Crystal Structure of Bovine Coronavirus Spike Protein Lectin Domain*⒮

Guiqing Peng[‡1], **Liqing Xu**[‡1], **Yi-Lun Lin**[‡], **Lang Chen**[‡], **Joseph R. Pasquarella**[‡], **Kathryn V. Holmes**[§], and **Fang Li**[‡2]

From the ‡Department of Pharmacology, University of Minnesota Medical School, Minneapolis, Minnesota 55455 and
§Department of Microbiology, University of Colorado School of Medicine, Aurora, Colorado 80045

Background: Coronavirus spike protein N-terminal domains (NTDs) bind sugar or protein receptors.
Results: We determined crystal structure of bovine coronavirus NTD and located its sugar-binding site using mutagenesis.
Conclusion: Bovine coronavirus NTD shares structural folds and sugar-binding sites with human galectins and has subtle yet functionally important differences from protein-binding NTD of mouse coronavirus.
Significance: This study explores origin and evolution of coronavirus NTDs.

The spike protein N-terminal domains (NTDs) of bovine coronavirus (BCoV) and mouse hepatitis coronavirus (MHV) recognize sugar and protein receptors, respectively, despite their significant sequence homology. We recently determined the crystal structure of MHV NTD complexed with its protein receptor murine carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1), which surprisingly revealed a human galectin (galactose-binding lectin) fold in MHV NTD. Here, we have determined at 1.55 Å resolution the crystal structure of BCoV NTD, which also has the human galectin fold. Using mutagenesis, we have located the sugar-binding site in BCoV NTD, which overlaps with the galactose-binding site in human galectins. Using a glycan array screen, we have identified 5-*N*-acetyl-9-*O*-acetylneuraminic acid as the preferred sugar substrate for BCoV NTD. Subtle structural differences between BCoV and MHV NTDs, primarily involving different conformations of receptor-binding loops, explain why BCoV NTD does not bind CEACAM1 and why MHV NTD does not bind sugar. These results suggest a successful viral evolution strategy in which coronaviruses stole a galectin from hosts, incorporated it into their spike protein, and evolved it into viral receptor-binding domains with altered sugar specificity in contemporary BCoV or novel protein specificity in contemporary MHV.

Coronaviruses are a family of large, enveloped, and positive-stranded RNA viruses. They infect mammalian and avian species and cause respiratory, enteric, systemic, and neurological diseases (1). Coronaviruses are classified into at least three major genetic genera: α, β, and γ. Bovine coronavirus (BCoV),[3]

human OC43 coronavirus (HCoV-OC43), and mouse hepatitis coronavirus (MHV) all belong to the β-genus. BCoV causes enteritis and respiratory disease in cattle, HCoV-OC43 causes respiratory disease in humans, and MHV causes hepatitis, enteritis, and neurological disease in mice. Genetically, BCoV and HCoV-OC43 are so closely related that HCoV-OC43 is believed to have resulted from zoonotic spillover of BCoV (2, 3). MHV is also genetically related to BCoV and HCoV-OC43, although not as closely as BCoV and HCoV-OC43 are to each other.

Coronaviruses use a variety of cellular receptors, including proteins and sugars. BCoV and HCoV-OC43 recognize a sugar moiety, 5-*N*-acetyl-9-*O*-acetylneuraminic acid (Neu5,9Ac2), on cell-surface glycoproteins or glycolipids (4, 5). In contrast, MHV does not use sugar as a receptor (6). Instead, it uses a protein receptor, murine carcinoembryonic antigen-related cell adhesion molecule 1a (mCEACAM1a) (7, 8), a member of the carcinoembryonic antigen (CEA) family in the immunoglobulin (Ig) superfamily (9). In addition, two other types of sugars, 5-*N*-glycolylneuraminic acid and 5-*N*-acetylneuraminic acid, can serve as receptors or co-receptors for some α-genus and γ-genus coronaviruses (10–12), whereas two other cell-surface proteins, angiotensin-converting enzyme 2 and aminopeptidase N, can serve as receptors for some α-genus and β-genus coronaviruses (13–18). How coronaviruses have evolved to recognize these diverse receptors presents an evolutionary conundrum.

The spike protein on coronavirus envelopes recognizes receptors through the activities of a receptor-binding subunit S1 before it fuses viral and host membranes through the activities of a membrane-fusion subunit S2 (19). S1 contains two independent domains, an N-terminal domain (NTD) and a C domain, both of which can function as viral receptor-binding domains (20). Crystal structures have been determined for the complexes of several coronavirus receptor-binding domains complexed with their respective receptors, including MHV NTD complexed with mCEACAM1a (21–24). Unexpectedly, MHV NTD contains the same fold as human galectins (galactose-binding lectins) (22), although it does not bind sugar (6). Instead, it binds mCEACAM1a through exclusive protein-pro-

---

tein interactions. In contrast, BCoV and HCoV-OC43 NTDs, both of which have significant sequence homology to MHV NTD, bind sugar and function as viral lectins. Consistent with the existence of a viral lectin in their spike proteins, BCoV and HCoV-OC43 also encode a hemagglutinin-esterase that functions as a receptor-destroying enzyme and aids viral detachment from sugar on infected cells (25). MHV also contains a hemagglutinin-esterase gene in its genome, but only some of the MHV strains actively express the hemagglutinin-esterase protein (26). These observations raise interesting questions about the origin and evolution of coronavirus spike protein lectin domains.

In this study, we have determined the structure of BCoV NTD by x-ray crystallography and mapped the sugar-binding site in BCoV NTD using mutagenesis. In addition, this study reveals the structural differences between BCoV and MHV NTDs, which lead to their respective receptor specificities. Based on these results, we speculate on the evolutionary relationships among BCoV NTD, MHV NTD, and host galectins.

