

Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies

Fang Li

Department of Pharmacology, University of Minnesota Medical School, Minneapolis, Minnesota, USA

Receptor recognition by viruses is the first and essential step of viral infections of host cells. It is an important determinant of viral host range and cross-species infection and a primary target for antiviral intervention. Coronaviruses recognize a variety of host receptors, infect many hosts, and are health threats to humans and animals. The receptor-binding S1 subunit of coronavirus spike proteins contains two distinctive domains, the N-terminal domain (S1-NTD) and the C-terminal domain (S1-CTD), both of which can function as receptor-binding domains (RBDs). S1-NTDs and S1-CTDs from three major coronavirus genera recognize at least four protein receptors and three sugar receptors and demonstrate a complex receptor recognition pattern. For example, highly similar coronavirus S1-CTDs within the same genus can recognize different receptors, whereas very different coronavirus S1-CTDs from different genera can recognize the same receptor. Moreover, coronavirus S1-NTDs can recognize either protein or sugar receptors. Structural studies in the past decade have elucidated many of the puzzles associated with coronavirus-receptor interactions. This article reviews the latest knowledge on the receptor recognition mechanisms of coronaviruses and discusses how coronaviruses have evolved their complex receptor recognition pattern. It also summarizes important principles that govern receptor recognition by viruses in general.

Coronaviruses (CoV) are a group of common, ancient, and diverse viruses. They infect many mammalian and avian species and cause respiratory, gastrointestinal, and central nervous system diseases (1, 2). Coronavirus virions contain an envelope, a helical capsid, and a single-stranded and positive-sense RNA genome. The length of their genomes, which are the largest among all RNA viruses, typically ranges between 27 and 32 kb. They were named “coronaviruses” because of the protruding spike proteins on their envelope that give the virions a crown-like shape (“corona” in Latin means crown). Coronaviruses belong to the *Coronaviridae* family in the order of *Nidovirales*. They can be classified into at least three major genera, α , β , and γ (formerly group 1, 2, and 3, respectively) (3). Prototypic α -genus coronaviruses include human coronavirus NL63 (HCoV-NL63), porcine transmissible gastroenteritis coronavirus (TGEV), and porcine respiratory coronavirus (PRCV). Prototypic β -genus coronaviruses include severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), mouse hepatitis coronavirus (MHV), and bovine coronavirus (BCoV). Prototypic γ -genus coronaviruses include avian infectious bronchitis virus (IBV). These three major coronavirus genera and their prototypic coronaviruses are the focus of this review article (Fig. 1).

Coronaviruses impose health threats to humans and animals. Two β -coronaviruses, SARS-CoV and MERS-CoV, are highly pathogenic human pathogens. SARS-CoV caused the SARS epidemic in 2002 to 2003, with over 8,000 infections and a fatality rate of ~10% (4–7). MERS-CoV emerged from the Middle East in 2012. As of 16 October 2014, MERS-CoV had caused 877 infections with a fatality rate of ~36% (<http://www.who.int/csr/don/16-october-2014-mers/en/>) (8, 9). HCoV-NL63 from the α -genus is a prevalent human respiratory pathogen that is often associated with common colds in healthy adults and acute respiratory diseases in young children (10, 11). Among the animal coronaviruses, TGEV from the α -genus and MHV from the β -genus cause close to 100% fatality in young pigs and young mice, respectively (12–15); BCoV from the β -genus and IBV from the γ -genus also cause significant health damage in cattle

and chickens, respectively (16–19). Therefore, research on coronaviruses has strong health and economic implications.

Receptor recognition by viruses is the first and essential step of viral infections of host cells (20). An envelope-anchored spike protein mediates coronavirus entry into host cells by first binding to a receptor on the host cell surface and then fusing viral and host membranes (21, 22). A member of the class I viral membrane fusion proteins (23–26), the coronavirus spike consists of three segments—an ectodomain, a single-pass transmembrane anchor, and a short intracellular tail (27, 28). The ectodomain can be divided into a receptor-binding S1 subunit and a membrane-fusion S2 subunit. The amino acid sequences of S1 diverge across different genera but are relatively conserved within each genus (29). S1 contains two independent domains, an N-terminal domain (S1-NTD) and a C-terminal domain (S1-CTD, also called S1 C-domain) (Fig. 1) (29). Either or both of these S1 domains can function as a receptor-binding domain (RBD). The binding interaction between coronavirus RBD and its receptor is one of the most important determinants of the coronavirus host range and cross-species infection (2, 30). In addition, coronavirus RBDs contain major neutralization epitopes, induce most of the host immune responses, and may serve as subunit vaccines against coronavirus infections (31–36). Knowledge about the receptor recognition mechanisms of coronaviruses is critical for understanding coronavirus pathogenesis and epidemics and for human intervention in coronavirus infections.

Coronaviruses recognize a variety of host receptors (Fig. 1).

Accepted manuscript posted online 26 November 2014

Citation Li F. 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* 89:1954–1964. doi:10.1128/JVI.02615-14.

Editor: S. P. Goff

Address correspondence to lifang@umn.edu.

Copyright © 2015, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JVI.02615-14

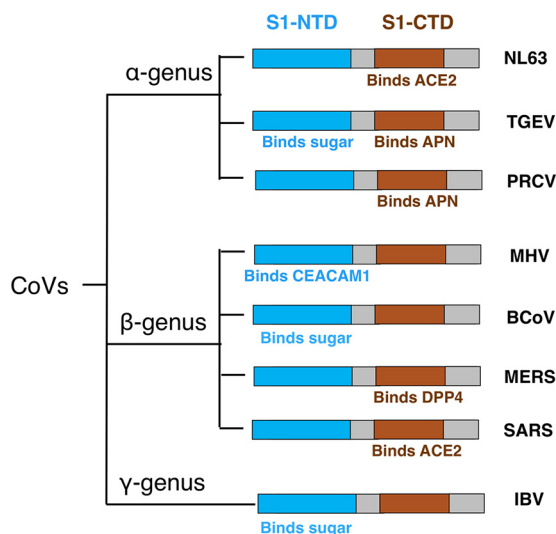


FIG 1 Receptor recognition pattern of coronaviruses.

Although HCoV-NL63 and SARS-CoV belong to the α -genus and β -genus, respectively, their S1-CTDs recognize the same receptor, angiotensin-converting enzyme 2 (ACE2) (37–43). Although HCoV-NL63, TGEV, and PRCV all belong to the α -genus, their S1-CTDs recognize different receptors—TGEV and PRCV S1-CTDs both recognize aminopeptidase N (APN) (44, 45). Similarly, although SARS-CoV and MERS-CoV both belong to the β -genus, their S1-CTDs recognize different receptors—MERS-CoV S1-CTD recognizes dipeptidyl peptidase 4 (DPP4) (46–48). Although MHV and BCoV both belong to the β -genus, their S1-NTDs recognize carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) and sugar, respectively (49–53). In addition, the S1-NTDs of α -genus TGEV and γ -genus IBV also recognize sugar (52, 54–58). Overall, coronaviruses have evolved a complex receptor recognition pattern: (i) coronaviruses use one or both S1 domains as RBDs; (ii) highly similar coronavirus S1-CTDs within the same genus can recognize different protein receptors, whereas very different coronavirus S1-CTDs from different genera can recognize the same protein receptor; and (iii) coronavirus S1-NTDs can recognize either protein or sugar receptors. Understanding the receptor recognition mechanisms of coronaviruses can provide critical insight into the origin, evolution, and receptor selection of coronaviruses.

In addition to their viral receptor functions, the receptors for coronaviruses have their own physiological functions. ACE2 is a zinc-dependent carboxypeptidase that cleaves one residue from the C terminus of angiotensin peptides and functions in blood pressure regulation (59–62). ACE2 also protects against severe acute lung failure, and SARS-CoV-induced downregulation of ACE2 promotes lung injury (63, 64). APN is a zinc-dependent aminopeptidase that cleaves one residue from the N terminus of many physiological peptides and plays multifunctional roles such as in pain regulation, blood pressure regulation, and tumor cell angiogenesis (65, 66). DPP4 is a serine exoprotease that cleaves two residues from the N terminus of many physiological peptides and functions in immune regulation, signal transduction, and apoptosis (67–70). CEACAM1 is a cell adhesion molecule and functions in cell-cell adhesion (71–73). Sugars decorate many

proteins and fats on cell surfaces and function in many biological processes such as immunity and cell-cell communication (57, 74, 75). How these cell-surface molecules are selected by viruses as their entry receptors has been a major puzzle in virology.

