ABSTRACT

Title of Document: STRUCTURAL ANALYSIS OF RIBOSOME BINDING ELEMENTS OF THE 3' UTR IN TWO POSITIVE-STRAND PLANT RNA VIRUSES

My Tra Le, Doctor of Philosophy, 2016

Directed By: Professor Anne E. Simon, Cell Biology and Molecular Genetics, and
Professor T. Kwaku Dayie, Chemistry and Biochemistry

Turnip crinkle virus (TCV) and *Pea enation mosaic virus* (PEMV) are two positive (+)-strand RNA viruses that are used to investigate the regulation of translation and replication due to their small size and simple genomes. Both viruses contain cap-independent translation elements (CITEs) within their 3′ untranslated regions (UTRs) that fold into tRNA-shaped structures (TSS) according to nuclear magnetic resonance and small angle x-ray scattering analysis (TCV) and computational prediction (PEMV). Specifically, the TCV TSS can directly associate with ribosomes and participates in RNA-dependent RNA polymerase (RdRp) binding. The PEMV kissing-loop TSS (kl-TSS) can simultaneously bind to ribosomes and associate with the 5′ UTR of the viral genome. Mutational analysis and chemical structure probing methods provide great insight into the function and secondary structure of the two 3′ CITEs. However, lack of 3-D structural information has limited our understanding of their functional dynamics.

Here, I report the folding dynamics for the TCV TSS using optical tweezers (OT), a single molecule technique. My study of the unfolding/folding pathways for the TCV TSS has provided an unexpected unfolding pathway, confirmed the presence of Ψ_3 and hairpin elements, and suggested an interconnection between the hairpins and pseudoknots. In addition, this study has demonstrated the importance of the adjacent upstream adenylate-rich sequence for the formation of H4a/Ψ_3 along with the contribution of magnesium to the stability of the TCV TSS.

In my second project, I report on the structural analysis of the PEMV kl-TSS using NMR and SAXS. This study has re-confirmed the base-pair pattern for the PEMV kl-TSS and the proposed interaction of the PEMV kl-TSS with its interacting partner, hairpin 5H2. The molecular envelope of the kl-TSS built from SAXS analysis suggests the kl-TSS has two functional conformations, one of which has a different shape from the previously predicted tRNA-shaped form.

Along with applying biophysical methods to study the structural folding dynamics of RNAs, I have also developed a technique that improves the production of large quantities of recombinant RNAs *in vivo* for NMR study. In this project, I report using the wild-type and mutant *E.coli* strains to produce cost-effective, site-specific labeled, recombinant RNAs. This technique was validated with four representative RNAs of different sizes and complexity to produce milligram amounts of RNAs. The benefit of using site-specific labeled RNAs made from *E.coli* was demonstrated with several NMR techniques.

STRUCTURAL ANALYSIS OF RIBOSOME BINDING ELEMENTS OF THE 3′ UTR IN TWO POSITIVE-STRAND PLANT RNA VIRUSES

By

My Tra Le

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Anne E. Simon, Chair
Associate Professor T. Kwaku Dayie
Professor Jonathan Dinman
Professor Jeffrey DeStefano
Professor James N. Culver, Dean's Representative

Dedication

I would like to dedicate this thesis to my family, especially my mother and my father, who have taught me of kindness and hard-working, my sister, who always believes in her little sister, my husband Chanh Kieu, for immense support throughout my studies, and my two sons An Kieu and Bach Kieu, who have brought joy and happiness to my life.

Acknowledgements

First and foremost, my deep gratitude goes to my two advisors, Dr. Anne Simon, and Dr. Kwaku Dayie, who expertly guided me through my graduate education. I would like especially thank Dr. Simon who allowed me to pursue my two exciting projects, and who has patiently taught me to write. I also would like to thank Dr. Dayie for sharing his expertise in NMR and for his vigorous training and the endless questions that always made me think.

A special thanks to my committee members Dr. Culver, Dr. Destefano and Dr. Dinman for their support of my pursuit of these two very challenging projects. I thank all the members of the Simon lab, the Dayie lab and the Seog lab past and present. They have been my friends and my supporters. Thank to Amy Lee, Adam Kracz for helping me in setting up the optical tweezers. Thanks to Dr. Thakur for teaching me to grow bacterial cells. Thanks to Regan LeBlanc for their expertise in NMR experiments. Thank Dr. Bin Chen for guidance in setting-up NMR experiments and SAXS data analysis. Thank Andrew Longhini for the numerous scientific discussions and for always making nucleotides for my experiments. Thank to my amazing undergraduate-student, Rachel Brown for your diligence in working RNA expression in bacterial cells. Special thanks to Vera Stupina, Micki Kuhlmann, Jared May, Owen Becette, Hyeyeon Nam, Leanna Hengst and Andrew Longhini for proof-reading my writing. Thank to Megan Young, Xue Feng Yuan, Rong Guo, Maitreyi Chattopadhyay, and Feng Gao for discussions about science and life. It has been my great pleasure to work with you, and have you as my friends.

Thanks my collaborator, Dr. Seog for his assistance in single molecule techniques. I would like to thank the Shapiro group in NCI who shared with me their expertise in computer modelling and single molecule dynamics. I also feel a debt to Dr. Fang, Dr. Lixin, Dr. Xiaobing Zou for helping collect my SAXS data.

Finally, I would like to thank my husband who has supported me throughout my grad-student life and helped to take care of my two sons. Thank for his love, patience, and understanding. Thanks to my two sons who endless say they love me and tell me their jokes to make me laugh. Also I would like to thank my parents-in-laws who helped me take care of the family when I was not able to be at home. I would like to thank my big sister and my grandmother who raised me and who always allowed me to pursue my dream to become a scientist.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

(A-)rich sequence	Adenylate -rich sequence
+	Positive
2-D	2-dimentional
3-D	3-dimentional
AD	Anti digoxigenin
BTE	Barley yellow dwarf virus-like element
BYDV	*Barley yellow dwarf virus*
CCFV	*Cardamine chlorotic fleck virus*
CGSC	Coli genetic stock center
CIRV	*Carnation Italian ringspot virus*
CITE	Cap-independent translational enhancer
CP	Coat protein
CrPv	*Cricket paralysis virus*
DIG	Digoxigenin
Dmax	Maximum particle size
DNAzymes	RNA-cleaving deoxyribozymes
DSS	^{1}H sodium 2,2-dimethyl-2-silapentane-5-sulfonate
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic initiation factors
EMCV	*Encephalomyocarditis* virus
EMSA	Electrophoretic mobility gel shift assays
FEC	Force extension curve
Fluo	*Bacillus anthracis* fluoride-binding riboswitch aptamer
FoXs	Fast X-Ray Scattering
HCV	*Hepatitis C virus*
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear single quantum correlation spectroscopy
IF	Initiation factor

IRES	Internal ribosome entry site
ISS	I-shaped structure
ITAP	IRES trans-activating factor
K_d	Dissociation constant
kl-TSS	kissing-loop tRNA-shaped structure
LB	Luria-Bertani-Miller
LMR	LeMaster-Richards
m^7Gppp	7-methylguanosine cap
MD	Molecular dynamic
Mg^{2+}	Magnesium
MNeSV	*Maize necrotic streak virus*
MNSV	*Melon necrotic spot virus*
MOPS	(3-(N-morpholino)propanesulfonic acid)
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhausser effect spectroscopy
NSD	Normalized spatial discrepancy
nt	nucleotide
OD	Optical density
ORF	Open reading frame
OT	Optical tweezers
PABP	Poly(A) binding protein
PAGE	Polyacrylamide Gel Electrophoresis
PBD	Protein bank data
PCR	Polymerase chain reaction
PDDF	Pair distance distribution function
PEG	Polyethylene glycol
PEMV	Pea enation mosaic virus
Phe-tRNA	Phenylalanine tRNA
pN	picoNewton
ppm	Part per million
Pr	Core promoter
PSIV	*Plauti stalin intestine virus*

PTE	*Panicum mosaic virus*-like translational enhancer
PV	*Poliovirus*
RdRp	RNA-dependent RNA polymerase
Rg	Radius of gyration
rNTPs	ribonucleotides
Rp	Ribosomal protein
RSE	Readthrough stimulatory element
SAM	S-adenosylmethionine
satC	Satellite RNA C
SAXS	Small angle X-ray scattering
SCV	*Saguaro cactus virus*
sg	Subgenomic
SHAPE	Selective 2´ hydroxyl acylation analyzed by primer extension
SL	Stem loop
SPG	Studier phosphate buffer
STNV	*Satellite tobacco necrosis virus*
TBSV	*Tomato bushy stunt virus*
TCV	Turnip crinkle virus
TED	Translation enhancer domain
tktA	Mutant transketolase *E.coli* strain
TSS	tRNA-shaped structure
UTR	Untranslated region
WLC	Worm like chain model
wt	widetype
YSS	Y-shaped structure
Ψ or PK	Pseudoknot

Chapter 1: Introduction

1.1. Viral RNA structures modulating translation initiation and ribosome interactions

Positive (+)-strand RNA viruses that directly use their RNA genomes for translation account for one third of all virus genera (1). The significant loss of human and animal life, as well as destruction of agricultural crops attributed to these genera are perpetrated by members such as hepaviruses, coronaviruses, flaviruses, and tombusviruses (2). Protein production in these viruses depends on expression of the viral genes using cellular ribosomes, in which translation initiation is a determinant step for the efficiency of translation. Due to the inherent size limitations of viral particles, (+)-strand RNA viruses contain compact genomes that encode for a small number of proteins and rely on RNA structural elements to regulate translation initiation. An increasing number of *cis*- and *trans*-acting RNA elements in (+)-strand RNA viral genomes have been found to function in translation initiation and ribosome interactions (2-9). Although many functionally important viral RNA structural elements have been identified to date, there is a paucity of critical structural information regarding these elements, which could provide valuable insight into their functions. Since RNA structural elements involved in translation initiation are critical for efficient viral propagation, they are the focus of this study.

1.1.1. Translation initiation

Translation is the process in which the genetic information in RNA is decoded into an amino acid sequence of proteins using cellular ribosomes. This process is often separated into three well defined phases: initiation, elongation and termination (10).

Among these three phases, initiation displays the greatest diversity of mechanisms among different organisms (3, 5). Whether under highly variable or limited environmental conditions, organisms tend to adopt optimal mechanisms for translation initiation in order to produce their proteins efficiently.

In general, translation initiation mechanisms are broadly categorized as either cap-dependent or cap-independent (2-5). Most eukaryotic mRNAs follow a cap-dependent mechanism for translation initiation, whereby the placement of a single 7-methylguanosine cap (m^7Gppp; Figure 1.1) directs the assembly of the ribosomal machinery on the 5´ capped mRNA (3, 10). The accessibility of the m^7Gppp to initiation factors (IFs), such as eIF4E, the cap-binding protein, directly determines the translation efficiency of mRNAs. For instance, a stable hairpin placed in the proximal region of the 5´ cap can block the interaction of IFs or ribosome scanning, which can inhibit translation (11-13). In the cap-dependent mechanism (Figure 1.2), translation initiation starts with the interaction of the small 40S ribosomal subunit with eukaryotic initiation factors 3 (eIF3) and eIF1A. This complex then binds the ternary complex containing eIF2-GTP and Met-tRNA to create the larger 43S complex. Association of the 43S complex with eIF4A helicase and eIF4F (composed of eIF4E and eIF4G), which interacts with capped mRNA through eIF4E, assists in the assembly of the 43S complex on capped mRNA (3). This process is facilitated by circularization of the mRNA, which is achieved by the interaction between the cap, eIF4E, the poly(A) binding protein (PABP), poly(A) tail of the mRNA and eIF4G (14). The 43S complex then scans mRNA in the 5´ to 3´

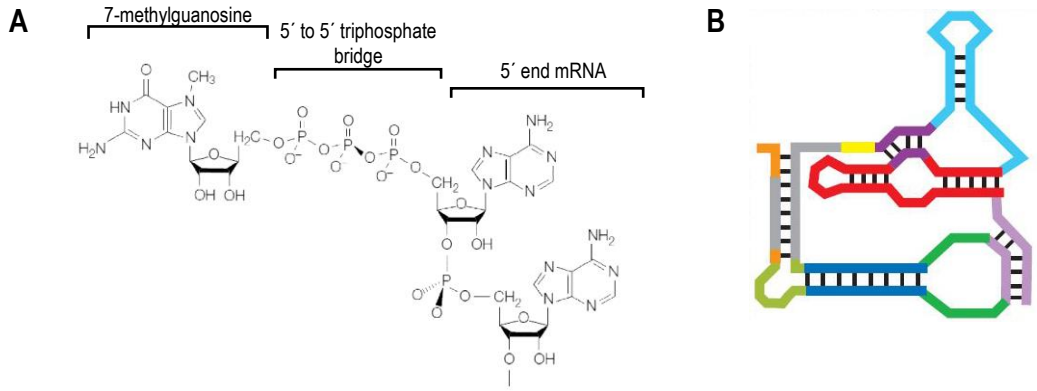


Figure 1.1. The simple structure of the 7-methylguanosine cap of cap-dependent translation versus the complex RNA structure of an internal ribosome entry site (IRES)
A. A 5´ cap is made of an inverted guanosine, which forms a 5′ to 5´ triphosphate linkage with the 5´ terminal nucleotide of the mRNA (15). **B.** Schematic secondary structure of the IRES of *Plauti stalin intestine virus* (PSIV) includes several stem-loops and pseudoknots (16).

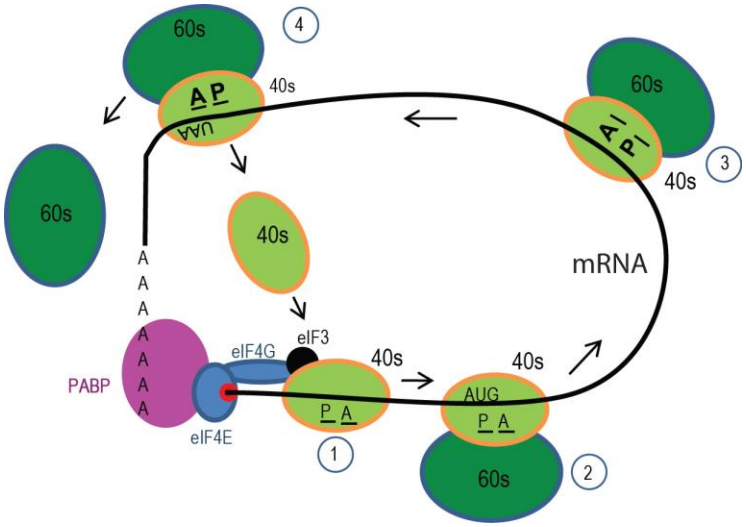


Figure 1.2. Model of cap-dependent translation
Translation includes three phases: initiation (step 1), elongation (steps 2 and 3) and termination (step 4). Translation initiation is supported by closed-loop mRNA (14). In cap-dependent translation, the closed-loop structure is formed through PABP (purple) creating a bridge between the poly(A) tail and an initiation factor eIF4E (blue), which interacts with the 5´ cap (red). Initiation begins with the interaction of eIF4F via eIF4E with the 5´ cap and the PAPB to assist the assembly of 43S complex on mRNA. The 43S complex includes 40S ribosomal subunit (lighter green), eIF3, and a ternary complex (not shown). Once assembled on the mRNA, the complex scans until it finds the first AUG in good context where it joins the 60S ribosomal subunit. Translation moves into elongation step until the ribosome reaches the stop codon where the ribosome dissembles and translation terminates. Image modified from Dreher et.al 2006 (17).

direction to find the first "start" codon in an optimal context (3). The translation initiation stage is completed after the large 60S ribosomal subunit joins the 43S complex at the

start codon to form the 80S ribosomal complex (3). Translation proceeds to the elongation stage in which amino acids are sequentially added to the growing polypeptide chain based on the encoded mRNA sequence. Protein synthesis terminates when the ribosome reaches a "stop" codon. The ribosome is then disassembled and recycled for the next round of translation (3).

Under adverse conditions, such as stress or starvation where caps or eIFs are limited or inhibited, translation initiation of a small subset of mRNAs can occur by a cap-independent mechanism (18, 19). Cap-independent translation is prevalent in (+)-strand animal and plant viruses since such mechanisms allow viral RNAs to compete efficiently with host mRNAs for the translational machinery (3-5, 7, 20). One common feature displayed by RNA templates that utilize cap-independent translation is the presence of structural elements, which can range from small and simple to large and complex (Figure 1.3) (5, 7, 20). Such elements appear to functionally replace the cap or the entire translation initiation complex by providing a scaffold for ribosome assembly (4, 5, 7, 20). A closed-loop structure formed by either a protein bridge or RNA:RNA interaction between RNA termini has also been suggested to assist cap-independent mechanisms of translation initiation (17, 21).

Two representative cap-independent translation mechanisms include the "internal ribosome entry site" (IRES) mechanism (Figure 1.3A), more commonly found in animal viruses and a small group of cellular mRNAs (19), and the "cap-independent translation enhancer" (CITE) mechanism (Figure 1.3B), which is usually found in plant viruses (5). Both IRESes and CITEs are RNA structural elements that can interact with translation initiation factors or directly bind to ribosomes, allowing assembly of translational

machinery on mRNAs. IRESes are typically located internally within the 5´ untranslated regions (UTRs) of RNAs. CITEs are mostly found within the 3´ UTRs of viral genomes. Different RNA sequences found in IRESes and CITEs tend to form conserved structures that can be categorized into structural subgroups. Current knowledge detailing 2-dimentional (2-D) or 3-D structural organization of these cap-independent RNA elements will be reviewed in the following sections.

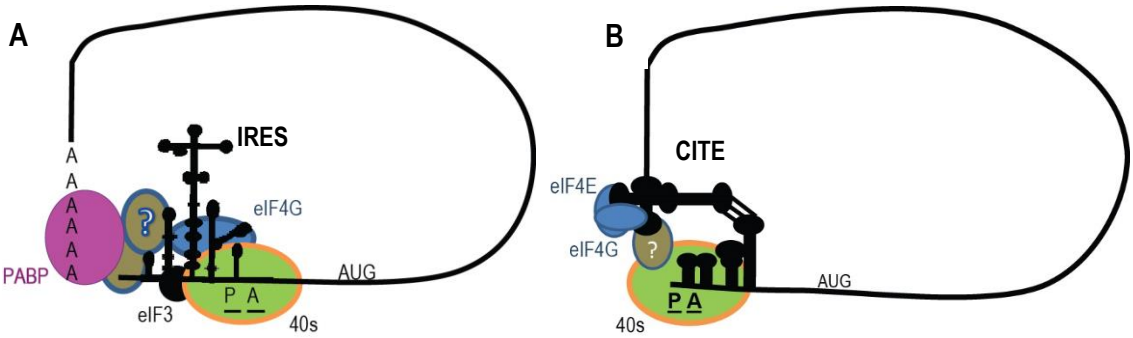


Figure 1.3. Examples of models of cap-independent translation initiation
Close-loop RNA structure proposed to form either by a protein bridge or RNA:RNA interaction to assist translation initiation in cap-independent. **A.** Model of translation initiation using internal ribosomal entry sites (IRES) in poliovirus, which have uncapped, poly-adenylated genomes. The IRES is complex and located in the 5´ UTR of the viral genome; assembly of the initiation complex is supported through the direct interaction of 5´ RNA structural elements with translation initiation factors and ribosomal subunits. PAPB is suggested to support the close-loop RNA structure (21). **B.** Model of translation using 3´ cap-independent translational enhancers (CITEs) in plant viruses. In many plant viruses, RNA structural elements located at the 3´ UTR of viral genome recruit translation initiation factors or ribosome/ribosome subunits. Circularizing of the genome through a long-range RNA:RNA interaction between 3´ CITEs and e 5´ hairpins or ribosome bridge assists in transferring of the translational machinery to the 5´ end to enhance translation of the virus (17, 22, 23). Images modified from Martinez-Salas et al, 2008 (21), Dreher et al, 2006 (17).

1.1.2. IRESes and their 3-dimentional (3-D) structures

IRESes were first discovered in two piconaviruses: *poliovirus* (PV) and *Encephalomyocarditis* virus (EMCV) in 1988 (24, 25). Since then, IRESes have been identified in at least 39 additional (+)-strand RNA viruses that infect mammals, invertebrates and plants (19, 26). IRES-dependent translation initiation mechanisms

allow many viruses to efficiently hijack the translation machinery of host cells. Additionally, 85 cellular mRNAs have been reported to use IRES elements for protein translation when cap-dependent translation initiation is compromised or inhibited (19, 26). IRESes are often found in the 5´ UTRs of mRNAs and contain structural elements that are highly efficient in recruiting initiation factors and ribosomal subunits to initiate translation in a cap-independent manner. Although IRESes are found in both viral and cellular mRNAs, they have different core structures. While viral IRESes tend to have conserved structures, little structural conservation has been found in IRESes of cellular mRNAs (27).

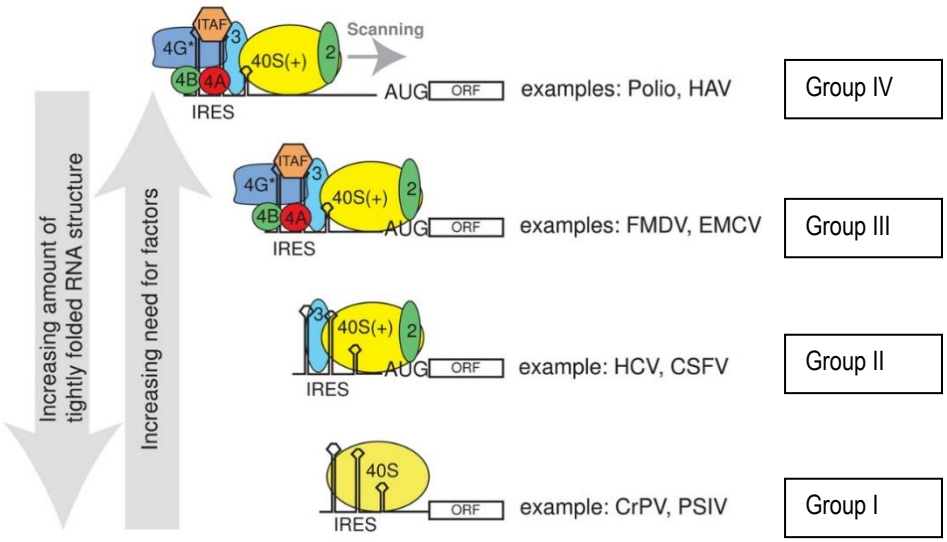


Figure 1.4. Diagram of four groups of viral IRESes recruiting translational machinery
Tightly folded IRES requires fewer elements of translational machinery. Compact group I IRESes require only the 40S ribosomal subunit. Group II IRESes require eIF2 and eIF3 in addition to the 40S ribosomal subunit. More translation initiation factors, including eIF2, eIF4A, eIF4B, eIF4G bind to extended IRESes of group III and IV. These two groups also require the presence of additional proteins called IRES trans-activating factors (ITAPs) (7).

Viral IRES elements are diverse in size, sequence and the requirements for trans-acting factors (8, 28). Highly structured IRESes can directly bind various translation initiation factors but they do not interact with the cap-binding protein, eIF4E (9, 21).

Interestingly, in viruses, tightly folded IRESes found in dicistroviruses, flaviviruses require fewer proteins to support their interaction with ribosomes while extended and largely flexible IRESes found in piconaviruses require more proteins including translation initiation factors and IRES trans-acting factors (Figure 1.4) (7).

Viral IRES elements can be quite large, with piconavirus IRESes ranging from 280 to 460 nt (9, 28). Due to the large size and complexity of IRES elements, the precise mechanisms by which they regulate translation initiation remain unclear. Structural knowledge of IRESes mostly comes from studies using picornaviruses, *Hepatitis C virus* (HCV), and dicistroviruses.

1.1.2.1. Group I IRES

Viral IRESes are classified into four groups based on conserved structures and requirements for initiation factors (Figure 1.4) (7, 21, 28, 29). Group I IRESes are found in dicistroviruses, and are the only IRES group located in the intergenic region (IGS) between two open reading frames (ORFs) (30). Highly structured group I IRESes require neither initiation factors nor an initiator methionyl tRNA. Group I IRESes can bind directly to the ribosome and translation initiates at the ribosome A-site rather than the P-site, which is used during canonical translation initiation (21, 31, 32). The length of IGS varies among Dicistroviruses, however secondary and tertiary structures of their IRESes are highly conserved and contain three different regions (1, 2, and 3). Regions 1 and 2 fold into a compact domain that is structurally independent from the variable region 3 (domain 3) (Figure1.5A) (33-35). Each region of group I IRESes is stabilized by a specific pseudoknot (PK), (PKI, PKII and PKIII) (Figure 1.5A) (34). Region 1 binds the 40S ribosomal subunit while region 2 binds the larger 60S ribosomal subunit (16, 33, 35,

36). This is the only IRES group for which a high-resolution structure is available for both ribosomal subunit bound and unbound (16, 37, 38). In the crystal structure of the *Plauti stali intestine virus* (PSIV) IRES (solved at 3.1 Å) (Figure 1.5B), the two conserved and distal IV and V stem-loops (SL) of region 1, responsible for direct binding to the 40S ribosome subunit, are found in close proximity and extend away from the folded IRES (16). Region 2 folds along the side of region 1 and interacts with the 60S ribosomal subunit (16). Region 3 is important for the initiation step. In PSIV,

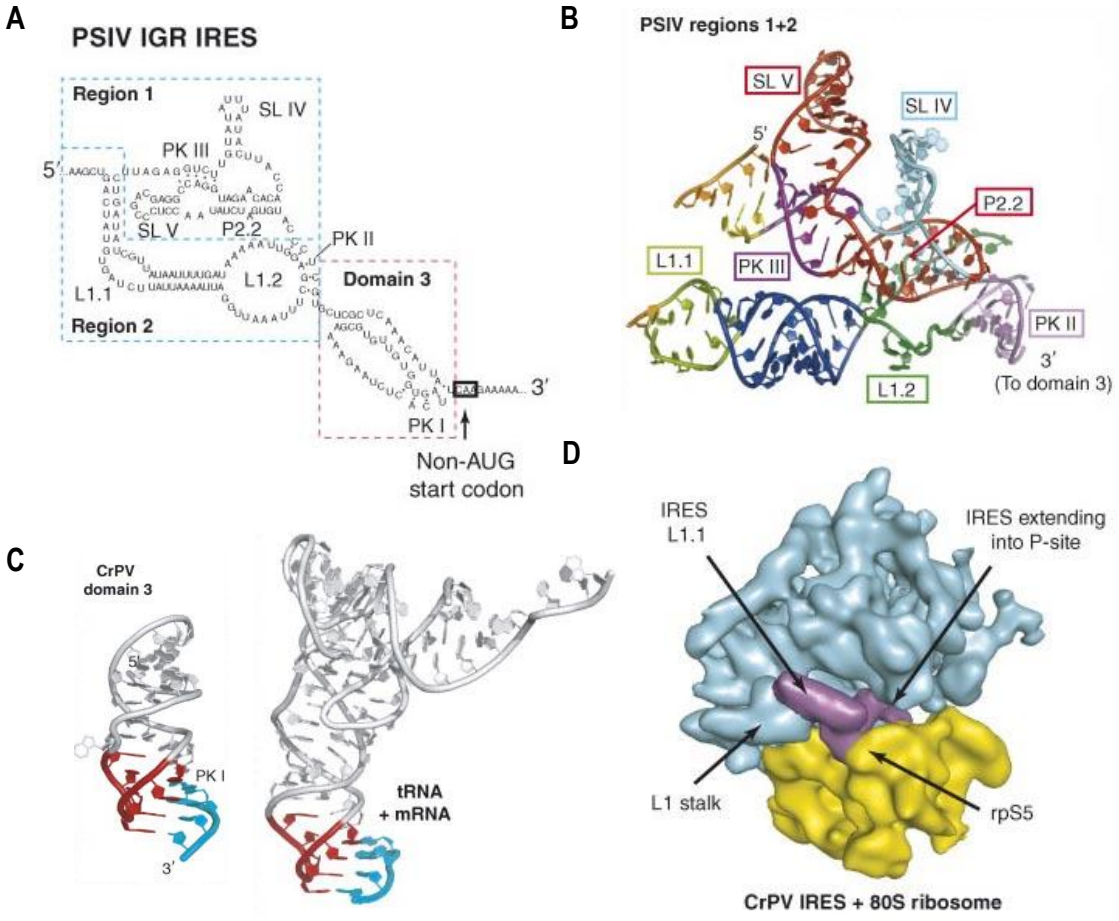


Figure 1.5. Model structure of group I IRES
A. Secondary structure of PSIV IRES. The structure of PSIV IRES is divided into three different regions (domains) in which each region is stabilized with one corresponding pseudoknot (PK). **B.** Crystal structure of region I and II of PSIV IRES. Stem-loop IV and V which are important for the interaction with the 40S ribosomal subunit are found in close proximity. Region 2 folds along the side of region 1 to support the interaction of region 2 with the 60S ribosomal subunit. **C.** The 3-D crystal structure of domain III of CrPV. Domain III mimics the interaction of tRNA with mRNA. **D.** The cryo-EM of CrPV IRES with 80S

ribosome. IRES is in purple, the 40S ribosomal subunit is in yellow and the 60S ribosomal subunit is in blue. IRES is found to extend to the P-site of the ribosome. Location of ribosomal L1 stalk and rpS5 are indicated (20). Images adapted from Jeffery S. Kieft 2008 (20).

mutations/deletions that disrupt base-pairing between PKI with a CUU of region 3 impair translation (Figure 1.5A, C) (39). Chemical structure probing of the PSIV IRES in the presence of the ribosome indicates that the interaction between PK I and CUU of region 3 lies in the P-site of the ribosome to position the initiation codon in the A-site, thus supporting initiation of translation at the A-site (35). Translation of group I IRES often initiates at a non-AUG codon including GCC, GCU, GCA, or CAA, which normally encode alanine or glutamine (21, 40). The crystal structure of the *Cricket paralysis virus* (CrPv) IRES (Figure 1.5C) indicates that region 3 mimics the intermolecular interaction between the initiator tRNA anticodon loop and the AUG start codon, supporting translation initiation at a non-AUG start codon (41). A structural comparison of the IRES with or without bound 80S ribosomes (Figure 1.5B-D) suggests that IRES domains fold prior to ribosome binding and, once bound to the ribosome, the structure of the group I IRES does not undergo a global structural rearrangement rather it slightly shifts local angles between helices of each region (16).

1.1.2.2. Group II IRES

An example of a group II IRES is the 5´ UTR of *Hepatitis C virus* (HCV), a member of the *Flaviviridae* family, in which viral RNA genomes lack a 5´ cap and 3´ poly(A)-tail (42, 43). HCV-like IRESes are also found in some piconaviruses (9). Translation initiation with group II IRESes involves binding of the 40S ribosomal subunit and interaction with the initiation factor eIF3 and ternary complex [eIF2-GTP-Met-

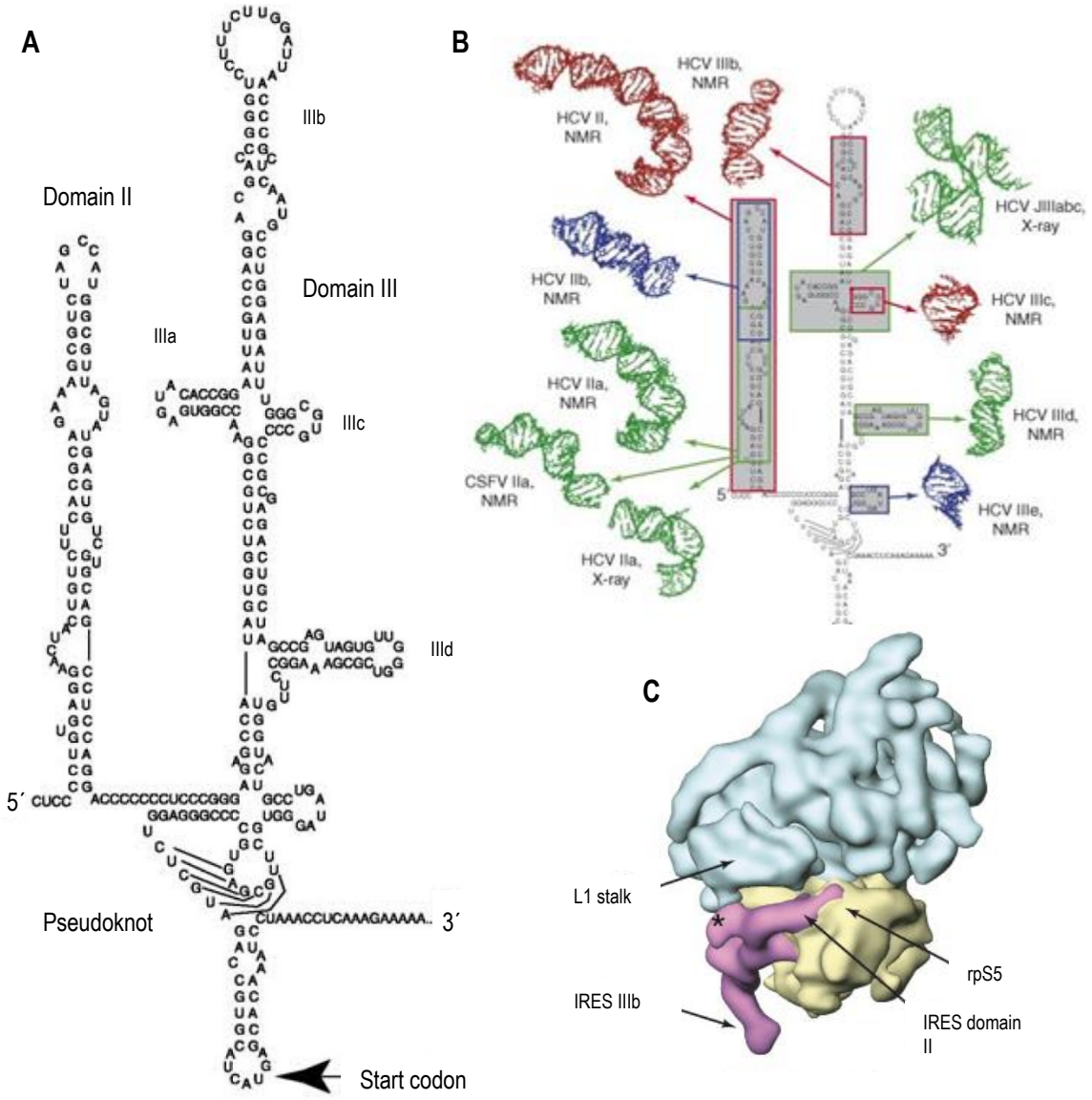


Figure 1.6. Model structure of group II IRES
A. Secondary structure of the HCV IRES. Arrangements of four domains are indicated. Domain IV is supported by a pseudoknot. **B**. Fragments of domain II and III have been solved by NMR and X-ray crystallography. **C**. Cryo-EM complex of HCV IRES (magenta) with 80S ribosome (yellow and blue). Locations of IRES domain II and III, ribosomal L1 stalk and rpS5 are indicated. (*) denotes contact between the HCV IRES and L1 stalk (7). Images adapted from Jeffery S Kieft 2008 (20).

tRNA] (20, 21, 44). Biochemical studies have established the secondary structure of the HCV 5´ UTR with one PK and four long stem-loops, which are divided into four domains (Figure 1.6A), with domains II–IV harboring RNA elements that interact with the

ribosomal subunit and translation initiation factors to facilitate translation (42, 43, 45). Domain II is an extended stem-loop that interacts with the 40S ribosomal subunit at the E-site and is important for tRNA binding while domain IV inserts its nucleotides into the mRNA-binding cleft of the 40S ribosomal subunit (18). High-resolution structures of truncated fragments of group II IRESes have been solved (Figure 1.6B). Structural analyses of nine fragments of HCV IRES domains II and III indicate that, in its unbound state, the HCV IRES adopts an extended conformation that contrasts the compact, globular conformation of group I IRESes (Figure 1.6B) (7). In addition, the extended conformation allows group II IRESes to interact with the ribosome at multiple sites, such as with the ribosomal L1 stalk, which is a flexible structure and responsible for tRNA movement within the ribosome during translocation (Figure 1.6C) (46, 47). The cryo-EM structure of the ribosome-bound HIV IRES indicates that, upon binding, there are almost no structural changes within the IRES and conformation of the free-form IRES is maintained.

1.1.2.3. Group III and IV IRESes

Group III and IV IRESes are common in the piconaviridae family, and are typically ~450 nt in length. These two groups of IRESes share an extended and flexible structure organized into multiple domains that interact with initiation factors eIF4A, eIF3, and the C-terminus of eIF4G (Figure 1.7). In addition, both IRES groups require IRES trans-activating factors (ITAFs) that possibly act as RNA chaperons; however, the actual functions of ITAFs with regard to group III and IV IRESes are still unknown (Figure 1.4) (28, 48). Group III IRESes, typified by *Encephalomyelitis virus* (ECMV) (Figure 1.7A), are found in genomic RNAs of cardio- and aphthovirus, while group IV IRESes, typified

by *poliovirus* (Figure 1.7B), are found in genomic RNAs of enteroviruses and rhinoviruses. Both groups contain an oligopyrimidine sequence located upstream of the initiation start codon AUG (Figure 1.7), where the distance between the oligopyrimidine sequence and the initiation start codon AUG is critical for translation initiation of EMCV (49). In viral genomes that contain group III IRESes, the IRES structures are

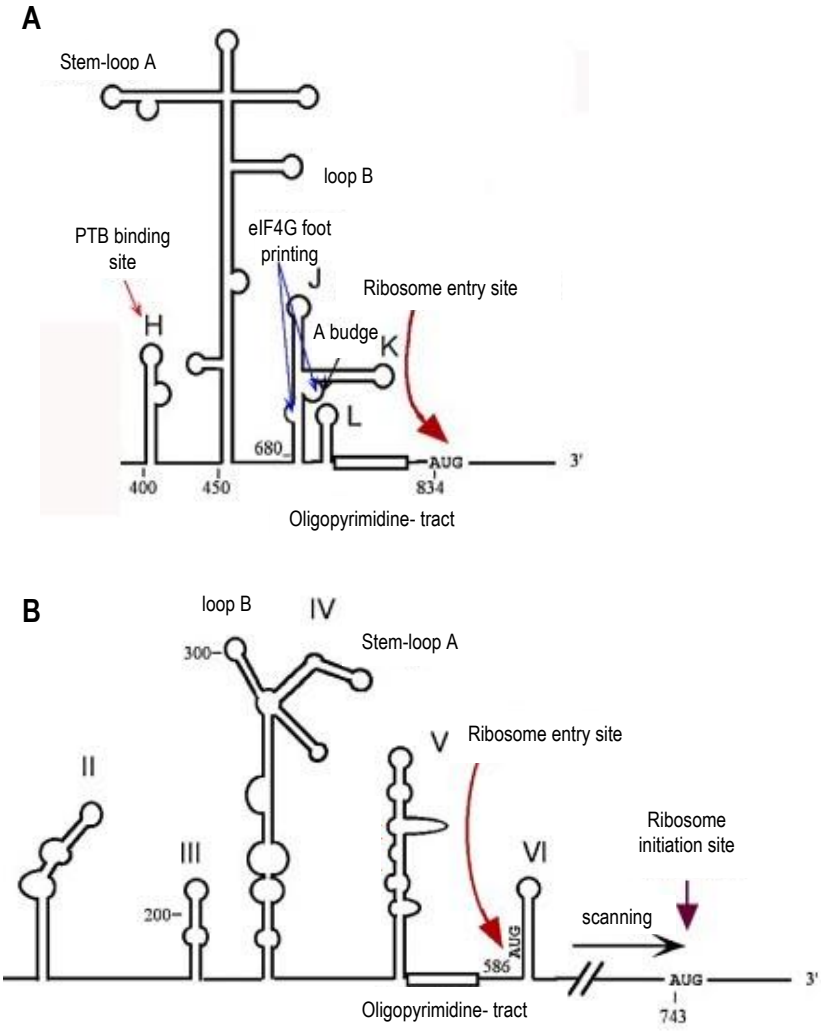


Figure 1.7. Schematic diagram of group III and IV IRESes
A. Group III IRES of EMCV. Ribosome is assembled directly at the AUG start codon. **B.** Group IV IRES in poliovirus. IRES positioned some distance away from the initiation site. Once assembled on the IRES, the ribosome scans until it reaches the AUG start codon. The binding sites of translation initiation factors and ribosome are indicated by arrows (7, 8). Images adapted from Balvay et al, 2009 (8).

located immediately upstream of an AUG start codon (Figure 1.7A), allowing the translational machinery to be assembled directly on the start codon. However, in viral genomes that contain group IV IRESes, the IRESes are located further upstream from the AUG start codon (Figure 1.7B). Thus, once assembled, the ribosome is required to scan down the RNA genome until it reaches the initiation AUG start codon (7, 8). So far, no 3-D structures are available for group III and IV IRESes owing to their large size, complexity and presence of extensive flexible regions.

1.1.3. Current structural knowledge of cap-independent translation enhancers (CITEs) and their interacting elements to facilitate translation initiation

Cap-independent translation enhancers (CITEes) were first discovered in *Satellite tobacco necrosis virus* (STNV) and later found in plant viruses from the *Luteoviridae* and *Tombusviridae* and *Umbravirus* genus (5, 50). CITEs differ from IRES elements in that they are located within or near the 3′ UTR of viral RNA genomes and act to enhance translation at the 5′ terminus. 3′ CITEs commonly contain RNA structural elements that can form long-range RNA:RNA interactions to assist in genome circularization by bringing the two ends in close proximity. Many of these 5′ and 3′ interacting sequences are conserved and are normally present in apical loops of hairpins (2, 4, 17). Unlike long multidomain IRESes, multiple functional RNA sequences of 3′ CITEs commonly cluster within one RNA structural domain, which allows the structure to simultaneously perform multiple functions. For instance, the stem of a 3′ CITE can recruit the translation machinery while its apical loop can simultaneously perform the long-range RNA:RNA interaction. These simultaneous interactions facilitate delivery of either the

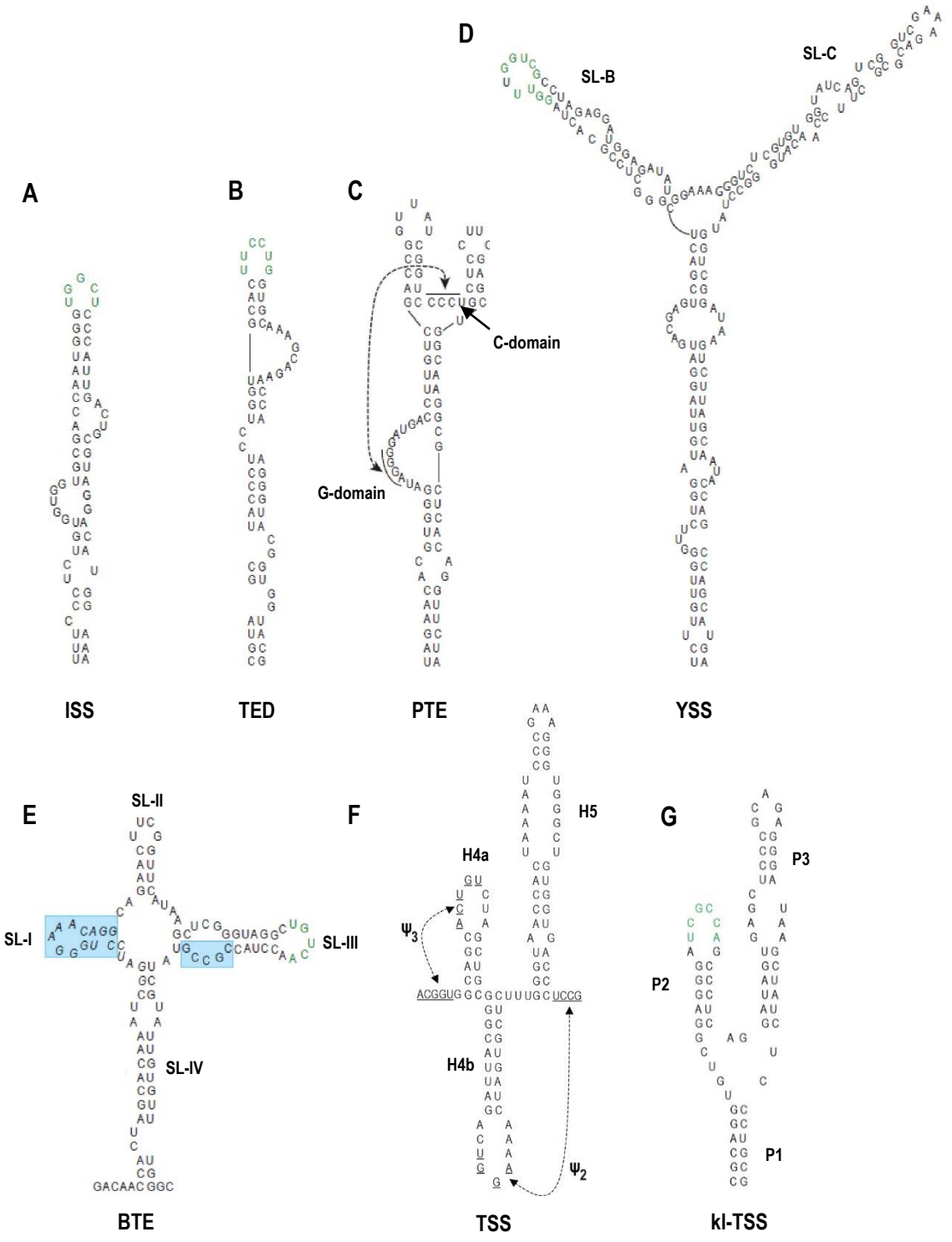


Figure 1.8. Sequences and secondary structures of 7 types of 3´ CITEs in plant viruses
A. ISS of MNSV-64. **B.** TED of STNV. **C.** PTE of PEMV. **D.** YSS of TBSV. **E.** BTE of BYDV. **F.** TSS of TCV. **G.** kl-TSS of PEMV. Sequences of known or predicted to form long-range RNA:RNA interactions are in green. The conserved 17-nt sequence of BTE is boxed in blue. Sequences that form pseudoknots are connected with an arrow. Images adapted and modified from Simon et al, 2013 (5).

3´bound initiation factors or ribosome to the 5´end of the genome (5, 51). Another common feature of CITEs is that they generally recruit either eIF4F, and/or subunits eIF4E and eIF4G, or/and ribosome subunits for their translational enhancing activities. Although they share common functions, seven structurally distinct classes of CITEs have been categorized so far based on their size, sequence, and structure (Figure 1.8) (4, 5, 52).

1.1.3.1. I-shaped structure (ISS)

The smallest and simplest of the 3´ CITEs is found in *Maize necrotic streak virus* (MNeSV) of tombusviruses (53) and *Melon necrotic spot virus* (MNSV) of carmoviruses (54). This type of 3´ CITE is predicted to form an I-shaped structure (ISS) with multiple internal stems separated by internal loops (Figure 1.8A) (55). The conserved interacting sequence of the ISS is located in the apical loop of the element, and is complementary to a sequence located in the 5´ UTR. The MNeSV ISS requires an intact eIF4F for tight binding (K_d = 190 nM), which is mediated by the eIF4E component (55). The MNeSV ISS can form a tripartite complex with the 5´ UTR and eIF4F *in vitro*, supporting the hypothesis that interaction between the 5´ UTR and the MNeSV ISS modulates delivery of 3´ bound initiation factors to the 5´ terminus (55).

1.1.3.2. Translation enhancer domain (TED)

Translation enhancer domain (TED), the first 3´ CITE discovered, was originally found in *Satellite tobacco necrosis virus* (STNV) (50, 56). The structure of the 120-nt TED was first computationally predicted and later confirmed by genetic analysis. Similar to the ISS, the TED is comprised of a long stem divided into multiple internal stems separated by bulges (Figure 1.8B). TED interacts with eIF4G with high affinity (K_d = 30

nM) (56, 57). Deletion of TED from the 3´ UTR of STNV reduces translation of reporter constructs by more than 20-fold and translational activity can only be rescued by the addition of a 5´ cap (56). Long-range RNA:RNA interaction of TED with 5´ sequences is likely but has not been confirmed since mutation of putative interacting sequences only modestly reduced translation and reestablishing the interaction with compensatory mutations did not restore translation to wildtype level (58).

1.1.3.3. *Panicum mosaic virus*-like translational enhancer (PTE)

Panicum mosaic virus-like translational enhancer (PTE) is another type of 3´ CITE first discovered in *Panicum mosaic virus* and later found in seven carmoviruses and aureusviruses (Figure 1.8C) (59, 60). Unlike other 3´ CITEs and IRESes, which bind eIF4F or eIF4G, the PTE interacts only with eIF4E with high affinity (K_d = 48 nM) (59, 60). The PTE is predicted to form a long basal stem containing a large G-bulge domain branching into two helical stem-loops connected by a C-rich sequence (C-domain) (Figure 1.8C). The C-domain forms a pseudoknot with the G-bulge of the main stem (59, 60). More detailed information about the PEMV PTE is presented in section 1.2. With the exception of the PTE of *Pea enation mosaic virus* 2 (PEMV), the 5' stem-loop of PTEs are predicted to form a long-range RNA:RNA interaction with complementary sequences located in the apical loop of a 5´ proximal hairpin (60). In *Saguaro cactus virus* (SCV), the long-range RNA:RNA interaction between the PTE and its 5´ interacting sequence, located in the coding region, has been confirmed. Insertion of these two SCV elements into a reporter construct enhanced translation by 75-fold (23). Relocation of the 5´ terminal hairpin closer to the 5´ terminus of the SCV genome reduces translation efficiency, indicating that the location of the 5´ terminal hairpin is important for efficient

translational enhancer activity of the PTE (61). Location of the PTEs in carmoviruses is not conserved; a few PTEs span the C terminus of the coat protein coding region while others are located within the 3´ UTR (23).

