Title: A structural view of coronavirus-receptor interactions

Author: Juan Reguera Gaurav Mudgal César Santiago José M. Casasnovas

PII: S0168-1702(14)00410-9
DOI: http://dx.doi.org/doi:10.1016/j.virusres.2014.10.005
Reference: VIRUS 96418

To appear in: *Virus Research*

Received date: 15-8-2014
Revised date: 6-10-2014
Accepted date: 6-10-2014

Please cite this article as: Reguera, J., Mudgal, G., Santiago, C., Casasnovas, J.M.,A structural view of coronavirus-receptor interactions, *Virus Research* (2014), http://dx.doi.org/10.1016/j.virusres.2014.10.005

A structural view of coronavirus-receptor interactions

Juan Reguera[1], Gaurav Mudgal[2], César Santiago and José M Casasnovas[2]*

[1]European Molecular Biology Laboratory, Grenoble Outstation, Grenoble Cedex 9, France

[2]Centro Nacional de Biotecnología (CNB-CSIC), Campus Universidad Autónoma, Darwin 3, 28049 Madrid, Spain

*Corresponding author: e-mail: jcasasnovas@cnb.csic.es

Tel: (+34) 91 5854917

Fax: (+34) 91 5854506

Abstract

In the coronavirus (CoV), the envelope spike (S) glycoprotein is responsible for CoV cell entry and host-to-host transmission. The S is a multifunctional glycoprotein that mediates both attachment of CoV particles to cell surface receptor molecules as well as membrane penetration by fusion. Receptor-binding domains (RBD) have been identified in the S of diverse CoV; they usually contain antigenic determinants targeted by antibodies that neutralize CoV infections. To penetrate host cells, the CoV can use various cell surface molecules, although they preferentially bind to ectoenzymes. Several crystal structures have determined the folding of CoV RBD and the mode by which they recognize cell entry receptors. Here we review the CoV-receptor complex structures reported to date, and highlight the distinct receptor recognition modes, common features, and key determinants of the binding specificity. Structural studies have established the basis for understanding receptor recognition diversity in CoV, its evolution and the adaptation of this virus family to different hosts. CoV responsible for recent outbreaks have extraordinary potential for cross-species transmission; their RBD bear large platforms specialized in recognition of receptors from different species, which facilitates host-to-host circulation and adaptation to man.

Key words: Coronavirus; virus entry; virus-receptor; virus neutralization; ectoenzymes; glycoproteins

Introduction

For productive entry into host cells, viruses attach to specific cell surface receptor molecules (Casasnovas, 2013; Marsh and Helenius, 2006). Selection of an entry receptor is governed by precise interactions that mediate efficient virus attachment to the cell surface as well as productive cell infection. Viruses can use a large number of cell surface molecules to penetrate host cells (Backovic and Rey, 2012); these molecules are the main determinants of virus tropism and pathogenesis. Receptor-binding motifs in viruses are subject to changes promoted by immune surveillance, which can target key receptor-binding residues during neutralization of virus infection. It is thus relatively common that a virus evolves to use distinct cell entry receptors over the course of an infection, or that related viruses use different cell surface molecules for host cell entry (Stehle and Casasnovas, 2009). This is the case of coronavirus (CoV), whose use of distinct entry receptor molecules is responsible for their broad host range and tissue tropism (Gallagher and Buchmeier, 2001; Masters, 2006). Some CoV have remarkable capacity for cross-species transmission which is linked to virus adaptation to the use of orthologous receptor molecules (Graham and Baric, 2010; Holmes, 2005).

The CoV are a large family of enveloped, positive single-stranded RNA viruses involved in respiratory, enteric, hepatic and neuronal infectious diseases in animals and in man. The CoV are subdivided into four genera, alpha, beta, gamma and delta (de Groot et al., 2011; de Groot et al., 2013). Prototype viruses in each genus are transmissible gastroenteritis virus (TGEV, alpha1-CoV), human coronaviruses (hCoV-229E and hCoV-NL63, alpha-CoV), mouse hepatitis virus (MHV, beta-CoV, lineage A), severe acute respiratory syndrome coronavirus (SARS-CoV, beta-CoV, lineage B), Middle East respiratory syndrome coronavirus (MERS-CoV, beta-CoV, lineage C), avian infectious bronchitis virus (IBV, gamma-CoV) and bulbul coronavirus (delta-CoV). The CoV have a major envelope

glycoprotein, the spike (S), which is responsible for CoV cell entry and interspecies transmission (Perlman and Netland, 2009). This glycoprotein mediates CoV particle attachment to cell surface molecules, as well as the fusion of virus and cell membranes (Masters, 2006). The S protein assembles into trimers, displayed as peplomers in the CoV envelope (Beniac et al., 2006); the protein has a membrane-distal globular N-terminal S1 portion and a stalk formed by the S2 region. The S1 region contains the receptor-binding determinants, whereas S2 mediates virus-cell fusion for membrane penetration (Fig. 1).

Like the class I fusion proteins, the S2 region adopts a helical structure, and is followed by the transmembrane domain (Bosch et al., 2003). S2 contains the fusion peptide and two conserved heptad repeat regions, HR1 (N-terminal) and HR2 (C-terminal) (Fig. 1), which form a coiled coil structure important for S trimerization and the fusion reaction during CoV cell entry (Supekar et al., 2004; Xu et al., 2004). The fusion peptide is N-terminal from the HR1 in the S2 sequence (Fig. 1), but the HR1-HR2 coiled coil structure places it close to the transmembrane region. As in other enveloped viruses, the initiation of the fusion reaction requires partial disassembly of the trimeric spikes and the exposure of the fusion peptide for binding to the host cell membrane (Belouzard et al., 2012; Beniac et al., 2007; Harrison, 2005). In some MHV variants and in the SARS-CoV, the S protein is processed into S1 and S2 fragments by cell proteases, which facilitate the fusion process and cell entry (Belouzard et al., 2012; Glowacka et al., 2011; Huang et al., 2006). The S of alpha-CoV is not processed. Receptor-mediated endocytosis and exposure to low pH is a necessary step for entry of TGEV, hCoV-229E and SARS-CoV (Masters, 2006). Other CoV, such as MHV and hCoV-NL63, do not require a low pH step for fusion, and the entry processes is mediated by receptor binding on the cell surface (Huang et al., 2006; Sturman et al., 1990). CoV can thus follow different entry pathways to penetrate host cells (Belouzard et al., 2012); receptor, low pH and proteases are three major inducers of membrane fusion, and CoV use them

differentially for cell entry. Mutations in the S1 and S2 fragments indicate that differences among CoV entry routes are probably related to variations in S trimer stability (Gallagher and Buchmeier, 2001). Nonetheless, the conformational changes in the CoV S that lead to membrane fusion and cell entry have not been defined.

The S1 region is largely variable in sequence and length, and is specialized in recognition of cell surface receptors (Fig. 1) (Li, 2012; Masters, 2006); it has several discrete modules or domains that can fold independently (Bonavia et al., 2003; Du et al., 2013; Godet et al., 1994; Li et al., 2005a; Reguera et al., 2011; Wu et al., 2009). Receptor-binding domains (RBD) can be located at the N- and/or C-terminal moieties of the S1 region (Li, 2012; Peng et al., 2011) (Fig. 1). The S glycoprotein N-terminal domain (NTD) can function as a RBD (N-RBD); it can be the only S1 domain engaged in receptor recognition or, in conjunction with C-terminal RBD, can broaden tissue tropism of certain CoV. As entry receptors, the N-RBD can recognize sialic acids in some cases (Fig. 1) (Peng et al., 2011), whereas it binds to carcinoembryonic antigen cell adhesion molecules (CEACAM) in MHV (Williams et al., 1991). The NTD in TGEV is responsible for its enteric tropism, absent in the related porcine respiratory CoV (PRCV) that lacks this domain (Sanchez et al., 1992). The NTD region adopts a galectin-like structure in two beta-CoV, and its fold might be conserved in alpha- and gamma-CoV, since glycan- binding activity has been reported for the three genera (Li, 2012; Schultze et al., 1996).

In most CoV, the major determinants of cell tropism are found in the C-terminal portion of the S1 region (Masters, 2006). These RBD can usually fold independently of the rest of the S, and can be expressed as a single domain with all receptor-binding determinants (Du et al., 2013; Reguera et al., 2011; Wong et al., 2004; Wu et al., 2009). Sequence and structure of the RBD vary considerably among CoV, and they recognize distinct receptors (Fig. 1). Several CoV of the genus alpha, including TGEV and hCoV-229E, use aminopeptidase N

(APN) for cell entry (Delmas et al., 1992; Yeager et al., 1992), whereas hCoV-NL63 binds to the human angiotensin-converting enzyme 2 (ACE2) (Wu et al., 2009). In the beta-CoV, the SARS- and the MERS-CoV use ACE2 and dipeptidyl peptidase 4 (DPP4, CD26) receptors, respectively (Li et al., 2003; Raj et al., 2013). APN, ACE2 and DPP4 are membrane-bound ectoenzymes with multiple functions such as angiogenesis, cell adhesion and blood pressure regulation (Boonacker and Van Noorden, 2003; Crackower et al., 2002; Mina-Osorio, 2008). The three proteins catalyze peptide-bond hydrolysis of short peptides. The reason for CoV use of ectoenzymes as entry receptors is unclear; it might be linked to their abundance on epithelial cells rather than on their peptidase function, which does not appear to be essential for CoV cell entry (Li et al., 2005c). Virus-binding regions in these ectoenzymes are distant from the catalytic site (Li et al., 2005a; Lu et al., 2013; Peng et al., 2011; Reguera et al., 2012; Wang et al., 2013; Wu et al., 2009).

The identification of the CoV entry receptors and the RBD in the S glycoprotein led to structural characterization of the CoV-receptor interaction. RBD-receptor complexes have been determined for prototype alpha- (TGEV and hCoV-NL63) and beta-CoV (MHV, SARS- and MERS-CoV). RBD regions are targets of antibodies (Ab) that neutralize CoV infection, and their epitopes overlap receptor-binding motifs (Godet et al., 1994; He et al., 2005; Hwang et al., 2006; Pak et al., 2009; Prabakaran et al., 2006; Reguera et al., 2012). Some structural studies have determined how neutralizing Ab prevent CoV cell entry and infection. In this review, we will summarize the currently determined CoV-receptor complex structures, highlighting the distinct receptor recognition modes in this virus family.

Alphacoronavirus recognition of cell entry receptors

The alphacoronavirus (alpha-CoV) genus is a group of important animal and human viruses subdivided into several lineages (de Groot et al., 2011). The alpha1 lineage comprises two types of canine (cCoV and cCoV-NTU336) and feline (fCoV and FIPV) CoV,

PRCV and TGEV; another lineage includes human CoV hCoV-229E and hCoV-NL63, and other members of the genus alpha are porcine epidemic diarrhea virus (PEDV) and some bat CoV.

TGEV, one of the most studied alpha-CoV, has enteric and respiratory tropism. The enteric tropism is linked to its NTD, since a deletion mutant of TGEV (the homologous PRCV) shows only respiratory tropism (Sanchez et al., 1992). NTD binding to an attachment factor (sialic acid) is thought to be responsible for its enteric tropism (Schultze et al., 1996). TGEV, PRCV and the related animal alpha1-CoV use APN for host cell entry (Fig. 1). APN is also the receptor for hCoV-229E (Delmas et al., 1992; Yeager et al., 1992), one of the first human CoV discovered, which is responsible for common colds (Kahn and McIntosh, 2005). The related hCoV-NL63 does not bind to APN and recognizes the cell surface ACE2 ectoenzyme (Fig. 1) (Smith et al., 2006), like the SARS-CoV (Li et al., 2003). The cell surface receptor of PEDV and other alpha-CoV are currently unknown.

The RBD in alpha-CoV

The alpha-CoV RBD are modules of ~150 residues that locate near the C-terminal portion of the S1 region (Fig. 1) (Breslin et al., 2003; Godet et al., 1994; Wu et al., 2009). The RBD can be expressed independently of the S; binding studies with receptors and Ab show that the RBD preserves its native conformation and binding specificity (Reguera et al., 2011; Wu et al., 2009). Preparation of single RBD proteins facilitates their crystallization in complex with receptors and Ab.

The crystal structures of hCoV-NL63, PRCV and TGEV RBD have been determined (Reguera et al., 2012; Wu et al., 2009). They show a single domain unit that has a β-barrel fold with two highly twisted β-sheets (Fig. 2). In one β-sheet, three β-strands (β1, β3 and β7) run parallel (Fig. 2A). The three RBD have three disulphide bonds. In the crystal structure of the TGEV RBD, solved at high resolution, the bent β-strand 5 (β5) crosses both β-sheets

(Fig. 2A). N-linked glycans cluster at one side of the β-barrel; the opposite side is not glycosylated and might be closer to other S protein domains. N- and C-terminal ends of the RBD are located on the same side of the domain (terminal side); at the opposite side, two β-turns form the tip of the barrel in the TGEV RBD (Fig. 2A). This region of the β-barrel domain contacts the receptor (see below) and its conformation in the APN-binding RBD of TGEV and PRCV differs from the ACE2-binding region in the hCoV-NL63 domain (Fig. 2B, 2C). These differences probably determine the distinct receptor-binding specificities of alpha-CoV. The TGEV or PRCV RBD tips are formed by two protruding β-turns (β1-β2 and β3-β4), each bearing a solvent-exposed aromatic residue (tyrosine or tryptophan) (Fig. 2A, 2B). In contrast, the hCoV-NL63 RBD tip has a slightly recessed conformation, with the aromatic residues at the center of the receptor-binding surface (Fig. 2C).

Alpha-CoV recognition of APN and ACE2 receptors

Crystal structures have been reported for complexes of alpha-CoV RBD with the APN and ACE2 ectodomains (Reguera et al., 2012; Wu et al., 2009). The RBD of these viruses contact receptor regions distal to the cell membrane (Fig. 3).