Analyses of crystal structures of coronavirus S1 domains and their complexes with their respective receptor have elucidated many puzzles associated with coronavirus-receptor interactions. Since the SARS epidemic, the crystal structures of five coronavirus S1 domains complexed with their respective receptor have been determined. These are the β -genus SARS-CoV S1-CTD complexed with human ACE2 (76), β -genus MERS-CoV S1-CTD complexed with human DPP4 (77, 78), α -genus HCoV-NL63 S1-CTD complexed with human ACE2 (79), α -genus PRCV S1-CTD complexed with porcine APN (80), and β -genus MHV S1-NTD complexed with murine CEACAM1 (81). In addition, the crystal structure of β -genus BCoV S1-NTD by itself has been determined, with its sugar-binding site identified through mutagenesis (53). These six representative structures not only reveal how coronaviruses recognize their receptors in atomic details but also shed light on how coronaviruses do so using complicated evolutionary strategies. Other than these six representative structures, several variant forms of these structures have also been determined, including S1-CTDs of different SARS-CoV strains complexed with ACE2 from animals and S1-CTD of a MERS-CoV-related bat coronavirus HKU4 complexed with human DPP4 (82–84). This article reviews these structural studies and their implications for the receptor recognition and evolution of coronaviruses.

S1-CTDs OF β -GENUS CORONAVIRUSES

β -genus SARS-CoV S1-CTD complexed with human ACE2 was the first crystal structure determined for a coronavirus S1 domain and S1 domain/receptor complex (Fig. 2A) (76, 85). SARS-CoV S1-CTD contains two subdomains: a core structure and an extended loop. The core structure consists of a five-stranded antiparallel β -sheet and several short connecting α -helices. The extended loop lies on one edge of the core structure and forms a gently concave surface with two ridges on both sides and a two-stranded antiparallel β -sheet sitting in the middle (Fig. 3A and B). Because this extended loop makes all the contacts with ACE2, it has been termed receptor-binding motif (RBM). On the other hand, the peptidase domain of ACE2 has a claw-like structure with two lobes. The enzymatic active site of ACE2 is buried in a cavity surrounded by the two lobes. SARS-CoV S1-CTD binds to the outer surface of the N-terminal lobe, away from the peptidase active site. Consequently, SARS-CoV binding has no effect on the enzymatic activity of ACE2 and vice versa. The SARS-CoV-binding region on the ACE2 surface has been termed virus-binding motif (VBM). The RBM and VBM complement each other in shape and chemical details. The structure of SARS-CoV S1-CTD/ACE2 complex provided the first view of coronavirus S1 and S1/receptor complex and laid the foundation for future structural and evolutionary comparisons with other coronavirus S1 and S1/receptor complexes.

Comparative studies of the interactions between the S1-CTD from different SARS-CoV strains and ACE2 from different host species have elucidated the molecular and structural mechanisms by which SARS-CoV transmitted from animals to humans and caused the SARS epidemic (30, 83, 84, 86–89). Two virus-binding hot spots have been identified in the VBM of ACE2, one centering on ACE2 residue Lys31 and the other centering on ACE2 residue

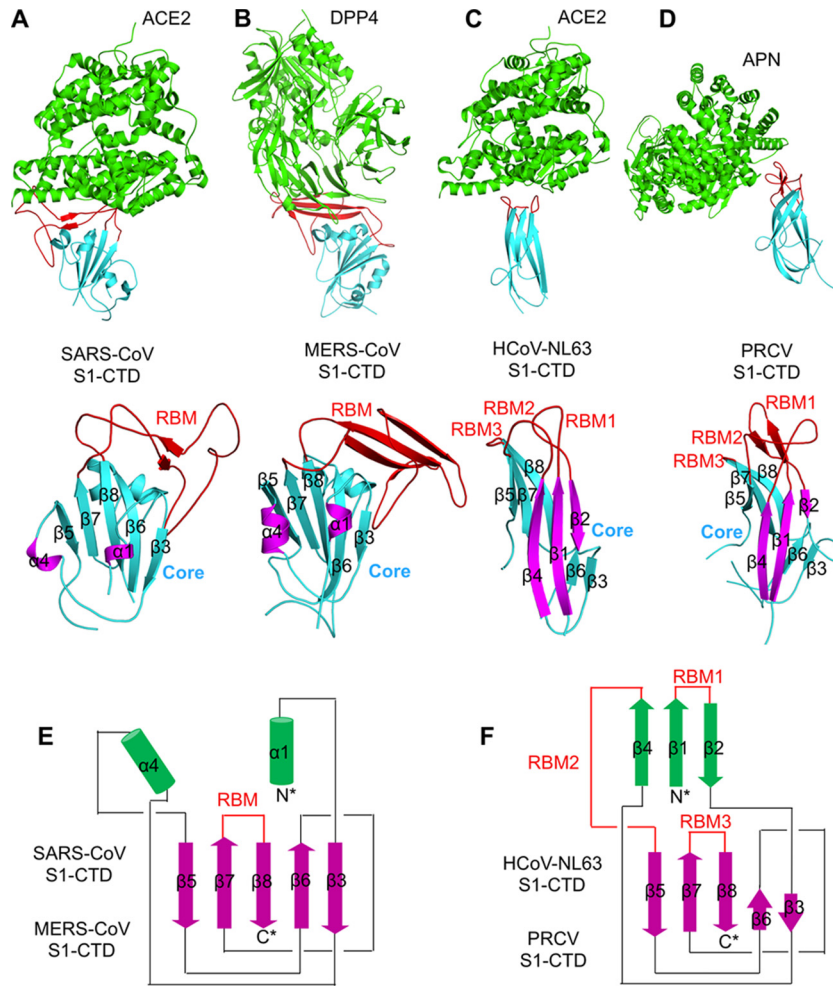


FIG 2 Crystal structures of coronavirus S1-CTDs complexed with their respective receptor. (A to D) These structures include β -genus SARS-CoV S1-CTD complexed with ACE2 (Protein Data Bank identifier [PDB ID]: 2AJF) (76) (A), β -genus MERS-CoV S1-CTD complexed with DPP4 (PDB ID: 4KR0) (77) (B), α -genus HCoV-NL63 S1-CTD complexed with ACE2 (PDB ID: 3KBH) (79) (C), and α -genus PRCV S1-CTD complexed with APN (PDB ID: 4F5C) (80) (D). In the structures of the complexes, the receptors are in green, and the cores and RBMs of S1-CTDs are in cyan and red, respectively. (E and F) The structural topologies of the four coronavirus S1-CTDs are shown as schematic illustrations, where β -strands are depicted as arrows and α -helices as cylinders. In the tertiary structures and structural topologies of the S1-CTDs, the secondary structures of all of the S1-CTDs are colored and numbered in the same way as for HCoV-NL63 S1-CTD.

Lys353 (Fig. 3C and D). Both of these virus-binding hot spots consist of a salt bridge that is buried in a hydrophobic environment. Structure-guided functional studies revealed that both virus-binding hot spots provide significant energy to the virus-receptor binding interactions (90). Indeed, all of the naturally selected viral mutations in SARS-CoV RBM surround the two hot spots, with significant impact on the structures of the hot spots, the ACE2 binding affinity, and the host immune responses (84, 91). One of these viral mutations, K479N, facilitated transmission of SARS-CoV from palm civets to humans. Another viral mutation, S487T, facilitated transmission of SARS-CoV from human to human. These two mutations contributed significantly to the SARS epidemic in 2002 to 2003. The S1-CTD of a SARS-CoV-related Rs3367 bat coronavirus contains two asparagines at these two positions (corresponding to positions 479 and 487 in human SARS-CoV strains) (92). The first asparagine is favorable for human ACE2 binding, and the second one is less favorable. Thus, Rs3367 recognizes human ACE2 but probably less well than the

human SARS-CoV strains do. For more details about how the structural analysis of SARS-CoV RBD/ACE2 interactions has provided insight into the SARS epidemic, please refer to another recent review article on this topic (30). These structural studies of SARS-CoV S1-CTD/ACE2 interactions demonstrate that it is critical to understand viral evolution, cross-species transmission, and epidemics within a detailed structural framework.

