Structural and Molecular Evidence Suggesting Coronavirus-Driven Evolution of Mouse Receptor

Guiqing Peng [1, 2, #]**, Yang Yang** [1, #]**, Joseph R. Pasquarella** [1]**, Liqing Xu** [1, 3]**, Zhaohui Qian** [4,5]**, Kathryn V. Holmes** [5]**, and Fang Li** [1, *]

From
[1] Department of Pharmacology, University of Minnesota Medical School, Minneapolis, MN 55455, USA
[2] State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, Hubei China
[3]Key Laboratory of Horticultural Plant Biology, Huazhong Agricultural University, Wuhan 4300 70, Hubei, China
[4] MOH Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China
[5] Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO 80045, USA

# These authors contributed equally to this work.

To whom correspondence should be addressed: Fang Li, Department of Pharmacology, University of Minnesota Medical School, Minneapolis, MN 55455, Telephone: (612) 625-6149; Fax: (612) 625-8408; Email: lifang@umn.edu.

Running title: Crystal structure of mouse CEACAM1b

ABSTRACT

Hosts and pathogens are locked in an evolutionary arms race. To infect mice, mouse hepatitis coronavirus (MHV) has evolved to recognize mouse CEACAM1a (mCEACAM1a) as its receptor. To elude MHV infections, mice may have evolved a variant allele from the *Ceacam1a* gene, called *Ceacam1b*, producing mCEACAM1b that is a much poorer MHV receptor than mCEACAM1a. Previous studies showed that sequence differences between mCEACAM1a and mCEACAM1b in a critical MHV-binding CC' loop partially account for the low receptor activity of mCEACAM1b, but detailed structural and molecular mechanisms for the differential MHV receptor activities of mCEACAM1a and mCEACAM1b remained elusive. Here we have determined the crystal structure of mCEACAM1b, and identified the structural differences and additional residue differences between mCEACAM1a and mCEACAM1b that affect MHV binding and entry. These differences include conformational alterations of the CC' loop as well as residue variations in other MHV-binding regions, including β-strands C' and C'' and loop C'C''. Using pseudovirus entry and protein-protein binding assays, we show that substituting the structural and residue features from mCEACAM1b into mCEACAM1a reduced the viral receptor activity of mCEACAM1a, whereas substituting the reverse changes from mCEACAM1a into mCEACAM1b increased the viral receptor activity of mCEACAM1b. These results elucidate the detailed molecular mechanism for how mice may have kept pace in the evolutionary arms race with MHV by undergoing structural and residue changes in

1

the MHV receptor, providing insight into this possible example of pathogen-driven evolution of a host receptor protein.

According to the Red Queen hypothesis, hosts and pathogens are in an evolutionary arms race to keep pace with each other for fitness and survival (1,2). Coronaviruses are a large family of ancient and diverse RNA virus pathogens that infect many mammalian and avian species (3,4). Different coronaviruses use a variety of cell surface receptors for entry into host cells through the activities of virus-surface spike proteins (5,6). The host receptor-adapting evolution of coronavirus spike proteins has been extensively studied (3-6), but coronavirus-driven evolution of host receptors is much less well understood. The current study investigates how a host receptor may undergo molecular changes under possible selective pressure from lethal coronavirus infections and how these changes may help the host to resist death from coronavirus infections.

As the prototypic member of the coronavirus family, mouse hepatitis coronavirus (MHV) presents a good model system for studying the co-evolutionary relationship between viruses and hosts. Depending on the strain, MHV can cause enteric, respiratory, or brain infections in mice. The enterotropic strains of MHV spread widely in susceptible mouse populations and are lethal in infant mice of many inbred strains (up to 100% fatality) (7). Infection with MHV is a major concern in laboratory mice because it can disrupt mouse-based research through clinical disease and/or alteration of immunologic responses (8). MHV uses a cell-surface protein, mouse Carcinoembryonic Antigen-related Cell Adhesion Molecule 1a (mCEACAM1a), as its host receptor (9,10). The CEACAM1 protein is widely expressed in all mammals on the membranes of epithelial cells, endothelial cells, and leukocytes (11). It mediates cell-cell adhesion and signaling, and participates in the differentiation and arrangement of tissue three-dimensional structure, angiogenesis, apoptosis, tumor suppression, cancer metastasis, and the modulation of innate and adaptive immune responses (12). The envelope-anchored MHV spike glycoprotein specifically recognizes mCEACAM1a through the N-terminal domain of its S1 subunit (S1-NTD) (9,13,14). Our previous structural studies revealed that coronavirus S1-NTDs have the same tertiary structural fold as human galectins (galactose-binding lectins), and that whereas MHV S1-NTD recognizes mCEACAM1, bovine coronavirus (BCoV) S1-NTD recognizes sugar (15,16). Thus, we proposed that coronaviruses acquired a host galectin gene and inserted it into its spike protein gene, and that whereas BCoV S1-NTD has kept its original sugar-binding lectin activity, MHV S1-NTD has evolved novel mCEACAM1a-binding affinity and lost its original sugar-binding lectin activity. These studies have provided insight into the host receptor-adapting evolution of coronaviruses (5,6).

To respond to the selective pressure from lethal MHV infections, mice may have evolved a variant allele from the *Ceacam1a* gene, called *Ceacam1b*; of the two gene products, mCEACAM1b is a much less efficient MHV receptor than mCEACAM1a (17-19). Correspondingly, mice homozygous for *Ceacam1b* (1b/1b) are resistant to death from MHV infections, while mice homozygous for *Ceacam1a* (1a/1a) are highly susceptible to lethal MHV infections (7,9,10,20,21). Other than their different MHV receptor activities, mCEACAM1a and mCEACAM1b appear to be functionally equivalent: neither *1a/1a* mice or *1b/1b* mice show any growth defects, while a dysfunctional *Ceacam1* gene leads to impaired insulin clearance, abnormal weight gain, and reduced fertility (22). Our previous structural studies of mCEACAM1a and its complex with MHV S1-NTD have delineated detailed interactions between mCEACAM1a and MHV S1-NTD (15,23). Moreover, previous studies identified the CC' loop (loop that connects β-strands C and C') in mCEACAM1a as critical for MHV binding; the sequence of this loop diverges in mCEACAM1b, partially accounting for the low MHV receptor activity of mCEACAM1b (17,24). However, due to

2

the lack of structural information about mCEACAM1b, it was not known what structural differences between mCEACAM1a and mCEACAM1b or whether additional residue differences between mCEACAM1a and mCEACAM1b account for the MHV resistance in mice homozygous for *Ceacam1b*.

