STRUCTURAL AND FUNCTIONAL INVESTIGATIONS OF CONFORMATIONALLY INTERCONVERTING RNA PSEUDOKNOTS

A Dissertation

by

SUZANNE NICOLE STAMMLER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Chemistry

STRUCTURAL AND FUNCTIONAL INVESTIGATIONS OF CONFORMATIONALLY INTERCONVERTING RNA PSEUDOKNOTS

A Dissertation

by

SUZANNE NICOLE STAMMLER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	David P. Giedroc
	Arthur E. Johnson
Committee Members,	Paul S. Cremer
	Paul Fitzpatrick
Head of Department,	David H. Russell

August 2009

Major Subject: Chemistry

ABSTRACT

Structural and Functional Investigations of Conformationally Interconverting RNA Pseudoknots. (August 2009)

Suzanne Nicole Stammler, B.S., California State University, Stanislaus

Co-Chairs of Advisory Committee: Dr. David P. Giedroc
Dr. Arthur E. Johnson

The biological function of RNA is often linked to an ability to adopt one or more mutually exclusive conformational states or isomers, a characteristic that distinguishes this biomolecule from proteins. Two examples of conformationally inconverting RNAs were structurally investigated. The first is found in the 3' untranslated region (UTR) of the coronavirus mouse hepatitis virus (MHV). A proposed molecular switch between mutually exclusive stable stem loop and pseudoknot conformations was investigated using thermal unfolding methods, NMR spectroscopy, sedimentation velocity ultracentrifugation and fluorescence resonance energy transfer (FRET) spectroscopy. Utilizing a "divide and conquer" approach we establish that the independent subdomains are folded as predicted by the proposed model and that a pseudoknotted conformation is accessible. Using the subdomains as spectral markers for the investigation of the intact 3' UTR RNA, we show that the 3' UTR is indeed a superposition of a double stem conformation and a pseudoknotted conformation in the presence of KCl and $MgCl_2$. In the absence of added salt however, the 3' UTR adopts exclusively the double stem conformation. Analysis of the pseudoknotted stem reveals only a marginally stable

folded state (ΔG^{25} = 0.5 kcal•mol^{-1}, t_m = 31 ^{o}C) which makes it likely that a viral or host encoded protein(s) is required to stabilize the pseudoknotted conformation.

A second conformationally interconverting RNA system investigated is an RNA element that stimulates -1 programmed ribosomal frameshifting in the human Ma3 gene. Structural analysis of the frameshifting element reveals a dynamic equilibrium between a functionally inactive double stem loop conformation and the active pseudoknotted conformation. Thermal melting and NMR spectroscopy reveal that the double stem loop is the predominant conformation in the absence of added KCl or $MgCl_2$. The addition of KCl and $MgCl_2$ results in the formation of a pseudoknot conformation. This conformation is dominant in solution only when the competing double stem loop conformation is abrogated by mutation. Functional studies of the Ma3 pseudoknot reveal that abrogation of double stem conformation increases frameshift stimulation by 2-fold and indicates that the pseudoknot is the active conformation.

DEDICATION

I would like to gratefully dedicate this work to my husband, Todd, and three sons, Bryce, Aiden and Devlin. They are the ones that make everything worthwhile and give me great joy and happiness. I would like to also dedicate this work to my mother who always provides all of the love and support that anyone could ever ask for.

ACKNOWLEDGEMENTS

Finishing graduate school has been a monumental task that has required patience and help from many. I would like to thank my advisor, Dr. David P. Giedroc. He is extremely dedicated and thorough in all aspects of research and provides an excitement to science that is contagious to all those around him. I would also like to thank my committee members, Dr. Arthur Johnson, Dr. Paul Cremer and Dr. Paul Fitzpatrick, for all of their patience and understanding. I would like to specifically thank my co-chair, Dr. Arthur Johnson, who called my house six years ago and told me about a biophysics program at Texas A&M University; he was very persuasive. I would like to thank all of the members of the Giedroc lab, past and present, especially Lichun Li and Dr. Xiaohua Chen for their helpful advice about science and life, especially cooking. I want to especially thank Dr. Xiangming Kong for the enormous amount of guidance and help with NMR spectroscopy that he has given to me over the years.

Finally, I would like to thank my family. My husband Todd and my three sons Bryce, Aiden and Devlin have always given me unconditional love and support which has meant the world to me especially when research was not going as well as I would have liked. I want to thank my mother, Susan, who sacrificed a year and a half of her life to help me take care of Devlin for which I will always be extremely grateful. I know it was hard for her but she always made sure that I had all of the help that I needed. I would also like to thank my mother-in-law Ellene and my father-in-law Egon for their

encouragement and faith which was instrumental in helping me to complete my research and dissertation.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

RNA research has moved beyond the simplistic view that RNA is strictly an intermediary between DNA and protein (*1, 2*). Experimental milestones such as the structural determination of the ribosome have clearly shown that RNA function is in reality quite complex and not as unassuming as it was perceived to be (*3-7*). Fortunately for RNA researchers, RNA structure and function has turned out to be far more exciting than could ever have been imagined. Presently, RNA is known to play pivotal roles in a range of diverse biochemical processes including protein synthesis (*4-6, 8-10*), gene regulation (*11-17*), and cellular differentiation (*12, 18-20*). This application diversity has required that RNA function in a variety of reactions including enzymatic catalysis (*3, 7, 9, 21-25*), binding recognition (*13, 26, 27*), and molecular switching (*13, 16, 17, 25-31*). Of interest to this thesis is the role that RNA specifically plays in translational frameshifting and viral replication.

RNA STRUCTURE

Though the molecular building blocks for RNA are only four nucleotides (adenine, guanine, cytosine and uracil), there seems to be a limitless supply of conformational diversity and functionality in this intriguing biomolecule. RNA folding

---

This dissertation follows the style of *The EMBO Journal*.

and stability dictate mechanistic reactivity. As such, a fundamental understanding of the driving forces involved in folding, knowledge of dynamic molecular interplay along with an understanding of the folding topologies exhibited by RNA, will allow insight into the RNA function.

Like DNA, RNA is composed of a polynucleotide chain. And, like DNA, the building blocks of the RNA chain are composed of a three component subunit consisting of a cyclic furanoside sugar (β-D-ribose in RNA or β-D-2'-deoxyribose in DNA), a phosphoryl group attached to the 5' position of the ribose sugar and finally a purine or pyrimidine base heterocycle group attached to the $C_{1'}$ position of the sugar moiety via a glycosidic $C_{1'}$-N linkage (see Fig. 1). Chemically, DNA and RNA are almost identical with only two slight changes. RNA contains a Uracil instead of a Thymine base functional group resulting in a change of the 5'methyl that is present on Thymine and absent in Uracil. And secondly, the ribose element differs in that RNA contains a 2' OH group instead of an hydrogen as found in DNA.

Though the two differences may seem inconsequential the ramifications for the biochemical function of these two molecules is overwhelmingly stark. DNA adopts largely double stranded arrangements with a highly flexible and extremely stable configuration. The helical structure adopts mainly a B-form display and the sugar moiety prefers a 2'endo conformation. In addition the helical geometry aligned through sugar conformation and the resultant helix base displacement from the axis, produces a helix with two distinct grooves, described as major and minor, with subsequently very different widths. The major groove is characterized by a width of 22 Å while the minor

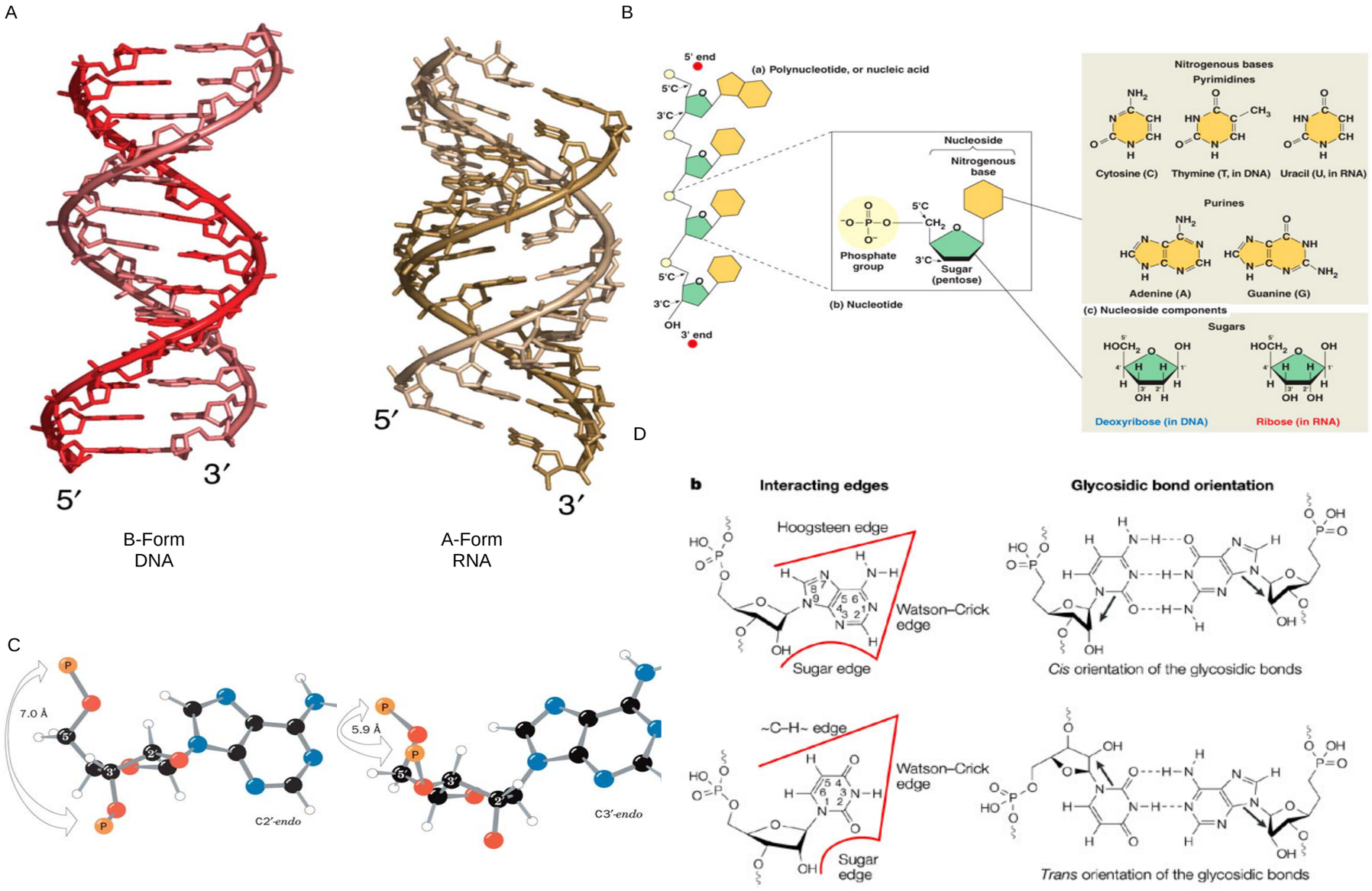

Figure 1 Basic RNA and DNA structure. A) RNA versus DNA double helix. B) RNA and DNA building blocks and how they are arranged. C) 2' Endo versus 3' Endo sugar pucker geometry. D) Base edges and trans versus cis orientation. (Figure modified from references (*26, 27, 29, 31*)).

groove is 12 Å. These structural constraints mean that protein binding is more likely to occur at the major groove rather than on the side of the minor groove. Not only is the major groove larger but it also contains base pairing information and thus can allow proteins to interact in a sequence specific manner.

RNA in contrast to DNA exists primarily as single stranded molecules which fold in highly complex arrangements. The presence of the 2'OH renders the RNA less flexibility and in addition can be a site of back bone hydrolysis thus providing less stability in highly flexible regions. RNA will often form helical geometries through intramolecular interactions. These helical regions generate A form helices, rather than the B form seen with DNA, as a result of the 2 'OH group found on the ribose of RNA. The main differences between the two geometries are the sugar conformation and the base displacement from the axis which generates extremely different groove depths and widths. RNA sugar conformation generally adopts the 3'-endo conformation as compared to the 2'-endo found in DNA. RNA, like DNA, also has two distinct helical grooves. In RNA the minor groove is wide (11 Å) and shallow (3 Å) while the major groove is narrow (3 Å) and deep (13.5 Å). As a consequence interactions in RNA occur mainly at the minor groove where base information is absent since this is the side with the ribose information. This aspect of RNA structure means that proteins do not generally bind with sequence specificity as seen in DNA but mainly through geometric interactions with the uniquely folded RNA.

Having addressed the basic features found in RNA molecules, it is important to understand how these characteristics generate structural diversity in RNA. Structural understanding of how the general features of RNA influence folding and function has progressed greatly in the last 30 years since the first structure of transfer RNA phenylalanine ($tRNA^{Phe}$) was determined *(28, 30)*. Progress was greatly enhanced by the determination of the 160 nucleotide P4-P6 domain of the group I ribozyme determined in 1996 *(32)*. Soon after, the structural determination of the ribosomal subunits 30S and 50S followed closely by the intact 70S ribosome revolutionized our understanding of RNA folding topology as well as protein-RNA interactions *(33-35)*. The combination of these and other RNA structures has lead researchers to devise an RNA folding strategy that seems to follow a hierarchical path *(31, 36, 37)* RNA folding is composed of secondary structure, generated mainly through Watson Crick (WC) base pairing producing A-form helices and tertiary structural interactions (*31*). Structural studies further indicate that RNA is essentially modular in design and that these secondary structures, represented by (see Fig. 2) helices, hairpins, bulges, and loops are the primary building blocks that presents a scaffold for tertiary interactions composed of coaxial stacking, kissing hairpin loops, ribose zippers, A-minor motifs, base triples as well as many others not discussed in this text (*36*). RNA folding is believed to proceed via the formation of proximal secondary elements (i.e. helices) followed by coaxial stacking of contiguous helices which then associate into bundles through short range tertiary interactions that are finally held together through long range tertiary interactions. Because RNA contains unpaired bases, bulges, loops and non-canonical base pairs, the

fundamental problem of groove interaction and molecular recognition is solved. As stated previously, the major groove is too narrow and deep to allow ligand interactions and the minor groove though accessible does not contain sequence specific information. As a result of these unpaired bases, bulges, loops and non-canonical base pairs, the inaccessible major groove widens and is now available for sequence specific interactions (*31*). In addition, such bulged bases are now available to engage in interactions (H-bonding) with other regions of the RNA. In this manner RNA is able to overcome the seemingly simple molecular building blocks and produce a surprisingly versatile and elegant variety of folding motifs.

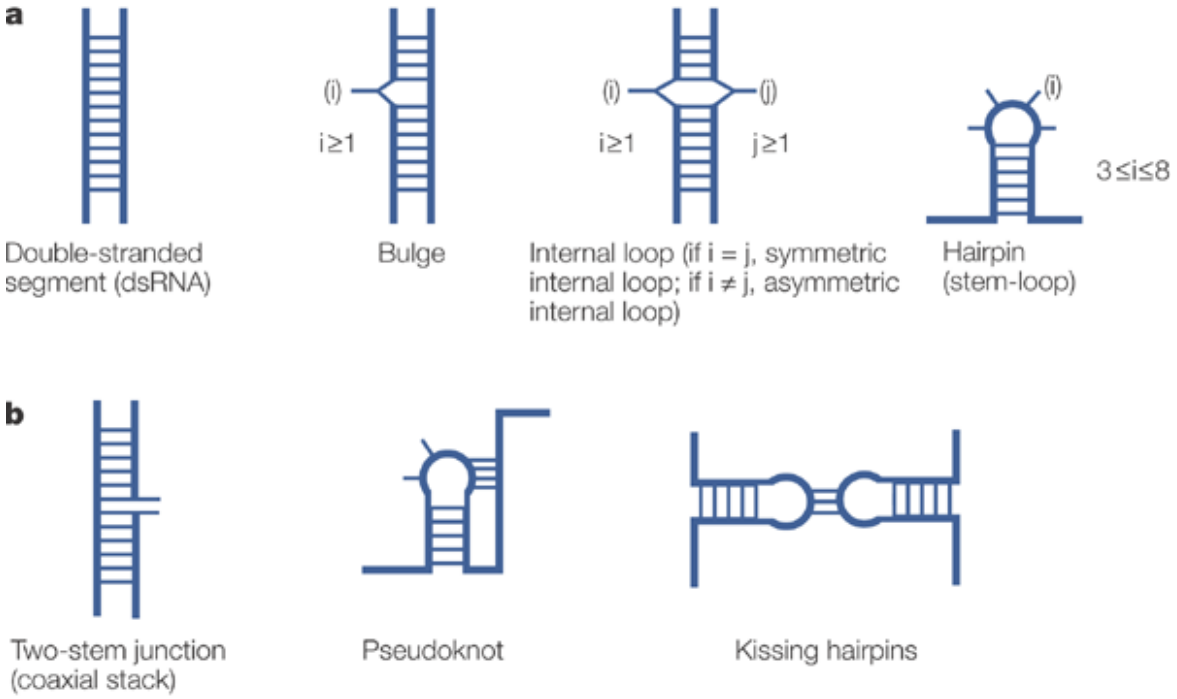

Figure 2 RNA secondary structure folding motifs. (Figure taken from reference(*38*)).

RNA PSEUDOKNOTS

Pleij and coworkers could be described as prescient in their ability to understand the importance of the RNA pseudoknot folding motif (*39*). Though there were few solved RNA structures and the overall understanding of RNA folding topologies were

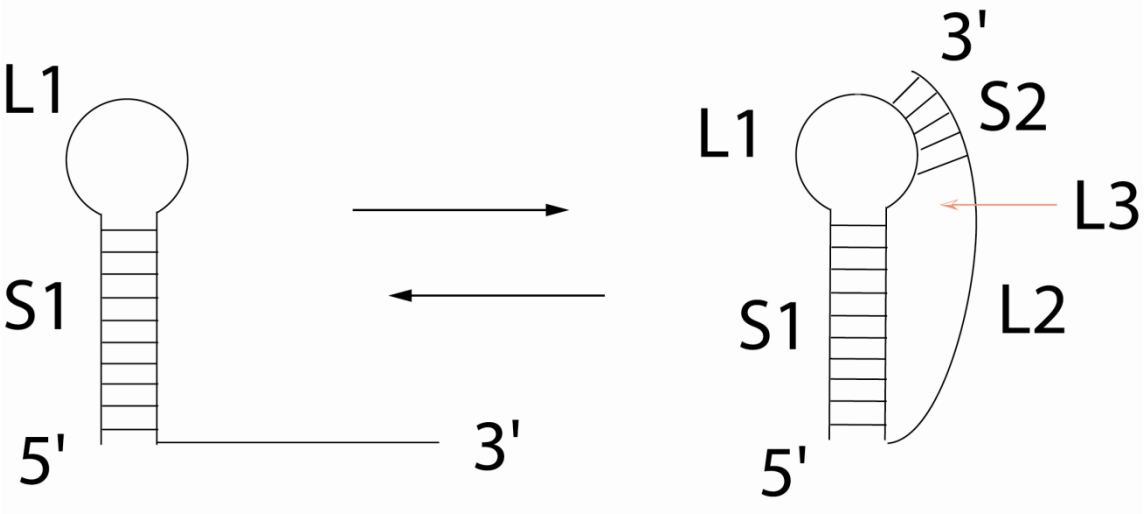

Figure 3 Secondary structure representation of a standard hairpin type pseudoknot.

quite limited, they clearly understood that this proposed new motif would be of fundamental importance in the ability of RNA to engage in long range interactions. Their research illustrated one method in which RNA is able to bridge gaps in folded RNA secondary structures to create complex tertiary interactions. Since this motif was first suggested, there have been numerous examples of the participation of pseudoknots in a diverse array of biological functions including translational recoding, structural scaffolding, transcriptional and translational regulation, and ribosomal rescue (*36, 40, 41*). Structurally pseudoknots are formed when the unpaired nucleotides present in a

loop region base pair with nucleotides elsewhere in the molecule. The unpaired loop region can take the form of a hairpin loop, bulge loop, interior loop or bifurcation loop.

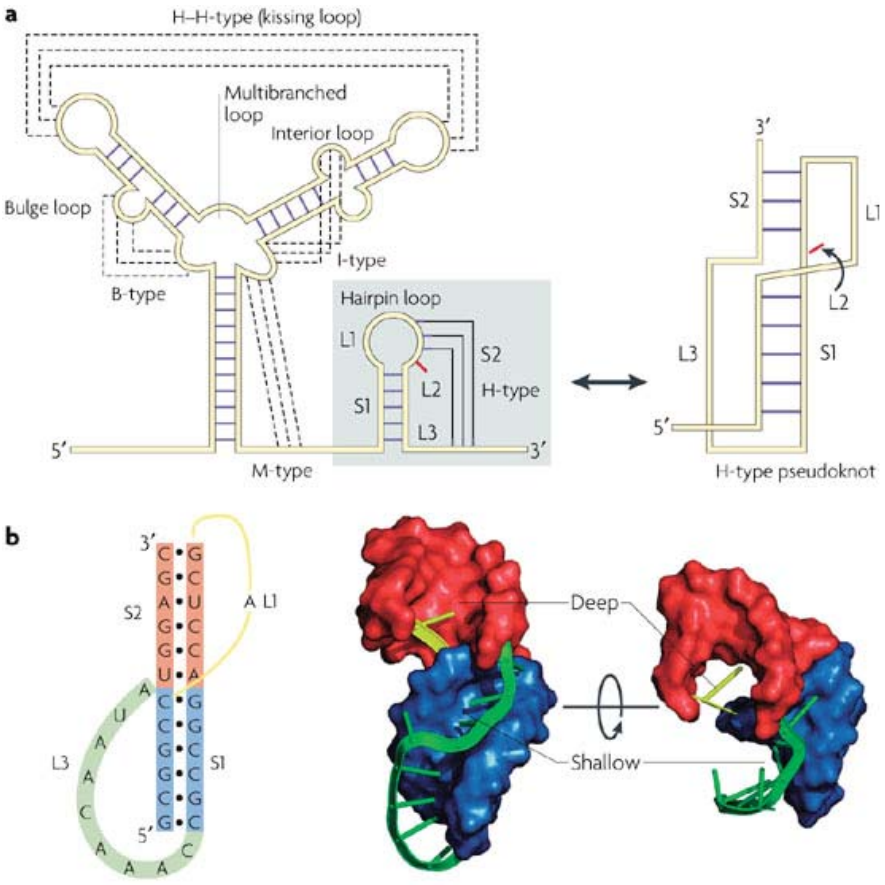

Figure 4 Structural characteristics of RNA pseudoknots. A) Example pseudoknot secondary structure subtypes. B) Schematic pseudoknot conformation. Loop length requirements are based on the length of the shallow groove and deep groove. Please note that the convention of L1, L2 and L3 assignments are different from the labeling of pseudoknot loops in the rest of the text which will use the nomenclature from Figure 3. (Figure taken from reference (*42*)).

The most common type of pseudoknot today is that formed from the loop of a hairpin and is pseudoknot may be formed and is depicted as a general diagram in Figure 3. As shown in Figure 4, a pseudoknot is made up of two stems that can potentially coaxially

stack on one another to create a nearly continuous stem of S1 and S2 base pairs. These stems are connected by two loops which are not equal in length. Structurally, in order for a pseudoknot with coaxially stacked stems to form there are physical requirements

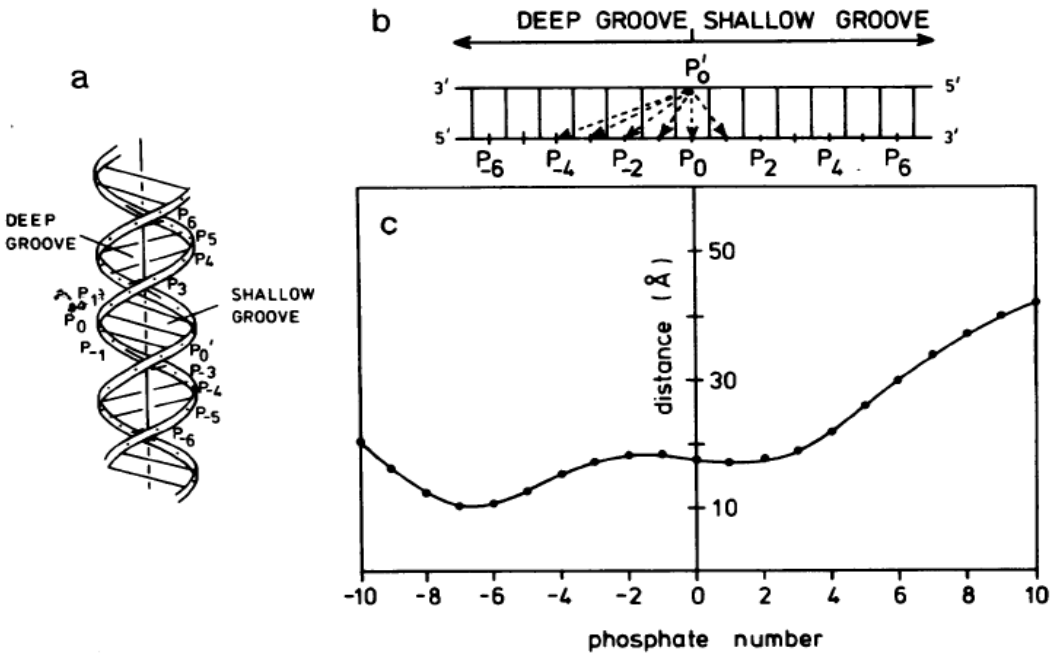

Figure 5 RNA pseudoknot loop and stem length folding requirements. A)RNA A form helix representation. B)distance requirements for phosphate backbone depending on location within the helix. (Figure taken from ref. (*39*)

for stem and loop lengths (see Fig. 5). Pleij and coworkers specifically detailed the lengths that are needed for the connecting loops to bridge the two stems(*39*). Pseudoknots are also defined by the helix-helix junction designated as the site where the two stems join together. This region has been found to be structurally important in regulation of biological function most notably in stimulating programmed ribosomal frameshifting, telomerase RNA structure and internal ribosome entry site (IRES)

function (*40, 41, 43-48*). Below, we briefly discuss representative examples of RNA pseudoknots from

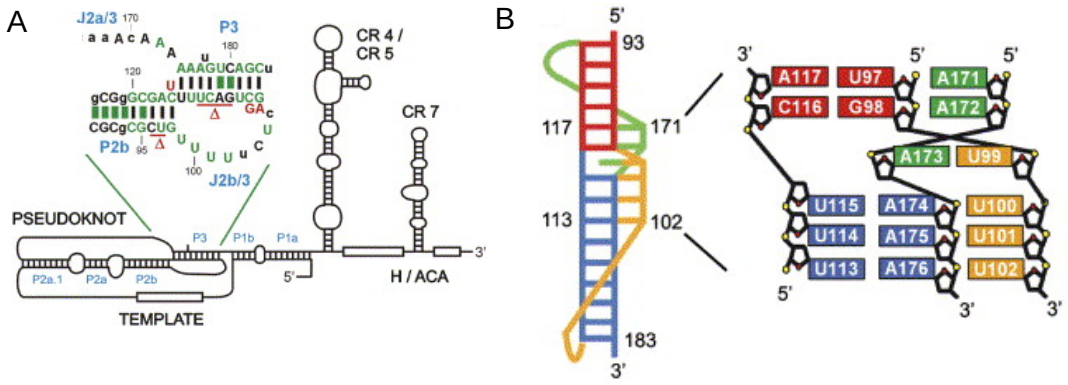

Figure 6 Human telomerase. A) Secondary structure diagram of human telomerase RNA. Green nucleotides indicate conservation, red nucleotides indicate disease related mutations. B) Schematic of the helical junction and an enlarged view of calculated helical junction structure. (Figure taken from reference(*41*).

telomeres, translation initiation, and ribosomal rescue. Translational recoding will be discussed in a separate section.

RNA PSEUDOKNOTS IN TELOMERASE

Every round of genomic transcription results in the shortening of the DNA. Teleomeres functions to maintain the length of genomic DNA by filling in the ends that are lost during the replication process. In addition to restoring the length of genomic DNA, teleomeres also serve as a protective cap to the ends of DNA protecting it from degradation and chromosomal rearrangements. The ribonucleoprotein complex,

telomerase, regulates function at the telomeres and is responsible for the maintenance of genomic length. Human telomerase consists of a 451 nucleotide RNA (hTR) molecule along with multiple proteins of which the catalytic component is characterized by 120 kD protein called human telomerase reverse transcriptase (hTERT). Structural investigations of the hTR have indicated that a highly conserved hairpin type pseudoknot, located in the 5' end of the RNA, is required for optimal telomerase function (*41*). The conserved pseudoknot structure was recently published by Theimer et al. whereby it was found that the 47 nt RNA molecule adopts an almost entirely triple helix orientation which surrounds the helical junction (*41*). Figure 6 shows the structure of the pseudoknot. Maintenance of the extensive triple helix conformation is tied together through a series of tertiary interactions which have been found to be required for telomerase function. These tertiary interactions include two minor groove base triples, a Hoogsteen base pair and three Hoogsteen base triples which span the stem 1-stem 2 junction (*41*). These interactions are necessary for pseudoknot stability and also correlate with disease causing mutations involved in telomerase inactivity. The pseudoknot has been shown to be necessary for TERT binding. In addition, the pseudoknot has also been shown to be important for repeat addition processivity of TERT but not for DNA binding. Theimer et al have proposed that a weak association between TERT and the TR pseudoknot allow for TERT to switch between binding pseudoknot and template binding in DNA synthesis.

RNA PSEUDOKNOTS IN INTERNAL RIBOSOMAL ENTRY SITE (IRES) TRANSLATION INITIATION

In a recent review, Brierly et al describes the function of RNA pseudoknots within the genome of positive sense RNA viruses as having very distinct roles depending on where in the genome the pseudoknot is located (*42*). In non-coding regions pseudoknots influence initiation of protein synthesis and viral replicase binding and recognition. Pseudoknots located in coding regions influence the elongation and termination steps of translation. One ingenious mechanism utilized by positive sense RNA viruses is the cap-independent initiation of translation of viral genomes using internal ribosomal entry sites located in the 5' non-coding region. Although there are many viruses that utilize this mechanism (and specifically pseudoknots) to get around the cellular requirements for translation, the flaviviruses are a good representative class of positive sense RNA viruses that successfully employ internal ribosomal entry sequences to initiate translation.

The flavivirus hepatitis C (HCV) is pathogen affecting 170 million people world wide of mortality and morbidity (*49*). HCV has adapted to human cellular machinery through incorporation of cap independent translation. This is accomplished through direct interactions between the 43S complex and the viral IRES and results in the recognition of the viral start codon without any scanning mechanism involved (*50*).

The HCV IRES is approximately 340 nucleotides long and is composed of three conserved domains (see Fig. 7) (*50-52*). Domain 1 represents the most 5' section and is

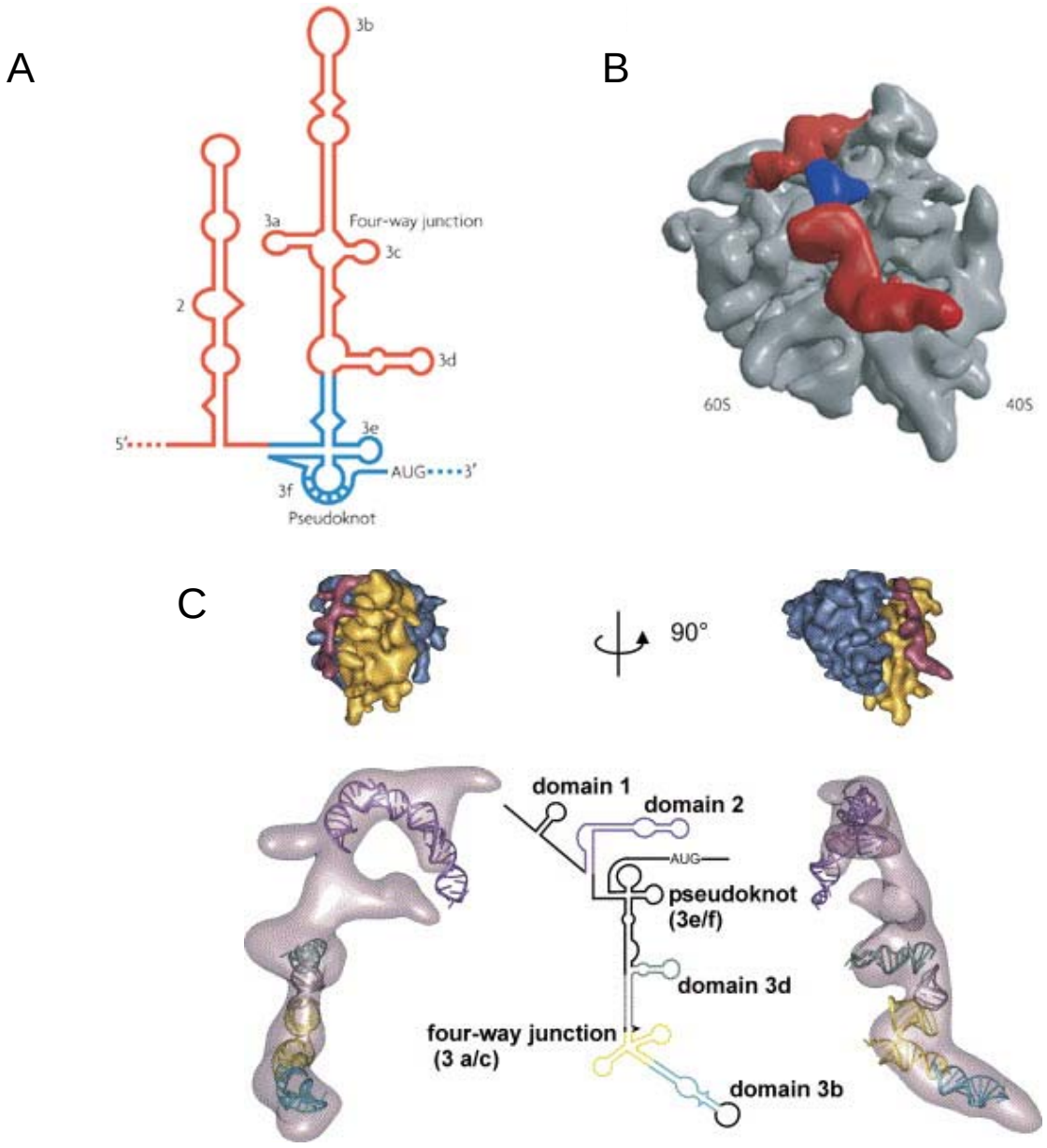

Figure 7 HCV IRES. A) Secondary structure model HCV IRES with pseudoknot shown in blue. B) Surface representation of human 80S (grey) complexed with HCV IRES (red, with the pseudoknot indicated in blue). C) Overlayed fitting of the known HCV IRES with the 80S ribosome. The lower diagram is the density fitted structure of the HCV IRES in both orientations as depicted in the above complex. (Figure taken from references (*42, 48*).

mainly single stranded. The rest of the UTR is well folded and has been characterized by phylogenetic comparison, biochemical probing and mutational analysis and has been found to be essential for translation initiation (*51*). The 3' part of the UTR contains domains 2 and 3. Domain 2 folds into an extended stem loop with multiple internal bulges. Domain 3 is also generally an extended stem loop that contains a 4-way junction and a pseudoknot located at the base of the domain (*42*). Though the pseudoknot structure has not been determined, it is known that it is an H-type pseudoknot with a long loop 1 that encompasses all of domain 3. Information as to the function of the pseudoknot has been suggested based off of a series of cryo-electronmicroscopy (cryo-EM) experiments in which the structure of the HCV IRES complex with the 80S ribosome was determined to 15 Å resolution(*51*). Through docking known IRES structural fragments, as determined by NMR, X-Ray crystallography and comparison with 40S complexes and translating ribosomal complexes, they were able to observe a network of contacts on the 40S and only one with the 60S subunits. In addition they proposed a model based off of their observations in which the pseudoknot which is located at the mRNA exit channel right next to ribosomal protein S5, is thought to be linked to binding of S5 and may play a role in positioning the AUG start codon into the peptidyl site of the ribosome. The suggestion is that the 4-way junction undergoes a conformational change that is transferred to the pseudoknot which positions it close to the mRNA exit channel and places the AUG start in the peptidyl site of the ribosome allowing for subunit joining between the 40S and the 60S subunits.

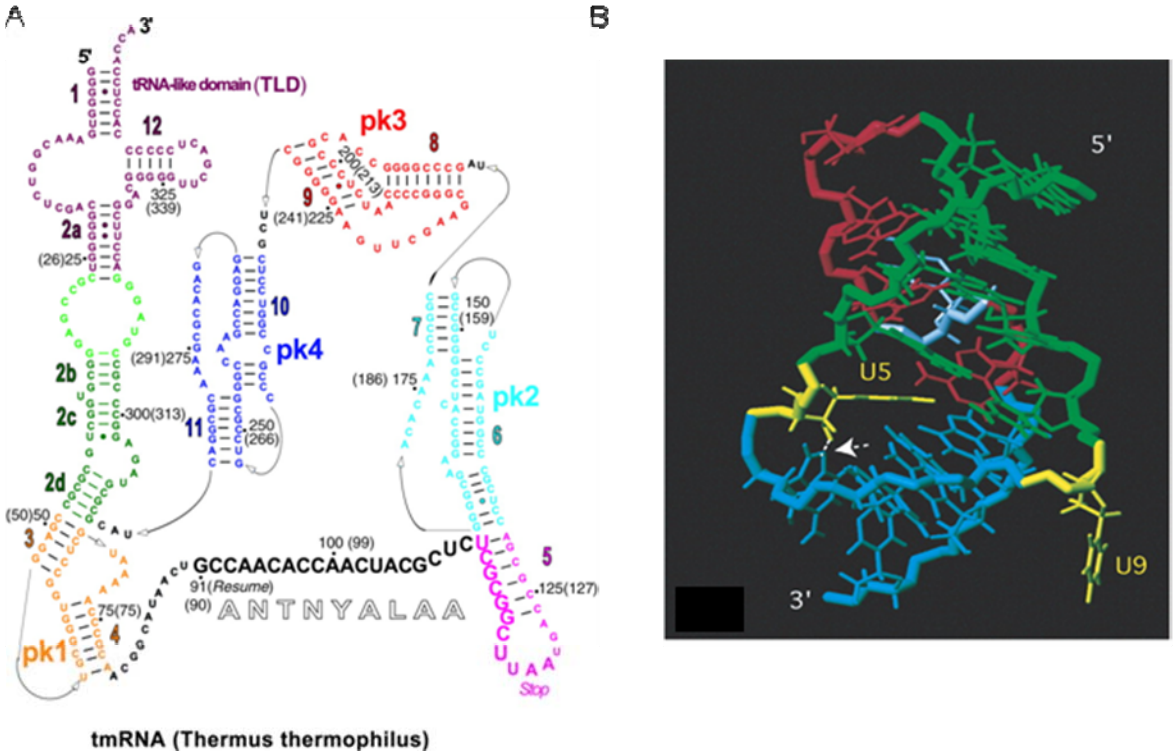

Figure 8 tmRNA. A) Secondary structure diagram of tmRNA B) CryoEM map of the 70S ribosome in complex with tmRNA, EF-Tu, and SmpB. (Figure taken from references(*53, 54*).

RNA PSEUDOKNOTS IN tmRNA MEDIATED RIBOSOMAL RESCUE

The HCV IRES pseudoknot is not the only pseudoknot intimately linked with ribosomal translation, another example involves the rescue of stalled ribosomes otherwise known as transfer messenger RNA (tmRNA). If an mRNA is missing a stop codon, or if there is any other defect in the mRNA which has caused the ribosome to become stuck, then there are potentially deleterious effects that can happen as a result. tmRNA is able to overcome these problems by recruiting the alanine charged tmRNA

into the A site of the stalled ribosome. This interaction would function in a similar manner as a normal tRNA and transfer the alanine to the nascent polypeptide. Translation is then switched from the defective mRNA to the tmRNA open reading frame (ORF) which encodes the degradation sequence. After the tag has been added the ribosome will come across the stop codon within the tmRNA sequence and terminate translation like normal (*55*).

The general structural architecture of tmRNA can be seen in Figure 8 and generally follows a circular loop pattern. The structure contains a tRNA like domain (TLD) which contains a D loop, T arm and acceptor arm that gets charged by alanine at the 3' OH. Following the loop around counter clockwise comes pseudoknot 1 (pk1), the ORF that contains the degradation tag and stop codon, and finally pseudoknots 2, 3, and 4 (pk2, pk3, and pk4)(*53-56*). While the function of the TLD and ORF domains are fairly well recognized the function of the nested set of pseudoknots is still unclear. It is known that the pk's are highly conserved sequences throughout all known tmRNA sequences but the exact role that they play is currently disputed. Some suggested roles for pseudoknot function are structural scaffolding in tmRNA folding, maintenance of RNA stability and prevention of degradation, positional orientation inducer for geometric translational switching or binding sites for proteins important for tmRNA function (*55*).

Although the pseudoknot function in tmRNA is still unclear, the structure for pk1 has been solved and has posed some interesting questions (*53*). The structure is extremely similar to eukaryotic viral frameshifting pseudoknots and in particular the

luteovirus class of frameshifting pseudoknots. It exhibits extensive stacking and stabilization of the junction through base triples, exclusion of a uridine or other flipped out nucleotide, and insertion of loop 2 into the minor groove of stem. In addition, they both contain GC rich stems interacting with A rich loops. Because these features have been shown to be important in functional frameshifting pseudoknots and because the location of the pseudoknots relative to the ribosome are identical, this suggests that interactions between the pseudoknots and the ribosome could be mechanistically similar and structural and functional aspects could be interchangeable.

RIBOSOMAL FRAMESHIFTING

The ribosome is an extraordinary molecular machine that contains highly regulated mechanisms for maintenance of accuracy in ribosomal decoding. The normal error rate is approximately 1 in every 3000 codons (*57*). However, even with all of these safeguards to ensure translational fidelity, ribosomal recoding events programmed by the mRNA itself do occur and in the case of many viruses are vital for maintenance of replication. Ribosomal recoding encompasses many variations on triplet code translation including hopping, stop codon read through, +1 frameshifting, incorporation of unusual amino acids and the one that will be discussed here, -1 programmed ribosomal frameshifting (-1 PRF) (*58*).

-1 PRF is the process by which a translating ribosome will move back one nucleotide in the 5' direction while translating an mRNA molecule (*40*). Because amino acids are translated as triplet codons there are essentially three reading frames to any one

mRNA molecule. By moving back one nucleotide the ribosome has in effect moved into another reading frame, the -1 frame, and will therefore translate an entirely different protein. For frameshifting to occur efficiently three factors must be present. The first element is a heptanucleotide slippery sequence. The general outline of a slip site is X XXY YYZ for the zero frame and XXX YYY Z for the -1 frame. X can be any stretch of three identical nucleotides, Y can be either A or U and Z must be A, C or U. The next element is a linker region that functions in between the first and the third elements to position them correctly relative to one another. It is a stretch of 6-8 nucleotides whose sequence is not highly regulated. The third element is an RNA structural element that is often an H-type pseudoknot (see Fig. 9).

The understanding of how these three elements work together to cause the ribosome to frameshift is not fully understood but recent structural and biophysical data have provided significant insight into possible mechanisms. The general understanding today is that the pseudoknot's ability to promote ribosomal frameshifting is directly related to the ability of the pseudoknot to withstand the unfolding forces of the ribosome (*59-62*). The longer the pseudoknot can stay folded the higher the efficiency of frameshifting. Though, it isn't clear how the ribosome interacts with the pseudoknot to promote frameshifting, the motivation for elucidating the mechanism exists on a fundamental level of first, understanding how the ribosome functions as a translating machine and second, how can this be used as a possible means of combating viruses whose viability is directly related to functioning frameshifting levels.

