Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor

Guiqing Peng[a], Dawei Sun[a], Kanagalaghatta R. Rajashankar[b], Zhaohui Qian[c], Kathryn V. Holmes[c], and Fang Li[a,1]

[a]Department of Pharmacology, University of Minnesota Medical School, Minneapolis, MN 55455; [b]Department of Chemistry and Chemical Biology, Cornell University, Northeastern Collaborative Access Team, Advanced Photon Source, Argonne, IL 60439; and [c]Department of Microbiology, University of Colorado School of Medicine, Aurora, CO 80045

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved May 19, 2011 (received for review March 17, 2011)

Coronaviruses have evolved diverse mechanisms to recognize different receptors for their cross-species transmission and host-range expansion. Mouse hepatitis coronavirus (MHV) uses the N-terminal domain (NTD) of its spike protein as its receptor-binding domain. Here we present the crystal structure of MHV NTD complexed with its receptor murine carcinoembryonic antigen-related cell adhesion molecule 1a (mCEACAM1a). Unexpectedly, MHV NTD contains a core structure that has the same β-sandwich fold as human galectins (S-lectins) and additional structural motifs that bind to the N-terminal Ig-like domain of mCEACAM1a. Despite its galectin fold, MHV NTD does not bind sugars, but instead binds mCEACAM1a through exclusive protein–protein interactions. Critical contacts at the interface have been confirmed by mutagenesis, providing a structural basis for viral and host specificities of coronavirus/CEACAM1 interactions. Sugar-binding assays reveal that galectin-like NTDs of some coronaviruses such as human coronavirus OC43 and bovine coronavirus bind sugars. Structural analysis and mutagenesis localize the sugar-binding site in coronavirus NTDs to be above the β-sandwich core. We propose that coronavirus NTDs originated from a host galectin and retained sugar-binding functions in some contemporary coronaviruses, but evolved new structural features in MHV for mCEACAM1a binding.

coronavirus evolution | galectin-like N-terminal domain of coronavirus spike proteins | carcinoembryonic antigen-related cell adhesion molecule 1 binding by coronaviruses | sugar binding by coronaviruses

Coronaviruses use a variety of cellular receptors and coreceptors, including proteins and sugars. The diverse use of receptors has allowed coronaviruses to infect a wide range of mammalian and avian species and cause respiratory, enteric, systemic, and neurological diseases. How coronaviruses have evolved to do so has been a major puzzle in virology. To solve this puzzle, we have investigated the structural basis for the complex receptor-recognition mechanisms of coronaviruses.

The Coronaviridae family of large, enveloped, positive-stranded RNA viruses consists of at least three major genera or groups (Table S1). Aminopeptidase-N (APN) is the receptor for porcine transmissible gastroenteritis virus (TGEV), porcine respiratory coronavirus (PRCoV), and human coronavirus 229E (HCoV-229E) from group 1 (1–3). Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), a member of the carcinoembryonic antigen family in the Ig superfamily, is the receptor for mouse hepatitis coronavirus (MHV) from group 2 (subgroup 2a) (4, 5). Angiotensin-converting enzyme 2 (ACE2) is the receptor for human coronavirus NL63 (HCoV-NL63) from group 1 and human severe acute respiratory syndrome coronavirus (SARS-CoV) from group 2 (subgroup 2b) (6, 7). Sugars serve as receptors or coreceptors for TGEV from group 1, bovine coronavirus (BCoV) and human coronavirus OC43 (HCoV-OC43) from group 2a, and avian infectious bronchitis virus (IBV) from group 3 (8–14). Receptor is unknown for some coronaviruses such as group 2a human coronavirus HKU1 (HCoV-HKU1). The diversity in receptor use is a distinctive feature of the Coronaviridae family and a few other virus families such as retroviruses and paramyxoviruses (15, 16).

The characteristic large spikes on coronavirus envelopes are composed of trimers of the spike protein. The spike protein mediates viral entry into host cells by functioning as a class I viral fusion protein (17). During maturation, the spike protein is often cleaved into a receptor-binding subunit S1 and a membrane-fusion subunit S2 that associate together through noncovalent interactions (Fig. 1*A*). S1 sequences are relatively well conserved within each coronavirus group, but differ markedly between different groups. S1 contains two independent domains, N-terminal domain (NTD) and C domain, that can both serve as viral receptor-binding domains (RBDs) (Table S1). C domain binds to APN or ACE2 in coronaviruses that use them as receptors (3, 18–24), whereas NTD binds to CEACAM1 in MHV or sugar in TGEV (9, 25). The sugar-binding domain has not yet been identified in the spike proteins of HCoV-OC43, BCoV, or IBV. The only atomic structures available for coronavirus S1 are C domains of HCoV-NL63 and SARS-CoV, each complexed with their common receptor ACE2 (23, 24). Despite marked differences in their structures, the two C domains bind to overlapping regions on ACE2 (23, 24). Structural information has been lacking for any coronavirus S1 NTD.

MHV, the prototypic and extensively studied coronavirus, causes a variety of murine diseases. Strain A59 (MHV-A59) is primarily hepatotropic, whereas strain JHM (MHV-JHM) is neurotropic. Murine CEACAM1 (mCEACAM1), whose primary physiological functions are to mediate cell adhesion and signaling, is the principal receptor for all MHV strains (26). mCEACAM1a is broadly expressed in epithelial cells, endothelial cells, and microphages, but its expression level is low in the central nervous system and is restricted to endothelial and microgolial cells (27). mCEACAM1 is encoded by two alleles to produce mCEACAM1a and -1b; mCEACAM1a is a much more efficient MHV receptor than mCEACAM1b (28, 29). mCEACAM1a contains either two [D1 and D4] or four [D1–D4] Ig-like domains in tandem, a result of alternative mRNA splicing (26). The crystal structure of mCEACAM1a [1,4] shows that a CC′ loop (loop connecting β-strands C and C′) in the V-set Ig-like domain D1 encompasses key MHV-binding residues including Ile41 (30). Curiously, although mammalian CEACAM1 proteins are significantly conserved (Tables S2 and S3), only murine CEACAM1a can serve as an efficient MHV receptor (31). Also, although group 2a coronavirus spike proteins are significantly conserved (Tables S2 and S3), only MHV spike protein interacts with mCEACAM1a (31). The molecular determinants for the viral and host specificities of coronavirus/CEACAM1 interactions remain elusive.

