African Journal of Biotechnology Vol. 11(75), pp. 13962-13968, 18 September, 2012 Available online at http://www.academicjournals.org/AJB DOI: 10.5897/AJB12.1657 ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Characterisation of human coronavirus-NL63 nucleocapsid protein

Michael Berry¹, Taryn-Lee Manasse¹, Yee-Joo Tan² and Burtram C. Fielding¹*

¹Molecular Biology and Virology Laboratory, Department of Medical BioSciences, Faculty of Natural Sciences, University of the Western Cape, Western Cape, South Africa.

²Department of Microbiology, MD4, 5 Science Drive 2, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597.

Accepted 1 August, 2012

Coronavirus N is a multifunctional protein that plays an essential role in enhancing the efficiency of virus transcription and assembly. This manuscript reports the analysis of HCoV-NL63 N protein by comparing the amino acid sequences of coronavirus N-homologues. A ~50 kDa protein was expressed in both a mammalian cell and bacterial cell system that is similar in size to the predicted ~42.6 kDa HCoV-NL63 N protein. PSORTII identified two putative nuclear localisations signals and PONDR identified one disordered region in HCoV-NL63 N. The reported protein analysis serves as a prelude to laboratory analysis to understand the processing of HCoV-NL63 N.

Key words: HCoV-NL63, nucleocapsid protein, protein analysis, nuclear localisation signals, protein ordered/disordered regions.

INTRODUCTION

Genome sequencing technologies have resulted in large genome databases and the rate of genome sequencing has surpassed our ability to determine the function of the proteins these genomes encode. In silico analysis provides tools to predict possible functions of these proteins by comparing novel proteins of unknown function with proteins of known function, or by analyzing the unknowns using an algorithm based on information from proteins of known function. Thus, the theoretical analysis of proteins provide a useful framework that has given researchers the ability to predict the function of these proteins, which can then be verified with the use of in vitro and in vivo studies (MacKenzie et al., 2008; Mahotka et al., 2002). Coronaviruses are enveloped viruses of the Coronaviridae family in the order Nidovirales. Recently, the International Committee for Taxonomy of Viruses (ICTV) reported that the three traditional coronavirus groups are replaced by four genera, namely Alphacoronaviruses, Betacoronaviruses,

Gammacoronaviruses and Deltacoronaviruses (http:// talk.ictvonline.org/files/ictv_documents/m/msl/4090.aspx). Coronaviruses have single-stranded, positive sense RNA genomes, ranging in size from 25 to 30 kbs. The genomes typically encode for four structural proteins, spike (S), envelope (E), membrane (M) and nucleocapsid (N). The coronavirus N protein is one of the most abundantly produced viral proteins within an infected cell. This protein has been shown to have several functions, includina binding to viral RNA to form the ribonucleocapsid and is proposed to have roles in replication, transcription and translation. To date, coronavirus N proteins have been shown to cause deregulation of the host cell-cycle (Surjit et al., 2006; Wurm et al., 2001), inhibit the production of interferon (Kopecky-Bromberg et al., 2007), up-regulate the production of COX2 (Yan et al., 2006), up-regulate AP1 activity (He et al., 2003), induce apoptosis in serum starved COS1-cells (Surjit et al., 2004a) and interact with various cellular proteins including cyclophilin A (Luo et al., 2004) and hnRNP-A1 (Luo et al., 2005; Wang and Zhang, 1999), which may have possible pathological consequences.

Even though several studies have shown that the N-

^{*}Corresponding author. E-mail: bfielding@uwc.ac.za. Tel: 27-21-959 3620. Fax: 27-21-959 3125.

protein of members of Nidovirales localizes to the nucleus and/or nucleolus (Cawood et al., 2007; Hiscox et al., 2001; Timani et al., 2005; Wurm et al., 2001), others have shown that N localizes predominantly or exclusively to the cytoplasm (Rowland et al., 2005; Stohlman et al., 1983; You et al., 2005). This indicates that the localization of N to the nucleus or nucleolus is not a conserved property of nidoviruses (Rowland et al., 2005). The N-protein of coronaviruses is highly antigenic and reliably induces an antibody response (Mourez et al., 2007) and has therefore been recognized as an important target in the development of coronavirus diagnostics (Blanchard et al., 2011; Lee et al., 2008).

HCoV-NL63 was first isolated from the nasopharyngeal aspirate of a 7 month old infant from Holland (van der Hoek et al., 2004). The virus has since been shown to have a global distribution and is observed primarily in the winter season in temperate climates. To date, the virus has been associated with acute respiratory illness and young croup in children, the elderly and immunocompromised individuals (Oosterhof et al., 2010; Sung et al., 2010; van der Hoek et al., 2005, 2006). Clinical manifestations most frequently associated with HCoV-NL63 infections include cough, coryza, fever, sore throat, bronchiolitis, bronchitis, croup and pneumonia (Fielding, 2011; van der Hoek, 2007). To date, three groups have reported a fatal outcome following HCoV-NL63 infection, indicating that this virus could be involved in more severe respiratory tract disease (Bastien et al., 2005; Cabeca and Bellei, 2012; Oosterhof et al., 2010); all of these cases had underlying suppressed or compromised immune systems.

Unfortunately, the virus is not well studied at a molecular level with only the spike (Lin et al., 2008, 2011; Mathewson et al., 2008) and ORF3 proteins (Fielding and Suliman, 2009; Muller et al., 2010) previously characterized. In this work, the HCoV-NL63 N was expressed in mammalian and bacterial systems. Various bioinformatics tools were used to analyze the HCoV-NL63 N amino acid sequence. The importance of identified putative nuclear localization signals and disordered motifs were discussed. This manuscript reports the analysis of HCoV-NL63 N as a prelude to the molecular characterization of N in *in vitro* mammalian and bacterial cell systems.

RESULTS AND DISCUSSION

The HCoV-NL63 genome is approximately 27 553 nucleotides in size, with the typical coronavirus genome organization. It produces six separate mRNAs, with all potential ORFs encoding for viral proteins. N is expressed from distinct subgenomic (sg) mRNA 6, the most abundant sg mRNA (Pyrc et al., 2004). To determine the size of HCoV-NL63 N, recombinant GST-N was expressed in a bacterial system, purified by

affinity column and treated with TEV to remove the GSTtag, according to the manufacturer's specification (Promega). Also, recombinant HCoV-NL63 N-HA was transfected and expressed in Cos-7 cells (Khan et al., 2006). SDS-PAGE analysis of purified bacterial expressed protein (Figure 1A) and Western Blot analysis of total mammalian cell lysate (Figure 1B) showed ~50 kDa proteins which is larger than the size of the 42.6 kDa protein predicted by our initial analysis (http://www.sciencegateway.org/tools/ proteinmw.htm). This discrepancy in size showed that the protein was probably pre-, co- or post-translationally modified; the type of modification is currently being investigated.

Next, the amino acid sequence of HCoV-NL63 N was compared to other coronavirus N proteins (Table 1). HCoV-NL63 was shown to be most similar to the human coronavirus 229E N (43% identity; 63% similarity). This was not totally unexpected, since previous molecular clock analysis estimate that HCoV-NL63 and HCoV-229E evolved from a most recent common ancestor in the 11th century (Pyrc et al., 2006). Interestingly, HCoV-NL63 N also showed high amino acid similarity to the bat coronavirus HKU7 N (40% identity; 57% similarity). As speculated previously (Fielding and Suliman, 2009), this high amino acid similarity raises interesting questions about the relationship between this human and bat coronaviruses. Woo and colleagues postulate that the ancestor of all present-day coronaviruses infected a bat and then jumped from that bat to a bird, or alternatively, first infected a bird and then jumped to a bat, evolving dichotomously. This bat coronavirus then jumped to another bat species, which in turn jumped to other bat species or other mammals, including humans. Each of these interspecies jumps resulted in the independent evolution of these coronaviruses and gave rise to the current coronavirus genera. In fact, present evidence suggests that the huge diversity of coronaviruses in bats make them excellent gene pools for group 1 and group 2 coronaviruses (Woo et al., 2009).

Comparative analysis using CLUSTAL X version 2.0 (Larkin et al., 2007) of 10 coronavirus species identified a amino acid region conserved between the N 9 that is FYY(L/T)GTGP 2). homologues, (Figure Conserved regions within a family of proteins is commonly indicative of structural and functional conservation, as well as an evolutionary relationship. The function of this sequence, if any, is yet to be elucidated. Since interactions between coronavirus RNA and N are assumed to predominantly take place at the nucleocapsid N-terminal domain, the possible role of this conserved motif in RNA-N interaction will be studied. However, several other binding sites within N have also previously been identified, indicating that both N- and C-terminal domains are probably involved in RNA binding (Chang et al., 2009).