1.1.3.4. Y-shaped structure (YSS)

The 3´ CITE Y-shaped structure (YSS) was first identified in *Tomato bushy stunt virus* (TBSV) and *Carnation Italian ringspot virus* (CIRV) in the *Tombusvirus* genus (Figure 1.8D) (62, 63). The YSS was later found in all but two tombusviruses. It forms a large three-way branched helix in which all stems are much longer than those of similar PTE. Sequence and structure of the YSS are critical for translational enhancer activity of the 3´ UTR in vivo (62). The apical loop of the 5´ YSS hairpin contains an interacting sequence that is complementary to a 5´ UTR hairpin loop located within a T-shaped structure element (62, 64). The 5´ T-shaped structure element is also required for replication, suggesting a mechanism that regulates the switch between translation and replication (65). Exchanging three CITEs (YSS, ISS and PTE) does not cause significant loss in translation and replication as long as their corresponding RNA:RNA interactions with the 5´ terminal sequences are maintained. Thus, these CITEs are proposed to act as one structural domain to deliver bound translation initiation factors to the 5´ end of the genome (63). In addition to forming a long-range RNA:RNA interaction with the 5´ end, the CIRV YSS specifically binds eIF4F and its isoform, isoeIF4F (63).

1.1.3.5. Barley yellow dwarf virus-like element (BTE)

Barley yellow dwarf virus-like element (BTE) is found in *Luteovirus*, *Diathovirus*, and *Necrovirus* genera as well as some umbraviruses and is one of best-characterized 3´

CITE (Figure 1.8E) (66-70). Phylogenetic comparisons, mutagenesis, and structural probing of the *Barley yellow dwarf virus* (BYDV) BTE has identified a major conserved stem-loop structure as well as a 17-nt conserved sequence [GGAU**CCUG**GgAaA**CAGG]** in which the bold nucleotides form a stable stem (66-69). Computer modeling predicts that the secondary structure of the BTE is comprised of a series of stem-loops radiating from a central hub (Figure 1.8E). The proposed structure of the BTE was later confirmed by structural probing (70). Mutation of the bases in the hub decreases translational activity of the BYDV BTE, suggesting the importance of these bases for BTE function. Interestingly, there are no apparent canonical base-pairs between nucleotides of the hub (67). From phylogenetic analyses of BTE elements in related viruses, the hub is responsible for holding two to five hairpins, which suggests that its nucleotides must have non-canonical base pairs to maintain a stable structure. Interaction between the first three bases of the 17-nt conserved sequence with nucleotides in the opposite sequence extends the basal stem (67). The long-range RNA:RNA interaction sequence in the BTE is located within a stable hairpin loop III (SL-III) (Figure 1.8E) that is positioned outside of the 17-nt conserved sequence (70, 71). Disrupting the long-range RNA:RNA interaction of the BTE reduced translational efficiency, which is restored by compensatory mutations (72). The BTE of BYDV binds to eIF4F via eIF4G to enhance translation (73). Stem-loop I (SL-I) of the BTE was inferred to interact with eIF4F as this sequence was protected by the presence of eIF4F in footprinting experiments. The relocation of BYDV BTE to the 5´ terminus maintains its translation enhancing function, indicating that the element functions in transferring translation machinery to the 5´ UTR (67) .

1.1.3.6. The tRNA-shaped structure (TSS) and the kissing-loop TSS

The 3´ CITE known as the tRNA-shaped structure (TSS) is found in *Turnip crinkle virus* (TCV) and *Cardamine chlorotic fleck virus* (CCFV) and contains a unique set of three hairpins and two pseudoknots (Figure 1.8F) (74). Unlike other 3´ CITEs, which are mostly found to interact with initiation factors, the TCV TSS interacts directly with the 60S ribosomal subunit and 80S ribosomes (75). A long-range RNA:RNA interaction sequence has not been identified for the TCV TSS, suggesting that the connection between the 5´ and 3´ ends in the TCV genome might be established through a ribosome bridge. The TSS TCV is the only CITE whose structure was validated by small angle X-ray scattering (SAXS) and NMR (details in TCV section) (76). Another structural analogue of the TSS, the kissing-loop (kl-) TSS, was later identified in umbravirus *Pea enation mosaic virus* RNA 2 (PEMV) (Figure 1.8G). Unlike the TCV TSS, the kl-TSS of PEMV contains an apical loop that interacts with a hairpin positioned within the coding sequence at the 5´ end (51, 77). More details about the 3´ CITEs of TCV and PEMV are presented in section 1.2.

1.2. *Turnip crinkle virus* and *Pea enation mosaic virus* and their 3´ CITEs as models for structural analysis of ribosome binding translation enhancers

1.2.1. Turnip crinkle virus: a model virus to study the conformational switch of the TCV TSS

Turnip crinkle virus (TCV) belongs to the carmovirus genus in the family *Tombusviridae* and is a model (+)-strand RNA virus with a compact genome of 4053 nt (Figure 1.9). The genome encodes five proteins: p28, a replication accessory protein; p88, the RNA-dependent RNA polymerase (RdRp); p8 and p9, two movement proteins; and p38, a coat protein and silencing suppressor (78). The p28 and p88 proteins are translated

directly from genomic RNA, in which p88 is the read-through extension product of p28. The p8 and p9 are translated from 1.72 kb subgenomic 1 RNA and the coat protein is translated from 1.45 kb subgenomic 2 RNA.

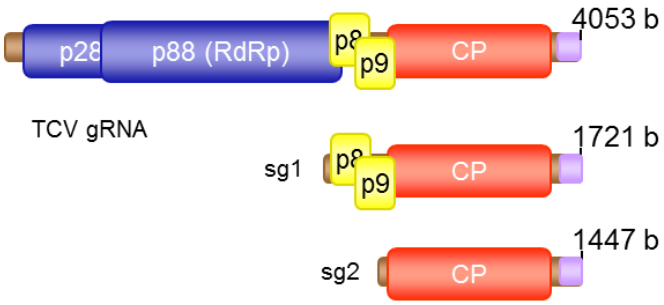


Figure 1.9. Genomic organization of TCV
TCV has one genomic RNA and two subgenomic RNAs (sg1 and sg2). The three RNAs are 3' coterminal. The TSS is in the 3' UTR (74). p28 and p88 are involved in replication, p8 and p9 are involved in movement and the coat protein (CP) participates in encapsidation and RNA silencing suppression.

TCV has been used as a model not only to investigate host-pathogen interactions in plants (79), but also to study RNA structural elements in viral genome involved in regulation of translation and replication (22, 74, 80-83). With the absence of both 5´ cap and 3´ poly(A) tail, translation of TCV genomic and subgenomic RNAs follows a non-canonical translation mechanism and is dependent on RNA structural elements found in the 3´ UTR (74, 84). The existence of five hairpins (H4, H4a, H4b, H5 and Pr) and four pseudoknots (Ψ_1, Ψ_2, Ψ_3 and Ψ_4) in the 3´ UTR of TCV and their specific functions in translation and replication have been established using genetic covariation, biochemical structure assays, and phylogenetic comparisons between related carmoviruses (74, 80, 85-87). These hairpins and pseudoknots are present in a region encompassing the 195-nt RNA sequence known as "the F4 fragment" (Figure 1.10) (22, 74, 80-83). TCV TSS, a cap-independent translation enhancer, which contains pseudoknots Ψ_3 and Ψ_2, as well as hairpins H4a, H4b and H5, is located between the H4 and Pr hairpins of the F4 fragment

(74). Two additional pseudoknots (Ψ_1 and Ψ_4) connect the TCV TSS to the Pr and H4 and allow these elements to interact for proper regulation of translation and replication (75, 85).

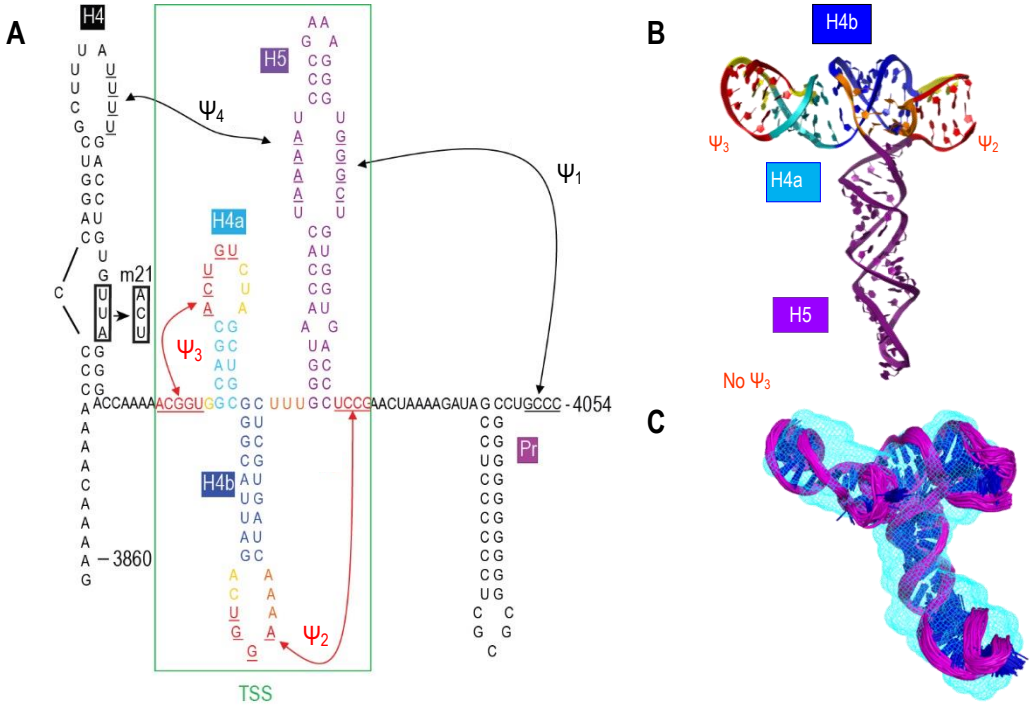


Figure 1.10. The secondary structure of the F4 fragment at the 3´ UTR of TCV
A. Secondary and tertiary interactions within the 3´ UTR of TCV. This 3´ UTR region is identified as F4. The TSS sequence is color-coded corresponding to its 3-D predicted structure in (B). Mutation m21 is boxed. **B.** The 3-D predicted structure of the TSS by RNA2D3-D. **C.** Structure of the TCV TSS by NMR and SAXS. Images adapted and modified from Mcormack et al, 2008 and Zuo et al, 2010 (74, 76).

1.2.1.1. The core promoter (Pr) and Ψ_1

The Pr of TCV is a stem-loop located at the 3´ terminus and is structurally conserved among all carmoviruses (Figure 1.10A) (6, 85-87). The Pr sequence was first discovered in a small, non-translated satellite RNA (satC) of TCV. SatC is a recombinant subviral RNA that shares 166 nt at its 3′ terminus with the TCV genomic RNA. Addition of the Pr sequence to a non-related template allowed it to be transcribed by the TCV RdRp (86). Several other RNA elements required for replication in TCV were also first

discovered in satC (85, 87-89). Deletion of the three C-terminal cytidylate residues in satC led to increased transcription, indicating that these nucleotides serve as a transcriptional suppressor (85, 90). In addition, deletion of these C residues led to increased flexibility of the three consecutive guanylates positioned within the large symmetrical loop of H5 hairpin in structure probing experiments, suggesting the presence of a PK (Ψ_1) connecting the two sequences. Disruption of Ψ_1 in satC by a point mutation resulted in increased accumulation of satC in protoplast while restoring the interaction by compensatory mutations results in wild-type levels of accumulation (85). This further confirmed the presence of Ψ_1 and its repressive activity on transcription. Stability of the Pr stem is important for *in vitro* and *in vivo* satC accumulation, since disruption of the Pr stem is detrimental (86). The Pr loop sequence is essential for TCV transcription and translation. Mutation of the Pr loop reduced *in vitro* transcription of a 3´ end fragment while increasing translation of a reporter construct (82). Transcription of TCV requires both 3´ UTR upstream sequence and the Pr, indicating a functional interconnection between Pr and upstream sequence (6, 80). Pr loop also modulates ribosome recoding required for RdRp synthesis by forming a long-range RNA:RNA interaction with the readthrough stimulatory element (RSE). Disruption of this interaction leads to undetectable levels of RdRp production (91).

1.2.1.2. Hairpin H4

Hairpin H4 of TCV contains an asymmetric loop near the base of the stem and a large apical loop that are both important for replication and translation (75, 80). Gel mobility shift analysis indicates that H4 and/or surrounding sequences interact with the RdRp and this interaction enhances transcription of both negative and positive strands of

satC and TCV (80, 92, 93). Specific mutations in the H4 terminal loop moderately repress translation of reporter constructs but can eliminate nearly all detectable virus accumulation in protoplasts. Mutations in the H4 asymmetric loop (m21), (Figure 1.10A) completely repressed translation *in vivo* and, like specific mutations in the apical loop and elsewhere in the 3'UTR, promote accumulation of compensatory second-site mutations clustered in several regions within the 3' UTR and within the coat protein coding region (75, 82, 94). Chemical structure probing of the 3´ UTR of the second-site mutants that are located in the coding region of the coat protein indicated that the few second-site mutants are clustered in a discreet RNA domain, suggesting a long-range RNA:RNA interaction between the genomic domain 1 and domain 2 of TCV (94). H4 exerts a negative effect on ribosome binding of a fragment containing TCV sequence from H4 to Ψ_2 ("F3" fragment). Disruption of Ψ_4, which connects the H4 terminal loop with the large symmetrical loop of H5, resulted in a two-fold increase in ribosome binding to the F3 fragment while compensatory mutations restoring Ψ_4 reduced ribosome binding (75). Local interactions between RNA elements within the 3´ UTR of TCV is further confirmed when a single point mutation (A3864U) upstream of H4 led to second-site changes in the H4b stem and the Pr loop (83).

1.2.1.3. The TCV TSS

As mention above, TCV contains a cap-independent translation enhancer (the TSS), comprised of pseudoknots Ψ_3 and Ψ_2 and hairpins H4a, H4b and H5 (Figure 1.10) (74). Mfold (95) prediction, structural probing and mutagenesis analysis confirmed the presence of H4a, H4b, H5 and Ψ_2 in the 3´ end of satC and TCV (74, 96, 97). H5 consists of an upper stem, large symmetrical loop and an apical loop, is conserved in all

carmoviruses, and likely acts as an RdRp chaperone (98). Ψ_2 is formed by interaction between the loop of H4b and the 3´ flanking region of H5, is conserved in four carmoviruses and is essential for replication of TCV and satC (99, 100). H4b and H4a are located just upstream of H5. H4b is predicted to be in eight other carmoviruses while H4a is only found in three carmoviruses (6). In both satC and TCV, H4a and its upstream sequence form a stable H-type pseudoknot (Ψ_3) (74, 75). However, disrupting Ψ_3 does not affect the accumulation of satC in protoplasts. Elements similar to the TCV TSS are also found in closely related *Cardamine chlorotic fleck virus* (CCFV). Substituting the entire core TSS structure in TCV with the one from CCFV, but not individual elements, led to viral accumulation similar to wildtype (wt) TCV, suggesting that these hairpins and pseudoknots of the TSS function together as a structural domain (74). H4a, H4b and Ψ_3 together are also viable as a unit in satC (97). Addition of the TCV TSS to the 3´ termini of luciferase construct that contains the 5´ UTR of TCV enhanced translation of reporter construct in protoplasts, indicating this element supports non-canonical translation (75).

Computational modeling using RNA2D3D predicted that TCV TSS hairpins and pseudoknots form a stable tRNA-shaped structure that has topology similar to a canonical tRNA. H4a and Ψ_3 are superimposed with the amino-acceptor arm of the phenylalanine tRNA (Phe-tRNA) while H5 is somewhat longer than the anti-codon stem-loop of the Phe-tRNA (Figure 1.10B) (74). Filter-binding and competition assays demonstrated that the TCV TSS binds to 80S ribosomes at the P-site through interaction with ribosomal 60S subunit (K_d = 0.45 μM) (75). The 40S ribosomal subunit interacts with a pyrimidine-rich sequence in the 5´ UTR of the genomic RNA in the vicinity of the start codon. In addition, the TCV TSS does not contain interacting sequences with the 5´ terminal

sequence. Thus, it is proposed that a ribosome bridge might connect the two ends of the TCV genome for optimal transfer of ribosomal subunits from the 3´ end to the 5´end during non-canonical translation (22, 74). The tRNA-shaped conformation of the TCV TSS was later confirmed by NMR and SAXS in which structures of H5 and H4b were established (Figure 1.10C). However in these structural studies, in order to support transcription by T7 RNA polymerase, two guanines residues were added to the 5´ end of the core TSS sequence. These two nucleotides (with surrounding sequence) formed a different hairpin that replaced Ψ_3 (Figure 1.10C) (76), placing the existence of this pseudoknot into question. However, disruption of Ψ_3 with a single point mutation, which increases the flexibility of H4a and the 5´ upstream adenylate-rich sequence, decreased ribosome binding to the TCV TSS, and reduced translation and accumulation in protoplasts (74, 75). Switching the adenylate (A)-rich sequence to a uridylate (U)-rich sequence decreased ribosome binding and translational efficiency in luciferase constructs (75). In addition, virus accumulation in protoplasts was undetectable when the adenylate-rich sequence was deleted (80). Thus, a possible role of the adenylate-rich sequence is to stabilize Ψ_3 (74, 75).

Chemical structure probing analysis of RNA fragments containing the TSS in the presence of 80S ribosomes detected only small conformational changes localized to the TSS region (22). Interestingly, substantial structural changes within the TSS and adjacent regions were observed in the presence of RdRp (Figure 1.11), suggesting that the TCV TSS has two functional conformations (81), and that regulation of the conformational switch within the TSS-containing region might regulate the switch between viral translation and replication. In addition, RNA elements of satC that resemble RNA

elements of the TCV TSS are also suggested to switch conformations to activate the replication of the satC template (93). Thus, the conformational change in the stable TCV TSS and its interaction with RNA elements in nearby sequences that regulate translation and replication render the TCV TSS as an exciting system for studies of RNA conformational changes.

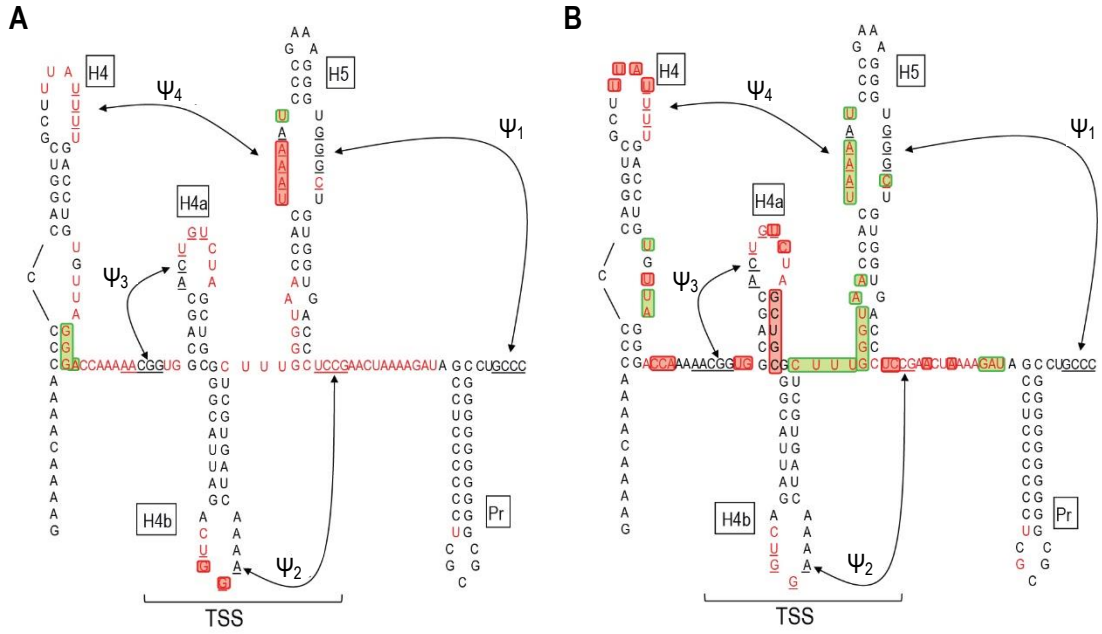


Figure 1.11. Structural probing of the TSS and surrounding sequence of TCV.
A. In-line structure probing of an RNA fragment (sequence shown) in the absence and presence of 80S ribosomes. **B.** In-line structure probing of F4 RNA in the absence and presence of RdRp. Flexible nucleotides are presented in red. Red boxes and green boxes denote nucleotides with increased or decreased flexibility, respectively, in the presence of either ribosomes (A) or RdRp (B) (22, 74, 81, 83). Images adapted and modified from Stupina et al, 2008 and Yuan et al, 2009 (75, 81).

1.2.2. *Pea enation mosaic virus* and multiple 3´ CITEs

PEMV RNA 2 (PEMV), a recombinant virus of the umbravirus genus (family not assigned but contains many elements similar to members of the *Tombusviridae*) has a 4.2 kb genome, which is uncapped and has no poly(A)-tail. The viral genome encodes a carmovirus-like RdRp and two overlapping, movement-associated proteins (p26 and p27) (Figure 1.12A) (101). PEMV RdRp allows the virus to replicate independently in

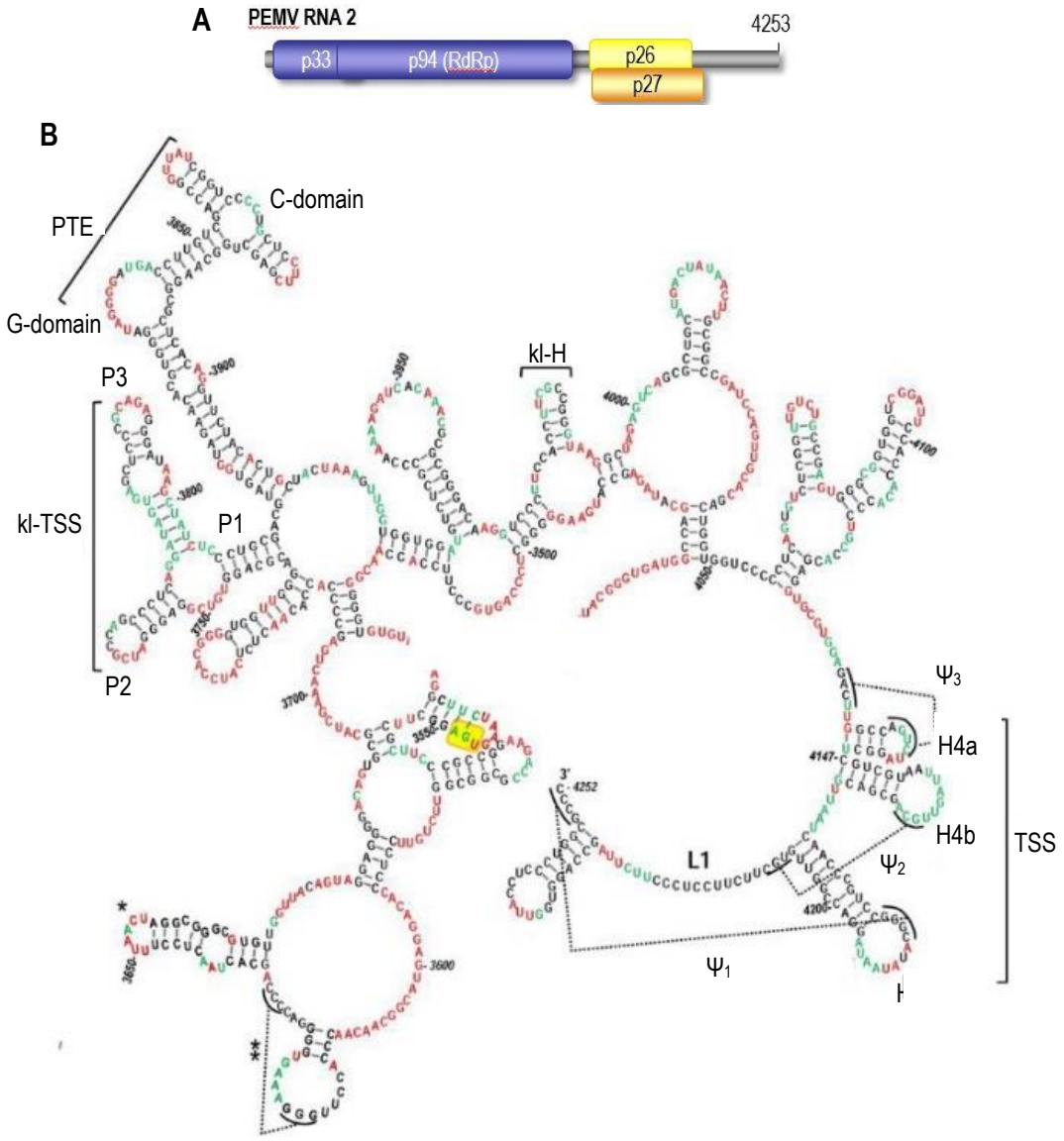


Figure 1.12. The secondary structure of the 3´ UTR of PEMV
A. Genomic organization of PEMV RNA 2. The PEMV genome encodes four proteins: p33, p94, p26 and p27. p94 is the product of ribosomal frameshifting of p33. **B.** Proposed secondary structure of the 3´ UTR of PEMV by SHAPE. Multiple RNA structural elements in the 3´ UTR of PEMV are indicated. Residues corresponding to high and low reactivities to NMIA are in red and green, respectively. Terminal stop codon of p27 is highlighted in yellow. Pseudoknots (known and predicted) are indicated with dashed lines (102).

protoplasts, however, due to the lack of a coat protein, encapsidation and transmission of PEMV from plant-to-plant requires the presence of gene products encoded by the associated viral RNA, PEMV RNA 1 (103). Unlike other viruses containing only single

3´ CITE in the viral 3´ UTR, examination of the long 3´ UTR (~ 705 nt) of PEMV has revealed multiple 3´ CITEs. Each 3´ CITE interacts with either eIF4E, ribosomal subunits or the 80S ribosome (Figure 1.12B), suggesting the presence of a novel translation enhancement mechanism (51, 59, 60, 77) .

1.2.2.1. The PTE of PEMV

As mentioned in the previous section, a *Panicum mosaic virus*-like translational enhancer (PTE) was the first 3´ CITE found in PEMV (Figure 1.8C) and is conserved in several carmoviruses (5, 59, 60). Translational enhancer activity of the PEMV PTE is supported through its interaction with eIF4E (59). In canonical translation, interaction of the mRNA with eIF4F helps to recruit other translation initiation factors. The PEMV PTE binds specifically to eIF4F through eIF4E rather than eIF4G, and likely recruits other translational initiation factors (59). Computational prediction and confirmation by structure probing shows that the PTE has a T-shaped structure that is presumably held through a magnesium-dependent pseudoknot connecting the C-domain and G-domain (Figure 1.13A) (60). Superimposing the computational model of the PEMV PTE with the 3-D structural model of eIF4E exposes the binding pocket of the PTE with eIF4E (Figure 1.13A), in which the binding pocket is formed by a pseudoknot between the C and G domains. The protrusion of the G domain is hypothesized to directly interact with eIF4E (60).

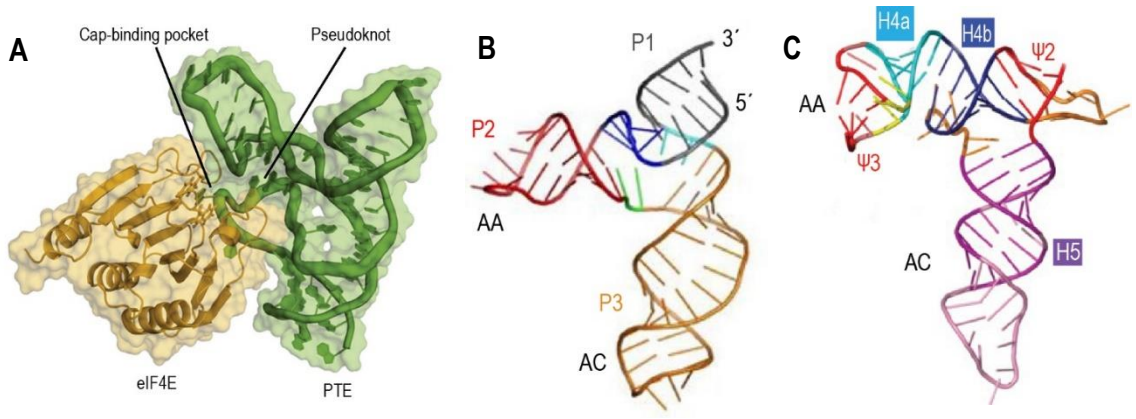


Figure 1.13. The predicted 3-D structure of the 3´ CITEs of PEMV.
A. The PTE (green) is docked with eIF3E (yellow). The pseudoknot and the cap-binding pocket are indicated. **B.** The kissing-loop TSS (kl-TSS) of PEMV. **C.** The TSS of PEMV. The stems of CITEs aligned with tRNA stems are indicated. AA: acceptor stem of tRNA, AC: anticodon stem of tRNA. Images adapted from Wang et al, 2011, and Gao et al, 2012, 2013 (60, 77, 102).

1.2.2.2. The kl-TSS of PEMV

The PEMV PTE does not form any discernible long-distance interaction with the 5´ UTR (59). However, an apical loop that is upstream of the PTE and located in the P2 stem of the PEMV kl-TSS interacts with an apical loop of a hairpin (H2) located in the 5´coding region (Figure 1.8G, 1.12B and 1.13B) (77). The apical loop of the P2 stem of the PEMV kl-TSS contains the conserved carmovirus kissing-loop motif GCCA (77). The presence of a terminal 5´ 89-nt RNA fragment retarded the mobility of radioactive-labeled PEMV kl-TSS in gel shift assays, indicating that the two RNA fragments interact (77). When combined with the 5´ 89-nt, the PEMV kl-TSS enhances *in vitro* translation of a luciferase construct by 14-fold (77). Disruption of the RNA:RNA interaction between P2 of the PEMV kl-TSS and 5´ H2 impaired *in vitro* translation of the reporter construct and *in vivo* accumulation of the full-length PEMV genome. These negative effects were restored by compensatory mutations, indicating the importance of the long-range RNA:RNA interaction of the PEMV kl-TSS (77). The PEMV kl-TSS is predicted

to form a three-way junction branching into two hairpins. Like the TCV TSS, the PEMV kl-TSS is predicted to assume a tRNA-shaped structure since its computational model is superimposed with Phe-tRNA (Figure 1.13B). The P2 stem is aligned with the amino-acceptor stem and the P3 stem is aligned with the anticodon stem. Unlike TCV TSS, in which H5 is slightly longer than the anticodon stem of the Phe-tRNA, the P3 stem of PEMV and the anticodon stem of the tRNA have similar length. The kl-TSS interacts individually with both ribosomal subunits as well as 80S ribosomes while the TCV TSS interacts with 60S ribosomal subunits and 80S ribosomes (22, 77). Unlike the TCV TSS, the PEMV kl-TSS does not occupy the ribosomal P-site, indicating that the PEMV kl-TSS might enhance translation through a different mechanism (51). The P1 stem and the apical loop of the PEMV kl-TSS are important for ribosome binding. Disrupting the P1 stem resulted in reduction of ribosome binding by several fold (77). A similar negative effect on ribosome binding was also observed using mutant kl-TSS fragments in which the P3 loop was either enlarged or truncated (77). The kl-TSS is able to interact simultaneously with the 5' 89-nt fragment and ribosomes (51). Interaction of PEMV kl-TSS with translation initiation factors has not been reported.

1.2.2.3. The TSS of PEMV

A third 3´ CITE of PEMV was recently found near the 3´ terminus of the PEMV genome. This independent 3´ CITE contains two pseudoknots and three hairpins (Figure 1.12 B), which are arranged similarly to the order of elements in the TCV TSS, and is predicted to form a similar tRNA-shaped structure. As such, this 3´ CITE of PEMV is called the PEMV TSS. However, Ψ_3 of the PEMV TSS is less stable (Figure 1.13C) (102). Like the TCV TSS, the PEMV TSS also binds to 80S ribosomes and 60S

ribosomal subunits. However, ribosome binding of the PEMV TSS is weaker (K_d =2.4 μM) as compared to the ribosome binding affinity of the TCV TSS (102). In the TCV TSS, Ψ_3 and H4a are critical for ribosome binding (74, 75). Thus, the less stable Ψ_3 of the PEMV TSS may be responsible for the weak ribosome binding of the element (102). Similar to the TCV TSS, the PEMV TSS does not contain any inherent sequence that interacts with the 5′ UTR sequence of the viral genomic RNA. However, the apical loop of a small hairpin (kl-H) located upstream of the PEMV TSS (Figure 1.12B) can form a long-range RNA:RNA interaction with the 5′ H2 that interacts with the PEMV kl-TSS in a reporter construct. The PEMV TSS and the kl-H with 5′H2 works synergistically to maintain translation of reporter constructs lacking functional kl-TSS and PTE (102). The presence of the kl-H in *trans* inhibits translation of the reporter construct containing the three 3′ CITEs of PEMV (102), indicating that the kl-H is important for translation. However, compensatory mutations between the kl-TSS and 5'H2 result in fully functional virus, indicating that the kl-H must have an alternative pairing partner in full-length virus.

The three 3′ CITEs of PEMV have significantly different structural, positional and functional properties and also appear to be different from the TCV TSS. The ribosome binding mode for the two 3′ CITEs is also different from the ribosome binding mechanisms of IRES. Since translation is a critical step in the virus life cycle, it is necessary to understand the structure of these 3′ CITEs to have a better insight into their modes of action.

1.3. An overview of structures, folding and conformational switches in RNAs

RNAs, despite having only four different bases, are capable of performing a variety of cellular functions that rival those of proteins, which contain 20 amino acid side chains. This considerable functional plasticity is in large part due to the ability of RNAs to fold into alternative functional structures (104, 105). RNAs are characterized by primary sequences, secondary structures and tertiary folds. Typical motifs found in the secondary structure of RNAs are hairpins, bulges, internal loops, and junctions. Typical motifs found in RNA tertiary interactions are pseudoknots, coaxial stacks, A-minor motifs, ribose zippers, and kissing loops (106, 107). Secondary structure of RNA is stable and dominated with canonical Watson-Crick base pairs, whereas higher order structure is less stable and dominated with non-canonical base pairs (108). RNA secondary structure can be predicted with reasonable accuracy using the nearest-neighbor model (95, 109) while prediction of tertiary structure is still limited due to the lack of thermal melting data (110). The folding process of RNA structures starts out with neutralization of the negatively charged phosphate backbone by cations such as sodium and magnesium (111), leading to formation of compact structural intermediates. Formation of a native state from compact intermediates can occur by a single step through the simultaneous collapse of multistages (112) or can undergo multiple transitional steps through the arrangement/rearrangement of intermediates (113). Tertiary interactions are proposed to guide the folding process of RNAs (114, 115). Once folded into the native state, RNA structures do not remain static but continue to be dynamic. Dynamic conformation changes in an RNA's structure allows it to perform either multiple functions or to signal the turning on and off of cellular processes (105, 116). Multiple regulatory RNAs have multiple conformations, for example, riboswitches (117, 118), 3′ CITEs (75, 81) (119,

120), and viral RNA packaging signals (121, 122). Conformational switch of RNA can be induced by environmental cues such as changes in pH, ion, magnesium, metabolites or proteins (123, 124). Knowledge of how RNA structures fold is essential for both understanding their functions and enabling the building of computational models that have the ability to simulate and predict the dynamic structure of new RNA elements.

1.4. An overview of biophysical methods used in this study

1.4.1. Optical tweezers (OT)–method of choice for studying RNA folding

Single molecule techniques have emerged as powerful tools for understanding the behavior of biological molecules such as DNA, RNA and proteins in terms of their folding pathways and energy landscape (125-127). In contrast to measuring *ensemble averaged* behaviors, single molecule techniques allow for the direct examination of activities of *individual* molecules, which enables the determination of the entire distribution of structural progression in real time and elucidation of folding variables (e.g. force, extension lengths, unfolding/folding kinetics) (110, 127-131).

A recently developed laser machine, optical tweezers (OT), can probe the structure of a single RNA molecule and provide an integrated picture of secondary structure and tertiary interactions (128, 132, 133). Most methods used to study RNA structure alter the environment (e.g. by heating). OT has the advantage of using force as a variable, which allows for examination of the stability and unfolding/folding behaviors of a single RNA molecule without affecting the surrounding environment, providing for a more relevant structure (128, 132-134). For example, OT has been used to reveal the unusual stability of a 2-nt kissing loop RNA interaction (128) or formation of a pseudoknot through the rearrangement of the two hairpins of the *rpsO* riboswitch (135). There is no nucleic acid

size limit using OT, and has been utilized to study the mechanical properties of a 16 μm DNA (136), which is roughly equivalent to the size of the largest viral RNA genome (125, 137).

OT can uncover hidden information and mechanical properties regarding RNA folding pathways, which may elucidate the mechanisms of biological processes. For example, OT analysis of the -1 ribosomal frameshifting pseudoknot of human *telomerase* RNA has revealed that an increase in the mechanical stability of the pseudoknot leads to an increase in -1 frameshifting efficiency (138). OT can also reveal short-lived intermediate folding states of nucleic acid structure (139) allowing the folding energy landscape of an RNA to be reconstructed (133, 139). Such data is revealing information on dynamic conformational states of RNA.

Two OT experiments, force ramping and force clamping, are used to elucidate molecular folding mechanisms and dynamic conformations of RNA (Figure 1.14). In force ramping, force is increased or decreased at a constant rate, while in force clamping, force is held constant. In both experiments, any changes in extension of the RNA are monitored. Force-dependent extension of the molecule is converted to force-free extension (contour length) using worm-like chain models. The rupture force range and the life-time intermediate states are also recorded so that they can be used to characterize the intermediate states, to calculate unfolding/folding kinetics of the RNA and to estimate RNA stretching/refolding work, thereby estimating the stability of the RNA. The stability of an RNA structure is equal to the value of work at which the magnitudes of probability distribution of unfolding work and the probability of refolding work of RNA is equal (130, 140).

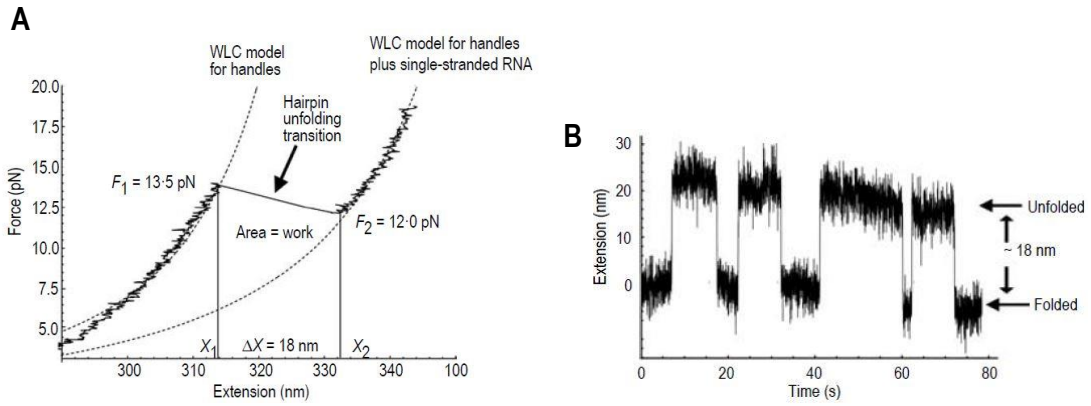


Figure 1.14. Representative data collected in OT experiments.
A. Force extension curve of an RNA hairpin in a force ramping experiment. A constant rate of 1 pN/s is applied to stretch the RNA hairpin flanked by 500 bp DNA/RNA handles at each end. At low force, the DNA/RNA handles are stretched monotonously. When the applied force reaches critical force (13.5 pN), the RNA hairpin opens, creating a sudden increase in extension ($\Delta X = X_2 - X_1$), called a rip, in the force extension curve. Force extension curves before and after the transition are fitted with the worm-like chain model to obtain the contour length of the handles and the handles plus single-stranded RNA. The area underneath the rip represents the work required to open the RNA hairpin and this work can be used to estimate RNA stability. **B**. Hopping behavior of an RNA hairpin in a force clamping experiment. Force is held constant and the change in extension of the hairpin is measured over time. Images adapted form Tinoco et al, 2006 (130).

As with proteins, RNA sequence is hypothesized to cluster into domains so that RNA elements can fold/unfold as a unit and act synergistically to perform functions due to their cooperative behaviors (141). The cooperativity between RNA elements is defined as the process during which unwinding of one RNA element affects the unfolding/folding behavior of the other RNA elements e.g., their stability, extension length, rupture force range, transition position, unfolding/folding kinetics and pathways (128, 129, 132-134). It has been reported that cooperative interactions of RNA structures to increase the stability of RNA structural domains and to shift the transition position of RNA close to the folded states (Figure 1.15) (129, 130, 132-134, 142). As such, secondary and tertiary RNA structures can be grouped into structural domains based on their cooperativity. These behaviors/mechanical properties of RNA can be obtained from OT experiments. The

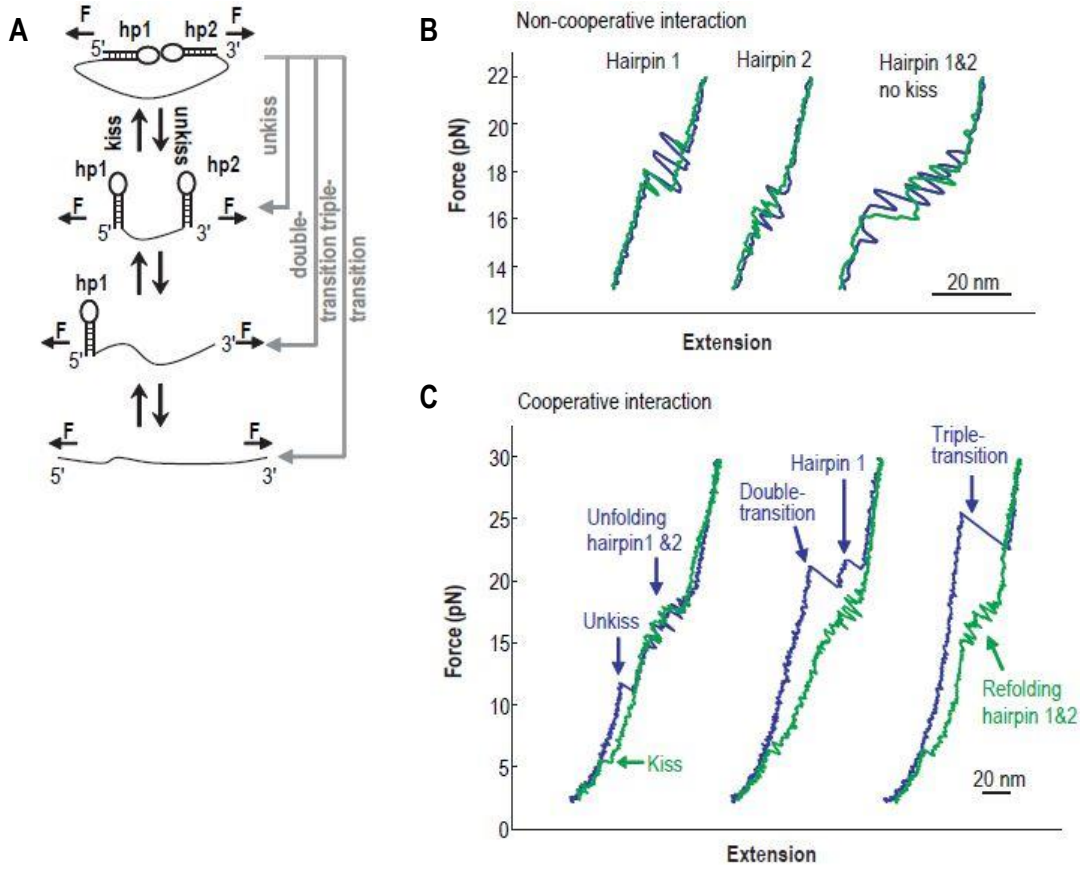


Figure 1.15. Representative mechanical properties of the cooperative tertiary interaction of kissing loops of two hairpins (hp) under applied force of OT

A. Diagram describes the rupture of the kissing complex by application of force in non-cooperative/no kissing interaction (B) and cooperative/kissing interaction situations (C). **B**. Force extension curves (FECs) of the two hairpins with non-cooperative interaction. **C.** Force extension curves of the two hairpins with cooperative interaction. The unfolding FECs are represented in blue and the refolding FECs are represented in green. No kissing/non-cooperative interaction of the two hairpins results in unfolding of the two hairpins sequentially. Cooperative tertiary interaction of the kissing complex stabilizes the complex, resulting in unfolding of the kissing complex with double transition or triple transition at higher force level. Images adapted from Li et al, 2006 (128).

biological relevance of the OT method is that it is capable of stretching the molecule until it completely unfolds and then allowing the molecule to refold. This process mimics the self-assembling process of an RNA sequence in nature (110, 127). The technique also provides information on interactions between RNA structures without the need to examine individual RNA elements. Thus, this method helps to reduce the amount of work required to study structural/functional domains of an RNA.

