me a chance.

Thank you to all the haters and saboteurs, who provided more obstacles for me to overcome, making me stronger.

Finally, I would like to thank my cats: Mister Pickles, my furry familiar, and Captain Pancakes, Avatar of Gluttony. They are great company… except when they are hungry.

Table of Contents

List of Figures

List of Abbreviations

Ψ	Psi packaging signal
(-)ssDNA	Strong-stop DNA
3’LTR	3’ long terminal repeat
5’LTR	5’ long terminal repeat
3’UTR	3’ untranslated region
5’UTR	5’ untranslated region
5NIA	5-nitroisatoic acid
BMH	Branched multiple hairpin
CA	Capsid
CBP80	Cap binding protein 80
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4
cDNA	Complementary deoxyribonucleic acid
CryoEM	Cryogenic electron microscopy
CTD	Carboxy-terminus domain
CXCR4	C-X-C chemokine receptor type 4
DIS	Dimer initiation site
DNA	Deoxyribonucleic acid
DMA	Dimethylamiloride
dsDNA	Double stranded deoxyribonucleic acid
eIF4E	Eukaryotic translation initiation factor 4E

eIF4G	Eukaryotic translation initiation factor 4G
Env	Envelope
G	Guanosine
Gag	Group specific antigen
gRNA	Genomic RNA
HIV	Human immunodeficiency virus
INI1	Integrase interactor 1
IN	Integrase
IP	Immunoprecipitation
IRES	Internal ribosome entry site
LDI	Long distance interaction
lrAID	Long-range adenosine interaction detection
LTR	Long terminal repeat
m7G	7-methylguanosine
MA	Matrix
mRNP	Messenger ribonucleoprotein complex
mTOR	Mechanistic target of torin
MS	Multiply spliced
NC	Nucleocapsid
NCBP3	Nuclear cap binding protein 3
NMR	Nuclear magnetic resonance
NPC	Nuclear pore complex
NTP	Nucleotide triphosphate

ORF	Original reading frame
PBS	Primer binding site
PCE	Post-transcriptional elements
PIC	Pre-integration complex
PolyA	Polyadenylation signal
Pol	Polymerase
PR	Protease
R	Repeat region
RHA	RNA helicase A
RNP	Ribonucleoprotein
RRE	Rev-response element
RT	Reverse transcriptase
RNA	Ribonucleic acid
SD	Splice donor
SHAPE	Selective 2′-hydroxyl acylation analyzed by primer extension
SS	Singly spliced
ssRNA	Single stranded ribonucleic acid
TAR	Trans-activation response
Tat	Trans-activator of transcription
TLE	tRNA like element
TmG	Trimethylated guanosine
tRNA	Transfer ribonucleic acid
U3	Unique 3

U5	Unique 5
US	Unspliced
Vif	Viral infectivity factor
Vpu	Viral protein U
Vpr	Viral protein R

Abstract

Human immunodeficiency virus (HIV) is among the most sinister pathogens plaguing man. As of now, we do not yet have a method to “cure” infection once the virus has integrated into the host genome. It has the highest mutation rate of any known biological entity, necessitating a persistent need of novel therapeutic options. Structure: function relationships derived from biochemical approaches have been long-standing method to understand the mechanisms of HIV infection and identifying targets for treatment. This dissertation presents work utilizing a wide range of biochemical approaches to explore the importance of the HIV-1 5’untranslated region (5’UTR) RNA in infection and strategies utilizing this gained knowledge for screening of novel therapeutic agents. The HIV-1 5’UTR structure is highly structured, structurally dynamic, and the most conserved region of the genome. In this dissertation we reveal work demonstrating the existence of a novel, alternate translation pathway adopted by HIV-1 to promote virus fitness to the detriment of the host. We show evidence implicating the 5’UTR structure in determining fate of RNAs between translation pathways. Additionally, we demonstrate that folding of the upstream elements of the HIV-1 5’UTR influence folding of downstream elements, suggesting the fate of the mRNA transcripts are determined when they are nascently transcribed. This work sets the foundation for further investigation into mechanism of 5’ UTR folding in early and late HIV-1 mRNAs and its relevance in the HIV-1 infectious lifecycle. In this dissertation, we map the binding interface of HIV-1 5’UTR TAR RNA with INI1, a core component of the SWI/SNF eukaryotic

chromatin remodeling complex. This is the first example of HIV-1 RNA surface structure mimicking a host protein surface, and the modeled structure will be vital for development of novel therapeutics targeting the interaction between INI1 and TAR. We also show evidence that defective INI1 results in decreased transcriptional elongation and decreased virion particle production. We then outline future directions to monitor HIV-1 5'UTR folding for screening of novel therapeutics.

CHAPTER 1. Background on Human Immunodeficiency Virus, Infectious Cycle, and HIV-1 5'UTR RNA Structure

1.1 Overview

This dissertation examines the role of human immunodeficiency virus 1 (HIV-1) 5' untranslated region (5'UTR) RNA structures in the infectious lifecycle and potential strategies to exploit these structures for therapeutic value. Chapter 2 centers on the how these 5'UTR RNA structures may be implicated in transcription and translation. Chapter 3 characterizes an interaction between the trans-activation response element (TAR), a critical structural element leading the 5' UTR RNA, with human INI1, a vital component in the eukaryotic chromatin remodeling complex machinery, via molecular mimicry that has evolved into HIV-1. Chapter 4 focuses on how the HIV-1 5'UTR structure elements can be used to identify structural motifs and markers for development of assays for novel therapeutics.

1.2 HIV-1 Lifecycle

HIV-1 is an enveloped, single-stranded, positive-sense RNA virus of genus lentivirus in the family Retroviridae. The HIV-1 virion's envelope is constituted of a host-derived lipid bilayer and contains three viral proteins: two surface-displaying glycoproteins, gp120 and gp41, and a structural support lattice that connects the bilayer to the dimeric genome-containing capsid core, matrix protein (MA). HIV-1 targets immune cells, including T-helper cells, monocytes, macrophages, and

dendritic cells by recognition of surface-layer, immunoglobulin-like glycoprotein CD4. This leads to a cascade of events, ultimately causing fusion of the cell and virion surface membranes. The lifecycle of HIV-infected cells is generally divided into two stages: early and late-stage replication. The switch from early to late-stage infection is generally marked by integration of the proviral DNA into the host genome, preceding the transcription of viral RNAs and translation of viral protein. Figure 1-1 illustrates a generalized schema of the HIV-1 lifecycle as we understand it (1).

Initialization of a HIV-1 virion's infection to a cell is marked by recognition of surface-layer immunoglobulin-like glycoprotein CD4 (cluster of differentiation 4) of mentioned cells via its gp120 envelope protein (2–4). Conformational change of gp120 arising from the interaction with CD4 triggers interaction with an additional coreceptor. Two of these receptors have been identified so far: chemokine receptors CCR5 and CXCR4 (5–7). These coreceptors belong to the G-protein-coupled receptor superfamily (8). This secondary co-receptor interaction leads to a cascade of conformational changes in gp120 and additional envelope protein gp41, resulting in a dimer of heterotrimers (three gp120 and three gp41) -the envelope "spike protein" complex, spear-headed by the N-terminus of the gp41, that facilitate membrane fusion between the cell and virion (9, 10).

Cell: virus membrane fusion results in release of the capsid core to within the cytoplasm of the infected cell. The capsid shell is described as cone-shaped, following a fullerene-like geometric hexagonal lattice of viral capsid protein (CA) (11). Housed within the CA are the protein and RNA starter-machinery needed to

start the replication factory: reverse transcriptase (RT), protease (PR), integrase (IN), nucleocapsid (NC), the two copies dimerized of the genome RNA (gRNA), accessory proteins, and host-derived factors needed for reverse transcription such as $tRNA^{Lys3}$ and RNA Helicase A (RHA) amongst many others (12–15). Reverse transcription is generally accepted to begin prior to CA uncoating (16). The 5'UTR-annealed $tRNA^{Lys3}$ is extended to form a strong-stop single-stranded DNA product ((-)ssDNA) (17). It is currently unknown "when" the uncoating occurs, three debated models are proposed: immediate uncoating, biphasic uncoating, and nuclear pore complex (NPC) uncoating. Immediate uncoating proposes full dissolution of the CA components after ejection into the infected cell. The biphasic uncoating model proposes a similar setup, but only a partial disassembly, with a portion of the CA associating with the reverse transcription complex. The NPC uncoating model proposes that the CA -with stalled reverse transcription complexes- remain intact as it migrates through the cytoplasm to a nuclear pore, where the disassembly leads to continuation of reverse transcription and product injection through said pore into the nucleus for integration (14). A fourth model has been recently proposed, where the CA remains intact through migration across the nuclear pore and then is disassembled (18). TRIM5a, a HIV restriction-factor, catalyzes CA deconstruction via recognition of the intact CA-lattice, activating innate immune signaling pathways (19).

Reverse transcription is performed by the viral RT contained within the by CA. The polymerase is a heterodimer composed of the p51 and p66 subunits. Subunit p51 provides structural support to the p66 subunit and has been

suggested to assist in loading p66 to the primer-template. Subunit p66 is the catalytic subunit, containing the active site for RNA polymerization from free NTPs at its N-terminus and the RNAse H active site located at its C-terminus (endonuclease, hydrolyzes RNA of RNA:DNA hybrid). Fully, the RT activity generates dsDNA from the ssRNA (gRNA) template, degrading it in the process.

Integration of the dsDNA follows RT activity. Localization to the host chromosome is contingent on the true nature of CA uncoating, as it may require trafficking across the nuclear pore complex. The pre-integration complex (PIC) is comprised of the viral dsDNA, IN, and several relevant host factors. Integration is catalyzed by the viral IN, which is composed of three domains: C-terminal zinc-finger domain, the catalytic core, and the N-terminal “beta-barrel”-like domain that binds DNA nonspecifically (20). IN performs viral genome integration by cleaving 3’ ends of host DNA and joining them to the viral dsDNA target 5’ phosphate. Integration is completed by host repair factors, repairing single-stranded gaps generated by IN (21). This establishes a templated HIV-encoding sequence within the host, dubbed the “provirus.”

Successfully integration marks the switch from early to late-phase viral replication. The integrated provirus serves as the template for transcription of viral RNA. This viral RNA will be used both as a template for viral protein production and packaged into CA’s of new virions for further infection. It is established that there are three major open reading frames (ORF) within the viral RNA for protein production: *gag*, *pol*, and *env* (1). The *gag* reading frame encodes the polyprotein Gag (group-specific antigen), a precursor that is cleaved into MA, CA, NC, and

others during virion maturation. The *pol* frame encodes another polyprotein, which is ultimately processed into PR, RT, and IN. The *env* frame encodes for GP160, which is cleaved into the previously mentioned glycoproteins gp41 and gp120. Ribosomal frameshifting, occurring with a frequency of 5-10%, also produces a Gag-pol fusion polypeptide (22). This occurs by a -1 positional frameshift, resulting in an overlap of the *gag* and *pol* reading frames. Differential splicing of nascent viral RNAs can also result in production of additional accessory proteins for infection: Tat, Rev, Vif, Vpu, Vpr, and Nef.

Transcription of the viral transcripts is stimulated from a promoter located in the 5' long terminal repeat (LTR). The LTR is composed of three distinct regions: unique 3 (U3), repeat region (R), and unique 5 (U5) region (7). The transcription initiation site is precisely at the junction between the U3 and R regions (23). The mentioned promoter, located in the U3 region, contains the canonical TATA-like element, which is used to recruit host transcription factors and associate host RNA polymerase II for initiation of transcription. During transcription, alternative splicing can generate multiply, singly (partial), or unspliced transcripts (24). Over time of infection within a cell, alternative splicing is modulated by an unknown mechanism, balancing production of viral elements for the relevant stage of infection. The early-generated transcripts are thought to be multiply-spliced for accumulation of viral accessory proteins, including Tat, Rev, Nef and Vpr, within the host cytoplasm (24). Tat is one of the earliest viral proteins synthesized upon infection. It is trafficked back into the nucleus to bind an RNA structural motif near the 5' end of nascently-transcribing RNAs, called the *trans*-activation region element (TAR).

This interaction stimulates transcription, resulting in a net rate-increase of RNA transcript accumulation. Singly-spliced transcripts are template for production of proteins Vpr, Vif, Vpu, Tat, and Env. Unspliced transcripts (gRNA) can be used for the production of Gag, Pol, and Gag-Pol polyproteins or dimerize for packaging into the CA of new virions.

Transport of the alternatively-spliced viral RNAs across the nuclear membrane are thought to occur by different pathways. It is not known how early multiply-spliced transcripts enter the cytoplasm, perhaps by simple diffusion due to their smaller size. These transcripts may not be ruled by the laws of trafficking for larger transcripts or may transport readily before the innate immune response modulates cellular trafficking. Later singly and unspliced RNAs are transported to the cytoplasm via viral protein Rev, which hijacks the host CRM1 nuclear transport pathway (1, 7, 25). Shuttling of these RNAs require Rev's recognition of the Rev Response Element (RRE). The RRE is a highly structured region, approximately 250 nucleotides in length, located with the *env* encoding region. Alternative splicing and different structural necessities of the various derived RNA species suggests the possibility of a dynamic HIV transcript, whose structural folding is core to proper infection.

The final phase of infection is highlighted by the production of components for new virions and assembly then budding of virions. The dimerized gRNA is coated in NC. Produced Gag and Gag-pol polyproteins associate with the NC, alongside Nef, Vif, and Vpr. The $tRNA^{Ly3}$ (reverse transcription primer) is loaded onto a structured region of the viral RNA known as the "primer binding site" (PBS)

(26). These are encapsulated within the assembled CA protein matrix. This CA is coated by an outer membrane, containing the Env heterodimer, with the surface-displaying gp120 and gp41 glycoproteins, and this outer membrane is stabilized by the subsurface MA protein matrix. The final budding requires interaction with the host endosomal sorting complex required for transport (ESCRT) via Gag (27). The HIV PR then cleaves Gag, maturing the virion into its final infectious state.

1.3 HIV-1 5'UTR RNA Structures

The HIV-1 genome consists of approximately 9.6 kb (28). The central region of the genome, containing the genes encoding protein, is flanked by a repeated sequence of approximately 634 bases, known as the 5' and 3' long terminal repeats (LTR). This palindromic sequence contains information needed for integration into the host genome and the proximal region of the repeat extends into the 5' UTR, encompassing TAR and PolyA. The 5'UTR RNA region of the HIV genome, upstream of the encoding genes, is the most conserved region within the HIV genome and is approximately 350 nucleotides in length (29). It is both highly structured and structurally dynamic, suggesting that conformational structure elements are key throughout the lifecycle. The 5'UTR region has documented involvement in reverse transcription initiation, transcription, transcript splicing, translation initiation, gRNA dimerization, and RNA encapsidation during virion assembly (30).

The exact structure(s) of the 5'UTR and their roles in infection are still debated (31, 32). Many structural models have been proposed, with some

agreement on structural features. The beginning of the 5'UTR is marked by the *trans*-activation region (TAR), which readily structures into a thermodynamically-stable hairpin loop. Immediately following the TAR sequence is the 5'polyadenylation signal (polyA). PolyA is utilized in the 3'UTR for polyadenylation of RNA transcripts and is theorized to also form a hairpin-like structure. PolyA's role in the 5'UTR and how it relates to overall 5'UTR structure and/or infection is unknown. Following PolyA is the primer binding site (PBS), which was named for binding $tRNA^{lys3}$ to for reverse transcription initiation. Its structure is dynamic and has not been sufficiently characterized. Notable within the PBS macrostructure is a stem loop-like structure, which has been called multiple names, including the tRNA-like element (TLE) and the U5-top hairpin. PBS is followed by four predicted stem loops (SL1-SL4). The most immediate (SL1) is the dimer initiation site (DIS), containing the palindromic "GCGCGC" sequence, the interface at which gRNA dimerization occurs. Following DIS is the splicing-donor (SD) signal hairpin (SL2), which is the 5' splicing donor for alternative splicing of nascent RNA transcripts and has been also proposed to bind Rev for nuclear transport. Following SD is the Psi (Ψ) hairpin (SL3), the packaging signal for the dimerized gRNA. Finally, is the AUG Gag start codon hairpin (SL4), is the beginning of the Gar protein ORF.

The current general consensus model is that the HIV 5'UTR structure adopts to predominant macrostructures: a "monomer" and "dimer" conformation. This understanding is born out the history of discovery in the field. It is well-established that the genome dimerizes for packaging, yet single transcripts are likely present throughout the lifecycle for trafficking and translation based on our

understanding of biology. The role of the “monomer” conformation in infection is completely unknown. Recent work has revealed that 5’ capping likely plays an important role in RNA folding and fate (33). Newly transcribed RNAs typically contain 1-3 guanines and are capped by a 5’,5’-triphosphate-linked 7-methylguanosine. ^{Cap}1G transcripts RNAs were found to preferentially dimerize, burying the cap within TAR and PolyA helices. Conversely, ^{Cap}2G and ^{Cap}3G were found to remain monomeric and with an exposed cap. Additionally, these ^{Cap}2G and ^{Cap}3G cap variants had a much higher affinity for eIF4e, the initiator of recruitment and assembly of eukaryotic translation machinery. This suggests that a monomeric 5’UTR may play an important role in translation.

Additionally, experimentation in laboratories with HIV 5’UTR RNA has repeatedly shown the existence of this second conformer, suggesting the existence of a kinetically or thermodynamically stable conformation that would likely play some role in infection (34). This conformation is likely reinforced by association with host or viral protein. One possibility is that its role is very early in infection, making it difficult to study. There has been little or no work establishing structure: function relationships with regards to transcription, nuclear export, splicing and alternatively spliced transcripts, or translation. Alternative conformations almost certainly have a role within one of these spaces. Current work examining the 5’ UTR structures- such as with NMR, X-ray crystallography, and cryoEM- have used refolded RNA, where an RNA is heated to denature higher order structures and base-pairing, then quickly cooling it and allowing it to refold in physiologically-relevant conditions (salt, pH, temperature). These

thermodynamically favorable structures may not be representative of the structures present within cells. It is plausible that nascently-transcribed RNAs are trapped in kinetically favorable conformations by host and viral factors, preventing equilibration to the lowest energy conformation.

1.4 HIV-1 5'UTR Structure and Translation

Note: This section is taken from a portion of a review article detailing HIV-1 translation in the process of submission.

The HIV-1 5'UTR coordinates various replication events, including translation (1). It is composed of several hairpins and structural motifs, including TAR, PolyA, PBS (primer binding site), DIS (dimerization initiation site), SD (splicing donor, or SL2), Ψ(SL3) and AUG (SL4) (35–38). Early studies have suggested some structural elements within the 5'UTR modulate translation. Nucleotides 104-336 were reported to harbor an internal ribosome entry site (IRES) to promote translation in a cap-independent manner (39, 40). IRES-dependent translation has been demonstrated to be active during G2/M phase, when the Cap-dependent translation is inhibited, but the IRES activity is weak or absent in cap-dependent translation initiation by the TAR-Poly(A) region (39). Deletions or mutations in PBS, DIS and SD altered the 5'UTR IRES activity in dual-luciferase reporter assays. While disruption of the upper hairpin of the stem loop upstream of PBS (nt 143-167) led to enhanced IRES activity, deletion of the loop residues down stream of PBS (nt 202-217), or mutations in DIS and SD, were reported to negatively modulate IRES activity (40, 41).

DHX9/RNA Helicase A (RHA) is a post-transcriptional effector that simulates translation of retroviral and selected cellular mRNAs through recognition of the post-transcriptional elements (PCE) in these mRNAs (42, 43). The HIV-1 RU5 region (nt 1-181) was identified as the PCE that is recognized by DHX9/RHA to facilitate polyribosome association and stimulate cap-dependent translation (41, 44). RNA co-precipitation data show that RHA weakly associated with RNA transcripts containing only R or U5 but tightly associated with RNA transcripts containing both R and U5, indicating that both R and U5 are required for the PCE activity (44). The PCE overlaps with the RHA binding region in the 5'UTR that recruits RHA during assembly (45). It is possible that the same protein:RNA interaction exists in both PCE-mediated translation modulation and recruitment of host RHA in to progeny virions.

While the structural elements within the 5'UTR could have a direct impact on translation, mutations in these regions may affect the overall folding of the 5'UTR and thus affect translation indirectly. *In vitro* refolded 5'UTR RNA can adopt both monomer and dimer conformations. Combining chemical probing and computational prediction, the Berkhout group proposed that the HIV-1 5'UTR folds into two mutually exclusive conformations, LDI (long distance interaction) and BMH (branched multiple hairpins) (46–48). In the LDI conformation, a long-distance base pairing interaction between the PolyA and DIS motifs is formed. Thus, the palindrome sequence in DIS is sequestered and cannot promote RNA dimerization. The Gag start codon forms base pairs with residues between splicing donor and the Ψ hairpin. In the BMH conformation, PolyA does not interact with

DIS and they both fold into local hairpin structures. The palindrome sequence in the loop of the DIS is exposed to promote RNA dimerization. The residues flanking the Gag start codon form long-range interactions with U5 residues, which was shown to be critical in directing genome packaging (13). To examine whether the LDI-BMH equilibrium regulates translation, a series of mutations were introduced to promote either the BMH or LDI conformation. However, while consistent monomer:dimer structural equilibrium was observed when the RNAs were refolded under in vitro conditions, no correlation between the translation efficiency and the BMH-LDI RNA structure was observed (49).

Later, NMR studies of the 5'UTR suggest a different monomer:dimer model (50–52). In the monomer conformation, DIS is paired with residues in U5, and AUG forms a hairpin at the 3'-end of the 5'UTR. In the dimer conformation, AUG forms long range interaction with residues in U5, displacing DIS to promote dimerization. Residues spanning SD are sequestered by participating in the formation of a three-way junction (50, 51, 53). It was reported that the nucleotides at the bottom of PolyA hairpin is critical in regulating the monomer:dimer equilibrium. For example, U103 substitutions are sufficient to shift monomer:dimer equilibrium toward monomer conformation in-solution, and increase the rate of *de novo* Gag synthesis (54, 55).

Recent studies probing the 5'-terminus of HIV-1 viral RNAs suggested that viral RNA transcripts with different numbers of guanosines have distinct functions in viral replication). Two groups independently reported that 1G capped RNA mainly existed in virions, and 2G/3G capped RNA mainly in cells (54, 56).

Additionally, 2G/3G capped RNAs were found enriched on polysomes that underwent active translation (54). Structural investigation of the capped RNA discovered that the heterogeneity of the 5’- transcription starting sites led to shifted monomer:dimer equilibria of the 5’UTR with different levels of cap exposure (52). RNA transcribed with three guanosines folded into a monomeric structure with the 5’- three guanosines forming base pairs with the three Cs at the bottom of TAR, disrupting the base pairings at the bottom of PolyA hairpin, and thus favoring the U5:DIS interaction to sequester the dimerization site. The m7G cap is therefore flexible and can be recognized by cap-binding protein eIF4E. In the RNA transcribed with one guanosine, the m7G cap together with G1 to form base pairings with C55-C56, without disrupting the bottom of PolyA stem. Therefore, the RNA preferentially folds into a dimer conformation with DIS exposed. Because m7G cap is paired with C56, it is no longer exposed and thus the RNA cannot bind eIF4e (52). These studies established a relationship between capped 5’UTR structures and function regulation in translation and genomic RNA packaging.

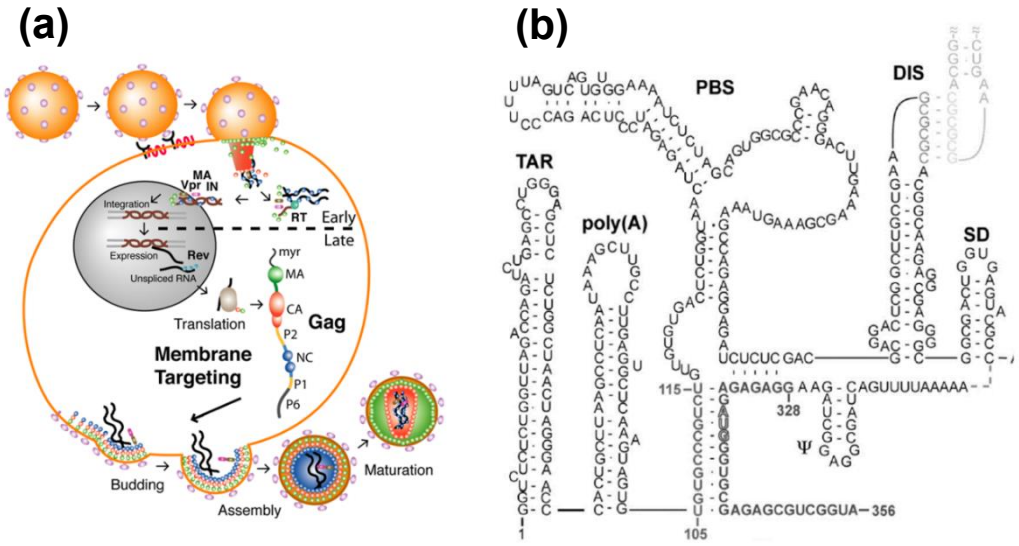

Figure 1-1. Cartoon Model of the HIV-1 Lifecycle and Model of the 5'UTR RNA Secondary Structure. (a) The infectious virion (top left) binds CD4 surface glycoprotein (black hook) of target cell via gp41 surface glycoprotein (purple). Structural rearrangement triggers formation of spike protein complex, releasing the capsid into the cell (red cone). Reverse transcription generates double-stranded DNA. The viral DNA is translocated into the host nucleus and integrated into host chromosome by viral integrase. **(b)** The HIV-1 UTR RNA is a highly-structured and conserved region. It is led by the TAR hairpin, which is traditionally known to bind Tat. Following is the polyA region, whose role is unknown in the 5'UTR, but provides the polyadenylation signal in the 3'LTR. Following the end of the 5'LTR is the PBS region, containing the TLE stem. This 5' UTR concludes with a series of 4 hairpins (SL1-4). SL1 contains the interface for dimerization (DIS), SL2 is the 5' splicing donor for nascent transcripts (SD), SL3 contains the packaging signal for RNA encapsidation (Ψ), and SL4 is the AUG Gag start codon hairpin (AUG).