Translocation of proteins from the cytoplasm through the nuclear pore complex and into the nucleoplasm is



Figure 1. Expression of HCoV-NL63 N in: (A) a bacterial cell system. Cleavage of the GST-tag from SARS-CoV and HCoV-NL63 full length N using ProTEV protease (Promega). SARS-CoV N and NL63 N were reverse transcribed from viral RNA and PCR amplified using nucleocapsid specific primers. The N-gene was cloned into the pFlexi[™] protein expression vector (Promega), which appends an N-terminal GST affinity tag. Protein expression was induced by addition of 0.1% Rhamnose (Promega). Lane 1: Pre-stained molecular weight protein marker; Lane 2: SARS-CoV GST-N; Lane 3: SARS-CoV N with GST cleaved; Lane 4: HCoV-NL63 GST-N; Lane 5: HCoV-NL63 N with GST cleaved. SARS-CoV N was included as a control. (B) a mammalian cell system. Recombinant plasmid pCAGGS-N-HA (HCoV-NL63 N gene cloned into vector pCAGGS which appends an N-terminal Haemagglutinin tag to expressed protein) was transfected with Lipofectamine 2000 (Invitrogen) into a COS-7 cell line (African green monkey cells) and grown at 37°C in 5% CO₂ for 16 h. Cells were then lysed and the target protein detected on a Western Blot using rabbit anti-HA polyclonal (1:250) as primary antibody and goat anti-Rabbit polyclonal secondary antibody (1:1000).

initiated by nuclear localization signals (NLSs) on the protein binding to a NLS-binding site on importin- α of the host cell (Rowland et al., 2005). NLSs are rich in basic amino acids and conform to one of three types, that is, (i) monopartite pat4, (ii) monopartite pat7 or (iii) bipartitemotifs. Pat4 NLSs are recognized by the

presence of 4 basic residues - either K or R or the presence of 3 basic residues (K/R) and/or (H/P). Pat7 NLSs are segments beginning with P and followed within 3 residues by a basic segment containing 3 K or R residues out of 4. On the other hand, bipartite NLSs begin with 2 basic residues, followed by a 10 residue

Table 1. Comparison of HCoV-NL63 N amino acid sequence to homologues from selected coronavirus isolates. Identity values (%) are shown in BOLD and similarity values (%) are shown in *ITALICS*.

Coronavirus	SARS	MHV	IBV	OC43	229E	TGEV	PEDV	HKU1	HKU7
NIL CO	21	18	17	18	43	32	35	20	40
INL03	34	31	31	32	63	51	50	34	57

To identify homologues, the amino acid sequence of HCoV-NL63 N 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). Subsequently, HCoV-NL63 N was aligned with selected coronavirus N proteins with CLUSTAL X version 2 (Larkin et al., 2007). The sequences aligned, with Genbank accession numbers in bold, were: HCoV-NL63: NL63 (DQ846901.1); SARS-CoV: SARS (AY360146.1); Mouse Hepatitis Virus: MHV (AAU06361); Infectious Bronchitis Virus: IBV (AAA46214), Human coronavirus OC43: HCoV-OC43 (AY585229.1); HCoV-229e: 229E (AAG48597); Transmissible gastroenteritis virus: TGEV (AAA47915); Porcine epidemic diarrhea virus:PEDV (AF353511); Human coronavirus HKU1: HKU1 (AAT98585) and Bat coronavirus HKU7: HKU7 (DQ666343).

