Coronavirus S and HE proteins: a two-component system for dynamic virion-sialoglycan interactions

Implications for embecovirus cross-species transmission, host adaptation and host exclusivity

Yifei Lang

Coronavirus S and HE proteins: a two-component system for dynamic virion-sialoglycan interactions

Implications for embecovirus cross-species transmission, host adaptation and host exclusivity

Coronavirus S en HE eiwitten: a twee-componenten systeem voor dynamische virion-sialoglycaan interacties

Implicaties voor cross-species transmissie, gastheertropisme, en gastheer-exclusiviteit van embecovirussen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 28 mei 2020 des ochtends te 10.30 uur

door

Yifei Lang

geboren op 30 juni 1993
te Liaocheng, China

Promotor:
Prof. dr. F.J.M. van Kuppeveld

Copromotor:
Dr. R.J. de Groot

The research described in this thesis was partly financially supported by the China Scholarship Council (CSC) under file number of 201403250042.

致我最爱的爸爸妈妈

To my beloved parents

古之欲明明德于天下者，先治其国。欲治其国者，先齐其家。欲齐其家者，先修其身。欲修其身者，先正其心。欲正其心者，先诚其意。欲诚其意者，先致其知。致知在格物。物格而后知至，知至而后意诚，意诚而后心正，心正而后身修，身修而后家齐，家齐而后国治，国治而后天下平。

----《礼记·大学》节选

Do not go gentle into that good night,
Old age should burn and rave at close of day;
Rage, rage against the dying of the light.

- Dylan Thomas

Table of contents

Chapter 1

General Introduction

General introduction

1. A brief introduction to virology

1.1 Viruses. Viruses are small, obligate intracellular parasites that are associated with disease in animals and plants. Unable to reproduce by themselves, they are fully reliant on the metabolic and biosynthetic processes in living host cells. In their extracellular stage, viruses occur as particles with an acellular organization that are called virions, the most essential component of which is the viral genome. The genome, which may be DNA- or RNA-based, is condensed and wrapped in a protective coat of virus-encoded proteins and, depending on the type of virus, also by a membrane comprised of host-derived lipids, called the envelope. In contrast to cellular organisms, viruses do not amplify through binary fission but through a process called replication. To this end, the viral genome contained in the virion is to be delivered into the host cell. Very much like a computer virus, the viral genome then acts as a hostile code to reprogram the host. The cell's macromolecular biosynthetic machinery is hijacked and tweaked to produce large amounts of viral nucleic acids –genome copies and mRNAs– and viral proteins, the key objective of which is to generate huge quantities of progeny virions and to promote their dissemination.

Viruses differ widely in virion morphology and virion composition, with particle sizes ranging from 10 to 400 nm for standard viruses to up to 1.5 µm for the giant viruses of aquatic amoeba [1,2]. They also display a staggering variation in genome composition, genetic complexity and replication strategies. Although viruses can be distinguished on the basis of any of these criteria, perhaps the most insightful method for classification was proposed by David Baltimore, who recognized that viral replication inevitably requires gene expression and thus mRNA synthesis [3]. By sorting viruses on the basis of how they arrive from their genome to *de novo* mRNA synthesis in the infected cell, they can be divided into 7 major groups (Fig. 1). This thesis will focus mainly on positive-strand RNA viruses (i.e. viruses with a single-stranded RNA genome that serves not only for storage of genetic information but also as an mRNA to direct viral protein synthesis) and on one group in particular, namely coronaviruses, which are pathogens of mammals and birds.

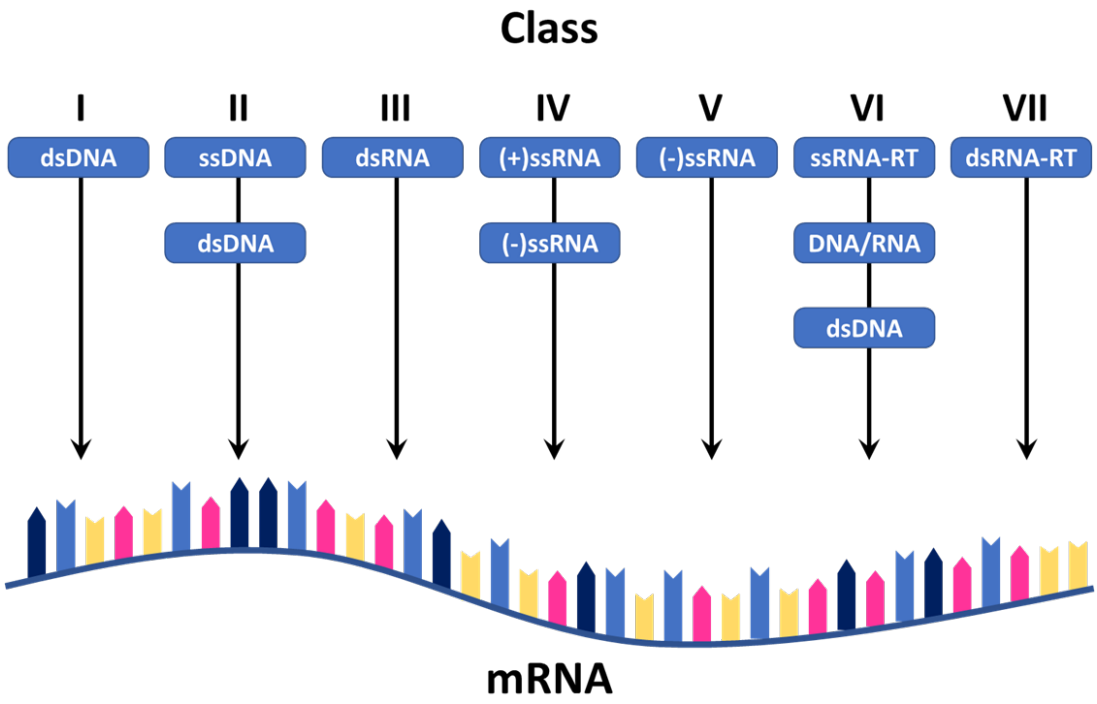


Fig. 1. The Baltimore classification clusters viruses into 7 different groups depending on their type of genome.

1.2 Cellular infection. For all viruses, no matter how different, the interaction with their host cell - from first contact to the escape of viral progeny can be divided into five or six common steps, collectively called the replication cycle, namely: viral attachment (or adsorption); penetration; uncoating; 'replication' (encompassing the biosynthesis of viral nucleic acids and proteins); virion assembly and release. For enveloped viruses that acquire their envelope by 'budding' at the plasma membrane, virion assembly and release occur in one step. In some viruses, virions undergo changes in structure even after their release, a process known as 'maturation'. One step in addition, rarely considered, but of potential importance nonetheless -particularly during natural infection in multicellular hosts- is 'pre-attachment', which comprises the events before a virus binds to its host cell. After all, to start an infection, a virus needs to get to a host cell first. The focus of this thesis is on virus-receptor interactions that occur during (pre-)attachment and less so on the events that follow thereafter.

Virus attachment is mediated by specialized protein(s) on the viral surface. Naked viruses generally bind to their host cells via their capsid proteins, through depressions in the capsid structure or via capsid associated fibers [4,5]. Enveloped viruses, to this end, encode receptor-binding transmembrane glycoproteins that are discernable in electron micrographs as surface projections and therefore commonly referred to as 'spikes'. Their natural binding partners, the 'receptors', are

cell surface glycoproteins and/or glycolipids [6]. The viral attachment proteins possess receptor-binding sites (RBSs) that ensure selective binding through high-specificity biomolecular interactions with 'receptor determinants', i.e. the substructures involved in binding. These may be protein-based, i.e. a specific patch on a particular cellular glycoprotein, or glycan-based with virions attaching to particular protein- and/or lipid-associated sugars [5–7]. The ultimate goal of viral attachment is the delivery of the viral genome into the cell ('entry'). This may occur directly at the plasma membrane, but more often the virion is taken up by endocytosis [5,6,8,9], upon which -triggered by any of several (bio)chemical cues- conformational changes take place in viral proteins and virions either culminating in the formation of a pore through the cellular membrane or disruption of the endosome (in naked viruses) [4,10,11] or in the fusion of viral and cellular membranes (in enveloped viruses) [12–14]. In some cases, viruses employ multiple receptors to gain access to their host cells such that a distinction can be made between 'attachment factors, the binding to which serves to concentrate virions on the cell surface and 'entry receptors', i.e. those molecules that through their interaction with virions allow introduction of the viral genome into the cell [5]. For some viruses, like human immunodeficiency virus HIV), a distinction is made between a primary receptor (CD4) essential for binding to the host cell and co-receptors (CXCR4 or CCR5), binding to which is essential for entry via fusion versus, respectively [15–19]. For Lassa and Ebola viruses, causative agents of hemorrhagic fever, entry is even more complicated and involves an intra-endosomal trigger-induced receptor switch [20–22].

1.3 Infection of multicellular organisms – port d'entrées and barriers. Vertebrate hosts like mammals and birds are well-protected against invading pathogens by physical barriers that constitute the first line of innate defense. The most obvious barrier is the skin, which in humans covers an average surface area of about 1.8 m^2. For a virus to enter the intact host across this barrier would require access via abrasions and wounds. This may occur passively or through a biting incident involving an infected animal acting as vector. A large group of viruses are transmitted via this latter route by hematophagous arthropods such as flies, mosquitoes and ticks. Most viruses, however, use another port of entry, and start infection at the mucosa (the 'wet' epithelia). The mucosal surfaces of the respiratory, gastrointestinal and urogenital tracts and the conjunctiva collectively account for more than 90% of the boundaries between the vertebrate body and the outside world [23,24]. Viruses may get access to the mucosae via the air, ingestion of food and fluids, and sexual intercourse. However, getting to the prospective epithelial cells and the underlying tissues is a different matter. The mucosal epithelia cells, lining the body cavities, are covered at their apical side by a dense forest of glycoproteins, the 'glycocalyx', comprised for a large part of membrane-bound mucins [25,26]. The mucins are glycosylated to the extreme, with in some domains O-linked sugar

chains attached to Ser and Thr at every second to fourth residue. In fact, glycans make up more than 80% of the mucin's molecular weight and the extensive glycosylation causes the protein backbone to be unfolded into a stretched, bottle brush-like structure extending from the cell surface like kelp from the seabed [27,28]. Overlying the epithelia is a viscoelastic hydrogel, known as the mucus layer, comprised of secretory mucins that are produced by goblet cells. The mucus layer and glycocalyx together form the actual barrier. The composition of the mucus layer as well as its thickness and viscosity differ between organs and even between different parts of the same organ system [24,26,29]. Its general purpose is to prevent physical, chemical and microbial damage of the epithelia. Bacteria and viral particles are trapped in the mucus, where they may be 'killed' by secretory antimicrobial products of host cells and microflora or neutralized by antibodies, and disposed of by mucus flow.

In the airways, the mucus prevents particles, like dust and pollen but also bacteria and virions, from entering the lungs. Cilia, i.e. microscopic, motile, hair-like cell surface organelles, beat up the mucus to the back of the throat, where it is swallowed, sending whatever was captured on a one way trip to the hostile environment of the stomach [30]. Pathogens of the gut also have to deal with mucus flow. In the intestinal tract, mucus constantly secreted in large quantities from crypts is moved towards the exit through peristaltic waves [31]. Thus, for a virus to successfully establish a mucosal infection, it must not only traverse the mucus layer and cover a huge distance in relation to its own size, it must also do so fast enough in order not to be swept away.

2. Virus-receptor interactions and factors that determine host and cell tropism (General)

A remarkably large number of mucosal viruses use glycan-based receptors determinant, often with sialic acid as an essential component of the glycotope (Table S1). Sialic acid (Sia) is in fact a collective namefor a diverse family of 9-carbon sugars that commonly occur as terminal residues at the non-reducing end of oligosaccharide chains on glycoproteins and glycolipids [32,33]. The variety among sialosides stems from differences in glycosidic linkage (α2,3; α2,6; α2,8), the composition of the underlying sugar chain, and from differential post-synthetic modifications at Sia carbons C4, C5, C7, C8 and C9 [32,33](Fig. 2). In birds and mammals, the key host species considered in this thesis, C5 may be *N*-acetylated or *N*-glycolylated, whereas *O*-acetylation is the most common modification at C4, C7, C8 and/or C9 [32–34].

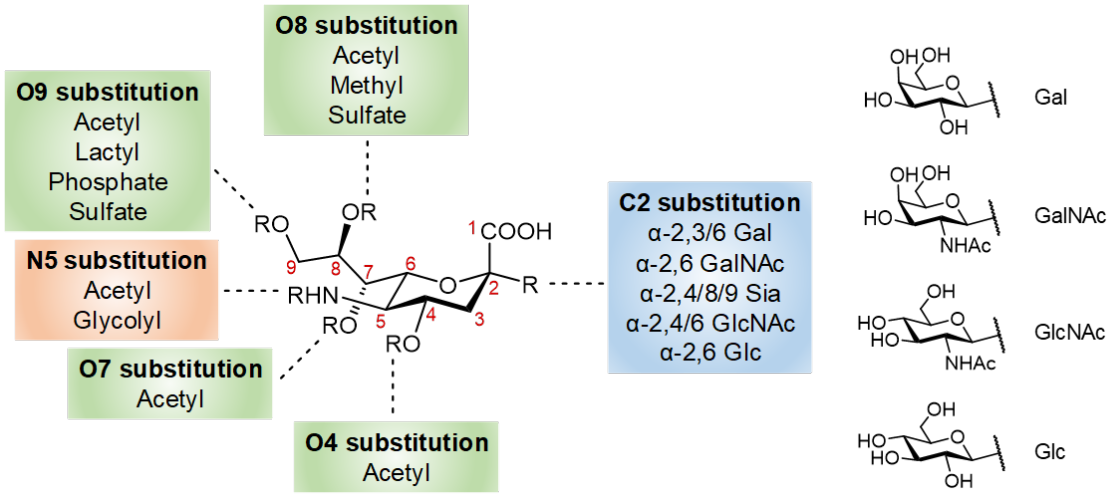


Fig. 2. Schematic overview of sialoside variety. The *N*-acetylneuraminic acid (Neu5Ac) scaffold functions as the chemical basis for different sialic acid modifications [33].

Viruses that use sialosides as receptor determinants do not recognize Sia *per se*, but have evolved to keenly distinguish between these glycotopes with binding promoted or hampered by the presence or absence of Sia modifications, Sia linkage type, and/or glycan composition [35,36]. They tend to be specialized in receptor usage to match the sialoglycome of target organ and host with implications for their disease-causing potential, the breadth of their host and organ tropism, and for their capacity to cross species-barriers [37–40]. In consequence, for these viruses a change in tropism typically requires adaptation to the sialoglycome in the new niche [41–43].

3.Coronaviruses

3.1 Coronaviruses; general charactieristics and taxonomy. The coronaviruses (subfamily *Orthocoronavirinae,* family *Coronaviridae*, suborder *Cornidovirineae*, order *Nidovirales*, realm *Riboviria*) are a diverse group of enveloped positive-strand RNA viruses of mammals and birds. Their virions are spherical, 80-120 nm in diameter, contain a helical nucleocapsid, and characteristically are decorated with petal-shaped spikes extending some 20 nm from the viral envelope [44,45]. Virus particles contain a minimal set of four structural protein species: the spike protein S, the triple-spanning membrane protein M, the small envelope protein E and the nucleocapsid phosphoprotein N (Fig. 3).

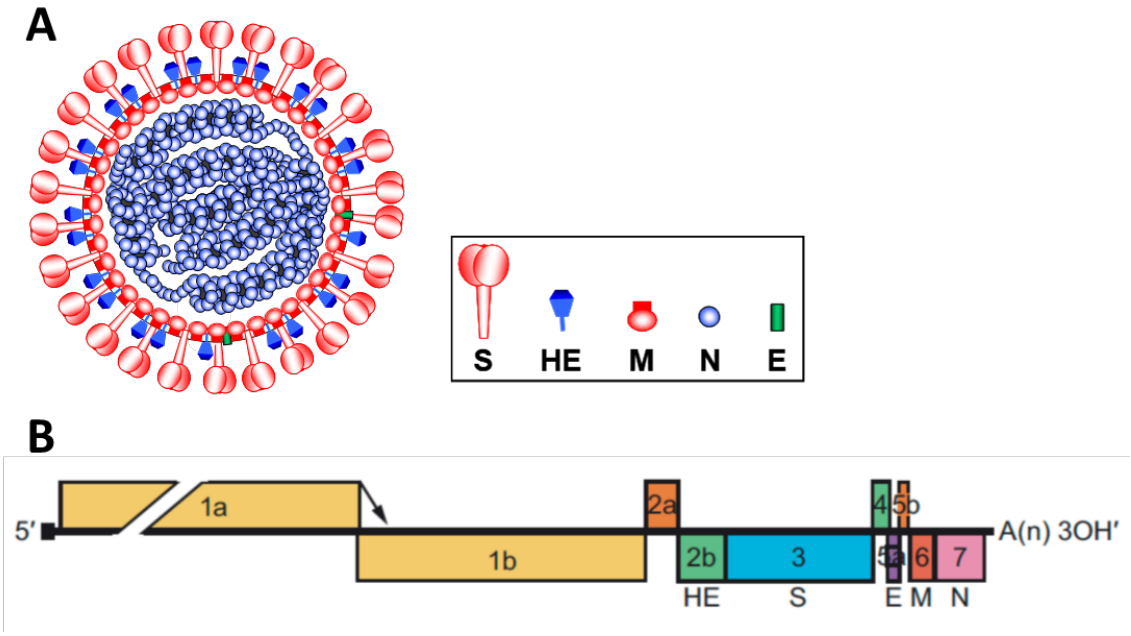


Fig. 3. **(A)** Schematic representation of a coronavirus virion. Particles are comprised of a helical nucleocapsid wrapped by a lipid bilayer, containing the spike protein (S), the hemagglutinin-esterase protein (HE), the membrane protein (M) and the minor envelope protein (E). **(B)** Schematic representation of the coronavirus genome organization. ORFs are represented by boxes; the ribosomal frameshift sites in ORF1 is indicated by an arrow; numbers refer to the mRNA species from which the ORFs are expressed; the poly(A) tail is indicated with A(n). The figure was adopted from [46].

Coronavirus genomes are of exceptional length (up to 32.7 kb) and among the largest RNA genomes known so far. The genome organization (Fig. 3) is conserved with a large replicase polyprotein gene, comprised of two overlapping open reading frames ORF1a and ORF1b and occupying the 5'-most two-thirds of the genome [47]. Processing of the Pol1a polyprotein and read-through polyprotein Pol1ab by viral and cellular proteases yields 15-16 mature products among which the RNA-dependent RNA polymerase, plus several functional intermediates [48]. Many of these nonstructural proteins are implicated in RNA synthesis; others act as antagonists of the antiviral response of the host (cell) [49]. Located downstream of ORF1b, in a strictly conserved order, are the genes for the structural proteins S, E, M and N. In addition, there is a variable set of accessory genes that may be genus-, subgenus, species or even subspecies-specific [44,45,50,51]. They encode proteins that, in general yet with important exceptions (see this thesis), are dispensable for virus replication in cultured cells. However, they contribute to viral fitness during natural infection with many of the accessory proteins acting as antagonists of the host defense [50–52].

Another distinctive feature of coronaviruses is their replication strategy, which entails the synthesis of a 3'-co-terminally nested set of chimeric subgenomic mRNAs, each carrying a 50-70 bp common

5'-leader sequence identical to the 5'-end of the genome [53,54]. These RNAs are produced by 'discontinuous transcription', a process that occurs during the synthesis of subgenomic minus-strand RNAs [44,55]. It involves the fusion of sequences that are non-contiguous on the genome via a mechanism that resembles similarity-assisted RNA recombination [56]. Yet another hallmark, coronaviruses engage in high frequency similarity-assisted 'homologous' RNA recombination also during genome synthesis, which acts to reduce the accumulation of deleterious mutations and contributes to viral diversity [52,57,58]. In addition, coronavirus genomes have been shaped by heterologous recombination, i.e. imprecise recombination events between genome sequences or even with non-coronaviral donor RNAs, which has led to loss and acquisition of new genes [45,59–61]. A very early study already concluded from differences in mRNA profiles between viruses, now recognized as representatives of the alpha-, beta- and gammacoronaviruses, that 'transcription units (one or more genes expressed from a single mRNA species) have been lost, gained or perhaps translocated as the coronaviruses diverged' [51].

At present, thirty-eight coronavirus species are formally recognized, divided over four genera - designated *Alpha-*, *Beta- Gamma-* and *Deltacoronavirus-* and twenty-four subgenera, but surveys into global coronavirus diversity indicate that the actual (sub)species richness vastly exceeds this number by at least one order of magnitude [62]. The paucity in coronavirus species classified so far by the International Committee of Taxonomy of Viruses stems from the fact (i) that only viruses for which a full genome sequence is available are taken into consideration and (ii) that many viruses that may well be considered separate biological entities are yet clustered in a single species on the basis of genetic relatedness [45,62].

The first coronavirus, avian infectious bronchitis virus (IBV; Subgenus: *Igacovirus*; Genus: *Gammacoronavirus*), was isolated from a chicken with "acute, fatal, respiratory disease" in 1931 [63]. The first human coronaviruses, 229E and OC43, were isolated in the mid 1960's from nasal washes from patients with common colds [64–66] and their virions were shown to be morphologically identical to those of IBV and mouse hepatitis virus (MHV; Subgenus: *Embecovirus*, Genus: *Betacoronavirus*) [65,67].

For more than three decades, the field was considered a 'virological backwater' as the human coronaviruses known at the time seemed associated merely with mild respiratory infections and not with life-threatening disease. Animal coronaviruses, however, were recognized to be of considerable veterinary interest, by causing economic damage in farm animals (for example, bovine coronavirus (BCoV); transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV) in swine; IBV in fowl) and high-mortality infections in pets (feline infectious peritonitis virus, FIPV).

Moreover, fundamental aspects of their biology were avidly studied albeit by a very small research community. This picture changed, however, in 2002, with the outbreak of severe acute respiratory syndrome (SARS), a highly lethal viral infection in humans, caused by a coronavirus (SARS CoV) that at time had not been seen ever before. Present evidence suggests that the virus may have arisen from recombination events among existing SARS-related (SARSr) coronaviruses in horseshoe bats (genus *Rhinolophus*) dwelling in caves in southwestern China [68]. Rhinolophus bats in one single cave in Yunnan province were recently found to harbor diverse populations of SARSr CoVs that together would comprise all building blocks of the human SARS-CoV genome, with individual viral genome sequences bearing ample evidence for frequent intratypic recombination [68,69]. SARS CoV eventually entered the human population, reportedly through the wet markets of Shenzhen, with masked palm civets (*Paguma larvata*) serving as accidental intermediate hosts [70]. The infection initially remained confined to mainland China, but after a visit from a SARS-infected urologist from a Guangdong hospital to Hong Kong, at least 15 individuals became infected allowing SARS CoV to spread to five continents and over two dozen countries. The virus was quickly contained within months after its discovery by rigorous quarantine measures, but not after 8,098 individuals became infected, 774 of which lethally, and causing havoc on South-East Asian economies. A mere decade later and in a different part of the globe, another coronavirus of humans emerged out of the blue to cause a lethal respiratory infection disturbingly similar to SARS. Comparative sequence analysis showed the new virus, eventually named Middle East Respiratory Syndrome (MERS) CoV [71], to be distinct from SARS CoV and to belong to different betacoronavirus clade [72]. MERS is a classical zoonotic infection with dromedary camels as natural reservoir host and limited human-to-human transmission, which provides an explanation for its generally restricted geographic occurrence. Yet, MERS CoV remains a pathogen of concern for global public health as illustrated by the 2015 outbreak in a South Korean hospital [73,74].

The SARS outbreak and advent of MERS have changed the field. Although the impact of SARS and MERS-CoV in terms of loss of human life is modest as compared to that of other infectious diseases, their emergence contributed to pandemic awareness and CoVs are now generally recognized as zoonotic threats with pandemic potential. The SARS outbreak led to a surge in coronavirus discovery spearheaded by Chinese group [75–81]. Their work has greatly expanded our insight in coronavirus phylogeny and evolution. While bats are recognized as ‘ecological drivers of CoV diversity’ [62], flightless mammals and birds constitute natural ‘chain’ reservoirs from which novel pathogens of humans and domestic animals may arise as illustrated by the case of MERS CoV [62,82].

Although SARS and MERS CoV captured the imagination and are intensely studied, they actually failed to become established in the human population. Four other coronaviruses, however, have

become true human pathogens [45]. Two of them, called 229E and NL63, are members of the genus *Alphacoronavirus*, the others -designated OC43 and HKU1- are betacoronaviruses. The latter viruses, both belonging to the subgenus *Embecovirus*, feature prominently in this thesis and aspects of their biology will be detailed further below. The impact of the true HCoVs is subject to reevaluation. They are estimated to cause 10% of all upper and lower respiratory tract infections with patients most often presenting common colds [83]. However, they are increasingly associated also with more severe clinical outcomes that require hospitalization in neonates and young children, in adults with underlying disease, in the elderly and, occasionally, even in healthy adults without comorbidities. Among the most severe conditions associated with HCoV infection are pneumonia, acute respiratory distress syndrome and fatal encephalitis [80,84–87].

Like SARS and MERS-CoV, all four HCoVs emerged from animal reservoirs, but in contrast to the former two they did pass the test and breached the species barrier successfully. They achieved a world-wide geographic distribution and are presently maintained through continuous human-to-human transmission. Two of them apparently originated from viruses in domestic animals. HCoV 229E, like MERS CoV, seems to have been transmitted from dromedary camels [88], whereas bovine coronavirus is considered as the ancestor of OC43. The zoonotic event that sparked OC43 emergence occurred very recently. To illustrate, the early OC43 isolate OC43-USA/1967 (ATCC VR759) and BCoV-Mebus, isolated in 1967, share 97% genome identity [89]. The split between BCoV and OC43 was estimated to have occurred at the end of the 19th century [89,90], but the actual introduction may even have been more recent still as the common ancestor of all OC43 strains was dated to the 1950's [91]. As a result of continuous evolution since the 1960's, contemporary OC43 field variants differ as much from OC43-USA/1967 as OC43-USA/1967 differs from BCoV Mebus [91]. The phylogenetic data indicate that the introduction of OC43 was a singular event and that adaptation to the human host has set the virus on a route towards speciation [89–91]. This notion raises important questions. Which factors determine whether or not embecoviruses like OC43 can cross the species barrier? What keeps BCoV from spilling over in humans continuously? Which factors prevent OC43 from spilling back into cattle? What adaptations are required for efficient replication in and transmission from human airways? This thesis looks into these questions albeit with a particular focus of virus-receptor interactions.

Coronavirus reverse genetics. In 1978, Taniguchi and coworkers reported that plasmid DNA, containing a cDNA copy of the RNA genome of coliphage Qβ, directed Qβ production when introduced in Escherichia coli [92]. Three years later, Racaniello and Baltimore were first to show the infectivity of a cloned cDNA copy of a mammalian RNA virus genome -that of poliovirus- by transfecting CV-1 and HeLa cells with PBR322-based plasmids [93]. The final breakthrough came

when Ahlquist and coworkers reported the infectivity of synthetic RNAs, transcribed *in vitro* from cloned brome mosaic virus (BMV) cDNA templates [94]. These findings heralded a revolution in RNA virus research. They provided a roadmap to dissect RNA viruses by reverse genetics and this methodology remains a key technique even to date. Be that as it may, the development of reverse genetics systems for coronaviruses proved difficult for a long time and was hampered by the extreme length of the genome and by the presence of 'toxic' sequences that prohibited cloning by traditional methods. To circumvent these problems, an alternative strategy was devised that made use of the high frequency RNA recombination during coronavirus replication. The first successful attempts to manipulate a CoV genome entailed the repair of a temperature sensitive (ts) mutation in the MHV N gene by transfecting virus-infected cells with a short non-replicating synthetic RNA [95]. Recombination between the MHV genome and the donor transcript resulted in a restorative insertion in N-gene and the concomitant introduction of a five-nucleotide tag in the 3'-nontranslated region. While the results demonstrated that the coronavirus genome is amenable to site directed mutagenesis, recombination efficiency was low and the construction of recombinant viruses depended on the strong counterselection provided by the thermolability of the acceptor virus. An independent study performed in parallel showed that the efficiency of site directed mutagenesis of the MHV genome could be increased significantly by using defective co-replicating RNAs as donor transcript [96], raising the frequency of recombinant offspring by three orders of magnitude from 0.001 to 1% and allowing the isolation of recombinant viruses with mutations in the N-terminus of ORF1a even without selection [97]. Targeted recombination really came of age, however, when Kuo and coworkers [98] developed a method that made use of the fact that most coronaviruses have a narrow host tropism and that host specificity is conferred by the receptor-specificity of the S protein. Proof of principle was provided by the construction of a recombinant MHV in which the S gene was replaced by that of feline coronavirus. The resultant chimeric virus exclusively replicated in feline cells and no longer in murine cell lines. By transfecting cells infected with this "acceptor virus" with co-replicating synthetic donor RNAs and selecting the progeny on murine cells, recombinant viruses could be obtained in which feline S was replaced again by autologous MHV S. The procedure allowed not only mutagenesis of S but also 'piggyback' insertion of mutations in the entire region of the MHV genome downstream of ORF1b. Although successfully applied to study various aspects of the replication and pathogenesis of MHV and other CoVs [99–104], targeted recombination cannot be readily used to introduce mutations in the replicase gene nor to produce defective viruses with lethal mutations. To circumvent these disadvantages, methods were eventually developed to construct full length cDNA clones as bacterial artificial chromosomes [105,106] or by using vaccinia virus as a carrier [107] and, least laborious and most effective and straightforward of all, by *in vitro* ligation of cDNA fragments [108]. Over the years, our laboratory established reverse genetics systems for a

string of CoVs. Some CoVs, for reasons not entirely clear but related at least in part to their *in vitro* propagation characteristics, remained refractory to reverse genetics. Over the years, several members of our group tried to develop a reverse genetics system for bovine coronavirus (BCoV), but the attempts did not meet with success. In **Chapter 5**, we report the production of recombinant BCoVs by targeted recombination. Reverse genetics once again showed its merits as it allowed us to discover exciting new aspects of the biology of BCoV and its relatives.

4. Embecoviruses

4.1 Embecoviruses aka 'lineage A' betacoronaviruses. Until recently, the ICTV recognized only five ranks (order, family, subfamily, genus, species), too few to adequately capture nidovirus diversity and the complex phylogenetic relationships between coronaviruses. As of February 2019, the number of formal ranks has been increased to seven, with consequences for CoV nomenclature [ICTV online]. The genus *Betacoronavirus* comprises four distinct phyloclusters that previously were referred to as lineage A through D and that now have been assigned the rank of subgenera called *Embecovirus*, *Sarbecovirus* (represented by SARS CoV), *Nobecovirus*, *Merbecovirus* (represented by MERS CoV, respectively. A fifth group, recently identified and represented so far by a single species *Bat Hp-betacorovirus*, has been included as an additional subgenus *Hibecovirus*. Note that in **Chapters 2** to **4**, we still use the term "lineage A betacoronavirus" to switch to the new taxonomy and nomenclature in **Chapter 5**.

The embecoviruses form only a minor monophyletic clade within the larger orthocoronavirus family with just four species recognized so far: *Betacoronavirus 1*, *Human coronavirus* HKU1, *Murine coronavirus 1* (MuCoV; represented by mouse hepatitis virus, MHV) and a newly discovered species *China Rattus coronavirus HKU24*. Nevertheless, representatives of this group have been of considerable importance to the field. Before the advent of SARS and MERS CoV, the molecular biology of bovine coronavirus (BCoV; a host range variant of *Betacoronavirus 1*) was studied vigorously and MHV in fact served as the leading CoV model [109]. Many fundamental features of coronavirus biology were discovered and analyzed in depth for MHV first [60,110]. Last but not least, two of the four true human coronaviruses known are embecoviruses, namely OC43 and HKU1.

4.2 Embecoviruses – not one but two types of surface projections. The embecoviruses differ from all other CoVs in gene content and virion composition. In addition to the 20-nm peplomers that are a coronavirus hallmark [44,45], embecovirus particles are typically decorated with an additional fringe of smaller 'granular' surface projections extending 5-8 nm from the viral envelope [111–115]. Analysis of virion proteins revealed a 65kD homodimeric glycoprotein initially called E3 (with S called

E2) that at the time had not been described for any other coronavirus [116]. E3, now known as the hemagglutinin-esterase (HE), the S protein, and their role during embecovirus infection is the main focus of this thesis as will be detailed further below.

Embecovirus S proteins. Coronavirus S proteins are heavily glycosylated class I membrane proteins that assemble into petal-shaped homotrimeric complexes ('peplomers'). In most cases, the trimers are processed by furin-like enzymes as a late step in S biosynthesis, cleaving the monomers into N-terminal subunits (S1) that form a the multidomain membrane distal globular part of the peplomers involved in receptor-binding and C-terminal subunits (S2) that form a membrane-anchored stalk comprising the fusion machinery [or a recent review on CoV S protein structure and function, see [117]]. The embecovirus S proteins differ in size from 1321 (MHV) to 1363aa (BCoV). For members of the species Betacoronavirus-1 (β1CoVs) and HKU1, cell surface sialoglycans are essential receptor determinants. Their S proteins specifically bind to 9-*O*-acetylated sialic acids (9-*O*-Ac-Sia) in a sialate-9-*O*-acetyl-dependent fashion [118–120]. Whether binding is to 9-*O*-Ac-Sia exclusively or whether there may be other proteinaceous receptors required for entry is under debate. Although the existence of an addition protein-based receptor cannot be excluded, there is as yet no convincing evidence in support nor would there be a mechanistic necessity for protein-protein interaction for virus entry. For example, Orthomyxoviruses, enveloped negative-strand RNA viruses and causative agents of flu, bind to sialosides only. Remarkably, however, not all embecoviruses use sialosides for S-mediated receptor-binding. The S proteins of members of the species *Murine Coronavirus 1*, including MHV and rat coronavirus, do not detectably bind to *O*-acetylated Sia [121,122] but bind to carcinoembryonic antigen-related cell adhesion molecule 1 instead [123,124].

Embecoviruses – unique features.

Embecovirus genomes contain two to four accessory genes, none of which found in any other type of CoV. Downstream of the MHV S gene there are two accessory genes, designated ORFs 4 and 5. ORF5 is conserved throughout [125,126], but in β1CoVs, ORF4 seems to have been inactivated and split into two very short ORFs, 4a and 4b [126,127], while in HKU1, ORF4 is missing altogether [80]. The ORF5 product, designated ns12.9, is a viroporin and apparently involved in virion morphogenesis [128]. Two larger accessory genes, both located between ORF1b and the S gene, encode proteins that have been studied in considerable detail and that play pivotal roles during natural infection. Saliently, both genes appear to have been acquired from non-related viruses through heterologous RNA recombination (see below).

4.3. NS2a – an antagonist of the OAS/RNAse L pathway. ORF2, immediately downstream of ORF1b, codes for a 30kD protein that is a member of the 2-His phosphoesterase family with 2',5'-

phosphodiesterase (2',5'-PDE) activity. By degrading 2'-5' oligoadenylate (2-5A), a second messenger produced upon dsRNA sensing by cellular oligoadenylate synthetases (OAS), it prevents RNAseL activation and thus functions as a powerful inhibitor of the OAS/RNAseL pathway. This is a cellular antiviral defense mechanism that upon activation directs the degradation of cellular and viral RNAs, with the killing of infected cells as ultimate consequence. In MHV, NS2A expression is dispensable for replication in cultured cells, but the replication of recombinant viruses, deficient for NS2A, is strongly reduced in infected mice, by 10,000-fold as compared to wildtype virus [129]. Nevertheless, NS2A is not universally conserved among embecoviruses. The gene is inactivated and fragmented in rabbit coronavirus HKU14 [79] and absent in HKU1 [80], which, according to the phylogenetic evidence, seems to have lost NS2A [79]. NS2A is unique to embecoviruses in sequence, but not in function. Nonrelated OAS/RNAseL antagonists have also been identified in merbecoviruses [130]. NS2A, however, shares sequence similarity with a domain at the very C terminus of torovirus pp1a [131], and with rotavirus A VP3 [132], raising the possibility that a proto-embecovirus acquired ORF2 through heterologous recombination. For the other embecovirus accessory protein, the hemagglutinin-esterase (HE), the evidence for a recombinant origin is even more compelling.

4.4. The hemagglutinin-esterase protein (HE). Luytjes et al. [60] while sequencing the genome of the MHV laboratory strain A59 discovered a gene, downstream of ORF2, that would have coded for a protein with 30% sequence identity to N-terminal hemagglutinin-esterase protein subunit 1 (HEF1) of influenza C virus (ICV) -a negative strand RNA virus in the family Orthomyxoviridae- were it not for a mutation in the transcription regulating sequence and a nonsense mutation. It was quickly established that this gene coded for the "E3" protein that had been observed in several related viruses, including bovine coronavirus and OC43 [133–136]. To explain the sequence similarity between MHV HE and ICV HEF, it was proposed that an MHV ancestor had acquired the gene from ICV by heterologous RNA recombination. ICV HEF is a multifunctional protein: it is a receptor-binding protein and mediates virion attachment to 9-*O*-acetylated sialic acid and it is a receptor-destroying enzyme with sialate-*O*-acetylesterase (SOAE) activity [137,138]. These functions reside in subunit HEF1 [139]. Furthermore, HEF mediates membrane fusion, with the fusion machinery largely comprised of C-terminal subunit HEF2 [139]. Accordingly, BCoV HE was shown to be a hemagglutinin and a SOAE, but not a fusion protein [133,140,141].

Remarkably, although among CoVs HE is unique to embecoviruses, an inactivated HE gene was discovered also in equine torovirus (EqToV Berne, a cell culture-adapted representative of a distinct group of in nidoviruses distantly related to CoVs (Order *Nidovirales*, Family: *Tobaniviridae*, Subfamily: *Torovirinae*) [142]. These observations were the start of a line of research at the Virology division that continues to date and of which this thesis describes the latest findings.

After the discovery of HE pseudogenes in the genomes of MHV-A59 and EqToV Berne, it was established that in field strains of porcine and bovine toroviruses, the HE gene is intact and functional, and, codes for an envelope protein additional to the torovirus equivalent of the S protein [143]. Comparative sequence analysis of MHV isolates other than A59 revealed that with few exceptions the HE gene had been rendered non-functional albeit through a variety of strain-specific mutations. Moreover, recombinant MHV-A59 derivatives in which the HE gene had been restored through reverse genetics rapidly lost HE expression upon serial passage *in vitro* [99]. Apparently, for MHV the expression of the HE protein is not merely dispensable, but in fact disadvantageous at least during propagation in cultured cells. Also, HE is not essential for infection of mice [144,145]. However, in MHV field variants, HE expression is conserved, indicating that HE does contribute to viral fitness during natural infection [99,134,146], apparently by promoting viral spread [147]. Remarkably, the role of HE and its relevance for infection differs among embecoviruses. In β1CoVs, HE apparently is an essential protein. Recombinant OC43 viruses, lacking HE sialate-*O*-acetylesterase activity, are severely hampered in their *in vitro* propagation. This defect can be restored by complementation in cells that stably or transiently express HE, showing that HE esterase activity is important for the production of infectious virus [148]. Moreover, the propagation of OC43-ancestor BCoV is abrogated by esterase-inhibitors [133] and infection is blocked by HE-specific monoclonal antibodies [149–151].

Initially, it was assumed that all nidovirus HEs resembled ICV HEF in substrate specificity and targeted 9-*O*-Ac-Sias. However, seminal work by Vlasak and coworkers [152,153] identified MHV HEs with sialate-4-*O*-acetylesterase activity. Following up on these observations, the Utrecht group showed that MHVs actually occur in two types based on HE esterase substrate preference either for 4- or 9-*O*-Ac-Sia [154]. A comprehensive biochemical comparison of corona- and torovirus HEs found most to be specific for 9-*O*-Ac-Sia, but even among these enzymes subtle yet significant differences in specificity were noted. For example, whereas porcine torovirus HEs cleave 9-mono-*O*-Ac-Sia exclusively, the HEs of toro-and coronaviruses of cattle display a preference for 7,9-di-*O*-Ac-Sias [36]. These observations provided a first indication that HE fine specificity might be related to virus host tropism and that cross-species transmission and viral adaptation to novel hosts might require adaptations in HE to match the sialoglycome in the target tissues of these hosts.

To better understand the molecular basis of HE substrate specificity and the general role of HE during infection, our group previously engaged in a series of experiments that combined biochemical, glycobiological and structure analysis and that relied heavily on the expertise of our collaborative partners. Crystal apo- and holostructures of HEs of BCoV [155] and MHVs [156,157] and those of bovine and porcine ToVs [158] have yielded a wealth of information. Like ICV HEF1 [139],

the corona- and toroviral HE monomers are comprised of a highly conserved SOAE domain with an SGNH-hydrolase fold and, appended to it, a receptor-binding lectin domain with an RBS comprised of variable loops grafted on a conserved 8-stranded "Swiss roll" core [155,158]. Importantly, but as would be expected, in each case the ligand (fine-)specificity of the lectin domain closely matches the substrate preference of the esterase [36,157].

Comparative analyses revealed a strong structural resemblance of the nidovirus HEs also to the receptor-binding domain of influenza A and B hemagglutinins (HAs) even though sequence identity has dwindled to a mere 7% [155]. Remarkably, the orthomyxoviral fusion proteins HEF and HA differ in substrate specificity and share a sequence identity of only 11%, yet their RBSs are similar in binding site topology and receptor orientation. Conversely, those in the corona- and torovirus HEs differ in these aspects from HEF and from each other [155,158] even though all three types of proteins bind similar sialosides and share 30% aa identity. The findings attest to the plasticity of the HE RBS and demonstrate that bindings sites that will bind *O*-Ac-Sias in a sialate-*O*-Ac-specific fashion are not necessarily strictly conserved in architecture and composition. However, The RBSs in HE lectin domains as well as the esterase substrate binding sites do conform to a general design [155–158]. Ligand/substrate specificity is based largely on shape complementarity and hydrophobic interactions, with binding essentially relying on the docking of the critical sialate-*O*-acetyl moiety into a deep hydrophobic pocket P1 and with the sialate-5-*N*-acyl group docking in an adjacent pocket/depression P2. P1 and P2 are typically separated by a perpendicularly oriented aromatic side chain. To increase RBS affinity and specificity, the binding is supported by additional protein-sugar interactions characteristically involving hallmark Sia functions like the sialate-5-*N*-acyl group, the glycerol side chains and, in many cases, the carboxylate. This basic design allowed the shift in HE ligand/substrate specificity from 9- to 4-*O*-Ac-Sia that occurred in murine coronaviruses [156,157]. Rather than by involving radical changes in protein architecture, it was brought about both in the lectin as well as in the esterase domain by altering ligand and substrate topology in the context of a largely conserved carbohydrate-binding site, facilitated by the stereochemical similarity between 4- and 9-*O*-Ac-Sia [157].

A major difference between the orthomyxo- and nidovirus HEs is that HEF, like HA assembles into trimers, whereas the CoV and ToV HEs are homo-dimeric. Although the nidovirus HEs do not mediate fusion, they contain elements structurally identical to key subdomain 'F2' of the fusion machinery as present in orthomyxoviral HEFs and HAs [139,155,158]. Apparently, the CoV and ToV HEs arose from HEF-like fusion proteins and have gone from a homo-trimeric to a dimeric organization, which would have necessitated the creation of new inter-monomer contacts. ToVs and CoVs are separated by an evolutionary distance and among CoVs, HE are found in embecoviruses exclusively. It can thus be

concluded that ToVs and CoVs must have acquired their HE genes through separate events, but it is an open question whether ToVs and CoVs obtained their HEs independently with an orthomyxovirus acting as donor in either case, or whether one of the two passed the protein on to the other after the shift from trimer-to-dimer. Although CoV and ToV HE dimers look highly similar at a superficial glance, tell-tale differences in monomer-monomer contacts, specifically in those contributed by the lectin domains, suggest that the transition from trimer-to-dimer may have happened independently as well [143].

Focus on Embecoviruses – Aim and outline of this thesis.

Cross-species transmission of coronaviruses and the emergence of new pathogens in humans and domestic animals are of concern to public health, animal health and food production. Thus, there is an evident need to understand the conditions that facilitate or hamper coronaviruses species jumping. Susceptibility, i.e. receptor compatibility, is generally considered a prerequisite if not the key factor. So far, studies on CoV cross species transmission primarily focused on proteinaceous receptors [121,122,159–169]. However, two of four established human CoVs, OC43 and HKU1, use 9-*O*-Ac-Sia for attachment and both emerged from the same minor clade of embecoviruses. Viral usage of glycan-based receptors comes with its own set of rules and complexities [170–172]. The objectives of this thesis were two-fold: to increase our general understanding of virus-glycan interactions with coronaviruses as a model system and to determine whether zoonotic introduction of embecoviruses and their adaptation to the human respiratory tract entailed or required alterations in their usage of glycan-based receptors that might explain their host specificity. We focused on bovine coronavirus and human coronavirus OC43, an established zoonotic pair [89–91] and hence an ideal model for our studies. In **Chapter 2,** we provide evidence that the HEs of human coronaviruses OC43 and HKU1 have lost their lectin function and thus have been reduced to receptor-destroying enzymes exclusively, apparently as an adaptation to replication in the human respiratory tract. In **Chapter 3**, we shifted our attention to the S protein and corrected existing literature [173] by identifying the true location of the β1CoV S RBS, using a comparative structural analysis approach in combination with automated ligand docking and mutational analysis. In **Chapter 4**, we joint efforts with the group of David Veesler (Department of Biochemistry, University of Washington, Seattle, WA, USA) to confirm our findings and structurally identify the RBS of HCoV-OC43 by single particle cryo-electron microscopy. In addition, we analyzed the kinetic binding characteristics of the S RBS by biolayer interferometry. In **Chapter 5**, we used reverse genetics and forced evolution/population dynamics analysis to study the consequences of loss of HE lectin function, using OC43-ancestor BCoV as a

model. In **Chapter 6**, the results presented in this thesis are summarized and discussed also in the context of observations that are as yet unpublished.

References

1. Abergel, C.; Legendre, M.; Claverie, J.M. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiol. Rev.* **2015**, *39*, 779–796.
2. Lefkowitz, E.J.; Dempsey, D.M.; Hendrickson, R.C.; Orton, R.J.; Siddell, S.G.; Smith, D.B. Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* **2018**, *46*, D708–D717.
3. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **1971**, *35*, 235–41.
4. Yamauchi, Y.; Helenius, A. Virus entry at a glance. *J. Cell Sci.* **2013**, *126*, 1289–95.
5. Marsh, M.; Helenius, A. Virus entry: open sesame. *Cell* **2006**, *124*, 729–40.
6. Grove, J.; Marsh, M. The cell biology of receptor-mediated virus entry. *J. Cell Biol.* 2011, *195*, 1071–1082.
7. Bartlett, A.H.; Park, P.W. Proteoglycans in host-pathogen interactions: molecular mechanisms and therapeutic implications. *Expert Rev. Mol. Med.* **2010**, *12*, e5.
8. Marsh, M.; Helenius, A. Virus entry into animal cells. *Adv. Virus Res.* **1989**, *36*, 107–51.
9. Mercer, J.; Schelhaas, M.; Helenius, A. Virus Entry by Endocytosis. *Annu. Rev. Biochem.* **2010**, *79*, 803–833.
10. Prchla, E.; Plank, C.; Wagner, E.; Blaas, D.; Fuchs, R. Virus-mediated release of endosomal content in vitro: Different behavior of adenovirus and rhinovirus serotype 2. *J. Cell Biol.* **1995**, *131*, 111–123.
11. Schober, D.; Kronenberger, P.; Prchla, E.; Blaas, D.; Fuchs, R. Major and minor receptor group human rhinoviruses penetrate from endosomes by different mechanisms. *J. Virol.* **1998**, *72*, 1354–64.
12. Lee, K.K. Architecture of a nascent viral fusion pore. *EMBO J.* **2010**, *29*, 1299–311.
13. Harrison, S.C. Viral membrane fusion. *Nat. Struct. Mol. Biol.* **2008**, *15*, 690–8.
14. Staring, J.; Raaben, M.; Brummelkamp, T.R. Viral escape from endosomes and host detection at a glance. *J. Cell Sci.* **2018**, *131*.
15. McDougal, J.S.; Kennedy, M.S.; Sligh, J.M.; Cort, S.P.; Mawle, A.; Nicholson, J.K. Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science* **1986**, *231*, 382–5.
16. Maddon, P.J.; Dalgleish, A.G.; McDougal, J.S.; Clapham, P.R.; Weiss, R.A.; Axel, R. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **1986**, *47*, 333–48.
17. Orloff, G.M.; Orloff, S.L.; Kennedy, M.S.; Maddon, P.J.; McDougal, J.S. Penetration of CD4 T cells by HIV-1. The CD4 receptor does not internalize with HIV, and CD4-related signal transduction events are not required for entry. *J. Immunol.* **1991**, *146*, 2578–87.
18. Berger, E.A.; Doms, R.W.; Fenyö, E.M.; Korber, B.T.; Littman, D.R.; Moore, J.P.; Sattentau, Q.J.; Schuitemaker, H.; Sodroski, J.; Weiss, R.A. A new classification for HIV-1. *Nature* **1998**, *391*, 240.
19. Wilen, C.B.; Tilton, J.C.; Doms, R.W. HIV: Cell binding and entry. *Cold Spring Harb. Perspect. Med.* **2012**, *2*.
20. Jae, L.T.; Raaben, M.; Herbert, A.S.; Kuehne, A.I.; Wirchnianski, A.S.; Soh, T.K.; Stubbs, S.H.; Janssen, H.; Damme, M.; Saftig, P.; et al. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* **2014**, *344*, 1506–10.
21. Moller-Tank, S.; Maury, W. Ebola Virus Entry: A Curious and Complex Series of Events. *PLoS Pathog.* **2015**, *11*.
22. Jae, L.T.; Brummelkamp, T.R. Emerging intracellular receptors for hemorrhagic fever viruses. *Trends Microbiol.* **2015**, *23*, 392–400.
23. Lamm, M.E. Interaction of antigens and antibodies at mucosal surfaces. *Annu. Rev. Microbiol.* **1997**, *51*, 311–40.
24. Helander, H.F.; Fändriks, L. Surface area of the digestive tract-revisited. *Scand. J. Gastroenterol.* **2014**, *49*, 681–689.
25. Linden, S.K.; Sutton, P.; Karlsson, N.G.; Korolik, V.; McGuckin, M.A. Mucins in the mucosal barrier to infection.

Mucosal Immunol. **2008**, *1*, 183–197.

26. Voynow, J.A.; Rubin, B.K. Mucins, mucus, and sputum. *Chest* **2009**, *135*, 505–512.
27. Ferez-Vilar, J.; Hill, R.L. The structure and assembly of secreted mucins. *J. Biol. Chem.* **1999**, *274*, 31751–31754.
28. van Putten, J.P.M.; Strijbis, K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. *J. Innate Immun.* **2017**, *9*, 281–299.
29. Leal, J.; Smyth, H.D.C.; Ghosh, D. Physicochemical properties of mucus and their impact on transmucosal drug delivery. *Int. J. Pharm.* **2017**, *532*, 555–572.
30. Zanin, M.; Baviskar, P.; Webster, R.; Webby, R. The Interaction between Respiratory Pathogens and Mucus. *Cell Host Microbe* **2016**, *19*, 159–68.
31. Johansson, M.E. V; Sjövall, H.; Hansson, G.C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10*, 352–61.
32. Schauer, R.; Kamerling, J.P. Exploration of the Sialic Acid World. In *Advances in Carbohydrate Chemistry and Biochemistry*; Academic Press, **2018**; Vol. 75, pp. 1–213 ISBN 9780128152027.
33. Cohen, M.; Varki, A. The sialome--far more than the sum of its parts. *OMICS* **2010**, *14*, 455–64.
34. Klein, A.; Roussel, P. O-acetylation of sialic acids. *Biochimie* **1998**, *80*, 49–57.
35. Neu, U.; Bauer, J.; Stehle, T. Viruses and sialic acids: Rules of engagement. *Curr. Opin. Struct. Biol.* **2011**, *21*, 610–618.
36. Langereis, M.A.; Bakkers, M.J.G.; Deng, L.; Padler-Karavani, V.; Vervoort, S.J.; Hulswit, R.J.G.; van Vliet, A.L.W.; Gerwig, G.J.; de Poot, S.A.H.; Boot, W.; et al. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* **2015**, *11*, 1966–1978.
37. Matrosovich, M.; Herrler, G.; Klenk, H.D. Sialic Acid Receptors of Viruses. *Top. Curr. Chem.* **2015**, *367*, 1–28.
38. Long, J.S.; Mistry, B.; Haslam, S.M.; Barclay, W.S. Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* **2019**, *17*, 67–81.
39. Herfst, S.; Imai, M.; Kawaoka, Y.; Fouchier, R.A.M. Avian influenza virus transmission to mammals. *Curr. Top. Microbiol. Immunol.* **2014**, *385*, 137–55.
40. de Graaf, M.; Fouchier, R.A.M. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J.* **2014**, *33*, 823–41.
41. de Graaf, M.; Fouchier, R.A.M. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J.* **2014**, *33*, 823–41.
42. Stencel-Baerenwald, J.E.; Reiss, K.; Reiter, D.M.; Stehle, T.; Dermody, T.S. The sweet spot: defining virus-sialic acid interactions. *Nat. Rev. Microbiol.* **2014**, *12*, 739–49.
43. Wasik, B.R.; Barnard, K.N.; Parrish, C.R. Effects of Sialic Acid Modifications on Virus Binding and Infection. *Trends Microbiol.* **2016**, *24*, 991–1001.
44. Masters, P.S. The Molecular Biology of Coronaviruses. *Adv. Virus Res.* **2006**, *65*, 193–292.
45. de Groot, R.; Baker, S.; Baric, R.; … L.E.-V.; 2012, U. Family coronaviridae. *Elsevier Amsterdam* **2011**, 806–828.
46. Groot, R. de Family Coronaviridae. *Virus Taxon. Ninth Rep. Int. Comm. Taxon. Viruses* **2011**, 806–828.
47. Gorbalenya, A.E.; Enjuanes, L.; Ziebuhr, J.; Snijder, E.J. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **2006**, *117*, 17–37.
48. Snijder, E.J.; Decroly, E.; Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. In *Advances in Virus Research*; Academic Press Inc., **2016**; Vol. 96, pp. 59–126.
49. Rabouw, H.H.; Langereis, M.A.; Knaap, R.C.M.; Dalebout, T.J.; Canton, J.; Sola, I.; Enjuanes, L.; Bredenbeek, P.J.; Kikkert, M.; de Groot, R.J.; et al. Middle East Respiratory Coronavirus Accessory Protein 4a Inhibits PKR-Mediated Antiviral Stress Responses. *PLoS Pathog.* **2016**, *12*.
50. Liu, D.X.; Fung, T.S.; Chong, K.K.L.; Shukla, A.; Hilgenfeld, R. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res.* **2014**, *109*, 97–109.
51. de Groot, R.J.; ter Haar, R.J.; Horzinek, M.C.; van der Zeijst, B.A. Intracellular RNAs of the feline infectious peritonitis coronavirus strain 79-1146. *J. Gen. Virol.* **1987**, *68 (Pt 4)*, 995–1002.
52. Herrewegh, A.A.; Smeenk, I.; Horzinek, M.C.; Rottier, P.J.; de Groot, R.J. Feline coronavirus type II strains 79-1683

and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *J. Virol.* **1998**, *72*, 4508–14.

53. Spaan, W.; Delius, H.; Skinner, M.; Armstrong, J.; Rottier, P.; Smeekens, S.; van der Zeijst, B.A.; Siddell, S.G. Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.* **1983**, *2*, 1839–44.

54. Lai, M.M.; Baric, R.S.; Brayton, P.R.; Stohlman, S.A. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 3626–30.

55. Sawicki, S.G.; Sawicki, D.L. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* **1990**, *64*, 1050–6.

56. Pasternak, A.O.; Spaan, W.J.M.; Snijder, E.J. Nidovirus transcription: how to make sense...? *J. Gen. Virol.* **2006**, *87*, 1403–21.

57. Keck, J.G.; Makino, S.; Soe, L.H.; Fleming, J.O.; Stohlman, S.A.; Lai, M.M. RNA recombination of coronavirus. *Adv. Exp. Med. Biol.* **1987**, *218*, 99–107.

58. Lai, M.M.; Baric, R.S.; Makino, S.; Keck, J.G.; Egbert, J.; Leibowitz, J.L.; Stohlman, S.A. Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* **1985**, *56*, 449–56.

59. Brian, D.A.; Spaan, W.J.M. Recombination and coronavirus defective interfering RNAs. *Semin. Virol.* **1997**, *8*, 101–111.

60. Luytjes, W.; Bredenbeek, P.J.; Noten, A.F.H.; Horzinek, M.C.; Spaan, W.J.M. Sequence of mouse hepatitis virus A59 mRNA 2: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **1988**, *166*, 415–422.

61. Lorusso, A.; Decaro, N.; Schellen, P.; Rottier, P.J.M.; Buonavoglia, C.; Haijema, B.-J.; de Groot, R.J. Gain, preservation, and loss of a group 1a coronavirus accessory glycoprotein. *J. Virol.* **2008**, *82*, 10312–7.

62. Anthony, S.J.; Johnson, C.K.; Greig, D.J.; Kramer, S.; Che, X.; Wells, H.; Hicks, A.L.; Joly, D.O.; Wolfe, N.D.; Daszak, P.; et al. Global patterns in coronavirus diversity. *Virus Evol.* **2017**, *3*.

63. Schalk, A.F.; Hawn, M.C. An Apparently New Respiratory Disease of Baby Chicks. *J. Amer. Vet. Med. Ass* **1931**, *78*, 413–423.

64. McIntosh, K.; Dees, J.H.; Becker, W.B.; Kapikian, A.Z.; Chanock, R.M. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci.* **1967**, *57*, 933–940.

65. McIntosh, K.; Becker, W.B.; Chanock, R.M. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci.* **1967**, *58*, 2268–2273.

66. Hamre, D.; Procknow, J.J. A new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.* **1966**, *121*, 190–3.

67. Becker, W.B.; McIntosh, K.; Dees, J.H.; Chanock, R.M. Morphogenesis of avian infectious bronchitis virus and a related human virus (strain 229E). *J. Virol.* **1967**, *1*, 1019–27.

68. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Huang, Y.; Tsoi, H.W.; Wong, B.H.L.; Wong, S.S.Y.; Leung, S.Y.; Chan, K.H.; Yuen, K.Y. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14040–14045.

69. Luk, H.K.H.; Li, X.; Fung, J.; Lau, S.K.P.; Woo, P.C.Y. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect. Genet. Evol.* **2019**, *71*, 21–30.

70. Guan, Y.; Zheng, B.J.; He, Y.Q.; Liu, X.L.; Zhuang, Z.X.; Cheung, C.L.; Luo, S.W.; Li, P.H.; Zhang, L.J.; Guan, Y.J.; et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **2003**, *302*, 276–8.

71. de Groot, R.J.; Baker, S.C.; Baric, R.S.; Brown, C.S.; Drosten, C.; Enjuanes, L.; Fouchier, R.A.M.; Galiano, M.; Gorbalenya, A.E.; Memish, Z.A.; et al. Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *J. Virol.* **2013**, *87*, 7790–7792.

72. van Boheemen, S.; de Graaf, M.; Lauber, C.; Bestebroer, T.M.; Raj, V.S.; Zaki, A.M.; Osterhaus, A.D.M.E.; Haagmans, B.L.; Gorbalenya, A.E.; Snijder, E.J.; et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* **2012**, *3*.

73. Lee, S.Y.; Khang, Y.H.; Lim, H.K. Impact of the 2015 Middle East Respiratory Syndrome Outbreak on Emergency Care Utilization and Mortality in South Korea. *Yonsei Med. J.* **2019**, *60*, 796–803.

74. Kim, K.H.; Tandi, T.E.; Choi, J.W.; Moon, J.M.; Kim, M.S. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. *J. Hosp. Infect.* **2017**,

95, 207–213.

75. Shi, Z. Emerging infectious diseases associated with bat viruses. *Sci. China. Life Sci.* **2013**, *56*, 678–82.

76. Wang, N.; Luo, C.; Liu, H.; Yang, X.; Hu, B.; Zhang, W.; Li, B.; Zhu, Y.; Zhu, G.; Shen, X.; et al. Characterization of a New Member of Alphacoronavirus with Unique Genomic Features in Rhinolophus Bats. *Viruses* **2019**, *11*.

77. Zhou, P.; Fan, H.; Lan, T.; Yang, X.-L.; Shi, W.-F.; Zhang, W.; Zhu, Y.; Zhang, Y.-W.; Xie, Q.-M.; Mani, S.; et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **2018**, *556*, 255–258.

78. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Yuen, K.Y. More and more coronaviruses: Human coronavirus HKU1. *Viruses* **2009**, *1*, 57–71.

79. Lau, S.K.P.; Woo, P.C.Y.; Yip, C.C.Y.; Fan, R.Y.Y.; Huang, Y.; Wang, M.; Guo, R.; Lam, C.S.F.; Tsang, A.K.L.; Lai, K.K.Y.; et al. Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J. Virol.* **2012**, *86*, 5481–96.

80. Woo, P.C.Y.; Lau, S.K.P.; Chu, C. -m.; Chan, K. -h.; Tsoi, H. -w.; Huang, Y.; Wong, B.H.L.; Poon, R.W.S.; Cai, J.J.; Luk, W. -k.; et al. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J. Virol.* **2004**, *79*, 884–895.

81. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Tsoi, H.-W.; Chan, K.-H.; Yuen, K.-Y. Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1. *J. Virol.* **2006**, *80*, 7136–7145.

82. Wong, A.C.P.; Li, X.; Lau, S.K.P.; Woo, P.C.Y. Global Epidemiology of Bat Coronaviruses. *Viruses* **2019**, *11*.

83. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A.C.K.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G.F. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016, *24*, 490–502.

84. Zeng, Z.Q.; Chen, D.H.; Tan, W.P.; Qiu, S.Y.; Xu, D.; Liang, H.X.; Chen, M.X.; Li, X.; Lin, Z.S.; Liu, W.K.; et al. Epidemiology and clinical characteristics of human coronaviruses OC43, 229E, NL63, and HKU1: a study of hospitalized children with acute respiratory tract infection in Guangzhou, China. *Eur. J. Clin. Microbiol. Infect. Dis.* **2018**, *37*, 363–369.

85. Morfopoulou, S.; Brown, J.R.; Davies, E.G.; Anderson, G.; Virasami, A.; Qasim, W.; Chong, W.K.; Hubank, M.; Plagnol, V.; Desforges, M.; et al. Human Coronavirus OC43 Associated with Fatal Encephalitis. *N. Engl. J. Med.* **2016**, *375*, 497–8.

86. Abdul-Rasool, S.; Fielding, B.C. Understanding Human Coronavirus HCoV-NL63. *Open Virol. J.* **2010**, *4*, 76–84.

87. Vassilara, F.; Spyridaki, A.; Pothitos, G.; Deliveliotou, A.; Papadopoulos, A. A Rare Case of Human Coronavirus 229E Associated with Acute Respiratory Distress Syndrome in a Healthy Adult. *Case Rep. Infect. Dis.* **2018**, *2018*, 6796839.

88. Corman, V.M.; Eckerle, I.; Memish, Z.A.; Liljander, A.M.; Dijkman, R.; Jonsdottir, H.; Ngeiywa, K.J.Z.J.; Kamau, E.; Younan, M.; Al Masri, M.; et al. Link of a ubiquitous human coronavirus to dromedary camels. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9864–9869.

89. Vijgen, L.; Keyaerts, E.; Moes, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604.

90. Vijgen, L.; Keyaerts, E.; Lemey, P.; Maes, P.; Van Reeth, K.; Nauwynck, H.; Pensaert, M.; Van Ranst, M. Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *J. Virol.* **2006**, *80*, 7270–7274.

91. Lau, S.K.P.; Lee, P.; Tsang, A.K.L.; Yip, C.C.Y.; Tse, H.; Lee, R.A.; So, L.-Y.; Lau, Y.-L.; Chan, K.-H.; Woo, P.C.Y.; et al. Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. *J. Virol.* **2011**, *85*, 11325–11337.

92. Taniguchi, T.; Palmieri, M.; Weissmann, C. QB DNA-containing hybrid plasmids giving rise to QB phage formation in the bacterial host. *Nature* **1978**, *274*, 223–8.

93. Racaniello, V.R.; Baltimore, D. Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science* **1981**, *214*, 916–9.

94. Ahlquist, P.; French, R.; Janda, M.; Loesch-Fries, L.S. Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 7066–70.

95. Koetzner, C.A.; Parker, M.M.; Ricard, C.S.; Sturman, L.S.; Masters, P.S. Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. *J. Virol.* **1992**, *66*, 1841–8.

96. van der Most, R.G.; Heijnen, L.; Spaan, W.J.M.; de Groot, R.J. Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucleic Acids Res.* **1992**, *20*, 3375–3381.

97. Masters, P.S.; Koetzner, C.A.; Kerr, C.A.; Heo, Y. Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis virus. *J. Virol.* **1994**, *68*, 328–37.

98. Kuo, L.; Godeke, G.-J.; Raamsman, M.J.B.; Masters, P.S.; Rottier, P.J.M. Retargeting of Coronavirus by Substitution of the Spike Glycoprotein Ectodomain: Crossing the Host Cell Species Barrier. *J. Virol.* **2000**, *74*, 1393–1406.

99. Lissenberg, A.; Vrolijk, M.M.; van Vliet, A.L.W.; Langereis, M.A.; de Groot-Mijnes, J.D.F.; Rottier, P.J.M.; de Groot, R.J. Luxury at a Cost? Recombinant Mouse Hepatitis Viruses Expressing the Accessory Hemagglutinin Esterase Protein Display Reduced Fitness In Vitro. *J. Virol.* **2005**, *79*, 15054–15063.

100. Li, C.; Li, Z.; Zou, Y.; Wicht, O.; van Kuppeveld, F.J.M.; Rottier, P.J.M.; Bosch, B.J. Manipulation of the porcine epidemic diarrhea virus genome using targeted RNA recombination. *PLoS One* **2013**, *8*, e69997.

101. van Beurden, S.J.; Berends, A.J.; Krämer-Kühl, A.; Spekreijse, D.; Chénard, G.; Philipp, H.-C.; Mundt, E.; Rottier, P.J.M.; Verheije, M.H. A reverse genetics system for avian coronavirus infectious bronchitis virus based on targeted RNA recombination. *Virol. J.* **2017**, *14*, 109.

102. Masters, P.S.; Rottier, P.J.M. Coronavirus Reverse Genetics by Targeted RNA Recombination. In *Coronavirus Replication and Reverse Genetics*; 2005; pp. 133–159.

103. de Haan, C.A.M.; van Genne, L.; Stoop, J.N.; Volders, H.; Rottier, P.J.M. Coronaviruses as vectors: position dependence of foreign gene expression. *J. Virol.* **2003**, *77*, 11312–23.

104. de Haan, C.A.M.; Haijema, B.J.; Masters, P.S.; Rottier, P.J.M. Manipulation of the coronavirus genome using targeted RNA recombination with interspecies chimeric coronaviruses. *Methods Mol. Biol.* **2008**, *454*, 229–36.

105. Almazán, F.; González, J.M.; Pénzes, Z.; Izeta, A.; Calvo, E.; Plana-Durán, J.; Enjuanes, L. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5516–21.

106. González, J.M.; Pénzes, Z.; Almazán, F.; Calvo, E.; Enjuanes, L. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. *J. Virol.* **2002**, *76*, 4655–61.

107. Eriksson, K.K.; Makia, D.; Thiel, V. Generation of recombinant coronaviruses using vaccinia virus as the cloning vector and stable cell lines containing coronaviral replicon RNAs. *Methods Mol. Biol.* **2008**, *454*, 237–54.

108. Yount, B.; Denison, M.R.; Weiss, S.R.; Baric, R.S. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. *J. Virol.* **2002**, *76*, 11065–78.

109. Navas-Martin, S.; Weiss, S.R. SARS: lessons learned from other coronaviruses. *Viral Immunol.* **2003**, *16*, 461–74.

110. Bosch, B.J.; van der Zee, R.; de Haan, C.A.M.; Rottier, P.J.M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *J. Virol.* **2003**, *77*, 8801–8811.

111. Callebaut, P.E.; Pensaert, M.B. Characterization and isolation of structural polypeptides in haemagglutinating encephalomyelitis virus. *J. Gen. Virol.* **1980**, *48*, 193–204.

112. Sugiyama, K.; Amano, Y. Hemagglutination and structural polypeptides of a new coronavirus associated with diarrhea in infant mice. *Arch. Virol.* **1980**, *66*, 95–105.

113. King, B.; Potts, B.J.; Brian, D.A. Bovine coronavirus hemagglutinin protein. *Virus Res.* **1985**, *2*, 53–59.

114. Hogue, B.G.; Brian, D.A. Structural proteins of human respiratory coronavirus OC43. *Virus Res.* **1986**, *5*, 131–144.

115. Dea, S.; Garzon, S.; Strykowski, H.; Tijssen, P. Ultrastructure and protein A-gold immunolabelling of HRT-18 cells infected with turkey enteric coronavirus. *Vet. Microbiol.* **1989**, *20*, 21–33.

116. King, B.; Brian, D.A. Bovine Coronavirus Structural Proteins. *J. Virol.* **1982**, *42*, 700–707.

117. Hulswit, R.J.G.; de Haan, C.A.M.; Bosch, B.J. Coronavirus Spike Protein and Tropism Changes. In *Advances in Virus Research*; 2016; Vol. 96, pp. 29–57.

118. Schultze, B.; Gross, H.J.; Brossmer, R.; Herrler, G. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J. Virol.* **1991**, *65*, 6232–7.

119. Huang, X.; Dong, W.; Milewska, A.; Golda, A.; Qi, Y.; Zhu, Q.K.; Marasco, W.A.; Baric, R.S.; Sims, A.C.; Pyrc, K.; et al. Human Coronavirus HKU1 Spike Protein Uses O -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J. Virol.* **2015**, *89*, 7202–7213.

120. Krempl, C.; Schultze, B.; Herrler, G. Analysis of cellular receptors for human coronavirus OC43. In Proceedings of the Advances in Experimental Medicine and Biology; 1995; Vol. 380, pp. 371–374.

121. Peng, G.; Sun, D.; Rajashankar, K.R.; Qian, Z.; Holmes, K. V; Li, F. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10696–701.

122. Langereis, M.A.; van Vliet, A.L.W.; Boot, W.; de Groot, R.J. Attachment of Mouse Hepatitis Virus to O-Acetylated Sialic Acid Is Mediated by Hemagglutinin-Esterase and Not by the Spike Protein. *J. Virol.* **2010**, *84*, 8970–8974.

123. Dveksler, G.S.; Pensiero, M.N.; Cardellichio, C.B.; Williams, R.K.; Jiang, G.S.; Holmes, K. V; Dieffenbach, C.W. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. *J. Virol.* **1991**, *65*, 6881–91.

124. Williams, R.K.; Jiang, G.S.; Holmes, K. V. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 5533–5536.

125. Mounir, S.; Labonté, P.; Talbot, P.J. Characterization of the nonstructural and spike proteins of the human respiratory coronavirus OC43: comparison with bovine enteric coronavirus. *Adv. Exp. Med. Biol.* **1993**, *342*, 61–7.

126. So, R.T.Y.; Chu, D.K.W.; Miguel, E.; Perera, R.A.P.M.; Oladipo, J.O.; Fassi-Fihri, O.; Aylet, G.; Ko, R.L.W.; Zhou, Z.; Cheng, M.-S.; et al. Diversity of dromedary camel coronavirus HKU23 in African camels revealed multiple recombination events among closely related Betacoronaviruses of the subgenus Embecovirus. *J. Virol.* **2019**.

127. Abraham, S.; Kienzle, T.E.; Lapps, W.E.; Brian, D.A. Sequence and expression analysis of potential nonstructural proteins of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane protein genes of the bovine coronavirus. *Virology* **1990**, *177*, 488–95.

128. Zhang, R.; Wang, K.; Ping, X.; Yu, W.; Qian, Z.; Xiong, S.; Sun, B. The ns12.9 Accessory Protein of Human Coronavirus OC43 Is a Viroporin Involved in Virion Morphogenesis and Pathogenesis. *J. Virol.* **2015**, *89*, 11383–11395.

129. Zhao, L.; Jha, B.K.; Wu, A.; Elliott, R.; Ziebuhr, J.; Gorbalenya, A.E.; Silverman, R.H.; Weiss, S.R. Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host Microbe* **2012**, *11*, 607–16.

130. Thornbrough, J.M.; Jha, B.K.; Yount, B.; Goldstein, S.A.; Li, Y.; Elliott, R.; Sims, A.C.; Baric, R.S.; Silverman, R.H.; Weiss, S.R. Middle East Respiratory Syndrome Coronavirus NS4b Protein Inhibits Host RNase L Activation. *MBio* **2016**, *7*, e00258.

131. Smits, S.L.; Snijder, E.J.; de Groot, R.J. Characterization of a Torovirus Main Proteinase. *J. Virol.* **2006**, *80*, 4157–4167.

132. Sui, B.; Huang, J.; Jha, B.K.; Yin, P.; Zhou, M.; Fu, Z.F.; Silverman, R.H.; Weiss, S.R.; Peng, G.; Zhao, L. Crystal structure of the mouse hepatitis virus ns2 phosphodiesterase domain that antagonizes RNase L activation. *J. Gen. Virol.* **2016**, *97*, 880–886.

133. Vlasak, R.; Luytjes, W.; Leider, J.; Spaan, W.; Palese, P. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J. Virol.* **1988**, *62*, 4686–90.

134. Shieh, C.K.; Lee, H.J.; Yokomori, K.; La Monica, N.; Makino, S.; Lai, M.M. Identification of a new transcriptional initiation site and the corresponding functional gene 2b in the murine coronavirus RNA genome. *J. Virol.* **1989**, *63*, 3729–36.

135. Makino, S.; Lai, M.M. Evolution of the 5'-end of genomic RNA of murine coronaviruses during passages in vitro. *Virology* **1989**, *169*, 227–32.

136. Zhang, X.; Kousoulas, K.G.; Storz, J. The hemagglutinin/esterase gene of human coronavirus strain OC43: Phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. *Virology* **1992**, *186*, 318–323.

137. Rogers, G.N.; Herrler, G.; Paulson, J.C.; Klenk, H.D. Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. *J. Biol. Chem.* **1986**, *261*, 5947–51.

138. Herrler, G.; Rott, R.; Klenk, H.D.; Müller, H.P.; Shukla, A.K.; Schauer, R. The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. *EMBO J.* **1985**, *4*, 1503–6.

139. Rosenthal, P.B.; Zhang, X.; Formanowski, F.; Fitz, W.; Wong, C.H.; Meier-Ewert, H.; Skehel, J.J.; Wiley, D.C. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **1998**, *396*, 92–96.

140. Popova, R.; Zhang, X. The spike but not the hemagglutinin/esterase protein of bovine coronavirus is necessary and sufficient for viral infection. *Virology* **2002**, *294*, 222–36.

141. Schultze, B.; Wahn, K.; Klenk, H.D.; Herrler, G. Isolated HE-protein from hemagglutinating encephalomyelitis virus and bovine coronavirus has receptor-destroying and receptor-binding activity. *Virology* **1991**, *180*, 221–8.

142. Snijder, E.J.; Den Boon, J.A.; Horzinek, M.C.; Spaan, W.J.M. Comparison of the genome organization of toro- and coronaviruses: Evidence for two nonhomologous RNA recombination events during berne virus evolution. *Virology* **1991**, *180*, 448–452.

143. de Groot, R.J. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* 2006, *23*, 59–72.

144. Navas, S.; Seo, S.H.; Chua, M.M.; Das Sarma, J.; Hingley, S.T.; Lavi, E.; Weiss, S.R. Role of the spike protein in murine coronavirus induced hepatitis: an in vivo study using targeted RNA recombination. *Adv. Exp. Med. Biol.* **2001**, *494*, 139–44.

145. Phillips, J.J.; Chua, M.M.; Rall, G.F.; Weiss, S.R. Murine coronavirus spike glycoprotein mediates degree of viral spread, inflammation, and virus-induced immunopathology in the central nervous system. *Virology* **2002**, *301*, 109–20.

146. Yokomori, K.; Banner, L.R.; Lai, M.M. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology* **1991**, *183*, 647–57.

147. Kazi, L.; Lissenberg, A.; Watson, R.; de Groot, R.J.; Weiss, S.R. Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence. *J. Virol.* **2005**, *79*, 15064–73.

148. Desforges, M.; Desjardins, J.; Zhang, C.; Talbot, P.J. The Acetyl-Esterase Activity of the Hemagglutinin-Esterase Protein of Human Coronavirus OC43 Strongly Enhances the Production of Infectious Virus. *J. Virol.* **2013**, *87*, 3097–3107.

149. Deregt, D.; Babiuk, L.A. Monoclonal antibodies to bovine coronavirus: Characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology* **1987**, *161*, 410–420.

150. Deregt, D.; Gifford, G.A.; Ijaz, M.K.; Watts, T.C.; Gilchrist, J.E.; Haines, D.M.; Babiuk, L.A. Monoclonal Antibodies to Bovine Coronavirus Glycoproteins E2 and E3: Demonstration of in vivo Virus-neutralizing Activity. *J. Gen. Virol.* **1989**, *70*, 993–998.

151. Milane, G.; Kourtesis, A.B.; Dea, S. Characterization of monoclonal antibodies to the hemagglutinin-esterase glycoprotein of a bovine coronavirus associated with winter dysentery and cross-reactivity to field isolates. *J. Clin. Microbiol.* **1997**, *35*, 33–40.

152. Strasser, P.; Unger, U.; Strobl, B.; Vilas, U.; Vlasak, R. Recombinant viral sialate-O-acetylesterases. *Glycoconj. J.* **2004**, *20*, 551–61.

153. Wurzer, W.J.; Obojes, K.; Vlasak, R. The sialate-4-O-acetylesterases of coronaviruses related to mouse hepatitis virus: a proposal to reorganize group 2 Coronaviridae. *J. Gen. Virol.* **2002**, *83*, 395–402.

154. Smits, S.L.; Gerwig, G.J.; van Vliet, A.L.W.; Lissenberg, A.; Briza, P.; Kamerling, J.P.; Vlasak, R.; de Groot, R.J. Nidovirus sialate-O-acetylesterases: evolution and substrate specificity of coronaviral and toroviral receptor-destroying enzymes. *J. Biol. Chem.* **2005**, *280*, 6933–41.

155. Zeng, Q.; Langereis, M.A.; van Vliet, A.L.W.; Huizinga, E.G.; de Groot, R.J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9065–9069.

156. Langereis, M.A.; Zeng, Q.; Heesters, B.; Huizinga, E.G.; de Groot, R.J. The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. *PLoS Pathog.* **2012**, *8*.