CHAPTER 2. HIV-1 5'UTR Structural Conformations and Their Implications in Transcription and Translation in the HIV-1 Infectious Life Cycle

Note: This chapter contains work of a research paper in the process of submission.

Abstract

The HIV-1 5'UTR structure is highly structured, structurally dynamic, and the most conserved region of the genome. It coordinates various viral replication steps such as splicing, translation, genome packaging, etc. Persistent translation of some viral proteins was observed when host eIF4E-initiated translation, the global mRNA translation pathway, is inhibited to restrict pathogens. However, it remained unclear how the viral mRNAs sharing the same 5'-sequence interact with distinct host factors and utilize different translation pathways. Here we show collaborative work that demonstrates the existence of a novel, alternate translation pathway adopted by HIV-1 to promote virus fitness to the detriment of the host. While the multiply spliced mRNA undergoes cap exchange to associate with eIF4E for translation, singly spliced and unspliced mRNAs form a complex with nuclear cap binding protein (NBCP3) and host RNA helicase A (RHA) for an alternative cap-dependent translation. The specialized translation pathway ensures efficient viral accessary and structural protein translation to bypass the host eIF4E restriction as a host defensive mechanism. We show evidence implicating the

5’UTR structure in determining fate of RNAs between these two translation pathways. Additionally, we demonstrate that folding of the upstream elements of the HIV-1 5’UTR influence folding of downstream elements, suggesting the fate of the mRNA transcripts are determined when they are nascently transcribed. This work sets the foundation for further investigation into mechanism of 5’ UTR folding in early and late HIV-1 mRNAs and its relevance in the HIV-1 infectious lifecycle.

Introduction

HIV-1 viral protein translation utilizes host translational machinery for synthesis of its viral proteins. Eukaryotic translation of mRNA is mostly controlled by regulation of the initiation step (57). Translation initiation is achieved by recognition of the 7-methylguanosine (7mG) cap present on the 5'UTR of the mRNAs by eukaryotic translation factor eIF4E, which is the central node for global translation regulation (58, 59). This factor recruits additional translational machinery, including eIF4G, and assembles messenger ribonucleoprotein complex (mRNP) to the 40s subunit of the ribosome.

The host Innate immune response inhibits eIF4e cap-binding by inhibition of serine-threonine kinase, mechanistic target of rapamycin (mTOR) (60). mTOR inhibition causes a rapid reduction in global mRNA-to-protein translation rate by cell signaling, leading to allosteric inhibition of eIF4E by 4E-BP (61–65). Treatment with mTOR catalytic site small molecule inhibitors, e.g. Torin-1 have demonstrated to decrease translation of certain mRNAs (66–68). Interestingly, some mRNAs have been found to bypass eIF4e-dependent translation initiation. Studies have shown the ability to initiate translation using an eIF4GI homolog, DAP-5, or CBP80 tethered to the m7G-cap by NCBP3 (69, 70). Recently, it has also been demonstrated that the number of 5' guanosines including the 7-methylguanosine cap influence HIV RNA folding and fate (33). Newly transcribed RNAs were revealed to contain 1-3 guanines and are capped by a 5',5'-triphosphate-linked 7-methylguanosine. ^{Cap}1G transcripts RNAs were found to preferentially dimerize, burying the m7G-cap within TAR and PolyA helices by forming a base pair with

C57. Conversely, ^{Cap}2G and ^{Cap}3G were found to remain monomeric and with an exposed cap when the RNA was refolded *in vitro*. Additionally, these ^{Cap}2G and ^{Cap}3G cap variants had a much higher affinity for eIF4e, the initiator of recruitment and assembly of eukaryotic translation machinery (52).

The current model of HIV infections supposes that the primary transcript can be differentially spliced into 3 class. In early expression, alternative splicing generates multiply spliced (MS) transcripts that encode Tat, Rev, and Nef. Later, singly spliced (SS) and unspliced (US) transcripts are predominant for synthesis of Env, Vpu, Vif, and Vpr and Gag and Gag-Pol protein, respectively (71, 72). These later transcripts utilize the host CRM1 nuclear transport system (Rev/RRE-dependent transport) and form polysomes with CBP80 (73). Additionally, translation of these SS and US transcripts have been found to persist during downregulation of eIF4E (40, 74).

Altogether, it has been established that HIV-1 has an alternate route for protein synthesis following downregulation of eIF4E that is facilitated by an alternate cap-binding initiation of translation, and the 5'-UTR guanosine/cap content affects 5'UTR folding and interaction. The nature of this pathway is currently unknown, as is the role of the 5'-UTR RNA structuring within translation pathways.

Results

HIV-1 Translation of Incompletely Spliced mRNAs Employs an Alternative, Non-eIF4E Pathway Which Requires RHA:PBS Interaction.

Our collaborator's work (Kathleen Boris-Lawrie Lab; University of Minnesota) has demonstrated with existence of an alternate translation pathway utilized for HIV-1 SS and US transcripts (Figure 2-1; in process of submission). These transcripts' 5' caps are trimethylated (TmG) and translation is initiated by CBP80 tethered to the cap by NCBP3 (Figure 2-1 Right Panel). This contrasts with the established, canonical translation pathway that is initiated by eIF4E binding of the 7mG cap of nascent transcripts (Figure 2-1 Left Panel). We have previously established (75) that host RHA binds to the three-way junction structure of HIV 5'UTR PBS, which contains a structured TLE stem, whose secondary structure is modeled in Figure 1-1b. Immunoprecipitation (IP) of eIF4G (canonical translation) vs. NCBP3 (non-canonical translation) demonstrated RHA association within the RNP complex in the non-canonical translation pathway (Figure 2-2a). Furthermore, western blot of HIV^{NL4-3} infected cells with showed that HIV-1 Gag translation (US mRNA) employs this specialized translation pathway and is resistant to mTOR inhibition (Figure 2-2b left). However, deletion (ΔPBS) or disruption (A140C) of PBS structure resulted in reduced Gag translation in the presence of Torin-1. This implicates proper PBS structuring in the non-eIF4E translation. Disrupting this structure by single point mutation A140C in the TLE stem of PBS was found to prevent RHA binding. Titration of RHA to WT vs A140C PBS RNA showed specificity for the WT RNA relative to A140C RNA (Figure 2-

2c). Using NMR, it was demonstrated that there are significant differences in PBS structure between WT and A140C (Figure 2-2d). The three-way junction structure was determined by NMR and using small-angle X-ray scattering (75). The secondary structure of the three-way junction PBS (Figure 2-2e) differs substantially from the predicted secondary structure of A140C (Figure 2-2f), which suggests an elongated PBS structure in which the TLE stem has been destroyed. Furthermore, it was demonstrated that 5'-cap trimethylation is abrogated in the absence of RHA (data not shown), suggesting the need of RHA for proper TmG generation. The Summers lab has recently shown that the 5'-cap of transcripts alters the 5'UTR structure and affects affinity for eIF4E (33). Altogether, this suggests that the 5'UTR structure of transcribing RNAs influences the fate of transcripts, potentially implicating the nascent structuring in splicing, nuclear export, and translation.

The structures of HIV-1 mRNAs engaged in NCBP3-RNP and eIF4e-RNP are different.

To determine differences in the folding of the 5'UTR of transcribed RNAs in each translation pathway, the structures of RNAs isolated from IP of each translation mRNP were investigated by Selective 2' Hydroxyl Acylation analyzed by Primer Extension (SHAPE). SHAPE experiments take advantage of the dynamic structure of folded RNAs, utilizing reagents that preferentially modify the 2'-OH of the RNA backbone in structurally flexible regions (76). The modified RNAs can then be reverse-transcribed in the presence of manganese, inducing mutations at labeled locations. Sequencing of the population gives a distribution of

mutational frequency across each position, yielding information that gives insight to the structure of the RNA at the time of labeling. To study the structures of HIV-1 mRNA engaged in various translation pathways, cells were treated with 5-nitroisatoic anhydride (5NIA) and RNA was purified from the IP of either CBP80 or eIF4E in HIV-1^{NL4-3} infected cells. The 5'-sequences prior to splicing donor site were reverse transcribed by SuperScript II in the presence of $MnCl_2$. The cDNA was then subjected to Illumina sequencing, and the mutation rates were measured to calculate SHAPE activities at each nucleotide position. We found significant differences in SHAPE activities of the RNAs, suggesting that the 5'UTR structures in these complexes of the two translation pathways are different (Figure 2-3a and Figure 2-3b). Notably, the SHAPE activities were drastically different in the TLE stem. Through collaboration with Dr. Shi-Jie Chen laboratory, computational modeling of transcribing RNAs PBS region with the SHAPE activity restraints generated a model remarkably similarity to the three-way junction-like structure of the CBP80 RNA that has been established within the field (Figure 2-3c). The same methods applied to eIF4E RNA generated a completely different structure that lacks many of the classically-defined motifs of the 5'UTR (Figure 2-3d). Our SHAPE and computational modeling are consistent with previous discoveries that the mRNAs' structures are different in eIF4E-RNP and NCBP3-RNP.

Structures of nascently transcribed RNAs could be different from their thermodynamic favorable conformations.

The differences in 5'UTR structures from SHAPE analysis of eIF4E and CBP80 pathways, established influence of the RNA structure led us to ask the

question of when these structural differences manifest in the RNAs lifetime. While it is possible that the US/SS mRNAs and MS mRNAs have different 5'UTR structures due to splicing, they share the same sequence 1-296 (TAR to first splicing donor site) and may kinetically trapped in the same conformation when they are nascently transcribed. Recent NMR analysis of dimeric 5'UTR RNA reveals that the AUG forms a long-range interaction with U5 residues and displaces DIS to promote dimerization (77). Thus, refolding of 5'UTR constructs lacking AUG element resulted in structural equilibrium shift towards monomer (Figure 2-4, left lanes). Interestingly, nascently transcribed 5'UTR constructs all migrated mainly as a dimer on a native agarose gel, regardless the 3'- boundary of the RNA (Figure 2-4, right lanes). These data demonstrate that 5'UTR RNA can have different structural properties depending on whether the RNA has been transcribed or refolded (Figure 2-4). Most structural studies on the HIV 5'UTR have utilized refolded RNA, which implicates the possibility that they are not representative of the infectious structure. The structure of the monomer and dimer in the nascently transcribed 5'UTR RNA was gel purified and analyzed by SHAPE. We found differences across multiple major structural elements in the refolded dimer (1-344) and nascently transcribed monomer, including TAR, PolyA, TLE, and PBS (Figure 2-5). Therefore, it is possible that a small portion of the nascently transcribed RNA adopts the monomer conformation without PBS three-way junction structure and majority of the RNA adopts the dimer conformation with PBS forming the three-way junction recognized by host RHA. The minor population undergoes multiply splicing and cap exchange to bind eIF4e for translation. The

major population harbors the RNA-shape dependent recognition site for host RHA, which forms NCBP3-RHA-RNA complex for hypermethylation of cap and employs the specialized translation pathway to produce viral accessary and structural proteins.

Transcriptional pause in the absence of Tat favors the RNA folding for eIF4e translation pathway.

Host RNA Polymerase II transcription of HIV-1 RNAs transitions from initiation, with synthesis of the beginning of the transcript, escaping the promoter, to the stable elongation complex. In the early stage of transcription when viral Tat protein is not available, RNA Polymerase II pauses at position U62, which is in the sequence immediately after TAR (Figure 2-6a). The nascently transcribed RNA folds into a hairpin structure that blocks the RNA exit tunnel of RNA Polymerase II, while approximately 18 nt single stranded RNA at the 3'-end is protected in exit tunnel (78). In the absence of viral Tat protein, transcription is significantly stalled by this pause structure and inefficient. Once Tat is transcribed from early MS RNAs, it releases positive transcription elongation factor P-TEFb from the 7SK small nuclear ribonucleoprotein complex (snRNP) (79). Together, these bind the paused RNA Polymerase II, leading to the phosphorylation of RNA Polymerase II-CTD, release of the pause, and greatly enhancing transcriptional rate (78). Because Tat binding to TAR is critical for formation of a stable elongation complex, and the binding interface is exposed during the U62 pause (80), this supposes that binding of Tat may influence the folding of the transcribing RNAs. However, Tat is absent in early transcription of ms mRNAs from the proviral DNA, suggesting that

these early transcribed RNAs may have a different 5'UTR structure than later SS and US RNAs. Immediately following TAR is the PolyA region of the 5'LTR. The function of the region is known to be the polyadenylation signal in the 3'LTR but has no known function in the 5'LTR. We propose that in the absence of Tat when RNA Polymerase II pauses at U62, the TAR RNA forms a shorter, alternate hairpin, whose 3' end structures with a portion of PolyA, which leads to alternative folding of the downstream elements, including obstruction of the DIS hairpin, decreasing dimerization efficiency (Figure 2-6b upper). In the presence of Tat, Tar forms a longer hairpin, leading to a PolyA self-structuring, resulting in different conformation of downstream elements, including proper structuring of DIS to promote dimerization of RNAs (Figure 2-6b lower).

To investigate the role of the PolyA sequence in downstream folding of PBS and mimic the RNA Polymerase II transcriptional pause, we created a series of 5' truncations of the PolyA region (Figure 2-7a). The RNAs were transcribed *in vitro* with T7 polymerase and then ran on native agarose gel. The RNA transcribed from different positions of PolyA showed differences in the ratio of monomer and dimer RNAs, indicating the available PolyA sequence play a key role in downstream folding (Figure 2-7b). Interestingly, the truncation immediately following U62 of full length 5'UTR RNA showed the highest ratio of monomer RNA.

We hypothesize that because the RNA exit tunnel of RNA Polymerase II protects the 18 nucleotides at the 3'-end of nascently transcribed RNA, maintain its single-strandness and preventing it from interacting in the folding pathway of the upstream, exposed RNA. This 5'- RNA could find its thermodynamic favorable

conformation during the pause, so the single stranded RNA released from pause cannot participate in forming TAR hairpin but instead could be trapped in local small hairpin structures that prevent formation of the PolyA hairpin. To investigate if TAR folding during the U62 pause affects downstream folding, we annealed biotin-5'-template to streptavidin beads and performed stepwise elongation with T7 polymerase of 3NTPs to near the pause site (Figure 2-8a). Following each elongation, the sample was washed before addition of the next 3NTP set. Once the pause site was reached, added all 4NTP's to allow extension to the end of the template, releasing the transcribing RNA in solution. Heparin was added to the 4NTP final elongation step to trap polymerase, preventing re-initiation of transcription. When ran on native agarose gel, we found that an increase in relative monomer concentration when termination occurred following Set 8, near the U62 pause site, relative to Set 7 -> termination, or transcription without controlled elongation of TAR (Figure 2-8b). These data show that the downstream RNA folding is sensitive to the transcriptional pausing site, and support the hypothesis that the transcriptional pause affects the co-transcriptional folding of nascent mRNA transcripts.

To further test the hypothesis in the RNA polymerase II system, we performed SHAPE analysis on RNAs transcribed in HeLa cell nuclear extracts under the control of the native HIV-1 5' promoter in the presence or absences of Tat. Briefly, DNA template (-450 to +612) was added to nuclear extract of HeLa cells. RNA transcription was initiated by incubating template in the buffered extract mixture at 37°C and then adding NTPs. To examine whether Tat transactivation

also affect RNA structure, nuclear extract was prepared from HeLa cells transfected with pCEP4-TAT (a kind gift of the Marc Johnson laboratory, University of Missouri) to express Tat. A 10-fold increase in the RNA extracted (measured by OD 260 nm) indicate the successful transactivation of Tat. SHAPE-seq was performed on RNAs produced in HeLa nuclear extracts without and with Tat. Due to the unexpected DNA contamination, the SHAPE modification efficiency was low and the data quality was poor. Nevertheless, notable differences in the PolyA region were observed between the two samples, indicating differences in structure of the PolyA region in the presence and absence of Tat. (Figure 2-9), supporting the hypothesis that the Tat-mediated release of transcriptional pause affects the folding of nascently transcribed mRNAs.

Discussion

Through collaboration with Dr. Kathleen Boris-Lawrie laboratory, we identified the HIV specialized translation by CBP80-NCBP3 and RHA, which provides an explanation for abundant literature that *in cellulo* translation of HIV RNA continues during shutdown of eIF4E-dependent translation by virus infection or mTOR inhibitors (81, 82). Hyper methylation of the 5'-cap of the Rev/RRE-dependent HIV mRNAs and binding of RHA to the structure of the 5'-UTR inhibits cap-exchange to eIF4E (Fig. 2-1). My studies focus on the transcriptional folding of the 5'UTR and reveal that mutually exclusive structures are formed when the 5'UTR is nascently transcribed and these structures are likely correlate with MS and SS/US mRNAs. In the absence of Tat, the strategic pausing after TAR hairpin leads to kinetic trapping of the RNA conformation that favors multiply splicing and

employs eIF4e translation pathway. Our data provide experimental evidence supporting the co-transcriptional folding pathways of the HIV-1 5'UTR that promote formation of biologically active RNA structures recognized by distinct host factors and assembles into different RNPs for efficient gene expression.

The controlled elongation of TAR by *in vitro* transcription by T7 polymerase yielded promising results, where we saw an increase in the monomer concentration of the nascently-folded 5'UTR RNA (Figure 2-8). These experiments needed to be repeated and quantified, but the results are apparent. An important consideration is the differences in transcription between T7 polymerase and RNA Polymerase II. T7 polymerase transcribes by itself, while RNA Polymerase II requires several protein cofactors and regulatory elements to facilitate effective transcription.

Additionally, The RNA Polymerase II has a longer RNA exit tunnel, which prevents RNA structuring during transcription (Figure 2-6a). This tunnel has been empirically determined to protect approximately 18 nt's (80). Examination of a published crystal structure of T7 polymerase (PDB:1CEZ) (83) suggests a shorter exit tunnel of approximately 10 nt's. This would be implicated in our data of controlled elongation of TAR (Figure 2-8b), which would affect the amount of available TAR sequence for folding of the transcribing RNA during our synthetic U62-like pause. Although imperfect, this experiment has provided proof-of-concept for the folding. Future experimentation will shift to doing this a more *in vivo*-like system. Our transcriptions in HeLa nuclear extract in the presence of Tat (Figure 2-9) were low quality due to DNA template contamination and using the wrong read

depth during deep sequencing. This needs to be repeated. We expect Tat binding TAR to alter the PBS structure.

Eukaryotic transcription has additional co-transcriptional processes that T7 polymerase lacks. These include 5'-capping, alternative splicing, and 3'-polyadenylation. It has been previously established that the 5'-guanosine content and 5'-7mG cap affects 5'UTR folding. Future experimentation will investigate how these elements will affect the 5'UTR folding. For capping, we will perform *in vitro* transcription without controlled elongation with 1, 2, 3, or 4G content at the start of the DNA template. It would be expected that the higher G content would decrease dimerization. As mentioned, it was shown that a lower G content promotes dimerization, so we would expect to see a G-concentration dependent shift in the monomer: dimer equilibrium. It is also possible to preload template-bound T7 with 7mG cap. We expect splicing and polyadenylation to have a much more subtle- if not at all- effect on 5'UTR folding, as we have established that downstream folding of elements is controlled by earlier elements in the transcription.

The mechanism that controls the 5' G content of new transcribed RNAs in the HIV infectious lifecycle. The integrated proviral template contains three G's at the start of the template. The heterogeneous 5'G content between transcripts has been proposed to be due to transcriptional slippage. However, because the G content affects folding- and likely RNA fate-, there is likely a unknown mechanism that regulates this during the lifecycle. We hypothesize that the higher G content is facilitated by the presence of a protein or RNA element from multiple possible sources. One option is that late-stage viral protein products (from SS or US RNAs),

such as Gag, are shuttled into the nucleus to modulate the transcription complex, promoting higher G content for non-canonical translation. Alternatively, the non-canonical transcription complex itself could create an alternate initiation complex. Another possibility is that any of the host factors from the innate immune response downregulating canonical eIF4E-initiated translation could lead to this change.

Author Contributions

Conceptualization X.H., K. BL., and S.S.; Methodology X. H., Z. S., S. S.; Nuclear extract preparation J.J.; 5NI1a-treated RNA's from IP for SHAPE G.S. and K.BL.; Computational modeling of the 5'->3' folding simulations of RNAs based on SHAPE data performed by The Shi-jie Chen lab.

Experimental Procedures

Plasmids

Plasmids pNL4-3 (NIH AIDS Reagent Program) and pUC19 containing the HIV-1 5'UTR sequence for DNA template synthesis were generated using methods described previously (84). The pNL4-3 HIV-1 sequence was fused into the pUC19 sequence for amplification. Plasmid sequences were verified by Sanger sequencing (DNA Core, University of Missouri, Columbia, MO).

In Vitro RNA Transcription

RNAs used for refolding were transcribed using purified T7 polymerase. Corresponding DNA templates were generated via PCR and purified prior to transcription. Transcriptions were performed at 5-20 mL volumes using ~0.5-2 mg of purified DNA template, 40 mM Tris-HCl (pH = 8.1), 5 mM DTT, 1 mM spermidine, 0.01%Triton X-100, 5-10 mM NTP's, and 5-25 mM $MgCl_2$. Optimal concentrations

of $MgCl_2$ and NTP's to maximize RNA yield was determined using small-scale reactions following by RNA staining on polyacrylamide gel. The transcriptions were incubated for 3 hours at 37C, and the reaction was then quenched using a Urea-EDTA mixture (7 M Urea pH = 8.0. 250 mM EDTA). The RNA was isolated on polyacrylamide denaturing gel by electrophoresis using FisherBiotech DNA sequencing system at 16W overnight. The RNA bands were visualized via UV-shadowing, excised, and then eluted overnight using the Elutrap electroelution system (Whatman) at 100 V. The purified RNA fraction was washed with 2M NaCl and then desalted within the Amicon Ultra-4 Centrifugal Filter Device (Millipore). RNA concentrations were determined by measuring absorbance at 260 nm using the NanoDrop 2000c (Thermo Fisher Scientific).

Transcription of Nascently-Folded RNAs

Corresponding DNA templates were generated via PCR and purified prior to transcription. Transcription of all RNAs whose folding during nascent transcription were monitored were performed under identical conditions. The transcription reaction was performed at a volume of 40 µL using 0.01 ρM final concentration of purified DNA template, 40 mM Tris-HCl (pH = 8.1), 5 mM DTT, 1 mM spermidine, 0.01%Triton X-100, 10 mM NTP's, 0.5 µM T7 polymerase, and 15 mM $MgCl_2$. The reaction was incubated for 1 hour at 37°C. Following this, 10 µL of transcription + 2 µL of 50% glycerol was loaded onto an ethidium-bromide-containing 2% native agarose gel (tris-borate) that was pre-chilled on ice, and then ran at 100 V for 60 minutes. The gel was then visualized using the Molecular

Imager Gel Doc XR+ System (Bio-Rad). Image quality was optimized using the Image Lab software (Bio-Rad).

HeLa Nuclear Extract Preparation

Nuclear lysate extract of HeLa and HeLa transfected with pCEP4-TAT was performed by Juan Ji (University of Missouri) according to standard methods (85).

