

# A PREVIOUSLY UNRECOGNIZED UNR STEM-LOOP STRUCTURE IN THE CORONAVIRUS 5' UNTRANSLATED REGION PLAYS A FUNCTIONAL ROLE IN REPLICATION

Pinghua Liu, Jason J. Millership, Lichun Li, David P. Giedroc,  
and Julian L. Leibowitz\*

## 1. INTRODUCTION

*Cis*-acting sequences in the MHV and bovine coronavirus (BCoV) 5' UTR are required for defective interfering RNA replication, subgenomic RNA synthesis, and presumably also for virus genome replication.<sup>1-3</sup> For the BCoV 5' UTR, the minimum free energy structure is predicted to contain three stem-loops.<sup>3</sup> We have analyzed the entire 5' UTR sequences of nine group 1 and group 2 coronaviruses, including the newly discovered SARS and HKU1 coronaviruses, using consensus, covariation secondary structure predictions. Our analysis shows that the predicted secondary structures of all coronavirus 5' UTRs are strikingly similar and contain three or four stem-loops, including a previously unrecognized highly conserved UNR stem-loop.

Computer-assisted modeling predicts an invariant and previously unrecognized UNR stem-loop among nine coronavirus UTR sequences, denoted SL2. This predicted conserved structure encompasses nucleotides 42-56 for both MHV and the SARS-coronavirus. NMR spectroscopy of the 16-nt RNA SL2<sup>sars</sup> reveals spectral features consistent with a UNR hairpin loop. Reverse genetics studies revealed that SL2 is required for MHV replication; MHV genomes containing a substitution of the required U48 with C in the UNR loop (U48C) were not viable. RT-PCR analysis of the U48C mutant indicated that negative sense genome sized RNAs were present in cells electroporated with this mutant; however, neither positive nor negative sense subgenomic RNAs were detected. Mutations that destabilized the stem of SL2 were viable but had moderately to severely impaired replication phenotypes. Mutants that maintained the stem-loop structure replicated similarly to wild-type MHV. These genetic data strongly support the existence of the predicted UNR stem loop and its functional importance in viral replication.

---

\* Pinghua Liu, Jason J. Millership, Julian L. Leibowitz, Texas A&M University System-HCS, College Station, Texas. Lichun Li, David P. Giedroc, Texas A&M University, College Station, Texas.

## 2. MATERIALS AND METHODS

The RNA secondary structure prediction algorithm Vienna RNA 1.5<sup>4</sup> was used to predict the secondary structures of group 1 and group 2 coronavirus 5' UTRs. A reverse genetic system based on *in vitro* assembly of cloned cDNAs (A-G) was used to recover wild-type MHV-A59 1000 and mutant viruses.<sup>5</sup> To construct mutants in SL2, a series of overlapping oligonucleotides spanning the sequence between MluI and SacII sites (1 to 106 nts) in plasmid A were synthesized and ligated together to form a 135-bp DNA fragment with MluI and Sac II sites at its 5' and 3' ends, respectively. The assembled fragment containing either the wild-type sequence, or with mutations in SL2, was ligated into plasmid A. cDNA corresponding to the entire MHV-A59 genome were assembled by *in vitro* ligation, transcribed *in vitro*, and the transcribed mutant genomic RNA or wild-type MHV-A59-1000 RNA was electroporated into BHK-R cells to recover infectious virus as described.<sup>5</sup> Viruses were plaque purified and expanded once in DBT cells. Mutants were tested three times, including three blind passages of lysates from electroporated cultures without recovering infectious virus, before being declared non-viable. Total cellular RNAs were extracted 4 and 8 hours post-electroporation and assayed by nested RT-PCR to detect negative sense genome sized RNA and positive and negative sense subgenomic RNAs, as described.<sup>6</sup>

## 3. RESULTS

A series of three stem-loops denoted I, II, and III had been predicted in the BCoV 5' UTR. These predictions were supported in part by nuclease mapping experiments and by DI replication assays.<sup>3,7</sup> However, it was puzzling to us that the predicted stem-loops I and II of BCoV are not conserved amongst group 2 coronaviruses. Here we used Vienna RNA 1.5<sup>4</sup> to predict the 5' UTR secondary structures of nine coronaviruses (summarized in Table 1). The results of Vienna RNA folding predictions carried out for MHV and SARS-CoV are shown in Figure 1. All coronavirus 5' UTR secondary structural models are strikingly similar, and are characterized by three major helical stems, denoted SL1, SL2, and SL4. Some sequences show a fourth stem-loop, denoted SL3, in which the leader TRS (TRS-L) is folded into a hairpin loop. SL2 is absolutely conserved and previously unrecognized. The (C/U)UUG(U/C) pentaloop sequence is the most highly conserved contiguous run of nucleotides in the entire 5' UTR outside of the core TRS-L, and covariation analysis clearly reveals that this loop is always stacked on a 5-bp helix. Analysis of the entire  $\approx 30$  kb MHV and SARS-CoV genomes reveals that SL2-like stem loops are extremely rare (appearing just 3 and 5 other times, respectively); this suggests an important role in coronavirus replication.