The crystal structures of β -genus MERS-CoV S1-CTD by itself and in complex with human DPP4 provided another view of coronavirus S1 and S1/receptor complex (Fig. 2B) (77, 78, 93). Like SARS-CoV S1-CTD, MERS-CoV S1-CTD also contains a core structure and an RBM. The core structures of MERS-CoV and SARS-CoV S1-CTDs are highly similar to each other, but their RBMs are markedly different, leading to different receptor specificities. The RBM of MERS-CoV S1-CTD mainly consists of a four-stranded β -sheet, in contrast to the loop-dominated RBM in SARS-CoV S1-CTD. Like the VBM for SARS-CoV on ACE2, the VBM for MERS-CoV is also located on the outer surface of DPP4,

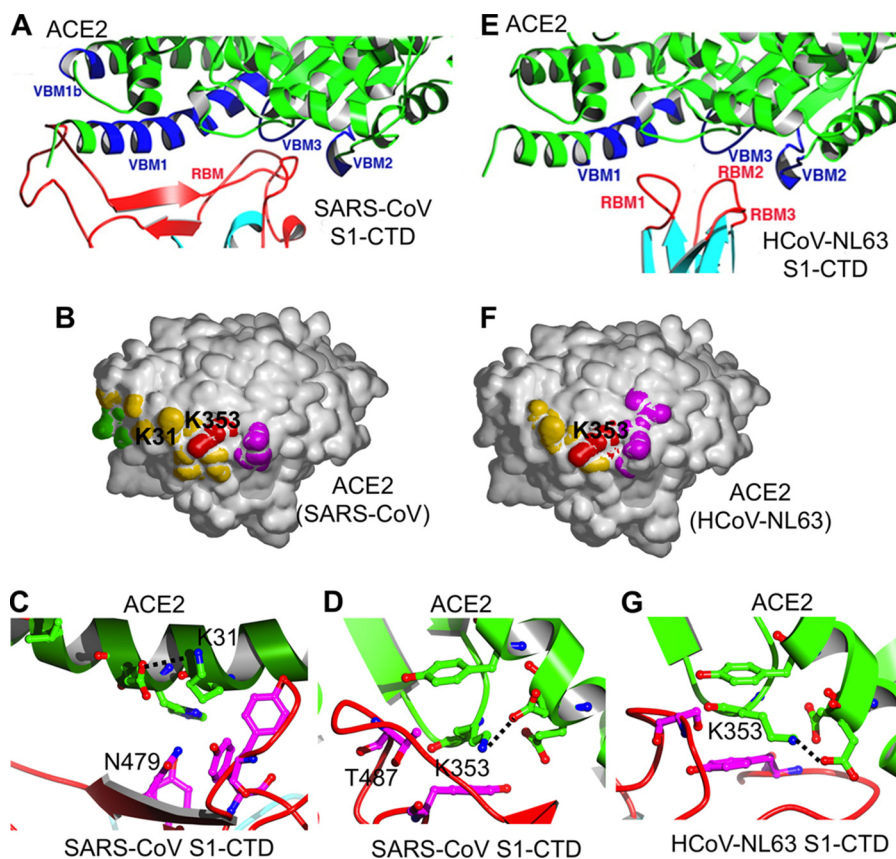


FIG 3 Virus-binding hot spots on ACE2 that are critical for the binding of SARS-CoV and HCoV-NL63. (A) Enlarged view of the SARS-CoV-ACE2 interface. VBMs on ACE2 and RBM on SARS-CoV S1-CTD are in blue and red, respectively. (B) Footprint of SARS-CoV on the surface of ACE2. The view is derived from the one in panel A by rotating ACE2 by 90° along a horizontal axis in such a way that the edge facing the viewer moves up. VBM1 residues are in orange, VBM2 residues in magenta, VBM3 residues in red, and VBM1b residues in green. (C) A virus-binding hot spot on ACE2 centering on Lys31 is critical for the binding of SARS-CoV S1-CTD. Mutation of residue 479 on SARS-CoV S1-CTD is critical for the transmission of SARS-CoV from palm civets to humans. (D) A second virus-binding hot spot on ACE2 centering on Lys353 is also critical for the binding of SARS-CoV S1-CTD. Mutation of residue 487 on SARS-CoV S1-CTD is critical for the transmission of SARS-CoV from human to human. (E) Enlarged view of the HCoV-NL63-ACE2 interface. (F) Footprint of HCoV-NL63 on the surface of ACE2. (G) The same virus-binding hot spot on ACE2 centering on Lys353 is also critical for the binding of HCoV-NL63 S1-CTD.

away from the peptidase active site. Whereas the conserved core structures of SARS-CoV and MERS-CoV S1-CTDs suggest a common evolutionary origin, the different RBMs of the two S1-CTDs indicate a divergent evolutionary pathway that has led to their recognition of different host receptors. The S1-CTDs of MERS-CoV and a highly related bat coronavirus HKU4 recognize DPP4 in very similar ways, suggesting a close evolutionary relationship between the two viruses (82, 94). In addition to enhancing the understanding of coronavirus evolution, the structure of MERS-CoV S1-CTD/DPP4 complex has important implications for understanding the host range and cross-species transmission of MERS-CoV (82, 94–97).

S1-CTDs OF α -GENUS CORONAVIRUSES

α -Genus HCoV-NL63 S1-CTD complexed with human ACE2 was the first crystal structure determined for an α -coronavirus S1 domain (Fig. 2C) (79). This structure, along with the structure of β -genus SARS-CoV S1-CTD complexed with ACE2, provided the first view of how two different viruses recognize their common host receptor. The finding was intriguing. At first glance, HCoV-NL63 and SARS-CoV S1-CTDs are very different. The core structure of HCoV-NL63 S1-CTD is a β -sandwich consisting of two

β -sheet layers stacked together through hydrophobic interactions, which is in contrast to the single β -sheet layer in the core structure of SARS-CoV S1-CTD. Their RBMs are also different. The RBMs of HCoV-NL63 S1-CTD are three short and discontinuous loops, whereas the RBM of SARS-CoV S1-CTD is a single long and continuous subdomain. Indeed, the protein-folding Dali server failed to detect any structural similarity between HCoV-NL63 and SARS-CoV S1-CTDs (98). However, structural topology analysis revealed that the secondary structural elements in HCoV-NL63 S1-CTD are connected in the same way as those in SARS-CoV S1-CTD, although two β -strands in the former (strands β -1 and β -4) become α -helices in the latter (helices α -1 and α -4) and another β -strand (strand β -1) in the former is missing altogether in the latter (Fig. 2E and F) (29). These results suggest that HCoV-NL63 and SARS-CoV S1-CTDs share an evolutionary origin and that the structural differences between the two S1-CTDs result from extensive divergent evolution.

Despite their different tertiary structures, HCoV-NL63 and SARS-CoV S1-CTDs bind to a common region on ACE2 (79, 90). The VBMs for the two viruses on ACE2 overlap, and a number of ACE2 residues interact with both S1-CTDs (Fig. 3E and F). Surprisingly, one of the two virus-binding hot spots on ACE2 for

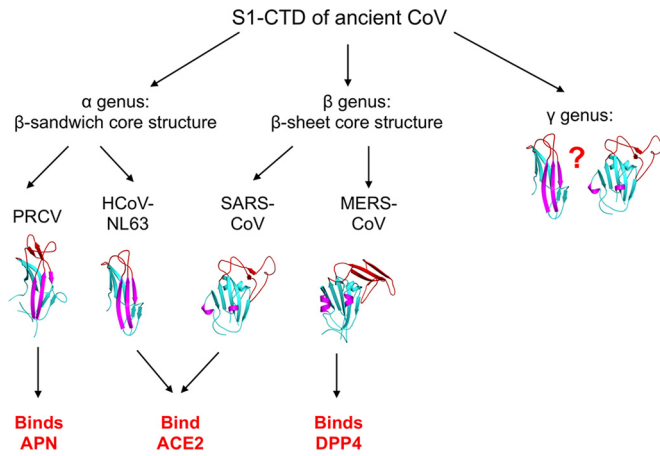


FIG 4 Proposed origin and evolution of coronavirus S1-CTDs. The question mark indicates possible tertiary structures of γ -coronavirus S1-CTD.

SARS-CoV binding, which centers on ACE2 residue Lys353, plays a similarly critical role in the binding of HCoV-NL63 (Fig. 3G). Disturbance of the hot spot structure via mutagenesis decreased or abolished the binding of both viruses. Hence, Lys353 and the nearby residues on ACE2 form a common virus-binding hot spot that is critical for the attachment of two different coronaviruses. On the other hand, among the three RBMs in HCoV-NL63 S1-CTD, only RBM1 and RBM2, but not RBM3, are involved in binding the common virus-binding hot spot on ACE2, despite the fact that RBM3 is topologically equivalent to the RBM in SARS-CoV S1-CTD (Fig. 2A, C, E, and F). The different molecular mechanisms used by the two S1-CTDs to recognize ACE2 suggest a convergent evolutionary relationship between the two S1-CTDs (i.e., the two S1-CTDs evolved independently to recognize the same virus-binding hot spot on ACE2), although a divergent evolutionary relationship cannot be completely ruled out (i.e., the two S1-CTDs both evolved from a common ancestral protein that bound ACE2). Therefore, after HCoV-NL63 and SARS-CoV S1-CTDs underwent divergent evolution to attain different structures, they might have further converged to recognize the same region on the same receptor. The common virus-binding hot spot on ACE2 might be the driving force for this later convergent evolution.

The crystal structure of α -genus PRCV S1-CTD complexed with porcine APN illustrated how another similar α -coronavirus S1-CTD recognizes a different host receptor (Fig. 2D) (80). Similarly to the structural relationship between SARS-CoV and MERS-CoV S1-CTDs, PRCV and HCoV-NL63 S1-CTDs also have highly similar core structures. However, their three RBMs are divergent, leading to different receptor specificities. Similarly to the VBMs on ACE2 and DPP4, the VBMs for PRCV on APN are also located on the outer surface of APN, away from the peptidase active site. Overall, these results suggest that PRCV and HCoV-NL63 S1-CTDs share an evolutionary origin but have diverged in their RBM loops to recognize different host receptors.

