INSIGHTS INTO SUBGENOMIC RNA SYNTHESIS IN CORONAVIRUSES FROM STRUCTURAL AND BIOPHYSICAL STUDIES

A Dissertation

by

LICHUN LI

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

December 2007

Major Subject: Biochemistry

INSIGHTS INTO SUBGENOMIC RNA SYNTHESIS IN CORONAVIRUSES FROM STRUCTURAL AND BIOPHYSICAL STUDIES

A Dissertation

by

LICHUN LI

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:

Chair of Committee,	David P. Giedroc
Committee Members,	C. Cheng Kao
	J. Martin Scholtz
	Gregory D. Reinhart
Head of Department,	Gregory D. Reinhart

December 2007

Major Subject: Biochemistry

ABSTRACT

Insights into Subgenomic RNA Synthesis in Coronaviruses from Structural and Biophysical Studies. (December 2007)

Lichun Li, B.S., Xiamen University;

M.S., Graduate School of Chinese Academy of Sciences

Chair of Advisory Committee: Dr. David P. Giedroc

The 5’ untranslated region (UTR) of coronaviral genomes contains *cis*-acting sequences necessary for replication, transcription and translation. A consensus secondary structural model of the 5' 140 nucleotides of the 5' UTRs of nine coronaviruses (CoVs) derived from all three major CoV groups is presented and characterized by three major stem loops, SL1, SL2 and SL4. SL2 is conserved in all CoVs, typically containing a pentaloop (C47-U48-U49-G50-U51 in MHV) stacked on a 5-bp stem, with some sequences containing an additional U 3' to U51. NMR structural studies of SL2 hairpin reveal that SL2 adopts a U-turn-like conformation. Parallel molecular genetic experiments reveal that SL2 plays an essential role in sgRNA synthesis as does SL1. We observe strong genetic selection against viruses that contain a deletion of A35, an extrahelical nucleotide that destabilizes SL1, in favor of genomes that contain a diverse panel of destabilizing second-site mutations, due to introduction of a collection of non-canonical base pairs near the deleted A35. Viruses containing destabilizing SL1-ΔA35 mutations also contain one of two specific single nucleotide

mutations in the 3' UTR. Thermal denaturation and imino proton solvent exchange experiments reveal that the lower half of SL1 is unstable and that second-site SL1-ΔA35 substitutions recover one or more features of the wild-type SL1. We propose a "dynamic SL1" model that supports viral replication; these characteristics of SL1 appear to be conserved in other coronaviral genomes.

The coronaviral nucleocapsid (N) protein contains two or more RNA binding domains. We investigated the RNA-binding properties of the N-terminal (NTD) and C-terminal (CTD) domain of MHV N. Our results reveal that the NTD specifically interacts with the TRS-L3 sequence. The role of conserved residues (Y127, Y129 and R110) for this specific interaction were systematically investigated. In contrast to the NTD, the MHV CTD is homodimeric in solution and binds single-strand RNA nonspecifically in a binding mode of the noncooperative large ligand lattice model. The CTD dimer binds with a site size, n=4 nucleotide and the appending of the NTD enhances the single-strand nucleic acid binding affinity.

DEDICATION

I would like to dedicate this work to my husband Hailong and my son Derek. Hailong is so smart, works very hard, and supports me in many ways. I would also like to dedicate this to my parents for good examples and inspiring me to work hard.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. David P. Giedroc for all of his faith in me for advice and support. I learned a lot from him in many areas. I would also like to thank Dr. Leibowitz and his lab members for their great collaboration. I greatly appreciate all the invaluable advice and encouragement from my committee members Dr. Kao, Dr. Scholtz, and Dr. Reinhart. I would like to thank Dr. Xiangming Kong for helping me with many NMR experiments. I thank all past and present members in the Giedroc lab. They helped me in many ways, and I have learned a great deal from them, especially Suzanne. I would also like to thank my classmates, particularly Cuijuan Tie, Chonghua Li, Hongjun Jin and my friend Feng Wang for their support and patience.

TABLE OF CONTENTS

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Ribonucleic acid (RNA) is no longer viewed, as embodied in the central dogma, as a passive molecule that enables the decoding of the genetic information in DNA into proteins. Indeed, recent genetic, biochemical and biophysical studies of RNA have led to an appreciation of RNAs as central regulatory molecules and catalytic entities within the cell. For example, RNA motifs embedded in larger RNAs can drive the regulation of gene expression at the transcriptional level as metabolite-sensing riboswitches (Mandal and Breaker, 2004), or at the translational level in formation "recoding" signals within protein-coding mRNAs (Grentzmann et al., 1998), and in internal ribosomal entry sites in RNA viruses that function in ribosome loading in translation initiation (Lukavsky et al., 2003). Other virally-encoded RNA motifs control genome replication and transcription and/or the switch between the two during RNA virus infection (Brian and Baric, 2005). Many naturally occurring RNAs function as true catalysts, coined ribozymes, that make and break covalent bonds during the process of mRNA maturation (splicing), tRNA processing, protein synthesis, and metabolic control (the glmS riboswitch is a ribozyme, Wakeman et al., 2007). Indeed, non-naturally occurring catalytic RNAs that catalyze a variety of chemical transformations have now been characterized, including the Diels-Alderase and an RNA ligase, for which high

---

This dissertation follows the style of *Embo Journal.*

resolution structures are now known (Scott, 2007). Finally, the seminal discovery of small noncoding RNAs, including microRNAs and small interfering RNAs that function to down-regulate gene expression in bacteria, plants and animals by targeting mRNAs for destruction, defines an entirely new regulatory network within the cell (Bartel, 2004). A detailed understanding of the structure and dynamics of RNA is required to fully understand the scope and mechanism of biological activities that are, or might be, carried out by this intriguing macromolecule.

RNA folding is hierarchical

RNA is transcribed in cells as single strands of nucleic acid. However, these linear biopolymers are not simply long strands of nucleotides. To be functional, RNA molecules undergo a hierarchical folding sequence, in which the secondary structure forms first from the primary sequence, followed by the formation of the complex tertiary structures characterized by the packing of helical secondary structural elements interacting with single-stranded regions. The basic secondary structural elements in RNA include helices, hairpin loops, internal and asymmetric loops, bulges and junctions. The regular A-form double-stranded helices formed by Watson-crick (WC) base pairs are the most dominant structural element in RNA (Figure 1-1A). Since the discriminatory edges of the bases are buried in the major groove, regular A-form RNA helices are not typically involved in the specific interactions with ligands. In contrast, the bases in loops, bulges and junctions are most often non-canonically base paired, and are involved in a variety of interactions. Those directed and ordered arrays of non-WC base

pairs induce distinctive folds into the phosphodiester backbones of the interacting RNA strands are defined as "RNA motifs" (Leontis and Westhof, 2003). Tertiary interactions are generally mediated by these RNA motifs. The availability of numerous diverse, RNA structure, such as tRNA, the 160 nucleotides P4-P6 domain of the group I intron (Cate et al., 1996), and especially the ribosomal subunits (Ban et al., 2000; Ramakrishnan, 2002; Wimberly et al., 2000) have greatly expanded our understanding, and have made possible the identification and classification of an increasing number of RNA motifs. The Structure Classification of RNA (SCOR, http://scor.lbl.gov/scor.html) database is a comprehensive resource of RNA motifs that classify and characterize RNA secondary and tertiary structural motifs assisted by literature descriptions. Comprehensive reviews about the RNA structural motifs are also available (Leontis and Westhof, 2003; Moore, 1999). Here, the structure of the specificity (S) domain of B-type ribonuclease P is used as an example to illustrate RNA motifs, particularly tertiary structural interactions involved in the complex structure (Figure 1-1B).

The model B-type Ribonuclease (RNase) P, from B. *subtilis*, consists of specificity domain (S domain, from nucleotides 86 to 239) and the catalytic domain (C domain). Inspection of the crystal structure of the S domain reveals that the structure is built of relatively rigid structural elements and stabilized by several common RNA motifs, including coaxial stacking of helices, an A-minor motif, a tetraloop-receptor motif, and cross-strand purine stacks.

The central element of the S domain is a rigid core (shaded red in Figure 1-1B) composed of continuous stacking of helical stems P10, P11 and the basal portion of

P10.1. P9 also coaxially stacks on P8 to form another continuous A-form helices. Coaxial stacking is a common RNA motif in which two separate helices stack to form coaxial helices as a pseudo-continuous helix. Coaxial stacking is a highly stabilizing tertiary interaction first identified in tRNA, and is common in rRNA (Noller, 2005). This core is stabilized by an A-minor motif in which the A177 and A178 on the loop of P10.1 enters the minor groove of the P7/P10 stack in a region including conserved base pairs G90-C235 and G132-C234 and form H-bonds between the 2' OH groups of the loop nucleotides with the 2' OH groups of the helical residues (Figure 1-1C). A-minor motifs appear widely as stabilizing contacts between RNA helices, interactions between loops and helices, and the conformations of junctions and tight turns (Nissen et al., 2001). A-minor motif is also found in many frameshifting RNAs (Cornish et al., 2005).

P12 interacts with P10.1 via a classical GAAA tetraloop (TL) -teraloop receptor (TLR) interaction, similar to that found in P4-P6 group I intron (Cate et al., 1996), but with an AC platform rather than AA found here instead (see Figure 1-1D). This 11 nucleotide tetraloop recepter motif is characterized by a consistent set of H-bonds between the tetraloop and the internal loop of the TLR: between the first A of the TL and a U·A base pair of the receptor to form a U·A·A base triple; between the second A of the TL and the backbone of the TLR C and U; between the third A of the TL and a C·G base pair of the TLR. The central part of P10.1 is stabilized by four non-WC base pairs of a cross-strand adenosine-stack and interacts with J11/12-J12/11 with a bulged G motif, in which the G168 stacks on A194 of J11/12-J12/11 (see Figure 1-1E).

In addition to the RNA motifs that have been previously found in other systems, the S-domain also contains a novel RNA element, termed base interchalation, again, involving unpaired adenosines. In this RNA element, the backbone around A230, an adenosine extruded from the P11 helix, turns nearly 180^{o}; as a result, A230 faces P9 and stacks on A130 which is bulged out of P9 helix (Figure 1-1F). This arrangement leads to the A230 and A130 being exposing to the solvent, and directly interacting with the pre-tRNA substrates.

The structure of B-type RNaseP S domain and other large RNA molecules indicates that simple, sometimes recurring RNA structural motifs are the basic building blocks of complex RNA tertiary structure. An increasing number of crystal and NMR structures of RNA and RNA-protein complexes are continuous to reveal many new structural elements. For example, the crystal structure of s2m from the 3' UTR of SARS-CoV reveals two novel RNA elements, termed a nucleotide quarter kink and a three-purine bulge (Robertson et al., 2005). The characterization of these new RNA motifs will not only help us to understand RNA tertiary structure but also its function.

In addition to the RNA structural motifs, metal ions are essential for RNA folding (for reviews, see DeRose, 2003; Draper, 2004; Sigel and Pyle, 2007; Woodson, 2005). The polyanion nature of RNA requires metal ions to acquire its functional three-dimensional structure. While the majority of metal ions are delocalized and non-specifically interact with RNA, some ions can specific bind to RNA through direct inner-sphere coordination and/or indirect outer-sphere coordination through water molecules (Draper, 2004). Specific pocket of the binding of divalent meal ions are found

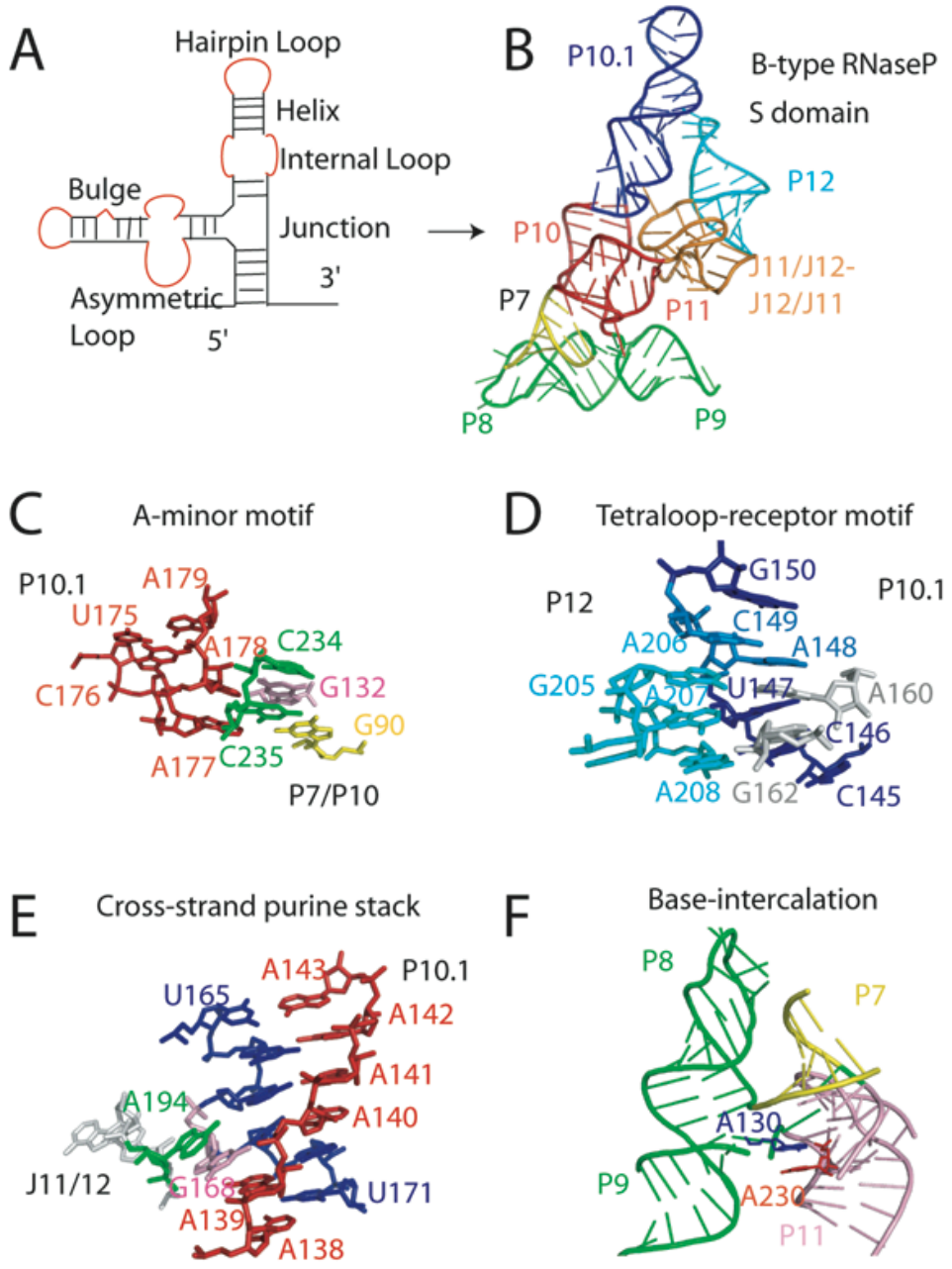


Figure 1-1 RNA structure and RNA motifs. Secondary structural elements in (A). Ribbon diagram of tertiary structure of B-type RNaseP S domain (PDB code: 1NBS) in (B). Sticks and ribbon diagrams of the A-minor motif (C), tetraloop receptor motif (D), cross-strand purine stacks motif (E) and base intercalation motif (F).

in a number of RNA structures (see review, Hermann and Patel, 1999). Interestingly, A recent study on a metal-sensor riboswitch has shown that Mg^{2+} is involved in genetic control via mediating multiple long-range interactions to stabilize a compact three-helical tertiary structure (Dann III et al., 2007).

RNA conformational dynamics (conformational rearrangements)

Although RNA can fold into complicated structures, static structures are not enough to account for the functional diversity of RNA. Rather, a number of RNA's functional diversity comes from the dramatic conformational changes that can be either self-induced or triggered by binding to cofactors, including metabolites, RNAs, proteins, sugars, lipids, metals and other molecules (Draper, 2004; Micura and Hobartner, 2003; Williamson, 2000).

Conformational dynamics is widely implicated in catalytic ribozymes (AI-Hashimi, 2005; Dunham et al., 2003; Williamson, 2000). Ribozymes exhibit a variety of complexities from the extreme spliceosome to the simple haipin ribozyme. As one extreme, the large molecule spliceosome is a dynamic ribonucleoprotein that consists of five small nuclear ribonucleoprotein particles (snRNPs) as well as numerous proteins, and mediates the excision of pre-mRNA introns and the ligation of exons to generate the mature mRNA. The assembly, rearrangement, disassembly and recycling of the snRNPs all require highly ordered RNA-RNA and RNA-protein rearrangements (Brow, 2002). Recent work on stem I of the U2 snRNA suggests that U2 adopt a number of mutually exclusive folded conformations during spliceosome assembly and catalysis (Sashital et

al., 2007), and supports the proposed model that the active-site conformations compete kinetically and that destablilization of one conformation drives the complex catalysis toward the other state (Query and Konarska, 2004). As the other extreme, one of the simplest ribozymes, the hairpin ribozyme has been shown to have complex conformational dynamics, in which the conformation starts with an extended "undocked" RNA fold, which is then triggered by substrate binding to undergo conversion into the catalytically active "docked" conformation. Following the cleavage, the RNA returns to an undocked conformation as the products are released (Zhuang et al., 2002).

A common RNA conformational rearrangement involves global reorientation of the helical domains mediated by linker motifs, e.g. junctions and bulges, comprising of the recognition site. The interruption of A-form helices caused by the junction and bulge normally results in high negative charge density at these sites. Therefore, as a result of these interruptions, free RNA molecule often avoids adopting the coaxial arrangement of helical domains. However, the binding of a basic protein or Mg^{2+} at these sites favors the coaxial conformation and leads to an RNA conformation change. For example, the S15 binds to the three-way junction of 16S ribosomal RNA and results in a coaxial alignment of the domain, which in turn, initiates the subsequent assembly of the central domain in the 30S ribosomal subunit (Orr et al., 1998). Mg^{2+} binding leads to similar conformational change.

A riboswitch involves a conformational rearrangement of a region of untranslated RNA that is induced by a small metabolite, and leads to the regulation of

gene expression (Mandal and Breaker, 2004). Riboswitchs are comprised of a ligand-binding aptamer domain and an expression platform whose conformation affects gene expression. The structural studies of purine riboswitches reveal that the ligand binding induces a folding of an unstructured aptamer domain, which in turn results in a complex tertiary RNA structure where the ligand is "engulfed" in aptamer domain. As a result of this conformational change, the expression platform also changes its structure, thereby regulating gene expression (Figure 1-2B) (Batey et al., 2004; Serganov et al., 2004).

Changes in conformational dynamics of RNA occur not only at the tertiary structural level, but also at the secondary structural level. For example, it has shown that an artificial RNA sequence can adopt two distinct ribozyme conformations with distinct catalytic activities (Figure 1-2C). One conformation is the hepatitis delta virus (HDV) ribozyme that carries out a site-specific phosperdiester bond cleavage reaction, while the other is a class III RNA ligase ribozyme that performs RNA ligation. These two conformations do not share even a single base pair in common, and minor variants of this sequence are highly active for one or the other activity, where each can be accessed from prototype ribozymes through a few nucleotide mutations (Micura and Hobartner, 2003; Schultes and Bartel, 2000). Very small RNAs (only 18-21 nucleotides) have also been shown to fold into more than one conformation (Micura and Hobartner, 2003).

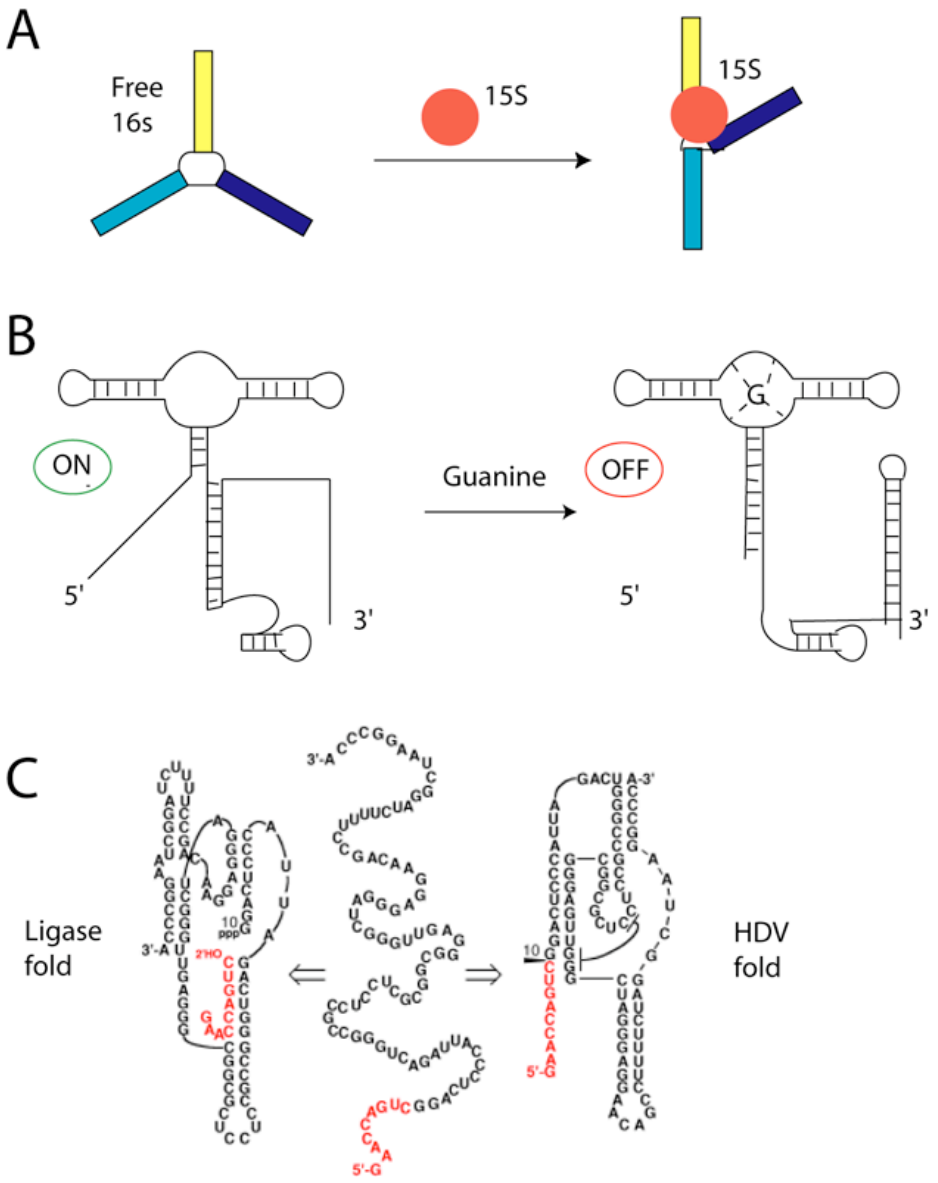


Figure 1-2 Cartoon diagrams of examples of conformational rearrangements of RNA molecules. (A) The orientation of the three helical domains of a 16S upon S15 binding. (B) A representative conformational rearrangement of the guanine mRNA riboswitch in gene regulation. (C) The artificial RNA sequence can adopts two different secondary structure with different ribozyme activity, adapted from (Micura and Hobartner, 2003).

Coronaviruses

The years 2002 and 2003 witnessed global outbreaks of severe acute respiratory syndrome (SARS), a then newly emerging life-threatening atypical pneumonia (Peiris et al., 2004). SARS-CoV, the etiological agent of SARS, received immediate attention because of its rapid spread, high infectivity, and significant mortality. According to WHO (http://www.who.int/csr/sars/country/en/), within months of the initial incidence, this disease had spread to 29 countries or regions and infected more than 8000 people, ultimately causing 774 deaths. Unlike previously characterized coronaviruses which were known to cause relatively mild infection in only one host species, SARS-CoV was found to be capable of crossing species lines as a result of a deletion, mutation, and recombination with previously existing coronaviruses (Li et al., 2005). The tragic ourtbreak and subsequent identification of SARS-CoV has stimulated the interest of the scientific community in coronavirus biology and mechanisms of replication in an effort to counteract even more life-threatening coronavirus infections that may ultimately emerge.

Coronaviruses are enveloped human and veterinary pathogens that can cause enteric and respiratory infections, and in some cases neurologic illness or hepatitis (Lai and Holmes, 2001). Coronaviruses and toroviruses are classified in *Coronaviridae* family, and together with *Arteriviridae* and *Roniviridae* families are in the order *Nidovirales*, which are characterized by a nested set arrangement of their subgenomic mRNAs (Gonzalez et al., 2003) (Figure 1-3). Coronaviruses are single-stranded positive-sense RNA viruses and have the largest genomes of any RNA virus, ranging from 27 kb

for Human Coronavirus (HCoV) strain 229E to 32 kb for the murine hepatitis virus (MHV), the subject of this study. The coronaviruses have been classified into three groups, denoted group 1, group 2, and group 3, based on serologic relatedness, genetic organization, and sequence similarity. Group 1 and 2 viruses are mammalian viruses, while group 3 viruses only infect avian hosts (Masters, 2006; Rota et al., 2003). SARS-CoV was first assigned as a new serological group of coronaviruses, but subsequent phylogenetic analysis of Orf1b which encodes the replicase polyprotein supports the contention that SARS-CoV is a group 2 coronavirus (Snijder et al., 2003) (Figure 1-3). After the discovery of SARS-CoV, two new human coronaviruses have also been identified: HCoV-NL63(van der Hoek et al., 2006), a group 1 virus and HCoV-HKU1 (Woo et al., 2005), a group 2 virus (for a review, see Masters, 2006).

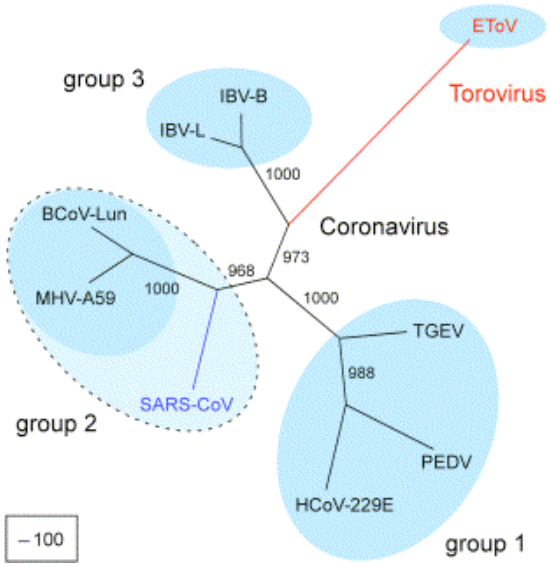


Figure 1-3 Phylogenetic analysis of coronaviruses replicase genes based on global sequence similarity of ORF1b. Adapted from (Snijder et al., 2003). Note that MHV-A59 is more closely relative to SARS-CoV.

Coronavirus morphology

Coronaviurses are named for their crown-like appearance in the electron micrographs. The virions are roughly spherical, with an approximate diameter of 80-120 nm. There is an outer envelope bearing distinctive, "club-shaped" peplomers (~20 nm long x 10 nm wide at the virus distal end). Three of four structural proteins encoded by all coronaviruses are embedded in the virion envelope, and include spike (S), envelope (E) and membrane (M) proteins. The internal component of the virion is a helically symmetric nucleocapsid although even moderate resolution images of this part are not yet available (Masters, 2006). The fourth structural protein, nucleocapsid protein (N), is the major virally encoded protein and is responsible for wrapping the positive-strand RNA genome into a helical nucleocapsid (see Figure 1-4).

The S protein (also previously denoted E2) is the most prominent structural protein in the envelope. It mediates receptor attachment and the fusion of viral and host cell membrane. The majority of S protein is composed of a highly glycosylated ectodomain, and the remaining small carboxy-terminal region composed of a transmembrane domain and an endodomain that assembles into a trimer. The S protein is processed by a trypsin-like host protease into two polypeptides, S1 and S2. It is thought that the S1 domains of a S protein oligomer form the bulb portion of the spike.

Both M and E are integral membrane proteins, and are essential for the maturation of newly formed virion particles. The E protein (also denoted sM) is a minor component of virion particle. Although the primary sequences of E from all three groups are divergent, they all share a common architecture: a short hydrophilic amino terminus,

a large hydrophobic region containing two to four cysteines and a long hydrophilic carboxy-terminal tail. The M glycoprotein (previously denoted E1) is responsible for the characteristic shape of the virion envelope. M protein is composed with a short ectodomain, followed by three transmembrane domains and a large endodomain. The M protein is the central organizer of assembly, because it not only self-associates, but also interacts with S for virion incorporation, as well as with the genomic RNA-bound form of N through an interaction between the M endodomain and the N C-terminus (Hurst et al., 2005).

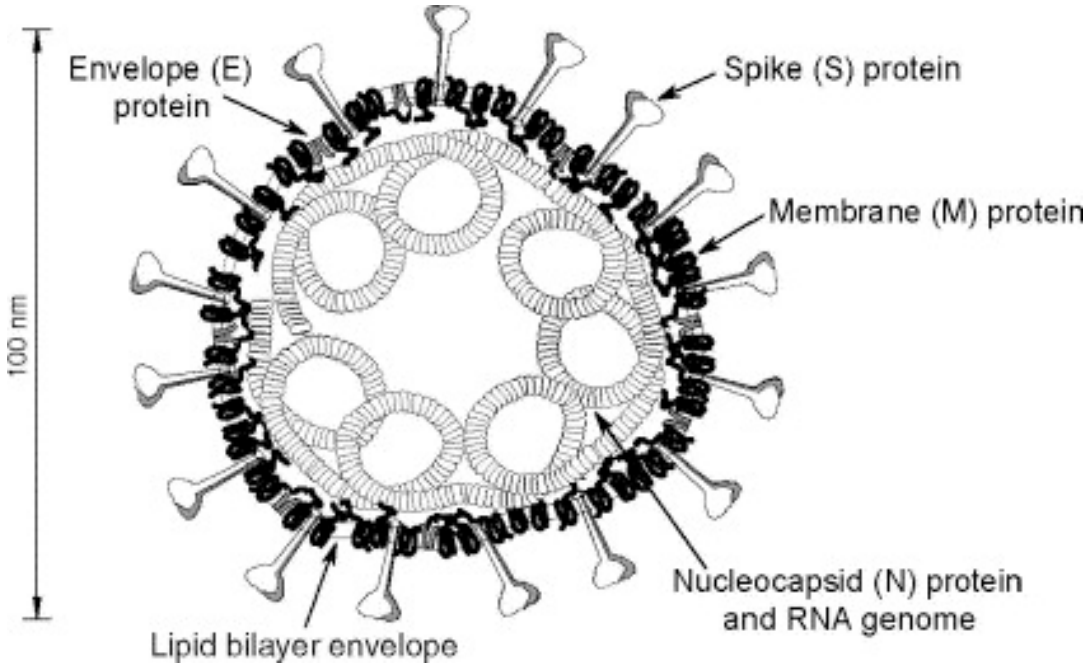


Figure 1-4 Schematic illustration of coronavirus virion particle. The minimal set of structural proteins are shown. Adapted from (Masters, 2006).

Genome structure of coronaviruses

The single-stranded positive-sense coronaviral genome contains both 5' and 3' untranslated regions (UTR) with a 5'-terminal cap and a 3'-terminal polyA tail. The 5' and 3' UTRs are ~200 to 500 nucleotides and contain *cis*-acting sequences and structural elements that play important roles in RNA replication and transcription (to be discussed in detail). The genomes of coronaviruses have a characteristic organization. As shown in Figure 1-5, the 5'-proximal two-thirds of the genome encodes the gene1 that gives rise to many of the nonstructural proteins (nsps) of the virus. In the SARS-CoV, there are 16 nsps (Figure 1-5). Gene 1 is composed with two large open reading frames, orf1a and orf1b, with a short region of overlap between orf1a and orf1b, mediating a -1 ribosomal frame shift that extends the orf1a polyprotein (denoted pp1a) to the orf1b region to produce a ~740 kDa polyprotein (denoted pp1ab). The efficiency of framshifting can approach 50%. A set of cleavage products derived from the pp1a and pp1ab precursors by the papain-like proteinase (PL^{pro}) or the 3C-like cysteine proteinase ($3CL^{pro}$) encoded in orf1a collectively function in RNA replication and transcription as a multiprotein replicase complex (Figure 1-5) (de Haan and Rottier, 2005; Lai and Holmes, 2001). Following orf1b are the four genes that are common to all coronaviruses that encode the structural proteins in the order 5'->3': S-E-M-N. The genomes of known coronaviruses contain a variable number of unique virus- or group-specific open reading frames (orfs) encoding nonstructural proteins, known as accessory proteins in this region of the genome. Those genes are scattered among the structural proteins, for example, in SARS-CoV (Figure 1-5), they are located between the S and E genes (orf3a, 3b in SARS-CoV),

between the M and N genes (orf6, 7a, 7b, 8a, 8b), or within the N gene itself (orf9b) (Enjuanes, 2005). Unlike gene 1 encoded polyproteins which are directly translated from the "positive" sense RNA genome, these structural and accessory proteins are translated from a nested set of 3' and 5' co-terminal subgenomic (sg) mRNAs.

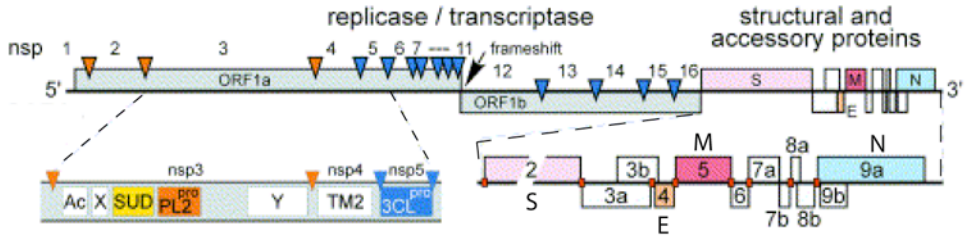


Figure 1-5 Overview of the SARS-CoV genome organization and expression. Adapted from (Snijder et al., 2003).

Coronavirus replication and transcription

As in other positive-sense RNA viruses, coronavirus replication requires a negative-sense RNA intermediate. Once the RNA genome is uncoated in the cytoplasm of the infected host cell, the positive-sense RNA functions as an mRNA to mediate the translation of the replicase polyproteins, pp1a and pp1ab by the host ribosome. The pp1a and pp1ab replicase polyproteins are then processed by the virally encoded proteinases located in pp1a and the processed products will assemble the viral replication machinery; this complex then synthesizes a full-length negative-sense or anti-genomic RNA. This negative-sense RNA subsequently functions as a template to produce full-length positive-sense genomic RNA in RNA replication (Figure 1-6). Meanwhile, a 3' co-terminal nested set of sgRNAs are generated through a discontinuous transcription

strategy and used as templates to translated the structural and other accessory proteins (Pasternak et al., 2006). During this replication cycle, RNA synthesis is highly asymmetric, with the ratio of positive-sense RNA to full-length negative-sense RNA ≈ 99:1. The ratio of synthesis of genomic individual to sgRNAs relative to one another appears to be also constant (Sawicki and Sawicki, 2005; Sawicki and Sawicki, 1990).

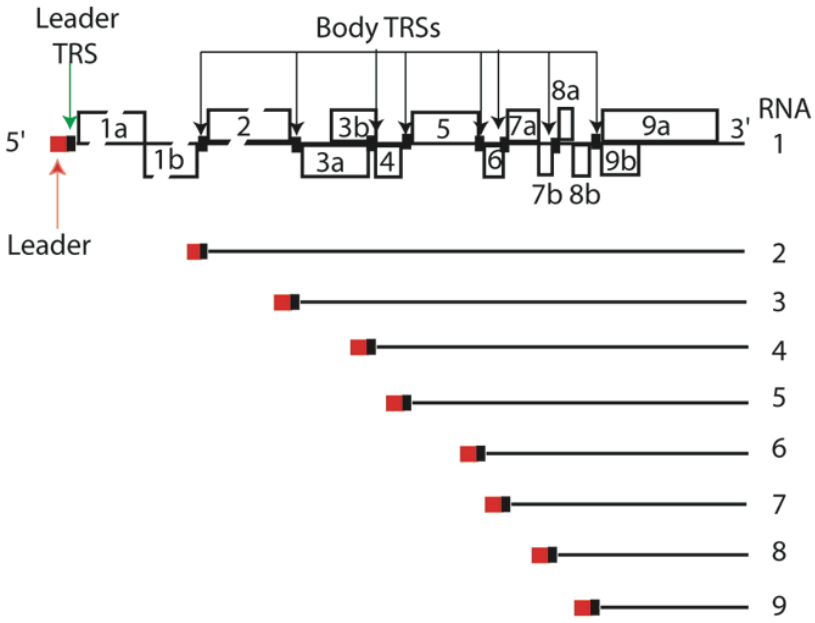


Figure 1-6 Organization of SARS-CoV genome and subgenome with leader and body TRSs indicated by small boxes.

The 5' UTR of the sgRNA is identical to the extreme 5' end of the 5' UTR of the genome (Lai et al., 1982). The highly conserved transcription-regulating sequences (TRSs) found in the leader and in the genome play important roles in sgRNA synthesis. TRSs are AU-rich elements that are located at the 3'end of the 5' leader sequence

(termed “TRS-L”), and upstream of each “body” gene segments (termed “TRS-B”) in the genomic RNA (Figure 1-6). This sequence defines the leader-to-body fusion site in the sgRNAs (Brian and Baric, 2005).

Different models have been proposed to explain the fusion of the 5’-leader to the body segments in sg mRNAs synthesis (Lai and Holmes, 2001; Sawicki and Sawicki, 2005). Although all replication models assume the fusion of the leader and body, the main disagreement is whether the discontinuous transcription step occurs during positive- or negative-strand synthesis. The commonly accepted model is the “discontinuous minus-strand RNA synthesis model” (Enjuanes et al., 2006; Sawicki and Sawicki, 1998), which proposes that subgenomic negative-sense RNAs are directly synthesized from the genome-sized RNA template in a discontinuous fashion. Here, the 3’ anti-body TRS-B of the nascent subgenomic RNA will be redirected to base pair with 5’ leader TRS-L sequence and extended through the anti-leader sequence, from which sg mRNAs are subsequently synthesized (Figure 1-7). This base-pairing requirement during transcription has been confirmed in the arterivirus equine arteritis virus (EAV) (Pasternak et al., 2001; van Marle et al., 1999) and in the coronaviruse avian transmissible gastroenteritis virus (TGEV) by mutagenesis of a full-length cDNA (Zuniga et al., 2004). It has been shown that the disruption of base-pairing dramatically influences the transcription efficiency and that compensatory mutations recover transcription. In addition, the “minimal free energy” of the corresponding putative TRS-L-TRS-B duplex also appears to contribute to the relative abundance of subgenomic mRNAs with sequences immediately flanking the core TRS-B playing an important role

in this process (Zuniga et al., 2004). Based on these results, Zuniga *et al.* have refined this model suggesting RNA-RNA interactions between the 5' and 3' UTRs, or proteins bound to these UTRs. In this refined three step model (Figure 1-7), the first step is the formation of 5'-3' circularized complex between leader TRS and 3'-end of the genome via viral and cellular proteins, which brings the 5'- and 3' -ends physically close; the second step is base-pairing scanning of the nascent negative RNA strand with the positive sense TRS-L; the last step is the template switching during synthesis of the negative strand to complete the negative sgRNA synthesis (Zuniga et al., 2004; Zuniga et al., 2007).

It is noted that in this model, the synthsis of full-length negative-sense genomic and sgRNAs occurs from the same full-length genomic RNA molecule. It is currently unknown what mediates switching between RNA transcription (sgRNA synthesis) via template jumping from individual TRS-B sequences and RNA replication in which full-length anti-genome is synthesized. It is clear however, in the work presented here, that these two processes have distinct structural requirements. Host-encoded or virally encoded proteins and RNA sequences with the 3' UTR and /or within the N gene are likely to be involved in this process (Brian and Baric, 2005).

