Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Virology 475 (2015) 15–27

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

SHAPE analysis of the RNA secondary structure of the Mouse Hepatitis Virus 5' untranslated region and N-terminal nsp1 coding sequences

Dong Yang [a,1], Pinghua Liu [a,2], Elyse V. Wudeck [a], David P. Giedroc [b], Julian L. Leibowitz [a,*]

[a] Department of Microbial Pathogenesis and Immunology, Texas A&M University, College of Medicine, 407 Reynolds Medical Building, 1114 TAMU, College Station, TX 77843-1114, USA
[b] Department of Chemistry, Indiana University, Bloomington, IN 47405-7102, USA

ARTICLE INFO

Article history:
Received 15 October 2013
Returned to author for revisions
21 November 2013
Accepted 3 November 2014
Available online 21 November 2014

Keywords:
Mouse Hepatitis Virus
Murine coronavirus
Betacoronavirus
RNA secondary structure
SHAPE analysis
5′ cis-acting region

ABSTRACT

SHAPE technology was used to analyze RNA secondary structure of the 5′ most 474 nts of the MHV-A59 genome encompassing the minimal 5′ *cis*-acting region required for defective interfering RNA replication. The structures generated were in agreement with previous characterizations of SL1 through SL4 and two recently predicted secondary structure elements, S5 and SL5A. SHAPE provided biochemical support for four additional stem–loops not previously functionally investigated in MHV. Secondary structure predictions for 5′ regions of MHV-A59, BCoV and SARS-CoV were similar despite high sequence divergence. The pattern of SHAPE reactivity of *in virio* genomic RNA, *ex virio* genomic RNA, and *in vitro* synthesized RNA was similar, suggesting that binding of N protein or other proteins to virion RNA fails to protect the RNA from reaction with lipid permeable SHAPE reagent. Reverse genetic experiments suggested that SL5C and SL6 within the nsp1 coding sequence are not required for viral replication.

Introduction

Coronaviruses (CoVs) comprise a group of enveloped RNA viruses, with large positive-sense, single-stranded RNA genomes of 25–31 kb, which cause respiratory, enteric, hepatic and neurological diseases in a broad range of vertebrate species, including humans (Stadler et al., 2003; Weiss and Leibowitz, 2007). Mouse Hepatitis Virus (MHV) is the most extensively studied prototypical CoV, and provides a platform to study CoV replication and transcription (Snijder et al., 2003; Weiss and Leibowitz, 2007). Cells infected with MHV contain genomic RNA and six to seven subgenomic mRNAs that make up a 3′ co-terminal nested set (Sawicki et al., 2007). At their 5′ ends subgenomic mRNAs contain a 72 nucleotide leader sequence identical to the 5′ 72 nucleotides of virion RNA (Sawicki et al., 2007). The entire leader sequence is not present in the genome sequence 5′-adjacent to the remainder (body) of the mRNA rather a short 7 nucleotide sequence known as a transcriptional regulatory sequence (TRS), also present at the 3′end of the leader sequence at the 5′end of the genome, is found at these positions (Sawicki et al., 2007). The TRS plays a crucial role in the discontinuous transcription mechanism by which the subgenomic mRNAs are synthesized, serving as signals (and sites) for leader body joining during transcription of the negative strand templates for positive sense subgenomic mRNAs (Sawicki et al., 2007). Several studies employing defective interfering (DI) RNAs as model replicons showed that approximately 400–800 nucleotides at the 5′ end and 400 nts at the 3′ end of MHV RNA genome act as *cis*-acting sequences necessary for replication (Kim et al., 1993; Lin and Lai, 1993; Luytjes et al., 1996), with the minimal length of 5′ sequence that supported DI replication being 474 nts (Kim et al., 1993). These *cis*-acting sequences in MHV RNA have been presumed to fold into secondary and higher-order structures functioning as signals important for RNA–RNA interactions and for binding of viral and cellular proteins during RNA replication, translation and potentially encapsidation (Brian and Baric, 2005; Liu and Leibowitz, 2010; Liu et al., 2009b).

MHV is closely related to Bovine Coronavirus (BCoV) and SARS-coronavirus (SARS-CoV) in the betacoronavirus genus (Gorbalenya et al., 2006). The secondary structures in the 5′-end-proximal genomic regions of these three viruses are largely conserved even though the nucleotide sequences are relatively divergent (Chen and Olsthoorn, 2010; Guan et al., 2012). A series of studies by consensus covariation modeling, chemical probing, and nuclear magnetic resonance (NMR) spectroscopy in conjunction with reverse genetics have been carried out to characterize the predicted secondary structures of *cis*-acting sequences in the 5′UTR

* Corresponding author. Tel.: +1 979436 0313; fax: +1 979 845 3479.
E-mail address: jleibowitz@tamu.edu (J.L. Leibowitz).
[1] Current address: Department of Microbiology, Immunology & Biochemistry, The University of Tennessee Health Science Center, College of Medicine, Memphis, TN 38163, USA.
[2] Current address: Center for Inflammation and Epigenetics, The Methodist Hospital Research Institute, 6670 Bertner Avenue, R9-460, Houston 77030, USA.

http://dx.doi.org/10.1016/j.virol.2014.11.001
0042-6822/

and the N-terminal nsp1 coding region of MHV and BoCV, and to identify their functional roles in viral replication (Brown et al., 2007; Chen and Olsthoorn, 2010; Guan et al., 2012, 2011; Li et al., 2008; Liu et al., 2009a, 2007; Yang et al., 2011). The structurally bipartite stem–loop SL1 (nts 5–40) (Li et al., 2008) and SL2 a uCUYG(U)a-like tetraloop stacked on a five base pair stem (nts 42–56) (Liu et al., 2009a, 2007) in the MHV 5′UTR (see Fig. 1A) have been functionally characterized and are required for subgenomic RNA synthesis; the structure of SL2 has been determined at atomic resolution (Lee et al., 2011). In MHV, a base pairing scheme can be drawn which places the TRS (transcriptional regulatory sequence) into the loop of SL3 (Chen and Olsthoorn, 2010), but the stem is not predicted to be stable at 37 °C (Liu et al., 2007). SL4 (nts 80–130) (see Fig. 1A) seems to function as a poorly understood "spacer" element that may optimally orient adjacent stem–loops; the presence of an SL4 is required to drive subgenomic RNA synthesis with little dependence on nucleotide sequence (Yang et al., 2011). Recently the Brian group determined that a stem loop that they designated SLIV (SL5A in our model; see Fig. 1A), spanning nts 171–225 and thus extending into the nsp1 coding sequence, was required for optimal replication of MHV (Guan et al., 2011). The individual stem–loop structures contained in the 5′UTRs of related betacoronaviruses are largely interchangeable, with the exception of their TRSs (Guan et al., 2011; Kang et al., 2006; Li et al., 2008). However, the 32 nts in the 5′ side of S5 (nts 142–173) in the BCoV 5′UTR, which has been predicted to be single-stranded in previous thermodynamic and phylogenetic models (Guan et al., 2011), cannot directly replace the equivalent sequences in MHV without genetic adaptation (Guan et al., 2012, 2011). These adaptations suggested that a long range interaction between nts 141 and 170 with nts 332 and 363 (S5, Fig. 1A) is required for optimal viral replication (Guan et al., 2012).

A chemical method termed SHAPE (selective 2′-hydroxyl acylation and primer extension) used to determine the overall secondary structure for any length of RNA sequence under different biological conditions has been developed (Merino et al., 2005; Watts et al., 2009; Wilkinson et al., 2008). In SHAPE chemistry an electrophilic reagent, 1M7 (1-methyl-7-nitroisatoic anhydride) or NMIA (N-methylisatoic anhydride), selectively acylates the ribose 2′-hydroxyl group of a flexible nucleotide in a RNA and forms a stable 2′-O-adduct (nucleotide 2′-ester) (Merino et al., 2005; Mortimer and Weeks, 2007). SHAPE has important advantages over traditional biochemical measurements of RNA secondary structure. SHAPE provides quantitative data for addressing structure–function relationships in RNA. SHAPE-derived RNA secondary structure determinations yield models of RNAs of known structure that include 85–95% of the accepted base pairs (Deigan et al., 2009; Kladwang et al., 2011). An additional advantage of SHAPE is that reactivity is not influenced by the base identity of the nucleotide sequence (Mortimer and Weeks, 2007; Wilkinson et al., 2009).

The hypothesis underlying this work is that the *cis*-acting 5′-most 474 nts of MHV-A59, including the 5′UTR (nts 1–209) and the immediately adjacent 5′ region of the nsp1 coding sequence (nts 210–474), form stem–loop structures important for viral replication and viral RNA synthesis. We use SHAPE methodology to determine the secondary structure of the 5′-most 474 nts of MHV-A59 genome, which corresponds to the minimal 5′ *cis*-acting regulatory region mapped in defective interference (DI) experiments (Kim et al., 1993; Luytjes et al., 1996). This region contains the 5′UTR and extends into nsp1, a 28-kDa amino-terminal protein (also known as p28) which is the first mature protein processed from the gene 1 polyprotein (Denison and Perlman, 1987). The N-terminus of the BCoV nsp1 coding region is predicted to contain stem–loop structures supported by RNase structure probing and by nucleotide covariation among closely related betacoronaviruses, and some of these secondary structures act as *cis*-acting elements required for BCoV DI RNA replication (Brown et al., 2007; Gustin et al., 2009). In this work we compare the RNA secondary structure of the 5′ 474 nts of MHV-A59 genome in the virion (*in virio*) to the structure of genomic RNA that has been gently extracted and deproteinized (*ex virio*) and to an *in vitro* RNA transcript targeting the 5′-most 474 nts, in an effort to identify differences in SHAPE reactivity among the three biological states of RNA from which we might infer putative protein binding sequences/structures within this region of the genome. To further assess the validity of the structural model generated by SHAPE and