We propose the following evolutionary scenario for coronavirus S1-CTDs (Fig. 4). All coronavirus S1-CTDs likely shared one evolutionary origin, as evidenced by their related structural topologies across different genera (Fig. 2E and F). Through divergent evolution, coronavirus S1-CTDs attained β -sandwich core struc-

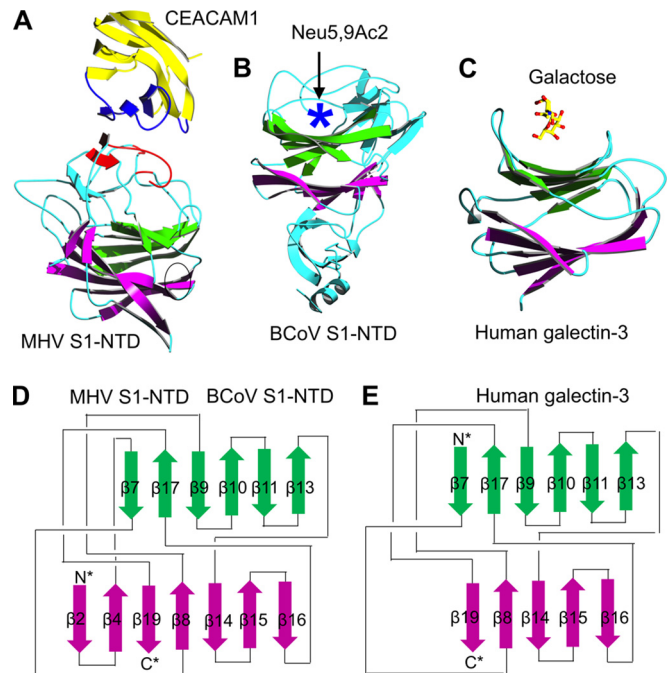


FIG 5 Crystal structures of coronavirus S1-NTDs. (A and B) These structures include β -genus MHV S1-NTD complexed with its receptor CEACAM1 (PDB ID: 3R4D) (81) (A) and β -genus BCoV S1-NTD by itself (PDB ID: 4H14) (53) (B). (C) The structure of human galectin-3 (PDB ID: 1A3K) is shown as a comparison. In the structure of the MHV S1-NTD-CEACAM1 complex, the VBM on CEACAM1 and the RBM on MHV S1-NTD are in blue and red, respectively. In the structure of BCoV S1-NTD, the sugar-binding pocket as identified by mutagenesis studies is indicated by a five-pointed star. (D and E) The structural topologies of the two coronavirus S1-NTDs (D) and human galectin-3 (E) are shown as schematic illustrations, where β -strands are depicted as arrows and α -helices as cylinders. In the structural topologies of these proteins, the secondary structures are colored and numbered in the same way as for the MHV S1-NTD.

tures in the α -genus and β -sheet core structures in the β -genus. Although the structures of γ -coronavirus S1-CTDs are not known, their core structures may also have a topology related to those of α - and β -coronavirus S1-CTDs. Furthermore, α -coronavirus S1-CTDs diverged in the three RBM loops to acquire different receptor specificities—ACE2 specificity for HCoV-NL63 and APN specificity for PRCV. β -Coronavirus S1-CTDs also diverged in the RBM subdomain to acquire different receptor specificities—ACE2 specificity for SARS-CoV and DPP4 specificity for MERS-CoV. The S1-CTDs of α -genus HCoV-NL63 and β -genus SARS-CoV first diverged into different tertiary structures but later converged to recognize the same receptor ACE2. In sum, coronavirus S1-CTDs have undergone convoluted structural evolutions, leading to their complex receptor recognition pattern.

S1-NTDs OF β -GENUS CORONAVIRUSES

β -Genus MHV S1-NTD complexed with mouse CEACAM1 was the first structure available for a coronavirus S1-NTD and S1-NTD/receptor complex (Fig. 5A) (81). Surprisingly, MHV S1-NTD contains a core structure that has the same structural fold as human galectins (galactose-binding lectins) (Fig. 5C) (99). The core structure of MHV S1-NTD is a thirteen-stranded β -sandwich consisting of two β -sheet layers of six and seven strands, respectively. The structural topologies of MHV S1-NTD and hu-

man galectins are identical, except that MHV S1-NTD contains two additional β -strands in one of the β -sheet layers (Fig. 5D and E). Compared with human galectins, MHV S1-NTD contains additional structural motifs on top of the core that form a ceiling-like structure. The outer surface of this ceiling-like structure functions as RBM by binding to the VBM on the N-terminal Ig-like domain of CEACAM1. Despite its galectin fold, MHV S1-NTD does not bind sugars, as revealed by sugar-binding assays. Moreover, neither the RBM on MHV S1-NTD nor the VBM on CEACAM1 contains any sugar at the binding interface. Instead, MHV S1-NTD binds to CEACAM1 through exclusive protein-protein interactions. A hydrophobic patch in the VBM of CEACAM1 functions as a virus-binding hot spot; mutations in this region significantly decreased the binding of MHV S1-NTD (81, 100–102). Taken together, these results suggest that MHV S1-NTD and host galectins share the same evolutionary origin; they also indicate that although MHV S1-NTD binds only a CEACAM1 protein receptor, other coronavirus S1-NTDs may bind sugar receptors and function as viral lectins.

Analysis of the crystal structure of β -genus BCoV S1-NTD provided the first view of a functional lectin domain in a coronavirus spike (Fig. 5B) (53). The overall structure of BCoV S1-NTD is highly similar to that of MHV S1-NTD, also containing a galectin-like core and a ceiling-like structure on top of the core. In contrast to MHV S1-NTD, which binds CEACAM1 but not sugars, BCoV S1-NTD binds a sugar receptor but not CEACAM1. Glycan screen arrays identified Neu5,9Ac2 (5-N-acetyl-9-O-acetylneuraminic acid) as the sugar receptor for BCoV S1-NTD. Although the structure of a sugar-bound BCoV S1-NTD is not available, structure-guided mutagenesis has revealed that the sugar-binding site is located in a pocket surrounded by the core and the ceiling-like structure on top of the core. The sugar-binding sites in BCoV S1-NTD and human galectins overlap, although human galectins recognize a different sugar receptor, galactose. Structural comparison between MHV and BCoV S1-NTDs revealed that subtle structural changes between the two S1-NTDs, mainly involving different conformations of RBM loops, explain why BCoV S1-NTD does not bind CEACAM1 and why MHV S1-NTD does not bind sugars. These results suggest that MHV and BCoV S1-NTDs are both evolutionarily related to human galectins but that they have diverged from human galectins with specificities for a novel protein receptor and a different sugar receptor, respectively.

We propose the following evolutionary scenario for coronavirus S1-NTDs (Fig. 6). Ancestral coronaviruses stole a host galectin gene and inserted it into the 5' end of their spike gene, which became coronavirus S1-NTD. Since then, coronavirus S1-NTDs have undergone divergent evolution in three genera. β -Genus BCoV S1-NTD has kept the lectin activity but evolved specificity for a different sugar receptor, Neu5,9Ac2. Although the crystal structures of α - and γ -coronavirus S1-NTDs are not available, they may also have the galectin fold for the following reasons. First, the conserved structural topology of S1-CTDs across different coronavirus genera strongly suggests a similarly conserved structural topology of S1-NTDs across different coronavirus genera. Second, the S1-NTDs of both α -genus TGEV and γ -genus IBV function as lectins, although the former recognizes both N-glycolylneuraminic acid (Neu5Gc) and N-acetylneuraminic acid (Neu5Ac) and the latter recognizes Neu5Gc. Hence, sugar-binding S1-NTDs across different coronavirus genera may share the

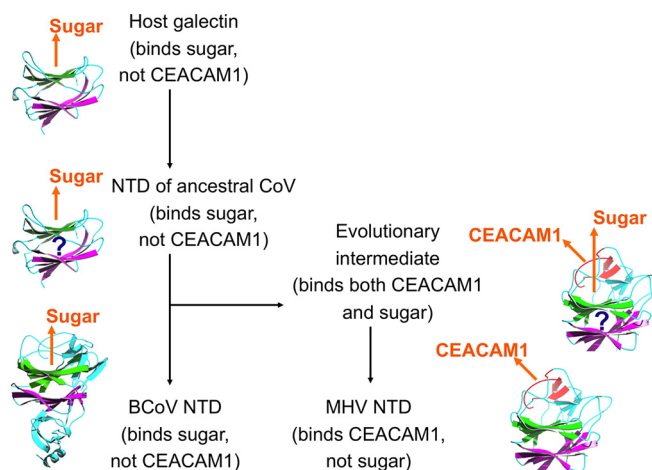


FIG 6 Proposed origin and evolution of coronavirus S1-NTDs. Orange arrows indicate the locations of CEACAM1 or sugar that binds coronavirus NTDs. Question marks indicate the postulated structures of hypothetical evolutionary intermediates.

same galectin fold but have diverged to recognize different sugar receptors. On the other hand, β -genus MHV S1-NTD has evolved specificity for a novel protein receptor, CEACAM1. Subsequently, MHV S1-NTD lost its lectin activity because proteins in general have advantages over sugars as viral receptors by providing higher affinity and specificity for viral attachment.

