

Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer

Alexandra C. Walls^{1*}, M. Alejandra Tortorici^{2,3*}, Berend-Jan Bosch^{4*}, Brandon Frenz¹, Peter J. M. Rottier⁴, Frank DiMaio¹, Félix A. Rey^{2,3} & David Veesler¹

The tremendous pandemic potential of coronaviruses was demonstrated twice in the past few decades by two global outbreaks of deadly pneumonia. Entry of coronaviruses into cells is mediated by the transmembrane spike glycoprotein S, which forms a trimer carrying receptor-binding and membrane fusion functions¹. S also contains the principal antigenic determinants and is the target of neutralizing antibodies. Here we present the structure of a mouse coronavirus S trimer ectodomain determined at 4.0 Å resolution by single particle cryo-electron microscopy. It reveals the metastable pre-fusion architecture of S and highlights key interactions stabilizing it. The structure shares a common core with paramyxovirus F proteins^{2,3}, implicating mechanistic similarities and an evolutionary connection between these viral fusion proteins. The accessibility of the highly conserved fusion peptide at the periphery of the trimer indicates potential vaccinology strategies to elicit broadly neutralizing antibodies against coronaviruses. Finally, comparison with crystal structures of human coronavirus S domains allows rationalization of the molecular basis for species specificity based on the use of spatially contiguous but distinct domains.

Coronaviruses are enveloped viruses responsible for 30% of mild respiratory infections and atypical pneumonia in humans worldwide⁴. The emergence of the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and of the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 demonstrated that these zoonotic viruses can transmit to humans from various animal species, and suggested that additional emergence events are likely to occur. The fatality rate of SARS-CoV and MERS-CoV infections are about 10–37%^{1,4} and there are no approved antiviral treatments or vaccines.

Coronaviruses use S homotrimers to promote cell attachment and fusion of the viral and host membranes. S determines host range, cell tropism and is the main target of neutralizing antibodies during infection¹. S is a class I viral fusion protein synthesized as a single chain precursor of about 1,300 amino acids that trimerizes upon folding. It is composed of an amino-terminal S₁ subunit, containing the receptor-binding domain, and a carboxy-terminal S₂ subunit, driving membrane fusion. Cleavage by furin-like host proteases at the junction between S₁ and S₂ (S₂ cleavage site) occurs during biogenesis for some coronaviruses such as mouse hepatitis virus (MHV, the prototypical and best-studied coronavirus)^{1,5}. The S₁ and S₂ subunits remain non-covalently associated in the metastable pre-fusion S trimer. After virion uptake by target cells, a second cleavage is mediated by endo-lysosomal proteases (S₂' cleavage site), allowing fusion activation of coronavirus S proteins⁶.

Crystal structures of coronavirus S post-fusion cores demonstrated that the fusogenic conformational changes lead to the formation of a so-called trimer of hairpins that is the hallmark of class I fusion proteins^{7–10}. These structures contain two heptad-repeat (HR) regions

present in S₂ assembled as an extended triple helical coiled-coil motif (HR1) surrounded by three shorter helices (HR2). Crystal structures of several coronavirus S receptor-binding domains in complex with their cognate receptors have also been reported^{11–14}. Finally, cryo-electron microscopy (cryoEM) of SARS-CoV virions provided a snapshot of the S glycoprotein at 16 Å resolution¹⁵. The lack of high-resolution data for any coronavirus S trimer has prevented a detailed analysis of the infection mechanisms.

We produced an MHV S ectodomain trimer with enhanced stability by mutating the S₂ cleavage site and fusing a GCN4 trimerization motif at the C-terminal end of the construct. The resulting MHV S ectodomain forms a trimer binding with high-affinity to the soluble mouse CEACAM1a receptor (Extended Data Fig. 1a, b). We used state-of-the-art cryoEM¹⁶ to determine the structure of the MHV S ectodomain trimer at 4.0 Å resolution (Fig. 1a–c and Extended Data Figs 2 and 3). We fitted the crystal structures of two S₁ domains^{11,13,17} and built *de novo* the rest of the polypeptide chain using Coot¹⁸ and Rosetta^{19,20} (Fig. 1d–f, Extended Data Figs 2–4 and Supplementary Tables 1 and 2). The final model includes residues 15 to 1118, with an internal break corresponding to a loop immediately upstream from the S₂' cleavage site (residues 827–863). The region connecting the S₁ and S₂ subunits (residues 718–754) features weak density that correlates with its accessibility for proteolytic cleavage *in vivo*. Residues 453–535 were modelled by density-guided homology modelling using Rosetta owing to the poor quality of the density in this region (Extended Data Fig. 3k).

The MHV S ectodomain is a 140 Å long trimer with a triangular cross-section varying in diameter from 70 Å, at the membrane proximal base, to 140 Å at the membrane distal end (Fig. 1d, e). The structure comprises two functional subunits (Fig. 2a–d): a distal moiety constituted by the S₁ subunits; and a central stem connecting to the viral membrane formed by the S₂ subunits.

The S₁ subunit has a 'V' shape contributing to the overall triangular appearance of the S trimer (Extended Data Fig. 5a). The S₁ N-terminal moiety comprises domain A, which is folded as a galectin-like β-sandwich decorated with extended loops on the viral membrane distal side, and a three-stranded antiparallel β-sheet plus an α-helix on the viral membrane proximal side. The S₁ C-terminal half folds as three spatially distinct β-rich domains, termed B, C and D (Fig. 2a–d).

The S₂ subunit connects to the viral membrane and is characterized by the presence of long α-helices (Figs 2b–d and 3a). A central helix (α₃₀) stretches 75 Å along the three-fold molecular axis towards the viral membrane (Fig. 3a). It is located immediately downstream of the HR1 motif, which folds as four consecutive α-helices (α₂₆–α₂₉; Fig. 3a and Extended Data Fig. 6a, b), in sharp contrast to the 120-Å long HR1 helix observed in the post-fusion S structures^{7–9} (Extended Data Fig. 6c–e). The 55-Å-long upstream helix (α₂₀), so named because it is located immediately upstream of the S₂' cleavage site, runs parallel to and is zipped against the central helix via hydrophobic contacts

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Institut Pasteur, Unité de Virologie Structurale, 75015 Paris, France. ³CNRS UMR 3569 Virologie, 75015 Paris, France. ⁴Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands.

*These authors contributed equally to this work.

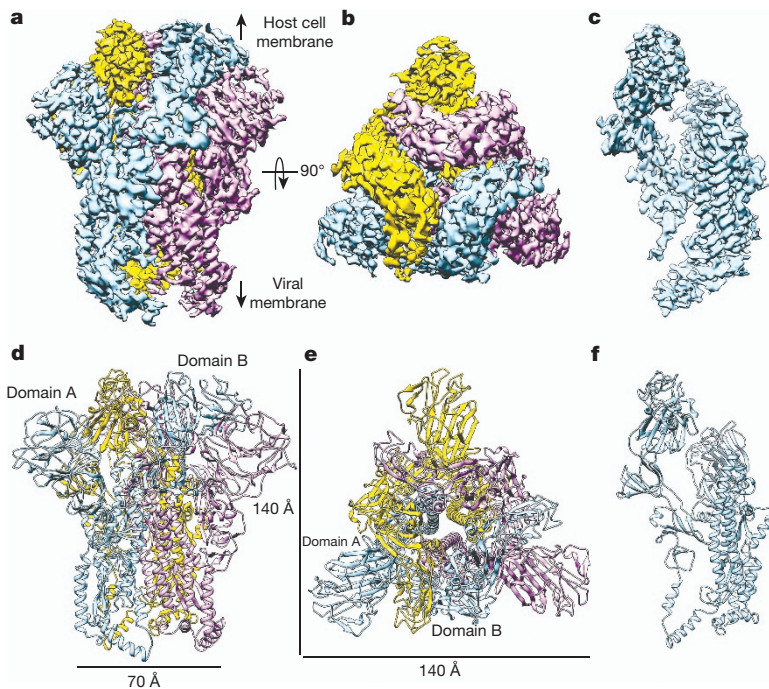


Figure 1 | 3D reconstruction of the MHV S trimer determined by single-particle cryoEM. **a–c**, 3D map filtered at 4.0 Å resolution coloured by protomer. Two different views of the S trimer (from the side (**a**) and from the top, looking towards the viral membrane (**b**)), and a side view of one S protomer (**c**) are shown. **d–f**, Ribbon diagrams showing the MHV S atomic model oriented as in **a–c**.

largely following a heptad-repeat pattern. A core antiparallel β -sheet (β_{46} – β_{49} – β_{50}) is present at the viral membrane proximal end and is assembled from an N-terminal β -strand (β_{46}), preceding the upstream helix, and a C-terminal β -hairpin (β_{49} – β_{50}), located downstream of the central helix.

MHV S_2 features a topology similar to the paramyxovirus F proteins (such as respiratory syncytial virus (RSV) F: root mean squared deviation (r.m.s.d.) 4 Å over 125 residues), with a comparable 3D organization of the core β -sheet, the upstream helix and the central helix (Fig. 3a, b). Importantly, these motifs were shown to remain invariant in the pre- and post-fusion F structures^{2,3}. The conservation of these motifs among coronavirus S and paramyxovirus F proteins suggests that these fusion proteins have evolved from a distant common ancestor. Although the density is too weak to trace the polypeptide chain downstream from β_{50} , secondary structure predictions suggest that the domain directly preceding HR2 could adopt a similar fold in coronavirus S and paramyxovirus F proteins.

