Journal of General Virology (2008), **89**, 1015–1024 DOI 10.1099/vir.0.83331-0

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

Han-Xin Lin,[1] Yan Feng,[1] Gillian Wong,[1] Liping Wang,[2] Bei Li,[1] Xuesen Zhao,[1] Yan Li,[3] Fiona Smaill[1] and Chengsheng Zhang[1]

Correspondence
Chengsheng Zhang
zhangch@mcmaster.ca

[1]Department of Pathology and Molecular Medicine, McMaster University and Department of Microbiology, St Joseph's Healthcare, Hamilton, ON L8N 4A6, Canada

[2]Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[3]National Microbiology Laboratory, Canadian Science Center for Human and Animal Health, Winnipeg, MB R3E 3R2, Canada

Human coronavirus NL63 (NL63), a member of the group I coronaviruses, may cause acute respiratory diseases in young children and immunocompromised adults. Like severe acute respiratory syndrome coronavirus (SARS-CoV), NL63 also employs the human angiotensin-converting enzyme 2 (hACE2) receptor for cellular entry. To identify residues in the spike protein of NL63 that are important for hACE2 binding, this study first generated a series of S1-truncated variants, examined their associations with the hACE2 receptor and subsequently mapped a minimal receptor-binding domain (RBD) that consisted of 141 residues (aa 476–616) towards the C terminus of the S1 domain. The data also demonstrated that the NL63 RBD bound to hACE2 more efficiently than its full-length counterpart and had a binding efficiency comparable to the S1 or RBD of SARS-CoV. A further series of RBD variants was generated using site-directed mutagenesis and random mutant library screening assays, and identified 15 residues (C497, Y498, V499, C500, K501, R518, R530, V531, G534, G537, D538, S540, E582, W585 and T591) that appeared to be critical for the RBD–hACE2 association. These critical residues clustered in three separate regions (designated RI, RII and RIII) inside the RBD, which may represent three receptor-binding sites. These results may help to delineate the molecular interactions between the S protein of NL63 and the hACE2 receptor, and may also enhance our understanding of the pathogenesis of NL63 and SARS-CoV.

Received 22 July 2007
Accepted 17 December 2007

INTRODUCTION

Coronaviruses consist of a large and diverse family of enveloped, positive-sense, single-stranded RNA viruses. They can infect a broad range of mammalian and avian species, causing a variety of diseases in the respiratory, gastrointestinal, hepatic and central nervous systems (Holmes & Lai, 1996). The outbreak of severe acute respiratory syndrome (SARS) in 2002–2003, which was caused by a highly pathogenic virus named SARS coronavirus (SARS-CoV), posed significant challenges to medical communities worldwide (Ksiazek *et al.*, 2003; Peiris *et al.*, 2003). Since then, two additional human coronaviruses, HCoV-NL63 (NL63) and HCoV-HKU1, have been identified (van der Hoek *et al.*, 2004; Woo *et al.*, 2005). NL63 has been found to be associated with diseases in the upper and lower respiratory tracts, such as bronchiolitis, conjunctivitis, croup and pneumonia in young children and immunocompromised adults (Arden *et al.*, 2005; Bastien *et al.*, 2005; Ebihara *et al.*, 2005; Gerna *et al.*, 2006; Kaiser *et al.*, 2005; van der Hoek *et al.*, 2005, 2006). So far, NL63 infections have been reported in 12 countries across Europe, Asia and North America, indicating that it is circulating among the human population worldwide (Arden *et al.*, 2005; Bastien *et al.*, 2005; Han *et al.*, 2007; Koetz *et al.*, 2006; Lau *et al.*, 2006; Suzuki *et al.*, 2005; Vabret *et al.*, 2005).

NL63 is a member of the group I coronaviruses, which also includes HCoV-229E (229E), feline infectious peritonitis virus 79-1146, feline enteric coronavirus 79-1683, canine coronavirus and porcine transmissible gastroenteritis virus (TGEV) (van der Hoek *et al.*, 2004, 2006). The genome of NL63 contains 27553 nt and encodes the polyproteins 1a

and 1b, spike (S), open reading frame 3 (ORF3), envelope (E), membrane (M) and nucleocapsid (N) proteins (van der Hoek *et al.*, 2004) (Fig. 1). The S protein of NL63 is a 1356 aa, type I membrane glycoprotein, which contains a signal peptide (1–20 aa), an ectodomain (1–1302 aa), a transmembrane domain (1303–1325 aa) and a cytoplasmic tail (1326–1356 aa). The S protein can also be divided into S1 (21–717 aa) and S2 (718–1356 aa) domains based on its similarity to S proteins of other coronaviruses (Jackwood *et al.*, 2001; Sturman & Holmes, 1984). The S1 domain of coronaviruses is responsible for binding to the cellular receptor, whilst the S2 domain is mainly involved in membrane fusion during viral entry into the host cell (Bonavia *et al.*, 2003; Gallagher & Buchmeier, 2001). Human aminopeptidase N (hAPN or CD13) and human angiotensin-converting enzyme 2 (hACE2) have been identified as the functional receptors for 229E and SARS-CoV, respectively (Li *et al.*, 2003; Yeager *et al.*, 1992). As all previously known group I coronaviruses use the APN of their own host species as functional receptors (Tresnan & Holmes, 1998; Tresnan *et al.*, 1996), and viral sequence analysis indicates that NL63 is most closely related to 229E (van der Hoek *et al.*, 2004), it has been proposed that NL63 may also employ hAPN as its receptor. However, hACE2 is a functional receptor for NL63 (Hofmann *et al.*, 2005). This finding was quite surprising given that the S protein of NL63 shows only limited sequence similarities to that of SARS-CoV.

Previous studies have demonstrated that the S1 domains of coronaviruses may contain so-called receptor-binding domains (RBDs) that could directly bind to the cellular receptors. The RBD of the S protein of SARS-CoV has been mapped to a fragment consisting of aa 318–510 in the S1 domain (Babcock *et al.*, 2004; Wong *et al.*, 2004). Whilst the first 330 residues at the N terminus of the mouse hepatitis virus (MHV) S1 subunit are able to bind efficiently to carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), the RBDs of 229E and TGEV are all located at the C terminus of their S1 domains (Ballesteros *et al.*, 1997; Bonavia *et al.*, 2003; Godet *et al.*, 1994; Laude *et al.*, 1995). A recent study suggested that the central and C-terminal domains of the NL63 S protein are required for hACE2 association (Hofmann *et al.*, 2006). However, the location of the NL63 RBD remains to be defined and the critical residues for receptor binding have not been characterized. In this study, we generated a series of S1-truncated variants, tested their association with the hACE2 receptor and subsequently mapped an RBD that consists of 141 aa (aa 476–616) towards the C terminus of the NL63 S1 domain. In addition, we identified 15 residues in the RBD that appear to be critical for receptor binding. Interestingly, these critical residues seemed to cluster into three separate regions (designated RI, RII and RIII) inside the RBD, suggesting that these sites might be associated with receptor binding.

