nature structural & molecular biology

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]*

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.
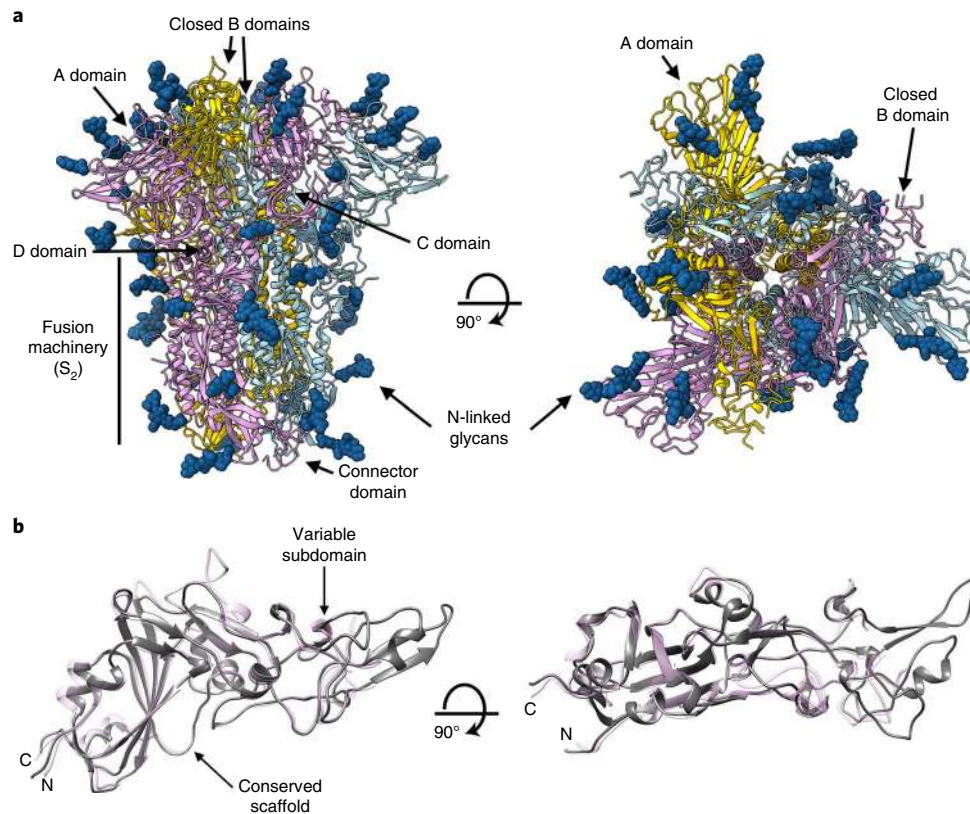
Coronaviruses are large, positive-sense enveloped RNA viruses in the Nidovirales order and are divided into four genera: α, β, γ and δ. Two β-coronaviruses have caused outbreaks of deadly pneumonia in humans since the beginning of the 21st century. The severe acute respiratory syndrome coronavirus (SARS-CoV) 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 S_1 and S_2, which remain non-covalently bound in the pre-fusion conformation. The N-terminal S_1 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 S_2 subunit is the metastable spring-loaded fusion machinery[9]. During entry, S_2 is further proteolytically cleaved at the S_2' 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

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.

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 S_1/S_2 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. 1a,b and Table 1). The S_1 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 S_2 subunit, which is more conserved than the S_1 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

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)
Data collection and processing		
Magnification	47,619	47,619
Voltage (kV)	300	300
Electron exposure (e⁻/Å²)	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 (Å²)	−61	−70
Model composition		
Nonhydrogen atoms	27,477	
Protein residues	3,519	3,519
Ligands	0	3
Waters	186	396
B factors (Å²)		
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

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].

Domain B shows the highest variability within S₁ 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 protein[53] (VP8*) to recognize their respective ligands (Supplementary Fig. 2a–c).