		*		20	*	40	*	60	*	80	*	100	
NL63			MAN.		VNVADDR.			AARKKEPPP	EYMPTI	VSSDKAPYRVI	PRNLVEIGKGN	K.DP	: 51
229E			NAT		VEDALASE	PO		RGROGRTPY	TYSPIT	VDSE OPWKUT	PRNTWETNER	K NK	. 53
2230	:		MCC		USTOCDC			DDDCCDUDE	NO A D T T	AUCNUDEMATN	DONGVENCKET		. 51
пкот	•		.1155.	•••••	VS QGDG.			FRRJGRVFF	BIAPLI	AIGNIFEWALM	PONGVENGRGL	R.N.	
PEDV	:	• • • • • • <mark>• •</mark> • • • • •	.MAS.		VSEQDRG.			RKRVPL	LYAPLE	RVTNDKPLSKVI	ANNAVETNKGN	JK.DQ	: 48
TGEV	:		.MANQ	G	QRVSNGDEST	KTRGRSNSR		GRKNNNIPL	FFNFIT	LQQGSKFWNLC	PRDFVFKGIGN	IR.DQ	: 65
MHV	:	MSEVPGQENAGO	RSSSVI	NRAGNGIL	KKTTNADQTE	RGPNNQNRGRR	NQPKQTATT	Q. PNSGSVVPHY	WESGIT	COFOKGKEFOFA	EGOGVEIANGI	FASP	: 99
0C43	:	MSFTPGKOSSS.	RASSGI	RSGNGIL	KWADOSD	OFRNVOTRGRR	ACPKOTATS	COPSGGNVVPYY	WESGI	OFOKGKEFEFA	EGOGVEIAPGV	FATP	: 96
HKI11		MSYTPG HYAG	RSSSG	NRSG TU	KKTSUADOSEI	RNYOTENRORK	TOPKETVST	O POGNTTPHY	RESCT	OFORGEDERES	DECEVETAFOU	PPSD	. 95
CADS Coll		norreo.mmoc	MODNICI	DOCNODEN	DDIMACCOMD	CEDNNONCCDN	C ADDV	ODDDOCT DNNEA		QUCK PETDEL	DCCCWETNENC	CDDD	. 00
SARS-COV	•		MSDNG	PUSNUKSA	PRII GGPID.	STUNNQNGGRN	IGARPA	QRRPQGLPNNIA	IAL IALI	QUCK. EELKEF	RGQGVEINING	GPUD	. 03
TRA	•	· · · · · · · · · · · · · · ·	.MASG		KAAGKTL	AFAPVIKLGGP	K	PPKVGSSGNA	WEÇALE	AKKLNTPPPKE	EGSGVEDNENI	KPS	: 00
		*		120	*	140	*	160	*	180	*	200	
NT.63		OTCVENUOER	DMDD	OPUD PD	WHEYYT CTC	HADIK PODS	DOWNARE	CART UNUSICN	KENOKI	TERESTATES	FLOWFFEDDO	NNSS	· 148
2205		T TOWNNUCTOR	TOTOT	VDUDUCD	TUEVVICEC		E CULUMUATIO	CHUT DD. CUCU	DVNODI	DINIUTNOW	CUT UPP DD	G	. 145
2296	•	LIGIWNVQKR	FRIKK	KRVD SP	ALHEIILGIG.	HKLAKTKERV	BGVVWVAVD	GERT. EPIGIGV	RKNSEI	PIEHENŐKTEN	GVIVVEEPD		: 145
HKU7	:	QIGYWNEQQR	WRTRK	NKETPS	KWHFYYLGTG.	HAEEPPRKRT	QGV FWVAKE	GAKT.EPHGIGTI	KKSESI	LIVERFSFRLEE	GLEUSEFASN.	.FAS	: 146
PEDV	:	QIGYWNEQIR	WRMRR	ERIEQPS	NWHFYYLGTG	HGDLRYRTRT	EGVEWVAKE	GAKT.EPINIGVI	KASEKI	PIIEKFSQQLES	SVVEIVEPNTP.	.FAS	: 143
TGEV	:	CIGYWNRCTR	YRMVK	ORKET PE	RWEFYYLGTG	HALAKKOKL	DGVVWVAKD	CAMN.KPUTUGS	GANNES	KALKFDGKVEG	SEFOI EVNOSR.	D	: 158
MHV		OKGYNYRHNRRS	EKTPD	COKOLLP	RWYFYYLGTG	HAG ASGDST	EGVEWVANS	CADTNTRSDTVE	DESSHE	TE TREAT	GTVI POGEYVE	G	. 192
0043	:	AKCVEVDHNDDO	TAD	NOPOLLP	DEVEVUI CTC	HAKDONGTOT	DOWNHUASN	CADUNTEADIVDI	DDSSDE		CTVI DOCVVTE	G	. 100
0043		ANGINIKINKK	TAL	NOR OT LE		INARDO GIDI	DOVINVESI	CODVINTERD VD	DESSDI		GIVHFQGIIII		. 100
HKUI		AKGIWIRHSRRS	ERTAD	ÇŐKŐLT H	RWMFYYLGTG.	TANASIGESL	BGVEWVANH	QADISTESUVSS	UPTTQE	ALSTREEF	GITT POGITVE	5G	: 189
SARS-COV	:	QIGYYRRATRR.	VRGGD	KMKELSP	RWYFYYLGTG	EASLPYCANK	EGIVWVATE	GALNTPKDHUGT	NPNNNA	ATVLQL	GTTEPKGFYAE	EGS	: 177
IBV	:	QHGYWRRCAR	EKPGK	GRKPVPL	AWYFYYTGTG	AADLNNGDTQ	DGIVWVAAK	CADTKSRSNQGT	DPDKFI	QYELRFSD	GGPDGNFRWDF	IPLN	: 161
						-							
					i								
		*		220	*	240	*	260	*	290	*	200	
		*		220	i *	240	*	260	*	280	*	300	0.05
NL63	:	* RAS <mark>S</mark> RSSTRNNS		220 5	i *	240	*	260 TSRQQSRTRSDSM	* 1	280 SSSDIVAAVTI	* Alknigfdn.ç	300 SKSP	: 205
NL63 229E	::	* RASSRSSTRNNS RAPSRSQSRSQS	RDSSR RGRGE	220 5	ا *	240	*	260 TSRQQSRTRSDSN KPQSRNPSSDRNF	* 11	280 QSSSDIVAAVTI NSQDDIMKAVAA	* ATKNIGFDN.Ç ATKSIGFDKPÇ	300 QSKSP QEKDK	: 205 : 203
NL63 229E HKU7		* RASSRSSTRNN RAFSRSQSRSQS RSSSCAFSRQQS	RDSSR RGRGE	220 5 5 TPTKQP	۱ *	240	*	260 TSRÇQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS	* 11 3KQKQSQ	280 SSSDIVAAVTI ISQDDIMKAVAA PQDDLVAAVRA	* ALKNIGFDN.Ç ALKSIGFDKPÇ ALLGYGISPQS	300 QSKSP QEKDK SSGGS	: 205 : 203 : 214
NL63 229E HKU7 PEDV		* RASSRSSTRNN RAPSRSQSRSQ RSSSCAPSRQQ RANSRSRSRGNG	RDSSR RGRGE KSRDP	220 5 5 FPTKQP SPSNNRGN	i *	240	*	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNOSNNRNOSM	* 1) 5KQKQS(1DRGGV1	280 2SSSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA SSDDIVAAVKI	* ATKNIGFDN.Ç ATKSIGFDKPÇ ATLGMGISPQS ATKSIGIGENE	300 QSKSP QEKDK SSGGS PDRHK	: 205 : 203 : 214 : 243
NL63 229E HKU7 PEDV TGEV		* RASSRSSTRNN RAPSRSQSRSQS RSSSCAFSRQQS RANSRSRSRGNC NSDSPSC5750	RDSSR RGRGE KSRDP NNRSR	220 5 5 FPTKQP SPSNNRGN	i * NQSRGNSQNR(240 GNNQGRGASQN	* 	260 TSRQQSRTRSDS1 KPQSRNPSSDRNI KSRDQSQTRKGQS KSRNQSNNRNQS1 BGPQOENNKKDDS	* 1N SKQKQS(NDRGGV1	280 QSSSDI VAAVTI NSQDDI MKAVAA QPQDDI VAAVRA ISRDDI VAAVKE VECAVLA	* ATKNIGFDN.Ç ATKSIGFDKPÇ ATLGVGISPQS ATKSIGIGENE ATKSIGIGENE	300 QSKSP QEKDK SSGGS PDRHK	: 205 : 203 : 214 : 243 : 211
NL63 229E HKU7 PEDV TGEV		* RASSRSSTENN RAFSRSCSRSC RSSSCAPSROC RANSRSRSFGNG NSRSRSCSRSS	RDSSR GRGE KSRDF NNRSR RNRSQ	220 5 5 5 5 5 5 5	i * NQSRGNSQNR(240 GNNQGRGASQN	*	260 TSRQQSRTRSDS1 KPQSRNPSSDRNI KSRDQSQTRKGQS KSRNQSNNRNQS1 RGRQQFNNKKDDS	* 1N SKQKQS(NDRGGV1	280 QSSSDI VAAVTI NSQDDI MKAVAA QPQDDI VAAVRA TSRDDI VAAVRI SRDDI VAAVRI 	* AD KNIGFDN.Ç AD KSIGFDKPÇ AD LGMGISPQS AD KSIGIGENE AD KKIGVDTEK	300 QSKSP QEKDK SSGGS PDRHK KQQQR	: 205 : 203 : 214 : 243 : 211
NL63 229E HKU7 PEDV TGEV MHV		* RAFSRSSTRNS RAFSRSSS RSSSCAFSRQC RANSRSRSFGNG NSRSRSSS SGRSAFASRSSS	RDSSR RGRGE KSRDF NNRSR RNRSQ RSQSR	220 5 PTKQP SPSNNRGN 5	i * NQSRGNSQNR(240 GNNQGRCASQN	* RGGNNNNN	260 TSRQQSRTRSDSM KPQSRNPSSDRNM KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS GPNNRARSSS(* H SKQKQSQ NDRGGV1 S Q	280 QSSSDI VAAVTI NSQDDI MKAVAA QPQDDI VAAVRA TSRDDI VAAVKI VEÇAVIA RFASTVKPDMAE	* ATKNIGEDN.Q ATKSIGEDKPQ ATLGYGISPQS ATKSIGIGENE ATKKIGVDTEK ETATVIAKLG	300 QSKSP QEKDK SSGGS PDRHK KQQQR GKLAG	: 205 : 203 : 214 : 243 : 211 : 247
NL63 229E HKU7 PEDV TGEV MHV OC43		* RASSRSSTENN RAPSRSCSFSC RANSRSRSFGN NSRSRSCSFSR SGRSAFASFSG SGRSAFNSRST	RDSSR RGRGE KSRDF NNRSR RNRSQ: RSQSR RSQSR	220 5 PTKQP SPSNNRGN 5	i * NQSRGNSQNR(240 GNNQGRGASQN	* IRGGNNNNNN	260 TSRQQSRTRSDSM KPQSRNPSSDRM KSRQQSQTRKGQS KSRNQSNNRNQSM RGRQQFNNKKDDS GPNNRARSSS SAGSRSRANSGM	* H SKQKQSQ NDRGGV1 S 2 R	280 2SSSDIVAAVTI NSCDDIMKAVAA PPCDIVAAVRA TSRDDIVAAVRA TSRDDIVAAVRA RFASTVKPDVAE PPTSGVTPDVAD	* AIKS GFDN.Ç AIKS GFDKPÇ AILGVGTSPQS AIKS GIGENF AIKS GVDTEK BIAA VIAKLG QIAS VIAKLG	300 SKSP EKDK SSGGS DRHK KQQQR SKLAG SKLAT	: 205 : 203 : 214 : 243 : 211 : 247 : 249
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1		* RASSRSSTENNS RAPSRSCSFSC RASSCAFSRCC RANSRSRSFGNC NSRSRSCSFSR SGRSAFASTSG SGRSAFNSFST SGRSAFNSFST SGRSASNSFFG	RDSSR RGRGE KSRDP NNRSR RNRSC RSQSR RTSSR RSQSR	220 5 FPTKQP 5.PSNNRGNI 5 5 5 5	i * NQSRGNSQNR(240 GNNQGRGASQN	* IRGGNNNNNN	260 TSRQQSRTRSDSN KPQSRNPSSDRNI KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSSQ SAGSRSRANSGNF NNRSLSRSNSNFF	* 1 SKQKQSQ NDRGGV1 5 2 2 3	280 2SSSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVKU VEÇAVLA RFASTVKPDVAE PFTSGVTPDVAL ISDSIVKPDVAE	* AIKNIGFDN.Ç AIKSIGFDRPÇ AIKSIGIGPQS AIKSIGIGTSPQS AIKSIGIQTSP AIKSIGIDTSP QIASIVIAKLG QIASIVIAKLG DIANIVIAKLG	300 SKSP EKDK SSGGS DRHK KQQQR SKLAG SKLAT SKDS.	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-COV		* RAFSRSSTRNN RAFSRSSS RSSSCAFSFQC RANSRSSSS RAFSRSSS SGRSAFASFSS SGRSAFNSFST SGRSAFNSFFS SGRSAFNSFFS RGGSCASFSS	RDSSR RGRGE KSRDP NNRSR RNRSC RSQSR RSQSR RSQSR RSQSR RSQSR	220 5 FPTKQP SPSNNRGNI 5 AS 5	I *	240 GNNQGRGASQN	* RGGNNNNNN	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF RNSTPGSSRG	* 1 SKQKQSQ NDRGGV1 5 2 2 3 3 4	280 SSSDIVAAVTI SQDDIMKAVAA SRDDIVAAVKI VEÇAVIA RFASTVKPDVAE ISDSIVKPDVAE ISDSIVKPDVAE	* AIKN GFDN.Q AIKS GFDKPQ AIKS GIGENF AIKS GIGENF AIKK GVDTEK EIAA VIAKLG DEIAN VIAKLG AIAI VIAKLG AIAI LLDRLN	300 QSKSP QEKDK SSGGS DRHK (QQQR GKLAG GKLAT GKDS. NQLES	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 247 : 233
NL63 229E HKU7 PEDV TGEV MHV oc43 HKU1 SARS-CoV TBV		* RASSRSSTENN RAPERSOSFSO RSSSCAFSFQC RANSRSSSSFGM SGRSAFASRSCS SGRSAFASRSCS SGRSAFASRSCS SGRSAFASRSTS SGRSASNSFFG RGGSCASSRSSS	RDSSR RGRGE NNRSR RNRSQ RSQSR RSQSR RSQSR RSQSR RSR RSR	220 SS PPTKQP SPSNNRGNI SS SS SS SS	I *	240 GNNQGRCASQN	* RGGNNNNN	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF RNSTPGSSRG	* HP SKQKQSQ NDRGGV1 5P 2P	280 2SSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA TSRDDIVAAVKI VEQAVIA RFASTVKPDVAE 1SDSIVKPDVAE SFARMASGGET GDDIIABAAK	* AIKN GFDN. Q AIKS GFDKPQ AILGYGISPQS AIKS GIGENE AIKK GVDTEK EIAA VIAKLG QIAS VIAKLG DIAN VIAKLG AIAL LLDRLN	300 2SKSP 2EKDK SSGGS 2DRHK (QQQR 5KLAG 5KLAT 5KDS. NQLES	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV		* RASSRSSTENN RAPSRSCSFSC RANSRSRSCSFSC NSRSRSCSFSR SGRSAFASFSG SGRSAFASFSG SGRSAFNSFSG RGGSCASSS RGRSGRSTAAS	R DSSR R GRGE K SRDP NNRSR R NRSQ R SQSR R SQSR R SQSR R SQSR AAASR	220 S PPTKQP SPSNNRGNI S SP SP S S S	*	240 GNNQGRGASQN	* RGGNNNNNN	260 TSRQQSRTRSDSN KPQSRNPSSDRNI KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS . GPNNRARSSS(SAGSRSRANSGNF NNRSLSRSNSNFF . RNSTPGSSRG . SREGSRGRRSDS	* H SKQKQSQ NDRGGV1 S 2 2 2 2 3 3	280 SSSDIVAAVTI SSCDIMKAVAA PPCDIVAAVRA TSRDDIVAAVRI AFASTVKPDVAE SSSIVKPDVAE SSSIVKPDVAE SSARMASGGGET .GDDIIARAAK	* AIKS GFDN.Q AIKS GFDKPQ AILGNGISPQS AIKS GIGENF AIKS GVDTEK DIAN VIAKLG DIAN VIAKLG DIAN VIAKLG AIAL LLDRLN VIAQ	300 2SKSP 2EKDK SSGGS 2DRHK (QQQR 5KLAG 5KLAT 5KDS. IQLES	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV		* RASSRSSTRNN RAFSRSSCR RSSSCAFSTQC RANGRSRSSCSR SGRSAFASSS SGRSAFASSS SGRSAFNSRST SGRSASNSRFS RGGSCASSSS RGGSGRSTAAS	R DS SR GRGE SRDP NNR SR R NR SQ SQ SR SQ SR SQ SR SRGN AAA SP	220 SSPTKQP SPSNNRGNI SS SPS SPS	*	240 GNNQGRGASQN	*	260 TSRQQSRTRSDSN KPQSRNPSSDRNH KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNH NNRSLSRSNSNFH .RNSTPGSSRG .SREGSRGRRSDS	* H BKQKQSQ NDRGGVI 5 2 R R S	280 QSSSDIVAAVTI NSQDDIMKAVAA PCDDIVAAVRA SRDDIVAAVKI VEÇAVIA FASTVKPDVAL ISDSIVKPDVAL ISDSIVKPDVAL SFARMASGGET GDDIIARAAK	* AIKN GFDN.Q AIKSI GFDKPQ AILGVGISPQS AIKSI GIGENF AIKKI GVDTEM EIAAI VIAKLG QIASI VIAKLG DIANI VIAKLG AIALI LLDRLN	300 2SKSP 2EKDK SSGGS 2DRHK KQQQR SKLAT SKDS. IQLES	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV		* RASSRSSTRN RAFRSSCAFSSC RSSSCAFSSCC RANSRSRSSGS SGRSAFASSSS SGRSAFASSSS SGRSASNSFPG RGGSCASSSSS RGRSGRSTAASS	R DS SR GRGE SRDP NNR SR R NR SQ SQ SR SQ SR SQ SR SRGN AAA SR	220 5 5 5.PSNNRGNN 5 5. 