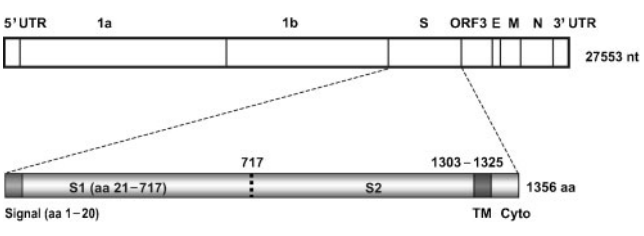

Fig. 1. Schematic diagram of the genome organization of NL63 and the deduced domains of its S protein.

METHODS

Construction of expression plasmids. A plasmid encoding hACE2 with a c-myc tag at the N terminus and a C9 tag at the C terminus of hACE2 has been described previously (Li *et al.*, 2005b). In this study, the domain structures of the NL63 S protein were defined using computer programs from the websites of SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html) and Pfam (http://www.sanger.ac.uk/Software/Pfam/). To construct the plasmids expressing S1 and its variants, we first synthesized a full-length, codon-optimized NL63 S1 gene (encoding aa 21–717) according to the method described by Babcock *et al.* (2004). All of the NL63 S1-truncated variants used in this study were generated based on this template. The RBD variants were generated by site-directed mutagenesis and/or random mutant library screening assays using a GeneMorph Random Mutagenesis kit (Stratagene) and overlapping PCR strategy. The cDNA inserts encoding the S1, RBD or a null (aa 12–327) fragment of SARS-CoV were amplified based on a codon-optimized S protein sequence of SARS-CoV described previously (Li *et al.*, 2003). For expression and production of soluble S1 or RBD variants, the cDNA inserts were cloned into a mammalian soluble protein expression vector, pSecTag/Hygro (Invitrogen), respectively, thereby producing c-myc- and His-tagged fusion proteins. In addition, some constructs were cloned into a modified pSecTag/Hygro-Ig vector that contains the mouse Igκ-chain leader sequence and human IgG Fc fragment, producing hACE2–Ig, S1–Ig, RBD–Ig or other Ig fusion proteins. All expression plasmids used in this study were confirmed by DNA sequencing.

Expression and purification of proteins. To make soluble hACE2, S1 or RBD proteins, 293T cells were transfected with the expression plasmids described above using a standard calcium phosphate transfection method. At 16–20 h after transfection, the cells were washed with PBS and cultured in serum-free medium (293 SFM II medium; Invitrogen). The protein-containing supernatants were collected on days 2, 3 and 4 post-transfection and pooled. An aliquot of the pooled protein supernatants was analysed by Western blotting (WB) to confirm the target protein expression. The supernatants were then passed through a 0.45 μm filter (Nalgene) and dialysed with binding buffer prior to affinity purification. The affinity column packed with charged nickel beads (Qiagen) was used to purify the 6 × His-tagged proteins. To purify the Ig fusion proteins (hACE2–Ig, S1–Ig or RBD–Ig), the supernatants were incubated with protein A/G–Sepharose beads at 4 °C for 16 h and then washed with 0.5 M NaCl in PBS. The samples were subsequently eluted with 50 mM sodium citrate/50 mM glycine (pH 2.0), neutralized with NaOH and dialysed against PBS as described previously (Sui *et al.*, 2004). The purified proteins were concentrated with Centricon filters (Amicon) and dialysed against PBS. Protein concentration was measured with a protein assay kit (Bio-Rad) and aliquots of the proteins were stored at –80 °C.

WB. To detect proteins by WB, purified proteins, cell culture supernatants or cell lysates were resolved by 10 % SDS-PAGE and

transferred to a nitrocellulose membrane (Amersham Biosciences). The membrane was blocked for 1 h at room temperature with blocking buffer (5 % skimmed milk, 0.05 % Tween 20 in PBS) and incubated with primary antibody (anti-c-myc antibody at a 1 : 2000 dilution or NL63-positive human serum at a 1 : 500 dilution) overnight at 4 °C, followed by three washes with washing buffer (0.05 % Tween 20 in PBS). The blot was then incubated with secondary antibody (HRP-conjugated anti-mouse IgG or HRP-conjugated anti-human IgG at a 1 : 4000 dilution; Amersham Bioscience) for 1 h at room temperature and washed three times with washing buffer and once with PBS. The protein bands were detected using ECL reagents (Amersham Biosciences). The density of the protein band was quantified using Scion Image (National Institutes of Health, Bethesda, MD, USA).

Immunoprecipitation (IP) assay. Cell lysate was prepared from hACE2-transfected 293T cells with lysis buffer consisting of 100 mM $(NH_4)_2SO_4$, 20 mM Tris/HCl (pH 7.5), 10 % glycerol, 1 % Cymal 5 and protease inhibitor cocktail. To detect an S–ACE2 association by IP, purified S1 or RBD proteins (1 μM) were mixed with 200 μg hACE2–C9 cell lysate and 2.5 μl anti-His antibody (Amersham Biosciences). The mixture was incubated at 4 °C overnight, followed by the addition of 30 μl protein A/G beads, and incubation was continued at 4 °C for 2 h. The precipitate was washed three times with Triton X-100 in PBS and resuspended in 40 μl protein loading buffer for detection of proteins by WB. Additionally, we performed IP using soluble hACE2–Ig proteins that were incubated with supernatant containing S1–c-myc–His or RBD–c-myc–His protein and protein A/G beads, followed by the steps described above.

Flow cytometry. To detect receptor expression and/or the binding interaction between S1 (or RBD) and the receptors on the cell surface, flow cytometry (or FACS) was performed as described previously (Zhang *et al.*, 1996). In brief, the target cells were detached with 5 mM EDTA in PBS and washed twice with immunofluorescence (IF) buffer (1 % fetal bovine serum and 0.02 % NaN_3 in PBS). The cells were incubated with anti-c-myc antibody (for the detection of receptor expression) or with S1–Ig or RBD–Ig protein (for the detection of binding between S1 or RBD and the receptor) on ice for 1 h, followed by three washes with IF buffer. They were further incubated with fluorescein isothiocyanate (FITC)-conjugated anti-mouse IgG or FITC-conjugated anti-human IgG on ice for 30 min. The cells were washed three times with IF buffer, fixed with 1 % formaldehyde and analysed using a Beckman Coulter FC5000 cytofluorometer.