Despite the lack of sequence homology between coronavirus spike proteins and any known sugar-binding proteins (lectins), sugar moieties on host cell membranes such as glycoproteins, glycolipids, and glycosaminoglycans play important roles in host cell infections by many coronaviruses (8). HCoV-OC43 and BCoV spike proteins recognize cell-surface components containing *N*-acetyl-9-*O*-acetylneuraminic acid (Neu5,9Ac2) (13, 14). These

Author contributions: G.P., K.V.H., and F.L. designed research; G.P., D.S., K.R.R., and Z.Q. performed research; G.P., K.V.H., and F.L. analyzed data; and F.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Crystallography, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank database (accession no. 3R4D).

[1]To whom correspondence should be addressed. E-mail: lifang@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104306108/-/DCSupplemental.

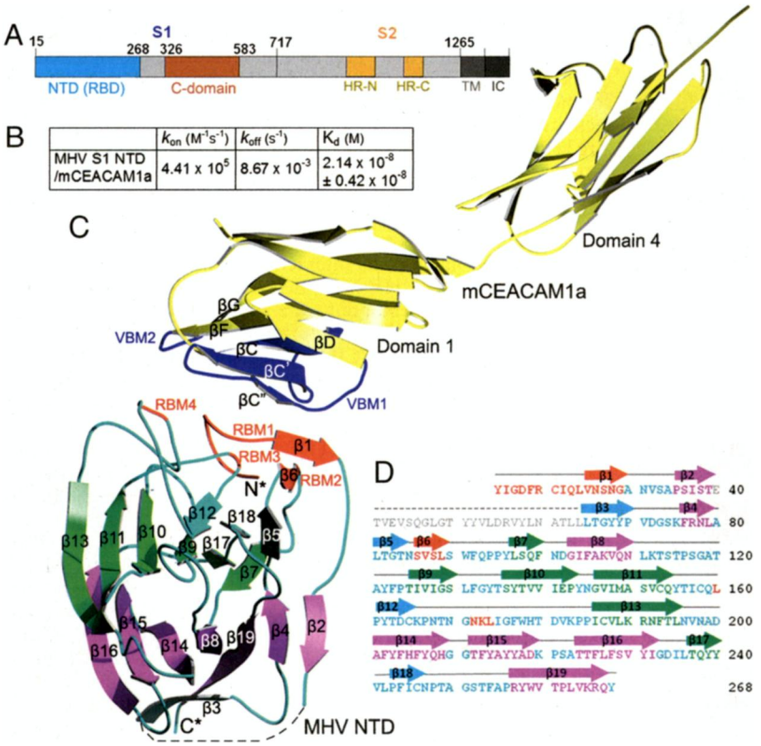

Fig. 1. Structure of MHV-NTD/mCEACAM1a complex. (*A*) Domain structure of MHV spike protein. NTD: N-terminal domain; RBD: receptor-binding domain; HR-N: heptad-repeat N; HR-C: heptad-repeat C; TM: transmembrane anchor; IC: intracellular tail. The signal peptide corresponds to residues 1–14 and is cleaved during molecular maturation (45). Structures and functions of gray areas have not been clearly defined. (*B*) Kinetics and binding affinity of NTD and mCEACAM1a. (*C*) Structure of NTD/mCEACAM1a complex. Two β-sheets of the NTD core are in green and magenta, respectively; receptor-binding motifs (RBMs) are in red; other parts of the NTD are in cyan; mCEACAM1a is in yellow; and virus-binding motifs (VBMs) are in blue. N*: N terminus; C*: C terminus. (*D*) Sequence and secondary structures of NTD. β-Strands are shown as arrows, and the disordered region as a dashed line.

viruses also contain a hemagglutinin–esterase (HE) that functions as a receptor-destroying enzyme and aids viral detachment from sugars on infected cells (32). MHV spike protein does not bind sugars (33), and the HE genes of many MHV strains are present but not expressed (34). TGEV spike protein recognizes *N*-glycolylneuraminic acid (Neu5Gc) and *N*-acetylneuraminic acid (Neu5Ac), and such sugar-binding activities are required for the enteric tropism of TGEV (9). PRCoV spike protein, an NTD-deletion mutant of TGEV spike protein, fails to bind sugars, and hence PRCoV has respiratory tropism only (3, 10). IBV spike protein recognizes Neu5Ac (11, 12). Overall, many coronavirus spike proteins can function as viral lectins, but the molecular nature of their lectin activities is a mystery.

How did coronavirus spike proteins originate and evolve, and how did the evolutionary changes in their spike proteins allow coronaviruses to explore novel cellular receptors and expand their host ranges? To address these questions, we have determined the crystal structure of MHV-A59 NTD complexed with mCEACAM1a[1, 4]. The structure has elucidated the receptor recognition mechanism of MHV and identified the determinants of the viral and host specificities of coronavirus/CEACAM1 interactions. Furthermore, the NTD structure has unexpectedly revealed the structural basis for the lectin activities of coronavirus spike proteins and provided structural insights into the origin and evolution of coronavirus spike proteins.