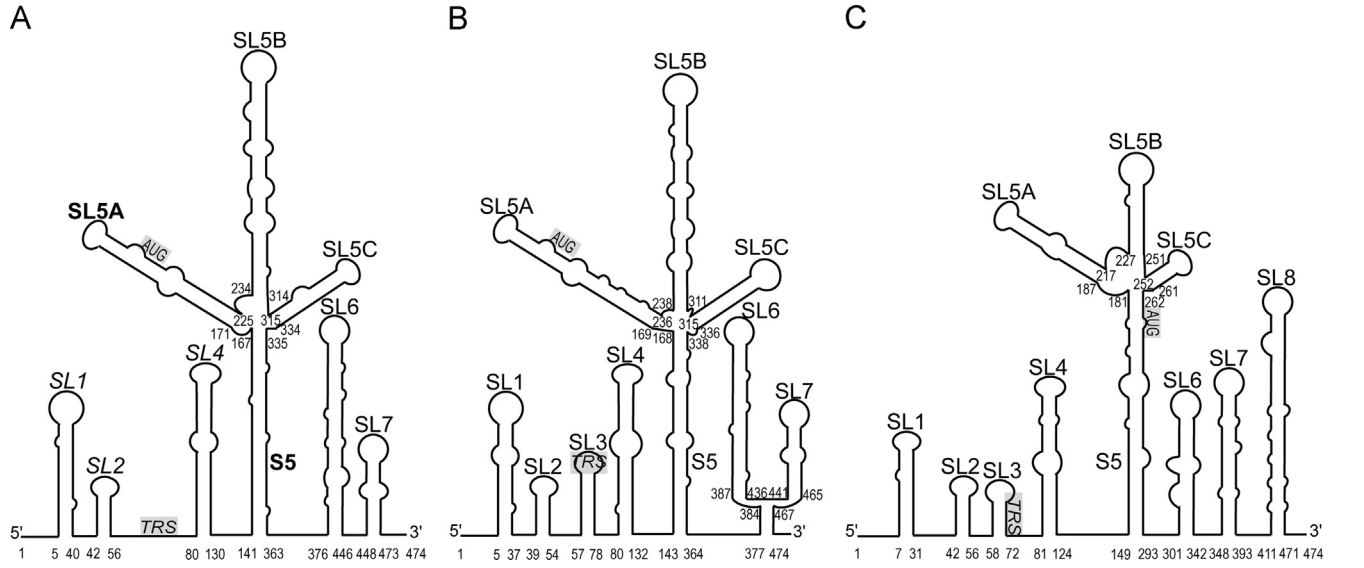

Fig. 1. Comparison of secondary structure models of the 5′- most 474 nts of MHV-A59, BCoV, and SARS-CoV. (A) The MHV-A59 model was generated by SHAPE analysis. Italicized labels indicate structural elements essential for virus recovery and subgenomic (−) sense RNA synthesis. Bolded labels indicate structural elements that when mutated result in large decreases in the ability of viral genomes to replicate, and thus have a significant role in viral replication. (B) The thermodynamically most stable models of the corresponding 5′ 474 nts of BCoV-Mebus and (C) SARS-CoV were generated by RNAstructure. The gray and italicized text denotes the core leader TRS regions. The gray AUGs represent the start codons of nsp1 in the three viruses. Note that the nomenclature for BCoV is not equivalent to that of previous studies by the Brian Laboratory (Brown et al., 2007; Guan et al., 2012, 2011); SL1 and SL2 correspond to SLI, SL3 to SLII, SL4 to SLIII, SL5A to SLIV, SL5B to SLV, SL5C to SLVI, S5 to a long-range RNA–RNA interaction, and SL6 to SLVIII.

to identify common replication signals in CoVs, we compare the secondary structures obtained by SHAPE with our current model of the first 140 nts of the MHV 5′UTR which contain SL1 (Li et al., 2008), SL2 (Liu et al., 2009a, 2007) and SL4 (Yang et al., 2011) and with the secondary structures in the 3′ adjacent region predicted by the Brian laboratory (Brown et al., 2007; Guan et al., 2012, 2011; Raman et al., 2003; Raman and Brian, 2005) and by Chen and Olsthoorn (2010). We employ a reverse genetic approach to examine the functional roles in virus replication of two novel secondary structures predicted by SHAPE analysis.

Results and discussion

SHAPE analysis of MHV-A59 5′ region of RNA

In vitro synthesized and refolded RNA corresponding to nts 1–489 plus a 24-nucleotide unstructured region were separately reacted with the SHAPE reagents 1M7 and NMIA, and SHAPE reactivity at each position was assessed by primer extension as described in *Materials and methods*. *In vitro* synthesized RNA was used as the initial substrate for SHAPE analysis because all previous biochemical analyses of the RNA secondary structure of the 5′ region of coronaviruses have used *in vitro* synthesized RNAs (Brown et al., 2007; Chen and Olsthoorn, 2010; Guan et al., 2012, 2011; Li et al., 2008; Liu et al., 2009a; Raman et al., 2003; Raman and Brian, 2005; Yang et al., 2011), thus allowing a direct comparison of results. Although 1M7 provides a faster reaction rate and more accurate information regarding RNA structure than NMIA (Mortimer and Weeks, 2007) we utilized both 1M7 and NMIA in order to compare the results obtained with the two chemicals and to provide a basis for subsequent *in virio* genomic RNA SHAPE experiments which require NMIA being able to penetrate virus envelopes (Wilkinson et al., 2008). The SHAPE reactivity profiles obtained with 1M7 and NMIA were highly reproducible and very similar. The electropherograms of a typical region at positions 20–74 obtained by 1M7 and NMIA modification are shown in Fig. 2A and B, respectively. Strikingly, the core leader-TRS sequence at nts 60–72 and the flanking region showed high SHAPE reactivity, as expected, because these regions in MHV are predicted to be weakly paired or unpaired (Chen and Olsthoorn, 2010; Liu et al., 2007).

Three independent experiments were carried out for each reaction and the raw electropherograms containing SHAPE reactivity *versus* nucleotide position were analyzed using ShapeFinder as described in *Materials and methods* to yield quantitative SHAPE reactivities at single nucleotide resolution. The mean SHAPE reactivity data for 1M7 modified *in vitro* transcribed RNA corresponding to nts 1–489 plus a 24 nts unstructured region is shown for nts 18–457 in Fig. 3. Quantitative SHAPE results were not obtained for the 5′- most 17 nts and for nts 458–474 at the 3′ end of the target sequence. The data shown in Fig. 3A is the average SHAPE reactivity obtained using a primer complementary to position 261–282 near the mid-point of the target sequence. Fig. 3B displays data obtained using a primer complementary to the unstructured primer binding site included in the *in vitro* transcribed RNA 3′ to nt 489. The experimental reproducibility was excellent.

SHAPE analysis informs a secondary structure model of the 5′ 474 nts cis-acting region required for replication of MHV-A59 DI RNA

The mean SHAPE reactivity results obtained with 1M7 for the 5′ 474 nts segment of *in vitro* transcribed and refolded RNA representing the MHV-A59 genome (see Fig. 3) were imported into RNAstructure, incorporating the SHAPE reactivity as pseudo-free energy constraints as described in *Materials and methods*. The

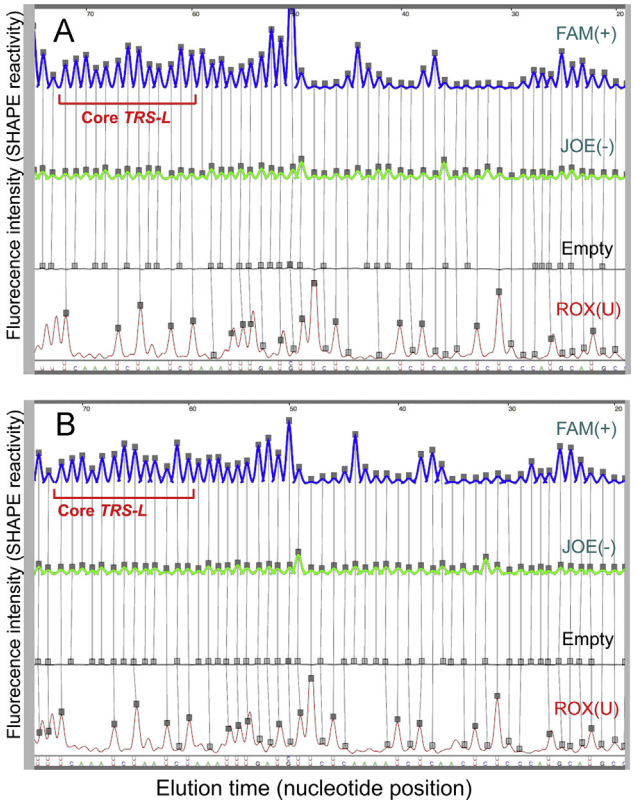

Fig. 2. Electropherograms of a typical region at position nts 20–74 of *in vitro* transcribed and refolded RNA. *In vitro* transcribed and refolded RNA was reacted with either (A) 1M7 or (B) NMIA and the samples were analyzed by primer extension using fluorescently labeled primers complementary to position 261–282. The leader core TRS region spans nts 60–72. The blue channel represents the modification reaction extended with the FAM-labeled primer; the green channel represents the control reaction with DMSO extended with the JOE-labeled primer; the red channel is the sequencing reaction with ddA terminator, displaying a U ladder, and the ROX-labeled primer. The gray channel is empty. The vertical axis, labeled SHAPE reactivity, represents fluorescence intensity and the horizontal axis is the elution time aligned to the nucleotide reference sequence (NC_001846).

secondary structure prediction from RNAstructure, annotated by the nucleotide resolution SHAPE reactivity, is shown in Fig. 4. A similar analysis was performed with the NMIA-derived mean SHAPE reactivity as the input into RNAstructure (Fig. S1), and the secondary structure generated for *in vitro* transcribed and refolded RNA was identical with that obtained using 1M7 as the modification reagent (compare Fig. 4 with Fig. S2).

The structure generated (Fig. 4) is in agreement with our previous characterization of SL1 (Li et al., 2008), SL2 (Liu et al., 2009a, 2007) and predicts the correct structure for SL4 at position nts 80–130 (Yang et al., 2011). These stem–loops serve as *cis*-acting elements required for driving subgenomic RNA synthesis. Interestingly, the stem predicted for SL3 by phylogenetic algorithms (Chen and Olsthoorn, 2010) in the TRS region was single-stranded with relatively high SHAPE reactivity, consistent with the prediction of Liu et al. (2007) that this region was weakly paired or unpaired. The structure was also in good agreement with the two recent models of MHV-A59 RNA secondary structure (Guan et al., 2012, 2011) that identified two additional *cis*-acting replication elements required for optimal viral replication. S5 (Fig. 4) which contains a long-range RNA–RNA interaction (nts 141–167 base paired with nts 363–335) was largely equivalent to the base pairing (nts 141–170 paired with nts 363–332) predicted by Guan et al. (2012). The sole exception is nts 332–334, which are unpaired in the Guan model but are part of the SL5C stem in our model. SL5A (nts 171–225) was identical to the stem–loop