1.4.2. Nuclear magnetic resonance (NMR) spectroscopy to study RNA structure at high resolution

The magnetic moment of a nucleus interacts with an external magnetic field, allowing the nucleus to absorb the energy at a corresponding resonant frequency. An NMR spectrometer can detect, amplify, and record the resonant frequency of the nucleus and transform it into a frequency signal. Since these signals depend on the strength of the applied magnetic field, they are measured against a standard compound and reported in units of parts per million (ppm). The resonant frequency of a nucleus varies based on its chemical environment and thus, is referred to as a chemical shift. For example, a proton that is covalently-bonded with a nitrogen has a different chemical shift compared to the chemical shift of a proton that is covalent-bonded with a carbon. NMR spectroscopy, which can determine chemical shifts of atoms, is a powerful biophysical tool for structural studies of RNAs in solution at atomic resolution. NMR has provided insight into the structures of numerous RNAs and the structures-based mechanism of multiple processes. For example, the small size structure of the 29-nt ribosomal A-site (143) and the relatively large size structure of 150-nt genome packing signal of HIV (144) have been solved by NMR. NMR spectroscopy provides information on the base-pair pattern (76), site-specific information of ligand binding (145), local structure and global structure of RNAs (146), and conformation of RNA structures (147). The Watson-Crick base-pair pattern of RNA is readily inferred from the imino signals of bases. When base-paired, RNA imino protons are normally protected from the solvent exchange, hence their chemical shifts can be observed in 1-D or 2-D NMR experiments (146). Each imino proton peak of a 1-D spectrum or each imino proton and nitrogen peak of a 2-D spectrum represents a base-pair in the RNA (Figure 1.16A-C). Protons that are within ~ 5 Å can

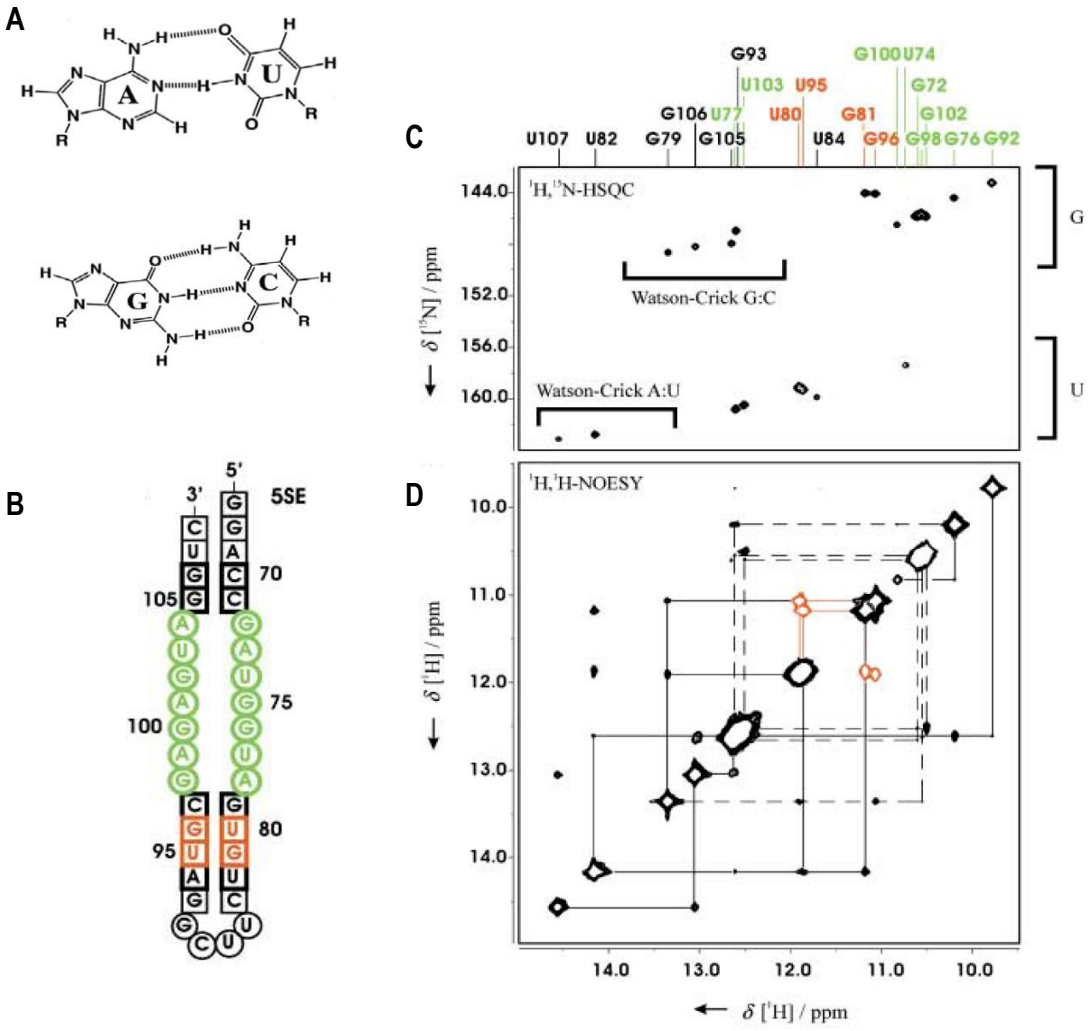


Figure 1.16. Examples of 1-D and 2-D NMR spectra of the E-loop of *E. coli* 5S rRNA
A. Schematic representation of Watson-Crick base-pairs AU and GC. Hydrogen bonds are represented by dashed lines. **B.** Secondary structure of the E-loop. Residues are color-coded corresponding to their resonance on 2-D spectra shown in C and D. **C.** 2-D ^{1}H,^{15}N-imino HSQC spectra of the E-loop. G and U regions are labeled. **D.** 2-D ^{1}H,^{1}H- NOESY spectra of the E-loop. Solid and dashed lines indicate sequential NOE contacts of the RNA. Images adapated from Furtig et al, 2003 (146).

transfer their magnetization to each other, thus producing a correlated peak/cross peak in the nuclear Overhausser effect spectroscopy (NOESY) experiment (Figure 1.16D) (146). The cross peaks that connect resonances of the different imino protons of different bases allow for sequential assignment of the imino proton resonances of bases in RNA molecules (Figure 1.16D). Signal intensity of imino proton peaks in NOESY spectra is

inversely proportional to the distance (r^{-6}) between protons (146). For example, two protons that are near each other produce a cross-peak with higher intensity than those that are further away from each other (Figure 1.16D). Thus, inherent distances between the correlated protons can be extracted. Scalar coupling constants ($J_{XY;}$ X,Y can be either protons (H) or heteroatoms-^{15}N, ^{13}C, or ^{31}P) and residual dipolar couplings provide valuable information on local angles or global angles in RNA, respectively (146, 148). Thus, information on distances and angles between atoms of RNA molecules can be used to reconstruct the structures of RNA molecules. Whereas NMR has emerged as the tool of choice to solve RNA structures and probe dynamics (149), three bottlenecks limit effective RNA analysis: (i) extensive chemical shift overlap of resonances, (ii) strong ^{13}C-^{13}C dipolar and scalar couplings of adjacent carbon atoms and (iii) rapid signal loss due to line broadening for RNA molecules bigger than 50 nt (150). These limitations can be overcome by using site-specifically labeled nucleotides (rNTPs) in combination with ultrahigh field NMR spectroscopy: the spectral resolution increases, the signal to noise ratio is enhanced due to the removal of ^{13}C-^{13}C coupling, and the structural assignment of RNAs is simplified (151-154).

1.4.3. Small angle X-Ray scattering for studying the global shape of RNA structure

Small angle X-Ray scattering (SAXS) is another structure determination method suitable for RNAs as the high electron density in RNA backbones make RNA molecules more sensitive for production of scattered profiles as compared with analysis of proteins

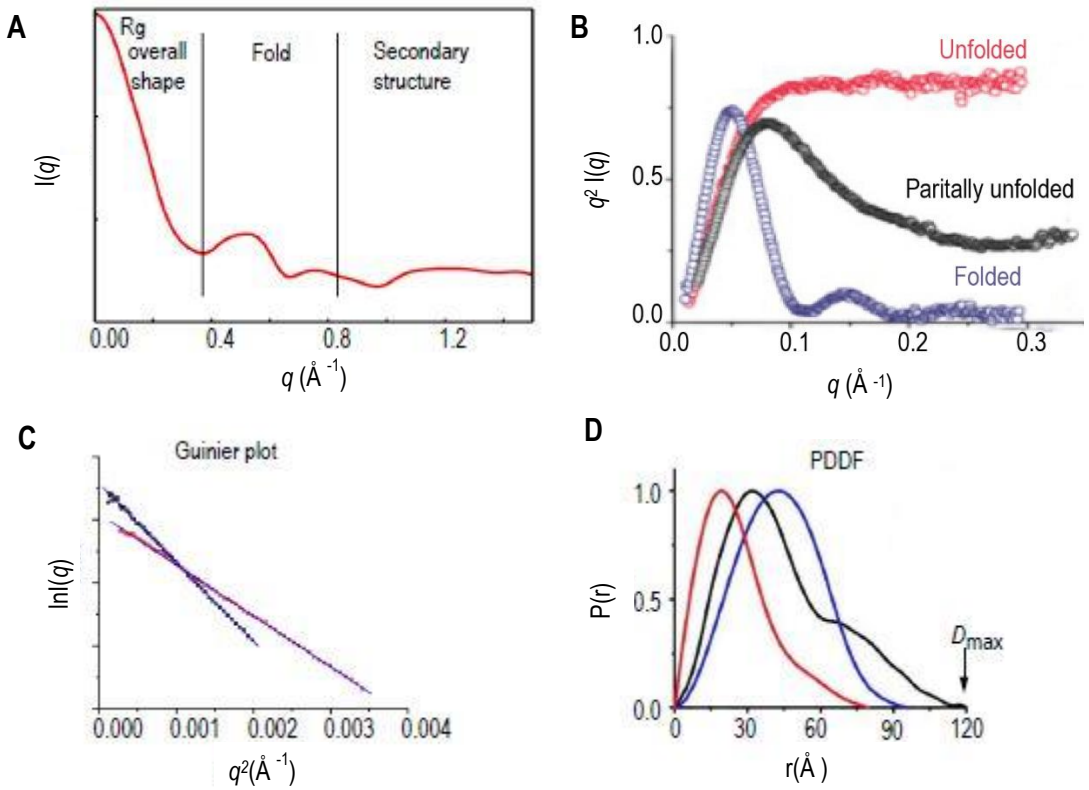


Figure 1.17. Examples of SAXS data analysis
A. Scattering profile of an RNA obtained using SAXS. **B.** Typical Kraky plots for molecules with different foldings. **C.** Guiner plots from which radius gyration and molecule weight of an RNA molecule can be deduced. **D.** Pair distance distribution function (PDDF) graph from which the maximum particle size (Dmax) can be deduced. PDDF of an RNA can be used to determine to the shape of the molecule. Images adapted from Putnam et al, 2007 and Fang et al, 2015 (155, 156).

(155). Unlike X-Ray crystallography, SAXS can work with RNA samples in solution (155, 157), allowing SAXS experiments to be performed in near physiological conditions. Scattering profiles of RNAs are obtained by shooting the small-angle collimated X-rays through the sample solution and are represented as signal intensity $I(q)$ plotted over a range for the scattering momentum vector (q) (Figure 1.17A). Analyzing linear Guiner region ($g.Rg < 1.3$) of a scattering profile for RNA can provide information on the radius of gyration and molecular mass of the molecule (Figure 1.17B). Applying Porod's law of globular macromolecules to plot q over $I(q).q^2$ provides information on RNA folding (Kraky plot) in which compact molecules tend to produce a bell-shaped

curve, while unfolded molecules create plateau curves at high q range. Extended molecules have a bell-shaped curve at low q range and an increase in signal at high q range (Figure 1.17C). Fourier transformation of scattering profiles of RNA molecules produces pair distance distribution of RNAs, which can be used to build the molecular envelope of the RNA (Figure 1.17D). Fitting the ensemble of computational models into the molecular envelope of RNA molecule can provide tentative 3-D models of RNAs. SAXS has been used to analyze a wide range of RNA sizes (158, 159). Conformational switches of specific RNA molecules were also reported by using SAXS (117, 160).

1.5. Thesis plan

Mutational analysis and chemical structure probing methods provide great insight into the function and secondary structure and of 3′ CITEs. However, lack of 3-D structural information for most of these 3′ CITEs has limited our understanding of their functional dynamics. TCV and PEMV, two (+)-strand RNA viruses, are well-established viral models used for investigating the regulation of translation and replication due to their small size and relatively simple genomes. Both viruses contain CITEs in their 3′ UTRs that fold into a tRNA-shaped structure (TSS). Specifically, the TCV TSS can directly associate with ribosomes and participates in RdRp binding. The PEMV kl-TSS can bind to ribosomes and interact with the 5’ UTR of the viral genome. In this thesis, multiple biophysical tools including optical tweezers (OT), small angle X-ray scattering (SAXS), and nuclear magnetic resonance spectroscopy (NMR) have been used to elucidate the folding dynamics and structure of two CITEs, the TSS TCV and PEMV kl-TSS, respectively.

In Chapter II, I report on the investigation of folding dynamics for the TCV TSS using optical tweezers. This study has provided us with an unexpected folding pathway for the TCV TSS, and confirmed the presence of Ψ_3 and hairpin elements of the TSS. In addition, this study has demonstrated the importance of the adjacent A-rich sequence to the formation of H4a/Ψ_3 along with contribution of magnesium to the stability of the TCV TSS. In Chapter III, I report on the structural analysis of the PEMV kl-TSS using NMR and SAXS. This study has re-confirmed the base-pair pattern of the PEMV kl-TSS and the interaction of the PEMV kl-TSS with 5´ hairpin H2. The molecular envelope of the kl-TSS built from SAXS shows that the kl-TSS might have two conformations and that one of the major conformation has a different shape than the previously predicted tRNA-shaped form. In Chapter IV, I report using wild-type and mutant *E.coli* strains to produce cost-effective, site-specific labeled, recombinant RNAs. The technique was validated with four RNAs of different sizes and complexity to produce milligram amounts of RNAs, which is sufficient for biophysical studies. Several NMR techniques were used to demonstrate the benefit of site-specific labeled RNAs made from *E.coli.*

Chapter 2: The folding pathway of an internal 3' proximal tRNA-shaped structure in *Turnip crinkle virus* using optical tweezers

2.1. Introduction

RNAs play key roles in a variety of cellular functions due to their ability to fold into a large number of functionally competent and dynamic structures (5, 8, 123, 161). Driven by conformational plasticity, RNA elements can assume alternative conformations with separate and specific biological functions (123, 162). Riboswitches, for example, are often found in the UTRs of bacterial genes and can cycle between active/inactive conformations by regulating direct binding/unbinding of small metabolites. When riboswitches shift conformations, the resulting structure can inhibit translation or attenuate the transcription of downstream genes (163, 164). Many regulatory RNA elements found in the 5′ and 3′ UTRs of viral RNAs also undergo conformational transformations to regulate translation and replication, two mutually exclusive processes (75, 81, 121, 122, 162, 165, 166). Despite the biological importance of regulatory RNA elements, information about unfolding and refolding remains elusive. In this report we investigate the folding pathway of a 3′ cap-independent translational enhancer (CITE) of *Turnip crinkle virus* (TCV), which has been reported to assume different conformations when bound to either ribosomes or RdRp (75, 81).

As mentioned in Chapter 1, TCV is a positive (+)-strand plant virus in the genus carmovirus, family *Tombusviridae*, with an uncapped, non-poly(A) genome of 4053 nt encoding five proteins (Figure 1.9). The TCV 3′ UTR contains five hairpins (H4, H4a, H4b, H5 and Pr) and four pseudoknots (Ψ_1, Ψ_2, Ψ_3 and Ψ_4) (Figure 1.10A) (5). Three hairpins (H4a/H4b/H5) and two pseudoknots (Ψ_3 and Ψ_2) fold into an internal tRNA-shaped structure (TSS) that serves as a translational enhancer for cap-independent

translation (Figure 1.10B) (74, 75). H4a and Ψ_3 are predicted to form an H-type pseudoknot that can superimpose with the amino-acceptor stem-loop of Phe-tRNA, whereas H5 is slightly longer than the anti-codon stem loop of Phe-tRNA (74). Ψ_2 superimposes with the structural region formed between the T-loop and D-loop of Phe-tRNA. The TSS specifically binds to 80S ribosomes at the P-site through an interaction with the 60S ribosomal subunit (75). Binding of the TSS to ribosomes correlates with translational activity of reporter constructs and mutations that reduce ribosome binding also have a negative effect on translation (75). Similar hairpins and pseudoknots are found in the closely related *Cardamine chlorotic fleck virus* (CCFV). Substituting the entire core TSS structure or Ψ3/H4a/H4b with the analogous regions from CCFV led to viral accumulation similar or greater than wild-type (wt) TCV. In contrast, TCV accumulated poorly when individual elements or other combinations of elements from CCFV were substituted. These results suggest that either all, or a combination of the hairpins and pseudoknots function together as a structural domain (74)

The TSS acts as a scaffold for multiple interactions with surrounding sequences (81, 97). In addition, the TSS region is also important for RNA RdRp binding to the 3´UTR (81) and for replication of a small, non-translated satellite RNA (satC) of TCV, which contains two 3´ regions of TCV (86). H5 is a key element in the replication of TCV and satC, and likely acts as an RdRp chaperone (98). Disruption of Ψ_2 enhances *in vitro* transcription of TCV using purified RdRp, but reduces accumulation of TCV and satC in protoplasts (74, 99, 100). Disruption of Ψ_3 also increases *in vitro* transcription of TCV (81) and does not affect the accumulation of satC in protoplasts. Interestingly, the satC TSS does not bind to ribosomes and is proposed to have an altered structure in satC

due to six different residues within the TSS region (167).

The overlap that exists in the TSS region between elements critical for translation and replication suggests that a mechanism exists for transitioning between different structures that supports individual processes. The TSS conformation does not change significantly when bound to ribosomes (Figure 1.11A) (75) but is substantially altered upon binding to the RdRp, and reverts upon removal of the RdRp (Figure 1.11B) (22, 82). Prominent changes include enhanced flexibility in the 3´ side of the H4a stem and reduced flexibility in the base of H5 (Figure 1.11B). The folding/unfolding dynamics that give rise to the TSS conformational changes are not known. Thus, the conformational change associated with translation and transcription, and the interaction with other RNA elements in the 3´UTR render the TSS an intriguing system for the study of RNA conformational switches.

In this study, I used optical tweezers (OT) to investigate the folding pathway of the TSS. As mentioned in Chapter 1, OT is a type of single molecule force spectroscopy (Figure 2.1B) that can be used to probe the structure of single RNA molecules providing an integrated picture of secondary and tertiary structures (128, 132, 133). OT uses force as a variable for examining the stability and unfolding/folding behavior of single RNA molecules without changing the surrounding environment (128, 129, 133, 134). By working with single molecules, OT has revealed the hierarchical folding of adenine riboswitches (133), and has demonstrated the rearrangement of two component hairpins to form a pseudoknot in the operator of the *Escherichia coli (E. coli) rpsO* gene (135). OT has also been used to examine the stepwise folding at low force and the cooperative folding at higher force of a telomerase pseudoknot (168).

Magnesium (Mg^{2+}) is a major factor associated with conformational switches (123, 169-171). Mg^{2+} functions through its electrostatic interactions with the RNA phosphate backbone to reduce the strong repulsion between negative-charged phosphodiester groups (111, 172-174) thus promoting formation and stabilization of RNA tertiary structures (111, 175-178). A second major factor, RNA pseudoknots, are formed from interactions between the loop of one stem and sequences adjacent to the stem, are hypothesized to function like "a switch" in many regulatory RNA structures (97, 179-181). Due to the critical roles of Mg^{2+} and pseudoknots in RNA folding, both are evaluated in this Chapter for their contribution to the stability and the folding dynamics of the TSS and its individual components.

Although the tRNA-shaped conformation of the TSS has been confirmed by NMR and SAXS (76), the existence of Ψ_3 as well as the role of the upstream adenylate (A)-rich sequence in the TSS structure needed further clarification. Unlike other hairpins and pseudoknot, Ψ_3 was not found in the TSS structure reported by NMR and SAXS (76) due to the addition of two guanylates at the 5′ end of the core TSS sequence to support transcription of T7 RNA polymerase. The nucleotide additions led to the formation of an alternative hairpin that replaced Ψ_3, while the TSS still retained comparable ribosome binding activity (Figure 1.10C) (76). However, disruption of Ψ_3 with a single point mutation increased flexibility of H4a and the 5′ A-rich sequence, which decreased ribosome binding, reduced translation of a luciferase reporter construct, and reduced accumulation of TCV in protoplasts (74, 75). Replacing the A-rich sequence with uridylates also decreased ribosome binding and translation efficiency of reporter constructs (75). In addition, virus accumulation in protoplasts was undetectable when the

A-rich sequence was deleted (80). Thus, a possible role of the upstream A-rich sequence is to stabilize Ψ_3 (74, 75).

Using optical tweezers, my results indicate that the TSS adopts an unexpected folding pathway in contrast to the folding pathway predicted by molecular dynamic simulations (MD), where H4a/Ψ_3 is the least stable and unfolds first. I found that, in the absence of Mg^{2+}, H4b unfolded first, followed by H5 and then H4a/Ψ_3. In the presence of Mg^{2+}, H4b unfolded together with H5, followed by H4a/Ψ_3. My results confirm the presence of Ψ_3 and reveal that H4a is unstable in the absence of the A-rich sequence and Ψ_3. Mg^{2+} enhanced the stability of all hairpins and pseudoknots within the TSS, but surprisingly, the H4a/Ψ_3 complex can still form in the absence of Mg^{2+}. In addition, in Mg^{2+}, cooperativity between H4b and H5 requires the presence of Ψ_2. Studying the folding pathway of the TSS illuminates the interconnection between its hairpins and pseudoknots and provides new insights in understanding its folding process and conformational switch dynamics.

2.2. Materials and Methods

2.2.1. Preparation of RNA molecules for OT

DNA fragments coding for wt and mutant TSS were amplified by polymerase chain reaction (PCR) using plasmids containing full length TCV genome. Fragments were inserted into pUC19 between EcoRI and SmaI restriction sites. TSS mutants were generated by site-directed mutagenesis using two complementary primers and high fidelity pfu polymerase following a protocol previously described (182). For site-directed mutagenesis, two fully complimentary mutant primers were designed to have the mutant nucleotides located at the center of each primer. A two-round PCR reaction was used to

perform site-directed mutagenesis. The first round PCR reaction contained 500 ng of plasmid, 1 µM of forward primer, 20 µM dNTPs, 3% DMSO, 1x of high GC buffer (NEB) and 2 unit of fusion pfu polymerase (NEB). The PCR reaction was denatured at 98ºC for 2 min, and polymerization performed for 18 cycles using the subsequent steps: 98ºC for 30 sec, 55ºC for 30 sec, and 72ºC for 45 sec. The PCR reaction was subsequently incubated at 72ºC for 5 min. After the first round of PCR, the reverse primer (2 µM) was added to the reaction. The PCR procedure was repeated. After the second round of PCR, Dpn I (10 U; NEB) was added and the reaction was incubated at 37ºC for 1 hour. Five microliters of the mixture was transformed into 50 µl of *E.coli* competent cells using a standard heat-shock protocol. The correct plasmid was selected by sequencing.

To prepare RNAs for OT, DNA templates of the RNA molecules were amplified from pUC19, in which the RNA cDNA sequence had been placed at the center flanked by 500 bp on either side corresponding to the DNA sequence of the two handles. The T7 promoter supporting transcription by T7 RNA polymerase was included in the DNA templates. The 100 µl volume transcription mixture included 1 µg of DNA template, 0.5 mM rNTPs, 5 mM $MgCl_2$, and 0.2 µg/µl T7 RNA polymerase. The reaction was performed in transcription buffer containing 40 mM Tris-HCl pH 8.0, 0.1 mM spermidine (Sigma), 0.01% Triton X-100, 10 mM DTT (Dithiothreitol), and supplemented with 2.0 U/mL thermostable inorganic pyrophosphatase (New England Biolabs, Inc.). After 3 h incubation at 37ºC, the reaction was quenched with an equal volume of 2x RNA loading dye (90% formamide, 5% glycerol, 0.1 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol FF). The nascent RNA was extracted from a

1% agarose gel using glass wool and centrifugation at 13,000 RPM followed by extraction with acid phenol:chloroform (pH 4.5) and ethanol precipitation. RNA was then dissolved in distilled water.

DNA handle A was PCR amplified from bases 2587 to 401 of the pUC19 plasmid and DNA handle B was PCR amplified from bases 412 to 892. Handle B was dual-biotinylated at its 5´ end using a dual-biotinylated primer. The digoxigenin (DIG) group was incorporated at the 5´ end of handle A using terminal deoxynucleotidyl transferase (Fermentas). In brief, 300 pmol of handle A was incubated with 7500 pmol of DIG-dUTP and 60 U of enzyme at 37ºC for 3 h. Enzymes were removed with phenol:chloroform:isoamyl (25:24:1), pH 8 and the excess DIG-dUTP was removed by using a 3K amicon filter column.

RNA fragments for OT analysis were annealed with DIG-handle A and dual biotin-handle B in which RNA (1.5 µg) was mixed with 3 µg of DNA-handle A and 3 µg of DNA-handles B and 80 µl annealing buffer (80% formamide, 1 mM EDTA, 40 mM PIPES pH 6.3, 0.4 mM NaCl). The mixture was heated to 85ºC for 10 min and slow cooled to 65ºC for 90 min, 55ºC for 90 min and 10ºC for 10 min. The annealed product was precipitated with 3 M sodium acetate and 100% ethanol, rinsed with 70% ethanol and dissolved in 100 µl H_2O.

2.2.2. Optical tweezers

Dual laser beam optical tweezers (183) with one optical trap was used to studying the folding properties of the TSS. Dual beam lasers were steered by mirrors, refocused using 100x magnification optical lenses, and passed through a flow chamber in the

opposite direction, creating an optical trap that holds a 4 μm bead coated with anti-DIG antibody. The position of the trapped bead was monitored by collecting the light scattered off the bead onto a position sensitive detector. Another 2 μm streptavidin-coated bead was mounted on a micro glass pipette using suction. The position of the micropipette was manipulated by controlling the piezoelectric flexure stage. Force was applied to the RNA molecule held between the two beads by moving the micropipette and optical trap away from each other. Changes in the extension of the RNA were measured by the relative movement of the trapped bead and the piezoelectric flexure stage. Experiments were performed in 10 mM Tris-HCl pH 7, 250 mM NaCl and with either 10 mM EDTA or 10 mM $MgCl_2$.

2.2.3. Force-ramping and force-clamping experiments

RNA molecules were subjected to a force range from 0 pN to 25 pN. In force-ramping, force was applied to a single RNA at a constant rate of 100 nm/s to unfold the molecule. This was followed immediately with refolding by decreasing the force at the same constant rate. In force-clamping, the RNA was held at a constant force for 1.5 or 2 min and change in extension of the RNA over the time was measured. Force was increased by increments of 0.5 pN and measurements retaken to determine the unfolding pathway. In both experiments, any change in extension of the RNA was monitored from the change in the position of the beads; the applied force was determined from the change in the momentum of the laser light passing through the trapped bead. Force and extension were recorded at a rate of 1000 Hz.

2.2.4. Data analysis

Data analysis was performed at 200 Hz for force-ramping and at 1000 Hz for force-clamping. Contour lengths of the handles were fitted using the Makko-Sigga WLC model (Equation 2.1), as the model is more accurate at a low force regime (184). When fitting the force extension curves of the DNA/RNA handles, the persistent length was limited between 8 and 10 nm (132, 142). Force-clamping data was analyzed by fitting the histogram of the unfolded and folded states with Gaussian distribution to obtain the displacement of the two states. The contour length of the RNA transition was estimated based on the Odjik WLC (equation 2.2) (184).

Equation 2.1

$$F = \left(\frac{k_B T}{L_p}\right)\left[\frac{1}{4\left(1-\frac{x}{L_0}\right)^2} - \frac{1}{4} + \frac{x}{L_0}\right]$$

Equation 2.2

$$x = L_0\left[1 - \frac{1}{2}\left(\frac{k_B T}{F L_p}\right)^{1/2}\right] + \frac{F}{K}$$

In which: F, force; x, extension; L_p, persistent lenth; L_0, contour length; K, elastic modulus; $k_B T$, Boltzmann's constant times absolute temperature.

2.3. Results

2.3.1. Folding and Unfolding of the TSS

Genetic analysis and biochemical probing suggest that the TSS and surrounding sequences can adopt two conformations (Figure 1.11). One conformation is favored for ribosome binding (Figure 1.11A) while the second is associated with RdRp binding (Figure 1.11B) (75, 81). To investigate how these different conformations might interconvert, I examined the mechanical properties and folding/unfolding pathway of the TSS using OT. The TSS, along with upstream A-rich sequence and 10 additional residues

and 8-nt downstream residues (118-nt fragment total) (Figure 2.1A) were joined at the 5´ and 3´ ends to 500-nt RNA handles, which were made double stranded (ds) after hybridization to cDNA tagged with biotin or DIG (Figure 2.1B). The biotin DNA handle was attached to a streptavidin-coated polystyrene bead and the other was attached to an anti-digoxigenin (AD)-coated bead. The streptavidin bead was held in place by a micropipette fixed on a piezoelectric stage and the AD bead was held in an optical trap. By moving the piezoelectric stage mechanically, the pipette was pulled away from the optical trap, allowing the connected single-stranded TSS to unfold. Alternatively, the pipette was repositioned proximal to the AD-bead, leading to TSS refolding (Figure 2.1B) (130, 185).

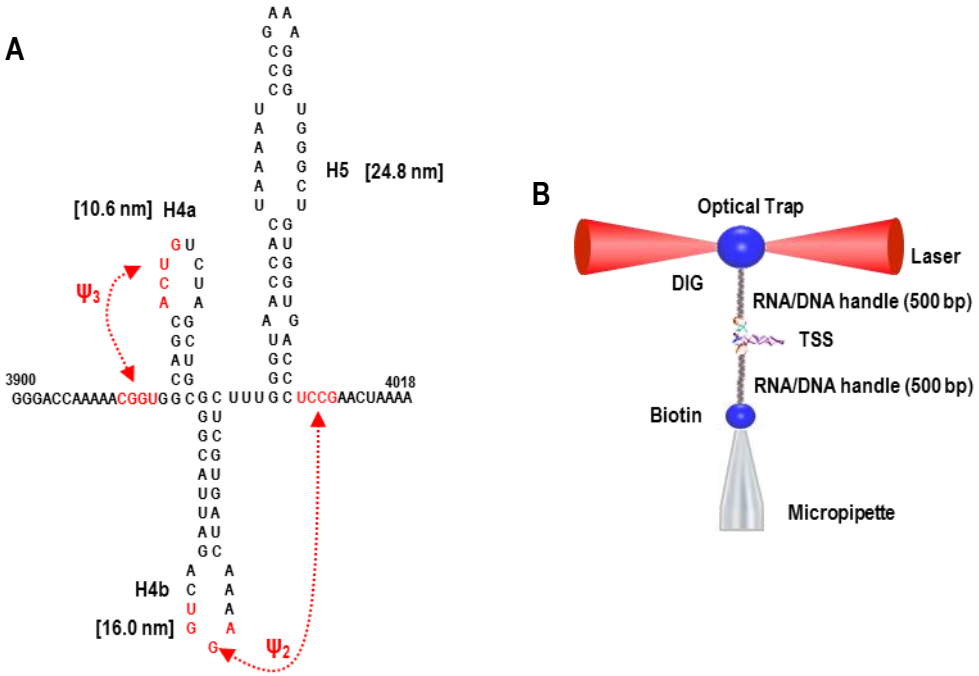


Figure 2.1. OT experimental set-up
A. The TSS fragment that was used displaying secondary structure and known tertiary interactions. Pseudoknot sequences are highlighted in red and denoted by arrows. Predicted lengths of the elements are in parentheses. **B.** Experimental OT set-up. TSS RNA was placed between 500-bp DNA/RNA handles that interacted with either the anti-DIG bead or the streptavidin bead. The former bead was trapped by lasers and the latter by suction. RNA stretching was accomplished by moving the two beads away from another, allowing measurement of the RNA extension length.

The unfolding pathway of the TSS was initially investigated using force ramping, a technique whereby RNA is unfolded by pulling the beads apart at a constant velocity (100 nm/s) and then releasing allowing the beads to move closer together, refolding the RNA. Figure 2.2A shows a typical force/extension curve for a single DNA/RNA handle. When the DNA/RNA hybrid was stretched, the force magnitude increased with a gently rising slope. However, when the extension of the handle reached its crystallographic length (contour length- the total bp length of the double helix), the slope became sharper. The stretching curve of a single handle molecule corresponds to the elastic response of a ds DNA/RNA hybrid following the worm-like chain model (WLC) and fitting the force/extension curves (FEC) of the DNA/RNA handle using the Marko-Sigga WLC model (Equation 2.1, Materials and Methods) (130, 184). The fitted contour length of the DNA/RNA handles was 290±1.07 nm (N = 10), which corresponds to its total length (0.29 nm/bp for 1000 bp) (130, 142). There were no transitions observed in the FECs of the DNA/RNA handle at forces ranging from 0-40 pN (Figure 2.2A), the typical force range used to stretch RNA structural elements (130).

The TSS contains three hairpins (H4a, H4b and H5), and two pseudoknots (Ψ_2 and Ψ_3) that join H4a and H4b loops with nearby sequences (Figure 2.1A) (74). Since tertiary interactions like pseudoknots are generally dependent on Mg^{2+} (130), the TSS folding pathway was investigated in both the absence and presence of Mg^{2+}. In the absence of Mg^{2+}, the TSS was expected to unfold with three transitions corresponding to the three hairpins. At low force (<10 pN), the TSS extended monotonically corresponding to the unfolding of the DNA/RNA handles (Figure 2.2B), indicating a

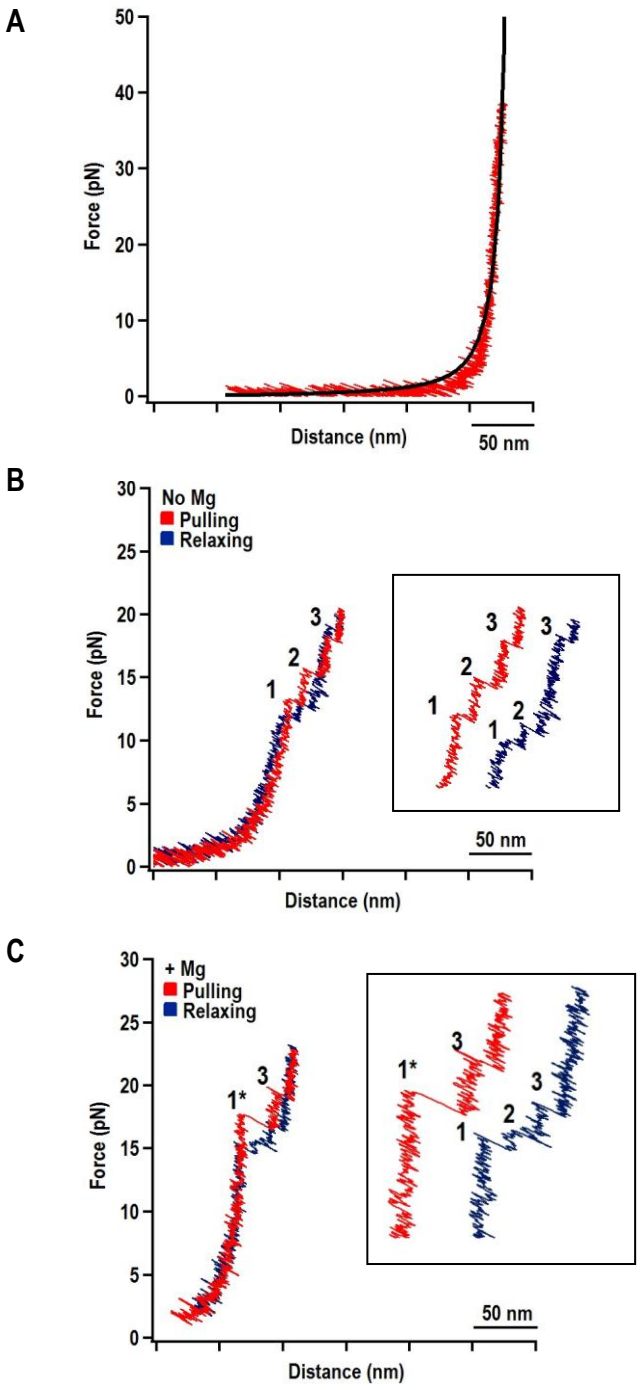


Figure 2.2. Representative force extension curves (FEC) for the TSS using force ramping
A. FEC of DNA/RNA handles (red) fitted with the Makko-Sigga WLC model (black). The extension length of a base-pair is 0.29 nm, and the persistent length of DNA/RNA was set between 8 and 10 nm. Total contour length of the DNA/RNA handles was 290±1.07 nm. **B**. Representative FEC of the TSS in the absence of Mg^{2+}. The three intermediates are labeled. The red line represents the unfolding process and blue line represents the refolding process. Insets show the transitions enlarged. Rips 1, 2, 3 are labeled. **C**. Representative FEC of TSS in the presence of Mg^{2+}. The large and small rips in the unfolding process are labeled 1* and 3, respectively.

single molecule was being stretched (it is possible to get more than a single nucleic acid attached to the beads). At higher force, three transitions (termed “rips”) labeled 1, 2, and 3 appeared in stepwise fashion (Figure 2.2B), corresponding to successive unfolding of three TSS structural elements. No rips were observed when the force exceeded 20 pN (data not shown). When the force was released by allowing the two beads to move back together, the TSS refolded in three successive steps corresponding to its three unfolding transitions (3, 2, and 1). The difference between refolding force and unfolding force for each transition denotes the hysteric characteristics of the two processes, suggesting that these transitions take place outside of their thermal equilibrium.

In the presence Mg^{2+}, the TSS unfolded with a new large rip (1*) and one small rip that appeared identical to rip 3 that was present in the absence of Mg^{2+}. The TSS refolded in Mg^{2+} with three small transitions (that were found to correspond to 3, 2, and 1, see below) (Figure 2.2C). These results indicate that, in the presence of Mg^{2+}, the TSS unfolded through the simultaneous rupture of multiple hairpins giving the large extended length of rip 1*. In addition, 80% of the FECs (N = 64) contained rip 1* in the presence of Mg^{2+} while only 14% of the curves (N = 86) contain rip 1* in the absence of Mg^{2+}. This indicates that Mg^{2+} allows for increased cooperativity between RNA elements of the TSS.

2.3.2. The order of TSS hairpin folding

In the previous section, the TSS was shown to unfold with three rips in the absence of Mg^{2+}. Based on the Mfold-predicted ΔG of the TSS hairpins, H4a should unfold first (ΔG= -6.9 kcal/mol), followed by H4b (ΔG= -9.2 kcal/mol) and H5 (ΔG= -21.8 kcal/mol). To investigate if this is the order of hairpin unfolding, a complementary

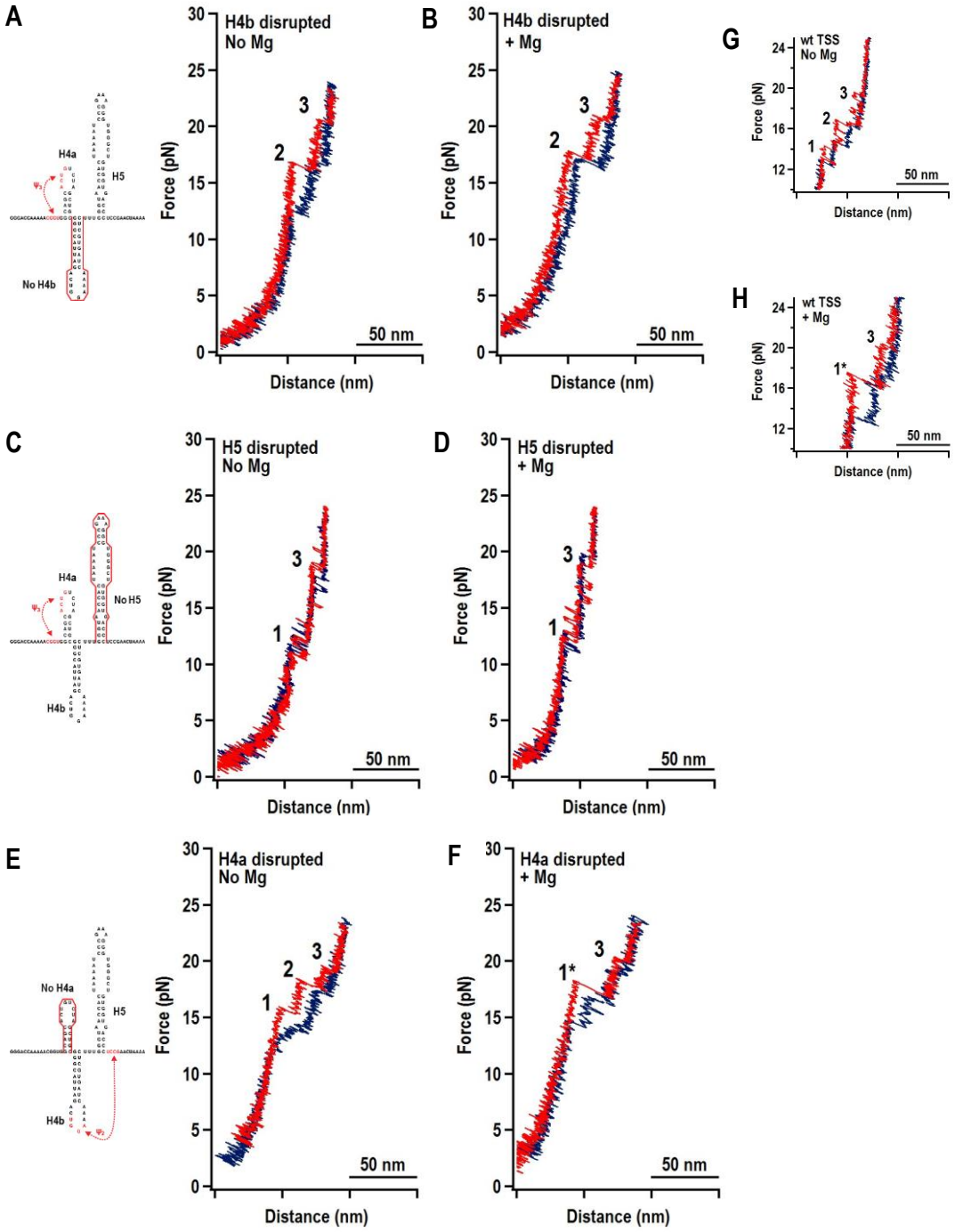


Figure 2.3. Representative FECs of TSS when disrupted by complimentary oligonucleotides
The TSS was annealed with an oligonucleotide complementary to H4b **(A, B)**; H5 (**C, D)**; H4a (**E, F)**, and OT conducted in the absence (A, C, and E) and presence (B, D, and F) of Mg^{2+}. The secondary structure of each disrupted TSS is presented to the left of the FECs, with the complimentary oligonucleotide denoted with a red line. Note that the oligonucleotide complementary to H4a has no effect on the FEC. **G, H,** FECs of wt TSS presented here for ease in comparisons. Rips are numbered as in wt TSS

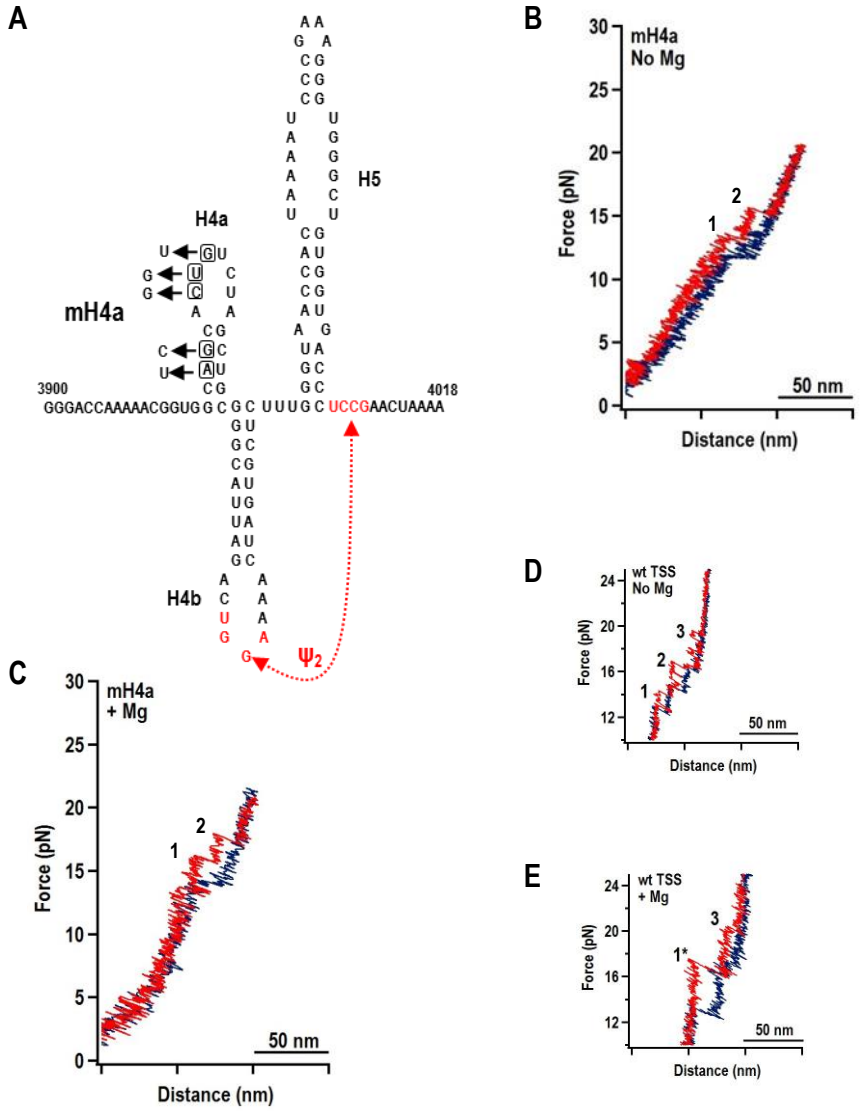


Figure 2.4. The folding pathway of the mutant TSS
A. Mutations generated in H4a (mH4a). **B, C.** FECs of mH4a in the absence and presence of Mg^{2+}. The two rips on the FECs are indicated. **E, F.** FECs of wt TSS are presented for ease in comparisons.

oligonucleotide approach was used to individually disrupt each hairpin and then the TSS was subjected to OT.

In the absence of Mg^{2+}, disrupting H4b eliminated rip 1, and disrupting H5 eliminated rip 2 (Figure 2.3). This suggests that the unfolding pathway of the TSS starts with H4b followed by H5. Disrupting H4a had no effect on the three rips, suggesting that

the oligonucleotide did not hybridize to the hairpin. When stretched in the presence of Mg^{2+}, disrupting H4b eliminated large rip 1*, and still generated rips 2 and 3. Disrupting H5 also eliminated rip 1*, and still generated rips 1 and 3 (Figure 2.3). These results strongly suggest that, in the presence of Mg^{2+}, H4b and H5 together generate large rip 1*. This large rip may also contain Ψ_2, which connects H4b with the base of H5. Thus, Mg^{2+} significantly increases the stability of H4b and facilitates cooperativity between at least H4b and H5.

Assignment of H4b and H5 to rips 1 and 2 suggested that H4a or H4a/Ψ_3 formed rip 3. However, the oligonucleotide complimentary to H4a did not alter the wt three TSS rips (Figure 2.3 E, F). One possible explanation is if stability of the H4a/Ψ_3 structure precluded oligonucleotide annealing to the H4a sequence. To identify whether rip 3 corresponds to the unfolding of H4a/Ψ_3, point mutations (mH4a) were introduced into the TSS by site-direct mutagenesis (Figure 2.4A). Subsets of these mutations were previously shown to disrupt Ψ_3 and H4a by in-line structure probing (75, 81). Using mH4a in the absence of Mg^{2+}, rip 3 was no longer observed above 18 pN (Figure 2.4B, C), suggesting that this rip represents H4a/Ψ_3. In the presence of Mg^{2+}, rip 3 was also absent. Curiously, in 10 repetitions, rip 1* was also absent and replaced with rip 1 and 2. This suggests that the absence of H4a/Ψ_3, affects the cooperation between H4b and H5.

No rips were observed when a fragment containing only H4a was stretched (Figure 2.5), suggesting that H4a is not stable without Ψ_3. This result is consistent with structure probing results that showed H4a is destabilized when Ψ_3 is disrupted (75). Since H4a by itself is not a stable hairpin, the high stability of H4a/Ψ_3 rip 3 can be attributed to the presence of Ψ_3. Since rip 3 is observed in the absence of Mg^{2+} (Figure 2.2B and 2.3A,

C), the pseudoknot can form and is stable in monovalent salt. In summary, in the absence of Mg^{2+}, the TSS unfolded in the order of H4b, H5, and H4a/Ψ_3. In the presence of Mg^{2+}, the TSS unfolded the H4b/H5 complex, followed by H4a/Ψ_3.

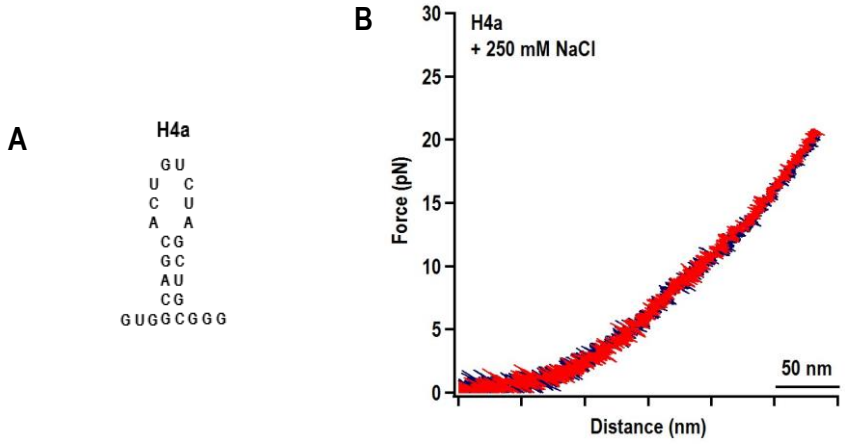


Figure 2.5. The unfolding/folding pathways of H4a in the absence of Ψ_3
A. H4a fragment in the absence of Ψ_3. **B.** FEC of H4a stretched in buffer containing 250 mM NaCl. No transition is observed, indicating that the hairpin by itself is not stable.

2.3.3. The effect of Mg^{2+} on stability and folding kinetics of the TSS elements

Since the three TSS rips were observed to unfold at a small range of different forces, force clamping experiments were performed to identify the limits of the force range. In these experiments, the RNA molecule was held at a constant force between 1.5 and 2 mins. At critical force levels, an RNA hairpin transitions reversibly ("hops") between its folded and unfolded states (Figure 2.6), creating upper and lower baselines in the extension/time curves.