The (C/U)UUG(U/C) sequence of SL2 has all the features of a classical UNR ( $U_0 \bullet N_{+1} \bullet R_{+2}$ ) hairpin loop, in which a UNR triloop stacks on a Y:Y, Y:A (Y-pyrimidine) or G:A noncanonical base pair.<sup>8</sup> The basic structural feature of the U-turn architecture is a sharp turn in the phosphate backbone between  $U_0$  and  $N_{+1}$  (*i.e.*, a uridine (U)-turn, first

**Table 1.** Predicted 5' UTR secondary structure for group 1 and 2 coronaviruses.

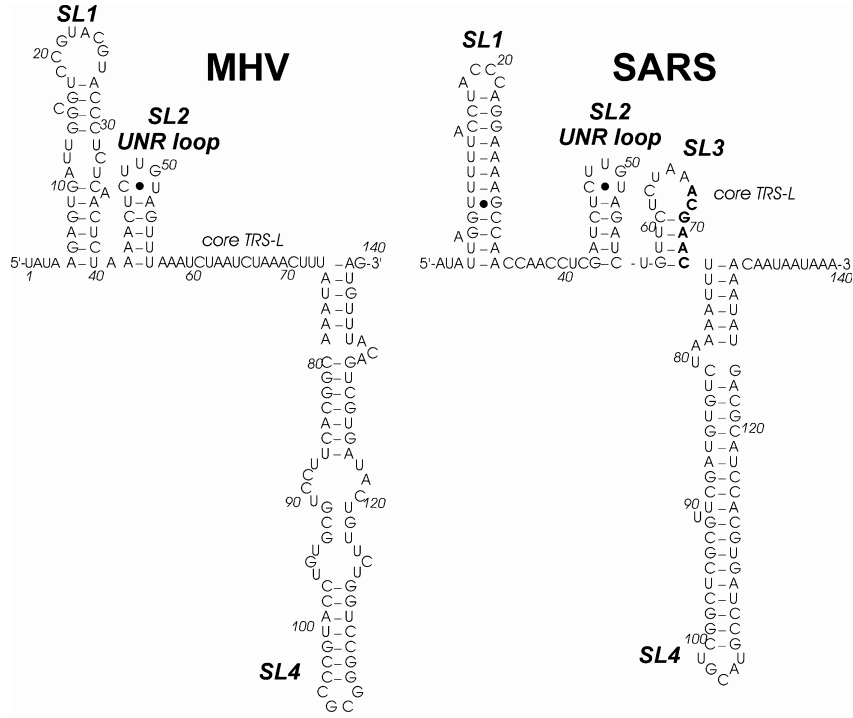
Viruses	Predicted secondary structure			
	SL1	SL2	SL3	SL4
HCoV-OC43	Predicted	Strongly predicted	Predicted	Predicted
BCoV	Predicted	Strongly predicted	Unfolded at 37°C	Predicted
MHV	Predicted	Strongly predicted	Not predicted	Predicted
HKU1	Predicted	Strongly predicted	Unfolded at 37°C	Predicted
SARS-CoV	Predicted	Strongly predicted	Predicted	Predicted
HCoV-NL63	Predicted	Strongly predicted	Not predicted	Predicted
HCoV-229E	Predicted	Strongly predicted	Not predicted	Predicted
TGEV	Predicted	Strongly predicted	Not predicted	Predicted

identified in the anticodon and T-loops of tRNA). In a UNR loop,  $U_0$  is stacked on the noncanonical base pair and is engaged in two critical hydrogen bonds: the  $U_0$  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  $R_{+2}$ . Substitution of  $U_0$  with any other nucleotide will abrogate formation of these hydrogen bonds and therefore destabilize the loop.

