Full Length Research Paper

Comparative analysis of human coronavirus-NL63 ORF3 protein homologues

Burtram C. Fielding* and Tasnim Suliman

Molecular Virology Laboratory, Department of Medical BioSciences, University of the Western Cape, Department of Natural Sciences, Western Cape, South Africa.

It has been reported in some studies that the newly discovered human coronavirus NL-63 (HCoV-NL63) is one of the most common coronaviruses associated with acute respiratory infections. HCoV-NL63 was first isolated in 2004 from a 7 month old infant in Holland. The HCoV-NL63 genome encodes for one accessory protein, ORF3. This reports the computational analysis of human coronavirus NL63 ORF3 by comparing the amino acid sequences of coronavirus ORF3-homologues. The HCoV-NL63 ORF3 gene was found to encode a putative protein ~25.6 kDa in size. ORF3 was predicted to contain three potential transmembrane regions. The amino acid sequence of HCoV-NL63 ORF3 was shown to be most similar to HCoV 229E ORF4 (43% identity; 62% similarity).

Key words: Human coronavirus, NL-63, accessory protein ORF3, gene analysis.

INTRODUCTION

Coronaviruses (CoVs) belong to the family Corona-viridae in the order nidovirales. They are divided into groups 1, 2 and 3, based mostly on genetic similarities. These viruses are enveloped, plus-stranded RNA viruses with large genomes. Coronavirus genomes encode for 5′ replicase polyproteins (ORF1a and ORF1b) and 3′ structural proteins in the conserved order, 5′-spike (S) - envelope (E) - membrane (M) - nucleocapsid (N) -3′ (Holmes and Lai, 1996). Accessory genes are found interspersed among the structural genes, which vary in number and location between the different coronavirus groups. These genes are poorly character-rized and the vast majority has no known function. Initial research into the functions of these genes has shown that they are non-essential and dispensable for virus growth in cell culture (Ontiveros et al., 2001). More recently though, using reverse genetic and targeted recombination systems, studies have shown that the accessory genes are required for *in vivo* infection and pathogenecity in the natural host (de Haan et al., 2002; Ortego et al., 2002; Haijema et al., 2004).

It has been reported that human coronaviruses account for a significant number of hospitalizations for children under 18 years of age, accounting for 4.4% of all admissions for acute respiratory infections. Of these, the newly discovered human coronavirus NL63 (HCoV-

NL63) was the most common coronavirus identified (Chiu et al., 2005). HCoV-NL63 was first isolated in 2004 from a 7 month old infant in Holland (van der Hoek et al., 2004). The virus has been shown to have a worldwide distribution and is observed primarily in the winter season in temperate climates, although excep-tions have been reported (Chiu et al., 2005). To date, the virus has been associated with acute respiratory illness and croup in young children, the elderly and immunocompromised individuals (van der Hoek et al., 2005; Ebihara et al., 2005). Furthermore, HCoV-NL63 can also present as asthma exacerbation, febrile seizures and high fever (Chiu et al., 2005).

Sequence analysis of HCoV-NL63 identified a genome of 27 553 nucleotides with the typical corona-virus genome organization. An accessory gene present at nucleotide position 24542 - 25219 of the genome, ORF3, was identified between the spike (S) and enve-lope (E) genes. This manuscript reports the computa-tional analysis of HCoV-NL63 ORF3, which encodes a putative membrane-bound accessory protein. Amino acid similarities to other related group 1 coronaviruses will also be discussed.

RESULTS AND DISCUSSION

The group 1 coronaviruses can be divided into two genetic subgroups, that is 1a and b (Gorbalenya et al., 2004). Whereas members of group 1a include canine coronavirus, Feline infectious peritonitis virus (FIPV), transmissible gastroenteritis virus (TGEV), and ferret

*Corresponding author. E-mail: bfielding@uwc.ac.za. Tel: 27-21-959 3620. Fax: 27-21-959 1339.

Table 1. Comparison of HCoV-NL63 ORF3 amino acid sequence to homologues from selected coronavirus isolates.

	HCoV229E[b]	BtCoV[b]	PEDV[b]	TGEV[a]	Canine CoV[a]	FCoV[a]
HCoVNL63[b]	**43**	**34**	**34**	**25**	**25**	**24**
	62	*60*	*56*	*50*	*48*	*48*