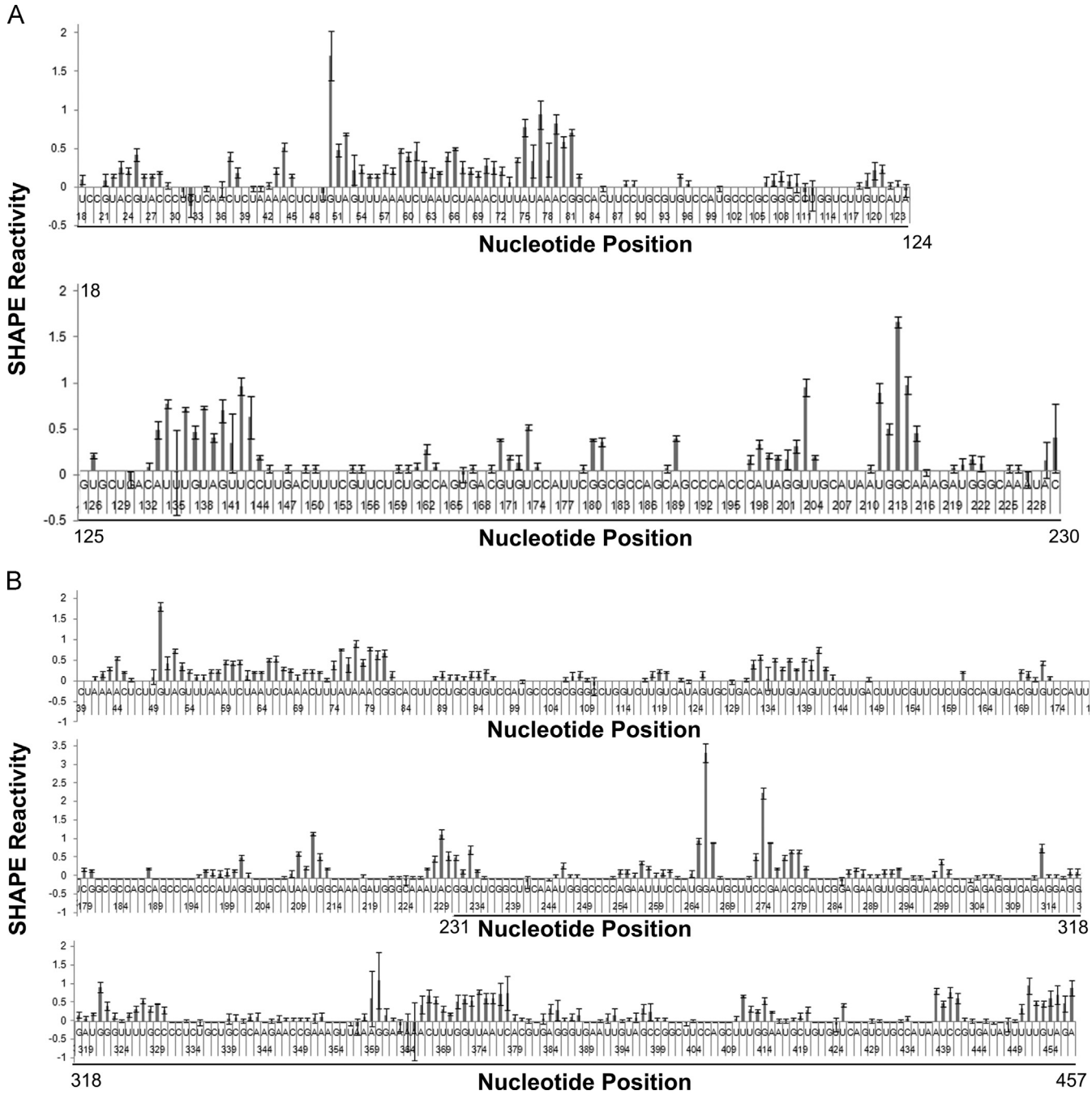

Fig. 3. The mean 1M7-derived quantitative SHAPE reactivity for *in vitro* transcribed and refolded RNA. (A) Data generated with the primer complementary to position 261–282; (B) data generated with the primer complementary to the 3′ end MHV-unrelated unstructured region present in *in vitro* transcribed RNA. All data are the means from independent triplicate experiments. Error bars represent standard deviations from the mean. Continuous SHAPE reactivity results for nts 18–457 (underlined) were obtained by combining data from the analyses with the two primers. We did not obtain reliable results for the 5′-most 17 nts and 3′ end nts 458–474. The nucleotide positions in MHV-A59 genome and their corresponding nucleotide sequences are shown in the horizontal axis.

designated as SLIV previously (Brown et al., 2007; Guan et al., 2011). Furthermore, SL5A is remarkably similar to a stem–loop, predicted and designated SL5 by Chen and Olsthoorn (2010) based upon a structural-phylogenetic analysis of the 5′UTRs of betacoronaviruses. SHAPE analysis has also provided biochemical support for SL5B, SL5C, SL6 and SL7 (although we have no SHAPE reactivity data for nts 458–474 in SL7), which, in MHV, previously lacked support from any genetic or biochemical studies.

Nucleotides 238–262 and nts 284–309 make up a segment of the bulged stem–loop designated SL5B in our SHAPE-generated model, and correspond to an identical stem–loop structure that was predicted as the basal segment of SLV (nts 238–262 and nts 284–309) for betacoronavirus by Mfold and by covariation analysis (Brown et al., 2007), and as part of an unnamed stem–loop by a structural-phylogenetic analysis of betacoronavirus (Chen and Olsthoorn, 2010). The SHAPE-generated model differs from earlier models in that, nts 234–237 are base-paired with nts 314–310 in the base of SL5B but these nucleotides are not included in SLV (Brown et al., 2007) or the corresponding region in Chen model (Chen and Olsthoorn, 2010). Moreover, the structure predicted by SHAPE analysis for nts 263–283 in the terminal part of SL5B agrees with that predicted by the Chen model (Chen and Olsthoorn, 2010), but is different from the stem–loop predicted for the terminal portion of SLV by Brown et al. (2007). BCoV SLV (nts 239–310) is also supported by RNase structure probing

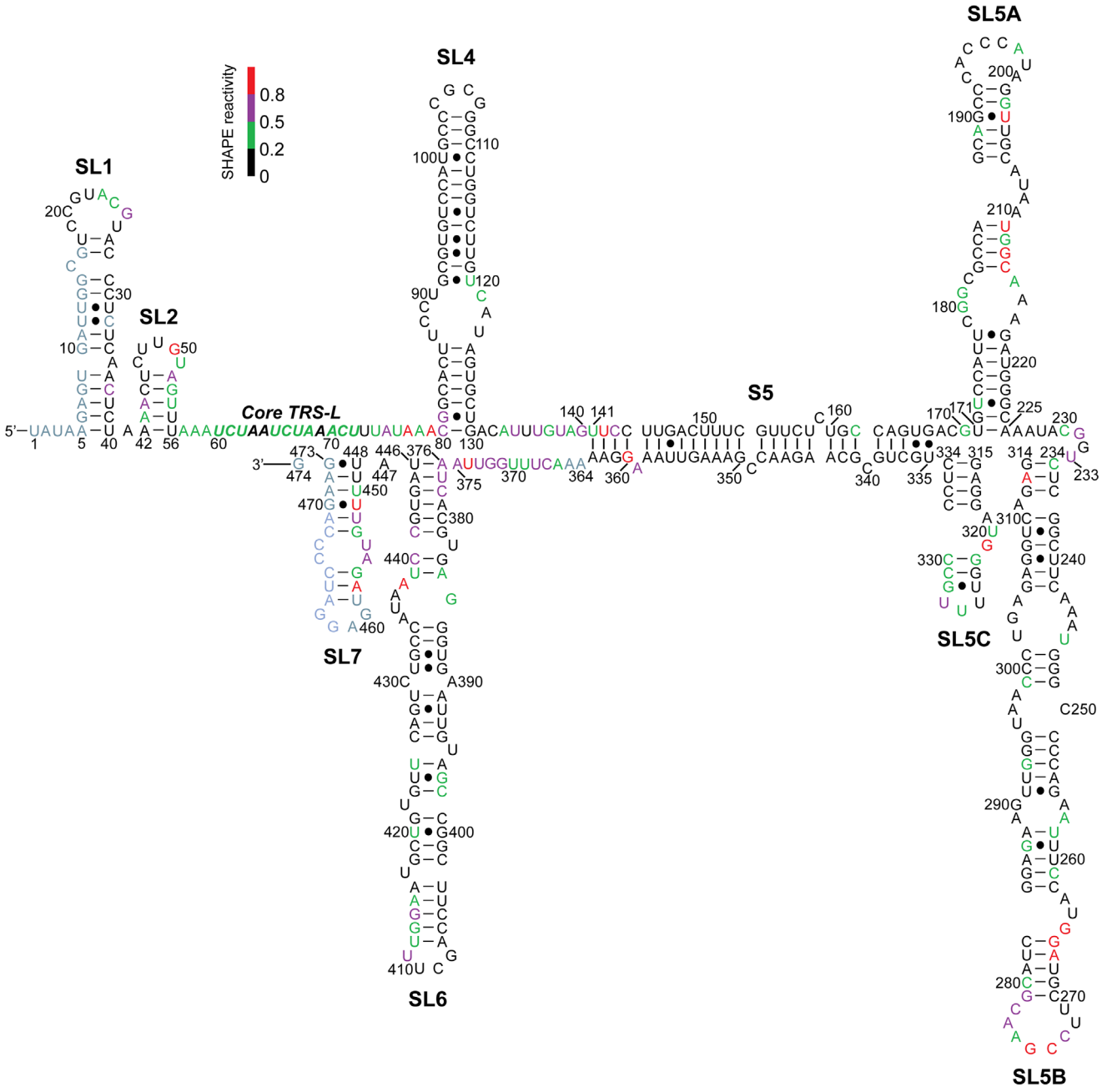

Fig. 4. The secondary structural model of 5′ 474 nts of MHV-A59 generated by SHAPE analysis. *In vitro* transcribed and refolded MHV RNA corresponding to the 5′-most 474 nts of the genome was modified by 1M7 and analyzed by SHAPE using two primers (see text). Mean SHAPE reactivity data was imported into RNAstructure and the structure shown was generated as described in *Materials and methods*. Nucleotides in red have average SHAPE reactivities >0.8 after normalization; nucleotides in purple have average SHAPE reactivities between 0.5 and 0.8; nucleotides in green have average SHAPE reactivities between 0.2 and 0.5; nucleotides in black have average SHAPE reactivities <0.2.

(Brown et al., 2007). SL5C (nts 315–334) is similar to the distal part of SLVI predicted by Brown et al. (2007) and to the corresponding region in the Chen model (Chen and Olsthoorn, 2010). BCoV SLVI has been identified as a *cis*-acting element required for DI RNA replication (Brown et al., 2007). The SHAPE-generated SL6 (nts 376–446) is remarkably similar to SLVIII, predicted by Mfold for betacoronavirus (Brown et al., 2007), but there is no previous structural or functional evidence to support this prediction.