In this study, we have determined the crystal structure of mCEACAM1b, and elucidated the structural differences and additional residue differences between mCEACAM1a and mCEACAM1b that impede the binding of MHV S1-NTD to mCEACAM1b. Moreover, we have performed structure-guided mutagenesis studies on mCEACAM1a and mCEACAM1b to investigate the significance of their structural and sequence differences upon their MHV receptor activities. These results provide insight into the possibility that MHV has driven the evolution of the mCEACAM1 protein in mice.

RESULTS

Because of alternative mRNA splicing, mCEACAM1 contains either two [D1 and D4] or four [D1-D4] Ig-like domains in tandem, in addition to a transmembrane anchor and a short intracellular tail at its C-terminus (12). mCEACAM1b[D1,D4] (residues 1-202) without the membrane anchor or the intracellular tail was expressed and purified as previous described for mCEACAM1a[D1,D4] (15). It was subsequently crystallized in space group $P3_121$, a=113.1Å, b=113.1Å, and c=64.4Å. Although each asymmetric unit of the crystal contains two mCEACAM1b[D1,D4] molecules, the protein is a monomer in solution based on gel filtration chromatography. The structure was determined by molecular replacement using the structure of mCEACAM1a[D1,D4] as the search template, and refined at 3.1Å resolution (Fig. 1A; Table 1). The final model contains all of the residues in domains D1 and D4, and a glycan N-linked to Asn270.

The overall structure of mCEACAM1b[D1,D4] is similar to that of mCEACAM1a[D1,D4], but the structural

similarity is uneven in different regions of the two proteins. In both mCEACAM1a and mCEACAM1b, the two Ig-like domains, D1 and D4, are arranged in tandem without any significant interactions with each other (Fig. 1A, 1B). In mCEACAM1a, the D1 domain binds to MHV S1-NTD, whereas the D4 domain has no contact with MHV S1-NTD (Fig. 1C). In the D1 domain of mCEACAM1a, several loops (CC', C'C", C"D, and FG) and β-strands (βC, βC', and βC") are directly involved in MHV binding, and thus these regions have been called the virus-binding motifs (VBMs) (Fig. 1C, 1D). Interestingly, the D1 domains of mCEACAM1a and mCEACAM1b are significantly more divergent in both primary structure (sequence identity = 74%) and tertiary structure (main chain RMSD = 1.11 Å) than the D4 domains (sequence identity = 98%; main chain RMSD = 0.77 Å) (Fig. 2). Furthermore, within the D1 domain, the VBMs of mCEACAM1a and mCEACAM1b are more divergent in both primary structure (sequence identity = 56%) and tertiary structure (main chain RMSD = 1.36 Å) than the non-virus-binding regions are (Fig. 2C, 2D). These results suggest that compared with the rest of the protein, the VBMs in the D1 domains of mCEACAM1a have been under strong selective pressure possibly from MHV infections.

Further inspection of the structures of mCEACAM1a and mCEACAM1b has identified detailed structural divergence between the VBMs of the two proteins. In mCEACAM1a, a critical CC' loop (loop that connects β-strands C and C') in the D1 domain interacts extensively with MHV S1-NTD and thus plays a prominent role in the virus/receptor binding interactions (Fig. 3A). These interactions include the multiple hydrophobic interactions between the side chain of receptor Ile41 and the side chains of MHV Tyr15, Leu89, and Leu160 as well as the hydrogen bonds between the carbonyl oxygen of receptor Thr39 and the side chain of MHV Arg20. Compared to mCEACAM1a, mCEACAM1b has undergone significant structural changes in loop CC' (Fig. 3B, 3C), which result from residue changes in this loop. For example, residue 38 is a threonine in

mCEACAM1a but a proline in mCEACAM1b. This residue change likely has a significant impact on the conformation of loop CC' because prolines are known to cause changes to protein secondary structures. Additionally, a number of residues in other VBM regions of mCEACAM1a also form critical interactions with MHV S1-NTD, but have been substituted with different residues in mCEACAM1b. For example, in mCEACAM1a, the side chain of receptor Arg47 in strand βC' forms hydrogen bonds with the carbonyl oxygens of MHV Gln23 and Val25, the side chains of receptor Met54 and Phe56 in strand βC' are part of a hydrophobic cluster at the S1-NTD/receptor interface, and the side chain of receptor Asn59 in strand βC' forms hydrophobic stacking with the Cα of S1-NTD Gly29 (Fig. 4A, 4B). However, Arg47, Met54, Phe56, and Asn59 in mCEACAM1a have been substituted with His47, Lys54, Thr56, and Pro59, respectively, in mCEACAM1b (Fig. 4C, 4D). The above structural and residue changes in the VBMs from mCEACAM1a to mCEACAM1b would lead to the loss of numerous energetically favorable interactions at the S1-NTD/receptor interface and disrupt the virus/receptor binding interactions. These structural analyses further suggest that the VBMs in the D1 domain of mCEACAM1 have been under strong selective pressure possibly from MHV infections.

To investigate how the structural and residue differences between mCEACAM1a and mCEACAM1b affect their functions as MHV receptor, we carried out structure-guided mutagenesis and introduced structural and residue features from mCEACAM1b into mCEACAM1a. These structural and residue changes include replacing loop CC' in mCEACAM1a with its counterpart from mCEACAM1b, and substituting residues 47, 54, 56, and 59 in mCEACAM1a with the corresponding residues from mCEACAM1b. A pseudovirus entry assay was performed where a lentiviral vector pseudotyped with the MHV spike protein was used to enter mammalian cells expressing either wild type or mutant mCEACAM1a on their surface. The results demonstrated that each of the structural

and residue features from mCEACAM1b introduced into mCEACAM1a significantly reduced the efficiency of pseudovirus entry (Fig. 5A), reflecting a weaker binding affinity between the MHV spike protein and the mutant mCEACAM1a. Thus, the structural and residue changes from mCEACAM1a to mCEACAM1b reduced the capability of mCEACAM1a to serve as the MHV receptor. These loss-of-function experiments mimic the possible loss-of-function evolution of the mouse *Ceacam1a* gene under the selective pressure from MHV infections.