In the S trimer, the three central helices are packed via their central portions whereas the two ends splay away from the three-fold axis (Extended Data Fig. 7a–c). Additional contacts between the

upstream and central helices participate to inter-protomer interactions. Furthermore, the S_1 subunits interlock to form a crown around the S_2 trimer stabilizing it in the pre-fusion conformation (Fig. 3c, d and Supplementary Table 3). This is illustrated by the large surface area buried at the interface between each S_1 subunit and the S_2 subunits of the three protomers (1,970 Å²). Many of these contacts involve the HR1 helices and the fusion peptide region. These polypeptide segments undergo major refolding during the fusogenic conformational changes (Extended Data Fig. 6a–e), which supports the notion that the S_1 subunits maintain the S_2 fusion machinery in its metastable state. Substitutions of the conserved alanine 994 by valine in helix α_{28} or of the conserved leucine 1062 by phenylalanine in the central helix were shown to attenuate fusogenicity^{21,22}. Our structure suggests that the former substitution would strengthen hydrophobic packing against the core β -sheet (Extended Data Fig. 7b), and that the later substitution could reinforce molecular stapling of the central helices (Extended Data Fig. 7a, c). The expected modification of the energy landscape between pre-fusion and post-fusion conformations would explain the reduction in fusion activity of these mutants^{21,22}.

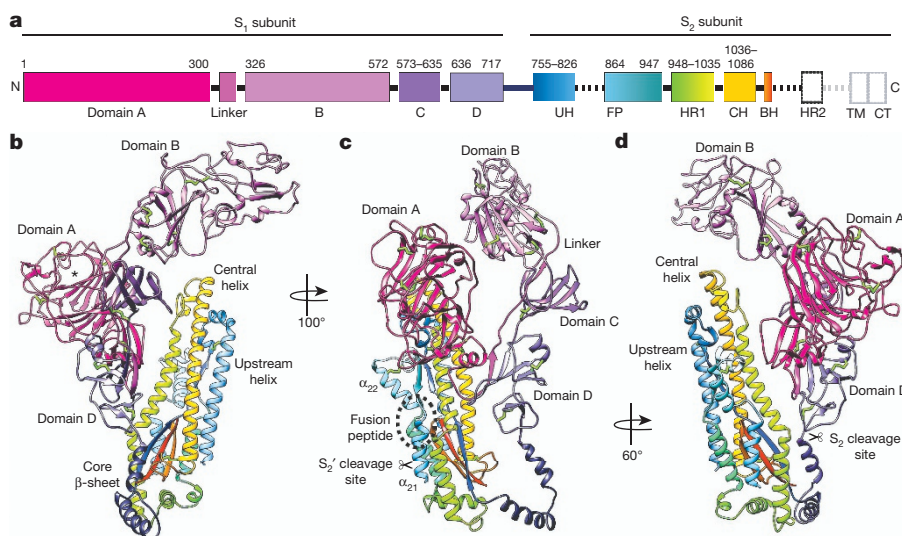


Figure 2 | Architecture of the MHV S protomer.

a, Schematic diagram of the S glycoprotein organization. Black and grey dashed lines denote regions unresolved in the reconstruction and regions that were not part of the construct, respectively. BH, β -hairpin (β_{49} – β_{50}); CH, central helix; CT, cytoplasmic tail; FP, fusion peptide; HR1/HR2, heptad-repeats; TM, transmembrane domain; UH, upstream helix. **b–d**, Ribbon diagrams depicting three views of the S protomer coloured as in **a**. Asterisk denotes the MHV S receptor-binding region. Disulfide bonds are shown as green sticks except for residues 453–535, for which they are not shown.

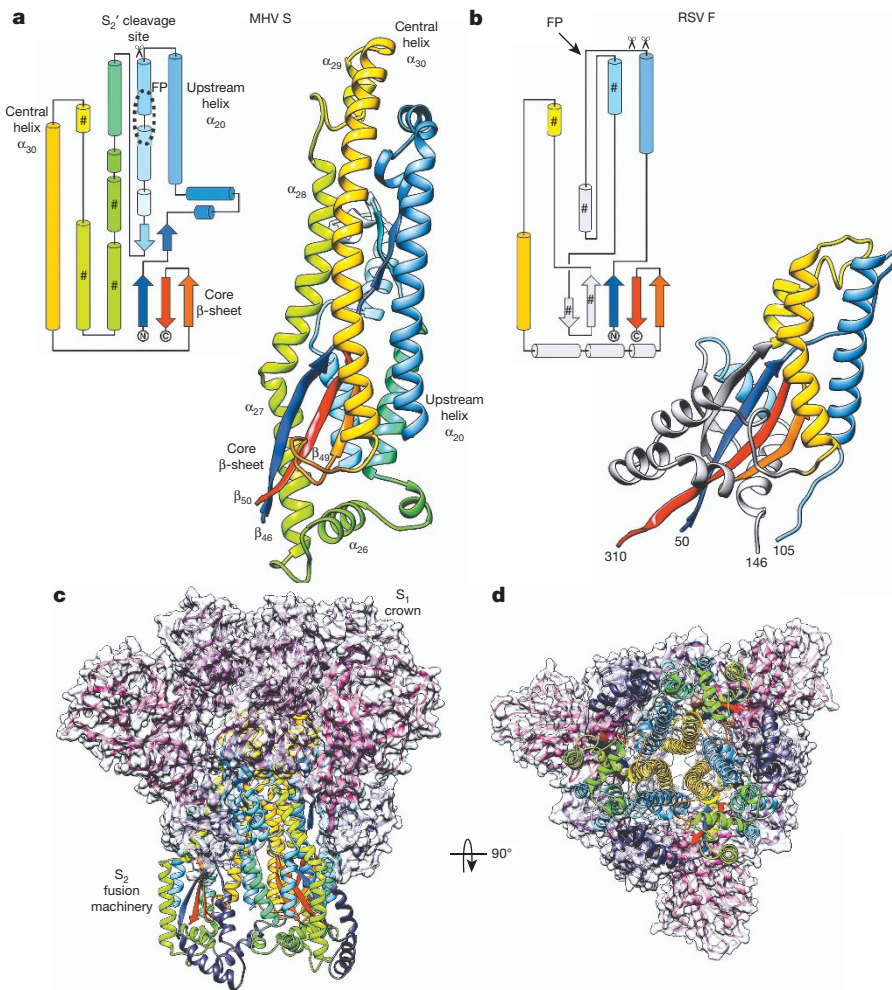


Figure 3 | Pre-fusion structure of the coronavirus fusion machinery. **a, b**, Topology and ribbon diagrams showing the structural similarity between coronavirus MHV S₂ (starting at residue 755) (**a**) and paramyxovirus RSV F (PDB 5C6B) (**b**). For clarity, only part of RSV F is shown, with conserved secondary structural elements coloured identically as for MHV S₂. ‘#’ denotes motifs participating to the post-fusion HR1 coiled-coil. **c, d**, Two different views of the MHV S trimer (from the side (**c**) and top, looking towards the host cell membrane (**d**)) highlighting how S₁ (ribbon diagram and semi-transparent surface) wraps around the S₂ fusion machinery (ribbon diagram) to stabilize it.

The predicted fusion peptide includes the C-terminal half of helix α_{21} and extends up to the N-terminal half of α_{22} (refs 6 and 23) (Fig. 2c). α_{21} is an amphipathic helix located at the periphery of the S trimer, burying hydrophobic side chains towards the S₂ centre and exposing charged residues to solvent (Fig. 2c and Extended Data Fig. 7b, c). In the case of porcine epidemic diarrhoea coronavirus, trypsin processing at the S₂' site can only occur after host cell attachment²⁴. This indicates that receptor binding could allosterically increase the accessibility of the S₂' site, which is located within helix α_{21} . The acidic pH of the endolysosomes could also contribute to exposing the S₂' cleavage site for coronaviruses requiring cleavage in this compartment. The fact that helix α_{21} appears dynamic and is found immediately downstream from

a disordered loop suggests that it could undergo considerable ‘breathing’ motions. Regardless of the mechanism promoting cleavage, the MHV S structure reported here explains the requirement for processing at the S₂' site, as it frees the fusion peptide from the S₂ N-terminal region, which is a prerequisite for its insertion ~200 Å away in the target membrane. The peripheral position of the fusion peptide is similar to what has been observed in the parainfluenza virus 5 F³ and HIV gp41 (ref. 25) prefusion structures (Extended Data Fig. 8a–c). The notable accessibility of the fusion peptide and its sequence conservation among coronaviruses^{6,23} suggest that it would be an ideal target for epitope-focused vaccinology initiatives aimed at raising broadly neutralizing antibodies against S glycoproteins (Fig. 4a–c and Extended Data Fig. 9). Major

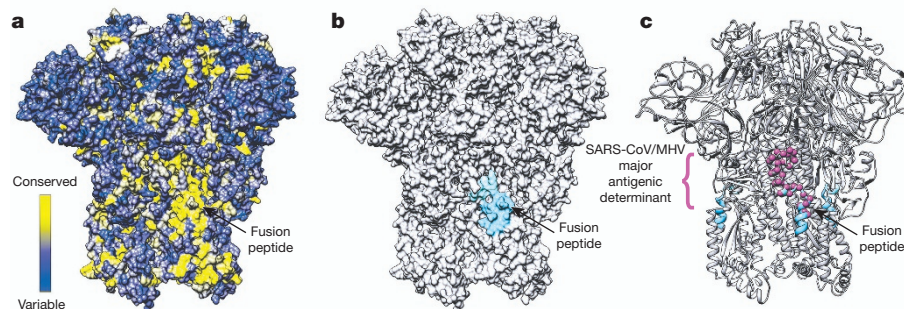


Figure 4 | Potential strategy for neutralizing coronavirus infections. **a**, Surface representation of the MHV S trimer coloured according to sequence conservation using the alignment presented in Extended Data Fig. 9. The fusion peptide sequence is highly conserved among coronavirus S proteins. **b**, Surface representation of the MHV S trimer

highlighting the peripheral position of the fusion peptide (blue and cyan). **c**, Ribbon diagrams of the MHV S trimer showing the overlapping positions of the fusion peptide (residues 870–887, blue and cyan) and of a major antigenic determinant identified for MHV and SARS-CoV (residues 875–905, magenta spheres).

antigenic determinants (inducing neutralizing antibodies) of MHV and SARS-CoV S proteins overlap with the fusion peptide region and support the suitability of this approach^{26,27}. Antibodies binding to this site will not only hinder insertion of the fusion peptide into the target membrane, but will also putatively prevent fusogenic conformational changes. This epitope-focused strategy has proven successful to obtain neutralizing antibodies against RSV F²⁸.