EXPERIMENTAL PROCEDURES

Structure Determination—BCoV NTD (residues 15–298) was expressed and purified as described previously for MHV NTD (residues 15–296) (22). Briefly, the BCoV NTD gene was inserted into insect cell expression vector pFastbac I. The protein, which contained a signal peptide (residues 1–14) and a C-terminal His$_6$ tag, was expressed in sf9 insect cells, secreted into cell culture medium, purified sequentially on nickel-nitrilotriacetic acid and gel-filtration columns, concentrated to 10 mg/ml, and stored in buffer containing 200 mM NaCl and 20 mM HEPES, pH 7.5. Crystals of BCoV NTD were grown in sitting drops at 20 °C, with 1 μl of protein solution and 1 μl well solution containing 2.0 M (NH$_4$)$_2$SO$_4$. Crystals diffracted to 1.55 Å resolution. H test for crystal twinning suggested that the data were twinned with a twinning fraction of 0.41 (27). The corresponding twinning operator (h+k, −k, −l) was applied to the following procedures, including molecular replacement and model refinement. The structure was determined by molecular replacement using Phaser software (28) with the structure of MHV NTD (Protein Data Bank code 3R4D) as the search model. The structure was refined to 1.55 Å using Refmac software (Table 1) (29).

Sugar-binding Assays of Coronavirus NTDs by ELISA—Sugar-binding assays of coronavirus NTDs were performed as described previously (22). Briefly, bovine submaxillary gland mucin (60 μg/ml in PBS) was coated in the wells of 96-well Maxisorp plates (Nunc). The wells were dried completely, blocked with BSA, and incubated with 1 μM coronavirus NTDs containing a C-terminal His$_6$ tag, washed five times with PBS, incubated with mouse anti-His$_6$ antibody (Invitrogen), washed five times with PBS, incubated with HRP-conjugated goat anti-mouse IgG antibody (1:5000), and washed five times with PBS. Finally, the bound proteins were detected using Femto-ELISA-HRP substrates, and the reaction was stopped with 1 N HCl. The absorbance of the resulting yellow color was read at 450 nm.

CEACAM1-binding Assays of Coronavirus NTDs by ELISA—CEACAM1-binding assays of coronavirus NTDs were performed the same way as the sugar-binding assays, except that

mammalian CEACAM1 proteins (60 μg/ml in PBS), instead of mucin, were coated in the wells of the plates. The CEACAM1 proteins used in this study were constructed and expressed the same way as mCEACAM1a that was previously crystallized in complex with MHV NTD (22). However, the CEACAM1 proteins used in this study all had a C-terminal Fc tag, while mCEACAM1a used in the previous study had a C-terminal His6 tag. Consequently, a protein G column instead of an nickel-nitrilotriacetic acid column was used as one of the purification steps for the CEACAM1 proteins used in this study. All of the CEACAM1 proteins were soluble in solution.

Substrate-binding Assays of Coronavirus NTDs by Surface Plasmon Resonance Using Biacore—The binding reactions between coronavirus NTDs and mucin or CEACAM1 were assayed by surface plasmon resonance using a Biacore 2000 as described previously (23). Briefly, mucin or CEACAM1 was directly immobilized on a C5 sensor chip. The surface of the sensor chip was first activated with *N*-hydroxysuccinimide; mucin or CEACAM1 was then injected and immobilized on the surface of the chip; finally, the remaining activated surface of the chip was blocked with ethanolamine. Soluble coronavirus NTD was introduced at a flow rate of 20 μl/min at different concentrations. Binding affinities were determined using BIA-EVALUATIONS software.

Mutagenesis—Site-directed mutagenesis was performed to introduce mutations into BCoV NTD (30). Briefly, the pFastbac I plasmid containing the BCoV NTD gene was PCR-amplified using two complementary oligonucleotides containing the desired mutations. The PCR product was digested by enzyme DpnI to remove the wild-type plasmid. The mutant plasmid that remained was transformed into DH5α competent cells, amplified, purified, and used to express the mutant protein in sf9 insect cells.

Glycan Screen Array—To determine the sugar-binding specificity of BCoV NTD, a glycan screen array was performed at the Consortium for Functional Glycomics. The printed glycan array (CFG version 5.0) was composed of 611 different natural and synthetic mammalian glycans (supplemental Table S1). In the binding assay, array slides were incubated with BCoV NTD with a C-terminal His$_6$ tag. The slides were then washed, and bound BCoV NTD was detected with mouse anti-His$_6$ antibody; readout was described arbitrarily as relative fluorescence unit. The intensity of binding to each of the 611 glycans on the array was graphed. Values represent means ± S.D.s of quadruplicate samples.