To further explore the functional significance of the structural and residue differences between mCEACAM1a to mCEACAM1b, we introduced the reverse substitutions (i.e., the features from mCEACAM1b introduced into mCEACAM1a). These structural and residue changes include replacing (i) loop CC', (ii) both loop CC' and strand βC', or (iii) all of loop CC', strand βC', loop C'C", and strand βC" in mCEACAM1b with the corresponding regions from mCEACAM1a. In addition to the pseudovirus entry assay, protein-protein binding assays were also performed between the MHV S1-NTD and wild type or mutant mCEACAM1b. The results showed that all of the structural and residue changes introduced into mCEACAM1b significantly enhanced both the pseudovirus entry efficiency and protein-protein binding affinity (Fig. 5A, 5B). Among the mutant mCEACAM1b molecules, the one containing changes in loop CC', strand βC', loop C'C", and strand βC" all together demonstrated the highest MHV receptor activity. More specifically, introduction of the above structural and residue changes into mCEACAM1b restored the receptor activity of mCEACAM1b up to ~67% of mCEACAM1a based on the pseudovirus entry efficiency and ~83% of mCEACAM1a based on the protein-protein binding affinity. It is worth noting that incorporation of the above structural and residue features from mCEACM1a did not fully restore the MHV receptor activity of mCEACAM1b to the same level as mCEACAM1a, suggesting that structural and/or residue differences elsewhere in domain D1 may account for the remaining

4

difference between mCEACAM1a and mCEACAM1b in their MHV receptor activities. Nevertheless, these gain-of-function experiments represent the reverse course of the possible loss-of-function evolution of the mouse *Ceacam1* gene under the selective pressure from MHV infections.

DISCUSSION

The Red Queen hypothesis states that hosts and pathogens are constantly in an evolutionary arms race. Previous structural studies of the coronavirus/receptor interactions have revealed how coronaviruses have evolved a variety of strategies to recognize different host receptors for host range expansion and cross-species infections (5,6,15,16,23,25-31). One of these strategies would be for coronaviruses to steal a host galectin, which became the S1-NTD of the coronavirus spike protein, and use it to bind sugar on host cell surfaces for viral attachment to host cells. While the S1-NTDs of many contemporary coronaviruses still recognize sugar receptors (32-38), MHV S1-NTD has evolved novel binding affinity for mCEACAM1a protein, which greatly enhanced the infection efficiency of MHV in mouse cells (15,16) (Fig. 6). Through the above evolution of its spike protein, MHV appears to have gained a significant edge in the evolutionary arms race with mice and become a highly infectious and pathogenic virus for mice.

How have mice evolved to keep pace in the evolutionary arms race with MHV? An interesting observation is that the mouse *Ceacam1* gene has diversified into two alleles, *Ceacam1a* and *Ceacam1b*. Their protein products, mCEACAM1a and mCEACAM1b, demonstrate different MHV receptor activities: mCEACAM1b is a much poorer MHV receptor than mCEACAM1a. Consequently, mice homozygous for *Ceacam1b* are highly resistant to death from MHV infections. The selective pressures that drive the evolution of mammalian *Ceacam1* genes could come from several sources. Mammalian CEACAM1 functions in many physiological processes including cell-cell adhesion, cell signaling, and cell development

(12). In addition, human CEACAM1 is a receptor for a variety of bacterial pathogens (39-41). Thus, the physiological functions of CEACAM1, host evasion of bacterial infections, or some other unknown selective pressures could potentially drive the evolution of mammalian *Ceacam1* genes. Although the physiological functions of mCEACAM1b need to be investigated further, mice homozygous for *Ceacam1b* apparently retain all the normal phenotypes of *Ceacam1a*, suggesting no major alterations of the physiological functions of mCEACAM1b. Moreover, mouse CEACAM1a is not a receptor for those bacterial pathogens that use human CEACAM1 as the receptor (42). On the other hand, MHV infections can be devastating to infant mice that express mCEACAM1a. Therefore, while other selective pressures cannot be ruled out, MHV infection is likely one of the major driving forces for the evolution of mouse *Ceacam1* gene. There are insufficient data on mouse genomes to prove which one of the mouse *Ceacam1* alleles evolved first. Based on the above discussion, we suggest that in the mouse population the *Ceacam1a* allele preceded the appearance and maintenance of the *Ceacam1b* allele in the presence of MHV epidemics.

This study investigates the structural and residue differences between mCEACAM1a and mCEACAM1b that render mCEACAM1b a less efficient MHV receptor. Previous studies identified a critical MHV-binding loop CC' that diverges in sequence between mCEACAM1a and mCEACAM1b, partially accounting for the different MHV receptor activities of the two proteins (17,24). The current study reveals the altered conformation of loop CC' in the crystal structure of mCEACAM1b, providing a structural basis for the critical role of loop CC' in the different MHV receptor activities of the two mCEACAM1 molecules. Furthermore, this study identifies residue variations in several other MHV-binding regions in mCEACAM1b that render mCEACAM1b a poor MHV receptor. These regions include β-strands C' and C'' and loop C'C''. Using structure-guided mutational and functional

5

assays, this study shows that the structural and residue substitutions from mCEACAM1a into mCEACAM1b cause a loss of receptor function in mCEACAM1a, whereas the reverse substitutions cause a gain of receptor function in mCEACAM1b. The structural and residue changes from mCEACAM1a to mCEACAM1b mimic the possible loss-of-function evolution in mCEACAM1a during pathogen-driven host evolution, which would result in less severe MHV infections in mice and partial alleviation of the selective pressure from MHV infections (Fig. 6). Therefore, it is likely that through divergent evolution of its *Ceacam1* gene to generates the *Ceacam1b* allele, mice may have gained the ability to keep pace in the evolutionary arms race with MHV for fitness and survival. Overall, the current study provides insight into a possible example of coronavirus-driven evolution of mouse receptor protein, lending molecular evidence to the Red Queen hypothesis.