RNA secondary structures predicted by SHAPE analysis of in vitro folded RNA and ex virio RNA are identical

Biochemical studies of *in vitro* transcribed and folded RNAs that represent subsequences of larger RNA species such as viral genomes are subject to difficulties due to inadvertently breaking up interacting base pairs by leaving out of the *in vitro* transcribed RNAs one side of a base-paired structure. SHAPE analysis using intact *ex virio* RNA obviates this particular source of error in using *in vitro* transcribed RNA for biochemical studies of RNA secondary structure. Thus we compared the SHAPE reactivity of the 5′ 474 nts region of interest of *in vitro* synthesized and refolded RNA and *ex virio* viral genomic RNA, both modified by 1M7. *Ex virio* viral genomic RNA that has been gently extracted and deproteinized before 1M7 modification is from authentic virion but contains no bound proteins. We obtained the quantitative single-nucleotide SHAPE reactivity for nts 17–461 by combining the information from two primers that provided overlapping datasets. All reactions and analyses were performed three times and were highly reproducible, and the mean SHAPE reactivity

for each nucleotide was calculated. The pattern of the SHAPE reactivity in electropherograms and the quantitative SHAPE profile of *ex virio* genomic RNA targeting 5′ 474 nts region (see Fig. S3) were very similar to those of *in vitro* transcribed RNA shown in Fig. 3. The resulting secondary structure model (Fig. S4) for *ex virio* RNA was identical to the model generated using the SHAPE reactivity of *in vitro* transcribed RNA (Fig. 4).

SHAPE reactivity of in virio RNA

We next compared the SHAPE reactivity profiles of *in vitro* transcribed RNA and *ex virio* genomic RNA to that of an *in virio* genomic RNA in an effort to identify the differences in SHAPE reactivity that might report on protection mediated by viral protein binding. For the *in virio* state of genomic RNA, purified virus containing the viral genome complexed to the nucleocapsid (N) protein in the virion was treated immediately with NMIA. The *in virio* NMIA modified genomic RNA was then extracted, followed by primer extension using the primer complementary to position 261–282 in order to focus our analysis on the core-leader TRS sequence (nts 60–72) which is known to bind to bacterially expressed N protein N-terminal domain with 10 nM affinity (Baric et al., 1988; Grossoehme et al., 2009; Keane et al., 2012; Nelson et al., 2000; Stohlman et al., 1988). We obtained quantitative single-nucleotide SHAPE reactivity for nts 21–230 from triplicate experiments. Surprisingly, the pattern of SHAPE reactivity in electropherograms and the quantitative SHAPE profile of *in virio* genomic RNA in this region are similar to those of *in vitro* transcribed RNA and *ex virio* genomic RNA (Fig. 5). The core and the adjacent region of leader-TRS demonstrated high SHAPE reactivity with NMIA (Fig. 5A), similar to those of *in vitro* transcribed RNA modified with NMIA (see Fig. 2B) and *ex virio* genomic RNA modified with 1M7 (Fig. 5B). The high SHAPE reactivity of the *in virio* TRS suggests that either intact N protein does not interact strongly with the TRS in RNA packaged into virions (Chang et al., 2013; Chen et al., 2007; Ma et al., 2010; Takeda et al., 2008), or any complex that is formed is such that the 2′ OH moeities are not strongly protected from NMIA modification. A structure of this complex is not yet known although nucleobase-aromatic stacking, which would place the 2′ OH on the "outside" of the protein–RNA complex, is known to occur and is critically important for sgRNA replication (Grossoehme et al., 2009).

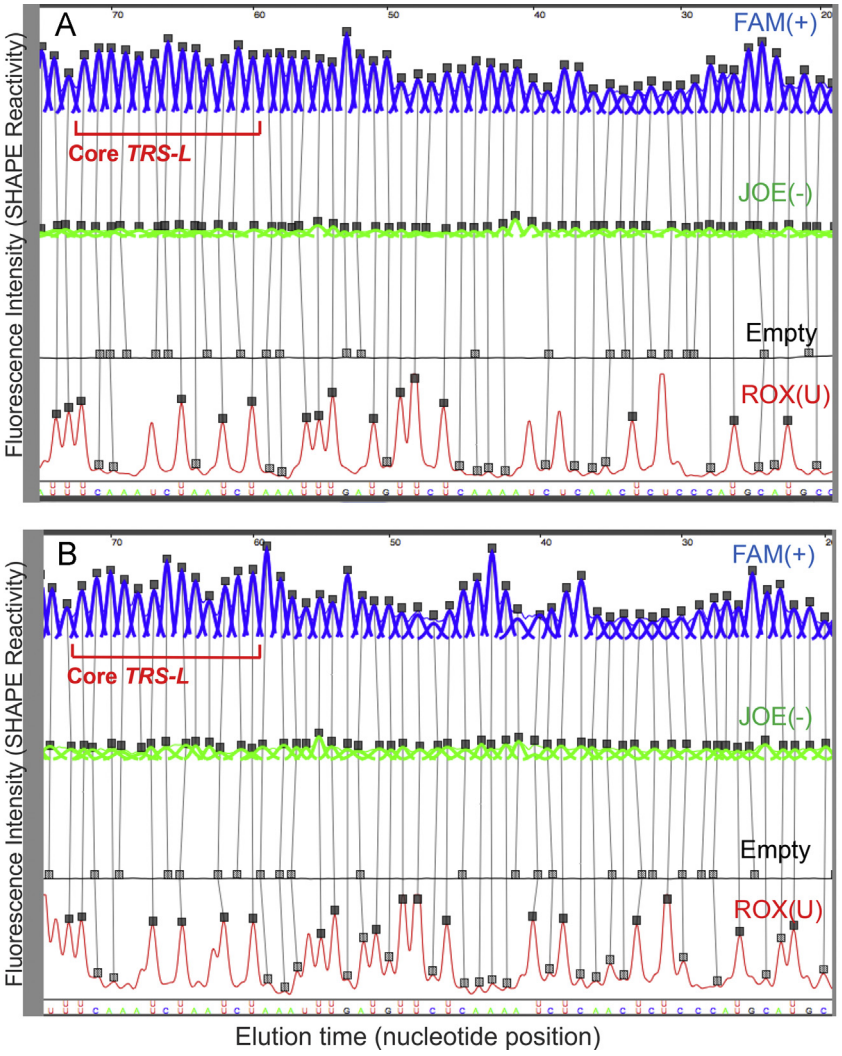

Fig. 5. A comparison of SHAPE reactivities of *in virio* genomic RNA and *ex virio* genomic RNA. The electropherograms of the region from nts 20–74 of (A) *in virio* genomic RNA after NMIA modification and (B) *ex virio* genomic RNA after 1M7 modification using the primer complementary to position 261–282 for primer extension are shown. The leader core TRS region spans nts 60–72. The blue channel represents the modification reaction extended with the FAM-labeled primer; the green channel represents the control reaction with DMSO extended with the JOE-labeled primer; the red channel is the sequencing reaction with ddA terminator, displaying a U ladder, and the ROX-labeled primer. The gray channel is empty. The vertical axis, labeled SHAPE reactivity, represents fluorescence intensity and the horizontal axis is the elution time aligned to the nucleotide reference sequence (NC_001846).

One of the goals in this work was to obtain biological insights of MHV by SHAPE analysis that may not be obtainable using *in vitro* synthesized and refolded RNAs. We compared the RNA secondary structure of the 5′ 474 nts of MHV-A59 genome in the virion (*in virio*) to the structure of genomic RNA that had been gently extracted and deproteinized (*ex virio*) and to the structure of an *in vitro* synthesized RNA, in order to identify differences in SHAPE reactivity among the three biological states of RNA that might allow us to infer protein binding sequences/structures within this region of the genome, as has been done for the HIV nucleocapsid interaction with the HIV genome (Wilkinson et al., 2008). One possibility was that encapsidation of genome RNA by the N protein (Sturman and Holmes, 1983) would decrease overall *in virio* SHAPE reactivity and that the TRS region would display a comparatively lower level of SHAPE reactivity due to tight binding to the N-terminal domain of the N protein (Baric et al., 1988; Grossoehme et al., 2009; Keane et al., 2012; Nelson et al., 2000; Stohlman et al., 1988). However, the pattern of the SHAPE reactivity of *in virio* genomic RNA for nts 21–230, which includes the TRS region, was similar to that of our *in vitro* transcribed RNA and to *ex virio* genomic RNA, and the three states of RNAs generated the same secondary structures in this region.

The secondary structure of the MHV-A59 5′ region obtained by SHAPE analysis is generally similar to the RNAstructure-based models for BCoV and for SARS-CoV

Fig. 1 shows the comparison of the secondary structure of MHV-A59 informed by SHAPE analysis (Fig. 1A) with the most thermodynamically stable models of the 5′ most 474 nts of BCoV-Mebus (Fig. 1B) and SARS-CoV (Fig. 1C) generated by RNAstructure software without incorporating any SHAPE reactivity data, highlighted by conserved SL1, SL2 and SL4 secondary structure elements and a four-helix junction, SL5ABC. In general, the overall configuration of the three models is remarkably similar despite the relatively high divergence of the nucleotide sequence. This includes a conserved SL5ABC four-helix junction, a finding strongly consistent with a conserved core architecture (Laing and Schlick, 2009). The lengths of connecting single-stranded junctions between the helices are also generally similar or identical, although the single stranded junctions between SL5A and the other two helices (S5 and SL5B; see Fig. 1) appear shorter for BCoV than they are in MHV or SARS-CoV. The majority of the RNAstructure-based model for BCoV-Mebus (Fig. 1B) is in good agreement with the model for the 5′UTR and N-terminal nsp1 coding region of BCoV generated by Mfold modeling, phylogenetic covariation, biochemical and genetic studies (Brown et al., 2007; Chen and Olsthoorn, 2010; Guan et al., 2012, 2011; Kang et al., 2006; Liu et al., 2007; Raman et al., 2003; Raman and Brian, 2005), although the nomenclature is not equivalent. Parts of the RNAstructure-based model for SARS-CoV (Fig. 1C) are also similar to the model for the 5′ proximal sequence of SARS-CoV developed through structural-phylogenetic analysis. This is especially true for substructures, SL5A, 5B and 5C, the first two of which have been proposed to function in genomic RNA packaging (Chen and Olsthoorn, 2010). Based upon prior functional studies of SL1-SL4, SL5ABC or the corresponding regions in three CoVs (Brown et al., 2007; Guan et al., 2012, 2011; Kang et al., 2007; Li et al., 2008; Liu et al., 2009a, 2007, 2006; Raman et al., 2003; Raman and Brian, 2005; Yang et al., 2011) we conclude that these secondary structures serve as common *cis*-acting signals important in CoVs replication and viral RNA synthesis, with the structural features of SL1, SL2 and SL4 appearing to be more important than the precise nucleotide sequences in these stem–loops.