Are coronaviruses the only viruses that stole a host lectin and integrated it into their spike? A survey of viral lectins with known tertiary structures revealed that galectin-like domains are present in a variety of viral spikes, including influenza virus hemagglutinin, whose galectin-like fold was previously unknown (24, 103). Moreover, these viral lectins display diverse sugar-binding modes, but they share a feature—their sugar-binding sites are all located in cavities and are not easily accessible to host antibodies and immune cells. As a comparison, the sugar-binding sites in host galectins are open and easily accessible (Fig. 5C). It was thus hypothesized that these viral lectins all originated from host galectins but have evolved to use hidden sugar-binding sites to evade host immune surveillance (104). The above analysis may explain why coronavirus S1-NTDs have evolved the ceiling-like structure on top of the core, which is used to protect the sugar-binding site in coronavirus S1-NTDs from the host immune system. Subsequently, MHV S1-NTD took advantage of the ceiling-like structure and evolved CEACAM1-binding RBM on the outer surface of this ceiling-like structure. In this sense, the evolution of CEACAM1-binding RBM in MHV S1-NTD might be an indirect outcome of the efforts of coronaviruses to battle the host immune attacks.

RECEPTOR BINDING BY CORONAVIRUSES

So far, we have reviewed the receptor recognition and evolution of coronavirus S1-NTDs and S1-CTDs separately. How do S1-NTDs and S1-CTDs work together in the receptor recognition and evolution of coronavirus spikes? Electron microscopic studies of the SARS-CoV spike revealed that it is a clove-shaped trimer, with three individual S1 heads and a trimeric S2 stalk (Fig. 7) (27, 28). ACE2 binds to the tip of the SARS-CoV spike trimer, where S1-CTD is located. Because the membrane-distal tips of the trimeric

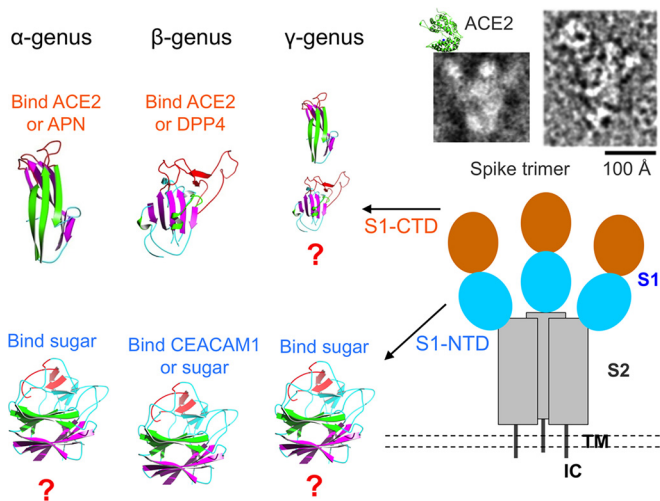


FIG 7 Summary of the receptor recognition mechanisms of coronaviruses in a three-dimensional view. The overall structure of trimeric SARS-CoV spike complexed with ACE2 is shown; it includes both the schematic topology of the spike and the negative-stain electron micrographs of the spike ectodomain (upper right). TM, transmembrane anchor. IC, intracellular tail. The structures and functions of coronavirus S1 domains are listed. The question marks indicate possible tertiary structures of coronavirus S1 domains.

spike are the most exposed and protruding region on the whole spike, S1-CTD is directly exposed to the host immune system, evolves at an increased pace to evade the host immune surveillance, and becomes hypervariable in primary, secondary, and tertiary structures. The RBM of S1-CTD is located on the very tip of the trimeric spikes and evolves at the fastest pace. On the other hand, S1-NTD is likely located underneath S1-CTD, is less exposed to the host immune system, and evolves at a slower pace than S1-CTD. Therefore, between the two S1 domains, the more conserved S1-NTDs may function as the more reliable RBDs that recognize sugar receptors, allowing coronaviruses to search for additional and high-affinity protein receptors using their fast-evolving S1-CTDs. Such dual-RBD structures in coronavirus spikes may give coronaviruses an evolutionary advantage in finding new receptors and expanding their host ranges.

Why were specific host cell surface molecules selected as coronavirus receptors? Among the known coronavirus receptors, sugars are probably the primordial and fallback receptors for coronaviruses. Sugars are abundant on host cell surfaces and are easy targets for viruses to grab. To use sugars as their receptors, a variety of viruses might have stolen a host galectin and used it as a viral lectin. On the other hand, using protein receptors may enhance the affinity and specificity of viral attachment, increase the efficiency of viral entry, and facilitate viruses to expand their host ranges and alter their tropisms (105). Host cell surface proteins have some common features as viral receptors. First, they frequently undergo endocytosis, which facilitates viral entry. Second, they contain VBM on their surfaces for high-affinity virus binding. In the VBMs of both ACE2 and CEACAM1, virus-binding hot spots have been identified and contribute significant energy to virus/receptor binding interactions (79, 81, 90). Therefore, host cell surface molecules are not randomly selected by viruses as their receptors. In fact, there are structural and evolutionary reasons behind these selections by viruses.

CONCLUDING REMARKS

The structural studies of coronavirus-receptor interactions described above have established the following virology principles. First, drastic structural changes in viral RBDs can still lead to recognition of a virus-binding hot spot on the same receptor protein. Supporting this principle is the finding that SARS-CoV and HCoV-NL63 recognize a common virus-binding hot spot on ACE2 using structurally divergent S1-CTDs. Second, subtle structural changes in viral RBDs can lead to a complete receptor switch. For example, HCoV-NL63 and PRCV recognize two different protein receptors using structurally conserved S1-CTDs with divergent RBMs, and so do SARS-CoV and MERS-CoV. Moreover, MHV and BCoV S1-NTDs recognize a protein receptor and a sugar receptor, respectively, through subtle conformational changes in receptor-binding loops. Third, it is a successful viral strategy to steal a host protein and evolve it into viral RBDs with novel protein receptor specificities or altered sugar receptor specificities. For example, MHV and BCoV S1-NTDs have the same structural fold as human galectins, but they recognize a novel protein receptor and a different sugar receptor, respectively. Fourth, a few residue changes at the receptor binding interface can lead to efficient cross-species infection and human-to-human transmission of a virus. For example, SARS-CoV needed only one or two mutations in its RBD to transmit from palm civets to humans. These virology principles may be extended from the *Coronaviridae* family to other virus families.

What are the remaining important questions regarding receptor recognition mechanisms of coronaviruses? First, what are the crystal structures of α -coronavirus S1-NTDs, γ -coronavirus S1-NTDs, and γ -coronavirus S1-CTDs? We have hypothesized that α -coronavirus and γ -coronavirus S1-NTDs have a galectin fold and that γ -coronavirus S1-CTDs have either a β -sandwich fold or a β -sheet fold. These hypotheses need to be tested using experimentally determined crystal structures of these S1 domains. Second, what are the detailed sugar-binding mechanisms for coronavirus S1-NTDs? The crystal structures of coronavirus S1-NTDs complexed with sugar receptors will reveal how sugar receptor specificities are achieved in these viral lectins across different coronavirus genera. Third, why do coronaviruses rely on peptidases as their receptors? Three of the four known protein receptors for coronaviruses are peptidases: ACE2, APN, and DPP4. They are all recognized by S1-CTDs of different coronaviruses. It is highly unlikely that the use of peptidases as coronavirus receptors is simply a coincidence. On the other hand, these receptors' peptidase activities have no effects on coronavirus entry, indicating that their common physiological function in degrading peptides was not the reason why they were selected as coronavirus receptors. To fully understand why peptidases became chosen receptors for coronaviruses, it will be important in the future to comprehensively examine the physiological functions of these peptidase receptors. Last, what was the evolutionary origin of coronavirus S1-CTDs? So far, coronavirus S1-CTDs appear to have a novel fold not related to any other proteins in the protein structure database. However, our previous structural studies of coronavirus spikes repeatedly showed that tertiary structures of viral proteins can deceive the currently available tertiary structural analysis software (98). Instead, our structural topology analysis is a powerful tool to identify structural homology among viral proteins (29, 103). This approach may help identify the evolutionary origin of

coronavirus S1-CTDs. To sum up, structural studies in the past decade have elucidated many puzzles surrounding receptor recognition, evolution, and cross-species transmission of coronaviruses. Future structural studies will continue to solve the remaining puzzles as well as new puzzles that may emerge regarding the receptor recognition mechanisms of coronaviruses.