MATERIALS AND METHODS

Protein preparation and crystallization- mCEACAM1b[D1, D4] (residues 1-202) was expressed and purified as previously described for mCEACAM1a[D1,D4] (15). Briefly, mCEACAM1b[D1, D4] containing a C-terminal His$_6$ tag was expressed in sf9 insect cells using the Bac-to-Bac expression system (Life Technologies), and was secreted into cell culture medium. The protein was harvested and 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). The protein was concentrated to 10 mg/ml and stored in buffer containing 20 mM Tris pH7.2 and 200 mM NaCl. Crystallization of mCEACAM1b[D1, D4] was set up using the sitting drop vapor diffusion method, with 1 μl protein solution added to 1 μl well buffer containing 0.1 M Tris pH6.2, 10% PEG4000 (v/v), and 1 M NaCl at 20°C. Crystals of mCEACAM1b[D1, D4] appeared in 2-3 days and were allowed to grow for another 2 weeks before they were harvested and flash-frozen in liquid nitrogen.

Data collection and structure determination- X-ray diffraction data was collected at the Advanced Light Source beamline 4.2.2 and processed using HKL2000 (43). The structure of mCEACAM1b[D1,D4] was determined by molecular replacement using mCEACAM1a[D1,D4] (PDB 3R4D) as the search template. The model was built using Coot (44) and refined with Refmac (45) to a final R$_{work}$ and R$_{free}$ of 0.216 and 0.273, respectively.

Pseudovirus entry assay- Lentiviruses pseudotyped with MHV spike protein were produced as previously described (15). Briefly, pcDNA3.1(+) plasmid encoding MHV spike protein (from MHV strain A59) was co-transfected into HEK293T cells with helper plasmid psPAX$_2$ and reporter plasmid pLenti-GFP at molar ratio 1:1:1 using lipofectamine 2000 (Life Technologies). 48 hours post-transfection, the produced pseudovirus particles were harvested and inoculated onto the HEK293T cells expressing mCEACAM1a or mCEACAM1b (wild type or mutant). 48 hours post-infection, cells were observed under fluorescent microscope, the percentage of GFP-expressing cells was calculated using ImageJ (National Institutes of Health). The expression levels of mCEACAM1a and mCEACAM1b in HEK293T cells were measured by Western blotting using antibodies against their C-terminal C9 tag, quantified using ImageJ, and presented as relative expressions in comparison to wild-type mCEACAM1a. The relative expression of each receptor was used to normalize pseudovirus entry efficiency. The experiments were further repeated twice, and similar results were obtained.

Protein-protein interaction assay using AlphaScreen- The interactions between recombinant MHV S1-NTD and recombinant mCEACAM1a or mCEACAM1b (wild type or mutant) were measured using AlphaScreen as previously described (46,47). Briefly, 300 nM MHV S1-NTD with a C-terminal His$_6$ tag were mixed with 30 nM mCEAMCAM1a or mCEACAM1b (wild type or mutant) with a C-terminal human IgG$_4$ Fc tag in OptiPlate-96 (PerkinElmer) for 1 hour at room temperature. AlphaScreen Nickel Chelate Donor Beads and

6

AlphaScreen protein A acceptor beads (PerkinElmer) were added to the mixture at final concentrations of 20 μg/ml. The mixture was incubated at room temperature for 1 hour, protected from light. The assay plates were read in an EnSpire plate reader (PerkinElmer). The experiments were further repeated twice, and similar results were obtained.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions

GP, YY, ZQ, KVH, FL designed experiments and wrote the manuscript; GP, YY, JRP, LX performed the experiments; GP, YY, JRP, LX, ZQ, KVH, and FL analyzed the data.

8

References

1.   Dawkins, R., and Krebs, J. R. (1979) ARMS RACES BETWEEN AND WITHIN SPECIES. *Proceedings of the Royal Society Series B-Biological Sciences* **205**, 489-511

2.   Van Valen, L. (1973) A new evolutionary law. *Evolutionary Theory* **1**, 1-30

3.   Perlman, S., and Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology* **7**, 439-450

4.   Li, W. H., 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. *Journal of Virology* **80**, 4211-4219

5.   Li, F. (2015) Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J Virol* **89**, 1954-1964

6.   Li, F. (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual review of virology* **3**, 237-261

7.   Homberger, F. R. (1997) Enterotropic mouse hepatitis virus. *Laboratory animals* **31**, 97-115

8.   Jacoby, R. O., and Lindsey, J. R. (1997) Health care for research animals is essential and affordable. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **11**, 609-614

9.   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. *Journal of Virology* **65**, 6881-6891

10.   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. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 5533-5536

11.   Godfraind, C., Langreth, S. G., Cardellichio, C. B., Knobler, R., Coutelier, J. P., Duboisdalcq, M., and Holmes, K. V. (1995) TISSUE AND CELLULAR-DISTRIBUTION OF AN ADHESION MOLECULE IN THE CARCINOEMBRYONIC ANTIGEN FAMILY THAT SERVES AS A RECEPTOR FOR MOUSE HEPATITIS-VIRUS. *Laboratory Investigation* **73**, 615-627

12. Beauchemin, N., Draber, P., Dveksler, G., Gold, P., Gray-Owen, S., Grunert, F., Hammarstrom, 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. *Experimental Cell Research* **252**, 243-249

13. Miura, H. S., Nakagaki, K., and Taguchi, F. (2004) N-terminal domain of the murine coronavirus receptor CEACAM1 is responsible for fusogenic activation and conformational changes of the spike protein. *J Virol* **78**, 216-223

14. Walls, A. C., Tortorici, M. A., Bosch, B. J., Frenz, B., Rottier, P. J., DiMaio, F., Rey, F. A., and Veesler, D. (2016) Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114-117

15. Peng, G. Q., Sun, D. W., Rajashankar, K. R., Qian, Z. H., Holmes, K. V., and Li, F. (2011) Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10696-10701

16. Peng, G. Q., Xu, L. Q., Lin, Y. L., Chen, L., Pasquarella, J. R., Holmes, K. V., and Li, F. (2012) Crystal Structure of Bovine Coronavirus Spike Protein Lectin Domain. *Journal of Biological Chemistry* **287**, 41931-41938

17. 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. *Journal of Virology* **72**, 1941-1948

18. Ohtsuka, N., Yamada, Y. K., and Taguchi, F. (1996) Difference in virus-binding activity of two distinct receptor proteins for mouse hepatitis virus. *Journal of General Virology* **77**, 1683-1692

