Virology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Virology

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

# Structural and functional conservation of *cis*-acting RNA elements in coronavirus 5'-terminal genome regions

Ramakanth Madhugiri[a], Nadja Karl[a], Daniel Petersen[a], Kevin Lamkiewicz[b,d], Markus Fricke[b,d], Ulrike Wend[a], Robina Scheuer[a], Manja Marz[b,c,d], John Ziebuhr[a,d,*]

[a] Institute of Medical Virology, Justus Liebig University, Giessen, Germany
[b] Faculty of Mathematics and Computer Science, Friedrich Schiller University, Jena, Germany
[c] FLI Leibniz Institute for Age Research, Jena, Germany
[d] European Virus Bioinformatics Center, Jena, Germany

ARTICLE INFO

*Keywords:*
Coronavirus
Replication
*cis*-acting RNA element
Stem-loop
RNA structure
Coronavirus phylogeny

ABSTRACT

Structure predictions suggest a partial conservation of RNA structure elements in coronavirus terminal genome regions. Here, we determined the structures of stem-loops (SL) 1 and 2 of two alphacoronaviruses, human coronavirus (HCoV) 229E and NL63, by RNA structure probing and studied the functional relevance of these putative *cis*-acting elements. HCoV-229E SL1 and SL2 mutants generated by reverse genetics were used to study the effects on viral replication of single-nucleotide substitutions predicted to destabilize the SL1 and SL2 structures. The data provide conclusive evidence for the critical role of SL1 and SL2 in HCoV-229E replication and, in some cases, revealed parallels with previously characterized betacoronavirus SL1 and SL2 elements. Also, we were able to rescue viable HCoV-229E mutants carrying replacements of SL2 with equivalent betacoronavirus structural elements. The data obtained in this study reveal a remarkable degree of structural and functional conservation of 5′-terminal RNA structural elements across coronavirus genus boundaries.

## 1. Introduction

*Cis*-acting RNA elements play important roles in the life cycle of plus-strand (+) RNA viruses, including RNA replication, viral gene expression and genome packaging (Barton et al., 2001; Liu et al., 2009b; Firth and Brierley, 2012; Goto et al., 2013; Kuo and Masters, 2013; Morales et al., 2013; Nicholson and White, 2014; Keane et al., 2015). Compared to many other +RNA viruses, information on *cis*-acting RNA elements of coronaviruses, including their specific functions, structures and interactions, remains limited. Particularly, this applies to viruses from genera outside the genus *Betacoronavirus* (for reviews, see Brian and Baric, 2005; Masters, 2007; Liu and Leibowitz, 2010; Madhugiri et al., 2014; Yang and Leibowitz, 2015; Madhugiri et al., 2016). Historically, RNA structures and sequences required for (beta)coronavirus RNA synthesis were characterized using defective interfering (DI) RNA-based systems (Chang et al., 1994, 1996; Raman et al., 2003; Raman and Brian, 2005; Brown et al., 2007; Gustin et al., 2009). Thus, for example, RNA structure probing studies of mouse hepatitis virus (MHV) and bovine coronavirus (BCoV)-derived RNAs led to the identification of up to four stem-loops within the 5′-terminal 215 nt of the genome (for recent reviews, see Liu and Leibowitz, 2010; Madhugiri et al., 2014, 2016; Yang and Leibowitz, 2015). In many cases, potential functional roles of RNA structural elements present in the 5'-terminal genome region could be confirmed by mutational analyses. More recently, genus- and subfamily-wide RNA structure-based alignments using all currently approved coronavirus species in the respective genera of the *Coronavirinae* were performed for this highly divergent genome region. The studies led to a model of three highly conserved stem-loop structures, called SL1, SL2, and SL4, in the 5′-terminal, ~150-nt genome region (Kang et al., 2006; Liu et al., 2007; Chen and Olsthoorn, 2010; Madhugiri et al., 2014). Furthermore, nuclear magnetic resonance (NMR) spectroscopy provided structural support for SL1 and SL2 in three betacoronaviruses, MHV, BCoV, and HCoV-OC43 (Liu et al., 2007, 2009a; Li et al., 2008). Also, a selective 2′-hydroxyl acylation and primer extension (SHAPE) analysis *in virio* and *ex virio* confirmed the predicted SL1, SL2, and SL4 structures for MHV-A59 (Yang et al., 2015).