SHAPE

Nascently-transcribed RNAs were transcribed by T7 polymerase identically to methods as in “Transcription of Nascently-Folded RNAs” and then treated with 5-nitroisatoic anhydride (5NIA). For 5NIA labeling 2’OH labeling, samples were treated with 25 mM of 5NIA and incubated at 37°C for 5 half-lives (8.33 minutes) and then put on ice to quench labeling. Refolded RNAs were diluted to 1 µM in 10 mM Tris-HCl (pH = 7.4) and boiled for 3 minutes. They were then snap-cooled on ice for three minutes. Salts were then added to the sample (1 mM $MgCl_2$, 140 mM KCl, 10 mM NaCl) and allowed to equilibrate for 1 hour at 37°C before labeling with 5NIA. HeLa Nuclear extract samples utilized nuclear extract from HeLa cells or HeLa cells transfected with pCEP4-TAT (a kind gift from Marc Johnson laboratory, University of Missouri). For a reaction volume of 250 µL, 2.5 µg DNA template containing the native 5’LTR promoter extending to the *gag* region, 60 µL of 1x HeLa TXN Buffer (20 mM HEPES, 100 mM KCl, 0.5 mM EDTA, 20% glycerol), and 3 mM $MgCl_2$ was added and incubated at 37°C for 10 minutes. Following incubation, 1.2 mM NTP’s were added, and the reaction was incubated at 37°C for 1 hour. The samples were then treated with 5NIA. Following treatment, the samples were put on ice for 30 minutes. The samples were then treated with

17.5 µL DNase I Reaction Buffer (NEB) and 5 ug of DNase I (NEB) for 30 minutes at 37°C. All samples' RNA was purified using standard TRIZol and EtOH-precipitation methodology and brought to a volume of 50 µL of RNase-free H2O. Following 5NIA-labeling and purification, samples were reverse-transcribed using Superscript II (Thermo Fisher Scientific). For each sample, to 10 µL was added 1 µM of primer, the sample was incubated at 65°C for 5 minutes, then cooled on ice. To the annealed mixture was added 8 µL of 2.5x MaP Buffer (250 mM Tris-HCl pH = 8.0, 187.5 mM KCl, 15 mM MnCl2, 25 mM DTT, 1.25 mM dNTP's) and then incubated for 2 minutes at 42°C. Superscript II was then added (1 µL of 200 U/µL) and incubated at 42°C for 3 hours. The cDNA was then purified using illustra MicroSpin G-25 Columns (GE Healthcare). Adapter sequences and barcodes were then annealed to each sample using PCR, and sequencing was performed in a single NovaSeq SP -PE150 flow cell using an Illumina NovaSeq (DNA Core, University of Missouri). Sequencing data was analyzed using ShapeMapper 2 (Weeks Lab; University of North Carolina).

Controlled In vitro Transcription Elongation of TAR

This method was adopted from the previously-established PLOR methodology (86). 5'-biotinylated DNA template was generated using a biotinylated forward primer (synthesized by IDT) and purified. To 500 µL 50% streptavidin agarose beads slurry (GoldBio) was added 1 µM of the template and incubated in Buffer A (40 mM Tris-HCl, 100 mM K_2SO_4, 6 mM $MgSO_4$, pH 8.0) for 3 days at 4°C. Following incubation, the sample was washed with Buffer B (40 mM Tris-HCl, 100 mM K_2SO_4, 6 mM $MgSO_4$, 10 mM DTT, pH 8.0) for removal of non-

annealed template. Following wash, to 200 µL of annealed beads was added 20 µM T7 polymerase and incubated at 37°C for 10 minutes. Following this, 2 mM of GTP, CTP, and UTP and the volume was brought to 1 mL with Buffer B and incubated at 37°C for 15 minutes. The sample was then loaded into a Poly-Prep column (Bio-Rad) and washed thoroughly with Buffer C (40 mM Tris-HCl, 6 mM $MgSO_4$, pH 8.0) to remove all non-complexed T7. Sequential 3 nucleotide elongation was performed by adding mixtures of 3 nucleotide at a time (20 µM per position) to 1 mL total volume in Buffer D (40 mM Tris-HCl, 6 mM $MgSO_4$, 10 mM DTT, pH 8.0). The mixture was incubated at room temperature for 10 minutes, and then washed with Buffer C. This was repeated until the RNA was extended to the desired length. For termination and release of the transcribing RNA, then all four NTP's were added for the remaining sequence length (20 µM per position) in addition to 150 µg/mL heparin to prevent reinitiation. The sample was incubated for 12 minutes at room temperature and then collected. Following this, 10 µL of transcription + 2 µL of 50% glycerol was loaded onto an ethidium-bromide-containing 2% native agarose gel (tris-borate) that was pre-chilled on ice, and then ran at 100 V for 60 minutes. The gel was then visualized using the Molecular Imager Gel Doc XR+ System (Bio-Rad). Image quality was optimized using the Image Lab software (Bio-Rad).

Polysome IP

Note: This is taken from a manuscript in the process of submission.

For IP, Dynabeads Protein G (30 µL) (Invitrogen) were washed two times in 10-bed volumes of cyto lysis buffer (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 2

mM EDTA, 1% NP-40) and antibodies in 10-bed volume of IP lysis buffer (20 mM TrisHCl [pH 7.4], 3 mM MgCl2, 150 mM NaCl, 2 mM DTT, 1× protease inhibitor cocktail EDTA-free, 5 µL/mL RNase Out, 0.2 M sucrose, 0.5% NP-40 and 0.1% Triton X-100) with 1 mM BSA were incubated 45 min at room temperature. Bead-antibody complexes were washed in IP wash buffer (20 mM Tris-HCl [pH 7.4], 300 mM NaCl, 0.5% NP-40) and incubated with 300 µg of cell lysate at 4°C for 2 h with rotation. Immune complexes were washed in IP wash buffer four times and precipitates were collected by boiling with 1× SDS sample buffer or isolated in TRIzol-LS. Total RNA or cytoplasmic RNA or co-precipitated RNA by anti-NCBP1 or eIF4E or RHA was resuspended in ice-cold TMG IP buffer (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 0.1% NP-40, 5 µL/mL RNase Out). TMG-antibody-bead complexes were washed and incubated with 5 µg of RNA or all RNA-IP at 4°C for 4 h with rotation, washed and isolated in TRIzol-LS. As described previously, cDNA was generated using Omniscript (Qiagen), random primers (Invitrogen) and cellular or cytoplasmic RNA (2 µg) or coprecipitated RNA samples(87) followed by RT-qPCR using gene-specific primers.

RNA:RHA EMSA

Note: This is taken from a manuscript in the process of submission.

The PBS-segment (nt 125-223 with two nonnative G-C pairs) and the A140C mutant RNAs were synthesized by in vitro T7 transcription, and purified by sequencing gel electrophoresis, elutrap elution and salt washes. For EMSA, RNA in 10 mM Tris-HCl, pH 7.5, was at 95 °C for 3 min, iced, mixed with buffer to reach final concentration of 10 mM Tris-HCl, pH 7.5, 140 mM KCl and 1 mM

MgCl2 (buffer A), and incubate at 37 °C for 30 min. Prefolded RNA was then mixed with recombinant RHA in buffer A (0.5 µM) and protein concentrations of 0 to 4.0 µM. The mixtures were incubated at room temperature for 30 min and resolved on a 1.5% agarose gel ran at a constant voltage of 150 Volts. Each EMSA experiment was repeated three times and the representative EMSA gels are shown. The intensities of the bands in each gel were quantified in ImageJ. Recombinant RHA containing a Mocr-tag and His-tag at the Nterminus was expressed in insect cells and purified as previously described (88).

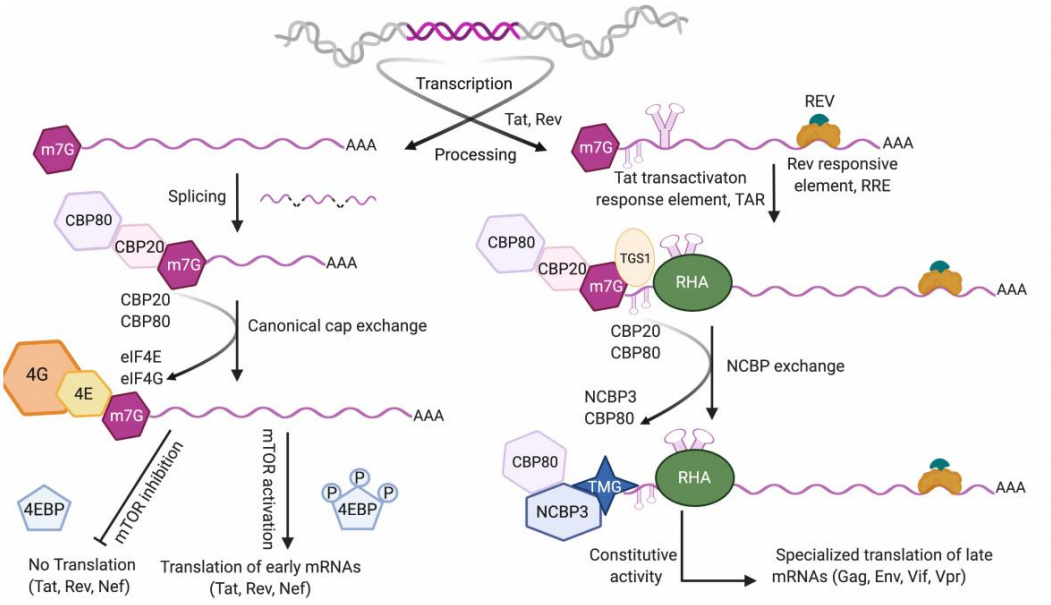

Figure 2-1. HIV-1 transcripts engage in two distinct cap-dependent translation pathways. Left panel. mTOR regulates translation of HIV Tat, Rev, and Nef. The m7G-cap of unspliced HIV RNA is co-transcriptionally bound by cap-binding complex (CBC) composed of CBP20-CBP80. Completely processed alternatively spliced mRNAs undergo canonical cap-exchange to eIF4E. Activity of eIF4E is activated by phosphorylated 4EBP, a substrate of serine-threonine kinase, mTOR. Inhibited mTOR results in hypophosphorylated 4EBP and downregulated eIF4E-dependent translation of host and viral regulatory proteins. **Right panel.** The translation of HIV structural and accessory proteins is not affected by mTOR. HIV late gene expression requires Tat and Rev trans-activation at Tat trans-activation response element (TAR) and Rev responsive element (RRE), respectively. Rev/RRE activate intron retention, and TGS1 binding 7mG of unspliced and incompletely spliced viral transcripts. Nuclear RNA helicase A (RHA) bound to 5'-UTR structure and TGS1 hypermethylation catalyzes CBC

exchange to NCBP3-CBP80. NCBP3-CBP80/RHA mRNPs support specialized translation of HIV late mRNAs unaffected by mTOR.

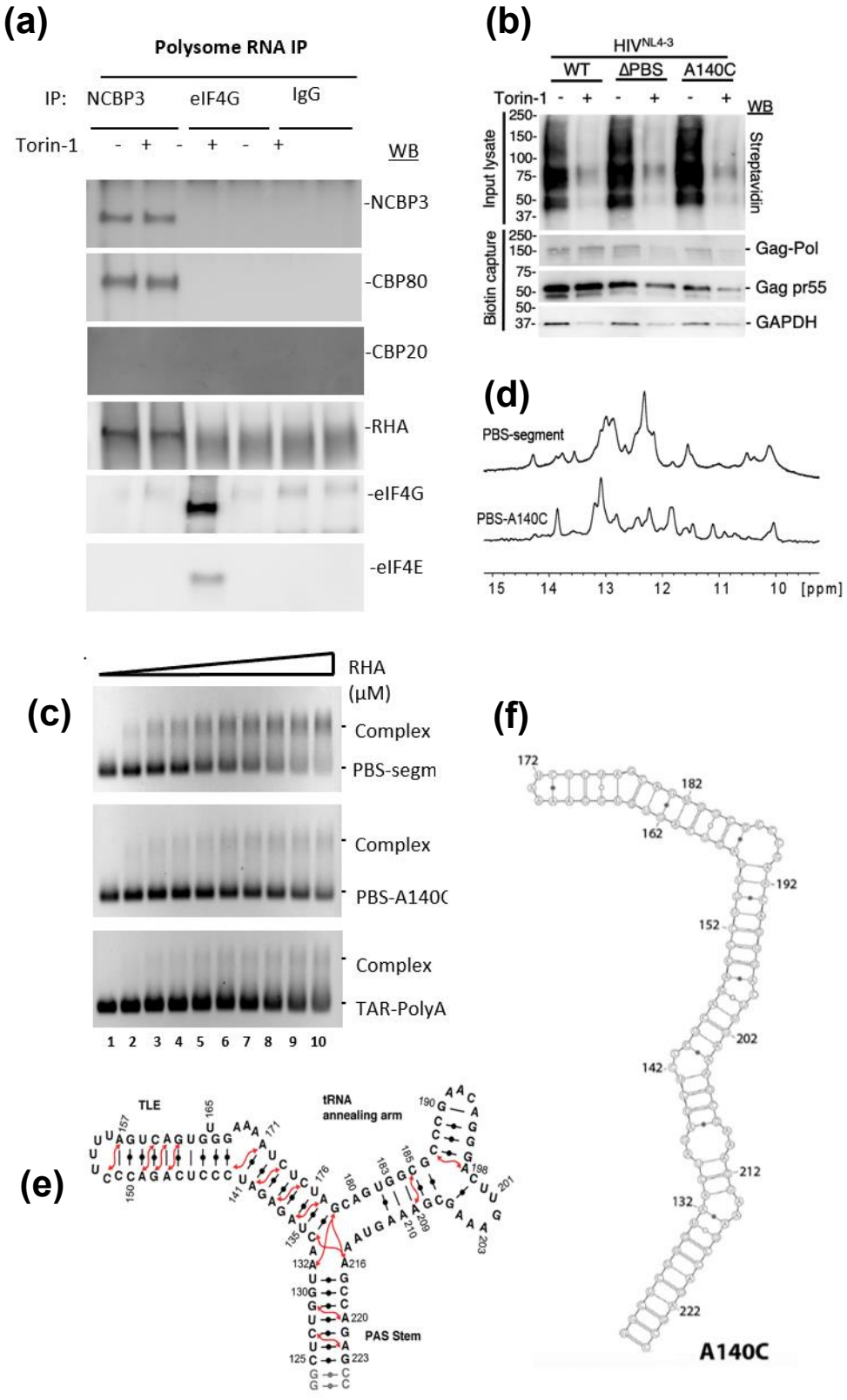
(a)
Polysome RNA IP
IP: NCBP3 eIF4G IgG
Torin-1 - + - + - +
WB
-NCBP3
-CBP80
-CBP20
-RHA
-eIF4G
-eIF4E
(b)
HIV NL4-3
WT ΔPBS A140C
Torin-1 - + - + - +
WB
Input lysate
Streptavidin
Biotin capture
- Gag-Pol
- Gag pr55
- GAPDH
(d)
PBS-segment
PBS-A140C
15 14 13 12 11 10 [ppm]
(c)
RHA (μM)
Complex
PBS-segm
Complex
PBS-A140C
Complex
TAR-PolyA
1 2 3 4 5 6 7 8 9 10
(f)
A140C
(e)
TLE
tRNA annealing arm
PAS Stem

Figure 2-2. Different HIV-1 5'UTR PBS Structures Affect RHA Binding and RHA is Associated to the Non-Canonical HIV-1 Translation Pathway. (a) Immunoprecipitation of NCBP3 (non-canonical mRNP), eIF4G (canonical mRNP), or IgG (control). **(b)** Translation of Gag became sensitive to Torin-1 treatment when the PBS structure was deleted (PBS) or disrupted (A140C). **(c)** Titration of RHA to HIV-1 5'UTR RNA's. RHA complexes with PBS three-way junction structure (top) and not PBS-A140C (middle) or upstream elements TAR, PolyA (bottom). **(d)** NMR spectra of 5'UTR PBS region, WT or A140C demonstrates differences in PBS structures. **(e)** PBS three-way junction secondary structure determined from previously published NMR and SAXs analysis (84). **(f)** Predicted PBS secondary structure with A140C mutation (84).

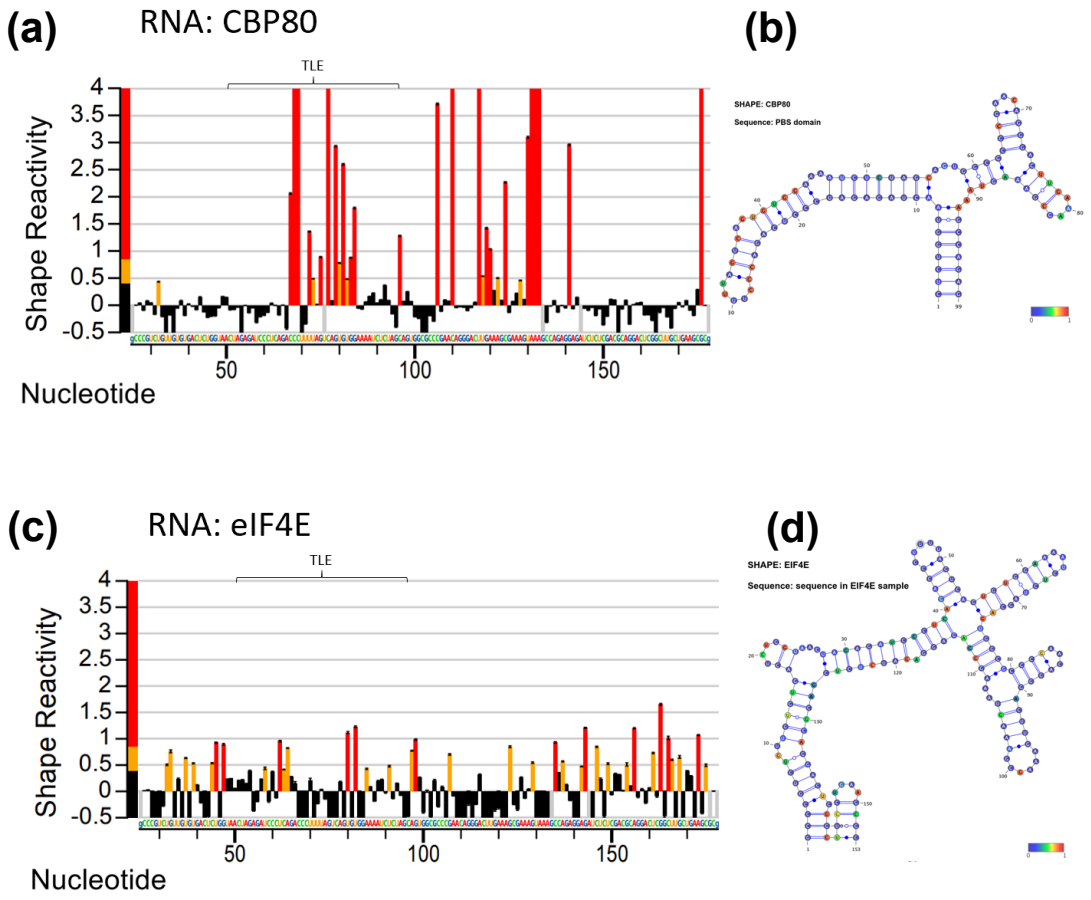

Figure 2-3. SHAPE Analysis of RNA Extracted from CBP80 and eIF4E IP Suggest Alternative HIV-1 5'UTR PBS Folding. (a) SHAPE reactivity profile of RNA purified from mRNP of CBP80 IP. Mutational frequency was high in the TLE stem and other elements of PBS similar to the three-way junction model. **(b)** Folding prediction of PBS based on SHAPE reactivity (a) suggests a three-way junction PBS in the non-canonical translation pathway. **(c)** SHAPE reactivity profile of RNA purified from mRNP of eIF4E IP. Mutational frequency was dissimilar to CBP80 RNA profile (a), suggesting alternative PBS structure. **(d)** Folding prediction of PBS based on SHAPE reactivity (b) generated a secondary structure lacking classical features of the three-way junction PBS, such as the structured TLE stem.

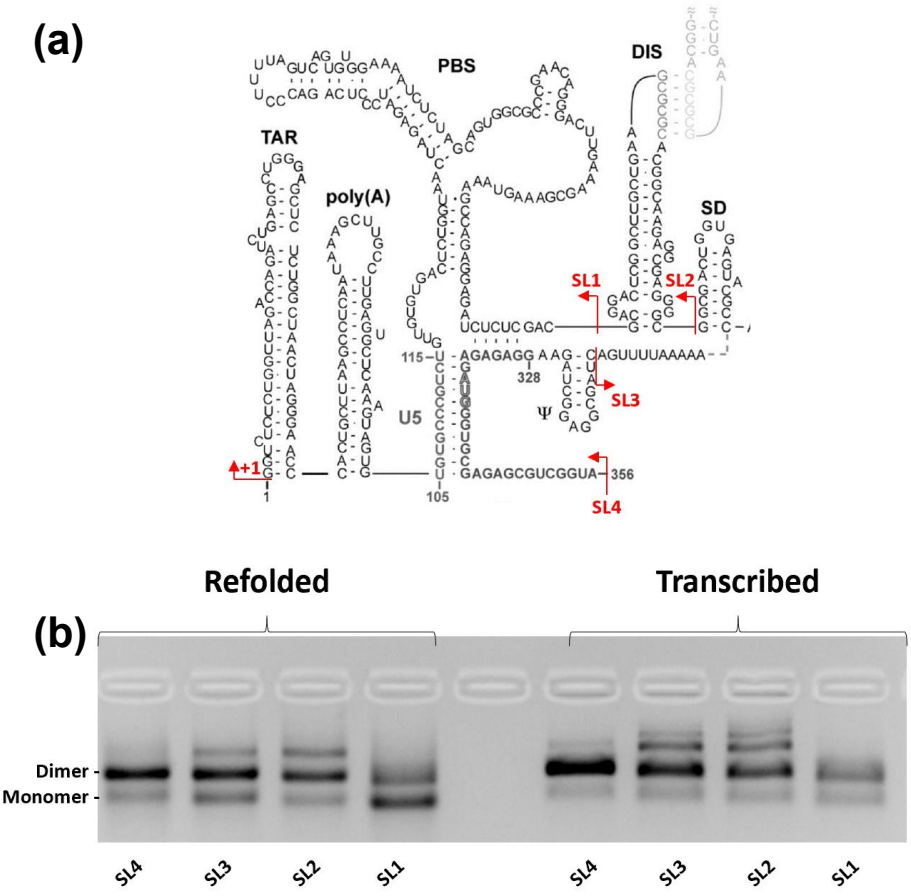

Figure 2-4. Transcribed 3' Truncations of HIV-1 5'UTR RNA Have a Different Monomer: Dimer Ratio When Refolded. (a) Template start and stop locations of transcribed RNA's. All RNA's were transcribed from the start of the 5'UTR (TAR), and the 3' end of the template corresponds to where indicated. **(b)** *In vitro* transcription of HIV-1 5'UTR RNA (right) with 3' truncations proceeding the relevant stem loop (left to right: SL4: full length; SL3: Ψ; SL2: SD; SL1 DIS) with T7 polymerase produces a mixed population of monomer and dimerized RNA'. In contrast, refolding these RNA's (left) produces a different distribution of monomer and dimer RNA, suggesting the folding pathway of transcribing 5'UTR RNA's proceed in a 5'->3' fashion.

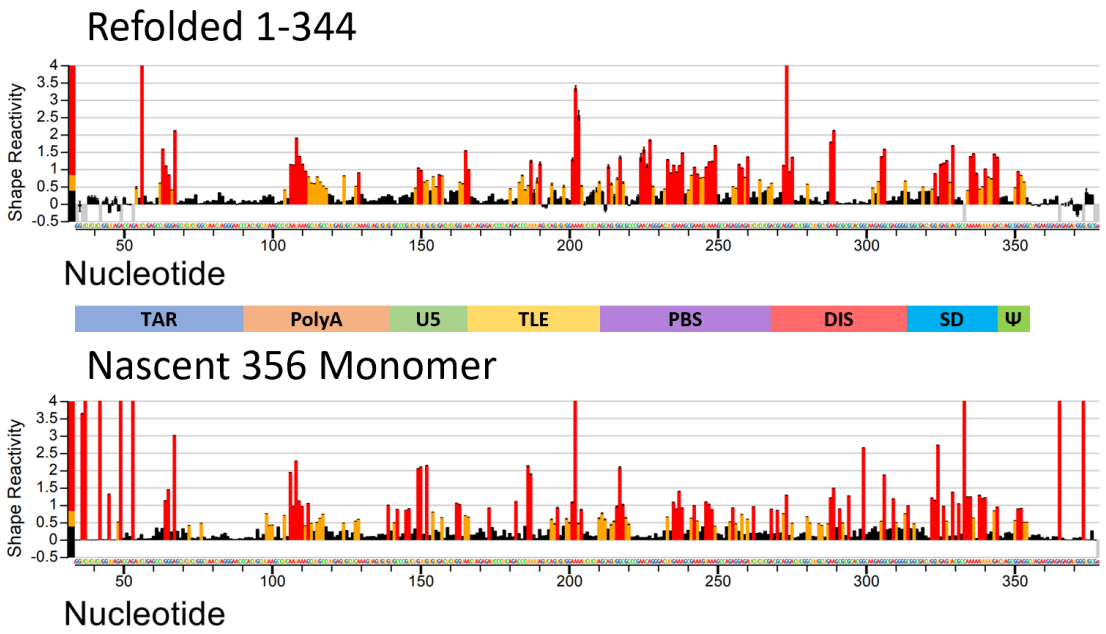

Figure 2-5. SHAPE Analysis of HIV-1 5'UTR Refolded 1-344 and Nascently-Transcribed WT Monomer. The 1-344 RNA was refolded and then treated with 5NIA, while the "nascent monomer" was transcribed, treated, then ran on a native agarose (excision of the monomer band). The generated SHAPE reactivity profiles reveal structural differences across the entire UTR of the gel-extracted monomer and refolded 1-344 RNA's, including TAR, PolyA, the TLE stem, and PBS.