Results and Discussion

Structure Determination. To prepare an MHV-A59 NTD fragment suitable for crystallization, we designed a series of C-terminal truncation constructs of MHV-A59 NTD based on its secondary structure predictions. One NTD fragment containing residues 1–296 was well expressed in insect cells and stable in solution. It bound to mCEACAM1a[1,4] to form a 1:1 heterodimeric complex with K_d of 21.4 nM (Fig. 1*B*). We crystallized this complex in space group $P6_122$, a = 76.4 Å, b = 76.4 Å, and c = 942.1 Å (Table S4), with two complexes per asymmetric unit (Fig. S1). The structure was determined by single-wavelength anomalous diffraction (SAD) phases using selenomethionine-labeled mCEACAM1a. The phases were subsequently improved by an averaging method (35). We refined the structure at 3.1Å resolution (Table S4). The final model contains residues 15–268 of NTD (except for a disordered loop from residues 40–64) and residues 1–202 of mCEACAM1a. The model also contains glycans *N*-linked to viral residue 192 and to receptor residues 37, 55, and 70.

Structure of MHV-NTD/mCEACAM1a Complex. The core of MHV NTD is a 13-stranded β-sandwich with two antiparallel β-sheets stacked against each other through hydrophobic interactions (Fig. 1 *C* and *D* and Fig. 2*A*). Surprisingly, this β-sandwich core of MHV NTD has a galectin fold, which will be discussed in detail later. Outside the core structure, MHV NTD contains several peripheral structural elements (Fig. 1*C* and Fig. 2*A*). Three loops from the "upper" β-sheet of the core converge with the N-terminal segment to form a distinct, receptor-contacting substructure; these four structural elements are termed receptor-binding motifs (RBMs). Three disulfide bonds, connecting cysteines 21–158, 165–246, and 153–187, reinforce MHV NTD (Fig. 2*B*).

MHV NTD binds to domain D1 of mCEACAM1a (Fig. 1*C* and Fig. 2*A*). There is no significant structural change in mCEACAM1a before and after MHV binding (Fig. S2). The four RBMs of MHV NTD contact two virus-binding motifs (VBMs) on the CC′C″ face of mCEACAM1a (Fig. 1*C* and Fig. 2*A*). A total of 14 residues in NTD interact with a total of 17 residues in mCEACAM1a (Fig. 3 *A*–*C*). The binding buries 1,500 Å^2 at the interface (Fig. 2*B*). This interface is intermediate between the SARS-CoV/ACE2 interface (1,700 Å^2) and the HCoV-NL63/ACE2 interface (1,300 Å^2), but all of the three viral receptor-binding domains bind to their respective protein receptors with similar affinities (23). Notably, none of the observed or predicted glycans is involved in MHV-NTD/mCEACAM1a interactions (Fig. 2*B*), and therefore MHV-NTD/mCEACAM1a binding depends exclusively on protein–protein interactions.

Detailed MHV-NTD/mCEACAM1a Interactions. The interface between MHV NTD and mCEACAM1a is dominated by hydrophobic interactions with scattered polar interactions. Two hydrophobic

MICROBIOLOGY

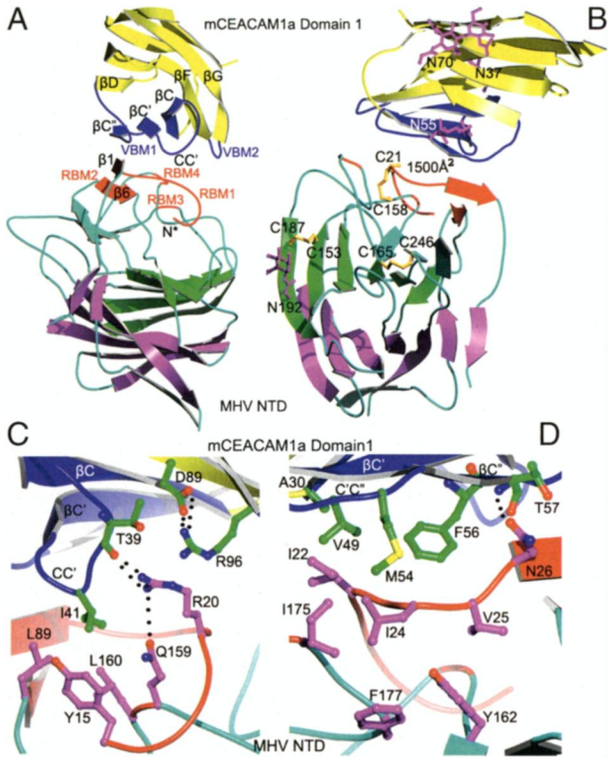

Fig. 2. Structural details of MHV-NTD/mCEACAM1a interface. (*A*) Another view of the MHV-NTD/mCEACAM1a structure, which is derived by rotating the one in Fig. 1*C* 90° clockwise along a vertical axis. Virus-binding motif 1 (VBM1) on MHV NTD includes strands βC, βC′, and βC″ and loops CC′, C′C″, and C″D. VBM2 on MHV NTD corresponds to loop FG. (*B*) Distribution of glycosylation sites and disulfide bonds. Glycans and glycosylated asparagines are in magenta, and cysteines are in yellow. The orientation of the structure is the same as in Fig. 1*C*. (*C*) A hydrophobic patch at the interface that is important for MHV-NTD/mCEACAM1a binding. MHV residues are in magenta, and mCEACAM1a residues are in green. The orientation of the structure is derived by rotating the structure in Fig. 1*C* 180° along a vertical axis. (*D*) Another hydrophobic patch at the interface that is important for MHV-NTD/mCEACAM1a binding. The orientation of the structure is slightly adjusted from the one in Fig. 1*C*.

