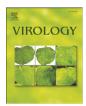
Contents lists available at ScienceDirect

### Virology



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

# Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs

#### Shih-Cheng Chen, René C.L. Olsthoorn \*

Leiden Institute of Chemistry, Department of Molecular Genetics, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

#### ARTICLE INFO

Article history: Received 22 December 2009 Returned to author for revision 6 January 2010 Accepted 5 February 2010 Available online 4 March 2010

Keywords: Coronavirus SARS-CoV RNA secondary structure Packaging signal 5' Untranslated region

#### Introduction

The emergence of the *Severe acute respiratory syndrome coronavirus* (SARS-CoV) in 2003 has boosted related research and led to the discovery of many novel coronaviruses (CoVs) from different hosts such as equines, whales, birds, and bats; the latter species are considered as the potential reservoir of SARS-CoV (Guan et al., 2003, Ksiazek et al., 2003; Li et al., 2005; Marra et al., 2003; Mihindukulasuriya et al., 2008; Woo et al., 2007, 2009; Zhang et al., 2007). In the past few years, also two novel human CoVs, NL63 and HKU1, have been identified causing rather severe symptoms in infants and the elderly (van der Hoek et al., 2004; Woo et al., 2005). The discovery of so many novel CoVs calls for a better understanding of the phylogeny of CoVs.

Based on serological patterns and genome organization, the genus *Coronavirus* has been classified into three major groups: group I, II and III (Lai and Cavanagh, 1997; Brian and Baric, 2005). More recently, these groups have been further subdivided into, in total, 9 subgroups, based upon amino acid similarity of structural and non-structural proteins (nsp) (Snijder et al., 2003; Woo et al., 2006, 2007; Woo et al., 2006, 2007). However, other studies propose at least 5 distinct lineages (Tang et al., 2006; Dong et al., 2007; Vijaykrishna et al., 2007), and even for SARS-CoV there is discussion whether it represents a separate lineage (Rota et al., 2003) or is an early split-off of group II CoVs (Snijder et al., 2004). Thus, in addition to the conventional pair-wise

#### ABSTRACT

Global predictions of the secondary structure of coronavirus (CoV) 5' untranslated regions and adjacent coding sequences revealed the presence of conserved structural elements. Stem loops (SL) 1, 2, 4, and 5 were predicted in all CoVs, while the core leader transcription-regulating sequence (L-TRS) forms SL3 in only some CoVs. SL5 in group I and II CoVs, with the exception of group IIa CoVs, is characterized by the presence of a large sequence insertion capable of forming hairpins with the conserved 5'-UUYCGU-3' loop sequence. Structure probing confirmed the existence of these hairpins in the group I *Human coronavirus-229E* and the group II *Severe acute respiratory syndrome coronavirus* (SARS-CoV). In general, the pattern of the 5' *cis*-acting elements is highly related to the lineage of CoVs, including features of the conserved hairpins in SL5. The function of these conserved hairpins as a putative packaging signal is discussed.

© 2010 Elsevier Inc. All rights reserved.

comparison of viral protein sequences, other genetic or structural features may be helpful in the classification of CoVs.

In the genome of CoVs, like that of most RNA viruses, the 5' and 3' untranslated regions (UTRs) usually harbor important structural elements which are involved in replication and/or translation (Chang et al., 1994; Raman et al., 2003; Raman and Brian, 2005; Goebel et al., 2007; Züst et al., 2008; Liu et al., 2009). In *Mouse hepatitis virus* (MHV), a group II CoV, a bulged stem-loop and a pseudoknot structure were identified in the 3' UTR (Goebel et al., 2004a). Similar pseudoknot structures were found in other group I and II CoVs, showing structural conservations of the CoV 3' UTR (Goebel et al., 2004a). However, the 3' UTR of MHV could be functionally replaced by the 3' UTR of group II SARS-CoV but not by that of the group I *Transmissible gastroenteritis virus* (IBV), indicating certain group-specific functions for the 3' UTR (Goebel et al., 2004b).

In this study the secondary structures of the 5' UTRs and the 5'proximal sequences of the ORF1ab gene in all known CoVs were predicted. The structural features of this region turned out to reflect the known grouping of CoVs, which is based on amino acid similarity. The unique and conserved features were further investigated in detail.

#### **Results and discussion**

### The clustering of the 5'-proximal sequence of CoV RNAs shows group specificity

The clustering of the CoV 5'-proximal 420 nucleotides (nts) obtained from the *Kalign* webserver (see Materials and methods)



<sup>\*</sup> Corresponding author. Leiden Institute of Chemistry, Department of Molecular Genetics, Gorlaeus Laboratories, Einsteinweg 55, 2333 CC, Leiden, The Netherlands. Fax: +31 715274357.

E-mail address: olsthoor@chem.leidenuniv.nl (R.C.L. Olsthoorn).