157. Bakkers, M.J.G.; Zeng, Q.; Feitsma, L.J.; Hulswit, R.J.G.; Li, Z.; Westerbeke, A.; van Kuppeveld, F.J.M.; Boons, G.-J.; Langereis, M.A.; Huizinga, E.G.; et al. Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proc. Natl. Acad. Sci.* **2016**, *113*, E3111–E3119.

158. Langereis, M.A.; Zeng, Q.; Gerwig, G.J.; Frey, B.; Itzstein, M. von; Kamerling, J.P.; de Groot, R.J.; Huizinga, E.G. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15897–15902.

159. Wang, Q.; Qi, J.; Yuan, Y.; Xuan, Y.; Han, P.; Wan, Y.; Ji, W.; Li, Y.; Wu, Y.; Wang, J.; et al. Bat origins of MERS-CoV supported by bat Coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **2014**, *16*, 328–337.

160. Delmas, B.; Gelfi, J.; Sjöström, H.; Noren, O.; Laude, H. Further characterization of aminopeptidase-N as a receptor for coronaviruses. *Adv. Exp. Med. Biol.* **1993**, *342*, 293–8.

161. Li, W.; Hulswit, R.J.G.; Kenney, S.P.; Widjaja, I.; Jung, K.; Alhamo, M.A.; van Dieren, B.; van Kuppeveld, F.J.M.; Saif, L.J.; Bosch, B.-J. Broad receptor engagement of an emerging global coronavirus may potentiate its diverse cross-species transmissibility. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E5135–E5143.

162. Delmas, B.; Gelfi, J.; L'Haridon, R.; Vogel, L.K.; Sjöström, H.; Norén, O.; Laude, H. Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. *Nature* **1992**, *357*, 417–420.

163. Tresnan, D.B.; Levis, R.; Holmes, K. V Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. *J. Virol.* **1996**, *70*, 8669–74.

164. Yeager, C.L.; Ashmun, R.A.; Williams, R.K.; Cardellichio, C.B.; Shapiro, L.H.; Look, A.T.; Holmes, K. V Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **1992**, *357*, 420–2.

165. Hofmann, H.; Pyrc, K.; van der Hoek, L.; Geier, M.; Berkhout, B.; Pöhlmann, S. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7988–93.

166. Kuhn, J.H.; Li, W.; Choe, H.; Farzan, M. Angiotensin-converting enzyme 2: A functional receptor for SARS coronavirus. *Cell. Mol. Life Sci.* **2004**, *61*, 2738–2743.

167. Williams, R.K.; Jiang, G.S.; Holmes, K. V Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *88*.

168. Raj, V.S.; Mou, H.; Smits, S.L.; Dekkers, D.H.W.; Müller, M.A.; Dijkman, R.; Muth, D.; Demmers, J.A.A.; Zaki, A.; Fouchier, R.A.M.; et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **2013**, *495*, 251–254.

169. Yang, Y.; Du, L.; Liu, C.; Wang, L.; Ma, C.; Tang, J.; Baric, R.S.; Jiang, S.; Li, F. 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.* **2014**, *111*, 12516–12521.

170. Dam, T.K.; Brewer, C.F. Multivalent lectin-carbohydrate interactions energetics and mechanisms of binding. *Adv. Carbohydr. Chem. Biochem.* 2010, *63*, 139–164.

171. Dam, T.K.; Brewer, C.F. Effects of clustered epitopes in multivalent ligand-receptor interactions. *Biochemistry* **2008**, *47*, 8470–6.

172. Dam, T.K.; Brewer, F.C. Maintenance of cell surface glycan density by lectin-glycan interactions: A homeostatic and innate immune regulatory mechanism. *Glycobiology* 2010, *20*, 1061–1064.

173. Peng, G.; Xu, L.; Lin, Y.L.; Chen, L.; Pasquarella, J.R.; Holmes, K. V.; Li, F. Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **2012**, *287*, 41931–41938.

Supplementary table. Viruses that use glycan-based receptors or initial attachment factors.

Family/Genus	Virus Species	Receptor/attachment factor	Ref
Adenoviridae	Adenovirus type 8	Sialic acid	[1]
	Adenovirus type 19a	Sialic acid	[1]
	Adenovirus type 37	Sialic acid	[2]
	Bovine adenovirus 3	Sialic acid	[3]
	Canine adenovirus 2	Sialic acid	[4]
Arenaviridae	Lassa virus	Sialic acid	[5]
Arteriviridae	Porcine reproductive and respiratory syndrome virus	heparan sulfate	[6]
Caliciviridae	Bovine norovirus	αGal epitope	[7]
	Canine norovirus	HBGA carbohydrate	[8]
	Human norovirus	HBGA carbohydrate	[9]
	Lagovirus	Glycosphingolipids	[10]
	Murine norovirus	Sialic acid	[11]
	Porcine Sapovirus	Sialic acid	[12]
	Vesivirus	Glycosphingolipids	[10]
Circoviridae	Porcine circovirus 2	heparan sulfate/ chondroitin sulfate B	[13]
Coronaviridae	Bovine coronavirus	9-*O*-acetylated sialic acid	[14–16]
	Canine respiratory coronavirus	9-*O*-acetylated sialic acid	This thesis
	Equine coronavirus	9-*O*-acetylated sialic acid	[17]
	Human coronavirus OC43	9-*O*-acetylated sialic acid	[14]
	Human coronavirus HKU1	9-*O*-acetylated sialic acid	[18]
	Porcine hemagglutinating encephalomyelitis virus	9-*O*-acetylated sialic acid	[19]
	Mouse hepatitis virus	4-*O*-/9-*O*-acetylated sialic acid	[20–23]
	Rat coronavirus	4-*O*-acetylated sialic acid	[24]
	Rabbit coronavirus	9-*O*-acetylated sialic acid	This thesis
	Middle East Respiratory Syndrome coronavirus	sialic acid	[25]
	Porcine transmissible gastroenteritis coronavirus	sialic acid (Neu5Gc)	[26]
	Infectious bronchitis coronavirus	sialic acid	[27]
	Porcine epidemic diarrhea virus	sialic acid	[28]
Deltavirus	Hepatitis D virus	heparan sulfate	[29]
Flaviviridae	Classical swine fever virus	heparan sulfate	[30]
	Dengue virus	heparan sulfate	[31]
	Japanese encephalitis virus	heparan sulfate	[32]
	Hepatitis C virus	heparan sulfate	[33]
	Yellow fever virus	heparan sulfate	[34]
Hepadnaviridae	Hepatitis B virus	sialic acid	[35]
Herpesviridae	Bovine herpesvirus	heparan sulfate	[36]
	Equine herpesvirus-1	heparan sulfate	[37]
	Herpes simplex virus type 1	heparan sulfate	[38]
	Herpes simplex virus type 2	heparan sulfate	[39]
	Human cytomegalovirus (HHV-5)	sialic acid	[40]
	Human herpesvirus 7 (HHV-7)	heparan sulfate	[41]
	Kaposi Sarcoma virus (HHV-8)	heparan sulfate	[42]
	Mouse cytomegalovirus	sialic acid	[43]
	Pseudorabies virus	heparan sulfate	[44]

	Varicella-zoster virus (HHV-3)	heparan sulfate	[45]
Microviridae	Enterobacteria phage PhiX174	lipopolysaccharide	[46]
Myoviridae	Bacteriophage Mu	Polysaccharide part of LPS	[47]
Orthomyxoviridae	Avian Influenza A virus	sialic acid	[48–52]
	Human Influenza A virus	sialic acid	
	Influenza B virus	sialic acid	[53,54]
	Influenza C virus	9-*O*-acetylated sialic acid	[55–57]
	Influenza D virus	9-*O*-acetylated sialic acid	[58,59]
	Isavirus	4-*O*-acetylated sialic acid	[60]
Paramyxoviridae	Human parainfluenza 1	sialic acid	[61]
	Human parainfluenza 3	sialic acid	[61]
	Human metapneumovirus	glycosaminoglycans	[62]
	Mammalian orthorubulavirus 5	sialic acid	[63]
	Mumps virus	sialic acid	[64]
	Newcastle disease virus	sialic acid	[65]
	Porcine rubulavirus LPM	sialic acid	[66,67]
	Respiratory syncytial virus	heparan sulfate	[68]
	Sendai virus	sialic acid	[69]
Parvoviridae	Adeno-associated viruses	sialic acid / heparan sulfate	[70–73]
	Bovine adeno-associated virus	sialic acid	[74]
	Bovine parvovirus	sialic acid	[75]
	Canine parvovirus	sialic acid	[76,77]
	Feline panleukopenia virus	sialic acid	[77]
	Minute virus of mice	sialic acid	[78]
Picornaviridae	Bovine enterovirus 261	sialic acid	[79]
	Encephalomyocarditis virus	sialic acid	[80]
	Enterovirus 70	sialic acid	[81]
	Enterovirus 71	sialic acid	[82]
	Enterovirus D68	sialic acid	[83]
	Equine rhinitis A virus	sialic acid	[84]
	Human coxsackievirus A24v	sialic acid	[85,86]
	Mengovirus	sialic acid	[87]
	Theiler's encephalomyelitis virus	sialic acid	[88]
Polyomaviridae	BK virus	sialic acid	[89]
	Human JC polyomavirus	sialic acid	[90]
	Merkel cell polyomavirus	sialic acid	[91]
	Monkey B-lymphotropic papovavirus	sialic acid	[92]
	Mouse polyomavirus mPy	sialic acid	[93]
	Simian virus 40	sialic acid	[94,95]
Poxviridae	Vaccinia virus	heparan sulfate	[96]
Reovirida	Bluetongue virus	sialic acid	[97]
	Bovine rotavirus	sialic acid	[98]
	Mammalian reovirus	sialic acid	[99–101]
	Rhesus Rotavirus	sialic acid	[102]
	Porcine Rotavirus	sialic acid/mucin core 2	[103–105]
	Simian rotavirus	sialic acid	[98,106]
Retroviridae	Human immunodeficiency virus 1	heparan sulfate	[107]
Rhabdoviridae	Rabies virus	sialic acid	[108]
	Vesicular stomatitis virus	sialic acid	[109]
Tobaniviridae	Bovine torovirus	9-*O*-acetylated sialic acid	[110]
	Porcine torovirus	9-*O*-acetylated sialic acid	[27,110]
Togaviridae	Sindbis virus	heparan sulfate	[111]

Supplemental References

1. Arnberg, N.; Kidd, A.H.; Edlund, K.; Olfat, F.; Wadell, G. Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. *J. Virol.* **2000**, *74*, 7691–3.
2. Arnberg, N.; Edlund, K.; Kidd, A.H.; Wadell, G. Adenovirus Type 37 Uses Sialic Acid as a Cellular Receptor. *J. Virol.* **2000**, *74*, 42–48.
3. Li, X.; Bangari, D.S.; Sharma, A.; Mittal, S.K. Bovine adenovirus serotype 3 utilizes sialic acid as a cellular receptor for virus entry. *Virology* **2009**, *392*, 162–168.
4. Seiradake, E.; Henaff, D.; Wodrich, H.; Billet, O.; Perreau, M.; Hippert, C.; Mennechet, F.; Schoehn, G.; Lortat-Jacob, H.; Dreja, H.; et al. The cell adhesion molecule "CAR" and sialic acid on human erythrocytes influence adenovirus in vivo biodistribution. *PLoS Pathog.* **2009**, *5*.
5. Jae, L.T.; Raaben, M.; Herbert, A.S.; Kuehne, A.I.; Wirchnianski, A.S.; Soh, T.K.; Stubbs, S.H.; Janssen, H.; Damme, M.; Saftig, P.; et al. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* **2014**, *344*, 1506–10.
6. Delputte, P.L.; Costers, S.; Nauwynck, H.J. Analysis of porcine reproductive and respiratory syndrome virus attachment and internalization: Distinctive roles for heparan sulphate and sialoadhesin. *J. Gen. Virol.* **2005**, *86*, 1441–1445.
7. Zakhour, M.; Ruvoën-Clouet, N.; Charpilienne, A.; Langpap, B.; Poncet, D.; Peters, T.; Bovin, N.; Le Pendu, J. The αGal epitope of the histo-blood group antigen family is a ligand for bovine norovirus Newbury2 expected to prevent cross-species transmission. *PLoS Pathog.* **2009**, *5*.
8. Caddy, S.; Breiman, A.; le Pendu, J.; Goodfellow, I. Genogroup IV and VI Canine Noroviruses Interact with Histo-Blood Group Antigens. *J. Virol.* **2014**, *88*, 10377–10391.
9. Cao, S.; Lou, Z.; Tan, M.; Chen, Y.; Liu, Y.; Zhang, Z.; Zhang, X.C.; Jiang, X.; Li, X.; Rao, Z. Structural basis for the recognition of blood group trisaccharides by norovirus. *J. Virol.* **2007**, *81*, 5949–57.
10. Taube, S.; Jiang, M.; Wobus, C.E. Glycosphingolipids as receptors for non-enveloped viruses. *Viruses* 2010, *2*, 1011–1049.
11. Taube, S.; Perry, J.W.; Yetming, K.; Patel, S.P.; Auble, H.; Shu, L.; Nawar, H.F.; Lee, C.H.; Connell, T.D.; Shayman, J.A.; et al. Ganglioside-Linked Terminal Sialic Acid Moieties on Murine Macrophages Function as Attachment Receptors for Murine Noroviruses. *J. Virol.* **2009**, *83*, 4092–4101.
12. Kim, D.S.; Hosmillo, M.; Alfajaro, M.M.; Kim, J.Y.; Park, J.G.; Son, K.Y.; Ryu, E.H.; Sorgeloos, F.; Kwon, H.J.; Park, S.J.; et al. Both α2,3- and α2,6-Linked Sialic Acids on O-Linked Glycoproteins Act as Functional Receptors for Porcine Sapovirus. *PLoS Pathog.* **2014**, *10*.
13. Misinzo, G.; Delputte, P.L.; Meerts, P.; Lefebvre, D.J.; Nauwynck, H.J. Porcine Circovirus 2 Uses Heparan Sulfate and Chondroitin Sulfate B Glycosaminoglycans as Receptors for Its Attachment to Host Cells. *J. Virol.* **2006**, *80*, 3487–3494.
14. Künkel, F.; Herrler, G. Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **1993**, *195*, 195–202.
15. Vlasak, R.; Luytjes, W.; Leider, J.; Spaan, W.; Palese, P. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J. Virol.* **1988**, *62*, 4686–90.
16. Zeng, Q.; Langereis, M.A.; van Vliet, A.L.W.; Huizinga, E.G.; de Groot, R.J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9065–9069.
17. Langereis, M.A.; Bakkers, M.J.G.; Deng, L.; Padler-Karavani, V.; Vervoort, S.J.; Hulswit, R.J.G.; van Vliet, A.L.W.; Gerwig, G.J.; de Poot, S.A.H.; Boot, W.; et al. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* **2015**, *11*, 1966–1978.
18. Huang, X.; Dong, W.; Milewska, A.; Golda, A.; Qi, Y.; Zhu, Q.K.; Marasco, W.A.; Baric, R.S.; Sims, A.C.; Pyrc, K.; et al. Human Coronavirus HKU1 Spike Protein Uses O -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J. Virol.* **2015**, *89*, 7202–7213.
19. Schultze, B.; Gross, H.J.; Brossmer, R.; Klenk, H.D.; Herrler, G. Hemagglutinating encephalomyelitis virus attaches to N-acetyl-9-O-acetylneuraminic acid-containing receptors on erythrocytes: comparison with bovine coronavirus and influenza C virus. *Virus Res.* **1990**, *16*, 185–94.
20. Langereis, M.A.; van Vliet, A.L.W.; Boot, W.; de Groot, R.J. Attachment of Mouse Hepatitis Virus to O-Acetylated Sialic Acid Is Mediated by Hemagglutinin-Esterase and Not by the Spike Protein. *J. Virol.* **2010**, *84*, 8970–8974.
21. Regl, G.; Kaser, A.; Iwersen, M.; Schmid, H.; Kohla, G.; Strobl, B.; Vilas, U.; Schauer, R.; Vlasak, R. The hemagglutinin-esterase of mouse hepatitis virus strain S is a sialate-4-O-acetylesterase. *J. Virol.* **1999**, *73*, 4721–7.

22. Smits, S.L.; Gerwig, G.J.; van Vliet, A.L.W.; Lissenberg, A.; Briza, P.; Kamerling, J.P.; Vlasak, R.; de Groot, R.J. Nidovirus sialate-O-acetylesterases: evolution and substrate specificity of coronaviral and toroviral receptor-destroying enzymes. *J. Biol. Chem.* **2005**, *280*, 6933–41.

23. Sugiyama, K.; Amano, Y. Hemagglutination and structural polypeptides of a new coronavirus associated with diarrhea in infant mice. *Arch. Virol.* **1980**, *66*, 95–105.

24. Bakkers, M.J.G.; Zeng, Q.; Feitsma, L.J.; Hulswit, R.J.G.; Li, Z.; Westerbeke, A.; van Kuppeveld, F.J.M.; Boons, G.-J.; Langereis, M.A.; Huizinga, E.G.; et al. Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proc. Natl. Acad. Sci.* **2016**, *113*, E3111–E3119.

25. Li, W.; Hulswit, R.J.G.; Widjaja, I.; Raj, V.S.; McBride, R.; Peng, W.; Widagdo, W.; Tortorici, M.A.; van Dieren, B.; Lang, Y.; et al. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci.* **2017**, *114*, E8508–E8517.

26. Schultze, B.; Krempl, C.; Ballesteros, M.L.; Shaw, L.; Schauer, R.; Enjuanes, L.; Herrler, G. Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. *J. Virol.* **1996**, *70*, 5634–7.

27. Schultze, B.; Cavanagh, D.; Herrler, G. Neuraminidase treatment of avian infectious bronchitis coronavirus reveals a hemagglutinating activity that is dependent on sialic acid-containing receptors on erythrocytes. *Virology* **1992**, *189*, 792–794.

28. Li, W.; van Kuppeveld, F.J.M.; He, Q.; Rottier, P.J.M.; Bosch, B.-J. Cellular entry of the porcine epidemic diarrhea virus. *Virus Res.* **2016**, *226*, 117–127.

29. Leistner, C.M.; Gruen-Bernhard, S.; Glebe, D. Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cell. Microbiol.* **2008**, *10*, 122–133.

30. Eymann-Häni, R.; Leifer, I.; McCullough, K.C.; Summerfield, A.; Ruggli, N. Propagation of classical swine fever virus in vitro circumventing heparan sulfate-adaptation. *J. Virol. Methods* **2011**, *176*, 85–95.

31. Chen, Y.; Maguire, T.; Hileman, R.E.; Fromm, J.R.; Esko, J.D.; Linhardt, R.J.; Marks, R.M. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. *Nat. Med.* **1997**, *3*, 866–871.

32. Liu, H.; Chiou, S.S.; Chen, W.J. Differential Binding Efficiency between the Envelope Protein of Japanese Encephalitis Virus Variants and Heparan Sulfate on the Cell Surface. *J. Med. Virol.* **2004**, *72*, 618–624.

33. Xu, Y.; Martinez, P.; Séron, K.; Luo, G.; Allain, F.; Dubuisson, J.; Belouzard, S. Characterization of hepatitis C virus interaction with heparan sulfate proteoglycans. *J. Virol.* **2015**, *89*, 3846–58.

34. Germi, R.; Crance, J.M.; Garin, D.; Guimet, J.; Lortat-Jacob, H.; Ruigrok, R.W.H.; Zarski, J.P.; Drouet, E. Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus. *Virology* **2002**, *292*, 162–168.

35. Komai, K.; Kaplan, M.; Peeples, M.E. The vero cell receptor for the hepatitis B virus small S protein is a sialoglycoprotein. *Virology* **1988**, *163*, 629–634.

36. Byrne, K.M.; Horohov, D.W.; Kousoulas, K.G. Glycoprotein B of bovine herpesvirus-1 binds heparin. *Virology* **1995**, *209*, 230–235.

37. Frampton, A.R.; Goins, W.F.; Cohen, J.B.; von Einem, J.; Osterrieder, N.; O'Callaghan, D.J.; Glorioso, J.C. Equine Herpesvirus 1 Utilizes a Novel Herpesvirus Entry Receptor. *J. Virol.* **2005**, *79*, 3169–3173.

38. Shukla, D.; Liu, J.; Blaiklock, P.; Shworak, N.W.; Bai, X.; Esko, J.D.; Cohen, G.H.; Eisenberg, R.J.; Rosenberg, R.D.; Spear, P.G. A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry. *Cell* **1999**, *99*, 13–22.

39. Cheshenko, N.; Herold, B.C. Glycoprotein B plays a predominant role in mediating herpes simplex virus type 2 attachment and is required for entry and cell-to-cell spread. *J. Gen. Virol.* **2002**, *83*, 2247–2255.

40. Lober, P.E.; Hober, D.; Dewilde, A.; Wattré, P. Cell membrane bound N-acetylneuraminic acid is involved in the infection of fibroblasts and phorbol-ester differentiated monocyte-like cells with human cytomegalovirus (HCMV). *Arch. Virol.* **1995**, *140*, 1357–1371.

41. Secchiero, P.; Sun, D.; De Vico, A.L.; Crowley, R.W.; Reitz, M.S.; Zauli, G.; Lusso, P.; Gallo, R.C. Role of the extracellular domain of human herpesvirus 7 glycoprotein B in virus binding to cell surface heparan sulfate proteoglycans. *J. Virol.* **1997**, *71*, 4571–80.

42. Hahn, A.; Birkmann, A.; Wies, E.; Dorer, D.; Mahr, K.; Sturzl, M.; Titgemeyer, F.; Neipel, F. Kaposi's Sarcoma-Associated Herpesvirus gH/gL: Glycoprotein Export and Interaction with Cellular Receptors. *J. Virol.* **2009**, *83*, 396–407.

43. Ravindranath, R.M.; Graves, M.C. Attenuated murine cytomegalovirus binds to N-acetylglucosamine, and shift to virulence may involve recognition of sialic acids. *J. Virol.* **1990**, *64*, 5430–40.

44. Sawitzky, D.; Hampl, H.; Habermehl, K.O. Comparison of heparin-sensitive attachment of pseudorabies virus (PRV)

and herpes simplex virus type 1 and identification of heparin-binding PRV glycoproteins. *J. Gen. Virol.* **1990**, *71*, 1221–1225.

45. Jacquet, A.; Haumont, M.; Chellun, D.; Massaer, M.; Tufaro, F.; Bollen, A.; Jacobs, P. The varicella zoster virus glycoprotein B (gB) plays a role in virus binding to cell surface heparan sulfate proteoglycans. *Virus Res.* **1998**, *53*, 197–207.

46. Kawaura, T.; Inagaki, M.; Karita, S.; Kato, M.; Nishikawa, S.; Kashimura, N. Recognition of receptor lipopolysaccharides by spike G protein of bacteriophage phiX174. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 1993–7.

47. Sandulache, R.; Prehm, P.; Kamp, D. Cell wall receptor for bacteriophage Mu G(+). *J. Bacteriol.* **1984**, *160*, 299–303.

48. Ito, T.; Suzuki, Y.; Mitnaul, L.; Vines, A.; Kida, H.; Kawaoka, Y. Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. *Virology* **1997**, *227*, 493–499.

49. Rogers, G.N.; Paulson, J.C. Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **1983**, *127*, 361–373.

50. Russell, R.J.; Stevens, D.J.; Haire, L.F.; Gamblin, S.J.; Skehel, J.J. Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj. J.* **2006**, *23*, 85–92.

51. Suzuki, Y.; Ito, T.; Suzuki, T.; Holland, R.E.; Chambers, T.M.; Kiso, M.; Ishida, H.; Kawaoka, Y. Sialic Acid Species as a Determinant of the Host Range of Influenza A Viruses. *J. Virol.* **2000**, *74*, 11825–11831.

52. Long, J.S.; Mistry, B.; Haslam, S.M.; Barclay, W.S. Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* **2019**, *17*, 67–81.

53. Suzuki, Y.; Nakao, T.; Ito, T.; Watanabe, N.; Toda, Y.; Guiyun, X.; Suzuki, T.; Kobayashi, T.; Kimura, Y.; Yamada, A.; et al. Structural determination of gangliosides that bind to influenza A, B, and C viruses by an improved binding assay: Strain-specific receptor epitopes in sialo-sugar chains. *Virology* **1992**, *189*, 121–131.

54. Xu, G.; Suzuki, T.; Tahara, H.; Kiso, M.; Hasegawa, A.; Suzuki, Y. Specificity of sialyl-sugar chain mediated recognition by the hemagglutinin of human influenza B virus isolates. *J. Biochem.* **1994**, *115*, 202–207.

55. Harms, G.; Reuter, G.; Corfield, A.P.; Schauer, R. Binding specificity of influenza C-virus to variably O-acetylated glycoconjugates and its use for histochemical detection of N-acetyl-9-O-acetylneuraminic acid in mammalian tissues. *Glycoconj. J.* **1996**, *13*, 621–630.

56. Herrler, G.; Szepanski, S.; Schultze, B. 9-O-acetylated sialic acid, a receptor determinant for influenza C virus and coronaviruses. *Behring Inst. Mitt.* **1991**, 177–84.

57. Rogers, G.N.; Herrler, G.; Paulson, J.C.; Klenk, H.D. Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. *J. Biol. Chem.* **1986**, *261*, 5947–51.

58. Hause, B.M.; Ducatez, M.; Collin, E.A.; Ran, Z.; Liu, R.; Sheng, Z.; Armien, A.; Kaplan, B.; Chakravarty, S.; Hoppe, A.D.; et al. Isolation of a Novel Swine Influenza Virus from Oklahoma in 2011 Which Is Distantly Related to Human Influenza C Viruses. *PLoS Pathog.* **2013**, *9*.

59. Song, H.; Qi, J.; Khedri, Z.; Diaz, S.; Yu, H.; Chen, X.; Varki, A.; Shi, Y.; Gao, G.F. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLoS Pathog.* **2016**, *12*, e1005411.

60. Hellebo, A.; Vilas, U.; Falk, K.; Vlasak, R. Infectious Salmon Anemia Virus Specifically Binds to and Hydrolyzes 4-O-Acetylated Sialic Acids. *J. Virol.* **2004**, *78*, 3055–3062.

61. Amonsen, M.; Smith, D.F.; Cummings, R.D.; Air, G.M. Human Parainfluenza Viruses hPIV1 and hPIV3 Bind Oligosaccharides with 2-3-Linked Sialic Acids That Are Distinct from Those Bound by H5 Avian Influenza Virus Hemagglutinin. *J. Virol.* **2007**, *81*, 8341–8345.

62. Thammawat, S.; Sadlon, T.A.; Hallsworth, P.G.; Gordon, D.L. Role of Cellular Glycosaminoglycans and Charged Regions of Viral G Protein in Human Metapneumovirus Infection. *J. Virol.* **2008**, *82*, 11767–11774.

63. Yuan, P.; Thompson, T.B.; Wurzburg, B.A.; Paterson, R.G.; Lamb, R.A.; Jardetzky, T.S. Structural studies of the parainfluenza virus 5 hemagglutinin-neuraminidase tetramer in complex with its receptor, sialyllactose. *Structure* **2005**, *13*, 803–815.

64. Merz, D.C.; Wolinsky, J.S. Biochemical features of mumps virus neuraminidases and their relationship with pathogenicity. *Virology* **1981**, *114*, 218–227.

65. J C Paulson, J E Sadler, and R.L.H. Restoration of specific myxovirus receptors to asialoerythrocytes by incorporation of sialic acid with pure sialyltransferases. *J. Biol. Chem* **1979**, 2120–2124.

66. Reyes-Leyva, J.; Hernández-Jáuregui, P.; Montaño, L.F.; Zenteno, E. The porcine paramyxovirus LPM specifically recognizes sialyl (alpha 2,3) lactose-containing structures. *Arch. Virol.* **1993**, *133*, 195–200.

67. Reyes-Leyva, J.; Espinosa, B.; Santos, G.; Zenteno, R.; Hernández, J.; Vallejo, V.; Zenteno, E. Purification and

characterization of the Hemagglutinin-Neuraminidase of Porcine rubulavirus LPMV. *Glycoconj. J.* **1999**, *16*, 517–522.

68. Krusat, T.; Streckert, H.J. Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells. *Arch. Virol.* **1997**, *142*, 1247–1254.

69. Markwell, M.A.; Svennerholm, L.; Paulson, J.C. Specific gangliosides function as host cell receptors for Sendai virus. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 5406–10.

70. Wu, Z.; Miller, E.; Agbandje-McKenna, M.; Samulski, R.J. 2,3 and 2,6 N-Linked Sialic Acids Facilitate Efficient Binding and Transduction by Adeno-Associated Virus Types 1 and 6. *J. Virol.* **2006**, *80*, 9093–9103.

71. Kaludov, N.; Brown, K.E.; Walters, R.W.; Zabner, J.; Chiorini, J.A. Adeno-Associated Virus Serotype 4 (AAV4) and AAV5 Both Require Sialic Acid Binding for Hemagglutination and Efficient Transduction but Differ in Sialic Acid Linkage Specificity. *J. Virol.* **2001**, *75*, 6884–6893.

72. Walters, R.W.; Yi, S.M.P.; Keshavjee, S.; Brown, K.E.; Welsh, M.J.; Chiorini, J.A.; Zabner, J. Binding of Adeno-associated Virus Type 5 to 2,3-Linked Sialic Acid is Required for Gene Transfer. *J. Biol. Chem.* **2001**, *276*, 20610–20616.

73. Summerford, C.; Samulski, R.J. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* **1998**, *72*, 1438–45.

74. Schmidt, M.; Chiorini, J.A. Gangliosides Are Essential for Bovine Adeno-Associated Virus Entry. *J. Virol.* **2006**, *80*, 5516–5522.

75. Thacker, T.C.; Johnson, F.B. Binding of bovine parvovirus to erythrocyte membrane sialylglycoproteins. *J. Gen. Virol.* **1998**, *79 (Pt 9)*, 2163–9.

76. Tresnan, D.B.; Southard, L.; Weichert, W.; Sgro, J.Y.; Parrish, C.R. Analysis of the cell and erythrocyte binding activities of the dimple and canyon regions of the canine parvovirus capsid. *Virology* **1995**, *211*, 123–132.

77. Barbis, D.P.; Chang, S.F.; Parrish, C.R. Mutations adjacent to the dimple of the canine parvovirus capsid structure affect sialic acid binding. *Virology* **1992**, *191*, 301–308.

78. Nam, H.J.; Gurda-Whitaker, B.; Wand, Y.G.; Ilaria, S.; McKenna, R.; Mehta, P.; Alvarez, R.A.; Agbandje-McKenna, M. Identification of the sialic acid structures recognized by minute virus of mice and the role of binding affinity in virulence adaptation. *J. Biol. Chem.* **2006**, *281*, 25670–25677.

79. Stoner, G.D.; Williams, B.; Kniazeff, A.; Shimkin, M.B. Effect of neuraminidase pretreatment on the susceptibility of normal and transformed mammalian cells to bovine enterovirus 261. *Nature* **1973**, *245*, 319–320.

80. Tavakkol, A.; Burness, A.T.H. Evidence for a Direct Role for Sialic Acid in the Attachment of Encephalomyocarditis Virus to Human Erythrocytes. *Biochemistry* **1990**, *29*, 10684–10690.

81. Nokhbeh, M.R.; Hazra, S.; Alexander, D.A.; Khan, A.; McAllister, M.; Suuronen, E.J.; Griffith, M.; Dimock, K. Enterovirus 70 Binds to Different Glycoconjugates Containing 2,3-Linked Sialic Acid on Different Cell Lines. *J. Virol.* **2005**, *79*, 7087–7094.

82. Yang, B.; Chuang, H.; Yang, K.D. Sialylated glycans as receptor and inhibitor of enterovirus 71 infection to DLD-1 intestinal cells. *Virol. J.* **2009**, *6*.

83. Uncapher, C.R.; Dewitt, C.M.; Colonno, R.J. The major and minor group receptor families contain all but one human rhinovirus serotype. *Virology* **1991**, *180*, 814–817.

84. Fry, E.E.; Tuthill, T.J.; Harlos, K.; Walter, T.S.; Rowlands, D.J.; Stuart, D.I. Crystal structure of equine rhinitis A virus in complex with its sialic acid receptor. *J. Gen. Virol.* **2010**, *91*, 1971–1977.

85. Nilsson, E.C.; Jamshidi, F.; Johansson, S.M.C.; Oberste, M.S.; Arnberg, N. Sialic Acid Is a Cellular Receptor for Coxsackievirus A24 Variant, an Emerging Virus with Pandemic Potential. *J. Virol.* **2008**, *82*, 3061–3068.

86. Baggen, J.; Hurdiss, D.L.; Zocher, G.; Mistry, N.; Roberts, R.W.; Slager, J.J.; Guo, H.; Van Vliet, A.L.W.; Wahedi, M.; Benschop, K.; et al. Role of enhanced receptor engagement in the evolution of a pandemic acute hemorrhagic conjunctivitis virus. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *115*, 397–402.

87. Anderson, K.; Bond, C.W. Biological properties of mengovirus: Characterization of avirulent, hemagglutination-defective mutants. *Arch. Virol.* **1987**, *93*, 31–49.

88. Zhou, L.; Lin, X.; Green, T.J.; Lipton, H.L.; Luo, M. Role of sialyloligosaccharide binding in Theiler's virus persistence. *J. Virol.* **1997**, *71*, 9701–12.

89. Dugan, A.S.; Eash, S.; Atwood, W.J. An N-Linked Glycoprotein with (2,3)-Linked Sialic Acid Is a Receptor for BK Virus. *J. Virol.* **2005**, *79*, 14442–14445.

90. Komagome, R.; Sawa, H.; Suzuki, T.; Suzuki, Y.; Tanaka, S.; Atwood, W.J.; Nagashima, K. Oligosaccharides as Receptors for JC Virus. *J. Virol.* **2002**, *76*, 12992–13000.

91. Erickson, K.D.; Garcea, R.L.; Tsai, B. Ganglioside GT1b Is a Putative Host Cell Receptor for the Merkel Cell Polyomavirus. *J. Virol.* **2009**, *83*, 10275–10279.

92. Keppler, O.T.; Herrmann, M.; Oppenländer, M.; Meschede, W.; Pawlita, M. Regulation of susceptibility and cell surface receptor for the B-lymphotropic papovavirus by N glycosylation. *J. Virol.* **1994**, *68*, 6933–9.

93. Cahan, L.D.; Singh, R.; Paulson, J.C. Sialyloligosaccharide receptors of binding variants of polyoma virus. *Virology* **1983**, *130*, 281–289.

94. Neu, U.; Woellner, K.; Gauglitz, G.; Stehle, T. Structural basis of GM1 ganglioside recognition by simian virus 40. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5219–5224.

95. Tsai, B.; Gilbert, J.M.; Stehle, T.; Lencer, W.; Benjamin, T.L.; Rapoport, T.A. Gangliosides are receptors for murine polyoma virus and SV40. *EMBO J.* **2003**, *22*, 4346–55.

96. Carter, G.C.; Law, M.; Hollinshead, M.; Smith, G.L. Entry of the vaccinia virus intracellular mature virion and its interactions with glycosaminoglycans. *J. Gen. Virol.* **2005**, *86*, 1279–1290.

97. Eaton, B.T.; Crameri, G.S. The site of bluetongue virus attachment to glycophorins from a number of animal erythrocytes. *J. Gen. Virol.* **1989**, *70*, 3347–3353.

98. Delorme, C.; Brussow, H.; Sidoti, J.; Roche, N.; Karlsson, K.-A.; Neeser, J.-R.; Teneberg, S. Glycosphingolipid Binding Specificities of Rotavirus: Identification of a Sialic Acid-Binding Epitope. *J. Virol.* **2001**, *75*, 2276–2287.

99. Helander, A.; Silvey, K.J.; Mantis, N.J.; Hutchings, A.B.; Chandran, K.; Lucas, W.T.; Nibert, M.L.; Neutra, M.R. The Viral 1 Protein and Glycoconjugates Containing 2-3-Linked Sialic Acid Are Involved in Type 1 Reovirus Adherence to M Cell Apical Surfaces. *J. Virol.* **2003**, *77*, 7964–7977.

100. Barton, E.S.; Connolly, J.L.; Forrest, J.C.; Chappell, J.D.; Dermody, T.S. Utilization of sialic acid as a coreceptor enhances reovirus attachment by multistep adhesion strengthening. *J. Biol. Chem.* **2001**, *276*, 2200–2211.

101. Gentsch, J.R.; Pacitti, A.F. Effect of neuraminidase treatment of cells and effect of soluble glycoproteins on type 3 reovirus attachment to murine L cells. *J. Virol.* **1985**, *56*, 356–64.

102. Dormitzer, P.R.; Sun, Z.-Y.J.; Blixt, O.; Paulson, J.C.; Wagner, G.; Harrison, S.C. Specificity and Affinity of Sialic Acid Binding by the Rhesus Rotavirus VP8* Core. *J. Virol.* **2002**, *76*, 10512–10517.

103. Haselhorst, T.; Fleming, F.E.; Dyason, J.C.; Hartnell, R.D.; Yu, X.; Holloway, G.; Santegoets, K.; Kiefel, M.J.; Blanchard, H.; Coulson, B.S.; et al. Sialic acid dependence in rotavirus host cell invasion. *Nat. Chem. Biol.* **2009**, *5*, 91–93.

104. Svensson, L. Group C rotavirus requires sialic acid for erythrocyte and cell receptor binding. *J. Virol.* **1992**, *66*, 5582–5.

105. Isa, P.; Arias, C.F.; López, S. Role of sialic acids in rotavirus infection. *Glycoconj. J.* 2006, *23*, 27–37.

106. Ludert, J.E.; Mason, B.B.; Angel, J.; Tang, B.; Hoshino, Y.; Feng, N.; Vo, P.T.; Mackow, E.M.; Ruggeri, F.M.; Greenberg, H.B. Identification of mutations in the rotavirus protein VP4 that alter sialic-acid-dependent infection. *J. Gen. Virol.* **1998**, *79*, 725–729.

107. Crublet, E.; Andrieu, J.P.; Vivès, R.R.; Lortat-Jacob, H. The HIV-1 envelope glycoprotein gp120 features four heparan sulfate binding domains, including the co-receptor binding site. *J. Biol. Chem.* **2008**, *283*, 15193–15200.

108. Conti, C.; Superti, F.; Tsiang, H. Membrane carbohydrate requirement for rabies virus binding to chicken embryo related cells. *Intervirology* **1986**, *26*, 164–168.

109. Superti, F.; Girmenta, C.; Seganti, L.; Orsi, N. Role of sialic acid in cell receptors for vesicular stomatitis virus. *Acta Virol.* **1986**, *30*, 10–8.

110. Langereis, M.A.; Zeng, Q.; Gerwig, G.J.; Frey, B.; Itzstein, M. von; Kamerling, J.P.; de Groot, R.J.; Huizinga, E.G. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15897–15902.

111. Byrnes, A.P.; Griffin, D.E. Binding of Sindbis virus to cell surface heparan sulfate. *J. Virol.* **1998**, *72*, 7349–56.

Chapter 2

Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity

Mark J.G.Bakkers[1], Yifei Lang[1], Louris J.Feitsma[2], Ruben J.G.Hulswit[1], Stefanie A.H.de Poot[1], Arno L.W.van Vliet[1], Irina Margine[1], Jolanda D.F.de Groot-Mijnes[3], Frank J.M.van Kuppeveld[1], Martijn A.Langereis[1], Eric G.Huizinga[2], and Raoul J.de Groot[1]

[1]Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, the Netherlands

[2]Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Sciences, Utrecht University, 3584 CH Utrecht, the Netherlands

[3]Department of Medical Microbiology, University Medical Center Utrecht, 3584 CX Utrecht, the Netherlands

Cell Host & Microbe. 2017 Mar 8;21(3):356-366

Abstract

Human beta1-coronavirus (β1CoV) OC43 emerged relatively recently through a single zoonotic introduction. Like related animal β1CoVs, OC43 uses 9-*O*-acetylated sialic acid as receptor determinant. β1CoV receptor binding is typically controlled by attachment/fusion spike protein S and receptor-binding/receptor-destroying hemagglutinin-esterase protein HE. We show that following OC43's introduction into humans, HE-mediated receptor binding was selected against and ultimately lost through progressive accumulation of mutations in the HE lectin domain. Consequently, virion-associated receptor-destroying activity toward multivalent glycoconjugates was reduced and altered such that some clustered receptor populations are no longer cleaved. Loss of HE lectin function was also observed for another respiratory human coronavirus, HKU1. This thus appears to be an adaptation to the sialoglycome of the human respiratory tract and for replication in human airways. The findings suggest that the dynamics of virion-glycan interactions contribute to host tropism. Our observations are relevant also to other human respiratory viruses of zoonotic origin, particularly influenza A virus.

Introduction

Coronaviruses (CoVs), long considered of high veterinary impact exclusively, are now generally recognized as zoonotic threats of pandemic potential in consequence of the 2002/2003 outbreak of severe acute respiratory syndrome (SARS) and the emergence of Middle East respiratory syndrome (MERS) in 2012 [1]. SARS-CoV was contained within 3 months after its discovery, while MERS-CoV causes a classical zoonotic infection with limited human-to-human spread and, as yet, is incapable of sustained community transmission [2,3]. However, four other respiratory coronaviruses - alphacoronaviruses NL63 and 229E and lineage A betacoronaviruses OC43 and HKU1 - successfully breached the species barrier and are currently maintained in the human population worldwide through continuous circulation [4,5]. Conceivably, the study of these genuine human coronaviruses (HCoVs) may yield clues to what is required for viral adaptation to the human host and thereby increase our understanding of the probabilities and risks of coronavirus cross-species transmission.

OC43 and HKU1, while generally associated with benign common colds in healthy immunocompetent individuals, may cause significant morbidity and even mortality in the frail [6,7]. OC43, the best-studied HCoV, apparently arose relatively recently, 120 to 70 years ago, with the most recent common ancestor of all extant OC43 variants dating to the 1950s [8–10]. OC43 groups in the species Betacoronavirus-1 (β1CoV), together with highly related viruses from ruminants (bovine coronavirus, BCoV), swine (porcine hemagglutinating encephalomyelitis virus, PHEV), equines (equine coronavirus, ECoV), leporids (rabbit coronavirus HKU14, RbCoV), and canines (canine respiratory coronavirus, CRCoV) ([5] ; Figure 1A). The extraordinary radiation of β1CoVs might be explained from their receptor usage, as they attach to 9-*O*-acetylated sialic acids (9-*O*-Ac-Sias) [11], i.e., glycan components common in mammals and birds [12]. Paradoxically, however, most β1CoVs, including OC43, have very narrow host ranges. They form distinct monophyletic clades congruent with host selectivity, and phylogenetic evidence strongly argues against recurrent inter-species transmissions ([8,10]; see also Figure 1A). With the emergence of OC43 seemingly sparked by a singular one-time zoonotic event, the founder virus likely possessed unique traits already to allow efficient infection of, and transmission among, humans. In turn, these traits and subsequent adaptations to the new niche must have closed the door on reintroduction of the virus into animals. Despite the close genetic relationship between OC43 and BCoV-Mebus, particularly (96.6% identity across their entire genomes), their high-prevalence endemicity/enzooticity, and the scale and frequency of interactions between their host species, there is no evidence of OC43 spreading to cattle nor of BCoV (or any other β1CoV) efficiently spreading into humans (Figure 1A).

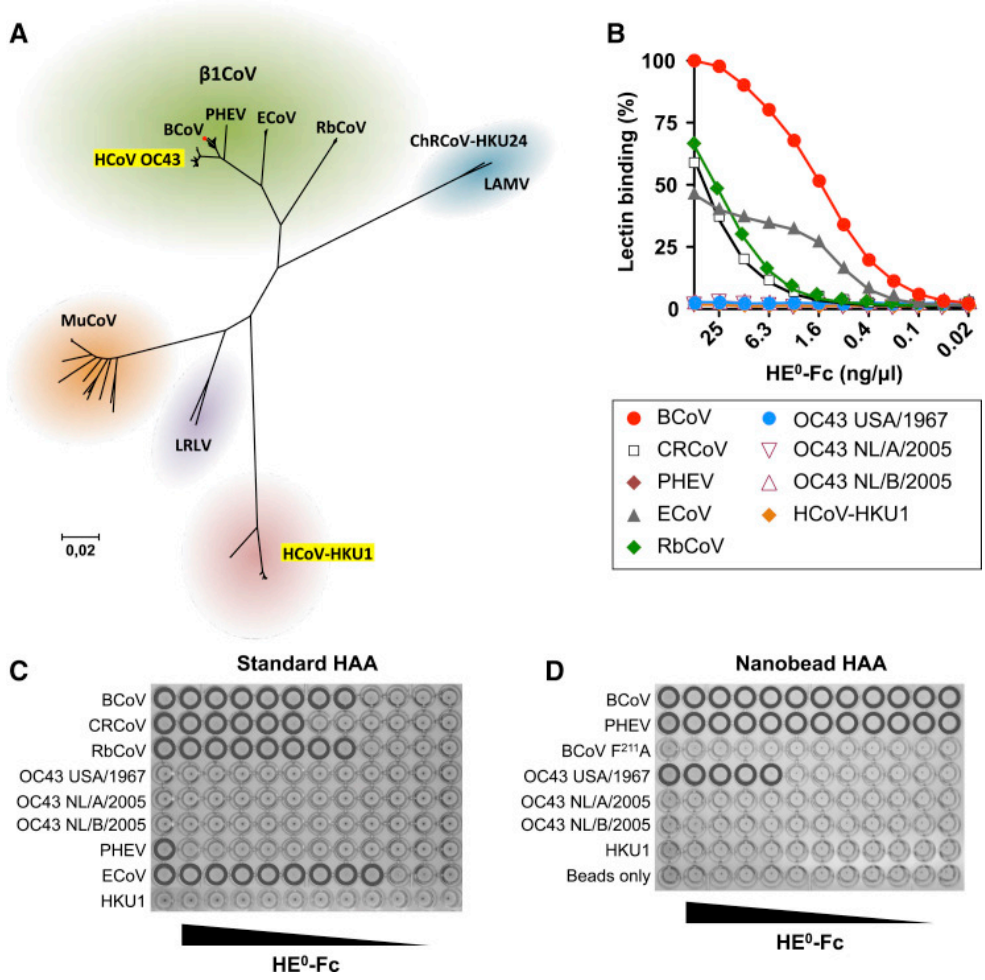


Figure 1. Loss of HE-Mediated Receptor Binding in Human Betacoronaviruses. (A) Evolutionary relationships among lineage A betacoronaviruses. Neighbor-joining tree based on βCoV lineage A ORF1b sequences in the NCBI database (n = 206), with 100% bootstrap support for all major branches. Evolutionary distances were computed using the Maximum Composite Likelihood method in MEGA6 [13]. The positions of human coronaviruses OC43 and HKU1 are highlighted relative to those of animal lineage A betacoronaviruses, including the various β1CoV subspecies and classical mouse hepatitis virus-type murine coronavirus (MuCoV). ChRCoV-HKU24, Chinese rat coronavirus HKU24 [14]; LAMV, Longquan Aa mouse coronavirus [15]; LRLV, Longquan RI rat coronavirus [15]. The position of CRCoV, a recent split-off of BCoV, is indicated by a red dot. (See also Figure S1.) (B) HE^0-Fc lectins (2-fold serial dilutions, starting at 50 ng/µL) were compared by sp-LBA for relative binding to BSM (with 50 ng/µL BCoV-HE^0-Fc set at 100%). (C) Conventional HAA with 2-fold serial dilutions of HE^0-Fc fusion proteins of β1CoV members and HCoV-HKU1 (starting at 25 ng/well). Wells positive for hemagglutination are encircled. (D) High-sensitivity nanobead HAA. Non-complexed nanobeads ("beads only") and nanobeads complexed with lectin-inactive mutant BCoV HE^0 $F^{211}A$ were included as negative controls.

β1CoV attachment involves two distinct types of surface projections: large 20-nm "spikes" comprised of homotrimers of the class I fusion protein S and stubby 8-nm protrusions that are homodimeric

assemblies of the hemagglutinin-esterase protein HE [5]. The S protein, common to all coronaviruses, mediates receptor binding and fusion of the viral and host membranes [16]. HEs are found in toroviruses and in influenza C and D viruses, but among coronaviruses, only in lineage A betacoronaviruses [5,11,17,18]. HE monomers have a bimodular structure with a carbohydrate-binding ("lectin") domain appended to an enzymatically active sialate-*O*-acetylesterase ("esterase") domain [19–21]. Typically, in β1CoVs, both the spikes and HEs bind 9-*O*-Ac-Sias, while the HE esterase domain promotes virus elution through receptor destruction. In consequence of the opposing activities of receptor binding and receptor destruction, β1CoV attachment to sialoglycans is dynamic and reversible. Thus, dead-end binding of virions to decoy receptors in oropharyngeal, respiratory, and gastrointestinal mucus may be prevented. Moreover, in infected host cells, HE-mediated receptor destruction is essential for efficient release of viral progeny [22].

Here we present a comprehensive structure-function study of OC43 HE. We demonstrate that over decades after OC43's introduction, its evolution was marked by a progressive loss of HE receptor-binding activity through the accumulation of select mutations in the HE lectin domain. The effect of these mutations on the organization of the carbohydrate-binding site (CBS) and on receptor binding is explained from the crystal structure of the BCoV HE-receptor complex [21] and visualized directly by the structure of a contemporary OC43 HE solved to 2.45Å resolution. Evidence is provided that loss of HE receptor-binding activity resulted in a reduction of virion-associated esterase activity toward multivalent clustered substrates. We propose that inactivation of the HE lectin domain altered the balance between virion binding and esterase-mediated virion elution, apparently as an adaptation to the sialoglycome of the human respiratory tract, and we speculate that this may be a contributing factor to host selectivity. This view is supported by our observation that the HE of HCoV-HKU1 also lost its receptor-binding properties. A mechanism is proposed in which accessibility of receptors to destruction depends on HE lectin function in relation to S-HE size differences and in which the balance between attachment and catalysis-driven virion elution is a determinant of host tropism.

Results and Discussion

Lectin Properties of β1CoV HEs

A comprehensive set of β1CoV HEs, expressed as esterase-inactive Fc-fusion proteins (HE^0-Fc), was compared for their Sia-binding properties. In accordance with previous findings [23], the HEs of most animal β1CoVs bound to bovine submaxillary mucin (BSM) in a 9-*O*-Ac-Sia-dependent fashion in solid-phase lectin-binding assays (sp-LBA; Figure 1B). Remarkably, however, those of PHEV strain VW572 and prototypic OC43 strain USA/1967 (also known as ATCC-VR759; [24]) showed no

detectable binding. In hemagglutination assays (HAA), more sensitive than sp-LBA, PHEV HE^0 tested positive, albeit weakly (Figure 1C). OC43 USA/1967 HE^0, however, did not hemagglutinate. To augment HAA sensitivity through multivalency-driven high-avidity binding, we complexed HE^0-Fc chimeras to protein A-coated nanobeads (Figure 1D). For PHEV HE^0, sensitivity was increased 250-fold as compared to the standard assay. Under these conditions, modest, but reproducible, hemagglutination by OC43 USA/1967 HE^0 was detected (Figure 1D). Apparently, it has lost most, though not all, of its lectin function.

Our observations prompted the question of whether loss of HE lectin activity is a strain-specific trait resulting from adaptation to in vitro propagation or a characteristic also shared by naturally occurring OC43 viruses. We therefore RT-PCR-amplified HE genes from more recent sputum-derived OC43 strains (respiratory season 2005). The encoded proteins OC43/NL/A/2005 and OC43/NL/B/2005 HE^0, representative for the two major HE lineages in OC43 (designated "A" and "B," Figure S1 and vide infra), did not show any detectable binding by sp-LBA or nanobead HAA (Figures 1B–1D). Apparently, loss of affinity for 9-*O*-Ac-Sias as seen for OC43/USA/1967 HE is not an artifact. In fact, the data suggest that HEs of contemporary OC43 variants have lost receptor-binding activity altogether.

Lectin Function of OC43 HE Impeded by a Combination of Mutations

The HE ectodomains of BCoV strain Mebus and OC43 USA/1967, each 365 residues in length, differ at 18 positions only (Figures 2A and 2B). To identify the differences accounting for loss of lectin activity in OC43 USA/1967 HE, we introduced OC43-specific substitutions in BCoV-Mebus HE^0-Fc, in sets and individually, and tested for a reduction in lectin activity by HAA (Figure S2A). In a complementary approach, OC43 HE residues were systematically replaced by BCoV HE orthologs (Figures S2B–S2D). The results show that the mutations in the esterase and MP domains of OC43 USA/1967 HE (Figures 2A–2C; Figure S2A) do not affect ligand binding. In fact, loss of 9-*O*-Ac-Sia binding can be attributed to four out of eight substitutions in the lectin domain ($T^{114}N$, $R^{177}P$, $E^{178}Q$, and $F^{247}L$; Figures S2A–S2D). Combined replacement of these residues in OC43 USA/1967 HE^0 by BCoV orthologs restored binding affinity to that of BCoV-Mebus HE (Figure 2C). Individual replacement of these residues in the context of a restored OC43 USA/1967 HE ("TREF") showed that each mutation affects receptor binding to more ($T^{114}N$, $E^{178}Q$) or lesser ($R^{177}P$, $F^{247}L$) extent (Figure 2C).

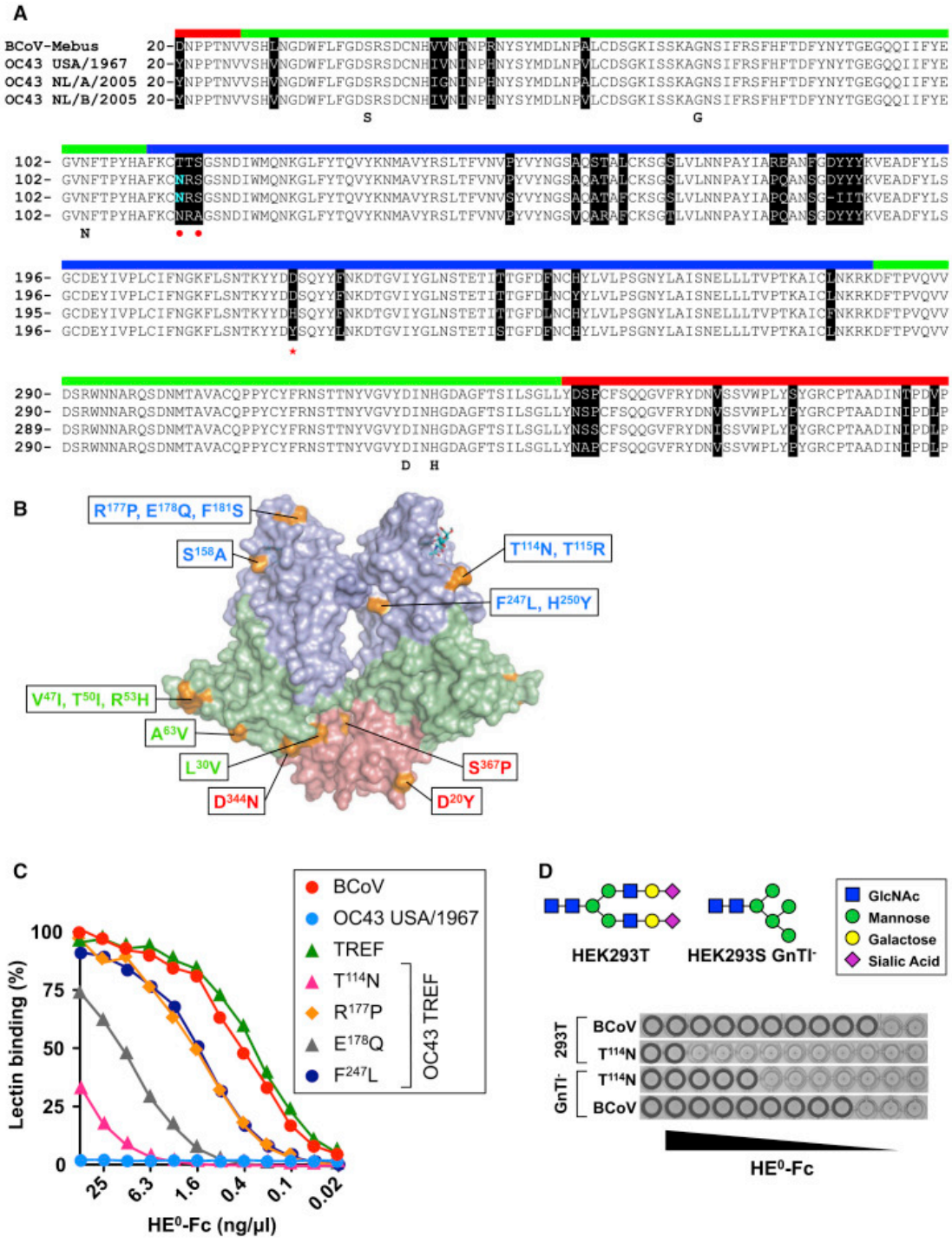


Figure 2 Loss of OC43 USA/1967 HE^0 Lectin Affinity Attributed to a Combination of Four Mutations. (A) Alignment of BCoV, OC43 USA/1967, OC43 NL/A/2005, and OC43 NL/B/2005 HE with sequences color coded by domain (membrane-proximal domain, red; esterase domain, green; lectin domain, blue). Residues crucial for esterase activity (SGNDH) are annotated. Amino acid differences are marked in black. N^{114} substituting for Thr is colored in cyan, and the resulting N-linked glycosylation site (NRS) marked with red dots. D^{220} is marked with a red asterisk. (B) Overall structure of BCoV HE (PDB: 3CL5) with mutations in OC43 USA/1967 HE indicated. Domain coloring as in (A). (C) Comparison of binding affinities of BCoV HE^0, OC43 USA/1967 HE^0, and derivatives by sp-LBA. BCoV, BCoV-Mebus HE^0; OC43 TREF, OC43 USA/1967 HE^0 with N^{114}, P^{177}, Q^{178} and L^{247}

replaced by BCoV-Mebus HE orthologs; $T^{114}N$, $R^{177}P$, $E^{178}Q$, and $F^{247}L$, TREF derivatives with indicated residues re-converted to the autologous OC43 orthologs. (D) Conventional HAA with BCoV-Mebus HE^0 $T^{114}N$ expressed in HEK293T or HEK293S GnTI$^-$ cells. (See also Figure S2.)

The impact of these mutations on 9-*O*-Ac-Sia binding can be understood from the crystal structure of the BCoV-Mebus HE-receptor complex [21]. The BCoV HE receptor-binding region is comprised of six surface loops, five grafted on the beta-sandwich core of the lectin domain and one emanating from the esterase domain (Figure S3A). The actual carbohydrate-binding site consists of a deep hydrophobic pocket P1 and a more shallow hydrophobic depression P2 that accommodate the methyl groups of the Sia-9-*O*- and −5-*N*-acetyl moieties, respectively (Figure 4E). The side chain of F^{211} (a residue in the β12/β13 β-hairpin) separates P1 from P2 and, in the HE-receptor complex, intercalates between the Sia-acetyl groups. Ligand binding, thus largely based on shape complementarity and hydrophobic interactions, is stabilized by extensive protein-sugar hydrogen bonding, particularly involving the β7–β10 β-hairpin, through main chain atoms of L^{212} and N^{214} to the sialate-5-*N*-acyl amide and the Sia carboxylate, respectively, and through the side-chain hydroxyl of S^{213} to the sialate-C8-hydroxyl and -C9-acyl oxygen. Y^{184} in the β5-β6 loop (residues 176–185) is particularly important for ligand binding as its side-chain hydroxyl group hydrogen bonds with the sialate-9-*O*-acetyl carbonyl, while its aromatic ring together with the side chains of F^{211}, L^{266}, and L^{267} walls the P1 pocket that is key to receptor recognition (Figure 4E).

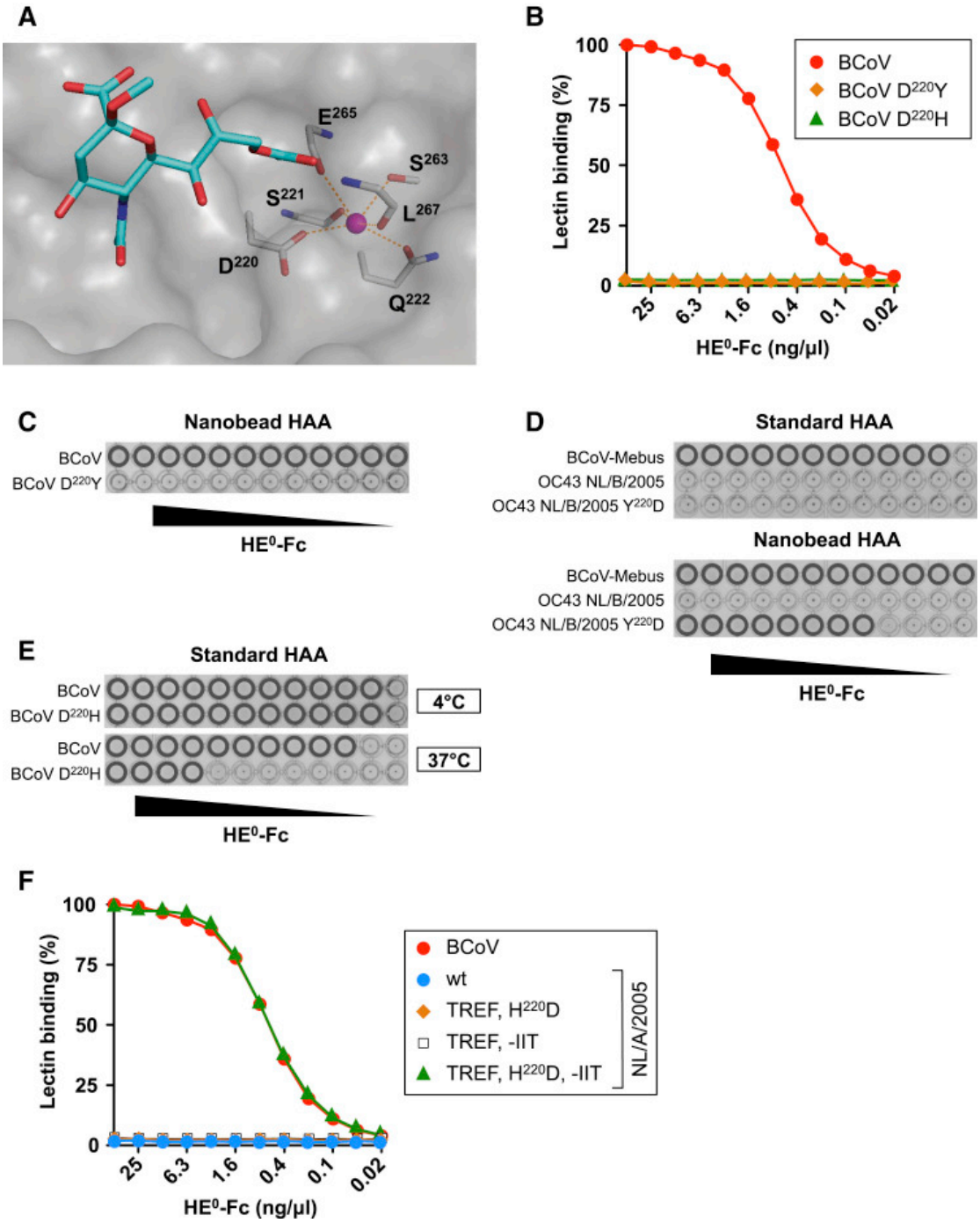


Figure 3 Complete Loss of HE Lectin Function during OC43 Evolution due to Progressive Accumulation of Mutations. (A) Close up of the BCoV-Mebus HE CBS in complex with α-Neu5,9Ac$_2$2Me (in sticks). Residues comprising the metal-binding site (MBS) are also shown in sticks, the potassium ion is shown as a magenta sphere, and interactions between the metal ion and coordinating amino acid residues as red dashed lines. (B) Disruption of the MBS in BCoV HE leads to loss of lectin affinity. sp-LBA as in Figure 1B. (C) Disruption of the BCoV HE lectin MBS through D^{220}Y substitution renders receptor binding non-detectable even by high-sensitivity nanobead HAA. (D) The Y^{220}D mutation partially restores lectin affinity of OC43 NL/B/2005 HE^0. (E) D^{220}H substitution in BCoV HE results in thermolability of receptor binding. Conventional HAA before (4°C) and after (37°C) a temperature shift up. (F) Receptor binding of OC43 NL/A/2005 HE^0 and derivatives as

determined by sp-LBA. In OC43 NL/A/2005 HE^0, N^{114}, P^{177}, Q^{178}, and L^{247} were replaced by BCoV-Mebus HE orthologs (TREF), in combination with (1) $H^{220}D$ substitution (TREF, $H^{220}D$), (2) repair of the β5-β6 loop by substituting $D^{183}YYY^{186}$ for IIT (TREF, −IIT), or (3) $H^{220}D$ and repair of the β5-β6 loop (TREF, $H^{220}D$, −IIT).

Residues of the α5-α6 loop (residues 111 – 118) do not directly contact the ligand. However, the T114N substitution in OC43 HE created a novel N-linked glycosylation site (TTS→NRS) (Figure 2A). A glycan attached to this site would extend into the CBS and potentially reduce receptor binding through steric hindrance. Indeed, the introduction of this glycosylation site in BCoV HE^0 resulted in an almost complete loss of binding as detected by sp-LBA (Figure 2C) and a 250-fold loss of apparent binding affinity as measured by HAA (Figure 2D; Figure S2A). The same mutant protein, but now expressed in N-acetyl glucosamine transferase I (GnTI)-deficient HEK293S $GnTI^-$ cells instead of HEK293T cells, showed a less pronounced loss of binding affinity (30-fold in HAA; Figure 2D). The data show that glycosylation at N^{114} indeed reduces the affinity of the HE lectin domain for 9-*O*-Ac-Sias and that this reduction is positively correlated with glycan size/complexity.

In BCoV HE, E^{178} fixates the β5-β6 loop by engaging in double hydrogen bonds with the backbone nitrogen atoms of S^{155} and A^{156} in the β4-β5 loop (Figure S3B). We offer that in OC43 HE, the replacement by Gln and consequential loss of inter-loop hydrogen bonding destabilizes the β5-β6 loop and that its increased flexibility impairs the Sia-binding site, presumably by affecting the critical P1 pocket. The consequences of the other two substitutions in OC43 USA/1967 HE are less evident, although it can be envisaged that the $R^{177}P$ replacement might again affect folding and/or stability of the β5-β6 loop and thereby indirectly the positioning of the critical Y^{184} residue. The $F^{247}L$ substitution is more difficult to explain as the affected residue locates to the core of the HE lectin domain, distal to the CBS, and the mutation apparently decreases ligand binding through indirect long-range effects. The combined data show for the early OC43 isolate USA/1967 that the reduced lectin activity of its HE resulted not from a single but from a combination of mutations.

Progressive Loss of HE Lectin Function during OC43 Evolution

In the course of this study, many full genome sequences from OC43 field variants from the US, Western Europe, and China, well documented with respect to place and date of virus sampling, became available in GenBank. This wealth of information allowed us to study the recent evolution of the OC43 HE protein. In phylograms, in accordance with the presumptive date of OC43 emergence, the HE of the USA/1967 strain is placed close to the root (Figure S1). In the decades thereafter, the HE proteins divided into three clades. Clade C, the least populous, is comprised of HEs from OC43 strains sampled in the early 1990s with no recent representatives and may have been replaced by clades A and B. Comparative sequence analysis in combination with structure-function analysis

showed that both type A and B HEs acquired several mutations in the lectin domain additional to those already present in OC43 USA/1967, which explains the loss of HE receptor binding in extant OC43 variants. Unexpectedly, in B-type HEs, but not in those of types A and C, glycosylation at N^{114} was lost again due to a $S^{116}A$ substitution (NRS→NRA). Further inspection revealed the replacement in B-type OC43 HEs of D^{220} by Y. D^{220} is part of a potassium ion-coordinating metal-binding site (MBS), a signature element of coronavirus HE lectin domains involved in the organization of the Sia-binding site (Figure 3A) [21]. The MBS is formed by main-chain carbonyl atoms of S^{221}, E^{265}, and L^{267}, together with the side chains of Q^{222}, S^{263}, and D^{220}, the latter of which is particularly important as it balances the positive charge of the metal ion. In BCoV HE^0, the D220Y mutation resulted in loss of all detectable binding to 9-*O*-Ac-Sia (Figures 3B and 3C). In turn, a $Y^{220}D$ back mutation in the type B HE of OC43 NL/B/2005 restored Sia-binding activity to levels detectable by high-sensitivity nanobead HAA (Figure 3D). Again, the structure of the BCoV HE-receptor complex provides an explanation for the detrimental effect of the mutation. As L^{266} and L^{267} are in a loop that is stabilized through the coordination of the potassium ion, disruption of the MBS might affect the folding of this loop, the positioning of the Leu residues, and thereby the structure of receptor-binding pocket P1.

Remarkably, also in A-type OC43 HEs, the MBS was targeted by substitution of D^{220}, now by His (Figure 2A). As evident from comparative sequence and phylogenetic analysis, the substitutions in A- and B-type HEs were independent events, and their selection must be ascribed to convergent evolution. Again, $D^{220}H$ mutation in BCoV HE resulted in loss of detectable Sia binding as measured by sp-LBA (Figure 3B). Unexpectedly, however, in conventional HAA routinely performed at 4°C, wild-type and mutant BCoV HEs hemagglutinated to similar extent (Figure 3E). The effect of the mutation became evident after a shift up to 37°C, upon which hemagglutination with wild-type BCoV HE remained stable while that of the mutant protein resolved (Figure 3E). Apparently, substitution of D^{220} by His is less disruptive than by Tyr, but results in thermolability of HE receptor binding, likely to cause reduced binding affinity under physiological conditions. Repair of the OC43-specific substitutions already present in the HE of the 1967 strain (Figure 2A), in conjunction with a $H^{220}D$ back substitution, did not restore lectin activity of OC43 NL/A/2005 HE (Figure 3F), suggesting the presence of at least one additional mutation to prevent receptor binding. Comparative sequence analysis revealed a major change uniformly shared by A-type HEs again in the β5-β6 loop, with four adjacent residues 183–186, Asp-Tyr-Tyr-Tyr, replaced by Ile-Ile-Thr (Figure 2A), apparently as a result of a double frameshift mutation (Figure S3C). Repair of this mutation, in combination with restorative changes at the other five sites, was required for OC43 NL/A/2005 HE to regain receptor-binding activity (Figure 3F). The impact of the β5-β6 loop mutation is easily understood

from the structure of the BCoV HE-receptor complex, as it involved loss of Y^{184}, an essential residue for ligand binding.

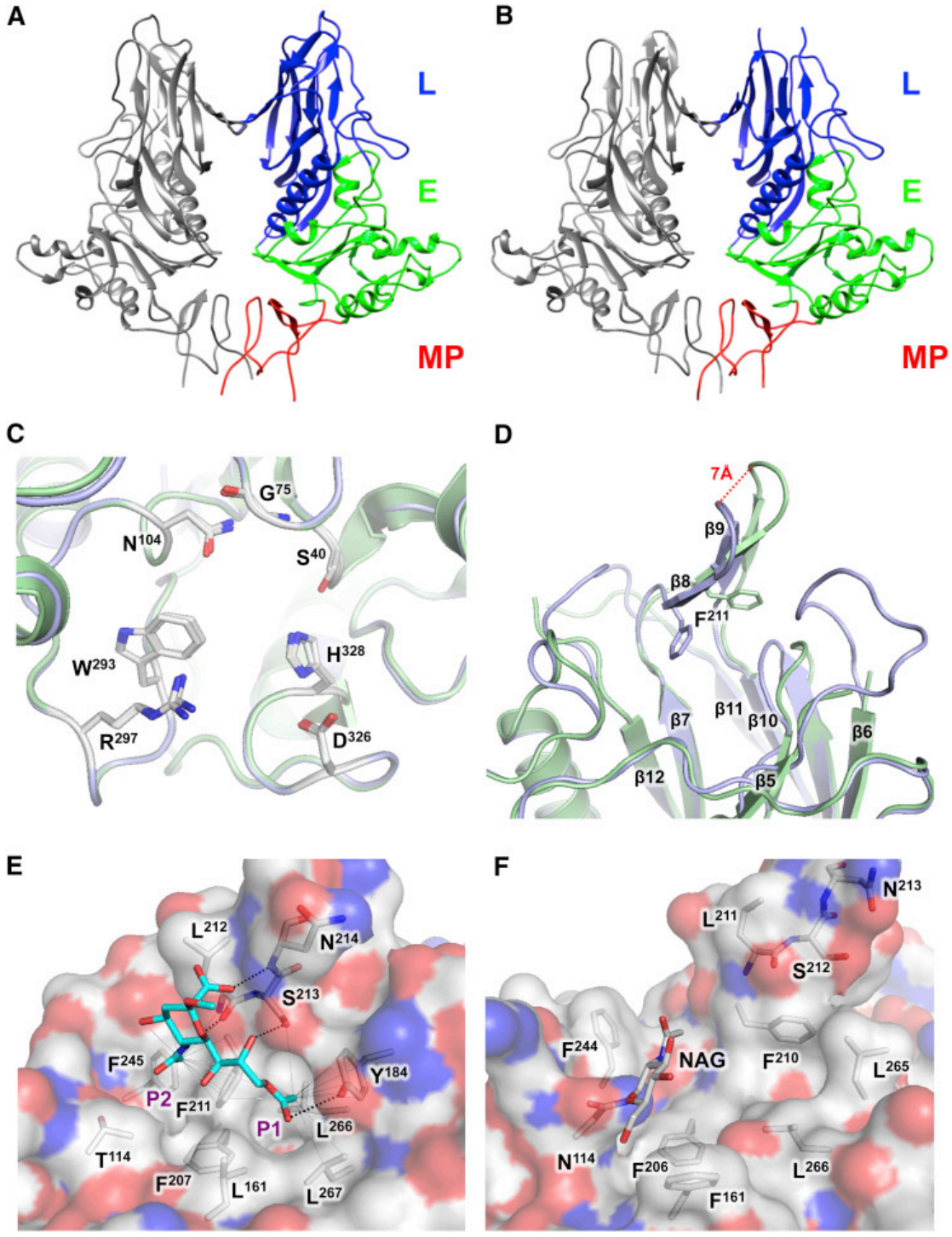


Figure 4. Crystal Structure of OC43 NL/A/2005 HE. (A and B) Side-by-side cartoon representations of the overall crystal structures of BCoV-Mebus HE (A) and OC43 NL/A/2005 HE (B). HE monomers are colored gray or by domain (as in Figure 2A). (C) BCoV and OC43 HE have identical esterase catalytic sites. Overlay of BCoV (blue)

and OC43 NL/A/2005 (green) HE esterase domains. Cartoon representations with residues crucial for activity indicated as sticks. (D) Overlay of BCoV and OC43 NL/A/2005 HE lectin domains. F^{211} indicated with sticks. (E) 9-*O*-Ac-Sia (with carbon atoms in cyan) binding in the BCoV-Mebus HE lectin CBS as observed in the crystal complex (PDB: 3CL5). Close up with contacting amino acid side chains shown in stick representation. Hydrogen bonds are shown as black dashed lines, and hydrophobic interactions with the Sia-9-*O*- and −5-*N*-methyl groups as thin gray lines. P1 and P2 indicate the pocket and hydrophobic depression, which accommodate the methyl groups of the Sia-9-*O*- and −5-*N*-acetyls, respectively. (F) Close up of the inactivated CBS of OC43 NL/A/2005 HE. Residues corresponding to those in (E) are in stick representation and colored by atom type. NAG, N-acetylglucosamine attached to N114, is shown in stick representation. (See also Figure S3 and Table S1.)

Crystal Structure of an OC43A-type HE

To assess the consequences of the combined mutations in the type A HE lectin domain, we determined the crystal structure of OC43 NL/A/2005 HE at 2.45Å resolution (for crystallographic details, see Table S1; PDB: 5N11). In overall fold, OC43 HE is highly similar to other coronavirus HEs (Figure 4). Differences in the esterase domain are limited to substitution of five surface residues, none likely to impair esterase function or activity. OC43 HE is also identical to BCoV HE in most of its lectin domain, particularly in the beta-sandwich core, but the loops shaping the receptor-binding site have changed dramatically (Figure 4D; Figure S3D). The β11-β12 and β5-β6 loops are disordered, suggestive of extensive flexibility, in the case of the latter loop in accordance with loss of E^{178}-mediated inter-loop hydrogen bonding. Those (parts of the) loops that can be visualized are reoriented and displaced with respect to their original position in BCoV HE, utterly destroying the CBS. For example the β7/β10 β-hairpin, comprising F^{211}, L^{212}, S^{213}, and N^{214} that are key to protein-ligand interaction in BCoV HE [21], has shifted by approximately 7Å (Figure 4D). The P1 pocket, arguably the most critical element of the CBS, no longer exists as (1) F^{211} is no longer at its original position and its side chain is rotated by 90°, (2) Y184 was lost as a result of the frameshift mutation in the β5-β6 loop, and (3) the side chains of L^{266} and L^{267} are reoriented apparently due to the loss of the MBS (Figures 4E and 4F). Finally, attached to N^{114} is an N-acetylglucosamine, providing formal evidence that the newly introduced glycosylation site in OC43 HEs is functional. Although the remaining residues of the sugar chain could not be visualized, the glycan would stretch across the remains of the CBS (Figure 4F).

Loss of HE Receptor Binding in HCoV-HKU1: A Case of Convergent Evolution?

OC43 and HKU1 occupy the same niche, prompting the question whether they have been subject to similar selective constraints. HCoV-HKU1 is also a lineage A betacoronavirus but distantly enough related to the β1CoVs to be assigned to a separate species [5] (Figure 1A). Like OC43, HCoV-HKU1

binds to 9-*O*-Ac-Sia receptor determinants via its spike [25], and it possesses an HE [26]. The HE of the prototypical HCoV-HKU1 2005 Hong Kong genotype A strain [27] tested negative for 9-*O*-Ac-Sia binding by sp-LBA, standard HAA, and nanobead HAA (Figures 1B–1D). Comparative sequence analysis, including genotypes A and B, of HCoV-HKU1 [27] shows that the HKU1 HE has undergone massive deletions, as a result of which most of the lectin domain was lost (Figure S4).