Possible biological functions of betacoronavirus 5′-terminal SL1 and SL2 structures in viral replication could be substantiated by reverse genetics studies (Kang et al., 2006; Liu et al., 2007, 2009a; Li et al., 2008). For example, an MHV study revealed that destabilization of the upper part of SL1 produces viruses with replication defects, while

* Correspondence to: Institute of Medical Virology, Biomedical Research Center, Justus Liebig University, Schubertstr. 81, 35392 Giessen, Germany.
*E-mail address:* john.ziebuhr@viro.med.uni-giessen.de (J. Ziebuhr).

https://doi.org/10.1016/j.virol.2017.11.025
Received 1 October 2017; Received in revised form 28 November 2017; Accepted 29 November 2017

compensatory mutations restoring these base-pairing interactions led to viruses with near-wildtype growth kinetics (Li et al., 2008). In contrast, disruption of the basal part of SL1 was largely tolerated, while compensatory mutations that restored these base-pairing interactions proved to be lethal, suggesting a critical role for the RNA sequence (rather than structure) in this lower part of SL1. Based on these and other data, SL1 was suggested to require an optimal stability suitable to establish transient long-range (RNA- and/or protein-mediated) interactions between the 5′- and 3′-UTRs that may be required for genome replication and subgenomic (sg) mRNA synthesis. Other reverse genetics studies confirmed that the 5′-terminal SL2 is also required for MHV RNA synthesis (Liu et al., 2007, 2009a). Based on phylogenetic analyses, the SL2 was proposed to be the most conserved RNA secondary structure in coronaviruses (Kang et al., 2006; Liu et al., 2007; Chen and Olsthoorn, 2010). It is composed of a 5-bp stem and a conserved loop sequence, 5′-CUUGY-3′, that was shown to adopt a 5′-uCUYG(U)a-3'- or a 5'-uYNMG(U)a-3′-like tetraloop structure (Liu et al., 2009a).

To extend these studies and corroborate predictions on alphacoronavirus-associated 5'-terminal RNA structural elements, we used a combination of bioinformatics, biochemical and reverse genetics approaches, focusing on structures and functions of the 5′-terminal SL1 and SL2 structures in the HCoV-229E genome (genus *Alphacoronavirus*). The data obtained in this study provide evidence for the existence of two SL structures (SL1 and SL2) in the ~80-nt, 5'-terminal HCoV-229E and HCoV-NL63 genome regions. The structures were found to be required for viral replication and appear to be (largely) conserved between alpha- and betacoronaviruses. Thus, for example, we were able to show that the HCoV-229E SL2 structure can be replaced with that of the betacoronaviruses BCoV and SARS-CoV, respectively, providing experimental support for our previous hypothesis that (some) RNA structural elements in coronavirus untranslated genome regions may be more conserved than previously thought, even across genus boundaries (Madhugiri et al., 2014).

## 2. Material and methods

### 2.1. Cells and viruses

Wildtype HCoV-229E and HCoV-229E mutants were propagated in Huh-7 cells. HCoV-229E titers were determined by plaque assay using Huh-7 cells. Recombinant vaccinia viruses were propagated in CV-1 and BHK-21 cells, and plaque purifications of single virus clones were performed using CV-1 and D980R cells as described previously (Isaacs et al., 1990; Thiel et al., 2001).

