Crystal Structure of Bovine Coronavirus Spike Protein Lectin Domain*\(\text{**}\)

Received for publication, September 10, 2012, and in revised form, October 19, 2012. Published, JBC Papers in Press, October 22, 2012, DOI 10.1074/jbc.M112.418210

Guiqing Peng \(^1\), Liqing Xu \(^1\), Yi-Lun Lin \(^3\), Lang Chen \(^1\), Joseph R. Pasquarella \(^3\), Kathryn V. Holmes \(^5\), and Fang Li \(^1,2\)

From the \(^3\)Department of Pharmacology, University of Minnesota Medical School, Minneapolis, Minnesota 55455 and \(^5\)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-acetyleneuraminic 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: \(\alpha\), \(\beta\), and \(\gamma\). Bovine coronavirus (BCoV), mouse OC43 coronavirus (HCoV-OC43), and mouse hepatitis coronavirus (MHV) all belong to the \(\beta\)-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-acetyleneuraminic 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-glycolyneuraminic acid and 5-N-acetyleneuraminic acid, can serve as receptors or co-receptors for some \(\alpha\)-genus and \(\gamma\)-genus coronaviruses (10–12), whereas two other cell-surface proteins, angiotensin-converting enzyme 2 and aminopeptidase N, can serve as receptors for some \(\alpha\)-genus and \(\beta\)-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-nitri-lotriacetic acid gel 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 l of protein solution and 1 l well solution containing 2.0 M (NH$_4$)$_2$SO$_4$. Crystals diffraction 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 μl of protein solution and 1 μl well solution containing 2.0 M (NH$_4$)$_2$SO$_4$. Crystals diffraction 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).

**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 P3,21, 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 NTD.
NTD, three of the C-terminal His6 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. 1A 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. 1C). 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.

<table>
<thead>
<tr>
<th>Data statistics</th>
<th>Collection</th>
<th>Space group</th>
<th>P3, 21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cell parameter (a, b, c (Å))</td>
<td>86.56, 86.56, 78.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>α, β, γ</td>
<td>90.00°, 90.00°, 120.00°</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wavelength</td>
<td>0.97918</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Resolution range (Å)</td>
<td>50.0–1.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. of reflections</td>
<td>46,536</td>
<td></td>
</tr>
<tr>
<td></td>
<td>% Completeness (last shell)</td>
<td>99.9 (99.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rmerge (last shell)</td>
<td>0.074 (0.497)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>completeness</td>
<td>33.0 (3.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundancy (last shell)</td>
<td>4.7 (4.0)</td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td>Resolution (Å)</td>
<td>50.0–1.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rwork/Rfree</td>
<td>16.3%/17.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. of atoms</td>
<td>2590</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein</td>
<td>2298</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carbohydrate</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ion</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Solvent</td>
<td>216</td>
<td></td>
</tr>
<tr>
<td>r.m.s.d.</td>
<td>Bond length (Å)</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bond angle</td>
<td>0.941°</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B factor</td>
<td>14.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramachandran plot: core, allow, disallow</td>
<td>96.49%, 3.51%, 0.00%</td>
<td></td>
</tr>
</tbody>
</table>

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, 2Fo – F, 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 β1 and part of β3 and the other at the C terminus, are disordered (22). Also in MHV NTD, the four highlighted and brown-colored regions 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.

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.
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 $\beta$-sheet and an $\alpha$-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, and bovine CEACAM1 exist in two slightly different forms, mCEACAM1a 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 (RBM 1–4) in MHV NTD, two of them (RBM 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 $\beta$-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 $\beta$-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 $\beta$-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-
measured. All of the sugar-binding activities have been calibrated against the sugar-binding activity of wild-type BCoV NTD.

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 \( \beta \)-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 \( \beta \)-sandwich core. Surface presentations of the pockets were shown as semi-transparent white surfaces. N.A., not available.

**TABLE 2**

<table>
<thead>
<tr>
<th>Sugar Type</th>
<th>( k_{in} ) (mol ( \cdot ) s ( ^{-1} ))</th>
<th>( k_{cat} ) (s ( ^{-1} ))</th>
<th>( K_s ) (mol ( \cdot ) l ( ^{-1} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCEACAM1a/MHV NTD</td>
<td>4.41 ( \times ) 10^4</td>
<td>9.44 ( \times ) 10^3</td>
<td>2.14 ( \times ) 10^-6</td>
</tr>
<tr>
<td>mCEACAM1b/MHV NTD</td>
<td>6.40 ( \times ) 10^4</td>
<td>3.14 ( \times ) 10^2</td>
<td>4.91 ( \times ) 10^-7</td>
</tr>
<tr>
<td>BSM/OC43 NTD</td>
<td>2.35 ( \times ) 10^3</td>
<td>9.33 ( \times ) 10^3</td>
<td>3.97 ( \times ) 10^-8</td>
</tr>
<tr>
<td>BSM/BCoV NTD 162Y/A</td>
<td>2.48 ( \times ) 10^2</td>
<td>1.45 ( \times ) 10^2</td>
<td>5.85 ( \times ) 10^-8</td>
</tr>
<tr>
<td>BSM/BCoV NTD 182E/A</td>
<td>1.58 ( \times ) 10^5</td>
<td>3.46 ( \times ) 10^2</td>
<td>2.19 ( \times ) 10^-7</td>
</tr>
<tr>
<td>BSM/BCoV NTD 184W/A</td>
<td>8.63 ( \times ) 10^4</td>
<td>1.65 ( \times ) 10^2</td>
<td>1.91 ( \times ) 10^-7</td>
</tr>
<tr>
<td>BSM/BCoV NTD 185H/A</td>
<td>4.77 ( \times ) 10^4</td>
<td>1.97 ( \times ) 10^2</td>
<td>4.14 ( \times ) 10^-4</td>
</tr>
</tbody>
</table>

**FIGURE 4. Interactions between BCoV NTD and sugar.** A, relative binding activities between BCoV NTD and sugar moieties on mucin-coated plates by ELISA. 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 \( \beta \)-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 \( \beta \)-sandwich core. Surface presentations of the pockets were shown as semi-transparent white surfaces. N.A., not available.

**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-acetylenamuramic acid (Neu5,9Ac2) shows the highest affinity for BCoV NTD. RFU, relative fluorescence unit.
12–13. As discussed earlier, loop 12–13 in MHV NTD is one of the mCEACAM1α-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. 1C). 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 mCEACAM1α-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
Structure, Function, and Evolution of Coronavirus Lectin Domain

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.

Acknowledgments—We thank the staff at APS beamline 24-ID for assistance in data collection, Dr. Robert Kammerer for cDNA clones expressing bovine CEACAM1 proteins, and the Consortium for Functional Glycomics for help in glycan screen arrays. Computer resources were provided by the Basic Sciences Computing Laboratory of the University of Minnesota Supercomputing Institute.

REFERENCES