RESULTS

The receptor-binding site is located in the C-terminal region of the S1 domain

To localize the region in the S1 domain of NL63 that is responsible for hACE2 binding, we first constructed a panel of expression plasmids that encoded full-length S1 ($S_{21–717}$) and six C-terminally truncated S1 variants: $S_{21–616}$, $S_{21–516}$, $S_{21–460}$, $S_{21–410}$, $S_{21–360}$ and $S_{21–260}$ (Fig. 2a). We then expressed and purified these S1 variants from human 293T cells. We demonstrated by WB that both anti-c-myc antibody (Fig. 2b, top panel) and NL63-positive human serum (Fig. 2b, middle panel) recognized these proteins, suggesting that they folded properly and retained their antigenicity. We also expressed and purified three S1 variants from SARS-CoV: full-length S1 ($S_{12–670}$), RBD ($S_{318–510}$) and a null fragment ($S_{12–327}$), which were detected by anti-c-myc antibody but not the NL63-positive human serum (Fig. 2b, top and bottom panels). To examine S1–hACE2 associations, we performed an IP assay using anti-His antibody, followed by WB using anti-C9

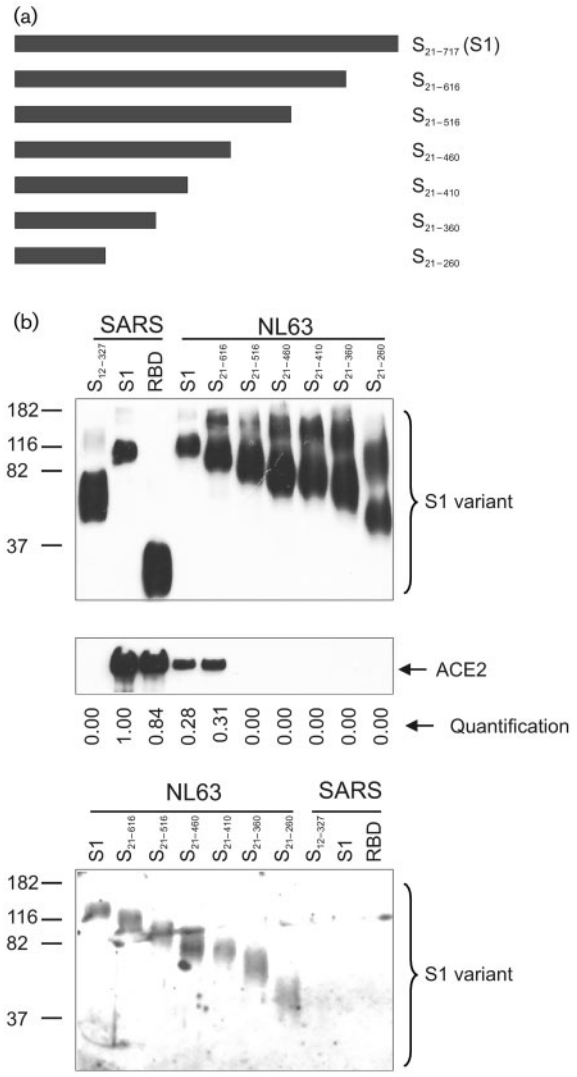

Fig. 2. Localization of the receptor-binding region of NL63 S1. (a) Diagram of the codon-optimized full-length S1 and six C-terminally truncated S1 variants constructed for this study. (b) Detection of purified S1 variant proteins by WB and S1–hACE2 associations by IP. All of the purified c-myc- and His-tagged S1 variants (500 ng each, from both SARS-CoV and NL63) were detected by WB using anti-c-myc antibody (top panel), whereas only the S1 variants of NL63 were recognized by NL63-positive human serum (bottom panel). For the IP assay (middle panel), S1 proteins (1000 nM each) were incubated with 100 μg C9-tagged hACE2 protein lysate, anti-His antibody and protein A/G beads, followed by WB detection of the precipitated hACE2 receptor using anti-C9 antibody. The data showed that hACE2 could bind to full-length S1 and to $S_{21–616}$, but not to other C-terminally truncated S1 variants of NL63.

antibody to detect precipitated ACE2 receptors. Fig. 2b (middle panel) shows that S1 and $S_{21–616}$ of NL63 were able to bind to hACE2, although their binding efficiencies appeared to be much lower than the S1 and RBD of SARS-CoV. In contrast, we did not detect an association between hACE2 and the other S1 variants ($S_{21–516}$, $S_{21–460}$, $S_{21–410}$, $S_{21–360}$ and $S_{21–260}$), suggesting that the region spanning aa 516–616 is required for receptor binding, whereas the region at the very end of the C terminus (aa 617–717) is dispensable for receptor binding. These results are consistent with previous findings reported by others (Hofmann *et al.*, 2006).

Mapping a minimal RBD

To map further the RBD, we constructed another panel of S1 variants with deletions in the N- and/or C-terminal regions of the S1 domain (Fig. 3a). Protein expression of three S1 variants ($S_{311–616}$, $S_{361–616}$ and $S_{411–616}$) was not detected in the supernatants of transfected 293T cells (data not shown), suggesting that the deleted regions could be important for stability, folding and/or secretion of the proteins. In contrast, the other five S1 variants were expressed at high levels in the supernatants of transfected 293T cells. We then purified these S1 variants that were detected by WB using anti-c-myc antibody (Fig. 3b, top panel) and performed IP to detect an S1–hACE2 association. As expected, S1 and RBD of SARS-CoV bound efficiently to hACE2, whilst the null fragment ($S_{12–327}$) did not (Fig. 3b, bottom panel). As for the NL63 S1 variants, we could detect an association between hACE2 and full-length S1 ($S_{21–717}$), $S_{210–717}$, $S_{311–717}$ and $S_{461–616}$, respectively, but not between hACE2 and $S_{311–566}$ or $S_{517–717}$. In particular, we demonstrated that the $S_{461–616}$ variant was able to bind to hACE2 much more efficiently than full-length S1 and had a binding efficiency comparable to that of S1 or RBD of SARS-CoV (Fig. 3b, bottom panel).