patches stand out. One centers on Ile41 from the CC′ loop of mCEACAM1a. Ile41 is surrounded by the hydrophobic side chains of MHV Tyr15, Leu89, and Leu160 and the aliphatic side chains of MHV Gln159 and Arg20 (Fig. 2*C*). Additionally, MHV Arg20 forms a bifurcated hydrogen bond with the main chain carbonyl group of receptor Thr39 and another hydrogen bond with MHV Gln159. Receptor Arg96 forms a bifurcated salt bridge with receptor Asp89, while stacking with MHV Arg20. The second critical hydrophobic patch involves multiple hydrophobic residues from both MHV and mCEACAM1a that include MHV residues Ile22 and Tyr162 and receptor residues Val49, Met54, and Phe56 (Fig. 2*D*). Additionally, MHV Asn26 forms a hydrogen bond with the main chain amide group of receptor Thr57. In protein–protein interactions, hydrophobic interactions contribute more to binding energy, whereas hydrophilic interactions contribute more to binding specificity. The above key hydrogen bonds between NTD side chains and the receptor main chain help bring the adjacent hydrophobic patches into place. These structural analyses suggest that the hydrophobic patches and additional polar interactions provide significant binding energy and specificity to MHV-NTD/mCEACAM1a binding interactions.

The importance of contact residues at the NTD/mCEACAM1a interface has been confirmed by mutagenesis data. Here we compared the efficiency of mCEACAM1a-dependent cell entry by lentiviruses pseudotyped with wild-type or mutant MHV-A59 spike protein (Fig. 3*D*). Our data, together with published data (Fig. 3*E*), showed that single mCEACAM1a substitution I41G and single NTD substitutions I22A, Y162A, Y162Q, Y162H, but not Y162F, abrogated viral infectivity (28, 36), underscoring the significance of the two hydrophobic patches. Furthermore, single NTD substitutions R20A, R20K, and N26A significantly decreased viral infectivity, confirming the significance of the hydrogen bonds between Arg20 and mCEACAM1a and between Asn26 and mCEACAM1a. A naturally occurring Q159L mutation in MHV NTD caused small viral plaques (Fig. 3*E*) (37), suggesting that the hydrogen bond between Gln159 and Arg20 helps position Arg20 to interact with mCEACAM1a. Therefore, through evolution MHV-A59 appears to have optimized many of its interactions with mCEACAM1a, and thus substitutions in MHV NTD that disrupt these specific molecular interactions weaken or abrogate viral infectivity.

Our study provides the structural basis for the viral and host specificities of coronavirus/CEACAM1 interactions. On the basis of structural analyses and mutagenesis data, we determined that MHV NTD contains Arg20, Ile22, Asn26, and Tyr162, all of which form energetically favorable interactions with mCEACAM1a, whereas other group 2a coronavirus NTDs contain residues at the corresponding positions that are expected to disrupt critical hydrophobic or polar interactions with mCEACAM1a (Fig. 3*B* and Fig. S3). On the other hand, on the basis of structural analyses, we found that mCEACAM1a contains Ile41, Val49, Met54, and Phe56, all of which form energetically favorable interactions with MHV, whereas mCEACAM1b, bovine CEACAM1a and -1b, and human CEACAM1 contain residues at the corresponding positions that likely disrupt these favorable interactions with MHV (Fig. 3*C* and Fig. S3). For example, hydrophobic residues Ile41 and Phe56 in mCEACAM1a become hydrophilic residues Thr41 and Thr56 in mCEACAM1b (Fig. 3*C* and Fig. S3). These results reveal the mechanisms whereby MHV uses only mCEACAM1a, and not mCEACAM1b or CEACAM1 from cattle or humans, as its receptor, and whereby other group 2a coronaviruses cannot use mCEACAM1a as a receptor.

Sugar Binding by Coronavirus NTDs. The β-sandwich core of MHV NTD shares the same 11-stranded fold as human galectins (S-lectins) and rotavirus VP4 (viral lectin) (38, 39), augmented by two additional β-strands in the "lower" β-sheet (Fig. 4 and Fig. S4). MHV NTD and human galectin-3 have a Dali *Z*-score of 7.8 and an rmsd value of 2.9 Å over 137 matching Cα atoms (40). Importantly, the topologies of their β-sandwich cores are identical (Fig. S4). This unexpected structural homology between MHV NTD and human galectins suggests that coronavirus NTDs may function as viral lectins. To test this possibility, we designed NTD constructs of other group 2 coronaviruses that correspond to the crystallized MHV-NTD fragment based on the sequence alignment of these spike proteins. We expressed and purified each of these coronavirus NTDs and detected their binding interactions with mucin, a mixture of highly glycosylated proteins containing all of the sugar moieties (Neu5,9Ac2, Neu5Gc, and Neu5Ac) recognized by the coronavirus spike proteins. Results showed that NTDs of HCoV-OC43 and BCoV bound sugars, whereas NTDs of MHV-A59, HCoV-HKU1, and SARS-CoV did not (Fig. 5). Removal of neuraminic acids from mucin by neuraminidase treatment prevented binding of HCoV-OC43 or BCoV NTD.

