Virus Research 206 (2015) 120–133

Contents lists available at ScienceDirect

Virus Research

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

Review

The structure and functions of coronavirus genomic 3′ and 5′ ends

Dong Yang[a], Julian L. Leibowitz[b,*]

[a] Department of Microbiology, Immunology & Biochemistry, The University of Tennessee Health Science Center College of Medicine, Memphis, TN 38163, USA
[b] Department of Microbial Pathogenesis and Immunology, Texas A&M University, College of Medicine, College Station, TX 77843-1114, USA

ARTICLE INFO

Article history:
Available online 28 February 2015

Keywords:
Coronaviruses
RNA secondary structure
cis-Acting sequences
Virus replication
RNA binding proteins

ABSTRACT

Coronaviruses (CoVs) are an important cause of illness in humans and animals. Most human coronaviruses commonly cause relatively mild respiratory illnesses; however two zoonotic coronaviruses, SARS-CoV and MERS-CoV, can cause severe illness and death. Investigations over the past 35 years have illuminated many aspects of coronavirus replication. The focus of this review is the functional analysis of conserved RNA secondary structures in the 5′ and 3′ of the betacoronavirus genomes. The 5′ 350 nucleotides folds into a set of RNA secondary structures which are well conserved, and reverse genetic studies indicate that these structures play an important role in the discontinuous synthesis of subgenomic RNAs in the betacoronaviruses. These *cis*-acting elements extend 3′ of the 5′UTR into ORF1a. The 3′UTR is similarly conserved and contains all of the *cis*-acting sequences necessary for viral replication. Two competing conformations near the 5′ end of the 3′UTR have been shown to make up a potential molecular switch. There is some evidence that an association between the 3′ and 5′UTRs is necessary for subgenomic RNA synthesis, but the basis for this association is not yet clear. A number of host RNA proteins have been shown to bind to the 5′ and 3′ *cis*-acting regions, but the significance of these in viral replication is not clear. Two viral proteins have been identified as binding to the 5′ *cis*-acting region, nsp1 and N protein. A genetic interaction between nsp8 and nsp9 and the region of the 3′UTR that contains the putative molecular switch suggests that these two proteins bind to this region.

Contents

* Corresponding author at: 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. Tel.: +1 979 436 0313; fax: +1 979 845 3479.
E-mail address: jleibowitz@tamu.edu (J.L. Leibowitz).

http://dx.doi.org/10.1016/j.virusres.2015.02.025
0168-1702/

1. Introduction

Coronaviruses (CoVs) are an important cause of illness in humans and animals. Most human coronaviruses commonly cause relatively mild respiratory illnesses; however two zoonotic coronaviruses, SARS-CoV and MERS-CoV, can cause severe illness and death. Investigations over the past 35 years have illuminated many aspects of coronavirus replication. The focus of this review is the structural and functional analyses of conserved RNA secondary structures in the 5′ and 3′ of the betacoronavirus genomes.

1.1. Classification and pathogenicity of coronaviruses

Coronaviruses belong to the subfamily Coronavirinae (http://ictvonline.org/virusTaxonomy.asp?version=2012), which together with Torovirinae make up the Coronaviridae family in the order Nidovirales. The Coronavirinae are classified into four genera Alphacoronavirus, Betacoronavirus, Deltacoronavirus and Gammacoronavirus. Torovirinae includes two genera, Torovirus and Bafinivirus. The Coronaviridae comprises a group of evolutionary related single-stranded, positive-sense, non-segmented, enveloped RNA viruses of vertebrates. The RNA genomes are 25–31 kb, the largest genomes of all known RNA viruses, and are infectious when introduced into permissive cells. However, unlike those of almost all other positive-strand RNA viruses, the RNA infectivity of transfected coronavirus genomes is greatly increased in the presence of a source of N protein (Casais et al., 2001; Grossoehme et al., 2009; Yount et al., 2000). Alphacoronaviruses include alphacoronavirus 1 (transmissible gastroenteritis virus, TGEV), porcine epidemic diarrhea virus (PEDV), bat coronavirus 1, BtCoV 512, BtCoV-HKU8, BtCoV-HKU2, human coronavirus HCoV-NL63 and HCoV-229E. Gammacoronaviruses include avian coronavirus and whale coronavirus SW1. Deltacoronaviruses include coronavirus HKU11, HKU12, and HKU13. The major emphasis of this review is on the betacoronaviruses, which has been the most studied genus. Within the genus Betacoronavirus, four lineages (A, B, C, and D) each with a unique set of accessory genes are commonly recognized. Lineage A includes HCoV-OC43 and HCoV-HKU1, betacoronavirus 1 (more commonly known as bovine coronavirus, BCoV), murine coronavirus (MHV); Lineage B includes severe acute respiratory syndrome-related SARS-CoV and various species recovered from bats; Lineage C includes Tylonycteris bat coronavirus HKU4 (BtCoV-HKU4), Pipistrellus bat coronavirus HKU5 (BtCoV-HKU5). Since April 2012, the Middle East Respiratory Syndrome MERS-CoV has emerged as a new member in lineage C of the betacoronaviruses, closely related to bat coronaviruses HKU4 and HKU5 (de Groot et al., 2013; Drexler et al., 2014; Zaki et al., 2012). MERS-CoV is the first Betacoronavirus lineage C member isolated from humans. Lineage D includes Rousettus bat coronavirus HKU9 (BtCoV-HKU9), which has only been detected in bats (http://www.ecdc.europa.eu/en/publications/Publications/novel-coronavirus-rapid-risk-assessment-update.pdf).