The spatial proximity of domains A and B in the S trimer allows rationalization of their alternative use among coronaviruses to interact with host receptors. MHV uses the viral membrane distal loops decorating domain A to interact with CEACAM1a (ref. 13), whereas MERS-CoV and SARS-CoV rely on the β -motif protruding from domain B to bind to DPP4 (ref. 11) or ACE2 (refs 12 and 14), respectively (Extended Data Fig. 5a–d). The poor sequence conservation of the B domain β -motif among coronavirus S proteins, its considerable length variation among MHV strains (Extended Data Fig. 9) and our density-guided homology model of this motif indicate structural and functional differences. These structural variations constitute the molecular basis underlying coronavirus species specificity and cell tropism using a single S architectural scaffold.

Sequence comparisons indicate that the MHV spike S₁ and S₂ subunits respectively share ~25% and ~40% sequence similarity with many other coronavirus S proteins (Extended Data Fig. 9). Therefore, the structure reported here is representative of the architecture of other coronavirus S such as those of MERS-CoV and SARS-CoV. This hypothesis is further supported by the structural similarity of (1) the MHV¹³ and bovine coronavirus¹⁷ A domains; (2) the MHV, MERS-CoV¹¹, SARS-CoV¹² and HKU4 (ref. 29) B domains (Extended Data Fig. 10); (3) the post-fusion cores of MHV⁷, SARS-CoV^{8,10} and MERS-CoV⁹; and (4) the isolation of infectious coronaviruses featuring a deletion of the A domain and using domain B as the receptor-binding domain³⁰. Our results now provide a framework to understand coronavirus entry and suggest ways for preventing or treating future coronavirus outbreaks.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 November 2015; accepted 13 January 2016.

Published online 8 February 2016.

- Du, L. *et al.* The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nature Rev. Microbiol.* **7**, 226–236 (2009).
- McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
- Yin, H. S., Wen, X., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**, 38–44 (2006).
- Coleman, C. M. & Frieman, M. B. Coronaviruses: important emerging human pathogens. *J. Virol.* **88**, 5209–5212 (2014).
- 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).
- Burkard, C. *et al.* Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* **10**, e1004502 (2014).
- Xu, Y. *et al.* Structural basis for coronavirus-mediated membrane fusion. Crystal structure of mouse hepatitis virus spike protein fusion core. *J. Biol. Chem.* **279**, 30514–30522 (2004).
- Duquerry, S., Vigouroux, A., Rottier, P. J., Rey, F. A. & Bosch, B. J. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* **335**, 276–285 (2005).
- Gao, J. *et al.* Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. *J. Virol.* **87**, 13134–13140 (2013).
- Supekar, V. M. *et al.* Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc. Natl Acad. Sci. USA* **101**, 17958–17963 (2004).
- Lu, G. *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).

- 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).
- 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).
- Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl Acad. Sci. USA* **106**, 19970–19974 (2009).
- Beniac, D. R., Andonov, A., Grudski, E. & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nature Struct. Mol. Biol.* **13**, 751–752 (2006).
- Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
- Peng, G. *et al.* Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **287**, 41931–41938 (2012).
- Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
- DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* **12**, 361–365 (2015).
- Wang, R. Y. *et al.* De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods* **12**, 335–338 (2015).
- Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R. & Gallagher, T. M. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J. Virol.* **75**, 2792–2802 (2001).
- Taguchi, F. & Matsuyama, S. Soluble receptor potentiates receptor-independent infection by murine coronavirus. *J. Virol.* **76**, 950–958 (2002).
- Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* **83**, 7411–7421 (2009).
- Wicht, O. *et al.* Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture. *J. Virol.* **88**, 7952–7961 (2014).
- Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
- Daniel, C. *et al.* Identification of an immunodominant linear neutralization domain on the S2 portion of the murine coronavirus spike glycoprotein and evidence that it forms part of complex tridimensional structure. *J. Virol.* **67**, 1185–1194 (1993).
- Zhang, H. *et al.* Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies. *J. Virol.* **78**, 6938–6945 (2004).
- Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201–206 (2014).
- Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **16**, 328–337 (2014).
- Reguera, J. *et al.* Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* **8**, e1002859 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Number T32GM008268 (A.C.W.). Part of this research was facilitated by the National Resource for Automated Molecular Microscopy (award number GM103310), and the Hyak supercomputer system at the University of Washington. We thank W. Bartelink, B. Tummers and B. van der Kooij for assistance in cloning, protein expression and characterization. We are also grateful to B. Baron for assistance with the Thermophoresis experiments.

Author Contributions M.A.T., B.-J.B., P.J.M.R., F.A.R. and D.V. designed the experiments. B.-J.B. and P.J.M.R. designed and cloned the protein constructs. B.-J.B. and M.A.T. carried out protein expression, purification and biophysical characterization. D.V. performed cryoEM sample preparation and data collection. A.C.W. and D.V. processed the cryoEM data. A.C.W., B.F., F.D. and D.V. built the atomic model. A.C.W., M.A.T., B.F., B.-J.B., F.D., F.A.R. and D.V. analysed the data. A.C.W., F.A.R. and D.V. prepared the manuscript, with input from all authors.

Author Information The cryoEM map and the atomic model have been deposited in the Electron Microscopy Data Bank (EMD) and the Protein Data Bank (PDB) under accession codes EMD-6526 and 3JCL, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.V. (dveesler@uw.edu) or F.A.R. (felix.rey@pasteur.fr).

METHODS

No statistical methods were used to predetermine sample size.

Plasmids. A human codon-optimized gene encoding the MHV spike gene (UniProt: P11224) was synthesized with an Arg717Ser amino acid mutation to abolish the furin cleavage site at the S₁-S₂ junction (S₂ cleavage site). From this gene, the fragment encoding the MHV ectodomain (residues 15–1231) was PCR-amplified and ligated to a gene fragment encoding a GCN4 trimerization motif (IKRMKQIEDKIEEIESKQKKEIENIARIKKIK)^{3,31}, a thrombin cleavage site (LVPRGSLE), an 8-residue long Strep-Tag (WSHPQFEK) and a stop codon. This construct results in fusing the GCN4 trimerization motif in register with the HR2 helix at the C-terminal end of the MHV S-encoding sequence. This gene was cloned into the pMT/BiP/V5/His expression vector (Invitrogen) in frame with the *Drosophila* BiP secretion signal downstream the metallothionein promoter. The D1 domain of mouse CEACAM1a (residues 35–142; gb NP_001034274.1) was amplified by PCR and cloned into a mammalian expression plasmid³², in frame with a CD5 signal sequence at the 5' end, and with a sequence encoding a thrombin cleavage site, a glycine linker and the Fc domain of human IgG1 at the 3' end, creating the pCD5-MHVR-T-Fc vector.

Production of recombinant CEACAM1a ectodomain by transient transfection.

293-F cells were grown in suspension using FreeStyle 293 Expression Medium (Life technologies) at 37 °C in a humidified 5% CO₂ incubator on a Celltron shaker platform (Infors HT) rotating at 130 r.p.m. (for 1 l culture flasks). Twenty-four hours before transfection, cell density was adjusted at 1.5×10^6 cells ml⁻¹, and culture grown overnight in the same conditions as mentioned above to reach $\sim 2.5 \times 10^6$ cells ml⁻¹ the day of transfection. Cells were collected by centrifugation at 1,250 r.p.m. for 5 min, and resuspended in fresh FreeStyle 293 Expression Medium (Life technologies) without antibiotics at a density of 2.5×10^6 cells ml⁻¹.