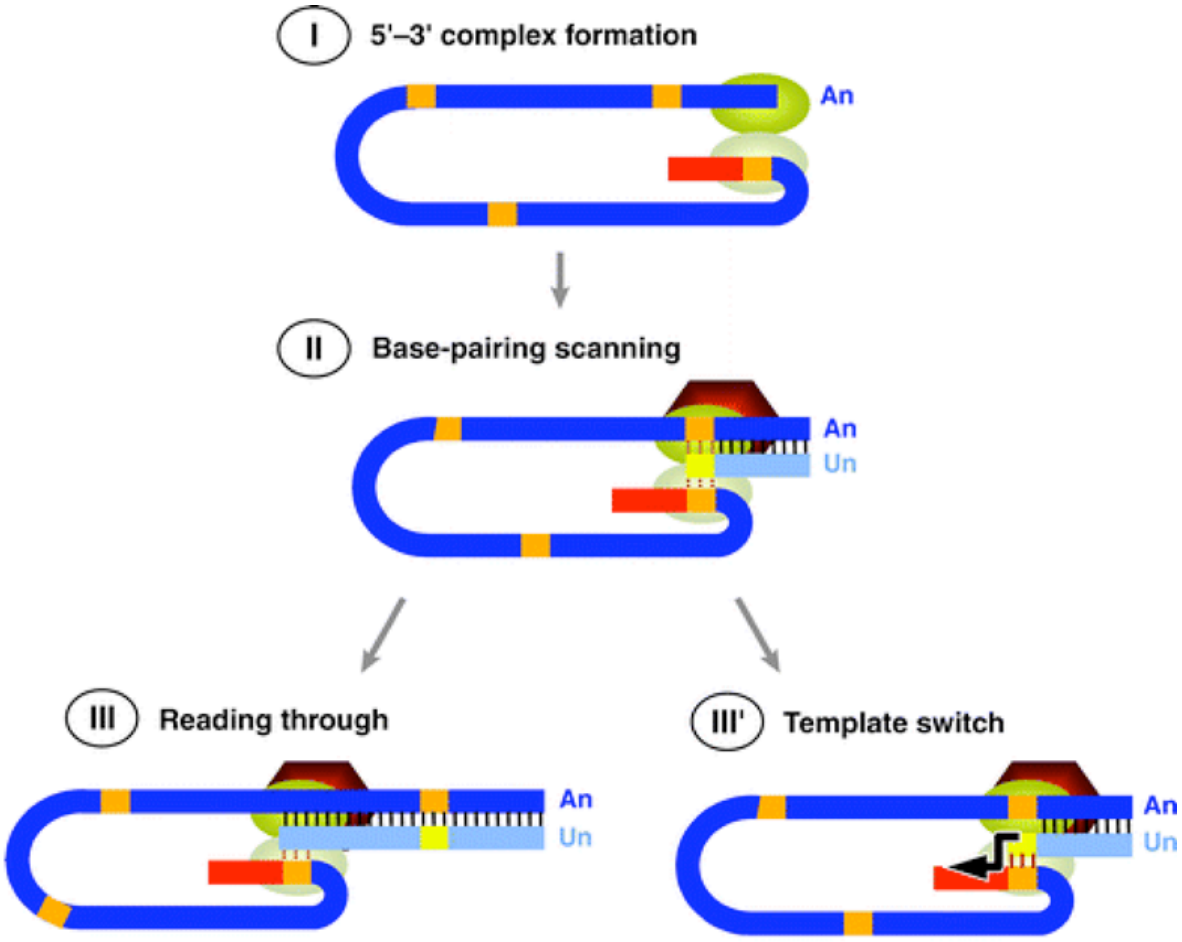


Figure 1-7 Three-step model of CoV discontinous transcription (Enjuanes et al., 2006). Details can be found in the text.

5' UTR of coronaviruses

The 5' UTR of the coronaviral genome ranges in size from ~200 to ~ 500 nucleotides. The 5' UTR possesses a 5' terminal methylated cap and the 5' "leader" sequence discussed above. The 5' cap structure suggests that the small ribosomal subunit locks onto the cap and initiates a scanning mechanism to ultimately assemble the

ribosomes at the orf1a initiation codon. There is some evidences in support of the contention that the 5' UTR and noncoding intergenic regions between the TRS-B and initiation codon of ORFs destined for subgenomic RNAs influence the translation efficiency in BCoV and MHV (Senanayake and Brian, 1999; Tahara et al., 1994). This translation initiation mechanism is distinguished from that employed by picornaviruses and flaviviruses, e.g., Hepatitis C virus, which initiate translation via an internal ribosome entry sites (IRESs), a RNA structure that directly recruits 40S ribosomal subunits in a cap-independent way (Fraser and Doudna, 2007; Vagner et al., 2001).

Evidence that 5'-UTR may also form a certain secondary or tetiary structure to regulate the translation and replication is not yet conclusive. The 5' UTRs in other viral systems have been shown to be highly structured. This has been well studied in the 5' UTR of tombusviruses within the context of a prototypical defective interfering (DI) RNA (Ray et al., 2003; Wu et al., 2001). Two major functional domains, the T-shaped domain (TSD) and downstream domain (DSD) have been identified. TSD is composed of three helices that form a three-way junction, and DSD forms a complex structure composed of a long helix, bulge and a large internal loop. These two domains are separated by an intervening hairpin structure, SL5. Interestingly, the TSD can interact with the 3' end of DSD through the formation of a pseudonknot, PK-TD1 (Ray et al., 2003). The TSD, DSD, SL5 and PK-TD1 have all been shown to contribute to efficient DI RNA replication (White and Nagy, 2004). Another example is the 5' UTR of piconaviruses. In piconaviruses, the 5' UTR contains the IRES structure. In addition, a cloverleaf structure in 5'UTR of enteroviruses and rhinoviruses have been shown to

serve as a binding site for viral protein 3CD and host protein PCBP and are involved in the regulation of RNA replication (Bedard and Semler, 2004). Interestingly, this interaction between the cloverleaf structure RNA and proteins are suggested to play a role in mediating an interaction between the 5' and 3' ends of the viral RNA.

The 5' UTR of positive-strand RNA virus genomes contains *cis*-acting elements to function in genome replication (Raman et al., 2003; Raman and Brian, 2005). To address whether the *cis*-acting elements in 5' UTR is sequence specific, or requires a higher-order RNA structure, some effort has gone into obtaining the secondary structure of the 5' UTR in group 2 coronaviruses BCoV. Four stem-loops I-IV have been predicted in 210 nucleotides 5' UTR of BCoV (Raman et al., 2003; Raman and Brian, 2005). These predicted stem loops are supported in part by enzyme mapping experiments and DI RNA replication assays. In this predicted secondary structure model, the stem-loops I and II are located within the 5' 84 nucleotides (Figure 1-8). This sequence has been shown to be required for DI RNA replication (Chang et al., 1994). However, the predicted stem-loop I and II are not conserved in other group 2 coronaviruses and the stem-loop I is predicted to be a very unstable structure due to the large loop and a number of bulges in the stem. In this model, the TRS-L is predicted to be positioned in the hairpin loop of stem-loop II (Figure 1-8). Similarly, in arterivirus EAV, the TRS-L is also predicted in a hairpin loop structure called "leader TRS hairpin" (LTH) (Van Den Born et al., 2004). Stem-loop III and IV, on the other hand, are phylogenetically conserved and well supported by nuclease mapping and DI replication results (Raman et al., 2003; Raman and Brian, 2005). Stem-loop IV has previously been

shown to be involved in the initiation of genome-length minus-strand synthesis (Raman and Brian, 2005).

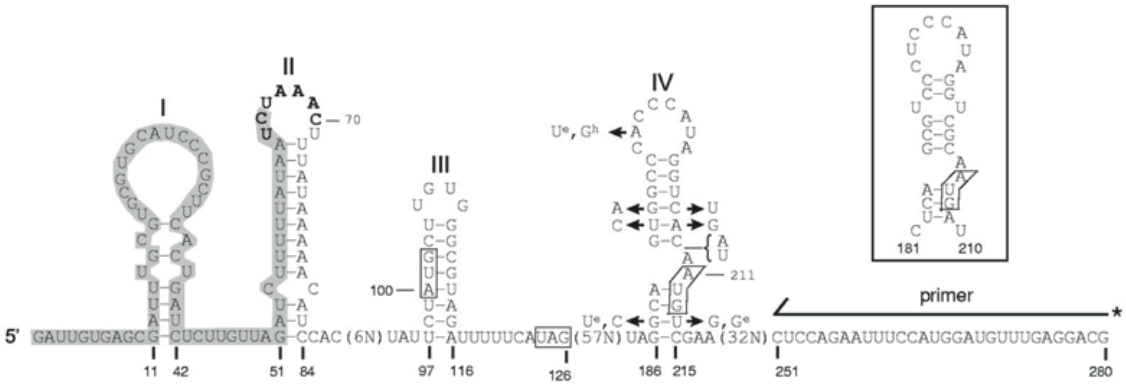


Figure 1-8 Predicted secondary structures within the 210-nt 5' UTR of the BCoV genome. The 65-nt BCoV leader sequence is shaded, and the TRS UCUAAAC intergenic sequence (nucleotides 64 to 70) is in bold. Stem-loops I, II, III and IV are identified. Figure is adapted from (Raman and Brian, 2005).

A more interesting question is whether the conformation of the 5' UTR is the same or different when participating in RNA replication, transcription or translation. If so, then how? Since the 5' UTR is suggested to be involved in regulating various stages of the viral replication cycle (Figure 1-6), and RNA conformational change between mutually exclusive conformations is not unprecedented, it is not difficult to envisage a role for conformational remodeling of the 5' UTR with or without the aid of virally encoded or cellular proteins during the viral life cycle. In coronaviruses, the 5' UTR may function by recruiting protein(s) to mediate circularization of genome via direct interaction with the 3' UTR, and/or undergo a conformational change triggered by cellular or virally encoded protein(s) or small molecule ligand to make sequence

specifically required for transcription, translation and/or replication accessible. Host and virally encoded proteins shown to bind 5' UTR or the complementary negative-sense strand include nucleocapsid protein, which is projected to specifically bind to the TRS in the MHV leader sequence (Nelson et al., 2000), the cellular polypyrimidine tract binding protein (PTB) also shown to bind to the leader sequence in MHV (Li et al., 1999), and heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) which binds the complementary negative-sense leader region to where PTB is known to bind (Li et al., 1997). However, our understanding of the structure of the 5'UTR of coronaviruses is still limited. More careful and accurate secondary structure prediction and higher-order structural studies of 5'UTR are needed (see Chapter II). In addition, the identification of the roles of 5'UTR-binding proteins will contribute to understanding the mechanism of viral transcription and replication.

Such a conformational rearrangement of the 5' UTR has been shown to occur in other viral systems. For example, the 5' UTR leader in the retrovirus HIV-1 is proposed to undergo an ordered structural rearrangement from an initial thermodynamically stable rodlike fold required for translation to a branched structure containing multiple hairpins which is involved in genome dimerization, packaging and viron assembly (Figure 1-9) (Huthoff and Berkhout, 2002). The autocomplementary loop in the dimerization initiation site (DIS) and the virally encoded nucleocapsid protein, a documented RNA chaperone plays a key role in this conformational rearrangment (Cruceanu et al., 2006).

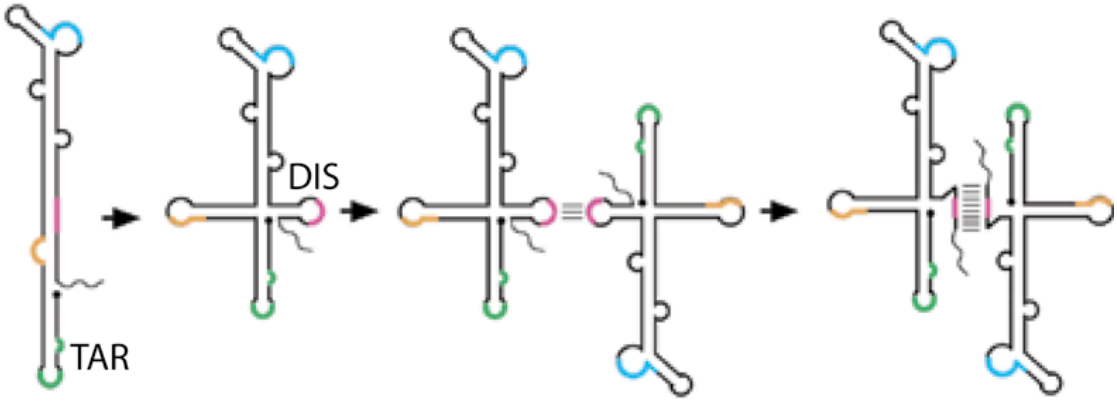


Figure 1-9 Schematic representation of a secondary structural conformational rearrangment of the HIV-1 leader RNA. Adapted from (Huthoff and Berkhout, 2002).

3' UTR of coronaviruses

The 3' UTRs of coronaviruses range in size from ~300 to ~500 nt, of which all possess a 3'-terminal poly (A) tail and an octameric sequence, ^{80}GGAAGAGC^{73} corresponding the residues 73 to 80 nucleotides upstream of the poly (A) tail (the 3' nucleotide is labeled 1) (Brian and Baric, 2005).

Coronaviruses contain 3'-terminal poly(A) tails similar in length to cellular poly(A) tails. This suggests that the 3' UTR poly(A) tail may stimulate translation by communicating with the 5' UTR through host encoded proteins, for example, poly(A)-binding protein (PABP). The interaction of PABP with poly(A) tail of DI RNAs is essential for coronavirus DI replication (Spagnolo and Hogue, 2000). PABP is involved in the circularization of the 5' and 3' ends of the mRNA during translation in both cap-dependent systems and cap-independent systems, e.g. caliciviruses (Herbert et al., 1997) and picornaviruses (Paulous et al., 2003). This circularization may enhance translation

efficiency by stabilizing the mRNA and the translation complex and/or increasing the efficiency of the recycling of ribosomes. PABP-mediated genome circularization may also play a role in regulating RNA synthesis (Shi and Lai, 2005) .

The 3' UTR of MHV and SARS-CoV (Goebel et al., 2004b) and of MHV and BCoV (Hsue and Masters, 1997) are functionally interchangeable suggesting that conserved functionally important structures are likely present in all of group 2 coronaviruses. *Cis*-acting elements and higher-order structural elements in the 3' UTR have been mapped in MHV and BCoV. The *cis*-acting elements of MHV are suggested to encompass the entire 301-nt 3' UTR (Kim et al., 1993; Lin and Lai, 1993; Luytjes et al., 1996). In MHV, the Masters group has identified a highly conserved 68 nucleotide bulged stem-loop structure immediately downstream of the N gene termination codon (Figure 1-10). This structure has been shown to be required for viral replication (Hsue et al., 2000; Hsue and Masters, 1997). At the downstream of this bulged stem-loop structure, Williams et al. have shown in BCoV a 54 nucleotides hairpin-type pseudoknot structure (Williams et al., 1999) is required for replication, and is phylogenetically conserved in all three coronaviruse groups. Intriguingly, stem F of bulged stem-loop structure partially overlap with the stem 1 of pseudoknot structure, which indicate that the two structures cannot be formed simultaneously (Figure 1-10). Therefore a dynamic conformational exchange between these two structures has been proposed to represent a molecular switch (Goebel et al., 2004b; Hsue et al., 2000). Downstream of these two structures, a 74 nucleotides complex bulged stem-loop structure, denoted "hypervariable

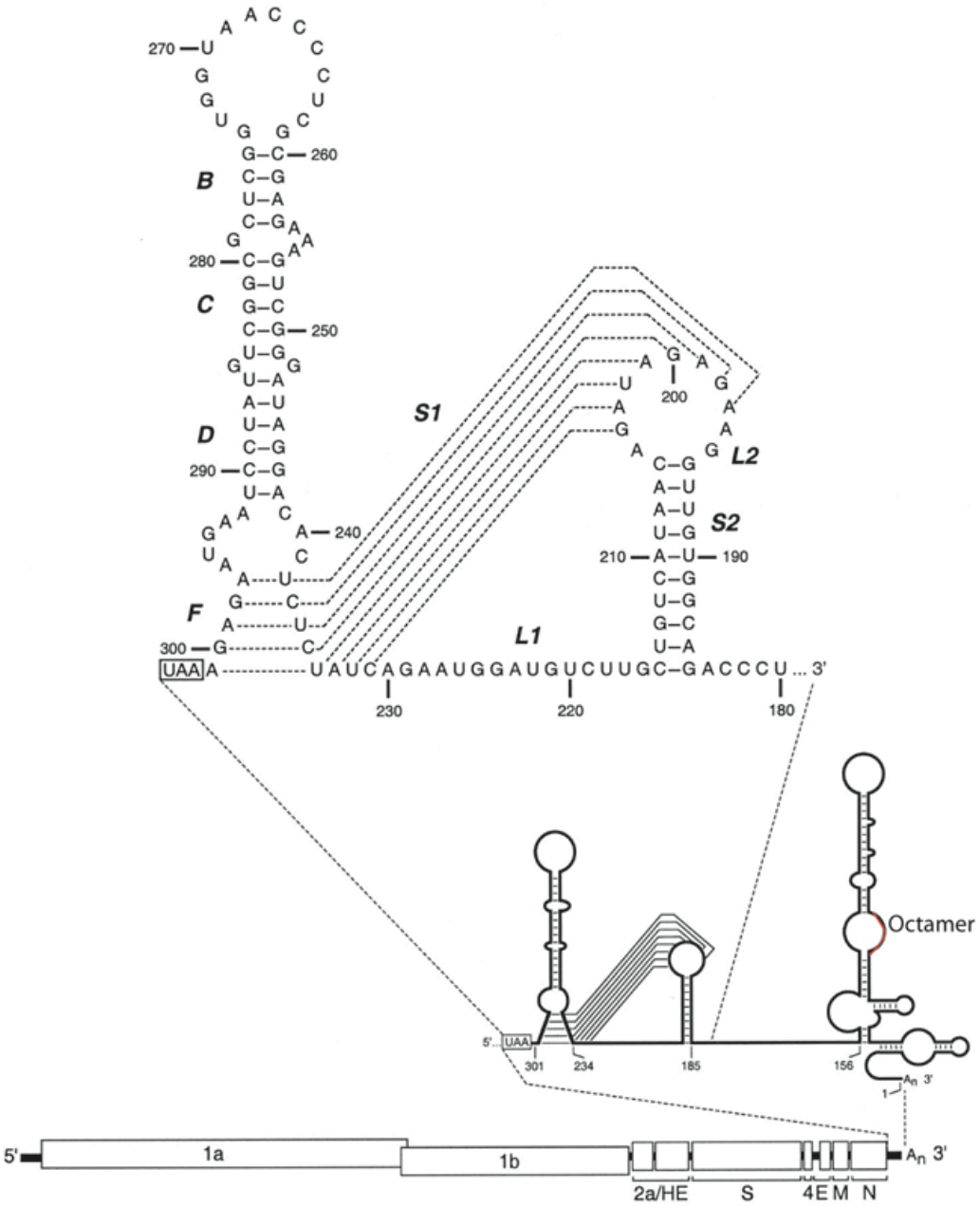


Figure 1-10 Proposed secondary structural elements within the 3' UTR of the coronavirus MHV. The expanded region represents the overlapping bulged stem loop and pseudoknot structure. The structures are mutually exclusive so that at any one point in time only one form is in existence. The conserved octamer is highlighted in red. Adapted from (Goebel et al., 2004a).

region" (HVR), makes up most of the rest of the 3' UTR in MHV (Liu et al., 2001) with the structure and sequence of this region poorly conserved in group 2 coronaviruses. However, this HVR contains a highly phylogenetically conserved octamer motif GGAAGAGC, considered as a signature motif for coronaviruses because it is highly conserved in all three group coronaviruses. To date, there are only three reported deviations in the 3' UTR of over 300 coronavirus strains from all three groups, GGGAGAGC in IBV isolate UK/919/68, GGAGGAGC in pigeon coronavirus, and GGAAGGGC in ferret enteric coronavirus (Goebel et al., 2007). Recent work by the Masters group shows that neither the HVR nor the octamer is essential for MHV RNA synthesis in tissue culture, but the HVR plays a significant as yet undefined role in viral pathogenesis (Goebel et al., 2007).

Virally encoded proteins in coronavirus replication/transcription

Two-thirds of the coronavirus genome is composed of two open reading frames termed orf1a and orf1b, with pp1ab expressed as a result of a -1 programmed ribosomal frameshift. Sixteen mature nonstructural proteins (designated nsp1-nsp16) are produced as a result of the proteolytic cleavage of pp1a and pp1ab (Brian and Baric, 2005). Nsp1-nsp11 are produced from pp1a, while nsp1-nsp10 and nsp12-nsp16 are produced from pp1ab. Functionally important proteins derived from pp1a include one or two papain-like cysteine protease (PL^{pro}s), a chymotrypsin/picornaviral 3C-like protease ($3CL^{pro}$, also termed main proteinase, designated M^{pro}) responsible for processing the polyprotein as well as other activities (Snijder et al., 2003), and a membrane-associated protein (MP). Orf1b is the most conserved region of the coronavirus genome and is typically

used for phylogenetic analysis of the coronaviral genomes. The RNA-dependent RNA polymerase (RdRp), a zinc-finger nucleic acid-binding protein (or metal-binding protein), and a nucleoside triphosphate (NTP)-binding helicase are encoded in this region.

Proteases: The viral proteases PL^{pro} and $3CL^{pro}$ are important targets for development of antiviral drugs that would inhibit viral replication. PL^{pro}s are early replicase products derived from nsp3 (Baker et al., 1993) (see Figure 1-5). Most coronaviruses contain two PL^{pro}s, termed $PL1^{pro}$ and $PL2^{pro}$, with only IBV and SARS-CoV containing a single PL^{pro}. The crystal structure of PL^{pro} from SARS-CoV is now available (Ratia et al., 2006), and reveals that the PL^{pro} has shape of a right hand with fingers, palm, thumb subdomains and an extend N-terminal Ub1 domain (Figure 1-11A). In MHV and HCoV-229E, $PL1^{pro}$ and $PL2^{pro}$ have been shown to recognize and process cleavage sites with different efficiencies (Kanjanahaluethai and Baker, 2000). *In vitro* studies indicate that $PL1^{pro}$ and $PL2^{pro}$ together process the N-terminal region of the polyproteins and generate nsp1, nsp2, nsp3 and N-terminus of nsp4. A reverse genetic study shows that $PL2^{pro}$ is essential, while $PL1^{pro}$ is not required for virus replication, although it does benefit virus fitness (Thiel and Siddell, 2005). Both coronavirus PLs^{pro} contain a transcription factor-like zinc finger domain (see Figure 1-11A), which suggest that these proteins may also directly bind nucleic acids (RNA) at some point in the life cycle. This idea is supported by findings from the arterivirus EAV, in which a papain-like protease, homologous to coronaviral PL^{pro}s, functions by coupling the sg-mRNA synthesis to genome translation (Tijms et al., 2001).

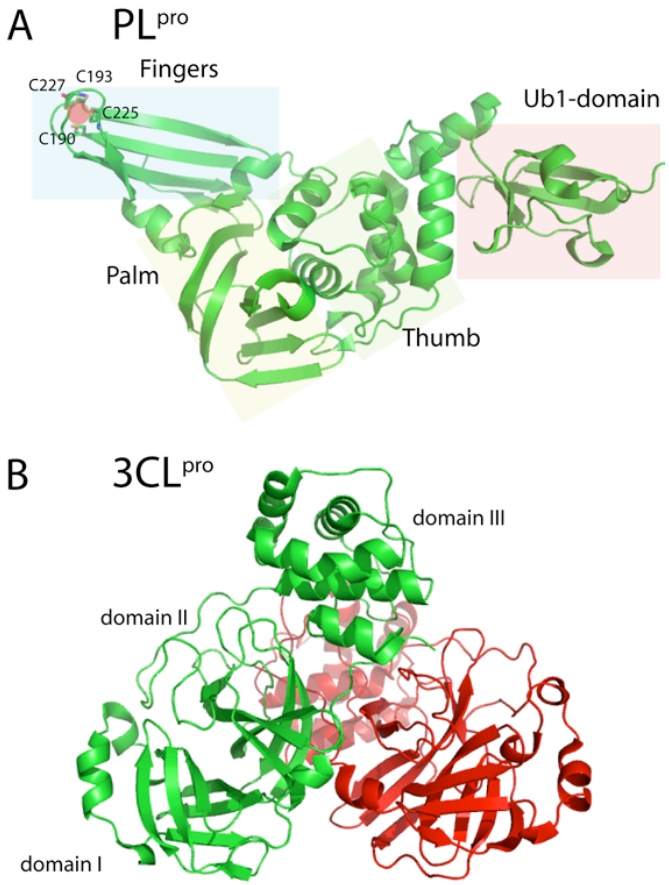


Figure 1-11 Ribbon diagrams of domain organizations of proteases from SARS-CoV. In (A) PL^{pro} (PDB code: 2FE8) structure, the structure motifs of fingers, palm, thumb and Ub1-domain are indicated, and the Zn^{2+} is shown as a red ball with the four coordinating Cys shown as sticks. (B) SARS-CoV $3CL^{pro}$ (PDB code: 1UK2) forms a homodimer with one monomer shaded red and the other one shaded green. Domain I, II, and III are shown in one monomer.

$3CL^{pro}$ localizes to nsp5 and shares low sequence similarity to other viral and cellular proteinases. $3CL^{pro}$ is believed to be the key viral protease responsible for the viral replicase complex formation. $3CL^{pro}$ mediates the processing of at least 12 protein products, including itself, the RdRp and the helicase. An inhibitor of the $3CL^{pro}$ inhibits

the proteolytic processing the MHV polyprotein (also termed the replicase) and therefore inhibits RNA synthesis (Kim et al., 1995). The crystal structures of $3CL^{pro}$ from TGEV (Anand et al., 2003), HCoV-229E (Anand et al., 2002), and SARS-CoV (Yang et al., 2003) are now available. SARS-CoV $3CL^{pro}$ forms a dimer, with each monomer composed of three domains. Domain I and II have an antiparallel β-barrel structure, a reminiscent of the trypsin-like serine proteases, and domain III contains five α-helices with a long loop connecting to domain II (Figure 1-11B).

Why do the coronaviruses require these two or three distinct viral proteinases? Although the reason for this is unknown, the large genome of coronavirus as well as the need for discontinuous sgRNA transcription that facilitates high-frequency recombination, may require distinct protein processing intermediates and/or macromolecular complexes at different stages in the viral life cycle. In addition, their activities may also be highly regulated. A good example of this is in the alphavirus, Sindbis virus (SV), in which only a single proteinase, nsp2, first functions in *cis* by processing the replicase polyprotein into intermediates that function in negative-strand RNA synthesis and then acts in *trans* to further process these intermediates into four mature products that function in positive-strand RNA synthesis (Shirako and Strauss, 1994).

Helicase: Helicases represent a diverse class of enzymes that unwind the duplex of both DNA and RNA by using the energy of NTP hydrolysis. Thus are involved in virtually every aspect of RNA and DNA metabolism (Kadare and Haenni, 1997; Singleton and Wigley, 2002). Helicases possess a common core structure with N- and/or

C-terminal accessory domains that confer differences in both activity and specificity (Singleton and Wigley, 2002). Coronavirus helicase is found in nsp13, a cleavage product of pp1ab flanked by two $3CL^{pro}$ cleavage sites (Snijder et al., 2003; Thiel et al., 2003a), and has been shown to possess multiple enzymatic activities, including RNA and DNA duplex-unwinding activities, NTPase and dNTPase activities likely important for genome RNA replication and sgRNA transcription, and an RNA 5′-triphosphatase activity which is thought to be involved in the formation of the 5′-cap structure of viral RNAs (Ivanov and Ziebuhr, 2004; Ivanov et al., 2004). A unique property of coronaviral helicases, as well as all nidovirus helicases, is that they localize in the viral polyprotein downstream of the RdRp. This arrangement makes the helicase protein generally precedes the RdRp in the viral polyprotein (Koonin and Dolja, 1993). Nidovirus helicases also differ from other positive-stranded RNA viral helicases in that they unwind duplex nucleic acids with a 5'->3' polarity in stead. This suggests that by effectively unwinding the double-stranded regions, the helicase can mediate the formation of the single-stranded RNA template and cooperates with the RdRp to facilitate RNA synthesis. This polarity might also be involved in the nascent negative-sense sgRNA release as well (Ivanov et al., 2004).

Coronaviral helicases contain two domains: a putative N-terminal zinc binding domain spanning from 1 to 77 amino acid and a C-terminal helicase domain extending from amino acid 279 to the C-terminus (Ivanov et al., 2004). In HCoV-229E, mutagenesis of the putative zinc binding domain resulted in a modulation of the enzymatic activity of the helicase (Seybert et al., 2005) and was found to be critical in

viral RNA replication and transcription. However, little is known about the function of the connecting region between the two domains. Fang *et al.* have shown that a R132P single mutation at nucleotide position 15526 is lethal to the infectivity of IBV on Vero cells and this mutation is proposed to block sgRNA synthesis (Fang et al., 2006).

Other viral proteins from the replicase polyprotein: Our knowledge about other components of the replicase polyprotein and their individual functional roles remains limited. Besides RdRp (nsp12) and helicase (nsp13), a 3'->5' exonuclease (designated ExoN) in nsp14, a U-specific endoribonuclease (designated EndoU) in nsp15 and an adenosylmethionine-dependent ribose 2'-O-methyltransferase (designated OMT) in nsp16 have also been predicted to be localized to pp1ab (Snijder et al., 2003). Each of these activities has been biochemically confirmed (Bhardwaj et al., 2004; Sperry et al., 2005; von Grotthuss et al., 2003). On the other hand, the cleavage products of pp1a are suggested to be involved in localizing the replicon complex to the membrane and preparing replication accessory factors that function in some aspect of viral RNA synthesis (Masters, 2006). In addition to PL^{pro}s in nsp3 and $3CL^{pro}$ in nsp5 described previously, nsp1 is thought to play a role in cell cycle arrest (Chen et al., 2004), and the putative transmembrane domains (TM) of nsp3, nsp4 and nsp6 are thought to mediate association of the replicase complex to intracellular membrane.

A "cassette" of four small conserved proteins nsp7-nsp10 encoded near the C-terminal end of orf1a are processed from the pp1a by $3CL^{pro}$ to yield mature products of 10, 22, 12.7 and 15 kDa, respectively. These proteins co-localize with the replication complex and are suggested to be involved in viral RNA synthesis (Bost et al., 2000; Lu

et al., 1998; Van der Meer et al., 1999). The structures of all four proteins have now been reported. The cocrystal structure of SARS-CoV nsp7 and nsp8 reveals that the two proteins form a hexadecameric super molecular complex resembling a hollow cylinder-like structure assembled with 8 copies of each protein (Figure 1-12A, D). The positive electrostatic potential properties of the inside channel of this complex suggests a role in nucleic acid binding, perhaps to encircle and stabilize the RNA template. A role as a processivity factor or "sliding damp" for the RdRp during replication and/or transcription has been proposed (Zhai et al., 2005). The SARS-CoV nsp9 crystal structure has shown that nsp9 forms a homodimer and resembles a carboxyl-extended oligonucleotide-oligosaccharide binding (OB) fold (Figure 1-12B).

Nsp9 is shown to interact with nsp8 and possess single-stranded RNA-binding properties. It has been suggested to serve to stabilize nascent and template RNA during replication, transcription, and processing (Campanacci et al., 2003; Egloff MP, 2004; Sutton et al., 2004). Complementation studies using temperature-sensitive mutants of nsp10 of MHV have shown that nsp10 is essential for the assembly of a functional replication/transcription complex (Sawicki, 2005). The crystal structure of nsp10 reveals a single domain protein with a novel fold, consisting of a pair of antiparallel N-terminal helices stacked against an irregular beta-sheet, a coil-rich C terminus, and two Zn fingers unique in coronaviruses (Joseph et al., 2006) (Figure 1-12C). Nsp10 exhibits nucleic acid binding activity, and can crystallize to form a spherical dodecameric structure of a hollow, positively charged cylinder that is thought to be a transcriptional regulator (Donaldson et al., 2007b; Joseph et al., 2006; Su et al., 2006). The Baric group have

recently employed a reverse-genetics approach to generate mutants of nsp10 in the context of the entire MHV genome and demonstrated that nsp10 is a critical regulator of coronavirus RNA synthesis. The analysis of temperature-sensitive mutant TS-LA6 suggests that nsp10 plays an important role in polyprotein processing as an activator of $3CL^{pro}$ activity (Donaldson et al., 2007a; Donaldson et al., 2007b). Althought more molecular details will emerge from future studies, the recent findings taken collectively reveal that nsp7-nsp10 play important role in coronavirus replication and transcription.

Nucelocapsid (N) protein: The N protein is the most abundant coronaviral structural proteins. Besides its primary structure role in the assembly of the virus particle, N is now widely accepted to play a role in viral RNA synthesis (Compton et al., 1987). N has been shown in several studies to bind specific coronavirus RNA sequences that are involved in the regulation of RNA synthesis, including the TRS sequences, the 5' leader region, and sequences located in the 3' UTR of the virus genome (Baric et al., 1988; Cologna and Hogue, 2000; Huang et al., 2004; Nelson et al., 2000). Recent studies reveal that N from TGEV functions as a RNA chaperon and can mediate the annealing of TRS-L and nascent anti-TRS-B sequences; this suggests that N might play a functionally important role in mediating template jumping thought to be important for sgRNA transcription (see Figure 1-7) (Zuniga et al., 2007). N has been demonstrated to be required in *cis* or in *trans* for optimal activity of TGEV replicons (Almazan, 2004), as well as in *cis* for efficient BCV DI RNA replication (Chang and Brian, 1996). N has been detected in the membrane fraction contains both the replicase and viral RNA (Bost et al., 2000; Sims et al., 2000), consistent with a role in transcription and assembly of

infected viral particles. In addition to its regulatory role in viral RNA synthesis, N also has been shown to involved in the translational control through an as yet undefined mechanism (Tahara et al., 1998).

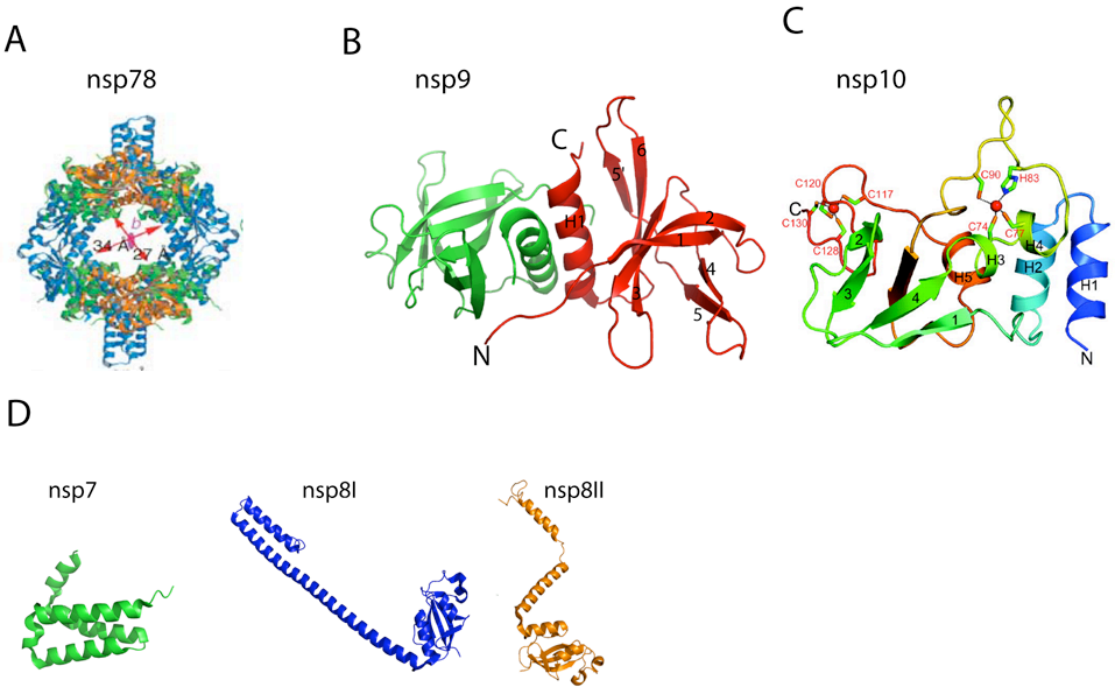


Figure 1-12 Ribbon diagrams of the crystal structures of SARS-CoV nsp7-nsp10. (A) nsp7-nsp8 hexadecameric complex (PDB code: 2AHM) is composed of 8 subunits of nsp7, 4 subunits of nsp8 with conformation of nsp8I and 4 subunits of nsp8 with conformation of nsp8II shown in (D). (B) nsp9 (PDB code: 1QZ8) forms a homodimer with one monomer shown in red and the other in green. The α helix and β strands are indicated in one monomer of nsp9 as well as in (C) nsp10 (PDB code: 2FYG). In nsp10, the Zn^{2+} are shown as red spheres, and the residues coordinating the Zn^{2+} are shown as sticks.

The multifunctional features of N derive directly from its RNA binding properties which at present remain partially defined. There are several known or projected RNA binding domains within N, including an N-terminal OB-fold like RNA-binding domain (NTD), a Ser/Arg (SR)-rich region, and the N-terminal region of C-terminal dimerization domain (CTD). Recent structural studies of the isolated NTD and CTD fragments have provided some insights into how N might function (Chen et al., 2007; Fan et al., 2005; Huang et al., 2004; Jayaram et al., 2006; Luo et al., 2006a) (to be discussed in more detail in Chapter IV); however, biochemical studies have been extremely limited.

Scope of this work

The work described in this dissertation is the result of a collaborate project with Dr. Leibowitz's laboratory, with the goal to characterize the structure and function of the 5' UTR, and in particular the 5' leader region of MHV, a well characterized model system for group 2 coronaviruses. These studies were initiated with our elucidation of a novel consensus secondary structure model of the 5' UTR that appears common to all coronavirus groups (Chapter II) with subsequent studies on an effort to test the structural validity and functional importance of the model. This model is characterized by three major RNA stem loops (SLs), SL1, SL2 and SL4, with the TRS positioned in a weakly base paired (SL3) or unstructured region positioned between SL2 and SL4. NMR experiments provide support for the structural model, with SL2 (Chapter II) and SL1 (Chapter III) characterized by unique structural and functional properties. SL2 is

invariant and was found to exhibit structural features consistent with a novel U-turn architecture. In Chapter III, the thermostability, the structural and dynamic properties of wild-type and mutant SL1 hairpins were systematically investigated. These experiments were carried out in an effort to understand second-site revertant mutants recovered from cells infected with a genetically unstable virus, SL1-ΔA35, in which a predicted extrahelical adenosine was deleted from SL1. These studies allowed us to propose that a "dynamic" SL1 structure is specifically required for the genome circularization required for the sgRNA synthesis. In Chapter IV, the first comprehensive RNA binding experiments for MHV N are presented, with the focus on understanding the ability of N to stimulate sgRNA synthesis through a specific interaction with the TRS-L. The work summarized in this dissertation provides new molecular level insight into how RNA structure and RNA-protein interaction associated with the 5' leader region stimulates sgRNA synthesis by coronaviruses.

CHAPTER II

A U-TURN MOTIF-CONTAINING STEM-LOOP IN THE CORONAVIRUS 5' UNTRANSLATED REGION PLAYS A FUNCTIONAL ROLE IN REPLICATION*

Introduction

The *cis*-acting sequences in the 5' UTR required for coronaviruses transcription and replication have been defined by studying defective interfering (DI) RNAs (Kim et al., 1993; Lin and Lai, 1993). However, the 5' UTR of only one group 2 coronavirus, BCoV, has been extensively studied to date. Four stem-loops, denoted I, II, III and IV, that map within the 210 nucleotides UTR of BCoV have been predicted and their existence was supported by RNase probing and functional studies (Chang et al., 1994; Chang et al., 1996; Raman et al., 2003; Raman and Brian, 2005; Wang and Zhang, 2000). The predicted stem-loop I (nts 11-42) contains just three contiguous Watson-Crick base pairs and a large 16-nt loop and is not conserved among group 2 coronaviruses. In addition, a mutational study designed to examine the requirement for stem-loop I was not definitive, in that all of the mutations in the DREP1 DI RNA construct were rapidly replaced by wild type sequences (presumably derived from helper virus by recombination) (Chang et al., 1994) irrespective of whether they were predicted to maintain or disrupt the stem-loop. The predicted stem-loop II (51-84 nucleotides) is an A-U base pair rich hairpin with a low

---

negative free energy that folds the TRS into the terminal loop (Raman et al., 2003; Raman and Brian, 2005). In contrast, stem-loop III is phylogenetically conserved among group 2 coronaviruses and appears to have homologs in coronavirus groups 1 and 3, and enzymatic structure probing and DI RNA replication assays supports its existence (Raman et al., 2003). The fourth predicted stem-loop, stem-loop IV, mapped to nucleotides 186 to 215, is also predicted to be conserved amongst group 2 coronaviruses (Raman and Brian, 2005). RNase mapping supports the existence of this stem-loop and DI RNA replication assays indicate that this structure likely plays a functional role in RNA replication, perhaps as a target for the binding of cellular proteins (Raman and Brian, 2005).

In this chapter, we present consensus secondary structure model of the 5'-most 140 nucleotides in the 5' UTR regions of nine group 1 and group 2 coronaviruses, including five human coronaviruses. All nine coronaviral genomes are predicted to fold into a similar secondary structure containing three or four stem-loops in this region, including a highly conserved 5-nucleotide hairpin loop (SL2) that possesses sequence features consistent with a U-turn motif containing a UNR sequence (Gutell et al., 2000). NMR studies of RNAs corresponding to SL1, SL2 and SL1/SL2 fragments from MHV, HCoV-OC43 and BCoV provide structural support for the general features of the model. A mutational analysis of SL2 in the context of the complete MHV genome supports the existence of this stem-loop structure and further reveals that SL2 has an essential role in MHV replication.

Materials and methods

Sequence alignment and RNA structure prediction

Nine coronaviruses 5' UTR sequences were obtained from GenBank with accession numbers listed. RNA nucleotide sequences were first aligned using Clustal W and manually inspected for the presence of strongly predicted co-variations that would support the presence of conserved terminal hairpin loop sequences. Each RNA sequence was then subjected to secondary structure prediction using ViennaRNA 1.5 (Hofacker, 2003; Hofacker et al., 2004), *mfold* 3.1 (Zuker, 2003), and PKNOTS (Rivas and Eddy, 1999). *mfold* 3.1 predicts both a minimum free energy structure as well as all structures within a certain energy from the minimum. PKNOTS is a dynamic programming algorithm that incorporates pseudoknot pairing. ViennaRNA uses updated nearest-neighbor thermodynamic parameters that incorporate the energetics of stabilizing tetraloops and unpaired dangling ends and generates both a minimum free energy (mfe) structure as well as equilibrium base pairing probabilities using the partition function (pf) algorithm of McCaskill (McCaskill, 1990). With the exception of MHV, the mfe and thermodynamic ensemble predictions corresponded in every case, taken as good evidence for one major lowest energy secondary structural conformer that speaks to the robustness of the prediction. For MHV, the structure shown is within 1.7 kcal mol^{-1} of the mfe structure; if SL2 is forced to pair as indicated, the structure shown represents the mfe structure.