A notable difference amongst the three RNAstructure-based models is the presence or absence of SL3. In BCoV and SARS-CoV, SL3 overlaps the leader (TRS-L) which defines the leader–body junction region for discontinuous sgRNA synthesis. In MHV-A59, this region was predicted to be weakly paired or unpaired (Chen and Olsthoorn, 2010; Liu et al., 2007), a prediction supported by our SHAPE analysis (Fig. 4). A second difference amongst the three structures is the position of the initiation codons for nsp1; they are similarly located for MHV and BCoV, in SL5A in MHV and BCoV, whereas the initiation codon for SARS-CoV nsp1 is in a more 3′ position, in the first part of the 3′ side of S5, as shown in Fig. 1. Previously an attempt to create an MHV chimeric virus in which the entire MHV 5′UTR was replaced by the SARS-CoV 5′UTR failed, even when the MHV leader TRS was substituted for the SARS-CoV TRS (Kang et al., 2006). The likely explanation for this failure is the different locations for the initiating AUGs of the two viruses. Thus the replacement of the MHV 5′UTR with the SARS-CoV 5′UTR would disrupt the long-range interaction in S5. Similarly, the BCoV 32 nts region (nts 142–173) could not directly replace the corresponding 30 nts sequences in MHV-A59 without genetic adaptation (Guan et al., 2011), which is likely explained by partial disruption of base-pairing in S5, a long-range interaction between this region (nts 142–173) and sequences approximately 200 nts downstream within the nsp1 coding sequence in both MHV and BCoV. Following S5, the structures generated by RNAstructure are somewhat more divergent amongst the three viruses with MHV-A59 and BCoV-Mebus being relatively similar, although SL6 and SL7 are predicted to be in a forked stem–loop structure for BCoV whereas for MHV they form separate stem loops, and for SARS-CoV, three stem–loops, SL6, SL7 and SL8, are predicted.

SL5C is not required for viral replication

We next performed a series of reverse genetic experiments to determine the functional role of SL5C (nts 315–334) and SL6 (nts 376–446) in MHV-A59 replication. SL5C (nts 315–334) is similar to the distal part of SLVI predicted by Brown et al. (2007) in BCoV and to the corresponding region in the Chen model (Chen and Olsthoorn, 2010). SLV (nts 239–310) and SLVI (nts 311–340) in BCoV RNA have been supported by RNase structure probing and by nucleotide covariation among closely related betacoronaviruses, and SLVI has been identified as a *cis*-acting element required for DI RNA replication (Brown et al., 2007). To determine the functional importance of the secondary structure and sequence of SL5C, four nucleotide mutations on both sides of the SL5C stem predicted by the Vienna RNA and Mfold webservers to disrupt its helical structure (Fig. 6A) were introduced into the cDNA representing the 5′ region of the viral genome as described in *Materials and methods*. These mutations were designed to be coding silent and preserve the translated amino acid sequence in nsp1.

An *in vitro* synthesized RNA containing the SL5C mutations but otherwise identical to the wild type *in vitro* synthesized RNA used previously was reacted with 1M7 and subjected to SHAPE analysis. SHAPE data was imported into RNAstructure, incorporating the SHAPE reactivity as pseudo-free energy constraints as described in *Materials and methods*, and RNA secondary structure generated. The SL5 portion of the predicted secondary structure is shown in Fig. S5. The configuration, position, and stability of the SL5Cstem–loop were significantly changed from those of the wild RNA with an increase in the spacing between SL5B and SL5C, the SL5C stem being shortened and less stable, and the length of the SL5C loop increasing from 3 to 8 nucleotides. Virus was generated from the mutant cDNA as described in *Materials and methods*. Viable viruses containing these mutations were recovered, plaque purified and sequenced to verify the presence of the introduced mutations in the recovered virus. No second-site substitutions were found in the mutated region or in the sequences flanking SL5C (nts 20–500). The recovered mutant virus had only moderate impairment of virus replication compared to that of the WT virus, as assessed by relative

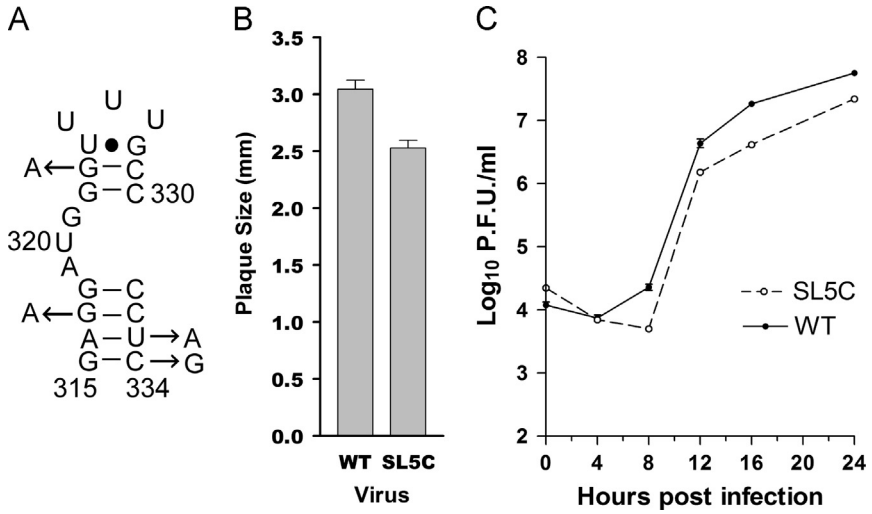

Fig. 6. Mutational study of SL5C. (A) The coding silent mutations introduced into SL5C are shown. (B) Average plaque sizes were determined as described in *Materials and methods*. (C) A one-step growth curve was determined at a MOI=3, as described in *Materials and methods*. Error bars represent the standard errors of the mean.

plaque size and one-step growth kinetics (Fig. 6B and C), suggesting that the wild type *cis*-acting SL5C is not required for viral replication. This contrasts to the data obtained with a BCoV DI model replicon where the destruction of the distal helix that corresponds to SL5C resulted in DI RNA which failed to replicate suggesting that SL5C functioned as a *cis*-acting structure in that context (Brown et al., 2007). We have previously found similar discordant results with MHV DI RNAs and recombinant viruses containing identical mutations in their 3′ UTRs (Johnson et al., 2005), and subsequently with recombinant MHVs containing SL4b mutations (Yang et al., 2011) when compared with a BCoV DI RNA containing similar mutations (Raman et al., 2003). Currently we reason that the nature of DI replication assays, in which a mutant DI RNA must compete with helper wild type virus and recombinant WT DI RNAs that arise during the experiment, allows DI replication experiments to magnify modest decreases in fitness observed in straight forward viral replication assays that focus on recovering viable viruses, making them much more severe or lethal in DI RNA replication assays. An alternative possibility not examined in this work is that any small stem loop is allowed in place of SL5C.

SL6 is also not required for viral replication

For this set of experiments we divided SL6 into four regions using internal bulges as landmarks (Fig. 7A): the base of the stem (nts 376–385 and nts 435–446), the middle portion of the stem (nts 386–398 and nts 423–434), the distal segment of the stem (nts 399–407 and nts 412–422) and the terminal loop (nts 408–411) to examine their functional role in viral replication separately.

In an earlier study of MHV nsp1 Brockway and Denison (2005) constructed a virus, VUSB4, containing a charge-to-alanine mutation [CAC→GCA (H57A)] at nts 378–380. This mutation is predicted by both Vienna RNA and Mfold to destabilize the base of SL6 while leaving the remainder of SL6 intact. VUSB4 was viable although its replication phenotype was somewhat impaired (Brockway and Denison, 2005), indicating that the stem at the base of SL6 is not required for viral replication.

To determine the functional importance of structure and sequence of the remainder of SL6, mutations were introduced into the middle (SL6-A mutations, Fig. 7A) and distal portions (SL6-B mutations) of the SL6 helix and the terminal loop of SL6 were targeted individually or together (SL6-AB mutations) by introducing multiple point mutations into the cloned cDNA representing the 5′

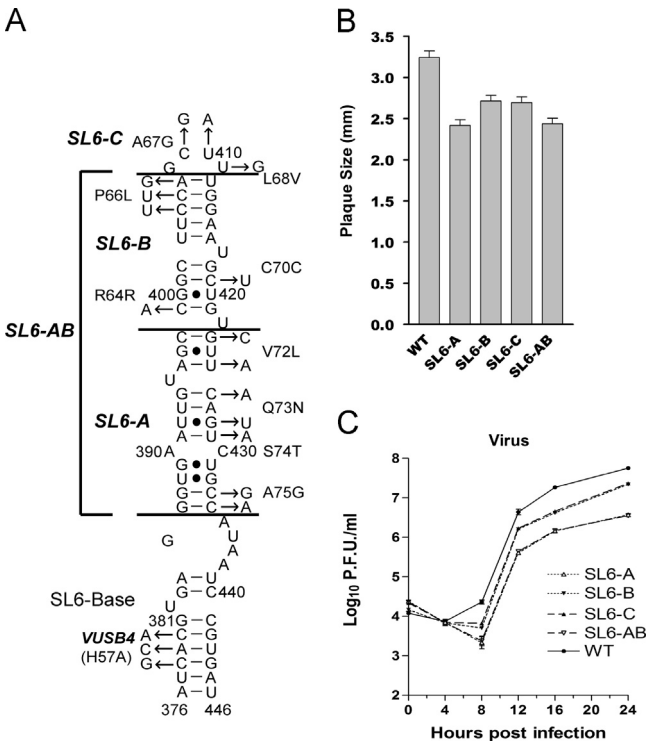

Fig. 7. Mutational study of SL6. (A) SL6 was divided into segments, the base (nts 376–385 and nts 435–446), the middle (nts 386–398 and nts 423–434), the distal portion (nts 399–407 and nts 412–422) and the terminal loop (nts 408–411) and these divisions are reflected by horizontal lines in the figure. The names of all the mutants in the figure are in a bold italicized font. Mutant SL6-A contains multiple mutations in the middle segment of SL6; SL6-B contains multiple mutations in the distal portion of SL6; SL6-C contains mutations in the terminal loop of SL6; SL6-AB includes the mutations from both SL6-A and SL6-B. The arrows show the introduced multiple mutations at the corresponding nucleotide sequences and the resulting amino acid substitutions in each part. Mutations in SL6-Base, CAC→GCA (H57A) at position nts 378–380 are from a charge-to-alanine mutant virus, VUSB4, constructed by Brockway and Denison (2005). The amino acid changes resulting from the introduced nucleotide changes are shown adjacent to the nucleotide changes. The potential structures after introducing multiple mutations at each part were predicted by the Vienna RNA and Mfold webservers. (B) Average plaque sizes of four mutant and wild-type viruses were determined as described in *Materials and methods*. (C) A one-step growth curve was determined at a MOI=3, as described in *Materials and methods*. The error bars represent the standard errors of the mean.