The APN ectodomain is composed of four domains (DI-DIV), is heavily glycosylated and forms dimers through extensive DIV-DIV interactions (Fig. 3A). Each APN monomer has an RBD bound in the crystal structure of the PRCV RBD-APN complex (Fig. 3A). The bidentate, protruding tip contacts the APN, and the exposed side chains of the tyrosine and tryptophan residues penetrate small cavities of the APN ectodomain. The tyrosine side chain fits between an α-helix and a carbohydrate N-linked to the APN, whereas the bulky tryptophan is in a narrow cavity formed at the DII-DIV junction (Fig. 3A). In addition to the tyrosine, other RBD residues contact the first N-acetyl glucosamine (NAG) linked to the porcine APN Asn736, and fix the glycan conformation. The CoV tyrosine and tryptophan residues are critical for TGEV RBD binding to the APN (Reguera et al., 2012), and

183 preliminary results indicated that they are essential for virus entry and infection (unpublished
184 data). CoV recognition of APN is species-specific, and specificity is linked to the APN N-
185 linked glycan that interact with the RBD β1-β2 turn in the structure (Reguera et al., 2012;
186 Tusell et al., 2007). Porcine, feline and canine alpha-CoV with a tyrosine at the β1- β2 turn
187 recognize APN proteins bearing the glycan. The large degree of sequence conservation in the
188 RBD tip of alpha1-CoV also suggests a highly conserved APN recognition mode (Reguera et
189 al., 2012). hCoV-229E does not have a tyrosine in its RBD β1- β2 turn, however, and it
190 recognizes the human APN that lacks this glycosylation (Reguera et al., 2012; Tusell et al.,
191 2007). The conformation of this alpha-CoV RBD tip differs from that of alpha1-CoV,
192 suggesting that hCoV-229E recognition of APN must be unique. It is nonetheless likely that
193 this human alpha-CoV preserves a protruding tip for binding to small APN cavities.
194 hCoV-NL63 RBD interacts with the ACE2 ectodomain opposite to the way that the alpha-
195 CoV bind to APN. The hCoV-NL63 RBD has a blunt tip that contacts protruding regions of
196 the receptor (Fig. 3B). In the middle of the interacting surface, the depressed center of the
197 RBD tip contacts a unique receptor β-turn (β4–β5), which interacts with a tyrosine and a
198 tryptophan in the virus protein (Fig. 3B). The rims of the RBD tip bind to two α-helices of
199 the ACE2 receptor. Specificity is determined by several hydrogen bonds that engage amino
200 and carbonyl groups in the main chains of the interacting molecules (Fig. 3B).
201 Alpha-CoV use protruding RBD regions to bind APN or recessed surfaces to recognize
202 exposed ACE2 motifs (Fig. 2, 3). Crystal structures demonstrate that the conformation of the
203 receptor-binding region in the alpha-CoV S must be the principal determinant of its receptor
204 recognition specificity. We recently demonstrated that the RBD tip is a principal antigenic
205 determinant (site A) in the S of TGEV and related alpha-CoV (Reguera et al., 2012). Potent
206 neutralizing Ab of porcine CoV cluster at site A (Delmas et al., 1990; Sune et al., 1990).
207 These Ab recognize the RBD tip and bind to the tyrosine or the tryptophan essential for APN

binding (Reguera et al., 2012). These data suggest that the conformation of the alpha-CoV receptor-binding region evolved under pressure from the immune system, particularly in humans, leading to small variations in the way hCoV-229E recognizes the APN protein (Tusell et al., 2007) or to radical changes that modified receptor specificity in hCoV-NL63.

Betacoronavirus recognition of cell entry receptors

The betacoronavirus (beta-CoV) genus comprises four lineages, A (MHV, hCoV-HKU1 and the beta1-CoV), B (SARS-CoV), C (batCoV and MERS-CoV), and D (batCoV HKU9) (de Groot et al., 2011). The most representative CoV prototypes of this genus are hCoV-OC43 (beta1-CoV), MHV, SARS-CoV and the recently identified MERS-CoV. Members of lineage A CoV incorporate an extra, short spike-like glycoprotein in their envelope, the hemagglutinin esterase (HE) (Masters, 2006; Qinghong et al., 2008).

hCoV-OC43 causes common cold and pneumonia in elderly populations, as well as severe lower respiratory tract infection in immunocompromised patients (Kahn and McIntosh, 2005). Like bovine CoV (bCoV), another beta1-CoV, it uses sialic acids (N-acetyl-9-O-acetylneuraminic acid, Neu5,9Ac2) as entry receptors (Fig. 1) (Krempl et al., 1995). Before SARS, MHV was the most studied beta-CoV *in vitro* and *in vivo*, especially in laboratory mouse. MHV strains cause specific inflammations in several mouse organs, such as the neurotropic strains JHM and A59 responsible for acute encephalitis and chronic demyelination in survivors, which serve as a model for the study of multiple sclerosis (Weiss and Leibowitz, 2011). The MHV cell entry receptor is a member of the CEACAM family (Williams et al., 1991).

The SARS-CoV brought coronavirology to the center of the research community's attention due to a worldwide epidemic with very high mortality rates (Gallagher and Perlman, 2013). It uses ACE2 as the entry receptor (Li et al., 2003). Epidemiologists believe that SARS virus originated in bats (natural reservoir), was then transmitted to palm civets, ferret

badgers, and raccoon dogs (amplification and transmission hosts) and then introduced into man (Li et al., 2005b). SARS-CoV adaptation to different species and its transmission to humans is linked to subtle changes in the S glycoprotein, which increased its binding affinity for human ACE2 (Li et al., 2005c).

MERS-CoV emerged in Saudi Arabia a decade after the SARS epidemic. It shares 90% sequence identity with batCoV-HKU4 and -HKU5, and it docks in beta-CoV lineage C (de Groot et al., 2013). Given this relationship, it is likely that MERS-CoV originated from bats (Raj et al., 2014). This virus uses DPP4 as a cell entry receptor (Raj et al., 2013). BatCoV-HKU4 recognizes the human DPP4 protein, indicating possible direct transmission from bats to humans (Wang et al., 2014; Yang et al., 2014). Recent evidence nonetheless shows involvement of dromedary camels as intermediates in virus transmission from bats to man (Doremalen et al., 2014; Haagmans et al., 2014). Human-to-human transmission is not frequent, probably because of low DPP4 expression in the human lower respiratory tract (Raj et al., 2014).

Receptor recognition by the SARS-CoV

Several crystal structures show the folding of the SARS-CoV RBD, the mode by which this virus recognizes its ACE2 entry receptor, and how Ab prevent virus binding to the receptor. These studies led to improved understanding of host-host transmission and adaptation of this CoV to humans, and also indicated strategies used by the SARS-CoV to evade neutralization by the immune system.

The SARS-CoV RBD

The SARS-CoV RBD is defined as a ~200-residue fragment in the C-terminal portion of the S1 region (Fig. 1)(Wong et al., 2004). It is composed of two subdomains; the core has a central five-stranded β-sheet surrounded by polypeptides that connect the β-strands (Fig. 4A, yellow). It has three small α-helices (A to C) and three disulphide bridges. A second

subdomain of ~65 residues inserts between two central β-strands of the core (β4 and β7), and is distal to the terminal side of the domain (Fig. 4A, dark-red). This inserted subdomain lies on one side of the core and comprises a central two-stranded β-sheet connected by a long loop region; one side of this loop and the β-sheet clamp the core. The β-sheet, the extensive interactions with the core, and a disulphide bond in the most solvent-exposed region of the subdomain stabilize its structure (Fig. 4A). One crystal structure of the isolated SARS-CoV RBD shows that it can form dimers through the terminal side (Hwang et al., 2006). The dimerization surface in these crystals is relatively large (~1000 $Å^2$ buried surface area, BSA/monomer) and the authors proposed that RBD dimers could crosslink S glycoprotein trimers. It is nonetheless unclear whether such oligomers are found on the virus envelope and could recognize ACE2.

SARS-CoV binding to ACE2

The ACE2 ectoenzyme is the cell entry receptor of SARS-CoV (Li et al., 2003). It is a type I membrane glycoprotein with an N-terminal extracellular domain built of two α-helical lobes; the catalytic site with a coordinated zinc ion is located between the two lobes (Fig. 3B, 4B). The ACE2 ectodomain shows some conformational movement, and substrate binding to the active site leads to a closed conformation (Towler et al., 2004). Drug binding to this active site does not affect SARS-CoV binding, in accordance with virus recognition of a single lobe (Li et al., 2005c) (Fig. 4B).

The SARS-CoV RBD inserted subdomain is the main S glycoprotein receptor-binding motif (Li et al., 2005a) (Fig. 4); the ACE2-binding subdomain region forms a curved, elongated surface with the two-stranded β-sheet at the bottom (Fig. 4A). The interaction buries 25 residues and about 860 $Å^2$ of the virus protein, and a similar surface (820 $Å^2$) of the ACE2 receptor. The ACE2-interactive surface of the SARS-CoV RBD is ~100 $Å^2$ larger that of hCoV-NL63, consistent with marked differences in kinetic dissociation rate constants,

283 which is an order of magnitude lower in SARS than in hCoV-NL63 (Li et al., 2005c; Wu et
284 al., 2009). Both viruses recognize overlapping ACE2 regions, including the N-terminal α-
285 helix (α1) and the β-turn formed by β4 and β5 strands (Fig. 3B, 4B). The central concave
286 SARS-CoV RBD surface cradles the ACE2 N-terminal α-helix, whereas the terminal side of
287 the subdomain interacts with the ACE2 β4-β5 turn and α10 (Fig. 4B, 4C). The interaction
288 includes at least 10 virus-receptor hydrophilic bonds, some of which engage the hydroxyl
289 groups of RBD tyrosines that also mediate non-polar interactions with the receptor (Fig. 4C).
290 There is an important virus-receptor hydrogen bond interaction between the ACE2 Lys353
291 carbonyl and the main chain amino group of RBD Gly488 (Fig. 4C) (Li et al., 2005a). The
292 lysine side chain amino interacts with RBD main chain carbonyl. This ACE2 lysine is absent
293 in mouse and rat ACE2 proteins, which are not recognized by the SARS-CoV. ACE2
294 glycosylation is also a determinant of SARS-CoV species specificity (Li et al., 2005c). A
295 glycan linked to rat ACE2 Asn82 prevents its use as an efficient virus receptor. Deletion of
296 the glycan and the His353/Lys substitution convert rat ACE2 into a SARS-CoV receptor,
297 showing that efficient ACE2 recognition is central to virus infection and host-to-host
298 transmission (Holmes, 2005; Li et al., 2005a; Li et al., 2005c).

299 SARS-CoV emerged from bat CoV and was transmitted through palm civet CoV; cross-
300 species transmission is linked to RBD changes that increased its affinity for human ACE2
301 (Holmes, 2005; Li, 2013; Li et al., 2005a; Li et al., 2005c). Of the residues involved in
302 SARS-CoV RBD binding to ACE2, only a few have a key role in SARS-CoV adaptation to
303 man (Fig. 4C). Lys479/Asn and Ser487/Thr mutations are two key changes in the SARS-
304 CoV S glycoprotein for infection of human cells. Substitutions in one of these residues
305 increases SARS-CoV RBD binding affinity to human ACE2 by 20- to 30-fold, whereas the
306 double mutation has a synergistic effect, with a 1000-fold increase in interaction affinity (Li
307 et al., 2005c). The Asn at position 479 is found in some civet CoV; it does not affect binding

to civet ACE2, but increases SARS-CoV RBD affinity for the human protein (Li et al., 2005c). Asn479 contacts the human ACE2 His34 and is relatively close to Lys31 in the N-terminal α-helix (Fig. 4C), which are Tyr34 and Thr31 in civet ACE2. The presence of a positively charged lysine rather than Asn in RBD position 479 does not complement the human ACE2 Lys31 and His34 residues. The crystal structure of SARS-CoV RBD in complex with human ACE2 demonstrates that the methyl group of the threonine at position 487 establishes specific contacts with the ACE2 Tyr41 and Lys535 side chains, increasing affinity for the human receptor (Fig. 4C) (Li et al., 2005a). The SARS-CoV that caused sporadic outbreaks in 2003-2004 has serine at position 487 and shows very poor human-to-human transmission. This phenotype was also associated with the Leu472/Pro substitution in the ACE2 contact region of the SARS-CoV RBD (Li et al., 2005a). Other RBD residues have some influence on cross-species transmission of SARS-CoV (Li, 2013).

Structural basis of SARS CoV neutralization by antibodies

The RBD is a major antigenic determinant in the S glycoprotein of the SARS-CoV (Du et al., 2009). Potent human and mouse SARS-CoV neutralizing Ab target the RBD and prevent virus infection by blocking its binding to the ACE2 receptor (He et al., 2005; Zhu et al., 2007). The RBD can elicit broadly neutralizing Ab against diverse isolates, and human monoclonal Ab (mAb) can protect from infection by various zoonotic and human SARS-CoV (He et al., 2006; Zhu et al., 2007). Several conformational epitopes (I-VI) have been defined in the RBD, some of which are conserved in different species (He et al., 2006). Epitopes of several neutralizing Ab have been identified by crystal structures of RBD-Ab complexes (Hwang et al., 2006; Pak et al., 2009; Prabakaran et al., 2006), which show that they overlap with the receptor-binding region (Fig. 5).

Neutralizing Ab bind to the RBD external subdomain that contacts ACE2 (Fig. 5). The human mAb m396 is a potent neutralizing Ab of several zoonotic and human SARS-CoV

(Zhu et al., 2007); it targets a region in the C-terminal side of the RBD inserted subdomain (residues 482-491) that is involved in ACE2 recognition, as well as residues in the RBD core (Fig. 5) (Prabakaran et al., 2006). The mAb epitope includes RBD residues Ile489 and Tyr491, which contact the receptor directly. A very similar epitope was described for the mouse mAb F26G19 (Fig. 5) (Pak et al., 2009), which contacts residues 486 to 492 of the RBD inserted subdomain and some regions of the core. Ile489 is a central residue in the F26G1 epitope (Fig. 5, black). Epitopes of mAb m396 and F26G19 are thus very similar, and include an exposed ridge in the RBD ACE2-binding region (Fig. 5); this S region must be a hot spot for SARS-CoV neutralization.