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 Å² 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.
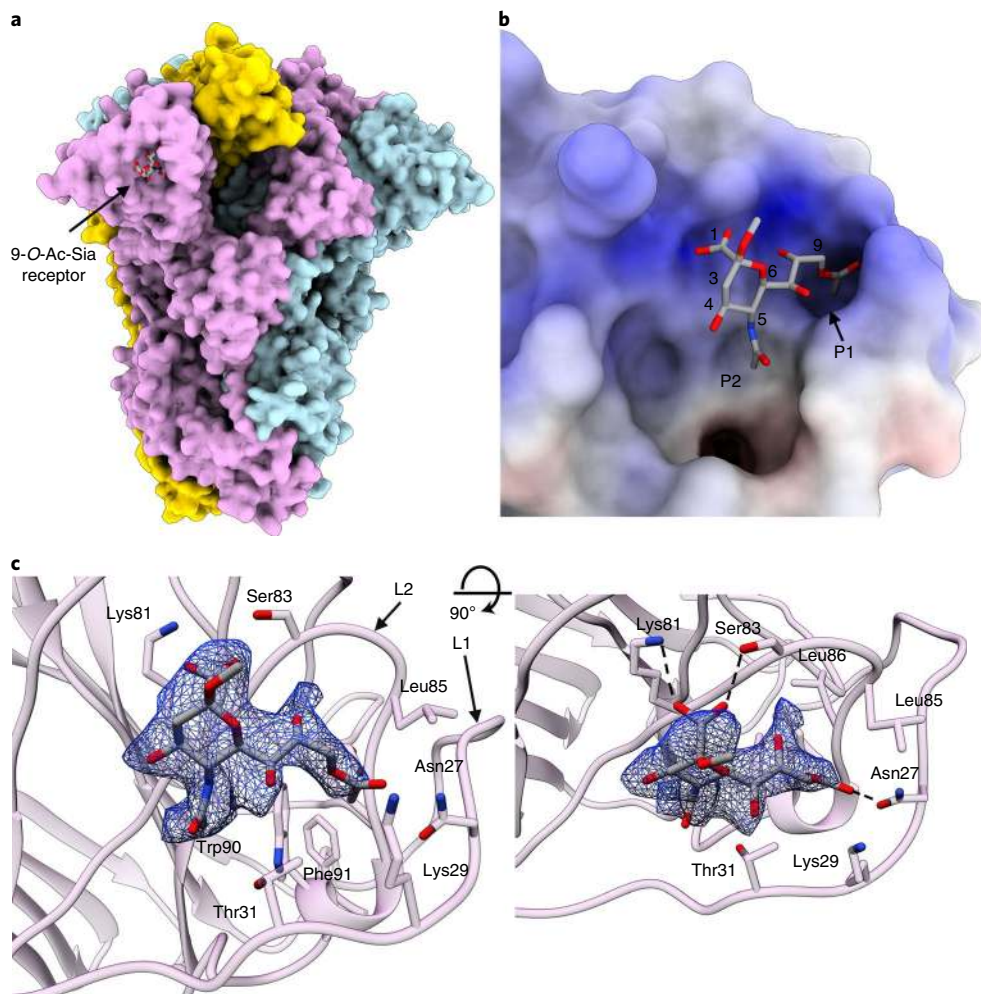
Fig. 2 | Identification of a sialoglycan-binding site in the holo-HCoV-OC43 S glycoprotein structure. a, Molecular surface representation of the holo-HCoV-OC43 S ectodomain trimer structure with the bound ligand shown as sticks. Protomers are individually colored. **b**, Surface representation of the ligand-binding site colored by electrostatic potential from −12 to +12 k_BT/e_c. **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.

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].