Identity values (%) are shown in bold and similarity values (%) are shown in italics.
The putative amino acid sequence of HCoV-NL63 ORF3 was compared to sequences in the GenBank database at the National Centre for Biotechnology by using the Basic Blast Search Server (Altschul et al., 1990). Identified HCoV-NL63 ORF3-homologues were aligned with CLUSTAL X v 1.81 (Thompson et al., 1997). The sequences aligned, with Genbank accession numbers in bold, were: (HCoV-NL63), Human coronavirus NL63 Amsterdam 1 ORF3 (**YP_003768**); (HCoV-229E), Human coronavirus 229E CCU T935 ORF4 (**ABM64808**); (BtCoV), Bat coronavirus HKU8 ORF3 (**YP_001718613**); (PEDV), Porcine epidemic diarrhea virus DX ORF3 (**ABS72124**); (TGEV), Transmissible gastroenteritis virus BW021898B ORF3b (**AAF02712**); (Canine CoV), Canine coronavirus type I Elmo/02 ORF3c (**AAR88615**); and (FCoV), Feline coronavirus FCoV/NTU2/R/2003 ORF3c (**AAZ86080**). [a]Group 1a coronavirus; [b]Group 1b coronavirus.

enteric coronavirus, group 1b includes porcine epidemic diarrhea virus (PEDV), bat corona-virus, human coronavirus NL63 (HCoV-NL63) and human coronavirus 229E (HCoV-229E) (Dijkman et al. 2006). Interestingly, all sequenced group 1b corona-viruses contain only one accessory gene between the S and E genes (Dijkman et al., 2006), with ORF3 protein for PEDV and HCoV-NL63, and ORF4 protein for HCoV-229E. To date, very little is known about the structure and function of the ORF3-like proteins.

The HCoV-NL63 genome produces at least six distinct mRNAs, with all potential ORFs encoding for viral proteins. ORF3 has been shown to be expressed from distinct subgenomic (sg) mRNA 3, which is 3121 nucleotides in length (Pyrc et al., 2004). The ORF3 gene encodes for a putative ~25.6 kDa protein, 225 amino acids in length. Pyrc et al. (2004) reports that HCoV-NL63 ORF3 has a unique nucleotide com-position and appears as a U-rich and A-poor region within the genome, indicating a recent gene transfer event from another viral or cellular origin.

The ORF3 protein of HCoV-NL63 was found to be homologous to proteins of the other group 1 corona-viruses (Table 1). The amino acid sequence of HCoV-NL63 ORF3 was shown to be most similar to HCoV-229E ORF4 (43% identity; 62% similarity). Although the HCoV-229E ORF4 gene was initially reported to consist of ORF4a and ORF4b, subsequent studies have shown that this observation is a cell culture adaptive mutation (Duarte et al., 1994; Dijkman et al., 2006). The number-ing of the ORFs in HCoV-229E is based on Northern blot analysis of sg RNAs (Thiel et al., 2001). The presence of an additional sg mRNA in HCoV-229E-infected cells (that is sg mRNA3) shifts the numbering from ORF3 to ORF4.

HCoV-NL63 ORF3 was highly similar to Bat corona-virus ORF3-homologue (34% identity and 60% similarity) as well. Nothing is known of the BtCoV ORF3 though, but this high similarity raises interesting questions about the relationship between this human and bat coronavirus. Does this suggest a common bat ancestor or possible recombination event? However, a more complete genome-wide comparison is needed to answer these types of questions.

Porcine epidemic diarrhea virus (PEDV) (34% identity and 56% similarity to HCoV-NL63 ORF3) serially passaged in cell culture also has a mutation in ORF3 which results in virus attenuation in the natural host (Song et al., 2003; Woods, 2001). Large deletions in ORF3 of cell culture adapted PEDV has been reported (Park et al., 2007), which would disrupt trans-membrane regions 2 and 3 (Figure 1) effectively abolishing the function of ORF3. In fact, whereas piglets inoculated orally with wild-type PEDV became sick and developed severe diarrhea (Debouck and Pensaert, 1980), piglets inoculated with serially passaged PEDV showed reduced disease and lesions (Kweon et al., 1999). Similarly, cell culture adapted attenuated strain DR13 exhibited reduced pathogenicity and induced immunogenicity in pigs (Song et al., 2003). These changes were reported to be a result from adaptation and atenuation through serial passage in Vero cell culture (Kweon et al., 1999; Song et al., 2003).

The ORF3 gene of transmissible gastroenteritis virus (TGEV) also bears a significant identity and similarity to HCoV-NL63 ORF3 (25 and 50% respectively, Table 1). It has been reported that serial passage of TGEV results in a truncated form of the ORF3 protein, which in turn results in decreased pathogeniticity in the natural host (Woods, 2001). In fact, TGEV ORF3 gene has been detected with insertions and deletions, sug-gesting that this gene could be involved in cell tropism and pathogenecity (McGoldrick et al., 1999; Vaughan et al., 1995). Some investigators have suggested this area of the genome might be involved in tropism and patho-genicity of TGEV. Taken together, current data indicate that even though group I coronavirus ORF3-homologue genes are not essential for replication *in vitro*; they play a role in pathogenecity in the natural host.