Why do the NTDs of HCoV-OC43 and BCoV, but not that of MHV, bind sugars, and where is the sugar-binding site located in coronavirus NTDs? In human galectin-3 that binds galactose, the sugar-binding site (site A) is located above the β-sandwich core and involves the 10–11 loop (loop connecting β10 and β11) (Fig. 4 *B* and *E*) (39). In rotavirus VP4, which binds sialic acids, site A is blocked by a two-stranded β-sheet; instead, the sugar-binding site (site B) is located in a groove between the two β-sheets of the β-sandwich core (Fig. 4 *C* and *F*) (38). In MHV NTD, site B is blocked due to the narrowed groove between the two β-sheets of the β-sandwich core, whereas site A is open and available (Fig. 4 *A* and *D*). However, compared with human galectin-3, MHV NTD has a markedly shortened 10–11 loop that may be responsible for its lack of lectin activity. HCoV-OC43 and BCoV NTDs likely share the same galectin fold as MHV NTD due to their high sequence similarities (Fig. S1), but they both contain longer 10–11

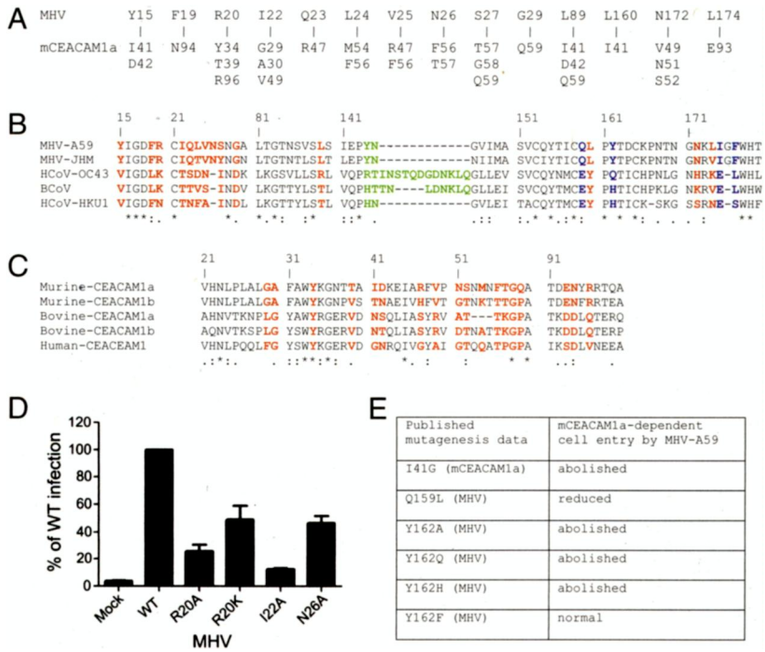

Fig. 3. Sequence analysis and mutagenesis studies of coronavirus/CEACAM1 interactions. (*A*) List of contact residues at the interface. (*B*) Partial sequence alignment of group 2a coronavirus NTDs. Contact residues are in red, important noncontact residues are in blue, and loop 10–11 is in green. Asterisks indicate positions that have fully conserved residues. Colons indicate positions that have strongly conserved residues. Periods indicate positions that have weakly conserved residues. (*C*) Partial sequence alignment of mammalian CEACAM1 proteins. (*D*) Structure-guided mutagenesis data on MHV NTD. Measured was mCEACAM1a-dependent cell entry by lentiviruses pseudotyped with wild-type or mutant MHV-A59 spike proteins. SEs are shown. (*E*) Published mutagenesis data on MHV NTD (28, 36, 37).

loops than MHV NTD (Fig. 3*B* and Fig. S3) and thus may use site A for sugar binding. To test this hypothesis, we modified the 10–11 loops in both BCoV and HCoV-OC43 NTDs, using MHV NTD as a reference (Fig. 3*B* and Fig. S3). For both BCoV and HCoV-OC43 NTDs, the mutant and wild-type proteins were equally well expressed and stable in solution, but the mutant proteins (OC43* and BCoV*) lacked sugar-binding activities (Fig. 5). These observations confirm that the 10–11 loops are critical for sugar binding in both BCoV and HCoV-OC43 NTDs. A more refined description of the sugar-binding site in BCoV and HCoV-OC43 NTDs awaits future structural and biochemical studies.

Coronavirus Receptor Use and Evolution. To date, three crystal structures are available for RBDs of coronavirus S1: group 2a MHV NTD, group 2b SARS-CoV C domain (24), and group 1 HCoV-NL63 C domain (23) (Fig. 6). Because of the significant sequence similarities of the S1 subunits of the spike proteins within each coronavirus group, the six-stranded β-sandwich core structure of the HCoV-NL63 C domain likely exists in other group 1 coronaviruses (23), and the 5-stranded β-sheet core structure of the SARS-CoV C domain likely exists in other group 2 coronaviruses (24). Similarly, the galectin-like NTD of MHV likely exists in other group 2 coronaviruses. The folds of group 1 and group 3 coronavirus NTDs are less clear. However, because both TGEV NTD and IBV S1 have lectin activities, the galectin-fold core structure of group 2a coronavirus NTDs may also be found in both group 1 and group 3 coronaviruses in similar or variant forms. The present study advances our understanding of the structures and functions of coronavirus spike proteins and the complex receptor-recognition mechanisms of coronaviruses.