Using steady-state analysis, we determined an equilibrium dissociation constant $K_D = 49.7 \pm 10.7\,\mu M$ 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 $k_{on} = 1.4 \times 10^4\,M^{-1}s^{-1}$. 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].
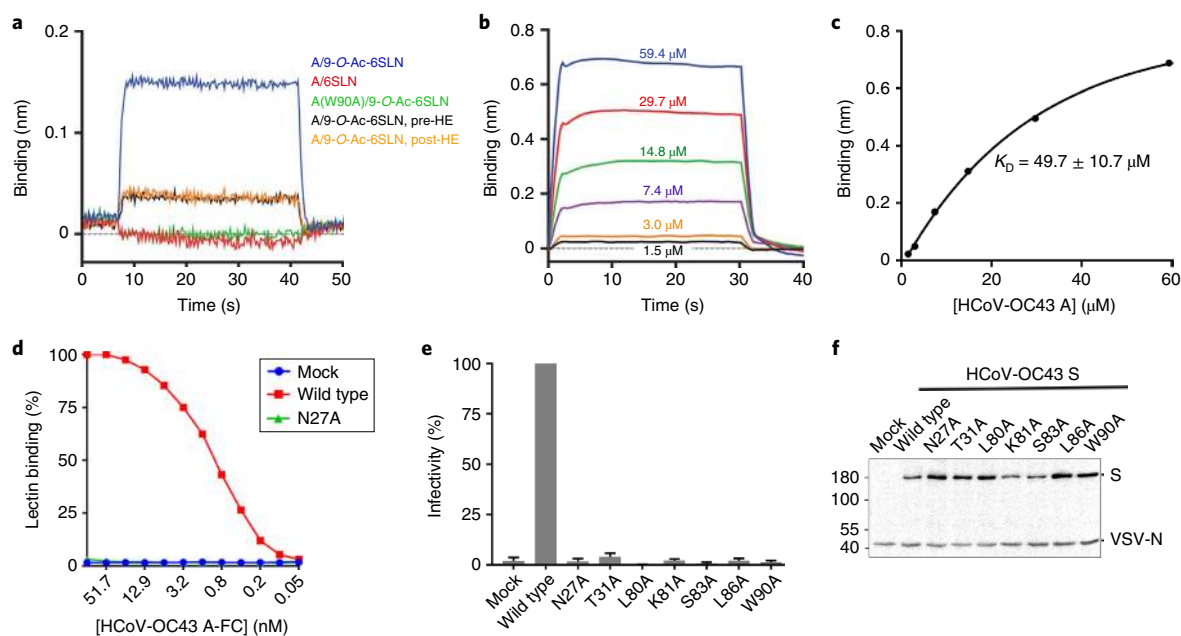
Fig. 3 | The identified HCoV-OC43 S interactions with sialosides are characterized by fast kinetics and are required for viral entry. a, Biolayer interferometry showing binding of wild-type or W90A monomeric HCoV-OC43 domain A to immobilized 6-sialyl-5-N-acetyl,9-O-acetyl-lactosamine (9-O-Ac-6SLN), 6-sialyl-5-N-acetyl-lactosamine (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 $K_D = 49.7 \pm 10.7\,\mu M$. **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.

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 S_1/S_2 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 μg•ml^{-1} (w/v), to recapitulate proteolytic priming[13], led to cleavage at the S_1–S_2 boundary, as observed via SDS-PAGE (Supplementary Fig. 4a). Incubation with 28 μg•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 S_2' 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
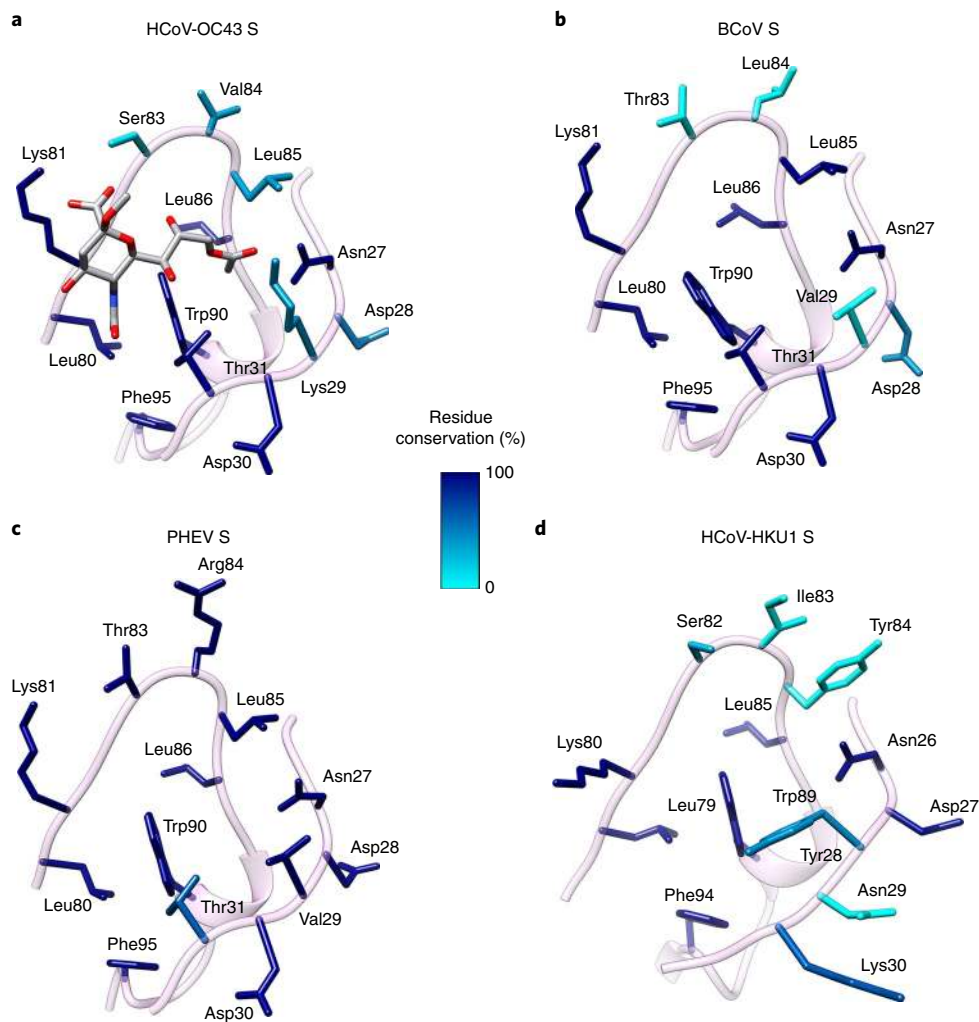
Fig. 4 | Conservation of the receptor-binding groove among all 9-O-Ac-sialoglycan-recognizing coronaviruses. a–d, Zoomed-in view of the binding sites rendered as ribbon diagrams with surrounding residues shown as sticks for HCoV-OC43 (**a**), BCoV (**b**), PHEV (**c**), HCoV-HKU1 (**d**). Residues are 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.