The crystal structure of the human R80 mAb shows a distinct mode of SARS-CoV neutralization that also prevents virus binding to ACE2 (Fig. 5) (Hwang et al., 2006). The R80 variable domains make extensive contact with the concave region of the RBD-inserted subdomain (Fig, 5), mimicking the way that RBD and ACE2 interact. The R80 epitope in the RBD overlaps with the region buried by the N-terminal α-helix of the receptor. The total surface buried by the R80-RBD interaction is larger than the ACE2-RBD surface and is responsible for its high affinity (in the nanomolar range). This mAb makes contact with 29 residues of the receptor-binding subdomain, 17 of which are involved in ACE2 recognition (Hwang et al., 2006).

All three SARS-CoV-neutralizing mAb epitopes overlap with the receptor-binding region in the S protein (Fig. 5); efficient virus neutralization is thus achieved by targeting receptor-binding residues and blocking virus binding to ACE2 and thus, cell entry. Virus mutants have been identified that escape mAb neutralization, although these mutants usually cause attenuated infection (Rockx et al., 2010); some of the escape mutations map to the RBD inserted subdomain (Fig. 5) and probably affect SARS-CoV binding to ACE2.

Receptor recognition by the MERS-CoV

MERS-CoV arose recently as a highly pathogenic virus in humans (Coleman and Frieman, 2014); it is thought to have emerged from bats and is transmitted to humans via dromedary camels. Cross-species transmission is determined mainly by the adaptability of this CoV for different hosts, mediated by subtle modifications in its envelope S protein. MERS and SARS-CoV RBD are structurally similar (Fig. 6), but use different cell entry receptors; MERS-CoV attach to a distinct ectoenzyme, DPP4 (Raj et al., 2013). Several crystal structures have defined MERS-CoV RBD and how it binds to its DPP4 receptor (Chen et al., 2013; Lu et al., 2013; Wang et al., 2013).

The MERS-CoV RBD

The MERS-CoV RBD is a fragment in the S1 region C-terminal portion (Fig. 1); its structure is remarkably similar to the SARS-CoV RBD (Fig. 6) (rmsd of 2.4 Å for 132 residues), although they show little sequence identity. The MERS-CoV RBD also has two subdomains (Fig. 6A), the core with a central five-stranded β-sheet and three disulphide bridges, as well as an inserted or external subdomain between two core β-strands (Chen et al., 2013; Lu et al., 2013; Wang et al., 2013). The central β-sheet of the core is surrounded by polypeptides that connect the β-strands and contain helical structures (Fig. 6A). The core has an overall globular shape. The inserted subdomain is distal from the RBD terminal side and has a four-stranded β-sheet (Fig. 6A). The β-sheet and a long loop that connects the β-strands at one edges of the sheet clamp the core subdomain, as in the SARS-CoV RBD (Fig. 4A). The cores are more similar in MERS- and SARS-CoV than the external subdomain (Fig. 6B), which is longer in the MERS (80 residues) than the SARS RBD (65 residues). Because of the extended β-sheet, the solvent-exposed region of the inserted subdomain is broader than that of SARS-CoV. The first (β6) and last (β9) β-strands of the MERS-CoV inserted subdomain align with the two β-strands of the SARS-CoV inserted subdomain, but the other two β-strands (β7 and β8) are absent in the SARS RBD (Fig. 6B). The MERS-CoV inserted

subdomain contains a concave surface or small "canyon" formed by the β-strands and the loop that connect β6 and β7 in the inserted RBD subdomain (Fig. 6A). This "canyon" is very distant from the terminal side and exposed for receptor recognition. It is absent in the SARS-CoV RBD, which contains a long loop in this location (Fig. 6B). Likely, these differences in the external subdomains are the major determinants of the distinct receptor-binding specificity between the MERS- and SARS-CoV.

MERS-CoV binding to its DPP4 receptor

DPP4 or CD26 is a multifunctional membrane-bound serine protease (Boonacker and Van Noorden, 2003). DPP4 is a type II membrane protein that forms homodimers on the surface of different cells (Fig. 7A). The DPP4 ectodomain has ~730 amino acids and is composed of two domains, an α/β-hydrolase domain and an eight-bladed β-propeller (Fig. 7A). The substrates bind to a pocket in a central cavity formed between the two domains (Boonacker and Van Noorden, 2003). The MERS-CoV contacts only the β-propeller domain (Fig. 7A, green).

Crystal structures of the MERS-CoV RBD bound to DPP4 demonstrate that the virus attaches to the most membrane-distal region of the β-propeller (Lu et al., 2013; Wang et al., 2013). One RBD binds to each of the DPP4 monomers in the dimer, away from the receptor dimerization interface (Fig. 7A). This dimeric virus-receptor complex is similar to the alpha-CoV RBD-APN structure described above (Fig. 3A). The bound RBD does not appear to interfere with DPP4 catalytic activity, binds only to the β-propeller subdomain and away from the regions at which the substrate accesses the active site.

The MERS-CoV RBD engages the DPP4 molecule through the solvent-exposed side of its external subdomain (Fig. 6A, 7A). It contacts the edges of DPP4 β-propeller blades IV and V, including N-linked carbohydrates at blade IV and a helix at the linker between the two blades (Fig. 7B). It is the largest CoV-receptor interface, and buries 32 residues of the RBD

(~1110 Å^2 surface) and of DPP4 (~1240 Å^2 surface). In the two structures reported (PDB ID 4KRO and 4L72) (Lu et al., 2013; Wang et al., 2013), the interaction includes between 9 to 14 hydrogen bonds and 2 to 3 salt bridges. A key spot in this virus-receptor interaction includes the RBD contact with the helix that bulges out at the N-terminus of blade V (Fig. 7B). The small "canyon" in the inserted subdomain cradles the DPP4 helix. This helix contains mostly hydrophobic residues (Ala291, Leu294, Ile295) that lie on a hydrophobic patch in the RBD "canyon", composed of the side chains of Lys502, Leu506, Tyr540, Arg542, Trp553 and Val555, residues located in the three main β-strands of the subdomain (Fig. 7B). The side chain amino groups of Lys504 and Arg542 are hydrogen-bonded to the main chain of DPP4. The loop at one rim of the small "canyon" forms polar interactions with the DPP4 β-strands in blade V (Fig. 7B).

An interesting feature of MERS-CoV binding to DPP4, also shown in the PRCV-APN complex (Fig. 3A, bottom), is the RBD interaction with N-linked receptor carbohydrates (Fig. 7B). The first three carbohydrates attached to DPP4 Asn229 are well defined in the crystal structures of the MERS-CoV RBD-DPP4 complex (Lu et al., 2013; Wang et al., 2013). They interact with several solvent-exposed residues in the virus protein (Fig. 7B). The first NAG residue is hydrogen-bonded to RBD Glu536, whereas the second NAG of the glycan stacks onto the aromatic ring of viral Trp535, which strengthens the glycan-virus interaction and probably stabilizes motif conformation. The third mannose residue in the DPP4 N-linked glycan also interacts with the RBD tryptophan. Another glycan at DPP4 Asn281 in blade IV is very close to the RBD (not shown), but does not interact with the virus protein. The conformation of this last glycan appears to be determined by its interaction with a tryptophan residue (Trp187) in the DPP4 protein (Lu et al., 2013; Wang et al., 2013), and could be critical for MERS-CoV RBD binding to DPP4. A highly flexible glycan in this

position could prevent this virus-receptor interaction, such as shown for glycosylation in the APN or ACE2.

The S glycoprotein N-terminal domain (NTD) in CoV receptor recognition

Folding of the NTD

The S glycoprotein NTD can mediate attachment of CoV particles to cell surface molecules (Peng et al., 2011; Peng et al., 2012; Schultze et al., 1996; Tsai et al., 2003). It thus function as an RBD (N-RBD) in certain CoV. The crystal structure of the MHV NTD shows a galectin-like fold (Fig. 8) (Peng et al., 2011); the homologous bCoV NTD also has a galectin fold (Peng et al., 2012). The galectins are a family of lectins with a common β-sandwich carbohydrate recognition domain (CRD) (Fig. 8A). They preferentially recognize N-acetyl lactosamine in cell surface proteins, which binds to conserved residues on one of the CRD β-sheets (Fig. 8A). The CoV NTD is also composed of a central β-sandwich formed by two long β-sheets with six and seven β-strands that is structurally similar to the galectin CRD (Fig. 8B).

The CoV is thought to incorporate this N-terminal galectin-like domain from the host (Peng et al., 2012). In several CoV such as TGEV, beta1-CoV and IBV, the NTD preserves glycan binding activity, whereas in MHV it binds to a protein receptor, CEACAM1 (Peng et al., 2012; Tsai et al., 2003). The CoV NTD has diverged from galectins and recognizes proteins or sialic acids rather than N-acetyl lactosamine; the mode of ligand recognition also differs (Peng et al., 2012). Although the side of the NTD that recognizes cell surface molecules is the same side as the galectin CRD that binds carbohydrates, the top of the carbohydrate-binding β-sheet is covered by polypeptides that shape the receptor-binding region in CoV (Fig. 8, 9A). In addition, a glycan N-linked to one edge of the β-sheet further prevents ligand binding to the carbohydrate-binding sheet in galectins. This region is similar

in MHV, which binds CEACAM1, and in bCoV, which binds sialic acid, showing that the NTD has evolved in CoV to specifically select cell entry receptors.

MHV binding to the CEACAM1 receptor

MHV is a prototype beta-CoV of the A lineage. It uses CEACAM receptors to enter host cells (Williams et al., 1991). CEACAM are type I membrane proteins of the immunoglobulin superfamily (IgSF), markers of colorectal tumors that contribute to tumorigenesis (Beauchemin et al., 1999); in contrast to other CoV receptor proteins, they are not peptidases. The CEACAM mediate homo- and heterophilic cell adhesion. There are two murine CEACAM genes, *CEACAM1* and *CEACAM2*. *CEACAM1* has four splice forms, which have two (D1, D4) or four (D1-D4) Ig-like domains in the extracellular region, as well as a transmembrane region and two distinct cytoplasmic tails (Beauchemin et al., 1999). All four CEACAM1 variants can be used as receptors by MHV (Dveksler et al., 1993). CEACAM1 is also a receptor for virulent Neisseria strains (Virji et al., 1999).

CEACAM1 is a member of the IgSF, and the MHV S protein recognizes the N-terminal Ig-like domain 1 (D1), which adopts a variable (V) fold (Tan et al., 2002). The virus interacts with the CFG β-sheet of D1 (Fig. 9B), the surface commonly engaged in intermolecular interactions by cell surface molecules of the IgSF. The CFG β-sheet is formed by the β-strands C, C′, C′′ on one side and the β-strands F and G on the other (Fig. 9B). About 25 receptor residues, 770 $Å^2$ of its surface, are buried by the MHV protein. Most of the virus-binding residues locate at the D1 C′′ edge and around the FG loop. CEACAM1 has a unique CC′ loop that protrudes from the CFG β-sheet of the Ig-like domain (Tan et al., 2002). This is a key structural determinant for CEACAM1 recognition by the MHV S protein (Peng et al., 2011).

The CEACAM1-binding surface is on top of the galectin-like β-sandwich in the MHV N-RBD (Fig. 9A). The N-terminal portion of the MHV N-RBD structure occupies the top of

481 the receptor-binding surface and contributes 50% of the 24 MHV residues buried by
482 interaction with the receptor. The N-terminal residues form a "socket" that contains a
483 hydrophobic amino acid, Leu160, at the bottom (Fig. 9A). Ile41 of CEACAM1 is exposed in
484 the D1 CC´ loop and penetrates the socket (Fig. 9B). MHV Tyr15, Leu89 and Leu160
485 contact the Ile41 side chain (Fig. 9C), and comprise a critical virus-receptor motif (Peng et
486 al., 2011; Tan et al., 2002). Surrounding residues in the CEACAM1 CC´ loop, Thr39 and
487 Asp42, form hydrogen bonds with the MHV N-RBD (Fig. 9C), which confirms the
488 importance of this receptor region in virus recognition.

489 The N-terminal portion of the MHV N-RBD also contacts other motifs in the C´´ edge of
490 D1. In the C´ β-strand, CEACAM1 Arg47 contributes to binding and establishes hydrogen
491 bonds with the main chain carbonyl oxygens of MHV N-terminal residues. Up to 10 polar
492 virus-receptor interactions contribute to virus-receptor specificity. MHV N-terminal residues
493 interact extensively with the receptor C´´ β-strand, which runs parallel to the β1-strand of the
494 virus domain. Phe56 in the C´´ β-strand appears to be an important residue for the interaction
495 and establishes van der Waals contacts with the virus protein. Another important receptor-
496 binding motif surrounds MHV Leu174 and contacts the loops at the top of the CFG β-sheet
497 (Fig. 9B). This N-RBD region protrudes slightly and is distant from the socket.

498 The crystal structure of the MHV NTD in complex with CEACAM1 shows how the
499 N-terminal module of a CoV S recognizes a protein receptor. This region has been
500 implicated in the recognition of sialic acids in alpha- (TGEV), beta- (bCoV) and gamma-
501 (IBV) CoV (Fig. 1). The NTD of these CoV were proposed have a similar fold, which was
502 confirmed by the crystal structure of the bCoV NTD (Peng et al., 2012). As in the MHV
503 structure (Fig. 8), the bCoV NTD has polypeptides on the top of the galectin-like β-sandwich.
504 The bCoV NTD structure nonetheless lacks the MHV NTD socket, a critical motif for
505 CEACAM1 binding. Differences in the conformation of exposed NTD regions could be

506 responsible for the distinct receptor-binding specificity observed among CoV that use the
507 N-terminal module to bind to cell surface molecules.
508

Discussion

The structural studies reviewed here established the basis for understanding receptor recognition diversity in CoV, its evolution and its adaptation to different hosts. CoV RBD folding, conformation of receptor-binding motifs and subtle changes in those motifs determine receptor binding specificity and CoV host range. Two domains of the multifunctional CoV S glycoprotein anchor the virus particles to cell surface molecules for virus penetration of cells (Fig. 1). The two domains might be exposed in the S1 region for CoV binding to host cell entry receptors (Fig. 10).