RNA preparation and NMR spectroscopy

RNAs were prepared by run off transcription using SP6 or T7 RNA polymerase, purified by denaturing PAGE and subjected to multiple cycles of ethanol precipitation as described previously (Cornish et al., 2005; Nixon et al., 2002a). $^{13}C,^{15}N$-[U]-labeled WT SL2 and $^{13}C,^{15}N$-[G]-labeled U48G RNAs were also prepared using $^{13}C,^{15}N$-[UTP] and $^{13}C,^{15}N$-[GTP] in transcription reactions, respectively. The final NMR buffer was 10 mM potassium phosphate, pH 6.0, unless otherwise indicated. The RNA concentration in each case ranged from 0.8-2.0 mM. All RNAs were checked for the presence of a monomer-dimer equilibrium by nondenaturing PAGE prior to extensive NMR analysis. 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}=200 ms) were acquired to obtain imino proton resonance assignments while ^{1}H-^{1}H D_2O NOESY (τ_{mix}=250 ms) and TOCSY experiments were performed to obtain non-exchangeable proton resonance assignments using standard methodologies (Furtig et al., 2003). A natural abundance 2D ^{1}H-^{13}C CT-HSQC spectrum was acquired to obtain ^{13}C and ^{1}H assignments for the adenosines in HCoV-OC43 SL1. Through-bond 2D HNCCCH (Simorre et al., 1995) H(CCN)H-TOCSY (Sklenar et al., 1996) experiments were used to correlate imino protons with uridine H6 and guanosine H8 protons, respectively, in $^{13}C,^{15}N$-[U]-labeled WT and $^{13}C,^{15}N$-[G]-labeled U48G SL2 RNAs, to provide unique starting points for sequential base-ribose connectivities in a D_2O NOESY spectrum (τ_m=150-250 ms for SL1 and

SL2-containing RNAs, respectively). A combination of sensitivity-optimized HCN-HMQC (ribose moiety, H1'-C1'-N1/9), HCN-TROSY (base moiety, H6/8-C6/8-N1/9), and HCNCH experiments provided unambiguous, through-bond sugar to base connectivities (Fiala et al., 2000) with complete ribose proton connectivities obtained with a ^{1}H-^{13}C HCCH-TOCSY experiment. Additional NOE restraints were obtained from analysis of a ^{13}C-edited NOESY spectra, a 2D ^{1}H-^{15}N CPMG-NOESY (60 and 250 ms) (Mueller et al., 1995), and $^{2}J_{HN}$-HSQC experiments (Simorre et al., 1996). Adenosine H2 protons were correlated with H8 protons using a HCCH-TOCSY experiment (Sklenar et al., 1996). All of the NMR data were processed using nmrPipe and analyzed using Sparky (Delaglio et al., 1995; Goddard and Kneller).

Luciferase assay

The plasmids pWFluc, pU48CFluc, pWRluc and pU48CRluc (supplied by Dr. Leibowitz's laboratory) were linearized with BamHI. One microgram of linearized plasmid was in vitro transcribed using the Ambion mMESSAGE mMACHINE High Yield Capped RNA Transcription Kit according to the manufacturer's instructions. The transcribed RNA was treated with 2U DNase I at 37 ºC for 30 minutes and the DNase I was heat inactivated at 70 ºC for 15 minutes. The treated RNA was purified with Ambion NucAway column, quantitated, and diluted to100 fmol/µl. One hundred fmoles of each transcribed RNAs was in vitro translated at 30ºC for 30 minutes individually or in different combinations using rabbit reticulocyte lysates (Promega). Translation reactions were stopped by the addition of cycloheximide and five microliters of the translation reaction was assayed for luciferase activity using the Promega Dual-

Luciferase Reporter Assay System. Each translation reaction was performed in triplicate and luciferase assays were performed for each translation reaction.

Results

RNA secondary structure prediction for group 1 and 2 coronaviruses

Vienna RNA 1.5 (Hofacker et al., 2004) was used to examine the secondary structures of nine coronavirus 5' UTRs, including five group 2 CoVs, BCoV and the closely related human coronavirus HCoV-OC43, MHV-A59, HKU1, and SARS-CoV (Figure 2-1C), as well as three representative group 1 CoVs, HCoV-NL63, HCoV-229E, and TGEV (Figure 2-1A). A similar secondary structure was predicted for the avian coronavirus IBV, a group 3 CoV. Minimum free energy (mfe) secondary structural models of the 5' 140 nucleotides of all CoVs are remarkably similar, and all contain three major helical stems, denoted SL1, SL2 and SL4. Some sequences are predicted to contain a fourth stem-loop, SL3, which folds the leader TRS (TRS-L) folding into a hairpin loop. Only for OC43 and SARS-CoV is SL3 predicted (at 37 ºC ΔG_{37}= -1.5 kcal mol^{-1} and ΔG_{37}= -2.2 kcal mol^{-1}, respectively); BCoV is capable of adopting the analogous SL3 stem-loop, corresponding to stem-loop II of (Raman and Brian, 2005) (see Figure 2-1B), although its T_m is predicted to be ≤37 ºC. The SARS-CoV 5' UTR differs from the other classical group 2 coronaviruses, in that SL2 is more distal from SL1. Extending these predictions to encompass the entire 5' UTR typically adds a few additional stem-loops to the mfe structure, as in the case of BCoV and OC43 (SL5-7), or another long multibranched stem structure in the 3' region as shown for SARS-CoV

(Figure 2-1C) and NL63 (Figure 2-1A), leaving the fold of the 5' ≈140 nts intact; this strongly suggests that our predictions are meaningful. We note that SL4b and SL7 in our complete BCoV prediction (Figure 2-1C) correspond to phylogenetically conserved and functionally important stem-loops III (Raman et al., 2003) and IV (Raman and Brian, 2005).

SL2 is absolutely conserved and strongly predicted to form in all coronaviruses examined. Except for the core TRS-L, the (C/U)UUG(U/C) sequence encompassing the predicted SL2 loop is the most conserved contiguous run of nucleotides in the entire 5' UTR of all coronaviruses examined. This conserved sequence only appears three or five times in the entire MHV or SARS-CoV genomes. Secondary structure analysis shows that this pentaloop is always stacked on a predicted 5-bp helix. This sequence conservation and the constancy of the predicted SL2 suggest an important functional role in coronavirus replication.

The (C/U)UUG(U/C) sequence of SL2 contains features of a canonical U-turn motif, in which the middle three nucleotides of the loop, UNR (U_0•N_{+1}•R_{+2}), forms a triloop that stacks on a Y:Y, Y:A or G:A noncanonical base pair (Gutell et al., 2000). U-turn motifs are widely distributed in transfer (Lescrinier et al., 2006; Quigley and Rich, 1976), ribosomal (Lebars et al., 2003), catalytic (Stallings and Moore, 1997) and viral (Puglisi and Puglisi, 1998) RNAs and often mediate RNA-RNA interactions between helical elements (Campbell and Legault, 2005). The basic structural feature of the canonical U-turn is a sharp turn in the phosphate backbone between U_0 and N_{+1}, with U_0 stacked on the noncanonical base pair and engaged in two critical hydrogen bonds: the

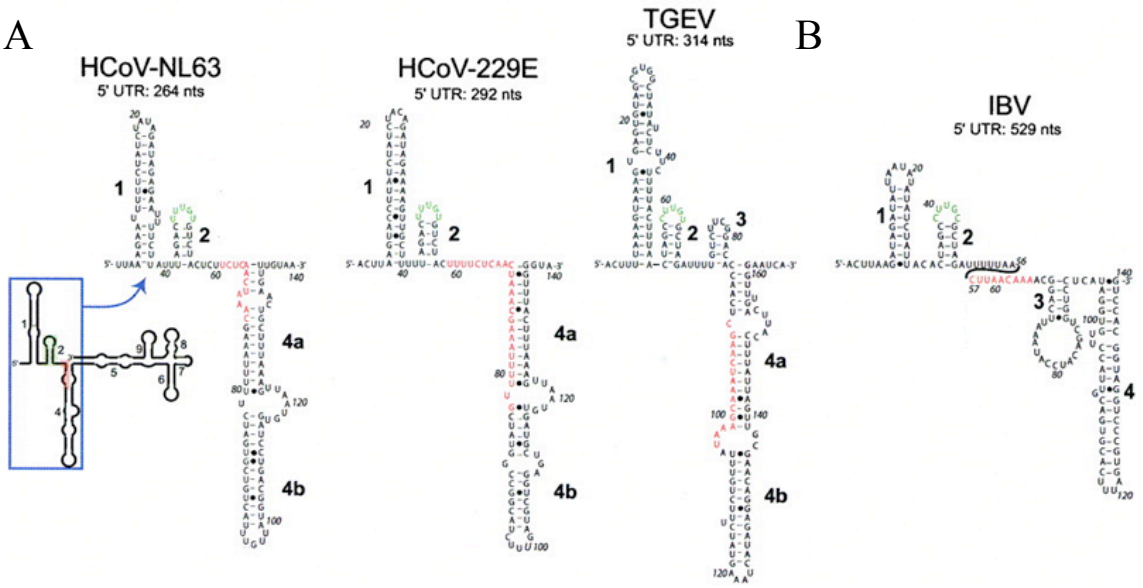
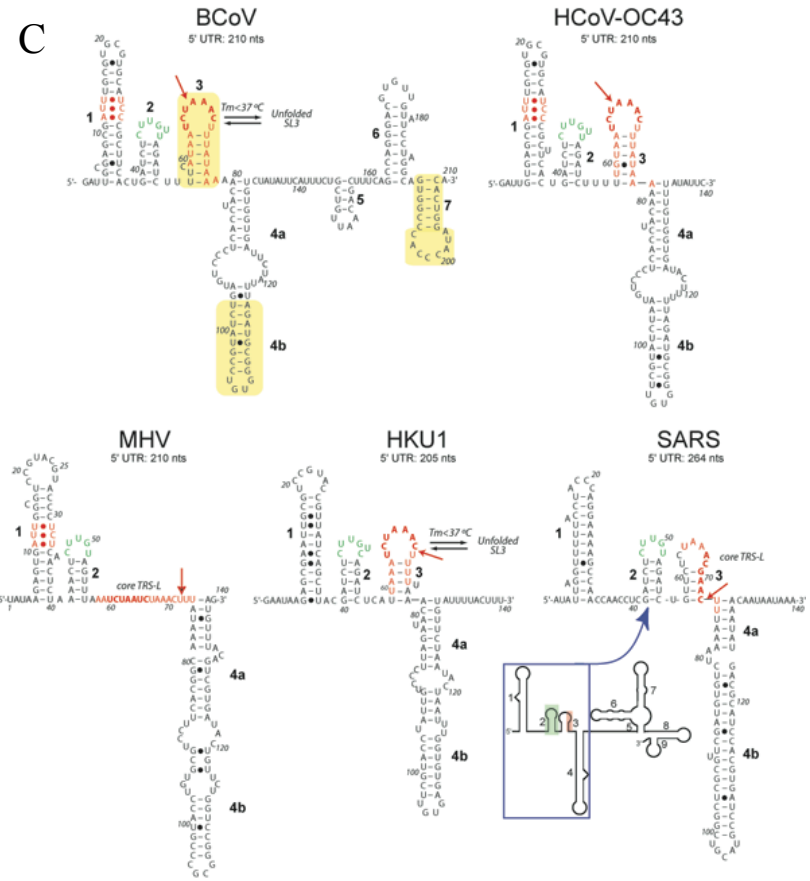


Figure 2-1 Predicted secondary structure models for the 5' UTR of nine coronaviruses. (A) Predicted secondary structure models for three group 1 coronaviruses. (B) Predicted secondary structure model for a group 3 coronavirus, avian infectious bronchitis virus (IBV). (C) Predicted secondary structure models for the entire 5' UTR of BCoV compared to the 5' 140 nucleotides of selected group 2 coronaviruses. Predicted stem loops SL1, SL2, SL3, and SL4 (4a and 4b) are indicated by the bold numbers. Leader TRS-L sequences shown in bold letters. SL-II, SL-III and SL-IV of (Raman et al., 2003; Raman and Brian, 2005) are shaded yellow. Nucleotide positions are numbered according to GenBank accession numbers (BCoV-LUN: AF391542; HCoV-OC43: NC_005147; MHV-A59: NC_001846; HKU1: NC_006577; SARS-CoV: NC_004718; HCoV-NL63: NC_005831; HCoV-229E: NC_002645; TGEV: NC_002306); IBV: NC_001451). All models except one represent mfe structures, and are predicted by both mfold, PKNOTS and ViennaRNA. The lone exception is MHV, which represents a structure within 1.7 kcal mol^{-1} of the mfe structure. If SL2, the strongest secondary structure in a co-variation analysis (not shown), is forced to pair as indicated, the structure shown represents the mfe structure.



(Figure 2-1 Continued)

U_0 imino proton donates a hydrogen bond to the nonbridging phosphate oxygen following R_{+2}, and the U_0 2'-OH proton donates a hydrogen bond to the N7 of R_{+2}. Single or multiple nucleotide insertions in a U-turn motif are not uncommon and, in some cases, the polynucleotide strand is diverted in a different direction (Gutell et al., 2000); in most cases, these breaks occur exclusively 3' to R_{+2} (G) like that in the hammerhead ribozyme (Feig et al., 1998). This is, in fact, where one additional uridine is inserted in the SL2s of BCoV, OC43 and HKU1 (Figure 2-1C). However, it is important to recognize that recent structural data suggest that U-turn motifs are conformationally

diverse in solution, and may be lacking one of more of the key structural features associated with canonical U-turn motifs (Campbell and Legault, 2005).

NMR spectroscopy of SL1 and SL2-containing RNAs

SL1s from BCoV/HCoV-OC43 and MHV are predicted to contain 13-14 base pairs capped by a four-nucleotide UGCG (YNMG) (Proctor et al., 2002; Theimer et al., 2003) (BCoV/OC43) or eight nucleotide (MHV) hairpin loop. Both SL1s are predicted to contain 2-3 non-canonical base pairs in the middle of the stem. To determine whether SL1 forms in solution, several RNAs were prepared and characterized by NMR spectroscopy. Figure 2-2A shows a sequence comparison of BCoV and OC43 SL1s and as can be seen, all sequence differences are localized to the base of SL1. The 1D imino proton region of an RNA corresponding to HCoV-OC43 SL1 termed OC43 SL1-Δ33 (see Figure 2-2B) is shown in Figure 2-2C, with resonance assignments obtained from analysis of a 150 ms Watergate NOESY spectrum acquired at 10 ºC at pH 6.0 (see Figure 2-3). The most notable feature of this spectrum is a U14•U27 base pair as predicted by the model, with both uridine imino protons strongly protected from exchange with solvent (Du et al., 2004; Ohlenschlager et al., 2004; Theimer et al., 2003). In addition, analysis of a ^{1}H-^{13}C HSQC experiment (Figure 2-2C, inset) clearly shows the presence of an adenosine residue protonated at N1, since the $^{13}C2$ chemical shift is strongly shifted upfield (Huppler et al., 2002). A8 and A26 H2 protons were assigned by virtue of the intense NOE to the U H3 imino proton of the A8-U34 and U15-U27 Watson-Crick base pairs, respectively, with the $A12^{+}$ H2 assigned by elimination. Consistent with this, analysis of thermal melting profiles for OC43 SL1-Δ33 acquired at

pH 5.5 and 8.0 reveal an ≈3 ºC shift in T_m upon protonation, with the highest T_m at ~ pH 6.0 (data not shown), consistent with previous studies of A+•C base pairs (Huppler et al., 2002). These studies establish that SL1 forms in BCoV/OC43 as predicted by the model.

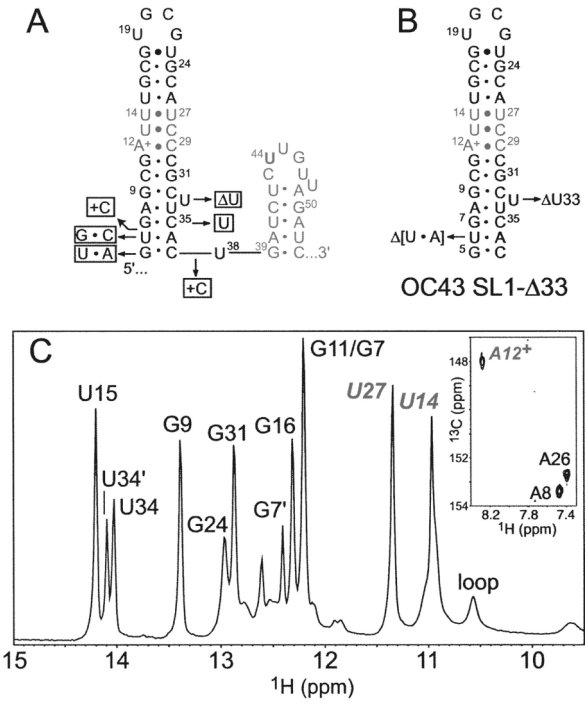


Figure 2-2 Secondary structure model and 1D spectrum HCoV-OC43 SL1-Δ33 RNA. (A) Nucleotide substitutions, insertions and deletions at the base of SL1 that correspond to the BCoV-Lun sequence are indicated in the adjacent boxes. (B) HCoV-OC43 SL1 construct used for NMR studies, denoted OC43 SL1-Δ33. The U6-A36 base pair was excised to enable transcription by T7 RNA polymerase; the extrahelical U33 was also deleted. C) Imino proton region of a 1D jump-return echo spectrum acquired at 10 ºC, pH 6.0 for OC43 SL1-Δ33. Resonance assignments were obtained from analysis of a homonuclear Watergate NOESY spectrum (τ_m=150 ms, see Figure 2-3). The U14-U27 base pair was verified by the presence of a strong crosspeak in a NOESY spectrum

acquired at a short mixing time (τ_m=50 ms). *Inset,* region of a natural abundance ^{1}H-^{13}C HSQC spectrum acquired for OC43 SL1-Δ33, with assigned adenosine ^{13}C2-^{1}H2 crosspeaks indicated.

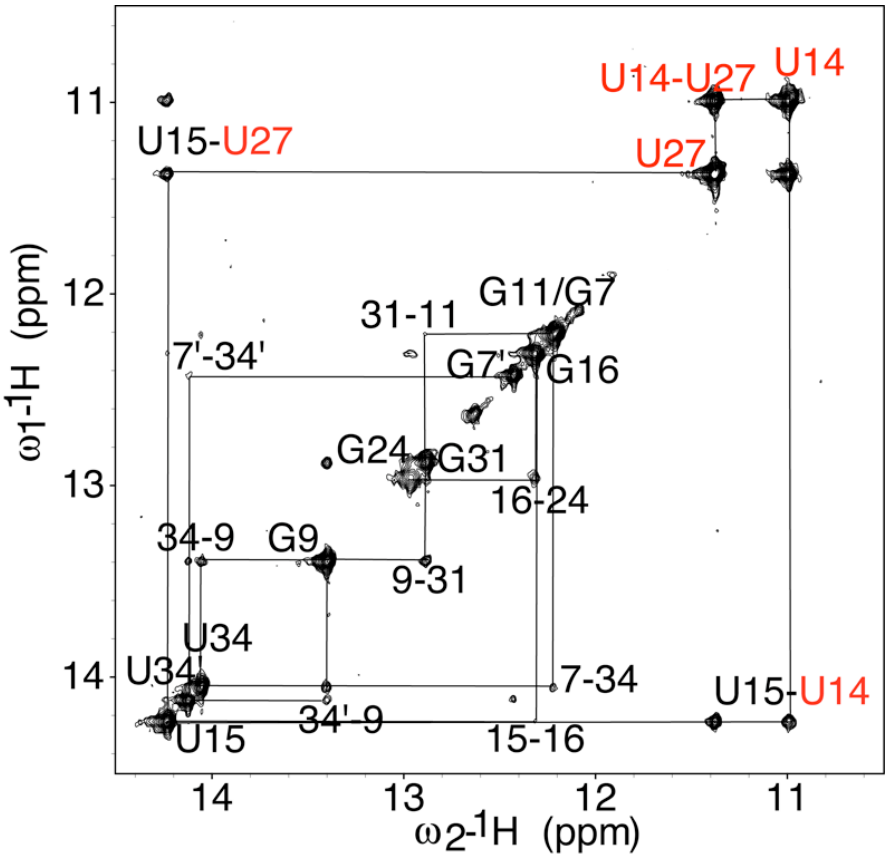


Figure 2-3 600 MHz Watergate NOESY spectrum of HCoV-OC43 SL1-Δ33 RNA. Spectrum is acquired at 10 mM potassium phosphate, pH 6.0, 10 ºC, (τ_m=150 ms). Note the heterogeneity at base of stem indicated by the G7 and U34 peak doubling.

We next prepared a series of SL1 RNAs corresponding to SL1 from MHV-A59. In contrast to BCoV/HCoV-OC43 SL1, MHV SL1 contains two unpaired nucleotides in the stem, C16 and A35, and is capped by an 8-nucleotide loop (Figure 2-4A).

Interestingly, if C19 and C20 are looped out, the MHV loop sequence becomes structurally identical to that of BCoV/HCoV-OC43, *i.e.,* a YNMG (U22-G25) tetraloop (Proctor et al., 2002) capped on a closing G21•U26 base pair (Figure 2-4B). 1D imino proton spectra are shown for three different variants of SL1, one with all four "looped-out" nucleotides deleted (SL1-$\Delta^{16/19/20/35}$; Figure 2-4C), SL1-$\Delta^{16/19/20/35}$ with A35 added back (SL1-$\Delta^{16/19/20}$; Figure 2-4D) and SL1-$\Delta^{16/19/20/35}$ with A35 and C16 added back (SL1-$\Delta^{19/20}$; Figure 2-4E). The spectra of all three RNAs are substantially identical, and all are characterized by a non-canonical U13•U31 base pair as predicted by the model (Figure 2-4A). A comparison of SL1-$\Delta^{16/19/20/35}$ (Figure 2-4C) with SL1-$\Delta^{16/19/20}$ (Figure 2-4D) reveals that the major change in the spectrum is dynamic in nature, with the imino protons corresponding to U9, G10 and U33, *i.e.*, those closest to the introduced A35, substantially broadened (at 10 ºC). Indeed, examination of the pyrimidine H5/H6 regions of homonuclear TOCSY spectra acquired for SL1-$\Delta^{16/19/20/35}$ and SL1-$\Delta^{16/19/20}$ (at 25 ºC) suggest that the two RNAs adopt similar structures with just 3 of 18 pyrimidine residues perturbation by the deletion of A35 (Figure 2-5); these results collectively suggest that A35 is extrahelical and that alternative pairing of U9 with A35 or A36 introduces a local destabilization and minor structural perturbation of this region of SL1. The same appears true of SL1-$\Delta^{19/20}$, except that the spectra degrade substantially in overall quality. The broader linewidths in this sample suggest that this is attributable to dimerization of the RNA.

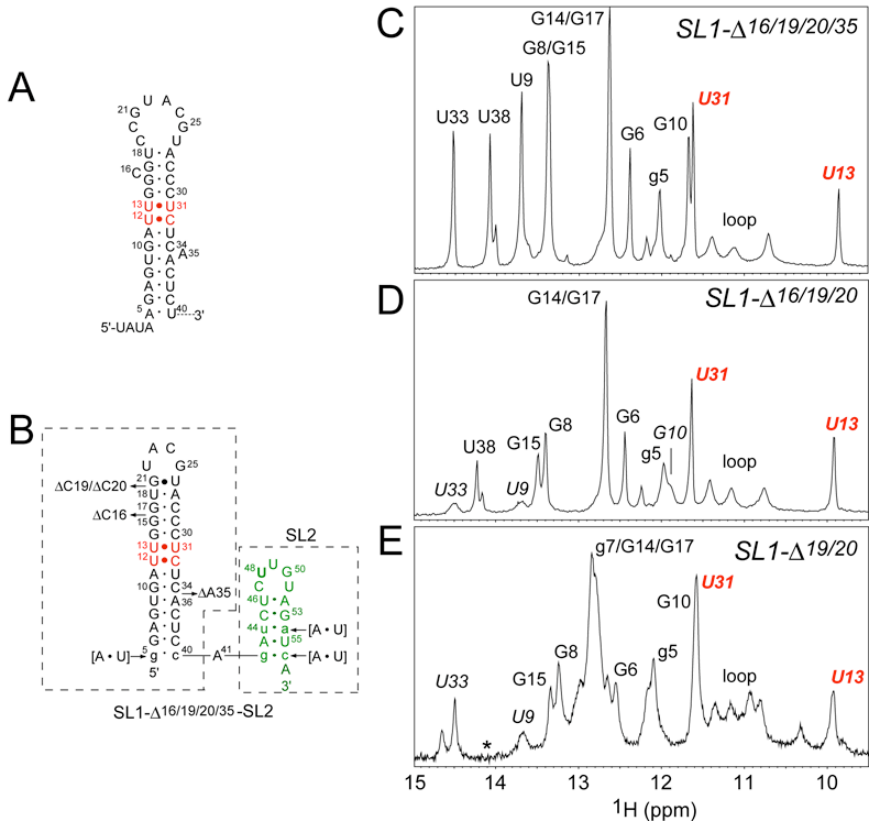


Figure 2-4 Imino proton regions of 1D spectra of MHV SL1 constructs. **(**A) Predicted secondary structure of MHV-A59 SL1. (B) Representation of the SL1, SL2 (boxed regions) and SL1-SL2 chimeras. All MHV SL1 constructs have a non-native g5-c40 base pair (native sequence shown in brackets) at the base of SL1 to facilitate transcription by T7 RNA polymerase, and incorporates a non-native base G-C pair base of SL2 and inverts the MHV A44-U54 pair in MHV to the U-A pair present in SARS-coronavirus (the native MHV sequence is shown in brackets, see text for details). Imino proton regions of 1D jump-return echo spectra acquired at 10 ºC unless otherwise indicated for (C) SL1-$\Delta^{16/19/20/35}$, (D) SL1-$\Delta^{16/19/20}$, and (E) SL1-$\Delta^{19/20}$ (5 ºC). Resonance assignments were obtained from analysis of homonuclear Watergate NOESY spectra (τ_m=150 ms) acquired for each RNA (300 ms for SL1-Δ19/20). Imino resonances for the U13-U31 base pair are indicated. The asterisk (14.2 ppm) in panel E indicates the expected absence of the imino resonance for the A7-U38 base pair due to substitution of a non-native g7-c38 base pair in this construct.

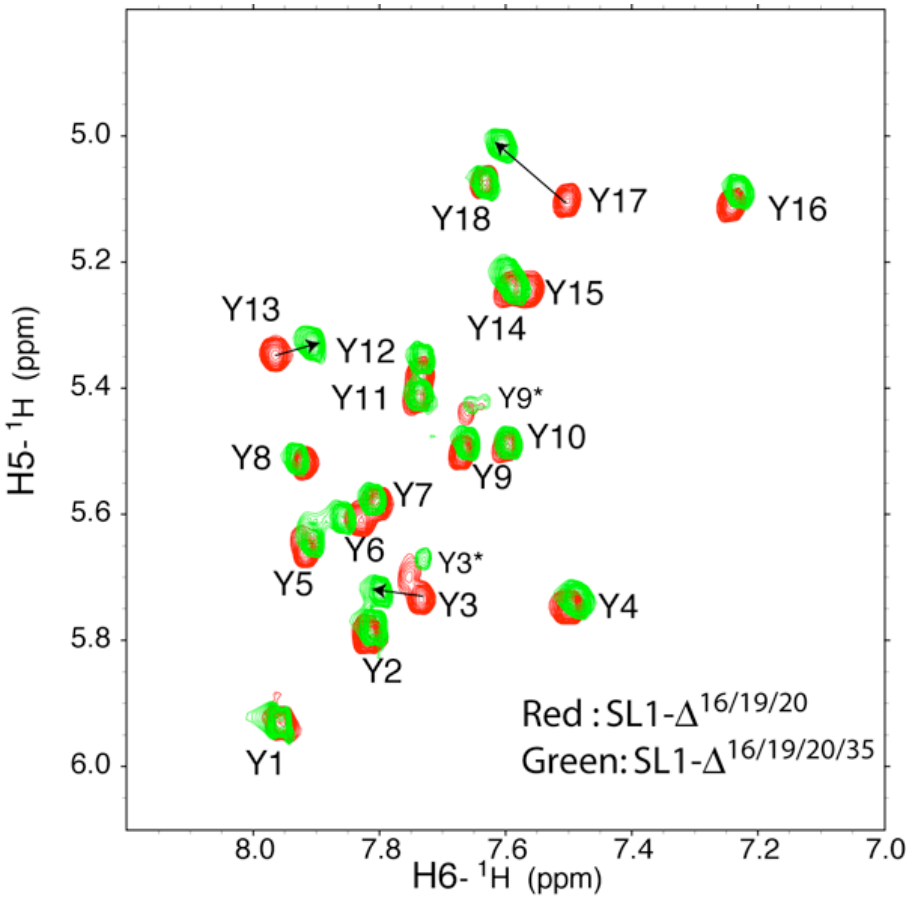


Figure 2-5 Superposition of the homonuclear TOCSY spectra acquired for SL1-$\Delta^{16/19/20/35}$ (green) and SL1-$\Delta^{16/19/20}$ (red) RNA. H5-H6 correlations of C and U residues are shown. 18 crosspeaks are expected with 18 observed (Y1→Y18), as yet unassigned. Minor conformers associated with Y3 and T9 are indicated with asterisk. 3 crosspeaks perturbations introduced by deletion of A35 are indicated by arrow. See Figure 2-4B for sequences for these RNA constructs. Conditions: 10 mM potassium phosphate, pH 6.0, at 25 ºC.

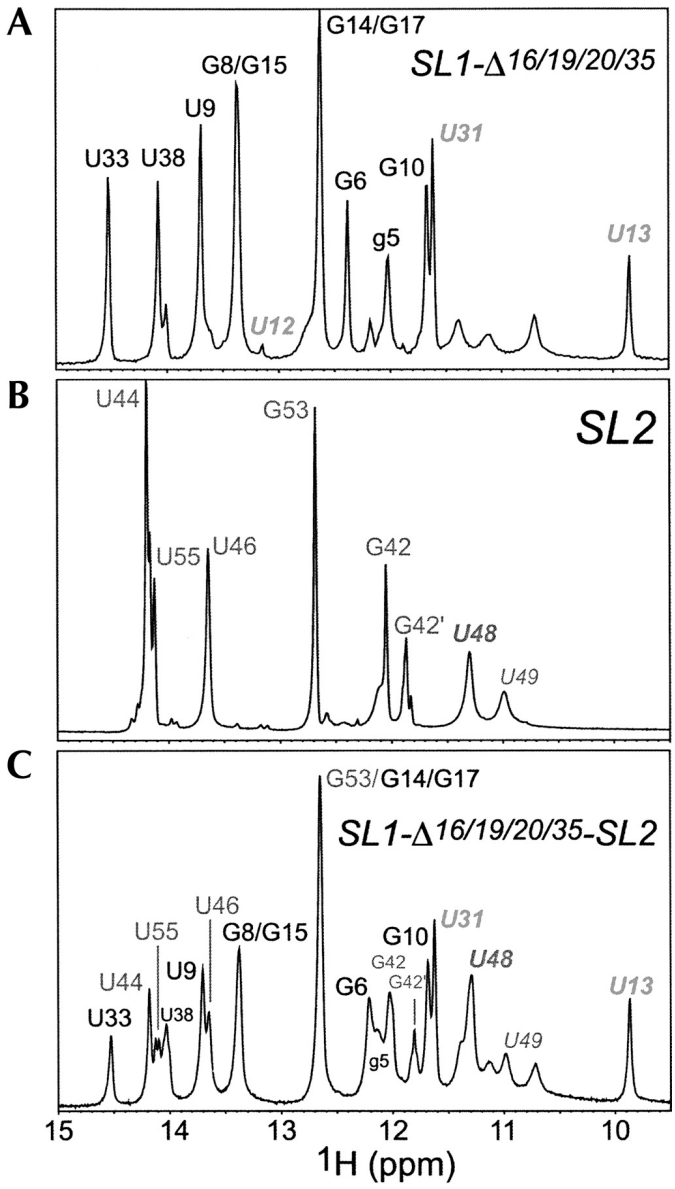


Figure 2-6 Imino proton regions of the 1D jump-return echo spectra of SL1-$\Delta^{16/19/20/35}$ (A), SL2 (B) and SL1-$\Delta^{16/19/20/35}$-SL2 (C) RNA. Conditions: 10 mM potassium phosphate, pH 6.0, 10 ºC. See Figure 2-4B for sequences for these RNA constructs. Note that the spectra for SL2 are characterized by slow conformational heterogeneity at the base of SL2 (G42-C56 and A43-U55 base pairs). The imino protons corresponding to U48 and U49 in the SL2 pentaloop are also indicated.

We next prepared a 16-nt SL2 RNA that corresponds to the SARS-CoV SL2 (see Figure 2-4B). This SL2 is more efficiently transcribed by SP6 RNA polymerase than the equivalent MHV SL2 and the stem will be more stable. Relative to the MHV sequence, SL2 contains A41-U56 to G41-C56 and A44-U54 to U44-A54 substitutions. A 3' adenosine cap was also added in an attempt to stabilize the adjacent stem (Theimer et al., 1998). We note that this SL2 stem sequence is also found in BCoV/HCoV-OC43 and HKU1 (Figure 2-1C, 2-2A), and replacement of the MHV SL2 with the SARS-CoV SL2 fully supports MHV replication (Kang et al., 2006). The 1D imino proton region of SL2 is shown in Figure 2-6B (pH 6.0, 10 ºC, no added salt), with the spectrum of SL1-$\Delta^{16/19/20/35}$ also reproduced here (Figure 2-6A) to facilitate comparison to the spectrum derived from a 49-nt RNA encompassing MHV SL1-$\Delta^{16/19/20/35}$ and SL2 (Figure 2-6C). The most notable feature of the SL2 spectrum is a relatively intense upfield shifted uridine imino proton, assigned to U48 (U_0 in the U_0-N_{+1}-R_{+2} nomenclature) (see below). This is consistent with the U48 H3 proton donating a hydrogen bond and thereby stabilizing a U-turn-like conformation. In addition, the imino proton of the closing U46-A52 base pair of the stem is also observable, suggesting the loop is structured and stacked on the U46-A52 base pair; this contrasts with a previous solution structure of a non-canonical U-turn motif (Campbell and Legault, 2005). In any case, the spectrum obtained for SL1-$\Delta^{16/19/20/35}$-SL2 RNA (Figure 2-6C) is very similar that expected for a superposition of component SL1 and SL2 hairpins. Although there are some difference, the superposition of the TOCSY spectra of these three RNA constructs is also consistent

with the idea SL1 and SL2 are independently folded and do not substantially interact with one another (Figure 2-7).

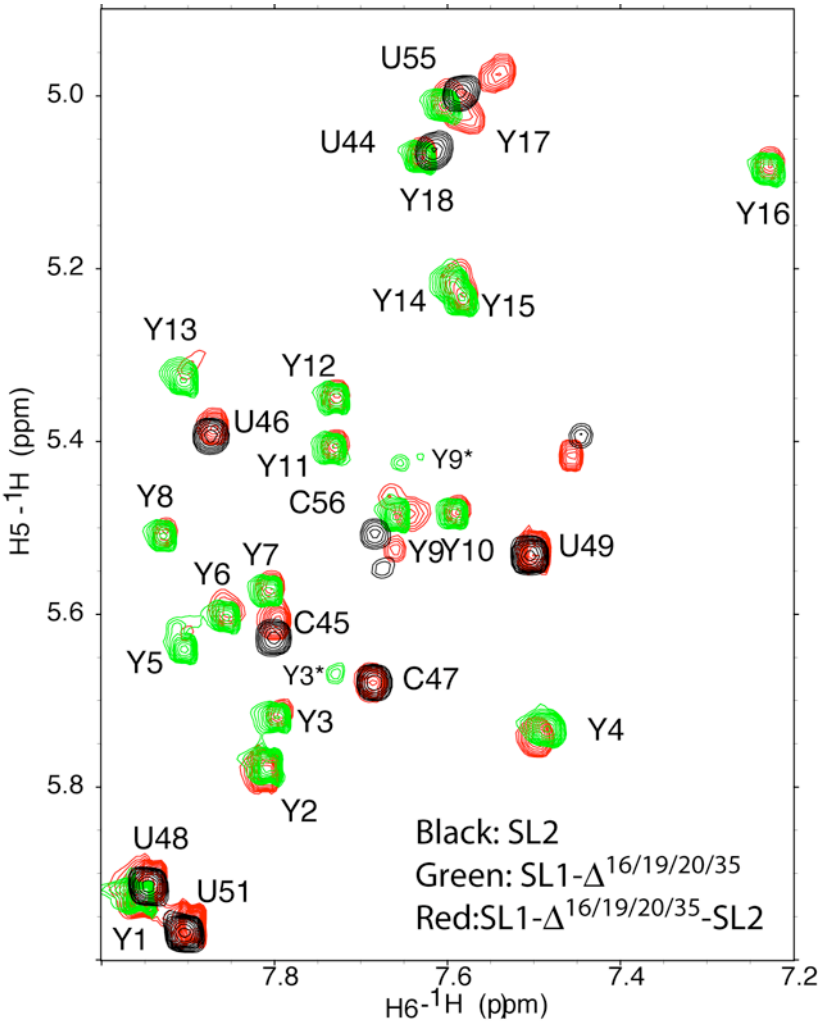


Figure 2-7 Superposition of the pyrimidine H5-H6 subregion of homonuclear TOCSY spectra acquired for SL1-$\Delta^{16/19/20/35}$ (green), SL2 (black) and SL1-$\Delta^{16/19/20/35}$–SL2 (red) RNAs. See Figure 2-4B for sequences for these RNA constructs. Conditions: 10 mM potassium phosphate, ph 6.0, at 25 ºC.

We also synthesized an RNA corresponding to HCoV-OC43 SL4, as well as the entire HCoV-OC43 140 nucleotides 5' UTR RNA. The 1D imino proton spectrum of HCoV-SL4 (Figure 2-8C) reveals that this RNA is not fully fold as predicted in the secondary structure model (Figure 2-1C). This may be with sufficient imino proton resonances to account for folding of the lower genetically confirmed SL4b (this spectrum was not assigned). Although the spectrum of the entire 5' UTR degrades substantially in overall quality (Figure 2-8D), the signature peaks in SL1 and SL4 appear to be observed on this spectrum. The signature peaks of SL2 are relatively difficult to be identified because of peak overlapping. In addition, the extra broad peak on ~ 14.6 ppm may indicate the folding of predicted SL3 with AU base pair rich stem in HCoV-OC43. In any case, although further investigation is required, these preliminary data would appear to indicate that under these solution conditions, little interactions exist between these stem loops since components of both SL1 and SL4 are clearly visible in the spectrum.

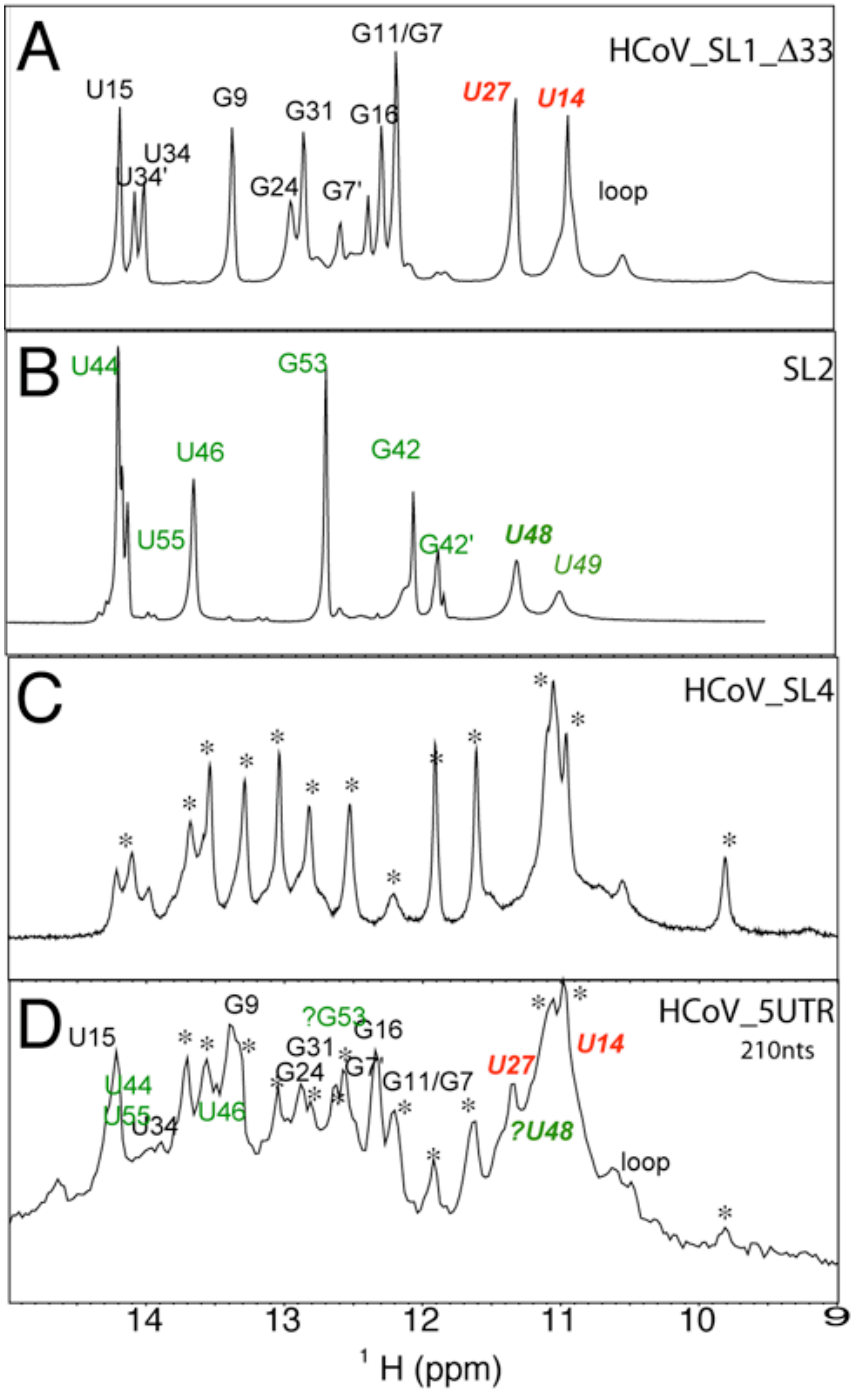


Figure 2-8 Imino proton regions of 1D jump-return echo spectra of (A) HCoV-SL1-Δ33, (B) SL2, (C) HCoV-SL4 and (D) HCoV-5' UTR (140 nucleotides) RNAs. See Figure 2-1C for sequences for these RNA constructs.