Loss of HE Lectin Function Reduces Receptor-Destroying Enzyme Activity toward Multivalent Substrates

In influenza C and D viruses, the hemagglutinin-esterase fusion protein is uniquely responsible for viral attachment and entry, and hence HE function is receptor binding first and foremost. The same holds for the HEs of MHV-type murine betacoronaviruses (MuCoV) that mediate virion binding to O-Ac-sialoglycan-based attachment factors [28], while S binds to the proteinaceous entry receptor CEACAM-1 [29]. At the other end of the spectrum, in OC43 and HKU1, HE's sole remaining function is receptor destruction, with virion attachment to 9-*O*-Ac-sialoglycans exclusively assigned to S. For the animal β1CoVs that retain a functional HE lectin domain, however, the role of HE is less obvious. On the one hand, HE might participate in attachment to increase virion binding avidity. On the other, its lectin domain may serve primarily to promote catalytic activity toward high-multivalency substrates by bringing the esterase in close proximity to and prolonged association with clustered glycotopes, in analogy with the carbohydrate-binding modules of cellular glycoside hydrolases [30]. In the latter case, the HE lectin CBS in the human lineage A βCoVs may have been lost to downregulate receptor-destroying activity. To study this, we compared soluble recombinant HEs of BCoV, OC43, and HKU1 for their relative enzymatic activity toward the monovalent substrate p-nitrophenol acetate (pNPA) and the multivalent substrate BSM. The latter glycoconjugate carries hundreds of O-linked glycans clustered in such densities to give the protein a filamentous bottlebrush appearance typical for mucins [31]. Whereas all three HEs showed comparable specific activities when assayed with pNPA (Figure 5 A), those of OC43 and HKU1 were more than 250-fold less active than BCoV HE in receptor-destruction assays with BSM (Figure 5B). Inactivation of the BCoV HE lectin CBS through a $F^{211}A$ substitution reduced its esterase activity toward BSM to that of wild-type OC43 HE. Conversely, restoration of the OC43 HE lectin domain increased esterase activity to that of BCoV HE (see "TREF," Figure 5B). Thus, BCoV and HCoV HEs indeed differ in their reactivity toward clustered receptors in consequence of the respective conservation or loss of lectin function.

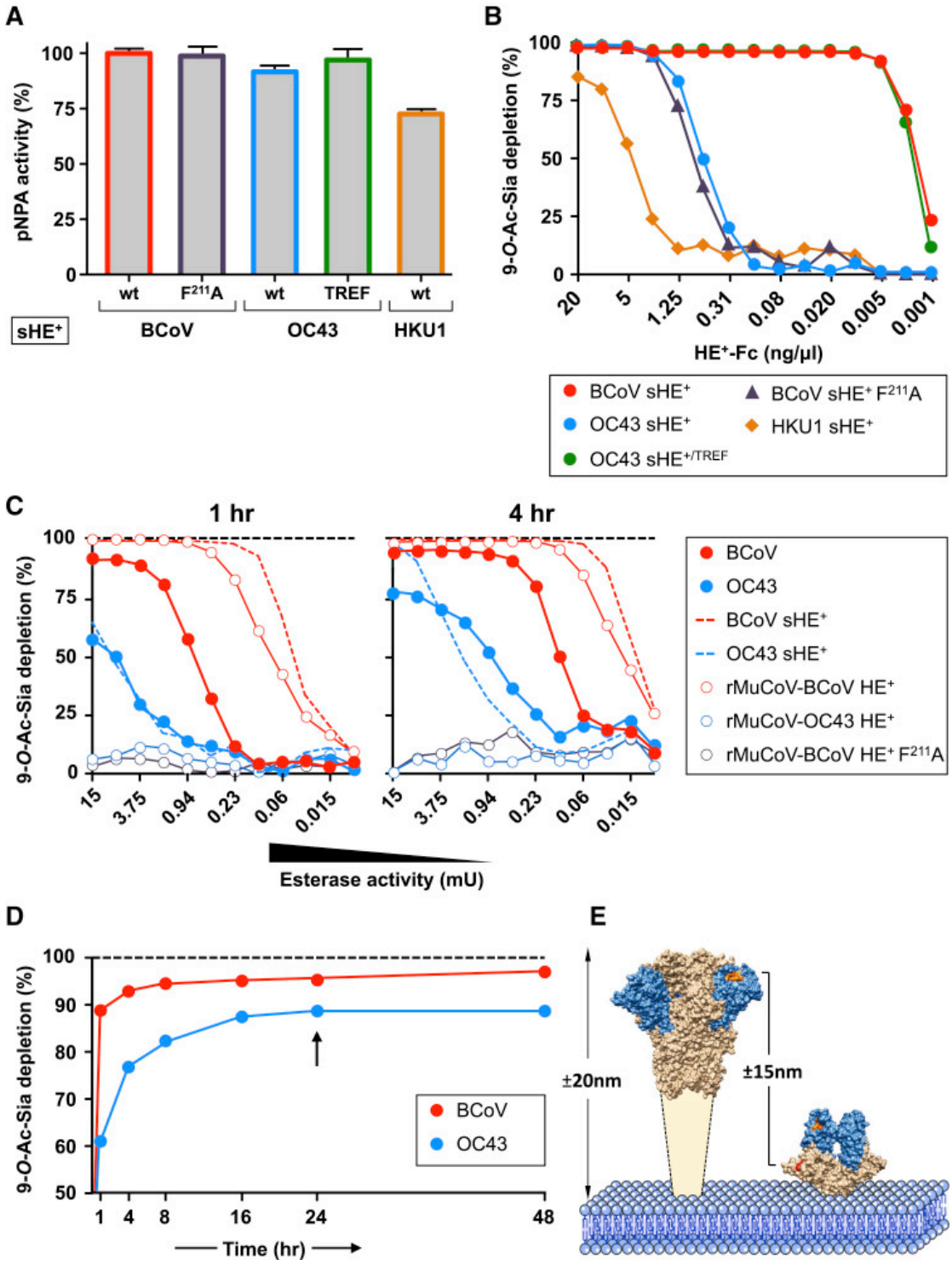


Figure 5. Loss of HE Lectin-Mediated Receptor Binding Alters Sialate-9-*O*-Acetylesterase Receptor, Destroying Activity toward Multivalent Substrates. (A) Esterase activity of soluble recombinant HE+-Fc fusion proteins (sHE$^+$) toward monovalent substrate pNPA. Enzymatic activity is shown as percentage of BCoV HE wild-type activity. WT, sHE$^+$ with wild-type HE ectodomains of BCoV-Mebus, OC43 USA/1967, or HCoV-HKU1 as indicated; F211A, BCoV-Mebus sHE$^+$ derivative with the lectin CBS inactivated through a F^{211}A substitution [21]; TREF, OC43 USA/1967 HE with the lectin CBS repaired (as in Figure 2C). Data are represented as mean ± SEM. (B) Loss of sHE$^+$-mediated receptor binding results in reduced esterase activity toward high-multivalency substrates as

determined by on-the-plate *O*-Ac-Sia depletion assay. BSM, coated on maxisorp plates, was incubated for 1 hr with 2-fold serial dilutions of sHE^{+}. Receptor destruction was assessed by sp-LBA with BCoV HE^{0}-Fc at fixed concentration (2 ng/µL). (C) Whole-virion receptor-destruction assays with 2-fold serial dilutions of purified viruses (starting at 15 mU). BCoV and OC43 sHE^{+}s were included for comparison. Receptor destruction was measured after 1 hr or 4 hr incubation. Black dotted lines indicate 100% receptor destruction as determined with excess amounts of BCoV or OC43 sHE^{+}. (D) Whole-virus-mediated receptor destruction over time with fixed amounts of BCoV-Mebus and OC43 USA/1967 virions (15 mU). To exclude "exhaustion" (inactivation of OC43 esterase and/or changes in the physicochemical properties of the virions over time), we removed BCoV and OC43 at t = 24 and replaced them with equal amounts of freshly thawed aliquots of the virus stocks (black arrow). (E) Scaled side-by-side representations of the structures of coronavirus S [32] and HE proteins. Indicated are the estimated heights of S and HE and the approximate distance separating S receptor-binding sites and HE catalytic sites. (See also Figure S4 and S5 and Table S2.)

Consequences of Loss of Lectin Function for Virion-Mediated Receptor Destruction

OC43 HE efficiently depletes endogenous pools of *O*-Ac-Sias within infected cells (Figure S5) as is critical for the release of viral progeny [22]. The loss of HE lectin function therefore is more likely to affect virus biology at another stage, namely that of virion (pre) attachment. Of note, the observations for recombinant soluble HEs cannot be extrapolated one-to-one to virion-associated HEs, because in virus particles, HE esterase activity will be affected by S-mediated attachment to 9-*O*-Ac-Sias as well as by the close packing of HE molecules in the viral envelope. Serial dilutions of intact BCoV and OC43 virions, adjusted for pNPA esterase activity (Table S2), were therefore compared in solid-phase receptor-destruction assays with BSM (Figures 5C and 5D). Remarkably, the difference between the two viruses as measured by the onset of detectable receptor destruction was smaller than for their respective soluble HEs.

The activity of virion-associated BCoV HE relative to that of soluble BCoV HE was reduced 8- to 16-fold. Conversely, the relative activity of virion-associated OC43 was increased at least in that destruction of receptors already became apparent at lower enzyme concentrations. (Figure 5C). To study whether these phenomena relate to the S protein and to S-mediated virion attachment to 9-*O*-Ac-Sias, we constructed recombinant murine coronaviruses (rMuCoVs) that express MuCoV S, which does not bind 9-*O*-Ac-Sias, in combination with either BCoV or OC43 HE. rMuCoV virions studded with OC43 HE lacked detectable esterase activity toward BSM. In contrast, the esterase activity of rMuCoVs expressing BCoV HE was 6- to 8-fold higher than that of wild-type BCoV virions (Figure 5C). Apparently, β1CoV S proteins, by binding to 9-*O*-Ac-Sias negatively or positively, affect HE esterase activity in BCoV and OC43 virions, respectively, depending on whether or not the HE lectin domain is functional. Most strikingly, BCoV and OC43 virions differed in the extent to which

receptors were depleted. The soluble HEs of OC43 and BCoV, despite their difference in enzyme activity, destroyed all *O*-Ac-Sias as detectable by sp-LBA. BCoV virions destroyed 95% of the receptors after 24 hr incubation, which increased to 98% after 48 hr (Figure 5D). For OC43 virions, the effect was more pronounced such that after 24 hr only 85% of the receptors were destroyed. The remaining receptors were resistant even to continued incubation (Figure 5D). The data suggest that, in the context of the virion, the absence of a functional HE lectin domain results not only in reduced enzyme activity toward clustered receptors, but also prohibits cleavage of particular receptor populations. As these "protected" receptor populations are readily depleted by soluble OC43 HE (Figures 5B–5D), their resistance to cleavage by OC43 virions suggests that they are inaccessible to HEs when embedded in the viral envelope and in the presence of S.

The data prompt a model in which the susceptibility of populations of clustered receptors to cleavage depends on (1) their accessibility, as determined by the distance by which they extend from a fixed surface—the bottom of the ELISA well in our artificial system and, for example, the host cell membrane in vivo—and (2) the considerable difference in height between S and HE (20 and 8 nm, respectively), with the CBSs of S and those of HE esterase domains separated by a distance of up to 15 nm (Figure 5E). S-mediated virion attachment to a 9-*O*-Ac-sialylated surface would bring the HE catalytic site in proximity of some clustered receptors (which would explain the stimulating effect of S on HE esterase activity in the case of OC43 virions, but not observed for rMuCoV-OC43 HE^+) but at the same time would keep the virus-associated enzyme at "arm's length" from other receptor populations, readily accessible to soluble HEs (which would explain the inhibiting effect of S on esterase activity in the case of BCoV virions and the apparent lack thereof in the case of rMuCoV-BCoV HE^+) (Figures 5C and 5D). Thus, a distinction arises between clustered receptor populations that come within immediate reach of the HE esterase and that therefore would be cleaved by virion-associated HE as efficiently as by soluble HE and those that are kept out of reach of virion-associated HE as a result of S-mediated receptor-binding and that are therefore cleaved at strongly reduced rates or even rendered non-cleavable unless HE is provided with a functional lectin domain (Figure 6). By virions grasping on to multivalent sialoglycoconjugates via the HE receptor-binding sites and "drawing them in" closer to the surface of the envelope, clustered receptors would yet become available to the HE esterase domain (Figure 6). Such a mechanism may be promoted by cooperativity among adjacent HE molecules and, membrane fluidity permitting, possibly even lead to redistribution of surface projections in the viral envelope, resulting in displacement of spikes, local recruitment of HEs, and formation of a "receptor-destruction patch." Saliently, some receptor populations may be completely resistant to destruction even by virion-associated HEs with a functional lectin domain, such as the 2% of detectable *O*-Ac-Sias

remaining after 48 hr of incubation with BCoV (Figure 5D). It is tempting to speculate that under natural conditions, the distinction between decoy and entry receptors is made by default on basis of their accessibility to cleavage by virion-associated HE.

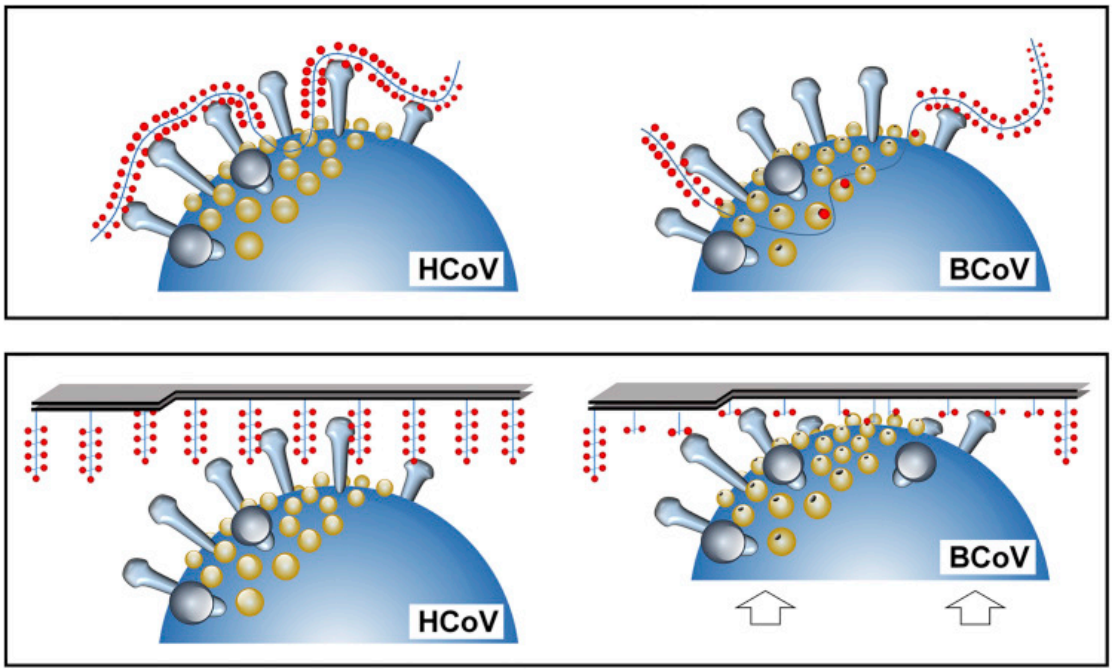


Figure 6. Hypothetical Model for the Interaction of HCoV and BCoV Virions with Multivalent Glycoconjugates. Schematically depicted are portions of HCoV (OC43 or HKU1) and BCoV virions, with large spikes comprised of S (in gray) and smaller protrusions comprised of HE (yellow) extending from the viral membrane (blue). Functional HE lectin CBSs are indicated by black holes (only one shown per HE dimer for reasons of simplicity). Also shown schematically are membrane-anchored (bottom) and non-anchored (top) mucin-type glycoconjugates of bottle-brush filamentous appearance with clustered receptors (9-*O*-Ac-Sias, red dots) arranged in linear arrays and with absence of red dots indicating receptor-destruction. The model, based on the size difference between S and HE, visualizes how loss of HE lectin function might alter virion-associated receptor-destroying activity, reducing the specific activity of virions as well as the rate and selectivity of receptor destruction. In virions with lectin-deficient HEs, clustered substrates will be largely kept at a distance from the HE esterase catalytic pocket as a result of S-glycoconjugate interaction. In contrast, HEs with intact lectin CBS may draw in portions of the glycoconjugates (or will draw the virion-associated HEs toward them), aided by cooperativity of binding between adjacent HEs within the viral envelope. Thus, clustered glycotopes become fixed within reach of the esterase catalytic sites and receptor destruction is promoted.

Conclusion

We showed that OC43—after its zoonotic introduction—has been under incessant selective pressure to adapt to the sialoglycome of the human host. HE-associated receptor binding was selected against, largely lost early on, and ultimately lost altogether through an accumulation of mutations in the lectin domain over a period of decades. In result, the balance between virion attachment and catalytically driven virion elution was reset, apparently to meet the specific requirements for optimal replication in human airways and to allow the virus in this particular niche to distinguish between decoys marked for destruction and functional receptors that should be preserved for cell entry. This view is reinforced by our observation that HCoV-HKU1, another respiratory lineage A βCoV of humans, followed a convergent evolutionary path and also lost HE-mediated receptor binding. We offer that differences in the sialoglycomes of bovids and humans, such as in the local expression, structure, and density of 9-*O*-acetylated sialoglycoconjugates, may pose incompatibilities during cross-species transmission of OC43 and BCoV and that the respective loss or preservation of HE-mediated receptor binding contributes to the host tropism of these viruses. This said, comparative sialoglycomics is still in its infancy, humans and bovids have not been studied for differences in *O*-Ac-sialoglycan expression in any detail, and the precise sialoglycan-based constraints that selected for the particular traits of the animal and human βCoVs therefore remain to be identified. At any rate, our findings do reveal an as yet unappreciated aspect of lineage A βCoV adaptation to humans. Of broader relevance, they provide a general paradigm for dynamic, catalysis-driven virus-sialoglycan interactions in relation to host selectivity. This notion is supported by observations for influenza A viruses, where neuraminidase (NA) activity toward multivalent substrates varies with differences in the length of the stalk domain and positively correlates with NA size [33,34], where NA catalytic activity is modulated by the absence or presence of a second Sia-binding site [35,36] and where changes in NA stalk length and NA lectin affinity have been implicated in host specificity [33,35–41]. These observations can be readily interpreted in the context of our data and model proposed.

Experimental Procedures

Materials and experimental procedures are detailed in the Supplemental Experimental Procedures.

Expression and Purification of HEs. Human codon-optimized sequences of the HEs of BCoV-Mebus, ECoV-NC99, PHEV-VW572, RbCoV-HKU14-1, CRCoV-240/05, HCoV-HKU1, and OC43 USA/1967 (ATCC VR-759) were cloned in expression plasmid pCD5-T-Fc [21]. The resulting constructs encode chimeric proteins comprised of the HE ectodomain fused to the human IgG1 Fc domain, with the domains

separated by a thrombin cleavage site. The fusion proteins were either expressed in an enzymatically active form (HE^+) or rendered inactive through a catalytic Ser-to-Ala substitution (HE^0). Plasmids coding for OC43 NL/A/2005 and OC43 NL/B/2005 HE were constructed through site-directed mutagenesis of the OC43 USA/1967 HE expression vector using the Q5 kit (New England Biolabs). For lectin affinity and esterase activity assays, HE-Fcs were produced by transient expression in HEK293T cells and purified from cell culture supernatants by protein A affinity chromatography and low pH elution as described [21]. For crystallization, OC43 NL/A/2005 HE^+-Fc was expressed in HEK293 GnTI($^-$) cells and purified by protein A affinity chromatography, followed by on-the-beads thrombin cleavage as described [21]. Beads were pelleted, and the HE ectodomain in the supernatant was concentrated to ∼15 mg/mL. HE genes from OC43 field strains were RT-PCR amplified with OC43 viral RNA, directly isolated from nasopharyngeal aspirates [42] as a template, and their sequences were used to construct codon-optimized versions.

Purification of Virions for Whole-Virus Receptor-Destruction Assays. Cell monolayers were infected at MOI 0.01. Virions were purified by ultracentrifugation through 20% (w/v) sucrose cushions and resuspended in PBS. Virus preparations were analyzed for particle content by qPCR, plaque assay, quantitative latex bead ratio EM, and pNPA esterase activity assay (Table S2).

Hemagglutination Assay. HAA was performed with rat erythrocytes (Rattus norvegicus strain Wistar; 50% suspension in PBS) and, in standard tests, with 2-fold serial dilutions of HE^0-Fc proteins (starting at 25 ng/well) as described [21]. For high-sensitivity HAA, based on multivalency-driven high-avidity binding, 5 μg HE^0-Fc was complexed to 2×10^9 protein A-coated 100 nm nanobeads (Chemicell) in 100 μL PBS for 45 min at 4°C prior to 2-fold serial dilution. Hemagglutination was for 2 hr at 4°C unless stated otherwise.

Solid-Phase Lectin-Binding Assay and On-the-Plate O-Ac-Sia Depletion Assay. sp-LBAs were performed with bovine submaxillary mucin (BSM) and 2-fold serial dilutions of HE0-Fc proteins as described [23,43]. Receptor-destroying esterase activities of soluble recombinant HEs (sHE^+) and whole viruses were measured by on-the-plate *O*-Ac-Sia depletion assay, performed essentially as described [23]. BSM, coated on Maxisorp flat-bottom 96-well plates, was (mock-)treated with 2-fold serial dilutions of sHE^+ in PBS (100 μL/well) starting at 20 ng/μL for 1 hr, or, for whole-virion receptor-destruction assays, with virions or sHE^+ starting at 15 mU esterase activity for up to 48 hr at 37°C. Depletion of *O*-Ac-Sia receptors was detected by sp-LBA with BCoV-LUN HE^0 (2 ng/μL; 100 μL/well). Esterase activity was plotted as the inverse of lectin-binding measurements, expressed in percentages.

Acknowledgments

We acknowledge the Paul Scherrer Institut, Villigen, Switzerland for provision of synchrotron radiation beam time at beamline PX of the Swiss Light Source (SLS) and its beamline scientists for assistance. We thank Jean-Luc Murk and Marco Viveen from the Diagnostic Electron Microscopy Unit, Medical Microbiology, University Medical Center Utrecht for the titration of virus stocks and Richard Wubbolts from the Center for Cell Imaging of the Utrecht Faculty of Veterinary Medicine for advice and assistance. This work was supported by ECHO grant 711.011.006 of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW).

References

1. De Wit, E.; Van Doremalen, N.; Falzarano, D.; Munster, V.J. SARS and MERS: Recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* 2016, *14*, 523–534.
2. Reusken, C.B.E.M.; Raj, V.S.; Koopmans, M.P.; Haagmans, B.L. Cross host transmission in the emergence of MERS coronavirus. *Curr. Opin. Virol.* 2016, *16*, 55–62.
3. Zumla, A.; Hui, D.S.; Perlman, S. Middle East respiratory syndrome. *Lancet* 2015, *386*, 995–1007.
4. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A.C.K.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G.F. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016, *24*, 490–502.
5. Groot, R. de; de Groot, R.; Baker, S.; Baric, R.; … L.E.-V.; 2012, U. Family Coronaviridae. *Virus Taxon. Ninth Rep. Int. Comm. Taxon. Viruses* **2011**, 806–828.
6. Morfopoulou, S.; Brown, J.R.; Davies, E.G.; Anderson, G.; Virasami, A.; Qasim, W.; Chong, W.K.; Hubank, M.; Plagnol, V.; Desforges, M.; et al. Human Coronavirus OC43 Associated with Fatal Encephalitis. *N. Engl. J. Med.* **2016**, *375*, 497–8.
7. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Huang, Y.; Tsoi, H.W.; Wong, B.H.L.; Wong, S.S.Y.; Leung, S.Y.; Chan, K.H.; Yuen, K.Y. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 14040–14045.
8. Lau, S.K.P.; Lee, P.; Tsang, A.K.L.; Yip, C.C.Y.; Tse, H.; Lee, R.A.; So, L.-Y.; Lau, Y.-L.; Chan, K.-H.; Woo, P.C.Y.; et al. Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. *J. Virol.* **2011**, *85*, 11325–11337.
9. Vijgen, L.; Keyaerts, E.; Moes, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604.
10. Vijgen, L.; Keyaerts, E.; Lemey, P.; Maes, P.; Van Reeth, K.; Nauwynck, H.; Pensaert, M.; Van Ranst, M. Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *J. Virol.* **2006**, *80*, 7270–7274.
11. Matrosovich, M.; Herrler, G.; Klenk, H.D. Sialic Acid Receptors of Viruses. *Top. Curr. Chem.* **2015**, *367*, 1–28.
12. Traving, C.; Schauer, R. Structure, function and metabolism of sialic acids. *Cell. Mol. Life Sci.* 1998, *54*, 1330–1349.
13. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729.
14. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Tsang, A.K.L.; Fan, R.Y.Y.; Luk, H.K.H.; Cai, J.-P.; Chan, K.-H.; Zheng, B.-J.; Wang, M.; et al. Discovery of a Novel Coronavirus, China Rattus Coronavirus HKU24, from Norway Rats Supports the Murine Origin of Betacoronavirus 1 and Has Implications for the Ancestor of Betacoronavirus Lineage A. *J. Virol.* **2015**, *89*, 3076–3092.
15. Wang, W.; Lin, X.D.; Guo, W.P.; Zhou, R.H.; Wang, M.R.; Wang, C.Q.; Ge, S.; Mei, S.H.; Li, M.H.; Shi, M.; et al. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* **2015**, *474*, 19–27.
16. Heald-Sargent, T.; Gallagher, T. Ready, set, fuse! the coronavirus spike protein and acquisition of fusion

competence. *Viruses* 2012, *4*, 557–580.

17. de Groot, R.J. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* 2006, *23*, 59–72.

18. Hause, B.M.; Ducatez, M.; Collin, E.A.; Ran, Z.; Liu, R.; Sheng, Z.; Armien, A.; Kaplan, B.; Chakravarty, S.; Hoppe, A.D.; et al. Isolation of a Novel Swine Influenza Virus from Oklahoma in 2011 Which Is Distantly Related to Human Influenza C Viruses. *PLoS Pathog.* **2013**, *9*.

19. Langereis, M.A.; Zeng, Q.; Gerwig, G.J.; Frey, B.; Itzstein, M. von; Kamerling, J.P.; de Groot, R.J.; Huizinga, E.G. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15897–15902.

20. Rosenthal, P.B.; Zhang, X.; Formanowski, F.; Fitz, W.; Wong, C.H.; Meier-Ewert, H.; Skehel, J.J.; Wiley, D.C. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **1998**, *396*, 92–96.

21. Zeng, Q.; Langereis, M.A.; van Vliet, A.L.W.; Huizinga, E.G.; de Groot, R.J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9065–9069.

22. Desforges, M.; Desjardins, J.; Zhang, C.; Talbot, P.J. The Acetyl-Esterase Activity of the Hemagglutinin-Esterase Protein of Human Coronavirus OC43 Strongly Enhances the Production of Infectious Virus. *J. Virol.* **2013**, *87*, 3097–3107.

23. Langereis, M.A.; Bakkers, M.J.G.; Deng, L.; Padler-Karavani, V.; Vervoort, S.J.; Hulswit, R.J.G.; van Vliet, A.L.W.; Gerwig, G.J.; de Poot, S.A.H.; Boot, W.; et al. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* **2015**, *11*, 1966–1978.

24. McIntosh, K.; Becker, W.B.; Chanock, R.M. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci.* **1967**, *58*, 2268–2273.

25. Huang, X.; Dong, W.; Milewska, A.; Golda, A.; Qi, Y.; Zhu, Q.K.; Marasco, W.A.; Baric, R.S.; Sims, A.C.; Pyrc, K.; et al. Human Coronavirus HKU1 Spike Protein Uses O -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J. Virol.* **2015**, *89*, 7202–7213.

26. Woo, P.C.Y.; Lau, S.K.P.; Tsoi, H.; Huang, Y.; Poon, R.W.S.; Chu, C.; Lee, R.A.; Luk, W.; Wong, G.K.M.; Wong, B.H.L.; et al. Clinical and Molecular Epidemiological Features of Coronavirus HKU1–Associated Community-Acquired Pneumonia. *J. Infect. Dis.* **2005**, *192*, 1898–1907.

27. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Tsoi, H.-W.; Chan, K.-H.; Yuen, K.-Y. Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1. *J. Virol.* **2006**, *80*, 7136–7145.

28. Langereis, M.A.; van Vliet, A.L.W.; Boot, W.; de Groot, R.J. Attachment of Mouse Hepatitis Virus to O-Acetylated Sialic Acid Is Mediated by Hemagglutinin-Esterase and Not by the Spike Protein. *J. Virol.* **2010**, *84*, 8970–8974.

29. Peng, G.; Sun, D.; Rajashankar, K.R.; Qian, Z.; Holmes, K. V; Li, F. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10696–701.

30. Boraston, A.B.; Bolam, D.N.; Gilbert, H.J.; Davies, G.J. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* 2004, *382*, 769–781.

31. Zappone, B.; Patil, N.J.; Madsen, J.B.; Pakkanen, K.I.; Lee, S. Molecular structure and equilibrium forces of bovine submaxillary mucin adsorbed at a solid-liquid interface. *Langmuir* **2015**, *31*, 4524–4533.

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

33. Castrucci, M.R.; Kawaoka, Y. Biologic importance of neuraminidase stalk length in influenza A virus. *J. Virol.* **1993**, *67*, 759–764.

34. Els, M.C.; Air, G.M.; Murti, K.G.; Webster, R.G.; Laver, W.G. An 18-amino acid deletion in an influenza neuraminidase. *Virology* **1985**, *142*, 241–247.

35. Uhlendorff, J.; Matrosovich, T.; Klenk, H.D.; Matrosovich, M. Functional significance of the hemadsorption activity of influenza virus neuraminidase and its alteration in pandemic viruses. *Arch. Virol.* **2009**, *154*, 945–957.

36. Varghese, J.N.; Colman, P.M.; van Donkelaar, A.; Blick, T.J.; Sahasrabudhe, A.; McKimm-Breschkin, J.L. Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl. Acad. Sci.* **1997**, *94*, 11808–11812.

37. Blumenkrantz, D.; Roberts, K.L.; Shelton, H.; Lycett, S.; Barclay, W.S. The Short Stalk Length of Highly Pathogenic Avian Influenza H5N1 Virus Neuraminidase Limits Transmission of Pandemic H1N1 Virus in Ferrets. *J. Virol.* **2013**,

87, 10539–10551.

38. Lai, J.C.C.; Garcia, J.M.; Dyason, J.C.; Böhm, R.; Madge, P.D.; Rose, F.J.; Nicholls, J.M.; Peiris, J.S.M.; Haselhorst, T.; Von-Itzstein, M. A secondary sialic acid binding site on influenza virus neuraminidase: Fact or fiction? *Angew. Chemie - Int. Ed.* **2012**, *51*, 2221–2224.

39. Li, Y.; Chen, S.; Zhang, X.; Fu, Q.; Zhang, Z.; Shi, S.; Zhu, Y.; Gu, M.; Peng, D.; Liu, X. A 20-amino-acid deletion in the neuraminidase stalk and a five-amino-acid deletion in the ns1 protein both contribute to the pathogenicity of h5n1 avian influenza viruses in mallard ducks. *PLoS One* **2014**, *9*.

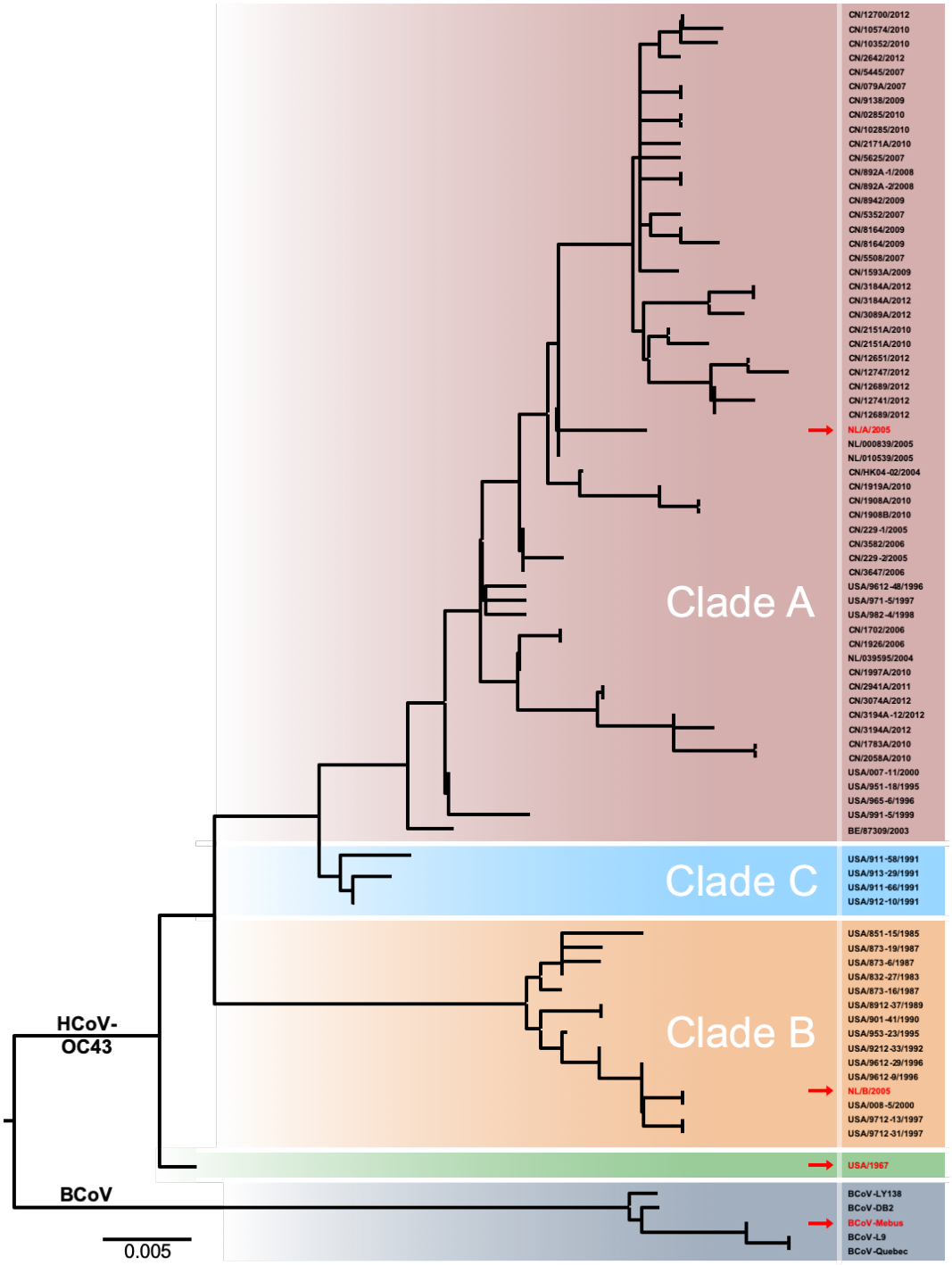
40. Matrosovich, M.; Zhou, N.; Kawaoka, Y.; Webster, R. The Surface Glycoproteins of H5 Influenza Viruses Isolated from Humans, Chickens, and Wild Aquatic Birds Have Distinguishable Properties. *J. Virol.* **1999**, *73*, 1146–1155.

41. Munier, S.; Larcher, T.; Cormier-Aline, F.; Soubieux, D.; Su, B.; Guigand, L.; Labrosse, B.; Cherel, Y.; Quéré, P.; Marc, D.; et al. A Genetically Engineered Waterfowl Influenza Virus with a Deletion in the Stalk of the Neuraminidase Has Increased Virulence for Chickens. *J. Virol.* **2010**, *84*, 940–952.

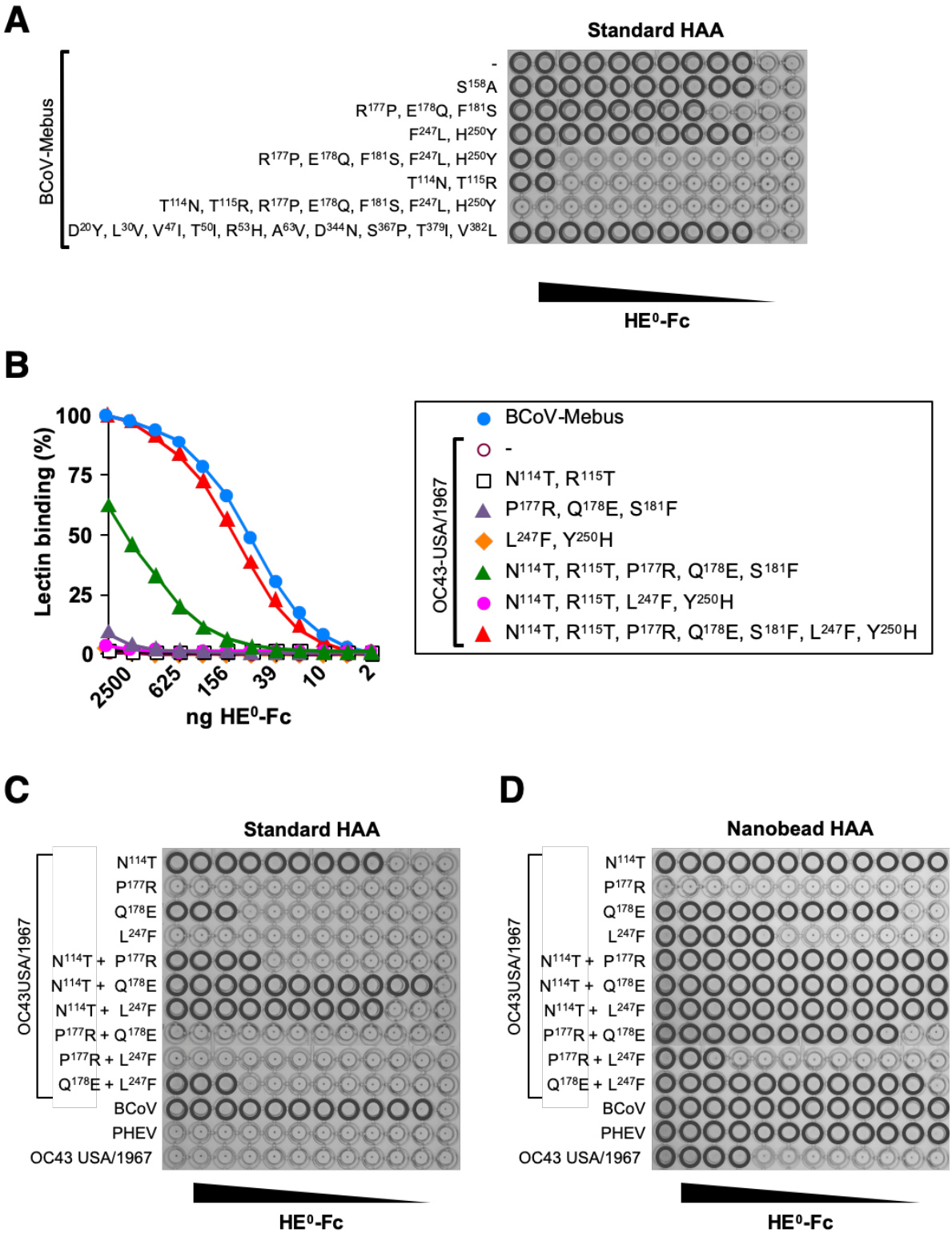
42. van de Pol, A.C.; Wolfs, T.F.W.; Jansen, N.J.G.; van Loon, A.M.; Rossen, J.W.A. Diagnostic value of real-time polymerase chain reaction to detect viruses in young children admitted to the paediatric intensive care unit with lower respiratory tract infection. *Crit. Care* **2006**, *10*, R61.

43. Langereis, M.A.; Zeng, Q.; Heesters, B.; Huizinga, E.G.; de Groot, R.J. The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. *PLoS Pathog.* **2012**, *8*.

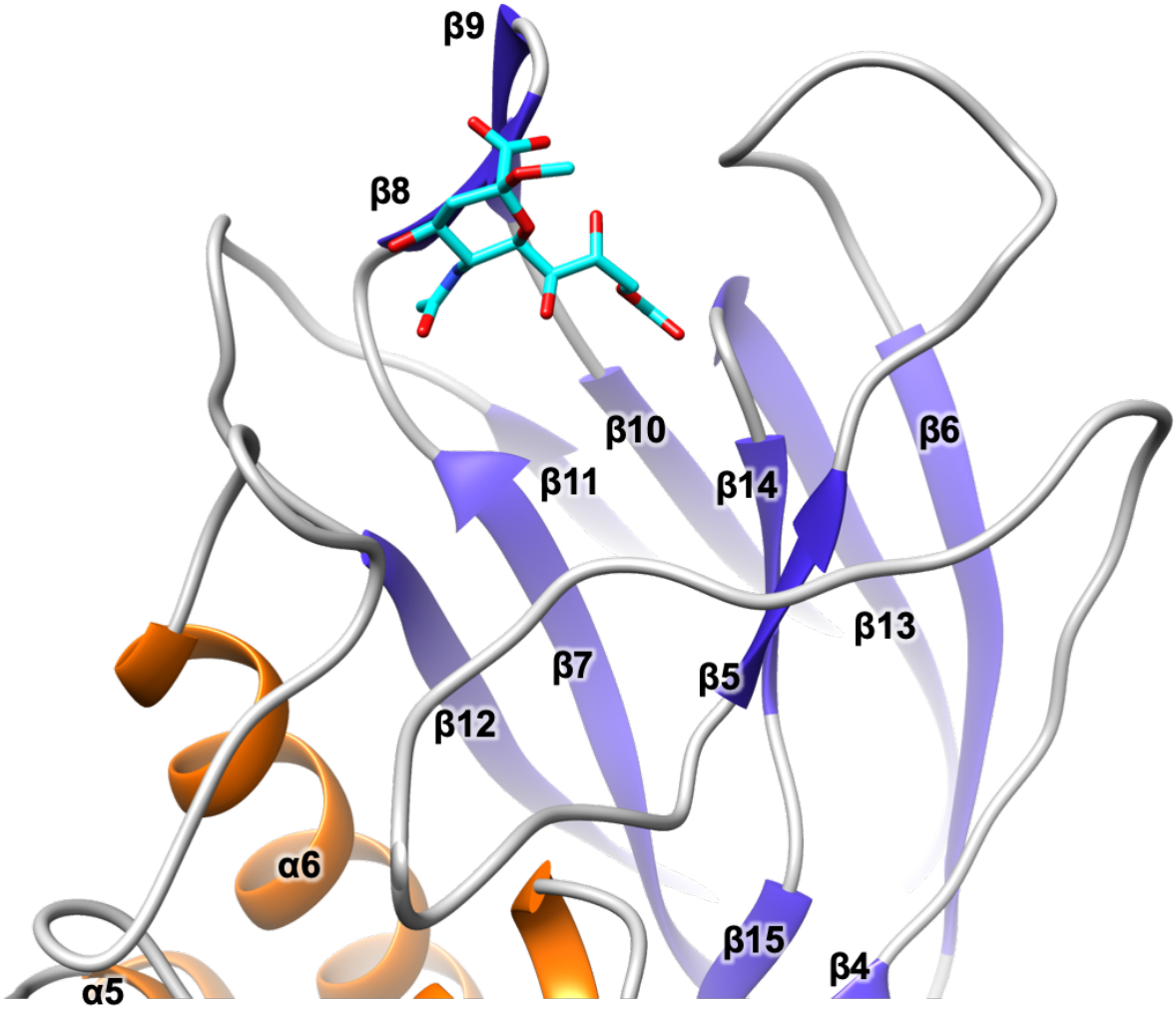
Supplementary Information



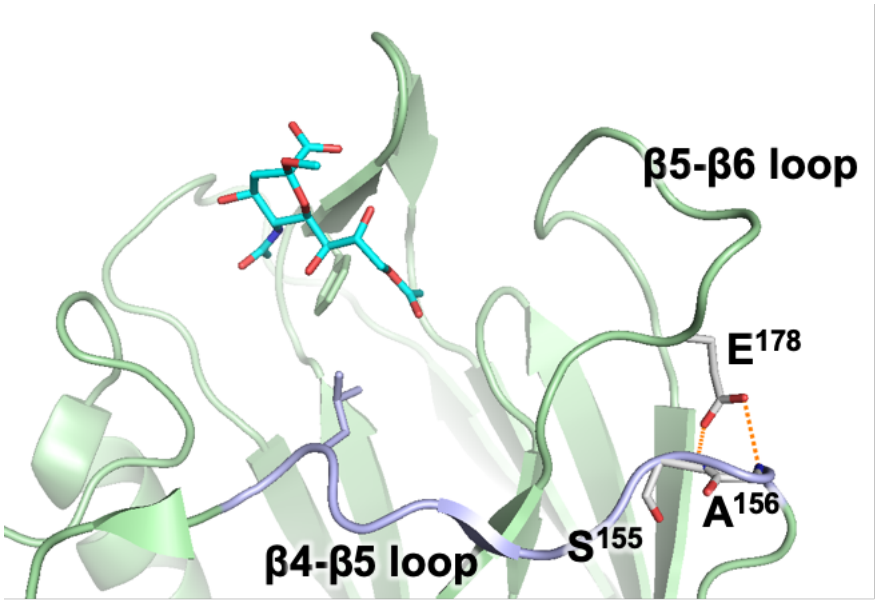
sFigure 1. Neighbor-Joining phylogenetic tree based on all OC43 HE sequences in the NCBI database, related to Figure 1. Evolutionary distances were computed using the Maximum Composite Likelihood method in MEGA6. BCoV HEs were included as an outgroup. OC43 HE sequences of unknown sampling date were omitted from the final tree. The different clades are color-coded with OC43 USA/1967 in green, and OC43 clades A-C in purple, orange and blue, respectively. A number of representative BCoV HEs are shaded in dark blue. HEs used in this study are shown in red and indicated with a red arrow.



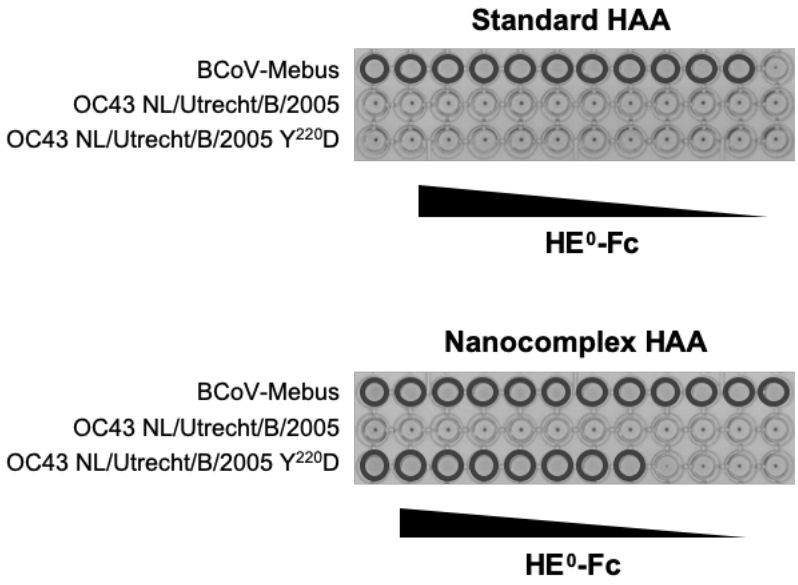
sFigure 2. Lectin affinity prescreen of OC43 USA/1967 HE specific mutations. (A) Mutations of OC43 USA/1967 HE were placed in the background of BCoV-Mebus HE^0 in sets and tested in standard HAA. (B) Corresponding residues from BCoV-Mebus HE were placed in thebackground of OC43 USA/1967 HE^0 in sets and tested by sp-LBA. (C, D) Single and double mutations from BCoV-Mebus HE were placed in the background of OC43 USA/1967 HE^0 and tested by standard HAA (C) and by nanobead HAA (D).



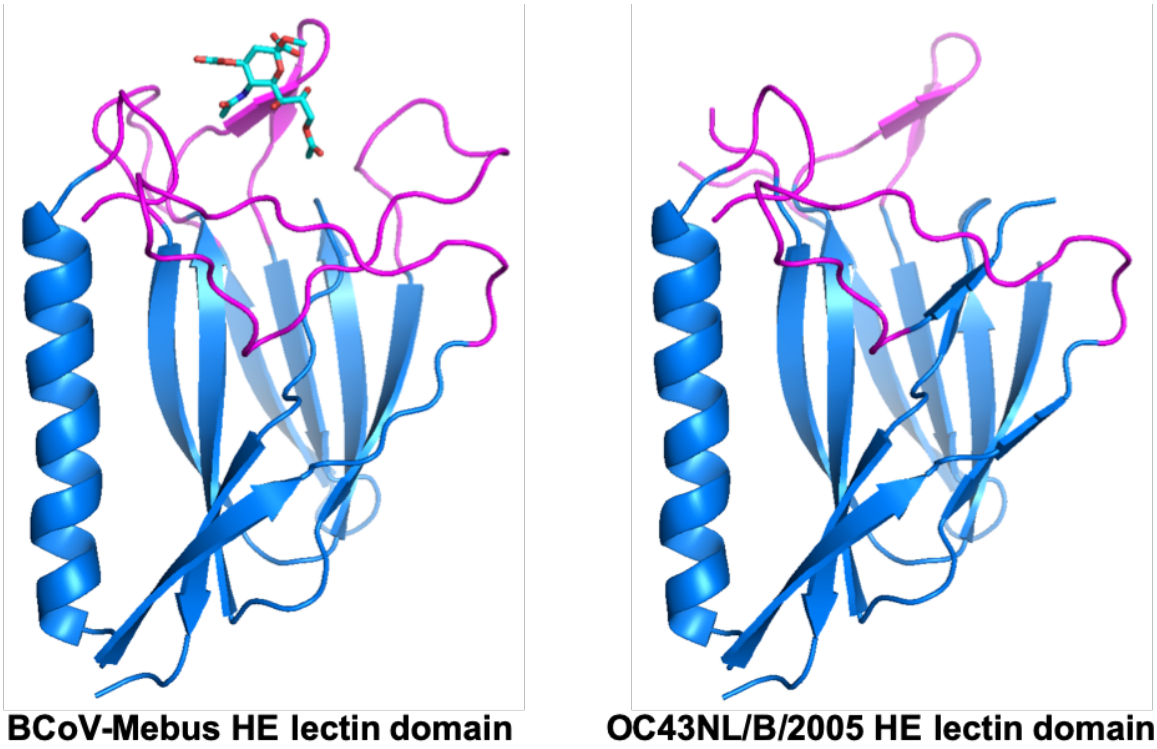
sFigure 3. Cartoon representation of the lectin domain of BCoV-Mebus HE in complex with 9-*O*-Ac-Sia (PDB ID: 3CL5). The structure is colored according to secondary structure with α-helices in orange and β-strands in blue. The receptor analogue in stick representation is colored according to atom type with carbon in cyan, oxygen in red and nitrogen in blue. Secondary structure elements are numbered sequentially.



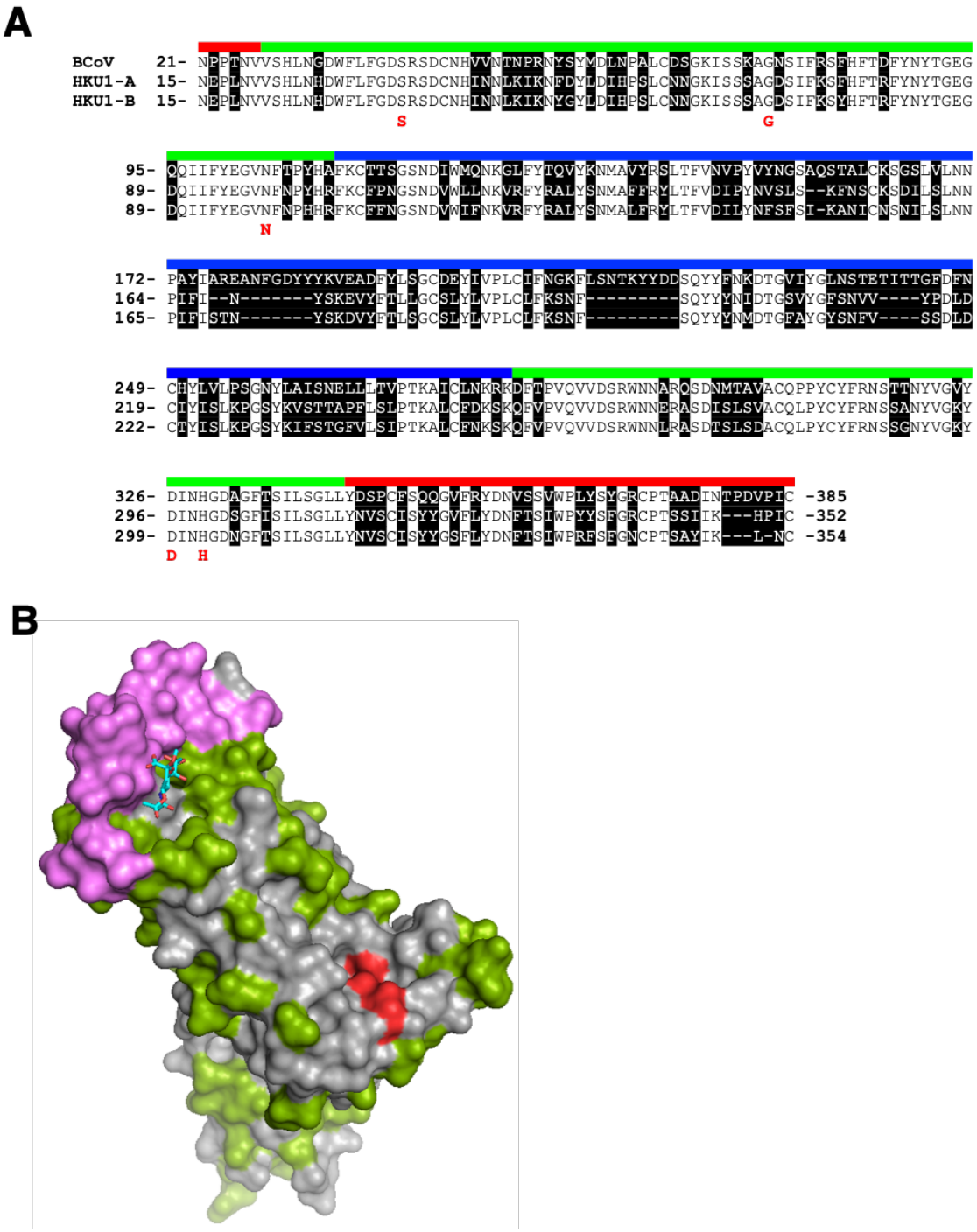
sFigure 4. E^{178} indicated in a cartoon representation of the lectin domain of BCoV-Mebus HE in complex with 9-*O*-Ac-Sia (PDB ID: 3CL5). E^{178} is located in the β5-β6 loop from where it can form hydrogen bonds with S^{155} and A^{156} present in the β4-β5 loop (shown in purple). The receptor analogue in stick representation is colored as in sFig. 3.



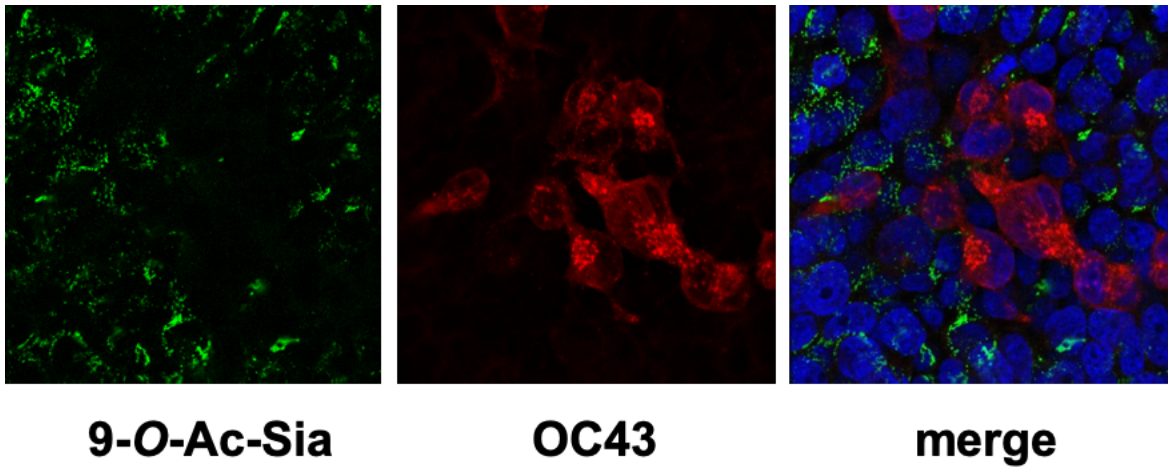
sFigure 5. A presumptive frameshift mutation in OC43 NL/A/2005 HE. Coding sequence of the DYYY to IIT mutation, with codons underlined alternatingly in black and red. Conserved bases are in blue, and mutations/deleted bases are in red.



sFigure 6. Comparison of the overall fold of the lectin domains of BCoV-Mebus HE in complex with 9-*O*-Ac-Sia (PDB ID: 3CL5) and of OC43 NL/A/2005 HE. Structures are shown as cartoons with the structurally conserved scaffold in blue and the extended loops that form the receptor binding site in pink. The receptor analogue is shown as in sFig. 3.



sFigure 7. (A) Amino acid alignment of BCoV HE with representatives of the two HCoV-HKU1 HE clades (designated clades A and B). Domain organization is color-coded (membraneproximal domain, red; esterase domain, green; lectin domain, blue). Residues crucial for esterase activity (SGNDH) are annotated below. Amino acid differences are marked in black. (B) Surface representation of BCoV-Mebus HE in complex with 9-*O*-Ac-Sia (PDB ID: 3CL5) with the amino acid mutations (green) and deletions (violet) that occurred in HKU1 HE visualized. The catalytic triad is indicated in red.



sFigure 8. Destruction of endogenous receptors in OC43-infected HEK293T cells. Cells were PFA-fixed, Triton-X_{100}-permeabilized and double-stained for 9-*O*-Ac-Sia with virolectin P4 and for OC43 proteins with serum from a BCoV-infected cow. Nuclei (N) stained with Hoechst-33258.

Chapter 3

Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor binding site in spike protein domain A

Ruben J.G. Hulswit[1#], Yifei Lang[1#], Mark J.G. Bakkers[1#], Wentao Li[1], Zeshi Li[2,3], Arie Schouten[4], Bram Ophorst[1], Frank J.M. van Kuppeveld[1], Geert-Jan P.H. Boons[2,3,5], Berend-Jan Bosch[1], Eric G. Huizinga[4], and Raoul J. de Groot[1]*

[1]Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CH Utrecht, the Netherlands;

[2]Department of Chemistry, University of Georgia, Athens, GA 30602;

[3]Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CG Utrecht, The Netherlands;

[4]Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Sciences, Utrecht University, 3584 CH Utrecht, the Netherlands;

[5]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602;

[#]equal contribution

Proc Natl Acad Sci U S A. 2019 Feb 12;116(7):2681-2690

Abstract

Human betacoronaviruses OC43 and HKU1 are endemic respiratory pathogens and, while related, originated from independent zoonotic introductions. OC43 is in fact a host range variant of the species *Betacoronavirus-1*, and more closely related to bovine coronavirus (BCoV) -its presumptive ancestor- and porcine hemagglutinating encephalomyelitis virus (PHEV). The beta1-coronaviruses (β1CoVs) and HKU1 employ glycan-based receptors carrying 9-*O*-acetylated sialic acid (9-*O*-Ac-Sia). Receptor binding is mediated by spike protein S, the main determinant of coronavirus host specificity. For BCoV, a crystal structure for the receptor-binding domain $S1^A$ is available and for HKU1 a cryo-EM structure of the complete S ectodomain. However, the location of the receptor-binding site (RBS) -arguably the single-most important piece of information- is unknown. Here we solved the 3.0 Å crystal structure of PHEV $S1^A$. We then took a comparative structural analysis approach to map the β1CoV S RBS, using the general design of 9-*O*-Ac-Sia-binding sites as blueprint, backed-up by automated ligand docking, structure-guided mutagenesis of OC43, BCoV and PHEV $S1^A$, and infectivity assays with BCoV-S-pseudotyped vesicular stomatitis viruses. The RBS is not exclusive to OC43 and related animal viruses, but apparently conserved and functional also in HKU1 $S1^A$. The binding affinity of the HKU1 S RBS towards short sialoglycans is significantly lower than that of OC43, which we attribute to differences in local architecture and accessibility, and which may be indicative for differences between the two viruses in receptor fine-specificity. Our findings challenge reports that would map the OC43 RBS elsewhere in $S1^A$ and that of HKU1 in domain $S1^B$.

Significance statement

Human coronaviruses OC43 and HKU1 are related, yet distinct respiratory pathogens, associated with common colds, but also with severe disease in the frail. Both viruses employ sialoglycan-based receptors with 9-*O*-acetylated sialic acid (9-*O*-Ac-Sia) as key component. Here, we identify the 9-*O*-Ac-Sia-specific receptor-binding site (RBS) of OC43 S and demonstrate it to be conserved and functional in HKU1. The considerable difference in receptor-binding affinity between OC43 and HKU1 S, attributable to differences in local architecture and RBS accessibility, is suggestive of differences between OC43 and HKU1 in their adaptation to the human sialome. The data will enable studies into the evolution and pathobiology of OC43 and HKU1 and open new avenues towards prophylactic and therapeutic intervention.

Introduction

Coronaviruses (CoVs; order *Nidovirales*, family *Coronaviridae*) are enveloped positive-strand RNA viruses of mammals and birds. So far, four coronaviruses of zoonotic origin are known to have successfully breached the species barrier to become true human pathogens [1–6]. These viruses, NL63, 229E, HKU1 and OC43, are persistently maintained in the human population through continuous circulation. Remarkably, the latter two both belong to a single minor clade, 'lineage A', in the genus *Betacoronavirus*. Although generally associated with common colds, HKU1 and OC43 may cause severe and sometimes fatal pulmonary infections in the frail [7, 8], and OC43, in rare instances, lethal encephalitis [9]. OC43 and HKU1 are distinct viruses that entered the human population independently to seemingly follow convergent evolutionary trajectories in their adaptation to the novel host [10]. OC43 is in fact more related to coronaviruses of ruminants, horses, dogs, rabbits and swine, with which it has been united in a single species *Betacoronavirus-1*.

Lineage A betacoronaviruses like HKU1 and OC43 differ from other CoVs in that their virions possess two types of surface projections both of which are involved in attachment: large 20 nm peplomers or 'spikes' that are very much a CoV hallmark and comprised of homotrimers of spike (S) protein, and 8 nm protrusions, unique to this clade, comprised of the homodimeric hemagglutinin-esterase (HE). S is central to viral entry and the key determinant of host and tissue tropism [11]. It mediates binding to cell surface receptors and, upon uptake of the virion by the host cell, fusion between the viral envelope and the limiting endosomal membrane [12]. In the case of HKU1 and beta-1 coronaviruses (β1CoVs), OC43 included, S binds to sugar-based receptor-determinants, specifically to 9-O-acetylated sialic acids (9-*O*-Ac-Sias) attached as terminal residues to glycan chains on glycoproteins and lipids [13, 14]. HE, a sialate-*O*-acetylesterase with appended 9-*O*-Ac-Sia-specific lectin domain, acts as a receptor-destroying enzyme (RDE) [15]. During pre-attachment, the RDE activity of HE averts irreversible binding of virions to the decoy receptors that are omnipresent in the extracellular environment. Furthermore, at the conclusion of the replication cycle, HE-mediated destruction of intracellular and cell surface receptors facilitates the release of viral progeny from the infected cell [16]. Remarkably, the HEs of OC43 and HKU1 lost their ability to bind 9-*O*-Ac-Sias, because their lectin domain, functional in all other HEs studied so far, was rendered inactive [10]. In consequence, the dynamics and extent of virion-mediated destruction of clustered receptor populations were altered presumably to match the sialome of the human respiratory tract and to optimize infection and/or transmission. Whether adaptation to the human host also entailed adaptations in S is not known and as yet cannot be assessed as for HKU1 the S receptor-binding site (RBS) has not been identified and the β1CoV S RBS has not been established with certainty [17].

Crystal structure analysis of orthomyxo-, toro- and coronavirus HEs complexed to receptor/substrate analogues identified 9-*O*-Ac-Sia binding sites, yielding exquisite insight into the architecture of these sites and into the general principles of ligand/substrate recognition at the atomic level [18–24]. Recently, cryo-EM structures were reported for several S proteins, including that of HKU1 [25–29]. The findings have greatly increased our understanding of the overall quaternary structure and function of the S homotrimers. Among others, the structures revealed how in betacoronavirus spikes the N-terminal S1 subunit of each S monomer folds into four individual domains – designated A through D as numbered from the N-terminus (see Fig. 1a) – of which domains A and B may function in receptor binding [11]. The spikes of β1CoVs OC43 and BCoV bind to *O*-acetylated sialic acid through domain A ($S1^A$) as determined by in vitro binding assays [30]. There is limited structural information, however, on how the spikes of HKU1 and the β1CoVs bind their ligands. The apo-structure of the BCoV $S1^A$ lectin domain was solved, but attempts to also solve the holo-structure reportedly failed [17]. Based on the galectin-like fold of the $S1^A$ domain and mutational analysis, the receptor-binding site was predicted [17]. Although this model remains to be confirmed, it has been widely accepted by the field [27, 28, 31–33]. Surprisingly, HKU1 S was recently reported to bind to its receptor via a domain other than S1A [31]. Despite the similarity of HKU1 domain A to that of BCoV and OC43, binding of HKU1 S1A to 9-*O*-Ac-Sia was -reportedly- not detectable [30, 31]. Prompted by these observations and the fact that the predicted RBS in β1CoV $S1^A$ in its design and architecture bears no resemblance to other 9-*O*-Ac-Sia binding sites, we sought to test the published model. The results led us to look for alternative binding sites in the β1CoV $S1^A$ domain through comparative structural analysis and in silico modelling using the general design of HE 9-*O*-Ac-Sia binding sites as a blueprint, backed up by structure-guided mutagenesis. Our findings show that the actual $S1^A$ RBS in β1CoV S maps elsewhere than currently believed. Moreover, we demonstrate that the newly proposed site is not exclusive to β1CoVs, but in fact conserved and functional also in the $S1^A$ domain of HKU1.

Results and Discussion

The S1^A RBS is located elsewhere than currently believed. To test the validity of the current model, we measured the effect of substitutions in the proposed RBS using S1^A-Fc fusion proteins. The binding properties of the mutated proteins were studied by hemagglutination assay (HAA) with rat erythrocytes and by solid phase lectin-binding enzyme-linked immune assay (sp-LBA) with bovine submaxillary mucin (BSM) as ligand. These assays are complementary: HAA is the more sensitive of the two, while the results of sp-LBA more precisely reflect differences in binding site affinity [10]. HAA and sp-LBA performed with S1A of BCoV strain Mebus confirmed the binding to 9-*O*-Ac-Sias (Fig. 1b-d). For comparison, the S1^A domains of related β1CoVs OC43 strain ATCC and porcine hemagglutinating encephalomyelitis virus (PHEV) strain UU were included. Again, 9-*O*-Ac-Sia-dependent binding was observed, but the affinity of the S1^A domain of OC43 was approximately 32-fold lower than that of BCoV as measured by sp-LBA and that of PHEV was lower still (~1,450-fold). Substitution in PHEV and OC43 S1A of Ala for Tyr^{162}, Glu^{182}, Trp^{184} and His^{185} (Fig. 2a), residues deemed critical for binding of BCoV S1A [17], indeed resulted in decreased binding (Fig. 2b, c), but, importantly and as in the original report, none of the mutations gave complete loss of function. The strongest effect was observed for Trp^{184}, but its substitution in the context of OC43 S1^A merely reduced binding affinity to that of wildtype PHEV S1^A. In a reverse approach, we attempted to identify substitutions that result in gain-of-function of OC43 S1^A, i.e. mutations that would raise its binding to that of BCoV S1^A. To this end, we systematically replaced residues within or proximal to the proposed binding site in OC43 S1^A by their BCoV orthologues (Fig. 2a; see also SI Appendix, Fig. S1). Individual substitutions Arg^{143}His, Lys^{181}Val, Leu^{186}Trp and Ile^{145}Thr (Fig. 2d), however, did not increase binding affinity, and neither did the replacement of residues 147-151 (Ser-Thr-Gln-Asp-Gly) by Leu, which resulted not only in a large deletion but also in the removal of an N-glycosylation site. In fact, an OC43 S1^A mutant, in which we combined all substitutions and thus essentially reconstructed the proposed BCoV S1^A RBS in the OC43 background, did not differ in its binding affinity from the parental wildtype OC43 protein as measured by HAA and sp-LBA (Fig. 2d). The combined results and the fact that the RBS proposed by Peng et al. [17] -henceforth referred to as 'site A'- does not conform to the typical anatomy of O-Ac-Sia binding sites [18–24], lead us to conclude that the actual RBS is located elsewhere.

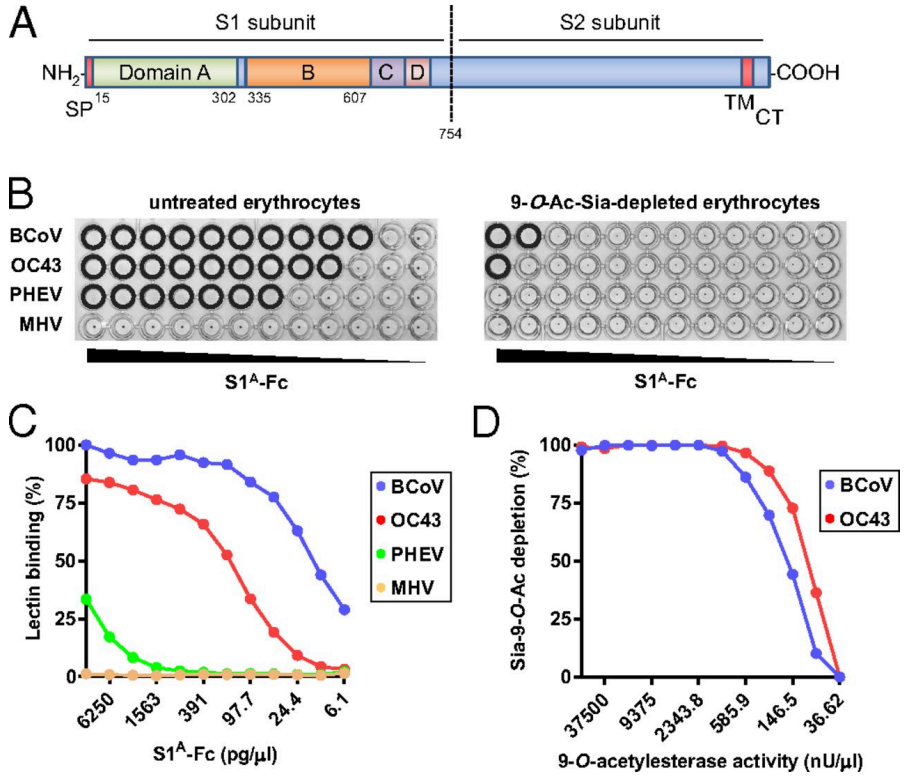


Fig. 1. S1^A domains of β1CoV host range variants all bind 9-*O*-Ac-Sias, but with different affinities. (A) Schematic representation of β1CoV spike protein with subunits S1 and S2, and S1 domains A through D indicated. Residue numbering is based on the OC43 strain ATCC/VR-759 spike protein (GenBank: AAT84354.1) with domain boundaries based on the MHV S structure). SP, signal peptide; TM, transmembrane domain; CT, cytoplasmic tail. (B) S1^A β1CoV variants differ in 9-*O*-Ac-Sia-binding affinity. (left panel) Conventional HAA with rat erythrocytes (diluted to a final concentration of 0.25% in PBS, 0.1% BSA) and 2-fold serial dilutions of S1^A-Fc fusion proteins of β1CoV variants BCoV-Mebus, OC43-ATCC and PHEV-UU. S1^A-Fc of MHV-A59 (starting at 25 ng/µl) was included as a negative control]. Hemagglutination was assessed after 2 hr incubation at 4°C. Wells scored positive for hemagglutination are encircled. HAAs were repeated at least 3 times. Representative experiments are shown. (right panel) S1^A-Fc-mediated hemagglutination is 9-*O*-Ac-Sia dependent. HAAs were performed as above, but now with rat erythrocytes, depleted for 9-*O*-Ac-Sias by prior sialate-*O*-acetylesterase treatment as in]. (C) Differences in S1^A-mediated 9-*O*-Ac-Sia-binding affinity among β1CoV variants S1^A protein as demonstrated by solid phase lectin-binding enzyme-linked immune assay (sp-LBA). S1^A-Fc fusion proteins (2-fold serial dilutions, starting at 12.5 ng/µl) were compared by sp-LBA for relative binding to bovine submaxillary mucin (BSM) at 37°C. MHV S1^A was included as a negative control. Experiments were performed at least three times, each time in triplicate, with each data point representing the average of the independent mean values. Mean ± *standard deviations* were less than 10%; error bars omitted for esthetical reasons. (D) S1^A-Fc-binding to BSM is 9-*O*-Ac-Sia dependent. BSM was specifically depleted for sialate-9-*O*-acetyl moieties by on-the-plate sialate-*O*-acetylesterase treatment for 2 hr with 2-fold serial dilutions of soluble hemagglutinin-esterase as in [10, 47] starting at 75 µU/µl. Receptor destruction was assessed by sp-LBA with fixed concentrations of BCoV and OC43 S1^A-Fc (0.3 ng/µl and 1.2 ng/µl, respectively).

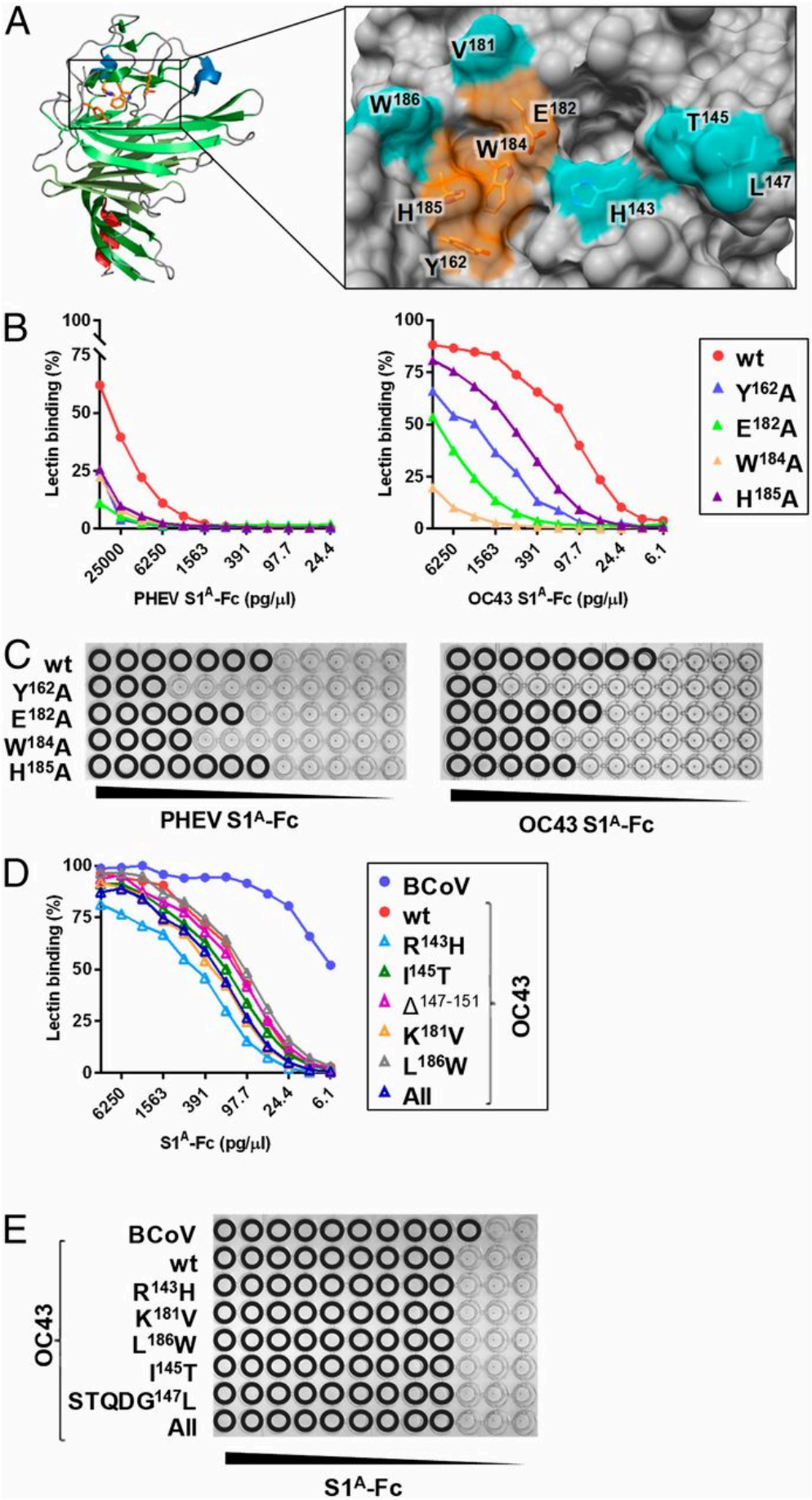

A
V181
W186
E182
T145
W184
L147
H185
H143
Y162
B
Lectin binding (%)
100
75
50
25
0
25000
6250
1563
391
97.7
24.4
PHEV S1A-Fc (pg/μl)
6.1
OC43 S1A-Fc (pg/μl)
wt
Y162A
E182A
W184A
H185A
C
wt
Y162A
E182A
W184A
H185A
PHEV S1A-Fc
OC43 S1A-Fc
D
Lectin binding (%)
S1A-Fc (pg/μl)
BCoV
wt
R143H
I145T
Δ147-151
K181V
L186W
All
OC43
E
BCoV
wt
R143H
K181V
L186W
I145T
STQDG147L
All
OC43
S1A-Fc

Fig. 2. Evidence from structure-guided mutagenesis against site A being the $S1^A$ RBS. (A) (left panel) Cartoon representation of the crystal structure of the BCoV $S1^A$ domain as determined by Peng and colleagues (PDB: 4H14). β-sheets colored green, the α-helix colored red, 3_{10}-helices colored blue, and side chains of site A residues, supposedly critical for 9-*O*-Ac Sia binding [17], indicated in sticks and colored orange. (right panel) Site A close-up in surface representation with side chains of residues, supposedly critical for 9-*O*-Ac-Sia binding, indicated as orange sticks. Side chains of residues that differ between BCoV Mebus and OC43 ATCC $S1^A$ are indicated in cyan. (B) Substitutions of site A residues in PHEV and OC43 $S1^A$-Fc result in partial, but not complete loss of RBS binding affinity. In PHEV and OC43 $S1^A$-Fc, orthologues of BCoV Tyr^{162}, Glu^{182}, Trp^{184} and His^{185} were substituted by Ala (See Table S1 for residue numbering). Mutant proteins were compared with parental (wt) $S1^A$ with maximum binding of BCoV $S1^A$-Fc set at 100%. sp-LBA was as in Fig. 1c, but with data points representing mean averages of independent duplicate experiments. (C) Residual binding of PHEV and OC43 $S1^A$-Fc mutants with Ala substitutions of site A residues as detected by HAA, performed as in Fig. 1b. (D) Substitutions of OC43 site A residues by BCoV orthologues do not increase RBS binding affinity as tested by sp-LBA and (E) by HAA.