5. 5. 5. 5. 5. 5.	*	240 GNNQGRCASQN	*	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG SREGSRGRRSDS	* HS SKQKQSG NDRGV1 5F 2F RF NS	280 2SSSDTVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA SSRDDIVAAVKI ARASTVKPDVAE PTSGVTPDVAL SSBIVKPDVAE SFARMASGGET .GDDIIARAAK	* ATKN GFDN.Q ATKS GFDKPQ ATKS GIENE ATKS GIENE ATKS GVDTEK ETAT VIAKLG QTAS VIAKLG DETAN VIAKLG ATAL LLDRLN UTQ	300 2SKSP 2EKDK 3SGGS 2DRHK 3QQQR 5KLAG 5KLAT 5KDS. 4QLES	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-COV IBV		* RASSRSSTENN RAPRSCSFSC RSSCAFSFQC RANSRSFGC SGRSAFASFSGN SGRSAFASFSG SGRSAFASFSG SGRSAFASFSG RGGSCASSSS RGRSGRSTASS *	R DS SR R GRGE K SRDP NNRSC R SQSR R SQSR R SQSR F SQSR P SRA AA SR	220 S IPTKQP SPSNNRGN S SPSNNRGN S SP AP AP	i NQSRGNSQNR(240 GNNQGRGASQN 340	* RGGNNNNNN	260 TSRQQSRTRSDSM KPQSRNPSSDRM KSRQQSQTRKGQS KSRQQSNNRNQSN RGRQQFNNKKDDS GPNNPARSSS(SAGSRSRANSGM NNRSLSRSNSNFF RNSTPGSSRGM .SREGSRGRRSDS 360	* HP SKQKQSQ UDRGGV1 SP RP NS SP	280 2SSSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA TSRDDIVAAVKD NEQAVIA RFASTVKPDVAE STARMASGGET GDDIIARAAK 380	* AIKS GFDN.Q AIKS GFDKPQ AILGVGISPQS AIKS GIGENE AIKK GVDTEK EIAA VIAKLG QIAS VIAKLG DIAN VIAKLG AIAL LLDRLN XIIQ	300 2SKSP 2EKDK SSGGS 2DRHK QQQR 5KLAG 5KLAT 5KDS. QLES 400	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63		* RASSRSSCAFSCOR RSSCAFSCOR RSSCAFSCOR RANGRSRSCSSR SGRAFASSSS SGRAFASSSS SGRAFNSSFS RGGSCASSSS RGRSGRSTAAS	R DSSR GRGE SRDP NNRSR R SQSR R SQSR R SQSR P SRGN AAASB	220 SSPSNNRGNI SSPSNNRGNI SS SPS A.PS 320 	*	240 SNNQGRGASQN 340 C. LKKERW	* RGGNNNNNN	260 TSRQQSRTRSDSN KPQSRNPSSDRNH KSRDQSQTRKGQ: KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSSG SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG .SREGSRGRRSD: 360 ENNIGCE FRDFN	* N	280 QSSSIVAAVTI ISQDDIWAAVAA PRODIVAAVKI VEÇAVIA VEQAVIA VEQAVIA SESIVKPDAL SISIVKPDAL SISIVKPDAL SISIVKPDAL SEARMASGGET GDDIIARAAK 380	* ATKN GFDN.Q ARKSI GFDRPQ ARLGY GISPQS ARKSI GIGENF AIKKI GVDTEM ETAA VIAKLG QTASI VIAKLG ATALI LLDRLN CIQ	300 2SKSP 2EKDK 3SGGS 2DRHK (QQQR 5KLAG 5KLAT 5KDS. 4QLES 400	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E		* RASSRSSTENN RAFERSOSESO RSSSCAFSEOC RANSRSRSSS SGRAFASESG SGRAFASESG SGRAFASESG RGGSCASSESS RGRGGRSTAAS * SSSGTSTPK	R DSSR RGRGE KSRDP NNRSR RNRSQ RSQSR RSQSR RSQSR RSQSR AAASR AAASR	220 55 5P5NNRGNN 5P5NNRGNN 5P5 5	I * NQSRGNSQNRG LSQPPADKPSG	240 GNNQGRGASQN 340 QIKKPRM	* RGGNNNNNN *	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSS RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG .SREGSRGRRSDS 360 ENVIQOFC FRDFN	* N	280 2SSSTVAAVTI NSQDDIMKAVAA 2PQDDIVAAVKI SSRDDIVAAVKI AEASTVKPDVAE PTSGVTPDVAI SSSIVKPDVAI SFARMASGGET 380 GDSDIVQNGVI GDSDIVQNGVI	* ATKN GFDN.Q ATKS GFDKPQ ATKS GIEDNE ATKS GIEDNE ATKS GIEDNE ATKS GVDTEK DTAN VIAKLG DETAN VIAKLG ATAL LLDRLN ITQ	300 2SKSP 2EKDK SSGGS 2DRHK (QQQR 5KLAG 5KLAT 5KDS. QLES 400 NCA	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205
NL63 229E HKU7 PEDV TGEV MHV oc43 HKU1 SARS-CoV IBV NL63 229E HVU7		* RASERSSTENN RAPERSOSTENN RSSCAFFSCQC RANSRSRSTENC SGRAFASRSCS SGRAFASRSCS SGRAFASRSCS SGRAFASRSCS SGRAFASRS SGRASS RASS SGRASS SGRASS SGRSTPK * SSSGTSTPK	R DSSR RGRGE SRDP INNRSS RNRSCI RSCSR RTSSR RSSR RSSR AAASR AAASR	220 5 5.SPSNNRGNN 5.SSS 3.SSS 3.SSSS 3.SSSSSSSSSSSSSSSSSS	I * NQSRGNSQNR LSQFRADKPS LARSQSSETKI	240 GNNQGRCASQN 340 Q. LKKPRW SQKHEMQKPRW	* RGGNNNNNN KRVPTRE KRVPTRE	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF RNSTPGSSRGN .SREGSRGRRSDS 360 ENVIQCFGPRDLI SNVTQCFGPRDLI	*	280 2SSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVKI SRDDIVAAVKI VEQAVIA RFASTVKPDVAE STSGVTPDVAE SFARMASGGGET .GDDIIARAAK 380 GDSDLVQNGVE GSAGVVANGVE	* ATKNIGFDN.Q AIKSIGFDKPQ AILGWGISPQS AIKSIGIGENE AIKKIGVDTEK ETAAIVIAKLG QTASIVIAKLG DITANIVIAKLG ATALILLDRLN XIIQ	300 QSKSP QEKDK SSGGS PDRHK QQQR SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKLAG SKSP QEKDK SSGGS SCORES S	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205 : 205 : 300
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 DDDV		* RASSRSSTRNN RAFSRSCAFSCC RSSCAFSCCS RANSRSSSGN SGRARAFASRSG SGRARSSRST SGRASNSTFG RGGSCASSRSS RGRSGRSTAAS * SSSGTSTFK	R DSSR R GRGE K SRDF NNRSR R NRSQ R SCSR R TSSR R SQSR R SRGN AAASR AAASR	220 SSPSNNRGNI SSSSNNRGNI SSSS A.PS 320 KPNKP SSQTSAKS TPREBSI	* NQSRGNSQNR LSQPRADKPS LARSQSSETKI KSPGKTSNQS	240 GNNQGRGASQN 340 QIKKPRW SQKHEMQKPRW DSKRQVDKPRW	* RGGNNNNNN KRVPTRE KRQPNDUVT KRVPNKS	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS(SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRGR .SREGSRGRRSDS 360 ENVIQCECFRDFN SNVTQCECFRDFN SNVTQCECFRDFN SNVTQCECFRDFN SNVTQCECFRDFN	* N	280 2555 LVAAVTI ISQDD LVAAVAA PCDD LVAAVAA PCDD LVAAVAA PCDD LVAAVAA PCDD LVAAVAA PCD LVAAVAA PTSGVTPD AL 15DS LVAAVAA 380 GDSD LVQNGVD GAGVVANGVA GDSD LVQNGVD	* ATKN GFDN.Q AIKSIGFDKPQ AIKSIGFDKPQ AIKSIGIGENF AIKKIGVDTEK ETAA VIAKLG QTASIVIAKLG DTAN VIAKLG AIALILLDRLN XIIQ	300 2SKSP 2EKDK SSGGS 2DRHK (QQQR SKLAG SKLAT SKDS. NQLES 400 ENÇA STA STA STA	: 205 : 203 : 214 : 243 : 247 : 249 : 247 : 233 : 205 : 285 : 300 : 301
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV		* RASSRSSTENN RAFSRSSCAFSCQC RSSSCAFSCQC RANGRSRSCSRS SGRSAFASSSS SGRSAFASSSS SGRSAFAS	R DSSR R GRGE K SRDF NNRSR R SQSR R SQSR R SQSR R SQSR R SRGN AAASR	220 	I * NQSRGNSQNR LSQPRADKPSQ LARSQSSETK KSPGKTSNQSJ PKKNKSRATSJ	240 GNNQGRGASQN 340 Q IKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDLKDIEEW	* RGGNNNNNN KRVFTR.E KRQFNDUTT KRVPNK.S RRIPKG.E	260 TSRQQSRTRSDSM KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSS RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG .SREGSRGRRSDS 360 ENVIQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN	*	280 2555 VAAVTI NSQDDIMKAVAA 2500 VAAVRA 5500 VAAVRA FASTVKPDVAE 1505 IVKPDVAE 1505 IVKPDVAE 55ARMASGGET 380 GDSDLVQNGVE GSAGVVANGVK GDGTLVREGVE GLAEFVEKGVE	* ATKNIGEDN.Q ATKSIGEDKEQ ATKSIGIGENE ATKSIGIGENE ATKSIGIGENE ATKSIGIGENE ATKIGVDTEX DETANIVIAKLO DETANIVIAKLO ATALILIDELN ATALILIDELN AKGFQIAETI AKGFQIAETI AKGYQIAETV ASGYAQIASIA	300 2SKSP 2EKDK SSGGS 2DRHK QQQR SKLAG SKLAT SKDS. QLES 400 ENCA ESTA ESTA ESTA	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 245 : 205 : 205 : 300 : 301 : 330
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV		* RASSRSSTENN RAFSRSQSRSQC RSSSCAFSFQQC RANSRSRSQSRSR SGRAFASRSGS SGRAFASRSG SGRASNSRPG RGGSCASSRSS RGRSGRSTAAS * SSSGTSTFK KSAKTGTPKPSF GAKPKKKSAPS. QQQKPKQEKSD.	R DSSR (RGRE) (SRDP) (NNRSC) R SCSR R TSSR R TSSR R SCSR SRGN AAASR NQSFA	220 SSPSNNRGNN S.SPSNNRGNN SSPSNRGNN SSPSS APSS S.QTSAKS S.SQTSAKS S.SQTSAKS S.SQTSAKS S.SQTSAKS S.SQKNT .NSGKNT	I * NQSRGNSQNRO LSQFRADKPSO LARSQSSETKI KSPGKTSNQSI PKKNKSRATSI KERSNSKTRD	240 GNNQGRCASQN 340 Q LKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDIKDI FEW TTPKNENKHTS	* RGGNNNNNN * KRVPTRE KRVPNZ.S RRIPKG.E KRTAGK	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG SRGSRGRRSDS 360 ENVIQCFG FRDLI EDVTCCFGRDLI EDVTCCFGRDLI EDVTCFGRRLAS NSVAACFG FRGDF GDVTRFYCARSSS	*	280 2SSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA SRDDIVAAVKI ARASTVKPDVAE PTSGVTPDVAE ISDSIVKPDVAE SFARMASGGET 380 GDSDLVQNGVE GDSDLVQNGVE GDSTLVREVE GDSTLVREVE GDTLVREVE GDTLVANGSS	* ATKNIGFDN.Q AIKSIGFDKPQ AIKSIGIENE AIKSIGIENE AIKSIGIENE AIKSIQITASIVIAKLO DIANIVIAKLO AIALILLDRLN IIQ * AKGFPQIAEII AKGFPQIAEII AKGFPQIAEII ASGYAQIASIA	300 2SKSP 2EKDK SSGGS 2DRHK QQQR SKLAG SKLAT SKDS. QLES 400 ENCA ESTA ENVA ESVS	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205 : 305 : 301 : 330 : 285