We next determined the extent to which the SL2 was amenable to substitution, since the U-turn hypothesis predicts that substitutions of U48 may be deleterious to the structure, in contrast to U49 (Gutell et al., 2000). 1D imino proton spectra are shown for WT (Figure 2-9A), U48G (Figure 2-9B), U48C (Figure 2-9C) and U49A (Figure 2-9D) SL2 RNAs acquired at low salt. We also show ^{1}H-^{15}N HSQC spectra that derive from ^{13}C,^{15}N-[U]-labeled WT SL2 (Figure 2-9A) and ^{13}C,^{15}N-[G]-labeled U48G (Figure 2-9B) RNAs. These spectra establish that the most intense upfield-shifted slowly exchanging imino proton corresponds to U48 and G48 in WT and U48G RNAs, respectively, and analysis of an ^{1h}J-HNN-COSY experiment acquired for these two RNAs suggest non-U or non-G nitrogen acceptors in the WT and U48G RNAs, respectively, as well (data not shown). These data, coupled with an analysis of the non-exchangeable NOE data support the contention that the G48 imino proton conserves the basic structure of the wild-type pentaloop, in a manner analogous to U48 in the WT sequence. The U48C loop structure may be disrupted due to the loss of U48 imino proton (Figure 2-9C). In contrast, as expected from the consensus UNR sequence, the U49A RNA appears to maintain a wild-type-like loop structure, given the identical chemical shift of the U48 H3 proton in both RNAs (Figure 2-9D).

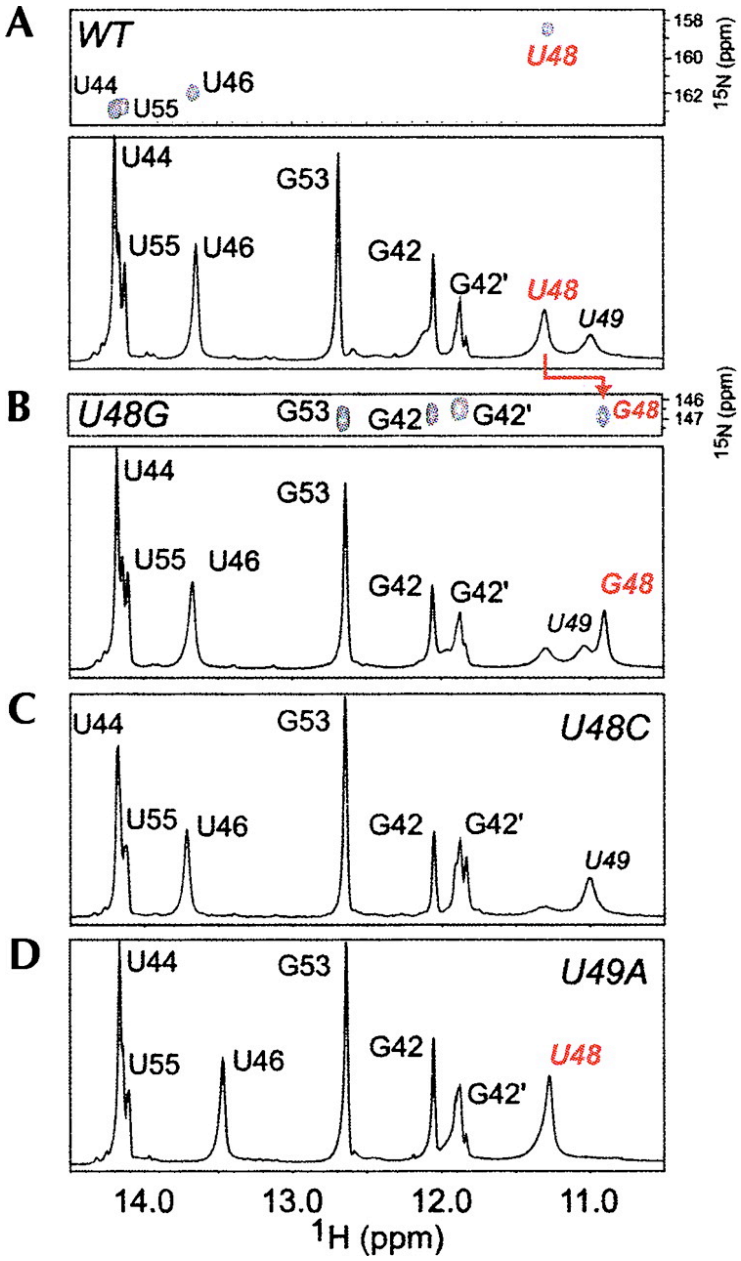


Figure 2-9 Imino proton regions of 1D jump-return echo spectra of SL2 variants. (A) WT SL2 (see Figure 2-4B); (B) U48G SL2; (C) U48C SL2; and (D) U49A SL2. ^{1}H-^{15}N-HSQC spectra for ^{13}C,^{15}N-[U]-labeled WT SL2 (A) and ^{13}C,^{15}N-[G]-labeled U48G SL2 (B) are shown. G42 and G42' represent alternative conformations for the terminal G42-C56 base pair. The immediately adjacent U55 resonance is also doubled, indicative of heterogeneity at the base of SL2.

We could find no evidence in either the WT or U48G SL2 RNAs in support of a C47•U51 base pair that would be expected for a canonical U-turn motif. In fact, the data suggest that U51 is extruded from the loop with G50 in close proximity to A52 (see Discussion); these data suggest a noncanonical U-turn conformation in CoV SL2. Although U-turn structures are capable of forming in the absence of divalent cations and at low salt (Puglisi and Puglisi, 1998), they can be stabilized by high concentrations of Mg^{2+} or $Co(NH_3)_6^{3+}$ (Cabello-Villegas et al., 2004). We find that the basic structural features of the pentaloop are unchanged in the presence of 0.1 M KCl, 5.0 mM $MgCl_2$, while the addition of mM $Co^{3+}(NH_3)_6$ leads to significant dimer formation (data not shown).

We are currently in the process of solving the solution structure of the SL2 RNA; however, careful inspection of the spectra seem to suggest that SL2 may be quite dynamic and/or composed of multiple loop conformations in rapid (>ms) exchange in the absence of an interacting partner (protein or RNA). Figure 2-10 shows the aromatic region of ^{1}H-^{13}C HSQC spectrum of ^{15}N,^{13}C-[U]-labeled WT SL2, in which a single $^{13}C6$-$^{1}H6$ crosspeak is observed for each of the six uridine. The spectrum reveals that the aromatic crosspeaks associated with the SL2 loop nucleotides (U48, U49, U51) are quite intense relative to those in helix, e.g. with U51, proposed to be extrahelical most intense. Since peak intensity is reporting on dynamics independent of and faster than macromolecular tumbling, this suggests the considerable flexibility or dynamics in SL2 loop structure under these conditions; this was also observed previously for VS ribozyme stem loop loop V U-turn-containing loop UGACU (Campbell and Legault,

2005). This severely complicates efforts to solve the structure, although these efforts are ongoing.

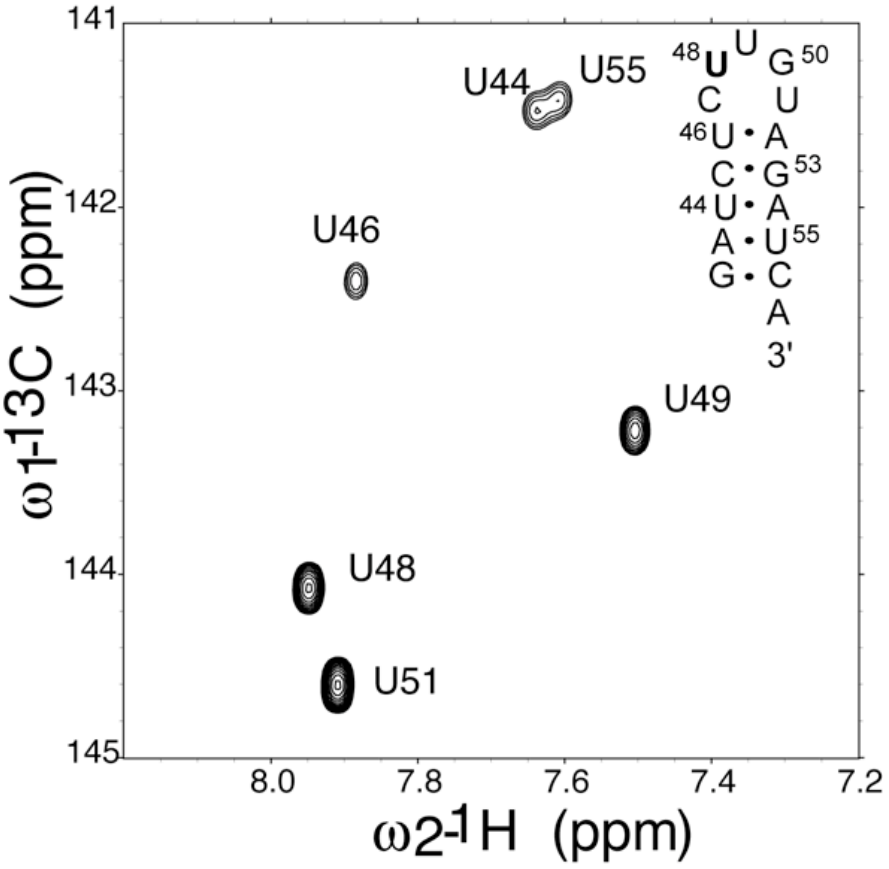


Figure 2-10 600 MHz ^{1}H-^{15}C-HSQC aromatic spectrum shown for ^{13}C,^{15}N-[U]-labeled WT SL2.

Mutations in SL2 affect RNA replication and in vitro translation

These structural studies were used to guide the design of mutant coronaviral primers which could then be functionally characterized in the Leibowitz's laboratory. The results show that MHV genomes containing point substitutions predicted to perturb

the SL2 structure of the pentaloop (U48C, U48A) are not viable, while those that maintain the structure (U48G and U49A) are viable. The U48C MHV mutant supports both positive and negative sense genome-size RNA synthesis, but fails to direct the synthesis of positive or negative sense subgenomic RNA (Liu et al., 2007).

We also investigated if mutations in SL2 affected translation, the MHV-A59 wild type 5'UTR and the U48C mutant 5'UTR were each cloned immediately upstream of the Renilla and firefly luciferase coding sequences fused to the MHV-A59 3'UTR. Capped RNAs corresponding to WT-5'UTR-Renilla luciferase-3'UTR (WT-Ren) and U48C-5'UTR-Renilla luciferase-3'UTR (U48C-Ren) were synthesized by in vitro transcription with T7 RNA polymerase and translated in reticulocyte lysates programmed with a subsaturating amount of RNA. The RNA containing the U48C 5'UTR was translated to yield approximately 19% of Renilla luciferase product as RNAs containing the wild type 5'UTR (Table 2-1). To be certain that this decrease in translation was not due to a different interaction of the mutant and wild type UTRs with the Renilla sequence, and to better control for small differences between individual in vitro translation reactions, we performed a series of ratiometric assays with *in vitro* translation reactions programmed with equal molar amounts of WT-Ren, U48C-Ren, WT-FF (WT-5'UTR-firefly luciferase-3'UTR), and U48C-FF (U48C-5'UTR-firefly luciferase-3'UTR) in the combinations shown in Table 2-1. These ratiometric assays confirmed that RNAs containing the wild type 5'UTR are translated in vitro ≈5-fold more efficiently than those containing a 5'UTR with a U48C mutation.

Table 2-1 Effect of the SL2 U48C mutation on *in vitro* translation.

Experiment	RNAs	RLU Firefly (FF) Luciferase	RLU Renilla (Ren) Luciferase	Relative Translational Efficiency U48C/WT
1	WT-Ren		17,400 ± 1000	
	U48C-Ren		3,400 ± 200	0.19 ± 0.08
2	WT-FF + WT-Ren	588,000 ± 22,000	10,600 ± 400	
	U48C-FF + U48C-Ren	89,000 ± 3000	2,700 ± 50	
	U48C-FF + WT-Ren	72,700 ± 2200	10,800 ± 300	0.12 ± 0.06*
	WT-FF + U48C-Ren	626,000 ± 15,000	3,000 ± 20	0.27 ± 0.05[#]

*Relative translational efficiencies calculated using the formula (U48C-FF/WT-Ren)/(WT-FF/WT-Ren)

[#]Relative translational efficiencies calculated using the formula (U48C-Ren/WT-FF)/(WT-Ren/WT-FF)

Discussion

In this chapter, we present a consensus RNA secondary structure model for the most 5' 140 nts of nine representative coronaviruses derived from all three major coronavirus groups. We also provide structural and functional support for the model, focusing here on the functional role that a pentaloop-containing SL2 plays in MHV replication. Although the SL1s of SARS-CoV and MHV have only 47.7% sequence identity, SL1 of SARS-CoV is capable of forming a stem loop structure of similar length to that of MHV SL1. Substitution of the SARS-CoV SL1 for the MHV SL1, which also increases the spacing between the predicted MHV SL1 and SL2 by 2 nucleotides, does not strongly affect viral viability but yields a virus with a smaller plaque size, impaired

RNA synthesis, and which replicated to lower titer than wild type MHV. Replacing the MHV SL4 with the SARS-CoV SL4 structure also generated a viable chimeric virus with a similar phenotype. SL2 is the most conserved secondary structure in coronavirus 5' UTR; replacing the SL2 of MHV with SARS SL2 resulted in a viable chimeric virus with a replication phenotype very similar to wild type MHV.

SL1 from MHV and BCoV/OC43 are uniquely characterized by two-three non-canonical U•U, U•C or A+•C base pairs in the middle of the stem (Figures. 2-2, 3, 4,). Their functional significance remains to be tested; however, they are obviously not absolutely required for replication since SARS-CoV SL1 can substitute, albeit weakly, for MHV SL1 (Kang et al., 2006). We point out, however, that the specific structural features of the upper two-thirds of BCoV/OC43 SL1 including the non-canonical base pairs and a predicted UNCG (YNMG; M=A or C) tetraloop stacked on a G•U wobble pair, are strikingly reminiscent of stem-loop D (SLD) in the 5' UTRs of picornoviruses which forms a binding site for the viral chymotrypsin-like protease ($3CL^{pro}$) (Du et al., 2004; Ihle et al., 2005; Ohlenschlager et al., 2004). A C•U Watson-Crick base pair in the triple pyrimidine mismatch widens the major groove of the stem by shortening the C1'-C1' distance across the helix (Ohlenschlager et al., 2004; Theimer et al., 2003). Our NMR experiments are consistent with the basic stem-loop structure, including the noncanonical pairing in the helical stem. However, it is clear that the anticipated 5'-GUGCGU tetraloop, where the closing base pair is 5' G•U, is destabilized with respect to the UUNCGG tetraloop in picornoviral SLD since characteristic imino proton resonances associated with the 3' guanosine in the loop and the closing G•U base pair are

absent or solvent-exchange broadened (Du et al., 2003; Ohlenschlager et al., 2004); this is consistent with previous thermodynamic studies that reveal that inversion of the 5'-C-G base pair to 5' G-C destabilizes a UUCG the tetraloop by 2.3 kcal mol^{-1} (Antao et al., 1991). Efforts are underway to solve the solution structure of SL1 (see Chapter III).

Our structural and functional data provide strong evidence in support of the formation of SL2. Ablation of the stem and introduction of compensatory mutations that restore SL2 base pairing argue strongly for the functional importance of SL2 (Liu et al., 2007). In addition, we also show that U48, which is predicted to stabilize the pentaloop conformation by virtue of formation of one or more hydrogen bonds, is a key determinant for SL2 function. This is in contrast to U49, where a nonconservative adenosine substitution is tolerated, in contrast to the U48A virus which, like the U48C virus, is nonviable. The precise structural basis for these findings is currently under investigation as is a more comprehensive mutagenesis analysis of the pentaloop; however, we note a 1:1 correspondence between presence of an hydrogen bond donor at position 48 (U48 H3 or G48 H1) with MHV viability. The mutagenesis and structural studies reported here coupled with the evolutionary conservation of the pentaloop (see Figure 2-1) suggests a novel, noncanonical U-turn motif conforming to the sequence 5'-YY(U/G)NR(Y)R, where a Y-R base pair closes a four-base Y(U/G)NR tetraloop with the (Y) nucleotide extruded from the structure (see Figure 2-11).

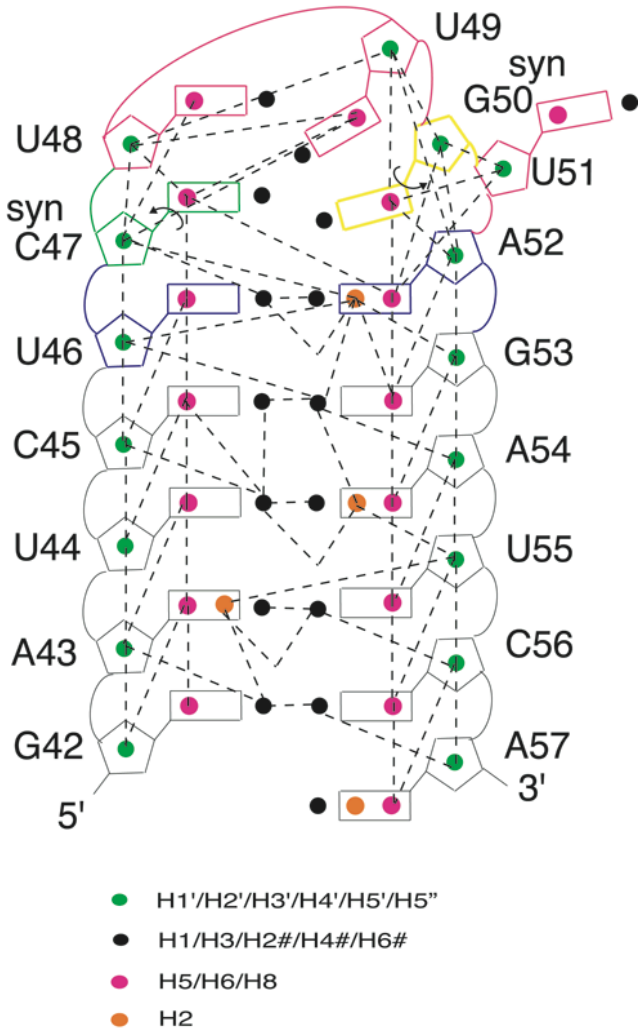


Figure 2-11 Summary diagram of the interresidue NOEs collected for SL2. Dashed lines represent one or more NOE(s) observed between two protons as defined in the legend.

Our secondary structural model leaves open the possibility that SL3 is single-stranded or folded into a weakly paired helical structure (see Figure 2-1C). In the latter case, the TRS sequence, which is absolutely required for template switching and discontinuous subgenomic RNA synthesis, would be found in the terminal loop of SL3 in BCoV/OC43, HKU1, and SARS-CoV (see Figure 2-1C). Studies of the structure of

the 5' UTR in a more distantly related member of the Nidovirus family, EAV (an arterivirus) suggest that the EAV leader TRS sequence is incorporated into the terminal loop of a stem-loop structure (stem-loop G), and this prediction is supported by enzymatic probing and functional data (Van Den Born et al., 2004; van den Born et al., 2005). The extent to which this is the case in any coronavirus has yet to be firmly established (see (Chang et al., 1996)); this is in contrast to sequence requirements of the both core and flanking nucleotides of the TRS, which have been extensively investigated (Sola et al., 2005; Zuniga et al., 2004).

Interestingly, the U48C substitution in SL2 also decreases the translational efficiency of heterologous RNAs bearing the MHV 5'UTR (Table 1-1). The precise mechanism by which SL2 influences translational efficiency is not clear. This could occur through interactions with proteins present in the *in vitro* translation reactions, by providing a structure which facilitates interactions between the 5' and 3'UTRs, favoring repetitive rounds of translation, or by promoting the folding of the UTR into a structure that is more easily traversed by ribosomes scanning for the initiating AUG. Our findings are consistent with earlier work implicating the leader 5'leader sequence containing SL2 in enhancing in vitro translational efficiencies of heterologous transcripts in lysates prepared from MHV infected cells (Tahara et al., 1994).

All mutations in SL2 affect virus-specific RNA synthesis, and similar findings characterize non-viable SL1 mutants as well (see Chapter III). Cells electroporated with non-viable mutants in SL2 all have defects in subgenomic RNA synthesis although genome replication occurs. These results are consistent with those observed by Wang

and Zhang in a DI RNA system (Wang and Zhang, 2000). It is currently unknown how SL2 would directly affect subgenomic (sg) RNA synthesis, but there are several possibilities that can be distinguished from one another on the basis of additional structural and functional studies. For example, it is often the case that U-turn motifs mediate long-range RNA-RNA interactions, and prominent examples of this occur within the tRNA anticodon loop-mRNA interaction, ribosomal RNA-RNA interactions (Gutell et al., 2000), and within the VS ribozyme, where a non-canonical U-turn motif base pairs with a non-adjacent hairpin loop in forming three consecutive loop-loop Watson-Crick base pairs (Campbell and Legault, 2005). even presenting the scissile phosphate to accomplish fast site-specific catalysis in the genomic HDV ribozyme (Sefcikova et al., 2007). This might be expected if, for example, the Watson-Crick edges of the bases within the loop are exposed to solvent. This will be more confirmed by our structure. Another possibility is that SL2 mediates a specific interaction with a viral-encoded or cellular protein. Previous studies suggest that polypyrimidine-tract binding protein (PTB) (hnRNP I) binds to the MHV leader to a region containing UCUAA repeats in the leader TRS where it is thought to function as a regulator of viral transcription (Choi et al., 2002; Li et al., 1999). Recent structural studies reveal the molecular basis by which PTB interacts strongly with short pyrimidine-rich (UCU, CUCU) sequences (Oberstrass et al., 2005), and these data suggest the possibility that that SL2 might provide a specific binding site for PTB. This interaction might play a role in mediating the circularization of the genome and/or directly facilitate template switching in discontinuous sgRNA synthesis.

CHAPTER III

STRUCTURAL LABILITY IN STEM-LOOP 1 DRIVES A 5' UTR-3' UTR INTERACTION IN CORONAVIRUS REPLICATION

Introduction

MHV transcription generates a set of nested subgenomic mRNAs, which contain 5'- and 3'-regions that are identical to those of the genomic RNA (Lai et al., 1984; Leibowitz et al., 1981; Spaan et al., 1983). The most widely accepted model to explain this unique discontinuous transcription mechanism posits that subgenomic negative strand RNAs are synthesized from genome-sized template, and that the subgenomic mRNAs are synthesized from subgenomic minus-strands (Baric and Yount, 2000; Sawicki and Sawicki, 1990; Sethna et al., 1989). Zuniga *et al.* have refined this model to take into account recent data suggesting the existence of RNA-RNA interactions between the 5' and 3' UTRs, and proteins bound to these UTRs (Zuniga et al., 2004). This 5'-3' communication has been shown to be required for efficient negative-strand RNA synthesis in *Poliovirus*, *Sindbis* virus and *Tombusvirus* (Barton et al., 2001; Frolov et al., 2001; Herold and Andino, 2001; White and Nagy, 2004).

We develop a consensus secondary structure model based on the sequences of nine coronaviruses, in which the 5' leader RNA, consisting of the 5' most 72 nucleotides of the 5' UTR in MHV, is characterized by two conserved stem-loop structures, denoted SL1 and SL2 (Chapter II). In this chapter, we extensively characterize the functional and structural properties of SL1. SL1 adopts a bipartite structure with the functional

boundary two consecutive pyrimidine-pyrimidine (pyr-pyr) base pairs in the middle of the SL1 helix (Chapter II). The upper region of the stem must be base paired to support viral replication to near wild-type levels. In striking contrast, characterization of mutant viruses in which an unpaired nucleotide, A35, is deleted, leads to the recovery of a diverse collection of mutant viruses that contain single-nucleotide second site mutations in the lower half of SL1, all of which are kinetically and thermodynamically destabilizing. In addition, these viruses contain one of two single nucleotide substitutions in the extreme 3' UTR of the genome. Nonviable mutants in this region of SL1, like SL2 mutants characterized in Chapter II, fail to support sgRNA synthesis. These data taken collectively suggest that the base of SL1 must adopt an optimized stability or kinetic lability to mediate a long-range interaction between the 5' and 3' UTRs that is an obligate step in the transcription of sgRNAs. The nature of this interaction is discussed.

Materials and methods

Preparation of RNA samples

RNAs were obtained by *in vitro* runoff transcription using SP6 RNA polymerase and purified by denaturing PAGE essentially as previously described (Nixon et al., 2002b). The NMR samples were subjected to exhaustive dialysis into a final buffer of 10 mM potassium phosphate (pH 6.0) and 100 μM DSS with RNA concentrations ranging from ≈1 to 2.5 mM in 300 μL (10% D_2O). The RNA samples for thermal denaturation experiments and calorimetry experiments were prepared by diluting into a

final dialysis buffer of 10 mM potassium phosphate (pH 6.0), 100 mM KCl, and 5mM $MgCl_2$. Before each experiment, the RNA samples were annealed by heating at 65^{o}C for 10 min, followed by slow cooling at room temperature. All samples were ≥90% monomeric as judged by nondenaturing PAGE.

Thermal denaturation experiments

RNA melts were collected on a Cary 1 scanning spectrophotometer operating in double beam mode. The RNA concentrations were between 1-20 μM, and all melting profiles were shown to be independent of RNA concentration over this range. The first derivative data of absorbance at 260 nm and 280 nm with respect to temperature (dA/dT) were subjected to a simultaneous non-linear least-squares fit of ΔH_i , $t_{m,i}$ and A_i, for each *i*th unfolding transition via sequential interacting two-state unfolding transition model using the t-melt program running on a Silicon Graphics O_2 workstation as previously described (Theimer and Giedroc, 1999). Melting profiles were subjected to single- or two-transition unfolding models as described in the text. Parallel thermal melts (50-100 μM RNA strand) were carried out on a Microcal VP-DSC scanning calorimeter under the same solution conditions as previously described (Theimer and Giedroc, 2000). Analysis of these baseline-corrected data with single-transition two-state unfolding model gave fitted parameters in qualitative agreement with those obtained from analysis of the optically monitored thermal unfolding experiments (data not shown).

Saturation transfer solvent exchange experiments

NMR spectra were acquired on a Varian Inova 500 MHz spectrometer at the

Biomolecular NMR Laboratory, Texas A&M University. The data were processed using NMRPipe (Delaglio et al., 1995) and analyzed using Sparky (Goddard and Kneller, 2001). The imino proton resonance were assigned by the jump-return echo 1D and Water-gate homonuclear ^{1}H-^{1}H NOESY spectra (τ_{mix}=300 ms), with reference to an internal standard DSS at 10^{o}C. The imino proton exchange rates were obtained by transfer of magnetization from water (Coman and Russu, 2004). The exchange was initiated by selectively inverting the water proton resonance using a Gaussian 180° pulse (5.3 ms duration) with exchange delay times ranging from 2 to 850 ms, followed by a weak gradient (0.1 G/cm) applied during the exchange delay to minimize effects due to radiation damping. At the end of the exchange delay, a second Gaussian pulse (2.8 ms) was applied to bring the water magnetization back to the z axis. Imino proton resonances were detected using a gradient-enhanced spin echo sequence. The acquired spectral array was processed using NUTS (Acorn NMR, Inc.) and the intensities of the imino proton resonances of interest were fitted to obtain the imino proton solvent exchange rate k_{ex} using Kaleidagraph (Synergy Software) as previously described (Cornish et al., 2006; Dhavan et al., 1999).

Results

SL1 folds in solution and adopts a functionally bipartite structure

A series of NMR studies of SL1-containing RNA constructs from both HCoV-OC43 and MHV document the key predicted structural features of SL1 in our model (Chapter II), the most notable of which is two consecutive pyr-pyr base pairs, U12•C32

and U13•U31 in the middle of the helix, capped on either side by Watson-Crick A11-U33 and G14-C30 base pairs (Figure 3-1A). In addition, these NMR studies also suggested that either A35 or A36 is an unpaired extrahelical nucleotide and that the alternative pairing of U9 with A35 or A36 introduces a local destabilization in this region of SL1 stem (Chapter II). Interestingly, the predicted SL1s in other coronaviral genomes also appear to conserve this general features of a rather weakly paired region at the base or middle of SL1, containing either or both non-canonical base pairs and extrahelical nucleotides, suggesting that this might be important for replication of MHV and perhaps other coronaviruses. This served as the motivation for the construction of a mutant MHV genome harboring a single nucleotide deletion of A35 (SL1-ΔA35).

Motivation behind the choice of MHV SL1 mutant constructs selected for structure studies

Since NMR data supported the existence of the SL1 stem-loop in 5' leader-containing RNA fragments (Chapter II), we collaborate with Dr. Leibowitz's laboratory, first investigated the functional importance of SL1 in viral replication. One set of mutations targeted the upper portion of SL1 above the pyr-pyr base pairs, while a second set targeted the lower region of SL1. C residues of the three cytidines unique to MHV (C16, C19 and C20, with C16 predicted to be extrahelical) are dispensable for viral replication justifying our use of the wild-type model (denoted WT*) in our physical studies presented below (see Figure 3-1B). Deletion of these cytidines results in three consecutive G-C base pairs in the upper segment of the stem-loop (Figure 3-1B). The

upper portion base pairing of SL1 was shown to be required for efficient virus replication.

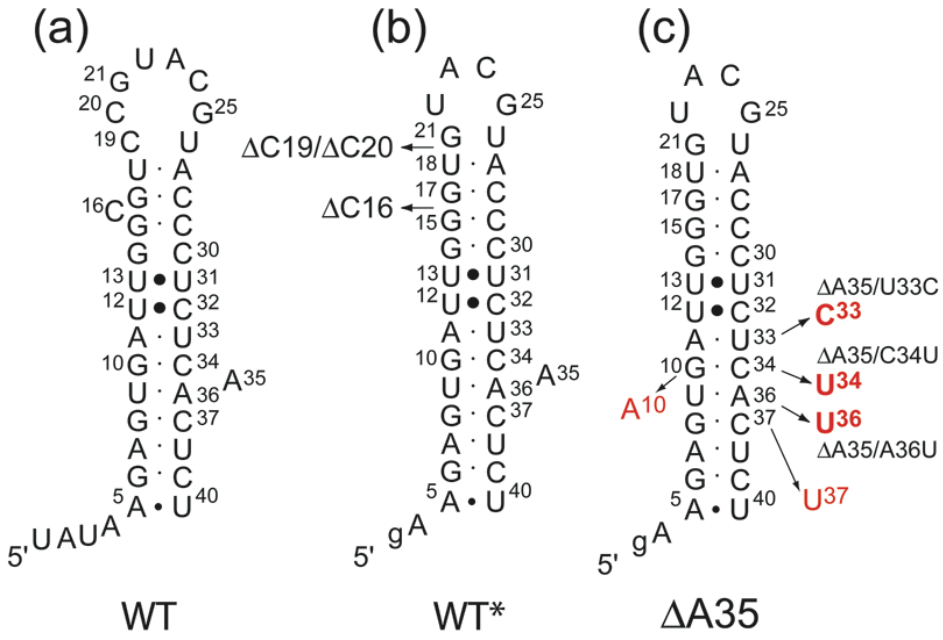


Figure 3-1 Predicted secondary structure model of MHV A59 SL1 constructs**.** (A) WT (Chapter II); (B) WT*, which corresponds to SL1-Δ(C16/C19/C20) with a 5'-GA overhang required for initiation of *in vitro* transcription by SP6 polymerase; and (C) Second-site mutant SL1s derived from infection with the SL1-ΔA35 virus (ΔA35 RNA sequence shown; recovered single-nucleotide substitutions highlighted in *red*). The constructs labeled WT*, ΔA35, ΔA35/U33C, ΔA35/C34U, and ΔA35/A36U RNAs (panels B and C) were used for the NMR and thermal unfolding studies. The 5' g is a non-native nucleotide, shown in lower case.

In contrast, the mutations introduced below the two consecutive pyr-pyr base pairs gave rise to completely opposite results. A-U to G-C base pair substitutions that are predicted to stabilize the lower region of SL1 is either nonviable (A5G/U40C) or

unstable (A7G/U38C). In the case of A5G/U40C, component single and double nucleotide substitutions that destroy the original base pair (A5G, U40C, or A5C/U40C) are all viable. Strikingly, the A7G/U38C virus is a true revertant, with the A7G substitution repaired in the first passage, and the U38C substitution repaired next to generate the original A7-U38 base pair. These information taken collectively suggest that the lower portion of SL1 is weakly base paired or not at all (*vide infra*), and that this region of the stem plays a critical role in viral replication.

To further investigate the functional importance of the lower SL1 region, the Leibowitz's group characterized mutant viruses containing a deletion of A35 (SL1-ΔA35). From two completely independent transfections 12 plaque purified viruses were recovered. Sequencing of the complete 5' and 3' UTRs of several SL1-ΔA35 plaque-purified viruses revealed only two of these viruses contained the SL1-ΔA35 sequence; the remaining ten viruses made up a diverse collection of single-nucleotide second-site substitutions in the immediate vicinity of the deleted nucleotide. These include C37U, A36U, C34U and A33C (from the base of SL1 toward the hairpin loop; see Figure 3-1C) in the context of the parent ΔA35 deletion. Additionally, viruses containing the SL1-ΔA35 sequence were completely replaced by viruses containing a second site mutation (G10A) in SL1 after only one passage in DBT cells, confirming that viruses containing only a deletion of A35 were genetically unstable and were subjected to strong negative selection. Remarkably, although diverse in sequence, all second site mutants possess a common property: each is predicted to change a Watson-Crick base pair in the wild-type SL1 to a non-canonical base pair (*vide infra*).

In addition to these substitutions in the 5'-UTR, all recovered viruses contained one of two second-site single nucleotide substitutions in the 3' UTR, 3'-A29G or 3'-A78G (3', counting from the 3' nucleotide of the 3' UTR in the genomic strand, before the poly(A) tail, near the very 3' end of the genome some ≈30 kb distant. We note that the latter mutation is within the really invariable octamer motif ^{80}GGAAGAGC^{73} discussed previously (Chapter 1). In one case (SL1-ΔA35/C34U) the passage 1 virus contained a mixture of the 3'-A29G and wild type sequences. The wild type sequence was lost after one additional passage. Both A29G and A78G mutations in the 3' UTR share the same sequence context, GAGA, relative to the same wild-type context GAAG. In addition, since in at least one case, C37U, both A29G and A78G substitutions were recovered; this is consistent with the idea that each mutation, which co-evolves with the 5' UTR mutation, is functionally equivalent. These data provide strong genetic evidence in support of a physical interaction between 5' UTR and 3' UTRs in MHV that is mediated by a "destabilized" or "dynamic" SL1.

A low thermostability of SL1 is crucial for virus stability and viability

The functional data presented above are consistent with what we term a "dynamic SL1" hypothesis, which posits that the lower region of SL1 must be thermodynamically destabilized and/or dynamically (kinetically) labile in a way that is dependent only on the general physical features of this region of the SL1, rather than the precise nucleotide sequence, in order to fully support virus replication. In order to test this hypothesis and gain additional insight into SL1 structure, we synthesized three representative second-site SL1 revertant mutant RNAs (see Figure 3-1C) and measured

their thermodynamic stabilities using quantitative optically-monitored thermal denaturation methods (Figure 3-2), and their structural and dynamic properties by NMR spectroscopy.

Optical melting profiles (dA/dT) are shown for the WT*, ΔA35, ΔA35/U33C, ΔA35/C34U, ΔA35/A36U SL1 RNAs in Figure 3-2 with the thermodynamic parameters described from a quantitative analysis of these melts compiled in Table 3-1. In contrast to all of the other RNAs, the optical melting profile of the ΔA35 SL1 RNA is well-modeled by a single two-state unfolding transition with a t_m of 80.3 ^oC with an van't Hoff enthalpy of unfolding (ΔH_{vH}) of 118 kcal mol^{-1} (Table 3-1). Thermodynamic parameters derived from melting curves acquired using differential scanning calorimetry (DSC) are in excellent agreement with the optical melts, as well as the expected ΔH estimated from the nearest neighbor model (Xia et al., 1998). In contrast, melting profiles obtained for the WT* RNA as well as the three ΔA35 second site revertant mutant RNAs are broader than that obtained for the ΔA35 RNA; they also unfold at significantly lower t_m. A broadened unfolding transition is consistent with a superposition of two sequential two-state unfolding transitions, *e.g.,* the lower and upper regions of the SL1 stem, or extreme non-two state unfolding behavior of the SL1 stem in a single transition. In both cases, fitting of these melting profiles to a single two-state transition would result in insufficient van't Hoff enthalpy of unfolding (ΔH_{vH}) to account for the unfolding of the entire 14-base pair stem (Figure 3-2). This is exactly what we observe, with the apparent ΔH_{vH} of ≈70-75 kcal mol^{-1} in each case (fits not shown). Invoking a sequential two-state unfolding transition model reveals two

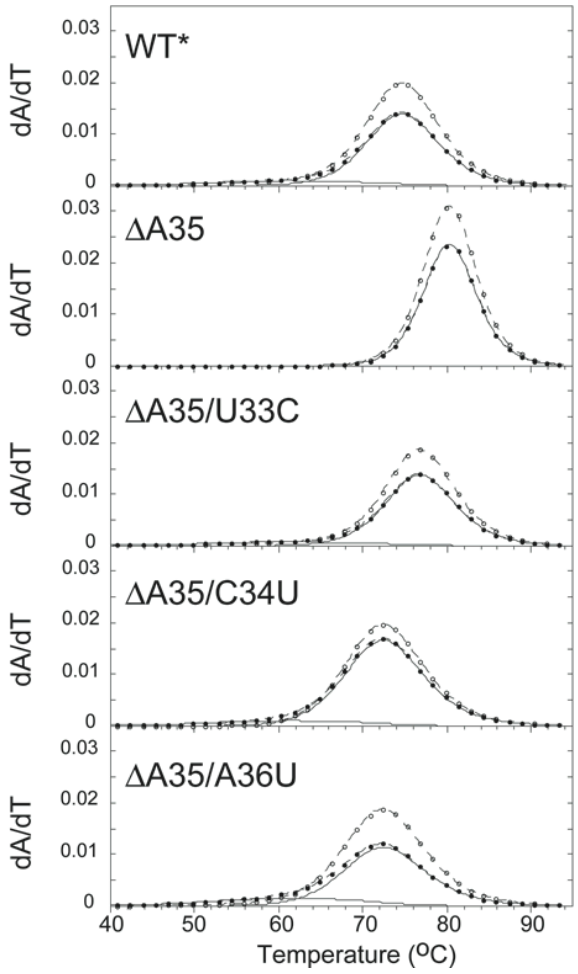


Figure 3-2 Comparison of the thermal unfolding of the WT*, ΔA35, ΔA35/U33C, ΔA35/C34U, ΔA35/A36U SL1 RNAs. The experimental optical melting profiles show every fifth data point collected at 260 nm (•) and 280 nm (o) with the calculated fits (*dashed lines*) shown. For the ΔA35 RNA, the nonlinear least squares simultaneous composite fit to a single transition unfolding model, and the transition is shown in *blue* solid line. For the WT* model and three recovered SL1 mutant ΔA35/U33C, ΔA35/C34U, ΔA35/A36U RNAs, a nonlinear least squares simultaneous composite fit to a two-transition unfolding model and component transition 1 and transition 2 (*solid lines*) are shown. The thermodynamic parameters derived from these fits are compiled in Table 3-1.

Table 3-1 Thermodynamic parameters derived for the unfolding of WT*, ΔA35, ΔA35/U33C, ΔA35/C34U, and ΔA35/A36U SL1 RNAs[a].

	Transition 1[b]		Transition 2			
RNA	ΔH_1	t_{mm1}	ΔH_2	t_{m2}	ΔG^o_{37}	$\Delta\Delta G^o_{37}$
WT*	37.9	60.8	81.0	74.5	-11.5	3.0
ΔA35	118.3	80.3	–	–	-14.5	–
ΔA35/U33C	37.9	60.8	85.7	76.7	-12.4	2.1
ΔA35/C34U	37.9	60.8	72.0	72.2	-10.0	4.5
ΔA35/A36U	37.9	60.8	69.8	72.2	-9.8	4.7

[a]Enthalpy and free energy are reported in kcal mol^{-1} and t_m is reported in ^oC. Thermodynamic parameters derived from the optical melting profiles determined using the two-state van't Hoff unfolding model with ΔC_p=0 (Theimer and Giedroc, 1999). [b]Parameters for the transition 1 were fixed to the predicted values for the unfolding of the bottom four base pairs of SL1, with the transition 2 parameters optimized during the fit (see text for details). The $\Delta G_{37,i}$ for the *i*th individual transition was obtained from $\Delta G^o_{37}=\Delta H - 310.15*\Delta S$ where $\Delta S = \Delta H/t_m$. $\Delta G^o37 = \Sigma\Delta G_{37}{}^i$ and $\Delta\Delta G^o_{37}$ is expressed relative to the ΔG^o_{37} measured for the ΔA35 RNA.

unfolding steps with different amplitudes. If we assign the first unfolding step to the denaturation of the lower four base pairs in SL1 (predicted ΔH_{vH}=38 kcal mol^{-1}; t_m =61 ºC) (Hofacker, 2003) then we recover sufficient ΔH_{vH} in the second (major) transition (70-86 kcal mol^{-1}) to account for unfolding of the entire molecule. From these fits we calculate that the wild-type RNA is destabilized by 3.0 kcal mol^{-1} relative the ΔA35 mutant, while the three second site revertants are destabilized by 2.1 (ΔA35/U33C), 4.5 (ΔA35/C34U) and 4.7 (ΔA35/A36U) kcal mol^{-1}. These data reveal that the each of the recovered revertant RNAs share a common physical property with the wild-type RNA,

i.e., all are thermodynamically destabilized relative to ΔA35 mutant.