<sup>0042-6822/\$ -</sup> see front matter © 2010 Elsevier Inc. All rights reserved. doi:10.1016/j.virol.2010.02.007

		•		0.55		Accession
Clustering of coronavirus		Group	TRS-L	. uORF	ORF1ab	number
		$\square$	$\square$	$\frown$	$\square$	$\frown$
	TGEV/Purdue	la	94	117-128	315	NC 002306
	TGEV/TS	la	94	117-128	315	DQ201447
	FIPV/79-1146	la	94	117-128	312	NC_007025
	FCoV/Black	la	93	116-127	312	EU186072
	FCoV/CIJe	la	92	115-126	311	DQ848678
	HCoV-229E/inf-1	lb	66	86-121,102-116	293	NC 002645
	HCoV-NL63/AMS-I	lb	66	101-118	287	NC_005831
	HCoV-NL63/AMS-496	lb	66	101-118	287	DQ445912
	BtCoV-HKU2	lb	69	98-118,119-130	297	NC_009988
	BtCoV-1A	lc	62	87-104,146-181	272	NC_010437
	BtCoV-1B	lc	63	88-105,147-182	273	NC_010436
	BtCoV-HKU8	lc	63	88-105,147-182	270	NC_010438
Ч	PEDV/CV777	ld	67	99-137	297	NC_003436
	PEDV/LZC	ld	67	99-179	297	EF185992
	BtCoV-512/2005	ld	70	97-135	294	NC_009657
	IBV/Beaudette	Illa	57	131-166	529	NC 001451
[]	IBV/Beaudette-p65	Illa	57	131-166	529	DQ001339
	IBV/Peafowl	Illa	57	131-166	529	AY641576
	TCoV/MG10	Illa	57	131-163	529	NC_010800
	CoV-SW1	IIIb	72	-	524	NC 010646
	BCoV	lla	65	100-126	211	NC_003045
	HEV	lla	65	100-126	211	NC 007732
	HCoV-OC43	lla	65	100-126	211	NC_005147
	HCoV-HKU1	lla	66	98-118	206	NC_006577
	MHV-A59	lla	66	99-125	210	NC_001846
	SARS-CoV/Tor2	llb	67	104-136	265	NC_004718
	SARS-CoV/TJF	llb	66	103-135	264	AY654624
	BtSCoV-Rp3	llb	67	104-136	265	NC 009693
	BtSCoV-HKU3-1	llb	65	102-134	262	NC_009694
	BtSCoV-279/2005	llb	65	104-136	262	DQ648857
	BtSCoV-Rm1	llb	66	104-133	261	NC_009696
	BtSCoV-273/2005	llb	67	104-133	261	DQ648856
	BtCoV-HKU9-4	lld	71	-	229	EF065516
	BtCoV-HKU9-2	lld	69	-	228	EF065514
	BtCoV-HKU9-3	lld	70	-	229	EF065515
	BtCoV-HKU9-1	lld	71	-	229	NC_009021
	BtCoV-HKU4-1	llc	63	140-163	267	NC_009019
	BtCoV-HKU4-4	llc	63	140-163	267	EF065508
	BtCoV-133/2005	llc	56	133-156	260	NC_008315
	BtCoV-HKU5-1	llc	61	141-164	261	NC_009020
	BtCoV-HKU5-2	llc	61	140-163	260	EF065510
	BuCoV-HKU11/796	llic	72	-	607	NC_011548
	ThCoV-HKU12/600	llic	65	-	592	NC_011549
L	MuCoV-HKU13/3514	llld	64	-	595	NC_011550

Fig. 1. Clustering and general features of the 5' 420 nucleotides of CoVs. The tree is based on a multiple sequence alignment using ClustalW2 at the European Bioinformatics Institute webserver. The phylogenetic group, the start of core TRS-L, the region of upstream ORF (uORF), the start of ORF1ab, and GenBank accession number of each CoV are listed.

basically resembled the current grouping system for CoVs (Fig. 1), though group I CoVs may be further subdivided into 4 subgroups, groups Ia to Id, according to their relatively large phylogenetic distances (Fig. 1). Sequence comparison further showed conserved and unique features for each CoV group, including: (i) the relative location of the core sequence of the leader transcription-regulating sequence (L-TRS) is quite conserved in all CoVs, except for the one in group Ia CoVs which has a rather long leader sequence upstream of the core TRS; (ii) the potentially translatable short ORF upstream of the genomic ORF1ab, the uORF, is present in most CoVs except for group Ild, IIIb, IIIc, and IIId CoVs; (iii) the 5' UTR in group III CoVs is substantially longer than that in group I and II CoVs, while group IIa CoVs have an exclusively short 5' UTR (Fig. 1). It has to be noted that in order to obtain a higher threshold of the phylogenetic distance, strains with the highest sequence variation were used for analysis (selected from the genomic sequences of all CoVs available in GenBank). This made it more promising if homology was found within a cluster. To further examine if particular features found in the RNA sequence in each group are relevant to specific organization of the 5' *cis*-acting elements, we globally predicted the secondary structures of the CoV 5' UTRs, predominantly using computational calculations at the *mfold* webserver (Zuker, 2003). We have identified several conserved stem–loop (SL) structures in this region, some of which are organized in a group-specific manner (see Figs. 2, 3, and 4).

**Fig. 2.** The structural-phylogenetic analysis of the 5'-proximal sequences in group I CoVs. The predicted secondary structures of the 5'-proximal sequence of (A) group Ia TGEVpurdue, (B) group Ib HCoV-229E-inf-1, (C) group Ic PEDV-CV777, and (D) group Id BtCoV-1A coronaviruses are shown. Nucleotide variations located in the conserved elements in the other representative CoVs of each subgroup are indicated. The start codon of the ORF1ab is boxed, the core sequence of the transcription-regulating leader (TRS-L CS) is bracketed, and the length of the sequence insertion in SL5 is indicated.

Download English Version:

## https://daneshyari.com/en/article/6141760

Download Persian Version:

https://daneshyari.com/article/6141760

Daneshyari.com