To determine whether $S_{461–616}$ is the minimal fragment that can bind to hACE2, we generated a further four truncated variants, $S_{476–616}$, $S_{491–616}$, $S_{461–601}$ and $S_{461–586}$. Whilst the expression levels of $S_{476–616}$ and $S_{461–601}$ were comparable to that of $S_{461–616}$, the other two variants ($S_{491–616}$ and $S_{461–586}$) did not express detectable proteins in the supernatants of transfected 293T cells (data not shown). We then purified $S_{476–616}$ and $S_{461–601}$ and used equal amounts of the proteins (Fig. 3c, top panel) for reciprocal IP assays in which either the S proteins (Fig. 3c, middle panel) or the hACE2 receptors (Fig. 3c, bottom panel) were precipitated and detected. In both IP assays, $S_{476–616}$ bound to hACE2 as efficiently as $S_{461–616}$, whilst $S_{461–601}$ did not bind. We also measured the S1–hACE2 association by flow cytometry using similar amounts of S1–Ig proteins of three NL63 S variants (S1, the minimal RBD $S_{476–616}$ and the null fragment $S_{21–460}$) and three SARS-CoV S constructs (S1, RBD and null) (Fig. 3d and e). The quantitative data presented in Fig. 3(e) showed dramatic differences in binding efficiency between NL63 RBD–hACE2 and S1–hACE2. However, we did not observe such differences among NL63 RBD–hACE2, SARS-CoV RBD–hACE2 and SARS-CoV S1–hACE2. Therefore, the flow cytometry data were consistent with the IP data. Collectively, our study defined the fragment $S_{476–616}$ as the minimal RBD of the S protein of NL63.

Identifying critical residues for RBD–hACE2 binding

A previous study on the RBD of SARS-CoV suggested that cysteine and some negatively charged residues appear to be important for receptor association (Wong *et al.*, 2004). To identify the critical residues in NL63 RBD for hACE2 binding, we first generated 20 RBD variants (Table 1) in which a single cysteine (C) or a negatively or positively charged residue was replaced by alanine (A). These mutants comprised six cysteine (C497A, C500A, C516A, C550A, C567A and C577A), two aspartic acid (D484A and D538A), three arginine (R518A, R525A and R530A), three glutamic acid (E572A, E582A and E602A) and six lysine (K501A, K532A, K546A, K556A, K562A and K564A) point mutations. The expression plasmids encoding these RBD variants were transiently transfected into 293T cells and protein expression levels in the cell culture supernatants were detected by WB. Six of the mutants (C516A, C550A, C567A, C577A, R525A and K556A) either did not express or had very low expression levels in the supernatants of transfected 293T cells, whereas the remaining 14 RBD mutants exhibited relatively high expression levels (Table 1 and Fig. 4a, top panel). Subsequently, similar amounts of protein were used for IP to detect RBD–hACE2 associations. As illustrated in the bottom panel of Fig. 4(a), alanine substitution of seven residues (C497, C500, D538, R518, R530, E582 and K501), respectively, could either abolish or dramatically decrease the receptor binding efficiency.

To identify additional residues that are critical for RBD–hACE2 binding, we constructed a random mutant library of NL63 RBD and screened protein expression of 143 mutants by WB. Surprisingly, most of the mutants (118/143, 83 %) had no or very low levels of protein expression in the supernatants of transfected 293T cells, suggesting that protein folding and/or stability of the RBD is vulnerable to mutagenic changes (data not shown). However, considerable levels of protein expression were observed for the remaining 25 mutants (Fig. 4b, top panels). Similar amounts of protein were used for the detection of RBD–hACE2 association by IP assay (Fig. 4b, bottom panels). Eleven RBD mutants (1-5, 1-22, 1-31, 1-36, 1-74, 2-10, 3-2, 3-7, 3-42, 3-46 and 3-55) were found to have undetectable or dramatically decreased binding activities to hACE2 receptor, whilst the other 14 mutants (1-10, 1-11, 1-19, 1-23, 1-30, 1-38, 1-66, 1-72, 1-77, 1-80, 2-1, 2-20, 3-15 and 3-40) could bind efficiently to hACE2. Sequence analysis revealed that three mutants carried a single amino acid substitution (1-22 with V531E, 3-42 with V499I and 3-46 with G534A), whereas the other eight mutants harboured two or more amino acid substitutions in RBD (Table 2). To pinpoint