To produce recombinant CEACAM1a ectodomain, 400 µg of pCD5-MHVR-T-Fc vector (purified using EndoFree plasmid kit from Qiagen) were added to 200 ml of suspension cells. The cultures were swirled for 5 min on shaker in the culture incubator before adding 9 µg ml⁻¹ of Linear polyethylenimine (PEI) solution (25 kDa, Polysciences). Twenty-four hours after transfection, cells were diluted 1:1 with FreeStyle 293 Expression Medium and the transfected cells were cultivated for 6 days. Clarified cell supernatants were concentrated tenfold using Vivaflow tangential filtration cassettes (Sartorius, 10-kDa cut-off) before affinity purification using a Protein A column (GE LifeSciences) followed by gel filtration chromatography using a Superdex 200 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The Fc tag was removed by trypsin cleavage in a reaction mixture containing 7 mg of recombinant CEACAM1a ectodomain and 5 µg of trypsin in 100 mM Tris-HCl, pH 8.0 and 20 mM CaCl₂. The reaction mixture was incubated at 25 °C overnight and re-loaded in a Protein A column to remove uncleaved protein and the Fc tag. The cleaved protein was further purified by gel filtration using a Superdex 75 column 10/300 GL (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The purified protein was quantified using absorption at 280 nm and concentrated to approximately 10 mg ml⁻¹.

Production of recombinant MHV S ectodomain in *Drosophila* S2 cells. To generate a stable *Drosophila* S2 cell line expressing recombinant MHV spike ectodomain, we used Effectene (Qiagen) and 2 µg of the plasmid encoding the MHV S protein ectodomain. A second plasmid, encoding blasticidin S deaminase was cotransfected as dominant selectable marker. Stable MHV S ectodomain expressing cell lines were selected by addition of 10 µg ml⁻¹ blasticidin S (Invivogen) to the culture medium 48 h after transfection.

For large-scale production of MHV S ectodomain the cells were cultured in spinner flasks and induced by 5 µM CdCl₂ at a density of approximately 10⁷ cells per ml. After a week at 28 °C, clarified cell supernatants were concentrated 40-fold using Vivaflow tangential filtration cassettes (Sartorius, 10-kDa cut-off) and adjusted to pH 8.0, before affinity purification using StrepTactin Superflow column (IBA) followed by gel filtration chromatography using Superose 6 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The purified protein was quantified using absorption at 280 nm and concentrated to approximately 4 mg ml⁻¹.

SEC-MALS. For size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis, samples (0.2 ml at 1 mg ml⁻¹) were loaded onto a Superdex 200 10/300 GL column (GE Life Sciences, 0.4 ml min⁻¹ in gel filtration buffer) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer coupled to a Wyatt Optilab TrEX differential refractive index detector. Data were analysed using Astra 6 software (Wyatt Technology Corp).

MicroScale Thermophoresis. Solution MicroScale Thermophoresis (MST) binding studies were performed using standard protocols on a Monolith NT.115 (Nanotemper Technologies). In brief, recombinant CEACAM1a ectodomain protein was labelled using the RED-NHS (Amine Reactive) Protein Labelling Kit (Nanotemper Technologies). The MHV S ectodomain protein was serially

diluted in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and the labelled recombinant CEACAM1a was added to a final concentration of 500 nM before overnight incubation at 4 °C. The CEACAM1a concentration was chosen such that the observed fluorescence was approximately 1,000 U at 40% LED power. The samples were loaded into standard-treated Monolith capillaries and were measured by standard protocols using a Monolith NT.115, NanoTemper. The changes in the fluorescent thermophoresis signal were plotted against the concentration of the serially diluted MHV spike protein, and K_d values were determined using the NanoTemper analysis software.

CryoEM sample preparation and data collection. Three microlitres of MHV spike at 1.85 mg ml⁻¹ was applied to a 1.2/1.3 C-flat grid (Protochips), which had been glow-discharged for 30 s at 20 mA. Thereafter, grids were plunge-frozen in liquid ethane using a Gatan CP3 and a blotting time of 3.5 s. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector. Coma-free alignment was performed using the Leginon software³³. Automated data collection was carried out using Leginon³⁴ to control both the FEI Titan Krios (used in microprobe mode at a nominal magnification of 22,500×) and the Gatan K2 Summit operated in counted mode (pixel size: 1.315 Å) at a dose rate of ~ 9 counts per physical pixel per s, which corresponds to ~ 12 electrons per physical pixels per s (when accounting for coincidence loss³⁵). Each video had a total accumulated exposure of 53 e Å⁻² fractionated in 38 frames of 200 ms (yielding movies of 7.6 s). A data set of $\sim 1,600$ micrographs was acquired in a single session using a defocus range of between 2.0 and 5.0 µm.

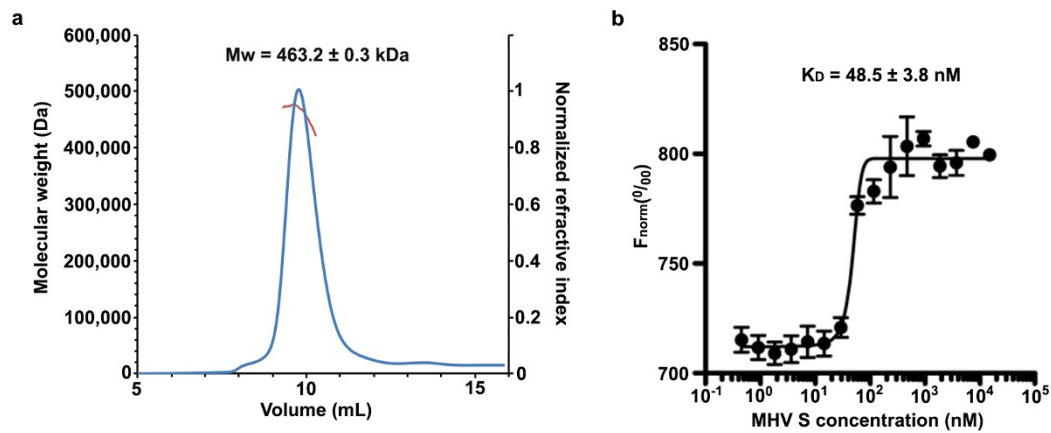
CryoEM data processing. Whole-frame alignment was carried out using the software developed previously³⁵, which is integrated into the Appion pipeline³⁶, to account for stage drift and beam-induced motion. The parameters of the microscope contrast transfer function were estimated for each micrograph using ctffind3 (ref. 37). Micrographs were manually masked using Appion to exclude the visible carbon supporting film for further processing. Particles were automatically picked in a reference-free manner using DogPicker³⁸. Extraction of particle images was performed using Relion 1.4 with a box size of 320 pixels² and applying a windowing operation in Fourier space to yield a final box size of 288 pixels² (corresponding to a pixel size of 1.46 Å). From the 1.2 million particles initially picked, a subset of 50,000 particles were randomly selected to generate class averages using RELION³⁹. An initial 3D model was generated using OPTIMOD⁴⁰ within the Appion pipeline. The entire data set was subjected to 2D alignment and clustering using RELION and particles belonging to the best-defined class averages were retained ($\sim 500,000$ particles). These $\sim 500,000$ particles were then subjected to RELION 3D classification with four classes (using c1 symmetry) starting with our initial model low-pass filtered to 40 Å resolution. We subsequently used the $\sim 230,000$ best particles (selected from the 3D classification) and the map corresponding to the best 3D class (low-pass filtered at 40 Å resolution) to run Relion 3D auto-refine (c3 symmetry), which led to a reconstruction at 4.4 Å resolution. We used the particle polishing procedure in RELION 1.4 to correct for individual particle movement and radiation damage^{41,42}. A second round of 3D classification with 6 classes (c3 symmetry) was performed using the polished particles resulting in the selection of 82,000 particles. A new 3D auto-refine run (c3 symmetry) using the selected 82,000 particles and the map corresponding to the best 3D class (low-pass filtered at 40 Å resolution) yielded a map at 4.0 Å resolution following post-processing in RELION. The final map was sharpened with an empirically determined B factor of -220 Å² using Relion post processing. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion⁴³, and Fourier shell correction curves were corrected for the effects of soft masking by high-resolution noise substitution⁴⁴. The soft mask used for FSC calculation had a 10 pixel cosine edge fall-off. The overall shape and dimensions of our reconstruction agree with previous data although the HR2 stem connecting to the membrane is not resolved¹⁵.

Model building and analysis. Fitting of atomic models into cryoEM maps was performed using UCSF Chimera⁴⁵ and Coot^{18,46}. We initially docked the MHV domain A structure (PDB 3R4D) and used a crystal structure of a bovine coronavirus domain A (PDB 4H14) to model the three-stranded β-sheet and the α-helix present on the viral membrane proximal side of the galectin-like domain. Next, the MERS-CoV domain B crystal structure (PDB 4KQZ) was also fit into the density, and rebuilt and refined using RosettaCM⁴⁷. Although we could accurately align the sequences corresponding to the core β-sheet of the MHV and MERS-CoV B domains, the ~ 100 residues forming the β-motif extension (residues 453–535, MERS-CoV/SARS-CoV receptor-binding moiety) could not be aligned with confidence. We used RosettaCM to build models of each of the 945 possible disulfide patterns into the density for domain B. For each disulfide arrangement, 50 models were generated, and there was a very clear energy signal for a single such arrangement (Extended Data Fig. 3k). Then, 1,000 models with this disulfide arrangement were sampled, and the lowest energy model (using the Rosetta force field

augmented with a fit-to-density score term) was selected. Owing to the poor quality of the reconstruction at the apex of the S trimer, the confidence of the model is lowest for the segment corresponding to residues 453–535, as homology modelling was used to fill in details missing in the map.