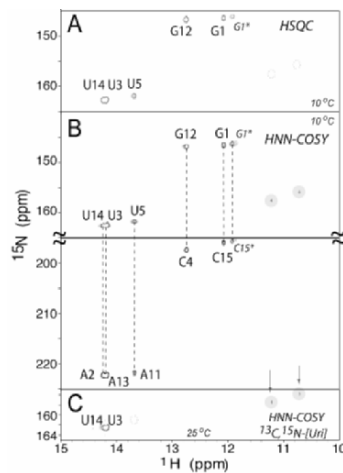
A 16-nt RNA termed SL2<sup>sars</sup> *in vitro* transcribed by SP6 RNA polymerase was subjected to NMR spectroscopy.  $^1\text{H}$ - $^{15}\text{N}$  HSQC and  $^2\text{J}$  HNN-COSY spectra acquired with uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled SL2<sup>sars</sup> (Fig. 2A, B) reveal the expected correlations for all five base pairs in the stem as well as two upfield-shifted  $^1\text{H}$ - $^{15}\text{N}$  correlations (Fig. 2B); their detection in the uridine-only  $^{13}\text{C}$ ,  $^{15}\text{N}$ -[U]-labeled sample reveal that they must correspond to U3-H3 cross-peaks (Fig. 2C) indicative of noncanonical hydrogen bonding. These uridines likely correspond to U48 and U51 (Fig. 1B).

Based on the predicted UNR stem-loop, a series of mutations were introduced into the SL2 region of MHV. Introduction of a conservative U48C substitution into the MHV genome resulted in a nonviable genome, a result consistent with our prediction of an invariant U at position 48. At 4 and 8 h post-electroporation of U48C or wild-type genomes, total RNA was extracted and analyzed by nested RT-PCR. As shown in Figure 3, negative sense genome sized RNAs were present in cells electroporated with U48C; however, neither positive nor negative sense subgenomic RNA6 and RNA7 were detected. A similar result was also obtained for RNA3 (not shown).

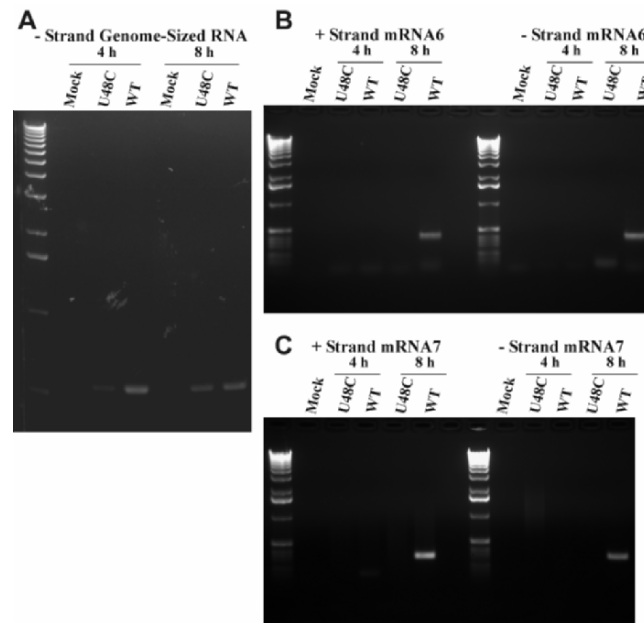
Consistent with the predicted structure of the UNR loop, the U49A mutant was viable and produced a virus with a near normal plaque size (Figs 4A-B). Three other mutant viruses, C45G, G53C, and C45G/G53C, with substitutions in the SL2 helical stem were recovered, although their plaque sizes are different. Mutants C45G and G53C, predicted to destabilize SL2, both form smaller plaques than wild-type virus. The double mutant C45G/G53C restores the SL2 stem and forms plaques slightly larger than wild-type MHV-A59. One step growth curves confirmed the growth phenotypes of these viruses (Figs. 4C-D). Denaturing gel electrophoresis of metabolically labeled RNA prepared from cells infected with wild-type MHV or various SL2 mutants demonstrated that a similar ratios of genome and subgenomic RNAs were synthesized, but that the amount of RNA synthesized correlated with virus growth phenotype. Thus cells infected with the G53C mutant synthesized much less RNA than wild-type MHV-A59 1000, cells infected with the C45G/G53C and U49A mutants synthesized near wild-type levels of



**Figure 1.** Predicted 5' UTR secondary structures of MHV and SARS-CoV.



**Figure 2.** (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC and (B)  $^2J$  HNN-COSY spectra acquired with uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled  $\text{SL2}^{\text{SARS}}$ . (C)  $^2J$  HNN-COSY spectrum acquired for a uridine-only  $^{13}\text{C}$ ,  $^{15}\text{N}$ -[U]-labeled sample. Conditions:  $^1\text{H}$  frequency 600 MHz, pH 6.0, 0.1 M KCl, 5 mM  $\text{MgCl}_2$ , 10–25°C.