19. 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

20. Boyle, J. F., Weismiller, D. G., and Holmes, K. V. (1987) Genetic resistance to mouse hepatitis virus correlates with absence of virus-binding activity on target tissues. *J Virol* **61**, 185-189

21. Ohtsuka, N., and Taguchi, F. (1997) Mouse susceptibility to mouse hepatitis virus infection is linked to viral receptor genotype. *J Virol* **71**, 8860-8863

22.   Hirai, A., Ohtsuka, N., Ikeda, T., Taniguchi, R., Blau, D., Nakagaki, K., Miura, H. S., Ami, Y., Yamada, Y. K., Itohara, S., Holmes, K. V., and Taguchi, F. (2010) Role of mouse hepatitis virus (MHV) receptor murine CEACAM1 in the resistance of mice to MHV infection: studies of mice with chimeric mCEACAM1a and mCEACAM1b. *J Virol* **84**, 6654-6666

23.   Tan, K. M., Zelus, B. D., Meijers, R., Liu, J. H., Bergelson, J. M., Duke, N., Zhang, R., Joachimiak, A., Holmes, K. V., and Wang, J. H. (2002) Crystal structure of murine sCEACAM1a 1,4 : a coronavirus receptor in the CEA family. *Embo Journal* **21**, 2076-2086

24.   Rao, P. V., Kumari, S., and Gallagher, T. M. (1997) Identification of a contiguous 6-residue determinant in the MHV receptor that controls the level of virion binding to cells. *Virology* **229**, 336-348

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

26.   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

27.   Chen, L., Lin, Y. L., Peng, G., and Li, F. (2012) Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc Natl Acad Sci U S A* **109**, 17966-17971

28.   Chen, Y., Rajashankar, K. R., Yang, Y., Agnihothram, S. S., Liu, C., Lin, Y. L., Baric, R. S., and Li, F. (2013) Crystal structure of the receptor-binding domain from newly emerged Middle East respiratory syndrome coronavirus. *J Virol* **87**, 10777-10783

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

30.   Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K. C., Chen, Y. H., Zhang, L., and Wang, X. (2013) Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* **23**, 986-993

31.   Reguera, J., Santiago, C., Mudgal, G., Ordono, 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**

32. 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. *Journal of Virology* **71**, 3285-3287

33. 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

34. Cavanagh, D., and Davis, P. J. (1986) CORONAVIRUS IBV - REMOVAL OF SPIKE GLYCOPOLYPEPTIDE-S1 BY UREA ABOLISHES INFECTIVITY AND HEMAGGLUTINATION BUT NOT ATTACHMENT TO CELLS. *Journal of General Virology* **67**, 1443-1448

35. Schwegmann-Wessels, C., and Herrler, G. (2006) Sialic acids as receptor determinants for coronaviruses. *Glycoconjugate Journal* **23**, 51-58

36. Kubo, H., Yamada, Y. K., and 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. *Journal of Virology* **68**, 5403-5410

37. Promkuntod, N., van Eijndhoven, R. E., de Vrieze, G., Grone, A., and Verheije, M. H. (2014) Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology* **448**, 26-32

38. Liu, C., Tang, J., Ma, Y., Liang, X., Yang, Y., Peng, G., Qi, Q., Jiang, S., Li, J., Du, L., and Li, F. (2015) Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. *J Virol* **89**, 6121-6125

39. Chen, T., and Gotschlich, E. C. (1996) CGM1a antigen of neutrophils, a receptor of gonococcal opacity proteins. *Proc Natl Acad Sci U S A* **93**, 14851-14856

40. Hill, D. J., Toleman, M. A., Evans, D. J., Villullas, S., Van Alphen, L., and Virji, M. (2001) The variable P5 proteins of typeable and non-typeable Haemophilus influenzae target human CEACAM1. *Molecular microbiology* **39**, 850-862

41. Toleman, M., Aho, E., and Virji, M. (2001) Expression of pathogen-like Opa adhesins in commensal Neisseria: genetic and functional analysis. *Cellular microbiology* **3**, 33-44

42. Voges, M., Bachmann, V., Kammerer, R., Gophna, U., and Hauck, C. R. (2010) CEACAM1 recognition by bacterial pathogens is species-specific. *BMC microbiology* **10**, 117

43.   Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326

44.   Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2126-2132

45.   Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta crystallographica. Section D, Biological crystallography* **55**, 247-255

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

47.   Liu, C., Yang, Y., Chen, L., Lin, Y. L., and Li, F. (2014) A unified mechanism for aminopeptidase N-based tumor cell motility and tumor-homing therapy. *J Biol Chem* **289**, 34520-34529

48.   Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL-W - IMPROVING THE SENSITIVITY OF PROGRESSIVE MULTIPLE SEQUENCE ALIGNMENT THROUGH SEQUENCE WEIGHTING, POSITION-SPECIFIC GAP PENALTIES AND WEIGHT MATRIX CHOICE. *Nucleic Acids Research* **22**, 4673-4680

49.   The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

50.   Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **66**, 486-501

13

Table 1. Data collection and refinement statistics

	mCEAMCAM1b
Data Collection	
Space group	P3$_1$21
Cell dimensions	
a, b, c (Å)	113.097, 113.097, 64.380
α, β, γ (°)	90, 90, 120
Resolution (Å)	16.97-3.10 (3.21- 3.10)[a]
R_{sym}	0.114 (0.507)
$I/\sigma I$	9.15 (1.78)
Completeness (%)	96.3 (86.4)
Redundancy	2.5 (2.1)
Refinement	
Resolution (Å)	16.97-3.10
No. reflections	8557
R_{work}/R_{free}	0.216 / 0.273
No. atoms	1579
Protein	1541
Ligand	28
Water	10
B-factors (Å^2)	40.49
Protein	40.57
Ligand	41.59
Water	23.83
RMSD	
Bond lengths (Å)	0.015
Bond angles (°)	1.46
Ramachandran plot	
Favored (%)	93.4
Outliers (%)	0.51

[a]Values in parentheses are for highest-resolution shell.