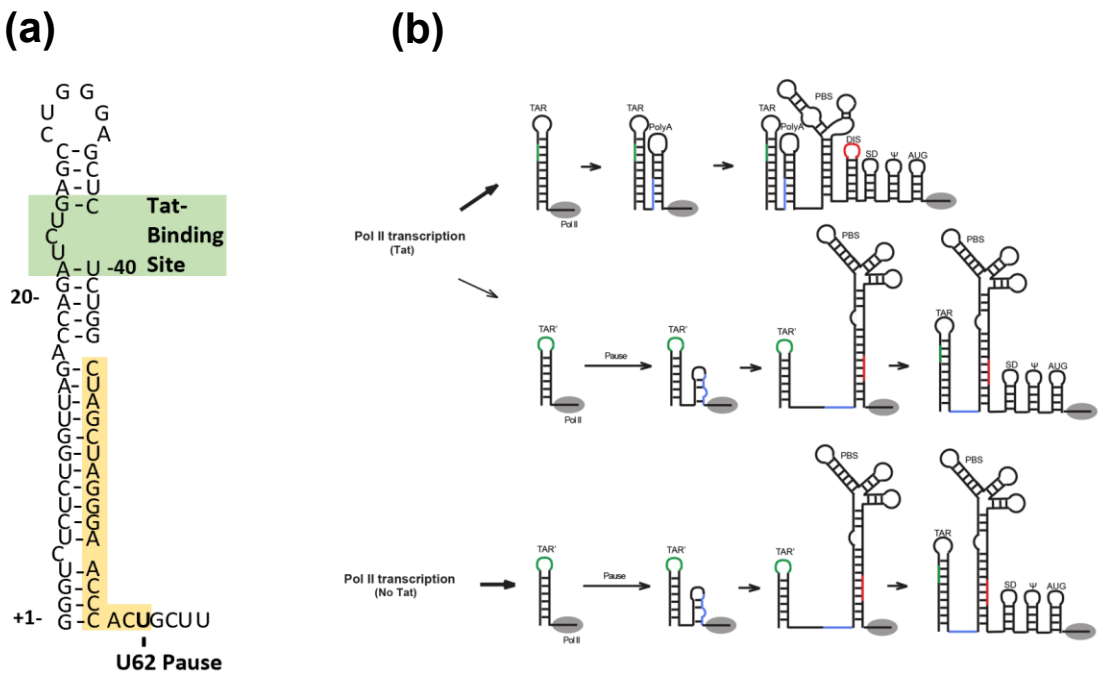

Figure 2-6. Folding of Transcribing HIV-1 5'UTR is Dependent on Structure of Upstream Elements TAR and PolyA Mediated By Tat Binding. (a) Elongated TAR RNA stem secondary structure with Tat-binding site (green) and region covered by exit tunnel of RNA Polymerase II (yellow) during transcription pausing at position U62. **(b)** Model of two folding pathways of transcribing 5'UTR RNA, influenced by binding of Tat to TAR. In the absence of Tat, TAR is predicted to form a shorter hairpin, with its 3'-end creating a small hairpin with the following PolyA sequence (top). This prevents PolyA base-pairing to the lower stem of PBS, leading to a distinct PBS structure and occlusion of an available DIS for dimerization (red). U62 pausing in the presence of Tat binding TAR (green) alters the folding pathway, preventing TAR from structuring from PolyA, leading to an alternate PBS structure and exposure of the DIS stem loop for dimerization.

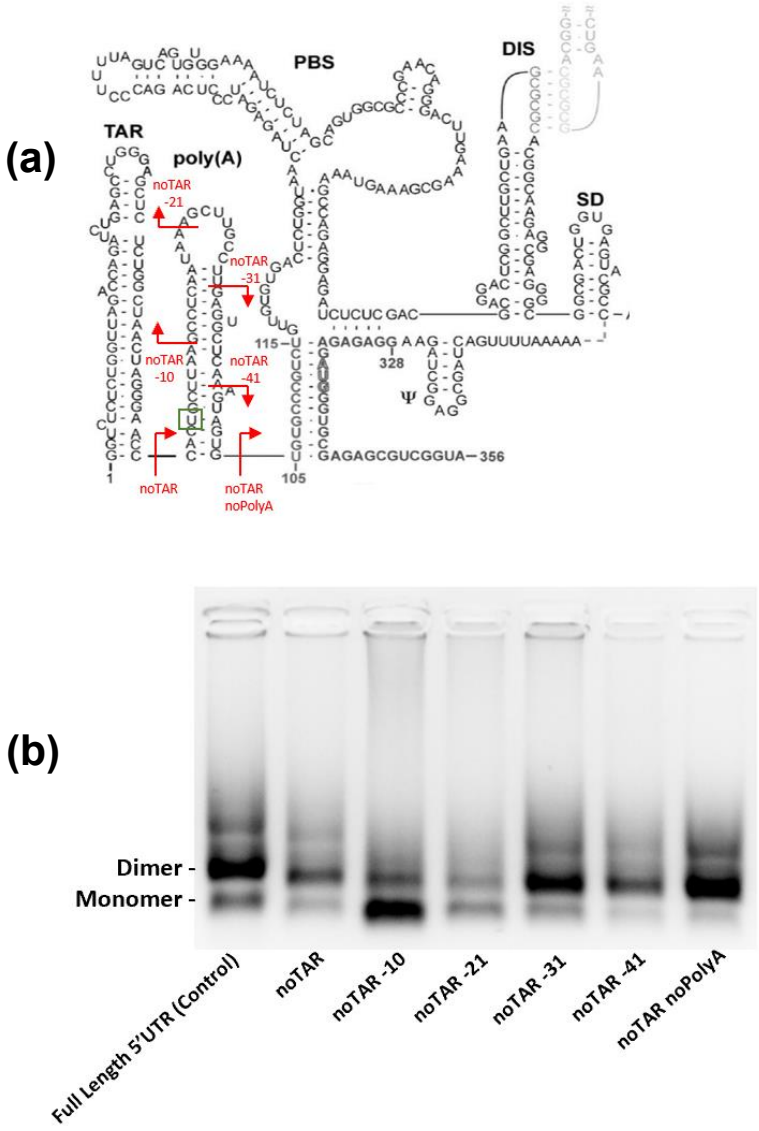

Figure 2-7. 5'Truncation of HIV-1 5'UTR RNA Alters the Dimerization of Transcribing RNA. (a) The HIV-1 5'UTR secondary structure with markers indicating the position of transcription initiation for (b). Position of transcriptional pausing, U62, is boxed in green. **(b)** *In vitro* transcription by T7 polymerase initiated from different positions of the 5'UTR affects dimerization, suggesting folding of transcribing RNAs occurs in a 5'->3' fashion. Sample noTAR -10, which begins after U62 (pause), produces a majority of monomeric RNA.

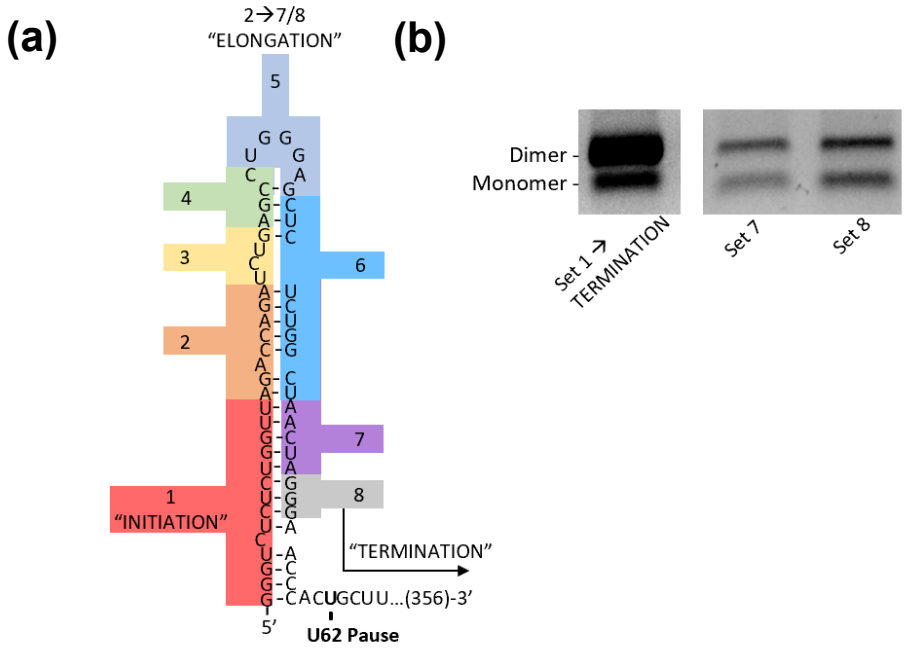

Figure 2-8. Controlled Elongation of Transcribing HIV'1 5'UTR TAR Affects Downstream Folding and Dimerization. (a) Biotin-5'-template was annealed to streptavidin-agarose beads and incubated with T7 polymerase and Set 1 NTP's (GCU; initiation). Each NTP set was added in a stepwise-fashion with washing between each addition to control elongation of the transcribing RNA. Following Set 7 or 8, the 4 NTP's (GACU) were added simultaneously (termination) alongside heparin (to prevent reinitiation) and the freed RNA was ran on a native agarose gel. **(b)** 5'UTR RNA on native gel with no elongation (left), termination after Set 7 (middle), and termination after Set 8 (right). Higher concentration with termination following Set 8 mimics the U62 transcriptional pulse by RNA Polymerase II *in vivo,* suggesting that transcriptional pausing influences folding of the TAR motif, which has downstream structural consequences, including 5'UTR dimerization.

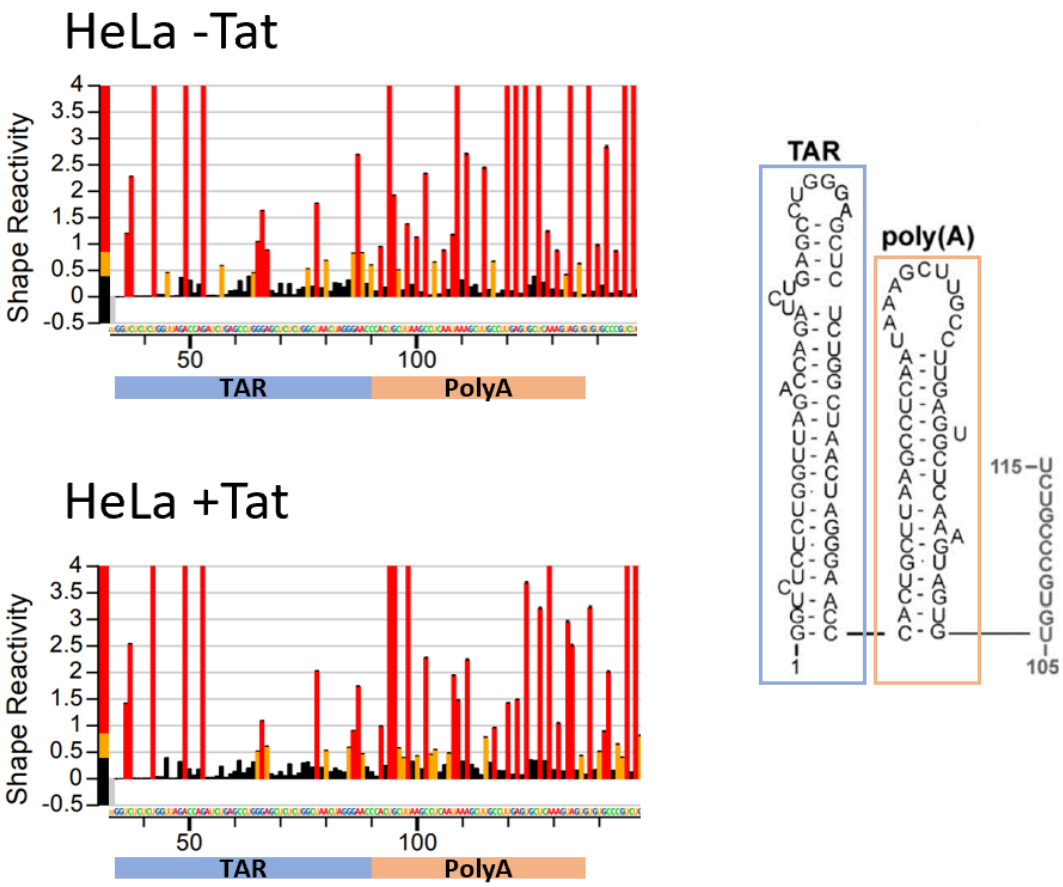

Figure 2-9. SHAPE Analysis of HIV-1 UTR in HeLa Nuclear Extract Reveals Difference in PolyA Structure in the Presence of Tat. The HIV-1 5'UTR was transcribed in HeLa nuclear extract in the presence or absence of Tat under the control of the native HIV-1 5'promoter. The SHAPE reactivity profiles demonstrate differences in mutational frequency in the PolyA region (orange), indicating difference in PolyA structure upon labeling. The positions shown on the reactivity profile are shown in the secondary structure orientation (right) and color-coded. The most pronounced differences are the nucleotides composing the leading nucleotides of the base-paired PolyA stem (left-half of PolyA stem).

CHAPTER 3. INI1a:TAR Complex: Rational Design of INI1a Mutations to Investigate the Virus:Host Interaction in HIV-1 Replication

Note: This chapter contains work for a research paper in the process of submission.

Abstract

INI1 is a core component of the SWI/SNF eukaryotic chromatin remodeling complex. It interacts with HIV-1 IN and is implicated in multiple stages of HIV -1 replication, including virion assembly, transcription, post-transcriptional RNA stability, and protein stability. It is led by an N-terminal winged helix DNA binding domain (WHD) and contains two conserved, imperfect repeat regions: Rpt1 and Rpt2. Through collaboration with Ganjam Kalpana laboratory, interaction between WHD and HIV-1 TAR RNA was identified, but the molecular details and the biological function of the interaction were unclear. NMR studies of the WHD:TAR interaction were carried out and the binding interface was mapped. The NMR-derived restraints were used to filter the complex models generated by molecular docking. Several key residues in INI1a for specific TAR RNA binding were identified. Mutation of these residues resulted in deficient TAR RNA binding without affecting INI1a's interaction with DNA. HIV-1 replication in cells expressing INI1a mutants was severely hindered as abortive viral mRNA transcription was observed and significantly less virion particles were produced. The structural

studies of INI1a:TAR enable rationale design of mutations to further mechanistic investigation of the virus:host interactions in virus replication events, and provide structural basis for the development of novel antiviral therapeutics.

Introduction

The HIV-1 infectious cycle utilizes multiple host factors at various stages of infection, such as RNA polymerase II in RNA synthesis, the spliceosome for splicing, and the ribosome for translation (1). HIV research has progressed substantially since the mid-1980's, but the mapping of the virusl:host interactome remains incomplete. The breadth of understanding of host-virus interactions continues to grow to establish a wholistic model of the retrovirus infection.

SMARCB1, also known as INI1 (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily B, Member 1; Integrase Interactor 1), is an essential component in the SWI/SNF eukaryotic chromatin remodeling complex, and is involved in a range of host process, including transcription, cell cycle development, and tumor suppression (89, 90). INI1 has been demonstrated to interact with host and viral proteins via two consecutive, highly conserved, imperfect repeat regions: Rpt1 and Rpt2. The Rpt1 domain (aa 183 – 248) is connected to the Rpt2 domain (aa 259 – 319) by a linker region (aa 249 – 258) (91–97) (Figure 3-1a).

Due to its role in host chromatin remodeling, INI1 was first identified as a binding partner for HIV-1 integrase (IN) (98). It was demonstrated that IN-INI1 interaction enhances HIV-1 genome integration into host nucleosomal targets. Subsequent studies have identified INI1's role in additional, multiple stages of HIV -1 replication, including virion assembly, transcription, post-transcriptional RNA stability, and protein stability (91, 99–112). HIV-1 IN binds INI1 via the Rpt1 domain, and the Rpt2 domain is not required (113). The Rpt1 domain is also

thought to be required for late-stage infection, as its trans-added inclusion has been shown to inhibit particle production by binding to the IN sequence within GagPol polyprotein (114). Recent work has suggested that INI1 specifically binds to the HIV-1 IN-CTD via its Rpt1 domain (PDB: 6AX5) (115). This study also found that the IN residues at the interface of the IN-INI1 complex were the same as those used for IN to bind HIV-1 TAR RNA (Figure 3-1b) (116). This revealed a remarkable similarity of the surface shape and charge between INI1-Rpt1 and HIV-1 Tar RNA. This would be the first example of an HIV-1 RNA mimicking a host protein's surface for hijacking host processes for viral lifecycle.

The N-terminal winged helix DNA binding domain (WHD, aa 1-110) of INI1 contains a series of basic amino acids that interact with dsDNA. Mutations within the DNA binding site has been identified in patients with familial schwannomatosis (117), demonstrating the essential dsDNA binding role of WHD in INI1. Preliminary work from our collaborators (Ganjam Kalpana Lab; Albert Einstein Institute) has shown that the INI1-WHD (aa 1-110) is capable of binding INI1-Rpt1 domain. As shown in (Figure 3-1c), the interaction between WHD and RPT1 was characterized in an Alpha assay, where the WHD and RPT1 were bound to either a donor or acceptor conjugate for fluorescent output. Because INI1-WHD can bind Rpt1, Rpt1 surface mimics Tar RNA, and HIV-1 IN binds INI1-Rpt1 and HIV-1 Tar RNA, this suggests the possibility that INI1 can bind HIV-1 TAR RNA via WHD (Figure 3-1a). To address this hypothesis, HIV-1 TAR RNA, as well as other regions of the viral RNA (nt 237-279) and scrambled RNAs, were added to the WHD:RPT1 complex in the alpha assays. An initial alpha assay with full length INI1 bound to

Rpt1 was competed against the viral RNA (nt 237-279), TAR RNA, and scrambled RNAs (Figure 3-2a). Only the addition of TAR RNA was found to be capable to compete with RPT1 binding of INI1, indicating the presence of a TAR-binding motif within INI1. To confirm that INI1 binds to TAR RNA, INI1:TAR EMSA (Figure 3-2b) was performed in the presence of unlabeled RNAs including TAR, Scrambled viral RNA (nt 237-279). Only the addition of TAR was capable of reducing the fluorescent signal of the initial INI1:TAR complex, suggesting binding. These data demonstrate that INI1 specifically binds to HIV-1 TAR RNA. To further identify the TAR binding binding domain in INI1, a series of INI1 truncations were cloned, expressed and purified (Figure 3-3a). Gel shift assays and Alpha assays reveal that the N-terminal WHD domain is necessary for TAR RNA binding, and INI1 mutant proteins lacking aa 1-110 are deficient in TAR RNA binding (Figure 3-3b and Figure 3-3c).

INI1 contains two isoforms, INI1a and INI1b, which are differentiated by a 9 amino acid deletion in the WHD domain of INI1b by alternative splicing (Figure 3-4a) (118). This deletion is located in a flexible loop region between helix three and strand four (119), and homology modeling of INI1b suggest that INI1b adopts the an almost identical structure as INI1a except lacking the flexible loop (Figure 3-4b, Figure 3-4c, and Figure 3-4d). INI1b protein was expressed and purified (Fig. 3-4e), and its interaction with HIV-1 TAR RNA was characterized by gel shift assays. Titration of TAR to both isoforms pre-incubated with nucleic acid polymer demonstrated that INI1a has a higher affinity for TAR compared to INI1b, suggesting that the loop (aa 69-77) is critical for this binding.

It currently remains unknown how the DNA binding domain INI1a-WHD is able to specifically bind the TAR RNA. It is well-established that the arginine residues of Tat interact with the bulge of TAR and form an arginine sandwich (120), so it is plausible that INI1a my bind TAR in a similar way. In this chapter we characterized the interaction between INI1a and HIV-1 TAR RNA and map the binding interface by solution NMR. Our data indicate that the binding mode of INI1a binding to TAR is different from the Tat:TAR interaction. Key interface residues were identified to guide mutagenesis studies to investigate the function of INI1a:TAR interaction in HIV-1 replication.

Results

Mapping the RNA binding interface on INI1a WHD.

To investigate interaction between the INI1a and HIV-1 TAR, the GST-tagged WHD domain (aa 1-110) was expressed in an *E. coli* vector and then purified, which included cleavage of the glutathione tag. HIV-1 5'-UTR TAR RNA (nt 1 – 56) was transcribed *in vitro* and purified. The INI1a was determined to bind the purified TAR RNA by gel-shift (not shown).

The HIV-1 TAR RNA was increasingly titrated to an equimolar ratio to the INI1a protein and ^{1}H-^{15}N TROSY NMR data were collected throughout the titration (Figure 3-5a). The assignment of the amide protons was made by assigning triple resonance spectra of the ^{15}N labeled INI1a WHD and referencing with the previously published chemical shifts (BMRB entry 26553). Several lysine residues were too weak for detection and many residues were broaden out upon RNA titration. Nevertheless, a number of peaks exhibited site-specific shifts during the

titration, suggesting a specific protein:RNA interaction. The chemical shift perturbation of INI1a residues was quantified and summarized (Figure 3-5b). Several significant perturbations at or near presumed basic residues for binding were resolved, including K8, T9, S30, R40, G41, R46 and Y47 (near K45). We also saw perturbations in T72 and T76, which are a position away from basic residues K73 and K77, respectively. This data indicated a number of basic residues that would be part of the binding interaction.

Mapping the INI1a WHD binding interface on HIV-1 TAR RNA.

NMR assignments of TAR RNA with equimolar INI1a was also performed (Figure 3-6a) to map the RNA residues near or on the protein binding interface. Chemical shift perturbations were observed for residues on the top of the TAR hairpin, including residues A19, A26, G27, C28, A34, G35, C38, C39, U40, G41, G42, C43, and U44 (Figure 3-6b). The chemical shifts of residues on the bottom of the TAR hairpin remained relatively unaffected upon INI1a WHD binding. These data suggest that WHD interacts with the top stem of TAR.

Molecular docking of INI1a:TAR complex

Through collaboration with Xiaoqin Zou laboratory (University of Missouri), the atomic structure of INI1a:TAR complex was modeled through molecular docking on the Mockup server. The solution NMR structures of INI1a (PDB code: 5AJ1) and TAR RNA (PDB code: 1ANR) molecule were downloaded from the Protein Data Bank (PDB) and used for docking. The NMR titration data was used to develop a preliminary docking model of INI1a to TAR based on a previously published model (Figure 3-7; PDB: 5AJ1). The lowest energy conformer of TAR

RNA (i.e., the one used in our previous publication) was used for docking. In contrast, all the 20 conformers of INI1a were used for docking due to the presence of a flexible loop spanning residues 68 - 87. The structures were filtered by NMR-derived distance restraints that the protein and RNA residues underwent significant chemical shift perturbations are near the interface. The rotamers of LYS8, LYS45, LYS73, LYS77 and ARG40 were adjusted to optimize their specific interactions with TAR RNA, considering that the NMR structure used for docking is not the native binding structure and that local conformational changes giving rise to favorable interactions could be induced upon binding.

Rational designed INI1a mutations abolished specific interaction with TAR RNA but maintained normal DNA binding affinity.

Based on the NMR data and preliminary docking model, sensible mutations were made for several basic residues within INI1a-WHD predicted to be key for binding. Using alpha assay of WHD and TAR, we found that mutations K8A, R40A, K45A, K73A, and K77A destroyed binding capability of INI1a to TAR, while mutants K13A and K61A only slightly reduced binding affinity below WT WHD (Figure 3-8a and Figure 3-8b). Furthermore, EMSA of INI1a with fluorophore-linked TAR showed similar results (Figure 3-8c). Mutants K13A and K61A did not appear to significantly affect binding of TAR. In contrast, R40A appeared to completely destroy binding capability, and K8A, K45A, K73A, and K77A mutants significantly reduced binding. Additionally, EMSA of fluorophore-linked DNA with INI1a showed that its DNA-binding capabilities was not significantly affected with any these mutations (Figure 3-8d). With this data, we then proceeded to generate a full

docking model (Figure 3-9), which better depicts the molecular interactions between INI1a-WHD and TAR.

Disruption of INI1a:TAR interaction led to abortive transcription and reduced production of viral particles.

To begin investigating the role of INI1-TAR interaction, monocyte cells deficient in INI1 (ΔINI1) were co-transfected with pNL4-3 and INI1-containing plasmid under constitutive expression. Using RT-PCR, the expression levels of viral mRNA's were measured (Figure 3-10a). The level of proviral expression was found to be consistent across all mutants (Figure 3-10b), as were levels of initial transcription products, indicated by TAR (Figure 3-10c). However, levels of unspliced transcripts (Figure 3-10d), and transcripts corresponding to singly-spliced and multiply-spliced transcripts (Figure 3-10e, Figure 3-10f, Figure 3-10g, Figure 3-10h) were found to be significantly diminished in the INI1a mutants demonstrated to be deficient at binding TAR (K73A, R40A, K77A, K8A, K45A). This suggests that INI1 binding to TAR is required for efficient transcriptional elongation from the 5'LTR promoter.

Due to the deficiency of transcription elongation, the viral replication is expected to be impaired, and the virus particle production was measured. As above, monocyte cells deficient in INI1 (ΔINI1) were co-transfected with pNL4-3 and INI1-containing plasmid under constitutive expression. After some time, the cells and supernatant were collected. All INI1 mutants were found to be expressed at similar levels by western blot (Figure 3-11a). An ELISA was performed to determine levels of CA (p24), corresponding to the amount of particle production.