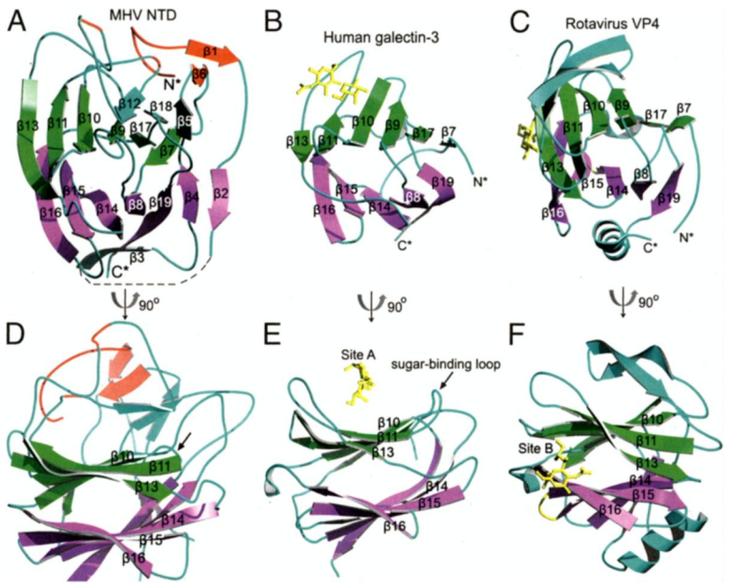

Fig. 4. Structural comparisons of MHV NTD, human galectins, and rotavirus VP4. (*A*) MHV NTD. The orientation of the structure is the same as in Fig. 1*C*. (*B*) Human galectin-3 [Protein Data Bank (PDB) 1A3K]. The β-sandwich core is labeled and colored the same as in MHV NTD. Bound galactose is in yellow. (*C*) Rotavirus VP4 (PDB 1KQR). Bound sialic acid is in yellow. (*D*) Another view of MHV S1 NTD, which is derived by rotating the structure in *A* counterclockwise along a vertical axis. Arrow indicates loop 10–11. (*E*) Another view of human galectin-3. Site A indicates its galactose-binding site. (*F*) Another view of rotavirus VP4. Site B indicates its sialic-acid–binding site.

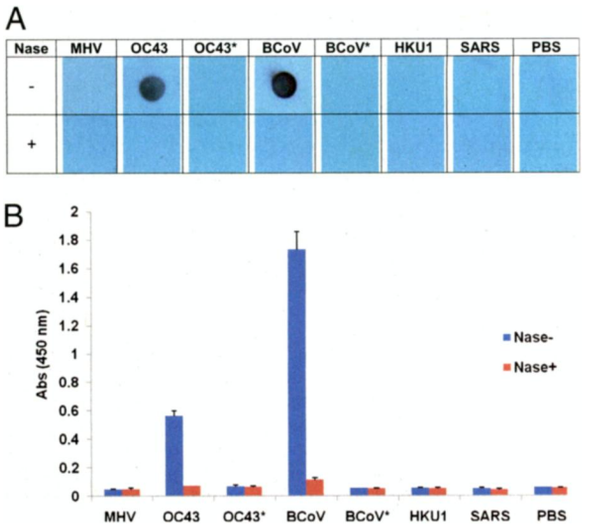

Fig. 5. Sugar-binding assays of group 2 coronavirus NTDs. (*A*) Dot-blot overlay assay. Measured were the binding interactions between coronavirus NTDs and sugar moieties on mucin-spotted nitrocellulose membranes. The membranes were either mock-treated or treated with neuraminidase (Nase) beforehand. Sugar-binding NTDs were detected using antibodies against their C-terminal His tags. (*B*) ELISA in which mucin-coated plates were used instead of nitrocellulose membranes. Sugar-binding NTDs were detected using ELISA substrates, and absorbance of the resulting yellow color was read at 450 nm. SEs are shown. BCoV* and OC43*: BCoV and HCoV-OC43 NTDs whose 10–11 loops have been replaced by that of MHV NTD.

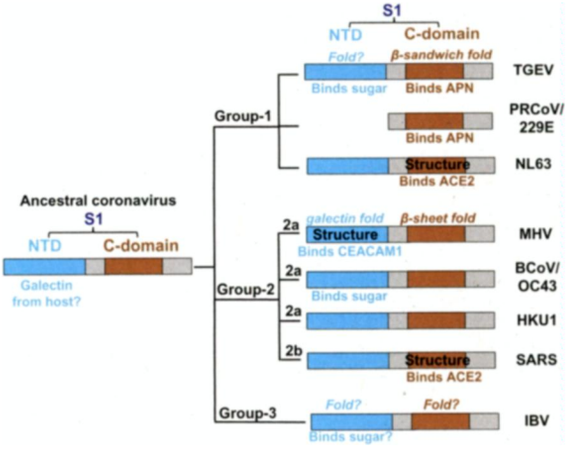

Fig. 6. Structures, functions, and evolution of S1 subunits of coronavirus spike proteins. The three known crystal structures are indicated by "Structure." Among these structures, MHV NTD has a 13-stranded galectin-like β-sandwich fold, HCoV-NL63 C domain has a six-stranded β-sandwich fold, and the SARS-CoV C domain has a five-stranded β-sheet fold.

How did coronavirus spike NTDs originate and evolve? We propose that an ancestral coronavirus acquired a galectin-like domain from its host. Subsequently, an ancestral group 2a coronavirus incorporated a HE gene into its genome to aid viral detachment from sugars on infected cells. Later, the galectin-like NTD of MHV evolved additional novel structural elements that allowed it to bind mCEACAM1a. Using a protein receptor instead of sugar receptors greatly enhanced the attachment affinity between MHV and murine cells, making sugar-binding functions dispensable. Accordingly, MHV-A59 underwent changes in the sugar-binding site of its NTD, lost its sugar-binding activity, and stopped expressing its HE gene. In contrast, the galectin-like NTDs of some contemporary coronaviruses such as HCoV-OC43, BCoV, and TGEV retain the lectin activity, although their sugar specificities have diverged in three coronavirus groups and differ from those of contemporary human galectins. Some TGEV strains deleted their NTD to become PRCoV after their C domain acquired APN-binding affinity. In addition to coronaviruses, there is evidence that paramyxoviruses may also have acquired their RBD from a host (although with a β-propeller fold) and used it to bind sugar or protein receptors (41, 42). It seems that viruses use a common evolutionary strategy by acquiring host proteins and evolving them into viral RBDs with different receptor specificity. This strategy allows viruses to explore novel cellular receptors and expand their host ranges. The current study provides critical structural information that illustrates how coronaviruses have successfully used this strategy.