The S glycoprotein NTD can function as an RBD in certain CoV (Fig. 1), and might have a conserved fold in alpha-, beta- and gamma-CoV (Peng et al., 2012). This domain has a galectin-like core, which indicates it was incorporated into the CoV S from a host (Li, 2012; Peng et al., 2011). It has evolved in some CoV to recognize cell surface molecules such as sialic acids, or in MHV to bind the CEACAM1 protein (Fig. 1). CoV NTD has integrated polypeptides and an N-linked glycan on the top of the flat galectin-like β-sandwich, which covers the galactose-binding β-sheet in galectins (Fig. 8). The virus-specific conformation of the polypeptides at the top of the NTD probably determine its receptor-binding specificity (Peng et al., 2012). The MHV NTD contains a socket for specific recognition of a unique structural feature in the CEACAM1 D1 (Fig. 9). Acquisition of the galectin-like NTD from the host probably expanded CoV host cell tropism, as shown for the TGEV NTD that confers enteric tropism (Schultze et al., 1996), although MHV and related beta-CoV only use the NTD for recognition of cell surface proteins (Fig. 1). The receptor-binding function of the S1 C-terminal portion appears to have been lost in these CoV. It would be interesting to explore the conformation of this region, which could provide clues to its presumed lack of function.

The S1 C-terminal RBD have unique structures unrelated to host proteins (Chen et al., 2013; Li et al., 2005a; Peng et al., 2011; Reguera et al., 2012; Wu et al., 2009) and can thus

be considered genuine CoV RBD. Alpha- and beta-CoV RBD adopt two distinct folds, but bind only to ectoenzymes; the NL63- and SARS-CoV bind to the same protein, ACE2 (Fig. 1). Some of these enzyme features must be essential for CoV entry into host cells. Perhaps they cluster with other proteases that facilitate fusion (Glowacka et al., 2011). ACE2, APN and DPP4 have distinct structures and functions, but their ectodomains share an inherent conformational flexibility (Boonacker and Van Noorden, 2003; Towler et al., 2004; Xu et al., 1997) that could assist in dissociation of the S1-S2 heterotrimer. Trimeric spikes that bind simultaneously to several receptor molecules could disassemble by pulling forces generated during ectodomain movement. The conformation and dynamics of the APN ectodomain vary with the pH (unpublished data), so that endosomal acidification can alter APN conformation during receptor-mediated endocytosis.

Alpha-CoV RBD adopt a conserved β-barrel fold (Fig. 2) (Reguera et al., 2012; Wu et al., 2009). S1 C-terminal fragments of the IBV gamma-CoV and the bulbul delta-CoV share certain sequence similarity with the alpha-CoV RBD, and could have a similar fold (Reguera et al., 2012). Crystal structures of alpha-CoV in complex with receptors identified the receptor-binding region in the RBD (Reguera et al., 2012; Wu et al., 2009), which has remarkable structural variability (Fig. 2). The conformation of the RBD tip dictates the receptor molecule used by alpha-CoV for host cell entry. RBD with protruding tips determine alpha-CoV attachment to APN, whereas those with blunt RBD tips recognize ACE2 and perhaps other yet uncharacterized receptor molecules. Structures of alpha-CoV RBD in complex with APN or ACE2 show two opposite modes of CoV-receptor recognition (Reguera et al., 2012; Wu et al., 2009) (Fig. 3). In viruses, recessed surfaces hide conserved receptor-binding residues from antibodies (Casasnovas, 2013; Rossmann, 1989); hCoV-NL63 uses a recessed surface to recognize exposed ACE2 motifs, following a receptor-binding strategy similar to the other beta-CoV reviewed here. CoV binding to APN is unique

among CoV, and contrasts with the mode of ACE2 and DPP4 recognition. The bidentate, protruding RBD tip of alpha1-CoV, which has two exposed aromatic residues, penetrates small cavities of the APN ectodomain (Fig. 3A). Similarly to MERS-CoV, the alpha1-CoV also recognize a dimeric cell surface protein.

The folding of the MERS- and SARS-CoV RBD are similar (Fig. 6) (Chen et al., 2013; Li et al., 2005a; Lu et al., 2013; Wang et al., 2013). Both have a core with a single β-sheet and an additional subdomain that recognizes cell entry molecules. MERS- and SARS-CoV show extraordinary potential for cross-species transmission, related to S binding to distinct orthologous receptor molecules. This is probably linked to the specific structure of their RBD, especially to the extended receptor-binding surfaces of the inserted subdomains (Fig. 4A, 6A). A few changes in those large surfaces increase affinity for receptor molecules in new hosts, while preserving virus growth (Holmes, 2005; Li et al., 2005c). Measles virus (MV) follows a similar strategy for recognition of several receptors that facilitate virus growth and transmission (Casasnovas, 2013). The MV hemagglutinin uses a broad concave surface to bind to three distinct receptor molecules, a unique feature of MV among the paramyxoviruses. The use of a large receptor recognition surface probably enables virus dissemination in tissues and host-to-host virus transmission.

The DPP4-binding surface in the MERS-CoV is larger (~300 $Å^2$) than the ACE2-binding surface in SARS-CoV, which correlates with a larger RBD inserted subdomain. The two CoV use concave surfaces to bind different receptors. MERS-CoV uses a small “canyon” to bind to an α-helix in the linker between blades IV and V of the DPP4 β-propeller (Fig. 7), whereas the curved inserted subdomain in SARS-CoV RBD cradles the N-terminal α-helix of ACE2 (Fig. 4). The mode by which these CoV bind to receptors shows similarities to other CoV-receptor interactions, particularly to hCoV-NL63, which also binds to ACE2 (Fig. 3B). NL63- and SARS-CoV recognize overlapping ACE2 regions, including two helices and a

β-turn in the virus-binding lobe of the receptor. The ACE2-binding surfaces in both CoV are concave and are distant from the terminal end of the RBD. The receptor-binding surface in SARS-CoV is more extended and curved than in hCoV-NL63, and interacts more extensively with the ACE2N-terminal α-helix; the two residues involved in SARS-CoV adaptation to humans (Asn479 and Thr487) interact directly with the α-helix.

MERS- and alpha1-CoV share recognition of carbohydrates N-linked to their receptors (Fig. 3A, 7B) (Lu et al., 2013; Reguera et al., 2012; Wang et al., 2013). In APN, the N-linked glycan is essential for binding and infection of TGEV and related alpha-CoV (Reguera et al., 2012; Tusell et al., 2007). Receptor glycosylations are important determinants of CoV-receptor recognition, as they can promote or hinder CoV binding to cell entry receptors in certain species (Holmes, 2005; Tusell et al., 2007), which delimits CoV host range.

The CoV RBD is a major target of neutralizing Ab that prevent virus infection by blocking virus binding to receptors (Hwang et al., 2006; Pak et al., 2009; Prabakaran et al., 2006; Reguera et al., 2012; Zhu et al., 2007). RBD protein can elicit potent neutralizing Ab and protective immune responses (Du et al., 2009). These neutralizing Ab recognize the exposed receptor-binding tyrosine or tryptophan in TGEV or PRCV (Reguera et al., 2012). In the SARS-CoV, structural studies showed that several neutralizing Ab bind to the receptor-binding subdomain (Fig. 5) (Hwang et al., 2006; Pak et al., 2009; Prabakaran et al., 2006). These results indicate that the receptor-binding regions are under selective pressure from the immune system. In alpha-CoV, this pressure could mediate the notable conformational changes in the RBD tip (Fig. 2), which alter receptor-binding specificity. The APN-binding tip in alpha-CoV RBD has exposed receptor-binding residues that are easily targeted by Ab, whereas the recessed ACE2-binding tip in hCoV-NL63 more efficiently hides conserved receptor-binding residues from immune surveillance.

Apha- and beta-CoV RBD folds are distinct but are both unique, with no known homology to host domains (Chen et al., 2013; Li et al., 2005a; Peng et al., 2011; Reguera et al., 2012; Wu et al., 2009). They are thought to have evolved from a common CoV RBD ancestor (Li, 2012). They share some common features, such as recognition of glycans N-linked to receptors, and the presence of parallel β-strands (β2-β11 in MERS and β1- β3- β7 in TGEV, Fig. 2A, 6A). It is tempting to speculate that this precursor RBD had a β-barrel fold similar to the alpha-CoV, with a variable tip that accommodated different receptor molecules. In SARS and MERS beta-CoV, the RBD lost the β-barrel fold, but maintained two β-sheets, one of which forms a large receptor-binding platform with recessed surfaces that bind to specific motifs in receptor molecules. The receptor-binding subdomains in SARS and MERS beta-CoV appear to specialize in recognition of orthologous receptor molecules. The beta-CoV RBD probably evolved to enhance host-to-host transmission, responsible for the recurrent CoV outbreaks in man.

Structural studies reviewed here have established the basis for understanding receptor recognition diversity in CoV, its evolution and adaptation to different hosts. These studies have identified sites of vulnerability in the CoV S that should guide the development of anti-virals and vaccines to prevent CoV infections.

Analysis and representation of crystal structures

Buried surfaces and residues at the molecular complex interfaces were determined with the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Figure 2A was prepared with Ribbons (Carson, 1987), Figure 10 with Chimera (Pettersen et al., 2004) and the other structure representations with PyMOL software (pymol.org).

Acknowledgements

We thank C. Mark for editorial assistance. GM is a PhD fellow of the La Caixa Foundation International Fellowship Program (La Caixa/CNB). The work was supported by a grant from

the Spanish Ministry of Science (BFU2011-23940 to JMC). We acknowledge support by the European Synchrotron Research Facility through the BAG-Madrid project.

References

Backovic, M., Rey, F.A., 2012. Virus entry: old viruses, new receptors. Curr. Opin. Virol. 2(1), 4-13.

Beauchemin, N., Draber, P., Dveksler, G., Gold, P., Gray-Owen, S., Grunert, F., Hammarström, S., Holmes, K., Karlsson, A., Kuroki, M., Lin, S., Lucka, L., Najjar, S., Neumaier, M., Obrink, B., Shively, J., Skubitz, K., Stanners, C., Thomas, P., Thompson, J., Virji, M., von Kleist, S., Wagener, C., Watt, S., Zimmermann, W., 1999. Redefined Nomenclature for Members of the Carcinoembryonic Antigen Family. Exp. Cell Res. 252(2), 243-249.

Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4(6), 1011-1033.

Beniac, D.R., Andonov, A., Grudeski, E., Booth, T.F., 2006. Architecture of the SARS coronavirus prefusion spike. Nat. Struct. Mol. Biol. 13(8), 751-752.

Beniac, D.R., deVarennes, S.L., Andonov, A., He, R., Booth, T.F., 2007. Conformational reorganization of the SARS coronavirus spike following receptor binding: implications for membrane fusion. PLoS ONE 2(10), e1082.

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(4), 2530-2538.

Boonacker, E., Van Noorden, C.J.F., 2003. The multifunctional or moonlighting protein CD26/DPPIV. Eur. J. Cell Biol. 82(2), 53-73.

Bosch, B.J., van der Zee, R., Haan, C.A.M., Rottier, P.J.M., 2003. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. J. Virol. 77(16), 8801-8811.

Breslin, J.J., Mork, I., Smith, M.K., Vogel, L.K., Hemmila, E.M., Bonavia, A., Talbot, P.J., Sjostrom, H., Noren, O., Holmes, K.V., 2003. Human coronavirus 229E: receptor binding domain and neutralization by soluble receptor at 37 degrees C. J. Virol. 77(7), 4435-4438.

Carson, M., 1987. Ribbon models of macromolecules. J. Mol. Graph. 5, 103-106.

Casasnovas, J.M., 2013. Virus-Receptor Interactions and Receptor-Mediated Virus Entry into Host Cells. In: Mateu, M.G. (Ed.), Structure and Physics of Viruses. Vol. 68. Springer Netherlands, pp. 441-466.

Chen, Y., Rajashankar, K.R., Yang, Y., Agnihothram, S.S., Liu, C., Lin, Y.-L., Baric, R.S., Li, F., 2013. Crystal Structure of the Receptor-Binding Domain from Newly Emerged Middle East Respiratory Syndrome Coronavirus. J. Virol. 87(19), 10777-10783.

Coleman, C.M., Frieman, M.B., 2014. Coronaviruses: Important Emerging Human Pathogens. J. Virol. 88(10), 5209-5212.