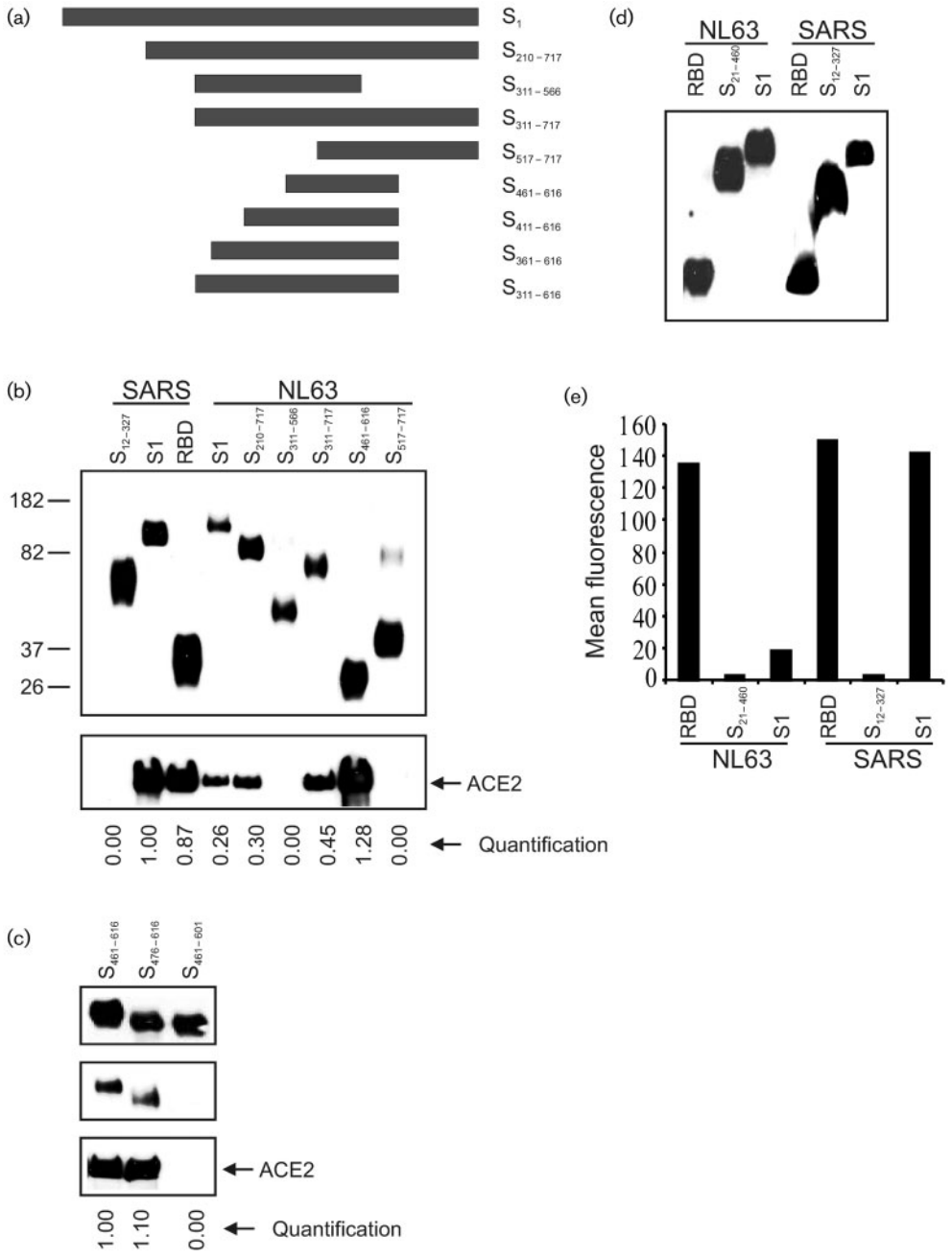

Fig. 3. Mapping of the minimal RBD. (a) Diagram of NL63 full-length S1 and eight S1 variants constructed for this study. (b) Detection of protein expression by WB and S1–hACE2 association by IP as described in Fig. 2. Three of the S1 variants ($S_{411–616}$, $S_{361–616}$ and $S_{311–616}$) were not detectable in the supernatants of transfected 293T cells (data not shown) and no further assays were conducted on these variants. Purified S1 proteins (500 ng each) were detected by WB using anti-c-myc antibody (top panel), and the precipitated hACE2 receptors were detected by WB using anti-C9 antibody (bottom panel). (c) Detection of minimal RBD–hACE2 association by reciprocal IP assays. Purified S1 proteins (1000 nM each), which were detected by WB using anti-c-myc antibody (top panel), were either incubated with soluble hACE2–Ig fusion protein and protein A/G beads followed by WB detection of the precipitated S1 proteins using anti-c-myc antibody (middle panel), or were incubated with C9-tagged hACE2 protein lysate, anti-His antibody and protein A/G beads followed by WB detection of the precipitated hACE2 receptors using anti-C9 antibodies (bottom panel). (d, e) Examination of the binding efficiency of RBD and S1 to hACE2 by flow cytometry. Human 293T cells transiently transfected with an empty plasmid (data not shown) or the plasmid encoding hACE2 were incubated with similar amounts of S1–Ig protein (d), followed by incubation with FITC-conjugated anti-human IgG Fc antibody and FACS analysis. The data are presented as the mean fluorescence value (e).

Table 1. Protein expression and receptor binding of NL63 RBD variants generated by site-directed mutagenesis with alanine substitution

Protein expression levels and receptor-binding efficiency were measured by WB. The density of protein bands was quantified using Scion Image. The density of the wt RBD band was set at 100 % and those of mutants were normalized against wt. ND, Not done.

RBD variant	Residue change	Protein expression (%)	Receptor binding (%)
Wt	None	100	100
C497A	C497→A	40	0
C500A	C500→A	48	0
C516A	C516→A	5	ND
C550A	C550→A	3	ND
C567A	C567→A	3	ND
C577A	C577→A	1	ND
D484A	D484→A	92	87
D538A	D538→A	80	0
R518A	R518→A	66	4
R525A	R525→A	0	ND
R530A	R530→A	78	0
E572A	E572→A	62	118
E582A	E582→A	85	0
E602A	E602→A	65	91
K501A	K501→A	72	10
K532A	K532→A	71	117
K546A	K546→A	100	133
K556A	K556→A	0	ND
K562A	K562→A	69	135
K564A	K564→A	79	143

further the role of individual amino acid substitutions in these mutants, we generated a series of new variants with only one amino acid substitution in each construct. Consequently, we identified an additional five residues (Y498, G537, S540, W585 and T591) that appeared to be critical for RBD–hACE2 binding (bottom panels of Fig. 4c and Table 2). The protein expression levels of two mutants (S523P and V515E) were undetectable in the supernatants and thus were not included in the IP assay (data not shown). Furthermore, we noticed that substitutions of some residues with different amino acids, such as E582A and E582K; V499I and V499L; K501A, K501M and K501N; T591A and T591I; and Y498H and Y498F exhibited similar effects on receptor binding (Tables 1 and 2 and Fig. 4), reinforcing the importance of these residues. Taken together, we identified 15 residues in the NL63 RBD (C497, Y498, V499, C500, K501, R518, R530, V531, G534, G537, D538, S540, E582, W585 and T591) that appear to be important for the RBD–hACE2 association.

DISCUSSION

In this study, we mapped residues 476–616 as the minimal RBD of the S protein of NL63. Interestingly, all known RBDs of group I coronaviruses, including 229E, TGEV and NL63, are located towards the C terminus of the S1 domain (Bonavia *et al.*, 2003; Godet *et al.*, 1994; Laude *et al.*, 1995). In contrast, one group II coronavirus, MHV, has the RBD located at the N terminus of its S1 subunit, and the SARS-CoV RBD is located near the central region of the S1 domain (Kubo *et al.*, 1994; Wong *et al.*, 2004) (Fig. 5). Therefore, the relative location of RBD in S1 of different coronaviruses seemingly correlates with their evolutionary and serological relationships. Whether or not this represents a conservative trait in the evolution of coronaviruses remains to be addressed.