region of the genome and then used the mutated cDNAs to regenerate virus, as described in *Materials and methods*. Mutations were coding silent whenever possible, and where that was not possible we restricted our mutations to those that resulted in conservative (structurally, or present in the corresponding position in BCoV) amino acid changes. Each of the potential structures in the tested regions was predicted using the Vienna RNA and Mfold RNA webservers. The SL6-A mutations (Fig. 7A) were predicted to open up the targeted middle helix region but also were predicted to disrupt the basal helix and form two new stem–loop structures, one in the left side (nts 377–390) of the base of SL6 and the second in the right side (nts 428–444) of the base of SL6, while keeping the structure of the distal region of SL6 unchanged. Mutant SL6-B targeting the distal helix was predicted to completely open up the distal region of the SL6 stem loop and alter the structure of the middle region while keeping the base of SL6 unchanged. To interrogate the terminal loop, mutant SL6-C (Fig. 7A) containing three point mutations (C409G/U410A/U411G) was made. Viable viruses containing the introduced mutations were recovered for each of these three sets of mutations (SL6-A, SL6-B, SL6-C). We then constructed an additional mutant, designated as SL6-AB, which contained the mutations of SL6-A in the middle part of the SL6 helix, with the mutations of SL6-B in the distal portion of the SL6 helix. This mutant was predicted to destroy the overall structure of SL6. We also recovered viable virus containing this mutation. All four mutant viruses had only small to moderate impairments of viral replication compared to those of the WT virus, as assessed by relative plaque size and one-step growth kinetics (Fig. 7B and C). The sequencing results for the mutated regions and the flanking regions (nts 20–500) of three or four plaque-purified isolates of each mutant at virus passage 1 showed that each contained the desired mutations with no other substitutions, with the exception of one isolate of SL6-A which contained a 27-nucleotide insertion between nts 70 and 71 which was a 27-nts repeat of nts 44–70, in addition to the introduced mutations. However, the phenotype of this isolate was identical to the other isolates containing only the introduced SL6-A mutations in terms of relative plaque size and titer achieved (data not shown).

To verify that the predicted mutations in SL6 did result in alterations in SHAPE reactivity and SHAPE informed secondary structure predictions, we separately *in vitro* synthesized three RNAs containing the SL6-A, SL6-B, and SL6-AB mutations, but otherwise were identical in sequence to the wild type *in vitro* synthesized RNA used previously. These RNAs were reacted with 1M7 and subjected to SHAPE analysis. SHAPE data was imported into RNAstructure, incorporating the SHAPE reactivity as pseudo-free energy constraints as described in *Materials and methods*, and RNA secondary structures generated. The corresponding SL6-A mutant SHAPE-informed RNAstructure predicted secondary structure is shown in Fig. S6. The mutations have induced a major structural change from the wild type structure, where the distal portion of SL6 now folds into a bifurcated structure containing two new loops. As can be seen in Fig. S7, the bracketed region, corresponding to the loop spanning nts 416–421, is more SHAPE reactive in the SL6-A mutant RNA than in the wild type RNA. A similar analysis was performed using the SHAPE data obtained with the SL6-B and SL6-AB mutant RNAs. As can be seen in Fig. S6, the SL6-B and SL6-AB mutations also produced large pertubations in the structure of SL6, with the SL6-AB mutation completely restructuring the region between positions 376 and 473. Nevertheless, as noted above, viral genomes containing each of these SL6 mutations were still able to regenerate infectious virus with only modest defects in their ability to replicate (Fig. 7B and C).

SL6 spans nts 376–446 in MHV-A59 and encodes nsp1 amino acid positions 56–79. The results of our mutational study (Fig. 7) demonstrate that MHV SL6 is not essential for viral replication. In a previous study by Brockway and Denison (2005) a genome (designated VUSB5) containing nsp1 charge-to-alanine mutations at nts 399–401, CGG→GCA (R64A), and at nts 414–416 GAA→GCA (E69A) was not viable. These mutations are predicted to destabilize the distal stem of SL6. A second mutation, VUSB6, includes substitutions at nts 441–446, CGUGAU→GCAGCA (R78A, D79A), was also not viable. Because the RNA secondary structure in these putative *cis*-acting regions was not known at the time, Brockway and Denison (2005) could not unequivocally assign the functional effects of these mutations on replication to the amino acid alterations in nsp1. Modeling the effects of these mutations on the secondary structure of SL6 indicates that VUSB5 would destabilize the distal portion of SL6 and VUSB6 is predicted to completely open up the bottom of SL6. In light of the viability of our mutations that greatly alter the structure of SL6 (Fig. S7), and the ability of Brockway and Denison (2005) to recover viable viruses containing mutations (VUSB4) in nsp1 that are predicted to destabilize the base of SL6 very much as did the lethal VUSB5, it is very likely that the lethality of VUSB5 and VUSB6 is due to their effects on nsp1 structure and function rather than due to effects on RNA secondary structure. The structural comparison in Fig. 1 shows that SL6-7 in MHV-A59 diverges somewhat from the corresponding structures in BCoV-Mebus, and is quite different from SL6-8 in SARS-CoV. This is consistent with our functional studies of SL6 which demonstrate that SL6 is not essential for MHV replication, in contrast to structural elements that are entirely within the 5′UTR (SL1, SL2, SL4) or which extend from the 5′UTR into the nsp1 coding sequence (S5), which are lethal or result in viruses that are crippled for viral replication (Guan et al., 2012, 2011; Li et al., 2008; Liu et al., 2009a, 2007; Yang et al., 2011).

Conclusions

In this work, we use SHAPE methodology to determine the secondary structures formed by the 5′-most 474 nts in the MHV-A59 genome. The secondary structures generated by SHAPE analysis are in agreement with those from previous studies. This work is the first description of the secondary structures encompassing the complete minimal 5′ *cis*-acting regulatory region required for replication of MHV-A59 DI RNAs used as model RNA replicons (Kim et al., 1993; Luytjes et al., 1996). The structural model generated for MHV is generally conserved for the other betacoronaviruses. SHAPE analysis can obtain structural information for any size RNA, including complete RNA genomes, and is able to identify structures involving long-range RNA–RNA interaction, such as S5 in this work. Additionally we have clarified our understanding of the *cis*-acting elements that essential for coronavirus replication, demonstrating that the SL5C and SL6 stem–loop structures previously implicated as being essential for DI replication (Brown et al., 2007) are not essential for replication of the intact virus. Furthermore, we successfully establish the use of SHAPE analysis (Merino et al., 2005; Watts et al., 2009) to probe the RNA secondary structure of genomic RNA purified from coronavirus particles rather than on a synthetic RNA fragment. Identical structures were generated by SHAPE analysis of MHV-A59 *ex virio* genomic RNA and an *in vitro* synthesized and refolded RNA representing the 5′-most 489 nts of the genome, giving us a high degree of confidence in the model. Because of its relatively high throughput nature, SHAPE analysis has the potential to provide a secondary structure model for the entire CoV RNA genome purified from virion particles (see Figs. S3 and S4). This methodology will be useful to map the detailed and long range secondary structures within genomic RNA from authentic viral particles, particularly to determine the interactions of the intergenic regions and their flanking sequences which likely play a role in regulating template switching during MHV sgRNA synthesis.

Materials and methods

Cells and virus production

DBT cells were maintained at 37 °C and 5% CO_2 in DMEM supplemented with 10% calf serum. MHV-A59 (Leibowitz et al., 1981) was used to infect DBT cells at MOI=0.1 and propagated for 24 h. The supernatants were harvested and clarified by centrifugation at 10,000 rpm for 30 min. Baby hamster kidney cells expressing the MHV receptor, Ceacam1a, (BHK-R) (Dveksler et al., 1991; Yount et al., 2002) were originally provided by Ralph Baric (University of North Carolina at Chapel Hill) and maintained at 37 °C and 5% CO_2 in DMEM supplemented with 10% calf serum, 10% tryptose phosphate broth, 800 μg/ml of Geneticin (G418 sulfate; Sigma) to select for cells expressing the MHV receptor, 100 I.U./ml of penicillin and 100 μg/ml of streptomycin. L2 cells were maintained at 37 °C and 3% CO_2 in DMEM supplemented with 10% calf serum. MHV-A59-1000 (Yount et al., 2002) was used as our wild type virus for mutagenesis and reverse genetic studies, and all mutant viruses were propagated in DBT cells.

Purification of viruses and extraction of ex virio RNA

Virus particles in clarified supernatants were pelleted by ultracentrifugation (25,000 rpm for 3 h) through a 3 ml 30% (w/w) sucrose cushion in an SW28 rotor. The pelleted viral particles were resuspended in 500 μl of MSE (10 mM MOPS at pH 6.8, 150 mM NaCl, 1 mM EDTA) and overlaid onto a 20–60% sucrose gradient and centrifuged in an SW41 rotor at 25,000 rpm overnight and the virus-containing opalescent band at a buoyant density of 1.17–1.19 was collected. The purified viruses were diluted in MSE, pelleted by ultracentrifugation in an SW41 rotor at 35,000 rpm for 3 h, and then resuspended in 140 μl of PBS. The viral genomic RNAs were extracted using the QIAamp Viral RNA Mini Kit (QIAGEN, Cat. No. 52904), quantitated by UV absorbance at 260 nm, aliquoted (1 pmol), followed by precipitation with 3 M sodium acetate (pH 5.5), ethanol and 20 μg of glycogen.

SHAPE reagents

NMIA was purchased from Invitrogen. 1M7 was synthesized in our laboratory following the procedure outlined in Mortimer and Weeks (2007). The yield of final product was above 70% as determined by NMR spectroscopy.

1M7 modification of ex virio RNA

Modification of *ex virio* genomic RNA was performed as described previously (Mortimer and Weeks, 2007; Wilkinson et al., 2008). *Ex virio* RNA (1 pmol) was dissolved in 14.4 μl of 1/2 × TE, adding 3.6 μl of prewarmed (37 °C) 5 × folding buffer (250 mM HEPES pH 8.0, 1 M potassium acetate pH 8.0, 25 mM $MgCl_2$) and incubated at 37 °C for 10 min. The *ex virio* RNA was divided equally between two tubes and treated with either 1M7 (1 μl, 100 mM in DMSO) or neat DMSO (1 μl) for 70 s at 37 °C. The RNA from each reaction was recovered by precipitation with sodium acetate, ethanol and glycogen, and resuspended in 10.5 μl of 1/2 × TE for primer extension.

Purification of viruses, modification by NMIA and extraction of in virio RNA

Virus particles were purified as described above for preparing *ex virio* RNA, except that the pelleted purified viruses were resuspended in 180 μl of prewarmed (37 °C) 1 × folding buffer (50 mM HEPES pH 8.0, 200 mM potassium acetate pH 8.0, 5 mM $MgCl_2$) and incubated at 37 °C for 15 min to ensure complete resuspension of the pellet. The virus particles were then divided equally between two tubes (90 μl/tube) and treated with NMIA (10 μl, 100 mM in DMSO) or neat DMSO (10 μl) for 45 min at 37 °C. EDTA (2 μl of 100 mM solution) was added to each tube to stop the reactions. The (+) and (−) NMIA reactions were added to 40 μl of PBS to give a total volume of 140 μl/tube. Viral genomic RNAs were extracted using a QIAamp Viral RNA Mini Kit as per the vendor's instructions. The (+) or (−) NMIA modified *in virio* RNA (0.5 pmol) was precipitated with sodium acetate, ethanol and glycogen.