To understand how the pseudoknot influences ribosomal frameshifting it is important to first understand our current knowledge of the mechanical aspects of

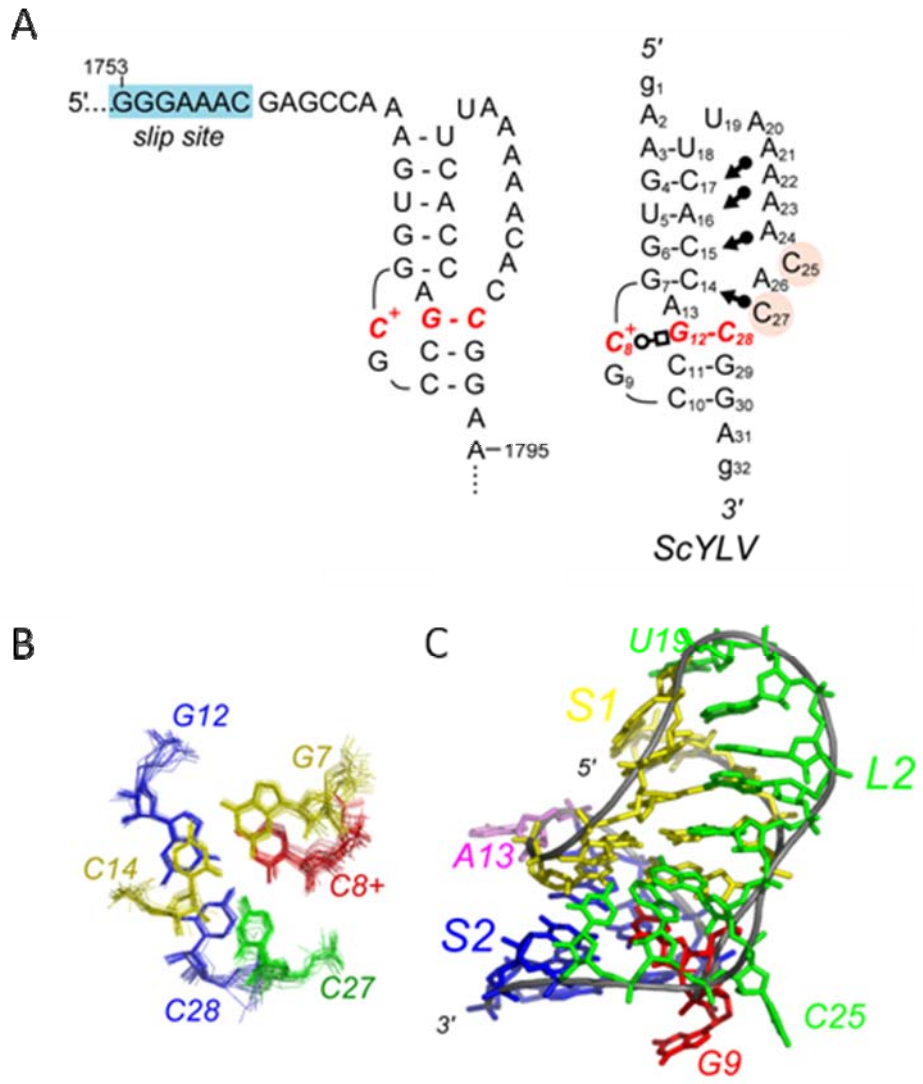

Figure 9 ScYLV and frameshifting. Elements required for frameshifting are a slippery sequence, a linker region and a downstream RNA structural element. Depicted here is the luteovirus sugar cane yellow leaf virus frameshifting pseudoknot (top) and NMR calculated structure (bottom) including the helical junction interactions. (Figure taken from reference (*44*).

ribosomal translocation. The theory describing the hybrid state of tRNA binding has recently become a very prominent model describing the state of the ribosome during translocation (*63*). The tRNA hybrid state refers to the idea that tRNA's, and the directly bound mRNA, can move independently relative to the ribosome. After the A site EF-Tu-GTP-aa-tRNA complex undergoes accommodation and following the subsequent peptidyl-transferase reaction, the tRNA's move from the classic A/A and P/P state into a hybrid A/P and P/E state. This intermediate state is defined by the fluctuating movement of the A and P tRNA acceptor loops moving between their current sites and their subsequent new target sites while their respective anticodon loops remain connected to the mRNA in the untranslated "classic" state. This hybrid state essentially amounts to movement within the 50S subunit only and a molecular anchoring within the 30S state. At this point the translocation factor EF-G complexed with GTP binds the existing hybrid state ribosome and catalyzes the movement of the 30S subunit into the classical P/P and E/E state producing an empty A/A state that is ready for the incoming aa-tRNA. This movement can be thought of as a tRNA induced dragging of the mRNA through the ribosome as they travel into the next ribosomal tRNA site (*63*).

Keeping the series of ribosomal translocation events in mind, there have been several models proposed explaining where and how frameshifting occurs. The models include simultaneous slippage, torsional restraint and a newly postulated "spring and ratchet system" (*59-62, 64*). Fundamentally, all of the models will need to address the movement from the X XXY YYZ zero frame state into a XXX YYY Z -1 shifted state.

This shift involves a basepairing difference in only the wobble site since the other two sites will inevitably bind the same type of nucleotide, but have simply moved over slightly. This infers that since the wobble site makes non-specific interactions and the other interactions will interact in the same manner, that energetically, the two states are the same. However, even though the overall net energy remains the same the activation energy to move from one state into the other must be rather large since frameshifting is specified to occur at certain events. To understand the influence of the pseudoknot on the promotion of frameshifting a description of the "spring and ratchet system will be discussed and the implications that it has for causing ribosomes to undergo such programmed events (*59*).

The "spring and ratchet" system has been proposed largely based on of the cryo-EM data in which a ribosome is stalled during the process of frameshifting using the IBV pseudoknot and a non-functioning slippery site (*59*). From of the data collected it was observed that the pseudoknot is indeed stuck at the entrance to the mRNA tunnel on the 30S subunit. In addition the pseudoknot appears to be interacting in the vicinity of the ribosomal helicases S3, S4, and S5. It was further seen that the elongation factor eEF2 (EF-G in eukaryotes) is bound in the A site and interacts with the P site tRNA. The eEF2 precludes the binding of any tRNA into the A site (*59, 65*). The P site tRNA was also found to be distorted in the cryo-EM structure with the D-arm strongly bent towards the A site where eEF2 is located. This D helix bending is very reminiscent to the flexibility observed in the aa-tRNA during accommodation which also depicted a spring-like orientation of the tRNA. All of these factors suggest that an interplay

between pseudoknot, eEF2, tRNA and subsequent ribosomal rearrangements are involved in the process of ribosomal frameshifting (*59, 65*). Taking all of the pieces together a mechanical explanation of frameshifting is proposed in which the elongating ribosomal helicases located in the mRNA entrance tunnel are unable to unwind the incoming pseudoknot and blockage is produced. When the ribosome attempts a translocation, movement of the tRNA through the ribosome is inhibited and a tension is produced between the ribosome and the translocating tRNA. Because the mRNA is bound to the tRNA through the anticodon-codon interactions and because the A site is occupied by eEF2, this tension results in a bending of the tRNA that is inevitably followed by a disconnection between the anticodon-codon interactions allowing for the repairing of the P site tRNA with the mRNA in the -1 frame. In short the tRNA is acting like a molecular spring that is released upon the hydrolysis of eEF2-GTP (*59, 61, 65*). This spring like motion releases the tension produced by the pseudoknot blockage through the breakage and subsequent relaxation of codon-anticodon interactions which results in final base pairing in the -1 frame. The end result is that a pseudoknot that can produce enough tension to allow breakage of the codon-anticodon interaction will result in ribosomal frameshifting.

CORONAVIRUSES

There has been a recognized growing need for the discovery of more antiviral drugs. As of 2004 there were approximately 40 FDA approved antiviral drugs with 19 of them specifically for HIV (*66*). With the extreme number and types of viruses present

in the world the demand will only continue to increase. This demand will require an in depth understanding of how viruses function including infectivity, replication and packaging.

Coronaviruses are a member of the order *Nidovirales* which includes the families *Roniviridae*, *Arteriviridae* and *Coronaviridae* (*67*). They are the causative agent of up to 30% of colds, however prior to severe acute respiratory syndrome (SARS) they rarely caused lower respiratory infections. Though in humans most coronaviruses are benign, they can cause devastation to animal livestock such as cows and pigs and are deadly to cats and chickens, as well as many other avian species (*68*). Coronaviruses are characterized by a unique image in cryo-EM that appears as though a crown surrounds the outer edges of the viral particle. This solar corona appearance exists as a result of the 20 nm spike protein imbedded in the outer membrane (*69*). Because the virion particules are pleiomorphic, ultrastructural analysis has been unsuccessful. The generally accepted particle structure, diagramed in Figure 10, consists of a nucleocapsid surrounded by a lipid envelope. Embedded in the envelope are three canonical coronaviral structural proteins called membrane protein (M), envelope protein (E), and spike protein (S). The ribonucleoprotein core contains one copy of the viral genome that is packaged into a helical structure by multiple copies of the nucleocapsid (N) protein.

There are three classified groups of coronaviruses based on phylogenetic and serological characterization. Group I consists of human coronavirus 229E (HCoV-229E), human coronavirus (HCoV-NL63), porcine respiratory coronavirus (PRCV),

porcine transmissible gastroenteritis virus (TGEV), porcine epidemic diarrhea virus (PEDV) feline infectious peritonitis virus (FIPV), feline enteritis virus (FEV) and canine

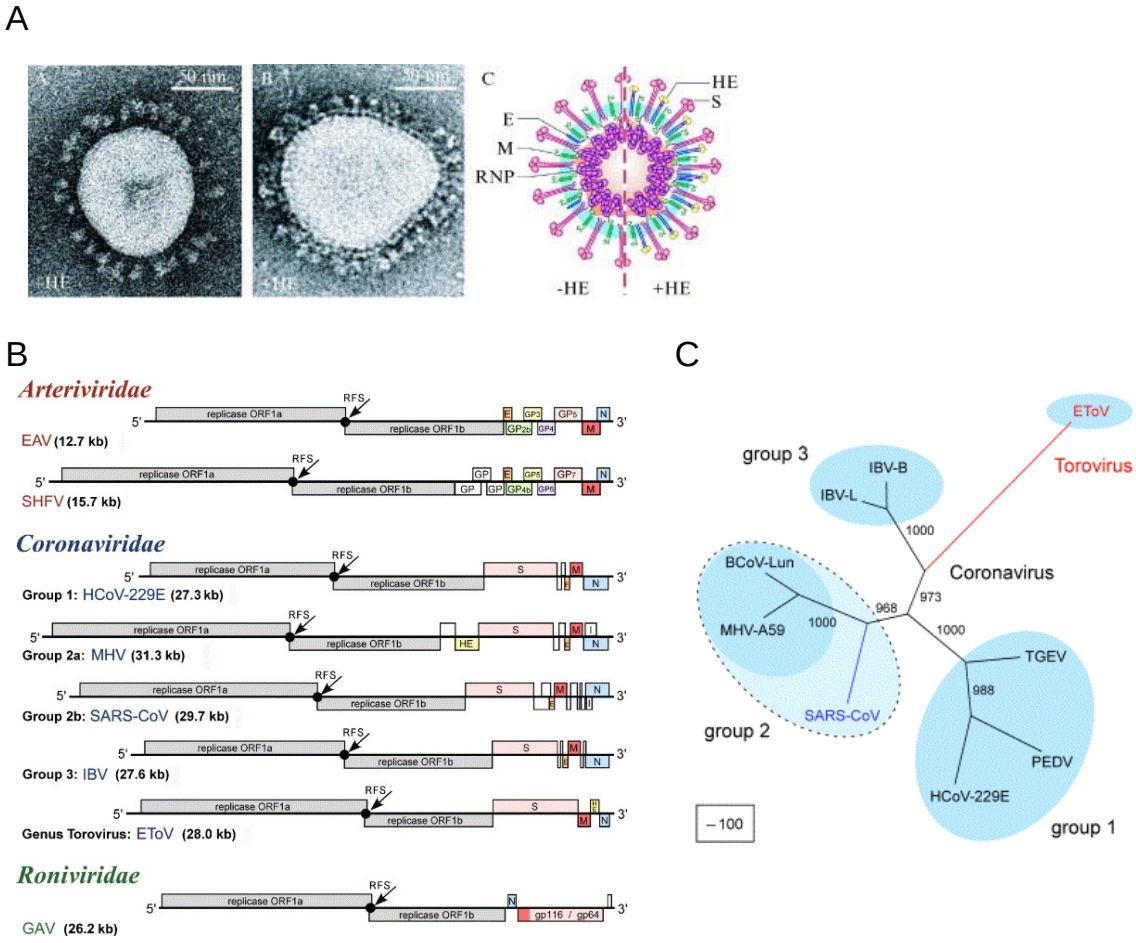

Figure 10 Overview of coronaviruses. A) Electron micrograph of coronavirus particle alongside a general schematic of the vital particle. The crown appearance results from the spike protein extending out of the membrane. B) Genomic outline of members of the nidovirus family. C) Coronalvirus subgroups. (Figure taken from (*69-71*).

coronavirus (CCoV). Group II consists of human coronavirus OC43 (HCoV-OC43), human coronavirus (HCoV-HKU1), bovine coronavirus (BCoV), mouse hepatitis virus

(MHV) and more recently, severe acute respiratory syndrome coronavirus (SARS-CoV). Finally the third group contains mainly avian species such as the chicken infectious bronchitis virus (IBV) and turkey coronavirus (TCoV) (*72*). In all cases, except for SARS-CoV, group assignments were clearly determined. The SARS-CoV however initially caused considerable debate as to group designation and whether it might constitute an entirely new group. Determination of group classification in the past was originally performed through antigenic response. This has proven erroneous in many situations and now a combination of serological response in concordance with genetic comparisons has become the method of choice. In the case of SARS-CoV however, comparison of genetic sequence placed it equidistant from each of the three groups. Sequence comparison of genes 1a and 1b, in addition to structural similarities found in the 3' untranslated region, clearly establish SARS as a group II coronavirus (*73*).

The order *Nidovirales* contains a number of highly complex unique RNA viruses typified by the production of a co-terminal nested set of subgenomic (sg) mRNAs transcribed from the 3' segment of the RNA genome (see Fig. 11). This characteristic quality is exhibited in the order name since *nidus* is Latin for nest. *Nidovirales* possess linear single stranded positive sense RNA genomes that contain a 5' cap structure and a 3' polyadenylated tail (*70*). The viruses contain untranslated regions located at the 5' and 3' ends of the genome with multiple open reading frames between them. Though the order nidovirales is drawn together by the hallmark the unique sg mRNA synthesis, many features are strikingly dissimilar. A major example is the genomic size variation

which can range from 12 kb for arteriviruses, to 32 kb for coronaviruses (*70*). The main discriminating feature of nidoviruses compared to all other RNA viruses is the large

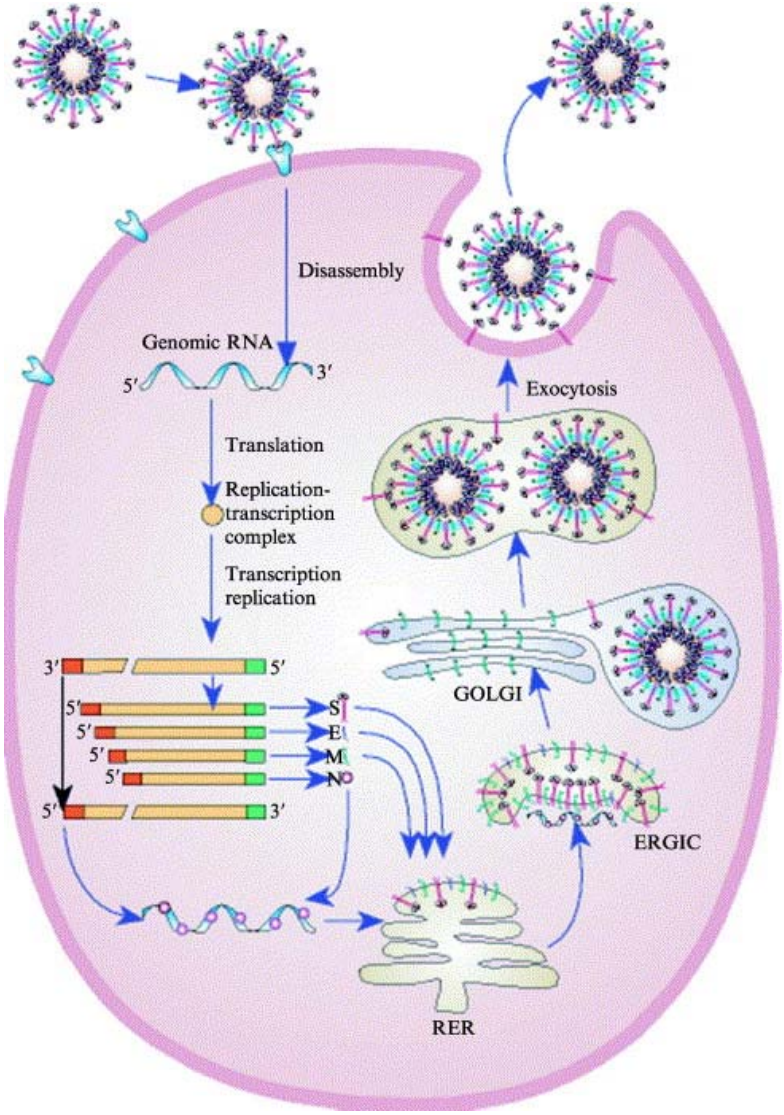

Figure 11 Viral replication cycle of coronaviruses. (Figure taken from reference (*69*)).

complex organization and composition of the multidomain replicase gene which encompasses more than two thirds of the viral genome.

Coronavirus genomes are extremely large nonsegmented single stranded positive-sense (i.e. they are the same sense as mRNA and are thus directly translated)

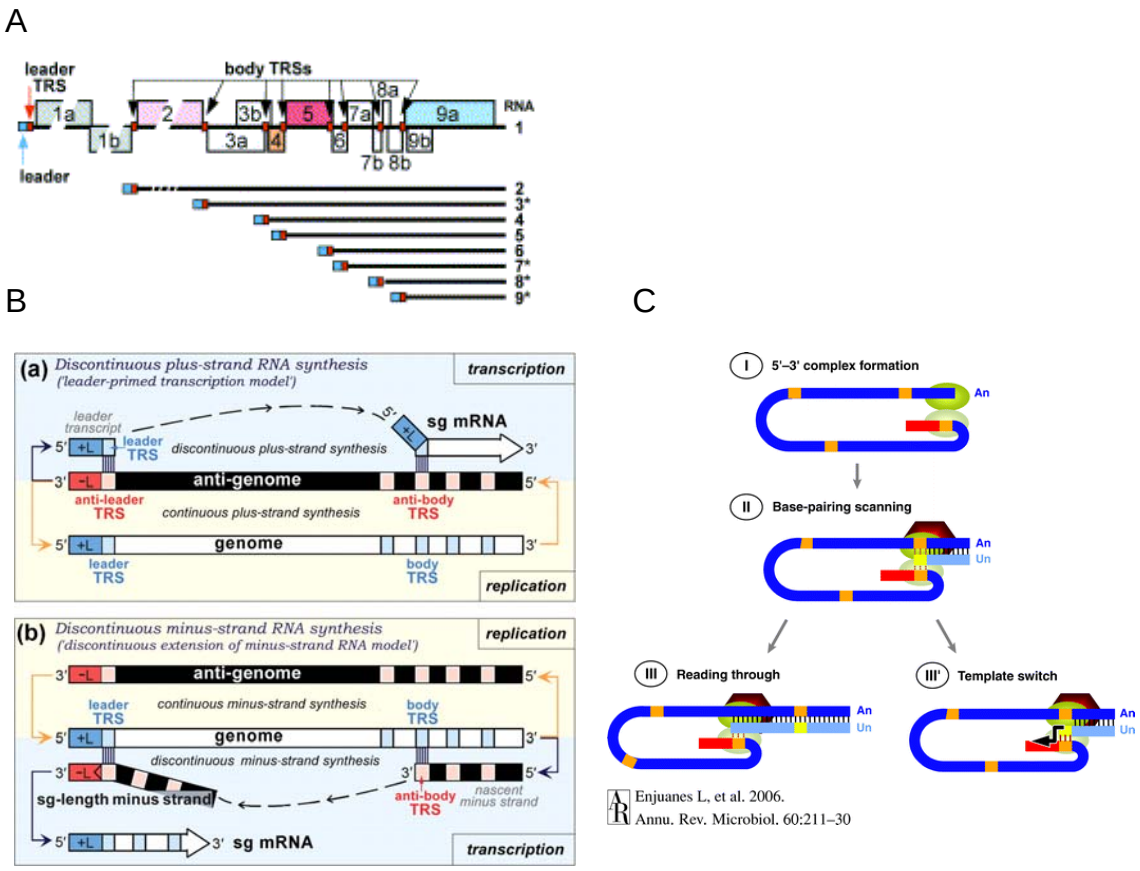

Figure 12 Coronaviral genomic replication. A) Diagram of subgenomic mRNA generated during MHV viral replication. B) Schematic of two conflicting models describing subgenomic mRNA synthesis. After the discovery of negative sense sg mRNA the Discontinuous minus strand RNA synthesis has become more favorable. C) Proposed sg mRNA synthesis. (Figure taken from references (*71, 74, 75*).

RNA molecules that are nearly three times the size of alphaviruses and flaviviruses and four times the size of picoronavirus genomes. With lengths ranging from 27.3 kb for HCoV-229E to 31.3 kb for MHV, they represent the largest mature RNA molecules

known to biology (*73*). As with all nidoviruses, the genome consists of multiple ORFs with the single viral replicase gene occupying the 5' most two thirds of the genome. The remaining structural proteins reside in the remaining one third clustered at the 3' end. The genomic gene order is invariant between all coronaviruses arranged as 5' replicase-S-E-M-N-3'. A number of accessory genes have also been determined spread throughout the structural genes. The full length genomic RNA functions as the template for translation for the replicase gene only with the remaining genes expressed from subgenomic RNAs. The coronaviral genome as with other positive sense RNAs are infectious when transfected into permissive host cells (*73*).

The coronaviral lifecycle, as with all viruses, begins with entry into the cell. The S protein is responsible for receptor binding and recognition and drives the conformational changes required for viral entry. The S protein is also responsible for species specificity (*76*). The main receptor required for viral entry in SARS is the angiotensin-converting enzyme 2 (ACE2). ACE 2 is a metalloprotease that is expressed in human heart, kidneys, testis, gastrointestinal tract and lungs. After the fusion of viral and cellular membranes, the nucleocapsid is delivered into the cellular cytoplasm. Coronaviral genome expression begins with the large ORF1a and ORF1b genes. ORF1a is composed of 4382 amino acid residues and is termed pp1a. ORF1b expresses the protein pp1ab which is 7073 amino acid residues and must be translated through a -1 translational frameshift (*67*). Through the participation of a downstream pseudoknot and the slippery sequence UUUAAAC, the efficiency is around 14% in vitro using rabbit reticulocyte lysate (*77*). pp1a and pp1ab are highly conserved sequences within

coronaviruses. Once they are translated they are cleaved by two virus encoded proteinases producing 16 individual proteins (*72*). The processed products collectively make up the replication-transcription complex. As a result of the many ORF1a hydrophobic protein subdomains, the replicase proteins interact with modified intracellular membranes to form a membrane associated RNA replication complex. This complex in concert with cellular factors is responsible for replication of the viral genome and the nested set of subgenomic RNA (*67*). The small genomic transcripts serve as translational templates for the structural proteins and the variety of accessory proteins located in the 3' region of the genome (*75*). Like all other positive sense RNA viruses, genomic replication proceeds through a negative strand intermediate which serves as the template for more positive sensed full length genomic sequences which are subsequently packaged into new viral particles (*78*). The newly transcribed genomes are encapsidated by the nucleocapsid N protein and the progeny viral particles are enveloped through cellular packaging involving the endoplasmic reticulum and golgi apparatus (*75*).

One of the characteristic features of coronaviruses is the co-terminal nested set of subgenomic (sg) mRNAs (Fig. 12A). A series of sg mRNA transcripts are produced which contain an identical 3' UTR and 5'leader RNA (the 5' most portion of the 5' UTR) that is found in the full length genomic RNA. The generated nested set of sg mRNA are produced such that a descending order of sizes are synthesized containing the ORF sequences required for expression of the structural proteins. The sg mRNAs are also characterized by the presence of a short conserved sequence termed transcription regulatory sequences (TRS) define the site of fusion of the 5' leader to a 3' RNA ORF-

containing fragment (Fig. 12B). The simultaneous presence of leader and genomic body TRSs in the genomic RNA are consistent with the concept of co-transcriptional leader to body fusion that occurs during the process of discontinuous sg RNA transcription (*75*) (Fig. 12 B-C).

Understanding the mechanism of discontinuous 5' and 3' UTR end fusion involved in production of the sg mRNA has required years of research and has generated two contrasting models of synthesis (*75*). Both models concur that the process involves the co-transcriptional fusion of leader and body segments but differ on whether the discontinuous step occurs during plus- or minus-strand synthesis. The two most prevalent models are termed the "leader primed transcription" and the "discontinuous extension of minus strand RNA synthesis" (*75, 79-81*). The leader primed model proposed by Lai and coworkers suggested that the viral transcriptase initiated replication at the 3' UTR and that after the initial production of the leader sequence along with the associated TRS, the transcriptase would jump from the 3' end of the minus strand to an internal complementary TRS located further down in the genome. This internal TRS would redirect the transcriptase and serve as an internal primer for transcription (*80*). The discovery of full-length negative-sense sg mRNAs in infected cells is inconsistent with leader-primed model which has therefore been replaced by the more generally accepted model of discontinuous minus strand RNA synthesis, to be discussed here in more detail.

Sawicki and co-workers first proposed the discontinuous extension of minus-strand RNA synthesis model as an answer to the newly determined data regarding sg

mRNA synthesis (*82*). In the model minus strand RNA synthesis was proposed as the event for discontinuous replication and not plus strand synthesis. The discontinuous event resulted from the presence of multiple TRSs scattered throughout the genomic template. The basic idea is that as the sg mRNA is being transcribed the nascent mRNA will come across a body TRS and the newly produced complementary TRS nascent mRNA will have the opportunity to base pair with the leader TRS located in the 5'UTR (Fig. 12B-C). If the 3' newly transcribed anti-body sequence basepairs with the 5' UTR leader TRS, than the replicase will switch strands and continue transcription on the leader 5' UTR sequence which has become the new template sequence (*82*). The process can be broken down into a series of events. First, the replicase complex must be recruited to the end of 3' UTR where minus strand synthesis is initiated. Second, the replicase complex moves along the template transcribing the nascent mRNA until it comes across a body TRS. Third, the replicase complex can either a) continue on past the body TRS transcribing the template as before or b) it can stop synthesis of the nascent mRNA and switch to a new template. This process of template switching, which can be looked at as the fifth step, can proceed through a relocation event in which the complementarity between the nascent mRNA and the upstream leader TRS can interact to allow for template switching. The switching event will allow the replicase to continue out the transcription of the 5' UTR essentially joining the nascent 3' UTR mRNA to the 5' leader sequence (*82*).

Though the model for discontinuous extension during subgenome length minus strand synthesis is intriguing and has recently had many experiments confirming the

basic tenets put forth by the model, there are still many areas that have not been addressed. Namely, the interactions between the 3' end and the 5' end of the genome must have some mediating force that brings them in close proximity to one another (Fig. 12C). Evidence in support of a physical interaction between the 5' and 3' UTRs has recently been obtained whereby mutations introduced in the 5' UTR of MHV coronavirus found to disrupt negative-strand sgRNA synthesis are rescued by compensatory or supressing mutations in the 3' UTR that restore RNA synthesis (*83*). While it is not clear how this is accomplished, there have been many proposed mechanisms. For example a protein-protein interaction might serve as a "bridge" that unites the two ends of the genome essentially circularizing it to allow for template switching. Another possibility is that the polymerase itself could actually bind both the 3' end as well as the 5' end of the genome at the same time and create a proof reading system in which the 5' UTR is scanned while the template is being synthesized. If this mechanism proved to be correct than it would also be true that only one polymerase could be functioning on any one genome at any one time (*81*).

MOLECULAR SWITCHES AND RNA-RNA INTERACTIONS

The inherent adaptability of RNA folding has been exploited by many organisms to control gene regulation and expression. One example is the riboswitch class of RNAs which are known to undergo complex large conformational changes to drive the expression of genes expression and suppression. Innovative use of RNA in biochemical regulation does not stop with riboswitches. Exploitation of conformationally switching

RNA has been occurring for quite some time in RNA viruses. Conformational switching is the optimal pathway for viruses to obtain the largest amount of diversity in the smallese amount of genetic material. The viral particle is limited in space; consequently the genetic material packaged within the particle must have all of the information necessary to carry out its propagation once inside of cell. When a virus first infects a cell it must translate the genome to synthesize the polymerase needed to replicate its genome. This means that the same sequence must contain signals which regulate when to switch from translation to transcription. In addition to signals which regulate the timing of events, many viruses contain signals which dictate either in increase or a decrease in the amount of product made. To illustrate these characteristics, two examples of conformationally dynamic RNA viruses will be discussed (Fig. 13).

The *tombusviridae* family of viruses are among the smallest and most simples positive sense RNA viruses. The genome of turnip crinkle virus is 4054 nucleotides long and encodes five proteins (*84-86*). Through genetic comparative analysis and computer modeling of the 3' UTR, conserved secondary elements were determined to be three positive stranded hairpins (Fig. 13A). Beginning at the 3' end,the first element is the core promoter, Pr, followed by H5 hairpin, which also contains a large symmetric loop LSL, and finally H4b which is a hairpin which has not been assigned a function. Some viruses in this family are also known to contain a fourth hairpin which has been designated H4a. The conserved LSL that is part of the H5 hairpin, also participates in a pseudoknot interaction with a sequence just 3' to the Pr (Fig. 13A).

Although in vivo data generally support the presence of these secondary structures, in vitro transcription of the 3' UTR sequence showed no sign of H5, Pr or the LSL 3' pseudoknot interaction in structural probing experiments. In vivo experiments effecting the 3' terminal cytosines or 5' terminal guanosines, as well as mutations

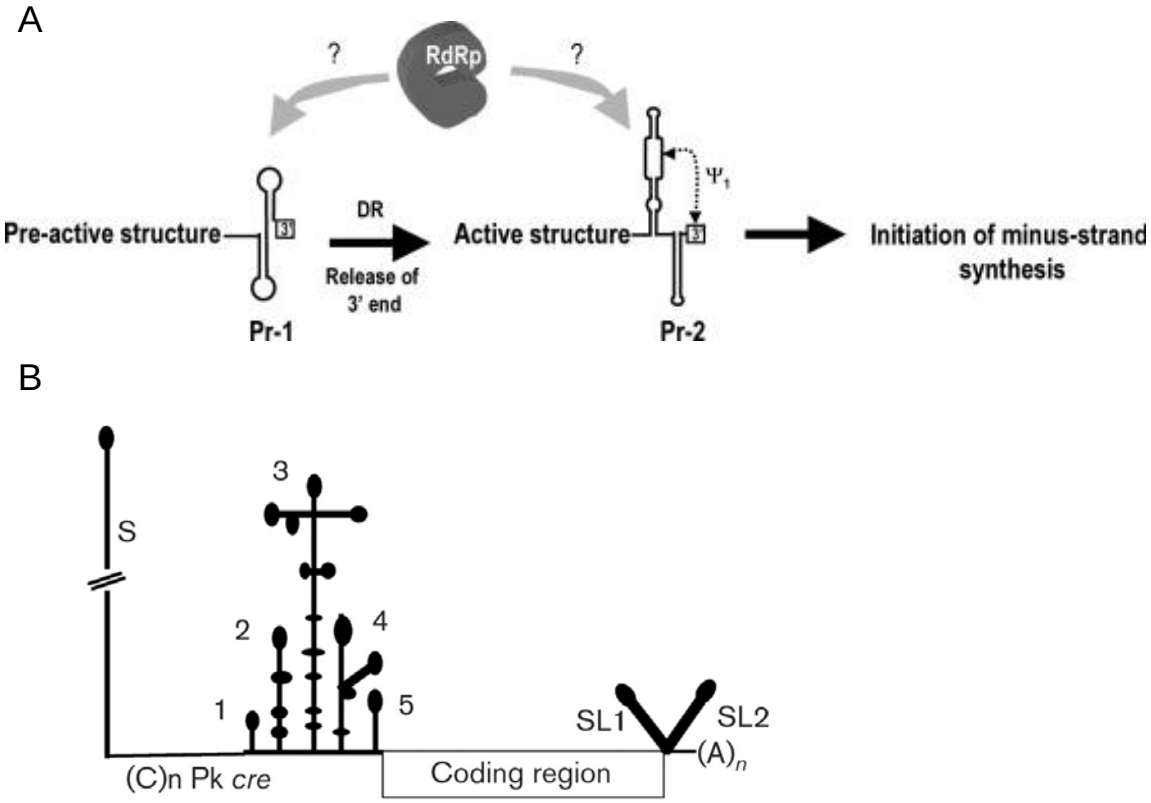

Figure 13 Turnip crinkle virus transcription. A) Proposed model of turnip crinkle virus initiation of transcription. Genome switches from a preactive non functional state to an activated transcription competent state. B) Genomic outline of FMDV. The genome is proposed to have two long distance interactions with the 3'UTR, one with the S domain and one with the IRES (*86*). (Figures taken from references (*75, 78*).

involving the H5 hairpin caused a 3' UTR rearrangement in the Pr element. The Pr converts from the initial, in vitro determined Pr-1 conformation which contains an alternate pseudoknot conformation, to the phylogenetically determined LSL 3'

pseudoknot containing Pr-2 conformation (Fig. 13A). The Pr-2 conformation is the transcriptionally active state exhibiting a 20 fold increase in replicated RNA concentration. In this way the virus is able to enlist the services of a conformational switch to turn on and off transcription (*86*).

Another example of conformational regulation of cellular function is in the positive sense picornavirus, foot and mouth disease (FMDV). The genome is 8500 nucleotides with the 5' and 3' UTR consisting of approximately 1300 and 100 nucleotides respectively (*87*). The 5' UTR contains several organizational domains including the S region, poly (C) tract, several pseudoknots, the *cis* acting replication element and the internal ribosomal entry site (IRES) (Fig. 13B). The 3' UTR is organized into two stem loops that are essential to virus replication. Many viruses mediate interactions between the 5' UTR and 3' UTR. In FMDV it has been shown that two long distance interactions are possible, with one involving the 3' UTR and the 5' IRES element and another involving the 3' UTR and 5' S domain interaction (see Fig. 13B). It was subsequently determined that the 3' UTR 5' IRES interaction is vital for viral translation and that the 3' UTR and 5' S domain interaction is required for replication initiation (*87*). The mechanism that mediates this conformational switch required for translation vs. replication is not fully understood. Nonetheless, much like the case for RNA riboswitches discussed previously, when anRNA is presented with an appropriate "trigger", be it a small molecule or a virally or host-encoded protein, it can be induced to form one conformation relative to another.

SCOPE OF DISSERTATION

In this dissertation, we present work designed to evaluate the degree to which RNA conformational switching impacts replication of a model coronavirus, mouse hepatitis virus (MHV), and the expression of human gene by –1 programmed ribosomal frameshifting. In Chapter II we investigate the thermodynamic parameters that control a putative molecular switch found in the 3' UTR of MHV utilizing optical melting profiles. In Chapter III this proposed molecular switch is conformationally investigated utilizing NMR spectroscopy. Chapter IV investigates the population distribution of proposed conformational states of the MHV 3' UTR through sedimentation velocity analysis as well as fluorescence resonance energy transfer (FRET) spectroscopy. Finally in Chapter V we present structural and thermodynamic investigations of a frameshifting pseudoknot found in the human Ma3 gene.

CHAPTER II

CHARACTERIZATION OF THE 3' UNTRANSLATED REGION IN MOUSE HEPATITIS VIRUS BY THERMAL UNFOLDING

INTRODUCTION

Coronaviruses are enveloped, positive sense RNA viruses that replicate intracytoplasmically through a membrane bound replication complex. During viral replication genomic length RNA as well as a 3' nested set of coterminal subgenomic mRNAs (sg mRNA) are generated. The sg mRNAs are transcribed such that the 5' and the 3' ends of the transcript are exactly identical to the genomic length 5' and 3' UTRs counterparts. The mechanism of replication of the genomic RNA and the sg mRNA requires cis acting sequences located in both the 5' UTR as well as the 3' UTR. These sequences are known to play essential roles in processes involving RNA synthesis, gene expression, translation and virion assembly.

Elucidation of the essential RNA elements required for replication has historically been determined through the use of enzymatic and chemical probing and the use of defective interfering RNAs (DI RNAs) (*88-91*). DI RNAs are extensively deleted genomic RNA elements capable of replication with the aid of a helper virus. More recently, full length cDNA clones have permitted the use of reverse genetic approaches that allow the introduction of genomic mutations anywhere within the viral genome. Infectious cDNA clones been constructed for MHV, SARS-CoV and NL63 (*88, 92*). Mouse hepatitis virus (MHV) is a group 2 coronavirus that has been studied extensively

and represents a well developed model system for the study of coronavirus replication and assembly. Since SARS-CoV is also a group 2 coronavirus, MHV has emerged as an excellent model system with which to obtain insight into the replication of SARS-CoV.

Both genomic RNA and sg RNA transcription initiate at the 3' UTR. As such, investigators have focused attention on determining the elements involved in transcription initiation and replication. Observations that MHV and the closely related group 2 coronavirus BCoV require the entire 3' UTR, beginning at the end of the N protein through the poly (A) tail, have lead investigators to focus on structural elements present within the 3' UTR. The 3' UTR can be divided into three structural regions as depicted in Figure 14. Immediately upstream of the poly (A) tail is a hypervariable (HVR) region whose sequence and predicted structure are not well conserved. This hypervariable region has recently been shown to function as a virulence enhancer but can be deleted with minimal impact on replication in tissue culture cells (*93*). The dispensable nature of the hypervariable region has been utilized for experiments outlined in Chapter IV. The remaining elements located immediately after the N protein are two highly conserved structures, a bulged stem loop and a pseudoknot that have been extensively studied using enzymatic probing and mutational replication analysis.

A remarkable feature of the conserved bulged stem loop and pseudoknot is that the 3' side of the lower stem in the 5' most hairpin, denoted stem P0a in Figure 14, are also complementary to loop nucleotides in the downstream P2 stem forming a new stem we have designated $P1^{pk}$. As can be seen, Watson-Crick base pairing between stem P0a and the alternate pseudoknot stem $P1^{pk}$, are mutually exclusive and require that only one

conformation be present at any one time. Extensive reverse genetics experiments have confirmed the presence and necessity of both conformations supporting the idea that both helical stems must form at some point during the viral replication cycle (*93-95*). Since the two conformations are mutually exclusive it has been proposed that this RNA element represents a "molecular switch" that could regulate differential initiation of RNA replication for either the full length anti-genomic RNA or for the nested set of subgenomic mRNA.

In this chapter, we present RNA unfolding (thermodynamic) experiments designed to determine if both conformations are in fact capable of forming and to use the subsequently derived absolute stability (ΔG^{25}) to characterize the pseudoknot-forming stem, $P1^{pk}$, in the presence and absence of competition from the alternately folded P0a stem. In order to accomplish this we have used a "divide and conquer" approach that characterizes individual subdomains excised from the entire 97-nt UTR fragment. Data collected from the individual subdomains is then used to interpret experiments from the entire 97-nt UTR fragment.

In subsequent chapters we probe the conformations of selected UTR-containing RNAs using NMR (Chapter III) and fluorescence resonance energy transfer (FRET) along with analytical ultracentrifugation (AUC) (Chapter IV), methods that provide additional insight into this conformational switch.

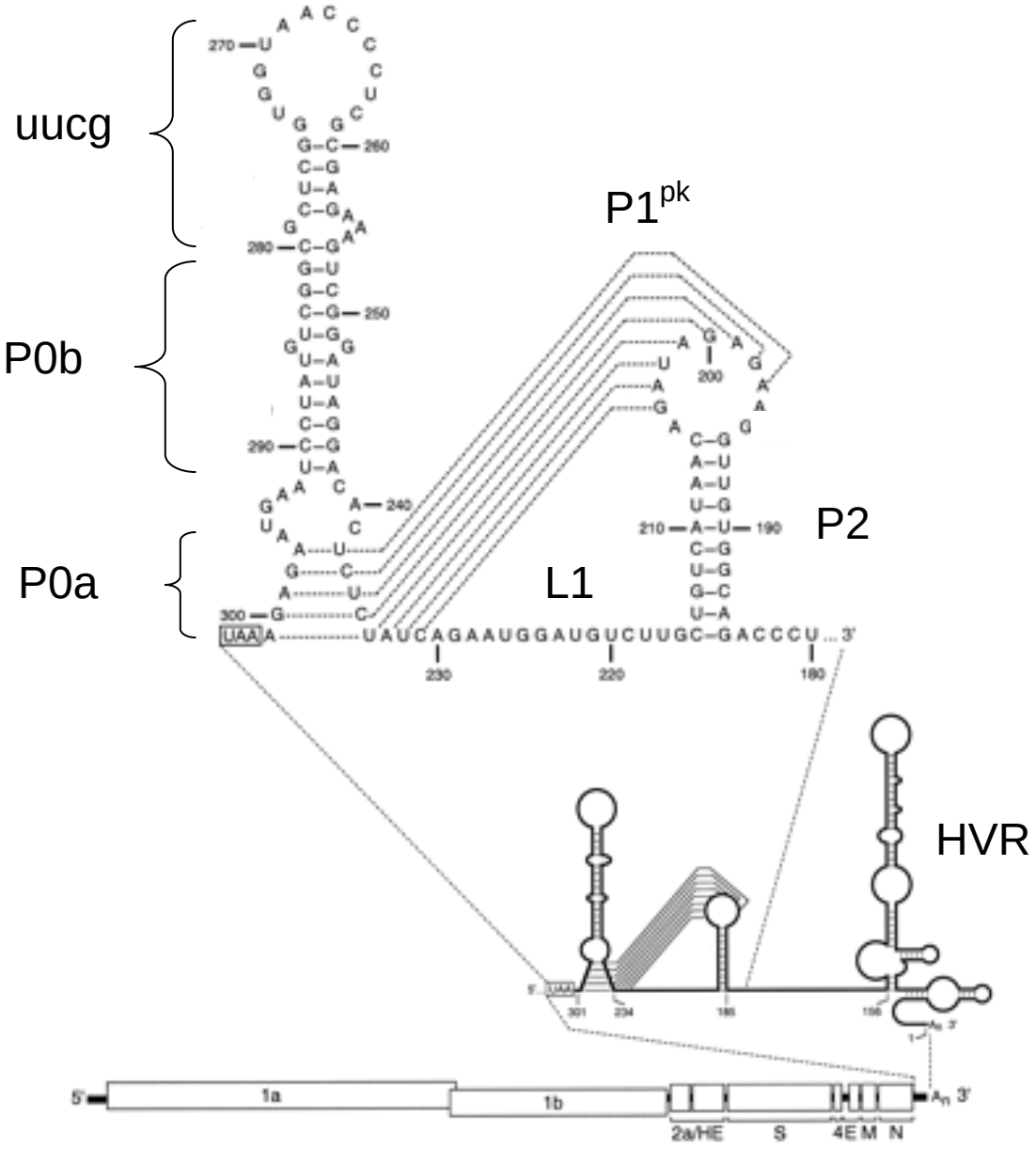

Figure 14 Schematic of experimentally defined secondary structures present in the 3' UTR of MHV. It has been shown that the upper stem loop is not required for viral replication so we have replaced this loop with a uucg tetraloop sequence. The stem just below the uucg tetraloop has been designated P0b and the competing lower stem is designated P0a. The pseudoknot stem is labeled as P1pk. Stem 2 of the pseudoknot is labeled as P2. (Figure taken from reference (*94*).

MATERIALS AND METHODS

RNA preparation. RNA constructs were prepared using run off transcription from synthetic double stranded DNA templates or blunt-end digested plasmids. Both SP6 and T7 RNA polymerase were used depending on the construct. Transcription conditions for individual templates were concentration optimized for the concentrations of nucleotides (2 mM – 10 mM), RNA polymerase (2 μL – 6 μL of prepared stock solution), and magnesium (10 mM – 20 mM) prior to large scale synthesis. All transcription reactions contained 5 μM template, 0.05 g/mL of 8000 MW polyethylene glycol, 4 units/mL of inorganic pyrophosphatase, and 1X transcription buffer (40 mM Tris-HCl (pH 8.0), 5 mM dithiothreitol, 1 mM spermidine, 50 μg/mL bovine serum albumin). Reactions were incubated at 40 °C for SP6 and 37 °C for T7 for 8 hours and quenched with a 1/10 volume of 0.5 M ethylenediamine tetraacetic acid (EDTA) (pH 8.0) and precipitated with absolute ethanol stored at -80 °C overnight. Centrifuged RNA pellet was resuspended in a solution of 7 M urea and 50 mM EDTA. Crude RNA is purified using denaturing polyacrylamide gel electrophoresis (PAGE) using 20% (19:1 bis-acrylamide:acrylamide), 7 M urea, 1X Tris:borate:EDTA (TBE) acrylamide gels run at 250 volts for 8-14 hours depending on size of RNA. Gel bands were visualized using UV shadowing and electroeluted in a 1X TBE solution using an S&S electroeluter. Recovered RNAs were buffer exchanged repeatedly using Millipore Amicon Ultra centrifugal filter devices. Final buffer conditions consisted of 10 mM phosphate pH 6.0 with varying amounts of monovalent KCl and magnesium. Annealing conditions

consisted of heating RNA at 95 °C for 3 min and then slow cooling at room temperature for 45 min.