ACKNOWLEDGMENT

This work was supported by NIH grant R01AI089728.

REFERENCES

1. Perlman S, Netland J. 2009. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol* 7:439–450. <http://dx.doi.org/10.1038/nrmicro2147>.
2. Li WH, Wong SK, Li F, Kuhn JH, Huang IC, Choe H, Farzan M. 2006. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *J Virol* 80:4211–4219. <http://dx.doi.org/10.1128/JVI.80.9.4211-4219.2006>.
3. González JM, Gomez-Puertas P, Cavanagh D, Gorbalenya AE, Enjuanes L. 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol* 148:2207–2235. <http://dx.doi.org/10.1007/s00705-003-0162-1>.
4. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong SX, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966. <http://dx.doi.org/10.1056/NEJMoa030781>.
5. Peiris JSM, Lai ST, Poon LLM, Guan Y, Yam LYC, Lim W, Nicholls J, Yee WKS, Yan WW, Cheung MT, Cheng VCC, Chan KH, Tsang DNC, Yung RWH, Ng TK, Yuen KY. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361:1319–1325. [http://dx.doi.org/10.1016/S0140-6736\(03\)13077-2](http://dx.doi.org/10.1016/S0140-6736(03)13077-2).
6. Marra MA, James SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, Khattri J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JE, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, et al. 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300:1399–1404. <http://dx.doi.org/10.1126/science.1085953>.
7. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong SX, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TCT, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus A, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300:1394–1399. <http://dx.doi.org/10.1126/science.1085952>.
8. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus A, Fouchier RAM. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820. <http://dx.doi.org/10.1056/NEJMoa1211721>.
9. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, Perlman S, Poon LL, Snijder EJ, Stephens GM, Woo PC, Zaki AM, Zambon M, Ziebuhr J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol* 87:7790–7792. <http://dx.doi.org/10.1128/JVI.01244-13>.
10. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJM, Wolthers KC, Wertheim-van Dillen PME, Kaandorp J, Spaargaren J, Berkhout B. 2004. Identification of a new human coronavirus. *Nat Med* 10:368–373. <http://dx.doi.org/10.1038/nm1024>.
11. Fouchier RAM, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, Osterhaus A. 2004. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A* 101:6212–6216. <http://dx.doi.org/10.1073/pnas.0400762101>.
12. Tajima M. 1970. Morphology of transmissible gastroenteritis virus of pigs. A possible member of coronaviruses. Brief report. *Arch Gesamte Virusforsch* 29:105–108.
13. Laude H, Rasschaert D, Delmas B, Godet M, Gelfi J, Charley B. 1990. Molecular biology of transmissible gastroenteritis virus. *Vet Microbiol* 23:147–154. [http://dx.doi.org/10.1016/0378-1135\(90\)90144-K](http://dx.doi.org/10.1016/0378-1135(90)90144-K).
14. Taguchi F, Hirano N, Kiuchi Y, Fujiwara K. 1976. Difference in response to mouse hepatitis virus among susceptible mouse strains. *Jpn J Microbiol* 20:293–302. <http://dx.doi.org/10.1111/j.1348-0421.1976.tb00991.x>.
15. Hirai A, Ohtsuka N, Ikeda T, Taniguchi R, Blau D, Nakagaki K, Miura HS, Ami Y, Yamada YK, Itohara S, Holmes KV, Taguchi F. 2010. Role of mouse hepatitis virus (MHV) receptor murine CEACAM1 in the resistance of mice to MHV infection: studies of mice with chimeric mCEACAM1a and mCEACAM1b. *J Virol* 84:6654–6666. <http://dx.doi.org/10.1128/JVI.02680-09>.
16. Langpap TJ, Bergeland ME, Reed DE. 1979. Coronaviral enteritis of young calves: virologic and pathologic findings in naturally occurring infections. *Am J Vet Res* 40:1476–1478.
17. Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, Nauwynck H, Pensaert M, Van Ranst M. 2006. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* 80:7270–7274. <http://dx.doi.org/10.1128/JVI.02675-05>.
18. Cavanagh D. 2007. Coronavirus avian infectious bronchitis virus. *Vet Res* 38:281–297. <http://dx.doi.org/10.1051/vetres:2006055>.
19. Cavanagh D. 2001. A nomenclature for avian coronavirus isolates and the question of species status. *Avian Pathol* 30:109–115. <http://dx.doi.org/10.1080/03079450120044506>.
20. Baranowski E, Ruiz-Jarabo CM, Domingo E. 2001. Evolution of cell recognition by viruses. *Science* 292:1102–1105. <http://dx.doi.org/10.1126/science.1058613>.
21. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. 2003. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *J Virol* 77:8801–8811. <http://dx.doi.org/10.1128/JVI.77.16.8801-8811.2003>.
22. Spaan W, Cavanagh D, Horzinek MC. 1988. Coronaviruses: structure and genome expression. *J Gen Virol* 69(Pt 12):2939–2952. <http://dx.doi.org/10.1099/0022-1317-69-12-2939>.
23. Skehel JJ, Wiley DC. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69:531–569. <http://dx.doi.org/10.1146/annurev.biochem.69.1.531>.
24. Wilson IA, Skehel JJ, Wiley DC. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289:366–373. <http://dx.doi.org/10.1038/289366a0>.
25. Eckert DM, Kim PS. 2001. Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* 70:777–810. <http://dx.doi.org/10.1146/annurev.biochem.70.1.777>.
26. Harrison SC. 2008. Viral membrane fusion. *Nat Struct Mol Biol* 15:690–698. <http://dx.doi.org/10.1038/nsmb1456>.
27. Li F, Berardi M, Li WH, Farzan M, Dormitzer PR, Harrison SC. 2006. Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J Virol* 80:6794–6800. <http://dx.doi.org/10.1128/JVI.02744-05>.
28. Beniac DR, Andonov A, Grudski E, Booth TF. 2006. Architecture of the SARS coronavirus prefusion spike. *Nat Struct Mol Biol* 13:751–752. <http://dx.doi.org/10.1038/nsmb1123>.
29. Li F. 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol* 86:2856–2858. <http://dx.doi.org/10.1128/JVI.06882-11>.
30. Li F. 2013. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Res* 100:246–254. <http://dx.doi.org/10.1016/j.antiviral.2013.08.014>.
31. Du LY, He YX, Zhou YS, Liu SW, Zheng BJ, Jiang SB. 2009. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol* 7:226–236. <http://dx.doi.org/10.1038/nrmicro2090>.
32. Du L, Zhao G, Yang Y, Qiu H, Wang L, Kou Z, Tao X, Yu H, Sun S, Tseng CT, Jiang S, Li F, Zhou Y. 2014. A conformation-dependent neutralizing monoclonal antibody specifically targeting receptor-binding domain in Middle East respiratory syndrome coronavirus spike protein. *J Virol* 88:7045–7053. <http://dx.doi.org/10.1128/JVI.00433-14>.
33. He YX, Lu H, Siddiqui P, Zhou YS, Jiang SB. 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* 174:4908–4915. <http://dx.doi.org/10.4049/jimmunol.174.8.4908>.
34. Ying T, Du L, Ju TW, Prabaharan P, Lau CC, Lu L, Liu Q, Wang L, Feng Y, Wang Y, Zheng BJ, Yuen KY, Jiang S, Dimitrov DS. 2014. Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies. *J Virol* 88:7796–7805. <http://dx.doi.org/10.1128/JVI.00912-14>.
 35. Jiang L, Wang N, Zuo T, Shi X, Poon KM, Wu Y, Gao F, Li D, Wang R, Guo J, Fu L, Yuen KY, Zheng BJ, Wang X, Zhang L. 2014. Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. *Sci Transl Med* 6:234ra59. <http://dx.doi.org/10.1126/scitranslmed.3008140>.
 36. Sui JH, Li WH, Murakami A, Tamin A, Matthews LJ, Wong SK, Moore MJ, Tallarico ASC, Olurinde M, Choe H, Anderson LJ, Bellini WJ, Farzan M, Marasco WA. 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. <http://dx.doi.org/10.1073/pnas.0307140101>.
 37. Hofmann H, Pirc 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. <http://dx.doi.org/10.1073/pnas.0409465102>.
 38. Li WH, Moore MJ, Vasilieva N, Sui JH, Wong SK, Berne MA, Somasundaram M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426:450–454. <http://dx.doi.org/10.1038/nature02145>.
 39. Wong SK, Li WH, Moore MJ, 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. <http://dx.doi.org/10.1074/jbc.C300520200>.
 40. Lin HX, Fen Y, Wong G, Wang LP, Li B, Zhao XS, Li Y, Smaill F, Zhang CS. 2008. Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction. *J Gen Virol* 89:1015–1024. <http://dx.doi.org/10.1099/vir.0.83331-0>.
 41. Hofmann H, Simmons G, Rennekamp AJ, Chaiphan C, Gramberg T, Heck E, Geier M, Wegele A, Marzi A, Bates P, Pohlmann S. 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. <http://dx.doi.org/10.1128/JVI.00560-06>.
 42. Babcock GJ, Eshaki DJ, Thomas WD, Ambrosino DM. 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. <http://dx.doi.org/10.1128/JVI.78.9.4552-4560.2004>.
 43. Xiao XD, Chakraborti S, Dimitrov AS, Gramatikoff K, Dimitrov DS. 2003. The SARS-CoV S glycoprotein: expression and functional characterization. *Biochem Biophys Res Commun* 312:1159–1164. <http://dx.doi.org/10.1016/j.bbrc.2003.11.054>.
 44. Delmas B, Gelfi J, Lharidon R, Vogel LK, Sjoström H, Noren O, Laude H. 1992. Aminopeptidase-N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature* 357:417–420. <http://dx.doi.org/10.1038/357417a0>.
 45. 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.
 46. Raj VS, Mou HH, Smits SL, Dekkers DHW, Muller MA, Dijkman R, Muth D, Demmers JAA, Zaki A, Fouchier RAM, Thiel V, Drosten C, Rottier PJM, Osterhaus A, Bosch BJ, Haagmans BL. 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495:251–254. <http://dx.doi.org/10.1038/nature12005>.
 47. Du L, Zhao G, Kou Z, Ma C, Sun S, Poon VK, Lu L, Wang L, Debnath AK, Zheng BJ, Zhou Y, Jiang S. 2013. Identification of a receptor-binding domain in the S protein of the novel human coronavirus Middle East respiratory syndrome coronavirus as an essential target for vaccine development. *J Virol* 87:9939–9942. <http://dx.doi.org/10.1128/JVI.01048-13>.
 48. Mou H, Raj VS, van Kuppeveld FJ, Rottier PJ, Haagmans BL, Bosch BJ. 19 June 2013. The receptor binding domain of the new MERS coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J Virol* <http://dx.doi.org/10.1128/JVI.01277-13>.
 49. Dveksler GS, Pensiero MN, Cardellichio CB, Williams RK, Jiang GS, Holmes KV, Dieffenbach CW. 1991. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. *J Virol* 65:6881–6891.
 50. Williams RK, Jiang GS, Holmes KV. 1991. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc Natl Acad Sci U S A* 88:5533–5536. <http://dx.doi.org/10.1073/pnas.88.13.5533>.
 51. Schultze B, Gross HJ, Brossmer R, Herrler G. 1991. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J Virol* 65:6232–6237.
 52. Kubo H, Yamada YK, 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.
 53. Peng GQ, Xu LQ, Lin YL, Chen L, Pasquarella JR, Holmes KV, Li F. 2012. Crystal structure of bovine coronavirus spike protein lectin domain. *J Biol Chem* 287:41931–41938. <http://dx.doi.org/10.1074/jbc.M112.418210>.
 54. Krempl C, Schultze B, Laude H, Herrler G. 1997. Point mutations in the S protein connect the sialic acid binding activity with the enteropathogenicity of transmissible gastroenteritis coronavirus. *J Virol* 71:3285–3287.
 55. Schultze B, Cavanagh D, Herrler G. 1992. Neuraminidase treatment of avian infectious-bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. *Virology* 189:792–794. [http://dx.doi.org/10.1016/0042-6822\(92\)90608-R](http://dx.doi.org/10.1016/0042-6822(92)90608-R).
 56. Cavanagh D, Davis PJ. 1986. Coronavirus IBV: removal of spike glycopolyptide S1 by urea abolishes infectivity and hemagglutination but not attachment to cells. *J Gen Virol* 67:1443–1448. <http://dx.doi.org/10.1099/0022-1317-67-7-1443>.
 57. Schwegmann-Wessels C, Herrler G. 2006. Sialic acids as receptor determinants for coronaviruses. *Glycoconj J* 23:51–58. <http://dx.doi.org/10.1007/s10719-006-5437-9>.
 58. Promkuntod N, van Eijndhoven RE, de Vrieze G, Grone A, Verheije MH. 2014. Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology* 448:26–32. <http://dx.doi.org/10.1016/j.virol.2013.09.018>.
 59. Towler P, Staker B, Prasad SG, Menon S, Tang J, Parsons T, Ryan D, Fisher M, Williams D, Dales NA, Patane MA, Pantoliano MW. 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J Biol Chem* 279:17996–18007. <http://dx.doi.org/10.1074/jbc.M311191200>.
 60. Donoghue M, Hsieh F, Baronas E, Godbout K, Gosselin M, Stagliano N, Donovan M, Woolf B, Robison K, Jayaseelan R, Breitbart RE, Acton S. 2000. A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* 87:E1–E9. <http://dx.doi.org/10.1161/01.RES.87.5.e1>.
 61. Keidar S, Kaplan M, Gamliel-Lazarovich A. 2007. ACE2 of the heart: from angiotensin I to angiotensin (1-7). *Cardiovasc Res* 73:463–469. <http://dx.doi.org/10.1016/j.cardiores.2006.09.006>.
 62. Boehm M, Nabel EG. 2002. Angiotensin-converting enzyme 2—a new cardiac regulator. *N Engl J Med* 347:1795–1797. <http://dx.doi.org/10.1056/NEJMcibr022472>.
 63. Imai Y, Kuba K, Rao S, Huan Y, Guo F, Guan B, Yang P, Sarao R, Wada T, Leong-Poi H, Crackower MA, Fukamizu A, Hui CC, Hein L, Uhlig S, Slutsky AS, Jiang CY, Penninger JM. 2005. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* 436:112–116. <http://dx.doi.org/10.1038/nature03712>.
 64. Kuba K, Imai Y, Rao SA, Gao H, Guo F, Guan B, Huan Y, Yang P, Zhang YL, Deng W, Bao LL, Zhang BL, Liu G, Wang Z, Chappell M, Liu YX, Zheng DX, Leibbrandt A, Wada T, Slutsky AS, Liu DP, Qin CA, Jiang CY, Penninger JM. 2005. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* 11:875–879. <http://dx.doi.org/10.1038/nm1267>.
 65. Chen L, Lin YL, Peng GQ, Li F. 2012. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc Natl Acad Sci U S A* 109:17966–17971. <http://dx.doi.org/10.1073/pnas.1210123109>.
 66. Mina-Osorio P. 2008. The moonlighting enzyme CD13: old and new functions to target. *Trends Mol Med* 14:361–371. <http://dx.doi.org/10.1016/j.molmed.2008.06.003>.
 67. Rasmussen HB, Branner S, Wiberg FC, Wagtman N. 2003. Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a