Crackower, M.A., Sarao, R., Oudit, G.Y., Yagil, C., Kozieradzki, I., Scanga, S.E., Oliveira-dos-Santos, A.J., da Costa, J., Zhang, L., Pei, Y., Scholey, J., Ferrario, C.M., Manoukian, A.S., Chappell, M.C., Backx, P.H., Yagil, Y., Penninger, J.M., 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart function. Nature 417(6891), 822-828.

de Groot, R.J., Baker, S.C., Baric, R., Enjuanes, L., Gorbalenya, A.E., Holmes, K.V., Perlman, S., Poon, L., Rottier, P.J.M., Talbot, P.J., Woo, P.C.Y., Ziebuhr, J., 2011. Coronaviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.),

679 Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of
680 Viruses. Elsevier Academic Press, San Diego, pp. 774-796.
681 de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier,
682 R.A.M., Galiano, M., Gorbalenya, A.E., Memish, Z.A., Perlman, S., Poon, L.L.M.,
683 Snijder, E.J., Stephens, G.M., Woo, P.C.Y., Zaki, A.M., Zambon, M., Ziebuhr, J.,
684 2013. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement
685 of the Coronavirus Study Group. J. Virol. 87(14), 7790-7792.
686 Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L.K., Sjostrom, H., Noren, O., Laude, H., 1992.
687 Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV.
688 Nature 357(6377), 417-420.
689 Delmas, B., Rasschaert, D., Godet, M., Gelfi, J., Laude, H., 1990. Four major antigenic sites
690 of the coronavirus transmissible gastroenteritis virus are located on the amino-
691 terminal half of spike glycoprotein S. J. Gen. Virol. 71, 1313-1323.
692 Doremalen, N.v., Miazgowicz, K.L., Milne-Price, S., Bushmaker, T., Robertson, S., Scott,
693 D., Kinne, J., McLellan, J.S., Zhu, J., Munster, V.J., 2014. Host Species Restriction of
694 Middle East Respiratory Syndrome Coronavirus through Its Receptor, Dipeptidyl
695 Peptidase 4. J. Virol. 88(16), 9220-9232.
696 Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., Jiang, S., 2009. The spike protein of SARS-
697 CoV—a target for vaccine and therapeutic development. Nat. Rev. Microbiol. 7(3),
698 226-236.
699 Du, L., Zhao, G., Kou, Z., Ma, C., Sun, S., Poon, V.K.M., Lu, L., Wang, L., Debnath, A.K.,
700 Zheng, B.-J., Zhou, Y., Jiang, S., 2013. Identification of a Receptor-Binding Domain
701 in the S Protein of the Novel Human Coronavirus Middle East Respiratory Syndrome
702 Coronavirus as an Essential Target for Vaccine Development. J. Virol. 87(17), 9939-
703 9942.
704 Dveksler, G.S., Dieffenbach, C.W., Cardellichio, C.B., McCuaig, K., Pensiero, M.N., Jiang,
705 G.S., Beauchemin, N., Holmes, K.V., 1993. Several members of the mouse
706 carcinoembryonic antigen-related glycoprotein family are functional receptors for the
707 coronavirus mouse hepatitis virus-A59. J. Virol. 67(1), 1-8.
708 Gallagher, T., Perlman, S., 2013. Public health: Broad reception for coronavirus. Nature
709 495(7440), 176-177.
710 Gallagher, T.M., Buchmeier, M.J., 2001. Coronavirus Spike Proteins in Viral Entry and
711 Pathogenesis. Virology 279(2), 371-374.
712 Glowacka, I., Bertram, S., Muller, M.A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I.,
713 Tsegaye, T.S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C.,
714 Pohlmann, S., 2011. Evidence that TMPRSS2 Activates the Severe Acute Respiratory
715 Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral
716 Control by the Humoral Immune Response. J. Virol. 85(9), 4122-4134.
717 Godet, M., Grosclaude, J., Delmas, B., Laude, H., 1994. Major receptor-binding and
718 neutralization determinants are located within the same domain of the transmissible
719 gastroenteritis virus (coronavirus) spike protein. J. Virol. 68(12), 8008-8016.
720 Graham, R.L., Baric, R.S., 2010. Recombination, Reservoirs, and the Modular Spike:
721 Mechanisms of Coronavirus Cross-Species Transmission. J. Virol. 84(7), 3134-3146.
722 Haagmans, B.L., Al Dhahiry, S.H.S., Reusken, C.B.E.M., Raj, V.S., Galiano, M., Myers, R.,
723 Godeke, G.-J., Jonges, M., Farag, E., Diab, A., Ghobashy, H., Alhajri, F., Al-Thani,
724 M., Al-Marri, S.A., Al Romaihi, H.E., Al Khal, A., Bermingham, A., Osterhaus,
725 A.D.M.E., AlHajri, M.M., Koopmans, M.P.G., 2014. Middle East respiratory
726 syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect.
727 Dis. 14(2), 140-145.

Harrison, S.C., 2005. Mechanism of membrane fusion by viral envelope proteins. Adv. Virus Res. 64, 231-261.

He, Y., Li, J., Li, W., Lustigman, S., Farzan, M., Jiang, S., 2006. Cross-Neutralization of Human and Palm Civet Severe Acute Respiratory Syndrome Coronaviruses by Antibodies Targeting the Receptor-Binding Domain of Spike Protein. J. Immunol. 176(10), 6085-6092.

He, Y., Lu, H., Siddiqui, P., Zhou, Y., Jiang, S., 2005. Receptor-Binding Domain of Severe Acute Respiratory Syndrome Coronavirus Spike Protein Contains Multiple Conformation-Dependent Epitopes that Induce Highly Potent Neutralizing Antibodies. J. Immunol. 174(8), 4908-4915.

Holmes, K.V., 2005. STRUCTURAL BIOLOGY: Adaptation of SARS Coronavirus to Humans. Science 309(5742), 1822-1823.

Huang, I.C., Bosch, B.J., Li, F., Li, W., Lee, K.H., Ghiran, S., Vasilieva, N., Dermody, T.S., Harrison, S.C., Dormitzer, P.R., Farzan, M., Rottier, P.J., Choe, H., 2006. SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. J. Biol. Chem. 281(6), 3198-3203.

Hwang, W.C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W.A., Liddington, R.C., 2006. Structural Basis of Neutralization by a Human Anti-severe Acute Respiratory Syndrome Spike Protein Antibody, 80R. J. Biol. Chem. 281(45), 34610-34616.

Kahn, J.S., McIntosh, K., 2005. History and recent advances in coronavirus discovery. Pediatr. Infect. Dis. J. 24(11 Suppl), S223-227, discussion S226.

Krempl, C., Schultze, B., Herrler, G., 1995. Analysis of cellular receptors for human coronavirus OC43. Adv. Exp. Med. Biol. 380, 371-374.

Li, F., 2012. Evidence for a Common Evolutionary Origin of Coronavirus Spike Protein Receptor-Binding Subunits. J. Virol. 86(5), 2856-2858.

Li, F., 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res. 100(1), 246-254.

Li, F., Li, W., Farzan, M., Harrison, S.C., 2005a. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309(5742), 1864-1868.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., Choe, H., Farzan, M., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426(6965), 450-454.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S., Wang, L.-F., 2005b. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. Science 310(5748), 676-679.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W.A., Guan, Y., Choe, H., Farzan, M., 2005c. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 24(8), 1634-1643.

Lu, G., Hu, Y., Wang, Q., Qi, J., Gao, F., Li, Y., Zhang, Y., Zhang, W., Yuan, Y., Bao, J., Zhang, B., Shi, Y., Yan, J., Gao, G.F., 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature 500, 227–231.

Marsh, M., Helenius, A., 2006. Virus entry: open sesame. Cell 124(4), 729-740.

Masters, P.S., 2006. The molecular biology of coronaviruses. Adv. Virus Res. 66, 193-292.

Mina-Osorio, P., 2008. The moonlighting enzyme CD13: old and new functions to target. Trends Mol. Med. 14(8), 361-371.

Pak, J.E., Sharon, C., Satkunarajah, M., Auperin, T.C., Cameron, C.M., Kelvin, D.J., Seetharaman, J., Cochrane, A., Plummer, F.A., Berry, J.D., Rini, J.M., 2009. Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. J. Mol. Biol. 388(4), 815-823.

Peng, G., Sun, D., Rajashankar, K.R., Qian, Z., Holmes, K.V., Li, F., 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc. Natl. Acad. Sci. USA 108(26), 10696-10701.

Peng, G., Xu, L., Lin, Y.-L., Chen, L., Pasquarella, J.R., Holmes, K.V., Li, F., 2012. Crystal structure of bovine coronavirus spike protein lectin domain. J. Biol. Chem. 287(50), 41931-41938.

Perlman, S., Netland, J., 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nat. Rev. Microbiol. 7(6), 439-450.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 25(13), 1605-1612.

Prabakaran, P., Gan, J., Feng, Y., Zhu, Z., Choudhry, V., Xiao, X., Ji, X., Dimitrov, D.S., 2006. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. J. Biol. Chem. 281(23), 15829-15836.

Qinghong, Z., Martijn, A.L., Arno, L.W.v.V., Eric, G.H., Raoul, J.d.G., 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc. Natl. Acad. Sci. USA 105(26), 9065-9069.

Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H.W., Muller, M.A., Dijkman, R., Muth, D., Demmers, J.A.A., Zaki, A., Fouchier, R.A.M., Thiel, V., Drosten, C., Rottier, P.J.M., Osterhaus, A.D.M.E., Bosch, B.J., Haagmans, B.L., 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495(7440), 251-254.

Raj, V.S., Osterhaus, A.D.M.E., Fouchier, R.A.M., Haagmans, B.L., 2014. MERS: emergence of a novel human coronavirus. Curr. Opin. Virol. 5(0), 58-62.

Reguera, J., Ordoño, D., Santiago, C., Enjuanes, L., Casasnovas, J.M., 2011. Antigenic modules in the N-terminal S1 region of the Transmissible Gastroenteritis Virus spike protein. J. Gen. Virol. 92, 1117–1126.

Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L., Casasnovas, J.M., 2012. Structural Bases of Coronavirus Attachment to Host Aminopeptidase N and Its Inhibition by Neutralizing Antibodies. PLoS Pathog. 8(8), e1002859.

Rockx, B., Donaldson, E., Frieman, M., Sheahan, T., Corti, D., Lanzavecchia, A., Baric, R.S., 2010. Escape from human monoclonal antibody neutralization affects in vitro and in vivo fitness of severe acute respiratory syndrome coronavirus. J. Infect. Dis. 201(6), 946-955.

Rossmann, M., 1989. The Canyon Hypothesis: Hiding the host cell receptor attachment site on a viral surface from immune surveillance. J. Biol. Chem. 264, 14587-14590.

Sanchez, C.M., Gebauer, F., Sune, C., Mendez, A., Dopazo, J., Enjuanes, L., 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. Virology 190(1), 92-105.

Schultze, B., Krempl, C., Ballesteros, M.L., Shaw, L., Schauer, R., Enjuanes, L., Herrler, G., 1996. Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. J. Virol. 70(8), 5634-5637.

Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S.H., Rini, J.M., 1998. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. J. Biol. Chem. 273(21), 13047-13052.

827 Smith, M.K., Tusell, S., Travanty, E.A., Berkhout, B., van der Hoek, L., Holmes, K.V., 2006.
828 Human angiotensin-converting enzyme 2 (ACE2) is a receptor for human respiratory
829 coronavirus NL63. Adv. Exp. Med. Biol. 581, 285-288.
830 Stehle, T., Casasnovas, J.M., 2009. Specificity switching in virus-receptor complexes. Curr.
831 Opin. Struct. Biol. 19(2), 181-188.
832 Sturman, L.S., Ricard, C.S., Holmes, K.V., 1990. Conformational change of the coronavirus
833 peplomer glycoprotein at pH 8.0 and 37ºC correlates with virus aggregation and
834 virus-induced cell fusion. J. Virol. 64(6), 3042-3050.
835 Sune, C., Jimenez, G., Correa, I., Bullido, M.J., Gebauer, F., Smerdou, C., Enjuanes, L.,
836 1990. Mechanisms of transmissible gastroenteritis coronavirus neutralization.
837 Virology 177(2), 559-569.
838 Supekar, V.M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A., Carfi, A., 2004.
839 Structure of a proteolytically resistant core from the severe acute respiratory
840 syndrome coronavirus S2 fusion protein. Proc. Natl. Acad. Sci. USA 101(52), 17958-
841 17963.
842 Tan, K., Zelus, B.D., Meijers, R., Liu, J.-h., Bergelson, J.M., Duke, N., Zhang, R.,
843 Joachimiak, A., Holmes, K.V., Wang, J.-h., 2002. Crystal structure of murine
844 sCEACAM1a[1,4]:a coronavirus receptor in the CEA family. EMBO J. 21(9), 2076-
845 2086.
846 Towler, P., Staker, B., Prasad, S.G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M.,
847 Williams, D., Dales, N.A., Patane, M.A., Pantoliano, M.W., 2004. ACE2 X-Ray
848 Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and
849 Catalysis. J. Biol. Chem. 279(17), 17996-18007.
850 Tsai, J.C., Zelus, B.D., Holmes, K.V., Weiss, S.R., 2003. The N-terminal domain of the
851 murine coronavirus spike glycoprotein determines the CEACAM1 receptor specificity
852 of the virus strain. J. Virol. 77(2), 841-850.
853 Tusell, S.M., Schittone, S.A., Holmes, K.V., 2007. Mutational analysis of aminopeptidase N,
854 a receptor for several group 1 coronaviruses, identifies key determinants of viral host
855 range. J. Virol. 81(3), 1261-1273.
856 Virji, M., Evans, D., Hadfield, A., Grunert, F., Teixeira, A.M., Watt, S.M., 1999. Critical
857 determinants of host receptor targeting by Neisseria meningitidis and Neisseria
858 gonorrhoeae: identification of Opa adhesiotopes on the N-domain of CD66 molecules.
859 Mol. Microbiol. 34(3), 538-551.
860 Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X.,
861 Arledge, K.C., Chen, Y.-H., Zhang, L., Wang, X., 2013. Structure of MERS-CoV
862 spike receptor-binding domain complexed with human receptor DPP4. Cell. Res.
863 23(8), 986-993.
864 Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J.,
865 Iwamoto, A., Woo, P.C.Y., Yuen, K.-Y., Yan, J., Lu, G., Gao, G.F., 2014. Bat Origins
866 of MERS-CoV Supported by Bat Coronavirus HKU4 Usage of Human Receptor
867 CD26. Cell Host Microbe 16(3), 328-337.
868 Weiss, S.R., Leibowitz, J.L., 2011. Chapter 4 - Coronavirus Pathogenesis. In: Karl
869 Maramorosch, A.J.S., Frederick, A.M. (Eds.), Adv. Virus Res. Vol. Volume 81.
870 Academic Press, pp. 85-164.
871 Williams, R.K., Jiang, G.S., Holmes, K.V., 1991. Receptor for mouse hepatitis virus is a
872 member of the carcinoembryonic antigen family of glycoproteins. Proc. Natl. Acad.
873 Sci. USA 88(13), 5533-5536.
874 Wong, S.K., Li, W., Moore, M.J., Choe, H., Farzan, M., 2004. A 193-amino acid fragment of
875 the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. J.
876 Biol. Chem. 279(5), 3197-3201.

Wu, K., Li, W., Peng, G., Li, F., 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc. Natl. Acad. Sci. USA 106(47), 19970-19974.

Xu, Y., Liu, Y., Lou, Z., Qin, L., Li, X., Bai, Z., Pang, H., Tien, P., Gao, G.F., Rao, Z., 2004. Structural Basis for Coronavirus-mediated Membrane Fusion: Crystal Structure of Mouse Hepatitis Virus Spike Protein Fusion Core. J. Biol. Chem. 279(29), 30514-30522.