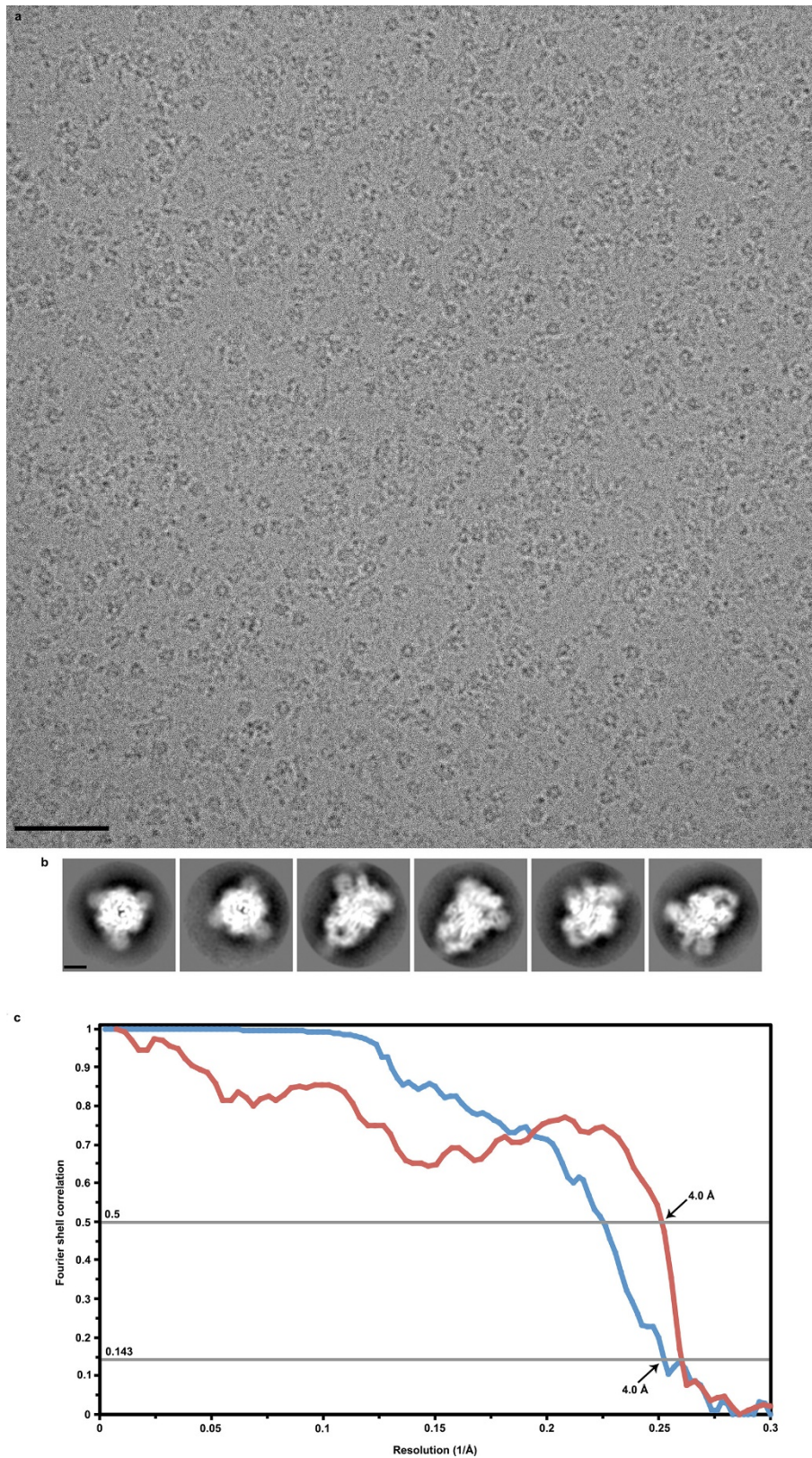
A backbone model was then manually built for the rest of the S polypeptide using Coot. Sequence register was assigned by visual inspection where side chain density was clearly visible. This initial hand built model was used as an initial model for Rosetta *de novo*²⁰. The Rosetta-derived model largely agreed with the hand-built model. Rosetta *de novo* successfully identified fragments allowing to anchor the sequence register for domains C and D as well as for helices α_{21} – α_{25} . Given these anchoring positions, RosettaCM⁴⁷ augmented with a novel density-guided model-growing protocol was able to rebuild domains C and D in full. The final model was refined by applying strict non-crystallographic symmetry constraints using Rosetta¹⁹. Model refinement was performed using a training map corresponding to one of the two maps generated by the gold-standard refinement procedure in Relion. The second map (testing map) was used only for calculation of the FSC compared to the atomic model and preventing overfitting⁴⁸. The quality of the final model was analysed with Molprobity⁴⁹. Structure analysis was assisted by the PISA⁵⁰ and DALI⁵¹ servers. The sequence alignment was generated using MultAlin⁵² and coloured with ESPript⁵³. All figures were generated with UCSF Chimera⁴⁵.

31. Eckert, D. M., Malashkevich, V. N. & Kim, P. S. Crystal structure of GCN4-pIQ1, a trimeric coiled coil with buried polar residues. *J. Mol. Biol.* **284**, 859–865 (1998).
32. 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).
33. Glaeser, R. M., Typke, D., Tiemeijer, P. C., Pulokas, J. & Cheng, A. Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. *J. Struct. Biol.* **174**, 1–10 (2011).
34. Suloway, C. *et al.* Automated molecular microscopy: the new Legimin system. *J. Struct. Biol.* **151**, 41–60 (2005).
35. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
36. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
37. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
38. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Lyumkis, D., Vinterbo, S., Potter, C. S. & Carragher, B. Optimod—an automated approach for constructing and optimizing initial models for single-particle electron microscopy. *J. Struct. Biol.* **184**, 417–426 (2013).
41. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2**, e00461 (2013).
42. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
43. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
44. 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).
45. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
47. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
48. DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* **22**, 865–868 (2013).
49. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
50. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
51. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
52. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).
53. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).



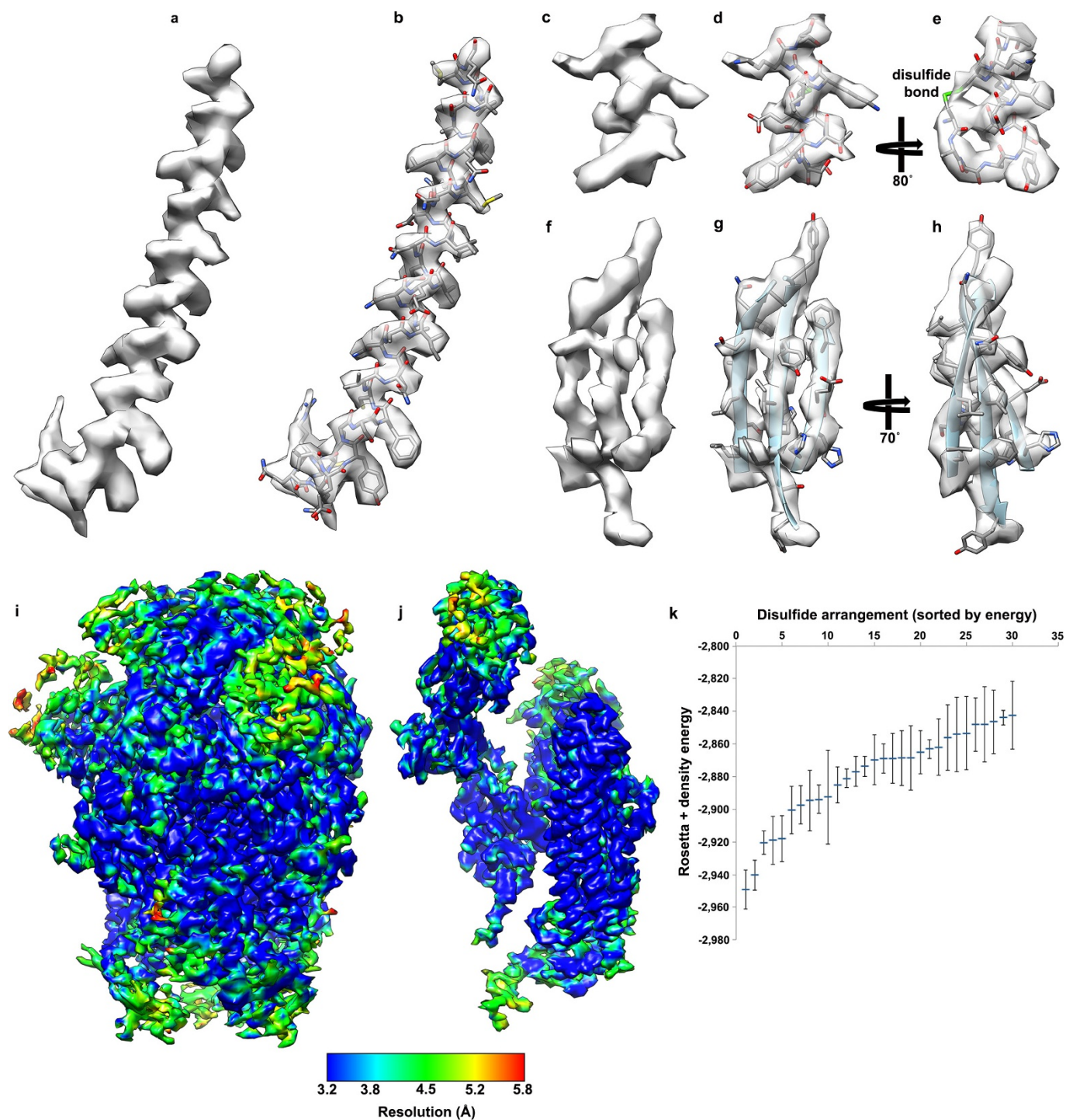
Extended Data Figure 1 | Biophysical characterization of the MHV S ectodomain. **a**, The MHV S molecular mass was determined to be 463.2 ± 0.3 kDa (mean \pm s.e.m.) (corresponding to a trimer) using size-exclusion chromatography coupled in-line with multi-angle light scattering and refractometry. The blue line represents the normalized refractive index (right ordinate axis) and the red line shows the estimated

molecular mass (expressed in Da, left ordinate axis). **b**, MHV S binds with high-affinity to the soluble mouse CEACAM1a receptor. Thermophoresis signal plotted against the MHV S concentration. The dissociation constant (K_D) was determined to be 48.5 ± 3.8 nM. Values correspond to the average of two independent experiments. The concentration of CEACAM1a used was 500 nM.