Thermal denaturation. A Cary spectrophotometer equipped with a temperature controller was using to monitor RNA unfolding simultaneously at 260 nm and 280 nm. RNA samples were between 2 and 4 μM in a total volume of 600 μL and placed in capped 10 mm pathlength cuvettes. The temperature was ramped between 5 and 95 °C in increments of 0.3 °C/min. Collected 260 nm and 280 nm transmittance data were converted to absorbance and a melting profile generated by taking the first derivative of absorbance with respect to temperature ($\partial A/\partial T$). The melting profiles were fit to a multiple sequential interacting transitions folding model that generated a value for t_m and ΔH for each of the transitions using the program t-melt. Both sets of derivative data ($\partial A_{260}/\partial T$ and $\partial A_{280}/\partial T$) were subjected to a simultaneous nonlinear least squares fit to a multiple sequential interacting two-state transition model that optimizes ΔH_i and $t_{m,i}$ for the unfolding of each i^{th} unfolding transition implemented by the algorithm t-melt, $\Delta C_p^o = 0$ (*96, 97*). The unfolding entropies were obtained from $\Delta S = \Delta H/t_m$ while $\Delta G^{25} = \Delta H - 298.15*\Delta S$.

RESULTS

Thermal denaturation of the MHV PK RNA. The minimal MHV 3' UTR fragment that contains all components of the proposed molecular switch is 97 nts and has been designated UTR (see Fig. 15A). The limiting conformations of this switch are conformations I and IV (see Fig. 15B), with conformations II and III depicting possible

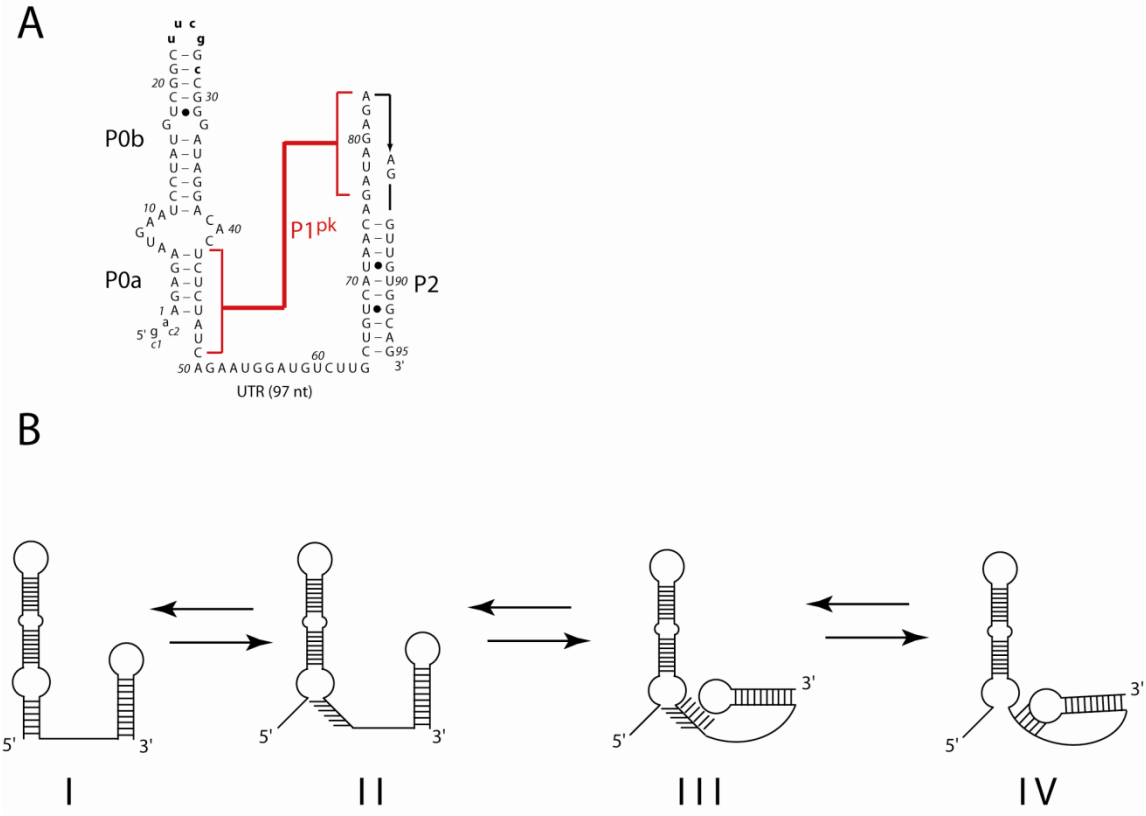

Figure 15 Secondary structure and model of 3' UTR molecular switch. A) Secondary representation of the minimal MHV 3' UTR fragment that contains all components of the proposed molecular switch and represents the RNA sequence utilized for experimentation. The 5' lower case ga are not native to the UTR but were added for transcription purposes. The UTR construct is 97 nucleotides and is referred to as UTR RNA. B) Proposed diagram depicting the possible steps involved in converting between the alternate conformations of the UTR. The limiting conformations of this switch are conformations I and IV with conformations II and III serving as possible intermediates during the interconversion.

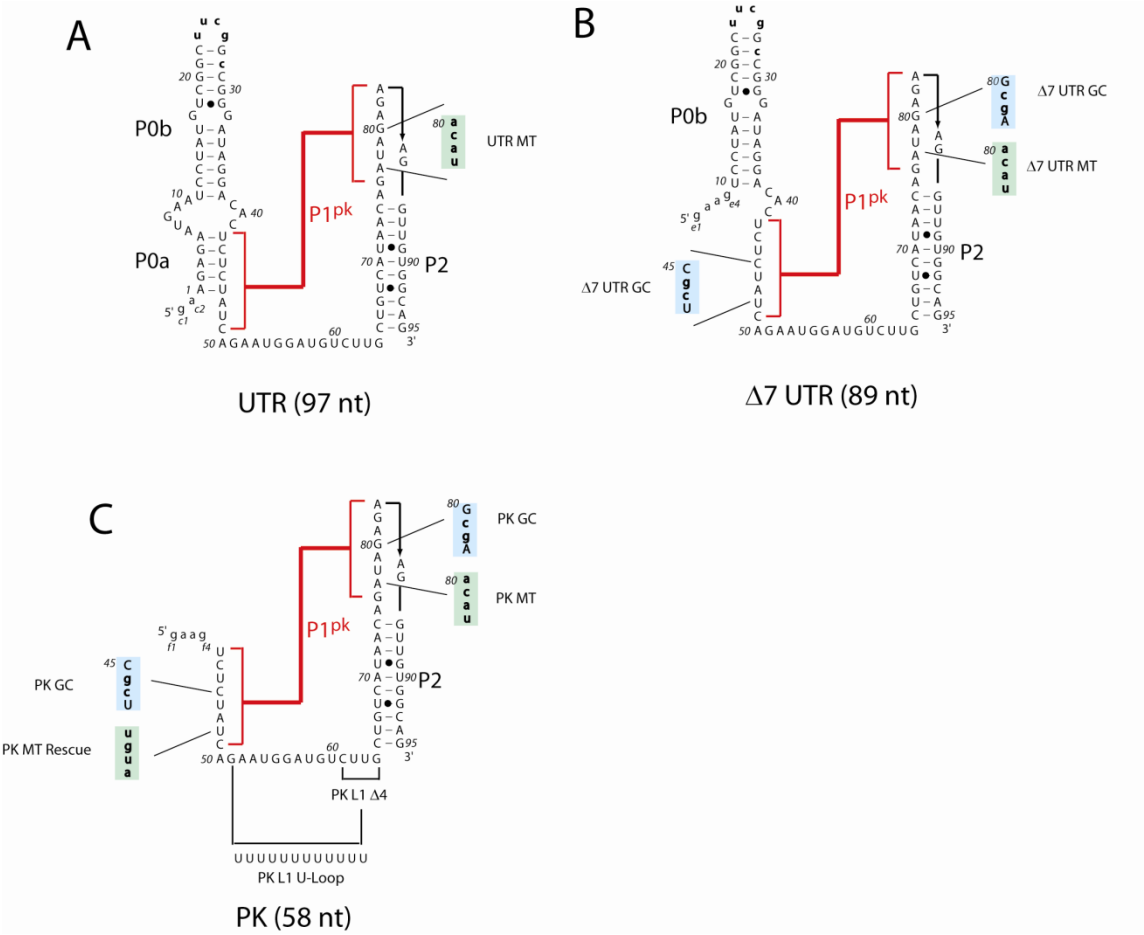

Figure 16 Diagram of constructs used in analysis of UTR RNA. A) UTR RNA construct. The expanded sequence represents the mutations used in the UTR MT construct which eliminates P1^{pk} pairing. B) Δ7 UTR RNA. Expanded sequences display the Δ7 UTR MT construct which disrupts P1pk pairing and Δ7 UTR GC replaces two A-U base pairs with two G-C base pairs. C) PK RNA. Expanded sequences display the PK MT which disrupts P1pk pairing and PK GC replaces two A-U base pairs with two G-C base pairs. In addition loop L1 mutants are indicated as PK L1 U-loop and PK L1 Δ4 which replace most of L1 with uridines and delete the first 4 nucleotides from the 3' end of L1, respectively.

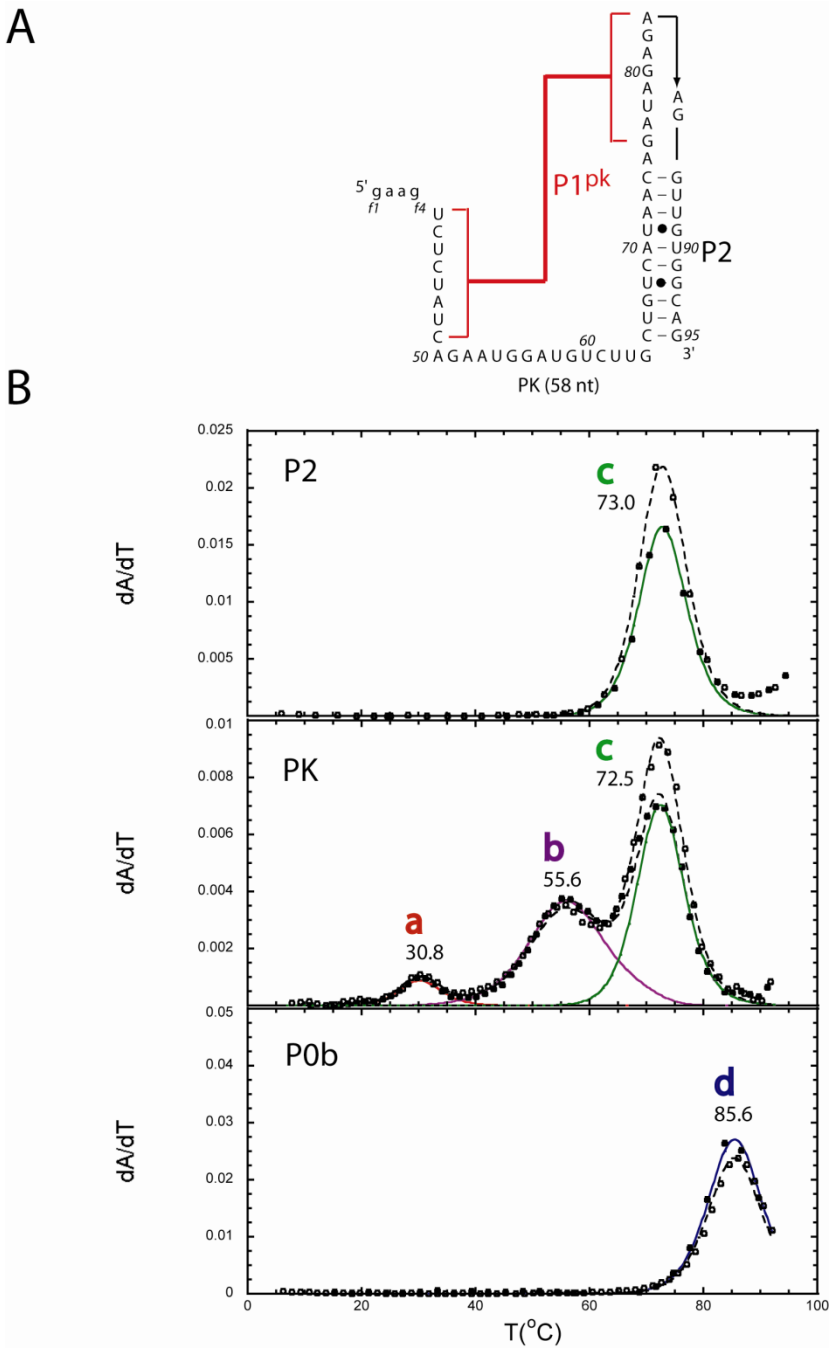

Figure 17 Thermal unfolding spectra of PK, P0b and P2 RNAs. A) Schematic of PK RNA secondary structure and sequence. B) Thermal melting profiles for P2, PK and P0b. Solid curves represent fitted curves of each transition with 260 and 280 nm data depicted by closed and open circles respectively. The solid unfolding curves are color coded according to red for transition a, purple for transition b, green for transition c and blue for transition d. Solution conditions were 10 mM phosphate buffer, pH 6 with 100 mM KCl and 5 mM Mg^{2+}.

intermediates in the interconversion (see Fig. 15B). Using this model of UTR folding we have designed various constructs that will stabilize various intermediates in the pathway. The UTR (97 nts) and two primary subdomain fragments denoted Δ7 UTR (89 nts) and PK (58 nts) were prepared and characterized. Note that both Δ7 UTR and PK RNAs prevent the competing stem P0a from forming. Individual component helical domains P0b, P0ab, and P2 were also prepared to enable their folding assignments in more complex RNAs (see below). Mutant RNAs based on the UTR, Δ7 UTR and PK constructs were designed to abrogate, stabilize or otherwise alter base pairing composition in either $P1^{pk}$ or P2 stems as indicated. In addition loop L1 substitutions and deletion mutations were prepared to first rule out the possibility of thermal unfolding transitions resulting from loop structures (PK L1 U-loop in which all of L1 loop nucleotides were converted to uridine residues, see Fig. 16C) and second to investigate loop length and $P1^{pk}$ pairing (PK L1 Δ4 in which four nucleotides were deleted from the 3' most sequence of L1, see Fig. 16C). The conformational integrity and homogeneity of each RNA were routinely checked by PAGE and imino proton NMR as outlined in Chapter III.

Thermal melting profiles are shown for PK RNA compared to that of the isolated stems P2 and P0b (see Fig. 17B) collected in the presence of 10 mM potassium phosphate buffer pH 6, 100 mM KCl and 5 mM $MgCl_2$. As can be see in panel B of Figure 17, the PK RNA is composed of three unfolding transitions, labeled as **a** (t_m = 30.8 ^oC), **b** (t_m = 55.6 ^oC) and **c** (t_m = 73.0 ^oC). Each melting profile obtained at 260 and 280 nm is simultaneously fit to a sequential two-state unfolding model that generates a

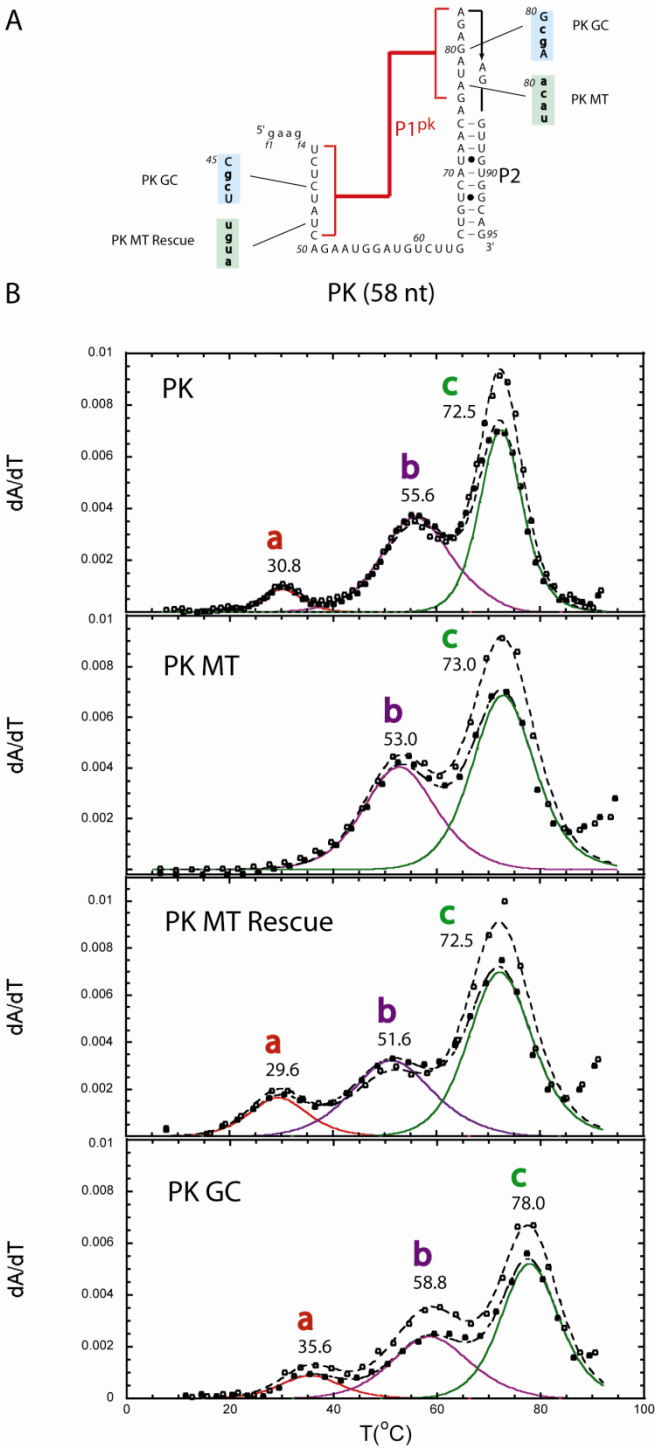

Figure 18 Thermal unfolding spectra of PK, PK MT, PK MT rescue and PK GC RNAs. Constructs designed to alter the folding of $P1^{pk}$. A) A schematic of the designed constructs displayed with mutations in the expanded regions. B) Thermal unfolding profiles of $P1^{pk}$ mutants displayed according to the conventions described in Fig. 27. Solution conditions were 10 mM phosphate buffer, pH 6 with 100 mM KCl and 5 mM Mg^{2+}.

Table 1 Thermodyanamic data for UTR constructs.

RNA	a			b			c			d			d*		
	t_m	ΔH	ΔG^{25}	t_m	ΔH	ΔG^{25}	t_m	ΔH	ΔG^{25}	t_m	ΔH	ΔG^{25}	t_m	ΔH	ΔG^{25}
P2	-	-	-	-	-	-	73.0	85.1	11.8	-	-	-	-	-	-
P0b	-	-	-	-	-	-	-	-	-	85.6	81.1	13.7	-	-	-
P0ab	-	-	-	-	-	-	-	-	-	-	-	-	81.5	91.1	14.5
MHV PK	30.8	50.6	1.0	55.6	67.7	6.3	72.5	77.2	10.6	-	-	-	-	-	-
MHV PK MT	-	-	-	53.0	45.4	3.9	73.0	58.4	8.1	-	-	-	-	-	-
MHV PK Rescue	29.6	48.3	0.7	51.6	40.8	3.3	72.5	57.1	7.9	-	-	-	-	-	-
MHV PK GCS1	35.7	49.0	1.7	58.8	44.1	4.5	78.1	64.6	9.8	-	-	-	-	-	-
SARSpk	24.6	72.4	-0.1	53.7	52.8	4.6	72.8	54.4	7.5	-	-	-	-	-	-
MHV PK U Loop	31.2	70.8	1.4	55.8	53.0	5.0	73.6	77.2	10.8	-	-	-	-	-	-
MHV PK LΔ4	17.2	25.5	-0.7	52.9	40.7	3.5	72.1	77.8	10.6	-	-	-	-	-	-
MHVpk90	30.4	57.8	1.0	56.1	45.5	4.3	85.3	87.5	14.7	-	-	-	-	-	-
MHVpk87	29.4	65.4	1.0	54.3	30.5	2.7	87.0	84.6	14.6	-	-	-	-	-	-
MHVpk44	44.5	29.7	1.8	63.2	52.4	6.0	83.8	80.2	13.2	-	-	-	-	-	-
MHVpk47	42.8	30.8	1.7	64.7	41.1	4.8	83.4	84.2	13.8	-	-	-	-	-	-
MHV Δ7UTR *	37.6	26.2	1.1	59.8	55.1	5.8	72.8	73.5	10.2	85.1	119.5	20.1	-	-	-
MHV GC Δ7UTR	20.2	41.8	-0.7	63.2	43.5	4.9	77.5	105.0	15.7	85.2	101.6	17.1	-	-	-
MHV Δ7UTR MT	-	-	-	50.8	20.7	1.6	74.3	68.6	9.7	90.3	82.0	14.7	-	-	-
UTR	29.3	32.5	0.5	-	-	-	70.1	67.3	8.8	-	-	-	79.2	79.8	12.3
Intact UTR MT	33.5	31.0	0.9	-	-	-	73.8	32.2	4.5	-	-	-	77.8	53.8	8.1

- All sample conditions were run at 10 mM potassium phosphate pH 6, 100 mM KCl and 5 mM $MgCl_2$.
- * Samples run at 10 mM potassium phosphate pH 6, 100 mM KCl and 30 mM $MgCl_2$.

t_m and van't Hoff ΔH (see Table 1 for fitted parameters) for each unfolding transition. Thermal melting profiles are presented with solid, color-coded curves derived from parameters determined by the model and drawn through each unfolding step. The data absorbance is drawn according to continuous dot-dash curves with filled circles used for 260 nm and open circles for 280 nm.

A comparison of thermal melts for PK and P2 allows the assignment of transition **c** to the unfolding of P2, given the close correspondence of thermodynamic parameters associated with the unfolding of this stem. Additionally, P0b stem (transition **d**) melts at a higher t_m compared to P2 and is obviously not found in the PK RNA unfolding profile. After assignment of transition **c** to P2 there are still two transitions remaining, **a** and **b**, that must be reporting on some aspect of PK unfolding, presumably $P1^{pk}$ and or possibly a loop L1 unfolding transition.

To determine if a $P1^{pk}$ transition is present a series of mutations were generated in an effort to specifically alter $P1^{pk}$ base pairing. The PK MT RNA substituted the pseudoknot base pairs with the mutations used by Brian et al. (*98*), designed to disrupt the formation of $P1^{pk}$. Two additional RNAs were also analyzed one designated PK MT Rescue, incorporates the compensatory base pairing mutations within the PK MT RNA to restore $P1^{pk}$ pairing and the other replaces two A-U base pairs within the projected $P1^{pk}$ stem with two G-C base pairs taking advantage of the hyperchromic shift observed for the unfolding of A-U base pairs as compared to G-C base pairs. An increase in 260 nm absorbance is observed for the unfolding of A-U base pairs as compared to G-C base

pairs that have an increase in 280 nm absorbance (*96, 97*). A stack plot of thermal melting profiles for these RNAs is shown in Fig. 18B.

The thermal unfolding profile for PK MT (Fig. 18B) results in two transitions that align with transitions **b** and **c** of the wild-type PK RNA, with transition **a** missing. Thermal unfolding of PK MT rescue RNA, which restores $P1^{pk}$ base pairing within the context of the PK MT, contains three transitions and aligns with to transitions **a**, **b** and **c** of the wildtype PK RNA. Finally, inspection of the PK GC RNA reveals the presence of all three transitions with transitions **a** and **b** characterized by an increase in absorbance at 280 nm as expected for a A-U to G-C replacement (*99*). Results from all three RNAs unambiguously indicate that transition **a** is reporting on unfolding of the $P1^{pk}$. Transitions **a** and **c** are therefore assigned to unfolding of $P1^{pk}$ and P2 stems respectively.

In an effort to assign transition **b**, a series of additional RNAs were analyzed to determine the origin of this unfolding transition. The genome of SARS-CoV and other group 2 CoVs have previously been predicted to contain a pseudoknot similar to that of MHV within the 3' UTR (*94*). The thermal unfolding profile of the PK^{SARS} RNA (Fig. 19) indicates three unfolding transitions very similar to that observed in the MHV PK RNA including an equivalent unfolding **b** transition. Note, however that the hyperchromic ratios are different, consistent with the different base pair compositions of each stem between RNAs. Potential Loop L1 base pairing interactions in the MHV PK RNA were probed using PK L1 U-loop and PK L1 Δ4 RNA constructs. PK L1 U-loop replaces all but five of the residues in L1 with uridines,that might be expected to

eliminate any potential secondary structures present within L1. For example a 4 base pair stem, helix P3, is possible even in the U-loop RNA since a G51-C61 base pair could be swapped for a G51-U61 non-canonical base pair and an A53-G55 non-canonical base pair could be swapped to the canonical A53-U55 base pair (see Fig. 19C). To test this idea, the PK L1 Δ4 RNA deletes the 3' most 4 nts in L1 and should also prevent secondary structure from forming in L1. The thermal unfolding profiles of PK L1 U-loop RNA shows that PK L1 U-loop contains all three unfolding transitions with nearly identical t_m values when compared to the wildtype PK (see Table 1). The PK L1 Δ4 RNA also shows three transitions, however, transition **a** is destabilized, with little effect on transition **b**, to t_m = 17.2 ^oC as compared to t_m = 30.8 for PK transition **a** (see Table 1). The major difference between PK L1 U-loop and the wildtype is a lower ΔH_b and ΔG^{25}_b for the **b** transition with virtually no effect on the **a** or P2 unfolding transitions as expected. The PK L1 Δ4 profile results in a marked destabilization of both transitions **a** and **b** with a $\Delta\Delta G^{25}$ of 1.7 kcal/mol and 2.8 kcal/mol respectively as compared to wildtype PK. Thermal unfolding of a PK L1 Δ8 RNA, which deletes the 3' most 8 nucleotides within L1, result in a complete absence of transition **a** with little or no effect on transition **b** or **c** (data not shown).

Mutations designed to alter the stabilities of $P1^{pk}$ and P2 separately were also analyzed using thermal melts. Constructs which stabilize $P1^{pk}$ interactions included PK 44 and 47 (see Fig. 20A) while constructs that stabilized P2 were designated PK 87 and 90 (see Fig. 20A). Thermal melts were analyzed under solution conditions containing 10 mM potassium phosphate pH 6, 100 mM KCl in the presence of and in the absence of 5

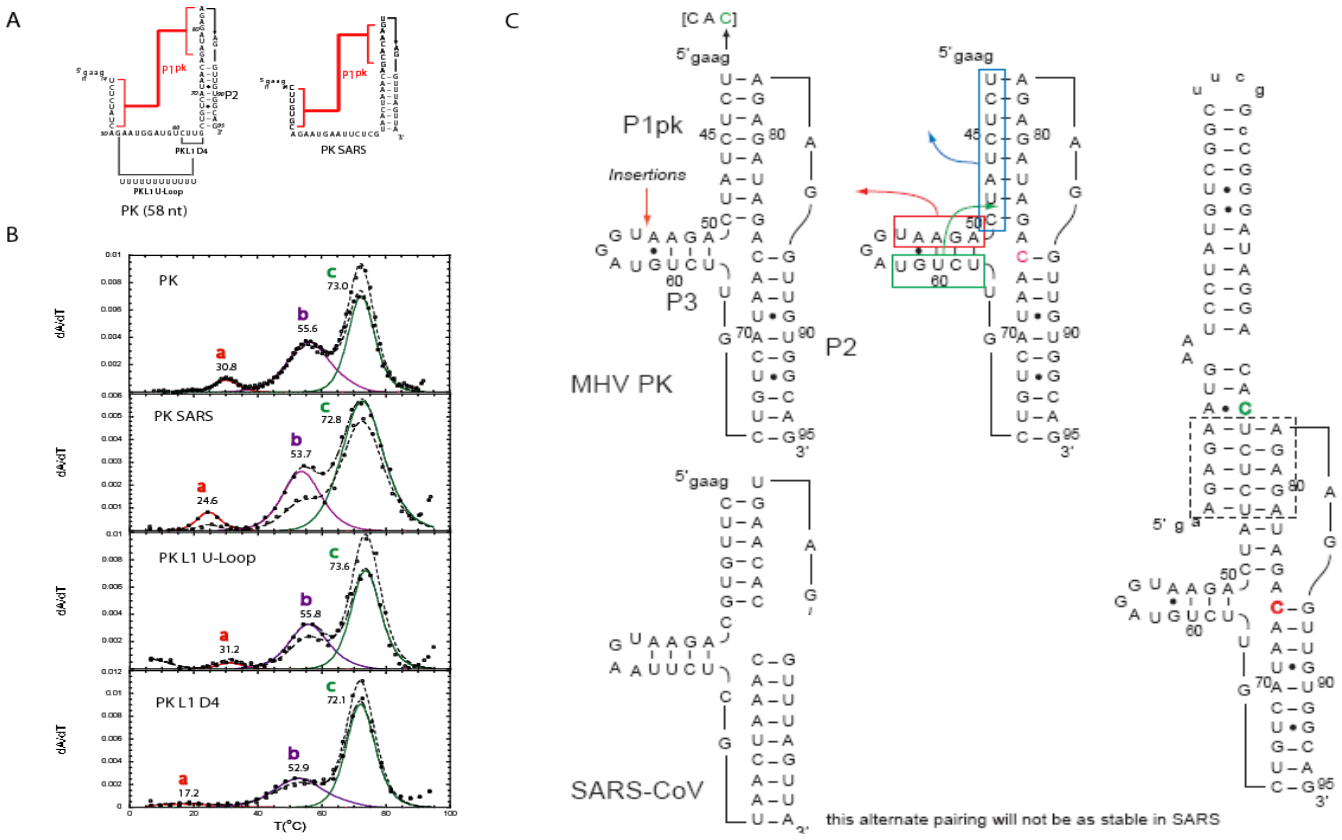

Figure 19 Thermal unfolding spectra of L1 mutations and proposed alternate pseudoknot model. A) Secondary structure and sequence of PK RNA with loop L1 mutations designated in the expanded regions. Beside the PK RNA is the secondary structure and sequence for PK SARS. B) Thermal unfolding profiles PK and PK mutants L1 U-loop and L1 Δ4. C) Secondary structure of proposed alternate pseudoknot folding. The top row of constructs represents the MHV PK RNA. Green, red and blue arrows indicate the conformational switch of the L1 to produce and alternately paired pseudoknot structure. The lower pseudoknot sequence represents the SARS PK which is proposed to be structurally similar to MHV PK and can most likely follow the same alternate pseudoknot folding. Beside the MHV PK constructs is the full length UTR RNA. As diagramed the alternate pseudoknot folding is not possible as a result of competition with stem P0a. Thermal unfolding profiles follow conventions used in Fig. 17. Solution conditions were 10 mM phosphate buffer, pH 6 with 100 mM KCl and 5 mM Mg^{2+}.

mM $MgCl_2$. Thermal melts were compared to the PK RNA as depicted in Figure 20B. Melting profiles of RNAs that are predicted to alter the stability of $P1^{pk}$ helix, PK 44 and PK 47, are characterized by an increase in hyperchromaticity at 280 nm (see Fig. 20B) as well as an increase in t_m for all three transitions (see Table 1). Mutations predicted to influence P2 pairing, PK 87 and PK 90, on the other hand, resulted in an increase in t_m for transition **c** only as well as a hyperchromic shift in the 280/260 ratio, consistent with a change in the compaction of P2 to more G-C base pairing. In contrast, transitions **a** and **b** remain relatively unaffected although, the apparent free energy of unfolding of the **b** transition is decreased somewhat (see Table 1).

Thermal melts of the PK RNA indicate that $P1^{pk}$ and P2 are folded consistent with the proposed pseudoknot model. Thermodynamic analysis of $P1^{pk}$ resulted in an unfolding transition occurring at a very low t_m indicating that $P1^{pk}$ is not very stable at room temperature. Mutational stabilization of $P1^{pk}$ results in a hyperchromic shift for transition **a** and **b** with a subsequent increase in t_m and calculated ΔG^{25} (see Table 1) as expected for the addition of G-C base pairs. Interestingly, although the mutations were only in the $P1^{pk}$ stem an increase in the t_m of transition **c** is also observed, suggesting there is some cooperativity of unfolding between all three transitions. In contrast, stabilization of P2 results in only a hyperchromic shift and subsequent increase in t_m for transition **c** indicating independent unfolding for P2 with no influence from transitions **a** or **b**.

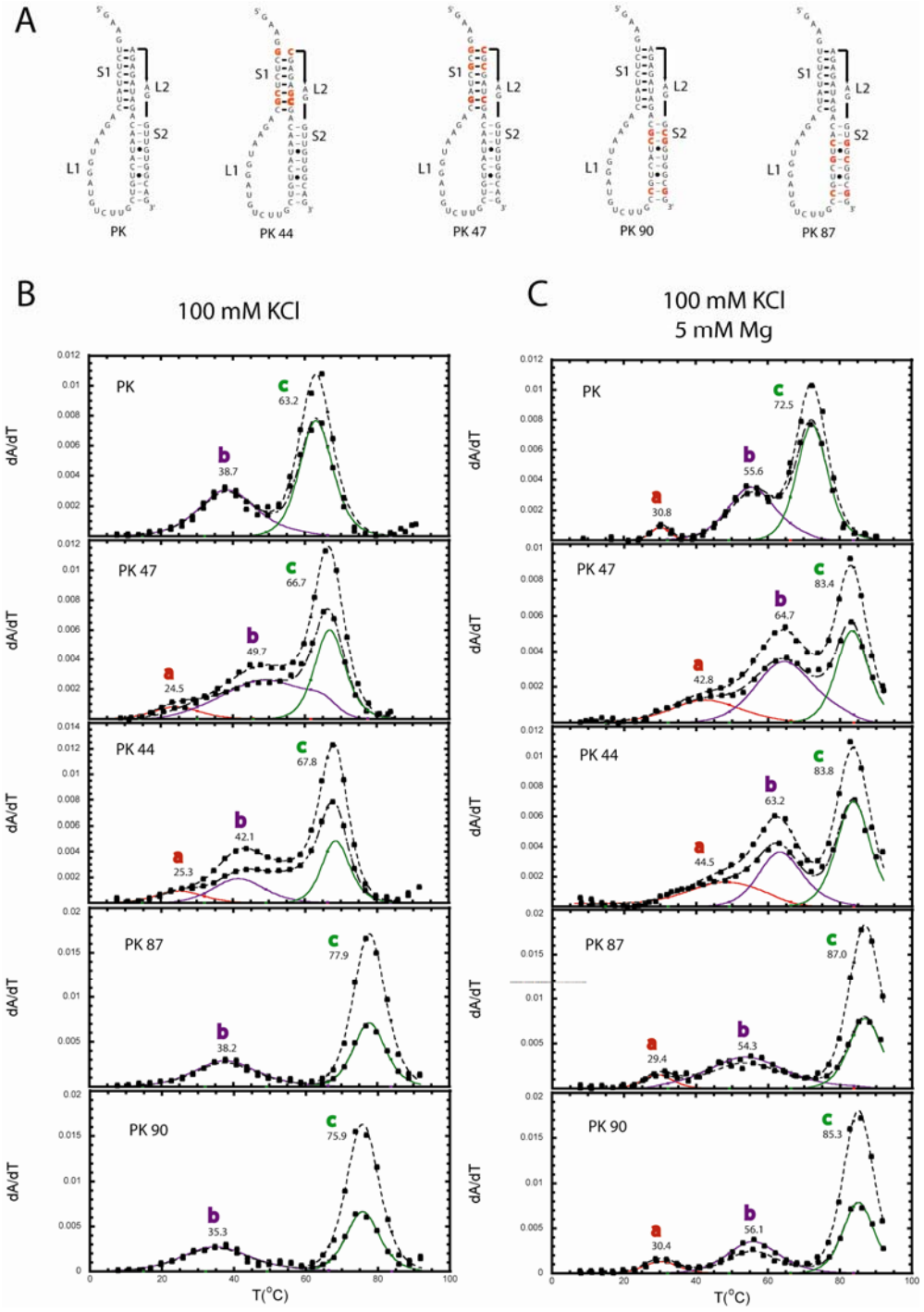

Figure 20 Thermal unfolding spectra for mutations involving stabilization of $P1^{pk}$ (S1) and P2 (S2). A) Diagram of pseudoknot folding for mutations where S1 represents $P1^{pk}$ and S2 represents P2. B) Thermal unfolding spectra of PK mutants. Spectra were analyzed under the conditions 10 mM potassium phosphate pH 6 with 100 mM KCl only or 100 mM KCl with 5 mM $MgCl_2$. Thermal unfolding profiles follow conventions used in Fig. 17.

Taking all of the PK thermal melts in total, we propose the following three transition unfolding model for the PK RNA:

$$\text{PK} \xrightarrow{\mathbf{a}} \text{PK*} \xrightarrow{\mathbf{b}} \text{P2} \xrightarrow{\mathbf{c}} \text{U}$$

As yet the nature of PK* cannot be determined from these data alone. Based on thermal melts it is clear that PK* is not reporting on the unfolding of secondary structure P3 helix in L1 unfolding since the PK L1 Δ4 RNA should prevent the formation of this stem. Since transition **b** is absent in P2 melting profiles unfolding intermediates other than P2 or $P1^{pk}$ base pairing directly. We hypothesize that PK* refers to an alternate pseudoknot conformation that pairs the 3' end of L1 with the same P2 hairpin loop and can only form when $P1^{pk}$ pairing is lost (see Discussion).

Thermal denaturation of MHV Δ7 UTR. Having investigated pseudoknot unfolding in the context of the minimal PK RNA, we next wanted to investigate pseudoknot folding within the context of the larger RNA without competition from the double hairpin conformation. To do this, the 5' most seven nucleotides of the UTR were deleted to prevent the formation of P0a, eliminating competition from the double hairpin conformation but leaving pseudoknot base pairing unaffected. The resulting construct, Δ7 UTR (89 nt) was analyzed according to similar experiments conducted for the PK RNA.

Examination of thermal melts obtained at 30 mM Mg^{2+}, pH 6 for Δ7 UTR RNA reveals four unfolding transitions, labeled **a-d** in Figure 21. Relative to PK RNA, the Δ7 UTR RNA melting profile differs in that transition **a** is very broad and only appears

under conditions of very high magnesium. In addition, a high t_m unfolding transition labeled transition **d** is present that is readily assigned to unfolding of P0b (see Fig. 21). Likewise, transition **c** can be identified as P2 unfolding. Transition **b** resembles the unfolding transition assigned to PK* from MHV PK unfolding.

To investigate the unfolding transitions further the same $P1^{pk}$ stabilizing mutations were incorporated into the Δ7 UTR RNA. Resulting thermal melts of construct the Δ7 UTR GC RNA reveals the same hyperchromic and t_m shifts observed for the PK RNA. Transitions **a** and **b** increase in 280 nm absorbance (see Fig. 21B) and transitions **a**, **b**, and **c** are each characterized by an increase in t_m relative to the wildtype Δ7 UTR RNA (see Table 1). Interestingly, comparison of the thermal melts of the Δ7 UTR GC RNA with the PK GC RNA reveals that transition **a** is destabilized by 1.0 kcal*mol^{-1} (100 mM KCl and 5 mM $MgCl_2$) (see Table 1).

Taking into account the broad nature of transition **a** in the wildtype Δ7 UTR RNA and the lower stability of the **a** transition in the two GC RNAs reveals that formation of the P0b stem although not directly involved in $P1^{pk}$ base pairing, destabilizes the formation of $P1^{pk}$ helix in the context of Δ7 UTR RNA relative to the PK RNA.

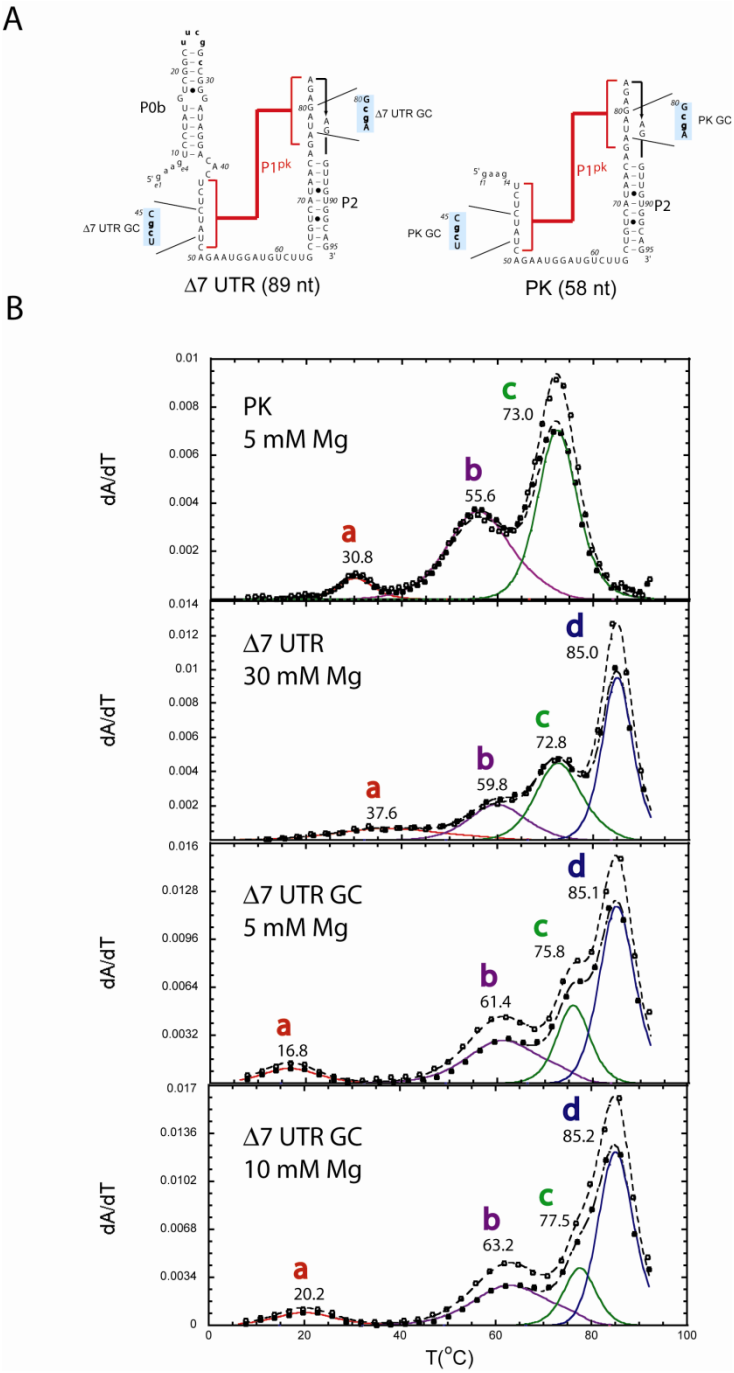

Figure 21 Thermal unfolding spectra of Δ7 UTR and mutant RNAs. A) Schematic secondary folding model and sequence of Δ7 UTR and PK RNAs with associated mutant constructs illustrated as expanded regions. B) Thermal unfolding spectra of PK, Δ7 UTR, and Δ7 UTR GC. Thermal unfolding profiles follow conventions used in Fig. 17. Solution conditions were 10 mM phosphate buffer, pH 6 with specified Mg^{2+} concentrations.