Xu, Y., Wellner, D., Scheinberg, D.A., 1997. Cryptic and regulatory epitopes in CD13/aminopeptidase N. Exp. Hematol. 25(6), 521-529.

Yang, Y., Du, L., Liu, C., Wang, L., Ma, C., Tang, J., Baric, R.S., Jiang, S., Li, F., 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc. Natl. Acad. Sci. USA 111(34), 12516-12521.

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(6377), 420-422.

Zhu, Z., Chakraborti, S., He, Y., Roberts, A., Sheahan, T., Xiao, X., Hensley, L.E., Prabakaran, P., Rockx, B., Sidorov, I.A., Corti, D., Vogel, L., Feng, Y., Kim, J.O., Wang, L.F., Baric, R., Lanzavecchia, A., Curtis, K.M., Nabel, G.J., Subbarao, K., Jiang, S., Dimitrov, D.S., 2007. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. Proc. Natl. Acad. Sci. USA 104(29), 12123-12128.

Figure legends

Figure 1. The CoV S glycoprotein and CoV cell surface receptors

Scheme of a CoV S glycoprotein with the functional domains in the S1 and S2 regions, which are exposed in the virus envelope. The N-terminal signal peptide and the transmembrane region are also shown. The N-terminal domain (NTD) that can act as a receptor-binding domain (N-RBD) and the canonical CoV RBD in the C-terminal portion of S1 are indicated. The heptad repeat regions (HR1 and HR2) and the putative fusion peptide (FP) are marked in S2. The arrowhead indicates the putative protease cleavage site in some CoV. Cell entry receptor molecules identified for the indicated CoV (right) are shown beneath their respective RBD regions. Sialic acids recognized by TGEV and IBV should be considered attachment factors.

Figure 2. Structures of alpha-CoV RBD and receptor-binding surfaces

A. Ribbon diagram of the TGEV RBD structure (PDB ID 4F2M) (Reguera et al., 2012). β-strands (numbered) are shown in light or dark blue, coils in orange, and the helix in red; a

β-bulge at β-strand 5 is in magenta. N- and C-terminal ends on the terminal side of the structure are indicated in lowercase letters. The Asn residues at glycosylation sites and the attached glycans defined in the structure are shown as a ball-and-stick model, with carbons in yellow. Cysteine residues and disulphide bonds are shown as green cylinders. The two β-turns at the β-barrel domain tip are labeled. Ribbon diagrams of the PRCV and hCoV-NL63 RBD structures are shown in **B** and **C**, respectively. The structures of these domains were determined in complex with the APN (PRCV, PDB ID 4F5C) and ACE2 receptors (NL63, PDB ID 3KBH) (Reguera et al., 2012; Wu et al., 2009). Receptor-binding surfaces in the RBD are shown in pink or red (tyrosine or tryptophan residues) and were generated by the RBD residues that contact the respective receptor molecules in the structures.

Figure 3. Alpha-CoV recognition of cell surface receptors

Crystal structures of alpha-CoV RBD in complex with the ectodomains of APN (**A**) and ACE2 (**B**).

A. Ribbon drawing of the dimeric structure of the PRCV RBD-APN complex (PDB ID 4F5C) (Reguera et al., 2012). Pig APN molecules are shown with domains in orange (N-terminal DI), yellow (DII), red (DIII) and green (C-terminal DIV), as well as the N-terminal ends near the putative location of the cell membrane. The RBD is shown as ribbon and surface drawings in blue and cyan, with the APN-binding tyrosine and tryptophan residues at the RBD tip in red.

B. Ribbon drawing of the hCoV-NL63 RBD-ACE2 complex (PDB ID 3KBH) (Wu et al., 2009). The ACE2 molecule is shown with the two lobes in green (N-terminal) and orange (C-terminal). The RBD is shown as ribbon and surface drawings in blue, with the ACE2-binding residue in pink and the aromatic residues that contact the receptor in red. The N- and C-terminal ends of the receptor molecules are marked in lowercase letters, N-linked glycans

are shown as sticks with carbons in yellow, and the zinc ion at the catalytic sites of APN and ACE2 as cyan spheres.

For A and B, details of key virus-receptor binding motifs are shown beneath the complex structures. Interaction of the PRCV RBD β1-β2 and β3-β4 turns (shown as sticks) at the domain tip with cavities in the APN (ribbon and surface drawings). The tyrosine at the β1-β2 turn contacts APN residues and the NAG carbohydrate (yellow surface), which is N-linked to pig APN Asn736. The tryptophan side chain at the β3-β4 turn penetrates between DII and DIV. Interaction of the concave center of the hCoV-NL63 RBD tip with the ACE2 β4-β5 turn. Lys535 at the tip of the ACE2 turn is labeled. The ACE2 α-helices α1 and α10 contact the most exposed regions of the RBD loops. Sides chains of buried residues in the virus-receptor interfaces are shown with oxygens in red and nitrogens in blue in this and the following figures; hydrogen bonds are dark dashed lines.

Figure 4. SARS-CoV RBD and binding to ACE2

A. Ribbon drawing of the SARS-CoV RBD (PDB ID 2AJF) (Li et al., 2005a), with the core subdomain in yellow and the inserted subdomain in dark red. The β-strands and α-helices are labeled with numbers and uppercase letters, respectively. Terminal ends are labeled in yellow and disulphide bonds in green; Asn residues at glycosylation sites and the attached glycans are shown as sticks, with carbons in yellow. SARS-CoV residues that bind to the ACE2 receptor and define the receptor-binding surface are pink.

B. Ribbon drawing of the SARS-CoV RBD-ACE2 complex (PDB ID 2AJF) (Li et al., 2005a). ACE2 is shown as in Fig. 3B and the RBD as in panel A. The three main ACE2 regions recognized by SARS-CoV are labeled in green.

C. Key virus-receptor binding motifs. ACE2 residues are shown, with carbons in green. In the RBD, receptor-binding tyrosines and an arginine are shown, with carbons in pink,

962 whereas the two critical residues for SARS-CoV adaptation to human ACE2 (Asn479 and
963 Thr487) are shown, with carbons in magenta.

964 **Figure 5. SARS-CoV neutralizing Ab bind to the RBD.**

965 Ribbon drawing of the SARS-CoV RBD in complex with three neutralizing Ab (Hwang et
966 al., 2006; Pak et al., 2009; Prabakaran et al., 2006). The three RBD-Ab crystal structures
967 were superimposed based on the RBD. The RBD is shown as in Figure 4A and the variable
968 domains of the Ab in green (R80, PDB ID 2GHW), blue (F26G19, PDB ID 3BGF) and cyan
969 (m396, PDB ID 2DD8). RBD Ile489, which is recognized by the m396 and F26G19 Ab (Pak
970 et al., 2009; Prabakaran et al., 2006), is black. Side chains of residues that change in scape
971 mutants to the neutralization are shown in red (Rockx et al., 2010).

972 **Figure 6. The MERS-CoV RBD and comparison with the SARS RBD.**

973 **A.** Ribbon drawing of the MERS-CoV RBD (PDB ID 4KRO) (Lu et al., 2013), shown as for
974 SARS-CoV RBD in Fig. 4A, but with the core subdomain in dark yellow. MERS-CoV
975 residues that bind to its DPP4 receptor define the receptor-binding surface (pink). The
976 arrowhead indicates the small "canyon" on one side of the DPP4-binding surface.

977 **B.** Stereo view of superimposed MERS- (yellow) and SARS-CoV (red) RBD, core
978 subdomain-based. The β-strands of the MERS-CoV inserted subdomain are labeled and the
979 two conserved in the SARS-CoV are red.

980 **Figure 7. MERS-CoV RBD binding to DPP4**

981 **A.** Ribbon drawing of the dimeric MERS-CoV RBD-DPP4 complex structure (PDB ID
982 4KRO) (Lu et al., 2013). The DPP4 monomers are shown with the N-terminal β-propeller
983 domain in green and the C-terminal α/β-hydrolase domain in orange. The RBD molecules are
984 as in Figure 6A. Labels and glycosylation are as in previous figures.

985 **B.** Key virus-receptor binding motifs. The virus-binding DPP4 β-propeller blades IV and V
986 are shown in light and dark green, respectively. DPP4 residues are shown, with carbons in

green. In the RBD, residues in the small "canyon" that interact with the exposed α-helix in the blade linker are shown, with carbons in magenta, whereas those that bind to the DPP4 N-linked glycan (Asn229) are shown, with carbons pink. Some residues in the two receptor-binding motifs and the external subdomain β-strands (β6–β9) are labeled.

Figure 8. Structure of the S glycoprotein NTD

A. Ribbon drawing of the human galectin-3 carbohydrate recognition domain (CRD) bound to galactose (PDB ID 1A3K) (Seetharaman et al., 1998). The β-strands in the β-barrel are in light or dark blue, and a galactose ligand on the top of the β-sheet is shown as sticks, with carbons in yellow. N- and C-terminal ends are indicated in lowercase letters.

B. Ribbon drawing of the MHV NTD structure (PDB ID 3R4D) (Peng et al., 2011). The β-strands in the central galectin-like β-barrel are in light or dark blue, and those on the top of the sheet are in pink. The Asn residues at glycosylation sites and the attached glycans defined in the structure are shown as sticks, with carbons in yellow. Cysteine residues and disulphide bonds are shown as green sticks.

Figure 9. MHV recognition of its CEACAM1 receptor

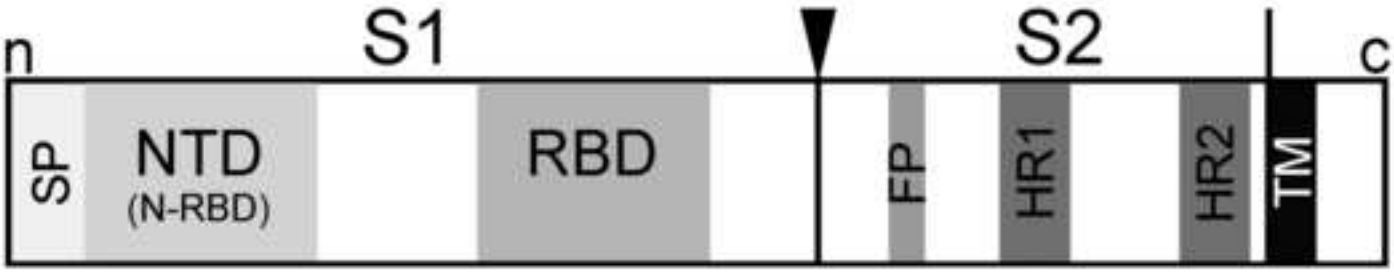
A. The MHV NTD structure with the CEACAM1-binding surface. The NTD ribbon diagram is shown as in Figure 8B. The surface of the N-terminal MHV residues that form a socket is shown in violet and that of the other receptor-binding residues is pink. MHV Leu160 in the bottom of the socket is shown in red.

B. The MHV NTD in complex with the CEACAM1 receptor (PDB ID 3R4D) (Peng et al., 2011). The CEACAM1 N-terminal D1 is shown in green, with the β-strands in the receptor-binding CFG β-sheet labeled. The side chain of CEACAM1 Ile41 that penetrates the NTD socket is shown as spheres. The MHV Leu160 in the socket and Leu174 that contacts the top of D1 are in red.

C. Key virus-receptor binding motifs. Side chains of some receptor-binding MHV residues are shown, with carbons in pink; the hydrophobic residues in the bottom of the socket and Leu174 are in magenta; the CEACAM1 residues are in green. Ile41 in the CC´ loop, the most important virus-binding motif in CEACAM-1 (Peng et al., 2011), is shown as spheres.

Figure 10. Structural view of the multifunctional CoV S with the two domains that bind to host cell surface receptors. The two domains, NTD and RBD, of the S1 region that CoV use for attachment to cell surface molecules (Fig. 1) docked into the cryo-electron microscopy map (grey) of the trimeric SARS-CoV S (EMD-1423) (Beniac et al., 2006). Ribbon representations of the SARS-CoV RBD (yellow) and the MHV NTD (blue) alone or bound to ACE2 (Fig. 4B) and to CEACAM1 D1 (Fig. 9B), respectively.

- Highlights
- Structural basis of coronavirus attachment to host cell entry receptors.
- Coronavirus-receptor complex structures.
- Evolution of receptor-recognition in coronavirus.
- Coronavirus host-to-host transmission and adaptation to man.
- Sites of vulnerability in the coronavirus spike glycoprotein.
- Antibody neutralization of coronavirus.