- substrate analog. *Nat Struct Biol* 10:19–25. <http://dx.doi.org/10.1038/nsb882>.
68. Reinhold D, Kahne T, Steinbrecher A, Wrenger S, Neubert K, Ansorge S, Brocke S. 2002. The role of dipeptidyl peptidase IV (DP IV) enzymatic activity in T cell activation and autoimmunity. *Biol Chem* 383:1133–1138.
 69. Kameoka J, Tanaka T, Nojima Y, Schlossman SF, Morimoto C. 1993. Direct association of adenosine deaminase with a T cell activation antigen, CD26. *Science* 261:466–469. <http://dx.doi.org/10.1126/science.8101391>.
 70. Wesley UV, McGroarty M, Homoyouni A. 2005. Dipeptidyl peptidase inhibits malignant phenotype of prostate cancer cells by blocking basic fibroblast growth factor signaling pathway. *Cancer Res* 65:1325–1334. <http://dx.doi.org/10.1158/0008-5472.CAN-04-1852>.
 71. Tan KM, Zelus BD, Meijers R, Liu JH, Bergelson JM, Duke N, Zhang R, Joachimiak A, Holmes KV, Wang JH. 2002. Crystal structure of murine sCEACAM1a 1,4: a coronavirus receptor in the CEA family. *EMBO J* 21:2076–2086. <http://dx.doi.org/10.1093/emboj/21.9.2076>.
 72. Beauchemin N, Draber P, Dveksler G, Gold P, Gray-Owen S, Grunert F, Hammarstrom S, Holmes KV, Karlsson A, Kuroki M, Lin SH, Lucka L, Najjar SM, Neumaier M, Obrink B, Shively JE, Skubitz KM, Stanners CP, Thomas P, Thompson JA, Virji M, von Kleist S, Wagener C, Watt S, Zimmermann W. 1999. Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res* 252:243–249. <http://dx.doi.org/10.1006/excr.1999.4610>.
 73. Hammarström S. 1999. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* 9:67–81. <http://dx.doi.org/10.1006/scbi.1998.0119>.
 74. Dove A. 2001. The bittersweet promise of glycobiology. *Nat Biotechnol* 19:913–917. <http://dx.doi.org/10.1038/nbt1001-913>.
 75. Ghazarian H, Idoni B, Oppenheimer SB. 2011. A glycobiology review: carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem* 113:236–247. <http://dx.doi.org/10.1016/j.acthis.2010.02.004>.
 76. Li F, Li WH, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309:1864–1868. <http://dx.doi.org/10.1126/science.1116480>.
 77. 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 GF. 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500:227–231. <http://dx.doi.org/10.1038/nature12328>.
 78. Wang N, Shi X, Jiang S, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X, Arledge KC, Chen YH, Zhang L, Wang X. 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* 23:986–993. <http://dx.doi.org/10.1038/cr.2013.92>.
 79. Wu K, Li WB, Peng G, Li F. 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding complexed with its human receptor. *Proc Natl Acad Sci U S A* 106:19970–19974. <http://dx.doi.org/10.1073/pnas.0908837106>.
 80. Reguera J, Santiago C, Mudgal G, Ordone D, Enjuanes L, Casasnovas JM. 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog* 8:e1002859. <http://dx.doi.org/10.1371/journal.ppat.1002859>.
 81. Peng GQ, Sun DW, Rajashankar KR, Qian ZH, Holmes KV, Li F. 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci U S A* 108:10696–10701. <http://dx.doi.org/10.1073/pnas.1104306108>.
 82. Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, Ji W, Li Y, Wu Y, Wang J, Iwamoto A, Woo PC, Yuen KY, Yan J, Lu G, Gao GF. 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 16:328–337. <http://dx.doi.org/10.1016/j.chom.2014.08.009>.
 83. Li F. 2008. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J Virol* 82:6984–6991. <http://dx.doi.org/10.1128/JVI.00442-08>.
 84. Wu KL, Peng GQ, Wilken M, Geraghty RJ, Li F. 2012. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem* 287:8904–8911. <http://dx.doi.org/10.1074/jbc.M111.325803>.
 85. Li F, Li WH, Farzan M, Harrison SC. 2006. Interactions between SARS coronavirus and its receptor. *Adv Exp Med Biol* 581:229–234. http://dx.doi.org/10.1007/978-0-387-33012-9_38.
 86. Li WH, Zhang CS, Sui JH, Kuhn JH, Moore MJ, Luo SW, Wong SK, Huang IC, Xu KM, Vasilieva N, Murakami A, He YQ, Marasco WA, Guan Y, Choe HY, Farzan M. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 24:1634–1643. <http://dx.doi.org/10.1038/sj.emboj.7600640>.
 87. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, Chen QX, Gao YW, Zhou HQ, Xiang H, Zheng HJ, Chern SW, Cheng F, Pan CM, Xuan H, Chen SJ, Luo HM, Zhou DH, Liu YF, He JF, Qin PZ, Li LH, Ren YQ, Liang WJ, Yu YD, Anderson L, Wang M, Xu RH, Wu XW, Zheng HY, Chen JD, Liang G, Gao Y, Liao M, Fang L, Jiang LY, Li H, Chen F, Di B, He LJ, Lin JY, Tong S, Kong X, Du L, Hao P, Tang H, Bernini A, Yu XJ, Spiga O, Guo ZM, et al. 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* 102:2430–2435. <http://dx.doi.org/10.1073/pnas.0409608102>.
 88. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, Rao X, Song HD, Wang SY, Zuo Y, Zheng AH, Luo M, Wang HL, Deng F, Wang HZ, Hu ZH, Ding MX, Zhao GP, Deng HK. 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J Biol Chem* 280:29588–29595. <http://dx.doi.org/10.1074/jbc.M500662200>.
 89. Hou YX, Peng C, Yu M, Li Y, Han ZG, Li F, Wang LF, Shi ZL. 2010. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch Virol* 155:1563–1569. <http://dx.doi.org/10.1007/s00705-010-0729-6>.
 90. Wu K, Chen L, Peng G, Zhou W, Pennell CA, Mansky LM, Geraghty RJ, Li F. 2011. A virus-binding hot spot on human angiotensin-converting enzyme 2 is critical for binding of two different coronaviruses. *J Virol* 85:5331–5337. <http://dx.doi.org/10.1128/JVI.02274-10>.
 91. Liu L, Fang Q, Deng F, Wang HZ, Yi CE, Ba L, Yu WJ, Lin RD, Li TS, Hu ZH, Ho DD, Zhang LQ, Chen ZW. 2007. Natural mutations in the receptor binding domain of spike glycoprotein determine the reactivity of cross-neutralization between palm civet coronavirus and severe acute respiratory syndrome coronavirus. *J Virol* 81:4694–4700. <http://dx.doi.org/10.1128/JVI.02389-06>.
 92. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Cramer G, Zhang SY, Wang LF, Daszak P, Shi ZL. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503:535–538. <http://dx.doi.org/10.1038/nature12711>.
 93. Chen YQ, Rajashankar KR, Yang Y, Agnihothram SS, Liu C, Lin YL, Baric RS, Li F. 2013. Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *J Virol* 87:10777–10783. <http://dx.doi.org/10.1128/JVI.01756-13>.
 94. Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, Baric RS, 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 U S A* 111:12516–12521. <http://dx.doi.org/10.1073/pnas.1405889111>.
 95. Cockrell AS, Peck KM, Yount BL, Agnihothram SS, Scobey T, Curnes NR, Baric RS, Heise MT. 2014. Mouse dipeptidyl peptidase 4 is not a functional receptor for Middle East respiratory syndrome coronavirus infection. *J Virol* 88:5195–5199. <http://dx.doi.org/10.1128/JVI.03764-13>.
 96. van Doremalen N, Miazgowicz KL, Milne-Price S, Bushmaker T, Robertson S, Scott D, Kinne J, McLellan JS, Zhu J, Munster VJ. 2014. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J Virol* 88:9220–9232. <http://dx.doi.org/10.1128/JVI.00676-14>.
 97. Barlan A, Zhao J, Sarkar MK, Li K, McCray PB, Jr, Perlman S, Gallagher T. 2014. Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus infection. *J Virol* 88:4953–4961. <http://dx.doi.org/10.1128/JVI.00161-14>.
 98. Holm L, Sander C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* 26:316–319. <http://dx.doi.org/10.1093/nar/26.1.316>.
 99. Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM. 1998. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution. *J Biol Chem* 273:13047–13052. <http://dx.doi.org/10.1074/jbc.273.21.13047>.
 100. Wessner DR, Shick PC, Lu JH, Cardellicchio CB, Gagneten SE, Beauchemin N, Holmes KV, Dveksler GS. 1998. Mutational analysis of the virus and monoclonal antibody binding sites in MHVR, the cellular receptor of the murine coronavirus mouse hepatitis virus strain A59. *J Virol* 72:1941–1948.