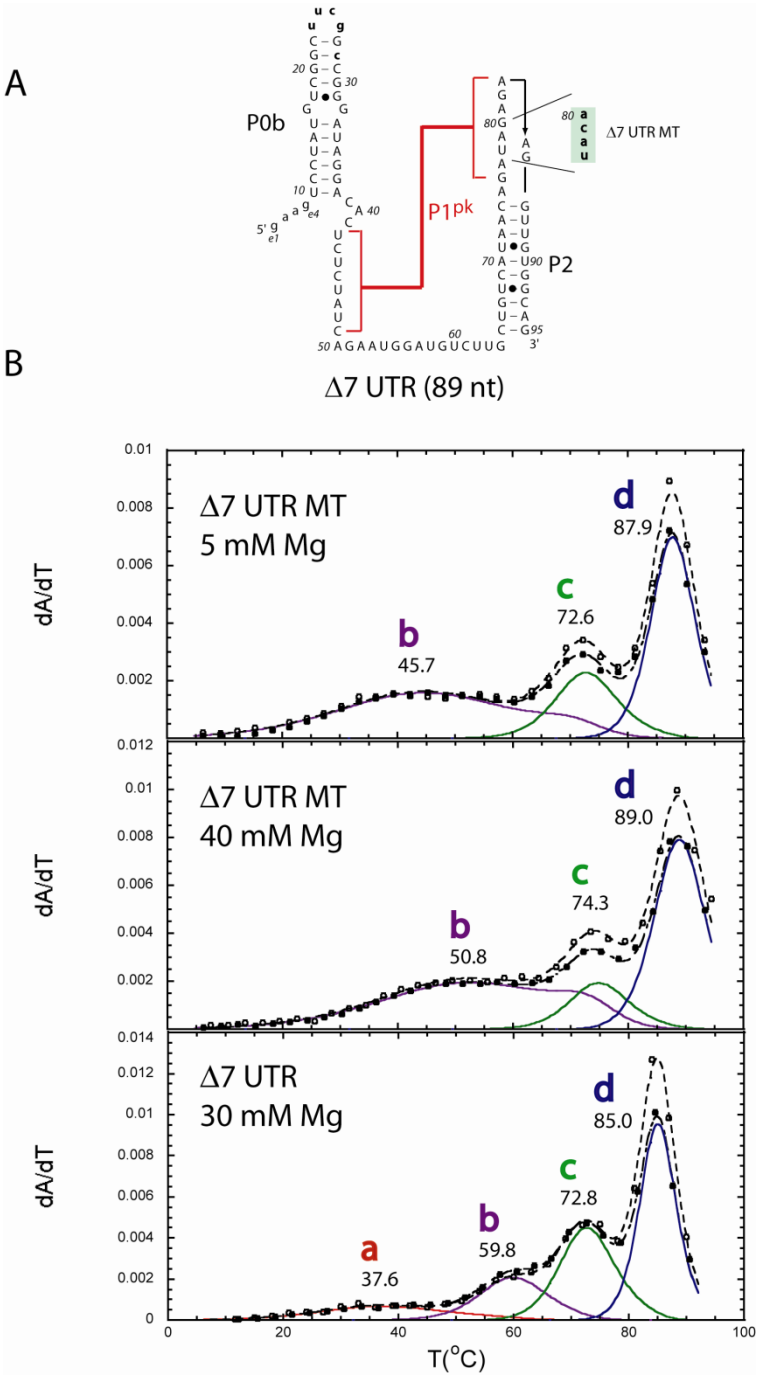

Figure 22 Thermal unfolding spectra of Δ7 UTR MT RNA. A) Schematic of Δ7 UTR secondary structure and sequence with mutants labeled as expanded regions. B) Thermal unfolding profiles comparing Δ7 UTR and P1^{pk} mutant Δ7 UTR MT. Thermal unfolding profiles follow conventions used in Fig. 17. Solution conditions were 10 mM phosphate buffer, pH 6 with specified Mg^{2+} concentrations.

Further investigation of $P1^{pk}$ pairing was conducted within the Δ7 UTR RNA by analyzing the unfolding profile of an equivalent pseudoknot stem mutant analyzed to probe for $P1^{pk}$ pairing. The Δ7 UTR MT RNA thermal melting profile is markedly perturbed relative to the Δ7 UTR RNA (see Fig. 22B). The melt consists of unfolding transitions readily assigned to P2 (transition **c**) and P0b (transition **d**) unfolding, and a featureless broad transition that remains unchanged even in the presence of high concentrations of magnesium (see Fig. 22B). The nature of this transition is unknown but magnesium reflects the unfolding of PK* non-native base pairs in the L1 loop. In any case transition **a** is missing in the melts.

Thermal denaturation of the UTR RNA. We next investigated the unfolding profiles of an intact UTR molecule denoted as UTR (97 nts) which allows the formation of the two mutually exclusive competing stems P0a and $P1^{pk}$. These thermal unfolding profiles reveal three unfolding transitions labeled as transition **a**, **c** and **d***. In order to deconvolute this unfolding pattern we compared it to those obtained for an RNA that abolishes $P1^{pk}$ pairing, UTR MT, as well as an RNA containing the full bulged stem-loop, P0ab RNA (see Fig. 23A). Unfolding of the P0ab RNA was modeled as a single unfolding transition labeled **d*** for identification purposes only (no attempt was made to extract independent thermodynamic parameters for P0a). Addition of the P0a stem reduced the apparent t_m of the P0b helix and this is reflected in the thermal melt of UTR RNA and has a t_m of ~80^{o}C versus 85 ^{o}C for the P0b helix. Transition **c** corresponds to P2 unfolding and is found precisely at previously determined t_m values.

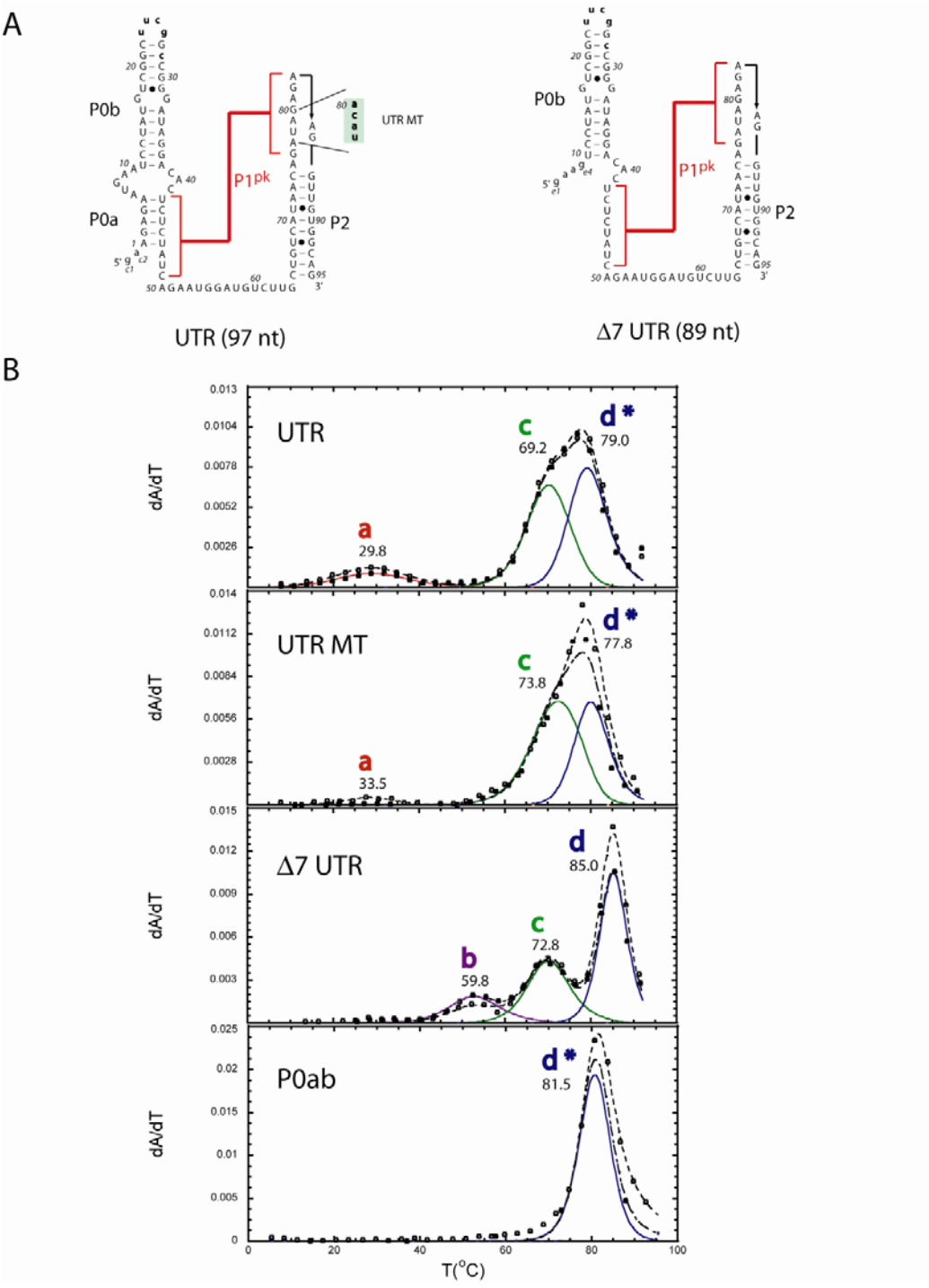

Figure 23 Thermal unfolding spectra of UTR, UTR MT, Δ7 UTR, and P0ab RNAs. A) Secondary structure and sequence of the UTR and Δ7 UTR RNA with expanded regions indicating mutant constructs. B) Thermal unfolding melts of UTR, Δ7 UTR and P0ab. Thermal unfolding profiles follow conventions used in Fig. 17. Solution conditions were 10 mM phosphate buffer, pH 6 with 100 mM KCl and 5 mM Mg^{2+}.

Comparison of the UTR and the UTR MT RNA reveal that the broad transition (t_m = 39 ^oC) is reporting primarily on $P1^{pk}$ pairing under these conditions.

An interesting and unique feature of both the UTR and UTR MT profile is that transition **b** appears completely absent. The important conclusion is that the presence of transition **a** reveals that pseudoknot pairing is possible in the context of the intact UTR RNA model. Somewhat surprisingly, the apparent stability (ΔG^{25}) of the $P1^{pk}$ helix appears relatively unaffected by the presence of absence of competition from the P0a helix although it requires much more Mg^{2+} to drive PK pairing in the Δ7 versus full length UTR RNAs (see Discussion).

Salt dependence of MHV UTR, Δ7 UTR and PK RNAs. The studies presented above reveal that salt (Mg^{2+} and K^+) has a significant effect on the formation and stability of $P1^{pk}$ pairing; depending on the context. To understand the folding requirements, thermal melts were conducted under varying conditions of both monovalent KCl and divalent $MgCl_2$. To first determine if magnesium is absolutely required for formation of $P1^{pk}$ pairing thermal melts were conducted in the presence of a high concentration of KCl (1 M). Thermal melting profiles indicate that PK RNA in 1 M KCl retains the same three transitions (see Fig. 24B) as observed in a solution containing 5 mM Mg and 100 mM KCl (compare with Fig. 17). Further, the UTR also indicates the same number of transitions present (see Fig. 23B) in the 1M KCl spectrum as was observed in 5 mM Mg and 100 mM KCl. This is consistent with Mg^{2+} functioning in a non-specific manner. The presence of transition **a** in solution conditions

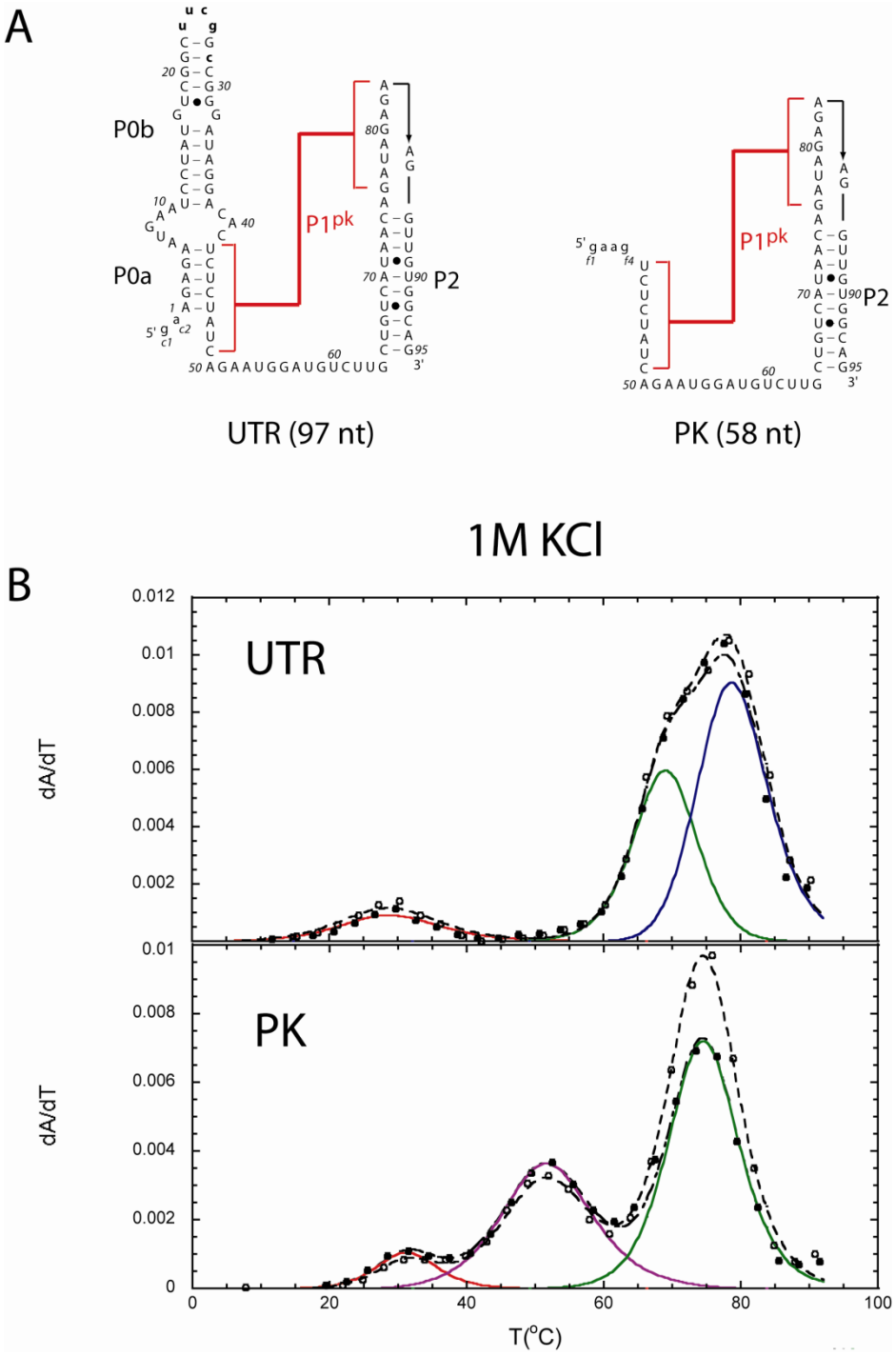

Figure 24 Salt dependence of PK and UTR RNAs. A) Secondary structure and sequence of UTR and PK RNAs. B) Thermal unfolding profiles of PK and UTR RNAs in the presence of 10 mM phosphate buffer, pH 6 with 1 M KCl. Thermal unfolding profiles follow conventions used in Fig. 17.

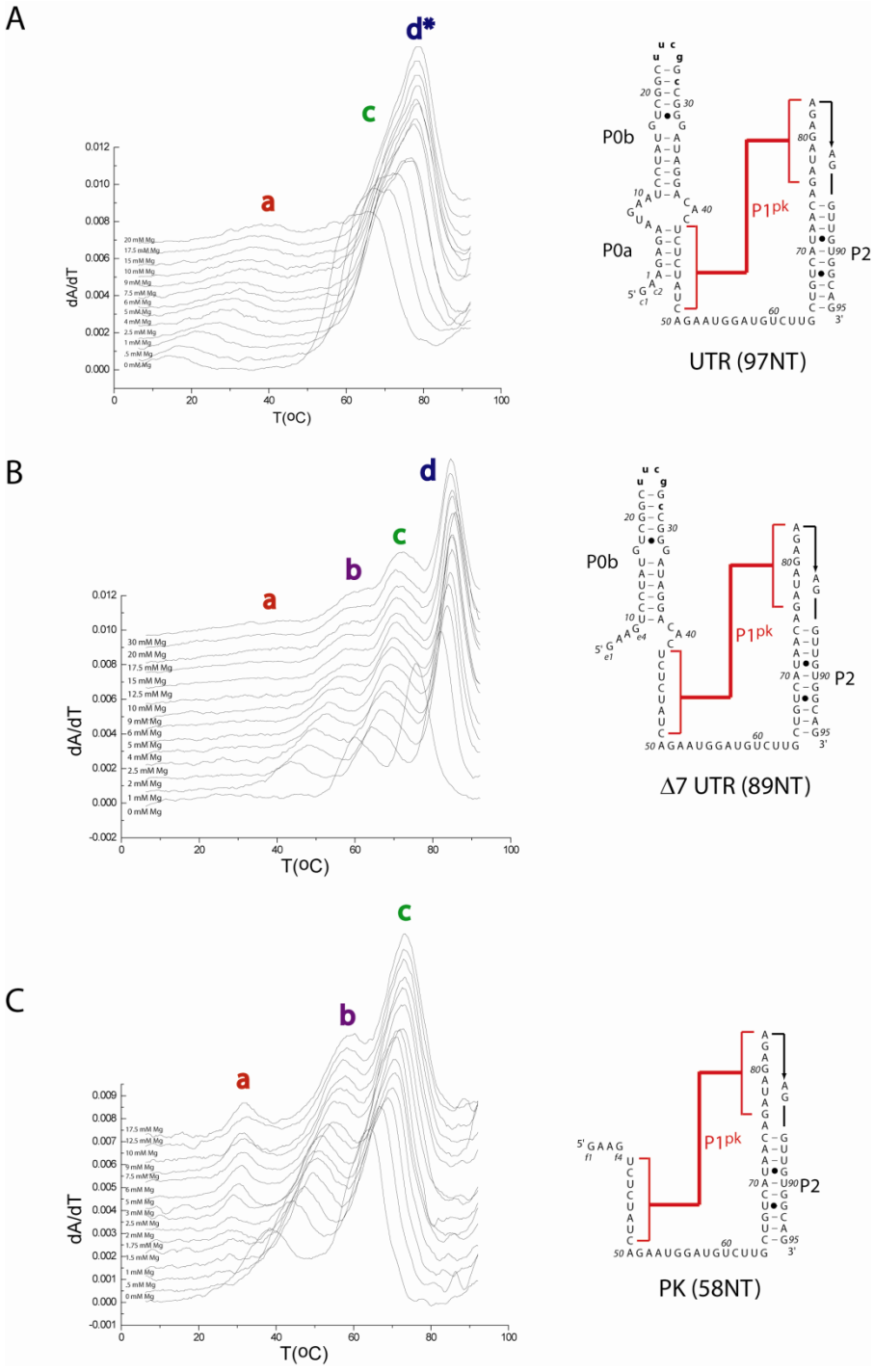

Figure 25 Stack plot of thermal unfolding profiles for UTR, PK and Δ7 UTR RNAs. Spectra collected in 10 mM phosphate pH 6, 100 mM KCl with varying concentrations of MgCl2. A) UTR RNA B) Δ7 UTR RNA C) PK RNA.

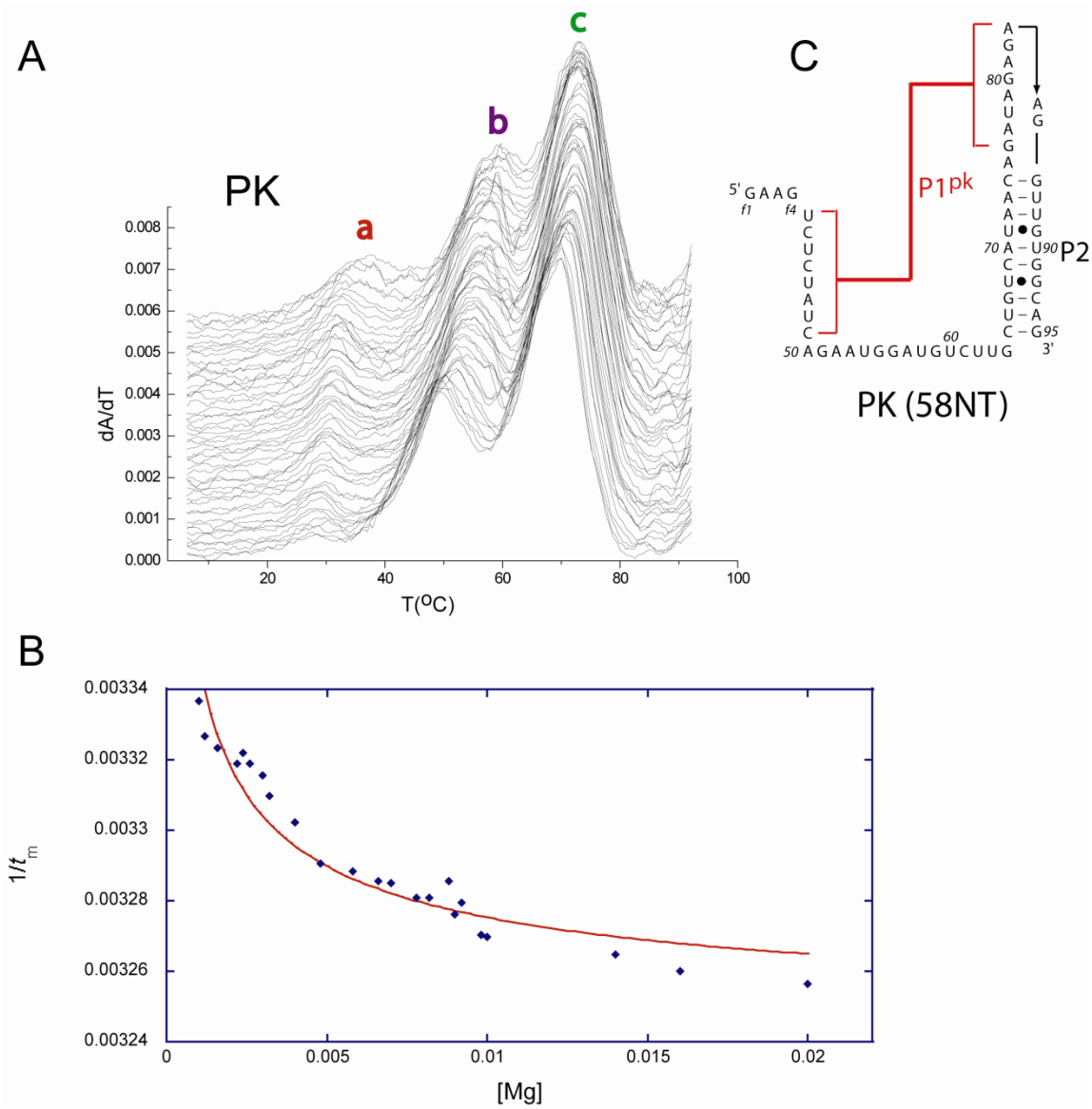

Figure 26 Stack plot of thermal unfolding profiles for PK and analysis of Mg^{2+} binding. A) Stack plot thermal melts of PK RNA used for the analysis of Mg^{2+} binding. Thermal unfolding spectra were measured in 10 mM phosphate pH 6, 100 mM KCl with varying concentrations of $MgCl_2$. B) Magnesium binding curve of PK $P1^{pk}$ base pairing.

only containing monovalent KCl reveals that $P1^{pk}$ folding is not strictly magnesium dependent.

Although the $P1^{pk}$ stem can fold in the absence of magnesium the stability is strongly dependent on the presence of salt. In order to address the salt dependence of $P1^{pk}$ folding, in the context of the three constructs, thermal melts were conducted where magnesium was varied from 0 – 30 mM $MgCl_2$ in a control background of 100 mM KCl. Folding stability is achieved with less salt using magnesium chloride than with only KCl, as expected. Stack plots were generated for the thermal unfolding profiles of each of the constructs with increasing magnesium concentration. Figure 25A indicates that transition **a** begins to form in the presence of 100 mM KCl within the full length UTR RNAs, in contrast a Mg^{2+} concentration of at least 10 mM $MgCl_2$ (see Fig. 25B) is required in the Δ7 UTR RNA before the appearance of the broad transition **a**. For the MHV PK RNA on the other hand transition **a** is clearly present at a concentration of 500 μM $MgCl_2$ (see Fig. 25C). Transition **b** also folds with less magnesium for MHV PK, 100 mM KCl only as compared to MHV Δ7 UTR where it appears fully formed under 2 mM $MgCl_2$. In contrast to other RNAs the full length UTR never appears to form transition **b** (see Fig. 25).

To understand the nature of the magnesium dependence of the unfolding to transition **a**, magnesium titration was conducted with wild type MHV PK using varying amounts of Mg^{2+} (Fig. 26A). The resulting RNA data were fit to a nonspecific binding model for divalent ions which generates magnesium ion affinities for the folded (K_f) and

unfolded (K_u) states of the RNA assuming a model in which one Mg^{2+} ion binds to two phosphate groups (*97*) (Fig. 26B). The affinity for the folded $P1^{pk}$ state, K_f was found to be 3800 +/- 400 (M^{-1}) while the affinity for the unfolded state was found to be 3300 +/- 300 (M^{-1}). These affinities reveal that magnesium binding is approximately the same for both the folded and unfolded states indicating that the nonspecific binding of Mg^{2+} ions to phosphate backbone interaction. In previous studies with model H-type pseudoknots, K_f was found to be significantly larger than the K_u consistent with pockets of deeper negative charge potential due to the close approach of phosphate chains at the helical junction. This seems not to be the case here.

DISCUSSION

Thermodynamic investigation of the MHV 3' UTR molecular switch was conducted using a "divide and conquer" approach that divides the MHV 3' UTR into three constructs PK, Δ7 UTR and UTR RNAs. The PK RNA represents the pseudoknot structure, Δ7 UTR RNA represents the pseudoknot structure as well as the appended P0b hairpin and UTR RNA represents the entire 3' UTR. Thermal unfolding profiles were used as an indicator of secondary structure formation through the analysis of isolated stems as unfolding standards in order to determine the thermodynamics of unfolding within the context of individual RNAs.

Quantitative analysis of the MHV PK RNA melting profiles reveals three unfolding transitions designated **a**, **b** and **c**. Through mutational analysis designed to disrupt the formation of the $P1^{pk}$ stem and through GC stabilization of $P1^{pk}$,

identification of transition **a** as representing the unfolding of the pseudoknot stem was concluded. Comparison of P2 stem unfolding to thermal melts of the PK RNA identified transition **c** as the unfolding of stem P2. The major conclusion from this work is that pseudoknot pairing is extremely unstable and melts at a very low t_m even under conditions where conformational competition has been eliminated and under conditions of high concentrations of magnesium. This is in contrast to other hairpin type pseudoknots such as the MMTV *gag-pro* pseudoknot, which is known to have a folding topology similar to that as MHV PK RNA. MMTV has a pseudoknot unfolding t_m of 73.5 ^oC versus MHV PK which has a t_m of 31.5 ^oC under equivalent 1 M monovalent salt solution conditions. The projected stability of an 8 base pair stem equivalent to the PK pseudoknot sequence is 9.9 kcal/mol. The experimentally calculated $P1^{pk}$ is only 1.0 kcal*mol^{-1}. This would suggest that the $P1^{pk}$ stem is not actually folding all eight base pairs but only incorporate a small fraction of the possible pseudoknot stem. Furthermore, this finding is consistent with what is known about the structural requirements for loop versus stem lengths as in the classic H-type pseudoknot by Pleij et al (see Chapter I - Pseudoknots) (*39*). In a standard pseudoknot topology Loop L1 will cross the deep major groove of stem 2 while loop L2 will cross the wide minor groove of stem 1. To span a pseudoknot stem 1 (here $P1^{pk}$) of 8 base pairs, a loop length of 7-8 nucleotides is required. The PK RNA has a proposed loop L2 length of just two nucleotides. Thus, in order to accommodate an 8 base $P1^{pk}$, the two stems would need to strongly bend towards one another or the actual length of $P1^{pk}$ is 4-5 base pairs rather

than 8 base pairs. Based on thermodynamic analysis of the $P1^{pk}$ stem it would suggest the latter.

While transition **b** could not be unambiguously identified from our mutational analysis, various characteristics of the unfolding transition could be defined. First, it is important to note that the unfolding of the P2 RNA does not contain a thermal unfolding transition for **b** indicating that it is not a function of isolated P2 unfolding. Other mutants designed to investigate loop L1 contributions to transition **b** indicate that a loop sequence that contains only uridine nucleotides (PK L1 U loop RNA) still maintains a **b** transition and further shortening of loop L1 (PK L1 Δ4 and PK L1 Δ8 RNAs) also does not abrogate transition **b**. When the $P1^{pk}$ pairing was mutated to prevent pseudoknot folding, transition **b** was also unchanged. Significant changes were observed in transition **b** in RNA PK mutants that altered the $P1^{pk}$ sequence. Increasing hyperchromaticity of transition **b** was observed concordantly with $P1^{pk}$ mutants PK GC, PK 44 and PK 47 which increased the G-C base pair content of $P1^{pk}$. Taking into account all of these observations we propose that transition **b** represents an alternately folded $P1^{pk}$ pairing that can fold in the context of the isolated pseudoknot construct but not in the presence of the full length UTR RNA (see Fig. 19). This alternate pairing would also be possible even if the entire loop L1 were converted to uridine nucleotides or if nucleotides were deleted from the 3' proximal side of loop L1 (mutants PK L1 Δ4 and PK L1 Δ8). This alternate base pairing is however not possible in the context of the UTR RNA because stem P0a competes for the 5' RNA sequence that is allowing for stem swapping in the isolated PK RNA (see Fig. 19).

Having determined that transition **a** is the unfolding of the pseudoknot stem $P1^{pk}$ in PK RNA, we next wished to investigate $P1^{pk}$ pairing in the context of the Δ7 UTR RNA. Δ7 UTR RNA contains stem $P1^{pk}$ as well as stem P0b and represents a truncated UTR RNA that is unable to form the competitive P0a stem. Thermal unfolding spectra of Δ7 UTR indicate that four transitions are present under conditions that employed large concentrations of magnesium. It was determined that transition **d** represented unfolding of P0b as assessed through the comparison of a standard unfolding spectrum of independent P0b. Based off of the same methodology, transition **c** could be shown as belonging to the unfolding of stem P2. In order to address pseudoknot pairing similar experiments, as were done for PK RNA, were conducted using Δ7 UTR RNA. Δ7 UTR GC RNA utilized the same G-C base pair mutations as conducted for PK RNA in order to determine if an increase in t_m and hyperchromaticity would result. Thermal profiles indicated that a shift in t_m and hyperchromaticity are observed for transition **a.** To further determine which transition is reporting on pseudoknot pairing a mutation was made to disrupt $P1^{pk}$ base pairing in the Δ7 UTR RNA. Δ7 UTR MT RNA has the same $P1^{pk}$ mutations utilized for the PK RNA. Thermal profiles of the Δ7 UTR MT indicate that transition **a** does not form even in the presence of large concentrations of magnesium. Although clear identification of P0b and P2 is possible, no presence of transition **a** is observed. Transition **b** is, however, present as a broad transition consistent with severe non-two state unfolding behavior. Based on the t_m and hyperchromaticity shift as well as the absence of transition **a** in the $P1^{pk}$ mutant Δ7 UTR MT, transition **a** can be identified as reporting on the unfolding of pseudoknot pairing.

Although pseudoknot pairing within the Δ7 UTR RNA could be observed, $P1^{pk}$ interactions were greatly destabilized relative to PK RNA. Calculation of $\Delta\Delta G^{25}$ for transition **a** between Δ7 UTR GC and the equivalent PK GC indicate that $P1^{pk}$ pairing is destabilized by 1.0 kcal/mol. Also, transition **a** within the Δ7 UTR RNA is extremely broad compared to transition **a** in PK RNA sugestive of non-two state unfolding. The origin of this destabilization is not completely clear. Perhaps the addition of P0b imparts a steric folding constraint that is not present in the PK RNA. Loop L1 is much larger than the required number of nucleotides needed to span a 9 base pair P2. Perhaps addition of the P0b stem to loop L1 hinders folding of the $P1^{pk}$ stem. The broadness of transition **a** unfolding indicates that RNA unfolding occurs through multiple pathways indicating a variation in $P1^{pk}$ pairing.

Finally, we investigated pseudoknot unfolding in the context of the intact UTR construct. This construct included $P1^{pk}$ pairing as well as the complete P0ab bulged stem which can form the competing alternate P0a stem in 3' UTR folding. Thermal melts indicate that the UTR RNA has three transitions designated **a**, **c** and **d***. Comparison of equivalent spectra from a standard P0ab stem indicated that transition **d*** represents cooperative unfolding of stem P0ab. Likewise transition **c** was shown as the unfolding of P2. Thermal melts of the UTR MT, an RNA that has abrogated $P1^{pk}$ folding in the same manner as PK MT and Δ7 UTR MT, indicated that transition **a** was greatly decreased illustrating the $P1^{pk}$ unpairing is represented by transition **a**. It is unclear why all of transition **a** is not completely eliminated in the UTR MT thermal melts. One explanation is that the four nucleotide mutation does not completely remove

$P1^{pk}$ pairing or that orientation factors in putative helical junctions could be orienting $P1^{pk}$ stem folding even in the absence of pseudoknot base pairing as has been observed in the folding of the hairpin ribozyme (*100*).

Although specific questions remain, it is clear that pseudoknot pairing exists in the context of each RNA construct. However, even though pseudoknot pairing is present, it is extremely weak and requires substantial additions of monovalent or divalent cations to be sufficiently stabilized to be observed in a melting profile. In addition, even when sufficient salt is added, melting temperatures never exceed 30 ^{o}C observed for the with wild-type $P1^{pk}$ stem in the presence of 100 mM KCl and 5 mM $MgCl_2$. These observations lead to questions regarding physiological function of the pseudoknot in the context of viral replication. Proteins are vital to viral replication and it is possible that protein binding plays a direct role in specifically stabilizing pseudoknot pairing. Genome circularization is a common event in viral replication and it is known that this circularization is mediated through proteins interacting on the 5' UTR with proteins on the 3' UTR. Protein binding in circularization of 3' and 5' ends of coronaviral genomes may also have an influence on the stability of pseudoknot pairing as well.

Finally, a remarkable feature of the coronaviral 3' UTR is the conservation of function in different viruses within the coronavirus family. It is known that the 3' UTR is functionally exchangeable between MHV and BCoV and more importantly between MHV and SARS-CoV (*95, 98*). Indeed, the remarkable similarity of thermal unfolding spectra between MHV PK and SARS-CoV PK RNAs certainly lead to intriguing possibilities that are well worth investigating. To obtain a better understanding of

coronaviral pseudoknot structural requirements, we next employed NMR spectroscopy (Chapter III) and FRET spectroscopy (Chapter IV) (*67, 73, 80*).

CHAPTER III

NMR CHARACTERIZATION OF THE 3' UNTRANSLATED REGION IN MOUSE HEPATITIS VIRUS

INTRODUCTION

In Chapter II, we quantitatively analyzed equilibrium unfolding or melting profiles for a series of RNA constructs derived from the 3' UTR of the MHV genome in an effort to determine the thermodynamic parameters (T_m, ΔH and ΔG^{25}) that characterize pseudoknot ($P1^{pk}$) pairing, as well as the degree to which added Mg^{2+} stabilized this conformation relative to competing conformation(s). These experiments establish the ΔG^{25} for pseudoknot pairing in the minimal PK RNA is –1 kcal mol^{-1} (pH 6.0, 5 mM Mg^{2+}) with a T_m of 31 ºC. Parallel experiments with the SARS-CoV PK RNA (PK^{sars}) allowed us to reach the same conclusion: $P1^{pk}$ pairing is weakly stable and is not expected to be formed at 37 ºC.

In this chapter, we employ NMR spectroscopy to identify specific structural features of the various RNA fragments that have been subjected to detailed thermodynamic (Chapter II) and fluorescence resonance energy transfer (FRET; Chapter IV) analysis. We note that NMR spectroscopy of large RNAs remains a considerable challenge. Although the largest RNA structure solved is 100 kDa, it represents an exceptional case. 15 kDa RNA structures still remain a challenge with normal NMR techniques (*101, 102*). Conformational dynamics also often plagues RNA NMR

spectroscopy as well, and this can be strongly dependent on solution conditions (*103, 104*).

Here, we use imino proton NMR spectroscopy to obtain unambiguous base pairing assignments that collectively establish formation of both P2 and P0b helical stems under conditions of low salt and low temperature (10 ºC) in larger 3' UTR RNAs. NMR studies of the 97-nt UTR RNA clearly show the presence of helical stem P0a under these conditions, consistent with the double hairpin as the dominant conformer in the intact UTR RNA under these conditions. These structural studies set the stage for the more global conformational studies of these RNAs using FRET and sedimentation velocity ultracentrifugation presented in Chapter IV.

MATERIALS AND METHODS

RNA preparation. RNA preparation followed procedures outlined in Chapter II. Purified RNA was suspended in a final NMR buffer of 10 mM potassium phosphate, pH 6.0, unless otherwise indicated. The RNA concentration in each case ranged from 0.8 to 2.0 mM.

NMR spectroscopy. All NMR experiments were performed on a Varian Inova 500- or 600-MHz spectrometer in the Texas A&M University Biomolecular NMR Laboratory. The proton resonances were referenced to an internal standard (100 μM DSS). The jump–return echo 1D and Watergate ^{1}H-^{1}H NOESY spectra (τ_{mix} = 150, short mixing time and τ_{mix} = 300 ms, long mixing time) were acquired to obtain imino proton

resonance assignments. All of the NMR data were processed using nmrPipe and analyzed using Sparky (*105, 106*).

RESULTS AND DISCUSSION

General approach. We acquired a series of one-dimensional (1D) and two-dimensional (2D) homonuclear (^{1}H-^{1}H) NMR spectra with the objective to obtain base pairing assignments within a series of increasingly complex MHV 3' UTR RNAs studied previously by thermal denaturation (Chapter II). The downfield region of a 1D spectrum acquired in 90% H_2O (9-15 ppm) largely reflects exchangeable imino protons that engage in hydrogen bonding interactions associated with Watson-Crick (W-C) and non-Watson-Crick (G•U, G•G, A•G, U•U and U•C) base pairs in helical segments of RNA. In unusual cases, the amino protons of a cytidine protonated at N3 and 2'-OH protons involved in stable hydrogen bonding interactions can also be found here (*43, 44*); however these are quite rare and typically not found in helical segments of RNA. In general, imino protons associated with noncanonical base pairs are shifted upfield of W-C base pairs, in the 9.5-12 ppm range. Imino protons associated with G-C W-C or U•C W-C-like base pairs are found between 11 and 13 ppm (*107-109*), while those imino protons associated with A-U base pairs are shifted most downfield, in the 13-15 ppm range.

As a result, the ≈10-15 ppm region of the spectrum of an RNA acquired in water will contain one imino proton resonance per Watson-Crick base pair, and one or two associated with each non-canonical pairing interaction, depending on the structure of the

base pair. A 2D ^{1}H-^{1}H Watergate NOESY spectrum acquired with a fairly long NOE mixing time (τ_m=200 ms) is used to perform an imino proton "walk" which relies on the fact that imino protons in successive base pairs in a helix are ≈4.0 Å apart; as a result, each imino proton will "see" (give detectable NOE cross peaks) to imino protons in the base pairs above and below it; terminal base pairs will give just one such connectivity. Finally, A-U base pairs are easily distinguished from G-C base pairs by virtue of strong intra-base pair connectivity between the uridine H3 imino proton and the H2 aromatic proton of adenosine within an A-U base pair. In contrast, the guanosine H1 imino proton will be close to the cytidine NH_2 protons in a G-C base pair; these resonances are easily identified in shorter mixing time NOESY experiments. In the data presented in this chapter, we only show the imino proton region of the 1D NMR spectrum compared to that of imino proton "walk" region; additional spectral regions or other spectra are not shown but have been used to corroborate the imino proton assignments given here.

Base pair dynamics also influence the ability to "walk" through a helical segment; this, in turn is largely dictated by the rate at which the imino proton exchanges with bulk solvent. Qualitatively, in 500-600 MHz spectrometers used for these studies, there will be some bleaching of the internuclear crosspeak intensity involving a pair of exchangeable imino protons during the mixing time (200 ms) of the NOESY experiment; this will attenuate crosspeak intensity, but not to a great degree since solvent exchange rates are relatively slow at 10 ºC for a "normal" base pair within a helical segment. As the imino proton exchange rate increases, the 1D spectrum might show a clear resonance, but this resonance will give no NOEs to other resonance since the

proton exchanges with solvent during the long τ_m. Thus, these base pairs are more dynamic than others. As the exchange rate increases further, the resonance is broadened in the 1D spectrum, and will not give appreciable NOE cross peaks in the long mixing time NOESY experiment, even when acquired at low temperature. We mention these general features here since they will come into play as we discuss the individual 3' UTR RNAs below.

NMR investigation of the MHV PK RNA. Since the major predicted helical stem of the PK RNA is the P2 stem-loop, we acquired 1D and 2D Watergate NOESY spectra (200 ms) in 90% H_2O for the P2 RNA (34 nts) (10 ºC, 10 mM potassium phosphate, pH 6.0, no added salt) (Fig. 27). Analysis of these spectra using the methods outlined above clearly reveals two noncanonical G•U base pairs (resonances connected by the *dashed* lines, upper panel), four A-U base pairs and three G-C base pairs with the assignments obtained from the imino proton walk indicated. These assignments are fully in accord with the proposed model of the P2 stem-loop.

Several additional features are worth noting however. The fact the G86 imino proton associated with the closing C74-G86 base pair at the "top" of P2 is protected from exchange with solvent and initiates a continuous "walk" to the penultimate base pair of the stem (U66) suggests that there is stable base pairing in the P2 loop, perhaps mediated by consecutive A•G base pairs at the base of the loop. In fact, there is a single U imino proton likely associated with A-U base pair just downfield of U66 (marked by a **?**; δ=13.9 ppm) that must derive from the hairpin loop. This may correspond to an U78-A83 base pair that is stacked on successive noncanonical G•A and A•G pairs on the P2

helix, which would close the loop by an GAGA tetraloop (see Fig. 27, *inset*). Since G imino protons associated with A•G pairs are often solvent exchange labile (*110, 111*), the structure of the P2 loop was not investigated further.

The same experiments were next performed on the 5' extended PK RNA (57 nts) under the same low salt conditions (Fig. 28). All resonances associated with the P2 stem could be readily identified in the 1D spectrum, with the exception of U66 in the penultimate U66-G94 base pair; this could be lost due to the immediately adjacent long L1 loop or it is simply shifted due to inversion of the immediately adjacent closing non-native g65-c95 base pair in the P2 RNA to native C65-G95 base pair in the PK RNA. Two additional upfield-shifted resonances could also be easily by identified in these spectra (indicated by the *red* asterisks, Fig. 28) that are not found in the P2 RNA; the protons that give rise to these new resonances are also close in space since they give a detectable NOE crosspeak (Fig. 28, *bottom*). These spectral features suggest a noncanonical G•U or U•U base pair (*109*).

Since these new resonances could be reporting on $P1^{pk}$ base pairing, an effort was made to assign these resonances by acquiring NMR spectra for the PK MT RNA, which thermal melts suggest cannot adopt the pseudoknotted conformation (Chapter II) (Fig. 18). The major conclusion from inspection of the PK MT spectra is that it is substantially identical to the wild-type PK RNA. Since both spectra contain the two additional "red" resonances, these cannot be reporting $P1^{pk}$ base pairing. They may be reporting on L1 base pairing or loop-stem interactions associated with P2. A major point of departure for both the PK and PK MT spectra, relative to the

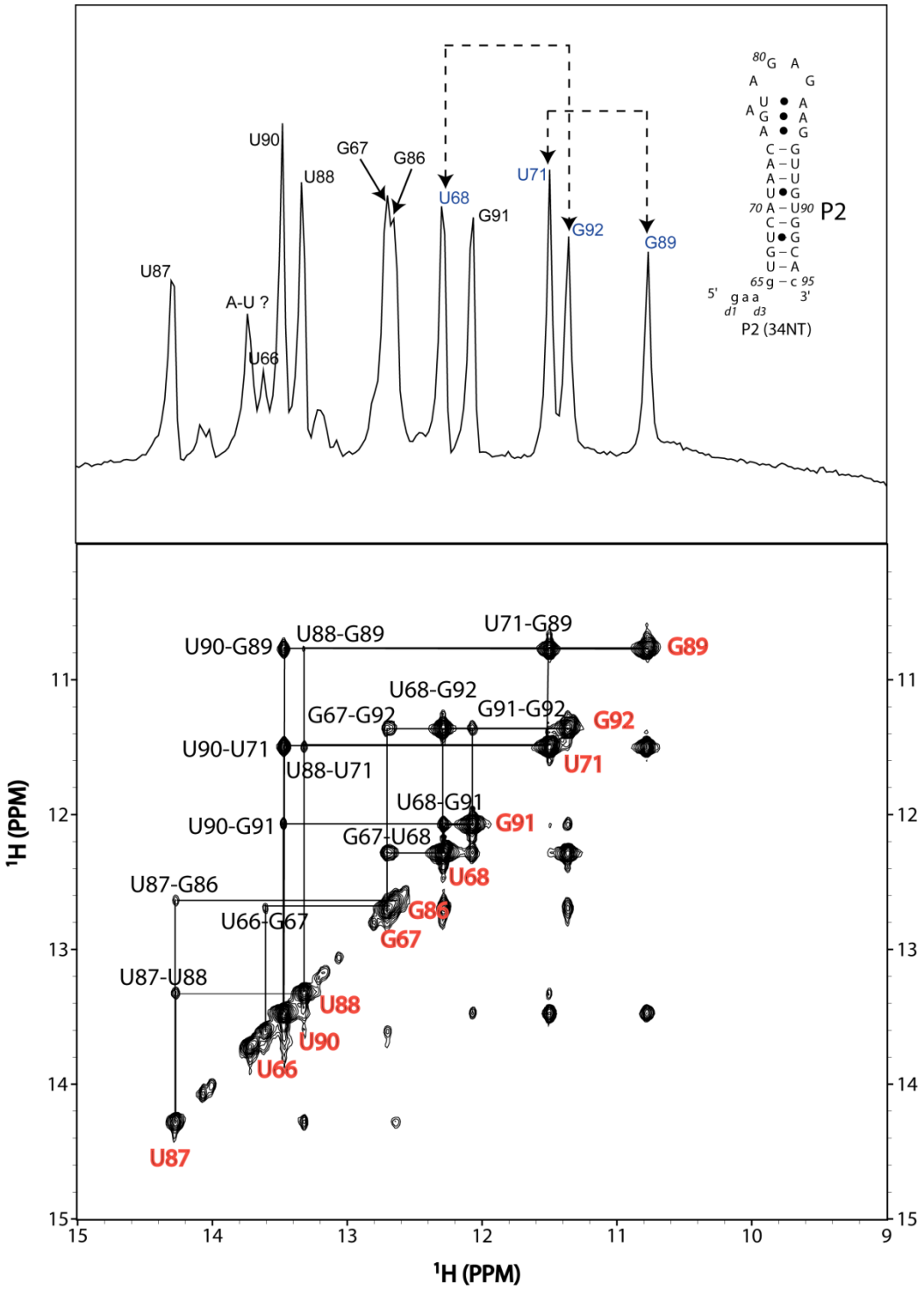

Figure 27 MHV P2 RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of P2 RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Solution conditions are 10 mM phosphate buffer pH 6 with 100 mM KCl and 5 mM Mg^{2+}. Imino proton resonance assignments are indicated.