Using TMHMM Server v. 2.0 (Krogh et al., 2001) to search for transmembrane regions in the ORF3 se-quence, we identified three such regions at amino acid position 39-61, 70-92 and 97-116. The comparative analysis shows regions ≥ 4 amino acids that are well conserved between the ORF3 homologues, that is
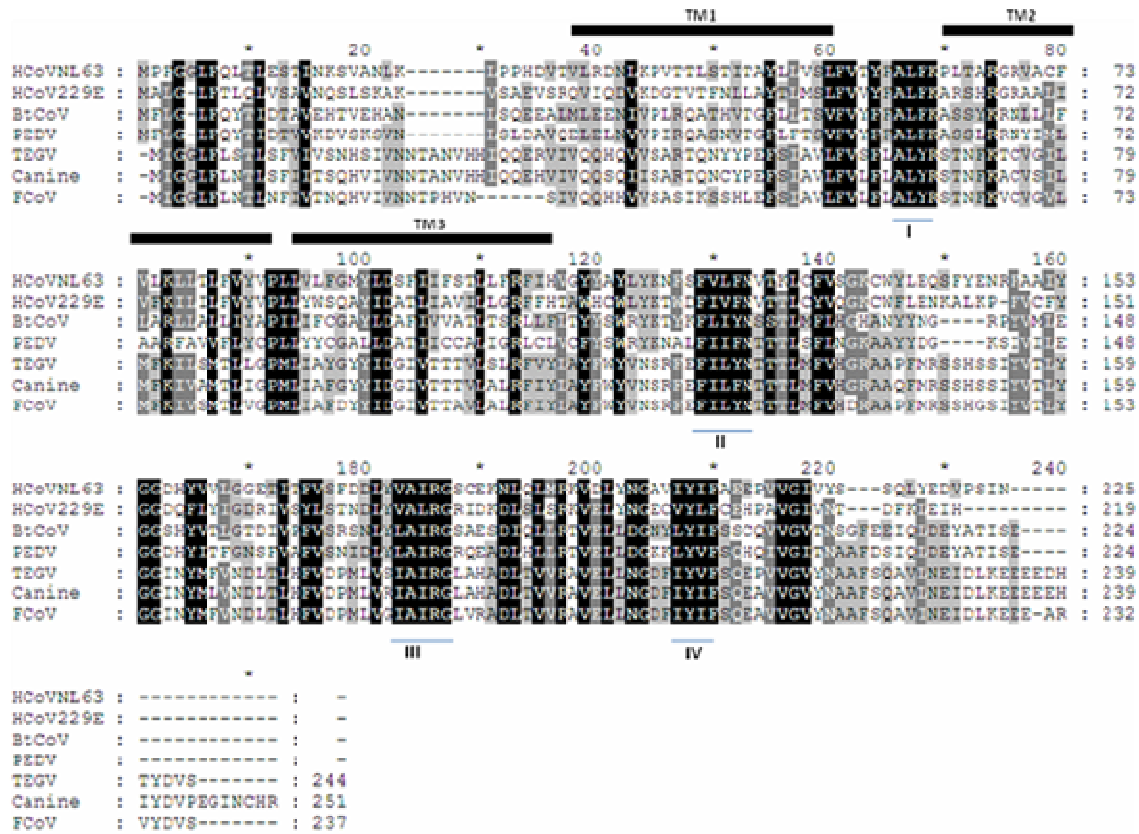
Figure 1. Predictive analysis of HCoV-NL63 ORF3. Alignment of the amino acid sequences of HCoV-NL63 ORF3 and selected coronavirus ORF3-homologues. Identified HCoV-NL63 ORF3-homologues were aligned with CLUSTAL X v 1.81 (Thompson et al., 1997) and viewed with GENEDOC version 2.6.002 software (Nicholas et al., 1997). Shading indicates conserved regions and gaps were introduced to align sequences. TMHMM Server v. 2.0 (Krogh et al., 2001) was used to search for transmembrane regions in the ORF3 sequence. Transmembrane (TM) regions are shown as black bars above the aligned sequences. Conserved regions identified: I:(ALF/YK/R), II:(FV/I/LL/I/VF/YN), III:(V/L/IAI/LRG) and IV:(I/L/VYI/L/VF).

(ALF/YK/R), (FV/I/LL/I/VF/YN), (V/L/IAI/LRG) and (I/L/VYI/L/VF). The functions, if any, of these regions are not known, but could warrant further research (Figure 1). HCoV-NL63 ORF3 homologues are found in both group 1a and b coronaviruses. Based on our preliminary *in silico* findings and comparative analysis, it appears as though ORF3 of HCoV-NL63 may also play a role in the infectivity and/or pathogenesis of the virus. It would be interesting to see if mutations in HCoV-NL63 ORF3 would also result in attenuation of this virus and whether this could lead to the development of a vaccine. However, the function of ORF3 protein in HCoV-NL63 is still not known. To our knowledge, no molecular analysis of ORF3 has been reported. In this study we have identified coronavirus HCoV-NL63 ORF3-homologues and used bioinformatic tools to compare these proteins.