Mutant SL1 RNAs experience enhanced conformational heterogeneity

Since thermodynamic destabilization is a global property of a molecule, we used NMR spectroscopy in an effort to localize changes in the structure and/or dynamics that could account for this. 1D imino proton spectra (10 ºC, pH 6.0) for all five RNAs are shown in Figure 3-3. What is immediately apparent for these RNAs is that the anticipated noncanonical base pairings, *i.e.,* G10•U34 in the ΔA35/C34U RNA (Figure 3-3d) and U9•U36 in ΔA35/A36U (Figure 3-3e) are present since the imino protons of U34 and U36 are strongly protected from exchange with solvent in each case (Du et al., 2004; Ohlenschlager et al., 2004; Theimer et al., 2003). Although we did not directly determine whether the A33U substitution resulted in a new A•C base pair, this is expected since this substitution would give rise to the exactly the same three noncanonical pairings that characterize SL1 of a related group 2 CoV HCoV-OC43; in that RNA, a protonated A+•C base pair is formed (Chapter II). In addition, the finding that this RNA is the least destabilized relative to the ΔA35 mutant is not compatible with the presence of extrahelical nucleotides in each of these positions. Thus, all three revertant RNAs are characterized by multiple noncanonical pairings in an otherwise perfectly base-paired helical stem.

Inspection of these spectra (Figure 3-3) reveals that thermodynamic destabilization manifests itself in multiple complex ways. First, the WT RNA imino proton spectrum reveals a superposition of at least two conformations in slow exchange on the ^{1}H NMR timescale as evidenced by peak doubling of the U40, G6 and U38 imino

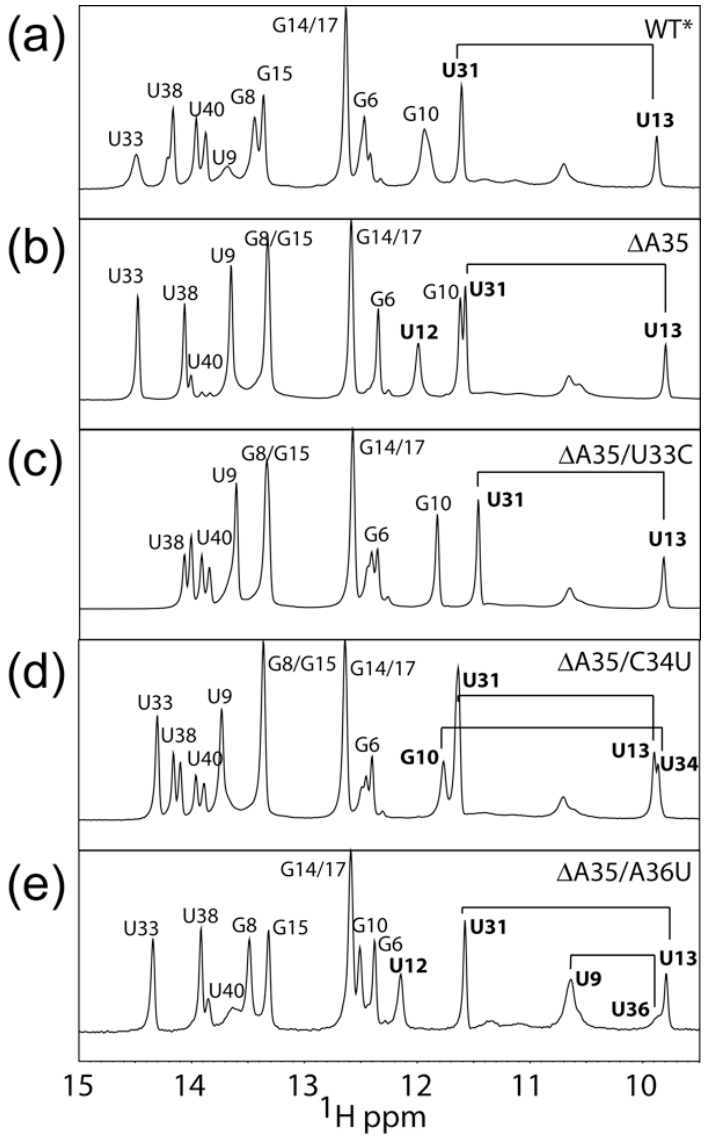


Figure 3-3 Imino proton regions of 1D jump-return echo spectra of WT* (a), ΔA35 (b), ΔA35/U33C (c), ΔA35/C34U (d) and ΔA35/A36U (e) SL1 RNAs. Imino protons corresponding to non-canonical base pairs are shown in bold. Note that some spectra (WT*, ΔA35/U33C, ΔA35/C34U) are characterized by slow conformational heterogeneity at the base portion of SL1 (A5-U40, G6-C39, and A7-U38 base pairs).

resonances at the base of the stem. Exactly the same type of heterogeneity is observed in the ΔA35/U33C and ΔA35/C34U RNAs but to a far greater degree, with the ΔA35/A36U RNA more like the WT* RNA in this regard. All heterogeneity is lost in the ΔA35 parent RNA.

Mutant SL1 RNAs experience enhanced dynamics

The imino proton solvent exchange rates, k_{ex}, reveal additional insight into SL1 dynamics. Although k_{ex} is a complex function of the rate constants for the opening (k_{op}) and closing (k_{cl}) of the base pair (where the two-state equilibrium constant for base pair opening, K_{op}, is defined by k_{op}/k_{cl}) as well as the intrinsic rate constant for proton exchange by base catalyst (k_{tr}), the magnitude of k_{ex} often tracks with the magnitude of k_{op} and K_{op} (Cornish et al., 2006; Dhavan et al., 1999; Lee and Pardi, 2007), since exchange will occur only very slowly from the base paired state below room temperature and at lower pH condition (Gueron and Leroy, 1995; Snoussi and Leroy, 2001). These data reveal that the WT* RNA is dynamically asymmetric, with imino protons above the U13-U31 base pair characterized by slow k_{ex} (≤ 2 s^{-1}), while those below this base pair are ≈2-50-fold faster. This is also confirmed by the Watergate NOESY experiment, in which the crosspeaks of G8, U9, U33 and G10 are greatly are broadened due to significant exchange with solvent during the NOE mixing time (300ms) (see Figure 3-5). For the ΔA35 RNA, imino proton solvent exchange rates are globally quenched, but

most strongly near the site of the deletion. For example, the k_{ex} values of U9, G10 and U31 are strongly attenuated in the ΔA35 RNA, by ≈10-, ≈32-, and ≈5-fold, respectively (Figure 3-4). Inspection of the base pair dynamics for the three mutant SL1 RNAs reveals that each re-introduces or accentuates one or more specific features of the kinetic lability that characterizes the WT* RNA. For example, in all three RNAs, the imino proton solvent exchange rates of the U13•U31 base pair are elevated relative to the ΔA35 RNA, by 4-10-fold, some more so than in the WT* RNA. This perturbation is local in the ΔA35/U33C RNA, but next-nearest neighbor and long-range in the ΔA35/C34U and ΔA35/A36U RNAs, respectively. The solvent exchange rate of G10 is also increased in all RNAs, most prominently in the ΔA35/C34U RNA in the context of the new G10•U34 base pairing, but also in the other two revertant RNAs (by 3-4 fold). These data argue that the base of SL1 through the pyrimidine base pairing region in both the WT* and second site mutant SL1s recovered from the genetically unstable ΔA35 virus must be conformationally heterogeneous and dynamically unstable; this, in turn, allows this region to become transiently unfolded so that a long-range interaction with the 3' UTR in MHV can occur.

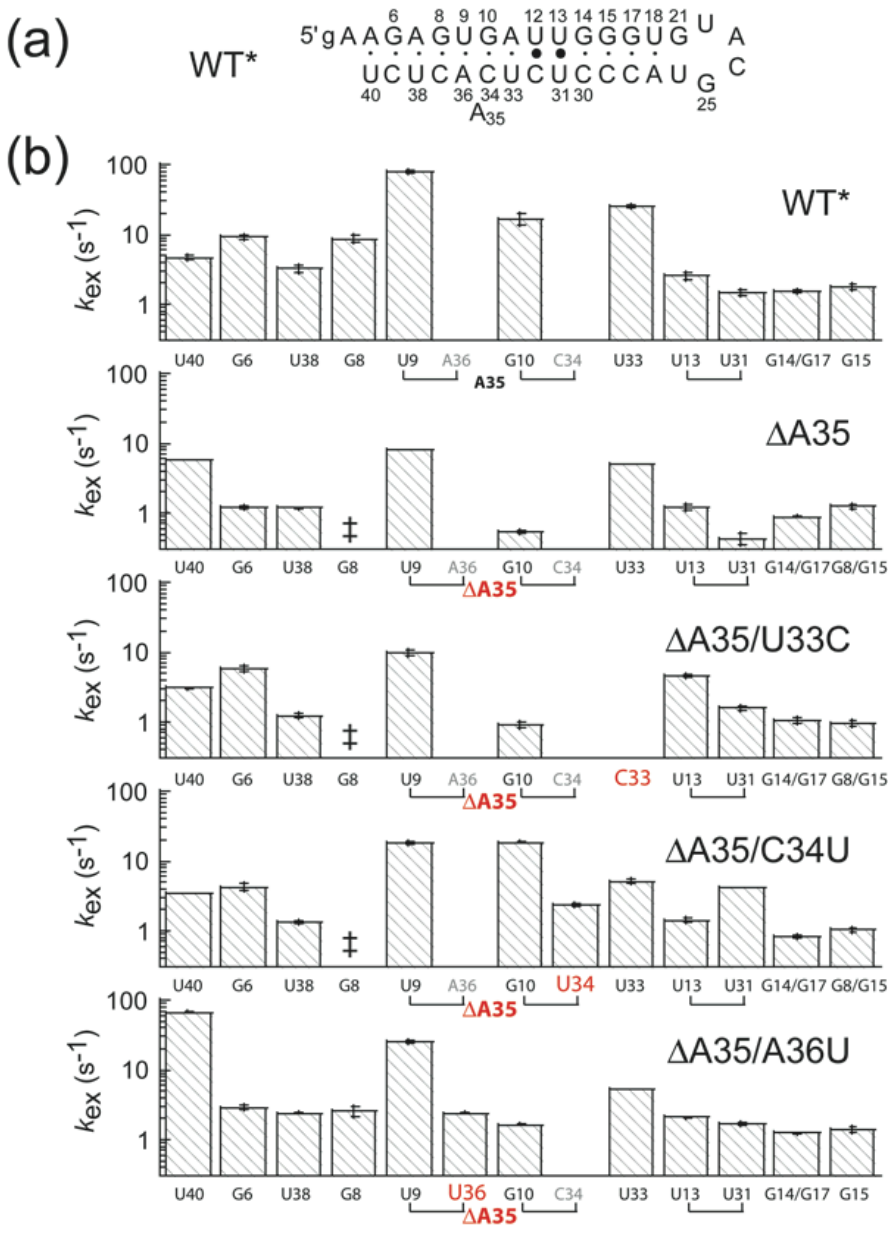


Figure 3-4 Graphical representation of the imino proton solvent exchange rates (k_{ex}) for the WT* and mutant SL1 RNAs. **(A)** Secondary structure of the WT* RNA; **(B)** k_{ex} is plotted vs. base pair position (from bottom to the top of the SL1 helix, from *left* to *right*). The mutations are shaded in *red*. k_{ex} could not be unambiguously measured for the G8 for the middle three RNAs due to spectral overlap with G15; the same is true of G14 and G17 imino protons in all spectra. In these cases, the average k_{ex} is plotted.

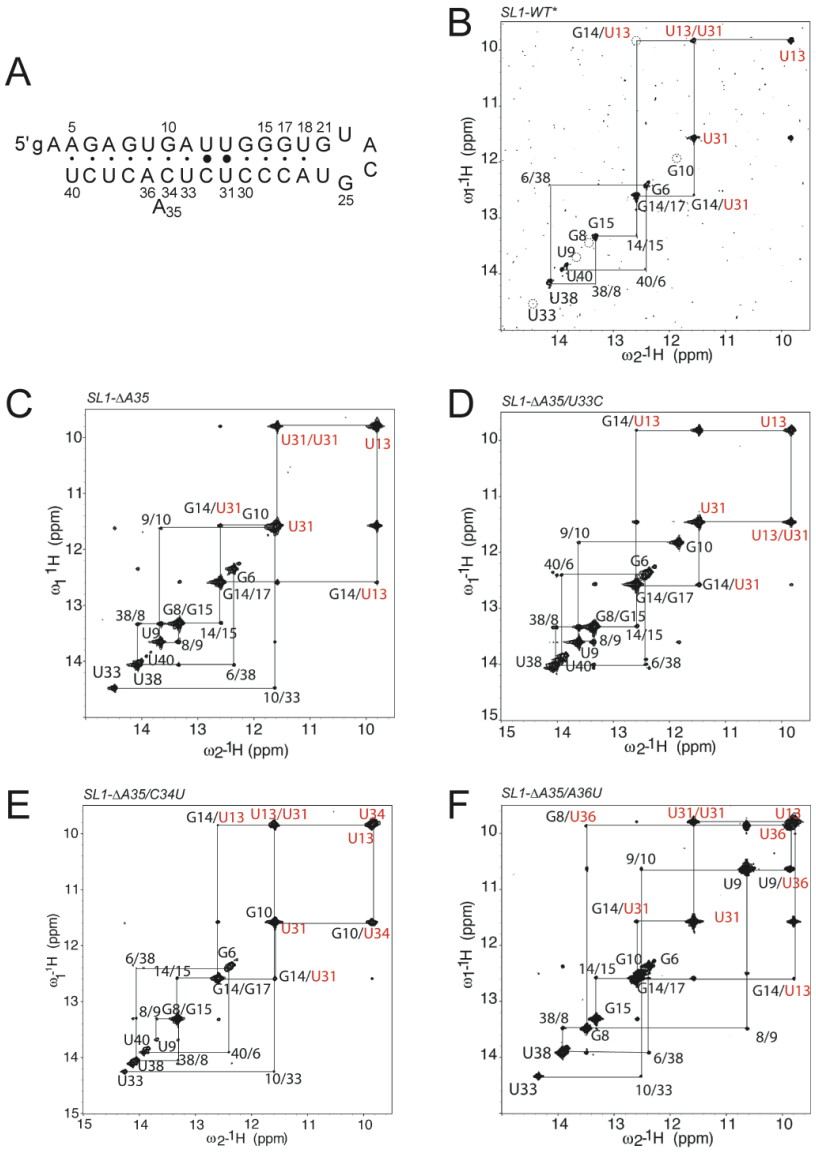


Figure 3-5 Imino proton NOE walks to obtain sequential resonance assignments for the RNAs used in this study. **(A)** Sequence and secondary structure of the WT* RNA; **(B)-(F)** Imino proton region of a 300 ms homonuclear NOESY spectrum (pH 6.0, 10 ºC) obtained each of five RNAs as indicated. Note that for the wild-type model RNA (SL1-WT*) (panel B), the imino protons in the immediate vicinity of A35, *e.g.,* G8, U9, G10 and U33, are broadened due to significant exchange with solvent during the NOE mixing time.

Discussion

The molecular mechanisms by which coronaviruses carry out sgRNA synthesis and ultimately coordinate this process with replication of the genomic RNA and translation of the genome remains poorly understood. An early event in these processes, however, is likely to be genome circularization which places the 5' and 3' termini of the viral genome in close physical proximity to allow for template switching during sgRNA "-" strand synthesis. This process is almost certainly regulated or facilitated by highly specific RNA structural motifs found in the 5' and 3' untranslated regions (UTRs) and associated interactions with virally encoded and/or host proteins. We have proposed that the 5' leader RNA appended to all sgRNAs and corresponding to the extreme 5' end of the 5' UTR, is composed of two stem-loop structures termed SL1 and SL2 just upstream of the transcriptional regulatory sequence (TRS) that defines the leader-body junction (Kang et al., 2006). Characterization of non-viable viruses that harbor mutations in SL1 (this Chapter) or SL2 (Chapter II) reveals that these genomes are generally wild-type with respect to synthesis of full-length genomic and anti-genomic RNAs; however, all are absolutely impaired in sgRNA synthesis.

A recent model for coronavirus replication (Zuniga et al., 2004) postulated that circularization of the genome is a necessary early step in sgRNA synthesis. However, there has been little direct evidence in support of or against a physical association of the 5' and 3' ends of the genome. During the course of testing the functional impact of mutations deposited in the MHV SL1, we have uncovered two critical aspects of SL1 that are required for sgRNA synthesis. Here, we present the first genetic evidence in

support of a direct interaction between SL1 and the extreme 3' end of the genome; furthermore, the physical analysis of RNAs harboring second site revertants in SL1 recovered from the genetically unstable SL1-Δ35 viruses, allows us to pinpoint clear structural requirements for the lower region of SL1 (below the pyr-pyr base pairs mapped by NMR spectroscopy) that are essential for supporting viral replication.

The base paring of the upper portion of the stem is required for optimal viral replication, while the lower portion is weakly base paired or even unpaired. Genomes containing A5G/U40C or A7G/U38C mutations that stabilize the extreme lower portion of the stem could not be recovered as viable viruses. While deletion of A35 is still viable, this virus is genetically unstable, and gives rise to second site mutations in both the 5' and 3'UTRs. Interestingly, all the second site mutations in the 5'UTR are located below the pyr-pyr base pairs, and introduce noncanonical base pairs to maintain stem formation. Each of these mutants SL1 stem structures are less thermodynamically stable compared with the genetically unstable ΔA35 mutant which forms a fully base-paired helical stem. These mutations also introduce conformational heterogeneity in this region as well. Interestingly, the MHV ΔA35 SL1 revertant mutants bear some resemblance to the predicted secondary structure of a wild-type SL1 from a related group 2 CoV, HKU1

(Kang et al., 2006; Woo et al., 2005). HKU1 SL1 does not contain an extrahelical nucleotide but instead is characterized by tandem predicted A•C and G•A mismatches in precisely the same region where 5' UTR SL1 second site mutations map in MHV.

In addition to the global thermodynamic stability information, the measurement of imino proton exchange gives important insights into the differences in dynamics or flexibility of individual base pairs in different ΔA35 RNAs. The k_{ex} results show that the lower half of the SL1 is kinetically labile, and the three second site revertant mutants recover one or more dynamic characteristics of the WT* RNA relative to ΔA35 mutant. Indeed, the closer the site of mutation localizes to A35, the more similar the dynamic properties become relative to the WT* RNA. Therefore, a less stable and more flexible SL1 might facilitate a specific interaction between the single-stranded RNA in this region and a host- or virally encoded protein(s) that is crucial for genome circularization and replication. A cartoon model that graphically illustrates this idea is shown in Figure 3-6, where the structure, conformational ensemble or longer-lived partially opened form(s) of SL1, schematized by A1' relative to fully base paired A structure, lowers the energy barrier for formation of a complex with hypothetical protein B, which in turn drives genome circularization and subgenomic RNA synthesis.

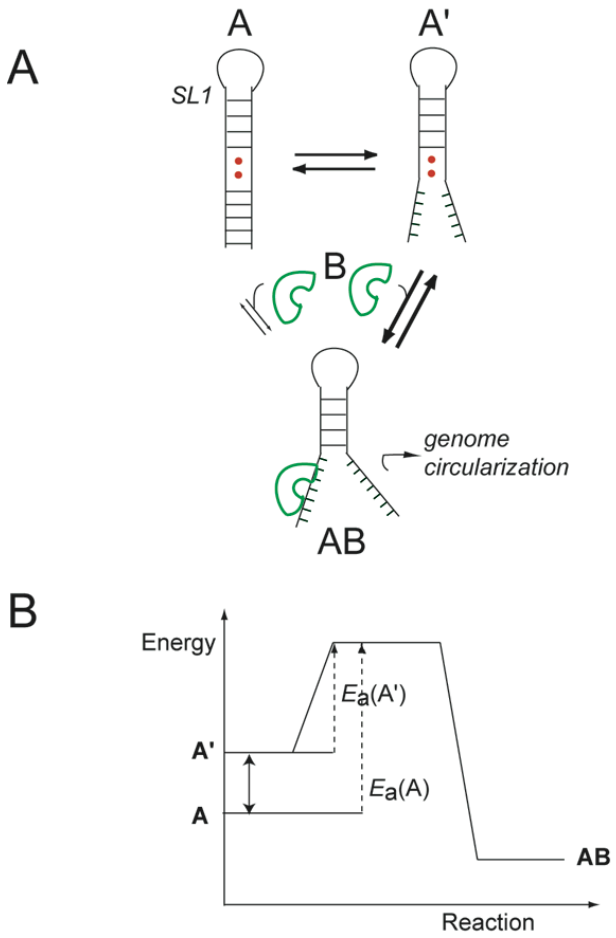


Figure 3-6 Model of a dynamic SL1. This model is consistent with the functional and structural data presented here. The fully based paired SL1 (A, modeled by the ΔA35 RNA) exists in equilibrium with one or more higher energy conformers (A', WT* and ΔA35 second site revertants) that are partially unfolded or experience dynamic destabilization as a result of non-canonical pairing. A hypothetical protein (B) binds to both A and A' to form the same partially unwound AB complex, but the affinity of B for A' will be higher since the full energetic cost of unfolding the lower stem will not have to be paid (energy barrier shown in B); this interaction then mediates a long distance RNA-RNA, RNA-protein or protein-protein interaction which is crucial for the viral replication.

Genetic interaction between the 5' and 3' UTRs

We note that four of five SL1 second site mutations recovered from ΔA35 viruses (U33C, C34U, A36U, C37U) map specifically to the 3' strand of SL1, the lone exception being G10A; all recovered viruses therefore maintain the 5'-^{6}GAGYR^{10} sequence in 5' portion of SL1 in MHV. This short sequence motif is conserved in all group 2 CoV genomes with the exception of SARS-CoV, which has an 5'-AGGU sequence in what is predicted to be a weakly paired region of SL1 with the A extrahelical (Kang et al., 2006). Strikingly, this SL1 5'-^{6}GAG^{9} sequence is identical to the 5'-GAG sequence recovered from second site mutations co-deposited in 3' UTR in the same viruses; this suggests the possibility that an oligomer of the same protein(s) capable of recognizing this short purine-rich motif may well play an important role in mediating a physical interaction between 5' and 3' UTRs. An excellent candidate for this role is hnRNAP A1 since both 3' UTR mutations occur in exactly the same sequence context, 5'-GAAG, a near-consensus binding site for A1, and each replaces the second A with a G, a consensus hnRNAP A1 binding site (Ding et al., 1999; Myers and Shamoo, 2004), with the 5'-AG the key specificity determinant. hnRNAP A1 possesses RNA chaperone activity and has been shown to bind the 3' UTR in MHV, while hnRNP A1 as well as PTB both bind to the complementary strands at the 5' end of MHV RNA. These proteins together mediate the formation of an RNP complex involving the 5'- and 3'-end fragments of MHV RNA *in vitro* (Huang and Lai, 2001).

Our data are compatible with the three-step model of coronavirus transcription recently refined by Zuniga *et al* (Zuniga et al., 2004), in which the first step is the formation of 5'-3' UTR form a complex through protein-RNA and protein-protein interactions, by which the TRS-L would be located in close proximity to sequences located at the 3' end of genomic RNA, and in turn, make template jumping possible. Overall, we demonstrate that SL1 must possess an optimum stability or kinetic lability required to mediate a key long-range physical interaction between the 5' and 3' UTR that is critical for subgenomic RNA synthesis. Such a functionally bipartite structure where a sub-region of a helical stem below tandem wobble pairs is tuned to an optimum stability is reminiscent of stem I of the U2 snRNA; here, these features have been hypothesized to allow U2 to adopt a number of mutually exclusive folded conformations during spliceosome assembly and catalysis (Sashital et al., 2007).

CHAPTER IV

THE RNA BINDING PROPERTIES OF MHV NUCLEOCAPSID PROTEIN

Introduction

RNA transcription in coronaviruses results in the formation of a nested 5'- and 3'-coterminal set of 6-7 sgRNAs of various sizes (Pasternak et al., 2006). sgRNA synthesis is a complex process involving discontinuous transcription. According to the well accepted minus-sense discontinuous transcription model (Sawicki and Sawicki, 1998; Zuniga et al., 2004), the synthesis of sgRNAs is mediated by a discontinuous synthesis of minus-strand sgRNA. In this process, the TRSs found in the body region of the genome are translocated to be in close physical proximity to the TRS in the leader of the 5' UTR, which in turn drives the "template switch" to generate template for sgRNA synthesis (Zuniga et al., 2004). RNA-protein and protein-protein interactions are suggested to be involved in the process (Choi et al., 2002; Huang and Lai, 2001; Shi et al., 2000).

Coronaviral N has been described as a multifunctional viral protein characterized by RNA-binding and protein-protein interaction activity. In addition to the primary structural role of N in packaging the RNA genome to form the ribonucleoprotein (RNP) complex, N also appears to play regulatory roles during the viral life cycle, including the aspects of viral RNA synthesis (replication and transcription), translation, as well as in modulating host cell metabolism (Almazan, 2004; Baric et al., 1988; Chang et al., 2005; Eleouet et al., 2000; He et al., 2003). In an early in vitro replication system, N-specific

antibodies were found to inhibit MHV RNA replication (Compton et al., 1987). In a system involving HCoV RNA vectors containing the complete *orf1ab* gene, the presence of N dramatically enhanced the efficiency of RNA replication in this system (Thiel et al., 2003b). N has been shown to bind to the leader RNA sequences of the 5' end of genomic RNA and of all sg mRNAs; this suggests that a high affinity binding site(s) is found in the leader RNA. N has also been shown to bind to the negative-stranded RNA in MHV, as well as a 32 nucleotides stem-loop structure found in the 3' UTR in infectious bronchitis virus (IBV) (Baric et al., 1988; Nelson et al., 2000; Stohlman et al., 1988; Zhou et al., 1996); the specificity of these interactions remain unclear. Interestingly, Zuniga *et al.* recently showed that N from TGEV and SARS-CoV have RNA chaperone activity, and TGEV N can promote viral TRS annealing (Zuniga et al., 2007). These data suggest that N plays an important accessory role in discontinuous transcription as well as perhaps other functions (Choi et al., 2002; Shi et al., 2000).

The molecular weights of N vary considerably between coronaviruses from different groups (377–455 amino acids). Early attempts to purify a coronavirus recombinant coronaviral N from the group 3 coronavirus IBV and from the group 2 SARS-CoV revealed that it is highly sensitive to proteolysis and resulted in two globular domains, an N-terminal RNA binding domain (NTD) and a C-terminal dimerization domain (CTD) (Chang et al., 2006). The NMR and crystal structures of the NTD from SARS-CoV N (Huang et al., 2004), as well as from IBV N have been reported. The crystal structure of the NTD (N29-160) of IBV N (Fan et al., 2005; Jayaram et al., 2006) was interpreted to suggest that the basic building block for nucleocapsid assembly was

an N dimer, although no evidence for or against this proposal has yet been reported (Fan et al., 2005). Biochemical data clearly reveal that the CTD is involved in oligomerization of N (Chang et al., 2006; Surjit et al., 2004; Yu et al., 2005). The crystal structures of CTD from IBV and SARS-CoV reveal a tightly intertwined, domain-swapped, stable dimer (Jayaram et al., 2006). Interestingly, the extreme N-terminal region of the IBV CTD is rich in basic amino acid, and the crystal structure of the SARS-CoV CTD containing this region has recently been implicated in nucleic acid binding activity (Chen et al., 2007). The carboxy terminus of N has been shown to interact with the C-terminal region of the M protein; this interaction seems to drive the encapsidation of viral RNA in the budding viral particle (Kuo and Masters, 2002; Luo et al., 2006b; Verma et al., 2006). This interaction is also suggested to require a conformational change in N which is triggered by RNA binding (Ortego et al., 2002). Interestingly, this region of N has also been shown to interact with hnRNP-A1 to regulate the MHV RNA synthesis (Wang and Zhang, 1999). The flexible linker between NTD and CTD is pictured as an interaction "hotspot". This region contains a Ser-Arg (SR)-rich region, which has been suggested to play an important role in regulating the oligomerization of SARS-CoV N (Luo et al., 2005b), and in MHV, this region has been reported to be involved in RNA binding (Nelson et al., 2000). In addition, recent reports suggest that hnRNP-A1 might bind to this region in SARS-CoV N have also appeared (He, 2004; Luo et al., 2005a).

Taking all of the information together, we propose that N has multiple structurally distinguishable RNA binding domains that play distinct functional roles in

different stages of the viral life cycle. We propose that the CTD binds genomic RNA, and mediates the assembly of the helical nucleoprotein filament. In contrast, the NTD plays an accessory role in stabilizing the RNP, as well as a regulatory role in RNA synthesis mediates a relatively high binding affinity of N for the TRS sequence. In the experiments presented here, we have tested this model in MHV using a variety of biochemical approaches.

Materials and methods

Preparation of RNA samples

Unlabeled TRS and TRS mutant RNAs were obtained by *in vitro* runoff transcription using SP6 RNA polymerase and purified by denaturing PAGE essentially as previously described (Nixon et al., 2002b). The 5'-fluorescein labeled TRS (*F*-5'-TRS-L3) with sequence *F*-5'-gAAUCUAAAC was purchased from Dharmacon and processed to deprotect the enclosed 2'-ACE according to the manufacturer's instructions. The final RNA samples concentrations range from 4-10 mM as determined by using the calculated molar extinction coefficient at 260 nm. Poly(A) and poly(dT) samples were obtained from Midland Certified Reagent Company (Midland, TX) and the samples were subjected to exhaustive dialysis into a final buffer of 25 mM potassium phosphate, 20 mM KCl, pH 8.0. The concentration of poly(A) and poly(dT) were spectrophotometrically quantified using the extinction coefficients ε_{260}= 10,300 M $(\text{nucleotide})^{-1}\bullet\text{cm}^{-1}$ and ε_{260}= 8100 M $(\text{nucleotide})^{-1}\bullet\text{cm}^{-1}$ respectively (Kowalczykowski et al., 1981).

Plasmid construction

For the plasmid constructions encoding various fragments of MHV-A59 nucleocapsid protein, the coding sequences were amplified from the full-length MHV N gene using standard PCR based approaches. The PCR products were digested by *Nde*I and *Bam*HI and ligated into either pET15b (Novagen) which will result in a N-terminal His_6-tag or pET3a expression vector without any tag. Only MHV N197 (residues 60-197) used for NMR studies was generated with pET3a vector; the other MHV N constructs were generated with pET15b. The plasmids encoding the substitution mutants were prepared using QuickChange PCR-based mutagenesis of the wild type overexpression plasmid as a template. The integrity of all the constructs of both strands of the plasmid were confirmed by DNA sequencing.

Protein expression and purification

Recombinant proteins were expressed from their respective pET3a-N and pET15b-N plasmids, in *Escherichia coli* BL21(DE3)/*pLysS* and were grown on 1.5% LB agar plates containing 0.1 mg/ml ampicillin at 37°C. The growth, expression, and purification of N fragments expressed from pET3a were carried out by using the procedure described previously (VanZile et al., 2000). For the His_6-tag fusion proteins expressed from pET15b, the growth and expression followed the same protocol as those proteins in pET3a. The harvested cell pellet was resuspended in lysis buffer (10 mM Tris, pH 8.0, 300 mM NaCl, 15 μM PMSF) and lysed by sonication. Cell lysate was centrifuged at 10,000 x g for 30 min twice at 4^oC, and the supernatant was filtered using 0.2 μm filter and then loaded onto a 10 ml Ni-NTA affinity HiTRAP column (Qiagen).

The column was then washed with lysis buffer with 20 mM imidazole followed by gradient elution from 20 mM imidazole to 500 mM imidazole. The fractions containing MHV Ns were pooled and extensively dialyzed against 10 mM Tris, pH 8.0, 100 mM NaCl to remove the imidazole at 4°C. Inspection of Coomassie-stained 18% Tris-glycine SDS-PAGE gels (Invitrogen) was used to estimate the protein purity to be >95%. The concentration of purified proteins was determined using the calculated molar extinction coefficient at 280 nm. Uniformly ^{15}N -labeled N fragments were expressed and purified similarly except that the cells were grown in minimal medium with ^{15}N $(NH_4)_2SO_4$ as the sole nitrogen sauce as described previously (Pennella et al., 2003). The purified proteins were stored at -80°C in concentrated aliquots.

Nmr experiments

^{1}H-^{15}N HSQC spectra were acquired on a Varian Unity Inova 500 or 600 MHz spectrometer in the Biomolecular NMR Laboratory at Texas A&M University. The sample contained 0.1-0.2 mM ^{15}N-labeled protein in 50 mM potassium phosphate, 100 mM KCl, pH 6.0, at 30°C. Chemical shift referencing is relative to 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS, from Sigma). All spectra were processed and analyzed using NMRPipe (Delaglio et al., 1995) and SPARKY (Goddard and Kneller).

Size exclusion chromatography based assay

The stoichiometry TRS-L3 binding of N-terminal N fragments were analyzed by size exclusion chromatography using an analytical Superdex G75 chromatographic column mounted on an AKTA Purifier 10 LC system (Amersham Pharmacia Biotech). The column was equilibrated at a flow rate of 1 mL/min with 10 mM Tris-HCl, pH 8.0,

100 mM NaCl buffer at room temperature. 100 μl of protein, or TRS-L3 RNA, or a mixture of protein and TRS-L3 RNA with a molar ratio of 1:1.5 was injected and detected by the absorbance at both 280 nm and 260 nm.

Analytical sedimentation equilibrium ultracentrifugation experiments

All experiments were run with a Beckman Coulter Optima XL-A analytical ultracentrifuge equipped with an An60 Ti rotor and two-channel 12-mm path length, charcoal-filled Epon Centerpieces, and quarts windows at 280 nm at 25^{o}C. Three samples with different concentrations were prepared to obtain an Abs_{280} of 0.2, 0.5 and 0.8. The samples were prepared by dilution into 25 mM potassium phosphate, 20 mM KCl, pH 8.0. The data were fitted to a single ideal species model with η=0.74 with ULTRASCAN 9.0.

Fluorescence anisotropy-based RNA-binding experiments

All fluorescence anisotropy experiments were carried out with an ISS PC1 photon counting spectrofluorometer operating in the steady-state mode. The *F*-5'-TRS-L3 concentration in all cases was fixed at 10.0 nM in 50 mM potassium phosphate, 100 mM KCl, pH 6.0. The binding of N variants to TRS-L3 was performed by monitoring the change in the apparent fluorescence anisotropy of the *F*-5'-TRS-L3. The excitation wavelength was 480 nm (2 mm slit width) with the total polarized emission intensity collected with a 530 nm cutoff filter. Nonlinear least-squares fits to these binding isotherms obtained for N variants were carried out using DynaFit (Kuzmic, 1996) with a 1:1 binding model (consistent with the size exclusion results) (see APPENDIX I). The binding of N variants to the unlabeled TRS and TRS mutants was followed using a

"competition" assay, in which N variant was first titrated to the *F*-5'-TRS-L3 until saturation was reached, followed by adding the unlabeled RNA to the same cell to compete off the *F*-5'-TRS-L3. The anisotropy of *F*-5'-TRS-L3 was monitored and fitted in DynaFit (Kuzmic, 1996) (see APPENDIX II).

Intrinsic fluorescence binding experiments

The intrinsic Trp fluorescence of C-terminal variants of N was monitored with an ISS PC1 photon counting spectrofluorometer with excitation at 292 nm (0.5 nm slid width) and emission at 347 nm (4 nm slid width). The intrinsic fluorescence titrations were carried out in 1.7 ml 10 mm pathlength quartz cuvettes in 25 mM potassium phosphate, 20 mM KCl, pH 8.0 buffer at 25^{o}C. Identical small aliquots of N variant were titrated in poly(A) or poly(dT) with indicated concentration (typically 0.3-1.0 μM nucleotide). The solutions were stirred and incubated for 2 min in the dark. The fluorescence intensity was measured by averaging ten 10 s integrations of the signal. Under these conditions there is essentially no photobleaching (<3%). The fluorescence intensity changes monitored were corrected for dilution and the inner filter correction, which is almost negligible under our conditions, and were presented in arbitrary units as a function of total added protein. The data were fitted using Scientist 3.0 using the McGhee-Von Hippel large ligand infinite lattice non-cooperative binding model to estimate the site-size, the apparent number of occluded nucleotide per MHV CTD monomer n_{obs},, K_{obs} and Q_{max}, the maximal fluorescence quenching (see APPENDIX III). Under these solution conditions, the CTD is a stable non-dissociable dimer (see below).

Results

Overview of the functional domain organization of the N

Our proposed domain organization of MHV N is shown in Figure 4-1. The NTD is based on previous structural studies of IBV and SARS-CoV N domain (Figure 4-1A). A multiple sequence alignment of various CoV N is shown in Figure 4-2. The ~130-residue MHV NTD (shaded orange; residue 60-194) is proposed to form a U-shaped β–platform composed of five short β-strands in the order of β5–β1–β3–β2–β4. The palm of the β –platform and the flexible hairpin loop extension (β2' and β3' in IBV NTD (Fan et al., 2005) between β2 and β3) contain a large number of basic and aromatic amino acids, and are therefore suggested to bind RNA (see below; Figure 4-2A).

The NTD is followed immediately by a highly conserved hydrophilic SRXX repeat region or Ser/Arg-rich domain (shaded blue) in the linker and extends to ~ residue 231, with the rest of the linker (to ~ residue 260) from various Ns not conserved. This SR-rich region on compassing residue 177-233 has been suggested previously to specifically bind to the TRS-L sequence (Nelson et al., 2000). We designed three different recombinant N-terminal domain fragments of MHV N, designated N197 (residues 60-197), N219 (residues 60-219) and N230 (residues 60-230) to evaluate the importance of the SR-rich region in RNA binding. N197 corresponds precisely to the region of SARS-CoV NTD (residues 45-181) previously studied by NMR spectroscopy (Huang et al., 2004). N219 appends the SR-rich region onto N197, while N230 corresponds to NTD appends with the further linker region.

Immediately following the linker region is the CTD, and the crystal structures of IBV and SARS-CoV show that the CTD (shaded green in Figure 4-1A) contains a central four-stranded β-sheet platform buttressed by peripheral α-helical elements giving rise to a tightly intertwined, domain-swapped, stable dimer (Jayaram et al., 2006). While the original structure of the SARS-CoV CTD contained residues beginning with the αA helix (see Figure 4-2C), subsequent structures of IBV N and SARS-CoV N CTDs (MHV residues 260-282; see Figure 4-1A) containing ~ 20 additional N-terminal residues rich in basic amino acids led to the identification of what we term the αF helix packed against the core dimer; this helix, in turn, greatly enhances the positive electrostatic potential in this region of the dimer (Chen et al., 2007; Jayaram et al., 2006). Limited biochemical data published earlier this year suggested that the CTD containing the N-terminal region was capable of binding single strand nucleic acids (Chen et al., 2007).

Characterization of TRS-containing RNA targets

The free energy of the duplex formation between TRS-L and complementary TRS-B has been suggested to play an important role in coronavirus discontinous transcription (Sola et al., 2005; Zuniga et al., 2004). Since the N possesses RNA chaperone activity and previous reports suggested that some region of N binds the TRS, this suggests to us that the NTD or N might form a high affinity complex with TRS-containing RNAs.

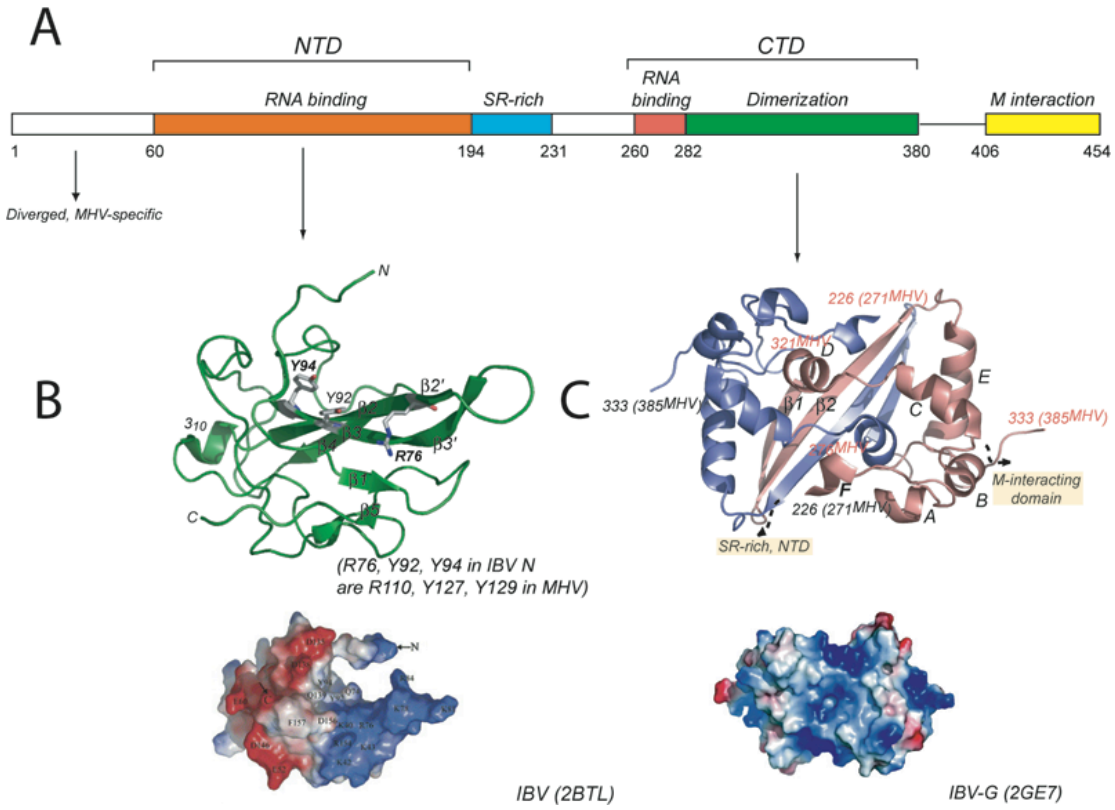


Figure 4-1 Overview of the proposed structural domains of MHV N. (A) Schematic diagram showing the primary structure representation of MHV N. The locations of the domains are MHV N residues corresponding to those known residues of NTD and CTD from IBV and SARS-CoV based on the sequence alignment. (B) is the crystal structure of NTD from IBV N (PDB code: 2BTL) with the surface electrostatic potential picture at the bottom; (C) is the crystal structure of the CTD dimer from IBV (PDB code: 2GE7) with the surface electrostatic potential picture shown at the bottom. In panels B and C, the N- and C-terminal residues in each domain are labeled as are residues that were targeted for substitution in MHV NTD and CTD fragments, based on homology with the IBV N.