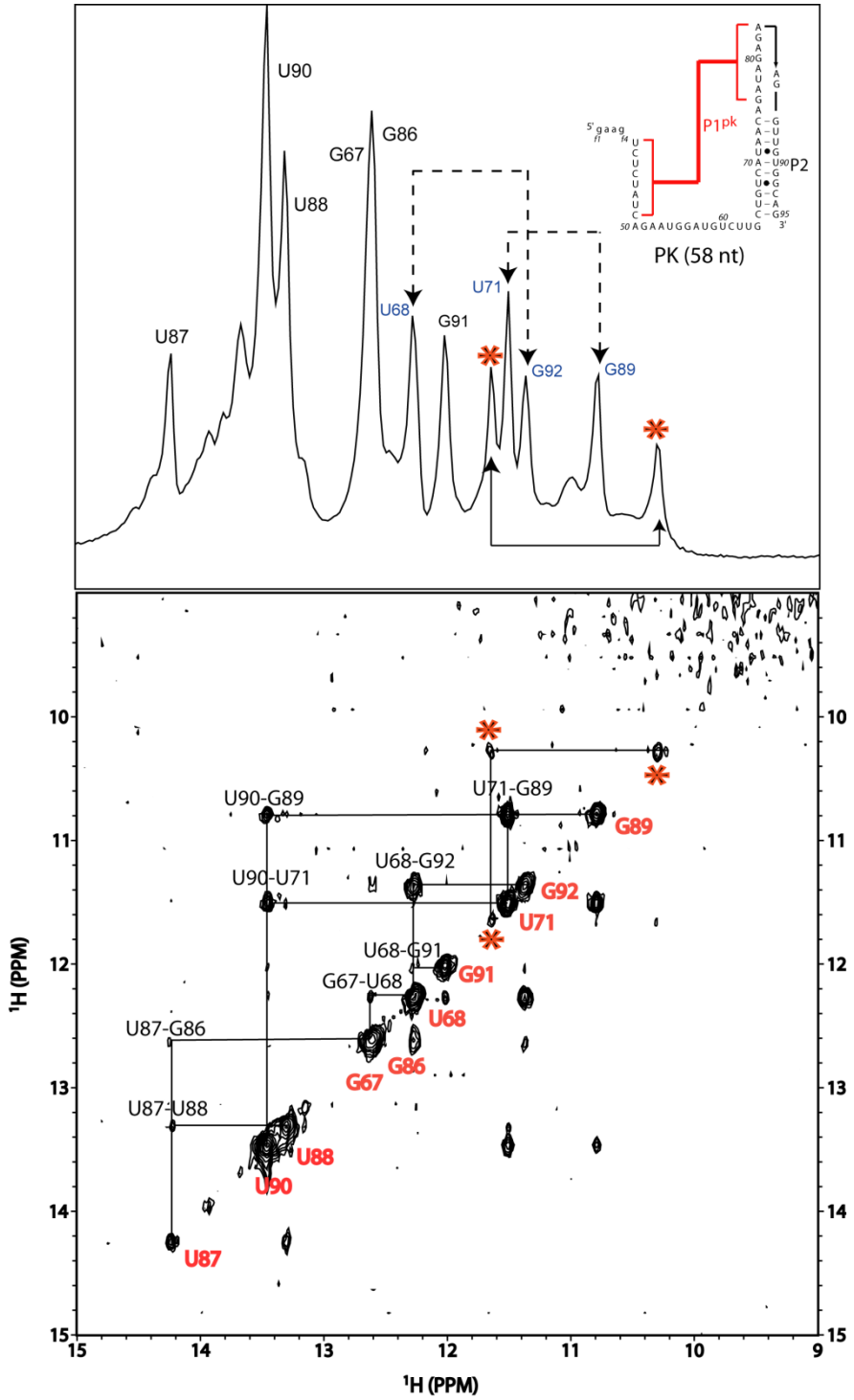

Figure 28 MHV PK RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of PK RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated.

component P2 stem-loop spectrum, however, is that the imino protons associated with the three P2 base pairs that are predicted to be closest to the helical junction in a pseudoknotted conformer, G86, U87 and U88, are characterized by increased solvent exchange rates since these two resonances are "washed out" to varying degrees along the diagonal and give no imino proton NOE connectivities in the 200 ms NOESY spectrum (Figs. 28-29). In fact, a continuous imino proton walk is observed for only the middle five base pairs of P2 (G67-U71) in both PK and PK MT RNA spectra. This situation contrasts sharply with the isolated P2 RNA (see Fig. 27).

These NMR studies establish that the PK and PK MT RNAs adopt identical conformations under these low salt conditions, a result fully in accord with thermal melting profiles of the PK RNA which reveal that the addition of 5 mM Mg^{2+} or high K^{+} is required to drive formation of appreciable the $P1^{pk}$ stem (see Fig. 20, Chapter II). In an effort to drive formation of the $P1^{pk}$ stem in the NMR tube, Mg^{2+} titrations were preformed with both the PK and PK MT RNAs with the spectra recorded at 25 ºC (Fig. 30) (spectra acquired a 10 ºC were severely line-broadened and gave no useful information). Although new imino proton resonances not observed previously grow into these spectra as Mg^{2+} is added, no off-diagonal NOE crosspeaks could be obtained that would permit their unambiguous assignment. Furthermore, at least a subset of the resonances appear to be associated with the MT RNA, and thus are unlikely to be reporting on $P1^{pk}$ pairing.

It is interesting to consider these spectra collected at 5 mM Mg^{2+} in the context of the thermal melting profiles for the PK and PK MT RNAs carried out under the same

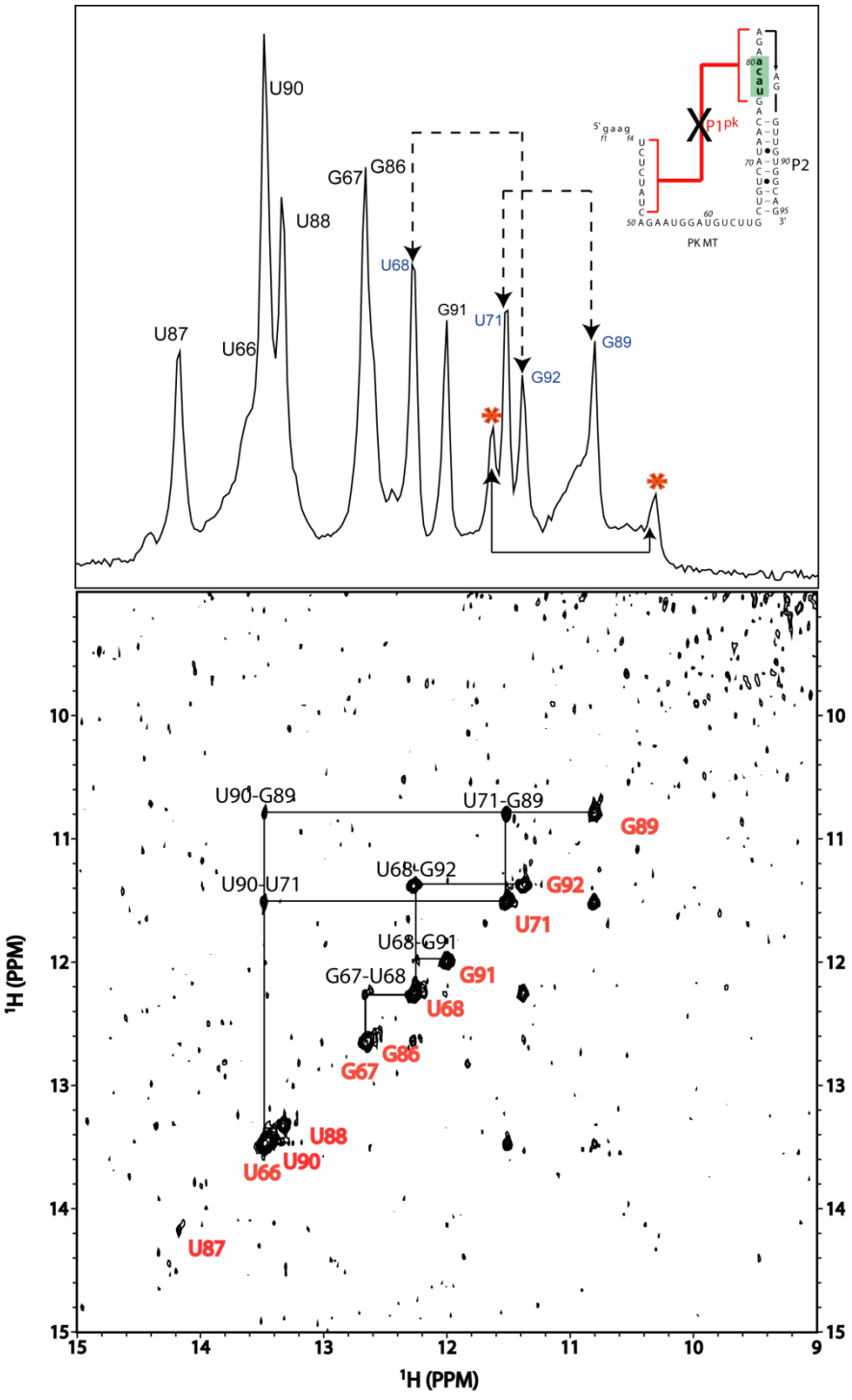

Figure 29 MHV PK MT RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of PK MT RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated.

conditions (Fig. 28, Chapter II Fig. 17). The unfolding of the $P1^{pk}$ stem occurs with a T_m of 31 ºC and this unfolding transition is only found in the PK RNA; thus, those imino protons associated with the $P1^{pk}$ stem may well be largely conformationally exchange broadened; this again serves to emphasize the dynamic and low-stability character of $P1^{pk}$ stem. In contrast, these new resonances might be reporting imino protons in base pairs that make up unfolding transition **b**, a subset of which would be present in both the PK and PK MT NMR spectra. This was not be investigated further.

We next decided to investigate whether stabilizing the $P1^{pk}$ stem would help in identifying resonances unambiguously associated with $P1^{pk}$ base pairing. To do this, we chose the PK 44 RNA (see Fig. 20B, Chapter II for thermal melt) which differs from the wild-type PK RNA by substitution of three G-C base pairs for predicted A-U base pairs; furthermore, $P1^{pk}$ stem would appear to be fully folded at 25 ºC (see Fig. 20; Chapter II). Inspection and analysis of these spectra (Fig. 31) (10 mM phosphate, pH 6, 10^{o}C) reveal the presence of approximately four additional imino proton resonances (labeled pk1-pk4) that correspond to 1-2 A-U and 2-3 G-C base pairs in addition to the full complement of P2 resonances; the A-U base pair labeled pk4 (δ=13.9 ppm) is also present in the wild-type PK RNA at 5 mM Mg^{2+}, but not in the PK MT RNA under the same conditions (Fig. 30). Two of these resonances, labeled pk1 and pk2, give a weak crosspeak in the 200 ms NOESY spectrum, thus revealing that they correspond to consecutive base pairs in the structure. The simplest interpretation of these spectra is that these new resonances are reporting on formation $P1^{pk}$ stem in the PK 44 RNA. Unfortunately, no additional imino-imino crosspeaks could be identified in these spectra;

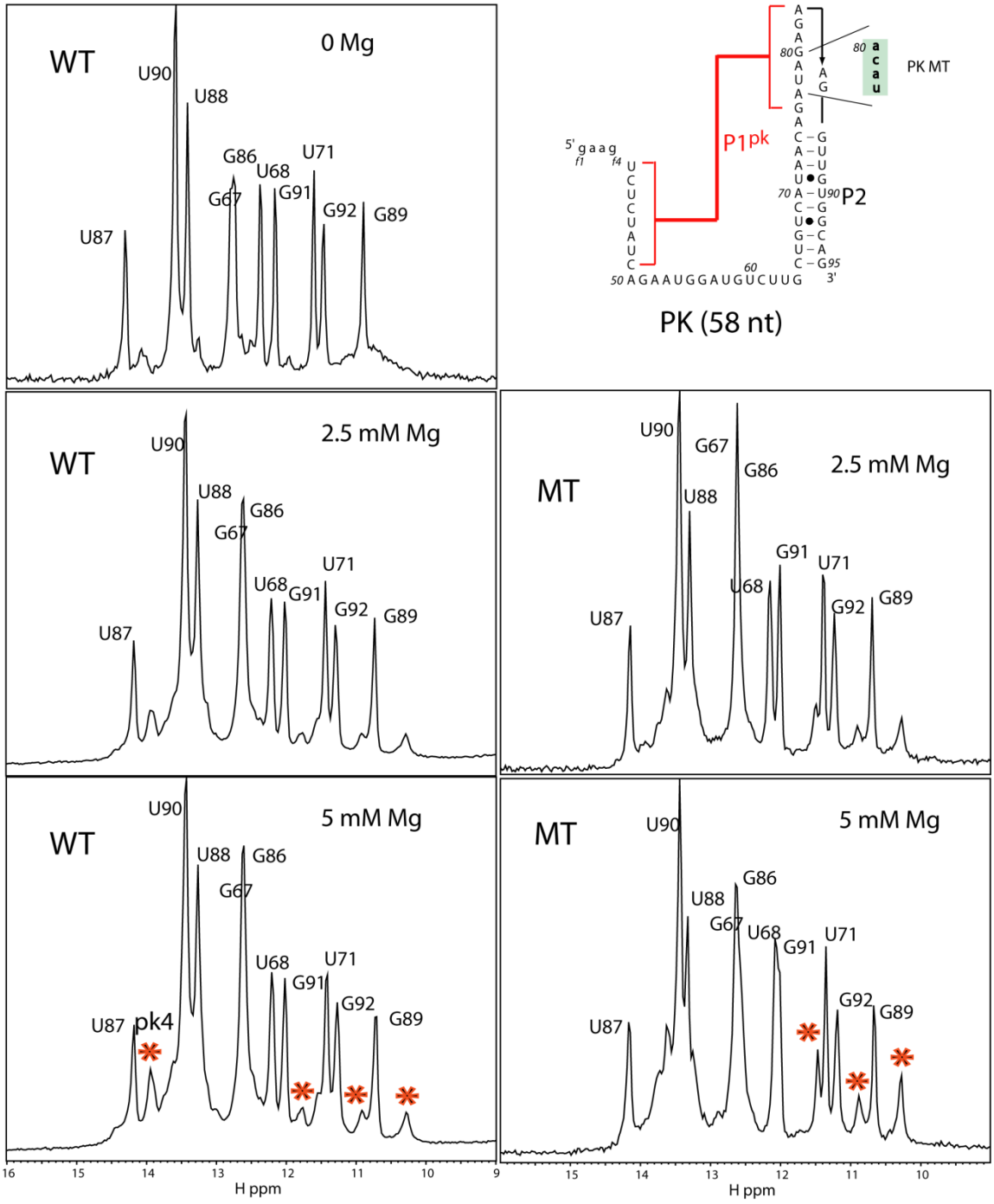

Figure 30 Mg^{2+} comparison of PK and PK MT RNAs NMR spectra. 1D NMR spectra comparison of PK and PK MT RNAs in a base solution of 10 mM phosphate pH 6 buffer 25 ^{o}C with added Mg^{2+} concentrations indicated. Red asterisks indicate differences in imino protons between PK and PK MT RNAs.

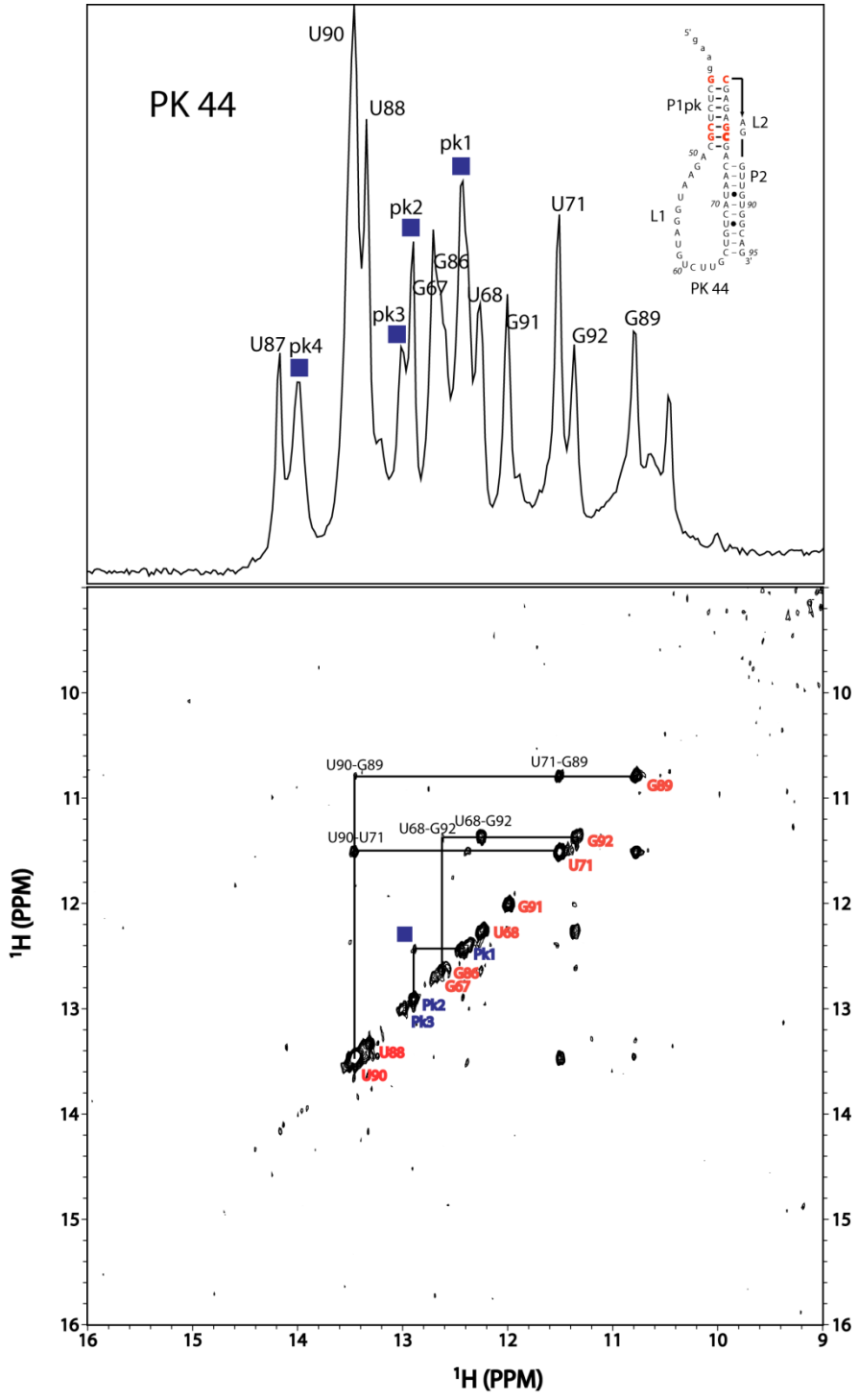

Figure 31 MHV PK 44 RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of PK 44 RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Blue squares indicate pseudoknot iminos. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated.

it was thus not possible to "walk" across the helical junction, for example (*43, 44, 112*). However, it is again interesting to note that top three base pairs of P2 (involving the G86, U87 and U88 protons) experience increased solvent exchange rates even when the $P1^{pk}$ stem is clearly formed.

The NMR data of the P2 and PK RNAs taken collectively reveal that the 5' extension of the P2 RNA enables $P1^{pk}$ base pairing and strongly enhances the solvent exchange rates of those P2 base pairs closest to the projected helical junction. Furthermore, $P1^{pk}$ pairing could only be confidently observed in the stabilized PK 44 RNA; however, even in this RNA in 5 mM Mg^{2+}, imino proton exchange rates for the majority of P2 and $P1^{pk}$ resonances are qualitatively more rapid than those in the central region of P2 helix. These data are consistent with a weakly folded and dynamic pseudoknot, perhaps one optimal for an RNA switching function.

NMR investigation of the MHV Δ7 UTR RNA. It was next of interest to determine the degree of $P1^{pk}$ pairing in the context of the Δ7 UTR RNA relative to the shorter PK RNA by NMR. As the Δ7 UTR is predicted to contain an intact P0b stem, base pairing assignments were obtained for an RNA encompassing the P0b stem only (Fig. 32). These spectra clearly confirm the secondary structure of the P0b hairpin; in fact, the presence of an unbroken imino proton walk from G31 to G32 to U16, and U16 to G17 on the opposite strand, coupled with a strong U18-G31 NOE, is consistent with consecutive noncanonical U18•G31 and G17•G32 base pairs incorporated into the helical P0b stem. The latter G17•G32 pair is consistent with biological experiments which reveal that substitution of the G•G pair with a G-C pair here is functionally silent

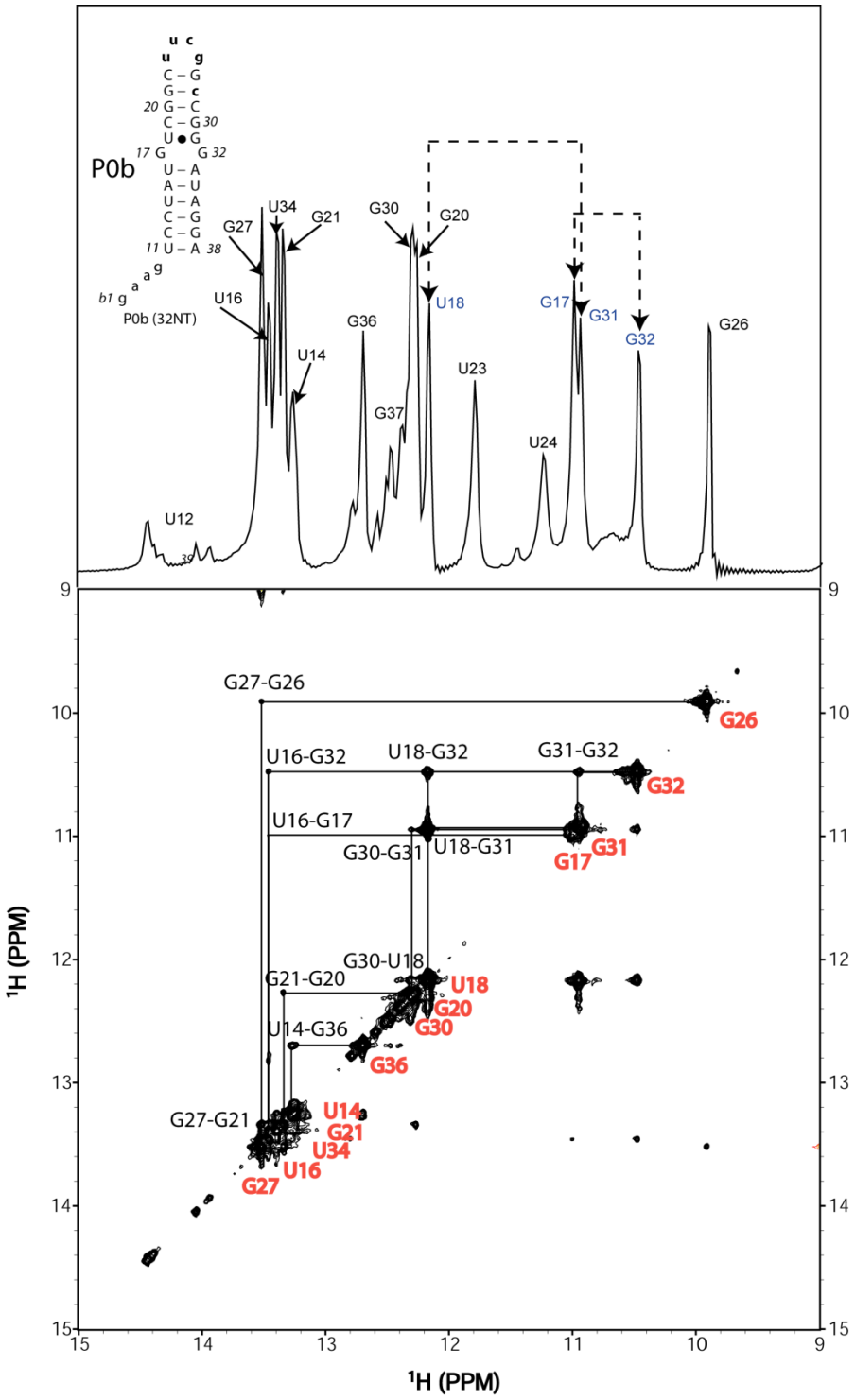

Figure 32 MHV P0b RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of P0b RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Solution conditions are 10 mM phosphate buffer pH 6 with 100 mM KCl and 5 mM Mg^{2+}. Imino proton resonance assignments are indicated.

(wild-type) (*94, 95*). In addition, the nonnative strongly stabilizing ^{23}uucg^{26} (also functionally allowed; (*94, 95*)) closes the top of P0b stem as evidenced by the characteristic g26 imino proton resonance (δ=9.9 ppm) (*108*) that gives an nearest neighbor NOE to G27, which initiates a continuous walk through G36. In contrast, imino protons at the base of the P0b stem (G37 and U11) appear conformationally exchange broadened while also existing in multiple environments. This is consistent with fraying of the helix terminus, a finding not surprising in the absence of stabilizing G-C cap (*108*).

Having assigned the resonances associated with the P0b stem, NMR spectra were recorded for the 87-nt Δ7 UTR RNA under the same solution conditions (10 ºC, 10 mM KP_i, pH 6 with no added KCl or $MgCl_2$) (Fig. 33). Inspection of the 1D imino proton spectrum under these conditions reveals that this spectrum is to a substantial degree simply a superposition of the P2 and P0b helical stems, with no clear evidence of additional resonances that could derive from P1^{pk} base pairing (compare the spectral sum of P2+P0b in Fig. 33B with that of the *bona fide* spectrum; Fig. 33A). This spectrum is exactly in accord with thermal melting profiles under these conditions in which we could obtain no evidence for P1^{pk} base pairing (Chapter II). As with the PK RNAs, no NOEs could be observed between imino protons at the top of the P2 stem, *e.g.,* U87 and U88, in this construct as well, again indicative of exchange broadening in this region of the RNA. Unfortunately, all efforts to stabilize P1^{pk} base pairing with the addition of K^+ or Mg^{2+} salts resulted in extreme spectral degradation due to excessive line broadening, and thus no additional information could be obtained from this RNA. Here, we are fighting

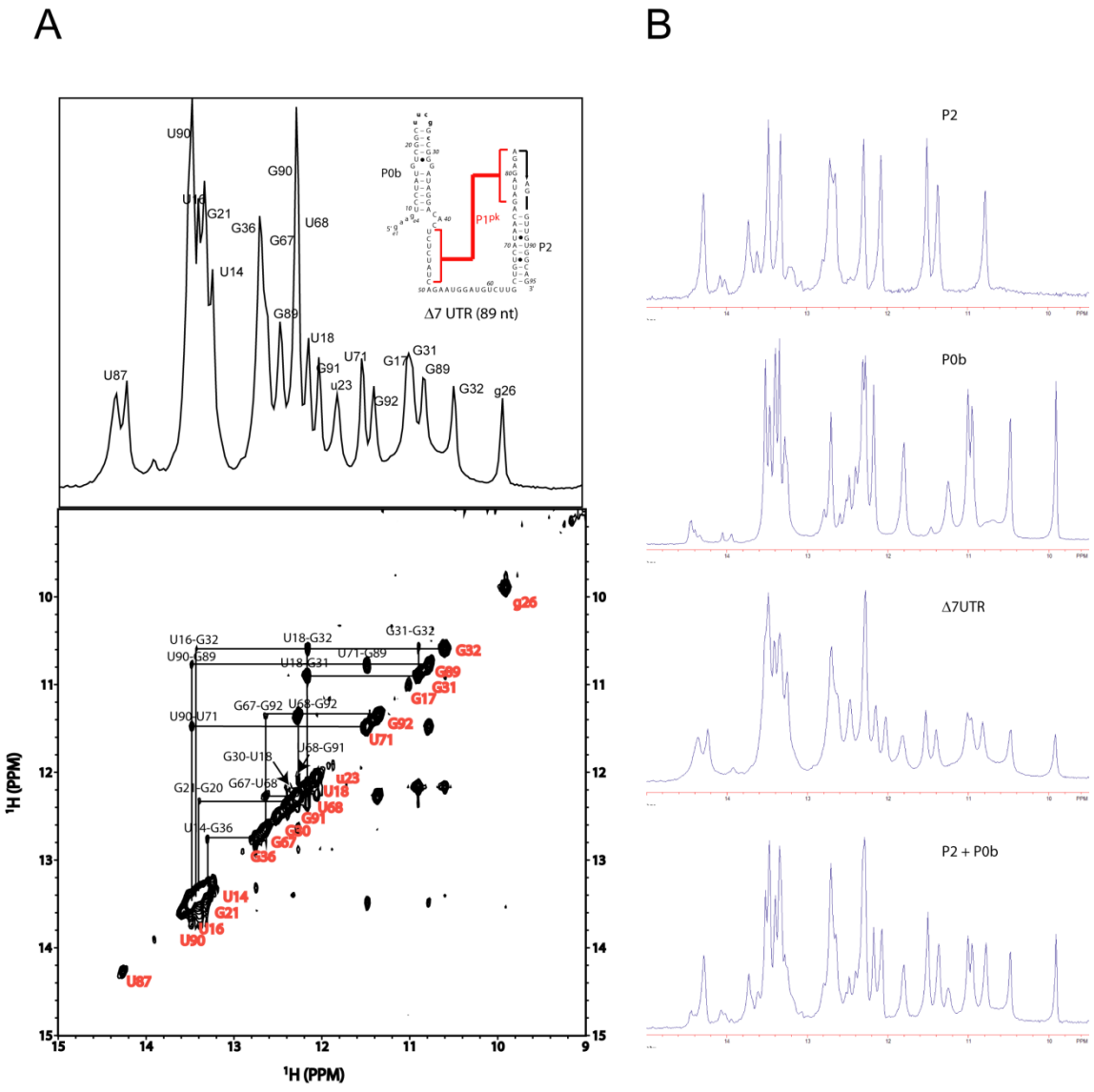

Figure 33 MHV Δ7 UTR RNA NMR spectra and addition spectra for P2 and P0b RNAs. A) MHV Δ7UTR RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of Δ7UTR RNA. Lower panel 2D NOESY mixing time 300 ms. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated. B) 1D spectral comparison of P2, P0b, and Δ7 UTR RNA relative to a synthetic spectral addition of P2 + P0b, bottom panel. Solution conditions are 10 mM phosphate buffer pH 6.

both conformational exchange broadening as well as very broad line widths (greatly decreased T_2) expected to be associated with the rotational tumbling of a large, folded RNA in solution (*101, 102*).

NMR investigation of the MHV UTR RNA. We next investigated the 97-nt UTR which is expected to adopt the double stem-loop conformation characterized by P0ab bulged stem-loop and P2 stems under low salt conditions and 10 ºC (Chapter II). In order to obtain spectral markers for the P0a stem, an isolated P0ab RNA (50 nts) was prepared; note that this construct contains two capping G-C base pairs and thus is designated P0ab GC RNA (Fig. 34, *inset*). Using the P0b base pair assignments as a guide, any remaining slowly exchanging imino proton resonances were assumed to be associated with the P0a helix. Four new imino proton resonances are readily identified in these spectra, with the NOESY spectra consistent two A-U and two G-C base pairs. Although the P0a stem capped with two G-C base pairs is predicted to contain three consecutive A-U/G-C base pair steps (see Fig. 34, *inset*), the observed U and G imino proton resonances are highly degenerate making it impossible to obtain their specific assignments; they are therefore designated Ua, Ub, Ga and Gb (see *yellow* circles). The imino proton region of the P0ab GC RNA (Fig. 34, *bottom*) reveals strong Ua-Gb and Ub-Ga NOE crosspeaks, a finding consistent with two consecutive A-U/G-C base pair steps. Note that we see no evidence for NOEs that connect the lower P0a and the upper P0b stems, suggesting that the asymmetric functions as a conformational "spacer" between the upper and lower stems under these conditions. We could also obtain no evidence for stable base pairing in the loop under these conditions, although this was

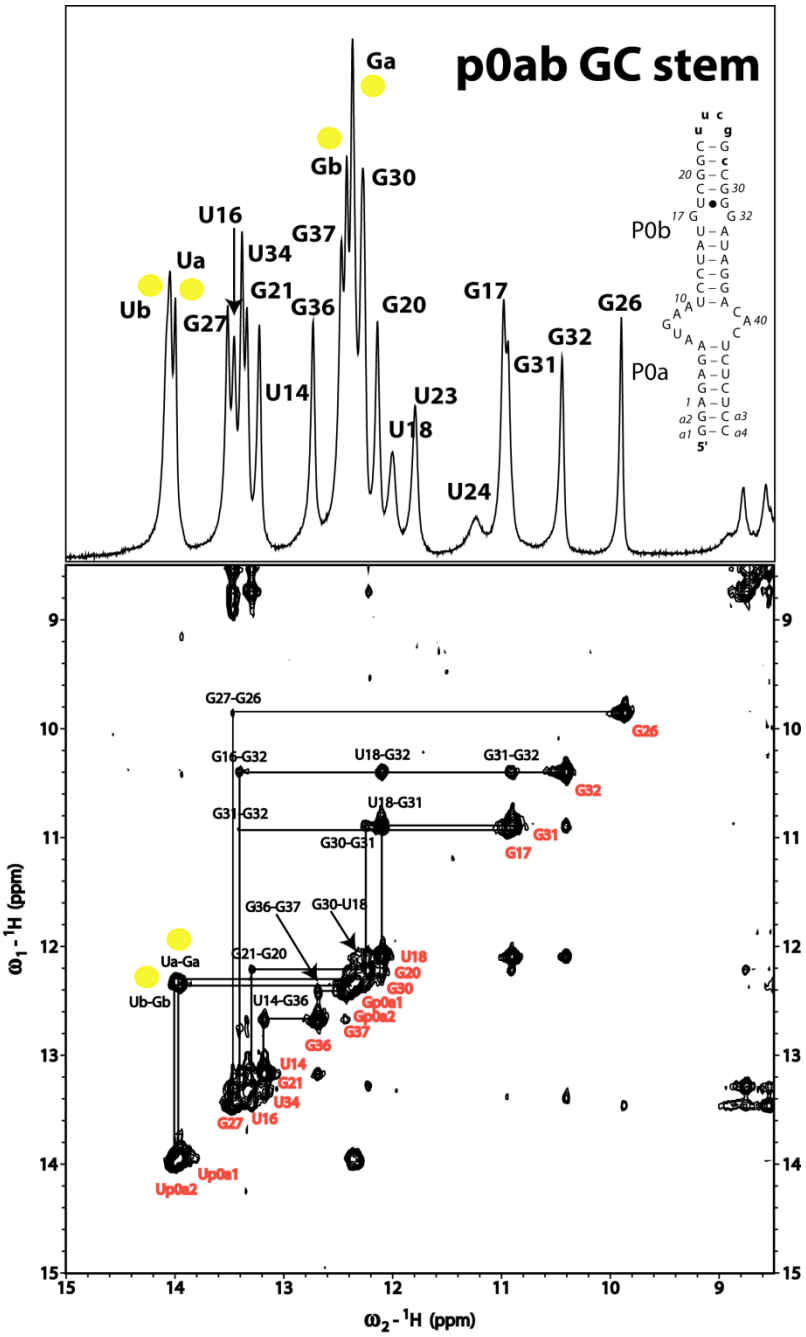

Figure 34 MHV P0ab GC RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of P0ab GC RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Yellow circles indicate stem P0a iminos. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated.

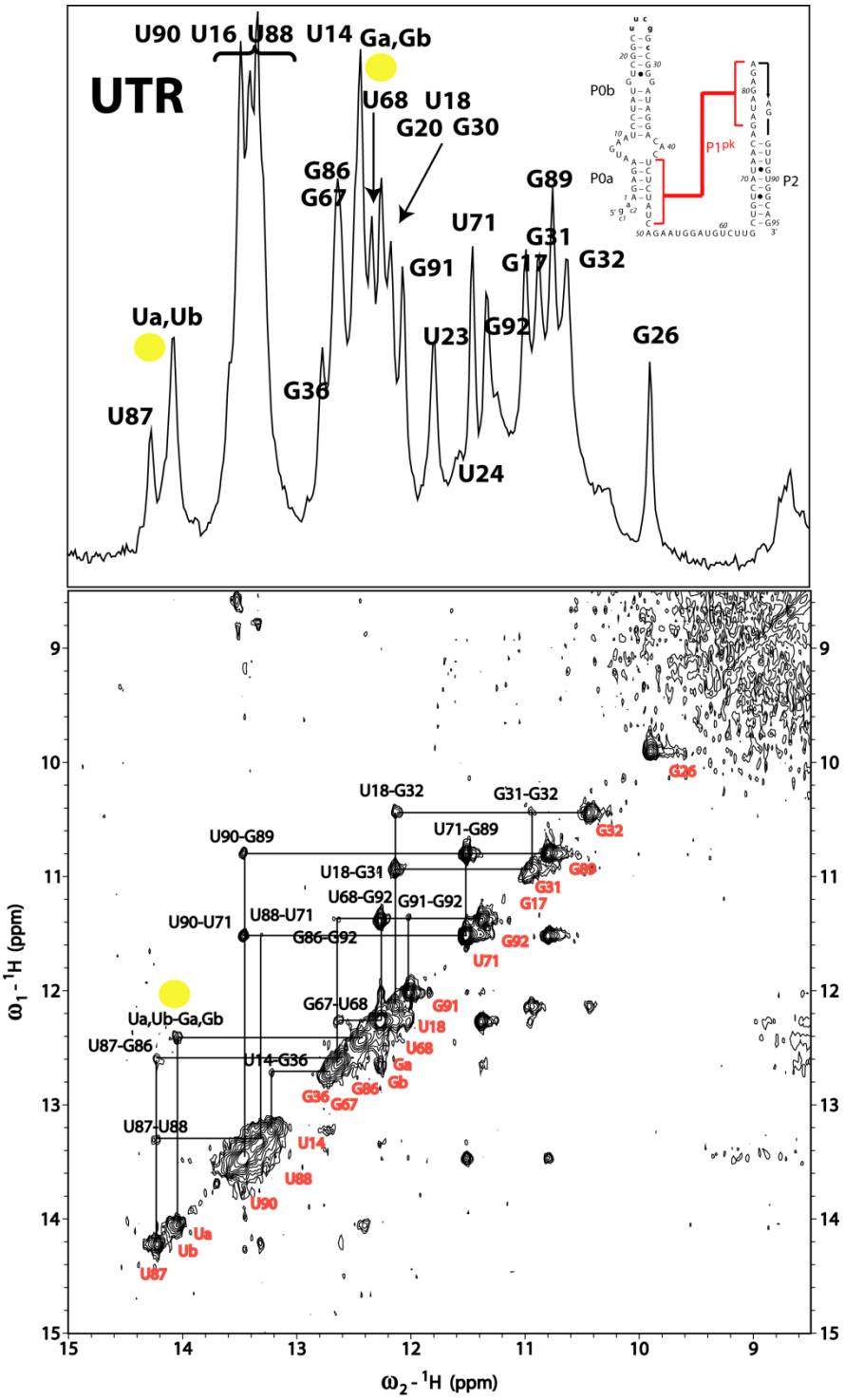

Figure 35 MHV UTR RNA NMR spectra. Upper panel 1D imino proton spectrum. Inset secondary structure diagram of UTR RNA. Lower panel 2D NOESY spectrum mixing time 300 ms. Yellow circles indicate stem P0a iminos. Solution conditions are 10 mM phosphate buffer pH 6. Imino proton resonance assignments are indicated.

not investigated further. These findings are consistent with a lack of conservation of residues in the J0ab asymmetric bulge in other CoV RNAs (*94, 95*).

Having determined marker imino proton resonances for the P0a and P0b stems, NMR spectra were acquired for the 97-nt UTR RNA under identical low salt solution conditions (10 ºC, 10 mM KP_i, pH 6.0) in an effort to determine if P0a and P0b resonances could be identified in these spectra (Fig. 35). Although the spectra are severely overlapped, the major conclusion is that the UTR RNA simply represents a superposition of P0a, P0b and P2 helices, with no other resonances that could be assigned to the $P1^{pk}$ base pairs. It is particularly clear that all four P0a resonances are present in these spectra (Fig. 35). This result is fully in accord with thermal melting profiles acquired under these conditions (Chapter II). Attempts to shift the equilibrium to more $P1^{pk}$ pairing with the addition of Mg^{2+} were unsuccessful, again to due to spectral degradation; nonetheless, Ua and Ub P0a resonances were still visible in these spectra (data not shown).

Strikingly, and in contrast to PK and Δ7 UTR RNAs, it is possible to observe a continuous imino proton walk from the top of the P2 helix, from G86 to G89. This result suggests that the P2 stem in the UTR RNA is characterized by structural and dynamical properties indistinguishable from the isolated P2 RNA stem-loop in strong contrast to the other RNAs.

CONCLUSION

The NMR data presented in this chapter provide important new insights into the structure and dynamics of the 3' UTR of MHV. Firstly, they establish the structural integrity of these large RNAs, and reveal that the PK, Δ7 UTR and UTR RNAs adopt homogenous secondary structures, with no evidence for misfolded conformers under these conditions. Secondly, they reveal that $P1^{pk}$ pseudoknot base pairing is very difficult to observe under the generally low-salt conditions that are compatible with detailed NMR studies of these large RNAs; this is *not* a typical feature for other H-type pseudoknots that stimulate translational frameshifting (*112, 113*) or perform some other biological role, *e.g.,* tmRNA-based tagging (*53-56, 114-116*) and translational autoregulation (*114-116*). Nonetheless, analysis of a stabilized PK RNA variant, PK 44, allowed us to obtain direct evidence that the $P1^{pk}$ stem can indeed be formed, at least in the context of the 57-nt PK RNA that lacks the P0b stem and any competition for folding from the 5' segment of the P0a helix. It is interesting to note that even when the $P1^{pk}$ stem forms however, the top of P2 helix closet to the helical junction region remains quite dynamic compared to the isolated P2 helix.

In addition, these studies provide strong evidence for the proposal that formation of the P0a helix essentially "locks down" the double hairpin conformer where the P0ab and P2 helices are fully folded and likely tumbling independent of one another in solution; when P0a is not allowed to form (in the PK and Δ7 UTR RNAs), the top of the P2 stem appears dynamically "poised" to form the pseudoknot under the right solution conditions or in the presence of a viral- or host-encoded protein or additional RNA

element (*117*) that can "throw the switch." These observations are collectively consistent with the basic tenets of the molecular switch model (*94, 95, 117*). Having established the fundamental thermodynamic (Chapter II) and structural features (this Chapter) of the MHV PK, Δ7 UTR and UTR RNAs, we next designed a fluorescence resonance energy transfer (FRET)-based approach to study the salt-induced global changes in structure and dynamics that accompanies pseudoknot pairing in the MHV 3' UTR region (Chapter IV).