RESULTS AND DISCUSSION

Structure Determination—We expressed BCoV NTD (residues 15–298) in insect cells, purified it from insect cell culture medium, and crystallized it in space group P3$_1$21, with one BCoV NTD in each asymmetric unit. The crystal diffracted to 1.55 Å. Although the crystal was a twin, application of the twinning operator allowed the structure to be determined by molecular replacement using MHV NTD as the search model (Protein Data Bank code 3R4D) (Fig. 1, *A* and *B*). The structure of BCoV NTD has been refined to R_{work} of 16.3% and R_{free} of 17.7% (Table 1), again after the application of the twinning operator. The final model contains all of the residues of BCoV
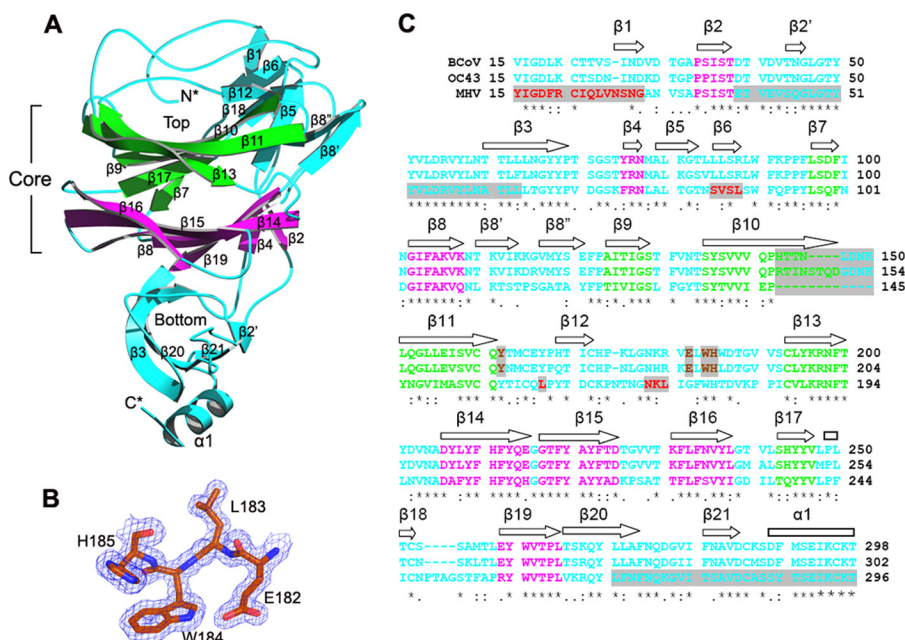
FIGURE 1. **Crystal structure of BCoV NTD.** *A*, overall structure of BCoV NTD. Two β-sheets of NTD core are colored *green* and *magenta*, respectively, and other parts of NTD are colored *cyan*. N[*], N terminus; C[*], C terminus. The β-sandwich core structure is indicated as "core." The two potential sugar-binding pockets above and underneath the core structure are indicated as *top* and *bottom*, respectively. *B*, $2F_o - F_c$ electron density of a portion of BCoV NTD at 1.5σ. This region includes three of the critical sugar-binding residues. *C*, secondary structures of BCoV NTD and sequence alignment of BCoV, HCoV-OC43, and MHV NTDs. β-Strands are shown as *arrows*, and α-helices are shown as *cylinders*. The sequences are colored the same way as the corresponding secondary structures in *A*. In MHV NTD, two highlighted regions, one covering β2′ and part of β3 and the other at the C terminus, are disordered (22). Also in MHV NTD, the four highlighted and *red-colored regions* are CEACAM1-binding RBMs (RBM1–4 from N to C terminus). In BCoV and HCoV-OC43 NTDs, the four highlighted and *brown*-colored residues between β11 and β13 are critical sugar-binding residues. In all three NTDs, the highlighted region covering part of β10 and loop 10–11 varies significantly in length. BCoV strain, Mebus; HCoV-OC43 strain, ATCC VR759; MHV strain, A59. *Asterisks* indicate positions that have fully conserved residues. *Colons* indicate positions that have strongly conserved residues. *Periods* indicate positions that have weakly conserved residues.

TABLE 1
Data statistics

Data collection	
Space group	$P3_121$
Cell parameter (a, b, c (Å))	86.56, 86.56, 78.07
α, β, γ	90.00°, 90.00°, 120.00°
Wavelength	0.97918
Resolution range (Å)	50.0–1.55
No. of reflections	46,536
% Completeness (last shell)	99.9 (99.5)
R_{merge} (last shell)	0.074 (0.497)
I/σ (last shell)	33.0 (3.0)
Redundancy (last shell)	4.7 (4.0)
Refinement	
Resolution (Å)	50.0–1.55
R_{work}/R_{free}	16.3%/17.7%
No. of atoms	2590
Protein	2298
Carbohydrate	71
Ion	5
Solvent	216
r.m.s.d.	
Bond length (Å)	0.005
Bond angle	0.941°
B factor	14.55
Ramachandran plot: core, allow, disallow	96.49%, 3.51%, 0.00%

FIGURE 2. **Stereo image of the superimposed structures of BCoV and MHV NTDs.** BCoV NTD is colored *blue*, and MHV NTD is colored *green*. Two of the mCEACAM1a-binding loops in MHV NTD are colored *red* and labeled as receptor-binding motifs 1 and 4 (RBM1 and RBM4). Sugar-binding residues in BCoV NTD are colored *brown* and shown in stick-and-ball presentation. *Bidirectional arrows* indicate different conformations of the receptor-binding loops in the two NTDs. *One-directional arrows* indicate the location of mCEACAM1a that binds MHV NTD.

NTD, three of the C-terminal His$_6$ tag, three *N*-linked glycans, five ions, and 216 solvent molecules.