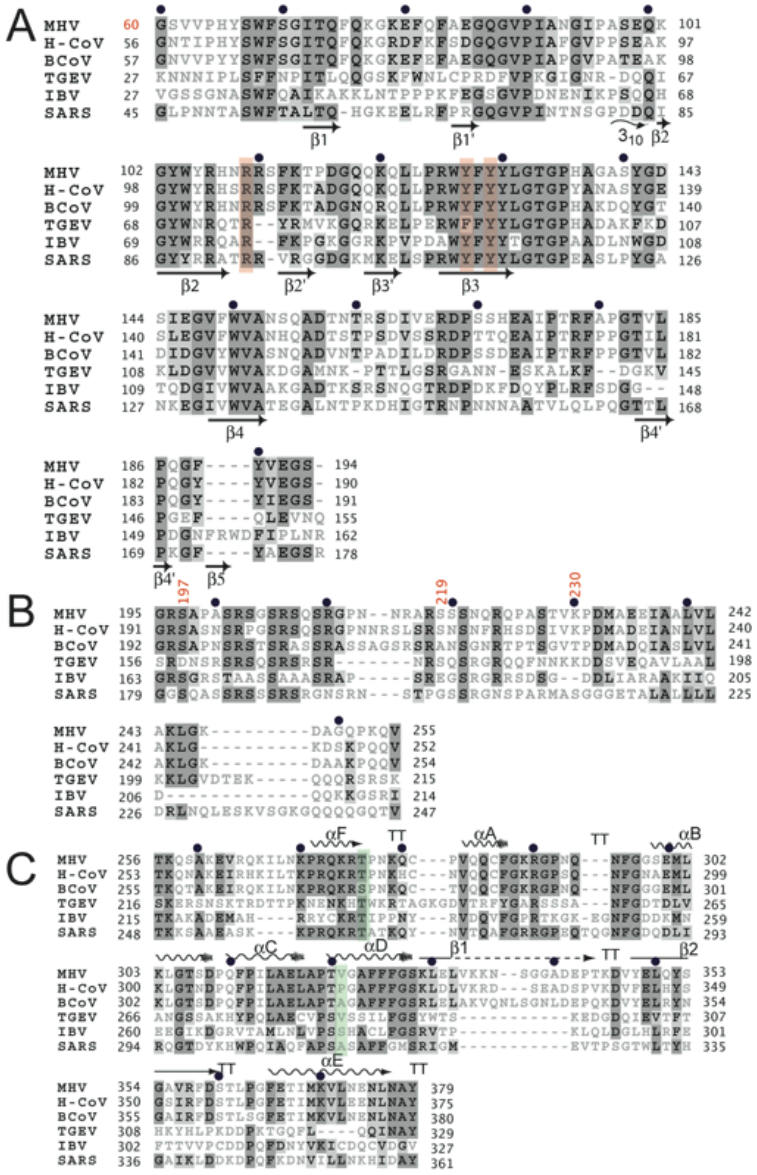


Figure 4-2 Structure-based multiple sequence alignment of coronavirus Ns from all three groups divided into three functionally important subdomains. (see text for details) (A) NTD, (B) liner region between the NTD and CTD, and (C) the CTD. The secondary structure elements from IBV N are indicated above the alignment (Fan et al., 2005). Conserved residues are shaded grey with the sites for mutation in the MHV NTD and CTD shaded red and green, respectively. The C-terminal residues of N197, N219 and N230 are labeled in panel B.

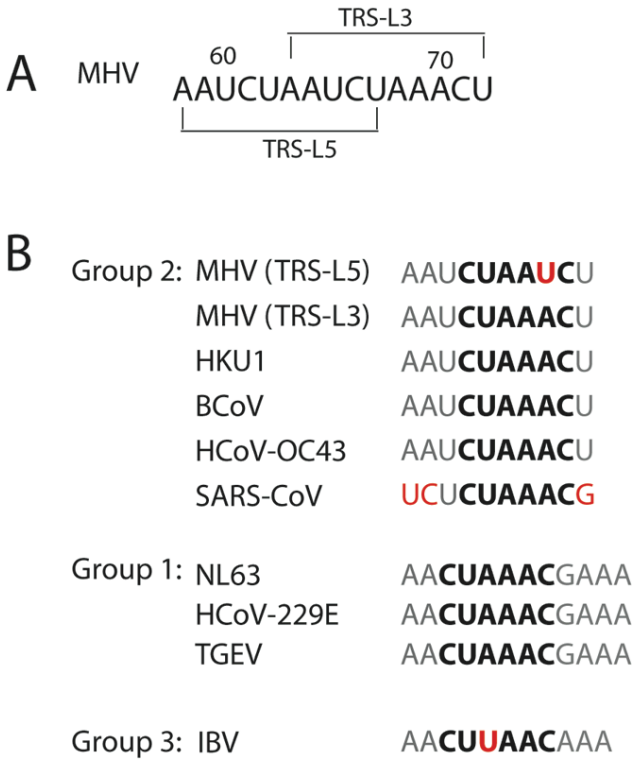


Figure 4-3 The TRS in the leader from three different coronavirus groups. (A) MHV TRS in the leader sequence. The overlapping TRS-L3 and TRS-L5 are indicated. (B) The core TRS with the flanking residues from all three groups. The core TRS are shown in bold face, while the sub-group specific substitutions are shown in red.

The core of TRS is a highly conserved hexanucleotide sequence, 5'-CUAAAC with the only exception IBV in group 3, which contains a single substitution 5'-CUUAAC. In addition, the flanking residues are also conserved within individual coronaviruse groups (Figure 4-3). In group 2, the MHV TRS region is unique in that this

region contains two partially overlapping TRS-like sequences, designated TRS-L5 5'-^{58}AAUCUAAUCU^{67} and TRS-L3 5'-^{63}AAUCUAAACU^{72}, immediately downstream from SL2 (Chapter II). Based on the conservation of the TRS in group 2 coronaviruses, we synthesized an RNA we denote as TRS-L3, 5'-gAAUCUAAAC with a nonnative 5'-g required for *in vitro* run-off transcription. We also produced a series of mutant RNAs in order to investigate the RNA binding specificity of the NTD of MHV N.

The TRS-L3 RNA and MHV NTD form a 1:1 tight-binding complex

We first showed that the NTD of MHV N is capable of binding the TRS within the MHV 72 nucleotides leader sequence using a crude native gel electrophoretic mobility experiment (data not shown). It was next of interest to determine the extent to which the NTD binds directly to the TRS. The stoichiometries of the interaction between TRS-L3 and two NTD variants (N197, N230) were determined to be 1:1 by the gel filtration experiments shown in Figure 4-4. Two peaks, at 11.5 mL and 13.5 mL, were observed in the gel filtration profile of the mixture of N230 and TRS-L3 at molar ratio of 1:1.5. Comparison of these peaks with the profiles of N230 and TRS-L3 alone indicates that the peak at 13.5 mL corresponds to that of the TRS-L3 RNA alone with absorbance at 260 nm being higher than that at 280 nm, whereas there is a 1-mL shift between the peak at 11.5 mL in the mixture and that of N230 alone. The faster migration rate of peak at 11.5 mL indicates that it contains a species larger than free N230 alone. Similar results are observed in the analytical gel filtration experiment carried out with N197 and TRS-L3 (data not shown). This 1:1 stoichiometry is also reflected in the ^{15}N-^{1}H HSQC-spectra of NTD variants with TRS-L3. Figure 4-5D shows the superposition of spectra

obtained for ^{15}N-N219 in the absence (red contours), the presence of 0.5 (green contours) and 1.0 (blue contours) molar equivalents of TRS-L3. The addition of one additional molar equivalent of TRS-L3 shows no further changes in the spectrum (data not shown).

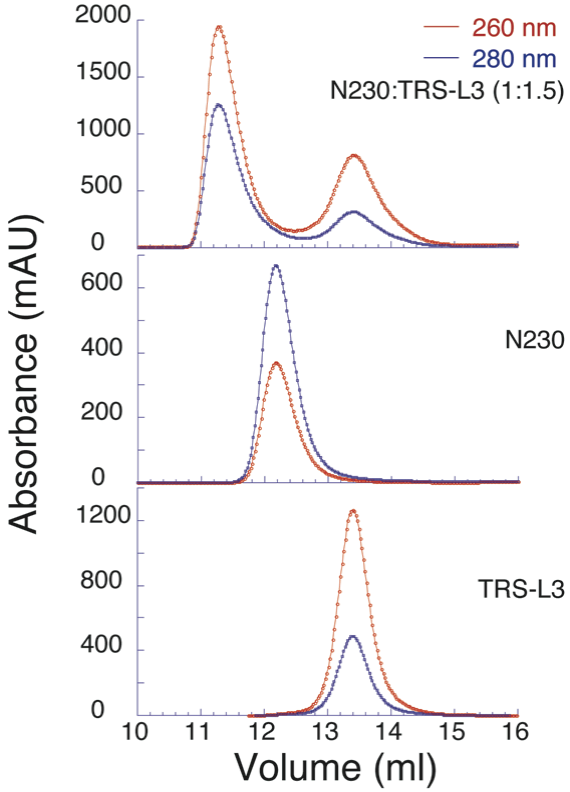


Figure 4-4 Gel filtration profiles of the mixture of MHV N230 and TRS-L3 at a 1:1.5 molar ratio. The profile of N230 and TRS-L3 alone reveals a single elution peak at 12.5 mL and 13.5 mL respectively. The profile of the mixture shows two peaks, one has an elution volume of 11.5 mL, and the other elutes at 13.5 mL, the latter identical to the TRS-L3 RNA alone.

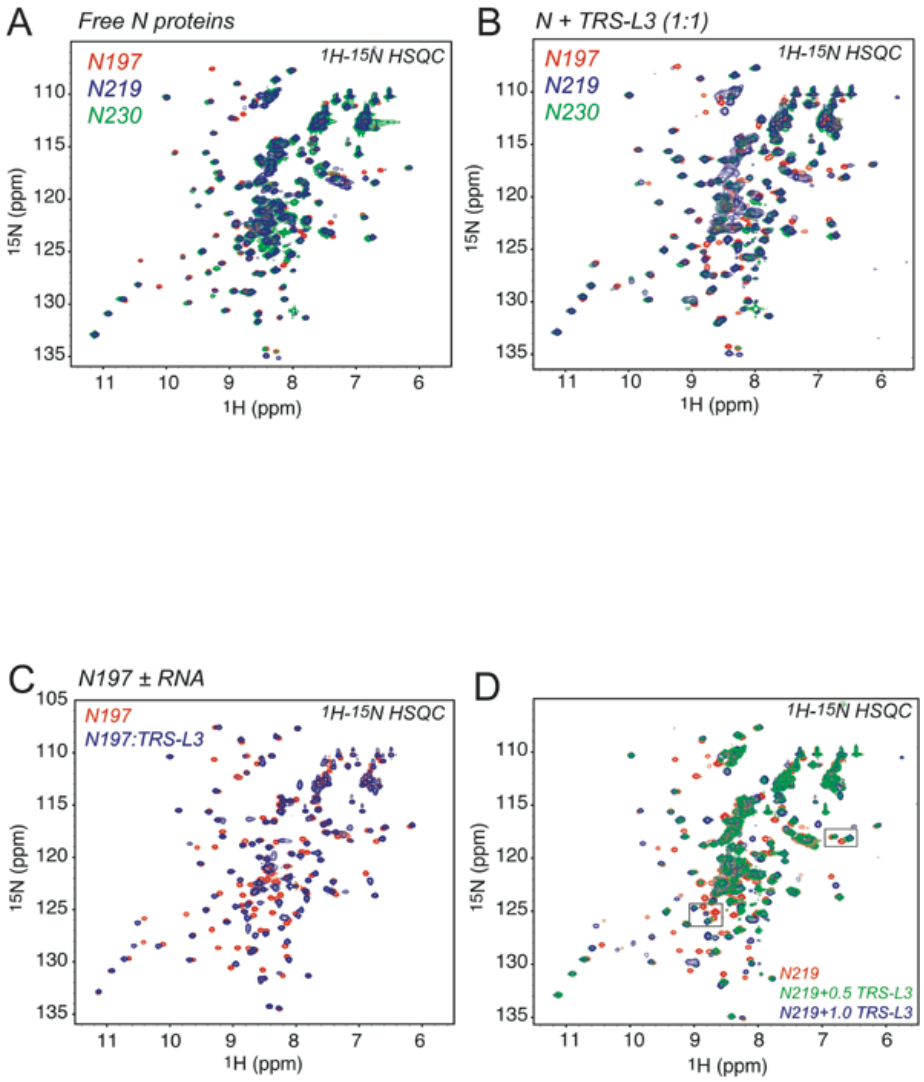
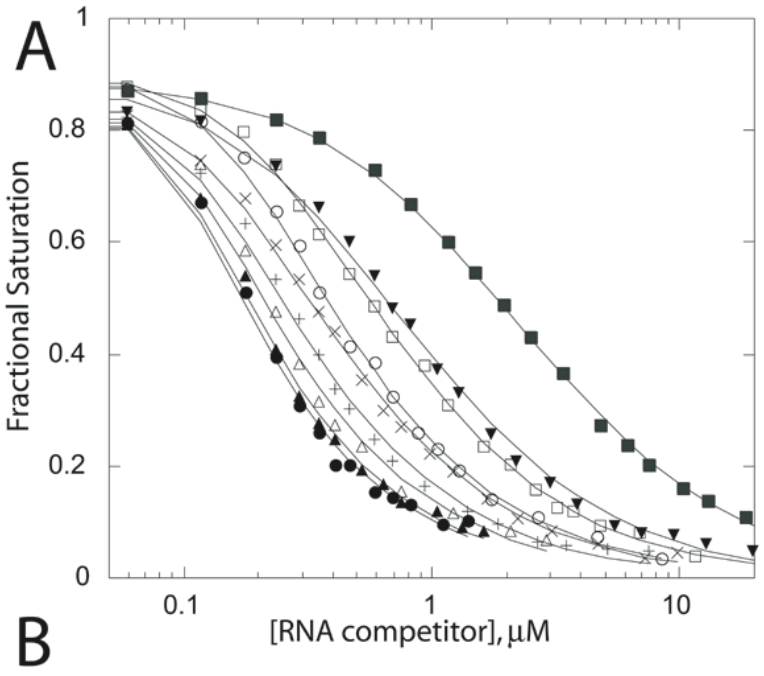


Figure 4-5 ^{1}H-^{15}N HSQC NMR spectroscopy of MHV N197, N219 and N230 variants. The superposition of (A) N197 (red), N219 (blue), N230 (green) in the absence of TRS-L3; (B) N197 (red), N219 (blue), N230 (green) in the presence of stoichiometry of TRS-L3; (C) N197 alone in the presence (blue) and absence (red) of TRS_L3; (D) N219 in the absence (red), presence of 0.5 molar ratio (green), and 1 molar ratio (blue) of TRS-L3. All experiments were performed in 50 mM potassium phosphate, 100 mM KCl, pH 6.0, at 30 ^{o}C. The protein concentration ranges from 100-200 μM.

Full-length MHV N has previously been shown to bind the TRS-containing leader sequence with a dissociation constant (K_d) of 14 nM, with the SR-rich region thought to provide most of the binding determinants (Nelson et al., 2000). We therefore measured the binding affinity of a 5' fluorescein-labeled TRS-L3 (*F*-5'TRS-L3), with N219, a NTD variant containing the intact SR-rich region, by fluorescence anisotropy under the solution conditions of 50 mM potassium phosphate, 100 mM KCl, pH 6.0, 25 ^{o}C. The binding isotherm obtained for N219 is shown in Figure 4-6 with the parameter values showing the results of the fit giving in Table 4-1. This complex is characterized by a K_{obs}=9.4(±0.9) x 10^7 M^{-1}, for ΔG_{obs}=-10.8 ± 0.1 k cal mol^{-1}.

NTD specifically binds to the TRS-L3 RNA

To address whether the TRS-L3 binding activity of NTD is nucleotide sequence specific or is sequence nonspecific, we carried out a series of fluorescence anisotropy-based RNA competition experiments with a collection of mutant TRS-L3 RNAs. In this experiment, incremental additions of the unlabeled mutant RNA is added to a complex formed by MHV N219 and *F*-5'-TRS-L3, and the fluorescence anisotropy of the *F*-5'-TRS-L3 is monitored. The data are fitted using DYNAFIT to a simple competition model (The Dynafit Scripps given in the Appendix II) fixing K_{obs} for *F*-5'-TRS-L3 to the value obtained in a direct titration. Note that the K_{obs} obtained for unlabeled wild type TRS-L3 generated from this competition experiment is indistinguishable within experimental error to that obtained from the direct anisotropy titration experiment with *F*-5'-TRS, 8.8(±0.9) x 10^7 M^{-1} vs 9.4(±0.9) x 10^7 M^{-1}; this reveals that the fluorescein moiety at the 5' end of the RNA does not influence the binding affinity as expected.



●	TRS-L3:	gAAUCUAAAC	K_{obs}: 9.39(±0.94) x10^7 M^{-1}
×	TRS-L3A68u:	gAAUCUuAAC	K_{obs}: 3.29(±0.12) x10^7 M^{-1}
▲	TRS-L3A69u:	gAAUCUAuAC	K_{obs}: 7.90(±0.65) x10^7 M^{-1}
□	TRS-L3A70u:	gAAUCUAAuC	K_{obs}: 2.47(±0.078) x10^7 M^{-1}
■	TRS-L3u:	gAAUCUcuuC	K_{obs}: 0.44(±0.012) x10^7 M^{-1}
○	TRS-L3a:	gAAgagAAAC	K_{obs}: 4.14(±0.19) x10^7 M^{-1}
+	TRS-L3g:	gAAUCUgAgC	K_{obs}: 4.75(±0.30) x10^7 M^{-1}
▼	TRS-L3gu:	gAAgagcuuC	K_{obs}: 1.36(±0.07) x10^7 M^{-1}
△	TRS-L3C71u:	gAAUCUAAAu	K_{obs}: 6.17(±0.74) x10^7 M^{-1}

Figure 4-6 Competition binding isotherms of WT TRS-L3 RNA and mutants with N219-*F*-5'-TRS-L3 complex from fluorescence anisotropy-based RNA competition experiments. Conditions: 50 mM potassium phosphate, 100 mM KCl, pH 6.0, 25 ^{o}C. (A) The normalized fractional saturation of *F*-5'-TRS is plotted vs the concentration unlabeled TRS-L3 mutant. The sequences and fitted K_{obs} of the mutants are listed in (B) with fitted binding parameters compiled in Table 4-1. The substitutions are shown in red, and the non-native residues shown in lower-case.

Table 4-1 Summary of TRS-L3 and mutant binding parameters of NTD variants as determined by fluorescence anisotropy titrations and competition experiments.

Proteins	RNAs	K_{app} ($x10^7 M^{-1}$)	ΔG(kcal mol^{-1})	ΔΔG(kcal mol-1)	ΔG_c(kcal mol^{-1})
N219	TRS-L3	9.39(0.94)	-10.88	0	
N197		1.86(0.09)	-9.92	0.96(0.07)	
Y129a		3.47(0.26)	-10.29	0.59(0.08)	
R110a		4.12(0.16)	-10.39	0.49(0.07)	
Y127a		0.45(0.06)	-9.08	1.80(0.10)	
Y129aR110a		2.37(0.18)	-10.06	0.82(0.06)	-0.27
Y127Y129		0.35(0.12)	-8.93	1.95(0.26)	-0.44
N219	TRS-L3	9.39(0.94)	-10.88	0	
	TRS-L3A68u	3.29(0.12)	-10.26	0.62(0.06)	
	TRS-L3A69u	7.90(0.65)	-10.77	0.10(0.08)	
	TRS-L3A70u	2.47(0.08)	-10.08	0.79(0.06)	
	TRS-L3u	0.44(0.01)	-9.06	1.81(0.06)	
	TRS-L3a	4.14(0.19)	-10.38	0.49(0.07)	
	TRS-L3g	4.75(0.30)	-10.47	0.40(0.07)	
	TRS-L3gu	1.36(0.07)	-9.73	1.14(0.07)	-1.16
	TRS-L3C71u	6.17(0.74)	-10.63	0.25(0.09)	
	TRS-L3L	13.46(2.28)	-11.09	-0.21(0.12)	
	TRS-L3LA70u	5.30(0.18)	-10.54	0.34(0.06)	
Y129a/R110a	TRS-L3	3.22(0.30)	-10.24	0.63(0.08)	
	TRS-L3A70u	0.77(0.10)	-9.39	1.48(0.10)	
	TRS-L3u	0.16(0.02)	-8.46	2.41(0.08)	

50 mM potassium phosphate, 100 mM KCl, pH 6.0, at 25°C. Numbers in the parenthesis are the standard errors.

All mutations in the TRS-L3 result in a decrease in binding affinity K_{obs}, but to varying degrees (Figure 4-6 and Table 4-1). For simplicity, we divided the TRS-L3 sequence into pyrimidine (^{65}UCU^{67}) and purine (^{68}AAA^{70}) sub-segments. While substitutions of ^{65}UCU^{67} with ^{65}gag^{67} results in only a ≈ 2-fold decrease in K_{obs}, or ΔΔG of 0.5 kcal mol^{-1}, complete replacement of the ^{68}AAA^{70} sequence with ^{68}cuu^{70} results in a

20-fold decrease in binding affinity or ΔΔG of 1.8 k cal mol^{-1}. Combining the two blocks of mutations into the same RNA (TRS-L3gu) gives rise to a ΔΔG of 1.16 kcal mol^{-1} or less than the sum of the component single triple-nucleotide substitutions, by ΔG_c of -1.14 kcal mol^{-1}. This suggests that the $^{68}AAA^{70} \rightarrow {}^{68}cuu^{70}$ substitution is globally destabilizing to the interface. Interestingly, individual pyrimidine (u) substitution of

$$\begin{array}{ccc} 0 & & 1.81 \\ WT & \rightarrow & {}^{68}cuu^{70} \\ \downarrow & & \downarrow \\ {}^{65}gag^{67} & \rightarrow & {}^{65}gagcuu^{70} \\ 0.49 & & 1.16 \end{array} \qquad \begin{array}{l} \Delta Gc = 1.16 - (1.81+0.49) \\ \quad = -1.14 \text{ kcal mol}^{-1} \end{array}$$

A68, A69 and A70 gives rise to ΔΔGs of 0.6, 0.1 and 0.8 kcal mol^{-1} respectively, the sum of which is less than the triple mutant ΔΔG=1.81 kcal mol^{-1}. Although the significance of this is not clear, only A69 appears nearly dispensable for high affinity binding. Purines are favored at at A68 and A70 as well, since in A68g/A70g/^{69}A mutant only gives ΔΔG of 0.4 kcal M^{-1}. Thus, the NTD shows a high purine bias in these positions with adenosines preferred. We note that the A70u mutant corresponds precisely to the upstream TRS-L5 (see Figure 4-3). Since this mutant results in a reduced binding affinity (ΔΔG=0.8 kcal M^{-1}), the data suggest that NTD effectively discriminates between very similar TRS sequences. Finally, to exclude the possibility that the A70u substitution is influenced by the 3' terminus, two additional control RNAs were designed, TRS-L with an extension on the 3' side of the RNA through actual leader-body joining segment (U72) via addition of $^{72}UUU^{74}$ sequence in the context of a WT and A70u substitution. The additional $^{72}UUU^{74}$ only leads to a small 1.5-fold increase in

affinity, and the relative K_{obs} values for the wild-type and A70u mutants are largely unchanged (see Figure 4-7 and Table 4-1). Collectively, these data suggest that all of the high affinity binding by N219 to the TRS is contained within those nucleotides just upstream of the junction region, toward the 3' edge of the leader. This is exactly where binding might be optimally positioned to enhance TRS-L-cTRS-B hybridization of template switching.

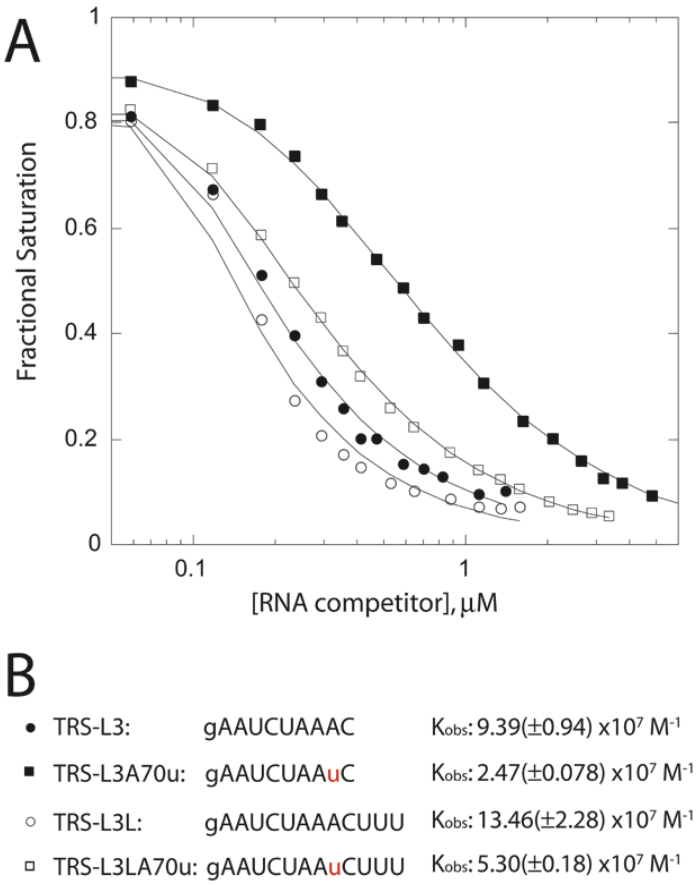


Figure 4-7 Competition of binding isotherms of TRS-L3, TRS-L3L and mutants with N219 derived from fluorescence anisotropy-based RNA competition experiments.

In addition, the superposition of ^{15}N-^{1}H HSQC spectra of MHV N197 with and without stiochiometric TRS-L3 shows chemical shift changes in a small number of amide resonances, with the remaining crosspeaks are insensitive to the bound TRS-L3 RNA (Figure 4-5C). This perturbation map, while not yet assigned, differs from that of SARS-CoV NTD bound to a nonspecific RNA (Huang et al., 2004); this is also consistent with specific binding.

The SR-rich does not engage in the specific interaction with the TRS-L3, but stabilizes the complex via electrostatic interactions

The SR-rich region of N has been previously suggested to specifically bind the leader in MHV (Nelson et al., 2000). However, the data presented below suggest that this is not the case. First, if this region is located in a proteinase sensitive flexible linker region, one might reasonably expect that RNA binding would induce the formation of a well folded structure. However, the inspection of the ^{15}N-^{1}H HSQC spectra of all three free NTD variants (N197, N219 and N230) reveals that the SR-rich tail in both N219 and N230 protein is conformationally averaged on the μs- timescale (see Figure 4-5A, B). In both N219 and N230 constructs, this is consistent with the fact that the crosspeaks expected for the tail are very weak or missing altogether, as are additional amide resonances in the core N197 structure as yet unassigned. However, the TRS-L3 induced perturbation maps of all three forms of the NTD are very similar to one another. The data suggest that base specific interactions are limited to the NTD core region, represented by MHV N197. In addition, the well-dispersed spectra of N197 in both free and TRS-L3-bound forms reveal that both conformations are in slow chemical exchange.

This is consistent with the K_{obs} for N197 measured by direct binding in the fluorescence anisotropy experiment. (K_{obs}= 1.86 ± 0.09×10^7 M^{-1} ; ΔG_{obs}=-9.9 kcal mol^{-1}) (see Figure 4-8 and Table 4-1).

Although the SR-rich region does not specifically bind the TRS-L3, this region does increase the binding affinity of the K_{obs} of N219 which is almost 5-fold of that of N197 (see Figure 4-8 and table 4-1). This increase in binding affinity may mainly be caused by the addition of large number of positive charges from the SR-rich region, which manifests itself a larger electrostatic component to the binding energy in N219 vs. N197. To test this, standard "salt-back" dissociation experiments were carried out to obtain information on the extent to which electrostatic interactions stabilize the NTD-TRS complex (Chen et al., 1998). Here, the NTD-*F*-5'-TRS-L3 complexes formed at a low salt are incrementally dissociated with the addition of small aliquots of a concentrated KCl stock solution. The fluorescence anisotropy change was monitored and fitted to obtain the SK_{obs}, the slope of the plot of log K_{obs} vs log [KCl]. The SK_{obs} , the dependence of K_{obs} on [K+], of N219 is large (SK_{obs} = -5.45), consistent with 7-8 ionic interactions in the complex for the RNA binding (Figure 4-9 and Table 4-2). At [KCl] = 0.15 M, ΔG_{PE} of N219 is -6.12 kcal mol^{-1}, revealing that 55% of the total binding free energy under this condition is contributed by the polyelectrolyte effect. In contrast, the SK_{obs} of N197 is only -3.92, and the polyelectrolyte contribution is only ~40%. In contrast to large difference in SK_{obs}, the log K_{obs} of 1 M K+, though of as the free energy of binding due to non-electronic interaction is logK≈ 4, more similar than different. This is perhaps contributed to aromatic residues in the NTD. Thus, in this simplified

polyelectrolyte model, N219 engages in 2-3 additional electrostatic interactions likely contributed by a subset of the additional C-terminal Arg residues.

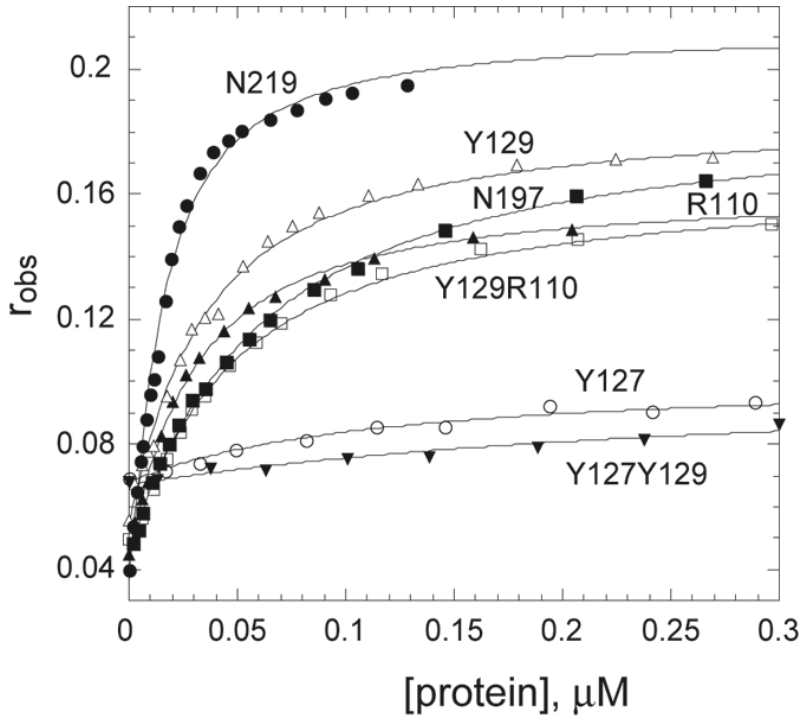


Figure 4-8 Fluorescence anisotropy titrations of *F*-5'-TRS with various MHV NTDs. N219 (solid circle), N197(solid square), and mutants based on N219 construct, Y129a (empty upside triangle), R110a (solid upside triangle), Y127a (empty circle), Y129a/R110a (empty square), and Y127a/Y129a (solid downside triangle). All experiments were performed at 25°C in 50 mM potassium phosphate, 100 mM KCl, pH 6.0. The anisotropy, r_{obs}, is plotted as a function of total protein concentration in monomer units. The solid lines through the experimental data represent the nonlinear least-squares fits according to a 1:1 binding model, and the derived values of K_{obs} and calculated ΔG, $\Delta\Delta G$ and ΔGc values are listed in Table 4-1.

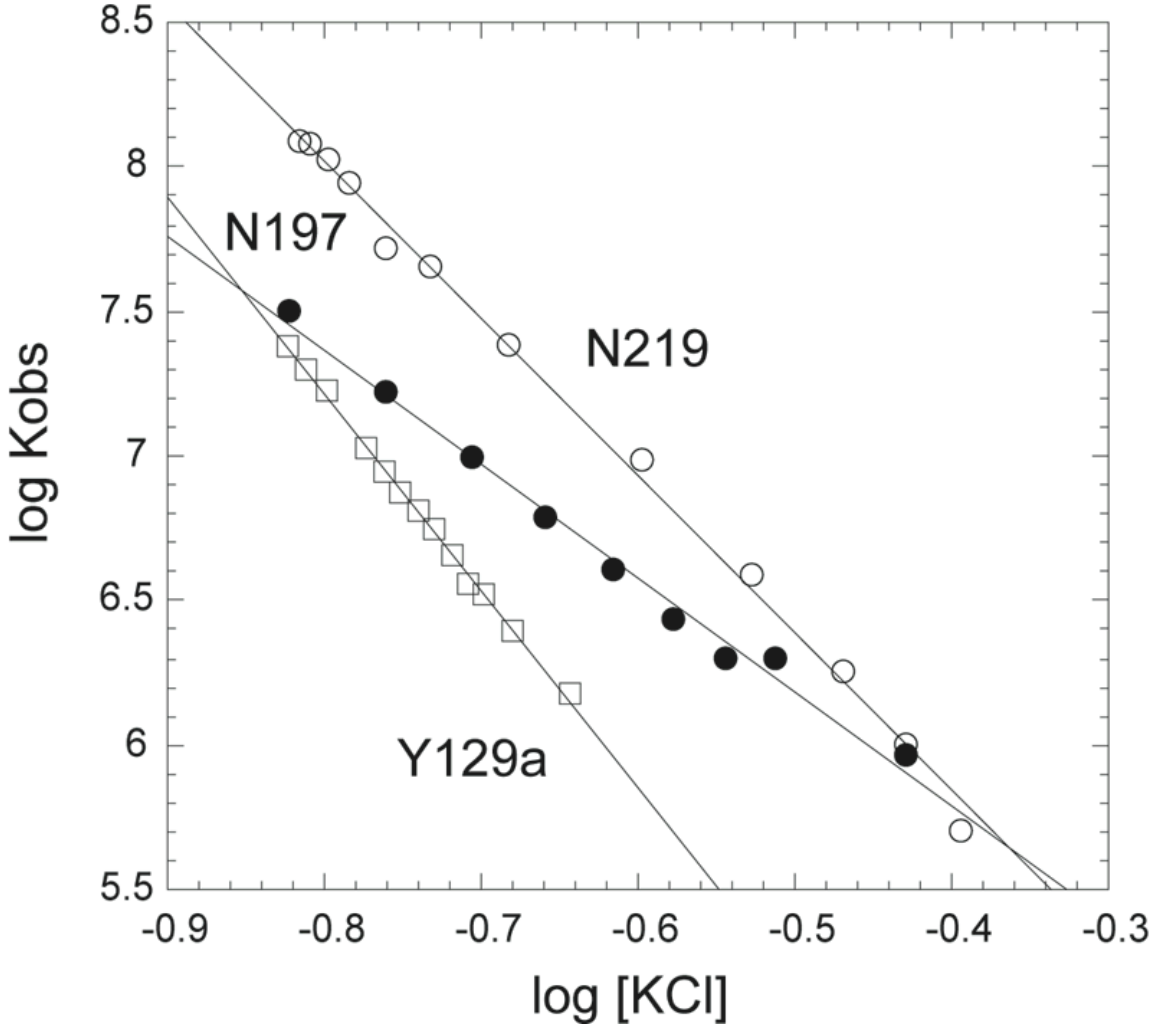


Figure 4-9 Log K_{obs} – log [KCl] plots of the effect of [KCl] on the binding affinity of NTD variants to the *F*-5'-TRS-L3 RNA. N219 (empty circle), N197 (solid circle), and Y129a (empty square). The experiments were performed in 50 mM potassium phosphate, 100 mM KCl, pH 6.0, at 25^{o}C. The protein was first titrated into the F-5'-TRS until saturation was achiedved, and small aqueous 4M KCl was titrated into the mixture, and the fluorescence anisotropy change was monitored and fitted. The fitted slope values (SK_{obs}) and calculated ΔG_{PE} for each protein are listed in Table 4-2.

Table 4-2 Summary of salt dependence of the binding of various NTD preparations to *F*-5'-TRS-L3 as determined by "salt-back" dissociation fluorescence anisotropy titrations.

Proteins	SK_{obs}	$\log K_{obs}^{1M}$	ΔG_{PE} (kcal mol $^{-1}$)	$\Delta G^{0.15M}$(kcal mol $^{-1}$)	%polyelectrolyte effect
N219	-5.45	3.66	-6.12	-11.1(-10.8)	55.28
N197	-3.92	4.22	-4.39	-10.1(-9.9)	43.33
Y129a	-6.82	1.76	-7.66	-10.1(-10.2)	76.14

SK_{obs} is fitted slope from "salt back" experiemtn, K_{obs}^{1M} is the K_{obs} at 1 M [KCl], ΔG_{PE} is calculated from $\Delta G_{PE}=-SK_{obs}RT\ln[KCl]$, with [KCl]=0.15M as in the experiment. $\Delta G^{0.15M}$ is calculated from $\Delta G=-RT\ln K_{app}^{1M}$, when [KCl]=0.15M as in this experiment, The number in parenthesis is calculated from the direct binding experiments.

Mutations in N219 influence the TRS-L3 binding affinity to varying degrees

Arg76 and Tyr94 in IBV N have been identified as residues that participated in RNA-binding and the substitution of one or both of which reduces viral infectivity (Tan et al., 2006) (see Figure 4-1). R76 and Y94 are located in the positively charged hairpin loop β2' and β3 strand on the hydrophobic platform, respectively, which has been implicated in RNA binding activity (Huang et al., 2004). From the multiple sequence alignment (Figure 4-2), R76 and Y94 are strongly conserved in Ns from all three coronavirus groups, and correspond to R110 and Y129 in MHV N. Upon examination of the structure of IBV and SARS-CoV NTD, we notice that Y127 in MHV N, corresponding Y92 in IBV N, is also solvent exposed and probably involved in RNA binding. In fact, Y92 (Y127) is located between Y94 (Y129) and R76 (R110 in MHV

N), and may function as a "bridge" residue that energetically couples these two residues. Interestingly, both Y92 and Y94 are part of the strongly conserved motif of W*Y*FF*Y*Y in coronavirus N (the positions of these two Y are shown in italics). We therefore targeted Y127, Y129 and R110 for substitution in N219 singly (Y127a, Y129a, R110a), doubly (Y127a/Y129a, Y127a/R110a, Y129a/R110a), and triply (Y127a/Y129a/R110a) combinations. A Preliminary determination of TRS-L3 binding affinity and salt dependence of each MHV Ns were then obtained.

The fluorescence anisotropy titrations are shown in Figure 4-8 with the fitted value of K_{obs} and r_i compiled in Table 4-1. Both Y129a and R110a result in about 3-fold decrease in binding affinity compared with wild type N219, corresponding to ΔΔG value of 0.59 and 0.49 kcal mol^{-1}, respectively. The Y129a/R110a double mutant gives ΔΔG of 0.82 kcal mol^{-1}, which gives a $\Delta Gc = \Delta\Delta G^{double} - \sum\Delta\Delta G^{Single}$ of -0.27 kcal mol^{-1}. Thus, although the effect of these individual mutations is small, the two residues are significantly energetically coupled, consistent with a long distance cooperativity. In addition, we also characterized the binding of the Y129a/R110a double mutant to two mutant RNAs, TRS-L3u ($^{68}AAA^{70} \rightarrow {}^{68}cuu^{70}$) and to TRS-L3A70u using RNA competition experiments. This double mutant shows the same extent of decreased binding affinity to each mutant RNA as is observed for the wild type N219 (see Fig 4-10). In other words, the ΔΔG for Y129a/R110a N on each of the mutant RNA is 0.6 kcal mol^{-1}, identical to that of the WT RNA. These given to the analysis suggests that more peripheral residue Y129 and R110 are unlikely to make base-specific contact with AAA sequence, and the RNA binding stabilizes the complex in an indirect way. However, the

possibility remains that Y129 and R110 may still directly interact with other residue(s) in TRS-L3 sequence.