Another interesting finding revealed by this study was the dramatic differences in receptor-binding efficiency between S1 and RBD of NL63 (Fig. 3). Whilst two previous studies of SARS-CoV have also suggested that RBD could bind to hACE2 more efficiently than full-length S1, they only found differences of less than twofold (Babcock *et al.*, 2004; Wong *et al.*, 2004). In contrast, we observed that hACE2 binding efficiency by the NL63 RBD was similar to that of SARS-CoV S1 or RBD, but was about six times higher than that of NL63 S1 (based on the mean fluorescent value in flow cytometry assays, Fig. 3e). This sharp difference in receptor-binding efficiency might have important implications for the pathogenesis, tropism and transmission of NL63 and SARS-CoV. It has been shown that hACE2 expression in cell lines correlates with susceptibility to SARS-CoV S-driven infection (Esposito *et al.*, 2006; Hofmann *et al.*, 2004). We have previously demonstrated that the receptor-binding affinity of the S1 protein of GD03T0013, a SARS-CoV strain associated with mild disease, was significantly lower than the S1 of Tor2, a highly pathogenic strain isolated from a patient who died of SARS (Li *et al.*, 2005b). Additional studies of the SARS-CoV-infected mouse model also strongly suggest that disease severity is correlated with S protein-mediated downregulation of ACE2 *in vivo* (Imai *et al.*, 2005; Kuba *et al.*, 2005). Therefore, the weak association between hACE2 and NL63 S1 might contribute to mild clinical symptoms in NL63-infected patients. Furthermore, the strong association of NL63 RBD with hACE2 suggests that the folding structure of RBD might be different from that of S1. It may also indicate that some efficient binding sites may be buried inside the structure of S1. This implies that NL63 possesses the potential to use hACE2 as efficiently as SARS-CoV. Mutations accumulate over time, some of which might alter the protein folding and expose the efficient binding sites on the surface. This may lead to the emergence of new NL63 strains that would be highly pathogenic to humans.

We have also conducted a series of mutagenesis studies on the NL63 RBD and identified 15 residues that appear to be critical for RBD–hACE2 interaction. We found that all of the six cysteine residues within the NL63 RBD are important for protein expression and/or hACE2 association (Table 1 and Fig. 4a), which is similar to the roles of cysteine residues in the RBD of SARS-CoV (Wong *et al.*,

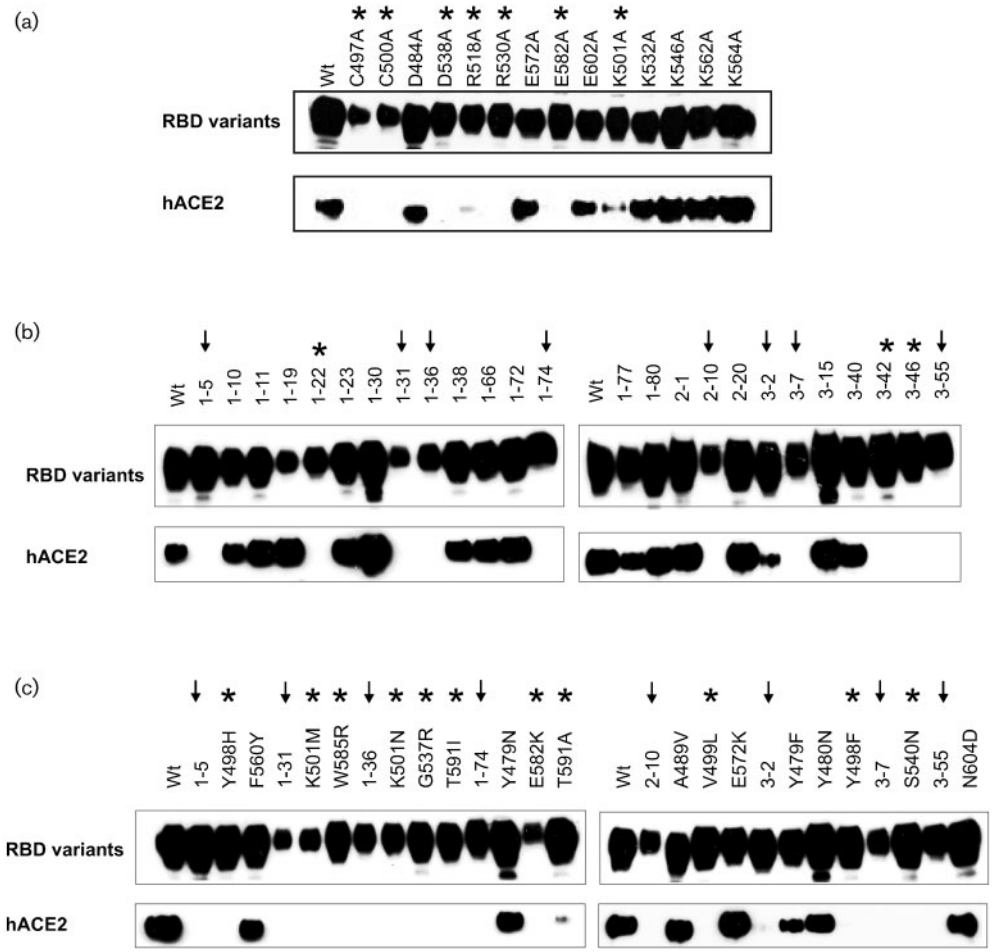

Fig. 4. Identifying critical residues for ACE2 binding. The top panels in Fig. 4(a), (b) and (c) show the wild-type (Wt) and mutant RBD protein expression levels detected by WB using anti-c-myc antibody, whilst the bottom panels show the precipitated hACE2 receptors detected by WB using anti-C9 antibody. The asterisks (*) and the arrow (↓) indicate the RBD variants that had no or dramatically decreased binding activity, respectively, to hACE2 receptor. In addition, the asterisk (*) indicates the RBD variants bearing only one amino acid substitution, whereas the arrow (↓) indicates the mutant harbouring two or more amino acid substitutions. (a) Detection of protein expression and receptor binding of 14 NL63 RBD variants generated by site-directed mutagenesis. (b) Detection of protein expression and receptor binding of 25 NL63 RBD variants generated by random mutant library-screening assays. (c) Further characterization of individual residues in RBD variants with multiple mutations found in (b).

2004). This previous study by Wong *et al.* (2004) identified two acidic residues (E452 and D454) that appeared to be important for the interaction between hACE2 and the RBD of SARS-CoV. Interestingly, two acidic residues (D538 and E582) in the NL63 RBD that appeared to be critical for receptor binding were also identified in our studies (Table 1 and Fig. 4a). In addition, five basic residues in the NL63 RBD (R518, R525, R530, K501 and K556) were identified to be critical for RBD protein expression and/or hACE2 binding (Table 1 and Fig. 4a). Furthermore, we identified seven critical residues from the screening assay of the random mutant library of NL63 RBD: two valines (V499 and V531), two glycines (G534 and G537), one tyrosine (Y498), one serine (S540) and one trytophan (W585). Our studies also suggest that, in addition to site-directed mutagenesis, screening of a PCR-generated random mutant library may become a powerful strategy for the identification of critical residues that are important for RBD–receptor binding or other protein–protein interactions.