As shown in Figure 2.6, slowly increasing the force extends the lifetime that an element stays in the unfolded state. For example, the lifetime of the unfolded state of H4b was < 20% of the total time at 11.5 pN, increasing to 96% at 13.5 pN (11.5-13.5 pN) and 100% at 14 pN and higher (14.5-20 pN) (Figure 2.6). Based on the increased

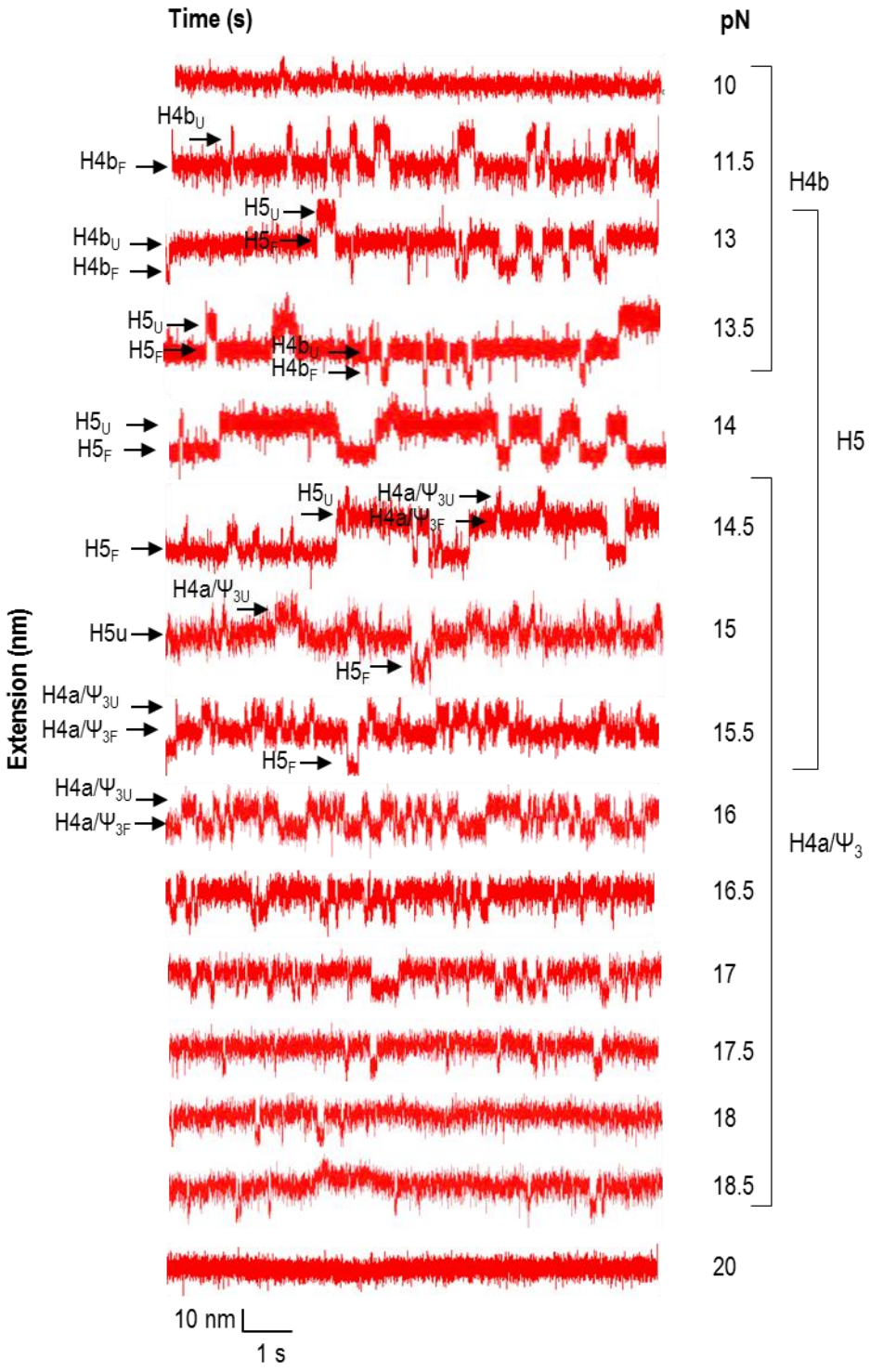


Figure 2.6. The extension versus time traces of the TSS at various force constants in 250 mM NaCl
The TSS was held at constant force for between 1.5 and 2 min. Force was increased by 0.5 pN intervals. The assigned hairpin for each transition is indicated. U and F are denoted unfolding state and folding state of the hairpins, respectively. Force range of each hairpin is indicated.

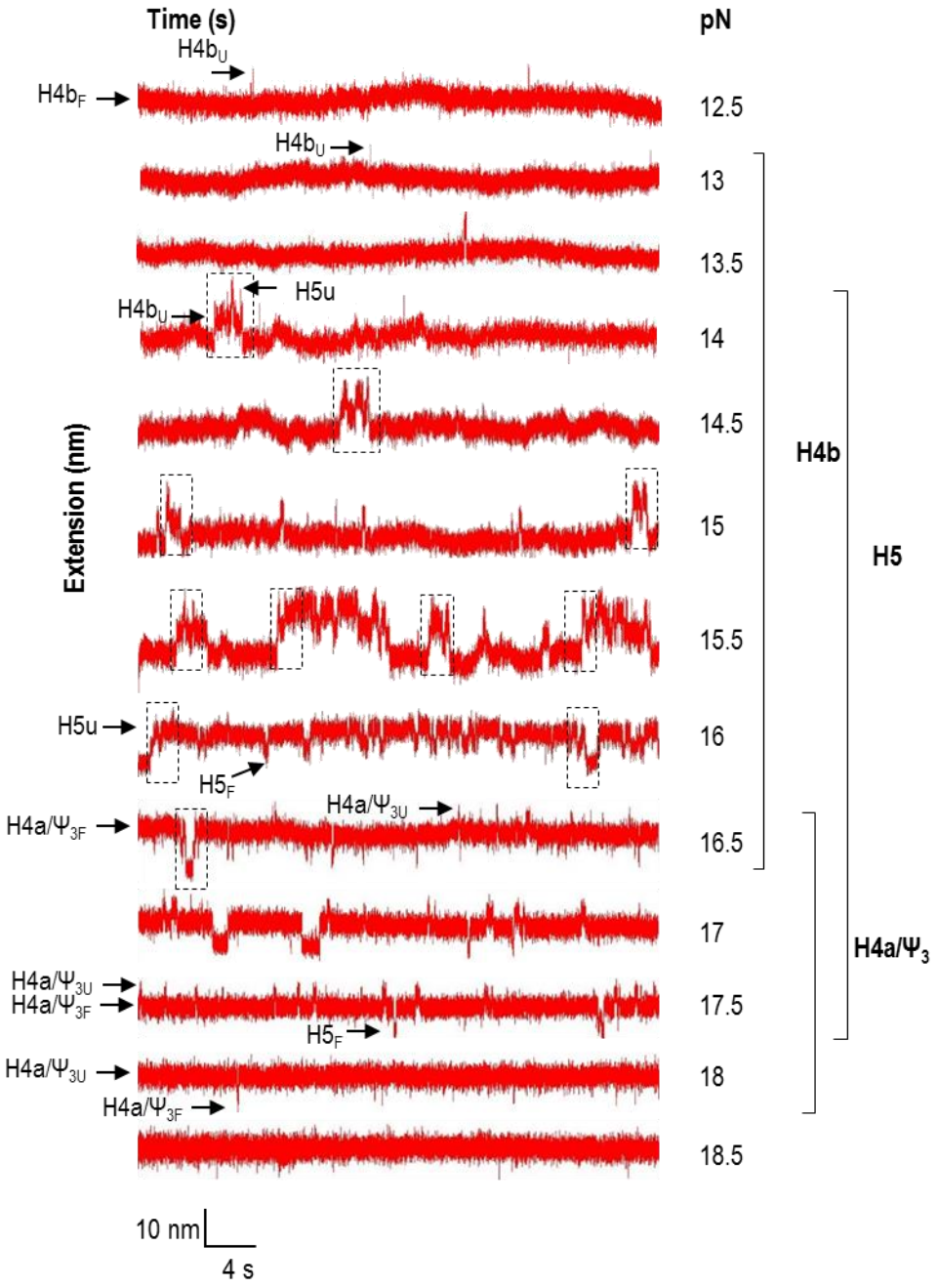


Figure 2.7. The extension versus time traces of the TSS in 10 mM $MgCl_2$
The TSS was held at constant force between 1.5 and 2 min. Force was increased by 0.5 pN. The assigned hairpin for each transition is indicated. Transitions in which the unfolded H4b led to the unfolding of H5 are boxed.

lifetime of the unfolded state of each transition, force ranges required to unfold the corresponding hairpins and pseudoknots of the TSS were determined. In the absence of

Mg^{2+} (Figure 2.6), when the force is held constant, three hopping transitions (H4b, H5 and H4a/Ψ_3) were observed, consistent with the three transitions found in force ramping experiments. The force ranges to unfold these hairpins and pseudoknots were 10-13.5 pN for H4b, 13-15.5 pN for H5 and 14.5-18.5 pN for H4a/Ψ_3 (Figure 2.6). In the presence of Mg^{2+}, H5 unfolded between 14-17.5 pN, a slightly higher force range then observed in the absence of Mg^{2+}, indicating that Mg^{2+} slightly increases the stability of H5 (Figure 2.7). The addition of Mg^{2+} also slightly increased the stability of H4b (13-16.5 pN) and H4a/Ψ_3 (16.5-18 pN) (Figure 2.7). These results suggest that Mg^{2+} increases the kinetic barriers of the TSS hairpins and pseudoknots.

In the absence of Mg^{2+}, an increased lifetime of the unfolded state for each transition was observed at each force increment. Multiple hopping of H4b and H4a/Ψ_3 transitions between unfolded and folded states was also detected at each force level (Figure 2.6). In the presence of Mg^{2+}, the lifetime of the H4b unfolded state was less than 1% of the total time at 15 pN, changed abruptly to 40% at 15.5 pN, and was nearly 100% at 16 pN (Figure 2.7). A similar sharp change in the lifetime of the unfolded state occurred for the H4a/Ψ_3 transition: less than 1% of the total time at 17.5 pN, and 100% of total time at 18 pN (Figure 2.7). Very few H4b and H4a/Ψ_3 hops were observed. The above results suggest that the unfolding kinetics for H4b and H4a/Ψ_3 differs from that of H5 when Mg^{2+} is present.

2.3.4. The extended length of the elements of the TSS

In the force clamping experiment, measuring the displacement of an RNA element's transitions between the unfolded state and the folded state allows for estimation of the extended length of an element (Figure 2.8). The extended length is then converted

into its contour length using Odjick WLC (Equation 2.2, Materials and Methods) (184). Details of the contour lengths for each transition measured at each force level are presented in Table 2.1. H4b and H4a/Ψ_3 maintained similar contour lengths at all force levels in the absence of Mg^{2+}, whereas H5 had a slightly longer length (~ 3 nm) at high force (Table 2.1A). The contour lengths of each transition at different force level were averaged and compared with their predicted lengths based on their sequences.

Based on the number of nt in the hairpins and pseudoknots, the contour length of H4b was estimated to be 15.93 nm and the experimental contour length (in the absence of Mg^{2+}) was 15.3±1.3 nm. For H4a/Ψ_3, the estimated length was 13.57 nm and the experimental value was 33% lower at 9.1±1.5 nm. The reduced length observed

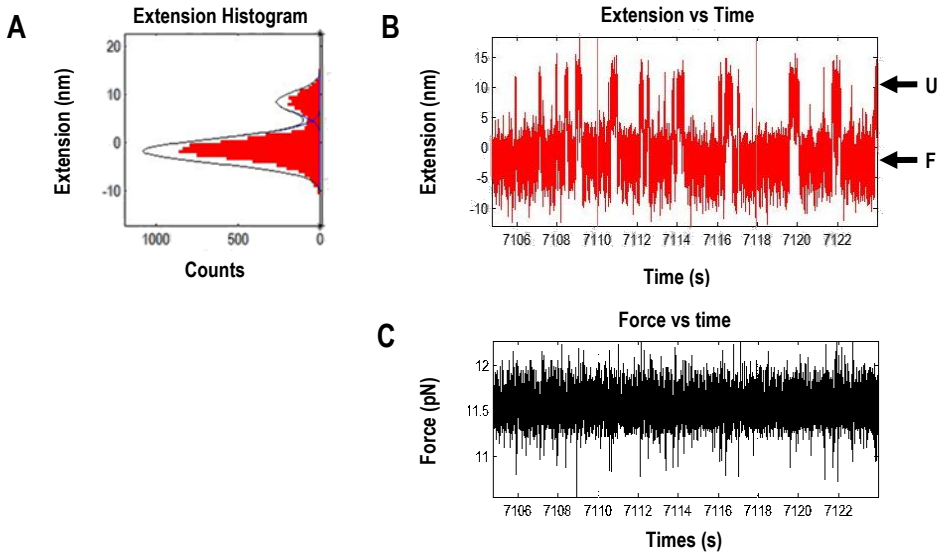


Figure 2.8. Measuring the displacement of RNA transition between its unfolded (U) and folded (F) states

A. Extension histogram (red) of each state was fitted with Gaussian function (black line). The length of each transition was obtained by subtracting the extension of the two states. **B.** The length versus time trace (red) of the TSS, which was used for extension analysis. **C.** Force versus time traces (black) during the force clamping experiments. During force clamping, the force feed-back was turned on, allowing force to be maintained constantly. At each force level, extension length of the intermediate states was obtained from three independent measurements of extension using the time trace of extension and then converted to contour length using the Odjik WLC model.

A

Force (pN)	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16	16.5	17	17.5	18	18.5	Ave
H4a/ Ψ_3										9.9	9.2	9.4	9.4	9.4	9.5	9.5	7.9	8.2	9.1 ± 1.5
										1.2	1.0	0.6	0.4	0.2	0.2	0.9	0.7	0.4	
H5							15.3	16.4	17.2	18.5	16.3	18.4							17.0 ± 1.8
							0.4	0.3	0.2	1.1	1.1	0.1							
H4b																			
Contour length (nm)	16.7	16.8	13.9	15.1	14.4	14.7	15.1	15.1											15.3 ± 1.5
Std	0.6	0.5	0.3	0.5	0.5	0.5	0.7	0.1											

B

Force (pN)	13	13.5	14.5	15	15.5	16	16.5	17	17.5	18	Ave
H4a/ Ψ_3							8.23	9.59	7.07	13.5	9.9 ± 3.6
							1.01	0.36	0.8	1.04	
H5						16	14.9	13.8	13.4	14.7	16.4	16.9	15.2 ± 2.5
						1.28	1.9	1.28	1.82	0.69	0.93	0.24	
H4b											
Contour length (nm)	13.7	13	15	16.9	14.6	15.8	16.6				15.8 ± 3.3
Std	-	0.46	1.51	2.1	3.04	3.72	0.85				

Table 2.1. Summary of contour lengths of TSS elements obtained by force-clamping
A. The contour lengths of each transition in 250 mM NaCl. **B. T**he contour lengths of each transition in 10 mM $MgCl_2$. Std: standard deviation obtained for three independent measurements. The contour lengths of each transition obtained at different force levels were averaged (Ave).

for H4a/Ψ_3 might reflect the formation of the shorter Ψ_3 stem (compared with its predicted structure by Mfold) that was observed in in-line structure probing (74).The estimated length of H5 was 24.78 nm and the experimental contour length was 30% lower at 17.0±1.8 nm. The reduced length observed for H5 might reflect the instability of the H5 lower stem (74), which might have unfolded at a low force range.

In the presence of Mg^{2+}, H4b had a similar average contour length (15.8 nm), however the contour lengths obtained at 14.5-16 pN were slightly longer (1-3 nm) than those obtained at 13 pN and 13.5 pN. In addition, significant fluctuation of the H5 contour length was observed in the presence of Mg^{2+}. Interestingly, full length H4a/Ψ_3 was only obtained at 18 pN in the presence of Mg^{2+}. These results indicate that in Mg^{2+}:

(1) the measured contour length of H4b, H5 and H4a/Ψ_3 depends on force; and (2) higher force causes unfolding with longer times at the transition states, indicating increased stability of the hairpins and pseudoknot.

2.3.5. Interconnection between RNA elements in the TSS

From the force ramping and force-clamping experiments, interconnections between RNA elements in the TSS were observed in the presence of Mg^{2+}. The unfolding force range for H4b (13-16.5 pN) and H5 (14-17.5 pN) partially overlapped (Table 2.1B), and unfolding of H4b promoted H5 unfolding at 14-16.5 pN (Figure 2.7). Since the only connection between H4b and H5 is Ψ_2, this suggests that Ψ_2 is contributing to the cooperativity between the two hairpins. In addition, since H4b is unfolding at a higher force in the cooperative structure, this suggests that Ψ_2 is also stabilizing H4b. To examine the effect of Ψ_2 on the unfolding of H4b, a force-clamping experiment was performed with H5 disrupted by oligonucleotide hybridization. My rationalization was that in the absence of H5, Ψ_2 would not form and H4b should then unfold at the same force range as it does in the absence of Mg^{2+}. Under these conditions, H4b unfolded at 11.5 pN in Mg^{2+} (Figure 2.9), which was the same force level as in the absence of Mg^{2+}, suggesting that Ψ_2 was absent (Figure 2.6). The above results indicate that in the absence of Ψ_2 and Mg^{2+}, H4b is less stable and that in the presence Ψ_2 and Mg^{2+}, unfolding of H4b is linked to the unfolding of H5.

An additional observation that arose from force clamping was that H5 appeared to influence the formation of H4a/Ψ_3. Under constant force, H4a/Ψ_3 unfolded between 14.5–18.5 pN in the absence of Mg^{2+} (Figure 2.7 and Table 2.1A), and between 16.5–18 pN in the presence of Mg^{2+} (Figure 2.8 and Table 2.1B). When H5 was disrupted and force was kept constant in the absence of Mg^{2+} (Figure 2.9B), no hopping of H4a/Ψ_3 was observed between 16–18 pN (Figure 2.9A). However, the hop transition of H4a/Ψ_3 was observed at

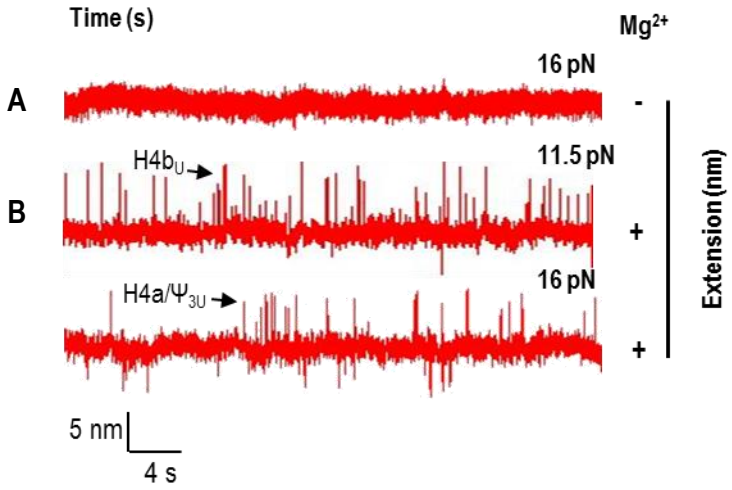


Figure 2.9. Length versus time traces of the H5-disrupted TSS
The complementary oligonucleotide of H5 was annealed to the TSS molecule and force clamping was performed on the H5-disrupted TSS using 5–20 pN force range. **A.** The hopping behavior of H4b in the H5-disrupted TSS was observed at 11.5 pN in the presence of Mg^{2+}. **B.** The hopping behavior of H4a/Ψ_3 in H5-disrupted TSS at 16 pN in the absence (**A**) or presence of Mg^{2+}. Note that, in absence of H5, no hopping of H4a/ Ψ_3 observed in the absence of Mg^{2+}, indicating that the structure is no longer forming.

a force range of 16–18 pN in the presence of Mg^{2+}, suggesting that Mg^{2+} countered the negative effect of the absence of H5, allowing the stable formation of H4a/Ψ_3 (Figure 2.9B). Although H4a/Ψ_3 was not observed by force clamping, a rip at high force was present in the force ramping FECs of disrupted H5 (Figure 2.3 C, D), which may reflect the force loading rate as others have found (134).

Using mH4, where H4a/Ψ_3 is disrupted, a new transition was observed between 10–13.5 pN in Mg^{2+}, the length of which was 7.38±0.98 nm at 10 pN (Figure 2.10). This transition is suggested to be unfolding of Ψ_2 and the first few bases of the H4b stem (Ψ_2^*), as the H4b transition was now observed at a lower force range (13.5 pN compared with 15 pN), and contained a shorter length (11.3±1.1 nm compared with 15.8 nm), which is expected in the absence of Ψ_2. The hopping behaviors of H4b and H5 in mH4a in Mg^{2+} (13-16.5 pN; Figure 2.10) were similar to those found for the two hairpins in the

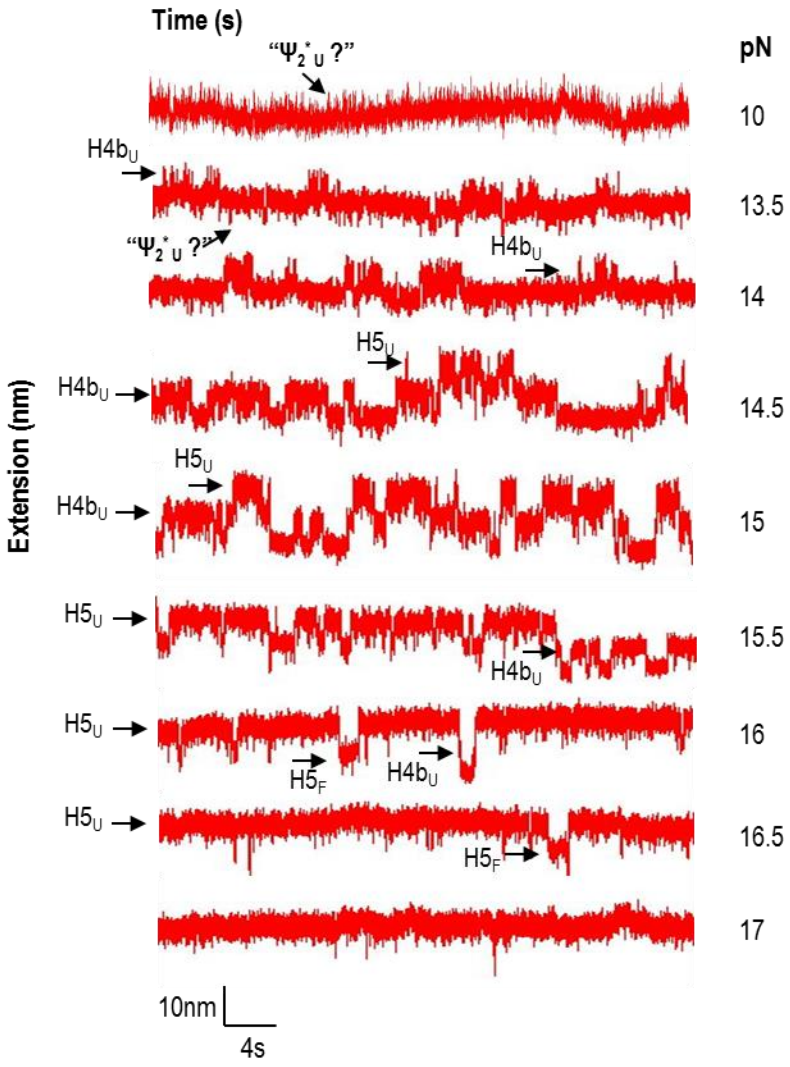


Figure 2.10. Length versus time traces using mH4a in the presence of Mg^{2+}
In mH4a, H4a and Ψ_3 are disrupted by mutations (Figure 2.4A). Ψ_2^* is denoted for the unfolding of Ψ_2 and the first few bases of the H4b stem.

wt TSS in the absence of Mg^{2+} (Figure 2.6, 13-15.5 pN), which was in contrast to their coupled behavior observed using wt TSS in Mg^{2+} (Figure 2.7, 14.5-16.5 pN). These results suggest that in the absence of H4a/Ψ_3, Ψ_2 is unstable and unfolds at low force creating the short transition at 10 pN (Figure 2.10).

2.3.6. Effect of Mg^{2+} and upstream A-rich sequence on the stability of H4a/Ψ_3

When the TSS fragment was subjected to MD simulations, the order of unfolding was predicted to be: Ψ_2, followed sequentially by Ψ_3, H4a, H4b and H5 (Shapiro and Jing, personal communication). In contrast, my data demonstrates that H4a/Ψ_3 is the most stable component of the TSS. One possible explanation is that H4a/Ψ_3 is stabilized by additional tertiary interactions. Previous reports indicated that mutations disrupting H4a/Ψ_3 also change the flexibility of the upstream A-rich sequence and vice versa (75, 81), suggesting that the A-rich sequence is interacting with H4a/Ψ_3. In the TSS structure solved by NMR-SAXS, this A-rich sequence was not included and two guanines were added to the 5' end of the H4a/Ψ_3 sequence, which were needed to stabilize the TSS (76). To better understand the folding of H4a/Ψ_3 and whether its enhanced stability is due to interaction with the A-rich sequence, the folding pathway of H4a/Ψ_3 was studied using a short RNA fragment that contained H4a/Ψ_3, the A-rich sequence, and several additional nucleotides on both sides to reduce any disturbance of the handles on H4a/Ψ_3 structure (Figure 2.11A).

Consistent with the unfolding/refolding pattern of H4a/Ψ_3 observed in FECs of the full-length TSS fragment (Figure 2.2B), a single H4a/Ψ_3 transition was observed in 10 mM Mg^{2+} at 19.95±0.06 pN with refolding at 16.29±0.04 pN (Figure 2.11B). Force ramping in 10 mM NaCl or 250 mM NaCl resulted in a single rip at forces that were 2-5 pN lower, suggesting that Mg^{2+} is exerting a stabilizing effect on the element (Figure 2.11C). In addition, the extended length of H4a/Ψ_3 was 10 nm in the absence of Mg^{2+} compared with 13 nm at 18 pN in Mg^{2+} (Figure 2.11B; Table 2.1A and B). This suggests that H4a/Ψ_3 may be nucleated in the absence of Mg^{2+}.

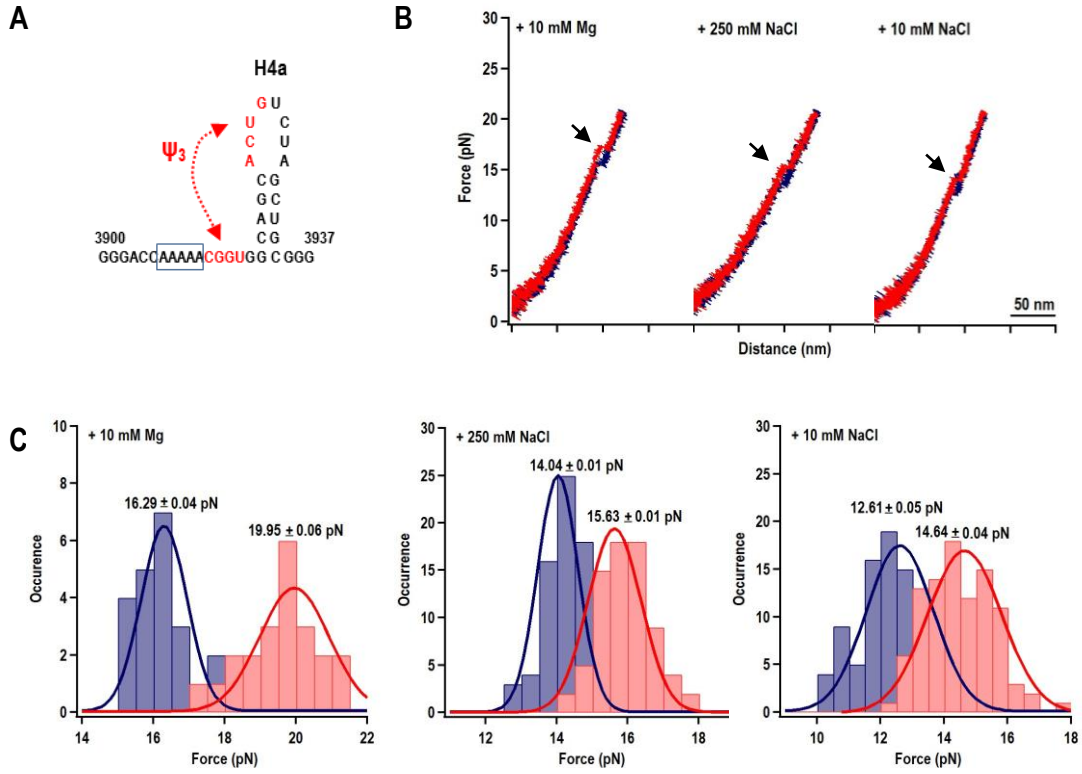


Figure 2.11. FECs of H4a/Ψ$_3$ under different salt conditions
A. H4a/Ψ$_3$ fragment. **B.** FECs of H4a/Ψ$_3$ in 10 mM Mg^{2+} (N = 22), 250 mM NaCl (N = 72), or 10 mM NaCl (N = 96). The H4a/Ψ$_3$ rip is indicated by an arrow. **C.** Distribution of the unfolding force (red) and folding force (blue) of H4a/Ψ$_3$. Critical force was obtained by Gaussian fitting. The A-rich sequence is boxed.

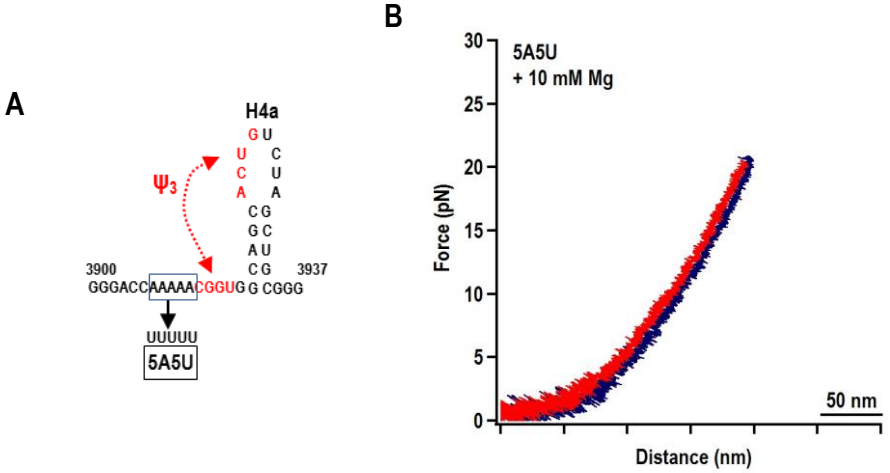


Figure 2.12. FEC of 5A5U mutant in the presence 10 mM Mg^{2+}
A. The five adenylates upstream (boxed) of H4a/Ψ$_3$ were changed to five uridylates using site-directed mutagenesis, generating 5A5U. **B.** FEC of 5A5U in 10 mM $MgCl_2$.

To examine the impact of the A-rich sequence on the folding dynamics of H4a/Ψ_3, the 5 adenylates in the H4a/Ψ_3 fragment were replaced with 5 uridylates, generating mutant fragment 5A5U. Using 5A5U, no transitions were observed in the presence or absence of Mg^{2+} (Figure 2.12). This result strongly suggests that the A-rich sequence plays a critical role in the stability of the H4a/Ψ_3 element.

2.4. Discussion

Conformational switches have been observed for many regulatory RNA elements found in 5´ and 3´UTR of (+)-strand RNA viruses (121, 162, 186). However, the unfolding/folding pathways for these elements remain elusive. To better understand the dynamics of one RNA structural switch, I investigated the unfolding/folding pathway of the TSS, a highly structured RNA located in the 3´UTR of TCV that adopts two biologically relevant conformations. The conformation assumed by the TSS under physiological conditions is required for ribosome binding (Figure 1.11A), whereas an alternate conformation is present upon RdRp binding (Figure 1.11B). Disruption of the native translation-promoting structure of the TSS resulted in reduced ribosome binding while enhancing *in vitro* transcription (75, 81). The TSS also reverts to the native conformation compatible for ribosome binding when the RdRp is degraded. Based on previous biochemical, genetic, and molecular modeling studies, the TSS was predicted to fold into a stable tRNA-shaped structure, which is the favorable conformation for ribosome binding to the P or E-site (75). Due to the highly stable structure of the TSS, the folding/unfolding mechanism that allows the element to adopt two conformations remained unclear.

Optical tweezers was used to unfold and refold a single TSS molecule to more

closely investigate its structural dynamics. The TSS consists of three hairpins (H4a, H4b, and H5) and two pseudoknots (Ψ_2 and Ψ_3) (74). Based on the Mfold-predicted ΔG for the three hairpins, H4a (in the absence of the pseudoknot) is the least stable hairpin (ΔG = -6.9 kcal/mol), followed by H4b (ΔG = -9.2 kcal/mol), and H5 (ΔG = -21.8 kcal/mol). The ΔGs of H4b and H5 correlate with the unfolding order of these hairpins in the absence of Mg^{2+}; H4b unfolded between 10–13.5 pN, and H5 unfolded at 13–15.5 pN (Figure 2.6). Curiously, no rip corresponding to H4a alone was found. Instead, my results indicate that Ψ_3 does not require Mg^{2+} to form and together with H4a is the most stable element of the TSS, unfolding at high force 14.5-18.5 pN (Figure 2.4 and 2.6). In Mg^{2+}, the unfolding pathway of the TSS starts with disassembly of a cooperative structural complex composed of Ψ_2, H4b and H5, followed by H4a/Ψ_3 (Figure 2.2C).

Data from previous studies indicated that RdRp binding to the 3´ end of TCV disrupts H4a/Ψ_3 (81). Destabilization of Ψ_3 also led to the disruption of the H4a stem, which increased minus-strand synthesis (81). Data from the above section indicated that unfolding of H4a/Ψ_3 affected the unfolding of Ψ_2 and H5. Thus, H4a/Ψ_3 may play an important role in the conformational switch of the TSS. A more compact H4a/Ψ_3 forms under low monovalent salt conditions (10 mM NaCl) in the absence of Mg^{2+} (Figure 2.11). Full-length H4a/Ψ_3 (13 nm) was obtained only in Mg^{2+} at high force (≥18 pN) (Table 2.1B), indicating that H4a/Ψ_3 is a very stable structure that can be nucleated in a wide spectrum of biological conditions possibly needed for association with ribosomes. Unexpectedly, stability of H4a/Ψ_3 depends upon the upstream A-rich sequence as the element does not form in its absence (Figure 2.12). This result is consistent with previous findings that replacing the A-rich sequence with a U-rich sequence decreased ribosome

binding and translational efficiency in luciferase constructs (75, 81). In addition, In addition, replacement of individual adenylates with cytidinates reduces TCV accumulation in protoplasts by ~50% (Xue Feng, unpublised) and replacement of two adenylates results in 20-fold decreases in accumulation (81). It is possible that the A-rich sequence functions like an A-minor motif, which promote stabilization of ribosomal RNA structures and *Tetrahymena thermophila* ribozymes through the insertion of its adenine bases into the partnering RNA structure (187, 188). The requirement for adjacent adenylates to stabilize H4a/Ψ_3 offers an explanation for why my results differ significantly from MD simulations, which do not incorporate the A-rich sequence and in which the TSS unfolds beginning with Ψ_2, followed sequentially by Ψ_3, H4a, H4b and H5 (Shapiro and Jing, personal communication).

Mg^{2+} is important for stabilizing RNA tertiary structures (111, 172-174) and has been shown to be a key factor in controlling the conformational switch in a number of RNAs (171, 189, 190). When Mg^{2+} was added, the stability of all three TSS transitions increased by 2-3 pN over the force required in the absence of Mg^{2+} (Table. 2.1 and Figure 2.7). These results indicate that Mg^{2+} has increased the kinetics barriers of these hairpins and pseudoknots. In addition, the hopping behavior of H4b and H4a/Ψ_3 changed abruptly from folded to unfolded states between 15.5 and 16 pN for H4b and between 17.5 pN and 18 pN for H4a/Ψ_3 (Figure 2.7). This result might serve as evidence to explain the refolding dynamics of the TSS, which assumes a conformation compatible with ribosome binding in the absence of the RdRp (81).

Additionally, there are several lines of evidence that indicate cooperation between individual RNA elements within the TSS. Ψ_2 is proposed to increase cooperativity

between H4b and H5 in Mg^{2+}, which are included in the large 1* rip in TSS FECs (Figure 2.2C). Disruption of H4b or H5 led to the disappearance of this large transition, regardless of the buffer condition (Figure 2.3). In addition, when H5 is disrupted by oligonucleotide hybridization, Ψ_2 would not form, H4b became less stable (transition at 11.5 pN in Mg^{2+} (Figure 2.9A), at which force level H4b transitions in the wt TSS were only observed in the absence of Mg^{2+} (Figure 2.6). In the presence of Mg^{2+}, H4b, which only connects promoted unfolding of H5 in a narrow force range between 14.0 pN and 16.5 pN (Figure 2.7). The second connection between elements was observed for H5 and H4a/Ψ_3 by force clamping (Figure 2.9). When H5 was disrupted, the H4a/Ψ_3 transition was no longer observed at its corresponding force range in the absence of Mg^{2+}, but was still observed in Mg^{2+} (Figure 2.9). These results indicate that H5 is required for folding of H4a/Ψ_3, which is countered by the addition of Mg^{2+}. The third connection was between the H4a/Ψ_3 and Ψ_2 in the force-clamping experiments. In mutant mH4, in which H4a/Ψ_3 was disrupted, a short transition that might correspond to the unfolding of Ψ_2^* was observed at 10 pN, 3 pN lower than the force obtained for the large transition 1* complex in the wt TSS in Mg^{2+} (Figure 2.7 and 2.10). The unfolding of Ψ_2 at a lower force also had an effect on the folding behavior of H4b and H5, in which the two transitions were observed to hop between the folded and unfolded states more frequently than they did in the wt TSS (Figure 2.7 and 2.10). The observed connections between RNA elements of the TSS in Mg^{2+} provide further information for how the TSS assumes a conformation favoring RdRp binding. It is known that in the presence of RdRp, Ψ_2 and H4a are disrupted (81). Thus, if Ψ_2 is disrupted first, this might promote H4b and H5 to hop back and forth between their unfolded and folded states more frequently. Due to the

structural connection between the H5 and H4a/Ψ_3, the unfolding of H5 can cause unfolding of H4a/Ψ_3. On the other hand, if RdRp disrupts H4a/Ψ_3 first, this may cause Ψ_2 to unfold, leading to the unfolding of H4b and H5. In contrast, since Mg^{2+} acts as a highly stabilizing agent for the hairpins, when the TCV enters cellular environment, Mg^{2+} can stabilize the TSS in the initial uncoated viral RNA for ribosome binding. Due to the unfolding of H4b and H4a/Ψ_3 at higher force range in presence of Mg^{2+}, these abrupt, high kinetic barriers for H4b and H4a/Ψ_3 might suppress unfolding of the TSS allowing the TSS to facilitate translation.

Chapter 3: The structural study of *Pea enation mosaic virus* kissing loop tRNA-shaped structure

3.1. Introduction

As presented in Chapter 1, the efficiency of translation initiation of many (+)-strand RNA viruses depends on structural element located within the 5' UTR and 3' UTRs (5, 7). Structural element involved in translation initiation known as CITEs are located at the 3' UTR (Chapter 1) and are commonly found in the *Tombusviridae* and *Luteoviridae* as well as in the genus Umbravirus (5, 191). CITEs are categorized into 7 structural groups: ISS, TED, PTE, YSS, BTE, TSS and kl-TSS based on their predicted secondary structures (Chapter 1) (5). These RNA elements are critical determinant for efficient translation in plant viruses and can enhance translation by 20-fold (5). CITEs interact with ribosomes or translation initiation factors and are proposed to transfer these factors to the 5' UTR by forming long-range RNA:RNA interactions or ribosome bridges (23, 75, 77). Due to their ability to function as a modular unit, some CITEs are interchangeable between viruses (74). The RNA structures of CITEs are most often determined by SHAPE and in-line probing analysis, or computational modeling because they are typically large in size (60-168 nt) (5). Lack of experimentally defined 3-D structures for CITEs limits our understanding of their role in translational mechanism. The TCV TSS is the only 3' CITE for which the structure is solved by both NMR and SAXS (76), and that structure included two guanylates for stability and did not include the upstream A-rich sequence that I showed was critical in Chapter 2. Thus, to obtain more information about structural basis for translation enhancement of 3' CITEs, I have carried out structural characterization of the PEMV kl-TSS, which plays a significant role in ribosome recruitment and translation enhancement in PEMV.

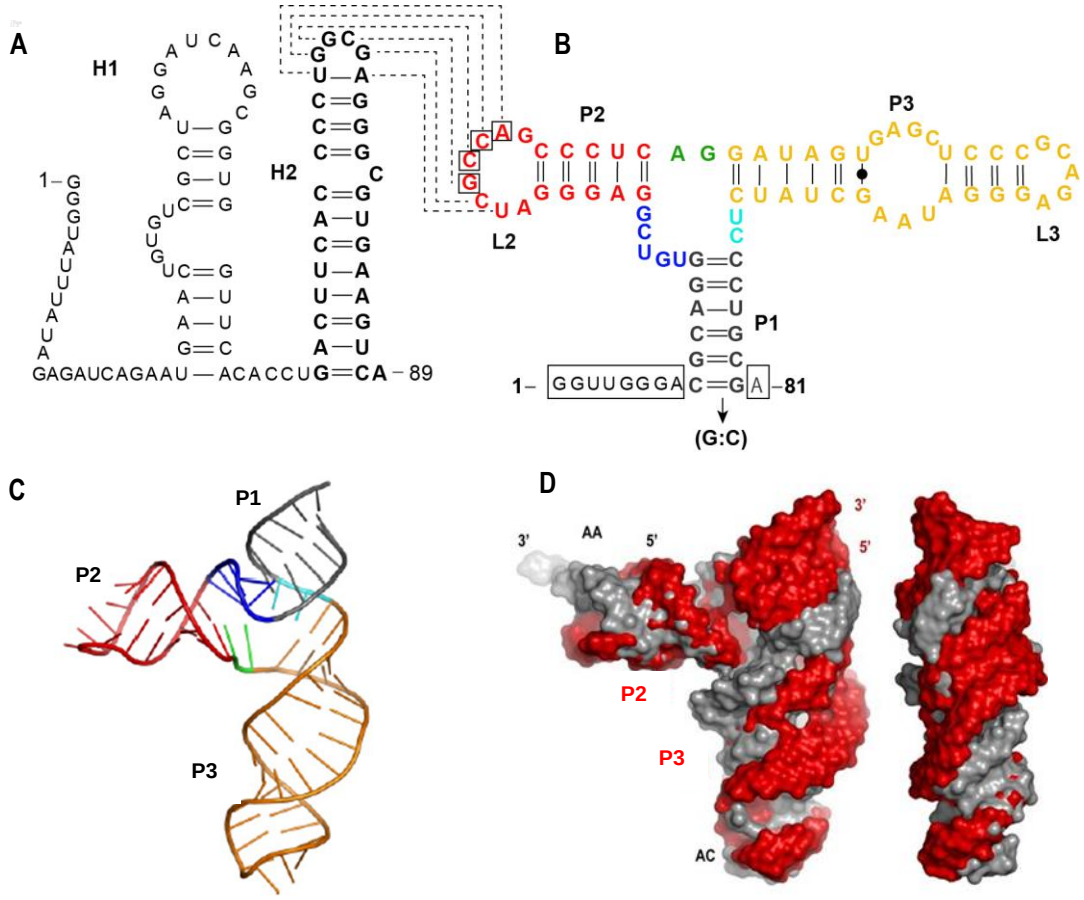


Figure 3.1. Predicted structure of the kl-TSS and the 5' 89-nt fragment by Mfold and in-line probing
A. Predicted structure of the 5' 89-nt fragment of PEMV. Hairpins 5H1 and 5H2 (bold) are indicated. Interactions between nucleotides of the 5H2 apical loop and L2 loop of the kl-TSS are connected by dashed lines. **B.** Predicted structure of the 81-nt kl-TSS of PEMV. Conserved residues in the L2 loop of the kl-TSS are boxed. The stems and loops of the kl-TSS are color-coded corresponding to its 3-D predicted structure presented in C. The boxed nucleotides at the 5' and 3' ends of the kl-TSS are not present in the 71-nt RNA fragment. In the 71-nt kl-TSS, the first C:G base-pair is replaced with G:C to support transcription by T7 RNA polymerase. **C.** 3-D model of the kl-TSS predicted with RNA2D3D. Stem P1, P2, P3 are indicated. **D.** Superimposition of the 3-D model of the kl-TSS (red) and Phe-tRNA (grey). Location of the amino-acceptor stem (AA), and anticodon stem of the tRNA are indicated. Structures of the kl-TSS are adapted from Gao et.al 2012 (51, 77).

As mentioned in Chapter 1, multiple translational enhancer element have been discovered in PEMV including the PTE, TSS and kl-TSS (192). The kl-TSS is positioned 9-nt upstream of a PTE CITE, which unlike other PTE, is incapable of interacting with 5' end sequences. Instead, the associated kl-TSS forms a long-distance kissing-loop (kl) interaction with a sequence located within the 5' 89-nt (51, 77) (Fig. 3.1A, B). The kl-TSS was predicted to form a tRNA-shaped structure that superimposes well on the 3-D

structure of the Phe-tRNA (PBD code 1EHZ) (Figure 3.1 C, D) and appears to be similar to the structure of the TCV TSS (51, 77). Like a variety of 3' CITEs found in other plant viruses, the kl-TSS has a conserved RNA:RNA interacting motif (GCCA) located in the apical L2 loop of the 5' P2 stem (59, 77). However, unlike 3' CITEs such as PTE that require binding to eIF4E for translation enhancement (60), translational enhancer activity of the kl-TSS is dependent on ribosome binding (77). The kl-TSS binds to both ribosomal subunits and to the 80S ribosome, but does not compete with the TSS for binding to the P-site (77). In contrast with the structure of the TCV TSS that contains two stabilized pseudoknots (74), the relative position of the three stems of the kl-TSS is dependent on the topology of its central 3-way junction (77). Mutations that disrupt the kl-TSS P1 or P3 stems, or truncate or enlarge the L3 loop, reduce ribosome binding by several fold (77). In contrast, compensatory mutations that reform the P1 stem allow for near wild-type (wt) levels of ribosome binding (77). From genetic and in-line probing data analyses, the kl-TSS L2 loop interacts with the loop of 5H2 within the 5' 89 nt fragment, and this interaction is compatible with ribosome binding (77). Solving the structure of the kl-TSS will provide insights into how this element interacts with ribosomes to enhance translation.

In this report, the interaction of the kl-TSS with 5H2 was confirmed using EMSA and NMR. The hydrogen bonding patterns of each stem in the kl-TSS were also established by NMR. By using SAXS in combination with computer modeling, tentative structural models of the kl-TSS are proposed. This analysis indicates that the kl-TSS might have two conformational states. The major conformation appears to differ

significantly from the previously predicted structure while the minor conformation appears somewhat similar to the predicted structure.

3.2. Material and methods

3.2.1. Preparation of RNA samples

RNAs used in this project were transcribed from either double-stranded or single-stranded DNA templates depending on the size of RNA fragment. The T7 promoter (5′ CTAATACGACTCACTATAG) was incorporated into DNA templates. In-house T7 RNA polymerase was prepared to perform *in vitro* transcription. To obtain the highest yields of transcribed RNAs, different transcription conditions were tested in 20 µl volumes. In the reaction, the concentration of $MgCl_2$ was varied from 4 to 15 mM, DNA templates from 0.3 to 0.6 µM, unlabeled rNTPs from 5 mM to 20 mM and the T7 RNA polymerase concentration from 0.1 µg/µl to 0.2 µg/µl. When isotopic ^{15}N, ^{13}C- rNTPs were used, their concentration was fixed at 5 mM. The reaction was performed in transcription buffer containing 40 mM Tris-HCl (pH 8.3), 0.1 mM spermidine (Sigma), 0.01% Triton X-100, 10 mM DTT, and supplemented with 2.0 U/mL thermostable inorganic pyrophosphatase (New England Biolabs, Inc.). A final concentration of 80 mg/mL PEG (8000MW) and (+)-strand T7 promoter oligonucleotide were added to the transcription reactions containing single-stranded DNA template. After 3 hours of incubation at 37ºC, the reaction was quenched with an equal volume of 2x RNA loading dye (90% formamide, 5% glycerol, 0.1 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol). To quantify the efficiency of RNA transcription, a 2.5 µl aliquot from each reaction was subjected to electrophoresis on 8M urea, 19:1 PAGE. The gels were stained with ethidium bromide and the RNA bands quantified using Gbox gel doc

(Biorad). Large-scale transcription (10 ml) was carried out with the optimal concentrations of Mg^{2+}, DNA template, rNTPs and T7 RNA polymerase. For large scale RNA preparations, the reaction was incubated at 37ºC for 3-8 hours and was then quenched with 50 mM of EDTA pH 8. DNA templates and proteins were removed by acid phenol:chloroform extraction (pH 4.5). The RNAs from the extraction were precipitated overnight with 0.3 M sodium acetate (pH 5.2) in combination with 3 volumes of 95% ethanol at -20ºC. Precipitated RNA was centrifuged and salt removed with 70% ethanol. The RNA was then re-dissolved with a minimal amount of water (3 ml) and purified using 8M urea denaturing PAGE (19:1). The percentage of acrylamide in the PAGE gel was selected based on the size of the purified RNAs. The corresponding RNA band was then visualized by UV shadowing and excised. The RNA was electro-eluted using an Elutrap (Owl) electro-separation system at 200 V for 8 hours at 4ºC. Collected RNA samples were combined and solvent exchanged using a spin filter with molecular weight cut-off (MWCO) of 3 kDa (Amicon ultra 3K, Millipore).

3.2.2. Native PAGE

RNA (1 μg) was heated in a water bath at 90ºC for 2 min and incubated on ice for 2 min. The RNA was subsequently added to a folding buffer containing 50 mM Herpes (pH 7.5), 50 mM NaCl, and various Mg^{2+} concentrations ranging from 0.1 mM to 20 mM. The mixture was incubated at 37ºC for 30 min. Loading dye (10% glycerol and 0.01 % xylene cyanol) was added to the RNAs before loading on a 12% native gel prepared with 29:1 acrylamide/bisacrylamide gel in a buffer containing 34 mM Tris, 66 mM Hepes (pH 7.5), 0.1 mM EDTA, 50 mM NaCl, and 1 mM $MgCl_2$. The native gel was

run at 4ºC and 300 V. Running buffer was re-circulated every hour. Gels were stained with ethidium bromide and observed with a Gbox (Biorad).