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. 4g,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$_{HCoV-OC43}$, 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
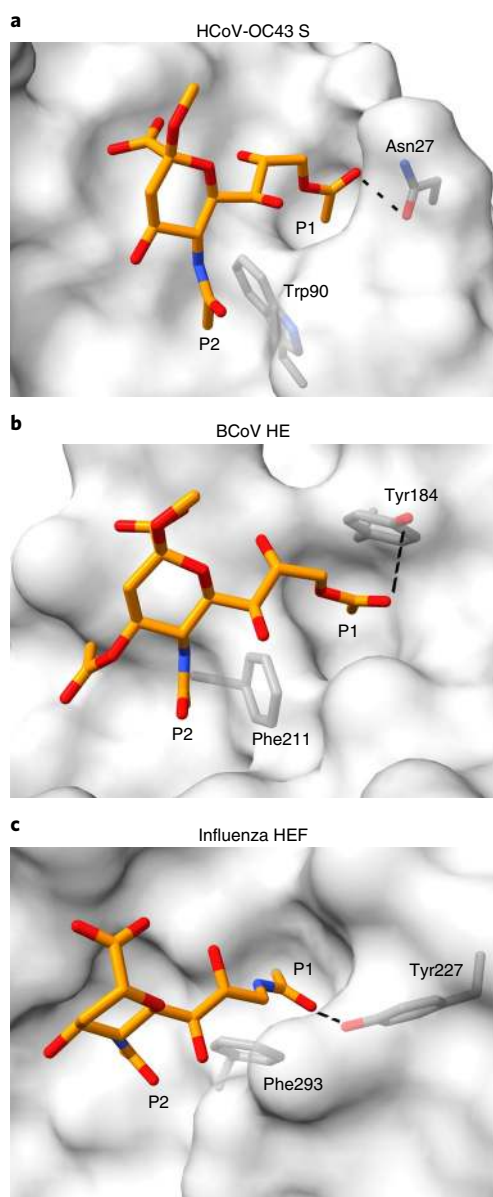
Fig. 5 | Conservation of the receptor-binding site architecture across coronavirus S, coronavirus HE and influenza virus HEF glycoproteins.
a, HCoV-OC43 S bound to 9-O-Ac-Me-Sia. **b**, BCoV HE bound to 5-N-acetyl-4,9-di-O-acetyl-neuraminic acid α-methylglycoside (PDB 3CL5). **c**, Influenza virus C 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.

(Leu80$_{HCoV-OC43}$), Lys80$_{HCoV-HKU1}$ (Lys81$_{HCoV-OC43}$), Leu85$_{HCoV-HKU1}$ (Leu86$_{HCoV-OC43}$), Trp89$_{HCoV-HKU1}$ (Trp90$_{HCoV-OC43}$) and Phe94$_{HCoV-HKU1}$ (Phe95$_{HCoV-HKU1}$) (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$_{HCoV-HKU1}$, Thr82$_{HCoV-HKU1}$ (Ser83$_{HCoV-OC43}$) or Trp89$_{HCoV-HKU1}$ 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.

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

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.

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$_{HCoV-HKU1}$ (Asn27$_{HCoV-OC43}$), Leu79$_{HCoV-HKU1}$

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 Å²) 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41594-019-0233-y.