CHAPTER IV

CHARACTERIZATION OF THE MOUSE HEPATITIS VIRUS 3' UTR BY ANALYTICAL ULTRACENTRIFUGATION AND FLUORESCENCE RESONANCE ENERGY TRANSFER

INTRODUCTION

We have previously characterized the MHV 3' UTR thermodynamically (Chapter II) and structurally using NMR (Chapter III). These techniques have confirmed the presence of weak pseudoknot pairing in the full length 3' UTR and have determined that the interaction is dynamic but maintains the tenets put forth by the proposed folding model. Here we have employed analytical ultracentrifugation (AUC) and fluorescence resonance energy transfer (FRET). These techniques provide insight into the global conformational distribution and extent of pseudoknot pairing in various UTR constructs used previously. Our analysis indicates that while a fragment of the population adopts the pseudoknotted state, it is not possible to convert the entire population of molecules to a single conformation, at least at magnesium concentrations less than 20 mM Mg^{2+} and 12 ^{o}C. Although we could not force the molecular ensemble to the fully pseudoknotted state, FRET experiments indicate that the population of molecules folded in the pseudoknot state is significant and result in a readily observable decrease in FRET signal is indicated by experiments that abrogate pseudoknot pairing.

MATERIALS AND METHODS

RNA preparation and analytical ultracentrifugation. RNA transcription and purification followed methods outlined in Chapter II. Analytical Ultracentrifugation samples were exchanged into a buffer containing 10 mM phosphate buffer, pH 6 with 100 mM KCl. Samples were then titrated with increasing amounts of magnesium chloride. Samples concentrations were prepared to produce an absorbance reading of 0.2 a.u. during analytical ultracentrifugation experiments.

Analytical ultracentrifugation experiments were run using a Beckman Coulter Optima XL-A ultracentrifuge equipped with an An60 Ti rotor. Sedimentation velocity experiments were carried out using a two channel 12-mm path length aluminum centerpiece with quartz windows. Data was collected at 10 ^{o}C, 60,000 rpm and monitored at a wavelength of 260 nm. Note that in nearly all runs the initial temperature of the samples varied from a start value of between 11-17 ^{o}C, with all samples reaching the set temperature of 10 ^{o}C at some point during the run. Data were analyzed using SedFit v9.4 with a continuous distribution model (*118*). The continuous distribution model is suited for mixtures of non-interacting, ideally sedimenting molecules. The continuous size distribution model determines the distribution of subpopulations of particles in solution using equation (1).

$$a(r, t) = \int c(M)L(M, r, t)\, dM + \epsilon \qquad (1)$$

In equation (1), *c(M)* is the particle population with mass, *M*, *L(M,r,t)* is the Lamm equation and *r* is the radius and *t* is the time. *a(r,t)* is the experimentally observed signal with associated error, ε. For sedimentation ultracentrifugation, the Lamm equation can be written as equation (2) for a dilute solution of a polymer.

$$\frac{d\chi}{dt} = \frac{1}{r}\frac{d}{dr}\left[rD(M)\frac{d\chi}{dr} - s(M)\omega^2 r^2 \chi \right] \qquad (2)$$

D(M) and *s(M)* are the diffusion and sedimentation coefficient of the particle, respectively, ω is the rotor angular velocity and χ is the particle concentration. The coefficients are related by the Svedberg equation (equation (3)).

$$s(M) = D(M)\frac{M(1-\bar{v}_M\rho)}{RT} \qquad (3)$$

Using the Stokes-Einstein relationship (equation (4)),where *k* is the Boltzman constant, *R* is the gas constant, *T* is the temperature, υ_M bar is the partial specific volume, $(f/f_o)_M$ is the frictional ratio and η_o and η_I are the is the standard and relative viscosity, an expression for *s(M)* can be obtained.

$$D(M) = \frac{kT}{6\pi\eta_0\eta_r(f/f_0)_M R(M,\bar{v}_M)} \qquad (4)$$

Finally, given *s(M)* and *D(M)* and their inverses, *M(s)* and *M(D)*, the size distribution *c(M)* can be transformed into a sedimentation coefficient c(s) and a diffusion coefficient, *c(D)*. The distributions *c(s)* are much more robust than *c(M)* in terms of unknown molecular shape. Whereas errors in the frictional ratio lead to overall changes in the determined values of *c(M)*, these errors only affect the resolution in *c(s)*, not the location of the peaks.

Using SEDFIT followed the general procedures outlined on the homepage www.analyticalultracentrifugation.com. This involved an initial fit of the data to optimize the frictional ratio and the meniscus and bottom values. This was then followed by a fit of the data to determine the sedimentation distribution. SEDFIT parameters used during the analysis were: resolution 100, tolerance 1, grid size 1000, confidence level .95, distance range 1-10, partial specific volume .53 cm^3 gm^{-1}, buffer density 1 gm/ml. In addition the fit was also conducted with the parameters optimizing for baseline, fit RI noise, and fit time independent noise. A table of the *C(S)* versus *S* value was generated and plotted using Kaleidagraph (Synergy).

RNA preparation and FRET analysis. RNAs were prepared using a segmental ligation of component RNA domains according to the method of Stark et al.(*119*). Component RNAs obtained from Dharmacon (Boulder, Co.) were given trivial names A, AΔ7, B, C, C MT and D (see Fig. 37 and 38). Dharmacon oligos were purified and internally labeled. Fragment B was internally labeled with an internal 6-carboxyfluorescein (F) while fragment C was internally labeled with a 5-carboxytetramethylrhodamine (R). Internal labels are located on C5 of cytidine

nucleotides linked through a three carbon amino group (-$CH_2CH_2CH_2NH_2$). Fragment C MT was also internally R labeled and represents a sequence in which four nucleotides of the predicted pseudoknot stem are mutated to prevent pseudoknot pairing (see Chapters II and III). Unlabeled fragments were also ordered from Dharmacon and consisted of A, AΔ7, and D. In addition B and C were also obtained as unlabeled RNAs to create donor only and acceptor only molecules. These were also used to create a synthetic PK RNA (B+C) to compare thermodynamically with the authentic PK RNA obtained through in vitro transcription. Fluorescently labeled ligated constructs were generated for ABC, ABC MT, AΔ7BC, AΔ7BC MT, BC, BC MT, BCD and BCD MT. Note that ABC is the UTR RNA studied previously, ABC MT is the UTR MT, AΔ7BC is the Δ7UTR, AΔ7BC MT is the Δ7UTR MT, BC is PK, BC MT is PK MT. BCD and BCD MT are 3' extended PK RNAs, based off of experiments by Zust et al. (*117*) designed to investigate 3' UTR long distance interactions with the upstream 97 NT putative molecular switch RNA.

Ligations were performed with T4 RNA ligase (New England Biolabs, Ipswich, MA.). Generic reaction conditions consisted of T4 RNA ligase buffer (diluted to 1X), 4 μM each of RNA fragment 1 and 2 and DNA linker. Solutions were heated at 95 ^{o}C for 5 minutes and cooled at room temperature for 1 hour. T4 RNA ligase was added (25 μL/ml of reaction from an enzyme stock of 20,000 U/ml) to the reaction and reacted overnight (~12 hours) at room temperature. Reactions were then DNase digested at 37 ^{o}C for 3 hours. Solutions were lyophilized to dryness and resuspended in a 7.5 M urea, 50 mM EDTA solution and gel purified according to techniques outlined in Chapter II.

Fluorescence experiments were collected on a Koala spectrofluorometer in photon counting mode (ISS, Champaign, IL, USA) with excitation from a xenon lamp source. Slit widths were 0.5 mm in excitation and 0.2 mm in emission. Emission intensities were determined in the L format using a single grating monochromator. Emission spectra were collected from 485 nm to 670 nm using excitation wavelength of 470 nm for fluorescently labeled RNA and from 565 nm 670 nm using an excitation wavelength of 542 nm for R-labeled RNAs. Samples were cooled to 12 ^{o}C using a temperature controlled circulating water bath and were run in a 10 mM phosphate buffer pH 6 with the indicated Mg^{2+} then added.

Data were analyzed using the enhanced fluorescence of acceptor method described by Clegg et al. (*120*) In this method the FRET efficiency is determined by measuring the intensity of the sensitized emission of the acceptor normalized to the fluorescence of the acceptor alone. Using equation (5), a ratio of the fluorescence of acceptor resulting from FRET excitation only $(F^{A}_{em}(\upsilon_1, \upsilon'))$ compared to total fluorescence of acceptor molecule in solution $(F^{A}_{ex}(\upsilon_2, \upsilon''))$, is generated. In equation (5) $F^{A}_{em}(\upsilon_1, \upsilon')$ is the subtracted difference between the total fluorescence emission excited at 470 nm from the fluorescence of the adjusted donor only spectrum excited at 470 nm. υ' is the excitation of the donor at 470 nm and υ'' is the excitation of acceptor only at 542 nm. υ_1 is the generated emission of the sample excited at the donor wavelength of 470 nm (υ') and υ_2 is the generated emission of the sample excited at acceptor only wavelength, 542 nm (υ''). The denominator, $F^{A}_{ex}(\upsilon_2, \upsilon'')$ is the

fluorescence of the double labeled sample excited at the acceptor wavelength 542 nm *(υ'')* and represents the total acceptor fluorescence of the sample.

$$(ratio)_{\mathrm{A}} = \frac{F^{\mathrm{A}}_{\mathrm{em}}(\nu_1, \nu')}{F^{\mathrm{A}}_{\mathrm{ex}}(\nu_2, \nu'')} \qquad (5)$$

The acceptor ratio $(ratio)_A$ is proportional to the FRET efficiency according to equation (6) where E is the FRET efficiency, ε is the molar absorptivity of the donor or acceptor at the two different excitation wavelengths (υ', υ''), Θ^A is the emission shape factor for the fluorescence quantum yield of the acceptor and d^+ is the percent labeling of the donor.

$$(ratio)_{\mathrm{A}} = \left\{ Ed^{+} \left[\frac{\varepsilon^{\mathrm{D}}(\nu')}{\varepsilon^{\mathrm{A}}(\nu'')} + \frac{\varepsilon^{\mathrm{A}}(\nu')}{\varepsilon^{\mathrm{A}}(\nu'')} \right] \right\} \frac{\Phi^{\mathrm{A}}(\nu_1)}{\Phi^{\mathrm{A}}(\nu_2)} \qquad (6)$$

RESULTS

Analytical ultracentrifugation. Analytical ultracentrifugation was used to determine population distribution of folded pseudoknot molecules in solution. Samples were run using increasing amounts of magnesium to induce pseudoknot folding (see Chapter II). Wild-type pseudoknot constructs were analyzed alongside the equivalent mutant pseudoknot constructs to obtain insight into the population distribution (see Figs. 15 and 16). Sedimentation velocity experiments were conducted and analyzed according

to a sedimentation distribution model which provides a population distribution of conformational states *C(S)* in solution according to sedimentation value *(S)*. Based on the analysis described previously, the sedimentation value rate is a function of diffusion which is determined in part by the frictional ratio of the molecule (see equation 3 and 4). The frictional ratio is dependent on the shape of the sedimenting molecule. Therefore the sedimentation value can provide information regarding conformational changes in solution.

The distribution plots for MHV PK, PK MT, Δ7UTR, Δ7UTR MT, UTR and UTR MT RNAs are shown in Figure 36 under 10 mM potassium phosphate, pH 6 and 100 mM KCl solution conditions with designated Mg^{2+} additions. Panel A in Figure 36 indicates the distribution obtained for the PK and PK MT RNAs. From the plot of PK under buffer only conditions (10 mM potassium phosphate, pH 6 and 100 mM KCl, no Mg^{2+}) a single conformation is predominant at an *S* value of ~3. Upon addition of magnesium, a second population is observed, with an *S* value of ~4.5. The relative amount of this species seems to increase as magnesium is added and plateaus at 10 mM Mg^{2+}. Inspection of the distribution plot for PK MT also displays a single population peak at $S \sim 3$ under buffer only conditions. In contrast to PK however, no second population is produced when the Mg^{2+} concentration is increased. In the absence of magnesium, based on thermal melts conducted under the same solution conditions (see Chapter II, Fig. 20), the $P1^{pk}$ stem does not form in the PK RNA. As a result, the single population distribution observed for the buffer only condition for both the PK and PK MT constructs must be reporting on the unfolded pseudoknot state. The observance of a

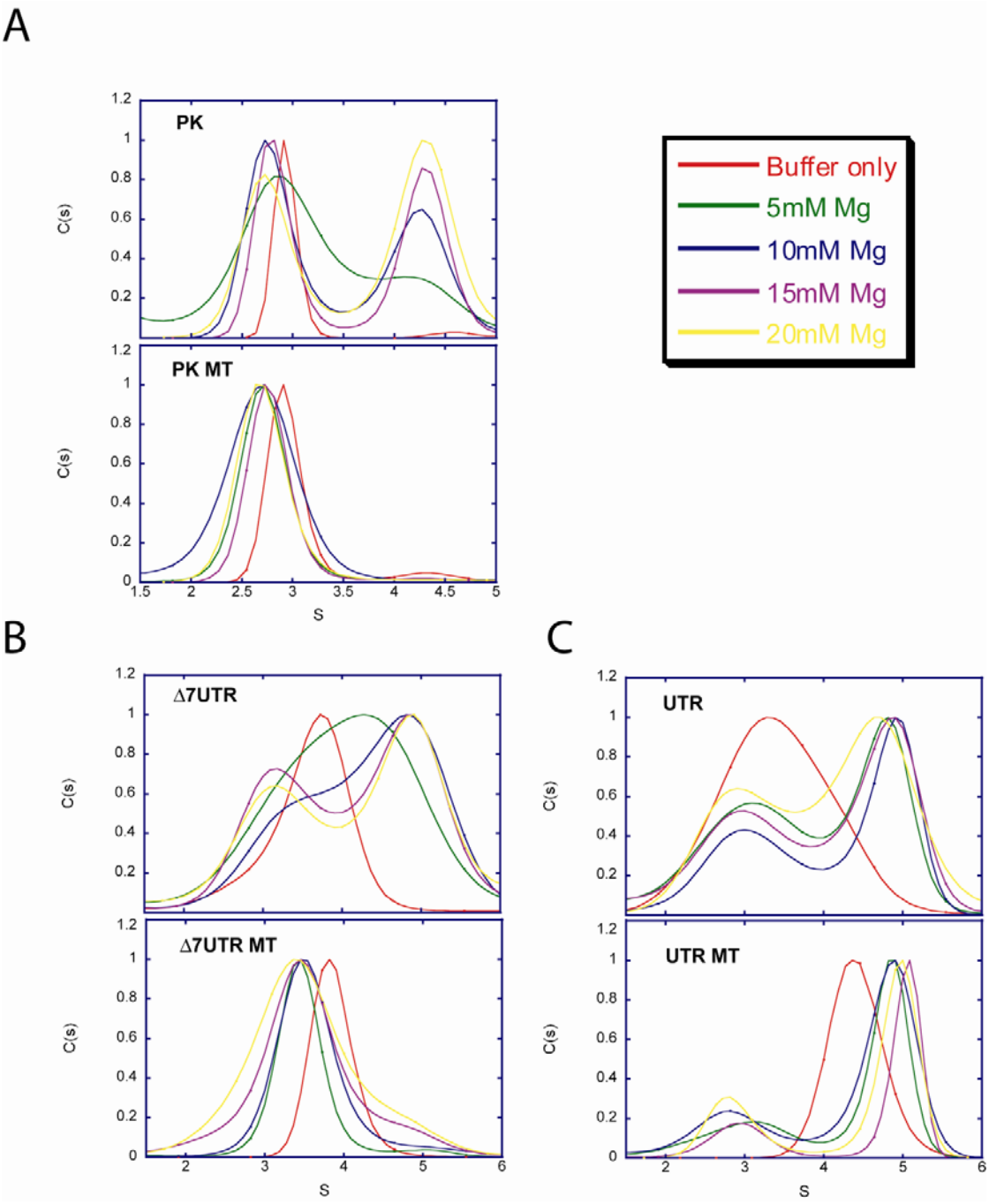

Figure 36 Analytical ultracentrifugation sedimentation velocity data for PK, Δ7 UTR and UTR RNAs. Constructs analyzed in the presence of increasing amounts of magnesium. Buffer only conditions are 10 mM KH2PO4 pH 6 with 100 mM KCl. A) PK and PK MT AUC data B) UTR and UTR MT AUC data C) Δ7 UTR and Δ7 UTR MT AUC data.

second population distribution in the PK RNA is therefore reporting on the formation of the $P1^{pk}$ state. This conclusion is corroborated by the absence of the second distribution in the PK MT RNA which is prevented from forming the $P1^{pk}$ stem.

To understand how the population distribution is effected by the presence of the P0b stem, the same sedimentation experiments were conducted for the Δ7UTR and Δ7UTR MT RNAs (see Fig. 36B). With the Δ7UTR RNA under buffer-only conditions, a single distribution at S ~3.8 is observed. As magnesium is added a second population distribution is observed that centers around S ~5. When performing the same titration for Δ7 UTR MT RNA, as in the case for PK MT RNA, no second population distribution is observed. Only a single peak is observed around S ~3.8. Under the same buffer-only conditions, thermal melts indicate that the $P1^{pk}$ stem is not folded. This indicates that the single distribution observed for Δ7 UTR and Δ7UTR MT RNAs, under buffer-only conditions, is reporting on the unfolded pseudoknot state. The addition of Mg^{2+}, as in the case of the PK MT RNA, does not produce a second population distribution in the Δ7 UTR MT RNA. This indicates that the second population observed in the Δ7 UTR RNA must also be reporting on the folding of the $P1^{pk}$ stem.

Figure 36C presents a distribution plot for UTR and UTR MT RNAs. The UTR RNA is characterized by a single conformation under buffer-only conditions; however in contrast to the PK and Δ7 UTR RNAs, the peak distribution is extremely broad and encompasses a large range of S values. With the addition of Mg^{2+}, two population distributions are observed, one centered around S ~3 and one at S ~5. The subsequent addition of magnesium does not significantly change the S value or the peak dimensions

of the two population distributions. The UTR MT RNA analyzed under buffer-only conditions also presents a single population distribution. In contrast to the UTR RNA, the peak is sharper and occupies a narrower range of *S* values. With the addition of Mg^{2+} two populations are also produced. This is strikingly different from the previous $P1^{pk}$ MT RNAs which only display a single population distribution regardless of the addition of Mg^{2+}. The two population distributions observed for the UTR MT RNA have *S* values similar to the UTR RNA (*S* ~3 and *S* ~5) and also indicate relatively little change with the addition of Mg^{2+} as observed for the UTR RNA. Although both UTR and UTR MT RNAs contain two population distributions, the relative areas under each peak varied in the two populations. In the UTR RNA, the population distributed around *S* ~3 contains a larger peak area relative to the peak distributed at *S* ~5. In contrast, the relative distribution between *S* ~3 and *S* ~5 in the UTR MT RNA indicates a decrease in area and intensity for the *S* ~3 population.

The formation of two population distributions for PK and Δ7 UTR RNAs indicate that two separately folded RNA conformations are present in solution after the addition of Mg^{2+}. The absence of the second population, at *S* ~4.5, in the PK MT and Δ7 UTR RNAs identifies this transition as belonging to the $P1^{pk}$ folded state. This is consistent with the thermal melts and the NMR data which indicate that the pseudoknot is unfolded in the absence of Mg^{2+} and in the $P1^{pk}$ mutant RNA constructs. Importantly, two population distributions were present even under the highest Mg^{2+} concentration of 20 mM. This indicates that there is always a mixture of states in solution involving the fully folded pseudoknot state and an unfolded $P1^{pk}$ state.

Although it is clear that the pseudoknot state is present in the PK and Δ7 UTR RNAs, its presence in the the UTR RNA is less clear. Both the UTR and UTR MT RNA constructs contain a single population distribution under buffer-only conditions; however, both RNAs generate two population distributions with the addition of Mg^{2+}. Based on the attenuation of the peak centered at S ~3 in the UTR MT RNA construct, it would seem possible that this is reporting on the pseudoknot folded state. However, it is expected based off of the PK MT and Δ7 UTR MT constructs where $P1^{pk}$ folding is eliminated. One possible explanation of the presence of $P1^{pk}$ folding in the UTR MT construct is that the four nucleotide $P1^{pk}$ mutation is insufficient to completely abrogate pseudoknot folding. This would be consistent with the thermal melt data of the UTR MT (see Fig. 23, Chapter II) which is characterized by an attenuated unfolding transition assignable to $P1^{pk}$ pairing.

FRET analysis of UTR, Δ7UTR and PK RNAs. FRET offers a method of spectroscopically detecting molecular interactions based a non-radiative transfer of energy from a donor molecule to an acceptor molecule. Because energy transfer is only possible within distances dictated by the donor acceptor pair, FRET offers a method of detecting specifically when the two fluorophores are close to one another in space. When this energy transfer is coupled with biomolecules the FRET pair can serve as a method of reporting on folding states of the biomolecule. We have employed this method in the determination of pseudoknot folding within the context of the UTR.

Fluorescence resonance energy transfer experiments were designed as outlined in Figure 37. FRET pairs were placed according to the proposed MHV UTR model

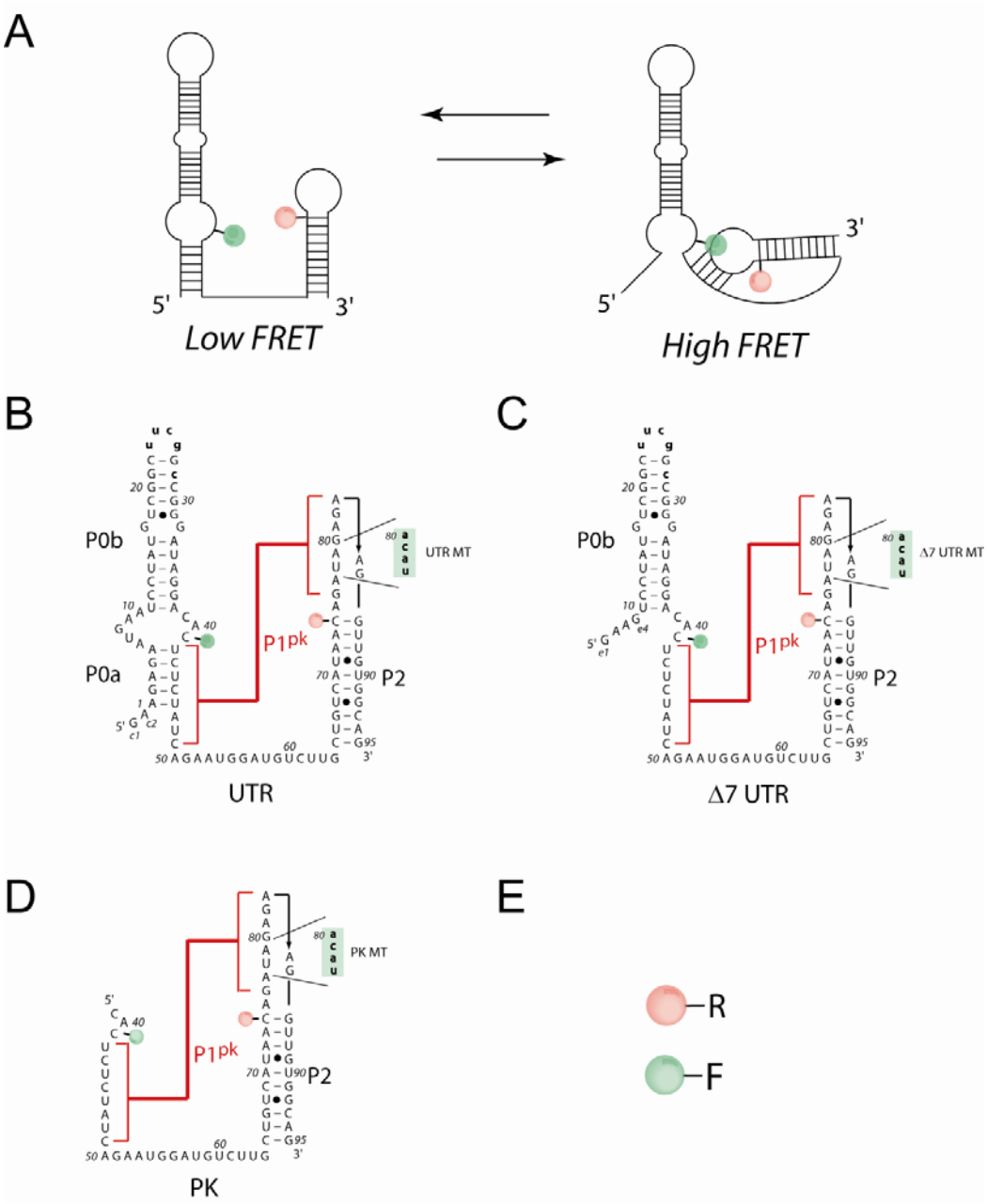

Figure 37 Schematic of FRET constructs used to measure pseudoknot pairing. A) Folding model of the UTR with representative fluorescent probes placed to report on pseudoknot folding. B) UTR RNA construct labeled with donor FAM placed on C41 and acceptor TAMRA placed on C74. Two constructs were generated with this donor fluorophore schematic one with a normal pseudoknot interaction and one with a mutated pseudoknot stem as indicated by the green boxed sequence. C) Represents the equivalent constructs used for D7 UTR and D) Represents the equivalent constructs for PK. E) Illustrates the donor and acceptor pair used for all fluorescence experiments.

focusing on the pseudoknot pairing interaction. FRET fluorescent pairs were placed at cytidines 41 and 74. Within the B and C RNAs, respectively, the donor F (green circle) was placed at C41 and the acceptor R (red circle) was placed at C74, each conjugated to the C5 position of cytidine. In order to synthesize the various RNAs depicted in Figure 37, a ligation method was employed using T4 RNA ligase and individual RNA segments as building blocks for the complete RNA constructs (see Fig. 38). The ligation method followed the protocol of Stark et al. in which a DNA linker is used to produce a bulged region at the ends of the RNA as indicated in Figure 38B (*119*). By making the DNA template shorter than the RNA segment, a five NT overhang on the 5' RNA fragment and a 2 nt overhang on the 3' RNA fragment is generated. The single stranded region is now functional for ligation by T4 RNA ligase, which requires single stranded regions to be active. In addition to the single-stranded ends, it also proved useful to maintain the 3' RNA fragment as a fully 2' ACE protected state as synthesized by Dharmacon. Using this method, RNAs were generated for the UTR, UTR MT (ABC and ABC MT), Δ7UTR, Δ7UTR MT (AΔ7BC and AΔ7BC MT) PK and PK MT (BC and BC MT) RNAs.

Analysis of the FRET signal has been accomplished through the determination of the acceptor ratio as outlined by Clegg et al (*120*). The acceptor ratio method normalizes the sensitized emission of the acceptor from donor excitation with the emission from the acceptor alone. To do this the spectra are excited at the donor wavelength excitation (470 nm for F) which provides the sensitized emission of the acceptor. The spectra are also excited at the acceptor excitation wavelength only (542

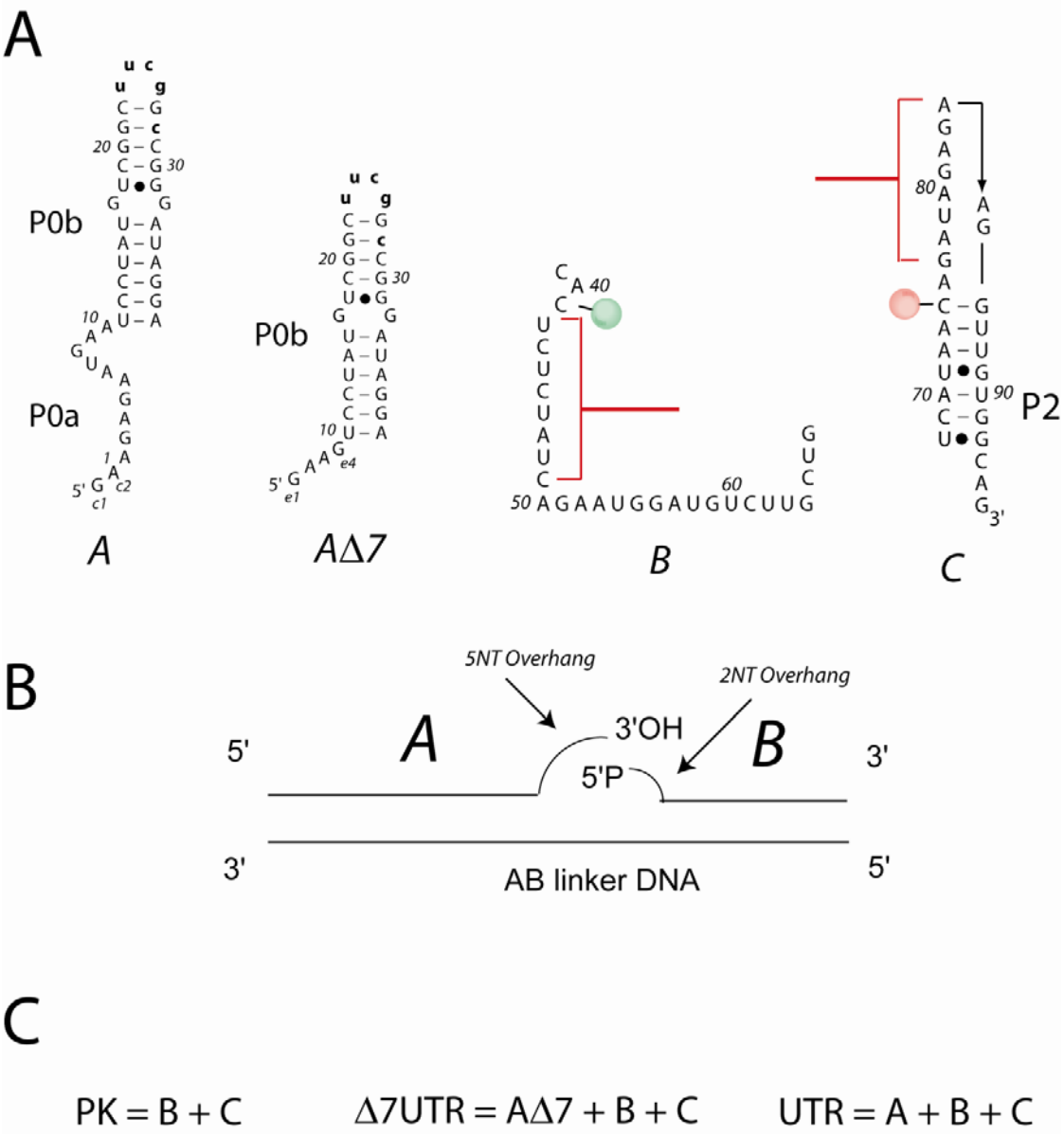

Figure 38 FRET constructs starting ligation fragments. A) Individual RNA segments used in the ligation of the MHV constructs ABC (UTR), AΔ7UTR (Δ7UTR) and BC (PK). To make the mutant pseudoknot constructs, C MT was used instead of segment C which contained a four nucleotide mutation disrupting pseudoknot pairing. B) Ligation method utilizing T4 RNA ligase which requires single stranded ligation ends. Using a DNA linker that is unpaired for the last 5 nt in the 5' fragment and unpaired for the first 2 nt in the 3' fragment, a bulged single stranded region could be used as a ligation source for T4 RNA ligase. C) Ligated constructs as defined using the independent RNA fragments.

nm for R) which provides a measure of the acceptor intensity only with no contribution from the donor. A donor only intensity allows a spectral standard that can be used to provide information regarding spectral shape only. By normalizing the height of the donor only sample to the sample containing both acceptor and donor and then subtracting the two, a spectrum containing just the sensitized emission from the acceptor can be obtained. The $(ratio)_A$ (see equation (5)), determined by dividing the two intensities, is directly proportional to the FRET efficiency, E, according to equation (6).

The advantage of this method is that taking the ratio circumvents the requirement of knowing the concentration of the labeled molecule, which is often erroneous, especially when sample concentrations are limited. Additionally this prevents the need of ensuring 100% labeling of acceptor molecule. The energy transfer ratio requires only two fluorescence measurements. This is extremely beneficial since the measurements are conducted on the same sample and with the same instrument. As a result, absorption measurements and concentration determinations are not necessary to calculate $(ratio)_A$.

FRET data were acquired for all of the samples outlined in Figure 39C and calculated according to the equations presented in Materials and Methods section. Figure 40A displays typical intensity excitation spectra for the PK RNAs BC and BC MT. BC-F^B is the donor only labeled spectrum and BC-R^C is the equivalent acceptor only labeled spectrum. FRET samples were annealed in a base solution consisting of 10 mM phosphate buffer pH 7.0 with 100 mM KCl. The spectra were measured at 12 ^oC with an excitation wavelength of 470 nm and over increasing magnesium concentrations.

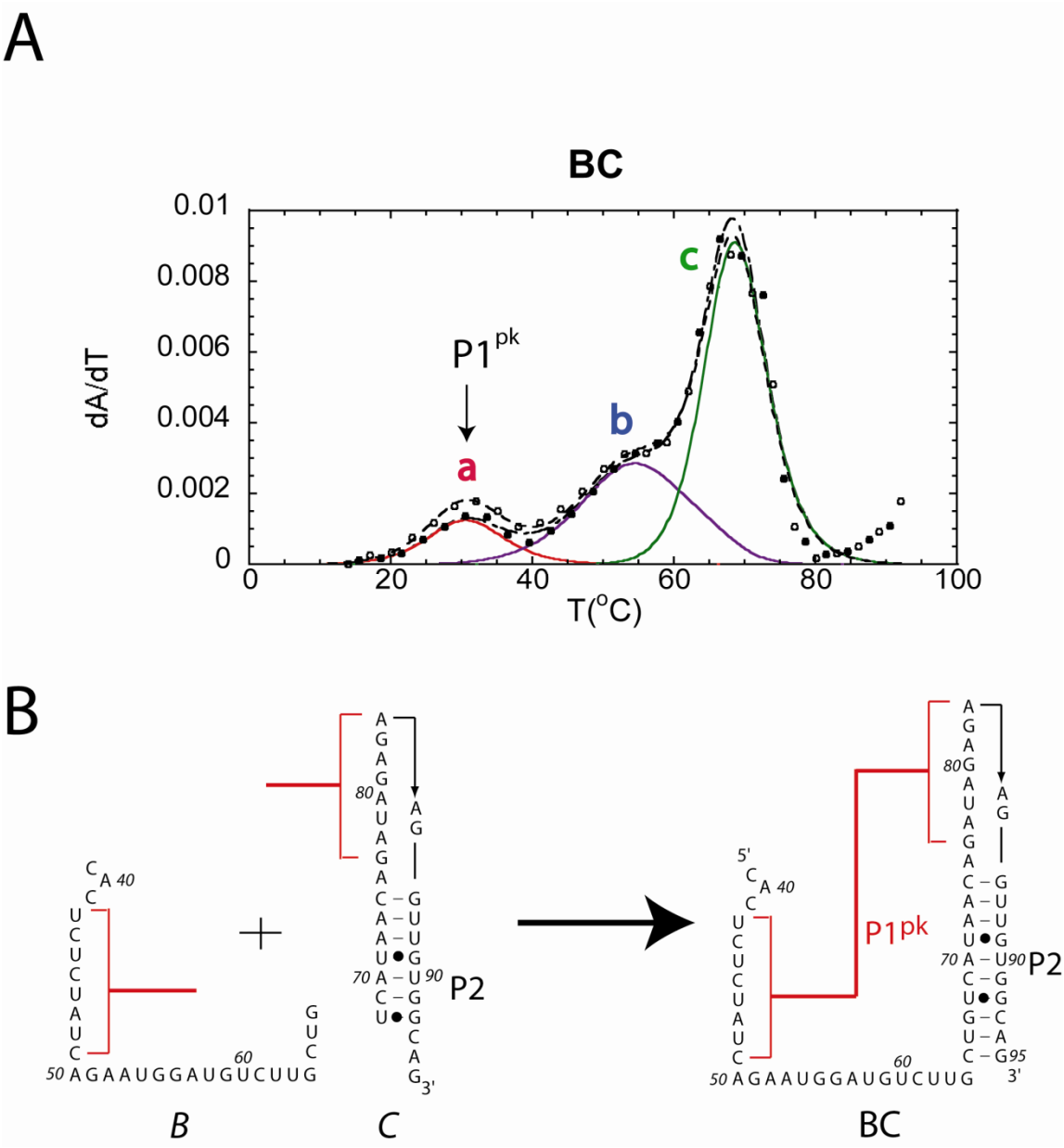

Figure 39 Ligation confirmation thermal melt. A) Thermal unfolding melt of the ligated PK construct composed of ligated B and C RNA fragments. B) RNA fragments used for ligation of PK RNA construct.

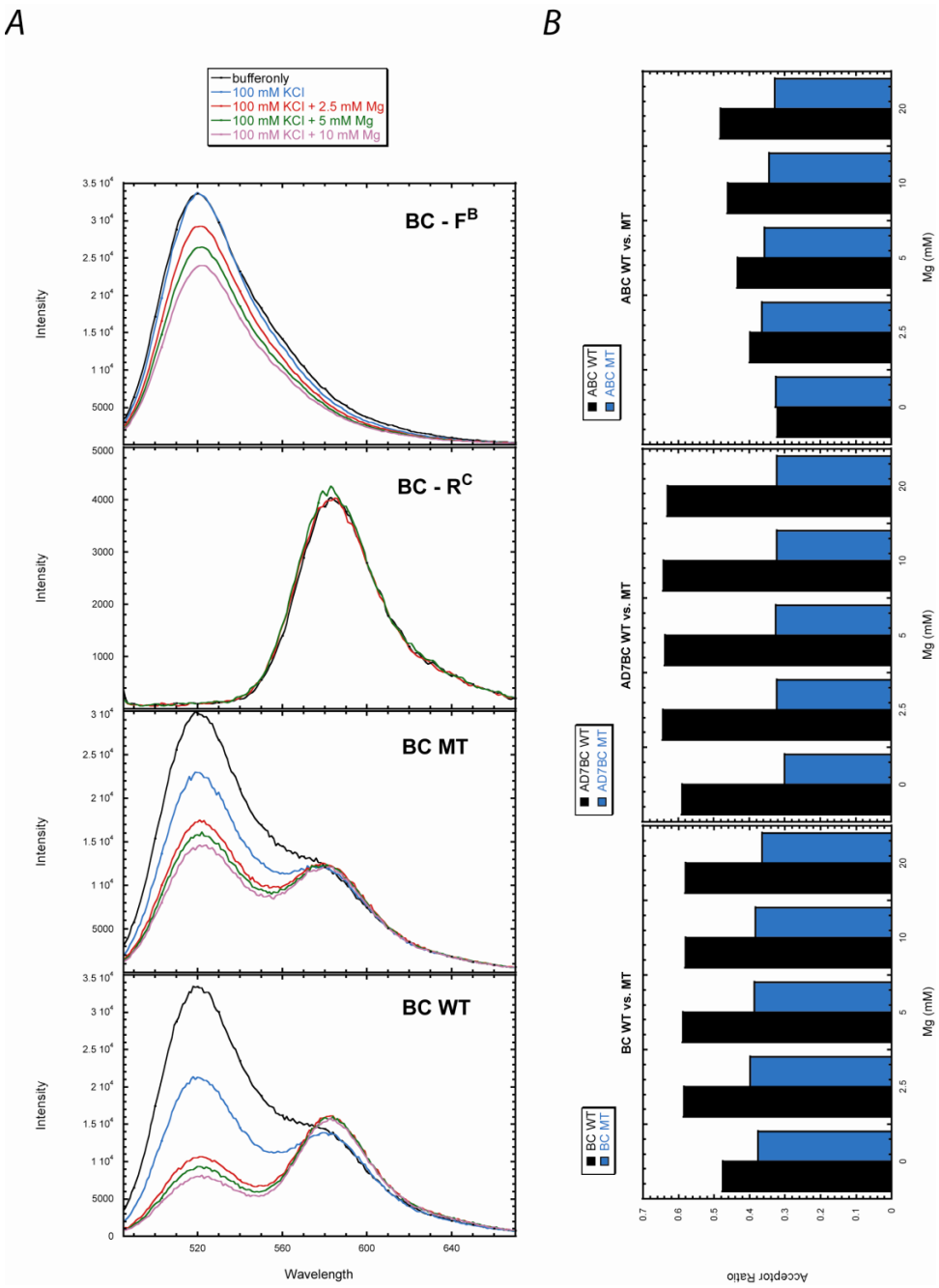

Figure 40 Fluorescence resonance energy transfer data of MHV PK. A) Fluorescence emission spectra for PK construct BC are plotted. A donor only spectra is presented in the top panel followed by an acceptor only spectrum. The lower panels display data for the mutant pseudoknot construct PK MT as well as a plot of the wildtype PK. B) Bar graph representation of acceptor ratio data for all ligated constructs where ABC is UTR, Δ7UTR is AΔ7BC and PK is BC. All samples contained a base solution of 100 mM KCl with 10 mM potassium phosphate pH 6. The x axis indicates the added amount of magnesium to the base solution. Acceptor ratios were calculated as described in Materials and Methods.

Based on the results obtained from thermal melts (see Chapter II), in the absence of Mg^{2+} the pseudoknot conformation should be unfolded. This will provide a spectroscopic baseline for the unfolded state. With subsequent addition of Mg^{2+}, an increase in FRET should be observed correlating with the formation of stem $P1^{pk}$.

Inspection of the BC-F^B spectra indicate that as Mg^{2+} is added to the donor only sample, a significant decrease in intensity is observed. Fluorescein dyes are known to be quenched under solution conditions containing metal and in solutions lower than pH 8. For the acceptor only sample, BC-R^C, addition of Mg^{2+} has no effect on the emission intensity. Spectra of BC WT RNA, which contains both fluorophores, includes emission peaks for both F and R dyes. As the Mg^{2+} concentration is increased, a decrease is observed in the intensity of the F peak with a corresponding increase in acceptor R emission intensity. Inspection of the BC MT (PK MT) RNA sample, also displays emission peaks representing both fluorophores F and R, identical to what was observed for BC WT. After the addition of magnesium however, the donor F intensity decreases but no subsequent increase in acceptor R is observed. The intensity of the acceptor never alters from the 10 mM phosphate 100 mM KCl buffer fluorescence intensity initially observed. From of the excitation spectra for the PK RNA samples it is apparent that magnesium quenches the fluorescence intensity of the F donor. F donor quenching is not complete so that remaining fluorophores in solution are present that can still participate in FRET. This is clear from the observed increase in acceptor R as Mg^{2+} is added. It is also clear that the acceptor is unaffected by the addition of magnesium so

the increase in intensity in the presence of donor must be coming from sensitized emission as a result of FRET.

To analyze the difference in FRET acceptor ratio among all RNAs, experiments similar to those outlined in Figure 39 were carried out for all of the UTR samples. Figure 40B displays bar graphs that compare the calculated acceptor ratio value as a function of Mg^{2+} ion for each of the six parts. Black bars represent the wild-type samples while the blue bars represent the corresponding pseudoknot mutant RNA which are impaired in $P1^{pk}$ folding. For the PK RNA (BC WT) an increase in acceptor ratio is observed with an increase in magnesium concentration. In contrast, the PK MT (BC MT) construct seen in blue, no increase is observed with an increase in Mg^{2+} ion up to 20 mM. Analogous observations of the Δ7UTR RNA (AΔ7BC) also illustrate an increase in acceptor ratio as compared to the mutant construct in blue (AΔ7BC MT). Inspection of the acceptor ratio at the highest Mg^{2+} concentration for both the WT PK and Δ7UTR RNAs indicate that the PK RNA has a slightly lower ratio than Δ7UTR RNA. In contrast, the mutant RNAs indicate that Δ7UTR RNA has a slightly lower acceptor ratio than PK MT RNA. Inspection of the UTR (ABC) RNA indicates that initially the wild-type and UTR MT (ABC MT) RNAs have similar FRET ratios. However after addition of magnesium, the FRET data of the UTR RNA increases while that of the UTR MT RNA remains essentially unchanged. A comparison of the maximum FRET ratio values obtained for the individual RNAs at the highest Mg^{2+} concentration tested (20 mM) reveals that Δ7 UTR and PK RNAs reach about the

same value (~0.60, ~0.65, respectively), while the intact UTR RNA rises to only 0.48 over the same solution conditions.

Using the individual MT RNAs as a baseline for the unfolded state of each wild-type RNA, it is clear that the initial FRET efficiency is decreased relative to that observed in the corresponding WT RNAs. In all three MT RNAs (PK, Δ7UTR and UTR) there is no increase in FRET efficiency with the addition of Mg^{2+}. Interestingly, a maximum FRET efficiency is observed in the PK WT and Δ7UTR WT RNAs with small additions of Mg^{2+} which is distinct from that which is observed for the UTR RNA. The UTR RNA initially starts with a ratio equivalent to the MT construct but only gradually increases with increasing Mg^{2+}. Since these FRET experiments are performed in an "ensemble" mode, this increase in FRET is likely reporting on a shift in the conformational distribution of folded molecules in solution. As the Mg^{2+} ion concentration increases, more UTR RNA molecules appear to adopt the $P1^{pk}$ folded state.