FOOTNOTES

FIGURE LEGENDS

FIGURE 1. Overall structure of mouse CEACAM1b. (A) Crystal structure of mCEACAM1b[D1, D4] containing the D1 and D4 domains. Secondary structures in domain D1 of mCEACAM1b[D1, D4] are named following the mCEACAM1a[D1, D4] structure (23). (B) Crystal structure of mCEACAM1a[D1, D4] (PDB 1L67) (23). Virus-binding motifs (VMBs) are in blue. (C) Crystal structure of mCEACAM1a[D1, D4] complexed with MHV S1-NTD (PDB 3R4D) (16). MHV S1-NTD is in cyan, with RBMs in red. (D) Sequence alignment of mCEACAM1b[D1, D4] and mCEACAM1a[D1, D4]. β-strands are shown as arrows. VBMs in mCEACAM1a and the corresponding regions in mCEACAM1b are in blue. Asterisks indicate positions that have fully conserved residues; colons indicate positions that have strongly conserved residues; periods indicate positions that have weakly conserved residues. The boundary between domains D1 and D4 is indicated by a black line. Sequence alignment was done using CLUSTAL W (48). Structural illustrations were done using PyMol (49).

FIGURE 2. Structural comparisons of mCEACAM1a and mCEACA1b. (A) Overlay of mCEACAM1a and mCEACAM1b in domain D1. mCEACAM1a is in black, and mCEACAM1b in yellow. VBMs in mCEACAM1a and the corresponding regions in mCEACAM1b are in blue. (B) Overlay of mCEACAM1a and mCEACAM1b in domain D4. (C) Structure of mCEACAM1b showing distribution of the residues that differ between mCEACAM1a and mCEACAM1b. Regions corresponding to VBMs in mCEACAM1a are in blue. Residues that differ between mCEACAM1a and mCEACAM1b are shown as balls and sticks. (D) Sequence identities and RMSDs between mCEACAM1a and mCEACAM1b in different regions. See Figure 1D for the residue ranges of domains and VBMs. RMSDs were calculated using Coot (50).

FIGURE 3. Structural comparisons of mCEACAM1a and mCEACAM1b in MHV-binding loop CC'. (A) Interactions between MHV S1-NTD (from MHV strain A59) and loop CC' of mCEACAM1a. mCEACAM1 residues are in green, and S1-NTD residues in magenta. Hydrophobic interactions are indicated as arrows, and hydrogen bonds as dotted lines. (B) Structure of loop CC' of mCEACAM1b. (C) Overlay of the CC' loops from mCEACAM1a and mCEACAM1b.

FIGURE 4. Structural comparisons of mCEACAM1a and mCEACAM1b in MHV-binding strands βC' and βC". (A) Interactions between MHV S1-NTD and Arg47 in strand βC' of mCEACAM1a. (B) Interactions between MHV S1-NTD and Met54, Phe56, and Asn59 in strand βC" of mCEACAM1a. (C) Conformation of the side chain of His47 in strand βC' of mCEACAM1b. (D) Conformations of the side chains of Lys54, Thr56, and Pro59 in strand βC" of mCEACAM1b.

FIGURE 5. Structure-guided mutational and functional characterizations of mCEACAM1a and mCEACAM1b. (A) Pseudovirus entry efficiency. Lentiviruses pseudotyped with MHV spike protein (from MHV strain A59) were used to enter HEK293T cells expressing mCEACAM1a or mCEACAM1b (wild type or mutant). The relative expression of each receptor was used to normalize pseudovirus entry efficiency. The pseudovirus entry mediated by wild-type

15

mCEACAM1a was taken as 100%. Error bars indicate standard errors of the means (n = 3). Comparisons between wild-type mCEACAM1a and mutant mCEACAM1a, or between wild-type mCEACAM1b and mutant mCEACAM1b in their capabilities to mediate pseudovirus entry were done using two-tailed t-test (**, $P<0.01$; ***, $P<0.001$). The mutant mCEACAM1a molecules contain single mutations R47H, M54K, F56T, Q59P, or loop CC' from mCEACAM1b (residues 38-43). The mutant mCEACAM1b molecules contain loop CC' from mCEACAM1a (residues 38-43), loop CC' and strand βC' from mCEACAM1a (residues 38-51), or loop CC', strand βC', loop C'C", and strand βC" from mCEACAM1a (residues 38-59). (B) Protein-protein binding assay. The interactions between MHV S1-NTD and mCEACAM1a or mCEACAM1b (wild type or mutant) were measured using AlphaScreen assay. MHV S1-NTD with a C-terminal His_6 tag and mCEAMCAM1a or mCEACAM1b (wild type or mutant) with a C-terminal human IgG_4 Fc tag were attached to AlphaScreen Nickel Chelate Donor Beads and Alpha Screen protein A acceptor beads, respectively. Error bars indicate standard errors of the means (n = 3). Comparisons between wild-type mCEACAM1b and mutant mCEACAM1b in their binding affinity for MHV S1-NTD were done using two-tailed t-test (***, $P<0.001$).