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-COV IBV NL63 229E HKU7 PEDV TGEV MHV		* RASSRSSTRNN RAFSRSCAFSCC RSSSCAFSCCS RANSRSSSSR SGRAARASSSS SGRAARASSS SGRAARSSS SGRAARSSS RGGSCASSSS RGGSCASSSS RGGSCASSSS RGRSGRSTAAS * SSSGTSTPK KSAKTGTPKPSF GAKPKKKSAFS. QQQKPKQEKSD.	R DSSR R GRGE K SRDF NNRSR R NRSQ R SQSR R TSSR R SQSR R SRGN AAASR NQSFA	220 SSPSNNRGNI SSSSNNRGNI SSSS SP.	I * NQSRGNSQNR LSQPRADKPS LARSQSSETK KSPGKTSNQSJ PKKNKSRATSJ KERSNSKTRD KQVTKQSAKE	240 GNNQGRGASQN 2IKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDLKDIFEW TTPKNENKHTS VRQKILNKFRQ	* * KRVPTR.E KRVPTR.S RRIPKG.S RRIPKG.E KRTACK.C	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRGR SBEGSRGRRSDS 360 ENVIQCFGFRDLI EDVTTCFGRTLAS NSVAACFGFRGEF GDVTRFGARSSS CFVQQCFGKRGFN	*	280 2555 LVAA TI SQDD IMKA AA PQDD LVAAVRA PSRDD VAAVRA SRDD VAAVRA SFARMASGGET GDD IARAAK 380 GDSD VQNGVE GSAGVVANGVK GIAL VRB VE GIAL VRB VE GIAL VRB VE GIAL VRB VE GSEMIKIGTS	* ATKNIGFDN.Q AIKSIGFDKPQ AIKSIGFDKPQ AIKSIGIGENF AIKKIGVDTEK ETAAIVIAKLG QIASIVIAKLG DIANIVIAKLG AIALILLDRLN XIIQ	300 2SKSP 2EKDK SSGGS 2DRHK (QQQR SKLAG SKLAG SKLAG SKLAS QLES 400 ENCA ESTA ESTA ESTA ESTA ESTA	: 205 : 203 : 214 : 247 : 249 : 247 : 233 : 205 : 205 : 300 : 301 : 330 : 330 : 319
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV MHV OC43		* RASSRSSTENN RAFSRSSC RSSSCAFSFQQ RANSRSRSCSFSR SGRSAFASSSS SGRSAFASSSS SGRSAFASS SGR	R DSSR R GRGE K SRDF NNRSR R SCSR R SCSR R SCSR R SCSR R SRGN AAASR NCSFA	220 SSPSNNRGNI SS SPSNNRGNI SS SPS SPS SSQTSAKS SSQTSAKS SRSKSI SRSKSI SRSKSI	I * NQSRGNSQNR LSQPRADKPS LARSQSSETK KSPGKTSNQSJ PKKNKSRATSJ KERSNSKTRD KQVTKQSAKE QOVTKHTAKE	240 GNNQGRGASQN 340 QLKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDLKDIEW TTPKNENKHTS VRQKIINKPRQ	* RGGNNNNNN KRVPTR.E KRQPNDVT KRVPNK.S RRIPKG.E KRIPKG.E KRIPKK.Q	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQf KSRNQSNNRNQSS RGRQQFNNKKDDS .GPNNRARSSS(SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG .SREGSRGRRSDS 360 ENVIQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SVAACFG FRGSS CFVQQCFG KRGPN CTVQCCFG KRGPN	*	280 25S5UVAAUTI ISQDDIMKAVAA 25RDDIVAAVRA SRDDIVAAVRA FASTVKPDVAE FPTSGVTPDVAI ISDSIVKPDVAE SFARMASGGET . GDDIIARAAK 380 CDSDLVQNGVE GSAGVVANGVK GDGTLVRBGVE GIAEFVERGVE GGESMIKIGTS GGESMIKIGTS	* ATKNIGFDN.Q ATKSIGFDKPQ ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATKIGVDTEM DIANIVIAKIG ATALILIDRIN XITQ	300 2SKSP 2EKDK SSGGS 2DRHK QQQR SKLAG SKLAT SKDS. VQLES 400 ENVA ESTA ESTA ESTA ESTA	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 233 : 205 : 205 : 300 : 301 : 301 : 330 : 319 : 321
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1		* RASSRSSTENN RAFSRSQSFSQC RSSSCAFSFQCQ RANSRSRSQSFSR SGRSAFASFSGN SGRSAFASFSGS SGRSASNSFPG RGGSCASSFSS RGRSGRSTAAS * SSSGTSTPK KSAKTGTPKPSF GAKPKKKSAFS QQQKPKQEKSD	R DSSR GRGE (SRDP) NNRSP RSCSR FSCSR FSCSR FSCSR AAASB NQSFA	220 	i * NQSRGNSQNR(LSQPPADKPS(LARSQSSETKI KSPGKTSNQSI KERSNSKTRD KQVTKQSAKE QQVTKHTAKE QQVTKHTAKE	240 GNNQGRGASQN 340 Q. LKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDLKDIPEW TTPKNENKHTS VRQKIINKPRQ VRQKIINKPRQ	* * KRVPTRE KRVPTR.S RRIPKG.E KRTPNK.Q KRTPNK.Q KRSPNK.Q KRSPNK.Q	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRDQSNRRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG SAGO ENVIQCFG FRDFN SNTQCFG FRDFN SNTQCFG FRDFN SNTQCFG FRGFF GDVTRFYGARSSS CFVQQCFG KRGFN CTVQQCFG KRGFN	*	280 2SSDIVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA SRDDIVAAVKI VEÇAVIA RFASTVKPDVAE ISDSIVKPDVAE ISDSIVKPDVAE GDDIIARAAK 380 GDSDIVQNGVI GJSTLVREVE GIAEFVEKGVI GDTLVRESE GGGEMIKIGTS GGGEMIKIGTS	* ATKNIGFDN.Q ATKSIGFDKPQ ATKSIGIENE ATKSIGIENE ATKSIGIENE ATKIGVIAKIG DETANIVIAKIG DETANIVIAKIG ATALILIDRIN ITQ	300 2SKSP 2EKDK SSGGS 2DRHK QQQR SKLAG SKLAT SKDS. QQLES 4000 ENCA ESTA ESTA ESTA ESTA ESTA ESTA	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205 : 300 : 301 : 330 : 319 : 319 : 319
NL63 229E HKU7 PEDV TGEV MHV oc43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV MHV oc43 HKU1		* RASSRSSTRNN RAFSRSCAFSCC RSSSCAFSCCS RANSRSRSCAFSCCS SGRSAFASRSS SGRSAFASRSS SGRSAFNSFSG RGGSCASSRSS RGRSGRSTAAS * SSSGTSTPK KSAKTGTPKPSF GAKPKKKSAPS QQQKPKQEKSD	R DSSR R GRGE K SRDE NNRSR R NRSQ R SQSR R TSSR R SQSR R SRGN AAASR/	220 5	k LSQPRADKPS LSQPRADKPS LARSQSSETK KSPGKTSNQSJ PKKNKSRATSJ KERSNSKTRD KQVTKQSAKE QQVTKQNAKE QQVTKQNAKE	240 GNNQGRGASQN 340 QIKKPRW SQKHEMQKPRW DSKRQVDKPRW DSKRQVDFRW KERDLKDIFEW FROMINKPRQ VRQKIINKPRQ VRQKIINKPRQ IRHKITKPRQ	* * KRVPTR.E KRVPNK.S RRIPKG.E KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q	260 TSRQQSRTRSDSN KPQSRNPSSDRNI KSRDQSQTRKGQS KSRDQSQTRKGQS SAGSRSRANSGNI NNRSLSRSNSNFI RNSTPGSSRGN .SREGSRGRRSDS 360 ENVIQCFGPBDLI EDVTTCFGRBLI EDVTCFGPBDLI EDVTCFGPBDLI EDVTCFGRBLI EDVTCFGRBLI SNSVACFGPBGGS GDVTRFYCARSSS CEVQQCFGRBGS CNVQQCFGRBGS	*	280 2555 LVAAVTI SQDD IMKA AA PQDD LVAAVKA PSRDD LVAAVKA SRDD LVAAVKA STORE PTSGVTPD AA SDS LVQNGVA GDSD LVQNA GDSD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA GDD LVQNA	* ATKNIGFDN.Q ATKSIGFDKPQ ATKSIGFDKPQ ATKSIGTENF ATKKIGVDTEK DIAAIVIAKLO QTASIVIAKLO DIANIVIAKLO ATALILLORIN TIQ	300 2SKSP 2EKDK SSGGS 2DRHK QQQR SKLAG SKLAT SKDS. QQLES 400 ENCA ESTA ESTA ESTA ESTA ESTA ESTA ESTA EST	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 233 : 205 : 300 : 301 : 330 : 330 : 3319 : 321 : 319 : 319
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV		* RASSRSSTENN RAFSRSCAFSTQC RSSSCAFSTQC RANGRSRSCSTSR SGRSAFASSSS SGRSAFASSFSG SGRSAFASSFSG SGRSAFASSFSS RGGSCASSFSS RGGSCASSFSS RGGSCASSFSS RGRSGRSTPK. KSAKTGTPKPSF GAKPKKKSAPS QQQKPKQEKSD	R DSSR R GRGE K SRDF NNRSR R SCSR R SCSR R SCSR R SCSR R SCSR AAASR AAASR	220 	* * LSQPRADKPS(LARSQSSETKI KSPGKTSNQSI PKKNKSRATSI KERSNSKTRD QQVTKQSAKE QQVTKQNAKE QQVTKQNAKE QQVTKQNAKE	240 GNNQGRGASQN 340 Q. LKKPRW EQKHEMQKPRW DSKRQVDKPRW KERDIKDIFEW TTPKNENKHTS VRQKIINKPRQ URQKIINKPRQ IRHKIITKPRQ SAAFASKKPRQ	* RGGNNNNNN KRVPTEE KRQPNDUT KRVPNKS RRIPKGE KRTPNKQ KRTPNKQ KRTPNKQ	260 TSEQQSRTESDSN KPQSRNPSSDRNF KSRDQSQTEKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRAESSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG SRGSRGRRSDS 360 ENVIQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRDFN SNVTQCFG FRGFS GDVTEFYCARSSS CFVQQCFG KRGPS CFVQQCFG KRGPS YNVTCAFG RBGFF	*	280 25S51VAAVTI ISQDDIMKAVAA 29QDDIVAAVRA SRDDIVAAVRA FSRDJVAAVKI SSSIVKPDVAE FPTSGVTPDVAI ISDSIVKPDVAI ISDSIVKPDVAI GDDIIARASGGET .GDDIIARASG GDSDLVQNGVE GSAGVVANGVK GDGTLVRBGVE GIAEFVERGVE GGSEMIKIGTS GGSEMIKIGTS GNAEMIKIGTD	* ATKNIGFDN.Q ATKSIGFDKPQ ATKSIGFDKPQ ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATKSIGIGENF ATALILIDELN AKGYPQTAEIV AKGYPQTAEIV AKGYPQTAEIV AKGYPQTAEIV AKGYPQTAEIV DPQFTIAEIA DPQFTIAEIA	300 2SKSP 2EKDK SSGGS SCDRHK QQQR SKLAG SKDS. VQLES 400 ENCA ESTA ESTA ESTA ESTA ESTA ESTA ESTA EST	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 249 : 245 : 300 : 301 : 330 : 301 : 330 : 319 : 313 : 313
NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV NL63 229E HKU7 PEDV TGEV MHV OC43 HKU1 SARS-CoV IBV		* RASSRSSTENN RAFSRSQSESQ RSSSCAFSEQQ RANSRSRSQSESR SGRAFASSSS SGRAFASSSS RGGSCASSESS RGGSCASSESS RGGSGRSTAAS * SSSGTSTPK KSAKTGTPKPSF GAKPKKKSAPS QQQKPKQEKSD	F DSSR GRGE KSRDF NRSC F SCSR TSCSR TSCSR SCSR SCSR SCSR SCSR NQSFA	220 5.	* * NQSRGNSQNR(* LSQPPADKPS(LSRSQSSETKI KSPGKTSNQSI PKKNKSRATSI KERSNSKTRD' KQVTKQSAKE' QQVTKQSAKE QQVTKQNAKE QQVTKQNAKE QQVTKQNAKE QQVTKQSRITKAI	240 GNNQGRGASQN QLKKPRW SQKHEMQKPRW DSKRQVDKPRW KERDLKDIFEW TTPKNENKHTS VRQKIINKPRQ VRQKIINKPRQ SAAEASKKPRQ KADEMAHRRYC	* RGGNNNNNN * KRVFTR.E KRQPNDVT KRVPNK.S RRIPKG.E KRTPKK.Q KRTPNK.Q KRSPNK.Q KRTPKK.H KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q KRTPNK.Q	260 TSRQQSRTRSDSN KPQSRNPSSDRNF KSRDQSQTRKGQS KSRNQSNNRNQSN RGRQQFNNKKDDS .GPNNRARSSS SAGSRSRANSGNF NNRSLSRSNSNFF .RNSTPGSSRG SNSTQCFG FRDFN SNYTQCFG FRDFN SNYTQCFG FRDFN SNYTQCFG FRDFN SNYTQCFG FRGFF GDVTRFYG ASSS CPVQQCFG KRGFN CTVQQCFG KRGFN CTVQCFG KRGFN CTVQQCFG KRGFN CTVQQCFG KRGFN CVQQCFG KRGFN CVQQCFG KRGFN CVQQCFG KRGFN CVQQCFG KRGFN CVQCFG KRGFN CVQQFG KRGFN CVQCFG KRGFN CVQCFN CVQCFG KRGFN CVQCFN CVQCFG KRGFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN CVQCFN	*	280 25S5TVAAVTI NSQDDIMKAVAA 2PQDDIVAAVRA SRDDIVAAVRA EASTVKPDVAE ISDSIVKPDVAE ISDSIVKPDVAE GDSDIVQNGVE GSAGVVANGVE GDSDIVQNGVE GSAGVVANGVE GGGEMIKIGTS GGGEMIKIGTS GGGEMIKIGTS GDQDLIRQCDI	* ATKNIGFDN.Q ATKSIGFDKPQ ATKSIGTENE ATKSIGTENE ATKSIGTENE ATKIGVDTEK ETAATVIAKLO DETANIVIAKLO ATATILLDRLN ITALILLDRLN ITQ	300 2SKSP 2EKDK SSGGS CDRHK CQQQR 5KLAG 5KLAT 5KDS. 1QLES 4000 4 NÇA 4 NÇA 4 STA 5 S	: 205 : 203 : 214 : 243 : 211 : 247 : 249 : 247 : 205 : 205 : 205 : 300 : 301 : 330 : 285 : 319 : 319 : 319 : 313 : 279

