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

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

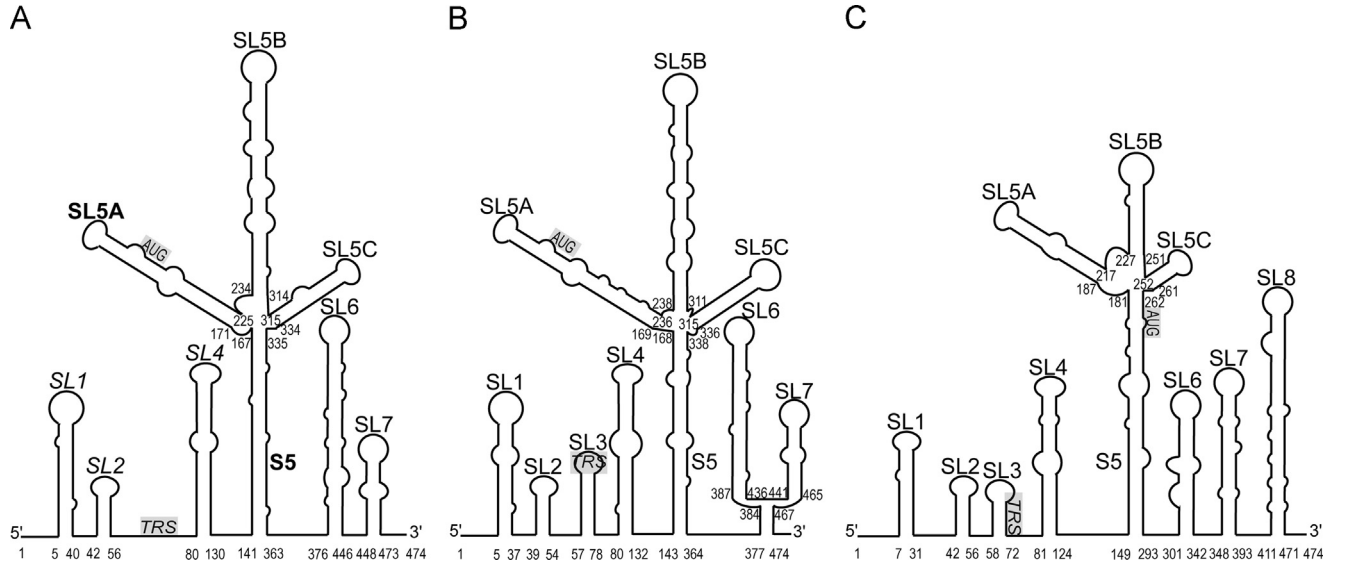

Fig. 1. Comparison of secondary structure models of the 5′- most 474 nts of MHV-A59, BCoV, and SARS-CoV. (A) The MHV-A59 model was generated by SHAPE analysis. Italicized labels indicate structural elements essential for virus recovery and subgenomic (−) sense RNA synthesis. Bolded labels indicate structural 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.