FIGURE 6. Proposed evolutionary arms race between MHV and mice. The race includes both MHV-driven evolution of mice (top) and mouse-adapting evolution of MHV (bottom). See text for detailed discussion.
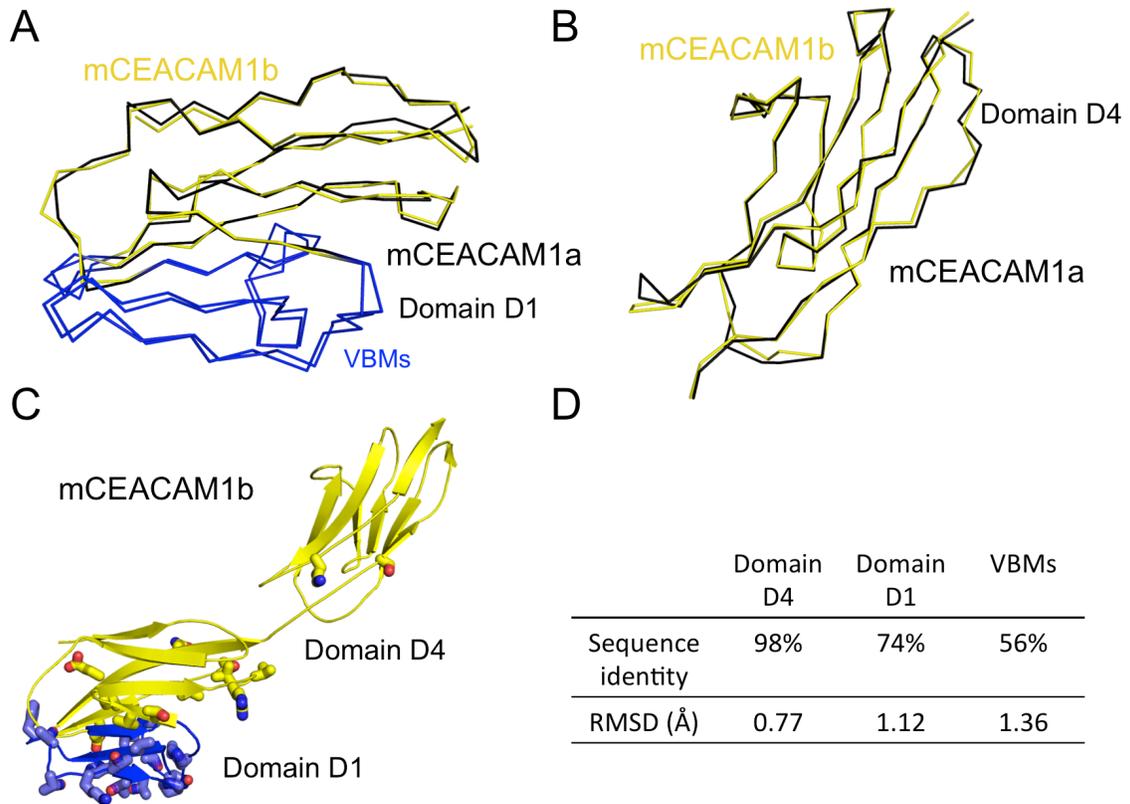
Figure 1:

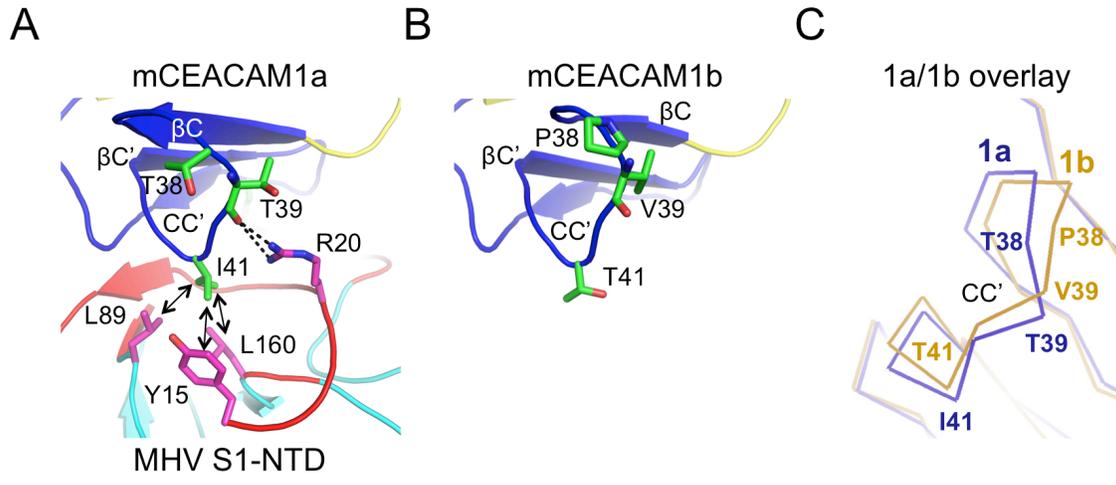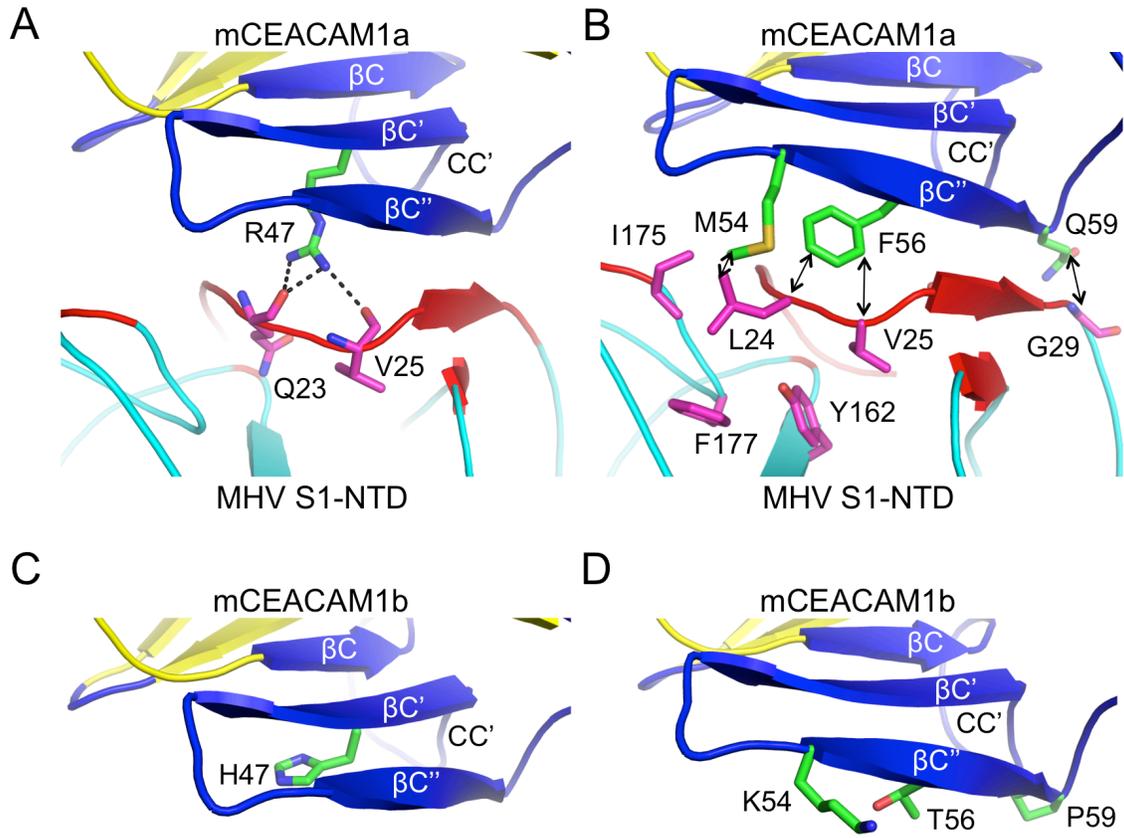**A** mCEACAM1b — Domain D4, C*, N*, βA, βB, βG, βE, βF, βC, βD, βC', βC'', Domain D1