3.2.3. RNA:RNA electrophoretic mobility gel shift assays (EMSAs)

EMSA was performed as previously described (77). In brief, in vitro transcribed kl-TSS RNA was dephosphorylated using calf intestinal phosphatase (CIP) and ^{32}P end-labeled using polynucleotide kinase. The labeled kl-TSS and the unlabeled full-length or truncated 5H2 RNA were folded separately as described above. ^{32}P-labeled kl-TSS RNA (2 pmol) was mixed with 20 pmol of unlabeled 5H2 RNA in RNA binding buffer (5 mM HEPES (pH 7.6), 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 3.8% glycerol). The mixture was incubated for 30 min at 25°C. RNA samples were resolved on 10% native polyacrylamide gels at 4°C for 4 to 5 h in 1× THEM buffer [34 mM Tris, 66 mM HEPES (pH 7.5), 10 mM $MgCl_2$, 0.1 mM EDTA]. Gels were subsequently dried and exposed to X-ray film.

3.2.4. Small angle X-rays scattering collection, processing and three-dimensional shape reconstruction

Small angle and wide angle X-ray scattering data was collected at three different RNA concentrations (0.8 mg/ml, 1.6 mg/ml, and 3.2 mg/ml). The folding buffer contained 50 mM Hepes (pH 7.5), 50 mM NaCl, and 3 mM $MgCl_2$. Data was collected at beam line 12-ID of the Advanced Photon Source at the Argonne National Laboratory. The wavelength (λ) of tX-Ray radiation was set to 1.033 Å. Procedures for SAXS/WAXS measurement were similar to those previously described (117, 193). For each RNA concentration and matching background buffer, 30 images were collected.

SAXS data was averaged and the background-subtracted using the NCI-SAXS program package. The averaged scattering profiles of three RNA concentrations were used to extrapolate to a concentration of zero using PRIMUS in the ATSAS program package (http://www.embl-hamburg.de/biosaxs/). The GUINIER plot was plotted with ln(I(q)) vs q^2 to check sample quality and to obtain I_0 and radius of gyration (R_g) within the range of q_{max}*R_g < 1.3. The data from each RNA concentration was then normalized with I_0. Conformation of the RNAs was examined using the Kratky plot for q < 0.3 $Å^{-1}$. Scattering profiles of RNAs were then Fourier-transformed using tGNOM of the ATSAS package to obtain the normalized pair distance distribution graph.

Ab in initio modeling was performed using the program DAMMIN in a slow mode (Svergun 1999). For each RNA, 20 models were obtained. These models were filtered and averaged using the DAMPUP, DAMFILT and DAMAVER of the ATSAS package (http://www.embl-hamburg.de/bioSAXS). Normalized spatial discrepancy (NSD) between each pair of the models was computed. The model with the lowest NSD value was selected as the reference model for superimposing onto other models. Outliner models (2 models) with an NSD above mean + 2*variation were removed before averaging.

3.2.5. Conformational ensemble analysis

MC-Fold (194) was used to generate the secondary structure of the 71-nt kl-TSS, in which base-pairs of the P1, P2 and P3 stems were constrained using imino assignment of NMR. The first 10 secondary structure models with the lowest energy were submitted for MC-SYM to generate 3-D conformational ensembles of the kl-TSS (194). Each secondary structure consisted of an average of 130, 3-D conformational models, with a

total of 1650 generated conformational models. These 3-D conformational ensembles were then relieved, refined, scored with Amber force field using MC-SYM (webserver), and sorted by the radius of gyration obtained from SAXS. All back-calculated scattering profiles of these conformational ensembles were generated by using Fast X-ray Scattering (FoXS) (webseve) (195). The difference of scattering profiles between experimental and back-calculated data was expressed as χ (195):

Equation 3.1:
$$\chi = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\frac{I_{\exp}(q_i) - cI_{cal}(q_i)}{\sigma(q_i)}\right)^2}$$

Where c is the scaling factor, $\sigma(q_i)$ is the experimental error, $I_{exp}(q_i)$ and $I_{cal}(q_i)$ are the experimental and back-calculated scattering intensities, respectively, at each *i* data point of the total M data point. To obtain the model that best represents , a sparse ensemble fit was performed using the back-calculated scattering profiles of the kl-TSS conformational ensembles using the sparse ensemble solutions (SES) (196). One state, two or three state models were scored using their χ values. To avoid overfitting, the data sets was resampled using a bootstrapping protocol (196).

3.2.6. NMR

Folding of the kl-TSS was performed as described above except that the RNA was heated to 75ºC to avoid RNA aggregation when a high RNA concentration was used. The concentrations of the RNA samples subjected to NMR ranged from 0.1 to 0.8 mM and were dissolved in a 250 µl of NMR buffer [50 mM deuterated MOPS (pH 6.8), 30 mM NaCl, 1 mM $MgCl_2$, 10 % D_2O] and transferred into a Shigemi NMR tube. For indirect referencing of chemical shifts, ^1H sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was added to the NMR buffer for a final concentration of 0.25 mM. NMR

experiment were performed using a 600 MHz or 800 MHz Bruker Avance III spectrometer equipped with a HCN triple resonance cryoprobe. NMR experiment were performed at either 25ºC for samples in water, or 37ºC for samples in D_2O. All NMR data was processed using TopSpin 3.2 (Bruker Biospin) and NMRviewJ (197).

3.3. Results

3.3.1. The conformational analysis of the kl-TSS

An RNA sequence can adopt multiple conformations (123). Thus, the conformation of the kl-TSS was initially analyzed on a native PAGE gel. The kl-TSS was heated and snap-cooled in water and either immediate loaded onto the gel or incubated in folding buffer at 37ºC. As shown in Figure 3.2, the sample that was loaded without incubation migrated as a large smear, indicating that the RNA sample had not folded into a single conformation. Tight kl-TSS RNA bands were observed when the RNA was further incubated in 37ºC (Figure 3.2A, lanes 3-6,), indicating that the kl-TSS was folded homogenously. Various concentration of Mg^{2+} did not discernibly affect migration of the kl-TSS (Figure 3.2A).

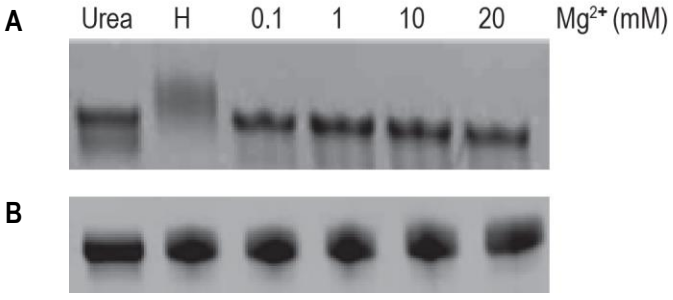


Figure 3.2. Folding of the kl-TSS
A. Native gel electrophoresis of the kl-TSS. RNA samples were folded in buffer containing either urea, or various concentrations of Mg^{2+}. Urea, RNA sample was heated in the presence of 6M urea. H, RNA sample that loaded immediately onto the gel after snap cooling. RNAs were resolved on 10% PAGE (29:1) in 34 mM Tris, 66 mM HEPES (pH 7.4), 0.1 mM EDTA, 50 mM NaCl, and 1 mM $MgCl_2$ at 4ºC. **B**. Denaturing 8M urea PAGE containing the same kl-TSS samples.

3.3.2. Base-pairing of the kl-TSS

In previous studies, the kl-TSS was used within an 81-nt fragment. However, only 71-nt of the RNA sequence was predicted to participate in forming a three way junction (Figure 3.1) (77). The upstream and downstream nucleotides were originally included with the 71-nt core kl-TSS (Figure 3.1B, boxed grey nucleotides) because the TCV TSS was found subsequent to NMR analysis to require the upstream A-rich sequence (74, 75). However, high flexibility of single-stranded regions may be problematic in structural studies, (e.g SAXS and NMR). To investigate the impact of these single-stranded nucleotides on the structure of the 71-nt kl-TSS, 2-D ^{1}H-^{15}N imino heteronuclear single quantum correlation (HSQC) NMR experiment were performed with the ^{15}N-labeled 71-nt and 81-nt PEMV kl TSS fragment. The first C:G base pair of the 71-nt kl-TSS was switched to a G:C base pair to facilitate *in vitro* transcription by T7 RNA polymerase. When base-paired, each imino proton of G and U is protected from solvent exchange, producing a single peak for GC and AU base-pairs (bp) and two peaks for GU bp on the 2-D ^{15}N-imino spectra (Figure 3.3A, B). When base-paired, C(N3) and A(N1) signals on the cross-strand are also obtained from heteronuclear *J(N,N)-HNN* correlation spectroscopy (COSY) NMR (Figure 3.3 A, C). In Figure 3.3B, the imino spectrum of the 81-nt kl-TSS contains 11 peaks of G(H1/N1), and 6 peaks of U(H3/N3) correspond to 10-GC, 5-AU and 1-GU base-pairs. The number of base-pairs obtained was close to the expected number of base-pairs (14-GC bps, 6-AU bps and 1-GU bp) from the predicted secondary structure of the 81-nt kl-TSS. The lower value for the number of the

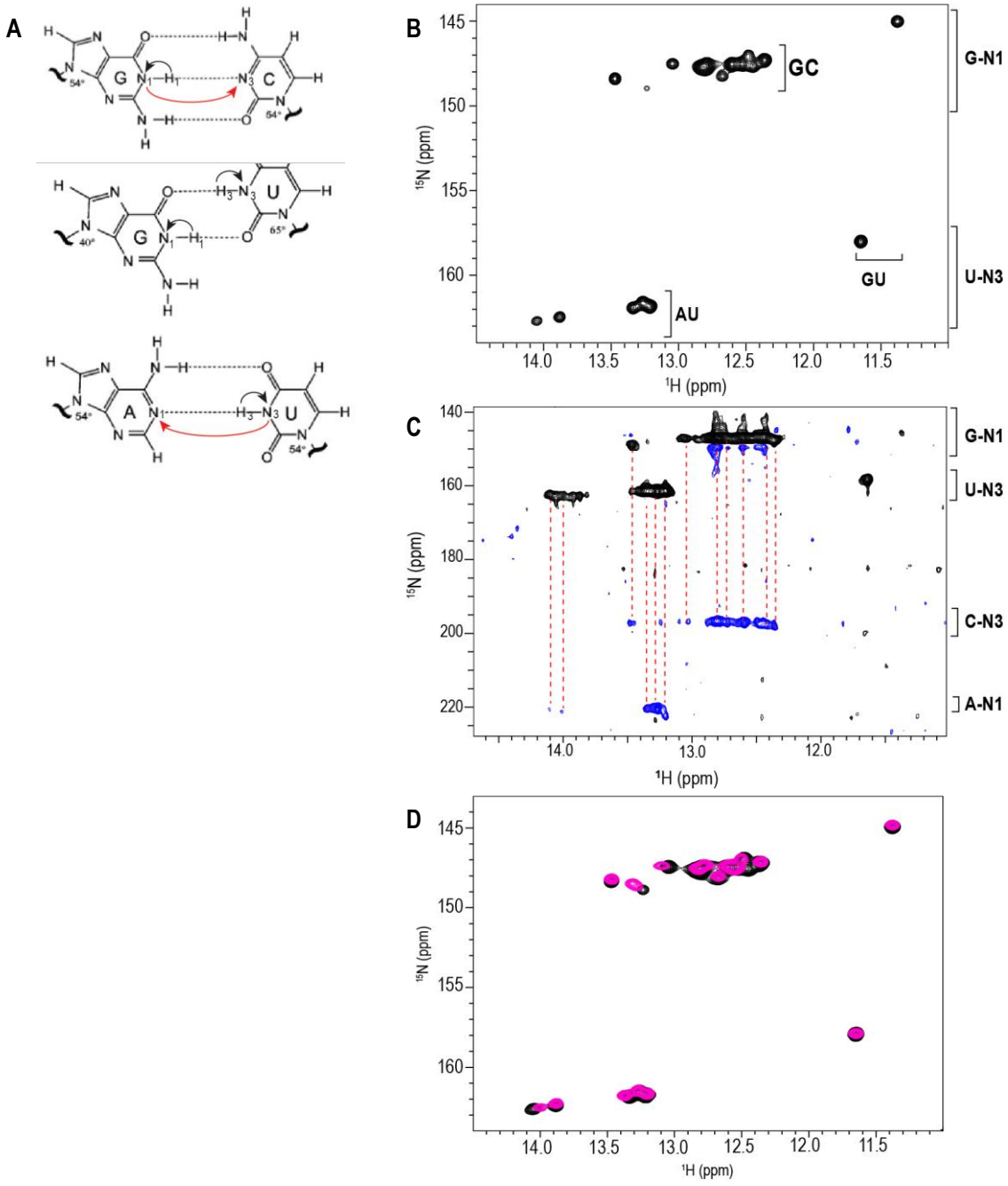


Figure 3.3. The ^{1}H-^{15}N HSQC spectra of the ^{15}N-labeled kl-TSS
A. Magnetization transfer between G-H1/N1, U-H3/N3 of GC, AU, and GU bps in ^{15}N-^{1}H HSQC (black arrows). In HNN COSY, magnetization was transferred between G-H1/N1 (black arrow)-C-N3 (red arrow) and between U(H3)/N3(black arrow)-A-N3(red arrow). **B.** ^{15}N-^{1}H HSQC spectrum of the 81-nt, ^{15}N-labeled kl-TS. The experiment was performed in 50 mM deuterated MOPS at pH 6.8, 30 mM NaCl, and 1 mM $MgCl_2$ at 37 ºC. Data was acquired on the 800 MHz spectrometer. The peaks represent resonances of proton G-H1 or U-H3 coupled to the resonances of nitrogen G-N1 or U-N3, respectively. Each peak corresponds to a bp. Regions for GC, AU and GU bp are indicated. **C.** The ^{15}N-^{1}H HNNCOSY spectrum of the 81-nt, ^{15}N-labeled kl-TSS. Resonance regions of G-N1, U-N3, C-N1 and A-N3 are labeled. Dashed lines indicate the magnetization transfer across the hydrogen bonds between G-N1 with C-N3 and U-N3

with A-N1. **D.** Overlap of the ^{15}N-^{1}H HSQC spectrum of the 81-nt, ^{15}N-labeled PEMV kl-TS (black) and the ^{15}N-^{1}H HSQC spectrum of the 71-nt, ^{15}N-labeled kl-TS (purple).

observed base-pairs could be due to spectral overlap. The N-H··N hydrogen bonds between the two helical strands of the 81-nt kl-TSS were established by detection of signals of C(N3) and A(N1) on the HNN COSY spectrum (Figure 3.3C). Peaks of C(N3) and A(N1) correlating with imino peaks in the spectrum of the 81-nt kl-TSS were also observed, further corroborating the hydrogen bonds present within the RNA. The ^{15}N spectra of the 81-nt and 71-nt kl-TSSes were superimposed and the overlaid spectra indicated that the two RNAs contain similar base pair patterns. Thus, the 5'- and 3' single-stranded tail sequences do not affect hydrogen bonding of the 81-nt fragment (Figure 3.3D). Therefore, the 71-nt kl-TSS RNA was chosen for further analysis.

3.3.3. The interaction of the 71-nt kl-TSS with 5' end hairpin 2 (5H2)

A previous study reported the interaction between the 81 nt kl-TSS and the 5' 89-nt genomic fragment that is co-terminal with the 5' end using electrophoretic mobility shift assays (EMSA) (77) (Figure 3.1A). Structure probing also revealed that the presence of the 5' 89-nt fragment altered the flexibility of the kl-TSS 6-nt apical loop (L2) of the P2 stem (77). Similar flexibility changes were also observed with the apical loop of 5H2. These results indicate that the apical loop of the 5H2 and L2 of the kl-TSS were responsible for the long-distance interaction (77). To determine if the interaction can be detected using isolated full-length and truncated 5H2 hairpin, EMSA was performed with ^{32}P kl-TSS and several truncated fragment (F1, F2 and F3) of the 5H2 (Figure 3.4A) (performed by Dr. Feng Gao). All 5H2 truncated fragment retarded the mobility of the

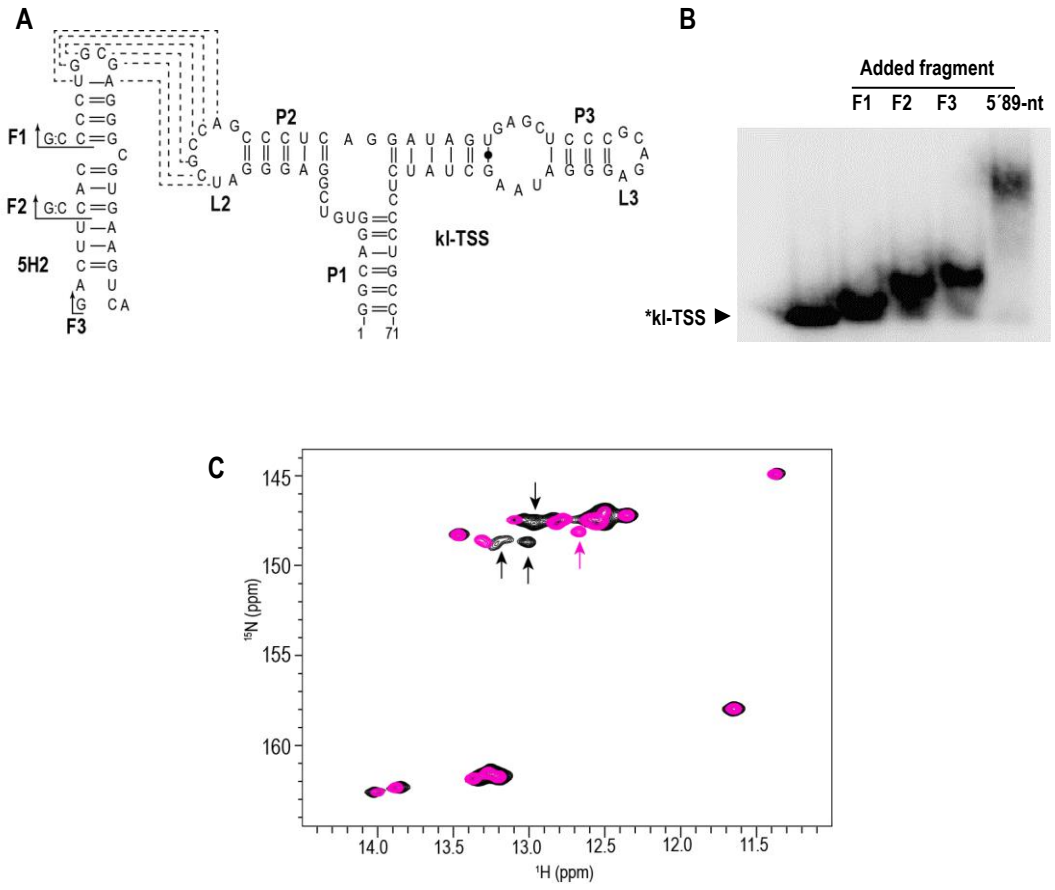


Figure 3.4. Interaction of the kl-TSS with truncated 5H2
A. Structure of 5H2 and kl-TSS. Interactions between the two loops are presented with dashed-lines. The locations of truncated fragment F1, F2 and F3 are indicated. The first CG base-pair of F2 and F3 fragment were replaced with G:C base-pairs to assist in transcription with T7 RNA polymerase. **B.** Gel shift assays between the ^{32}P-labeled kl-TSS (*kl-TSS) and truncated 5H2 fragment and the 5'89-nt genome fragment of PEMV. **C**. Spectra of the ^{15}N-labeled kl-TSS collected in the presence (black)/absence (pink) of the F1 fragment of 5H2. One peak disappears as indicated by the pink arrow. The shifted peaks are indicated with black arrows.

^{32}P kl-TSS, indicating that the complete 5'89 fragment or full-length 5H2 is not required for the interaction (Figure 3.4B).

Interaction between 5H2 and the kl-TSS was expected to form four G:C and two A:U base-pairs, for which the kl-TSS contributed one G and one U (Figure 3.4A). Thus, a ^{1}H-^{15}N HSQC experiment was utilized to investigate these interactions. New peaks were expected to appear on the 2-D ^{1}H,^{15}N-imino spectrum of the kl-TSS. In Figure 3.4C, the

2D ^{1}H-^{15}N HSQC spectrum of the kl-TSS obtained in the presence of the unlabeled F3 fragment of 5H2, was overlaid with the 2-D ^{1}H-^{15}N HSQC spectrum of the kl-TSS by itself. On the spectrum of the ^{15}N-labeled kl-TSS collected in the presence of unlabeled F3 fragment, one G-imino peak was absent and several G-imino peaks appeared to be shifted (Figure 3.4C), suggesting that new hydrogen bonds had formed between 5H2 and the kl-TSS. Further detailed analysis of the interaction between these two RNAs will be presented in section 3.3.5.

3.3.4. "Divide and conquer" approach to assign chemical shift resonances of the kl-TSS

In the imino spectra of the kl-TSS (Figure 3.3D), 11 out of 15 G and 6 out of 7 U imino resonance peaks were observed. The absence of peaks could be the result of spectral overlap, which would interfere with resonance assignments. Therefore, the "divide and conquer" approach was used to assign resonances. This approach was successfully applied to build a high-resolution structure of 26 kDa HCV IRES (198). Truncated fragment of the kl-TSS were designed based on its RNA2D3D structural model in which the P3 stem loop of the kl-TSS likely functions as an independent structural unit and the P1 stem and P2 stem loop are positioned close to each other due to folding of the junction (77). The first fragment (P3) was 32 nt long and contained only the P3 stem loop (Figure 3.5A). The second fragment (P1P2) was 53 nt long and contained P1, P2 and the basal stem of the P3 to maintain the conformational topology of the kl-TSS three-way junction. In P1P2, the GC and GU base-pairs just below the internal loop of the P3 lower stem were replaced with a single A:U bp and then capped with a GAAA tetraloop (Figure 3.5A) to maintain a P3-like hairpin. An additional 49-nt

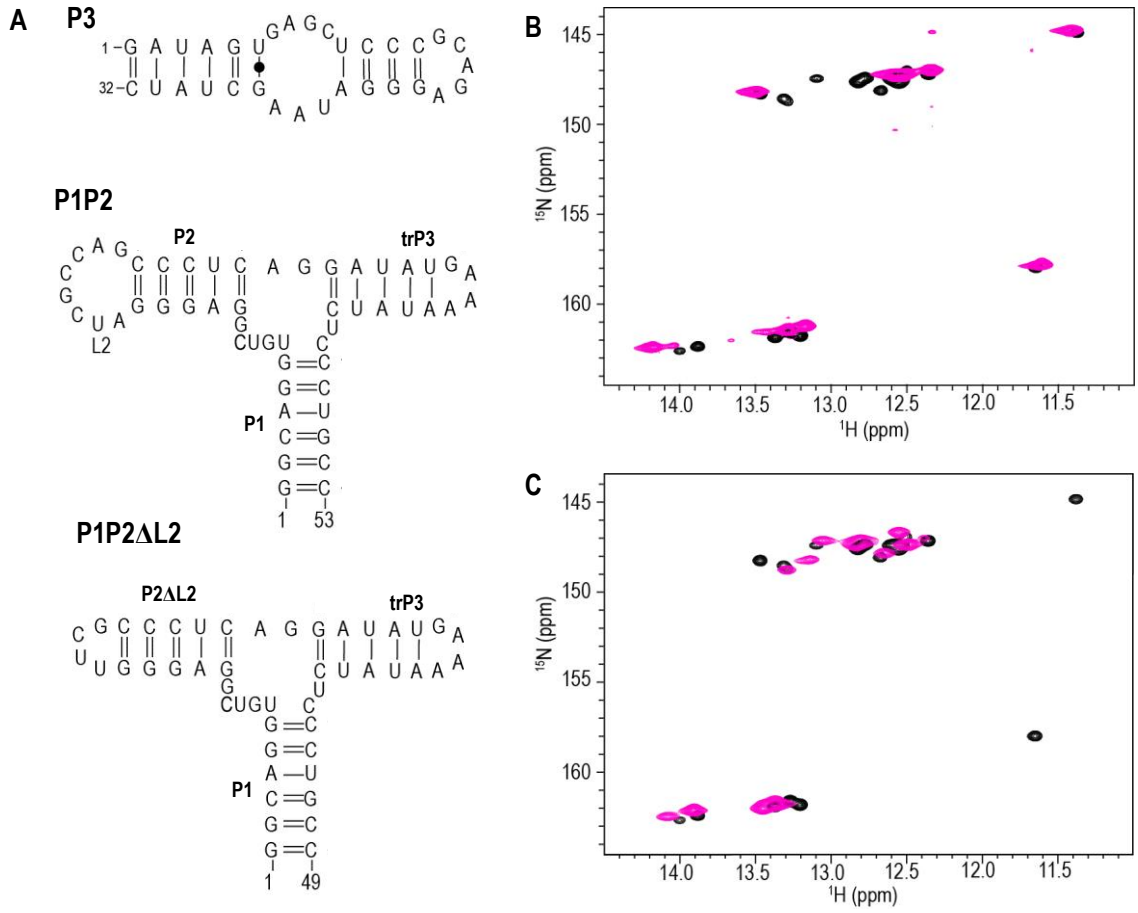


Figure 3.5. P3, P1P2 and P1P2ΔL2 fragments of the kl-TSS
A. Structures of P2, P1P2 and P1P2ΔL2 as predicted by Mfold. A GAAA loop replaced the top stem of P3 in the P1P2 construct. UUCG replaced loop L2 of P2 in the P1P2ΔL2 construct. **B.** Superimposition of the imino spectrum of ^{15}N-labeled P3 fragment (purple) and the imino spectrum of ^{15}N-labeled kl-TSS. **C.** Superimposition of the imino spectrum of ^{15}N-labeled P1P3 fragment (purple) and the imino spectrum of the ^{15}N-labeled kl-TSS. All imino spectra were collected in 50 mM deuMOPS (pH 6.8), 30 mM NaCl, 1 mM $MgCl_2$, 10% D_2O at 37°C.

fragment labeled P1P2ΔL2 that has the same hydrogen bond pattern of P1P2 was also designed, in which the 8-nt loop of P2 was replaced with a stable UUCG loop.

2-D ^{1}H-^{15}N HSQC was performed to compare base-pair pattern of these fragments with full-length kl-TSS. As shown in Figure 3.5B, the imino spectrum of P3 overlaid well with that of the kl-TSS, indicating that P3 alone has a similar base pairing pattern as P3 within full length kl-TSS. Based on the 2-D ^{1}H-^{15}N HSQC spectrum of the

^{15}N-labeled P3 (Figure 3.5B), nearly all base-pairs predicted from the secondary structure were observed (4 out of 5 G-imino peaks and 3 out of 4 U-imino peaks were observed). One G and one U had imino protons resonates at 11.5 ppm, indicating that they might form non-canonical base-pair interactions. In addition, the superimposed imino spectra between the ^{15}N-labeled kl-TSS and P1P2 (Figure 3.5C) indicated that P1P2 also maintained a similar base pair pattern as the kl-TSS. In the spectrum of P1P2, 9 out of 10 G-imino peaks were observed and 4 out of 6 U-imino peaks were observed (Figure 3.5C). The imino resonances of P3 and P1P2, which overlapped well with the imino resonances of the kl-TSS, rendered the spectra of these two fragment suitable for assignment of the full-length kl-TSS. In addition, the NOESY spectrum of P1P2ΔL2 is comparable with that of the P1P2 fragment (see below), rendered the NOESY spectrum of the P1P2ΔL2 suitable for assignment of the P1P2 (the ^{1}H-^{13}C 2-D spectra at ^{13}C6 and ^{13}C8 region of P1P2ΔL2 also overlaid well with that of the kl-TSS (Appendix)).

3.3.5. Resonance assignments for the imino protons of P3, P1P2 and the kl-TSS

Since space magnetization transfer can occur for imino protons within a distance of 5 Å, this transfer results in the appearance of cross-peaks seen in NOESY NMR spectra (Chapter 1) (146). The imino resonance assignment of the RNA is accomplished by sequentially connecting imino peaks on the same strand and on the cross strand. To assign imino chemical shifts for P3, P1P2, P1P2ΔL2 and the kl-TSS, NOESY experiments were performed on each RNA (Figure 3.6, 3.7 and 3.8).

Sequences within the basal stem of P3 were used for starting the assignment based on two prominent imino cross-peaks between a G (11.3 ppm) and a U (11.5 ppm) observed in the NOESY spectrum (Figure 3.6C). These up-field chemical shifts and their

cross-peaks suggest that the G and U form a wobble base pair. Since there is only one predicted wobble pair in the secondary structure of P3, these two chemical shifts likely arise from U37(H3) and G58(H1) resonances. G36 is covalently bonded with U37 and forms a cross-strand with G58 (Figure 3.6A). In addition, the chemical shift of G36(H1) also has cross-peaks with both U37(H3) and G58(H1). Thus, these connections provide an unambiguous starting point for assignment of the P3 stem. U60(H3) with a chemical shift of 13.3 ppm was assigned through the cross-peak connecting it to G36(H1). The chemical shift of U62(H3) was assigned based on its connectivity with G32 on the NOESY spectra of P1P2 and P1P2ΔL2 (Figure 3.7) and the kl-TSS (Figure 3.8). The

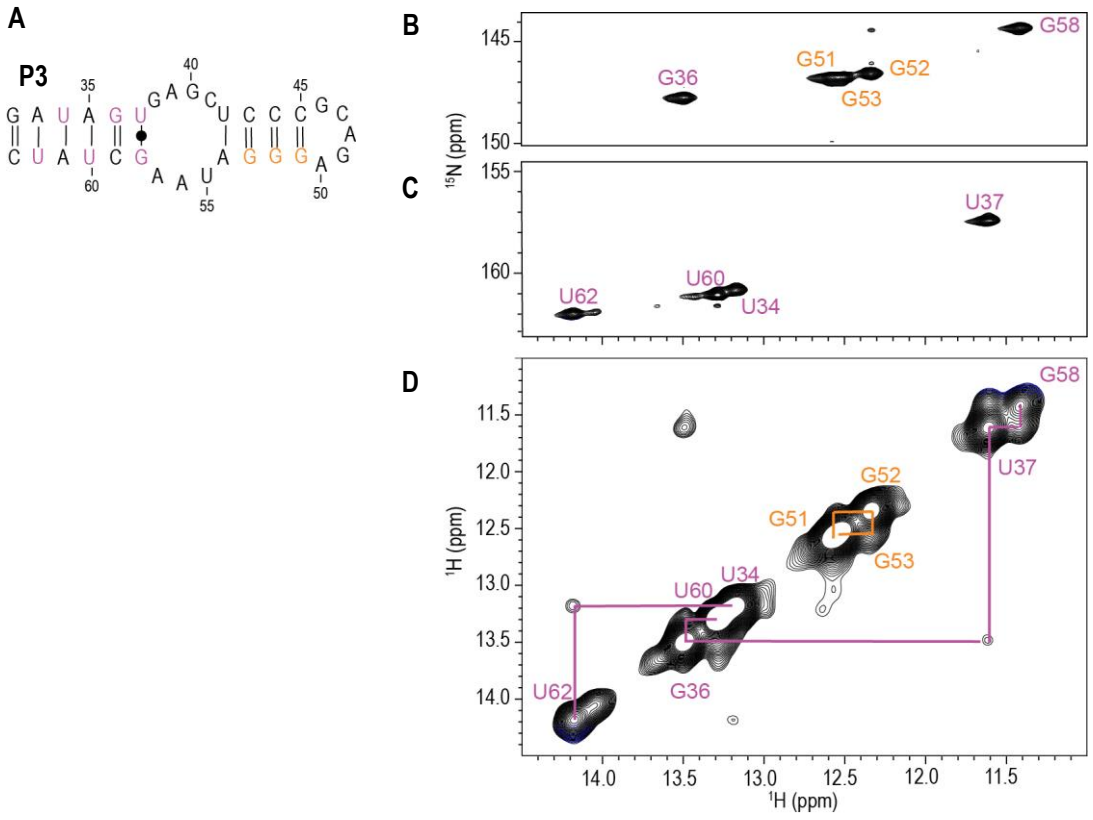


Figure 3.6. Imino assignment in P3
A. Secondary structure of P3. Nucleotides are color-coded corresponding to their assigned resonances on the 2-D ^{1}H-^{15}N HSQC spectrum (B) and 2-D NOESY spectrum (C). **B, C.** Strips of the 2-D ^{1}H $^{-15}$N HSQC spectrum of ^{15}N-labeled P3. **D. 2**-D NOESY spectrum. The number of assigned nucleotides in B, C and D are indicated.

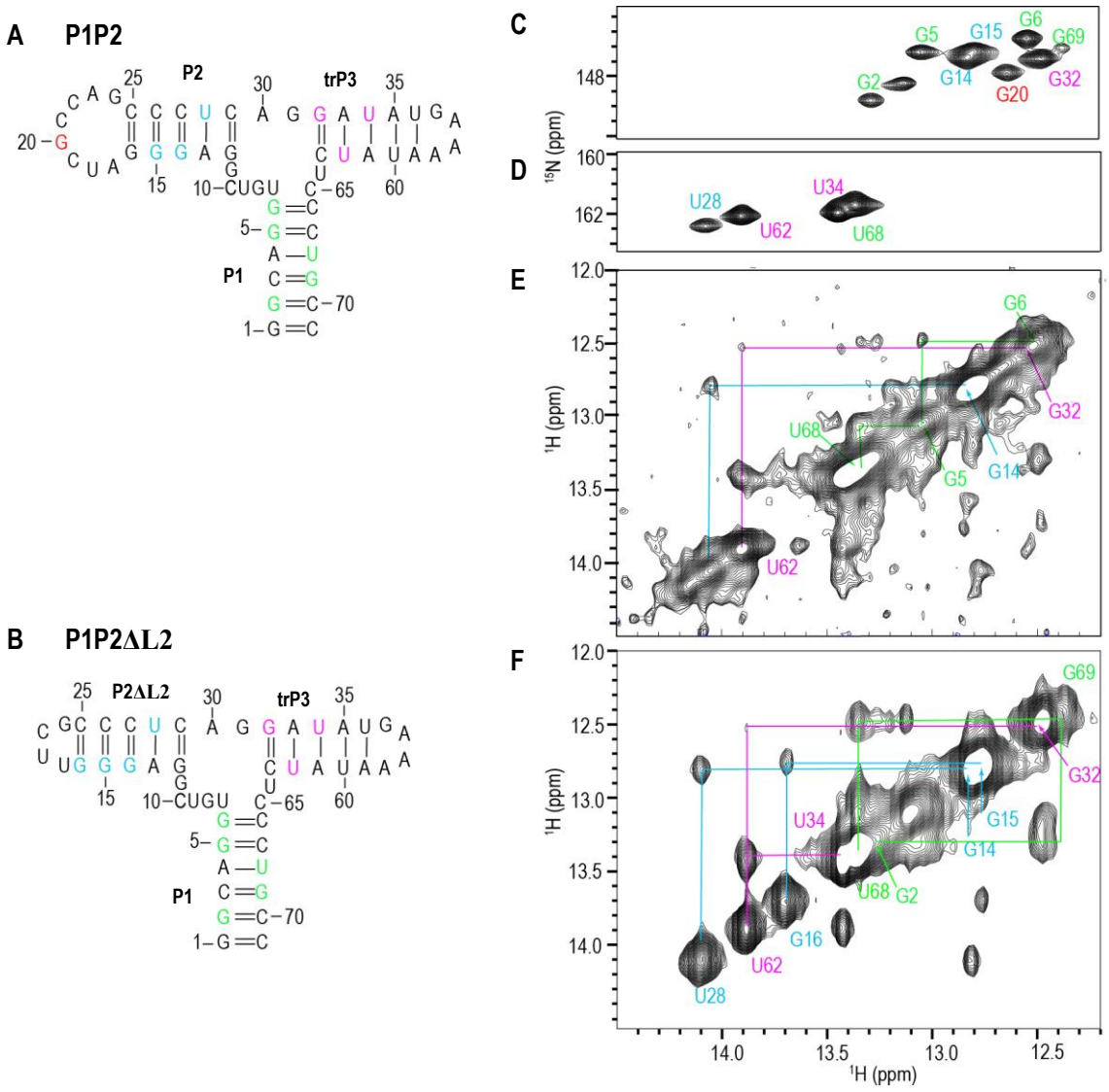


Figure 3.7. Assignment of imino chemical shifts for P1P2 and P1P2ΔL2
A, B. Secondary structures of P1P2 and P1P2ΔL2, respectively. Nucleotides are color-coded corresponding to their assigned resonances on the 2-D ^{1}H-^{15}N HSQC spectrum (C, D) and 2-D NOESY spectrum of P1P2 and P1P2ΔL2 (E, F). **C, D.** Chemical-shift strips of the 2-D ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled P1P2. **E, F**. 2-D NOESY spectra of P1P2 and P1P2ΔL2. Cross-peaks are connected using color-coded solid lines. Number of the assigned nucleotides is indicated. G20 imino resonance (red) was only observed in the spectrum of P1P2 (E) while G16 imino resonance (blue) was only observed in the spectrum of P1P2ΔL2 (F), allowing assignment of chemical shifts of G20(H1) and G16(H1) respectively.

chemical shift of U62(H3) has a strong cross-peak with U34(H3), allowing assignment of the chemical shift of U34(H3) resonance.

Assignment of the apical stem of P3 began with the G52 imino peak, which had

cross-peaks to the chemical shifts of both G51(H1) and G53(H1). The cross-peak resonance between G53(H1) and U42(H3) in the NOESY spectrum of the kl-TSS (Figure 3.8C) allowed for the confident assignment of the chemical shift of G53 on the apical P3 stem. The partial assignment of P3 is presented in Figure 3.6. The assigned chemical shifts for P3 were then used to assign the chemical shifts on the spectrum of the kl-TSS.

Assignments from P1P2ΔL2, in which the 8-nt L2 loop of P1P2 was replaced with a stable UUCG tetraloop, assisted with the imino chemical shift assignment for P1P2 (Figure 3.7A). Similar patterns for the imino cross- and diagonal- peaks of P1P2 and P1P2ΔL2 were observed (Figure 3.7E, F), suggesting that these two RNAs possessed similar pattern of hydrogen-bonding. Interestingly, under similar 2-D NOESY experimental conditions, P1P2ΔL2 produced more cross-peaks with higher signal intensity for the imino protons than those in the NOESY spectrum of P1P2 (Figure 3.7F). These results suggest that the 8-nt L2 loop of the P2 stem might destabilize hydrogen bonds in the stems of P1P2. As such, the cross peaks between the chemical shifts of U28(H3) and G14(H1), and G16(H1) and G15(H1) were only observed in the spectrum of P1P2ΔL2 since the L2 loop was replaced with a UUCG tetraloop (Figure 3.7F). Thus, chemical shift assignment of the P2 stem begins with U28(H3) and G16(H1). The cross-peak connections of U28(H3) and G16(H1) with G14(H1) and G15(H1) respectively, allowed for assignment of 4 out 5 bp of the P2 hairpin (Figure 3.7C, D and F).

High exposure to solvent eliminates the imino signal from the base pair at the termini of an RNA helix. By applying this rule, assignment of the chemical shift of the imino protons of the upper P1 stem initiates with U68(H3) and connects with G5(H1) and G6(H1) on the P1P2 spectrum (Figure 3.7E). The remainder of the P1 stem was assigned

using the NOESY spectrum of P1P2ΔL2 in which the cross peak signal from U68 to G69 is much stronger than the one in the NOESY spectrum of P1P2. A weak cross-peak connecting G69(H1) with G2(H1) allowed for assignment of the chemical shift of G2(H1). G20(H1) was assigned as its chemical shift only appeared in the spectra of the kl-TSS and P1P2 fragment, where the loop of the P2 hairpin is present.

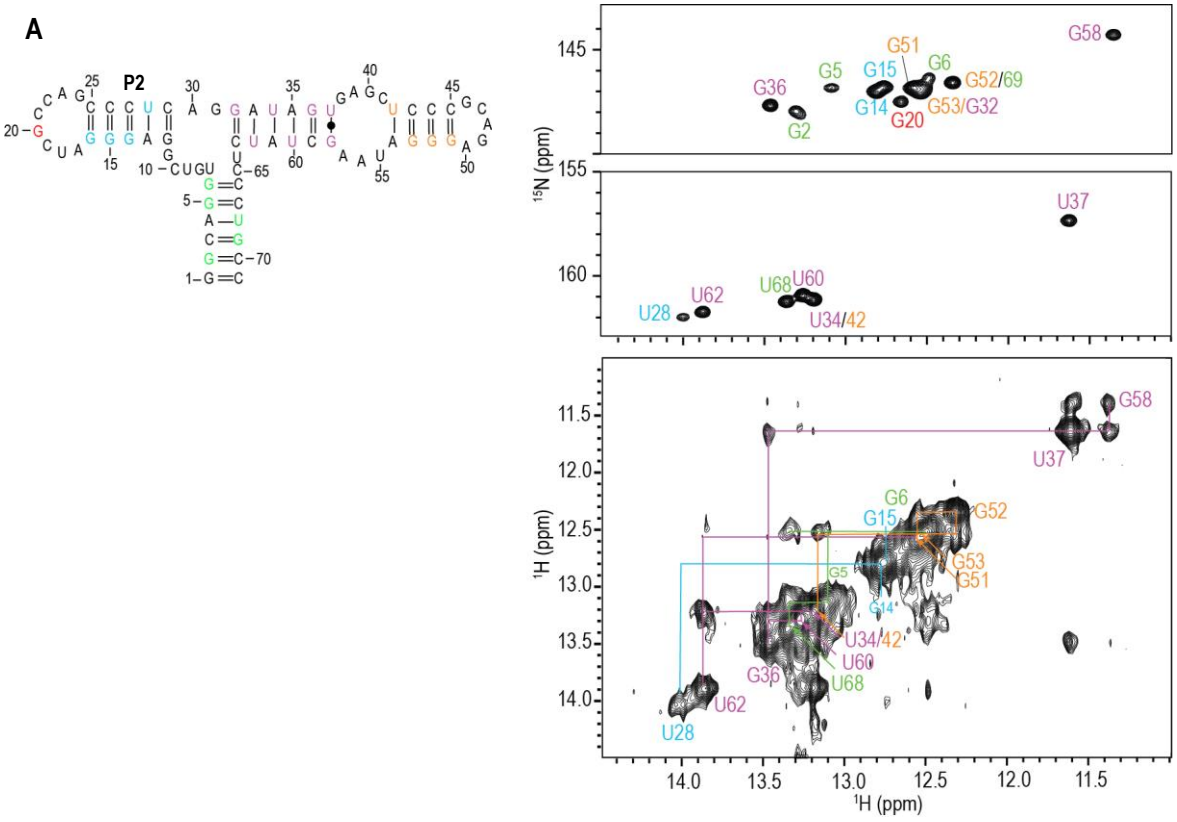


Figure 3.8. Assignment of the imino spectrum for the kl-TSS
A. Secondary structure of the kl-TSS. Nucleotides were labeled corresponding to their assigned imino proton. **B.** 2-D ^{1}H- ^{15}N HSQC of the kl-TSS. **C.** NOESY spectrum of the kl-TSS. The cross-peaks are linked using color-coded solid lines corresponding to specific stem. The assigned peaks are indicated. P1 basal stem: purple, P1 upper stem: orange; P2 stem: blue, P1 stem: green.

The chemical shift assignment for the spectra of P3, P1P2 and P1P2ΔL2 were combined to assign the imino resonance of the kl-TSS (Figure 3.8). Based on the chemical shift assignment of the imino peaks for the kl-TSS, P3 stem residues are not

perturbed by the presence of 5H2, confirming that these residues are not involved in interaction with 5H2 (Figure 3.4C). Conversely, the G15(H1) and G20(H1) peaks disappeared while the G14(H1) peak was slightly perturbed, indicating that the P2 loop residues are responsible for the interaction with 5H2 (Figure 3.4C). In addition, in the presence of the 5H2, three new G imino peaks were observed on the spectrum of the kl-TSS. One of the chemical shift might be the product from formation of a new base-pair and the other two may be the shifted signals from G15(H1) and G20(H1). No new chemical shift of U17(H3) was observed, ruling out the possibility of the U17(H3) imino being involved in the interaction. These results confirmed formation of the hydrogen bonds between 5H2 and the kl-TSS, and that the G residues of the P2 hairpin play important roles in these interactions.

3.3.6. Initial global structural analysis of the wt kl-TSS

Small angle X-ray scattering (SAXS) data was collected to investigate the global structure of the kl-TSS. The scattering signals of the kl-TSS were collected using three different RNA concentrations (SAXS data was collected by Dr. Lin Xin and Dr. Fang, NCI). The signals of these three scattering profiles were zero-extrapolated to obtain an average scattering profile that is independent of RNA concentration. The average scattering intensity of the kl-TSS was normalized with its $I_{0,}$ allowing comparison between different RNA samples. In Figure 3.9B, the averaged scattering profile of the kl-TSS was plotted as the average scattering intensity, $I(q)$, versus momentum transfer, q. The scattering profile of the kl-TSS displayed the typical diffraction feature found in a double helix, with the P1 and P2 peaks occurring at a high-q region (Figure 3.9B) (199), suggesting that the kl-TSS contains helical regions. However,

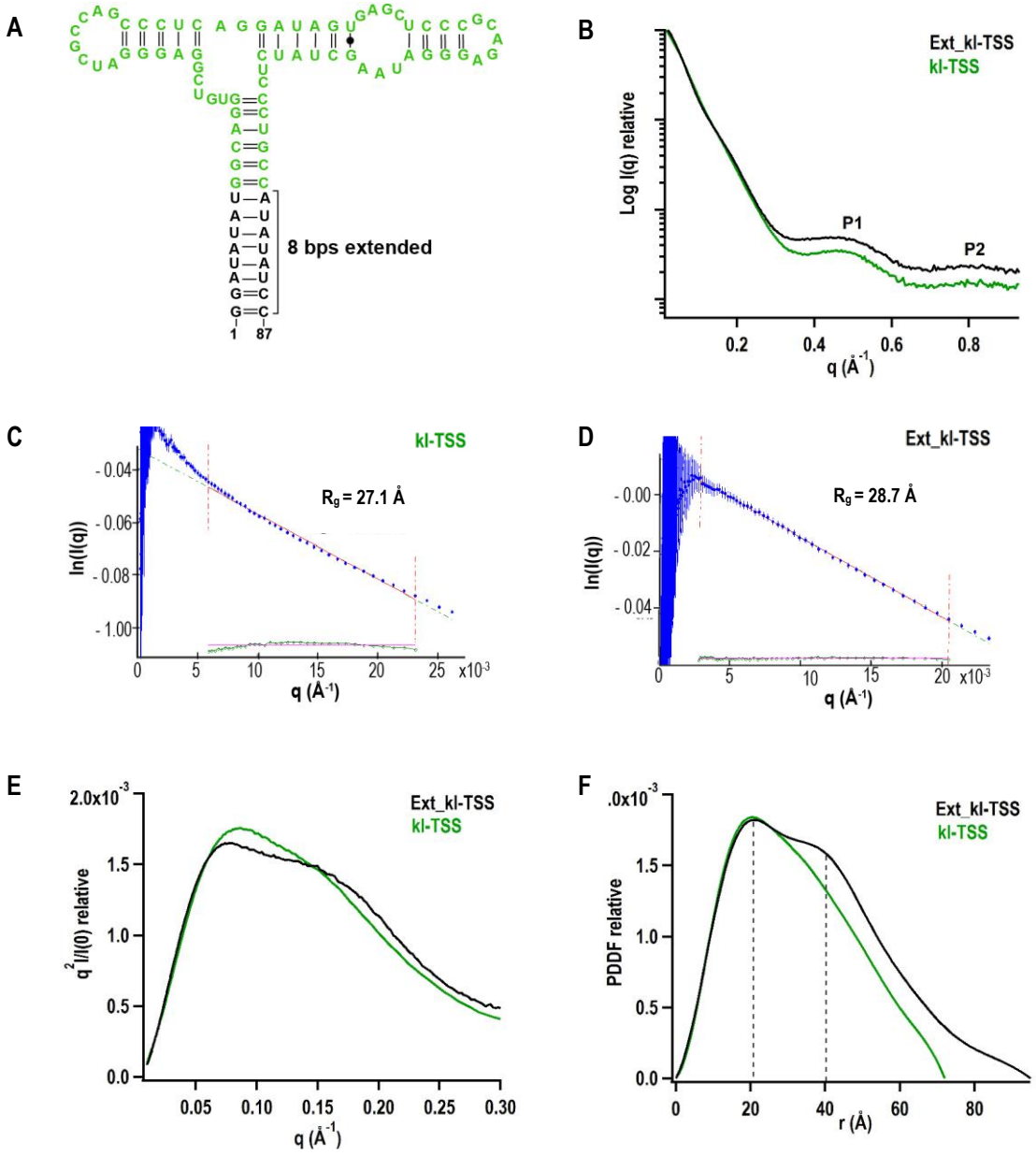


Figure 3.9. SAXS analysis of the wt and extended kl-TSS

A. Secondary structure of wt kl-TSS (green). The 8 bp extended version of the kl-TSS (ext_kl-TSS, black) was used to locate stems within the SAXS envelope. **B.** Ccattering profiles of the wt kl-TSS (green) and ext_kl-TSS (black). Signal intensities were normalized with their I_0s. P1 and P2 peaks feature the scattering profile of electrons within the major and minor groves of the two helical strands. **C**. Guinier plots (blue) for cattering curves of wt kl-TSS and **D.** ext_ kl-TSS. The Guinier region (between the two red lines) was linear-fitted (green dashed line) to obtain R_g. Errors of the fitting is presented in the green dotted curve at the bottom **E.** Kraky plots for the ext_kl-TSS (black) and wt kl-TSS (green). **F.** Pair distance distribution function plots (PDDFs) of the ext_kl-TSS (black) and wt kl-TSS (green) indicate helical characteristics of the two RNAs with major distance population of 20 Å. The 40-Å peak of the ext_kl-TSS is indicated.

the peak intensities of P1 and P2 were much lower than those found in a simple duplex (199). This may be due to the presence of non-duplex regions within the structure or because of the dynamic conformation of the RNA. The Guiner region for the kl-TSS scattering curve is linear, indicating that the RNA is monodispersed and homogeneous in solution (Figure 3.9C) (200). Using the Guinier approximation ($\ln[I(q)] \sim \ln[I(0)] - R_g^2 q^2/3$) within the low range of q values ($qR_g < 1.3$), the obtained radius of gyration (R_g) for the kl-TSS was 27.1 Å. In Figure 3.9E, the Kratky plot for the kl-TSS has one prominent peak with raised intensity at a higher q range, suggesting that this RNA is extended and partially folded (157). The D_{max} of the kl-TSS obtained from the pair distance distribution plot (PDDF) was 72 Å. In addition, the PDDF plot for the kl-TSS revealed that the most populated distance of the kl-TSS was 20 Å (Figure 3.9F), which is close to the diameter (25 Å) of an A-form RNA duplex.