Crystallization attempts to identify the $S1^A$ RBS. The most obvious and direct approach to identify the RBS would be by crystallographic analysis of an $S1^A$-ligand complex. In their report [17], Peng and colleagues stated that all efforts to determine the $S1^A$ holo-structure had been unsuccessful. We considered that the complications the authors encountered might be related to crystal packing and therefore opted to perform crystallization trials not with BCoV or OC43 $S1^A$, but with that of PHEV. As its sequence varies from that of BCoV $S1^A$ by 22%, we hoped that crystals of different packing topology would be produced, as was indeed the case. Crystals formed under a variety of conditions, but all with space group P3121 with two $S1^A$ molecules per asymmetric unit. These crystals consistently disintegrated within seconds during soaking attempts with receptor analogue methyl-5-*N*-acetyl-4,9-di-*O*-acetyl-α-neuraminoside. Crystals flash-frozen immediately upon ligand addition showed poor diffraction that did not extend beyond 10 Å resolution. The observations are suggestive of ligand-binding interfering with crystal stability (see also SI Appendix, Fig. S2).

As an alternative, we performed extensive co-crystallization screenings with PHEV $S1^A$, both with native and EndoHf-treated protein samples, together with methyl-5,9-di-*N*-acetyl-α-neuraminoside (Neu5,9Ac2α2Me), a receptor analogue chemically more stable than the 9-*O*-acetylated compound [22, 34]. These attempts also remained unsuccessful as no crystals were formed. In a final effort to obtain crystals of altered packing that perchance would be compatible with ligand soaking, each of the four N-glycosylation sites in PHEV $S1^A$ were systematically removed, separately and in combination, by Asn-to-Gln substitutions. However, the resulting proteins were prone to aggregation and hence unsuitable for crystallization.

Table 1. X-ray data statistics PHEV S1[A].

Data Collection	
Wavelength (Å)	0.9677
Space group	$P3_121$
Cell dimensions a, b, c (Å)	112.57, 112.57, 141.44
α, β, γ (°)	90, 90, 120
Resolution range (Å)*	97.5 - 3.0 (3.07-2.97)
No. unique reflections	20445 (1640)
Redundancy	7.3 (4.1)
Completeness (%)	93.5 (79.9)
R_{merge}	0.154 (1.56)
I/σI	14.0 (1.0)
$CC_{1/2}$	0.994 (0.462) ‡
Refinement	
R_{work} / R_{free}	0.218 / 0.253
No. molecules in the asymmetric unit	2
No. atoms	
Protein	4536
Carbohydrate	193
Water	30
Average B / Wilson B ($Å^2$)	98.1 / 94.5
RMS deviations	
Bond lengths (Å)	0.0078
Bond angles (°)	1.33
NCS-restrained atoms	0.054
Ramachandran plot; favoured, allowed, outliers (%)	92.6, 6.9, 0.5

*Numbers between brackets refer to the outer resolution shell.

‡Diffraction is highly anisotropic: CC1/2 in the outer resolution shell is 0.845 for reflections within 20°from c*, whereas it is 0.00 within 20°from the a*b* plane.

The diffraction data obtained for non-complexed PHEV $S1^A$ eventually allowed us to solve the apo-structure to 3.0 Å resolution (see SI Appendix, Fig. S1). While the results do not provide direct clues to the location of the β1CoV S RBS, they do permit a side-by-side comparison of the $S1^A$ domains of two divergent β1CoVs. The difficulties met by us and others to identify the RBS by crystallography led us to switch strategy and to follow an alternative approach based on comparative structural analysis and visual inspection of $S1^A$ domains using the general design of HE 9-*O*-Ac-Sia binding sites as a query.

An alternative RBS candidate identified by comparative structural analysis. As explicated by Neu et al. (2011), structural comparison of viral attachment proteins in complex with their sialoglycan-based ligands may aid to define common parameters of recognition [35], and in turn these 'rules of engagement' may be used to predict the location of Sia-binding sites in viral proteins for which such structural information is still lacking. Studies by us and others on corona-, toro- and orthomyxoviral HE(F) proteins provide a wealth of structural data on how proteins recognize O-Ac-Sias [18–24]. The *O*-acetyl Sia binding sites as they occur in HE lectin domains and esterase catalytic sites -despite major differences in structure and composition- conform to a common design in which ligand/substrate recognition is essentially based on shape complementarity and hydrophobic interactions, supported by protein-sugar hydrogen bonding involving characteristic Sia functions such as the glycerol side chain, the 5-*N*-acyl moiety and, very often, the carboxylate. Typically, the critical O-acetyl moiety docks into a deep hydrophobic pocket 'P1' and the 5-*N*-acyl in an adjacent hydrophobic pocket/depression 'P2'. P1 and P2, approximately 7 Å apart in 9-*O*-Ac-Sia RBSs, are separated by an aromatic side chain, placed such that in the bound state the side chain intercalates between the *O*- and *N*-acyl groups, often with the ring structure positioned to allow for CH/π interactions with the *O*-acetyl-methyl moiety [18–20, 22–24]. Site A previously proposed as the BCoV $S1^A$ RBS [17] clearly does not conform to this signature. Visual inspection of the PHEV and BCoV $S1^A$ structures, however, identified a region distal from site A, that in many aspects does resemble a typical *O*-Ac-Sia binding site, with two hydrophobic pockets separated by the Trp^{90} indole (Fig. 3a). We will refer to this location as site B. Interestingly, in the BCoV $S1^A$ apo-structure, at the rim of site B, a sulfate ion is bound through hydrogen bonding with Lys^{81} and Thr^{83}. Its presence is of considerable significance as oxoanions in apo-structures often are indicative of and informative on interactions between the RBS and ligand-associated carboxylate moieties [36–38]. Thus, in apo-structures of other Sia-binding proteins, sulfate and phosphate anions were found to mimic the

Sia carboxylate in topology and sugar-protein hydrogen bonding [37, 38]. Automated docking of 9-*O*, 5-*N*-Ac-Sia with the Sia carboxylate anchored at the position of the sulfate ion showed that the ligand would fit into site B, with the 9-*O*-acetyl group, most critical for ligand recognition [15, 39], docking into the more narrow and deeper pocket of the two (Fig. 3b). The 5-*N*-acyl moiety would be accommodated by a hydrophobic patch within the adjacent pocket, which is wide enough to also accept a 5-*N*-glycolyl group.

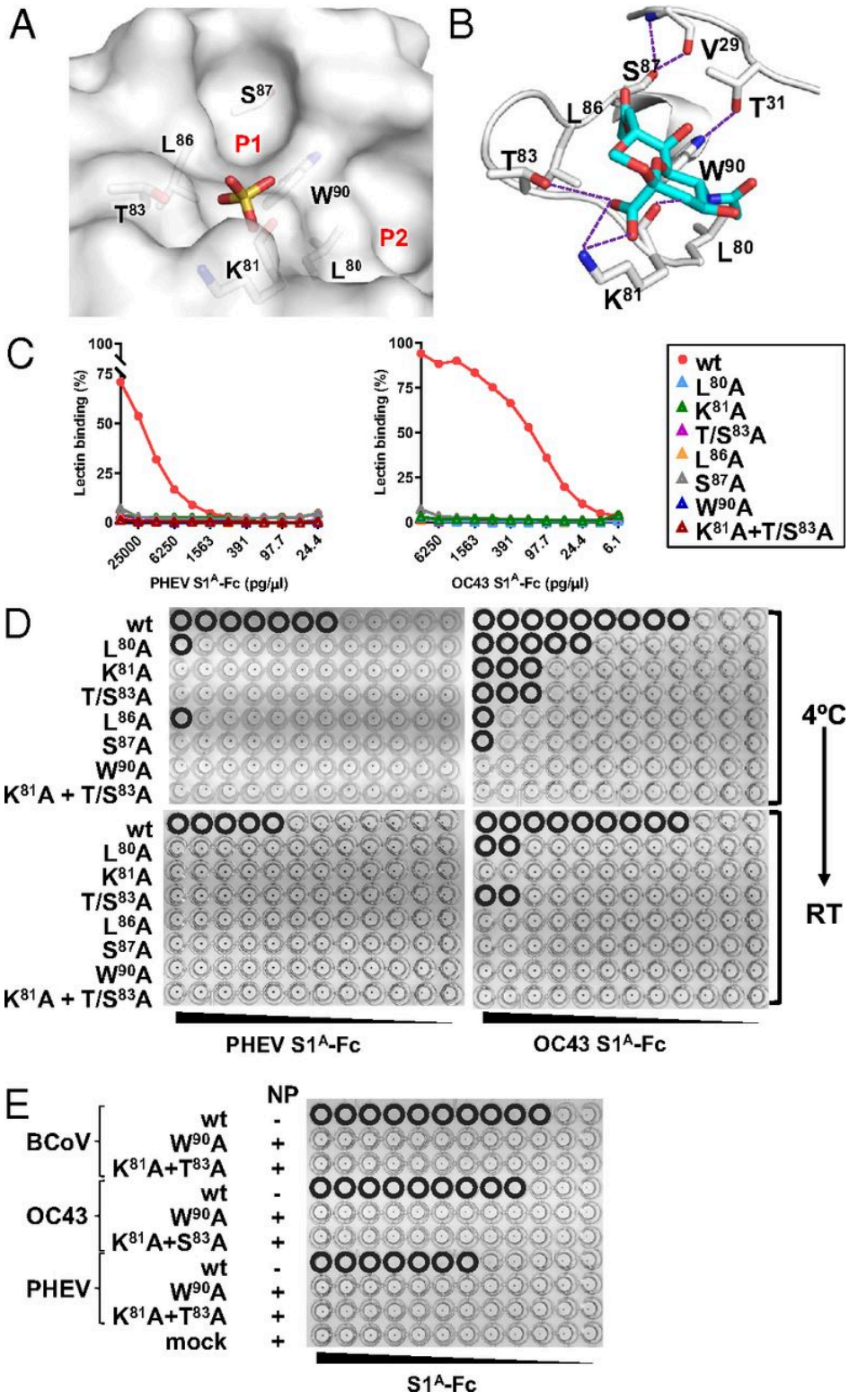


Fig. 3. Evidence from comparative structural analysis and structure-guided mutagenesis to suggest that the S1^{A} RBS locates at site B. (A) Close-up of BCoV S1^{A} site B in surface representation with hydrophobic pockets P1 and P2 indicated, and with side chains of residues, proposedly involved in receptor-binding, shown in sticks and colored by element. A sulfate ion bound to site B in the BCoV S1^{A} apo structure (PDB: 4H14) is also shown (oxygen, red; nitrogen, blue; carbons, gray; sulfur, yellow). Hydrogen bonds between the sulfate ion and site B residues Lys^{81} and Thr^{83} are shown as purple dashed lines. (B) Cartoon representation of BCoV S1^{A} in complex

with 9-*O*-Ac-Sia as modelled by automated molecular docking with Autodock4). Residues predicted to be involved in ligand-binding are shown in stick representation and colored as in A. Sia-9-*O*-Ac is also shown in sticks, but with carbon atoms colored cyan. Predicted hydrogen bonds between Lys^{81} and Thr^{83} side chains and the Sia carboxylate moiety, and between Lys^{81} main chain and the Sia-5-*N*-Ac moiety are shown as purple dashed lines. (C) Substitutions of site B residues in PHEV and OC43 $S1^A$-Fc result in complete loss of binding as detected by sp-LBA, performed as in **Fig. 2b**. (D) (upper panel) Substitutions of site B residues in PHEV and OC43 $S1^A$-Fc result in substantial to complete loss of binding as measured by conventional HAA at 4°C. (lower panel) Mutations, resulting in residual binding, render the $S1^A$ RBS thermolabile. HAA shown in upper panel after a temperature shift-up to 21°C and continued 16 hr-incubation. (E) Substitution of proposed ligand contacting residues Trp^{90}, Lys^{81} and Thr^{83}/Ser^{83} results in total loss of detectable binding of BCoV, OC43 and PHEV $S1^A$-Fc even when assayed by high-sensitivity nanoparticle HAA. Assays performed with $S1^A$-Fc protein multivalently presented on 60-meric Protein A-lumazine synthase icosahedral shells (pA-LS) as in ref 40.

To explain the architecture of Site B, Fig. S1 (SI Appendix) presents the overall structure of BCoV $S1^A$ and the designation of secondary structure elements. Central to $S1^A$ is a β-sandwich scaffold, with one face packing against a separate domain containing the C-terminal region, and the other covered by loop excursions that form a topologically distinct layer. The putative binding site is located within this layer at an open end of the β-sandwich with at its heart the β5-3101 region (residues 80-95). Leu^{80}, Leu^{86}, Trp^{90}, Phe^{91}, Phe^{95} form a well-packed hydrophobic core that on one side interacts with the underlying β-sandwich. On the other side, along the protein surface, it is wrapped by residues from the N-terminal loop 1- β1-loop 2 segment (L1β1L2) to form site B pocket P1. L1β1L2 is locked in position through disulfide bonding of Cys^{21} and β12-residue Cys^{165}, and through extensive intersegment hydrogen bonding with β5-3101 residues, involving Ser^{25} and Asn^{27} via main chain interactions with Leu^{86} and Arg^{88}, Val^{29} via main chain-side chain interaction with Ser^{87}, and Thr^{31} via side chain-side chain interaction with Trp^{90}.

Site B is conserved in all three β1CoVs and, from S cryo-EM structures obtained for other lineage A betacoronaviruses [26, 27], should be readily accessible also in the context of the fully folded intact spike (see SI Appendix, Fig. S3). In the PHEV and BCoV $S1^A$ crystals [17], however, site B -but notably not site A- is occluded by packing contacts, in the case of BCoV $S1^A$ even via coincidental inter-monomeric site B-site B interactions (see SI Appendix, Fig. S2 and S4). Thus, the rapid disintegration of PHEV $S1^A$ crystals during soaking may well be explained by disruption of crystal contacts in result of ligand-binding.

Mutational analysis of β1CoV S1A site B. To test whether site B represents the true RBS, we again performed structure-guided mutagenesis of PHEV and OC43 $S1^A$ (Fig. 3). Of note, the mutant proteins tested were expressed and secreted to wildtype levels, arguing against gross folding defects

(for a quantitative comparison of the expression levels of key mutant proteins and their melting curves, see SI Appendix, Fig. S5). From the docking-model, Trp^{90}, Lys^{81} and Thr^{83} (Ser^{83} in OC43 $S1^{A}$) are predicted to be critically involved in ligand binding. In accordance, their replacement by Ala led to complete loss of detectable binding by sp-LBA (Fig. 3c). Upon substitution of Trp^{90} by Ala, binding was no longer observed neither for PHEV nor for OC43 $S1^{A}$, not even by HAA be-it conventional or high-sensitivity nanoparticle (NP-) HAA [40] (Fig. 3d and 3e). Substitution of either Lys^{81} or Thr^{83} in PHEV S1A led to loss of detectable binding also as measured by conventional HAA at 4°C, but for OC43 $S1^{A}$-Ser^{83}Ala and OC43 $S1^{A}$-Lys^{81}Ala residual binding was detected. Saliently, hemagglutination by OC43 S1A-Lys^{81}Ala fully resolved after a shift-up to room temperature, indicating that the mutation renders the receptor-ligand interaction thermolabile and that binding affinity is more severely affected under physiological conditions (Fig. 3d). Importantly, combined substitution of Lys^{81} and Ser^{83} in OC43 $S1^{A}$ resulted in complete loss of detectable binding even by NP-HAA (Fig. 3e). Also, Ser^{87}Ala substitution, which disrupts the hydrogen bond with Val^{29} and thereby weakens the association between L1β1L2 and β5-3101, caused loss of binding as detected by sp-LBA, and a considerable decrease in binding as measured by HAA. Similar results were obtained upon substitution of Leu^{86} and Leu^{80}, that line the proposed P1 and P2 pockets, respectively (Fig. 3). Finally, replacement of L1β1L2 residue Thr^{31}, thus abolishing the hydrogen bond with β5-3101 residue Trp^{90}, reduced binding affinity in sp-LBA by more than a 1,000-fold as demonstrated for OC43 S1A (see SI Appendix, Fig. S6).

To further test our model, we asked whether we might also identify gain-of-function mutations, i.e. substitutions within or proximal to site B that would increase $S1^{A}$ binding affinity. Among the three β1CoVs, the proposed RBS itself is highly conserved. OC43 and BCoV $S1^{A}$, for instance, differ at site B only at position 83, which is either a Thr (in BCoV) or Ser (in OC43). Despite this near identity, BCoV $S1^{A}$ is the stronger binder of the two as indicated by a consistent 32-fold difference in binding efficiency as measured by sp-LBA (Fig. 1c). The difference at position 83 is particularly intriguing as this residue is predicted to be involved in ligand binding through hydrogen bond formation with the Sia carboxylate (Fig. 3b). Indeed, Ser^{83}Thr substitution in OC43 $S1^{A}$ resulted in a profound increase in binding affinity almost to that of BCoV $S1^{A}$ (Fig. 4a).

Comparative sequence analysis of BCoV and PHEV $S1^{A}$ revealed several amino acid differences proximal to site B, among which Thr^{88}Arg, Thr^{22}Asn and Val^{24}Ser, that would alter hydrogen bonding within L1 and between L1 and β6-3101 (see SI Appendix, Fig. S1). Moreover, the latter two together create an additional N-glycosylation site in PHEV $S1^{A}$. To test whether and how these differences affect receptor binding, PHEV residues were replaced by their BCoV orthologues (Fig. 4b). The mutations again led to gain of function. Thr^{88}Arg substitution, disrupting a Thr^{88}-Asp^{28} hydrogen bond

absent in BCoV $S1^A$, resulted in a consistent two-fold increase in binding affinity as measured by sp-LBA. $Asn^{22}Thr$ substitution, eliminating the glycosylation site in PHEV $S1^A$, raised binding affinity even 50-fold. However, $Ser^{24}Val$ substitution, also destroying the glycosylation site, did not have an effect in isolation. Yet, $Ser^{24}Val$ and $Asn^{22}Thr$ in combination raised PHEV $S1^A$ binding affinity by a further 30-fold almost to that of BCoV (Fig. 4b). Apparently, the difference in $S1^A$ binding affinity between PHEV and BCoV can be ascribed to the architecture of site B as determined by the L1 - β5-3101 hydrogen bonding network rather than to the absence or presence of a glycan at position 22 (Fig. 4b). Changes in $S1^A$ L1 - β5-3101 association may alter pocket P1 and thereby affect ligand binding (see SI Appendix, Fig. S7; note the difference in the topology of Leu^{86} and Trp^{90} side chains in PHEV and BCoV $S1^A$).

One remaining question is how Tyr^{162}, Glu^{182}, Trp^{184} and His^{185} in site A affect the RBS and why their substitution reduces ligand binding affinity, albeit modestly as compared to mutations within site B. Within the structure the sites are spaced relatively closely together and indirect effects of site A mutations can be envisaged. One possible explanation is the location of strand β12', comprising Glu182 and Trp184 right next to the first residues of L1. As we show, mutation of L1 residues Asn^{22}, Val^{24}, Thr^{31} or of residues that interact with L1 like Ser^{87} have a drastic effect on ligand binding. Through a similar mechanism, mutations that destabilize strand β12' may in turn affect L1 and ligand binding at site B.

Summarizing, our analyses revealed both loss-of-function and gain-of-function mutations within or in close proximity of site B, each of which in accord with our model for $S1^A$ binding of 9-*O*-Ac-Sia in β1CoVs. Moreover, they provide a plausible explanation for the modest yet reproducible loss of binding affinity upon substitution of residues previously identified by Peng and coworkers. The results therefore lead us to conclude that site B identified here represents the true $S1^A$ RBS.

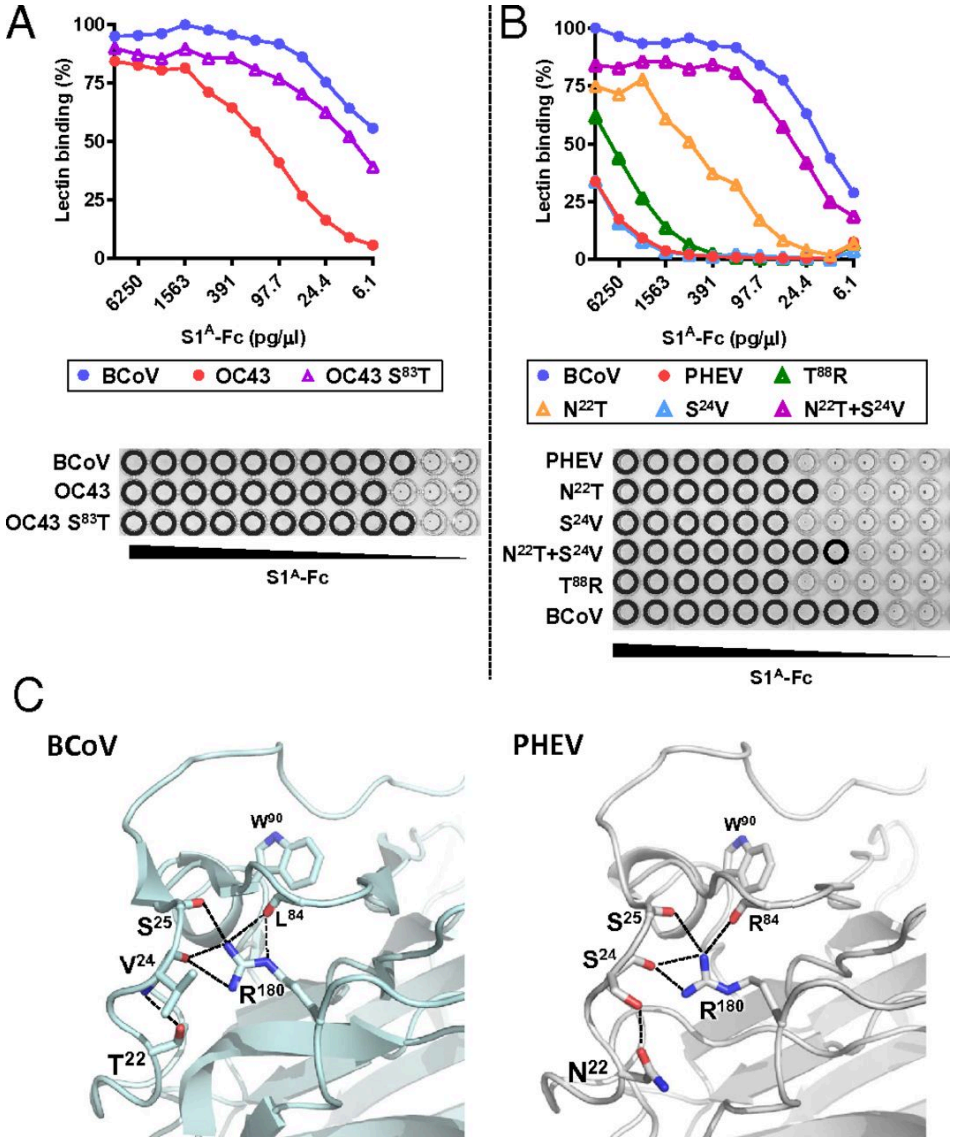
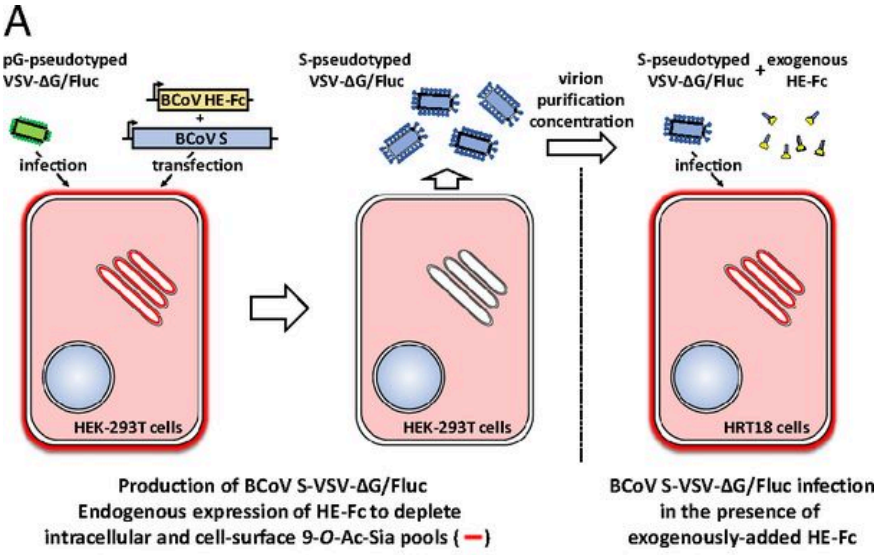
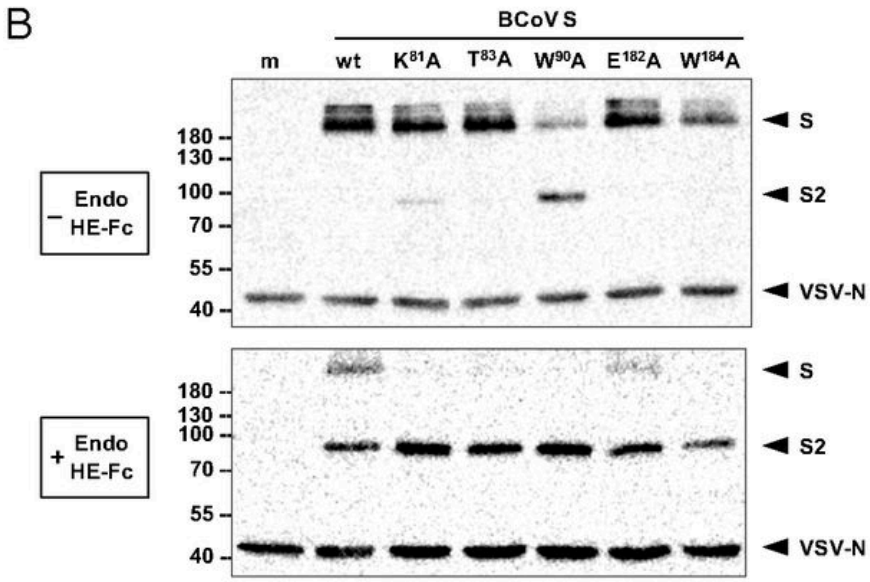


Fig. 4. Increased RBS binding affinity of OC43 and PHEV S1^{A}-Fc upon replacement of site B residues by BCoV orthologues. (A) Comparison of binding affinity of OC43 S1^{A}-Ser^{83}Thr to that of parental OC43 S1^{A} and BCoV S1^{A} by sp-LBA and conventional HAA. (B) Comparison of binding affinity of PHEV S1^{A} mutants (T^{88}R, N^{22}T, S^{24}V individually, and N^{22}T and S^{24}V combined) to that of parental PHEV S1^{A} and BCoV S1^{A} by sp-LBA and conventional HAA. Assays performed and results shown as in Fig. 1. (C) Differences in PHEV and BCoV S1^{A} binding affinity correlate with aa variations in loop L1 and the resulting changes in an intricate site B-organizing hydrogen bonding network. Side-view of S1^{A} site B in BCoV (left) and PHEV (right) in combined cartoon and sticks representation to highlight differences in intra- and intersegment hydrogen bonds that fix the central site B β5-3$_{10}$1 segment through interactions with loop L1 and with the 3$_{10}$2-β12' loop (see SI Appendix, Fig. S1). Note the pivotal role of the 3$_{10}$2-β12' loop Arg^{180} side chain bridging the L1 and β5β6 loops through multiple interactions with main chain carbonyls (depicted in sticks), and the difference between BCoV and PHEV S1^{A} in intra-loop L1 hydrogen bonding between residues 24 and 26.


A
pG-pseudotyped VSV-ΔG/Fluc
BCoV HE-Fc
+
BCoV S
infection
transfection
HEK-293T cells
S-pseudotyped VSV-ΔG/Fluc
virion purification concentration
HEK-293T cells
S-pseudotyped VSV-ΔG/Fluc + exogenous HE-Fc
infection
HRT18 cells
Production of BCoV S-VSV-ΔG/Fluc
Endogenous expression of HE-Fc to deplete intracellular and cell-surface 9-O-Ac-Sia pools (—)
BCoV S-VSV-ΔG/Fluc infection in the presence of exogenously-added HE-Fc


B
BCoV S
m
wt
K81A
T83A
W90A
E182A
W184A
180
130
100
70
55
40
– Endo HE-Fc
+ Endo HE-Fc
S
S2
VSV-N

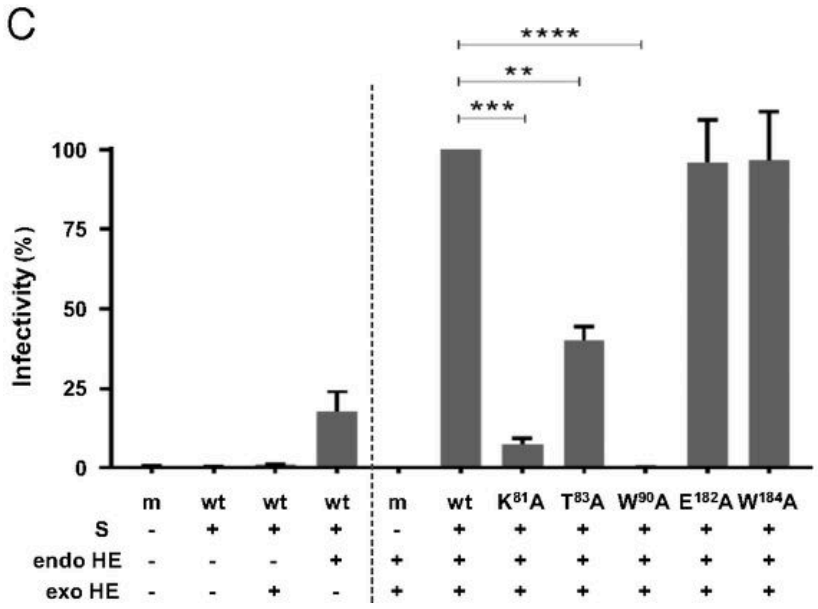

C
****
**
***
Infectivity (%)
100
75
50
25
0
m
wt
wt
wt
m
wt
K81A
T83A
W90A
E182A
W184A
S
endo HE
exo HE

Fig. 5. Mutations in site B but not site A reduce infectivity of BCoV S-pseudotyped VSV particles. (A) Schematic outline of the production of BCoV S-pseudotyped VSV-ΔG/Fluc and of the infection experiments. (B) Protein composition of (mock) pseudotyped VSV particles. Western blot analysis of purified and concentrated virus particles, secreted by mock-transfected, VSVΔG/Fluc-transduced cells (m) or by transduced cells, transfected to transiently express Flag-tagged BCoV strain Mebus S (wildtype, 'wt') or mutants thereof. BCoV S was either expressed exclusively (- Endo HE-Fc; top panel) or co-expressed with HE-Fc (+ Endo HE-Fc; bottom). Arrowheads indicate noncleaved S (S), the C-terminal subunit S2 resulting from intracellular proteolytic cleavage of S (S2), and the N protein of VSV. Note that co-expression of HE-Fc promotes cleavage of BCoV S, presumably by preventing inadvertent receptor association during S biosynthesis. Also note that in the absence of endogenous HE-Fc, site B mutants are cleaved to extents inversely related to RBS affinity. (C) Infectivity assays with BCoV S-pseudotyped viruses. Virus particles, pseudotyped with BCoV S or mutants thereof and produced in the presence or absence of intracellular HE-Fc (endo HE) were used to inoculate HRT18 cells, either with (+) or without (-) exogenous HE (exo HE) added to the inoculum. "Infectivity" is expressed in relative light units (RLU) detected in lysates from inoculated cells at 18 hr after inoculation. RLU values were normalized with those measured for VSV-S^{wt} set at 100%. The data shown are averages from three independent experiments, each of which performed with technical triplicates. Standard deviations and significant differences, calculated by Welch's unequal variances *t*-test, are indicated (**, P≤0.01; ***, P≤0.001, ****, p≤0.0001).

Site B is essential for infectivity. To directly test the relevance of sites A and B for infectivity, we pseudotyped G-protein-deficient, luciferase-expressing recombinant vesicular stomatitis virus (VSV) with Flag-tagged BCoV S or with mutant derivatives (see Fig. 5a for a schematic outline of the experiment). Considering that for β1CoVs, HE is an essential protein [10, 16], we expressed wildtype BCoV S either exclusively or together with a secretory HE-Fc fusion protein. Because HE-Fc is not membrane-associated, it will not be included in the VSV envelope, but would deplete endogenous 9-*O*-Ac-Sia pools in the exocytotic compartment and from the cell surface. Thus, virus release would be facilitated, and inadvertent receptor association during S biosynthesis and receptor-mediated virion aggregation would be prevented. Wildtype S in the absence of HE-Fc was incorporated in VSV particles in its non-cleaved 180K form (Fig. 5b, top). However, the resulting particles, purified and concentrated from the tissue culture supernatant by sucrose cushion ultracentrifugation, were non-infectious as measured by luciferase assay (Fig. 5c). VSV particles, produced in cells that co-expressed S and HE-Fc, carried spikes that resembled those described for BCoV and OC43 virions in that a large proportion had been proteolytically cleaved into subunits S1 and S2 (Fig. 5b, bottom) [41, 42]. Moreover, these particles were infectious. Values of relative light units (RLUs) detected in inoculated cells were more than 40-fold above background. Strikingly, their infectivity was increased by a further 5.7-fold upon addition of 'exogenous' HE-Fc to the inoculum (Figs. 5a, 5c). Hence, these conditions were chosen to determine the effects of mutations in BCoV S. Site B mutations caused

complete loss of infectivity ($Trp^{90}Ala$), or significantly reduced infectivity to extents correlating with the effect of each mutation on S binding affinity ($Trp^{90}Ala$ > $Lys^{81}Ala$ > $Thr^{83}Ala$; see Figs. 3, 5c). Particles pseudotyped with site A mutants were as infectious as S^{wt}-VSV. The combined data reinforce our conclusion that site B is the RBS and provide direct evidence that this site is essential for BCoV infectivity.

HCoV-HKU1 also binds to 9-*O*-Ac-Sia via $S1^A$ site B. HKU1 is a lineage A betacoronavirus but separated from the β1CoVs by a considerable evolutionary distance. To illustrate, $S1^A$ of HKU1 is only 55-60% identical to those of BCoV and OC43. Yet, HKU1, like the β1CoVs, uses 9-*O*-Ac-sialoglycans for attachment and the binding to these receptors is essential for infection [14]. According to recent publications, the RBS would not be in $S1^A$, but in a downstream domain [31, 32], the main argument being that HKU1 $S1^A$ does not detectably bind to glycoconjugates [30]. Indeed, we also did not detect binding of HKU1 $S1^A$-Fc, at least not by standard sp-LBA and conventional HAA (Fig. 6a, e). However, when HKU1 $S1^A$-Fc was tested by high-sensitivity-NP-HAA with multivalent presentation of the HKU1 $S1^A$-Fc proteins, agglutination of rat erythrocytes was observed (Fig. 6a). Moreover, depletion of cell surface 9-*O*-Ac-Sia receptors by pretreatment of the erythrocytes with BCoV HE completely prevented hemagglutination (Fig. 6a). These findings conclusively demonstrate that, in fact, HKU1 $S1^A$ does bind to 9-*O*-Ac-Sia in a 9-*O*-Ac-dependent fashion apparently through low-affinity high-specificity interactions. The detection of such interactions critically depends on multivalency of receptors and receptor-binding proteins [10, 40], and thus they are easily missed. This requirement for multivalency may also explain why preincubation of cells with soluble $S1^A$-Fc fusion proteins did not block HKU1 infection [32].

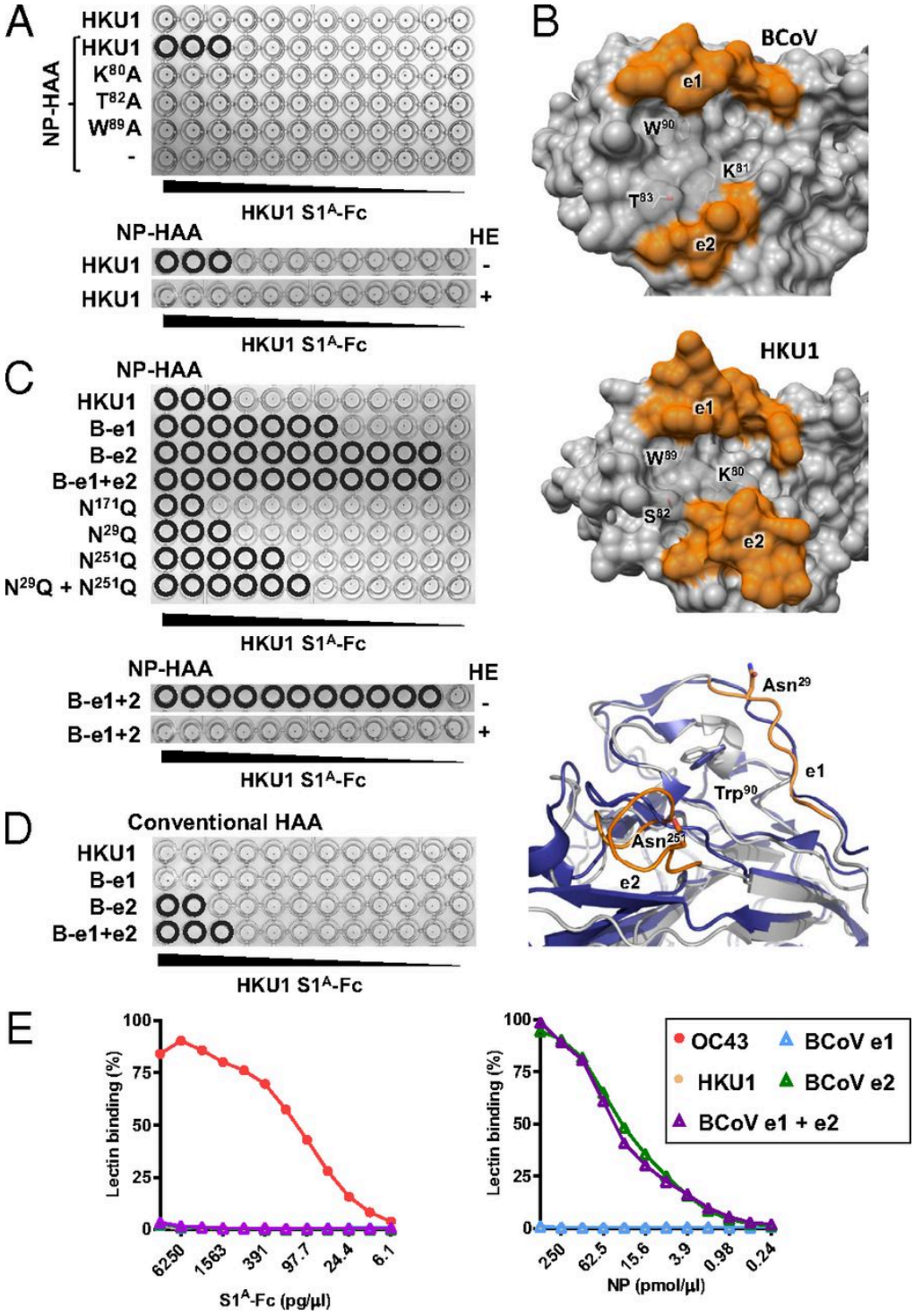


Fig. 6. HKU1 S1^{A} site B mediates binding to 9-*O*-Ac-Sias. (A) HKU1 S1^{A}-Fc binds to 9-*O*-Ac-Sia through low-affinity, high-specificity interactions. (upper panel) Conventional HAA (top row) and high-sensitivity pA-LS nanoparticle HAA (NP-HAA, bottom rows) with native HKU1 S1^{A}-Fc (HKU1) or mutants thereof (K^{80}A, T^{82}A, W^{89}A). Non-complexed nanoparticles were included as negative control (-). (lower panel) NP-HAA with erythrocytes (mock)depleted for 9-*O*-Ac-Sias by prior sialate-*O*-acetylesterase treatment. (B) Surface representation of site B in BCoV (top panel) and HKU1 (middle panel) (PDB: 5I08 [27]), respectively. Protein structures were aligned with site B residues Lys^{81}, Thr^{83} and Trp^{90} as query. Side chains of site B residues shown

in sticks. BCoV residues 29-34 (element e1) and 246-253 (e2), and corresponding elements in HKU1 indicated in orange. (bottom panel) Overlay of cartoon representations of BCoV (purple) and HKU1 (white) $S1^A$ centered on site B. In the HKU1 protein, e1 and e2 are colored orange and marked. Side chains of Asn^{29} and Asn^{251}, acceptors of *N*-linked glycosylation, indicated in sticks. BCoV Trp^{90} and its HKU1 orthologue, also in sticks, shown as a reference point to site B. (C) (upper panel) Replacement of HKU1 $S1^A$ e1 and e2 by corresponding BCoV elements increases binding affinity to 9-*O*-Ac-sialoglycans present on rat erythrocytes. NP-HAA with native HKU1 $S1^A$-Fc (HKU1) or with HKU1 $S1^A$-Fc derivatives, with e1 and e2 replaced by the corresponding BCoV elements separately (B-e1, B-e2) or in combination (B-e1+e2). For comparison, mutant HKU1 $S1^A$-Fc proteins were included with *N*-glycosylation sites eliminated in site A ($N^{171}Q$) or in elements e1 ($N^{29}Q$) and e2 ($N^{251}Q$), individually and in combination ($N^{29}Q$ + $N^{251}Q$). (lower panel) NP-HAA with erythrocytes (mock)depleted for 9-*O*-Ac-Sias. (D) Replacement of HKU1 $S1^A$ e1 and e2 by corresponding BCoV elements increases binding affinity to 9-*O*-sialoglycans as to allow detection by conventional HAA. (E) Replacement of HKU1 $S1^A$ e1 and e2 increases binding affinity as to allow detection of 9-*O*-Ac-sialoglycans by nanoparticle sp-LBA. (left) Conventional sp-LBA with soluble $S1^A$-Fc. (right) Nanoparticle sp-LBA with $S1^A$-Fc complexed to pA-LS icosahedral shells in 2-fold serial dilutions, starting at 50 nmol pA-LS/well. Data points are mean averages of independent duplicate experiments, each performed in triplicate.

Interestingly, the RBS identified for β1CoV $S1^A$ is conserved in HKU1 in sequence and structure (Fig. 6b; see SI Appendix, S1b and S8), raising the question as to whether binding of HKU1 S to its sialoglycan receptor occurs through the same site. Indeed, Ala substitution of HKU1 $S1^A$ Lys^{80}, Thr^{82} or Trp^{89} -orthologues of Lys^{81}, Thr^{83} and Trp^{90} in BCoV $S1^A$- resulted in a complete loss of detectable binding (Fig. 6a; see SI Appendix, Fig. S9 for an analysis in the context of S1-Fc). To further study whether this site is involved in 9-*O*-Ac-Sia binding and to understand the structural basis for the considerable difference in affinity between HKU1 and β1CoV $S1^A$ domains, we performed a side-by-side structure comparison. While in β1CoV $S1^A$ domains the RBS is readily accessible, the corresponding site in HKU1 $S1^A$ is located at the bottom of a canyon as it is flanked by protruding parallel ridges comprised of HKU1 S1A residues 28 through 34 (element (e)1 corresponding to BCoV L2β2-residues 29-35), and 243 through 252 (e2 corresponding to BCoV residues 253-256 in the β18-β19 loop), each decorated with an N-linked glycan (Fig. 6b; see also SI Appendix, S1b)[27]. Conceivably, this assembly would hamper binding particularly to short sialoglycans such as the O-linked STn sugars (Sia-α2,6GalNAc-α1-O-Ser/Thr) predominantly present on BSM [43]. Individual exchange of e1 and e2 in HKU1 $S1^A$ by the corresponding segments from BCoV indeed increased receptor-binding as measured by NP-HAA by 16- or 256-fold for e1 and e2, respectively (Fig. 6c). Exchange of e2 in fact enhanced receptor affinity to levels that now also allowed detection of binding by conventional HAA, while simultaneous replacement of both elements increased binding affinity even further (Fig. 6d). Importantly, pretreatment of the erythrocytes with BCoV HE to deplete

receptors by Sia-de-*O*-acetylation prevented hemagglutination (Fig. 6c), indicating that the enhanced binding was still 9-*O*-Ac-Sia-dependent. Replacement of e2 increased receptor affinity to such extent that also binding to BSM by sp-LBA assay became detectable, albeit only with nanoparticle-bound and not with free $S1^A$ (Fig. 6e).

To study whether the N-glycans attached to e1 and e2 hinder receptor-binding and thus contribute to the apparent low affinity of HKU1 $S1^A$, we expressed the protein in N-acetylglucosaminyltransferase I-deficient HEK293S $GnTI^-$ cells [44]. Replacement of complex N-glycans by high-mannose sugars resulted in an 8-fold increase in binding affinity as measured by NP-HAA (see SI Appendix, Fig. S9). In agreement, combined disruption of the N-linked glycosylation sites in e1 and e2 through Gln substituting for Asn^{29} and Asn^{251} increased binding affinity to similar extent, with the latter mutation exerting the largest effect (Fig. 6c). The deletion of glycosylation sites, however, did not enhance affinity to that of the HKU1-BCoV S1A/e1+e2 chimera, indicating that receptor binding is hampered not only by the N-linked glycans, but also by the local protein architecture. Of note, removal of the glycan attached to Asn^{171}, proximal to the binding site originally predicted by Peng et al. [17] did not enhance but, actually, lower the binding affinity of HKU1 $S1^A$ (Fig. 6c) possibly by affecting protein folding.

In summary, our results provide direct evidence that HKU1 binds to 9-*O*-Ac-Sias via S domain A. The data again argue against a role for site A [17], and on the basis of loss and gain of function mutations strongly indicate that HKU1 binds it sialoglycan receptor via the novel site B here identified for β1CoVs. On a critical note, we are aware that in the published HKU1 S apo-structure, the orientation of the Lys81 side chain differs from that in BCoV $S1^A$ such that hydrogen bonding with the Sia carboxylate would be precluded. Whether this is an inaccuracy in the structure, whether the orientation of the side chain changes during ligand binding, or whether this is indeed a factual difference between HKU1 and β1CoV RBSs remains to be seen. Be-it-as-it-may, our findings do imply that in two human coronaviruses, related yet separated by a considerable evolutionary distance, the RBS and general mode of receptor-binding has been conserved. It is tempting to speculate that the apparently low affinity of HKU1 $S1^A$ is a virus-specific adaptation to replication in the human tract. Conceivably, occlusion of the HKU1 S RBS by e1 and e2 might allow selective high-affinity binding to particular sialoglycans with the 9-*O*-Ac-Sia terminally linked to extended glycan chains as to prevent non-productive binding to short stubby sugars -such as present on BSM and other mucins- that otherwise would act as decoys and inhibitors. In analogy, influenza A H3N2 variants and 2009 pandemic H1N1 (Cal/04), initially assumed to carry low affinity HAs, were recently reported to have evolved a preference for a subset of human-type α2,6-linked sialoglycan-based receptors comprising branched sugars with extended poly-*N*-acetyl-lactosamine (poly-LacNAc) chains [45]. Whether such

an adaptation would translate in distinctive differences between OC43 and HKU1 with respect to dynamic virion-receptor interactions and cell tropism merits further study.

Materials & Methods

Protein design, expression and purification. 5'-terminal sequences of the S genes of BCoV strain Mebus (GB: P15777.1), OC43 strain ATCC VR-759 (GB: AAT84354.1), PHEV strain UU (GB: ASB17086.1), HKU1 strain Caen1 (GB: ADN03339.1) and mouse hepatitis virus (MHV) strain A59 (GB: P11224.2), encoding the signal peptide and adjacent S1A domain, were cloned in expression vector pCAGGS-Tx-Fc [46]. Domain A, defined on the basis of the cryoEM structure of the MHV-A59 S ectodomain [26], corresponds to S aa residues 15-294, 15-298 and 15-294 of BCoV, OC43 and PHEV, respectively. The amino acid sequences of these S1A-Fc fusion proteins were deposited at GenBank (MG999832-35). Recombinant $S1^A$ proteins, genetically fused to the Fc domain of human IgG, were purified by protein A affinity chromatography from the supernatants of transiently transfected HEK293T cells as in [18].

Hemagglutination Assay (HAA). Standard HAA was performed as in [18]. Two-fold serial dilutions of CoV $S1^A$-Fc proteins (starting at 2.5 µg/well; 50 µl/well) were prepared in V-shaped 96-well microtiter plates (Greiner Bio-One) unto which was added 50 µl of a rat erythrocyte suspension (Rattus norvegicus strain Wistar; 0.5% in Phosphate Buffered Saline (PBS)). High sensitivity nanoparticle HAA (NP-HAA) was performed as in [40]. In brief, $S1^A$-Fc proteins were complexed with pA-LS (a self-assembling 60-meric lumazine synthase nanoparticle, N-terminally extended with the immunoglobulin Fc-binding domain of the S. aureus protein A) at a 0.6:1 molar ratio for 30 min on ice, after which complexed proteins were 2-fold serially diluted and mixed 1:1 with rat erythrocytes (0.5% in PBS). Hemagglutination was assessed after 2 hr incubation on ice unless stated otherwise. For specific depletion of cell-surface O-Ac-Sias, erythrocytes (50% in PBS pH 8.0) were (mock-) treated with soluble HE+ (0.25 µg/µl BCoV-Mebus and 0.25 µg/µl PToV-P4) for 4 hr at 37°C comparable to as described in [47].

Solid-Phase Lectin-Binding Assay (sp-LBA]. 96-Well Maxisorp® microtitre ELISA plates (Nunc) were coated with 0.1 µg bovine submaxillary mucin (BSM; Sigma-Aldrich) per well. Conventional sp-LBAs were performed using 2-fold serial dilutions of CoV $S1^A$-Fc proteins as described [20, 47]. In nanoparticle sp-LBA experiments, BSM-bound nanoparticles were detected with StrepMAb-HRP via the C-terminally appended Strep-tag of the pA-LS proteins as described previously [40]. Receptor-depletion assays were performed by (mock-)treatment of coated BSM with 2-fold serial dilutions of soluble HE^+ in PBS (100 µl/well) starting at 0.6 ng/µl (BCoV-Mebus) for 2 hr at 37°C as in

[47]. Receptor destruction was assessed by sp-LBA with β1CoV S1A-Fc at fixed concentrations (BCoV 0.3 ng/µl; OC43 1.2 ng/µl).

Glycoside Synthesis and NMR Analysis. The starting material Neu5Ac2αMe [48] was treated with p-toluenesulfonyl chloride (1.8 eq.) in pyridine overnight and then with sodium azide (5 eq.) in DMF at 70°C. The resulting intermediate was subjected to Staudinger reduction by 1M trimethylphosphine in toluene (3 eq.) in the presence of potassium hydroxide. Upon completion, acetyl chloride (4 eq.) was added directly. After 5 min, 1M potassium hydroxide solution was added to quench the reaction. The mixture was neutralized with acidic resin. The residue was purified with silica gel and then p-2 biogel to give pure Neu5,9NAc22αMe at a yield of 31% over 4 steps. The product was dissolved in 50 mM ammonium bicarbonate and freeze-dried, which was repeated three times to give the ammonium salt form. The final product was analyzed by nuclear magnetic resonance (NMR) spectroscopy (see SI Appendix, Fig. S10).

Crystallization. Crystallization conditions were screened by the sitting-drop vapor diffusion method using a Gryphon (Art Robins]. Drops were set up with 0.15 µl of S1A dissolved to 11 mg/ml in 10 mM BisTris-propane, 50 mM NaCl, pH 6.5 and 0.15 µl reservoir solution at room temperature. Diffracting crystals were obtained from the JCSG+ screen (200 mM MgCl2, 100 mM BisTris pH 5.5, 25% (w/v) PEG3350) (JCSG Technologies) at 18°C. Crystals were flash-frozen in liquid nitrogen using reservoir solution with 30% (w/v) glycerol as the cryoprotectant. Crystals were soaked with methyl-5,9-di-*N*-acetyl-α-neuraminoside and methyl-5-*N*-acetyl-4,9-di-*O*-acetyl-α-neuraminoside as described, for the latter ligand, with batches previously used to solve HE holo-structures [18–20, 22].

Data Collection and Structure Solution. Diffraction data of PHEV $S1^A$ crystals were collected at European Synchrotron Radiation Facility station ID30A-3. Diffraction data were processed using XDS [49] and scaled using Aimless from the CCP4 suite [50]. Molecular replacement was performed using PHASER [51] with BCoV $S1^A$ residues 35–410 as template (Protein Data Bank (PDB) accession code 4H14). NCS-restrained structure refinement was performed in REFMAC [52], alternated with model building in Coot [53]; water molecules were built in difference density peaks of at least 5.0 σ located at buried sites conserved and occupied by water in the 1.5 Å-resolution structure of BCoV $S1^A$. Molecular graphics were generated with PYMOL (http://pymol.sourceforge.net).

Molecular Docking. Molecular docking of 9-*O*-Sia in the crystal structure of apo-BCoV $S1^A$ (PDB ID code 4H14) was performed with AutoDock4 [22] similar to the procedure described by Bakkers et al. [22]. The ligand molecule was extracted from BCoV HE (PDB ID code 3CL5). During docking, the protein was considered to be rigid. An inverted Gaussian function (50-Å half-width; 15-kJ energy at infinity) was used to restrain the Sia-carboxylate near the position occupied by the sulfate ion in the

BCoV $S1^A$ crystal structure. The initial ligand conformation was randomly assigned, and 10 docking runs were performed.

Preparation of BCoV S-pseudotyped vesicular stomatitis virus particles and infection experiments. Recombinant G-protein-deficient VSV particles were pseudotyped as described [54] with the S protein of BCoV strain Mebus. To allow transport of S to the cell surface and its incorporation into VSV particles, a truncated version of S was used from which the C-terminal 17 residues, comprising an ER-retention signal, had been removed. To facilitate detection, the S protein was provided with a Flag-tag by cloning its gene in expression vector pCAGGS-Flag. HEK 293T cells at 70% confluency were transfected with PEI-complexed plasmid DNA as described [18]. For co-expression of BCoV S and HE-Fc, S expression vectors and pCD5-BCoV HE-Fc [18] were mixed at molar ratios of 8:1. At 48 hr after transfection, cells were transduced with VSV-G pseudotyped VSVΔG/Fluc [54] at MOI 1. Cell-free supernatants were harvested at 24 hr after transduction, filtered through 0.45 µm membranes and virus particles were purified and concentrated by sucrose cushion ultracentrifugation at 29,000 rpm for 3 hr [10]. Pelleted virions were resuspended in PBS and stored at -80°C until further use. Relative virion yields were determined on the basis of VSV-N content by Western blot analysis by using anti-VSV-N monoclonal antibody 10G4 (kerafast). Uptake of BCoV S into virus particles was detected by Western blot analysis with monoclonal antibody ANTI-FLAG® M2 (sigma). Inoculation of HRT18 monolayers in 96-well cluster format was performed with equal amounts of S-pseudotyped VSVs, as calculated from VSV-N content (roughly corresponding to the yield from $2x10^5$ cells from each transfected and transduced culture), diluted in 10% fetal bovine serum-supplemented DMEM. For virus infections with 'exogenous' HE, 12.5 mU/ml of BCoV HE-Fc protein was added to the inoculum. At 18 hours post infection, cells were lysed using passive lysis buffer (Promega]. Firefly luciferase expression was measured using a homemade firefly luciferase assay system as described [55]. Infection experiments were performed independently in triplicate, each time with three technical replicates.

Acknowledgements

We acknowledge the European Synchrotron Radiation Facility for providing beamline facilities and the beamline scientists at ID30A-3 for help with data collection. We thank Louis-Marie Bloyet for his help in preparing the structure heat map conservation figure. This study was supported by TOP Project Grant 40-00812-98-13066 funded by ZonMW provided to FJMvK and BJB, CSC grant 2014-03250042 to YL and by ECHO grant 711.011.006 of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW) to RJdG.

References

1. Corman VM, et al. (2016) Link of a ubiquitous human coronavirus to dromedary camels. *Proc Natl Acad Sci U S A* 113(35):9864–9869.
2. Vijgen L, et al. (2006) Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* 80(14):7270–7274.
3. Vijgen L, et al. (2005) Complete Genomic Sequence of Human Coronavirus OC43 : Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event Complete Genomic Sequence of Human Coronavirus OC43 : Molecular Clock Analysis Suggests a Relatively Recent. *J Virol* 79(3):1595–1604.
4. Huynh J, et al. (2012) Evidence Supporting a Zoonotic Origin of Human Coronavirus Strain NL63. *J Virol* 86(23):12816–12825.
5. Tao Y, et al. (2017) Surveillance of Bat Coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history. *J Virol* 91(5):1–16.
6. Yuen K, Lau S, Woo P (2012) Wild animal surveillance for coronavirus HKU1 and potential variants of other coronaviruses. *Hong Kong Med J* 18(2):25–26.
7. Woo PC, et al. (2005) Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. *J Infect Dis* 192(September):1898–1907.
8. McIntosh K, et al. (1970) Seroepidemiologic studies of coronavirus infection in adults and children. *Am J Epidemiol* 91(6):585–92.
9. Morfopoulou S, et al. (2016) Human Coronavirus OC43 Associated with Fatal Encephalitis. *N Engl J Med* 375(5):497–498.
10. Bakkers MJG, et al. (2017) Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. *Cell Host Microbe* 21(3):356–366.
11. Hulswit RJG, de Haan CAM, Bosch B-J (2016) Coronavirus Spike Protein and Tropism Changes. *Adv Virus Res* 96:29–57.
12. Heald-Sargent T, Gallagher T (2012) Ready, set, fuse! the coronavirus spike protein and acquisition of fusion competence. *Viruses* 4(4):557–580.
13. Matrosovich M, Herrler G, Klenk HD (2015) Sialic Acid Receptors of Viruses. *Top Curr Chem* 367:1–28.
14. Huang X, et al. (2015) Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J Virol* 89(14):7202.
15. Vlasak R, Luytjes W, Spaan W, Palese P (1988) Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci U S A* 85(12):4526–4529.
16. Desforges M, Desjardins J, Zhang C, Talbot PJ (2013) The acetyl-esterase activity of the hemagglutinin-esterase protein of human coronavirus OC43 strongly enhances the production of infectious virus. *J Virol* 87(6):3097–107.
17. Peng G, et al. (2012) Crystal structure of bovine coronavirus spike protein lectin domain. *J Biol Chem* 287(50):41931–41938.
18. Zeng Q, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ (2008) Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc Natl Acad Sci U S A* 105(26):9065–9069.
19. Langereis MA, et al. (2009) Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc Natl Acad Sci U S A* 106(37):15897–15902.
20. Langereis MA, Zeng Q, Heesters B, Huizinga EG, de Groot RJ (2012) The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. *PLoS Pathog* 8(1):1–8.
21. Cook JD, Sultana A, Lee JE (2017) Structure of the infectious salmon anemia virus receptor complex illustrates a unique binding strategy for attachment. *Proc Natl Acad Sci*:201617993.
22. Bakkers MJG, et al. (2016) Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proc Natl Acad Sci U S A* 113(22):E3111-3119.
23. Song H, et al. (2016) An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from

Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLoS Pathog* 12(1). doi:10.1371/journal.ppat.1005411.

24. Rosenthal PB, et al. (1998) Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* 396(6706):92–96.
25. Walls AC, et al. (2016) Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat Struct Mol Biol* 23(10):899–905.
26. Walls AC, et al. (2016) Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531(7592):114–117.
27. Kirchdoerfer RN, et al. (2016) Pre-fusion structure of a human coronavirus spike protein. *Nature* 531(7592):118–121.
28. Yuan Y, et al. (2017) Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun* 8:15092.
29. Gui M, et al. (2016) Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res* 27(1):119–129.
30. Peng G, et al. (2011) Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci U S A* 108(26):10696–10701.
31. Ou X, et al. (2017) Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat Commun* 8(May):15216.
32. Qian Z, et al. (2015) Identification of the Receptor-Binding Domain of the Spike Glycoprotein of Human Betacoronavirus HKU1. *J Virol* 89(17):8816–8827.
33. Li F (2016) Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 3(1):237–261.
34. Brossmer R, Isecke R, Herrler G (1993) A sialic acid analogue acting as a receptor determinant for binding but not for infection by influenza C virus. *FEBS Lett* 323(1–2):96–98.
35. Neu U, Bauer J, Stehle T (2011) Viruses and sialic acids: Rules of engagement. *Curr Opin Struct Biol* 21(5):610–618.
36. Kaiser JT, et al. (2000) Crystal Structure of a NifS-like Protein from Thermotoga maritima : Implications for Iron Sulphur Cluster Assembly. doi:10.1006/jmbi.2000.3581.
37. Pronker MF, et al. (2016) Structural basis of myelin-associated glycoprotein adhesion and signalling. *Nat Commun* 7(13584):1–13.
38. van der Vries E, et al. (2012) H1N1 2009 Pandemic Influenza Virus: Resistance of the I223R Neuraminidase Mutant Explained by Kinetic and Structural Analysis. *PLoS Pathog* 8(9):1–8.
39. de Groot RJ (2006) Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj J* 23(1–2):59–72.
40. Li W, et al. (2017) Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci* 114(40):E8508–E8517.
41. King B, Brian DA (1982) Bovine Coronavirus Structural Proteins. *J Virol* 42(2):700–707.
42. Hogue BG, King B, Brian DA (1984) Antigenic Relationships Among Proteins of Bovine Coronavirus, Human Respiratory Coronavirus OC43, and Mouse Hepatitis Coronavirus A59. *J Virol* 51(2):384–388.
43. Kozak RP, Royle L, Gardner RA, Fernandes DL, Wuhrer M (2012) Suppression of peeling during the release of O-glycans by hydrazinolysis. *Anal Biochem* 423(1):119–128.
44. Reeves PJ, Callewaert N, Contreras R, Khorana HG (2002) Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc Natl Acad Sci* 99(21):13419–13424.
45. Peng W, et al. (2017) Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. *Cell Host Microbe* 21(1):23–34.
46. Li W, et al. (2003) Angiotensin-converting enzyme 2 is a functional receptor for SARS coronavirus. *Nature* 426(6965):450–454.
47. Langereis MA, et al. (2015) Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep* 11(12):1966–1978.
48. Lubineau A, Gallic J Le (1991) Stereoselective Syntheses of ALKYL- and ALKYL-2-THIO-α-Sialosides. *J Carbohydr*

Chem 10(2):263–268.

49. Kabsch W (2010) XDS. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):125–132.

50. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.

51. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(4):658–674.

52. Vagin AA, et al. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr Sect D Biol Crystallogr* 60(12 I):2184–2195.

53. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* 66(4):486–501.

54. Park J-E, et al. (2016) Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc Natl Acad Sci* 113(43):12262–12267.

55. Burkard C, et al. (2014) Coronavirus Cell Entry Occurs through the Endo-/Lysosomal Pathway in a Proteolysis-Dependent Manner. *PLoS Pathog* 10(11):e1004502.

56. Langereis MA, van Vliet ALW, Boot W, de Groot RJ (2010) Attachment of mouse hepatitis virus to O-acetylated sialic acid is mediated by hemagglutinin-esterase and not by the spike protein. *J Virol* 84(17):8970–4.

Supplementary Information

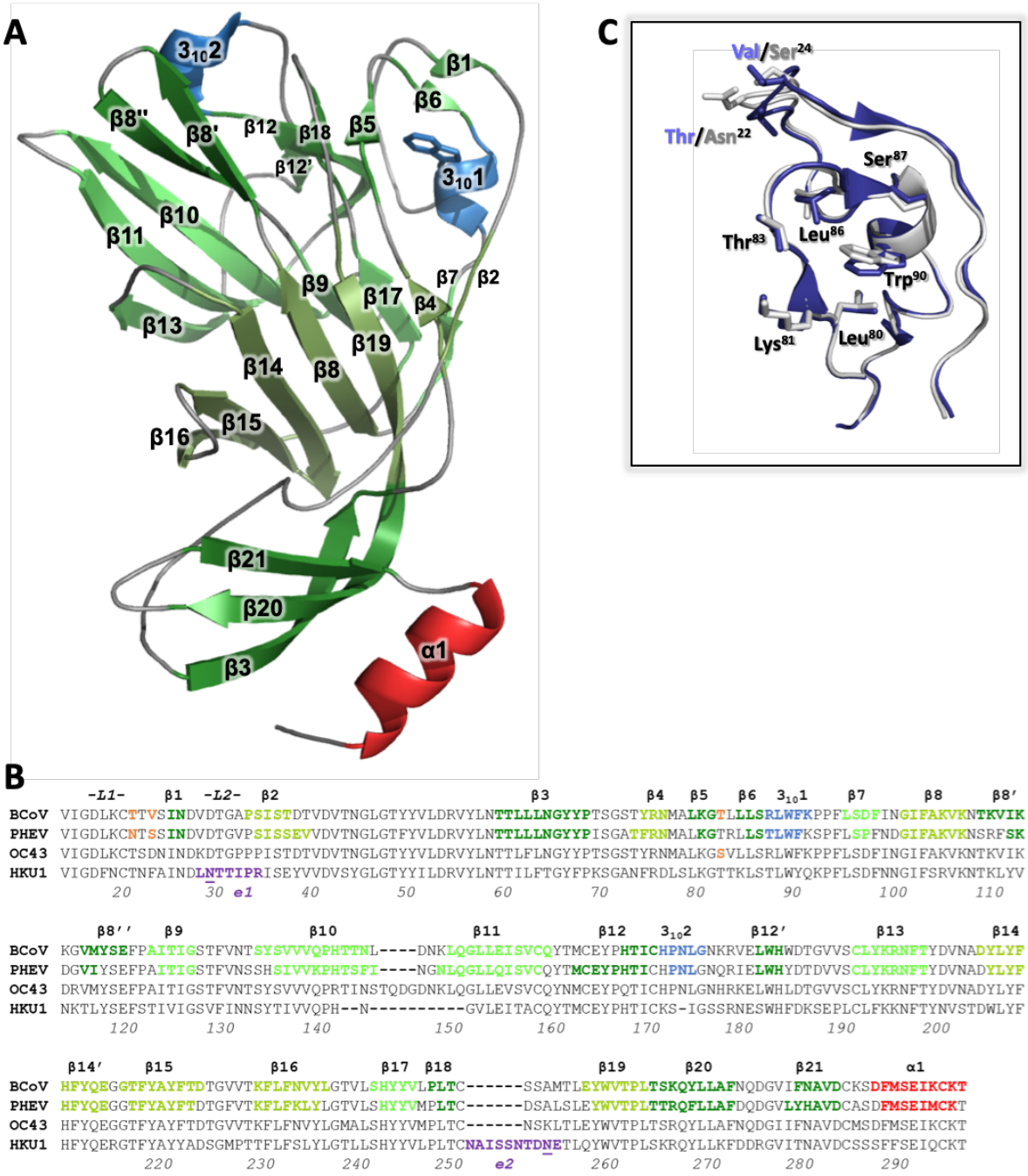


Fig. S1. (A) BCoV S1^{A} structure with secondary structure elements indicated as determined in DSSP [1] and color-coded accordingly (β-strands (forest green); β-sheet 1 (split pea); β-sheet 2 (chartreuse); α-helix (red); 3_{10}-helix (blue). The side chain of Trp^{90}, central to site B, is indicated in sticks. (B) Structure-based sequence alignment of lineage A betaCoV S1^{A} proteins (BCoV strain Mebus (GB: P15777.1), OC43 strain ATCC VR-759 (GB: AAT84354.1), PHEV strain UU (GB: ASB17086.1) and HKU1 strain Caen1 (GB: ADN03339.1)) used in this study. Secondary structure elements of BCoV (PDB: 4H14) and PHEV (PDB: 6QFY) are color-coded as under A) and numbered above alignment. OC43 (Ser^{83}) and PHEV (Asn^{22}, Ser^{24}, Val^{33}) residues targeted in gain-of-function

experiments and corresponding BCoV amino acids are colored orange. HKU1 elements (e) 1 and 2 (indicated below the alignment) are colored purple and *N*-glycosylation site Asn residues are underlined. Residue numbering according to BCoV S1^{A}. (C) Superimposition of cartoon representations of BCoV (purple) and PHEV (grey) S1^{A} site B with the conserved site B residues and divergent second ring residues indicated in sticks.

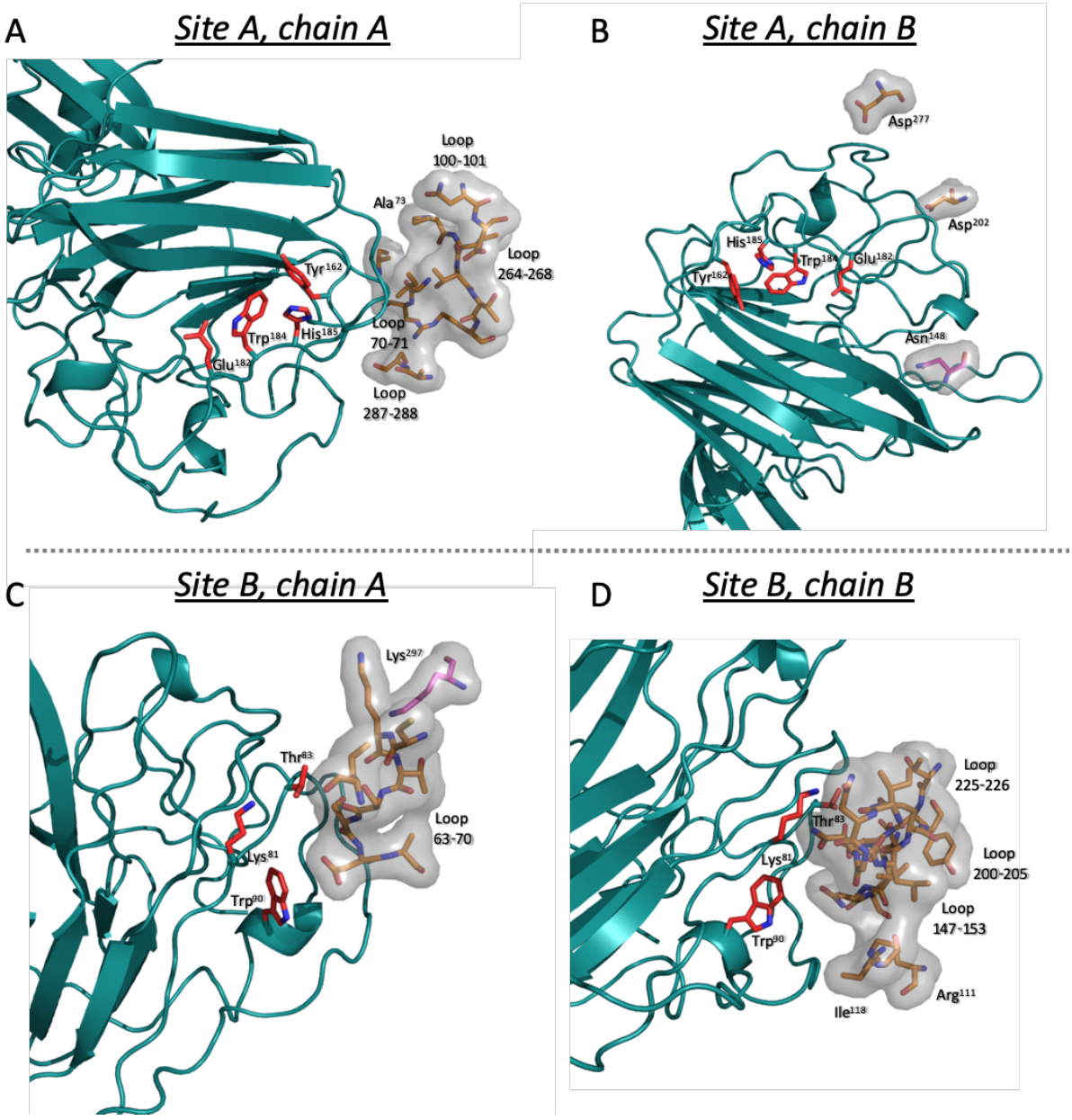


Fig. S2. Crystal packing of PHEV S1^{A} reveals accessible site A but occluded site B. (A) Site A within chain A of PHEV S1^{A} crystals is accessible. Cyan cartoon representation of PHEV S1^{A} with site A residues Tyr^{162}, Glu^{182}, Trp^{184} and His^{185} indicated as red sticks. Atoms of neighboring S1^{A} molecules within 15 Å of any atom of residues forming site A are indicated in orange sticks and colored by element (oxygen, red; nitrogen, blue). No atoms of neighboring chains are within 10 Å of site A (minimal distance 11.4 Å). (B) Site A within chain B of PHEV S1^{A} crystals is accessible. S1^{A} representation as under (A). Atoms of neighboring molecules within 15 Å of

site A are indicated in orange or magenta sticks (different color for different chains) and colored by element. No atoms of neighboring chains are within 10 Å of site A (minimal distance 12.9 Å). (C) Site B within chain A of PHEV $S1^A$ crystals is occluded by loop 63-70 of a neighboring $S1^A$ molecule. Cyan cartoon representation of PHEV $S1^A$ with site B residues Lys^{81}, Thr^{83} and Trp^{90} indicated as red sticks. Atoms of other chains within 10 Å of any atom of residues forming site B are indicated in orange or magenta sticks (different color for different chains) and colored by element. (D) Site B within chain B of PHEV $S1^A$ crystals is occluded by loops of a neighboring molecule. $S1^A$ representation as under (C). Atoms of other chains within 10 Å of site B are indicated in orange sticks and colored by element.

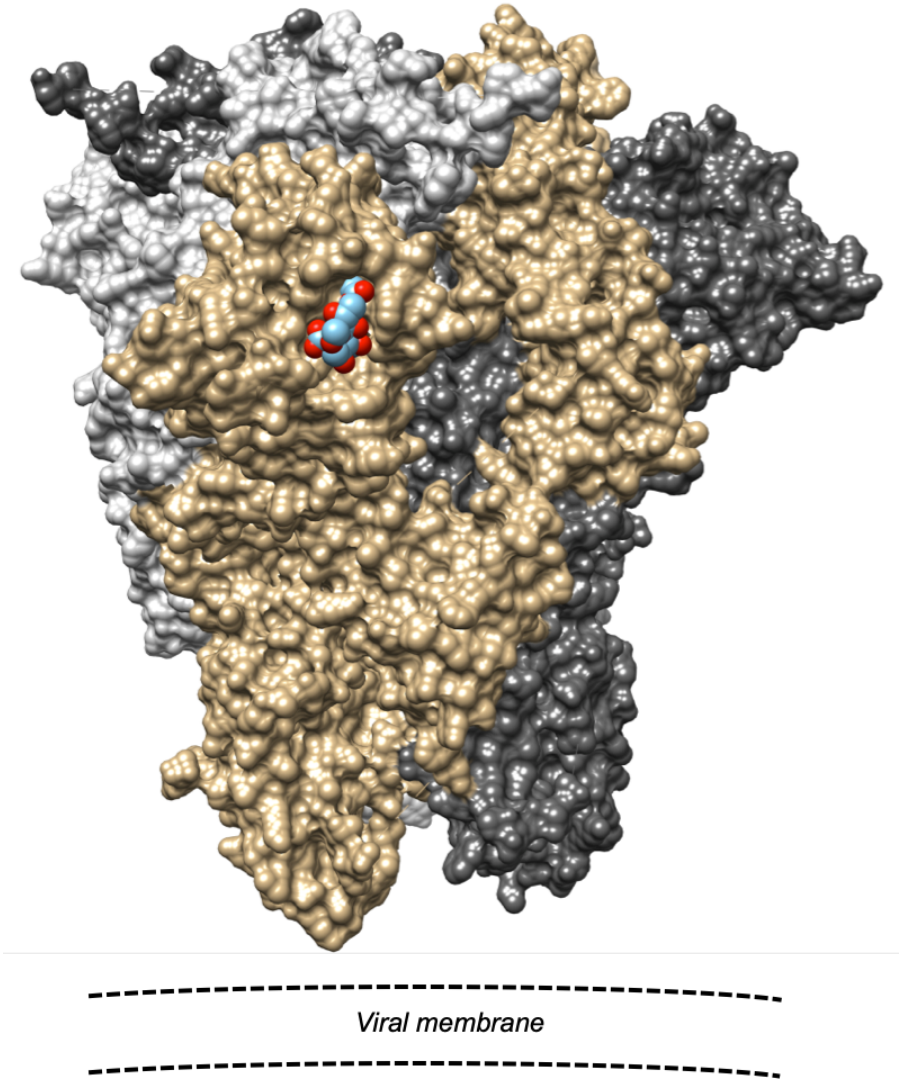


Fig. S3. Surface representation of trimeric HKU1 spike protein ectodomain as determined by cryo-electron microscopy [2] with 9-*O*-Ac-Sia ligand modeled into site B. The monomers of the trimeric spike peplomer are indicated in different colors (wheat, white, grey). The position and orientation of the ligand ($Neu5,9Ac_2$) corresponds to the automated molecular docking as performed for BCoV $S1^A$ (depicted in **Fig. 3b**). Atoms of the sialoglycan receptor are colored according to element (oxygen, red; carbons, cyan; nitrogen, blue).

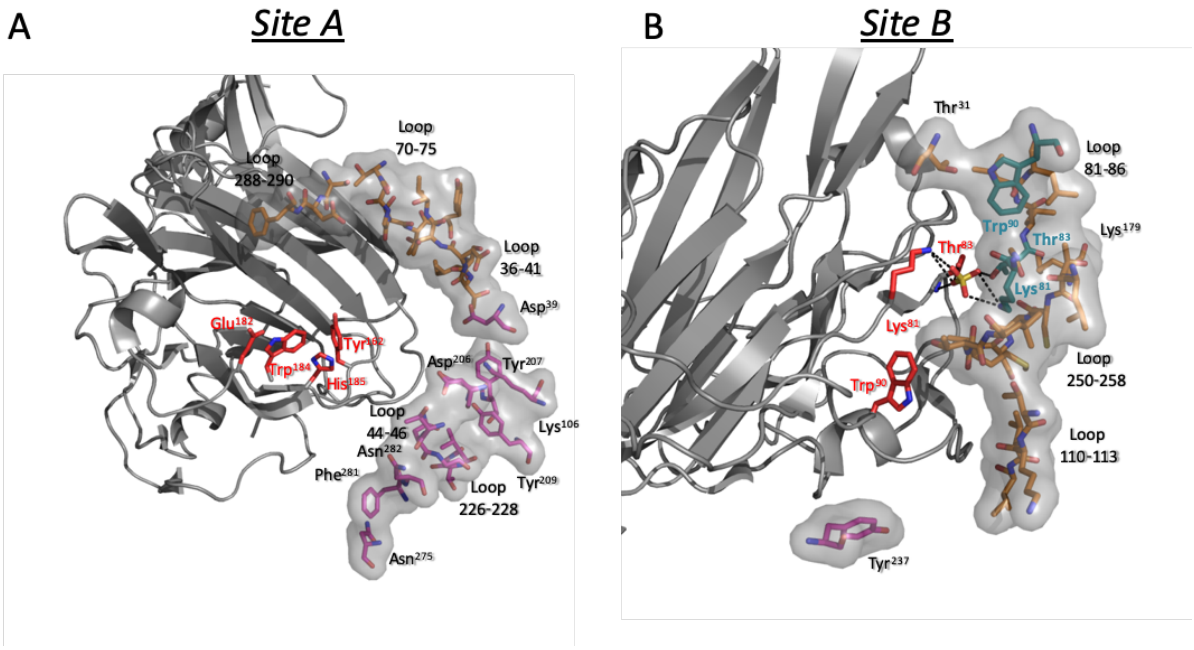


Fig. S4. Crystal packing of BCoV S1^{A} reveals accessible site A but occluded site B. (A) Site A is accessible in BCoV S1^{A} crystals. Grey cartoon representation of BCoV S1^{A} with site A residues Tyr^{162}, Glu^{182}, Trp^{184} and His^{185} indicated as red sticks. All atoms of neighboring S1^{A} molecules within 15 Å of any atom of residues forming site A are indicated in orange or magenta sticks (different color for different chains) and colored by element (oxygen, red; nitrogen, blue). No atoms of neighboring chains are within 10 Å of site A (minimal distance 10.1 Å). (B) Site B is occluded by the equivalent region of a neighboring molecule in BCoV S1^{A} crystals. Grey cartoon representation of BCoV S1^{A} with site B residues Lys^{81}, Thr^{83} and Trp^{90} indicated as red sticks. All atoms of molecules within 10 Å of site A are indicated in orange or magenta sticks (different color for different chains) and colored by element. Site B residues of the neighboring chain are indicated as cyan sticks and colored by element. Note how the sulfate oxoanion (yellow sticks) forms hydrogen-bonds with the Lys^{81} and Thr^{83} residues of both molecules.