ACKNOWLEDGEMENTS

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Chiu SS, Chan KH, Chu KW, Kwan SW, Guan Y, Poon LLM, Peiris JSM (2005). Human Coronavirus NL63 Infection and Other Corona-virus Infections in Children Hospitalized with Acute Respiratory Disease in Hong Kong, China CID 2005: p. 40.

Debouck P, Pensaert M (1980). Experimental infection of pigs with a new porcine enteric coronavirus, CV 777, Am. J. Vet. Res. 41(2): 219-23.

de Haan CA, Masters PS, Shen X, Weiss S, Rottier PJ (2002). The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. Virol. 296: 177-189.

Dijkman R, Jebbink MF, Wilbrink B, Pyrc K, Zaaijer HL, Minor PD, Franklin S, Berkhout B, Thiel V, van der Hoek L (2006). Human coronavirus 229E encodes a single ORF4 protein between the spike and the envelope genes. Virol J. 3: p. 106.

Duarte M, Tobler K, Bridgen A, Rasschaert D, Ackermann M, Laude H (1994). Sequence analysis of the porcine epidemic diarrhea virus genome between the nucleocapsid and spike protein genes reveals a polymorphic ORF. Virol. 198: 466-476.

Ebihara T, Endo R, Ma X, Ishiguro N, Kikuta H (2005). Lack of association between New Haven coronavirus and Kawasaki disease. J. Infect. Dis. 192: 351-352.

Gorbalenya AE, Snijder EJ, Spaan WJ (2004). Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. J. Virol. 78: 7863-7866.

Haijema BJ, Volders H, Rottier PJ (2004). Live, attenuated corona-virus vaccines through the directed deletion of group-specific genes provide protection against feline infectious peritonitis. J. Virol. 78: 3863-3871.

Holmes KV Lai MMC (1996). Coronaviridae: The viruses and their replication. In: Fields BN, Knipe DM, Howley PM (Editors). Fields Virology. Third. Philadelphia: Lippincott-Raven Publishers. pp. 1075–1093.

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305(3): 567-580.

Kweon CH, Kwon BJ, Lee JG, Kwon GO, Kang YB (1999). Derivation of attenuated porcine epidemic diarrhea virus (PEDV) as vaccine candidate. Vaccine, (20-21): 2546-2553. Erratum in: Vaccine 1999 Aug 20, 18(1-2): p. 201.

McGoldrick A, Lowings JP, Paton DJ (1999). Characterisation of a recent virulent transmissible gastroenteritis virus from Britain with a deleted ORF 3a. Arch Virol. 144(4): 763-770.

Nicholas KB, Nicholas Jr HB, Deerfield DW (1997). GeneDoc: analysis and visualization of genetic variation. EMBNEW News, 4: 14.

Ontiveros E, Kuo L, Masters PS, Perlman S (2001). Inactivation of expression of gene 4 of mouse hepatitis virus strain JHM does not affect virulence in the murine CNS. Virol. 289: p. 230.

Ortego J, Escors D, Laude H, Enjuanes L (2002). Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome. J. Virol. 76: 11518-11529.

Park SJ, Moon HJ, Luo Y, Kim HK, Kim EM, Yang JS, Song DS, Kang BK, Lee CS, Park BK (2008). Cloning and further sequence analysis of the ORF3 gene of wild- and attenuated-type porcine epidemic diarrhea viruses. Virus Genes, 36(1): 95-104.

Pyrc K, Jebbink MF, Berkhout B, van der Hoek L (2004). Genome structure and transcriptional regulation of human coronavirus NL63. Virol. J. 1: 7.

Song DS, Yang JS, Oh JS, Han JH, Park BK (2003). Differentiation of a Vero cell adapted porcine epidemic diarrhea virus from Korean field strains by restriction fragment length polymorphism analysis of ORF 3. Vaccine, 21:1833-1842.

Thiel V, Herold J, Schelle B, Siddell SG (2001). Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J. Gen. Virol. 82: 1273-1281.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.

van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B (2004). Identification of a new human coronavirus. Nat. Med. 10(4): 368-73.

van der Hoek L, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Uˇberla K (2005). Croup Is Asso-ciated with the Novel Coronavirus NL63. PLoS Med. 2(8): p. 764.

Vaughan EM, Halbur PG, Paul PS (1995). Sequence comparison of porcine respiratory coronavirus isolates reveals heterogeneity in the S, 3, and 3-1 genes. J. Virol. 69(5): 3176-3184.

Woods RD (2001). Efficacy of a transmissible gastroenteritis coronavirus with an altered ORF-3 gene. Can. J. Vet. Res. 65: 28-32.