**Figure 3.** Nested RT-PCR analysis of cells electroporated with wild-type or U48C mutant genomes.

RNA, and cells infected with the C45G mutant synthesized somewhat less RNA than wild-type infected cells (data not shown). Taken together, these results are consistent with a role for SL2 in MHV RNA replication and transcription.

#### 4. DISCUSSION

The genetic and NMR studies presented above support the proposed secondary structure model for the coronavirus 5' UTR, and the functional importance of SL2 in subgenomic RNA synthesis. All coronaviruses contain similar SL1, SL2, and SL4 stem-loops. This covariation-based model differs in several respects from an earlier minimum free energy structure predicted for the BCoV 5' UTR.<sup>3</sup> The highly conserved SL2 structure we have studied here is not present in the prior BCoV model. Our genetic and biophysical studies support the prediction of SL2 being a UNR U-turn stem-loop. One feature shared by both models is a stem-loop designated as stem-loop III in the earlier BCoV model, which corresponds to stem-loop 4B in our model. The previous BCoV minimum free energy model is also supported by nuclease mapping studies, particularly for stem-loop III, and by genetic studies using DI RNAs. This raises the remote possibility, not addressed here, that the two competing confirmations may exist in equilibrium and perhaps have different functional roles in coronavirus replication.

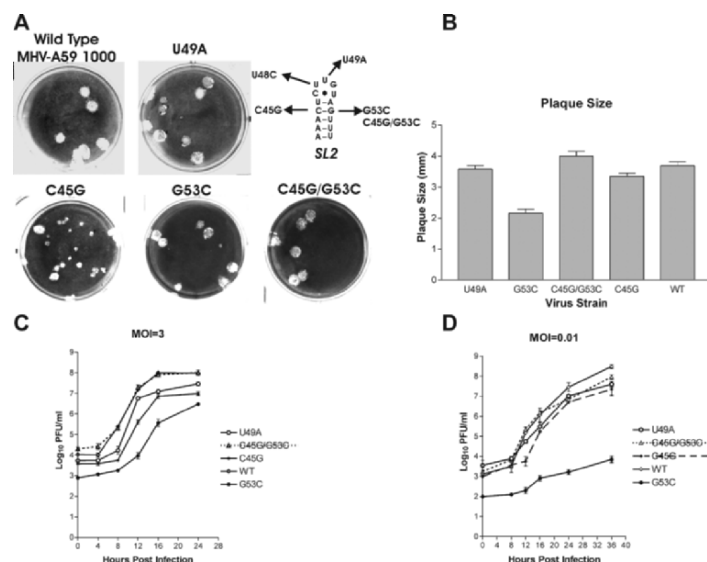


Figure 4. Characterization of growth phenotypes for viable SL2 mutant viruses.

## 5. ACKNOWLEDGMENTS

This work was supported by NIH grants AI051493 (J.L.L.) and AI040187 (D.P.G.).

## 6. REFERENCES

1. R.-Y. Chang, M. A. Hofmann, P. B. Sethna, and D. A. Brian, A cis-acting function for the coronavirus leader in defective interfering RNA replication, *J. Virol.* **68**, 8223-8231 (1994).
2. Y. Kim, Y. S. Jeong, and S. Makino, Analysis of cis-acting sequences essential for coronavirus defective interfering RNA replication, *Virology* **197**, 53-63 (1993).
3. S. Raman, P. Bouma, G. D. Williams, and D. A. Brian, 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 (2003).
4. I. L. Hofacker, Vienna RNA secondary structure server, *Nucleic Acids Res.* **31**, 3429-3431 (2003).
5. B. Yount, M. R. Denison, S. R. Weiss, and R. S. Baric, Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59, *J. Virol.* **76**, 11065-11078 (2002).
6. R. F. Johnson, M. Feng, P. Liu, J. J. Millership, B. Yount, R. S. Baric, and J. L. Leibowitz, The effect of mutations in the mouse hepatitis virus 3'(+)-42 protein binding element on RNA replication, *J. Virol.* (in press).
7. R.-Y. Chang, R. Krishnan, and D. A. Brian, The UCUAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA, *J. Virol.* **70**, 2720-2729 (1996).
8. R. R. Gutell, J. J. Cannone, D. Konings, and D. Gautheret, Predicting U-turns in ribosomal RNA with comparative sequence analysis, *J. Mol. Biol.* **300**, 791-803 (2000).