Figure 2. Comparative analyses of HCoV-NL63 N. Selected HCoV-NL63 N-homologues were aligned with CLUSTAL X version 2.0 (Larkin et al. 2007) and viewed with GENEDOC version 2.6.002 software (Nicholas et al; 1997). Shading indicates conserved regions and gaps were introduced to align sequences. Conserved region identified: (i) FYY(L/T)GTGP. N amino acid sequences were obtained from NCBI (accession numbers in brackets): HCoV-NL63 (DQ846901.1), SARS-CoV (AY360146.1), MHV (AAU06361), IBV (AAA46214), HCoV-OC43 (AY585229.1), HCoV-229E (AAG48597), TGEV (AAA47915), PEDV (AF353511), HKU1 (AAT98585) and HKU7 (DQ666343).

			*	420	*	440	*	460	*	480	*	500	
NL63	:	ALFEDSEV	STDEVG.		DNVQITYT .	YKMLVAK	DNKNIPKFIE	ISAFTKPSS:	IKEMQSQSSHV	VQNT		VL	: 351
229E	:	AMLFDSHI	VSKESG.		NTVVLTET.	TRVTVPK	DHPHIGKFLEI	LNAFTREN	MQQHPLLNPSA	LEFN		PS	: 364
HKU7	:	ALLEDSEI	STKEAG.		EDVIITYH.	YRMKVPK	SDKNISRFLQ	VSAYAD.SS	SLDGSKLDSNA	PEF		VPS	: 366
PEDV	:	ALLFGGNV	AVRELA.		DSYEITYN.	YKMTVPK	SDPNVELLVS	VEAFKTGNA	KLORKKEKKNK	RETTLOOH.	EEAIYDD	VGAPS	: 410
TGEV	:	SILFGSYW	TSKEDG.		DOIEVIFT.	HKYHLFK	DDPKTGOFL	INAYARPSEV	AKEORKRKSR	SKSAER		SEQ	: 354
MHV	:	AFFFGSKI	ELVKKN.	. SGGADEPTI	UVYELOYSO	AVREDSTLPG	FETIMKVLNE	LNAYQKI	DGGADVVSPKP	QRKGRRÇA.	QEKKDEVI	DNVSV	: 408
OC43	:	AFFFGSRI	ELAKVQN	NLSGNPDEPQI	DVYELRYNG	AIRFDSTISG	FETIMKVLNEI	NLNAYQQ	DGMMNMSPKP	QRQRGH	KNGQGENI	DNISV	: 411
HKU1	:	AFFFGSKI	DLVKRD.	SEADSPVI	UVFELHYS	SIRFDSTLFG	FETIMKVLEE	LNAYVNSNQ		QRKRGV.	KQLPEQFI	DSLNL	: 409
SARS-CoV	:	AFFGMSRI	GMEVTP.		SGTWLTYHO	AIKLDDKDPQ	FKDNVILLNK	LICAYKTEPP	TEPKKDKKKKT	DEAQPLP		QR	: 386
IBV	:	ACLEGSRV	TPKLQL.		DGLHLRFEE	TTVVPCDDPQ	FDNYVKICDQ	VD <mark>GVGTRP</mark> KI	DDEPKPKSRSS	SREATRGNSE	APRQQRPKKE	KKLKK	: 368
						-							
			*	520	*	540							
NL63	:	NASIPE	SKPLADI	DSAIIEIVNI	VLH	:	377						
229E	:	QTSFAT	AEPVRDE	EVSIETDIIDI	v <mark>n</mark>	:	389						
HKU7	:	GAVGG	FEPEQD.	VIVEMVDI	VKEAEV	:	391						
PEDV	:	DVTHANLE	WDTAVD	GDTAVE I I NI	IFDTGN		441						
TGEV	:	DVVPEALI	ENYTOVE	DDTQVEIIDI	VTN	:	382						
MHV	:	AKPKSSVQ	RNVSREI	TPEDRSLIA	ILDDGVVPI	GLEDDSNV :	450						
OC43	:	AVPKSRVQ	QNKSREI	TAEDISLIK	MDEP	YTEDTSEI :	448						