FRET analysis of BCD. Recently, Masters et al. reported the results of molecular probe experiments that suggest that sequences downstream of the pseudoknot stem P2 in MHV base pair with sequences within loop L1 (*117*). This base pairing interaction appears to be conserved in other CoV 3' UTRs. A cartoon that incorporates these findings into a model for the initiation of viral replication is shown in Fig. 41. In a

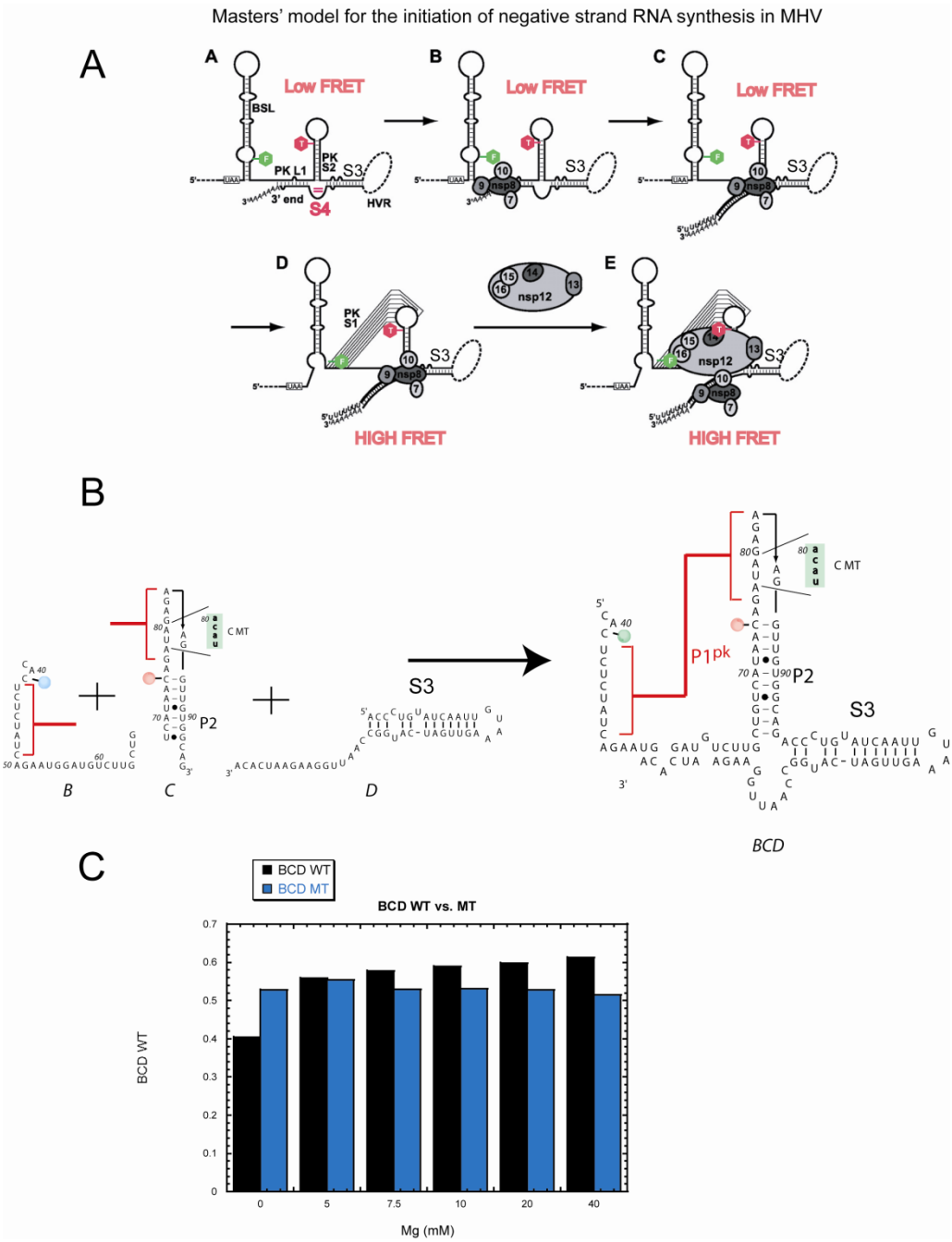

Figure 41 Extended UTR FRET data and acceptor ratio analysis. A) Revised model based off of Zust et al. The presence of the extended stem loop S3 interacts with the loop L1 of the pseudoknot (*117*). The interaction may be mediated by Protein interactions which stabilize one conformation over another. In the pseudoknot unfolded state the FRET will be low and if pseudoknot pairing is present FRET will be high. The presence of S3 and its effect on pseudoknot pairing was monitored using FRET. B) RNA constructs used in the ligation of the extended UTR. Fragments B and C were added onto fragment D creating the extended UTR. C) Acceptor ratio data for the BCD construct plotted as wild-type BCD versus pseudoknot mutant BCD MT RNA.

preliminary experiment to address the degree to which these putative base pairing interactions (*121*) influence formation of the pseudoknot, we employed a FRET ligation scheme similar to that discussed above. Specifically, we extended pseudoknot fragment BC to include a D fragment representing the extension proposed by Zust et al.(*117*) (see Fig. 41) which we propose results in the formation of a pseudo-4-helix junction that incorporates P2 and three new helical stems denoted L1S, S3 and S4 (*117*). In this proposal, note that proposed stem S3 is a hairpin in which the entire hypervariable region (HVR) was replaced with a GNRA (GUAA) tetraloop with no effect on viral replication in culture (see Fig. 41A). Utilizing this information, we substituted the HVR the same GUAA tetraloop to create BCD RNA which is capable of formation of the pseudo-4-helix junction with intact $P1^{pk}$ pairing. These new base pairing interactions could be indicative of a potential site of regulation mediated by conformational exchange involving the interactions of viral and host encoded proteins which could drive this minimal pseudoknotted state but only after L1S pairing has been broken due to replication of the extreme ends of the genome (see Fig. 41B). To determine if pseudoknot pairing in the BC RNA is influenced by the 3' extended sequence, BCD and BCD MT RNA were synthesized and FRET efficiencies analyzed exactly as described (see Fig. 41). A preliminary analysis of the FRET acceptor ratio reveals that in the absence of divalent cations, the BCD RNA is comparable to that of the BC and Δ7 UTR RNAs. Further, this value increases in a way that is comparable to the RNAs where Mg^{2+} is added. Unexpectedly, the mutant BCD RNA, BCD MT, in which pseudoknot pairing is abrogated, is characterized by a higher acceptor ratio value relative to WT

RNA in the absence of Mg^{2+}, but does not change significantly upon addition of Mg^{2+} (see Discussion).

DISCUSSION

Analytical ultracentrifugation can provide information about population distribution within a sample of macromolecules in solution. Under favorable conditions, it allows for the detection of multiple folded states. This method is ideal for conformationally switching states and provides a probe for determining how mutations effect population distributions (*121*).

Using the subdivided constructs from the MHV 3' UTR, we find that the PK and PK MT RNAs contain only one conformation in solution that is identified as the unfolded $P1^{pk}$ state in the absence of added divalent cation. The subsequent addition of magnesium produces two conformational states only in the PK RNA. The absence of this second population within the PK MT RNA demonstrates that this population is reporting on the pseudoknotted state. Within the PK RNA two folded states are still present even in under conditions of high concentrations of Mg^{2+} (20 mM) indicating that Mg^{2+} is not sufficient to stabilize one conformation to the exclusion of another. When the same titration was conducted with the PK MT RNA, no substantial alternately folded conformation was detected, even at the highest Mg^{2+} concentration tested. Furthermore, the initial population present under buffer-only conditions in the absence of magnesium in both PK and PK MT RNAs is in fact the unfolded state of PK, which NMR experiments reveal contains only stem P2. The formation of a pseudoknotted state only

upon addition of magnesium is consistent with both the optical melting profiles and the NMR spectra.

AUC analysis of the Δ7 UTR RNAs, which contains the added P0b segment, also give rise to results that are similar to those of for PK RNA pair. Under conditions containing buffer with no added Mg^{2+} ion, only a single species is present in solution for both the Δ7 UTR and the Δ7UTR MT RNAs. After the addition of Mg^{2+} a second population emerges only for the Δ7 UTR RNA analogous to results observed for the PK RNA constructs. In the case of the pseudoknot-deficient Δ7 UTR MT RNA, the presence of only one population of folded molecules again reveals that the pseudoknotted state is absent. The results observed for Δ7 UTR can be interpreted similarly to those obtained for the PK RNA constructs. The conditions employing buffer without divalent ions is reporting on the unfolded pseudoknot state. The addition of Mg^{2+}induces the formation of the pseudoknotted state for wild-type RNA only.

AUC experiments for the full length UTR, which is capable of folding in either the pseudoknotted state or the double stem state, reveal that under buffer conditions without divalent ions, only one folded state seems to be present for both the UTR WT and UTR MT RNAs. Though the UTR RNA exhibited one folded state for conditions containing buffer only, the population distribution for the peak is extremely broad. This is suggestive of multiple folding states or two states (PK vs. non-PK) that are exchanging slowly in the time course of the experiment. This broad peak is not observed for the UTR MT RNA construct indicating a narrower range of conformational states. With the addition of magnesium the UTR construct clearly establishes two

folding populations. Unexpectedly, the UTR MT RNA also produces two folded states in contrast to the previous subdivided MT RNA constructs which contain a single folding state regardless of the amount of Mg^{2+}added. Although the UTR MT RNA contains a second population, it is attenuated when compared to the population intensity observed for the UTR WT RNA. This would suggest that the larger population state is reporting on the presence of the pseudoknotted conformation. The apparent presence of the pseudoknot folded state within the context of the UTR MT RNA suggests that there may be some residual pseudoknotted or PK like structure despite the incorporation of the $P1^{pk}$ mutation designed to abrogate pseudoknot pairing. This could be reflective of the possibility that the four nucleotide mutation did not completely remove pseudoknot pairing in the context of the intact UTR construct. This is consistent with the data observed for the thermal melt of the UTR MT in Chapter II. The UTR MT RNA still contains an unfolding transition for the $P1^{pk}$ stem as indicated by the presence of transition **a** (see Fig. 23). The presence of the attenuated pseudoknot state could be the result of the intact molecule providing a structural scaffold which promotes the positional proximity for pseudoknot folding even in the absence of a fully base paired pseudoknotted stem.

Investigation of all three sets of AUC data suggests that multiple populations exist even in the absence of competition of base pairing of the lower P0a stem. For example, the PK and Δ7 UTR RNAs that lack the ability to make a P0a helix show more than one conformation in solution regardless of the amount of Mg^{2+} added (up to 20 mM) in solution. When a mutation is introduced into the pseudoknot pairing region

however, both constructs seem to occupy the unfolded state. In the case of the full length UTR RNA, two conformations are observed in both the wildtype construct and the mutant construct, however the mutant indicates a greatly attenuated pseudoknotted state. In order to obtain additional insight into AUC experiments, fluorescence resonance energy transfer (FRET) was employed to investigate the extent of pseudoknot folding in the three RNA constructs. FRET markers specific for pseudoknotted folding were introduced into each RNA, and fluorescence spectra acquired to assess their conformational characteristics. Analysis of the FRET acceptor ratio provided a means to compare the different constructs under similar solution conditions over a series of Mg^{2+} concentrations.

Analysis of the PK and PK MT RNAs acceptor ratio values provides a baseline for the folded and the unfolded pseudoknotted state. Clearly, less FRET is observed for the PK MT construct as compared to the PK state as would be expected since the markers are positioned to report on pseudoknot pairing. Using this baseline as a comparison for the Δ7 UTR folded state, it can be seen that once again, the Δ7 UTR MT construct has a decreased level of FRET relative to the Δ7 UTR construct which contains a fully functional pseudoknotted stem. These results are consistent with the presence of the folded pseudoknot after the addition of magnesium as was observed in the thermal melts and the NMR experiments. With the baseline established for the folded and unfolded states of the pseudoknot, analysis of the UTR constructs reveals that FRET is occurring in the UTR RNA although it appears slightly attenuated when compared to the fully folded PK RNA even in the presence of the same Mg^{2+} conditions. Alternatively,

the UTR MT RNA adopts an unfolded state consistent with the baseline established by the PK MT RNA. Taken together, these data suggest that the pseudoknot conformational state is indeed present but that under these steady state conditions, the UTR never adopts a level of folding observed for the PK and Δ7 UTR RNAs. This is consistent with the idea that competition exists within the UTR RNA decreasing the population of the folded state as observed for the level of the non-competitive PK-only RNAs.

Zust et al. (*117*) proposed that the extreme 3' end of the UTR, continuous with the poly(A) tail could base pair with nucleotides in the PK loop L1 and thereby perhaps stabilize the double hairpin conformation(*117*). To test this prediction and determine the degree to which pseudoknot folding is still possible, we performed analogous FRET experiments with the BCD RNA. Analysis of the FRET acceptor ratio obtained for this RNA in the absence and the presence of increasing Mg^{2+} reveals a qualitative similarity to that which occurs in BC (PK) RNA in the absence of the 3' extension. This suggests that the 3' extension has little influence on the degree of PK pairing. Consistent with our other constructs, Mg^{2+} has no effect on the FRET acceptor ratio in the BCD MT RNA. Somewhat puzzling however is the folding for the initial value for the BCD MT RNA is quite high (0.52). The origin of this is unknown but a possibility is that the mutant sequence favors the formation of a different folded or misfolded RNA. Additional experiments will be required to understand this in some detail.

To fully understand how the function of the S3, S4 and L1 helices influence pseudoknot pairing in 3' UTR, FRET experiments will need to be conducted with D7

ABCD and ABCD RNAs the latter of which represents a model for the entire 3' UTR. This will allow us to test the influence of the extended pseudoknot sequences in the context of a steady state ensemble arrangement. Additionally, it would prove beneficial to analyze the system using a single molecule approach. The steady state system is hampered by the averaging of conformations so that we are not assessing any singly folded conformation rather a series of states which can complicate the analysis. For instance, the UTR RNA never attained the FRET ratio observed for the PK RNA construct because it is apparently in a dynamic average between the double stem and the pseudoknot conformations. Our FRET experiments are only able to report the average distribution of distance. With a single molecule approach to the 3' UTR, the entire conformational switch progression could be observed and quantified allowing for real time analysis of folding stability and orientation which would answer the impossible questions leftover from the steady state approach.

CHAPTER V

CHARACTERIZATION OF THE HUMAN MA3 PSEUDOKNOT

INTRODUCTION

In the previous chapters we utilized experiments designed to introduce mutations in the 3' UTR of MHV coronavirus as a means to understand the alteration in the proportion of pseudoknot to hairpin conformation and how that would transfer to the replication of the single stranded RNA genome. In this chapter, we present the results of our studies of a frameshifting pseudoknot found in a human gene. Like the 3' UTR of MHV and other coronaviruses, our structural studies support the presence of a pseudoknot-double hairpin equilibrium in solution.

It is well known that RNA viruses utilize programmed ribosomal frameshifting (PRF) during the expression of their genomes during viral replication. Viruses typically contain small genomes within viral particles that are limited in capacity of packaged cellular replication machinery. As a result of the lack of prepackaged replication proteins and the small RNA genome size, alternative mechanisms that allow for translation of multiple proteins from a single mRNA have evolved and are highly evolutionarily advantageous to viruses. Viral frameshifting systems such as Beet Western Yellow Virus (BWYV) (*122, 123*), Mouse Mammary Tumor Virus (MMTV) (*124, 125*) and Infectious Bronchitis Virus (*126, 127*) have been characterized and have lead to a basic understanding of the requirements necessary for efficient frameshifting. All frameshifting systems require a slippery site composed of seven nucleotides of the

sequence XXX YYY Z where the ribosome pauses, a linker region composed of six to eight nucleotides, and a downstream RNA structure that is often a hairpin type pseudoknot (*40*). The slippery sequence is the site of ribosomal recoding in which the zero frame triplet codon currently being read by the ribosome is induced to slip back one nucleotide in the 5' direction generating a new decoding frame and translating a new protein. The general motif of the slippery site consists of a zero frame according to X XXY YYZ and -1 frame according to XXX YYY Z, where X can be any stretch of three identical nucleotides, Y can be either A or U and Z must be A, C or U (*40*).

While our understanding of RNA viral, bacterial and yeast frameshifting systems has increased dramatically, the degree to which human gene expression is regulated by frameshifting is largely unknown. The most well-studied frameshifting system in higher eukaryotes utilizes a +1 frameshifting event and encodes genes involved in antizyme expression; this is part of a feed back loop that allows for the regulation of polyamine biosynthesis (*128*). -1 frameshifting events have also been suggested to occur in mammalian cells but our progress in understanding of these systems has not progressed to the same extent as viral systems. Although -1 frameshifting may be responsible for diseases such as Hungtingtins and Machado-Joseph disease (MJD), the synthesis of polyglutamine repeats in these diseases represent non-conventional or non-programmed cases of frameshifting (they lack slippery sites) and thus will not be discussed further (*129, 130*).

There are currently two known mammalian genes for which programmed -1 frameshifting is projected to occur in a manner analogous to viral RNA systems. These

are the mouse embryonal carcinoma differentiation regulated gene (Edr) (human ortholog PEG10) and the human Ma3 gene (*131*). In mice, Edr is expressed at high levels during embryogenesis with a temporal linkage to musculoskeletal development (*131*). Although Edr plays a role in mammalian development its precise role in this process is unclear. A more recently discovered human gene whose expression appears to be regulated through a -1 frameshifting event is Ma3 (*130*). The protein product of Ma3 has been implicated in immunity disorders indirectly involved in tumor development. The resulting autoimmune response to antigens expressed by the tumor cells erroneously target Ma proteins in affected cells (*130, 132*). That is, the cancer cells produce an antigen that binds to Ma3 gene products that is subsequently attacked by the immune system. The end result is that not only do patients have a tumor but they also suffer from severe autoimmune response whose effects have been linked to brain encephalitis and cerebral degeneracy (*133-135*).

The human Ma family of genes include Ma1, Ma2, Ma3, MAOP-1(Ma4) Ma5 and PNMA6A (Ma6). Orthologs have been found in rat, mouse, chimpanzee and macaque. All of the Ma family members contain features that are similar to retrotransposable elements that encode for viral Gag proteins (*130, 132*). Ma3 specifically contains a CCHC zinc finger found in retroviral nucleocapsid proteins, as well as other features of the retroviral Gag poly protein. Ma3 and Ma5 both contain a heptanucleotide slip site and a potential downstream pseudoknot in a genomic position analogous to the *gag-pro* junction of MMTV and related retroviruses. In contrast to Ma3, Ma6 does have the potential to form a stable stem loop or a pseudoknot near the

end of ORF1. Although the Ma3 gene product clearly plays a role in tumor progression, the extent to which the tumor pathology is connected to -1 PRF is not yet known (*130*).

Furthermore, biochemical characterization of the predicted downstream Ma3 frameshifting stimulatory signal reveals that it is characterized by a frameshifting efficiency of approximately 20% (*130*). These studies are consistent with a classic hairpin (H)-type pseudoknot which is characterized by features predicted to be similar to other retroviral pseudoknots, in particular with the *gag-pro* pseudoknot of MMTV (*124, 125*). Both RNA structural elements have similar stem and loop lengths and each is projected to incorporate an intercalated adenosine between the two helical stems, deletion of which abolishes frameshifting. Given the ongoing controversy as to the solution structure of retroviral *gag-pro* pseudoknots and the potential to characterize a pseudoknot associated with human disease, NMR analysis was conducted with the intent to solve the solution structure of this RNA. A series of experiments were initiated with the long-term goal to investigate the structure and dynamics of the Ma3 pseudoknot. To our surprise, the wild-type Ma3 RNA was found to be in equilibrium with a competing double stem conformation (see Fig. 42), with the latter clearly the predominant species under most solution conditions. The results of these studies are presented here.

MATERIALS AND METHODS

RNA preparation. RNA constructs were prepared using methods outlined in Chapter II.

NMR spectroscopy. NMR spectra were collected and analyzed using the same procedures outlined in Chapter III.

Thermal denaturation. Thermal melts were collected and analyzed using the same procedures outlined in Chapter II.

RESULTS

General approach. One-dimensional (1D) and two-dimensional (2D) homonuclear (^{1}H-^{1}H) NMR spectra were obtained to determine base pairing assignments within a series of Ma3 RNA structural domains and mutants RNAs. Mutants were designed to differentiate between the double stem and pseudoknot conformational states of the Ma3 RNA (see Fig. 42). The RNA construct used for Ma3 RNA was designed based on the frameshifting studies conducted by Wills et al. (*130*). Using the same methodology outlined in Chapter III, we obtained imino ^{1}H NMR as well as 2D ^{1}H-^{1}H Watergate NOESY which were then used to assign conformational states.

NMR investigation of the Ma3 RNA. Both the pseudoknot conformation and the double stem conformation contain a common hairpin designated as S1 in Figure 42. To provides a baseline for peak assignments we acquired 1D and 2D Watergate NOESY spectra (300 ms) in 90% H_2O (10 ºC, 10 mM potassium phosphate, pH 6.0, no added salt) (Fig. 43). Analysis of these spectra using the methods outlined in Chapter III indicate that one noncanonical G•U base pair (resonances connected by the *dashed* lines, upper panel), two A-U base pairs and five G-C base pairs are present and can be

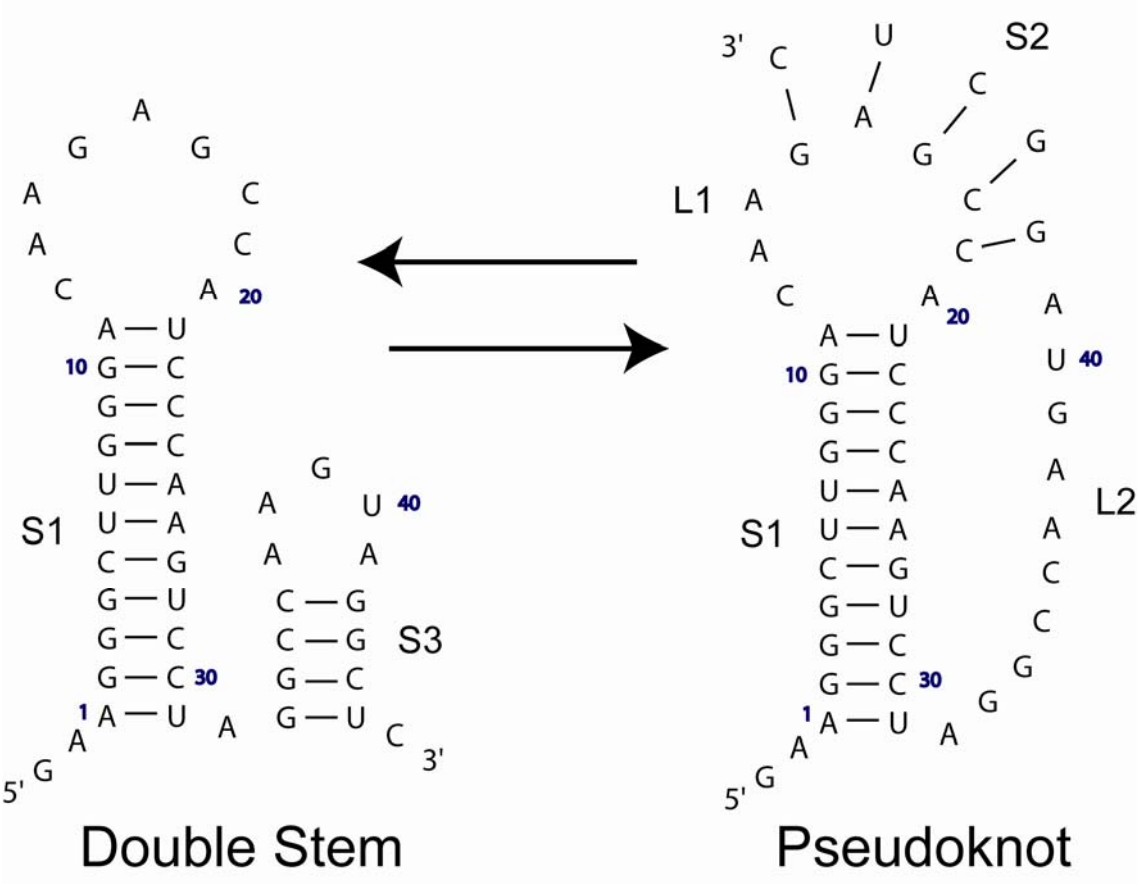

Figure 42 Proposed Ma3 RNA pseudoknot and stem loop conformational switch. Model based off of the frameshifting sequence used by Wills et al. (130)

confirmed through the imino "walk". These assignments are consistent with the proposed model of the S1 stem loop.

To understand how the folding of the S1 stem loop fits into the context of the Ma3 RNA, the same NMR experiments were conducted on the wildtype RNA sequence (Watergate NOESY (300 ms) in 90% H_2O, 10 ºC, 10 mM potassium phosphate, pH 6.0, no added salt) (see Fig. 44). S1 stem imino protons are readily assigned and are labeled in blue in Figure 44. In addition to the S1 imino protons, three additional G-C base pairs are observed (labeled in red in Fig. 44). Imino NOE data indicate that three G-C base

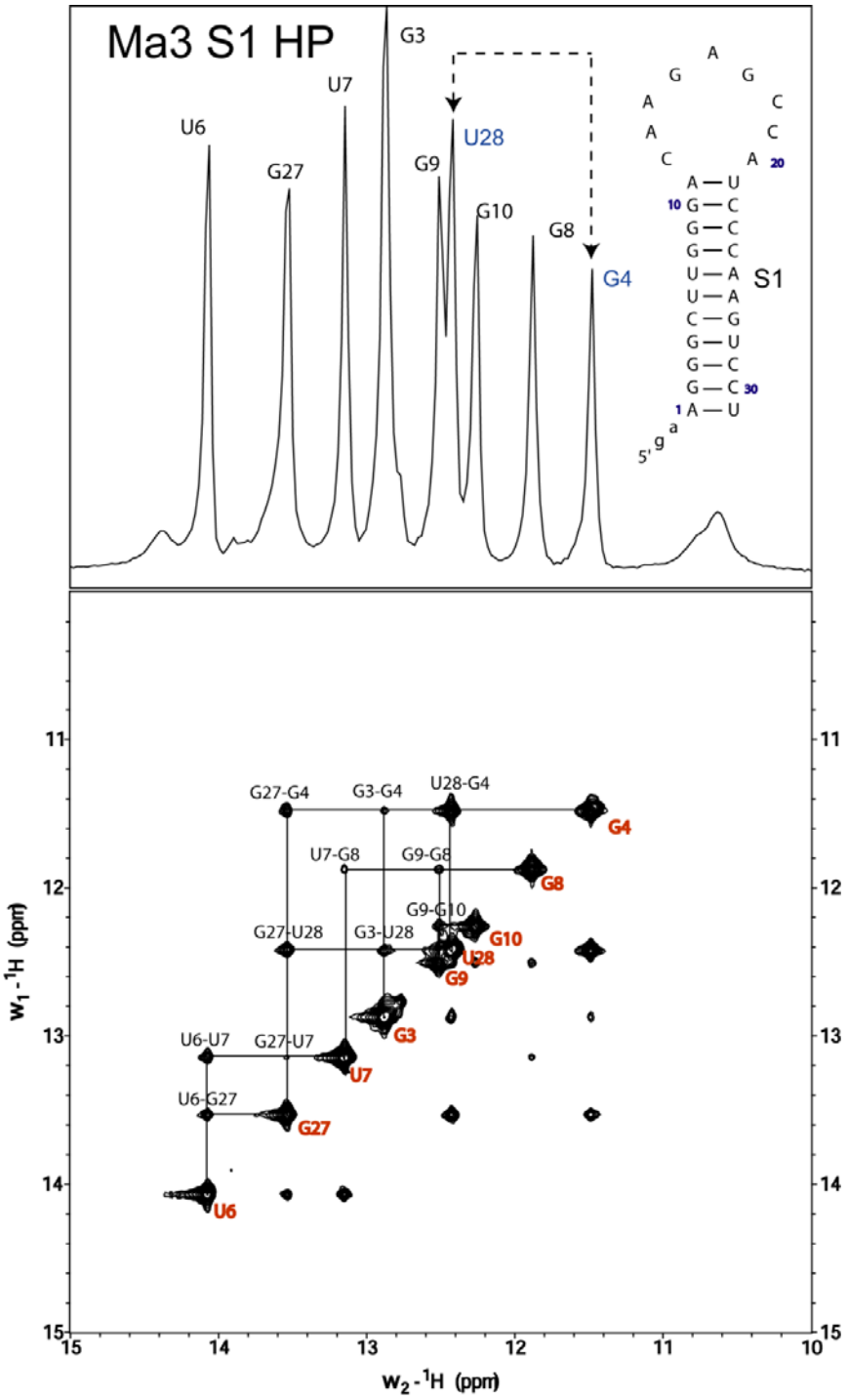

Figure 43 Ma3 S1 RNA NMR spectra. Upper panel, imino 1D spectra. Lower panel, 2D Watergate NOESY spectra. Inset into the upper panel is the RNA sequence used for the analysis. At the 5' end, 2 GA nucleotides were added for transcription purposes.

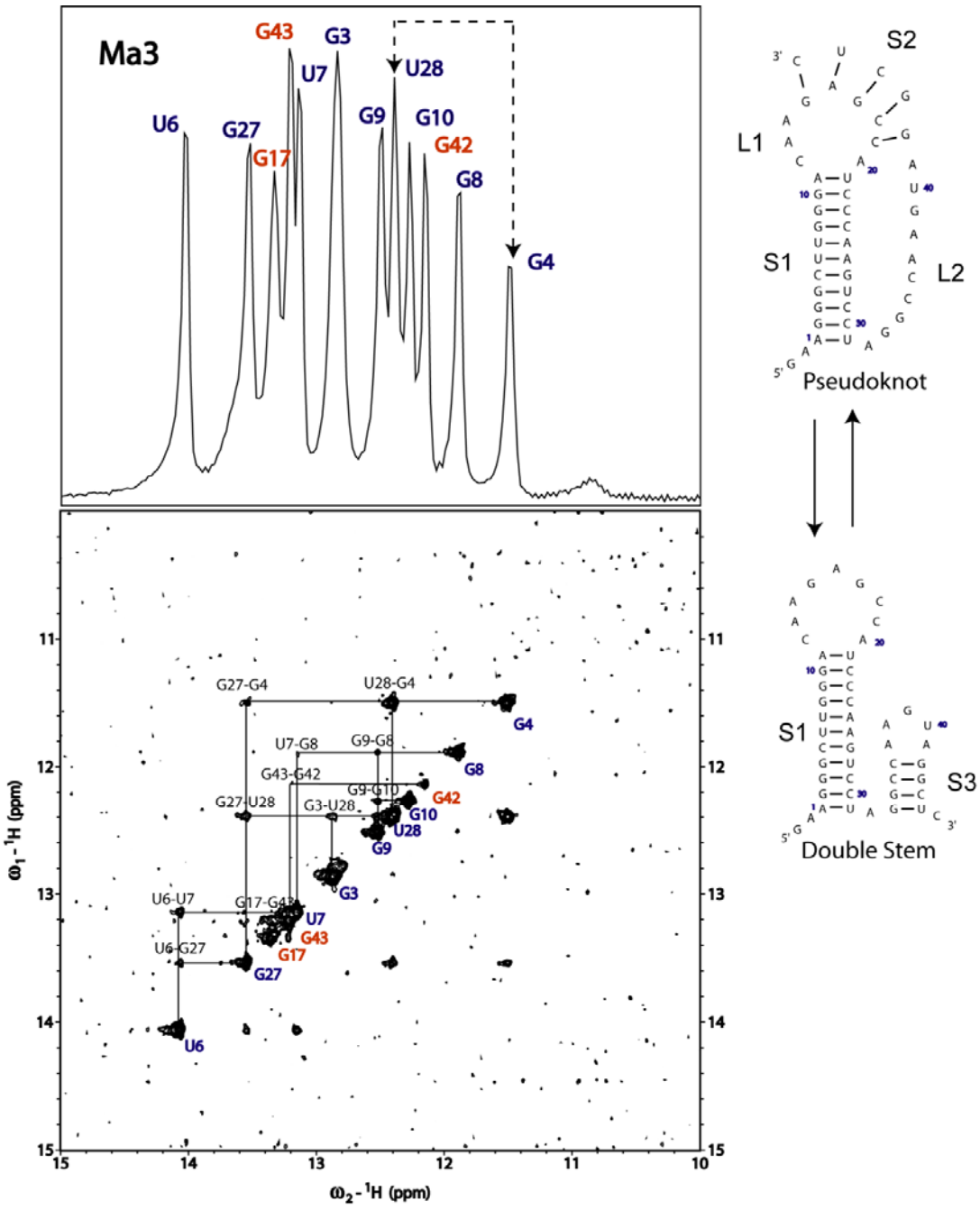

Figure 44 Ma3 RNA NMR spectra. Upper panel, imino 1D spectra. Lower panel, 2D Watergate NOESY spectra. Inset into the upper panel is the RNA sequence used for the analysis. It is represented as an equilibrium between the two possible states of the Ma3 RNA. Two 5' GA nucleotides were added for transcriptional purposes.

pairs are consecutive stem base pairs as revealed by the imino proton "walk" between NOE peaks.

Although it is clear that the Ma3 RNA contains a second folded helical stem structure other than the S1 stem, it is unclear if this stem corresponds to the pseudoknotted stem (S2) or the downstream stem S3, associated with the double stem conformation. Both the pseudoknot stem and the double stem conformations contain three potential G-C base pairs and will therefore give rise to similar 1D and 2D NOESY spectra. To unambiguously determine the conformational state of the Ma3 RNA, mutant RNAs were prepared to stabilize the folding of one conformation over the other. NMR spectra were then collected for each mutant and assigned according to the same methodology. In order to eliminate the function of pseudoknot stem S2, a Ma3 PK MT RNA was designed which changes nucleotides ^{16}AGCC^{19} to ^{16}UCGG^{19} (see Fig. 45B). S1 resonances are clearly defined as are three additional G-C base pairs (see Fig. 45 blue and red labeled peaks). Comparison of the imino protein spectra of the Ma3 and Ma3 PK MT RNAs reveal that they are essentially identical; this is consistent with the idea that the RNA is adopting primarily the double stem conformation rather than the pseudoknotted conformation.

To assess if formation of the pseudoknotted stem is possible another mutant RNA, designated Ma3 DS MT, which is predicted to abolish the formation of stem S3 in the double conformation, was prepared and characterized. The DS MT RNA converts nucleotides ^{33}GGCC^{36} to ^{33}CCGG^{36} (see Fig. 45C). Inspection of the NMR 1D and 2D Watergate NOESY spectra indicate that in addition to the S1 resonances two additional

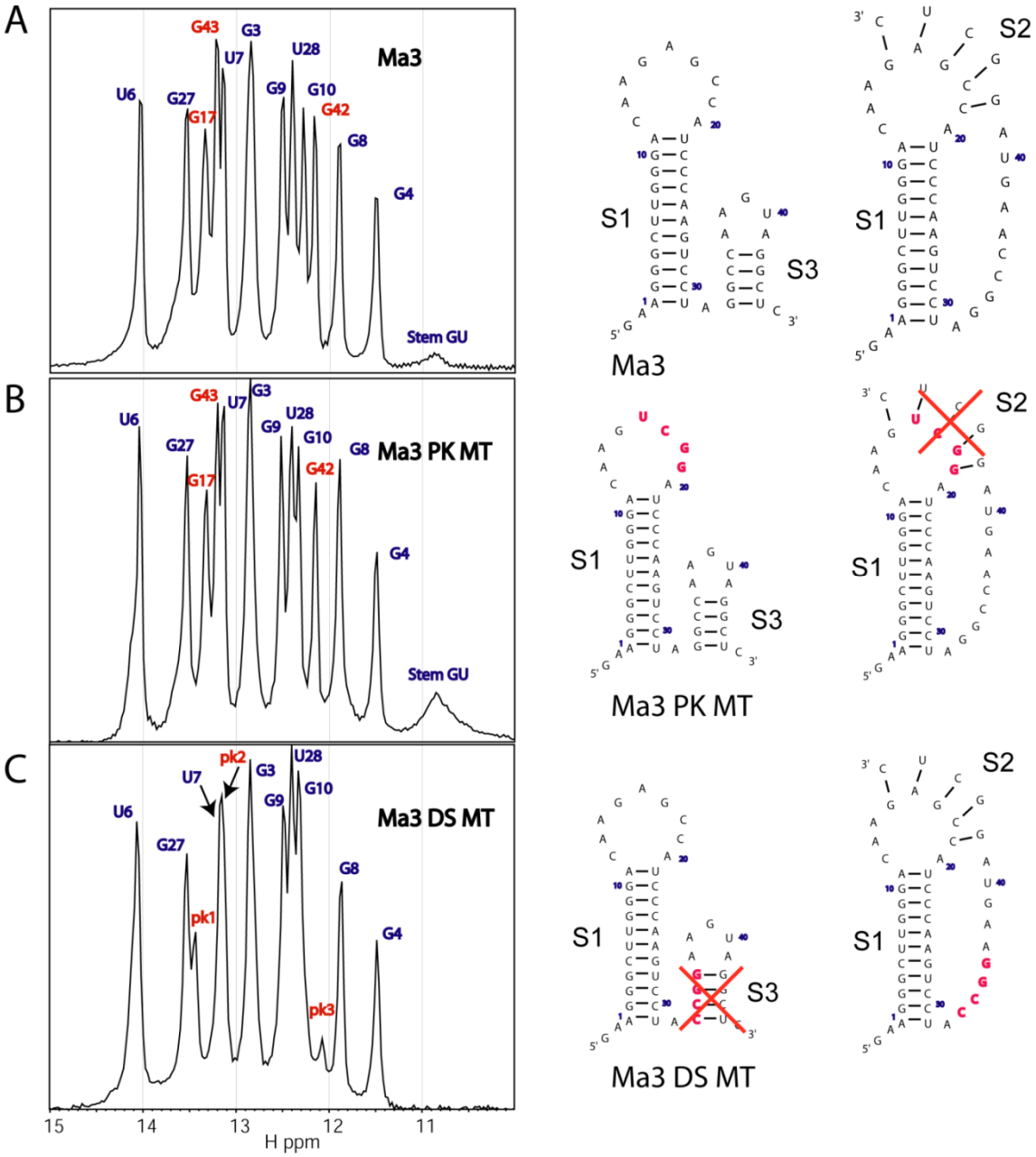

Figure 45 1D imino NMR spectral comparison for Ma3, Ma3 PK MT and Ma3 DS MT RNAs. A)Ma3 RNA wildtype 1D NMR spectra. Nucleotides labeled in blue are assigned to the HP construct. Nucleotides labeled in red belong to the double stem base pairs. B) Ma3pk MT RNA which has mutated the base pairs in the pseudoknot stem. Nucleotides labeled in blue are assigned to the HP stem and nucleotides labeled in red are assigned to the double stem base pairs. C) Double stem MT RNA has mutated the double stem conformation so that only the pseudoknot stem can fold. Nucleotides labeled in blue are assigned to the HP RNA and nucleotides labeled in red are assigned to the pseudoknot stem base pairs.

G-C base pairs are present. These likely derive from the pseudoknotted stem S2 (see Fig. 45C blue and red labeled peaks). These additional G-C base pairs are near to one another in space since they exhibit a sequential NOE connectivity (spectra not shown). These data reveal that the pseudoknotted conformation can form only when the double stem conformation is abrogated by mutation.

Thermal denaturation of Ma3 RNA. The Ma3 RNA predominantly adopts the double stem conformation in 10 mM potassium phosphate with no monovalent salt added. When magnesium is titrated into the sample and analyzed by 1D imino NMR, spectral changes including peak broadening and peak shifting occur (data not shown). These experiments suggest that there two or more conformational states of RNA are in exchange with one another, but could not be pursued further as a result of spectral degradation. To gain a better understanding of which conformer is preferred by the Ma3 RNA under various conditions, thermal melting experiments were carried out on the wildtype and mutant RNA (Ma3 PK MT and Ma3pk DS MT) structures by NMR. In addition to these RNAs, we also characterized Ma3 A33-36 (mutates ^{33}GGCC^{36} to ^{33}AAAA^{36} abrogates the double stem conformation) and Ma3 G34A RNAs (mutates ^{34}G to ^{34}A disrupts one base pair within the double stem conformation).

Thermal melting profiles were obtained for each RNA under two solution conditions: 10 mM potassium phosphate pH 6 (buffer only)or 10 mM potassium phosphate buffer pH 6 with 100 mM KCl and 5 mM $MgCl_2$ (Fig. 46). Inspection of the wildtype Ma3 RNA melting profiles reveals that under buffer only conditions, two unfolding transitions are present labeled **a** and **b** (see Fig. 46 left column). Transition **b**

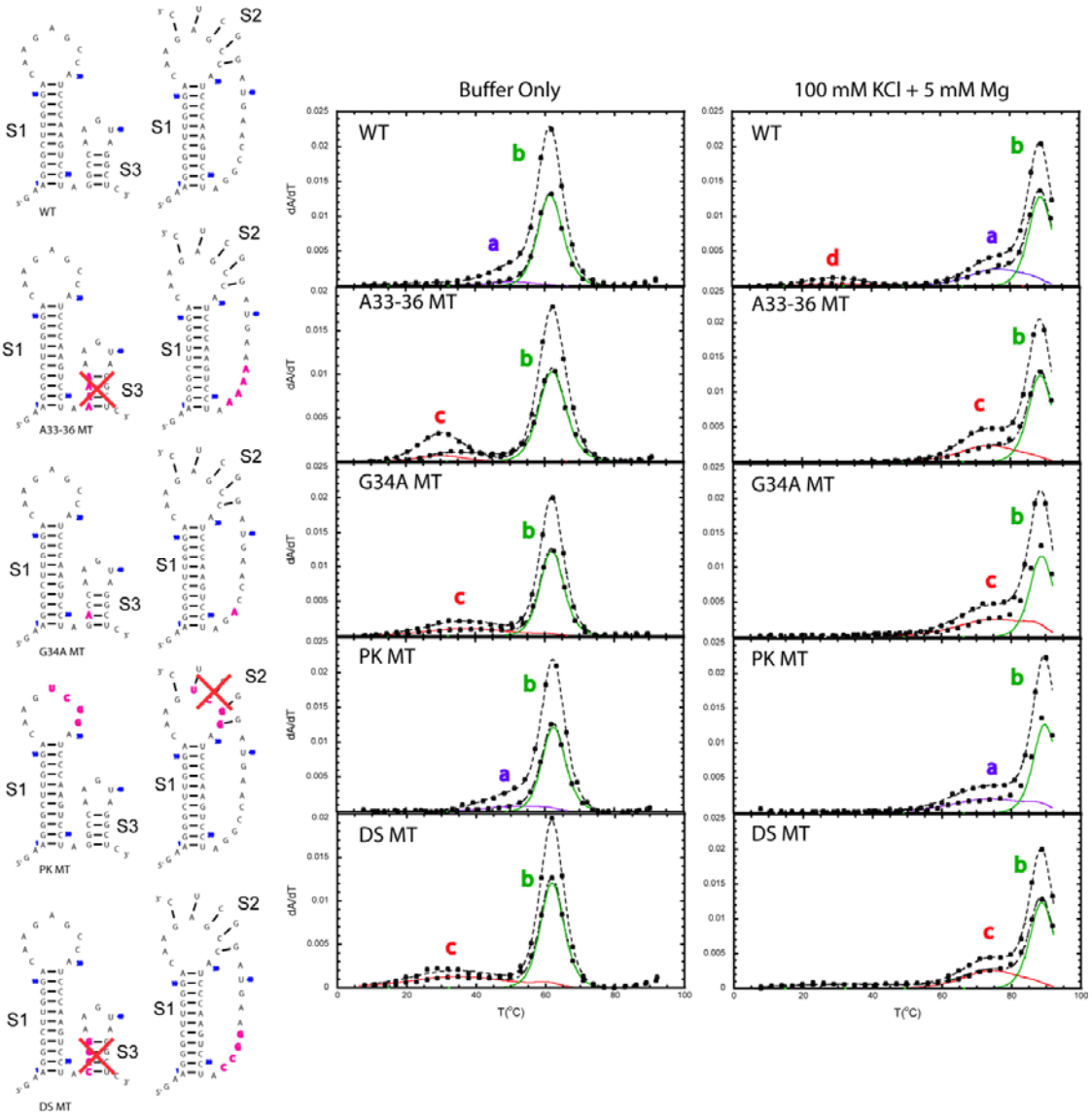

Figure 46 Thermal unfolding spectra of Ma3 RNA and mutant constructs. Left column represents the RNA sequences of the mutants analyzed. Each sequence is aligned with the corresponding optical melt spectra. Middle column contains the thermal melt spectra for each respective RNA construct under 10 mM phosphate buffer pH 6 only conditions. Right column contains the thermal melt spectra for each respective RNA construct under 10 mM phosphate buffer pH 6 with 100 mM KCl and 5 mM $MgCl_2$ conditions. Mutants A33-36, G34A, and Double Stem MT have disrupted the double stem conformation so that only the pseudoknot conformation is able to fold. Mutant PK MT has disrupted the pseudoknot stem so that only the double stem conformation is able to fold.

represents the unfolding of the stem S1 through comparison with standard thermal melts of the isolated S1 hairpin RNA (data not shown). Transition **a** is reporting on unfolding of the S3 stem exclusively as determined by imino proton NMR spectroscopy. Upon the addition of magnesium and KCl, three unfolding transitions are observed labeled **a**, **b** and **d** (see Fig. 46 right column). Once again the **b** transition represents the unfolding of the stem S1. Transition **a** melts at a temperature much higher than transition **d** which has a t_m of only 35 ^oC. Based on NMR spectra analyzed under the same conditions, the double stem conformation is the predominant form indicating that transition **a** is reporting on the unfolding of S3. The identity of transition **d** is unknown, but may be reporting on the unfolding of the pseudoknot stem S2.