B mCEACAM1a — Domain D4, C*, N*, βA, βB, βG, βE, βF, βD, βC, βC', βC'', Domain D1, VBMs

C mCEACAM1a — Domain D4, Domain D1, VBMs, RBMs, MHV S1-NTD

D

```
      βA          βB          βC          βC'          βC"          βD
1b  EVTIEAVPPQ VAEDNNVLLL VHNLPLALGA FAWYKGNPVS TNAEIVHFVT GTNKTTTGPA HSGRETVYSN 70
1a  EVTIEAVPPQ VAEDNNVLLL VHNLPLALGA FAWYKGNTTA IDKEIARFVP NSNMNFTGQA YSGREIIYSN 70
    ********** ********** ********** *******..:  : **.:**. .:* . ** * :**** :***

      βE          βF          βG
1b  GSLLIQRVTV KDTGVYTIEM TDENFRRTEA TVQFHVHQPV TQPSLQVTNT TVKELDSVTL TCLSNDIGAN 140
1a  GSLLFQMITM KDMGVYTLDM TDENYRRTQA TVRFHVHQPV TQPFLQVTNT TVKELDSVTL TCLSNDIGAN 140
    ****:* :*: ** ****::* ****:***:* **:****** *** ****** ********** **********

1b  IQWLFNSQSL QLTERMTLSQ NNSILRIDPI KREDAGEYQC EISNPVSVKR SNSIKLDIIF DP 202
1a  IQWLFNSQSL QLTERMTLSQ NNSILRIDPI KREDAGEYQC EISNPVSVRR SNSIKLDIIF DP 202
    ********** ********** ********** ********** *********:* ********** **
```

17

Figure 2:

A   mCEACAM1b

mCEACAM1a

Domain D1

VBMs

B   mCEACAM1b

Domain D4

mCEACAM1a

C   mCEACAM1b

Domain D4

Domain D1

D

	Domain D4	Domain D1	VBMs
Sequence identity	98%	74%	56%
RMSD (Å)	0.77	1.12	1.36

Figure 3:

A
mCEACAM1a

βC
βC'
T38
CC'
T39
I41
R20
L89
L160
Y15
MHV S1-NTD

B
mCEACAM1b

βC
P38
βC'
V39
CC'
T41

C
1a/1b overlay

1a
1b
T38
P38
CC'
V39
T41
T39
I41

19

Figure 4:

20

Figure 5:

A

B

21

Figure 6:

Sugar

mCEACAM1a

mCEACAM1b

MHV-driven
evolution
of mice

**Receptor
recognition**

**Receptor
recognition**
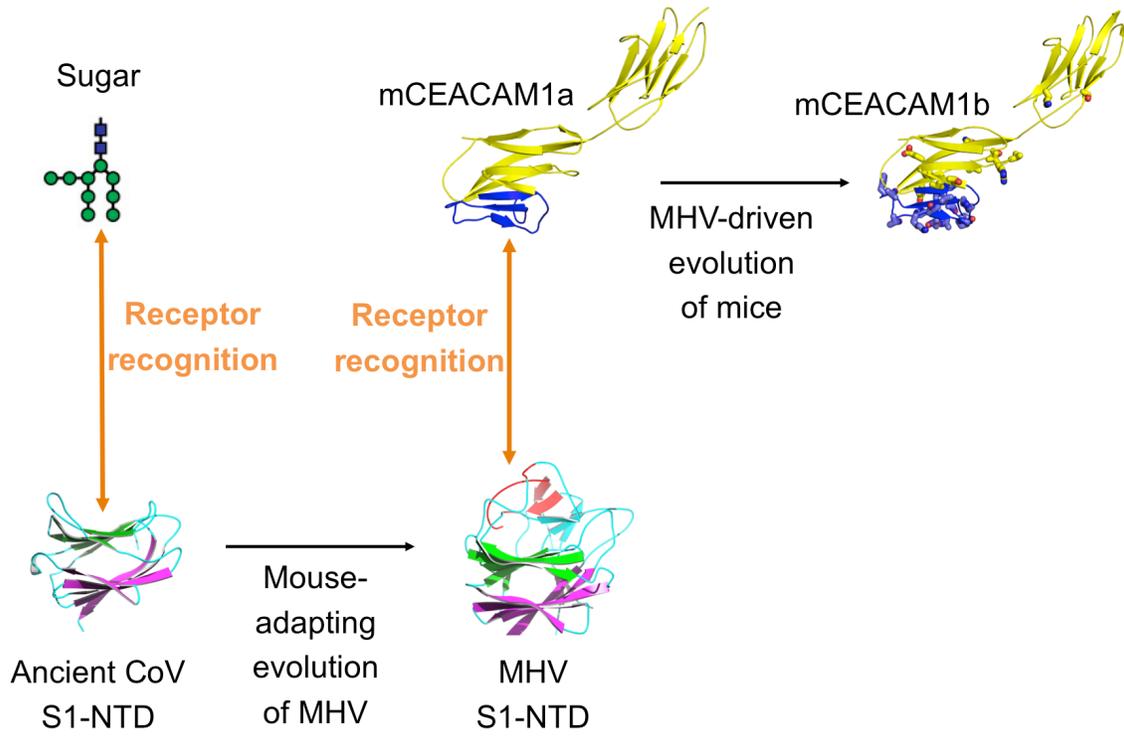
Ancient CoV
S1-NTD

Mouse-
adapting
evolution
of MHV

MHV
S1-NTD

Structural and Molecular Evidence Suggesting Coronavirus-Driven Evolution of Mouse Receptor

Guiqing Peng, Yang Yang, Joseph R. Pasquarella, Liqing Xu, Zhaohui Qian, Kathryn V. Holmes and Fang Li

Access the most updated version of this article at doi: 10.1074/jbc.M116.764266

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

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2016/12/29/jbc.M116.764266.full.html#ref-list-1