Overall Structure—The overall structure of BCoV NTD is similar to, but significantly more complete than, that of MHV NTD (Figs. 1*A* and Fig. 2) (22), despite that the two NTDs have equivalent N and C termini (residues 15–298 for BCoV NTD and 15–296 for MHV NTD) (Fig. 1*C*). Similar to MHV NTD, BCoV NTD contains a β-sandwich core structure consisting of
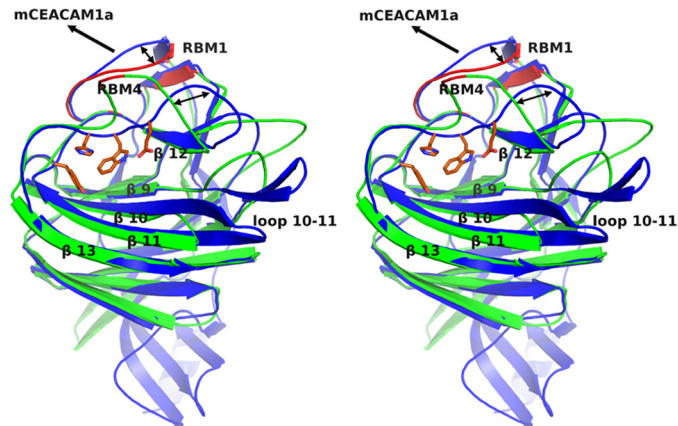
one six-stranded β-sheet and one seven-stranded β-sheet that are stacked together through hydrophobic interactions. This core structure has the same structural topology as human galectins. Also similar to MHV NTD, BCoV NTD contains several peripheral structural elements, mostly long loops and short β-sheets, on top of the core structure. Different from the MHV NTD structure, however, the BCoV NTD structure contains additional peripheral structural elements underneath the core
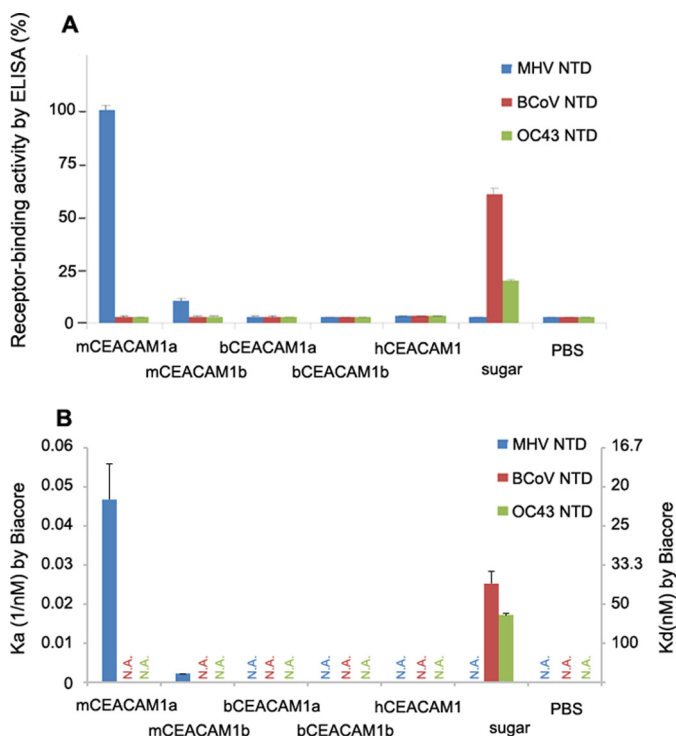
FIGURE 3. **Interactions between coronavirus NTDs and mammalian CEACAM1 proteins.** *A,* relative receptor-binding activities of coronavirus NTDs by ELISA. Measured were relative binding activities between coronavirus NTDs and mammalian CEACAM1 proteins that had been coated on 96-well Maxisorp plates. CEACAM1-binding NTDs were detected using antibodies against their C-terminal His$_6$ tags. As a comparison, binding activities between coronavirus NTDs and sugar moieties on mucin-coated plates were also shown. PBS buffer was used as a negative control. All of the binding activities have been calibrated against the binding activity between MHV NTD and mCEACAM1a. *B,* receptor-binding affinities of coronavirus NTDs by surface plasmon resonance using Biacore. Mammalian CEACAM1 proteins were immobilized on Biacore chips, and coronavirus NTDs were flown through. *N.A.* indicates that the binding affinity is too low to be reliably measured. As a comparison, binding affinities between coronavirus NTDs and sugar moieties on mucin-immobilized sensor chips were also shown. *mCEACAM1,* murine CEACAM1; *bCEACAM1,* bovine CEACAM1; *hCEACAM1,* human CEACAM1.

structures that were disordered in the MHV NTD structure (residues 39–63 and 271–298) (Fig. 1C). These additional structural elements of BCoV NTD form a four-stranded β-sheet and an α-helix that may in involved in interacting with other parts of the trimeric spike protein. Additionally, the MHV NTD structure in complex with mCEACAM1a was refined to 3.1 Å resolution, whereas the BCoV NTD structure has been refined to 1.55 Å resolution. The BCoV NTD structure should be highly homologous to the HCoV-OC43 NTD structure, which has not yet been determined, due to the high sequence homology between the two proteins (Fig. 1C). Overall, compared with the previous MHV NTD structure, the current BCoV NTD structure presents a significantly more complete and a much higher resolution view of a coronavirus NTD.

CEACAM1 Binding—We systematically characterized the interactions between coronavirus NTDs and mammalian CEACAM1 proteins *in vitro*, which had not been well characterized previously (Fig. 3, *A* and *B*). Both murine CEACAM1 and bovine CEACAM1 exist in two slightly different forms, CEACAM1a and CEACAM1b, which are encoded by two alleles (31–33). Conversely, human CEACAM1 has only one

form that is encoded by one allele. We expressed and purified each of these mammalian CEACAM1 proteins as well three coronavirus NTDs (BCoV, HCoV-OC43, and MHV) and performed NTD/CEACAM1 and NTD/sugar binding assays using both ELISA and Biacore. Our results show that MHV NTD binds mCEACAM1a with high affinity and mCEACAM1b with low affinity, which is consistent with previous studies (31, 33, 34). Our results also show that MHV NTD does not bind sugar or any of the CEACAM1 proteins from bovine or human and that BCoV and HCoV-OC43 NTDs only bind sugar, but not any of the CEACAM1 proteins from bovine, murine, or human (Fig. 3, *A* and *B*).