To confirm these assignments of unfolding transitions **a**, **b**, and **d**, a PK MT RNA was prepared and subjected to thermal denaturation. Figure 46 reveals that under buffer only conditions, the PK MT RNA is characterized by two unfolding transitions (labeled **a** and **b** Fig. 46), apparently identical to that of the wild-type Ma3 RNA under the same buffer only conditions. Upon addition of magnesium and KCl, the same two transitions are present (labeled **a** and **b** Fig. 46). The absence of transition **d**, in the PK MT RNA that specifically disrupts stem S2, is consistent with the idea that the lower stability **d** transition is reporting on some aspect of pseudoknot unfolding. To investigate this further, mutants designed to abrogate double stem hairpin S3 (denoted A33-36, G34A and DS MT) RNAs were also prepared and characterized, again under "buffer only" and KCl/Mg^{2+}conditions (see Fig. 46). Inspection of DS MT, A33-36

Table 2 Thermodynamic data for Ma3 RNA constructs.

RNA	Conditions/Transition	t_m	ΔH	ΔG^{25}
	10 mM KH_2PO_4, pH 6			
WT	a	50	45.6	3.5
	b	61.4	89.1	9.7
PK MT	a	54.2	46.5	4.1
	b	63.3	96.9	11.0
DS MT	c	35.8	19.6	0.7
	b	61.7	97.3	10.7
A33-36	c	30.7	40.7	0.8
	b	62.1	83.7	9.3
G34A	c	38.2	23.6	1.0
	b	62	98.5	10.9
ΔA20	a	54.2	46.5	4.1
	b	63.3	96.9	11.0
	10 mM KH_2PO_4, pH 6, 5 mM Mg^{2+},100 mM KCl			
WT	d	29.9	32.5	0.5
	a	76.6	44.8	6.6
	b	88.7	109	19.2
PK MT	a	75.5	33.7	4.9
	b	88.6	110.9	19.5
DS MT	c	74.2	42.7	6.1
	b	88.7	108.7	19.1
A33-36	c	74.1	43.1	6.1
	b	88.5	108.3	19.0
G34A	c	75.5	33.7	4.9
	b	88.6	110.9	19.5
ΔA20	d	13.2	16.5	-0.7
	a	76.5	32.1	4.7
	b	89.4	107.6	19.1

and G34A RNA unfolding profiles reveal a second transition labeled **c** with t_m in the 30-40 ^oC range depending on the RNA. When KCl and Mg^{2+} are added, transition **c** moves to $t_m \sim 74\ ^oC$ ($\Delta H \sim 40$ kcal*mol^{-1}) in all three cases (see Table 2 for thermodynamic parameters), or to a position very similar to the unfolding transition of stem S3 in the WT and PK MT RNAs. This is consistent with the idea that the stability of stems S2 and S3 maybe expected to be quite similar. RNA folding simulations carried out by Dr. Shi-Jie Chen (University of Missouri) are largely consistent with this expectation where predicted unfolding transitions for S2 and S3 are 62 and 75 ^oC, respectively at 1 M NaCl for the wild-type and A33-36 RNAs (personal communication; data nt shown). Considering the t_m for S2, determined through the unfolding of RNAs A33-36, G34A and DS MT, and taking into account the corresponding predicted t_m for S2 calculated through computer simulations, this would seem to suggest that transition **d** (as observed in the wild-type RNA, see Fig. 46 top) is not reporting on the unfolding of stem S2. The shift in t_m between the WT transition **d** (at 5 mM Mg^{2+}) and the A33-36 mutant transition **c** is over 40 ^oC which leads to the conclusion that transition **d** is most likely not reporting on the unfolding of stem S2 (pseudoknot pairing).

Having used the mutant RNA constructs to assign unfolding transitions in the Ma3 RNA, another mutant was analyzed that investigated the stability of the pseudoknot folded RNA. Previous frameshifting assays, conducted by Wills et al. (*130*) on the mutant ΔA20 RNA, indicate that frameshifting is essentially eliminated as a result of the loss of this adenosine. This mutation involves an adenosine structurally homologous to the intercalated adenosine found in the helical junction of the similar frameshifting

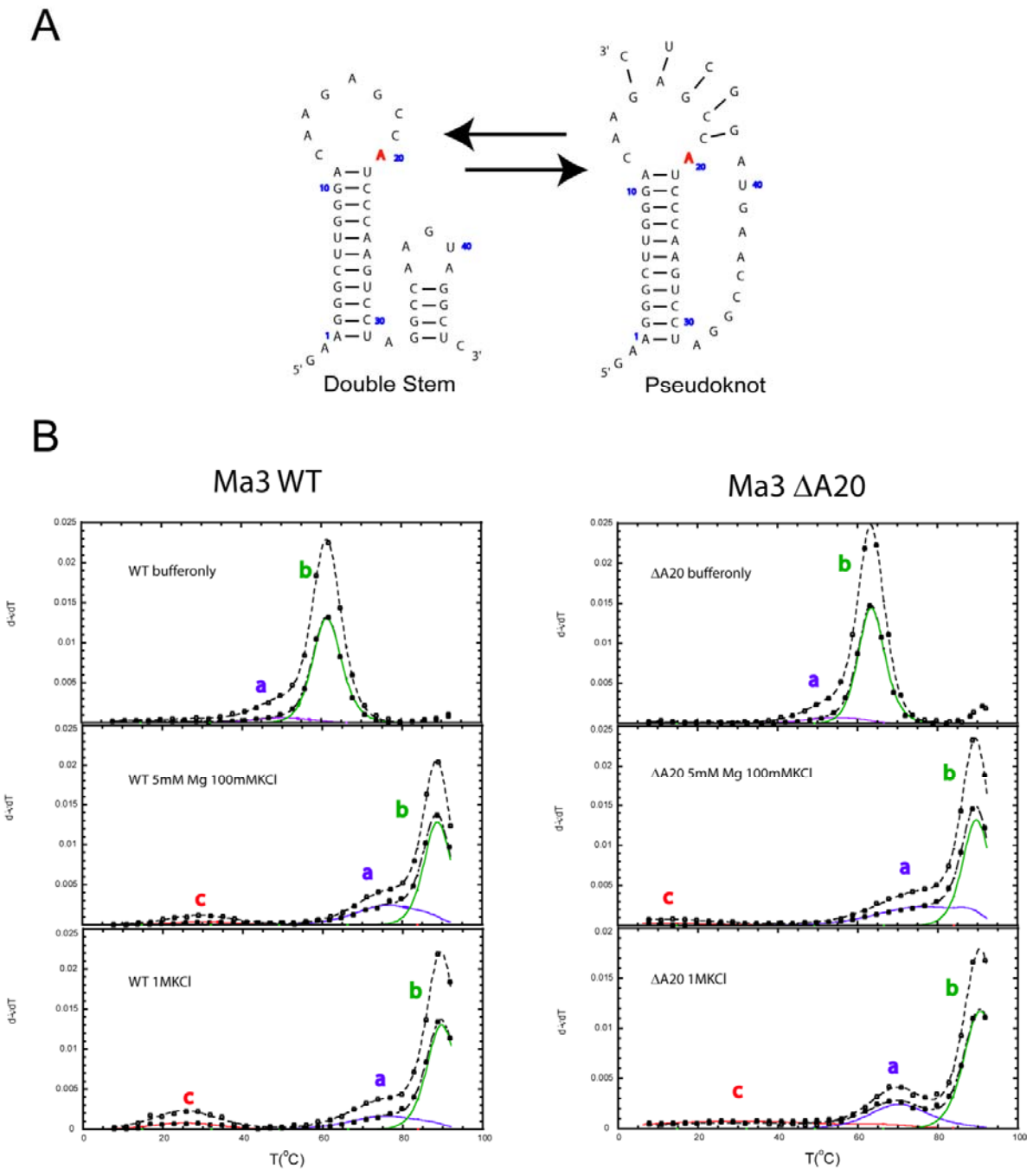

Figure 47 Thermal unfolding spectra of Ma3 and Ma3 ΔA20 MT RNAs. A) RNA secondary structure sequence of the Ma3 RNA switch with nucleotide A20 highlighted in red. B) Displays the optical melts of the RNA sequences under varying salt conditions. Ma3 wildtype RNA is in the left column and Ma3 ΔA20 MT is on the right.

pseudoknot found in MMTV the presence of which stabilized the pseudoknot by ~0.5 kcal*mol^{-1} (*97*). To understand how the stability of the RNA is effected by removing the adenosine, we decided to compare thermal melting spectra for the wildtype Ma3 RNA and the Ma3 ΔA20 RNA constructs. The spectra were analyzed under three different salt solutions. Comparison of the Ma3 RNA with the Ma3 ΔA20 RNA (see Fig. 47 left column and right column, respectively) indicates that under 10 mM phosphate buffer-only conditions, both spectra are identical with only transition **a** (S3 unfolding) and transition **b** (S1 unfolding). After addition of 5 mM $MgCl_2$ and 100 mM KCl three transitions are clearly observed (labeled **a**, **b** and **d**, Fig. 47) in the Ma3 RNA and the Ma3 ΔA20 RNA. While Ma3 ΔA20 RNA also contains three transitions (labeled **a**, **b** and **d**, Fig. 47) only a weak transition **d** is observed when comparing the wild-type Ma3 RNA. Quantitative analysis of the thermal melts reveals that transition **d** is destabilized by 1.2 kcal*mol^{-1} at 25 ^{o}C as compared to Ma3 RNA (see Table 2). Although transition **d** may not be directly reporting on stem S2 unfolding it could be reporting on the unfolding of tertiary structural interactions that specifically localize to the helical junction. The deletion of A20 may directly impact this tertiary interaction which is translated through the thermodynamic reduction in stability of transition **d** (*97*).

DISCUSSION

These physical studies reveal, contrary to expectations from functional studies, that the 48 nucleotide wild-type Ma3 RNA is capable of adopting two mutually exclusive RNA conformations and that a double stem loop conformation predominates at

very low salt conditions. Pseudoknot folding can only be induced with mutants that disrupt the secondary stem S3. These findings are consistent with RNA folding simulations carried out in collaboration with Dr. Shi-Jie Chen (University of Missouri) which reveal that the S1-S3 double stem loop conformer is the predominant form representing greater than 90% of the folded conformation in solution under 1 M NaCl and at 25 ^{o}C (data not shown). The same folding simulations carried out with the RNA mutants analyzed in Figures 46 and 47 exactly parallel the findings discussed here.

Wills and coworkers have previously carried out a mutational study of the Ma3 pseudoknot and have shown that deletion of A20 reduces frameshifting to just 3% of the wild-type level, *i.e.,* to a level just above the shift site alone. We show here that the unfolding of transition **d** in Ma3 is most likely reporting on tertiary structural interactions occurring in the helical junction. The stability of these tertiary interactions is significantly reduced, by 1.2 kcal*mol^{-1} for the Ma3 ΔA20 RNA, a value that is comparable to the 0.6 kcal*mol^{-1} ΔΔG observed for the analogous viral system in the MMTV *gag-pro* pseudoknot. These results are consistent with previous findings that indicate that the stability of the helical junction is important for frameshifting efficiencies (*45, 46, 96, 97, 113, 126, 136, 137*).

In a collaborative series of experiments, Wills and coworkers recently measured frameshifting efficiencies for the Ma3 mutants characterized here, in an effort to determine the functional significance of the structural findings summarized here (*130*). They employed a heterologous system in which the entire frameshifting signal is placed between glutathione *S*-transferase, and rabbit β-globin with efficiencies measured by

functional gel electrophoresis. These studies reveal that Ma3 DS MT RNA stimulates frameshifting to a level of 7% relative to Ma3 WT signal which has a frameshifting efficiency of 17%. In contrast, the Ma3 A33-36 RNA is characterized by a frameshifting efficiency of 33%, or a ≈2-fold *increase* in frameshift stimulation relative to the wild-type RNA. These findings are fully consistent with the conclusion reached earlier that the pseudoknotted conformation is the active structure since disruption of pseudoknot stem S2 results in an almost complete loss of activity, in contrast to the DS RNA which disrupts only S3. However, it is interesting to note that Ma3 A33-36 RNA which also abolishes the formation of stem S3 results in a significant increase in frameshifting efficiency. This suggests one or two possibilities. One is that simple elimination of competition of formation of the non-frameshifting S3 increases the amount of the functional amount of folded pseudoknot; this in turn increases frameshifting efficiency. However this result is inconsistent with findings of the DS MT RNA which mutates the same four base pairs, albeit to something other than AAAA. An alternative interpretation is that interaction of the run of the adenosine at the 5' side of the loop L2*, i.e.,* that region of L2 that is projected to interact directly with the translating ribosome (*40*), stabilizes the RNA through specific L2-S1 minor groove hydrogen bonding interactions observed in other frameshift stimulating pseudoknots (*112*). In any case, the parameters in Table 2 of the unfolding of transition **c** seem to argue against this. These functional and thermodynamic studies argue that the A33-36 Ma3 RNA represents an excellent structural target for detailed NMR structural and mechanical unfolding (*138*) experiments.

CHAPTER VI

SUMMARY AND PERSPECTIVES

SUMMARY

The inherent structural plasticity of RNA is vital to many biochemical systems (*13, 16, 21, 139*). The research summarized in this dissertation provides structural and thermodynamic insight into two distinct RNA systems, one being essential for viral replication of coronaviruses (*93-95*) and the other required for –1 translational frameshifting in the expression of a human gene (*130*). Although these systems are functionally very different, the work presented here provides direct support for the contention that each RNA is capable of adopting mutually exclusive folded conformations that are in equilibrium with one another. Although this conformational heterogeneity significantly complicates efforts to obtain high resolution solution structures of each RNA (the initial goal of this project), we have employed complementary approaches that collectively provide physical insights into the local (NMR spectroscopy, thermal melting profiles) and global (FRET spectroscopy; analytical ultracentrifugation) structural features of the conformational ensemble in each RNA.

The first RNA system is a proposed molecular switch between an extended double stem loop and a downstream pseudoknotted conformation, present in the 3' UTR of MHV, and in many other coronaviruses, including SARS-CoV (*93-95*). Goebel et al. (*94*) proposed that both mutually exclusive conformations are vital to viral replication

and viability of MHV. Our studies of the proposed molecular switch focused first on confirmation of the secondary structural model, including the pseudoknot, followed by structural and thermodynamic characterization. We also wished to determine if a conformational exchange or interconversion between conformers could be measured, and if so, define the molecular interactions involved in this process.

Our physical studies allow us to conclude that both an extended stem loop, P0ab, as well as a pseudoknot stem $P1^{pk}$ can indeed fold in the context of excised RNA subdomains (in PK and P0ab RNAs, respectively) as well as within the intact 3' UTR model (UTR RNA). Imino proton NMR spectroscopy was used to obtain the base pairing assignments and to confirm the presence of the competing of stem P0a within the UTR RNA. These studies establish that the double stem conformation is the predominant folded conformation in the absence of added Mg^{2+}. Unambiguous NMR assignments for the $P1^{pk}$ stem were only possible for a pseudoknot mutant that stabilized the $P1^{pk}$ stem, PK 44, but not for the wild-type PK RNA, due presumably to complications from chemical exchange broadening. NMR experiments also indicated that major helical stems of the UTR not involved in conformational switching, P0b and P2, were folded in the UTR RNA as well as appropriate subdomain fragments, thus satisfying minimal tenets of the model proposed by Goebel et al. (*93-95*).

Thermodynamic studies of individual UTR subdomains also generally provide support for the formation of the $P1^{pk}$ stem; however, thermodynamic studies reveal that it is quite weakly stable (ΔG^{25} = 1.0 kcal*mol^{-1}) even in the minimal PK RNA, and is likely not folded at a physiological temperature of 37 ^{o}C. Direct competition with

formation of the P0a stem reduces its "effective" stability even more so. These findings would seem to argue that a viral and/or host encoded protein is required to stabilize or transiently fold the pseudoknotted conformation in a regulatory manner. Alternatively, RNA sequences 3' to the PK region may have a significant impact on the stability of the pseudoknotted $P1^{pk}$ stem; however, a preliminary investigation into this possibility suggests that this is unlikely to be the case.

We also employed global hydrodynamic (analytical ultracentrifugation) and spectroscopic (FRET) methods to investigate the distribution of folded conformers in solution. These studies reveal that the PK RNA exists as an equilibrium between the folded $P1^{pk}$ pseudoknotted state and an unfolded P2 state even in the presence of Mg^{2+} at concentrations as high as 20 mM. The UTR RNA is also characterized by population distribution that is in equilibrium between a double stem conformation and a pseudoknot conformation. This population distribution was similar to that found for the PK RNA, in that a single, fully folded pseudoknot conformation could not be induced with Mg^{2+} concentrations up to 20 mM.

The second structurally interconverting RNA investigated in this work is a proposed hairpin (H)-type pseudoknot reported to in stimulate –1 translational frameshifting in the expression of the human Ma3 gene (*130*). Previous experiments identified what appeared to be a conventional retroviral-type pseudoknot (*40*) positioned downstream of the XXXYYYZ slip-site in the Ma3 mRNA (*130*). Our initial goal was to elucidate the solution structure of the Ma3 pseudoknot, a proposal motivated by the fact that the Ma3 pseudoknot was of mammalian cell origin, as well as inconsistencies in

the solution structures of previously characterized retroviral-type *gag-pro* pseudoknots (*96, 97, 126*). However, it became clear immediately that an alternatively folded, nearly thermodynamically isoenergetic, double stem conformation was the predominant conformation under solution conditions amenable to detailed NMR studies (*i.e.,* low salt conditions). It was predicted however, that the addition of Mg^{2+} ions would fold this RNA into the anticipated pseudoknot; this turned out not to the case as judged by preliminary imino proton NMR spectra acquired up to 20 mM Mg^{2+} ions (data not shown). These data therefore reveal that the pseudoknot is thermodynamically less stable when in direct competition with the double stem structure, a point reinforced by RNA folding simulations (S. Chen, personal communication). These findings remain puzzling since mutational studies strongly support the pseudoknotted RNA conformation as fully responsible for stimulation of –1 frameshifting in an entirely heterologous, non-native context, rather than the double hairpin (*130*) (N. Wills, personal communication). One possible explanation for this is that ribosome somehow enforces a pseudoknotted conformation and/or the lifetime of the pseudoknot is sufficiently long under the solution conditions of translocation so as to efficiently stimulate recoding.

PERSPECTIVES

The work presented in this dissertation establishes that the two RNAs studied here are capable of adopting at least two mutually exclusive conformations, presumably in equilibrium with one another. Simply changing the solution conditions or altering the RNA by mutation and/or deletion can apparently shift the equilibrium from one state to

the other. What remains unclear from these studies is a more quantitative description of the conformational heterogeneity of the ensemble, as well as the *rates* at which individual conformations interconvert. For example, using a highly sensitive FRET-based assay, we have documented the ability of Mg^{2+} to drive the formation of the pseudoknotted conformation of the MHV 3' UTR using a standard ensemble-based method. A more detailed description of the conformational ensemble, as well as rates of interconversion between states can only be obtained when individual molecules are interrogated at the single-molecule level, either using confocal microscopy-based diffusion or immobilization on a surface (*140*). Such an approach could also be used to directly measure the effect of solution conditions and protein binding on the conformational distribution of the states that make up the MHV 3' UTR molecular switch.

In addition, given that the double hairpin conformation of the MHV 3' UTR predominates under most solution conditions, it seems likely that viral or host-encoded protein(s) might drive the formation of the pseudoknot. Indeed, our FRET-based assay of this conformational switch could be used directly to probe fractionated crude cell lysates from infected vs. noninfected cells in order to identify a protein or protein complexes that induce the high FRET state, relative to a control RNA in which $P1^{pk}$ pairing is abolished, using mass spectrometry-based protein identification techniques. Furthermore, such an assay could be used in a high throughput screening approach to interrogate small molecule or combinatorial peptide libraries that would preferentially and irreversible stabilize the pseudoknotted conformation of the 3' UTR. Such

molecules would represent potential, broad-spectrum anticoronaviral agents used to specifically target human pathogenic coronaviruses, SARS-CoV and NL63 (*67, 68, 72, 73, 78, 141*). Simultaneously monitoring of the folding of stem P0a may also prove advantageous using a similar assay.

Finally, although our studies of the human Ma3 frameshifting pseudoknot were complicated by what is unlikely to be a biologically relevant or phylogenetically conserved competing equilibrium, they do identify mutant RNA constructs, *e.g.,* the A33-36 RNA, that could be used for detailed NMR structural and mechanical unfolding studies (*44, 142*). Recent "pulling" experiments suggest that the unfolding of frameshifting pseudoknots is strongly kinetically controlled (*143*) and an investigation of the Ma3 pseudoknot and selected mutants would seem to represent an excellent model system with which to investigate the generality of these findings.

REFERENCES

1. Crick, F. (1970) Central dogma of molecular biology, *Nature* **227**: 561-563.

2. Mattick, J. S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms, *Bioessays* **25:** 930-939.

3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell* **35:** 849-857.

4. Korostelev, A., Trakhanov, S., Asahara, H., Laurberg, M., Lancaster, L., and Noller, H. F. (2007) Interactions and dynamics of the Shine Dalgarno helix in the 70S ribosome, *Proc Natl Acad Sci U S A* **104**: 16840-16843.

5. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 A resolution, *Science* **310:** 827-834.

6. Selmer, M., Dunham, C. M., Murphy, F. V. t., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA, *Science* **313:** 1935-1942.

7. Zaug, A. J., and Cech, T. R. (1992) The intervening sequence RNA of Tetrahymena is an enzyme. 1986, *Biotechnology* **24:** 57-62.

8. Korostelev, A., and Noller, H. F. (2007) The ribosome in focus: new structures bring new insights, *Trends Biochem Sci* **32:** 434-441.

9. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis, *Science* **289:** 920-930.

10. Noller, H. F. (2006) Biochemical characterization of the ribosomal decoding site, *Biochimie* **88:** 935-941.

11. Callis, T. E., Deng, Z., Chen, J. F., and Wang, D. Z. (2008) Muscling through the microRNA world, *Exp Biol Med (Maywood)* **233:** 131-138.

12. Chu, C. Y., and Rana, T. M. (2007) Small RNAs: regulators and guardians of the genome, *J Cell Physiol* **213:** 412-419.

13. Edwards, T. E., Klein, D. J., and Ferre-D'Amare, A. R. (2007) Riboswitches: small-molecule recognition by gene regulatory RNAs, *Curr Opin Struct Biol* **17:** 273-279.

14. Grundy, F. J., and Henkin, T. M. (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements, *Crit Rev Biochem Mol Biol* **41:** 329-338.

15. Rana, T. M. (2007) Illuminating the silence: understanding the structure and function of small RNAs, *Nat Rev Mol Cell Biol* **8:** 23-36.

16. Sashital, D. G., and Butcher, S. E. (2006) Flipping off the riboswitch: RNA structures that control gene expression, *ACS Chem Biol* **1:** 341-345.

17. Tucker, B. J., and Breaker, R. R. (2005) Riboswitches as versatile gene control elements, *Curr Opin Struct Biol* **15:** 342-348.

18. Mehler, M. F., and Mattick, J. S. (2007) Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease, *Physiol Rev* **87:** 799-823.

19. Zhao, Y., and Srivastava, D. (2007) A developmental view of microRNA function, *Trends Biochem Sci* **32:** 189-197.

20. Stefani, G., and Slack, F. J. (2008) Small non-coding RNAs in animal development, *Nat Rev Mol Cell Biol* **9:** 219-230.

21. Fedor, M. J., and Williamson, J. R. (2005) The catalytic diversity of RNAs, *Nat Rev Mol Cell Biol* **6:** 399-412.

22. Spirin, A. S. (2004) The ribosome as an RNA-based molecular machine, *RNA Biol* **1:** 3-9.

23. Steitz, T. A. (2008) A structural understanding of the dynamic ribosome machine, *Nat Rev Mol Cell Biol* **9:** 242-253.

24. Steitz, T. A., and Moore, P. B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme, *Trends Biochem Sci* **28:** 411-418.

25. Strobel, S. A., and Cochrane, J. C. (2007) RNA catalysis: ribozymes, ribosomes, and riboswitches, *Curr Opin Chem Biol* **11:** 636-643.

26. Bartholomew, B. (Southern Illinois School of Medicine) Sugar pucker figure.

27. Mallery, C. H. (2007) Polynucleotide figure, in *Biology 150 - Introductory Biology for Majors*, Department of Biology, University of Miami.

28. Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H., Seeman, N. C., and Rich, A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA, *Science* **185:** 435-440.

29. Reiter, N. J., Maher, L. J., 3rd, and Butcher, S. E. (2008) DNA mimicry by a high-affinity anti-NF-kappaB RNA aptamer, *Nucleic Acids Res* **36:** 1227-1236.

30. Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F., and Klug, A. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution, *Nature* **250:** 546-551.

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

32. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing, *Science* **273:** 1678-1685.

33. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science* **289:** 905-920.

34. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit, *Nature* **407:** 327-339.

35. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution, *Science* **292:** 883-896.

36. Holbrook, S. R. (2005) RNA structure: the long and the short of it, *Curr Opin Struct Biol* **15:** 302-308.

37. Noller, H. F. (2005) RNA structure: reading the ribosome, *Science* **309:** 1508-1514.

38. Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., and Mathews, M. B. (2004) The double-stranded-RNA-binding motif: interference and much more, *Nat Rev Mol Cell Biol* **5:** 1013-1023.

39. Pleij, C. W., Rietveld, K., and Bosch, L. (1985) A new principle of RNA folding based on pseudoknotting, *Nucleic Acids Res* **13:** 1717-1731.

40. Giedroc, D. P., Theimer, C. A., and Nixon, P. L. (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting, *J Mol Biol* **298:** 167-185.

41. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function, *Mol Cell* **17:** 671-682.

42. Brierley, I., Pennell, S., and Gilbert, R. J. (2007) Viral RNA pseudoknots: versatile motifs in gene expression and replication, *Nat Rev Microbiol* **5:** 598-610.

43. Cornish, P. V., Giedroc, D. P., and Hennig, M. (2006) Dissecting non-canonical interactions in frameshift-stimulating mRNA pseudoknots, *J Biomol NMR* **35:** 209-223.

44. Cornish, P. V., Hennig, M., and Giedroc, D. P. (2005) A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting, *Proc Natl Acad Sci U S A* **102:** 12694-12699.

45. Kim, Y. G., Su, L., Maas, S., O'Neill, A., and Rich, A. (1999) Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency, *Proc Natl Acad Sci U S A* **96:** 14234-14239.

46. Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W., Heus, H. A., and Hilbers, C. W. (1998) NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA, *Science* **280:** 434-438.

47. Pfingsten, J. S., Costantino, D. A., and Kieft, J. S. (2006) Structural basis for ribosome recruitment and manipulation by a viral IRES RNA, *Science* **314:** 1450-1454.

48. Schuler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P. A., Westhof, E., and Spahn, C. M. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA, *Nat Struct Mol Biol* **13:** 1092-1096.

49. Gallego, J., and Varani, G. (2002) The hepatitis C virus internal ribosome-entry site: a new target for antiviral research, *Biochem Soc Trans* **30:** 140-145.

50. Fraser, C. S., and Doudna, J. A. (2007) Structural and mechanistic insights into hepatitis C viral translation initiation, *Nat Rev Microbiol* **5:** 29-38.

51. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D., and Stark, H. (2005) Structure of the hepatitis C Virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES, *Structure* **13:** 1695-1706.

52. Ji, H., Fraser, C. S., Yu, Y., Leary, J., and Doudna, J. A. (2004) Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA, *Proc Natl Acad Sci U S A* **101:** 16990-16995.

53. Nonin-Lecomte, S., Felden, B., and Dardel, F. (2006) NMR structure of the *Aquifex aeolicu*s tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation, *Nucleic Acids Res* **34:** 1847-1853.

54. Valle, M., Gillet, R., Kaur, S., Henne, A., Ramakrishnan, V., and Frank, J. (2003) Visualizing tmRNA entry into a stalled ribosome, *Science* **300:** 127-130.

55. Moore, S. D., and Sauer, R. T. (2007) The tmRNA system for translational surveillance and ribosome rescue, *Annu Rev Biochem* **76:** 101-124.

56. Haebel, P. W., Gutmann, S., and Ban, N. (2004) Dial tm for rescue: tmRNA engages ribosomes stalled on defective mRNAs, *Curr Opin Struct Biol* **14:** 58-65.

57. Wilson, D. N., and Nierhaus, K. H. (2003) The ribosome through the looking glass, *Angew Chem Int Ed Engl* **42:** 3464-3486.

58. Namy, O., Rousset, J. P., Napthine, S., and Brierley, I. (2004) Reprogrammed genetic decoding in cellular gene expression, *Mol Cell* **13:** 157-168.

59. Moran, S. J., Flanagan, J. F. t., Namy, O., Stuart, D. I., Brierley, I., and Gilbert, R. J. (2008) The mechanics of translocation: a molecular "spring-and-ratchet" system, *Structure* **16:** 664-672.

60. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J., and Brierley, I. (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting, *Nature* **441:** 244-247.

61. Plant, E. P., and Dinman, J. D. (2005) Torsional restraint: a new twist on frameshifting pseudoknots, *Nucleic Acids Res* **33:** 1825-1833.

62. Plant, E. P., Jacobs, K. L., Harger, J. W., Meskauskas, A., Jacobs, J. L., Baxter, J. L., Petrov, A. N., and Dinman, J. D. (2003) The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting, *RNA* **9:** 168-174.

63. Puglisi, J. D., Blanchard, S. C., and Green, R. (2000) Approaching translation at atomic resolution, *Nat Struct Biol* **7:** 855-861.

64. Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region, *Cell* **55:** 447-458.

65. Ogle, J. M., Carter, A. P., and Ramakrishnan, V. (2003) Insights into the decoding mechanism from recent ribosome structures, *Trends Biochem Sci* **28:** 259-266.

66. De Clercq, E. (2004) Antivirals and antiviral strategies, *Nat Rev Microbiol* **2:** 704-720.

67. Ziebuhr, J. (2004) Molecular biology of severe acute respiratory syndrome coronavirus, *Curr Opin Microbiol* **7:** 412-419.

68. Holmes, K. V. (2003) SARS coronavirus: a new challenge for prevention and therapy, *J Clin Invest* **111:** 1605-1609.

69. de Haan, C. A., and Rottier, P. J. (2005) Molecular interactions in the assembly of coronaviruses, *Adv Virus Res* **64:** 165-230.

70. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., and Snijder, E. J. (2006) Nidovirales: evolving the largest RNA virus genome, *Virus Res* **117:** 17-37.

71. Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J., and Gorbalenya, A. E. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage, *J Mol Biol* **331:** 991-1004.

72. Groneberg, D. A., Hilgenfeld, R., and Zabel, P. (2005) Molecular mechanisms of severe acute respiratory syndrome (SARS), *Respir Res* **6:** 8.

73. Masters, P. S. (2006) The molecular biology of coronaviruses, *Adv Virus Res* **66:** 193-292.

74. Enjuanes, L., Almazan, F., Sola, I., Zuniga, S., Alvarez, E., Reguera, J., and Capiscol, C. (2006) Biochemical aspects of coronavirus replication, *Adv Exp Med Biol* **581:** 13-24.

75. Pasternak, A. O., Spaan, W. J., and Snijder, E. J. (2006) Nidovirus transcription: how to make sense...?, *J Gen Virol* **87:** 1403-1421.

76. de Haan, C. A., and Rottier, P. J. (2006) Hosting the severe acute respiratory syndrome coronavirus: specific cell factors required for infection, *Cell Microbiol* **8:** 1211-1218.

77. Plant, E. P., Perez-Alvarado, G. C., Jacobs, J. L., Mukhopadhyay, B., Hennig, M., and Dinman, J. D. (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal, *PLoS Biol* **3:** e172.

78. Shi, S. T., and Lai, M. M. (2005) Viral and cellular proteins involved in coronavirus replication, *Curr Top Microbiol Immunol* **287:** 95-131.

79. Baric, R. S., Stohlman, S. A., and Lai, M. M. (1983) Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains, *J Virol* **48:** 633-640.

80. Sawicki, S. G., and Sawicki, D. L. (2005) Coronavirus transcription: a perspective, *Curr Top Microbiol Immunol* **287:** 31-55.

81. Sawicki, S. G., Sawicki, D. L., and Siddell, S. G. (2007) A contemporary view of coronavirus transcription, *J Virol* **81:** 20-29.

82. Sawicki, S. G., and Sawicki, D. L. (1995) Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands, *Adv Exp Med Biol* **380:** 499-506.

83. Li, L., Kang, H., Liu, P., Makkinje, N., Williamson, S. T., Leibowitz, J. L., and Giedroc, D. P. (2008) Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication, *J Mol Biol* **377:** 790-803.

84. Zhang, G., Zhang, J., George, A. T., Baumstark, T., and Simon, A. E. (2006) Conformational changes involved in initiation of minus-strand synthesis of a virus-associated RNA, *RNA* **12:** 147-162.

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

86. Zhang, J., Zhang, G., Guo, R., Shapiro, B. A., and Simon, A. E. (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.

87. Serrano, P., Pulido, M. R., Saiz, M., and Martinez-Salas, E. (2006) The 3' end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA-RNA interactions with the 5' end region, *J Gen Virol* **87:** 3013-3022.

88. Almazan, F., Dediego, M. L., Galan, C., Escors, D., Alvarez, E., Ortego, J., Sola, I., Zuniga, S., Alonso, S., Moreno, J. L., Nogales, A., Capiscol, C., and Enjuanes, L. (2006) Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis, *J Virol* **80:** 10900-10906.

89. Baric, R. S., and Yount, B. (2000) Subgenomic negative-strand RNA function during mouse hepatitis virus infection, *J Virol* **74:** 4039-4046.

90. Pasternak, A. O., van den Born, E., Spaan, W. J., and Snijder, E. J. (2001) Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis, *EMBO J* **20:** 7220-7228.

91. Zhu, H., Liu, Y., Cai, Y., Yu, D., Pu, Y., Harmon, L., and Zhang, X. (2006) Toward the development of an infectious cDNA clone of a human enteric coronavirus, *Adv Exp Med Biol* **581:** 527-530.

92. Yount, B., Denison, M. R., Weiss, S. R., and Baric, R. S. (2002) Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59, *J Virol* **76:** 11065-11078.

93. Goebel, S. J., Miller, T. B., Bennett, C. J., Bernard, K. A., and Masters, P. S. (2007) A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis, *J Virol* **81:** 1274-1287.

94. Goebel, S. J., Hsue, B., Dombrowski, T. F., and Masters, P. S. (2004) Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome, *J Virol* **78:** 669-682.

95. Goebel, S. J., Taylor, J., and Masters, P. S. (2004) The 3' cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome, *J Virol* **78:** 7846-7851.

96. Theimer, C. A., and Giedroc, D. P. (1999) Equilibrium unfolding pathway of an H-type RNA pseudoknot which promotes programmed -1 ribosomal frameshifting, *J Mol Biol* **289:** 1283-1299.

97. Theimer, C. A., and Giedroc, D. P. (2000) Contribution of the intercalated adenosine at the helical junction to the stability of the gag-pro frameshifting pseudoknot from mouse mammary tumor virus, *RNA* **6:** 409-421.

98. Williams, G. D., Chang, R. Y., and Brian, D. A. (1999) A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication, *J Virol* **73:** 8349-8355.

99. Theimer, C. A., Wang, Y., Hoffman, D. W., Krisch, H. M., and Giedroc, D. P. (1998) Non-nearest neighbor effects on the thermodynamics of unfolding of a model mRNA pseudoknot, *J Mol Biol* **279:** 545-564.

100. Tan, E., Wilson, T. J., Nahas, M. K., Clegg, R. M., Lilley, D. M., and Ha, T. (2003) A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate, *Proc Natl Acad Sci U S A* **100:** 9308-9313.

101. Foster, M. P., McElroy, C. A., and Amero, C. D. (2007) Solution NMR of large molecules and assemblies, *Biochemistry* **46:** 331-340.

102. Tzakos, A. G., Grace, C. R., Lukavsky, P. J., and Riek, R. (2006) NMR techniques for very large proteins and rnas in solution, *Annu Rev Biophys Biomol Struct* **35:** 319-342.

103. Davis, J. H., Tonelli, M., Scott, L. G., Jaeger, L., Williamson, J. R., and Butcher, S. E. (2005) RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex, *J Mol Biol* **351:** 371-382.

104. Peterson, R. D., Theimer, C. A., Wu, H., and Feigon, J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes, *J Biomol NMR* **28:** 59-67.

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

106. Goddard, T. D., and Kneller, D. G. (2001) *Sparky 3,* University of California, San Francisco.

107. Cromsigt, J., van Buuren, B., Schleucher, J., and Wijmenga, S. (2001) Resonance assignment and structure determination for RNA, *Methods Enzymol* **338:** 371-399.

108. Furtig, B., Richter, C., Bermel, W., and Schwalbe, H. (2004) New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop, *J Biomol NMR* **28:** 69-79.

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

110. Chen, G., and Turner, D. H. (2006) Consecutive GA pairs stabilize medium-size RNA internal loops, *Biochemistry* **45:** 4025-4043.

111. Cojocaru, V., Klement, R., and Jovin, T. M. (2005) Loss of G-A base pairs is insufficient for achieving a large opening of U4 snRNA K-turn motif, *Nucleic Acids Res* **33:** 3435-3446.

112. Nixon, P. L., Cornish, P. V., Suram, S. V., and Giedroc, D. P. (2002) Thermodynamic analysis of conserved loop-stem interactions in P1-P2 frameshifting RNA pseudoknots from plant Luteoviridae, *Biochemistry* **41:** 10665-10674.

113. Cornish, P. V., and Giedroc, D. P. (2006) Pairwise coupling analysis of helical junction hydrogen bonding interactions in luteoviral RNA pseudoknots, *Biochemistry* **45:** 11162-11171.

114. Du, Z., Giedroc, D. P., and Hoffman, D. W. (1996) Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots, *Biochemistry* **35:** 4187-4198.

115. Du, Z., Holland, J. A., Hansen, M. R., Giedroc, D. P., and Hoffman, D. W. (1997) Base-pairings within the RNA pseudoknot associated with the simian retrovirus-1 gag-pro frameshift site, *J Mol Biol* **270:** 464-470.

116. Qiu, H., Kaluarachchi, K., Du, Z., Hoffman, D. W., and Giedroc, D. P. (1996) Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA, *Biochemistry* **35:** 4176-4186.

117. Zust, R., Miller, T. B., Goebel, S. J., Thiel, V., and Masters, P. S. (2008) Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome, *J Virol* **82:** 1214-1228.

118. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, *Biophys J* **78:** 1606-1619.

119. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A., and Rader, S. D. (2006) An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs, *RNA* **12:** 2014-2019.

120. Clegg, R. M. (2002) FRET tells us about proximities, distances, orientations and dynamic properties, *J Biotechnol* **82:** 177-179.

121. Dann, C. E., 3rd, Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., and Winkler, W. C. (2007) Structure and mechanism of a metal-sensing regulatory RNA, *Cell* **130:** 878-892.

122. Csaszar, K., Spackova, N., Stefl, R., Sponer, J., and Leontis, N. B. (2001) Molecular dynamics of the frame-shifting pseudoknot from beet western yellows virus: the role of non-Watson-Crick base-pairing, ordered hydration, cation binding and base mutations on stability and unfolding, *J Mol Biol* **313:** 1073-1091.

123. Pallan, P. S., Marshall, W. S., Harp, J., Jewett, F. C., 3rd, Wawrzak, Z., Brown, B. A., 2nd, Rich, A., and Egli, M. (2005) Crystal structure of a luteoviral RNA pseudoknot and model for a minimal ribosomal frameshifting motif, *Biochemistry* **44:** 11315-11322.

124. Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., Jr., and Varmus, H. E. (1995) Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting, *EMBO J* **14:** 842-852.

125. Shen, L. X., and Tinoco, I., Jr. (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus, *J Mol Biol* **247:** 963-978.

126. Liphardt, J., Napthine, S., Kontos, H., and Brierley, I. (1999) Evidence for an RNA pseudoknot loop-helix interaction essential for efficient -1 ribosomal frameshifting, *J Mol Biol* **288:** 321-335.

127. Napthine, S., Liphardt, J., Bloys, A., Routledge, S., and Brierley, I. (1999) The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting, *J Mol Biol* **288**: 305-320.

128. Ivanov, I. P., and Atkins, J. F. (2007) Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation, *Nucleic Acids Res* **35:** 1842-1858.

129. Wills, N. M., and Atkins, J. F. (2006) The potential role of ribosomal frameshifting in generating aberrant proteins implicated in neurodegenerative diseases, *RNA* **12:** 1149-1153.

130. Wills, N. M., Moore, B., Hammer, A., Gesteland, R. F., and Atkins, J. F. (2006) A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene, *J Biol Chem* **281:** 7082-7088.

131. Shigemoto, K., Brennan, J., Walls, E., Watson, C. J., Stott, D., Rigby, P. W., and Reith, A. D. (2001) Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting, *Nucleic Acids Res* **29:** 4079-4088.

132. Rosenfeld, M. R., Eichen, J. G., Wade, D. F., Posner, J. B., and Dalmau, J. (2001) Molecular and clinical diversity in paraneoplastic immunity to Ma proteins, *Ann Neurol* **50:** 339-348.

133. Dalmau, J., and Bataller, L. (2006) Clinical and immunological diversity of limbic encephalitis: a model for paraneoplastic neurologic disorders, *Hematol Oncol Clin North Am* **20**: 1319-1335.

134. Darnell, R. B., and Posner, J. B. (2003) Paraneoplastic syndromes involving the nervous system, *N Engl J Med* **349:** 1543-1554.

135. Darnell, R. B., and Posner, J. B. (2006) Paraneoplastic syndromes affecting the nervous system, *Semin Oncol* **33:** 270-298.

136. Lilley, D. M. (2000) Structures of helical junctions in nucleic acids, *Q Rev Biophys* **33:** 109-159.

137. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif, *Proc Natl Acad Sci U S A* **98:** 4899-4903.

138. Onoa, B., and Tinoco, I., Jr. (2004) RNA folding and unfolding, *Curr Opin Struct Biol* **14:** 374-379.

139. Gilbert, S. D., and Batey, R. T. (2006) Riboswitches: fold and function, *Chem Biol* **13:** 805-807.

140. Roy, R., Hohng, S., and Ha, T. (2008) A practical guide to single-molecule FRET, *Nat Methods* **5:** 507-516.

141. Pyrc, K., Berkhout, B., and van der Hoek, L. (2007) The novel human coronaviruses NL63 and HKU1, *J Virol* **81:** 3051-3057.

142. Tinoco, I., Jr., Li, P. T., and Bustamante, C. (2006) Determination of thermodynamics and kinetics of RNA reactions by force, *Q Rev Biophys* **39:** 325-360.

143. Green, L., Kim, C. H., Bustamante, C., and Tinoco, I., Jr. (2008) Characterization of the mechanical unfolding of RNA pseudoknots, *J Mol Biol* **375:** 511-528.

VITA

Name:	Suzanne Nicole Stammler
Address:	722 Lazy Lane Bryan, TX 77802
Email Address:	stammler@medicine.tamhsc.edu
Education:	B.S., Chemistry, California State University, Stanislaus, 2000 Ph.D., Chemistry, Texas A&M University, 2009