A

[protein] (µg/ml)	wt	$K^{80/81}A$	$T^{82/83}A$	$W^{89/90}A$
OC43 $S1^A$	59,0+/-6,7	62,3+/-17,0	80,0+/-11,5	111,9+/-18,3
HKU1 $S1^A$	60,2+/-8,72	46,2+/-5,7	48,4+/-1,6	63,6+/-6,7

B

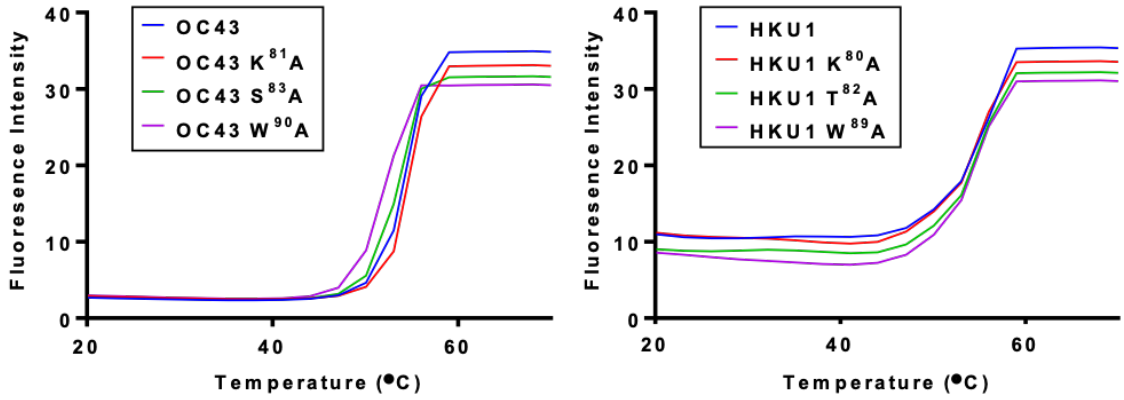


Fig. S5. Expression levels and thermal stability of OC43 and HKU1 $S1^A$-Fc and mutants thereof. (A) HEK293T cells grown to 70% confluency in 35mm wells were transfected with equal amounts of the expression vectors (1 mg plasmid DNA/well) as described [3]. Cell-free culture supernatants, harvested at seven days post transfection, were two-fold serially diluted and coated in Nunc Maxisorp 96-well ELISA plates (Fisher Scientific). $S1^A$-Fc fusion proteins were detected by ELISA with HRP-conjugated goat anti-human IgG antiserum (Southern Biotech) as described previously [4]. Standard curves were generated with purified OC43 and HKU1 $S1^A$-Fc. Protein concentrations (mean ± SD; µg/ml cell culture supernatant) were calculated by 4 parameter logistic regression (4PL) method (GraphPad Prism version 7.04) from three independent transient transfection experiments. (B) Thermal protein stability as an indicator for proper folding was assessed by fluorescence-based thermal shift assay with fluorescent dye Sypro Orange as described in [5].

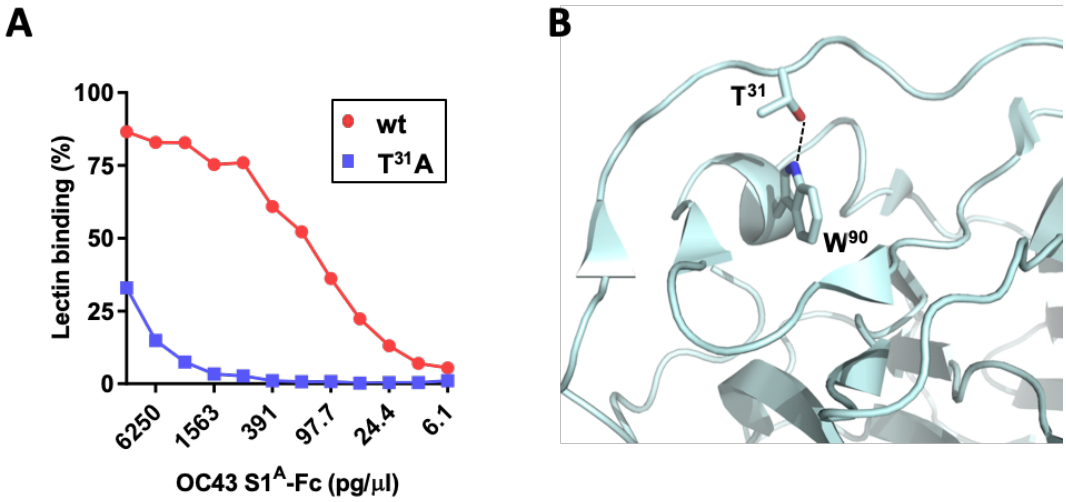


Fig. S6. Effect of Thr31Ala mutation on lectin activity of OC43 S1^{A} protein domain. (A) S1^{A}-hFc fusion proteins (2-fold serial dilutions, starting at 12.5 µg/µl) were compared by sp-LBA for relative binding to BSM at 37°C. sp-LBA performed and results presented as in **Fig. 1**. (B) Cartoon representation of BCoV site B with Thr^{31} and Trp^{90} side chains indicated in sticks and colored by element and the hydrogen bond between the side chains of residue 31 and 90 indicated as a black dashed line.

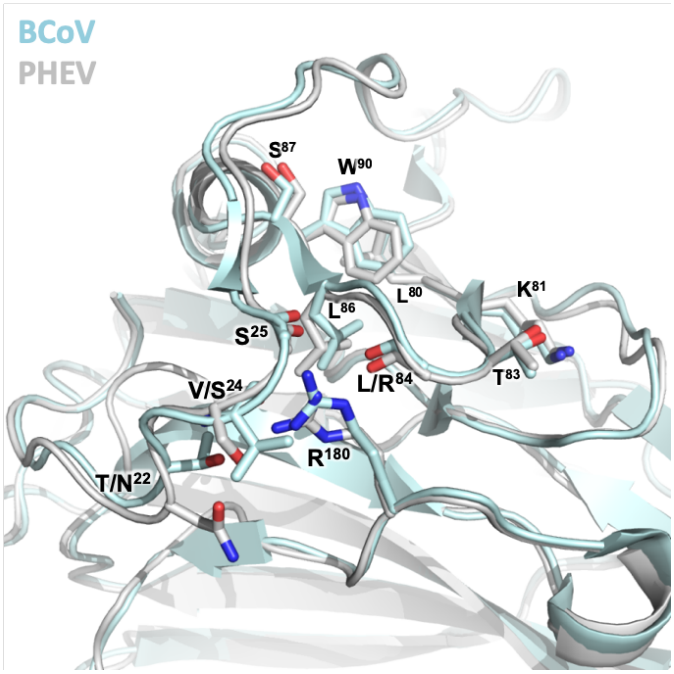


Fig. S7. Structural alignment of β1CoV S1^{A} site B. Cartoon representation of superimposed BCoV (blue) and PHEV (white) S1^{A} structures focusing on site B and the N-terminal loop adjacent to it. Side chains of site B residues 80, 81, 83, 84, 86, 87 and 90 are indicated in sticks and colored by element. Side chains of N-terminal loop residues 22 and 24 as well as carboxy-termini of residues 24 and 25, all of which are involved in H-bond interactions between the N-terminal loop and site B (see **Fig. 4c**), are indicated in sticks and colored by element.

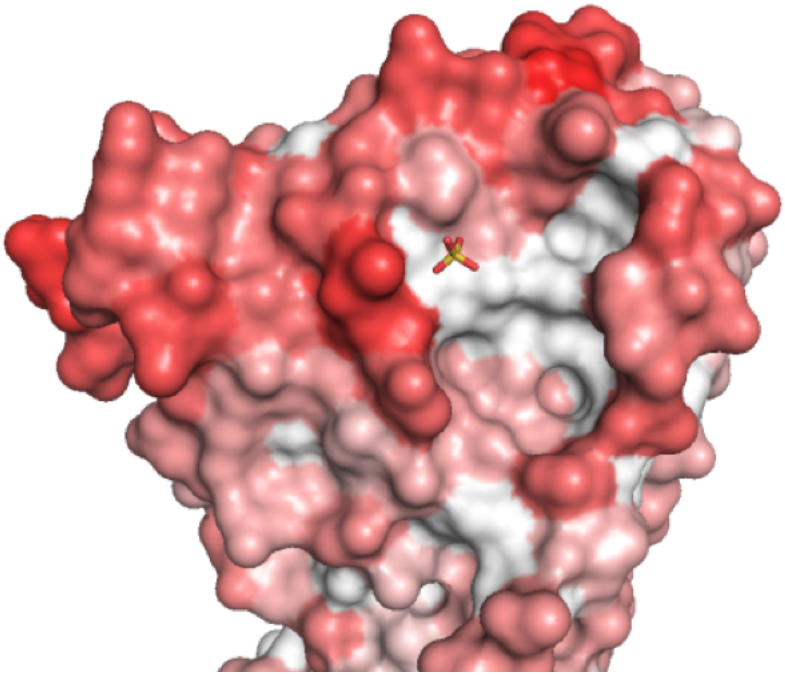


Fig. S8. Amino acid conservation heat map of lineage A betaCoV S1^{A} proteins. Sequence conservation determined by alignment of βCoV lineage A S1^{A} sequences in the NCBI database (n=158) of viruses known and presumed to bind *O*-acetylated sialic acids (BCoV, HCoV-OC43, PHEV, CRCoV, ECoV, RbCoV, HCoV-HKU1, HKU23, HKU24). For this reason, MHV S1^{A} sequences were excluded. Conserved positions are colored white and highly variable positions are indicated in dark red. Structure of BCoV S1^{A} (PDB: 4H14) was used as a template.

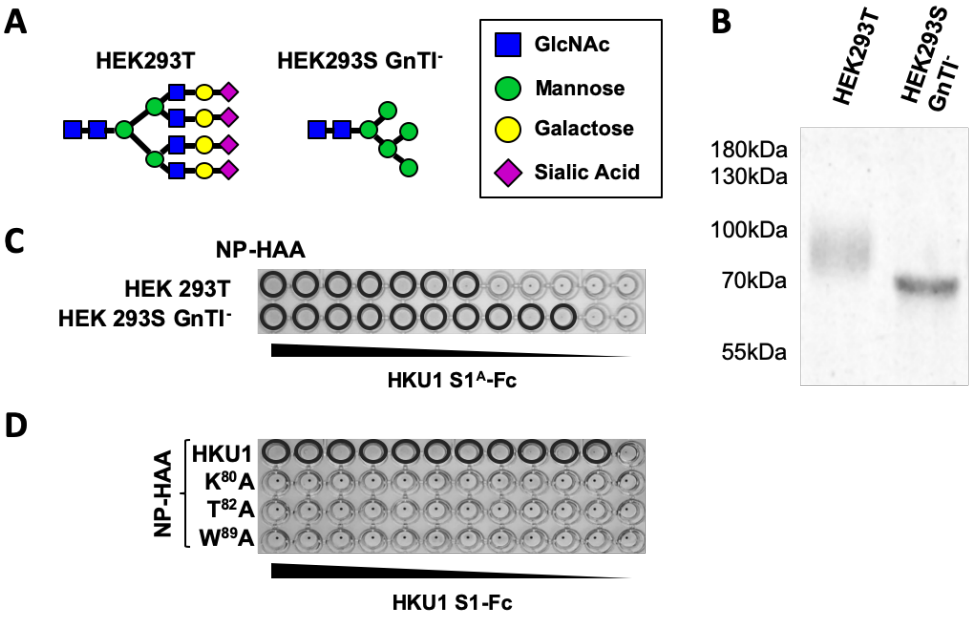


Fig. S9. N-glycans attached to HKU1 S1^{A} hamper binding to 9-*O*-Ac-Sias. (A) *N*-glycan types produced in HEK 293T and 293S GnTI^{-} cells: a heterogeneous mixture of complex-type bi-, tri- and tetra-antennary forms versus

homogeneous high-mannose chains. (B) HKU1 $S1^A$-Fc produced in HEK 293T and 293S cell lines as analyzed by SDS-PAGE and Coomassie Brilliant Blue staining. (C) Comparison of binding affinities of HKU1 $S1^A$ proteins produced in HEK 293T or HEK 293S GnTI$^-$ cells. (D) Comparison of binding affinities of HKU1 S1 and mutant derivatives thereof produced in HEK 293S GnTI$^-$ cells. The results indicate that HKU1 binding to 9-*O*-Ac-Sia is mediated by site B in domain $S1^A$. NP-HAA performed and presented as in **Fig. 6**.

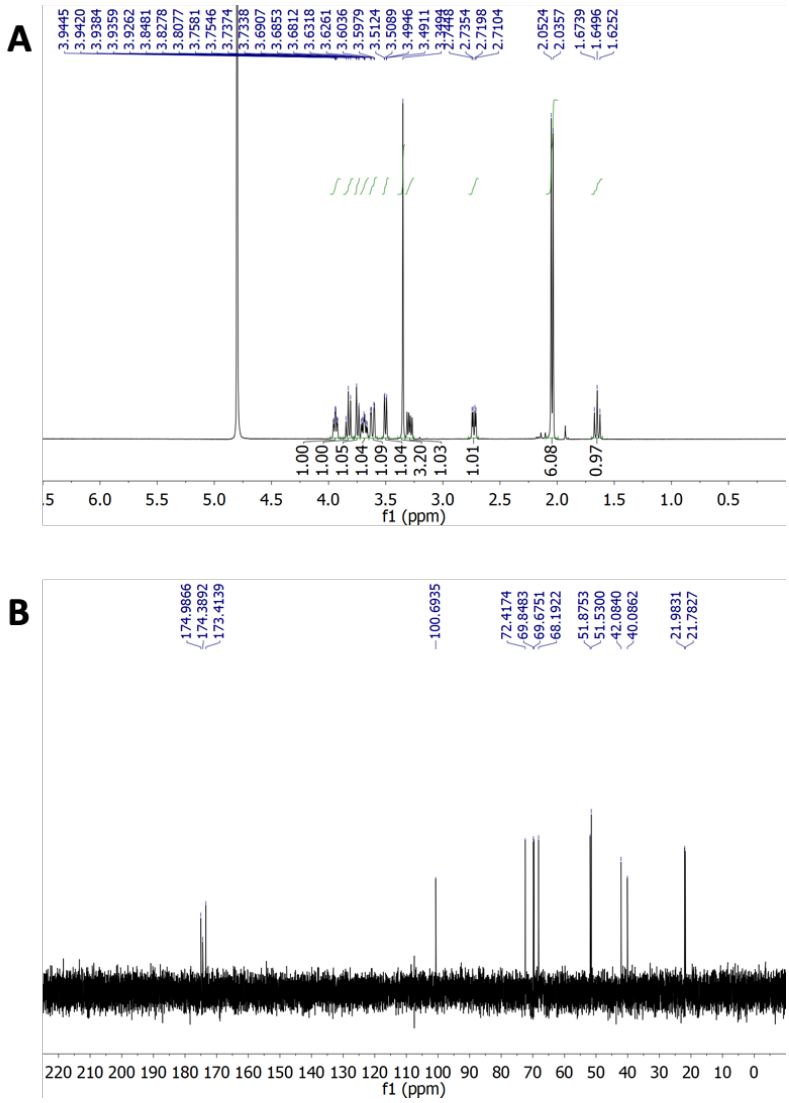


Fig. S10. Nuclear Magnetic Resonance (NRM) spectra of synthesized Neu5,9*N*Ac$_2$2αMe. Proton (A) and carbon (B) NMR spectra were collected on Varian Inova 500. Hi-resolution mass data were collected on Agilent 6560 Ion Mobility Q-TOF LC/MS. NMR data was processed using MestReNova. Proton and carbon chemical shifts are listed below.

^{1}H NMR (500 MHz, D_2O) δ 3.94 (ddd, *J* = 8.9, 7.6, 2.8 Hz, 1H, H-8), 3.83 (t, *J* = 10.1 Hz, 1H, H-5), 3.75 (dd, *J* = 10.4, 1.8 Hz, 1H, H-6), 3.69 (ddd, *J* = 11.9, 9.8, 4.7 Hz, 1H, H-4), 3.61 (dd, *J* = 14.1, 2.9 Hz, 1H, H-9a), 3.50 (dd, *J* =

8.9, 1.8 Hz, 1H, H-7), 3.35 (s, 3H, OCH_3), 3.29 (dd, *J* = 14.1, 7.8 Hz, 1H, H-9b), 2.73 (dd, *J* = 12.5, 4.7 Hz, 1H, H-3_{eq}), 2.05 (s, 3H, COCH_3), 2.04 (s, 3H, COCH_3), 1.65 (t, *J* = 12.2 Hz, 1H, H-3_{ax}).

^{13}C NMR (126 MHz, D_2O) δ = 174.99, 174.39, 173.41, 100.69, 72.42, 69.85, 69.68, 68.19, 51.88, 51.53, 42.08, 40.09, 21.98, 21.78. ESI-MS (pos.) Calculated for $C_{14}H_{23}N_2O_9Na$: 387.1380; found: 387.1386 [M + Na], 409.1205 [M – H + 2Na].

References Supplemental Information

1. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–637.

2. Kirchdoerfer RN, et al. (2016) Pre-fusion structure of a human coronavirus spike protein. *Nature* 531(7592):118–121.

3. Zeng Q, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ (2008) Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc Natl Acad Sci U S A* 105(26):9065–9069.

4. Langereis MA, Zeng Q, Heesters B, Huizinga EG, de Groot RJ (2012) The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. *PLoS Pathog* 8(1):1–8.

5. Lo MC, et al. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* 332(1): 153–159 (2004).

Chapter 4

Structural basis for human coronavirus attachment to sialic acid receptors

M. Alejandra Tortorici[1,2,3], Alexandra C. Walls[1], Yifei Lang[4], Chunyan Wang[4], Zeshi Li[5,6], Danielle Koerhuis[4], Geert-Jan Boons[5,6,7,8], Berend-Jan Bosch[4], Félix A. Rey[2,3,] Raoul J. de Groot[4], and David Veesler [1]

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA.

[2]Institut Pasteur, Unité de Virologie Structurale, Paris, France.

[3]CNRS UMR 3569, Unité de Virologie Structurale, Paris, France.

[4]Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine,Utrecht University, Utrecht, the Netherlands.

[5]Department of Chemical Biology and Drug Discovery, Utrecht University, Utrecht, the Netherlands.

[6]BijvoetCenter for Biomolecular Research, Utrecht University, Utrecht, the Netherlands.

[7]Department of Chemistry, University of Georgia, Athens, GA, USA.

[8]Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA

Nat Struct Mol Biol. 2019 Jun;26(6):481-489.

Abstract

Coronaviruses cause respiratory tract infections in humans and outbreaks of deadly pneumonia worldwide. Infections are initiated by the transmembrane spike (S) glycoprotein, which binds to host receptors and fuses the viral and cellular membranes. To understand the molecular basis of coronavirus attachment to oligosaccharide receptors, we determined cryo-EM structures of coronavirus OC43 S glycoprotein trimer in isolation and in complex with a 9-*O*-acetylated sialic acid. We show that the ligand binds with fast kinetics to a surface-exposed groove and that interactions at the identified site are essential for S-mediated viral entry into host cells, but free monosaccharide does not trigger fusogenic conformational changes. The receptor-interacting site is conserved in all coronavirus S glycoproteins that engage 9-*O*-acetyl-sialogycans, with an architecture similar to those of the ligand-binding pockets of coronavirus hemagglutinin esterases and influenza virus C/D hemagglutinin-esterase fusion glycoproteins. Our results demonstrate these viruses evolved similar strategies to engage sialoglycans at the surface of target cells.

Coronaviruses are large, positive-sense enveloped RNA viruses in the *Nidovirales* order and are divided into four genera: α, β, γ and δ. Two β-coronaviruses have caused outbreaks of deadly pneumonia in humans since the beginning of the 21st century. The severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in 2002 and was responsible for an epidemic that spread to five continents with a fatality rate of 10% before being contained in 2003 (with additional cases reported in 2004). The Middle East respiratory syndrome coronavirus (MERS-CoV) emerged in the Arabian Peninsula in 2012 and has caused recurrent outbreaks in humans with a fatality rate of 35%. SARS-CoV and MERS-CoV are zoonotic viruses that crossed the species barrier using bats/palm civets [1] and dromedary camels [2], respectively. Four other coronaviruses of zoonotic origin are endemic in the human population, accounting for up to 30% of mild respiratory tract infections and causing severe complications or fatalities in young children, the elderly and immunocompromised individuals [3,4]. These viruses are HCoV-NL63 and HCoV-229E (α-coronaviruses) and HCoV-OC43 and HCoV-HKU1 (β-coronaviruses). Currently, no specific antiviral treatments or vaccines are available to combat any human coronavirus. Furthermore, future cross-species transmission events of coronaviruses seem likely, given the large reservoir found in bats [5–7]. Studying coronaviruses will therefore help in understanding the principles governing cross-species transmission and adaptation to humans and in preparing for putative future zoonotic outbreaks.

Coronaviruses use homotrimers of the spike (S) glycoprotein to promote host attachment and fusion of the viral and cellular membranes for entry. S is the main antigen present at the viral surface and is the target of neutralizing antibodies during infection. As a result, it is a focus of vaccine design. S is a class I viral fusion protein synthesized as a single polypeptide chain precursor of approximately 1,300 amino acids [8]. For many coronaviruses, S is processed by host proteases to generate two subunits, designated S1 and S2, which remain non-covalently bound in the pre-fusion conformation. The N-terminal S1 subunit comprises four β-rich domains, designated A, B, C and D, with domain A or B acting as receptor-binding domains in different coronaviruses. The transmembrane C-terminal S2 subunit is the metastable spring-loaded fusion machinery [9]. During entry, S2 is further proteolytically cleaved at the S2' site, immediately upstream of the fusion peptide [10]. This second cleavage step occurs for all coronaviruses and is believed to activate the protein for membrane fusion, which takes place via irreversible conformational changes [11–14]. In recent years, cryo-EM work led to the determination of coronavirus S glycoprotein ectodomain structures in the pre-fusion and post-fusion states, providing snapshots of the start and end points of the fusion reaction [9,13,15–24]. Cryo-EM structures of the SARS-CoV and MERS-CoV S glycoproteins in complex with human neutralizing antibodies also informed about the mechanism of fusion activation [25].

HCoV-OC43 was isolated for the first time in 1967 from volunteers at the Common Cold Unit in Salisbury, United Kingdom. Molecular clock analysis of genome sequences suggested that HCoV-OC43 originated from a zoonotic transmission event of a bovine coronavirus (BCoV) and dated their most recent common ancestor between the 1890s and the 1950s [26,27]. HCoV-OC43, HCoV-HKU1, BCoV and porcine hemagglutinating encephalomyelitis virus (PHEV) use 9-*O*-acetyl-sialic acid (9-*O*-Ac-Sia) as a receptor, which is terminally linked to oligosaccharides decorating glycoproteins and gangliosides at the host cell surface [28,29]. The S glycoprotein of these viruses mediates 9-*O*-Ac-Sia binding, whereas the hemagglutinin-esterase (HE) protein acts as receptor-destroying enzyme, via sialate-*O*-acetyl-esterase activity, to facilitate release of viral progeny from infected cells and escape from attachment to non-permissive host cells or decoys [30–33]. These properties are shared with the hemagglutinin-fusion-esterase (HEF) glycoproteins of influenza C and D viruses [28,34–36].

Sialic acids are ubiquitous terminal residues of glycoconjugates and occur in a wide variety as a result of modifications of the core *N*-acetyl neuraminic acid molecule and of differences in glycosidic linkages [37–39]. Previous biochemical work established that domain A of coronavirus S glycoproteins mediates attachment to oligosaccharide receptors, such as for HCoV-OC43 and BCoV, which interact with 9-*O*-Ac-Sia [28,40,41], or MERS-CoV, which binds to α2,3-linked (and to a lesser extent to α2,6-linked) sialic acids, with sulfated sialyl-Lewis X being the preferred binder [42]. On the basis of the galectin-like fold of domain A of coronavirus S and mutational analyses, it was suggested that key saccharide-binding residues locate to the viral membrane distal side of the BCoV β-sandwich. Our recent work, however, indicated that the 9-*O*-Ac-Sia binding site of HCoV-OC43, HCoV-HKU1, BCoV and PHEV is conserved among these viruses and resides at a distinct location of domain A [43]. Although we validated the findings using mutagenesis and BCoV infectivity assays, no structural information is available on the mechanism of coronavirus interaction with saccharides aside from in silico modeling [43]. This knowledge gap limits our understanding of the roles of these receptors in viral infection or zoonosis and hinders the rational design of inhibitors.

To understand attachment of coronaviruses to sialic acids at the surface of host cells, we determined cryo-EM structures of a stabilized HCoV-OC43 S glycoprotein trimer in isolation and in complex with 5-*N*-acetyl,9-*O*-acetyl-neuraminic acid α-methyl glycoside (9-*O*-Ac-Me-Sia) at 2.9-Å and 2.8-Å resolution, respectively. We show that the ligand binds with fast association/dissociation kinetics in a groove on HCoV-OC43 S located at the surface of domain A. Site-directed mutagenesis combined with binding experiments validated our structural findings, and infectivity assays showed that the residues involved in 9-*O*-Ac-Sia binding are essential for HCoV-OC43 S–mediated entry into host cells. Our results further show that binding to free 9-*O*-Ac-Me-Sia and/or acidic pH did not induce

fusogenic conformational changes of S, suggesting that multivalent interactions with sialoglycans and/or further attachment to a putative proteinaceous receptor [44] are essential to promote membrane fusion. The receptor-interacting site is conserved in all coronavirus S glycoproteins known to attach to 9-*O*-Ac-sialoglycans and shares architectural similarity with the ligand-binding pockets of coronavirus HEs and influenza virus C/D HEF glycoproteins, thus highlighting common structural principles of recognition [45,46].

Results

Structure of the apo-HCoV-OC43 S glycoprotein. We determined a 2.9-Å resolution cryo-EM reconstruction of an apo-HCoV-OC43 S ectodomain trimer mutant, in which the S1/S2 furin cleavage site was abrogated to prevent proteolytic processing during biogenesis. HCoV-OC43 S folds as a 150-Å high and 130-Å wide compact trimer (Fig.1a, Supplementary Fig. 1 a, b and Table 1). The S1 subunit has a V-shaped architecture resulting from the 3D arrangement of its four domains (A, B, C and D), similarly to other β-coronavirus S structures [9,17–25]. The S2 subunit, which is more conserved than the S1 subunit among coronaviruses, folds as a mostly helical, elongated trimeric unit with a connector domain appended at its C-terminal end [9,16] (Fig. 1a). Among available coronavirus S glycoprotein structures, HCoV-OC43 S is most similar to mouse hepatitis virus (MHV) S [9] (r.m.s. deviation (r.m.s.d.) 4.7 Å over 979 aligned Cα positions) and to HCoV-HKU1 S [18] (r.m.s.d. 4.5 Å over 949 aligned Cα positions), sharing 62% and 68% sequence identity, respectively. The cryo-EM reconstruction resolves 14 N-linked glycans extending from the surface of each protomer. The HCoV-OC43 S oligosaccharide density is comparable to that of SARS-CoV S and MERS-CoV S, with all three viruses belonging to the β-genus, but lower than the glycan density of the porcine delta coronavirus S (δ-genus) or the HCoV-NL63 S (α-genus) glycoproteins [16,17,25].

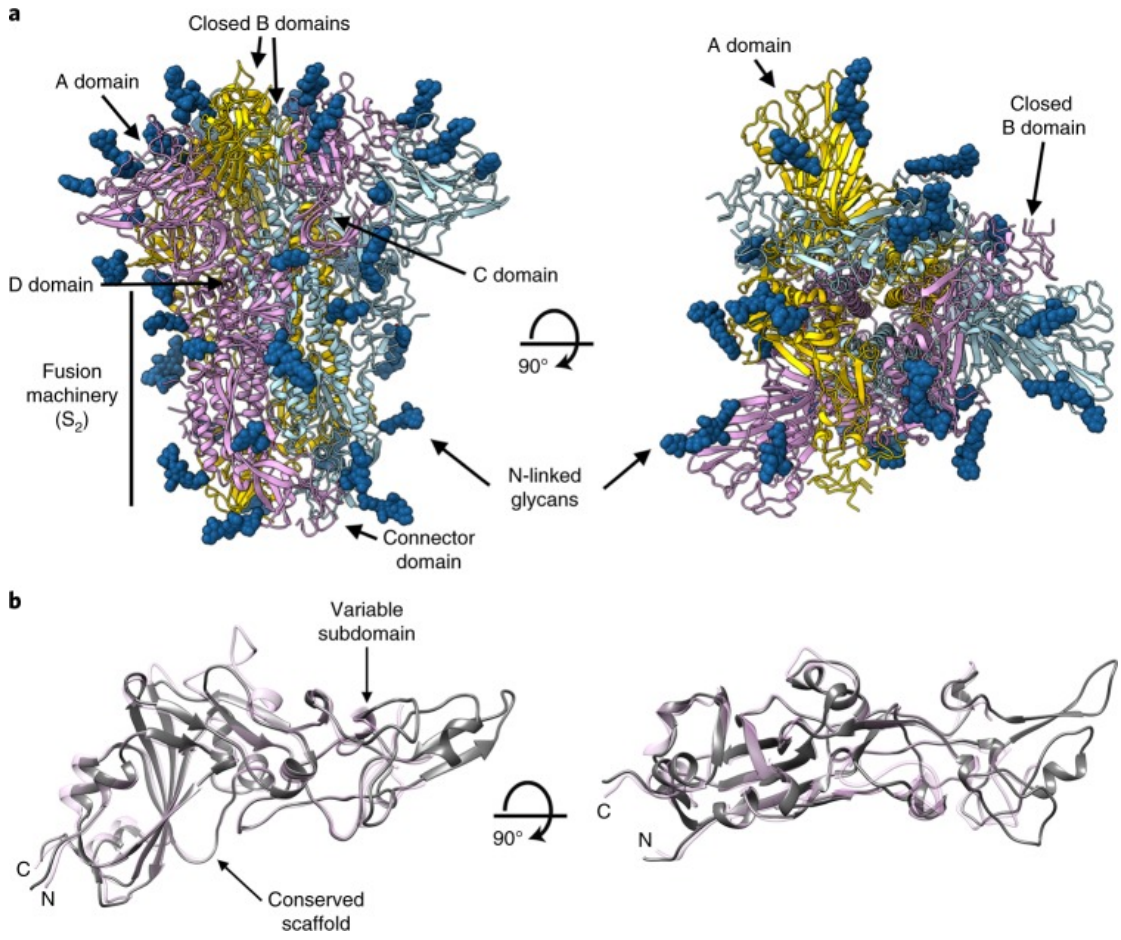


Fig. 1. Cryo-EM structure of the apo-HCoV-OC43 S glycoprotein. a, Ribbon diagrams of the HCoV-OC43 S ectodomain trimer in two orthogonal orientations. The individual protomers are each in a different color, and the glycans are rendered as dark blue spheres. b, Ribbon diagrams of the superimposed HCoV-OC43 (light pink) and HCoV-HKU1 (dark gray) B domains in two orthogonal orientations. The N and C termini are labeled.

Table 1. Cryo-EM data collection, refinement and validation statistics

	Apo-HCoV-OC43 S (EMD-20070, PDB 6OHW)	**Holo-HCoV-OC43 S (EMD-0557, PDB 6NZK)**
Magnification	47,619	47,619
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	70	70
Defocus range (μm)	0.3–4.8	0.4–2.8
Pixel size (Å)	0.525	1.05
Symmetry imposed	C3	C3
Initial particle images (no.)	197,791	332,912
Final particle images (no.)	69,648	105,919
Map resolution (Å)	2.9	2.8
FSC threshold	0.143	0.143

Refinement		
Initial model used (PDB code)	3JCL	3JCL
Model resolution (Å)	3.0	2.9
FSC threshold	0.5	0.5
Map sharpening B factor (Å^2)	−61	−70
Model composition		
Nonhydrogen atoms	27,477	
Protein residues	3,519	3,519
Ligands	0	3
Waters	186	396
B factors (Å^2)		
Protein	18.6	12.3
Ligand	–	33.9
R.m.s. deviations		
Bond lengths (Å)	0.026	0.025
Bond angles (°)	1.80	1.82
Validation		
MolProbity score	0.7	0.8
Clashscore	0.6	1
Poor rotamers (%)	0.4	0.4
Ramachandran plot		
Favored (%)	98.1	98.3
Allowed (%)	100	99.9
Disallowed (%)	0	0.1

Domain B shows the highest variability within S_1 subunits across coronaviruses, which correlates to the ability of different viruses to interact with distinct host receptors. For β-coronaviruses, the canonical architecture of domain B comprises a conserved five-stranded anti-parallel β-sheet, decorated with α-helices on both sides, and a highly variable external subdomain that can mediate receptor engagement for SARS-CoV [19,47] or MERS-CoV [48]. Domain B of HCoV-OC43 and HCoV-HKU1 are structurally similar and can be superimposed with an r.m.s.d. of 1.0 Å over 251 aligned Cα positions [49], with differences largely restricted to the external subdomain (Fig. 1b). The current consensus in the field is that HCoV-OC43 S does not rely on receptors other than 9-*O*-Ac-sialoglycans for promoting viral entry into host cells. In contrast with the MERS-CoV S and SARS-CoV S glycoproteins [19–21,23,25,50], in which domain B adopts alternative conformations, we detected a single closed conformation of domain B in the HCoV-OC43 S structure (Fig. 1a). Only the closed domain B conformation was also observed for the MHV S [9], HCoV-NL63 S [16], HCoV-HKU1 S [18], PDCoV S [17,24] and IBV S [22] glycoprotein structures.

Cryo-EM identification of a sialoside-binding site in the HCoV-OC43 S glycoprotein. HCoV-OC43, HCoV-HKU1, BCoV and PHEV attach to the surface of target cells by binding to 9-*O*-Ac-sialoglycans [28,29]. To directly visualize the binding site and characterize the molecular details of the interactions, we incubated the HCoV-OC43 S protein with 100 mM 9-*O*-Ac-Me-Sia, prior to vitrification and cryo-EM data collection. We determined a 3D reconstruction of the stabilized HCoV-OC43 S protein in complex with its receptor at 2.8-Å resolution, hereafter referred to as holo-HCoV-OC43 S (Supplementary Fig. 1c,d and Table 1). The resolution estimate is supported by the visible ordered water molecules interacting with the S glycoprotein, as expected at this resolution [51]. The structure reveals that the ligand interacts with a groove at the periphery of domain A, in agreement with the biochemical observations reported by Hulswit et al [43] (Fig. 2a–c). The receptor therefore docks into a distinct groove from those used by either human galectin-3 (ref. 52) or the rhesus rotavirus sialic acid-attachment protein53 (VP8*) to recognize their respective ligands (Supplementary Fig. 2a–c).

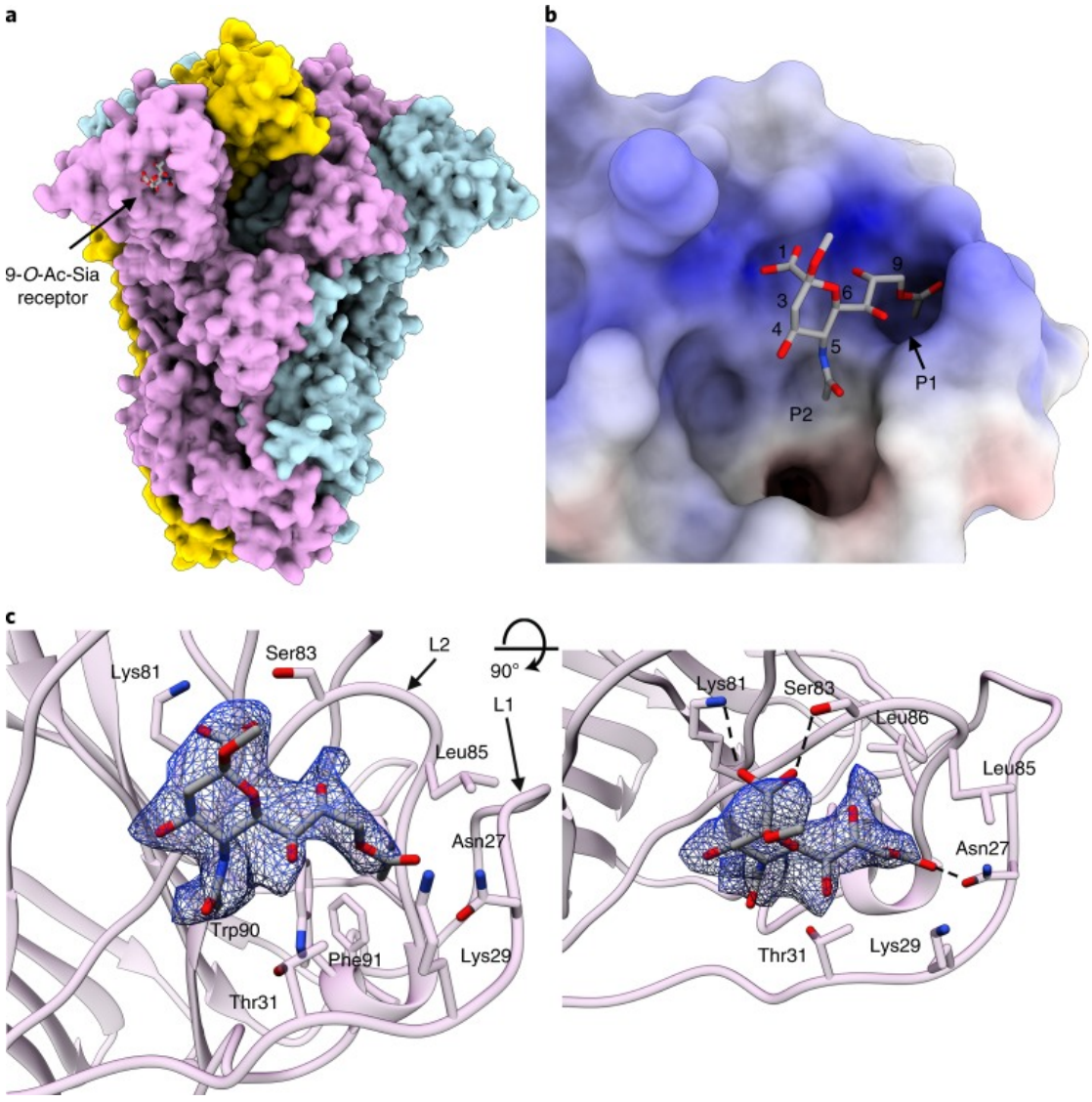


Fig. 2. Identification of a sialoglycan-binding site in the holo-HCoV-OC43 S glycoprotein structure. a, Molecular surface representation of the holo-HCoV-OC43 S ectodomain trimer structure with the bound ligand shown as sticks. Protomers are individually colored. b, Surface representation of the ligand-binding site colored by electrostatic potential from −12 to +12 kBT/ec. c, Two orthogonal views of the 9-*O*-Ac-Me-Sia binding site. The A domain is rendered as a ribbon diagram with the side chains of surrounding residues shown as sticks. The cryo-EM density is shown as a blue mesh. In a−c, the ligand is rendered as sticks with atoms colored by element (carbon, gray; nitrogen, blue; oxygen, red). Dashed lines show a salt bridge and hydrogen bonds formed between the ligand and domain A.

The sialoside-interacting groove defines two hydrophobic pockets, designated P1 and P2 (according to the nomenclature defined by Hulswit et al [43]), separated by the Trp90 indole side chain, and is delineated by two loops forming the rims of the binding site, termed L1 (27-Asn-Asp-Lys-Asp-Thr-Gly-32) and L2 (80-Leu-Lys-Gly-Ser-Val-Leu-Leu-86) (Fig.2c). The 9-*O*-Ac-Me-Sia C1-carboxylate forms a salt bridge with the Lys81 side chain amine and a hydrogen bond with the Ser83 side chain hydroxyl (Fig. 2c and Supplementary Fig. 3). The 5-nitrogen atom of the ligand makes a hydrogen bond with the Lys81 backbone carbonyl (Fig..2c). The ligand *N*-acetyl methyl inserts into the P2 hydrophobic pocket, defined by residues Leu80, Trp90 and Phe95. The ligand 9-*O*-acetyl methyl docks in the P1 hydrophobic pocket, which comprises Leu85, Leu86 and Trp90, whereas the 9-*O*-acetyl carbonyl makes a hydrogen bond with the Asn27 side chain amide. These observations rationalize the specificity of HCoV-OC43 S for this sialoside, because the 9-*O*-acetyl group is accommodated by a combination of hydrogen bonding and shape complementarity (Fig. 2b, c), similarly to 9-*O*-Ac-Sia binding sites of coronavirus, torovirus and orthomyxovirus HEs/HEFs [32–35,43,54]. Although most interactions occur with the same side of the ligand, the side chain hydroxyl of residue Thr31, which faces the 9-*O*-Ac-Me-Sia solvent-exposed side, forms a hydrogen bond with the Trp90 indole nitrogen. This interaction participates in stapling the A domain N-terminal segment to the β-sandwich core and contributes to defining the shape of the ligand-binding groove (Fig. 2c). Overall, the ligand buries 350 $Å^2$ of its surface upon binding to the HCoV-OC43 S protein, corresponding to approximately 62% of the 9-*O*-Ac-Me-Sia total solvent-accessible surface area. The observed binding mode is compatible with interactions with longer oligosaccharides, including α2,3- and α2,6-linked sialoglycans found on cell surfaces.

HCoV-OC43 S binds 9-*O*-Ac-Sia with fast association and dissociation rates. To characterize the binding kinetics and affinity of an individual HCoV-OC43 S binding site for 9-*O*-Ac-Sia receptors, we recombinantly produced the monomeric HCoV-OC43 S domain A and used biolayer interferometry to analyze its attachment to biotinylated oligosaccharides immobilized on the surface of streptavidin-coated biosensors. Domain A bound to and dissociated from 6-sialyl-5-*N*,

9-*O*-acetyl-lactosamine (9-*O*-Ac-6SLN) with fast on and off rates. (Fig. 3a). The observed binding was specific, as it was critically dependent on the presence of the sialate-9-*O*-acetyl moiety, in accordance with previous observations [28,32,43,55]. Domain A did not detectably bind to the corresponding non-*O*-acetylated oligosaccharide, 6SLN. This finding is explained by the absence of the 9-*O*-acetyl moiety in 6SLN, which contributes one-third of the total ligand buried surface area by contacting Asn27 and the P1 pocket of the glycoprotein, as revealed in our structure (Fig. 2b,c). Moreover, binding was largely abolished by de-*O*-acetylation of biosensor-bound 9-*O*-Ac-6SLN with porcine torovirus HE (Fig. 3a). Finally, substitution of Trp90 with alanine abrogated interactions with 9-*O*-Ac-6SLN (Fig. 3a), thereby confirming the central role for sialoside attachment of this amino acid residue that defines the floor of the ligand-binding groove [43].

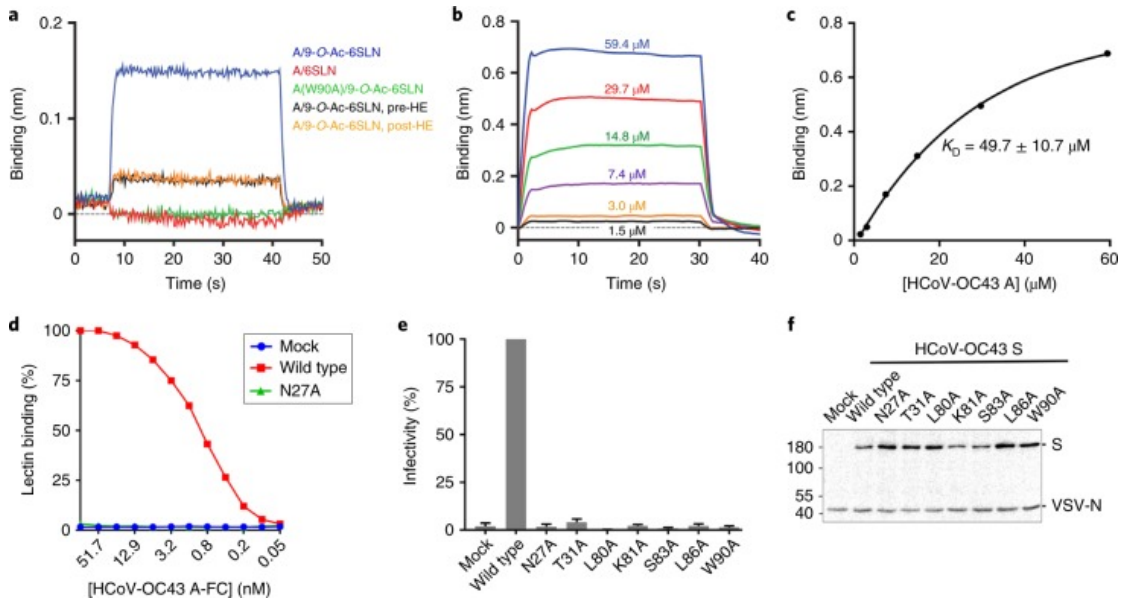


Fig. 3 The identified HCoV-OC43 S interactions with sialosides are characterized by fast kinetics and are required for viral entry. a, Biolayer interferometry showing binding of wild-type or W90A monomeric HCoV-OC43 domain A to immobilized 6-sialyl-5-N-acetyl, 9-*O*-acetyl-lactosamine (9-*O*-Ac-6SLN), 6-sialyl-5-*N*-acetyl-lactosamine (6SLN) or HE-pre-treated 6-sialyl-5-*N*-acetyl,9-*O*-acetyl-lactosamine before binding (9-*O*-Ac-6SLN, pre-HE) or after a successful association/dissociation event (9-*O*-Ac-6SLN, post-HE). b, Binding of different concentrations of wild-type monomeric A domain to immobilized 9-*O*-Ac-6SLN. c, Steady-state affinity determination using the curves shown in b. HCoV-OC43 A engages 9-*O*-Ac-6SLN with a K_D = 49.7 ± 10.7 µM. d, Asn27, a key 9-*O*-Sia-interacting residue visualized in the holo-HCoV-OC43 S glycoprotein structure was substituted with alanine, and binding was assessed using a solid-phase lectin binding assay. Data points are averages from three independent technical triplicates. The data are normalized relative to the wild type. e, Sialoside binding to the identified site is necessary for HCoV-OC43 S-mediated entry of pseudotyped VSV-ΔG particles into host cells. n = 3 pseudovirus experiments (technical replicates). Data are normalized relative to wild type and shown as mean and s.d. of technical triplicates. f, Western-blot analysis of

VSV-ΔG pseudotyped with wild-type or mutant HCoV-OC43 S. VSV-N was used as a quantitative control for the amount of virions analyzed.

Using steady-state analysis, we determined an equilibrium dissociation constant K_D = 49.7 ± 10.7 µM for the HCoV-OC43 domain A−9-*O*-Ac-6SLN complex (Fig. 3b,c). We calculated a half-life of $t_{1/2}$ = 0.7 s from the dissociation curves, a dissociation rate constant k_{off} = 1 s^{-1} ($k_{off} = t_{1/2}/ln_2$) and an association rate constant $k_{on} = 1.4 \times 104$ $M^{-1}s^{-1}$. These values predict rapid S-mediated virion attachment, particularly in high-density receptor environments such as the mucus layer, glycocalyx and cell surfaces. On the basis of these results, the mean life ($1/k_{off}$) of the 1:1 complex is predicted be short, in the order of 1 s, much shorter than the mean life of an individual influenza A hemagglutinin receptor-binding domain in complex with sialic acid, which ranges between 7 and 13.5 s [56]. In the context of authentic virions, however, the large number of S glycoproteins at the surface of coronaviruses is likely to increase the apparent binding affinity for sialoglycans through avidity, as described for influenza virus [57]. We posit that HCoV-OC43 and related β-coronavirus S glycoproteins evolved to dynamically interact with host sialosides and avoid irreversible attachment to decoy receptors via HE-mediated virion elution. Dynamic binding in combination with receptor destruction could promote virion motility by directional sliding diffusion through high-density interaction sites, as recently reported for influenza A and C viruses [58–60].

HCoV-OC43 S attachment to 9-*O*-Ac-sialoglycans is necessary for viral entry. Our structure rationalizes the results of our previous study in which the effect of individual HCoV-OC43 S domain A substitutions was assessed using a solid-phase lectin binding assay [43]. Substitution of Lys81 or Ser83 with alanine completely abrogated binding, as expected on the basis of our holo-HCoV-OC43 S structure, owing to disruption of the aforementioned electrostatic interactions with 9-*O*-Ac-Sia. Moreover, mutations of Leu80, Leu86 or Trp90 also disrupted binding, probably as a result of alteration of the P1 and/or P2 hydrophobic pockets accommodating the ligand 9-*O*-acetyl and 5-*N*-acetyl methyl groups, respectively. On the basis of our structure, we predicted that substitution of Asn27 with alanine would also inhibit binding, owing to loss of a hydrogen bond between the ligand 9-*O*-acetyl carbonyl and the Asn27 side chain amide. Using the same solid-phase lectin-interaction assays, we show that this substitution resulted in a loss of detectable binding, further validating our cryo-EM results (Fig. 3d).

We subsequently evaluated the importance of the identified interactions for HCoV-OC43 S-mediated infectivity using pseudotyped G-deficient vesicular stomatitis virus (VSV-ΔG). Substitutions at Asn27, Thr31, Leu80, Lys81, Ser83, Leu86 and Trp90 led to complete abrogation of viral entry (Fig. 3e, f), in agreement with our structural data, biolayer interferometry and solid-phase lectin binding assays, as

well as the literature [43]. These findings (i) support the importance of the identified residues for interacting with 9-*O*-Ac-Sia in the context of a full-length, membrane-embedded, HCoV-OC43 S glycoprotein and (ii) indicate that attachment to oligosaccharide receptors using the binding site visualized via cryo-EM plays a critical role in promoting HCoV-OC43 S-mediated viral entry.

Free 9-*O*-Ac-Sia does not trigger fusogenic conformational changes. Comparison of the stabilized apo- and holo-HCoV-OC43 S glycoprotein structures did not reveal conformational rearrangements upon binding to 9-*O*-Ac-Sia (the two structures can be superimposed with a Cα r.m.s.d. of 0.2 Å). To validate this finding, we investigated the effect of ligand binding to wild-type HCoV-OC43 S (that is, with a native S1/S2 cleavage site sequence) in various biochemical conditions. Importantly, the HCoV-OC43 S ectodomain trimer remained uncleaved after secretion (Supplementary Fig. 4a), perhaps owing to the paucity of furin present in the secretory pathway of HEK293F cells [61]. Incubation of the wild-type HCoV-OC43 S ectodomain trimer with trypsin at concentrations ranging from 0.2 to 28 $\mu g \cdot ml^{-1}$ (w/v), to recapitulate proteolytic priming [13], led to cleavage at the S1–S2 boundary, as observed via SDS-PAGE (Supplementary Fig. 4a). Incubation with 28 $\mu g \cdot ml^{-1}$ trypsin also led to cleavage of a small fraction of S at a second site, yielding a band with an apparent molecular weight of ~55 kDa, which could be consistent with cleavage at the S2' site (Supplementary Fig. 4a), an event believed to be restricted to fusion triggering upon receptor engagement for SARS-CoV S [11] or MERS-CoV S [12,62]. EM analysis of negatively stained samples, however, showed that the HCoV-OC43 S trimers remained in the pre-fusion conformation and were highly stable, even at the highest trypsin concentration tested (Supplementary Fig. 4b). Furthermore, we did not detect conformational changes (i) of pre-cleaved wild-type HCoV-OC43 S incubated with 100 mM 9-*O*-Ac-Me-Sia, (ii) after trypsin cleavage of 9-*O*-Ac-Me-Sia-bound wild-type HCoV-OC43 S or (iii) of pre-cleaved wild-type HCoV-OC43 S incubated at pH 4.5 (Supplementary Fig. 4c–f). Therefore, 9-*O*-Ac-Me-Sia binding and pH acidification of the medium, such as the one occurring in the endosomal compartment, did not trigger HCoV-OC43 S fusogenic conformational changes.

To evaluate the ability of our purified glycoprotein construct to undergo fusogenic conformational changes, we incubated the pre-cleaved wild-type HCoV-OC43 S ectodomain at 50 °C for 25 min in absence or presence of isopropanol (used to dissolve the trypsin inhibitor added to stop the proteolytic reaction) (Supplementary Fig. 4 g, h). In the latter conditions, we noticed the formation of HCoV-OC43 S rosettes arising from the nonspecific interactions of multiple post-fusion trimers via the hydrophobic fusion peptides (Supplementary Fig. 4h). These biochemical conditions lowered the energy barrier between the metastable pre-fusion state and the post-fusion (ground) state, acting as a surrogate for receptor-mediated fusion activation. This finding indicated that the wild-type HCoV-OC43 S ectodomain trimer could refold to the post-fusion conformation, although neither free

9-*O*-Ac-Me-Sia nor pH acidification triggered this transition. It has been previously established that caveolin-mediated endocytosis is a major route of HCoV-OC43 entry into host cells [63]. Because we demonstrated interactions of sialoglycans with the identified site are necessary for S-mediated viral entry, we hypothesize that membrane fusion occurs upon formation of multivalent interactions with sialoglycans (via mechanical destabilization of the pre-fusion trimers) and/or binding to a putative proteinaceous receptor [44], before or after virus internalization. In conclusion, 9-*O*-Ac-Sia-containing receptors appear to differ from the proteinaceous SARS-CoV receptor, because addition of monomeric angiotensin-converting enzyme 2 ectodomain to wild-type SARS-CoV S trimers, in the presence of trypsin, promoted refolding to the post-fusion state [23,25].

A conserved sialoside attachment strategy. HCoV-OC43, BCoV, PHEV and HCoV-HKU1 are the four coronaviruses known to engage 9-*O*-Ac-Sia-capped sialoglycans to initiate infection of target cells. The A domain of their S glycoproteins share strikingly similar structures that can be superimposed with a Cα r.m.s.d. between 0.8 and 2.0 Å (Supplementary Fig. 5a–d).

Virtually all residues participating in interactions with 9-*O*-Ac-Me-Sia or the formation of the binding groove are conserved in BCoV S and PHEV S, such as Asn27, Leu80, Lys81, Leu85, Leu86, Trp90, Phe95 and Thr31 (Fig. 4a–c). $Ser83_{HCoV\text{-}OC43}$, however, is substituted with $Thr83_{BCoV/PHEV,}$ and both side chains are expected to form a hydrogen bond with the C1-carboxylate of the ligand (Fig. 4a–c). These findings and the abrogation of BCoV and PHEV domain A–mediated hemagglutination of rat erythrocytes upon substituting Lys81/Thr83 or Trp90 with alanine [43] indicate that these two viruses interact with 9-*O*-Ac-Sia in an identical manner to HCoV-OC43 S. The binding pocket seems to be compatible with recognition of 9-*O*-Ac-Sia and of 9-*O*-acetyl-glycolyl-neuraminic acid. Although the latter saccharide is not found in humans, it is present at the termini of oligosaccharides, decorating other mammalian and avian glycoproteins and glycolipids, and could be a receptor for BCoV and PHEV.

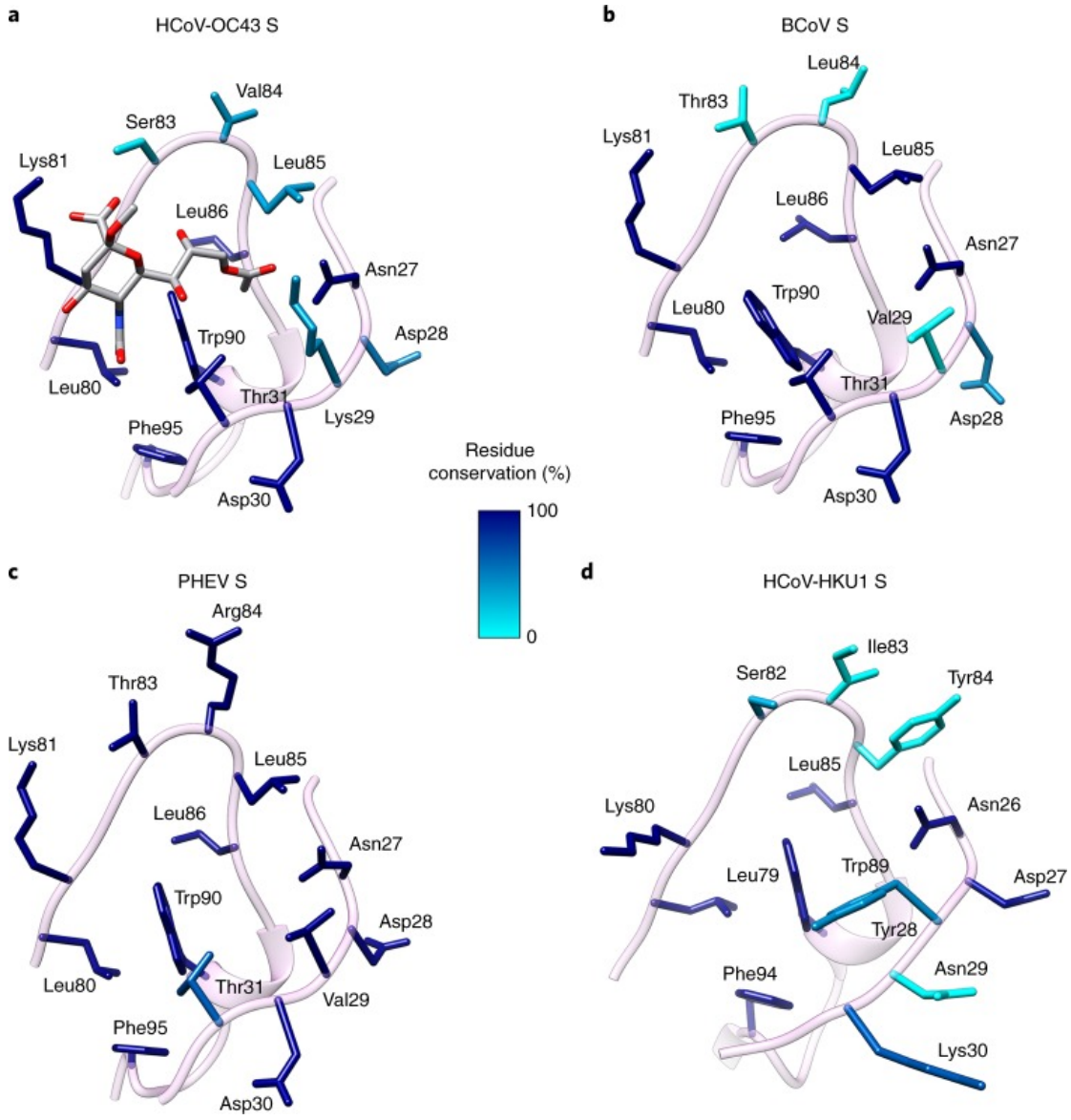


Fig. 4. Conservation of the receptor-binding groove among all 9-*O*-Ac-sialoglycan-recognizing coronaviruses. a–d, Zoomed-in view of the binding sites rendered as ribbon diagrams with surrounding residues shown as sticks for HCoV-OC43 (a), BCoV (b), PHEV (c), HCoV-HKU1 (d). Residues are colored by conservation, based on the analysis of all the S glycoprotein sequences available for each virus. In a, the 9-*O*-Ac-Me-Sia ligand is rendered as sticks with atoms colored by elements (carbon, gray; nitrogen, blue; oxygen, red). HCoV-OC43, 192 sequences; BCoV, 150 sequences; PHEV, 12 sequences; HCoV-HKU1, 28 sequences.

Many of the ligand-interacting residues or residues indirectly involved in formation of the recognition site identified in the holo-HCoV-OC43 S structure are also strictly conserved in HCoV-HKU1 S, such as $Asn26_{HCoV\text{-}HKU1}$ ($Asn27_{HCoV\text{-}OC43}$), $Leu79_{HCoV\text{-}HKU1}$ ($Leu80_{HCoV\text{-}OC43}$), $Lys80_{HCoV\text{-}HKU1}$ ($Lys81_{HCoV\text{-}OC43}$), $Leu85_{HCoV\text{-}HKU1}$ ($Leu86_{HCoV\text{-}OC43}$), $Trp89_{HCoV\text{-}HKU1}$ ($Trp90_{HCoV\text{-}OC43}$) and $Phe94_{HCoV\text{-}HKU1}$ ($Phe95_{HCoV\text{-}HKU1}$) (Fig. 4a,d), suggesting that HCoV-HKU1 S interacts with 9-*O*-Ac-sialoglycans using the same binding site as that identified for HCoV-OC43 S. This hypothesis is supported by site-directed mutagenesis

experiments showing that substitution of $Lys80_{HCoV\text{-}HKU1}$, $Thr82_{HCoV\text{-}HKU1}$ ($Ser83_{HCoV\text{-}OC43}$) or $Trp89_{HCoV\text{-}HKU1}$ with alanine abrogated HCoV-HKU1 domain A–mediated hemagglutination of rat erythrocytes [43].

Our results show that all coronaviruses recognizing host cell 9-*O*-Ac-sialoglycans share a conserved binding pocket and bind to the ligand via virtually identical interactions. Strikingly, BCoV HE and influenza HEF similarly interact with 9-*O*-Ac-Sia, despite ample differences in the architecture of their ligand-binding pockets [33,34,43]. Specifically, the two methyl groups of the ligand are docked into two hydrophobic depressions separated by an aromatic amino acid side chain, and hydrogen bonds are formed with the 5-nitrogen of the neuraminic acid core and the 9-*O*-acetyl carbonyl (Fig. 5a–c). The similarity across the three binding sites is reinforced by the observation that 9-*O*-Ac-Sia buries a comparable surface area at the interface with each of these glycoproteins and that the 9-*O*-acetyl moiety makes a major contribution to it in all three cases (~110 $Å^2$). One notable difference, however, is that the C1 carboxylate anchors the ligand to HCoV-OC43 S via a salt bridge and a hydrogen bond, whereas it relies on the formation of one or two hydrogen bonds with the BCoV HE or influenza HEF lectin domains, respectively. These results expand on our previous biochemical work [43] to demonstrate that BCoV HE and influenza HEF use structural principles similar to those of other 9-*O*-Ac-sialoglycan-recognizing human and animal coronaviruses for engagement to host cell receptors.

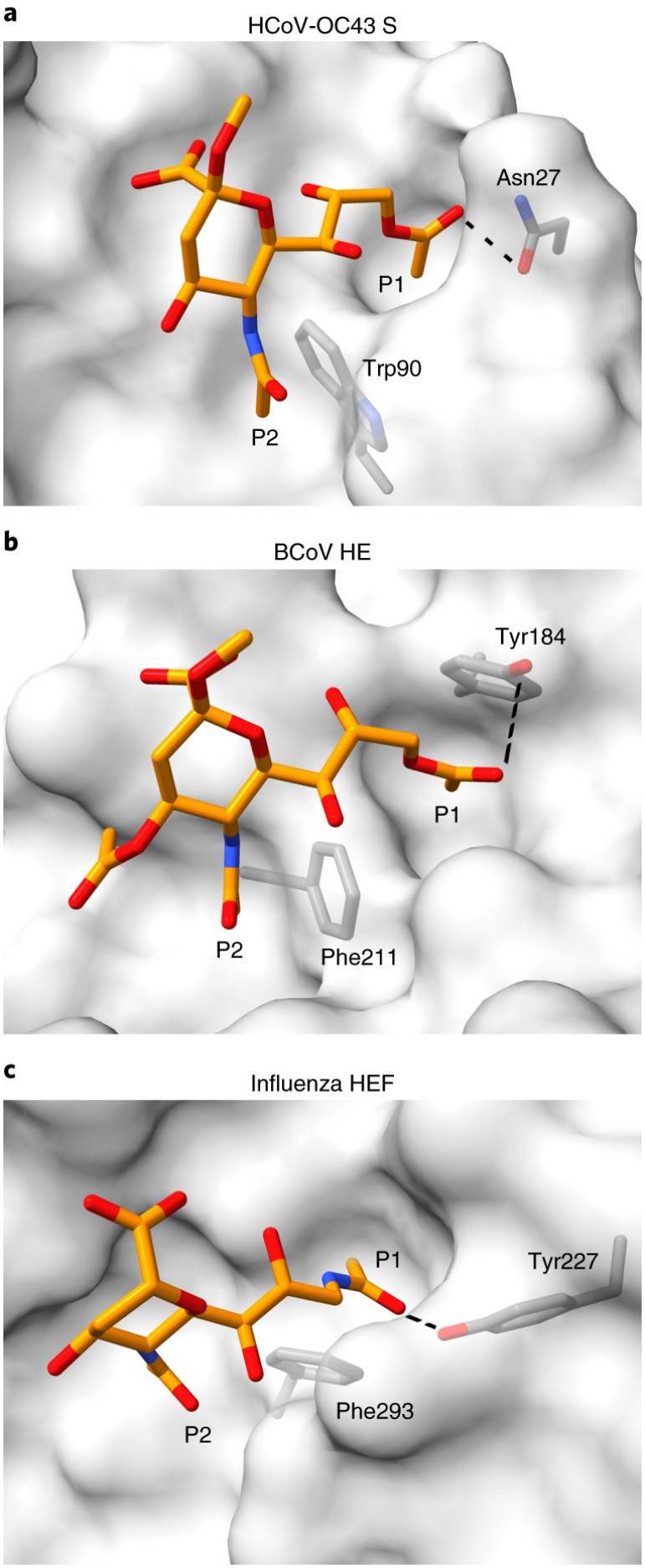

a
HCoV-OC43 S
Asn27
P1
Trp90
P2
b
BCoV HE
Tyr184
P1
P2
Phe211
c
Influenza HEF
P1
Tyr227
Phe293
P2

Fig. 5. Conservation of the receptor-binding site architecture across coronavirus S, coronavirus HE and influenza virus HEF glycoproteins. a, HCoV-OC43 S bound to 9-*O*-Ac-Me-Sia. b, BCoV HE bound to 5-*N*-acetyl-4,9-di-*O*-acetyl-neuraminic acid α-methylglycoside (PDB 3CL5). c, Influenza virus C HEF in complex with 9-*N*-Ac-Sia. In all panels, the glycoprotein is rendered as a gray surface with the bound ligand shown as sticks. The hydrogen bond formed with the carbonyl of the 9-*O*/*N*-acetyl group is shown by dashed lines.

Discussion

We structurally identified and characterized with unprecedented detail the HCoV-OC43 S sialoglycan-binding site, which is located in a groove at the surface of domain A. This site is conserved in all other coronaviruses known to attach to 9-*O*-Ac-Sia, including HCoV-HKU1 S (another endemic human coronavirus), and BCoV S (the presumptive zoonotic ancestor of HCoV-OC43). Our results provide a molecular framework explaining the specific recognition of 9-*O*-Ac-Sia-decorated oligosaccharides present at the surface of host cells targeted by these viruses. The β-sandwich architecture of domain A is conserved among all coronaviruses, and some of them feature a duplication of this domain at the S glycoprotein N-terminal region [16]. Other coronaviruses like MERS-CoV (β-coronavirus), infectious bronchitis virus (IBV, γ-coronavirus), porcine epidemic diarrhea virus (α-coronavirus) and transmissible gastroenteritis virus (α-coronavirus) have been described to also bind to sialoglycans (distinct from 9-*O*-Ac-sialosides) via their A domains during host cell infection [42,64–66]. The ligand-binding pocket identified in the holo-HCoV-OC43 S structure is not conserved in the MERS-CoV or in the IBV A domains, for which structures are available, suggesting that host attachment of this subset of viruses involve different interactions. The conserved topology of domain A among coronavirus S glycoproteins indicate that it derived from divergent evolution of an ancestral galectin domain. Viral evolution and adaptation thus lead to the use of distinct binding residues on the same domain putatively to acquire different ligand specificities such as 9-*O*-Ac-sialosides versus non-*O*-acetylated-sialoglycans. This evolutionary plasticity is reminiscent of what has been described for the BCoV HE lectin domain in comparison with influenza A/B hemagglutinin and influenza C/D HEF [33,35].

Sialic acids cap numerous oligosaccharides found at the surface of eukaryotic cells and constitute an important class of receptors for several human pathogens [35,37,38]. Modulation of attachment to sialoglycans can therefore have profound effects on zoonotic transmission, tropism and virulence of many viruses. For instance, a single point mutation in the highly pathogenic H5N1 avian influenza virus hemagglutinin was proposed to account for most of the preference switch from avian enteric tract receptors (α2,3-linked sialic acid) to human respiratory tract receptors (α2,6-linked sialic acid)

[57]. Although influenza A/B hemagglutinin, influenza C/D HEF and coronavirus HE have distinct architectures compared with those of coronavirus S glycoproteins, common rules of ligand engagement emerge. These rules also appear to extend to the interactions of sialoglycans with adenoviruses [67] and reoviruses [68]. In all cases, sialic acid binding involves burying a small surface area (300–400 $Å^2$) through contacts with a solvent-exposed groove of the protein. One face of the sialic acid ligand makes extensive interactions with the viral proteins, whereas the opposite, solvent-exposed face, makes few contacts. The binding affinity for sialic acids usually ranges between the micromolar and millimolar range, and the aforementioned viruses display numerous oligomeric spikes to enhance adsorption to target receptors through avidity [57].

Despite these similarities, marked differences in the 3D organization of the binding sites explain the selectivity of different viruses for unmodified or modified sialic acids. The ligand-binding sites of BCoV HE, influenza HEF and a subset of coronavirus S glycoproteins have evolved to specifically recognize 9-O-Ac-Sia via hydrogen bonding with the 9-*O*-acetyl carbonyl moiety and formation of a hydrophobic pocket accommodating the 9-*O*-acetyl methyl [32–34,43,54]. In contrast, influenza hemagglutinin cannot accommodate 9-*O*-acetylated neuraminic acids, owing to steric restrictions, but a subset of hemagglutinins can bind to N-glycolyl neuraminic acids [57,69]. The HCoV-OC43 S, HCoV-HKU1 S, BCoV S and PHEV S glycoproteins therefore share the ligand specificity of influenza C/D HEF but are functionally more similar to influenza A/B hemagglutinin, by carrying receptor attachment and membrane fusion functions, whereas a dedicated HE (coronaviruses) or neuraminidase (influenza A/B) is responsible for the receptor-destroying activity. In conclusion, our results illuminate how coronaviruses recognize 9-*O*-Ac-sialosides to enable infection of susceptible cells and show that a conserved strategy is utilized to engage such ligands across coronaviruses and orthomyxoviruses.

Methods

Construct design. The fragment encoding the HCoV-OC43 S ectodomain (residues 15–1263, UniProtKB: Q696P8) was amplified by (RT-)PCR from the viral genome and placed into a modified pCAGGS mammalian expression vector with a CD5 N-terminal signal peptide (MPMGSLQPLATLYLLGMLVASVLA) and an engineered C-terminal extension encoding a GCN4 trimerization motif (IKRMKQIEDKIEEIESKQKKIENEIARIKKIK), a thrombin cleavage site (underlined) (LVPRGSLE), and an eight-residue Strep-tag (WSHPQFEK) followed by a stop codon, as previously described [9,15–17]. This construct results in fusing the GCN4 trimerization motif in register with the HR2 helix at the C-terminal end of the HCoV-OC43 S-encoding ectodomain sequence. A mutant gene carrying three R-to-G amino acid mutations to abolish the furin cleavage (754-RRSRG-758 →

754-GGSGG-758) at the S1 – S2 junction (S2 cleavage site) was also generated following the same strategy. A pCAGGS vector encoding the HCoV-OC43 S domain A (residues 1 – 306) C-terminally extended with a thrombin cleavage site followed by the Fc region of human IgG was generated as described previously [70].

Protein expression and purification. HEK293F cells were grown in suspension using FreeStyle 293 Expression Medium (Life Technologies) at 37 °C in a humidified 8% CO2 incubator rotating at 130 r.p.m. Wild-type or mutant HCoV-OC43 S ectodomain construct were transfected into 250 ml cultures with cells grown to a density of 1 million cells per milliliter using 293fectin (ThermoFisher Scientific). After 4 d, supernatant was collected, and cells were kept in culture for an additional 4 d, yielding two harvests per transfection. Recombinant wild-type or mutant HCoV-OC43 S ectodomain was purified from clarified supernatants using a 1 ml StrepTrap column (GE Healthcare). Purified proteins were concentrated and flash frozen in Tris-saline buffer (20 mM Tris, pH 8.0, 150 mM NaCl) prior to negative staining and cryo-EM analysis.

Negative stain electron microscopy. Protein samples were adsorbed to glow-discharged carbon-coated copper grids for 30 s prior to 2% uranyl formate staining. Micrographs were recorded using the Leginon software71 on a 120 kV FEI Tecnai G2 Spirit with a Gatan Ultrascan 4000 CCD camera at 67,000 nominal magnification. The defocus ranged from 1.0 to 2.0 µm, and the pixel size was 1.6 Å.

Conformational change analysis using negative-staining electron microscopy and SDS-PAGE. Wild-type HCoV-OC43 S ectodomain trimer at 1 mg·ml^{-1} (6.6 µM spike monomer) was digested or not with trypsin at 14 µg·ml^{-1} for 30 min at room temperature, after which 1.5 mM PMSF was added to stop the reaction. The samples were subsequently incubated: either overnight at 4 °C with 100 mM 9-*O*-Ac-sia; 25 min at 50 °C; or 30 min at pH 4.5 using 20 mM sodium citrate buffer before being analyzed via negative-staining EM and SDS-PAGE.

Cryo-EM sample preparation and data collection. Three microliters of HCoV-OC43 S at 1 mg·ml^{-1} was applied to a 2/2 C-flat grid (Protochips) that had been glow discharged for 30 s at 20 mA. After preferential orientation was observed, 2.7 µl of HCoV-OC43 S at 10 mg·ml^{-1} was mixed with 0.3 µl of n-Octyl-β-D-glucopyranoside (OG) 180 mM immediately before being applied to a glow-discharged grid. Thereafter, grids were plunge frozen in liquid ethane using an FEI Mark IV Vitrobot with a 6.5–7.5 s blot time at 100% humidity and 20 °C. Incubation of 1.1 µM HCoV-OC43 S with 100 mM 9-*O*-acetylated sialic acid (9-*O*-Ac-sia) was performed overnight at 4 °C, and immediately before vitrification, OG was added to the mixture reaction at a final concentration of 18 mM. Data were

acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. Automated data collection was carried out using Leginon71 at a nominal magnification of 130,000× with a pixel size of 0.525 Å for apo-HCoV-OC43 S (super-resolution mode) and 1.05 Å for holo-HCoV-OC43 S (counted mode). The dose rate was adjusted to 8 counts/pixel/s, and each movie was dose-fractionated in 50 (apo) or 60 (holo) frames of 200 ms. A total of 2,211 and 2,402 micrographs were respectively collected for apo- and holo-HCoV-OC43 S, with a defocus range between 1.3 and 1.8 µm.

Cryo-EM data processing. Movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking and extraction were carried out using Warp [72]. Particle images were extracted with a box size of 800 binned to 400 for apo-HCoV-OC43 S or with a box size of 400 for holo-HCoV-OC43 S, both yielding a pixel size of 1.05 Å. Reference-free 2D classification in Relion was used to parse particles from the original 197,791 and 332,912 for apo- and holo-HCoV-OC43 S, respectively. The MHV S cryo-EM map [9] was used to generate an initial model of apo-HCoV-OC43 S. The initial model of holo-HCoV-OC43 S was generated using the apo-HCoV-OC43 S map. Relion 3D classification without symmetry was used to select ~83,000 and ~178,000 particles from apo- and holo-HCoV-OC43 S, respectively. CTF refinement in Relion3.0 (ref. [73]) was used to refine per-particle defocus values. Particle images were subjected to the Bayesian polishing procedure implemented in Relion3.0 (refs. [73,74]) before performing another round of per-particle defocus refinement. The particles were then subjected to 3D classification without refining angles and shifts using the same soft mask as that used during 3D refinement and with a tau value of 30. Final 3D refinement of the apo- and holo-HCoV-OC43 S datasets imposing C3 symmetry was carried out using non-uniform refinement in cryoSPARC75 and yielded reconstructions at 2.9- and 2.8-Å resolution, respectively. Local resolution estimation, filtering and sharpening was carried out using CryoSPARC. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) of 0.143 criterion76, and FSC curves were corrected for the effects of soft masking by high-resolution noise substitution [77].

Cryo-EM model building and analysis

UCSF Chimera [78] and Coot [79] were used to fit the MHV atomic model (PDB 3JCL) into the holo-HCoV-OC43 S cryo-EM map. The models were subsequently manually rebuilt using Coot [79]. N-linked glycans were hand built into the density where visible, and the models were rebuilt and refined using Rosetta80–83. Models were analyzed using MolProbity [84], Privateer [85] and PISA

[86]. Figures were generated using UCSF Chimera [78] and ChimeraX [87]. Analysis of the ligand-binding site electrostatic surface potential was performed using PDB 2PQR [88] and APBS [89].