Synthesis and purification of in vitro transcribed RNA

Because detection of RNA adducts by primer extension typically results in the loss of structural data for 40–60 nts at the 3′ end of the RNA (Merino et al., 2005; Steen et al., 2011), the sequence of *in vitro* transcribed RNA was extended to position 489 nts of MHV-A59, and a 24 nts unstructured sequence (5′- GAG TGA CTA CCT CCA AGG CCC TTT) was added to the 3′ end to obtain additional SHAPE data for the target 5′- 474 nts of MHV-A59. A DNA template for *in vitro* transcription encoding the 5′ 489 nts of the MHV-A59 genome plus the 24 nts unstructured region was generated by PCR. The forward primer incorporated a SpeI site (underlined) and a promoter for T7 RNA polymerase in the 5′ end: 5′- CGG CGG CAT GCT AAT ACG ACT CAC TAT AGG GTA TAA GAG TGA TTG GCG TCC G; the reverse primer incorporated a XbaI site (underlined) in the 5′ end: 5′- CGC GTT CTA GAA AAG GGC CTT GGA GGT AGT CAC TCT AGT TGA GGC CTC CAC CTT CTG GGG ATC CTC ATC TAC. Plasmid A (Yount et al., 2002) containing the first 4882 nts of the MHV-A59 genome under the control of a T7 promoter was mixed with the forward and reverse primers and amplified by using 5 PRIME Master Mix (5PRIME) to generate a DNA product of 555 bp, which after SpeI and XbaI digestion was cloned into plasmid pUC18 similarly digested with SpeI and XbaI. The resulting plasmid was then linearized by XbaI and RNA was synthesized *in vitro* following the manufacturer's protocol for the AmpliScribe™ T7-Flash™ Transcription Kit (Epicentre). The RNA transcript was resolved and purified by denaturing polyacrylamide gel electrophoresis (6% polyacrylamide, 7 M Urea, 1 × TBE). Full-length RNA transcript was located by UV-shadowing, excised from the gel, and eluted at 4 °C into 0.5 M sodium acetate (pH 5.2) overnight. The eluted RNA was ethanol precipitated, quantitated and stored at −80 °C. The homogeneity of the *in vitro* transcribed RNA was analyzed by non-denaturing gel electrophoresis and a single band of RNA was observed suggesting that the RNA was unlikely to have folded into multiple conformations.

Modification of in vitro transcribed RNA

In vitro transcribed RNA (2 pmol) in 14.4 μl of 1/2 × TE was refolded by heating at 95 °C for 3 min, placing on ice for 2 min, adding 3.6 μl of prewarmed (37 °C) 5 × folding buffer (250 mM HEPES pH 8.0, 1 M potassium acetate pH 8.0, 25 mM $MgCl_2$) and incubating at 37 °C for 30 min. The folded RNA was divided equally between two tubes and treated with either 1M7 (1 μl, 100 mM in DMSO) or neat DMSO (1 μl) for 70 s at 37 °C. The RNA from the (+) and (−) 1M7 reagent experiments was recovered by precipitation and resuspended in 10.5 μl of 1/2 × TE for primer extension. In parallel, the same amount of folded RNA was also treated by either NMIA (1 μl, 100 mM in DMSO) or neat DMSO (1 μl) for 45 min at 37 °C and recovered by ethanol precipitation for primer extension.

Primer synthesis and labeling for SHAPE analysis

Two sets of primers were designed for detection of *ex virio* RNA and *in vitro* transcribed RNA targeting the MHV-A59 5′-most 474 nts. A primer (5′- ATG CGT TCG GAA GCA TCC ATG G) complementary to position 261–282 at the mid-point of the target sequence was used

for both *in vitro* transcribed and virion RNAs. A primer (5′- CTA CCG AAC TGC AAT GCC ATC A) complementary to position 490–511 was used for analyses of genome RNA modified *ex virio* or *in virio* to obtain the SHAPE data for the target 5′ 474 nts of MHV-A59, and the primer (5′- GGC CTT GGA GGT AGT CAC TC) complementary to the non-MHV unstructured extension of the 3′ end of *in vitro* transcribed RNA. Primers contained a 5′ six carbon linker terminating in a primary amine. The amine-tethered DNA primers were labeled with one of three fluorophores, 5-FAM, 6-JOE, or 5-ROX purchased from AnaSpec, Inc. following a protocol from Invitrogen for labeling amine-modified oligonucleotides. Labeled primers were precipitated and purified by denaturing gel electrophoresis (20% 29:1 acrylamide/bisacrylamide). Spectrophotometric measurements indicated labeling efficiencies were greater than 95% as determined by the [dye]/[DNA] absorbance ratios.

Primer extension and capillary electrophoresis for SHAPE analysis

In vitro transcribed RNA (1 pmol), *ex virio* RNA or *in virio* RNA (0.5 pmol) from the (+) and (−) 1M7 or NMIA reactions were used as templates for primer extension. 5-FAM fluorescently labeled primers (1 μl, 2 pmol/μl) were mixed with the 1M7 (+) or NMIA (+) treated RNAs, whereas 6-JOE labeled primers (1 μl, 2 pmol/μl) were added to the control RNAs (−) treated with vehicle (DMSO). Primers were annealed to the RNA by incubating at 65 °C for 5 min followed by 35 °C for 10 min, and then placed on ice. Primer extension was initiated by addition of reverse transcriptase mix (4 μl of 5 × first-strand buffer, 2 μl of 0.1 M DTT, 1 μl of 10 mM dNTPs), and annealing at 52 °C for 1 min, followed by the addition of Superscript III reverse transcriptase (1 μl, 200 U, Invitrogen) and continued incubation at 52 °C for 15 min. Reactions were terminated with EDTA (2 μl, 100 mM). *In vitro* transcribed RNA not treated with 1M7 or NMIA served as template for a single dideoxy (ddATP) sequencing reaction to align peaks of SHAPE reactivity to the sequence of the *in vitro* transcribed RNA. For the sequencing reaction of *ex virio* RNA, a longer *in vitro* transcribed RNA of approximate 900 nts (1 pmol, 9.5 μl in 1/2 × TE) was used as template. This RNA was the *in vitro* transcription product from ScaI linearized Plasmid A DNA template and was purified by PAGE. The sequencing reactions were carried out using the same procedure used for primer extension with (+) or (−) 1M7/NMIA reactions with the exception that a 5-ROX labeled primer (1 μl, 2 pmol/μl) was used and ddATP (1 μl, 10 mM) (Invitrogen) was added to the reverse transcriptase mix. For each state of RNA, the three primer extension products: the (+) 1M7/NMIA reactions, the (−) 1M7/NMIA reactions, and the ddA sequencing reaction, were combined and precipitated with ethanol in the presence of sodium acetate and glycogen. cDNA fragments were resolved by capillary electrophoresis on an Applied Biosystems ABI 3130 Genetic Analyzer. Three independent experiments were done using independently prepared RNA (*in vitro* transcribed or from purified virus) as starting material.

SHAPE data processing

Raw electropherograms containing fluorescence intensity *versus* elution time were analyzed using ShapeFinder software that was generously provided by the authors (Vasa et al., 2008). Raw sequencer traces were processed for fitted baseline adjustment, signal decay correction, filter convolution and scale factor. Quantitative SHAPE reactivities for individual datasets were normalized by first excluding the negative peak intensities which contained high background in the (−) reagent control reaction, calculating the average of the 8% most reactive peak intensities, and then dividing all of positive peak intensities by this average. SHAPE reactivities from each primer extension reaction were normalized independently. For each state of RNA, the continuous SHAPE reactivity results targeting the 5′ 474 nts of the MHV-A59 genome were obtained by combining the information from two overlapping reads produced from the two primers covering the target, meaning that the first part of the SHAPE data was generated from the primer complementary to positions 261–282, and the second part of the data was generated from the 3′ primer corresponding to nts 490–511 for genome RNAs. For *in vitro* transcribed RNAs, a primer corresponding to the added 3′ unstructured region was used in place of the 490–511 primer.

Incorporation of SHAPE intensities into RNAstructure

The combined SHAPE reactivity results targeting the 5′ 474 nts of MHV-A59 genome for each state of RNA were imported into RNAstructure (software can be downloaded from http://rna.urmc.rochester.edu/RNAstructureDownload.html) incorporating pseudo-free energy constraints using $\Delta G_{SHAPE} = m\ln$ [SHAPE reactivity + 1.0] + b. To determine the pairing persistence, structures were computed with two sets of values for the b and m parameters, $m = 1.7$ and $b = -0.6$ kcal/mol (Gherghe et al., 2008) or $m = 2.6$ and $b = -0.8$ kcal/mol (default in RNAstructure software). Maximum pairing distance was set to No limitations. In order to generate an accurate secondary structure for SL1 and SL2, SL2 (see Fig. 1) was forced to be single-stranded at positions 47 through 51 and base-paired between positions 42 through 46 and positions 56 through 52. In previous studies, SL2 has been predicted to be highly conserved in all coronaviruses examined by minimum free energy secondary structural modeling using ViennaRNA and by covariation analysis of the stem. Its existence and functional importance in subgenomic RNA synthesis and virus replication have been demonstrated by biophysical and reverse genetic studies (Lee et al., 2011; Liu et al., 2009a, 2007). However, if SL2 is not forced to form, RNAstructure generates alternative structures in the SL1 and SL2 region (not shown) without affecting the rest of the secondary structures in the 1-474 target RNA sequence. These same alternative structures were also previously predicted in BCoV by Mfold (Chang et al., 1994, 1996). The final structure for each modification condition was based on the most stable structure generated. Three short (2–4 nts) base-pairings with high SHAPE reactivity were manually removed because they were predicted to have Tm values ≈ 20 °C and thus would not be stable under physiologic conditions, and were not present when SL2 was not forced to base pair.

Construction of mutations in SL5C and SL6

Plasmid A was used as a substrate for site-directed mutagenesis (Yount et al., 2002). Mutations were introduced into the desired regions of SL5C or SL6 in MHV-A59 with the QuikChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The sequences of the oligonucleotides used for mutagenesis are available on request. Mutagenized plasmids were sequenced between unique MluI and BamHI restriction sites to verify the introduced mutations. To minimize the possibility that the plasmids used to generate mutant viruses contain unintended mutations introduced outside the MluI–BamHI fragment, the sequenced MluI–BamHI fragments containing the introduced mutations were excised and ligated with MluI–BamHI-digested parental unmutagenized plasmid A to generate the plasmids used to produce mutant viruses. The sequences of these plasmids between the MluI and BamHI sites were all verified again by sequencing.