When we tried to understand how these critical residues may interact with the hACE2 receptor, it was interesting to find that, with the exception of R518, the other 14 critical residues clustered into three separated regions (designated RI, RII and RIII) in the RBD. As shown in Fig. 5, the RI region contains residues C497, Y498, V499, C500 and K501, and RII contains residues R530, V531, G534, G537, D538 and S540, whilst RIII contains residues E582, W585 and T591. These regions may represent three receptor-binding sites inside the NL63 RBD. Previous studies of the SARS-CoV RBD–hACE2 interaction

Table 2. Protein expression and receptor binding of NL63 RBD variants generated by random library screening and site-directed mutagenesis assays

Protein expression levels and receptor-binding efficiency were measured by WB. The density of protein bands was quantified using Scion Image. The density of the wt RBD band was set at 100 % and those of mutants were normalized against wt.

RBD variant	Residue change	Protein expression (%)	Receptor binding (%)
Wt	None	100	100
1-5	Y498→H, F560→Y	108	0
Y498H	Y498→H	91	0
F560Y	F560→Y	89	66
1-22	V531→E	62	0
1-31	K501→M, W585→R	38	0
K501M	K501→M	43	0
W585R	W585→R	82	0
1-36	K501→N, G537→R, T591→I	60	0
K501N	K501→N	60	0
G537R	G537→R	84	
T591I	T591→I	77	0
1-74	Y479→N, E582→K, T591→A,	76	0
Y479N	Y479→N	98	68
E582K	E582→K	27	0
T591A	T591→A	95	5
2-10	A489→V, V499→L, E572→K	50	0
A489V	A489→V	87	92
V499L	V499→L	94	0
E572K	E572→K	77	106
3-2	Y479→F, Y480→N, Y498→F	99	12
Y479F	Y479→F	77	43
Y480N	Y480→N	104	78
Y498F	Y498→F	93	0
3-7	W585→R, S540→N	53	0
S540N	S540→N	100	0
3-15	L476→Q	101	133
3-42	V499→I	103	0
3-46	G534→A	101	0
3-55	K501→M, N604→D	74	0
N604D	N604→D	103	101

and the crystal structure of the SARS-CoV RBD–hACE2 complex have identified a panel of residues in the RBD region that are critical for the RBD–hACE2 association, and demonstrated that some of these residues may play important roles in the cross-species transmission and pathogenesis of SARS-CoV (Li *et al.*, 2005a, b). Therefore, the current study and further characterization of NL63 RBD–hACE2 interactions, including examination of the crystal structure of the NL63 RBD–hACE2 complex may not only help our understanding of the molecular interactions between the S protein of NL63 and the hACE2 receptor, but may also advance our understanding of the pathogenesis of NL63 and SARS-CoV.

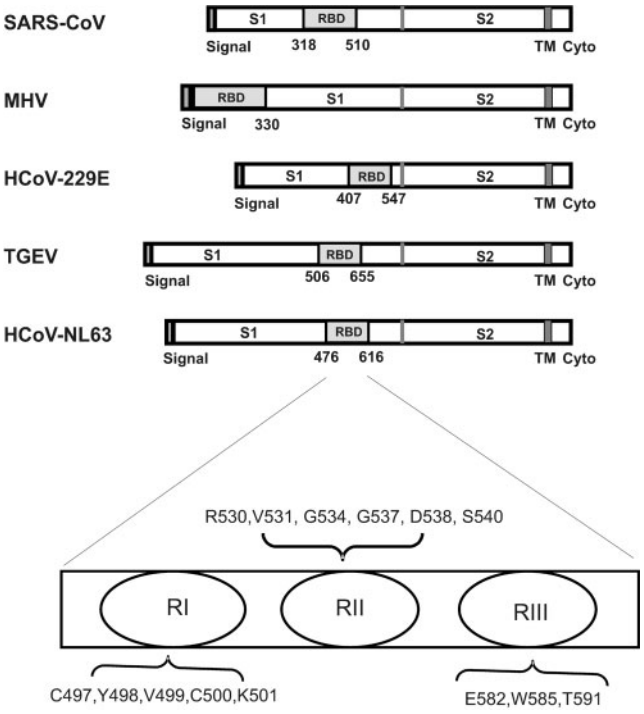

Fig. 5. Schematic diagram of the locations of the RBD of different coronaviruses and the proposed receptor-binding sites in the NL63 RBD. The locations of the RBDs of the S proteins of SARS-CoV, MHV, 229E, TGEV and NL63 are shown here. In addition, 14 of the 15 critical residues identified in this study clustered into three separate regions in the NL63 RBD, designated regions I (RI), II (RII) and III (RIII), which may represent three receptor-binding sites.

ACKNOWLEDGEMENTS

We are grateful to Dr Michael Farzan and Dr Wenhui Li from Harvard Medical School for providing reagents, Dr Gregory Babcock for providing a protocol for synthesis of the codon-optimized S gene of NL63, and Dr Lizhi He and Ms Jenny Yip for assistance with the preparation of the manuscript. This work was supported by operating grant MOP 68955 (C. Z.) from the Canadian Institute of Health Research (CIHR).

REFERENCES

Arden, K. E., Nissen, M. D., Sloots, T. P. & Mackay, I. M. (2005). New human coronavirus, HCoV-NL63, associated with severe lower respiratory tract disease in Australia. *J Med Virol* **75**, 455–462.

Babcock, G. J., Esshaki, D. J., Thomas, W. D., Jr & Ambrosino, D. M. (2004). Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J Virol* **78**, 4552–4560.

Ballesteros, M. L., Sanchez, C. M. & Enjuanes, L. (1997). Two amino acid changes at the N-terminus of transmissible gastroenteritis

coronavirus spike protein result in the loss of enteric tropism. *Virology* **227**, 378–388.

Bastien, N., Anderson, K., Hart, L., Van Caeseele, P., Brandt, K., Milley, D., Hatchette, T., Weiss, E. C. & Li, Y. (2005). Human coronavirus NL63 infection in Canada. *J Infect Dis* **191**, 503–506.