References

1. Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
2. Haagmans, B. L. et al. Middle east respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect. Dis.* **14**, 140–145 (2014).
3. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **24**, 490–502 (2016).
4. Isaacs, D., Flowers, D., Clarke, J. R., Valman, H. B. & MacNaughton, M. R. Epidemiology of coronavirus respiratory infections. *Arch. Dis. Child* **58**, 500–503 (1983).
5. Menachery, V. D. et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015).
6. Menachery, V. D. et al. SARS-like WIV1-CoV poised for human emergence. *Proc. Natl Acad. Sci. USA* **113**, 3048–3053 (2016).
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.* **13**, e1006698 (2017).
8. Bosch, B. J., van der Zee, R., de Haan, C. A. & Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**, 8801–8811 (2003).
9. Walls, A. C. et al. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114–117 (2016).
10. Millet, J. K. & Whittaker, G. R. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015).
11. Belouzard, S., Chu, V. C. & Whittaker, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl Acad. Sci. USA* **106**, 5871–5876 (2009).
12. Millet, J. K. & Whittaker, G. R. Host cell entry of middle east respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl Acad. Sci. USA* **111**, 15214–15219 (2014).
13. Walls, A. C. et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl Acad. Sci.* **114**, 11157–11162 (2017).
14. Burkard, C. et al. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* **10**, e1004502 (2014).
15. Walls, A. et al. Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy. *Protein Sci.* **26**, 113–121 (2017).
16. Walls, A. C. et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
17. Xiong, X. et al. Glycan shield and fusion activation of a deltacoronavirus spike glycoprotein fine-tuned for enteric infections. *J. Virol.* **92**, e01628-17 (2018).
18. Kirchdoerfer, R. N. et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* **531**, 118–121 (2016).
19. Kirchdoerfer, R. N. et al. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.* **8**, 15701 (2018).
20. Pallesen, J. et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl Acad. Sci. USA* **114**, E7348–E7357 (2017).
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.* **27**, 119–129 (2017).
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.* **14**, e1007009 (2018).
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.* **14**, e1007236 (2018).
24. Shang, J.et al. Cryo-EM structure of porcine delta coronavirus spike protein in the pre-fusion state. *J. Virol.* https://doi.org/10.1128/JVI.01556-17 (2017).
25. Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion.*Cell* **176**, 1026–1039.e15 (2019).
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.* **79**, 1595–1604 (2005).
27. Lau, S. K. 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.* **85**, 11325–11337 (2011).
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* **85**, 4526–4529 (1988).
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.* **89**, 7202–7213 (2015).
30. de Groot, R. J. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* **23**, 59–72 (2006).
31. 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.* **87**, 3097–3107 (2013).
32. Bakkers, M. J. et al. Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. *Cell Host Microbe* **21**, 356–366 (2017).
33. Zeng, Q., Langereis, M. A., van Vliet, A. L., Huizinga, E. G. & de Groot, R. J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl Acad. Sci. USA* **105**, 9065–9069 (2008).
34. Rosenthal, P. B. et al. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **396**, 92–96 (1998).
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.* **12**, e1005411 (2016).
36. Herrler, G. et al. The receptor-destroying enzyme of influenza C virus is neuraminate-O-acetylesterase. *EMBO J.* **4**, 1503–1506 (1985).