It was found that INI1a mutants demonstrated to be deficient at binding TAR (K73A, R40A, K77A, K8A, K45A) also had greatly reduced levels of virion production (Figure 3-11b). Together, these data demonstrate the critical role of INI1a:TAR interaction in HIV-1 replication.

Discussion

Our work has uncovered the first example of an HIV-1 RNA structure (TAR), mimicking the protein surface of a host protein (INI1a-Rpt1). The mapping this binding interface will provide valuable insights for developing novel therapeutics to target this interaction, as the work has demonstrated the necessity of INI1a binding of TAR for both transcriptional elongation and virion production. The wholistic work demonstrates the ability of INI1a-WHD to bind HIV IN, INI1a-Rpt1, and HIV TAR RNA. The role of this interplay is not yet fully understood. Because INI1a-Rpt1 binds TAR RNA and HIV IN in the same manner, it could act as a placeholder for IN, preventing steric hindrance of proper encapsidation for ordered assembly. It is possible that INI1a may aid in transcriptional elongation in a manner similar to Tat, as they both bind TAR. We speculate that INI1a may play a role in Tat transactivation of elongation for transcriptional enhancement. INI1a aid in the recruitment of PTEF-b from the 7sK snRNP, aid in the association of Tat to PTEF-b, or be part of the transactivation complex formed during the U62 pause.

NMR analysis of titration of TAR RNA to INI1a yielded valuable data to aid in the modeling of the interaction. However, we suffered a number of difficulties while performing this interaction. The INI1a:TAR complex solution was very viscous, and the viscosity of the complex increased as more TAR RNA was titrated

into the sample. In addition, the complex precipitated out of solution as more RNA was added. These problems can be mitigated in multiple ways, such as altering pH and increasing salt concentration. Unfortunately, NMR data quality decreases with increasing salt concentration. We had trouble resolving a number of key lysine and arginine residues, and it has been documented that lowering the pH of the sample can increase the signal of these residues during NMR analysis. However, we found that lowering the pH of our sample caused rapid precipitation of the sample, preventing it from being a viable method. Further NMR analysis would seek to mitigate these problems. One option is to add a solubility tag. GST, used to purify our protein, is a solubility tag. However, keeping the protein tag during NMR analysis has the side effect of significantly complicating the NMR spectrum. Collecting NMR data of the GST tag alone may allow the analyzer to assign these peaks out of the GST-INI1a spectrum, which might make data analysis possible.

Modeling of the INI1a:TAR interaction generated a sensible binding mode that corroborates the collected NMR data. We found significant shifts in the RNA spectra for residues A20, G27, A34, G35, C41, and U42 of the TAR RNA. This aligned with data collected from the protein spectra, where we saw shifts at or near the residues K8, K45, R40, K77, and K78. Additional NMR analysis of the protein would be ideal for further refinement of the model. Our model shows that the binding mode of INI1a-WHD to TAR differs from the arginine sandwich mode of Tat binding TAR.

Author Contributions

Rationale for work G.K. laboratory; NMR experimentation X.H, Z.S, S.S.; All biological and binding data U.D., G.K. laboratory; Protein Purification advice Z.S.; NMR collection methods R.K., S.VD.; Computational modeling X.Z. laboratory

Experimental Procedures

Plasmids

Plasmids pNL4-3 (NIH AIDS Reagent Program) and pUC19 containing the HIV-1 5'UTR sequence for DNA template synthesis were generated using methods described previously (84). The pNL4-3 HIV-1 sequence was fused into the pUC19 sequence for amplification. Plasmid sequences were verified by Sanger sequencing (DNA Core, University of Missouri, Columbia, MO).

In Vitro RNA Transcription

RNAs used for refolding were transcribed using purified T7 polymerase. Corresponding DNA templates were generated via PCR and purified prior to transcription. Transcriptions were performed at 5-20 mL volumes using ~0.5-2 mg of purified DNA template, 40 mM Tris-HCl (pH = 8.1), 5 mM DTT, 1 mM spermidine, 0.01%Triton X-100, 5-10 mM NTP's, and 5-25 mM $MgCl_2$. Optimal concentrations of $MgCl_2$ and NTP's to maximize RNA yield was determined using small-scale reactions following by RNA staining on polyacrylamide gel. The transcriptions were incubated for 3 hours at 37C, and the reaction was then quenched using a Urea-EDTA mixture (7 M Urea pH = 8.0. 250 mM EDTA). The RNA was isolated on polyacrylamide denaturing gel by electrophoresis using FisherBiotech DNA

sequencing system at 16W overnight. The RNA bands were visualized via UV-shadowing, excised, and then eluted overnight using the Elutrap electroelution system (Whatman) at 100 V. The purified RNA fraction was washed with 2M NaCl and then desalted within the Amicon Ultra-4 Centrifugal Filter Device (Millipore). RNA concentrations were determined by measuring absorbance at 260 nm using the NanoDrop 2000c (Thermo Fisher Scientific).

Purification of INI1a

The GST-INI1a pGEX-3x plasmid (aa 1 – 110; N-terminal glutathione) was provided as a kind gift by Dr. Ganjam Kalpana's laboratory (Albert Einstein Institute; Boston, MA). The plasmid was transformed into *E. coli* BL21(DE3). Cells were grown at 37°C to an OD_{600} = 0.1 in 50 mL LB broth (121). Following this, the cells were spun by centrifugation. The pellet was added to either 1 L LR media (122) (24 g KH_2PO_4, 5 g NaOH, 200 µL 0.5 M $CaCl_2$, 2.2 mL 1M $MgSO_4$, 0.5 g $^{15}NH_4Cl$, 2.5 g ^{13}C-glucose; for protein NMR analysis) or 1 L LB broth (for RNA NMR analysis). The culture was grown to an OD_{600} = 0.8, then treated with 1 mM IPTG and expressed overnight at 16°C, 220 RPM. Cells were then centrifuged and resuspended in 40 mL PBS buffer. Following this, the homogenate was treated with lysozyme for 30 minutes at room temperature, and then phenylmethylsulfonyl fluoride (PMSF) for 30 minutes on ice. Sonication was performed with a Sonics Vibra-Cell. Following centrifugation, the supernatant was loaded onto a glutathione-agarose resin, washed with PBS buffer, and then eluted with PBS + 20 mM reduced glutathione. The eluant was concentrated with Amicon Ultra-4 Centrifugal Filter Device (Millipore), and then the GST tag was removed by

cleavage of Factor XA overnight at 4°C. The cleaved protein was purified using NGC Quest 10 Chromatography System (Bio-Rad).

NMR Experiments

All NMR spectra were obtained using a Bruker AVANCE III 800 MHz spectrometer equipped with a TCI cryoprobe (NMR Core, University of Missouri). Sequence specific and side-chain assignments of INI1a were obtained using standard triple resonance methods (HNCACB, HNCOCACB, HSQC) using 850 µM ^{13}C-^{15}N-INI1a in physiological salt buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, 140 mM KCl, 10 mM NaCl, 10% D_2O, pH = 5.5). For generation of ^{1}H-^{15}N HSQC spectra of titration of TAR RNA to INI1a, 400 µM of ^{15}N-INI1a was incubated in physiological salt buffer (10 mM Tris-HCl, 10 mM $MgCl_2$, 140 mM KCl, 10 mM NaCl, 10% D_2O, pH = 6.5) and then the sample was analyzed. Concentrated TAR RNA was added to a relative molar concentration (0% → 10% → 20% → 40% → 60% → 80% → 100%), incubated for 15 minutes at 37°C, then analyzed before addition of the next portion of TAR RNA. Data was processed using NMRPipe/NMRDraw (123) and analyzed using NMRFAM-Sparky (protein) (124) or NMRViewJ (RNA) (125).

Methods for Figure 3-1, 3-2, 3-3, 3-4, 3-7, 3-8, 3-9, 3-10, 3-11

At the time of writing, these methods were not made available. Many similar methods are outlined in the precursor paper to this work (115).

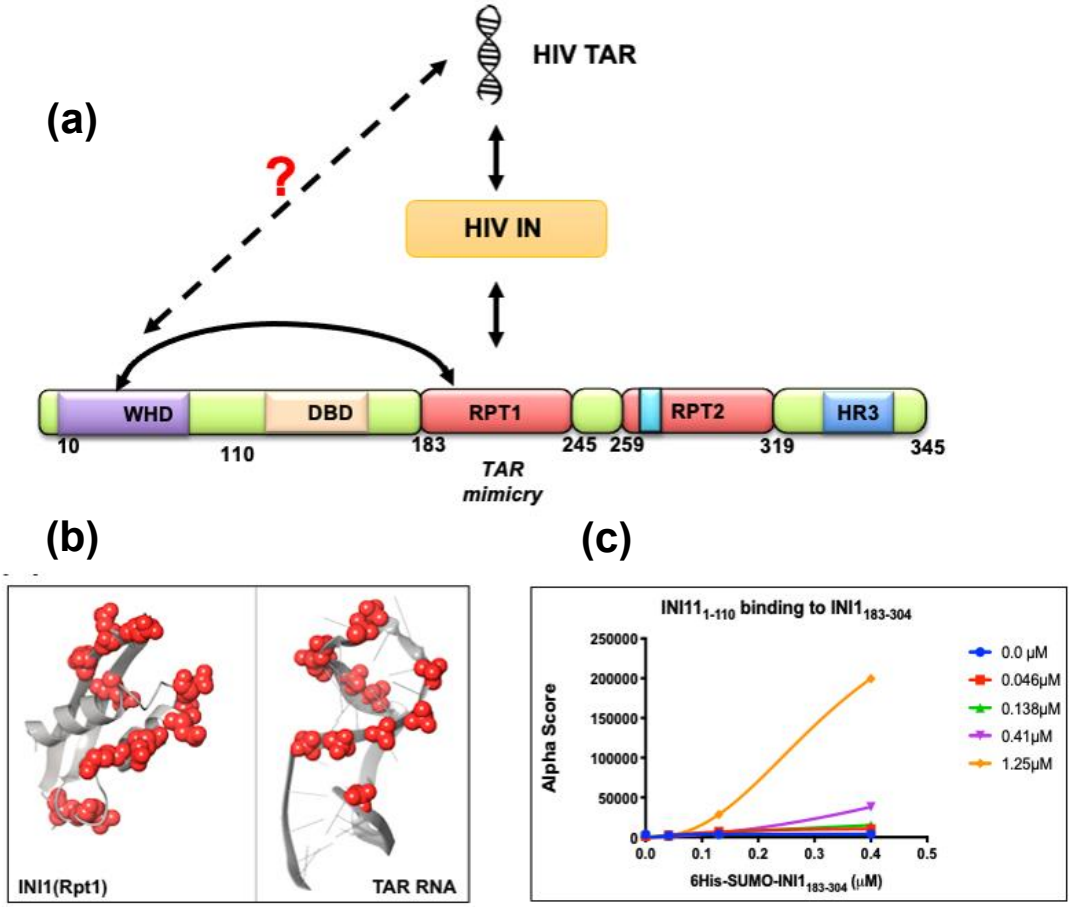

Figure 3-1. The SMARC1 (INI1) Protein Contains an Rpt1 Domain Whose Surface Mimics HIV-1 5'UTR TAR RNA. (a) The INI1 protein sequence contains an N-terminal winged helix domain (WHD), followed by a DNA-binding domain (DBD), followed by two consecutive, imperfect repeat domains (Rpt1 and Rpt2), and at the C-terminus is a coiled coil domain (homology region 3; HR3). RPT1 interacts with the HIV-1 IN, which binds HIV-1 TAR RNA for integration. RPT1 also interacts with the WHD. We therefore hypothesized that WHD interacts with HIV-1 TAR. **(b)** Mapping negative charged surface residues of the INI1-Rpt1 (red spheres) reveals similarity to spacing of phosphate backbone charges of

structured TAR RNA hairpin. Adapted from precursor publication of this work(115). **(c)** Titration of INI1 WHD (aa 1-110) to Rpt1 (aa 183-304) demonstrates binding between the two domains.

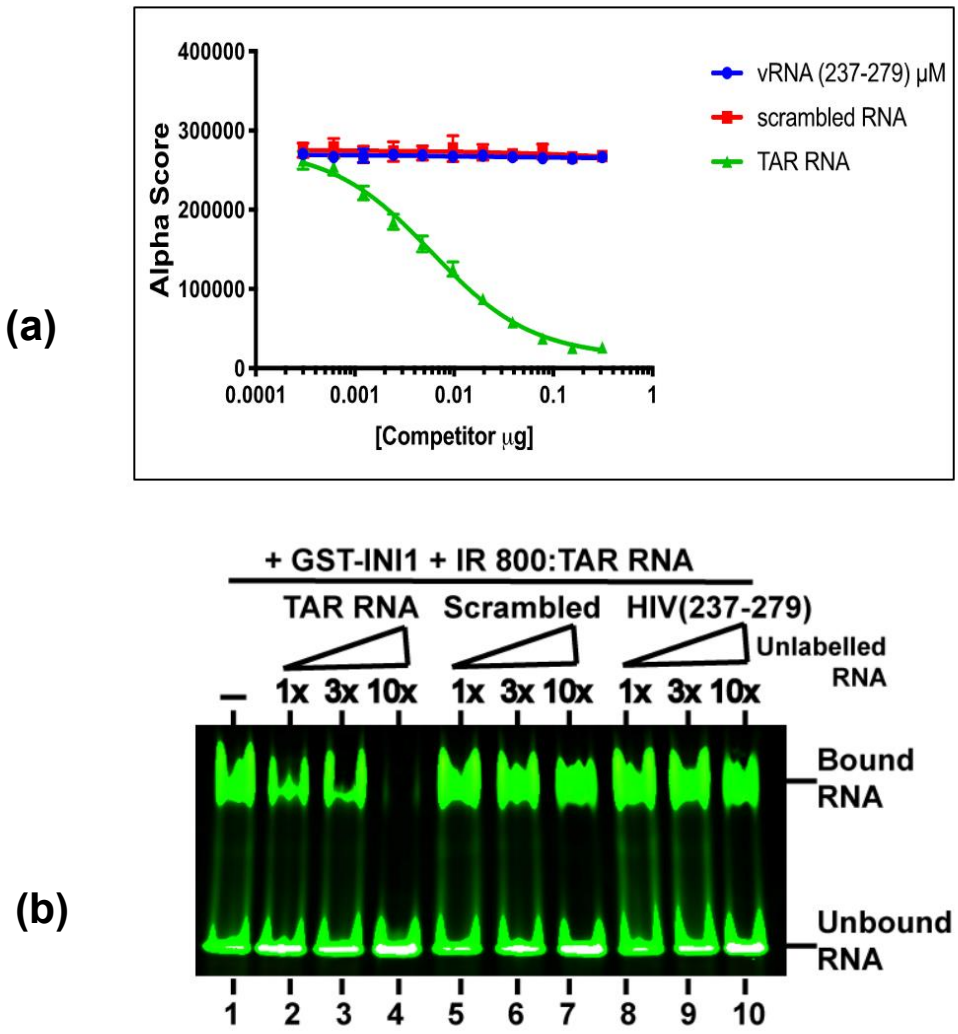

Figure 3-2. INI1 protein specifically interacts with the HIV-1 5'UTR TAR RNA. **(a)** Competitive binding of INI1 to TAR RNA or control RNA's (vRNA, scrambled RNA) reveal specific affinity of INI1 to the TAR RNA sequence. **(b)** Gel-shift assay of GST-INI1 to TAR RNA (left), scrambled RNA (middle) or HIV-1 5'UTR PBS (right) demonstrates specific affinity of INI1 to the TAR RNA sequence.

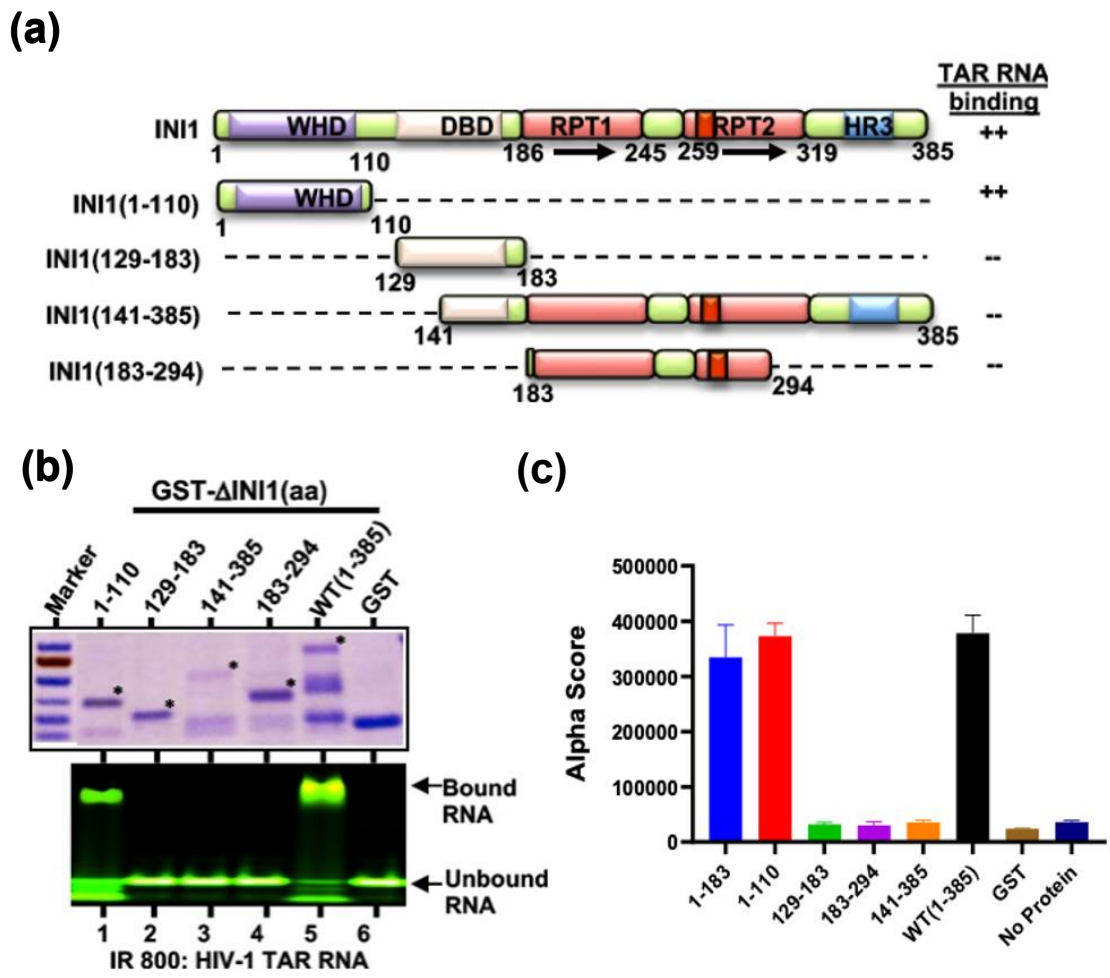

Figure 3-3. The INI1 N-terminal Winged Helix Domain (WHD) is the INI1 Element Responsible for HIV-1 5'UTR TAR Binding, and Not Other Domains. **(a**) The INI1 protein sequence truncations used to test binding of TAR RNA in (b) and (c). **(b)** Gel-shift assay of truncated INI1 reveals demonstrates that the WHD (aa 1-110) binds the TAR RNA. **(c)** Binding of TAR RNA by WHD is not enhanced by the INI1 DNA-binding domain (DBD), which immediately follows the WHD in the INI1 protein sequence.

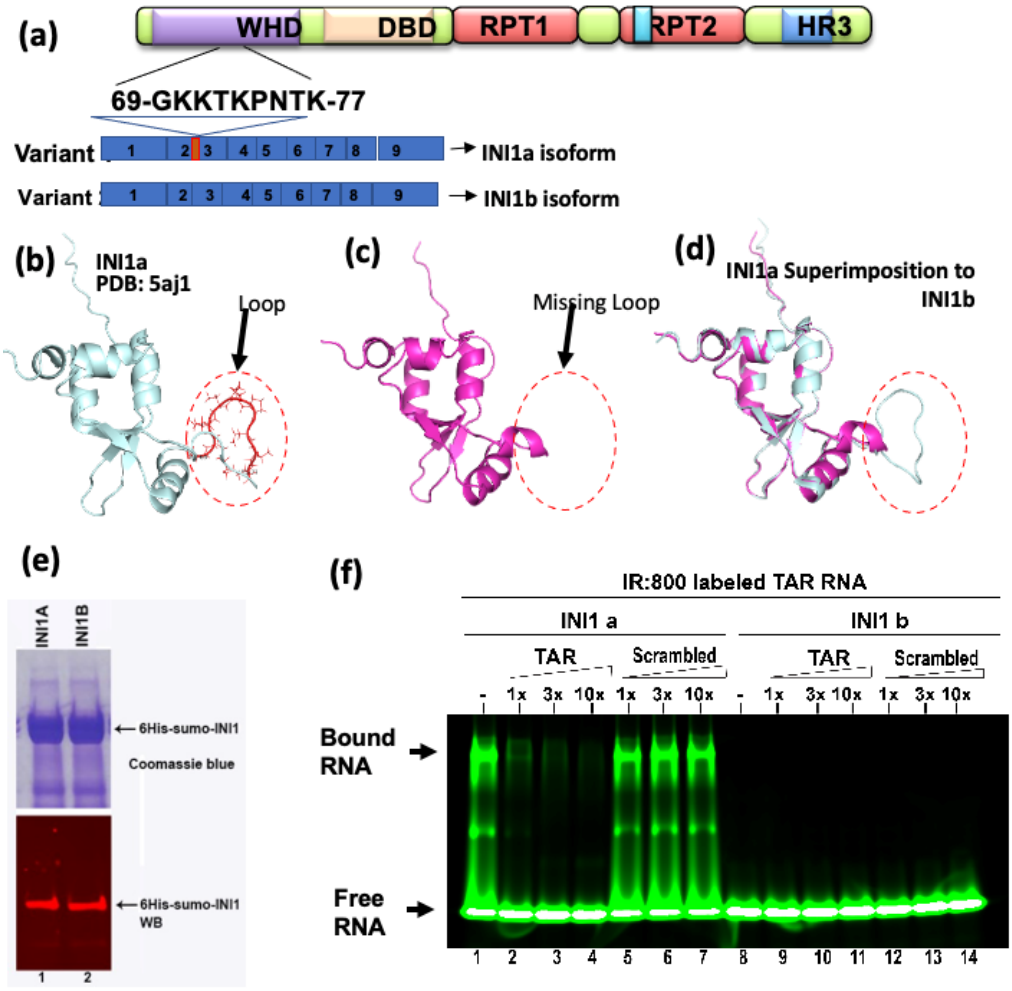

Figure 3-4. The INI1a Isomer, and Not INI1b, Binds TAR RNA. (a) INI1 is translated as two different isomers, INI1a and INI1b. They vary by a 9 amino acid deletion in the N-terminal winged helix domain, which is achieved by alternative splicing during transcription. **(b**) Crystal structure of INI1a, which differs from INI1b by the presence of an unstructured loop. **(c)** Crystral structure of INI1b. **(d)** Superimposition of INI1a to INI1b. **(e)** Gel-shift assay of INI1a and INI1b reveals that only the INI1a isomer is capable of binding HIV-1 5'UTR TAR RNA.

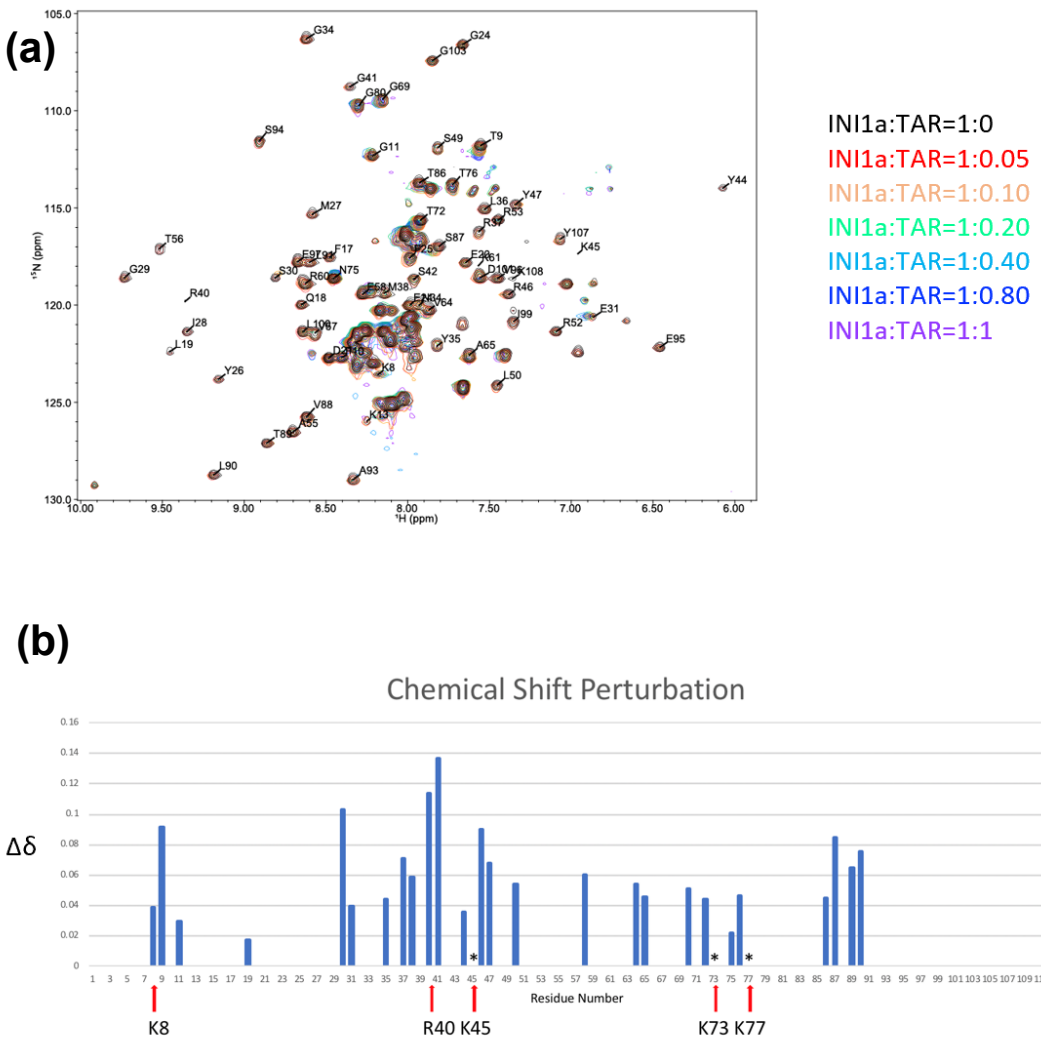

Figure 3-5. Titration of HIV-1 5'UTR TAR RNA to INI1a Shows Shifts Residue Shift with NMR Analysis. (a) TAR RNA was titrated to INI1a, allowed to incubate, then analyzed by NMR. Different residues positions shifted by different amount. **(b)** RMS analysis (0% vs. 40% molarity TAR RNA) shows heterogeneous shifting of residue positions, suggesting their presence at or near the binding interface.