The differences in CEACAM1-binding specificities of coronavirus NTDs can be readily explained by the structural differences between BCoV and MHV NTDs (Fig. 2). Among the four mCEACAM1a-binding loops (RBMs 1–4) in MHV NTD, two of them (RBMs 1 and 4) have significantly different conformations from their counterparts in BCoV NTD. The more significant conformational difference is in RBM4, which is located in loop 12–13 (loop connecting β-strands 12 and 13). These structural differences between MHV and BCoV NTDs explain why BCoV and HCoV-OC43 NTDs cannot bind any of the mammalian CEACAM1 proteins.

Sugar Binding—Our efforts to determine the crystal structure of BCoV NTD complexed with sugar have been unsuccessful so far. Instead, to identify the sugar-binding site in BCoV NTD, we systematically performed alanine substitutions of residues in two potential sugar-binding pockets, one above the β-sandwich core and one underneath. We also grafted loop 10–11 from MHV NTD into BCoV NTD (Fig. 1C). This was based on the observation that compared with MHV NTD, both BCoV and HCoV-OC43 NTDs contain a long insertion in this region, and thus, we thought it may be involved in sugar binding (22). We expressed and purified each of these mutant BCoV NTDs. All of the mutant proteins showed the same expression levels, solubility, and chromatographic behaviors as the wild-type BCoV NTD. We performed sugar-binding assays on these mutant proteins using ELISA. Our results show that single alanine substitution for each of four residues, Tyr-162, Glu-182, Trp-184, and His-185, significantly decreased the sugar-binding affinity of BCoV NTD and that replacement of loop 10–11 abolished the sugar-binding affinity of BCoV NTD (Fig. 4, *A*, *C*, and *D*). We further confirmed these results by surface plasmon resonance using Biacore (Fig. 4B; Table 2). Mutations elsewhere in BCoV NTD did not affect sugar binding (Fig. 4, *A*, *C*, and *D*). These mutagenesis studies suggest that the pocket above the β-sandwich core is the sugar-binding site in BCoV NTD.

What type of sugar is preferred by BCoV NTD? Previous virus infection studies have shown that Neu5,9Ac2 can function as a receptor or co-receptor for BCoV (4, 5). However, it is not clear whether any other type of sugar may also have high affinity for BCoV NTD. In this study, we performed glycan screen arrays to evaluate the binding affinity between BCoV NTD and different types of sugar (Fig. 5 and supplemental Table S1). Of the 611 types of sugar that were screened, only Neu5,9Ac2 showed high affinity for BCoV NTD. Hence, BCoV NTD and BCoV hemagglutinin-esterase have the same pre-
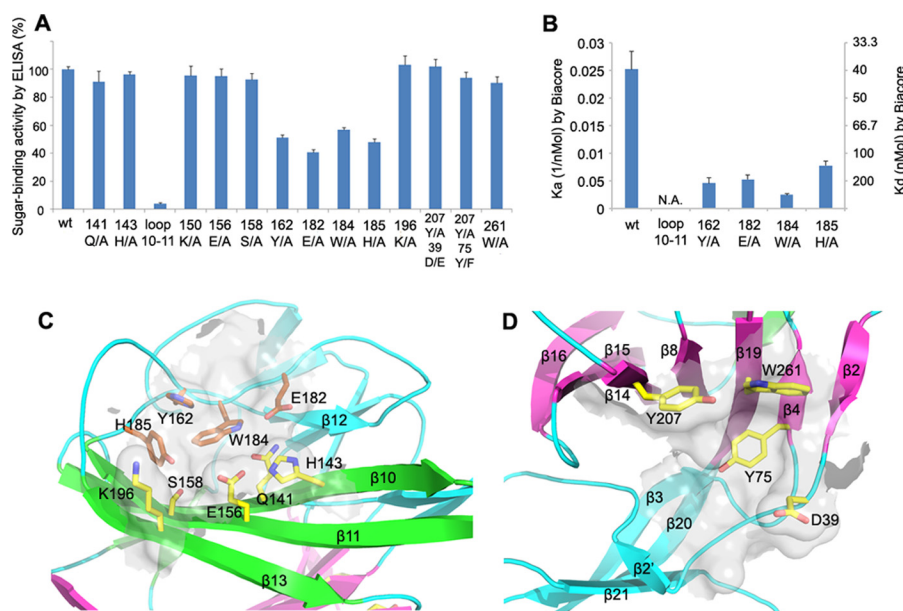
FIGURE 4. **Interactions between BCoV NTD and sugar.** *A*, relative binding activities between BCoV NTD and sugar moieties on mucin-coated plates by ELISA. Sugar-binding BCoV NTD was detected using antibodies against its C-terminal His$_6$ tag. Sugar-binding activities of both wild-type and mutant BCoV NTDs were measured. All of the sugar-binding activities have been calibrated against the sugar-binding activity of wild-type BCoV NTD. *B*, binding affinity between BCoV NTD and sugar moieties on mucin by surface plasmon resonance using Biacore. Mucin was immobilized on Biacore chips, and BCoV NTD was flown through. *C*, distribution of mutated residues in the pocket above the β-sandwich core. Critical sugar-binding residues are colored *brown*, and non-critical residues are colored *yellow*. *D*, distribution of mutated residues in the pocket underneath the β-sandwich core. Surface presentations of the pockets were shown as *semi-transparent white surfaces*. *N.A.*, not available.