37. 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* **12**, 739–749 (2014).
38. Neu, U., Bauer, J. & Stehle, T. Viruses and sialic acids: rules of engagement. *Curr. Opin. Struct. Biol.* **21**, 610–618 (2011).
39. Schauer, R. & Kamerling, J. P. Exploration of the sialic acid world. *Adv. Carbohydr. Chem. Biochem* **75**, 1–213 (2018).
40. Peng, G. et al. Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **287**, 41931–41938 (2012).
41. Peng, G. et al. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl Acad. Sci. USA* **108**, 10696–10701 (2011).
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* **114**, E8508–E8517 (2017).
43. 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. USA* **116**, 2681–2690 (2019).
44. Collins, A. R. HLA class I antigen serves as a receptor for human coronavirus OC43. *Immunol. Invest* **22**, 95–103 (1993).
45. Langereis, M. A. et al. Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. *Cell Rep.* **11**, 1966–1978 (2015).
46. Bakkers, M. J. et al. Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *Proc. Natl Acad. Sci. USA* **113**, E3111–E3119 (2016).
47. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
48. Lu, G. et al. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).
49. Ou, X. et al. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nat. Commun.* **8**, 15216 (2017).
50. Yuan, Y. et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).
51. Campbell, M. G., Veesler, D., Cheng, A., Potter, C. S. & Carragher, B. A resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *eLife* **4**, e06380 (2015).
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.* **273**, 13047–13052 (1998).
53. Dormitzer, P. R., Sun, Z. Y., Wagner, G. & Harrison, S. C. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21**, 885–897 (2002).
54. Langereis, M. A., Zeng, Q., Heesters, B. A., 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.* **8**, e1002492 (2012).
55. Kunkel, F. & Herrler, G. Structural and functional analysis of the surface protein of human coronavirus OC43. *Virology* **195**, 195–202 (1993).
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* **5**, 1480–1498 (2015).
57. Xiong, X. et al. Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **497**, 392–396 (2013).
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.* **14**, e1007233 (2018).
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.* https://doi.org/10.1128/JVI.01522-17 (2018).
60. Sakai, T., Nishimura, S. I., Naito, T. & Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Sci. Rep.* **7**, 45043 (2017).
61. Mukherjee, S. et al. Mechanism and significance of cell type-dependent neutralization of flaviviruses. *J. Virol.* **88**, 7210–7220 (2014).
62. Park, J. E. et al. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc. Natl Acad. Sci. USA* **113**, 12262–12267 (2016).
63. Owczarek, K. et al. Early events during human coronavirus OC43 entry to the cell. *Sci. Rep.* **8**, 7124 (2018).
64. Wickramasinghe, I. N., de Vries, R. P., Grone, A., de Haan, C. A. & Verheije, M. H. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J. Virol.* **85**, 8903–8912 (2011).
65. Liu, C. et al. Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. *J. Virol.* **89**, 6121–6125 (2015).
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.* **70**, 5634–5637 (1996).
67. Nilsson, E. C. et al. The GD1a glycan is a cellular receptor for adenoviruses causing epidemic keratoconjunctivitis. *Nat. Med* **17**, 105–109 (2011).
68. Reiter, D. M. et al. Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. *PLoS Pathog.* **7**, e1002166 (2011).
69. Ito, T. et al. Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. *Virology* **227**, 493–499 (1997).

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.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0233-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.V.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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% CO_2 incubator rotating at 130 r.p.m. Wild-type or mutant HCoV-OC43 S ectodomain construct was 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 software[71] 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 Leginon[71] 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 cryoSPARC[75] 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 criterion[76], 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 Rosetta[80–83]. 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 S1A-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, K_D, 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 described[43]. 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).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The cryo-EM maps and atomic models have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank with accession codes EMD-0557 and PDB ID 6NZK (holo-HCoV-OC43 S) and EMD-20070 and PDB ID 6OHW (apo-HCoV-OC43 S).

References

70. Raj, V. S. et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495**, 251–254 (2013).

71. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).

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. *eLife* **7**, e42166 (2018).

74. Zivanov, J., Nakane, T. & Scheres, S.H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

75. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

76. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

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* **135**, 24–35 (2013).

78. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).

79. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

80. Wang, R. Y. et al. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *eLife* **5**, e17219 (2016).

81. Frenz, B. D. R. et al. Automatically fixing errors in glycoprotein structures with Rosetta. *Structure* **27**, 1–6 (2019).

82. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & Andre, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6**, e20450 (2011).

83. DiMaio, F. et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365 (2015).

84. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D. Biol. Crystallogr* **66**, 12–21 (2010).

85. Agirre, J. et al. Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834 (2015).

86. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

87. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

88. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665–W667 (2004).

89. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).

90. Kaname, Y. et al. Acquisition of complement resistance through incorporation of CD55/decay-accelerating factor into viral particles bearing baculovirus GP64. *J. Virol.* **84**, 3210–3219 (2010).

nature research

Corresponding author(s):	David Veesler
Last updated by author(s):	Apr 16, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see <u>Authors & Referees</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

Software and code

Policy information about <u>availability of computer code</u>

Data collection	N/A
Data analysis	N/A

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research <u>guidelines for submitting code & software</u> for further information.

Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The cryoEM maps and atomic models have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank with accession codes EMD-0557 and PDB-6NZK (holo-HCoV-OC43 S), and EMD-20070 and PDB-6OHW (apo-HCoV-OC43 S).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☐	☒ Eukaryotic cell lines
☒	Palaeontology
☒	Animals and other organisms
☒	Human research participants
☒	Clinical data

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Life technologies HEK293F, ATCC HEK293T, ATCC HRT18
Authentication	None of the cell lines used were authenticated
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	N/A