Assembly of full-length MHV-A59 cDNAs and recovery of mutant viruses

cDNAs representing the entire MHV-A59 genome, carrying either the wild-type sequence or mutations in SL5C or SL6, were

constructed by sequential ligation of digested and gel purified cDNAs A to G and the ligated cDNA transcribed *in vitro* using the Ambion T7 mMESSAGE mMACHINE kit (Applied Biosystems), as described previously (Johnson et al., 2005; Yount et al., 2002). The transcription reactions were then electroporated into BHK-R cells that were then overlaid onto DBT cells. Cultures were incubated for up to 72 h and monitored by phase microscopy for the development of cytopathic effect (CPE). Recovered viruses were plaque purified on L2 cell monolayers and subsequently propagated in DBT cells. The mutated region and the flanking region (nts 20–500) were amplified by RT-PCR and sequenced to confirm the presence of the introduced mutations and to identify any second-site mutations in the neighborhood that might have arisen during virus recovery and propagation. For each mutant, at least two (usually 3) plaque-purified isolates at passage 1 (P1) were selected for sequencing.

Plaque size determination and growth curve assays

Plaque assays were performed in L2 cells as described previously (Johnson et al., 2005). Plaque diameters were calculated from 30 well-formed plaques for each virus. Growth curves were performed by infecting replicate wells of DBT cells in 96-well plates with mutant or wild type viruses at a multiplicity of infection (MOI) of 3. Triplicate infected cultures were harvested at 0, 4, 8, 12, 16, and 24 h postinfection (h.p.i). Virus titers were determined by plaque assays. The Kruskal–Wallis test and one-way ANOVA were used for statistical comparisons amongst mutant and WT viruses. Error bars represent the standard errors of the mean.

Acknowledgments

The authors gratefully acknowledge support from the US National Institutes of Health Grants, AI067416 and AI051493. We would like to thank Dr. Kevin Weeks and Philip Homan at the University of North Carolina for providing the protocol and advice for the synthesis of 1M7 and for providing us with the SHAPE-Finder software. We also want to thank Dr. Veronica Sanchez at Texas A&M HSC for allowing us to install and use SHAPEFinder software on her computer. We also thank Dr. Xiangyu Shi and Larry Harris-Haller of the Texas A&M Gene Technologies Laboratory for allowing us to use the ABI3130 analyzers located there and for their help in operating the instrument. We also thank Drs. Zhen Mahoney and Mitchell Jamal of Life Technologies for advice on standardizing the ABI3130 for the SHAPE analyses.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2014.11.001.

References

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

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

Brockway, S.M., Denison, M.R., 2005. Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. Virology 340, 209–223.

Brown, C.G., Nixon, K.S., Senanayake, S.D., Brian, D.A., 2007. An RNA stem–loop within the bovine coronavirus nsp1 coding region is a cis-acting element in defective interfering RNA replication. J. Virol. 81, 7716–7724.

Chang, C.K., Chen, C.M., Chiang, M.H., Hsu, Y.L., Huang, T.H., 2013. Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. PLoS One 8, e65045.

Chang, R.Y., Hofmann, M.A., Sethna, P.B., 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., 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.Y., Chang, C.K., Chang, Y.W., Sue, S.C., Bai, H.I., Riang, L., Hsiao, C.D., 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, S.C., Olsthoorn, R.C., 2010. Group-specific structural features of the 5′-proximal sequences of coronavirus genomic RNAs. Virology 401, 29–41.

Deigan, K.E., Li, T.W., Mathews, D.H., Weeks, K.M., 2009. Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. USA 106, 97–102.

Denison, M., Perlman, S., 1987. Identification of putative polymerase gene product in cells infected with murine coronavirus A59. Virology 157, 565–568.

Dveksler, G.S., Pensiero, M.N., Cardellichio, C.B., Williams, R.K., Jiang, G.S., Holmes, K.V., Dieffenbach, C.W., 1991. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. J. Virol. 65, 6881–6891.

Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G., Weeks, K.M., 2008. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. J. Am. Chem. Soc. 130, 12244–12245.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res. 117, 17–37.

Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann 3rd, C.E., Leibowitz, J.L., Giedroc, D. P., 2009. Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS–cTRS RNA duplexes. J. Mol. Biol. 394, 544–557.

Guan, B.-J., Su, Y.-P., Wu, H.-Y., Brian, D.A., 2012. Genetic evidence of a long-range RNA–RNA interaction between the genomic 5′ untranslated region and the nonstructural protein 1 coding region in murine and bovine coronaviruses. J. Virol. 86, 4631–4643.

Guan, B.J., Wu, H.Y., Brian, D.A., 2011. An optimal cis-replication stem–loop IV in the 5′ untranslated region of the mouse coronavirus genome extends 16 nucleotides into open reading frame 1. J. Virol. 85, 5593–5605.

Gustin, K.M., Guan, B.J., Dziduszko, A., Brian, D.A., 2009. Bovine coronavirus nonstructural protein 1 (p28) is an RNA binding protein that binds terminal genomic cis-replication elements. J. Virol. 83, 6087–6097.

Johnson, R.F., Feng, M., Liu, P., Millership, J.J., Yount, B., Baric, R.S., Leibowitz, J.L., 2005. Effect of mutations in the mouse hepatitis virus 3′(+)42 protein binding element on RNA replication. J. Virol. 79, 14570–14585.

Kang, H., Bhardwaj, K., Li, Y., Palaninathan, S., Sacchettini, J., Guarino, L., Leibowitz, J. L., Kao, C.C., 2007. Biochemical and genetic analyses of murine hepatitis virus Nsp15 endoribonuclease. J. Virol. 81, 13587–13597.

Kang, H., Feng, M., Schroeder, M.E., Giedroc, D.P., 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.

Keane, S.C., Liu, P., Leibowitz, J.L., Giedroc, D.P., 2012. Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. J. Biol. Chem. 287, 7063–7073.

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

Kladwang, W., VanLang, C.C., Cordero, P., Das, R., 2011. Understanding the errors of SHAPE-directed RNA structure modeling. Biochemistry 50, 8049–8056.

Laing, C., Schlick, T., 2009. Analysis of four-way junctions in RNA structures. J. Mol. Biol. 390, 547–559.

Lee, C.W., Li, L., Giedroc, D.P., 2011. The solution structure of coronaviral stem–loop 2 (SL2) reveals a canonical CUYG tetraloop fold. FEBS Lett. 585, 1049–1053.

Leibowitz, J.L., Wilhelmsen, K.C., Bond, C.W., 1981. The virus-specific intracellular RNA species of two murine coronaviruses: MHV-a59 and MHV-JHM. Virology 114, 39–51.

Li, L., Kang, H., Liu, P., Makkinje, N., Williamson, S.T., Leibowitz, J.L., Giedroc, D.P., 2008. Structural lability in stem–loop 1 drives a 5′ UTR-3′ UTR interaction in coronavirus replication. J. Mol. Biol. 377, 790–803.

Lin, Y.J., Lai, M.M., 1993. Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontiguous sequence for replication. J. Virol. 67, 6110–6118.

Liu, P., Leibowitz, J.L., 2010. RNA higher-order structures within the coronavirus 5′ and 3′ untranslated regions and their roles in viral replication. In: Lal, S.K. (Ed.), Molecular Biology of the SARS-Coronavirus. Springer, Heidelberg, Dordrecht, London, New York, pp. 47–61.

Liu, P., Li, L., Keane, S.C., Yang, D., Leibowitz, J.L., Giedroc, D.P., 2009a. Mouse hepatitis virus stem–loop 2 adopts a uYNMG(U)a-like tetraloop structure that is highly functionally tolerant of base substitutions. J. Virol. 83, 12084–12093.

Liu, P., Li, L., Millership, J.J., Kang, H., Leibowitz, J.L., 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, P., Millership, J.J., Li, L., Giedroc, D.P., 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.

Liu, Y., Wimmer, E., Paul, A.V., 2009b. Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochim. Biophys. Acta 1789, 495–517.

Luytjes, W., Gerritsma, H., Spaan, W.J., 1996. Replication of synthetic defective interfering RNAs derived from coronavirus mouse hepatitis virus-A59. Virology 216, 174–183.

Ma, Y., Tong, X., Xu, X., Li, X., Lou, Z., Rao, Z., 2010. Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA–protein binding mechanism in coronavirus. Protein Cell 1, 688–697.
Merino, E.J., Wilkinson, K.A., Coughlan, J.L., Weeks, K.M., 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J. Am. Chem. Soc. 127, 4223–4231.
Mortimer, S.A., Weeks, K.M., 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J. Am. Chem. Soc. 129, 4144–4145.
Nelson, G.W., Stohlman, S.A., 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.
Raman, S., Bouma, P., Williams, G.D., 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., 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.
Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. J. Virol. 81, 20–29.
Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., 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.
Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.D., Rappuoli, R., 2003. SARS—beginning to understand a new virus. Nat. Rev. Microbiol. 1, 209–218.
Steen, K.A., Siegfried, N.A., Weeks, K.M., 2011. Selective 2′-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) for direct analysis of covalent adducts and of nucleotide flexibility in RNA. Nat. Protoc. 6, 1683–1694.
Stohlman, S.A., Baric, R.S., Nelson, G.N., Soe, L.H., Welter, L.M., Deans, R.J., 1988. Specific interaction between coronavirus leader RNA and nucleocapsid protein. J. Virol. 62, 4288–4295.
Sturman, L.S., Holmes, K.V., 1983. The molecular biology of coronaviruses. Adv. Virus Res. 28, 35–112.
Takeda, M., Chang, C.K., Ikeya, T., Guntert, P., Chang, Y.H., Hsu, Y.L., Huang, T.H., Kainosho, M., 2008. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J. Mol. Biol. 380, 608–622.
Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., Giddings, M.C., 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA 14, 1979–1990.
Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess Jr., J.W., Swanstrom, R., Burch, C.L., Weeks, K.M., 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711–716.
Weiss, S.R., Leibowitz, J.L., 2007. Pathogenesis of murine coronavirus infections. In: Perlman, S., Gallagher, T., Snijder, E.J. (Eds.), Nidoviruses. ASM Press, Washington, DC, pp. 259–278.
Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., Weeks, K.M., 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol. 6, e96.
Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., Giddings, M.C., Weeks, K.M., 2009. Influence of nucleotide identity on ribose 2′-hydroxyl reactivity in RNA. RNA 15, 1314–1321.
Yang, D., Liu, P., Giedroc, D.P., Leibowitz, J., 2011. Mouse hepatitis virus stem–loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. J. Virol. 85, 9199–9209.
Yount, B., Denison, M.R., Weiss, S.R., Baric, R.S., 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. J. Virol. 76, 11065–11078.