TABLE 2
Binding kinetics by Biacore
Each experiment was repeated five times, and S.D. for measuring K_d were calculated and shown.

|	k_{on}	k_{off}	K_d
	$mol^{-1}\,s^{-1}$	s^{-1}	mol
mCEACAM1a/MHV NTD	4.41×10^5	9.44×10^{-3}	$2.14 \times 10^{-8} \pm 0.42 \times 10^{-8}$
mCEACAM1b/MHV NTD	6.40×10^4	3.14×10^{-3}	$4.91 \times 10^{-7} \pm 0.26 \times 10^{-7}$
BSM/BCoV NTD	2.35×10^5	9.33×10^{-3}	$3.97 \times 10^{-8} \pm 0.50 \times 10^{-8}$
BSM/OC43 NTD	2.48×10^5	1.45×10^{-2}	$5.85 \times 10^{-8} \pm 0.18 \times 10^{-8}$
BSM/BCoV NTD 162Y/A	1.58×105	3.46×10^{-2}	$2.19 \times 10^{-7} \pm 0.438 \times 10^{-7}$
BSM/BCoV NTD 182E/A	8.63×104	1.65×10^{-2}	$1.91 \times 10^{-7} \pm 0.28 \times 10^{-7}$
BSM/BCoV NTD 184W/A	4.77×104	1.97×10^{-2}	$4.14 \times 10^{-7} \pm 0.414 \times 10^{-7}$
BSM/BCoV NTD 185H/A	5.97×105	7.7×10^{-2}	$1.29 \times 10^{-7} \pm 0.134 \times 10^{-7}$
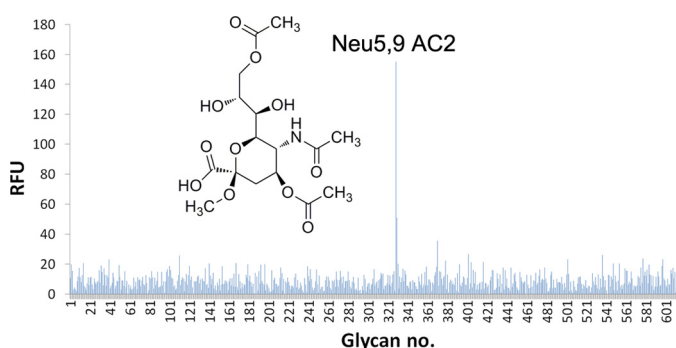
FIGURE 5. **Glycan screen array to identify substrate sugar type for BCoV NTD.** See supplemental Table S1 for glycans used in the experiment. Among these glycans, 5-*N*-acetyl-9-*O*-acetylneuraminic acid (Neu5,9Ac2) shows the highest affinity for BCoV NTD. *RFU*, relative fluorescence unit.

ferred sugar substrate (25), suggesting an elegant co-evolutionary relationship between the two proteins that allows coordination between viral attachment and detachment from host cells. It is also worth noting that galactose, the sugar substrate for human galectins, is not recognized by BCoV NTD (supplemental Table S1). Thus, BCoV NTD and human galectins recognize different types of sugar despite sharing the same fold in their core structures.

Based on the mutagenesis data and the structural comparison between BCoV NTD and human galectins, we suggest that Neu5,9Ac2 binds into the pocket above the β-sandwich core in the BCoV NTD structure and has direct contacts with residues Tyr-162, Glu-182, Trp-184, and His-185 (Fig. 6*A*). Although BCoV NTD and human galectins bind different types of sugar, the sugar-binding sites in the two proteins overlap (Fig. 6*B*). This is in contrast to rotavirus VP4, another viral lectin that also has a human galectin fold, but binds its sugar substrate 5-*N*-acetylneuraminic acid in a groove between the two β-sheet layers of its β-sandwich core structure (22, 35, 36). Therefore, although the human galectin fold is conserved in different viral lectins, the sugar-binding sites and sugar-binding specificity may vary depending on the viral lectin.

Structural comparison between BCoV and MHV NTDs explains why MHV NTD does not use sugar as a receptor (Fig. 2) (6). The four critical sugar-binding residues in BCoV NTD are distributed on two sugar-binding loops: Tyr-162 is located on loop 11–12, and Glu-182, Trp-184, and His-185 are on loop
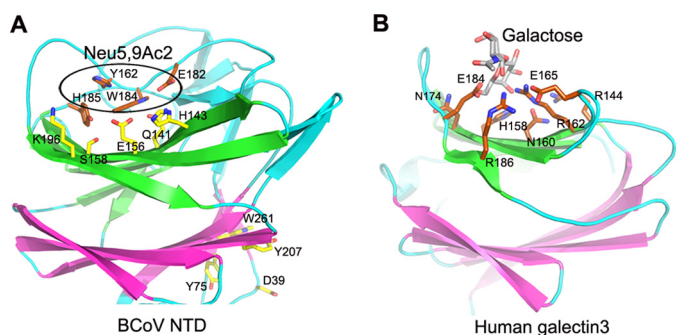
FIGURE 6. **Comparison of sugar-binding sites in BCoV NTD and human galectins.** *A*, sugar-binding site in BCoV NTD. *Black circle* indicates the Neu5,9Ac2-binding site in BCoV NTD, as identified by mutagenesis studies. Critical sugar-binding residues are colored *brown*, and non-critical residues are colored *yellow*. *B*, galactose-binding site in human galectin 3 (Protein Data Bank code 1A3K). Galactose is colored *gray*, and critical galactose-binding residues are colored *brown*.

FIGURE 7. **Proposed origin and evolution of coronavirus spike protein lectin domain.** *Orange arrows* indicate the locations of CEACAM1 or sugar that binds coronavirus NTDs. *Question marks* indicate the postulated structures of hypothetical evolutionary intermediates.