Receptor		CoV	
Sialic acid	APN	TGEV (alpha1)	alpha-CoV
	APN	PRCV (alpha1)	
	APN	fCoV (alpha1)	
	APN	cCoV (alpha1)	
	APN	hCoV-229E	
	ACE2	hCoV-NL63	
Neu 5,9 Ac2		bCoV (beta1)	beta-CoV
Neu 5,9 Ac2		hCoV-OC43 (beta1)	
CEACAM1		MHV	
	ACE2	SARS-CoV	
	DPP4	MERS-CoV	
Sialic acid		IBV (gamma-CoV)	

Figure2

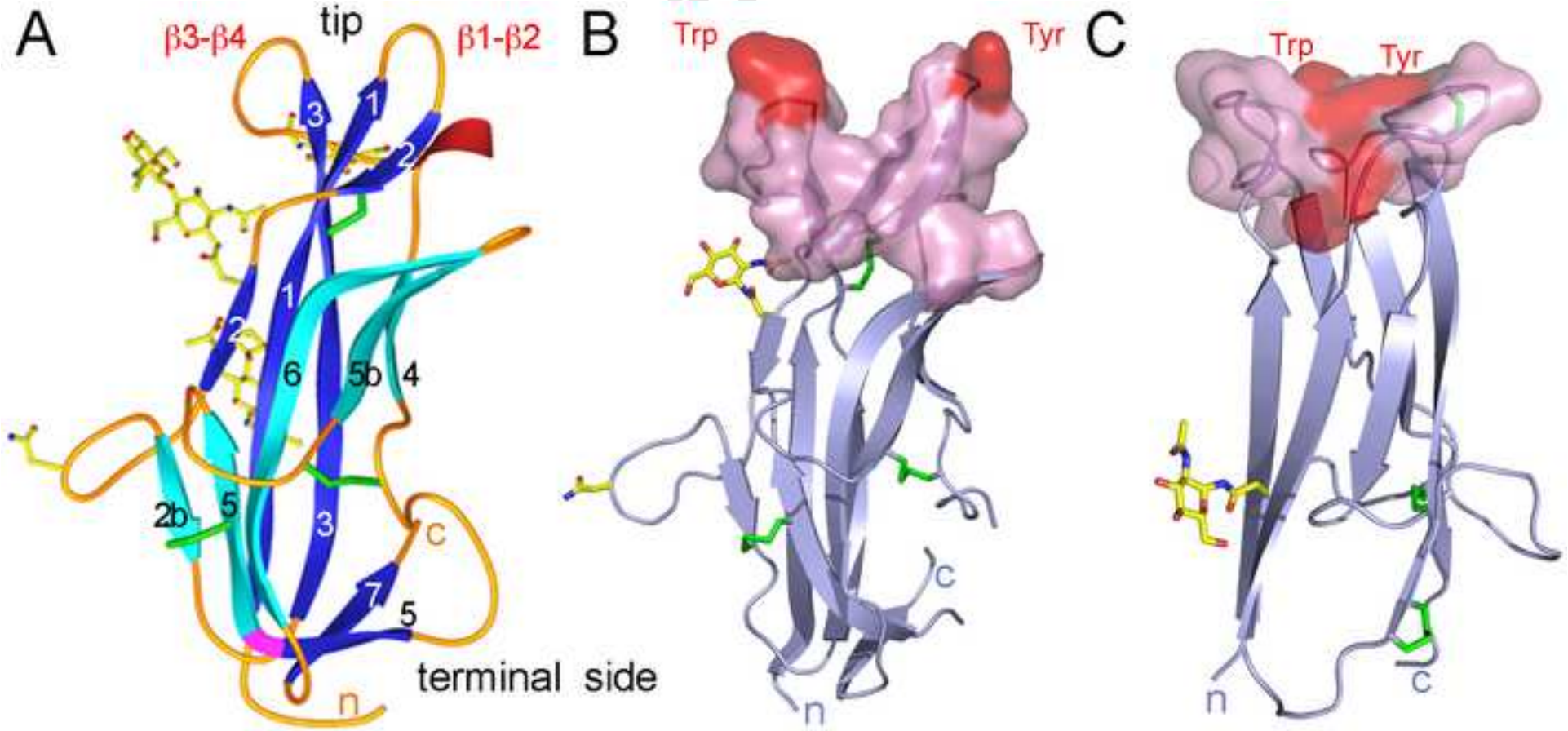
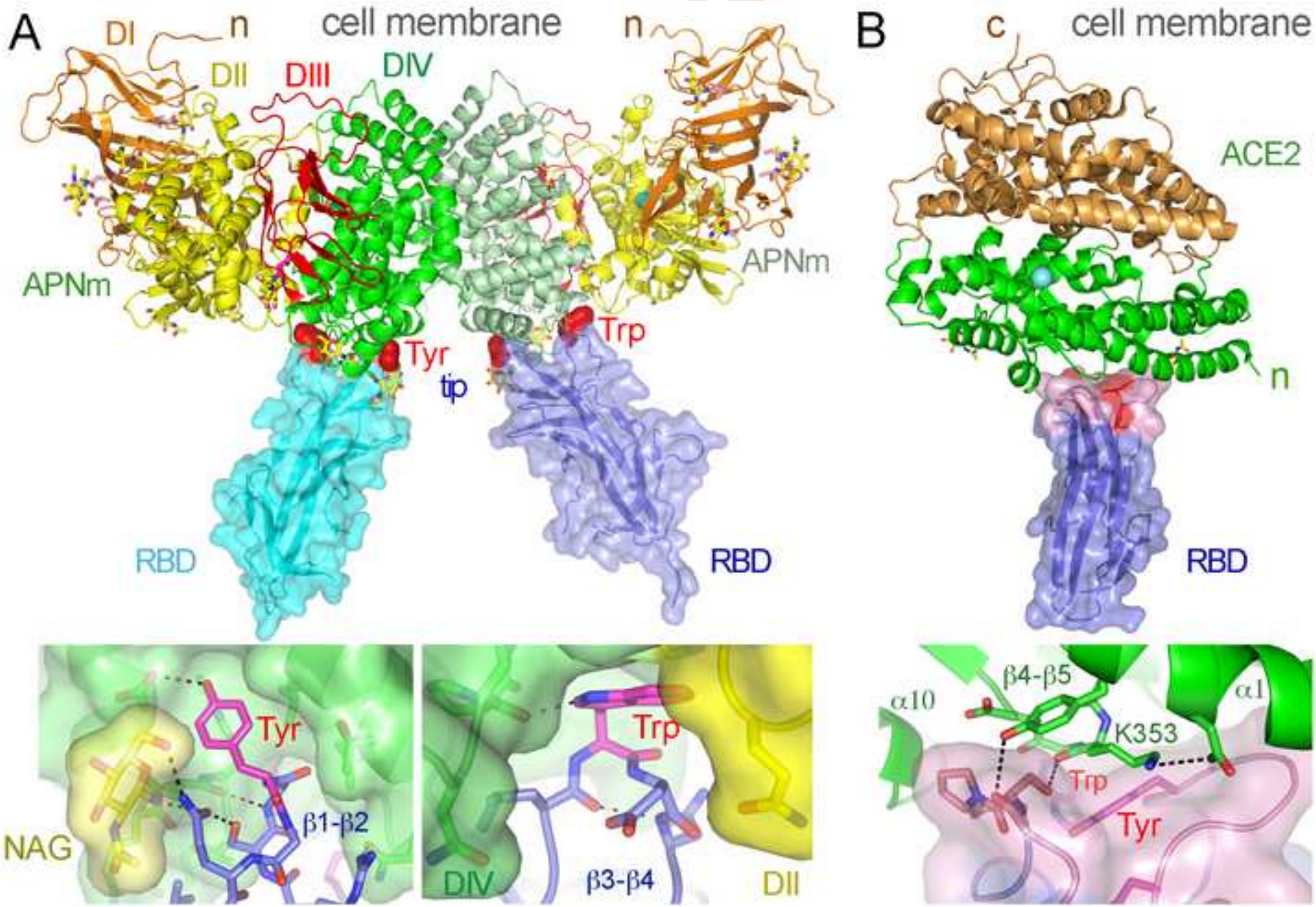

A
β3-β4
tip
β1-β2
3
1
2
1
2
6
5b
4
2b
5
3
c
7
5
n
terminal side
B
Trp
Tyr
c
n
C
Trp
Tyr
n
c

Figure3


A
DI
n
cell membrane
n
DII
DIII
DIV
APNm
APNm
Tyr
tip
Trp
RBD
RBD
B
c
cell membrane
ACE2
n
RBD
Tyr
β1-β2
NAG
Trp
DIV
β3-β4
DII
α10
β4-β5
α1
K353
Trp
Tyr