3.3.7. Hairpin assignment and low resolution *ab initio* modeling

To identify the positions of the individual stem-loops within the SAXS envelope of the wt kl-TSS RNA, its P1 stem was extended by 8 bp. The extended kl-TSS (ext_kl-TSS) was expected to have a longer P1 stem, which would allow differentiating between the P1 and P2 stems within the SAXS envelope (Figure 3.9A). The signal intensity of the scattering profiles for the ext_kl-TSS was normalized using its I_0 value. The signal intensities of P1 and P2 peaks for the ext_kl-TSS are slightly higher than those of the kl-TSS (Figure 3.9B). This could be due to the increased number of base-pairs in this extended RNA. The R_g of the ext_kl-TSS is 28.7 Å (Figure 3.9D), which is slightly longer than the R_g of wt kl-TSS (27.1 Å). The Kraky profile of the ext_kl-TSS exhibits two clear peaks whereas there is only one peak for the wt kl-TSS (Figure 3.9E). This

A

Recommendation	NSD	Model no.
Include	0.651	12
Include	0.657	18
Include	0.66	3
Include	0.665	7
Include	0.666	11
Include	0.671	19
Include	0.673	9
Include	0.674	4
Include	0.677	10
Include	0.68	5
Include	0.681	2
Include	0.682	17
Include	0.684	11
Include	0.685	18
Include	0.69	8
Include	0.693	20
Include	0.693	6
Include	0.695	16
Discard	0.74	1
Discard	0.746	15
	Avg. =	0.683
	Std . =	0.024

B

Recommendation	NSD	Model no.
Include	0.728	14
Include	0.732	5
Include	0.739	9
Include	0.745	4
Include	0.75	1
Include	0.753	8
Include	0.756	12
Include	0.756	3
Include	0.758	2
Include	0.761	19
Include	0.764	10
Include	0.786	17
Include	0.79	6
Include	0.793	13
Include	0.797	20
Include	0.799	7
Include	0.799	15
Include	0.802	18
Include	0.811	16
Include	0.824	11
	Avg. =	0.772
	Std. =	0.028

Table 3.1. Normalized spatial discrepancy (NSD) of the kl-TSS (A) and the ext_kl-TSS (B)
The NSD of the 20 conformations of each RNA was quantified using DAMSEL of the ATSAT package (REF). Avg: Average; Std: standard deviation. If the NSD value of the model is larger than mean + 2* std., the model is discarded. A value of 0 indicates that the two models are perfectly superimposed and a value of 1 indicates that the two models are completely different.

indicates that the mass of the ext_kl-TSS is distributed through a larger area than that of the wt kl-TSS. When comparing the PDDFs of the two RNAs (Figure 3.9F), the graphs clearly show that the ext_kl-TSS has an additional populated distance (40 Å) that is about two-fold greater than the diameter of an A-form helix, indicating that the extended P1 might stack on one of the other two helices.

To visualize the 3-D shapes of the ext_ kl-TSS and wt kl-TSS, the PDDF of the two RNAs were submitted to the program DAMMIN (201) to reconstruct 3-D models in 20 independent runs. The resulting 20 models for each RNA were then pairwise-scored using the program DAMSEL (201) to obtain the normalized spatial discrepancy (NSD) which is a quantitative measure of similarity between sets of 3-D point ranging from 0-1 (0 meaning two models superimposed ideally and 1 meaning the two models are completely different). The average NSD of ext_kl-TSS and wt kl-TSS were 0.72 and 0.68, respectively, indicating that there was a moderate convergence of their individual models with the overall ensembles of the two RNAs (Table 3.1A, B). The two outliner models of the kl-TSS were removed and the remaining models were then averaged with the DAMAVER/DAMFILT program to obtain the averaged/filtered envelope of the RNA. By definition, the averaged envelope contains the space of all 20 models while the filtered envelope only contains the common space shared by the 20 models. The filtered envelope of the wt kl-TSS shows that the kl-TSS has an "N" shape with the thickness part of the envelope ranging from 15 Å to 20 Å (Figure 3.10A). The filtered envelope of the ext_kl-TSS was obtained using a similar process of data analysis as mentioned above (Figure 3.10B). Superimposing the filtered models of the ext_kl-TSS and wt kl-TSS offers a suggestion for the position of the individual stem-loops of the RNA.

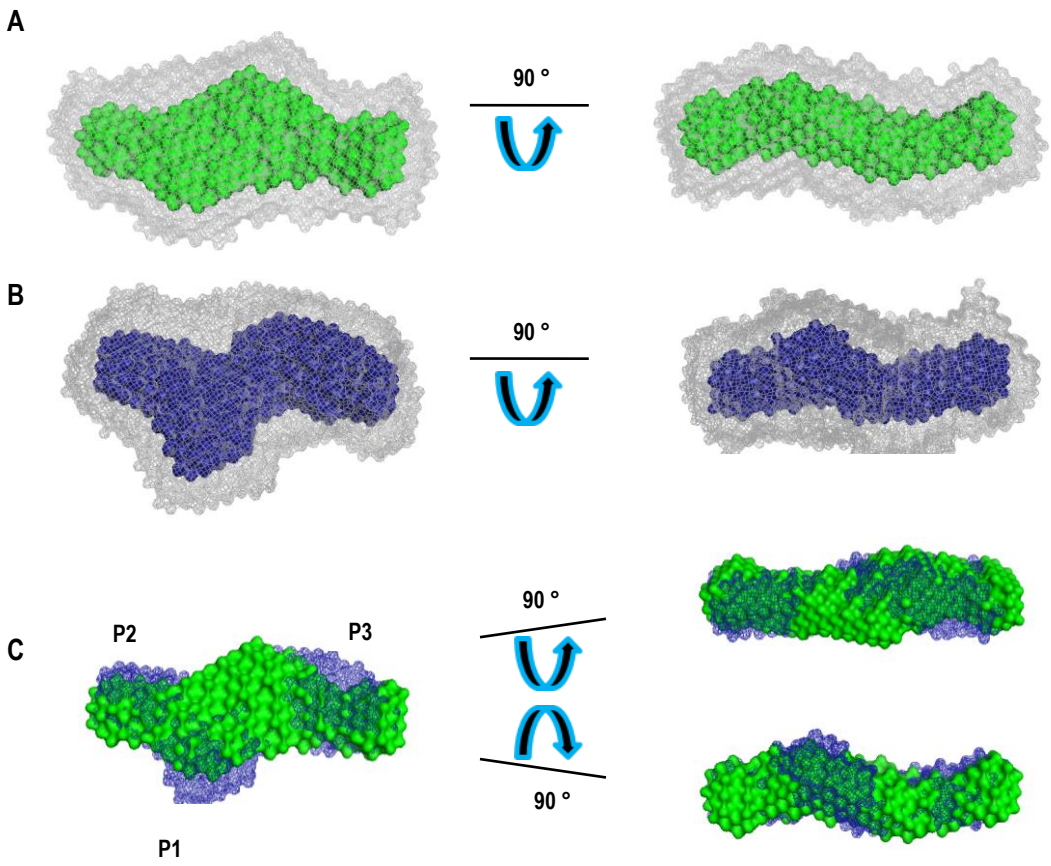


Figure 3.10. Molecular envelopes of the ext_kl-TSS and the wt kl-TSS
A. Superimposition of the averaged (grey) and filtered envelops (green) of the wt kl-TSS and 90° side view. Side view shows the "N" shape of the wt kl-TSS. Envelopes are presented in surface mode. **B.** Superimposition of the averaged (dark grey) and filtered envelopes (blue) of the ext_kl-TSS and 90° side view. **C.** Superposition of the filtered envelopes of the wt kl-TSS (green) and the ext_kl-TSS (blue). The two envelopes were aligned using the sumpcom program of ATSAS. Location of P1, P2 and P3 stem are indicated.

Figure 3.10C shows a protrusion region for the ext_kl-TSS not found in the wt kl-TSS, while the remaining parts of the two constructs superimposed well, indicating that this protrusion region is a part of the P1 stem. Difference in lengths between the P2 and P3 stems made it possible to assign the positions for these two hairpins within the envelope (Figure 3.10C). These assignment show that the P1 and P2 stem loops are positioned in close proximity with one another while the P3 stem loop protrudes away from the central hub. The P2 and P3 stems stretch across the molecule and serve as the

two main contributors to the D_{max} of the wt kl-TSS. This explains the slightly increased R_g of the ext_kl-TSS when its P1 stem loop was extended by 8 bp.

3.3.8. 3-D structural model of the kl-TSS guided by its SAXS scattering profile

Conformational ensemble analysis was performed to obtain the 3-D model of the kl-TSS. 1650 conformational models of the kl-TSS were generated using MC-Sym and MC-Fold (194). The scattering profiles of these 3-D models were back calculated using fast X-ray scatting (FoXS) (195). Similarity between the back-calculated scattering profiles and experimental scattering profile was scored using the χ value (Materials and Methods). Figure 3.11A present the overlay of the best ten back-calculated scattering profiles in which their average χ value is 3.46±0.25. The SAXS envelope and these ten 3-D models of the kl-TSS were aligned using the Sumpcom program (201). Figure 3.10B present the overlay of the kl-TSS envelope with the three best-fit models, supporting the localization of the hairpins within the kl-TSS. However, when these 3-D models were compared with the 3-D structure of Phe-tRNA (Figure 3.11C), which was used to superimposed with the computational model of the kl-TSS (77), the new 3-D models of the kl-TSS were more extended. Further analysis of the conformational ensembles of the kl-TSS using SES (196) indicated that the two state model gave moderate improvement in the χ value (analyzed by Andrew Longhini). The lowest χ value obtained from the single state model was 2.94 while the χ value for the two state model was 2.57. However, the χ value for the three state model did not improve (χ = 2.54). In addition, to further confirm the result of the analysis, a bootstrapping technique was applied. In this technique, synthetic data sets of experimental and back-calculated scattering data were created by randomly drawing out and adding back the original data sets. The repeated

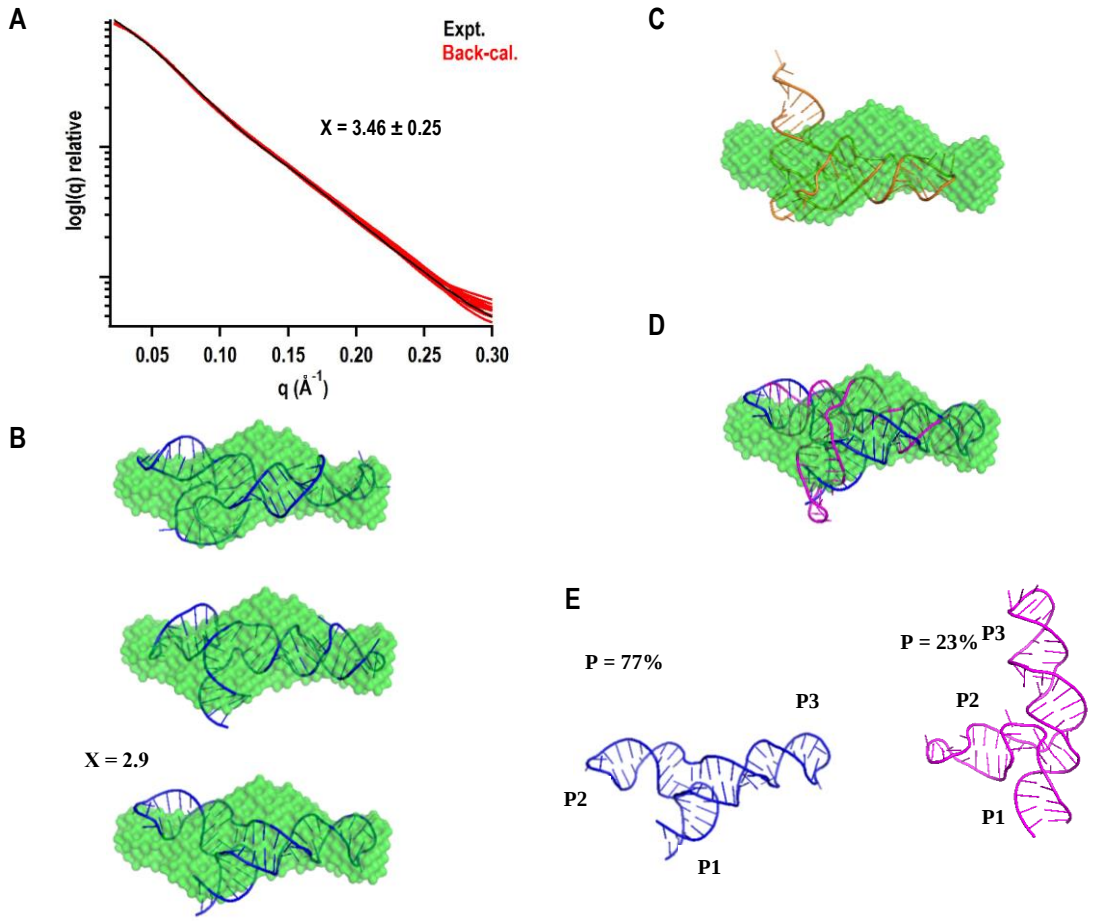


Figure 3.11. Ensemble analysis of 3-D models of the kl-TSS
A. Back-calculated scattering profiles of the 3-D models of the kl-TSS (red lines). The back-calculated scattering profiles were obtained from FoXS (webserver) (195). The black line is the experimental scattering profile of the kl-TSS obtained from SAXS. **B.** Superimposition of three best fit models of the kl-TSS in the SAXS envelope. The envelope and 3-D model of the kl-TSS were superimposed using the sumpcom program of the ATSAS package (201). The surface envelope is presented in green and the 3-D models are in blue with base-pairs in blue lines. The best fit models have a χ value of 2.9. **C.** Superimposition of Phe-tRNA (organge) and the SAXS envelope of the kl-TSS. **D.** Superimposition of the two state models within the SAXS envelope of the kl-TSS. **E.** Two conformation states of the kl-TSS. % populated values are shown.

process was performed until the synthetic data sets had similar total data point as the original data sets. Interestingly, after multiple rounds of testing, the same two initial models were obtained (Figure 3.11D, E). Thus, it is likely that the kl-TSS has two conformational states which are presented in Figure 3.11 D, E. Of these two kl-TSS conformations, one is represented by 77 percent of the population while the other

conformation is represented by 23 percent of the population (Figure 3.11 E). Interestingly, when comparing the two state models with the previously predicted model of the kl-TSS, the previously predicted model of the kl-TSS has similar topology with the low populated state conformation while the major conformation has a significantly different topology.

3.4. Discussion

The discovery of 3'CITEs in plant viruses has increased the diversity of translation mechanisms. Different from IRESs, these robust 3'CITEs can enhance translation of viral proteins when canonical translation is not shut off. Many new 3' CITEs have been recently discovered and are currently divided into 7 classes based on their secondary structure (5). However, the limited knowledge of 3-D structures for 3´ CITEs has limited our understanding of the mechanisms of these translational elements. Therefore, in this Chapter, I have reported the structural characterization of the kl-TSS, representative 3' CITE of PEMV using SAXS and NMR.

From my study, the structure of the kl-TSS, unlike the TCV TSS, is independent of the adjacent 5' and 3' sequences. Removal of these adjacent nucleotides from the core kl-TSS sequence did not affect base-pairing of the kl-TSS (Figure 3.3D). The number of imino resonance peaks in the 2-D imino HSQC spectrum of the kl-TSS is representative of the number of formed hydrogen bonds within the structure (Figure 3.3D). The detection of NH···N cross peaks between A(N1) and U(N3/H3) and G(N1/H1) and C(N3) in 2D *HNN* COSY spectrum of the 81-nt kl_TSS (Figure 3.3C) directly showed covalent bond features of the hydrogen bonds within this RNA. The number of imino resonance peaks in the 2-D imino NMR spectrum of the kl-TSS (Figure 3.3) is close to

the number of the base-pairs predicted by Mfold (77, 95), indicating that the two structures have the same base-pair pattern.

Due to signal overlaps (Figure 3.3), resonances of the imino protons in the full-length kl-TSS were assigned by using the "divide and conquer" approach. All of the imino resonances were observed in the spectrum of the ^{15}N-labeled kl-TSS fragments without any significant chemical shift changes, solving the resonance overlap problem and improving the confidence in assignment of the imino protons in the full-length kl-TSS (Figure 3.5-3.8). Probable resonance assignment in the lower stem of the P3 fragment was initiated with a key resonance signature of G58:U37 wobble base-pair (Figure 3.6), in which both bases had cross-peaks with the imino proton of G36. The terminal G32 imino proton in the P3 fragment was highly susceptible to the solvent and thus was only assigned within the spectra of the P1P2 and P1P2ΔL2 fragments due to appearance of its cross peak with U62(H3) (Figure 3.7E, F). Resonance assignment for the imino protons in the P1 and P2 stems resulted from the replacement of the L2 loop in P1P2ΔL2 with the stable tetra loop UUCG. The cross- and diagonal-peaks of the imino protons in this fragment had much higher signal intensity and were well-resolved as compared with those in the P1P2 fragment (Figure 3.7E, F), allowing the assignment all of its observed imino protons (Figure 3.7D). The appearance of resonance for G20(H1) in the imino spectrum of the P1P2 fragment but not in the spectrum of the P1P2ΔL2 indicated that G20 likely forms a hydrogen bond with one of the C nucleotides within the L2 loop (Figure 3.7B). The assignment of the observable imino protons in the full-length kl-TSS was completed by combining the imino proton assignments for P3, P1P2 and P1P2ΔL2 fragments (Figure 3.8D). In the spectrum of the full-length kl-TSS, 13 out of

15 imino resonances for G nucleotides and all imino resonances for U nucleotides were assigned (Figure 3.8).

Previous studies have shown that interaction between the kl-TSS and the 5' 89-nt fragment is important for efficient translation (51, 77). This interaction was determined to only require 5H2 through retarded mobility of ^{32}P-labeled kl-TSS with associated truncated fragments of 5H2 using EMSA (Figure 3.4). The hydrogen bonds between 5H2 and the kl-TSS were further confirmed by the presence of a new G imino resonance peak and a few shifted imino peaks for the P2 stem in the imino spectrum of the kl-TSS (Figure 3.4C). The un-disturbed resonances of the imino protons in the the P1 and P3 stems in the presence of 5H2 (Figure 3.4 and 3.8) provide strong evidence that these two stems are not involved the interaction, which is consistent with the previous study (77).

The bootstrapping technique was used for the first time to analyze and score the number of 3-D models of the kl-TSS that can be fitted within the SAXS molecular envelop. Interestingly, after multiple rounds of testing, the same two initial models were obtained, increasing the confidence for this model fitting. Based on the fitting results, the kl-TSS is proposed to have two conformational states when using its SAXS scattering profile to analyze the conformational ensembles of the kl-TSS (Figure 3.11). The low-populated conformation of the kl-TSS has similar topology with the previously predicted structure for the kl-TSS. However, the P3 stem of the low-populated conformation is slightly longer than the one in the previously predicted structure (Figure 3.11). The highly-populated conformation of the kl-TSS differs from the tRNA-shaped structure of the TCV TSS. This result supports the previous observation that the kl-TSS binds ribosomes with a slightly higher Kd (77) and at a different site than the P-site like in the

TCV TSS (51). However, the remaining question of why the kl-TSS adopts two conformations and which of these two conformations preferentially binds ribosomes will require further investigation.

Chapter 4: *In vivo*, site-specific labeling of homogeneous, recombinant RNA in wt and mutant *E. coli* for NMR structural studies

4.1. Introduction

Among the various biophysical tools developed to solve RNA structures, such as high resolution X-ray crystallography and low resolution small-angle X-ray scattering (SAXS), only nuclear magnetic resonance (NMR) spectroscopy can probe sparsely (<5%) and highly (>90%) populated RNA structures in solution at atomic resolution (202-209). In chapter 3, I have presented the characterization of structure of the PEMV kl-TSS using NMR. However, to solve large RNA structures effectively using NMR, at least two inherent limitations of NMR must be circumvented: signal overlap and rapid signal decay. For example, extremely poor chemical shift dispersion of the ribose atoms and the large ^{13}C-^{13}C multiplets that arise from uniformly enriched $^{13}C/^{15}N$-labeled RNAs decrease the overall sensitivity and resolution of NMR experiments (150).

Four approaches have been proposed previously to address these limitations: total solid-phase chemical synthesis of RNA, and *de novo*, biomass, or selective-biomass biosynthesis of NTPs (152, 210-223). Other than total chemical synthesis, which is limited to the production of ≤ 40-nt RNAs, these other methods require *in vitro* T7 bacteriophage polymerase-based transcription that is subject to addition of nontemplated nucleotides at the 5′- and 3′-termini of the transcribed RNA (224, 225).

An alternative, cost-effective approach is to insert the RNA of interest in the anticodon loop of a tRNA-scaffold, express the chimeric construct in *E.coli* and then cleave off the recombinant RNA after purification (226-228). Due to its location within the tRNA scaffold, the RNA of interest is protected from degradation by cellular nucleases. Alternative approaches applied to limit the degradation of the recombinant

RNA include replacement of the tRNA-scaffold with the 5S ribosomal RNA (rRNA) (229) or an induced T7 RNA promoter with a constitutive lipoprotein promoter (230). After purification from *E.coli*, the RNA of interest is cleaved from the chosen scaffold using RNAse H, hammerhead ribozymes, or RNA-cleaving deoxyribozymes (DNAzymes). However, these methods have thus far only been used to synthesize unlabeled and ^{15}N and/or ^{13}C uniformly labeled RNAs (228, 230).

In this report, I combine the tRNA-scaffold approach with wt and mutant *E.coli* strains whose metabolic pathways have been modified to produce various site-specific labeling patterns of rNTPs when grown in specifically labeled carbon sources (211, 214). This combined approach allows for the rapid and uniform labeling of RNAs with ^{15}N isotopes and site-specific labeling with ^{13}C isotopes. Using wt K-12 *E. coli* grown in [1-^{13}C]-acetate, synthesized RNAs are specifically labeled with ^{13}C isotopes at the C2 and C4 carbons of the pyrimidine nucleobases, at the C4 and C6 carbons of the purine nucleobases and at the C3′ ribose carbon atoms (211). When using mutant transketolase *E.coli* strain (tktA), which does not produce the pentose phosphate shunt enzyme transketolase, growth in [1-^{13}C]-glucose generates RNAs labeled with ^{13}C isotopes at the C5 and C6 carbons of the pyrimidine nucleobases, the C2 and C8 carbons of the purine nucleobases and the C1′ and C5′ ribose carbons (214, 231, 232).

I demonstrate the general applicability of this method by successfully expressing and site-specifically labeling four RNAs with different complexities and sizes ranging from 48 nt to 118 nt (Figure 4.1), using 2-D HSQC and 3-D HNCO (146). I also present the first example of an NMR experiment that correlates the imino proton with carbonyl carbon atoms in large RNAs (>50nt) to provide local structural information about the

nucleobases in the SAM-II riboswitch aptamer. Compared to uniform labeling, our site-selective labeling approach provides clear advantages of improved resolution and sensitivity.

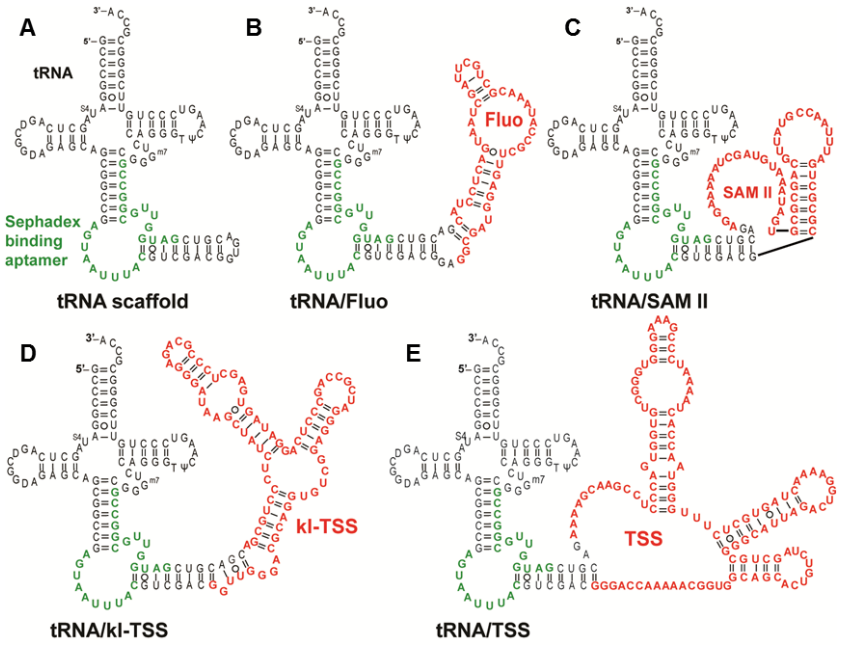


Figure 4.1. Sequence and predicted structures of the recombinant tRNA-scaffolds
A. The construct for the human $tRNA_{lys}$-scaffold, which includes the Sephadex aptamer (green) for affinity purification and restriction sites necessary for subcloning RNAs of interest. The following four recombinant RNAs were expressed to demonstrate the utility of our labelling method in RNAs of varying lengths: **B.** the *B. anthracis* fluoride-binding riboswitch aptamer (tRNA/Fluo); **C.** S-adenosylmethionine (SAM)-II riboswitch (tRNA/SAM); **D.** the PEMV tRNA/kl-TSS; and **E**. the TCV tRNA/TSS. Inserted RNAs are highlighted in red.

4.2. Materials and Methods

4.2.1. Vector construction

Vectors were constructed as previously described (228). The plasmid pBSKrnaSeph containing a sephadex aptamer at the anticodon stem of human lysine lysine tRNA ($tRNA_{lys}$) was kindly provided by Dr. Frédéric Dardel. The sequences of computationally-predicted *Bacillus anthracis* fluoride-binding riboswitch aptamer (Fluo) (48 nt) (233, 234), the S-adenosylmethionine (SAM) metabolite binding (SAM-II)

riboswitch aptamer (54 nt) from the Sargasso Sea metagenome (117), the PEMV kl-TSS (81 nt) (51, 77) and the TSS (118 nt) (74, 76) were subcloned into the plasmid at *Aat II* and *Sal I* sites (Figure 4.1). The recombinant plasmids were confirmed by sequencing. Pilot expression of each vector was carried out using RNA minipreps as previously described (228).

4.2.2. Bacterial strains

The wild type *E. coli* strain K-12 NCM 3722 (CGSC #4401:*F*+) from the Coli Genetic Stock Center (CGSC) was used for cloning, optimization, and site-specific labeling of RNAs. The mutant *E. coli* strain tktA (CGSC # 11606, *F-Δ(araD-araB)567, Δ lacZ4787(::rrnB-3), λ⁻,ΔtktA783::kan, rph-1,Δ(rhaD-rhaB)568, hsdR514*) was also used for site-specifically labeling of RNAs (214, 235). These bacterial strains were made chemically competent using $CaCl_2$.

4.2.3. *E.coli* growth media

Media used in this project were Luria-Bertani-Miller (LB), 2x TY, Studier phosphate buffer (SPG), and LeMaster-Richards (LMR) media. These were prepared as described previously (215, 228, 236-238). To obtain ^{15}N labeled RNAs, media was supplemented with 25 mM $(^{15}NH_4)_2SO_4$ and 0.4 % weight/volume (w/v) glucose. To obtain ^{13}C site-specifically labeled RNAs, media was supplemented with 0.3 % (w/v) sodium [1-^{13}C]-acetate for K-12 (211) or with 0.2 % (w/v) [1-^{13}C]-glucose for tktA (214). Cultures were also supplemented with 100 µg/ml amp and minimal media were supplemented with 0.25x BME vitamins 100x solution (Sigma). Media for growing tktA was also supplemented with 25 µg/ml shikimic acid and 30 µg/ml kanamycin (214).

4.2.4. Buffer selection for minimal media with a limited carbon source

The recombinant tRNA/Fluo construct was used to test the growth rate and the tRNA-scaffold expression of K-12 grown in two commonly used minimal media buffers, SPG-amp and LMR-amp, supplemented with 0.2 % (w/v) glucose. A single *E.coli* colony was used to inoculate 10 ml of LB-amp and incubated at 37°C and 300 RPM until the cell density measured by the absorbance at 600 nm (OD_{600}) reached approximately 0.5 (215). Cells were pelleted by centrifugation, resuspended, and used to inoculate 100 ml of SPG-amp medium supplemented with 0.4 % (w/v) glucose. At an OD_{600} of 0.5, cells were pelleted, resuspended, divided into equal portions, and used to inoculate 1 L of SGP-amp and 1 L of LMR-amp media supplemented with 0.2 % (w/v) glucose. To monitor growth, the OD_{600} of each culture was measured hourly, and cells from 10 ml of culture were pelleted hourly. To analyze recombinant tRNA-scaffold expression on denaturing PAGE, each pellet was resuspended in 500 µl of 10 mM Tris-HCl (pH 7.4) and 10 mM $MgCl_2$ (lysis buffer). Total cellular RNA was extracted by addition of an equal volume of acid phenol:chloroform (pH 4.5) (Life Technologies) with agitation at 4°C. Expression of the recombinant tRNA-scaffold in each sample was analyzed on 10 % (w/v) denaturing PAGE. The data were collected in triplicate.

4.2.5. RNA production in *E.coli*

I adapted a previously described double selection process (239) and large-scale growth protocol (214, 215, 228) to optimize the expression of the recombinant tRNA-scaffold.

Double selection process

Transformed mutant *E. coli* cells were plated on SPG-amp agar containing the desired carbon source for optimization. These plates were incubated at 37°C until colonies became visible. Five colonies were randomly selected and grown in 1 ml of LB-amp. After 3 h, cells were pelleted, resuspended, and transferred to 10 ml of SPG-amp supplemented with the desired carbon source. The recombinant tRNA-scaffold is expressed during the log phase of *E. coli* growth (240), so 10 µl of each culture was spread on a new SPG-amp agar plate when they reached an OD_{600} of 0.5, again supplemented with the desired carbon source and incubated at 37°C. At an OD_{600} of 1.0, the total RNA of each culture was extracted using the lysis buffer and acid phenol:chloroform as described previously. The expression levels of the recombinant tRNA-scaffold of each colony were compared using analytical denaturing PAGE. The selection process was repeated using five progeny colonies originating from the plate of the colony with best expression of the recombinant tRNA-scaffold. In addition, at an OD_{600} of 0.5, 500 µl of each culture was added to 500 µl of 30 % volume/volume (v/v) glycerol and stored at -80°C as a glycerol stock.

Large-scale growth and recombinant tRNA-scaffold expression

Five microliters of the double selected colony glycerol stock was used to inoculate 10 ml of LB-amp medium, and this culture was incubated at 37°C, 300 RPM until the OD_{600} reached approximately 0.5 (~ 4 h). Cells from this culture were pelleted, resuspended, and used to inoculate 100 ml of SPG-amp supplemented with 0.4 % (w/v) unlabeled glucose to quickly enrich cell density in a shorter period of time. The initial OD_{600} of the100-ml culture was 0.05 to achieve a consistent growth rate (215). Depending on the *E. coli* strain, cells took 4-8 h to reach an OD_{600} of 0.5. Cells were then

pelleted and resuspended in 1 L of SPG-amp supplemented with the appropriate isotopically labeled carbon source. Cells were grown until the maximum OD_{600} was reached (12-18 h). When cultivating K-12 in acetate, cells were grown in separate 500 ml batch cultures in 4-L flasks to maintain a sufficient aeration rate (239, 241). Cultures were pelleted and resuspended in 2.5 ml lysis buffer/g of pellet to allow for sufficient suspension of cells. Total cellular RNA was partitioned from chromosomal DNA and proteins using an equal volume of acid phenol:chloroform (pH 4.5). Total RNA in the aqueous phase was precipitated using 0.1 volumes 5 M NaCl and 3 volumes 100 % ethanol. The RNA was pelleted using high-speed centrifugation, air-dried, and dissolved in distilled water. The RNA solution was further centrifuged at high speeds to remove any remaining fine protein debris. The supernatant, containing total cellular RNA, was then purified by anion-exchange or affinity chromatography

4.2.6. RNA purification using anion-exchange chromatography and affinity-size exclusion chromatography

The anion-exchange method was modeled after the protocol developed by Nelissen *et. al.* (230) with several modifications to optimize the purification of the recombinant tRNA-scaffold with less contamination from smaller cellular RNAs. After equilibration with purification buffer (40 mM potassium phosphate buffer, pH 7), the in-house pre-packed 50-ml Source 15Q anion exchange column was loaded with the total cellular RNA from the 1 L culture using a 50-ml superloop (GE Healthcare). RNAs with different charge densities were separated using a segmented and linear gradient from 0.44 M to 1M NaCl in purification buffer. The extended washing step consisted of 5 CV of 0.44 M NaCl in purification buffer followed by 5 CV of a gradient from 0.44 M to 0.52 M NaCl

in purification buffer allowed for the removal of smaller cellular RNAs. The recombinant tRNA-scaffold was then eluted with 5 CV of 0.52 M NaCl in purification buffer. rRNAs were removed from the column with 2 CV of a linear gradient from 0.52 M to 1M NaCl followed by 2 CV of 1 M NaCl in purification buffer (Figure 4.6). Fractions containing the purified recombinant tRNA-scaffold were confirmed by 10 % (w/v) denaturing PAGE. Positive fractions were combined using a 3 KDa molecular weight cutoff (MWCO) spin column (Millipore). The RNA was then solvent-exchanged into distilled water or appropriate buffer for either cleavage with DNAzymes or NMR experiments.

Due to its capability of binding to the sephadex tag in the tRNA-scaffold, the size exclusion HiLoad 26/600 Superdex 75 pg column (GE healthcare) was utilized for an automated affinity purification to replace manual purification using Sephadex G-100/G-200 beads (226, 230). Total cellular RNA from 1 L of culture was loaded onto the Superdex column. The unbound RNA was removed with 1.5 CV purification buffer containing 100 mM NaCl. The recombinant tRNA-scaffold was eluted with purification buffer containing 4 M urea. Fractions containing the recombinant tRNA-scaffold were combined and solvent-exchanged as described above (Figure 4.7).

4.2.7. 8-17 DNAzyme cleavage

Two 8-17 DNAzymes were designed to excise each inserted RNA of interest from the recombinant tRNA-scaffold (242). 8-17 DNAzymes contain a catalytic core sequence that cleaves the RNA substrate at a dinucleotide, flanked by 12-nt “handle” sequences on either side that anneal to the RNA substrate. Optimal DNAzyme core sequences were selected according to Schlosser et al., 2008 (60). The two core sequences of the DNAzymes used were 5′-TGTCAGCGACTCGAA-3′ and 5′-GATAGCATTCCCGAG-3′

to cleave the dinucleotide sites 5′-GG-3′ and 5′-GA-3′, respectively. List of DNAzymes used to cleave the recombinant tRNA scaffolds were listed in Table 4.1

RNA insert	5' DNAzyme	3' DNAzyme
Bacillus anthracis fluoride-binding riboswitch aptamer	5'-AACTCCATCGCC**GATAGCATTCCCGA G**CGTCGACGTAAA-3'	5'-CAACCATCGACG**GATAGCATTCCC GAG**GTAGGAGTCATT-3'
SAM-II riboswitch aptamer	5'-AATCAGCGCGCG**GATAGCATTCCCG AG**GACGTAAATTAC-3'	5'-AACCATCGACGT**TGTCAGCGACTC GAA**CCTTTTTAGCTA-3'
PEMV kissing loop T-shaped structure	5'CCTGCGTCCCAA**TGTCAGCGACTC GAA**GTCGACGTAAAT-3'	5'CAACCATCGACG**GATAGCATTCCC GAG** GTCGCAGGGAGA-3'
TCV T-shaped structure	5'-CCGTTTTTGGTC**TGTCAGCGACTCGA A** GTCGACGTAAAT-3'	5'-AACCATCGACGT**TGTCAGCGACTC GAA** TTTAGTTCGGAG-3'

Table 4.1. Sequence of DNAzymes
The core sequences of each DNAzyme indicated with the bold, underlined letters were flanked with two substrate-recognition domains at the 5' and 3' ends

The recombinant tRNA-scaffold was combined with the appropriate DNAzymes at ratios of 1:2:2 to 1:4:4 recombinant tRNA-scaffold:5′-DNAzyme:3′-DNAzyme depending on the inserted RNA of interest. An additional 18-mer oligonucleotide (5′-GCCCGAACAGGGACTTGAA-3′), complementary to the 3′ side of the tRNA acceptor stem and T arm of the tRNA-scaffold, was added to the reaction in a ratio of 1:2 (recombinant tRNA-scaffold:18-mer oligonucleotide) to assist in annealing of the DNAzymes to the recombinant tRNA-scaffold. The mixture was heated at 90°C for 3 min and immediately snap-cooled on ice for 10 min. An equal volume of 100 mM MOPS (pH 7.2), 1 M NaCl, 250 mM KCl, 15 mM $MgCl_2$, 30 mM $MnCl_2$, and 1 mM spermidine) was added. Depending on the construct, the reaction was left at room temperature for 48 to 60 h and quenched with 80 mM EDTA. The RNA of interest was separated from the tRNA-scaffold, DNAzymes, and 18-mer oligonucleotide using

preparative denaturing PAGE. The RNAs were eluted from the gel using the Elutrap electroelution system (Whatman) and solvent-exchanged as described above.

4.2.8. NMR spectroscopy

The recombinant tRNA-scaffold samples were diluted with NMR buffer (as described in chapter 3) without any further folding step, while the RNAs of interest prepared from the recombinant tRNA-scaffold were re-folded using a slow-cool method in which RNAs were heated at 75°C for 3 min before the NMR buffer was added, followed by incubation at 37°C for 15 min. NMR spectra were collected at either 25°C or 37°C with either a Bruker Avance 600 or 800 MHz spectrometer equipped with an HCN triple resonance cryoprobe. The imino ^{1}H-^{15}N HSQC the tRNA/kl-TSS, the tRNA-scaffold, and kl-TSS were recorded with 1024 (t_2) x 256 (t_1) points using spectral width of 17 kHz (^{1}H) and 29 kHz (^{15}N). The ^{1}H-^{13}C HSQC of the kl-TSS RNA synthesized by tktA were recorded at 1024 (t_2) x 128 (t_1) points using a spectral width of 6 kHz (^{1}H) and 3.8 kHz (^{13}C) at the base C8 region and 8 kHz (^{1}H) and 1.8 kHz (^{13}C) for the ribose C5´region. The 2D ^{15}N-edited HNCO was collected with 2048 (t_3) x 48 (t_2) points with a spectral width of 12 kHz (^{1}H) by 1.8 kHz (^{15}N). The 2D ^{13}C-edited HNCO was collected with 2048 (t_3) x 80 (t_1) points with a spectral width of 12 kHz (^{1}H) by 2.7 kHz (^{13}C). All NMR data were processed using TOPSPIN 3.2 and NMRview (81).

4.3. Results

4.3.1. Pilot expression of the recombinant tRNA scaffolds in wt *E. coli*

NMR requires large quantities of RNA, and thus there is a need for protocols to synthesize RNA cost-effectively and efficiently. Previous work by Ponchon and Dardel

(2007) established that RNAs can be rapidly expressed *in vivo* and then purified to homogeneity (226). To obtain milligram quantities of homogeneous, uniform ^{15}N-labeled and site-specifically ^{13}C-labeled recombinant RNAs, previous schemes for recombinant RNA expression and purification were modified by incorporating the biomass production of site-specifically labeled rNTPs using wt and mutant *E.coli* strains. These highly and site-specifically labeled RNAs were obtained through selection of the appropriate growth media, application of a double selection process to select the best colony expressing the RNA, extension of a previously published anion-exchange purification method, and the optimization of 8-17 DNAzyme cleavage (Figure 4.2).

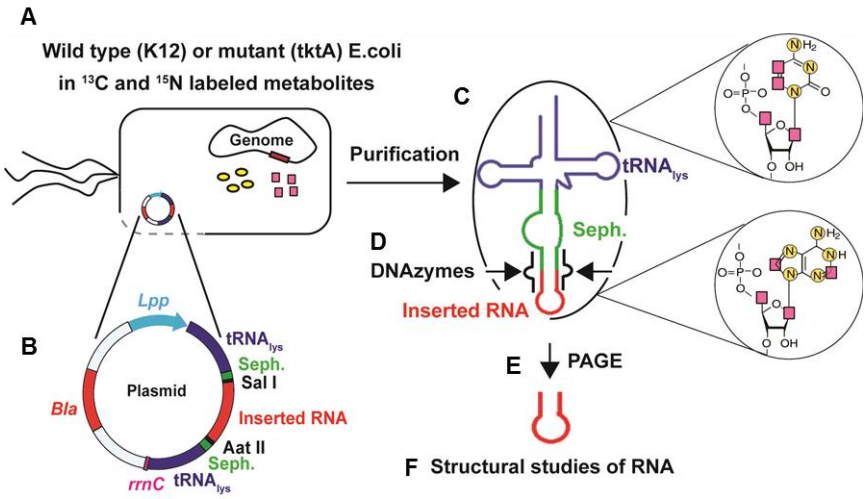


Figure 4.2. General scheme for the expression and purification of the recombinant tRNA-scaffold in wild type and mutant *E. coli* strains.
A. Wildtype K12 and mutant tktA *E. coli* strains containing the tRNA scaffold plasmid and grown in ^{13}C (pink squares) and ^{15}N (yellow circles) labeled metabolites. **B**. The anticodon loop of $tRNA_{lys}$-scaffold was replaced with the RNA of interest using *Sal I* and *Aat II* sites. The chimeric RNA transcript was under the control of the lipoprotein promoter (*lpp*) and terminated with a ribosomal RNA operon transcription terminator (*rrnC).* The sephadex tag was included for affinity purification. **C.** Incorporation of NMR-active isotopes into the recombinant tRNA-scaffold: *E coli* grown in minimal media supplemented with ^{15}N-ammonium sulfate and/or [^{13}C-1]-glucose or [^{13}C-1]-acetate. ^{13}C isotopes are denoted with purple squares and ^{15}N isotopes are denoted with yellow circles. This labeled chimeric tRNA-scaffold was then purified by anion exchange or affinity chromatography. **D.** The inserted RNA was excised from the tRNA-scaffold using two 8-17 DNAzymes. **E.** Purification of RNA of interest using denature PAGE. **F.** The final RNA product was utilized for structural and dynamic experiments using NMR spectroscopy.

To demonstrate the feasibility of such a method, four RNAs of various sizes, structural complexities, biological functions, and origins were generated (Figure 4.1). These include a 48-nt *Bacillus anthracis* fluoride-binding riboswitch aptamer (Fluo) (233, 234), a 54-nt S-adenosylmethionine (SAM) metabolite binding (SAM-II) riboswitch aptamer from the Sargasso Sea metagenome (117), an 81-nt PEMV kl-TSS, which is a 3'-UTR ribosome-binding cap-independent translation enhancer (51, 77) and a 118-nt TCV TSS, which is a similar ribosome-binding translation enhancer (74, 76). The RNAs were sub-cloned into the tRNA-scaffold plasmid (pBSKrnaSeph), which has a constitutive *lipoprotein* (*lpp*) promoter, causing the recombinant tRNA-scaffolds to be continuously expressed in wild type and mutant *E.coli* (226-228). Unlike the earlier approach, in which whole mixtures of transformed bacteria were used (228), a single wt K-12 colony containing each construct was initially used to inoculate 10 ml 2x TY cultures. The total extracted cellular RNA of each construct was analyzed using denaturing PAGE. All four constructs were efficiently expressed in rich media with the expected length (as indicated by arrows) (Figure 4.3).

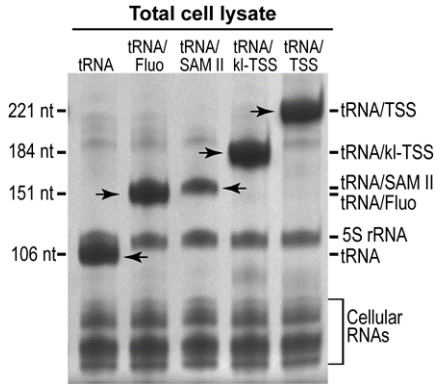


Figure 4.3. Expression of different recombinant tRNA-scaffolds in K12 grown in 2x TY rich media
The length of the tRNA scaffold is 106-nt. The tRNA/Fluo is 151-nt and Fluo RNA is 48-nt. The tRNA/SAM II is 154-nt and the SAM II is 52-nt. The tRNA/kl-TSS is 184-nt and kl-TSS is 81-nt. The tRNA/TSS is 221-nt and the TSS RNA is 118-nt. Expressed recombinant tRNA-scaffold were indicated with arrows.

4.3.2. Buffer selection for minimal media with a limited carbon source

In agreement with a previous study that reported variability in the growth rate and total RNA transcription of *E.coli* when grown in different minimal media (215), several constructs were initially poorly expressed in minimal media. To overcome these problems, conditions were explored that affected the growth rate and expression level of recombinant tRNA/Fluo, one of the poorly expressed constructs. In wt K-12 *E.coli*, this RNA was grown in two minimal media SPG-ampicillin (amp) and LMR-amp supplemented with 0.2 % (w/v) glucose. K-12 had a 5 h lag phase in both SPG-amp and LMR-amp (Figure 4.4A), consistent with a previous report (215). Surprisingly, after the lag phase, cells in SPG-amp grew much more rapidly and the culture reached a maximum

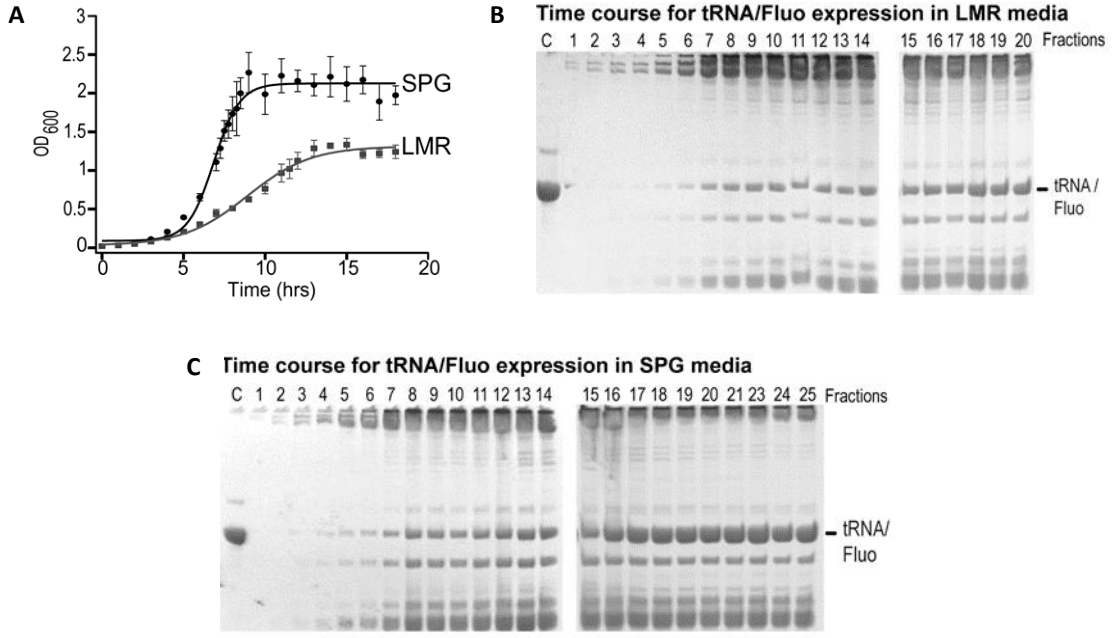


Figure 4.4. The growth and the yield of RNA expression of K12 on Studier phosphate (SPG) and Le Master (LMR) media
A. Optical density of *E. coli* K12 in batch culture in SPG (filled black circle) and LMR (filled gray square) media supplemented with 0.2% v/v glucose. The growth of K12 in each media was measured for every hour. Due to the fast growth of K12 in SPG, the growth at the end of log phase was measured every 15 min. The growth of K12 in each media was done in triplicate. **B** and **C**. Expression of tRNA/Fluo in LMR and SPG, respectively. Each fractions were collected at every time point. Total extracted RNA of each fractions were loaded on 10 % denaturing PAGE.

cell density after 5 h of rapid growth in log phase (Figure 4.4A). Cells in LMR-amp grew more slowly with 10 h in log phase and the maximum cell density was two-fold lower than the maximum cell density of cells grown in SPG-amp. Additionally, the expression of tRNA/Fluo was significantly higher in cells grown in SPG-amp (Figures 4.4B, C). Therefore, SPG-amp was chosen as the preferred buffer to make minimal media for all subsequent cultures.