Biolayer interferometry. HCoV-OC43 $S1^A$-Fc was expressed in HEK293T cells and purified from the cell culture supernatant by protein A chromatography, as described [43]. Monomeric domain A, wild type or with a W90A substitution, was subsequently obtained by on-the-bead thrombin cleavage [33], after which the proteins were concentrated to up to 3.8 mg·ml^{-1} in PBS, aliquoted and stored at −80 °C until further use. Biolayer interferometry analysis was performed on an Octet RED384 machine. All assays were performed using Fortebio Kinetics Buffer (KB; PBS supplemented with 0.1% BSA, 0.02% Tween20 and 0.05% sodium azide) at 30 °C. Synthetic biotinylated 6-sialyl-5-N-, 9-*O*-acetyl-lactosamine (9OAc6SLN) or 6-sialyl-5-*N*-acetyl-lactosamine (6SLN) dissolved to 100 nM were loaded onto streptavidin (SA) biosensors to maximum loading levels (until no further increase in reflection was observed). Sensors were washed in KB until a stable baseline was obtained. Binding of monomeric HCoV-OC43 S domain A was performed by moving receptor-loaded sensors to wells containing 100 µl of purified protein, dissolved in KB to various concentrations, for up to 3 min, then dissociating for 3 min dissociation. To abolish unspecific binding, sensors were subjected to five successive association/dissociation cycles. To test whether binding of domain A was sialate-9-*O*-acetyl-dependent, biosensors loaded with 9OAc6SLN were de-*O*-acetylated by dipping them in wells containing 20 µg·ml^{-1} porcine torovirus P4 HE-Fc [45] in KB for 30 min, then washing prior to association/dissociation (pre-HE) or after a cycle of association/dissociation, upon which the biosensors were subjected to a final cycle (post-HE). The equilibrium dissociation constant, KD, was determined from three independent experiments with the 'Response' option of the Octet Data Analysis software. The half-life of the domain A–9OAc6SLN complex was calculated manually from the dissociation curves.

Pseudovirus entry assays. HCoV-OC43 S–pseudotyped VSV-ΔG particles were prepared as previously described43. Briefly, HEK293T cells at 70% confluency were transfected with PEI-complexed plasmid DNA. For coexpression of HCoV-OC43 S and BCoV HE-Fc, S expression vectors and pCD5-BCoV HE-Fc were mixed at molar ratios of 8:1. At 48 h after transfection, cells were transduced with VSV-G–pseudotyped VSVΔG/Fluc [90] at a multiplicity of infection of 1. Cell-free supernatants were harvested at 24 h after transduction and filtered through 0.45-µm membranes, and virus particles were purified and concentrated via sucrose cushion ultracentrifugation at approximately 100,000g for 3 h. Pelleted virions were resuspended in PBS and stored at −80 °C until further use. Inoculation of HRT18 monolayers in 96-well format was performed with equal amounts of S-pseudotyped virions, as calculated from VSV-N content (roughly corresponding to the yield from 2×10^5 transfected and transduced cells), diluted in 10% FBS-supplemented DMEM. At 18 h post infection,

cells were lysed using passive lysis buffer (Promega). Firefly luciferase expression was measured using a firefly luciferase assay system. Infection experiments were performed independently in triplicate, each time with three technical replicates. Pseudovirus incorporation of flag-tagged OC43 S was determined for the parental type and each of the mutants via Western blotting and by calculating the S content (measured with monoclonal antibody ANTI-FLAG M2; Sigma) relative to that of VSV-N (measured with anti-VSV-N monoclonal antibody 10G4; Kerafast).

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences (R01GM120553 to D.V.), a Pew Biomedical Scholars Award (D.V.) and an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund (D.V.), the Zoonoses Anticipation and Preparedness Initiative (IMI115760, F.A.R. and B.-J.B.), the Pasteur Institute (M.A.T. and F.A.R.), the Centre National de la Recherche Scientifique (F.A.R.), the LabEx Integrative Biology of Emerging Infectious Diseases (F.A.R.), the Netherlands Organization for Scientific Research (NWO TOP-PUNT 718.015.003, G.-J.B.) and CSC grant 2014-03250042 (Y.L.). This work was also partly supported by the Arnold and Mabel Beckman cryo-EM center and the Proteomics Resource (UWPR95794) at the University of Washington, and the Electron Imaging Center for NanoMachines supported by the NIH (1U24GM116792, 1S10RR23057 and 1S10OD018111), NSF (DBI-1338135) and CNSI at UCLA. We thank Y. Matsuura (Osaka University, Japan) for provision of the VSV-G pseudotyped VSVΔG/Fluc plasmids.

Contributions

M.A.T., A.C.W., Y.L., R.J.d.G. and D.V. designed the experiments. C.W. and B.-J.B. designed and cloned the protein constructs. M.A.T. and C.W. carried out protein expression and purification. Z.L. and G.-J.B. provided key reagents; M.A.T. and A.C.W. performed cryo-EM sample preparation and data collection. M.A.T., A.C.W. and D.V. processed the cryo-EM data. M.A.T. and D.V. built and refined the atomic models. Y.L. and D.K. carried out the binding and pseudovirus assays. M.A.T., A.C.W., Y.L., R.J.d.G., F.A.R. and D.V. analyzed the data. M.A.T., A.C.W. and D.V. prepared the manuscript with input from all authors.

References

1. Ge XY, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013;503:535–538.
2. Haagmans BL, et al. Middle east respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect. Dis. 2014;14:140–145.
3. Su S, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol. 2016;24:490–502.
4. Isaacs D, Flowers D, Clarke JR, Valman HB, MacNaughton MR. Epidemiology of coronavirus respiratory infections. Arch. Dis. Child. 1983;58:500–503.
5. Menachery VD, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat. Med. 2015;21:1508–1513.
6. Menachery VD, et al. SARS-like WIV1-CoV poised for human emergence. Proc. Natl Acad. Sci. USA. 2016;113:3048–3053.
7. Hu B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog. 2017;13:e1006698.
8. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 2003;77:8801–8811.
9. Walls AC, et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature. 2016;531:114–117.
10. Millet JK, Whittaker GR. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. Virus Res. 2015;202:120–134.
11. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl Acad. Sci. USA. 2009;106:5871–5876.
12. Millet JK, Whittaker GR. Host cell entry of middle east respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc. Natl Acad. Sci. USA. 2014;111:15214–15219.
13. Walls AC, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc. Natl Acad. Sci. 2017;114:11157–11162.
14. Burkard C, et al. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog. 2014;10:e1004502.
15. Walls A, et al. Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. Protein Sci. 2017;26:113–121.
16. Walls AC, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat. Struct. Mol. Biol. 2016;23:899–905.
17. Xiong X, et al. Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections. J. Virol. 2018;92:e01628-17.
18. Kirchdoerfer RN, et al. Pre-fusion structure of a human coronavirus spike protein. Nature. 2016;531:118–121.
19. Kirchdoerfer RN, et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Sci. Rep. 2018;8:15701.
20. Pallesen J, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc. Natl Acad. Sci. USA. 2017;114:E7348–E7357.
21. Gui M, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. Cell Res. 2017;27:119–129.
22. Shang J, et al. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. PLoS Pathog. 2018;14:e1007009.
23. Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog. 2018;14:e1007236.
24. Shang, J.et al. Cryo-EM structure of porcine delta coronavirus spike protein in the pre-fusion state. J. Virol.10.1128/JVI.01556-17 (2017).
25. Walls AC, et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. Cell.

2019;176:1026–1039.e15.

26. Vijgen L, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. 2005;79:1595–1604.
27. Lau SK, et al. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. J. Virol. 2011;85:11325–11337.
28. Vlasak R, Luytjes W, Spaan W, Palese P. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. Proc. Natl Acad. Sci. USA. 1988;85:4526–4529.
29. Huang X, et al. Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. J. Virol. 2015;89:7202–7213.
30. de Groot RJ. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. Glycoconj. J. 2006;23:59–72.
31. Desforges M, Desjardins J, Zhang C, Talbot PJ. The acetyl-esterase activity of the hemagglutinin-esterase protein of human coronavirus OC43 strongly enhances the production of infectious virus. J. Virol. 2013;87:3097–3107.
32. Bakkers MJ, et al. Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. Cell Host Microbe. 2017;21:356–366.
33. Zeng Q, Langereis MA, van Vliet AL, Huizinga EG, de Groot RJ. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc. Natl Acad. Sci. USA. 2008;105:9065–9069.
34. Rosenthal PB, et al. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. Nature. 1998;396:92–96.
35. Song H, et al. An open receptor-binding cavity of hemagglutinin-esterase-fusion glycoprotein from newly-identified influenza D virus: basis for its broad cell tropism. PLoS Pathog. 2016;12:e1005411.
36. Herrler G, et al. The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. EMBO J. 1985;4:1503–1506.
37. Stencel-Baerenwald JE, Reiss K, Reiter DM, Stehle T, Dermody TS. The sweet spot: defining virus-sialic acid interactions. Nat. Rev. Microbiol. 2014;12:739–749.
38. Neu U, Bauer J, Stehle T. Viruses and sialic acids: rules of engagement. Curr. Opin. Struct. Biol. 2011;21:610–618.
39. Schauer R, Kamerling JP. Exploration of the sialic acid world. Adv. Carbohydr. Chem. Biochem. 2018;75:1–213.
40. Peng G, et al. Crystal structure of bovine coronavirus spike protein lectin domain. J. Biol. Chem. 2012;287:41931–41938.
41. Peng G, et al. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc. Natl Acad. Sci. USA. 2011;108:10696–10701.
42. Li W, et al. Identification of sialic acid-binding function for the middle east respiratory syndrome coronavirus spike glycoprotein. Proc. Natl Acad. Sci. USA. 2017;114:E8508–E8517.
43. Hulswit RJG, et al. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proc. Natl Acad. Sci. USA. 2019;116:2681–2690.
44. Collins AR. HLA class I antigen serves as a receptor for human coronavirus OC43. Immunol. Invest. 1993;22:95–103. [PubMed] [Google Scholar]
45. Langereis MA, et al. Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. Cell Rep. 2015;11:1966–1978.
46. Bakkers MJ, et al. Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. Proc. Natl Acad. Sci. USA. 2016;113:E3111–E3119.
47. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science. 2005;309:1864–1868. [PubMed] [Google Scholar]
48. Lu G, et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature. 2013;500:227–231.
49. Ou X, et al. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. Nat. Commun. 2017;8:15216.
50. Yuan Y, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat. Commun. 2017;8:15092.

51. Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B. A resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. eLife. 2015;4:e06380.

52. Seetharaman J, et al. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. J. Biol. Chem. 1998;273:13047–13052. [PubMed] [Google Scholar]

53. Dormitzer PR, Sun ZY, Wagner G, Harrison SC. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J. 2002;21:885–897.

54. Langereis MA, Zeng Q, Heesters BA, Huizinga EG, de Groot RJ. The murine coronavirus hemagglutinin-esterase receptor-binding site: a major shift in ligand specificity through modest changes in architecture. PLoS Pathog. 2012;8:e1002492.

55. Kunkel F, Herrler G. Structural and functional analysis of the surface protein of human coronavirus OC43. Virology. 1993;195:195–202.

56. Fei Y, et al. Characterization of receptor binding profiles of influenza a viruses using an ellipsometry-based label-free glycan microarray assay platform. Biomolecules. 2015;5:1480–1498.

57. Xiong X, et al. Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature. 2013;497:392–396. [PubMed] [Google Scholar]

58. Guo H, et al. Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. PLoS Pathog. 2018;14:e1007233.

59. Sakai, T., Takagi, H., Muraki, Y. & Saito, M. Unique directional motility of influenza C Virus controlled by its filamentous morphology and short-range motions. J. Virol.10.1128/JVI.01522-17 (2018). [PMC free article] [PubMed]

60. Sakai T, Nishimura SI, Naito T, Saito M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. Sci. Rep. 2017;7:45043.

61. Mukherjee S, et al. Mechanism and significance of cell type-dependent neutralization of flaviviruses. J. Virol. 2014;88:7210–7220.

62. Park JE, et al. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. Proc. Natl Acad. Sci. USA. 2016;113:12262–12267.

63. Owczarek K, et al. Early events during human coronavirus OC43 entry to the cell. Sci. Rep. 2018;8:7124.

64. Wickramasinghe IN, de Vries RP, Grone A, de Haan CA, Verheije MH. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. J. Virol. 2011;85:8903–8912.

65. Liu C, et al. Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. J. Virol. 2015;89:6121–6125.

66. Schultze B, et al. Transmissible gastroenteritis coronavirus, but not the related porcine respiratory coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. J. Virol. 1996;70:5634–5637.

67. Nilsson EC, et al. The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. Nat. Med. 2011;17:105–109. [PubMed] [Google Scholar]

68. Reiter DM, et al. Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. PLoS Pathog. 2011;7:e1002166.

69. Ito T, et al. Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. Virology. 1997;227:493–499. [PubMed] [Google Scholar]

70. Raj VS, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature. 2013;495:251–254.

71. Suloway C, et al. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 2005;151:41–60. [PubMed] [Google Scholar]

72. Tegunov, D. C., P. Real-time cryo-EM data pre-processing with Warp. Preprint at: https://www.biorxiv.org/content/10.1101/338558v1 (2018).

73. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife7, e42166 (2018). [PMC free article] [PubMed]

74. Zivanov J, Nakane T, Scheres SHW. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ. 2019;6:5–17.

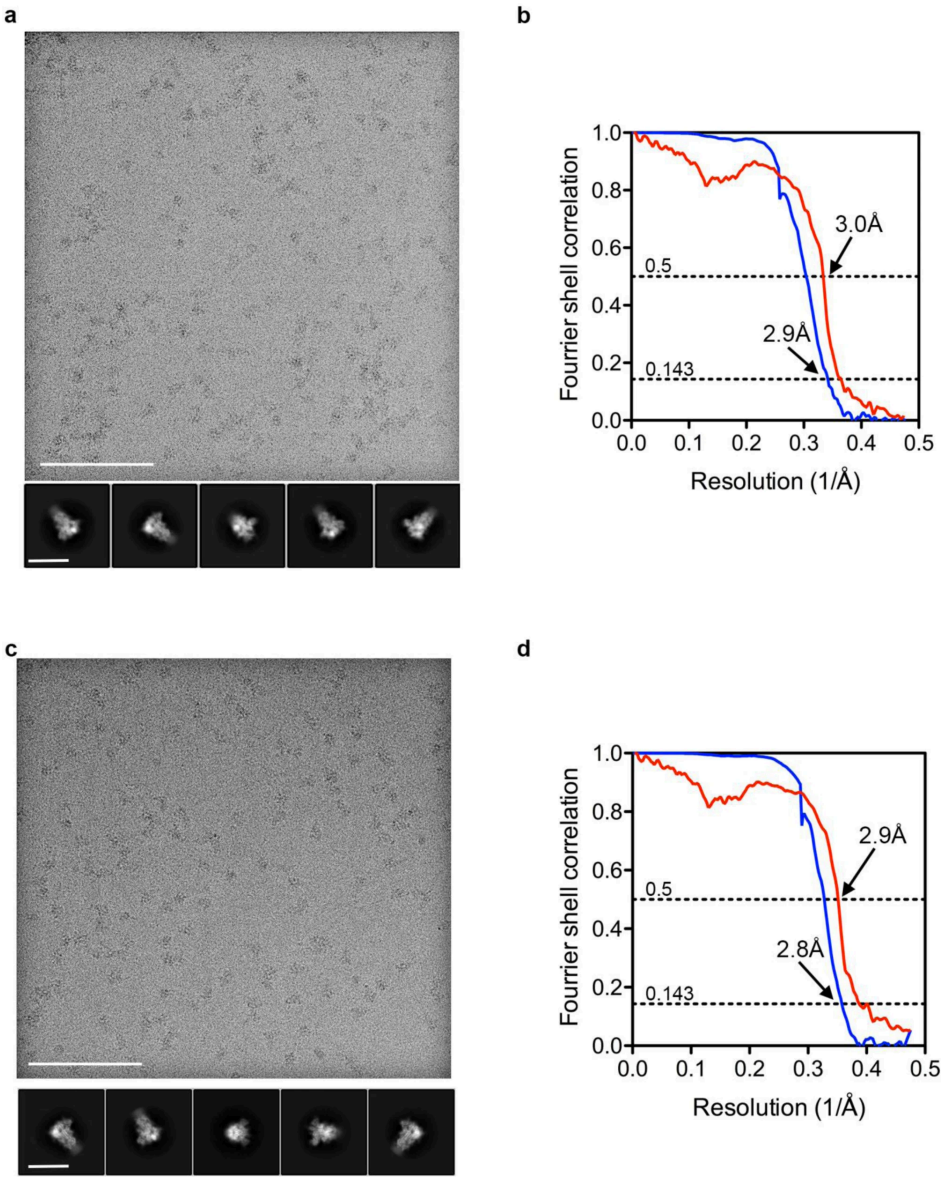
75. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods. 2017;14:290–296. [PubMed] [Google Scholar]

76. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in

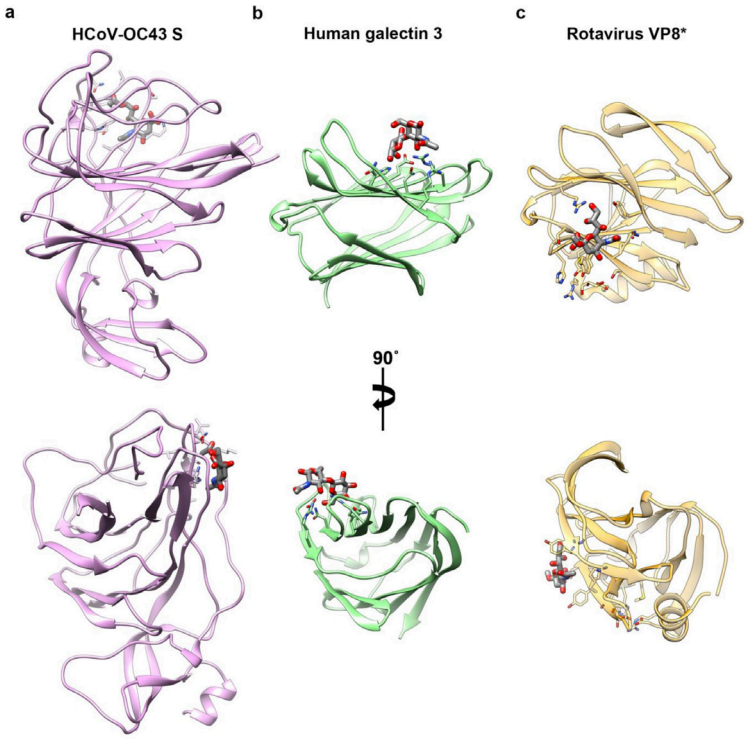
single-particle electron cryomicroscopy. J. Mol. Biol. 2003;333:721–745. [PubMed] [Google Scholar]

77. Chen S, et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy. 2013;135:24–35.
78. Goddard TD, Huang CC, Ferrin TE. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 2007;157:281–287. [PubMed] [Google Scholar]
79. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 2010;66:486–501.
80. Wang RY, et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. eLife. 2016;5:e17219.
81. Frenz BDR, et al. Automatically fixing errors in glycoprotein structures with Rosetta. Structure. 2019;27:1–6.
82. DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I. Modeling symmetric macromolecular structures in Rosetta3. PLoS One. 2011;6:e20450.
83. DiMaio F, et al. Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. Nat. Methods. 2015;12:361–365.
84. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D. Biol. Crystallogr. 2010;66:12–21.
85. Agirre J, et al. Privateer: software for the conformational validation of carbohydrate structures. Nat. Struct. Mol. Biol. 2015;22:833–834. [PubMed] [Google Scholar]
86. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 2007;372:774–797. [PubMed] [Google Scholar]
87. Goddard TD, et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 2018;27:14–25.
88. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004;32:W665–W667.
89. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl Acad. Sci. USA. 2001;98:10037–10041.
90. Kaname Y, et al. Acquisition of complement resistance through incorporation of CD55/decay-accelerating factor into viral particles bearing baculovirus GP64. J. Virol. 2010;84:3210–3219.

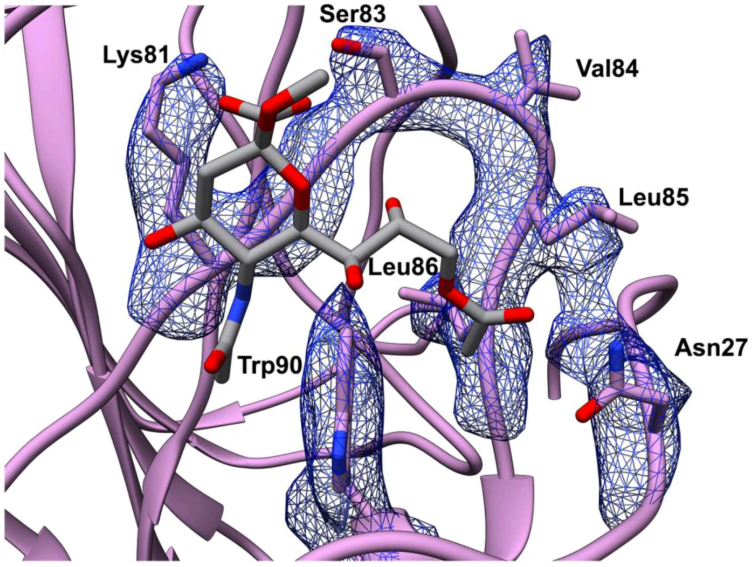
Supplementary Information



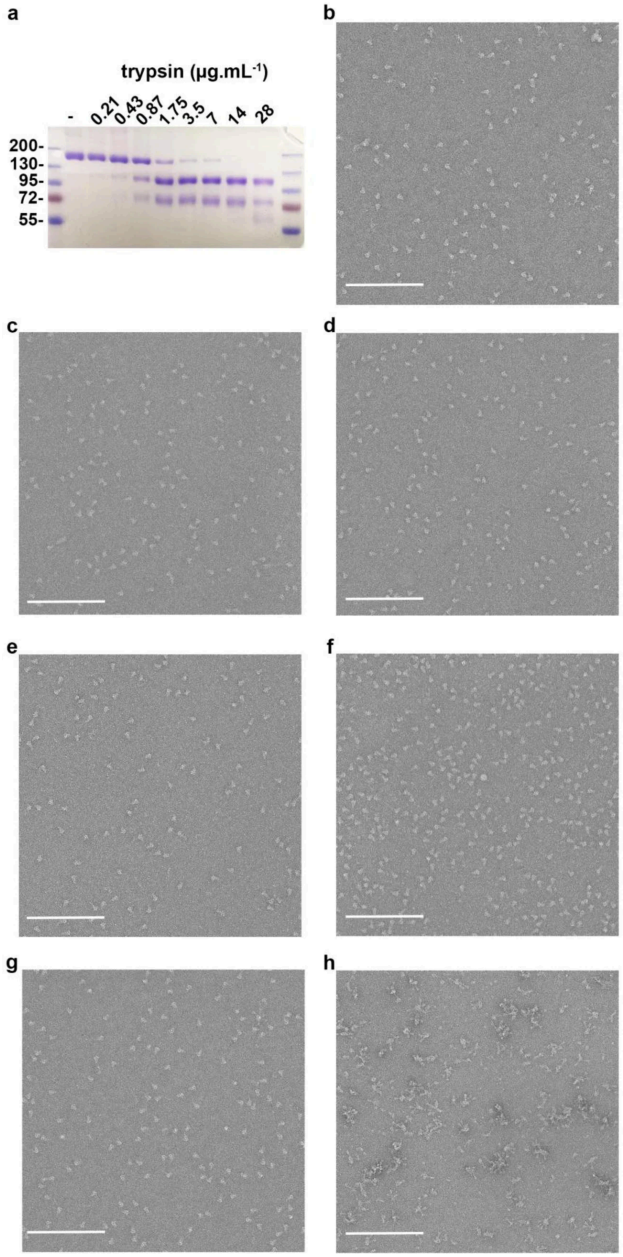
Supplementary Figure 1. CryoEM analysis of apo and holo HCoV-OC43 S glycoprotein. a, c, Representative electron micrographs (defocus: 1.4 µm and 1.2 µm) and class averages of apo (a) and holo (c) HCoV-OC43 S trimer embedded in vitreous ice. Scale bars: 100 nm (micrograph) and 200 nm Å (class averages). b, d Gold-standard (blue) and model/map (red) Fourier shell correlation curves. The resolution was determined to 2.9 Å and 2.8 Å for the apo and holo HCoV-OC43 S glycoprotein maps, respectively. The 0.143 and 0.5 cut-off values are indicated by horizontal dashed lines.



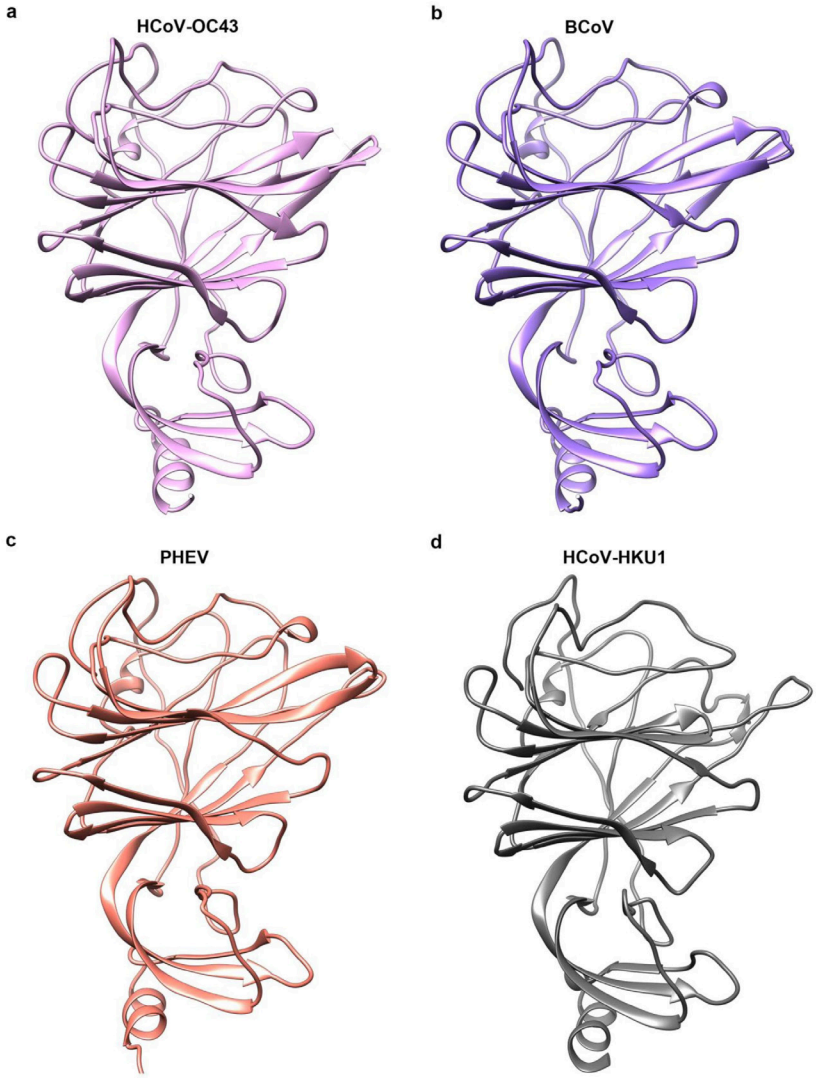
Supplementary Figure 2. Carbohydrate ligands are recognized by distinct regions of the HCoV-OC43 S, human galectin-3 or rotavirus VP8* β-sandwich. (a-c) Ribbon diagrams in two orthogonal orientations showing 9-O-Ac-Me-Sia bound to the HCoV-OC43 S domain A (a), galactose bound to human galectin-3 (b, PDB 1A3K) and sialic acid bound to rotavirus VP8* (c, PDB 1KQR). Each domain is shown in the same two orientations for comparison.



Supplementary Figure 3. Zoomed-in view of the sialoglycan-binding site in the holo-HCoV-OC43 S glycoprotein structure. The A domain is rendered as a ribbon diagram with the side chains of key ligand-interacting residues shown as sticks and the corresponding cryoEM density shown as a blue mesh. The ligand is rendered as sticks with atoms colored by elements (carbon: grey, nitrogen: blue, oxygen: red).



Supplementary Figure 4. Free 9-*O*-Ac-Me-Sia receptor does not trigger HCoV-OC43 S fusogenic conformational changes. (a) SDS-PAGE. (b) Purified wild-type HCoV-OC43 S ectodomain trimer in the prefusion conformation. (c-f) The wild-type HCoV-OC43 S ectodomain trimer remained in the prefusion conformation after cleavage with 28 µg.mL−1 trypsin (c), incubation with 100 mM 9-*O*-Ac-Me-Sia of pre-cleaved trimer (d) or trypsin cleavage of 9-*O*-Ac-Me-Sia-bound trimer (e), cleavage with 28 µg.mL^{-1} trypsin and incubation at pH 4.5 for 1 hour at room temperature (f), as visualized by single-particle electron microscopy of negatively stained samples. The wild-type HCoV-OC43 S ectodomain trimer was cleaved with 28 µg.mL^{-1} trypsin and heated for 20 minutes at 50 °C in the absence (g) or presence (h) of 10% isopropanol (used to dissolve the trypsin inhibitor added to stop the proteolytic reaction). Only the latter condition led to the formation of postfusion rosettes. Scale bars: 200 nm.



Supplementary Figure 5. Coronavirus S galectin-like A domains have a conserved architecture. (a–d) Ribbon diagrams of the HCoV-OC43 (a), BCoV (b, PDB 4H14), PHEV (c, PDB 6QFY) and HCoV-HKU1 (d, PDB 5I08) A domains.

Chapter 5

Coronavirus hemagglutinin-esterase and spike proteins co-evolve for functional balance and optimal virion avidity

Yifei Lang[1], Wentao Li[1], Zeshi Li[2], Danielle Koerhuis[1], Arthur C.S. van den Burg[1], Erik Rozemuller[3], Berend-Jan Bosch[1], Frank J.M. van Kuppeveld[1], Geert-Jan P.H. Boons[2,4,5], Eric G. Huizinga[6], Hilde M. van der Schaar[3], and Raoul J. de Groot[1]*

[1]Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CH Utrecht, the Netherlands

[2]Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CG Utrecht, The Netherlands

[3]GenDx B.V., Yalelaan 48, 3584 CM Utrecht, the Netherlands

[4]Department of Chemistry, University of Georgia, Athens, GA 30602

[5]Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602

[6]Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Sciences, Utrecht University, 3584 CH Utrecht, the Netherlands

Proc Natl Acad Sci U S A. 2020 in revision

Abstract

Human coronaviruses OC43 and HKU1 are respiratory pathogen of zoonotic origin that have gained worldwide distribution. OC43 apparently emerged from a bovine coronavirus (BCoV) spill-over. All three viruses attach to 9-*O*-acetylated sialoglycans via spike protein S with hemagglutinin-esterase HE acting as a receptor-destroying enzyme. In BCoV, an HE lectin domain promotes esterase activity towards clustered substrates. OC43 and HKU1, however, lost HE lectin function as an adaptation to humans. Replaying OC43 evolution, we knocked-out BCoV HE lectin function and performed forced evolution-population dynamics analysis. Loss of HE receptor-binding selected for second-site mutations in S, decreasing S binding affinity by orders of magnitude. Irreversible HE mutations selected for cooperativity in virus swarms with low-affinity S minority variants sustaining propagation of high-affinity majority phenotypes. Salvageable HE mutations induced successive second-site substitutions in both S and HE. Apparently, S and HE are functionally interdependent and co-evolve to optimize the balance between attachment and release. This mechanism of glycan-based receptor usage, entailing a concerted, fine-tuned activity of two envelope protein species, is unique among CoVs, but reminiscent of that of influenza A viruses (IAVs). Apparently, general principles fundamental to virion-sialoglycan interactions prompted convergent evolution of two important groups of human and animal pathogens.

Significance Statement

Human coronavirus OC43 arose relatively recently, presumably from a bovine coronavirus (BCoV) spill-over. Both viruses use 9-*O*-acetylated sialoglycans as receptors to which they attach via spike protein S. Another envelope protein, hemagglutinin-esterase (HE), serves as a receptor-destroying enzyme (RDE). In BCoV but not OC43, HE also mediates receptor-binding, which enhances RDE activity towards clustered glycotopes in mucus and on the cell surface. We demonstrate the existence of a functional balance between S-mediated attachment and HE lectin-controlled catalytic virion elution, which required resetting upon cross-species transmission. Such a two-protein mechanism for dynamic attachment is unique among CoVs, but reminiscent of that of influenza A viruses. Apparently, general principles fundamental to virion-sialoglycan interactions prompted convergent evolution of two zoonotically-relevant groups of pathogens.

Introduction

The subfamily *Orthocoronavirinae* comprises a group of enveloped positive-strand RNA viruses of clinical and veterinary significance. Adding to the socio-economic impact of coronaviruses (CoVs) already extant in humans and livestock, the emergence of 'new' CoVs through cross species transmission poses an ever-looming threat to public health, animal health, and food production.

Seven coronaviruses are known to infect humans, but not all of them have become established. The introduction of SARS CoV in 2002 from horseshoe bats with masked palm civets as incidental transient hosts, was rapidly contained through quarantine measures [1]. MERS CoV, natural to dromedary camels, causes a classical zoonotic infection with limited human-to-human spread [2]. December 2019, a member of the species *Severe acute respiratory syndrome related coronavirus*, 79.5% identical to the 2002 SARS CoV variant, emerged in Wuhan, China [3,4]. The outcome of this outbreak, which may range from containment to full scale pandemicity, will become clear in the coming weeks. There is a distinct possibility new variant SARS-CoV eventually becomes established in the human population.

Four other respiratory coronaviruses of zoonotic origin already did succeed in becoming true human viruses with world-wide distribution [5–7]. Among them are HKU1 and OC43 (subgenus *Embecovirus*, genus *Betacoronavirus*), related yet distinct viruses that arose from different zoonotic progenitors and entered the human population independently. OC43 is far more related to bovine coronavirus (BCoV), its presumptive ancestor, with early isolates sharing 97% genome identity [8,9]. Together with viruses of swine, canines, equines and lagomorphs, OC43 and BCoV are considered host range variants of the virus species *Betacoronavirus-1* (collectively referred to as β1CoVs throughout) [7]. OC43 apparently emerged 70 to 130 years ago from a single cross species transmission event that gave rise to a human-only virus [8–10]. Like OC43, other β1CoVs also exhibit host specificity [8,11]. While these observations attest to the host promiscuity and zoonotic potential of embecoviruses and β1CoVs in particular, they are also indicative for the existence of host barriers, the breaching of which selects for adaptive mutations that result in host specialization and, ultimately, virus speciation. Conceivably, comparative studies of BCoV and OC43 may identify factors that promote or restrict cross species transmission of CoVs and thus further our understanding of the requirements for colonization of the human host.

Embecoviruses, OC43 and BCoV included, differ from other CoVs in that they encode two types of surface projections. Homotrimeric 'peplomers' comprised of spike protein S and extending 20 nm from the viral membrane, mediate virion attachment to entry receptors and membrane fusion [12]. Interspersed are stubby 8-nm homodimeric projections comprised of the hemagglutinin-esterase

(HE) [13–15], a dual function protein typically encompassing a receptor-binding lectin domain specific for *O*-acetylated sialic acid (*O*-Ac-Sia) and a receptor destroying sialate-*O*-acetylesterase domain [16–20]. The HE lectin domain contributes to virion attachment, but at the same time enhances sialate-*O*-acetylesterase activity towards clustered sialoglycotopes [11].

Some embecoviruses, like mouse hepatitis virus (MHV) and related CoVs in rodents, attach to 4- or 9-*O*-acetylated sialosides (4- or 9-*O*-Ac-Sias) via HE [21–25] and to a proteinaceous entry receptor via S [26,27]. Others, animal β1CoVs included, bind to 9-*O*-Ac-Sias via HE [28] but, remarkably, also via S [29,30] or, in the case of human coronaviruses OC43 and HKU1, even exclusively via S [11,31,32].

Structure function analyses of HE and S proteins have yielded a wealth of data on ligand binding, substrate selection and protein-glycan interactions. The receptor binding sites (RBSs) of CoV HE lectin domains and those in related proteins of toro- and influenza C/D viruses [22,33–35] differ in sequence and structure yet conform to a common architectural design with a deep hydrophobic pocket ('P1') to accommodate the crucial sialate-*O*-acetyl moiety, and an adjacent pocket or depression ('P2') to accept the 5-*N*-acyl group [17,21,22,33]. Characteristically, P1 and P2 are separated by an aromatic side chain and binding of the ligand is stabilized further through electrostatic protein-glycan interactions typically involving distinctive Sia functions such as the Sia glycerol side chain, the 5-*N*-Acyl and/or the Sia carboxylate. The RBS for 9-*O*-Ac-Sia in the S proteins of BCoV and OC43, identified by comparative structural analysis [32] and confirmed by the cryo-EM holostructure of OC43 S [36], conforms to this blueprint. Moreover, this site is structurally and functionally conserved in HKU1 [32].

Much less is known about the functional relationship between S and HE, and the role of HE in particular remains poorly understood. In MHV, HE expression is dispensable for replication and rapidly lost during cell culture propagation [15]. Conversely, in β1CoVs, HE seems critical for infection. In OC43, loss of HE-associated acetyl esterase activity abrogates the production of infectious virus and virus dissemination in cell culture [37]. Moreover, acetyl-esterase inhibitors impede BCoV replication [19], and antibodies against HE neutralize the virus *in vitro* and *in vivo* [38–40]. Still, even among β1CoVs there are differences in HE function apparently correlating with host specificity. Whereas HE lectin activity is strictly maintained in BCoV [28], OC43 lost this function through progressive accumulation of mutations in the HE RBS, apparently as an adaptation to replication in the human respiratory tract [11]. Neverthless, isolates of either virus propagate in cultured cells. To better understand the consequences of loss of HE lectin function as it occurred during OC43 and also HKU1 evolution, we took a reverse genetics/forced evolution approach with BCoV as a model. The findings reveal that HE and S are functionally interdependent and that the

acquisition of HE by a proto-embecovirus allowed its β1CoV descendants to adopt strategies for reversible virion-sialoglycan attachment, remarkably similar to those of influenza A viruses.

Results

Disruption of HE lectin function selects for mutations in S. To study the role and importance of HE in β1CoV propagation, we developed a reverse genetics system for BCoV strain Mebus (sFig. 1A) based on targeted RNA recombination [41,42]. Recombinant 'wildtype' BCoVs (rBCoV) with parental type HE and S, but with accessory ORF4a replaced by the *Renilla* luciferase gene (rBCoV-Rluc), were readily generated upon seeding acceptor-virus-infected, donor RNA-transfected mouse LR7 cells onto monolayers of feeder HRT18 cells. rBCoV-Rluc arose and within seven days grew to final titers routinely obtained for wildtype BCoV (~10^8 TCID50/ml).

Generating BCoV-Rluc derivatives defective in HE lectin function proved more cumbersome. To abolish the HE RBS, we substituted Phe^{211}, which is key to ligand binding [17](sFig. 1B), by Ala via two nucleotide substitutions to block reversion. Mutant viruses were recovered eventually, but, in 3 out of 4 successful trials, a multistep 160-hr rescue did not suffice and an additional 72-96 hr blind passage was required (Fig. 1A).

Sequence analysis of clonal virus populations, obtained by endpoint dilution, confirmed the presence of the HE Phe^{211}Ala substitution in all cases. Surprisingly, the purified viruses all suffered single site mutations in S, clustering in domain $S1^A$ (aa 15-302) in proximity of the RBS (Figs. 1A-C; sFig. 2). Two of the trials yielded multiple S variants and some variants -Thr^{83}Ile and Leu^{89}Pro- emerged independently in separate experiments (Fig. 1A). The mutations map to three distinct S RBS elements (Fig. 1C, sFig. 2C; nomenclature according to [32]). Ile^{26}Ser and Asn^{27}Ser locate in the $S1^A$ β1 element within the N-terminal L1-β1-L2 segment (aa 15-33) that walls pocket P1; P1 is crucial for ligand binding as it accommodates the all-important sialate-9-*O*-acetyl moiety [32,36]. Moreover, in the OC43 S cryo-EM holostructure, the Asn^{27} side chain hydrogen bonds with the 9-*O*-acetyl carbonyl [36]. Leu^{89}Pro in $S1^A$ element $3_{10}1$ is immediately adjacent to Trp^{90}. The latter is arguably the most critical residue in the RBS as its indole side chain separates the P1 and P2 pockets, and its replacement precludes receptor-binding and virus infectivity [32]. Finally, Gly^{82}Glu and Thr^{83}Ile substitutions occurred in $S1^A$ element β5 that interacts with the sialate carboxylate through hydrogen bonding with Lys^{81} and Thr^{83}.

As measured by solid phase lectin binding assay (sp-LBA) with $S1^A$-Fc fusion proteins and bovine submaxillary mucin (BSM) as ligand, all mutations significantly reduced S binding to 9-*O*-Ac-Sia albeit to widely different extents. $S1^A$-Fc binding affinities of the mutants were 500-fold ($Asn^{27}Ser$) to more than 10.000-fold ($Ile^{26}Ser$; $Gly^{82}Glu$) lower than that of parental BCoV $S1^A$-Fc (Fig. 1D).

Loss of the HE lectin RBS would both abolish HE-assisted virion attachment to 9-*O*-Ac-Sia receptor determinants as well as reduce virion-associated receptor-destroying sialate-*O*-acetylesterase activity towards clustered glycotopes [11]. The findings suggest that the HE-$Phe^{211}Ala$ substitution creates a fitness defect that can be alleviated not by increasing but by dramatically reducing the affinity of the S RBS. Thus, the defect apparently is caused by loss of virion-associated RDE activity rather than by loss of overall virion avidity.

Loss of HE lectin function exerts a fitness cost by affecting reversible virion attachment. In principle, downregulation of HE esterase activity consequential to loss of lectin function, could affect virus propagation at two stages of the infectious cycle, namely virus release, which would require depletion of intracellular and cell surface receptor pools, and (pre)attachment. We recently reported that G-deficient vesicular stomatitis (VSV) virions pseudotyped with wildtype BCoV S require exogenous HE for efficient infection [32]. $S1^A$ mutations that reduce S affinity inhibit infection, but, as we now show, only when HE is present. VSV virions pseudotyped with low affinity mutant S proteins were less reliant on or even inhibited by exogenous HE with decreasing S RBS affinity (Fig. 1E). The findings provide direct proof that the $S1^A$ mutations act at the level of virion attachment and support the notion that the $S1^A$ mutations, selected for in HE-defective rBCoVs, restore reversibility of receptor binding by lowering S affinity and thereby protect against inadvertent attachment to decoy receptors.

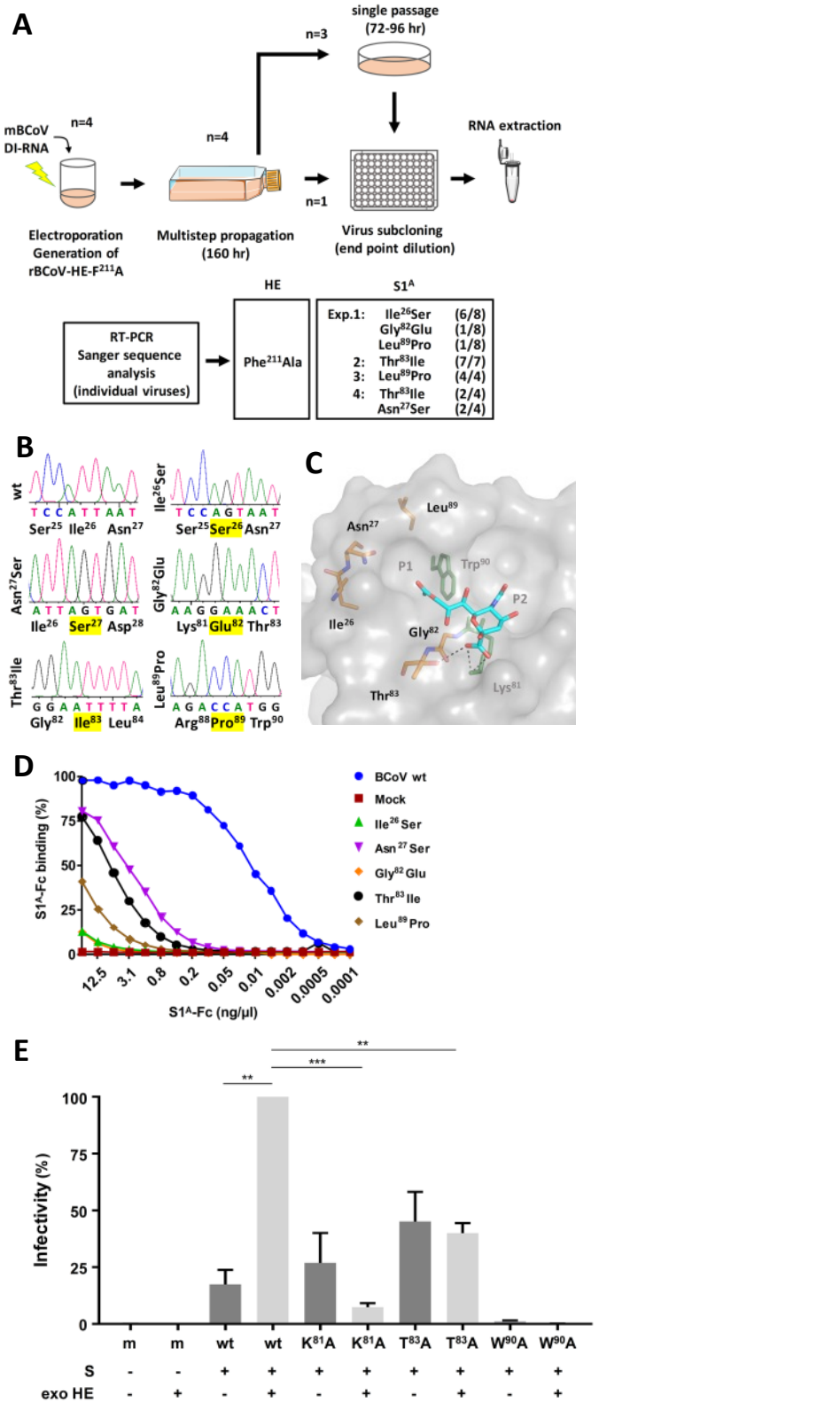


Fig. 1. Site-directed mutagenesis of BCoV HE by targeted recombination; loss of HE lectin activity selects for second-site mutations in domain 1^A of the spike protein. (*A*) Schematic outline of four independent experiments, depicting each step from targeted RNA recombination and virus rescue to virus purification and genetic analysis of the resultant clonal populations by RT-PCR and Sanger sequencing. The number of virus clones, found to contain a particular $S1^A$ mutation, relative to the total number of virus clones analyzed are

given (between parenthesis). (*B*) Relevant portions of Sanger DNA sequencing chromatograms of the $S1^A$ coding region in recombinant wildtype BCoV and in cloned rBCoV-HE-Phe^{211}Ala derivatives. Amino acid substitutions marked in yellow. (*C*) Second-site mutations in $S1^A$ locate in close proximity of the RBS. Close-up of the BCoV $S1^A$ RBS (in surface representation; PDB:4H14), with 9-*O*-Ac-Sia (in sticks, colored by element; oxygen, red; nitrogen, blue; carbons, cyan) modeled in the RBS [32], showing the locations of the mutations. Key elements of the RBS (hydrophobic pockets P1 and P2, and the side chains of RBS residues Lys^{81} and Trp^{90} in sticks, colored green) are indicated. Side chains and connecting main chains of RBS mutations also shown in sticks, but with carbon colored orange. Predicted hydrogen bonds between the Sia carboxylate moiety and the side chains of Lys^{81} and Thr^{83} shown as black dashed lines. (*D*) The $S1^A$ mutations strongly reduce binding to 9-*O*-Ac-Sia. Mutant $S1^A$-Fc fusion proteins in twofold serial dilutions, starting at 2.5 µg/well, were tested by sp-LBA for their binding to BSM relative to that of wildtype $S1^A$-Fc. Binding expressed in percentages with maximum binding of wildtype $S1^A$-Fc set to 100%. (*E*) Infectivity of VSV particles, pseudotyped with low affinity BCoV S variants, is inhibited rather than promoted by soluble 'exogenous' sialate-*O*-acetylesterase. HRT18 cells were inoculated with equal amounts of G-deficient VSV particles, pseudotyped with BCoV S or mutants thereof, either with (+) or without (-) soluble exogenous sialate-*O*-acetylesterase ('exo HE') added to the inoculum. 'Infectivity' expressed in RLUs in cell lysates at 18 h p.i., normalized to those measured for VSV-S^{wt}. The data shown are averages from three independent experiments, each of which performed with technical triplicates. SDs and significant differences, calculated by Welch's unequal variances *t* test, are indicated (**$P \leq 0.01$; ***$P \leq 0.001$).

HE lectin-deficient recombinant BCoVs are genetically stable when grown in the presence of exogenous receptor-destroying enzyme. To test whether loss of virion-associated RDE activity in rBCoV-HE-F^{211}A/S^{wt}/Rluc might be compensated for by adding exogenous soluble HE to the culture medium, we seeded infected/transfected LR7 cells onto HRT18 cell monolayers, supplemented the cell culture supernatant with BCoV HE-Fc [17] to final concentrations of 1 pg to 10 µg/ml, and allowed infection to proceed for 120 hr. While in the absence of HE-Fc there was no sign of virus propagation as detectable by IFA, concentrations of exogenous sialate-*O*-acetylesterase as low as 1 ng/ml to up to 1 µg/ml promoted virus growth (Fig. 2A).

To determine whether these conditions would allow isolation of rBCoV-HE-F^{211}A without mutations in $S1^A$, we performed targeted recombination and rescued recombinant viruses by 160 hr multistep propagation as before, but now with culture supernatant supplemented with 100 ng/ml HE-Fc (Fig. 2B). Sanger sequence analysis of RT-PCR amplicons showed that all viruses cloned by endpoint dilution of the 160-hr stock (n=4) coded for mutant HE-Phe^{211}Ala in combination with wildtype $S1^A$. To assess the stability of clonal rBCoV-HE-F^{211}A/S^{wt}/Rluc, the virus population resulting from a subsequent 120-hr amplification in the presence of exogenous HE-Fc was analyzed by Next-Generation Sequencing (NGS), which allows for the detection of low frequency mutants.

Sequence variation in HE and $S1^A$ was distributed randomly and did not exceed background levels (<0.15%). More than 99.5% of the viruses coded for HE-$Phe^{211}Ala$, while preserving parental type $S1^A$ (Fig. 2B).

Loss of HE lectin function gives rise to mixed virus population with competition and cooperativity among S affinity variants. With a clonal, virtually pure stock of rBCoV-HE-$F^{211}A/S^{wt}$ available, we performed controlled forced evolution experiments. The virus was serially passaged involving three consecutive 120-hr multistep propagation rounds in HRT18 cells but now in the absence of exogenous HE-Fc, with the initial infection at multiplicity of infection (MOI) of 0.005 (Fig. 2C). In trial 1, viral titers in passage 1 (p1) increased only slowly to 3×10^4 and 2×10^4 TCID50/ml (measured with or without exogenous HE-Fc, respectively). The withdrawal of exogenous RDE during viral passage immediately selected for mutations in $S1^A$. Virus cloning by endpoint dilution of the 120-hr p1 sample yielded S RBS mutants $Asn^{27}Ser$, $Thr^{83}Ile$ and $Ile^{26}Ser$ (Fig. 2C) – all three of which had been seen before (Fig. 1A). NGS analysis revealed the true complexity of the p1 population (Fig. 2C) and identified two additional $S1^A$ variants with substitutions -$His^{173}Tyr$ and $Arg^{197}Cys$- more distal from the RBS (sFigs. 2D, 3A, B). $His^{173}Tyr$ also reduced the relative binding affinity of $S1^A$-Fc albeit less dramatically than the other mutations, namely by 30-fold (Table 1). The $Arg^{197}Cys$ mutation seemingly falls in a separate category as it reduced $S1^A$-Fc expression levels by more than 90% suggestive of defective folding (sFig. 3C). Apparently, aberrant disulfide-bonding causes most of the fusion protein to be retained in the ER with only a minor, presumably properly folded fraction slipping through to become secreted.

All in all, the p1 population was comprised for virtually 100% of HE-$Phe^{211}Ala$ mutants, 40% of which in combination with parental BCoV S, the remaining 60% with second-site mutations in $S1^A$ (Fig. 2C; Table 1). Of the latter, the ultra-low affinity variant $S1^A$-$Ile^{26}Ser$ was the most abundant at 46% and the $Asn^{27}Ser$ variant the least at less than 1%. However, upon a subsequent round of 120-hr multistep propagation, the tables were turned with $S1^A$-$Asn^{27}Ser$ now comprising almost 40% of the p2 population and the $Ile^{26}Ser$ variant reduced to 0.7%. In addition, four other $S1^A$ variants emerged. One of these had a mutation in $S1^A$ RBS loop L1, $Val^{29}Gly$, and a relative binding affinity close to that of the $Asn^{27}Ser$ mutant (Fig. 2C; Table 1). We also identified at position 75 a second $S1^A$ Cys-substitution mutant, which like $Arg^{197}Cys$, presumably disrupts the RBS through aberrant disulfide bonding (sFig. 3). Remarkably, two other $S1^A$ variants arose with mutations -$Arg^{88}Thr$ and $Pro^{174}Leu$- that affected the relative binding affinity only modestly to 0.25 and 0.5 of that of wildtype $S1^A$-Fc, respectively (Table 1). Even more remarkably, upon further passage these mutants increased to dominate the p3 population, effectively outcompeting variants with low affinity spikes as well as those with parental spikes (Fig. 2C, D). However, when the p1, p2 and p3 stocks were cloned by

endpoint dilution in the absence of exogenous HE-Fc, only virus variants with low affinity mutations in $S1^A$ were isolated (Fig. 2C; Table 1). Strikingly, from the p3 stock, the $S1^A$-$Asn^{27}Ser$ variant was isolated exclusively against all odds (10/10 tested; $p < 10^{-6}$) when calculated purely from its frequency in the population (21%). Conversely, virus cloning by endpoint dilution in the presence of exogenous HE-Fc yielded high affinity $S1^A$ mutants $Pro^{174}Leu$ (5/11) and $Arg^{88}Thr$ (4/11), parental virus rBCoV-HE-$Phe^{211}Ala/S^{wt}$ (1/11), and intermediate S affinity variant $His^{173}Tyr$ (1/11).

Notably, the conditions selected not only for mutations in S but also in HE. Variants with an Ala211Val substitution in HE emerged in p2, rising to 17% of the p2 end population, to stabilize around this frequency in p3. As a result of this mutation, HE lectin affinity was regained albeit to levels solely detectable by high-sensitivity nanobead hemagglutination assay, while esterase activity towards clustered glycotopes in BSM increased 4-fold as compared to HE-Phe211Ala, but still remained 125-fold lower than that of wildtype HE (sFig. 4). Apparently, the increase in HE function, minor as it may be, provides a selective advantage, but apparently one that benefits both low and high affinity S variants, because the mutation was found in cloned viruses of either type.

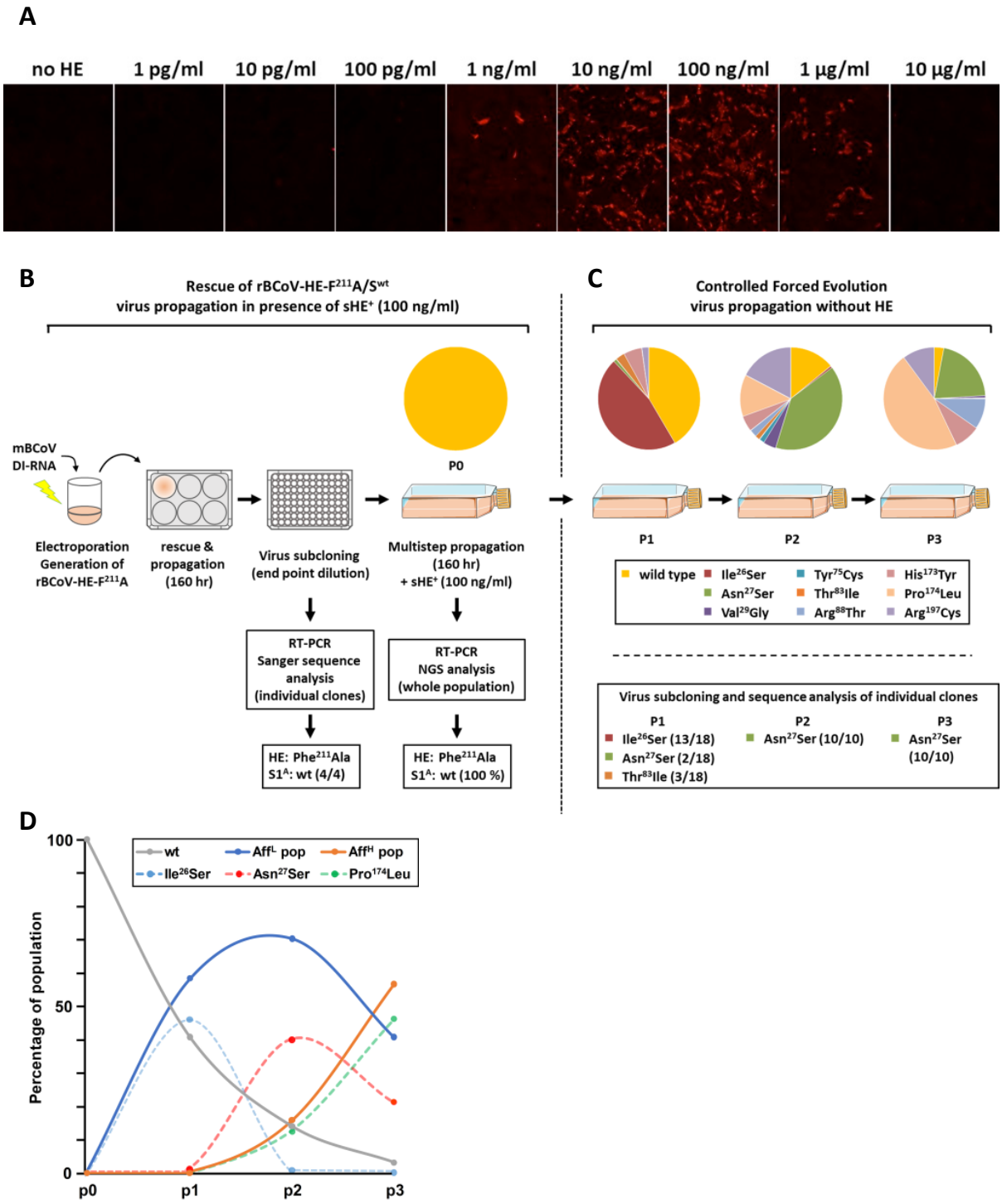


Fig. 2. Stable propagation and controlled directed evolution of rBCoV-HE-F^{211}A. (A) rBCoV-HE-F^{211}A propagation and spread is enhanced by soluble exogenous sialate-*O*-acetylesterase. mBCoV-infected LR7 cells, donor RNA-transfected to generate rBCoV-HE-F^{211}A, were seeded on HRT18 cell monolayers to rescue recombinant viruses with cell culture supernatants supplemented with purified BCoV HE^{+}-Fc at concentrations indicated. Cell supernatants, harvested 120 hr after seeding, were inoculated onto HRT18 cells grown on glass coverslips. After a single infectious cycle (12 hr p.i.), infected cells were identified by immunofluorescence assay. Infected cells stained red. (B) Stable maintenance of wildtype S protein in rBCoV-HE-F^{211}A in the

presence of exogenous HE and (C) forced evolution in the absence thereof. Visual representation of experimental procedures and findings. Generation of recombinant rBCoV-HE-F^{211}A by targeted recombination was as in Fig. 1, but with rescue, cloning and virus amplification steps performed with tissue culture media supplemented with 100ng/ml exogenous HE-Fc. Rescued virus was purified by endpoint dilution. Individual clonal populations were characterized for HE and S1^{A} master sequences by extracting viral genomic RNA from the cell culture supernatant followed by RT-PCR and Sanger sequencing. One clonal population was used to grow a p0 stock of rBCoV-HE-F^{211}A and HE and S1^{A} diversity was assessed by next gen illumina sequence analysis (NGS). The virus was used to inoculate 5 × 10^{6} HRT18 cells at an MOI of 0.005 TCID50/cell and serially passaged. Cell culture supernatants were harvested at 120 hr p.i. and virus diversity was assessed by subcloning and genetic analysis of purified viruses as in Fig. 1. In addition, viral RNA was extracted, and diversity determined by RT-PCR amplification and NGS. Frequencies of S1^{A} variants are presented in pie charts with individual color coding as indicated. (D) Propagation of rBCoV-HE-F^{211}A quasispecies selects for variants with (near) wildtype S affinity. Stylized graph representation depicting the emergence and decline of viral variants during serial passage of rBCoV-HE-F^{211}A in the absence of exogenous HE. Changes in the frequencies of variants with wildtype S and groups of variants with low affinity S (sum of Ile^{26}Ser, Asn^{27}Ser, Val^{29}Gly, Thr^{83}Ile, His^{173}Tyr and Arg^{197}Cys) and high affinity S (sum of Arg^{88}Thr and Pro^{174}Leu) are depicted with solid lines, colored in gray, blue and orange, respectively. Those of individual S variants, Ile^{26}Ser, Asn^{27}Ser and Pro^{174}Leu, are shown in dashed lines and colored light blue, red and green, respectively.

Loss of HE lectin function selects for virus swarms with low affinity S escape mutants promoting the emergence and propagation of high-affinity S variants. To corroborate our observations, the controlled forced evolution experiment was repeated (sFig. 5). As compared to the first trial, there was a much faster population built-up already in p1 at 120 hr p.i. with final titers reaching 4 x 10^{8} and 3.4 x 10^{7} TCID50/ml, when measured in the presence or absence of exogenous HE-Fc, respectively. Surprisingly, in stark contrast to trial 1, the trial 2 p1 population was comprised for about 94% of viruses expressing wildtype BCoV S. Less than 6% consisted of variants with mutations in S1^{A}, four of low receptor binding affinity (Thr^{22}Ile, Asn^{27}Tyr, Val^{29}Gly, His^{173}Tyr), one of near-wildtype binding affinity (Pro^{174}Leu), and, with the exception of Thr^{22}Ile, all at positions seen before (Tables 1, 2). Consistent with our previous findings, however, virus purification through endpoint dilution in the absence of exogenous HE-Fc yielded low affinity mutants (11/11 tested) exclusively (sFig. 5). If the variants in the trial 2 p1 population were all of equal replicative fitness under the conditions applied, the odds of this result would be less than 1.10^{-12}.

At first glance, the two trials would seem to differ in their outcomes. We offer, however, that the results are in fact consonant and that the main difference is in the speed with which the virus populations increased and evolved. There is an inherent stochastic element to the experimental approach and whether the developing quasispecies undergoes slow track (experiment 1) or fast track

evolution (experiment 2) is likely dependent on the time of advent of the first mutant virus and its properties, for instance whether it is an ultralow (like Ile^{26}Ser) or low affinity variant (like Asn^{27}Ser). The findings allow for several conclusions. (i) They confirm and firmly establish that loss of HE lectin function selects for mutations in S1^{A} that reduce S receptor-binding affinity and virion avidity. (ii) The possibilities to reduce the affinity of the S RBS through single site mutations are finite. In several independent experiments, substitutions in S1^{A} occurred at a limited number of positions albeit not necessarily by the same residue. For example, Asn^{27} was replaced both by Ser and Tyr. (iii) The mutations that reduce S affinity fall into different categories. Most map within or in close proximity of the RBS to affect receptor-ligand interaction directly. Others, like His^{173}Tyr and Pro^{174}Leu, are more distal from the RBS and apparently affect ligand binding indirectly through long range conformational effects. A third type of mutations, quasi-random Cys substitutions, apparently disrupt S1^{A} folding by promoting non-native disulfide-bonding. While an Arg^{197}Cys substitution strongly decreased secretion of the S1^{A}-Fc fusion protein, the biosynthesis and intracellular transport of native trimeric spikes was seemingly affected to lesser extent. At least, the uptake of S-Arg^{197}Cys into VSV pseudotypes was not noticeably impaired as compared to that of wildtype S (sFig. 3E). Still, the mutation did alter the infectivity of the pseudotyped particles making them less dependent on exogenous HE-Fc (sFig. 3F) presumably by reducing the avidity of S trimers through local S1^{A} misfolding and consequential disruption of the RBS in one or more monomers. (iv) Perhaps most surprisingly, quasispecies developed in which loss of HE lectin function was compensated at the level of the viral population with minority low affinity variants, constituting less than 6% of the swarm, not only sustaining the replication of high affinity variants but actually allowing the latter to flourish and amplify to become the majority phenotype.

S and HE proteins co-evolve to attain functional balance and optimal virion avidity. Among the first mutations fixed upon zoonotic introduction and early emergence of OC43, was a HE-Thr^{114}Asn substitution, which created a glycosylation site at the rim of the lectin domain RBS [11] (Fig. 3A). Glycans attached to HE Asn^{114} hamper binding to 9-*O*-Ac-Sia through steric hindrance, causing a 500-fold reduction in HE avidity (Fig. 3B) and a 125-fold in sialate-*O*-acetylesterase-activity, respectively (Fig. 3C). We introduced the HE Thr^{114}Asn substitution in BCoV, expecting that the glycosylation site would be rapidly lost through any of several single-nucleotide restorative mutations in HE. Indeed, NGS analysis of the virus swarm arising after targeted recombination showed the glycosylation site to be destroyed but only in 10% of the population and exclusively by Ser^{116}Phe substitution (Fig. 3D). This mutation partially restores HE receptor binding and receptor destruction to 0.125 and 0.17 of that of wildtype HE, respectively (Fig. 3B, C). In the vast majority of viruses, the newly introduced HE glycosylation site was retained and, instead, S1^{A} mutations that

reduced S affinity were selected again, with S-RBS Thr^{83} replaced either by Ile (69%) -as seen before (Figs. 1A, 2C; Table 1)- or by Asn (10%) (Fig. 3D). The latter mutation reduces S1^{A} affinity to 0.008 of that of wildtype.

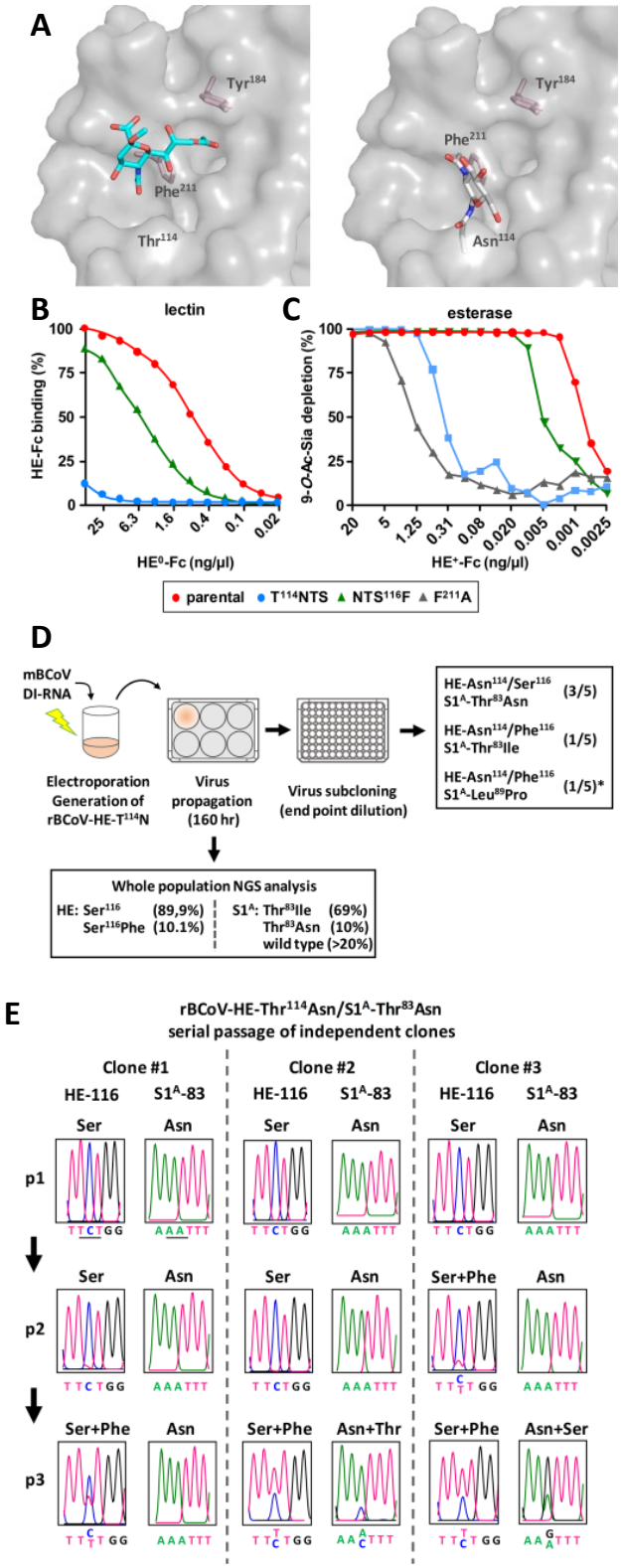


Fig. 3. BCoV S and HE co-evolve to restore functional balance and optimal virion avidity. (A) Introduction of an N-glycosylation site at the rim of the HE RBS. Side-by-side close ups of the BCoV HE holostructure [17] (PDB 3CL5) in surface representation with 9-*O*-Ac-Sia (in sticks, colored by element; oxygen, red; nitrogen, blue; carbons, cyan) bound to the RBS (left) or without the ligand and with the $Thr^{114}Asn$ substitution and N-linked glycan modelled in the RBS (in sticks, colored by element as above but with carbons in white) (right). Modelling performed by superpositioning of the OC43 strain NL/A/2005 HE structure (PDB 5N11) and confirmed in Coot. HE RBS key residues are indicated, with side chains of Tyr^{184} and Phe^{211} shown in sticks. (B) Loss of HE lectin function upon introduction of an N-glycosylation site through HE-$Thr^{114}Asn$ substitution and partial restoration through a second-site $Ser^{116}Phe$ mutation. Sp-LBA with serial dilutions of enzyme-inactive HE^{0}-Fc as in Fig. 1D. (C) Consequences for HE esterase activity towards clustered ligands. On-the-plate receptor depletion assay with bovine submaxillary mucin as substrate as in [11]. The assay was performed with 2-fold serial dilutions of enzymatically active HE^{+}-Fc and residual 9-*O*-Ac-Sia measured by sp-LBA with a fixed amount of HE^{0}-Fc. (D) Visual representation of experimental procedures and findings as in Figs. 1 and 2. Note that of five virus clones purified by endpoint dilution, the identity of one isolate could not be established and was deduced from subsequent propagation experiments as explained in the text (marked with *; see also Fig. 4). (E) Serial passage of rBCoV-HE-$Thr^{114}Asn$/$S1^{A}$-$Thr^{83}Asn$ selects for successive mutations in HE and S to restore their function to near wildtype levels. Results are shown for three independent isolates. Viral RNA extracted from tissue culture supernatants collected at the end of each passage was characterized by RT-PCR and Sanger DNA sequencing. Relevant portions of Sanger DNA sequencing chromatograms are presented to show changes in the master sequence of the virus populations at the coding sequence for the N-glycosylation site (HE codon 116) and for the site of the low affinity S mutation selected initially (S1A codon 83).

Virus cloning by endpoint dilution yielded, in three out of five isolates, $S1^{A}$-$Thr^{83}Asn$ variants with the newly introduced N-glycosylation site in HE intact (HE-$Thr^{114}Asn$). Furthermore, a single $S1^{A}$-$Thr^{83}Ile$ variant was isolated, but this virus in addition had the N-glycosylation site in HE destroyed (HE-$Thr^{114}Asn$/$Ser^{116}Phe$) (Fig. 3D). The observations led us to entertain the possibility that the mutations in $S1^{A}$ and HE did not occur independently and that, even in viruses expressing low affinity spikes, partially restorative mutations in HE would yet provide a selective advantage. To test this, the clonal $S1^{A}$-$Thr^{83}Asn$/HE-$Thr^{114}Asn$ variants were serially passaged. All three viruses independently lost HE Asn^{114} glycosylation over time and, saliently, through $Ser^{116}Phe$ substitution exclusively. Even more remarkably, with HE-$Ser^{116}Phe$ mutants gaining dominance, variants emerged that had restored S affinity to (near)wildtype through substitution of $S1^{A}$-Asn^{83} either by Thr or by Ser (Fig. 3E).

Table 1. Virus composition in passages p0 through p3 of controlled forced evolution experiment 1. [1]Percental occurrence of BCoV $S1^A$ mutations and [2]their relative binding affinities as measured by equilibrium endpoint solid phase binding assay with S1A-Fc fusion proteins with that of parental BCoV $S1^A$-Fc ('wildtype') set at 1.0. $S1^A$ variants $Thr^{22}Ile$ and $Asn^{27}Tyr$ emerged only in experiment 2 (see sFig. 5), but their affinities relative to that of wildtype $S1^A$ are shown for comparison. (NA, not applicable).

	Frequency in population[1]				
$S1^A$	**p0**	**p1**	**p2**	**p3**	***r*Aff[2]**
wildtype	100	40.90	13.90	3.00	1.0
$Ile^{26}Ser$	0	45.83	0.70	0.03	0.0000625
$Asn^{27}Ser$	0	0.94	39.80	21.09	0.004
$Val^{29}Gly$	0	0	4.10	0.65	0.008
$Tyr^{75}Cys$	0	0	1.60	0.02	ND
$Thr^{83}Ile$	0	2.90	1.60	0.21	0.002
$Arg^{88}Thr$	0	0	2.30	9.69	0.25
$His^{173}Tyr$	0	5.77	5.00	8.10	0.03
$Pro^{174}Leu$	0	0	13.00	46.72	0.5
$Arg^{197}Cys$	0	2.24	17.25	10,.27	0.25
$Thr^{22}Ile$	NA	NA	NA	NA	0.015
$Asn^{27}Tyr$	NA	NA	NA	NA	0.002

For one of the five clonal populations obtained by endpoint dilution, we unfortunately failed to determine its genotype for technical reasons. From the NGS analysis of the p1 population, we deduced that the starting mutant must have been a low affinity S1A-$Leu^{89}Pro$ variant that, like the S1A-Thr^{83}Asn/Ile variants described above, quickly lost the HE-Asn^{114} glycan through an HE-$Ser^{116}Phe$ substitution. Oddly enough, the $Leu^{89}Pro$ substitution had not been detected by NGS in the pre-cloning virus stock. Note, however, that this mutation had been selected before twice independently in trials with rBCoV-HE-$F^{211}A$ (Table 2). Possibly, it arose spontaneously during end point dilution procedure. Be that as it may, its in vitro evolution proved informative (Fig. 4A). NGS analysis of a passage p1 population, resulting from 120-hr multistep propagation, showed that 100% of the viruses coded for HE-Thr^{114}Asn/$Ser^{116}Phe$ in combination with S1A-Pro^{89} (53.3%), - Thr^{89} (40.5%), or -Ser^{89} (2%). Note that the relationship between these variants and the course of evolution –from Leu^{89} in the parental recombinant virus to Pro and from Pro to Thr or Ser– is evident from the codon sequences (CTA→CCA→T/ACA) and that the Thr^{89} and Ser^{89} substitutions restored S RBS affinity almost to that of wildtype RBS (Fig. 4B). All three variants - S1A-Pro^{89}, -Thr^{89} and -Ser^{89}, were

readily cloned and isolated by standard endpoint dilution, and propagated independently without a requirement for exogenous RDE. p1 also contained a minor population of viruses with parental $S1^A$, presumably regenerated from S1A-Pro^{89}, which as for the S1A-Thr^{89} and Ser^{89} variants would have required only a single nucleotide substitution (CCA→CTA). Apparently, with HE lectin function partially restored, viruses that regained (near) wildtype S affinity had a selective advantage. At the end of passage p2, S1A-Pro^{89} variants had dwindled to less than 1.5%, S1A -Thr^{89} had become dominant at 75% and viruses with parental S1A-Leu^{89} had rapidly risen from 2.25% in p1 to 23% (Fig. 4A).

Table 2. Summary of $S1^A$ mutations identified in rBCoVs upon i. introduction of a HE-Phe^{211}Ala substitution and virus rescue by straight forward targeted recombination (E1 through E4, see also Fig. 1), ii. passage of rBCoV-HE-F^{211}A/Swt in the absence of exogenous HE (E5 and E6) or iii. introduction of a HE-Asn^{114}Thr substitution and subsequent viral passage (E7).

$S1^A$ residue	E1	E2	E3	E4	E5	E6	E7
Thr-22						Ile	
Ile-26	Ser				Ser		
Asn-27				Ser	Ser	Tyr	
Val-29					Gly	Gly	
Tyr-75					Cys		
Gly-82	Glu						
Thr-83		Ile		Ile	Ile		Ile/Asn
Arg-88					Thr		
Leu-89	Pro		Pro				Pro
His-173					Tyr	Tyr	
Pro-174					Leu	Leu	
Arg-197					Cys		

In summary, the introduction of a glycosylation site in the HE lectin domain that reduced receptor-binding affinity and, thereby, reduced sialate-*O*-acetylesterase activity towards clustered glycotopes, triggered a series of successive mutations in S and HE. Thus, the data directly demonstrate HE-S co-evolution. Moreover, the findings suggest that virions are under selective pressure not only to balance receptor-binding and receptor-destroying activities in apparent relation to cell-surface receptor-densities, but also, within these constrains, to maximize virion avidity.