Materials and Methods

Protein Purification and Crystallization. Both MHV-A59 S1 NTD (residues 1–296) and mCEACAM1a[1,4] (residues 1–202) were expressed in Sf9 insect cells using the bac-to-bac system (Invitrogen) and then purified as previously described (24). Briefly, the proteins containing C-terminal His tags were harvested from Sf9 cell supernatants, loaded onto a nickel-nitrilotriacetic acid (Ni-NTA) column, eluted from the Ni-NTA column with imidazole, and further purified by gel filtration chromatography on Superdex 200 (GE Healthcare). To purify the NTD/mCEACAM1a complex, NTD was incubated with excess mCEACAM1a before the complex was purified by gel filtration chromatography and concentrated to 10 mg/mL. Crystals of the NTD/mCEACAM1a complex were grown in sitting drops at 4 °C over wells containing 10% PEG6000 and 0.2 M CaCl2. Crystals were harvested in 2 wk; stabilized in 10% PEG6000, 0.2 M CaCl2, and 30% ethylene glycol; and flash-frozen in liquid nitrogen.

Selenomethionine-labeled mCEACAM1a was expressed in Sf9 cells as previously described (43). Briefly, 24 h postinfection, cells were transferred into medium without methionine for methionine depletion. After 4 h, cells were transferred into medium without methionine but supplemented with 50 mg/mL selenomethionine for selenomethionine labeling. After 36 h, cells were harvested and selenomethionine-labeled protein was purified using the same procedure as above.

Structure Determination and Refinement. X-ray data were collected at Advanced Photon Source Northeastern Collaborative Access Team beamlines. The crystal contains two complexes per asymmetric unit. From a selenomethionine-labeled crystal, 12 selenomethionine sites in mCEACAM1a were identified. SAD phases were then calculated. The phases were subsequently improved by twofold noncrystallographic symmetry (NCS) averaging within the crystal and cross-crystal averaging with the mCEACAM1a crystal (35). During both density modification and structure refinement, the twofold NCS restraint was applied to each of the three domains—NTD and domains D1 and D4 of mCEACAM1a—between the two complexes in each asymmetric unit. This is because domains D1 and D4 undergo a hinge movement relative to each other in the two complexes due to the flexibility of the domain linker. The structure was refined to a final R_{free} of 30.8% and R_{work} of 24.8%. Data, phasing, and refinement statistics are shown in Table S4. Software used for data processing, structure determination, and refinement is also listed in Table S4.

Kinetics and Binding Affinity of MHV S1 NTD and mCEACAM1a by Surface Plasmon Resonance Using Biacore. The kinetics and binding affinity of MHV S1 NTD and mCEACAM1a were measured by surface plasmon resonance using a Biacore 3000. The surface of a C5 sensor chip was first activated with *N*-hydroxysuccinimide, MHV S1 NTD was then injected and immobilized to the surface of the chip, and the remaining activated surface of the chip was blocked with ethanolamine. Soluble mCEACAM1a was introduced at a flow rate of 20 μL/min at different concentrations. Kinetic parameters were determined using BIA-EVALUATIONS software.

mCEACAM1a-Dependent Cell Entry of Lentiviruses Pseudotyped with MHV-A59 Spike Protein. Lentiviruses pseudotyped with MHV-A59 spike protein were produced as previously described (44). Briefly, plasmid encoding wild-type or mutant MHV-A59 spike protein was cotransfected into 293T cells with helper plasmid psPAX2 and reporter plasmid pLenti-GFP at molar ratio 1:1:1 using polyethyleneimine (Polysciences Inc.). Forty-eight hours posttransfection, the resulting pseudotyped viruses were harvested and inoculated onto the 293 cells expressing mCEACAM1a. After overnight, cells were washed twice with PBS and lifted by 100 μL of 0.25% trypsin and 0.38 mg/mL EDTA. After being washed twice with PBS, cells were fixed with 1% paraformaldehyde and analyzed for GFP expression by flow cytometry. All experiments were repeated at least three times. The expression levels of spike proteins in pseudotyped

viruses were measured by Western blotting using polyclonal goat antibody to MHV-A59 spike protein (Fig. S5).

Sugar-Binding Assays of Coronavirus S1 NTDs. Sugar-binding assays of coronavirus S1 NTDs were performed as previously described (13, 33). Briefly, for the dot-blot overlay assay, 10 μg of bovine submaxillary gland mucin (BSM) (Sigma-Aldrich) was spotted onto nitrocellulose membranes. The membranes were dried completely, blocked with BSA at 4 °C overnight, and either mock-treated or treated with 20 mU/mL *Arthrobacter ureafaciens* neuraminidase (Roche Applied Science) at 25 °C for 2 h. The membranes were then incubated with 1 μM coronavirus S1 NTDs containing a C-terminal His tag at 4 °C for 2 h, washed five times with PBS, incubated with anti-His antibody (Invitrogen) at 4 °C for 2 h, washed five times with PBS again, incubated with HRP-conjugated goat anti-mouse IgG antibody (1:5,000) at 4 °C for 2 h, and washed five times with PBS. Finally, the bound proteins were detected using a chemiluminescence reagent (ECL plus, GE Healthcare). For the ELISA, BSM (60 μg/mL in PBS) was coated at 4 °C overnight in the wells of 96-well Maxisorp plates (Nunc). The wells were treated in the same way as in the dot-blot overlay assay. Femto-ELISA-HRP substrates were added and incubated at room temperature. The reaction was stopped with 1 N HCl, and the absorbance of the resulting yellow color was read at 450 nm.