Coronaviruses (CoVs) cause respiratory, enteric, hepatic and neurological diseases in a broad range of vertebrate species (Stadler et al., 2003; Weiss and Leibowitz, 2007). Most human coronaviruses commonly cause relatively mild respiratory disease, however two coronaviruses, SARS-CoV (Rota et al., 2003) and MERS-CoV (Zaki et al., 2012) can cause severe illness and death. SARS-CoV was first recognized in China in November 2002 causing a worldwide outbreak including 774 deaths from 2002 to 2003. MERS-CoV is a novel coronavirus first reported in Saudi Arabia in 2012 and has caused illness in hundreds of people from several countries (http://www.cdc.gov/coronavirus/about/index.html). As of July 23, 2014, 837 laboratory-confirmed cases of MERS-CoV infection have been reported by WHO, including 291 deaths. Both SARS-CoV and MERS-CoV are thought to have originated in bats and spread to humans through intermediate hosts (Coleman and Frieman, 2014). Human coronaviruses also have been detected in human CNS and are able to replicate in CNS derived cells (Arbour et al., 1999; Murray et al., 1992) as well as having been isolated from patients with gastroenteritis and diarrhea (Gerna et al., 1985; Resta et al., 1985), and more seriously, causing neonatal necrotizing enterocolitis (Rousset et al., 1984).

2. Genome organization and replication

2.1. Genome organization

Coronaviruses are roughly spherical with a fringe of large, bulbous surface projections. Coronaviruses infect cells primarily by binding of the spike protein to its host specific cell receptors (Delmas et al., 1992; Hofmann et al., 2005; Li et al., 2003; Raj et al., 2013; Williams et al., 1991; Yeager et al., 1992). Virus enters cells by fusion at the cell surface or by an endocytotic pathway depending upon the strain of virus and the target cell (Nash and Buchmeier, 1997; Wang et al., 2008). After entering the cytoplasm and uncoating, the virus particle releases the RNA genome. For all coronaviruses the genomes are organized into 5′ non-structural protein coding regions comprising the replicase genes, which are two-thirds of the genome, and 3′ structural and nonessential accessory protein coding regions (Masters, 2006). Infected cells contain seven to nine virus specific mRNAs with coterminal 3′ ends, the largest of which is the genomic RNA (Masters, 2006). All of the mRNAs carry identical 70–90 nts leader sequences at their 5′ ends (Lai et al., 1983, 1984; Leibowitz et al., 1981; Spaan et al., 1982). The 3′ end of the leader sequence contains the transcriptional regulatory sequence (TRS-L), which is also present in the genome just upstream of the coding sequence for each transcription unit [TRS-B (body)], where it acts as a *cis*-regulator of transcription (Budzilowicz et al., 1985). All coronavirus TRSs include conserved 6–8 nucleotides core sequence (CS) plus variable 5′ and 3′ flanking sequences (Sola et al., 2005). Betacoronaviruses contain a consensus heptameric sequence, 5′-UCUAAAC-3′, with the SARS-CoV TRS having 5′-ACGAAC-3′ as the core sequence (Marra et al., 2003; Rota et al., 2003). Replication occurs shortly after entry and uncoating of the virion through production of full-length genomic and subgenomic negative strand intermediates (Baric and Yount, 2000; Sawicki and Sawicki, 1990; Sethna et al., 1989). Translation of subgenomic mRNAs gives rise to structural and nonstructural viral proteins. The replicated RNA genome is then encapsidated and packaged into virions. A minimal 69-nts packaging signal has been characterized in MHV that maps within ORF1b approximately 20 kb from the 5′ end of the genome, which is sufficient for RNA to be incorporated into virions (Fosmire et al., 1992; Kuo and Masters,