### 2.2. Mutagenesis of the HCoV-229E full-length cDNA clone

HCoV-229E mutants (HCoV-229E_C11G, _C16G, _G45C, _C47G, _C11G-G34C, _C16G-G29C, _G45C-C55G, and _C47-G53C) were generated using the recombinant vaccinia virus vHCoV-inf-1, which contains a full-length HCoV-229E cDNA (GenBank accession number NC_002645). Site-directed mutagenesis of the HCoV-229E cDNA insert in vHCoV-inf-1 was done using previously described methods (Thiel et al., 2001). To construct vHCoV-inf-1 derivatives containing nucleotide substitutions in the HCoV-229E 5′-UTR, we used the plasmid pBS-5′GPT for recombination with vaccinia virus vHCoV-inf-1. This pBluescriptII-derived plasmid was constructed to contain the *E. coli gpt* gene flanked by (i) a 500-bp fragment representing the vaccinia DNA sequence located upstream of the HCoV-229E cDNA insert in vHCoV-inf-1 and (ii) a 500-bp fragment representing the cDNA sequence of nts 1001–1500 of the HCoV-229E genome RNA. Next, a *gpt*-positive vHCoV-inf-1 derivative, called vRec-5′GPT, was selected from CV-1 cells infected with vHCoV-inf-1 and transfected with pBS-5′ plasmid DNA. In a second selection step, D980R cells were infected with vRec-5′GPT and transfected with an appropriate pBS-5′UTR-mut plasmid DNA. Using appropriate selection conditions (Hertzig et al., 2004), *gpt*-negative vHCoV-inf-1 derivatives (called vHCoV_5'UTR-mut) that contained the desired mutation(s) in the 5' UTR cDNA sequence were isolated. The pBS-5′UTR-mut plasmid constructs used to produce the recombinant vHCoV_5'UTR-mut vaccinia viruses contained the 500-bp vaccinia virus sequence described above followed by a cDNA copy of HCoV-229E nts 1–1500, with appropriate mutations being introduced by PCR-based mutagenesis. Sequences of vHCoV_5'UTR-mut vaccinia virus constructs were verified by Southern blotting and sequence analysis as described (Thiel et al., 2001). The vHCoV-inf-1 derivatives generated in this study were called vHCoV_5′UTR-C11G, vHCoV_5′UTR-C16G, vHCoV_5′UTR-G45C, vHCoV_5′UTR-C47G, vHCoV_5′UTR-C11G + G34C, vHCoV_5′UTR-C16G + G29C, vHCoV_5′UTR-G45C + C55G, and vHCoV_5′UTR-C47G + G53C. Genome-length HCoV-229E RNAs were prepared by T7-based *in vitro* transcription (RiboMAX Large Scale RNA Production System, Promega) using purified genomic DNA from vHCoV-inf-1 and its mutant derivatives, respectively. 1.25 μg of *in vitro*-transcribed genome-length RNAs and 0.75 μg of *in vitro*-transcribed HCoV-229E nucleocapsid (N) protein mRNA (Schelle et al., 2005; Almazan et al., 2006) were used to transfect $1 \times 10^6$ Huh-7 cells using the TransIT® mRNA transfection kit according to the manufacturer's instructions (Mirus Bio LLC). At 72 h posttransfection (p.t.), cell culture supernatants were collected to determine viral titers, and total RNA was isolated for subsequent Northern blot and genome sequence analyses.

### 2.3. RNA extraction and Northern blot analysis

At 72 h p.t., intracellular RNA was extracted using TRizol reagent (Invitrogen) according to the manufacturer's instructions. To analyze viral RNAs by Northern blot hybridization, 10 μg total RNA was denatured for 10 min at 65 °C in loading buffer (50% deionized formamide, 18% formaldehyde, 1x MOPS) and separated in a 1% (w/v) agarose and 2.2 M formaldehyde-containing, 1x MOPS-buffered gel at 16 V for 16–17 h. The gel was soaked in buffer A (50 mM NaOH, 150 mM NaCl) for 30 min and then in buffer B (100 mM Tris-HCl / pH 7.5, 150 mM NaCl) for 30 min. Next, the RNA was transferred onto a positively charged nylon membrane by vacuum blotting. The RNA was cross-linked to the membrane and hybridized with an [α-$^{32}$P]dCTP-labeled DNA probe specific for HCoV-229E nucleotides 26857–27277 and the negative-strand complement of this sequence (TaKaRa Bio Inc). Following hybridization, membranes were rinsed 2 times with 2x SSC/0.01% (w/v) SDS at room temperature and 2 times with 0.2x SSC/0.01% (w/v) SDS at 55 °C for 30 mins. Hybridization signals were visualized by autoradiography using a Typhoon 9200 imager (GE Healthcare).

### 2.4. Genome sequence analysis of virus progeny

At 72 h p.t., cell culture supernatants were collected (passage zero [p0]) and used to determine virus titers and plaque sizes (see below). From the cell pellet, total RNA was extracted using TRIzol reagent (Invitrogen). Following reverse transcription (RT)-PCR amplification, the 5' and 3'-terminal HCoV-229E genome regions (nts 25–750 and nts 25323–27317, respectively) were sequenced. The following primer pairs were used to produce two amplicons for subsequent sequence analyses: (1) HCoV-229E-25up (5′-ACTTAAGTACCTTATCTATCTACAG-3′) and HCoV-229E-750dn (5′-GAAATTATCATCAATGGTCATACTTAC-3′) and (2) HCoV-229E-25323up (5′-CATGGAATCCTGAGGTTAATGCAATC-3′) and HCoV-229E-oligo(dT) (5′-TTTTTTTTTTGTGTATCCATATCG-3′). To determine the 5′-terminal nts 1–25 of progeny virus genomes, the FirstChoice™ RLM-RACE kit was used according to the manufacturer's instructions (Invitrogen). PCR products used for sequence analyses were gel purified (innuPREP Gel Extraction Kit, Analytik Jena) and subjected to automated Sanger sequencing (LGC Genomics). The 5′-UTR and 3'-UTR amplicons were sequenced using oligonucleotides HCoV-229E-750dn and HCoV-229E-27317dn,