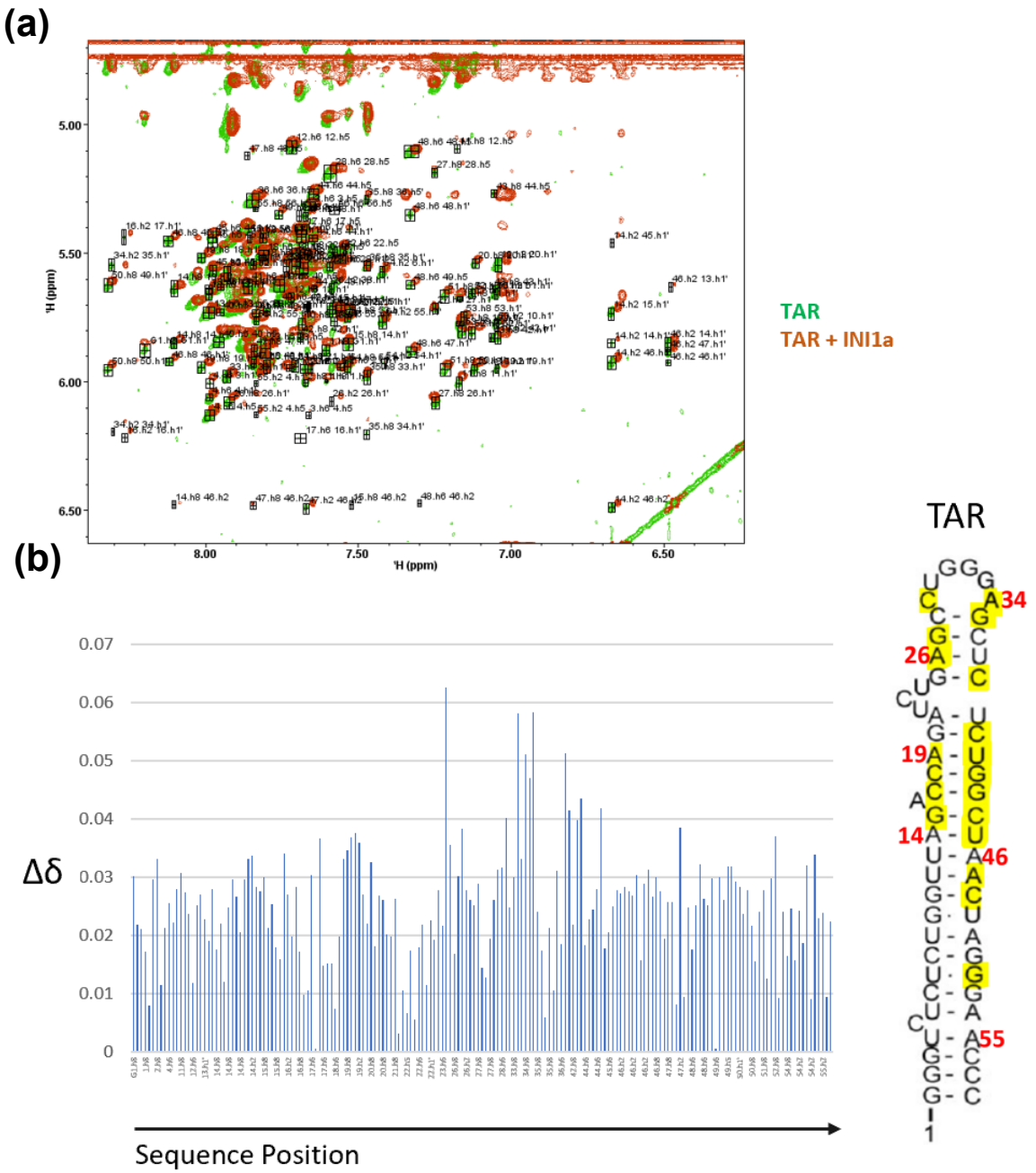
(a)
TAR
TAR + INI1a
(b)
Δδ
Sequence Position
TAR

Figure 3-6. NMR Analysis of TAR RNA with INI1a Shows Shifts Residue Shift. **(a)** TAR RNA was added to INI1a at equimolar concentrations, allowed to incubate, then analyzed by NMR. Residue positions were assigned to the TAR RNA only spectrum (green) and then with INI1a (orange) **(b)** RMS analysis (0% vs. 40% molarity TAR RNA) shows heterogeneous shifting of residue positions, suggesting their presence at or near the binding interface. Significantly perturbed residue NOESY's are marked on the TAR RNA structure (yellow).

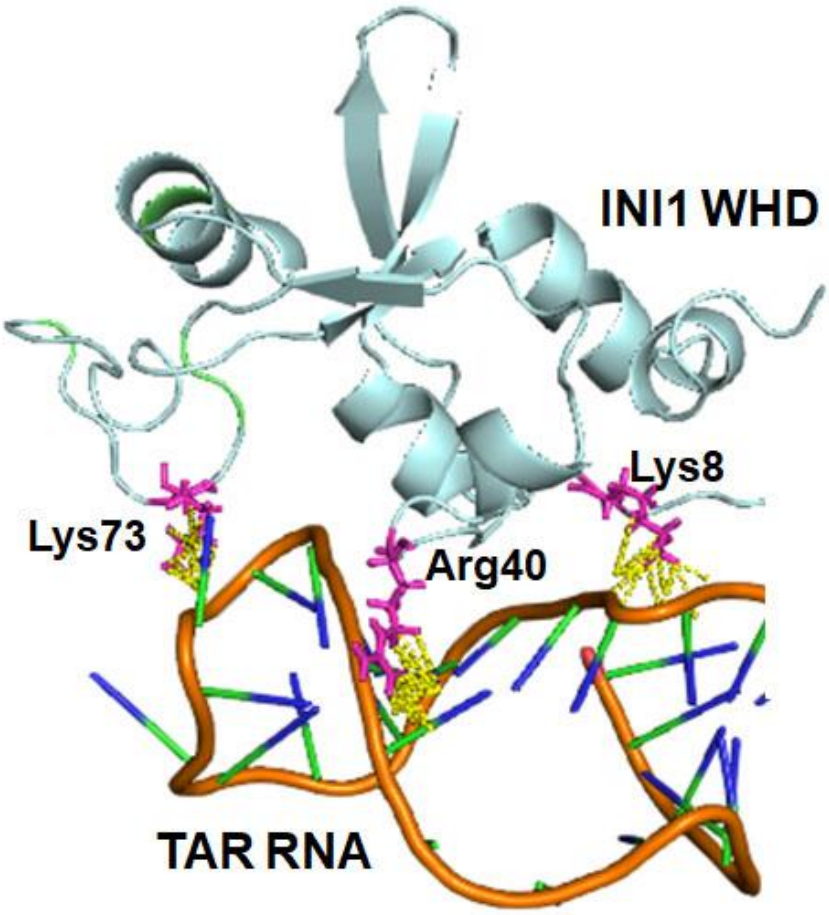

Figure 3-7. Modeling of TAR RNA Binding by INI1a WHD Domain. Computational model of INI1a WHD binding to TAR RNA, highlighting key amino acid and nucleotide residues of the binding interface.

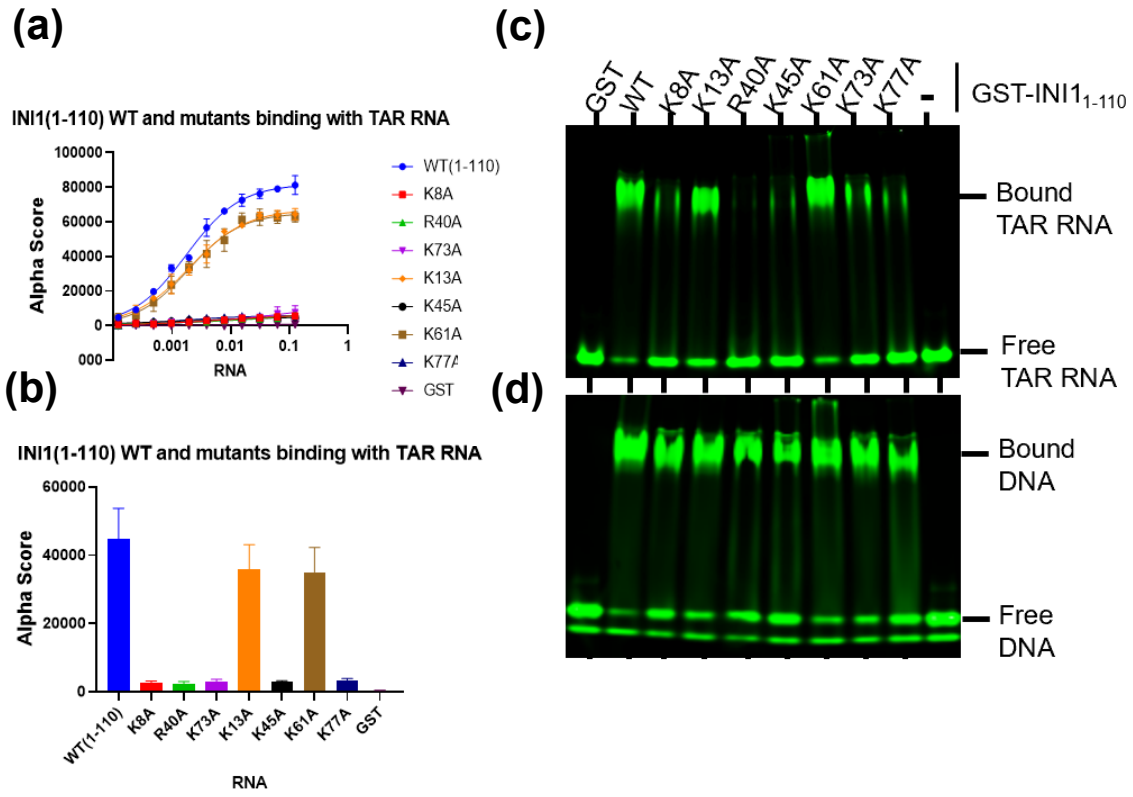

Figure 3-8. Basic Residues of the INI1a Loop Region are Required for TAR RNA Binding. A series of INI1a WHD mutants were made to test their binding affinity for TAR RNA. (a) The WHD domain binding was destroyed with mutants R40, K73A, K45A, K77A. Mutants K13A and K61A only slightly reduced binding affinity below WT. **(b)** Bar graph representation of (a). **(c)** Gel-shift assay shows similar results as (a) and (b) for TAR RNA binding. **(d)** Mutant DNA binding was similar between WT and all mutants.

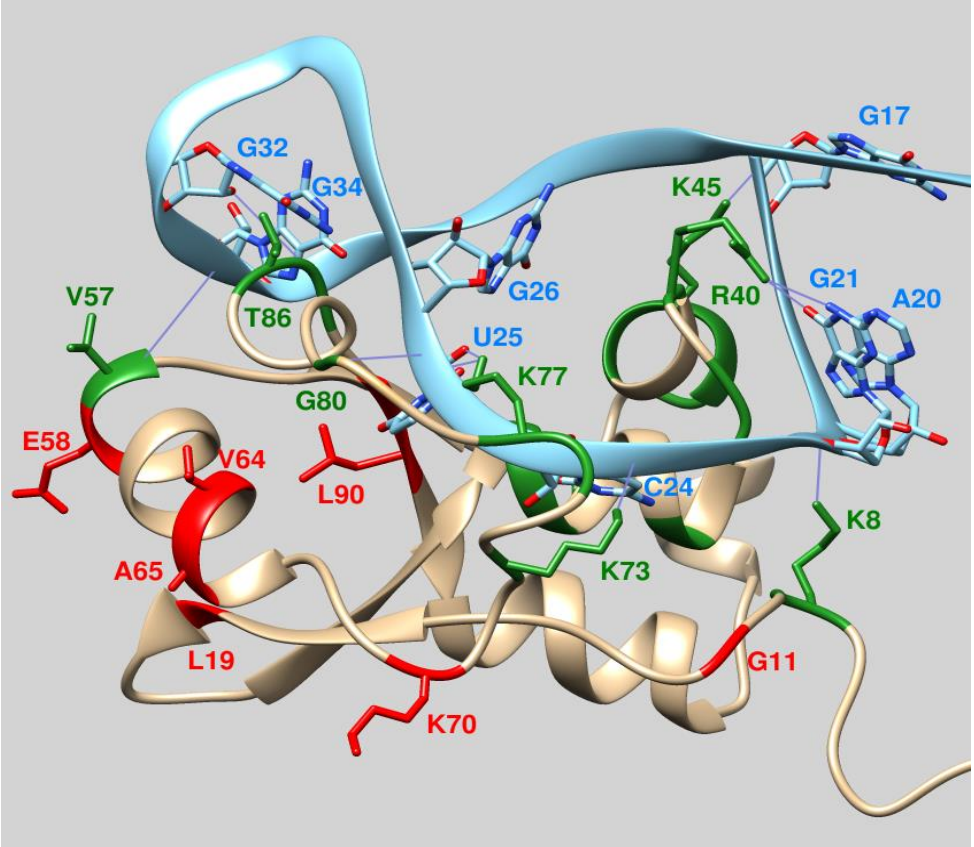

Figure 3-9. Modeling of TAR RNA Binding by INI1a WHD Domain. Computational model of INI1a WHD binding to TAR RNA, highlighting key amino acid and nucleotide residues of the binding interface.

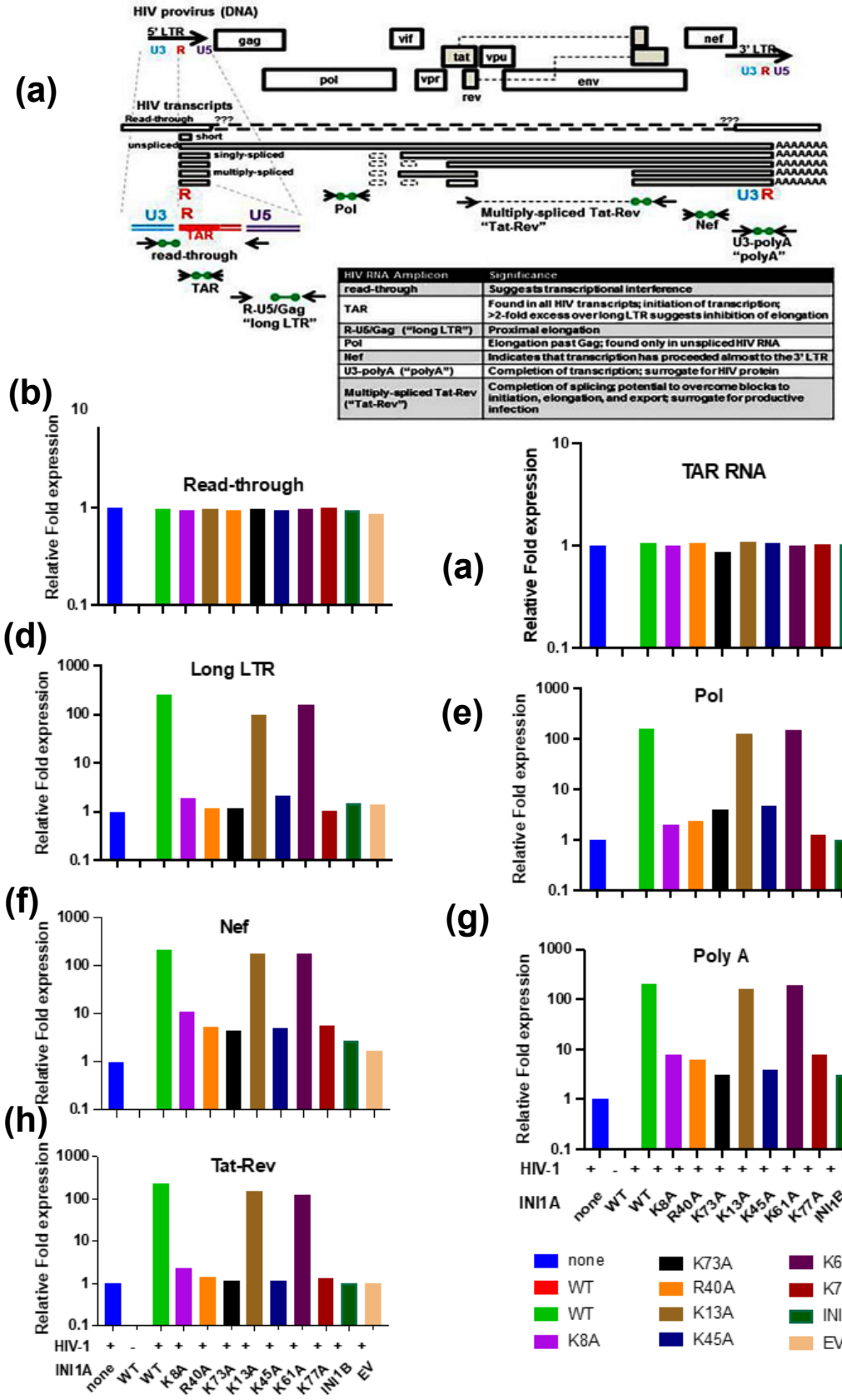
(a)
HIV provirus (DNA)
5' LTR
U3 R U5
gag
vif
tat
vpu
nef
3' LTR
U3 R U5
pol
vpr
rev
env
HIV transcripts
Read-through
???
short
unspliced
singly-spliced
multiply-spliced
AAAAAAA
R
U3
R
U5
TAR
Pol
Multiply-spliced Tat-Rev
"Tat-Rev"
Nef
U3-polyA
"polyA"
read-through
TAR
R-U5/Gag
"long LTR"
HIV RNA Amplicon
Significance
read-through
Suggests transcriptional interference
TAR
Found in all HIV transcripts; initiation of transcription; >2-fold excess over long LTR suggests inhibition of elongation
R-U5/Gag ("long LTR")
Proximal elongation
Pol
Elongation past Gag; found only in unspliced HIV RNA
Nef
Indicates that transcription has proceeded almost to the 3' LTR
U3-polyA ("polyA")
Completion of transcription; surrogate for HIV protein
Multiply-spliced Tat-Rev ("Tat-Rev")
Completion of splicing; potential to overcome blocks to initiation, elongation, and export; surrogate for productive infection
(b)
Read-through
Relative Fold expression
10
1
0.1
(a)
TAR RNA
(d)
Long LTR
1000
100
(e)
Pol
(f)
Nef
(g)
Poly A
(h)
Tat-Rev
HIV-1
INI1A
none
WT
WT
K8A
R40A
K73A
K13A
K45A
K61A
K77A
INI1B
EV

Figure 3-10. TAR RNA binding by INI1 is required for transcriptional elongation from HIV LTR promoter. Monocyte cells deficient in INI1 (ΔINI1) were co-transfected with pNL4-3 and INI1-containing plasmid under constitutive expression. **(a)** Levels of transcripts corresponding to spliced variants of were quantified using RT-PCR and correlated to read-through RNA levels **(b)** Quantification of proviral expression from plasmid, indicating similar levels or proviral template for transcription across samples. **(c)** Quantification of TAR RNA expression, demonstrating similar transcription rates near the 5'LTR promoter. **(d)** Elongation rate for 5'UTR region proximal to initiation of transcription elongation. **(e)** Elongation rate corresponding to unspliced RNA. **(f)** Elongation rate corresponding to near end of the provirus. **(g)** Complete elongation of a proviral rna. **(h)** Elongation rate corresponding to a full-spliced transcript.

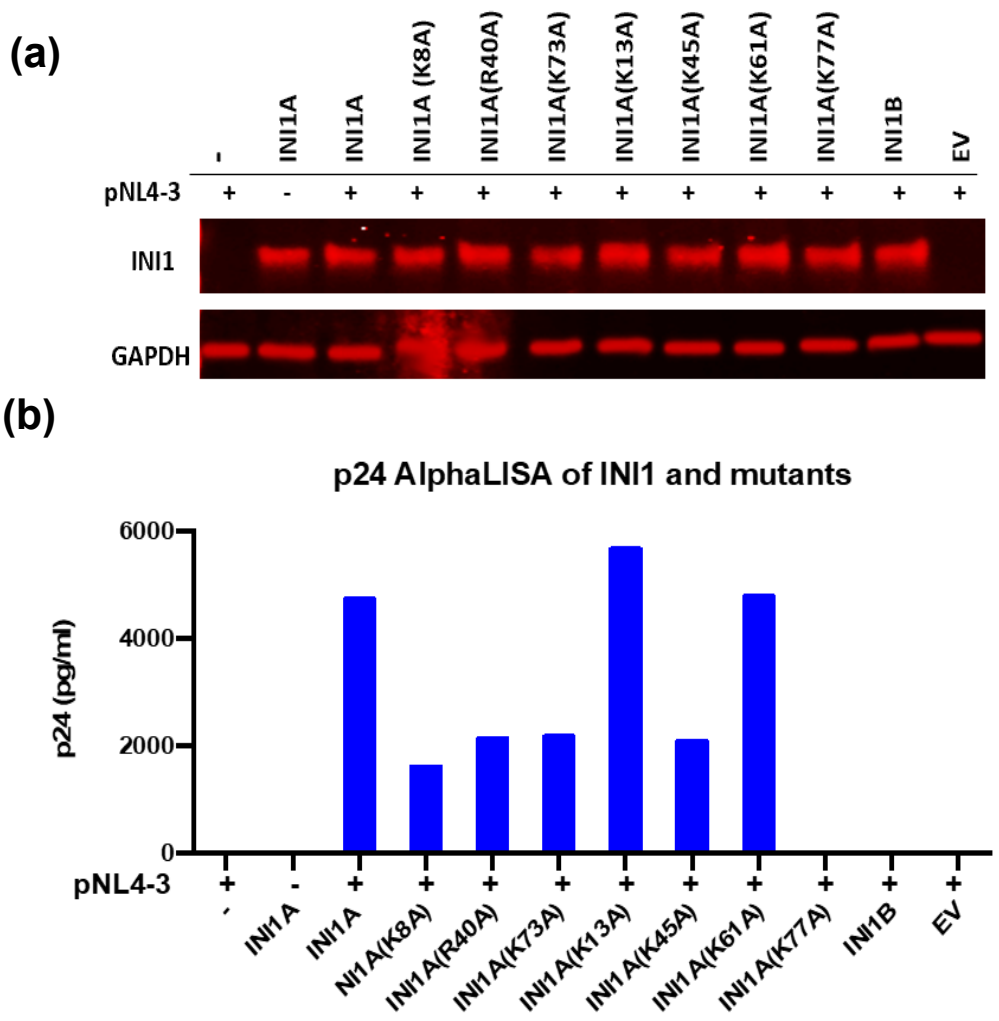

Figure 3-11. TAR RNA binding activity of INI1a is required for particle production of HIV-1. Monocyte cells deficient in INI1 (ΔINI1) were co-transfected with pNL4-3 and INI1-containing plasmid under constitutive expression. **(a)** Western blot of INI1 mutants and control (GAPDH) shows similar protein levels across all samples. **(b)** AlphaLISA of viral CA shows reduced virion production in INI1a binding-defective mutants (K8A, R40A, K73A, K45A).