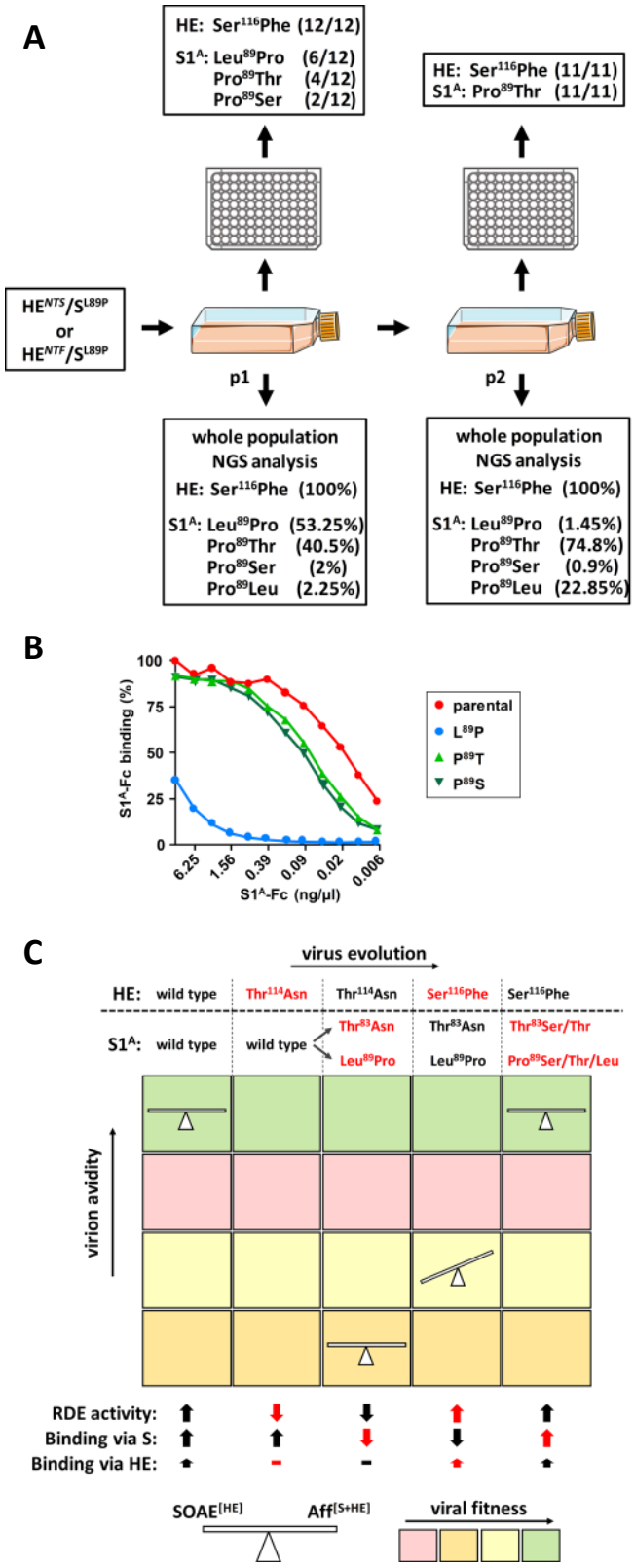


Fig. 4. Serial passage of rBCoV-HE-T^{114}N selects for successive mutations in HE and S to optimize HE/S functional balance and overall virion avidity. (A) Continued serial passage of rBCoV-HE-T^{114}N+S^{116}F/S1A-L^{89}P. Schematic outline of the experiment and presentation of results of NGS analysis and genetic characterization of subcloned variants as in Figs. 1, 2 and 3. (B) Re-substitution of S1A-Leu^{89}Pro by Thr or Ser restores S1^{A} binding to near wildtype levels. Sp-LBA as in Fig. 1D. (C) BCoV is under selective pressure for an optimal functional

balance between virion attachment and catalysis-driven release as well as for optimal virion avidity. Schematic summary of the evidence for HE-S coevolution and our interpretation thereof in a two-dimensional chart. The course of evolution of rBCoV-HE-Thr^{114}Asn (direction indicated by arrow) is shown for two types of low affinity S escape variants (Thr^{114}Asn and Leu^{89}Pro) with the succession of mutations in HE and S selected for during serial passage (top) brought in relation to (i) overall virion avidity, as mediated by S and HE, on the Y-axis from low to high as indicated by the arrow, (ii) viral fitness, color-coded from low (pink) to high (green) as indicated in the color legend at the bottom, (iii) the effect of the mutations on the function of S (attachment) and HE (receptor-destroying enzyme (RDE) activity and attachment) as indicated by thick arrows (arrows pointing up, near wildtype activity; arrows pointing down, decreased function; – , total loss of function; the size difference between arrows for S and HE reflect the difference in their contribution to virion binding; to indicate the effect of newly emerging mutations corresponding arrows for function are colored red) and (iv) their effect on HE-S functional balance (as indicated by the position of the scale).

Cell culture adapted BCoV and OC43 strains differ in their set point of the S/HE balance. The impact of loss of function mutations in the BCoV HE lectin domain was unexpected, because this defect seems well tolerated by the prototype OC43 laboratory strain USA/1967. In HRT18 cells, it grows to titers comparable to those of BCoV reference strain Mebus. Next gen sequencing of OC43 stocks revealed heterogeneity, but no indications for the existence of low S affinity minority variants that would support replication of majority high S affinity viruses. Also, clonal virus populations obtained by end point dilution (10/10) all conformed to the S1^{A} master sequence. We offer that instead OC43-USA/1967 may have reached a viable S/HE balance compatible with efficient *in vitro* propagation through adaptations in S that reduced receptor-binding affinity and/or altered receptor fine-specificity. When measured by solid phase assay with bivalent S1^{A}-Fc fusion proteins, binding of the S protein of OC43 USA/1967 to bovine submaxillary mucin, containing both mono- and di-*O*-acetylated α2,6-sialoglycans, is 16 to 32-fold lower than that of BCoV-Mebus [32]. sp-LBA with BSM preparations, selectively depleted for either 9-*O*- or 7,9-di-*O*-Sias, showed that BCoV S, like BCoV HE [28], preferentially binds to 7,9-di-*O*-Ac-Sias (sFig. 6A). OC43 USA/1967 S1^{A} may not share this preference. Apparently due to its low affinity, detectable binding to BSM was lost upon depletion of either type of Sia (sFig. 6A). Moreover, even though BCoV-Mebus S preferably binds to 7,9-di-*O*-Ac-Sia, monovalent one-on-one binding of the BCoV S1^{A} domain to α2,6-linked 9-*O*-acetylated Sia is still 3-fold stronger than that of OC43-USA/1967 as measured by biolayer interferometry (sFig. 6B). On a cautionary note, the isolation and complex passage history of OC43-USA/1967 [43,44] entailed several passages in human tracheal organ culture, suckling mouse brain and many rounds of replication in cultured cells [9], which would have given the virus ample opportunity to adapt to the *in vitro* conditions. Thus, the binding characteristics of its spike may not

faithfully reflect those in circulating field variants. Indeed, OC43 variants in sputum samples, contrary to the OC43 USA/1967, replicate in airway epithelial cell cultures but not in tissue culture cells [45].

Dissussion

Co-evolution and functional interdependence of embecovirus S and HE proteins. Our findings demonstrate that in the prototypic β1CoV BCoV the envelope proteins S and HE are functionally entwined and co-evolve. We posit that the same holds for other members of the species Betacoronavirus-1, including its zoonotic descendant human coronavirus OC43 and related viruses of swine, rabbits, dogs and horses, as well as for other Embecovirus species, most prominently among which human coronavirus HKU1. The data lead us to conclude that the respective activities of S and HE in receptor-binding and catalysis-driven virion elution are balanced to ensure dynamic reversible virion attachment and, thereby, efficient virus propagation. In consequence, for the viruses listed above, the roles of S and HE during natural infection cannot be understood in isolation but must be considered in unison.

Using a reverse genetics-based forced evolution approach with BCoV as a model system, we showed that loss of HE lectin function causes an offset in S-HE balance, practically incompatible with virus propagation and spread. With the HE lectin domain as modulator of esterase activity, mutations that decrease or abolish HE RBS affinity reduce virion-associated sialate-*O*-acetylesterase activity towards clustered glycotopes on hypervalent glycoconjugates [11] such as are present in the mucus and glycocalyx in natural tissues and on the surface of cultured cells. The extent of the resultant defect is such that compensatory second-site mutations in S are selected for that dramatically reduce S RBS affinity, apparently to restore reversibility of binding as an escape ticket from inadvertent virion attachment to non-productive sites.

The single-amino acid mutations in receptor-binding domain $S1^A$ were limited to a finite number of positions, either within or proximal to the RBS to directly affect protein-ligand interactions, or more distal to reduce RBS affinity through long range effects or by disrupting local folding through aberrant disulfide-bonding. Whereas the parental recombinant viruses, defective in HE lectin function but with wildtype S RBS affinity, require an external source of receptor-destroying enzyme for propagation, their progeny escape mutants regained propagation-independence by lowering S affinity.

In expanding clonal populations of HE-defective rBCoV-HE-F^{211}A, propagated in the absence of exogenous receptor-destroying enzyme activity, viruses with reduced S affinity gained a selective

advantage initially. Upon prolonged passage, however, quasispecies developed in which loss of HE lectin function was compensated at the population level. Variants that combined the HE-Phe^{211}Ala mutation with (near) wildtype affinity S proteins increased to dominate the swarm at least numerically. Still, these high affinity S variants for their proliferation were strictly reliant on minority low affinity S variants. This relationship extends beyond cooperativity and group selection described for other systems [46–52] and amounts to a state of dependency. We propose that the virions of the low affinity minority variants provide aid by serving as a source of exogenous sialate-*O*-acetylesterase activity. They themselves evade decoy receptors through enhanced reversibility of virion attachment, but this phenomenon increases their motility -whether by sliding diffusion or binding-rebinding- causing them to deplete cell surface 9-*O*-Ac-Sias, decoy receptors and functional receptors alike. With increasing concentrations of low affinity virions in the culture supernatant, high affinity variants would profit progressively, whereas falling cell surface receptor densities would put the low affinity viruses increasingly at a disadvantage.

The forced evolution trials performed with rBCoV-HE-F^{211}A were restricted in course and outcome by design, because full reversion would require simultaneous mutation of two adjacent nucleotides. Moreover, the crucial role of the Phe^{211} in ligand binding (sFig. 1) obviates conservative substitutions [17]. Although rBCoV-HE A^{211}V variants did emerge in two separate experiments, this mutation only marginally increases HE RBS affinity and sialate-*O*-acetylesterase activity.

In contrast to the HE-Phe^{211}Ala mutation, the deleterious effect of N-glycosylation at HE-Asn^{114} can be reversed, completely or partially, through various single-nucleotide substitutions in codons 114 and 116 and would therefore more readily allow for compensatory mutations also in HE. Indeed, serial passage of the rBCoV-HE-Thr^{114}Asn resulted in a succession of mutations alternatingly in HE and S. The order of appearance of these mutations and their effect on protein function indicated that they were not fixed to merely restore the balance between attachment and catalysis-driven virion elution. The HE-Thr^{114}Asn substitution initially selected for second-site mutations that reduced S affinity (Thr^{83}Ile, Thr^{83}Asn and Leu^{89}Pro), but with propagation thus recovered, derivatives rapidly emerged with increased HE lectin and esterase activity through a Ser^{116}Phe mutation. Apparently, this created an HE-S disbalance that in turn favored the selection of viruses with revertant mutations in S that raised S RBS affinity again to wildtype (Thr^{83} → Ile → Thr; Leu^{89} → Pro → Leu) or near wildtype levels (Thr^{83} → Ile → Ser; Leu^{89} → Pro→ Thr/Ser). Conjointly, our findings indicate that through an initial sharp reduction in overall avidity, compensatory to loss of HE function, virus particles regained the capacity of eluding non-productive attachment to decoy receptors, but at a fitness penalty. The decrease in S RBS affinity would predictably lower the specific infectivity of virus particles through a decrease in productive host cell attachment. The rapid selection of the

HE-Ser116Phe mutation in a low-affinity S background can thus be understood to have increased virion avidity, albeit through HE and rather than through S. HE does have a dual function after all and in influenza viruses C and D as well as in murine coronavirus-1, it is a receptor-binding protein first and foremost [22,35,53–56]. Of note, the partial $Ser^{116}Phe$ reversion of HE consistently seen in multiple independent experiments, suggests that a return to (near) wildtype lectin and esterase activity along with a low affinity S would have tipped the scale too much towards catalytic virion release. We posit that in addition to an optimal balance between receptor-binding and receptor-destruction, the system strives towards optimal virion avidity (Fig. 4C).

Under natural circumstances, the set-point of the S/HE balance would be tailored to conditions met in the target tissues of the intact host. The spontaneous loss of HE lectin function in OC43 and HKU1 may thus be understood to have arisen through convergent evolution as an adaptation to the sialoglycan composition of the mucus in the human upper respiratory tract and that of the glycocalyx of the respiratory epithelia [11]. This change, which would predictably reduce virion-associated receptor-destruction and hence decrease virion elution/increase or prolong virion attachment, might have been selected for by low density occurrence of 9-*O*-Ac-sialoglycans in the human upper airways. In accordance, limited tissue array analyses with HE-based virolectins suggested that these sugars are not particularly prevalent in the human respiratory tract and by far not as ubiquitous as in the gut [28]. However, full understanding of how the S-HE balance was reset in OC43 and HKU1 upon their zoonotic introduction and why awaits further analysis of the binding properties and ligand fine-specificity of the S proteins of naturally occurring variants as well as more quantitative and comprehensive inter-host comparative analyses of airway sialoglycomes. As an added complication, virion particles encounter widely different circumstances while traversing the mucus layer, at the epithelial cell surface, during local cell-to-cell dissemination, and during transmission. It is an open question whether this selects for majority phenotypes that can cope individually and independently with each of these different conditions by striking an uneasy compromise with regard to HE/S balance and overall virion avidity, or whether there is loco-temporal selection for swarms of variants that collectively allow the virus population as a whole to overcome each hurdle.

Similarities between embeco- and influenza A viruses point to common principles of virion-sialoglycan receptor-usage. The embecovirus HE gene originated from a horizontal gene transfer event, presumably with an influenza C/D-like virus as donor [17,57]. Like the orthomyxovirus hemagglutinin-esterase fusion proteins, the newly acquired coronavirus HE protein provided the acceptor virus with an opportunity to reversibly bind to 9-*O*-Ac-sialoglycans [26]. This in turn would seem to have prompted a shift in the receptor-specificity of S through adaptations in $S1^A$ that created a 9-*O*-Ac-Sia binding site *de novo* so that virions could now attach to these receptor determinants

also via S. The embecoviruses thus adopted a strategy of receptor usage entailing a concerted and carefully fine-tuned activity of two envelope proteins that is unique among coronaviruses, but uncannily similar to that of influenza A viruses. In the latter, the hemagglutinin (HA), as a pendant of S, mediates binding to either α2,3- or α2,6-linked sialosides, while the neuraminidase, like HE, is a receptor-destroying enzyme with a substrate fine-specificity that closely matches HA ligand preference [58]. For influenza A virus, the existence and biological relevance of a functional balance between receptor-binding and receptor destruction is well recognized [59–62]. This balance is critical for receptor-associated virus motility through the mucus and at the cell surface [63–68]. Complete or partial loss of NA activity -whether invoked spontaneously, through reverse genetics or by viral propagation in the presence of NA inhibitors- selects for mutations around the HA receptor-binding pocket that reduce HA affinity [60,69,70]. Furthermore, as proposed here for HE, NA contributes to virion attachment and even compensates for loss of virion avidity in mutant viruses with reduced HA affinity [71]. Different from HE, NA may do so via its catalytic pocket which doubles as a Sia-binding site [72]. However, NA also possesses a second Sia binding site [73,74], which like the HE lectin domain, regulates NA activity and which, in further analogy, is conserved or lost in apparent correlation with host tropism [75–77]. Finally, among many other similarities to embecoviruses, influenza A variants with different set points in their HA-NA functional balance may cooperate to support their propagation in cultured cells [48]. Our observations establish that there are common principles of virion-sialoglycan interactions that prompted convergent evolution of β1CoVs and influenza A viruses. Although these two groups of viruses essentially differ in genome type and replication strategy, envelope proteins, and receptors, they seem to be subject to the same rules of engagement with respect to dynamic receptor-binding, the differences between them constituting variations on a theme. This implies that observations made for the one system are informative for the other. Perhaps more importantly, insight into the overriding principles of virus-glycan interactions may open avenues to common strategies for antiviral intervention.

Materials & Methods

Cells and viruses. Human rectal tumor (HRT) 18 (ATCC® CCL-244™) and mouse LR7 [41] cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS), penicillin (100 IU/ml) and streptomycin (100 µg/ml). BCoV strain Mebus and OC43 strain USA/1967, purchased from the American Type Culture Collection (ATCC), were propagated in HRT18 cells.

Reverse genetics through targeted recombination. A reverse genetics system based on targeted RNA recombination was developed for BCoV strain Mebus essentially as described [15,41,78]. Using

conventional cloning methods, RT-PCR amplicons of the 5'-terminal 601 nts and 3'-terminal 9292 nts of the BCoV strain Mebus genome (reference Genbank sequence U00735.2) were fused and cloned in plasmid pUC57, downstream of a T7 RNA polymerase promotor and upstream of a 25-nt poly(A) tract and a *Pac*I site, yielding pD-BCoV1. From this construct, BCoV ORF 4a was deleted (nts 27740-27853) and replaced by the *Renilla* luciferase (Rluc) gen, yielding pD-BCoV-Rluc. A second pD-BCoV1 derivative, pD-mBCoVΔHE, was created by replacing the coding sequence for the ectodomain of BCoV S (nts 23641-27433) by the corresponding MHV-A59 sequence and by deleting the BCoV HE gene (nts 22406-23623). The nucleotide sequences of pD-BCoV1, pD-BCoV-Rluc and pD-mBCoVΔHE, determined by bidirectional Sanger sequence analysis, were deposited in Genbank (accession codes: XXX).

To generate a recombinant chimeric acceptor virus, mBCoVΔHE, HRT18 cells were infected with BCoV-Mebus at a MOI of 10 TCID50/cell and trypsinized and re-suspended in PBS. An aliquot of this suspension, containing 1.5×10^6 cells in 0.8 ml, was mixed with capped synthetic RNA that had been produced by *in vitro* transcription using the mMESSAGE mMACHINE™ T7 Transcription Kit (Thermo fisher) with *Pac*I-linearized mBCoVΔHE vector as template. The mixture was subjected to two consecutive electrical pulses of 850 V at 20 µF with a Gene Pulser II electroporator (Bio-Rad) and the cells were then seeded on a confluent monolayer of LR7 feeder cells in a 35mm dish. Incubation was continued at 37°C, 5% CO_2 for 18 hours post transfection until wide-spread cytopathic effect (CPE) was apparent. The cell culture supernatant was harvested and cleared by low speed centrifugation at 1200 rpm, and mBCoVΔHE was purified by end-point dilution and used to generate stocks for future usage in LR7 cells.

To generate luciferase-expressing rBCoVs with the BCoV HE and S genes reconstituted, i.e. $rBCoV^{wt}$ or rBCoV-HE-Phe^{211}Ala, LR7 cells, infected with mBCoVΔHE at MOI 5, were electroporated as described above with synthetic RNA transcribed from pD-BCoV-Rluc and derivatives thereof. The infected and transfected cells were then seeded on HRT18 cell monolayers in 35-mm plates for up to 160 hr. For rescue and propagation of rBCoV-HE-Phe^{211}Ala without second-site mutations in S, the cell culture supernatants were supplemented with 100 ng/ml of BCoV HE-Fc protein [17]. After 5-7 days of incubation at 37°C, samples of the cell culture supernatants were tested for infectivity by transferring them to HRT18 cell monolayers grown on 12-mm glass coverslips in 15.6-mm wells. Incubation was continued for 12 hr after which the cells were fixed with paraformaldehyde and immunofluorescence staining was performed with polyclonal antiserum from a BCoV-infected cow.

Virus titration, purification and characterization of viral populations. mBCoV was titrated and cloned by endpoint dilution on LR7 cells with cytopathic effect as read-out. rBCoVs were titrated and

cloned in HRT18 cells. To identify infected wells, cell supernatants were analyzed by hemagglutination assay with rat erythrocytes [26] and by *Renilla* luciferase assay (Dual-Luciferase® Reporter Assay System, Promega). Titers were calculated by the Spearman-Kaerber formula. Clonal virus populations were characterized by isolating viral RNA from 150 µl aliquots of the cell culture supernatant with the NucleoSpin® RNA Virus kit (MACHEREY-NAGEL) followed by conventional RT-PCR and bidirectional Sanger sequence analysis.

Controlled forced evolution experiments. Confluent HRT18 monolayers (5×10^6 cells) grown in 25-cm^2 flasks, were inoculated with rBCoV-HE-F^{211}A at an MOI 0.005 in PBS for 1 hr at 37°C. The cells were washed three times with PBS to remove residual exogeneous HE-Fc and incubation was continued in DMEM + 10% FCS at 37°C, 5% CO_2 for 120 hr post infection (pi) with samples collected every 24 hours (passage 1). Subsequent 120-hr passages were performed by adding 10 µL of supernatant to new cultures of HRT18 cells in 25-cm^2 flasks.

Expression and purification of HE-Fc and $S1^A$-Fc proteins. BCoV HE, either enzymatically-active (HE^+) or rendered inactive through a Ser^{40}Ala substitution (HE^0), and OC43 $S1^A$ were expressed as Fc fusion proteins in HEK293T cells and purified from the cell supernatant by protein A affinity chromatography as detailed [17,32]. Monomeric $S1^A$ was obtained by on-the bead thrombin cleavage [32]. pCD5-BCoVHE-T-Fc vectors [17] encoding mutant BCoV HE derivatives were constructed with the Q5® Site-Directed Mutagenesis Kit per the instructions of the manufacturer.

Pseudovirus entry assays.

The production of BCoV S-pseudotyped VSV-ΔG particles, their characterization by Western blot analysis, and infectivity assays in HRT18 cells were as described [32].

Solid-phase lectin binding assay (sp-LBA). sp-LBA was performed as described [32] with bovine submaxillary mucin (BSM; Sigma-Aldrich), coated to 96-Well Maxisorp® microtitre ELISA plates (Nunc, 0.1 µg BSM per well), serving as a ligand. Binding assays were performed with 2-fold serial dilutions of HE^0-Fc, $S1^A$-Fc, or mutated derivatives thereof. Receptor-destroying esterase activities of soluble HEs were measured by on-the-plate 9-*O*-Ac-Sia depletion assays as described [11,28].

Hemagglutination assay (HAA). HAA was performed with rat erythrocytes (*Rattus norvegicus* strain Wistar; 50% suspension in PBS). Standard HAA was done with two-fold serial dilutions of HE^0-Fc proteins (starting at 25 ng/well) as described [17]. High sensitivity nanoparticle HAA (NP-HAA) was performed as in [32,79]. Briefly, self-assembling 60-meric nanoparticles, comprised of lumazine synthase (LS), N-terminally extended with the immunoglobulin Fc-binding domain of the *S. aureus*

protein A, were complexed with HE^0-Fc proteins at a 1:0.6 molar ratio for 30 min on ice. The HE^0-Fc-loaded nanoparticles were then 2-fold serially diluted and mixed 1:1 (vol/vol) with rat erythrocytes (0.5% in PBS). Incubation was for 2 hr at 4°C after which HAA titers were read.

NGS analysis. Viral RNA from culture supernatants was isolated as described above. HE and $S1^A$ coding regions from viral genome of different virus populations were obtained by RT-PCR with primer sets HE_F 5'-TTAGATTATGGTCTAAGCATCATG-3' and HE_R 5'-TTAGATTATGGTCTAAGCATCATG-3', $S1^A_F$ 5'-ACCATGTTTTTGATACTTTTA-3' and $S1^A_R$ 5'-AGATTGTGTTTTACACTTAATCTC-3', respectively. Amplicons were processed in the NGSgo® workflow for Illumina according to the Instructions for Use (Edition 4), except that the fragmentation was prolonged to 40 min at 25°C (protocol 3A). Briefly, amplicons were subjected to fragmentation and adapter ligation using NGSgo-LibrX (GenDx). Size selection and clean-up of the samples was performed with SPRI beads (Machery-Nagel). Unique barcodes were ligated to each sample using NGSgo-IndX (GenDx), after which all samples were pooled and subsequently purified with SPRI beads, resulting in a library of fragments between ~400 and 1000 bp. The DNA fragments were denatured and paired-end sequenced on a MiSeq platform (Illumina) using a 300 cycle kit (V2). FASTQ files were analyzed in NGSengine® (GenDx), which aligned the reads to the reference sequences of HE and S (reference Genbank sequence U00735.2 for BCoV strain Mebus, and NC_006213.1 for OC43 strain USA/1967). For the characterization of each virus sample, amplicons from five independent RT-PCR reactions were analyzed in parallel and mutation frequencies were determined by averaging the results from these five replicates.

Author contributions

Y.L., W.L. and R.J.d.G. conceived the study and designed research; Y.L., W.L., D.K. and A.C.S.v.B. performed research; E.R. and H.M.v.S. performed NGS analysis; Y.L., W.L., Z.L., D.K., A.C.S.v.B., E.R., G.J.P.H.B., F.J.M.v.K., B.J.B., E.G.H., H.M.v.S. and R.J.d.G. analyzed data; W.L., Z.L. and G.J.P.H.B. provided and synthesized reagents; Y.L. and R.J.d.G. wrote the paper. All authors discussed the results and W.L., Z.L., G.J.P.H.B., F.J.M.v.K., B.J.B., E.G.H. and H.M.v.S. commented on the manuscript.

References

1. De Wit, E.; Van Doremalen, N.; Falzarano, D.; Munster, V.J. SARS and MERS: Recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* 2016, *14*, 523–534.
2. Reusken, C.B.E.M.; Raj, V.S.; Koopmans, M.P.; Haagmans, B.L. Cross host transmission in the emergence of MERS coronavirus. *Curr. Opin. Virol.* 2016, *16*, 55–62.
3. Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**.
4. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**.
5. Woo, P.C.Y.; Lau, S.K.P.; Yip, C.C.Y.; Huang, Y.; Yuen, K.Y. More and more coronaviruses: Human coronavirus HKU1. *Viruses* 2009, *1*, 57–71.
6. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A.C.K.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G.F. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* 2016, *24*, 490–502.
7. de Groot, R.; Baker, S.; Baric, R.; … L.E.-V.; 2012, U. Family coronaviridae. *Elsevier Amsterdam* **2011**, 806–828.
8. Vijgen, L.; Keyaerts, E.; Lemey, P.; Maes, P.; Van Reeth, K.; Nauwynck, H.; Pensaert, M.; Van Ranst, M. Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *J. Virol.* **2006**, *80*, 7270–7274.
9. Vijgen, L.; Keyaerts, E.; Moes, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604.
10. Lau, S.K.P.; Lee, P.; Tsang, A.K.L.; Yip, C.C.Y.; Tse, H.; Lee, R.A.; So, L.-Y.; Lau, Y.-L.; Chan, K.-H.; Woo, P.C.Y.; et al. Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. *J. Virol.* **2011**, *85*, 11325–11337.
11. Bakkers, M.J.G.; Lang, Y.; Feitsma, L.J.; Hulswit, R.J.G.; de Poot, S.A.H.; van Vliet, A.L.W.; Margine, I.; de Groot-Mijnes, J.D.F.; van Kuppeveld, F.J.M.; Langereis, M.A.; et al. Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. *Cell Host Microbe* **2017**, *21*, 356–366.
12. Hulswit, R.J.G.; de Haan, C.A.M.; Bosch, B.J. Coronavirus Spike Protein and Tropism Changes. In *Advances in Virus Research*; 2016; Vol. 96, pp. 29–57.
13. King, B.; Brian, D.A. Bovine Coronavirus Structural Proteins. *J. Virol.* **1982**, *42*, 700–707.
14. Kienzle, T.E.; Abraham, S.; Hogue, B.G.; Brian, D.A. Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. *J. Virol.* **1990**, *64*, 1834–8.
15. Lissenberg, A.; Vrolijk, M.M.; van Vliet, A.L.W.; Langereis, M.A.; de Groot-Mijnes, J.D.F.; Rottier, P.J.M.; de Groot, R.J. Luxury at a Cost? Recombinant Mouse Hepatitis Viruses Expressing the Accessory Hemagglutinin Esterase Protein Display Reduced Fitness In Vitro. *J. Virol.* **2005**, *79*, 15054–15063.
16. de Groot, R.J. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* 2006, *23*, 59–72.
17. Zeng, Q.; Langereis, M.A.; van Vliet, A.L.W.; Huizinga, E.G.; de Groot, R.J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9065–9069.
18. Schultze, B.; Wahn, K.; Klenk, H.D.; Herrler, G. Isolated HE-protein from hemagglutinating encephalomyelitis virus and bovine coronavirus has receptor-destroying and receptor-binding activity. *Virology* **1991**, *180*, 221–8.
19. Vlasak, R.; Luytjes, W.; Leider, J.; Spaan, W.; Palese, P. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J. Virol.* **1988**, *62*, 4686–90.
20. King, B.; Potts, B.J.; Brian, D.A. Bovine coronavirus hemagglutinin protein. *Virus Res.* **1985**, *2*, 53–59.
21. Bakkers, M.J.G.; Zeng, Q.; Feitsma, L.J.; Hulswit, R.J.G.; Li, Z.; Westerbeke, A.; van Kuppeveld, F.J.M.; Boons, G.-J.; Langereis, M.A.; Huizinga, E.G.; et al. Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proc. Natl. Acad. Sci.* **2016**, *113*, E3111–E3119.
22. Langereis, M.A.; Zeng, Q.; Heesters, B.; Huizinga, E.G.; de Groot, R.J. The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. *PLoS Pathog.* **2012**, *8*.
23. Smits, S.L.; Gerwig, G.J.; van Vliet, A.L.W.; Lissenberg, A.; Briza, P.; Kamerling, J.P.; Vlasak, R.; de Groot, R.J.

Nidovirus sialate-O-acetylesterases: evolution and substrate specificity of coronaviral and toroviral receptor-destroying enzymes. *J. Biol. Chem.* **2005**, *280*, 6933–41.

24. Regl, G.; Kaser, A.; Iwersen, M.; Schmid, H.; Kohla, G.; Strobl, B.; Vilas, U.; Schauer, R.; Vlasak, R. The hemagglutinin-esterase of mouse hepatitis virus strain S is a sialate-4-O-acetylesterase. *J. Virol.* **1999**, *73*, 4721–7.

25. Wurzer, W.J.; Obojes, K.; Vlasak, R. The sialate-4-O-acetylesterases of coronaviruses related to mouse hepatitis virus: a proposal to reorganize group 2 Coronaviridae. *J. Gen. Virol.* **2002**, *83*, 395–402.

26. Langereis, M.A.; van Vliet, A.L.W.; Boot, W.; de Groot, R.J. Attachment of Mouse Hepatitis Virus to O-Acetylated Sialic Acid Is Mediated by Hemagglutinin-Esterase and Not by the Spike Protein. *J. Virol.* **2010**, *84*, 8970–8974.

27. Williams, R.K.; Jiang, G.S.; Holmes, K. V Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl. Acad. Sci.* **2006**, *88*, 5533–5536.

28. Langereis, M.A.; Bakkers, M.J.G.; Deng, L.; Padler-Karavani, V.; Vervoort, S.J.; Hulswit, R.J.G.; van Vliet, A.L.W.; Gerwig, G.J.; de Poot, S.A.H.; Boot, W.; et al. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* **2015**, *11*, 1966–1978.

29. Schultze, B.; Gross, H.J.; Brossmer, R.; Herrler, G. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J. Virol.* **1991**, *65*, 6232–7.

30. Huang, X.; Dong, W.; Milewska, A.; Golda, A.; Qi, Y.; Zhu, Q.K.; Marasco, W.A.; Baric, R.S.; Sims, A.C.; Pyrc, K.; et al. Human Coronavirus HKU1 Spike Protein Uses O -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J. Virol.* **2015**, *89*, 7202–7213.

31. Künkel, F.; Herrler, G. Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **1993**, *195*, 195–202.

32. Hulswit, R.J.G.; Lang, Y.; Bakkers, M.J.G.; Li, W.; Li, Z.; Schouten, A.; Ophorst, B.; van Kuppeveld, F.J.M.; Boons, G.-J.; Bosch, B.-J.; et al. Human coronaviruses OC43 and HKU1 bind to 9- O -acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci.* **2019**, *116*, 2681–2690.

33. Langereis, M.A.; Zeng, Q.; Gerwig, G.J.; Frey, B.; Itzstein, M. von; Kamerling, J.P.; de Groot, R.J.; Huizinga, E.G. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15897–15902.

34. Rosenthal, P.B.; Zhang, X.; Formanowski, F.; Fitz, W.; Wong, C.H.; Meier-Ewert, H.; Skehel, J.J.; Wiley, D.C. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **1998**, *396*, 92–96.

35. Song, H.; Qi, J.; Khedri, Z.; Diaz, S.; Yu, H.; Chen, X.; Varki, A.; Shi, Y.; Gao, G.F. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLoS Pathog.* **2016**, *12*, e1005411.

36. Tortorici, M.A.; Walls, A.C.; Lang, Y.; Wang, C.; Li, Z.; Koerhuis, D.; Boons, G.J.; Bosch, B.J.; Rey, F.A.; de Groot, R.J.; et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **2019**, *26*, 481–489.

37. Desforges, M.; Desjardins, J.; Zhang, C.; Talbot, P.J. The Acetyl-Esterase Activity of the Hemagglutinin-Esterase Protein of Human Coronavirus OC43 Strongly Enhances the Production of Infectious Virus. *J. Virol.* **2013**, *87*, 3097–3107.

38. Deregt, D.; Babiuk, L.A. Monoclonal antibodies to bovine coronavirus: Characteristics and topographical mapping of neutralizing epitopes on the E2 and E3 glycoproteins. *Virology* **1987**, *161*, 410–420.

39. Deregt, D.; Gifford, G.A.; Ijaz, M.K.; Watts, T.C.; Gilchrist, J.E.; Haines, D.M.; Babiuk, L.A. Monoclonal Antibodies to Bovine Coronavirus Glycoproteins E2 and E3: Demonstration of in vivo Virus-neutralizing Activity. *J. Gen. Virol.* **1989**, *70*, 993–998.

40. Milane, G.; Kourtesis, A.B.; Dea, S. Characterization of monoclonal antibodies to the hemagglutinin-esterase glycoprotein of a bovine coronavirus associated with winter dysentery and cross-reactivity to field isolates. *J. Clin. Microbiol.* **1997**, *35*, 33–40.

41. Kuo, L.; Godeke, G.-J.; Raamsman, M.J.B.; Masters, P.S.; Rottier, P.J.M. Retargeting of Coronavirus by Substitution of the Spike Glycoprotein Ectodomain: Crossing the Host Cell Species Barrier. *J. Virol.* **2000**, *74*, 1393–1406.

42. Masters, P.S.; Rottier, P.J.M. Coronavirus Reverse Genetics by Targeted RNA Recombination. In *Coronavirus Replication and Reverse Genetics*; 2005; pp. 133–159.

43. McIntosh, K.; Becker, W.B.; Chanock, R.M. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci.* **1967**, *58*, 2268–2273.

44. McIntosh, K.; Dees, J.H.; Becker, W.B.; Kapikian, A.Z.; Chanock, R.M. Recovery in tracheal organ cultures of novel

viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci.* **1967**, *57*, 933–940.

45. Dijkman, R.; Jebbink, M.F.; Koekkoek, S.M.; Deijs, M.; Jonsdottir, H.R.; Molenkamp, R.; Ieven, M.; Goossens, H.; Thiel, V.; van der Hoek, L. Isolation and Characterization of Current Human Coronavirus Strains in Primary Human Epithelial Cell Cultures Reveal Differences in Target Cell Tropism. *J. Virol.* **2013**, *87*, 6081–6090.

46. Vignuzzi, M.; Stone, J.K.; Arnold, J.J.; Cameron, C.E.; Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **2006**, *439*, 344–348.

47. Bordería, A. V.; Isakov, O.; Moratorio, G.; Henningsson, R.; Agüera-González, S.; Organtini, L.; Gnädig, N.F.; Blanc, H.; Alcover, A.; Hafenstein, S.; et al. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS Pathog.* **2015**, *11*.

48. Xue, K.S.; Hooper, K.A.; Ollodart, A.R.; Dingens, A.S.; Bloom, J.D. Cooperation between distinct viral variants promotes growth of h3n2 influenza in cell culture. *Elife* **2016**, *5*.

49. Domingo-Calap, P.; Segredo-Otero, E.; Durán-Moreno, M.; Sanjuán, R. Social evolution of innate immunity evasion in a virus. *Nat. Microbiol.* **2019**, *4*, 1006–1013.

50. Ciota, A.T.; Ehrbar, D.J.; Van Slyke, G.A.; Willsey, G.G.; Kramer, L.D. Cooperative interactions in the West Nile virus mutant swarm. *BMC Evol. Biol.* **2012**, *12*.

51. Cao, L.; Wu, C.; Shi, H.; Gong, Z.; Zhang, E.; Wang, H.; Zhao, K.; Liu, S.; Li, S.; Gao, X.; et al. Coexistence of Hepatitis B Virus Quasispecies Enhances Viral Replication and the Ability To Induce Host Antibody and Cellular Immune Responses. *J. Virol.* **2014**, *88*, 8656–8666.

52. Shirogane, Y.; Watanabe, S.; Yanagi, Y. Cooperation between different RNA virus genomes produces a new phenotype. *Nat. Commun.* **2012**, *3*.

53. Herrler, G.; Rott, R.; Klenk, H.D.; Müller, H.P.; Shukla, A.K.; Schauer, R. The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. *EMBO J.* **1985**, *4*, 1503–6.

54. Rogers, G.N.; Herrler, G.; Paulson, J.C.; Klenk, H.D. Influenza C virus uses 9-O-acetyl-N-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells. *J. Biol. Chem.* **1986**, *261*, 5947–51.

55. Hause, B.M.; Ducatez, M.; Collin, E.A.; Ran, Z.; Liu, R.; Sheng, Z.; Armien, A.; Kaplan, B.; Chakravarty, S.; Hoppe, A.D.; et al. Isolation of a Novel Swine Influenza Virus from Oklahoma in 2011 Which Is Distantly Related to Human Influenza C Viruses. *PLoS Pathog.* **2013**, *9*.

56. Vlasak, R.; Krystal, M.; Nacht, M.; Palese, P. The influenza c virus glycoprotein (HE) exhibits receptor-binding (hemagglutinin) and receptor-destroying (esterase) activities. *Virology* **1987**, *160*, 419–425.

57. Luytjes, W.; Bredenbeek, P.J.; Noten, A.F.H.; Horzinek, M.C.; Spaan, W.J.M. Sequence of mouse hepatitis virus A59 mRNA 2: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **1988**, *166*, 415–422.

58. Long, J.S.; Mistry, B.; Haslam, S.M.; Barclay, W.S. Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* 2019, *17*, 67–81.

59. Xu, R.; Zhu, X.; McBride, R.; Nycholat, C.M.; Yu, W.; Paulson, J.C.; Wilson, I.A. Functional Balance of the Hemagglutinin and Neuraminidase Activities Accompanies the Emergence of the 2009 H1N1 Influenza Pandemic. *J. Virol.* **2012**, *86*, 9221–9232.

60. Mitnaul, L.J.; Matrosovich, M.N.; Castrucci, M.R.; Tuzikov, A.B.; Bovin, N. V; Kobasa, D.; Kawaoka, Y. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* **2000**, *74*, 6015–20.

61. Wagner, R.; Matrosovich, M.; Klenk, H.D. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* 2002, *12*, 159–166.

62. Baum, L.G.; Paulson, J.C. The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* **1991**, *180*, 10–15.

63. Cohen, M.; Zhang, X.Q.; Senaati, H.P.; Chen, H.W.; Varki, N.M.; Schooley, R.T.; Gagneux, P. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virol. J.* **2013**, *10*.

64. Guo, H.; Rabouw, H.; Slomp, A.; Dai, M.; van der Vegt, F.; van Lent, J.W.M.; McBride, R.; Paulson, J.C.; de Groot, R.J.; van Kuppeveld, F.J.M.; et al. Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. *PLoS Pathog.* **2018**, *14*.

65. Matrosovich, M.N.; Matrosovich, T.Y.; Gray, T.; Roberts, N.A.; Klenk, H.-D. Neuraminidase Is Important for the Initiation of Influenza Virus Infection in Human Airway Epithelium. *J. Virol.* **2004**, *78*, 12665–12667.

66. Sakai, T.; Nishimura, S.I.; Naito, T.; Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Sci. Rep.* **2017**, *7*.

67. Yang, Y.; Du, L.; Liu, C.; Wang, L.; Ma, C.; Tang, J.; Baric, R.S.; Jiang, S.; Li, F. 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.* **2014**, *111*, 12516–12521.

68. de Vries, E.; Du, W.; Guo, H.; de Haan, C.A.M. Influenza A Virus Hemagglutinin–Neuraminidase–Receptor Balance: Preserving Virus Motility. *Trends Microbiol.* 2019.

69. McKimm-Breschkin, J.L.; Blick, T.J.; Sahasrabudhe, A.; Tiong, T.; Marshall, D.; Hart, G.J.; Bethell, R.C.; Penn, C.R. Generation and characterization of variants of NWS/G70C influenza virus after in vitro passage in 4-amino-Neu5Ac2en and 4-guanidino-Neu5Ac2en. *Antimicrob. Agents Chemother.* **1996**, *40*, 40–6.

70. Hughes, M.T.; Matrosovich, M.; Rodgers, M.E.; McGregor, M.; Kawaoka, Y. Influenza A Viruses Lacking Sialidase Activity Can Undergo Multiple Cycles of Replication in Cell Culture, Eggs, or Mice. *J. Virol.* **2002**, *74*, 5206–5212.

71. Wen, F.; Wan, X.F. Influenza Neuraminidase: Underrated Role in Receptor Binding. *Trends Microbiol.* 2019, *27*, 477–479.

72. Hooper, K.A.; Bloom, J.D. A Mutant Influenza Virus That Uses an N1 Neuraminidase as the Receptor-Binding Protein. *J. Virol.* **2013**, *87*, 12531–12540.

73. Webster, R.G.; Air, G.M.; Metzger, D.W.; Colman, P.M.; Varghese, J.N.; Baker, A.T.; Laver, W.G. Antigenic structure and variation in an influenza virus N9 neuraminidase. *J Virol* **1987**, *61*, 2910–2916.

74. Varghese, J.N.; Colman, P.M.; van Donkelaar, A.; Blick, T.J.; Sahasrabudhe, A.; McKimm-Breschkin, J.L. Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl. Acad. Sci.* **1997**, *94*, 11808–11812.

75. Lai, J.C.C.; Garcia, J.M.; Dyason, J.C.; Böhm, R.; Madge, P.D.; Rose, F.J.; Nicholls, J.M.; Peiris, J.S.M.; Haselhorst, T.; Von-Itzstein, M. A secondary sialic acid binding site on influenza virus neuraminidase: Fact or fiction? *Angew. Chemie - Int. Ed.* **2012**, *51*, 2221–2224.

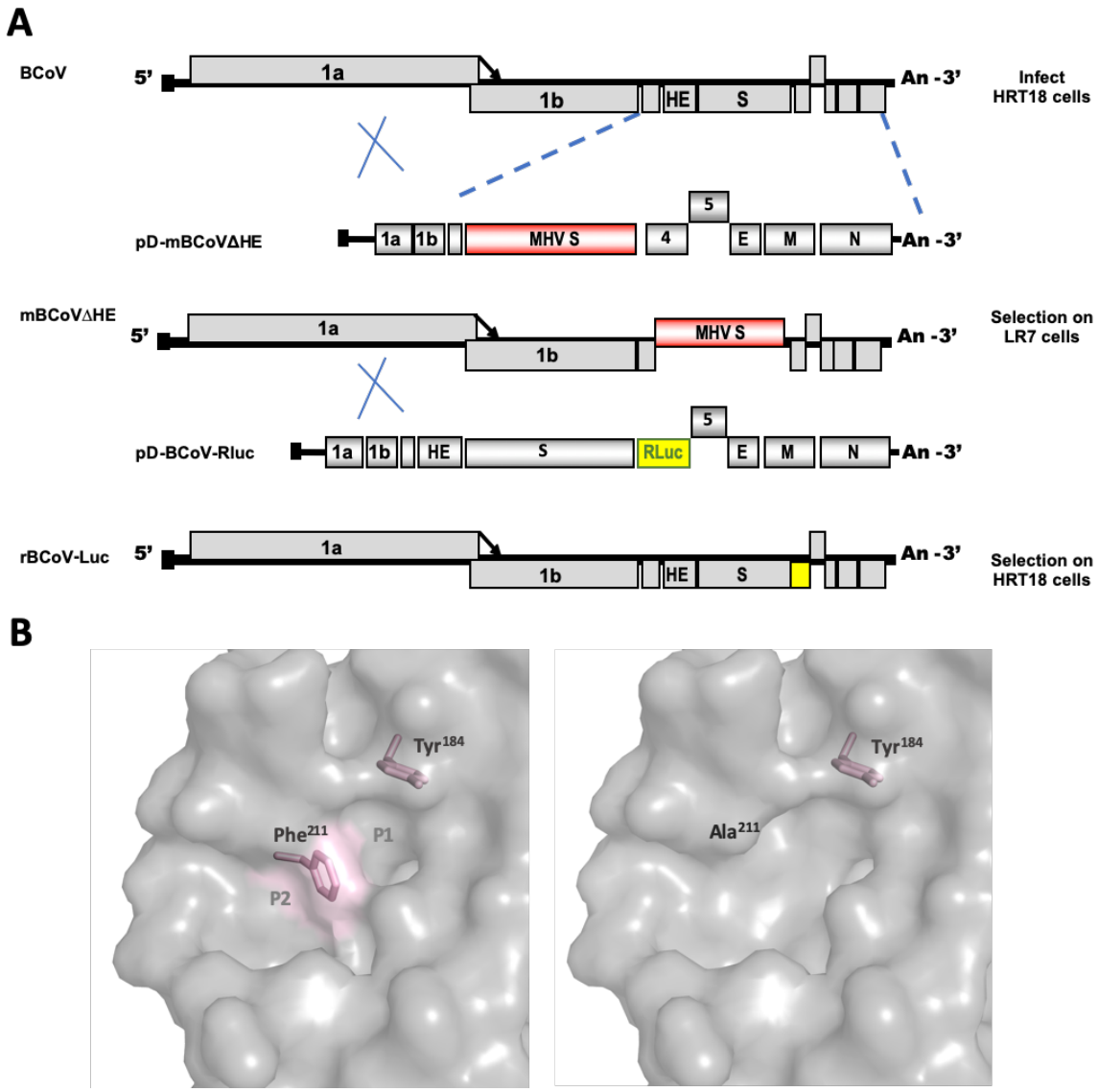
76. Dai, M.; McBride, R.; Dortmans, J.C.F.M.; Peng, W.; Bakkers, M.J.G.; de Groot, R.J.; van Kuppeveld, F.J.M.; Paulson, J.C.; de Vries, E.; de Haan, C.A.M. Mutation of the Second Sialic Acid-Binding Site, Resulting in Reduced Neuraminidase Activity, Preceded the Emergence of H7N9 Influenza A Virus. *J. Virol.* **2017**, *91*.

77. Du, W.; Guo, H.; Nijman, V.S.; Doedt, J.; van der Vries, E.; van der Lee, J.; Li, Z.; Boons, G.J.; van Kuppeveld, F.J.M.; de Vries, E.; et al. The 2nd sialic acid-binding site of influenza A virus neuraminidase is an important determinant of the hemagglutinin-neuraminidase-receptor balance. *PLoS Pathog.* **2019**, *15*, e1007860.

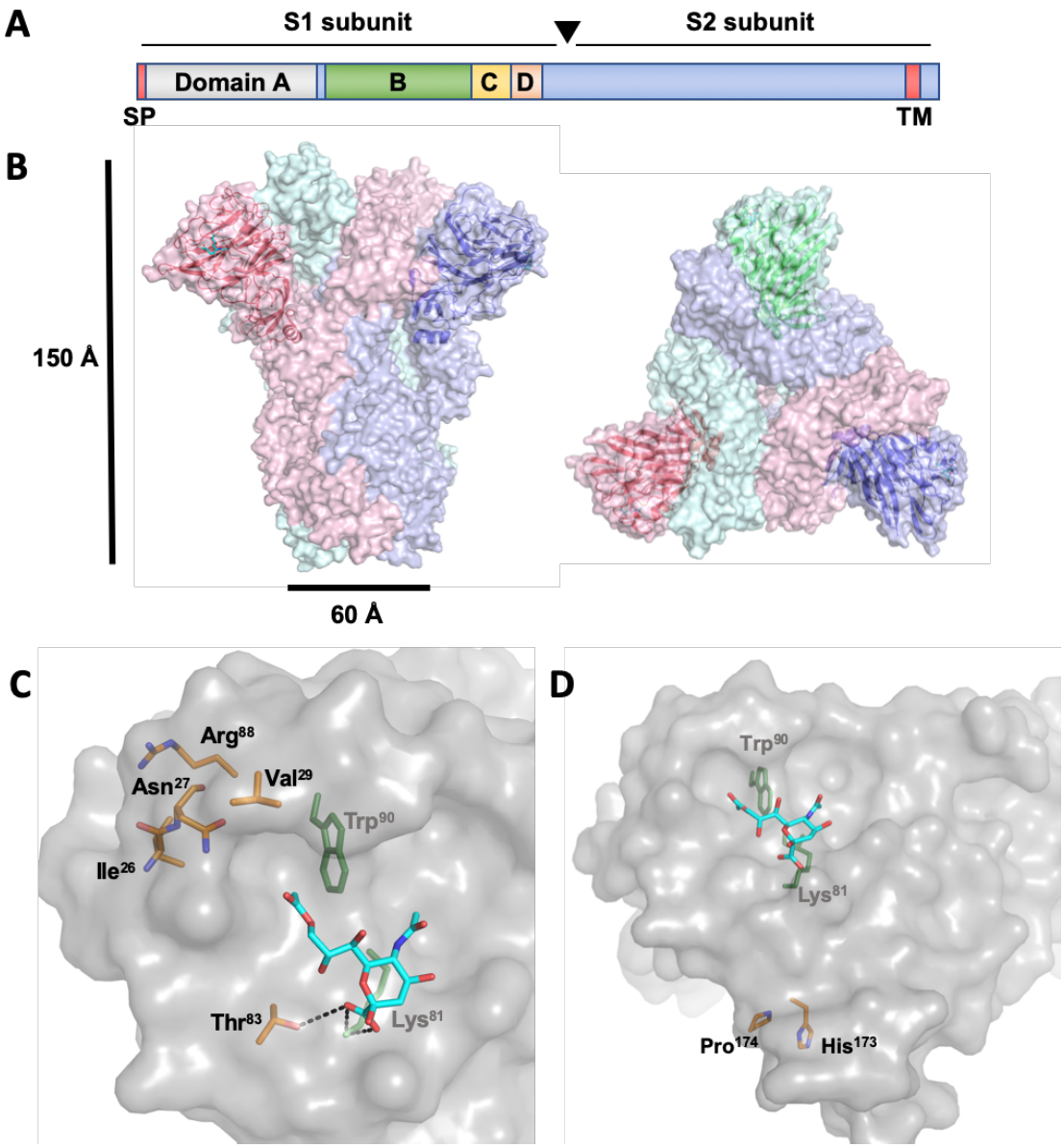
78. Haijema, B.J.; Volders, H.; Rottier, P.J.M. Switching Species Tropism: an Effective Way To Manipulate the Feline Coronavirus Genome. *J. Virol.* **2003**, *77*, 4528–4538.

79. Li, W.; Hulswit, R.J.G.; Widjaja, I.; Raj, V.S.; McBride, R.; Peng, W.; Widagdo, W.; Tortorici, M.A.; van Dieren, B.; Lang, Y.; et al. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci.* **2017**, *114*, E8508–E8517.

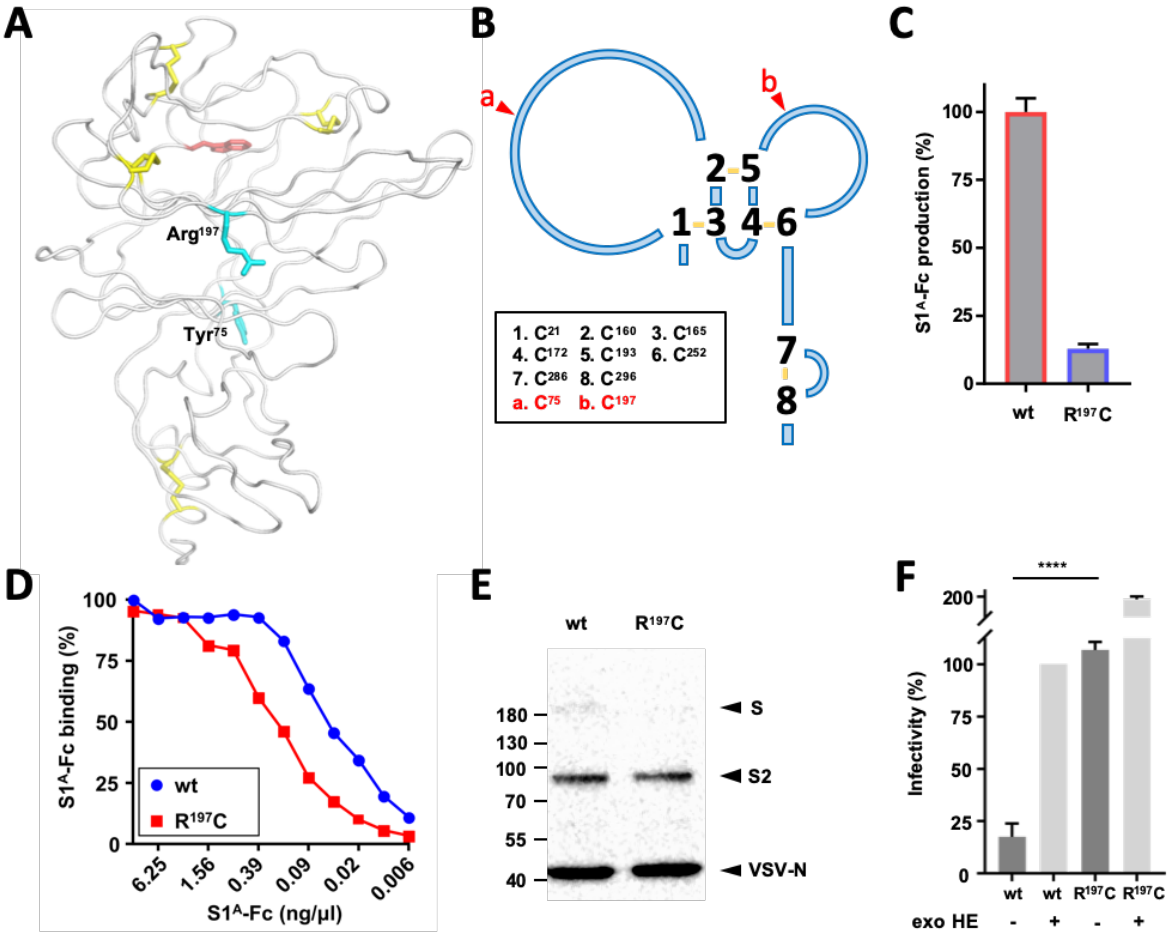
Supplemental Information



sFig. 1. Construction of recombinant bovine coronaviruses via targeted RNA recombination. (A) Schematic representation of the genomes of BCoV (strain Mebus), the acceptor virus mBCoVΔHE, and recombinant 'wildtype' rBCoV-Luc. Structures of the synthetic donor RNAs, transcribed from pD-mBCoVΔHE to generate the acceptor virus, and from pD-BCoV1 to generate the various rBCoV derivatives, are also shown. Genes are shown as boxes, with those for the polymerase polyproteins (POL1a and POL1b) and the HE, S, ORF4a, ORF5, E, M, and N proteins indicated. MHV-A59 spike ectodomain colored in red and the gene for Renilla luciferase colored in yellow. Cap structures indicated by black rectangles, ribosomal frameshifting sites by arrows, and poly(A) tails by 'An'. (B left) A close-up of the BCoV HE structure [1] (PDB code 3CL5) with the region containing the RBS in surface representation. Key elements of the RBS, pockets P1 and P2, are indicated. Also indicated are residues Phe211 (colored pink) and Tyr184 with their side chains in sticks (also in pink). (Right) $Phe^{211}Ala$ substitution alters the architecture of *O*-Ac-Sia binding site as modelled by Coot [2] and leads to a complete loss of HE lectin function.

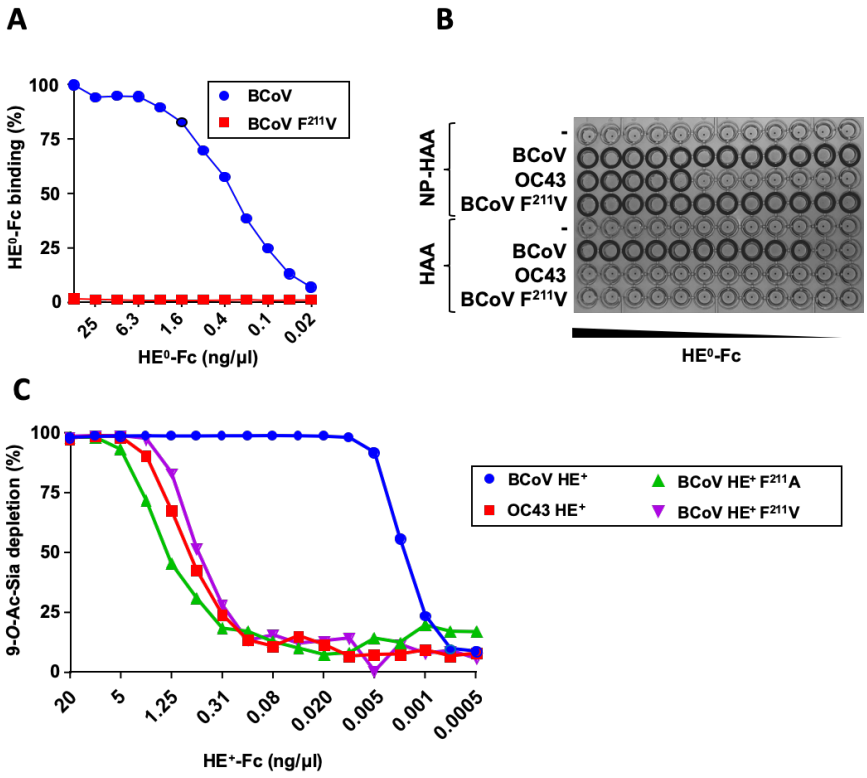


sFig. 2. Loss of HE lectin function selects for single site S1^{A} substitutions both proximal to and more distal from the RBS. (A) Linear presentation of β1CoV S with subunits S1 and S2, and S1 domains A through D indicated. SP, signal peptide; TM, transmembrane domain; arrowhead, cleavage site. (B) Surface representation of the HCoV-OC43 S ectodomain trimer [3] (PDB 6NZK) in side (left) and top view (right) with individual protomers in different colors, S1A domains superimposed in cartoon and the associated 9-*O*-Ac-Sia shown in sticks. (C) Close-up of the S1^{A}-9-*O*-Ac-Sia complex in surface representation as in Fig. 1. Side chains of Lys^{81}, Trp^{90} shown in sticks and colored green. Residues in the RBS, substituted upon loss of HE lectin function, are shown in sticks and colored by element (oxygen, red; nitrogen, blue; carbons, orange). Hydrogen bonds between Lys^{81} and Thr^{83} side chains and the Sia carboxylate are shown as black dashed lines. (D) Wide angle view of the S1^{A} domain with RBS and bound ligand indicated, and side chains of His^{173} and Pro^{174} shown in sticks, colored by element.

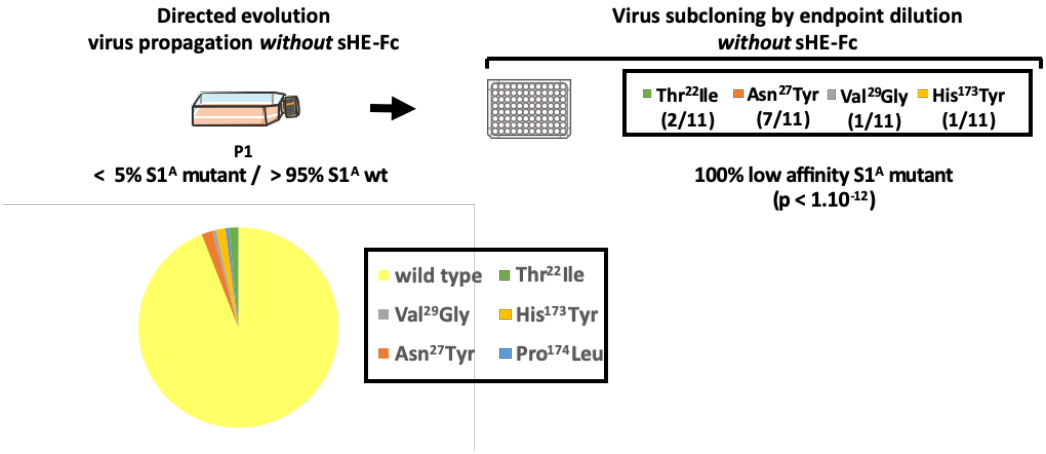


sFig. 3. Consequences of the introduction of unpaired cysteine residues in BCoV S1^{A}. (*A*) Cartoon representation of BCoV S1^{A} with Cys residues and disulfide bonds depicted in sticks and colored yellow; the location of the RBS, indicated by the Trp^{90} (colored red, side chain in sticks) and of the residues substituted for by Cys (Tyr^{75} and Arg^{197}) are also shown (colored blue, side chains in sticks). (*B*) Schematic representation of the disulfide-bonded structure of S1^{A}. Native Cys residues are in black and numbered according to the legend with disulfide bonds indicated in yellow. Newly introduced Cys are in red, their positions indicated by red arrowheads. (*C*) Arg^{197}Cys substitution strongly reduces secretion of S1^{A}-Fc fusion protein in HEK293T cells. Expression levels in the supernatant were determined as in [4]. HEK293T cells grown to 70% confluency in 35-mm wells were transfected with equal amounts of the expression vectors (1 mg plasmid DNA/well). Cell-free culture supernatants, harvested at seven days post transfection, were two-fold serially diluted and coated in Nunc Maxisorp 96-well ELISA plates (Fisher Scientific). S1^{A}-Fc fusion proteins were detected by ELISA with HRP-conjugated goat anti-human IgG antiserum (Southern Biotech). Protein concentrations (mean ± SD; µg/ml cell culture supernatant) were calculated using a standard curve, generated with purified OC43 S1^{A}-Fc, by 4 parameter logistic regression (4PL) method (GraphPad Prism version 7.04) from three independent transient transfection experiments. (*D*) Secretory S1^{A}-R^{197}C-Fc is of near-wildtype affinity. Parental type S1^{A}-Fc ('wt') and mutant S1^{A}-R^{197}C-Fc ('R^{197}C'), purified from supernatants from transiently-transfected cells, were two-fold serially diluted and compared for their binding to BSM by sp-LBA (starting at 2.5 µg virolectin/well.

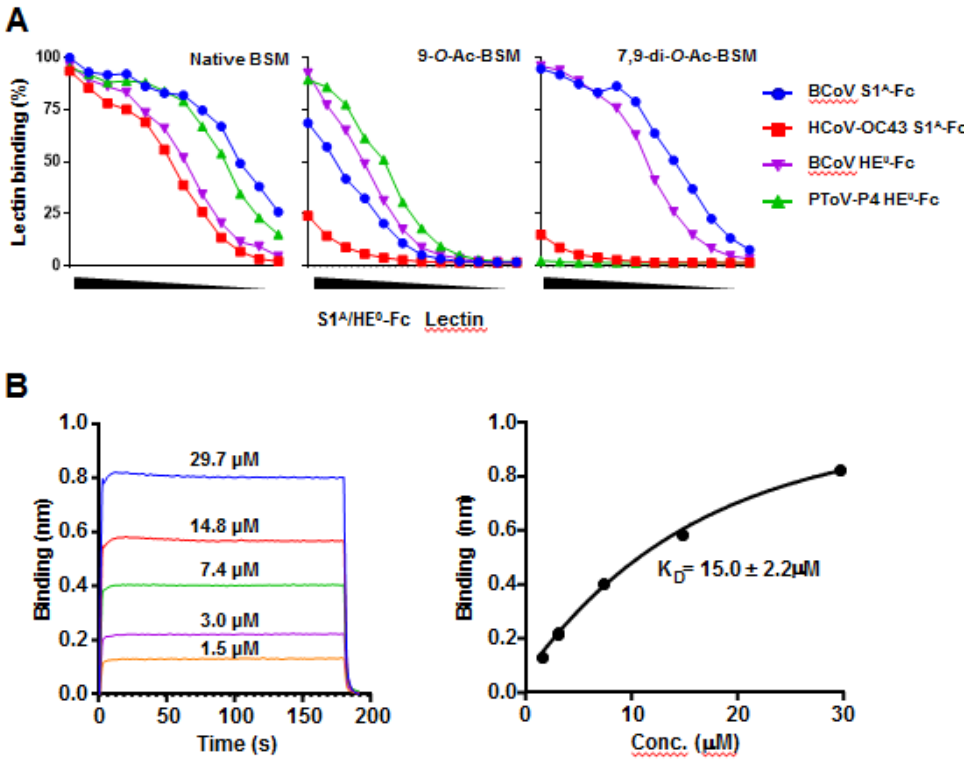
Binding expressed in percentages with maximum binding of wildtype $S1^{A}$-Fc set to 100%. (*E*) $S1^{A}$-$R^{197}C$ substitution does not noticeably affect biosynthesis, and proteolytic cleavage of full-length S as measured by its incorporation in G-deficient VSV particles. Pseudotyped VSV particles were produced and analyzed for relative content of N and S by western blot analysis as in [4]. Note that uptake of S in pseudotyped particles requires export of S homotrimers from the endoplasmic reticulum to the budding site at the plasma membrane. Apparently, $Arg^{197}Cys$ substitution affects exocytotic transport of full-length trimeric S to much lesser extent than that of the homodimeric $S1^{A}$-Fc fusion protein. (*F*) S-$Arg^{197}Cys$ substitution increases the infectivity of S-pseudotyped VSV particles and decreases their dependency on exogenous HE. VSV particles pseudotyped with BCoV S (wt) or BCoV-S-$R^{197}C$ were used to inoculate HRT18 cells, either with (+) or without (–) exogenous (exo) HE added to the inoculum. Infectivity measured and results presented as in Fig. 1E. The observations fit a model in which locally-misfolded S monomers are yet incorporated in homotrimers together with properly folded S monomers causing a decrease in average avidity of S trimers and, consequently, in overall virion-avidity.



sFig. 4. HE-$Ala^{211}Val$ substitution restores HE lectin and esterase function only marginally. (A) sp-LBA performed with HE-Phe^{211}Val-Fc conducted and presented as in Fig. 1D. (B) Receptor binding activity of the HE-Phe^{211}Val-Fc measured by conventional HAA (lower panel) and by high-sensitivity nanoparticle HAA (NP-HAA, higher panel). BCoV and OC43 HE-Fc were included as controls. (C) On-the-plate receptor depletion assay with bovine submaxillary mucin as substrate performed as in [5]. The receptor destroying activity of HE-Phe^{211}Val-Fc was compared to that of the HE+-Fc fusion proteins of BCoV and OC43 and of BCoV HE lectin-deficient mutant $F^{211}A$.



sFig. 5. Propagation of rBCoV-HE-F^{211}A in the absence of exogenous HE selects for mixed viral populations with minority low affinity S mutants promoting amplification of majority high affinity S variants. Low MOI inoculation of HRT18 cells with rBCoV-HE-F^{211}A and 120-hr multistep propagation in the absence of exogeneous HE, and characterization of the viral swarm by NGS and virus subcloning as in Fig. 2. Frequencies of S1^{A} variants are presented in in a pie chart as in Fig. 2. Although the virus population was comprised for 95% of variants that expressed wildtype S at a titer of 5 x 10^{8} TCID50 units/ml as measured in the presence of exogenous HE, virus subcloning in the absence of exogenous HE yielded low affinity S variants exclusively.



sFig. 6. S protein of OC43-USA/1967 differs from BCoV-Mebus S in receptor-binding affinity and ligand preference. *(A)* Binding of BCoV and OC43 $S1^A$-Fc fusion proteins to BSM was compared by sp-LBA as in Fig. 1. BSM is exceptionally rich in a variety of mono- and di-*O*-acetylated, 5-*N*-acetylated or -glycolylated Sias that can be selectively depleted by sialate-*O*-acetylesterase treatment with torovirus and coronavirus HE^+-Fc proteins [6]. To assess potential differences in ligand preference between BCoV and OC43 $S1^A$ binding was tested using BSM preparations, prepared and characterized previously [6], that were depleted for mono-*O*-acetylated Sia species with porcine torovirus P4 HE to contain 7,9-di-*O*-Ac-Sias exclusively (7,9-di-*O*-Ac-BSM) or for 7,9-di-*O*-Ac-Sias with bovine coronavirus HE to contain 9-mono-*O*-Ac-Sias only (9-*O*-Ac-BSM). Virolectins, specific 9-*O*-Ac, 5-*N*-Ac-Sia (porcine torovirus PToV P4 HE^0-Fc) or preferentially binding to 7,9-di-*O*-Ac-Sias (BCoV HE^0-Fc) were included as controls. Note that BCoV $S1^A$ binds preferentially to 7,9-di-*O*-Ac-Sia and that depletion of 9-mono-*O*-Ac-Sia by PToV P4 HE affects binding only modestly. Also note that binding of OC43 is strongly reduced by depletion of either type of *O*-Ac-Sia. (*B*) Binding affinities (K_D) of BCoV measured by biolayer interferometry (BLI) with α2,6-linked 9-mono-*O*-Ac-SiaLacNAc. To allow monovalent 1 on 1 binding, $S1^A$ domains were produced and purified by thrombin cleavage of $S1^A$-Fc proteins at an engineered cleavage site. BLI was performed and binding affinities determined as described[3]. Note that although BCoV $S1^A$ preferentially binds to 7,9-di-*O*-Ac-Sia, its K_D for α2,6-linked 9-mono-*O*-Ac-Sia such as are present in BSM [6] is still 3-fold lower than that of OC43 $S1^A$ [3].

Supplemental references

1. Zeng, Q., Langereis, M. A., van Vliet, A. L. W., Huizinga, E. G. & de Groot, R. J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **105**, 9065–9069 (2008).
2. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
3. Tortorici, M. A. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* **26**, 481–489 (2019).
4. Hulswit, R. J. G. *et al.* Human coronaviruses OC43 and HKU1 bind to 9- O -acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci.* 116, 2681–2690 (2019).
5. Bakkers, M. J. G. *et al.* Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. *Cell Host Microbe* 21, 356–366 (2017).
6. Langereis, M. A. *et al.* Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* 11, 1966–1978 (2015).

Chapter 6

Summary and General Discussion

Summary and General Discussion

The orthocoronaviruses (CoVs) comprises a highly diverse group of enveloped positive-strand RNA viruses [1]. Their propensity to cross species barriers allowed them to radiate and colonize a wide variety of avian and mammalian host species, humans included, in some cases not once but multiple times. For example, swine are host to at least four types of coronaviruses belonging to different genera and subgenera: porcine hemagglutinating encephalomyelitis virus (subgenus *Embecovirus*, genus *Betacoronavirus*), transmissible gastroenteritis virus, porcine epidemic diarrhea virus (subgenera *Tegacovirus* and *Pedacovirus*, respectively; genus *Alphacoronavirus*) and porcine coronavirus HKU15 (subgenus *Buldecovirus*, genus *Deltacoronavirus*). The latter virus species was discovered only in 2012 [1–3]. Humans can be infected by at least seven different coronaviruses, but at the time of writing only four of them have become established in the general population. Whether SARS-CoV-2, which emerged in humans in the fall of 2019 to give rise to a massive outbreak with Wuhan, China as epicenter [4,5], will succeed in colonizing the novel host, hangs in the balance. The outcome of the SARS-CoV-2 epidemic, extinction as in the case of the 2002 SARS-CoV-1 outbreak or pandemicity, will become clear in the next coming weeks and months. The four human coronaviruses that did become established so far are all associated with respiratory infections, but in contrast to SARS and MERS CoVs, they generally cause common colds. Two of them, 229E and NL63, belong to two different subgenera (*Duvina*- and *Setracoronavirus*, respectively) in the genus *Alphacoronavirus*. The other two, OC43 and HKU1, are different species, but both group in a minor clade, subgenus *Embecovirus,* in the genus *Betacoronavirus* [1]. The embecoviruses invaded the human population independently, HKU1 approximately 740 years ago as estimated by molecular clock analysis [6], OC43 far more recently, some 70-120 years ago [7–9]. OC43 belongs to a closely-knit group of viruses, grouped in a single species *Betacoronavirus 1* (β1CoV). These viruses have been particularly successful in cross-species transmission and in addition to humans invaded various host species, ruminants, canines, swine, lagomorphs and equines [2,10]. Interestingly, each introduction appears to have been a one-way street (Chapter 2 and references therein). The phylogenetic evidence suggests that in each case cross-species transmission required host adaptation amounting to host exclusivity sending the different viruses on a route towards speciation. Given that embecoviruses, and β1CoVs especially, seem to be even more promiscuous than other coronaviruses, the question arises what makes them special. In the studies described in this thesis, the focus was on OC43 and its presumptive ancestor bovine coronavirus (BCoV). The genomes of early strains of these viruses share 97% overall sequence identity, but it is clear from the phylogenetic record that the introduction of OC43 was a singular event and that both viruses have since gone their separate ways ([7,8], see also chapter 2). An explanation for β1CoV promiscuity may be their choice of receptor. All β1CoVs attach

to glycan-based receptor determinants with 9-*O*-acetylated sialic acid (9-*O*-Ac-Sia) as main component ([11–14], see also chapter 2). Sialate-9-*O*-acetylation is a common modification in vertebrates and 9-*O*-Ac-Sias are abundantly present in the respiratory and enteric tracts of mammals and birds, in the mucus layer protecting the epithelial cells and also, at the epithelial cell surface, in the glycocalyx [15–18]. Although receptor choice may explain why β1CoVs invaded so many different species, it begs the question as to how and why these viruses became host specific. Here we tested the hypothesis that host specificity may still be determined at the level of receptor-binding during pre-attachment, i.e. in the mucus layer en route to the target cells, and/or during attachment, at the cell surface. In other words, we asked whether OC43 and BCoV might differ in the way they 'deal' with their sialoglycan receptors and/or whether they might yet differ in receptor-fine specificity. After all, post-synthetically sialosides occur in a wide variety [19,20]. Receptor preference and consequences thereof may well be determined by Sia modifications additional to the critical Sia-9-*O*-acetyl moiety, like *O*-acetylation at Sia carbons 4, 7 and 8, the type of acylation (glycolylation or acetylation) at Sia carbon 5, and the glycosidic linkage to the penultimate sugar residue. To answer these questions, we took a multidisciplinary approach, combining classical and molecular virology, phylogenetic analysis, structural analysis and biochemistry, to study the properties of the two viral envelope glycoproteins involved in embecovirus receptor-binding, namely the hemagglutinin-esterase HE and the spike protein S.

In **Chapter 2,** we focused on HE and showed that, in contrast to the HEs of all animal corona- and toroviruses studied so far, those of human coronaviruses OC43 and HKU1 have lost their lectin function. By comparative sequence analysis of the HEs of BCoV strain Mebus and OC43 ATCC strain 1967/USA in combination with mutational analysis, we showed that loss of HE lectin function in early OC43 strains was not due to a single substitution but rather to a combination of four mutations proximal to the HE receptor-binding site. Importantly, we showed that even these four mutations did not destroy the HE RBS completely as residual binding to 9-*O*-Ac-Sia was observed, albeit detectable only with the most sensitive of assays, i.e. nanobead hemagglutination, based on multivalency-driven high-avidity binding. Apparently, even this low level HE-mediated attachment to 9-*O*-Ac-Sia was disadvantageous to propagation and/or human-to-human spread, because during continued OC43 evolution since 1967, additional mutations were selected that completely obliterated the HE receptor binding site. In two separate lineages of OC43 HEs, a metal-binding site, crucial to the organization of the RBS was targeted. In both HE types, the most critical residue, Asp^{220}, was replaced but saliently by different amino acids, either Tyr or His.

These observations in isolation could already be taken as evidence for convergent evolution of OC43 lineages and that loss of HE lectin function occurred as an adaptation to replication in human

airways. Strong additional support for this notion came from our observations made for HKU1 HE. Loss of HE lectin function in OC43, which occurred over a few decades, apparently was a replay of HKU1 evolution. HKU1 HE also lost its lectin function even to an extent that in the course of centuries of circulation in the human population, large parts of the lectin domain were deleted. We have made several attempts to also solve the crystal structure of HKU1 HE, but these unfortunately did not meet with success. Only recently, in a collaborative effort with Daniel Hurdiss (Utrecht University, Utrecht, The Netherlands), did we succeed in solving the HKU1 HE structure to a resolution of 3.5 Å by cryo-EM single particle analysis. Functional HE RBSs are formed by five variable loops, four grafted on the conserved core of the lectin domain -an eight-stranded 'swiss roll'- and one emanating from the esterase domain. In HKU1 HE, all of these loops have been deleted and the top of the molecule is covered by closely-spaced *N*-glycans (Fig. 1; Hurdiss, Drulyte, Lang, de Groot et al., *under review*).

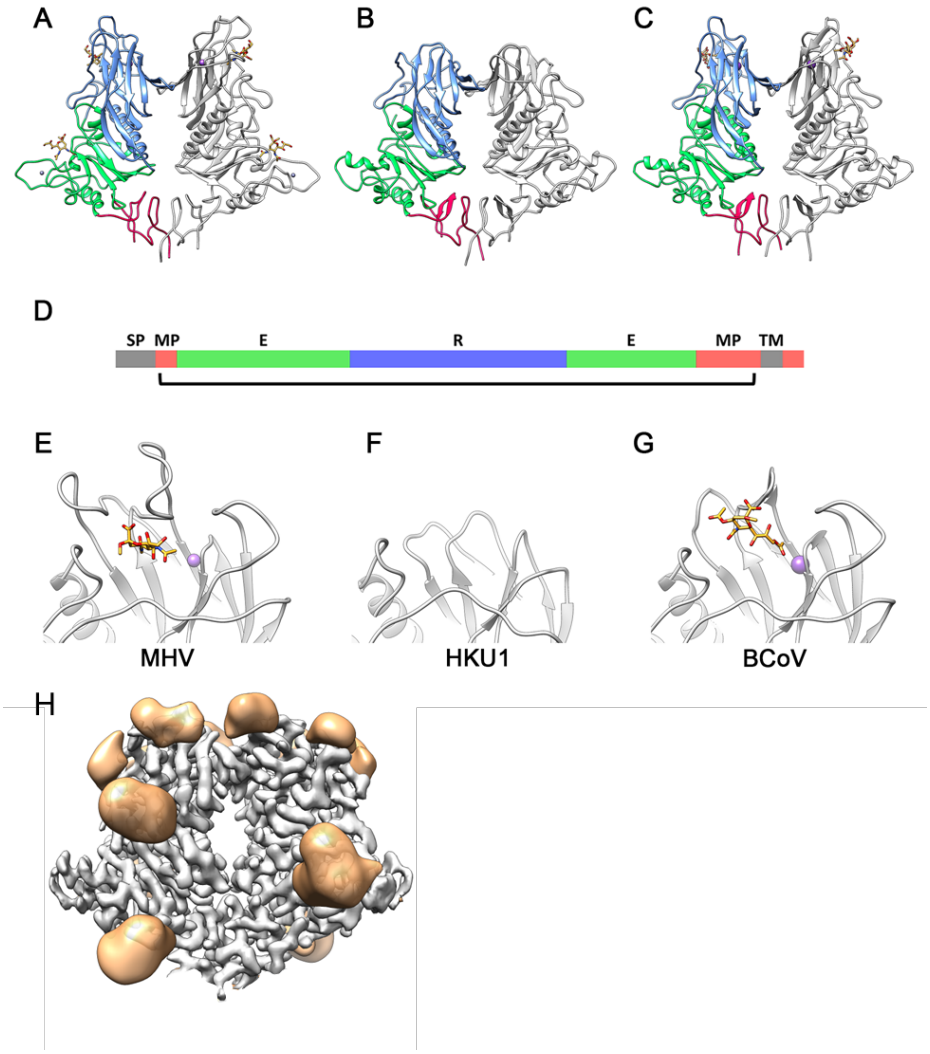


Fig. 1. The 3.5Å cryo-EM structure of HKU1 HE. (Upper panel) Side-by-side comparison of cartoon representations of (A) the crystal structure of MHV-DVIM (PDB: 5JIF), (B) the cryo-EM structure of HKU1 HE and (C) the crystal structure of BCoV-Mebus HE (PDB: 3CL5). (D) Linear schematic representation of HKU1 HE

domain organization. Domains and domain segments are color coded; membrane proximal domain (MP), red; esterase domain segments (E), green; receptor-binding lectin domain (R), blue; signal peptide (SP) and transmembrane domain (TM) indicated in grey. (E-G) Close-ups of the 9-*O*-Ac-Sia binding sites in the HEs of (E) MHV-DVIM and (G) BCoV-Mebus and (F) the corresponding structure in HKU1 HE, all shown in cartoon representation. Ligands (9-*O*-Ac-Sia) in the receptor-binding sites are shown in stick representation and colored by atom type. Note that the loops that constitute the receptor-binding sites in BCoV and MHV HE have been lost in HKU1 HE (F). (H) HKU1 HE glycosylation. HE dimer shown in surface representation (grey) with N-glycans shown as blobs and colored light orange.