101. **Thackray LB, Turner BC, Holmes KV.** 2005. Substitutions of conserved amino acids in the receptor-binding domain of the spike glycoprotein affect utilization of murine CEACAM1a by the murine coronavirus MHV-A59. *Virology* 334:98–110. <http://dx.doi.org/10.1016/j.virol.2005.01.016>.
102. **Leparc-Goffart I, Hingley ST, Chua MM, Jiang XH, Lavi E, Weiss SR.** 1997. Altered pathogenesis of a mutant of the murine coronavirus MHV-A59 is associated with a Q159L amino acid substitution in the spike protein. *Virology* 239:1–10. <http://dx.doi.org/10.1006/viro.1997.8877>.
103. **Chen L, Li F.** 2013. Structural analysis of the evolutionary origins of influenza virus hemagglutinin and other viral lectins. *J Virol* 87:4118–4120. <http://dx.doi.org/10.1128/JVI.03476-12>.
104. **Rossmann MG.** 1989. The canyon hypothesis. Hiding the host-cell receptor attachment site on a viral surface from immune surveillance. *J Biol Chem* 264:14587–14590.
105. **Wang J.** 2002. Protein recognition by cell surface receptors: physiological receptors versus virus interactions. *Trends Biochem Sci* 27:122–126. [http://dx.doi.org/10.1016/S0968-0004\(01\)02038-2](http://dx.doi.org/10.1016/S0968-0004(01)02038-2).

Fang Li is an Associate Professor of Pharmacology at the University of Minnesota. He received his Ph.D. in Structural Biology from Yale University and postdoctoral training in Structural Virology from Harvard Medical School. He started to work on structural biology of coronaviruses in 2003, motivated by the SARS outbreak that swept the world that year. Since his publication of the crystal structure of SARS-CoV receptor-binding protein complexed with its human receptor in 2005, he has determined a number of crystal structures of other coronavirus receptor-binding proteins complexed with their respective receptor. In addition to solving structures, he has identified the host receptor for bat coronavirus HKU4 and revealed the cell entry mechanism of MERS coronavirus. His research interests cover how viruses explore different host receptors and other host factors to expand their host ranges and how they transmit from animals to humans to cause epidemics.