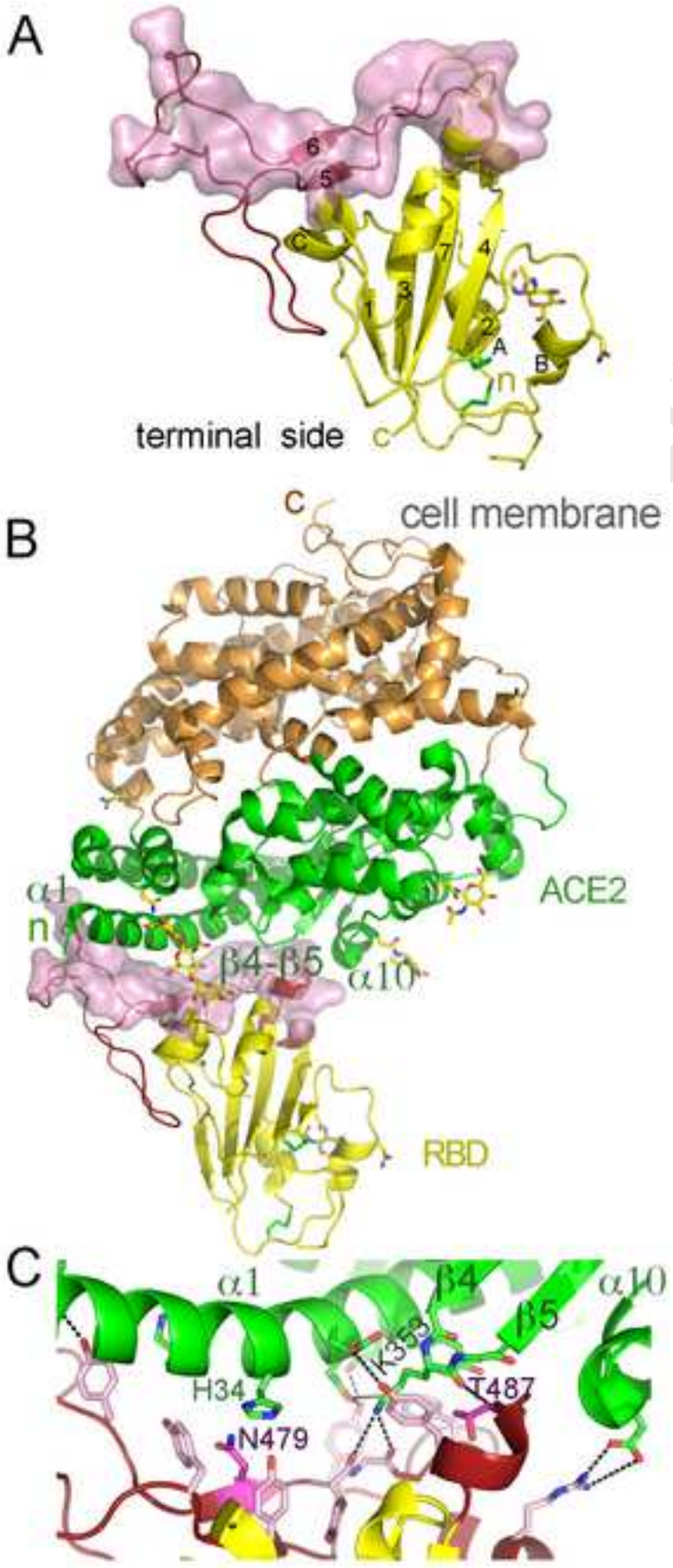


Figure5

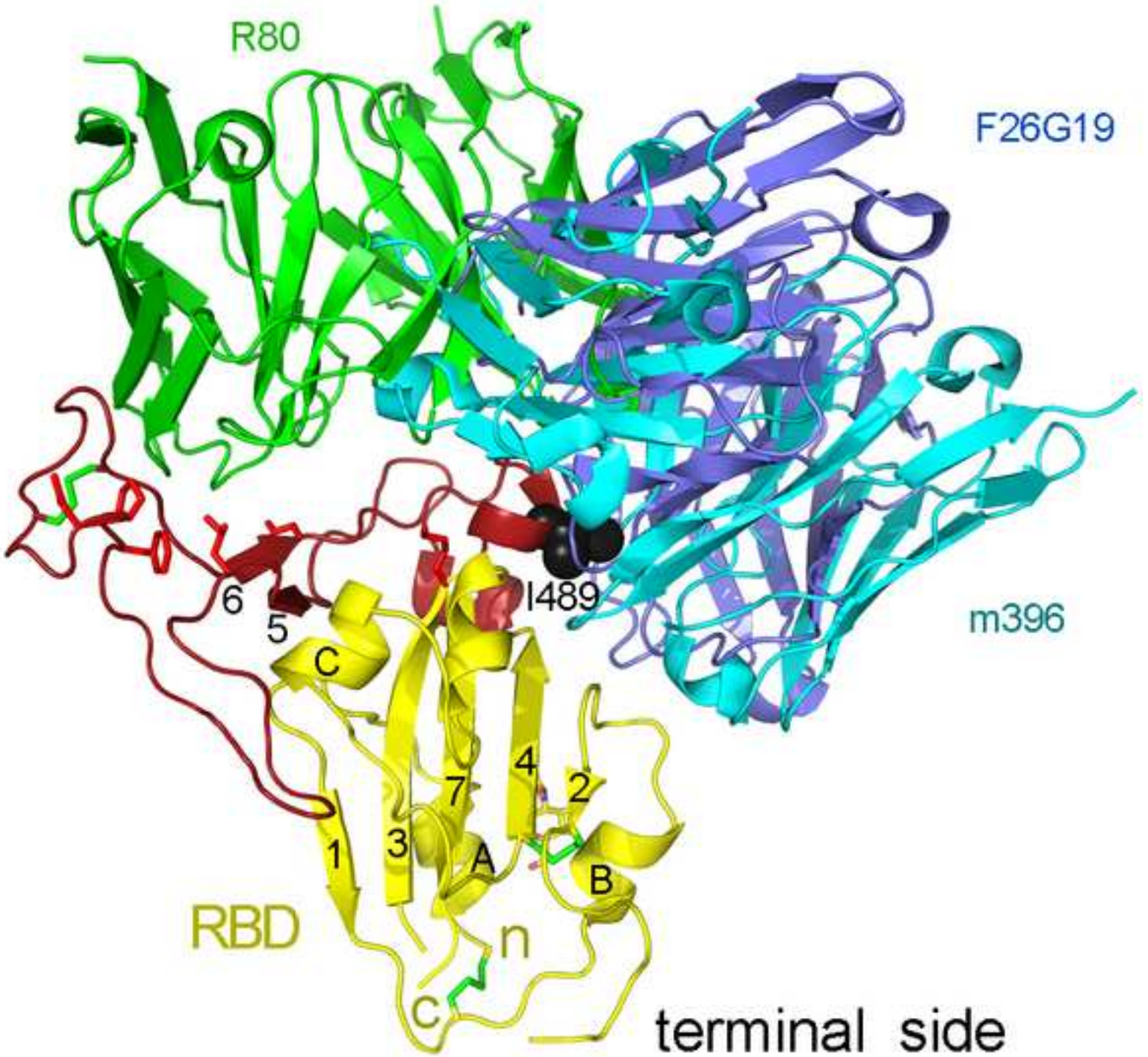
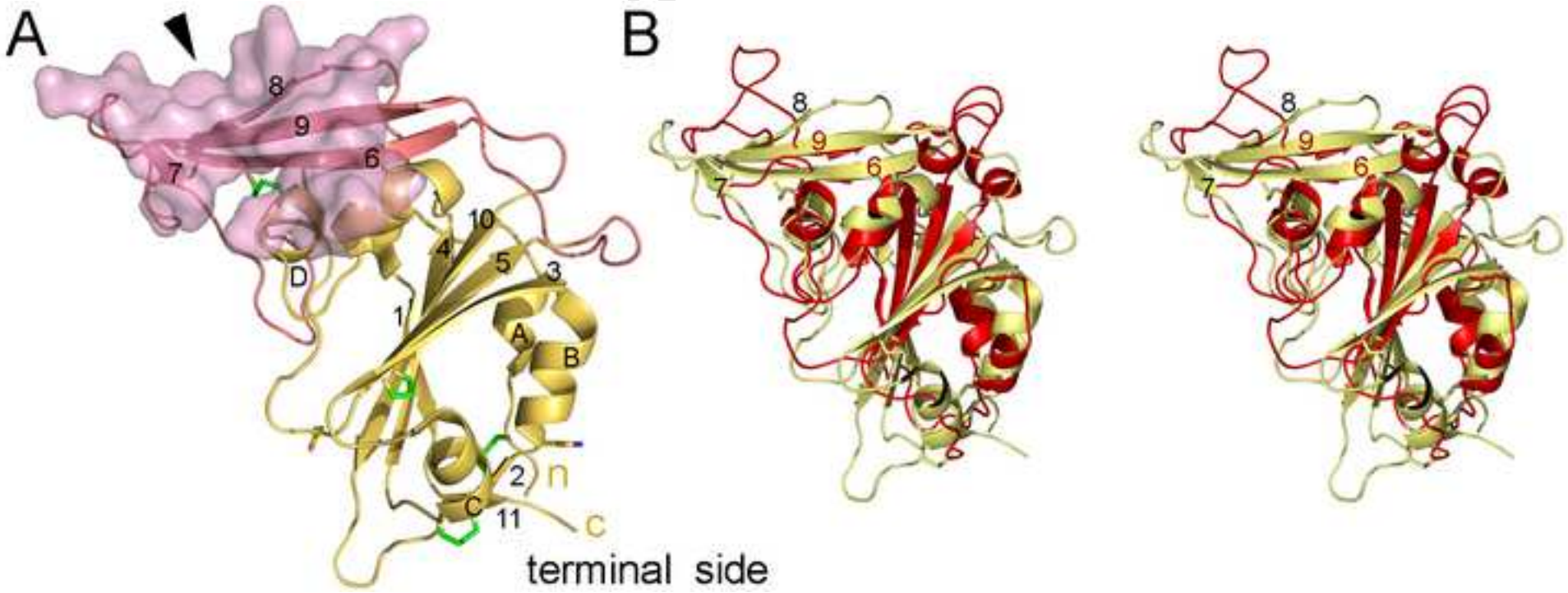
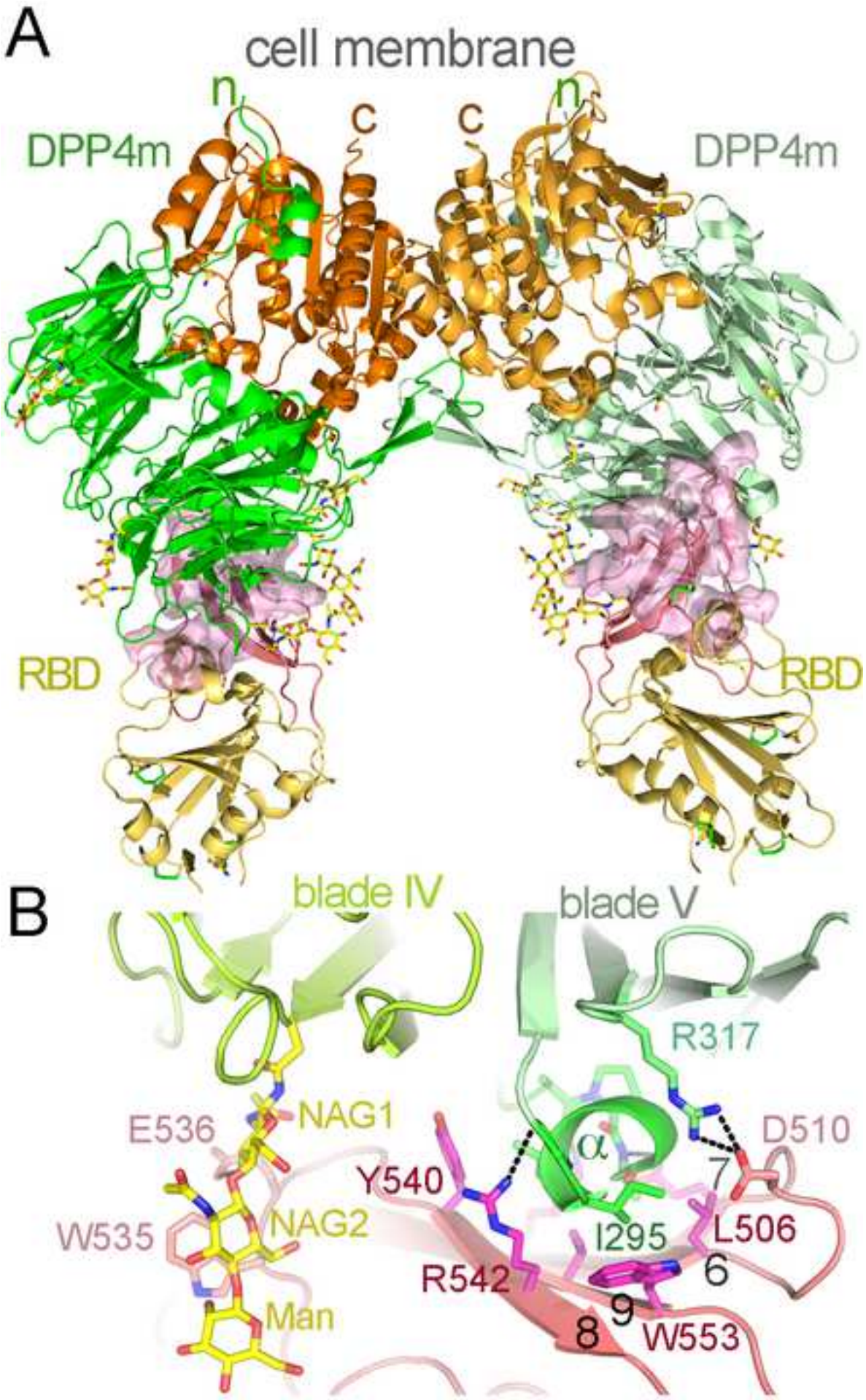


Figure6





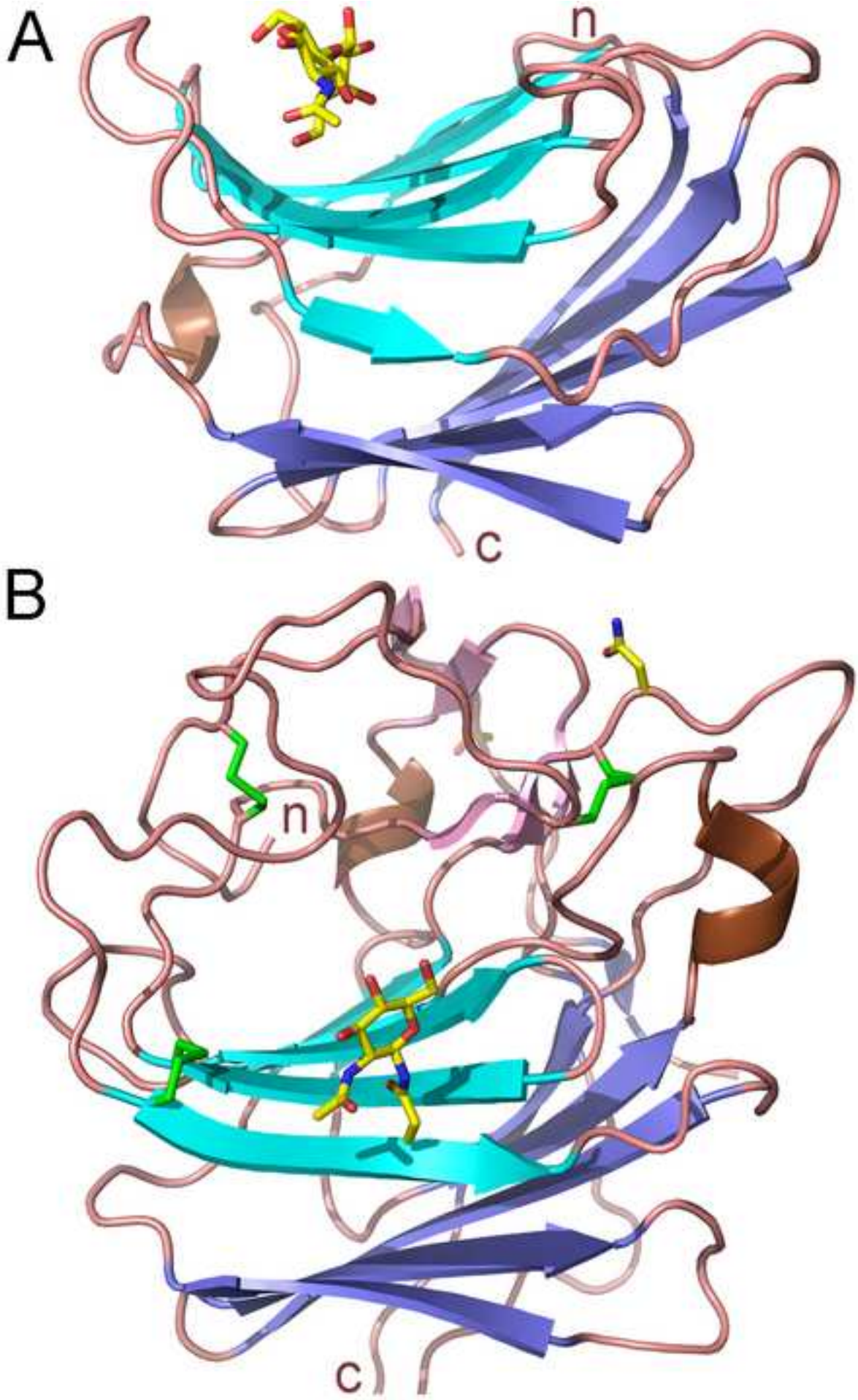


Figure9

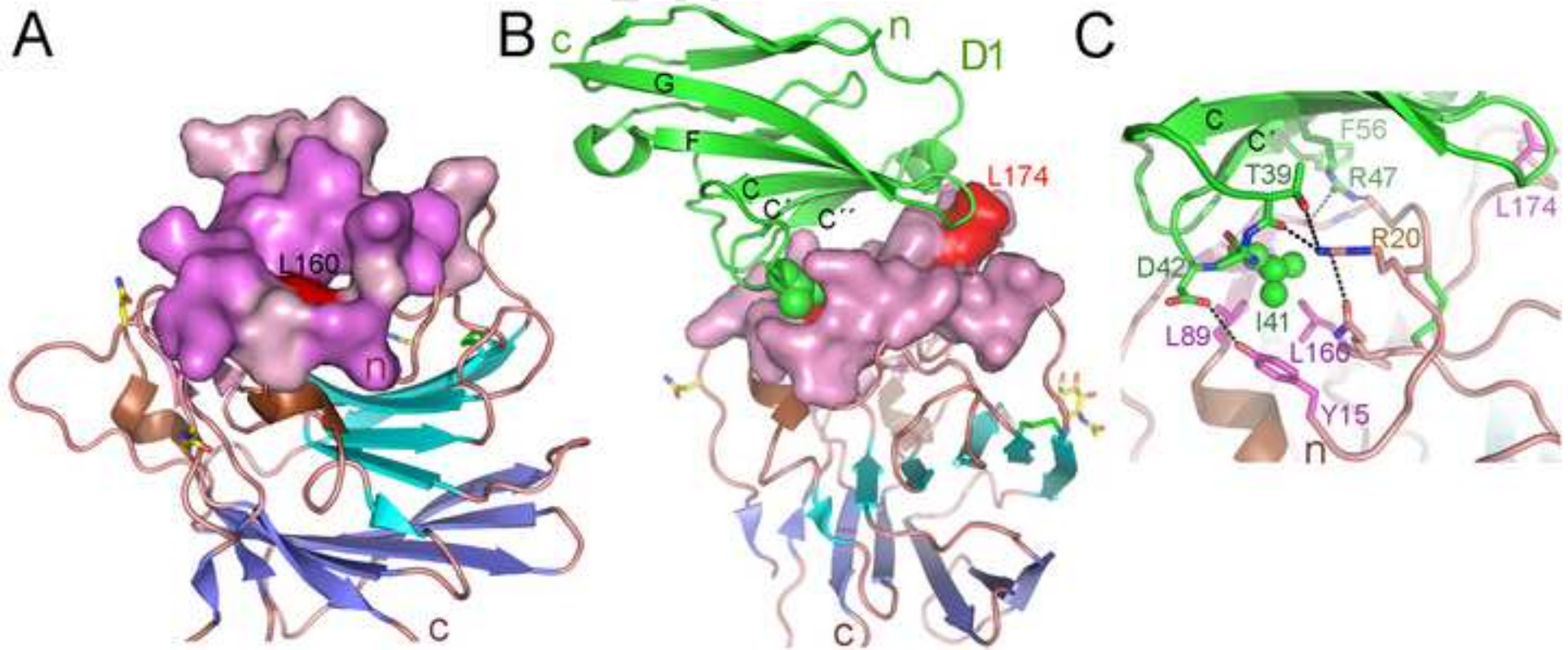


Figure10