In contrast, substitution of the "bridge" residue Y127 with Ala (Y127a) gives rise to ≈ 20-fold decrease in binding affinity relative to N219, or $\Delta\Delta G$ = 1.8 kcal mol^{-1}, a large effect. Since this free energy determined is identical to that of the $^{68}AAA^{70} \rightarrow {}^{68}cuu^{70}$ mutant, one might speculate that Y127 anchors the site-specificity of the N219-TRS complex. Interestingly, the K_{obs} for the Y127a/Y129a mutant is ≈ 27-fold less relative to WT ($\Delta\Delta G$ = 1.95 kcal mol^{-1}). Strikingly however, ΔGc for Y127 and Y129 is -0.44 kcal mol^{-1}, consistent with the idea that Y127a substitution is globally disruptive to complex formation. Additional experiments are required to address the effect of Y127a substitution in the specific binding of N.

Preliminary "salt-back" dissociation experiments were also performed. The Y129a substitution seems to reveal a significant increase in electrostatic contribution to the total RNA binding free energy, indicated by the fact that the polyelectrolyte contributed free energy (ΔG_{PE}) accounts for ~76% of the total binding energy whereas in WT N219, it accounts for only ~55%. This might suggest a more global recognition of the binding interaction despite the relative small decrease in ΔG at 0.15 M KCl. This indicates that Y129 may also interact with the TRS but perhaps upstream of the AAA sequence.

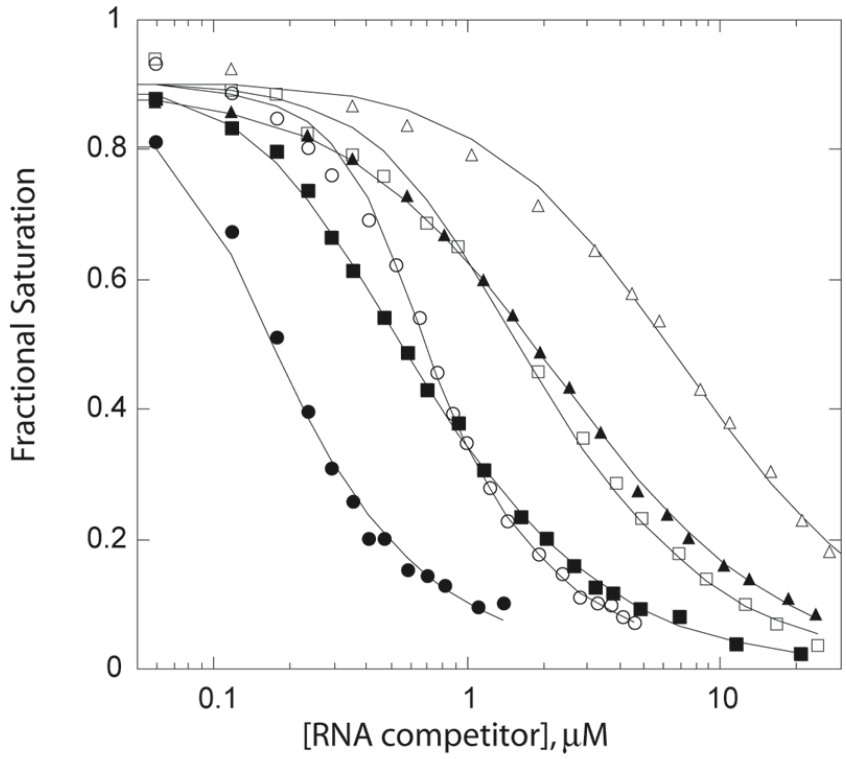


Figure 4-10 Fluorescence-anisotropy-based competition binding isotherms of TRS-L3 (circle) and mutants TRS-L3A70U (square), TRS-L3u (triangle) with N219 (solid) and Y129a/R110a mutant (empty) respectively. The fitted values of K_{obs} and calculated ΔG and $\Delta\Delta G$ values are listed in Table 4-1.

MHV CTD forms a dimer in solution

The available crystal structures of the CTD (Jayaram et al., 2006; Yu et al., 2006) show that CTD exists as a tightly intertwined dimer with two antiparallel β strands and a α helix from one monomer extensively interacting with the other monomer. The electrostatic potential surface of the CTD dimer shows that αC, αD and αF helices of

the dimer, combine to form a shallow groove rich in basic residues but lacking conserved aromatic residues that could interact with RNA (see Figure 4-1). From this observation, it has been speculated that nucleocapsid filament formation may be driven either by CTD dimer-dimer interactions or higher-order (octameric) oligomers of CTD (Surjit et al., 2004) or NTD-NTD interactions. However, little biochemical information on the interaction of the CTD with nucleic acids currently exists. To gain direct insight into this proposed CTD-nonspecific RNA binding mode, we designed an MHV CTD construct consisting of residues 256 to 385 (see Figure 4-1, 4-2), which corresponds to the SARS-CoV CTD (from 248 aa) as recently studied crystallographically (Chen et al., 2007) (see Figure 4-2), and carried out simple RNA binding experiments with homopolymeric nucleic acids.

The sequence alignment of CTD shows that the predicted RNA-binding helices are more conserved than the dimerization domains (Figure 4-2). We therefore first needed to confirm that recombinant MHV CTD is largely dimeric in solution. Analytical sedimentation equilibrium ultracentrifugation was carried out to determine the quaternary structure of CTD (Figure 4-11). A simultaneous analysis of multiple scans of V321w CTD acquired at concentrations varying from 20-80 μM reveals that these data can be fit to a single component species model, with a MW = 32.9 kDa. Since the expected molecular weight is 33.2 kDa, this reveals that the MHV CTD is predominantly dimeric under these solution conditions (25 mM potassium phosphate, 20 mM KCl, pH 8.0).

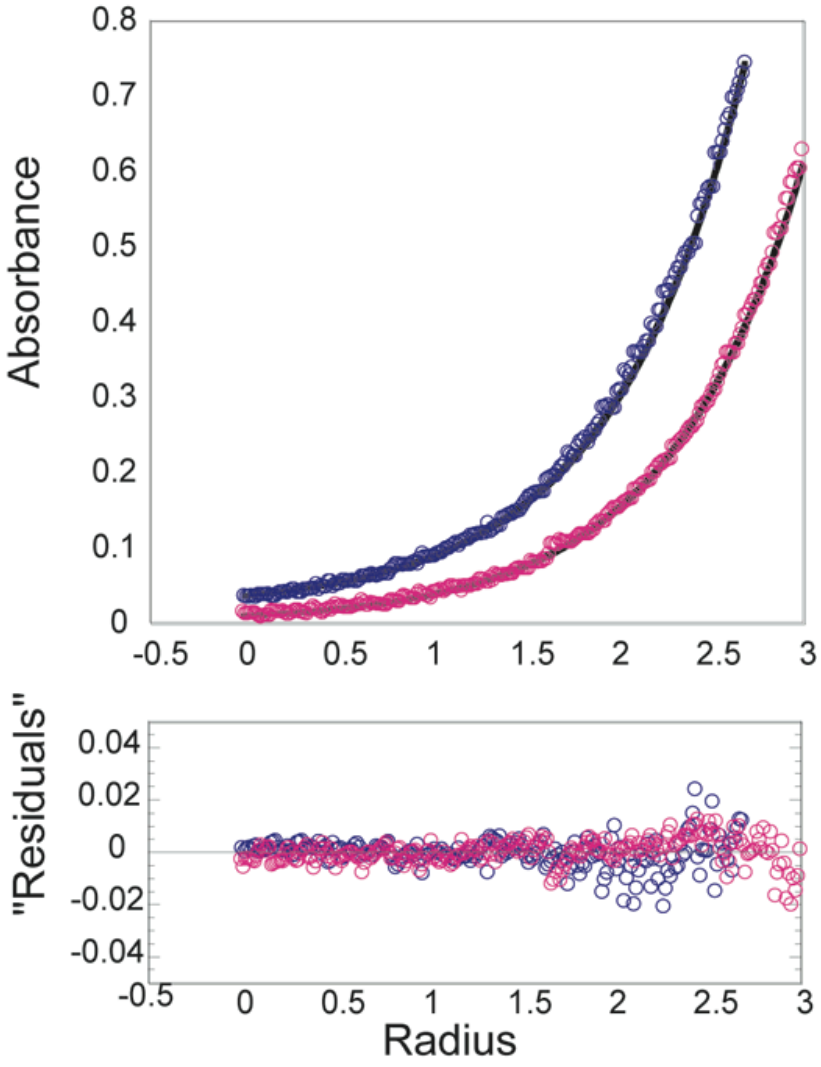


Figure 4-11 Assembly state of V321W CTD in solution. A representative AUC sedimentation equilibrium experiment (80 μM V321w CTD in 20 mM potassium phosphate, 20 mM KCl, pH 8.0, speed 27,000 rpm) scanned at 260 (blue) and 280 (red) nm is shown MW= 32.9 kDa (33.2 kDa expected).

Nonspecific RNA binding properties of V321W and T276W MHV CTDs

Since our MHV CTD (256-385) contains no aromatic (Trp or Tyr) residues, we incorporated one Trp into each of two predicted solvent exposal sites, V321 in the αD helix (V321w CTD) and one in the more N-terminal αF helix, T276 (T276w) (Figure 4-2), in or near a putative RNA-binding surface (Figure 4-1C). Our strategy was to monitor binding of Trp-mutant CTDs by measuring the quenching of the intrinsic Trp fluorescence upon binding model homopolymeric nucleic acids, e.g., poly(A) or poly(dT). This is a well-established methodology to examine the binding site affinity, occluded site size, and degree of cooperativity of single-stranded nucleic acid binding protein without complications from nucleic acid structure (Giedroc et al., 1990).

Representative reverse titrations of V321W and NV321W (Figure 4-12A) and T276W and NT276W (Figure 4-12B) proteins with poly(A) are shown, plotted as Q_{obs} vs [poly(A)]/[protein monomer] ratio. The data were fit to the non-cooperative McChee-Von Hippel large ligand infinite lattice model to obtain Q_{max}, the maximal fluorescence quenching, the apparent saturation point, n_{obs}, and K_{obs} (Table 4-3), see Appendix III. The solid line represents an unconstrained fit to the data with Q_{max}, n_{obs}, and K_{obs} optimized; these fits suggest an occluded site size of ≈ 3.8 ± 0.2. Thus, each monomer binds about 4 nucleotides (8 nucleotides/dimer) with K_{obs} of 2.4×10^6 M^{-1} (4.8×10^6 M^{-1} per dimer). Additional titrations with different concentrations of V321W (0.6 μM and 0.3 μM monomer) were also carried out, and the same occluded site size and Q_{max} were obtained (data not shown). A similar site-size (n_{obs}=3.3 ± 0.2 nucleotides) and K_{obs} within 3-fold of V321W CTD are obtained for the T276W CTD (Figure 4-12B),

suggesting that the Trp substitution are not influential to binding.

Analogous experiments carried out for corresponding intact N-domain models, NV321W and NT276W reveal somewhat unexpectedly, a smaller apparant site size (n_{obs} ≈ 1.4 ± 0.2, 3-4 nucleotides/dimer) for this larger protein on the basis of an unconstrained fit. This prompted us to refit the binding isotherms for V321w and T276w CTDs with n_{obs} fixed at 2 nucleotides/monomer (dashed lines, Figure 4-12). The fit suggests that the site size is not unambiguously defined by this experiment. If one assumes the same site size (3-4 nucleotides per dimer) for both intact N model and the CTD, then K_{obs} is ~6-fold larger for the intact N-model than for the CTD alone. The Q_{max} is also considerably smaller, consistent with the fact that the NTD contributes four additional Trp residues, none of which are sensitive to single strand RNA binding (data not shown).

Similar results are obtained with another homopolymer DNA lattice poly(dT) (Figure 4-13), with V321W and NV321W proteins, with the occluded site size perhaps slightly smaller (n_{obs} = 3.0 ± 0.2, 1.4 ± 0.2/monomer; 6 and 3 per dimer). As expected from other ss nucleic acid binding proteins, the Q_{max} is considerably larger for poly(dT) vs. poly(A) (Kumaran et al., 2006). However, in striking contrast to other single strand nucleic acid binding proteins, K_{obs} for poly(A) and poly(dT) are apparently similar. The reason for this is unknown but suggests a binding mode dominated by electrostatic interactions rather than base-specific interaction (see Discussion).

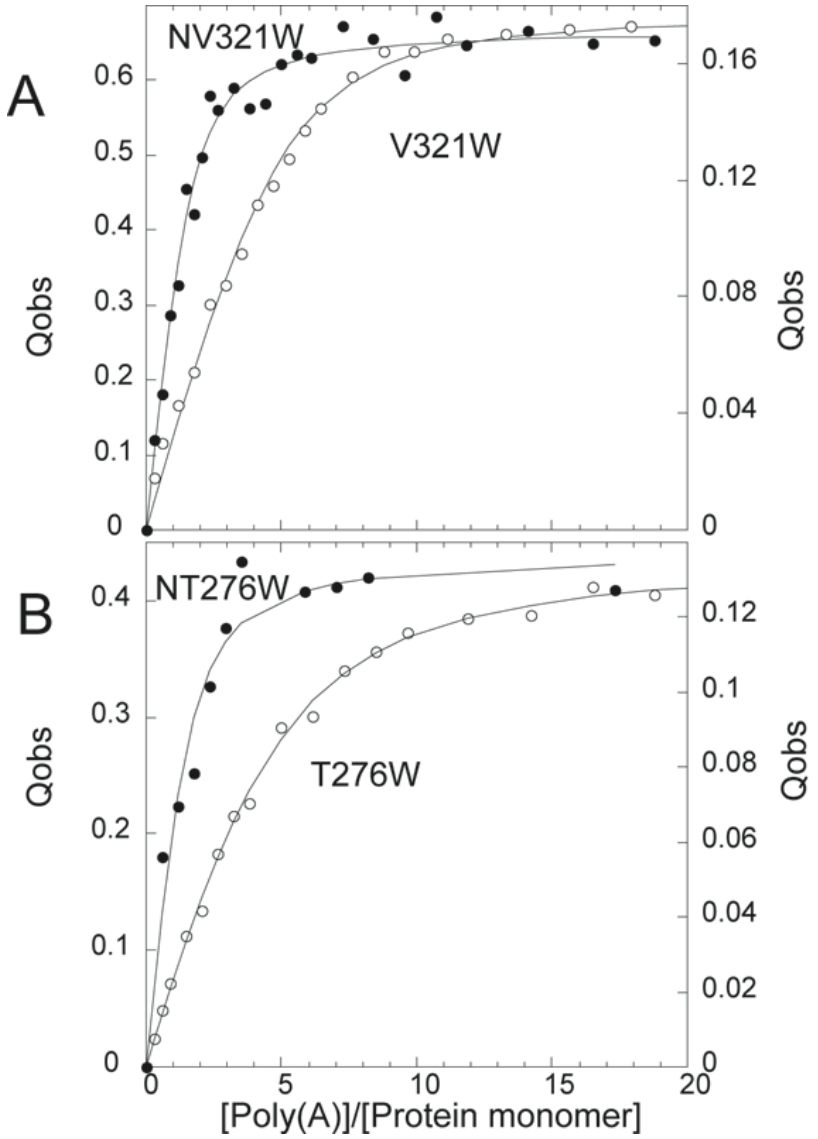


Figure 4-12 Reverse fluorescence quenching experiments carried out with 1 μM protein monomer as indicated with poly(A). Conditions: 25 mM potassium phosphate, 20 mM KCl, pH 8.0. Q_{obs} for the left axis is for V321W and T276W (empty circle), and Q_{obs} for the right axis is for NV321W and NT276W (solid circle). The solid line represents a fit in which Q_{max}, K_{obs} and n_{obs} were allowed to float in the fit; the dashed line refers to a fit when n_{obs} is fixed at 2 and only Q_{max} and K_{obs} are optimized. Parameters are compiled in Table 4-3.

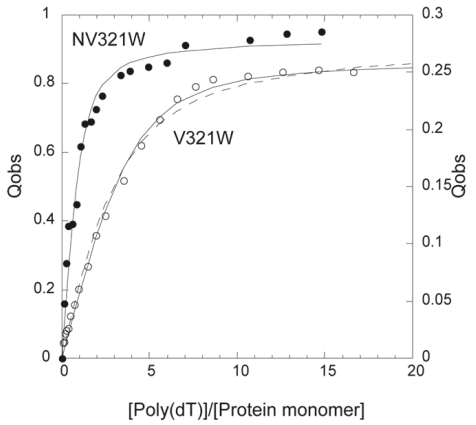


Figure 4-13 Reverse fluorescence quenching experiments carried out with 1 μM protein monomer as indicated with poly(dT). Conditions: 25 mM potassium phosphate, 20 mM KCl, pH 8.0. Q_{obs} for the left axis is for V321W (empty circle), and Q_{obs} for the right axis is for NV321W (solid circle). The solid line represents a fit in which Q_{max}, K_{obs} and n_{obs} were allowed to float in the fit; the dashed line refers to a fit when n_{obs} is fixed at 2 and only Q_{max} and K_{obs} are optimized. Parameters are compiled in Table 4-3.

Talbe 4-3 Summery of nonspecific nucleic binding parameters for CTD-containing constructs.

Lattice	Protein	Q_{obs}	$K_{obs}(x10^6\ M^{-1})$	n_{obs}
poly(A)	V321w	0.69(0.006)/0.73(0.01)[a]	2.4(0.6)/0.7(0.06)[a]	3.8(0.2)/2[a]
	T276w	0.44(0.004)/0.46(0.005)[a]	1.0(0.1)/0.5(0/03)[a]	3.3(0.2)/2[a]
	NV321w	0.17(0.003)	4.0(1.6)	1.4(0.2)
	NT276w	0.14(0.009)	2.9(2)	1.3(0.5)
poly(dT)	V321w	0.88(0.015)/0.92(.0.015)[a]	2.0(0.6)/0.8(0.09)[a]	3.0(0.2)/2[a]
	T276w	0.72(0.028)	0.4(0.1)	5.0 (0.7)
	NV321w	0.27(0.021)	5.5(1.0)	1.9(0.7)

[a] indicates the data obtained when n=2 is fixed. Standard error is shown in parenthesis.

Discussion

N is involved in many functions in the coronavirus life cycle. Despite moderate sequence similarity, all are predicted to share a number of common features. The abundant arginine and lysine residues make N very basic, with an isoelectric point of ~9-10, consistent with its primary role in RNA binding. Although the atomic resolution structures of NTD and CTD from several coronaviruses are now available, knowledge of the RNA binding properties of N is rather limited. These studies establish that there are multiple RNA binding domains in N that are structurally and functionally distinct, including the NTD, the SR-rich domain, and the N-terminal region of the CTD. In this work, we use a divide-and-conquer strategy to characterize the RNA binding properties of several N domain fragments.

Although previous work suggested that SR-rich region of MHV N specifically bound the TRS-leader sequence (Nelson et al., 2000), our studies of N197 and N219 are more consistent with a model where OB-fold like NTD provides the specificity determinant for recognition of the TRS, with the SR-rich region stabilizing the complex via non-specific electrostatic interactions. Any substitution in TRS sequence results in a reduced binding affinity of the MHV NTD. The triple adenosine motif ^{68}AAA^{70} on the 3' side of the TRS is a critical determinant for mediating high affinity binding to the NTD; however, the upstream pyrimidine triplet ^{65}cuu^{67} and the downstream ^{68}AAA^{70} appear to work in concert to maintain high affinity binding, since mutations in one or the other site are not additive in free energy terms. A68 and A70 appear more important than the center A69 since A68g and A70g substitutions only weakly destabilizing ($\Delta\Delta G$=0.4 kcal mol^{-1}); on the other hand the effect of right substitutions are nearly additive. This is

consistent with a minimum RxR motif as the preferred NTD interaction site.

These studies also establish that Y127 on the predicted β-sheet β3 is a major determinant for specific interaction with the TRS RNA. Y127 is one Tyr in a highly conserved region of W^{127}-Y-F^{129}-Y-Y in N. Based on the structure of IBV N, the side chains of F128 and Y130 are predicted to point into the hydrophobic core and stabilize the fold of the NTD; this leaves Y127 (Y92 in IBV N) and Y129 (Y94 in IBV N) pointing into the solvent where they could engage in specific interactions with nucleotides, via hydrogen bonding or base stacking, as is commonly observed in other proteins that recognize single strand nucleic acids (Tan et al., 2006). W126 (see Figure 4-14A), in contrast, points away from the surface. It is striking that R110 (R76 in IBV N), Y127 (Y92 in IBV N) and Y129 (Y94 in IBV N) in MHV N are predicted to form a nearly contiguous surface on the β-platform (Figure 4-14B). This picture provides a molecular level explanation for our mutagenesis experiments and suggests that those three residues function as a cooperative unit with a nearest neighbor coupling free energy between Y127 and Y129 of 0.44 kcal mol^{-1}, and a next nearest neighbor coupling free energy between R110 and Y129 of 0.27 kcal mol^{-1}; in this picture, Y127 likely mediates this long-range coupling. Strikingly, a Y127a substitution would be expected to globally perturb this network; this is exactly what is found. Interestingly, an Ala substitution of the residue analogous to Y127 in IBV N, Y92A, leads to a profound reduction in viral replication and in particular, subgenomic RNA synthesis, which is abolished. In the same study, Ala substitutions of R76 (R110 in MHV) and Y94 (Y129 in MHV) were also shown to lead to reduction in RNA-binding activity on the basis of a

dot-plot experiment (Tan et al., 2006). Thus, the data taken collectively strongly suggest that the NTD-TRS complex studied here is a structurally required intermediate during subgenomic RNA synthesis mediated by template switching. The high-resolution structure of NTD-TRS-L3 complex will allow us to elucidate the molecular mechanism of NTD associated regulation of sgRNA transcription. The high quality of the ^{15}N-^{1}H HSQC spectrum of N197-TRS-L3 complex (Figure 4-5C) would appear to suggest that this represents an ideal structural target.

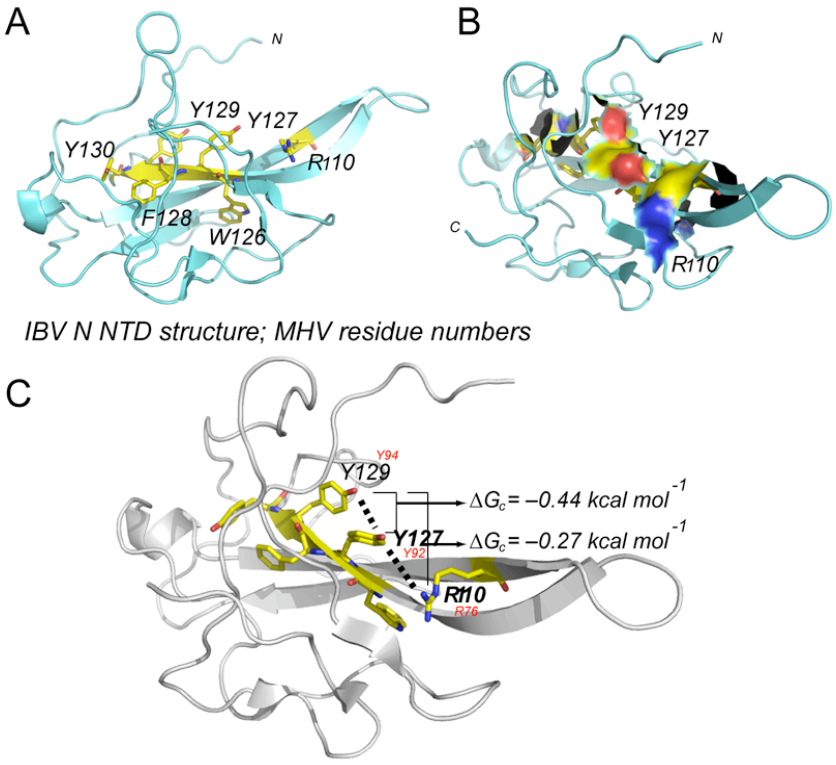


Figure 4-14. Ribbon diagrams of the crystal structure of the NTD from IBV N (PDB code: 2BTL). Residues are labeled according to the MHV sequence (see Figure 4-2 for sequence alignment). The structure of MHV N is as yet unknown.

It is also known that coronavrius N encapsidates viral RNA into ribonucleoprotein (RNP) particle (Narayanan et al., 2003) and SARS-CoV N has been indicated in playing an essential role in viral RNA packaging (Hsieh et al., 2005); however, the mechanism of RNA packaging is far from clear. The most recent structural study of the SARS-CoV CTD led the authors to speculate that the CTD plays key role in the helical nucleocapsid filament assembly (Chen et al., 2007). Consistent with this, we show that the CTD binds single stranded nucleic acid with an affinity in the μM range, and an occluded site size of ~2-4 residues per monomer or 4-8 per dimer. Interestingly, our full length N model (residues 60-385) incorporating the NTD, SR-rich region and the CTD, appears to increase the non-specific RNA binding affinity of the CTD subject to the content that the site size by fixed at the same value (~ 2 nucleotides/monomer or 4 nucleotides/dimer). Interestingly, potential RNA binding path across the surface of the IBV CTD reveals an approximate end-to-end distance of 24 - 48 Å, depending on the precise orientation of the RNA binding groove in the CTD (see Figure 4-15). At 5.5 Å per nucleotide in an extended conformation, this is large enough to accommodate 4-8 nucleotides/dimer (2-4 per monomer), consistent with results presented here.

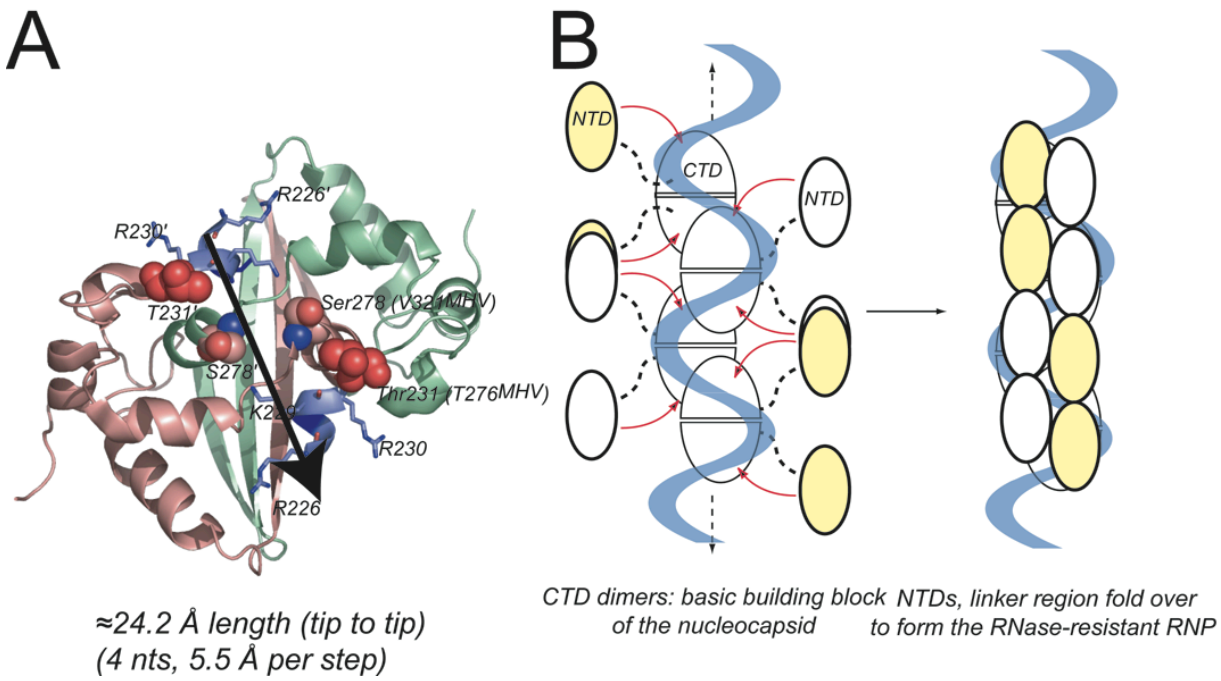


Figure 4-15 Genomic RNA binding model of N in the formation of the helical nucleocapsid. (A) Schematic view of a potential RNA binding path on the CTD, consistent with significant quenching of the T276w and V321w MHV CTD constructs. T231 (T276 in MHV) and S278 (V321 in MHV) are shown. Using IBV CTD (pdb: 2GE7) as a model. The shortest length from tip to tip is measured about 24 Å, which will accounts about 4 nucleotides (2 nucleotides per monomer). (B) Cartoon model of CTD-driven non-specific RNA binding mode.

Although a detailed molecular understanding of assembly requires further investigation, we speculate that N binds the phosphate backbone of the viral RNA nonspecifically through the CTD dimer building block mainly through the electrostatic interactions (Figure 4-15B). The flexible linker between the CTD and NTD may then allow the NTD to form the helical core and provide additional binding affinity via interaction with the bases. This will allow the NTD to form a cap on the helical filament

to prevent degradation by intracellular RNases and perhaps mediate compaction of the structure. We note that the phosphorylation of the Ser residue(s) in the SR-rich region documented to occur in TGEV N (Calvo et al., 2005) could potentially modulate this conformational switch. In the absence of a competing high affinity RNA binding site, the NTD may only be loosely held and thus can search for other high affinity sites in close physical proximity, e.g., a TRS-L or TRS-B. This binding in turn might induce a conformation change of the TRS that enhancing the rate of TRS-L/cTRS-B annealing required for template switching in the subgenomic RNA synthesis (Figure 4-16). Fluorescence resonance energy transfer (FRET)-based approaches would seem to be ideal way to monitor interdomain NTD-CTD interactions as well as the assembly/disassembly equilibrium on a suitably labeled nucleic acid substrate. The later approach has been successfully applied to the investigation of the various ssDNA binding modes of *E. coli* SSB protein (Roy et al., 2007).

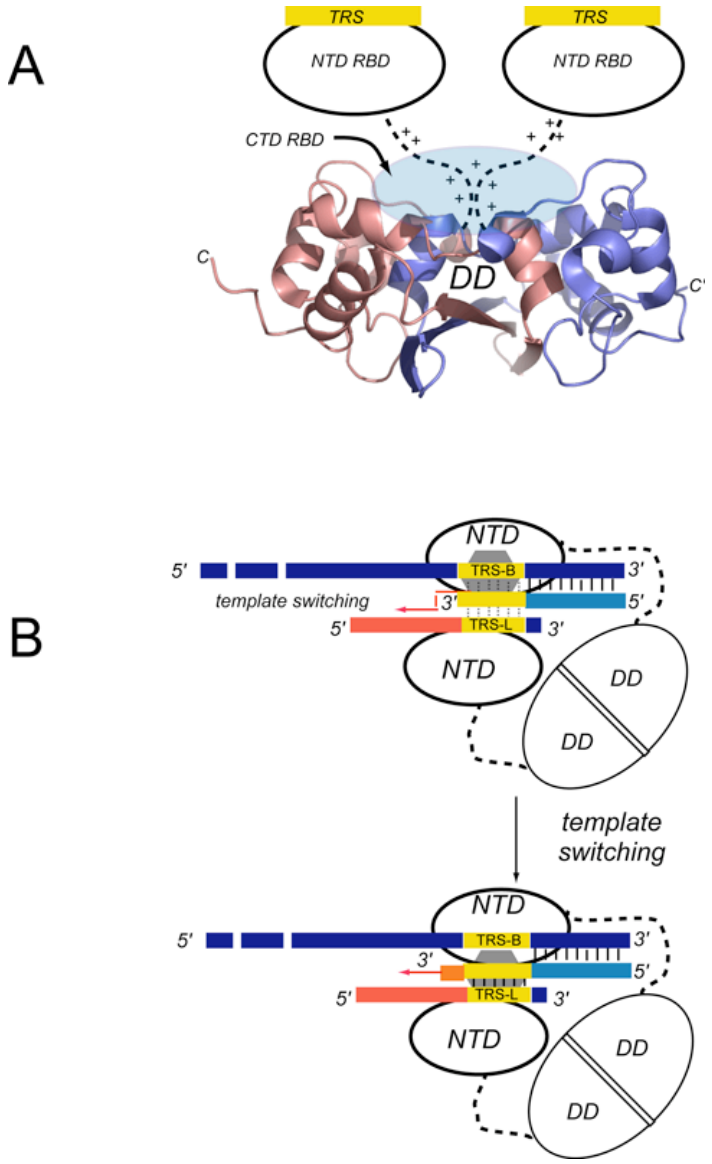


Figure 4-16 Wrapping/unwrapping model. In this binding model (A) the N binds the RNA nonspecifically, mainly through CTD and/or SR-region nonspecific binding. The NTD then undergoes a search for higher affinity ligand and binds specifically once the TRS is located. The fact that the CTD and the intact N is a dimer makes it possible that two long-distance TRS can be brought in close juxtaposition by specific binding with the NTD. (B) The N then functions as a RNA chaperone to medicate TRS-L-cTRS-B hybridization and thus template switching (Enjuanes et al., 2006).

CHAPTER V
SUMMARY AND PERSPECTIVE

Summary

NMR spectroscopy provides a powerful tool with which to quickly establish base-pairing schemes, significant insight into the base-paring formation, the degree of conformation heterogeneity, as well as residue-specific dynamic properties that influence the structure and function of RNA molecules. In the work presented in this dissertation, NMR data was used to establish experimental support for SL1 and SL2 in three group 2 CoVs, including MHV, BCoV and HCoV-OC43. SL2 is absolutely conserved in all CoVs, typically containing a pentaloop (C47-U48-U49-G50-U51 in MHV) stacked on a 5-bp stem, with some sequences containing an additional uridine residue 3' to U51; SL2 therefore possesses sequence features consistent with a U-turn-like conformation. The imino protons of U48 in the WT RNA, and G48 in the U48G SL2 mutant RNA, are significantly protected from exchange with solvent, consistent with a hydrogen bonding interaction critical to the hairpin loop architecture. In addition, an extensive mutagenesis analysis *in vivo* carried in collaboration with Dr. Leibowitz's group was also used to support that SL2 forms a U-turn-like structure and is involved in the regulation of RNA synthesis. Determination of the high resolution solution structure is in progress.

NMR data for SL1 from MHV and BCoV/OC43 provided experimental support for the presence of two or three non-canonical U•U, U•C or A+•C base pairs in the middle of the helical stem. Our NMR data are consistent with the basic stem-loop

architecture, including the noncanonical pairing in the helical stem. The SL1 structure is reminiscent of Stem-loop D (SLD) found in the 5' UTR of piconaviruses; however, NMR spectra show that SL1 is more dynamic and/or conformationally averaged than SLD. Interestingly, a mutant MHV genome containing a single nucleotide deletion of A35, an extrahelical nucleotide that destabilizes SL1 by 3.0 kcal mol^{-1}, gives rise to genetically unstable virus, which results in second-site mutations in SL1 itself as well as in a specific sequence in the 3' UTR. The second site mutations in SL1 are collectively characterized by one property, that they are predicted to be destabilized relative to the parent ΔA35 virus. Thermal denaturation revealed that the lower half of SL1 is thermodynamically unstable, in that the wild-type RNA is destabilized by 3.0 kcal mol^{-1} relative the ΔA35 mutant, while the three second site revertants are destabilized by 2.1 (ΔA35/U33C), 4.5 (ΔA35/C34U) and 4.7 (ΔA35/A36U) kcal mol^{-1} relative to parent ΔA35 virus respectively. Imino proton spectra revealed conformational heterogeneity in wild-type and the second-site mutants; and imino proton solvent exchange experiments indicated that second-site SL1-ΔA35 substitutions recover one of more characteristic features of the wild-type SL1. A recent model for coronavirus replication (Zuniga et al., 2004) postulated that circularization of the genome is a necessary early step in sgRNA synthesis. However, there had been little direct evidence in support for or against a physical association of the 5' and 3' ends of the genome. Through characterization of a ΔA35 virus, we present the first genetic evidence in support of a direct interaction between SL1 and the extreme 3' end of the genome; furthermore, the physical analysis of RNAs harboring second site revertants in SL1 recovered from the genetically unstable

SL1-Δ35 viruses, allows us to propose a "dynamic SL1" model, in which the conformational ensemble or longer-lived partially opened form(s) of SL1, lowers the energy barrier for formation of a complex with a hypothetical protein, which in turn drives genome circularization and sgRNA synthesis. These general characteristics of SL1 appear to be conserved in other coronaviral genomes.

In the virus life cycle, there are many viral and cellular proteins involved in the RNA replication. Nucleocapsid protein (N) plays a critical role in multiple functions. Recent studies on N suggest that it contains multifunctional RNA binding domains. In the work presented in this dissertation, we investigate the RNA-binding properties of different RNA-binding domains in the MHV N. Our results clearly indicate that the N-terminal domain (NTD) of MHV N specifically binds to the MHV TRS-L3 sequence. Y127 on β3 strand plays a significant role in this specific interaction. Compared with previously reported NTD studies in IBV, our data strongly suggest that MHV N directly regulates sgRNA synthesis by specifically interacting with aTRS-containing RNA. The SR-rich region, previously suggested to be the sole region for the specific interaction with TRS, simply increases the TRS binding affinity of N. The putative nonspecific RNA-binding property of MHV C-terminal domain (CTD) (residues 256-385) was also investigated. The MHV CTD was found to form a stable dimer in solution. The CTD binds large poly(A) with site-association constant of 10^6 M^{-1}, with the presence of NTD and SR-rich region changing the structure of the complex and increasing the affinity of N for nonspecific nucleic acids. We speculate that the NTD play an accessory role in the

genome RNA assembly, and plays a key role in the regulation of subgenomic transcription.

Perspective

Although our strategy of dissection of a large RNA into component subdomains is a standard approach to structure determination, many interesting questions remain that require additional investigation. For example, how are these stem-loop structures arranged relative to one another in the context of the entire 5'UTR? Are there significant tetiary interactions between individual domains that would stabilize a more compact form of the UTR? Certainly, the 5' UTR of retrovirus HIV-1 and MMTV appears to function as a series of stem loop structures that undergo global conformational change required for genome dimerization and packaging (D'Souza and Summers, 2004; Huthoff and Berkhout, 2002). Preliminary NMR experiments of the entire 210-nt 5'UTR of HCoV-OC43 suggest that at least under the conditions examined, there may be little interactions between component stem loops of the UTR. This is clearly the case for a 48 nucleotides SL1-SL2 containing fragment (Chapter II). Investigation of the global structure of the 5' UTR can be addressed in a number of ways. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) can probe the structure of folded RNA with an acylating agent N-methylisatoic anhydride (NMIA) which reacts preferentially with more flexible 2' OH found in unpaired or unstructured regions at single-nucleotide resolution under diverse solution conditions (Wilkinson et al., 2006). The development of segmental isotopic labeling strategies provides a useful tool to study the larger RNA

molecules by NMR spectroscopy. By the segmental ^{15}N-labeling of the 25KDa HCV IRES domain II, the Puglisi group has shown that the isolated extended helical domain II of the HCV IRES adopts the same fold in the context of the entire 100 KDa HCV IRES (Kim et al., 2002). We could potentially use the same strategy to segmentally isotope label SL1 & SL2, and investigate its conformation in the context of the entire 5' UTR. Furthermore, by segmentally labeling more than one helical element, we could define the orientation of each domain relative to one another in the entire 5' UTR using residual dipolar coupling (RDC) (D'Souza and Summers, 2004). FRET experiments could also be used to investigate the conformation of the 5'UTR as well as the kinetics of the folding of 5' UTR. We might also crystallize the entire 5'UTR RNA and to get the crystal structure of it. These experiments might be more insightful once virally encoded or host protein factors are identified which interact with the 5' UTR. Recently, the Zhuang group used single molecule FRET to investigate the step-wise protein-induced RNA folding assembly mechanism of the tolemerase RNP (Stone et al., 2007). Thus, it might be reasonable to speculate that the 5' UTR of coronavirus behaves in a similar manner; indeed a loosely folded structure dominated by secondary structure might more readily undergo conformational change required for various biological processes in the viral life cycle.

Since our data clearly reveal a genetic interaction mediated in some way by SL1 between the 5' and 3' ends of the coronavirus genome, the next step is to attempt to biochemically characterize a RNA complex between the intact 5' UTR and 3' UTR associated protein factors. A good place to start the experiments would be to use RNA

ligation strategies to introduce single nucleotide substitution with “zero-length” photoactivatible cross-linking groups at specific sites guided by our structural studies. These might be placed on the 5’ side of base of SL1, U49 in the SL2 loop, in the TRS region itself, or at the site of the reversion mutation in the 3’ UTR. This would allow us to determine if these nucleotides mediate specific protein-RNA or RNA-RNA interactions that drive circularization of the genome. In coronaviruses, PTB and hnRNP A1 have been suggested to bind the complimentary sequences in the 5’ UTR and 3’ UTR, respectively, and they may function together in promoting circularization. Our data suggest that SL1 may interact with hnRNP A1 to drive the circularization.

The mechanisms and factors involved in the replication of positive stranded RNA viruses remain unclear. In the vast majority of current studies from other positive strand RNA viruses, circularization of the genome is a very efficient way for the virus to perform translation, subgenomic RNA transcription, and genome replication, and may be a general replication mechanism used by these viruses. In tombusviruses, small plus-strand RNA viruses, a long-distance RNA-RNA interaction has been shown to stimulate the replication using a very interesting assay, in which the two RNA strands were tethered together with a replication enhancer (REN) present in the nontemplate RNA and a promoter located in the template RNA (Panavas and Nagy, 2005). In poliovirus, RNA replication requires a protein-protein bridge to mediate genome circularization (Bedard and Semler, 2004). In fact, the initiation of negative strand RNA synthesis requires a 3' poly(A) tail, while also requiring a cloverleaf-like 4-helix junction RNA structure located in the 5’ UTR. An RNP complex formed around the 5' cloverleaf RNA structure

through $3CL^{pro}$ bound to SLD and PCBP2 bound to a C-rich region in the same structure recruits the poly(A) binding protein bound to the 3' poly(A) tail, thus linking the ends of the viral RNA and effectively circularizing it.