Figure 2. Continued.

HKU1

TBV

spacer, which is then followed by another basic region composed of 3 basic residues out of 5 residues. PSORT II (http://www.genscript.com/psort/psort2.html) identified two possible NLSs for HCoV-NL63 N. A predicted pat4 motif was identified at amino acid position 232-235 (KKPR) and a predicted pat7 motif at position 234-240 (PRWKRVP); no bipartite motifs were identified for NL63 N-protein.

: SAG....TQHISNDFTPEDHSLIATIDDP.....YVEDSVA. : 441

: QDDEADKALTSDEERNNAQLEFYDEPKVINWGLAALGENEL. : 409

SARS-COV : QKKQPTVTLLFAADMDDFSRQLQNSMSGASADSTÇA..... : 422

Predictors of Natural Disordered Regions (PONDR) was used to predict possible disordered regions within HCoV-NL63 N. PONDR (http://www.pondr.com) uses several algorithms, including VL-XT, XL1-XT, VL3 and VSL1 to predict possible disordered regions. In this study, the VL-XT and VSL1 algorithms were used. VL-XT incorporates 3 feed-forward neural networks which are based on long disordered regions characterized by x-ray chrystallographic data (Li et al., 1999; Romero et al., 1997, 2001). The VSL1 utilizes two algorithms for short (< 30aa) and long (> 30aa) disordered regions and is trained on 1335 protein sequences (Obradovic et al., 2005). Even though in silico analysis is a good predictor of protein structure, errors can occur and findings must be substantiated experimental evidence. with The order/disorder state of HCoV-NL63 N was different to that of other coronavirus N-proteins, with HCoV-NL63 N having only one disordered region (residues 109 to 248) compared to the 2 to 3 disordered regions within other coronavirus N-proteins (Chen et al., 2005). Interestingly, many of the motifs and/or regions important for the interaction of SARS-CoV N with other proteins fall within these disordered regions (He et al., 2004; Huang et al., 2004; Luo et al., 2004; Surjit et al., 2004b). For this reason, residues 109 to 248 of HCoV-NL63 N could be crucial for interactions between HCoV-NL63 and other

proteins.

In this study, molecular and bioinformatics tools were used to analyse HCoV-NL63 nucleocapsid protein by comparing it to selected coronavirus N proteins. Putative motifs that could be essential for the functioning of this protein have been identified; these are currently being verified in our laboratory. To our knowledge, this is the first manuscript reporting the molecular analysis of HCoV-NL63 N.

ACKNOWLEDGEMENTS

The authors wish to thank Dr M Muller (University of Bonn, Germany) for the recombinant HCoV-NL63 N plasmid used for mammalian expression. HCoV-NL63 RNA was a kind gift from Prof Dr Lia van der Hoek (Netherlands). This research was funded by the National Research Foundation (NRF), South Africa. MB received financial support from NRF. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regard thereto.