ACKNOWLEDGMENTS. We thank Dr. Stephen C. Harrison, Dr. Robert J. Geraghty, and Dr. Doug Ohlendorf for comments. Computer resources were provided by the Basic Sciences Computing Laboratory of the University of Minnesota Supercomputing Institute. This work was supported by National Institutes of Health Grants R01AI089728 (to F.L.), P01AI059576 (to K.V.H.), and RR-15301 (to the Advanced Photon Source Northeastern Collaborative Access Team beamlines); by a University of Minnesota Academic Health Center Faculty Research Development grant (to F.L.); and by a Minnesota Partnership for Biotechnology and Medical Genomics grant (to the University of Minnesota).

1. Delmas B, et al. (1992) Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* 357:417–420.
2. Yeager CL, et al. (1992) Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* 357:420–422.
3. Delmas B, Gelfi J, Sjöström H, Noren O, Laude H (1993) Further characterization of aminopeptidase-N as a receptor for coronaviruses. *Adv Exp Med Biol* 342:293–298.
4. Dveksler GS, et al. (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.
5. 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 USA* 88:5533–5536.
6. Hofmann H, et al. (2005) Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc Natl Acad Sci USA* 102: 7988–7993.
7. Li WH, et al. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426:450–454.
8. Schwegmann-Wessels C, Herrler G (2006) Sialic acids as receptor determinants for coronaviruses. *Glycoconj J* 23:51–58.
9. 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.
10. Schultze B, et al. (1996) Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. *J Virol* 70:5634–5637.
11. Cavanagh D, Davis PJ (1986) Coronavirus IBV: Removal of spike glycopolypeptide S1 by urea abolishes infectivity and haemagglutination but not attachment to cells. *J Gen Virol* 67:1443–1448.
12. 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.
13. 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.
14. Künkel F, Herrler G (1993) Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* 195:195–202.
15. Bowden TA, Crispin M, Jones EY, Stuart DI (2010) Shared paramyxoviral glycoprotein architecture is adapted for diverse attachment strategies. *Biochem Soc Trans* 38:1349–1355.
16. Overbaugh J, Miller AD, Eiden MV (2001) Receptors and entry cofactors for retroviruses include single and multiple transmembrane-spanning proteins as well as newly described glycophosphatidylinositol-anchored and secreted proteins. *Microbiol Mol Biol Rev* 65:371–389.
17. 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.
18. 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.
19. Breslin JJ, et al. (2003) Human coronavirus 229E: Receptor binding domain and neutralization by soluble receptor at 37 degrees C. *J Virol* 77:4435–4438.
20. Hofmann H, et al. (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.
21. Lin HX, et al. (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.
22. 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.
23. Wu K, Li WK, Peng G, Li F (2009) Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc Natl Acad Sci USA* 106:19970–19974.
24. Li F, Li WH, Farzan M, Harrison SC (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309:1864–1868.
25. 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.
26. Beauchemin N, et al. (1999) Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res* 252:243–249.
27. Godfraind C, et al. (1995) Tissue and cellular distribution of an adhesion molecule in the carcinoembryonic antigen family that serves as a receptor for mouse hepatitis virus. *Lab Invest* 73:615–627.
28. Wessner DR, et al. (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.
29. Ohtsuka N, Yamada YK, Taguchi F (1996) Difference in virus-binding activity of two distinct receptor proteins for mouse hepatitis virus. *J Gen Virol* 77:1683–1692.
30. Tan KM, et al. (2002) Crystal structure of murine sCEACAM1a[1,4]: A coronavirus receptor in the CEA family. *EMBO J* 21:2076–2086.
31. Compton SR, Stephensen CB, Snyder SW, Weismiller DG, Holmes KV (1992) Coronavirus species specificity: Murine coronavirus binds to a mouse-specific epitope on its carcinoembryonic antigen-related receptor glycoprotein. *J Virol* 66:7420–7428.
32. Zeng QH, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ (2008) Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc Natl Acad Sci USA* 105:9065–9069.
33. Langereis MA, van Vliet ALW, Boot W, de Groot RJ (2010) Attachment of mouse hepatitis virus to O-acetylated sialic acid is mediated by hemagglutinin-esterase and not by the spike protein. *J Virol* 84:8970–8974.
34. Yokomori K, Banner LR, Lai MMC (1991) Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology* 183:647–657.
35. Li WK, Li F (2011) Cross-crystal averaging with search models to improve molecular replacement phases. *Structure* 19:155–161.
36. 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.
37. Leparc-Goffart I, et al. (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.
38. Dormitzer PR, Sun ZYJ, Wagner G, Harrison SC (2002) The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J* 21: 885–897.
39. Seetharaman J, et al. (1998) X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. *J Biol Chem* 273:13047–13052.
40. Holm L, Sander C (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* 26:316–319.
41. Bowden TA, et al. (2008) Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. *Nat Struct Mol Biol* 15:567–572.
42. Santiago C, Celma ML, Stehle T, Casasnovas JM (2010) Structure of the measles virus hemagglutinin bound to the CD46 receptor. *Nat Struct Mol Biol* 17:124–129.
43. Bellizzi JJ, Widom J, Kemp CW, Clardy J (1999) Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure* 7:R263–R267.
44. Qian ZH, Wang HZ, Empig C, Anderson WF, Albritton LM (2004) Complementation of a binding-defective retrovirus by a host cell receptor mutant. *J Virol* 78:5766–5772.
45. Taguchi F, Ikeda T, Shida H (1992) Molecular cloning and expression of a spike protein of neurovirulent murine coronavirus JHMV variant cl-2. *J Gen Virol* 73:1065–1072.