A direct RNA-RNA interaction between the 5’ UTR and 3’UTR is also possible. Since the SL1, SL2 and SL4 of MHV and SARS-CoV are functionally interchangeable (Kang et al., 2006), the possible direct interaction site might involve the highly conserved SL2 and/or the TRS region. A highly conserved overlapping bulged stem loop/ pseudoknot structure in the 3’ UTR has been suggested to participate in a conformational switch function in viral replication. Interestingly, the sequence in the long loop 1 (L1) of the pseudoknot conformation (3’-UAGGUAAGA) is quite complementary to a region that partially overlaps core TRS sequence, 5’-AAUCUAAUCUAAC; this sequence correspond TRS-L5. Regulated base pairing between the TRS-L5 of the 5’ UTR and L1 of 3’UTR may drive 3’UTR into one or the other conformation (pseudoknot vs. double stem loop), thus facilitating the “switch” between replication and transcription via an as yet undefined mechanism. In fact, previous studies of L1 in MHV reveal that this region is functionally important since nucleotide insertions here lead to nonviable virus, for reasons that yet not immediately obvious from inspection of the 3’ UTR.

The work presented in this dissertation further refines our understanding of the *cis*-acting RNA motifs in the 5’ leader region of MHV, and makes predictions as to what viral and cellular protein might be involved in regulation of subgenomic RNA synthesis. One significant contribution to our understanding of this process is that our data strongly

suggest that N directly regulates subgenomic RNA transcription through specific interactions with the TRS. hnRNP-A1 has been suggested to interact with the PTB protein and the two together might help regulate RNA transcription. The putative binding sites for N and PTB in the 5'UTR are partially overlapped in the leader sequence. Does N interact directly with PTB as well? Or do they bind to the leader region in different stages of RNA replication and/or transcription? Further investigation of these and other questions will greatly enhance our understanding of the molecular mechanism of this complex process.

REFERENCES

AI-Hashimi, H. (2005) Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBioChem*, **6**, 1506-1519.

Almazan, F., Galan, C. and Enjuanes, L. (2004) The nucleoprotein is required for efficient coronavirus genome replication. *J Virol*, **78**, 12683–12688.

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J. and Hilgenfeld, R. (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *Embo J*, **21**, 3213-3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. and Hilgenfeld, R. (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763-1767.

Antao, V.P., Lai, S.Y. and Tinoco, I., Jr. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res*, **19**, 5901-5905.

Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V. and Lai, M.M.C. (1993) Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J Virol*, **67**, 6056-6063.

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.

Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N. and Stohlman, S.A. (1988) Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *J Virol*, **62**, 4280-4287.

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

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.

Barton, D.J., O'Donnell, B.J. and Flanegan, J.B. (2001) 5' cloverleaf in poliovirus RNA is a *cis*-acting replication element required for negative-strand synthesis. *Embo J,* **20**, 1439-1448.

Batey, R.T., Gilbert, S.D. and Montange, R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411-415.

Bedard, K.M. and Semler, B.L. (2004) Regulation of picornavirus gene expression. *Microbes and Infection*, **6**, 702-713.

Bhardwaj, K., Guarino, L. and Kao, C.C. (2004) The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor. *J Virol*, **78**, 12218-12224.

Bost, A.G., Carnahan, R.H., Lu, X.T. and Denison, M.R. (2000) Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J Virol*, **74**, 3379-3387.

Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.

Brow, D.A. (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet*, **36**, 333-360.

Cabello-Villegas, J., Tworowska, I. and Nikonowicz, E.P. (2004) Metal ion stabilization of the U-turn of the A37 N6-dimethylallyl-modified anticodon stem-loop of Escherichia coli tRNAPhe. *Biochemistry*, **43**, 55-66.

Calvo, E., Escors, D., Lopez, J.A., Gonzalez, J.M., Alvarez, A., Arza, E. and Enjuanes, L. (2005) Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells. *J Gen Virol*, **86**, 2255-2267.

Campanacci, V., Egloff, M.P., Longhi, S., Ferron, F., Rancurel, C., Salomoni, A., Durousseau, C., Tocque, F., Bremond, N., Dobbe, J.C., Snijder, E.J., Canard, B. and Cambillau, C. (2003) Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein. *Acta Crystallogr D Biol Crystallogr*, **59**, 1628-1631.

Campbell, D.O. and Legault, P. (2005) Nuclear magnetic resonance structure of the Varkud satellite ribozyme stem-loop V RNA and magnesium-ion binding from chemical-shift mapping. *Biochemistry*, **44**, 4157-4170.

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

Chang, C.K., Sue, S.C., Yu, T.H., Hsieh, C.M., Tsai, C.K., Chiang, Y.C., Lee, S.J., Hsiao, H.H., Wu, W.J., Chang, C.F. and Huang, T.H. (2005) The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett*, **579**, 5663-5668.

Chang, C.K., Sue, S.C., Yu, T.H., Hsieh, C.M., Tsai, C.K., Chiang, Y.C., Lee, S.J., Hsiao, H.H., Wu, W.J., Chang, W.L., Lin, C.H. and Huang, T.H. (2006) Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci*, **13**, 59-72.

Chang, R.Y. and Brian, D.A. (1996) cis requirement for N-specific protein sequence in bovine coronavirus defective interfering RNA replication. *J Virol*, **70**, 2201-2207.

Chang, R.Y., Hofmann, M.A., Sethna, P.B. and Brian, D.A. (1994) A *cis*-acting function for the coronavirus leader in defective interfering RNA replication. *J Virol*, **68**, 8223-8231

Chang, R.Y., Krishnan, R. and Brian, D.A. (1996) The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. *J Virol*, **70**, 2720-2729.

Chen, C.J., Sugiyama, K., Kubo, H., Huang, C. and Makino, S. (2004) Murine coronavirus nonstructural protein p28 arrests cell cycle in G0/G1 phase. *J Virol*, **78**, 10410-10419.

Chen, C.Y., Chang, C.K., Chang, Y.W., Sue, S.C., Bai, H.I., Riang, L., Hsiao, C.D. and Huang, T.H. (2007) Structure of the SARS coronavirus nucleocapsid protein

RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol*, **368**, 1075-1086.

Chen, X., Agarwal, A. and Giedroc, D.P. (1998) Structural and functional heterogeneity among the zinc fingers of human MRE-binding transcription factor-1. *Biochemistry*, **37**, 11152-11161.

Choi, K.S., Huang, P. and Lai, M.M. (2002) Polypyrimidine-tract-binding protein affects transcription but not translation of mouse hepatitis virus RNA. *Virology*, **303**, 58-68

Cologna, R. and Hogue, B.G. (2000) Identification of a bovine coronavirus packaging signal. *J Virol*, **74**, 580-583.

Coman, D. and Russu, I.M. (2004) Site-resolved stabilization of a DNA triple helix by magnesium ions. *Nucleic Acids Res*, **32**, 878-883.

Compton, S.R., Rogers, D.B., Holmes, K.V., Fetsch, D., Remenick, J. and McGowan, J.J. (1987) In vitro replication of mouse hepatitis virus strain A59. *J Virol*, **61**, 1814-1820.

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.

Cornish, P.V., Stammler, S.N. and Giedroc, D.P. (2006) The global structures of a wild-type and poorly functional plant luteoviral mRNA pseudoknot are essentially identical. *RNA*, **12**, 1959-1969.

Cruceanu, M., Gorelick, R., Musier-Forsyth, K., Rouzina, I. and Williams, M. (2006) Rapid kinetics of protein–nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol*, **363**, 867-877.

Dann III, C.E., 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.

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

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.

DeRose, V.J. (2003) Metal ion binding to catalytic RNA molecules. *Curr Opin Struct Biol*, **13**, 317-324.

Dhavan, G.M., Lapham, J., Yang, S. and Crothers, D.M. (1999) Decreased imino proton exchange and base-pair opening in the IHF-DNA complex measured by NMR. *J Mol Biol*, **288**, 659-671.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R. and Xu, R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev*, **13**, 1102-1115.

Donaldson, E.F., Graham, R.L., Sims, A.C., Denison, M.R. and Baric, R.S. (2007a) Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-

LA6 suggests that nsp10 plays a critical role in polyprotein processing. *J Virol*, **81**, 7086-7098.

Donaldson, E.F., Sims, A.C., Graham, R.L., Denison, M.R. and R.S., B. (2007b) Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *J Virol*, **81**, 6356-6368.

Draper, D.E. (2004) A guide to ions and RNA structure. *RNA*, **10**, 335-343.

D'Souza, V. and Summers, M.F. (2004) Structural basis for packaging the dimeric genome of moloney murine leukaemia virus. *Nature*, **431**, 586-590.

Du, Z., Yu, J., Andino, R. and James, T.L. (2003) Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackievirus B3. *Biochemistry*, **42**, 4373-4783.

Du, Z., Yu, J., Ulyanov, N.B., Andino, R. and James, T.L. (2004) Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry*, **43**, 11959-11972.

Dunham, C., Murray, J. and Scott, W. (2003) A helical twist-induced conformational switch activates cleavage in the hammerhead ribozyme. *J Mol Biol*, **332**, 327-336.

Egloff MP, F.F., Campanacci V, Longhi S, Rancurel C, Dutartre H, Snijder EJ, Gorbalenya AE, Cambillau C, Canard B. (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding

subunit unique in the RNA virus world. *Proc Natl Acad Sci U S A*, **101**, 3792-3796.

Eleouet, J.F., Slee, E.A., Saurini, F., Castagne, N., Poncet, D., Garrido, C., Solary, E. and Martin, S.J. (2000) The viral nucleocapsid protein of transmissible gastroenteritis coronavirus (TGEV) is cleaved by caspase-6 and -7 during TGEV-induced apoptosis. *J Virol*, **74**, 3975-3983.

Enjuanes, L. (2005) *Coronavirus Replication and Reverse Genetics*. Springer Berlin Heidelberg.

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.

Fan, H., Ooi, A., Tan, Y.W., Wang, S., Fang, S., Liu, D.X. and Lescar, J. (2005) The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure*, **13**, 1859-1868.

Fang, S., Chen, B., Tay, F.P., Ng, B.S. and Liu, D.X. (2006) An arginine-to-proline mutation in a domain with undefined functions within the helicase protein (Nsp13) is lethal to the coronavirus infectious bronchitis virus in cultured cells. *Virology*, **358**, 136-147.

Feig, A.L., Scott, W.G. and Uhlenbeck, O.C. (1998) Inhibition of the hammerhead ribozyme cleavage reaction by site-specific binding of Tb. *Science*, **279**, 81-84.

Fiala, R., Czernek, J. and Sklenar, V. (2000) Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *J Biomol NMR*, **16**, 291-302.

Fraser, C. and Doudna, J. (2007) Structural and mechanistic insights into hepatitis C viral translation initiation. *Nature Reviews Microbiology*, **5**, 29-38.

Frolov, I., Hardy, R. and Rice, C.M. (2001) *Cis*-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA*, **7**, 1638-1651.

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

Giedroc, D.P., Khan, R. and Barnhart, K. (1990) Overexpression, purification, and characterization of recombinant T4 gene 32 protien22-301 (g32P-B). *J Biol Chem*, **265**, 11444-11455.

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

Goebel, S., Miller, T., Bennett, C., Bernard, K. and Masters, P. (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.

Goebel, S.J., Hsue, B., Dombrowski, T.F. and Masters, P.S. (2004a) 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.

Goebel, S.J., Taylor, J. and Masters, P.S. (2004b) 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.

Gonzalez, J.M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A.E. and Enjuanes, L. (2003) A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol*, **148**, 2207-2235.

Grentzmann, G., Ingram, J.A., Kelly, P.J., Gesteland, R.F. and Atkins, J.F. (1998) A dual-luciferase reporter system for studying recoding signals. *RNA*, **4**, 479-486.

Gueron, M. and Leroy, J.L. (1995) Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol*, **261**, 383-413.

Gutell, R.R., Cannone, J.J., Konings, D. and Gautheret, D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J Mol Biol*, **300**, 791-803.

He, R., Leeson, A., Andonov, A., Li, Y., Bastien, N., Cao, J., Osiowy, C., Dobie, F., Cutts, T., Ballantine, M. and Li, X. (2003) Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Commun*, **311**, 870-876.

He, R.L.A., Ballantine M, Andonov A, Baker L, Dobie F, Li Y, Bastien N, Feldmann H, Strocher U, Theriault S, Cutts T, Cao J, Booth TF, Plummer FA, Tyler S, Li X. (2004) Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. *Virus Res*, **105**, 121-125.

Herbert, T., Brierley, I. and Brown, T. (1997) Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. *J Gen Virol*, **78**, 1033-1040.

Hermann, T. and Patel, D. (1999) Stiching together RNA tertiary architectures. *J Mol Biol*, **294**, 829-849.

Herold, J. and Andino, R. (2001) Poliovirus RNA replication requires genome circularization through a protein-protein bridge. *Mol Cell*, **7**, 581-591.

Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429-3431.

Hofacker, I.L., Stadler, P.F. and Stocsits, R.R. (2004) Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics*, **20**, 1495-1499.

Hsieh, P.K., Chang, S.C., Huang, C.C., Lee, T.T., Hsiao, C.W., Kou, Y.H., Chen, I.Y., Chang, C.K., Huang, T.H. and Chang, M.F. (2005) Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol*, **79**, 13848-13855.

Hsue, B. and Masters, P.S. (1997) A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. *J Virol*, **71**, 7567-7578.

Hsue, B., Hartshorne, T. and Masters, P.S. (2000) Characterization of an essential RNA secondary structure in the 3' untranslated region of the murine coronavirus genome. *J Virol*, **74**, 6911-6921.

Huang, P. and Lai, M.M.C. (2001) Heterogeneous nuclear ribonucleoprotein A1 binds to the 3'-untranslated region and mediates potential 5'-3'-end cross talks of mouse hepatitis virus RNA. *J Virol*, **75**, 5009-5017.

Huang, Q., Yu, L., Petros, A.M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S.W. and Olejniczak, E.T. (2004) Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*, **43**, 6059-6063.

Huppler, A., Nikstad, L.J., Allmann, A.M., Brow, D.A. and Butcher, S.E. (2002) Metal binding and base ionization in the U6 RNA intramolecular stem- loop structure. *Nat Struct Biol*, **9**, 431-435.

Hurst, K.R., Kuo, L., Koetzner, C.A., Ye, R., Hsue, B. and Masters, P.S. (2005) A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. *J Virol*, **79**, 13285-13297.

Huthoff, H. and Berkhout, B. (2002) Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry*, **41**, 10439-10445.

Ihle, Y., Ohlenschlager, O., Hafner, S., Duchardt, E., Zacharias, M., Seitz, S., Zell, R., Ramachandran, R. and Gorlach, M. (2005) A novel cGUUAg tetraloop structure with a conserved yYNMGg-type backbone conformation from cloverleaf 1 of bovine enterovirus 1 RNA. *Nucleic Acids Res*, **33**, 2003-2011.

Ivanov, K. and Ziebuhr, J. (2004) Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J Virol*, **78**, 7833-7838.

Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J. and Ziebuhr, J. (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol*, **78**, 5619-5612.

Jayaram, H., Fan, H., Bowman, B.R., Ooi, A., Jayaram, J., Collisson, E.W., Lescar, J. and Prasad, B.V. (2006) X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol*, **80**, 6612-6620.

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M., Moy, K., Yadav, M.K., Velasquez, J., Buchmeier, M.J., Stevens, R.C. and Kuhn, P. (2006) Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J Virol*, **80**, 7894-7901.

Kadare, G. and Haenni, A. (1997) Virus-encoded RNA helicases. *J Virol*, **71**, 2583-2590.

Kang, H., Feng, M., Schroeder, M.E., Giedroc, D.P. and Leibowitz, J.L. (2006) Putative cis-acting stem-loops in the 5' untranslated region of the severe acute respiratory syndrome coronavirus can substitute for their mouse hepatitis virus counterparts. *J Virol*, **80**, 10600-10614.

Kanjanahaluethai, A. and Baker, S.C. (2000) Identification of mouse hepatitis virus papain-like proteinase 2 activity. *J Virol*, **74**, 7911-7921.

Kim, I., Lukavsky, P.J. and Puglisi, J.D. (2002) NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J Am Chem Soc*, **124**, 9338-9339.

Kim, J.G., Spence, R., Currier, P., Lu, X. and Denison, M. (1995) Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor E64d. *Virology*, **208**, 1-8.

Kim, Y.-N., Jeong, Y.S. and Makino, S. (1993) Analysis of cis-acting sequences essential for coronavirus defective interfering RNA replication. *Virology*, **197**, 53-63.

Koonin, E.V. and Dolja, V.V. (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol,* **28**, 375-430.

Kowalczykowski, S.C., Lonberg, N., Newport, J.W. and von Hippel, P.H. (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. I. Characterization of the binding interactions. *J Mol Biol*, **145**, 75-104.

Kumaran, S., Kozlov, A.G. and Lohman, T.M. (2006) Saccharomyces cerevisiae replication protein A binds to single-stranded DNA in multiple salt-dependent modes. *Biochemistry*, **45**, 11958-11973.

Kuo, L. and Masters, P.S. (2002) Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J Virol*, **76**, 4987-4999.

Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem*, **237**, 260-273.

Lai, M.M.C., Baric, R.S., Brayton, P.R. and Stohlman, S.A. (1984) Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc Natl Acad Sci U S A*, **81**, 3626-3630.

Lai, M.M.C. and Holmes, K.V. (2001) Coroniviradae: the viruses and their replication. In *Fields Virology*, Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B. and Straus, S.E. (eds.). Philadelphia, Lippincott, Williams and Wilkins, 1163-1189.

Lai, M.M.C., Patton, C.D. and Stohlman, S.A. (1982) Further characterization of mRANAs of mouse hepatitis virus: presence of comon 5'-end nucleotides. *J Virol*, **41**, 557-565.

Lebars, I., Yoshizawa, S., Stenholm, A.R., Guittet, E., Douthwaite, S. and Fourmy, D. (2003) Structure of 23S rRNA hairpin 35 and its interaction with the tylosin-resistance methyltransferase RlmAII. *Embo J*, **22**, 183-192.

Lee, J.H. and Pardi, A. (2007) Thermodynamics and kinetics for base-pair opening in the P1 duplex of the Tetrahymena group I ribozyme. *Nucleic Acids Res*, **35**, 2965-2974.

Leibowitz, J.L., Wilhelmsen, K.C. and Bond, C.W. (1981) The virus-specific intracellular RNA species of two murine coronaviruses: MHV-A59 and MHV-JHM. *Virology*, **114**, 39-51.

Leontis, N.B. and Westhof, E. (2003) Analysis of RNA motifs. *Curr Opin Struct Biol*, **13**, 300-308.

Lescrinier, E., Nauwelaerts, K., Zanier, K., Poesen, K., Sattler, M. and Herdewijn, P. (2006) The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res*, **34**, 2878-2886.

Li, F., Li, W., Farzan, M. and Harrison, S.C. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, **309**, 1864-1868.

Li, H.P., Huang, P., Park, S. and Lai, M.M. (1999) Polypyrimidine tract-binding protein binds to the leader RNA of mouse hepatitis virus and serves as a regulator of viral transcription. *J Virol*, **73**, 772-777.

Li, H.P., Zhang, X., Duncan, R., Comai, L. and Lai, M.M. (1997) Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription- regulatory region of mouse hepatitis virus RNA. *Proc Natl Acad Sci U S A*, **94**, 9544-9549.

Lin, Y.-J. and Lai, M.M.C. (1993) Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontinous sequence for replication. *J Virol*, **67**, 6110-6118.

Liu, P., Li, L., Millership, J.J., Kang, H., Leibowitz, J.L. and Giedroc, D.P. (2007) A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA*, **13**, 763-780.

Liu, Q., Johnson, R.F. and Leibowitz, J.L. (2001) Secondary structural elements within the 3' untranslated region of mouse hepatitis virus strain JHM genomic RNA. *J Virol*, **75**, 12105-12113.

Lu, X.T., Sims, A.C. and Denison, M.R. (1998) Mouse hepatitis virus 3C-like protease cleaves a 22-kilodalton protein from the open reading frame 1a polyprotein in virus-infected cells and in vitro. *J Virol*, **72**, 2265-2271.

Lukavsky, P.J., Kim, I., Otto, G.A. and Puglisi, J.D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol*, **10**, 1033-1038.

Luo, H., Chen, Q., Chen, J., Chen, K., Shen, X. and Jiang, H. (2005a) The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett*, **579**, 2623-2628.

Luo, H., Ye, F., Chen, K., Shen, X. and Jiang, H. (2005b) SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. *Biochemistry*, **44**, 15351-15358.

Luo, H., Chen, J., Chen, K., Shen, X. and Jiang, H. (2006a) Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry*, **45**, 11827-11835.

Luo, H., Wu, D., Shen, C., Chen, K., Shen, X. and Jiang, H. (2006b) Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. *Int J Biochem Cell Biol*, **38**, 589-599.

Luytjes, W., Gerritsma, H. and Spaan, W.J.M. (1996) Replication of synthetic defective interfering RNAs derived from coronavius mouse hepatitis virus-A59. *Virology*, **216**, 174-183.

Mandal, M. and Breaker, R.R. (2004) Adanine riboswitches and gene activation by disruption of a transcription termination. *Nat Struct Mol Biol*, **11**, 29-35.

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

McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymer*, **29**, 1105-1119.

Micura, R. and Hobartner, C. (2003) On secondary structure rearrangements and equilibria of small RNAs. *Chem Bio Chem*, **4**, 984-990.

Moore, P. (1999) Structural motifs in RNA. *Annu Rev Biochem*, **68**, 287-300.

Mueller, L., Legault, P. and Pardi, A. (1995) Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino protons. *J Am Chem Soc*, **117**, 11043-11048.

Myers, J.C. and Shamoo, Y. (2004) Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM). *J Mol Biol*, **342**, 743-756.

Narayanan, K., Kim, K.H. and Makino, S. (2003) Characterization of N protein self-association in coronavirus ribonucleoprotein complexes. *Virus Res*, **98**, 131-140.

Nelson, G.W., Stohlman, S.A. and Tahara, S.M. (2000) High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J Gen Virol*, **81**, 181-188.

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.

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

Nixon, P.L., Rangan, A., Kim, Y.G., Rich, A., Hoffman, D.W., Hennig, M. and Giedroc, D.P. (2002b) Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol*, **322**, 621-633.

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

Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. and Allain, F.H. (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054-2057.

Ohlenschlager, O., Wohnert, J., Bucci, E., Seitz, S., Hafner, S., Ramachandran, R., Zell, R. and Gorlach, M. (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure*, **12**, 237-248.

Orr, J., Hagerman, P. and Williamson, J. (1998) Protein and Mg2+-induced conformational changes in the S15 binding site of 16 s ribosomal RNA. *J Mol Biol*, **275**, 453-464.

Ortego, J., Escors, D., Laude, H. and Enjuanes, L. (2002) Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. *J Virol*, **76**, 11518-11529.

Panavas, T. and Nagy, P.D. (2005) Mechanism of stimulation of plus-strand synthesis by an RNA replication enhancer in a tombusvirus. *J Virol*, **79**, 9777-9785.

Pasternak, A., Spaan, W. and EJ, S. (2006) Nidovirus transcription: how to make sense...? *Journal of General Virology*, **87**, 1403-1421.

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.

Paulous, S., Malnou, C., Michel, Y., Kean, K. and Borman, A. (2003) Comparison of the capacity of different viral internal ribosome entry segments to direct translation initiation in poly(A)-dependent reticulocyte lysates. . *Nucleic Acids Res*, **31**, 722-733.

Peiris JS, GuanY and KY, Y. (2004) Severe acute respiratory syndrome. *Nature Medicine*, **10**, S88 - S97

Pennella, M.A., Shokes, J.E., Cosper, N.J., Scott, R.A. and Giedroc, D.P. (2003) Structural elements of metal selectivity in metal sensor proteins. *Proc Natl Acad Sci U S A,* **100**, 3713-3718.

Proctor, D.J., Schaak, J.E., Bevilacqua, J.M., Falzone, C.J. and Bevilacqua, P.C. (2002) Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry*, **41**, 12062-12075.

Puglisi, E.V. and Puglisi, J.D. (1998) HIV-1 A-rich RNA loop mimics the tRNA anticodon structure. *Nat Struct Biol*, **5**, 1033-1036.

Query, C.C. and Konarska, M.M. (2004) Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell*, **14**, 343-354.

Quigley, G.J. and Rich, A. (1976) Structural domains of transfer RNA molecules. *Science*, **194**, 796-806.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557-572.

Raman, S., Bouma, P., Williams, G.D. and Brian, D.A. (2003) Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J Virol*, **77**, 6720-6730.

Raman, S. and Brian, D.A. (2005) Stem-loop IV in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J Virol*, **79**, 12434-12446.

Ratia, K., Saikatendu, K., Santarsiero, B., Barretto, N., Baker, S., Stevens, R. and Mesecar, A. (2006) Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proc Natl Acad Sci U S A,* **103**, 5717-5722.

Ray, D., Wu, B., White, K.A. (2003) A second functional RNA domain in the 5' UTR of the Tomato bushy stunt virus genome: intra- and interdomain interactions mediate viral RNA replication. *RNA*, **9**, 1232-1245.

Rivas, E. and Eddy, S.R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Bio*, **285**, 2053-2068.

Robertson, M., Igel, R., Baertsch, D., Haussler, M., Ares, J. and Scott, W. (2005) The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol*, **3**, e5.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A.D., Drosten, C., Pallansch, M.A., Anderson, L.J. and Bellini, W.J. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, **300**, 1394-1399.

Roy, R., Kozlov, A.G., Lohman, T.M. and Ha, T. (2007) Dynamic structural rearrangements between DNA binding modes of E. coli SSB protein. *J Mol Biol*, **369**, 1244-1257.

Sashital, D.G., Venditti, V., Angers, C.G., Cornilescu, G. and Butcher, S.E. (2007) Structure and thermodynamics of a conserved U2 snRNA domain from yeast and human. *RNA*, **13**, 328-338.

Sawicki, S.G. and Sawicki, D.L. (1990) Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J Virol*, **64**, 1050-1056.

Sawicki, S.G. and Sawicki, D.L. (1998) A new model for coronavirus transcription. *Adv Exp Med Biol*, **440**, 215-219.

Sawicki, S. and Sawicki, D. (2005) Coronavirus transcription: a perspective. *Curr Top Microbiol Immunol*, **287**, 31-35.

Sawicki, S.G., D. L. Sawicki, D. Younker, Y. Meyer, V. Thiel, H. Stokes, and S. G. Siddell. (2005) Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLos Pathog.*, **1**, e39.

Schultes, E.A. and Bartel, D.P. (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science*, **289**, 448-452.

Scott, W. (2007) Ribozymes. *Curr Opin in Structural Biology*, **17**, 280-286.

Sefcikova, J., Drasovska, M.V., Sponer, J. and Walter, N.G. (2007) The genomic HDV ribozyme utilizes a previously unnoticed U-turn motif to accomplish fast site-specific catalysis. *Nucleic Acids Res*, **35**, 1933-1946.

Senanayake, S. and Brian, D. (1999) Translation from the 5' untranslated region (UTR) of mRNA 1 is repressed, but that from the 5' UTR of mRNA 7 is stimulated in coronavirus-infected cells. *J Virol*, **73**, 8003-8009.

Serganov, A., Yuan, Y., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A., Hobartner, C., Micura, R., Breaker, R. and Patel, D. (2004) Structural basis for

discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol*, **11**, 1729-1741.

Sethna, P.B., Hung, S.L. and Brian, D.A. (1989) Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc Natl Acad Sci U S A*, **86**, 5626-5630.

Seybert, A., Posthuma, C.C., van Dinten, L.C., Snijder, E.J., Gorbalenya, A.E. and Ziebuhr, J. (2005) A complex zinc finger controls the enzymatic activities of nidovirus helicases. *J Virol*, **79**, 696-704.

Shi, S.T., Huang, P., Li, H.-P. and Lai, M.M.C. (2000) Heterogeneous nuclear ribonucleoprotein A1 regulates RNA synthesis of a cytoplasmic virus. *Embo J*, **19**, 4701-4711.

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

Shirako, Y. and Strauss, J.H. (1994) Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. *J Virol*, **68**, 1874-1885.

Sigel, R. and Pyle, A. (2007) Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem Rev*, **107**, 97-113.

Simorre, J.P., Zimmermann, G.R., Pardi, A., Farmer, B.T.n. and Mueller, L. (1995) Triple resonance HNCCCH experiments for correlating exchangeable and

nonexchangeable cytidine and uridine base protons in RNA. *J Biomol NMR*, **6**, 427-432.

Simorre, J.P., Zimmermann, G.R., Mueller, L. and Pardi, A. (1996) Triple-resonance experiments for assignment of adenine base resonances in C-13/N-15-labeled RNA. *J Am Chem Soc*, **118**, 5316-5317.

Sims, A.C., Ostermann, J. and Denison, M.R. (2000) Mouse hepatitis virus replicase proteins associate with two distinct populations of intracellular membranes. *J Virol*, **74**, 5647-5654.

Singleton, M.R. and Wigley, D.B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol*, **184**, 1819-1826.

Sklenar, V., Dieckmann, T., Butcher, S.E. and Feigon, J. (1996) Through-bond correlation of imino and aromatic resonances in C-13-,N-15-labeled RNA via heteronuclear TOCSY. *J Biomol NMR*, **7**, 83-87.

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.

Snoussi, K. and Leroy, J.L. (2001) Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry*, **40**, 8898-8904.

Sola, I., Moreno, J.L., Zuniga, S., Alonso, S. and Enjuanes, L. (2005) Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol*, **79**, 2506-2516.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B.A. and Siddell, S.G. (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *Embo J*, **2**, 1839-1844.

Spagnolo, J.F. and Hogue, B.G. (2000) Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. *J Virol*, **74**, 5053-5065.

Sperry, S.M., Kazi, L., Graham, R.L., Baric, R.S., Weiss, S.R. and Denison, M.R. (2005) Single-amino-acid substitutions in open reading frame (ORF) 1b-nsp14 and ORF 2a proteins of the coronavirus mouse hepatitis virus are attenuating in mice. *J Virol*, **79**, 3391-3400.

Stallings, S.C. and Moore, P.B. (1997) The structure of an essential splicing element: stem loop IIa from yeast U2 snRNA. *Structure*, **5**, 1173-1185.

Stohlman, S.A., Baric, R.S., Nelson, G.N., Soe, L.H., Welter, L.M. and Deans, R.J. (1988) Specific interaction between coronavirus leader RNA and nucleocapsid protein. *J Virol*, **62**, 4288-4295.

Stone, M.D., Mihalusova, M., O'Connor, C.M., Prathapam, R., Collins, K. and Zhuang, X. (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature*, **446**, 458-461.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M. and Z., R. (2006) Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *J Virol*, **80**, 7902-7908.

Surjit, M., Liu, B., Kumar, P., Chow, V.T. and Lal, S.K. (2004) The nucleocapsid protein of the SARS coronavirus is capable of sel-association through a C-terminal 209 amino acid interaction domain. *Biochem. Biophys. Res. Commun.*, **317**, 1030-1036.

Sutton, G., Fry, E., Cater, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L.L., Diprose, J., Alderton, D., Walsh, M., Grimes, J.M. and Stuart, D.I. (2004) The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure*, **12**, 341-353.

Tahara, S.M., Dietlin, T.A., Bergmann, C.C., Nelson, G.W., Kyuwa, S., Anthony, R.P. and Stohlman, S.A. (1994) Coronavirus translational regulation: leader affects mRNA efficiency. *Virology*, **202**, 621-630.

Tahara, S.M., Dietlin, T.A., Nelson, G.W., Stohlman, S.A. and Manno, D.J. (1998) Mouse hepatitis virus nucleocapsid protein as a translational effector of viral mRNAs. *Adv Exp Med Biol*, **440**, 313-318.

Tan, Y.W., Fang, S., Fan, H., Lescar, J. and Liu, D.X. (2006) Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells. *Nucleic Acids Res*, **34**, 4816-4825.

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.

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.

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.

Theimer, C.A., Finger, L.D. and Feigon, J. (2003) YNMG tetraloop formation by a dyskeratosis congenita mutation in human telomerase RNA. *RNA*, **9**, 1446-1455.

Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., Gorbalenya, A.E. and Ziebuhr, J. (2003a) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol*, **84**, 2305-2315.

Thiel, V., Karl, N., Schelle, B., Disterer, P., Klagge, I. and Siddell, S.G. (2003b) Multigene RNA vector based on coronavirus transcription. *J Virol*, **77**, 9790–9798.

Thiel, V. and Siddell, S. (2005) Reverse genetics of coronavirusesusing vaccinia virus vectors. *Curr Opin Microbiol Immunol*, **287**, 199-227.

Tijms, M., van Dinten, L., Gorbalenya, A. and Snijder, E. (2001) A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus. *Proc Natl Acad Sci U S A*, **98**, 1889-1894.

Vagner, S., Galy, B. and Pyronnet, S. (2001) Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep*, **2**, 893-898.

Van Den Born, E., Gultyaev, A.P. and Snijder, E.J. (2004) Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. *RNA*, **10**, 424-437.

van den Born, E., Posthuma, C.C., Gultyaev, A.P. and Snijder, E.J. (2005) Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J Virol*, **79**, 6312-6324.

van der Hoek, L., Pyrc, K. and Berkhout, B. (2006) Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol Rev*, **30**, 760-773.

Van der Meer, Y., Snijder, E.J., Dobbe, J.C., Schleich, S., Denison, M.R., Spaan, W.J.M. and Locker, J.K. (1999) Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication. *J Virol*, **73**, 7641-7657.

van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J. and Snijder, E.J. (1999) Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc Natl Acad Sci U S A*, **96**, 12056-12061.

VanZile, M.L., Cosper, N.J., Scott, R.A. and Giedroc, D.P. (2000) The zinc metalloregulatory protein Synechococcus PCC7942 SmtB binds a single zinc ion per monomer with high affinity in a tetrahedral coordination geometry. *Biochemistry*, **39**, 11818-11829.

Verma, S., Bednar, V., Blount, A. and Hogue, B.G. (2006) Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. *J Virol*, **80**, 4344-4355.

von Grotthuss, M., Wyrwicz, L.S. and Rychlewski, L. (2003) mRNA cap-1 methyltransferase in the SARS genome. *Cell*, **113**, 701-702.

Wakeman, C., Winkler, W.C. and Dann III, C. (2007) Structural features of metabolite-sensing riboswitches. *Trends in Biochemical Sciences*, **32**, 415-424.

Wang, Y. and Zhang, X. (1999) The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 in vitro and in vivo. *Virology*, **265**, 96-109.

Wang, Y. and Zhang, X. (2000) The leader RNA of coronavirus mouse hepatitis virus contains an enhancer-like element for subgenomic mRNA transcription. *J Virol*, **74**, 10571-10580.

White, K.A., Nagy, P.D. (2004) Advances in the molecular biology of tombusviruses: gene expression, genome replication, and recombination. *Prog Nucleic Acid Res Mol Bio*, **78**, 187-226

Wilkinson, K., Merino, E. and Weeks, K. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols*, **1**, 1610-1616.

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.

Williamson, J.R. (2000) Induced fit in RNA-protein recognition. *Nat Struct Biol*, **7**, 834-837.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M.J., 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.

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S. and Yuen, K.Y. (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*, **79**, 884-895.

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

Wu, B., Vanti, W.B., White, K.A. (2001) An RNA domain within the 5' untranslated region of the tomato bushy stunt virus genome modulates viral RNA replication. *J Mol Biol*, **305**, 741-756.

Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G.F., Anand, K., Bartlam, M., Hilgenfeld, R. and Rao, Z. (2003) The

crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* **100**, 13190-13195.

Yu, I.M., Gustafson, C.L., Diao, J., Burgner, J.W., 2nd, Li, Z., Zhang, J. and Chen, J. (2005) Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J Biol Chem*, **280**, 23280-23286.

Yu, I.M., Oldham, M.L., Zhang, J. and Chen, J. (2006) Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. *J Biol Chem*, **281**, 17134-17139.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. and Rao, Z. (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. *Nature structural & Molecular Biology*, **12**, 980-986.

Zhou, M., Williams, A.K., Chung, S.I., Wang, L. and Collisson, E.W. (1996) The infectious bronchitis virus nucleocapsid protein binds RNA sequences in the 3′ terminus of the genome. *Virology*, **217**, 191-199.

Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, **296**, 1473-1476.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nuc Acid Res*, **31**, 3406-3415.

Zuniga, S., Sola, I., Alonso, S. and Enjuanes, L. (2004) Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol*, **78**, 980-994.

Zuniga, S., Sola, I., Moreno, J.L., Sabella, P., Plana-Duran, J. and Enjuanes, L. (2007) Coronavirus nucleocapsid protein is an RNA chaperone. *Virology*, **357**, 215-227.

NOTES

[1]Although A35 or A36 (or both, one at a time) could theoretically be extruded from the SL1 helix, extrusion of either one gives rise to the same structure of paired bases in the SL1 stem. In this work, we characterized ΔA35 MHV mutants, but ΔA36 viruses would give rise to exactly the same sequence in SL1 and thus would not be functionally distinguishable from the ΔA35 virus.

APPENDIX I

In the fluorescence anisotropy titration experiments, the binding isotherms for the N protein variants were fit by using DynaFit to a 1:1 binding model. The script file is listed bellow:

```
[task]
    data = equilibria
    task = fit
[mechanism]
    P + R* <==> P.R* : K1 assoc.
[concentrations]
    R* = 0.01
[constants]
    K1 = 100 ?
[responses]
    R* = 7.5 ?
    P.R* = 20 ?
[equilibria]
    variable P
file ./data/N197_trsL3.txt
[output]
    directory ./output/N197_trsL3
```

[end]

Here, P represents the N protein variant and R is the *F*-5'-TRS-L3. In this fit, the concentration of R is fixed at 10 nM. A nonlinear least-squares fit is carried out on the anisotropy data to obtain K_{obs} (K1), r_0 and $r_{complex}$.

APPENDIX II

The script file in DynaFit for fluorescence anisotropy-based competition experiment is listed below:

```
[task]
    data = equilibria
    task = fit
[mechanism]
    P + R* <==> P.R* : K1 assoc.
    P + S <==> P.S : K2 assoc.
[concentrations]
    R* = 0.01
    P=0.13066
[constants]
    K1 = 93.9  , K2 = 100 ?
[responses]
    R* = 5.4?
    P.R* = 19.3?
[equilibria]
    variable S
    file ./data/N219_trsL3_trsa.txt
[output]
```

Here, P is the N variant, R is the *F*-5'-TRS-L3, and S is the unlabeled TRS-L3 WT or mutant RNA construct. In this fit, the concentration of R is fixed at 10 nM, the concentration of P is fixed at saturation, and K_1 is fixed at the number determined in normal fluorescence anisotropy titration experiment (listed in APPENDIX I). Nonlinear least-squares fits are carried out to fit the anisotropy data to obtain K_2 , which represents K_{obs} .

APPENDIX III

The reverse intrinsic fluorescence titration data were fit by SCIENTIST according to the McGhee and von Hippel model, in which the poly(A) /poly(dT) (termed M here) is considered as a linear lattice of N sites. Each bound protein (ligand, termed X here) occupies n sites. The script file used in SCIENTIST is given as following:

```
// MicroMath Scientist Model File
// Von'Hippel model (no cooperativity) for reverse fluorecsence titrations
//signal from the ligand (X) titrated with the macromolecule (M)
IndVars:Mtot
DepVars:Qobs,Rtot,v,Xf,Scatch
Params:Xtot,Qmax,k,n
Xf=Xtot-v*Mtot
v=Xf*(K*(1-n*v)*((1-n*v)/(1-n*v+v))^(n-1))
0<v<1/n
Qobs=v*Qmax*(Mtot/Xtot)
Rtot=Mtot/Xtot
Scatch=v/Xf
***
```

X_{tot} is fixed, M_{tot} is the titrated macromolecule (poly(A) or poly(dT) in our case). Q_{obs} is normalized fluorescence quenching data. The fits of Q_{obs} data will generate parameters of Q_{max}, K_{obs}, and n.

VITA

Name Lichun Li

Education B.S., Biochemistry (1999), Xiamen University
M.S., Biochemistry and Molecular Biology (2002)
Graduate School of Chinese Academy of Sciences
Ph.D. Biochemistry (2007), Texas A&M University

Publications

Li, L*., Kang, H*., Makkinje, N., Liu, P., Williamson, S.T., Leibowitz, J.L. and Giedroc, D.P. Structural lability in stem-loop 1 (SL1) drives a 5' UTR-3'UTR interaction required for replication of Mouse Hepatitis Virus (MHV) (*submitted*) (*Authors contributed equally)

Liu, P*., Li, L.*, Millership, J.J., Kang, H., Leibowitz, J.L. and Giedroc, D.P. (2007) A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA*, **13**, 763-780 (*Authors contributed equally).

Liu, P., Millership, J.J., Li, L., Giedroc, D.P. and Leibowitz, J.L. (2006) A previously unrecognized UNR stem-loop structure in the coronavirus 5' untranslated region plays a functional role in replication. *Adv Exp Med Biol*, **581**, 25-30.

Tian, W.X., Li, L.C., Wu, X.D. and Chen, C.C. (2004) Weight reduction by Chinese medicinal herbs may be related to inhibition of fatty acid synthase. *Life Sci*, **74**, 2389.

Li, L.C., Wu, X.D. and Tian, W.X. (2003) Inhibition to fatty acid synthase with extract of tuber fleeceflower root. *Chinese Journal of Biochemistry and Molecular Biology*, **19**, 297-304.

Li, L.C. and Tian, W.X. (2001) New evolution: inhibitors of fatty acid synthase and fat-reducing study. *Chinese Science Bulletin,* **46**, 1673.

Address Department of Biochemistry and Biophysics
College Station, TX 77843-2128