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-410.
- Bastien N, Anderson K, Hart L, Van Caeseele P, Brandt K, Milley D, Hatchette T, Weiss EC, Li Y (2005). Human coronavirus NL63 infection in Canada. J. Infect. Dis. 191:503-506.
- Blanchard EG, Miao C, Haupt TE, Anderson LJ, Haynes LM (2011). Development of a recombinant truncated nucleocapsid protein based immunoassay for detection of antibodies against human coronavirus OC43. J. Virol. Methods 177:100-106.

- Cabeca TK, Bellei N (2012). Human coronavirus NL-63 infection in a Brazilian patient suspected of H1N1 2009 influenza infection: description of a fatal case. J. clin. virol. official publ. Pan Am. Soc. Clin. Virol. 53:82-84.
- Cawood R, Harrison SM, Dove BK, Reed ML, Hiscox JA (2007). Cell cycle dependent nucleolar localization of the coronavirus nucleocapsid protein. Cell Cycle 6:863-867.
- Chang CK, Hsu YL, Chang YH, Chao FA, Wu MC, Huang YS, Hu CK, Huang TH (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J. Virol. 83:2255-2264.
- Chen H, Gill A, Dove BK, Emmett SR, Kemp CF, Ritchie MA, Dee M, Hiscox JA (2005). Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance. J. Virol. 79:1164-1179.
- Fielding BC (2011). Human coronavirus NL63: a clinically important virus? Future microbial. 6:153-159.
- Fielding BC, Suliman T (2009) Comparative analysis of human coronavirus-NL63 ORF3 protein homologues. Afr. J. Biotechnol. 8:3175-3178.
- He R, Dobie F, Ballantine M, Leeson A, Li Y, Bastien N, Cutts T, Andonov A, Cao J, Booth TF, Plummer FA, Tyler S, Baker L, Li X (2004). Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochem. Biophys. Res. Commun. 316:476-483.
- He R, Leeson A, Andonov A, Li Y, Bastien N, Cao J, Osiowy C, Dobie F, Cutts T, Ballantine M, Li X (2003). Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein. Biochem. Biophys. Res. Commun. 311:870-876.
- Hiscox JA, Wurm T, Wilson L, Britton P, Cavanagh D, Brooks G (2001). The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. J. Virol. 75:506-512.
- Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW, Olejniczak ET (2004). Structure of the N-terminal RNAbinding domain of the SARS CoV nucleocapsid protein. Biochemistry 43:6059-6063.
- Khan S, Fielding BC, Tan TH, Chou CF, Shen S, Lim SG, Hong W, Tan YJ (2006) Over-expression of severe acute respiratory syndrome coronavirus 3b protein induces both apoptosis and necrosis in Vero E6 cells. Virus. Res. (122)1-2: 20-27.
- Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P (2007). Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J. Virol. 81:548-557.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.
- Lee HK, Lee BH, Dutta NK, Seok SH, Baek MW, Lee HY, Kim DJ, Na YR, Noh KJ, Park SH, Kariwa H, Nakauchi M, Mai le Q, Heo SJ, Park JH (2008). Detection of antibodies against SARS-Coronavirus using recombinant truncated nucleocapsid proteins by ELISA. J. Microbiol. Biotechnol. 18:1717-1721.
- Li X, Romero P, Rani M, Dunker AK, Obradovic Z (1999). Predicting Protein Disorder for N-, C-, and Internal Regions. Genome informatics. Workshop Genome Inform. 10:30-40.
- Lin HX, Feng Y, Tu X, Zhao X, Hsieh CH, Griffin L, Junop M, Zhang C (2011). Characterization of the spike protein of human coronavirus NL63 in receptor binding and pseudotype virus entry. Virus Res. 160:283-293.
- Lin HX, Feng Y, Wong G, Wang L, Li B, Zhao X, Li Y, Smaill F, Zhang C (2008). Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction. J. Gen. Virol. 89:1015-1024.
- Luo C, Luo H, Zheng S, Gui C, Yue L, Yu C, Sun T, He P, Chen J, Shen J, Luo X, Li Y, Liu H, Bai D, Yang Y, Li F, Zuo J, Hilgenfeld R, Pei G, Luo H, Chen Q, Chen J, Chen K, Shen X, Jiang H (2005). The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1.

FEBS Lett. 579:2623-2628.

- MacKenzie JJ, Gomez ND, Bhattacharjee S, Mann S, Haldar K (2008). A *Plasmodium falciparum* host-targeting motif functions in export during blood stage infection of the rodent malarial parasite *Plasmodium berghei*. PLoS One. 3: e2405.
- Mahotka C, Liebmann J, Wenzel M, Suschek CV, Schmitt M, Gabbert HE, Gerharz CD (2002). Differential subcellular localization of functionally divergent survivin splice variants. Cell Death Differ. 9: 1334-1342.
- Mathewson AC, Bishop A, Yao Y, Kemp F, Ren J, Chen H, Xu X, Berkhout B, van der Hoek L, Jones IM (2008). Interaction of severe acute respiratory syndrome-coronavirus and NL63 coronavirus spike proteins with angiotensin converting enzyme-2. J. Gen. Virol. 89:2741-2745.
- Mourez T, Vabret A, Han Y, Dina J, Legrand L, Corbet S, Freymuth F (2007). Baculovirus expression of HCoV-OC43 nucleocapsid protein and development of a Western blot assay for detection of human antibodies against HCoV-OC43. J. Virol. Methods 139:175-180.
- Muller MA, van der Hoek L, Voss D, Bader O, Lehmann D, Schulz AR, Kallies S, Suliman T, Fielding BC, Drosten C, Niedrig M (2010). Human coronavirus NL63 open reading frame 3 encodes a virionincorporated N-glycosylated membrane protein. Virol. J. 7:6.
- Nicholas KB, Nicholas HB Jr., Deerfield DW (1997). GeneDoc: analysis and visualization of genetic variation. EMBNEW News 4:14.
- Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK (2005). Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61(Suppl 7):176-182.
- Oosterhof L, Christensen CB, Sengelov H (2010). Fatal lower respiratory tract disease with human corona virus NL63 in an adult haematopoietic cell transplant recipient. Bone marrow transplant. 45:1115-1116.
- Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, van der Hoek L (2006). Mosaic structure of human coronavirus NL63, one thousand years of evolution. J. Mol. Biol. 364:964-973.
- Pyrc K, Jebbink MF, Berkhout B, van der Hoek L (2004). Genome structure and transcriptional regulation of human coronavirus NL63. Virol. J. 1:7.
- Romero P, Obradovic Z, Dunker K (1997). Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. Genome informatics. Workshop Genome Inform. 8:110-124.
- Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001). Sequence complexity of disordered protein. Proteins 42:38-48.
- Rowland RR, Chauhan V, Fang Y, Pekosz A, Kerrigan M, Burton MD (2005). Intracellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein: absence of nucleolar accumulation during infection and after expression as a recombinant protein in vero cells. J. Virol. 79:11507-11512.
- Stohlman SA, Fleming JO, Patton CD, Lai MM (1983). Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein. Virology 130:527-532.
- Sung JY, Lee HJ, Eun BW, Kim SH, Lee SY, Lee JY, Park KU, Choi EH (2010). Role of human coronavirus NL63 in hospitalized children with croup. Pediatr. infect. dis. J. 29:822-826.
- Surjit M, Liu B, Chow VT, Lal SK (2006). The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells. J. Biol. Chem. 281:10669-10681.
- Surjit M, Liu B, Jameel S, Chow VT, Lal SK (2004a). The SARS coronavirus nucleocapsid protein induces actin reorganization and apoptosis in COS-1 cells in the absence of growth factors. Biochem. J. 383:13-18.
- Surjit M, Liu B, Kumar P, Chow VT, Lal SK (2004b). The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. Biochem. Biophys. Res. Commun. 317:1030-1036.
- Timani KA, Liao Q, Ye L, Zeng Y, Liu J, Zheng Y, Yang X, Lingbao K, Gao J, Zhu Y (2005). Nuclear/nucleolar localization properties of Cterminal nucleocapsid protein of SARS coronavirus. Virus Res. 114:23-34.
- van der Hoek L (2007). Human coronaviruses: what do they cause? Antivir. ther. 12:651-658.
- 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. 10:368-373.

- van der Hoek L, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Uberla K (2005). Croup is associated with the novel coronavirus NL63. PLoS medicine 2: e240. van der Hoek L, Sure K, Ihorst G, Stang A, Pyrc K, Jebbink MF, Petersen G, Forster J, Berkhout B, Uberla K (2006). Human coronavirus NL63 infection is associated with croup. Adv. Exp. Med. Biol. 581:485-491.
- Wang Y, Zhang X (1999). The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 *in vitro* and *in vivo*. Virology 265:96-109.
- Woo PC, Lau SK, Huang Y, Yuen KY (2009). Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. (Maywood) 234:1117-1127.

- Wurm T, Chen H, Hodgson T, Britton P, Brooks G, Hiscox JA (2001). Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J. Virol. 75:9345-9356.
- Yan X, Hao Q, Mu Y, Timani KA, Ye L, Zhu Y, Wu J (2006). Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein. Int. J. Biochem. Cell Biol. 38:1417-1428.
- You J, Dove BK, Enjuanes L, DeDiego ML, Alvarez E, Howell G, Heinen P, Zambon M, Hiscox JA (2005). Subcellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein. J. Gen. Virol. 86:3303-3310.