Extended Data Figure 2 | CryoEM analysis of the MHV S trimer.
a, b, Representative electron micrograph (defocus: $4.6\ \mu\text{m}$) (**a**) and class averages (**b**) of the MHV S trimer embedded in vitreous ice. Scale bars: $573\ \text{\AA}$ (micrograph) and $44\ \text{\AA}$ (class averages). **c,** Gold-standard (blue) and

model/map (red) Fourier shell correlation (FSC) curves. The resolution was determined to $4.0\ \text{\AA}$. The 0.143 and 0.5 cut-off values are indicated by horizontal grey bars.

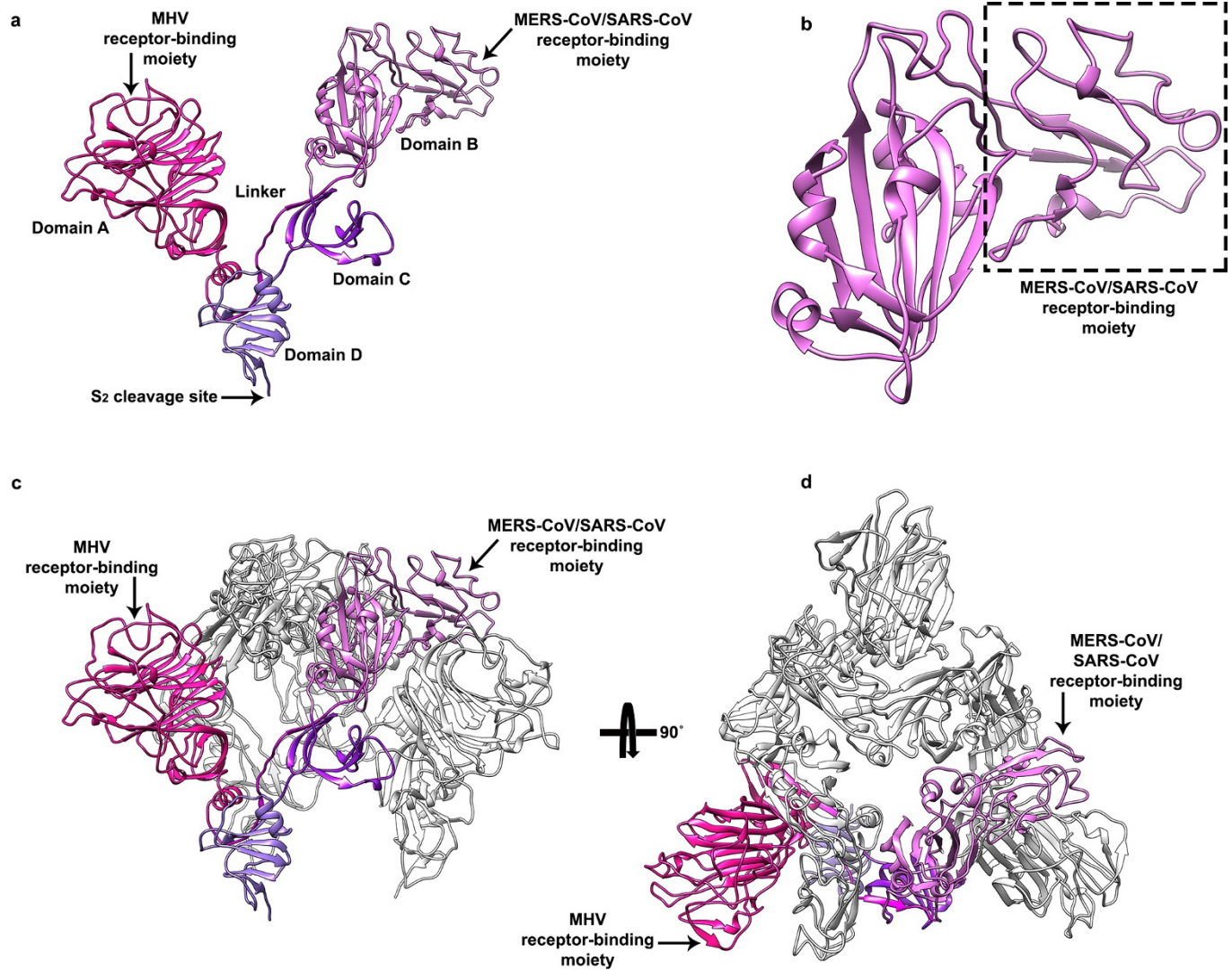


Extended Data Figure 3 | CryoEM density for selected regions of the MHV S reconstruction, local resolution analysis and density-guided homology modelling of residues 453–535. The atomic model is shown with the corresponding region of the map. **a, b**, Upstream helix. **c–e**, Helix belonging to domain A (residues 284–296). **f–h**, Core β -sheet. **i, j**, CryoEM density corresponding to the MHV S trimer (**i**) and a single

protomer (**j**), coloured according to local resolution determined with the software Resmap. We interpret Resmap results as a qualitative (rather than quantitative) estimate of map quality. **k**, Rebuilding of the MHV S domain B using RosettaCM. Plot showing the energy mean and s.d. of the models corresponding to the 30 lowest energy disulfide arrangements (out of 945) for domain B.

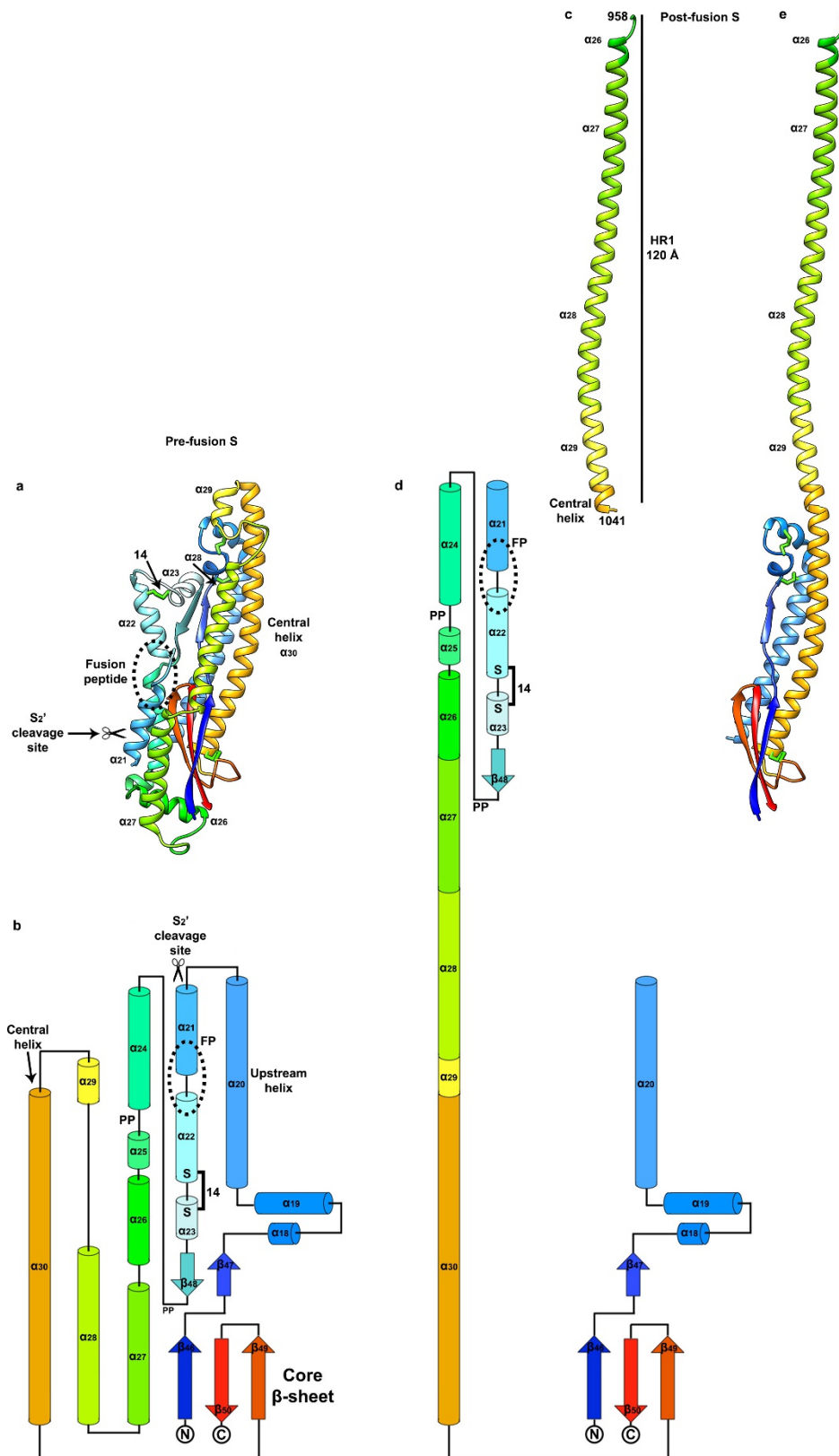
Data collection	
Number of particles	82,000
Pixel size (Å)	1.315 (rescaled to 1.46)
Defocus range (µm)	2-5
Voltage (kV)	300
Electron dose (e ⁻ /Å ²)	53
Refinement	
Resolution	4.0
Map sharpening B factor (Å ²)	-220
Model validation	
Molprobrity score (percentile)	1.55 (94 th)
All-atom clashscore (percentile)	3.68 (97 th)
Poor rotamers (%)	0
Ramachandran favored (%)	94.26
Ramachandran allowed (%)	99.44
Ramachandran outliers (%)	0.56
r.m.s.d bonds (Å)	0.018
r.m.s.d angles (°)	1.8

Extended Data Figure 4 | Refinement and model statistics.