CHAPTER 4. Future Directions: Identifying HIV-1 5'UTR Structural Features for Potential Therapeutic Exploration and Development

Abstract

Highly active antiretroviral therapy (HAART) is the current therapeutic strategy for HIV-1 infection, administering 3-5 agents targeting viral proteins (reverse transcriptase, protease, integrase, and envelope) to inhibit virion production, as there is no known cure for HIV-1 once the virus has integrated to within the host genome. Due to the high mutation rate of HIV-1 and thus the emerging antiviral resistance, the development of additional therapeutic compounds is of persistent need. Our work has established the importance of HIV-1 5'UTR structure for optimal infection and presents an opportunity for therapeutic targeting. Here we outline three preliminary future directions to utilize 5'UTR structure information for screening of novel therapeutic compounds to alter 5'UTR folding and/or structure and decrease viral fitness. First, we demonstrate the use of an inserted ^{1}H-NMR signal marker to monitor structuring of the 5'UTR TLE stem. Second, we test the ability of dimethylamiloride compounds to affect the folding pathway of nascently transcribed 5'UTR RNA. Third, we incorporate the fluorogenic Mango IV aptamer into the 5'UTR to provide a spectrophotometric output for high-throughput screening of novel compounds. These strategies

provide a rational foundation for further exploration of HIV-1 5'UTR folding as a novel therapeutic target.

Introduction

There is no known cure for HIV-1 after the virus has integrated into the host genome. Current therapeutics aim to inhibit different stages of the infection. The Highly active antiretroviral therapy (HAART) is successful in the management of HIV keeping virion levels low, preventing death of the infected human. The two main targets of retroviral therapy are the viral RT and PR. Despite a number of viable drugs, the need for additional therapeutic options remains. The full HIV genome is approximately 10 kb, and HIV has a mutation rate of approximately 1:10,000 (the highest of any known biological entity) (126). HIV will quickly mutate around any single therapeutic target, although this effect is somewhat mitigated with the invention of multi-drug therapy. In addition, HIV therapeutics have a number of off-target interactions with host biology, leading to a lowered quality of life in HIV+ patients.

Modern drug development approaches use three main strategies: molecular iteration of established effective compounds, structure-guided compound development based on presumed target interaction site, and high-through put screening (127). The most successful approach by far has been high-throughput screening. This strategy relies on development of a platform to quickly test tens to hundreds to thousands of compounds against the candidate with an easily-observable output signal such as colorimetric spectroscopy.

RNA and DNA aptamers have an emerging role in biotechnology. Aptamers have evolved naturally, the most poignant example being small molecule-aptamer modulation of transcription. This functionality is being extended to sensors, affinity

reagents, controlled release of therapeutics, and metabolic engineering, amongst other applications. Fluorogenic aptamers, aptamers that activate fluorescence of a dye upon binding, have value for studying RNA structures. Integrating their structure with an RNA of dynamic structure provides the opportunity for a visual output signal.

As established in Chapter 2, The HIV-1 5'UTR is folds into multiple structures that present opportunity for therapeutic targeting. Capturing transcribing RNA structures with small molecule binding has the opportunity to interfere with all elements downstream of transcription initiation, potentially inhibiting successful transcription, splicing, nuclear transport, translation, packaging, and virion production. Here we outline strategies to exploit the HIV 5'UTR for therapeutic targeting.

Future Direction 1: Monitor the RNA co-transcriptional folding by NMR.

As established in Chapter 1 and prior work, the HIV 5'UTR RNA adopts a number of conformational structures and motifs, and the nature of these structures, such as the DIS and PBS, varies depending on 5'G content, structural of upstream elements, and protein association. We have done a number of NMR studies in an attempt to find structural signal markers that could be used to identify structuring of these elements of the 5'UTR. We have attempted to monitor transcribing RNA by T7 polymerase within an NMR tube, but the effort was unsuccessful due to low signal intensity and signal overlap (data not shown). To build in a NMR signal marker for the PBS-segment structure, we have incorporated the previously

published lrAID signal (77) into the TLE stem of PBS (Figure 4-1a). This 8 nucleotide base-pairing causes a characteristic downshift of the central adenosine in 1D-^{1}H-NMR to 6.4 ppm, outside the signal of most other nucleotides. We incorporated this motif into the TLE stem of point mutants that have been previously established to be monomer (J1) or dimer-prone (J8) when refolded (49). We found that the dimer-prone variant had an intact TLE stem, while the monomer-prone mutant did not (Figure 4-1b). These data suggest that the folding of the PBS-segment is correlated with the folding of DIS, that the three-way junction structure of PBS-segment exist when the DIS is exposed and an alternative PBS-segment structure is adopted when DIS is sequestered. NMR's utility lies in its ability to gain information about binding and folding after a potential molecule has been identified. It has the additional benefit of being the primary method of obtaining atomic-level resolution of molecules in solution.

Future Direction 2: Small compound screening to identify compounds that alter RNA co-transcriptional folding pathway.

Due to the complex folding pathway of the HIV-1 5'UTR, and the necessity of this to be properly structured for function, we aimed to identify novel compounds that could bind to transcribing HIV-1 5'UTR RNA, locking it into a non-productive folding pathway. Although NMR studies provide detailed RNA folding information, but they have the downside of requiring a large sample concentration, and a time-intensive sample preparation and analysis. Instead, since the folding of PBS and DIS is coupled, we employed EMSA to quickly screen some small compounds through collaboration with Dr. Amanda Hargrove (Duke). The Hargrove laboratory

has developed and synthesized small compounds libraries targeting RNAs (128, 129). There has been some preliminary success with dimethylamiloride compounds (DMA), and we chose a handful of these for analysis (Figure 4-2a) (130). We attempted to transcribe both full length HIV-1 5'UTR and noTAR -10, which was established in Chapter 1 to fold into a monomer. We found no significant differences with the tested compounds, with the exception of DMA-192, which may have produced a slight alteration of the folding pathway (Figure 4-2b). DMA-177 and DMA-178, although appeared to inhibit T7 transcription, changed the noTAR-10 transcripts to mainly dimers. How these compounds interact with the HIV-1 RNA and affect the transcriptional folding pathway required further investigation. Many tested DMA compounds did not affect nascently-transcribed RNA folding. It is possible that these compounds are non-interacting, but that falls in contradiction to previously established results of these molecules capability to bind to TAR. It is possible that they do bind but are unable to affect the overall folding of the 5'UTR, rendering them therapeutically null. Alternatively, these compounds could be effective, but were not tested at viable concentrations, and should be tested at higher concentrations. Testing drug candidates on transcribing 5'UTR RNA is a novel concept and has merit due to its ease of performance. These can be performed at small volumes and are not particularly expensive. Additionally, tying the folding to an easily observable output, such as fluorescence, permits the possibility of significantly speeding up screening time.

Future Direction 3: Establish a fluorescence-based platform for high-throughput screening of compounds affecting RNA co-transcriptional pathway.

Due to observe structural differences in the HIV-1 5'UTR (Chapter 1), we saw an opportunity to exploit the differences in the 5'UTR by incorporation of the Mango IV aptamer to within the RNA structure (Figure 4-1). We predicted because the DIS stem is structured when dimerized and folded differently in the absence when the RNA is not multimerized. We minimized the Mango IV aptamer to the active region, with the stabilizing stem being a portion of the DIS, and the sequence ended with a complementary sequence to anneal to the DIS. This structural switch would fluoresce when the DIS is not structured (RNA is not dimerized), and fluoresce when the RNA was monomeric. Our constructs differed by 5' template truncation, with the monomer sequence being analogous to the noTAR10, and the dimer sequence analogous to the noTAR noPolyA outlined in Chapter 1. We found that the nascently transcribed RNAs did not fluoresce as expected, with the WT construct being comparable to the monomer (Figure 4-3b). When refolded, we also found that the WT construct seemed to fluoresce brighter than the monomer control (Figure4-3c). Additionally, the full length 5'UTR (no aptamer) fluoresced stronger than our aptamer constructs. We did not have success with preliminary testing of the incorporation of the Mango IV aptamer into the DIS sequence of the 5'UTR. One main issue is that the concentration of nascently-transcribed RNA varied between the WT, monomer, and dimer-Mango IV constructs varied, while the concentration of dye was fixed (data not shown). The amount of output RNA

must be controlled or measured to correlate to the fluorescent intensity. Additional SHAPE experiments of nascently transcribed RNA, as outlined in Chapter 1, will provide more insight to the 5'UTR structure, allowing for the potential for iteration of the platform by incorporation of the aptamer into a more productive location. Our refolded data (Figure 4-3c) suggests the possibility of nonspecific interaction, due to high fluorescence of our WT construct without the aptamer present. It is likely that the refolded data is of less utility than the nascently-transcribed RNA data, due to the increased thermodynamic favorability of binding the dye while the RNA is folding, whereas the nascently transcribed RNA would not bind dye until the full sequence is transcribed.

Author Contributions

Platform design S.S., X.H.; Plate reader assistance J.J.

Experimental Procedures

Plasmids

Plasmids pNL4-3 (NIH AIDS Reagent Program) and pUC19 containing the HIV-1 5'UTR sequence for DNA template synthesis were generated using methods described previously (84). The pNL4-3 HIV-1 sequence was fused into the pUC19 sequence for amplification. Plasmid sequences were verified by Sanger sequencing (DNA Core, University of Missouri, Columbia, MO).

In Vitro RNA Transcription

RNAs used for refolding were transcribed using purified T7 polymerase. Corresponding DNA templates were generated via PCR and purified prior to

transcription. Transcriptions were performed at 5-20 mL volumes using ~0.5-2 mg of purified DNA template, 40 mM Tris-HCl (pH = 8.1), 5 mM DTT, 1 mM spermidine, 0.01%Triton X-100, 5-10 mM NTP's, and 5-25 mM $MgCl_2$. Optimal concentrations of $MgCl_2$ and NTP's to maximize RNA yield was determined using small-scale reactions following by RNA staining on polyacrylamide gel. The transcriptions were incubated for 3 hours at 37C, and the reaction was then quenched using a Urea-EDTA mixture (7 M Urea pH = 8.0. 250 mM EDTA). The RNA was isolated on polyacrylamide denaturing gel by electrophoresis using FisherBiotech DNA sequencing system at 16W overnight. The RNA bands were visualized via UV-shadowing, excised, and then eluted overnight using the Elutrap electroelution system (Whatman) at 100 V. The purified RNA fraction was washed with 2M NaCl and then desalted within the Amicon Ultra-4 Centrifugal Filter Device (Millipore). RNA concentrations were determined by measuring absorbance at 260 nm using the NanoDrop 2000c (Thermo Fisher Scientific).

Transcription of Nascently-Folded RNAs with DMA Derivatives

Corresponding DNA templates were generated via PCR and purified prior to transcription. Transcription of all RNAs whose folding during nascent transcription were monitored were performed under identical conditions. The transcription reaction was performed at a volume of 40 µL using 0.01 ρM final concentration of purified DNA template, 40 mM Tris-HCl (pH = 8.1), 5 mM DTT, 1 mM spermidine, 0.01%Triton X-100, 10 mM NTP's, 0.5 µM T7 polymerase, and 15 mM $MgCl_2$. Each DMA compound was added to a final concentration of 50 µM. The reaction was incubated for 1 hour at 37°C. Following this, 10 µL of transcription

+ 2 µL of 50% glycerol was loaded onto an ethidium-bromide-containing 2% native agarose gel (tris-borate) that was pre-chilled on ice, and then ran at 100 V for 60 minutes. The gel was then visualized using the Molecular Imager Gel Doc XR+ System (Bio-Rad). Image quality was optimized using the Image Lab software (Bio-Rad).

NMR Analysis of J1 and J8 lrAID RNAs

Refolded RNAs at a concentration of 200 µM in 200 µL D_2O with 10 mM dTris-dCl (pH = 8.0) were boiled for three minutes. They were then snap-cooled on ice for three minutes. Salts were then added to the sample (1 mM $MgCl_2$, 140 mM KCl, 10 mM NaCl) and allowed to equilibrate for 1 hour at 37°C. NMR analysis was performed using a Bruker AVANCE III spectrometer equipped with a TCI cryoprobe (800 MHz, ^2H, 37°C; NMR Core, University of Missouri). Data was processed using NMRPipe/NMRDraw (123) and analyzed using NMRViewJ (125).

Mango IV RNA Fluorescence Assay

RNAs were transcribed as described in "*Transcription of Nascently-Folded RNAs with DMA Derivatives,"* but in the absence of DMA compounds. Instead, 5 µM of a biotinylated thiazole orange derivative, TO1-3PEG-Biotin (ABM), was added at the start of transcription or at the end of the transcription (60 minutes). Samples were then analyzed on a transparent 96-well plate using a plate reader (Life Sciences Center, University of Missouri) at excitation λ= 510 nm, emission λ= 535 nm.

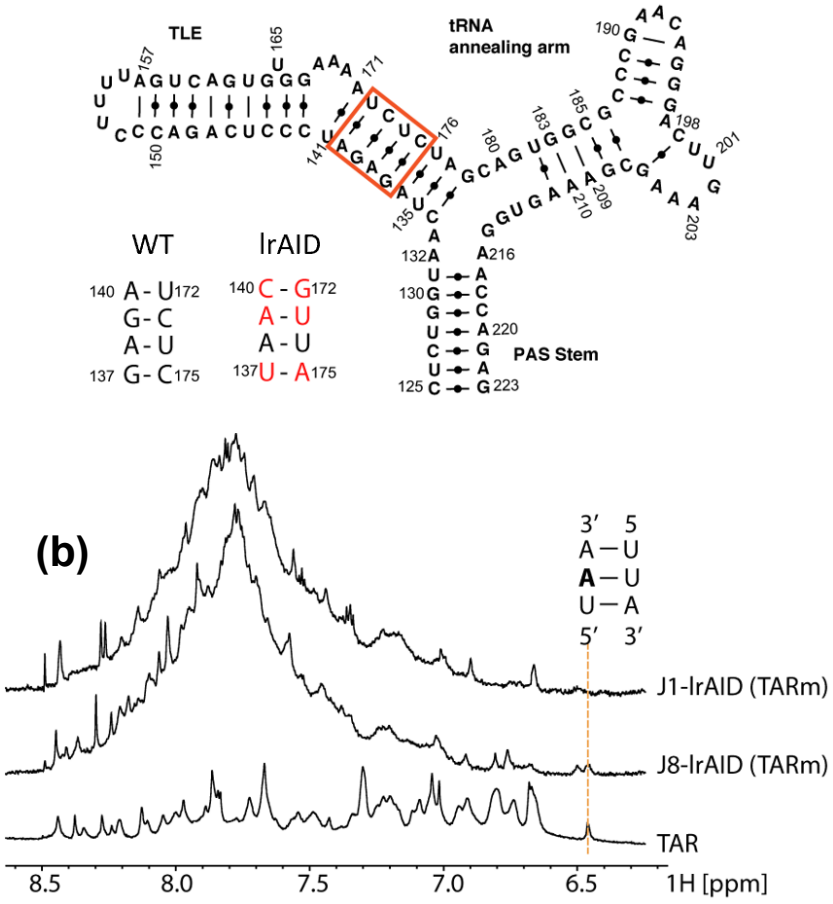

Figure 4-1. Probing the secondary structure of the PBS-segment by NMR. (a) The paired region G137-A140:U172-C175 was selected to probe the structural changes in the PBS-segment in 5'UTR mutants. The boxed residues were mutated to an lr-AID sequence so that A138-H2 is expected have a chemical shift at 6.47 ppm if the base pairs are formed. **(b)** 1D ^{1}H-NMR of J1-lrAID and J8-lrAID are shown. The signal is present in J8-lrAID but absent in J1-lrAID. TAR RNA (nt 1-58) containing a native lrAID signal (A46-H2) was shown as a control.

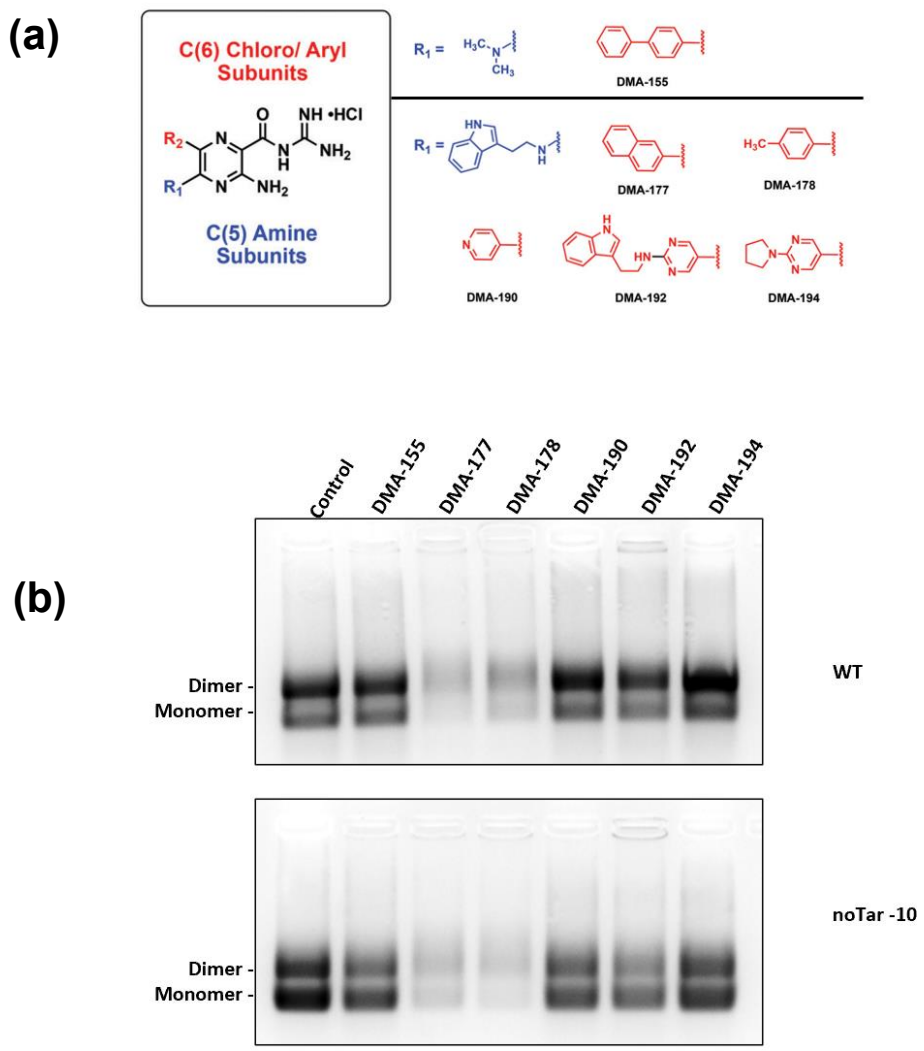

Figure 4-2. Test of Compounds for Alteration of HIV-1 5'UTR Transcriptional Folding. (a) Structure of molecules tested. **(b)** *In vitro* transcription of full length HIV-1 5'UTR or noTar -10 (as in Figure 2-6) using T7 polymerase. DMA-192 showed slight alteration of monomer:dimer ratio.

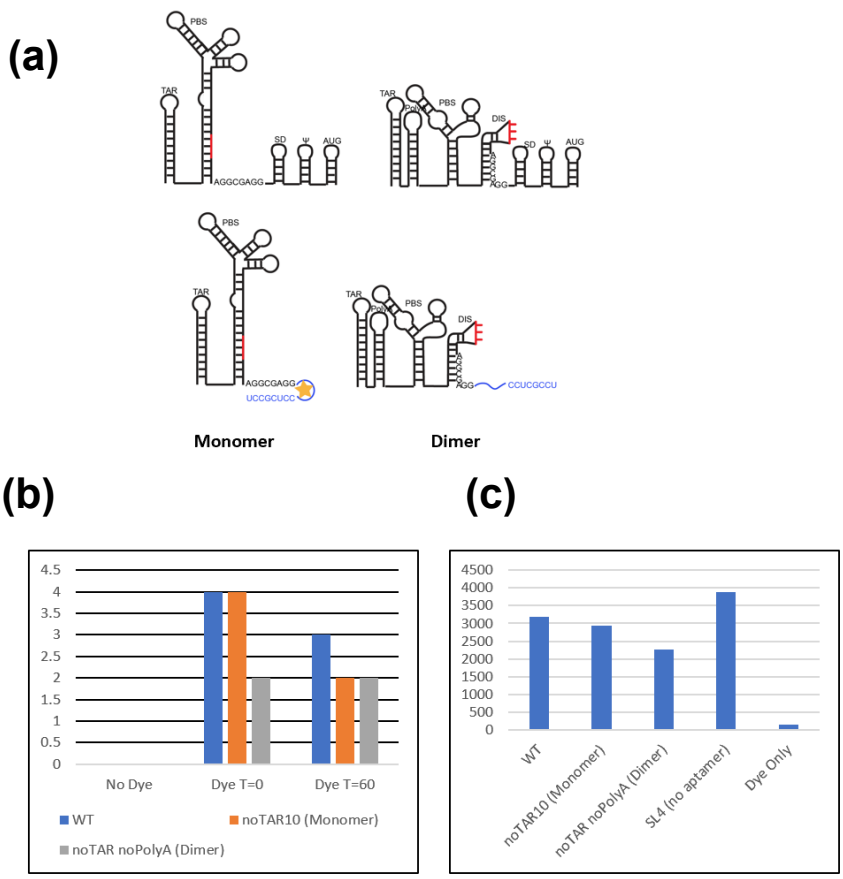

Figure 4-3. Mango IV 5'UTR Fluorescent Reporter. (a) Chapter 1: HIV-1 5'UTR folds into different shapes, exposing or occluding the DIS stemloop (SL1). The fluorophore-binding region of the mango IV aptamer was incorporated into hairpin of the DIS stem, followed by a consensus sequence to the DIS stem, allowing formation of the active complex. Monomer 5'UTR structure should permit fluorescence, while the dimer structure should not permit fluorescence. **(b)** RNA was transcribed *in vitro* by T7 polymerase for 1 hr. Dye was added either at the start of transcription or at 60 minutes. Fluorescent intensity of samples were then measured. **(c)** Mango IV-RNAs were refolded in the presence of dye and then the fluorescent intensity was measured.

Concluding Remarks

The work presented in this dissertation utilized functional biochemical approaches, structure studies, computational approaches, cell-based approaches, and computational approaches to explore the importance of the HIV-1 5'UTR RNA in infection. In Chapter 2 of this dissertation, we revealed a soon-to-be published non-canonical, non-eIF4E translation pathway, and showed evidence implicating the proper 5'UTR folding as a potential mechanism to direct RNA fate. Additionally, we showed evidence demonstrating that upstream elements of the HIV-1 5'UTR affect the folding of downstream elements, suggesting the fate of mRNAs are determined as they are nascently transcribed. The work provides foundational evidence for further investigation into the folding of mRNAs and their role in the infectious lifecycle. In Chapter 3 of this dissertation, we mapped the binding interface of HIV-1 5'UTR TAR RNA with INI1, a core component of the SWI/SNF eukaryotic chromatin remodeling complex. This is the first example of HIV-1 RNA surface structure mimicking a host protein surface, and the modeled structure will be vital for development of novel therapeutics targeting the interaction between INI1 and TAR. We also showed evidence that defective INI1 results in decreased transcriptional elongation and decreased virion particle production. In Chapter 3 of this dissertation, we outlined three future directions to monitor HIV-1 5'UTR folding for screening of novel therapeutics. The 5'UTR folding is itself a novel target, and the work presented here in this dissertation is an essential basis for therapeutic success.