4.3.3. Double selection of high-expressing *E. coli* clones

Several attempts to grow *E.coli* in SPG-amp produced low and inconsistent yields of the recombinant tRNA-scaffold. For example, the yield of tRNA/kl-TSS from randomly selected K-12 colonies grown in SPG-amp supplemented with 0.4 % (w/v) glucose varied from 18 mg/L to 26 mg/L. To minimize this variability and obtain a consistent yield of the recombinant tRNA-scaffold, a double selection process was adapted that had previously been used to obtain high yields and stable expression of recombinant proteins (239). In the first round of selection, five colonies were randomly

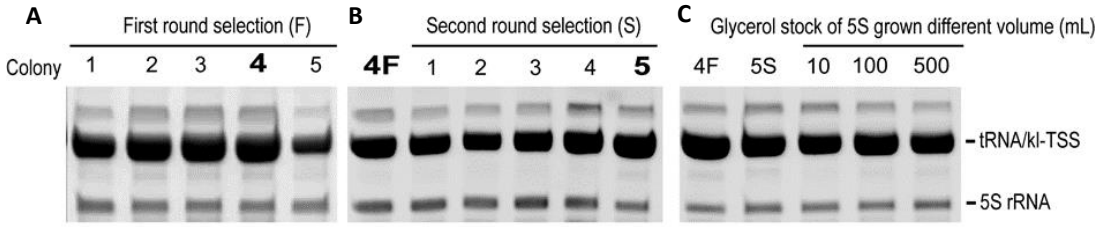


Figure 4.5. Double colony selection of K12 expressing tRNA/kl-TSS, grown in 0.4% glucose and D_2O
A. First round selection (F): five colonies were inoculated into 10 ml of media supplemented with 0.4% glucose and D_2O. Expression of tRNA/kl-TSS are shown using 10% denaturing PAGE. Colony 4 (bold) with best expression of tRNA/kl-TSS was spread on the second agar plates for second round selection. **B.** Second round selection (S): five progeny colonies of colony 4 (4F) were inoculated into target media. Colony 5 (bold) was selected to create the glycerol stock. **C.** Progeny bacteria of glycerol stock of colony 5 (5S) was grown in different volumes and expression of tRNA/kl-TSS of each volume were compared.

selected from the transformed *E.coli* initially grown on SPG-agar plates and subsequently inoculated into small-scale batch cultures supplemented with the target carbon source. The colony which expressed the highest RNA yield was spread on the second SPG-agar plate. In the second round of selection, the best-expressing progeny colony derived from the colony selected in the first round was used to make glycerol stocks. As expected, the first round of selection resulted in variable recombinant tRNA/kl-TSS expression in different colonies (Figure 4.5A). Using the 5S rRNA accumulating in each colony as an internal standard, colony 4 was selected as the best-expressing colony in the first selection round (Figure 4.5A) and generated stable and consistent expression of progeny colonies in the second round of selection (Figure 4.5B). Progeny colony 5 which expressed slightly higher yield than other colonies in the second round of selection was used to make the glycerol stock to inoculate future batch cultures. When used to inoculate higher volume cultures, bacteria of this glycerol stock produced stable RNA expression (Figure 4.5C).

4.3.4. Large-scale production of the recombinant tRNA-scaffold in wt K-12 and mutant *E. coli* strains

A judicious choice of cell strain, carbon source, and optimized production conditions enabled the production of RNAs with the desired labeling pattern of interest and in high yield. *E.coli* strains were chosen for their ability to synthesize RNA with site-specific labeling patterns useful for providing NMR structural assignments and base-pairing information. For example, K-12 grown in [1-^{13}C]-acetate as the only carbon source creates ^{13}C site-specifically labeled rNTPs at the carbonyl position of each base. These carbonyl carbon chemical shifts provide valuable base-pairing information (146),

which helps to resolve overlap in the imino region in a 3D experiment. In addition, when grown in [1-^{13}C]-glucose, the tktA strain, whose metabolic flux is redirected mainly through the oxidative pentose phosphate pathway, is capable of site-specifically labeling the bases at the C2 (~75%) and C8 (44%) carbons of purines, and C5 (43%) and C6 (28%) carbons of pyrimidines (214), which are important atomic sites in structural assignments and dynamics experiments of RNA molecules (151). For these reasons, K-12 and tktA were chosen to produce uniformly ^{15}N labeled and site-specifically ^{13}C labeled RNAs. To create useful labeling patterns, the K-12 and tktA strains were grown in SPG-amp supplemented with $^{15}N_2$-ammonium sulfate in combination with [1-^{13}C]-acetate or [1-^{13}C]-glucose, respectively, as the sole carbon source.

To maintain consistent recombinant tRNA-scaffold expression levels, a two-step enrichment process was used. First, cells from the doubly selected glycerol stock were seeded into 10 ml LB-amp, followed by transfer of these cells into 100 ml SPG-amp supplemented with 0.4 % (w/v) glucose. Second, these cells were then transferred into 1 L of media supplemented with the selected carbon source for final growth. This simple procedure greatly improved yields. Since RNA is best expressed in log phase, and to minimize degradation that occurs in the stationary phase (240), bacteria in each seed culture were transferred when cells were in the log phase (OD_{600} of 0.5). Maintaining the initial OD_{600} of subsequent growth at 0.05 preserved the growth rate and gave consistent expression of the recombinant tRNA-scaffold, as observed in a previous study (215). Aeration was also an important factor that contributed to high level expression of the recombinant tRNA-scaffold in mutant *E.coli* grown in minimal media (239, 241). A previous study optimizing protein expression in minimal media showed that maintaining

the aeration by growing *E.coli* in multiple small volumes (5 x 50 ml) instead of one large volume (250 ml) of batch culture increased cell density significantly (239). In our hands, growing 10 ml cultures in 50 ml flasks, 50-100 ml cultures in 250 ml flasks, and 500-1000 ml cultures in 4 L flasks maintained the growth rate and increased the RNA yield to milligram quantities (Table 4.2). The highest expression of tRNA/Fluo was obtained 2 h into the stationary phase (Figure 4.4C). Therefore, cells were collected 2 h into the stationary phase during large-scale production of the recombinant tRNA-scaffolds.

Compared to wild type K-12, yields of the recombinant tRNA-scaffolds from mutant *E. coli* were reduced when both *E. coli* were grown in SPG-amp supplemented with 0.2 % (w/v) glucose. However, after double selection and a two-step scale-up growth, a yield of 8.6 mg/L was obtained for ^{13}C-labeled tRNA/Fluo in tktA. This value was 2-fold greater than that obtained for ^{15}N-labeled tRNA/Fluo in K-12 without colony selection. The benefit of double selection was also observed in K-12 colonies grown in SPG supplemented with acetate. In the absence of double selection, cells grew poorly and expressed little detectable RNA. However, using the double selection strategy, 1.25 mg/L tRNA/kl-TSS and 0.7mg/L tRNA/SAM-II were obtained in K-12 grown in [1-^{13}C]-acetate (Table 4.2), and usable NMR spectra were generated using these RNAs.

	E.coli strain			
	K12			tktA
RNA type	2x TY	^{15}N-amonium sulfate, 0.4% glucose	^{15}N-amonium sulfate,0.3 % ^{13}C-1-acetate	^{15}N-amonium sulfate, 0.2% ^{13}C-1-glucose
tRNA/Fluo	10.6 mg	3.2 mg	NA	8.6 mg*
tRNA/SAM II	5 mg	NA	0.7 mg*	NA
tRNA/kl-TSS	38 mg	18-26 mg	1.25 mg*	4.4 mg

Table 4.2. Yield of chimera RNA per liter of culture
*: represents sample prepared from glycerol stock of with doubly selected colony and two-step growth. And without selection, no yield or little yield was obtained. NA: not produced

4.3.5. RNA purification

To improve the purity of the recombinant tRNA-scaffolds, anion exchange chromatography (230) was adapted by extending the washing step and including a linear low salt concentration gradient. The majority of small cellular RNAs were removed using 5 column volumes (CV) of 0.44 M NaCl and 5 CV of a linear gradient from 0.44 to 0.52 M NaCl (Figure 4.6). As a result, the recombinant tRNA-scaffold was well-resolved with little to no contamination by cellular RNAs and was eluted in an extended step of 5 CV of 0.52 M NaCl. Larger RNAs, such as 23S rRNAs, were removed using a gradient from 0.52 M to 1 M NaCl. A final step of 2 CV of 1 M NaCl was added to remove any contamination before reusing the column.

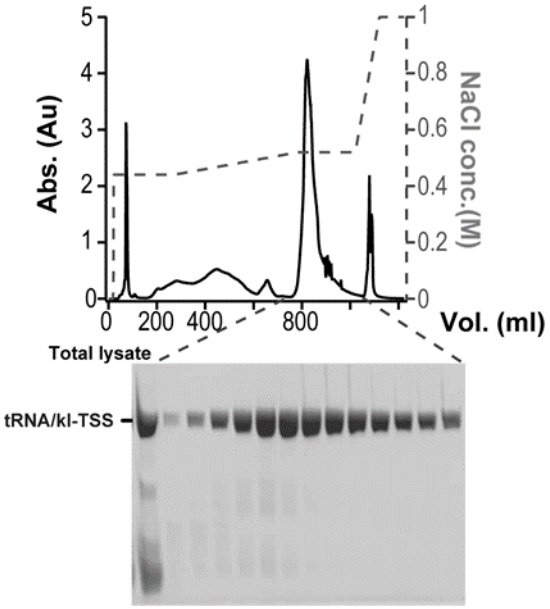


Figure 4.6. Purification of recombinant ^{15}N-tRNA/kl-TSS using the Source 75 Q anion exchange column

The total cellular RNA was extracted from 1 L of culture and loaded onto the column. The UV absorbance was monitored at 260 nm and is shown as a solid, black line. The salt concentration is presented as dashed, grey line. By using a combination of steps and different gradients of NaCl, the ^{15}N-tRNA/klTSS was eluted in a single peak. The pure ^{15}N-tRNA/kl-TSS from different eluted fractions is shown using 10% denaturing PAGE.

To overcome the problem of purifying recombinant tRNAs with sizes similar to cellular RNAs, the Sephadex tag in the tRNA-scaffold was utilized for affinity purification. Previously, Sephadex G-100 or G-200 beads were packed manually into a low-pressure column for affinity purification (226, 230). However, this procedure was labor intensive. I therefore used a HiLoad 26/600 Superdex 75 pg column (GE Healthcare) as an automatic affinity column to purify the recombinant tRNA-scaffold (Figure 4.7). The majority of cellular RNAs were removed with 1 CV purification buffer, and bound tRNA/kl-TSS was eluted in well-resolved fractions using 4 M urea (Figure 4.7). RNAs that bound non-specifically to the column, and thus were eluted with the tRNA/kl-TSS, were removed in the downstream purification by denaturing PAGE. Both anion exchange and affinity chromatography provided highly efficient purification. After solvent-exchange, the purified RNA was appropriate for NMR experiments or cleavage using DNAzymes.

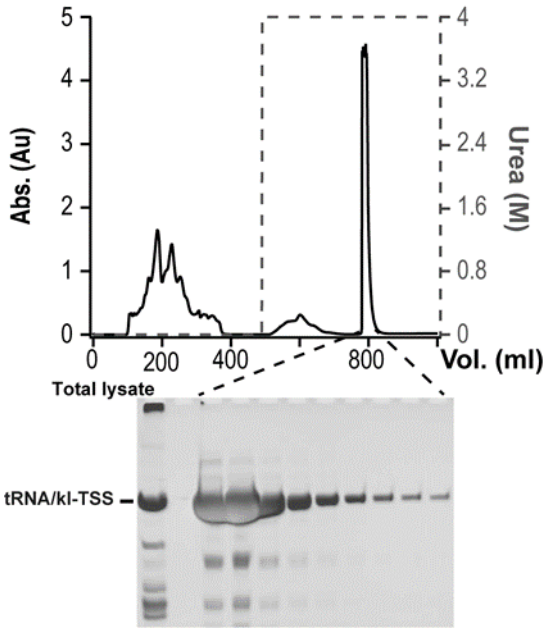


Figure 4.7. Purification of recombinant ^{15}N-tRNA/kl-TSS using affinity chromatograph
Total cellular RNA extracted from 1 L of culture was loaded onto HiLoad 26/600 Superdex 75 pg column (GE Healthcare). The UV absorbance was monitored at 260 nm and is shown as a solid, black line. Cellular

RNAs were washed with purification buffer containing 100 mM NaCl. The ^{15}N-tRNA/kl-TSS was eluted using 4 M urea, represented as a dashed, grey line. The ^{15}N-tRNA/kl-TSS from different eluted fractions is shown using 10% denaturing PAGE.

4.3.6. RNA excision using 8-17 DNAzymes

To resolve the problem of overlapped NMR signals from the recombinant tRNA-scaffold, the RNA of interest was removed from the tRNA-scaffold. Several trans-acting enzymes have been previously used for this excision step, including RNase H and catalytic DNAs known as DNAzymes (228, 229). With the capability of fast cleavage and less requirement of RNA substrate, RNase H was deemed preferable in a previous study. However, RNase H can cleave RNA nonspecifically in the absence of costly modified 2´-O-methyl RNA/DNA chimera (243-245).

I therefore chose to use 8-17 DNAzymes, which cleave RNA specifically and are cost-effective due to their ease of synthesis and reusability. 8-17 DNAzymes contain a catalytic domain that cleaves RNAs at specific, unpaired dinucleotides, which are flanked by two substrate-recognition domains that form Watson-Crick base pairs with the RNA substrate (Figure 4.2) (242, 246). In our study, the catalytic sequence of 8-17 DNAzymes that specifically cleave unpaired dinucleotide cleavage sites (5′-GG-3′ and 5′-GA-3′) were chosen based on their rapid cleavage capability (242). Two DNAzymes were designed to cleave each tRNA-scaffold, one to cleave at the 5′ end and the other to cleave at the 3′ end of the inserted RNA of interest (Table 4.1, Materials and Methods).

Accessibility of the two substrate-recognition domains of each DNAzyme to the recombinant tRNA-scaffold determines cleavage efficiency. To optimize the interaction of the DNAzymes with the recombinant tRNA-scaffold substrate, an 18-mer oligonucleotide was added that annealed to the acceptor stem and the T arm of $tRNA_{lys}$.

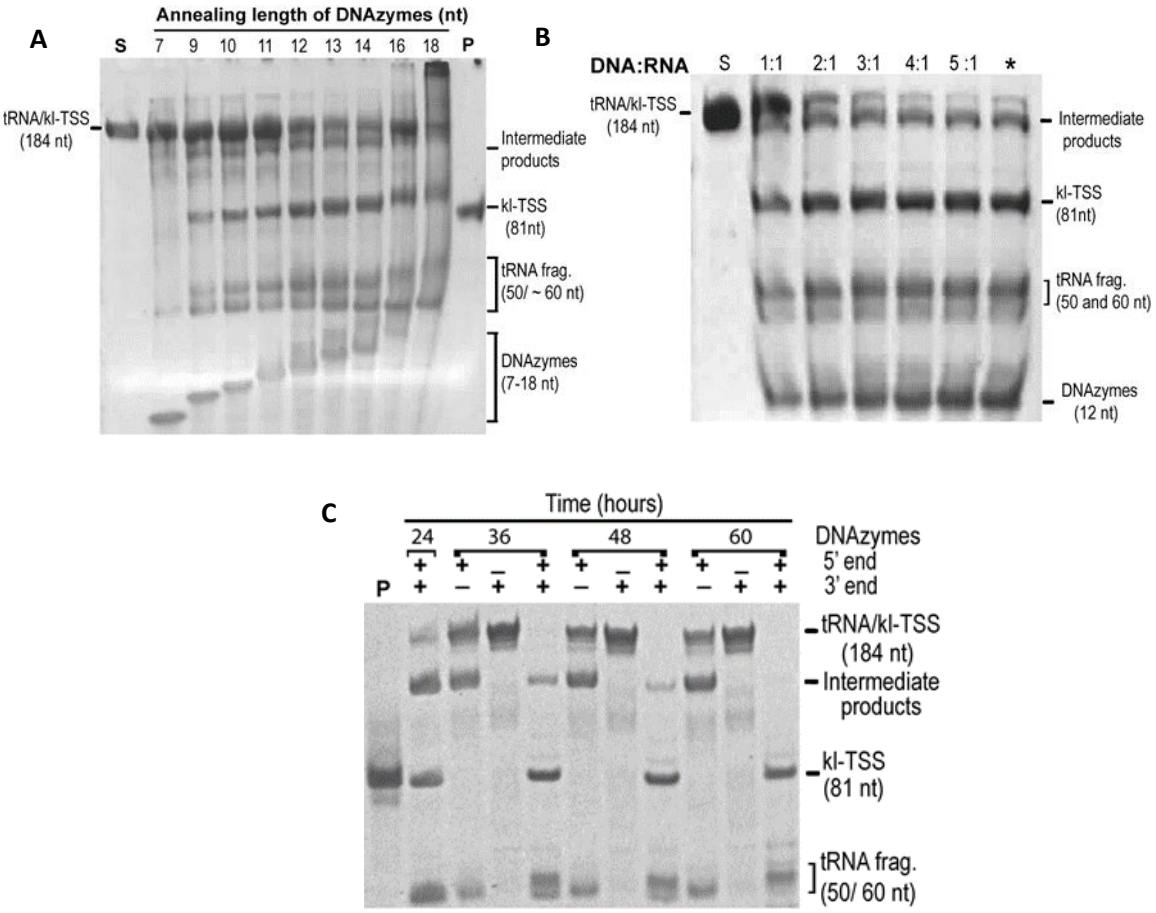


Figure 4.8. Optimization of 8-17 DNAzyme cleavage of recombinant tRNA/kl-TSS
A. The cleavage of 8-17 DNAzymes with different lengths of substrate-recognition domains from 7-nt to 18-nt of both ends. Cleavage products were separated by 10% denaturing PAGE. The tRNA/kl-TSS substrate (S) in buffer without DNAzymes was used as negative control. The kl-TSS RNA transcribed *in vitro* was used as the positive control (P). **B.** Various ratios of 8-17 DNAzyme:tRNA were tested. The reaction was stopped at 48 h. The 18-mer oligonucleotide was added in an equal molar amount to the 8-17 DNAzymes except in the final lane (*), in which the DNAzymes:oligonucleotide:tRNA ratio was 5:2:1. **C.** The efficiency of cleavage of 12-nt DNAzymes. Mixture was then incubated in DNAzymes buffer at room temperature. Aliquots were collected at 24, 36, 48, 60 hours. After 60 hours, the cleavage is nearly 100% complete.

Comparison of the cleavage efficiencies of DNAzymes with substrate-recognition domains ranging from 7-18 nt in length showed that DNAzymes with short (<11 nt) substrate-recognition domains cleaved less efficiently than DNAzymes with long substrate-recognition domains (Figure 4.8A). However, DNAzymes with long (>13 nt) substrate-recognition domains also cleaved the recombinant tRNA-scaffold

nonspecifically (Figure 4.8A). In the case of tRNA/kl-TSS, DNAzymes with 12-nt substrate-recognition domains were utilized in a ratio of 4:4:2:1 5′ DNAzyme:3′ DNAzyme:18-mer oligonucleotide:recombinant tRNA-scaffold, respectively (Figure 4.8B). This combination produced the most efficient cleavage. However, for the less structured tRNA/Fluo, efficient and complete cleavage was obtained using DNAzymes with substrate-recognition domains that are 11-12 nt in length. In this case, a ratio between components of the reaction as low as 2:2:2:1 was sufficient for optimal cleavage (data not shown). The cleavage of tRNA/kl-TSS by DNAzymes was nearly 100 % at 60 hours of incubation (Figure 4.8C).

4.3.7. NMR spectroscopic analysis of recombinant RNAs

Analysis of the purified *in vivo* labeled RNAs by ^{1}H-^{15}N HSQC correlation experiments revealed a unique set of peaks in the NMR spectrum suggestive of one product with a single conformation in solution. The peaks of tRNA/kl-TSS overlaid well with the peaks of the tRNA-scaffold and the kl-TSS, again confirming that the recombinant tRNA-scaffold folds into a conformation that maintains the structure of the kl-TSS. The presence of the kl-TSS did not appear to affect the folding of the tRNA-scaffold (Figure 4.9). Using previous $tRNA_{lys}$ spectral assignments (247), peaks corresponding to modified rNTPs (e.g. $m^{7}G46$) were clearly observed in the spectra of the tRNA/kl-TSS (Figure 4.9, top and bottom insets). In the ^{1}H-^{15}N HSQC spectrum of the ^{15}N-kl-TSS prepared in K-12 (Figure 4.9), we observed five A-U and ten to twelve G-C base pairs, consistent with the predicted secondary structure. Additionally, the spectrum indicates a strong non-canonical Watson-Crick GU wobble base-pair, also

consistent with the predicted secondary structure. However, the spectrum of the tRNA/kl-TSS was heavily overlapped, precluding its direct use for other downstream applications.

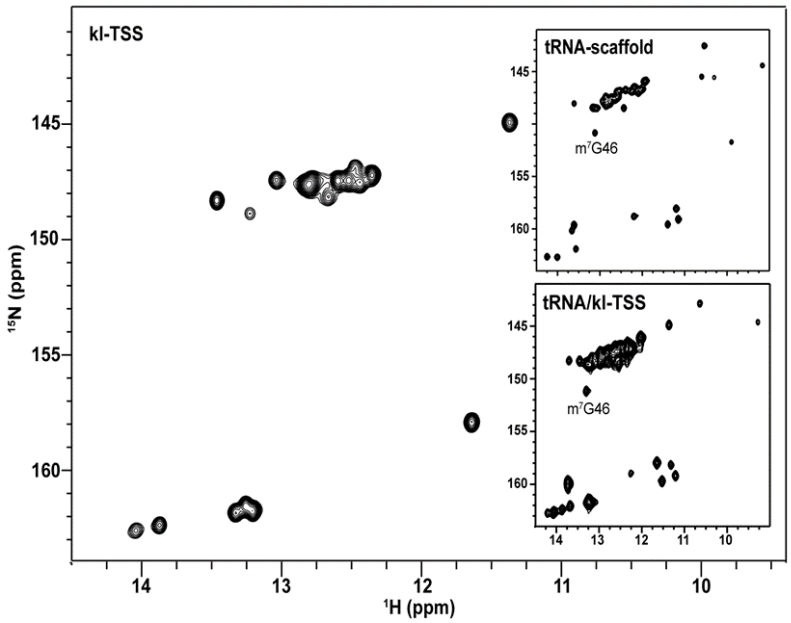


Figure 4.9. 2D ^{1}H-^{15}N imino HSQC spectra of purified ^{15}N-labeled kl/TSS
The ^{15}N-labeled tRNA/kl-TSS (bottom inset) was made in K12 grown SPG supplemented with ^{15}N-ammonium sulfate. The imino ^{1}H-^{15}N spectra of tRNA is shown at the top right inset. The ^{15}N-labeled kl-TSS RNA was prepared from the cleavage of the ^{15}N-labeled tRNA/kl-TSS using two 8-17 DNAzymes. All peaks of the tRNA/kl-TSS spectra overlay with either peaks of the tRNA scaffold or the kl-TSS spectra. Modified peak m^{7}G46 of the $tRNA_{lys}$ is indicated in the spectra of the ^{15}N-labeled tRNA/kl-TSS and the ^{15}N-labeled tRNA scaffold.

Compared to ^{13}C labeling, ^{15}N labeling has limited utility, and yet till now all *in vivo* RNA labeling has focused on ^{15}N labeling. Unlike the ^{15}N sites in RNA, the ^{13}C sites are more widely distributed among the different structural elements (248), making the ^{13}C probes of significant interest. I report here the first example of such an *in vivo* site specific labeling using an *E. coli* strain deficient in the transketolase gene (tktA), which we had earlier shown shunts most of the metabolic flux via the oxidative pentose phosphate pathway (214). In agreement with those earlier studies (214), *in vivo* production of RNA using this *E. coli* strain on [1-^{13}C]-glucose produced very high levels

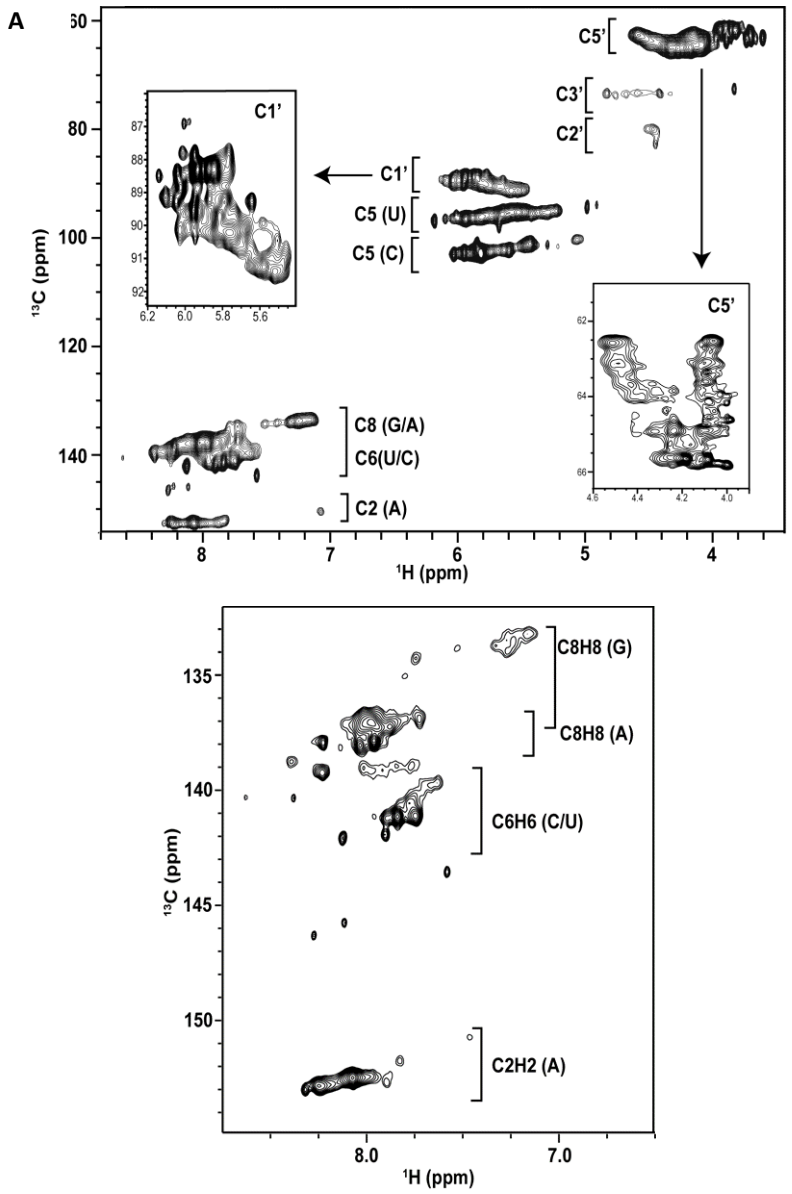


Figure 4.10. 2D ^{1}H-^{13}C HSQC of the kl-TSS
2D ^{1}H-^{13}C HSQC spectra obtained at the ribose region (**A**) and C8 region (**B**) of the kl-TSS extracted from tktA grown in SPG with ^{15}N-amonium sulfate and [^{13}C-1]-glucose. Labeled carbon regions were bracketed.

(~75%) of enrichment of the ribose C5′ positions and moderate enrichment (~20%) of the ribose C1′ positions without unwanted one-bond ^{13}C1′-^{13}C2′ or ^{13}C4′-^{13}C5′ scalar and dipolar couplings (Figure 4.10A). RNAs made by tktA grown in [1-^{13}C]-glucose also had useful nucleobase labeled sites. Ten out of fourteen C2 carbon peaks of adenine as well as the C8 carbon of all the purines in the kl-TSS RNA were labeled and readily

identifiable (Figure 4.10B). The C5 and C6 carbons of pyrimidines were also labeled without any residual carbon-carbon coupling (Figures 4.10A, B), as reported previously. In a previous study, the C6 and C8 carbons of RNA were employed in 3D NOE experiments to resolve the overlap problem of proton-proton cross-peaks, allowing structural assignment and providing distant restraint of the RNA. I anticipate this would be a useful application of the labels reported herein.

In addition to the protonated carbon sites typically probed in NMR studies, the non-protonated carbonyl carbon sites, though less used, are useful probes of hydrogen-bonding and ligand binding. The chemical shifts of these carbonyl groups, such as the C2 and C4 of uracil nucleobases, involved in hydrogen bonding appear to be sensitive to the nature of the bond such that uracil C2 and C4 carbons have different chemical shift signatures in an AU or GU or UU base pairs (146). Thus, being able to label these sites is of great advantage. To showcase the potential utility of this approach for *in vivo* labeling, tRNA/SAM-II was produced in K-12 grown in SPG-amp supplemented with $^{15}N_2$-ammonium sulfate and [1-^{13}C]-acetate as the sole carbon source. I were able to quickly verify with 1D ^{13}C HSQC that the SAM-II RNA had ^{13}C isotopes at the C4 carbons of all rNTPs, as well as the C2 carbons of cytosine and uracil, the C6 carbons of guanine and adenine (Figure 4.11A), and the C2′ and C3′ ribose carbons (Figure 4.11B. As expected, all of these labeling patterns were consistent with those of rNTPs prepared from *E.coli* grown in [1-^{13}C]-acetate (211). To link the imino protons with the carbonyl carbons, we ran a 3D HNCO experiment (249-251) on the 52 nt SAM-II RNA (Figures 4.11D, E). The magnetization transfer steps during the HNCO experiment in UTP and GTP are shown in Figure 4.11C. The proton and nitrogen peaks of the 3D ^{15}N-edited HNCO

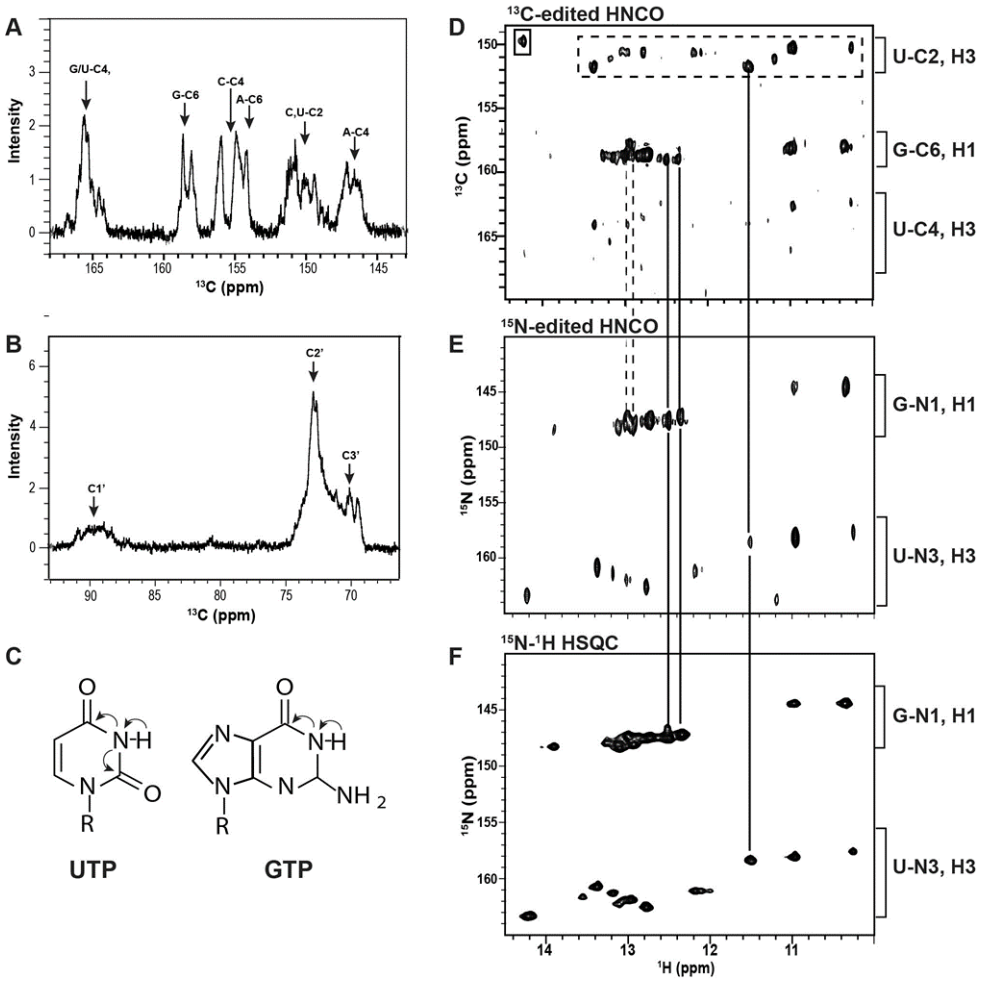


Figure 4.11. NMR spectra of SAM-II extracted from K12 grown in SPG supplemented with ^{15}N-ammonium sulfate and [^{13}C-1]-acetate

A, B. 1D carbon spectra of base region and ribose region of SAM-II. Resonance of each carbon region were indicated. **C.** Magnetization transfer of bases of the UTP and GTP in 3-D HNCO experiment. The R stands for ribose. **D.** 2-D ^{13}C-edited HNCO NMR at imino region of SAM-II. **E.** 2D ^{15}N-edited HNCO NMR at imino region of SAM-II. **F.** 2-D ^{1}H-^{15}N HSQC spectra of the imino region of SAM II. Dashed lines indicate the connectivity of nitrogen and carbon resonance of the same base. Two black dashed lines presents that the two imino peaks have similar proton and nitrogen ppm but connect with two well separated carbons.

spectrum of SAM-II corresponded with those obtained in 2D ^{1}H-^{15}N HSQC spectrum of the RNA, and the 2D HNCO experiment correlating the imino ^{1}H and carbonyl ^{13}C showed the labeled carbon positions of the UTPs and GTPs of SAM-II RNA (Figures 4.11D, F). An important use of the HNCO for signal readout is to spread the ^{1}H-^{15}N correlation map along a third dimension. In that way overlapped proton and nitrogen

resonances in the 2D ^{1}H-imino HSQC of the SAM-II RNA are now be well-resolved (Figures 4.11D, F). The resonances of the carbonyl groups of uracils (C2 and C4) derived from the HNCO experiment can allow the identification of the geometry of base pairing (146). The chemical shifts of C2 of uracils that engage in non-canonical hydrogen bonding in AU, UU and GU are typically shifted downfield with respect to that of non base-paired U-C2 of canonical base-paired AU. There are also clear differences of resonance between base-paired and non-base-paired U-C4. Canonically base-paired U-C4 are normally shifted downfield (146). In case of SAM II riboswitch, many non-canonical base-paired U of the RNA (117, 252) were not detected in normal ^{15}N-^{1}H HSQC (Figure 4.11F) but were clearly observed in ^{13}C-edited HNCO spectra (Figure 4.11D). Therefore, labeling at C2 and C4 position of uracil together with application of HNCO experiment can bring much benefit to probing hydrogen bonding of nucleotides.

4.4. Discussion

Advances in RNA labeling technologies are necessary for continuing to move forward the field of RNA NMR structural biology. Towards this goal, we present here a versatile method to prepare uniformly ^{15}N isotopic and site-specifically ^{13}C isotopic labeled RNAs *in vivo*. To demonstrate the versatility of our approach, we optimized RNA expression and performed 2D and 3D-edited NMR experiments on a 52-nt SAM-II riboswitch RNA and a 71-nt cap-independent translation enhancer RNA element located in the 3' UTR of *Pea enation mosaic virus* (kl-TSS RNA).

4.4.1. Optimization of *in vivo* labeling

The *in vivo* labeling technologies introduced here combine the best attributes of the tRNA-scaffold approach with the selective biomass bio-production method (214, 228). Five key features were introduced to optimize yields. Use of SPG media allowed for twice the cell density and higher RNA yield than that of other commonly used minimal media (Figure 4.4). Similarly, the adoption of a double colony selection protocol previously used to increase protein expression in *E.coli* (239), boosted the yields of the tRNA/kl-TSS, Fluo and SAM-II riboswitch RNA, which initially had either low or no yield of recombinant tRNA-scaffold expression (Table 4.2). Double selection also provided for consistency in RNA expression. To solve the problem of slow growth of *E.coli* cells in minimal media, the ratio between the culture volume and the flask volume was optimized, finding that a 1:8 ratio in a normal shaker with a maximum speed of 300 RPM was most effective. To maintain consistency in RNA production, seed cells were grown without exceeding an OD_{600} of 0.5, with the initial OD_{600} of the following culture fixed at 0.05 (215). Finally, to excise the labeled RNA from the recombinant tRNA-scaffold, two 8-17 DNAzyme were tested containing core sequences designed to cleave the dinucleotides 5′-GG and 5′-GA. Based on the optimization of the DNAzymes cleavage reactions, we found that the optimal length of the substrate-recognition domains of the DNAzymes should be 12 nt. Shorter or longer lengths caused less efficient or nonspecific cleavage.

4.4.2. NMR of ^{15}N and selective ^{13}C labeled RNAs

RNA NMR suffers from extensive chemical shift overlap of the constituent nuclei (^{1}H, ^{13}C, ^{15}N, ^{31}P). In the past, the chemical shift overlap problem was partially addressed

by the introduction of uniform labeling, and use of this approach has enabled spectacular success in studying the structure and dynamics of RNAs of up to 50 nt (150). However, uniform labeling introduces direct one-bond and residual dipolar ^{13}C–^{13}C couplings that degrade the attainable resolution and sensitivity. Ultimately, this labeling strategy prevents accurate measurement of ^{13}C relaxation parameters such as longitudinal relaxation rate (R_1), transverse relaxation rates (R_2), and heteronuclear Overhauser effect (hNOE), and CPMG (Carl-Purcell-Meiboom-Gill) relaxation measurements. Although for CEST (Chemical Exchange Saturation Transfer) and $R_{1\rho}$ (rotating-frame relaxation rate) measurements (253-257), these couplings are more of a nuisance than an obstacle because elegant spectroscopic strategies exists to circumvent the coupling problem. Nonetheless, these couplings can complicate and limit the range of applicability of CEST and rotating-frame relaxation rate ($R_{1\rho}$) measurements and their effects must be explicitly accounted for in data analysis (255-257).

Of the five approaches introduced to address the limitations of uniform labeling (151, 258), selective biomass production of NMPs overcomes the isotopic scrambling problem with adequate suppression of ^{13}C–^{13}C couplings. Nonetheless, low overall yields remain a limiting issue (212, 214, 216, 259). The *in vivo* labeling technologies introduced here, which combine the recombinant tRNA-scaffold and selective biomass bio-production approaches, is one way to alleviate these problems.

The *in vivo* production of ^{15}N uniformly and ^{13}C site-specifically labeled RNAs is cost-effective by using inexpensive precursors as metabolic substrates (215). Using the wild type K-12, 18-26 mg of ^{15}N-labeled tRNA/kl-TSS was obtained per liter culture. This is equivalent to ~1 mM of 81-nt ^{15}N labeled RNA in 250 µl. ^{15}N-only labeling in K-

12 affords a rapid delineation of the number of A-U and G-C Watson-Crick base pairs within the tRNA/kl-TSS, and these overlay well with the peaks of the tRNA-scaffold alone and those of the kl-TSS RNA excised from the tRNA-scaffold alone (Figure 4.9). However, compared to ^{15}N-only labeling, ^{13}C labeling allows access to more widely distributed sites within the different RNA structural elements, among other benefits (146, 150). *In vivo* labeling of RNA using the tktA *E. coli* strain enables site-specific placement of ^{13}C isotopes at predictable sites within the ribose (~75% of C5′ positions, ~20% of C1′ positions, and negligible labeling elsewhere in the sugar ring) and nucleobase moieties (~75% of C2 and ~44% of C8 carbon atoms of purines and ~43% of C5 and ~28% of C6 carbon atoms of pyrimidines) (214). A significant benefit of this labeling is that unwanted one-bond scalar and dipolar couplings (e.g. ^{13}C1′-^{13}C2′ or ^{13}C4′-^{13}C5′) are completely eliminated (Figure 4.10). These labeling patterns add to the growing arsenal of technologies that will be useful for many structural, dynamic, and functional RNA studies.

This *in vivo* labeling approach can also provide direct access to non-protonated carbonyl carbon sites. The *in vivo* production of tRNA/SAM-II using K-12 grown in SPG-amp supplemented with $^{15}N_2$-ammonium sulfate and [1-^{13}C]-acetate as the sole carbon source enabled site-specific enrichment of C4 carbons of all nucleotides, as well as the C2 carbons of cytosines and uridines, C6 carbons of guanosines and adenosines, and C2′ and C3′ ribose carbons with ^{13}C isotopes (Figures 4.11A, B). Given that the C2, C4, and C6 chemical shift resonances appear to correlate with the nature of the hydrogen bond (e.g. A-U, G-U, or U-U base pairs) (146), the ability to observe these shifts will go a long way in mapping the sites of ligand or drug binding to RNAs of biological interest.

Furthermore, using 3D HNCO NMR experiments, the imino protons can be readily linked with their carbonyl carbons (Figure 4.11D). Thus, using 3D ^{15}N-edited HNCO for signal readout promises to remove spectra overlap by spreading the ^{1}H-^{15}N correlation map along a third dimension to resolve overlapped proton and nitrogen resonances.

Chapter 5: Conclusions and future directions

RNA is a key factor involved in various cellular processes. The diverse functions of RNA are predicated upon its ability to assume different structural conformations associated with different properties. Information on folding dynamics for RNA structures is thus essential for understanding their biological function. Although detailed information of chemical composition and RNA secondary structure is currently available, the rules that govern RNA folding into 3-D structures, or how RNA structures switch between folding states to perform different functions remains to be fully elucidated. Due to their small simple genomes, TCV and PEMV, two (+)-strand plant RNA viruses, have been extensively used as models for analyses on how viruses maximize the use of their short 3´ UTRs for translation and replication. The aim of this thesis is to analyze the dynamic structures of two 3'UTR-containing CITEs to shed light on tertiary RNA structure and folding.

In Chapter 2, by using OT to analyze the dynamics of structural folding of the TCV TSS, I determined that the TSS pseudoknots and Mg^{2+} play important roles in increasing the stability of TSS hairpins. Pseudoknotted H4a/Ψ_3, in association with the upstream A-rich sequence (now to be referred to as AR-H4a/Ψ_3), was the most stable RNA element while its hairpin component alone, H4a, was the least stable. Previous studies have shown that the formation of RNA tertiary structures depends on the presence of high concentrations of a monovalent salt or the presence of Mg^{2+} (173, 260). Interestingly, my study provided the first evidence that pseudoknotted H4a/Ψ_3 was able to nucleate at 10 mM NaCl and formed the nearly full-length sized pseudoknot at a concentration of 250 mM NaCl, indicating that H4a/Ψ_3 is able to form in the absence of

Mg^{2+}. In contrast, the formation of Ψ_2 in the TSS is Mg^{2+} -dependent. The possible presence of Ψ_2 in Mg^{2+} increased the stability of H4b and coupled the unfolding process of H4b with H5. When induced by force in the absence of Mg^{2+}, the TCV TSS initiates unfolding with H4b followed by H5 and AR-H4a/Ψ_3. Conversely, in the presence of Mg^{2+}, the complex containing Ψ_2, H4b and H5 unfolds as a unit and is subsequently followed by the unfolding of AR-H4a/Ψ_3. tRNA-like structures are found in the 3' UTR of many positive-strand viruses and are proposed to function as conformational switches to regulate translation and replication (120, 179, 261). Recently, the first atomic resolution tRNA-like structure in turnip yellow mosaic virus is solved at 2 Å, in which the two pseudoknots of this RNA were also proposed to be involved in the regulation of the conformational switch (262). Analysis of the folding pathway for the TCV TSS also suggested the following structural connections: (1) between H4b and H5 via Ψ_2; (2) between AR-H4a/Ψ_3 and H5; and (3) between AR-H4a/Ψ_3 and Ψ_2. These connections might possibly allow the individual RNA elements within the TSS to facilitate each other's folding. One explanation includes the possibility that the TSS (and its surrounding sequences) might undergo a conformational switch upon the RdRp binding. Similar interconnections are likely present in other viral tRNA-like structures, allowing these RNAs to efficiently switch between conformations. However, regarding the folding pathway of the TCV TSS, a number of questions still remain to be addressed: (1) How does the TSS fold in vivo? (2) How do salt concentrations or RdRp binding induce the conformational changes in the TSS? and (3) If the RdRp is the main factor promoting the conformation switch, is this due to the physical interaction of the RdRp with its target

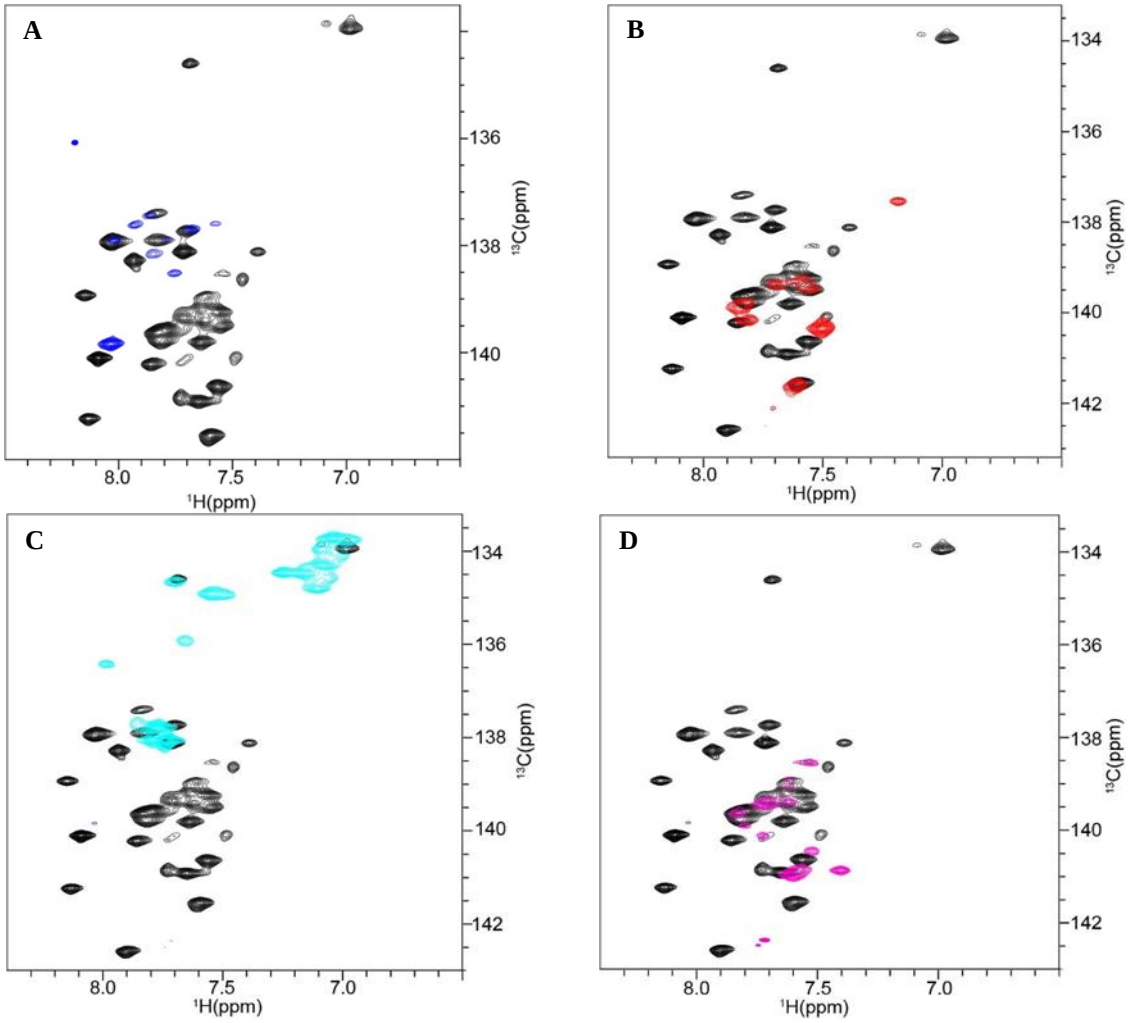
RNA sequence, or does the RdRp merely select the populations of RNAs already folded in a binding-competent conformation?