12–13. As discussed earlier, loop 12–13 in MHV NTD is one of the mCEACAM1a-binding sites (RBM4 for CEACAM1 binding). Compared with MHV NTD, loop 12–13 in BCoV NTD has a markedly different conformation that allows it to function as a sugar-binding loop and precludes its CEACAM1-binding capability. Additionally, a critical sugar-binding residue in BCoV NTD, Glu-182, is a glycine in MHV NTD (Fig. 1*C*). Compared with Glu-182, an alanine at this position in BCoV NTD significantly decreased sugar binding affinity (Fig. 4, *A* and *B*, and Table 2); thus, a glycine here may also decrease the sugar binding affinity due to the loss of the interactions between the glutamate side chain and the sugar. Curiously, despite being implicated previously as critical for sugar binding in BCoV NTD (22), loop 10–11 does not appear to be directly involved in sugar binding. Close inspection of the BCoV NTD structure suggests that loop 10–11 has extensive contacts with other loops over the β-sandwich core including the sugar-binding loop 12–13 (Fig. 2). Hence, loop 10–11 in BCoV NTD probably contributes indirectly to sugar binding by stabilizing the structure of the sugar-binding pocket, whereas a shortened loop 10–11 in MHV NTD abolishes sugar binding by altering the conformations of the sugar-binding loops. Overall, compared with BCoV NTD, different conformations of sugar-binding loops and substitution of critical sugar-binding residues together abolish any potential lectin function of MHV NTD.

Evolution of Coronavirus Spike Protein Lectin Domain—In this study, we have determined the crystal structure of BCoV spike protein NTD at 1.55 Å, characterized its sugar-binding activity and specificity, and compared its structure and function to those of CEACAM1-binding MHV NTD and galactose-binding host galectins. First, the high-resolution and complete structural view of coronavirus NTDs reveal that they have evolved additional peripheral structural elements that are not found in host galectins. These structural elements may interact with other parts of coronavirus spike proteins and/or may be used to recognize specific host receptors. Second, subtle structural differences between BCoV and MHV NTDs, primarily involving conformational differences in their receptor-binding loops, have significant functional outcomes. For example, one of the sugar-binding loops in BCoV NTD is an mCEACAM1a-binding loop in MHV NTD. As a result, MHV NTD does not
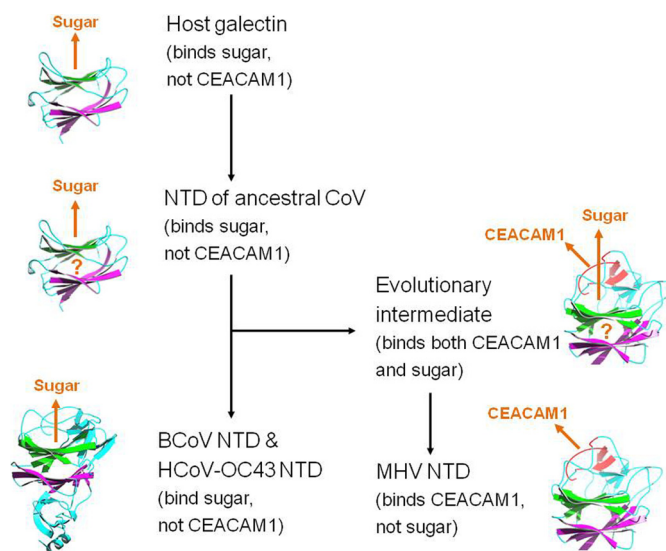
recognize sugar, whereas BCoV NTDs does not recognize CEACAM1. Third, although BCoV NTD and host galectins recognize different types of sugars, they share the same sugar-binding site. This finding supports the common evolutionary origin of these proteins but also suggests that coronavirus sugar-binding NTDs have diverged from host galectins in their sugar substrate specificities as part of viral adaptations to their host ranges and tropisms. Therefore, this study provides insights into the structures, functions, and evolution of coronavirus NTDs.

Whereas our previous structural study on MHV NTD suggested that coronavirus NTDs may have originated from a host galectin (22), the current study allows us to draw a clearer picture of how the evolution of coronavirus NTDs may have occurred (Fig. 7). Acquiring a lectin domain from their host cell and inserting it into their spike protein may have enabled ancestral coronaviruses to use sugars on the cell surface as their receptors, which enhanced cell entry efficiency of these viruses. Thus, the lectin function has been conserved in the NTDs of some contemporary coronaviruses such as BCoV and HCoV-OC43. It is unlikely that the sugar-binding specificity of contemporary BCoV and HCoV-OC43 NTDs evolved from CEACAM1-binding MHV NTD because it would be an evolutionary detour for coronaviruses to evolve lectin functions twice, first from host galectin and second from CEACAM1-binding NTD. Instead, we propose the opposite: the CEACAM1-binding specificity of contemporary MHV NTD evolved from sugar-binding coronavirus NTDs. In fact, as this study has demonstrated, no dramatic structural evolution of their NTDs was necessary for coronaviruses to switch from sugar-binding specificity to CEACAM1-binding specificity. There might even have existed some coronaviruses that were evolutionary intermediates between sugar-binding coronaviruses and CEACAM1-binding coronaviruses. These evolutionary intermediates might have been able to use both CEACAM1 and sugar as receptors. Because protein receptors in general

provide higher affinity and specificity for virus binding than sugar receptors do, the spike protein NTDs of these hypothetical evolutionary intermediates may have subsequently lost their lectin function, leading to the emergence of contemporary MHV. The existence and maintenance of an hemagglutinin-esterase gene in the genomes of many MHV strains, whether silent or active expressing, support the hypothesis that the spike protein NTD of ancestral MHV could function as a viral lectin. Overall, it appears that coronaviruses adopted a successful evolutionary strategy when they stole a host protein and evolved it into viral receptor-binding domains with altered sugar receptor specificity as in contemporary BCoV or novel protein receptor specificity as in contemporary MHV.