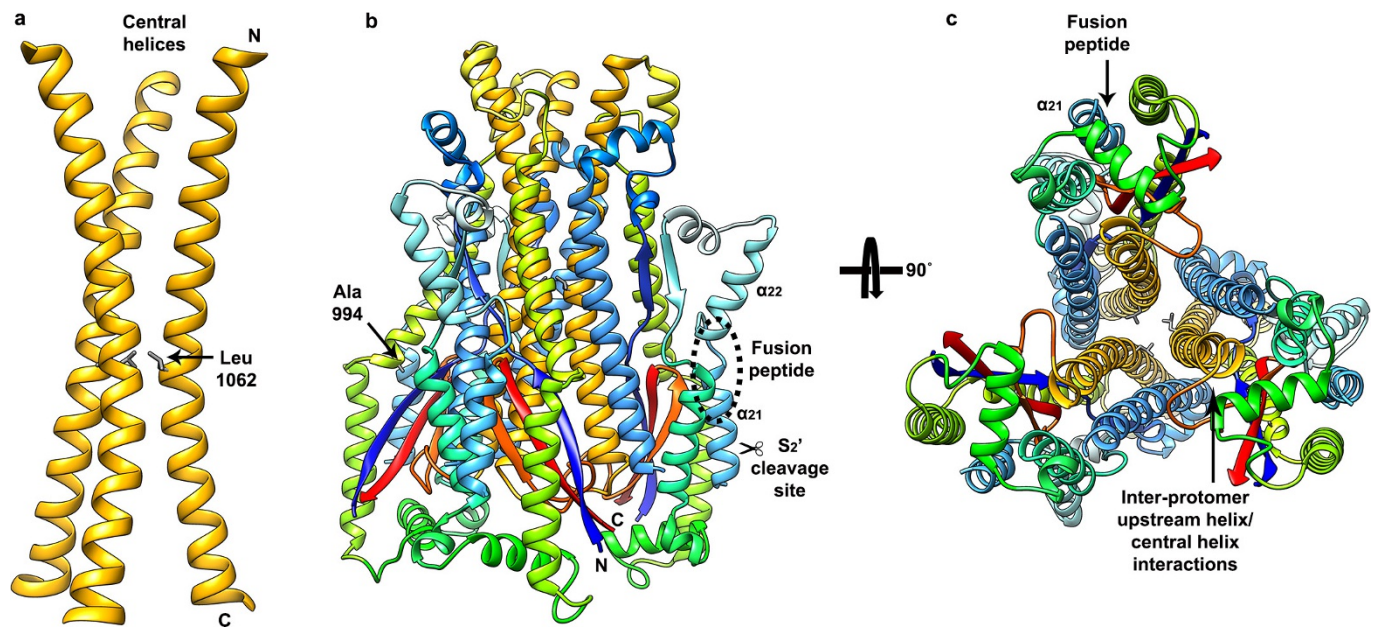
Extended Data Figure 5 | Structural organization of the S₁ subunit. **a**, Ribbon diagram showing a single S₁ protomer. **b**, Close-up view of the MHV S domain B. The structural motif used as a receptor-interacting moiety by MERS-CoV and SARS-CoV is indicated. The density was too

weak to allow tracing of this segment (residues 453–535), which has been traced by density-guided homology modelling using Rosetta. **c**, **d**, Ribbon diagrams of the S₁ trimer viewed from the side (**c**) and from the top (looking towards the viral membrane) (**d**).

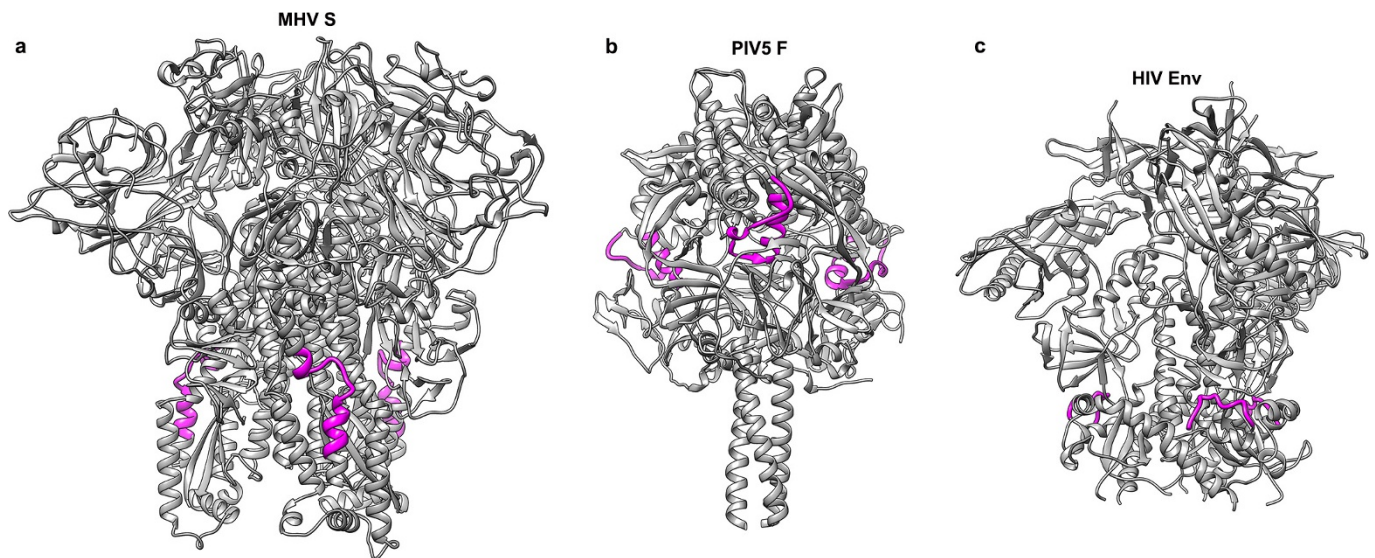


Extended Data Figure 6 | Mechanisms of membrane fusion promoted by coronavirus S glycoproteins. **a**, Ribbon diagram of the MHV S_2 pre-fusion structure. Disulfide bonds are shown as green sticks. **b**, Topology diagram of the MHV S_2 pre-fusion structure. PP, di-proline that will act as a helix breaker. The presence of these di-proline motifs indicates that the post-fusion HR1 coiled-coil could not extend up to the fusion peptide as a single helix. This hypothesis is further supported by the observation of a conserved disulfide bond formed between residues Cys894 and Cys905 (labelled 14 in **a** and **b**), which will prevent refolding of helices α_{22} and