In Chapter 3, EMSA was used to show the interaction between the PEMV kl-TSS and 5H2. The hydrogen-bond pattern of the kl-TSS was confirmed using a 2D ^{1}H-^{15}N HSQC experiment. Based on the resonance assignment of the PEMV kl-TSS, the interaction between kl-TSS with 5H2 further showed that the P2 loop, and not P1 or P3, of the kl-TSS is solely responsible for the interaction. The SAXS envelope of the kl-TSS suggests that the PEMV kl-TSS does not form a tRNA shape, as was previously predicted. The sparse ensemble selection and bootstrapping techniques were used for the first time to score the 3-D models of the kl-TSS that were fitted in its SAXS molecular envelope. After vigorous resampling using bootstrapping, similar two models of the kl-TSS were obtained, indicating that the kl-TSS may have two conformational states. One conformation contributes 77% to the overall population while the other conformer accounts for 23%. Differences between the conformations of the PEMV kl-TSS and the TCV TSS can explain the difference in the ribosome binding pattern for these two RNAs. However, several questions regarding the PEMV kl-TSS structure still remain to be answered: (1) Which conformational state of the PEMV kl-TSS is responsible for its binding with ribosomes and 5H2? (2) What is the ribosome binding site in the PEMV kl-TSS? and (3) What factors cause the conformational switch in the PEMV kl-TSS?

Chapter 4 presented a fast and cost-effective methodology for obtaining milligram quantities of uniformly ^{15}N-labeled and site-specifically ^{13}C-labeled RNAs for structural and dynamic studies using NMR. This method has several advantages, including: (1) use of inexpensive starting materials to prepare homogeneous RNAs *in vivo*; (2) site-specific

labeling patterns that help to overcome chemical shift overlap problems and remove scalar coupling and ^{13}C-^{13}C dipolar couplings; (3) labeled RNAs that can be recycled by nuclease digestion followed by rephosphorylation to produce site-specifically labeled NTPs for making labeled RNAs; (4) a variety of multidimensional NMR experiments (e.g. 2D ^{1}H-^{15}N/^{1}H-^{13}C correlation spectra as well as 3D HNCO) can be applied using our labeling techniques to illuminate RNA function. I anticipate that this hybrid selective biomass and *in vivo* tRNA-scaffold approach for synthesizing RNAs will enable new types of RNA studies using multidimensional NMR technologies to better characterize the structural and dynamic basis of RNA function.

Appendix



Superimposition of ^{1}H-^{13}C TROSY spectra at the base region of the P1P2ΔL2 and the kl-TSS

The P1P2ΔL2 and kl-TSS were transcribed using site-specific labeled ribonucleotides, in which adenine and guanine were labeled at ^{13}C8-carbon; cytidine and uracil were labeled at ^{13}C6-carbon. The P1P2ΔL2 were transcribed with all four types of site-specific labeled ribonucleotides and its spectra.is presented in black color in panel A, B,C, and D. 2-D ^{13}C-^{1}H TROSY spectrum of the kl-TSS was made using ^{13}C8 adenine (A, blue), ^{13}C6 uracil (B, red), ^{13}C8 guanine (C, cyan), and ^{13}C8 cytidine(D, purple). The overlaid chemical shifts of the P1P2ΔL2 and the kl-TSS indicates that the P1, P2 and basal P3 stem in the P1P2ΔL2 maintains similar topology as those in the kl-TSS.

Bibliography

1. **Koonin EV, Dolja VV, Krupovic M.** 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology **479-480:**2-25.

2. **Nicholson BL, White KA.** 2014. Functional long-range RNA-RNA interactions in positive-strand RNA viruses. Nat Rev Microbiol **12:**493-504.

3. **Jackson RJ, Hellen CU, Pestova TV.** 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol **11:**113-127.

4. **Nicholson BL, White KA.** 2011. 3' Cap-independent translation enhancers of positive-strand RNA plant viruses. Curr Opin Virol **1:**373-380.

5. **Simon AE, Miller WA.** 2013. 3' cap-independent translation enhancers of plant viruses. Annu Rev Microbiol **67:**21-42.

6. **Simon AE.** 2015. 3'UTRs of carmoviruses. Virus Res **206:**27-36.

7. **Filbin ME, Kieft JS.** 2009. Toward a structural understanding of IRES RNA function. Curr Opin Struct Biol **19:**267-276.

8. **Balvay L, Soto Rifo R, Ricci EP, Decimo D, Ohlmann T.** 2009. Structural and functional diversity of viral IRESes. Biochim Biophys Acta **1789:**542-557.

9. **Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, Diaz-Toledano R.** 2015. Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res **206:**62-73.

10. **Merrick WC.** 2010. Eukaryotic protein synthesis: still a mystery. J Biol Chem **285:**21197-21201.

11. **Godefroycolburn T, Thivent C, Pinck L.** 1985. Translational discrimination between the four RNAs of alfalfa mosaic-virus - a quantitative-evaluation. Eur J Biochem **147:**541-548.

12. **Lawson TG, Ray BK, Dodds JT, Grifo JA, Abramson RD, Merrick WC, Betsch DF, Weith HL, Thach RE.** 1986. Influence of 5' proximal secondary structure on the translational efficiency of eukaryotic messenger-RNAS and on their interaction with initiation-factors. J Biol Chem **261:**13979-13989.

13. **Gray NK, Hentze MW.** 1994. Regulation of protein synthesis by mRNA structure. Mol Biol Rep **19:**195-200.

14. **Costello J, Castelli LM, Rowe W, Kershaw CJ, Talavera D, Mohammad-Qureshi SS, Sims PF, Grant CM, Pavitt GD, Hubbard SJ, Ashe MP.** 2015. Global mRNA selection mechanisms for translation initiation. Genome Biol **16:**10.

15. **Fechter P, Brownlee GG.** 2005. Recognition of mRNA cap structures by viral and cellular proteins. J Gen Virol **86:**1239-1249.

16. **Pfingsten JS, Costantino DA, Kieft JS.** 2006. Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. Science **314:**1450-1454.

17. **Dreher TW, Miller WA.** 2006. Translational control in positive strand RNA plant viruses. Virology **344:**185-197.

18. **Hellen CU, Sarnow P.** 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev **15:**1593-1612.

19. **Martínez-Salas E, Piñeiro D, Fernández N.** 2012. Alternative Mechanisms to Initiate Translation in Eukaryotic mRNAs. Comp Funct Genomics **2012:**391546.

20. **Kieft JS.** 2008. Viral IRES RNA structures and ribosome interactions. Trends Biochem Sci **33:**274-283.

21. **Martinez-Salas E, Pacheco A, Serrano P, Fernandez N.** 2008. New insights into internal ribosome entry site elements relevant for viral gene expression. J Gen Virol **89:**611-626.

22. **Stupina VA, Yuan X, Meskauskas A, Dinman JD, Simon AE.** 2011. Ribosome binding to a 5' translational enhancer is altered in the presence of the 3' untranslated region in cap-independent translation of turnip crinkle virus. J Virol **85:**4638-4653.

23. **Chattopadhyay M, Shi K, Yuan X, Simon AE.** 2011. Long-distance kissing loop interactions between a 3' proximal Y-shaped structure and apical loops of 5' hairpins enhance translation of Saguaro cactus virus. Virology **417:**113-125.

24. **Jang SK, Krausslich HG, Nicklin MJ, Duke GM, Palmenberg AC, Wimmer E.** 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J Virol **62:**2636-2643.

25. **Pelletier J, Sonenberg N.** 1988. Internal initiation of translation of eukaryotic messenger-RNA directed by a sequence derived from poliovirus RNA. Nature **334:**320-325.

26. **Baird SD, Turcotte M, Korneluk RG, Holcik M.** 2006. Searching for IRES. RNA **12:**1755-1785.

27. **Fitzgerald KD, Semler BL.** 2009. Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. Biochim Biophys Acta **1789:**518-528.

28. **Belsham GJ.** 2009. Divergent picornavirus IRES elements. Virus Res **139:**183-192.

29. **Hellen CUT, de Breyne S.** 2007. A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse Picornavirus genera: Evidence for modular exchange of functional noncoding RNA elements by recombination. J Virol **81:**5850-5863.

30. **Jan E.** 2006. Divergent IRES elements in invertebrates. Virus Res **119:**16-28.

31. **Wilson JE, Pestova TV, Hellen CU, Sarnow P.** 2000. Initiation of protein synthesis from the A site of the ribosome. Cell **102:**511-520.

32. **Jan E, Thompson SR, Wilson JE, Pestova TV, Hellen CU, Sarnow P.** 2001. Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. Cold Spring Harb Symp Quant Biol **66:**285-292.

33. **Costantino D, Kieft JS.** 2005. A preformed compact ribosome-binding domain in the cricket paralysis-like virus IRES RNAs. RNA **11:**332-343.

34. **Kanamori Y, Nakashima N.** 2001. A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. RNA **7:**266-274.

35. **Nishiyama T, Yamamoto H, Shibuya N, Hatakeyama Y, Hachimori A, Uchiumi T, Nakashima N.** 2003. Structural elements in the internal ribosome entry site of Plautia stali intestine virus responsible for binding with ribosomes. Nucleic Acids Res **31:**2434-2442.

36. **Jan E, Sarnow P.** 2002. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. J Mol Biol **324:**889-902.

37. **Spahn CM, Jan E, Mulder A, Grassucci RA, Sarnow P, Frank J.** 2004. Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. Cell **118:**465-475.

38. **Schuler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CM.** 2006. Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nat Struct Mol Biol **13:**1092-1096.

39. **Kamoshita N, Nomoto A, RajBhandary UL.** 2009. Translation initiation from the ribosomal A site or the P site, dependent on the conformation of RNA pseudoknot I in dicistrovirus RNAs. Mol Cell **35:**181-190.

40. **Sasaki J, Nakashima N.** 1999. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. J Virol **73:**1219-1226.

41. **Costantino DA, Pfingsten JS, Rambo RP, Kieft JS.** 2008. tRNA-mRNA mimicry drives translation initiation from a viral IRES. Nat Struct Mol Biol **15:**57-64.

42. **Brown EA, Zhang HC, Ping LH, Lemon SM.** 1992. Secondary structure of the 5' nontranslated regions of Hepatitis-C Virus and Pestivirus genomic RNAs. Nucleic Acids Res **20:**5041-5045.

43. **Wang CY, Sarnow P, Siddiqui A.** 1993. Translation of human Hepatitis-C Virus-RNA in cultured-cells is dediated by an internal ribosome-binding mechanism. J Virol **67:**3338-3344.

44. **Kieft JS, Zhou KH, Jubin R, Doudna JA.** 2001. Mechanism of ribosome recruitment by hepatitis CIRES RNA. RNA **7:**194-206.

45. **Wang C, Le SY, Ali N, Siddiqui A.** 1995. An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. RNA **1:**526-537.

46. **Otto GA, Lukavsky PJ, Lancaster AM, Sarnow P, Puglisi JD.** 2002. Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction. RNA **8:**913-923.

47. **Cornish PV, Ermolenko DN, Staple DW, Hoang L, Hickerson RP, Noller HF, Ha T.** 2009. Following movement of the L1 stalk between three functional states in single ribosomes. Proc Natl Acad Sci U S A **106:**2571-2576.

48. **Belsham GJ, Sonenberg N.** 2000. Picornavirus RNA translation: roles for cellular proteins. Trends Microbiol **8:**330-335.

49. **Kaminski A, Belsham GJ, Jackson RJ.** 1994. Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. EMBO J **13:**1673-1681.

50. **Danthinne X, Seurinck J, Meulewaeter F, Van Montagu M, Cornelissen M.** 1993. The 3' untranslated region of satellite tobacco necrosis virus RNA stimulates translation in vitro. Mol Cell Biol **13:**3340-3349.

51. **Gao F, Gulay SP, Kasprzak W, Dinman JD, Shapiro BA, Simon AE.** 2013. The kissing-loop T-shaped structure translational enhancer of Pea enation mosaic virus can bind simultaneously to ribosomes and a 5' proximal hairpin. J Virol **87:**11987-12002.

52. **Miller WA, Wang Z, Treder K.** 2007. The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. Biochem Soc Trans **35:**1629-1633.

53. **Scheets K, Redinbaugh MG.** 2006. Infectious cDNA transcripts of Maize necrotic streak virus: infectivity and translational characteristics. Virology **350:**171-183.

54. **Nieto C, Morales M, Orjeda G, Clepet C, Monfort A, Sturbois B, Puigdomenech P, Pitrat M, Caboche M, Dogimont C, Garcia-Mas J, Aranda MA, Bendahmane A.** 2006. An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. Plant J **48:**452-462.

55. **Nicholson BL, Wu B, Chevtchenko I, White KA.** 2010. Tombusvirus recruitment of host translational machinery via the 3' UTR. RNA **16:**1402-1419.

56. **Timmer RT, Benkowski LA, Schodin D, Lax SR, Metz AM, Ravel JM, Browning KS.** 1993. The 5' and 3' untranslated regions of satellite tobacco necrosis virus RNA affect translational efficiency and dependence on a 5' cap structure. J Biol Chem **268:**9504-9510.

57. **Gazo BM, Murphy P, Gatchel JR, Browning KS.** 2004. A novel interaction of cap-binding protein complexes eukaryotic initiation factor (eIF) 4F and eIF(iso)4F with a region in the 3 '-untranslated region of satellite tobacco necrosis virus. J Biol Chem **279:**13584-13592.

58. **Meulewaeter F, Danthinne X, Van Montagu M, Cornelissen M.** 1998. 5'- and 3'-sequences of satellite tobacco necrosis virus RNA promoting translation in tobacco. Plant J **14:**169-176.

59. **Wang Z, Treder K, Miller WA.** 2009. Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. J Biol Chem **284:**14189-14202.

60. **Wang Z, Parisien M, Scheets K, Miller WA.** 2011. The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. Structure **19:**868-880.

61. **Chattopadhyay M, Kuhlmann MM, Kumar K, Simon AE.** 2014. Position of the kissing-loop interaction associated with PTE-type 3'CITEs can affect enhancement of cap-independent translation. Virology **458-459:**43-52.

62. **Fabian MR, White KA.** 2004. 5'-3' RNA-RNA interaction facilitates cap- and poly(A) tail-independent translation of tomato bushy stunt virus mrna: a potential common mechanism for tombusviridae. J Biol Chem **279:**28862-28872.

63. **Nicholson BL, Zaslaver O, Mayberry LK, Browning KS, White KA.** 2013. Tombusvirus Y-shaped translational enhancer forms a complex with eIF4F and can be functionally replaced by heterologous translational enhancers. J Virol **87:**1872-1883.

64. **Fabian MR, White KA.** 2006. Analysis of a 3'-translation enhancer in a tombusvirus: a dynamic model for RNA-RNA interactions of mRNA termini. RNA **12:**1304-1314.

65. **Ray D, Na H, White KA.** 2004. Structural properties of a multifunctional T-shaped RNA domain that mediate efficient tomato bushy stunt virus RNA replication. J Virol **78:**10490-10500.

66. **Shen R, Miller WA.** 2004. The 3' untranslated region of tobacco necrosis virus RNA contains a barley yellow dwarf virus-like cap-independent translation element. J Virol **78:**4655-4664.

67. **Guo L, Allen E, Miller WA.** 2000. Structure and function of a cap-independent translation element that functions in either the 3' or the 5' untranslated region. RNA **6:**1808-1820.

68. **Meulewaeter F, van Lipzig R, Gultyaev AP, Pleij CW, Van Damme D, Cornelissen M, van Eldik G.** 2004. Conservation of RNA structures enables TNV and BYDV 5' and 3' elements to cooperate synergistically in cap-independent translation. Nucleic Acids Res **32:**1721-1730.

69. **Mizumoto H, Tatsuta M, Kaido M, Mise K, Okuno T.** 2003. Cap-independent translational enhancement by the 3' untranslated region of red clover necrotic mosaic virus RNA1. J Virol **77:**12113-12121.

70. **Wang Z, Kraft JJ, Hui AY, Miller WA.** 2010. Structural plasticity of Barley yellow dwarf virus-like cap-independent translation elements in four genera of plant viral RNAs. Virology **402:**177-186.

71. **Kraft JJ, Treder K, Peterson MS, Miller WA.** 2013. Cation-dependent folding of 3' cap-independent translation elements facilitates interaction of a 17-nucleotide conserved sequence with eIF4G. Nucleic Acids Res **41:**3398-3413.

72. **Guo L, Allen EM, Miller WA.** 2001. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. Mol Cell **7:**1103-1109.

73. **Treder K, Kneller EL, Allen EM, Wang Z, Browning KS, Miller WA.** 2008. The 3' cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the eIF4G subunit to initiate translation. RNA **14:**134-147.

74. **McCormack JC, Yuan X, Yingling YG, Kasprzak W, Zamora RE, Shapiro BA, Simon AE.** 2008. Structural domains within the 3' untranslated region of Turnip crinkle virus. J Virol **82:**8706-8720.

75. **Stupina VA, Meskauskas A, McCormack JC, Yingling YG, Shapiro BA, Dinman JD, Simon AE.** 2008. The 3' proximal translational enhancer of Turnip crinkle virus binds to 60S ribosomal subunits. RNA **14:**2379-2393.

76. **Zuo X, Wang J, Yu P, Eyler D, Xu H, Starich MR, Tiede DM, Simon AE, Kasprzak W, Schwieters CD, Shapiro BA, Wang YX.** 2010. Solution structure of the cap-independent translational enhancer and ribosome-binding element in the 3' UTR of Turnip crinkle virus. Proc Natl Acad Sci U S A **107:**1385-1390.

77. **Gao F, Kasprzak W, Stupina VA, Shapiro BA, Simon AE.** 2012. A ribosome-binding, 3' translational enhancer has a T-shaped structure and engages in a long-distance RNA-RNA interaction. J Virol **86:**9828-9842.

78. **Hacker DL, Petty IT, Wei N, Morris TJ.** 1992. Turnip crinkle virus genes required for RNA replication and virus movement. Virology **186:**1-8.

79. **Cao M, Ye X, Willie K, Lin J, Zhang X, Redinbaugh MG, Simon AE, Morris TJ, Qu F.** 2010. The capsid protein of Turnip crinkle virus overcomes two separate defense barriers to facilitate systemic movement of the virus in Arabidopsis. J Virol **84:**7793-7802.

80. **Sun X, Simon AE.** 2006. A cis-replication element functions in both orientations to enhance replication of Turnip crinkle virus. Virology **352:**39-51.

81. **Yuan X, Shi K, Meskauskas A, Simon AE.** 2009. The 3' end of Turnip crinkle virus contains a highly interactive structure including a translational enhancer that is disrupted by binding to the RNA-dependent RNA polymerase. RNA **15:**1849-1864.

82. **Yuan X, Shi K, Young MY, Simon AE.** 2010. The terminal loop of a 3' proximal hairpin plays a critical role in replication and the structure of the 3' region of Turnip crinkle virus. Virology **402:**271-280.

83. **Yuan X, Shi K, Simon AE.** 2012. A local, interactive network of 3' RNA elements supports translation and replication of Turnip crinkle virus. J Virol **86:**4065-4081.

84. **Qu F, Morris TJ.** 2000. Cap-independent translational enhancement of turnip crinkle virus genomic and subgenomic RNAs. J Virol **74:**1085-1093.

85. **Zhang G, Zhang J, Simon AE.** 2004. Repression and derepression of minus-strand synthesis in a plus-strand RNA virus replicon. J Virol **78:**7619-7633.

86. **Song C, Simon AE.** 1995. Requirement of a 3'-terminal stem-loop in in vitro transcription by an RNA-dependent RNA polymerase. J Mol Biol **254:**6-14.

87. **Guan H, Song C, Simon AE.** 1997. RNA promoters located on (-)-strands of a subviral RNA associated with turnip crinkle virus. RNA **3:**1401-1412.

88. **Simon AE, Howell SH.** 1986. The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3' end of the helper virus genome. EMBO J **5:**3423-3428.

89. **Guan H, Carpenter CD, Simon AE.** 2000. Analysis of cis-acting sequences involved in plus-strand synthesis of a turnip crinkle virus-associated satellite RNA identifies a new carmovirus replication element. Virology **268:**345-354.

90. **Guan H, Simon AE.** 2000. Polymerization of nontemplate bases before transcription initiation at the 3' ends of templates by an RNA-dependent RNA polymerase: an activity involved in 3' end repair of viral RNAs. Proc Natl Acad Sci U S A **97:**12451-12456.

91. **Cimino PA, Nicholson BL, Wu B, Xu W, White KA.** 2011. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. PLoS Pathog **7:**e1002423.

92. **Nagy PD, Pogany J, Simon AE.** 2001. In vivo and in vitro characterization of an RNA replication enhancer in a satellite RNA associated with turnip crinkle virus. Virology **288:**315-324.

93. **Zhang G, Zhang J, George AT, Baumstark T, Simon AE.** 2006. Conformational changes involved in initiation of minus-strand synthesis of a virus-associated RNA. RNA **12:**147-162.

94. **Chattopadhyay M, Stupina VA, Gao F, Szarko CR, Kuhlmann MM, Yuan X, Shi K, Simon AE.** 2015. Requirement for host RNA-silencing components and the virus-silencing suppressor when second-site mutations compensate for structural defects in the 3' untranslated region. J Virol **89:**11603-11618.

95. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31:**3406-3415.

96. **Zhang J, Stuntz RM, Simon AE.** 2004. Analysis of a viral replication repressor: sequence requirements for a large symmetrical internal loop. Virology **326:**90-102.

97. **Zhang J, Zhang G, Guo R, Shapiro BA, Simon AE.** 2006. A pseudoknot in a preactive form of a viral RNA is part of a structural switch activating minus-strand synthesis. J Virol **80:**9181-9191.

98. **McCormack JC, Simon AE.** 2004. Biased hypermutagenesis associated with mutations in an untranslated hairpin of an RNA virus. J Virol **78:**7813-7817.

99. **Guo R, Lin W, Zhang J, Simon AE, Kushner DB.** 2009. Structural plasticity and rapid evolution in a viral RNA revealed by in vivo genetic selection. J Virol **83:**927-939.

100. **Zhang J, Zhang G, McCormack JC, Simon AE.** 2006. Evolution of virus-derived sequences for high-level replication of a subviral RNA. Virology **351:**476-488.

101. **Demler SA, Rucker DG, de Zoeten GA.** 1993. The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. J Gen Virol **74 (Pt 1):**1-14.

102. **Gao F, Kasprzak WK, Szarko C, Shapiro BA, Simon AE.** 2014. The 3' untranslated region of Pea Enation Mosaic Virus contains two T-shaped, ribosome-binding, cap-independent translation enhancers. J Virol **88:**11696-11712.

103. **Ryabov EV, Robinson DJ, Taliansky M.** 2001. Umbravirus-encoded proteins both stabilize heterologous viral RNA and mediate its systemic movement in some plant species. Virology **288:**391-400.

104. **Steitz TA.** 2008. A structural understanding of the dynamic ribosome machine. Nat Rev Mol Cell Biol **9:**242-253.

105. **Serganov A, Nudler E.** 2013. A decade of riboswitches. Cell **152:**17-24.

106. **Moore PB.** 1999. Structural motifs in RNA. Annu Rev Biochem **68:**287-300.

107. **Holbrook SR.** 2005. RNA structure: the long and the short of it. Curr Opin Struct Biol **15:**302-308.

108. **Woodson SA.** 2010. Compact intermediates in RNA folding. Annu Rev Biophys **39:**61-77.

109. **Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA.** 2007. Thermodynamic analysis of interacting nucleic acid strands. Siam Review **49:**65-88.

110. **Li PTX, Vieregg J, Tinoco I.** 2008. How RNA unfolds and refolds. Annu Rev Biochem **77:**77-100.

111. **Draper DE.** 2004. A guide to ions and RNA structure. RNA **10:**335-343.

112. **Roh JH, Guo L, Kilburn JD, Briber RM, Irving T, Woodson SA.** 2010. Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering. J Am Chem Soc **132:**10148-10154.

113. **Wu M, Tinoco I, Jr.** 1998. RNA folding causes secondary structure rearrangement. Proc Natl Acad Sci U S A **95:**11555-11560.

114. **Chauhan S, Caliskan G, Briber RM, Perez-Salas U, Rangan P, Thirumalai D, Woodson SA.** 2005. RNA tertiary interactions mediate native collapse of a bacterial group I ribozyme. J Mol Biol **353:**1199-1209.

115. **Behrouzi R, Roh JH, Kilburn D, Briber RM, Woodson SA.** 2012. Cooperative tertiary interaction network guides RNA folding. Cell **149:**348-357.

116. **Woodson SA.** 2011. Pathways of RNA folding and ribosome assembly from time-resolved footprinting. Acc Chem Res **44:**1312-1319.

117. **Chen B, Zuo X, Wang YX, Dayie TK.** 2012. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. Nucleic Acids Res **40:**3117-3130.

118. **Rieder R, Lang K, Graber D, Micura R.** 2007. Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. Chembiochem **8:**896-902.

119. **Matsuda D, Yoshinari S, Dreher TW.** 2004. eEF1A binding to aminoacylated viral RNA represses minus strand synthesis by TYMV RNA-dependent RNA polymerase. Virology **321:**47-56.

120. **Dreher TW.** 2009. Role of tRNA-like structures in controlling plant virus replication. Virus Res **139:**217-229.

121. **Patel N, Dykeman EC, Coutts RHA, Lomonossoff GP, Rowlands DJ, Phillips SEV, Ranson N, Twarock R, Tuma R, Stockley PG.** 2015. Revealing the density of encoded functions in a viral RNA. Proc Natl Acad Sci USA **112:**2227-2232.

122. **D'Souza V, Summers MF.** 2004. Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. Nature **431:**586-590.

123. **Dethoff EA, Chugh J, Mustoe AM, Al-Hashimi HM.** 2012. Functional complexity and regulation through RNA dynamics. Nature **482:**322-330.

124. **Schroeder R, Barta A, Semrad K.** 2004. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol **5:**908-919.

125. **Neuman KC, Nagy A.** 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods **5:**491-505.

126. **Greenleaf WJ, Woodside MT, Block SM.** 2007. High-resolution, single-molecule measurements of biomolecular motion. Annu Rev Biophys Biomol Struct **36:**171-190.

127. **Moffitt JR, Chemla YR, Smith SB, Bustamante C.** 2008. Recent advances in optical tweezers. Annu Rev Biochem **77:**205-228.

128. **Li PT, Bustamante C, Tinoco I, Jr.** 2006. Unusual mechanical stability of a minimal RNA kissing complex. Proc Natl Acad Sci USA **103:**15847-15852.

129. **Liphardt J, Onoa B, Smith SB, Tinoco I, Jr., Bustamante C.** 2001. Reversible unfolding of single RNA molecules by mechanical force. Science **292:**733-737.

130. **Tinoco I, Jr., Li PT, Bustamante C.** 2006. Determination of thermodynamics and kinetics of RNA reactions by force. Q Rev Biophys **39:**325-360.

131. **Tinoco I, Chen G, Qu X.** 2010. RNA reactions one molecule at a time. Cold Spring Harb Perspect Biol **2:**a003624.

132. **Onoa B, Dumont S, Liphardt J, Smith SB, Tinoco I, Jr., Bustamante C.** 2003. Identifying kinetic barriers to mechanical unfolding of the *T. thermophila* ribozyme. Science **299:**1892-1895.

133. **Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM.** 2008. Direct observation of hierarchical folding in single riboswitch aptamers. Science **319:**630-633.

134. **Green L, Kim CH, Bustamante C, Tinoco I, Jr.** 2008. Characterization of the mechanical unfolding of RNA pseudoknots. J Mol Biol **375:**511-528.

135. **Wu YJ, Wu CH, Yeh AY, Wen JD.** 2014. Folding a stable RNA pseudoknot through rearrangement of two hairpin structures. Nucleic Acids Res **42:**4505-4515.

136. **Diez S, Reuther C, Dinu C, Seidel R, Mertig M, Pompe W, Howard J.** 2003. Stretching and transporting DNA molecules using motor proteins. Nano Letters **3:**1251-1254.

137. **Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ.** 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res **117:**17-37.

138. **Chen G, Chang KY, Chou MY, Bustamante C, Tinoco I, Jr.** 2009. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. Proc Natl Acad Sci U S A **106:**12706-12711.

139. **Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, Block SM.** 2006. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. Science **314:**1001-1004.

140. **Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I, Jr., Bustamante C.** 2005. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. Nature **437:**231-234.

141. **Sattin BD, Zhao W, Travers K, Chu S, Herschlag D.** 2008. Direct measurement of tertiary contact cooperativity in RNA folding. J Am Chem Soc **130:**6085-6087.

142. **Wen JD, Manosas M, Li PT, Smith SB, Bustamante C, Ritort F, Tinoco I, Jr.** 2007. Force unfolding kinetics of RNA using optical tweezers. I. Effects of experimental variables on measured results. Biophys J **92:**2996-3009.

143. **Lynch SR, Puglisi JD.** 2001. Structure of a eukaryotic decoding region A-site RNA. J Mol Biol **306:**1023-1035.

144. **Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson BA, Salemi M, Telesnitsky A, Summers MF.** 2015. RNA structure. Structure of the HIV-1 RNA packaging signal. Science **348:**917-921.

145. **Noeske J, Richter C, Grundl MA, Nasiri HR, Schwalbe H, Wohnert J.** 2005. An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. Proc Natl Acad Sci U S A **102:**1372-1377.

146. **Furtig B, Richter C, Wohnert J, Schwalbe H.** 2003. NMR spectroscopy of RNA. Chembiochem **4:**936-962.

147. **Puglisi EV, Puglisi JD.** 2007. Probing the conformation of human $tRNA_3^{Lys}$ in solution by NMR. FEBS Lett **581:**5307-5314.

148. **Getz M, Sun X, Casiano-Negroni A, Zhang Q, Al-Hashimi HM.** 2007. NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings. Biopolymers **86:**384-402.

149. **Latham MP, Brown DJ, McCallum SA, Pardi A.** 2005. NMR methods for studying the structure and dynamics of RNA. Chembiochem **6:**1492-1505.

150. **Dayie KT.** 2008. Key labeling technologies to tackle sizeable problems in RNA structural biology. Int J Mol Sci **9:**1214-1240.

151. **Alvarado LJ, Longhini AP, LeBlanc RM, Chen B, Kreutz C, Dayie TK.** 2014. Chemo-enzymatic synthesis of selectively $^{13}C/^{15}N$-labeled RNA for NMR structural and dynamics studies. Methods Enzymol **549:**133-162.

152. **Thakur CS, Dayie TK.** 2011. Asymmetry of ^{13}C labeled 3-pyruvate affords improved site specific labeling of RNA for NMR spectroscopy. J Biomol NMR **52:**65-77.

153. **Longhini AP, LeBlanc RM, Becette O, Salguero C, Wunderlich CH, Johnson BA, D'Souza VM, Kreutz C, Dayie TK.** 2015. Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. Nucleic Acids Res doi:10.1093/nar/gkv1333.

154. **Le MT, Brown RE, Simon AE, Dayie TK.** 2015. In vivo, large-scale preparation of uniformly ^{15}N- and site-specifically ^{13}C-labeled homogeneous, recombinant RNA for NMR studies. Methods Enzymol **565:**495-535.

155. **Fang X, Stagno JR, Bhandari YR, Zuo X, Wang YX.** 2015. Small-angle X-ray scattering: a bridge between RNA secondary structures and three-dimensional topological structures. Curr Opin Struct Biol **30:**147-160.

156. **Putnam CD, Hammel M, Hura GL, Tainer JA.** 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys **40:**191-285.

157. **Reyes FE, Schwartz CR, Tainer JA, Rambo RP.** 2014. Chapter eleven - methods for using new conceptual tools and parameters to assess RNA structure by small-angle X-ray scattering, p 235-263. *In* Donald HB-A (ed), Methods Enzymol, vol Volume 549. Academic Press.

158. **Baird NJ, Ferre-D'Amare AR.** 2010. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. RNA **16:**598-609.

159. **Gopal A, Zhou ZH, Knobler CM, Gelbart WM.** 2012. Visualizing large RNA molecules in solution. RNA **18:**284-299.

160. **Jenkins JL, Krucinska J, McCarty RM, Bandarian V, Wedekind JE.** 2011. Comparison of a preQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. J Biol Chem **286:**24626-24637.

161. **Serganov A, Patel DJ.** 2007. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet **8:**776-790.

162. **Simon AE, Gehrke L.** 2009. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. Biochim Biophys Acta **1789:**571-583.

163. **Winkler WC, Breaker RR.** 2003. Genetic control by metabolite-binding riboswitches. Chembiochem **4:**1024-1032.

164. **Soukup JK, Soukup GA.** 2004. Riboswitches exert genetic control through metabolite-induced conformational change. Curr Opin Struct Biol **14:**344-349.

165. **Huthoff H, Berkhout B.** 2001. Two alternating structures of the HIV-1 leader RNA. RNA **7:**143-157.

166. **Stammler SN, Cao S, Chen SJ, Giedroc DP.** 2011. A conserved RNA pseudoknot in a putative molecular switch domain of the 3'-untranslated region of coronaviruses is only marginally stable. RNA **17:**1747-1759.

167. **Guo R, Meskauskas A, Dinman JD, Simon AE.** 2011. Evolution of a helper virus-derived, ribosome binding translational enhancer in an untranslated satellite RNA of Turnip crinkle virus. Virology **419:**10-16.

168. **Chen G, Wen JD, Tinoco I, Jr.** 2007. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. RNA **13:**2175-2188.

169. **Cromie MJ, Shi Y, Latifi T, Groisman EA.** 2006. An RNA sensor for intracellular Mg^{2+}. Cell **125:**71-84.

170. **Garcia-Sacristan A, Moreno M, Ariza-Mateos A, Lopez-Camacho E, Jaudenes RM, Vazquez L, Gomez J, Martin-Gago JA, Briones C.** 2015. A magnesium-induced RNA conformational switch at the internal ribosome entry site of hepatitis C virus genome visualized by atomic force microscopy. Nucleic Acids Res **43:**565-580.

171. **Reining A, Nozinovic S, Schlepckow K, Buhr F, Furtig B, Schwalbe H.** 2013. Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature **499:**355-359.

172. **Pyle AM.** 2002. Metal ions in the structure and function of RNA. J Biol Inorg Chem **7:**679-690.

173. **Tan ZJ, Chen SJ.** 2011. Salt contribution to RNA tertiary structure folding stability. Biophys J **101:**176-187.

174. **Woodson SA.** 2005. Metal ions and RNA folding: a highly charged topic with a dynamic future. Curr Opin Chem Biol **9:**104-109.

175. **Butcher SE, Pyle AM.** 2011. The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. Acc Chem Res **44:**1302-1311.

176. **Haller A, Rieder U, Aigner M, Blanchard SC, Micura R.** 2011. Conformational capture of the SAM-II riboswitch. Nat Chem Biol **7:**393-400.

177. **Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA.** 2006. Folding of the adenine riboswitch. Chem Biol **13:**857-868.

178. **Cate JH, Hanna RL, Doudna JA.** 1997. A magnesium ion core at the heart of a ribozyme domain. Nat Struct Biol **4:**553-558.

179. **Hammond JA, Rambo RP, Filbin ME, Kieft JS.** 2009. Comparison and functional implications of the 3D architectures of viral tRNA-like structures. RNA **15:**294-307.

180. **Klein DJ, Edwards TE, Ferre-D'Amare AR.** 2009. Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. Nat Struct Mol Biol **16:**343-344.

181. **Chen SC, Olsthoorn RC.** 2010. In vitro and in vivo studies of the RNA conformational switch in Alfalfa mosaic virus. J Virol **84:**1423-1429.

182. **Zheng L, Baumann U, Reymond JL.** 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res **32:**e115.

183. **Smith SB, Cui Y, Bustamante C.** 2003. Optical-trap force transducer that operates by direct measurement of light momentum. Methods Enzymol **361:**134-162.

184. **Wang MD, Yin H, Landick R, Gelles J, Block SM.** 1997. Stretching DNA with optical tweezers. Biophys J **72:**1335-1346.

185. **Vilfan ID, Kamping W, van den Hout M, Candelli A, Hage S, Dekker NH.** 2007. An RNA toolbox for single-molecule force spectroscopy studies. Nucleic Acids Res **35:**6625-6639.

186. **Abbink TE, Ooms M, Haasnoot PC, Berkhout B.** 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry **44:**9058-9066.

187. **Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA.** 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc Natl Acad Sci U S A **98:**4899-4903.

188. **Ferre-D'Amare AR, Doudna JA.** 1999. RNA folds: insights from recent crystal structures. Annu Rev Biophys Biomol Struct **28:**57-73.

189. **Choudhary PK, Sigel RK.** 2014. Mg^{2+}-induced conformational changes in the btuB riboswitch from E. coli. RNA **20:**36-45.

190. **Suddala KC, Wang J, Hou Q, Walter NG.** 2015. Mg^{2+} shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. J Am Chem Soc **137:**14075-14083.

191. **Kneller EL, Rakotondrafara AM, Miller WA.** 2006. Cap-independent translation of plant viral RNAs. Virus Res **119:**63-75.

192. **Batten JS, Desvoyes B, Yamamura Y, Scholthof KB.** 2006. A translational enhancer element on the 3'-proximal end of the Panicum mosaic virus genome. FEBS Lett **580:**2591-2597.

193. **Wang YX, Zuo X, Wang J, Yu P, Butcher SE.** 2010. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. Methods **52:**180-191.

194. **Parisien M, Major F.** 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature **452:**51-55.

195. **Schneidman-Duhovny D, Hammel M, Sali A.** 2010. FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res **38:**W540-544.

196. **Berlin K, Castaneda CA, Schneidman-Duhovny D, Sali A, Nava-Tudela A, Fushman D.** 2013. Recovering a representative conformational ensemble from underdetermined macromolecular structural data. J Am Chem Soc **135:**16595-16609.

197. **Johnson BA, Blevins RA.** 1994. NMR View: A computer program for the visualization and analysis of NMR data. J Biomol NMR **4:**603-614.

198. **Lukavsky PJ, Otto GA, Lancaster AM, Sarnow P, Puglisi JD.** 2000. Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. Nat Struct Biol **7:**1105-1110.

199. **Zuo X, Cui G, Merz KM, Jr., Zhang L, Lewis FD, Tiede DM.** 2006. X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. Proc Natl Acad Sci U S A **103:**3534-3539.

200. **Jacrot B.** 1976. The study of biological structures by neutron scattering from solution. Rep Prog Phys **39:**911.

201. **Svergun DI.** 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). Biophysical Journal **77:**2896-2896.

202. **Dethoff EA, Petzold K, Chugh J, Casiano-Negroni A, Al-Hashimi HM.** 2012. Visualizing transient low-populated structures of RNA. Nature **491:**724-728.

203. **Bouvignies G, Kay LE.** 2012. A 2D ^{13}C-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding. J Biomol NMR **53:**303-310.

204. **Bouvignies G, Kay LE.** 2012. A 2D C-13-CEST experiment for studying slowly exchanging protein systems using methyl probes: an application to protein folding. J Biomol NMR **53:**303-310.

205. **Fawzi NL, Ying JF, Ghirlando R, Torchia DA, Clore GM.** 2011. Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. Nature **480:**268-272.

206. **Forsen S, Hoffman RA.** 1963. Study of moderately rapid chemical exchange reactions by means of nuclear magnetic double resonance. J Chem Phys **39:**2892-2901.

207. **Vallurupalli P, Bouvignies G, Kay LE.** 2012. Studying "invisible" excited protein states in slow exchange with a major state conformation. J Am Chem Soc **134:**8148-8161.

208. **Zhao B, Hansen AL, Zhang Q.** 2014. Characterizing slow chemical exchange in nucleic acids by carbon CEST and low spin-lock field R-1 rho NMR spectroscopy. J Am Chem Soc **136:**20-23.

209. **Palmer AG, 3rd, Grey MJ, Wang C.** 2005. Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. Methods Enzymol **394:**430-465.

210. **Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR.** 1992. Preparation of isotopically labeled ribonucleotides for multidimensional NMR-spectroscopy of RNA. Nucleic Acids Res **20:**4515-4523.

211. **Holland DWHaJA.** 1995. Preparation of ^{13}C-labeled ribonucleotides using acetate as istopic source. Nucleic Acids Res **23:**3361-3362.

212. **Johnson JE, Jr., Julien KR, Hoogstraten CG.** 2006. Alternate-site isotopic labeling of ribonucleotides for NMR studies of ribose conformational dynamics in RNA. J Biomol NMR **35:**261-274.

213. **Nikonowicz EP, Sirr A, Legault P, Jucker FM, Baer LM, Pardi A.** 1992. Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimensional NMR studies. Nucleic Acids Res **20:**4507-4513.

214. **Thakur CS, Luo Y, Chen B, Eldho NV, Dayie TK.** 2012. Biomass production of site selective ^{13}C/^{15}N nucleotides using wild type and a transketolase *E. coli* mutant for labeling RNA for high resolution NMR. J Biomol NMR **52:**103-114.

215. **Thakur CS, Brown ME, Sama JN, Jackson ME, Dayie TK.** 2010. Growth of wildtype and mutant *E. coli* strains in minimal media for optimal production of nucleic acids for preparing labeled nucleotides. Appl Microbiol Biotechnol **88:**771-779.

216. **Thakur CS, Sama JN, Jackson ME, Chen B, Dayie TK.** 2010. Selective ^{13}C labeling of nucleotides for large RNA NMR spectroscopy using an *E. coli* strain disabled in the TCA cycle. J Biomol NMR **48:**179-192.

217. **Dayie TK, Thakur CS.** 2010. Site-specific labeling of nucleotides for making RNA for high resolution NMR studies using an *E. coli* strain disabled in the oxidative pentose phosphate pathway. J Biomol NMR **47:**19-31.

218. **Tolbert TJ, Williamson JR.** 1996. Preparation of specifically deuterated RNA for NMR studies using a combination of chemical and enzymatic synthesis. J Am Chem Soc **118:**7929-7940.

219. **Milecki J.** 2002. Specific labelling of nucleosides and nucleotides with ^{13}C and ^{15}N. J Labelled Comp Radiopharm **45:**307-337.

220. **Schultheisz HL, Szymczyna BR, Scott LG, Williamson JR.** 2011. Enzymatic de novo pyrimidine nucleotide synthesis. J Am Chem Soc **133:**297-304.

221. **SantaLucia J, Jr., Shen LX, Cai Z, Lewis H, Tinoco I, Jr.** 1995. Synthesis and NMR of RNA with selective isotopic enrichment in the bases. Nucleic Acids Res **23:**4913-4921.

222. **Scott LG, Tolbert TJ, Williamson JR.** 2000. Preparation of specifically ^{2}H- and ^{13}C-labeled ribonucleotides. Methods Enzymol **317:**18-38.

223. **Wunderlich CH, Spitzer R, Santner T, Fauster K, Tollinger M, Kreutz C.** 2012. Synthesis of $^{13}C6$ pyrimidine nucleotides as spin-labels for RNA dynamics. J Am Chem Soc **134:**7558-7569.

224. **Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC.** 1987. Oligoribonucleotide synthesis using T7 RNA-polymerase and synthetic DNA templates. Nucleic Acids Res **15:**8783-8798.

225. **Pleiss JA, Derrick ML, Uhlenbeck OC.** 1998. T7 RNA polymerase produces 5 ' end heterogeneity during in vitro transcription from certain templates. RNA **4:**1313-1317.

226. **Ponchon L, Dardel F.** 2007. Recombinant RNA technology: the tRNA scaffold. Nat Methods **4:**571-576.

227. **Ponchon L, Dardel F.** 2011. Large scale expression and purification of recombinant RNA in *Escherichia coli*. Methods **54:**267-273.

228. **Ponchon L, Beauvais G, Nonin-Lecomte S, Dardel F.** 2009. A generic protocol for the expression and purification of recombinant RNA in *Escherichia coli* using a tRNA scaffold. Nat Protoc **4:**947-959.

229. **Liu Y, Stepanov VG, Strych U, Willson RC, Jackson GW, Fox GE.** 2010. DNAzyme-mediated recovery of small recombinant RNAs from a 5S rRNA-derived chimera expressed in *Escherichia coli*. BMC Biotechnol **10:**85.

230. **Nelissen FH, Leunissen EH, van de Laar L, Tessari M, Heus HA, Wijmenga SS.** 2012. Fast production of homogeneous recombinant RNA-towards large-scale production of RNA. Nucleic Acids Res **40:**e102.

231. **Josephson BL, Fraenkel DG.** 1974. Sugar metabolism in transketolase mutants of *Escherichia coli*. J Bacteriol **118:**1082-1089.

232. **Josephson BL, Fraenkel DG.** 1969. Transketolase mutants of *Escherichia coli*. J Bacteriol **100:**1289-1295.

233. **Ren A, Rajashankar KR, Patel DJ.** 2012. Fluoride ion encapsulation by Mg^{2+} ions and phosphates in a fluoride riboswitch. Nature **486:**85-89.

234. **Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR.** 2012. Widespread genetic switches and toxicity resistance proteins for fluoride. Science **335:**233-235.

235. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:**2006 0008.

236. **Studier FW.** 2005. Protein production by auto-induction in high-density shaking cultures. Protein Express Purif **41:**207-234.

237. **LeMaster DM, Richards FM.** 1982. Preparative-scale isolation of isotopically labeled amino acids. Anal Biochem **122:**238-247.

238. **Paliy O, Bloor D, Brockwell D, Gilbert P, Barber J.** 2003. Improved methods of cultivation and production of deuteriated proteins from *E-coli* strains grown on fully deuteriated minimal medium. J Appl Microbiol **94:**580-586.

239. **Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q.** 2009. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. Protein Sci **18:**936-948.

240. **Emilsson V, Naslund AK, Kurland CG.** 1993. Growth-rate-dependent accumulation of twelve tRNA species in *Escherichia coli*. J Mol Biol **230:**483-491.

241. **Juergensmeyer MA, Nelson ES, Juergensmeyer EA.** 2007. Shaking alone, without concurrent aeration, affects the growth characteristics of *Escherichia coli*. Lett Appl Microbiol **45:**179-183.

242. **Schlosser K, Gu J, Sule L, Li Y.** 2008. Sequence-function relationships provide new insight into the cleavage site selectivity of the 8-17 RNA-cleaving deoxyribozyme. Nucleic Acids Res **36:**1472-1481.

243. **Crooke ST, Lemonidis KM, Neilson L, Griffey R, Lesnik EA, Monia BP.** 1995. Kinetic characteristics of *Escherichia coli* RNase H1: cleavage of various antisense oligonucleotide-RNA duplexes. Biochem J **312 (Pt 2):**599-608.

244. **Duss O, Maris C, von Schroetter C, Allain FH.** 2010. A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. Nucleic Acids Res **38:**e188.

245. **Inoue H, Hayase Y, Iwai S, Ohtsuka E.** 1987. Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. FEBS Lett **215:**327-330.

246. **Santoro SW, Joyce GF.** 1997. A general purpose RNA-cleaving DNA enzyme. Proc Natl Acad Sci USA **94:**4262-4266.

247. **Tisne C, Roques BP, Dardel F.** 2001. Heteronuclear NMR studies of the interaction of tRNA$(Lys)_3$ with HIV-1 nucleocapsid protein. J Mol Biol **306:**443-454.

248. **Dayie KT, Brodsky AS, Williamson JR.** 2002. Base flexibility in HIV-2 TAR RNA mapped by solution ^{15}N, ^{13}C NMR relaxation. J Mol Biol **317:**263-278.

249. **Grzesiek S, Bax A.** 1992. Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. J Magn Reson **96:**432-440.

250. **Schleucher J, Sattler M, Griesinger C.** 1993. Coherence selection by gradients without signal attentuation-application to the 3-dimentional HNCO experiment. Angew Chem Int Ed Engl **32:**1489-1491.

251. **Kay LE, Xu GY, Yamazaki T.** 1994. Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. J Magn Reson **109:**129-133.

252. **Gilbert SD, Rambo RP, Van Tyne D, Batey RT.** 2008. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol **15:**177-182.

253. **Hansen AL, Bouvignies G, Kay LE.** 2013. Probing slowly exchanging protein systems via ^{13}Calpha-CEST: monitoring folding of the Im7 protein. J Biomol NMR **55:**279-289.

254. **Vallurupalli P, Kay LE.** 2013. Probing Slow Chemical Exchange at Carbonyl Sites in Proteins by Chemical Exchange Saturation Transfer NMR Spectroscopy. Angew Chem Int Ed Engl **52:**4156-4159.

255. **Zhou Y, Yang D.** 2014. Effects of J couplings and unobservable minor states on kinetics parameters extracted from CEST data. J Magn Reson **249:**118-125.

256. **Bouvignies G, Vallurupalli P, Kay LE.** 2014. Visualizing side chains of invisible protein conformers by solution NMR. J Mol Biol **426:**763-774.

257. **Zhou Y, Yang D.** 2015. ^{13}C-(alpha) CEST experiment on uniformly ^{13}C-labeled proteins. J Biomol NMR **61:**89-94.

258. **Alvarado LJ, LeBlanc RM, Longhini AP, Keane SC, Jain N, Yildiz ZF, Tolbert BS, D'Souza VM, Summers MF, Kreutz C, Dayie TK.** 2014. Regio-selective chemical-enzymatic synthesis of pyrimidine nucleotides facilitates RNA structure and dynamics studies. Chembiochem **15:**1573-1577.

259. **LeMaster DM, Kushlan DM.** 1996. Dynamical mapping of E-coli thioredoxin via ^{13}C- NMR relaxation analysis. J Am Chem Soc **118:**9255-9264.

260. **Takamoto K, He Q, Morris S, Chance MR, Brenowitz M.** 2002. Monovalent cations mediate formation of native tertiary structure of the Tetrahymena thermophila ribozyme. Nat Struct Biol **9:**928-933.

261. **Dreher TW.** 2010. Viral tRNAs and tRNA-like structures. Wiley Interdiscip Rev RNA **1:**402-414.

262. **Colussi TM, Costantino DA, Hammond JA, Ruehle GM, Nix JC, Kieft JS.** 2014. The structural basis of transfer RNA mimicry and conformational plasticity by a viral RNA. Nature **511:**366-369.