REFERENCES

1. Li, W., Wong, S. K., Li, F., Kuhn, J. H., Huang, I. C., Choe, H., and Farzan, M. (2006) Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. *J. Virol.* **80,** 4211–4219
2. Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A. M., and Van Ranst, M. (2005) Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J. Virol.* **79,** 1595–1604
3. Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M., and 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
4. Künkel, F., and Herrler, G. (1993) Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **195,** 195–202
5. Schultze, B., Gross, H. J., Brossmer, R., and 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
6. Langereis, M. A., van Vliet, A. L., Boot, W., and de Groot, R. J. (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
7. Dveksler, G. S., Pensiero, M. N., Cardellichio, C. B., Williams, R. K., Jiang, G. S., Holmes, K. V., and Dieffenbach, C. W. (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
8. Williams, R. K., Jiang, G. S., and Holmes, K. V. (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
9. Beauchemin, N., Draber, P., Dveksler, G., Gold, P., Gray-Owen, S., Grunert, F., Hammarström, S., Holmes, K. V., Karlsson, A., Kuroki, M., Lin, S. H., Lucka, L., Najjar, S. M., Neumaier, M., Obrink, B., Shively, J. E., Skubitz, K. M., Stanners, C. P., Thomas, P., Thompson, J. A., Virji, M., von Kleist, S., Wagener, C., Watt, S., and Zimmermann, W. (1999) Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp. Cell Res.* **252,** 243–249
10. Cavanagh, D., and Davis, P. J. (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
11. Schultze, B., Cavanagh, D., and 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
12. Krempl, C., Schultze, B., Laude, H., and 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
13. Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L. K., Sjöström, H., Norén, O., and Laude, H. (1992) Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* **357,** 417–420
14. Yeager, C. L., Ashmun, R. A., Williams, R. K., Cardellichio, C. B., Shapiro, L. H., Look, A. T., and Holmes, K. V. (1992) Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357,** 420–422
15. Delmas, B., Gelfi, J., Sjöström, H., Noren, O., and Laude, H. (1993) Further characterization of aminopeptidase-N as a receptor for coronaviruses. *Adv. Exp. Med. Biol.* **342,** 293–298
16. Hofmann, H., Simmons, G., Rennekamp, A. J., Chaipan, C., Gramberg, T., Heck, E., Geier, M., Wegele, A., Marzi, A., Bates, P., and Pöhlmann, 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
17. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H., and Farzan, M. (2003) Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454
18. Chen, L., Lin, Y. L., Peng, G. Q., and Li, F. (2012) *Proc. Natl. Acad. Sci. U.S.A.* **109,** 17966–17971
19. Li, F., Berardi, M., Li, W., Farzan, M., Dormitzer, P. R., and Harrison, S. C. (2006) Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J. Virol.* **80,** 6794–6800
20. Li, F. (2012) Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J. Virol.* **86,** 2856–2858
21. Li, F., Li, W., Farzan, M., and Harrison, S. C. (2005) Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309,** 1864–1868
22. Peng, G., Sun, D., Rajashankar, K. R., Qian, Z., Holmes, K. V., and 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
23. Wu, K., Li, W., Peng, G., and Li, F. (2009) Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19970–19974
24. Reguera, J., Santiago, C., Mudgal, G., Ordoño, D., Enjuanes, L., and Casasnovas, J. M. (2012) Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLOS Pathog.* **8,** e1002859
25. Zeng, Q., Langereis, M. A., van Vliet, A. L., Huizinga, E. G., and de Groot, R. J. (2008) Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 9065–9069
26. Yokomori, K., Banner, L. R., and Lai, M. M. (1991) Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology* **183,** 647–657
27. Yeates, T. O. (1988) Simple statistics for intensity data from twinned specimens. *Acta Crystallogr. A* **44,** 142–144
28. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
30. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.* **32,** e115
31. Ohtsuka, N., Yamada, Y. K., and Taguchi, F. (1996) Difference in virus-binding activity of two distinct receptor proteins for mouse hepatitis virus. *J. Gen. Virol.* **77,** 1683–1692
32. Kammerer, R., Popp, T., Singer, B. B., Schlender, J., and Zimmermann, W. (2004) Identification of allelic variants of the bovine immune regulatory molecule CEACAM1 implies a pathogen-driven evolution. *Gene* **339,**

99–109

33. Wessner, D. R., Shick, P. C., Lu, J. H., Cardellichio, C. B., Gagneten, S. E., Beauchemin, N., Holmes, K. V., and Dveksler, G. S. (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

34. Compton, S. R., Stephensen, C. B., Snyder, S. W., Weismiller, D. G., and Holmes, K. V. (1992) Coronavirus species specificity: murine coronavirus binds to a mouse-specific epitope on its carcinoembryonic antigen-related receptor glycoprotein. *J. Virol.* **66,** 7420–7428

35. Dormitzer, P. R., Sun, Z. Y., Wagner, G., and Harrison, S. C. (2002) The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21,** 885–897

36. Haselhorst, T., Blanchard, H., Frank, M., Kraschnefski, M. J., Kiefel, M. J., Szyczew, A. J., Dyason, J. C., Fleming, F., Holloway, G., Coulson, B. S., and von Itzstein, M. (2007) STD NMR spectroscopy and molecular modeling investigation of the binding of *N*-acetylneuraminic acid derivatives to rhesus rotavirus VP8* core. *Glycobiology* **17,** 68–81

PROTEIN STRUCTURE AND FOLDING
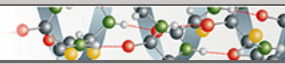
Access the most updated version of this article at doi: 10.1074/jbc.M112.418210

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:

http://www.jbc.org/content/suppl/2012/10/22/M112.418210.DC1.html

This article cites 36 references, 23 of which can be accessed free at
http://www.jbc.org/content/287/50/41931.full.html#ref-list-1