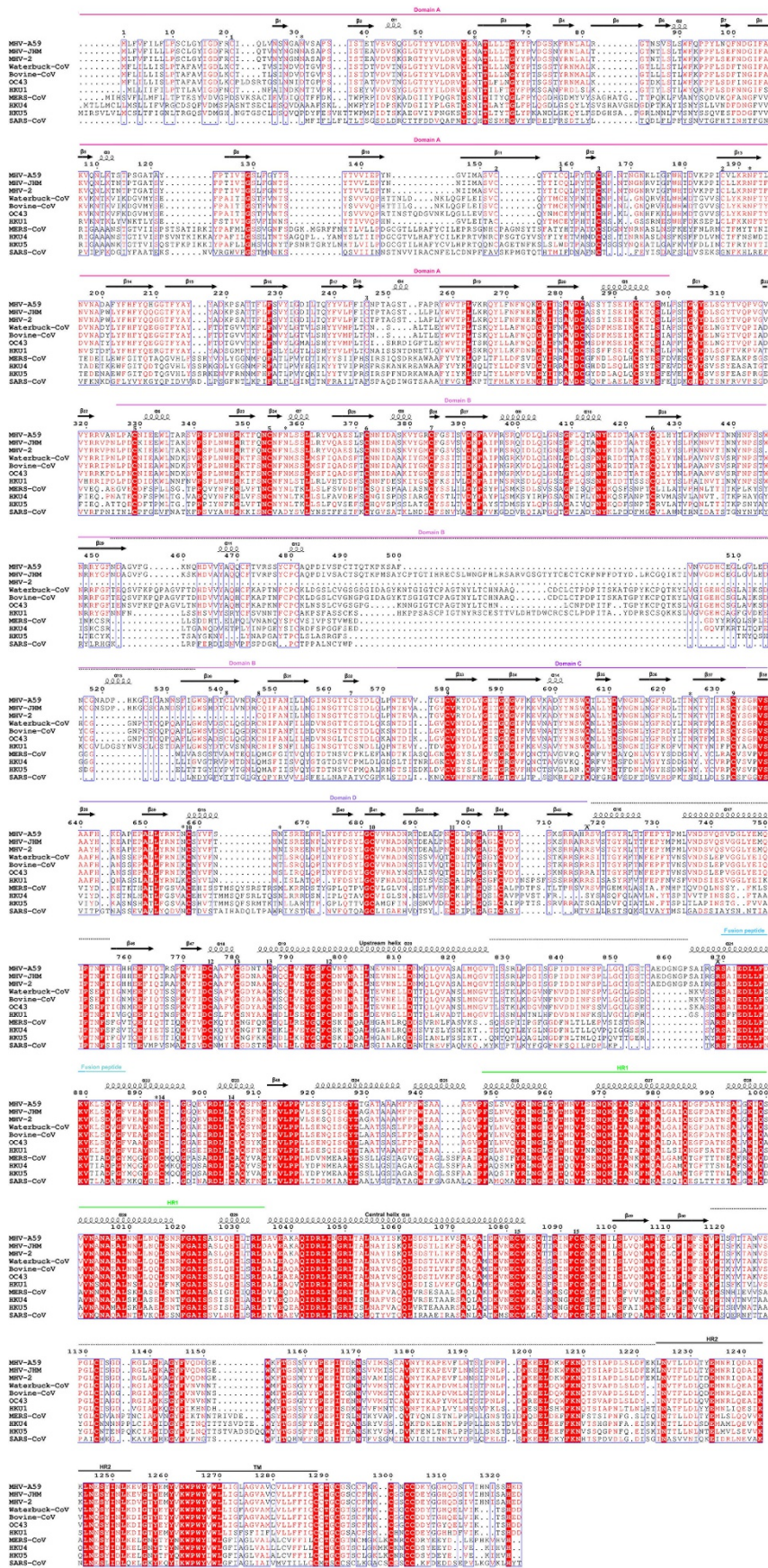
α_{23} as a single extended helix. **c**, Ribbon diagram of the SARS-CoV post-fusion HR1 helix obtained by X-ray crystallography (PDB 1WYY). The residue numbers corresponding to the MHV A59 sequence are indicated. **d**, Topology diagram showing the expected coronavirus S post-fusion conformation derived from our MHV S structure and the SARS-CoV post-fusion core crystal structure shown in **c**. **e**, Ribbon diagram of a model of the MHV S_2 post-fusion conformation. Residues belonging to α_{21} , α_{22} , α_{23} , β_{48} , α_{24} and α_{25} are not represented owing to a lack of structural information.



Extended Data Figure 7 | Structural organization of the S₂ fusion machinery. **a**, Ribbon diagram of the trimer of central helices. **b**, **c**, Ribbon diagrams of the S₂ trimer (starting at residue 755) viewed from the side (**b**) and from the bottom (looking towards the host cell membrane) (**c**). Residues Ala994 and Leu1062, which are discussed in the text, are shown in stick format.



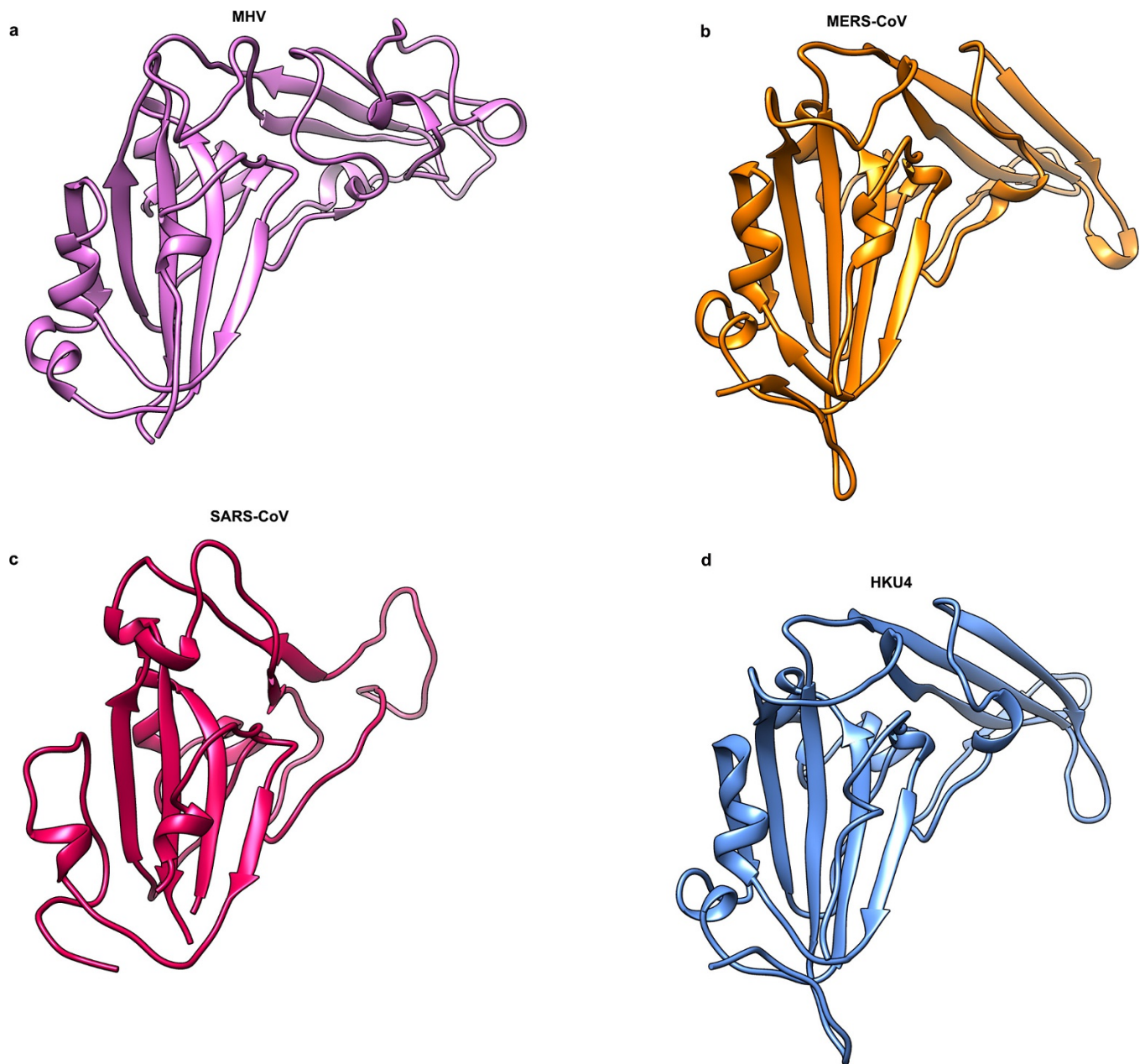
Extended Data Figure 8 | Class I viral fusion proteins with exposed fusion peptide. **a**, MHV S (residues 870–887). **b**, Parainfluenza virus 5 F (PIV5 F, residues 103–128, PDB 2B9B). **c**, HIV-1 gp41 (residues 518–528, PDB 4TVP). The trimeric fusion proteins are shown as grey ribbon diagrams with the fusion peptides rendered in magenta.



Extended Data Figure 9 | See next page for figure caption.

Extended Data Figure 9 | Sequence conservation among coronavirus S glycoproteins. a, Sequence alignment of coronavirus S proteins. Bovine-CoV, bovine respiratory coronavirus AH187 (gi 253756585); HKU1, human coronavirus HKU1 (gi 545299280); HKU4, tyloncteris bat coronavirus HKU4 (gi 126030114); HKU5, pipistrellus bat coronavirus HKU5 (gi 126030124); MERS-CoV, Middle East respiratory syndrome coronavirus (gi 836600681); MHV-A59, mouse hepatitis virus A59 (gi 1352862); MHV-JHM, mouse hepatitis virus JHM (gi 60115395); MHV-2, mouse hepatitis virus 2 (gi 5565844); OC43, human coronavirus OC43 (gi 744516696); SARS-CoV, severe acute respiratory syndrome coronavirus ZJ01 (gi 39980889); Waterbuck-CoV, waterbuck coronavirus

US/OH-WD358-TC/1994 (gi 215478096). Asparagine residues featuring N-linked glycan chains visible in the MHV S reconstruction are indicated with a star. The S_2 and S_2' cleavage sites are indicated with scissors at positions corresponding to the MHV S sequence. Cysteine residues involved in the formation of disulfide bonds are numbered according to Supplementary Table 2. The secondary structure elements observed in our MHV S reconstruction are indicated above the sequence. The black dotted lines above the sequence indicate regions poorly defined in the density. Although the viral membrane distal loops of the A domains are weakly defined in the density, the availability of a crystal structure of this domain from the same virus (PDB 3R4D) helped with the modelling.



Extended Data Figure 10 | Structural similarity of B domains among coronavirus S glycoproteins. a, MHV (pink). b, MERS-CoV (orange, PDB 4KQZ). c, SARS-CoV (red, PDB 2AJF). d, HKU4 (blue, PDB 4QZV).