References

1. Coffin, J.M., Hughes, S.H. and Varmus, H. (1997) *Retroviruses*

2. Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*. **312**, 763–767

3. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIV gp 120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*. **393**, 648–659

4. Mcdougal, J. S., Kennedy, M. S., Sligh, J. M., Cort, S. P., Mawle, A., and Nicholson, J. K. A. (1986) Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science (80-.).* **231**, 382–385

5. Zaitseva, M. B., Lee, S., Rabin, R. L., Tiffany, H. L., Farber, J. M., Peden, K. W. C., Murphy, P. M., and Golding, H. (1998) CXCR4 and CCR5 on human thymocytes: Biological function and role in HIV- 1 infection. *J. Immunol.* **161**, 3103–31013

6. Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* **17**, 657–700

7. Freed, E. O. (2001) HIV-1 replication. *Somat. Cell Mol. Genet.* **26**, 13–33

8. Alkhatib, G. (2009) The biology of CCR5 and CXCR4. *Curr. Opin. HIV AIDS*. **4**, 96–103

9. Pancera, M., Majeed, S., Ban, Y. E. A., Chen, L., Huang, C. C., Kong, L., Kwon, Y. Do, Stuckey, J., Zhou, T., Robinson, J. E., Schief, W. R., Sodroski, J., Wyatt, R., and Kwong, P. D. (2010) Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1166–1171

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

11. Acton, O., Grant, T., Nicastro, G., Ball, N. J., Goldstone, D. C., Robertson, L. E., Sader, K., Nans, A., Ramos, A., Stoye, J. P., Taylor, I. A., and Rosenthal, P. B. (2019) Structural basis for Fullerene geometry in a human endogenous retrovirus capsid. *Nat. Commun.* **10**, 1–13

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

13. Nikolaitchik, O., Rhodes, T. D., Ott, D., and Hu, W.-S. (2006) Effects of

Mutations in the Human Immunodeficiency Virus Type 1 gag Gene on RNA Packaging and Recombination. *J. Virol.* **80**, 4691–4697

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

15. Santos, S., Obukhov, Y., Nekhai, S., Bukrinsky, M., and Iordanskiy, S. (2012) *Virus-producing cells determine the host protein profiles of HIV-1 virion cores*, 10.1186/1742-4690-9-65

16. Isel, C., Ehresmann, C., and Marquet, R. (2010) Initiation of HIV reverse transcription. *Viruses*. **2**, 213–243

17. Cosnefroy, O., Murray, P. J., and Bishop, K. N. (2016) HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*. 10.1186/s12977-016-0292-7

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

19. Pertel, T., Hausmann, S., Morger, D., Züger, S., Guerra, J., Lascano, J., Reinhard, C., Santoni, F. A., Uchil, P. D., Chatel, L., Bisiaux, A., Albert, M. L., Strambio-De-Castillia, C., Mothes, W., Pizzato, M., Grütter, M. G., and Luban, J. (2011) TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature*. **472**, 361–365

20. Delelis, O., Carayon, K., Saïb, A., Deprez, E., and Mouscadet, J. F. (2008) Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology*. 10.1186/1742-4690-5-114

21. Craigie, R., and Bushman, F. D. (2012) HIV DNA integration. *Cold Spring Harb. Perspect. Med.* 10.1101/cshperspect.a006890

22. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*. **331**, 280–283

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

24. Emery, A., Zhou, S., Pollom, E., and Swanstrom, R. (2017) Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* 10.1128/jvi.02515-16

25. Pollard, V. W., and Malim, M. H. (1998) The HIV-1 Rev protein. *Annu. Rev. Microbiol.* **52**, 491–532

26. Freed, E. O. (2015) HIV-1 assembly, release and maturation HHS Public Access. *Nat Rev Microbiol*. **13**, 484–496

27. Schmidt, O., and Teis, D. (2012) The ESCRT machinery. *Curr. Biol.* **22**, R116

28. Seitz, R. (2016) Human Immunodeficiency Virus (HIV). *Transfus. Med. Hemotherapy*. **43**, 203–222

29. Brigham, B. S., Kitzrow, J. P., Reyes, J. P. C., Musier-Forsyth, K., and Munro, J. B. (2019) Intrinsic conformational dynamics of the HIV-1 genomic RNA 5'UTR. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10372–10381

30. Bieniasz, P., and Telesnitsky, A. (2018) Multiple, switchable protein: RNA interactions regulate human immunodeficiency virus type 1 assembly. *Annu. Rev. Virol.* **5**, 165–183

31. Berkhout, B. (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **54**, 1–34

32. Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E. L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S. S., LaCotti, C., Barton, S., Tummillo, D., Hosic, A., Edme, K., Albrecht, S., Telesnitsky, A., and Summers, M. F. (2011) NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science (80-.).* **334**, 242–245

33. Brown, J. D., Kharytonchyk, S., Chaudry, I., Iyer, A. S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S. H., Singh, K., Lopresti, M. W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F. G., Stewart, K., Francis, D., Edwards, B., Chen, P., Case, D. A., Telesnitsky, A., and Summers, M. F. *Structural basis for transcriptional start site control of HIV-1 RNA fate*, [online] http://science.sciencemag.org/ (Accessed June 17, 2021)

34. Abbink, T. E. M., Ooms, M., Haasnoot, P. C. J., and Berkhout, B. (2005) The HIV-1 Leader RNA Conformational Switch Regulates RNA Dimerization but Does Not Regulate mRNA Translation †. 10.1021/bi0502588

35. Baudin, F., Marquet, R., Isel, C., Carlix, J. L., Ehresmann, B., and Ehresmann, C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* **229**, 382–397

36. Clever, J., Sassetti, C., and Parslow, T. G. (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* **69**, 2101–2109

37. Harrison, G. P., and Lever, A. M. (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* **66**, 4144–4153

38. McBride, M. S., and Panganiban, A. T. (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA

element composed of functional hairpin structures. *J. Virol.* **70**, 2963–2973

39. Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J.-L., and Sonenberg, N. (2003) The Leader of Human Immunodeficiency Virus Type 1 Genomic RNA Harbors an Internal Ribosome Entry Segment That Is Active during the G 2 /M Phase of the Cell Cycle . *J. Virol.* **77**, 3939–3949

40. Gendron, K., Ferbeyre, G., Heveker, N., and Brakier-Gingras, L. (2011) The activity of the HIV-1 IRES is stimulated by oxidative stress and controlled by a negative regulatory element. *Nucleic Acids Res.* **39**, 902–912

41. Plank, T.-D. M., Whitehurst, J. T., Cencic, R., Pelletier, J., and Kieft, J. S. (2014) Internal translation initiation from HIV-1 transcripts is conferred by a common RNA structure. *Translation*. **2**, e27694

42. Hartman, T. R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D. R., and Boris-Lawrie, K. (2006) RNA helicase A is necessary for translation of selected messenger RNAs. *Nat. Struct. Mol. Biol.* **13**, 509–516

43. Bolinger, C., Yilmaz, A., Hartman, T. R., Kovacic, M. B., Fernandez, S., Ye, J., Forget, M., Green, P. L., and Boris-Lawrie, K. (2007) RNA helicase A interacts with divergent lymphotropic retroviruses and promotes translation of human T-cell leukemia virus type 1. *Nucleic Acids Res.* **35**, 2629–2642

44. Bolinger, C., Sharma, A., Singh, D., Yu, L., and Boris-Lawrie, K. (2010) RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res.* **38**, 1686–1696

45. Boeras, I., Song, Z., Moran, A., Franklin, J., Brown, W. C., Johnson, M., Boris-Lawrie, K., and Heng, X. (2016) DHX9/RHA Binding to the PBS-Segment of the Genomic RNA during HIV-1 Assembly Bolsters Virion Infectivity. *J. Mol. Biol.* **428**, 2418–2429

46. Huthoff, H., and Berkhout, B. (2001) Two alternating structures of the HIV-1 leader RNA. *RNA*. **7**, 143–157

47. Abbink, T. E. M., and Berkhout, B. (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 rna occludes the gag start codon. *J. Biol. Chem.* **278**, 11601–11611

48. Ooms, M., Huthoff, H., Russell, R., Liang, C., and Berkhout, B. (2004) A Riboswitch Regulates RNA Dimerization and Packaging in Human Immunodeficiency Virus Type 1 Virions. *J. Virol.* **78**, 10814–10819

49. Abbink, T. E. M., Ooms, M., Haasnoot, P. C. J., and Berkhout, B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*. **44**, 9058–9066

50. Lu, K., Heng, X., and Summers, M. F. (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–633

51. Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., Bolden, N. C., McCowin, S., Case, D. A., Johnson, B. A., Salemi, M., Telesnitsky, A., and Summers, M. F. (2015) Structure of the HIV-1 RNA packaging signal. *Science (80-.).* **348**, 917–921

52. Brown, J. D., Kharytonchyk, S., Chaudry, I., Iyer, A. S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S. H., Singh, K., Lopresti, M. W., Orellana, M., Rodriguez, T., Oboh, U., Hijji, J., Ghinger, F. G., Stewart, K., Francis, D., Edwards, B., Chen, P., Case, D. A., Telesnitsky, A., and Summers, M. F. (2020) Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science (80-.).* **368**, 413–417

53. Keane, S. C., Van, V., Frank, H. M., Sciandra, C. A., McCowin, S., Santos, J., Heng, X., and Summers, M. F. (2016) NMR detection of intermolecular interaction sites in the dimeric 5′-leader of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13033–13038

54. Kharytonchyk, S., Monti, S., Smaldino, P. J., Van, V., Bolden, N. C., Brown, J. D., Russo, E., Swanson, C., Shuey, A., Telesnitsky, A., and Summers, M. F. (2016) Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13378–13383

55. Boeras, I., Seufzer, B., Brady, S., Rendahl, A., Heng, X., and Boris-Lawrie, K. (2017) The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Sci. Rep.* **7**, 1–10

56. Masuda, T., Sato, Y., Huang, Y. L., Koi, S., Takahata, T., Hasegawa, A., Kawai, G., and Kannagi, M. (2015) Fate of HIV-1 cDNA intermediates during reverse transcription is dictated by transcription initiation site of virus genomic RNA. *Sci. Rep.* **5**, 17680

57. Piserà, A., Campo, A., and Campo, S. (2018) Structure and functions of the translation initiation factor eIF4E and its role in cancer development and treatment. *J. Genet. Genomics*. **45**, 13–24

58. Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., Shi, E. Y., Stumpf, C. R., Christensen, C., Bonham, M. J., Wang, S., Ren, P., Martin, M., Jessen, K., Feldman, M. E., Weissman, J. S., Shokat, K. M., Rommel, C., and Ruggero, D. (2012) The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*. **485**, 55–61

59. Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S., and Sabatini, D. M. (2012) A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*. **485**, 109–113

60. Chi, H. (2012) Regulation and function of mTOR signalling in T cell fate decisions. *Nat. Rev. Immunol.* **12**, 325–338

61. Borden, K. L. B., and Volpon, L. (2020) The diversity, plasticity, and

adaptability of cap-dependent translation initiation and the associated machinery. *RNA Biol.* **17**, 1239–1251

62. Liu, G. Y., and Sabatini, D. M. (2020) mTOR at the nexus of nutrition, growth, ageing and disease. *Nat. Rev. Mol. Cell Biol.* **21**, 183–203

63. Pópulo, H., Lopes, J. M., and Soares, P. (2012) The mTOR signalling pathway in human cancer. *Int. J. Mol. Sci.* **13**, 1886–1918

64. Vadysirisack, D. D., and Ellisen, L. W. (2012) MTOR activity under hypoxia. *Methods Mol. Biol.* **821**, 45–58

65. Chee, N. T., Lohse, I., and Brothers, S. P. (2019) MRNA-to-protein translation in hypoxia. *Mol. Cancer*. **18**, 1–13

66. Hong, S., Freeberg, M. A., Han, T., Kamath, A., Yao, Y., Fukuda, T., Suzuki, T., Kim, J. K., and Inoki, K. (2017) LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. *Elife*. 10.7554/eLife.25237

67. Jefferies, H. B. J., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. (1997) Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70(s6k). *EMBO J.* **16**, 3693–3704

68. Meyuhas, O., and Kahan, T. (2015) The race to decipher the top secrets of TOP mRNAs. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849**, 801–811

69. de la Parra, C., Ernlund, A., Alard, A., Ruggles, K., Ueberheide, B., and Schneider, R. J. (2018) A widespread alternate form of cap-dependent mRNA translation initiation. *Nat. Commun.* **9**, 1–9

70. Singh, G., Fritz, S. E., Seufzer, B., and Boris-Lawrie, K. (2020) The mRNA encoding the JUND tumor suppressor detains nuclear RNA-binding proteins to assemble polysomes that are unaffected by mTOR. *J. Biol. Chem.* **295**, 7763–7773

71. Sharma, A., Yilmaz, A., Marsh, K., Cochrane, A., and Boris-Lawrie, K. (2012) Thriving under stress: Selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eiF4E translation activity. *PLoS Pathog.* **8**, e1002612

72. Leblanc, J., Weil, J., and Beemon, K. (2013) Posttranscriptional regulation of retroviral gene expression: Primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA. *Wiley Interdiscip. Rev. RNA*. **4**, 567–580

73. Sharma, A., Yilmaz, A., Marsh, K., Cochrane, A., and Boris-Lawrie, K. (2012) Thriving under stress: Selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eiF4E translation activity. *PLoS Pathog.* **8**, e1002612

74. Carvajal, F., Vallejos, M., Walters, B., Contreras, N., Hertz, M. I., Olivares,

E., Cáceres, C. J., Pino, K., Letelier, A., Thompson, S. R., and López-Lastra, M. (2016) Structural domains within the HIV-1 mRNA and the ribosomal protein S25 influence cap-independent translation initiation. *FEBS J.* **283**, 2508–2527

75. Song, Z., Gremminger, T., Singh, G., Cheng, Y., Li, J., Qiu, L., Ji, J., Lange, M. J., Zuo, X., Chen, S.-J., Zou, X., Boris-Lawrie, K., and Heng, X. (2021) The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity. *Nucleic Acids Res.* **49**, 5925–5942

76. Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231

77. Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E. L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., Divakaruni, S. S., LaCotti, C., Barton, S., Tummillo, D., Hosic, A., Edme, K., Albrecht, S., Telesnitsky, A., and Summers, M. F. (2011) NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. *Science (80-.).* **334**, 242–245

78. Gu, W., Wind, M., and Reines, D. (1996) Increased accommodation of nascent RNA in a product site on RNA polymerase II during arrest. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6935–6940

79. Kim, J. B., and Sharp, P. A. (2001) Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *J. Biol. Chem.* **276**, 12317–12323

80. Palangat, M., Meier, T. I., Keene, R. G., and Landick, R. (1998) Transcriptional pausing at +62 of the HIV-1 nascent RNA modulates formation of the TAR RNA structure. *Mol. Cell*. **1**, 1033–1042

81. Amorim, R., Costa, S. M., Cavaleiro, N. P., Da Silva, E. E., and Da Costa, L. J. (2014) HIV-1 transcripts use ires-initiation under conditions where cap-dependent translation is restricted by poliovirus 2A protease. *PLoS One*. 10.1371/journal.pone.0088619

82. Monette, A., Valiente-Echeverría, F., Rivero, M., Cohen, É. A., Lopez-Lastra, M., and Mouland, A. J. (2013) Dual Mechanisms of Translation Initiation of the Full-Length HIV-1 mRNA Contribute to Gag Synthesis. *PLoS One*. **8**, 68108

83. Cheetham, G. M. T., Jeruzalmi, D., and Steltz, T. A. (1999) Structural basis for initiation of transcription from an RNA polymerase- promoter complex. *Nature*. **399**, 80–83

84. Song, Z., Gremminger, T., Singh, G., Cheng, Y., Li, J., Qiu, L., Ji, J.,

Lange, M. J., Zuo, X., Chen, S.-J., Zou, X., Boris-Lawrie, K., and Heng, X. (2021) The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity. *Nucleic Acids Res.* **49**, 5925–5942

85. Nuclear Extraction Protocol | Thermo Fisher Scientific - US [online] https://www.thermofisher.com/us/en/home/references/protocols/cell-and-tissue-analysis/elisa-protocol/elisa-sample-preparation-protocols/nuclear-extraction-method-.html (Accessed June 28, 2021)

86. Liu, Y., Holmstrom, E., Zhang, J., Yu, P., Wang, J., Dyba, M. A., De Chen, Ying, J., Lockett, S., Nesbitt, D. J., Ferré-D'Amaré, A. R., Sousa, R., Stagno, J. R., and Wang, Y.-X. (2015) Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature*. 10.1038/nature14352

87. Singh, G., Rife, B. D., Seufzer, B., Salemi, M., Rendahl, A., and Boris-Lawrie, K. (2018) Identification of conserved, primary sequence motifs that direct retrovirus RNA fate. *Nucleic Acids Res.* **46**, 7366–7378

88. Brady, S., Singh, G., Bolinger, C., Song, Z., Boeras, I., Weng, K., Trent, B., Brown, W. C., Singh, K., Boris-Lawrie, K., and Heng, X. (2019) Virion-associated, host-derived DHX9/RNA helicase A enhances the processivity of HIV-1 reverse transcriptase on genomic RNA. *J. Biol. Chem.* **294**, 11473–11485

89. Savas, S., and Skardasi, G. (2018) The SWI/SNF complex subunit genes: Their functions, variations, and links to risk and survival outcomes in human cancers. *Crit. Rev. Oncol. Hematol.* **123**, 114–131

90. Kadoch, C., and Crabtree, G. R. (2015) Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, 2021

91. Sorin, M., Cano, J., Das, S., Mathew, S., and Wu, X. (2009) Recruitment of a SAP18-HDAC1 Complex into HIV-1 Virions and Its Requirement for Viral Replication. *PLoS Pathog*. **5**, 1000463

92. Ariumi, Y., Serhan, F., Turelli, P., Telenti, A., and Trono, D. (2006) The integrase interactor 1 (INII) proteins facilitate Tat-mediated human immunodeficiency virus type 1 transcription. *Retrovirology*. 10.1186/1742-4690-3-47

93. Cheng, S. W. G., Davies, K. P., Yung, E., Beltran, R. J., Yu, J., and Kalpana, G. V. (1999) c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat. Genet.* **22**, 102–105

94. Hwang, S., Lee, D., Gwack, Y., Min, H., and Choe, J. (2003) Kaposi's sarcoma-associated herpesvirus K8 protein interacts with hSNF5. *J. Gen. Virol.* **84**, 665–676

95. Lee, D., Sohn, H., Kalpana, G. V., and Choe, J. (1999) Interaction of E1 and hSNF5 proteins stimulates replication of human papillomavirus DNA. *Nature*. **399**, 487–491

96. Wu, D. Y., Kalpana, G. V, Goff, S. P., and Schubach, W. H. (1996) Epstein-Barr virus nuclear protein 2 (EBNA2) binds to a component of the human SNF-SWI complex, hSNF5/Ini1. *J. Virol.* **70**, 6020–6028

97. Morozov, A., Yung, E., and Kalpana, G. V. (1998) Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1120–1125

98. Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R., and Goff, S. P. (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science (80-.).* **266**, 2002–2006

99. Cano, J., and Kalpana, G. V. (2011) Inhibition of Early Stages of HIV-1 Assembly by INI1/hSNF5 Transdominant Negative Mutant S6. *J. Virol.* **85**, 2254–2265

100. Parissi, V., Caumont, A., Richard De Soultrait, V., Dupont, C. H., Pichuantes, S., and Litvak, S. (2000) Inactivation of the SNF5 transcription factor gene abolishes the lethal phenotype induced by the expression of HIV-1 integrase in yeast. *Gene*. **247**, 129–136

101. Elliott, J., Eschbach, J. E., Koneru, P. C., Li, W., Puray-Chavez, M., Townsend, D., Lawson, D., Engelman, A. N., Kvaratskhelia, M., and Kutluay, S. B. (2020) Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *Elife*. **9**, 1–56

102. Elliott, J. L., and Kutluay, S. B. (2020) Going beyond Integration: The Emerging Role of HIV-1 Integrase in Virion Morphogenesis. *Viruses*. 10.3390/v12091005

103. Engelman, A. N. (2019) Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. *J. Biol. Chem.* **294**, 15137–15157

104. Kessl, J. J., Kutluay, S. B., Townsend, D., Rebensburg, S., Slaughter, A., Larue, R. C., Shkriabai, N., Bakouche, N., Fuchs, J. R., Bieniasz, P. D., and Kvaratskhelia, M. (2016) HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. **166**, 1257-1268.e12

105. Das, S., Cano, J., and Kalpana, G. V. (2009) Multimerization and DNA binding properties of INI1/hSNF5 and its functional significance. *J. Biol. Chem.* **284**, 19903–19914

106. La Porte, A., Cano, J., Wu, X., Mitra, D., and Kalpana, G. V. (2016) An Essential Role of INI1/hSNF5 Chromatin Remodeling Protein in HIV-1 Posttranscriptional Events and Gag/Gag-Pol Stability. *J. Virol.* **90**, 9889–9904

107. Mathew, S., Nguyen, M., Wu, X., Pal, A., Shah, V. B., Prasad, V. R., Aiken, C., and Kalpana, G. V (2013) *INI1/hSNF5-interaction defective HIV-1 IN mutants exhibit impaired particle morphology, reverse transcription and integration in vivo*, 10.1186/1742-4690-10-66

108. Sorin, M., Yung, E., Wu, X., and Kalpana, G. V. (2006) HIV-1 replication in cell lines harboring INI1/hSNF5 mutations. *Retrovirology*. **3**, 1–15

109. Yung, E., Sorin, M., Pal, A., Craig, E., Morozov, A., Delattre, O., Kappes, J., Ott, D., and Kalpana, G. V. (2001) Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1. *Nat. Med.* **7**, 920–926

110. Yung, E., Sorin, M., Wang, E.-J., Perumal, S., Ott, D., and Kalpana, G. V. (2004) Specificity of Interaction of INI1/hSNF5 with Retroviral Integrases and Its Functional Significance. *J. Virol.* **78**, 2222–2231

111. Mahmoudi, T., Parra, M., Vries, R. G. J., Kauder, S. E., Verrijzer, C. P., Ott, M., and Verdin, E. (2006) The SWI/SNF chromatin-remodeling complex is a cofactor for Tat transactivation of the HIV promoter. *J. Biol. Chem.* **281**, 19960–19968

112. Lesbats, P., Botbol, Y., Chevereau, G., Vaillant, C., Calmels, C., Arneodo, A., Andreola, M. L., Lavigne, M., and Parissi, V. (2011) Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes. *PLoS Pathog.* **7**, e1001280

113. Morozov, A., Yung, E., and Kalpana, G. V. (1998) Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1120–1125

114. Marchand, C., Johnson, A. A., Semenova, E., and Pommier, Y. (2006) Mechanisms and inhibition of HIV integration. *Drug Discov. Today Dis. Mech.* **3**, 253–260

115. Dixit, U., Bhutoria, S., Wu, X., Qiu, L., Spira, M., Mathew, S., Harris, R., Adams, L. J., Cahill, S., Pathak, R., Rajesh Kumar, P., Nguyen, M., Acharya, S. A., Brenowitz, M., Almo, S. C., Zou, X., Steven, A. C., Cowburn, D., Girvin, M., and Kalpana, G. V. (2021) INI1/SMARCB1 Rpt1 domain mimics TAR RNA in binding to integrase to facilitate HIV-1 replication. *Nat. Commun.* 10.1038/s41467-021-22733-9

116. Kessl, J. J., Kutluay, S. B., Townsend, D., Rebensburg, S., Slaughter, A., Larue, R. C., Shkriabai, N., Bakouche, N., Fuchs, J. R., Bieniasz, P. D., and Kvaratskhelia, M. (2016) HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. **166**, 1257-1268.e12

117. Caltabiano, R., Magro, G., Polizzi, A., Praticó, A. D., Ortensi, A., D'Orazi, V., Panunzi, A., Milone, P., Maiolino, L., Nicita, F., Capone, G. L., Sestini, R., Paganini, I., Muglia, M., Cavallaro, S., Lanzafame, S., Papi, L., and

Ruggieri, M. (2017) A mosaic pattern of INI1/SMARCB1 protein expression distinguishes Schwannomatosis and NF2-associated peripheral schwannomas from solitary peripheral schwannomas and NF2-associated vestibular schwannomas. *Child's Nerv. Syst.* **33**, 933–940

118. Pyeon, D., Price, L., and Park, I. W. (2015) Comparative molecular genetic analysis of simian and human HIV-1 integrase interactor INI1/SMARCB1/SNF5. *Arch. Virol.* **160**, 3085–3091

119. Allen, M. D., Freund, S. M. V., Zinzalla, G., and Bycroft, M. (2015) The SWI/SNF Subunit INI1 Contains an N-Terminal Winged Helix DNA Binding Domain that Is a Target for Mutations in Schwannomatosis. *Structure*. **23**, 1344–1349

120. Pham, V. V., Salguero, C., Khan, S. N., Meagher, J. L., Brown, W. C., Humbert, N., de Rocquigny, H., Smith, J. L., and D'Souza, V. M. (2018) HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat. Commun.* **9**, 1–12

121. Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007) Escherichia coli physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749

122. LeMaster, D. M., and Richards, F. M. (1982) Preparative-scale isolation of isotopically labeled amino acids. *Anal. Biochem.* **122**, 238–247

123. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–293

124. Lee, W., Tonelli, M., and Markley, J. L. (2015) NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*. **31**, 1325–1327

125. Johnson, B. A., and Blevins, R. A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*. **4**, 603–614

126. Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J., and Sanjuán, R. (2015) Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biol.* 10.1371/journal.pbio.1002251

127. Schneider, G. (2018) Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97–113

128. Patwardhan, N. N., Cai, Z., Umuhire Juru, A., and Hargrove, A. E. (2019) Driving factors in amiloride recognition of HIV RNA targets. *Org. Biomol. Chem.* **17**, 9313–9320

129. Hargrove, A. E. (2020) Small molecule-RNA targeting: Starting with the fundamentals. *Chem. Commun.* **56**, 14744–14756

130. Patwardhan, N. N., Cai, Z., Umuhire Juru, A., and Hargrove, A. E. (2019)

Driving factors in amiloride recognition of HIV RNA targets. *Org. Biomol. Chem.* **17**, 9313–9320

Vita

Seth A. Staller was born in Kokomo, Indiana to Steve and Shannon Staller. He attended Western High School in rural Russiaville, Indiana, on the outskirts of Kokomo. He was born curious, always interested in learning more about the world, the universe, the constructs of reality.

During attendance at Indiana University Bloomington, he initially became fascinated with chemistry through learning organic chemistry: he liked pushing those arrows. It was reminiscent of mental activities performed in video games, which he loved when he was younger. This led him into biochemistry, where he learned about the RNA world hypothesis as an explanation for life's origins on Earth. He earned several awards in his undergraduate career at IUB and graduated with a BS in Biochemistry and a BS in Microbiology.

Due to his commitment to the pursuit of knowledge and his perceived state of the United States economy, Seth decided to go to graduate school to learn more about RNA chemistry. he was awarded the Mizzou NMR Facility Mini-grant in 2020, near the start of the Covid-19 pandemic, which ironically prevented him from effectively using this award to do research on viruses due to restrictions. On July 7th, 2021, Seth successfully defended his Ph.D. in Biochemistry under the supervision of Dr. Xiao Heng.