Bonavia, A., Zelus, B. D., Wentworth, D. E., Talbot, P. J. & Holmes, K. V. (2003). Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. *J Virol* **77**, 2530–2538.

Ebihara, T., Endo, R., Ma, X., Ishiguro, N. & Kikuta, H. (2005). Detection of human coronavirus NL63 in young children with bronchiolitis. *J Med Virol* **75**, 463–465.

Esposito, S., Bosis, S., Niesters, H. G., Tremolati, E., Begliatti, E., Rognoni, A., Tagliabue, C., Principi, N. & Osterhaus, A. D. (2006). Impact of human coronavirus infections in otherwise healthy children who attended an emergency department. *J Med Virol* **78**, 1609–1615.

Gallagher, T. M. & Buchmeier, M. J. (2001). Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **279**, 371–374.

Gerna, G., Campanini, G., Rovida, F., Percivalle, E., Sarasini, A., Marchi, A. & Baldanti, F. (2006). Genetic variability of human coronavirus OC43-, 229E- and NL63-like strains and their association with lower respiratory tract infections of hospitalized infants and immunocompromised patients. *J Med Virol* **78**, 938–949.

Godet, M., Grosclaude, J., Delmas, B. & Laude, H. (1994). Major receptor-binding and neutralization determinants are located within the same domain of the transmissible gastroenteritis virus (coronavirus) spike protein. *J Virol* **68**, 8008–8016.

Han, T. H., Chung, J. Y., Kim, S. W. & Hwang, E. S. (2007). Human coronavirus-NL63 infections in Korean children, 2004–2006. *J Clin Virol* **38**, 27–31.

Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G. H., Gramberg, T. & Pohlmann, S. (2004). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Commun* **319**, 1216–1221.

Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B. & Pohlmann, S. (2005). Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci U S A* **102**, 7988–7993.

Hofmann, H., Simmons, G., Rennekamp, A. J., Chaipan, C., Gramberg, T., Heck, E., Geier, M., Wegele, A., Marzi, A. & other authors (2006). Highly conserved regions within the spike proteins of human coronaviruses 229E and NL63 determine recognition of their respective cellular receptors. *J Virol* **80**, 8639–8652.

Holmes, K. V. & Lai, M. M. C. (1996). *Coronaviridae*: the viruses and their replication. In *Fields Virology*, 3rd edition, pp. 1075–1093. Philadelphia: Lippincott–Raven.

Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., Yang, P., Sarao, R., Wada, T. & other authors (2005). Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112–116.

Jackwood, M. W., Hilt, D. A., Callison, S. A., Lee, C. W., Plaza, H. & Wade, E. (2001). Spike glycoprotein cleavage recognition site analysis of infectious bronchitis virus. *Avian Dis* **45**, 366–372.

Kaiser, L., Regamey, N., Roiha, H., Deffernez, C. & Frey, U. (2005). Human coronavirus NL63 associated with lower respiratory tract symptoms in early life. *Pediatr Infect Dis J* **24**, 1015–1017.

Koetz, A., Nilsson, P., Linden, M., van der Hoek, L. & Ripa, T. (2006). Detection of human coronavirus NL63, human metapneumovirus and respiratory syncytial virus in children with respiratory tract infections in south-west Sweden. *Clin Microbiol Infect* **12**, 1089–1096.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A. & other authors (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953–1966.

Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y. & other authors (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* **11**, 875–879.

Kubo, H., Yamada, Y. K. & Taguchi, F. (1994). Localization of neutralizing epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J Virol* **68**, 5403–5410.

Lau, S. K., Woo, P. C., Yip, C. C., Tse, H., Tsoi, H. W., Cheng, V. C., Lee, P., Tang, B. S., Cheung, C. H. & other authors (2006). Coronavirus HKU1 and other coronavirus infections in Hong Kong. *J Clin Microbiol* **44**, 2063–2071.

Laude, H., Godet, M., Bernard, S., Gelfi, J., Duarte, M. & Delmas, B. (1995). Functional domains in the spike protein of transmissible gastroenteritis virus. *Adv Exp Med Biol* **380**, 299–304.

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K. & other authors (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454.

Li, F., Li, W., Farzan, M. & Harrison, S. C. (2005a). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868.

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K. & other authors (2005b). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* **24**, 1634–1643.

Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W. & other authors (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325.

Sturman, L. S. & Holmes, K. V. (1984). Proteolytic cleavage of peplomeric glycoprotein E2 of MHV yields two 90K subunits and activates cell fusion. *Adv Exp Med Biol* **173**, 25–35.

Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L. J., Wong, S. K., Moore, M. J., Tallarico, A. S., Olurinde, M. & other authors (2004). Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci U S A* **101**, 2536–2541.

Suzuki, A., Okamoto, M., Ohmi, A., Watanabe, O., Miyabayashi, S. & Nishimura, H. (2005). Detection of human coronavirus-NL63 in children in Japan. *Pediatr Infect Dis J* **24**, 645–646.

Tresnan, D. B. & Holmes, K. V. (1998). Feline aminopeptidase N is a receptor for all group I coronaviruses. *Adv Exp Med Biol* **440**, 69–75.

Tresnan, D. B., Levis, R. & Holmes, K. V. (1996). Feline aminopeptidase N serves as a receptor for feline, canine, porcine and human coronaviruses in serogroup I. *J Virol* **70**, 8669–8674.

Vabret, A., Mourez, T., Dina, J., van der Hoek, L., Gouarin, S., Petitjean, J., Brouard, J. & Freymuth, F. (2005). Human coronavirus NL63, France. *Emerg Infect Dis* **11**, 1225–1229.

van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-van Dillen, P. M., 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, M. F., Petersen, G., Forster, J., Berkhout, B. & Uberla, K. (2005). Croup is associated with the novel coronavirus NL63. *PLoS Med* **2**, e240.

van der Hoek, L., Pyrc, K. & Berkhout, B. (2006). Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol Rev* **30**, 760–773.

Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. (2004). A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* **279**, 3197–3201.

Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., Wong, B. H., Poon, R. W., Cai, J. J. & other authors (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* **79**, 884–895.

Yeager, C. L., Ashmun, R. A., Williams, R. K., Cardellichio, C. B., Shapiro, L. H., Look, A. T. & Holmes, K. V. (1992). Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420–422.

Zhang, C., Cui, Y., Houston, S. & Chang, L. J. (1996). Protective immunity to HIV-1 in SCID/beige mice reconstituted with peripheral blood lymphocytes of exposed but uninfected individuals. *Proc Natl Acad Sci U S A* **93**, 14720–14725.