More important than how HE lectin function was lost are the consequences of this loss for dynamic virion attachment. In Chapter 2, we demonstrated that the HE lectin domain promotes esterase activity and particularly enhances the destruction of clustered glycotopes such as are present on mucins. Loss of HE lectin function in the HCoVs altered virion-associated receptor-destroying activity: cleavage of 9-*O*-Ac-Sias in bovine submaxillary mucin was strongly reduced and became fully dependent on receptor-binding by S. Intriguingly, and potentially more important for our understanding of virus-glycan interactions, it resulted in selective receptor destruction at least under our in vitro conditions. Certain receptor populations, readily destroyed by soluble HE-Fc fusion proteins, seemed resistant to cleavage by virion-associated HE. These observations led us to propose a model in which the size difference between S and HE, extending 20 and 8 nm from the viral envelope, respectively, determines whether or not particular receptor populations can be cleaved by virion-associated HE. Cleavage would be determined by substrate accessibility. Receptor populations located close to a matrix -the bottom of the ELISA wells in our in vitro assays or theoretically the plasma membrane in the natural setting- may be more difficult to reach with the large spikes bound to such receptors keeping HE at a distance. Conceivably, this effect would be more pronounced upon loss of HE lectin function.

The findings, described in Chapter 2, led us to conclude that adaptation to replication in and transmission from the human respiratory tract involved a reset in the balance between virion attachment and receptor destruction. However, our findings did not allow the conclusion that adaptive changes occurred exclusively in HE. In other words, the possibility of additional adaptations also in the human coronavirus spike proteins was wide open. At the time, however, there was limited information on the structure of embecovirus S proteins. Although the spikes of embecoviruses were known to bind to 9-*O*-Ac-Sias via N-terminal S domain 1A [11,21–23] and a crystal apostructure of BCoV $S1^A$ was available [24], the RBS had not been identified unambiguously. Peng and coworkers [24] proposed a location for the binding site based on a resemblance in folding between $S1^A$ and galectins in combination with limited mutational analysis. Although their conclusions were embraced

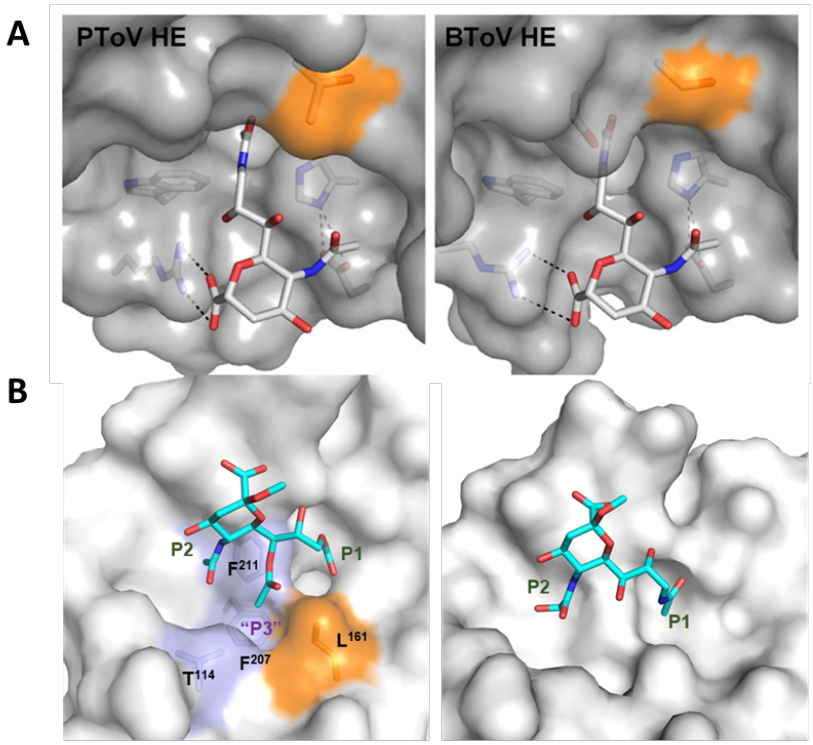
by others in publications in leading journals [25–31], both the reasoning and experimental results in support of the proposed RBS location were not convincing. The hypothesis rested on the assumption that the S proteins and galectins bind their ligands via a site that had been conserved across a vast evolutionary distance, despite the difference in ligands -galectosides and 9-*O*-acetylated Sias- with evident implications for the structural basis of ligand recognition, and despite the fact that during virus evolution, i.e. across relatively short evolutionary distances, glycan-binding sites may readily arise de novo. For example, toro-, corona and orthomyxovirus HEs share 30% sequence identity and recognize the same type of receptor, but their RBSs, though similar in overall design (see below), are not structurally conserved [32–34]. Polyoma VP1 proteins are even more closely related at more than 50% sequence identity yet engage the very same ligand, 5-*N*-Ac-Sia, in at least four different orientations [35–38]. Perhaps the most striking example of plasticity in virus-glycan interactions, different reovirus strains all bind to 5-*N*-Ac-Sia via their attachment protein σ1 but through physically distinct binding sites [35]. Adding to our skepticism regarding the proposed $S1^A$ RBS, substitution of purported key residues did not fully abolish binding. Finally, the proposed site did not resemble typical RBSs for 9-*O*-Ac-Sias as they occur in the HE proteins of corona-, toro- and influenza C viruses [33,39–42]. These considerations prompted us to entertain the possibility that the RBS was located elsewhere. Thus, in Chapter 3, we took a comparative structural analysis approach in combination with automated ligand docking and mutational analysis to identify the true location of the β1CoV S RBS. We show that this site is not only present in the S proteins of BCoV, OC43 and PHEV, but also functionally conserved in the S protein of HKU1. Direct evidence that the RBS is essential for infectivity was provided by pseudotyping vesicular stomatitis virus (VSV) with embecovirus S proteins and RBS-defective mutants thereof. In Chapter 4, we extended our findings by engaging in a collaborative study with the group of David Veesler (Department of Biochemistry, University of Washington, Seattle, WA, USA) to structurally identify the RBS of HCoV-OC43. A holostructure was determined by cryo-electron microscopy of OC43 S in complex with 9-*O*-acetylated sialic acid at 2.8 Å resolution. Moreover, we were first to analyze the binding kinetics of the S RBS to 9-*O*-Ac-Sia in monovalent 1:1 binding assays using biolayer interferometry and synthetic sialosides as ligand. We showed that binding is highly dynamic with association and release occurring within tenths of seconds. While the structural data are informative, they only become relevant when considered in biological context. With the structures of HE and S solved, the stage was set to study their interaction during infection. As described in Chapter 5, we developed a reverse genetics system for BCoV and, replaying OC43 evolution, used the system to abolish the HE RBS. We show that loss of HE lectin function consistently selects for second-site mutations in the S RBS that dramatically reduce S binding affinity. In some mutants, bivalent binding of $S1^A$-Fc fusion proteins to the clustered sialoglycans of bovine submaxillary mucin was reduced by four orders of magnitude. We present

data to demonstrate that HE and S are in functional balance and co-evolve, with selection not only for optimal balance between attachment and receptor-destruction, but also for maximal virion avidity within the given constraints. The data lead us to conclude that the acquisition of an HE gene has set the embecoviruses on an evolutionary path that selected for mechanisms of virus-glycan interactions that are unique among coronaviruses, but remarkably similar to those seen in influenza A viruses (IAVs), as also detailed in Chapter 2.

Embecovirus adaptation to the host sialome. Host adaptation in Influenza A virus (IAVs) has been studied extensively. IAV cross-species transmission generally requires an adaptation to the sialome of the novel host, commonly involving shifts in receptor preference and resetting of the HA/NA functional balance. As a textbook example, avian IAVs prefer α2,3-linked sialosides, those adapted to efficient human-to-human spread prefer α2,6-linkage [43–45] (reviewed in [46–48]). In addition, IAV variants distinguish between 5-*N*-acetylated and 5-*N*-glycolylated Sia [49,50] or display show (in)tolerance for other post-synthetic Sia modification, such as *O*-acetylation, apparently in correlation with host specificity [50,51].

The various β1CoV host range variants as well as HKU1 all target 9-*O*-Ac-Sias in a sialate-9-*O*-Ac-dependent fashion [11–14,52]. The question arises how such a seemingly narrow receptor specificity with binding critically hinging on a single *O*-acetyl group would play out in host specificity, particularly as sialate-9-*O*-acetylation is a very common modification in birds and mammals. Previous studies by Langereis et al [15], with synthetic α2,3- and α2,6-linked sialosides and a panel of HEs from different embecovirus host range variants as well as from bovine and porcine toroviruses and influenza C virus, suggested that glycosidic linkage affects binding affinity albeit to limited extent. HE mediated receptor-binding and receptor-destruction, however, was strongly promoted or hampered by Sia modifications, additional to 9-*O*-acetylation. Some HEs, like those of porcine torovirus, bound to 9-mono-*O*-acetylated, 5-*N*-acetylated ligands exclusively, whereas others like those of bovine corona- and toroviruses were tolerant of either form of 5-*N*-acylation but strongly preferred 7,9-di-*O*-Ac-Sias both as ligands and substrates [33]. Previous HE structure analyses showed esterase substrate specificity for 9-mono- or 7,9-di-*O*-Ac-Sia to be controlled by an inconspicuous single-residue difference in the catalytic site either facilitating or sterically blocking cleavage of di-*O*-acetylated substrates (Fig. 2A, [33]). In the course of the present PhD project, we determined the structural basis for BCoV HE receptor preference by solving holostructures of HE complexed with 7,9-*O*, 5-*N*-Ac- and 9-*N*, 5-*N*-Gc methyl glycosides (Fig. 2B; Lang, Huizinga, de Groot *et al*., *in preparation*). These as yet unpublished data confirmed and extended previous conclusions from biochemical analysis and automated modelling [15] and revealed that

preferential binding of 7,9-di-*O*-Ac-Sia can be ascribed to the docking of the 7-*O*-Ac moiety in a dedicated hydrophobic pocket P3.



C

	Kd (μM)	α-2,3	α-2,6
BCoV S1^{A}	9-*O*-Ac	11	15
	7,9-di-*O*-Ac	7.9	3.7

Fig. 2. Nidovirus HE lectin/substrate fine specificities. (A) Close-ups of the HE esterase catalytic sites of porcine torovirus (PToV) and bovine torovirus (BToV) HE in surface representation. Ligand and contacting amino acid residues shown in sticks representation and colored by atom type (oxygen, red; nitrogen, blue; carbon, gray). Hydrogen bonds between HE and substrate were shown as black, dashed lines. The crucial residues that determine substrate specificity, Thr^{73} (PToV HE) and Ser^{64} (BToV HE), are colored in orange. Figure was adapted from [33]. (B) close-up of the HE lectin RBS of BCoV-Mebus HE in complex with 7,9-di-*O*-Ac-Sia (left) or 9-*N*-Ac, 5-*N*-Gc sia (right). Side chains of residues, contributing to the new hydrophobic pocket (P3), were depicted as sticks and colored in light blue (T^{114}, F^{207}, F^{211}) or orange (L^{161}). (C) Equilibrium dissociation constants (K_D) of different BCoV S1^{A}-SLN complexes determined by steady-state analysis of monovalent 1:1 binding assays using biolayer interferometry.

The observation that HEs of corona- and toroviruses of cattle display a strong ligand and substrate preference for 7,9-di- over 9-mono-*O*-Ac-Sia prompted the hypothesis that this is a host-specific adaptation to replication in ruminant respiratory and gastrointestinal tracts that occurred

independently in two distinct, distantly related nidoviruses [15,33]. In **Chapter 5**, we extended these findings and provided further support for this notion by showing that also the S protein of BCoV preferentially binds 7,9-di-*O*-Ac-Sia. Recombinant dimeric $S1^A$-Fc fusion proteins, when tested by solid phase lectin binding assay with bovine submaxillary mucin, showed that binding avidity to 9-mono-*O*-Ac-Sia was 32-64-fold lower. A more quantitative analysis by biolayer interferometry showed that the BCoV $S1^A$ RBS in monovalent 1:1 binding assays with synthetic sialosides as ligands binds 1.5-4 times more strongly to 7,9-di-*O*-Ac-Sia than 9-*O*-Ac-Sia (Fig. 2C; Lang, Li, Boons and de Groot; *unpublished*). Conversely, the S protein of OC43 ATCC laboratory strain USA/1967 did not show a preference and bound to either receptor type to similar extent albeit with much lower affinity than BCoV S. It is tempting to speculate that the differences between BCoV and OC43 in virion-receptor interaction, i.e. downregulation of HE receptor-destruction, S receptor specificity/promiscuity, and reduced S receptor binding affinity arose from adaptation of OC43 to replication in and transmission from the human respiratory tract. Predictably, a reduction in virion receptor-destroying activity would be advantageous in low receptor-density environments. In accordance, tissue array analysis with HE-based virolectins did detect 9-*O*-Ac-Sias in sections of human airways as expected, but in much lower abundance than in the gastrointestinal tract. As explained in **Chapter 5**, however, the data presented for OC43 should be considered with some caution as the analyses were done for a highly cell-culture adapted isolate, obtained more than 50 years ago [53,54]. Saliently, comparative sequence analysis of OC43 S proteins revealed that considerable changes occurred in $S1^A$, including in the segment L1, that we showed to be critical for receptor-binding (Fig. 3). Future research should establish the S binding properties of contemporary OC43 field strains and determine whether and if so, how these evolved over the last decades.

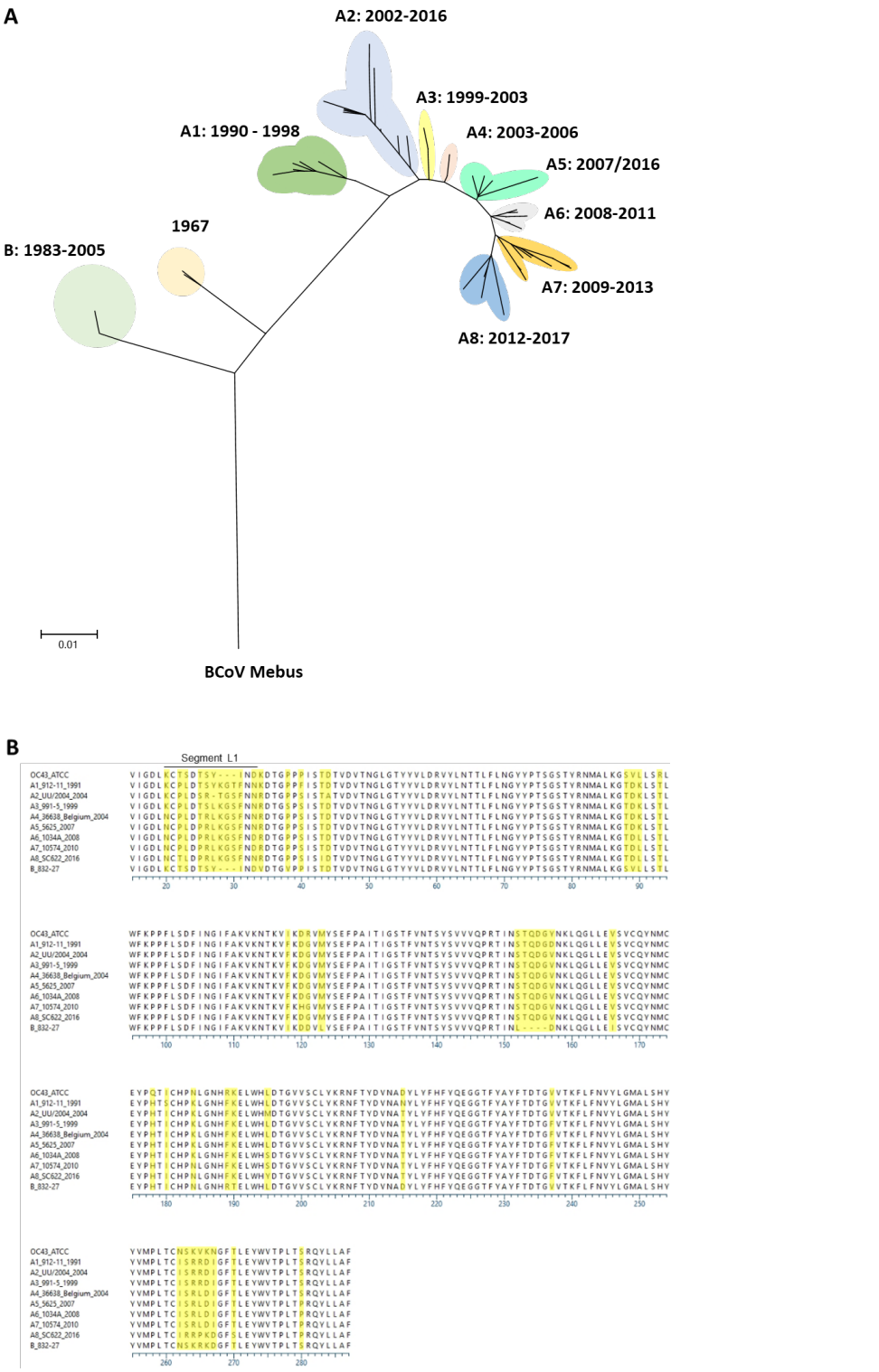


Fig. 3. Continued evolution of HCoV-OC43 S from 1967 onwards. (A) Neighbor-Joining tree based on OC43 $S1^{A}$ sequences separating different OC43 groups with isolation date indicated. (B) Amino acid alignment (15-287) of OC43 $S1^{A}$ from representatives of different groups. Non-conserved residues are marked yellow. Residues comprising segment L1 are indicated.

A receptor is a receptor is a receptor? As explained in Chapter 1, viruses of the mucosal epithelia must traverse the mucus layer and glycocalyx to reach and enter their prospective target cells, while avoiding entrapment through irreversible binding to decoy receptors. The question arises how virus particles without apparent means of energy-driven locomotion manage to pass through layers of mucus that may be tens (respiratory tract) to hundreds (gastrointestinal tract) of micrometers thick, depending on the organ system [55]. A series of recent observations showed that influenza A virus may do so via a mechanism of HA-mediated reversible binding and NA-mediated catalytic virus release, the repetitive action of which culminates in virion motility with receptor-destruction providing directionality, i.e. away from environments depleted for receptors (for a review, see [56]). Diffusion of virus particles in the mucus, was both demonstrated and shown to be NA dependent by single particle tracking and by a "virus in-capsule-mucus penetration system", visualizing the translocation of the virions with time by immunofluorescence staining [57]. Moreover, two-dimensional NA-dependent virus movement on receptor-coated glass surfaces was demonstrated by total internal reflection fluorescence microscopy (TIRFM) combined with single particle tracking analysis [58] and, indirectly, by biolayer interferometry [59].

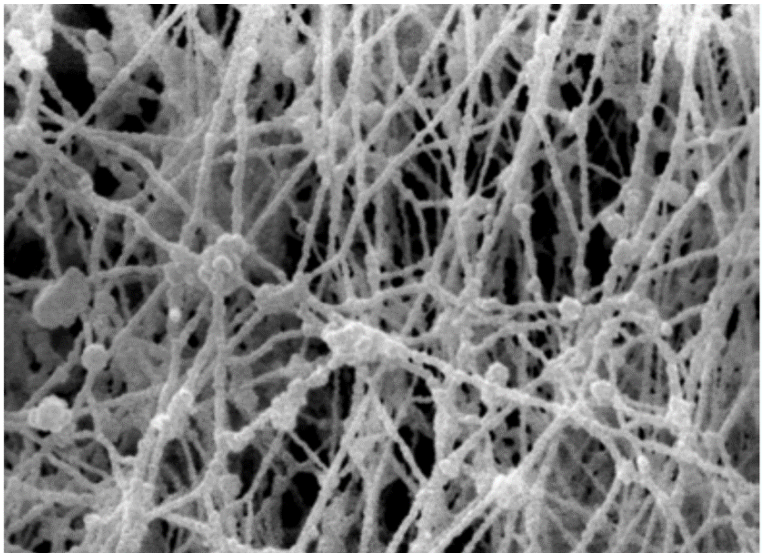


Fig. 4. Mucus from mouse jejunum at 15,000X magnification. Mouse jejunum specimens were treated with RR-saponin-T/O/ T/O method (RR, ruthenium red; T, distilled water; O, 1.0% osmium tetroxide in distilled water) and visualized by high resolution scanning electron microscopy (SEM). At a magnification of 15,000X, mucus "displays filaments that are arranged in a regular meshwork, revealing a globule-bearing structure" as cited from [60]. Figure was adapted from [61].

In high resolution scanning electron micrographs, mucus appears as a mesh of filaments, measuring 30-50 nm in thickness, arranged in regular networks (Fig. 4; [55,61]) and one might envision virions moving along the densely glycosylated mucin cables via directional receptor-assisted diffusion from

the periphery of the mucus gel towards the underlying cells. Virus motility has also been demonstrated for influenza C virus, driven by the receptor-binding and receptor-destroying activities of HEF [58]. It stands to reason that embecoviruses may use comparable mechanisms to avoid irreversible attachment to decoy receptors and to travel to their host cells, like the orthomyxoviruses by relying on a balance between receptor-binding and receptor-destruction that is tailored to the composition of the host sialome. It has been proposed that in orthomyxoviruses, both for IAVs and ICVs, filamentous particles rather than spherical ones are prone to move, which in the case of IAVs would be aided by an asymmetric distribution of NA and HA in their membranes [58,62]. Given that ICV only has one single envelope protein, required for attachment as well as receptor-destruction, such an asymmetric distribution of activities does not seem to be a prerequisite for motility. As far as we are aware, CoVs have been described exclusively as roughly spherical [2]. Although we are skeptical as to whether a filamentous virion morphology is a strict requirement, embecovirion motility remains to be formally demonstrated.

For passage through mucus and penetration of the glycocalix, dynamic, i.e. reversible, binding rather than stable high affinity binding would be key as the latter would obviously result in immobilization. On the cell surface, the situation might be different. Here, low affinity binding in combination with receptor-destruction would predictably hamper successful entry with viruses eluting from the cell rather than being taken up. Our observations in Chapter 5 may be interpreted in this light. We showed that in populations of BCoV mutants with defective HE lectin domain, the consequential loss of HE esterase activity was initially compensated by reducing S affinity to restore S-HE balance. Yet, upon prolonged passage, variants were selected, through successive mutations in both HE and S, of which the overall virion avidity was increased and maximized -not in absolute terms, but to maximum levels as would be allowed by the given constraints- presumably to stabilize attachment and enhance infection.

The question arises whether functional attachment -that is attachment leading to entry- is subject only to overall virion avidity and the balance between binding and release. On their way through the mucus towards the epithelial cells as well as on the cell surface, virions may encounter different sialosides to which it may bind with higher or lower affinity and that may be destroyed, with higher or lower efficiency by the virion-associated receptor-destroying enzymes (RDEs), depending on the presence or absence of particular post-synthetic Sia modifications and or the glycosidic linkage and structure of the underlying glycan chain. Theoretically, functional attachment promoting cell entry may also be determined at the level of receptor fine specificity and/or receptor accessibility. Might there be a distinction between transient receptors that promote virus motility and specific cell surface receptor variants and/or receptor populations for functional attachment? Here, our

observations described in Chapter 2 come to mind, where we proposed that virion-mediated destruction of receptor populations may be limited by substrate accessibility due to the size difference between S and HE. Although this phenomenon was particularly pronounced for OC43, which lacks HE lectin function, it was also observed for BCoV particles (Chapter 2, Fig. 5). A premise for such a mechanism to be relevant under natural conditions is that on the cell surface, such inaccessible receptor populations do in fact exist. One group of molecules that would qualify are O-acetylated disialogangliosides like GD3, the glycan chain of which would extend less than 15Å from the lipid bilayer. Actually, gangliosides previously were shown to serve as entry receptors for influenza C virus [63] as well as for a range of other viruses (for a review see [64]) . Support that such glycolipids may also act as embecovirus receptors comes from a series of as yet unpublished experiments in which we transfected HEK293T cells with alpha2,8 sialyltransferase 1 (ST8SIA1) aka GD3 synthase [65], and observed a stark increase in binding of OC43 $S1^A$-Fc as tested by immunofluorescence assay (Fig. 5A). Binding also to mock-transfected cells was not unexpected as HEK293T-cells are susceptible to infection by OC43-1967/USA. HKU1 can only be propagated in epithelial cell cultures and select cell lines, but to our knowledge, not in HEK293T cells. In accordance, HKU1 $S1^A$-Fc does not detectably bind to mock-transfected cells. Remarkably, however, it strongly bound to cells overexpressing ST8SIA1 (Fig. 5A). Substitution of the critical $S1^A$ RBS residue Trp90 by Ala abolished binding. Binding was also strongly reduced upon sialate-de-9-*O*-acetylation of transfected cells with BCoV HE (Fig. 5A). Thus, the data unambiguously show specific binding of both OC43 and HKU1 $S1^A$ to 9-*O*-acetylated GD3. Specific 9-*O*-Ac-Sia-dependent binding of these virolectins was not only observed to transfected cultured cells, but also to lung epithelium (Fig. 5B). The strong binding of HKU1 $S1^A$ is particularly surprising given that in previous studies attachment of HKU1 S to 9-*O*-Ac-Sias was either undetectable [52] or exceedingly difficult to show using standard in vitro assays. In Chapter 3, we explained the poor binding of HKU1 $S1^A$ from the architecture of its RBS, which is located at the bottom of a canyon. We showed that the ridges surrounding the RBS sterically hinder ligand binding and speculated that the binding site may only be accessible to long extended glycan chains. In view of the current data, it seems more likely that binding is specific for α2,8-linked 9-*O*-Ac-Sias. Glycosidic linkage to the C8 atom in the glycerol side chain of the adjacent Sia residue as compared to the C3 or C6 atom would increase ligand flexibility as well as increase the distance between the two pyranose rings. Binding of α2,3- and α2,6-linked ligands might be hampered by steric hindrance. The most important remaining question is whether GD3 also supports entry. This is indeed the case, as we showed by testing VSV particles pseudotyped with the S proteins of OC43 and HKU1 and RBS-inactive derivatives thereof as we did in Chapters 3 and 4. As shown in Fig. 5C, mock-transfected HEK293T cells were susceptible to infection with OC43 S-pseudotyped VSV as would be expected, but the infection increased significantly upon ST8SIA1 over-expression.

Mock-transfected HEK293T cells were not detectably infected by VSV pseudotyped with HKU1 S, but, strikingly, became susceptible when transfected to overexpress ST8SIA1. VSV particles pseudotyped with RBS-defective S proteins were not infectious. Moreover, infection was prohibited in transfected 9-*O*-Ac-Sia deficient HEK293T-CASD1^{-} cells.

The data show that OC43 as well as HKU1, like influenza C virus [63], can use gangliosides decorated with 9-*O*-acetylated α2,8-linked di-Sias as entry receptors and suggest that perhaps indeed a distinction should be made between 9-*O*-Ac-Sia receptor types that allow and promote virus motility through the mucus and those that provide entrance to the cell. If so, this would be a departure from our current understanding of virus-glycan interactions with potential implications for antiviral intervention.

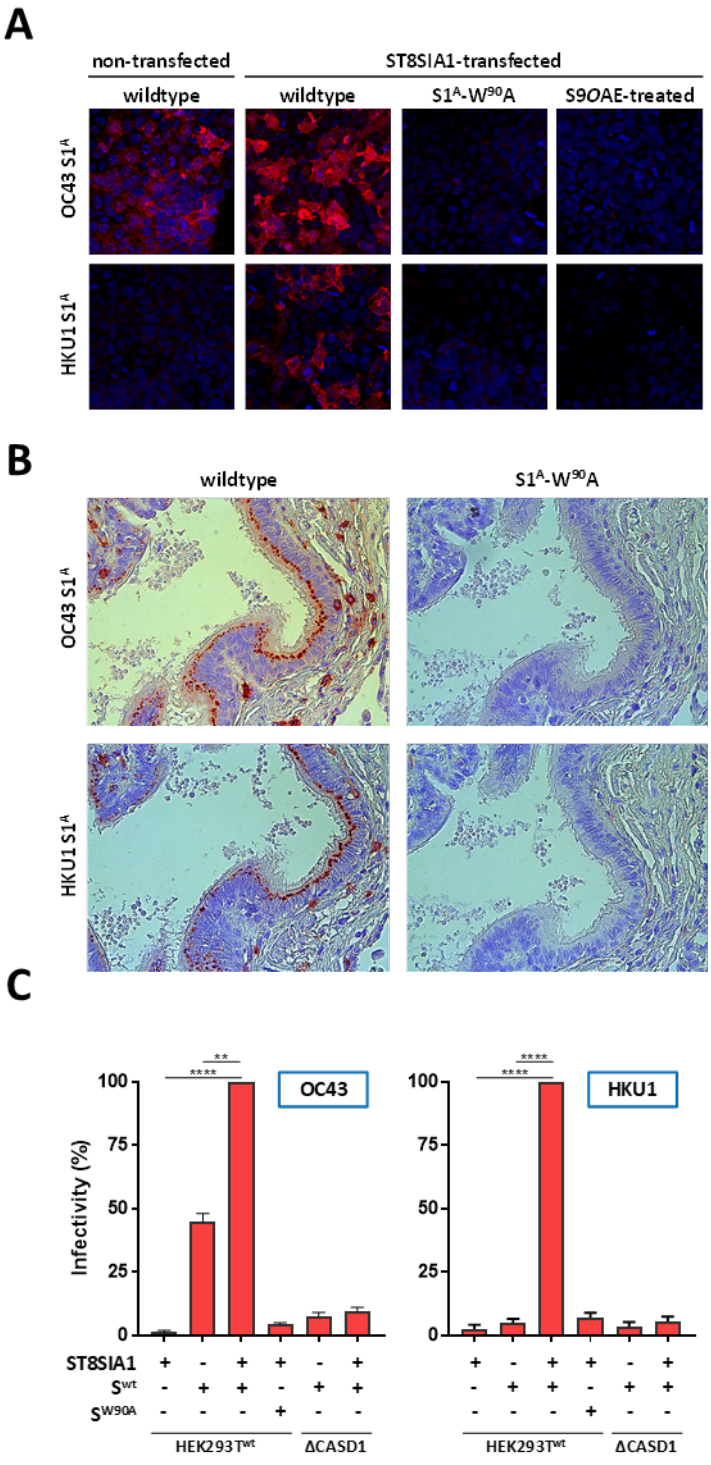


Fig. 5. 9-*O*-acetylated GD3 serves as a receptor for human coronavirus OC43 and HKU1. (A) Binding of OC43 and HKU1 $S1^A$-Fc fusion proteins to HEK293T cells, mock-transfected or transfected to overexpress ST8SIA1 and 9-*O*-Ac-GD3 as shown by immunofluorescence. Binding is abolished upon substitution of the critical $S1^A$ RBS residue Trp90 by Ala ($W^{90}A$), and upon depletion of 9-*O*-Ac-Sias by sialate-9-*O*-acetyl esterase (S9OAE) treatment with BCoV HE. (B) Specific binding of OC43 and HKU1

$S1^{A}$-Fc fusion proteins to lung epithelium as shown by immunohistochemistry. Note that binding is abolished by RBS $W^{90}A$ substitution. Paraffin-embedded, formalin-fixed tissue sections kindly provided by Dr. Debby van Riel, EMC. (C) GD3 supports entry of VSV particles, pseudotyped with the S proteins of OC43 or HKU1. Note that infection is dependent on a functional S RBS and prohibited in ST8SIA1-transfected 9-*O*-Ac-Sia deficient HEK293T-$CASD1^{-}$ cells.

Concluding remarks. Our studies with BCoV, HCoV-OC43 and HKU1, extending previous work, highlight the crucial role of virion binding to 9-*O*-Ac-Sia during pre-attachment and attachment. They add to our understanding of embecovirus evolution, cross-species transmission and host exclusivity. Amongst others, the findings demonstrate the existence of a critical functional balance between S-mediated attachment and HE-mediated catalytic virion elution and suggest that embecovirus cross-species transmission requires a resetting of this balance as well as optimization of overall viral avidity to match the sialoglycome in the target organs of the novel host. This mechanism of receptor usage, entailing a concerted and carefully fine-tuned activity of two envelope proteins, S and HE, is unique among CoVs, but reminiscent in many aspects of that of influenza A viruses (IAVs) [58,59,66–74]. Importantly, however, our observations go beyond those made for IAVs that are scattered over many papers published over many years. Our studies bring embecoviruses from the backfield to the front line to show that general principles fundamental to virion-sialoglycan interactions prompted convergent evolution of two important zoonotically-relevant groups of human and animal pathogens. Our data even further our understanding of other viruses, including some CoVs like MERS CoV, that attach to Sia but that lack RDEs. As shown in **Chapter 5** for VSV particles bearing mutant embecovirus S proteins, a sharp reduction in receptor binding affinity, to levels barely detectable by *in vitro* assays, preserved infectivity but caused infection to become HE-independent or even to be inhibited by exogenous esterase activity. Apparently, ultralow receptor binding affinity ensures reversibility of attachment, thus preventing entrapment by decoy receptors and, presumably, still enables virus motility albeit through non-directional receptor-assisted browsing. Having said this, many aspects of embecovirus-glycan interactions remain incompletely understood. We established for β1CoVs and HKU1 the importance of a functional HE-S balance, lifting the status of HE from an exotic accessory bystander to an essential partner in a two-component system for dynamic virion-receptor interactions. We also provide evidence that it is not only about balance between binding and release but also about optimal overall virion avidity. On a sobering thought, however, it is far from clear precisely how this balance and optimal virion avidity relates to differences in Sia diversity and Sia density in the target organs of different host species. We extend previous observations that among embecoviruses differences in receptor fine specificity exist and for a number of S and HE proteins determined the structural basis for receptor preference. Using synthetic

sialosides and biolayer interferometry we even provided first quantitative analyses of the binding properties of the S RBS. While these observations allow general sweeping speculations about the factors and conditions that might favor or hamper cross species transmission and host adaptation, a true understanding of the phenomena that occur during pre-attachment and attachment and how these play into host tropism, would require further research, including quantitative comparative analyses of Sia composition in the mucus and at the cell surface in different host species. An assessment of how evolution during continued OC43 circulation since 1967 has shaped the properties of the S protein may provide further insight. On a final note, our findings indicate that for embecoviruses, efforts into prophylactic and therapeutic intervention should not necessarily focus on S exclusively but might consider also HE as vaccine candidate and druggable target.

References

1. International Committee on Taxonomy of Viruses (ICTV). **2020**.
2. de Groot, R.; Baker, S.; Baric, R.; … L.E.-V.; 2012, U. Family coronaviridae. *Elsevier Amsterdam* **2011**, 806–828.
3. Woo, P.C.Y.; Lau, S.K.P.; Lam, C.S.F.; Lau, C.C.Y.; Tsang, A.K.L.; Lau, J.H.N.; Bai, R.; Teng, J.L.L.; Tsang, C.C.C.; Wang, M.; et al. Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavi. *J. Virol.* **2012**, *86*, 3995–4008.
4. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, NEJMoa2001017.
5. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv* **2020**, 2020.01.22.914952.
6. Lau, S.K.P.; Woo, P.C.Y.; Li, K.S.M.; Tsang, A.K.L.; Fan, R.Y.Y.; Luk, H.K.H.; Cai, J.-P.; Chan, K.-H.; Zheng, B.-J.; Wang, M.; et al. Discovery of a Novel Coronavirus, China Rattus Coronavirus HKU24, from Norway Rats Supports the Murine Origin of Betacoronavirus 1 and Has Implications for the Ancestor of Betacoronavirus Lineage A. *J. Virol.* **2015**, *89*, 3076–3092.
7. Vijgen, L.; Keyaerts, E.; Lemey, P.; Maes, P.; Van Reeth, K.; Nauwynck, H.; Pensaert, M.; Van Ranst, M. Evolutionary History of the Closely Related Group 2 Coronaviruses: Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human Coronavirus OC43. *J. Virol.* **2006**, *80*, 7270–7274.
8. Vijgen, L.; Keyaerts, E.; Moes, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604.
9. Lau, S.K.P.; Lee, P.; Tsang, A.K.L.; Yip, C.C.Y.; Tse, H.; Lee, R.A.; So, L.-Y.; Lau, Y.-L.; Chan, K.-H.; Woo, P.C.Y.; et al. Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. *J. Virol.* **2011**, *85*, 11325–11337.
10. Lau, S.K.P.; Woo, P.C.Y.; Yip, C.C.Y.; Fan, R.Y.Y.; Huang, Y.; Wang, M.; Guo, R.; Lam, C.S.F.; Tsang, A.K.L.; Lai, K.K.Y.; et al. Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J. Virol.* **2012**, *86*, 5481–96.
11. Vlasak, R.; Luytjes, W.; Spaan, W.; Palese, P. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci.* **1988**, *85*, 4526–4529.
12. Vlasak, R.; Luytjes, W.; Leider, J.; Spaan, W.; Palese, P. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J. Virol.* **1988**, *62*, 4686–90.
13. Schultze, B.; Gross, H.J.; Brossmer, R.; Klenk, H.D.; Herrler, G. Hemagglutinating encephalomyelitis virus attaches to N-acetyl-9-O-acetylneuraminic acid-containing receptors on erythrocytes: comparison with bovine coronavirus and influenza C virus. *Virus Res.* **1990**, *16*, 185–94.
14. Schultze, B.; Gross, H.J.; Klenk, H.D.; Brossmer, R.; Herrler, G. Differential reactivity of bovine coronavirus (BCV) and influenza C virus with N-acetyl-9-O-acetylneuraminic acid (NEU5, 9AC2)-containing receptors. In Proceedings of the Advances in Experimental Medicine and Biology; 1990; Vol. 276, pp. 115–120.
15. Langereis, M.A.; Bakkers, M.J.G.; Deng, L.; Padler-Karavani, V.; Vervoort, S.J.; Hulswit, R.J.G.; van Vliet, A.L.W.; Gerwig, G.J.; de Poot, S.A.H.; Boot, W.; et al. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep.* **2015**, *11*, 1966–1978.
16. Shen, Y.; Tiralongo, J. Regulation of sialic acid O-acetylation in human colon mucosa. *Biol. Chem.* **2004**, *385*, 145–152.
17. Muchmore, E.A.; Varki, N.M.; Fukuda, M.; Varki, A. Developmental regulation of sialic acid modifications in rat and human colon. *FASEB J.* **1987**, *1*, 229–235.
18. Klein, A.; Roussel, P. O-acetylation of sialic acids. *Biochimie* **1998**, *80*, 49–57.
19. Cohen, M.; Varki, A. The sialome--far more than the sum of its parts. *OMICS* **2010**, *14*, 455–64.
20. Schauer, R.; Kamerling, J.P. Exploration of the Sialic Acid World. In *Advances in Carbohydrate Chemistry and Biochemistry*; Academic Press, 2018; Vol. 75, pp. 1–213 ISBN 9780128152027.
21. Künkel, F.; Herrler, G. Structural and functional analysis of the surface protein of human coronavirus OC43.

Virology **1993**, *195*, 195–202.

22. Schultze, B.; Gross, H.J.; Brossmer, R.; Herrler, G. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-O-acetylated sialic acid as a receptor determinant. *J. Virol.* **1991**, *65*, 6232–7.

23. Peng, G.; Sun, D.; Rajashankar, K.R.; Qian, Z.; Holmes, K. V; Li, F. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10696–701.

24. Peng, G.; Xu, L.; Lin, Y.L.; Chen, L.; Pasquarella, J.R.; Holmes, K. V.; Li, F. Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **2012**, *287*, 41931–41938.

25. Shang, J.; Zheng, Y.; Yang, Y.; Liu, C.; Geng, Q.; Luo, C.; Zhang, W.; Li, F. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. *PLoS Pathog.* **2018**, *14*, e1007009.

26. Xiong, X.; Tortorici, M.A.; Snijder, J.; Yoshioka, C.; Walls, A.C.; Li, W.; McGuire, A.T.; Rey, F.A.; Bosch, B.-J.; Veesler, D. Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections. *J. Virol.* **2017**, JVI.01628-17.

27. Kirchdoerfer, R.N.; Cottrell, C.A.; Wang, N.; Pallesen, J.; Yassine, H.M.; Turner, H.L.; Corbett, K.S.; Graham, B.S.; McLellan, J.S.; Ward, A.B. Pre-fusion structure of a human coronavirus spike protein. *Nature* **2016**, *531*, 118–121.

28. Qian, Z.; Ou, X.; Góes, L.G.B.; Osborne, C.; Castano, A.; Holmes, K. V; Dominguez, S.R. Identification of the Receptor-Binding Domain of the Spike Glycoprotein of Human Betacoronavirus HKU1. *J. Virol.* **2015**, *89*, 8816–8827.

29. Ou, X.; Guan, H.; Qin, B.; Mu, Z.; Wojdyla, J.A.; Wang, M.; Dominguez, S.R.; Qian, Z.; Cui, S. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat. Commun.* **2017**, *8*.

30. Yuan, Y.; Cao, D.; Zhang, Y.; Ma, J.; Qi, J.; Wang, Q.; Lu, G.; Wu, Y.; Yan, J.; Shi, Y.; et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **2017**, *8*, 15092.

31. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol.* **2016**, *3*, 237–261.

32. de Groot, R.J. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* 2006, *23*, 59–72.

33. Langereis, M.A.; Zeng, Q.; Gerwig, G.J.; Frey, B.; Itzstein, M. von; Kamerling, J.P.; de Groot, R.J.; Huizinga, E.G. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15897–15902.

34. Langereis, M. a. *Viral hemagglutinin-esterases: Mediators of dynamic virion-glycan interactions*; 2011; ISBN 978-90-393-5629-6.

35. Ströh, L.J.; Stehle, T. Glycan Engagement by Viruses: Receptor Switches and Specificity. *Annu. Rev. Virol.* **2014**, *1*, 285–306.

36. Stehle, T.; Harrison, S.C. High-resolution structure of a polyomavirus VP1-oligosaccharide complex: Implications for assembly and receptor binding. *EMBO J.* **1997**, *16*, 5139–5148.

37. Neu, U.; Khan, Z.M.; Schuch, B.; Palma, A.S.; Liu, Y.; Pawlita, M.; Feizi, T.; Stehle, T. Structures of B-Lymphotropic Polyomavirus VP1 in Complex with Oligosaccharide Ligands. *PLoS Pathog.* **2013**, *9*, e1003714.

38. Neu, U.; Hengel, H.; Blaum, B.S.; Schowalter, R.M.; Macejak, D.; Gilbert, M.; Wakarchuk, W.W.; Imamura, A.; Ando, H.; Kiso, M.; et al. Structures of merkel cell polyomavirus VP1 complexes define a sialic acid binding site required for infection. *PLoS Pathog.* **2012**, *8*, 8.

39. Zeng, Q.; Langereis, M.A.; van Vliet, A.L.W.; Huizinga, E.G.; de Groot, R.J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9065–9069.

40. Bakkers, M.J.G.; Zeng, Q.; Feitsma, L.J.; Hulswit, R.J.G.; Li, Z.; Westerbeke, A.; van Kuppeveld, F.J.M.; Boons, G.-J.; Langereis, M.A.; Huizinga, E.G.; et al. Coronavirus receptor switch explained from the stereochemistry of protein–carbohydrate interactions and a single mutation. *Proc. Natl. Acad. Sci.* **2016**, *113*, E3111–E3119.

41. Rosenthal, P.B.; Zhang, X.; Formanowski, F.; Fitz, W.; Wong, C.H.; Meier-Ewert, H.; Skehel, J.J.; Wiley, D.C. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **1998**, *396*, 92–96.

42. Bakkers, M. Sweet attachment: structural and functional studies on nidovirus hemagglutinin-esterases, Utrecht University, 2017.

43. Rogers, G.N.; Paulson, J.C. Receptor determinants of human and animal influenza virus isolates: Differences in

receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **1983**, *127*, 361–373.

44. Rogers, G.N.; Paulson, J.C.; Daniels, R.S.; Skehel, J.J.; Wilson, I.A.; Wiley, D.C. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **1983**, *304*, 76–78.

45. Rogers, G.N.; D'Souza, B.L. Receptor binding properties of human and animal H1 influenza virus isolates. *Virology* **1989**, *173*, 317–322.

46. Lipsitch, M.; Barclay, W.; Raman, R.; Russell, C.J.; Belser, J.A.; Cobey, S.; Kasson, P.M.; Lloyd-Smith, J.O.; Maurer-Stroh, S.; Riley, S.; et al. Viral factors in influenza pandemic risk assessment. *Elife* **2016**, *5*.

47. Shi, Y.; Wu, Y.; Zhang, W.; Qi, J.; Gao, G.F. Enabling the "host jump": Structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* 2014, *12*, 822–831.

48. Long, J.S.; Mistry, B.; Haslam, S.M.; Barclay, W.S. Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* **2019**, *17*, 67–81.

49. Song, X.; Yu, H.; Chen, X.; Lasanajak, Y.; Tappert, M.M.; Air, G.M.; Tiwari, V.K.; Cao, H.; Chokhawala, H.A.; Zheng, H.; et al. A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. *J. Biol. Chem.* **2011**, *286*, 31610–31622.

50. Higa, H.H.; Rogers, G.N.; Paulson, J.C. Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N,O-diacetyineuraminic acids. *Virology* **1985**, *144*, 279–282.

51. Bradley, K.C.; Galloway, S.E.; Lasanajak, Y.; Song, X.; Heimburg-Molinaro, J.; Yu, H.; Chen, X.; Talekar, G.R.; Smith, D.F.; Cummings, R.D.; et al. Analysis of Influenza Virus Hemagglutinin Receptor Binding Mutants with Limited Receptor Recognition Properties and Conditional Replication Characteristics. *J. Virol.* **2011**, *85*, 12387–12398.

52. Huang, X.; Dong, W.; Milewska, A.; Golda, A.; Qi, Y.; Zhu, Q.K.; Marasco, W.A.; Baric, R.S.; Sims, A.C.; Pyrc, K.; et al. Human Coronavirus HKU1 Spike Protein Uses O -Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J. Virol.* **2015**, *89*, 7202–7213.

53. McIntosh, K.; Dees, J.H.; Becker, W.B.; Kapikian, A.Z.; Chanock, R.M. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci.* **1967**, *57*, 933–940.

54. Becker, W.B.; McIntosh, K.; Dees, J.H.; Chanock, R.M. Morphogenesis of avian infectious bronchitis virus and a related human virus (strain 229E). *J. Virol.* **1967**, *1*, 1019–27.

55. Bansil, R.; Turner, B.S. The biology of mucus: Composition, synthesis and organization. *Adv. Drug Deliv. Rev.* 2018.

56. de Vries, E.; Du, W.; Guo, H.; de Haan, C.A.M. Influenza A Virus Hemagglutinin–Neuraminidase–Receptor Balance: Preserving Virus Motility. *Trends Microbiol.* 2019.

57. Yang, X.; Steukers, L.; Forier, K.; Xiong, R.; Braeckmans, K.; Van Reeth, K.; Nauwynck, H. A beneficiary role for neuraminidase in influenza virus penetration through the respiratory mucus. *PLoS One* **2014**, *9*, e110026.

58. Sakai, T.; Nishimura, S.I.; Naito, T.; Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Sci. Rep.* **2017**, *7*.

59. Guo, H.; Rabouw, H.; Slomp, A.; Dai, M.; van der Vegt, F.; van Lent, J.W.M.; McBride, R.; Paulson, J.C.; de Groot, R.J.; van Kuppeveld, F.J.M.; et al. Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. *PLoS Pathog.* **2018**, *14*.

60. Familiari, G.; Relucenti, M.; Familiari, A.; Battaglione, E.; Franchitto, G.; Heyn, R. *Visualization of the real microarchitecture of glycoprotein matrices with scanning electron microscopy*;

61. Familiari, G.; Nottola, S.A.; Macchiarelli, G.; Familiari, A.; Motta, P.M. A technique for exposure of the glycoproteic matrix (zona pellucida and mucus) for scanning electron microscopy. *Microsc. Res. Tech.* **1992**, *23*, 225–229.

62. Vahey, M.D.; Fletcher, D.A. Influenza a virus surface proteins are organized to help penetrate host mucus. *Elife* **2019**, *8*.

63. Herrler, G.; Klenk, H.D. The surface receptor is a major determinant of the cell tropism of influenza C virus. *Virology* **1987**, *159*, 102–108.

64. Matrosovich, M.; Herrler, G.; Klenk, H.D. Sialic Acid Receptors of Viruses. *Top. Curr. Chem.* **2015**, *367*, 1–28.

65. Baumann, A.M.T.; Bakkers, M.J.G.; Buettner, F.F.R.; Hartmann, M.; Grove, M.; Langereis, M.A.; De Groot, R.J.; Muhlenhoff, M. 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. *Nat. Commun.* **2015**, *6*.

66. Varghese, J.N.; Colman, P.M.; van Donkelaar, A.; Blick, T.J.; Sahasrabudhe, A.; McKimm-Breschkin, J.L. Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl. Acad. Sci.* **1997**, *94*, 11808–11812.

67. Xu, R.; Zhu, X.; McBride, R.; Nycholat, C.M.; Yu, W.; Paulson, J.C.; Wilson, I.A. Functional Balance of the Hemagglutinin and Neuraminidase Activities Accompanies the Emergence of the 2009 H1N1 Influenza Pandemic. *J. Virol.* **2012**, *86*, 9221–9232.

68. Mitnaul, L.J.; Matrosovich, M.N.; Castrucci, M.R.; Tuzikov, A.B.; Bovin, N. V; Kobasa, D.; Kawaoka, Y. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *J. Virol.* **2000**, *74*, 6015–20.

69. Wagner, R.; Matrosovich, M.; Klenk, H.D. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* 2002, *12*, 159–166.

70. Baum, L.G.; Paulson, J.C. The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* **1991**, *180*, 10–15.

71. Cohen, M.; Zhang, X.Q.; Senaati, H.P.; Chen, H.W.; Varki, N.M.; Schooley, R.T.; Gagneux, P. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virol. J.* **2013**, *10*.

72. Matrosovich, M.N.; Matrosovich, T.Y.; Gray, T.; Roberts, N.A.; Klenk, H.-D. Neuraminidase Is Important for the Initiation of Influenza Virus Infection in Human Airway Epithelium. *J. Virol.* **2004**, *78*, 12665–12667.

73. Hooper, K.A.; Bloom, J.D. A Mutant Influenza Virus That Uses an N1 Neuraminidase as the Receptor-Binding Protein. *J. Virol.* **2013**, *87*, 12531–12540.

74. Xue, K.S.; Hooper, K.A.; Ollodart, A.R.; Dingens, A.S.; Bloom, J.D. Cooperation between distinct viral variants promotes growth of h3n2 influenza in cell culture. *Elife* **2016**, *5*.

Addendum

Nederlandse Samemvatting

Acknowledgement

Curriculum Vitae

List of publications

Nederlandse samenvatting

Coronavirussen (CoVs) staan erom bekend dat zij van gastheer kunnen wisselen. De overdracht van CoVs vanuit natuurlijke reservoirs -vleermuizen, vogels, knaagdieren- naar landbouwhuisdieren en gezelschapsdieren kan leiden tot het ontstaan van nieuwe veterinair-relevante ziekten. Zoönotische overdracht van CoVs -van dier naar mens- vormen echter ook een bijzonder ernstige bedreiging voor de mondiale volksgezondheid, zoals pijnlijk geïllustreerd door de huidige pandemie van het type 2 severe acute respiratory syndrome coronavirus. Dit nieuwe virus is de humane populatie binnengekomen laat in de herfst van 2019 en heeft na een aanvankelijk lokale uitbraak in Wuhan (provincie Hubei, China), zich razendsnel verspreid, eerst binnen China en vervolgens over de rest van de wereld, met dramatische gevolgen. Het ziet er naar uit dat wij voorlopig niet van dit virus af zijn. Naar alle waarschijnlijkheid zal het zich voegen bij vier andere respiratoire humane coronavirussen (HCoVs), die eerder met succes de sprong vanuit het dier hebben gemaakt en die zich blijvend in de humane populatie hebben gevestigd. Het is van groot belang om inzicht te verkrijgen in de factoren, die zoönotische introductie van CoVs en kolonisatie van de mens als gastheer mogelijk maken. Allesbepalend hierbij is de susceptibiliteit van de nieuwe gastheer, d.w.z. of de cellen van de nieuwe gastheersoort al dan niet oppervlaktemoleculen dragen, die door het virus gebruikt kunnen worden om binnen te dringen. Tot dusver heeft onderzoek naar receptorgebruik bij CoVs in relatie tot cross species transmissie zich vooral gericht op eiwit-gebaseerde receptoren. Echter, van de vier thans bekende endemische humane coronavirussen binden er maar liefst twee niet aan eiwitten, maar aan suikerketens, waarmee cellulaire oppervlakte-eiwitten en lipiden gedecoreerd zijn. Deze virussen, HCoV-OC43 en -HKU1 genaamd, binden aan een eindstandig suikerresidu, siaalzuur, en dan met name aan gemodificeerd siaalzuren, die postsynthetisch voorzien zijn van een *O*-acetyl groep op de koolstof 9 positie (9-*O*-geacetyleerd siaalzuur, 9-*O*-Ac-Sia). Wat het extra bijzonder maakt is dat deze virussen nauw verwant zijn, behorend tot een kleine coronavirus subgroep (het subgenus *Embecovirus* in het genus *Betacoronavirus*), en er afzonderlijk in geslaagd zijn om de mens te koloniseren. Andere leden van dit subgenus zijn, direct of indirect vanuit knaagdieren, terecht gekomen in onder andere konijnen, paarden, varkens, honden, en dromedarissen. HCoV-OC43 lijkt recentelijk, midden vorige eeuw, te zijn voortgekomen uit een embecovirus van het rund, het bovine coronavirus (BCoV).

Embecovirussen verschillen behalve in hun receptorvoorkeur ook nog van alle andere CoVs in hun virioncompositie. Naast het spike eiwit S, dat de voor CoV karakteristieke 20-nm lange spikes vormt en verantwoordelijk is voor receptorbinding en membraanfusie, coderen embecovirussen voor een uniek envelop eiwit, het hemagglutinine-esterase (HE). HE klieft de *O*-acetylgroep van 9-*O*-Ac-Sia residuen af en fungeert daarmee als virion-geassocieerd receptor-vernietigend enzym.

Kennelijk zijn embecovirussen uiterst bedreven in het veroveren van nieuwe gastheersoorten, maar merkwaardig genoeg zijn alle genoemde virussen ook in hoge mate gastheerspecifiek. Zij springen dus

niet voortdurend heen en weer tussen verschillende gastheersoorten, maar blijven na de oversprong hun nieuwe gastheersoort trouw. Vermoedelijk hangt dit samen met hun receptorgebruik. Het gebruik van suikers als receptor komt met een eigen set aan regels en complexiteiten, die vooralsnog slechts ten dele begrepen worden. De doelstellingen van het onderzoek, beschreven in dit proefschrift, was tweeledig: i) het verkrijgen van meer inzicht in virus-suikerinteracties met CoVs als modelsysteem en ii) om te bepalen of de zoönotische introductie van OC43 en HKU1 en hun aanpassing aan replicatie in de bovenste luchtwegen van de mens, gepaard is gegaan met veranderingen in receptorgebruik, waaruit wellicht hun nieuw verworven gastheerspecificiteit verklaard kan worden. Wij hebben ons daarbij vooral gericht op het bovine coronavirus (BCoV) en HCoV-OC43, die een zoönotisch paar vormen en daarom een uitstekend model boden voor onze studies.

In **Hoofdstuk 2** laten wij zien, dat de HEs van HCoV-OC43 en HKU1 essentieel verschillen van die van embecovirussen bij dieren. HE bestaat oorspronkelijk uit twee domeinen: een suikerbindend lectine domein, dat net als S bindt aan 9-*O*-Ac-Sia, en een enzymatisch-actief esterase domein. Bij dierlijke embecovirussen is het lectine domein intact en functioneel, maar bij de HCoVs is deze functie verloren gegaan. Bij OC43 is dat geleidelijk gegaan door accumulatie van puntmutaties in het lectinedomein, waardoor de receptorbindingsplaats (RBP) uitgeschakeld is. Bij HCoV-HKU1 dat langer in de mens heeft gecirculeerd zijn grote delen van het lectinedomein verdwenen. Wij laten zien dat het HE lectinedomein de esterase-activiteit reguleert en met name de vernietiging bevordert van geclusterde suikerreceptoren, zoals die voorkomen op mucines (zwaar geglycosyleerde eiwitten en de belangrijkste component van de beschermende slijmlaag van de luchtwegen en het maagdarmkanaal). Verlies van HE lectine functie, kennelijk in samenhang met verschil in grootte tussen S en HE, leidt tot een verandering in de virion-geassocieeerde enzymactiviteit. Deze is sterk gereduceerd, afhankelijk van activiteit van receptor-binding door S en selectief voor bepaalde receptorpopulaties. Receptoren, die vernietigd kunnen worden door vrij, niet virion-gebonden HE, zijn kennelijk voor virion-geassocieerd HE niet langer toegankelijk. Wij concluderen hieruit dat aanpassing aan replicatie in en transmissie vanuit de humane respiratoire tract een verandering vergde in de balans tussen S-gemediëerde virusaanhechting en HE gemedieerde receptorvernietiging. Onze bevindingen sloten echter niet uit dat behalve HE ook de S eiwitten van deze HCoVs veranderd waren. Het belangrijkste stukje informatie om dit te kunnen beoordelen, de locatie van de RBP van S ontbrak vooralsnog.

Hoofdstuk 3 beschrijft de identificatie van de RBP middels vergelijkende structuuranalyse in combinatie met mutagenese en geautomatiseerde *in silico* 'docking' van het ligand. We laten zien dat de RBP niet alleen aanwezig is in het S eiwit van HCoV-OC43 en van nauw verwante virussen van runderen en varkens, maar ook functioneel geconserveerd in dat van HCoV-HKU1. Bovendien leveren wij bewijs dat de RBP essentieel is voor de infectiviteit van deze virussen middels experimenten met virusdeeltjes van het vesiculaire stomatitis virus, die voorzien waren van ("gepseutotypeerd" met) het BCoV S eiwit.

Hoofdstuk 4 betreft een studie uitgevoerd in samenwerking met de groep van David Veesler (Department of Biochemistry, University of Washington, Seattle, WA, USA). Hierin bevestigen wij de resultaten uit Hoofdstuk 3 door het ophelderen van de 3D structuur van het HCoV-OC43 S eiwit in complex met 9-*O*-Ac-Sia met een resolutie van 2,8 Å met behulp van cryo-electronenmicroscopie. Bovendien presenteren wij de allereerste analyse van de kinetiek, waarmee de S RBP bindt aan zijn ligand, in monovalente 1:1 bindingsassays, gebaseerd op biolayer interferometrie en gebruikmakend van synthetische sialosiden. Wij laten zien dat deze binding uiterst dynamisch is en dat associatie en dissociatie van S-ligand complexen plaatsvinden binnen tienden van seconden.

Hoewel de structuurdata informatief zijn, worden zij slechts relevant wanneer beschouwd in de biologische context. Met de opheldering van de structuur van het S eiwit en met die van HE bekend, was het moment daar om de interactie tussen deze eiwitten tijdens de infectie te bestuderen. In **Hoofdstuk 5** beschrijven wij de ontwikkeling van een 'reverse genetics' methode voor BCoV. Hiervan gebruikmakende werd de RBP van het BCoV HE eiwit (en daarmee de lectine functie van HE) uitgeschakeld. In feite herhaalden wij hiermee de evolutie van HCoV-OC43 onder laboratoriumcondities. Terwijl HCoV-OC43 zich in weefselkweekcellen uitstekend vermeerderd, bleek het uitschakelen van de HE lectine functie in BCoV een bijzonder nadelig effect voor de "fitness" van het virus te hebben. Bij pogingen deze virussen te genereren en te vermeerderen werd er steevast geselecteerd op compenserende mutaties in of nabij de S RBP, die de bindingsaffiniteit van het S eiwit enorm verlaagden. Door sommige van deze mutaties werd de bivalente binding van S1A-Fc fusie eiwitten aan de geclusteerde sialoglycanen op een model mucine-eiwit zelfs meer dan 10.000 maal verminderd. Wij laten zien, opnieuw gebruikmakende van pseudotypering, dat het defect in BCoV tengevolge van verlies van de HE lectine functie te wijten is aan processen die plaatsvinden tijdens de virusaanhechting aan de gastheercel. Onze waarnemingen leveren bewijs voor een functionele balans tussen S-gemedieerde aanhechting en HE-gemedieerde catalytische virus elutie, waarbij er niet alleen selectie is op een optimaal evenwicht tussen associatie en dissociatie, maar ook -binnen gegeven grenzen- voor maximale bindingsaviditeit van het virion. Onze waarnemingen brengen ons tot de conclusie dat de verwerving van het HE eiwit, de embecovirussen op een evolutionair pad heeft gebracht dat uiteindelijk geresulteerd heeft in een mechanisme voor virus-suiker interacties, dat uniek is onder CoVs, maar opmerkelijk sterk gelijkend op dat van influenza A virussen.

In **Hoofdstuk 6** worden ten slotte de resultaten van dit proefschrift samengevat en bediscussieerd, mede in de context van bevindingen die nog niet gepubliceerd zijn.

It was an unusual sunny day in Utrecht, and the clocks were striking twelve. Yifei Lang, the PhD student in self-isolation, is sitting in his foxhole, performing imaginary pipetting. While everything in his apartment stays relatively the same, the outside world had changed, drastically. For the first time in his life, cinemas were closed because life is more dramatic than movies. Looking back in time, he realized he should start writing down things before he forgets: his unchained, mixed and tangled feelings. He took out an old dusty notebook, turned to the front page and sharpened his pencil. *To mark the paper was the decisive act.* In small, clumsy letters, he wrote---

Acknowledgement

Finale. *Fine*. De Einde. 终章。Once and for all I have reached the "most viewed" part of my thesis, and probably one of the few theses that will be read after the promotion. The year 2020 had marked the world's memory with the word "coronavirus". Thanks to everyone listed down below, my journey with coronaviruses started a little earlier than that.

First of all, I would like to express my deepest gratitude to my promoter Prof. Frank van Kuppeveld and my co-promoter Dr. Raoul de Groot.

Frank, I consider you answering my first email as the best thing happened to me in my career. Without that, I would never come here and had this fantastic adventure across a quarter of the globe. To be honest, I contacted you while still doubting my way to science; but after staying in our group I consolidated it as a life-long career and belief. Thank you for your support during these years: work discussions, conferences, and the kind extensions that allowed me to finish.

Raoul, I cannot believe it had been so long already. It seems to me like yesterday we just saw each other via Skype, with Noodle (the cat) in your arms. You surprised me first at Schipol with your Dutch height, then continue surprising me until this very moment you are reading it (and counting) with your scientific spirit and dedication. The sparkle of scientific inspiration lightened up by the day, when the "Caribbean salsa" met the "Chinese Kungfu". Looking back in time, it is hard for someone to believe that how you managed to forge a naïve, reckless boy to the current me. As you know I was not so lucky with getting ready for my holidays, that's because I was destined of putting all my chips on the Wheel of Supervisors. I know how easy it is for others to think and take everything for granted; but I vividly remember every moment that we overcame the hurdles and setbacks, and how I was treated as a family member. To sum up my PhD life in our language (Verdi style): New life welcomed me with the brutal experience of *la donna e mobile*, me being heartbroken like Tosca in *vissi d'arte*; after couples of turns of *Frühlingsstimmen* and *What Power Art Thou,* I started to agree that *L'amour est un oiseau rebelle*, until I found my *una furtiva lagrima; M'appari tutt' amor* and let me understand the essence of *Voi che sapete,* my life turned into the term of *Sempre Libera*. Encounter of wonderful science I always get shaky *Che gelida manina*; and I had been shining under the *O' Sole Mio* of it. During these years instead of me being a lab *Les oiseaux dans la charmille,* we turned to fought so close together as *Di quella pira*--- there had been lots of *Nessun dorma* 's and leaving the lab when it was *E lucevan le stelle*---finally here we come to the *Votre toast, je peux vous le rendre*. (*Deh vieni, non tardar…!*) You are a *o Mio Babbino Caro* to me, it is so hard to face the *Donde lieta usc*ì! (After my leave) I will miss you deeply especially your *Mein Herr Marquis* and let's *Libiamo ne' lieti calici* anytime when we see each other again!

Standing at the crossroad of chemistry, crystallography and virology, I am very grateful to work together with our great collaborators. Professor **Geert-Jan Boons**, it is a blessing for our whole glycovirology society that you started this group in Utrecht. I learned and benefited so much from our collaboration. **Zeshi (Jack)**, my life brightened up after you were airborne by 2016 in Utrecht. Before I met you here I thought you were an ABC (American-born Chinese) based on your English level shown in our corresponding emails; after I got to know you I understand that you are an actual ABC (Amazingly Brilliant Chemist). Tables and chairs in the UMC cafeteria had witnessed so many meaningful and inspiring discussions that we had; trips to Texel and Cologne had marked the friendship we built beyond work and ourselves. I admire your hard-working spirit and also your easy-going personality. Everything happened here in Utrecht is just a start; there will be a lot more in the future. I wish you all the best with finishing your PhD and the next steps.

Eric and Arie, many thanks for bringing me into the reciprocal world! **Eric**, it is my great fortune to serve as a part of this chain of collaboration which already lasted over a decade and still goes strong. We had a lot of fun working together, solving chemical/structural puzzles and putting virtual/real stamps on all those structures. **Arie**, my crystallography mentor, although our masterpieces of crystallography work were not included as chapters of this thesis, but there is no doubt this book wouldn't have its current shape without you. I am constantly touched, encouraged and motivated by your craftsmanship: you work really precise, organized and you are always willing to go the extra mile to all those tiny little details, then compose a whole symphony out of it. I do cherish a lot of both the theoretical and practical training sessions you provided for me, and most of all your trust and friendship. It is such an honor to have you in my PhD journey.

Richard, I am "stormborn" now thanks to your training and the time we spent at the Nikon together. Thanks so much for teaching me all the cool microscopy stuff, and also for your patience with me. Valar morghulis.

During my work at the **Viro** group I have had so many joyful and inspiring times with many colleagues. Many thanks to the "alumni" of team Raoul: **Martijn**, although I never had the privilege to work side by side with you, but I still benefit a lot from the expertise you built: protocols, standards, even RNA/proteins stored in the freezers. I admired a lot of your achievements, and it is a great and historical coincidence to the sialoglycobiology/virology field that you and Raoul teamed up back in time. Same also goes for **Arno**: thanks so much for always help me with my puzzles in the lab, and your positive attitude always encourages me. I would also like to give my highest appreciation to you and **Nancy** on behalf of all the AIO's for your effort to maintain our working platforms. **Mark**, thanks for working me in when I started.

My dear roomies, **Nancy** and **Ivy**, we had made our office the most stable one in years except the offices of the PI's. Thanks so much for tolerating a young boy in the office for so long☺. I enjoyed a lot answering the phone with "I don't speak Dutch".

Lucian, the lighthouse of the lab and my climbing partner, hopefully all is well with you. **Linda**, I am listening to Eva Cassidy right now. Thanks for all the time sitting together doing cell culture and the cheer talks when I am blue. **Huib,** it is so nice to have you around for my whole PhD. Thanks for your help throughout the years and let's see when you wake up and become Prof. Balugawhale. **Dan,** we had a lot of fun with the HKU1 HE project and also the trip to Ireland. I learned quite some structural knowledges from you. **Cristina, Floor** and **Maryam:** I don't think I laugh that much anymore since you

guys left. I miss so much hearing your laughter from my office. And as quoted: *every lab needs a Maryam…*

Peter, Herman, Berend-Jan, Xander, Erik, Jim, Hendrik, Jeroen, Fiona, Hilde, Lisa, Esther, Rachel, Qian, Erion, Ieva, Mirte, Malte, Erhard, Anja, Clasin, Susanne, Kyra, it is my pleasure to meet you here in Utrecht. I wish you all the best in the future.

It is so nice to have a family far away from home. Special thanks goes to all my Chinese colleagues here in the Viro group. **Wentao**, my big brother, after all these years it seems like you are the only close and constant friend that I had in my PhD time. I can't imagine my life without you: all those lab skills you taught me and discussions we had on the daily basis; taking care of me while I were gravely injured; all the fishes we net, all the miles we run and all the racket strings we broke (I only won like 5 rallies against you in 5 years); the icy view of the Arctic Circle and the tropical view of the Canary Islands… I could go on but I am sure I would not like to pay for the printing of the extra 10 pages of this book. *We lucky few, we band of brothers, for he who today sheds his blood with me shall be my brother.* There is a bright future ahead of us where band of brothers will show its strength; please allow me to return your favor of all these years as a life-long friend, and a brother, and a first mate. **Hongbo (and Wenshi),** many thanks for teaching me how to use the Octet ("outlet"?) and for your hospitality in Rotterdam. Also thanks to other Chinese colleagues working and worked in our lab: **Ruisong, Chunhua, Qiushi, Huihui, Meiling, Rui, Lianci, Wenjun, Chunyan, Shan, Yongtao, Xinyi, Mengying, Yongle and Xiaoyao**.

Many thanks to the students I had worked together with: **Arthur, Angelic, Danielle, Gijs, Coco, David, Geerte, Robin** and **Natalia**; the time I spent with you and the experience I have got will be a nice boost to my future career. Special thanks for Arthur *Skywalker* (the force is strong with this one!), Danielle *Gump* (run Dani, run!) and Gijs *Beerinhand* (If one Duvel ain't solving the problem, how about two?): your internships really colored my PhD life!

Special merits to Prof. Jaap Wagenaar: **Jaap**, it is an honor to meet you here in Utrecht. Thanks so much to let us internationals to feel the Dutch culture. I had a lot of fun with the trips we had together.

Outside the lab, it had been a joyful time being a part of the IB graduate school. I really enjoyed the time spent as a member of the IB AIO committee; thanks to **Marjolein**, **Marsa**, **Tania** and **Feline**: it has been a great pleasure together organizing all the IB AIO evenings, IB AIO dags and the retreat.

费尽心血不敢妄言成功，侥幸渡劫不敢奢求眷顾。这个博士读下来实属不易，这里再一次感谢在此期间给过我帮助的人。

文思隽永，怒海听涛；

泽被苍生，春华秋实。

涛哥，岛爷，再一次用荷兰人看不懂的语言向你们表示感谢！哥几个在荷兰这些年都是敢于直面惨淡的课题，敢于正视淋漓的 diss 的真正勇士。一路走过来，大家都经历过挫折的洗礼，也都经历了雨后彩虹的喜悦。以后的路还长，我们最好的时间也刚刚开始。希望我们以后灵心不倦，初心不远，互相勉励的走在科研的羊肠小道上。

感谢在乌村认识的公派同学们：CBDD 的小伙伴们，龙爷，凤姐，抖森，孟爷，燕燕还有你们的家属们，大家聚一起玩的时候的时间总是过得又快又开心。他日江湖相见，五升啤酒走起。UMC 的妹子们，娜娜，玉玺，宇洁，慧颖和常晓，我们差不多一起来到荷兰，到现在都还非常怀念当时一起上英语课和聚餐的日子。感谢一起打羽毛球的小伙伴们：金鑫，玉珑，杨敏，一粒，超哥，栋哥，渭东，德朝，晓彬，刘鑫，天松，肖玲，孙季。和你们一起打球的时间最开心，也成功拖慢了一点点我长胖的步伐。感谢中国好舍友婧姐和张淼，一起在一个屋檐下的日子一去不复返又那么让人怀念。包包，谢谢你邀请我做你的 paranimph，以后你在荷兰的日子要一直开心幸福下去哦。范山和增慧，好久不见甚是想念，希望以后再见你们的时候我的普通话能好点。

感谢四川农业大学曹三杰教授、颜其贵教授、黄小波教授，四川大学华西医院黄伟教授、刘杰教授和杨靖梅博士在职业生涯规划方向上给我的建议和帮助。

吃水不忘挖井人。在这里再次感谢我读研究生期间的导师常惠芸研究员和师兄高闪电副研究员，谢谢你们在我读研期间给我的鼓励、支持和包容。

亲爱的爸爸妈妈，谢谢你们对我的养育和教导。没有你们无私的奉献和支持就没有我的今天。养儿一百岁，常忧九十九。希望现在博士毕业的我，可以让你们放心我以后可以从容面对人生新的征程。老爸，“狼博士”这个外号，我接棒替你戴几年；老妈，很幸运做你的儿子，以后的日子还要多多继续唠叨下去。

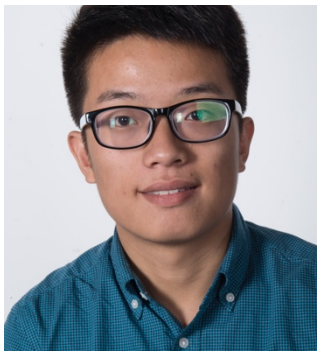
老赵同学，虽然见过了太多写进论文致谢然后翻车的案例，但是我还是鼓起勇气在这里好好感谢你一下。欧洲有几百所学校，我们却都选择了乌村大学；此前没有任何交集，甚至未曾去过同一个城市的我们，却因为在风车国的遇见，彼此认取了未来生活的模样。人生如戏，浮生若梦。在荷兰读书的我们，邂逅已实属不易；也祝愿以后的我们，可以从容面对长相守的考验。“当你知道你想和谁一起度过余生时，你会希望余生开始的越早越好。”He is looking at you, kid.

To myself: it is not the critic who counts.

Yifei Lang （郎一飞）

Utrecht，2020

Curriculum Vitae



Yifei Lang was born on June 30th, 1993 at Liaocheng, Shandong province, P.R. China. He enrolled in No.3 Middle school of Liaocheng in 2003. After receiving his middle school diploma in 2006, he studied his bachelor's program in Biotechnology at the College of Life Sciences, Shandong Agricultural University in Taian, Shandong province. After obtaining his bachelor's degree in 2010, he chose to continue in science and academia by enrolling in the master program of Chinese Academy of Agricultural Sciences. He performed his master research in State key laboratory on Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, China. During his master research he focused on the molecular epidemiology and novel vaccine design of bovine viral diarrhea virus. Since then, his interest grew for virology. He obtained his master's degree in July 2014. In September 2014, as a PhD candidate he began his further research in Virology Division, Department of Infectious Diseases & Immunology, Faculty of Veterinary Medicine, Utrecht University, the Netherlands. His PhD project is partially funded by the China Scholarship Council. Under the supervision of Dr. Raoul J. de Groot and Prof. Dr. Frank J.M. van Kuppeveld, his research focused on the cross-species transmission and host adaptation in embecoviruses.

List of publications

Lang, Y., Li, W., Li, Z., Koerhuis, D., van den Burg, A., Rozemuller, E., Bosch, B.J., van Kuppeveld, F.J.M., Boons, G.J., Huizinga, E.G. van der Schaar, H.M. & de Groot, R.J. (2020) Coronavirus hemagglutinin-esterase and spike proteins co-evolve for functional balance and optimal virion avidity. ***Proceedings of the National Academy of Sciences.*** In revision.

Tortorici, M.A., Walls A. C., **Lang, Y.,** Wang. C, Li, Z, Koerhuis, D., Boons, GJ., Bosch, B.J., Rey, FA., de Groot, R.J. & Veesler,D. (2019). Structural basis for human coronavirus attachment to sialic acid receptors. ***Nature Structural & Molecular Biology***, 26, 481–489

Hulswit, R. J.[#], **Lang, Y.[#],** Bakkers, M. J.[#], Li, W., Li, Z., Schouten, A., Ophorst, B., van Kuppeveld, F.J.M., Boons, G.J., Bosch, B.J., Huizinga, E.G. & de Groot, R.J. (2019). Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. ***Proceedings of the National Academy of Sciences***, 116 (7), 2681-2690 ([#]Co-first author)

Li, W., Hulswit, R. J., Widjaja, I., Raj, V. S., McBride, R., Peng, W., Widagdo, W., Tortorici, M.A., van Dieren, B., **Lang, Y.**, van Lent, J. W., Paulson, J.C., de Haan, C.A.M., de Groot, R.J., van Kuppeveld, F.J.M., Haagmans, B.L & Bosch, B.J. (2017). Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. ***Proceedings of the National Academy of Sciences*** 114(40), E8508-E8517.

Bakkers, M.J.G., **Lang, Y.**, Feitsma, L.J., Hulswit, R.J.G., de Poot, S.A.H., van Vliet, A.L.W., Margine, I., Mijnes- de Groot, J.D.F., van Kuppeveld, F.J.M., Langereis, M.A., Huizinga, E.G. & de Groot, R.J. (2017). Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. ***Cell Host & Microbe*** 21(3), 356-366.

Lang, Y., Gao, S., Du, J., Shao, J., Cong, G., Lin, T., Zhao F., Liu L. & Chang, H. (2014). Polymorphic genetic characterization of E2 gene of bovine viral diarrhea virus in China. ***Vet. Microbiol.*** 174(3), 554-559.

Gao, S., Du, J., Shao, J., **Lang, Y.,** Lin, T., Cong, G., Zhao F., Belák S., Liu L., Chang H. & Yin, H. (2014). Genome analysis reveals a novel genetically divergent subgenotype of bovine viral diarrhea virus in China. ***Infect. Genet. Evol.*** 21, 489.

Gao, S., Luo, J., Du, J., **Lang, Y.,** Cong, G., Shao, J., Lin T, Zhao F., Belák S., Liu L., Chang H. & Yin H. (2013). Serological and molecular evidence for natural infection of Bactrian camels with multiple subgenotypes of bovine viral diarrhea virus in Western China. ***Vet. Microbiol.*** 163(1), 172-176.

Oral Presentation

November 3-4, 2016, IB Conference on Biomembranes, Utrecht, The Netherlands, “Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity”

June 4-9, 2017, XIVth International Nidovirus Symposium (Nido2017), Kansas City, Kansas, United States, “Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity”

March 8, 2019, Dutch Annual Virology Symposium, Amsterdam, The Netherlands, “Betacoronavirus S and HE proteins are in functional balance; evidence for convergent evolution of betacorona- and influenzaviruses”

June 8-14, 2019, The Keystone Symposia conference on positive-strand RNA viruses, Killarney, Ireland, “Betacoronavirus S and HE proteins are in functional balance; evidence for convergent evolution of betacorona- and influenzaviruses”

Poster Presentation

May 2-4, 2018, 1st International Symposium on Glycovirology, Schöntal, Germany, “An atomic level explanation for ligand/substrate fine specificity of nidovirus hemagglutinin-esterase proteins”